

https://dspace.org/

Insights into the Physiochemical Properties of the Interaction between the L38NL+4 HIV-1 Hinge Mutant Subtype C Protease and a Related Gag Cleavage Site

Adams, Taryn Racheal

2023-09

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/41969

Adams, Taryn Racheal. (2023). Insights into the Physiochemical Properties of the Interaction between the L38NL+4 HIV-1 Hinge Mutant Subtype C Protease and a Related Gag Cleavage Site. [Master's dissertation, University of the Witwatersrand, Johannesburg]. https://hdl.handle.net/10539/41969

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



Insights into the Physiochemical Properties of the Interaction between the L38↑N↑L+4 HIV-1 Hinge Mutant Subtype C Protease and a Related Gag Cleavage Site

by

Taryn Adams

(1615069)

MSc Dissertation

Submitted in fulfilment of the requirements for the degree

Masters of Science

in

Molecular and Cell Biology

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

Supervisor: Prof. Yasien Sayed

September 2023

Declaration

I, Taryn Adams (student number: 1615069), am a student registered for the degree of Master of Science at The University of the Witwatersrand in the academic year 2023.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where 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: 13 / 09 / 2023

Abstract

Frequent mutations in HIV protein drug targets such as the protease (PR), have led to a rise in resistance to clinically available treatments and inquiries into the associated biochemical mechanisms. In this study, the L38↑N↑L^{+4} PR (mutant) and related gag, were isolated from a PR inhibitor naïve child for further study. Wild-type and mutant PRs were successfully overexpressed and purified using ion exchange chromatography. Intrinsic fluorescence studies probing tryptophan residues located at the hinge revealed variations in tertiary structure. This coincided with significant differences in refolding efficiency (mutant PR recovery ~ 10% and wild-type PR recovery ~ 34%). Furthermore, differences in catalytic efficiency (mutant-specific activity ~ 7.4 µmol.min^{-1}.mg^{-1}, mutant k_{cat} ~ 2.71 sec^{-1}; wild-type specific activity ~ 31.6 µmol.min^{-1}.mg^{-1}, k_{cat} ~ 11.60 sec^{-1}). Thermal shift assays revealed reduced mutant PR structural stability (T_m ~ 67 ℃) compared to the wild-type PR (T_m ~ 64 ℃). Furthermore, reduced stability of inhibitor-mutant PR complexes (T_m ~ 73.5 ℃ Acetyl pepstatin (AP), 90 ℃ Darunavir (DRV), and 88 ℃ Saquinavir (SQV)) compared to complexes involving the wild-type PR (T_m ~ 70 ℃ (AP), 82 ℃ (DRV), and 82 ℃ (SQV)). Overall, the L38↑N↑L^{+4} PR mutations were found to influence the tertiary structure of the hinge, gag processing, and PI stability within the mutant PR active site. Computational docking studies highlighted the potential role of a gag single nucleotide polymorphism (SNP), located at the 4^{th} amino acid (P4) position of the peptide-2-nucleocapsid (p2/NC) gag cleavage site, for future studies as a compensatory mutation aiding PR polymorphisms. Further studies should focus on the gag-PR functional pair to build a more accurate understanding of HIV drug resistance mechanisms.

I would like to dedicate this work to my family members, my mother, Lauren Adams, and my sister, Jade Adams, who have supported me unconditionally through my MSc journey.

I would also like to thank Colin Kinghorn and his family members (Anthony Kinghorn, Lesley Kinghorn, and Julia Kinghorn) for their generosity and support as a second family to me.

Acknowledgments

I would like to express my gratitude to Prof. Y. Sayed for supporting my undertaking of this project.

Thank you to Dr Mia Færch for her continuous advice and assistance as well as Dr. Maya Makatini and Dr Peme for their assistance in peptide design and synthesis. I would also like to thank Dr Sheik Ismail, Dr Lakhi, and Dr Donald for their mentorship within the laboratory.

I am also grateful to the University of the Witwatersrand for the financial assistance, enabling me to continue my studies further. Furthermore, funding from the National Research Foundation (NRF) which make research like this possible

Contents

List of Abbreviations 10

List of Figures 8

CHAPTER 1. INTRODUCTION 11

1.1 HIV epidemic: advancements and prospective concerns 11

1.2 HIV Replication cycle and drug targets 12

1.3 Structural and functional characteristics of the HIV PR and gag 14

1.3.1 PR structure and function 14

1.3.2 Gag Structure and Function 16

1.3.3 PR-Gag Binding Structural Elements 17

1.3 Mutations in the HIV PR and Gag 20

1.4.1 PR Polymorphisms 20

1.4.2 Gag Polymorphisms 23

1.5 Obstacles in HIV Treatment Efficacy and Prospects 26

1.5.1 FDA-Approved PR Inhibitors 26

1.5.2 Addressing HIV Drug Resistance Mechanisms 27

1.6 The L38↑N↑L^{+4} PR-Gag Pair 28

1.6.1 L38↑N↑L^{+4} PR 28

1.6.2 L38↑N↑L^{+4} Gag 28

1.7 Aim and Objectives 29

Chapter 2. Experimental Procedure 30

2.1 Heterologous overexpression of inactive and active constructs of the Subtype C wild-type and the L38↑N↑L^{+4} hinge mutant PRs 30

2.1.1 Transformation of pET-11a mutant HIV-1 PR plasmid constructs into BL21 (DƐ3) pLysS *E. coli* cells 30

2.1.2 IPTG induction trials of transformed BL21 (DƐ3) pLysS *E. coli* cells containing inactive wild-type PR, inactive mutant PR, and active mutant PR expression vectors 31

2.1.3 Heterologous overexpression of wild-type and mutant HIV-1 PRs 32

2.2 Ion-exchange chromatographic purification and chemical renaturation of HIV-1 PRs 32

2.2.1 Purification of HIV-1 PRs using a dual ion-exchange chromatographic technique and chemical renaturation through two-step dialysis 33

2.2.2 Absorbance-based concentration determination and quality assessment of HIV-1 PRs 35

2.3 Structural assessment of inactive and active constructs of the Subtype C Wild Type and the L38↑N↑L^{+4} hinge mutant PRs 36

2.3.1 Primary amino acid evaluation of DNA extracted from transformed overexpression BL21 (DƐ3) pLysS *E. coli* cell lines 36

2.3.2 Secondary structural analysis of HIV-1 PRs using far-UV CD 36

2.3.3 Tertiary structural evaluation of HIV-1 PRs by probing tryptophan and tyrosine probes using intrinsic fluorescence 37

2.3.4 Quaternary structural evaluation of HIV-1 PRs using SE-HPLC 37

2.4 Global functional assessment of inactive and active wild-type and mutant PRs 38

2.4.1 Active site titrations of purified HIV-1 PRs against acetyl-pepstatin using ITC 38

2.5 Global stability studies of active wild type and active mutant HIV-1 PRs in the absence and presence of known inhibitors using thermal shift assays 41

2.6 Design, synthesis, and purification of HIV-1 Gag p2/NC-mimicking peptides 41

2.6.1 Design and optimisation of peptide empirical properties 41

2.6.2 Synthesis and purification of HIV-1 Gag p2/NC-mimicking peptides 42

2.6.3 Evaluation of purified peptides using LC-MS and far-UV CD 44

2.7.1 Isothermal titration of a soluble HIV-1 Gag p2/NC-mimicking peptide against active wild-type and mutant HIV-1 PR 44

2.7 Evaluation of a soluble HIV-1 Gag p2/NC-mimicking peptide binding to active Subtype C 45

2.7.2 Monte Carlo-based docking simulations of the mutant HIV-1 Gag p2/NC-mimicking peptide against active wild-type and mutant HIV-1 PR 45

3.1 Comparative optimisation of heterologous overexpression of wild-type and mutant PRs in BL21 (DƐ3) pLysS *E. coli* cells 46

3.1.1 Cell viability assessment of transformed cell lines by growth curve analysis 46

3.1.2 Identification and comparison of optimal IPTG concentrations and induction periods using SDS-PAGE analysis 48

3.4 Functional assessment of inactive and active constructs of the Subtype C wild-type and the L38↑N↑L^{+4} hinge mutant PRs 73

3.4.1 Active site titrations of purified HIV-1 PRs against acetyl-pepstatin using ITC 73

3.4.2 Enzyme activity evaluation using a FRET-based activity assay 75

3.5 Stability evaluation of active wild-type and mutant HIV-1 PRs in the absence and presence of known inhibitors using a thermal shift assay 77

3.6 Generating HIV-1 Gag p2/NC-mimicking peptides for functional studies with active wild-type and active mutant PRs 81

3.6 Evaluation of the mutant-related peptide binding to active wild Type and active mutant PR 85

3.6.1 Isothermal titration of a soluble HIV-1 Gag p2/NC-mimicking peptide against active wild type and mutant HIV-1 PR 85

3.6.2 Hierarchal docking simulations of the mutant-related peptide against active wild type and mutant HIV-1 PR 85

3.7 Concluding Remarks 91

5.1 Active site titrations of purified inactive HIV-1 PRs against acetyl-pepstatin using ITC .. 99

5.2 LC-MS profiles of purified mutant p2/NC peptide from fractions two and three 100

5.2 Active site titrations of purified active HIV-1 PRs against purified mutant p2/NC-mimicking peptide using ITC 102

List of Figures

Figure 1 taken from the US office of AIDS research (OAR) ("About HIVinfo.NIH.gov | NIH," n.d.): Stages of the HIV replication cycle and corresponding drug targets……...**13**

Figure 2: HIV PR structure, gag structure, and PR-gag structural association. (A) WT C-SA PR structure. (B) Schematic of gag structure.……..**15**

Figure 3: Structural Elements of PR-gag binding……..…**19**

Figure 4: The L38↑N↑L^{+4} mutant PR and gag pair. (A) Mutations in the L38↑N↑L^{+4}PR, visualised in PyMOL, by Schrödinger. (B) Schematic of the mutations in the L38↑N↑L^{+4}gag. (C) Alignment of L38↑N↑L^{+4}PR and the WT C-SA PR visualised in the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.……...**22**

Figure 5: A schematic of the cation-anion exchange chromatography and two-step renaturation procedure required for purification of HIV-1 PR.……...……...**34**

Figure 6: A schematic of the HIV-1 PR reaction mechanism in relation to the FRET-based enzyme assay……...…**40**

Figure 7: A schematic of solid phase peptide synthesis, SEC-purification and evaluation using LC-MS.……...……..……...**43**

Figure 8: Growth curves of HIV-1 PR expression cell lines …….…...…...…**47**

Figure 9: SDS-PAGE analysis of crude extracts of PR expression under varying concentrations of IPTG and varying induction periods.…..……..….**49**

Figure 10: Chromatograms of ion exchange purification of the inactive wild-type HIV-1 PR……...**51**

Figure 11: SDS-PAGE analysis of samples taken along the purification of the inactive wild-type PR.……...……...**52**

Figure 12: Chromatograms of ion exchange purification of the inactive mutant HIV-1 PR. …..…....**53**

Figure 13: SDS-PAGE analysis of samples taken along the purification of the inactive mutant PR.……...**54**

Figure 14: Chromatograms of ion exchange purification of the active wild-type HIV-1 PR.…….....**55**

Figure 15: SDS-PAGE analysis of samples taken along the purification of the active wild-type PR.……...……..….**56**

Figure 16: Chromatograms of ion exchange purification of the active mutant HIV-1 PR.…….....**57**

Figure 17: SDS-PAGE analysis of samples taken along the purification of the inactive wild-type PR.……...…**58**

Figure 18: UV-absorbance spectra purified, working concentrations of the inactive mutant PR (A), inactive wild-type PR (B), active mutant PR (C), and active wild-type PR (D)……...**60**

Figure 19: (A) Amino acid sequences of inactive and active wild-type PRs translated from DNA sequencing of extracted plasmid DNA. (B) Alignment of amino acid sequences of inactive and active wild-type and mutant PRs. ……..……...……**62**

Figure 20: Far-UV CD spectra of the inactive mutant PR (A) and inactive wild-type PR (C) and the corresponding HT (V) plots for the inactive mutant PR (B) and inactive wild-type PR (D). …..……**64**

Figure 21: Far-UV CD spectra of the active mutant PR (A) and active wild-type PR (C) and the corresponding HT (V) plots for the active mutant PR (B) and active wild-type PR (D)……..……**65**

Figure 22: Crystal Structure of the wild-type HIV-1 PR representing B-factors of probes (A), B-factors across the PR structure (B), and hydrophobicity of probes (C) and hydrophobicity across the PR structure (D)…….……**67**

Figure 23: Intrinsic fluorescence spectra of inactive and active wild-type and mutant HIV-1 PRs….**68**

Figure 24: SE-HPLC profiles of the inactive mutant PR with protein standards (A) and in isolation (B) in comparison to the SE-HPLC of the inactive wild-type PR with protein standards (C) and in isolation (D). A standard curve was generated from the SE-HPLC of protein standards (E)……..….**70**

Figure 25: SE-HPLC profiles of the active mutant PR with protein standards (A) and in isolation (B) in comparison the SE-HPLC of the active wild-type PR with protein standards (C) and in isolation (D). A standard curve generated from the SE-HPLC of protein standards (E)……..…..……..**72**

Figure 26: The fit of an independent binding model to the raw thermal data from active site titrations of AP against the active wild-type PR (A) and active mutant PR (B)……...**74**

Figure 27: Determination of the specific activity of the active wild-type (A) and active mutant (C) PR as well as the k_{cat} of active wild-type (B) and mutant (D) PRs under steady-state kinetics. (E) Catalytic activity assessment of inactive PRs……...**76**

Figure 28: Thermal shift assays of the active mutant PR at varying concentrations of AP (A), DRV (B), and SQV (C)……...**78**

Figure 29: Thermal shift assays of the active wild-type PR at varying concentrations of AP (A), DRV (B), and SQV (C).……..**79**

Figure 30: Predicted empirical properties of wild-type and mutant p2/NC peptides in comparison to the FRET-based assay peptide. ……....**82**

Figure 31: LC-MS profiles of purified mutant p2/NC peptide. ……..…..…**83**

Figure 32: (A) Far-UV CD spectra of purified mutant p2/NC peptide and the corresponding HT (V) plot (B).……..**84**

Figure 33: (A) Conserved polar contacts across multiple subtype B PR-gag peptide crystal complexes (Prabu-Jeyabalan *et al.,* 2002). (B) Alignment of predicted Subtype B PR-peptide complex and the corresponding crystal structure.……...……..…**86**

Figure 34: Binding energy scores (kcal/mol) generated using molecular docking simulations in HPEPDOCK. ……**88**

Figure 35: HPEPDOCK model of the complexes of the wild-type PR (green) and mutant PR (purple) in complex with the synthetic mutant peptide synthesised and utilised in ITC binding studies……..……...**89**

Supplementary Figure 1: Raw thermal data of active site titrations involving AP against inactive wild-type PR (A) and inactive mutant PR (B)…….……..**99**

Supplementary Figure 2: LC-MS profile of purified mutant p2/NC peptide……...**100**

Supplementary Figure 3: LC-MS profile of purified mutant p2/NC peptide……..….**101**

Supplementary Figure 3: Raw thermal data of active site titrations involving the mutant p2/NC peptide against active wild-type PR (A) and active mutant PR (B)…….**102**

List of Abbreviations

AIDS: Acquired immunodeficiency syndrome

Arg: Arginine

Asp: Aspartate

ARVs: Antiretrovirals

CA: Capsid Protein

CD: Circular Dichroism

CM: Carboxymethyl

C-SA: South African Subtype C

CSM: Cleavage site mutation

DCM: Dichloromethane

DIPEA: *N*,*N*-Diisopropylethylamine

DMF: Dimethylformamide

DMSO: Dimethyl Sulfoxide

DRV: Darunavir

dsDNA: Double-stranded DNA

DTG: Aspartate-Threonine-Glycine

Far-UV CD: Far-ultraviolet circular dichroism

FDA: Food and Drug Administration

FRET: Fluorescence resonance electron transfer

Gag: Gag polyprotein

HAART: Highly active antiretroviral therapy

HBTU: Hexafluorophosphate benzotriazole tetramethyl uranium

HIV-1: Human immunodeficiency virus

IEX: Ion exchange chromatography

IN: Integrase

IPTG: Isopropyl thiogalactopyranoside

ITC: Isothermal Titration Calorimetry

LC-MS: Liquid chromatography mass spectrometry

LPV: Lopinavir

MA: Matrix Protein

MD: Molecular Dynamics

NC: Nucleocapsid

NCSM: Non-cleavage site mutation

NOP: Naturally Occurring Polymorphism

NMR: Nuclear Magnetic Resonance

OD: Optical density

OPLS3: Optimised potentials for liquid simulations force field

p6: The Late Domain

PDB: Protein data bank

PIs: PR inhibitors

PR: Protease

RMSD: Root mean square deviation

RT: Reverse Transcriptase

RTV: Ritonavir

SDS-PAGE: Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

SE-HPLC: Size exclusion-high performance liquid chromatography

SIV: Simian immunodeficiency virus

SOC: Super Optimal Broth with Catabolite Repression

SP1: Spacer Peptide 1

SP2: Spacer Peptide 2

SQV: Saquinavir

TFA: Trifluoroacetic acid

TIS: Triisopropylsilane

UV: Ultraviolet radiation

WHO: World Health Organisation

WT: Wild type

CHAPTER 1. INTRODUCTION

1.1 HIV epidemic: advancements and prospective concerns

The human immunodeficiency virus (HIV) remains within the global collection of communicable threats due to a suitable cure still eluding the medical and drug discovery community. Despite worldwide cases declining by 23% since 2010, HIV still holds its global pertinence, with an estimated 38.4 million people living with the virus in 2021, of which 1.5 million account for new infections. Furthermore, around two-thirds of global infections occur in Africa, with 7.8 million infections being approximated in South Africa (roughly, 13% of the population) (UNAIDS, 2021). HIV is classified into two main types, HIV-1 and HIV-2, with the former attributing to the majority of global infections while the latter is localised to West Africa (Fee & Parry, 2008). Within HIV-1, group M contains the subtypes B and C. Subtype B displays prevalence in developed countries such as the US, Western Europe, and Australia whereas subtype C is predominant in sub-Saharan Africa, India, China, and Brazil. Despite that half of the total global infections are subtype C viral infections, the research for this subtype is severely underdeveloped in comparison to subtype B due to the geographic distribution of these subtypes (de Oliveira *et al.,* 2003; Naicker *et al.,* 2013). Limited resources have slowed the progress in research concerning the predominant subtype (subtype C) in developing countries such as India, China, Brazil, and South Africa. Additionally, the subtype C virus displays a wider global spread. This implies the subtype has a higher viral fitness at a population level furthering the need for continual research of this HIV subtype (de Oliveira *et al.,* 2003).

The advancement of antiretrovirals (ARVs) has rapidly improved the treatment of disease progression in infected individuals by controlling viral replicative capacity. According to UNAIDS (2021), more than 50% of global HIV-positive cases have access to ARVs. However, the efficacy of ARVs is continually threatened by the rise in viral resistance. HIV possesses notable genetic diversity which can be accredited to a fallible reverse transcriptase combined with a high viral replication rate. This results in the continuous generation and selection of naturally occurring polymorphisms (NOPs). As a result, HIV proteins display significant mutation rates which result in the accumulation of drug-resistant mutations in drug-selecting environments and eventually the introduction of these mutant strains into drug naïve populations (Naicker *et al.,* 2013; Williams *et al.,* 2019).

Failure to address the continually mutating virus will lead to an eventually reduced capacity to slow disease progression due to the depleting long-term sustainability of currently available treatments. Furthermore, evidence has shown that the accumulation of NOPs and subsequent selection for drug-resistant mutations are inherently different between subtypes. Increasing evidence is suggesting this inherent difference in genetic diversity introduces differences in disease progression and accepted and novel mechanisms to drug resistance between subtypes (Taylor *et al.,* 2008). As a result, this will

ultimately impact treatment and vaccine development now and in the future (Taylor *et al.,* 2008). Current global predicaments, such as the SARS-CoV-2 epidemic, emphasise the need for continual research regarding viruses especially those prone to generating highly contagious and treatment-resistant strains such as HIV. The search for efficient cures and treatment strategies as well as structure-based efforts of drug design will increasingly rely on data related to understanding molecular mechanisms behind potential drug evasion coupled with the corresponding viral fitness restoration mechanisms.

1.2 HIV Replication cycle and drug targets

The initial drug targets of HIV were discovered through a thorough understanding of the virus replication cycle. HIV targets $CD4^+$ T-lymphocytes and attaches to the CD4 glycoprotein receptors as well as co-receptors CCR5 and CXCR4 to facilitate viral entry (Figure 1). Following fusion into the cell, the viral reverse transcriptase (RT) reverse transcribes the single-stranded viral RNA genome to generate double-stranded DNA. This double-stranded DNA is then inserted into the host genome by the viral integrase (IN) protein. Subsequent exploitation of the host transcriptional and translational machinery facilitates the expression of the HIV gag and gag-pol polyproteins. These polyproteins, in addition to replicates of the viral genome, are transported and covalently anchored to the target cell membrane to assemble immature virions. Following the budding of the non-infectious immature virions, the viral protease (PR) plays the critical role of cleaving the polyproteins into functional and structural viral elements, resulting in the formation of infectious mature virions. Therefore, both the PR and gag viral proteins are essential for the proliferation, maturation, and replicative capacity of HIV and thus pose as major drug targets. As such, the focus has been placed on the development of protease inhibitors (PIs) to disrupt the cleavage of the gag and gag-pol polyprotein by PR (Chan & Kim, 1998; Wyatt and Sodroski, 1998; Neagu *et al.*, 2018).

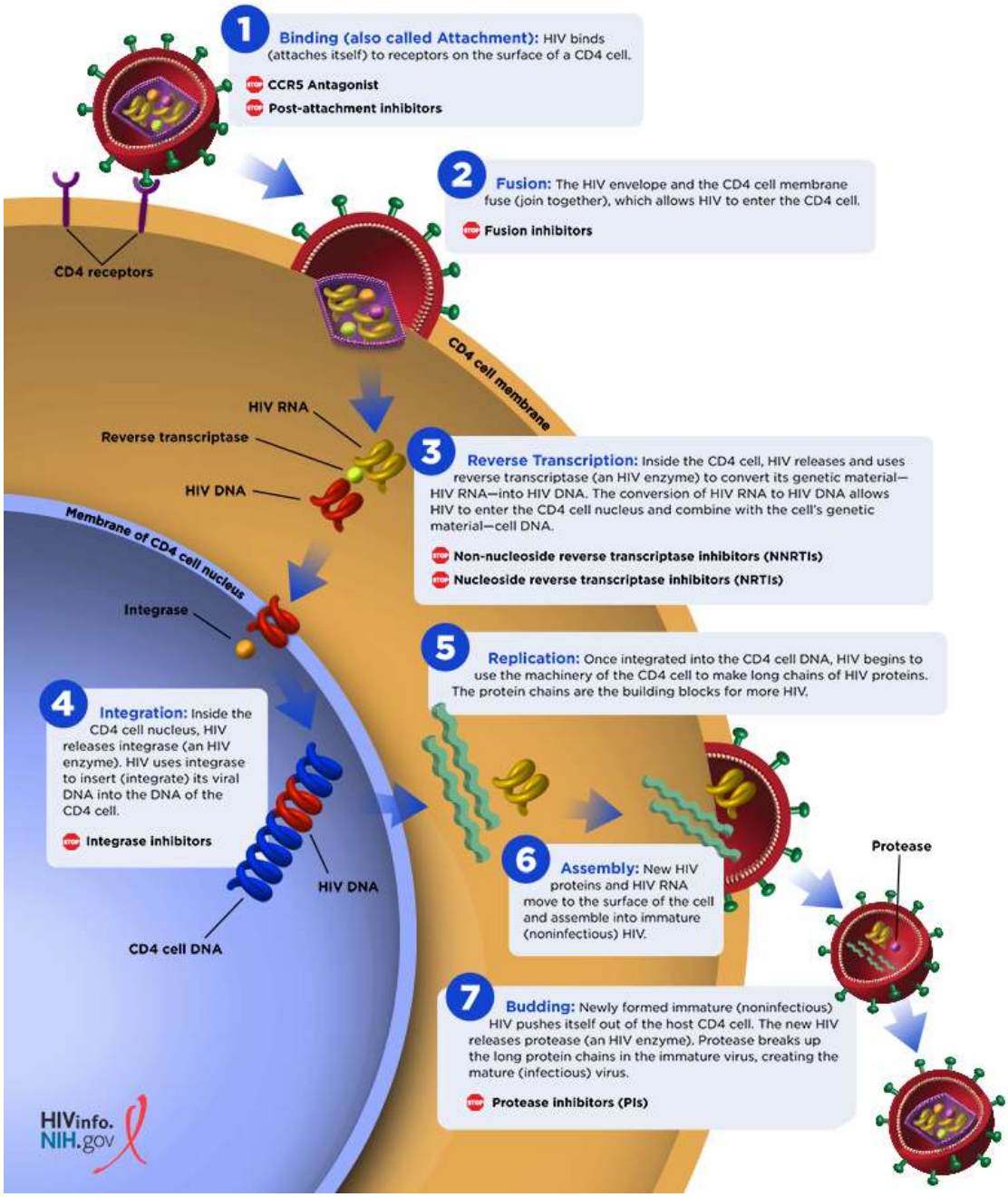


Figure 1 taken from the US office of aids research (OAR) ("About HIVinfo.NIH.gov | NIH," n.d.): Stages of the HIV replication cycle and corresponding drug targets. A depiction of the various stages of the HIV replication cycle and corresponding drug targets to potentially disrupt viral replicative capacity. The interactions between PR and the gag polyprotein are potent targets due to the essential nature of gag polyprotein cleavage by PR for the formation of infectious mature virions (Chan & Kim, 1998; Wyatt and Sodroski, 1998; Neagu *et al.*, 2018).

1.3 Structural and functional characteristics of the HIV PR and gag

1.3.1 PR structure and function

The 22 kDa HIV PR is a homodimeric aspartyl protease which on a secondary structural level, displays a significant number of β-sheets, with only 2 prominent α-helices (Mager, 2001). The functional PR exists as a homodimer composed of two identical polypeptide chains (with 99 amino acids each) enclosing a C2-symmetric active site (Kempf *et al.,* 1991). Two regions of significant intermolecular interactions are imperative to dimer stabilisation. The latter include salt bridges between Arg8, Asp29, and Arg87 at the active site interface (Figure 2A). Secondly, the formation of 4-stranded antiparallel β-sheets between the N- and C-terminals of the PR monomers contribute the most to a stable dimeric PR, in comparison to the latter active site interface. Both sets of intermolecular interactions are essential to the formation of a functional enzyme possessing a flap-enclosed active site where ligands can bind (Weber, 1990).

The PR flaps, hinge & elbow regions, fulcrum, and cantilever are crucial structural elements having significant structural flexibility. The PR flaps (Figure 2A) consist of amino acid residues 46-54 in each monomeric chain and display remarkable structural and dynamic changes upon ligand binding. The dynamics of the flaps are attributed to the equally flexible hinge region of the PR composed of residues 35–42 and 57–61 (Zondagh *et al.*, 2018). Similarly, the cantilever and fulcrum (Figure 2A) cohesively move downwards during flap movement to support ligand entry (Zondagh *et al.*, 2018). Together, the above flexible PR constituents allow for a dynamic equilibrium of semi-open, open, and closed PR forms (Hornak *et al.,* 2006). Subsequently, these PR forms allow for the controlled entry of a ligand (i.e., PI or gag) followed by the formation of key interactions for binding, catalysis, and product release. Conserved interactions include those at a Gly-rich region within the flaps, which are crucial for substrate recognition and binding, as well as conserved pockets of other active site residues (Gustchina & Weber, 1991).

To emphasise the importance of flap control and an ideal active site landscape, research into flap dynamics and drug binding has shown the consequences of mutations within the hinge region and the active site landscape. Consequentially, the latter seems to impact the stability of PIs and the gag in complex with the PR active site (Naicker *et al.,* 2013; Zondagh *et al.,* 2018; Maseko *et al.,* 2019; Williams *et al.,* 2019). Other than the vital and dynamic structural elements previously mentioned, the highly conserved DTG catalytic triad in the active site is also crucial for PR activity. PR performs a nucleophilic attack of the P1/P1' gag peptide bond subsequently separating and releasing the gag domains. Either an Asp25-activated water molecule participates as the negatively charged nucleophile and interacts with the peptide bond, alternatively, a nucleophilic atom of an active site amino acid side-chain satisfies the latter function (Shen *et al.*, 2012).

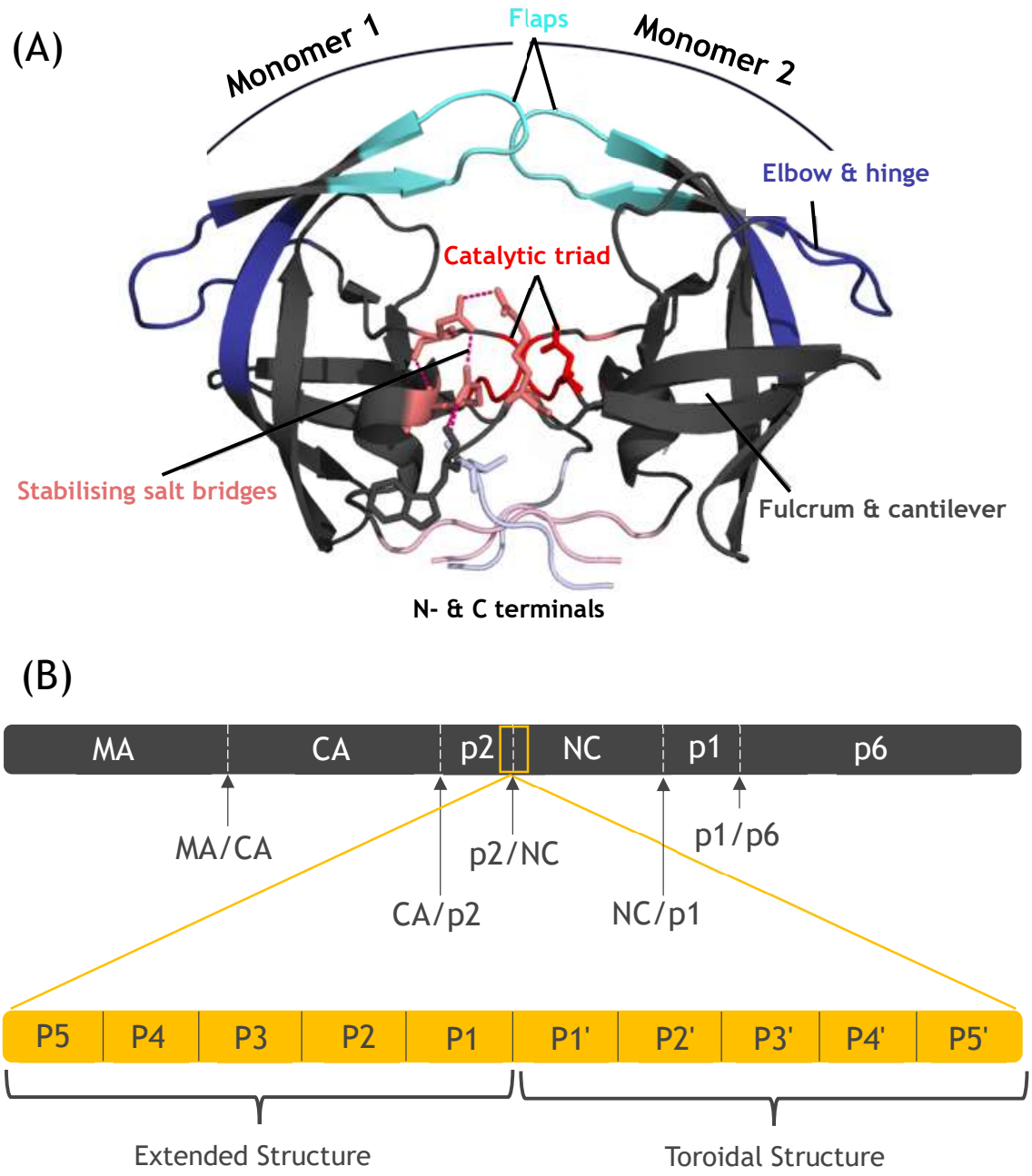


Figure 2: HIV PR structure, gag structure, and PR-gag structural association. (A) WT C-SA PR structure. (B) Schematic of gag structure. (A) The PR possesses many dynamic structural elements which support the controlled entry of peptide, gag, and inhibitor, followed by binding, catalysis, and product release. (B) The gag possesses six domains linked by cleavage sites that display a conserved asymmetric 3D shape, which is characterised by half-toroidal and half-extended structures. This 3D shape is subsequently recognised and processed by the PR at the peptide bond between the P1 and P1' residue (Sundquist & Kräusslich, 2012; Naicker *et al.,* 2013).

1.3.2 Gag Structure and Function

The PR-related gag polyprotein encompasses most of the viral structural and functional protein components. These domains require separation through cleavage by the viral PR to generate virulent particles. The 56 kDa gag is expressed from the host-integrated viral genome as a 500 amino acid polypeptide chain (Sundquist & Kräusslich, 2012). Overall, the gag consists predominantly of coiled structures, several helices, and extended secondary structures; however, there is no complete model of an isolated gag model on the Protein Data Bank (PDB) (Wright *et al.,* 2007). The five gag cleavage sites exhibit the most consistent secondary structure and possess a superimposable form that fits within the active site, understood as the substrate envelope in biological terms. Specifically, the PR processes the gag at these cleavage sites after the gag polyproteins have assembled into tertiary and quaternary structures associated with immature virions. Structural data of immature virions at the plasma membrane of the infected T-lymphocyte cell has shown this higher-order assembly to consist of homotrimers complexed further as hexamers of trimers (Sundquist & Kräusslich, 2012).

The processed cleavage sites are found between folded domains connected by flexible polypeptide regions wherein the HIV PR-targeted cleavage sites exist (Figure 2B). The highly regulated processing of the gag begins at the p2/NC cleavage site and continues in the following order: MA/CA,p1/p6, CA/p2, and then the NC/p1(Sundquist & Kräusslich, 2012). Subsequent cleavage sites are processed at rates that are consecutively several folds slower than the p2/NC site (Pettit *et al.,* 2005). Gag cleavage sites display almost no sequence homology but do possess a conserved asymmetric, superimposable form. The reliance of PR specificity on gag 3D shape corresponds with the difficulty to mimic this in a drug scenario. However, this does explain the diverse cleavage site sequences the PR can identify and process (Prabu-Jeyabalan *et al.,* 2002). The resultant structural proteins, denoted MA, CA, Sp1/p2, NC, Sp2/p1, and p6, have vital structural or functional roles in the formation of mature infectious HIV particles. The MA, NC, and CA domains form the matrix, nucleocapsid, and capsid structural proteins. The p1 and p2 domains are thought to have regulatory functions in the virus life cycle while the p6 domain is essential to the viral assembly at the T-lymphocyte membrane and the promotion of budding (Sundquist & Kräusslich, 2012).

Given the above, mature HIV virion production and maintenance is not only propelled by the PR but by the partnering gag polyprotein. Given the structural characteristics of the gag, it is likely that changes like the gag cleavage sites would directly impact PR processing. Especially considering the different gag cleavage sites with no sequence homology are processed at varying efficiencies. Secondly, each gag domain assumes an indispensable role in virion budding and maturation. Thirdly, various domains of the gag (such as the p1, p2, and p6 domains) participate in allosteric mechanisms that regulate viral replication. Furthermore, this is most likely achieved cooperatively to support immature virion

generation. Thus, any changes in these gag domains would most likely impact viral fitness too. Therefore, the gag can be considered a functional partner to the PR in ensuring the replicative capacity of HIV.

1.3.3 PR-Gag Binding Structural Elements

Many key structural changes occur when the gag interacts with the PR. PR displays significant symmetry in the apoenzyme form. Upon gag binding, that symmetry is broken resulting in 1 000 $Å^2$ of buried PR surface area. Each PR monomer makes significant structural adaptations to accommodate the asymmetric structure of the gag cleavage site side chains (Prabu-Jeyabalan *et al.,* 2002). Complexes of the PR and gag-mimicking peptides in the PDB (1KJ4, 1F7A, 1KJ7, 1KJF, 1KJG, and 1KJH) have shown the gag cleavage site to possess a toroidal form on the primed half and an extended structure on the other half of the cleavage site. This allows for the P1 and P1' residues to bridge the Asp25 residue of the catalytic triad, in the active site (Figure 2B & Figure 3).

In the bound form, the gag forms many hydrogen bonds with the PR. These bonds are mainly between backbone atoms and a few hydrogen bonds between side chain atoms. The gag backbone atoms also form hydrogen bonds with the PR flaps in parallel and antiparallel β-sheet conformations, as illustrated in Figure 3. These hydrogen bonds with the PR flaps are a key structural factor in substrate recognition (Prabu-Jeyabalan *et al.,* 2002). The latter hydrogen bonds with the PR flaps are found to be maintained between all gag cleavage sites and consistently involve gag cleavage site residues P4-P4' in contact with PR residues Asn25', Gly 27/27', Asp29/29' and Gly48/48' (Prabu-Jeyabalan *et al.,* 2002). There are some interactions between gag side-chains and PR backbone atoms which are only partially conserved. These often involve the PR resides Asp29/29' and Asp30/30'. At the cleavage site, the P1/P1' peptide bond, is a sharp deviation in backbone torsional angles compared to the rest of the gag cleavage site backbone torsional angles. The toroidal-extended 3D form and peptide backbone accommodation are necessary for strong hydrogen bonding between the P1 amino acid and Asp25 catalytic residue. This is the intermolecular contact that initiates PR catalysis. Between all cleavage sites, there are also conserved water molecules that bridge backbone hydrogen bonds between the PR and ligand as seen in Figure 3 (Prabu-Jeyabalan *et al.,* 2002). This is a common binding strategy used to assist bond formation between atoms that cannot be at a close enough distance necessary for bond formation without introducing steric clashing. A partially conserved Gly residue at P4' that binds to the Asp29' amino group, is one unique characteristic, between the PR and the p2/NC site. Overall, there are common hydrogen bonding patterns between all the cleavage sites as well as other common structural elements associated with the accommodations and limitations of pockets of residues in the PR active site (Prabu-Jeyabalan *et al.,* 2002).

The structural accommodations the PR binding pockets endure to allow the fit of the asymmetric gag cleavage site are remarkable. However, there are some variations in the limitations between some PR active site pockets designated to accommodate corresponding gag cleavage site residues. Commonly all PR pockets are largely hydrophobic, correspondingly the gag cleavage sites possess many hydrophobic residues. Between PR pockets there are conserved structural limitations that correspondingly place conserved restrictions on the identity of amino acids at specific positions on the gag cleavage site. One conserved example is the lower degree of accommodation at the P2/P2' binding pockets in comparison to the flap region and P1 loop (Figure 3) (Prabu-Jeyabalan *et al.,* 2002). The P2 pocket has a more compact molecular space than other PR pockets and thus mutations in the P2 residue of the gag cleavage site are more restricted (Shen *et al.*, 2012).

The design of FDA-approved PIs has been based on a similar substrate envelope as the gag cleavage sites (i.e., mimic gag cleavage site 3D structure). PIs aim to disrupt the PR-gag interaction, eliminating the formation of mature infectious viral particles from immature viral particles. However, primary mutations and secondary mutations in the PR as well as mutations in the gag cleavage and non-cleavage sites accumulate and are selected throughout treatment. These selected mutations often result in drug resistance and an increase in viral fitness. The inherent flexibility of peptides, such as that associated with the gag, contributes to an ability to adapt to changes in the active site landscape, because of accumulated mutations. Furthermore, the preservation of gag binding in the event of mutations could also be supported by compensatory gag mutations which restore PR-gag associations and subsequently viral fitness. Both the latter phenomena have been characterised as some of the key qualities the gag may possess to adapt to substrate envelope changes, where rigid synthetic PIs could not. Structure-based drug design, especially given the increasing mutations within the PR and the gag, will increasingly benefit from insights regarding the gag cleavage sites in association with the PR. This is because a more thorough understanding of drug resistance mechanisms and scenarios will be achieved through continuous probing of the extensions and limitations of the PR and gag relationship.

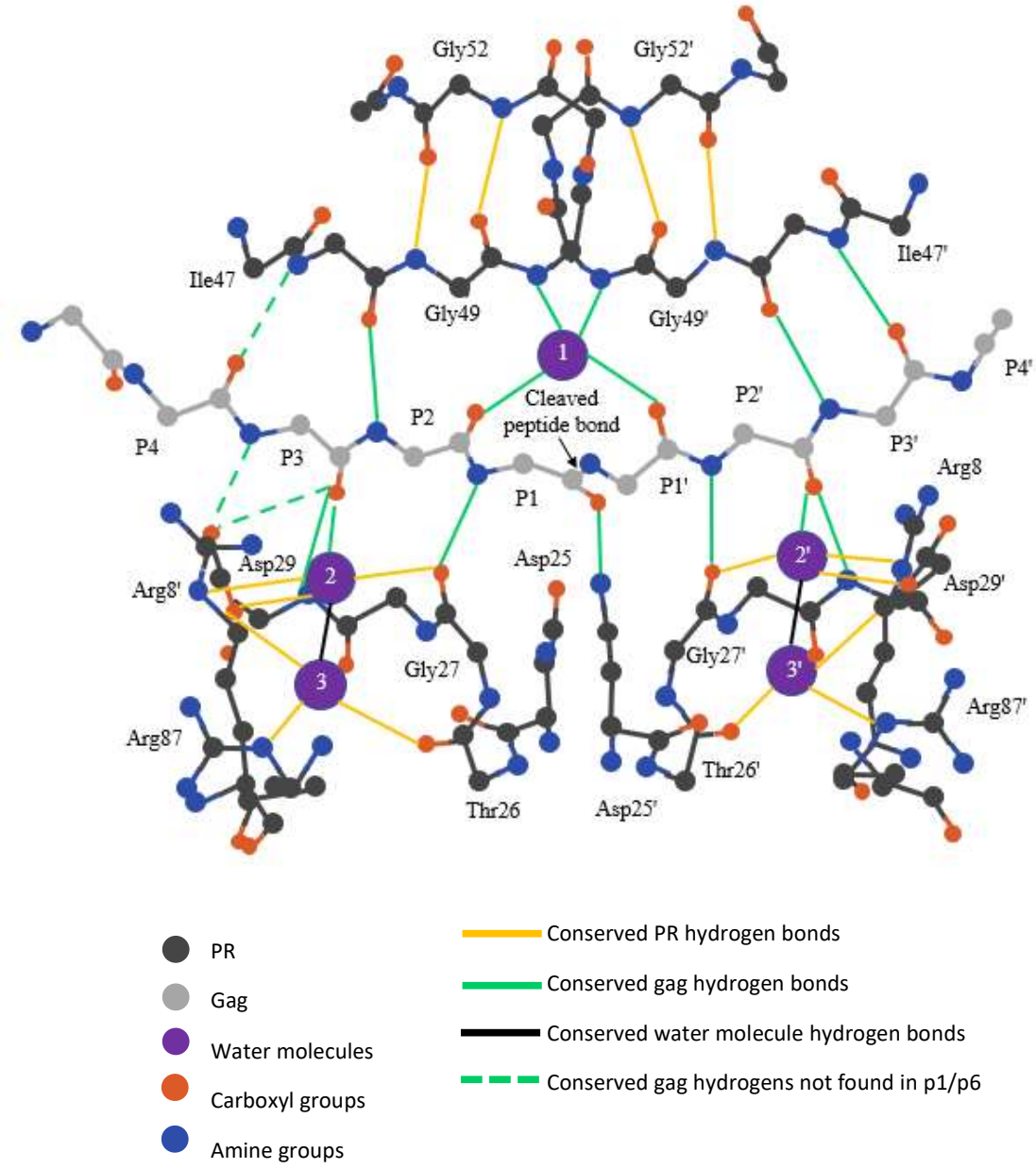


Figure 3: Structural elements of PR-gag binding. The PR-gag complex is characteristic of several conserved hydrogen bonds, asymmetric shape, and limitations across the association to all the gag cleavage sites. This schematic depicts the conserved hydrogen bonds dictated in section 1.3.3 as well as the minimal number of hydrogen bonds that are not conserved across all cleavage sites. This image was adapted from (Prabu-Jeyabalan *et al.,* 2002).

1.3 Mutations in the HIV PR and Gag

The PR and gag NOPs that are selected in the presence of drug pressure often increase drug resistance. As the gag is inherently more flexible it may be able to accommodate active site mutations in the PR that elicit changes in the substrate envelope. Synthetic rigid molecules do not possess the same inherent flexibility and thus stable associations with the PR become compromised. PR active site mutations are also detrimental to PR-gag interactions; however, over time compensatory mutations in both the PR and gag are selected and restore PR-gag interaction and thus viral fitness. This generates a fit viral strain with a high resistance phenotype, rendering the effects of PIs obsolete. The vast accumulation of mutations in HIV proteins can be attributed to the error-prone replication machinery of the virus coupled with a fast replication rate (William *et al.,* 2019). NOPs, therefore, accumulate as often as each genome reverse transcription event occurs, making the virus highly polymorphic (Ammaranond *et al.,* 2003; Mansky and Temin, 1995). This alone has resulted in the HIV genome differing by approximately 35% between subtypes.

Polymorphisms generate conformational variations and dynamic differences in subtype-specific PR elements, such as the flaps (Kear *et al.,* 2009). The NOPs T12S, I15V, L19I, M36I, R41K, H69K, L89M, and I93L appear between the subtype C and B PRs (Mosebi *et al.,* 2008). As a result, it has been shown that the subtype C PR can accommodate a wider range of conformations due to an inherent increase in flap flexibility (Naicker *et al.*, 2013). This has ultimately affected the potency of some PIs, originally designed to target the subtype B PR (Huang *et al.,* 2014; Velazquez-Campoy *et al.,* 2002). Early studies have revealed a significant increase in the concentration needed to inhibit 90% of infectious virions of subtype C PRs (compared to subtype B PRs) in the context of FDA-approved inhibitors (Shafer *et al.*, 1999). Accumulated mutations confer differences in viral subtypes and mutants which are highly relevant to the medical and drug discovery community. Increasing evidence suggests that the NOPs between subtypes elicit differing degrees of resistance and subsequently influence different pathways to acquire new resistance mutations (Wainberg & Brenner, 2012). Originally drug resistance has been primarily investigated in the PR; however, many studies are suggesting the selection for mutations in the gag should be considered equally as potent in supporting drug resistance phenotypes (Su *et al.,* 2019). A strategy the virus may employ to maintain sufficient catalytic efficiency in primary resistance carrying-PRs.

1.4.1 PR Polymorphisms

Mutations that occur within the PR active site envelope are classified as primary mutations, while those occurring elsewhere within the protein are classified as secondary mutations. The mechanisms of primary PR drug resistance mutations are the most well-understood of the potential mutations in HIV

proteins. Since primary mutations change the active site geometry, drugs may no longer be able to bind due to steric hindrance and the loss of essential interactions. Therefore, primary mutations at residues such as V82 can directly confer drug resistance. However, since changes in the active site may also deter gag binding, the virus relies on the support of secondary mutations. Secondary mutations can assist in improving overall PR kinetic efficiency, and consequently viral fitness, thereby contributing to cross-resistance. Despite the imperative nature of secondary mutations in generating highly resistant phenotypes, their mechanisms are less well understood (Manolescu *et al.,* 2010; Huang *et al.,* 2014).

As aforementioned, the effect of NOPs, in the form of primary mutations, on the structural, functional, and dynamic features of PRs may lead to drug resistance. Mutations in the active site influence the shape of the substrate envelope which is occupied by the PIs that aimed to mimic it. Simultaneously, the same primary mutations may alter the nature of the PR-gag interaction and thus processing efficiency (Velázquez-Campoy *et al.*, 2003; Bandaranayake *et al.*, 2010). To compensate for this, the PR may accumulate and select compensatory secondary mutations during viral infection to regain enzymatic efficiency. Secondary substitution mutations are attributed to the mistakes in reverse transcription while mutations that alter the number of amino acids in the PR can be attributed to reverse transcriptase slippage contributing to the duplications or deletion of nearby sections of sequences (Penno *et al.,* 2017). Secondary mutations located in PR structural elements like the critical hinge region alter the length and overall structure of these PR elements (Figure 4). Which in turn changes how the hinge controls the flexibility of the all-important PR flaps. The resultant structural changes often impact the dynamic behaviour of the PR overall.

Despite the historical scarcity of insertion mutations, the number of documented cases has increased. This is especially true between residues 32 and 42 which mark the hinge region (Adamson, 2012). Understanding PR structural elements and the contributing factors that influence regions that control substrate entry has pioneered investigations into alternative strategies for targeting the PR, such as the use of allosteric inhibitors. Rather than a directly competitive mechanism aimed at eliminating the PR-gag interaction, these strategies would aim to hinder the ability of the PR structural elements to support an open conformation, thereby compromising catalytic function. Such have included computational investigations into a peptide with an allosteric binding site such as the epitope regions, also connected to flap flexibility (Badaya & Sasidhar, 2020). This is an ideal example of exploring alternative treatment strategies based on alternative understandings of how HIV evades drugs and generates resistance.

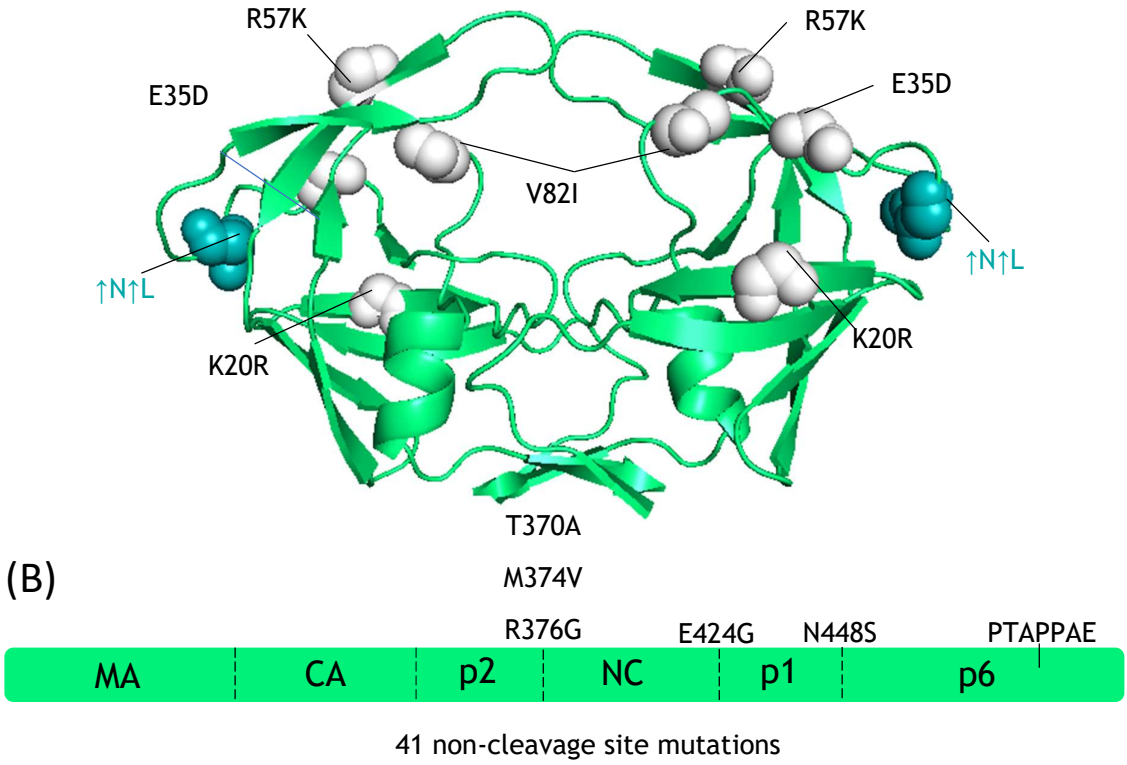


Figure 4: The L38↑N↑L^{+4} mutant PR and gag pair. (A) Mutations in the L38↑N↑L^{+4}PR, visualised in PyMOL, by Schrödinger. (B) Schematic of the mutations in the L38↑N↑L^{+4}gag. (C) Alignment of L38↑N↑L^{+4}PR and the WT C-SA PR visualised in the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. (A) The L38↑N↑L^{+4} PR is characterised by 1 hinge insertion, one primary, and 3 background mutations. (B) The L38↑N↑L^{+4} gag is characterised by 41 non-cleavage site mutations, one PTAPPAE duplication in the p6 domain, and five cleavage site substitutions. Three substitution mutations are concentrated in the p2/NC site. (C) Hinge insertions influence the structure of the hinge region and thus the dynamics of the PR flaps.

Accumulated secondary mutations, such as a hinge insertion, are more than just relevant in drug-selecting environments but have also shown long-term persistence in PI-naïve individuals. This long-term persistence outside of drug-selecting environments suggests hinge mutations could provide some viral benefit (Amiel *et al.*, 2011). If this is true, hinge mutation frequency could increase. Furthermore, if coupled with PR and gag primary mutations could generate highly resistant phenotypes in infected individuals which current inhibitors could not address. The biochemical relevance of hinge insertion mutants has been studied concerning PIs, displaying PI interactions with less favourable Gibbs free binding energies (Zondagh *et al.,* 2018; Maseko *et al.,* 2019; Williams *et al.,* 2019).

Similarly, biochemical consequences are evident with the mutant of interest in this study, L38↑N↑L^{+4} PR. In comparison to the wild-type PR, the L38↑N↑L^{+4} PR displays a decreased affinity for the conserved WT-mimicking gag cleavage site in addition to decreased catalytic efficiency (Williams *et al.,* 2019). Therefore, to restore viral fitness, mutations elsewhere in the PR would need to form to restore effective gag binding. Although the conserved gag cleavage site is not efficiently processed, a mutated gag cleavage site was shown to alter PR drug susceptibility. Thus, the research community is beginning to accept the potential role of compensatory gag mutations in restoring PR-gag interactions and/or generating high-resistance phenotypes. Essentially, viral strains that accumulate significant drug resistance through multiple mutations in the PR, eventually compensate for any loss in viral fitness through the accumulation of gag cleavage site and non-cleavage site mutations.

1.4.2 Gag Polymorphisms

In recent years, interest has grown regarding a potential compensatory relationship between viral counterparts such as the PR and gag. This has developed from evidence that has suggested that PR-related gag proteins could have a role in reducing susceptibility to PIs. Gag mutations in cleavage sites have been associated with restoring PR-gag interactions subsequently being associated with increases in gag binding affinity over the PIs where the gag is mutant PR-related. Essentially the gag is being realised as a potential tool exploited by the virus in rendering PR mutations and mutation combinations plausible that would be expected to be ineffective alone. Considering the intricacy of PR-gag binding and restoration in the understanding of drug resistance development will ultimately support more effective strategies in drug design, as evidence increasingly suggests the gag could be exploited as a potential path to generating high resistance phenotypes.

The gag, a significantly polymorphic HIV element, possesses conserved nucleocapsid zinc-finger domains, as well as conserved capsid N- and C-terminal domains. In recent years, significant investigations have gone into resolving the polymorphic patterns associated with gag domains and such between subtypes too. Such understanding of these genetic patterns would include which regions have

fewer restrictions on accumulated mutations, which are prone to harbouring resistance mutations, and how these may relate to the same phenomena occurring in related PRs. Domains such as the MA NC, Sp2, and p6 domains of the gag have been considered more polymorphic and prone to a high number of resistance cleavage site mutations (CSM). On the contrary, other studies indicate that the generation and selection of NOPs are more prominent between p2 and NC domains. Higher resistance mutation frequency within the MA/CA, NC/p1, and p1/p6 cleavage sites could be correlated with cleavage rate. The latter sites are processed at a lower rate than the p2/NC cleavage site. Potentially the sequence and structural nature of the p2/NC site is already evolutionarily enhanced to support high affinity binding affinity and ideal processing efficiency in WT viruses. While other cleavage sites associated with lower binding affinity and rate of cleavage are more likely tolerable of the accumulation of a diverse range of mutations.

However, available databases have limited genotypic insights regarding the gag genotype, patterns of polymorphisms susceptibility within domains, and the corresponding differences amongst subtypes. Genotypic data regarding the gag is significantly limited in comparison to other viral elements. Thus, understandings of the polymorphic landscape of gag domains and cleavage sites are too. Gag resistance pathways, NOPs, and the consequences concerning response to treatment would likely vary amongst subtypes as has PR polymorphisms and drug resistance mechanisms. The strongest consensus that can be reached across eight major HIV-1 subtypes, is that NOPs vary between subtype and domain. The structural and functional relevance as well as in connection with the corresponding PR is still relatively underdeveloped.

Gag cleavage and non-cleavage site mutations (NCSM) have been connected to subtype B PR mutations involved in conferring PI resistance. CSMs include SNPS within the designated gag cleavage sites which seem to rarely occur without PR mutations first and more likely as a range of secondary mutations. These secondary mutations assist in restoring PR-gag interactions that have been negatively impacted by primary drug resistance mutations in the PR active site. Other studies have also suggested some gag mutations can generate primary resistance in the absence of initial PR resistance, which would call for major concern. For example, mutant viruses with certain mutations in the p2/NC cleavage site, when compared to subtype C wild-type virus, displayed a higher PI concentration needed to inhibit 50% of the infectious virions (Nijhuis *et al.,* 2007). Reversing these gag mutations resulted in a significant reduction of the viral mutant drug resistance (Dam *et al.*, 2009). NCSMs are the substitutions, deletions, and duplications within gag domains excluding the cleavage site residue positions. Similarly, to secondary PR mutations, NCSMs are also thought to be predominantly compensatory.

Other investigations have suggested gag mutations use allosteric mechanisms to support reduced PI susceptibility in the absence of PR resistance mutations (Gatanaga *et al.,* 2002; Lastere *et al.,* 2004;

Parry *et al.,* 2011). One potential way could be through allosteric communication of the NCSMs with upstream cleavage sites, contributing to PI resistance (Su *et al.,* 2018). For example, viruses of the C subtype, have several distinct biological properties from other subtypes. Specifically, the occurrence of the NCSM, and the PTAPPAE duplication in the p6 domain. The Pro-rich p6 domain is responsible for the recruitment of cellular factors involved in budding. Thus, a duplication could enhance viral budding, conferring a replication advantage (Williams *et al.,* 2019). Since NCSMs do not accumulate in PR cleavage sites, it is most likely allosteric mechanisms for achieving resistance would involve multiple mutations in various domains. These mutations would most likely differ in identity and even per domain. Allosteric mechanisms are not well understood.

There are many mechanistic roles of gag mutations that could be considered in viral adaption. These include the role of CSMs concerning PR mutations or acting in a primary resistance context of a certain PI, and even allosteric communication between functional and non-functional domains containing mutations to enhance the viral life cycle. For example, one study found that a PR of sufficient hydrolytic efficiency had accumulated approximately 19 mutations and as many mutations located close to or within the p2/NC, NC/p1, and p6pol/PR domains (Kožíšek *et al.,* 2012). It is becoming more evident that many mechanisms and combinatorial ones are being evolutionarily tailored by the virus and will eventually be a prominent part of the viral population. Only a few studies have investigated these events in the context of subtype C PR-gag pairs.

In a subtype C context, PR resistance to Ritonavir (RTV has been associated with the majority of cases attributed to PR mutations but also instances in which patients possess gag substitutions without PR mutations (Giandhari *et al.,* 2015). Furthermore, subtype C infections have shown lower PI sensitivity in the absence and presence of PR substitutions (M46I, I54V, V82A, K20R, and L33F) while frequently possessing gag substitution mutations at cleavage site residues 431, 436, and 437 (Giandhari *et al.*, 2016; Nijhuis *et al.*, 2007; Fun *et al.*, 2011). Some of the above PR mutations show a potential link between hinge region polymorphisms and high resistance phenotypes involving the support of gag CSMs (Giandhari *et al.,* 2016). Relevant to this project is the landscape of the p2/NC cleavage site which in the above investigation displayed mutations in p2/NC cleavage site positions 373, 374, and 378. The first two p2/NC mutations were linked to PR V82 substitutions in the active site and hinge mutations at residues 35, 36, and 37. Specifically, substitutions at gag residue 374 coupled with PR mutations have been isolated from patients with resistance to the PR inhibitors, RTV, and LPV (Giandhari *et al.*, 2016). These findings are all relevant to the mutational content of the L38↑N↑L^{+4} mutant virus. The combinatorial association of gag mutations as well as with PR mutations, suggests a potential increase in the level of intricacy in future resistant viruses (Fun *et al.,* 2012).

1.5 Obstacles in HIV Treatment Efficacy and Prospects

1.5.1 FDA-Approved PR Inhibitors

The most effective treatment method for HIV/AIDS is highly active antiretroviral therapy (HAART), which includes the use of highly potent PIs (Lv *et al.,* 2015). The biggest disadvantage to PIs from a treatment perspective is lower than ideal molecular selectivity which causes many side effects. From a biochemical perspective, it is the rigid nature of synthetic drugs that do not readily adapt to changes in the corresponding binding site i.e., the substrate envelope in the active site of the PR. The FDA has approved 26 anti-HIV compounds which include ten PIs (Voshavar, 2019). These PIs include saquinavir (SQV), indinavir, ritonavir, nelfinavir, amprenavir (APV), fosamprenavir, lopinavir (LPV), atazanavir (ATV), tipranavir, and darunavir (DRV) (Voshavar, 2019). These inhibitors have been designed to mimic the structural elements of PR-gag interaction discussed in Section 1.3.3, with chemical modifications to increase potency, address resistance, and reduce side effects. Effectively, PIs are competitive inhibitors, binding at the active site of the PR in direct competition with the gag cleavage sites.

Current inhibitors mimic the substrate/gag transition state by possessing a hydroxyl group that forms hydrogens with the carboxyl group of PR active site residues, Asp25 and Asp25′, in the catalytic triad (Lv *et al.,* 2015). PR residues such as Gly 27, Asp 29, Asp 30, and Gly 48, are relatively conserved, but the accumulation of drug-resistance mutations renders PI-PR interactions less favourable inducing treatment resistance. Over the years as resistance has developed, newer PIs have been designed to accommodate patient resistance to original PIs. For example, lopinavir is closely related to ritonavir except for the P2 and P2′ groups. These groups targeting the P2 and P2′ PR pockets are smaller, decreasing the contact with highly variable residues at the 82 sites of HIV-1 protease (Sham *et al.,* 1998). Other PIs such as ATV were designed to improve bioavailability and reduce the side effects associated with other PIs.

The most recently designed PI aimed to accommodate a reduction in both side effects while addressing viral resistance. Essentially, DRV forms hydrogen bonds with the backbone of the PR that slow the development of drug resistance. The structure of DRV, based on the structure of APV includes a tetrahydrofuran P2 group that allows the drug to form more hydrogen bonds with the Asp29 residues of the PR (Lefebvre and Schiffer, 2008). By targeting interactions with critical PR residues, the impact of mutations on drug binding is limited in the PR. However, attempts to target the limitations of the PR could be compromised by gag resistance mechanisms. Thus, future drug design efforts will most probably benefit from a consideration of the PR and gag as functional units not only in catalysis but in drug evasion too.

1.5.2 Addressing HIV Drug Resistance Mechanisms

The pool of information achieved through HIV research shows that HIV viral proteins will continue to accumulate drug-resistant mutations in drug-selecting environments and these resistant strains will eventually be introduced into drug naïve populations. The PR accumulates primary mutations which impact inhibitor and substrate binding, but high resistance phenotypes are dependent on the coupling of drug-resistant primary mutations and compensatory secondary mutations. Increasing evidence is suggesting that PR is not the only route to drug evasion but rather that gag could participate in many pathways that lead to drug resistance (Su *et al.,* 2011).

Promisingly, a few studies have probed PR PI susceptibility in the connection of related gag peptides possessing both CSMs and NCSMs. Kinetic analysis of mutant gag-mimicking peptides against wild-type PRs and mutant PRs has revealed the consequences of mutations at gag cleavage site residues. Further, structural analyses increased some understanding of the mechanisms involved in gag polypeptide-induced resistance to PIs (Fehér *et al.*, 2002). In comparing various combinations of mutant peptides and PRs the major determinant of specificity in the PR-gag cleavage site interaction was found to be the maximisation of intermolecular interactions between the two. A continuation that addresses this relationship structurally and functionally would address many questions. Potentially, how single mutations act locally within the PR-gag association, what the differences in molecular interactions are between mutant functional pairs to that of wild-type functional pairs, and how dynamic balance between regulated and ordered processing of cleavage sites maintained in the presence of these mutations (Fehér *et al.*, 2002; Ho *et al.,* 2008; Taylor *et al.,* 2008; Dam *et al.,* 2011; Su *et al.,* 2019; Williams *et al.,* 2019).

Furthermore, it would be invaluable to understand in which domains PI-resistant mutations accumulate, which gag mutations produce primary resistance, which acts to compensate losses in PR-gag interactions, and the mechanisms of mutation sites in synergy within the gag (de Oliveira *et al.,* 2003). Subsequently, all the above concerning different subtypes. Structural and binding investigations are imperative to uncovering the possible synergistic relationships between PR and gag. This understanding can be exploited to reveal strategies for designing more sustainable gag inhibitors and PR inhibitors. In the scheme of keeping the HIV epidemic under control: genotypic data, structural, and functional data concerning gag mutations will assist identification of mechanisms to resistance and differentiate between the source of resistance that elicits PI resistance in patients. Knowledge regarding these gag mutations in functional synergy with the related mutant PRs could reveal limitations in the PR-gag relationship that could be exploited in drug design. The cumulation of this information regarding the PR-gag interrelationship will prove insightful to novel therapeutics and sustainable solutions to disrupting critical protein interactions in the viral replication cycle (Su *et al.,* 2019).

1.6 The L38↑N↑L^{+4} PR-Gag Pair

1.6.1 L38↑N↑L^{+4} PR

The L38↑N↑L^{+4} PR and L38↑N↑L^{+4} gag proteins were isolated from a PR inhibitor naïve child whose mother had never been exposed to PIs either. This indicates the potential increase of variant pairs within the drug-naïve infected population, with multiple mutations in the PR and the related gag that could be structurally and functionally influential. The PR is characterised by three background mutations (K20R, E35D, and R57K), one primary mutation (V82I), and an insertion of asparagine and leucine at residue position 38, located in the hinge. The mutation at amino acid position 82 is an accepted primary resistance mutation site while the insertion is considered a rare but increasingly apparent type of mutation. In previous investigations, the L38↑N↑L^{+4} PR displayed overall reduced catalytic efficiency (Williams *et al.,* 2019). This was concluded after a ±50% reduction in the L38↑N↑L^{+4} PR specific activity and catalytic turnover was identified (in comparison to the wild-type PR) in assays against a conserved wild-type gag mimicking peptide. The maximal velocity (V_{max}) of the L38↑N↑L^{+4} PR remained comparable to that of the wild-type PR and unexpectedly the L38↑N↑L^{+4} PR had a greater affinity for the wild-type peptide.

In contrast, the catalytic efficiency (k_{cat}/K_M) and specific activity were 12-fold and 2-fold less, respectively, in comparison to the wild-type PR. Studies such as this support the revelation that mutations in the PR reduce viral fitness and subsequently need the support of compensatory mutations in either the gag or PR to generate a high-resistance phenotype. Inhibition studies with the PR alone did not reveal any significant changes in drug susceptibility. The inhibitory constant for L38↑N↑L^{+4} PR for PIs such as LPV and DRV of the PR alone indicated increased susceptibility to the drugs in comparison to wild-type PR. Overall there was no significant decrease in PI susceptibility related to the L38↑N↑L^{+4} PR (Williams *et al.*, 2019). This lent to an investigation of the gag to give a reason as to why this mutant virus persisted in the PI-naïve patient.

1.6.2 L38↑N↑L^{+4} Gag

Although the L38↑N↑L^{+4} PR revealed no significant decreased susceptibility to PIs, aberrant susceptibilities were displayed in the context of the related gag. A phenotypic assay involving the L38↑N↑L^{+4} gag revealed no change in LPV and ATV susceptibility. On the contrary, susceptibility to DRV was found to be compromised. These studies suggested that the gag processing efficiency of the L38↑N↑L^{+4} PR was reduced, however, the related gag polypeptide could be evolving as a contributor to a potential resistance phenotype in an *in vivo* context. Mutations in the related gag cleavage sites included T370A, M374V, and R376G in the p2/NC cleavage site, E424G in the NC/p1 cleavage site, and N448S in the p1/p6 cleavage site. It was unexpected to discover the site of initiation and regulation,

the p2/NC site, was 30% mutated compared to the wild-type sequence. NCSMs have been implicated in drug resistance scenarios; however, potential allosteric mechanisms still require thorough investigation. In a subtype C context, this would involve the investigation of duplications such as the PTAPPAE motif found within the p6 domain of the L38↑N↑L^{+4} gag (Williams *et al.*, 2019).

The existence of functional pairs that still maintain substrate-enzyme interaction and sufficient functionality could be related to novel pathways to drug resistance and restored viral fitness. The coexistence of the L38↑N↑L^{+4} functional pair highlights PR resistance to DRV, which does not occur in *in vitro* isolation of the PR (Williams *et al.,* 2019). The related gag contribution to the latter is still to be confirmed as gag contributions to restored PR activity are less well characterised. Thus, investigation of the L38↑N↑L^{+4} PR-gag functional pair could reveal potential insights into PR-gag-induced drug resistance and viral restoration.

1.7 Aim and Objectives

The L38↑N↑L^{+4} PR (referred to as mutant PR from here onwards) is characterised by a significant hinge insertion mutation in addition to four background mutations. Hinge mutations are increasing in frequency in the HIV-infected population. In comparison to the wild-type PR, the mutant PR has been characterised by decreased catalytic efficiency coupled with a reduced DRV susceptibility in the presence of the related mutant gag protein. Therefore, the main aim of this study was to probe the structural and functional differences between the mutant PR and wild-type PR and probe any functionality associated with mutations present in the p2/NC cleavage site.

The objectives were as follows:

1. Perform heterologous overexpression and purification of the wild-type and mutant PRs

2. Perform a global structural evaluation of wild-type and mutant PRs using:
   a. Absorbance-based concentration determination
   b. SDS-PAGE
   c. Far-Ultraviolet (UV) Circular Dichroism (CD)
   d. Intrinsic Fluorescence
   e. Size exclusion high performance chromatography (SE-HPLC)

3. Determine the active PR concentrations achieved following purification using isothermal titration calorimetry (ITC)

4. Perform a global functional evaluation of wild-type and mutant PRs using FRET-based enzyme assays

5. Assess the stability of wild-type and mutant PRs in apo form and in the presence of inhibitors, DRV, and SQV using thermal shift assays

6. Synthesise wild-type and mutant-related p2/NC peptides using solid-phase peptide synthesis and liquid chromatography mass spectrometry

7. Assess prospective binding of soluble peptides to wild-type and mutant PRs using ITC

8. Use computational docking studies to provide molecular context to results in objective 7 and provide insights into p2/NC mutations

Chapter 2. Experimental Procedure

2.1 Heterologous overexpression of inactive and active constructs of the Subtype C wild-type and the L38↑N↑L^{+4} hinge mutant PRs

2.1.1 Transformation of pET-11a mutant HIV-1 PR plasmid constructs into BL21 (DƐ3) pLysS *E. coli* cells

Sequence data for the wild-type PR was obtained from Professor Lynn Morris of the National Institute of Communicable Disease, South Africa (NICD, Johannesburg, South Africa). Sequence data for the mutant PR was obtained from Dr Gillian Hunt and Ms Johanna Ledwaba (NICD, Johannesburg, South Africa). The mutant PR sequence was isolated from a PI-naive child in a government mother-to-child transmission program. In this study, it was considered for further investigation of the structural and functional influences of hinge insertion mutations, which are increasing in frequency in drug naïve populations (Kuhn *et al.,* 2014). One of the practical obstacles associated with the experimental study of HIV-1 PR is autoproteolysis; the PR's ability to non-specifically cleave itself (Rosé *et al.,* 1993). Inactive HIV-1 PRs (containing a D25A mutation) have been found to express in higher yields and without autoproteolytic abilities, improving the *in vitro* practicality of these PR constructs (Prabu-Jeyabalan *et al.,* 2002). Therefore, gene constructs of both the active and inactive versions of the wild-type and mutant PRs were included in this investigation to assess whether inactive constructs could be utilised in place of active constructs to overcome a bottleneck associated with PR yield and experimental practicality. Transformed cell stocks, from previous studies (Naicker *et al.,* 2013 & Sherry *et al.,* 2020), containing active and inactive wild-type PR plasmids, respectively, were utilised for this investigation. The active and inactive mutant PR plasmids were designed for synthesis by Genscript Biotech (Jiangsu Province, China).

The inactive and active mutant PR vectors were subsequently transformed into the same cell line as the inactive and active wild-type PRs, chemically competent BL21 (DƐ3) pLysS *E. coli* cells, using the following procedure. BL21 (DƐ3) pLysS *E. coli* cells were transformed with the pET-11a plasmids

containing the active and inactive (possessing a D25A mutation) mutant PR sequences using the heat shock method. Firstly, 6 µL of 162,0 ng/µL of the vector was introduced to thawed BL21 (DƐ3) pLysS *E. coli* cells, thereafter, mixed by flicking and incubated on ice for 35 min. The vector-cell mixtures were heat-shocked for 45 s at 42 ℃ before being placed on ice for 5 min. Subsequently, the cells were transferred to 950 µL of SOC media (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose). The transformed cells were recovered through incubation within the SOC suspension at 37 ℃ for 1 hour at 200 rpm.

After recovery, 100 µL of the above transformation reactions were spread, using aseptic techniques, onto LB agar plates (1.5% (w/v)) containing 30 $\mu g.mL^{-1}$ chloramphenicol and 100 $\mu g.mL^{-1}$ ampicillin. The remaining suspensions of both transformation reactions were concentrated through centrifugation at 5000 rpm for 5 min using a Spectrafuge 24D microcentrifuge (Labnet International Inc.) followed by being re-suspended in 100 µL of sterile Milli-Q® water. These concentrated suspensions were plated as above on LB agar plates. Similarly, negative controls containing untransformed cell suspensions were plated on LB agar plates. All LB plates were incubated overnight at 37 °C. After 20 hrs, discrete colonies from each plate were used to inoculate fresh LB (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl) supplemented with the necessary antibiotics. These cultures were incubated at 37 °C overnight at 200 rpm. Overnight culture volumes of 500 µL were pelleted at 5000 rpm for 5 min and resuspended in sterile glycerol (80% (v/v)) in a 1:1 ratio before storage at -80 °C.

2.1.2 IPTG induction trials of transformed BL21 (DƐ3) pLysS *E. coli* cells containing inactive wild-type PR, inactive mutant PR, and active mutant PR expression vectors

BL21 (DƐ3) pLysS *E. coli* cells were utilised as expression hosts for inactive and active wild-type PRs throughout the duration of this investigation. Expression conditions for the active and inactive wild-type PRs were adapted from previous publications (Naicker *et al.,* 2013 & Sherry *et al.,* 2020). Expression conditions for both active and inactive mutant PRs were optimised by investigating various concentrations of IPTG and induction periods. The expression conditions of the inactive wild-type were re-trialled for comparison. Induction profiling was performed through IPTG induction over 1-22 hours (0, 2, 4, 6, and 22 hours) at IPTG concentrations ranging from 0 mM to 1 mM. Temperature (37 ℃) and culture media (Luria Broth (LB)) remained fixed. Cell cultures were grown at 37 °C in LB media containing antibiotics (100 $\mu g.mL^{-1}$ ampicillin and 30 $\mu g.mL^{-1}$ chloramphenicol). A growth curve for transformed cells was constructed from time course monitoring of the optical density (OD) at 600 nm of the culture to assess cell viability and identify the incubation period required to reach early log-phase (OD_{600} ~ 0.4) (Hall *et al.,* 2014). Once the early log-phase was achieved, PR expression was induced with the appropriate IPTG concentrations in the above range.

Before induction with IPTG, 1 mL culture samples corresponding to 0 hours of induction were sampled. Subsequently, 1 mL culture samples were taken after each induction period (0, 2, 4, 6, and 22 hours). Expression trail samples were processed for SDS-PAGE analysis as follows. Each sample was pelleted through centrifugation at 10 000 xg for 5 min in the Spectrafuge 24D microcentrifuge (Labnet International Inc.). The pellets were re-suspended in 50 µL of Buffer A (10 mM Tris-HCl, pH 8, 2 mM EDTA disodium, 1 mM PMSF). Cell suspensions were then sonicated thrice for 30 s at 15 V using the XL-2000 series sonicator (Misonix Ultrasonic Liquid Processors). In between sonication cycles, the samples were exposed to 30 s freeze-thaw cycles on ice. Soluble and insoluble cellular fractions were separated through centrifugation at 13 000 xg for 10 minutes. The soluble fraction was combined with reducing sample buffer (12% (w/v) SDS, 30% (v/v) 100 % Glycerol, 150 mM Tris, 6% (w/v) BME, 0.05% (w/v) Coomassie brilliant blue G250) in a 1:1 ratio. The remaining insoluble fraction was resuspended in 50 µL of Buffer A combined with reducing sample buffer in a 1:1 ratio. The prepared SDS-PAGE samples were boiled at 90 ℃ and separated at 120 V on a tricine SDS-PAGE polyacrylamide gel composed of a 16% separating gel and 4% stacking gel (Schägger & von Jagow, 1987). Tricine gels are more suitable for the high-resolution separation of proteins ranging from 1-100 kDa molecular weights (ideal for HIV-1 PR with a predicted reduced molecular weight of 11 kDa) (Schägger and von Jagow, 1987).

2.1.3 Heterologous overexpression of wild-type and mutant HIV-1 PRs

An overnight culture was prepared through inoculation of sterile LB media with a transformed glycerol stock (see section 2.1.1). A total of 8 L of LB media was inoculated with 6% v/v of overnight culture for expression of active wild-type and mutant PRs while 4 L of LB media was inoculated with 6% v/v of overnight culture for expression of inactive wild-type and mutant PRs. The media was incubated at 200 rpm at 37 ˚C until an OD_{600} of 0.4 (Hall *et al.,* 2014). The culture was cold-shocked on ice for 5 minutes to reach the optimal induction temperature of IPTG, 28 ℃ (Mühlmann *et al.,* 2017). PR expression was induced with the relevant IPTG concentration and induction period according to results obtained in section 2.1.2 and established protocols (Naicker *et al.,* 2013 & Sherry *et al.,* 2020). Active and inactive wild-type PR expression was induced with 0.5 mM IPTG for 4 hrs and 0.6 mM for 6 hrs, respectively. Active and inactive mutant PR expression was induced with 0.5 mM IPTG for 4 hrs and 0.2 mM IPTG for 2 hrs, respectively. Subsequently, the cells were harvested through centrifugation at 5000 xg for 5 min at 4 °C using the Sorvall RC6 centrifuge (Thermo Scientific®). The cell pellets were resuspended in 20 mL of Buffer A for storage at -20 °C.

2.2 Ion-exchange chromatographic purification and chemical renaturation of HIV-1 PRs

2.2.1 Purification of HIV-1 PRs using a dual ion-exchange chromatographic technique and chemical renaturation through two-step dialysis

To achieve high yields of 98% pure inactive and active wild-type and mutant HIV-1 PRs, the purification protocol was adapted from a published protocol by Sherry *et al*., 2020 (Figure 5). Following the thawing of cell lysates previously stored at -80 ℃, 50 µL of DNase I (1 mg/mL) and 250 µL of $MgCl_2$ (10 mM) were added. Subsequently, these lysates were incubated at 25 ℃ for 1 hr with continuous rotation to reduce viscosity associated with DNA contamination. $MgCl_2$ supplementation has been shown to improve the DNase I efficiency necessary for effective DNA breakdown (Sherry *et al.,* 2020). The cell lysates were then frozen at -80 ℃ and subsequently thawed three times. Sonication of the cell lysates was performed eight times for 30 s at 70 A with equilibration on ice between sonication cycles. Inclusion bodies were pelleted through centrifugation for 30 min at 18 000 xg and 4 ℃ using the Megafuge 16 centrifuge (Thermo Scientific®). Overexpression of the PR results in the formation of inclusion bodies, thus the supernatant was discarded, and denaturant-based methods were employed to isolate the PR. Firstly, the pellet was resuspended in 100 mL wash buffer (10 mM Tris-HCl, pH 8 2 mM EDTA 2% (v/v) Triton X-100) and centrifuged at 18 000 xg, for 30 min, at 4 ℃. This was intended to rid contaminating proteins non-specifically adhering to PR inclusion bodies. Subsequently, PR-pellets were resuspended in Buffer B (10 mM Tris, 8 M urea, 5 mM DTT, pH 8) and incubated for 60 min at 25 ℃ with rotation. Urea and DTT unfold the inclusion bodies by disrupting non-covalent interactions stabilising the inclusion bodies. Residual insoluble constituents were removed through centrifugation at 18 000 xg, for 30 min, at 4 ℃.

The denatured supernatant was then purified through cation-anion-exchange chromatography using the GE AKTA Start FPLC purification system (GE Healthcare Life Sciences, Chicago). The 5-mL HiTrap® DEAE Fast Flow column and 5-mL HiTrap® CM Fast Flow column columns (GE Healthcare Life Sciences, Chicago) were placed in tandem and equilibrated with Buffer B. As the pH of Buffer B is lower than the PR theoretical pI (~9.1), the net positive surface charge of the PR promotes the binding of the PR to the (-)-CM-Sepharose column (Gasteiger *et al.,* 2005). Before elution, the (+)-DEAE column, whereon negatively charged contaminants were trapped, was removed. The bound PR was eluted from the CM-sepharose column using a single-step elution with Buffer B containing 1M NaCl. The semi-purified, unfolded eluent was diluted to 50 mL with Buffer B and refolded in a dual buffer dialysis system using SnakeSkin dialysis tubing (with a 3.5 kDa exclusion limit) (Thermo Scientific®). The eluent was dialysed in Buffer C (10 mM Formic acid, pH 3, 10% (v/v) 100% Glycerol, 0.01% (w/v) sodium azide) for 4 hr at 4°C using a 1:100 sample/buffer ratio. The low pH denatures non-specific PRs, but not HIV-1 PR, an aspartyl PR with a low pH preference (Volontè *et al.,* 2011).

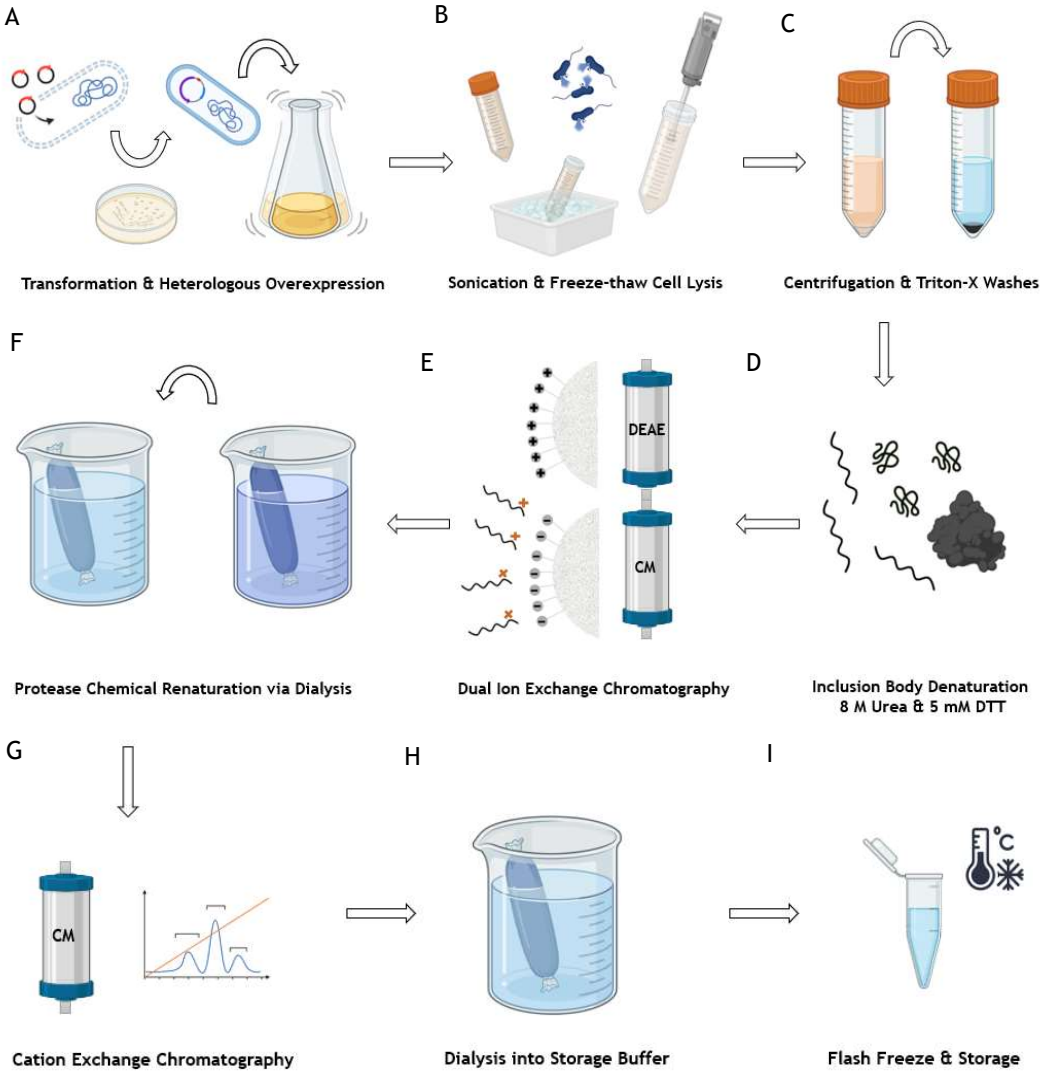


Figure 5: A schematic of the cation-anion exchange chromatography and two-step renaturation procedure required for purification of HIV-1 PR. Key steps include unfolding the PR inclusion bodies using a denaturation buffer B (10 mM Tris, 8 M urea, 5 mM DTT, pH 8) (D). Subsequently, trapping negatively charged contaminants in the positively charged DEAE column while positively charged unfolded proteins, such as HIV-1 PR, bind to the negatively charged CM column (E). After the first purification step, the PR is refolded using two-step dialysis (F). Lastly, the PR is purified using cation exchange and a salt gradient elution to remove any remaining contaminants before dialysing the PR into a storage buffer and flash freezing for storage (G-I). Created with BioRender.com.

The PR was subsequently placed in a second dialysis Buffer D (10 mM sodium acetate, pH 5, 0.01% (w/v) sodium azide) with 2 mM DTT for 16 hr at 4°C using a 1:100 sample/buffer ratio. After dialysis, the protein solution was centrifuged at 18 000 xg for 30 min, at 4 °C to remove any aggregation. After refolding the PR was kept on ice, consistently, to minimise autoproteolysis. This was especially relevant to the active PRs. The refolded solution was purified further using a 5-mL HiTrap® CM Fast Flow column and the GE AKTA Start FPLC system to remove any remaining contaminants. The CM column was equilibrated with Buffer D and 2 mM DTT before the PR was passed through the column. As the pH of Buffer D is below the theoretical pI of PR, the positively charged PR will bind to the (-)-CM-Sepharose column. A salt gradient elution was utilised to separate the remaining contaminants from the purified PR. The collected eluant was dialysed against Buffer D for 4 hours at 4 °C and concentrated using an Amicon® 50-mL stirred-cell concentrator (Merck, Darmstadt, Germany). Samples were taken throughout the purification protocol for assessment of purification steps through tricine SDS-PAGE analysis. The samples were processed and analysed for SDS-PAGE analysis as per section 2.1.2.

2.2.2 Absorbance-based concentration determination and quality assessment of HIV-1 PRs

The Jasco V-550 UV/vis absorbance spectrophotometer (Jasco, Easton, Pennsylvania, USA) was used to measure the UV/vis absorbance spectrum of the four purified PRs between 240 and 360 nm. Scatter contributions as a result of buffer constituents (230 nm absorbance maximum) and protein aggregates (340 nm absorbance maximum) in the absorbance at 280 nm (A_{280}) measurements were corrected according to (Birdsall *et al.,* 1983). The absorbance value at 280 nm was used to estimate the concentration of PR determined from the slope of the linear regression model according to the equation defined by Beer Lambert's Law:

$$A_{280} = \varepsilon c l \qquad \text{[Equation 1]}$$

where A_{280}, ε, c, and l represent absorbance, extinction coefficient ($M^{-1}.cm^{-1}$), concentration (M), and length of the path the light must travel (cm), respectively. Lastly, the A_{280}: A_{260} ratio was calculated and assessed for potential DNA contamination of protein stock.

2.3 Structural assessment of inactive and active constructs of the Subtype C Wild Type and the L38↑N↑L^{+4} hinge mutant PRs

2.3.1 Primary amino acid evaluation of DNA extracted from transformed overexpression BL21 (DƐ3) pLysS *E. coli* cell lines

Plasmid DNA was extracted from the transformed cultures using a GeneJET Plasmid Miniprep Kit (Thermo Scientific®) before being sequenced by Inqaba Biotech® (Pretoria, South Africa) to confirm the genetic contents of the transformed BL21 (DƐ3) pLysS *E. coli* glycerol stocks containing each of the four protease plasmid constructs. The sequencing data was cleaned and analysed in MEGA Software version 11 and aligned using the Clustal Omega tool (Sievers and Higgins, 2014; Tamura *et al.,* 2007).

2.3.2 Secondary structural analysis of HIV-1 PRs using far-ultraviolet circular dichroism (Far-UV CD)

Far-UV CD is a sensitive technique used to determine the secondary structural content of proteins by monitoring the unequal absorption of left-handed and right-handed circularly polarised light that interacts with the peptide backbone (Greenfield, 2006). The main absorbing element in the far-UV region (240 nm – 190 nm), the peptide backbone, is present in ordered secondary structures which when exposed to circularly polarised light generate distinct CD-spectra. Thus far-UV CD allows for the assessment of the global secondary structure content of a protein sample, confirming the protein is correctly folded at a secondary level and revealing any impacts mutations may have on the global secondary structure. The Jasco J-1500 CD Spectrophotometer (Jasco, Easton, Pennsylvania, USA) was used to perform far-UV CD analysis on the purified inactive and active wild-type and mutant PR samples at a protein concentration of approximately 15 µM. Raw data were collected using a 0.5 nm bandwidth with a data pitch of 0.2 nm using a 2 mm path length quartz cuvette at 25 °C. Raw CD data was recorded between 205 nm and 250 nm at 25 °C and manipulated to mean residue ellipticity (MRE) using the following equation:

$$\mathrm{MRE} = \frac{\theta \mathrm{x} 100}{\mathrm{cn}l} \qquad \text{[Equation 2]}$$

Where θ is signal data generated (mdeg), c is the concentration of the protein sample used (mM), n is the total number of amino acids in each PR and l is the path length of light through the cuvette (cm).

2.3.3 Tertiary structural evaluation of HIV-1 PRs by probing tryptophan and tyrosine probes using intrinsic fluorescence

Intrinsic fluorescence is a sensitive technique used to assess the tertiary structural integrity of proteins through the excitation of intrinsic probes containing aromatic rings within their chemical structures such as tryptone, tyrosine, and phenylalanine (Longworth, 1971). The indole ring found within the chemical structure of tryptophan is a dominant source of ultraviolet (UV) absorbance and emission within proteins, with tyrosine having a similar quantum yield. The emission of aromatic structures found in tryptophan and tyrosine is highly sensitive to the local environment and thus can be used to probe potential protein structural changes in the presence of mutations and assess the overall structural integrity of renatured protein at a tertiary level. To investigate the structural integrity of purified and refolded PRs at a tertiary level, a 2-5 µM sample of each PR was excited at both 280 nm and 295 nm. Subsequently, emission spectra between 310 nm and 460 nm were measured with the Jasco J-1500 CD Spectrophotometer (Jasco, Easton, Pennsylvania, USA) using a 5 nm bandwidth with a data pitch of 0.5 nm in a quartz fluorescence cuvette at 25 ℃. Excitation at 280 nm was used to probe both tryptophan and tyrosine residues while excitation at 295 nm was used to probe tryptophan residues alone. The wavelength corresponding to the maximum fluorescence intensity values produced at each excitation wavelength (280 nm and 295 nm) for each protease was determined and compared between PRs and published literature (Zoltán Szeltner and Polgár, 1996).

2.3.4 Quaternary structural evaluation of HIV-1 PRs using size exclusion-high performance liquid chromatography (SE-HPLC)

SE-HPLC makes use of a column composed of pores of a specific size. Solutes are differentially excluded from the pores according to molecular weight (Kilz, 1999). Solutes of larger molecular weight are excluded from the column and elute first while those of smaller molecular weight elute later, due to retardation effects when flowing through the pores. SE-HPLC was used to confirm the predominating quaternary form of each PR, expected to be a homodimer. The PR sample was assessed using a Shimadzu UFLC system (Shimadzu, Kyoto, Japan) and TSKgel® Super SW2000 column (TOSOH Bioscience, Tokyo, Japan) that was equilibrated with Buffer D containing 300 mM NaCl. A higher salt concentration, although not physiologically relevant, was necessary to prevent the PRs from adhering to the column resin and eluting in a broad indistinguishable peak. The column had a calibration range of 5000 - 1.5×10^5 Da to allow for the separation of proteins below 100 kDa. The elution profile of the protein samples was generated through continuous UV-absorbance measurements of eluent samples at 280 nm at a flow rate of 0.2 mL/min for 30 min. Subsequently, the molecular weight of distinguishable peaks was determined through the substitution of the peak-corresponding retention times into the linear regression equation of a standard curve constructed from the elution profile of protein standards size-excluded under the same conditions.

2.4 Global functional assessment of inactive and active wild-type and mutant PRs

2.4.1 Active site titrations of purified HIV-1 PRs against acetyl-pepstatin using Isothermal titration calorimetry (ITC)

ITC is a thermodynamic technique used to determine thermodynamic parameters of interactions in solution such as protein-drug binding. To study protein-protein/protein-ligand interactions without the complications of attaching a probe to one of the biomacromolecules of interest isothermal titration calorimetry (ITC) can be used. The ITC machinery measures the heat released or absorbed upon said interactions (Srivastava and Yadav, 2019). Within the thermal core of an isothermal calorimeter resides a reference cell containing pure buffer and a sample cell containing the sample of interest solubilised in the relevant buffer. A highly calibrated syringe is used to inject stepwise concentrations of the ligand of interest into the sample cell. Thereafter, changes in heat energy as the substrate binds to the target protein of interest as well as the heat required to return the temperature between reference and sample cells to the same is detected.

As PR displays autocatalytic activity, cleaved peptides of the protein may appear in solution and hence obscure the true concentration estimated through spectrophotometric techniques. Therefore, ITC is employed to determine the concentration of active PR in solution by exploiting thermodynamic activities in the manner described above. In this investigation, both active and inactive wild-type and mutant proteases were titrated against acetyl-pepstatin (AP), a weak known inhibitor using the Nano ITC Instrument (TA Instruments, New Castle, USA) (Sayer and Louis, 2009). A concentration of 30 µM of inactive wild-type PR and inactive mutant PR were titrated against 200 µM of acetyl-pepstatin at 20 ℃ with a stirring rate of 200 rpm for a minimum of 30 injections. Similarly, 30 µM of active wild-type PR and 30 µM active mutant PR were titrated against 200 µM and 135 µM of acetyl-pepstatin, respectively, at 20 ℃ with a stirring rate of 200 rpm until the reaction had reached saturation (minimal binding heat is given off as less heat is required to maintain the temperature between the reference and sample cell). Raw thermal data of µcal/sec against time was integrated, where possible, under the respective curve to render an isotherm plot where integration values are plotted against the molar ratio of substrate to the target protein, using the software NanoAnalyze (TA Instruments, New Castle, USA) (Srivastava and Yadav, 2019). The baseline signal was corrected for in each run, by subtracting the heats of dilution achieved from the raw thermal data obtained by titrating the corresponding concentration of acetyl pepstatin used in the active site titration into the buffer. From the isotherm plot, the thermodynamic parameters such as affinity constant (K_d), the molar ratio (n), enthalpy (ΔH), and ΔS (Entropy) were derived and the ΔG (Gibbs Free Energy) and were calculated where possible using the following equation (Matthews, 2000):

$$\Delta G^\circ = \Delta H^\circ - T\,\Delta S^\circ \qquad \text{[Equation 3]}$$

where ΔG° is the Gibbs free energy (J), ΔH° is enthalpy (J), T is the temperature (K) and ΔS° is entropy ($J.K^{-1}$).

Specifically, the molar ratio (n) at the center of the binding isotherm that reveals the reaction stoichiometry was used to correct the concentration of PRs as follows (Sherry *et al.,* 2020):

$$[C]_{\text{active PR}} = [C]_{\text{observed}} \times n \qquad \text{[Equation 4]}$$

where $[C]_{\text{active PR}}$ (M) is the true concentration of active PR, $[C]_{\text{observed}}$ (M) is the concentration determination by spectrophotometric methods, and n is the molar ratio. The corrected concentration represented the true concentration of correctly folded and active PR necessary for the optimisation of downstream experiments.

2.4.2 PR activity evaluation using a FRET-based activity assay

Enzyme activity of the inactive and active wild-type and mutant PRs was measured by monitoring the hydrolysis of the fluorogenic substrate that mimics a cleavage site of the subtype C wild-type Gag polypeptide (Figure 6). The sequence of the fluorogenic substrate (Abz-Arg-Val-Nle-Phe-(NO_2)-Glu-Ala-Nle-NH2) (Synpeptide, Shanghai, China), mimics the capsid/p2 cleavage site while making use of fluorescence resonance electron transfer (FRET) in the following manner. Before cleavage the fluorogenic 2-aminobenzoyl group (Abz) remains in close contact with the quencher p-nitrophenyl (Phe(NO_2)), reducing fluorogenic emission. Cleavage of the peptide minimizes the quenching effect and results in a fluorescent signal from Abz emission which was monitored through excitation at a wavelength of 337 nm and an emission wavelength of 425 nm. All spectrofluorometric enzyme assays were performed under steady-state kinetics and in triplicate for averaging. All fluorescence data was baselined using the reaction buffer. In this investigation, the standardised fluorogenic substrate was used to determine the specific activity and k_{cat} of the active wild-type and mutant PRs. Furthermore, used to assess any catalytic activity associated with the inactive wild-type and mutant PRs. The HIV-1 PR performs optimally at pH 5, at 25 °C, with ionic strength no lower than 0.1 M and no higher than 1 M, with higher ionic strength preferred (Szeltner & Polgár, 1996). To determine the specific activity of the active PRs, enzyme concentrations of 0, 10, 20, 30, 40, and 50 nM were incubated with 500 µM of the substrate in reaction buffer (0.1 M NaCl, 50 mM sodium acetate, pH 5.0) for a total of 60 seconds. The initial reaction velocity ($\mu mol.min^{-1}$) was plotted as a function of the amount of enzyme (mg) to produce a linear regression model from wherein the slope represents the specific activity of the enzyme ($\mu mol.min^{-1}.mg^{-1}$). Subsequently a derived plot of initial reaction velocity ($\mu mol.min^{-1}$) as a function of enzyme concentration (µM) allowed for the determination of the k_{cat}.

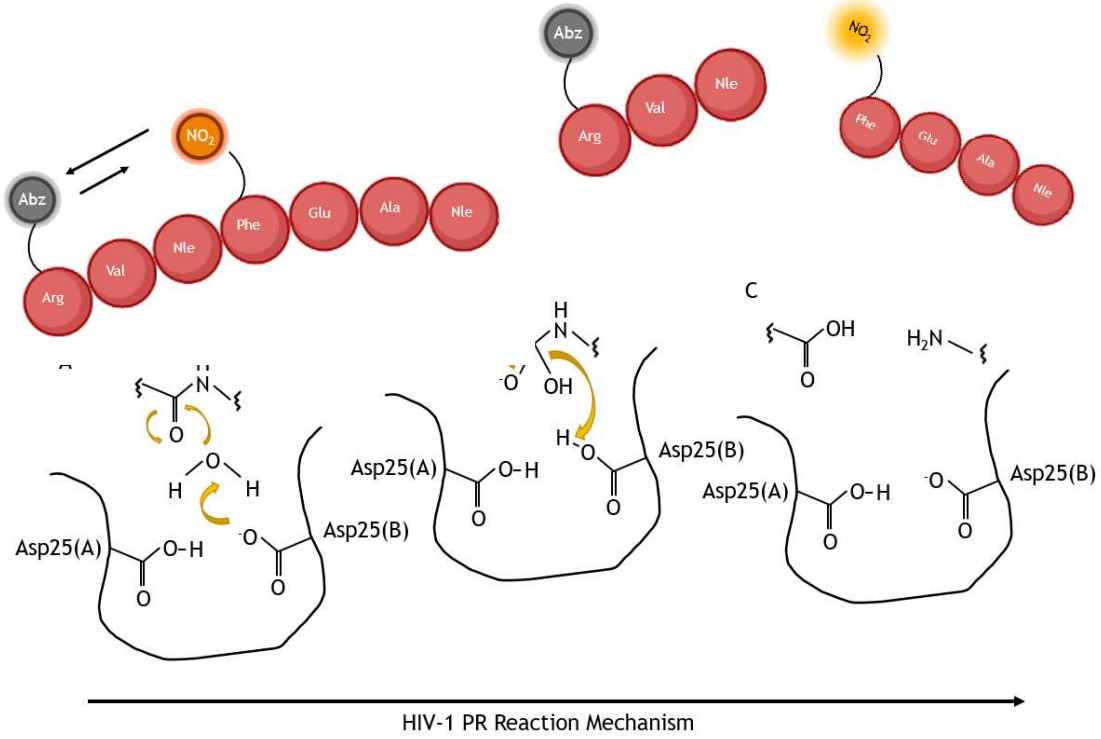


Figure 6: A schematic of the HIV-1 PR reaction mechanism in relation to the FRET-based enzyme assay. The PR makes use of water in a hydrolysis reaction that results in amid bond cleavage. The fluorogenic substrate (Abz-Arg-Val-Nle-Phe-(NO_2)-Glu-Ala-Nle-NH_2) (Synpeptide, Shanghai, China), that mimics the capsid/p2 cleavage site, makes use of fluorescence resonance electron transfer (FRET) in the following manner. Before cleavage the fluorogenic 2-aminobenzoyl group (Abz) remains in close contact with the quencher p-nitrophenyl (Phe(NO2)), reducing fluorogenic emission. Cleavage of the peptide minimises the quenching effect and results in a fluorescent signal from Abz emission which was monitored through excitation at a wavelength of 337 nm and an emission wavelength of 425 nm.

2.5 Global stability studies of active wild type and active mutant HIV-1 PRs in the absence and presence of known inhibitors using thermal shift assays

The thermal stability of active wild-type and active mutant PRs was assessed in the presence of one weak inhibitor (acetyl pepstatin) and two strong inhibitors (DRV and SQV) using a thermal shift assay. The thermal stability of inactive PRs was not assessed due to the lack of binding to the acetyl pepstatin in section 2.4.2. Two 96-well qPCR plates, one for each active PR, were set up as follows. A total of 25 µL of the sample was pipetted into each well containing 10 µM of the respective PR, 5 µL of SYBR orange dye, and varying concentrations of ligand (0, 1, 2, 5, 10, 30, 50, 75 µM). Samples for PRs and acetyl pepstatin were run in duplicate while samples for DRV and SQV were run in triplicate. First, the PR was pipetted, following the relevant ligand, and the remaining volume was made up with a storage buffer. The photosensitive dye was added to the wells in a dark room and incubated for 1 hr. SYBR orange is a protein-specific dye that binds to hydrophobic patches on the protein. The melting of the protein can be monitored by the fluorescence signal produced as the dye binds to more hydrophobic patches exposed as the protein population denatures due to increasing temperature. The plate was then analysed using the CFX 96 Thermocycler (Bio Rad, California, USA), by monitoring the fluorescence signal as the samples were melted from 20 ℃ to 95 ℃. The relative fluorescence units (RFU) from each melt curve were derived and plotted against temperature (℃) using the CFX Maestro software (Bio Rad, California, USA).

2.6 Design, synthesis, and purification of HIV-1 Gag p2/NC-mimicking peptides

2.6.1 Design and optimisation of peptide empirical properties

The sequence data of the wild-type was obtained from Prof. Lynn Morris (NICD, Johannesburg, South Africa) and the mutant gag proteins were obtained from Dr. G. Hunt and Ms. J. Ledwaba (NICD, Johannesburg, South Africa). An alignment of the sequence data revealed several substitution mutations within the gag cleavage sites as well as a PTAPPAE duplication (Williams *et al.,* 2019). The p2/NC gag cleavage site within the mutant gag displays only 70% sequence homology to the wild-type p2/NC gag cleavage site. Furthermore, genetic studies have revealed that the p2/NC site is significantly polymorphic within subtype C and between subtypes of the virus. The p2/NC site is also the site of gag cleavage initiation and plays a role in regulating the rate of cleavage of the other gag cleavage sites. Given all these factors, the p2/NC cleavage was placed under investigation in this study. A peptide mimicking the p2/NC cleavage site of the wild-type gag and mutant gag was designed using computational tools that predict the empirical properties of peptides. To ensure that the peptides exhibit ideal empirical properties for synthesis and *in vitro* application, the following parameters need to be taken into consideration: sequence homology to the mutant p2/NC and wild-type p2/NC sites, amino acid composition, extinction capabilities, sequence length, pI, net charge, acidic or basic qualities,

buffer solubility, hydrophobicity, aliphatic percentage, and estimated half-life. The following prediction tools were used to assist in obtaining the optimal properties for peptide synthesis: Pep-Calc.com by Innovagen, GenScript Peptide Property Calculator, Thermo Fisher Scientific Peptide Analysing Tool, and ProtParam on the ExPASy portal (Gasteiger *et al.,* 2005; Lear & Cobb, 2016).

2.6.2 Synthesis and purification of HIV-1 Gag p2/NC-mimicking peptides

The p2/NC mimicking peptides of both the wild-type and mutant p2/NC gag cleavage sites were synthesised using solid phase peptide synthesis (Figure 7). The peptide synthesis was headed by Dr. Thabo Peme at the School of Chemistry (University of the Witwatersrand, Johannesburg), and the protocol was adapted from standard laboratory protocols (Peme *et al.*, 2021). Peptide synthesis was initiated by swelling 600 mg Fmoc-rink amide resin in 10 mL of Dimethylformamide (DMF) for 20 min. The DMF was subsequently removed by suction whereafter a 20% piperidine solution in DMF was added and bubbled for 5 min with inert nitrogen gas. The resin solution was filtered out and was subjected to 30 sec 5 mL DMF washes. The first amino acid was coupled to the resin as follows. A concentration of 0.2 M of the first amino acid was mixed with 0.19 M of a coupling reagent (containing HBTU and 1M of DIPEA in 6 mL of DMF). When added to the reaction vessel, along with an extra 6 mL of DMF, the reaction was bubbled with inert nitrogen gas for 45 min. Subsequently, the resin was washed three times with DMF and two times with Dichloromethane (DCM) and dried with suction. The loading capacity was performed to calculate the exact moles of the amino acid attached to the resin. The resin-bound amino acid was then treated with 20% piperidine in DMF for 15 min to remove the Fmoc group protecting the first amino acid carboxyl group. Subsequently, 5 mL of DMF was used to wash the resin 5 times. The second amino acid as above until the full peptide was synthesised. For larger amino acids which can be challenging to couple, a double coupling was performed. Between the coupling of amino acids samples were taken for analysis by liquid chromatography mass spectrometry (LC-MS) to ensure the peptide chain was growing with each coupling.

The final resin-bound peptides were washed 3 times with 5 mL of DMF and 3 times with 5 mL of DCM. To cleave the peptides off the resin, the peptide-resin was incubated for 3 hrs in a 10 mL solution of 94% trifluoroacetic acid (TFA), 2.5% of distilled water, and 1% triisopropylsilane (TIS). The cleavage solution was filtered out and the peptide-resin was washed with 5 mL of TFA and filtered before being aliquoted into centrifuge tubes. Cold ether was used to precipitate out crude peptide whereafter it was isolated by centrifugation at 5000 rpm for 10 min. The cold ether precipitation was repeated until no precipitant was present. The crude mutant peptide was then solubilised in 40% acetonitrile; whereas the highly insoluble wild-type peptide was subjected to a solubility trial using a range of acid and basic solutions, organic and inorganic, before being solubilised in a large volume of 100% acetonitrile.

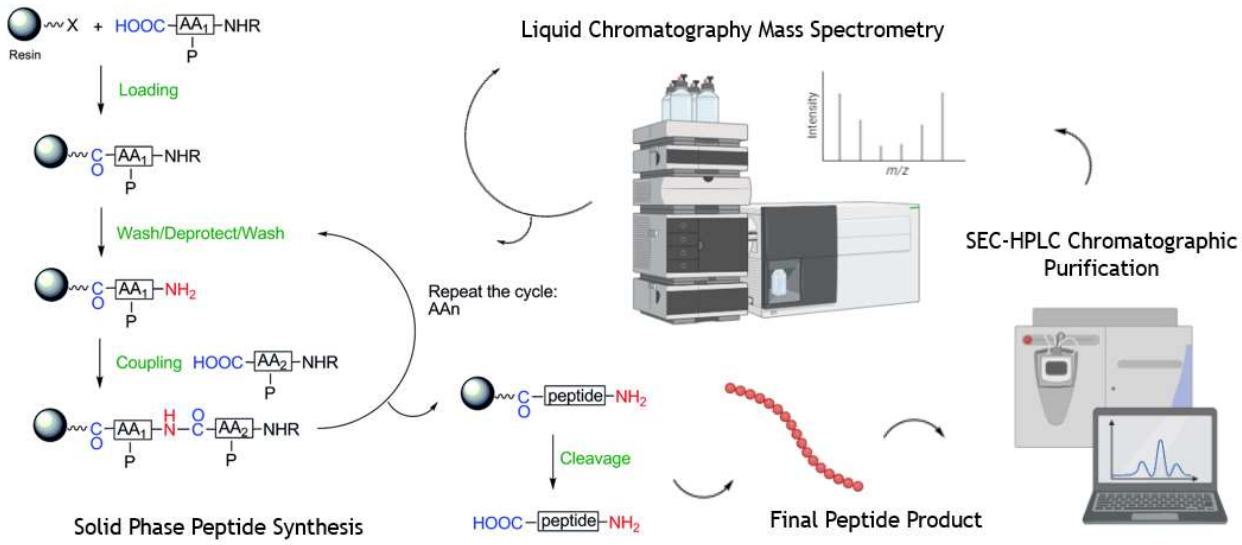


Figure 7: A schematic of solid phase peptide synthesis, SEC-purification and evaluation using LC-MS. LC-MS was used to monitor the coupling of every 2-3 amino acids. After the complete synthesis of the peptides, purification occurred via SE-HPLC followed by the generation of an LC-MS profile to evaluate the final purified peptide product.

Purification of the mutant peptides was achieved via an Agilent 1260 Infinity semi-preparative HPLC system with a UV/VIS detector and an automated fraction collector on a Kinetix® 5 μm BC18, 100 Å (250 ×230 nm) column. A two-buffer system was employed; Buffer A consisted of 0.1% formic acid in distilled water and a buffer consisted of 0.1% formic acid in acetonitrile. A flow rate of 20 mL/min or 15 mL/min and UV wavelengths of 215 and 254 nm were utilised. The wild-type peptide could not be purified due to its propensity to aggregate in volumes lower than 150 mL.

2.6.3 Evaluation of purified peptides using LC-MS and far-UV CD

Analytical LC-MS analysis was carried out on an Ultra High Performance Liquid Chromatography (Thermo Scientific Ultimate 3000, RS diode array detectors)-High Resolution Mass Spectrometer (Bruker Compact quadruple time of-flight) coupled to Diode Array (215 and 254 nm). A two-buffer system was used with a flow rate of 0.3 mL/min on a C18 column (5 μm, 100 Å, 4.60 mm × 150 mm). The buffer system included buffer 1 containing water with 0.1% formic acid and buffer 2 containing acetonitrile with 0.1% formic acid. Each signal produced was assessed as to whether it was a multiple of the total molecular weight of the peptide and thus the desired peptide primary amino acid sequence. Subsequently, the purified peptide was subjected to Far-UV CD analysis as per section 2.3.2 to decipher any predominating secondary structure.

2.7.1 Isothermal titration of a soluble HIV-1 Gag p2/NC-mimicking peptide against active wild-type and mutant HIV-1 PR

To assess whether the newly synthesised peptides, representing an alternative gag cleavage site to section 2.6, could be used as a biological tool — binding studies were carried out using ITC. As the wild-type peptide could not be solubilised and reliably purified, only the mutant p2/NC mimicking peptide was used for binding studies using ITC against both the active wild-type and active mutant PRs. The inactive PRs were not used for binding studies as the D25A mutation was found to influence acetyl pepstatin binding within the active site and thus could impact the quantification of other binding molecules, especially those that could have a weak and only marginally stable binding interactions such as peptide-protein interactions. To preliminarily probe a potential interaction between wild-type and mutant active PRs with the mutant p2/NC mimicking peptide, 20 μM PR was titrated against 50 μM, 100 μL and 200 μM of peptide in the Nano ITC Instrument (TA Instruments, New Castle, USA). The titrations were carried out at 20 °C with a stirring rate of 200 rpm for a minimum of 30 injections. Raw thermal data were processed and analysed as in section 2.4.1 to deduce whether the peptide was binding to the PRs (Sayer & Louis, 2009).

2.7 Evaluation of a soluble HIV-1 Gag p2/NC-mimicking peptide binding to active Subtype C

2.7.2 Monte Carlo-based docking simulations of the mutant HIV-1 Gag p2/NC-mimicking peptide against active wild-type and mutant HIV-1 PR

To further evaluate the results obtained from peptide-PR ITC-based binding studies, both active wild-type and mutant PR structures and models, respectively, were subjected to peptide-protein docking simulations using HPEPDOCK (Zhou *et al.,* 2018). The HPEPDOCK program is a hierarchal algorithm designed for the docking of peptides to protein receptors. Within the program, 1000 possible peptide conformations are generated for rigid-body docking against a rigid receptor. The large ensemble of peptide conformations generated and docked against the receptor is necessary to account for the inherent flexibility of the peptide that small molecular docking programs cannot. This program also performs a global search for the most likely binding site before docking multiple peptide conformations before complex reduction and generation and evaluation of binding energy scores. In this study, the synthetic mutant peptides were subjected to docking simulations as well as theoretical models of p2/NC sequences.

An I-TASSER-generated mutant p2/NC peptide model was docked to an active wild-type PR structure and active mutant PR homology model was generated in SWISS-MODEL using the 3u71 crystal structure as a template (Yang *et al.,* 2015). The docking tool was first validated by redocking a subtype B PR-p2/NC peptide complex and comparing the generated model to the solved crystal structure (1kj7). Both structural characteristics as well as a calculated RMSD were evaluated using Chimera. Thereafter, the synthetic mutant peptide models (used in ITC studies), models of the lone p2/NC sequences (corresponding to the wild-type and mutant sequence), and models of the p2/NC sequences with each mutant-related polymorphism in isolation were subjected to docking simulations against wild-type and mutant PRS in HPEPDOCK. The s (kcal/mol), a score representative of predicted complex stability, of the highest-ranked complex models were compared. The structural orientation and other characteristics such as polar contacts were compared between systems and to the solved subtype B PR-peptide complexes (Prabu-Jeyabalan *et al.,* 2002).

Chapter 3. Results and Discussion

3.1 Comparative optimisation of heterologous overexpression of wild-type and mutant PRs in BL21 (DƐ3) pLysS *E. coli* cells

Expression of HIV-1 PRs is characterised by cytotoxic effects and low expression yields (especially PRs with a complex profile of mutations). Within this study, optimisation of (both active and inactive) mutant PR expression was investigated in BL21 (DƐ3) pLysS *E. coli*. The choice of this cell line was imperative to controlling for expression "leakage", a cytotoxic factor that could influence cell viability during overexpression and thus overall PR yields. This strategy had been adopted for the expression of active and inactive wild-type PRs, previously (Naicker *et al.,* 2013, Sherry *et al.,* 2020). As the expression and purification protocols for the inactive and active wild-type PRs were established, resident glycerol stocks were used for downstream overexpression of these PRs. While optimising the expression of mutant PRs, this study, assessed the use of inactive constructs in parallel. The experimental use of inactive constructs could be a potential strategy for overcoming bottlenecks in PR yields and downstream complications in experimental studies such as PR autocatalysis and instability.

3.1.1 Cell viability assessment of transformed cell lines by growth curve analysis

Following mutant PR transformations, cell growth of the inactive wild-type, inactive mutant, and active mutant PR cell lines were monitored as a function of optical density at 600 nm over 9 hrs (Figure 8). The inactive wild-type PR expression trials were reperformed as a comparative control to assess the overexpression performance associated with the newly transformed mutant PR cell stocks. All three cell lines reached the early exponential growth phase (OD_{600} of 0.4) at the same rate, 2 hrs after inoculation. The inactive wild-type PR cell line grew at a greater rate than that of the inactive and active mutant PR cell lines, reaching an OD_{600} of 1 (stationary phase of growth) approximately 35 minutes earlier (Figure 8). An essential factor was whether all cell lines presented a healthy and consistent growth rate up until induction goals. Cell growth variation after induction is natural amongst separately transformed cell lines and can be attributed to the differences in the metabolic state across cell stocks (Hall *et al.,* 2014). No notable differences in cell growth before and after induction goals suggested the cell line was controlling for unnecessary basal PR expression. Thus, induction was set to be carried out 2 hrs after inoculation (OD_{600} of 0.4) across all PR-producing cell lines.

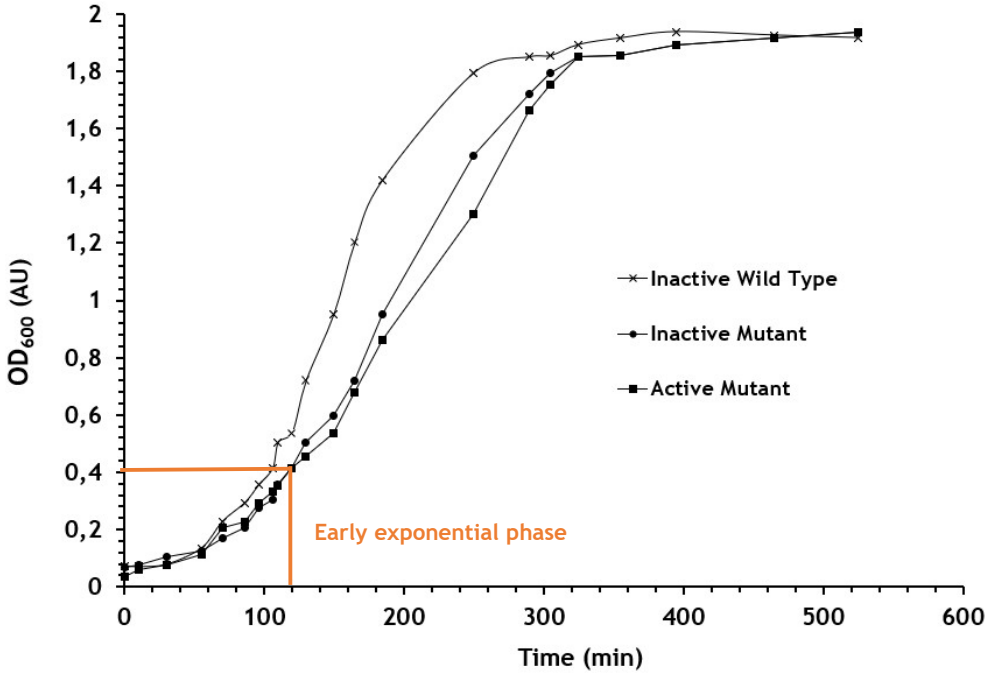


Figure 8: Growth curves of HIV-1 PR expression cell lines. The growth of inactive wild-type PR as well as inactive and active mutant PR cell lines was monitored as a function of OD_{600} over 9 hrs. All three cell lines reached early exponential growth phase (OD_{600} of 0.4) at approximately the same rate, 2 hrs after inoculation. Thus, induction was set to be carried out 2 hours after inoculation (OD_{600} of 0.4) across all PR-producing cell lines.

3.1.2 Identification and comparison of optimal IPTG concentrations and induction periods using SDS-PAGE analysis

Following cell viability, inducer (IPTG) concentration and induction period were investigated in relation to optimal PR yield. Concentrations ranging from 0 to 1 mM of IPTG as well as induction periods of 0, 2, 4, 6, and 22 hours were considered. Overnight expression (22 hrs) was excluded due to the increased protein contamination that would need to be purified out, not suitable for a non-specific purification technique such as ion exchange. The highest yield of all PRs was found in the insoluble fraction as expected (Volontè *et al.,* 2011). Furthermore, across the 0 mM IPTG concentrations and 0 induction period, there was no noticeable PR expression supporting the purpose of the BL21 (DƐ3) pLysS *E. coli* strain in minimising any unnecessary basal expression, which would generate cytotoxic effects.

Overall inactive wild-type and inactive mutant PRs expressed at higher yields than the active mutant PR suggesting the D25A mutation influences PR expression yields (Figure 9A-H). HIV-1 PRs possess proteolytic activity that is not sequence-specific by the variety of processable gag cleavage site sequences (Fun *et al.,* 2012). Knocking out the activity of the enzyme could downregulate any cytotoxicity associated with residual soluble and active HIV-1 PRs in the expression cell line. Furthermore, amino acid availability could be an influential factor. Future trials could involve the supplementation of active PR-expressing cell cultures with aspartate. The inactive wild-type PR expressed noticeably higher yields than the inactive mutant PR suggesting PR identity can influence expression yields (Figure 9A-D). Thus, the use of inactive constructs of HIV-1 PR could be a potential strategy for increasing working concentrations of PR for downstream experimental procedures.

The optimal inducer concentration and periods were decided upon qualitative analysis of the 11 kDa protein yield from crude extracts of overexpression culture resolved on a 16% tricine SDS-PAGE gel. Inactive wild-type PR displayed the highest level of expression after 6 hrs of induction with an IPTG concentration of 0.6 mM, as expected (Sherry *et al.,* 2020) (Figure 9D). The PR bands resided closer to the 15 kDa standard as opposed to the 10 kDa standard bands. It was determined that the highest yield of inactive mutant PR could be achieved after 2 hrs of induction at an IPTG concentration as low as 0.2 mM Figure 9F). The yield of active mutant PR was the lowest, displaying no noticeable expression until after 4 hrs of induction. As longer induction periods showed little impact on the yield of active mutant PR obtained, the induction time was limited to 4 hrs at 0.5 mM IPTG (Figure 9K). An increase in cytotoxicity associated with the active mutant PR could be delaying PR production. Future considerations would include cell systems that are preferred for toxic proteins (lemo21 (DƐ3) possessing tuneable expression) or recombinant tags that promote PR secretion. Overall induction periods were limited to 6 hours or less as well as IPTG concentrations to 0.6 mM or less to minimise cell cytotoxicity and maximise PR yield in an economic time-effective manner.

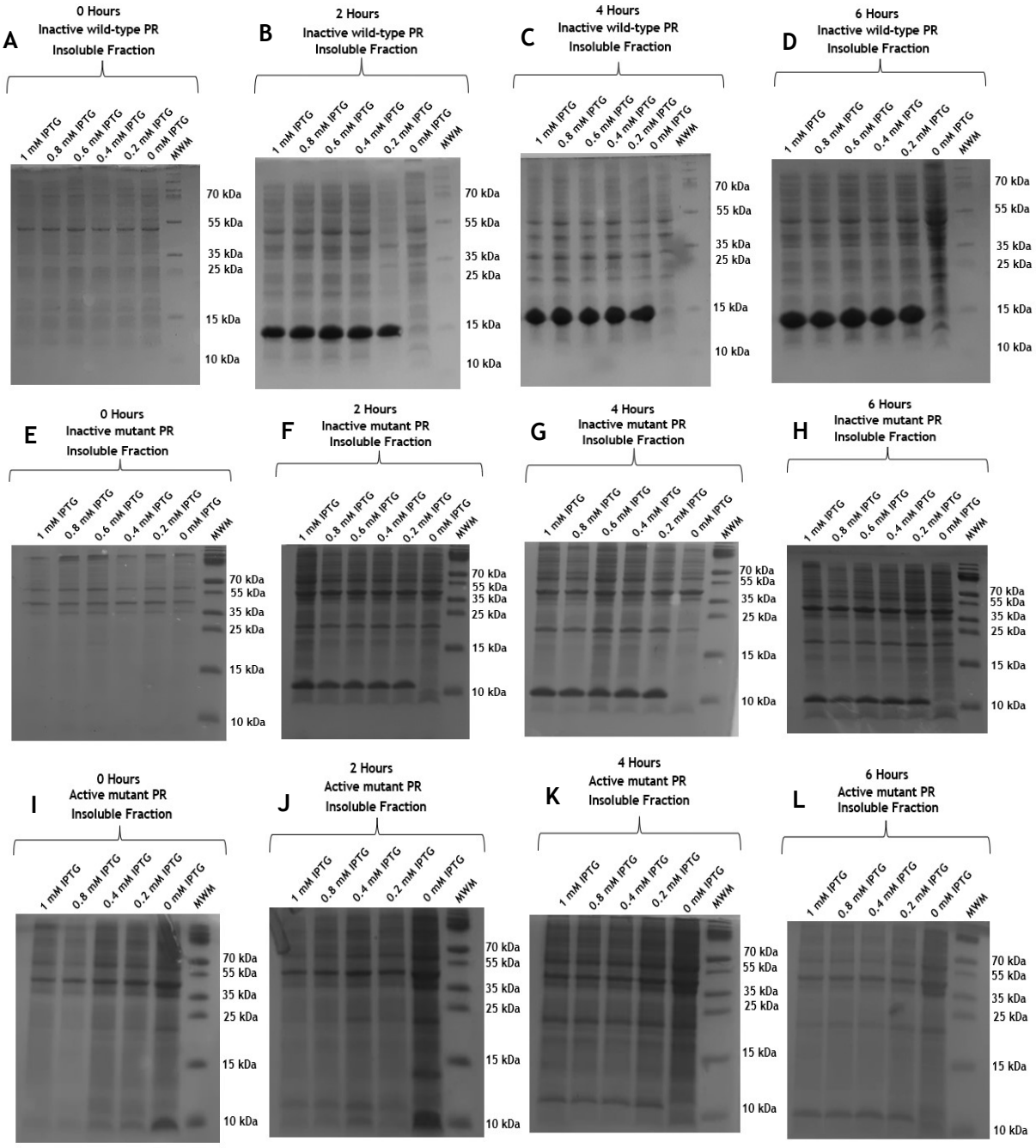


Figure 9: SDS-PAGE analysis of crude extracts of PR expression under varying concentrations of IPTG and varying induction periods. Inactive wild-type PR displayed the highest level of expression after 6 hours of induction with an IPTG concentration of 0.6 mM (D) (Sherry *et al.,* 2020). The highest yield of inactive mutant PR could be achieved after 2 hrs of induction at an IPTG concentration as low as 0.2 mM (F). The yield of active mutant PR was the lowest, displaying no noticeable expression until after 4 hrs of induction. As longer induction periods showed little impact on the yield of active mutant PR obtained, induction time was limited to 4 hrs at an IPTG concentration of 0.5 mM (K).

3.2 Ion-exchange chromatographic purification and chemical renaturation of HIV-1 PRs

3.2.1 Purification and renaturation of inactive and active wild type and mutant PRs using cation-anion exchange and a two-step dialysis

Following heterologous overexpression under the appropriate conditions, each PR was purified using a two-step ion exchange procedure. Protein contamination and PR yields could be monitored throughout the purification via UV-absorbance (280 nm) and SDS-PAGE. Levels of protein contamination varied across the four different PR expressions in both steps 1 and 2 of ion exchange purification. The first purification step was not a suitable indicator of PR yield as this eluant still consisted of unfolded protein contaminants that were subsequently precipitated out in chemical renaturation via dialysis. Furthermore, as the protein content is unfolded and aromatic residues are largely solvent exposed, the absorbance signal will appear exaggerated. After step 1, the protocol by Sherry *et al.,* 2020 was adapted to reduce protein contamination, prevent PR aggregation, and promote mutant PR stability upon refolding. The Trizma-based refolding buffers were replaced with a formic acid buffer, at pH 3, followed by a sodium acetate buffer, at pH 5. Changes to the buffer systems acted to reduce aggregation during PR refolding. The first step followed by the dialysis steps made significant contributions to remove protein contamination and concentrating PR yield as seen in SDS-PAGE analysis (Figures 11, 13, 15 & 17).

The second purification step was crucial to ensuring 98% pure HIV-1 PRs, by resolving residual contamination from semi-purified PR solution using a salt gradient elution, after refolding. Furthermore, the protocol by Sherry *et al.,* 2020 was adapted to include a CM column in this step as opposed to a DEAE column to increase purity and concentrate the PR upon collection as opposed to collecting dilute samples through the DEAE flow through. Final PR yields were consistently low as seen by the final mAu signal from each PR's second purification step chromatogram at approximately 10-30% salt in the gradient elution (Figures 10, 12, 14 & 16). The inactive PRs displayed higher intensity bands and greater absorbance values (with the mutant PR lower than the wild-type PR) most probably due to the knockout of activity which minimised any autocatalysis events throughout the purification that would reduce PR yield (Figures 11,13,15 & 17). The active PR band yields (Figures 15 & 17) and UV-absorbance (Figures 14 & 16) detectability were notably lower than inactive PRs approaching the end of purification. This elicited the need for a stepwise 100% salt elution step to concentrate PR elution to a mAu signal that could be reliably detected. The inactive wild-type PR resolved at a higher molecular weight than the other three PRs as seen by its direct comparison against the active wild-type PR (Figure 11). The discrepancy was attributed to the plasmid design which included intrinsic gag regions specifically cleaved during the refolding of the active wild-type PR (see section 3.3.1). Overall, the purification protocol was successful in achieving 98% pure HIV-1 PR for both inactive and active wild-type and mutant PRs as seen by predominating 11 kDa protein bands as well as the resolved peaks in chromatograms from the second purification step.

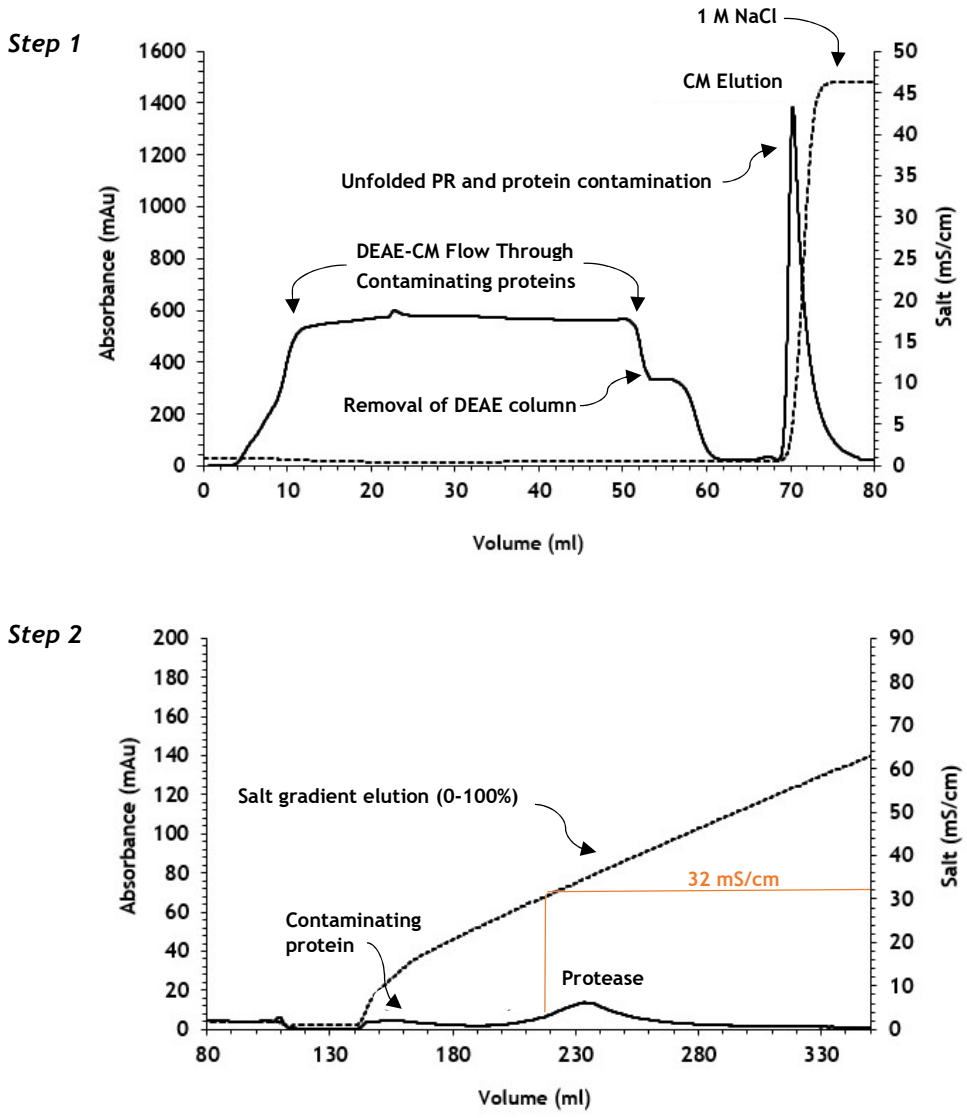


Figure 10: Chromatograms of ion exchange purification of the inactive wild-type HIV-1 PR. Step 1 involved the purification of the unfolded protein solution using a dual cation-anion exchange before a 100% salt elution and refolding of the semi-purified PR solution. Step 2 included a salt gradient elution of residual protein contaminants and PR bound to a CM column. The protein contaminants were eluted at 10 mS/cm followed by gradual elution of the PR starting at 32 mS/cm. As the level of expressed PR was high enough to be detected reliably in mAu, a gradual gradient (0-100 % salt) was used throughout to resolve the PR from contaminating proteins.

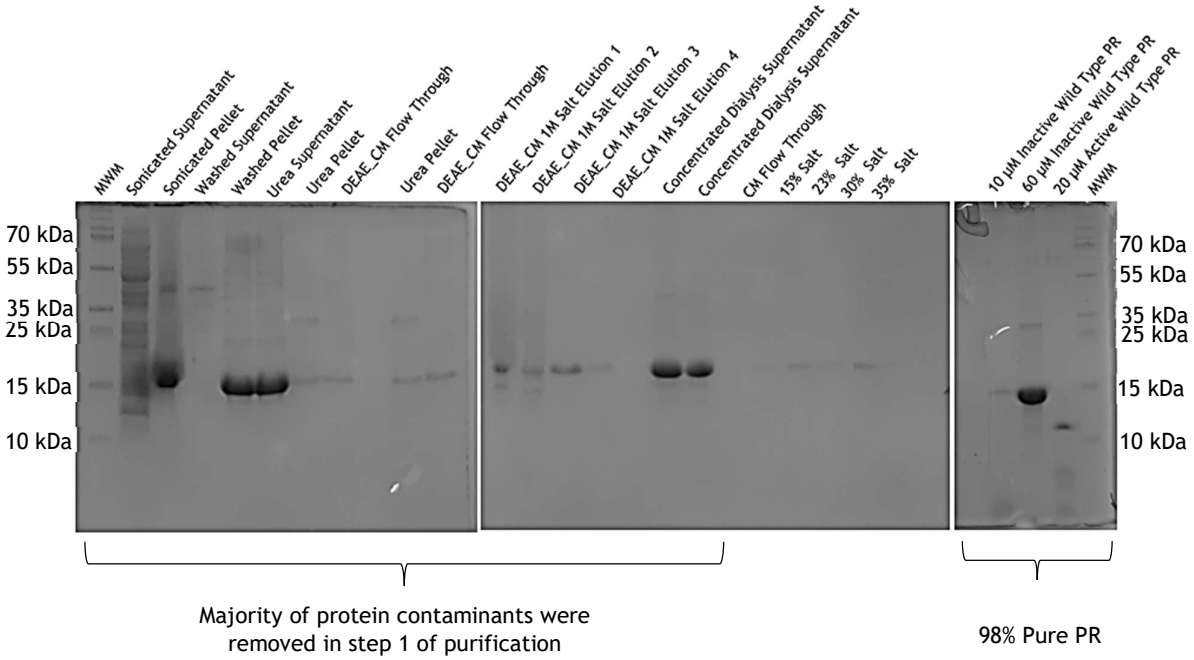


Figure 11: SDS-PAGE analysis of samples taken along the purification of the inactive wild-type PR. PR loss was minimised across the purification profile as seen by the maintained band intensity from the beginning of purification to the end of purification. The discrepancy in molecular weight was noted when purified inactive wild-type was compared to purified active wild-type. The discrepancy was explained by primary sequence analysis of extracted, sequenced, and translated plasmid DNA. This size discrepancy was attributed to included gag regions that could not be cleaved upon refolding as is achieved during the refolding of the active wild-type PR (*see section 3.3.1*). This PR was found to elute around 32 mS/cm (at approximately 30% of the salt gradient) during purification step 2. The inactive wild-type PR was found to be 98% pure with the highest yield of all four PRs.

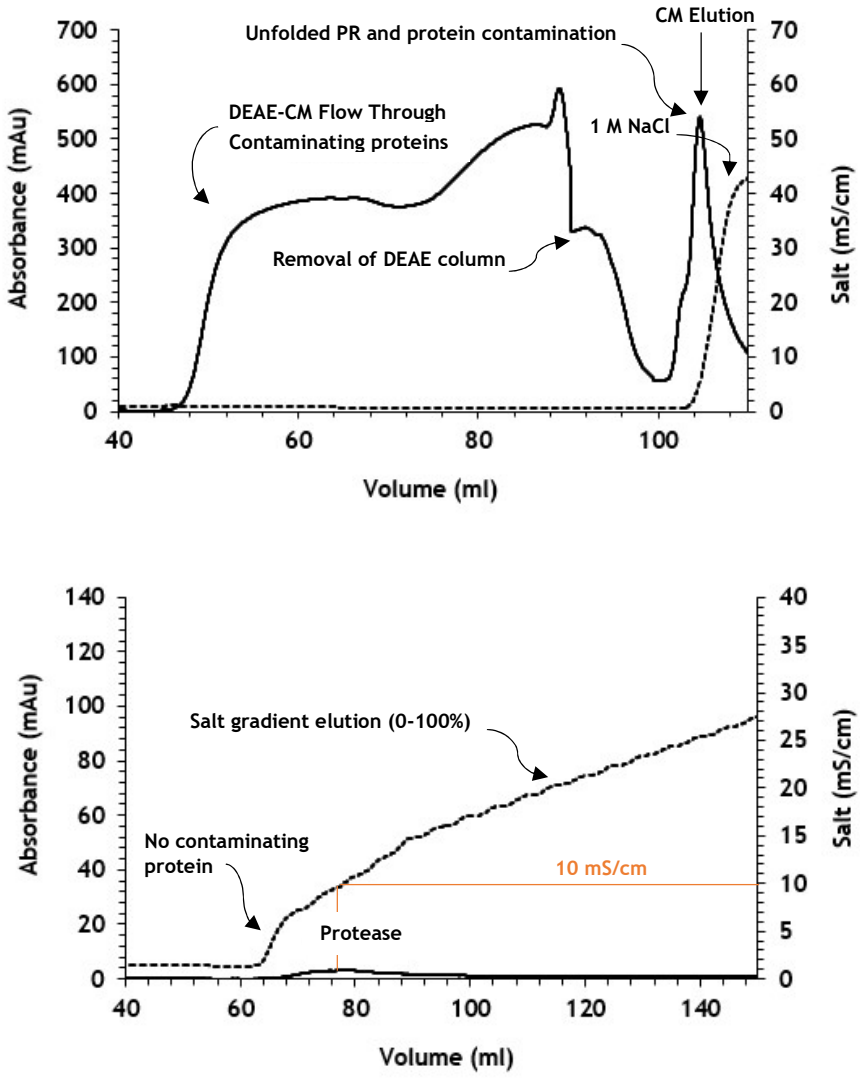


Figure 12: Chromatograms of ion exchange purification of the inactive mutant HIV-1 PR. Step 1 involved the purification of the unfolded protein solution using a dual cation-anion exchange before 100% salt elution and refolding of the semi-purified PR solution. Step 2 included a salt gradient elution of residual protein contaminants and PR bound to a CM column, which was proven unnecessary. No noticeable contamination eluted before this PR, in step 2, which eluted earlier than expected at approximately 10 mS/cm. Step 1 was effective at removing all protein contaminants associated with the inactive mutant PR, thus only a single purification step was required. The low IPTG concentration and short induction period could contribute to a limited load of contaminating proteins expressed at basal levels in the cells. As the level of expressed PR was high enough to be detected reliably in mAu, a gradual gradient (0-100% salt) was used throughout to resolve the PR from contaminating proteins.

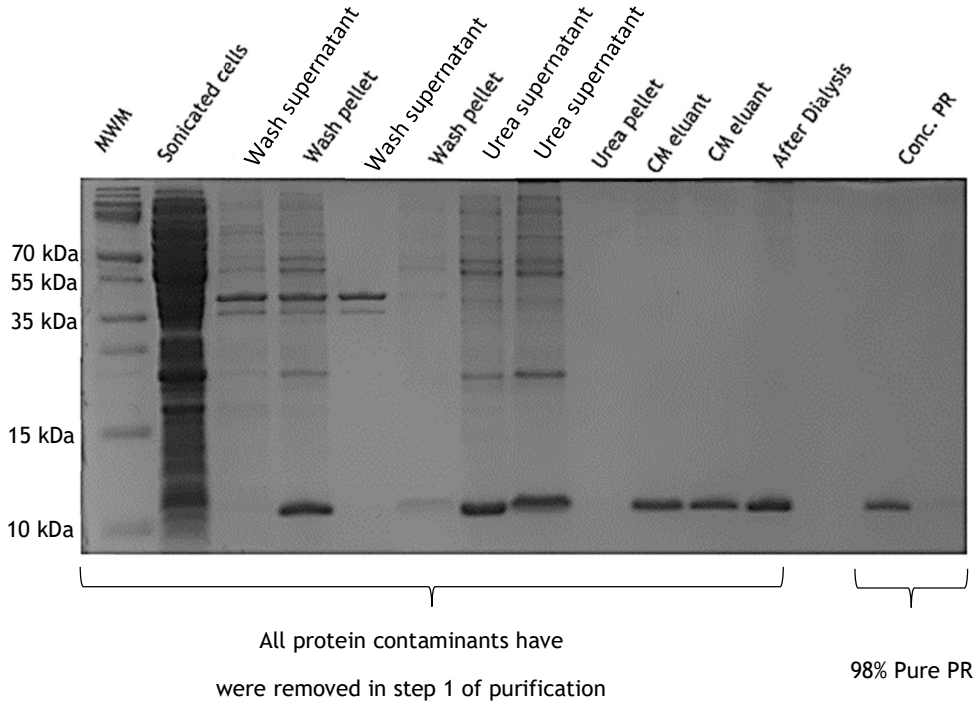


Figure 13: SDS-PAGE analysis of samples taken along the purification of the inactive mutant PR. PR loss was minimised across the purification profile as seen by the maintained band intensity from the beginning of purification to the end of purification. The inactive mutant PR was found to be 98% pure and resolved at the expected molecular weight of 11 kDa. This PR was found to elute earlier than expected at approximately 10 mS/cm with no noticeable residual contaminating protein eluting before this PR. The low IPTG concentration and short induction period could contribute to limited load of contaminating proteins expressed at basal levels in the cells. The inactive mutant PR was found to be 98% pure with the second-highest yield of all four PRs.

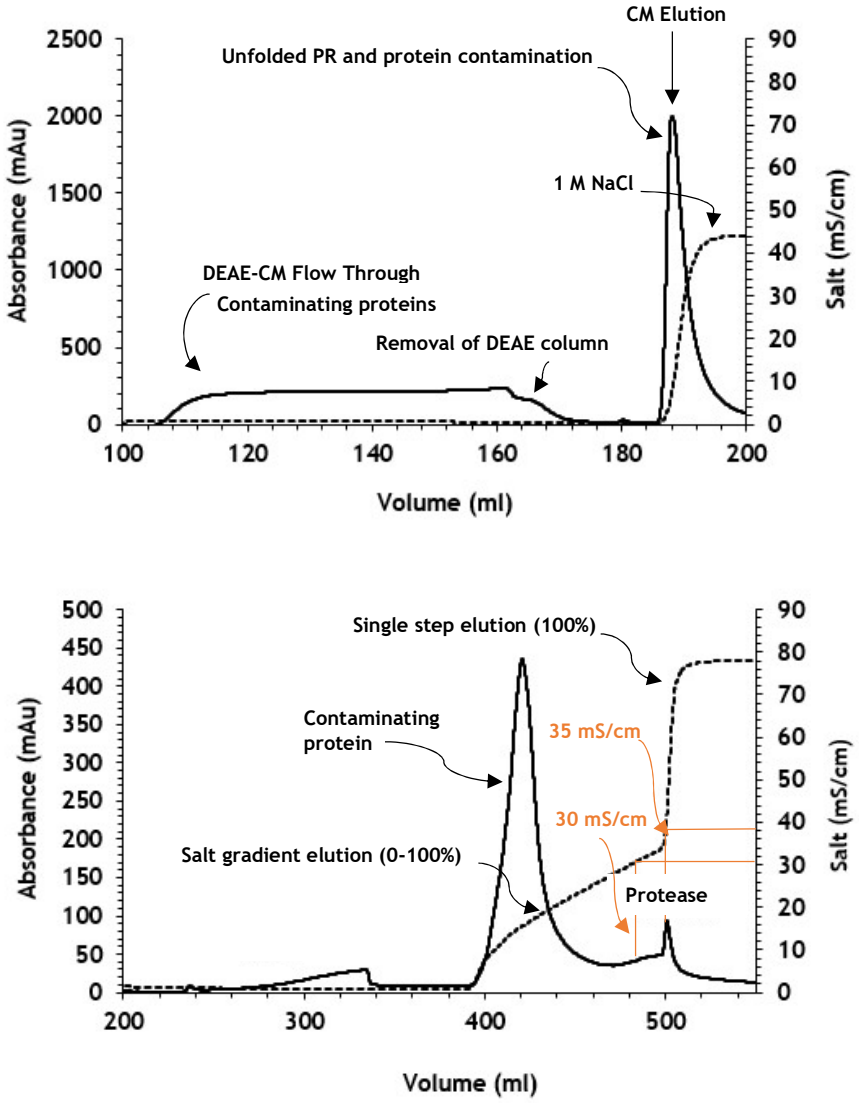


Figure 14: Chromatograms of ion exchange purification of the active wild-type HIV-1 PR. Step 1 involved purification of the unfolded protein solution using a dual cation-anion exchange before 100% salt elution and refolding of the semi-purified PR solution. Step 2 included a salt gradient elution of residual protein contaminants and PR bound to a CM column. The protein contaminants were eluted at approximately 10 mS/cm followed by stepwise elution of the PR starting at 30 mS/cm, whereafter this PR elution peaked at 35 mS/cm. As the level of expressed PR was not high enough to be detected reliably in mAu over a gradual elution, the stepwise salt elution (0-30%, 100% salt) was used to generate a concentrated elution of this PR.

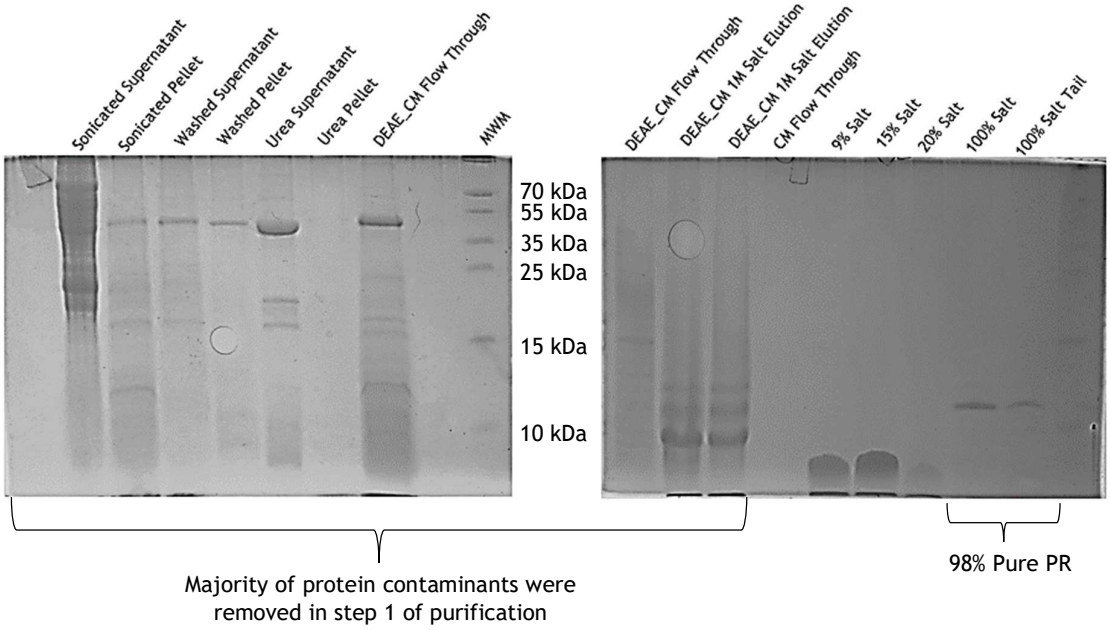


Figure 15: SDS-PAGE analysis of samples taken along the purification of the active wild-type PR. PR loss was minimised across the purification profile as seen by the maintained band intensity from the beginning of purification to the end of purification. The active wild-type PR was found to be 98% pure and resolved at the expected molecular weight of 11 kDa. This PR elution began at 30 mS/cm, whereafter a single-step elution was introduced to generate a concentrated peak that could be reliably detected using SDS-PAGE. The active wild-type PR was found to be 98% pure with the third-highest yield of all four PRs.

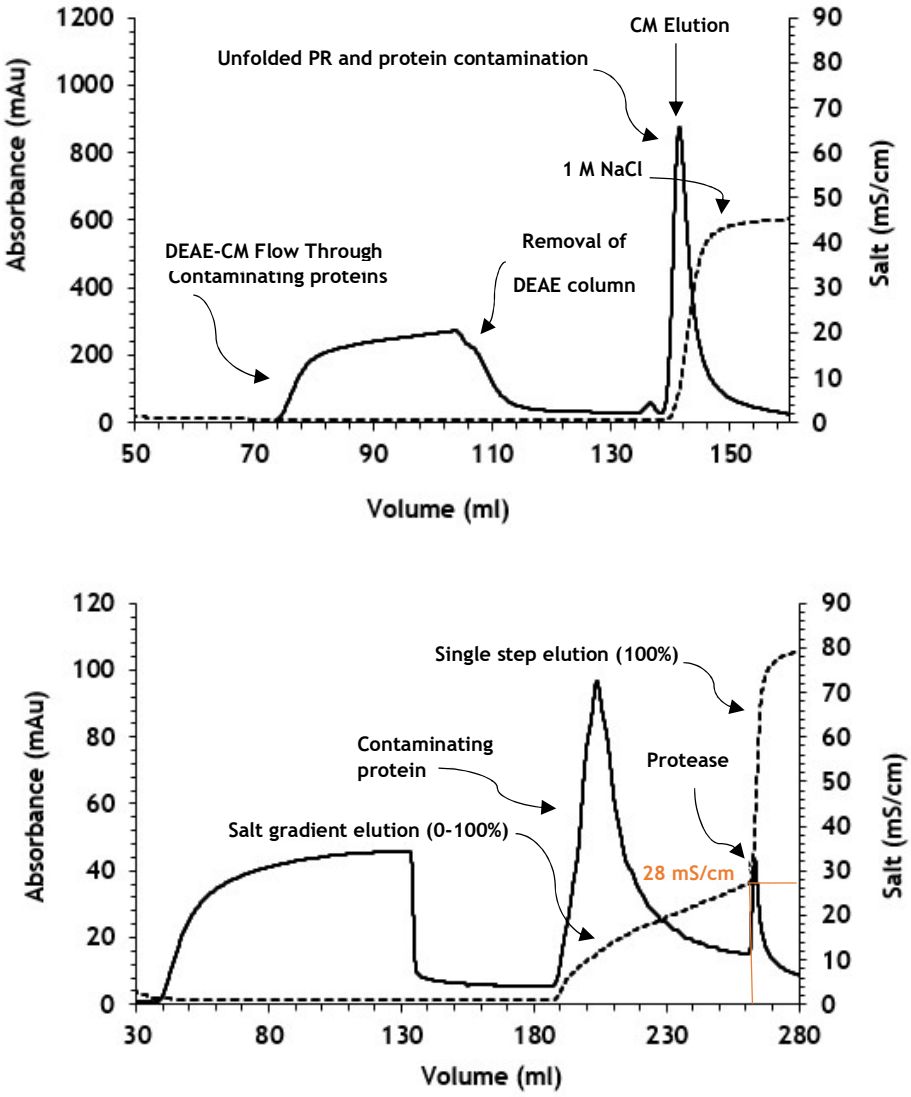


Figure 16: Chromatograms of ion exchange purification of the active mutant HIV-1 PR. Step 1 involved purification of the unfolded protein solution using a dual cation-anion exchange before 100% salt elution and refolding of the semi-purified PR solution. Step 2 included a salt gradient elution of residual contaminants and PR bound to a CM column. The contamination was eluted at approximately 10 mS/cm followed by stepwise elution of the PR starting at 28 mS/cm, whereafter this PR peaked at 35 mS/cm. As the level of expressed PR was not high enough to be detected reliably in mAu over a gradual elution, the stepwise salt elution (0-30%, 100% salt) was used to generate a concentrated elution of this PR.

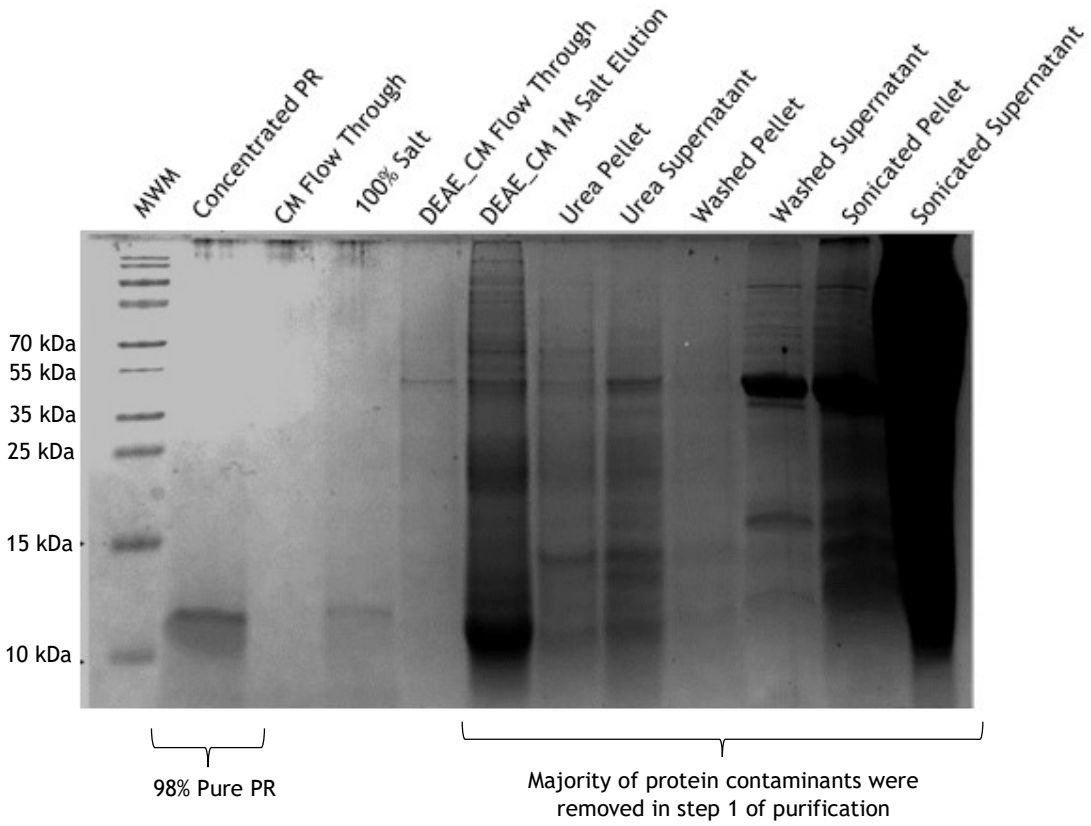


Figure 17: SDS-PAGE analysis of samples taken along the purification of the inactive wild-type PR. PR loss was minimised across the purification profile as seen by the maintained band intensity from the beginning of purification to the end of purification. The protein contamination was significant concerning this PR, proving the second purification step to be essential in comparison to the other PR purifications. The active wild-type PR was found to be 98% pure and resolved at the expected molecular weight of 11 kDa. This PR elution began at 28 mS/cm, whereafter a single-step elution was introduced to generate a concentrated peak that could be reliably detected using SDS-PAGE. The active wild-type PR was found to be 98% pure with the third-highest yield of all four PRs.

3.2.2 Determination and comparison of purified PR concentrations and sample quality using absorbance spectrophotometry

UV-absorbance spectra of all four purified PRs were taken between wavelengths 240 nm and 360 nm to assess protein concentration (at 280 nm), DNA contamination (at 260 nm), and protein aggregation (at 340 nm) (Figure 22) (Grimsley & Pace, 2003). Both the inactive wild-type and inactive mutant PRs could be concentrated down to a working volume of 4 mL with a UV-absorbance reading of approximately 0.6 Au. This value converted using Beer Lambert's law rendered an approximated PR concentration of 27 µM. The active mutant and active wild-type PRs could be concentrated to a working volume of 2.5 mL with UV-absorbance values of approximately 0.3 Au and 0.5 Au, respectively. Using Beer Lambert's law, the working concentrations of active mutant and active wild-type PR could be determined to be approximately 13 µM and 22 µM, respectively. Overall, the inactive PRs could be expressed and purified to higher concentrations and volumes than active PRs, with the active mutant PR achieving the lowest working concentration and working volume. This emphasised the strategic advantage of inactive constructs of the two PRs in an experimental setting, should the D25A mutation not affect downstream evaluations and the clinical relevance of subsequent experimental studies.

Aside from A_{280}, UV-absorbance readings at 340 nm were a minimal 0,058 Au, 0,062 Au, 0,043 Au, and 0,057 Au across the inactive mutant, inactive wild-type, active mutant, and active wild-type PRs, respectively. On average the overall estimated concentration of aggregation after PRs were concentrated and before storage at -80 °C, was 2.5 µM. Lastly, to ensure the quality of the PR samples used for downstream studies, the ratio of UV-absorbance at 260 nm to UV-absorbance at 280 nm was calculated for each PR. A ratio above 1 would be indicative of DNA contamination, negatively impacting the signal-noise ratio of downstream analysis of the PRs. The A_{260}:A_{280} ratio for the working samples of the inactive mutant, inactive wild-type, active mutant, and active wild-type PRs were found to be approximately 0.75, 0.75, 0.85, and 0.79, respectively. As these values did not exceed a value of 1, DNA contamination was kept to a minimum and the DNase I coupled with $MgCl_2$ supplementation effectively degraded and reduce the presence of DNA in the protein samples. The higher ratios associated with the active PRs (with the active mutant PR having the highest ratio) could be attributed to the general reduction in achievable protein concentrations. The following working concentrations were used in downstream global structural comparisons of the HIV-1 PRs as these techniques are not concentration dependent.

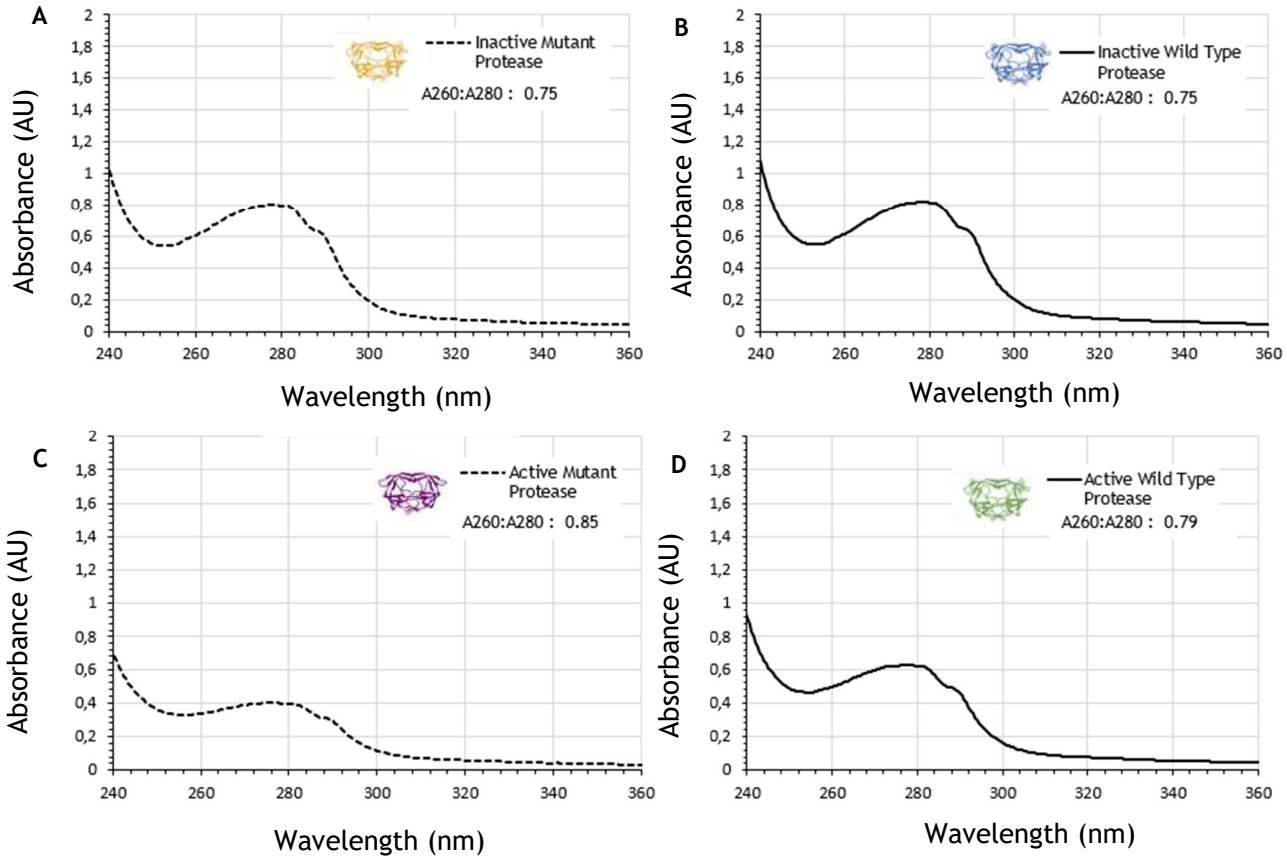


Figure 18: UV-absorbance spectra purified, working concentrations of the inactive mutant PR (A), inactive wild-type PR (B), active mutant PR (C), and active wild-type PR (D). UV-absorbance spectra of all four purified PRs were taken between wavelengths 240 nm and 360 nm to assess protein concentration (at 280 nm), DNA contamination (at 260 nm), and protein aggregation (at 340 nm) (Grimsley and Pace, 2003). Both the inactive wild-type and inactive mutant PRs could be concentrated to a working volume of 4 mL with an A_{280} of 0.6 Au, approximately 27 µM. The active mutant and active wild-type PRs could be concentrated down to a working volume of 2.5 mL with A_{280} readings of 0.3 Au and 0.5 Au, respectively. The working concentrations of active mutant and active wild-type PR were estimated to be 13 µM and 22 µM, respectively. On average the overall estimated concentration of aggregation after concentration and before storage at -80 °C, was 2.5 µM. The A_{260}:A_{280} ratio for the working samples of the inactive mutant, inactive wild-type, active mutant, and active wild-type PRs were found to be approximately 0.75, 0.75, 0.85, and 0.79, respectively. Overall, DNA contamination and protein aggregation were minimised while adequate concentrations and volumes of all four HIV-1 PRs were achieved.

3.3 Global structural comparison of inactive and active wild-type and mutant PRs

3.3.1 Primary amino acid evaluation of DNA extracted from BL21 (DƐ3) pLysS *E. coli* expression cell lines

DNA was extracted and sequenced from all four PR-plasmid-containing cell lines to ensure the resident cell stocks contained the wild-type inactive and wild-type active PR plasmids. Furthermore, to ensure the newly transformed cells contained the mutant inactive and mutant active PR plasmids. DNA sequences were prepared and translated using ExPASy Translate and aligned using ClutslOmega to confirm and compare the amino acid content corresponding to each PR, *in silico* (Figure 19A). Upon alignment, the inactive PR sequences were confirmed to possess the characteristic D25A mutation when compared to the active PR sequences (Figure 19B). Furthermore, the mutant PR amino acid sequences consisted of the four background mutations (Figure 19B 1,3,5 & 6) and double insertion mutation (Figure 19B 4) when compared to the wild-type sequences. Directly between the start and stop codon for both mutant PRs was the primary amino acid sequence characteristic of HIV-1 PR, beginning with proline 1 (P1) and ending with phenylalanine 101 (F101), denoted phenylalanine 99 (F99) in the wild-type sequence not possessing an insertion mutation. The active wild-type primary amino acid differed by possessing 10 extra amino acids before P1 and F99, a representation of the autocatalytic PR cleavage sites found in the viral gag polypeptide (SFNF/PQIT and TLNF/PISP).

During the replication cycle of the virus, the PR is translated within a single, long polypeptide chain possessing all functional and structural viral protein domains (Chatterjee *et al.,* 2005). The domain associated with the PR folds and auto-processes itself from the polypeptide by targeting and cleaving the bridging sequences (Figure 19A). The original wild-type active PR plasmid construct was designed to mimic this process (Naicker *et al.,* 2013). The inactive wild-type PR plasmid construct was previously adapted from this sequence by mutating the D25 to an A25 (Sherry *et al.,* 2020). Considering the D25A mutation was implemented to knock out the catalytic activity of the PR, the inactive wild-type is unable to cleave off the extra 20 amino acids providing a reason for the size discrepancy associated with SDS-PAGE analysis (and SE-HPLC in section 2.3.4). Evidentially, it can be concluded that the D25A mutation effectively knocks out PR activity. However, further studies such as global structural and functional evaluations would have to be performed to confirm the extra 20 amino acids do not have an impact on the other structural and functional characteristics of the inactive wild-type PR. The mutant inactive PR, within this study, was designed to exclude these bridging sequences as should the wild-type inactive for future reference. In conclusion, the source for the size discrepancy associated with the inactive wild-type PR could be attributed to the plasmid design which was adapted from the active wild-type PR plasmid design. Overall, DNA sequencing and subsequent translational manipulation, reliably confirmed, that all PR plasmid constructs possessed the appropriate single nucleotide polymorphisms and insertion mutations characteristic of the PR phenotypes under investigation.



Figure 19: (A) Amino acid sequences of inactive and active wild-type PRs translated from DNA sequencing of extracted plasmid DNA. (B) Alignment of amino acid sequences of inactive and active wild-type and mutant PRs. Inactive PRs possessed the characteristic D25A mutation (B2) when compared to the active PR sequences. The active wild-type primary amino acid differed by having a stretch of 10 extra amino acids before P1 and F99, a representation of the autocatalytic cleavage sites found in the viral gag polypeptide (SFNF/PQIT and TLNF/PISP) (Naicker *et al.,* 2013). The wild-type PR plasmid was adapted from this sequence by mutating the D25 an A25 (Sherry *et al.,* 2020). The D25A mutation was implemented to knock out the catalytic activity of the PR, thus, inactive wild-type is unable to cleave off the extra 20 amino acids corresponding to the size discrepancy in SDS-PAGE analysis. The mutant PR amino acid sequences consisted of the four background mutations (1,3,5 & 6) and double insertion mutation (4) under investigation in this study.

3.3.2 Secondary structural analysis of inactive & active wild-type & mutant PRs using far-UV CD

To confirm the secondary structural integrity and assess the influences of mutations between all four PRs, far-UV CD spectra were generated between wavelengths of 190 nm and 250 nm. Circularly polarised light interacts with a chromophore which in the case of proteins is the peptide backbone. This backbone is present in ordered regular structures such as α-helices and β-strands and turns, which each produce a specific CD spectrum allowing for confirmation and identification of the principal secondary structure/s of a protein (Greenfield, 2006). To ensure the spectra produced by each PR were not influenced by the noise produced from buffer artifacts the corresponding HT (V) readings were monitored simultaneously with the CD readings, ensuring the spectra fell within the threshold of 700 V (Figure 20B, Figure 20D, Figure 21B & Figure 21D) (J. Miles & A. Wallace, 2016). This ensures the readings are fitted within a low noise: signal ratio and few artifacts were contributing to the overall trend found within the far-UV CD spectra.

Far-UV CD analysis of all four PRs produced spectra characteristic of a protein that is predominantly composed of β-strands as indicated by the negative trough between 216 nm and 220 nm (Figure 20A, Figure 20C, Figure 21A & Figure 21C) (Rodger, 2013). These results were expected after consideration of crystal structures and published data that suggest the structure of HIV-1 PR is around 48% β-sheet, 4% α-helix, and 48% random coils (Mosebi *et al.,* 2008; Naicker *et al.,* 2013; Zondagh *et al.,* 2018; Williams *et al.,* 2019). Overall, no significant difference in global secondary structure between inactive and active PRs as well as between the wild-type and the mutant PRs was found. Thus, the mutations attributed to the mutant PRs (K20R, E35D, R57K, and V82I, as well as L38↑N↑L) do not disrupt the overall secondary structure of the PR as the characteristic HIV-1 PR spectra is maintained concerning the mutant PRs (Figure 20A & Figure 21A). Furthermore, the D25A mutation and the extra amino acids of the inactive wild-type do not influence the overall secondary structure either (Figure 20C). All four PRs were found to be structurally sound and characteristic of HIV-1 PR at a global secondary level.

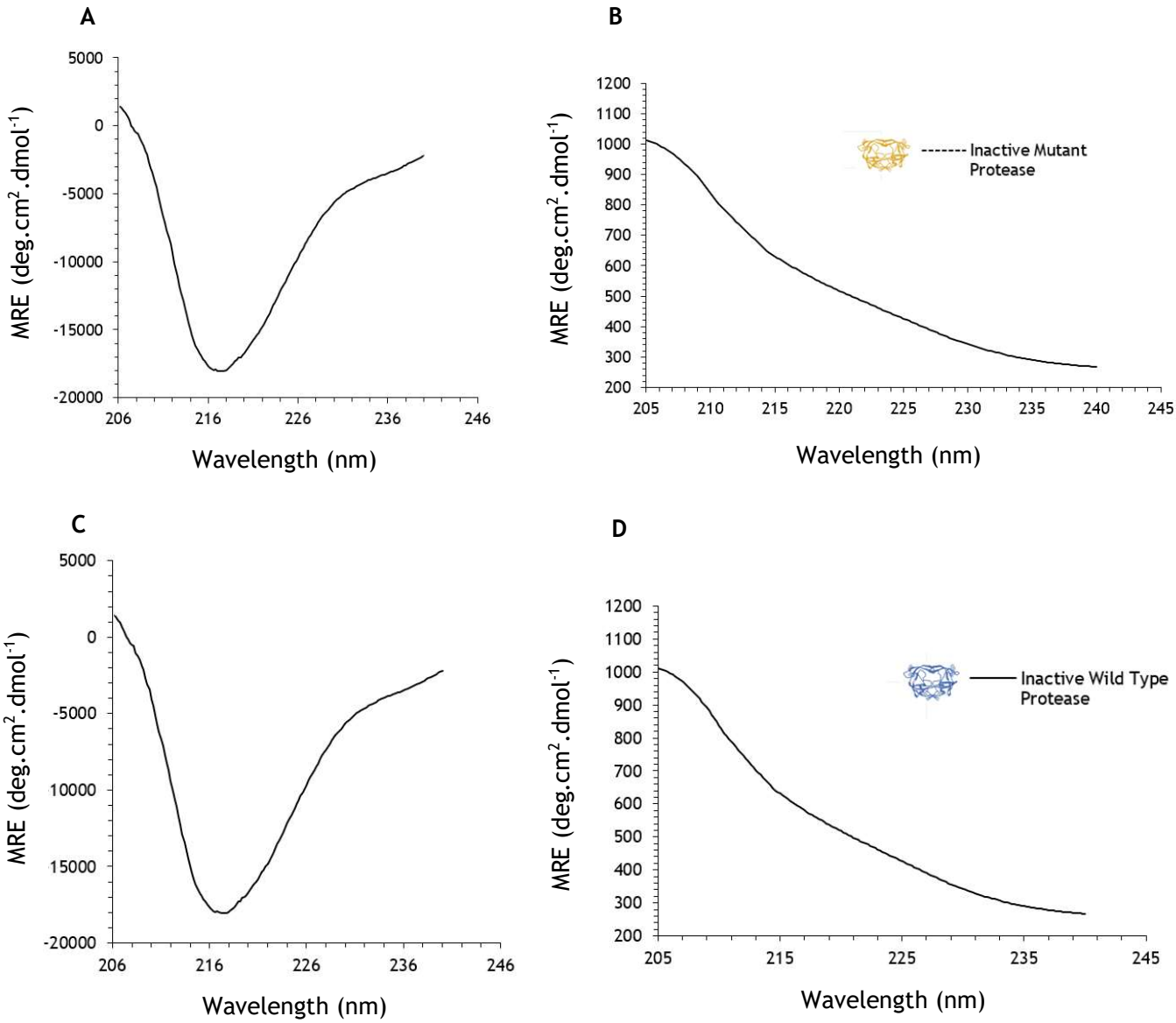


Figure 20: Far-UV CD spectra of the inactive mutant PR (A) and inactive wild-type PR (C) and the corresponding HT (V) plots for the inactive mutant PR (B) and inactive wild-type PR (D). The spectra were reliably fitted within the HT threshold of 700 V, wherein there is a low signal-noise ratio. Both spectra were characteristic of a protein that is predominantly composed of β-strands as indicated by the negative trough between 216 nm and 220 nm. These results were expected after consideration of published data of wild-type and mutant HIV-1 PRs which stipulated that the structure of PR is approximately 48% β-sheet, 4% α-helix, and 48% random coils (Mosebi *et al.,* 2008; Naicker *et al.,* 2013; Zondagh *et al.,* 2018).

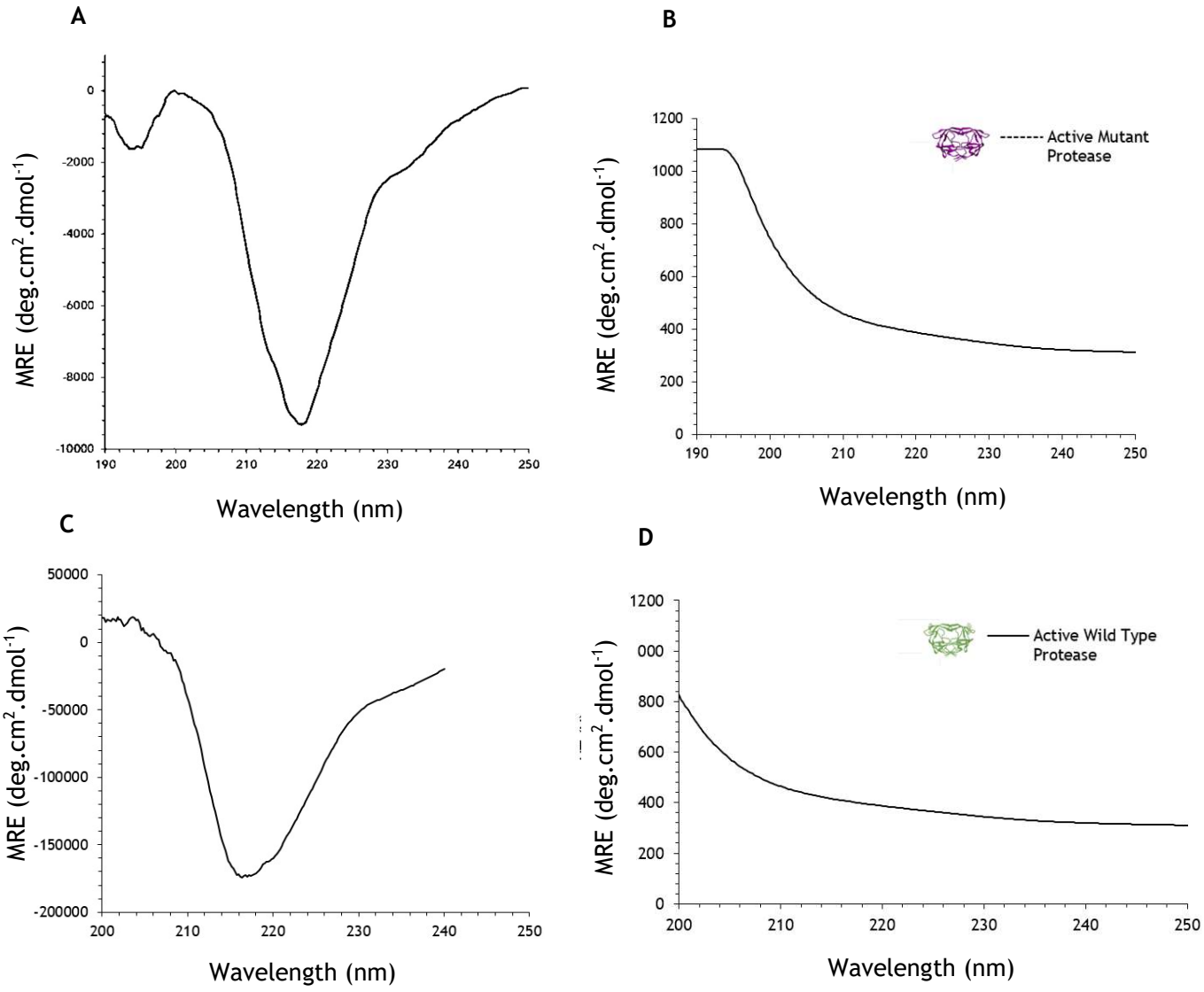


Figure 21: Far-UV CD spectra of the active mutant PR (A) and active wild-type PR (C) and the corresponding HT (V) plots for the active mutant PR (B) and active wild-type PR (D). The spectra were reliably fitted within the HT threshold of 700 V, wherein there is a low signal-noise ratio. Both spectra were characteristic of a protein that is predominantly composed of β-strands as indicated by the negative trough between 216 nm and 220 nm. These results were expected after consideration of published data of wild-type and mutant HIV-1 PRs which stipulated that the structure of PR is approximately 48% β-sheet, 4% α-helix, and 48% random coils (Mosebi *et al.,* 2008; Naicker *et al.,* 2013; Zondagh *et al.,* 2018).

3.3.3 Tertiary structural analysis of inactive & active wild-type & mutant PRs using intrinsic fluorescence

Intrinsic fluorescence was utilised to assess differences in global tertiary structure between the four PRs. All four HIV-1 dimeric PRs possess four tryptophan residues (located in the hinge region and dimer interface) and two tyrosine residues (located in the hinge region) (Figure 18A). According to the subtype C wild-type structure (3u71), all 6 intrinsic probes resided close to the surface of the PR in local environments that are relatively solvent exposed. The degree of hydrophobicity (Figure 18C) of the four tryptophan residues and 2 tyrosine residues is consistent while a noticeable difference can be attributed to the B-factors (Figure 18A) associated with each residue. The B-factor can be indicative of the relative flexibility of a region of a protein (Sun *et al.,* 2019). Trp6 on each chain is located within the dimer interface, a relatively static area with lower overall B-factors in comparison to the hinge region. The Tyr59 is almost buried within the hinge, reducing flexibility, but the Trp42 residues can be denoted with the highest B-factors of all 6 probes and thus significantly more flexibility. All 6 probes were used to indicate whether the PR is natively folded, aggregated, or misfolded across the inactive, active, wild-type, and mutant PR identities. However, the Tyr59 (a residue largely buried in the hinge already) and more likely the Trp42 (displaying greater flexibility and surface exposure) were essential to evaluating whether the insertion mutation, located near those probes, is influencing the overall tertiary structure and stability of the hinge in the mutant PRs when compared to the wild-type PRs.

Upon excitation of the inactive PRs at a wavelength of 280 nm, minimal differences were seen in the emission spectra (Figure 23). Both inactive PRs produced an emission spectrum with a maximum fluorescence intensity at a wavelength of 346.5 nm (Figure 23A), characteristic of most HIV-1 PRs (Szeltner & Polgár, 1996). Similarly, both the active wild-type and mutant PR produced a predominating fluorescence peak at a wavelength of 347 nm (Figure 23B), indicating a native global tertiary fold. This suggested that under excitation at 280 nm, the influence of the substitution mutations (K20R, E35D, R57K, and V82I) and insertion mutation (L38↑N↑L) on the global tertiary structure of the PR is minimal. Furthermore, when comparing active PR spectra to inactive PR spectra, a minimal influence on the global tertiary structure of the PR was associated with the D25A mutation. However, a shoulder at 331 nm was noticeable in the spectrum of the active mutant PR suggesting some species of tryptophan and tyrosine residues existed in different local environments (Figure 23B2). A decrease in wavelength is associated with a decrease in local environment polarity (i.e., burying of the tryptophan and/or tyrosine residues). This could be attributed to an increased propensity of the active mutant PR to aggregate under *in vitro* conditions, potentially reduced protein stability associated with the presence of the mutations, or a reduced refolding efficiency during the chemical renaturation via dialysis. The high sensitivity of the active mutant PR to *in vitro* refolding was further noted in subsequent analyses, by a large degree of unfolded species producing fluorescence maxima at red-shifted wavelengths, 349 nm, 366 nm, and 385 nm (Figure 23B2).

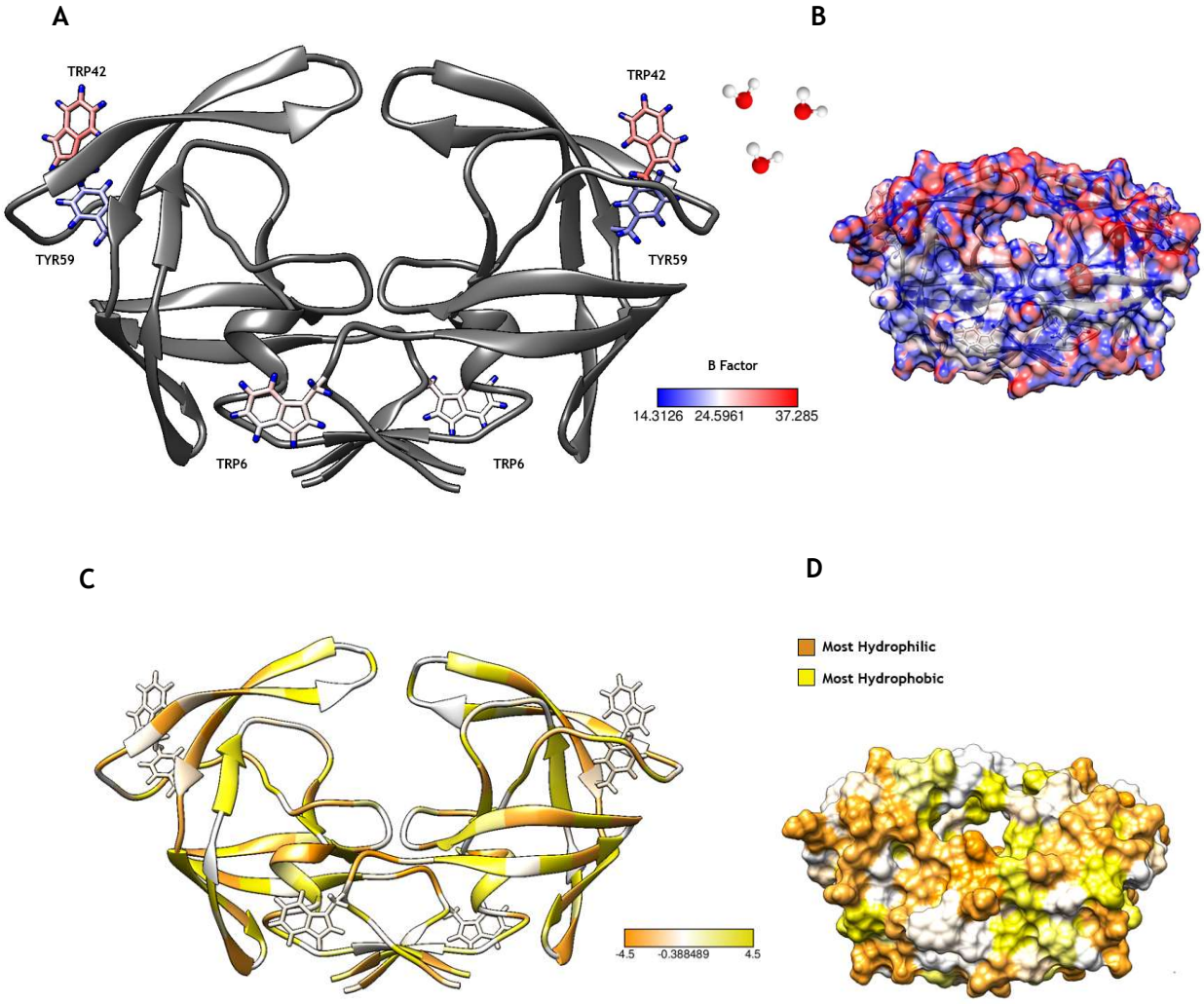


Figure 22: Crystal Structure of the wild-type HIV-1 PR representing B-factors of probes (A), B-factors across the PR structure (B), and hydrophobicity of probes (C) and hydrophobicity across the PR structure (D). All four HIV-1 dimeric PRs possess four tryptophan residues (located to the hinge region and dimer interface) and two tyrosine residues (located in the hinge region). The degree of hydrophobicity of the four tryptophan residues and two tyrosine residues is consistent while a noticeable difference can be attributed to the B-factors associated with each residue. The B-factor can be indicative of relative flexibility of a region of a protein. Trp6 on each chain is located to the dimer interface, a relatively static area with low B-factor in comparison to the hinge region. The Tyr59 is almost buried within the hinge, reducing flexibility but the Trp42 residues can be denoted with the highest B factors of all six probes and thus significantly more flexibility.

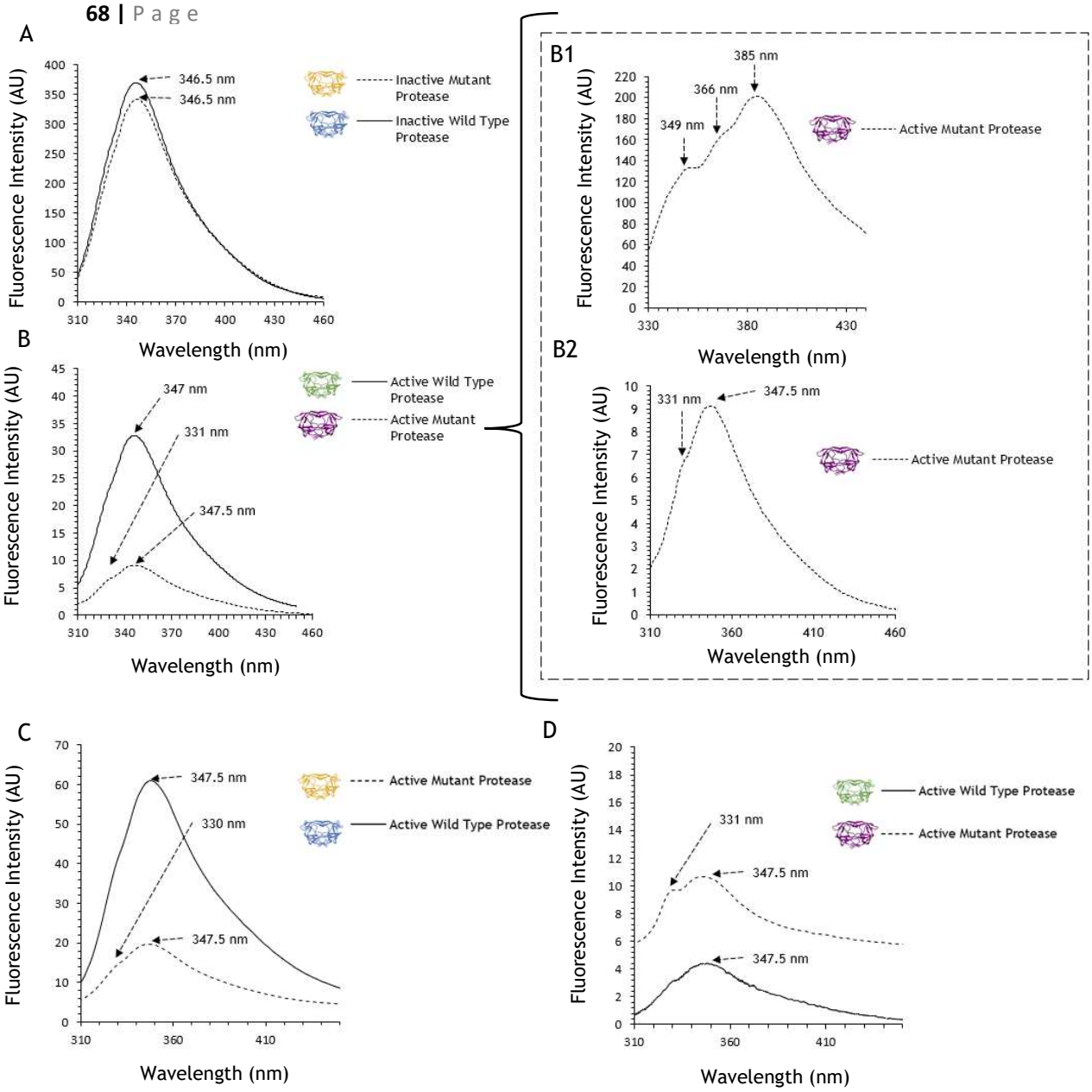


Figure 23: Intrinsic fluorescence spectra of inactive and active wild-type and mutant HIV-1 PRs. Upon excitation of the inactive PRs at a wavelength of 280 nm, both produced emission spectra with a maximum fluorescence intensity at 346.5 nm, characteristic of most HIV-1 PRs (A) (Szeltner & Polgár, 1996). Similarly, both the active wild-type and mutant PR produced a predominating fluorescence peak at a wavelength of 347 nm, indicating native global tertiary fold (B). Further investigations under excitation at 295 nm (probing only tryptophan residues considering the crucial location and characteristics of Trp42), revealed both the inactive mutant and active mutant PR intrinsic fluorescence spectra produced a shoulder at 330-331 nm (C & D). Overall, the tertiary structural analysis revealed, the mutations (substitutions and insertion) were found to produce no significant impact on the global tertiary structure of the PR when probed at 280 nm. The D25A mutations do not impact global tertiary structure of the PR. Targeted probing of tryptophan species at 295 nm, suggested the insertion mutations could be influencing hinge tertiary structure, fold, and potentially dynamics.

Further investigations at an excitation at 295 nm (probing only tryptophan residues considering the crucial location of Trp42), revealed both the inactive mutant and active mutant intrinsic fluorescence spectra produced a shoulder at 330-331 nm, with greater apparency in the active mutant PR spectra (Figure 23C and Figure 23D). The more targeted evaluation of the global tertiary structure suggests the insertion mutations associated with the mutant PRs (in both active and inactive constructs) could be influencing the local environment of the tryptophan residues residing in the hinge region when compared to the wild-type PRs. Furthermore, computational studies have suggested that hinge insertions can reduce PR flap dynamics (Zondagh *et al.,* 2018). Furthermore, the D25A mutation could be acting to reduce misfolding and aggregation propensity of the mutant PR during *in vitro* chemical renaturation. Coupled with minimal influence on the global tertiary structure, as previously deduced, the D25A mutation could be employed as a tool to reduce the empirical obstacles associated with PRs. Intrinsic fluorescence is a highly sensitive technique that provides global insights into whether a profile of mutations is contributing to differences in the structure, fold, and stability of a protein at a global tertiary level when compared to a well-established control i.e., wild-type protein. However, such insights are limited to a higher-level understanding of these structural influences from a static perspective. To probe the mechanisms associated with these changes at local structural and functional regions of the protein, real-time analytical techniques such as stop-flow-based chemical folding studies and molecular dynamics modelling would be investigated in the future.

Overall, the tertiary structural analysis revealed, the mutations (substitutions and insertion) were found to produce no significant impact on the global tertiary structure of the PR when probed at 280 nm. Furthermore, the D25A mutation had no significant impact on the global tertiary structure of the PRs. The active mutant PR was denoted to be less stable in solution and to fold less efficiently under this study's chemical renaturation procedure than its wild-type and inactive counterpart. Targeted probing of tryptophan species at 295 nm, suggested the insertion mutations could be influencing hinge tertiary structure, fold, and thus potentially dynamics (Zondagh *et al.,* 2018).

3.3.4 Global quaternary structure evaluation of inactive & active wild-type & mutant PRs using SE-HPLC

SE-HPLC was used to evaluate the predominating quaternary structure adopted by the active and inactive mutant PRs. Both the inactive and active wild-type and mutant proteases eluted between 18 and 20 minutes between the appropriate standard proteins (Figure 24 A & Figure 24 C). The elution profile in isolation depicted a single peak for both inactive PRs (Figure 24B & Figure 24D). Upon construction of a standard curve, the linear regression equation was used to analyse the inactive wild-type PR and inactive mutant PR molecular weights to ~26 kDa and ~22 kDa, respectively (Figure 24E).

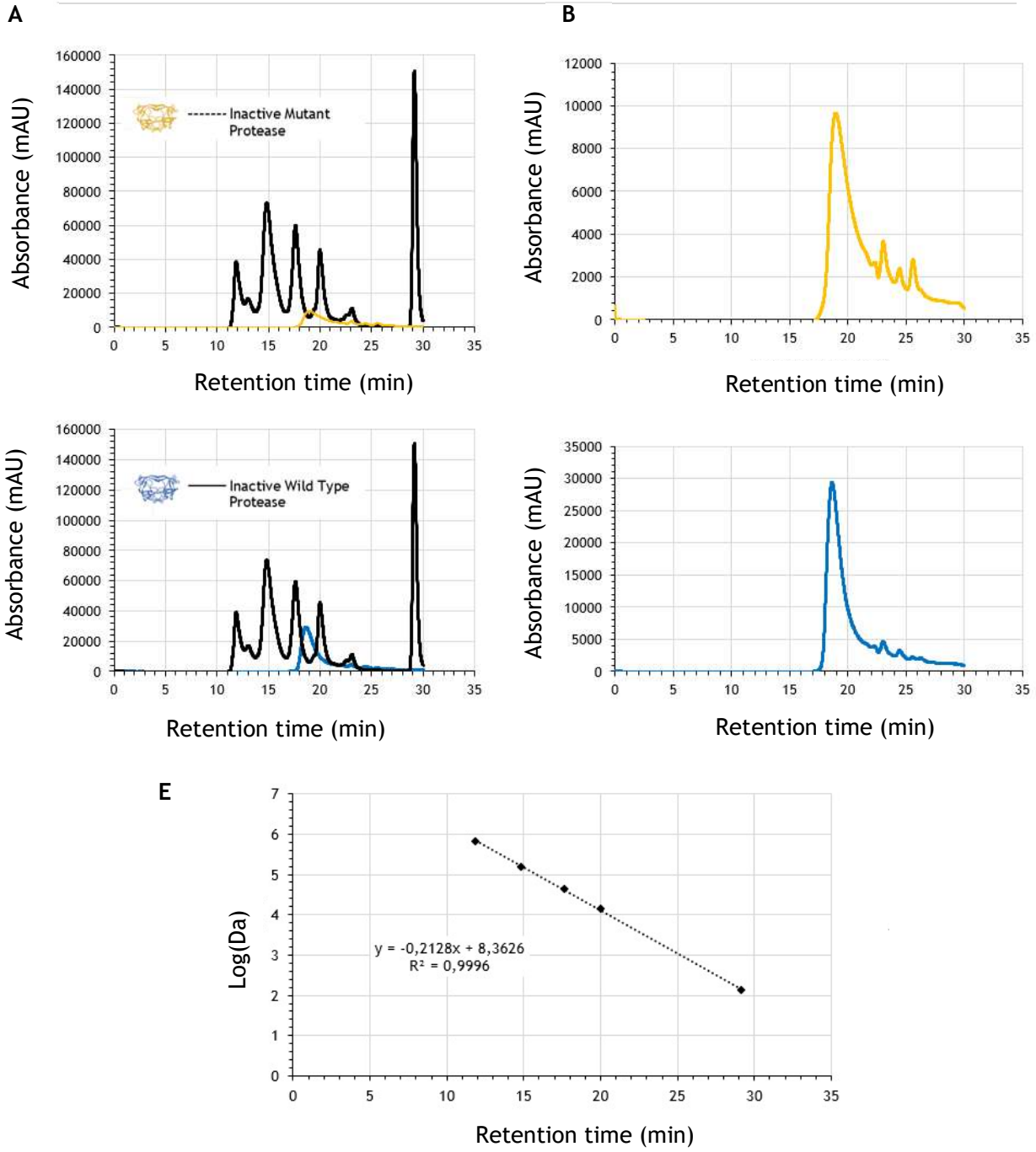


Figure 24: SE-HPLC profiles of the inactive mutant PR with protein standards (A) and in isolation (B) in comparison to the SE-HPLC of the inactive wild-type PR with protein standards (C) and in isolation (D). A standard curve was generated from the SE-HPLC of protein standards (E). Both PRs were found to elute between the appropriate standard proteins when compared to the elution profile of standard proteins (A & C). The isolated elution profile of each inactive PR depicted a single peak (B & D). The inactive wild-type PR and inactive mutant PR were found to correspond to a molecular weight of ~26 kDa and ~22 kDa, characteristic of the HIV-1 dimer when corresponding retention times were interpolated along a standard curve, with a reliable correlation between data points (R^2~0.99) (E).

The inactive mutant PR corresponded directly to the native size of a dimeric PR, 22 kDa while the inactive wild-type PR corresponded to a slightly higher kDa. However, the difference in size, 4 kDa, associated with the inactive wild-type does correspond to the average molecular weight of the extra 20 amino acids on the N and C terminal ends of the PR as discussed in section 2.3.1. Both inactive PRs were considered to preferentially adopt the native dimeric state of the HIV-1 PR (Hayashi *et al.,* 2014).

Similarly, the active wild-type and mutant PRs eluted between the appropriate protein standards (Figure 25A & Figure 25C). Although SE-HPLC is more sensitive than quaternary analysis by SDS-PAGE, fluctuations in column and machine pressure can generate variations in the rate of elution. This necessitates the use of standards for each SE-HPLC run to be able to size the peaks and decipher molecular weights with greater consistency. The SEC profiles associated with the active PRs varied from the inactive PRs by the presence of a shoulder and subsequent peak as opposed to a single peak (Figure 25B & Figure 25D). Upon construction and utilisation of a standard curve, the shoulder was found to correspond to a molecular weight of ~22 kDa and the peak corresponded to a molecular weight of ~10 kDa (characteristic of a PR monomer) (Figure 25E). Thus, under high salt concentrations, 300 mM, the dimer: monomer ratio in the active PRs is shifted in the favour of the monomer, whereas the inactive PRs remain in the dimeric state. This could be due to the hydrophobic nature of the alanine in the inactive PRs which will be preferentially buried in polar environments, favouring the dimeric state. Under extreme salt conditions, the negative aspartate residues in the monomers of the active PRs may preferentially interact with the ions under high salt concentrations as opposed to being brought into proximity in the dimeric PR state, increasing less favourable electrostatic repulsive forces. These behavioural differences are important to evaluating physiological differences between the use of inactive PRs and active PRs in certain studies as well as the design of downstream investigations. For example, ensuring the appropriate concentration of salt is used in enzyme assays, enough to promote activity (100 mM) but not disrupt the dimeric state (less than 300 mM) (Z. Szeltner & Polgár, 1996).

Overall, all four PRs adopt a quaternary structure considered to be that of a native 22 kDa HIV-1 dimeric PR. Thus, the substitution and insertion mutations associated with the mutant PRs do not influence the global quaternary structure of the PR when compared to the wild-type PRs as all adopt a size of 22 kDa. With those considerations, the substitution and insertion mutations associated with the various PRs do not seem to affect any of the PR's ability to adopt the native 22 kDa quaternary state necessary for enzyme function.

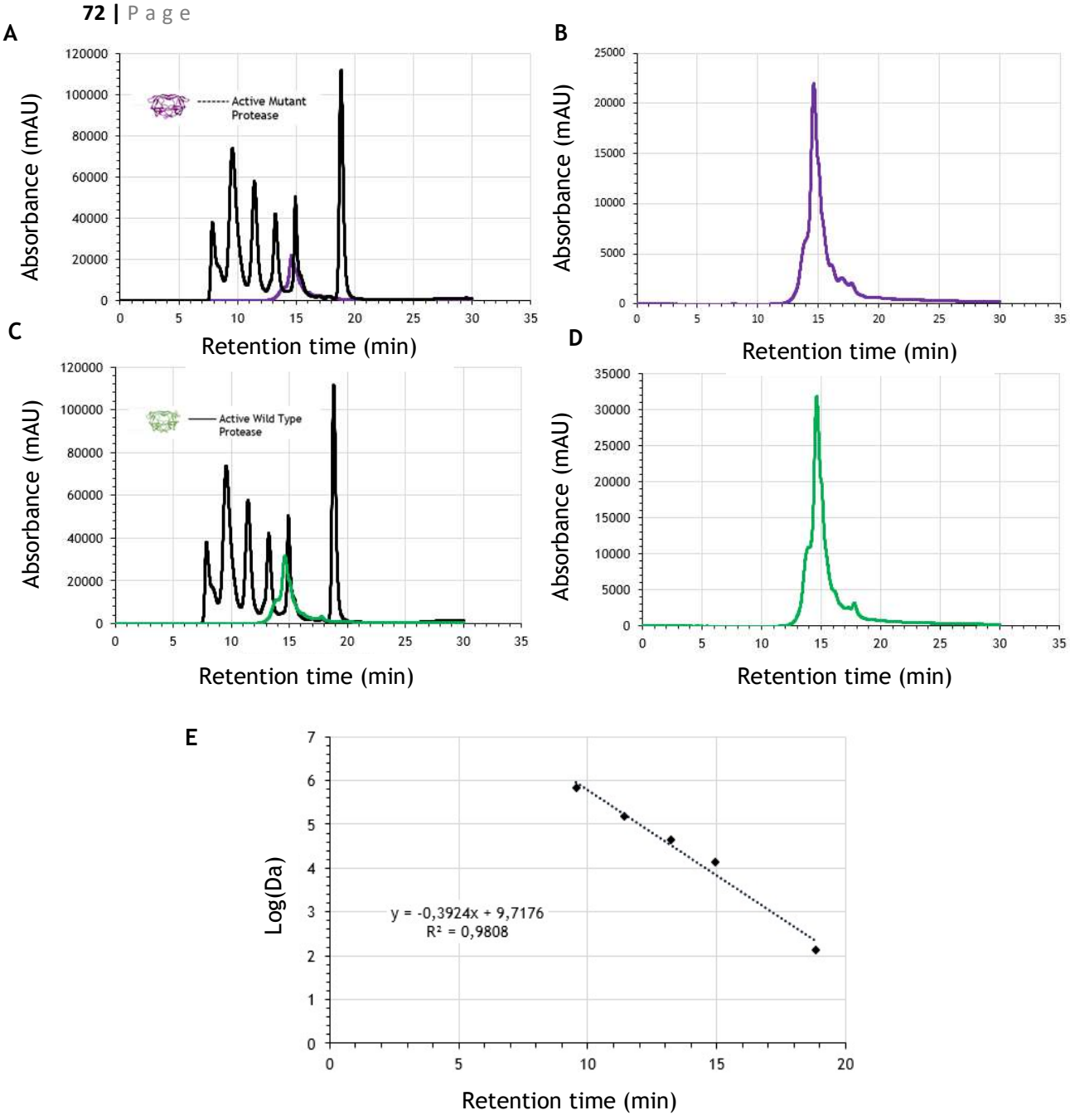


Figure 25: SE-HPLC profiles of the active mutant PR with protein standards (A) and in isolation (B) in comparison to the SE-HPLC of the active wild-type PR with protein standards (C) and in isolation (D). A standard curve was generated from the SE-HPLC of protein standards (E). Both PRs were found to elute between the appropriate standard proteins when compared to the elution profile of standard proteins (A & C). The isolated elution profile of each active PR depicted a shoulder followed by a single peak (B & D). Upon construction of a standard curve possessing reliable correlation between data points (R^2~0.98) (E), the shoulder and peak were found to correspond to molecular weights of ~22kDa (the dimeric size) and ~10 kDa (the monomeric size), respectively. The dimer: monomer ratio shift was attributed to the influence of extreme salt concentrations on the active PRs.

3.4 Functional assessment of inactive and active constructs of the Subtype C wild-type and the L38↑N↑L^{+4} hinge mutant PRs

3.4.1 Active site titrations of purified HIV-1 PRs against acetyl-pepstatin using ITC

Active site titrations are necessary to establish the population of PR that is native and functionally folded, which cannot be determined by spectrophotometric methods. Titration of AP against the inactive PRs revealed no noticeable binding (Supplementary Figure 1). All raw heat data generated within the reactions were characteristic of heats of dilution indicating no binding interaction between AP and both the inactive wild-type and inactive mutant PRs. This suggested that the D25A mutation could be compromising the stable interaction of acetyl-pepstatin, characterised as a weak inhibitor, within the active site. Either a D25-activated water molecule participates as the negatively charged nucleophile and interacts within the hydrogen bond, alternatively, a nucleophilic atom of an active site amino acid side-chain satisfies the latter function (Shen *et al.*, 2012). Inhibitors like AP have been designed to mimic the substrate/gag transition state by possessing a hydroxyl group that forms hydrogen bonds with the carboxyl group of PR active site residues, Asp25 and Asp25′ (Lv *et al.,* 2015). Thus, the replacement of the Asp residue for a hydrophobic alanine residue could influence the structural nature and bridging of water molecules necessary to forming these polar contacts. As an n-value could not be derived from AP active site titrations with the inactive PRs, the concentration determined by absorbance spectrophotometry could not be corrected and *in vitro* refolding success could not be evaluated. This limited further studies of inactive PRs concerning inhibitors and other substrates.

On the contrary, the binding of AP to both the active wild-type and mutant PRs could be detected (Figure 26). The profile of the curve agreed with previous studies that suggest the reaction is endothermic, the interaction is weak (low K_d values (10^{-8}) compared to strong inhibitors such as SQV ~ 133 nM), associated with few polar contacts (low ΔH value) and entropically unfavoured (ΔS) (Sherry *et al.,* 2020). However, these values can only be used to make broad inferences as the AP has not been titrated against a corrected concentration of PR (Sherry *et al.,* 2020). The wild-type-AP active site titration revealed that the population of PR recovered by 34%, as indicated by the n-value of 0.342. The mutant-AP titration revealed that the population of PR was only 10% recovered, as indicated by the n-value of 0.1. This relates to findings in the global structural evaluation which suggests the profile of mutations associated with the mutant PR impacts the fold of the mutant PR increasing its propensity to misfold. Neither PRs recovered to desirable concentrations (wherein one study suggested the wild-type could be recovered to as close as 90%) suggesting future changes to the chemical refolding would be needed to optimise functional yield (Sherry *et al.* 2020*)*. Subsequently, the concentrations of the active wild-type and active mutant PR were corrected to 7.5 µM and 1.3 µM, respectively. Thus, the protein stocks could be concentrated to more accurate working concentrations for downstream experiments that are concentration dependent such as enzyme assays and further ligand-PR studies.

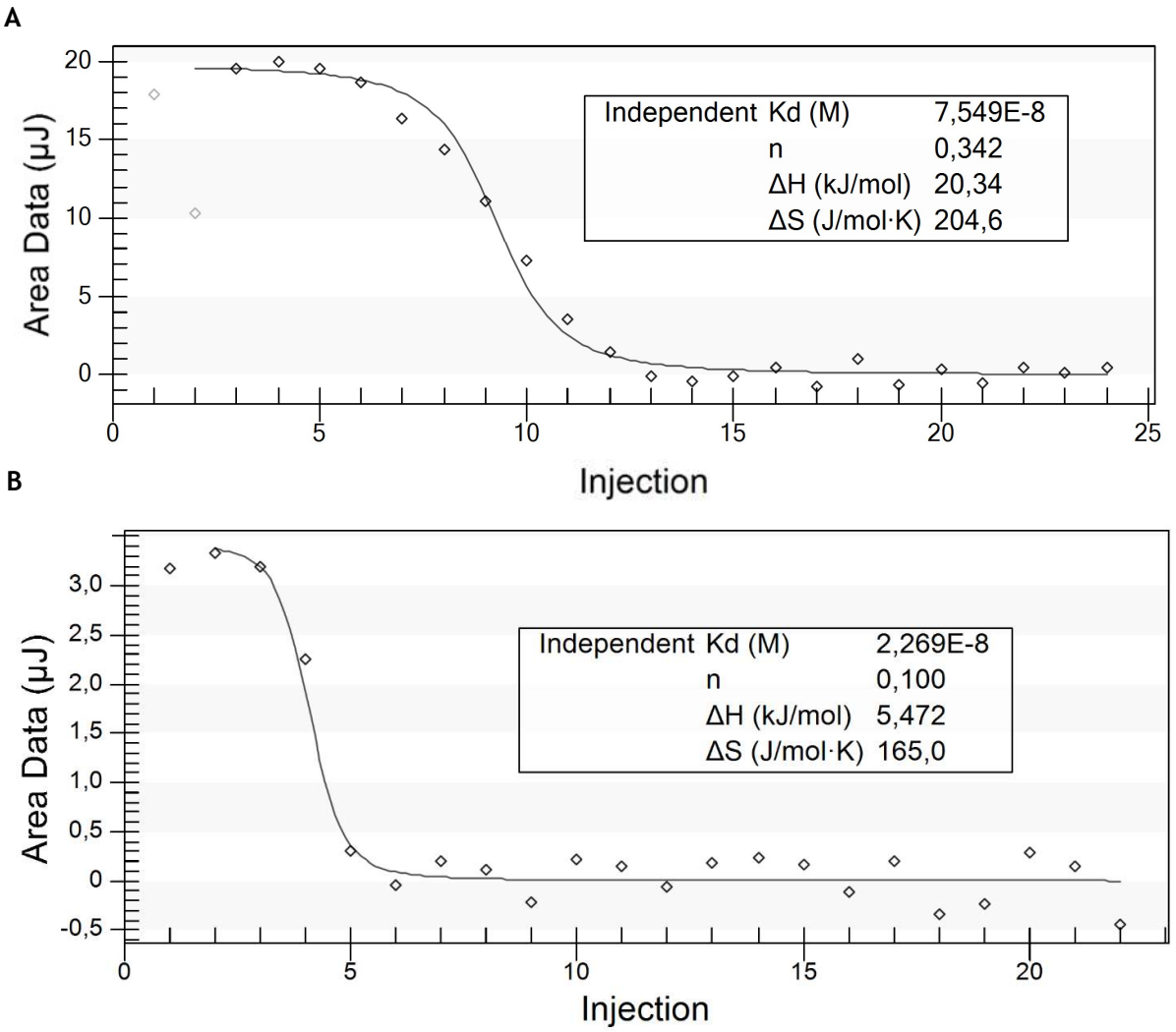


Figure 26: The fit of an independent binding model to the raw thermal data from active site titrations of AP against the active wild-type PR (A) and active mutant PR (B). The wild-type-AP active site titration revealed that the population of PR recovered by 34%, as indicated by the n-value of 0.342. The mutant-AP titration revealed that the population of PR was only 10% recovered, as indicated by the n-value of 0.1. This cooperates with the findings in the global structural evaluation which suggests the profile of mutations associated with the mutant PR impacts the fold mutant PR increasing its propensity to misfold. Neither PRs recovered to desirable concentrations (wherein in 1 study suggested the wild-type could be recovered to as close as 90%) suggesting further changes to the chemical refolding would be needed in the future to optimise functional yield (Sherry *et al.,* 2020).

3.4.2 Enzyme activity evaluation using a FRET-based activity assay

Enzyme activity of the inactive and active wild-type and mutant PRs was measured by monitoring the hydrolysis of the fluorogenic substrate that mimics a capsid/p2 cleavage site of the Gag polypeptide. Cleavage of the peptide minimises the established quenching effect and results in a fluorescent signal from Abz emission which is monitored through excitation at a wavelength of 337 nm and an emission wavelength of 425 nm. The D25A mutation was proven to knockout PR activity by revealing no change in fluorescence (indicative of no substrate hydrolysis) at a fixed concentration when compared to the wild-type active (Figure 27E). Following, the catalytic capabilities of the active mutant PR were compared to the wild-type PR, in relation to conclusions made about the structural influences of the profile of mutations (Figure 27A & Figure 27B).

The specific activity of the active wild-type PR was found to be 31.6 $\mu mol.min^{-1}.mg^{-1}$ with minimal error (±0.0004) and a strong correlation (R^2 of 0.98) between data points (Figure 27A). The specific activity of the active mutant PR was found to be 7.4 $\mu mol.min^{-1}.mg^{-1}$ with minimal error (±0.0003) and a strong correlation (R^2 of 0.99) between data points (Figure 27C). Specific activity is equal to the rate of reaction multiplied by the volume of the reaction divided by the mass of total protein. Thus, specific activity refers to the amount of active enzyme present in a sample that possesses the ability to catalyse the transformation of the substrate into a product. The k_{cat} of the wild-type PR could then be reliably determined through manipulation of the specific activity data. The wild-type k_{cat} was found to be 11.60 sec^{-1} with a minimal error of $6.75x10^{-6}$ and a strong correlation of 0.98 between data points (Figure 27B). The k_{cat} of the mutant PR was found to be 2.71 sec^{-1} with a minimal error of $3.97x10^{-6}$ and a strong correlation of 0.99 between data points (Figure 27D). The k_{cat} indicates the number of substrate molecules that are processed per unit of time by a single enzyme. These values correlate with published data (specific activity~ 12.1 ± 1.1 $\mu mol.min^{-1}.mg^{-1}$ & 4.5 ± 0.4 sec^{-1}) that indicated the specific activity and k_{cat} of the L38↑N↑L^{+4} mutant PR is lower than that of the wild-type PR (WT PR specific activity~ 21.0 ± 1.1 $\mu moles.min^{-1}.mg^{-1}$ & k_{cat}~ 7.7 ± 0.4 sec^{-1}) (Williams *et al.,* 2019). Molecular dynamic studies have revealed the double insertion mutation, L38↑N↑L, reduces flap dynamics in the mutant PR (Williams *et al.,* 2020). Thus, the profile of mutations may not only act as a mechanism for drug resistance but potentially impacts the catalytic efficacy of the mutant PR as the flaps play a crucial role in controlling substrate entry and binding.

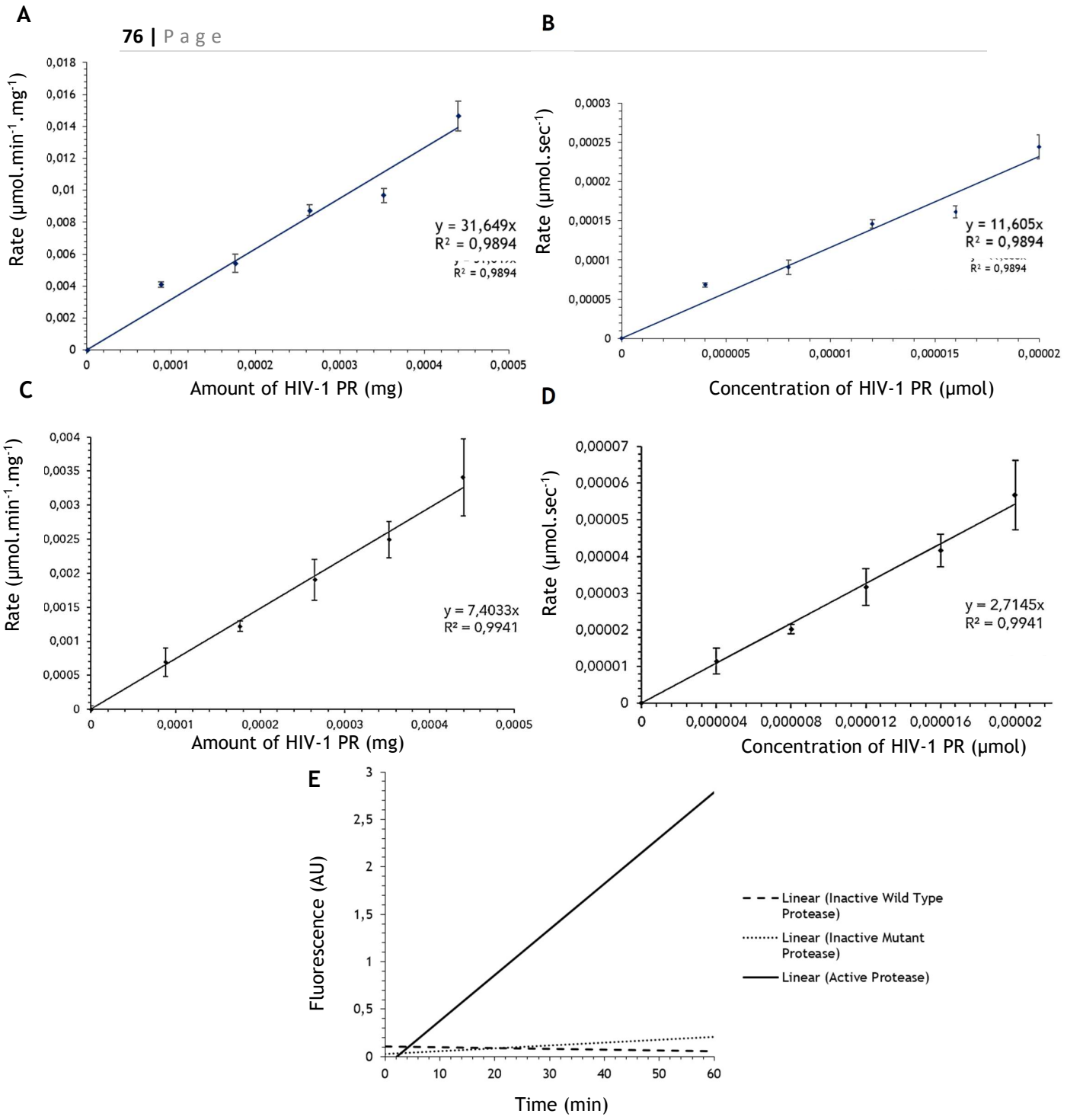


Figure 27: Determination of the specific activity of the active wild-type (A) and active mutant (C) PR as well as the k_{cat} of active wild-type (B) and mutant (D) PRs under steady-state kinetics. (E) Catalytic activity assessment of inactive PRs. The specific activity of the active mutant PR was found to be 7.4 µmol.min^{-1}.mg^{-1} with minimal error (±0.0003) and a strong correlation (R^2 of 0.99) between data points. The specific activity of the active wild-type PR was found to be 31.6 µmol.min^{-1}.mg^{-1} with minimal error (±0.0004) and a strong correlation (R^2 of 0.98) between data points. The k_{cat} could then be reliably determined through manipulation of the specific activity data and found to be 2.71 sec^{-1} with a minimal error of 3.97x10^{-6} and a strong correlation of 0.99 between data points. The wild-type PR k_{cat} was found to be 11.60 sec^{-1} with a minimal error of 6.75x10^{-6} and a strong correlation of 0.98 between data points. (E) No activity associated with inactive PRs.

3.5 Stability evaluation of active wild-type and mutant HIV-1 PRs in the absence and presence of known inhibitors using a thermal shift assay

Following active site titrations, active PRs were subjected to thermal stability studies to assess whether mutant-related polymorphisms affected apo PR stability and the stability of known inhibitors within the PR active site (Andreotti *et al.,* 2015). Inactive PRs were not included in this study due to the lack of AP binding noted in section 3.4.1. Thermal melting of the wild-type and mutant PRs alone revealed reduced thermal stability associated with the mutant PR (possessing a melting temperature of 64 ℃) when compared to the wild-type PR (possessing a higher melting temperature of 67 ℃) (Figure 28 & Figure 29). Previous differential scanning calorimetry (DSC) studies revealed melting temperatures in similar ranges including a wild-type PR melting temperature of 64 ℃ and a related mutant melting temperature of 69 ℃ (Sheik-Ismail *et al.,* 2023 *manuscript under review*). The difference in melting temperatures can be attributed to the reduced sensitivity associated with the thermal shift technique compared to DSC (Lo *et al.,* 2004). An overall 4 ℃ difference in melting temperature between wild-type and mutant PR in this study can be interpreted as the extra thermal energy required to denature the wild-type PR compared to the mutant PR (Andreotti *et al.,* 2015). This suggests that mutant-related polymorphisms affect the structural stability, thus thermal stability, causing the PR to denature at lower melting temperatures. This reduced structural stability can be associated with the mutant PR propensity to misfold or adopt structural variations in the hinge region due to the insertion mutation as seen in the global tertiary structural evaluation (Williams *et al.,* 2019). Furthermore, these results agree with the noticeable reduction in the mutant PR catalytic ability. Thus, to probe the clinical relevance of structural alterations associated with the mutant PR, thermal stability studies were conducted in the presence of known inhibitors AP (as a comparative control), DRV, and SQV.

In thermal shift assays with AP, the initial thermal shift was seen at 5 µM of AP for both the wild-type and mutant PRs. AP was seen to stabilise both the wild-type and mutant complex, marginally, generating a 1 ℃ increase in melting temperature at 5 µM and an overall 6 ℃ increase in melting temperature at maximum AP concentration (75 µM). Although AP-mutant PR melting temperatures were lower overall across AP concentrations compared to that of the wild-type complex, the overall difference of 6 ℃ was the same at the maximum concentration of AP (75 µM). This suggests the AP-wild-type complex is not significantly more stable than the AP-mutant complex under saturating AP concentrations. These findings are all characteristic of a weak inhibitor, generating a consistent but marginal shift in melting temperature (℃) with increasing inhibitor concentrations (Marciniszyn *et al.,* 1976).

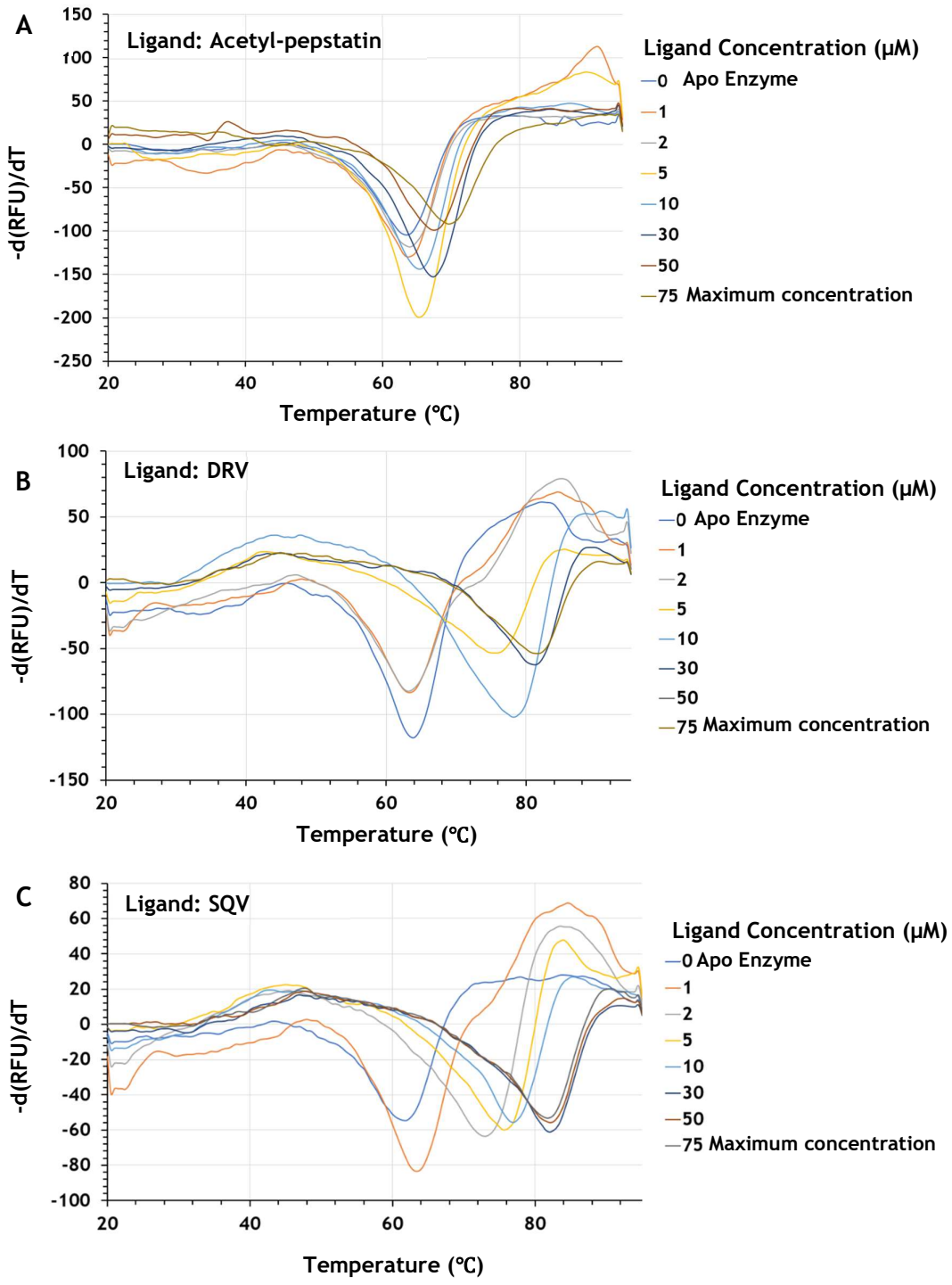


Figure 28: Thermal shift assays of the active mutant PR at varying concentrations of AP (A), DRV (B), and SQV (C). The apo mutant PR melting temperature was found to be 67 °C. A concentration of 5 µM of AP was required for the first melting temperature shift, followed by 1 µM for DRV and 1 µM for SQV. At maximum inhibitor concentration (75 µM) the melting temperature was found to be 73.5 °C, 90 °C, and 88 °C in the presence of AP, DRV, and SQV, respectively.

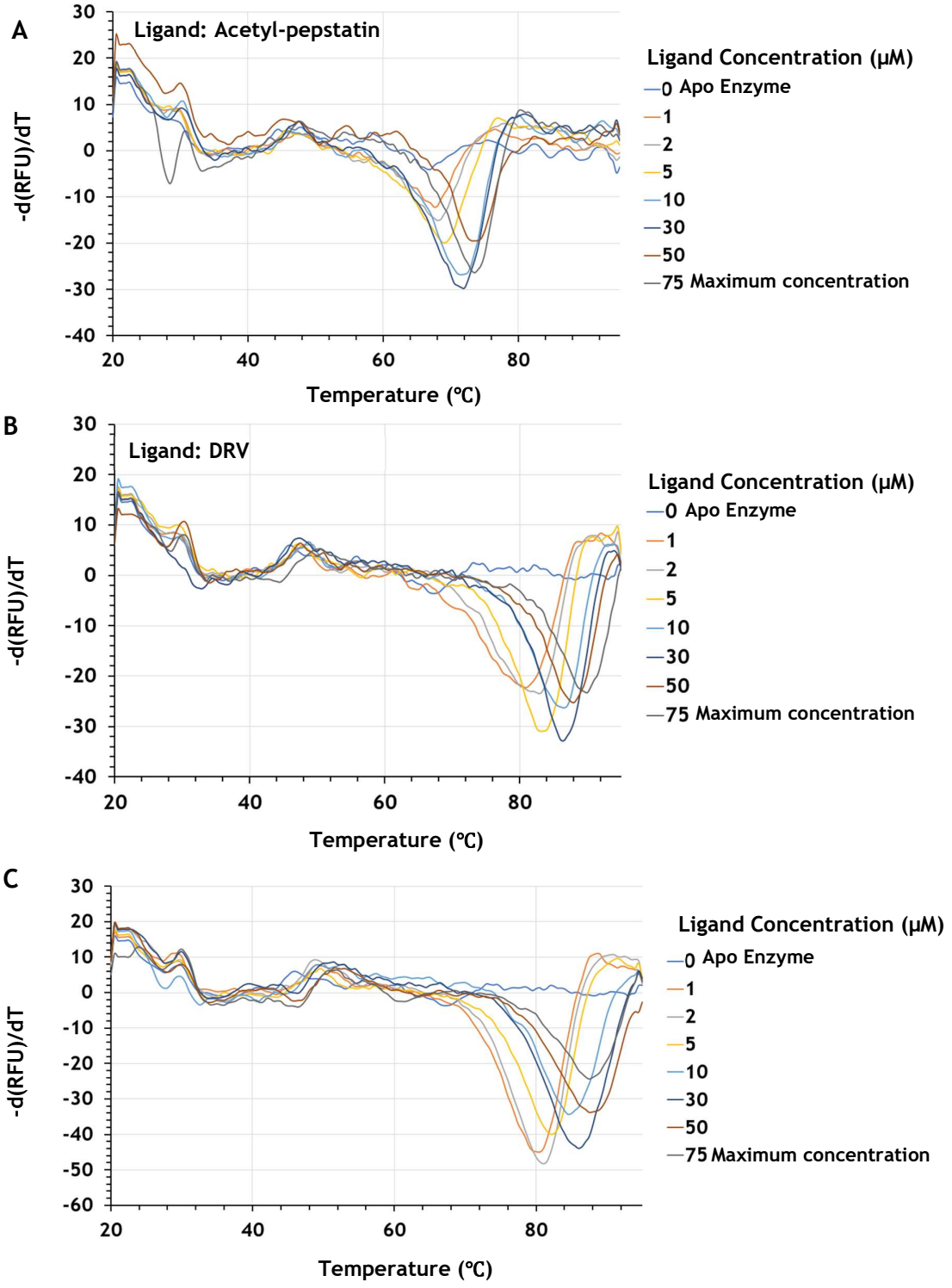


Figure 29: Thermal shift assays of the active wild-type PR at varying concentrations of AP (A), DRV (B), and SQV (C). The apo wild-type PR melting temperature was found to be 64 °C. A concentration of 5 µM of AP was required for the first melting temperature shift, followed by 5 µM for DRV and 2 µM for SQV. At the maximum concentration of inhibitor (75 µM) the melting temperature was found to be 70 °C, 82 °C, and 82 °C in the presence of AP, DRV, and SQV, respectively.

In contrast to AP, DRV produced stronger PR stabilising effects as seen by the significant increase in overall melting temperatures in both PR thermal shift assays. DRV is a stronger inhibitor than AP, producing a melting temperature difference of 23 ℃ between the apo wild-type PR and DRV complex (at 75 µM DRV) as well as a difference of 18 ℃ between the apo mutant PR-DRV complex (at 75 µM DRV). The initial shift was seen at a concentration as low as 1 µM DRV in the wild-type assay, while the initial shift was seen at a higher concentration of 5 µM DRV in the mutant assay. At these respective concentrations, the melting temperature of the wild-type complex and the mutant complex was found to be 81 ℃ and 76 ℃, respectively. Thus, when compared to the wild-type scenario, DRV binds with lower affinity at low concentrations to the mutant PR, and the stability of DRV within the mutant active site is reduced (as seen by the complex denaturing 5 ℃ earlier). At a maximum concentration of 75 µM DRV, the wild-type-DRV complex denatured at 90 ℃ while the mutant-DRV complex denatured earlier at 82 ℃. This further emphasises the reduced stability of DRV binding in the mutant PR active site. Previous inhibition studies have shown that the PIs, lopinavir, atazanavir, and darunavir, do bind and inhibit the mutant PR (Williams *et al.,* 2019). However, molecular dynamics simulations depicted reduced flap dynamics, important to regulating entry into the active site, in the mutant PR compared to the wild-type PR. Furthermore, induced-fit docking studies showed that the DRV binding can be characterised by reduced hydrophobic contacts and hydrogen bonds. These thermal shift assays support the notion that the mutant polymorphisms impact the stability of DRV within the active site when compared to the wild-type PR.

Similarly, SQV produced stronger PR stabilising effects as seen by the drastic increase in melting temperatures across concentrations in both PR thermal shift assays. SQV presence produced melting temperature differences of 21 ℃ between the apo wild-type PR and SQV complex (at 75 µM SQV) as well as a difference of 18 ℃ between the apo mutant PR and SQV complex (at 75 µM SQV). The initial shift was seen at a concentration as low as 1 µM SQV for the wild-type PR, while at a higher concentration of 2 µM of SQV for the mutant PR. At these respective concentrations, the melting temperature of the wild-type complex was found to be 80 ℃ and 73 ℃ for the mutant complex. This suggests SQV binds with less affinity at low concentrations to the mutant PR than the wild-type PR and that the stability of SQV within the mutant active site is reduced (as seen by the complex denaturing 7 ℃ earlier). Melting temperatures were like those of the DRV assay suggesting SQV does not bind with any greater stability than DRV in both wild-type and mutant PR active sites. At a maximum concentration of 75 µM, the wild-type-SQV complex denatured at 88 ℃ while the mutant-DRV complex denatured earlier at 82 ℃, further emphasising the reduced stability of SQV binding in the mutant PR. Thermal shift assays support the notion that mutant polymorphisms impact the stability of SQV within the active site when compared to the wild-type scenario. Thus, thermal shift assays support the notion that the inclusion of hinge insertions in the PR could be a potential drug resistance mechanism in the context of not only DRV but SQV as well.

3.6 Generating HIV-1 Gag p2/NC-mimicking peptides for functional studies with active wild-type and active mutant PRs

The HIV-1 PR can process a variety of gag cleavage sites, possessing no sequence homology, at varying rates (Kuhn *et al.,* 2014). Thus, subsequent studies trialled the generation of peptides corresponding to alternative gag cleavage sites (i.e., p2/NC) that could be used in enzymatic studies to generate a more comprehensive picture of mutant PR catalytic capability. Furthermore, previous studies had suggested the mutant PR displays a level of DRV resistance when exposed to the complementary mutant gag (Williams *et al.,* 2019). Thus, a peptide mimicking the mutant p2/NC cleavage site, containing three mutations, was designed for synthesis and binding studies. The peptides mimicking the p2/NC gag cleavage site of the wild-type and mutant HIV-1 gag were empirically optimised using computational prediction tools that assessed qualities based on the peptide primary sequence (Smialowski *et al.,* 2007) (Figure 30). These properties were compared to the peptide used in FRET-based enzyme assays to improve the probability of peptide solubility and stability in the established PR-compatible *in vitro* environments. The computational tools predicted both the peptide mimicking the mutant p2/NC site and the wild-type p2/NC site with similar characteristic to the FRET-based assay peptide (Figure 30). Some of the crucial modifications to the peptide included the addition of a YKK tail. The Tyr was intended as a probe to monitor peptide concentration in solution while the two Lys were necessary to increase the peptide solubility.

Following solid-phase synthesis, the mutant peptide could be readily solubilised in H_2O and the PR storage buffer. However, the wild-type peptide remained highly insoluble in a range of acidic, basic, organic, and inorganic solvents with the highest solubility achieved in 100% DMSO, conditions not compatible with experimental studies involving the PR. Due to the insoluble nature of the wild-type peptide, purification and subsequent studies with this peptide could not be continued. The mutant peptide was purified through SE-HPLC into three separate fractions. The three fractions were subjected to LC-MS to evaluate whether the peptide fractions consisted of complete 13 amino acid peptides with few contaminating peptide species. LC-MS of the purified mutant peptide revealed sized and charged fragments that equalled multiples of the larger 13 amino acid peptide molecular weight. Furthermore, the LC-MS showed peaks corresponding to the total molecular weight of the mutant peptide, 1466 Da (Figure 31). LC-MS analysis was repeated for two other fractions, similarly indicative of a complete mutant peptide (as seen by the sizes of the fragments) and ideal purity (as seen by no predominating contaminating fragments) (Supplementary Figure 2 & Figure 3). Subsequently, far-UV CD was used to assess any predominating secondary structure associated with the mutant peptide, revealing the predominating random coil structure. This is like the gag protein which is largely disordered in structure until bound to the PR (Greenfield, 2006; Prabu-Jeyabalan *et al.,* 2002) (Figure 32). This was necessary to ensure the peptide models used in docking studies possessed the appropriate secondary structure. Such insights were also necessary to the overall understanding of the results in ITC binding studies.

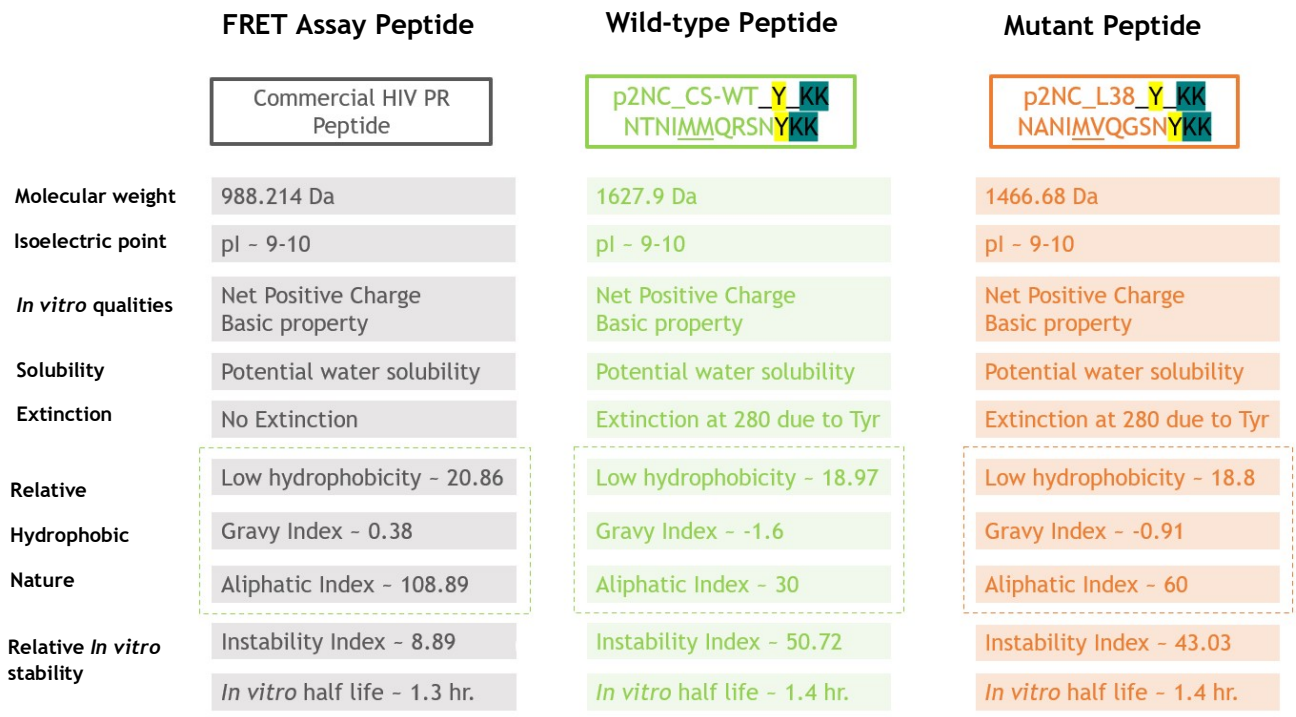


Figure 30: Predicted empirical properties of wild-type and mutant p2/NC peptides in comparison to the FRET-based assay peptide. The computational tools predicted both the peptide mimicking the mutant p2/NC site and the wild-type p2/NC site as having similar characteristic to each other and the FRET-based assay peptide. Some of the crucial modifications to the peptide included the addition of a YKK tail. The Tyr was needed as a probe to monitor peptide concentration in solution while the two Lys were necessary to increase the peptide solubility.

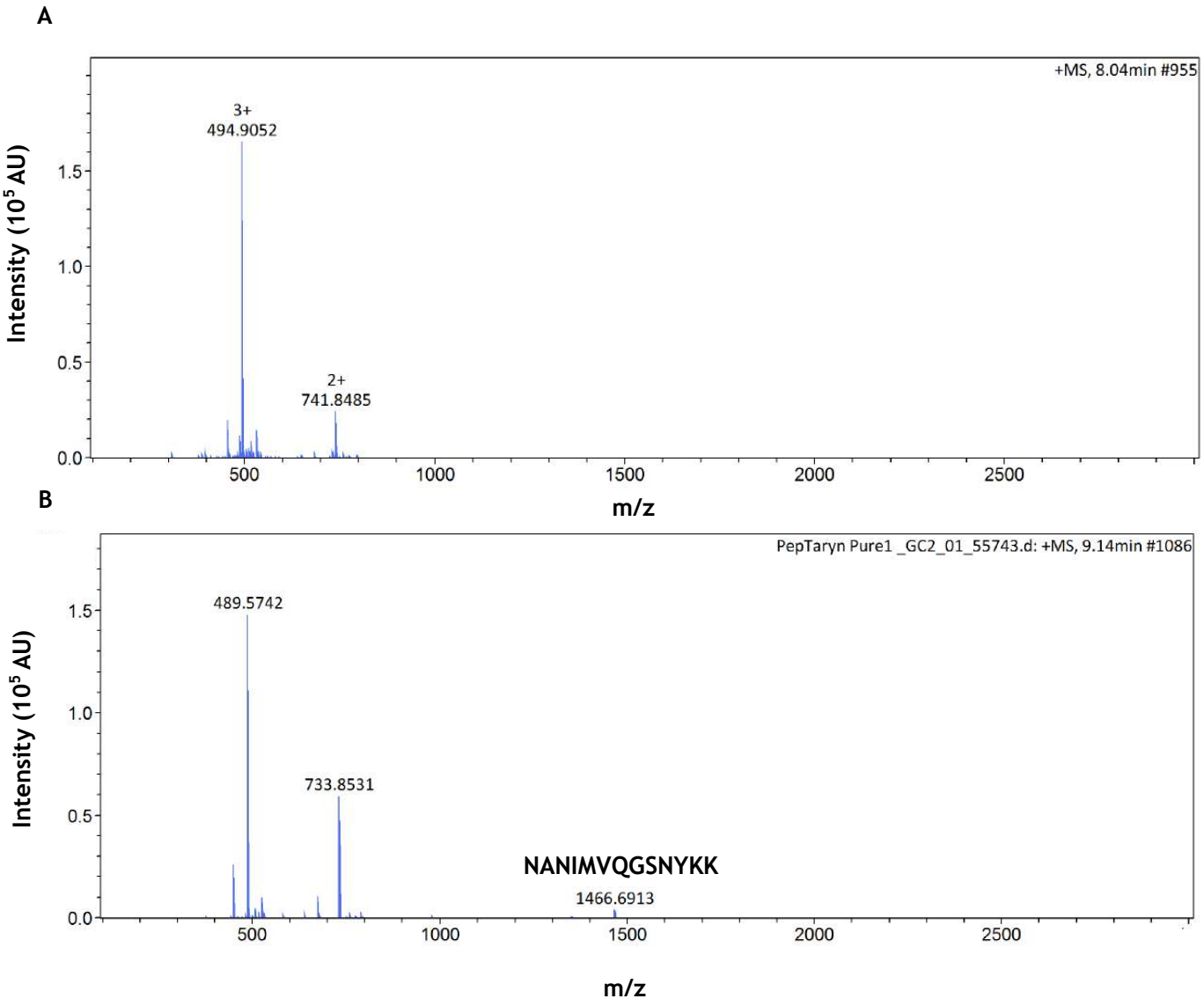


Figure 31: LC-MS profiles of purified mutant p2/NC peptide. LC-MS of the purified mutant peptide revealed sized and charge fragments that equalled multiples of the larger 13 amino acid peptide (A & B). Furthermore, the LC-MS showed peaks corresponding to the total molecular weight of the mutant peptide, 1466 Da (B).

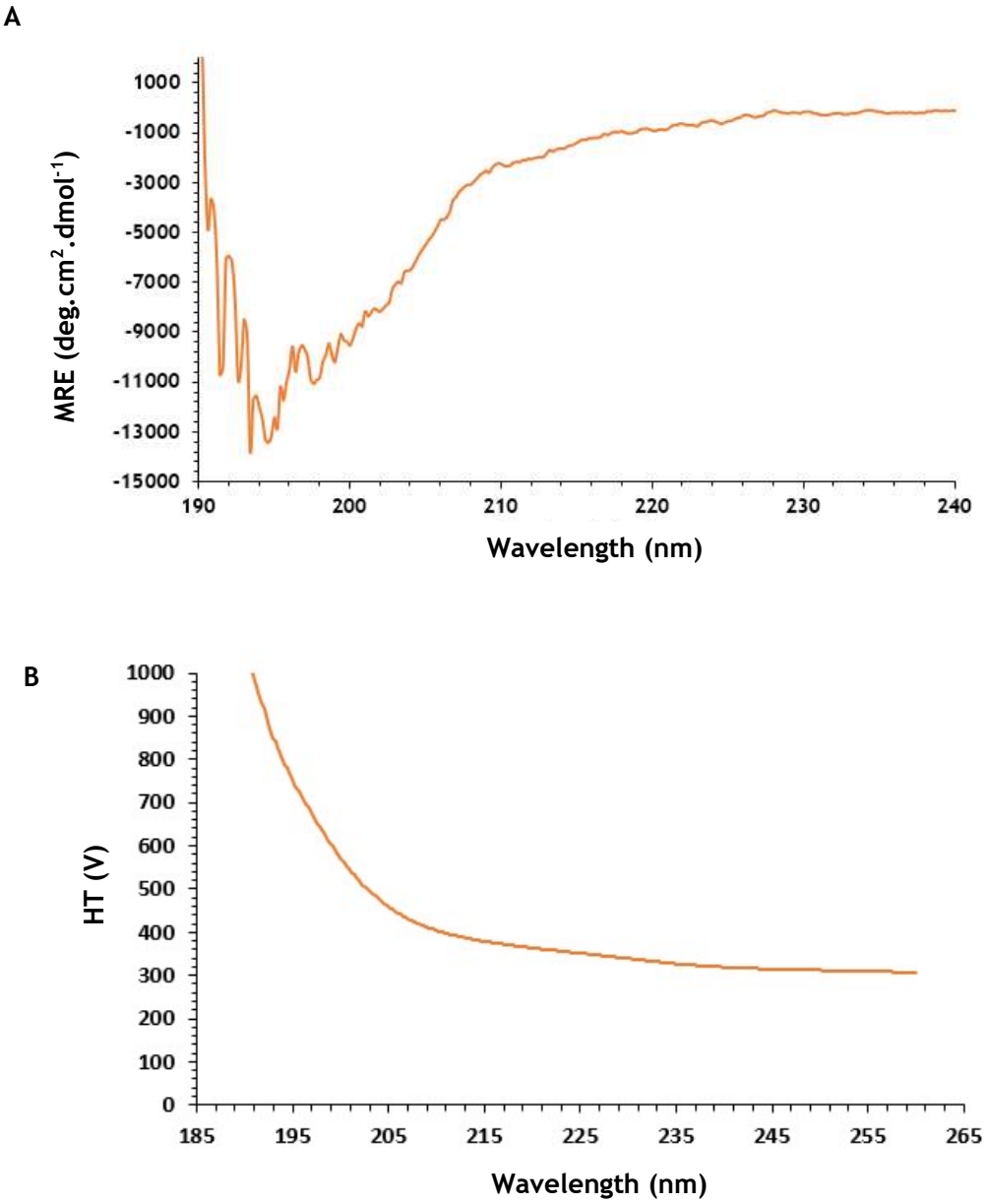


Figure 32: (A) Far-UV CD spectra of purified mutant p2/NC peptide and the corresponding HT (V) plot (B). Far-UV CD revealed the predominating structure of the peptide was that of a random coil, as seen by the characteristic trough below 200 nm. The HIV gag has been associated with a largely disordered structure until bound to the PR (Greenfield, 2006; Prabu-Jeyabalan *et al.,* 2002).

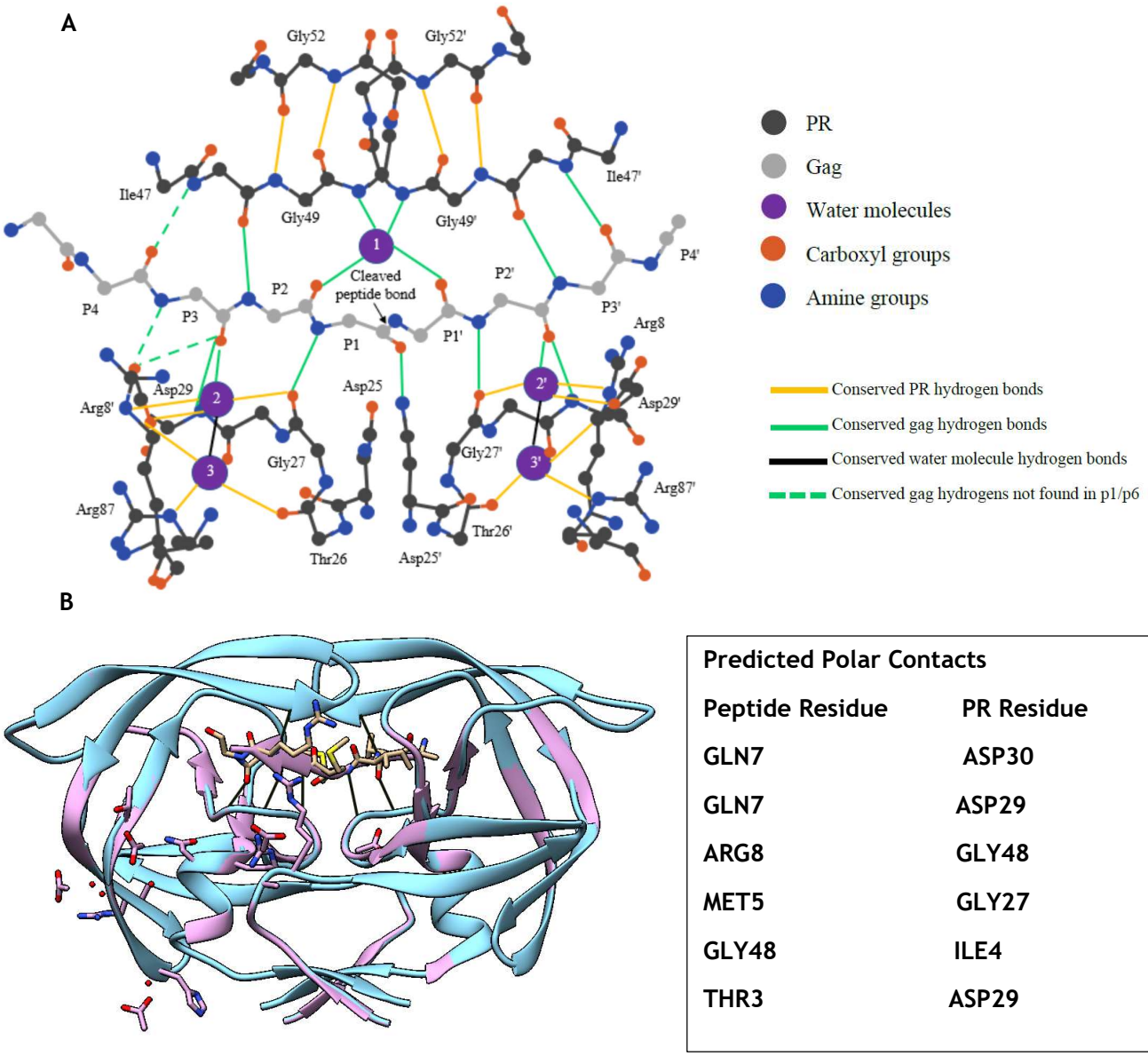
3.6 Evaluation of the mutant-related peptide binding to active wild Type and active mutant PR

3.6.1 Isothermal titration of a soluble HIV-1 Gag p2/NC-mimicking peptide against active wild type and mutant HIV-1 PR

To assess whether the newly synthesised peptides, representing an alternative gag cleavage site to that assessed in FRET-based assays, could be used as a biological tool — binding studies were carried out using ITC. The mutant p2/NC peptide, which could be purified and solubilised in PR storage buffer unlike the wild-type p2/NC peptide, was titrated at varying concentrations against 20 µM of the wild-type and mutant PR. The inactive PRs were not included in this study due to the concentration dependence of this study and the lack of AP binding in section 3.4.1. No noticeable binding was associated between the mutant p2/NC peptide and the wild-type PR nor the mutant PR, as seen by the raw thermal data which was characteristic of heats of dilution (Supplementary Figure 4). Due to the lack of *in vitro* success in studying binding between the generated peptide and the PRs, molecular docking studies were employed to assess how the experimental design of the peptides could be improved upon in the future, by comparison to published subtype B PR-peptide crystal structures.

3.6.2 Hierarchal docking simulations of the mutant-related peptide against active wild type and mutant HIV-1 PR

Docking simulations provided a predicted binding energy score (kcal/mol), a reflection of the predicted stability of a complex, as well as structural models of various peptide sequences probed in this investigation. The subtype B crystal structure (1kj7), a complex of the subtype B PR and related p2/NC gag peptide, was used as a control to investigate future considerations for optimising subtype C peptide design for improved binding study results. The molecular docking system was validated through redocking of the subtype B crystal structure (1kj7) in the HPEPDOCK program. After validation, further docking studies with the synthetic mutant peptide were undertaken to evaluate predicted structural complexes with the peptides and the structural obstacles that may have reduced mutant peptide and PR interactions, identified by no strong binding interaction within the ITC study (section 3.6.1). To validate the docking system as stated above, the subtype B PR structure and a model of the related p2/NC peptide, generated in I-TASSER, were docked using the hierarchal docking program HPEPDOCK. Upon comparison of the crystal structure and the generated model, a backbone RMSD was calculated to a value of 0 Å between the complexes, validating the docking system for further use. The p2/NC adopted a similar conformation and orientation within the active site of the subtype B PR when compared to the published crystal structure. Furthermore, several of the conserved hydrogen bonds were re-generated in the model when compared to the crystal complex (Figure 33). Thus, it was concluded that the docking system had reliably reproduced the 1kj7 crystal structure and could be used for the generation of further HIV PR-peptide complexes.



Predicted Polar Contacts	
Peptide Residue	**PR Residue**
GLN7	ASP30
GLN7	ASP29
ARG8	GLY48
MET5	GLY27
GLY48	ILE4
THR3	ASP29

Figure 33: (A) Conserved polar contacts across multiple subtype B PR-gag peptide crystal complexes (Prabu-Jeyabalan *et al.*, 2002). (B) Alignment of predicted Subtype B PR-peptide complex and the corresponding crystal structure. The backbone RMSD between the HPEPDOCK modelled complex and crystal structure was found to be 0 Å. Furthermore, many of the predicted polar contacts within the modelled complex corresponded to conserved polar contacts found in (A).

To assess potential reasons for poor binding of the synthetic mutant peptide to the subtype C PRs, models of the potential complex between wild-type PR and mutant synthetic peptides as well as mutant PR and mutant synthetic peptide were generated using the validated HPEPDOCK program and compared to the subtype B PR-p2/NC crystal complex. Firstly, the mutant PR model was generated in SWISS-MODEL using the wild-type crystal structure (3u71) and validated for structural reproducibility. Docking of the mutant synthetic peptide to the wild-type and mutant PRs revealed less favourable binding energy (-156,148 kcal/mol and -160,34 kcal/mol, respectively) than the reproduced subtype B complex (-276.126 kcal/mol) (Figure 34). There was a small difference between the binding energies of the synthetic peptide-mutant PR complex (-160,34 kcal/mol) and the peptide-wild-type complex (-156,148 kcal/mol). This correlated with the ITC results suggesting neither wild-type nor mutant peptide bound strongly to the synthetic mutant peptide and that the overall interaction was probably extremely weak, much weaker than the peptide-PR interaction that was crystalised to generate the subtype B complex. The minimal difference in binding energies could suggest that the mutations associated with the mutant peptide would generate a more stable complex with the mutant PR as opposed to the wild-type PR, however, this was not considered significant as neither PRs were proven to bind to the peptides in ITC studies.

Upon further evaluation, it was noted that the size of the synthetic mutant peptide could be associated with limited entry into the active site compared to the subtype B peptide which was kept below 10 amino acid residues (Prabu-Jeyabalan *et al.,* 2002). Although the YKK tail was necessary to improve solubility and other empirical qualities of the peptide, this large tail could be generating a steric obstruction, preventing the synthetic mutant peptide from entering the active site completely and orientating the cleavage site over the catalytic triad (Figure 35). Both these factors are necessary for establishing the appropriate bonds that are seemingly conserved across subtype B-peptide complexes and would have promoted a more stable interaction between the synthetic mutant peptide and the PRs of interest (Figure 33) (Prabu-Jeyabalan *et al.,* 2002). In the generated models of wild-type and mutant PRs with the synthetic mutant peptide, the cleavage site is not orientated over the catalytic triad (Figure 35). Furthermore, a large volume or surface area of the peptide extends outside the PR active site. Subsequently, the flaps are largely open, and upon further assessment few conserved bonds were predicted to form within these complexes (Figure 35). Both the wild-type PR-peptide and mutant PR-peptide complex displayed evidence of only one potentially conserved bond between the P1ˈV residue of the peptide and the PR Asp29 residue.

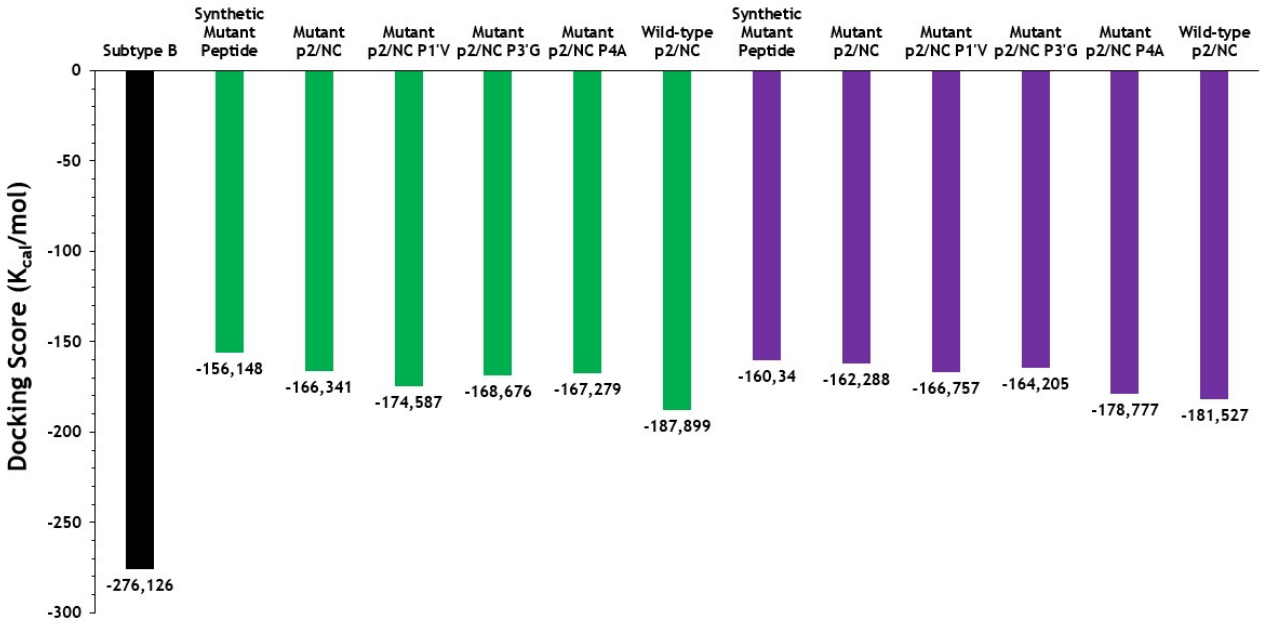


Figure 34: Predicted binding energies of peptides to PRs (kcal/mol) generated using molecular docking simulations in HPEPDOCK. The wild-type PR (green) was docked to the p2/NC mutant synthetic peptide, p2/NC wild-type sequence, p2/NC mutant sequence, and each mutation (P1ˈV, P3ˈG, and P4A) in isolation. The binding energies of these models was compared to docking studies with the same theoretical ligands and the mutant PR (purple). Furthermore, the 1kj7 subtype B crystal structure was re-docked and included as a validation tool and control (black). These docking scores roughly represent the ligand/protein binding energy. The lower the value the better the docking is and thus more stable the complex is.

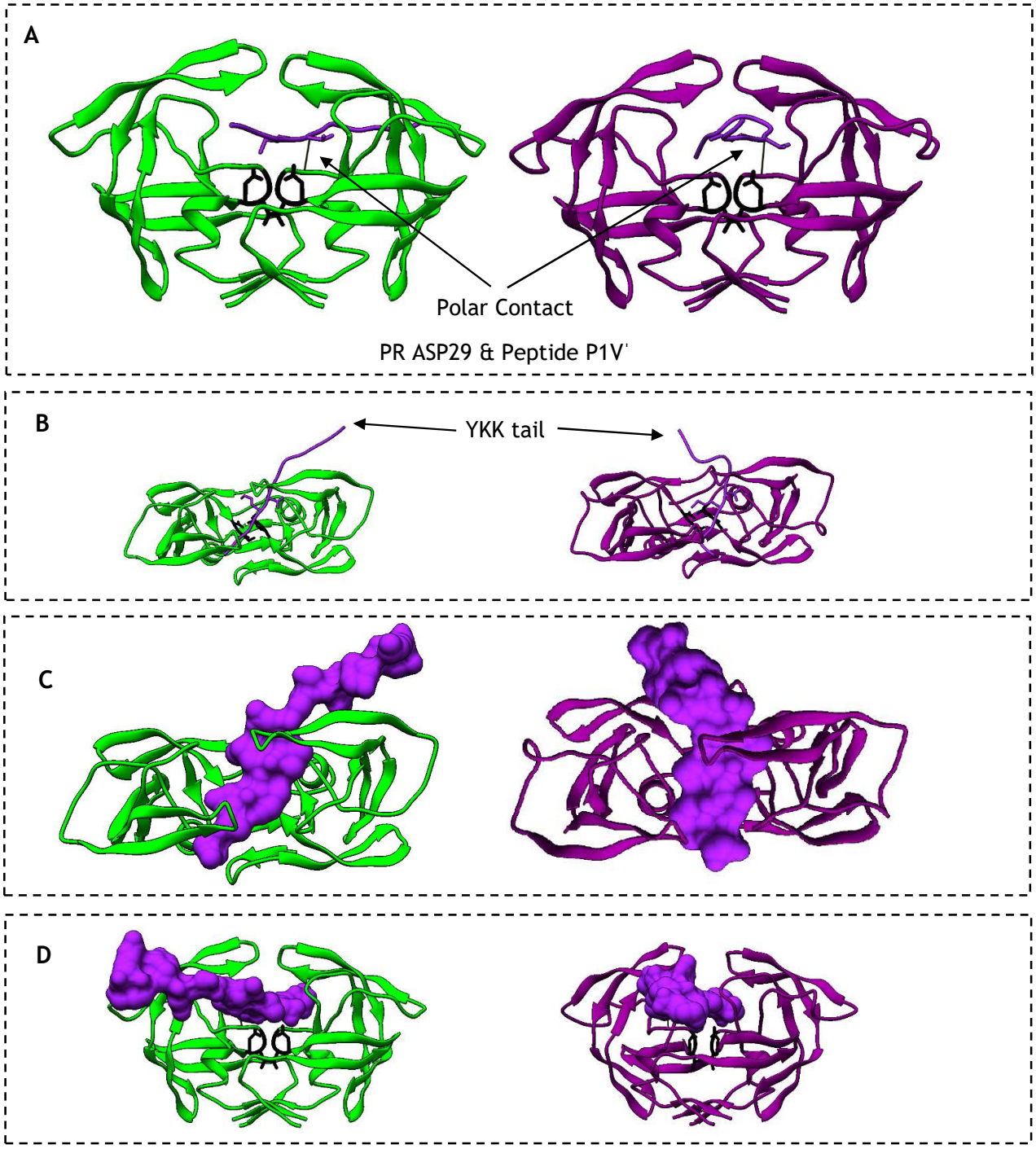


Figure 35: HPEPDOCK model of the complexes of the wild-type PR (green) and mutant PR (purple) in complex with the synthetic mutant peptide synthesised and utilised in ITC binding studies. The cleavage site is not orientated over the catalytic triad (A & B), at least 50% of the peptide resides outside the active site (C), and few conserved bonds were predicted to form (A). The size and YKK tail could be generating a steric obstruction that doesn't promote PR-peptide interaction (C & D).

Docking of the mutant peptide sequence alone (i.e., excluding the YKK tail) showed an improvement in binding energy suggesting greater stability upon binding. The wild-type PR-peptide complex generated a score of -166,341 kcal/mol and the mutant PR-peptide complex generated a score of -162,288 kcal/mol upon removal of the YKK from the sequence, emphasising its obstructive nature. This docking investigation re-iterated the need to re-evaluate the peptide design in the terms of size and limit large terminal residues. Future considerations would have to be made regarding the size of the peptide and the inclusion or exclusion of residues that may obstruct binding in an experimental context. In the future molecular dynamic simulations could be applied in parallel to the design of the peptide while optimising empirical qualities.

Docking simulations were further applied to assess in an *in-silico* environment whether mutations associated with the mutant peptide produced any compensatory effect towards the mutations in the related mutant PR. Furthermore, which mutations should be carried forward for *in vitro* and molecular dynamics studies in the future, if any noticeable difference between wild-type p2/NC and mutant p2/NC interactions is noted. According to Figure 34, the wild-type p2/NC sequence associates with greater stability within the wild-type PR active site than the mutant p2/NC sequence producing predicted binding energies of -187,899 kcal/mol and -166.341 kcal/mol, respectively. Similarly, the wild-type p2/NC sequence associates with greater stability within the mutant PR active site than the mutant p2/NC sequence producing binding energies of -181,527 kcal/mol and -160,34 kcal mol, respectively. This suggests that the triple mutation within the mutant p2/NC site may not act to compensate for mutations present in the mutant PR as the wild-type p2/NC sequence is still associated with greater stability.

Upon examination of the mutations in isolation with the wild-type PR, it was found that the presence of each mutation (P1ˈV, P3ˈG, and P4A) contributes to a less stable complex than the wild-type PR with the wild-type p2/NC sequence. The inclusion of the P1ˈV, P3ˈG, and P4A peptide mutations in complex with the wild-type PR produced binding energies of -174,587 kcal/mol, -168,676 kcal/mol, and -167, -279 kcal/mol, respectively, all higher than the binding energy of the wild-type PR in complex with the wild-type p2/NC sequence of -187,899 kcal/mol. This suggests these mutations confer less stability in the context of wild-type PR binding. These binding energies were still lower than the wild-type PR in complex with the triple mutant p2/NC sequence, with a binding energy of -166,341 kcal/mol. Thus, reverse substitution within the docking analysis suggests these mutations do not have a functional role to play in complex with the wild-type PR.

Furthermore, the reverse substitution of the mutations in the complex with the mutant PR revealed that the P1’V (with an associated binding energy of -166,757 kcal/mol) and P3G (with an associated binding energy of -164,205 kcal/mol) mutations contributed to complexes that were not significantly more stable than the mutant PR in complex with the triple mutant sequence (with an associated binding energy of -162,288 kcal/mol). Furthermore, these complexes were found to be less stable than the mutant PR-

wild-type p2/NC complex (with an associated binding energy of -181,527 kcal/mol), the mutant PR complex with the lowest binding energy and greater predicted stability. Unexpectedly the P4A mutation in isolation was shown to restore the complex binding energy and predicted stability to -178,777 kcal/mol, almost equal to the mutant PR-wild-type p2/NC complex scoring -181,527 kcal/mol. Molecular docking studies can only be limited to preliminary probing of these mutations due to their computational limitations. Future studies would thus include improvements to the synthetic peptide designs and isolated study of this P4A mutation for compensatory effects in the context of the mutant PR. Such studies would include experimental analysis of this mutation using wet-lab binding techniques and molecular dynamic simulations.

In conclusion, the docking studies suggested poor results in ITC binding studies could be attributed to the size of the synthesised peptide and steric obstruction caused by the YKK tail. The peptide would need to be reduced in size, and studies focused on critical residues and improvements to solubility and other empirical properties should be generated in more subtle ways. This way even includes further solubility trials aimed at the PR to identify new storage buffers that support both the PR and the peptide as opposed to preferentially the PR alone. Future studies may focus on the P4A mutation as a compensatory mutation to promote p2/NC and mutant PR binding as identified in the docking studies as a mutation that restored complex stability to that of the mutant PR in complex with the wild-type p2/NC sequence. Reduced flap flexibility attributed to the insertion mutation could correlate with the need to replace bulky residues such as the Thr at position P4 with a smaller residue Ala, to maintain mutant PR binding with the p2/NC gag cleavage site. This can only be considered as a preliminary insight, due to the inherent limitations of computational docking studies and should be supported with molecular dynamics investigations before being considered for the wet-lab study.

3.7 Concluding Remarks

The advancement of antiretrovirals (ARVs) has rapidly improved the treatment of disease progression in infected individuals by controlling viral replicative capacity. HIV possesses notable genetic diversity which can be accredited to a fallible reverse transcriptase combined with a high viral replication rate. This results in the continuous generation and selection of naturally occurring polymorphisms (NOPs) in viral proteins that pose as major drug targets, currently. For example, the interactions between PR and the gag polyproteins are potent targets due to the essential nature of polyprotein cleavage by PR for the formation of infectious mature virions (Chan & Kim, 1998; Wyatt and Sodroski, 1998; Neagu *et al.*, 2018). As a result, HIV proteins display significant mutation rates which result in the accumulation of drug-resistant mutations in drug-selecting environments and eventually the introduction of these mutant strains into drug naïve populations (Naicker *et al.,* 2013; Williams *et al.,* 2019).

Failure to address the continually mutating virus will lead to an eventually reduced capacity to slow disease progression due to the depleting long-term sustainability of currently available treatments. Furthermore, evidence has shown that the accumulation of NOPs and subsequent selection for drug-resistant mutations are inherently different between subtypes. Increasing evidence is suggesting this inherent difference in genetic diversity introduces differences in disease progression and accepted and novel mechanisms to drug resistance between subtypes (Taylor *et al.,* 2008). As a result, this will ultimately impact treatment and vaccine development now and in the future (Taylor *et al.,* 2008).

The PR accumulates primary mutations which impact inhibitor and substrate binding, but high resistance phenotypes are dependent on the coupling of drug-resistant primary mutations and compensatory secondary mutations. Increasing evidence is suggesting that PR is not the only route to drug evasion but rather that gag could participate in pathways that lead to drug resistance (Su *et al.,* 2011). The L38↑N↑L^{+4} PR and L38↑N↑L^{+4} gag protein sequences were identified in PR inhibitor naïve child whose mother had not been treated with PIs either. This indicates the potential increase of variant pairs within the drug-naïve infected population, with multiple mutations in the PR and the related gag that could be structurally and functionally influential. The L38↑N↑L^{+4} PR (referred to as mutant PR from here onwards) is characterised by a significant hinge insertion mutation in addition to four background mutations. Hinge mutations are increasing in frequency in the HIV-infected population (Kuhn *et al.,* 2014). In comparison to the wild-type PR, the mutant PR has been characterised by decreased catalytic efficiency coupled with a reduced DRV susceptibility in the presence of the related mutant gag protein (Williams *et al.,* 2019). Therefore, the main aim of this study was to probe the structural differences and functional differences (in association with a related gag cleavage site) between the mutant PR and wild-type PR. Furthermore, preliminary assess any functionality associated with mutations present in the p2/NC cleavage site.

Inactive and active wild-type and mutant PRs were successfully expressed in BL21 (DƐ3) pLysS *E. coli* cells and purified to 98 % purity using dual ion-exchange chromatography coupled with a two-step chemical renaturation system using dialysis. Inactive PRs proved to express and purify to higher concentrations and working volumes than the active PRs suggesting their suitability as practical replacements for active PRs associated with bottlenecks in yield and obstacles such as proteolysis and *in vitro* instability. Global structural analysis via sequencing showed that all four PRs possessed the suitable SNPs and insertion mutations related to their PR identity (mutant or wild-type and inactive or active). The size discrepancy in the inactive wild-type was associated with intrinsic bridging sequences (related to the natural autoproteolytic cleavage sites in the gag polyprotein) that are cleaved by the active wild-type. This is not possible in the inactive wild-type as the D25A mutation knocks out the proteolytic activity.

Global structural analysis using far-UV CD suggested that secondary structure across all four PRs was characteristic of HIV-1 PR and conserved. Therefore, neither mutant SNPS (M46I, I54V, V82A, K20R, and L33F), insertion mutation (L38↑N↑L), nor the D25A mutation affect PR secondary structure. On the contrary, global tertiary structural evaluation by far-UV CD at an excitation wavelength of 295 nm suggested the insertion mutation impacts the tertiary structure, fold, and more than likely the dynamics of the PR hinge region. Similarly, to secondary structural analysis, SE-HPLC revealed the quaternary structure of the PR is conserved across all four PRs. Thus, neither the mutant SNPs, nor insertion mutation, nor D25A mutation were found to affect PR dimerisation. Based on global structural analysis, the D25A PR mutants were considered suitable for use in place of active PRs that express lower yields and present autoproteolytic obstacles. Furthermore, the global structural analysis provided further reason to assess the impact of the hinge insertion mutation on global functional characteristics of the PR in relation to processing of a related gag cleavage site.

Firstly, active site titrations were necessary to establish the true concentration of functional PR, before samples were utilised in concentration dependent enzyme assays. Active site titrations using ITC revealed that D25A mutations impact AP binding and that inactive PR constructs may not be suitable or clinically relevant to drug binding studies. However, inactive PRs could be utilised in other experimental investigations where autoproteolysis presents major obstacles such as chemical refolding studies. On the contrary, active site titrations of the active wild-type PR and mutant PR revealed binding to AP. Thus, the active wild-type and active mutant PRs could be found to be 34% and 10% recovered, respectively. This suggested that the chemical renaturation process via dialysis would need further optimisation to increase PR recovery. Post correction of the absorbance-based concentrations of the active PRs, FRET-based assays revealed a reduction in the mutant PR catalytic efficiency, when processing a related gag cleavage site. This was deduced to be related to the structural changes in the hinge region, as seen in tertiary structural analysis. This hinge is directly connected to the flaps which control substrate entry and form critical polar contacts with the natural gag cleavage sites, which the FRET-based peptide mimics. The functional influence of the hinge insertion was further evaluated in thermal stability studies in the context of clinical inhibitors.

Thermal shift assays showed that the apo mutant as well as the mutant-inhibitor complexes (including those with DRV and SQV) were less stable than the apo wild-type and wild-type-inhibitor complexes. Thus, the maintenance of PR hinge insertion mutations within the HIV population could be a novel route to future drug resistance. This was confirmed by previous computational studies that showed inhibitors such as DRV bound with reduce polar contacts within the active site of the PR more than likely linked to a change in the structural nature of the active site and reduce flap dynamics. Thus, thermal shift assays support the notion that the inclusion of hinge insertions in the PR could be a potential drug resistance mechanism in the context of not only DRV but SQV as well.

Previous studies also suggested that the presence of the related gag can confer drug resistance not noticed within the PR alone (Williams *et al.,* 2019). Peptide studies and computational docking studies were carried out to evaluate whether mutations in the p2/NC gag site could be acting in a compensatory fashion to support the complex profile of mutations associated with the mutant PR. Peptides mimicking the p2/NC sequence of the wild-type and mutant gag were designed and synthesised. However, only the mutant synthetic peptide could be subjected to further studies as the wild-type peptide remained insoluble in conditions that were PR-compatible. Due to the lack of *in vitro* success in studying ITC binding studies between the generated mutant peptide and the PRs, molecular docking studies were employed to assess how the experimental design of the peptides could be improved upon in the future, by comparison to published subtype B PR-peptide crystal structures. The docking studies suggested poor results in ITC binding studies could be attributed to the size of the synthesised peptide and steric obstruction caused by the YKK tail. The peptide design process would be coupled with more intensive computational analysis such as docking and dynamics studies in the future to promote a design strategy orientated around promoting PR-peptide interaction. Theoretical docking studies suggested that the P4A mutation could be the mutation in the mutant p2/NC site generating compensatory effects in the context of the mutant PR. This was shown by the restoration of the binding energy score to a score close to the binding score associated with the most stable mutant PR complex, the mutant PR, and the wild-type p2/NC sequence.

In conclusion, the hinge insertion does impact the structural and functional nature of the PR, alone and in the context of related gag cleavage sites. Furthermore the mutagenic profile associated with this mutant PR could be acting to promote drug resistance to known clinical inhibitors. Hinge insertions could be generating PR resistance followed by restored catalytic support by background compensatory mutations in the PR. Furthermore, at least one of the mutations located in the p2/NC cleavage of the mutant gag could be present to compensate for PR mutations, potentially the P4A mutation. Thus although PR mutations have been shown to compromise processing of related gag cleavage sites, there is the potential functional role of gag cleavage site mutations to act in a compensatory fashion. The increase of these mechanisms will render drastic mutations such as PR insertion mutations possible, highly supporting drug resistant phenotypes in the future. Such conclusions about the gag are limited to theoretical models and should be further investigated using molecular dynamics and improved peptide binding studies. Further studies should focus on the gag and PR as a functional pair, potential compensatory drug resistance mechanisms, and how to exploit this in the design of more flexible and sustainable inhibitors.

5. References

About HIVinfo.NIH.gov | NIH [WWW Document], n.d. URL https://hivinfo.nih.gov/about (accessed 3.16.23).

Adamson, C.S., 2012. Protease-Mediated Maturation of HIV: Inhibitors of Protease and the Maturation Process. Mol Biol Int 2012, 604261. https://doi.org/10.1155/2012/604261

Ammaranond, P., Cunningham, P., Oelrichs, R., Suzuki, K., Harris, C., Leas, L., Grulich, A., Cooper, D.A., Kelleher, A.D., 2003. Rates of transmission of antiretroviral drug-resistant strains of HIV-1. J Clin Virol 26, 153–161. https://doi.org/10.1016/s1386-6532(02)00114-2

Andreotti, G., Monticelli, M., Cubellis, M.V., 2015. Looking for protein stabilizing drugs with thermal shift assay. Drug Test Anal 7, 831–834. https://doi.org/10.1002/dta.1798

Birdsall, B., King, R.W., Wheeler, M.R., Lewis, C.A., Goode, S.R., Dunlap, R.B., Roberts, G.C., 1983. Correction for light absorption in fluorescence studies of protein-ligand interactions. Anal Biochem 132, 353–361. https://doi.org/10.1016/0003-2697(83)90020-9

Chatterjee, A., Mridula, P., Mishra, R.K., Mittal, R., Hosur, R.V., 2005. Folding Regulates Autoprocessing of HIV-1 Protease Precursor*. Journal of Biological Chemistry 280, 11369–11378. https://doi.org/10.1074/jbc.M412603200

Fee, E., Parry, M., 2008. Jonathan Mann, HIV/AIDS, and Human Rights. Journal of Public Health Policy 29, 54–71.

Fun, A., Wensing, A.M., Verheyen, J., Nijhuis, M., 2012a. Human Immunodeficiency Virus gag and protease: partners in resistance. Retrovirology 9, 63. https://doi.org/10.1186/1742-4690-9-63

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server, in: Walker, J.M. (Ed.), The Proteomics Protocols Handbook, Springer Protocols Handbooks. Humana Press, Totowa, NJ, pp. 571–607. https://doi.org/10.1385/1-59259-890-0:571

Gatanaga, H., Suzuki, Y., Tsang, H., Yoshimura, K., Kavlick, M.F., Nagashima, K., Gorelick, R.J., Mardy, S., Tang, C., Summers, M.F., Mitsuya, H., 2002. Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. J Biol Chem 277, 5952–5961. https://doi.org/10.1074/jbc.M108005200

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

Giandhari, J., Basson, A.E., Sutherland, K., Parry, C.M., Cane, P.A., Coovadia, A., Kuhn, L., Hunt, G., Morris, L., 2016. Contribution of Gag and Protease to HIV-1 Phenotypic Drug Resistance in Pediatric Patients Failing Protease Inhibitor-Based Therapy. Antimicrob Agents Chemother 60, 2248–2256. https://doi.org/10.1128/AAC.02682-15

Global HIV & AIDS statistics — 2020 fact sheet [WWW Document], n.d. URL https://www.unaids.org/en/resources/fact-sheet (accessed 4.14.21).

Greenfield, N.J., 2006a. Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc 1, 2876–2890. https://doi.org/10.1038/nprot.2006.202

Grimsley, G.R., Pace, C.N., 2003. Spectrophotometric Determination of Protein Concentration. Current Protocols in Protein Science 33, 3.1.1-3.1.9. https://doi.org/10.1002/0471140864.ps0301s33

Hall, B.G., Acar, H., Nandipati, A., Barlow, M., 2014. Growth Rates Made Easy. Mol Biol Evol 31, 232–238. https://doi.org/10.1093/molbev/mst187

Hayashi, H., Takamune, N., Nirasawa, T., Aoki, M., Morishita, Y., Das, D., Koh, Y., Ghosh, A.K., Misumi, S., Mitsuya, H., 2014. Dimerization of HIV-1 protease occurs through two steps relating to the mechanism of protease dimerization inhibition by darunavir. Proc Natl Acad Sci U S A 111, 12234–12239. https://doi.org/10.1073/pnas.1400027111

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

Huang, X., Britto, M.D., Kear-Scott, J.L., Boone, C.D., Rocca, J.R., Simmerling, C., Mckenna, R., Bieri, M., Gooley, P.R., Dunn, B.M., Fanucci, G.E., 2014. The Role of Select Subtype Polymorphisms on HIV-1 Protease Conformational Sampling and Dynamics. J Biol Chem 289, 17203–17214. https://doi.org/10.1074/jbc.M114.571836

J. Miles, A., A. Wallace, B., 2016. Circular dichroism spectroscopy of membrane proteins. Chemical Society Reviews 45, 4859–4872. https://doi.org/10.1039/C5CS00084J

Kear, J.L., Blackburn, M.E., Veloro, A.M., Dunn, B.M., Fanucci, G.E., 2009. Subtype Polymorphisms Among HIV-1 Protease Variants Confer Altered Flap Conformations and Flexibility. J. Am. Chem. Soc. 131, 14650–14651. https://doi.org/10.1021/ja907088a

Kempf, D.J., Marsh, K.C., Paul, D.A., Knigge, M.F., Norbeck, D.W., Kohlbrenner, W.E., Codacovi, L., Vasavanonda, S., Bryant, P., Wang, X.C., 1991. Antiviral and pharmacokinetic properties of C2 symmetric inhibitors of the human immunodeficiency virus type 1 protease. Antimicrob Agents Chemother 35, 2209–2214.

Kilz, P., 1999. 9 - Design, Properties, and Testing of Polymer Standards Service Size Exclusion Chromatography (SEC) Columns and Optimization of SEC Separations, in: Wu, C. (Ed.), Column Handbook for Size Exclusion Chromatography. Academic Press, San Diego, pp. 267–303. https://doi.org/10.1016/B978-012765555-0/50010-5

Kožíšek, M., Henke, S., Šašková, K.G., Jacobs, G.B., Schuch, A., Buchholz, B., Müller, V., Kräusslich, H.-G., Řezáčová, P., Konvalinka, J., Bodem, J., 2012. Mutations in HIV-1 gag and pol Compensate for the Loss of Viral Fitness Caused by a Highly Mutated Protease. Antimicrob Agents Chemother 56, 4320–4330. https://doi.org/10.1128/AAC.00465-12

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

Lastere, S., Dalban, C., Collin, G., Descamps, D., Girard, P.-M., Clavel, F., Costagliola, D., Brun-Vezinet, F., NARVAL Trial Group (ANRS 088), 2004. Impact of insertions in the HIV-1 p6 PTAPP region on the virological response to amprenavir. Antivir Ther 9, 221–227.

Lefebvre, E., Schiffer, C.A., 2008. Resilience to resistance of HIV-1 protease inhibitors: profile of darunavir. AIDS Rev 10, 131–142.

Lo, M.-C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M., Ellestad, G., 2004. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal Biochem 332, 153–159. https://doi.org/10.1016/j.ab.2004.04.031

Longworth, J.W., 1971. Luminescence of Polypeptides and Proteins, in: Steiner, R.F., Weinryb, I. (Eds.), Excited States of Proteins and Nucleic Acids. Springer US, Boston, MA, pp. 319–484. https://doi.org/10.1007/978-1-4684-1878-1_8

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

Mansky, L.M., Temin, H.M., 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J Virol 69, 5087–5094. https://doi.org/10.1128/JVI.69.8.5087-5094.1995

Marciniszyn, J., Hartsuck, J.A., Tang, J., 1976. Mode of inhibition of acid proteases by pepstatin. Journal of Biological Chemistry 251, 7088–7094. https://doi.org/10.1016/S0021-9258(17)32945-9

Matthews, M.A., 2000. A to Z of Thermodynamics By Pierre Perrot (Université des sciences et technologies de Lille). Oxford University Press: Oxford, New York, and Tokyo. 1998. vi + 329 pp. $65.00. ISBN 0-19-856556-9 (Hardback). J. Am. Chem. Soc. 122, 3799–3800. https://doi.org/10.1021/ja995706b

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

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

Biomolecular Structure and Dynamics 31, 1370–1380. https://doi.org/10.1080/07391102.2012.736774

Nijhuis, M., van Maarseveen, N.M., Lastere, S., Schipper, P., Coakley, E., Glass, B., Rovenska, M., de Jong, D., Chappey, C., Goedegebuure, I.W., Heilek-Snyder, G., Dulude, D., Cammack, N., Brakier-Gingras, L., Konvalinka, J., Parkin, N., Kräusslich, H.-G., Brun-Vezinet, F., Boucher, C.A.B., 2007. A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. PLoS Med 4, e36. https://doi.org/10.1371/journal.pmed.0040036

Parry, C.M., Kolli, M., Myers, R.E., Cane, P.A., Schiffer, C., Pillay, D., 2011. Three Residues in HIV-1 Matrix Contribute to Protease Inhibitor Susceptibility and Replication Capacity. Antimicrobial Agents and Chemotherapy 55, 1106–1113. https://doi.org/10.1128/AAC.01228-10

Peme, T., Brady, D., Juma, W., Makatini, M., 2021. Development of fructose-1,6-bisphosphate aldolase enzyme peptide mimics as biocatalysts in direct asymmetric aldol reactions. RSC Adv. 11, 36670–36681. https://doi.org/10.1039/D1RA06616A

Penno, C., Kumari, R., Baranov, P.V., van Sinderen, D., Atkins, J.F., 2017. Specific reverse transcriptase slippage at the HIV ribosomal frameshift sequence: potential implications for modulation of GagPol synthesis. Nucleic Acids Res 45, 10156–10167. https://doi.org/10.1093/nar/gkx690

Pettit, S.C., Lindquist, J.N., Kaplan, A.H., Swanstrom, R., 2005. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. Retrovirology 2, 66. https://doi.org/10.1186/1742-4690-2-66

Prabu-Jeyabalan, M., Nalivaika, E., Schiffer, C.A., 2002a. Substrate Shape Determines Specificity of Recognition for HIV-1 Protease: Analysis of Crystal Structures of Six Substrate Complexes. Structure 10, 369–381. https://doi.org/10.1016/S0969-2126(02)00720-7

Rodger, A., 2013. Far UV Protein Circular Dichroism, in: Roberts, G.C.K. (Ed.), Encyclopedia of Biophysics. Springer, Berlin, Heidelberg, pp. 726–730. https://doi.org/10.1007/978-3-642-16712-6_634

Rosé, J.R., Salto, R., Craik, C.S., 1993. Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions. J Biol Chem 268, 11939–11945.

Sayer, J.M., Louis, J.M., 2009. INTERACTIONS OF DIFFERENT INHIBITORS WITH ACTIVE-SITE ASPARTYL RESIDUES OF HIV-1 PROTEASE AND POSSIBLE RELEVANCE TO PEPSIN. Proteins 75, 556–568. https://doi.org/10.1002/prot.22271

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

Sham, H.L., Kempf, D.J., Molla, A., Marsh, K.C., Kumar, G.N., Chen, C.M., Kati, W., Stewart, K., Lal, R., Hsu, A., Betebenner, D., Korneyeva, M., Vasavanonda, S., McDonald, E., Saldivar, A., Wideburg, N., Chen, X., Niu, P., Park, C., Jayanti, V., Grabowski, B., Granneman, G.R., Sun, E., Japour, A.J., Leonard, J.M., Plattner, J.J., Norbeck, D.W., 1998. ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. Antimicrob Agents Chemother 42, 3218–3224. https://doi.org/10.1128/AAC.42.12.3218

Sievers, F., Higgins, D.G., 2014. Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol 1079, 105–116. https://doi.org/10.1007/978-1-62703-646-7_6

Smialowski, P., Martin-Galiano, A.J., Mikolajka, A., Girschick, T., Holak, T.A., Frishman, D., 2007. Protein solubility: sequence based prediction and experimental verification. Bioinformatics 23, 2536–2542. https://doi.org/10.1093/bioinformatics/btl623

Srivastava, V.K., Yadav, R., 2019. Chapter 9 - Isothermal titration calorimetry, in: Misra, G. (Ed.), Data Processing Handbook for Complex Biological Data Sources. Academic Press, pp. 125–137. https://doi.org/10.1016/B978-0-12-816548-5.00009-5

Su, C.T.-T., Kwoh, C.-K., Verma, C.S., Gan, S.K.-E., 2018. Modeling the full length HIV-1 Gag polyprotein reveals the role of its p6 subunit in viral maturation and the effect of non-cleavage site mutations in protease drug resistance. J Biomol Struct Dyn 36, 4366–4377. https://doi.org/10.1080/07391102.2017.1417160

Su CT, Koh DW, Gan SK., 2019. Reviewing HIV-1 Gag Mutations in Protease Inhibitors Resistance: Insights for Possible Novel Gag Inhibitor Designs. Molecules. 2019 Sep 6;24(18):3243. doi: 10.3390/molecules24183243

Sun, Z., Liu, Q., Qu, G., Feng, Y., Reetz, M.T., 2019. Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. Chem Rev 119, 1626–1665. https://doi.org/10.1021/acs.chemrev.8b00290

Sundquist, W.I., Kräusslich, H.-G., 2012. HIV-1 Assembly, Budding, and Maturation. Cold Spring Harb Perspect Med 2. https://doi.org/10.1101/cshperspect.a006924

Szeltner, Zoltán, Polgár, L., 1996. Conformational Stability and Catalytic Activity of HIV-1 Protease Are Both Enhanced at High Salt Concentration (∗). Journal of Biological Chemistry 271, 5458–5463. https://doi.org/10.1074/jbc.271.10.5458

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24, 1596–1599. https://doi.org/10.1093/molbev/msm092

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

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

Voshavar, C., 2019. Protease Inhibitors for the Treatment of HIV/AIDS: Recent Advances and Future Challenges. Curr Top Med Chem 19, 1571–1598. https://doi.org/10.2174/1568026619666190619115243

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

Wright, E.R., Schooler, J.B., Ding, H.J., Kieffer, C., Fillmore, C., Sundquist, W.I., Jensen, G.J., 2007. Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. EMBO J 26, 2218–2226. https://doi.org/10.1038/sj.emboj.7601664

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. Nat Methods 12, 7–8. https://doi.org/10.1038/nmeth.3213

Zhou, P., Jin, B., Li, H., Huang, S.-Y., 2018. HPEPDOCK: a web server for blind peptide-protein docking based on a hierarchical algorithm. Nucleic Acids Res 46, W443–W450. https://doi.org/10.1093/nar/gky357

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

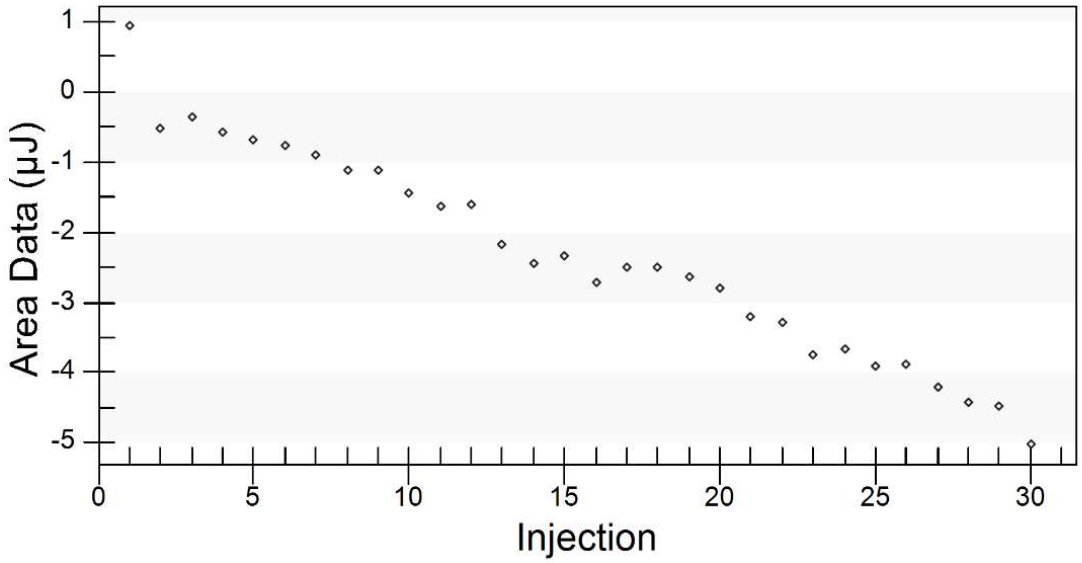
Zondagh, J., Williams, A., Achilonu, I., Dirr, H.W., Sayed, Y., 2018b. Overexpression, Purification and Functional Characterisation of Wild-Type HIV-1 Subtype C Protease and Two Variants Using a Thioredoxin and His-Tag Protein Fusion System. Protein J 37, 369–379. https://doi.org/10.1007/s10930-018-9779-5

5. Supplementary Section

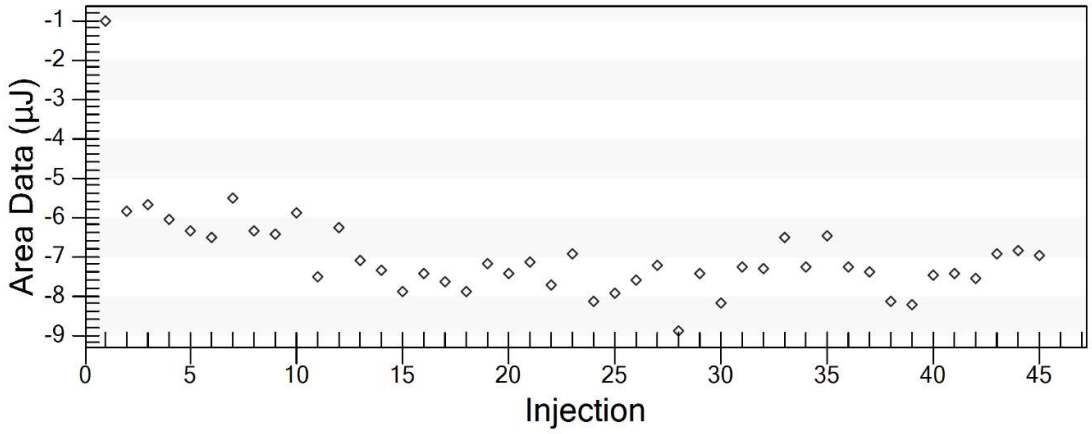
5.1 Active site titrations of purified inactive HIV-1 PRs against acetyl-pepstatin using ITC

Active site titrations are necessary to establish the population of PR that is native and functionally folded, which cannot be determined by spectrophotometric methods. Titration of AP against the inactive PRs revealed no noticeable binding.

A



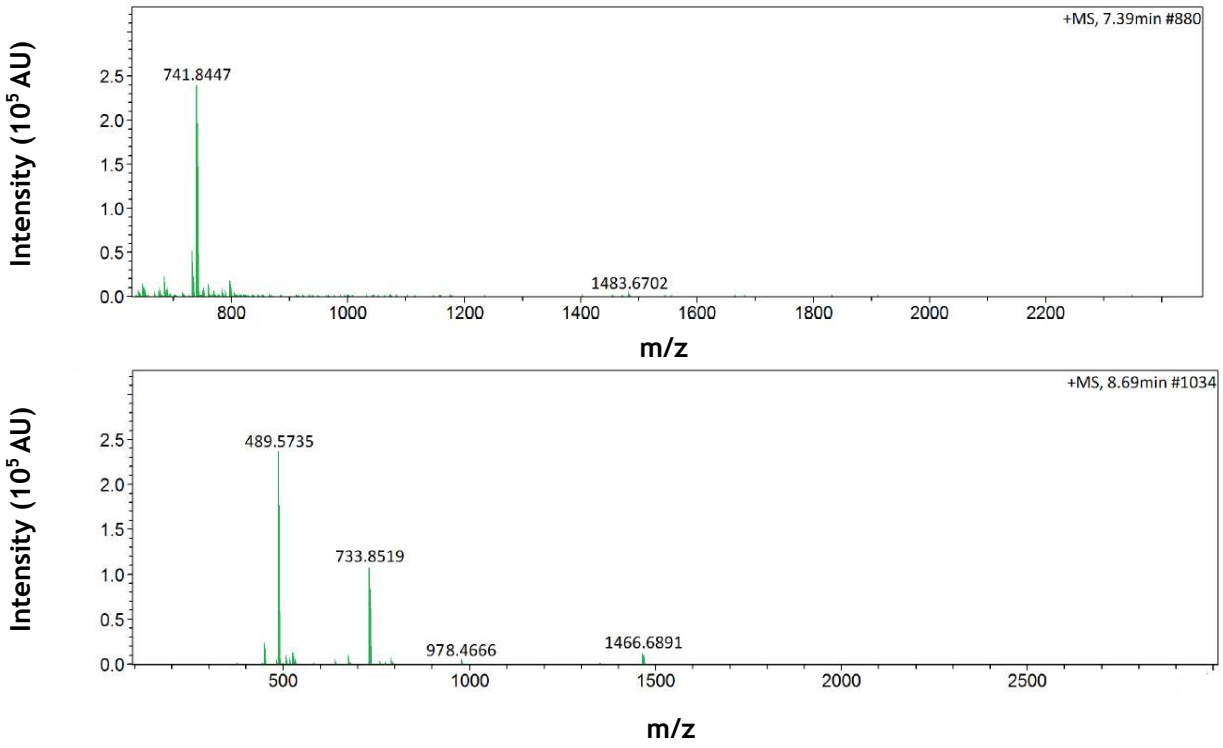
B



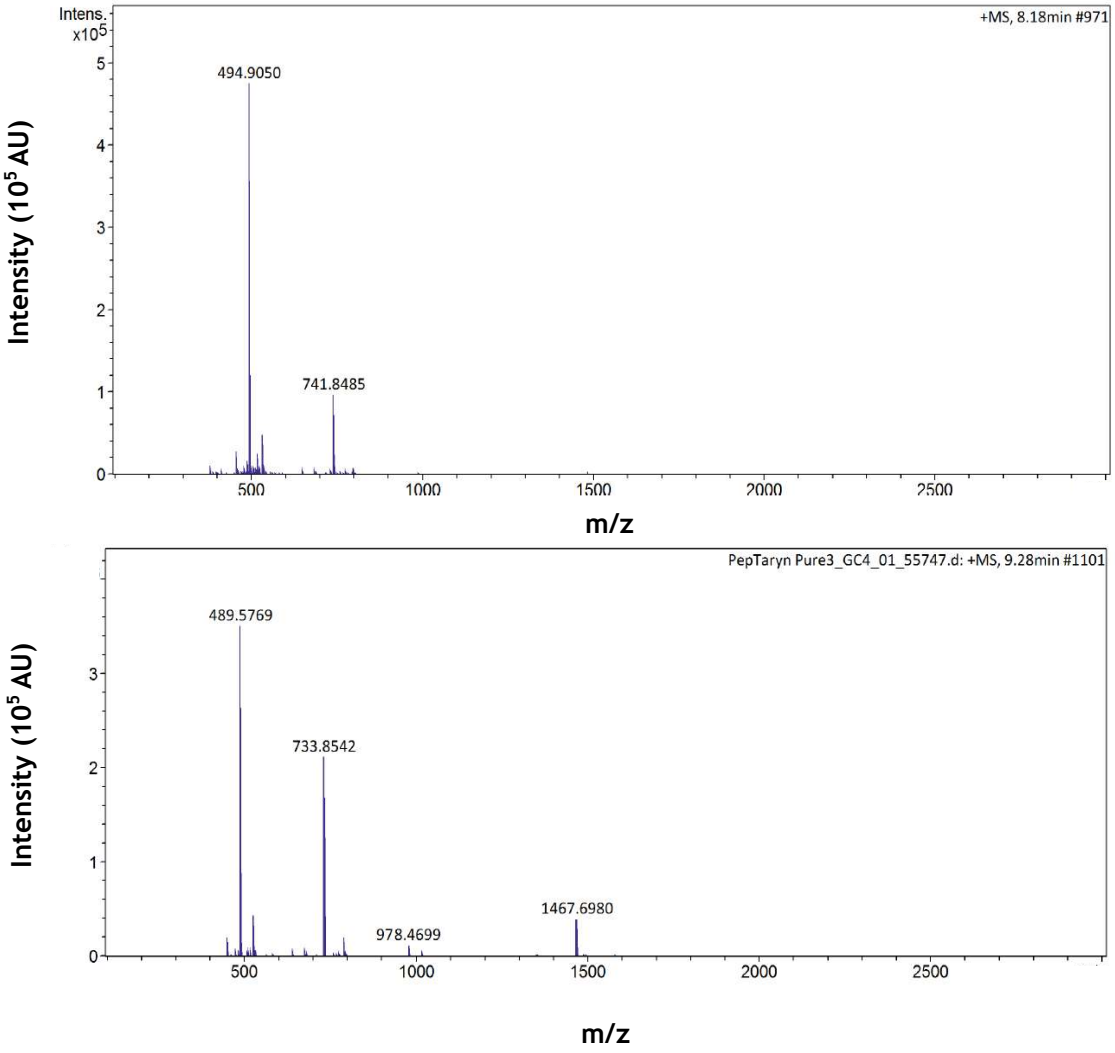
Supplementary Figure 1: Raw thermal data of active site titrations involving AP against inactive wild-type PR (A) and inactive mutant PR (B). No noticeable binding was detected between AP and inactive PRs suggesting the D25A mutation could knock out a polar contact important to the stable interaction of the AP with the PR active site. All raw heat data generated within the reactions were characteristic of heats of dilution indicating no binding interaction between AP and both the inactive wild-type and inactive mutant PRs. This suggested that the D25A mutation could be compromising the stable interaction of acetyl-pepstatin, characterised as a weak inhibitor, within the active site.

5.2 LC-MS profiles of purified mutant p2/NC peptide from fractions two and three

The three fractions of purified mutant peptide were subjected to LC-MS to evaluate whether the peptide fractions consisted of complete 13 amino acid peptides with few contaminating peptide species. LC-MS of the purified mutant peptide revealed sized and charged fragments that equalled multiples of the larger 13 amino acid peptide molecular weight. Furthermore, the LC-MS showed peaks corresponding to the total molecular weight of the mutant peptide, 1466 Da



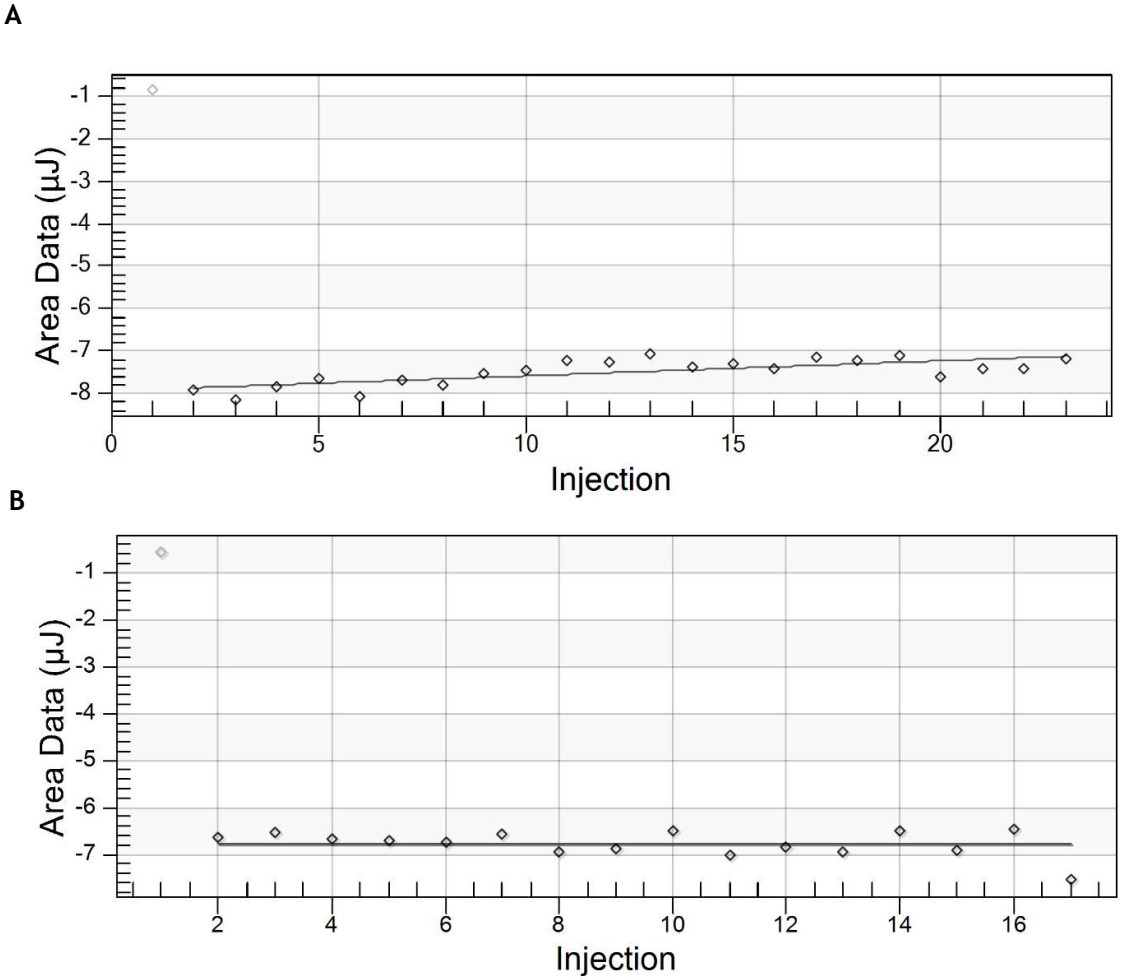
Supplementary Figure 2: LC-MS profile of purified mutant p2/NC peptide. LC-MS of the purified mutant peptide revealed sized and charge fragments that equalled multiples of the larger 13 amino acid peptide (A & B). Furthermore, the LC-MS showed peaks corresponding to the total molecular weight of the mutant peptide, 1466 Da (B).



Supplementary Figure 3: LC-MS profile of purified mutant p2/NC peptide. LC-MS of the purified mutant peptide revealed sized and charge fragments that equalled multiples of the larger 13 amino acid peptide (A & B). Furthermore, the LC-MS showed peaks corresponding to the total molecular weight of the mutant peptide, 1467 Da (B).

5.2 Active site titrations of purified active HIV-1 PRs against purified mutant p2/NC-mimicking peptide using ITC

To assess whether the newly synthesised peptides, representing an alternative gag cleavage site to that assessed in FRET-based assays, could be used as a biological tool — binding studies were carried out using ITC. No noticeable binding was associated between the mutant p2/NC peptide and the wild-type PR nor the mutant PR, as seen by the raw thermal data which was characteristic of heats of dilution.



Supplementary Figure 3: Raw thermal data of active site titrations involving the mutant p2/NC peptide against active wild-type PR (A) and active mutant PR (B). No noticeable binding was detected between the mutant peptide and either PR suggesting the peptide may not be forming a stable associated with the PR active site and the design may have to be revisited in the future. Both raw heat data profiles were characterised as heats of dilution generating only negative raw heat data and no significant change in the overall trend.