MOLECULAR MODELLING AND DRUG SUSCEPTIBILITY OF THE L38↑N↑L HIV-1 SUBTYPE C PROTEASE

Alison Williams
348481
May 2018

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

Supervisor: Professor Yasien Sayed

Co-supervisor: Doctor Ikechukwu Achilonu

Declaration

I, Alison Williams (Student number: 348481), am a student registered for the degree of Doctor of Philosophy (PhD) in the academic year 2018.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that all the work submitted for examination 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: 28 May 2018

-For Michael and Elsabé Williams-

“Success is not final, failure is not fatal: it is the courage to continue that counts.”

-Winston Churchill

Abstract

Human Immunodeficiency Virus (HIV) is a global concern due to the 36 million people infected worldwide. HIV is genetically diverse consisting of nine subtypes. Subtype C infections predominate in sub-Saharan Africa and this subtype has been under-investigated in comparison to subtype B. Great advances have been made to combat this disease, particularly in South Africa, but drug resistance still remains a concern. HIV protease cleaves the Gag and Gag-Pol polyproteins into their functional forms, making it indispensable to the production of infectious virions. It is, as such, a major drug target. The proteolytic enzyme accumulates mutations associated with drug resistance due to the high replication rate of the virus and the error prone reverse transcriptase. These include mutations in the active site and distal regions. Insertion mutations are rarely incorporated into the hinge region and because of that the effect of these insertions are poorly characterised. The variant protease in this study (L38↑N↑L) contains a double insertion of Asparagine and Leucine at position 38, in the hinge region. For the first time L38↑N↑L protease was successfully overexpressed and purified using a thioredoxin-hexahistidine tag fusion system. Molecular dynamics simulations showed that the flap region of L38↑N↑L was less dynamic than that of a wild-type protease, suggesting a possible mechanism to evade drug binding. Induced-fit docking studies showed that the drugs lopinavir, atazanavir and darunavir do bind L38↑N↑L albeit with reduced hydrophobic contacts and hydrogen bonds. *In vitro* inhibition studies confirmed that these drugs do bind and inhibit L38↑N↑L. The catalytic efficiency of L38↑N↑L was diminished compared to wild-type, which resulted in reduced replication capacity of the virus. Phenotypic assays showed that L38↑N↑L had reduced susceptibility to darunavir in the presence of a Gag sequence thus confirming that this region does contribute to drug resistance.

Research outputs

Papers accepted for publication

Protein Journal

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

Jake Zondagh[1], Alison Williams[1], Ikechukwu Achilonu, Heini W. Dirr, Yasien Sayed*
Protein Structure-Function Research Unit, School of Molecular and Cell Biology, University of Witwatersrand, Johannesburg, 2050, South Africa
[1]Both authors contributed equally to this work
*Corresponding author: Yasien Sayed

Papers Submitted for publication:

Submitted to the journal: Molecular Simulation

Molecular dynamic simulations of L38↑N↑L HIV-1 protease shows reduced flap dynamics

Alison Williams[a], Vijayakumar Balakrishna[a], Ikechukwu Achilonu[a], Heini Dirr[a] and Yasien Sayed[a*]
[a] Protein Structure-Function Research Unit, School of Molecular and Cell Biology, University of Witwatersrand, Johannesburg, 2050, South Africa
[*]Corresponding author

Conferences Attended

- Title of poster presented: Impact of L38↑N↑L insertions on expression of HIV-1 protease subtype C in South Africa

  Presented at:

  1) 24th South African Society of Biochemistry and Molecular Biology conference in July 2014 hosted by the University of Cape Town at Goudini Spa Resort, Rawsonville, Western Cape, South Africa
  2) The 6th Cross Faculty Symposium in October 2014 at the University of the Witwatersrand, Johannesburg, South Africa

- Title of poster presented: Novel expression system for HIV-1 subtype C protease

  Presented at the Biophysical Society Thematic Meeting: Biophysics in the Understanding, Diagnosis and Treatment of Infectious Diseases in November 2015 held at Spier Wine Farm, Stellenbosch, South Africa

Acknowledgments

I would like to thank my supervisor, Professor Yasien Sayed, and co-supervisor, Dr Ikechukwu Achilonu, for the continuous support, patience, motivation, and excellent supervision and mentorship during my studies.

Thank you to Adriaan Basson for his training and assistance in performing the phenotypic assays.

Thank you to all members of the Protein Structure-Function Research Unit. The discussions were insightful and the help invaluable. A special thank you to Jake Zondagh, who walked this road with me.

I am also grateful to the University of the Witwatersrand, and the South African National Research Foundation for their generous financial support that has enabled my studies.

Thank you to my friends and family who supported me through this journey. A special thank you to Mark Ilsley for always being generous with your support, hugs and Sunday afternoons.

Importantly, thank you to my parents (Mike and Elsabé Williams) who always support and motivate me and have always encouraged me never stop learning and asking questions.

Contents

Note on Referencing

References for chapters 1 and 5 appear at the end of the thesis.

References for chapters 2, 3 and 4 appear at the end of each chapter.

List of Figures

List of Tables

List of Abbreviations

AIDS:	Acquired Immunodeficiency Syndrome
APV:	Amprenavir
ARV:	Antiretroviral
ATV:	Atazanavir
CA:	Capsid protein
CRF:	Circulating Recombinant Form
DRV:	Darunavir
FDA:	Food and Drug Administration
HAART:	Highly Active Antiretroviral Therapy
HIV:	Human Immunodeficiency Virus
IC_{50}:	Concentration at which 50% of enzyme inhibited
k_{cat}:	Turnover number
k_{cat}/K_M:	Catalytic efficiency
K_i:	Inhibitory constant
K_M:	Michaelis-Menten constant
L38↑N↑L:	Insertion of Asn and Leu at position 38 indicated by the upward arrows. The arrow indicates an insertion event.
LPV:	Lopinavir
MA:	Matrix protein
NC:	Nucleocapsid protein
NNRTI:	Non-nucleoside Reverse Transcriptase Inhibitor
NRTI:	Nucleoside Reverse Transcriptase Inhibitors
PI:	Protease Inhibitor
RTV:	Ritonavir
Trx:	Thioredoxin
V_{max}:	Maximum velocity

CHAPTER 1 INTRODUCTION

1.1. HIV and AIDS

Human immunodeficiency virus (HIV) causes acquired immunodeficiency syndrome (AIDS), which is the extreme suppression of the immune system. Globally, in 2016, there were 36.7 million people living with HIV [1]. Worldwide, in 2016, there were 1.8 million new infections and 1.0 million people died from AIDS-related causes [1]. Since HIV was first identified in 1983 [2], 76.1 million people have become infected and 35 million people have died from AIDS-related illnesses. HIV is a particular problem in Africa because 69% of people living with HIV come from Africa [1]. Eastern and southern Africa account for 43% of the global total of new HIV infections. Sub-Saharan Africa is the epicentre of this epidemic as 51% of people infected with HIV reside here. It was estimated that 13% of the South African population in 2016 was living with HIV. However, there have been great improvements in treatment in Africa, particularly South Africa. In 2015 South Africa had 3.4 million people on treatment, more than any other country [3]. There are, however, still populations within South Africa that are at high risk of HIV infection; these include sex workers, people who inject drugs, transgender people, prisoners, gay men and men who have sex with men [3]. In 2015, these populations accounted for 20% of all new infections in sub-Saharan Africa [3] and it was seen that the prevalence of HIV among sex workers in Johannesburg was 71.8% [4].

Human Immunodeficiency Virus (HIV), a highly mutable lentivirus, is the causative agent for Acquired Immunodeficiency Syndrome (AIDS). HIV is a member of the lentivirus genome and belongs to the *Retroviridae* family. HIV-1 is enveloped by a lipid bilayer that is derived from the membrane of the host cell [5]. The glycoprotein, gp120, is found on the surface of the membrane and is anchored in the membrane by gp41 [5]. The membrane is lined with matrix proteins (MA) [5]. At the centre of the virus is a cone-shaped capsid core [5]. The conical capsid consists of capsid protein (CA) monomers assembled predominately into hexamers as well as a few pentamers which facilitate the curvature at the top and the bottom, closing the capsid [6–8]. Within the capsid core are two copies of unspliced viral RNA which is stabilised by the nucleocapsid protein (NC) [5]. The genome of

this virus consists of *gag/pol/env* genes which is typical of other retroviruses [9]. Also contained within the capsid core are the virally-encoded enzymes; reverse transcriptase, integrase and protease. The virus also packages the accessory proteins Nef, Vif and Vpr [5].

1.2. Subtypes

HIV is genetically diverse and is one of the most variable human pathogens [10]. Within a patient, HIV exists as many highly related but non-identical genomes called quasispecies [10]. The genetic diversity of HIV is attributed to three factors. Firstly, viral replication is very rapid - an estimated 10^{10} virions are produced a day [11]. Secondly, reverse transcriptase is error-prone and introduces on average $3.4x10^{5}$ mutations per base pair per replication cycle [10, 12]. Thirdly, recombination between the two RNA molecules packaged into the virion occurs which generates a mosaic DNA genome occurring at a frequency of 7 to 30 per replication round [13]. This recombination is a major force in viral evolution both within a patient as well as globally [10].

HIV is made up of two types: HIV-1 and HIV-2 [14]. HIV-2 infections are very rare and are limited to West Africa as it is believed to be less transmissible and pathogenic than HIV-1 [10]. HIV-1 is further divided into groups M, N, O and P [14]. Group M is the main group and groups O and N are very rare and limited to Cameroon in Central Africa [15, 16]. Group M is divided into nine subtypes, A to H, J, K and L and many circulating recombinant forms (CRFs) [14]. CRFs are as a result of recombination between subtypes within a dually infected individual that is then transmitted to other people. A CRF is only classified if three or more people with no direct epidemiological linkage are infected with it [12]. Genetic variation between subtypes can be from 25 to 35% and within a subtype 15 to 20% [17]. The different subtypes occur in different regions of the world, for example subtype A is found in East Africa, Central Asia and Eastern Europe [10]. Subtype B, the most studied subtype, is found in America, Western Europe and Australia [14]. Subtype C is found predominately in southern Africa, the horn of Africa and India [10] and thus it is the subtype that infects most of the people living with HIV.

1.3. Replication cycle

The HIV replication cycle is divided into an early and a late phase. The early phase begins with the recognition of the target cell and ends with the integration of the viral genome. The late phase

begins with the expression of the proviral genome and ends with maturation. The replication cycle is shown in Figure 1.

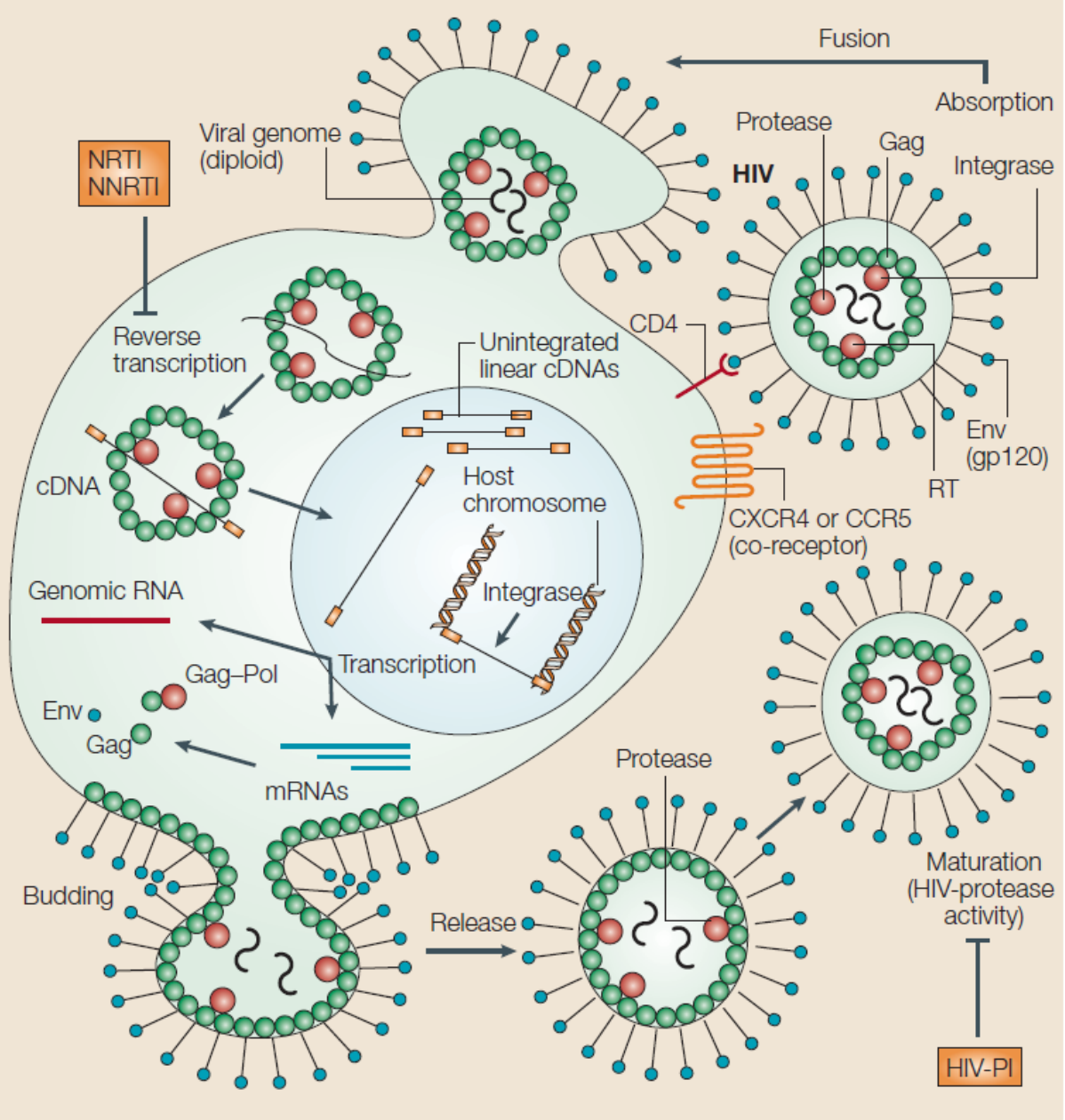


Figure 1: Replication cycle of HIV. The major targets for drug therapy are represented by the orange boxes. Nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs) target the reverse transcriptase enzyme. HIV-1 protease inhibitors (HIV-PIs) target the HIV-1 protease enzyme. Integrase inhibitors are newer targets for drug therapy. Image taken from Monini *et al*., (2004) [18].

1.3.1. Early Phase

HIV-1 binds to the cells bearing the CD4 receptor via an interaction between gp120 and the amino-terminal immunoglobulin domain of CD4 [5]. This interaction does not cause fusion; rather a secondary interaction with the chemokine receptors CXCRC or CCR5 is required [19]. Fusion releases the viral core into the cytoplasm and is followed by an uncoating event [5]. The RNA viral genome is then reverse transcribed to double-stranded DNA by reverse transcriptase. The viral cDNA genome forms part of a pre-integration complex consisting of integrase, matrix protein, reverse transcriptase and Vpr [20]. This complex interacts with various cellular transportins and nucleoporins to drive import into the nucleus [21]. Vpr directs nuclear localisation [22, 23] by connecting the complex to cellular nuclear import machinery [22, 24, 25]. Integrase and the cellular cofactor: lens epithelium-derived growth factor, then catalyse the integration of the viral DNA into the host genome [5, 26].

1.3.2. Late Phase

The host cell machinery transcribes the integrated DNA (the provirus). This includes the full length and various spliced mRNAs encoding viral proteins. Spliced and unspliced mRNA is transported out of the nucleus for translation [5]. The export of unspliced mRNA (encoding Gag and Gag-Pol polyproteins) and singly-spliced mRNA (encoding Env, Vpu, Vif and Vpr) is mediated by the accessory protein Rev because unspliced cellular RNA molecules are ordinarily retained in the nucleus [5]. The highly spliced mRNA that contains Tat, Rev and Nef are able to leave the nucleus. The Env precursor gp160 is synthesised in the endoplasmic reticulum, the same location that CD4 molecules are synthesised [5]. To prevent premature binding of Env and CD4, the accessory protein Vpu binds to CD4 and signals their degradation [5]. The surface CD4 are degraded by endosomal degradation which is signalled by Nef [5, 26]. Env is glycosylated in the endoplasmic reticulum and Golgi apparatus and cleaved into gp120 and gp40. It is then transported to the plasma membrane for viral assembly.

The Gag polyprotein is synthesised in ribosomes from unspliced mRNA. Gag-Pol is produced by a rare -1 ribosomal frameshift that takes place 5-10% of the time [27]. A “slippery” sequence near the 5’-end of the p6 coding region is responsible for the frameshift [28]. This frameshift results in the translation of the coding regions for protease, reverse transcriptase and integrase. Gag and Gag-Pol become localised to the cell membrane directed by the matrix protein [5]. RNA binds to Gag at the plasma membrane and promotes the multimerisation and assembly of the immature Gag lattice [29]. Approximately 2400 copies of Gag [30] and 120-240 copies of Gag-Pol [27] bud to form an

immature particle, which encapsidates two RNA copies of the unspliced viral RNA genome [5]. An immature virion buds from the plasma membrane and maturation of the virus is triggered by protease-mediated cleavage of Gag-Pol and Gag polyproteins. A dimerisation event of the Gag-Pol polyprotein activates the protease [31]. This dimer is extremely unstable [32] and exhibits much lower enzymatic activity than the free enzyme [33, 34]. This embedded dimer cleaves intramolecularly at first until it is free and then is able to cleave the Gag polyprotein intermolecularly [34–36]. Once cleaved, the structural proteins (matrix and capsid) rearrange to form an infectious virus particle [5]. The maturation event is critical for the infectivity and fusogenicity of the virus. Should maturation not occur the virus will not be able to infect cells and also the fusogenicity of Env glycoproteins will be reduced [37, 38].

1.4. Gag and Gag-Pol polyproteins

HIV's genome is divided into three open reading frames (ORFs); *gag*, *pol* and *env*. This is typical of other retroviruses [39]. *Pol* encodes the three essential retroviral enzymes: integrase, protease and reverse transcriptase. The *env* gene encodes the surface envelope proteins. Gag (group specific antigen) is the major structural protein of HIV and comprises approximately 50% of the mass of the viral particle [28]. *Gag* encodes the structural proteins that form the virus capsid, nucleocapsid and matrix. Gag is translated from the 9 kb fully unspliced mRNA which also encodes the Pol polyprotein. Gag is structurally divided into four domains; beginning at the N-terminus, matrix, capsid, nucleocapsid and p6 at the C-terminus (Figure 2). Flanking the nucleocapsid are two smaller "spacer" regions: p2 (N-terminus) and p1 (C- terminius) [28].

The order of cleavage of the Gag polyprotein is essential for proper assembly of the virus. The amino acids in the cleavage sites differ which explains the different cleavage rates; however, these cleavage sites are similar in 3-dimensional structure [40, 41]. The order of cleavage of the Gag polyprotein begins with the separation of the nucleocapsid from the capsid, which on the C-terminal end of the spacer peptide p2 [42]. The capsid protein is then separated from the matrix protein which remains associated with the virion membrane [42]. Almost simultaneously the C-terminal sequence of p6 is released, this is on the C-terminal end of the linker p1 between the nucleocapsid and p6 [42]. p1 and p2 are then trimmed from the nucleocapsid and the capsid, respectively [42]. p2 is important for maturation of the Gag and regulation of ordered cleavage [43, 44] and p1 is important for incorporation of Gag and Pol into the virus [28]. The natural substrates of the protease have a

variable and weaker interaction with the catalytic site compared to HIV-1 protease inhibitors [45]. This

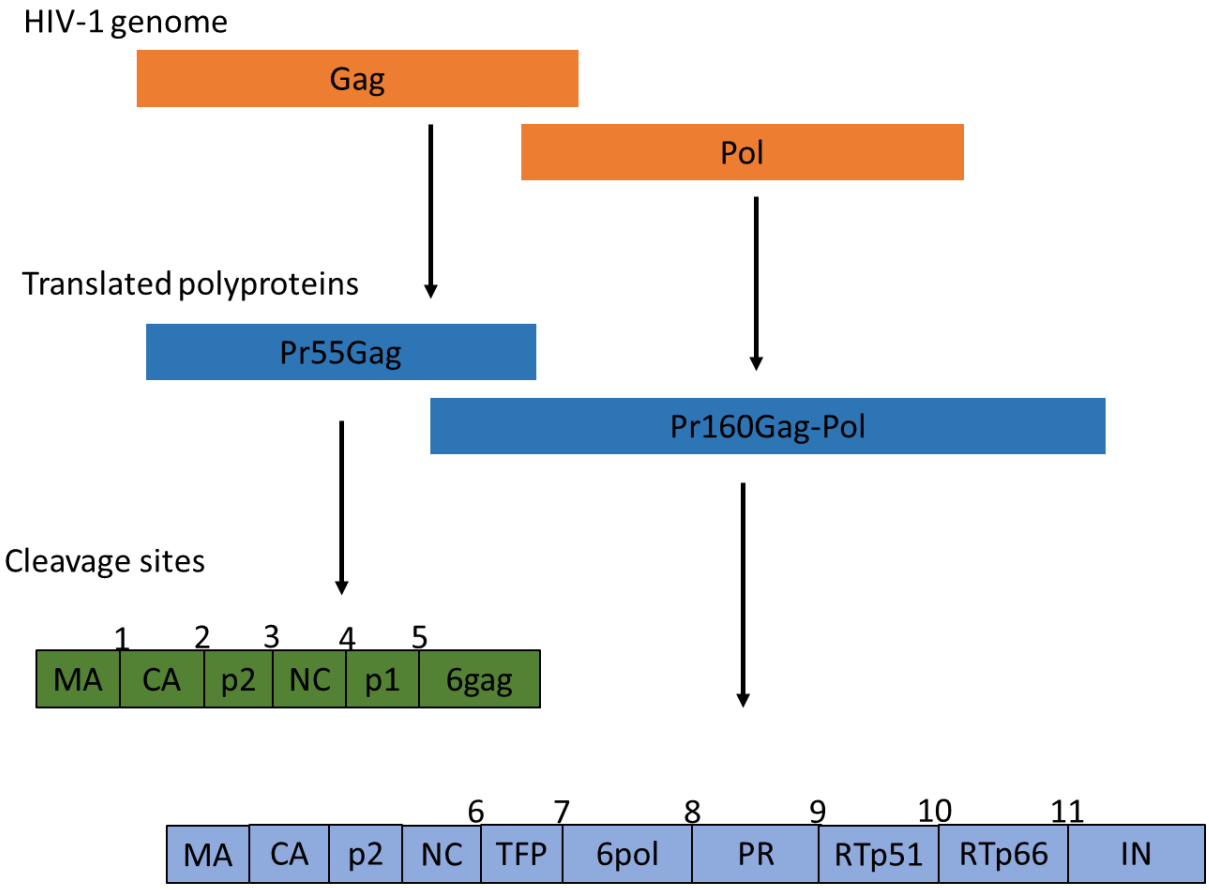


Figure 2: Gag and Gag-Pol polyproteins. The order of cleavage of Gag and Gag-Pol is indicated by the numbered cleavage sites

phenomenon causes the ordered sequential cleavage of the polyproteins [45]. Resistance mutations that cause the enlargement of the active site would, therefore, have a greater effect on the protease inhibitors than on the cleavage of the Gag polyprotein [45].

1.5. HIV protease

It was postulated by Ratner *et al*. (1985) that the protease present in the HIV virion was an aspartyl protease due to the presence of the characteristic active site amino acid triplet; Asp-Thr-Gly. This was confirmed by pepstatin inhibition of the protease [47–49], which is characteristic of enzymes in this class [50]. Further confirmation was the loss of enzymatic activity with the deletion of the catalytic residue Asp25 [49]. The first crystal structure of HIV-1 protease was solved in 1989 by Navia *et al*. [51]. This was followed by a crystal structure showing the side chain locations published by Wlodawer *et al*., (1989) [52].

1.5.1. Structure

HIV-1 protease is a 22 kDa symmetrical homodimer with 99 amino acids per monomer. In order to function, HIV-1 protease must be in the dimeric form as the monomeric form is inactive [53]. This obligation to dimerise is an important regulatory function for proteolytic cleavage of the polyproteins. The subunits are noncovalently associated and the active site is formed at the dimer interface [52].The secondary structure consists mainly of β-sheets and contains only one α-helix per monomer. The dimer interface consists of four short anti-parallel β-strands, unlike other aspartyl proteases, which consist of six β-strands. It is formed by the N- and C-termini from each monomer consisting of residues 1-4 and 96-99, respectively (Figure 3).

The active site is formed at the dimer interface and one Asp residue from each monomer (Asp25) is contributed. The active site loop is relatively rigid due to a network of hydrogen bonds called the fireman's grip [54]. The active site consists of residues 23-30 of each monomer. Two β-turns known as the flap region (residues 43-58) cover the active site. These flaps are glycine rich and undergo major structural changes. The flaps exist in the open, semi-open and closed conformations depending on whether substrate or inhibitor is bound. The flap region plays a role in substrate binding because the flaps open to allow substrate to bind to the active site. The movement of the flaps is aided by the hinge region, made up of residues 35-42. The movement of the flaps can occur due to compensatory motions in residues 59-75, which act as a cantilever, the residues 11-22 act as the fulcrum to the

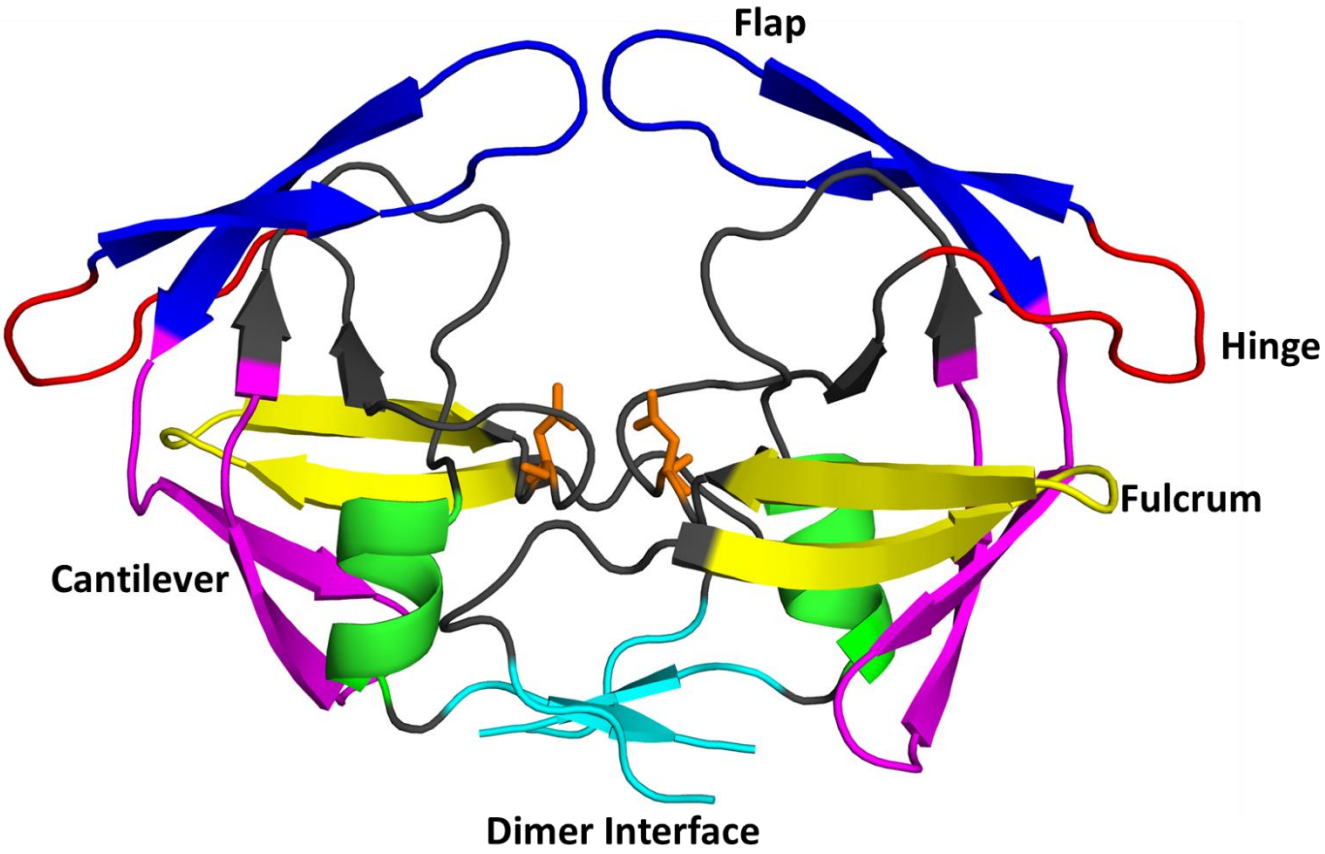


Figure 3: Structure of wild-type subtype C protease (PDB ID: 3U71). HIV-1 protease is a homodimer and contains mainly β-sheets and one α-helix per monomer. There are five regions defined within the structure: the flap region (blue), the hinge region (red), the fulcrum region (yellow), the cantilever region (magenta) and the dimer interface (cyan). The catalytic Asp25 residues are shown in the active site (orange). PyMOL was used to generate the figure (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

cantilever [55]. During opening of the flaps the fulcrum region and the cantilever region assist by moving downwards [56].

1.5.2. Catalytic Mechanism of HIV-1 protease

Proteases use one of two mechanisms to cleave peptides. This first makes use of an activated water molecule, which acts a nucleophile and attacks the amide bond carbonyl carbon of the substrate's scissile bond. The water molecule can either be activated by a Zinc atom (Zinc metallo-proteinases) or by two aspartyl β-carboxylate groups at the active site (aspartyl proteases such as HIV-1 protease). The second mechanism makes use of a nucleophilic atom of an amino acid side chain to hydrolyse the amide bond.

The bond being hydrolysed on the substrate is called the scissile bond or cleavage site and is positioned between the P1 and P1' site [57], Figure 4A. The flanking amino acids on the substrate, towards the N-terminus, are termed P1, P2, and P3 and towards the C-terminus the amino acids are designated P1', P2' and P3' [58]. The protein subsites (designated S) that interact non-covalently with the corresponding side chains of the substrate are termed S1, S2, S3 and S1', S2', S3' respectively. The S1 subsite (Arg8, Leu23, Asp25, Gly27, Gly48, Gly49, Ile50, Thr80, Thr81 and Val82) is highly hydrophobic with the exception of the active site aspartate residues [58]. The S2 subsite (Val23, Ala28, Asp29, Asp30, Ile47, Gly49, Ile50, Leu76, Ile84) is mostly hydrophobic with the exception of Asp29 and Asp30 [58]. This site is smaller than S1 and S3 and as a result is more specific as the size and type of residues at P2 is restricted. The S3 subsite is adjacent to S1 and is mostly hydrophobic. S3 has a broad specificity and will accept residues of different types and sizes [58]. Not much information is available about the S4 and S5 subsites due to the limited number available structures of retroviral proteases in complex with ligands that extend beyond P3.

In the absence of substrate binding the active site in filled with water molecules. The catalytic aspartate residues (Asp25 and Asp25') interact with a water molecule. These residues perform a general acid-base role and activate the water molecule to act as a nucleophile. When a substrate binds to the HIV-1 protease, it is converted to a gem-diol intermediate through nucleophilic attack by the activated water molecule. The covalent linkages are broken and two products are formed, which are released sequentially. Near atomic-resolution crystal structures have shown these vital steps in the reaction mechanism of HIV-1 protease [59] and confirm that this mechanism proceeds through a gem-diol intermediate [60]. The mechanism is represented in Figure 4B.

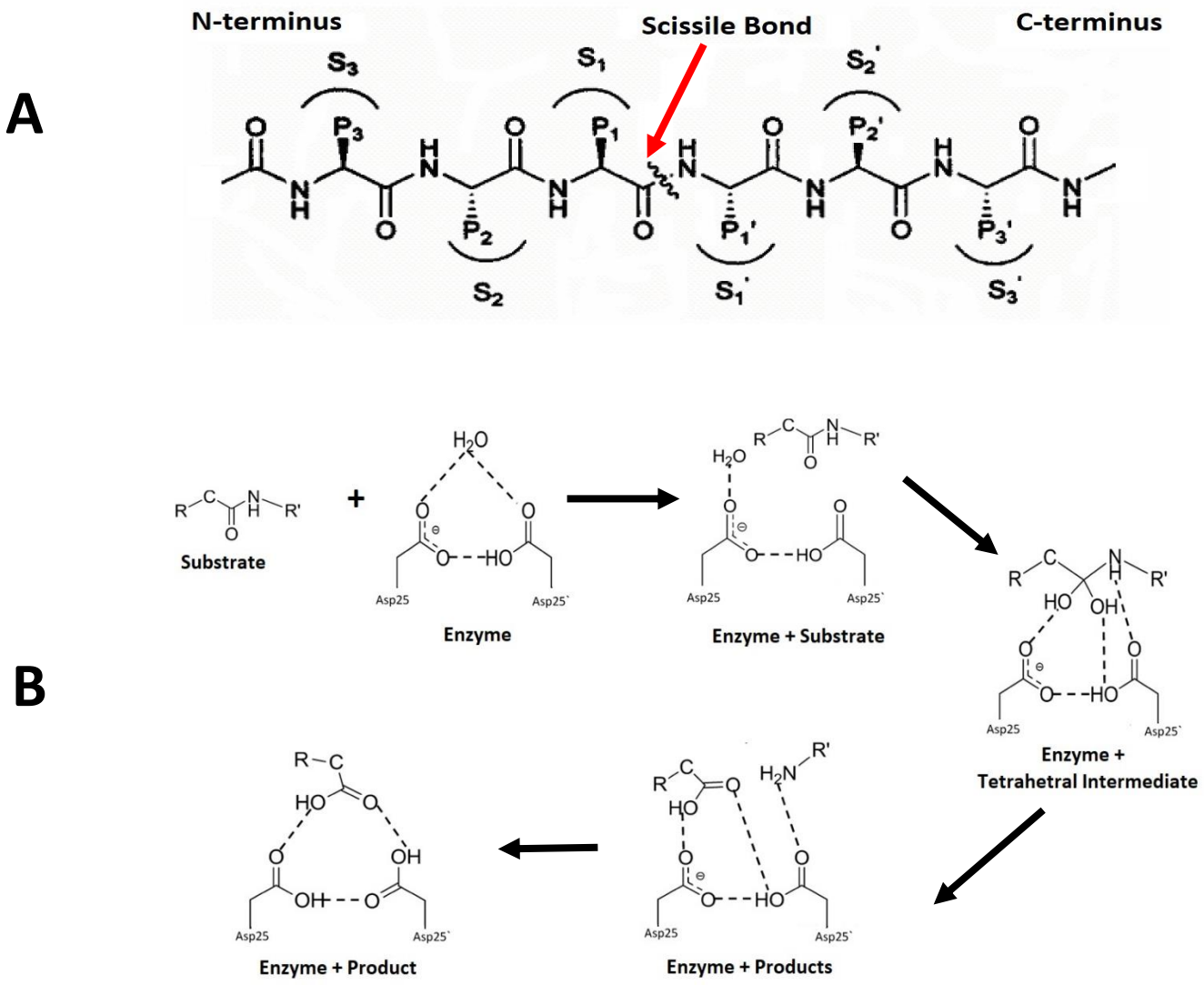


Figure 4: Active site nomenclature and reaction mechanism of HIV-1 protease. The nomenclature used for binding sites in the HIV-1 protease and residues on the substrate is shown in A. This was adapted from Wlodawer and Vrondasek, 1998 [58]. The reaction mechanism is shown in B, adapted from Shen *et al*., (2012) [59].

1.6. Antiretroviral Therapy

In the absence of treatment, a patient can harbour the HI-virus for a decade or more without showing symptoms. As the CD4 T-cells are depleted to below 200 cells/cm^3 opportunistic infection occurs [61]. Disease prevention through behavioural changes has played a significant part in controlling the spread of the pandemic as well as the international effort to develop therapies to prevent viral replication and restore immune function in patients [62, 63]. The resulting drugs have significantly improved life expectancy for people with HIV-1 and reduced the capacity to transmit the virus [62].

Antiretrovirals (ARVs) are divided into six different classes depending on the target: Nucleoside reverse transcriptase inhibitors (NRTIs), Non-nucleoside reverse transcriptase inhibitors (NNRTIs), HIV-1 protease inhibitors (PIs), Integrase inhibitors, Fusion inhibitors and Chemokine receptor antagonists. Currently, the recommended AIDS therapy uses a mixture of drugs from the different classes in highly active antiretroviral therapy (HAART). There are ten FDA approved HIV-1 protease inhibitors and the first one was introduced into clinical practice in 1995 [64]. The inclusion of HIV-1 protease inhibitors in therapy has resulted in prolonged viral suppression, control, reduced morbidity and mortality for HIV infected people [64].

1.6.1. HIV-1 protease inhibitors

HIV-1 protease was selected as a drug target due to its importance in the replication cycle of the virus. If the HIV-1 protease activity is altered it leads to defective viral particles and reduced infectivity [65, 66]. Inactivating the HIV-1 protease leads to virus particles that are not infective [67]. HIV-1 protease inhibitors were based on structure-based inhibitor design. The first generation of HIV-1 protease inhibitors were peptidomimetic molecules with non-cleavable isosteres presented to the active site [68]. These inhibitors were developed based on knowledge of renin and pepsin aspartyl proteases. First generation inhibitors mimic the transition state of the natural substrate and thus bind tightly to the enzyme [69]. First generation HIV-1 protease inhibitors were designed with polar groups that resemble those of the natural substrate peptide main chain and have a central hydroxyl that interacts with the catalytic aspartates and mimics the hydroxyl of a tetrahedral reaction intermediate [64]. These contain hydroxyethelene or hydroxyethylamine isosteres [68]. Saquinavir was the first drug to be approved and indinavir and ritonavir (RTV) soon followed. Other first generation inhibitors include nelfinavir and amprenavir (APV). First generation drugs suffered

short-comings due to their peptide-like structural features which resulted in poor bioavailability, low metabolic stability, heavy pill burden and debilitating side effects [70].

Once drug resistance towards HIV-1 protease inhibitors emerged, the drug design strategy changed to targeting drug resistant variants [64]. The aim was to increase the drug resistance barrier and cross-resistance profiles of inhibitors, improve bioavailability and reduce drug toxicity. Second generation drugs were designed with less peptide backbone features but retained the central hydroxyl group [71, 72]. The peptide-like carbonyl in the earlier HIV-1 protease inhibitors is replaced by sulfonamide [64]. Second generation drugs include lopinavir, atazanavir, tipranavir and darunavir. HAART therapies currently recommend the use of three HIV-1 protease inhibitors lopinavir, atazanvir and darunavir and thus these were focused on in this study (Figure 5).

1.6.2. Lopinavir

Lopinavir (LPV) is a second generation drug developed by Abbott [73]. LPV is the most widely used in drug naïve patients and is administered with the booster RTV [73]. RTV is used as a booster because it enhances pharmacokinetic properties of LPV [74, 75] by binding and inhibiting the CYP34A enzyme that is responsible for metabolism of certain PIs. The design of LPV was based on RTV but eliminates the P3 isopropylthiazolyl group of RTV [73]. This was done to remove the interaction with the V82 residue, which is mutated to alanine in drug resistant HIV-1 proteases. Furthermore, the thiazolylmethoxycarbonyl moiety was replaced with the dimethylphenoxyacetyl group [73]. The P1-P1' positions are still occupied by the same hydroxyethylene peptidomimetic as RTV [73]. LPV/r is the first choice PI for an antiretroviral regimen because it has a high genetic barrier for resistance and exhibits long-term efficacy.

1.6.3. Atazanvir

Atazanavir (ATV) was developed by Bristol-Myers Squibb [76] and approved for use in 2003. It was approved as an alternative option for initial PI-based HAART. ATV was developed to have a higher genetic barrier for resistance and greater bioavailability than first generation drugs. ATV is a 2-hydroxy-1,3-diaminopropane transition-state isostere with an aza-dipeptide core [77]. ATV has a high potency against wild-type protease with an inhibitory constant (K_i) of 10 pm [78]. The good bioavailability of ATV is due to the pyridylbenzyl moiety [76]. One advantage of ATV use for patients is it's once-a-day dosing.

1.6.4. Darunavir

Darunavir (DRV) is the most recent PI to be approved for use in HAART. It was developed by Tibotec Inc. (now Johnson and Johnson) [79] and was approved for use in 2006. DRV design resulted from

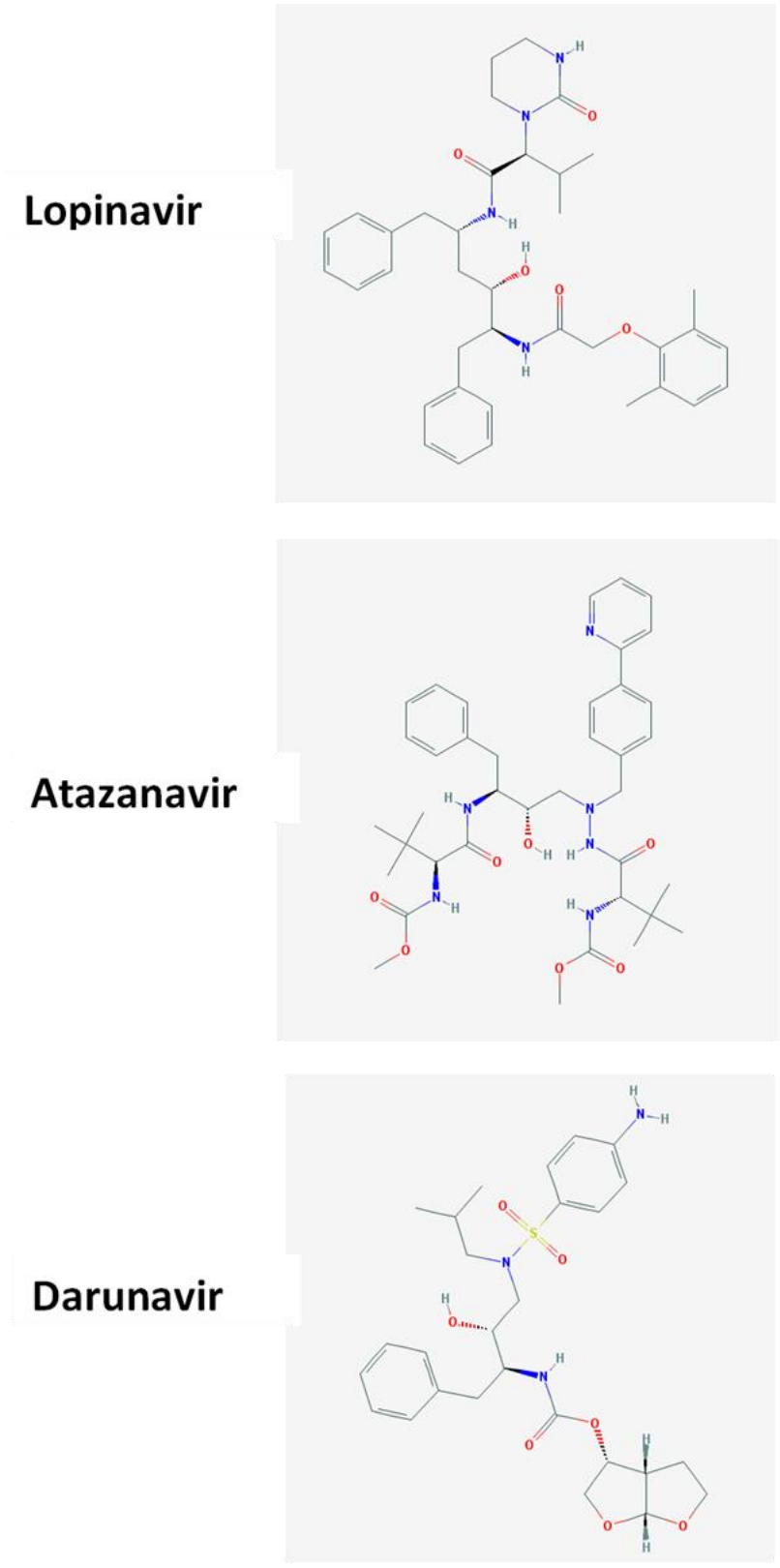


Figure 5: Chemical structures of lopinavir atazanavir and darunavir. Images obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov)

structure-based design approach, which focused on maximising interactions of the inhibitor with most conserved domains of the HIV-1 protease i.e. the backbone [80, 81]. DRV is chemically related to the first generation drug APV. The sulfonamide isostere of APV was replaced with chiral bicyclic *bis*-tetrahydrofuran (*bis*-THF). The *bis*-THF moiety was incorporated to introduce additional polar interactions with main-chain atoms on the HIV-1 protease dimer. DRV displays a high affinity for HIV-1 protease and a high K_i (16 pM) [79]. DRV was designed to inhibit drug resistant strains and is indeed a broad-spectrum potent inhibitor against isolates containing multi-drug resistant strains [72]. *In vitro* selection of resistance appears to be slower and less frequent than other HIV-1 protease inhibitors and HIV-1 protease inhibitor naïve viruses can escape drug pressure via an alternative Gag substrate-based resistance [82].

1.7. Resistance

During the asymptomatic phase in an individual infected with HIV, the viral loads in the plasma are 10^4-10^5 copies/ml [83]. The viral half-life of free and cell-associated viruses are short meaning that there is a rapid turnover of the viral population. It has been estimated that ~10^9 new virions are synthesised each day [84]. Reverse transcription has a nucleotide mis-incorporation rate of ~1:10 000 with no proof reading capacity [83]. The genome size (~10 kb) means that the mis-incorporation rate allows one point mutation each time the HIV genome is transcribed [85] and combined with high viral turnover results in a high mutation rate. This leads to the existence of viral quasispecies. The most common of which would be the wild-type virus as it would be the most fit [86]. Under normal conditions, a quasispecies would be less fit than wild-type but under drug pressure the genetic flexibility allows the population to respond to the different selection pressure [82]. The population of infectious progeny harbouring mutations is, however, relatively small as the majority of mutations would produce non-infectious virions [87–89].

Mutations that occur are not a direct consequence of drug action but are due to viral replication. Antiviral drug resistance is selected through elimination of most drug susceptible viruses from the replicating pool leaving only the fittest viruses to survive in its presence [90]. In the absence of ARVs HIV shows genetic diversity, especially the HIV-1 protease gene [91]. There are extensive mutations that confer cross-resistance to HIV-1 protease inhibitors [92]. More mutations are selected by HIV-1 protease inhibitors than by any other class of drug [64]. This is a particular problem because the drug resistance to a particular HIV-1 protease inhibitor may lead to cross resistance to other HIV-1 protease inhibitors [64]. The degree of cross resistance depends on the mutations selected and the number of mutations [93].

1.7.1. Primary Mutations

Mutations that affect amino acids directly involved in substrate binding are termed primary mutations. These mutations have an effect on drug susceptibility [84, 94]. Primary mutations occur in the binding site and do not involve residues involved in catalysis (e.g. Asp25) but do involve residues that have direct interactions with the HIV-1 protease inhibitors. These mutations cause conformational changes in the binding site [95], which affect the drug/target interface by loss of interactions or steric effects created by altered geometries [96, 97]. Primary mutations cause an enlargement of the catalytic active site which decreases binding to an inhibitor and, in parallel, the natural substrate which could decrease viral replication [98–100]. Primary mutations result in several fold (two to five) decrease in susceptibility to one or more HIV-1 protease inhibitors [101–103] and generally initiate resistance to HIV-1 protease inhibitors [42]. Primary mutations are selected for early in the process of resistance mutation accumulation and tend to be specific for each compound [83]. The advantage of primary mutations is that they are able to confer drug resistance but are disadvantageous in that they have been known to impact enzyme function and viral fitness [42]. Primary mutations decrease enzyme activity [101] and impair viral replication *in vitro* [98, 103–105]. Primary mutations are not natural polymorphisms as they would not be advantageous to the virus in the absence of drug pressure.

1.7.2. Secondary Mutations

Mutations that occur distal to the substrate binding site are known as secondary mutations [95]. These mutations are natural polymorphisms and do not cause resistance individually [95, 103]. Secondary mutations occur after primary mutations and are compensatory [106]. These mutations individually have little effect on drug susceptibility and are generally accumulated to compensate for reduced viral fitness [83]. Secondary mutations help restore original viral fitness by increasing enzyme activity [42, 101, 102]. Mutations, in subtype B, such as L63P and N88D restore HIV-1 protease activity, which is often diminished due to primary mutations [107]. However, secondary mutations are not able to fully restore the viral fitness as most high-level resistant viruses display various degrees of fitness loss [99, 100, 105, 108]. Secondary mutations are important for drug resistance because when they are accumulated they can cause a stepwise reduction in susceptibility [103]. The accumulation of secondary mutations (10-20) can be responsible for the loss of inhibitor potency, giving these mutations an additional role to compensation for loss of activity [106, 109]. In combination with primary mutations, secondary mutations are critical for high levels of resistance.

1.7.3. Insertion mutations

Amino acid insertions in HIV-1 protease have been identified in patient-derived HIV strains near residues 18, 25, 36, 70 and 95 [110–112]. Most insertions appear to be due to duplications of the neighbouring genetic sequence [113]. This is caused by reverse transcriptase stalling and slippage [114, 115]. HIV-1 protease insertions are very rare, tenfold less common than insertions in reverse transcriptase [116]. The frequency of inserts between the years 1999 and 2001 was 0.1% [110, 111]. Insertions in HIV-1 protease are primarily located at externally exposed loops and turns, which would be more likely to accommodate the extra residues by extending the amino acids outwards from the molecule [113]. This also prevents extreme modification of the enzyme which would decrease or destroy enzyme function [113]. Recently, the frequency of insertions has increased, especially between residues 32 and 42, the hinge region [117]. Insertions have been associated with HIV-1 protease inhibitor resistance when in combination with other well-described HIV-1 protease inhibitor resistance mutations by imposing minor structural changes to the flap and substrate binding cleft [118]. Insertions not only occur as a result of HIV-1 protease inhibitor therapy, they can occur even in treatment naïve individuals [111, 116]. In the case of treatment-naïve individuals the presence of the insertion can persist for a long time, implying a selective advantage for the virus, and can be transmitted [116].

1.8. L38↑N↑L variant

The variant used in this study was found in an infant in Johannesburg. The mother was part of the Prevention of Mother to Child Transmission programme [119]. The mother was HIV-1 protease inhibitor naïve but was exposed to reverse transcriptase inhibitors. The variant contained the following subset of mutations: K20R, E35D, R57K and V82I (Figure 6). The only active site mutation is the V82I. The prevalence of the mutations is shown in Table 1. The variant also contains a double insertion of Asn and Leu at position 38, which is in the hinge region of the protein. The prevalence of these specific insertion mutations is not known.

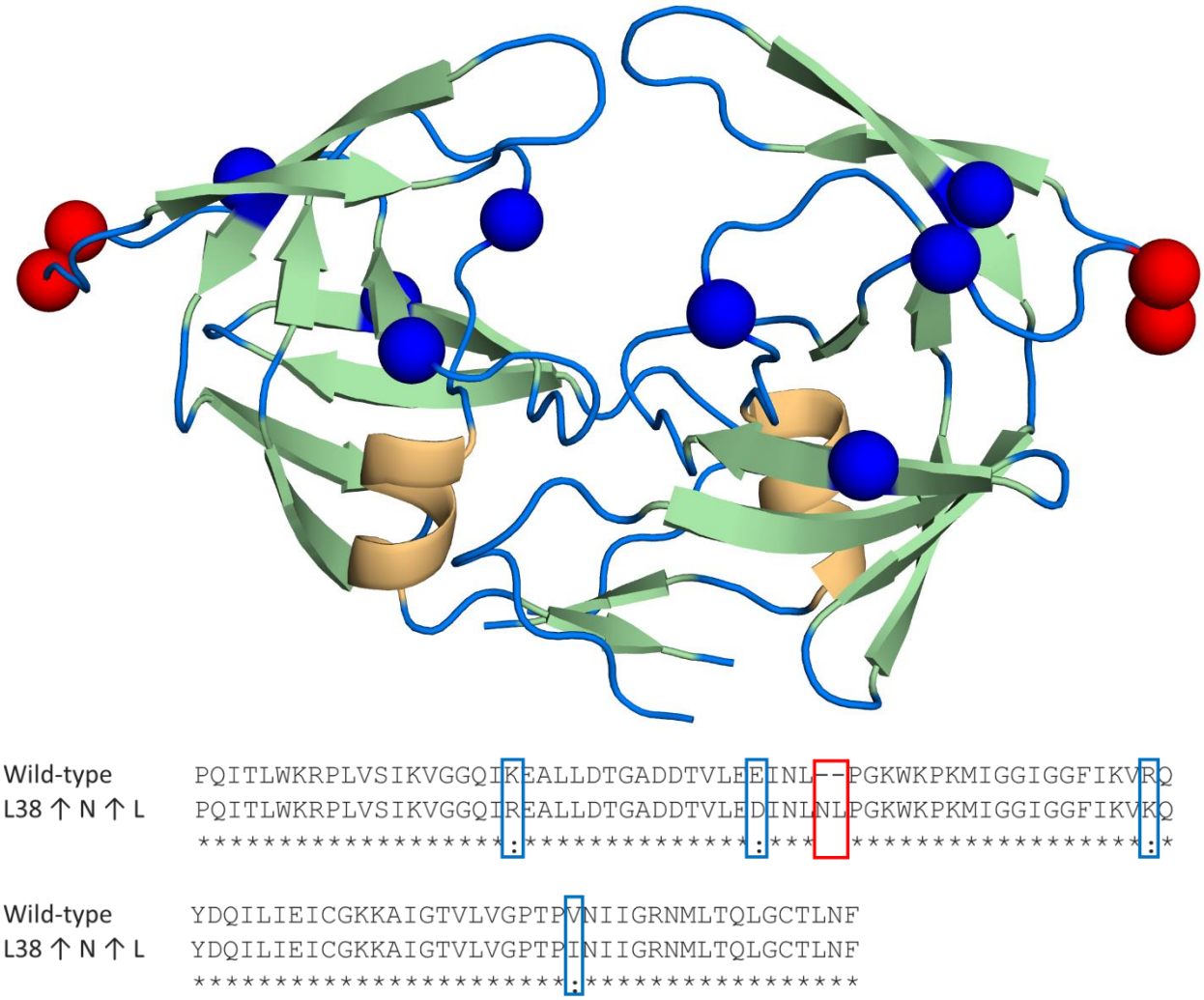


Figure 6: Homology model of L38↑N↑L protease and sequence data. HIV-1 protease contains mainly β-sheets (green) and one α-helix per monomer (gold). The blue spheres on the structure and the blue boxes on the sequence represent the relative positions of the subset of mutations, K20R, E35D, R57K and V82I. The red spheres on the structure and the red box in the sequence alignment represent the double insertion of Leu and Asn. The homology model was generated using SWISS Model using data from the Protein Data Bank (PDB ID: 3U71). The figure was created using PyMol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC). The sequence alignment was generated using the Clustal Omega tool (EMBL-EBI).

Table 1: Frequency of mutations within the L38↑N↑L protease as obtained from the Stanford University HIV Drug Resistance Database (https://hivdb.stanford.edu).

	Frequency of mutation (%)	
Mutation	**Protease inhibitor naïve**	**Protease inhibitor treated**
K20R	20	27
E35D	25	29
R57K	4.1	6.4
V82I	6.7	6.7

1.9. Aim and objectives

1.9.1. Aim

Determine the impact of the L38↑N↑L insertion mutations as well as the subset of mutations (K20R, E35D, R57K and V82I) on the structure and function of HIV-1 protease subtype C.

1.9.2. Objectives

1. Express and purify both wild-type and L38↑N↑L protease using a Trx-6His tag
2. Perform molecular dynamic simulations of wild-type and L38↑N↑L protease
3. Determine the steady-state kinetic values (V_{max}, K_M, k_{cat}, k_{cat}/K_M) of L38↑N↑L
4. Determine the IC_{50} values of both enzymes with LPV, ATV, DRV
5. Perform phenotypic assays in the presence of protease inhibitors

CHAPTER 2 OVEREXPRESSION AND PURIFICATION USING TRX-HIS TAG

Overexpression, purification and functional characterisation of HIV-1 subtype C protease and two variants using a novel thioredoxin and His-tag protein fusion system

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

[1]Both authors contributed equally to this work.

Protein Journal (accepted)

In this publication the overexpression and purification of the wild-type and two variant HIV-1 proteases (N37T↑V and L38↑N↑L) using a thioredoxin-hexahistidine fusion system is described. The thioredoxin moiety coupled with a hexahistidine tag successfully improved the overexpression of all three proteases.

Author contributions: Jake Zondagh performed experimental work on wild-type and N37T↑V proteases and prepared the manuscript. Alison Williams performed experimental work on wild-type and L38↑N↑L proteases and prepared the manuscript. Ikechukwu Achilonu assisted in experimental design and manuscript revision. Heini W. Dirr assisted in manuscript revision. Yasien Sayed supervised the project and assisted in data analysis and interpretation

Overexpression, purification and functional characterisation of HIV-1 subtype C protease and two variants using a novel thioredoxin and His-tag protein fusion system

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

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

[1]Both authors contributed equally to this work

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

Acknowledgements

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

Abstract

In recent years, various strategies have been used to overexpress and purify HIV-1 protease because it is an essential drug target in anti-retroviral therapy. Obtaining sufficient quantities of the enzyme, however, remains challenging. Overexpression of large quantities is prevented due to the enzyme's autolytic nature and its inherent cytotoxicity in *Escherichia coli* cells. Here, we describe a novel HIV-1 protease purification method using a thioredoxin-hexahistidine fusion system for the wild-type and two variant proteases. The fusion proteases were overexpressed in *Escherichia coli* and recovered by immobilised metal ion affinity chromatography. The proteases were cleaved from the fusion constructs using thrombin. When compared to the standard overexpression and purification protocol in use in our laboratory, the expression of the fusion-derived wild-type protease was increased from 0.83 to 2.5 mg/L of culture medium. The expression levels of the two variant proteases ranged from 1.5 to 2 mg/L of culture medium. The fusion wild-type and variant proteases were inactive before the cleavage of the thioredoxin-hexahistidine fusion tag as no enzymatic activity was observed. The proteases were, however, active after cleavage of the tag. The novel

thioredoxin-hexahistidine fusion system, therefore, enables the successful overexpression and purification of catalytically active HIV-1 proteases.

Keywords

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

Abbreviations

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

1. Introduction

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

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

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

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

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

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

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

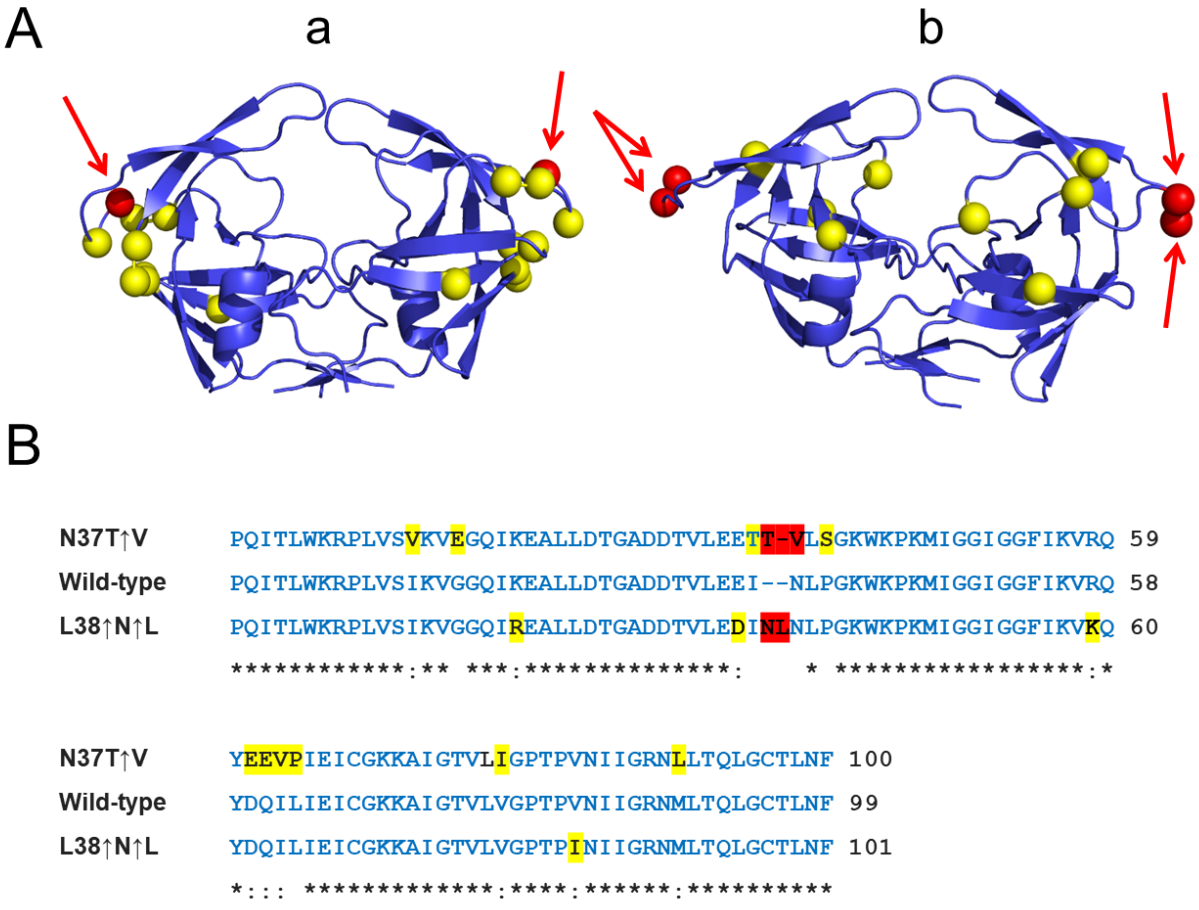


Fig. 1 A Homology models of the (a) N37T↑V and (b) L38↑N↑L proteases. The secondary structural elements of the homology models are rendered as ribbons. The relative positions of the amino acid insertions (red spheres) are indicated by arrows. Yellow spheres without arrows represent background mutations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

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

2. Materials and Methods

2.1 Construction of the fusion plasmids

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

2.2 Expression and purification

The control wild-type protease was purified using a standard protease purification system routinely used in our laboratory [23]. Briefly, *E.coli* BL21 (DE3) pLysS cells were transformed with a pET-11b vector encoding the control wild-type protease gene. The cells were induced for four hours with 1 mM IPTG, and the protease was recovered from inclusion bodies after cell disruption. Recovery buffer contained 8 M urea, 10 mM Tris-HCl and 2 mM DTT (pH 9). The sample was incubated for one hour in the urea buffer before recovery by centrifugation, 23 000xg for 30 minutes at 20 ℃. The sample was dialysed (and refolded) against 10 mM sodium acetate (pH 5) and purified using CM-Sepharose ion exchange chromatography with a 0-1 M NaCl gradient. The control protease was included to measure the success of the new purification strategy.

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

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

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

The L38↑N↑L fusion protease was overexpressed in six litres of LB media at 20 °C overnight using 1 mM IPTG. The cells were resuspended in 40 ml of 20 mM Tris-HCl buffer (pH 7.4). The cells were sonicated as described earlier and the soluble fraction was isolated by centrifugation at 24 000×g. The protease was purified from the soluble fraction using a 5 ml IMAC column and eluted using an imidazole gradient (0-500 mM). Following thrombin cleavage (as described earlier), and 10 mM formic acid precipitation, the sample was dialysed against 10 mM sodium acetate buffer (pH 5). The sample was passed through a CM-Sepharose column to remove any unwanted protein. The protease was eluted using a 0-1 M NaCl gradient and dialysed against 10 mM sodium acetate buffer (pH 5) at 4 °C overnight to remove any residual NaCl. The absence of the salt decreases autolysis (the ability

to undergo autoproteolysis in solution). The purity of the protease was evaluated by a 16% tricine-SDS-PAGE [24,25]. The final concentration of pure protease was determined using the absorbance value at 280 nm and the extinction coefficient of the protein according to the Beer-Lambert equation. The extinction coefficients used were: 25 480 $M^{-1}.cm^{-1}$ for wild-type, 24 980 $M^{-1}.cm^{-1}$ for N37T↑V and 25 230 $M^{-1}.cm^{-1}$ for L38↑N↑L. Molar extinction coefficients were calculated using the following equation [26]:

$$\varepsilon = 550 * \sum Trp + 1340 * \sum Tyr + 150 * \sum Cys$$

2.3 Structural characterisation

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

2.4 Functional characterisation

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

An enzyme assay was conducted during thrombin cleavage to determine whether the protease was catalytically active. The increase in fluorescence intensity attributed to the cleavage of the fluorogenic substrate: Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2 was measured. The sample was

excited at 337 nm and the fluorescence emission monitored at 425 nm. The assay was performed on a Jasco FP-6300 Spectrofluorometer.

3. Results

3.1 Construction of fusion plasmids

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

3.2 Overexpression of fusion proteases

Figure 3 represents the whole-cell lysates. The gel shows the improved expression profile of the fusion wild-type (Fig. 3, lane 2, ~25 kDa) compared to that of the control wild-type protease (Fig 3, lane 1, ~11 kDa). This is seen by a thicker band at ~25 kDa in lane 2 compared to the band at ~11 kDa in lane 1, indicated by the arrows. Samples were normalised before electrophoresis to ensure that equal amounts of cell lysate were loaded onto each gel. The size of the fusion-product corresponds to the predicted size of the reduced, monomeric fusion protein 25.1 kDa (Figure 2, ProtParam tool, http://www.expasy.ch/tools/protparam.html) [27].

3.3 Protease purification

The control wild-type was overexpressed and purified by ion exchange chromatography as previously described by Naicker *et al*. (2014) [28]. The three fusion proteases were purified by immobilised metal ion affinity chromatography (IMAC). The steps involved in the purification of the fusion wild-type, N37T↑V and L38↑N↑L proteases are shown in Figure 4A, B and C, respectively. The last lane in each gel (IMAC 2 peak) shows the pure cleaved protein. The amount of protein was not normalised.

The insoluble cell fractions used in this study were incubated in buffer containing 8 M urea, 10 mM Tris-HCl and 2 mM DTT (pH 9). The cell debris was collected by centrifugation, and the resultant supernatant was diluted to a final concentration of 4 M urea before the first IMAC step. The dilution was performed to prevent spontaneous crystallisation of the urea. The L38↑N↑L fusion construct was purified from the soluble fraction by metal ion affinity chromatography. The first chromatographic step yielded high concentrations of all three fusion proteases. The final yield of

free protease is represented in milligrams per litre of culture. The data are represented in Figure 5. The yield of fusion wild-type increased 3-fold compared to the control wild-type. The N37T↑V expression yielded 2 mg/L culture and L38↑N↑L yielded 1.5 mg/L culture.

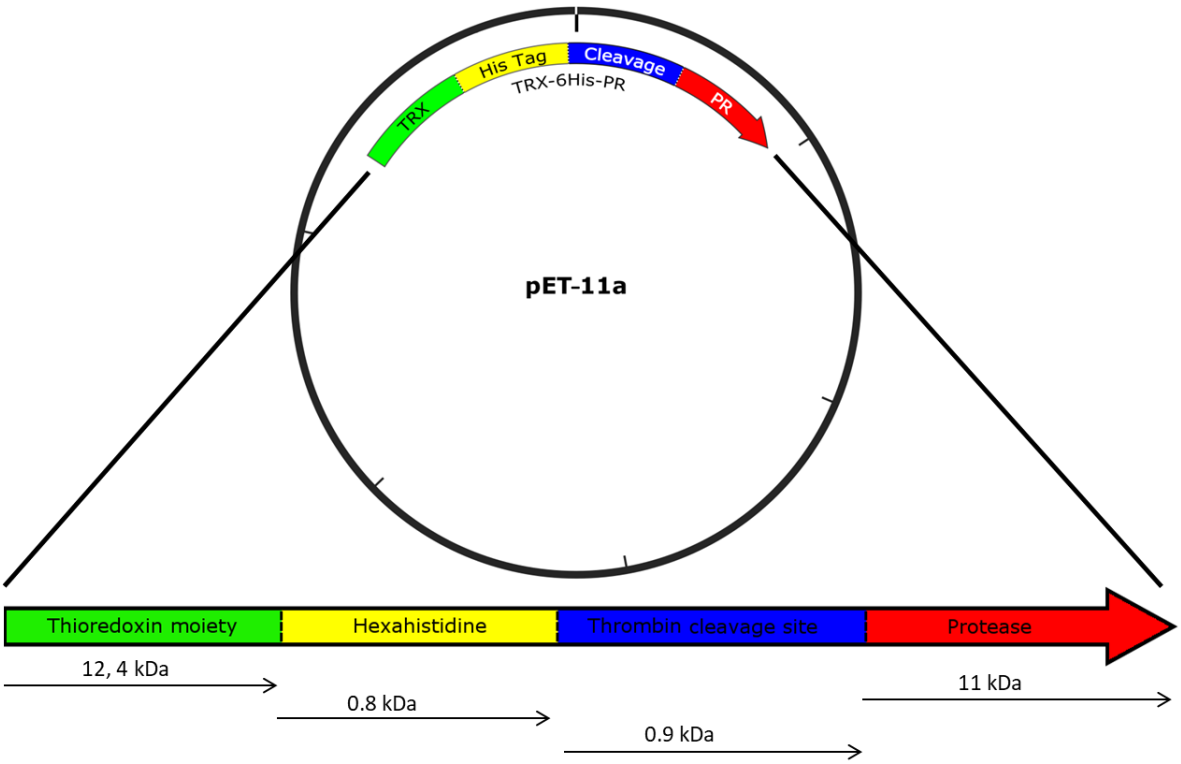


Fig. 2 Plasmid construct of TRX-6His-TCS-PR. The abbreviation TRX-6His-TCS-PR denotes the thioredoxin-like moiety (TRX), hexahistidine (6His) tag and protease (PR) enzyme. A thrombin cleavage site (TCS) is present between the hexahistidine tag and the protease. The size of each constituent is shown and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

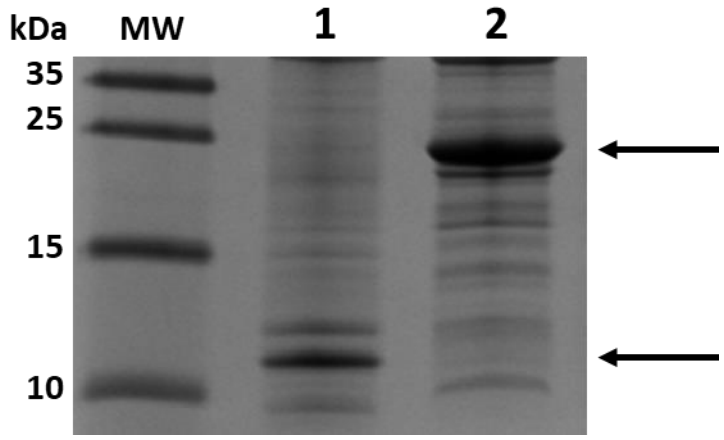


Fig. 3 Schagger Tris-Tricine SDS-PAGE (16%) gel showing the overexpression of control wild-type and fusion wild-type. Whole lysates are shown. Transformed BL21 (DE3) pLysS *E. coli* cells were grown to early exponential phase and induced for six hours with 1 mM IPTG. MW: molecular weight marker, lane 1: control wild-type protease overexpression, lane 2: fusion wild-type overexpression. The positions of the fusion wild-type protease (lane 2, ~25 kDa) and the control wild-type protease (lane 1, ~11 kDa) are indicated by arrows

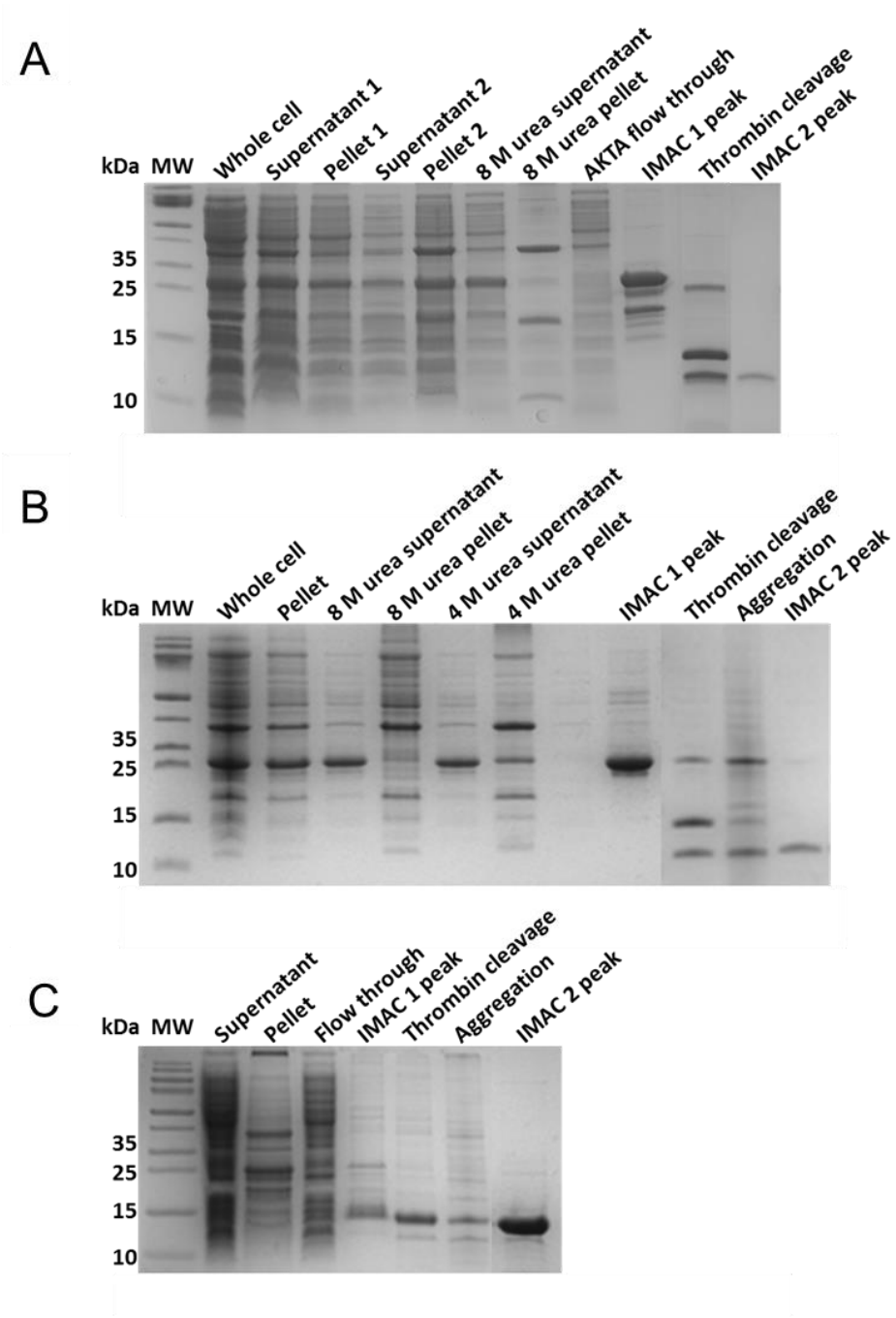


Fig. 4 Overexpression profiles of fusion wild-type and variant proteases from IPTG-induced cell lysates. The purification steps, from cell lysis to final product, are shown from left to right. MW: molecular weight marker. A Fusion wild-type purification profile. B N37T↑V fusion purification profile.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

.

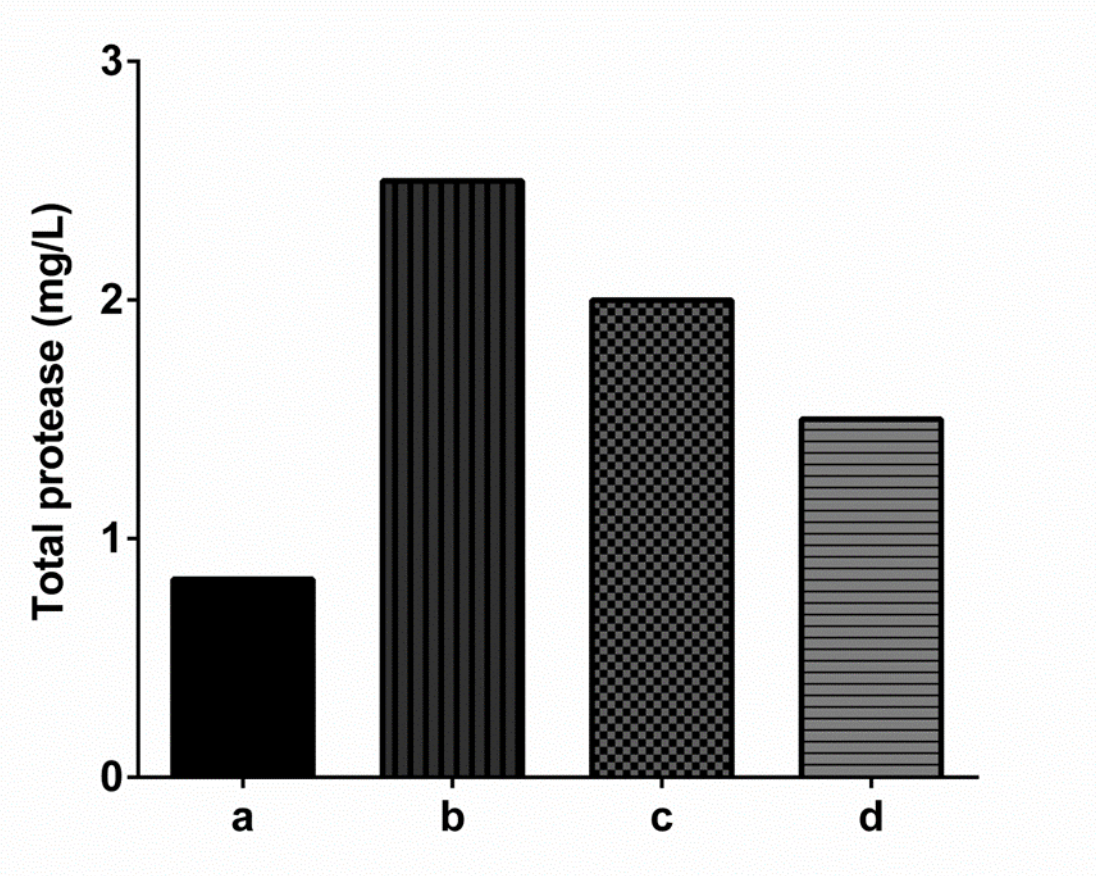


Fig. 5 Quantity of fusion-derived HIV-1 protease produced per litre of culture media compared to an ion exchange purification method. Bar (a) 0.83 mg/L control wild-type protease, (b) 2.5 mg/L fusion-derived wild-type protease, (c) 2 mg/L N37T↑V protease, and (d) 1.5 mg/L L38↑N↑L protease.

3.4 In vitro fusion protease processing

Thrombin cleavage trials were conducted on the TRX-6His-TCS-PR construct to determine the optimal time, temperature and amount of thrombin required for optimal cleavage (Fig. 6A). We found that ideal cleavage occurred overnight at 20 °C with 1 U/ml thrombin. As thrombin cleavage progressed, protease activity (Fig. 6B) was measured by conducting enzyme assays which followed the cleavage of Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2. This peptide mimics the cleavage site between the capsid and p2 proteins of the Gag-Pol polyprotein. To assess whether the protease samples possessed functional activity, we monitored the activity of the proteases using linear progress curves. The fluorescence intensity increased with time as the fusion protein was cleaved (Fig 6A) as seen by the linear progress curve in Fig 6B. The progress curve was expected to be linear as the substrate was in saturating concentration.

3.5 Structural analysis

The quaternary structures of the cleaved fusion proteases were analysed using high-performance liquid chromatography (refer to supplementary figures). The size of fusion wild-type, N37T↑V and L38↑N↑L proteases are indicated by an X in supplementary Figure 1. The results indicate that the dimeric sizes of the proteins were: 22 kDa, 23 kDa and 22 kDa for the fusion wild-type, N37T↑V and L38↑N↑L proteases, respectively. These sizes correspond with the expected sizes of the fully folded homodimeric molecule.

3.6 Enzyme activity determination

An active site titration was performed on each purified protease sample using a VP-ITC Microcalorimeter (data not shown). It is important to assess the percentage active protease in a sample preparation because HIV-1 protease possesses autolytic activity [12]. This procedure, therefore, allows the experimenter to correct the concentration of active protease in a sample. Obtaining the concentration via absorbance spectroscopy at 280 nm and applying the Beer-Lambert equation is insufficient. Acetyl pepstatin is a naturally occurring weak inhibitor, and it was titrated against each protease sample. Since the stoichiometry of acetyl pepstatin binding to HIV-1 protease is known (1:1), it is possible to determine the concentration of active protease in a sample as a function of the total measured protease concentration [29, 30]. Upon titration, the percentage of each protease in the active conformation was; 13% fusion wild-type, 32% N37T↑V and 9% L38↑N↑L.

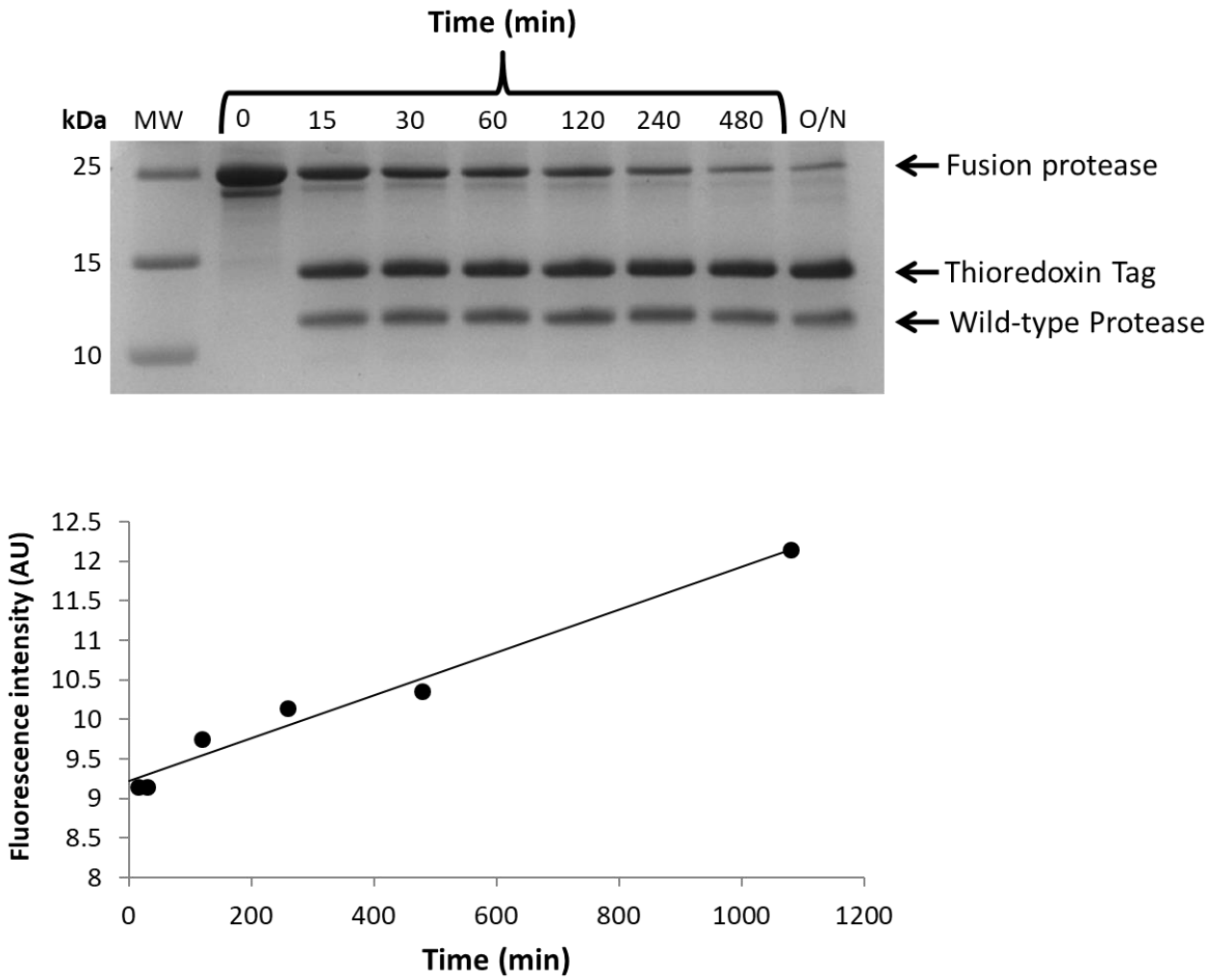


Fig. 6 A time-course thrombin cleavage assay of fusion wild-type protease. B Fusion wild-type protease activity over time. Protease activity was monitored during thrombin (1 U/ml) cleavage by following fluorogenic substrate (Abz-Arg-Val-Nle/Phe(NO_2)-Glu-Ala-Nle-NH_2) processing at a wavelength of 425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

4 Discussion

HIV-1 protease represents a major drug target in the treatment of HIV/AIDS. To study this enzyme, it is important to obtain sufficient quantities for use in biochemical and biophysical studies. Heterologous overexpression of the viral enzyme does not occur readily. In fact, HIV-1 protease exhibits cytotoxic effects when expressed in a variety of host cells, including bacteria, yeast and mammalian cells. Due to the cytotoxic nature of HIV-1 protease, it is difficult to obtain large quantities of the enzyme. In this paper, we describe the overexpression and purification of the wild-type and two variant HIV-1 proteases using a thioredoxin-hexahistidine fusion system. A thioredoxin moiety coupled with a hexahistidine tag successfully improved the overexpression of all three proteases.

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

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

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

The overexpression of the fusion proteases was notably greater than that of the control wild-type protease, which was purified using ion exchange chromatography. Conventionally, a large volume of

culture media (6-8 L) is required to generate a yield of 0.83 mg/L of culture. In this paper, we demonstrated that one litre of culture could produce 2.5 mg/L of fusion wild-type, 2 mg/L of N37T↑V and 1.5 mg of L38↑N↑L. Our novel HIV protease fusion purification method, therefore, produced significantly higher yields of pure protease than our control method (Gag-Pol derived protease).

The quaternary structure of each protease was analysed by determining the relative hydrodynamic volume using high-performance liquid chromatography. An online tool was also used to predict the sizes of the dimeric proteases for comparison. The predicted sizes of all the fusion proteases were 22 kDa (ProtParam tool) [27]. HIV-1 protease is an obligate homodimer and must be conformationally stable to function correctly. The experimentally determined sizes of the proteases were as follows: fusion wild-type, 22 kDa; N37T↑V, 23 kDa and L38↑N↑L, 22 kDa. The sizes correspond to the homodimeric size of the HIV-1 protease.

HIV-1 protease is autolytic. Therefore, it is crucial to determine the percentage of active enzyme in a prepared sample. Active site titrations, determined using ITC, showed the percentage of active enzyme in each protease sample and also verified that all the enzymes possessed enzyme activity [32]. Thirteen percent of the fusion wild-type enzyme sample was active and available to the natural ligand; whereas, N37T↑V and L38↑N↑L had 32% and 9% of active proteases in these samples, respectively. The low percentage of active proteases in the samples could be explained by high levels of autolytic activity that often occurs when proteases are incubated for an extended period (e.g. thrombin cleavage). The observed autolytic activity is particularly interesting because these proteases contain a Q7K mutation that should minimise autolysis [33]. L38↑N↑L was expressed in the soluble fraction thus indicating that it was most likely active and able to undergo autolysis which could have contributed to the lower percentage of active protease in this sample. It is not clear whether the low percentage active sites can be attributed to increased autolysis due to the mutations present in each protease or if it is due to other factors.

Other groups have investigated the effectiveness of different HIV-1 protease fusion expression systems. In those systems, autocatalysis occurred despite the presence of the tags [34]. In our study, we postulate that the relative size of the TRX-6His-TCS moiety does not interfere with protease dimer formation – dimerisation is essential for autocatalytic activity (removal of itself from the Gag-Pol polyprotein). The thioredoxin moiety, however, sufficiently mimics the structure of the Gag protein from which HIV-1 protease cleaves itself. Here, autocatalysis (autoexcision from the Gag-Pol

precursor) must not be confused with autolysis - which is the ability of a protease to undergo autoproteolytic activity in solution. Our results are, therefore, different to the studies from others [34]. In our case, we postulate that the TRX-6His-TCS moiety and the protease form higher order oligomeric states where steric hindrance effects inhibit the autocatalytic activity of the protease. This postulation is demonstrated by the observed increase substrate cleavage as a function of thrombin cleavage time as the protease is released from the TRX-6His tag (Fig. 5). The presence of higher order oligomeric states could be determined using size exclusion chromatography, analytical ultracentrifugation and static light scattering.

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

5 Conclusion

The procedure described in this study highlights a quick and easy method of HIV-1 protease purification. In addition to the smaller volume of culture media needed, the total wild-type protease yield from this fusion system exceeds our control purification method by 250%. Because the fusion proteases are autolytic, a suitable method of inhibition could be included during the purification step so that higher yields of active protease are obtained. Although a subtype C protease was used for this study, the system could also be applied to HIV-1 proteases from other subtypes.

References

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

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

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

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

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

6. McCutchan FE (2006) Global epidemiology of HIV. J Med Virol 78:7–12. doi: 10.1002/jmv.20599

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

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

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

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

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

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

13. Nutt RF, Brady SF, Darke PL, et al. (1988) Chemical synthesis and enzymatic activity of a 99-residue peptide with a sequence proposed for the human immunodeficiency virus protease. Proc Natl Acad Sci U S A 85:7129–7133.

14. Cheng YS, McGowan MH, Kettner CA, et al. (1990) High-level synthesis of recombinant HIV-I protease and the recovery of active enzyme from inclusion bodies (Expression; synthetic genes; refolding; proteolysis; recombinant DNA). Gene 87:243–248.

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

16. Yang H, Nkeze J, Zhao RY (2012) Effects of HIV-1 protease on cellular functions and their potential applications in antiretroviral therapy. Cell Biosci 2:32. doi: 10.1186/2045-3701-2-32

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

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

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

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

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

22. Arnau J, Laurtitzen C, Petersen GE, et al. (2005) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr Purif 48: 1–13. doi:10.1016/j.pep.2005.12.002

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

24. Laemmli, U. K. (1970). Cleavage of structural proteins duringthe assembly of the head of bacteriophage T4. Nature 227:680–685.

25. Schägger, H. (2006). Tricine-SDS-PAGE. Nature Protocols, 1:16–23.

26. Perkins S (1986) Protein volumes and hydration effects. Eur J Biochem 157:169-180

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

28. Naicker P, Sayed Y (2014) Non-B HIV-1 subtypes in sub-Saharan Africa: Impact of subtype on protease inhibitor efficacy. Biol. Chem. pp 1151–1161

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

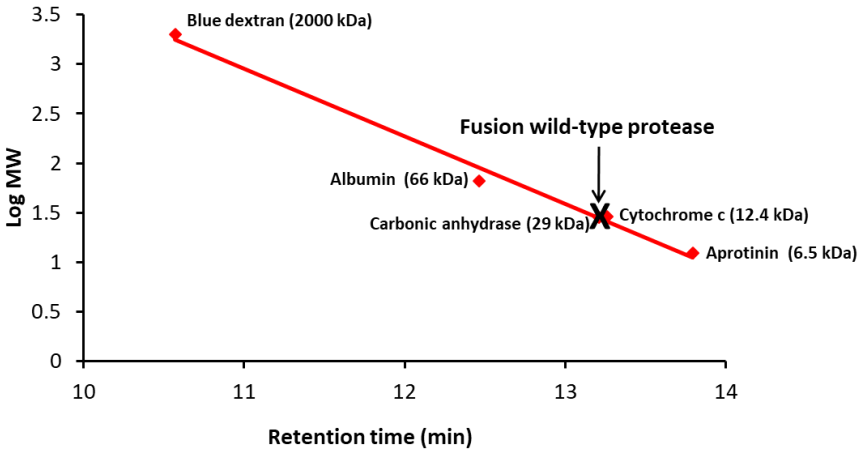
30. Velazquez-Campoy a, Kiso Y, Freire E (2001) The binding energetics of first- and second-generation HIV-1 protease inhibitors: implications for drug design. Arch Biochem Biophys 390:169–175. doi: 10.1006/abbi.2001.2333

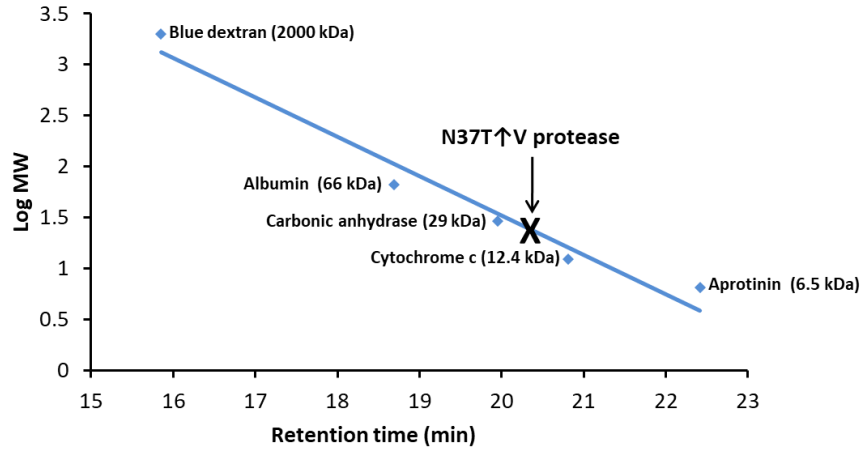
31. Sawyer JR, Schlom J, Kashmiri S V (1994) The effects of induction conditions on production of a soluble anti-tumor sFv in Escherichia coli. Protein Eng 7:1401–1406.

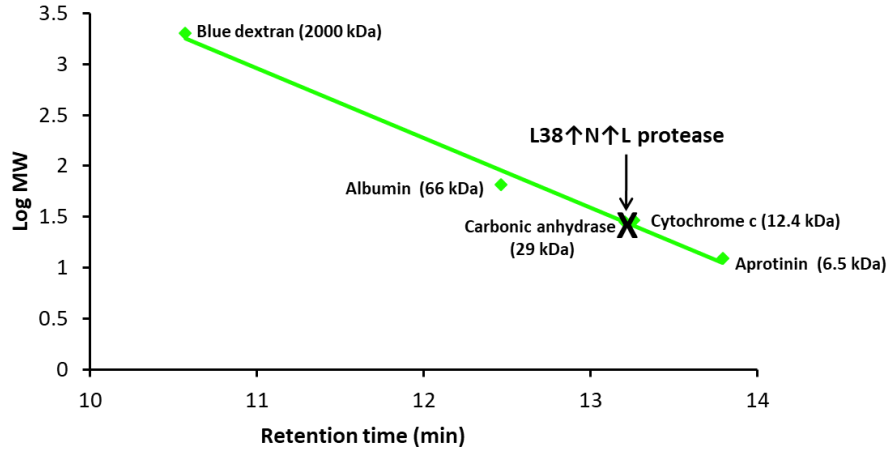
32. Tomasselli AG, Olsen MK, Hui JO, et al. (1990) Substrate analogue inhibition and active site titration of purified recombinant HIV-1 protease. Biochemistry 29:264–269. doi: 10.1021/bi00453a036

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

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







Supplementary Fig. 1: Size exclusion-HPLC retention times of fusion wild-type, N37T↑V and L38↑N↑L proteases. The molecular standards consisted of blue dextran (2000 kDa), serum albumin (66 kDa), carbonic anhydrase (29 kDa), cytochrome C (12.4 kDa) and aprotinin (6.5 kDa). The retention time of each protease is indicated. The relative molecular weight of each protease was calculated from the standard curve. The elution of fusion wild-type and L38↑N↑L was 0.25 ml/min and the elution of N37T↑V was 0.20 ml/min.

CHAPTER 3
MOLECULAR DYNAMICS

Molecular dynamic simulations of L38↑N↑L HIV-1 protease shows reduced flap dynamics

Alison Williams, Vijayakumar Balakrishnan, Ikechukwu Achilonu, Heini Dirr and Yasien Sayed

Molecular Simulation (Under review)

In this publication the dynamics of an HIV-1 protease containing a double insertion in the hinge region was analysed. The flap region of L38↑N↑L protease was not as dynamic as the wild-type and is stabilised by a number of salt bridges. Induced-fit docking of Lopinavir and Darunavir to L38↑N↑L showed reduced hydrophobic contacts and docking Atazanavir showed reduced hydrogen bonding.

Author contributions: Alison Williams performed and analysed the molecular simulations using GROMACS, analysed the docking data and prepared the manuscript. Vijayakumar Balakrishnan performed the docking using Glide Schrödinger. Ikechukwu Achilonu assisted in experimental design and manuscript revision. Heini Dirr assisted in manuscript revision. Yasien Sayed supervised the project and assisted in data analysis and interpretation.

Molecular dynamic simulations of L38↑N↑L HIV-1 protease shows reduced flap dynamics

Alison Williams[a], Vijayakumar Balakrishnan[a], Ikechukwu Achilonu[a], Heini Dirr[a] and Yasien Sayed[a*]

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

[*]Corresponding author email: yasien.sayed@wits.ac.za

Abstract

HIV-1 infects 37 million people and most infections are due to the subtype C virus, the subtype of interest in this study. The enzyme HIV-1 protease is essential for maturation of the virus and is an important drug target. The variant studied here, L38↑N↑L, contains insertions of Asn and Leu in the hinge region. Molecular dynamics simulations were conducted on the wild-type subtype C HIV-1 protease and L38↑N↑L using GROMACS. The flap region, two anti-parallel β-sheets, covers the active site. These flaps curl and open to allow substrate to bind to the active site and then close upon substrate binding. The results showed that the flaps in L38↑N↑L were not as dynamic as the wild-type and this is attributed to the presence of the stabilising salt bridges: Asp30-Lys47, Asp62-Lys45 and Asp35-Arg20 in L38↑N↑L. We propose that reduced flap dynamics is part of a mechanism to evade the binding of HIV-1 protease inhibitors in this protease. The HIV-1 protease inhibitors lopinavir, atazanavir and darunavir were docked to L38↑N↑L using Schrödinger Glide. Lopinavir and Darunavir showed reduced hydrophobic contacts while Atazanavir showed reduced hydrogen bonding to L38↑N↑L. The reduction in the number of hydrophobic contacts could be an indication of reduced susceptibility to the HIV-1 protease inhibitors.

Key words

HIV-1, subtype C, protease, hinge region, flap dynamics, molecular dynamics

1.1 Introduction

Human Immunodeficiency Virus (HIV) is a global health problem with 37 million people infected worldwide [1]. HIV infection is associated with acquired immunodeficiency syndrome (AIDS), which is the extreme suppression of the immune system. HIV-1 is genetically diverse and is divided into

three groups [2]. The main group, group M, is divided into nine subtypes and a number of CRFs [2]. Subtype C is the most prevalent subtype occurring in sub-Saharan Africa, the region with the highest rate of HIV-1 infections world-wide [3,4]. Subtype C is of interest for this study because the variant is of subtype C origin.

HIV utilises three enzymes in its replication cycle; integrase, reverse transcriptase and protease. HIV-1 protease is important for the maturation of the virus as it cleaves the Gag and Gag-Pol polyproteins into their functional forms resulting in mature, infectious virions. Inhibition of this enzyme results in immature non-infectious virions [5] making it an ideal drug target. HIV-1 protease is a dimeric protease consisting of two subunits, 99 amino acids each, which are noncovalently associated [6]. It is an obligate homodimer and is inactive as a monomer[7]. It is an aspartyl protease and the active site, formed at the dimer interface, consists of the characteristic Asp-Thr-Gly sequence (catalytic triad). One catalytic triad from each monomer is contributed.

HIV-1 protease is composed predominately of β sheets. The active site is covered by two glycine-rich β-hairpin structures (residues 43-58), termed the flaps [8]. Residues within the flap region form numerous interactions with the substrate and thus the flap region plays an important role in substrate binding [9]. This region is highly flexible and can undergo large conformational changes during substrate binding and release [10]. It has been shown by NMR that the flaps are highly mobile and exist in open, semi-open and closed conformations [11]. In the apo-form the flaps exist in an open conformation 12-14 Å apart. Upon substrate binding, these flaps close to approximately 5.9 Å apart [12]. Naicker, et al., [12] showed that the flaps of wild-type subtype C protease are more dynamic than those of wild-type subtype B protease. The flap tips, residues 46-54, are dynamic but move on a smaller time scale when compared to the entire flap [11]. The flap tips are glycine-rich and this accounts for the increased mobility [11]. The flap tips have been shown to curl inward, towards the active site, before the flaps open [13]. This curling motion allows the peptide chain enough space to access the active site. The movement and stability of the flaps is aided by the presence of loops extending from residues 35 to 42 and are termed the hinge region.

The use of HIV-1 protease inhibitors in antiretroviral therapy has caused mutations to occur in the HIV-1 protease gene. The HIV-1 protease inhibitors approved by the FDA for use in patients are targeted towards subtype B. This is problematic as many of the polymorphisms found in the subtype C protease have been shown to cause drug resistance. The wild-type subtype C protease has a reduced affinity toward HIV-1 protease inhibitors [14]. Primary mutations mainly occur in the active

site and directly affect substrate binding thus decreasing the affinity for the drug. These mutations are selected early and can be inhibitor specific [15]. Primary mutations decrease the efficiency of the HIV-1 protease and render high levels of drug resistance. Secondary mutations are usually distal from the active site and occur to compensate for the loss of efficiency. Rarely ,1-6 amino acids are inserted in the region of residues 35-42, the hinge region. This is the region that contains many of the naturally occurring polymorphisms in subtype C such as E35D, M36I, S37N, R41K and R57K [16]. The hinge region aids the flaps to close upon substrate or inhibitor binding. Mutations in this region, therefore, may contribute towards reduced binding free energy [16]. Some studies have suggested a role for amino acid insertions in the drug resistance to HIV-1 protease inhibitors [17–19]. Due to the rarity of these insertions the effect on the dynamics of the protein have not been studied extensively.

The variant protease in the current study contains an insertion of Asn and Leu at postion 38 resulting in a monomer of 101 amino acids. We have a developed a nomenclature to indicate the insertion with the use of upward arrows in the nomenclature L38↑N↑L (Yasien Sayed and Ikechukwu Achilonu, Protein Structure-Function Research Unit, University of the Witwatersrand, South Africa). The variant also contains the following subset of polymorphisms K20R, E35D, R57K and V82I. This study deals with the altered flap dynamics of L38↑N↑L and docking of the drugs LPV, ATV and DRV to L38↑N↑L.

1.2 Methods

1.2.1 Molecular dynamics simulation

A homology model of L38↑N↑L was created using SWISS-model. The South African subtype C protease structure (PDB code 3U71, resolved at a 2.7 Å resolution [12]) was used as a template due to its high homology. This template was chosen over other subtype C structures because the variant in this study is of South African origin and contains the correct polymorphisms. The homology model was validated using PROCHECK [20,21]. The simulations were carried out using GROMACS version 5.07 [22,23] on an Intel core i7 5960x extreme edition (3.3 GHz, 20 M cache 16x cores) equipped with GTX 750Ti graphics card, 32 GB DDR4-2133 MHz memory on a MSI X99 motherboard. The AMBER99sb force field was used for both models [24,25]. The simulations were conducted with explicit solvent in a cubic box universe. The long-range electrostatics were handled by the particle-mesh Ewald (PME) method [26]. The solvated systems were relaxed with energy minimisation to less than 100 kJ/mol/nm followed by 5 ns of MD simulation under the NPT ensemble (constant number

of particles (N), constant pressure (P) and constant temperature (T)). The temperature was increased linearly from 10-300 degrees Kelvin. The positional restraints were calculated and the original values annealed to relax the models and ensure stable temperature. The temperature-stable models were subjected to 5 ns of simulation under the NVT ensemble (constant number of particles (N), constant volume (V) and constant temperature (T)) where the pressure of each system was equilibrated. After equilibration the MD simulations were performed for 20 ns under a constant temperature of 300 K with a Berendsen thermostat and an average pressure of 1 atm maintained by Parrinello-Rahman barostat algorithm [27]. Analysis of the trajectories was carried out using the script-based utilities in GROMACS version 5.07, Visual Molecular Dynamics [28] and the Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081) [29]. Molecular graphics were generated using Chimera and PyMoL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

1.2.2 Induced fit docking

Ligand binding to an active site causes side-chain or backbone (or both) conformational changes in proteins. These conformational changes allow the active site in the protein to conform closely to the shape of the ligand. This process is known as flexible docking (IFD) and uses the Glide and Prime modules (Schrödinger LLC, 2009, USA). The Glide software was used to perform induced fit docking and calculate binding free energies. The docking simulations were carried out on a CentOS EL-5 workstation. The Schrodinger modules Glide, Prime, QSite, Liaison and MacroModel were used for protein preparation. Each model was modified by correcting bond orders while ionisable residues were assigned a charge corresponding to a solution at pH 5 (experimentally determined optimal pH). Models were subjected to energy minimisation until the average RMSD reached 0.3 Å. The three HIV-1 protease inhibitors used in HAART in South Africa; Lopinavir (LPV), Atazanavir (ATV) and Darunavir (DRV) were obtained from PubChem Compound database. Ligand energy minimisation was done using the Ligprep tool (Schrödinger LLC, 2009, USA). Ligprep converted 2D structures to 3D structures, added hydrogens, bond angles and lengths, chose correct chirality and performed energy minimisation. The Epik too was used to choose the lowest energy tautomers and ring structures. Energy minimised wild-type and L38↑N↑L protease complexed with either LPV, ATV or DRV was loaded into the workspace and the ligand was selected to specify the active site. van der Waals radii on non-polar atoms of the HIV-1 protease and the ligand were scaled by a factor of 0.50 and 20 conformational poses were calculated (system default setting). The best conformation, based on docking score, Glide energy and Glide E-model, was chosen for further analysis. The hydrophobic

interactions and hydrogen bonds between the ligand and the HIV-1 protease were viewed using Ligplot+ [30].

1.3 Results

1.3.1 Molecular dynamics simulation

The L38↑N↑L homology model was obtained using SWISS-model and validated using PROCHECK. Ramachandran analysis showed 93.2% of residues were in the most favoured regions, 6.8% were in the allowed regions and 0% were in the generously allowed regions and disallowed regions. Superimposing the variant model on the template (3U71) resulted in a root mean square deviation (RMSD) of 0.097 Å. Fig. 1 shows the superimposed structures and the region where the L38↑N↑L model deviated from the template is seen in the hinge region (circled in red).

The root mean square fluctuations (RMSF) shows the fluctuations of each residue in the simulation, Fig. 2. The residues with the greatest fluctuations (35-61) are represented in the red boxes in Fig. 2. These residues correspond to the flap and hinge region of the HIV-1 protease as is seen in the structure in Fig. 2. The greatest difference in RMSF was seen in residues 35-44 of L38↑N↑L, which differed (at its maximum) from wild-type by 1.2 Å in chain A and 1.1 Å in chain B.

The RMSD value during the simulation infers the overall dynamic motion of the Cα backbone structure. The RMSD remained stable for both wild-type (1.8 Å) and L38↑N↑L (2.4 Å) during the simulation (Fig. 3). The radius of gyration (Rg), which is an index if the compactness of the protein structure, shows that both proteins maintained compactness during the simulation. Wild-type began the simulation more compact (Rg of 1.74 nm) than L38↑N↑L and decreased in compactness until an Rg of 1.68 nm at 5.5 ns but increased to an average of 1.77 nm for the remainder of the simulation. L38↑N↑L began the simulation less compact than wild-type (Rg of 1.82 nm) and decreased to an average Rg of 1.74 nm for the remainder of the simulation (Fig. 3).

The distance between the Ile50 residues were measured to determine when the flaps open during the simulation as these residues occur on the tips of the flaps. The distance between Ile52 residues in L38↑N↑L were measured as these are the residues corresponding to Ile50 in wild-type. The wild-type protease is initially in the semi-closed conformation and then the flaps open, at 6 ns until 15 ns, and close at 15 ns, ending the simulation closed (Fig. 4). The L38↑N↑L protease began the simulation in an open conformation and then proceeded to the semi-closed position (2-8 ns). The flaps then adopted the closed conformation and remained closed for the rest of the simulation (Fig.

4). To determine the distance between the flaps in the open and closed conformations the distance between Ile50 in each chain (Ile52 in L38↑N↑L) was measured. The flaps of L38↑N↑L opened wider (20.8 Å)

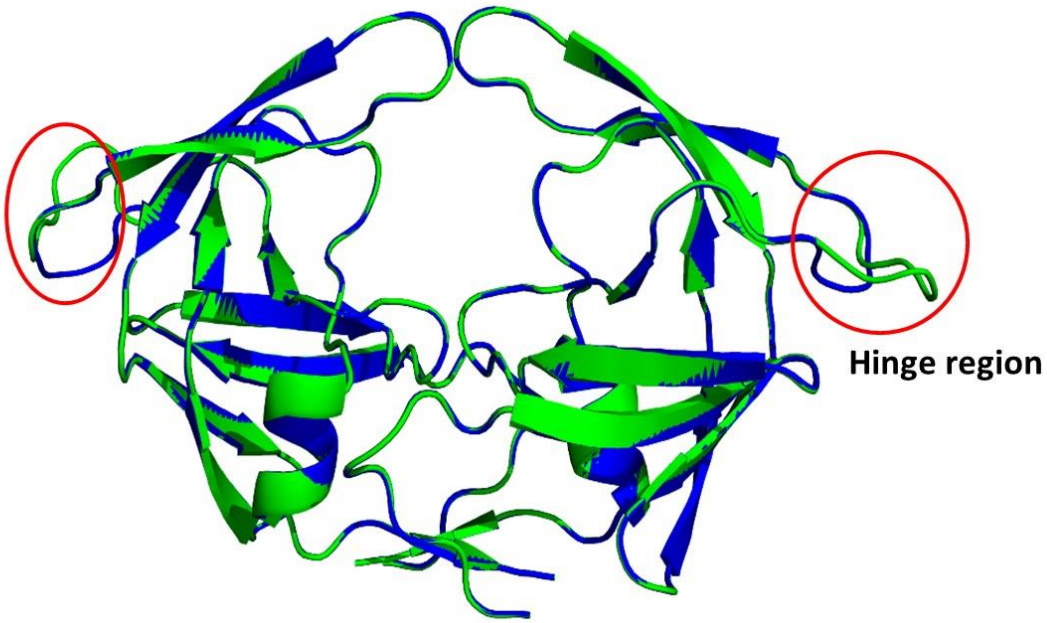


Figure 1: Superimposition of template (blue) and L38↑N↑L (green) protease. The hinge region is circled in red and is the region where the model deviated from the template. The RMSD for the superimposition is 0.097 Å. Figure generated using PyMol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

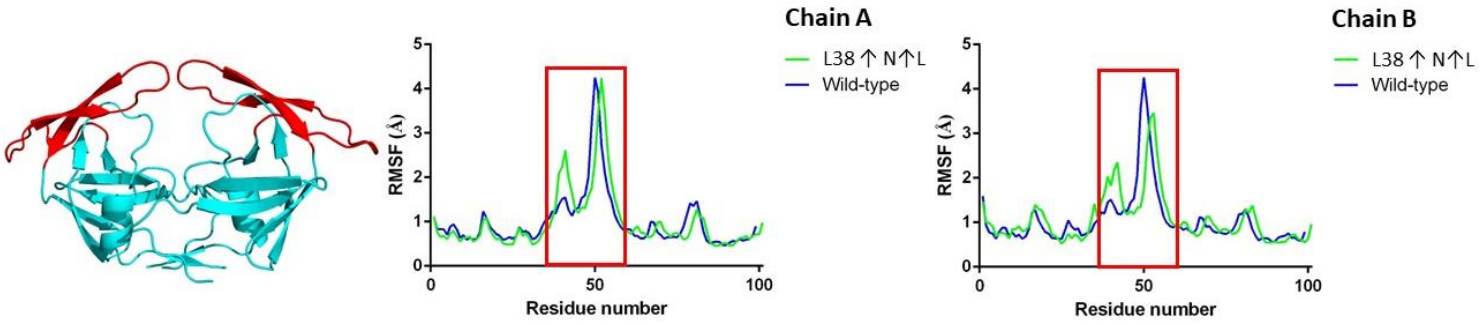


Figure 2: The RMSF of wild-type and L38↑N↑L in chain A and chain B. The residues with the greatest fluctuations are boxed in red and represented in the structure in red. Molecular graphic generated using PyMol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

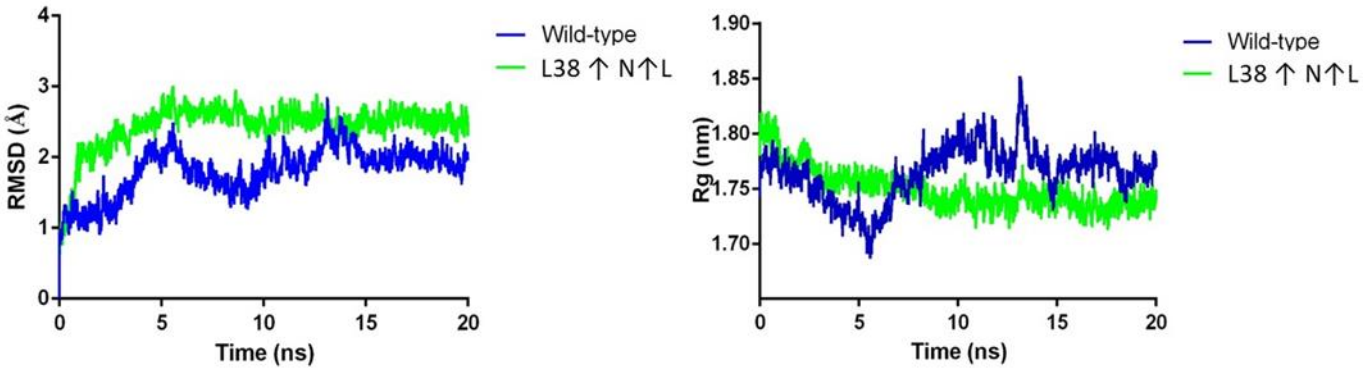


Figure 3: The RMSD and radius of gyration of both the wild-type and L38↑N↑L protease during the 20 ns simulation. The RMSD remained stable for both proteins during the simulations. The radius of gyration showed that L38↑N↑L was more compact during the simulation compared to wild-type.

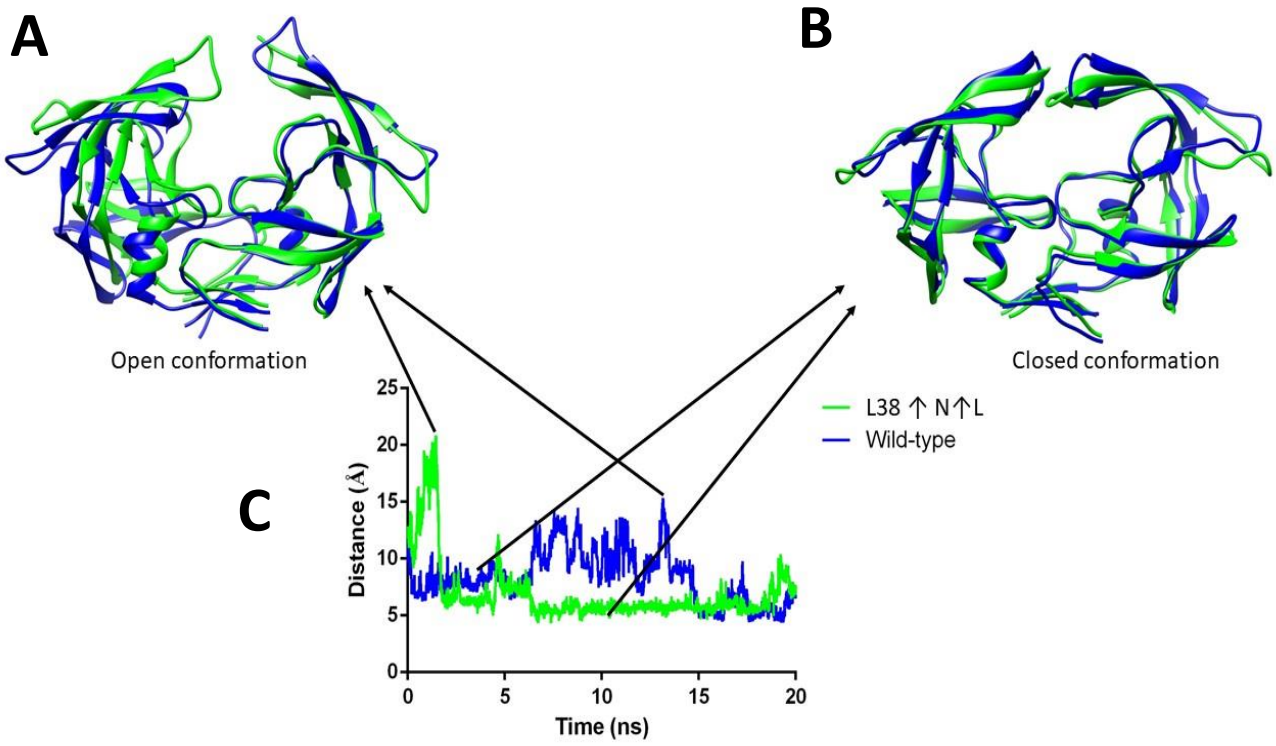


Figure 4: The different conformations of the wild-type and L38↑N↑L protease. The distance between the Ile50 residues (wild-type) and Ile52 residues (corresponding residue in L38↑N↑L) were measured shown (C). The open conformation (A) and closed conformation (B) are represented. Figure generated using Chimera (Pettersen, *et al.*, 2004) [29].

than the wild-type protease (15.4 Å) but in the closed positon the flaps were the same distance apart (4.3 Å).

The curling of the flap tips triggers the flaps to open. The angles between three adjacent alpha carbon atoms (termed the Tri-Cα angles) in the flap tip was measured to determine the degree of flap curling. The Tri-Cα angles in residues Gly48, Gly49 and Ile50 in wild-type, for both chains, were measured and the corresponding residues in L38↑N↑L, for both chains, were also measured (Gly50, Gly51 and Ile52). Wild-type showed greater fluctuations of curling in (~115°) and curling out (~145°) than the L38↑N↑L protease (Fig. 5C and D). Chain B of L38↑N↑L showed greater fluctuations of flap curling (100°-145°) than Chain A, which remained in the curled out conformation (~145°) Fig. 5C and D.

1.3.2 Salt-bridges

A salt-bridge is considered formed when the distance between the positively charged and negatively charged residue is 4 Å or less. Due to the insertions in the hinge region corresponding residues in L38↑N↑L will have a different number after residue 38. If the residue has a different number in L38↑N↑L it will be indicated after a forward slash. Several salt-bridges are found in the hinge region and the flap region of the two proteases. These include Asp30-Lys45/47 (47 indicates the position of the Lys in L38↑N↑L) which exists over a longer time period in the L38↑N↑L protease than the wild-type protease, shown in Fig. 6A. The Glu35-Arg57 (wild-type) or Asp35-Lys59 (L38↑N↑L) salt-bridge, Fig. 6B, exists longer in the wild-type protease than the L38↑N↑L protease and exists in wild-type chain B almost throughout the entire simulation. The salt-bridge Asp60-Lys43 (wild-type) or Asp62-Lys45 (L38↑N↑L) occurs transiently in chain B of both proteases but is maintained between 10 and 15 ns in chain A of L38↑N↑L, Fig. 6C. The Asp35-Arg20 salt bridge only forms in chain A of L38↑N↑L and only at 10 ns of the simulation and exists for 8 ns, Fig. 6D. The Glu34-Lys20 (wild-type) salt-bridge existed for the first 8 ns of the simulation in chain A but only exists transiently in chain B. The corresponding salt-bridge in L38↑N↑L (Glu34-Arg20) does not form in either chain.

1.3.3 Induced-Fit docking

The inhibitors currently in use in South Africa, LPV, ATV and DRV, were docked to both the wild-type and L38↑N↑L protease. The docking scores, Glide E-model and Glide scores are represented in Table 1. The interactions between the wild-type and L38↑N↑L protease and the HIV-1 protease inhibitors are represented in Fig. 7. The hydrophobic contact of the HIV-1 protease inhibitor made

with the protein are represented by half moon circles and the hydrogen bonds are represented by a dotted green line. A reduction in hydrophobic contacts was observed between LPV and L38↑N↑L; two

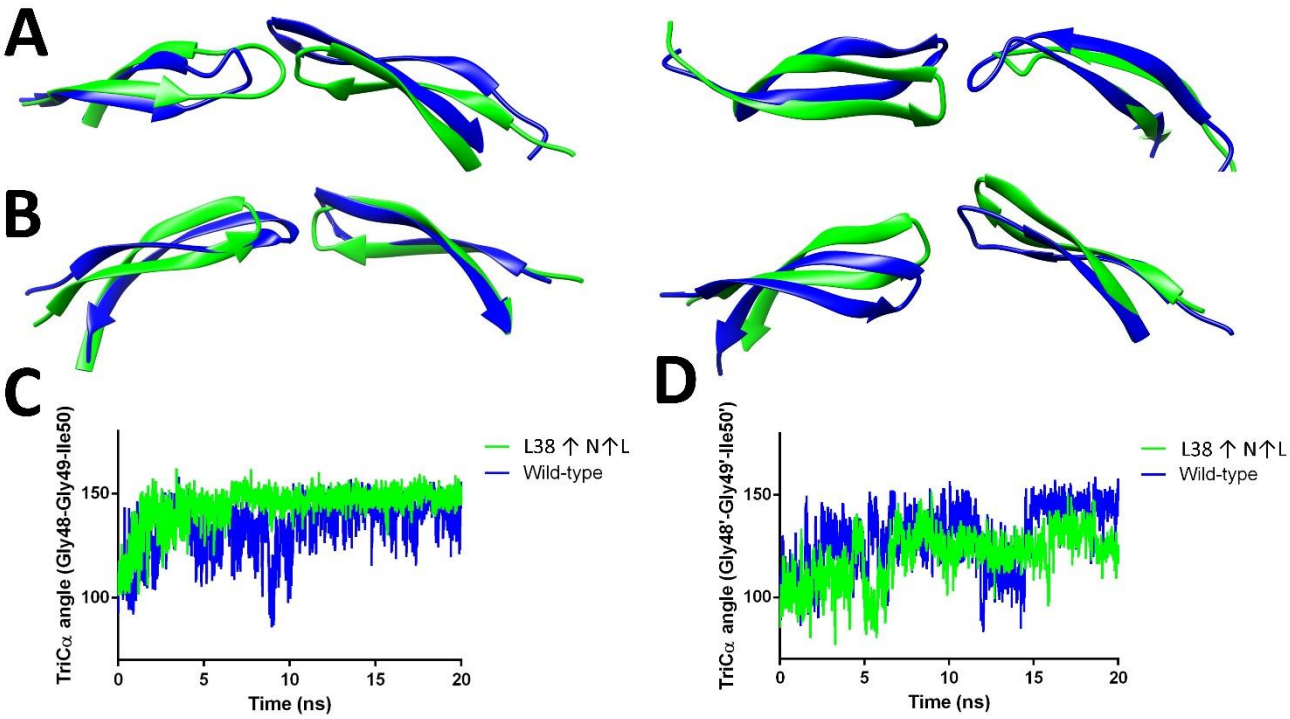


Figure 5: The curled forms of wild-type (blue) and L38↑N↑L (green) protease. Figure 5A represents the curled out forms of wild-type and L38↑N↑L and B represents the curled in forms. The Tri-C_α angles of Gly48-Gly49-Ile50 of wild-type and the corresponding residues in L38↑N↑L were measured for the entire simulation for chain A (C) and Chain B (D). Molecular graphics generated using Chimera (Pettersen, *et al*., 2004) [29].

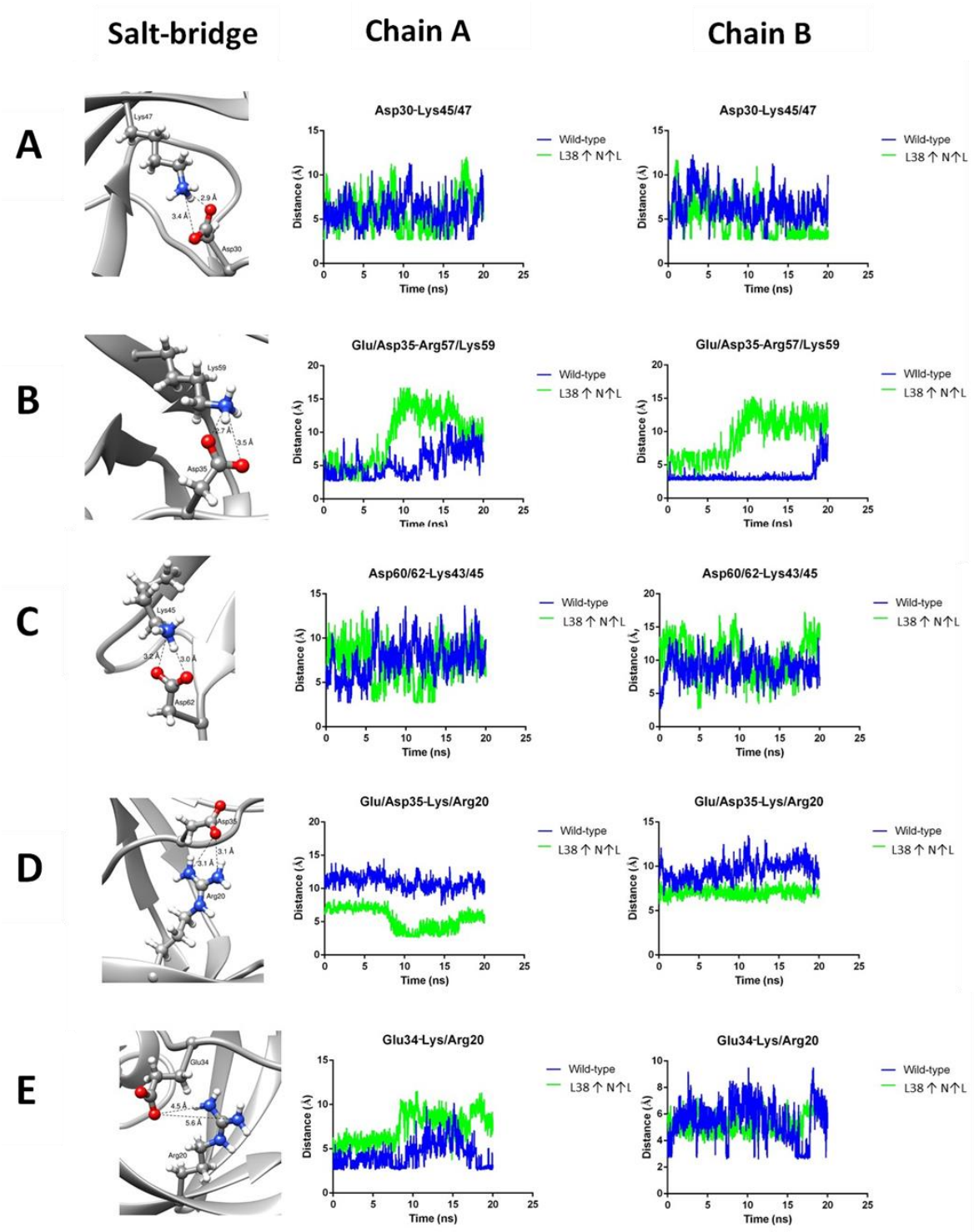


Figure 6: Salt-bridges found within the hinge and flap region of chain A and B of wild-type and L38↑N↑L protease. The first panel shows the salt-bridges formed within 4 Å in the L38↑N↑L protease. The second and third panels show the distances between the two residues in chain A and chain B of each protease respectively for the entire simulation. The residue after the forward slash represents the mutations or altered position in L38↑N↑L. Salt-bride graphic generated using Chimera (Pettersen, *et al.*, 2004) [29].

Table 1: The docking scores, Glide E-model, and Glide energy of LPV, ATV and DRV docked to wild-type and L38↑N↑L protease

Inhibitor	**Docking Score (kcal/mol)**		**Glide E-model (kcal/mol)**		**Glide energy (kcal/mol)**	
	Wild-type	L38↑N↑L	Wild-type	L38↑N↑L	Wild-type	L38↑N↑L
LPV	-12.286	-14.368	-135.386	-160.989	-84.988	-87.513
ATV	-11.589	-10.141	-84.815	-124.038	-140.329	-81.178
DRV	-10.187	-8.282	-108.784	-86.512	-70.780	-60.442

hydrophobic contacts were lost but there is an addition of hydrogen bonds between the side chains of Arg8, Asp29 and Asp30 and LPV. L38↑N↑L and ATV gained five hydrophobic contacts and lost hydrogen bonding between the side chains of Arg8 and Asp30. There was a loss of only one hydrophobic contact for L38↑N↑L and DRV and no loss of hydrogen bonding. No loss of interaction between the catalytic residues (Asp25) was observed in L38↑N↑L protease for any of the HIV-1 protease inhibitors.

1.4 Discussion

1.4.1 Molecular dynamics simulation

The hinge region of HIV-1 protease contains many of the naturally occurring polymorphisms in subtype C and mutations in this region may reduce binding free energy of HIV-1 protease inhibitors [16]. Sometimes amino acids are inserted in this region and this may play a role in drug resistance to HIV-1 protease inhibitors [17–19]. The dynamics of HIV-1 proteases containing these insertions have not been extensively studied. Here we modelled the L38↑N↑L protease containing an insertion of Asn and Leu at postion 38 as well as the following subset of mutations: K20R, E35D, R57K and V82I. A homology model was generated by SWISS-model using the subtype C template PDB: 3U71. This template was chosen because of its high homology to the variant. The template and the variant are both of subtype C origin and thus contain the polymorphisms present in the South African subtype C protease. Validation of the model showed that the residues were all in allowed conformations. Superimposition of the model and the template (Fig. 1) resulted in an RMSD of 0.097 Å, indicating a good fit. It shows that the subset of mutations did not affect the overall structure of the model. The region that deviated in postion was the hinge region; this is unsurprising as the insertion of two amino acids occurs here.

The residues that showed the greatest RMS fluctuations were those that occur in the hinge and flap regions of the proteases (residues 35-45). This confirms that these regions are the most flexible within both enzymes. The RMSD values of both enzymes (wild-type: 1.8 Å and L38↑N↑L: 2.4 Å) remained stable throughout the simulation although more fluctuation was seen in the wild-type protease. This is due this structure being more dynamic in the simulation. The radius of gyration (Fig. 3) of L38↑N↑L (1.74 nm) was smaller than that of the wild-type (1.77 nm). This is because the flaps of L38↑N↑L do not open as often as the wild-type protease (Fig. 4). In order for a substrate or inhibitor to bind to the HIV-1 protease the flaps need to open. The closure of the flaps in L38↑N↑L

may be a mechanism for HIV-1 protease to evade binding HIV-1 protease inhibitors. It has been shown that the flaps are

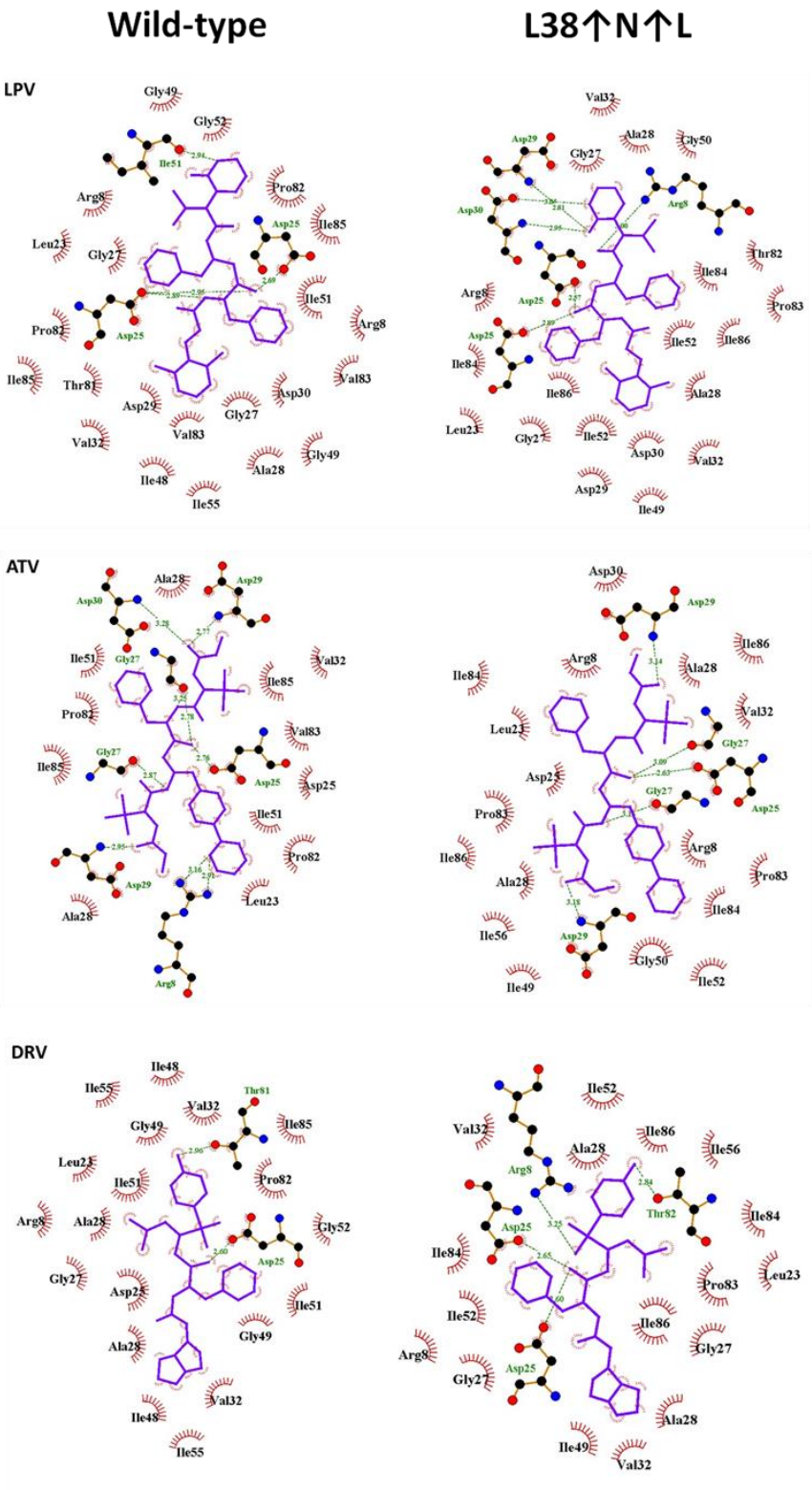


Figure 7: Interaction of LPV, ATV and DRV with both wild-type and L38↑N↑L protease. The half circles represent the hydrophobic contacts between the drug and the protease. Figure generated using Ligplot+ (Laskowski and Swindells, 2011) [30].

necessary for efficient catalytic activity of the HIV-1 protease [31]. The extended period of closure of the flaps would result in a reduction of the catalytic efficiency of the L38↑L↑L. We have demonstrated this to be true in vitro (manuscript in preparation). This finding contrasts Perryman *et al*., [32] who showed than mutations can cause a shift in equilibrium from closed to semi-open conformations, although these mutations did not occur in the hinge region. Zondagh *et al*., [33] showed that mutations in the hinge region can cause the rate of flap opening to increase and insertions in this region cause the flaps to open to a greater extent. This confirms the finding here that the flaps of the L38↑N↑L protease are further apart in the open conformation (20.8 Å) compared to wild-type (15.4 Å).

Flap curling is an important process because it buries the conserved Ile50 residue, which ensures that opening of the flap is thermodynamically favourable. The residues of the flap tips (wild-type and L38↑N↑L) curl inward (~115°) just before the flap opens confirming observations made by Scott and Schiffer [13]. Curling that is more frequent would suggest more flap opening. The flap tip curling fluctuates more in the wild-type protease than the L38↑N↑L protease (Fig. 4) and this, too, is explained by the flaps in L38↑N↑L not opening as often as the wild-type protease.

1.4.2 Salt-bridges

The Asp30-Lys45/47 salt-bridge occurs between the hinge (Asp30) and the flap (Lys45/47) region of the HIV-1 protease. It exists transiently in both chains of the wild-type protease throughout the simulation (Fig. 6A). In L38↑N↑L protease, this salt-bridge exists transiently at the beginning of the simulation but is maintained in chain A between 8-18 ns and chain B 15-20 ns. This is the time scale when the flaps of L38↑N↑L protease are in the closed position, which indicates that this salt bridge may play a role in stabilising the closed structure.

The Asp60/62-Lys43/45 (Fig. 6C) salt-bridge exists transiently during the closed stage of the L38↑N↑L protease and thus may also be a factor in stabilising the closed structure. This salt-bridge is only formed at the beginning of the simulation for wild-type and never forms again because this structure does not remain closed for most of the simulation.

The Glu34-Lys20 salt-bridge seen in chain A of wild-type (Fig. 6E) is not present in the L38↑N↑L protease and may be disrupted by the K20R mutation within L38↑N↑L. K20R is part of the subset of mutations found within the variant. The K20R polymorphism occurs in 20% of drug naïve and 27% of drug treated patients infected with subtype C and is not associated with drug resistance (http://hivdb.stanford.edu/index.html). The Glu34-Lys20 salt-bridge is, however, replaced by the

salt-bridge Asp35-Arg20 (Fig. 6D) in chain A of L38↑N↑L. The salt-bridge is not present in wild-type and only forms in L38↑N↑L when the flaps are in the closed conformation. Therefore, it could possibly aid in maintaining the closed conformation.

In the wild-type protease, the Glu35-Arg57 salt bridge is present in both chains. The E35D mutation found in L38↑N↑L is a common polymorphism in subtype C and occurs in 25% of drug naïve patients and 29% of drug treated patients infected with subtype C (http://hivdb.stanford.edu/index.html). The L38↑N↑L also contains an R57K mutation. These two mutations effectively knock-out the salt-bridge Glu35-Arg57 as can be seen in Fig. 6B. Liu et al., showed that the flap tips exhibited diminished motion due to the E35D mutation [34] which is also seen in the L38↑N↑L protease as it remains closed for most of the simulation (Fig. 4). We show here that the salt-bridge, Glu35-Arg57, is eliminated for most of the simulation, which is in agreement with Liu et al., [34]. It was proposed by Liu et al., that this elimination is compensated for by interactions between Asp35 and Pro79 and Gly48 and interactions between Lys57 with Val77 [34].

1.4.3 Induced-Fit Docking

Conventional rigid body docking assumes that the receptor is rigid when in reality receptors often alter their conformation to bind the ligand. Induced fit docking considers all the possible binding modes and associated conformational changes that occur between the ligand and protein upon ligand binding. This would more accurately predict the binding of a ligand to a receptor as it more closely mimics what would be occurring in vivo. The docking results (Table 1) show that the drugs LPV, ATV and DRV do all bind to L38↑N↑L. The Glide energy is an estimation of the free energy of binding and is composed of a combination of the posed ligand and receptor [35]. L38↑N↑L had a Glide energy score of -84.988 kcal/mol for LPV compared to -87.513 kcal/mol for wild-type. This difference, however, is not considered to be significant as it only differs by 3 kcal/mol. L38↑N↑L shows reduced binding free energy to ATV and DRV and this could be due to loss of hydrophobic contacts as seen in Fig. 7.

The majority of the amino acids forming the substrate binding site are hydrophobic with the exception of Asp25 and Asp29 which form hydrogen bonds with the peptide main chain groups as well as Arg8, Asp30 and Lys45 which are able to interact with polar side chains [15]. Due to the active site being mainly hydrophobic the HIV-1 protease inhibitors are also mainly hydrophobic. A loss in hydrophobic contacts is seen in many drug resistant mutants [15] and so the reduction of

hydrophobic contacts seen with L38↑N↑L and LPV and DRV could be indicative of the beginning of reduced drug susceptibility but does not indicate drug resistance. The L38↑N↑L variant does not contain any major drug resistant mutations as confirmed by the Stanford HIV database [36–38]. The variant, however, does contain the secondary mutation V82I [39]. The docking results show that all the drugs bind to the variant protease and so this minor mutation has no impact. This is not unusual because secondary mutations by themselves have very little effect on drug resistance and are selected for to compensate for viral fitness [40]. The ability of the variant protease to bind HIV-1 protease inhibitors is unaffected by the insertion of amino acids at position 38. The L38↑N↑L protease remains susceptible to the inhibitors and will be inhibited. We will discuss, in a forthcoming manuscript, how drug binding of the variant protease is affected by the presence of the Gag and Gag-Pol regions.

In summary, the L38↑N↑L protease shows reduced flap dynamics when compared to the wild-type subtype C protease. This is due to the existence of Asp30-Lys47, Asp62-Lys45 and Asp35-Arg20 salt-bridges, which stabilise the closed flap postion in L38↑N↑L. We propose that the reduced flap dynamics is a possible mechanism in which this enzyme can evade binding to HIV-1 protease inhibitors. LPV, ATV and DRV were successfully docked to L38↑N↑L. The negative Glide scores indicated that binding occurs. There were, however, reduced hydrophobic contacts made to LPV and DRV, which could be indicative of the beginning of reduced drug susceptibility. The insertion of the amino acids in the hinge region do not appear to affect drug binding but as a result of these insertions multiple salt-bridges form which hinder the movement of the flap region resulting in reduced flap dynamics of this variant protease. Further studies, to be investigated in our laboratory, include crystallising this variant to determine if it resembles the homology model. As well as crystallising the variant in the presence the HIV-1 protease inhibitors.

Acknowledgments

The South African Medical Research Council under a Self-Initiated Research Grant to Yasien Sayed supported this research. The research was also supported by a grant to Heini Dirr via the DST/NRF South African Research Chair Initiative Programme. The views and opinions expressed are those of the authors and do not necessarily represent the official views of the SA MRC. The authors would like to thank the National Research Foundation and Professor Lynn Morris (Head: HIV Research, National Institute for Communicable Diseases, South Africa).

References

[1] UNAIDS. UNAIDS Data 2017. UNAIDS, Geneva, Switz. [Internet]. 2017

[2] Robertson DL, Anderson JP, Bradac J a, et al. HIV-1 nomenclature proposal. Science. 2000;288:55–56.

[3] World Health Organization (WHO). Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. WHO Press. Geneva, Switz. 2016;

[4] McCutchan FE. Global Epidemiology of HIV. J. Med. Virol. 2006;78:7–12.

[5] Kohl NE, Emini EA, Schleif WA, et al. Active human immunodeficiency virus protease is required for viral infectivity. Proc. Natl. Acad. Sci. U. S. A. 1988;85:4686–4690.

[6] Wlodawer A, Erickson J. Structure-based Inhibitors of HIV-1 Protease. Annu. Rev. B. 1993;62:543–585.

[7] Naicker P, Seele P, Dirr HW, et al. F99 is critical for dimerization and activation of South African HIV-1 subtype C protease. Protein J. 2013;99:560–567.

[8] Wlodawer A, Miller M, Jasklski M, et al. Conserved Folding in Retroviral Proteases: Crystal Structure of a Synthetic HIV-1 Protease. Science (80-.). 1989;245:616–621.

[9] Gustchina A, Weber IT. Comparison of inhibitor binding in HIV-1 protease and in non-viral aspartic proteases: the role of the flap. FEBS Lett. 1990;269:269–272.

[10] Collins JR, Burt SK, Erickson JW. Flap opening in HIV-1 protease simulated by “activated” molecular dynamics. Nat. Struct. Biol. 1995;2:334–338.

[11] Ishima R, Freedberg DI, Wang YX, et al. Flap opening and dimer-interface flexibility in the free and inhibitor- bound HIV protease, and their implications for function. Structure. 1999;7:1047–1055.

[12] Naicker P, Achilonu I, Fanucchi S, et al. Structural insights into the South African HIV-1 subtype C protease : impact of hinge region dynamics and flap flexibility in drug resistance. J. Biomol. Struct. Dyn. 2013;37–41.

[13] Scott WR, Schiffer C a. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. Structure. 2000;8:1259–1265.

[14] Mosebi S, Morris L, Dirr HW, et al. 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. 2008;82:11476–11479.

[15] Weber IT, Agniswamy J. HIV-1 Protease: Structural Perspectives on Drug Resistance. Viruses. 2009;1:1110–1136.

[16] Veázquez-Campoy A, Vega S, Freire E, et al. Amplification of the effects of drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes. Biochemistry. 2002;41:8613–8619.

[17] Tramuto F, Bonura F, Mancuso S, et al. Detection of a new 3-base pair insertion mutation in the protease gene of human immunodeficiency virus type 1 during highly active antiretroviral therapy (HAART). AIDS Res Hum Retroviruses. 2005;21:420–423.

[18] Kozísek M, Sasková KG, Rezácová P, et al. Ninety-nine is not enough: molecular characterization of inhibitor-resistant human immunodeficiency virus type 1 protease mutants with insertions in the flap region. J. Virol. 2008;82:5869–5878.

[19] Pereira-Vaz J, Duque V, Trindade L, et al. Detection of the protease codon 35 amino acid insertion in sequences from treatment-naïve HIV-1 subtype C infected individuals in the Central Region of Portugal. J. Clin. Virol. 2009;46:169–172.

[20] Laskowski RA, MacArthur MW, Moss DS, et al. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 1993;26:283–291.

[21] Laskowski R, Rullmann JA, MacArthur M, et al. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR. 1996;8:477–486.

[22] Abraham MJ, Murtola T, Schulz R, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1–2:19–25.

[23] Van Der Spoel D, Lindahl E, Hess B, et al. GROMACS: Fast, flexible, and free. J. Comput. Chem. 2005;26:1701–1718.

[24] Hornak V, Okur A, Rizzo RC, et al. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. Proc. Natl. Acad. Sci. 2006;103:915–920.

[25] Lange OF, Van Der Spoel D, De Groot BL. Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR Data. Biophys. J. 2010;99:647–655.

[26] Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 1993;98:10089–10092.

[27] Okumura H, Itoh SG, Okamoto Y. Explicit Symplectic Integrators of Molecular Dynamics Algorithms for Rigid-Body Molecules in the Canonical, Isothermal-Isobaric, and Related Ensembles. J. Chem. Phys. 2007;126:84103.

[28] Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J. Mol. Graph. 1996. p. 33–38.

[29] Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. 2004;25:1605–1612.

[30] Laskowski RA, Swindells MB. LigPlot + : Multiple Ligand À Protein Interaction Diagrams for Drug Discovery. J. Chem. Inf. Model. 2011;51:2778–2786.

[31] Tözsér J, Yin FH, Cheng YS, et al. Activity of tethered human immunodeficiency virus 1 protease containing mutations in the flap region of one subunit. Eur. J. Biochem. 1997;244:235–241.

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

[33] Zondagh J, Balakrishnan V, Achilonu I, et al. The N37T ↑ V hinge region insertion mutation alters the Molecular dynamic landscape of HIV - 1 Subtype C protease Molecular Dynamics and Ligand Docking of a Hinge Region Variant of South African HIV - 1 Subtype - C protease. J. Mol. Graph. Model. 2018;82:1–11.

[34] Liu Z, Huang X, Hu L, et al. Effects of hinge-region natural polymorphisms on human immunodeficiency virus-type 1 protease structure, dynamics, and drug pressure evolution. J. Biol. Chem. 2016;291:22741–22756.

[35] Friesner RA, Banks JL, Murphy RB, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J. Med. Chem. 2004;47:1739–1749.

[36] Rhee S-Y, Kantor R, Katzemstein DA, et al. HIV-1 pol mutation frequence by subtype and treatment experience: extension of the HIVseq program to seven non-B subtypes. AIDS. 2006;20:643–651.

[37] Rhee S-Y, Gonzales M., Kantor R, et al. Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res. 2003;31:298–303.

[38] Shafer RW, Jung DR, Betts BJ. Human immunodeficiency virus type 1 reverse transcriptase and protease mutation search engine for queries. Nat. Med. 2000;6:1290–1292.

[39] Wensing AM, Calvez V, Gunthard HF, et al. 2017 Update of the Drug Resistance Mutations in HIV-1. Top. Antivir. Med. 2017;24:132–133.

[40] Ammaranond P, Cunningham P, Oelrichs R, et al. Rates of transmission of antiretroviral drug resistant strains of HIV-1. J. Clin. Virol. 2003;26:153–161.

CHAPTER 4
DRUG SUSCEPTIBILITY

Double trouble? Gag in combination with double insert in HIV-1 protease contributes to reduced DRV susceptibility

Alison Williams, Adriaan Basson, Ikechukwu Achilonu, Heini Dirr, Lynn Morris and Yasien Sayed

Manuscript to be submitted

In this publication the drug susceptibility of L38↑N↑L protease to LPV, ATV and DRV, is assessed in the presence and absence of a Gag sequence using a phenotypic assay and an enzyme assay. L38↑N↑L was less catalytically efficient and with its associated Gag displayed reduced replicative capacity. Phenotypic assays showed L38↑N↑L had reduced DRV susceptibility.

Author contributions: Alison Williams performed all the experimental work, analysed the data and prepared the manuscript. Adriaan Basson assisted in analysis of phenotypic data and manuscript revision. Ikechukwu Achilonu assisted in experimental design and manuscript revision. Heini Dirr assisted in manuscript revision. Lynn Morris assisted in analysis of phenotypic data and manuscript revision. Yasien Sayed supervised the project and assisted in data analysis and interpretation.

Double trouble? Gag in combination with double insert in HIV-1 protease contributes to reduced DRV susceptibility

A. Williams[1], A. Basson[2,3], I. Achilonu, H. W. Dirr[1], L. Morris[2] and Y. Sayed[1]

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

[2] Centre for HIV and STIs, National Institute for Communicable Diseases (NICD), Johannesburg, South Africa.

[3]HIV Pathogenesis Research Unit, Department of Molecular Medicine and Haematology, Faculty of Health Sciences, University of the Witwatersrand, 7 York Road, Parktown, 2193, Johannesburg, South Africa

Abstract

The use of combined antiretroviral therapy has made a significant contribution to increasing the life expectancy and quality of life in people living with HIV. HIV-1 protease is essential for processing the Gag polyprotein to produce infectious virions and is a major target in antiretroviral therapy. However, due to the error prone reverse transcriptase, resistance-associated mutations can accumulate in the HIV-1 protease. We have identified an unusual HIV-1 subtype C variant protease that contains insertions of leucine and asparagine (L38↑N↑L) in the hinge region of protease at position 38. Isothermal titration calorimetry showed that the L38↑N↑L protease had diminished activity (80%) compared to wild-type subtype C protease (90%), with ±50% reduction in K_M and k_{cat}. There was a 42% reduction of specific activity for the variant using the substrate Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2. Although the V_{max} of L38↑N↑L and wild-type were similar, the variant showed a 10-fold reduction in catalytic efficiency (k_{cat}/K_M). An *in vitro* phenotypic drug susceptibility assay to determine the phenotypic consequences of amino acid changes in Gag and HIV-1 protease was conducted. It showed the L38↑N↑L protease to be susceptible to lopinavir, atazanavir and darunavir in the presence of a wild-type subtype C Gag. However, in the presence of the related Gag, L38↑N↑L showed reduced susceptibility to darunavir while remaining susceptible to lopinavir and atazanavir. Furthermore, a reduction in viral replication capacity was observed in combination with the related Gag. The reduced susceptibility to darunavir and decrease replication capacity may be due a duplication of the proline rich domain within p6 that is responsible for recruiting Tsg101, PTAPP, in the related Gag.

Introduction

HIV remains a global health problem with 36.7 million people living with HIV globally [1]. HIV is particularly a problem in Africa as 52% of people living with HIV are from southern Africa and 45% of all new infections occur in this region [1]. It was estimated that 13% of the South African population in 2016 was living with HIV. However, there have been great improvements in treatment in Africa, particularly South Africa. In 2015 South Africa had 3.4 million people on treatment, more than any other country. There are, however, still populations within South Africa that are at a high risk of HIV infection; these include sex workers, people who inject drugs, transgender people, prisoners, gay men and men who have sex with men. In 2015 these populations accounted for 20% of all new infections in sub-Saharan Africa [1].

HIV is divided into two types HIV-1 and HIV-2 with HIV-1 being the main type. HIV-1 is divided into groups M, N, O and P. Group M is the main group and is further divided into subtypes A, B, C, D, F, G, H, J, and K [2]. Subtype C is found in sub-Saharan Africa, India, Brazil and China [2] and accounts for approximately 50% of global infections [3]. The variant in this study is of subtype C origin. The great diversity among HIV is attributed to the high replication rate as well as the low fidelity of reverse transcriptase [4].

HIV-1 protease, a homodimeric aspartyl protease, cleaves the Gag and Gag-Pol polyproteins to produce three enzymes (reverse transcriptase, integrase and protease) and the structural proteins (capsid, matrix, nucleocapsid, p6, gp120 and gp41) needed for capsid assembly [5]. The secondary structure of HIV-1 protease consists mainly of β-sheets and one α-helix per monomer. The active site contains the characteristic Asp-Gly-Thr sequence of an aspartyl protease [6]. The structure of HIV-1 protease and the position of the Asp25 residues are shown in Figure 1. The hydrophobic active site is covered by two β-turns, termed the flap region, which open up to allow substrate to bind to the active site and then close upon substrate binding. The movement of the flaps is aided by the hinge region (residues 32-42) of the protein, a region known to contain several polymorphisms in subtype C.

HIV-1 protease plays a critical role in viral replication since failure to cleave the Gag and Gag-Pol polyproteins results in immature virions that are non-infectious [7]. This vital step is one of the drug targets of second-line antiretroviral therapy (ART) that makes use of a ritonavir (RTV)-boosted HIV-1 protease inhibitors (PIs) backbone of lopinavir (LPV), atazanavir (ATV) or darunavir (DRV), in addition to two nucleot(s)ide inhibitors (NRTIs) [8]. HIV is notorious for the development of drug resistance

to the antiretrovirals (ARVs). Resistance to PIs is no exception. Resistance to PIs is a result of amino acid substitutions in the substrate-binding sites as well as distal sites [9]. These substitutions directly or indirectly alter the interactions between the inhibitors and the HIV-1 protease, which reduces the affinity for the enzyme [10]. Rarely, amino acid insertions in the HIV-1 protease gene are selected for during antiretroviral therapy. Insertion mutations are not unusual in reverse transcriptase [11] but are rare in HIV-1 protease (approximately 0.1%) [12]. Most insertions arise due to duplications of neighbouring DNA sequences due to primer/template slippage during reverse transcription. Resistance mutations in HIV-1 protease have been well characterised [13]. However, little is known about the effect of amino acid insertions in the hinge region on drug binding.

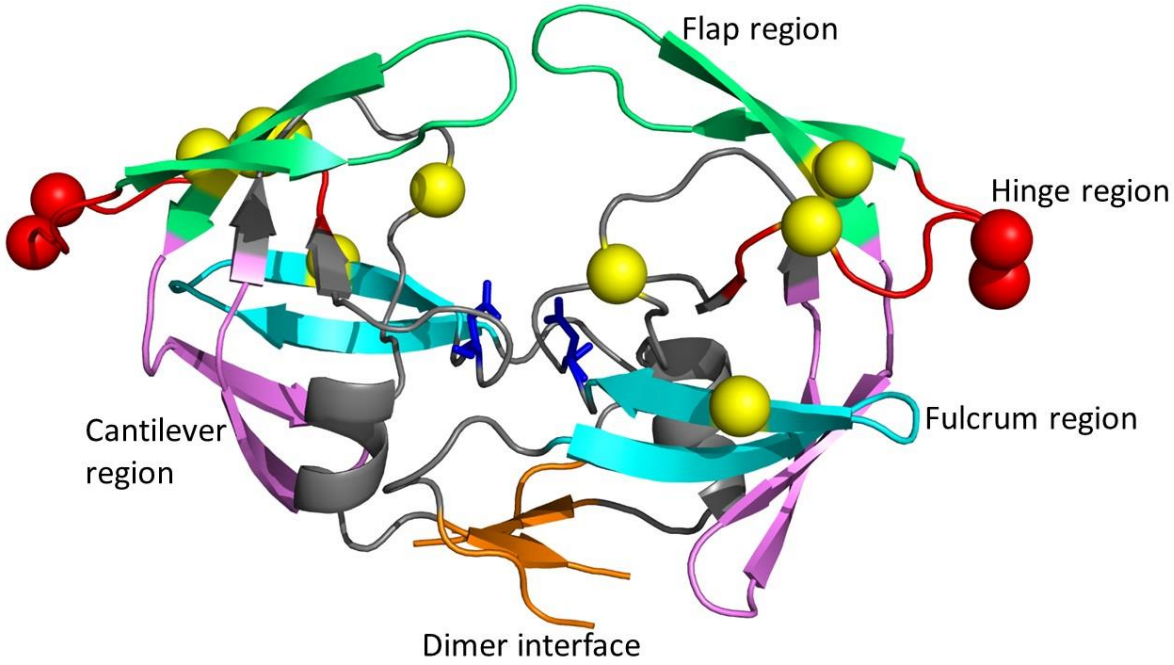
The HIV-1 protease variant in this study (Figure 1) was isolated from an infant whose mother was part of the Prevention of Mother to Child Transmission programme in South Africa [14] and was HIV-1 protease inhibitor naïve. This sequence is a subtype C sequence and the variant contains the following subset of polymorphisms; K20R, E35D, R57K and V82I as well as a double insertion of leucine and asparagine at position 38. The variant will be referred to as L38↑N↑L - the upward arrows indicate the insertion of two amino acids at position 38. The present study shows a comparison of the drug susceptibility of the L38↑N↑L protease with its accompanying Gag sequence and the wild type protease with Gag to LPV, ATV and DRV.

Methods

Expression and Purification

The wild-type subtype C and L38↑N↑L protease sequence data were obtained from Prof Lynn Morris (AIDS Virus research Unit, NICD of Johannesburg, South Africa). The pET-11b plasmid encoding the HIV-1 wild-type subtype C protease was previously generated in our laboratory [15]. The pET-11a plasmid containing the L38↑N↑L gene was purchased from GenScript (HongKong). The plasmids containing the wild-type and L38↑N↑L sequence were used to transform *Escherichia coli* BL21 (DE3) pLysS and *Escherichia coli* Rosetta^{TM} cells, respectively. The proteases were overexpressed by the addition of 1 mM IPTG at mid-log phase (A_{600nm} = 0.6). Overexpression was allowed to continue for 4 hours for wild-type and 6 hours for L38↑N↑L at 37 ℃. The cells were pelleted by centrifugation at 5 000xg and resuspended in Buffer A (10 mM Tris, 5 mM EDTA, pH 8). The bacterial cell membranes were disrupted by sonication, 12 V for 10x 30 s intervals. The cell lysates were centrifuged at 23 000xg for 40 min to separate the soluble and insoluble fractions. The

insoluble fractions were resuspended in buffer A+ (10 mM Tris, 2 mM EDTA and 2% Triton X-100 at pH 8). This was centrifuged at 23 000xg




```
Wild-type       PQITLWKRPLVSIKVGGQIKEALLDTGADDTVLEEI--NLPGKWKPKMIGGIGGFIKVRQ
L38↑N↑L         PQITLWKRPLVSIKVGGQIREALLDTGADDTVLEDINLNLPGKWKPKMIGGIGGFIKVKQ
                ********************:**************:*  **********************:*

Wild-type       YDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF
L38↑N↑L         YDQILIEICGKKAIGTVLVGPTPINIIGRNMLTQLGCTLNF
                ************************:****************
```

Figure 1. Homology model of L38↑N↑L protease and sequence data.

HIV-1 protease is a homodimer and contains mainly β-sheets and one α-helix per monomer. There are five regions defined within the structure: the flap region (green), the hinge region (red), the fulcrum region (cyan), the cantilever region (violet) and the dimer interface (orange). The catalytic Asp25 residues are shown in the active site (blue). The yellow spheres and the yellow boxes on the sequence represent the relative positions of the subset of mutations, K20R, E35D, R57K and V82I. The red spheres on the structure and the red box in the sequence alignment represent the double insertion of Leu and Asn. PyMOL was used to generate the homology model using data from the Protein Data Bank (PDB ID: 3U71). The sequence alignment was generated using the Clustal Omega tool (EMBL-EBI).

for 30 min and the process was repeated. The proteases were recovered from inclusion bodies by re-suspending the pellet in buffer B (8 M urea, 10 mM Tris, 2 mM DTT, pH 8) and incubating at 20 ℃ for an hour. The unfolded proteases were centrifuged at 23 000xg and refolded by dialysing against 10 mM formic acid containing 10% glycerol (v/v) for 4 hours at 4 ℃. The proteases were then dialysed against Buffer D (10 mM sodium acetate, 2 mM DTT pH 5) overnight and purified using a CM-Sepharose column. Elution was performed using a salt gradient of 0-1 M NaCl. Finally, the protease was dialysed against 10 mM sodium acetate (pH 5), and the purity assessed using a 16% Tricine gel [16]. The concentration of the proteases was determined using the molar extinction co-efficient 25480 $M^{-1}.cm^{-1}$ from absorbance spectra obtained on a Jasco V-630 spectrophotometer. The purified proteases were aliquoted and store at -80°C until used.

Active site titration

To assess the percentage activity of the proteases, isothermal titration calorimetry (ITC) was performed. Acetyl pepstatin (200 µM), an inhibitor of aspartyl proteases, was titrated against 17 µM wild-type protease and 18 µM L38↑N↑L protease in 10 µL injections at 293.8 K using a Nano-ITC instrument (TA Instruments, Delaware USA). The heat due to the dilution of acetyl pepstatin was subtracted from the data set and the baseline adjusted using NITPIC [17]. The changes in heats were integrated and fitted using ITCsy [17]. The percentage of active sites was determined from the stoichiometry value, with a value of 1 indicating activity of 100%.

Steady-state and Inhibition Kinetics

The kinetic parameters K_M, k_{cat}, k_{cat}/K_M and the specific activity were determined in separate experiments. K_M and V_{max} were determined using a hyperbolic plot, linear plots were used to determine k_{cat}, k_{cat}/K_M and the specific activity. All plots were fitted using SigmaPlot (Systat Software, San Jose, CA). The hydrolysis of the HIV-1 protease fluorogenic substrate Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2 was monitored. The substrate used represents a conserved cleavage site capsid/p2 within Gag. For all kinetic measurements, an excitation wavelength of 337 nm and an emission wavelength of 425 nm were used at 1 min measurement intervals during steady-state. All activity assays were performed in 50 mM sodium acetate buffer containing 1 M sodium chloride (pH 5.0) at 20 ℃. All kinetic experiments were performed in triplicate using a Jasco FP-6300 spectrofluorometer (Easton, MD, USA).

An active enzyme concentration of 50 nM and substrate concentrations ranging from 5-200 µM was used to determine the K_M. To determine the k_{cat}/K_M, a substrate concentration ranging from 1-10 µM

was used and an active enzyme concentration of 50 nM. The specific activity and the k_{cat} were determined using active enzyme amounts ranging from 1-10 pmol and a constant substrate concentration of 50 µM. The FDA-approved drugs (LPV, ATV and DRV) competitively inhibit HIV-1 protease with dissociation constants in the nanomolar range [18]. The inhibition constants (K_i) for LPV, ATV and DRV were determined using the equation for tight-binding inhibitors [19]:

$$= \frac{\left(IC_{50} - \frac{[E]}{2}\right)}{\left(\frac{[S]}{K_M} + 1\right)}$$

E - active enzyme concentration (50 nM),

S - substrate concentration (50 µM)

IC_{50} - concentration of inhibitor at which there is half-maximal activity of the HIV-1 protease. IC_{50} values were determined using inhibitor concentrations ranging from 0-200 µM. A final concentration of 2% (v/v) dimethylsulfoxide was used during IC_{50} determinations for inhibitor solubility.

Construction of expression vectors

The vector p8.9NSX containing the wild-type subtype C *Gag* and *protease* gene (p8.9MJ4GP) was digested using Not I (cleavage site upstream of *Gag*) and Xho I (Cleavage site downstream of HIV-1 protease gene). The reaction mixture was incubated at 37 ℃ for two hours and heat inactivated at 65 ℃ for 15 minutes. Successful restriction was confirmed by 0.7% agarose gel electrophoresis at 100 V for 1.5 hours. An uncut control plasmid was included. The restricted p8.9MJ4GP was purified from the agarose gel using PureLink Quick Gel Extraction Kit (Life Technologies, Germany) according to the manufacturer's protocol. The *Gag* and HIV-1 *protease* gene to be inserted into p8.9MJ4GP was previously isolated from RNA and reverse transcribed into cDNA. The insert was ligated into p8.9MJ4GP and the vector dephosphorylated using the Rapid DNA Dephos and Ligation Kit (Roche, Basel, Switzerland). This resulted in a vector encoding mutated Gag and L38↑N↑L protease. The resultant pseudovirion was termed L38NL. A schematic of this process is shown in Figure 2. The vector encoding wild-type subtype C Gag was created by restricting p8.9MJ4GP with Apa I (upstream of protease) and Xho I. The L38↑N↑L gene was ligated into the restricted p8.9MJ4GP vector using the Rapid DNA Dephos and Ligation Kit (Roche, Basel, Switzerland). This resulted in a vector encoding wild-type Gag and L38↑N↑L (Figure 2). The resultant pseudovirion was termed WTGAG L38NL.

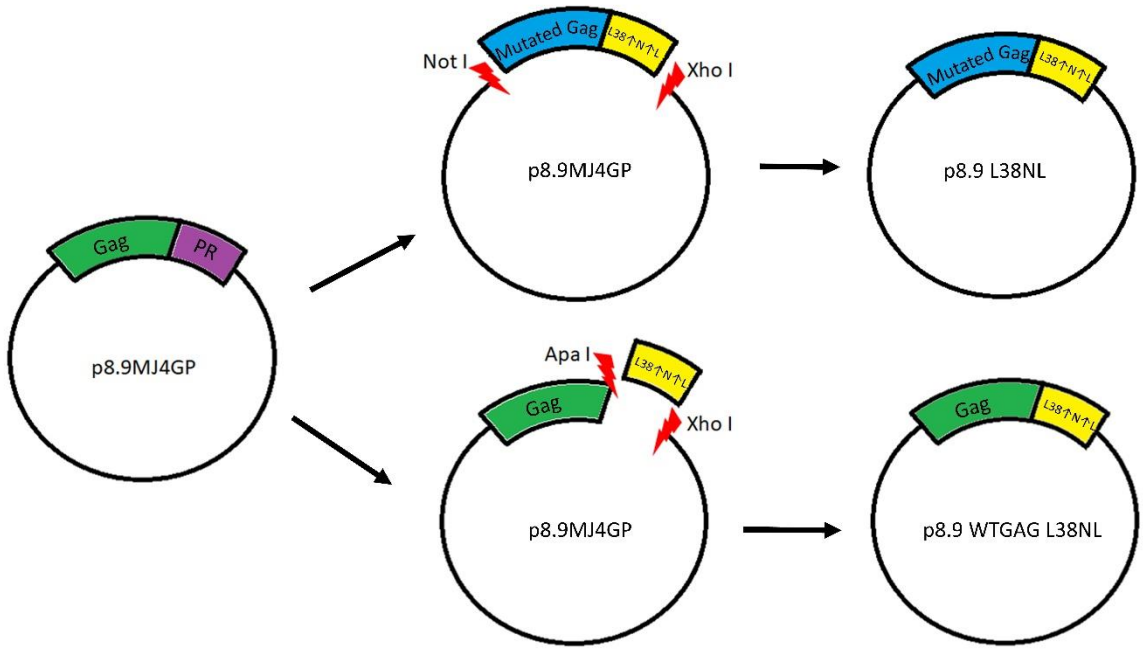


Figure 2: Cloning of the *gag* and *protease* genes into the p8.9MJ4GP expression vector. Restriction enzymes (Not I and Xho I) were used to digest p8.9MJ4GP and allow for cloning of the mutated *Gag* and *L38↑N↑L* gene resulting in the p8.9 L38NL expression vector. Restriction enzymes (Apa I and Xho I) were used to digest p8.9MJ4GP and allow for cloning of the *L38↑N↑L* gene resulting in the p8.9 WTGAG L38NL expression vector.

Phenotypic assessment of protease inhibitor susceptibility

A phenotypic assay using a single infection event per virion was conducted as previously described [20]. Briefly, HEK293T cells were transfected with 300 ng of plasmid pMDG (encodes for vesicular stomatitis virus G protein for entry), 500 ng of plasmid pCSFLW (encodes for firefly luciferase for quantification) and 300 ng of the HIV expression vector (encodes HIV-1 gag-pol: p8.9MJ4GP, p8.9 L38NL, p8.9 WTGAG L38NL) using 3.3 µg polyethylenimine (PEI, Polysciences, Inc., Warrington, PA, USA). The transfected cells were harvested after 18 hours and seeded in the presence of serially diluted LPV (60 nM – 3 pM), ATV (40 nM – 2 pM) and DRV (60 nM – 3 pM). The supernatants were transferred to corresponding 96-well culture plates that contained fresh HEK293T cells after 24 hours. The degree of infection was determined 48 hours later by measuring the expression of firefly luciferase using a BrightGlo luciferase assay system (Promega, Madison, USA) on the Victor^3 multi-label plate reader (PerkinElmer, Waltham, USA). For each drug-pseudovirus combination, the IC_{50} was calculated using Microsoft Excel (Microsoft, Redmond, USA). A pseudovirus is only able to infect in a single round. The phenotypic susceptibility was expressed as the fold change in the IC_{50} relative to that of the wild-type pseudovirus (wild-type subtype C Gag and protease). The assay specific fold change cut-off value for each drug was determined using the 99th percentile of the average IC_{50} for the wild-type pseudovirus assessed in multiple repeat screens for each drug. Two-way analysis of variance (ANOVA) and Bonferroni's post-test were used to identify significant differences in IC_{50} values between wild-type (wild-type subtype C Gag and protease), L38NL (mutated Gag and L38↑N↑L protease) and WTGAG L38NL (wild-type subtype C Gag and L38↑N↑L protease) pseudoviruses. The p8.MJ4GP wild-type pseudovirus contained the gag-pol of an HIV-1 subtype C reference isolate while the p8.L38NL contained the gag-protease from L38↑N↑L. The WTGAG L38NL pseudovirus contained the MJ4 gag-L38↑N↑L protease. The drug resistant control consisted of an insert of a Gag and protease sequence found within a patient failing therapy and contained many drug resistance mutations (all protease mutations present: L10I, K20R, E35D, M46I, I54V,Q61H, I62V, L63P, T74S and V82A).

Replication capacity

To assess the replication capacity (RC) of each of the pseudovirions, a p24 enzyme-linked immunosorbent assay (ELISA) was used according to Aalto Bioreagents (Dublin, Ireland). A 96 well plate was coated with p24 antibody and incubated overnight. The plate was blocked with 2% BSA for 1 hour and washed with TBS. Cells containing pseudovirions generated in the absence of HIV-1

protease inhibitors were disrupted using 2% Empigen in TBS. Supernatants were submitted in triplicate for p24 ELISA and incubated for 3 hours. The plate was washed three times with TBS and the conjugate (EH12AP, 20% sheep serum, 0.05% Tween) was applied. The plate was washed with TROPIX buffer and TROPIX Sapphire Substrate was added. The plate was incubated for 45 minutes and the ratio of relative light units (RLU)-to-p24 was calculated. The percent replication capacity (%RC) was calculated relative to the wild-type control.

Sequencing of Gag and Protease genes

Sequencing was performed using BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, California, USA). The primers used for the HIV-1 *gag* and *protease* genes are shown in Table 1 and Figure 3. The sequence reaction was plated in a MicroAmp Optical 96-well Reaction Plate (Applied Biosystems, China) sealed with Microseal plate covers (Bio-Rad, UK). The reaction was performed using a GeneAmp 9700 thermal cycler (Applied Biosystems, Carlsbad, CA). The following programme was used, initial denaturation for 1 minute at 96 ℃, 35 cycles of 10 s at 96 ℃ for denaturation, 5 s at 50 ℃ for annealing, 4 minutes at 60 ℃ for extension and a final hold at 4 ℃.

Results

Expression and Purification

The wild-type subtype C and the L38↑N↑L proteases were successfully purified from the insoluble fraction as shown in Figure 4. Both proteases were purified to >95% purity and a yield 2.7 mg per 28 g of wet cell weight of *E.coli* was obtained for L38↑N↑L protease.

Steady-state and Inhibition Kinetics

HIV-1 protease has the ability to autolyse. The peptides produced contribute towards the concentration determined spectroscopically but are not catalytically active. Therefore, to determine the percentage activity of each of the proteases, isothermal titration calorimetery (ITC) was conducted. Acetyl pepstatin was titrated against each protease. The stoichiometry of binding was used to determine the percentage active sites. Since a binding ratio of 1:1 is expected, a stoichiometry of 1 would represent a sample that is 100% active. The percentage of active protein in each sample preparation was 90% and 80% for wild-type and L38↑N↑L, respectively. This percentage was used to adjust the concentrations obtained spectroscopically.

Table 1: Primers used to sequence the *gag* and *protease* genes

Primer name	Sequence (5'-3')
GagNot	GCGGCGGCCGCAAGGAGAGAGATGGGTGCG
Gag1F	TTAGACAAGATAGAGGAAGA
Gag1.5F	TCTATCCCATTCTGCAGC
Gag2F	ATGATGACAGCATGTCAGGG
DuGagIn2	ACATGGGTATTAGCTCTGGGCT
ProXhoR2	CTGGTACAGTCTCGAGRGGACTRATKGG
Gag5F	CTTTAAGAGCTGAACAAGCT
Gag6R	AAAATGGTCTTACAATCTGG

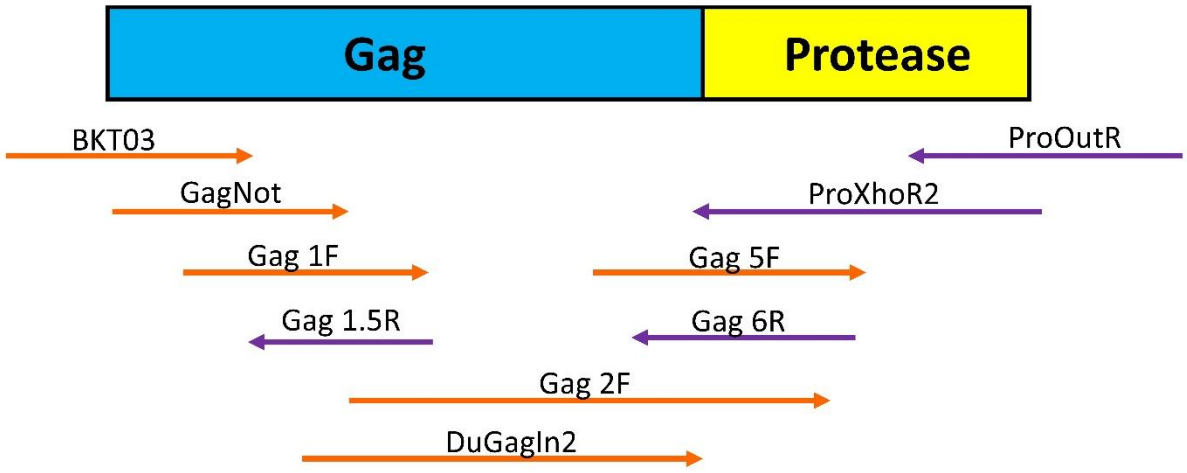


Figure 3: Diagram of positions of the primers used for sequencing. Forward primers shown in orange and reverse primes shown in purple

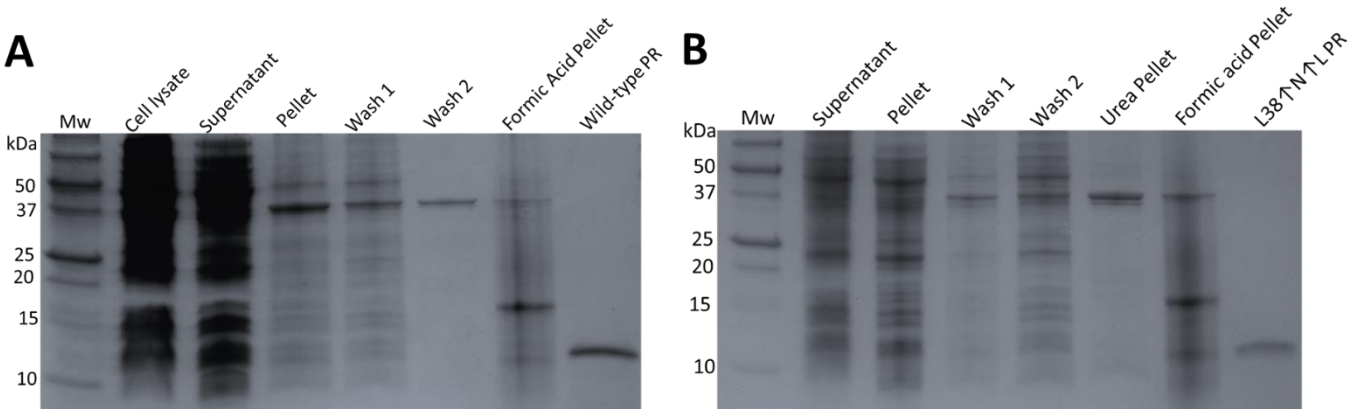


Figure 4: SDS PAGE showing the purification profile of (A) wild-type protease and (B) L38↑N↑L protease. The molecular weight is shown in the first lane of each gel. The purification steps, from cell lysis to final product, are shown from left to right.

Using a fluorogenic substrate (Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2) that mimics the capsid/p2 cleavage site, the steady-state kinetics of each protease was determined (Table 2). The K_M, specific activity and k_{cat} for L38↑N↑L was approximately half that of the wild-type while the V_{max} was similar between the two enzymes. The catalytic efficiency (k_{cat}/K_M) of the L38↑N↑L protease is 12-fold less than that of the wild-type protease. The L38↑N↑L specific activity is approximately 2-fold lower than that of the wild-type. IC_{50} values were used to calculate the inhibition constants for each drug (Table 3). The K_i for L38↑N↑L for LPV was 10-fold less than the wild-type protease and 3-fold less for DRV. The K_i for ATV was, however, three-fold higher for L38↑N↑L than for the wild-type protease.

Phenotypic Susceptibility and Replication Capacity

A phenotypic viral assay was conducted to determine the susceptibility of L38↑N↑L to protease inhibitors in the presence of a Gag sequence. It was found that the L38↑N↑L protease was susceptible to LPV (L38NL: 4.0 ± 0.8 nM and WTGAG L38NL: 4.4 ± 1.2 nM) and ATV (L38NL: 4.2 ± 0.7 nM and WTGAG L38NL: 4.4 ± 0.4 nM) with either the mutated or wild-type subtype C Gag sequences. This was due to the IC_{50} not being 2.9-fold greater than the wild-type (LPV and ATV: 1.4 ± 0.5 nM) (Figure 5). The L38↑N↑L protease showed reduced susceptibility to DRV with both wild-type and mutated Gag (Figure 5). In the presence of the mutated Gag L38↑N↑L protease was less susceptible to DRV than when in the presence of the wild-type subtype C Gag. The DRV IC_{50} for L38NL (1.6 ± 0.2 nM) was 5-fold higher than wild-type (0.3 ± 0.05 nM), above the 1.3-fold cut off. WTGAG L38NL IC_{50} for DRV (1.0 ± 0.2 nM) was 3-fold higher than wild-type. The resistant control showed high-level resistance (FC>10) for all three PIs. An analysis of the RC (Figure 6) showed the L38NL pseudovirion to have a reduced RC (24%) while the WTGAG L38NL pseudovirion showed an increased RC (120%) when compared to the wild-type sample. The drug resistant control showed the lowest RC (13%).

Analysis of L38↑N↑L Gag

The accompanying Gag of L38↑N↑L was sequenced to determine if there were any drug resistance mutations present. The following mutations were found in the cleavage sites: T370A, M374V, R376G in the p2/NC cleavage site, E424G in the NC/p1 cleavage site and N448S in the p1/p6 cleavage site. A duplication of the PTAPP motif, was seen downstream of p1/p6 cleavage site. Multiple polymorphisms were also observed in the non-cleavage site regions as seen in Figure 7.

Table 2: The steady-state kinetic parameters determined for wild-type and L38↑N↑L protease

Parameter	Wild-type PR	L38↑N↑L
K_M (μM)	14 ±1.7	7 ±0.9
V_{max} (μmol.min^{-1})	0.01 ± 0.0003	0.01 ± 0.0003
Specific activity (μmol.min^{-1}.mg^{-1})	21.0 ± 1.1	12.1 ± 1.1
k_{cat} (s^{-1})	7.7±0.4	4.5 ± 0.4
k_{cat}/K_M (s^{-1}.μM^{-1})	12.2 ± 0.48	1.0 ± 0.004

Table 3: The K_i (nM) values obtained for wild-type and L38↑N↑L protease against the drugs LPV, ATV and DRV

Protease	LPV	ATV	DRV
Wild-type	2.1 ± 0.2	1.2 ± 0.1	1.4 ± 0.2
L38↑N↑L	0.2 ± 0.02	3.5 ± 0.7	0.4 ± 0.02

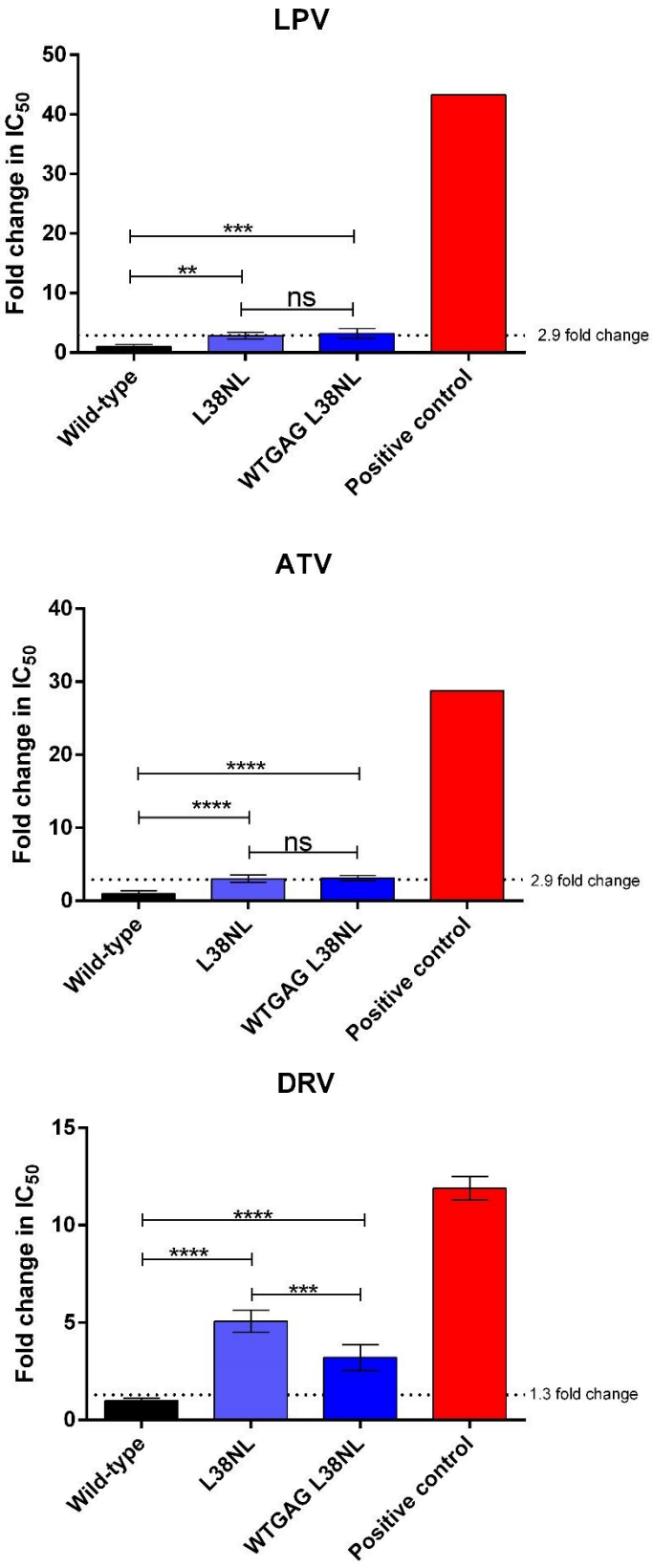


Figure 5: Phenotypic susceptibility of pseudoviruses to approved protease inhibitors.

The phenotypic susceptibility of the wild-type control, L38NL sample with the L38↑N↑L protease and the accompanying Gag sequence (L38NL), the L38↑N↑L protease with a wild-type subtype C Gag (WTGAG L38NL) and drug resistant control (protease mutations: L10I, K20R, E35D, M46I, I54V,Q61H, I62V, L63P, T74S and V82A) is shown. ***p<0.001, ****p<0.0001, ns = not significant.

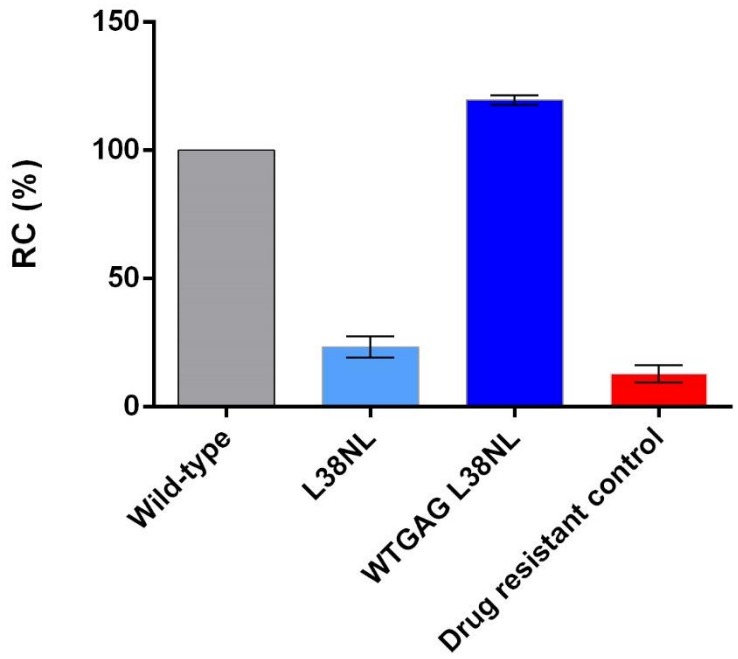


Figure 6: Relative replication capacity (RC) of the pseudovirions.

The relative replication capacity of each pseudovirus was compared to the wild-type. The L38NL pseudovirus showed a reduction in replication capacity of 76% while WTGAG L38NL showed an increase of 20%. The drug resistant control had a replication capacity of 13%.

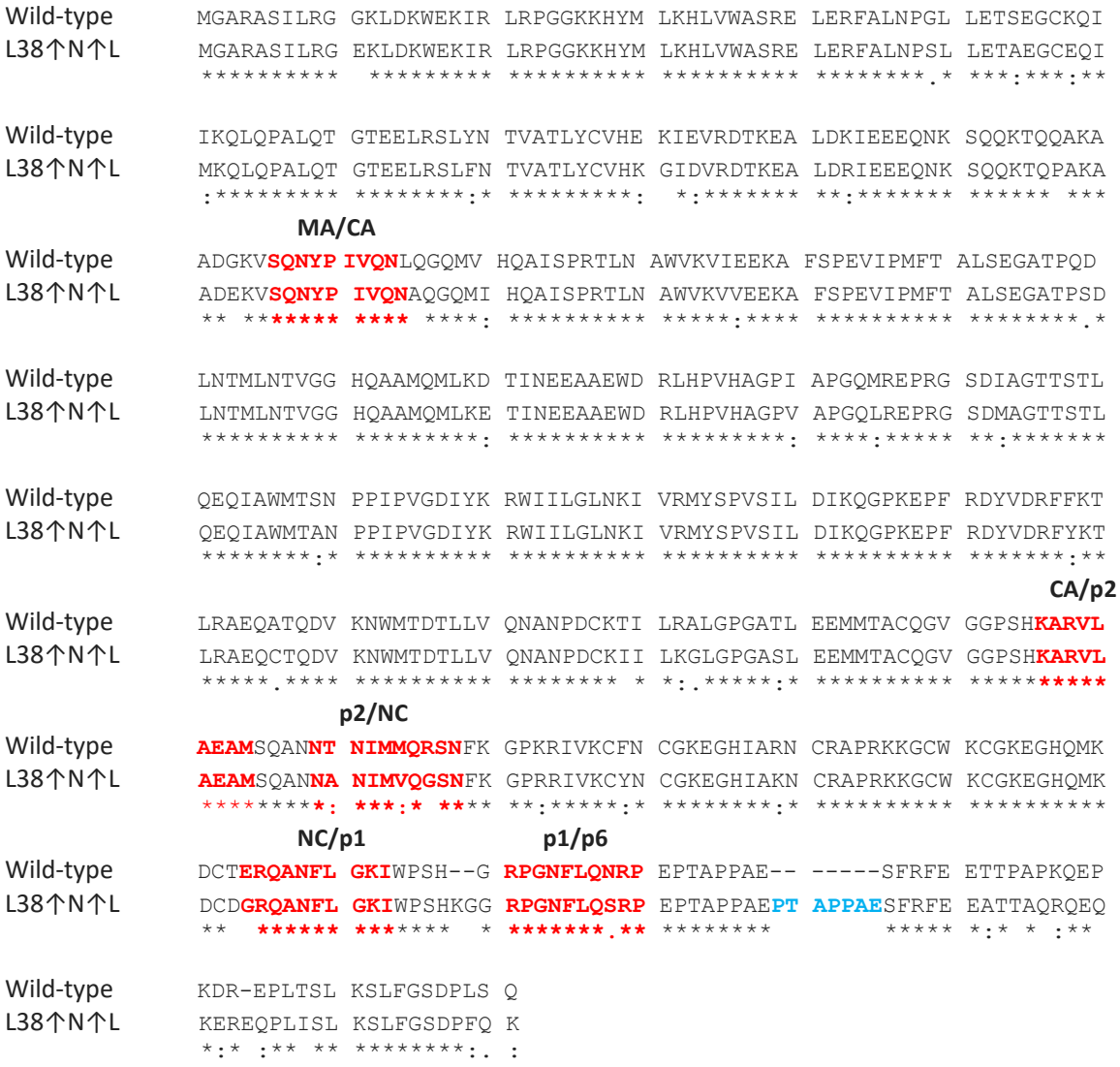


Figure 7: Alignment of Gag sequences of wild-type Subtype C and the L38↑N↑L variant. The cleavage sites are indicated in red and PTATPPAE insertion indicated in blue. The sequence alignment was generated using the Clustal Omega tool (EMBL-EBI).

Discussion

Use of HIV-1 protease inhibitors has greatly improved the response to ARV regimes. However, resistance to HIV-1 protease inhibitors is a problem. Mutations occur rapidly due to reverse transcriptase lacking proof reading capabilities. Mutations occurring in HIV-1 protease may be major (occurring at the active site), minor (occurring in regions distal to the active site) or may be insertions of 1 – 6 amino acids. Insertion mutations occur rarely and thus have not been well characterised, especially in subtype C viruses. Here we show the drug susceptibility of a subtype C protease containing a double insertion.

The L38↑N↑L and wild-type proteases were expressed and purified from inclusion bodies, with an 80% and 90% activity, respectively. An analysis of the K_M indicated that the L38↑N↑L variant had an increased affinity for the substrate mimicking the CA/p2 cleavage site. Although the maximum velocity (V_{max}) was approximately the same as that of the wild-type protease, the turnover number was approximately 2-fold less. This indicates that L38↑N↑L cleaves 2-fold less substrate per second than wild-type. The catalytic efficiency (k_{cat}/K_M) of L38↑N↑L was 10-fold less than that of wild-type and shows a smaller proportion of substrate being converted to product when compared to wild-type. This confirms the findings of Kozísek *et al*. (2008) [12], who showed that amino acid insertions in the hinge region of HIV-1 protease reduce the catalytic efficiency. A mechanistic reason for the reduced efficiency could be due to the flap region remaining closed for a longer period than the wild-type, as shown by molecular dynamics simulations studies (manuscript under review). This, in turn, would hinder substrate binding and result in decreased product formation. The reduced specific activity of the L38↑N↑L confirms that the catalytic efficacy of L38↑N↑L was reduced compared to wild-type. The affinities of LPV and DRV were increased for L38↑N↑L as displayed by the decrease in inhibitory constant (K_i). However, in the presence of ATV, a 3-fold decrease in affinity was observed for the L38↑N↑L, although this was not large enough to indicate resistance. This indicates that these HIV-1 protease inhibitors would inhibit variant protease without a Gag region present.

Resistance to PIs occurs with the accumulation of mutations in the HIV-1 protease, which may lead to a reduced replication capacity of the virus as the affinity for its natural substrate has decreased [9, 20, 21]. There are, however, mutations that occur in Gag that will partially restore the replicative capacity of the virus and thus Gag mutates to compensate for mutations in HIV-1 protease [20, 22]. These mutations can occur in cleavage sites or elsewhere [20]. Resistance to HIV-1 protease inhibitors is not only due to the HIV-1 protease itself but also due to amino acid substitutions in the

Gag cleavage sites [20, 23]. However, when patients are failing therapy, drug resistance mutations in Gag are typically not considered.

A phenotypic assay was conducted to determine the effect of the L38↑N↑L insertion mutation on the *in vitro* drug susceptibility to PIs. This was performed in conjunction with an unrelated wild-type control Gag, as well as the related Gag, to assess the phenotypic impact of Gag on drug susceptibility in the context of the L38↑N↑L insertion mutation. For LPV and ATV, no significant reduction in drug susceptibility was observed for either Gag-containing version of the L38↑N↑L protease. However, for DRV, both Gag-containing version of L38↑N↑L protease showed a small but significant decrease in DRV susceptibility relative to the wild-type. A reduced susceptibility to DRV for an insertion mutation at position 35 was also observed [12]. Furthermore, the decrease in DRV susceptibility was significantly greater for the related Gag-containing L38↑N↑L protease than for the unrelated Gag-containing L38↑N↑L protease, implicating Gag as a contributor towards drug resistance in the context of the L38↑N↑L insertion mutation. This underlines the importance of including the Gag region in PI resistance screening, as was also shown by Giandhari *et al*. [24].

Mutations in Gag cleavage sites have been linked to drug resistance [23] and several cleavage site mutations were observed in the related Gag sequence, including mutations in the p2/NC, NC/p1 and p1/p6 cleavage sites. Mutations in the non-cleavage site regions of Gag have also been linked to drug resistance; these include H219Q and R409K [25] and K437V and K436E [26]. The mutations reported here, to our knowledge, have not been reported to be linked with drug resistance. The PTAPP motif within p6-gag is a proline rich domain that is responsible for recruiting Tsg101, a cellular factor involved in budding of the virus. The PTAPP duplication, a common polymorphism in Gag, is more common in subtype C viruses [27]. This could be due to the absence of the Alix-binding YPXnL motif in p6-gag, which aids in budding and so a duplication of the PTAPP is a compensatory mechanism for viral budding [28]. The Alix motif is mutated in more than 95% of subtype C viruses [29]. PTAPP duplications are selected for in viral isolates from HAART treatment [30] and PTAPP duplications accumulate in subtype C isolates with a smaller number of HIV-1 protease inhibitor resistance mutations as compared to subtype B [28]. It has been suggested that PTAPP duplications may play a role in subtype C virus susceptibility to HIV-1 protease inhibitors due to its position close to the cleavage site as it may be responsible for an important secondary structure conferred by proline residues [28]. The PTAPP duplication has been linked to poor virological response to the PI amprenavir (APV) [31]. The loss of susceptibility to PIs, or any ARV drug, may be due to enhanced budding in the presence of drugs [31].

The reduction in viral fitness is not uncommon in viruses that are evolving to escape drug pressure [21, 32]. The selection for drug resistance mutations in HIV-1 protease often results in a loss of replication capacity, which is then compensated for by the selection of additional compensatory mutations in Gag [33]. We have observed a reduction in the relative replicative capacity of the L38↑N↑L protease in the context of its related Gag. Kozísek *et al*., (2008) [12] also showed that a HIV-1 protease containing amino acid insertions in the hinge region display reduced replicative capacity in the presence of a mutated Gag. It is unclear whether the mutations/polymorphisms observed in the related Gag were selected for as a consequence of the L38↑N↑L insertion mutation in the HIV-1 protease, especially since our RC studies have shown a debilitating effect of the related Gag on the replication capacity of the L38↑N↑L protease. However, the possibility exists that changes in the related Gag occurred independent of the L38↑N↑L insertion mutation. In fact, it has been shown that mutations can develop in Gag without any mutations being present in the HIV-1 protease [26].

In summary, catalytically active L38↑N↑L variant protease was successfully purified and found to have reduced efficiency when compared to wild-type protease. L38↑N↑L protease is susceptible to LPV, ATV and DRV without a Gag sequence present. Phenotypic assays with a Gag sequence showed that the L38↑N↑L variant has reduced susceptibility to DRV. The L38↑N↑L protease was less susceptible to DRV when the mutated Gag was present indicating that Gag plays a role in drug susceptibility here. The duplication of the PTAPP motif may play a role. The mutated Gag lowered the replication capacity indicating mutations may have occurred in this region before the HIV-1 protease insertion mutations, possibly implicating an alternative mechanism for drug resistance developing in this patient.

References

1. UNAIDS (2016) Gobal report: UNAIDS report on global AIDS epidemic 2016. UNAIDS, Geneva, Switz.

2. McCutchan, F. E. (2006) Global Epidemiology of HIV. *J. Med. Virol*. ***78***, 7–12

3. Hemelaar, J., Gouws, E., Ghys, P. D., and Osmanov, S. (2011) Global trends in molecular epidemiology of HIV-1 during 2000– 2007. *AIDS*. ***25***, 679–689

4. Roberts, J. D., Bebenek, K., and Kunkel, T. A. (1988) The Accuracy of Reverse Transcriptase from HIV-1. *Science*. ***242***, 1171–1173

5. Turner, B. G., and Summers, M. F. (1999) Structural biology of HIV. *J. Mol. Biol*. ***285***, 1–32

6. Wlodawer, A., Miller, M., Jasklski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kentt, S. B. H. (1989) Conserved Folding in Retroviral Proteases: Crystal Structure of a Synthetic HIV-1 Protease. *Science*. ***245***, 616–621

7. Kohl, N. E., Emini, E. a, Schleif, W. a, Davis, L. J., Heimbach, J. C., Dixon, R. a, Scolnick, E. M., and Sigal, I. S. (1988) Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. U. S. A.* ***85***, 4686–90

8. World Health Organization (WHO) (2016) Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. WHO Press. Geneva, Switz.

9. Condra, J. H., Schleif, W. A., Blahy, O. M., Gabryelski, L. J., Graham, D. J., Quintero, J. C., Rhodes, A., Robbins, H. L., Roth, E., and Shivaprakash, M. (1995) In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature*. ***374***, 569–571

10. Ridky, T. W., Kikonyogo, A., Leis, J., Gulnik, S., Copeland, T., Erickson, J., Wlodawer, A., Kurinov, I., Harrison, R. W., and Weber, I. T. (1998) Drug-resistant HIV-1 proteases identify enzyme residues important for substrate selection and catalytic rate. *Biochemistry*. ***37***, 13835–13845

11. Masquelier, B., Race, E., Tamalet, C., Descamps, D., Izopet, J., Buffet-janvresse, C., Ruffault, A., Mohammed, A. L. I. S. I., Cottalorda, J., Schmuck, A., Calvez, V., Dam, E., Virology, T., and Bernard, P. C. (2001) Genotypic and Phenotypic Resistance Patterns of Human

Immunodeficiency Virus Type 1 Variants with Insertions or Deletions in the Reverse Transcriptase (RT): Multicenter Study of Patients Treated with RT Inhibitors. *Antimicrob. Agents Chemother*. ***45***, 1836–1842

12. Kozísek, M., Sasková, K. G., Rezácová, P., Brynda, J., van Maarseveen, N. M., De Jong, D., Boucher, C. a, Kagan, R. M., Nijhuis, M., and Konvalinka, J. (2008) Ninety-nine is not enough: molecular characterization of inhibitor-resistant human immunodeficiency virus type 1 protease mutants with insertions in the flap region. *J. Virol*. ***82***, 5869–78

13. Clavel, F., and Hance, A. J. (2004) HIV drug resistance. *N. Engl. J. Med*. ***350***, 1023–1035

14. Department of Health, South Africa (2010) CLINICAL GUIDELINES : PMTCT (Prevention of Mother-to- Child Transmission) National Department of Health , South Africa ;

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

16. Schägger, H. (2006) Tricine-SDS-PAGE. *Nat. Protoc*. ***1***, 16–22

17. Brautigam, C. A., Zhao, H., Vargas, C., Keller, S., and Schuck, P. (2016) Integration and global analysis of isothermal titration calorimetry data for studying macromolecular interactions. *Nat. Protoc*. ***11***, 882–894

18. Velázquez-Campoy, A., Vega, S., Fleming, E., Bacha, U., Sayed, Y., Dirr, H. W., and Freire, E. (2003) Protease Inhibition in African Subtypes of HIV-1. *AIDS Rev.* **5**, 165–171

19. Copeland, R. A., Lombardo, D., Giannaras, J., and Decicco, C. P. (1995) Estimating KI values for tight binding inhibitors from dose-response plots. *Bioorganic Med. Chem. Lett*. ***5***, 1947–1952

20. Parry, C. M., Kohli, A., Boinett, C. J., Towers, G. J., McCormick, A. L., and Pillay, D. (2009) Gag determinants of fitness and drug susceptibility in protease inhibitor-resistant human immunodeficiency virus type 1. *J. Virol*. ***83***, 9094–9101

21. Croteau, G., Doyon, L., Thibeault, D., Mckercher, G., Pilote, L., Croteau, G., Doyon, L., Thibeault, D., Kercher, G. M. C., Pilote, L., and Lamarre, D. (1997) Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors .

Impaired Fitness of Human Immunodeficiency Virus Type 1 Variants with High-Level Resistance to Protease Inhibitors. J*. Virol*. ***71***, 1089–1096

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

23. Clavel, F., and Mammano, F. (2010) Role of gag in HIV resistance to protease inhibitors. *Viruses*. ***2***, 1411–1426

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

25. Gatanaga, H., Suzuki, Y., Tsang, H., Yoshimura, K., Kavlick, M. F., Nagashima, K., Gorelick, R. J., Mardy, S., Tang, C., Summers, M. F., and Mitsuya, H. (2002) Amino Acid Substitutions in Gag Protein at Non-cleavage Sites Are Indispensable for the Development of a High Multitude of HIV-1 Resistance against Protease Inhibitors . *J. Biol. Chem*. ***277***, 5952–5961

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

27. Marlowe, N., Flys, T., Hackett, J., Schumaker, M., Jackson, J. B., and Eshleman, S. H. (2004) Analysis of insertions and deletions in the gag p6 region of diverse HIV type 1 strains. *AIDS Res. Hum. Retroviruses.* ***20***, 1119–25

28. Martins, A. N., Arruda, M. B., Pires, A. F., Tanuri, A., and Brindeiro, R. M. (2011) Accumulation of P(T/S)AP Late Domain Duplications in HIV Type 1 Subtypes B, C, and F Derived from Individuals Failing ARV Therapy and ARV Drug-Naive Patients. *AIDS Res. Hum. Retroviruses*. ***27***, 687–692

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

30. Fun, A., Wensing, A. M. J., Verheyen, J., and Nijhuis, M. (2012) Human Immunodeficiency Virus gag and protease : partners in resistance. *Retrovirology.* ***9***, 1–14

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

32. Mammano, F., Petit, C., and Clavel, F. (1998) Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *J. Virol*. ***72***, 7632–7

33. Dauber, D. S., Ziermann, R., Parkin, N., Maly, D. J., Mahrus, S., Harris, J. L., Ellman, J. A., Petropoulos, C., and Craik, C. S. (2002) Altered Substrate Specificity of Drug-Resistant Human Immunodeficiency Virus Type 1 Protease. *J. Virol*. ***76***, 1359–1368

CHAPTER 5
GENERAL DISCUSSION

5.1 Overexpression and purification of HIV-1 protease using Trx-His tag

The use of protein and peptide tags are very popular for purifying recombinant proteins. Affinity tags have the following features; a one-step adsorption purification, minimal effect on tertiary structure and biological activity of the recombinant protein and easy and specific removal of the tag [120]. A tag can be a small peptide such as poly-Arg-, FLAG-, poly-His-, c-myc-, S-, or Strep II-tag. The advantage of the small peptide tag is that under certain conditions it is not necessary to remove the tag.

A polyhistidine tag (His-tag) is widely used in the purification of recombinant proteins. Immobilised metal affinity chromatography (IMAC) is used to purify the His-tagged protein. IMAC is based on the interaction between a transition metal (commonly Co^{2+}, Ni^{2+}) immobilised on a matrix and the histidine side chain as it exhibits the strongest interaction. The electron donor groups of the histidine ring form coordination bonds with the immobilised transition metal. Once bound to the immobilised transition metal, the tagged recombinant protein can be eluted with the use of imidazole. A disadvantage of using imidazole is that it can influence NMR experiments, competition studies, crystallographic trials and its presence may cause protein aggregation [121]. The placement of the His-tag, on the N- or C-terminus, is dependent on the function of protein and can be placed at either end. The success of this tag is seen in the multiple expression systems it has been used with, which include; bacteria, yeast, mammalian cells, baculovirus-infected insect cells [120].

In cases where solubility is a problem, as with HIV-1 protease, a fusion protein may be used to improve solubility. These include maltose binding protein [122], thioredoxin (Trx) [123] or glutathione S-transferase [124]. It is unclear how these proteins enhance solubility but several hypotheses have been proposed [125]. These include the fusion protein acting as an electrostatic shield and reducing aggregation by electrostatic repulsion between highly charged soluble polypeptide extensions [126]. It has been proposed that the fusion proteins are chaperone magnets and that solubility results from interactions with endogenous chaperones [127]. Solubility enhancers

may also have an innate, passive chaperone-like quality that manifests itself by intramolecular binding to the fusion partner in a manner that prevents their self-association and aggregation [128–131]. The present study used a Trx fusion protein to improve solubility and reduce toxicity of the HIV-1 protease. Both improvement in the solubility and a reduction in toxicity, as observed by increased expression, were seen for L38↑N↑L, which was expressed in the soluble fraction. Overall, a reduction in toxicity was observed as there was a great improvement in expression from the previous expression system, which involved expression of the HIV-1 protease as part of a gag-pol construct and purification using ion exchange chromatography. The amount of wild-type protease purified was increased from 0.83 to 2.5 mg/L of culture medium and 1.5 mg/L of L38↑N↑L was obtained.

Commonly, fusion tags need to be removed to carry out down-stream processes such as crystallography. The most common proteases that are used for cleavage of the tag are enterokinase, tobacco etch virus, thrombin, and factor Xa. The protease utilised in this study was thrombin. Thrombin cleavage can be carried out between 20-37 °C. Thrombin cleavage results in the retention of two amino acids on the C-terminal side of the cleavage site of the recombinant protein. The thrombin can be removed from the target protein by affinity chromatography using Benzamidine Sepharose™. A minimum concentration of 150 mM NaCl is required to prevent non-specific binding of the HIV-1 protease to the Benzamidine Sepharose™, which proved problematic for this study. HIV-1 protease activity is increased with increased ionic strength [132–134]. Polgár *et al*. (1994) showed that the activity of HIV-1 protease significantly increased (4-fold) from 0-1 M NaCl [132] and Hyland *et al*., (1991) showed that this effect with NaCl is not saturable [60]. This increased activity of the HIV-1 protease may be attributed to the increased stability that the NaCl provides the enzyme [133, 134]. Therefore, incubation of the fusion protease with thrombin in a buffer containing 150 mM NaCl resulted in enhanced autolytic activity due to the presence of the NaCl.

Determining the concentration of HIV-1 protease active sites by isothermal titration calorimetry showed that the proteases purified using the fusion method had a low percentage of active sites, 9% for L38↑N↑L and 13% for wild-type. This is attributed to the overnight incubation in 150 mM NaCl during thrombin cleavage, which would increase HIV-1 protease activity and, therefore, increase autolytic activity. Although a Q7K mutation was included to reduce autolytic activity [135], it does not entirely knock this process out. A possible solution is to include the following substitutions in the HIV-1 protease; L33I and L63I. These have shown to reduce autolytic activity while still maintaining the functionality of the HIV-1 protease [136]. Mildner *et al*., showed that after incubation of the

Q7K/L33I/L63I mutant at room temperature for 11 days at pH 5.5 (a pH at which the HIV-1 protease is most active) there was no loss of activity and minimal loss of protein due to autolysis as opposed to wild-type which lost 40% of its activity [136].

5.2 Molecular Dynamics and Induced-fit docking

Molecular dynamics simulations have been vital in our understanding of proteins as dynamic structures rather than rigid structures. The first molecular simulation of a protein was published in 1977 [137] and in the last 40 years the technique has advanced greatly due to the increase in computing power. Molecular dynamics simulations have become an important tool in studying proteins. Regions of proteins that may be difficult to view using crystallography, such as loops can be modelled using molecular dynamics. The movement of proteins in solution can be modelled and the movement of flexible regions for example the flap region in HIV-1 protease can be viewed. Molecular dynamics, however, cannot replace *in vitro* studies and it is important to confirm results found *in silico, in vitro*.

The dynamics of many proteins have been studied using simulations and HIV-1 protease is no exception. The flap region has been of particular interest [138–142] due to its importance in substrate binding: the flaps open to allow substrates to bind and close upon substrate binding [139]. In the apo-form the flaps can exist in three conformations; open, semi-open and closed [140]. The dominant conformation in solution is thought to be the semi-open [140]. There are different ways of measuring the opening and closing of the flap region: the distance between two Ile50 residues (one in each flap tip) can be measured and the distance between the catalytic residue (Asp25) and Ile50 [141].

The distance between Ile50 residues [141] in the flap tips is measured to determine whether the flaps are in the open or closed conformation. This residue is chosen because it is highly conserved and occurs on the tip of the flap. The distance between the Asp25 and Ile50 can be chosen instead of the distance between Ile50 residues due to flap curling that occurs [141]. A comparison of the difference between these two approaches is seen in Figure 7A (distance between Ile50 residues) and Figure 7B (distance between Asp25 and Ile50). The flaps need to curl in order to open. Curling, however, does not have to occur symmetrically and one flap may curl more than the other. This can be seen in Figure 7C. The flap of chain B in both wild-type and L38↑N↑L curls inward (~115°) more

than the chain A flap and so the distances between Ile50 and Asp25 in chain B are more varied than that of Chain A.

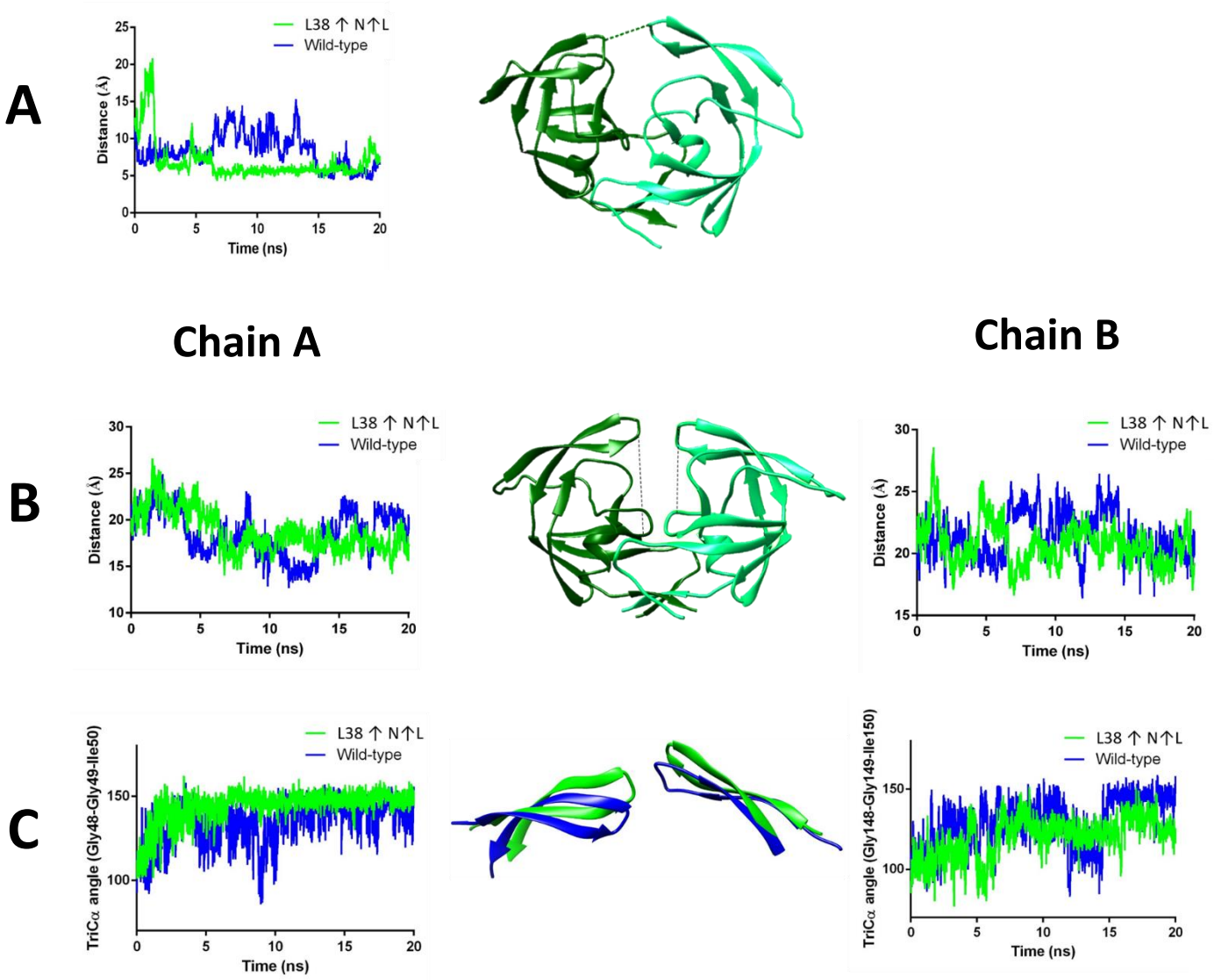


Figure 7: Comparison of methods to determine flap conformation. The distance between Ile50 residues in each monomer was measured in (A). The distance between Ile50 and Asp25 was measured for each chain in (B). The flap curling for each chain was measured in (C). Structures were generated in PyMol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

For this reason, the distances between Ile50 residues in each residue was chosen to determine when the flaps are in the open and closed conformations.

The flap region of L38↑N↑L was found to be less dynamic than that of the wild-type protease and remained in the closed conformation (4 Å apart) for most of the simulation. It has long been thought that the flap regions of HIV-1 protease remain in the semi open conformation in solution when not bound to an inhibitor. Upon inhibitor binding the flap region adopts and remains in the closed conformation. This closed conformation is maintained due to the high energy barrier the HIV-1 protease would have to overcome to open the flaps [143]. Recently, it has been shown using NMR studies that inhibitor-free HIV-1 protease can adopt a closed conformation in solution [144], confirming the finding shown here by molecular simulation.

Docking ligands to protein active sites is an important technique in drug discovery. It aids in the identification of novel compounds that may possibly bind to the target without the expense of *in vitro* testing. However, it cannot be used in isolation and *in vitro* testing must still occur. It is advantageous in that it will minimise the number of compounds that need to be tested and so only the best hits are used. Traditionally, docking was performed using a flexible ligand that is positioned into a rigid binding site. The main reason for the rigid protein is due to computational feasibility. This method of docking is representative of the lock-and-key model of ligand binding to a protein suggested by Emil Fischer in 1890. This, however, is not a true representation of what occurs *in vitro*. Proteins and ligands are dynamic in solution and have to be modelled as such. It has become clear that protein flexibility is crucial to the receptor-ligand complex formation and must be considered *in silico* [145, 146]. Induced-fit docking has the benefit of including the motions of both the protein and ligand. This is representative of the induced-fit model of ligand binding to a protein. Induced-fit docking is more representative of what is occurring *in vitro* and thus is a preferred method of docking drugs to a protein. Docking of the drugs LPV, ATV and DRV to L38↑N↑L showed a reduction in the number of hydrogen bonds formed and hydrophobic contacts made. Two hydrophobic contacts were lost between LPV and L38↑N↑L but there was an addition of hydrogen bonds between the side chains of Arg8, Asp29 and Asp30 and LPV. L38↑N↑L and ATV gained five hydrophobic contacts and lost hydrogen bonding between the side chains of Arg8 and Asp30. There was a loss of only one hydrophobic contact for L38↑N↑L and DRV and no loss of hydrogen bonding. The pattern of hydrogen bonding is shown in Figure 8.

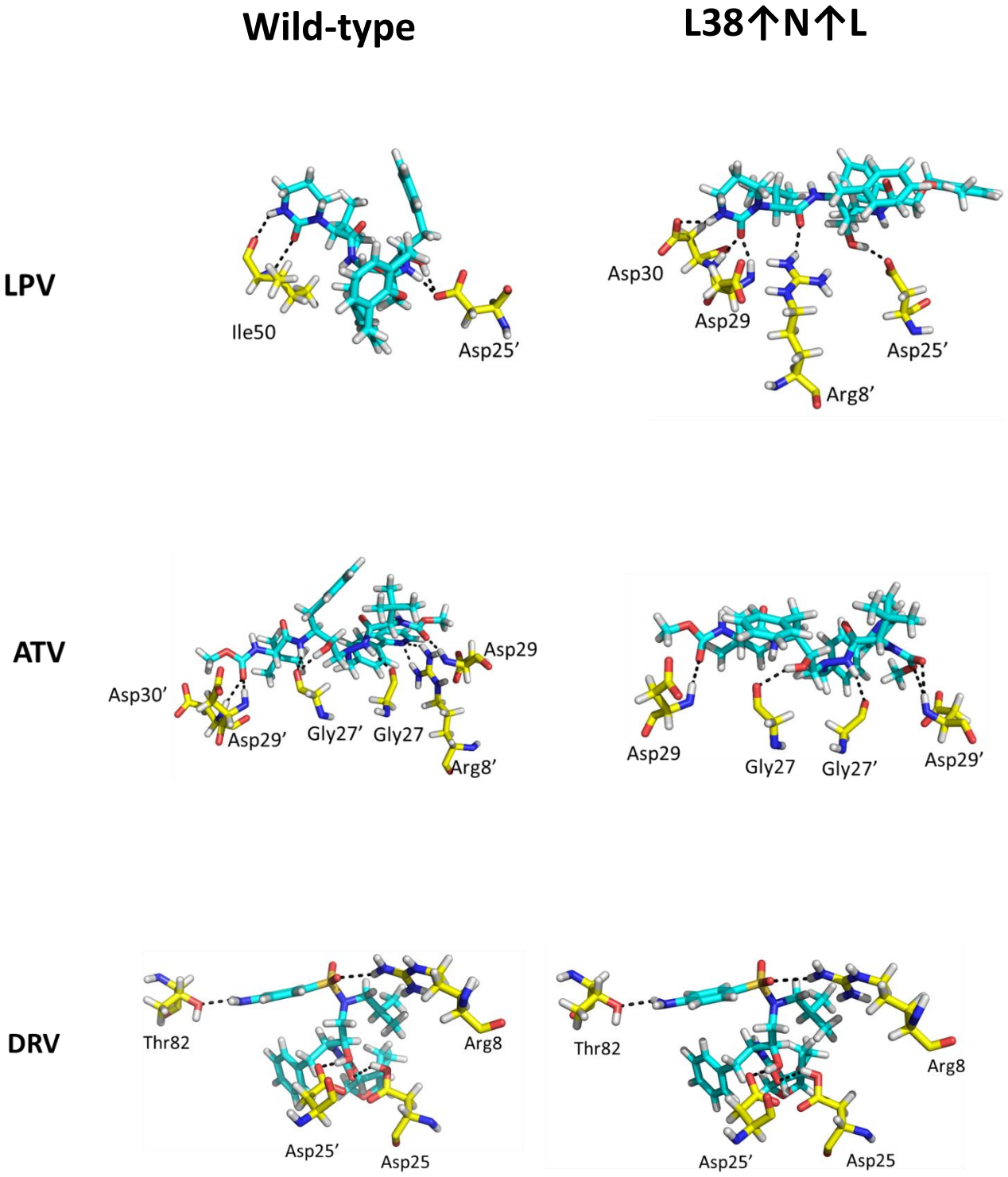


Figure 8: Hydrogen bonds formed between wild-type and L38↑N↑L protease and LPV, ATV and DRV. LPV binding to L38↑N↑L results in the loss of the hydrogen bond between Ile50 and LPV but new hydrogen bonds between Arg8, Asp29 and Asp30 and LPV are formed. ATV binding to L38NL resulted in a loose of hydrogen bonds between ATV and Asp30' and Arg8'. The hydrogen bonds between DRV and L38↑N↑L are maintained. Image generated using PyMol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

Future work resulting from these *in silico* experiments would be using ITC to determine the thermodynamic parameters and dissociation constant for each of the drugs to L38↑N↑L. To confirm the flap dynamics observed, hydrogen-deuterium exchange mass spectrometry can be used as was done previously in our laboratory [147]. NMR studies can be performed such as Louis and Roche (2016) [143], who showed that binding of HIV-1 protease inhibitors traps the enzyme in an energy minimum of the closed postion . Importantly, a crystal structure of the variant needs to be obtained in the apo- and drug-bound forms.

5.3. Drug susceptibility

The challenge faced by treatment of HIV and AIDS is the occurrence of drug resistance [148]. WHO estimates that half of the people infected with HIV are on therapy [149] and this means many more people will be faced with resistance to drugs than in the past. There are two ways of determining drug resistance, genotyping and with phenotypic assays. The phenotypic assay is expensive and time consuming as it involves measuring replication of the virus (with clinically-derived RNA) in the presence of different drugs. Genotyping, which is more efficient, involves sequencing the viral genome and identifying mutations that have been associated with drug resistance. A list of all mutations that cause drug resistance has been compiled and is regularly updated [150]. Databases, such as the Stanford University HIV Drug Resistance Database, are usual tools for identifying whether a given sequence contains a drug resistance mutation [151–153]. Upon analysing the protein sequencing of L38↑N↑L, using the HIV Drug Resistance Database, it was found that there were no drug resistant mutations present. This database, however, does not include any amino acid insertions for HIV-1 protease and so it is important to determine what effect the presence of these insertions may have on drug resistance phenotypically.

The functionality of L38↑N↑L was analysed using Michaelis-Menten steady-state enzyme kinetics. L38↑N↑L displayed a lower K_M, which normally would be indicative of a greater catalytic efficiency but this was not the case. The catalytic efficiency of L38↑N↑L was 12-fold lower than the wild-type protease. This confirms the findings of Kozísek *et al.,* (2008) [118] and Sasková *et al*., (2014) [154] and also explains the reduced replicative capacity of the pseudovirus as a less efficient enzyme would result in reduced infectivity of the virus. If the variant protease is not as efficient as the wild-type protease the virus will not be able to replicate as efficiently as a wild-type virus. The low

catalytic efficiency could be attributed to the reduced flap dynamics seen by L38↑N↑L. If the flap regions remain predominantly in the closed conformation the HIV-1 protease will not be able to bind the substrate. The inhibitory constant for L38↑N↑L in the presence of LPV, ATV and DRV was determined and it was found that these drugs would inhibit L38↑N↑L.

Using phenotypic assays, it was shown that a virus containing the L38↑N↑L protease was susceptible to LPV and ATV with or without a mutated Gag. The L38↑N↑L IC_{50} values for LPV and ATV were both increased from wild-type IC_{50} values but were not significantly different to indicate a reduced susceptibility. The virus containing L38↑N↑L showed reduced susceptibility to DRV, which was not expected because of the K_i value determined in the absence of a Gag sequence. This finding implies that the Gag region may play a role in reduced DRV susceptibility that was seen.

The importance of the Gag region to drug susceptibility was highlighted by the small but significant increase in DRV IC_{50} when the mutated Gag (IC_{50} =1.6 ± 0.2 nM, 5-fold higher than wild-type), as opposed to the wild-type Gag (IC_{50}= 1.0 ± 0.2 nM, 3-fold higher than wild-type), was present. Mutations in Gag cleavage sites [42] and non-cleavage site positions [155, 156] can contribute to HIV-1 protease inhibitor resistance. Upon evaluation of the variant Gag sequence, it was found not to include any mutations in the cleavage sites, and elsewhere, that are known to cause drug resistance. It does, however, contain a duplicated PTAPP region. The PTAPP motif is within p6 and is responsible for recruiting Tsg101, a cellular factor involved in budding of the virus. This region has been contentious in the drug resistance debate. Some suggest that this region does contribute to drug resistance [157–161] while others report that this is a common polymorphism in different subtypes [162–164]. PTAPP duplication is common to subtype C viruses and this is attributed to the deletion of the Alix-binding YPXnL motif in p6, which aids in budding [159, 165]. Poor virological response to APV has been linked to the PTAPP duplication [158]. It cannot be discounted as the reason for the decreased susceptibility to DRV seen here but neither can it be confirmed. The PTAPP motif may result in a loss of susceptibility to HIV-1 protease inhibitors due to enhanced budding in the presence of drugs [158].

Replicative capacity of the pseudovirus containing L38↑N↑L and its associated Gag was greatly diminished. Kozísek *et al.,* (2008) [118] also showed that a HIV-1 protease containing amino acid insertions reduced replicative capacity in the presence of a mutated Gag. However, it was not shown what the effect of this HIV-1 protease alone had on replicative capacity by using a wild-type Gag sequence. The present study showed that replicative capacity was restored when the wild-type Gag

was present. This finding makes it difficult to suggest that HIV-1 protease acquired mutations before Gag and that the mutations found in Gag occurred to compensate for mutations in HIV-1 protease. However, it does suggest that Gag may have acquired mutations before HIV-1 protease and that HIV-1 protease mutated to compensate. This is not unusual as it has been shown that Gag can acquire mutations without any mutations being present in HIV-1 protease [156].

Genotypic analysis of a HIV-1 protease sequence is not enough to determine if the virus is resistant or displays reduced susceptibility to HIV-1 protease inhibitors. This is because it does not consider mutations in regions other than reverse transcriptase, integrase and HIV-1 protease. The importance of considering the Gag region when looking at drug resistance is not only highlighted in this study but also by Giandhari *et al*. [166]. Future work would include phenotypic analysis of the mutated Gag sequence with a wild-type HIV-1 protease to determine whether the reduced DRV susceptibility observed is solely due to the Gag sequence.

CONCLUSIONS

L38↑N↑L and wild-type protease were both successfully overexpressed and purified using a Trx-His tag fusion system. This fusion system reduced cytotoxicity during overexpression and improved yield during purification. Molecular dynamics showed that the flap region of L38↑N↑L was less dynamic than that of wild-type protease. This reduced movement could be a possible mechanism to evade drug binding. Induced-fit docking showed that the drugs LPV, ATV and DRV would bind L38↑N↑L albeit with a reduction in hydrogen bonds and hydrophobic contacts. This finding was confirmed by determining the K_i of each drug binding to L38↑N↑L, which indicated that all drugs would inhibit L38↑N↑L. Catalytic efficiency of this variant was reduced and may be attributed to the extended time the flap region of this HIV-1 protease spends in the closed conformation as shown by molecular dynamics. The associated Gag region of L38↑N↑L possibly contributes to reduced DRV susceptibility seen in phenotypic assays. The replication capacity of a pseudovirus containing L38↑N↑L and its associated Gag was greatly reduced. Upon replacement of the Gag with a wild-type Gag, replication capacity was restored. This implicates Gag as having mutated first and then the HIV-1 protease mutating to compensate for the mutated Gag substrate.

REFERENCES

1. UNAIDS (2017) UNAIDS Data 2017. *UNAIDS, Geneva, Switz.* 978-92-9173-945-5

2. Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983) Isolation of a T-Lymphotopic Retrovirus from a Patient at Risk for Aquired Immune Deficiency Syndrome (AIDS). *Lancet*. **220**, 868–871

3. UNAIDS (2016) Gobal report: UNAIDS report on global AIDS epidemic 2016. *UNAIDS, Geneva, Switz.*

4. UCSF, Anova Health Institute, and WHRI (2015) South African Health Monitoring Study (SAHMS), Final Report: The Integrated Biological and Behavioural Survey among Female Sex Workers, South Africa 2013-2014. *San Fr. UCSF*

5. Turner, B. G., and Summers, M. F. (1999) Structural biology of HIV. *J. Mol. Biol.* **285**, 1–32

6. Campbell, E. M., and Hope, T. J. (2015) HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat. Rev. Microbiol.* **13**, 471–483

7. Ganser, B. K., Li, S., KlishkoV.Y., Finch, J. T., and Sundquist, W. . (1999) Assembly and Analysis of Conical Models for the HIV-1 Core. *Science (80-.).* **283**, 80–83

8. Li, S., Hill, C. P., Sundquist, W. I., and Finch, J. T. (2000) Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature*. **407**, 409–413

9. Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C. V, Lasky, L. A., and Capon, D. J. (1985) Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. *Nature*. **313**, 450–8

10. McCutchan, F. E. (2006) Global Epidemiology of HIV. *J. Med. Virol.* **78**, 7–12

11. Ho, D. D. (1997) Perspectives series: host/pathogen interactions. Dynamics of HIV-1 replication in vivo. *J. Clin. Invest.* **99**, 2565–7

12. Taylor, B. S., Sobieszczyk, M. E., McCutchan, F. E., and Hammer, S. M. (2008) The challenge of HIV-1 subtype diversity. *N. Engl. J. Med.* **359**, 1965–6

13. Levy, D. N., Aldrovandi, G. M., Kutsch, O., and Shaw, G. M. (2004) Dynamics of HIV-1 recombination in its natural target cells. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4204–4209

14. Robertson, D. L., Anderson, J. P., Bradac, J. a, Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S., and Korber, B. (2000) HIV-1 nomenclature proposal. *Science*. **288**, 55–6

15. Roques, P., Robertson, D. L., Souquière, S., Apetrei, C., Nerrienet, E., Barré-Sinoussi, F., Müller-Trutwin, M., and Simon, F. (2004) Phylogenetic characteristics of three new HIV-1 N strains and implications for the origin of group N. *AIDS*. **18**, 1371–1381

16. Yamaguchi, J., Bodelle, P., Vallari, A. S., Coffey, R., McArthur, C. P., Schochetman, G., Devare, S. G., and Brennan, C. A. (2004) HIV Infections in Northwestern Cameroon: Identification of HIV Type 1 Group O and Dual HIV Type 1 Group M and Group O Infections. *AIDS Res. Hum. Retroviruses*. **20**, 944–957

17. Hemelaar, J., Gouws, E., Ghys, P. D., and Osmanov, S. (2006) Global and regional distribution of HIV-1

genetic subtypes and recombinants in 2004. *AIDS*. **20**, 13–23

18. Monini, P., Sgadari, C., Toschi, E., Barillari, G., and Ensoli, B. (2004) Antitumour effects of antiretroviral therapy. *Nat. Rev. Cancer*. **4**, 861–75

19. Doranz, B. J., Rucker, J., Yi, Y., Smyth, R. J., Samson, M., Peiper, S. C., Parmentier, M., Collman, R. G., and Doms, R. W. (1996) A dual-tropic primary HIV-1 isolate that uses fusin and the β-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell*. **85**, 1149–1158

20. Miller, M. D., Farnet, C. M., and Bushman, F. D. (1997) Human immunodeficiency virus type 1 preintegration complexes : studies of organization and composition . Human Immunodeficiency Virus Type 1 Preintegration Complexes : Studies of Organization and Composition. *J. Virol.* **71**, 5382–5390

21. Matreyek, K. A., and Engelman, A. (2013) Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. *Viruses*. **5**, 2483–2511

22. Fouchier, R. A. M., Meyer, B. E., Simon, J. H. M., Fischer, U., and Malim, M. H. (1997) HIV-1 infection of non-dividing cells: Evidence that the amino-terminal basic region of the viral matrix protein is important for Gag processing but not for post-entry nuclear import. *EMBO J.* **16**, 4531–4539

23. Freed, E. O., and Martin, M. A. (1995) Virion incorporation of envelope glycoproteins with long but not short cytoplasmic tails is blocked by specific, single amino acid substitutions in the human immunodeficiency virus type 1 matrix. *J. Virol.* **69**, 1984–9

24. Popov, S., Rexach, M., Zybarth, G., Railing, N., Lee, M. A., Ratner, L., Lane, C. M., Moore, M. S., Blobel, G., and Bukrinsky, M. (1998) Viral protein R regulates nuclear import of the HIV-1 pre-integration complex. *EMBO J.* **17**, 909–917

25. Vodicka, M. A., Koepp, D. M., Silver, P. A., and Emerman, M. (1998) HIV-1 Vpr interacts with the nuclear transport pathway to promote macrophage infection. *Genes Dev.* **12**, 175–185

26. Frankel, A. D., and Young, J. A. (1998) HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **67**, 1–25

27. Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*. **331**, 280–283

28. Bell, N. M., and Lever, A. M. L. (2013) HIV Gag polyprotein: Processing and early viral particle assembly. *Trends Microbiol.* **21**, 136–144

29. Muriaux, D., and Darlix, J.-L. (2010) Properties and functions of the nucleocapsid protein in virus assembly. *RNA Biol.* **7**, 744–753

30. Carlson, L. A., Briggs, J. A. G., Glass, B., Riches, J. D., Simon, M. N., Johnson, M. C., Müller, B., Grünewald, K., and Kräusslich, H. G. (2008) Three-Dimensional Analysis of Budding Sites and Released Virus Suggests a Revised Model for HIV-1 Morphogenesis. *Cell Host Microbe*. **4**, 592–599

31. Kaplan, A., Manchester, M., and Swanstrom, R. (1994) The Activity of the Protease of Human Immunodeficiency Virus Type 1 Is Initiated at the Membrane of Infected Cells before the Release of Viral Proteins and Is Required for Release To Occur with Maximum Efficiency. *J. Virol.* **68**, 6782–6786

32. Tang, C., Louis, J. M., Aniana, A., Suh, J., and Clore, G. M. (2008) Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease. *Nature*. **455**, 693–696

33. Partin, K., Zybarth, G., Ehrlich, L., DeCrombrugghe, M., Wimmer, E., and Carter, C. (1991) Deletion of sequences upstream of the proteinase improves the proteolytic processing of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 4776–4780

34. Louis, J. M., Clore, G. M., and Gronenborn, a M. (1999) Autoprocessing of HIV-1 protease is tightly coupled to protein folding. *Nat. Struct. Biol.* **6**, 868–75

35. Sluis-Cremer, N., Arion, D., Abram, M. E., and Parniak, M. A. (2004) Proteolytic processing of an HIV-1 pol polyprotein precursor: Insights into the mechanism of reverse transcriptase p66/p51 heterodimer formation. *Int. J. Biochem. Cell Biol.* **36**, 1836–1847

36. Pettit, S. C., Everitt, L. E., Choudhury, S., Dunn, B. M., and Kaplan, A. H. (2004) Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. *J. Virol.* **78**, 8477–85

37. Murakami, T., Ablan, S., Freed, E. O., and Tanaka, Y. (2004) Regulation of human immunodeficiency virus type 1 Env-mediated membrane fusion by viral protease activity. *J. Virol.* **78**, 1026–1031

38. Wyma, D. J., Jiang, J., Shi, J., Zhou, J., Lineberger, E., Miller, M. D., Aiken, C., and Lineberger, J. E. (2004) Coupling of Human Immunodeficiency Virus Type 1 Fusion to Virion Maturation : a Novel Role of the gp41 Cytoplasmic Tail Coupling of Human Immunodeficiency Virus Type 1 Fusion to Virion Maturation : a Novel Role of the gp41 Cytoplasmic Tail. *J. Virol.* **78**, 3429–3435

39. Weiss, A., Hollander, H. and, and Stobo, H. (1985) Acquired Immuno-Deficiency Syndrome: Epidemiology, Virology and Immunology. *Annu. Rev. Med.* **36**, 535–564

40. Pettit, S. C., Henderson, G. J., Schiffer, C. A., and Swanstrom, R. (2002) Replacement of the P1 Amino Acid of Human Immunodeficiency Virus Type 1 Gag Processing Sites Can Inhibit or Enhance the Rate of Cleavage by the Viral Protease. *J. Virol.* **76**, 10226–10233

41. Prabu-jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2002) Substrate Shape Determines Specificity of Recognition for HIV-1 Protease : Analysis of Crystal Structures of Six Substrate Complexes. *Structure*. **10**, 369–381

42. Clavel, F., and Mammano, F. (2010) Role of gag in HIV resistance to protease inhibitors. *Viruses*. **2**, 1411–1426

43. Pey, S. C., Moody, M. D., Wehbie, R. S., Kaplan, A. H., Nantermet, P. V, Kleinmil, C. A., and Swanstrom, R. (1994) The p2 Domain of Human Immunodeficiency Virus Type 1 Gag Regulates Sequential Proteolytic Processing and Is Required To Produce Fully Infectious Virions. *J. Virol.* **68**, 8017–8027

44. Shehu-Xhilaga, M., Kraeusslich, H. G., Pettit, S., Swanstrom, R., Lee, J. Y., Marshall, J. A., Crowe, S. M., and Mak, J. (2001) Proteolytic Processing of the P2 / Nucleocapsid Cleavage Site Is Critical for Human Immunodeficiency Virus Type 1 RNA Dimer Maturation. *J. Virol.* **75**, 9156–9164

45. Clavel, F., and Hance, A. J. (2004) HIV drug resistance. *N. Engl. J. Med.* **350**, 1023–1035

46. Ratner, L., Starcich, B., Josephs, S. F., Hahn, B. H., Reddy, E. P., Livak, K. J., S.R., P., M.L., P., Haseltine, W. ., Arya, S. ., and Wong-Staal, F. (1985) Polymorphism of the 3' open reading frame of the virus associated with the acquired immune deficiency syndrome, human T-lymphotropic virus type III. *Nucleic Acids Res.* **13**, 8219–8229

47. Darkes, P. L., Leu, C., Davis, J., Heimbach, J. C., Diehl, R. E., Hill, W. S., Dixon, R. A. F., and Irving, S. (1989) Human Immunodeficiency Virus Protease Bacterial expression and characterization of the purified aspartic protease. *J. Biol. Chem.* **264**, 2307–2312

48. Hansen, J., Billich, S., Schulze, T., Sukrow, S., and Moelling, K. (1988) Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. *EMBO J.* **7**, 1785–1791

49. Seelmeier, S., Schmidt, H., Turk, V., and von der Helm, K. (1988) Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 6612–6

50. Umezawa, H., Aoyagi, T., Morishima, H., Matsuzaki, M., Hamada, M., and Takeuchi, T. (1970) A NEW PEPSIN PRODUCED A specific inhibitor of an enzyme might be useful for the analysis of certain genic phenomena and also perhaps for the treatment of certain diseases . It was postu- lated that

microorganisms because they produce catobolic enzymes may. *J. Antibiot. (Tokyo).* **XXIII**, 259–262

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

52. Wlodawer, A., Miller, M., Jasklski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kentt, S. B. H. (1989) Conserved Folding in Retroviral Proteases: Crystal Structure of a Synthetic HIV-1 Protease. *Science (80-.).* **245**, 616–621

53. Naicker, P., Seele, P., Dirr, H. W., and Sayed, Y. (2013) F99 is Critical for Dimerization and Activation of South African. *Protein J.* **99**, 560–567

54. Pearl, L., and Blundell, T. (1984) The active site of aspartic proteinases. *FEBS Lett.* **174**, 96–101

55. Harte, W. E., Swaminathan, S., Mansuri, M. M., Martin, J. C., Rosenberg, I. E., Beveridge, D. L., Harte Jr., W. E., Swaminathan, S., Mansuri, M. M., Martin, J. C., Rosenberg, I. E., and Beveridge, D. L. (1990) Domain communication in the dynamical structure of human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8864–8868

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

57. Schechter, I., and Berger, A. (1972) On the size of the active site in proteases: Papain. *Biochem. Biophys. Res. Commun.* **46**, 1956–1960

58. Wlodawer, a, and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249–84

59. Shen, C. H., Tie, Y., Yu, X., Wang, Y. F., Kovalevsky, A. Y., Harrison, R. W., and Weber, I. T. (2012) Capturing the reaction pathway in near-atomic-resolution crystal structures of HIV-1 protease. *Biochemistry*. **51**, 7726–7732

60. Hyland, L. J., Tomaszek, T. A., Roberts, G. D., Carr, S. A., Magaard, V. W., Bryan, H. L., Fakhoury, S. A., Moore, M. L., Minnich, M. D., Culp, J. S., Desjarlais, R. L., and Meek, T. D. (1991) Human Immunodeficiency Virus- 1 Protease . 1 . Initial Velocity Studies and Kinetic Characterization of Reaction Intermediates by 180 Isotope Exchange. *Biochemistry*. **30**, 8441–8453

61. Swanstrom, R., and Coffin, J. (2012) HIV-1 Pathogenesis : The Virus. *Cold Spring Harb. Perspect. ub Med.* **2**, 1–19

62. Arts, E. J., and Hazuda, D. J. (2012) HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med.* **2**, a007161

63. Fonner, V. A., Denison, J., Kennedy, C. E., O Reilly, K., and Sweat, M. (2012) Voluntary counseling and testing (VCT) for changing HIV- related risk behavior in developing countries. *Cochrane Database Syst. Rev.* 10.1002/14651858.CD001224.pub4.www.cochranelibrary.com

64. Weber, I. T., and Agniswamy, J. (2009) HIV-1 Protease: Structural Perspectives on Drug Resistance. *Viruses*. **1**, 1110–1136

65. Kaplan, A. H., Michael, S. F., Wehbie, R. S., Knigge, M. F., Paul, D. A., Everitt, L., Kempf, D. J., Norbeck, D. W., Erickson, J. W., and Swanstrom, R. (1994) Selection of multiple human immunodeficiency virus type 1 variants that encode viral proteases with decreased sensitivity to an inhibitor of the viral protease. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5597–601

66. Rosé, J. R., Babé, L. M., and Craik, C. S. (1995) Defining the level of human immunodeficiency virus type 1 (HIV-1) protease activity required for HIV-1 particle maturation and infectivity. *J. Virol.* **69**, 2751–2758

67. Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M., and Sigal, I. S. (1988) Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4686–90

68. Wlodawer, A., and Erickson, J. . (1993) Structure-based Inhibitors of HIV-1 Protease. *Annu. Rev. B*. **62**, 543–585

69. Roberts, N., Martin, J., Kinchington, D., Broadhurst, A., Craig, J., Duncan, I., Galpin, S., Handa, B., Kay, J., Krohn, A., and et, al. (1990) Rational design of peptide-based HIV proteinase inhibitors. *Science (80-.).* **248**, 358–361

70. Ghosh, A. K., Bilcer, G., and Schiltz, G. (2001) Syntheses of FDA approved HIV protease inhibitors. *Synthesis (Stuttg).* **15**, 2203–2229

71. Mehandru, S., and Markowitz, M. (2003) Tipranavir: a novel non-peptidic protease inhibitor for the treatment of HIV infection. *Expert Opin. Investig. Drugs*. **12**, 1821–1828

72. Koh, Y., Nakata, H., Maeda, K., Bilcer, G., Devasamudram, T., Kincaid, F., Boross, P., Wang, Y., Tie, Y., Volarath, P., Gaddis, L., Harrison, R. W., Weber, T., Ghosh, A. K., Mitsuya, H., Koh, Y., Nakata, H., Maeda, K., Ogata, H., Bilcer, G., Devasamudram, T., Kincaid, J. F., Boross, P., Wang, Y., Tie, Y., Volarath, P., Gaddis, L., Harrison, R. W., Weber, I. T., Ghosh, A. K., and Mitsuya, H. (2003) Nonpeptidic Protease Inhibitor (PI) UIC-94017 (TMC114) with Potent Activity against Multi-PI-Resistant Human Immunodeficiency Virus In Vitro Novel bis -Tetrahydrofuranylurethane-Containing Nonpeptidic Protease Inhibitor (PI) UIC-94017 (TMC114) wit. *Antimicrob. Agents Chemother.* **94017**, 3123–3129

73. Pokorná, J., Machala, L., Řezáčová, P., and Konvalinka, J. (2009) *Current and novel inhibitors of HIV protease*, 10.3390/v1031209

74. Abdel-Meguid, S. S., Murthy, K. H. M., Winborne, E., Zhao, B., Metcalf, B. W., Carr, T. J., Tomaszek, T. A., Dreyer, G. B., Meek, T. D., Pitts, W. J., Demarsh, P., Ivanoff, L., Lambert, D. M., Petteway, S. R., DesJarlais, R. L., Fisher, S., and Green, D. W. (1994) An Orally Bioavailable HIV-1 Protease Inhibitor Containing an Imidazole-Derived Peptide Bond Replacement: Crystallographic and Pharmacokinetic Analysis. *Biochemistry*. **33**, 11671–11677

75. Hoog, S. S., Towler, E. M., Zhao, B., Doyle, M. L., Debouck, C., and Abdel-Meguid, S. S. (1996) Human immunodeficiency virus protease ligand specificity conferred by residues outside of the active site cavity. *Biochemistry*. **35**, 10279–10286

76. Bold, G., Fässler, A., Capraro, H. G., Cozens, R., Klimkait, T., Lazdins, J., Mestan, J., Poncioni, B., Rösel, J., Stover, D., Tintelnot-Blomley, M., Acemoglu, F., Beck, W., Boss, E., Eschbach, M., Hürlimann, T., Masso, E., Roussel, S., Ucci-Stoll, K., Wyss, D., and Lang, M. (1998) New aza-dipeptide analogues as potent and orally absorbed HIV-1 protease inhibitors: Candidates for clinical development. *J. Med. Chem.* **41**, 3387–3401

77. Robinson, B. S., Riccardi, K. A., Gong, Y. F., Guo, Q., Stock, D. A., Blair, W. S., Terry, B. J., Deminie, C. A., Djang, F., Colonno, R. J., and Lin, P. F. (2000) BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents. *Antimicrob Agents Chemother*. **44**, 2093–2099

78. Gong, Y.-F., Robinson, B. S., Rose, R. E., Deminie, C., Spicer, T. P., Stock, D., Colonno, R. J., and Lin, P. (2000) In Vitro Resistance Profile of the Human Immunodeficiency Virus Type 1 Protease Inhibitor BMS-232632. *Antimicrob. Agents Chemother.* **44**, 2319–2326

79. Ghosh, A. K., Kincaid, J. F., Cho, W., Walters, E., Krishan, K., Hussain, K. A., Koo, Y., Cho, H., Rudall, C., Holland, L., and Buthod, J. (1998) Potent HIV protease inhibitors incorporating high-affinity P2-ligands and (R)-(hydroxyethylamino) sulfonamide isostere. *Bioorganic Med. Chem. Lett.* **8**, 687–690

80. Ghosh, A. K., Dawson, Z. L., and Mitsuya, H. (2007) Darunavir, a Conceptually New HIV-1 Protease

Inhibitor for the Treatment of Drug-resistant HIV. *Bioorganic Med. Chem.* **31**, 1713–1723

81. Ghosh, A. K., Chapsal, B. D., Weber, I. T., and Mitsuya, H. (2008) Design of HIV protease inhibitors targeting protein backbone: An effective strategy for combating drug resistance. *Acc. Chem. Res.* **41**, 78–86

82. Wensing, A. M. J., van Maarseveen, N. M., and Nijhuis, M. (2010) Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res.* **85**, 59–74

83. Ammaranond, P., Cunningham, P., Oelrichs, R., Suzuki, K., Harris, C., Leas, L., Grulich, A., Cooper, D. a, and Kelleher, A. D. (2003) Rates of transmission of antiretroviral drug resistant strains of HIV-1. *J. Clin. Virol.* **26**, 153–61

84. Hirsch, M. S., Conway, B., Aquila, R. T. D., Johnson, V. A., Clotet, B., Demeter, L. M., and Hammer, S. M. (1998) Antiretroviral Drug Resistance Testing in Adults With HIV Infection Implications for Clinical Management. *Int. AIDS Soc. USA Panel. - Jama*. **279**, 1984–1991

85. Mansky, L. M., and Temin, H. M. (1995) Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse Lower In Vivo Mutation Rate of Human Immunodeficiency Virus Type 1 than That Predicted from the Fidelity of Purified Reverse Trans. *J. Virol.* **69**, 5087–5094

86. Domingo, E., Menéndez-Arias, L., and Holland, J. J. (1997) RNA virus fitness. *Rev. Med. Virol.* **7**, 87–96

87. Leigh Brown, A. J., and Richman, D. D. (1996) HIV-1: Gambling on the evolution of drug resistance? *Nat. Med.* **3**, 268–271

88. Leigh Brown, A. J. (1997) Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1862–5

89. Nijhuis, M., Boucher, C. A. B., Schipper, P., Leitner, T., Schuurman, R., and Albert, J. (1998) Stochastic processes strongly influence HIV-1 evolution during suboptimal protease-inhibitor therapy. *Proc. Natl. Acad. Sci.* **95**, 14441–14446

90. Condra, J. H., Schleif, W. A., Blahy, O. M., Gabryelski, L. J., Graham, D. J., Quintero, J. C., Rhodes, A., Robbins, H. L., Roth, E., and Shivaprakash, M. (1995) In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature*. **374**, 569–571

91. Vergne, L., Peeters, M., Mpoudi-ngole, E., Bourgeois, A., Liegeois, F., Toure-kane, C., Mboup, S., Mulanga-kabeya, C., Saman, E., Jourdan, J., Reynes, J., Delaporte, E., Liegeois, F., Toure-kane, C., Mboup, S., Mulanga-kabeya, C., Saman, E., Jourdan, J., Reynes, J., and Delaporte, E. (2000) Genetic Diversity of Protease and Reverse Transcriptase Sequences in Non-Subtype-B Human Immunodeficiency Virus Type 1 Strains : Evidence of Many Minor Drug Resistance Mutations in Treatment-Naive Patients Genetic Diversity of Protease and Reverse Transcr. *J. Clin. Microbiol.* **38**, 3919–3925

92. Hertogs, K., Bloor, S., Kemp, S. D., Den, C. Van, Alcorn, T. M., Pauwels, R., Houtte, M. Van, Staszewski, S., Miller, V., and Larder, B. A. (2000) Phenotypic and genotypic analysis of clinical HIV-1 isolates reveals extensive protease inhibitor cross- resistance : a survey of over 6000 samples. *AIDS*. **14**, 1203–1210

93. Miller, V. (2001) *Resistance to protease inhibitors*, 10.1097/00042560-200103011-00005

94. Hanna, G. J., and Caliendo, A. M. (2001) Testing for HIV-1 drug resistance. *Mol. Diagnosis*. **6**, 253–263

95. Kim, R., and Baxter, J. D. (2008) Protease inhibitor resistance update: where are we now? *AIDS Patient Care STDS*. **22**, 267–77

96. Velazquez-Campoy, A., Kiso, Y., and Freire, E. (2001) The binding energetics of first- and second-

generation HIV-1 protease inhibitors: implications for drug design. *Arch. Biochem. Biophys.* **390**, 169–75

97. Ohtaka, H., Velazquez-Campoy, A., Xie, D., and Freire, E. (2002) Overcoming drug resistance in HIV-1 chemotherapy: The binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease. *Protein Sci.* **11**, 1908–1916

98. Croteau, G., Doyon, L., Thibeault, D., Mckercher, G., Pilote, L., Croteau, G., Doyon, L., Thibeault, D., Kercher, G. M. C., Pilote, L., and Lamarre, D. (1997) Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors . Impaired Fitness of Human Immunodeficiency Virus Type 1 Variants with High-Level Resistance to Protease Inhibitors. *J. Virol.* **71**, 1089–1096

99. Mammano, F., Trouplin, V., Zennou, V., and Clavel, F. (2000) Retracing the evolutionary pathways of human immunodeficiency virus type 1 resistance to protease inhibitors: virus fitness in the absence and in the presence of drug. *J. Virol.* **74**, 8524–8531

100. Nijhuis, M., Schuurman, R., De Jong, D., Erickson, J., Gustchina, E., Albert, J., Schipper, P., Gulnik, S., and Boucher, C. A. B. (1999) Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *Aids*. **13**, 2349–2359

101. Hirsch, M. S., Brun-Vezinet, F., D'Aquila, R. T., Hammer, S. M., Johnson, V. a, Kuritzkes, D. R., Loveday, C., Mellors, J. W., Clotet, B., Conway, B., Demeter, L. M., Vella, S., Jacobsen, D. M., and Richman, D. D. (2000) Antiretroviral drug resistance testing in adult HIV-1 infection: recommendations of an International AIDS Society-USA Panel. *J. Am. Med. Assoc.* **283**, 2417–2426

102. Dumans, A. T., Soares, M. A., Pieniazek, D., Kalish, M. L., Vroey, V. De, Hertogs, K., and Tanuri, A. (2002) Prevalence of Protease and Reverse Transcriptase Drug Resistance Mutations over Time in Drug-Naïve Human Immunodeficiency Virus Type 1-Positive Individuals in Rio de Janeiro , Brazil Prevalence of Protease and Reverse Transcriptase Drug Resistance Mutatio. *Antimicrob. Agents Chemother.* 10.1128/AAC.46.9.3075

103. Randolph, J., and DeGoey, D. (2004) Peptidomimetic Inhibitors of HIV Protease. *Curr. Top. Med. Chem.* **4**, 1079–1095

104. Mammano, F., Petit, C., and Clavel, F. (1998) Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *J. Virol.* **72**, 7632–7

105. Martinez-Picado, J., Savara, A. V, Sutton, L., and D'Aquila, R. T. (1999) Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *J Virol*. **73**, 3744–3752

106. Muzammil, S., Ross, P., and Freire, E. (2003) A major role for a set of non-active site mutations in the development of HIV-1 protease drug resistance. *Biochemistry*. **42**, 631–638

107. Sugiura, W., Matsuda, Z., Yokomaku, Y., Hertogs, K., Larder, B., Oishi, T., Okano, A., Shiino, T., Tatsumi, M., Matsuda, M., Abumi, H., Takata, N., Shirahata, S., Yamada, K., Yoshikura, H., and Nagai, Y. (2002) Interference between D30N and L90M in selection and development of protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* **46**, 708–715

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

109. Olsen, D. B., Stahlhut, M. W., Rutkowski, C. A., Schock, H. B., Aimee, L., and Kuo, L. C. (1999) Non-active Site Changes Elicit the HIV-1 Protease to Inhibitors. *J. Biol. Chem.* **274**, 23699–23701

110. Kim, E., Winters, M., Kagan, R., and Merigan, T. (2001) Functional correlates of insertion mutations in

the protease gene of human immunodeficiency virus type 1 isolates from patients. *J. Virol.* **75**, 11227–11233

111. Sturmer, M., Staszewski, S., Doerr, H. W., and Hertogs, K. (2003) A 6-Base Pair Insertion in the Protease Gene of HIV Type 1 Detected in a Protease Inhibitor-Naive Patient Is Not Associated with Indinavir Treatment Failure. *AIDS Res. Hum. Retroviruses*. **19**, 967–968

112. Winters, M. A., Kagan, R. M., Heseltine, P. N. R., and Merigan, T. C. (2005) New Two-Amino Acid Insertion Near Codon 70 of the HIV Type 1 Protease Gene. *AIDS Res. Hum. Retroviruses*. **21**, 311–313

113. Winters, M. a, and Merigan, T. C. (2005) MINIREVIEW Insertions in the Human Immunodeficiency Virus Type 1 Protease and Reverse Transcriptase Genes : Clinical Impact and Molecular Mechanisms. *Society*. **49**, 2575–2582

114. Bebenek, K., Abbotts, J., Roberts, J. D., Wilson, S. H., and Kunkel, T. a (1989) Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. *J. Biol. Chem.* **264**, 16948–56

115. Harrison, G. P., Mayo, M. S., Hunter, E., and Lever, A. M. L. (1998) Pausing of reverse transcriptase on retroviral RNA templates is influenced by secondary structures both 5' and 3' of the catalytic site. *Nucleic Acids Res.* **26**, 3433–3442

116. Amiel, C., Charpentier, C., Désiré, N., Bonnard, P., Lebrette, M. G., Weiss, L., Pialoux, G., and Schneider, V. (2011) Long-term follow-up of 11 protease inhibitor (PI)-naïve and PI-treated HIV-infected patients harbouring virus with insertions in the HIV-1 protease gene. *HIV Med.* **12**, 138–144

117. Adamson, C. S. (2012) Protease-Mediated Maturation of HIV: Inhibitors of Protease and the Maturation Process. *Mol. Biol. Int.* **2012**, 1–13

118. Kozísek, M., Sasková, K. G., Rezácová, P., Brynda, J., van Maarseveen, N. M., De Jong, D., Boucher, C. a, Kagan, R. M., Nijhuis, M., and Konvalinka, J. (2008) Ninety-nine is not enough: molecular characterization of inhibitor-resistant human immunodeficiency virus type 1 protease mutants with insertions in the flap region. *J. Virol.* **82**, 5869–78

119. Africa, D. of H. S. (2010) CLINICAL GUIDELINES : PMTCT (Prevention of Mother-to- Child Transmission) National Department of Health , South Africa ;

120. Terpe, K. (2003) Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **60**, 523–533

121. Hefti, M. H., Van Vugt-Van Der Toorn, C. J. G., Dixon, R., and Vervoort, J. (2001) A novel purification method for histidine-tagged proteins containing a thrombin cleavage site. *Anal. Biochem.* **295**, 180–185

122. Kapust, R. B., and Waugh, D. S. (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668–1674

123. LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F., and McCoy, J. M. (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. *Bio/Technology*. **11**, 187–193

124. Smith, D. B., and Johnson, K. S. (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene*. **67**, 31–40

125. Rosano, G. L., and Ceccarelli, E. A. (2014) Recombinant protein expression in Escherichia coli : advances and challenges. *Front. Microbiol.* **5**, 1–17

126. Zhang, Y. B., Howitt, J., McCorkle, S., Lawrence, P., Springer, K., and Freimuth, P. (2004) Protein aggregation during overexpression limited by peptide extensions with large net negative charge.

Protein Expr. Purif. **36**, 207–216

127. Ahn, J. Y., Choi, H., Kim, Y., Han, K., Park, J., Han, S., and Lee, J. (2005) Heterologous gene expression using self-assembled supra-molecules with high affinity for HSP70 chaperone. *Nucleic Acids Res.* **33**, 3751–3762

128. Fox, J., Routzahn, K. M., M.H., B., and Waugh, D. S. (2000) Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. *FEBS Lett.* **483**, 6–10

129. Shaki-Loewenstein, S., Zfania, R., Hyland, S., Wels, W. S., and Benhar, I. (2005) A universal strategy for stable intracellular antibodies. *J. Immunol. Methods*. **303**, 19–39

130. Bach, H., Mazor, Y., Shaky, S., Shoham-Lev, A., Berdichevsky, Y., Gutnick, D. L., and Benhar, I. (2001) Escherichia coli maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. *J. Mol. Biol.* **312**, 79–93

131. Kim, C. W., Han, K. S., Ryu, K.-S., Kim, B. H., Kim, K.-H., Choi, S. Il, and Seong, B. L. (2007) N-terminal domains of native multidomain proteins have the potential to assist de novo folding of their downstream domains in vivo by acting as solubility enhancers. *Protein Sci.* **16**, 635–43

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

133. Richards, A. D., Phylip, L. H., Farmerie, W. G., Scarborough, P. E., Alvarez, A., Dunn, B. M., Hirel, P., Konvalinka, J., Kostka, V., and Kay, J. (1990) Sensitive, Substrates Soluble Chromogenic for HIV- 1 Proteinase. *J. Biol. Chem.* **265**, 7733–7736

134. Szeltner, Z., and Polgár, L. (1996) Conformational stability and catalytic activity of HIV-1 protease are both enhanced at high salt concentration. *J. Biol. Chem.* **271**, 5458–5463

135. Rose, J. R., Salto, R., and Craik, C. S. (1993) Regulation of Autoproteolysis of the HIV-1. *J. Biol. Chem.* **268**, 11939–11945

136. Mildner, A. M., Rothrock, D. J., Leone, J. W., Bannow, C. A., Lull, J. M., Reardon, I. M., Sarcich, J. L., Smith, C. W., Heinrikson, R. L., Tomasselli, A. G., Howe, W. J., and Tomich, C. S. C. (1994) The HIV-1 Protease as Enzyme and Substrate: Mutagenesis of Autolysis Sites and Generation of a Stable Mutant with Retained Kinetic Properties. *Biochemistry*. **33**, 9405–9413

137. McCammon, J. A., Gelin, B. R., and Karplus, M. (1977) Dynamics of folded proteins. *Nature*. **267**, 585–590

138. York, D. M., Darden, T. A., Pedersen, L. G., and Anderson, M. W. (1993) Molecular Dynamics Simulation of HIV-1 Protease in a Crystalline Environment and in Solution. *Biochemistry*. **32**, 1443–1453

139. Collins, J. R., Burt, S. K., and Erickson, J. W. (1995) Flap opening in HIV-1 protease simulated by 'activated' molecular dynamics. *Nat. Struct. Biol.* **2**, 334–338

140. Ishima, R., Freedberg, D. I., Wang, Y. X., Louis, J. M., and Torchia, D. A. (1999) Flap opening and dimer-interface flexibility in the free and inhibitor- bound HIV protease, and their implications for function. *Structure*. **7**, 1047–1055

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

142. Zondagh, J., Balakrishnan, V., Achilonu, I., Dirr, H. W., and Sayed, Y. (2018) The N37T ↑ V hinge region insertion mutation alters the Molecular dynamic landscape of HIV - 1 Subtype C protease Molecular Dynamics and Ligand Docking of a Hinge Region Variant of South African HIV - 1 Subtype - C protease. *J. Mol. Graph. Model.* **82**, 1–11

143. Louis, J. M., and Roche, J. (2016) Evolution under drug pressure remodels the folding free-energy landscape of mature HIV-1 protease. *J. Mol. Biol.* **21**, 129–139

144. Louis, J. M., and Roche, J. (2016) Evolution under Drug Pressure Remodels the Folding Free-Energy Landscape of Mature HIV-1 Protease. *J. Mol. Biol.* **428**, 2780–2792

145. Carlson, H. A., and Mccammon, J. A. (2000) Accommodating Protein Flexibility in Computational Drug. *Mol. Pharmacol.* **218**, 213–218

146. Teague, S. J. (2003) Implications of Protein Flexibility for Drug Discovery. *Nat. Rev.* **2**, 527–541

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

148. Weber, I. T., and Robert, W. (2017) Decoding HIV resistance: from genotype to therapy. *Future Med. Chem.* **9**, 1529–1538

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

150. Wensing, A. M., Calvez, V., Gunthard, H. F., Johnson, V. A., Paredes, R., Pillay, D., Shafer, R. W., and Richman, D. D. (2017) 2017 Update of the Drug Resistance Mutations in HIV-1. *Top. Antivir. Med.* **24**, 132–133

151. Shafer, R. W., Jung, D. R., and Betts, B. J. (2000) Human immunodeficiency virus type 1 reverse transcriptase and protease mutation search engine for queries. *Nat. Med.* **6**, 1290–2

152. Rhee, S.-Y., Gonzales, M. ., Kantor, R., Betts, B. ., Ravela, J., and Shafer, R. . (2003) Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **31**, 298–303

153. Rhee, S.-Y., Kantor, R., Katzemstein, D. A., Camacho, R., Morris, L., Sirivichayakul, S., Jorgensen, L., Brigido, L. F., Schapiro, J. M., Shafer, R. W., and Group, I. N. S. B. H.-1 W. (2006) HIV-1 pol mutation frequence by subtype and treatment experience: extension of the HIVseq program to seven non-B subtypes. *AIDS*. **20**, 643–651

154. Grantz Saskova, K., Kozisek, M., Stray, K., de Jong, D., Rezacova, P., Brynda, J., van Maarseveen, N. M., Nijhuis, M., Cihlar, T., and Konvalinka, J. (2014) GS-8374, a Prototype Phosphonate-Containing Inhibitor of HIV-1 Protease, Effectively Inhibits Protease Mutants with Amino Acid Insertions. *J. Virol.* **88**, 3586–3590

155. Gatanaga, H., Suzuki, Y., Tsang, H., Yoshimura, K., Kavlick, M. F., Nagashima, K., Gorelick, R. J., Mardy, S., Tang, C., Summers, M. F., and Mitsuya, H. (2002) Amino Acid Substitutions in Gag Protein at Non-cleavage Sites Are Indispensable for the Development of a High Multitude of HIV-1 Resistance against Protease Inhibitors *. *J. Biol. Chem.* **277**, 5952–5961

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

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

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

159. Martins, A. N., Arruda, M. B., Pires, A. F., Tanuri, A., and Brindeiro, R. M. (2011) Accumulation of P(T/S)AP Late Domain Duplications in HIV Type 1 Subtypes B, C, and F Derived from Individuals Failing ARV Therapy and ARV Drug-Naive Patients. *AIDS Res. Hum. Retroviruses*. **27**, 687–692

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

161. Martins, A. N., Waheed, A. A., Ablan, S. D., Huang, W., Newton, A., Petropoulos, C. J., Brindeiro, R. de M., and Freed, E. O. (2015) Elucidation of the Molecular Mechanism Driving Duplication of the Hiv-1 Ptap Late Domain. *J. Virol.* **90**, 768–119

162. Gallego, O., Mendoza, C. De, Corral, A., and Soriano, V. (2003) Changes in the Human Immunodeficiency Virus p7-p1-p6 gag Gene in Drug-Naive and Pretreated Patients Changes in the Human Immunodeficiency Virus p7-p1-p6 gag Gene in Drug-Naive and Pretreated Patients. **41**, 5–8

163. Marlowe, N., Flys, T., Hackett, J., Schumaker, M., Jackson, J. B., and Eshleman, S. H. (2004) Analysis of insertions and deletions in the gag p6 region of diverse HIV type 1 strains. *AIDS Res. Hum. Retroviruses*. **20**, 1119–25

164. Flys, T., Marlowe, N., Hackett, J., Parkin, N., Schumaker, M., Holzmayer, V., Hay, P., and Eshleman, S. H. (2005) Analysis of PTAP Duplications in the gag p6 Region of Subtype C HIV Type 1. *AIDS Res. Hum. Retroviruses*. **21**, 739–741

165. Patil, A., and Bhattacharya, J. (2012) Natural deletion of L35Y36 in p6 gag eliminate LYPXnL/ALIX auxiliary virus release pathway in HIV-1 subtype C. *Virus Res.* **170**, 154–158

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

348481:Inline_Submission

ORIGINALITY REPORT

8%	3%	7%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Naicker, Previn, Stoyan Stoychev, Heini W. Dirr, and Yasien Sayed. "Amide hydrogen exchange in HIV-1 subtype B and C proteases - insights into reduced drug susceptibility and dimer stability", FEBS Journal, 2014. Publication	<1%
2	www.who.int Internet Source	<1%
3	Thabiso N. Tshabalala, Mihai-Silviu Tomescu, Allan Prior, Vijayakumar Balakrishnan, Yasien Sayed, Heini W. Dirr, Ikechukwu Achilonu. "Energetics of Glutathione Binding to Human Eukaryotic Elongation Factor 1 Gamma: Isothermal Titration Calorimetry and Molecular Dynamics Studies", The Protein Journal, 2016 Publication	<1%
4	Naicker, Previn, Palesa Seele, Heini W. Dirr, and Yasien Sayed. "F99 is Critical for Dimerization and Activation of South African HIV-1 Subtype C Protease", The Protein	<1%

Journal, 2013.
Publication

5	Giandhari, Jennifer, Adriaan E. Basson, Katherine Sutherland, Chris M. Parry, Patricia A. Cane, Ashraf Coovadia, Louise Kuhn, Gillian Hunt, and Lynn Morris. "Contribution of Gag and Protease to HIV-1 Phenotypic Drug Resistance in Pediatric Patients Failing Protease Inhibitor-Based Therapy", Antimicrobial Agents and Chemotherapy, 2016. Publication	<1%
6	Bell, Neil M., and Andrew M.L. Lever. "HIV Gag polyprotein: processing and early viral particle assembly", Trends in Microbiology, 2013. Publication	<1%
7	Adamson, Catherine S.. "Protease-Mediated Maturation of HIV: Inhibitors of Protease and the Maturation Process", Molecular Biology International, 2012. Publication	<1%
8	www.ncbi.nlm.nih.gov Internet Source	<1%
9	www.freepatentsonline.com Internet Source	<1%
10	Ammaranond, P.. "Rates of transmission of antiretroviral drug resistant strains of HIV-1", Journal of Clinical Virology, 200302	<1%

	Publication	
11	Clavel, FranÃ§ois, and Fabrizio Mammano. "Role of Gag in HIV Resistance to Protease Inhibitors", Viruses, 2010. Publication	<1%
12	Maseko, Sibusiso B., Satheesh Natarajan, Vikas Sharma, Neelakshi Bhattacharyya, Thavendran Govender, Yasien Sayed, Glenn E.M. Maguire, Johnson Lin, and Hendrik G. Kruger. "Purification and characterization of naturally occurring HIV-1 (South African subtype C) protease mutants from inclusion bodies", Protein Expression and Purification, 2016. Publication	<1%
13	www.mdpi.com Internet Source	<1%
14	link.springer.com Internet Source	<1%
15	www.ccsenet.org Internet Source	<1%
16	Adriaan E. Basson, Soo-Yon Rhee, Chris M. Parry, Ziad El-Khatib et al. "Impact of Drug Resistance-Associated Amino Acid Changes in HIV-1 Subtype C on Susceptibility to Newer Nonnucleoside Reverse Transcriptase	<1%

Inhibitors", Antimicrobial Agents and Chemotherapy, 2015
Publication

17	Federica Volontè. "Optimizing HIV-1 protease production in Escherichia coli as fusion protein", Microbial Cell Factories, 2011 Publication	<1%
18	M. A. Winters. "Insertions in the Human Immunodeficiency Virus Type 1 Protease and Reverse Transcriptase Genes: Clinical Impact and Molecular Mechanisms", Antimicrobial Agents and Chemotherapy, 07/01/2005 Publication	<1%
19	Lockhat, Husain A., José R. A. Silva, Cláudio N. Alves, Thavendran Govender, Jerônimo Lameira, Glenn E. M. Maguire, Yasien Sayed, and Hendrik G. Kruger. "Binding free energy calculations of nine FDA-approved protease inhibitors against HIV-1 subtype C I36T↑T containing 100 amino acids per monomer", Chemical Biology & Drug Design, 2015. Publication	<1%
20	Umar Muzaffer, V. I. Paul, N. Rajendra Prasad. "Molecular docking of selected phytoconstituents with signaling molecules of Ultraviolet-B induced oxidative damage", In Silico Pharmacology, 2017 Publication	<1%

21	Adrian Velazquez-Campoy. "Structural and Thermodynamic Basis of Resistance to HIV-1 Protease Inhibition: Implications for Inhibitor Design", Current Drug Targets - Infectious Disorders, 12/01/2003 Publication	<1%
22	Pawar, S.A.. "Synthesis and molecular modelling studies of novel carbapeptide analogs for inhibition of HIV-1 protease", European Journal of Medicinal Chemistry, 201207 Publication	<1%
23	F. Clavel. "HIV Drug Resistance", New England Journal of Medicine, 3/4/2004 Publication	<1%
24	Salman Muzammil, Patrick Ross, Ernesto Freire. " A Major Role for a Set of Non-Active Site Mutations in the Development of HIV-1 Protease Drug Resistance ", Biochemistry, 2003 Publication	<1%
25	Bhuvaneshwari Mahalingam. "Combining mutations in HIV-1 protease to understand mechanisms of resistance", Proteins Structure Function and Genetics, 07/01/2002 Publication	<1%

science.bard.edu

26	Internet Source	<1%
27	"Substrate-Envelope-Guided Design of Drugs with a High Barrier to the Evolution of Resistance", Handbook of Antimicrobial Resistance, 2014. Publication	<1%
28	Advances in Experimental Medicine and Biology, 1991. Publication	<1%
29	Raquel Blanco, Luis Carrasco, Iván Ventoso. "Cell Killing by HIV-1 Protease", Journal of Biological Chemistry, 2003 Publication	<1%
30	tampub.uta.fi Internet Source	<1%
31	John M. Louis, Lalit Deshmukh, Jane M. Sayer, Annie Aniana, G. Marius Clore. "Mutations Proximal to Sites of Autoproteolysis and the α-Helix That Co-evolve under Drug Pressure Modulate the Autoprocessing and Vitality of HIV-1 Protease", Biochemistry, 2015 Publication	<1%
32	Junko Shibata, Wataru Sugiura, Hirotaka Ode, Yasumasa Iwatani et al. "Within-host co-evolution of Gag P453L and protease	<1%

	D30N/N88D demonstrates virological advantage in a highly protease inhibitor-exposed HIV-1 case", Antiviral Research, 2011 Publication	
33	Francine E. McCutchan. "Global epidemiology of HIV", Journal of Medical Virology, 2006 Publication	<1 %
34	www.cabsa.org.za Internet Source	<1 %
35	Molla, A.. "Recent developments in HIV protease inhibitor therapy", Antiviral Research, 199807 Publication	<1 %
36	doc.utwente.nl Internet Source	<1 %
37	Mildner, Ana M., Donna J. Rothrock, Joseph W. Leone, Carol A. Bannow, June M. Lull, Ilene M. Reardon, Jean L. Sarcich, W. Jeffrey Howe, and Che-Shen C. Tomich. "The HIV-1 Protease as Enzyme and Substrate: Mutagenesis of Autolysis Sites and Generation of a Stable Mutant with Retained Kinetic Properties", Biochemistry, 1994. Publication	<1 %
38	Turner, B.G.. "Structural biology of HIV", Journal of Molecular Biology, 19990108 Publication	<1 %

39	lib.dr.iastate.edu Internet Source	<1%
40	www.dndi.org Internet Source	<1%
41	Edison Castro, Zachary S. Martinez, Chang-Soo Seong, Andrea Cabrera-Espinoza et al. " Characterization of New Cationic , -Dimethyl[70]fulleropyrrolidinium Iodide Derivatives as Potent HIV-1 Maturation Inhibitors ", Journal of Medicinal Chemistry, 2016 Publication	<1%
42	bit.edu.pl Internet Source	<1%
43	mcb.asm.org Internet Source	<1%
44	www.ideals.illinois.edu Internet Source	<1%
45	Copeland, . "Tight Binding Inhibition", Evaluation of Enzyme Inhibitors in Drug Discovery A Guide for Medicinal Chemists and Pharmacologists, 2013. Publication	<1%
46	www.intechopen.com Internet Source	<1%

47	Jean-Marie Zimmer. "Impact on Replicative Fitness of the G48E Substitution in the Protease of HIV-1", JAIDS Journal of Acquired Immune Deficiency Syndromes, 07/2008 Publication	<1%
48	www.jbc.org Internet Source	<1%
49	Gary J. Robertson, Stoyan H. Stoychev, Yasien Sayed, Ikechukwu Achilonu, Heini W. Dirr. "The effects of mutating Tyr9 and Arg15 on the structure, stability, conformational dynamics and mechanism of GSTA3-3", Biophysical Chemistry, 2017 Publication	<1%
50	blanco.biomol.uci.edu Internet Source	<1%
51	Ali, Akbar, Rajintha M. Bandaranayake, Yufeng Cai, Nancy M. King, Madhavi Kolli, Seema Mittal, Jennifer F. Murzycki, Madhavi N.L. Nalam, Ellen A. Nalivaika, AyÅŸegÃ¼l Ã–zen, Moses M. Prabu-Jeyabalan, Kelly Thayer, and Celia A. Schiffer. "Molecular Basis for Drug Resistance in HIV-1 Protease", Viruses, 2010. Publication	<1%
52	Hermans, Sabine Boulle, Andrew Caldwell,. "Temporal trends in TB notification rates during	<1%

	ART scale-up in Cape Town: an ecological analysis.(R", Journal of the International AIDS Societ, Jan 2015 Issue Publication	
53	Dwivedi, Shailendra, Suraj Singh Yadav, Manish Kumar Singh, Shailja Shukla, Sanjay Khattri, and Kamlesh Kumar Pant. "Pharmacogenomics of Viral Diseases", Omics for Personalized Medicine, 2013. Publication	<1%
54	www.theglobalfund.org Internet Source	<1%
55	www.researchgate.net Internet Source	<1%
56	www.jneuroinflammation.com Internet Source	<1%
57	www.tandfonline.com Internet Source	<1%
58	birc.au.dk Internet Source	<1%
59	sentebale.org Internet Source	<1%
60	D. S. Dauber. "Altered Substrate Specificity of Drug-Resistant Human Immunodeficiency Virus Type 1 Protease", Journal of Virology,	<1%

02/01/2002
Publication

61	Xu, Meng, Wenfan Xie, and Minren Huang. "Overexpression of PeRHD3 alters the root architecture in Populus", Biochemical and Biophysical Research Communications, 2012. Publication	<1%
62	repository.kulib.kyoto-u.ac.jp Internet Source	<1%
63	Monique Nijhuis. "HIV protease resistance and viral fitness", Current Opinion in HIV and AIDS, 03/2007 Publication	<1%
64	Byoungkook Kim. "Solution structure and backbone dynamics of the XPC-binding domain of the human DNA repair protein hHR23B", FEBS Journal, 5/2005 Publication	<1%
65	Kang, Yun (Kenneth) Ng, Stanley Lee, Jul. "Development of an alternative monoclonal antibody polishing step.(Antibody Polishing)", Biopharm International, May 2012 Issue Publication	<1%
66	Huang, L.. "Production of bioactive human beta-defensin 5 and 6 in Escherichia coli by soluble fusion expression", Protein Expression and Purification, 200810	<1%

	Publication	
67	intl-jvi.asm.org Internet Source	<1%
68	eprints.maynoothuniversity.ie Internet Source	<1%
69	www.chinafrica.cn Internet Source	<1%
70	scholar.sun.ac.za Internet Source	<1%
71	jci.org Internet Source	<1%
72	van Maarseveen, N.M.. "A novel real-time PCR assay to determine relative replication capacity for HIV-1 protease variants and/or reverse transcriptase variants", Journal of Virological Methods, 200605 Publication	<1%
73	rutchem.rutgers.edu Internet Source	<1%
74	tubdok.tub.tuhh.de Internet Source	<1%
75	Rossenkhan, R., I. J. MacLeod, T. K. Sebunya, E. Castro-Nallar, M. F. McLane, R. Musonda, B. A. Gashe, V. Novitsky, and M. Essex. "tat Exon 1 Exhibits Functional Diversity during HIV-1	<1%

	Subtype C Primary Infection", Journal of Virology, 2013. Publication	
76	www.j3.jstage.jst.go.jp Internet Source	<1%
77	Bottcher, J.. "Structural and Kinetic Analysis of Pyrrolidine-Based Inhibitors of the Drug-Resistant Ile84Val Mutant of HIV-1 Protease", Journal of Molecular Biology, 20081107 Publication	<1%
78	Zhao, Li, Pupu Zhang, Shiyang Long, Linlin Wang, Hanyong Jin, Weiwei Han, and Pu Tian. "Investigating the substrate binding mechanism of sulfotransferase 2A1 based on substrate tunnel analysis: a molecular dynamics simulation study", Journal of Molecular Modeling, 2016. Publication	<1%
79	www.nature.com Internet Source	<1%
80	Federico Pulido. "Noninferiority and lopinavir/ritonavir monotherapy trials", AIDS, 08/2008 Publication	<1%
81	www.proteinsocthai.net Internet Source	<1%

82	docserv.uni-duesseldorf.de Internet Source	<1%
83	Angela M. Gronenborn. Nature Structural Biology, 09/01/1999 Publication	<1%
84	Kolli, M., A. Ozen, N. Kurt-Yilmaz, and C. A. Schiffer. "HIV-1 PROTEASE SUBSTRATE CO-EVOLUTION IN NELFINAVIR RESISTANCE", Journal of Virology, 2014. Publication	<1%
85	van Maarseveen, Noortje M., Annemarie M. J. Wensing, Dorien de Jong, Maaike Taconis, Jan C. C. Borleffs, Charles A. B. Boucher, and Monique Nijhuis. "Persistence of HIV-1 Variants with Multiple Protease Inhibitor (PI)–Resistance Mutations in the Absence of PI Therapy Can Be Explained by Compensatory Fixation", The Journal of Infectious Diseases, 2007. Publication	<1%
86	Journal of Science and Technology Policy Management, Volume 7, Issue 3 (2016) Publication	<1%
87	Dieter Hofmann. "Detailed-atomistic molecular modeling of small molecule diffusion and solution processes in polymeric membrane materials", Macromolecular Theory and Simulations, 07/01/2000	<1%

Publication

88	Toor, J.S.. "Prediction of drug-resistance in HIV-1 subtype C based on protease sequences from ART naive and first-line treatment failures in North India using genotypic and docking analysis", Antiviral Research, 201111 Publication	<1%
89	"UNAIDS WELCOMES UNITED STATES OF AMERICA'S ANNOUNCEMENT OF A NEW US$ 100 MILLION INVESTMENT FUND TO ", States News Service, June 9 2016 Issue Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		