2016-12-13

# The Structural Basis for the Interdependence of Drug Resistance in the HIV-1 Protease

Debra A. Ragland
*University of Massachusetts Medical School*

# THE STRUCTURAL BASIS FOR THE INTERDEPENDENCE OF DRUG RESISTANCE IN THE HIV-1 PROTEASE

A Dissertation Presented

By

DEBRA ANN RAGLAND

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

DECEMBER 13, 2016

INTERDISCIPLINARY GRADUATE PROGRAM

THE STRUCTURAL BASIS FOR THE INTERDEPENDENCE OF DRUG RESISTANCE IN THE HIV-1 PROTEASE

A Dissertation Presented By

DEBRA ANN RAGLAND

This work was undertaken in the Graduate School of Biomedical Sciences
INTERDISCIPLINARY GRADUATE PROGRAM

The signature of the Thesis Advisor signifies
validation of Dissertation content

---

CELIA A. SCHIFFER, PH.D., THESIS ADVISOR

The signatures of the of the Dissertation Defense Committee signify
Completion and approval as to style and content of the Dissertation

---

WILLIAM ROYER, PH.D., Member of Committee

---

KONSTANTIN ZELDOVICH, PH.D., Member of Committee

---

JEREMY LUBAN, M.D., Member of Committee

---

BRUCE TORBETT, PH.D., External Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation
meets the requirements of the Dissertation Committee

---

DAN BOLON, PH.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences
signifies that the student has met all graduation requirements of the school

---

ANTHONY CARRUTHERS, PH.D.,
Dean of the Graduate School of Biomedical Sciences
DECEMBER 13, 2016

# Dedication

To Josiah, someday you'll read this and hopefully be proud of your mom.

To Jean, my biggest supporter and sole caretaker, I held on and it only took a minute.

To all my childhood friends who also lost their parents to AIDS related illnesses, how lucky we were to have each other.

And last but not least,
To Bridgett, this is the apex of our lifelong journey together and although you won't be able to read this, I hope it makes you proud.

## Acknowledgements

I would like to personally thank several people who have helped me throughout my graduate career be it via scientific input or kindness and friendship. I do not have family support so there were many (many, many) days that I just wanted to give up but I was encouraged to keep going.

First, I would like to thank Ellen Nalivaika, Dr. Anthony Cura and Dr. Jay Sage for all your help with various scientific endeavors and writing. I would not have made it thru my first two years in the laboratory without your help. Your support was invaluable and there was never a dull moment at lunch.

Next, I would like to thank Dr. Hong Cao for all of your help with enzyme kinetics and the personal time and effort you put into teaching me enzymology. I would also like to thank Dr. Akbar Ali for all of your kindness and willingness to help answer any questions I had. And Dr. Nese Kurt-Yilmaz for translating my writing into English and helping with various tasks along the way. Drs. Yufeng Cai, Shivender Shandilya, Aysegül Ozen, Woody Sherman and the staff at Schrödinger, Janet Paulsen and Kristina Prachanranarong, without you all I'd probably still run from the command line or have a meltdown at the every error message.

Thank you to Troy Whitfield, for without your help, I'd probably never be able to finish graduate school. I have learned a tremendous amount from you about statistics and machine learning. And I simply would not have made it through the second leg of this thesis without your insight. I would like to thank my committee members, Dr. William Royer, Dr. Konstantin Zeldovich, Dr. Jeremy

Luban, Dr. Dan Bolon and Dr. Bruce Torbett for their time and dedication to ensuring my success as a scientist. I would also like to thank Dr. Swanstrom at UNC-Chapel Hill and the members of his laboratory insight for all of their help with experiments, writing, training and resources necessary for me to advance my research.

To Dr. Madhavi Kolli, my first mentor in graduate school, thank you for all of your kindness and training and warmth and friendship. I would also like to thank my UMass family of friends, faculty and admins who always had the ability to make my days just a little brighter and encouraged me to push through. In no particular order; Fransenio Clark, Ashley Matthew, Asia Matthew, Dr. Brian Lewis, Dr. Djade Soumana, Alan Lucia, Karen Welch, Irene Couture, Dr. Caroline Duffy, Dr. Nick Rhind and Dr. Mary Munson. And I would like to thank my general "family of friends" who have supported me over the last six years through all the ups and downs, again, in no particular order; Raymond Beamon, Latisha Scarborough, Latia Franks, Kristen Chambers, Jamie King, Dr. Darkus Jenkins, Portia Mason, Dr. Carlos Crawford, Dr. Alfred Faulkener, Dr. Ashalla Freeman, and my sister Sharika Hall.

I would like to thank Dr. Lesley Brown, who took the initiative to be my mentor, offering both scientific and general life and career advice since the start of my summer internship in 2010. I would also like to thank Dr. Robert Siliciano and Dr. Maame Sampah, who helped me determine that science was something that I could do and for your scientific correspondence and support. Drs. Kai McKinstry and Tara Strutt, without the two of you an your wonderful mentoring

and guidance through my rotation, I don't know where I would be. Dr. Zerihun Assefa and Dr Mary Smith for all your guidance and training throughout my scientific career.

Finally, I would like to thank my thesis advisor Dr. Celia Schiffer. It has been a pleasure to work for you these last few years. I am so grateful to have been apart of your laboratory even though I told you I was not coming back initially, you have always been understanding. Your kindness and compassion for your students helps us all in ways you'll probably never know I feel like we've been through everything together. Without you, I would not have been able to fulfill my dream of becoming a scientist. Thank you.

# Abstract

The human immunodeficiency virus type 1 (HIV-1) protease (PR) is a critical drug target as it is responsible for virion maturation. Mutations within the active site (1°) of the PR directly interfere with inhibitor binding while mutations distal to the active site (2°) to restore enzymatic fitness. Increasing mutation number is not directly proportional to the severity of resistance, suggesting that resistance is not simply additive but that it is interdependent. The interdependency of both primary and secondary mutations to drive protease inhibitor (PI) resistance is grossly understudied.

To structurally and dynamically characterize the direct role of secondary mutations in drug resistance, I selected a panel of single-site mutant protease crystal structures complexed with the PI darunavir (DRV). From these studies, I developed a *network hypothesis* that explains how mutations outside the active site are able to perpetuate changes to the active site of the protease to disrupt inhibitor binding.

I then expanded the panel to include highly mutated multi-drug resistant variants. To elucidate the interdependency between primary and secondary mutations I used statistical and machine-learning techniques to determine which specific mutations underlie the perturbations of key inter-molecular interactions. From these studies, I have determined that mutations distal to the active site are able to perturb the global PR hydrogen bonding patterns, while primary and secondary mutations cooperatively perturb hydrophobic contacts between the PR and DRV. Discerning and exploiting the mechanisms that underlie drug

resistance in viral targets could proactively ameliorate both current treatment and inhibitor design for HIV-1 targets.

# Table of Contents

# List of Tables

# List of Figures

# List of Third Party Copyrighted Material

**Figure 1.2** Reprinted by permission from Macmillan Publishers Ltd: Nature, Volume 410, Issue 6831, pages 963-967. Robin A. Weiss. Gulliver's travels in HIVland. Copyright 2001, with permission from Nature Publishing Group (License Number: 3975920680164).

**Figure 1.3** Reprinted with permission from Cold Spring Harbor Laboratory Press. Cold Spring Harbor Perspectives in Medicine, Volume 2, Issue 7, pages 1-24. Wesley I. Sundquist and Hans-Georg Kräusslich. HIV-1 Assembly, Budding, and Maturation. Copyright 2012.

**Figure 1.5** Reproduced from Organic & Biomolecular Chemistry, Volume 1, Issue 1, pages 5-14. Ashraf Brik and Chi-Huey Wong. HIV-1 protease: mechanism and drug discovery. Copyright 2003, with permission from the Royal Society of Chemistry (License Number: 3990140566774)

**Figure 1.6** Reprinted by permission from Macmillan Publishers Ltd: Nature, Volume 410, Issue 6831, pages 995-1001. Douglas D. Richman. HIV chemotherapy. Copyright 2001, with permission from Nature Publishing Group (License Number: 3975920831848).

**Figure 4.1** Adapted from Elsevier Chemistry and Biology, Volume 10, Issue 9, pages 787-797. Amy C. Anderson. The Process of Structure-Based Drug Design. Copyright 2003, with permission of Elsevier.

# List of Abbreviations

| | |
|---|---|
| HIV-1 | Human Immunodeficiency Virus Type 1 |
| AIDS | Acquired Immune Deficiency Syndrome |
| PR | Protease |
| PI | Protease Inhibitor |
| DRV | Darunavir |
| MD | Molecular Dynamics |
| RMSD | Root Mean Square Deviation |
| RMSF | Root Mean Square Fluctuation |
| PCA | Principal Components Analysis |
| vdW | van der Waals |
| PDB | Protein Data Bank |
| WT | Wild Type |
| KI | Inhibition Constant |
| $IC_{50}$ | half maximal inhibitory concentration |
| $EC_{50}$ | half maximal effective concentration |

# Preface

Chapter II is a collaborative study that has been previously published as:

Debra A. Ragland, Ellen A. Nalivaika, Madhavi N. L. Nalam, Kristina L. Prachanronarong, Hong Cao, Rajintha M. Bandaranayake, Yufeng Cai, Nese Kurt-Yilmaz, and Celia A, Schiffer. Drug Resistance Conferred by Mutations Outside the Active Site through Alterations in the Dynamic and Structural Ensemble of HIV-1 Protease. Journal of the American Chemical Society. 2014, 136, 11956-11963. DOI: 10.1021/ja504096m

Author Contributions: D.A.R., N.K.Y., and C.A.S designed research; D.A.R., E.A.N., M.N.L.N and R.M.B., H.C., Y.C., and K.P., performed research; D.R. analyzed data; and D.A.R., N.K.Y, and C.A.S, wrote the paper.

Contributions from Debra A. Ragland: I performed structural and dynamic analyses of the five crystal structures in this study with help from Yufeng Cai and Kristina Prachanronarong. I created all figures and tables for this manuscript. With the guidance of my advisor Celia Schiffer I interpreted the data and wrote the manuscript with the assistance of Nese Kurt-Yilmaz and Celia Schiffer.

Chapter III* is a collaborative study that is in preparation for publication as:

Debra A. Ragland, Troy W. Whitfield, Nese-Kurt Yilmaz Konstantin B. Zeldovich, and Celia A. Schiffer. (2017) Elucidating the interdependence of drug resistance from combinations of mutations. [*in preparation*]

Author contributions: D.A.R. and T.W.W contributed equally to this work. D.A.R and T.W.W designed and performed research and interpreted data. D.A.R, T.W.W., N.K.Y., and C.A.S wrote the paper.

Contributions from Debra A. Ragland: I selected all variants from the panel and performed all homology modeling and molecular dynamics simulations and interpreted the resulting data. With the guidance of Troy W. Whitfield and my advisor Celia A. Schiffer I was able to use statistics and machine learning techniques to further interpret the data. I created all figures and tables for the manuscript and wrote the manuscript with the assistance of Nese Kurt-Yilmaz, Troy W. Whitfield, and Celia A. Schiffer.

*Chapter III also contains information from a collaborative study that has been previously published as:

Marc Potempa, Ellen Nalivaika, Debra Ragland, Sook-Kyung Lee, Celia A. Schiffer and Ronald Swanstrom. A Direct Interaction with RNA Dramatically Enhances the Catalytic Activity of the HIV-1 protease *In Vitro*. Journal of Molecular Biology. 2015. 427, 2360-2378. DOI:10.1016/j.jmb.2015.05.007

Author Contributions: M.P. and S.K.L. carried out predominantly all experimentation. M.P. wrote the paper with the guidance of his advisor Ronald Swanstrom.

Contributions from Debra A. Ragland: I expressed and purified all five of the patient-derived variants used in this article and I assisted with making figures for the publication.

# Chapter I Introduction

## 1.1 HIV-1 history and life cycle

### 1.1.1 History of HIV-1

In 1981 the CDC published in their Morbidity and Mortality Weekly Report (MMWR) about a group of five young, previously healthy, homosexual men that had undergone biopsy were confirmed to have *Pneumocystis carinii* pneumonia (PCP) at three different hospitals in Los Angeles, California [1]. Before the report was published, two of the men had already expired. Although the CDC did not know at the time that this group of five men had succumbed to what we now know as acquired immune deficiency syndrome (AIDS), this report was the beginning of the collaborative efforts to identify AIDS and determine its causative agent HIV. Shortly after issuing this report, the CDC was overwhelmed with more instances of both PCP and a rare aggressive cancer known as Kaposi's sarcoma (KS) from both New York and California [2].

The following year, in 1982, the CDC published in MMWR an update on AIDS, which became the first official use of the term [3]. It would however, take another two years to identify the viral etiology of AIDS. The hunt for the human immunodeficiency virus started with several false leads. Because AIDS infection appeared to be a chronic disease marked by a long lag time between exposure and immune suppression and manifested itself by secondary infections and cancers in the clinic, there were many factors from fungi to chemicals that were considered as possible causes. However, for the scientists at the National Institutes of Health (NIH) in Bethesda, MD and at the Pasteur Institute in Paris,

France there were many clues. These included the decreased levels of CD4 T-cells, the ability of the virus to be transmitted through blood, sexual activity and from mother to infant. Cases being reported in hemophiliacs left scientists searching for what they thought was a human T-cell leukemia virus (HTLV) and possibly related to HTLV-II. There were even several early publications in *Nature* and *Science* that reported the complete sequence and gene expression for HTLV-III and AIDS-Associated Retrovirus (ARV-2) [4, 5].

That the causative agent of AIDS was a retrovirus like HTLV was correct, but the assumption that the virus was a relative of HTLV proved to be incorrect. Sequencing and cloning of an AIDS-causing virus isolate from a patient with lymphadenopathy at the Pasteur Institute helped shift the hunt from HTLV to what was then known as the lymphodenopathy AIDS virus (LAV) because it too was a retrovirus but was not evolutionarily similar to HTLV I or II [6, 7]. With subsequent studies, including the ability to grow the virus in T-Cell lines, human immunodeficiency virus (HIV) was accepted as the causative agent for AIDS in 1984 and it was given a formal name in 1986 [8, 9].

Once the linkage between HIV and AIDS was determined, the search for its origin began. Very early on François Clavel and coworkers were able to isolate a new human retrovirus from West African AIDS patients that was more similar to simian T-Cell lymphotrophic virus in macaques (STLV-$III_{mac}$) than it was to the AIDS-related virus reported in America [10]. Because the new STLV-$III_{mac}$ like virus was similar to LAV with the exception of some differences in the

envelope glycoproteins, Clavel and coworkers referred to this virus as LAV-II, which is what we now know as HIV-2. Continued work led to findings that STLV-III was in fact simian immunodeficiency virus (SIV), a lentivirus causing chronic and persistent infection in mammals similar to its relative HIV with differing antigenicity, and ability to induce a disease similar to human AIDS [11, 12]. Like HIV, SIV too can be transmitted through sexual intercourse and mother to child transmission. SIV was originally thought to not impact its natural host, but it was later discovered that primate communities impacted by SIV had higher immune system suppression and mortality rates than those that were not infected [13-15].

One SIV in particular is genetically closer to HIV-1 than the rest and that is SIV from chimpanzees (SIVcpz). Morever, the human evolved HIV-1 can be specifically traced back to the *Pan troglodytes troglodytes (P.t. troglodytes)* subspecies of chimpanzee. Although HIV-1 could be traced back to chimpanzees, the origin of SIVcpz remained a mystery until characterization of its genome revealed that not only was it closely related to SIV found in red capped mangabeys (SIVrcm). It was also derived from non-human primatesrelated to red-capped mangabeys (other monkeys from the *Cercopithecus* species)[16]. These findings suggested that there was cross-species transmission from smaller monkeys to chimpanzees possibly due to chimpanzees' hunting behavior[17]. SIVcpz also had cross-species infection into gorillas but the origin of this has yet to be discovered. Both SIVcpz and SIVgor

have been identified as reservoirs for human HIV, possibly due to human bushmeat hunting [18, 19].

In recent years findings show that HIV-1 can be broken further into four different lineages also known as groups. In order of discovery and prevalence, there is the main group M, followed by groups N, O (outlier) and P [20, 21]. Group M also represents the pandemic form of HIV-1 and has been traced back to Kinshasa in the Democratic Republic of Congo [12, 22]. HIV-1 group M can be further broken into nine subtypes known as clades (A-D, F-H, J & K) [23] and due to recombination in people already infected with one subtype, there are also over 40 circulating recombinant forms (CRFs) [24]. Subtype B accounts for the majority of HIV-1 group M infections in Europe and the Americas and viral sequencing studies indicate that this specific subtype left Africa and moved onward to Haiti then to the United States [25].

### 1.1.2 The HIV-1 viral life cycle

The HIV-1 genome encodes nine open reading frames, three of which contain the gag, pol and env polyprotein common to all retroviruses (Figure 1.1).

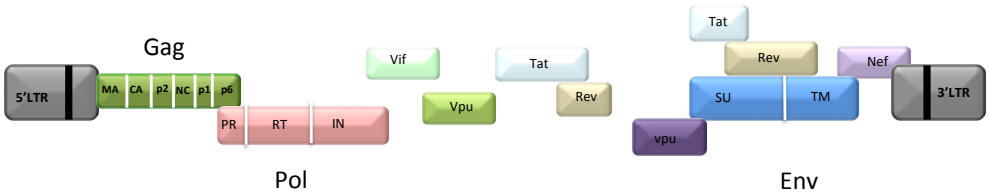



**Figure 1.1 The HIV-1 genome. The Gag, Pol, and Env polyproteins are labeled accordingly with individual proteins shown separated by white bars. Accessory proteins and 5' and 3' LTR are labeled accordingly.**

There are four proteins encoded within the gag precursor polyprotein, matrix (MA), capsid (CA), nucleocapsid (NC), and p6. There are three proteins encoded within the pol region of the gag-pol precursor polyprotein, protease (PR), reverse transcriptase (RT), and integrase (IN). There are two proteins encoded in the envelope precursor polyprotein, the surface protein (SU, gp120) and the transmembrane protein (TM, gp41). The gag and envelope proteins make up the structural components of the viral particle while the enzymes within the pol polyprotein provide essential functions necessary to create fully infectious viral particles. There are also six other proteins contained within the virus known as accessory proteins, which undertake various tasks to ensure that viral replication succeeds. These are Vif, Vpr, Nef, Tat, Rev and Vpu.

The viral genome itself is encoded by two copies of a 9kb positive sense RNA that gets packaged into the virion with the help of zinc finger containing NC protein. The myristolation site at the N-terminus of MA allows anchoring to the plasma membrane of the host cell along with the rest of the full-length gag-pro-pol polypeptide precursor. The capsid proteins, once cleaved from the aforementioned precursor, form a conical shaped structure at the center of the virus that encapsulates the RNA and NC proteins. p6 contains binding domains for viral protein Vpr as well as host proteins from the endosomal sorting complexes required for transport (ESCRT) pathway [26]. The viral envelope proteins are targeted to the cell membrane after cleavage by the cellular

protease furin into gp120 and gp41, after being co-translationally inserted in the endoplasmic reticulum (ER) membrane and travelling through the host cells secretory pathway via vesicular transport.

Both accessory proteins Nef and Vpu play a role in down-regulating CD4 with Nef working to down-regulate CD4 at the cell surface via endocytosis and Vpu working to down-regulate CD4 in the ER. Both Nef and Vpu also antagonize host proteins such as tetherin [17, 27]. Rev binds the rev response element (RRE) and is required to transport the unspliced viral mRNAs from the cell nucleus to the cell cytoplasm, which would otherwise be retained in the nucleus and degraded. Tat is required to bind to the trans-activating response element or TAR in order to promote transcription. Finally, the protein Vpr is important for nuclear localization in general but it is specifically responsible for facilitating transport of nucleoprotein complexes and partially reverse transcribed DNA into the host cell nucleus [27].

For entry into a host cell, the envelope protein gp120 binds to host cell receptor CD4. This binding event leads to a conformational change in the viral envelope protein, which allows it to then expose gp41. gp41 binds to a co-receptor, either CXCR4 or CCR5, for entry. The viral membrane and the host membrane then fuse, which allows the entry of components within the viral particle into the cell. At some point during entry, the viral capsid uncoats to release the viral RNA and proteins into the cell cytoplasm where the RNA is subsequently reverse transcribed by the RT into a duplex linear DNA and

transferred to the cell nucleus by Vpr. Inside the nucleus, integrase performs a series of steps to integrate the viral DNA into the host cell genome. Viral transcripts are then expressed using Tat to enhance binding to the promoter in the 5' LTR to enhance transcription initiation. The newly transcribed viral mRNAs are then transferred from the nucleus to the cytoplasm mostly unspliced by Rev where they are subsequently translated and targeted to various locations in the cell. As the viral particles assemble at the cell membrane, the viral protease is activated and is able to cleave itself out of the polyprotein, dimerize, then cleave other proteins to ensure that the viral particle is mature and infectious.

All necessary viral components and a lysyl tRNA from the host cell are used to prime the cDNA are packaged into the budding virion. Although the HIV-1 Gag polyprotein is largely able to mediate the events necessary for viral budding at the plasma membrane, it is the recruitment and binding of host proteins Tsg101 and ALIX from the ESCRT pathway to p6 that drive the membrane fission between virion and host cell (Figure 1.3) [28-31]. Once the process is complete, the virion moves on to infect another CD4+ T-cell. This process can be seen in Figure 1.2.

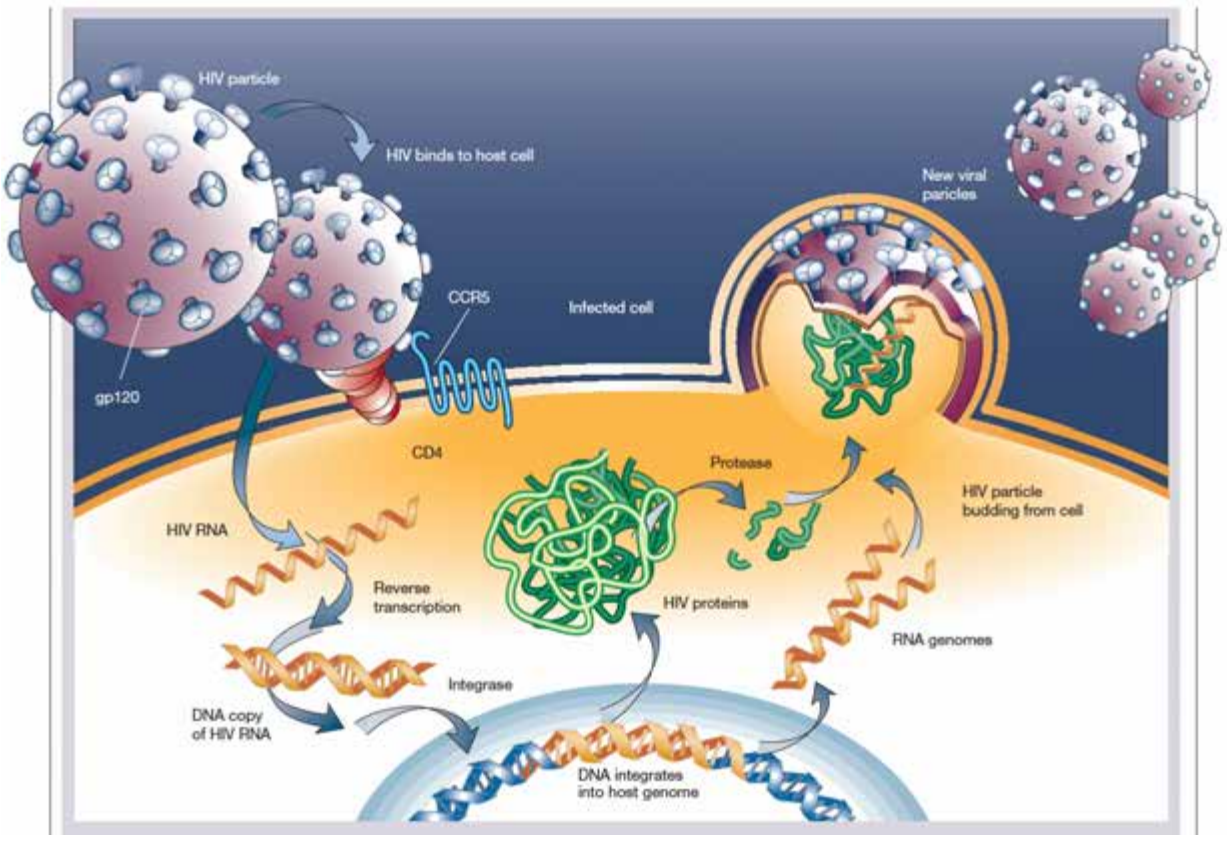



**Figure 1.2 The HIV-1 viral life-cycle [32].**

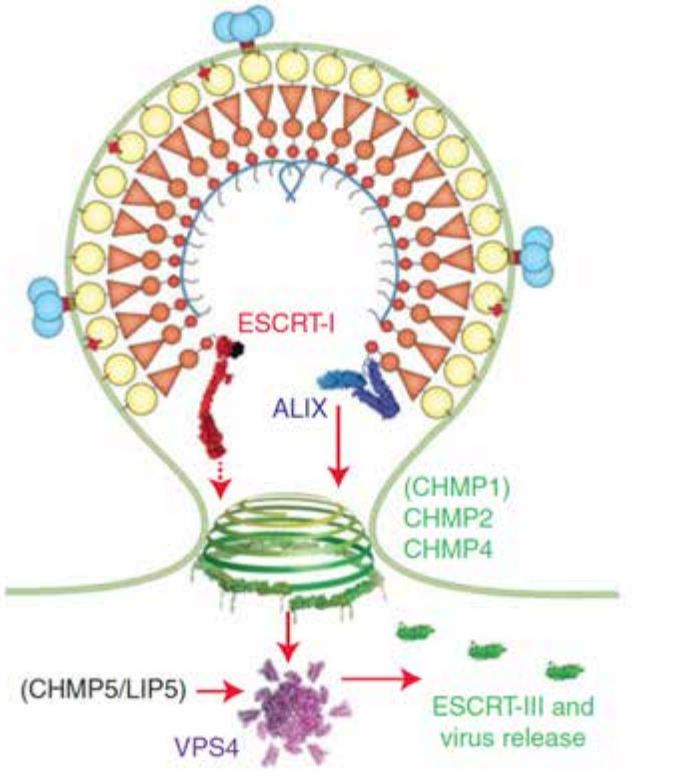



**Figure 1.3 ESCRT machinery used for membrane fission in HIV-1 budding [26]. Auxiliary factors are shown in parentheses. The p6 (protein) region within Gag contains two different late domain motifs, the primary domanin consists of residues Pro-Tyr-Ala-Pro ("PTAP") and binds the TSG101 subunit of heterotetrameric ESCRT-1 complex (red, and complexed with ubiquitin black). The secondary "YPXL"(Tyr-Pro-X=variable-Leu) domain binds the ESCRT factor ALIX (blue). These interactions recruit ESCRT-III (made of CHMP 1, 2 and 4, green) which polymerize into a dome and promotes closure of the membrane neck and VPS4 ATP-ases(purple) which uses the energy from ATP to release ESCRT-III back into the cytoplasm.**

### 1.1.3 HIV heterogeneity and in vivo evolution

Simply put, the human immunodeficiency virus is a replication machine. HIV-1 is able to evolve at a rate several orders of magnitude faster than host T-cells, taking just ~2 days to produce an entirely new generation of viral particles on the order of $10^{10}$ virions per day [33-36]. The secret weapon behind this mass

viral production is the RT. RT is naturally error-prone due to its lack of a proofreading mechanism, the virus can replicate without interruption even though the consequence of this replication is 3-4 $x10^{-5}$ mutations per base per cycle. That is, the RT introduces a mutation every 1000-10,000 nucleotides [37-39]. Because the HIV-1 genome is only 9-10 kb, the rate of mutation ranges from 1 to 10 per genome for every replication cycle.

The high mutation rate caused by RT's lack of proof-reading is only one means by which the virus is able to substantiate genetic diversity while replicating within its host. Due to its ability to produce genetically distinct yet highly related viruses, HIV-1 is said to be a quasispecies [40]. Because HIV-1 is a quasispecies, an infected patient may contain a population of viruses with high variation of genetic information [41-44]. HIV-1 is also able to recombine with other subtypes of virus replicating within the same cell [24, 45]. Recombination can occur at one or more steps during viral replication. For instance, since there are two copies of RNA packaged into each virion the RT is able to switch between the RNAs, promoting strand transfer, and because there are no checkpoints for the virus as there are for the host cell, the RNAs that get packed into a budding virion can contain information from two different viral sequences. By default, recombination can also occur during repair of any damage done to either of the viral RNAs [46]. There has even been evidence of viral recombination between two major groups, M and O [47]. The inherent genetic

diversity within the virus allows multiple pathways for viral adaptation and poses a challenge to HIV-1 chemotherapy.

### 1.1.4 HIV-1 and the immune system

In the midst of early HIV-1 discoveries in the 1980s and 1990s, there was one more observation that was not well understood: exactly how did HIV-1 cause CD4+ T-cell depletion in infected patients? At first, there were reports that HIV-1 did in fact cause T-cell depletion. Ho and coworkers showed that prior to any chemotherapies, continuous rounds of HIV-1 infection caused the turnover of ~ $2x10^9$ CD4+ T-cells per day [35, 48]. How this turnover was driven was unclear. The lack of a well-defined answer to these questions led to many theories such as, HIV-1 was somehow able to alter the natural homing response of CD4+ and CD8+ T-cells out of lymphoid tissue and into circulation upon the presentation of antigen such that there would be more T-cells for HIV-1 to infect and destroy [49-51]. Another theory stated that uninfected T-cells activated for immune response would be infected by HIV-1 and not replenished due to thymic depletion [52, 53]. Finally, another theory stated that HIV-1 was able to suppress hematopoiesis [54].

It was not until relatively recently that Warner C. Greene and colleagues at the Gladstone Institute of Virology and Immunology established the cause of death for CD4+ T-cells as a result of HIV infection. In both instances, the groups stated that CD4 T-cells die via mechanistic caspase-1 mediated pyropoptosis and not the canonical caspase-3 mediated apoptosis [55]. This group previously

published that only 5% of CD4s are productively infected (the virus is able to integrate and replicate fully) and circulating in the blood while 95% of the CD4+ T-cells within the lymphoid tissue are abortively infected (the virus cannot integrate and replicate resulting in abortive transcripts in the cytoplasm)[56]. This "by-stander" population in the lymphoid tissue however is most affected by HIV-1 infection. Greene and co-workers found that cell death of the circulating population of T-cells that produce productively infective virus is triggered through caspase-1 mediated apoptosis. However, when HIV-1 abortively infects nonpermissive CD4 T-cells cell death occurs by caspase-3 mediated pyropoptosis, which accounts for the majority of cell death within the lymphoid tissue [57, 58]. Because HIV-1 causes pathogenic inflammation, the signals between cells to promote clearance are disrupted and instead of clearing the infection, more cells are attracted to infected tissue and meet the same fate as resident CD4 T-cells, which in turn causes more inflammation and the cycle continues.

The above scenario precisely underscores why antiretrovirals (ARVs) are vital to HIV-1 suppression: replication must be halted to disrupt the infection cycle. Because ARVs suppress viral replication by inhibiting entry, reverse transcription, integration and polyprotein cleavage, they subsequently prevent the above processes from occurring and in turn allow for the continued proliferation of T-cells in the presence on infection. ARVs are subsequently able to block the

host response by preventing viral replication and genetic diversity to overcome the lack of proper host response.

## 1.2 The HIV-1 Protease

### 1.2.1 HIV-1 protease structure and function

HIV-1 is a retrovirus with the same basic composition as other retroviruses; a *gag* gene encoding structural proteins, a *pol* gene encoding enzymes necessary for replication and an *env* gene encoding proteins for viral entry. HIV-1 is slightly different than other retroviruses in that it also encodes for several proteins unique only to the virus [59]. Of the proteins common to all retroviruses, HIV-1 contains a protease encoded within the *pol* gene used by the virus to cleave peptide bonds between various proteins within the precursor polyproteins of the virus and even some within the host that interfere with viral replication or necessary to induce cell death [55, 60, 61].

The HIV-1 protease, or retropepsin as classified within the enyme classification system (E.C. 3.4.23.16), is a hydrolase. More specifically, it is an aspartic endopeptidase related to cathepsin D, renin and human endogenous retrovirus K10 protease [62, 63]. As with all eukaryotic aspartic proteases, HIV-1 protease contains two catalytic aspartic acids within its active site necessary for the enzyme to function. Early structural and expression and purification studies confirmed that HIV-1 protease functions as a dimer, with each monomer contributing one aspartate to the active site [64, 65]. In addition to the HIV-1 protease's catalytic dyad each monomer contributes a threonine and a glycine in

the order $Asp^{25}$-$Thr^{26}$-$Gly^{27}$, which are necessary for protein dimerization as the two monomers are not covalently linked [66]. The protease contains two flaps, one from each monomer, which open and close to allow substrates into the active site cleft (Figure 1.4) [67]. Studies showed that the HIV-1 protease was highly specific with respect to the cleavage sites within its substrates [68]. However, due to the diversity of residues surrounding the scissile bond, questions remain as to why or how the protease performed hydrolysis on these specific sites.

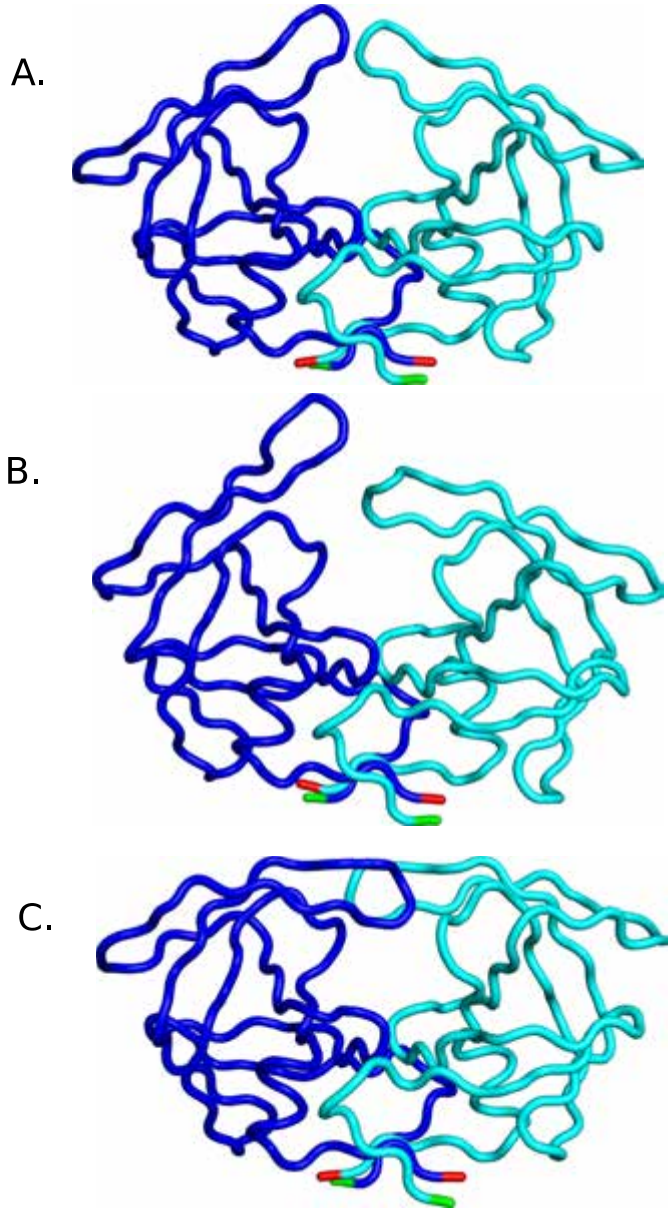



**Figure 1.4 Open, partially open and closed conformations of HIV-1 protease. N and C termini are colored in green and red respectively. "A and B" monomers are colored in blue and cyan to differentiate. Conformations in A & B were originally described in [69] accession codes 3UHL and 3UF3. Conformation in C was originally described in [70] accession code 1T3R.**

Of the twelve substrates within the virus that undergo cleavage by the protease, the majority of them contain a small hydrophobic residue at the P1 position while the residue on the other side of the scissile bond at the P1' site is generally a proline but it often varies. It would not be until twenty years later that the determinant of the specificity of the protease was elucidated.

Like all other hydrolases the HIV-1 protease catalyzes the hydrolysis of peptide bonds in the form;

$$A - B + H_2O \rightarrow A - OH + B - H$$

where A−B represents the peptide bond between two amino acids. The widely accepted mechanism behind this reaction is as follows; the two aspartic acid residues within the active site begin a water mediated attack on a nearby nucleophilic hydroxyl group from the backbone of a residue. Due to the pH dependent reaction of the protease, the nucleophilic attack begins from the aspartic acid residue that is not protonated. The water that is attacked by the unprotonated Asp subsequently attacks the carbonyl carbon, which allows for release of the oxygen of the carbonyl group. This forms an unstable tetrahedral intermediate where the originally protonated Asp residue is intermediately hydrogen bonded to the substrate. Subsequent protonation of the nitrogen from the amide group of the original peptide bond and rearrangement causes the tetrahedral intermediate to breakdown and hydrolysis succeeds (Figure 1.5) [71]. Understanding and exploiting the mechanism of action used by the protease to

hydrolyze substrates has proven to be incomparably vital to treatment of individuals with HIV-1 via inhibiting the protease.

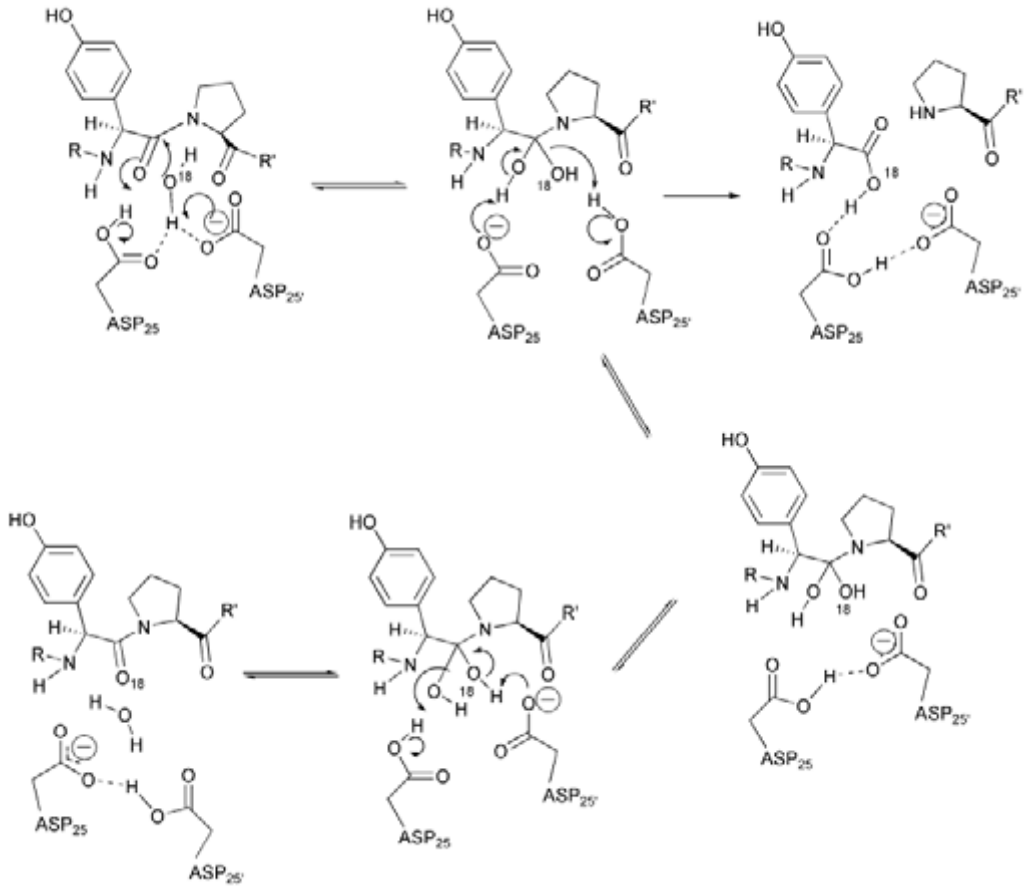


**Figure 1.5 HIV-1 protease mechanism of action [72].**

### 1.2.2 The HIV-1 protease as a drug target

Soon after the discovery that HIV-1 protease was in fact an aspartyl protease that functioned as a dimer, it was identified as a potential target for therapeutics after an experiment showed that inhibition via a frame shift mutation

resulted in the failure of the protease to process the gag polypeptide [73]. With this information, the search began for molecules capable of inhibiting the protease. Naturally, given that the HIV-1 protease was an aspartyl protease, some researchers tried to use known inhibitors of other aspartic proteases on the HIV-1 protease such as various forms of pepstatin and renin inhibitors [74].

Although these general aspartic protease inhibitors proved not to be effective on HIV-1 protease, the rationale taken from preliminary studies, that HIV-1 protease inhibitors should transition state mimics, proved to be vital in the years and trials to come. A focus on improving the non-hydrolyzable hydroxyethylene isosteres was undertaken by many groups since this was the preliminary scaffold for pepstatin inhibitors [75, 76]. With new structural information and armed with the knowledge that HIV-1 protease had the unique ability to cleave between sites containing phenylalanine/tyrosine and proline residues—which are not susceptible to cleavage by mammalian endopetidases—creating mimics of these residues around the pepstatin-based scaffolds proved successful. However, what proved to be even more successful was the use of a hydroxyethylamine isostere scaffold instead of the reduced amide scaffold to better serve as a mimetic of the Phe-Pro and Tyr-Pro scissile bonds [77]. The use of large hydrophobic side chain mimics served to fill the subsites within the protease, which played into the protease's elusive specificity. The hydroxyl based scaffolds (e.g. hydroxyethylene, -ethylamine, or –ethylamino sulfonamides) were more potent against the HIV-1 protease because the presence of the hydroxyl

group displaced the active site water and engaged both Asp residues into a transition state like interaction. Compounds made using this scaffolding were also found to be highly specific to targeting the protease, which in turn reduced toxicity and mistargeting of host aspartyl proteases. These notable advancements in protease inhibitor design are arguably the start of rational structure-based drug design.

Using the techniques and chemical scaffolds mentioned above, the first protease inhibitor, saquinavir (Fortovase™/Invirase™), was approved as a protease inhibitor by the FDA in 1995 followed by ritonavir (Norvir™) in 1996. Prior to this, there was only the reverse transcriptase inhibitor azidothymidine (AZT/ Zidovudine, ZDV™) approved in 1986. The protease is active late in viral maturation and its ability to process substrates is crucial for maturation and infectivity of newly budding virions. Treatment with protease inhibitors prevents the necessary cleavage events from occurring thereby rendering virus particle incapable of infecting another cell (schematic Figure 1.6). This feature had given protease inhibitors a unique advantage over the likes of AZT, which halts viral replication in newly infected cells, as protease inhibitors are also able to disrupt virus generated from chronically infected cells. As an RT inhibitor, AZT is only able to inhibit it's target in acutely infected cells [78].

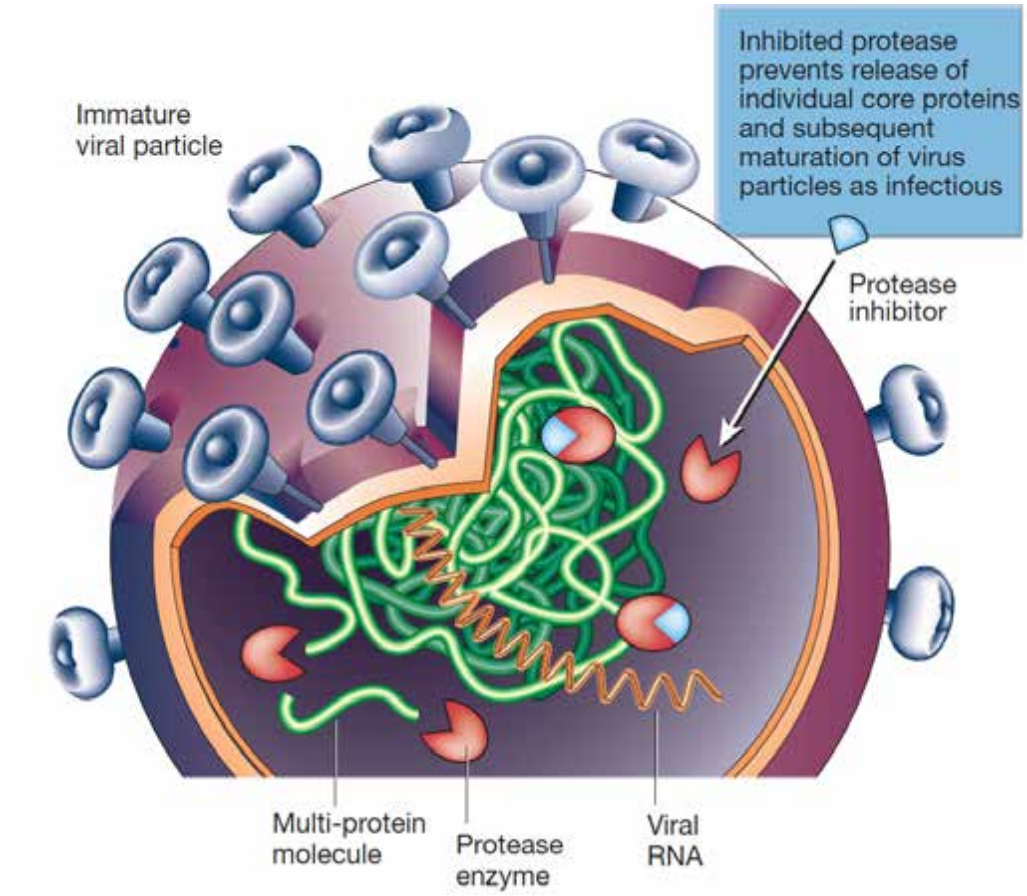



**Figure 1.6 Schematic diagram of HIV-1 protease inhibition [79].**

## 1.3 History of HIV-1 treatment and generations of protease inhibitors

### 1.3.1 The progression of antiretroviral treatment from monotherapies toward a vaccine

Dr. Anthony Fauci of the National Institutes of Health (NIH, Bethesda, MD) accurately categorized the treatment history of HIV-1 treatment into three phases: the “dark ages” from 1981–1986, the “pre-highly active antiretroviral therapy (HAART) era from 1987–1996, and the “HAART” era from 1996–

2006[80]. Certainly, at the discovery of AIDS in 1981 scientists were far behind in piecing together the puzzle of the relation of HIV and AIDS but once HIV was found to be a retrovirus, the work of finding inhibitors to slow viral replication began. As mentioned previously, researchers started with known aspartic acid protease inhibitors and soon found that these were ineffective. Initially, HIV-1 therapy constituted of AZT monotherapy-which was approved by the FDA in a record 21 months, followed by a series of other nucleoside reverse transcriptase inhibitors (NRTIs) including didanosine (ddI, Videx™) in 1991, 3TC(Lamivudine, Epivir™) in 1995.

With information gained from protease/inhibitor co-crystal complex structures, the commencement of protease inhibitor (PI) design was initiated [81-83]. Concurrent with clinical development of PIs was the refinement of intra-host HIV-1 quantitation assays. These new methods for determining the amount of virus within an individual were based on viral RNA copy number within the plasma instead of CD4 count, and were thereby more sensitive to both viral load and viral inhibition [33, 35, 84, 85]. Although highly potent in vitro, early protease inhibitors had low bioavailability, which means that patients had to inject themselves in order for the drug to be administered. With some refinement of solubility, the bioavailability of protease inhibitors improved, allowing patients the ability to take them orally [86].

The issue of viral heterogeneity proved highly problematic in the very early stages of HIV therapeutics. Because there is simply not one single viral strain

within a host, many groups sought to figure out how to quantitatively account for this in inhibitor design [87-89]. The viral propensity to mutate and recombine in addition to the short term effective inhibition of early monotherapies called for re-evaluation of how to best administer antiretrovirals to effectively inhibit all pre-existing mutant populations [90]. The interpretation of several mathematical models and experimental findings concluded that at least two or more drugs should be given to patients to help both with longer lasting and more effective antiretroviral treatment better suited to ward off the rise of mutations and drug resistance. Initial combination therapy consisted of the administration of NRTIs such as AZT and 3TC which in time ultimately led to failure after a slightly extended period of time when compared to monotherapies with either drug [91, 92].

It was not until the introduction of PIs to therapeutic regimens either as additional monotherapeutics or in combination with failing regimens (i.e. 2 NRTIs +1 PI) that HAART was born [93]. Over time, the types of combinations have increased with the number of approved inhibitors. Because of the potency, propensity to slow viral progression and lack of other interfering parameters (low bioavailabity, high toxicity etc.) the combination of 2 NRTIs + 1 PI has become the standard for HIV-1 treatment in naïve patients [94]. There have also been many studies conducted with PIs in combination with other PIs and with non-NRTIs (NNRTIs) however these have proven to be less optimal except in

instances where ritonavir is used as a “booster” (cytochrome P-450 inhibitor) to increase the half-life of other protease inhibitors.

The advancement of HIV-1 chemotherapy over the past few decades has been met with many challenges and setbacks but has progressed to be highly effective and has prolonged the life expectancy of those living with the virus. Because HAART is typically a life-long treatment and cannot cure individuals, there have been strides made to find an HIV-1 vaccine to prevent and eradicate the virus once and for all. However, like the early days of drug discovery for the virus, the hunt for a vaccine is met with challenges. In recent years, major effort has been devoted to T-cell vaccine trials with moderate success. One of the major setbacks to vaccine trials has been the lack of ability to generate broadly neutralizing antibodies (BNAbs) due to HIV-1 eliciting only a limited response from these kinds of antibodies years after initial infection [95-97]. The fact that the virus can go years being undetected by the immune system with respect to antibody response is a major problem in and of itself due the virus’ archetypical ability to establish heterogeneity. Thus, in order to produce an effective vaccine, we must first understand how to enhance antibody response [98].

### 1.3.2 Outcomes of Prolonged Treatment and Aging with HIV

For the 34 million people living with HIV-1 today, the efficacy of combinatorial antiretroviral therapy is essentially allowing individuals to live full lives. However, with prolonged survival come challenges to the typical aging process. Some of these challenges include the onset of comorbidities at an

earlier age than in non-infected individuals (diabetes, heart disease, kidney disease, etc.). As previously mentioned, antiretroviral therapy is life long and this means that some individuals will be exposed to therapy for many decades. With this prolonged exposure, the chances of toxicity are increased due to change in physiological conditions of the human body with age. Because of this, there may be need to work towards new inhibitors that are more tolerable, have lowered interactions with drugs taken for other age related morbidities, or that are longer lasting and thereby taken less frequently [99].

With age too, comes the contraction of the thymus, the organ responsible for T-cell development. As the thymus shrinks due to age, it also produces less naïve T-cells. The increased levels of T-cells within infected individuals with HIV is a hallmark of treatment, however, the loss of T-cells or their function can cause HIV to progress more rapidly in the elderly. There is also the problem of neurocognitive abnormalities, such as AIDS related dementia in the elderly living with HIV. Simply put, the increased life expectancy for those infected with HIV comes at a cost. The natural aging defects seen in HIV negative individuals are compounded in those afflicted with the disease. This means that antiretroviral therapy should not be the last stop in HIV treatment, but instead should be the first step of many to ensure that HIV remains controllable and tolerable throughout the life of the afflicted.

### 1.3.3 Non-adherence to antiretrovirals leads to resistance

Several clinical studies have made the connection clear: adherence to antiretroviral regimens is the key to continued viral suppression. Adherence to treatment regimens even reduces the opportunistic infections such as PCP [100]. Combination therapy is the most effective means by which doctors can help patients keep their viremia low, so low in fact that with adherence viral levels can be suppressed below detection levels. The reason why HAART is so effective at viral suppression lies in the fact that using a combination of drugs decreases the probability of selecting a viral population that is resistant to all the drugs in one regimen. However, there are several factors that sometimes make regimen adherence difficult for patients. These include intolerance to one or more of the drugs within the cocktail, lack of access to therapy, high baseline viral loads, or poor pharmacokinetics of ARVs [101-103]. Although deviating from prescribed treatment regimens or discontinuing treatment altogether is not advised, some patients choose not to adhere to therapy and this in turn leads to viral resistance. Selection of drug resistant viral populations due to lack of chemotherapy undermines the efficacy of ARVs and subsequently leads patients on a path to complete ARV failure via viral rebound [103-105].

As mentioned previously, the addition of protease inhibitors significantly increases the efficacy of ARV regimens, even for patients failing on NRTI or NNRTI based regimens, due to their very high viral suppression activity [78]. Amazingly, it has been shown in several clinical trials that regimens containing a

PI in addition to some combination of other viral enzyme inhibitors do not select for protease mutations [103, 106, 107]. Instead mutations at targets of other drugs within the regimen are more prevalent[108]. Although in theory this is how HAART should work, even with mutations against two of three drugs, it is not understood why failure occurs in the presence of a protease inhibitor if there are no mutations within the protease itself [106-109]. In a 2012 modeling study, David Rosenbloom and coworkers developed a mathematical model to predict how long after the last treatment dose (before non-adherence) would the virus (either mutant or WT) begin to rebound. This study they found that the level of adherence needed avoid viral rebound is dependent upon the class of drug. For instance, patients on NNRTIs should adhere to their regimens as much as possible beacause of an increased risk of mutant viral growth while patients on a boosted protease inhibitor regimen have a bit more room for non-adherence due to wildtype viral regrowth. This result consistent with clinical findings that patients on NNRTI and PI therapy typically encounter virological failure [86]. They also found, specifically for protease inhibitors, that the sharp dose-response curve and rapid decay of PI concentrations leaves little room for mutant population growth in a period of non-adherence. This suggests that patients who fail protease inhibitor therapy should be able to re-suppress their virus with improved adherence [110].

### 1.3.4 Improvement on protease inhibitors with each generation

To date there are nine FDA approved competitive HIV-1 protease inhibitors, eight of which are peptidomimetics based on the early transition state mimic conceptual foundation in the 1980s [111]. The inhibitors can be broken into generations depending on their time of discovery and approval. The first generation inhibitors saquinavir (SQV), ritonavir (RTV), indinavir (IDV), nelfinavir (NFV) and amprenavir (APV) are largely not prescribed anymore. The second generation inhibitors lopinavir (LPV), atazanavir (ATV), tipranavir (non-peptidomimetic, TPV), and darunavir (DRV) are largely still in clinical use with boosted LPV or ATV being among the most prescribed PIs in antiretroviral regimens. The majority of these approved inhibitors share strong chemical commonalities; for instance, the second generation LPV was largely based on the chemical scaffold of RTV, while APV and DRV differ by just the addition of a tetrahydrofuran group to the latter. Although the protease inhibitors as a class are highly potent against their target, side effects that they may cause make them very intolerable to some patients. For instance TPV, it is only to be prescribed at absolute failure of other PIs due to its hepatotoxicity and its necessity to be double boosted by RTV. Only recently has DRV been approved to be given as a first line PI in chemotherapy regimens.

Protease inhibitors essentially inactivate the enzyme by binding at the active site and locking the flaps down, mimicking prolonged transition state. To overcome this inhibition, the protease develops mutations that prevent the

inhibitors from binding. However, the accumulation of mutations typically comes at a cost; the enzyme loses the efficiency of processing natural substrates, also known as fitness. To aid in fitness recovery, the virus also develops additional mutations within either within the protease or elsewhere within Gag and Gag-Pro-Pol polypeptide, which improve cleavage of substrates, known as substrate co-evolution [112]. Studies have shown that not only do the mutations within and surrounding cleavage sites improve enzymatic fitness, they are also able, in conjunction with the protease, to help drive resistance to PIs [113, 114]. In fact, in a study by Parry et al, the non-peptidomimetic TPV was the only PI assayed that both a mutant protease and mutant Gag remained susceptible to even though TPV interacts with the protease in a similar way as the other PIs [115]. In any case, resistance to all nine FDA-approved inhibitors has been documented and mutations that underlie PI resistance never appear alone. Some of the resistance-causing mutations within the protease directly confer resistance to PIs while others arise to restore viral fitness loss via the presence of mutations that are detrimental to substrate turnover.

### 1.3.5 The role of structure based drug design in creating more effective PIs

With the precedent set in the late 1980s and the early 1990s, rational structure based drug design (SBDD) has become the superior technique in the conquest to combat HIV-1. Although, in the case of HIV-1, use of rational drug design was a means to an end race against time, it has advanced to become ends to a mean tool in the field of pharmacology, with entire companies built on

this foundation [116]. The process now, iterative but certainly more robust, typically involves expression/purification, structural determination of the target to be inhibited, identification of possible inhibitors followed by biochemical assays to determine potency, and structure determination of a target-inhibitor complex to unveil molecular interactions that can possibly be enhanced to improve potency [117].

With respect to rational drug design of antiretrovirals against HIV-1, there is no question that the design of protease inhibitors has triumphed over other target inhibitors. It was pure serendipity that led to the discovery of Enfuvirtide™, the first HIV-1 inhibitor of gp41 preventing fusion and small molecule screening that led to the discovery of its counterpart maraviroc which inhibits gp120 from binding to the CCR5 co-receptor [118, 119]. Specifically for the enzymes encoded within the *pol* gene, the initial administration of AZT was given to patients without the structure of RT or RT in complex with an inhibitor being known [120]. And while integrase's structure was elucidated in 1994, the first integrase inhibitor, raltegravir (RAL), was approved in 2008 with some chemical based intuition and some true to form rational design. Still with the advancement of such parameters as computational power, small molecule screening, and crystallographic screening automation, inhibitors of the HIV-1 protease remain the unsurpassed prototypes for structure based rational design.

The mechanistic approaches used in conjunction with SBDD have further been improved upon. For instance, Lee and coworkers have found success using

a solvent anchoring approach utilizing a covalently attached phosphonate moiety to increase the potency of inhibitors while others such as Tidor and coworkers have found success using inverse design methods utilizing the substrate volume as a restraint to dictate inhibitor design [121, 122]. The most successful mechanistic approach in the past decade appears to be maximizing the interactions between the enzyme and the inhibitor, such as that used by Mitsuya and coworkers [123]. With this strategy, the most potent FDA approved HIV-1 protease inhibitor to date, darunavir (DRV/Prezista™) has been discovered. While both fairly recently approved PIs darunavir and tipranavir(TPV/Aptivus™) are highly potent against the protease and highly resistant variants, DRV has better bioavailability and slightly lower levels of toxicity. By targeting the backbone atoms of residues in the protease's active site, Mitsuya and coworkers unlocked a key facet lacking in previously approved inhibitors. That DRV is able to form many hydrogen bonds with the backbone atoms of the protease allows it to overcome drug resistance in a manner that is different than that of TPV, which directly interacts with the flaps of the protease.

## 1.4 Drug resistance to ARVs

### 1.4.1 Why drug resistance occurs in HIV-1

HIV-1 is masterfully erroneous. With high levels of productivity and the inherently inaccurate nature of the very enzyme responsible for transcribing its genome, drug resistance in HIV-1 is a matter of eventuality par for the course. With its implicit heterogeneity, ARVs are essentially able to target the most

susceptible quasispecies within the host. However, with the targeting and subsequent downfall of one viral population comes the uprising of a population that is resistant to the antiviral agents administered. Though it is not a simple process, if the population containing the appropriate mutation(s) is prevalent enough, it can provide a selective advantage to the virus in the form of resistance. There is also the issue of cross-resistance within those targets for which there are multiple inhibitors. Cross-resistance occurs when a mutation selected for by one inhibitor is inherently resistant to another inhibitor from the same class. If combination therapy is given and works effectively then viral replication is suppressed; however, if for any reason the ARVs are not as effective and the virus is allowed to replicate in the presence of these drugs then it is quite possible that the drug may posses the ability to apply selective pressure to the target warranting a mutation. And once viral resistance occurs, it is almost impossible to control expansion of the viral population [124].

Within each target, signature mutations develop in response to exposure to the antiviral agent used. For example, the M184V mutation lies in the heart of the RT active site and arises in response to targeting by NRTIs. This change from methionine to valine disallows the incorporation of thymidine analog inhibitors and thereby disabling chain termination [125]. Because NNRTIs bind directly to the RT, they are able to block the flexibility of the enzyme and its ability to synthesize DNA; consequently, the residues within this binding site mutate to overcome this targeting [126]. Resistance to entry inhibitors such as enfuvirtide

occurs through the development of mutations in hydrophobic region 1 (HR1) within gp41 that allow for the continued rearrangement of gp41 and its fusion to the host cell via restored communication between HR1 and HR2 [127]. Integrase, being a highly complex multi-domain enzyme, develops resistance via multiple pathways. Perhaps the most well studied is resistance to RAL, which involves mutations at Q148 and N155 within the catalytic core domain (CCD) that restore the coordination of metals by the active site Asp residues blocked by integrase inhibitors [128, 129].

### 1.4.2 Drug resistance in the HIV-1 protease

Resistance to HIV-1 protease inhibitors is multi-faceted, involving mutations within the and outside the active site of the enzyme as well as at protease cleavage sites. Like other targets for which there are multiple drugs within the same class, the protease is also able to develop cross-resistance to multiple PIs, with some mutations in the active site conferring resistance to all PIs such as mutations at V82 of I84 [130]. Because inhibitors are conformationally constrained and make highly specific interactions with active site residues, they exhibit high affinity for the wild-type target but cannot accommodate changes due to mutations unlike substrates or inhibitors that are less constrained [131]. Molecular mechanisms of PI resistance consist of reduced interactions with the inhibitor, shifting of backbone atoms due to mutation, alterations of the dimer interface, transmission of changes perpetuated from outside of the active site and mutations that cause a more relaxed transition state intermediate [132].

With so many non-homologous cleavage sites, the mystery underlying how the protease is not only able to recognize its cleavage sites but also how it is still able to process these substrates despite a vast array of accumulated mutations has only been elucidated in recent years. First came the discovery that the protease's molecular recognition is not in fact guided by amino acid sequence but instead but instead a conserved shape or volume that overlapping substrates occupy when bound at the active site, termed the substrate envelope[133].

The finding of the substrate envelope provided understanding for how some mutations within the protease are detrimental to inhibitor binding but do not affect substrate turnover. The comparison of overlapping of the inhibitors — termed the inhibitor envelope—to the substrate envelope revealed that the inhibitors occupy a certain volume within the active site but also protrude outside the substrate envelope [134]. The sites at which the inhibitors protruded beyond the substrate envelope corresponded to amino acid sites where mutations to inhibitors generally arise. PIs are smaller than natural substrates for bioavailability purposes with shapes different than that of the substrates but containing similar moieties. Because of their size and tendency to fit similarly within the subsites of the enzyme, it was clear that multiple inhibitors could select for the same mutation, thereby causing cross-resistance.

To overcome resistance or otherwise shift the balance between substrate recognition and drug targeting in favor of the latter, it is advantageous that an

inhibitor is able to fit within the substrate envelope. In this case the inhibitor should retain maximum potency despite mutations, as with DRV, which was not designed with the substrate envelope fit as a constraint but does fit well within the envelope. DRV is described as having a high genetic barrier to resistance, which means that multiple mutations within and outside of the active site need to accumulate to confer resistance to DRV [135]. Still, even with this knowledge, the issue of how high levels of accumulated mutations within the protease work together to drive drug resistance to even the most potent of inhibitors remains elusive.

## 1.5 The importance of understanding the role of mutations in PI resistance

### 1.5.1 The role of active site mutations within the protease

No sooner than the first protease inhibitors were developed and administered in clinical trials did the problem of resistance in patients become apparent [136]. Protease inhibitors represent a large class of drugs in ARV therapy, and a larger number of mutations are selected against PIs than other ARV classes. Furthermore, each inhibitor selects for its very own set of primary (major) mutations—those that directly interfere with inhibitor binding—and secondary (minor) mutations—those that do not directly interfere with inhibitor binding. For example, resistance to the first FDA approved inhibitor SQV was driven by mutations G48V and L90M, while RTV selects for mutations at 82, 54, 71 and 36 [137]. Over the years there have been extensive research efforts to

elucidate key mutations in conferring resistance to protease inhibitors. With the improvements in SBDD, more effort has been put forth to predict how a specific mutation may impact inhibitor binding as the inhibitor is designed. The increased level of cross-resistance between PIs has made this objective increasingly necessary. Predicting the order and patterns with which mutations occur still remains a formidable challenge even with the knowledge of the mutations selected by different PIs.

The protease's active site is comprised of the catalytic residues D25, T26 & G27 along with several other residues including R8, L23, A28, D29, D30, V32, K45, M46, I47, G48, G49, I50, T80, P81, V82 & I84 [132]. All nine FDA-approved PIs select for a mutation of at least one, if not more, of these residues. Because these mutations make up the active site, their role in substrate turnover is crucial. Changes at any of these sites not only interfere with inhibitor binding but also with substrate turnover. For instance, the mutation I50V is a major mutation affecting DRV binding while I50L is a major mutation in resistance to ATV, and neither inhibitor is a match for the flexibility and adaptability allowed by the co-evolved substrate [138, 139].

It should be noted that active site mutations impair enzymatic fitness in the absence of drug. However, in the presence of drug in some cases mutations within the active site render the virus as fit or even more fit than the initial wild-type population. Active site mutations often get outcompeted by mutations that do not cause considerable fitness impairment in the absence of drug [140]. Still,

the virus has to find a balance so that it is dually able to carry out its biological function while not being effectively inhibited, and the most readily identifiable ways to achieve this balance is to develop mutations distal to the active site to compensate for any fitness loss endured due to primary mutations.

### 1.5.2 The role of distal mutations in enzyme fitness compensation

Whether PIs selected for their resistance specific major mutations or mutations within the active site, one observation was made clear very early on; those mutations never developed alone. In addition, those major or active site mutations that did arise alone were eventually outcompeted by viral populations containing a combination of major resistance mutations and minor secondary mutations. Some studies early on made the effort of tracking the ordered accumulation of mutations in the protease and in a very precise manner, the protease developed mutations within the active site first and then further developed several minor mutations outside of the active site [137]. More experimental studies showed that after a certain point the additions of more mutations no longer aided in replicative fitness but began to hamper the enzyme's replicative capacity in the presence of drug [141].

Some studies specifically looked at the enzymatic fitness of protease with mutations developed against an inhibitor that selects for them, while others sought to determine if fitness could be restored by mutations not selected for by a specific inhibitor in the absence of drug [140]. There are even some studies that give secondary mutations a larger role in viral replication by examining their

ability to increase virulence and the progression of disease in the absence of therapy [142, 143]. In this regard, strides have been made to predict the evolution and individual fitness scores of these distal mutations over the course of treatment, which showed that patients that accumulated many secondary mutations had higher fitness scores, which equaled higher viral loads, and lower CD4 counts [144]. For secondary mutations, the theories all center around one focus; compensatory mutations have the sole purpose of increasing enzymatic fitness to perpetuate viral replication and disease progression. While the role of secondary mutations in this faction of viral replication has been more or less proven, the role in drug resistance in the absence of primary mutations- remains grossly understudied

### 1.5.3 The role of distal mutations in drug resistance

As mentioned above, protease variants that are able to remain resistant to inhibitors while retaining biological function tend to harbor a number of secondary mutations. Many reports have shown through crystallographic or enzyme kinetics studies that secondary mutations may aid in drug resistance via widening the active site, expanding the substrate envelope itself, and changing the conformation of the flaps from closed to partially open [69, 145-147]. Interestingly, proteases that contain many secondary mutations generally only contain 1-2 active site mutations. Several questions regarding secondary mutations arise that have yet to be fully addressed: Why does the protease accumulate so many non-active site mutations? How do these mutations

communicate changes to the active site and flap regions of the protease? Are there overlapping patterns used by distal mutations to circumvent inhibitor binding?

There have been studies, though few and far in between, that examine the role of secondary mutations in the presence of active site mutations [148, 149]. Despite supporting evidence of distal mutations being able to drive, albeit low levels of, resistance in the form of reduced binding energies and in the absence of active site mutations, very little has been done to decompose the interdependency of mutations beyond the active site alone, collectively, and in conjunction with active site mutations [150-153]. There are a number of studies on distal mutations individually or in pairs, there have been only a limited number of attempts to elucidate the roles of secondary mutations in larger (>10), highly complex combinations of mutations [154-156].

Although there has been a recent surge of studies on clinically derived isolates trying to decipher the role of distal mutations in complex combinations, the issue of elucidating how these mutations work together interdependently remains a mystery. Acquiring a high number of mutations does not directly correlate with resistance suggesting that the effect of mutations in conferring resistance is not additive. For this reason there must be significant effort put forth to gain a complete understanding of resistance in terms of how mutations in small or large mutational contexts are able to drive resistance. Elucidating the

roles of distal mutations in drug resistance may be the key to finally preventing failure of ARV therapy.

### 1.5.4 Possible mechanisms used by non-active site mutations to propagate changes to the active site and flaps of the protease

Our laboratory has previously found that 19 residues outside of the HIV-1 protease active site lie within a highly hydrophobic region, or the hydrophobic core, and conformational dynamic changes involving these residues are intimately linked to flap opening and protease function [157]. The residues within this core facilitate conformational changes via the sliding of residues past one another with minimal energetic expense, termed hydrophobic sliding. Many of the residues within this core region are associated with drug resistance and are thought to alter the conformational flexibility of the protease, which subsequently impact the binding of inhibitor or substrate. Our laboratory has also confirmed that restricting these core residues undergoing hydrophobic sliding resulted in a dramatic loss of function. [158]. Given their roles in protein flexibility and function, studies such as these have laid the groundwork for elucidating the roles of distal mutations in drug resistance.

The examination of distal mutations dynamically has proven to be very insightful. Molecular dynamics studies allow for the following of details that cannot readily be seen experimentally or crystallographically. Through molecular dynamics simulations, Mittal et al were able to see just how much enzyme

flexibility was reduced when residues within the hydrophobic core were engineered in affixed positions, causing reduction in enzymatic function [158]. It is with dynamics that Appadurai et al were able to ascertain that protease distal mutations cause the active site to become uncoupled with the flaps causing a loss of contacts with the inhibitor [153].

### 1.5.5 Secondary mutations and drug resistance in non-protease targets

The HIV-1 protease is not the only target that makes use of secondary mutations to drive resistance. Secondary mutations pose a threat to resistance for many disease targets within and outside of HIV-1. The HIV-1 integrase, whose first inhibitor has been approved within this decade, develops secondary mutations to circumvent inhibitor binding. A study by Nakahara et al. found that primary mutations in the integrase enzyme Q148K or Q148R reduced the efficiency by which integrase was able to incorporate viral DNA and that this loss of function was restored upon the addition of secondary mutations E138K or G140S. They also found that the combination of G140S and Q148R was able to drive a greater and 6 fold increase in resistance to one of the compounds tested [159]. Because of the increased fold change in resistance and the ability to restore viral replication and infectivity, they deduced that the combinations of primary and secondary mutations within the integrase evolved due to the specific abilities that the presence of the secondary mutations could restore.

Beyond HIV-1, anti-cancer therapies are thwarted by the development of drug resistance. A well-studied example is the epidermal growth factor receptor (EGFR) family of receptor tyrosine kinases (RTKs), which stimulates non-small cell lung cancer (NSCLC). This transmembrane protein is activated upon binding ligands like EGF, which in turn causes the homodimerization of two EGFRs and the autophosphorylation of the TK domain and downstream cell signaling. However, the signaling events initiated by EGFR can be dysregulated due to activating mutations within the TK domain. Although some activating mutations make the TK domain more sensitive to the inhibitors gefitinib and erlotinib, there are some that cause the TK domain to remain in a constitutively activated state and become resistant to inhibitors. To overcome the hypersusceptibility of some activating mutations, EGFR develops a secondary mutation T790M, which is important for regulating inhibitor specificity and increasing affinity for ATP within the binding pocket of EGFR allowing ATP to outcompete the inhibitors [160, 161].

More cancer-based examples include secondary mutations within BRCA1 and BRCA2 in breast cancer that allow functions to be restored in BRCA1/2 mediated tumors along with resistance to inhibitors, and secondary mutations in FLT-3 in acute myeloid leukemia. Given that secondary mutations in any drug resistant target may primarily work in the same manner, there is a need to develop methods to combat resistance caused by both primary and secondary mutations. One interesting methodology being explored in improvement of cancer therapy is potentiating EGFR for degradation. Another general strategy

would be developing inhibitors that would work better against targets with both types of mutations, and specifically for cancer targets, developing inhibitors that work in targets downstream in signaling pathways while simultaneously inhibiting EGFR.

## 1.6 The importance of protein dynamics in drug binding and resistance

### 1.6.1 Visualizing protein dynamics to understand resistance

To truly and completely understand how a protein functions, one must be able to probe both structural and dynamic properties. X-ray crystallography has been invaluable to visualizing, understanding and targeting with small molecules of many proteins. Crystallography has even been able to capture some dynamic movements of the HIV-1 protease at high resolution [69, 135]. Still, a great deal of information is missed by crystallography in that only a snapshot of the protein or biomolecule movement within the crystal lattice can be captured. Because of this limitation, methods that allow for the study of protein dynamics have enhanced our acumen of molecular recognition. There are both experimental and computational methods used to study protein dynamics, including using X-ray crystallography in a time resolved manner [162]. Other experimental methods to study protein dynamics include nuclear magnetic resonance (NMR), electron paramagnetic resonance (EPR), and hydrogen-deuterium exchange [163-165].

Although these experimental methods offer the ability to monitor protein dynamics, probing dynamics computationally via molecular dynamics (MD)

simulations allows for the most highly detailed visualization. When used in combination, experiments and MD simulations complement each other in a manner by which most researchers can expound upon structural changes relevant for function. For example, the 150-loop in influenza neuraminidase contains the catalytic residue D151 and is highly flexible. The 150-loop is so flexible in some strains that it is not ordered in the crystal lattice and hence cannot be seen in x-ray crystal structures. However, Bush and coworkers found that the 150-loop can be stabilized by a salt-bridge through use of crystallography and MD simulations thereby making this site a possible target for inhibitors [166].

While all the methods mentioned above have their limitations, the method chosen may depend on several factors that are important to the experimentalist. For instance, the time scale for which one wishes to run the experiment, the size of the macromolecule and properties of the system set in place during the experiments should all be taken into consideration prior to choosing a method. With advancements in computation (the advent of GPUs and computing clusters), increased power of synchrotrons and the automation of several methodologies have catapulted these methods into a territory where there are very few limitations remaining.

### 1.6.2 The role of dynamics in HIV-1 protease resistance

The early structural studies on HIV-1 protease provided a foundation for HIV-1 treatment today; however, what no one could glean from the structures of

the protease was its flexibility and just how this flexibility, or lack thereof, modulates activity. In the decades since those initial structural studies, technology has brought the HIV-1 treatment field a long way by allowing for increased insight into protein dynamics. From gaining the now straightforward knowledge that the proteases movement is not of the simple garden-variety hedge shear to the ability to understand the plasticity with which substrates are cleaved, it is clear that protein dynamics have catapulted our understanding of this small but mighty enzyme.

The protease flap region is highly flexible with intricate workings designed to retain this feature in the absence and presence of mutations [157]. Studies have shown that while both wild-type and mutant protease variants are highly flexible in the absence of inhibitor, the presence of mutations increases this flexibility extensively within the flaps [167]. The protease's flexibility is not limited to its unbound form; it remains flexible even when bound to ligand as well [168-170]. As mentioned previously, the presence of mutations causes the flaps and active site of the protease to become uncoupled, resulting in a loss of contacts specifically to the inhibitor. It is with the understanding of both static and dynamic alterations afforded by mutations within the protease that we will be able to further elucidate the interdependencies of the mechanisms underlying drug resistance.

### 1.6.3 The use of computational methods to probe dynamics in relation to resistance

The use of MD simulations to aid in structure based drug design and general visualization of small molecule dynamics is ever increasing. For instance, just this year Nagasundaram and co-workers were able to identify several anti-malarial compounds that may be able to combat resistance via virtual screening and MD simulations [171]. However, the use of MD simulations alone may only be able to provide finite detailed information about conformational changes that may regulate drug resistance. Although there are copious amounts of information that can be determined from resultant MD trajectories, it is up to the user to identify which interactions may be key in understanding drug resistant mechanisms used by their target.

In the case of HIV-1 protease, there is a wealth of knowledge pertaining to drug resistance including which mutations arise to which inhibitors, which mutations underlie cross-resistance and that there is no direct correlation with the number of mutations to the severity of resistance. We have been able to identify which interactions play a key role in inhibitor binding and have tried to take advantage of these interactions and improve upon them [172]. This approach does seem a bit retroactive in that there are still no defined mechanisms of resistance used by distal mutations to drive drug resistance. Some studies have used machine learning techniques and algorithms to predict genotypic and phenotypic resistance from large datasets [173]. However, we have learned with HIV-1 protease one simple rule of thumb: resistance begets resistance and if

viral suppression does not occur, this cycle will continue. The simple prediction of resistance to better shuffle ARVs for a more effective regimen will eventually lead to yet another shuffling in a matter of time. In addition to trying to improve inhibitors, we must try to decompose the interdependency employed by distal mutations to propagate changes to the active site of the enzyme. Identifying and tracking key interactions between a drug resistant protease or a set of proteases throughout the duration of MD simulations may only be the first steps on the road to unlocking drug resistant mechanisms used by mutations outside of the active. Going a step further and using machine learning and statistical techniques to parse through any patterns used by mutations may prove to be the pivotal in understanding drug resistance from an omniscient perspective.

## 1.7 I.VII. Thesis Scope

This thesis attempts to fill the gaps surrounding the role of mutations distal to the active site in drug resistance of the HIV-1 protease. Moreover, this thesis tries to elucidate the interdependent mechanisms employed by distal mutations to drive resistance to the highly potent protease inhibitor darunavir and how the elucidation and exploitation of these interdependent mechanisms can proactively meliorate both current treatment and inhibitor design for targets within and outside of HIV-1.

First I demonstrate that mutations outside of the active site are able to drive resistance independently of active site or other distal mutations using information from static and dynamic structural studies of a small panel of

protease variants. I also identify and probe key inter-molecular interactions that are manipulated by the presence of distal mutations to alter the dynamic ensemble of mutant protease-inhibitor complexes (Chapter II).

I then expand the panel of protease variants to include clinically-derived multi-drug resistant (MDR) variants that contain many distal mutations in complex combinations. I monitor the key interactions I identified previously via MD simulations in these highly mutated proteases. I present that, with the use of unsupervised machine learning and statistical techniques, mutations outside of the active site are able to operate interdependently to drive resistance to the protease inhibitor darunavir (Chapter III).

How RNA may be able to enhance the catalytic efficiency of MDR proteases and the significance of these findings with respect to inhibitor binding in non HIV-1 targets will be discussed in Chapter IV.

# Chapter II

# Drug Resistance Conferred by Mutations Outside the Active Site Through Alterations in the Dynamic and Structural Ensemble of HIV-1 Protease

## 2.1 Abstract

HIV-1 protease inhibitors are part of the highly active anti-retroviral therapy effectively used in the treatment of HIV infection and AIDS. Darunavir (DRV) is the most potent of these inhibitors, soliciting drug resistance only when a complex combination of mutations occur both inside and outside the protease active site. The role of mutations outside the active site in conferring resistance remains elusive. Through a series of DRV–protease complex crystal structures, inhibition assays, and molecular dynamics simulations we find that single and double site mutations outside the active site often associated with DRV resistance alter the structure and dynamic ensemble of HIV-1 protease active site. These alterations correlate with the observed inhibitor binding affinities for the mutants, and suggest a *network hypothesis* on how the effect of distal mutations are propagated to pivotal residues at the active site and may contribute to conferring drug resistance.

## 2.2 Introduction

In the absence of a vaccine and in lieu of a cure, antiretroviral combination therapy has been the main form of treatment for individuals infected with HIV. As is the case with treatment of most rapidly evolving viruses/diseases, drug resistance decreases the effectiveness of treatment. The high replicative capacity of HIV and the infidelity of the reverse transcriptase quickly lead to a

heterogeneous population of viruses within patients, from which resistance has emerged to all 30 of the currently used anti-viral drugs.

HIV-1 protease inhibitors (PIs) have recently emerged as the most effective drugs in the treatment of HIV [174-176]. PIs are competitive active site inhibitors that mimic the transition state of the enzyme and are the most potent antiretroviral drugs for the treatment of HIV/AIDS [110]. These drugs are ideal for therapy as they target the viral protease responsible for viral maturation and thus the spread of the virus. Unfortunately, the rapid evolution of HIV-1, coupled with the selective pressure of therapy, results in many viable multidrug resistant variants. In fact mutations at 45 of the 99 residues that make up HIV-1 protease have been implicated in drug resistance [177]. While resistance due to mutations at 11 of these 45 residues can be explained as direct changes within the active site, the resistance mechanisms for the remaining mutations outside the active site of the enzyme mostly remain elusive.

Drug resistance mutations in HIV-1 protease allow the enzyme to become less susceptible to inhibition while retaining enzymatic activity. Points of inhibitor–protease contact at residues within the active site where the inhibitor protrudes beyond the substrate envelope are sites selected for resistance, as their interactions are more critical for inhibitor binding than substrate turnover[134]. While mutations at some active site residues, such as 82 and 84, lead to resistance to all PIs, other mutations are signatures of specific inhibitors, such as D30N for nelfinavir and I47A for lopinavir [178]. These mutations directly impact

inhibitor binding by altering or reducing contacts necessary for inhibiting the enzyme, but can also simultaneously decrease the catalytic efficiency or enzymatic fitness. The mutations at the remaining 34 of the 45 residues associated with drug resistance occur outside the active site. These changes have often been considered secondary or accessory mutations, and are thought to indirectly impact inhibitor binding while assisting in enzyme fitness or stability. Structural studies on the effect of several HIV-1 protease secondary mutations have provided insights into how inhibitor binding may be affected [156, 157, 179-181]. However for the most part, their specific role in protease inhibitor resistance or mechanism of action has not been elucidated.

Darunavir (DRV) is the most potent of the FDA approved HIV-1 protease inhibitors. This high potency combined with the inhibitor's fit within the substrate envelope appears to account for DRV's robustness against drug resistance[182, 183]. Drug resistance to DRV usually occurs only in patients who have high levels of pre-existing PI resistance, requiring at least seven mutations to simultaneously occur for therapeutic failure. In fact, DRV is being investigated as a potential mono-therapy in treatment-naïve patients [184].

In DRV-resistant HIV variants, many changes occur outside the active site of the enzyme in complex combinations. Single site mutations cannot confer high levels of resistance to DRV, and a combination of multiple mutations including those outside the active site are needed to decrease potency. However the role of these mutations in conferring resistance is not well understood: some may be

enhancing enzymatic activity, while others may directly confer drug resistance and still others may be residual mutations from previous therapy history. In this study, we examine some of the most common of these mutations—V32I, L33F, L76V, and L90M (as a control; not a signature of DRV resistance but frequent in multidrug resistance [180]—for their impact on DRV inhibition. Using a combination of static and dynamic structural analyses, by determining crystal structures of complexes and performing molecular dynamics simulations, we elucidate the possible roles of these secondary mutations both independently (L76V, L90M, V32I) and in combination (V32I/L33F) in conferring resistance. We find how mutations at residues with no direct contact with the inhibitor can alter the structure and dynamics of the protease to affect inhibitor binding through common mechanisms, which we define through a “network hypothesis”.

## 2.3 Results

To determine how mutations remote from the active site contribute to DRV resistance in HIV-1 protease, the impact of four mutations (L76V, L90M, V32I, and V32I/L33F; Figure 2.1) in a subtype B background was investigated in terms of enzyme inhibition, inhibitor-bound crystal structures, and molecular dynamics simulations.

### 2.3.1 Enzyme Inhibition

The enzyme inhibition constant for DRV was measured against each of the protease mutants, in addition to WT subtypes B and C for comparison (Table 2.1). DRV is highly potent against WT subtype B protease with a $K_i$ of 2 pM, as

we previously reported [182]. The level of inhibition for the mutants varied from 2 pM to 45 pM, with the L90M mutant being inhibited as potently as the WT protease and the V32I/L33F double mutant exhibiting the greatest decrease in susceptibility to DRV with a fold-change greater than 20. Hence, single mutations are not enough to confer high levels of DRV resistance, as expected, and the mutations had varying degrees of effects on DRV susceptibility.

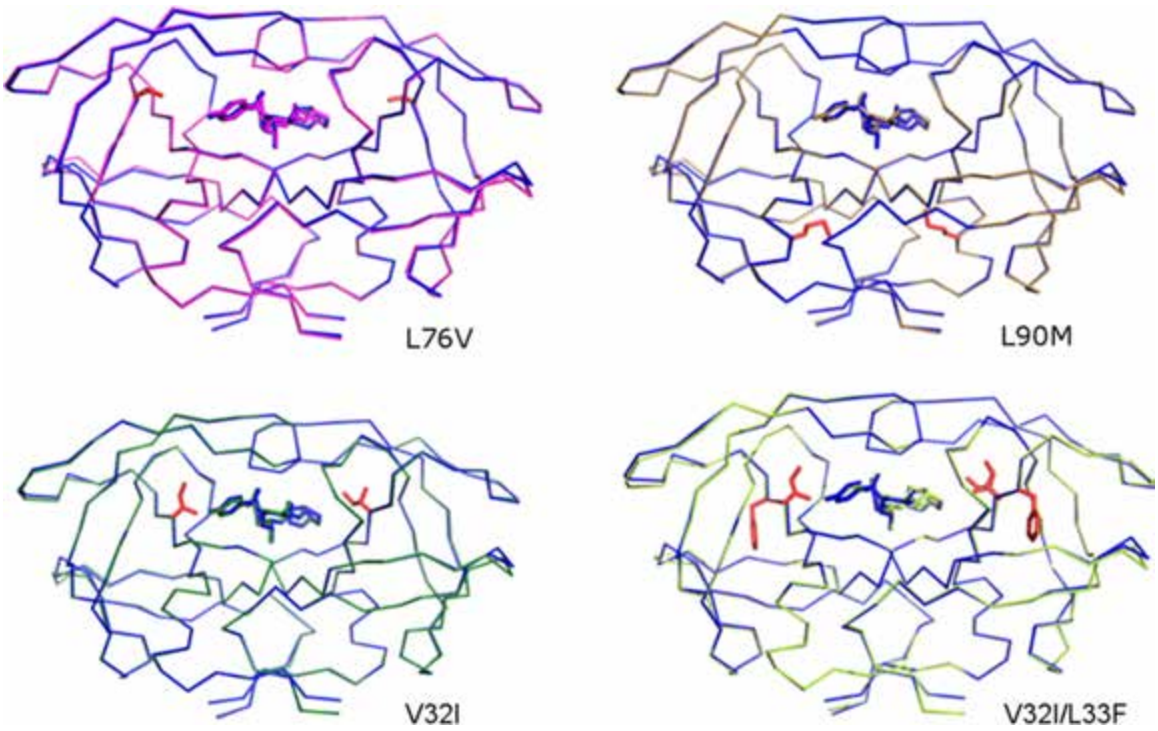



**Figure 2.1 Structure of HIV-1 protease variants bound to DRV. Crystal structures of mutant protease variants superimposed with the WT protease complex structure in blue. The side chains od mutation sites are in red sticks.**

**Table 2.1 DRV interaction and susceptibility of HIV-1 protease variants. DRV inhibition constants (Ki) of HIV-1 protease variants, with fold changes relative to subtype B WT protease in parentheses. The overall vdW interaction energy between inhibitor and protease was determined from crystal structures.**

| protease variant | $K_i$ (pM) | vdW (kcal/mol) | ΔvdW (kcal/mol) |
|---|---|---|---|
| subtype C | 5 ± 2 (2.5) | | |
| WT | 2 ± 1 (1.0) | −44.5 | |
| L76V | 3 ± 2 (1.5) | −43.0 | 1.5 |
| L90M | 2 ± 2 (1.0) | −44.4 | 0.1 |
| V32I | 7 ± 9 (3.5) | −44.2 | 0.2 |
| V32I/L33F | 45 ± 18 (22.5) | −43.3 | 1.2 |

### 2.3.2 Crystal Structures

To structurally characterize the effects of the mutations on DRV binding, we determined the crystal structures of variants L76V, L90M, V32I, and V32I/L33F, which diffracted to resolutions of 1.5–1.9 Å in the $P2_12_12_1$ space group (Table 2.2). Alignment of the four complex structures on our previously determined structure of the WT protease–DRV complex (1T3R [70]) showed that the variants had only minor backbone variations, mainly in the 20s loop likely due to crystal packing differences (Figure 2.1). Therefore, the mutations had very little impact on the overall backbone structure of the protease.

**Table 2.2 Crystallographic statistics for DRV-Bound HIV-1 protease structures. *a* denotes structure used from [70].**

| | WT[a] | L76V | L90M | V32I | V32I/L33F |
|---|---|---|---|---|---|
| PDB code | 1T3R | 3OY4 | 4Q1W | 4Q1X | 4Q1Y |
| data collection | | | | | |
| space group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| *a* (Å) | 50.75 | 50.72 | 50.74 | 50.91 | 50.81 |
| *b* (Å) | 57.81 | 57.71 | 57.55 | 58.28 | 58.23 |
| *c* (Å) | 62.01 | 62.03 | 61.81 | 61.78 | 61.82 |
| *b*-angle | 90 | 90 | 90 | 90 | 90 |
| *Z* | 2 | 2 | 2 | 2 | 2 |
| temperature (K) | 190 | 100 | 100 | 100 | 100 |
| resolution (Å) | 1.20 | 1.76 | 1.85 | 1.90 | 1.50 |
| total reflections | 55 056 | 133 462 | 107 138 | 100 241 | 192 411 |
| unique reflections | 52 226 | 17 434 | 15 822 | 14 168 | 29 633 |
| *R* merge (%) | 3.8 | 8.5 | 6.4 | 6.3 | 4.8 |
| completeness (%) | 95.5 | 97.8 | 98.6 | 93.9 | 98.7 |
| crystallographic refinement | | | | | |
| *R* value (%) | 14.0 | 18.0 | 17.3 | 18.0 | 18.3 |
| *R* free (%) | 17.9 | 21.9 | 20.6 | 23.4 | 20.7 |
| RMSD in: | | | | | |
| bond length (Å) | 0.039 | 0.006 | 0.008 | 0.008 | 0.009 |
| bond angles | 1.593 | 1.572 | 1.342 | 1.474 | 1.497 |

[a]From Surleraux et al.

### 2.3.3 Detailed Structural Analysis of DRV Binding from Co-crystal Structures

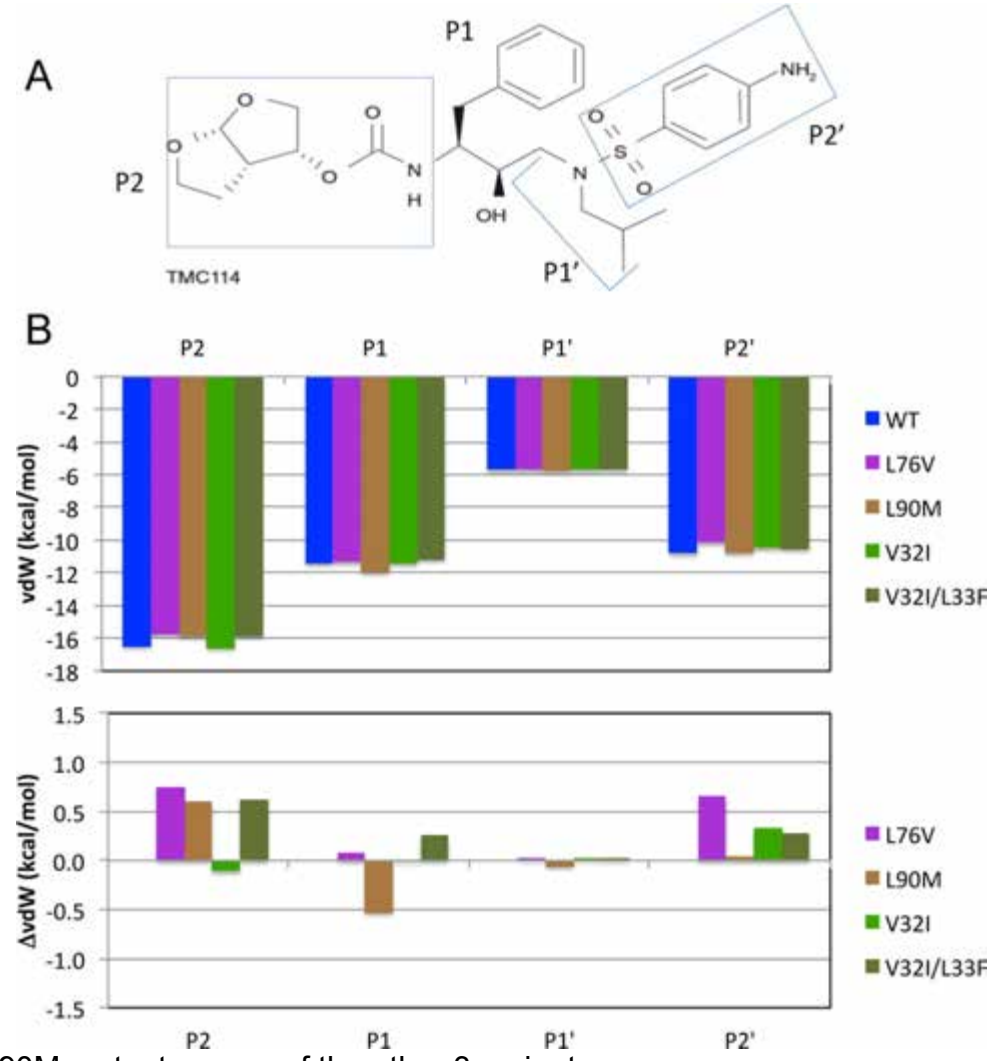
The high-resolution co-crystal structures enabled detailed analysis of protease–DRV contacts in each of the five complexes. The WT complex had the most extensive van der Waals (vdW) contacts with the inhibitor with a favorable energy of –44.5 kcal/mol, similar to V32I and L90M variants (Table 2.1). The L76V variant and V32I/L33F double mutant lost more than 1 kcal/mol in vdW contact energy with DRV relative to the WT complex. Thus despite no large-scale changes in the protease backbone, subtle changes in repacking occurred around

DRV in these two complexes to weaken protease interactions with the inhibitor. However, the extent of contacts lost with DRV in the mutant crystal structures with respect to WT protease does not correlate completely with the fold-change loses in $K_i$ values (Table 2.1).

Contacts involving specific DRV moieties (Figure 2.2) and protease active site residues (Figure. 2.3) were analyzed in detail. In general, the impact of mutations on DRV contacts are larger at the P2 and P2' than the central P1 and P1' moieties. The bis-THF group of DRV P2 moiety forms the most extensive contacts in all of the complexes (Figure 2.2), but also loses considerable contacts due to the mutations, except in the V32I structure. In the case of V32I, DRV contacts are retained as in the WT complex, consistent with no significant change in total vdW or $K_i$ values (Table 2.1). When this mutation occurs together with L33F in the double mutant though, contacts are lost in all three of P2, P1 and P2' moieties. In L90M variant, although interactions get weaker at the P2 position, gain of contacts at P1 compensate for this loss yielding comparable total vdW contacts and susceptibility to DRV as WT protease.

While the apo form of the protease is a symmetric homodimer, DRV induces asymmetry to the complex and thus despite identical residues mutating in both monomers, the effect of these mutations on protease–inhibitor contacts is distinct in the two monomers (Figure 2.3). Specifically, L76V and L90M mutations cause considerable loss of contacts at I47, but to a lesser extent at I47'. Other active site residues whose contacts are altered in mutant structures include I50

at the tip of the flaps, and 81-82-84 at the 80s loop. Contrary to previous [132], we do not see any major enhancement of DRV contacts with the catalytic D25 in




the L90M mutant, or any of the other 3 variants.

**Figure 2.2 Contacts of DRV moieties with HIV-1 protease variants. A. Chemical structure of DRV(TMC114) with the inhibitor moieties P2-P2' indicated. B. vdW interaction energy(kcal/mol) of DRV moieties for contacts with the protease active site in the crystal structures, and changes in vdW interaction energy in mutant structures relative to the WT complex. Positive values indicate loss of contacts.**

Residue 32 is at the periphery of the active site, and V32I mutation causes a unique pattern of rearrangement of inhibitor contacts than the other variants studied. Unlike L76V and L90M, contacts with 47 are retained in V32I. Although the backbone is not shifted significantly, the proximity of residue 32 to the 80s loop causes subtle rearrangements to result in repositioning of the DRV away from I84's and more towards I50's at the tip of the flaps. As a result, DRV contacts with residues 84, 81' and 84' are lost but those with 50 and 50' are enhanced. The larger isoleucine also forms additional contacts with the inhibitor in the unprimed-side monomer. In the V32I/L33F variant, which loses an additional 7-fold in binding affinity relative to V32I (Table 2.1), the contacts are rearranged again. In contrast with V32I alone, additional loss of interactions at residues 47 and 50 are observed. These losses of contacts are similar to the alterations observed in the L76V and L90M variants. Thus the double mutant V32I/L33F variant alters the active site in a synergistic manner, leveraging both alterations similar to L76V and L90M, and some changes from V32I. The change in variants' affinity is not simply due to a loss of van der Waals contacts, but an interdependent change in optimal contacts.

In the WT complex, DRV forms a network of hydrogen bonds within the active site involving both backbone and side chains. Most of these bonds, including the two water-mediated ones with I50, are conserved in the variant complexes. Two exceptions occurred in the L76V and V32I complexes:

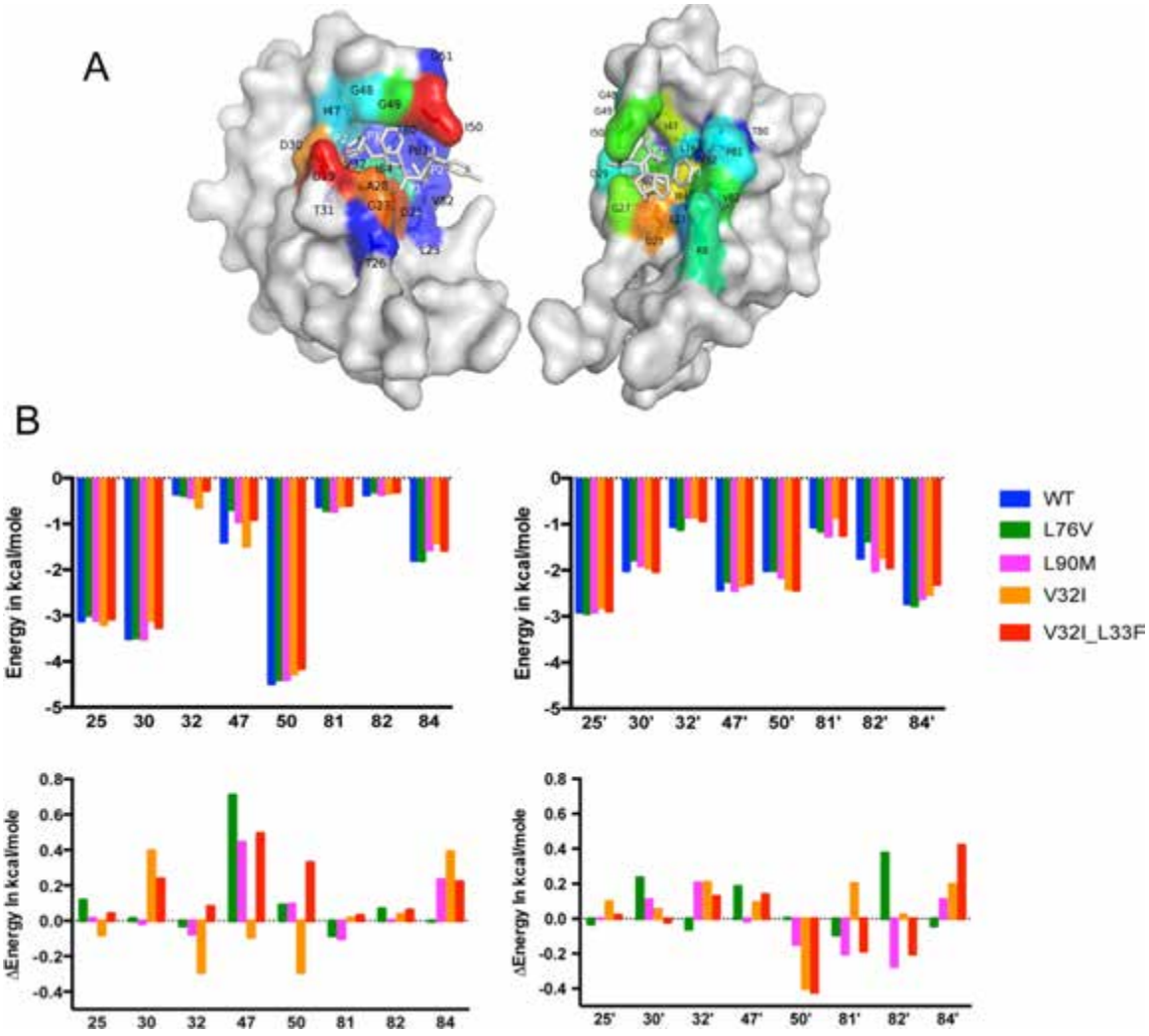



**Figure 2.3 Contacts of protease active site residues with DRV in the crystal structures. A. The two monomers of WT protease in surface representation with the bound DRV displayed as sticks. Active site residues are colored from blue to red for increasing vdW contacts with the inhibitor. The monomer that interacts mostly with the P2-P1 moieties of DRV is on the left, and the primed-side monomer is on the right. B. The vdW interaction energy of active site residuesin crystal structures (top), and changes in mutant complexes relative to the WT structure (bottom). Only the residues displaying considerable changes relative to WT are included for both monomers. See Figure 2.10 for information on changes in all active site residues. Positive values indicate loss of contacts.**

Consistent with the loss of vdW contacts, in the L76V complex a hydrogen bond to the backbone of D30 is lengthened from 2.0 to 3.0 Å. In the V32I variant, an additional water-mediated hydrogen bond with the side chain of D30 is formed. Nevertheless, overall the hydrogen bonds with DRV within the various complexes are conserved.

### 2.3.4 Dynamic Simulations of Complexes

Analysis of crystal structures above revealed that the mutations away from the active site are able to influence interactions of DRV-contacting residues at the active site. The alterations in vdW contacts or hydrogen bonds lost, however, only partly correlate with the experimentally determined enzyme inhibition constants. Another possible mechanism by which these secondary mutations could alter inhibitor binding is by influencing the dynamic ensemble of the enzyme.

Starting from the crystal structures of the DRV complexes, three replicates of fully hydrated 10 ns MD simulations of each DRV complex were performed and analyzed. Root-mean-square deviations (RMSD) of Cα atoms during the simulation and the average root-mean-square fluctuations (RMSF) about their mean positions readily reveal that the secondary mutations alter the overall enzyme dynamics (Figure 2.4, Top). The L90M and V32I/L33F variants display larger fluctuations throughout the enzyme compared to the other variants, although the catalytic D25 stays relatively rigid in both monomers. These altered

fluctuations are not restricted to the sites of mutation, but propagate throughout the enzyme.

To further analyze the impact of mutations on the dynamic ensemble sampled by the protease, the distance distributions were calculated across the active site at a variety of positions (Figure 2.4, Bottom and Figure 2.5). In all the variant complexes, the dynamic ensemble sampled by the protease is altered relative to WT. Many of the distances displaying a significant change are longer than the WT distance, indicating a widening of the active site. In the L76V complex, the changes are highly asymmetric, with one side of the active site constricting and the other widening (Figure 2.4 Bottom). In all variant complexes, alterations involve residues in the 80s loops. The 80s loops in both monomers form the “side walls” of the active site. Relative to the WT complex, the distance between residues 81 in the two monomers are shorter, and that between 84–84′ are longer in all variants. Hence, the “upper” part of the side walls are closer and the “lower” part is farther away in the mutant complexes compared to the WT. In addition to the 80s loop, certain distances involving residue 50 at the tip of the flaps, and even the catalytic D25 are altered in the variant complexes. The catalytic site is the most invariant and dynamically restricted region of the protease, both when different crystal structures are compared and dynamics analyzed by simulations and NMR experiments [167, 168, 185]. Therefore, widening of the D25–D25′ distance in the V32I/L33F double mutant is an unusually profound impact of remote changes on the catalytic region.

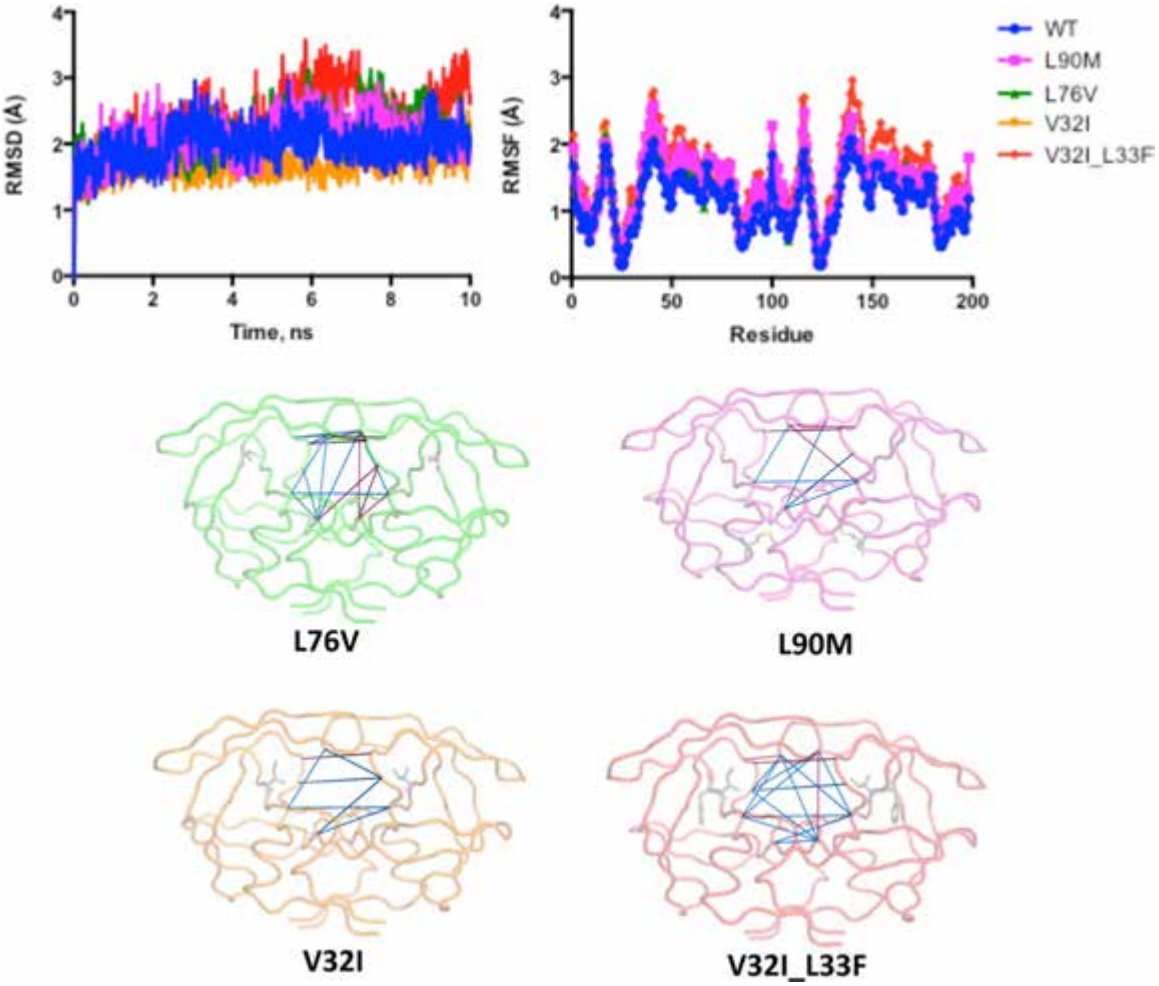



**Figure 2.4 Molecular dynamics simulations of DRV-HIV-1 protease complexes. Top. RMSD of Cα atoms from the initial positions, and RMS fluctuations of residues averaged over three 10ns trajectories. Bottom. Significantly altered change in distance between residue pairs around the active site relative to WT complex, smapled during the MD simulations; increased and decreased distances are indicated by blue and red respectively.**

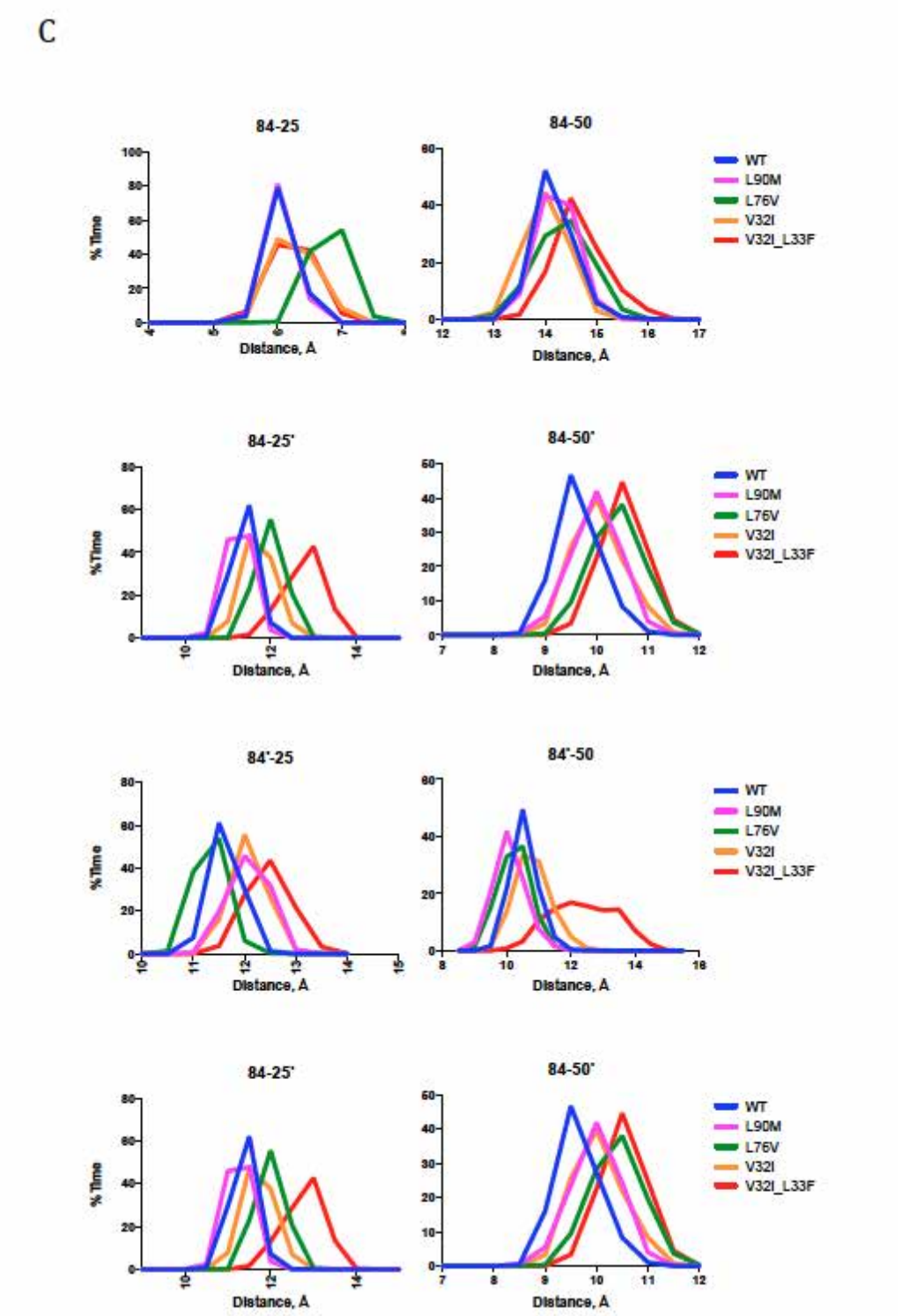



**Figure 2.5 Sample of distance distributions between residue pairs around the active site over three 10ns MD trajectories of DRV-bound HIV-1 protease complexes. Distances involving residue 84 shown specifically.**

The MD simulations also permitted a detailed analysis of the interaction network both for direct interactions within the active site to DRV and the internal hydrogen-bonding network throughout the enzyme (Figure 2.6 and Figures 2.7-9). Throughout the MD simulations the WT complex maintains a network of stable hydrogen bonds. Starting from the bottom of the enzyme the c-terminal α-helix forms a network of hydrogen bonds that links the termini of the protein to the flap regions. The backbone of residue 95 links to residue 90 which in turn contacts residue 86, residue 88 bridges to residues 29, 31, and 74, and residue 76 bonds to both residues 31 and 33 which is bonded to residue 78. Residues 29 and 30 make direct hydrogen bonds to DRV in both monomers. The hydrogen bonds linking residues 47 and 54 within the flaps stay tightly hydrogen bonded throughout the simulation. Thus, as we previously observed[157], the hydrogen bonding network is stably retained within the WT MD simulation.

In comparing the simulations of the variant DRV complexes with the WT, subtle changes are seen in the vdW contacts within the active site (Figure 2.9), similar to what was observed in the crystal structures. However, in each of the four variants, with the notable exception of the 47–54 linkages, the hydrogen bond network is disrupted to a greater or lesser extent asymmetrically, including the direct hydrogen bonds with DRV (Figure 2.6B and 2.7-8). The V32I/L33F variant is the most disrupted with 12 hydrogen bonds changing by greater than 20% relative to the WT complex throughout the dimer, with 11 being weakened (Figure 2.6D) including most dramatically the interactions of the side chain of Asn

88. Eight of these changes are within the monomer that coordinates the highly rigid bis-THF moiety including weakened interactions at points of contact between the protein and DRV. Thus, mutations distal to the active site often weaken the strength of the hydrogen bonds in the network, which is propagated through to the active site including altering vdW packing, pushing the flaps, and thereby the contact of I47 with DRV. Taken together, one can decipher how contacts between protein and inhibitor are affected by these changes even for residues that are packed through vdW contacts or covalently linked along the backbone and are not directly involved in the hydrogen bonding network (Figure 2.6A and C).

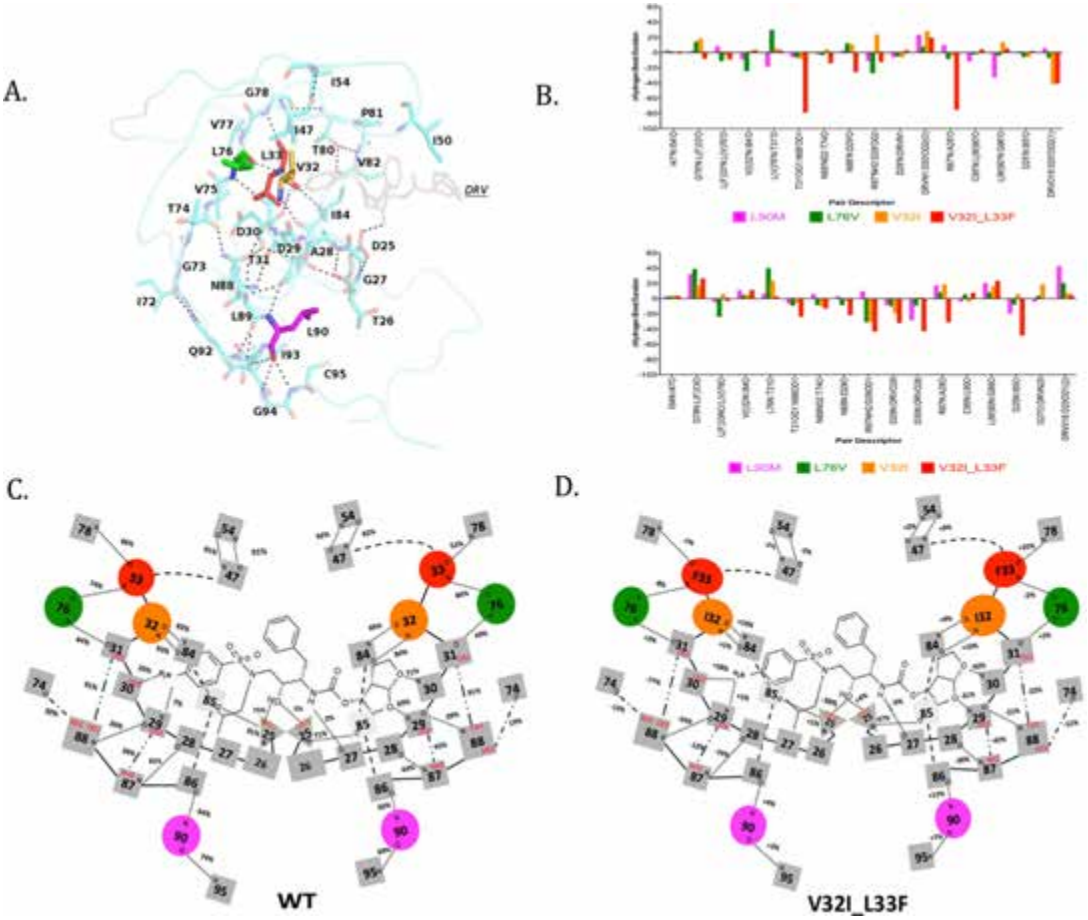



**Figure 2.6 Network of hydrogen bonds within HIV-1 protease. (A) Crystal structure of DRV bound to the active site, and only one monomer of the protease is shown for clarity. The sites of mutation (L76, L90, V32, and L33) have colored side chains. (B) Histograms of the changes of the percentage time hydrogen bonds are formed relative to the WT simulation for each of the complexes. (C) Schematic hydrogen bond network of the HIV-1 protease dimer with the percentage time hydrogen bonds are formed during the WT simulation (Figure 2.7). (D) Schematic representation of the V32I_L33F complex simulation with the change in hydrogen bonding relative to the WT simulations. The remaining variants schematic are shown in Figure 2.8.**

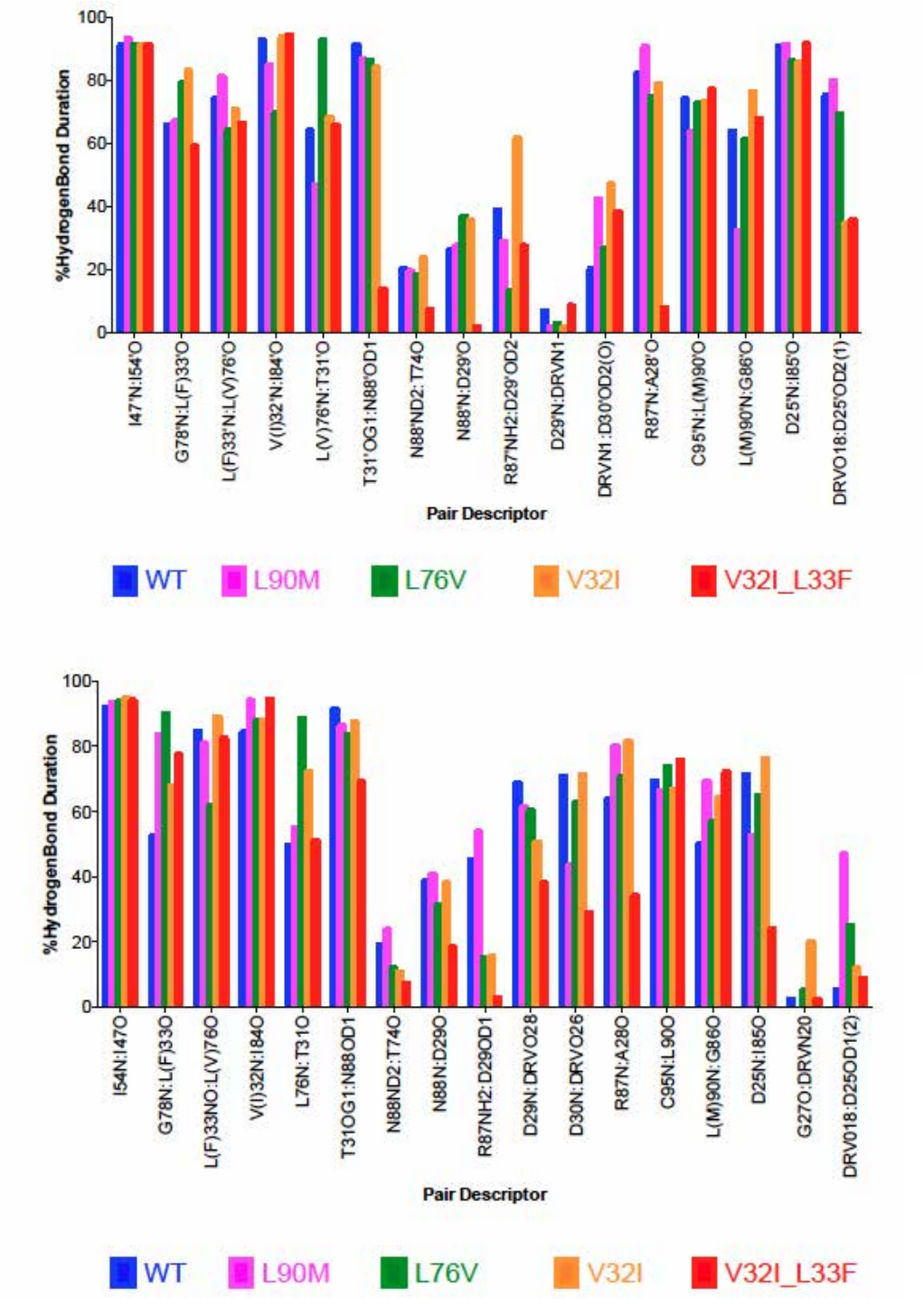



**Figure 2.7 Histograms of total hydrogen bond duration as percentage of time during MD simulations. Top: Duration between specified residue pairs in the monomer binding the P2'moiety of DRV. Bottom: hydrogen bond duration in the monomer binding the P2 moiety of DRV**

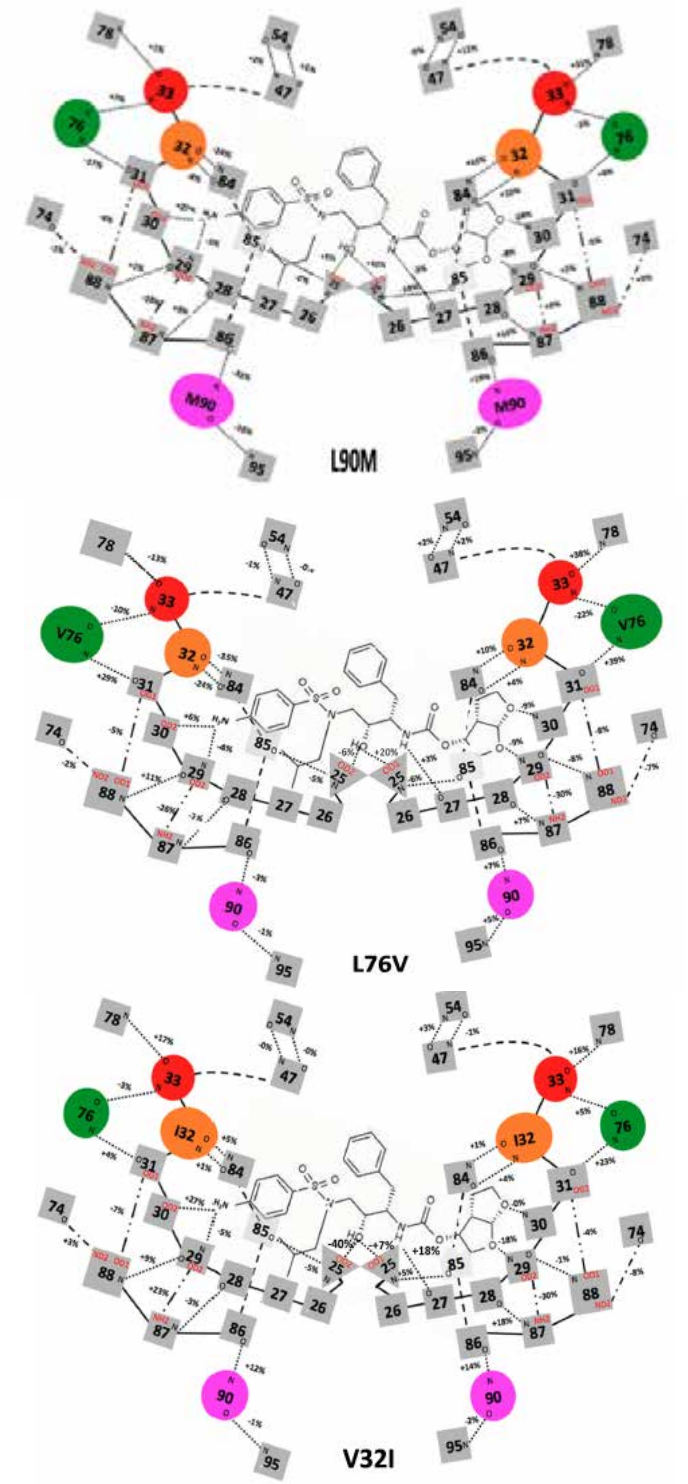
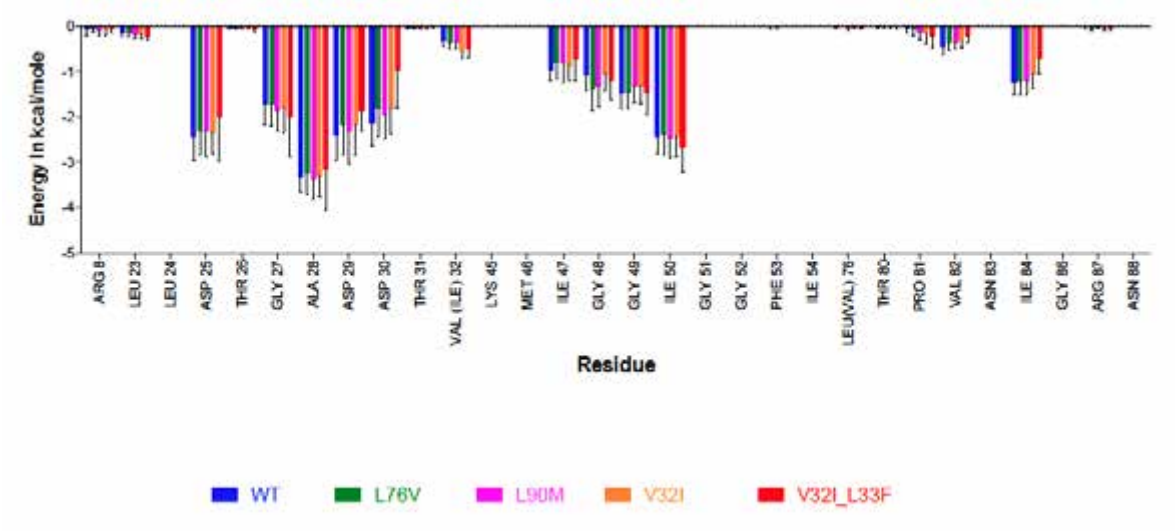



**Figure 2.8 Schematic representation of hydrogen bonding over the course of the simulation for L90M, L76V and V32I, percentages are relative to the length of time observed for the WT simulation (Fig 2.6 B, C and 2.7).**



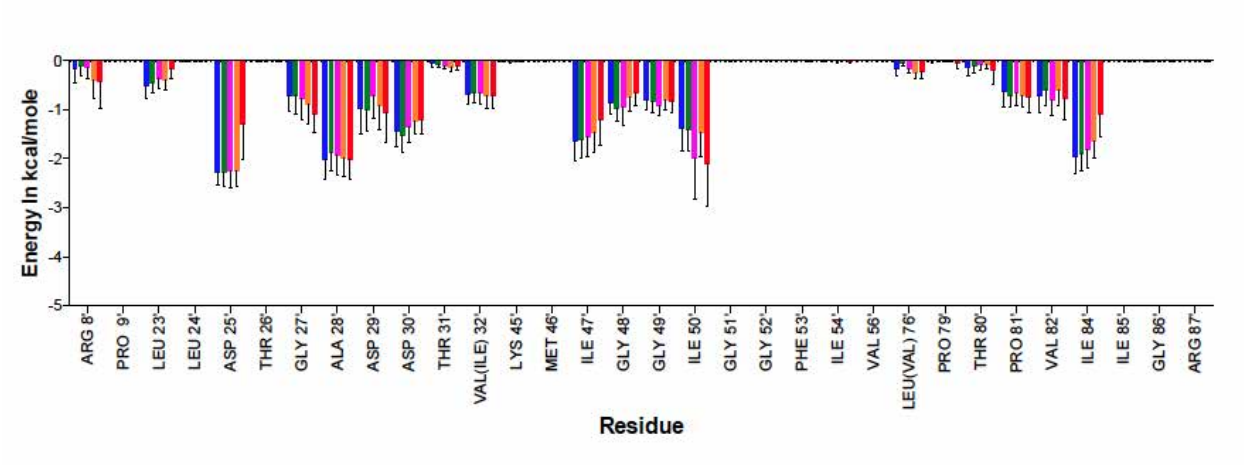




**Figure 2.9 Histograms of vdW between DRV and protease during MD simulations.**

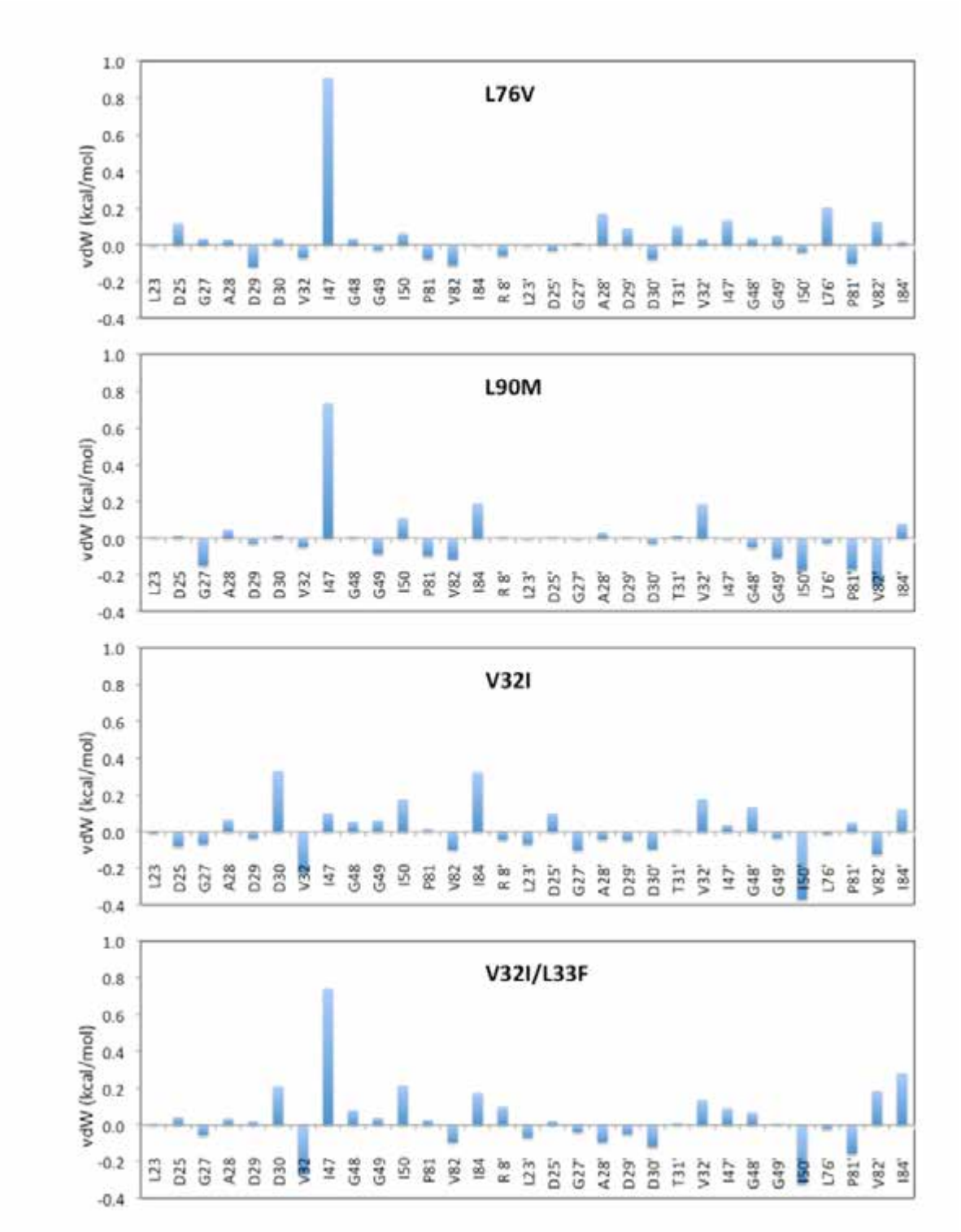



**Figure 2.10 Changes in vdW interaction energy (kcal/mol) of active site residues with DRV in HIV-1 protease crystal structures relative to the WT complex.**

## 2.4 Discussion

HIV-1 protease evolves in complex combinations to evade inhibition, but still maintains biological function. The active site mutations have a relatively straightforward mechanism of disturbing the inhibition–function balance, which is effectively explained by the substrate envelope[134]. However, in highly resistant variants, active site mutations often coexist with mutations outside the active site. This is particularly necessary when resistance is achieved to the highly potent inhibitor DRV, which fits well within the substrate envelope. However, the role of these changes outside the active site has long been thought to be only in recovering viral fitness, or protease stability.

In the current study, we have primarily chosen enzyme variants that are associated with DRV resistance: L76V, V32I, and L33F[186] Although L76V causes only a 1.5-fold decrease in DRV binding affinity (Table 2.1), this mutation is often observed in highly mutated DRV-resistant variants[154, 187], as well as variants with hyper-susceptibility to other PIs. L33F is a highly networked mutation co-occurring with many others in highly drug-resistant patient isolates, often together with V32I [188]. Therefore, comparison of V32I and the V32I/L33F double mutant permits the context-dependence of mutational effects in drug resistance. While not directly associated with DRV resistance, L90M is a canonical highly networked mutation that typically arises in multidrug resistant proteases. The large and rigid P1/P1′ moieties in NFV and SQV have been implicated in susceptibility to L90M, a feature lacking in DRV [180] L90M has

been found in more than half of patient isolates with at least one PI resistance substitution, and hence is often present in patients needing DRV-based salvage therapy [188]. Thus, elucidating the physical impact of these secondary mutations on DRV binding provides a detailed perspective on how the enzyme accommodates such frequently observed changes.

Specifically, we find that mutations outside the active site impact inhibitor binding thereby playing a direct role in conferring drug resistance. Compared to the WT complex, the overall structure and backbone conformations are very similar in the co-crystal structures of the variant complexes. However, the mutations cause subtle but significant rearrangements in the structure to cause altered interactions with the bound inhibitor, as well as impacting the dynamic ensembles of these complexes. We had previously hypothesized [157] and tested [158] that alterations in the hydrophobic core of the enzyme could alter the conformational dynamic ensemble through changes in the hydrophobic sliding of internal residues potentially impacting drug resistance. This impact on dynamics is not localized to the points of mutation but would propagate throughout the enzyme. In the present study we hypothesize these mutations outside the active site share a common pathway of altering the overall enzyme dynamics and propagating their effects to the active site.

Although the resistance-associated mutations are located at a variety of positions in the protease and away from the active site, they all may utilize a common mechanism or pathway of altering the protease–inhibitor interactions.

The mutations cause subtle changes through the repacking of the active site; in particular, these are observed in the crystal structures at residues 47, 50, and 84 in both monomers, and also observed in the MD analysis (Figure 2.9). Within the crystal structures of both the L76V and L90M complexes residue I47, which is located in the flap, loses contact with DRV. In contrast, in the V32I complex I47 contacts are retained, while this loss is restored when L33F occurs in V32I/L33F (these changes are also observed in subtle differences in the MD simulation Figure 2.9). Interestingly, V32I and I47V are the second most frequent pair of residues often found to coevolve, thus compensating for each other [188]. Mutations at I47, together with I54, which its backbone is hydrogen bonded with, is a major DRV resistance site. Among about 30 total active site residues that contact DRV, I47 is consistently the residue whose contacts are affected the most in L76V, L90M, and V32I/L33F variants (Figure 2.10). These results suggest that the interactions of residue 47 with inhibitors within the active site may represent a pivotal site in conferring drug resistance to PIs, and these interactions can be altered by changes propagated through the enzyme from remote sites.

In addition to repacking around the inhibitor in the crystal structures, the secondary mutations share a common pattern of altering the dynamic ensemble sampled by the protease, and the shape of the active site. Overall in the dynamic ensemble of the V32I and the V32I/L33F variants the active site is expanded, with the double mutant expanding the active site more, while L76V active site

contracts and the L90M active site displays asymmetric changes. Hence, even though not located at the active site, mutations at all these remote sites affect the shape of the active site in the dynamic ensemble.

How are single mutations at a remote site able to alter interactions and dynamics of the active site with highly common molecular mechanisms? We propose a “network hypothesis” where the perturbation introduced by mutation of a distal residue is propagated to the active site through a network of interactions within the protease structure (Figure 2.6). The distal mutation sites we studied are all part of a hydrogen-bonded network connecting to the active site where the inhibitor binds. Our network hypothesis postulates that the mutations have similar effects and common mechanisms as they all cause a rearrangement of this same network. This hypothesis is supported by the alterations observed during the MD simulations in the stability of the hydrogen bonding networks (Figure 2.6), where changes propagate from residues 74–78 and 87–90, through 28–33, to 84–85 and 25. This altered interaction network includes repacking of the vdW contacts with residues 47 and 54, which are pivotal in linking the networked residues to the rearrangement of the flaps, residues 29 and 30 that directly hydrogen bond to DRV, and 82 and 84 that are key sites within the active site cavity. We hypothesize that all these are the active site residues where the impact of distal mutations is propagated as a common mechanism of resistance in all variants and their subtle rearrangements can cause inhibitor specific resistant changes.

These common mechanisms provide an explanation for why some mutations are redundant and thus are not observed together in patient sequences, while others are synergistic and occur together to confer higher levels of drug resistance as they impact one another at pivotal sites that confer resistance often through expanding the active site. This hypothesis does not exclude the possibility that some changes may still provide additional stability, increasing the combined fitness of the variants. Most significantly, our findings show that all of the mutations we have studied, although outside the active site, still directly alter the shape and flexibility of the active site, thus likely play a direct role in conferring resistance.

## 2.5 Methods

### 2.5.1 Protease Gene Construction

Each of the four protease mutants was constructed using a standard site-directed mutagenesis protocol on a WT-SF2 protease gene with a codon sequence optimized for *E. coli* expression. The WT PR gene contained the amino acid substitution Q7K to minimize the enzyme's autoproteolytic activity.

### 2.5.2 Protease Expression and Purification

Each PR mutant was expressed and purified as previously described[189]. Briefly, the mutant HIV-1 protease gene was cloned into the pXC-35 plasmid, which was then transformed into the TAP56 strain of *Escherichia coli*. Transformed cells were grown in 6 × 1 L cultures from which cell pellets were

harvested 3 h after induction. The cell pellets were lysed and the protease was retrieved from inclusion bodies with 100% glacial acetic acid. The protease was separated from higher molecular weight proteins by size-exclusion chromatography on a Sephadex G-75 column. The purified protein was refolded by rapid dilution into a 10-fold volume of 0.05 M sodium acetate buffer at pH 5.5, containing 10% glycerol, 5% ethylene glycol, and 5 mM dithiothreitol (refolding buffer). The protease solution was concentrated, followed by dialysis to remove any remaining acetic acid. Protease used for crystallization was further purified with a Pharmacia Superdex 75 fast-performance liquid chromatography column equilibrated with refolding buffer.

### 2.5.3 Protease Crystallization

Crystals were set up with a 5-fold molar excess of inhibitor to protease, which ensures ubiquitous binding. The final protein concentration ranged from 0.8 to 1.6 mg/mL in refolding buffer. The hanging-drop method was used for crystallization as previously described[189]. For the L76V, L90M, and V32I mutants, the reservoir solution consisted of 126 mM phosphate buffer at pH 6.2, 63 mM sodium citrate, and ammonium sulfate at a range of 24–29%. For the V32I/L33F double mutant, the reservoir solution consisted of 0.1 M citrate-phosphate buffer, 7% DMSO, and 25–30% ammonium sulfate.

### 2.5.4 Enzyme Kinetics

Enzyme inhibition studies were carried out using a PerkinElmer Envision multilabel plate reader. A substrate peptide mimicking the MA-CA (p17-p24) cleavage site labeled with K-E(EDANS)-S-Q-N-Y-P-I-V-Q-K(DABCYL)-R (0.5 μM, final concentration) was added just prior to the reading to each well containing 50 nM of PR and varying concentrations of inhibitor. The FRET pair (EDANS, the donor and DABCYL, the quencher) was attached to the indicated amino acids of the peptide (Molecular Probes). Fluorescence intensity increase upon hydrolysis of the fluorogenic substrate was monitored at 490 nm (emission of EDANS) from the highest inhibitor concentration to the lowest, as well as the no inhibitor control well. Each inhibitor titration included at least 12 inhibitor concentration points. Initial velocities were obtained from the progress curves and plotted against inhibitor concentration to get inhibition curves. Resulting curves were globally fitted to Morrison's equation to obtain the $K_i$ value, as described previously[190].

### 2.5.5 Evaluation of Hydrogen Bonding and van der Waals interactions

The Maestro component of the Schrödinger software suite was used to analyze the hydrogen bonds between the inhibitor and the protease residues and neighboring waters after optimization of the complex structure. Briefly, a hydrogen bond was defined by a distance between donor and acceptor of <3.5 Å and a donor-hydrogen acceptor angle of >120°. The vdW contacts between the

inhibitor and protease were calculated using a simplified Lennard–Jones potential, following previously published protocols[191].

### 2.5.6 MD Simulations

The MD simulations were performed using the program Sander in the Amber 8 package, as previously described[167]. A set of three simulations was run for each of the four mutants and the WT-PR yielding a total of 15 trajectories for analysis. Each simulation was assigned initial velocities according to the Maxwellian distribution and random seeds were assigned with five different values for each PR. An in-house script was used to determine the intra and intermonomeric Cα distances between various residues using the trajectories. To calculate the hydrogen bond duration between various residues within the network from the simulations, the Visual Molecular Dynamics (VMD) version 1.9.1 was used[192, 193]. VMD was used to write out the trajectory in a pdb format using the coordinates and trajectory files generated by PTraj from the AMBER simulation software. VMD was also used to generate the trajectory pdb files to determine the vdW contact energies over the simulations. The in-house vdW script was then modified to assess vdW contacts from the simulations. The script was run to determine vdW contacts for each of the trajectories. Once the robustness of the system was assessed, the three trajectories for each system were concatenated into one file containing 1500 frames and the total vdW contacts were analyzed [166].

# Chapter III

# Elucidating the Interdependence of Drug Resistance from Combinations of Mutations

## 3.1 Abstract

The HIV-1 protease is responsible for the cleavage of 12 non-homologous sites within the gag and gag-pro-pol polyproteins in the viral genome. Under the selective pressure of inhibition, the protease is able to develop mutations within (primary) and outside (secondary) of the active site allowing the protein to facilitate substrate processing while simultaneously countering inhibition. The primary protease mutations impede inhibitor binding, while the secondary mutations are considered accessory mutations that compensate for enzymatic fitness loss. However, the role of secondary mutations in conferring drug resistance remains a largely unresolved topic. We have shown previously that while darunavir is superior to preceding protease inhibitors in viral suppression and has a higher barrier to resistance, mutations distal to the active site are able to perturb darunavir binding by disruption of the protein's internal hydrogen-bonding network [169]. In this study we show that mutations distal to the active site can interdependently play a role in darunavir resistance although they are set in complex mutational backgrounds. Using a combination of statistical and machine learning techniques, we identify individual and paired residue positions within the protease that are key in perturbing the dynamic ensemble of the protein. These findings reveal that primary mutations are not solely responsible for driving resistance. Furthermore, the techniques used in this study are applicable to larger and more diverse drug resistant protein variants. Identifying

which variable positions have the largest impact on drug resistance may be useful in the future of structure based drug design.

## 3.2 Introduction

The HIV-1 protease inhibitor (PI) darunavir (DRV) is highly potent against its target. Unlike first generation PIs, DRV is able to withstand many mutations both within and outside of the proteins' active site [135, 194]. Given its high barrier to resistance, single mutations do not pose a threat to DRV targeting. Underlying substitutions responsible for cross-resistance to other PIs are still fairly susceptible to DRV inhibition [195]. Contributing factors to DRV's high genetic barrier to resistance include the tight binding affinity (Kd = 4.5 $x10^{-12}$ M [182]), extensive hydrogen bonding with several active site backbone atoms [196], hydrophobic contacts within the active site and a good fit within the substrate envelope [183]. However, even with all these key attributes the protease is still able to develop complex interdependent mutational patterns to evade DRV inhibition.

In such complex multiple mutational patterns, while active site mutations physically alter inhibitor binding and are readily identified, the role of mutations beyond the active site is not completely clear. The widely accepted notion is that accessory mutations had the sole purpose of balancing the destabilizing effects of primary active site mutations [86]. However, accessory mutations' direct role in drug resistance has not been extensively probed [69, 148, 152, 197] and even less well known are which specific variable positions outside the active site play a

role in resistance [188]. Specifically, some mutations within the active site that arise to darunavir are readily explained by the substrate envelope hypothesis. For instance, DRV resistance associated mutations I84V, I50V, V32I and I47V all lay in positions where the inhibitor atoms protrude beyond the substrate envelope at the active site [134]. However, in clinical trials DRV selected for several non active site (secondary or accessory) mutations at positions 11, 33, 54, 73, 76, 85, and 89 among others, [186, 198] and the role of these mutations in DRV resistance is not clear.

Previously, to gauge how secondary mutations away from the protease active site, both associated with DRV resistance and not associated with DRV resistance, could play a role in protease inhibitor susceptibility, we examined several single mutations and one double mutation variant of HIV-1 protease. These mutations included V32I at the periphery of the active site, and a combination of V32I/L33F. In addition we examined the distal DRV resistance associated mutation (RAM) L76V and the non-DRV RAM L90M. A careful investigation of the crystal structures and molecular dynamics simulations of these variants bound to DRV showed that while these distal mutations alone do not drive significant levels of resistance, they were all able to perturb the network of hydrogen bonds within the protein, thereby propagating the effect to the protease active site causing slight loss of affinity. This network model provided a general understanding as to how mutation of residues may communicate with

one another and why some co-mutant relationships may be synergistic or redundant.

In this study we further investigate the role of mutations both at and away from the active site in complex combinatorial backgrounds of a set of protease variants, and compare to the WT and single/double site mutants examined previously. Specifically, this study seeks to determine which variable positions in the protease are responsible for the distinguishing patterns of resistance within the heavily mutated variants. Several DRV-resistant protease variants were selected from viral passaging experiments as well as patient-derived variants from the HIV Drug Resistance Database at Stanford University (hivdb.stanford.edu). Extensive (triplicate) 100 ns MD simulations were conducted on the panel of 15 protease variants (SF-2 and NL4-3 WTs and 13 variants) all bound to DRV. From the resultant MD trajectories, interactions within protease and with DRV were monitored by hydrogen bond occupancies, internal active site Cα-Cα distances and van der Waals energies. For each of the data sets a combination of statistical and machine-learning techniques was used to identify which residue positions contribute the most to the calculated properties and protease resistance. These analyses find that the hydrogen bonding patterns among the panel of protein variants distinguish single/double mutants from the more complex variants. Mutations in the hydrophobic core (A71V) and near the flap regions (R41K) also impact the hydrogen-bonding network. In addition to the hydrogen bonding, the primary mutation I84V in conjunction with the accessory

M46I and several other accessory mutations give rise to distinguishing patterns of interdependency in hydrophobic (vdW) contacts with DRV. Finally, we find via protein structure network analyses that, some highly mutated patient variants exhibit similar dynamic cross-correlations to both WT enzymes. This result implicates some accessory mutations as having a dual role, aiding in both resistance and the enzyme's ability to recapitulate more WT like enzymatic behavior in the presence of drug.

## 3.3 Results

To determine which variable positions specifically impact the structural and dynamic properties of protease-DRV binding, a combination of inhibitor bound crystal structures and homology models were used as input for molecular dynamics simulations. Using the resulting trajectories, hydrophobic contacts and internal Cα-Cα distances were calculated, and the hydrogen bond occupancies were monitored for the panel of protease variants.

### 3.3.1 Convergent evolution drives protease resistance in independent viral lineages

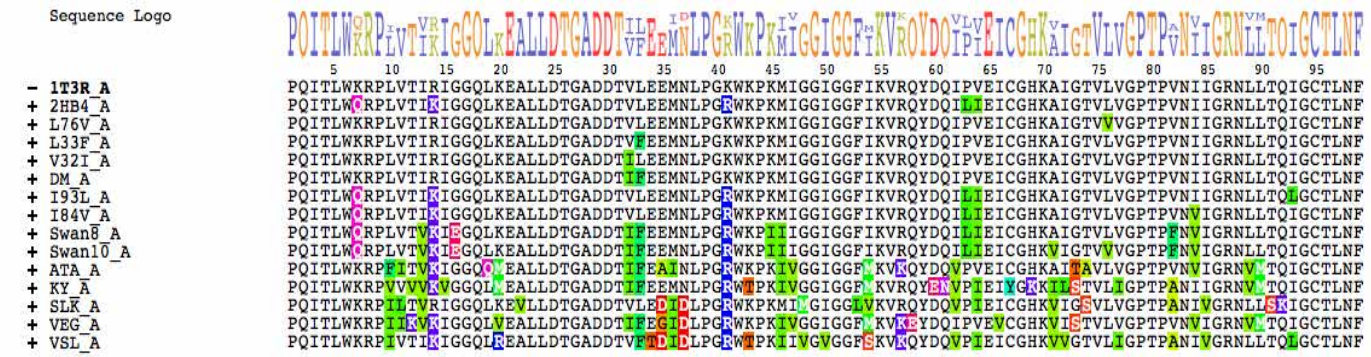
A diverse panel of fifteen HIV-1 protease variants were chosen with a broad range of sequence substitutions containing single site mutants and more heavily mutated multi-drug resistant proteases (MDR-PRs). Both the SF-2 and NL4-3 wild-type proteases were used as controls for the variants in the panel. The wild-type proteins share 95% identity, varying at positions 7, 14, 41, 63 and 64, and they both have high susceptibilities to DRV in single-digit picomolar

values (SF-2 with a $K_I$ of 2 pM [169], and NL4-3 with an $EC_{50}$ of 4 nM), (Table 3.1).

**Table 3.1 Ki, $IC_{50}$ and $EC_{50}$ values as determined by the Schiffer laboratory, Monogram Biosciences and the Swanstrom laboratory, respectively. Note that $IC_{50}$ value for NL4-3 WT protease is the median value for inhibition with DRV.**

| | Ki | | Monogram (IC50) | | Swanstrom EC50 | |
|---|---|---|---|---|---|---|
| | Raw | Fold Change | Raw | Fold Change | Raw | Fold Change |
| SF-2 | 2pM | - | | | | |
| NL4-3 | | | ~0.8nM | | 3.98nM | - |
| L76V | 3pM | 1.5 | - | - | - | - |
| L33F | - | - | - | - | - | - |
| V32I | 7pM | 3.5 | - | - | - | - |
| V32I/ L33F(DM) | 45pM | 22 | - | - | - | - |
| I84V | - | - | - | - | - | - |
| I93L | - | - | - | - | - | - |
| Swan8 | - | - | - | - | - | - |
| Swan10 | - | - | - | - | - | - |
| $ATA_{21}$ | - | - | | >200 | - | - |
| $KY_{26}$ | - | - | 89.6nM | 112 | 1.16M | 291 |
| $SLK_{19}$ | - | - | 19.2nM | 24 | 32.5nM | 8 |
| $VEG_{23}$ | - | - | | >200 | 7.8M | 1959 |
| $VSL_{23}$ | - | - | 31.2nM | 39 | 320nM | 80 |

Of the MDR proteases in the panel, two were obtained from long-term viral passaging experiments, conducted under DRV selective pressure by the Swantsrom laboratory at UNC-Chapel Hill. The remaining MDR proteases were obtained from the HIV Drug Resistance Database at Stanford




(hivdb.stanford.edu). The patient-derived MDR-PRs contain between 19 and 26 substitutions when compared to the SF-2 WT protease. Taken together, the panel of 15 proteases has sequence variations at a total of 50 of the 99 amino acid positions within each monomer (Figure 3.1).

**Figure 3.1 Proteases derived from multiple ancestors converge evolutionarily to drive resistance to DRV. Sequence alignment of all proteases included in the panel. Mutations are highlighted compared to the SF-2 WT protease (PDB code 1T3R. PDB code 2HB4 denotes NL4-3 WT).**

Although we cannot infer viral population ancestry and temporal treatment history for each of the patient derived proteases, they share common evolutionary changes with one another and also with those MDR-PRs derived from viral passaging experiments. Based on the years that the samples were isolated and the high resistance to DRV inhibition, there is a possibility that one

patient, $VEG_{23}$, may have been treated with DRV. Interestingly, the patient-derived $ATA_{21}$ variant is highly similar to the patient-derived $VEG_{23}$ and serially passaged Swan 8 variants, sharing sequence identities of 88% and 80% respectively (Figure 3.2).

These shared sequence identities within the panel emphasize the role of underlying cross-resistance to DRV among the patient isolates. Although we do not know whether variant $ATA_{21}$ was exposed to DRV, we do know that variant Swan8 was exposed exclusively to DRV. The level of resistance to DRV that these variants exhibit, along with their high percentage identity suggest that under the selective pressure of inhibition the phenotypes of these proteases have converged perhaps driving similar mechanisms of resistance. The relations between all sequences can be seen in the phylogenetic tree in Figure 3.3.

| | SF#2% | NL4#3% | L76V% | L33F% | V32I% | V32I_L33F% | I84V% | I93L% | Swan8% | Swan10% | ATA% | KY% | SLK% | VEG% | VSL% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SF#2% | 100# | 95# | 99# | 99# | 99# | 98# | 94# | 94# | 87# | 85# | 79# | 74# | 81# | 77# | 77# |
| NL4#3% | 95# | 100# | 94# | 94# | 94# | 93# | 99# | 99# | 92# | 90# | 78# | 75# | 80# | 76# | 78# |
| L76V% | 99# | 94# | 100# | 98# | 98# | 97# | 93# | 93# | 86# | 86# | 78# | 73# | 80# | 76# | 76# |
| L33F% | 99# | 94# | 98# | 100# | 98# | 99# | 93# | 93# | 88# | 86# | 80# | 75# | 80# | 78# | 78# |
| V32I% | 99# | 94# | 98# | 98# | 100# | 99# | 93# | 93# | 88# | 86# | 80# | 75# | 80# | 78# | 76# |
| V32I_L33F% | 98# | 93# | 97# | 99# | 99# | 100# | 92# | 92# | 89# | 87# | 81# | 76# | 79# | 79# | 77# |
| I84V% | 94# | 99# | 93# | 93# | 93# | 92# | 100# | 98# | 93# | 91# | 79# | 74# | 79# | 77# | 77# |
| I93L% | 94# | 99# | 93# | 93# | 93# | 92# | 98# | 100# | 91# | 89# | 77# | 74# | 79# | 75# | 79# |
| Swan8% | 87# | 92# | 86# | 88# | 88# | 89# | 93# | 91# | 100# | 98# | 80# | 76# | 75# | 78# | 75# |
| Swan10% | 85# | 90# | 86# | 86# | 86# | 87# | 91# | 89# | 98# | 100# | 78# | 75# | 75# | 78# | 75# |
| ATA% | 79# | 78# | 78# | 80# | 80# | 81# | 79# | 77# | 80# | 78# | 100# | 79# | 73# | 88# | 74# |
| KY% | 74# | 75# | 73# | 75# | 75# | 76# | 74# | 74# | 76# | 75# | 79# | 100# | 68# | 77# | 74# |
| SLK% | 81# | 80# | 80# | 80# | 80# | 79# | 79# | 79# | 75# | 75# | 73# | 68# | 100# | 73# | 80# |
| VEG% | 77# | 76# | 76# | 78# | 78# | 79# | 77# | 75# | 78# | 78# | 88# | 77# | 73# | 100# | 75# |
| VSL% | 77# | 78# | 76# | 78# | 76# | 77# | 77# | 79# | 75# | 75# | 74# | 74# | 80# | 75# | 100# |

**Figure 3.2 Sequence identity matrix for all 15 proteases in the panel. High % identity values are colored blue, low % identities are colored red.**

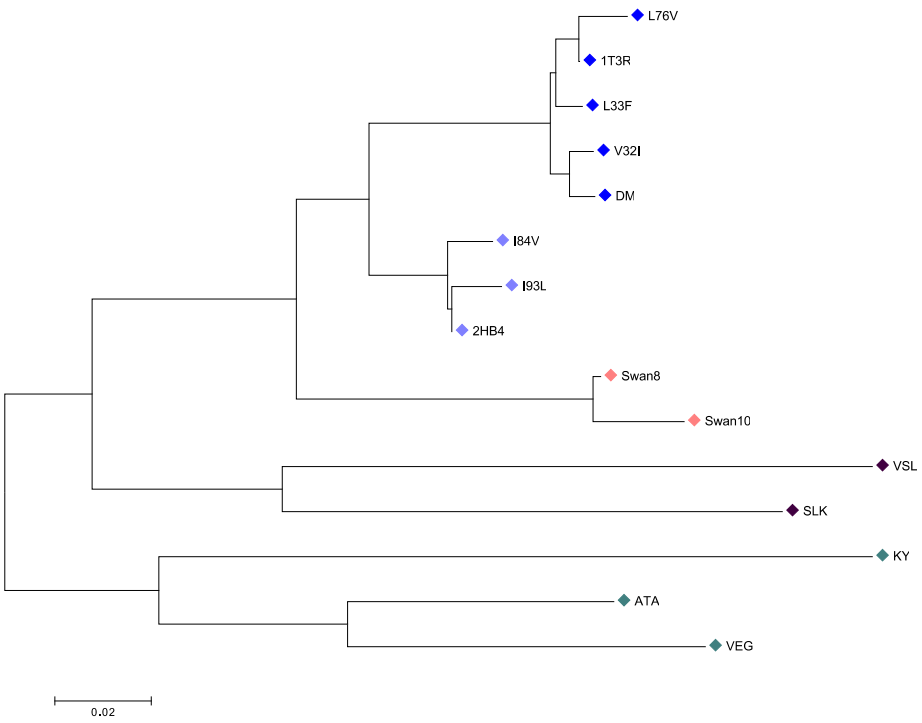



**Figure 3.3 Neighbor-joining phylogenetic tree of all 15 protease variants. Color annotations are similar between variants with high sequence similarity.**

### 3.3.2 Sequences of multi-drug resistant mutants are predictive of protease dynamics

The SF-2 WT (PDB: 1T3R), L76V, V32I and V32I/L33F were the variants with available crystal structures. The NL4-3 WT and remaining variants were modeled based on the DRV bound wild-type 1T3R structure. The crystallographic water molecules, including the important bridge water between the inhibitor and the protease flaps, were preserved in each model. Three replicates of fully

hydrated 100 ns MD simulations of each DRV complex were performed and analyzed.

The root mean square deviations (RMSD) reveal that the accumulation of mutations from single site to patient-derived variants progressively alters the dynamics of the proteins. When compared to the SF-2 WT, the L76V, $ATA_{21}$, $KY_{26}$ and $VEG_{23}$ variants deviate the most from their respective starting structures, whereas each of the variants derived from viral passaging experiments only deviate slightly over the duration of the simulation when compared to the NL4-3 WT (Figure 3.4).

The changes in Cα root mean square fluctuations (RMSF) about the mean increase from single to patient-derived variants with the most pronounced changes in fluctuation seen in the flaps, flap elbow, and lower loop regions of the L76V, I84V, $ATA_{21}$ and $VEG_{23}$ proteins while the catalytic aspartate residues remain rigid across the simulations (Figure 3.5). Interestingly, hierarchical clustering of per-residue RMSF for the fifteen variants results in groupings of the variants in a similar fashion as within the phylogenetic tree in Figure 3.3. This overlapping of clustering suggests that sequence similarity alone may be a predictor of similar backbone dynamics (Figure 3.6).

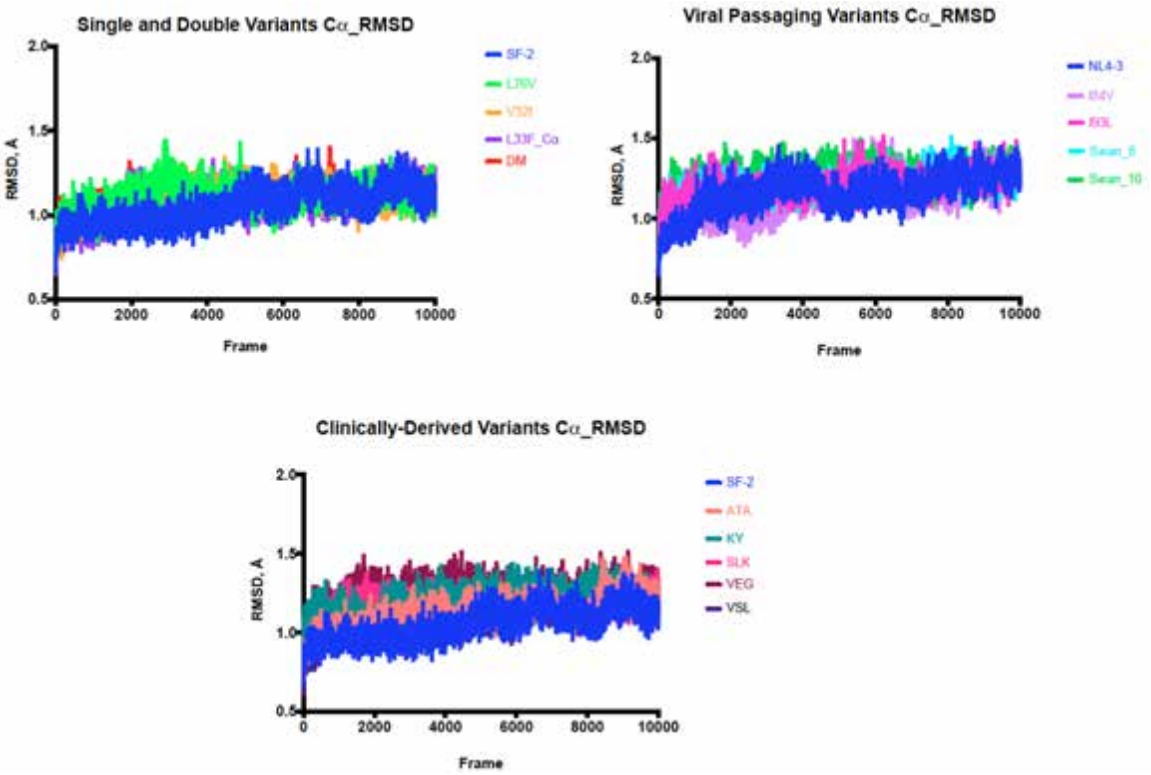



**Figure 3.4 Protease dynamics is impacted by high number of mutations. Top left: Single mutants L76V, V32I, L33F, and V32I+L33F as well as patient-derived variants are compared to the SF-2 WT(Bottom). Top right: Variants I84V, I93L, Swan8 and Swan10 are compared to the NL4-3 WT. For graphing purposes half of the total amount of frames incurred over the 100ns simulations are shown. "Frame" on the x-axis is interchangeable with "Time" as 10,000 frames is the equivalent of 50ns.**

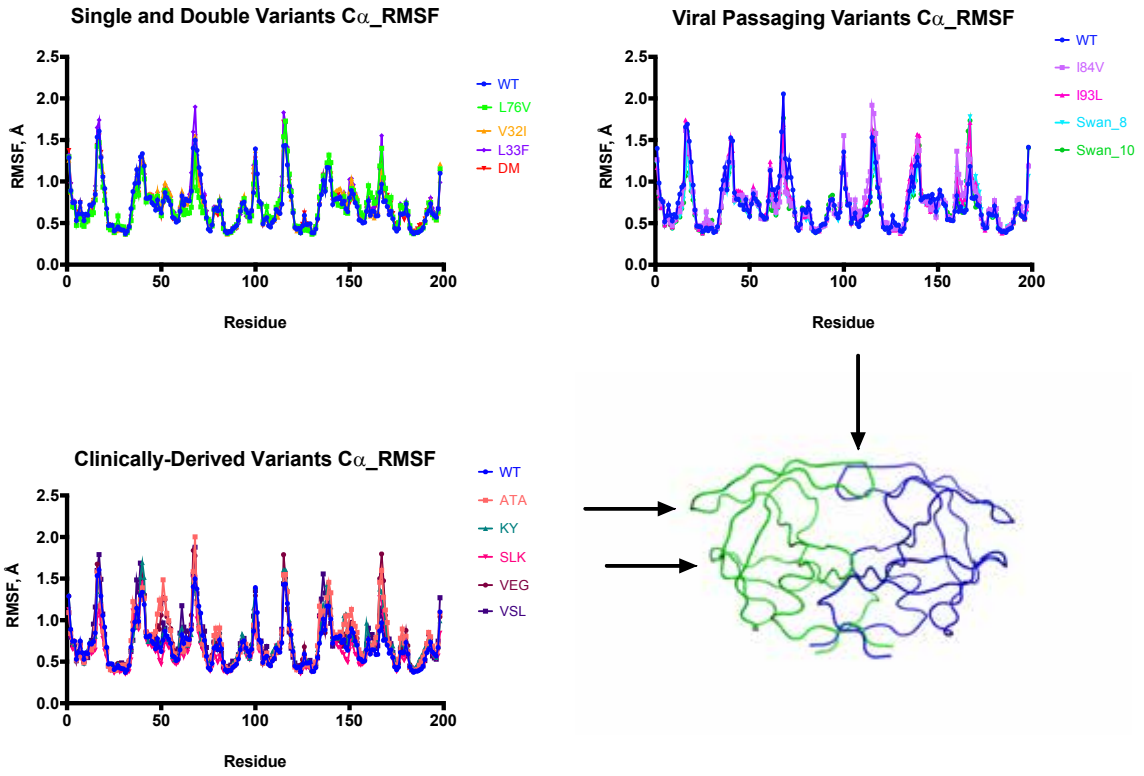



**Figure 3.5 Overlay of Cα RMSF with same groupings as in Figure 3.4, where arrows pointing to protease highlight areas of significant change.**

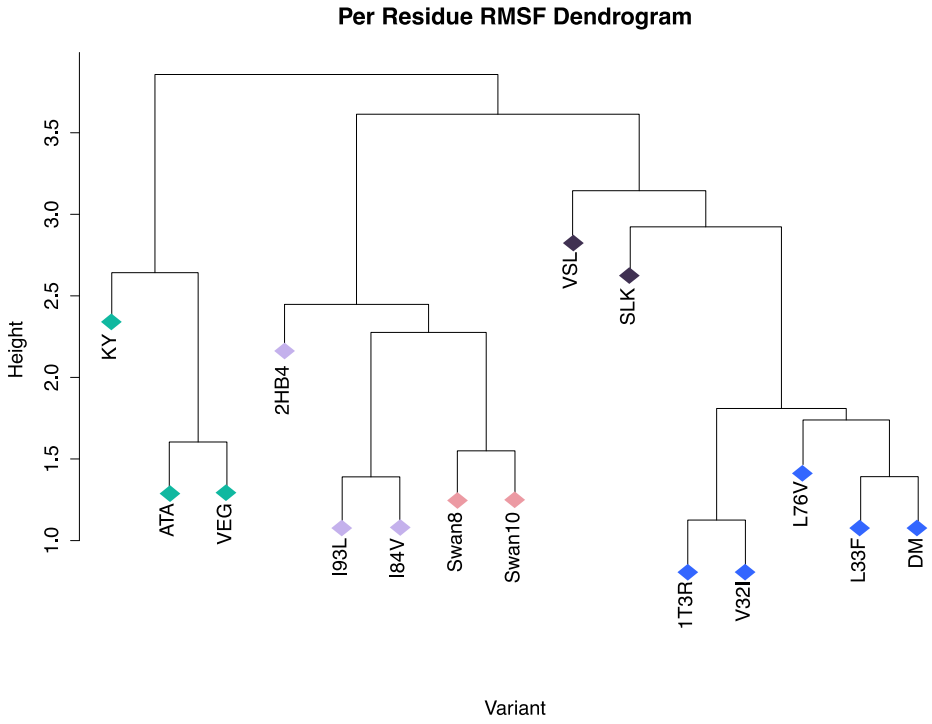



**Figure 3.6 Dendrogram of per-residue RMSF for the 15 variants in the panel. All heavy atoms are included in the calculation of per-residue RMSF but hydrogen atoms are excluded.**

### 3.3.3 Hydrogen bonds elicit changes in dynamics brought about by combinations of mutations

To characterize alterations in the hydrogen bonding observed among the different variants and to determine which variable positions best explain these alterations in hydrogen bond network among the proteases in the panel, a total of 143 hydrogen bonds were monitored over the course of the simulation. Included in these measurements were 111 main chain hydrogen bonds including those from protein to DRV and the 32 side chain hydrogen bonds formed both with DRV and with other residues. With an expanded panel of protease variants,

relative to our earlier studies, and such a long list of hydrogen bonds, the use of algorithms for detecting patterns of altered occupancies and identifying specific mutations that may underlie these alterations becomes essential. Accordingly, we employed a combination of principal component analysis (PCA), for helping to detect alterations, followed by hypothesis testing based upon amino acid substitution at specific sites in the protease. To begin, a 15x15 correlation matrix of mean occupancies for these 143 hydrogen bonds was computed and used for PCA (see methods for details). We note that 1) the first principal component accounts for nearly all (>85%, Figure 3.7) of the inter-variant variance in hydrogen bond occupancies and 2) the distribution of variants along this component is bimodal.

Using the 111 main chain hydrogen bonds, in the plot of the first two principal components, we note that the resistant variants tend to have higher values of the first principal component (denoted as $u_1$) than the more susceptible variants, including the two WT strains. In order to infer that substitutions at specific positions account for the distribution of variants (density, $\rho(u_1)$) along $u_1$ we conducted hypothesis tests (Wilcoxon rank sum) for pairs of distributions defined by the presence or absence of a mutation. The null hypothesis is no difference between the means of these two distributions. There is a lower bound on the p-value, defined by $\rho(u_1)$. The protease panel was partitioned into two groups such that their difference along the first principal component was maximal (low PC1 and high PC1) (Figure 3.6). Table 3.2 summarizes the results of these

tests, identifying A71V as the single mutation that best accounts for the spread in hydrogen bonding patterns among the variants in our panel. Segregating the variants based on this mutation, we find that Swan 10, $KY_{26}$, $SLK_{19}$, $VEG_{23}$ and $VSL_{23}$ all contain changes at position 71.

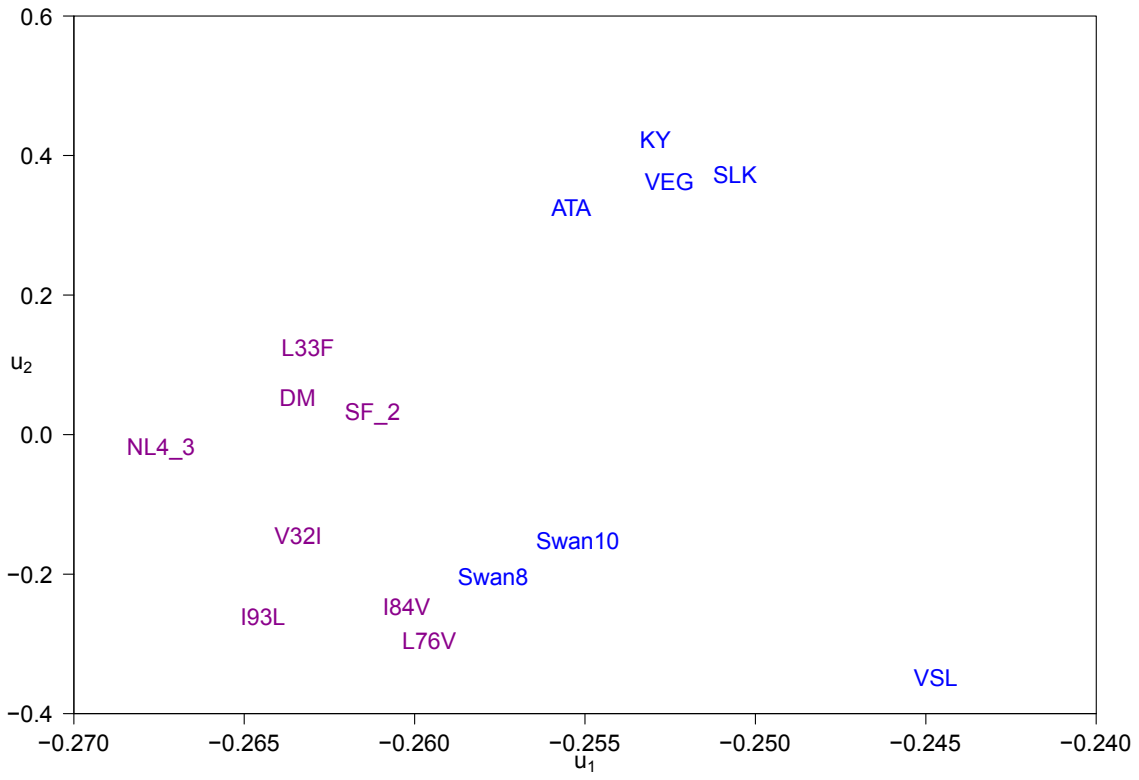



**Figure 3.6 Fifteen sequence variants projected on to the first two principal components for the correlation matric of 111 mean main chain hydrogen bond occupancies. Variants colored in blue have high PC1 values and variants colored in violet have low PC1 values.**

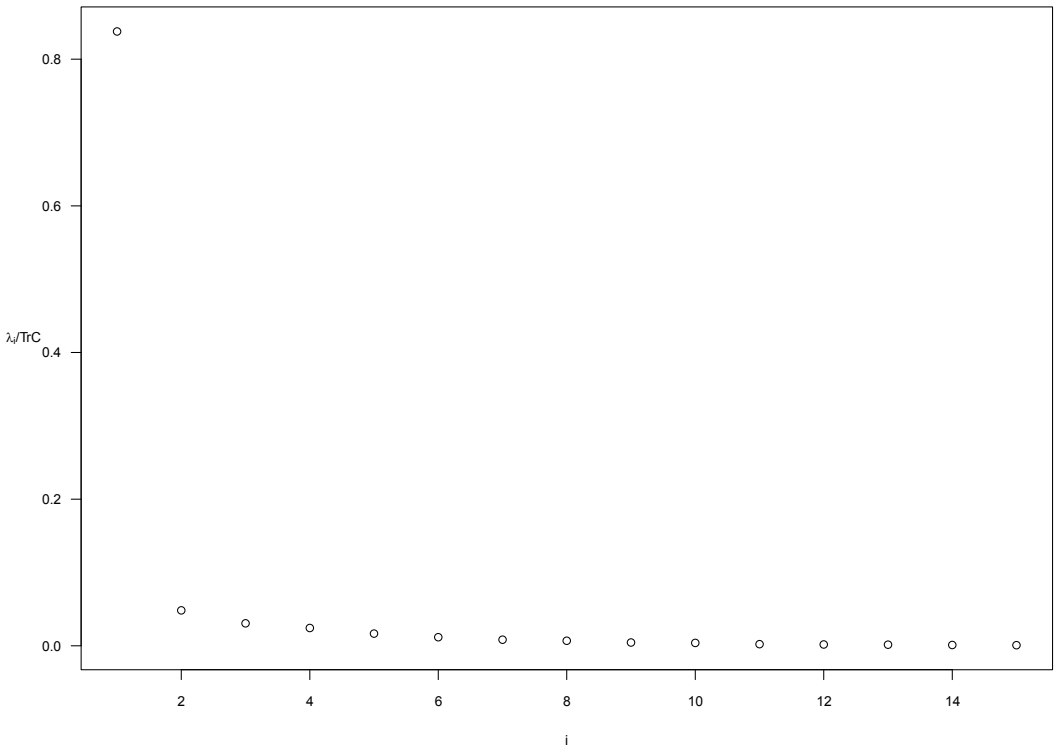



**Figure 3.7 The variance of hydrogen bond occupancies is accounted for by the first principal component (eigenvalue spectrum).**

**Table 3.2 List of top five positions that underlie changes seen in hydrogen bonding patterns.**

| Residue Position | P. Value |
|---|---|
| 71 | 0.000666 |
| 10 | 0.00133 |
| 54 | 0.00133 |
| 37 | 0.00439 |
| 35 | 0.00586 |

To investigate further, combination of substitutions that better recapitulate the segregation of variants into low and high PC1 were identified. We found no pair or other combination of substitutions that can better account for the observed distribution of variants along $u_1$ than does the A71V mutation alone (Table 3.3 and Figure 3.8).

**Table 3.3 List of top five position pairs found to most likely underlie changes seen in hydrogen bonding patterns.**

| Residue Position 1 | Residue Position 2 | P. Value |
|---|---|---|
| 41 | 71 | 0.000666 |
| 10 | 41 | 0.00133 |
| 10 | 54 | 0.00133 |
| 41 | 54 | 0.00133 |
| 10 | 71 | 0.00146 |

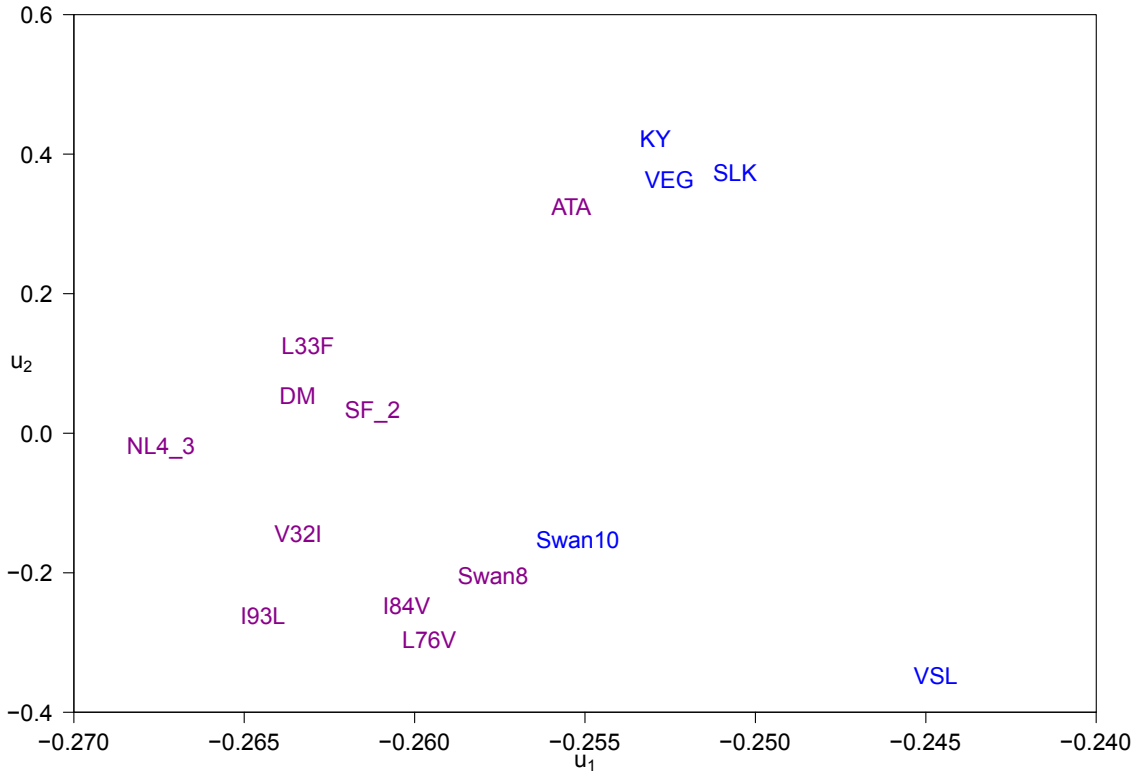



**Figure 3.8 Fifteen sequence variants projected as in Figure 3.6. Variants colored blue contain substitutions at residues 71 and 41 with variants $KY_{26}$, $SLK_{19}$, $VEG_{23}$, and $VSL_{23}$ also containing substitutions at positions 10 and 54.**

Moreover, the variance within this panel can be better explained by classifying the variants into those containing mutations at 10, 54, 71 and 41 simultaneously, and those that do not. Only the patient variants mentioned above contain this combination of mutations. While the importance of residue 41 initially seems miniscule, it is one of five residue changes that distinguish those variants in our panel in an NL4-3 background from those in the SF-2 background. This finding suggests that perhaps mutations at positions 71, 54, and 10 have a greater effect on the hydrogen-bonding network of those variants in the NL4-3

background relative to those variants with an SF-2 background in this panel. Also, because mutations 10, 54, 71 and 41 are very distal to the active site, these additional mutations may perhaps be relaying information about the global dynamics of the protein consistent with our original hypothesis that mutations perturb the dynamic ensemble of the protease via the network of hydrogen bonds. The findings here correlate well with the raw hydrogen bond occupancy data, as the majority of variants in the panel contain very pronounced changes about the high 60s/lower 70s beta strand-loop-beta strand region (Figure 3.9).

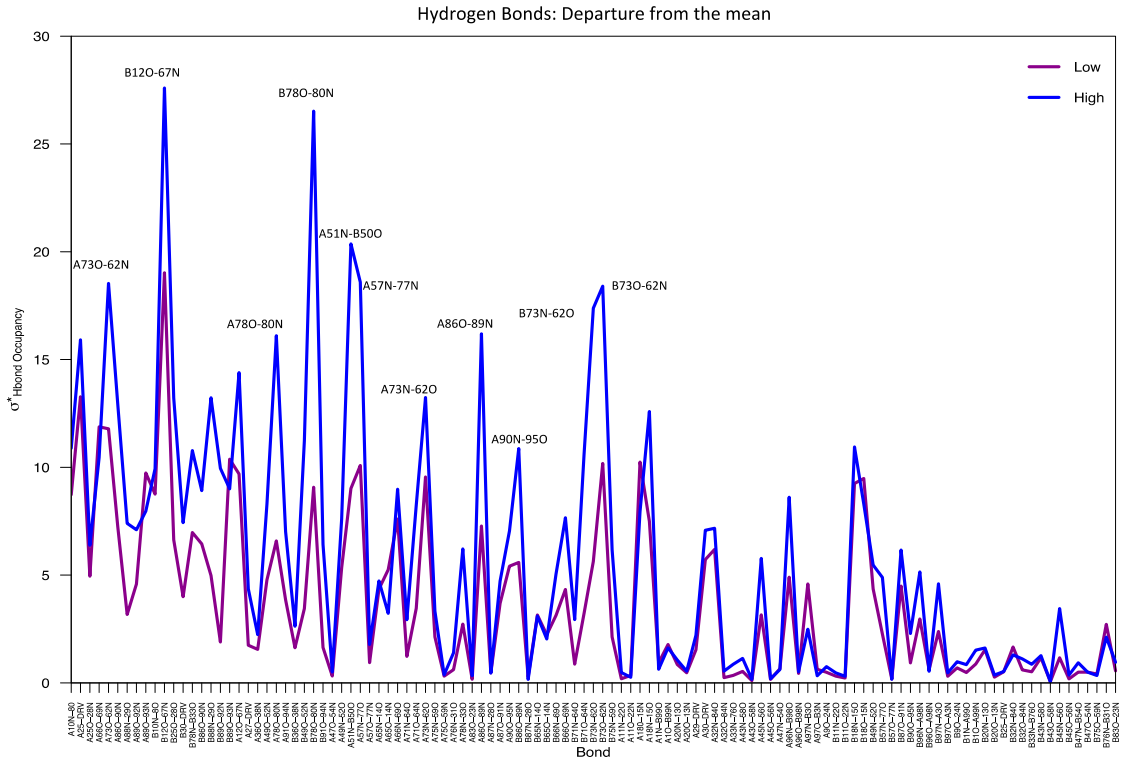



**Figure 3.9 Departure from the mean calculated from variants as segregated in Figure 3.6.**

## 3.4 The role of non-active site mutations elucidated via van der Waals contacts

Using the aforementioned methods, amino acid substitutions that played key roles in altering the hydrophobic contacts of the protease with DRV were determined. The mean van der Waals (vdW) contact energies between the protease active site residues and DRV were calculated over the trajectory for each of the variants in the panel. Energies were collected for 64 amino acids within the protease active site during simulation. Inspecting the distribution of the variants along the first principal component we find the variants segregate differently than the internal hydrogen bonding. Interestingly, the variants $KY_{26}$ and $SLK_{19}$ segregate with both the WT proteases and some single mutants close to this group as well, while the I84V and L76V segregate with the more complex variants (Figure 3.11).

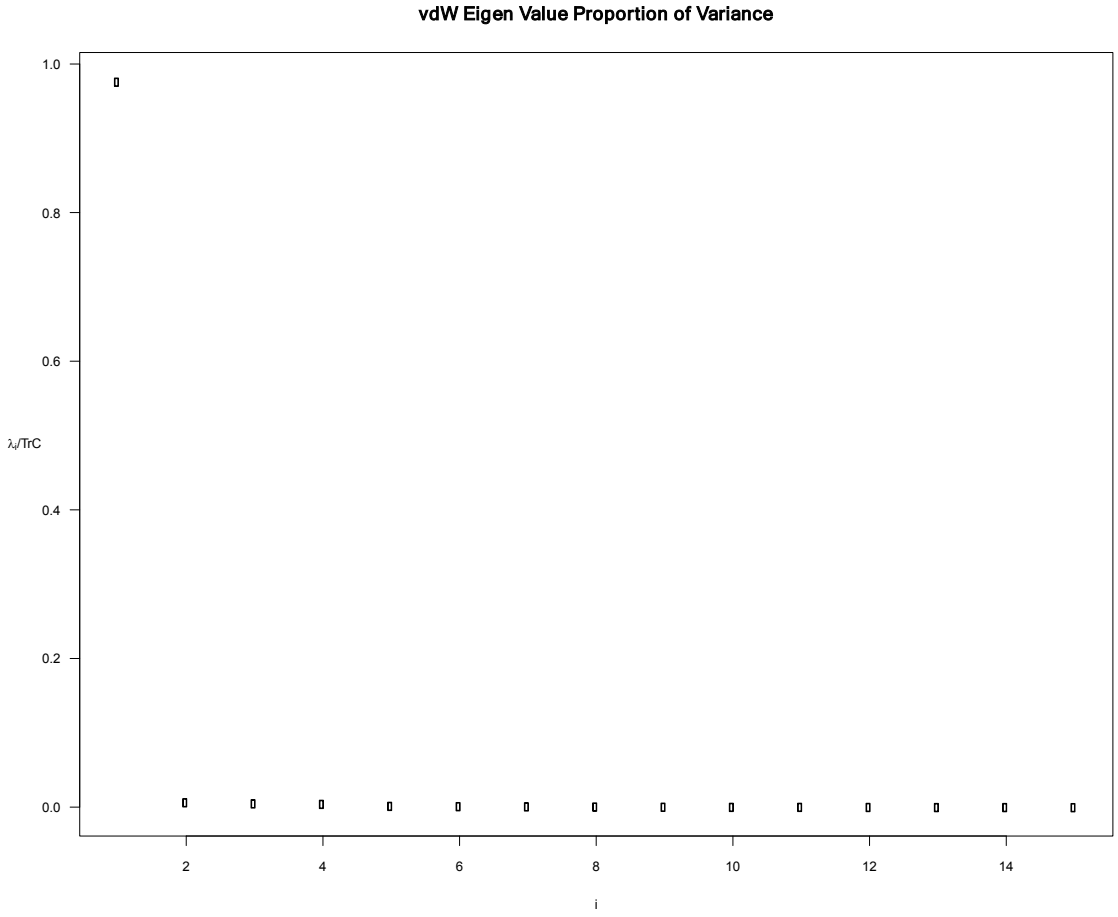



**Figure 3.10 Eigenvalues (proportion of variance) for the fifteen protease variants.**

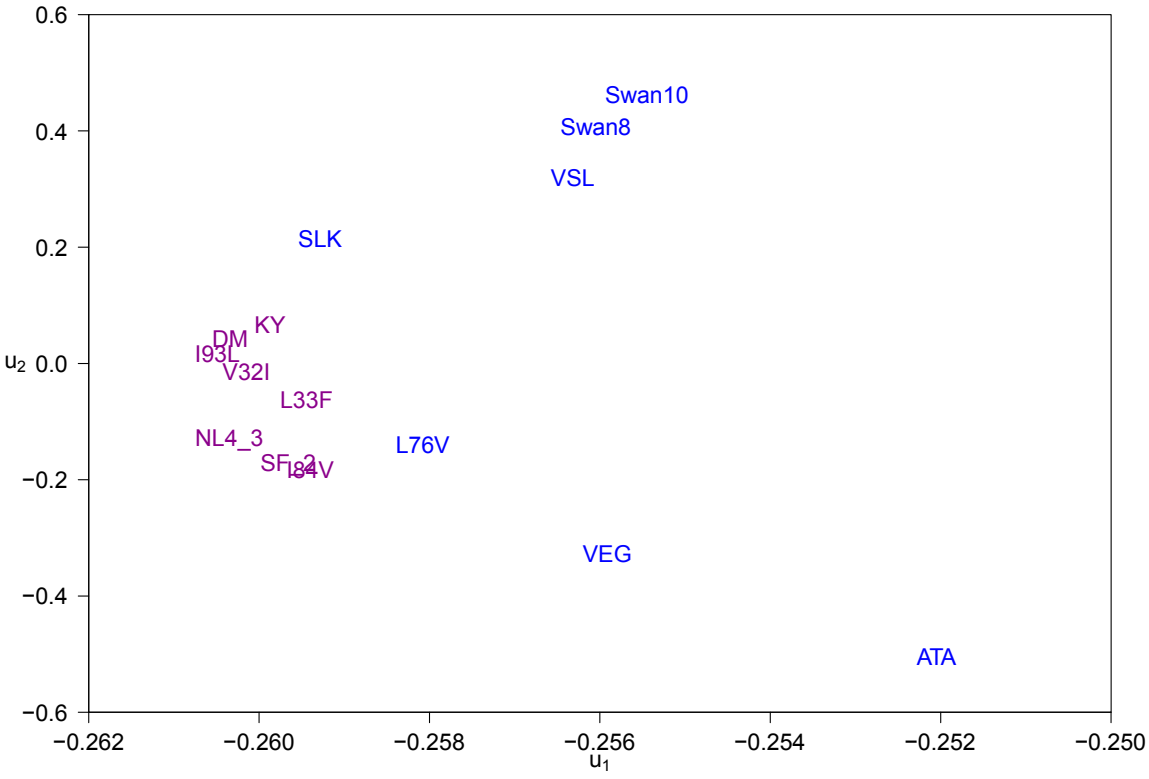



**Figure 3.11 Protease variants projected onto top two principal components and partitioned such that their differences along the first principal component are maximal. Variants colored in violet have low PC1 values and those colored blue have high PC1 values.**

Following the positional scanning and hypothesis testing approach that we used to examine the hydrogen bond occupancies we determine that I84V is the most predictive single site substitution for classifying the variants (Table 3.4). Segregating the variants based solely on the presence of I84V captures most of the base partitioning, such that variants $ATA_{21}$, $VEG_{23}$, $KY_{26}$, Swan8, Swan10 and, necessarily the I84V single mutant, are distinguished from the remainder of

the protease panel (Figure 3.13). Among the candidate pairs of mutational sites, the pair of residues that is most predictive of the perturbations of the vdW contact energies is I84V and M46I (Table 3.5). Segregating the variants based on this combination of mutants, $ATA_{21}$, $VEG_{23}$, Swan8 and Swan10 can be distinguished from the rest of the panel (Figure 3.14).

**Table 3.4 List of top five residue positions that most likely to explain variance in the vdW data.**

| Residue Position | P. Value |
|---|---|
| 84 | 0.00466 |
| 46 | 0.00759 |
| 13 | 0.0175 |
| 57 | 0.0307 |
| 35 | 0.0395 |

**Table 3.5 List of top five residue position pairs found to most likely explain the variance in the vdW data.**

| Residue Position 1 | Residue Position 2 | P. Value |
|---|---|---|
| 13 | 84 | 0.00146 |
| 32 | 84 | 0.00146 |
| 33 | 84 | 0.00146 |
| 46 | 84 | 0.00146 |
| 14 | 84 | 0.00466 |

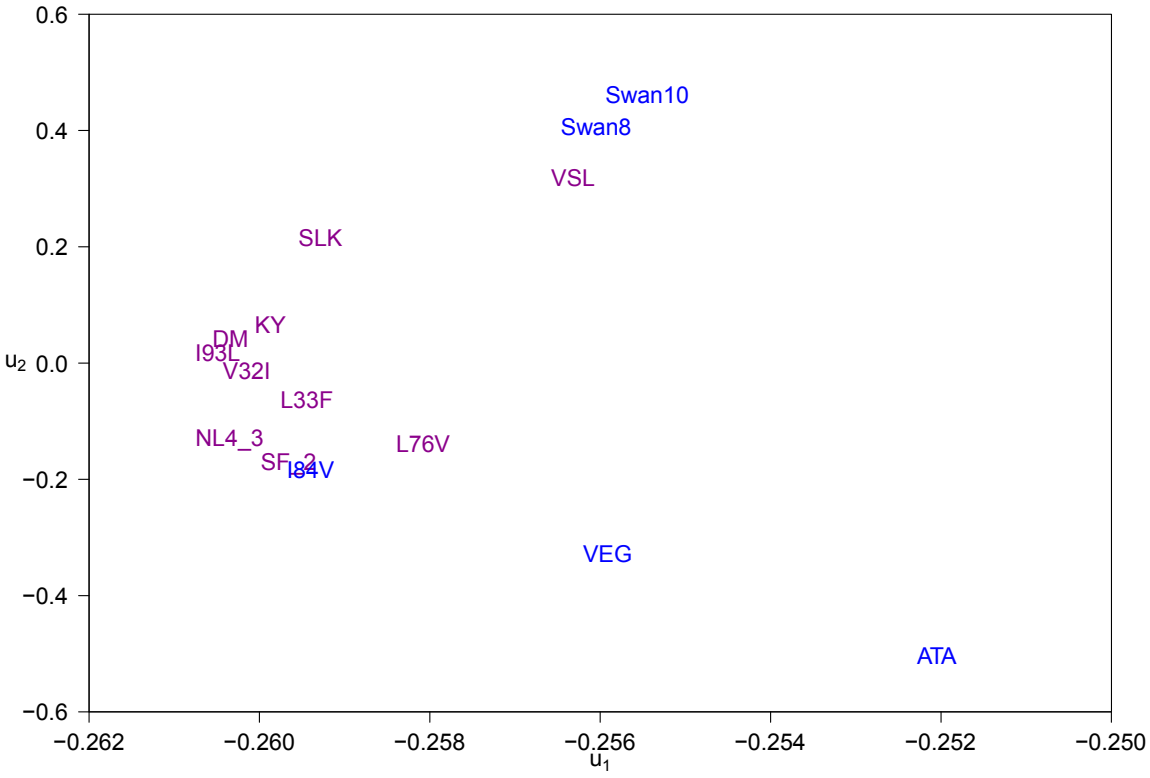



**Figure 3.12 Protease variants projected as in Figure 3.12 and partitioned based on the presence (blue) or absence (violet) of I84V.**

Furthermore, combinations of substitutions at positions 13, 32, and 33 combined with I84V were predictive of the distinguishing patterns within the vdW data. All four of the variants containing a change at positions 84 and 46 also contain mutations I13V, V32I and L33F. This finding suggests that once mutations are developed at I84 and M46 the subsequent changes may become

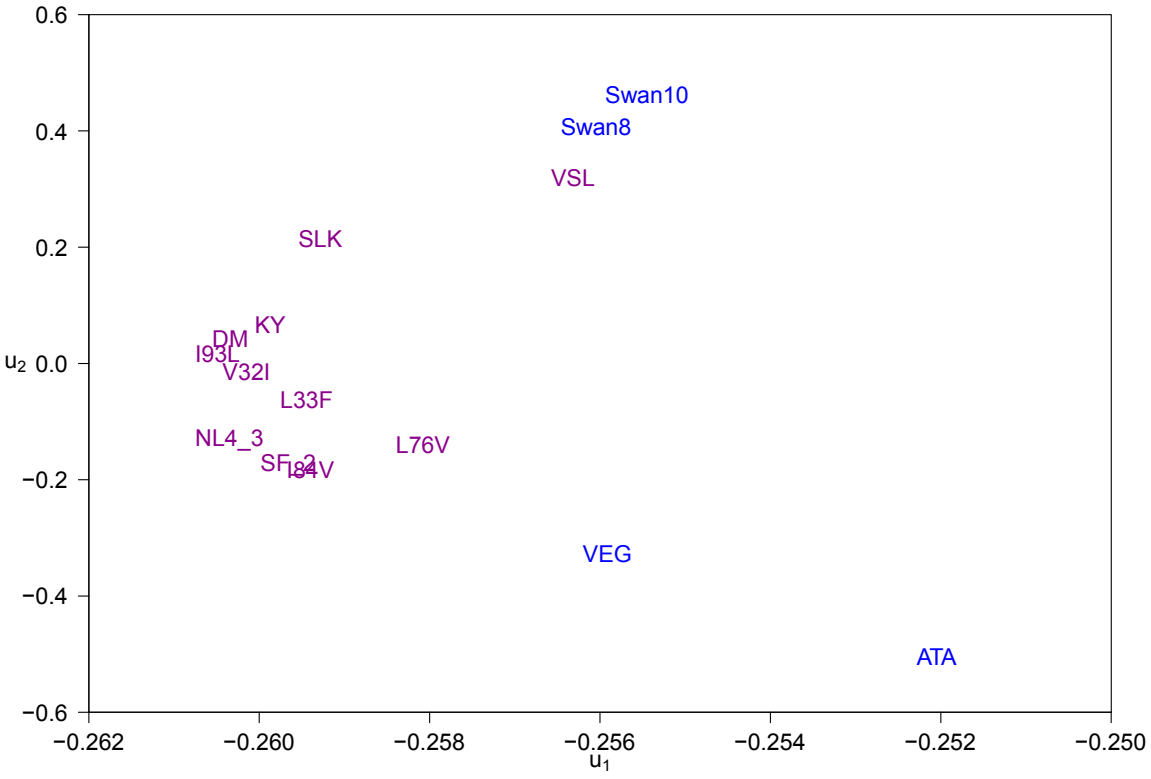



**Figure 3.13 Protease variants as in 3.12 and partitioned based on the presence (blue) or absence (violet) of I84V and M46I.**

obligatory in the obstruction of protease-DRV binding. Among the 64 residues that make hydrophobic contacts with the protease, residues within or adjacent to the active site, specifically D30, I50 and I84 in chain A and R8, D29, D30, G27, G48, V82 and I84 in chain B, are perturbed the most due to the presence of accessory mutations, as indicated by the departure from the mean for all residues across the 15 proteases (Figure 3.14, see methods).

That the distinguishing variations in hydrophobic contacts mostly impact generally immutable active site residues suggests that the role of preserving

hydrophobic contacts upon inhibitor binding is crucial for sustained targeting (Figure 3.14). Mapping the difference in departure from the mean values onto the structure further details how active site residues are impacted by mutations (Figure 3.15). The findings from the vdW data set are in good agreement with DRV binding in general, in that DRV employs hydrogen bonding and hydrophobic contacts to maximize inhibition. Figure 3.16 inset displays the exemplary relationship between I84' and its proximity to the P1' phenylalanine mimicking moiety in DRV. The change from isoleucine to valine would reduce the close packing of the isoleucine side chain and the inhibitor.

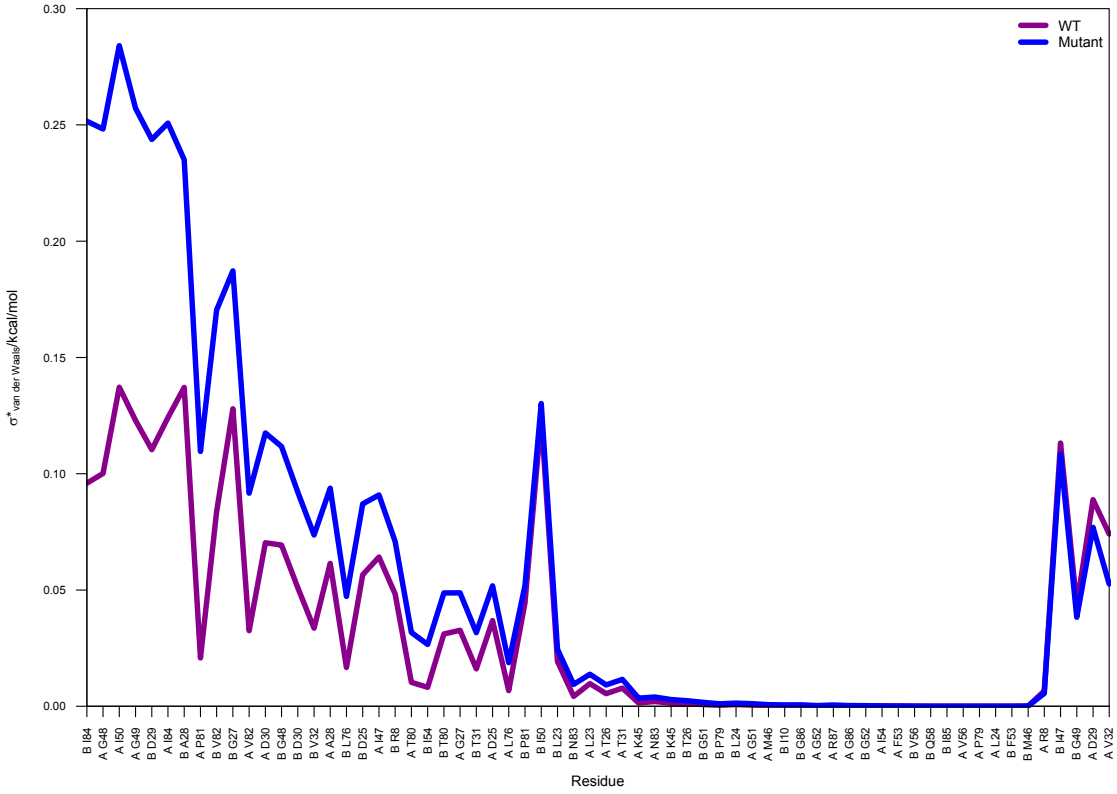



**Figure 3.14 Departure from the mean across all 15 variants for enzyme-inhibitor van der Waals energies. Colors are preserved from initial partitioning in Figure 3.14.**

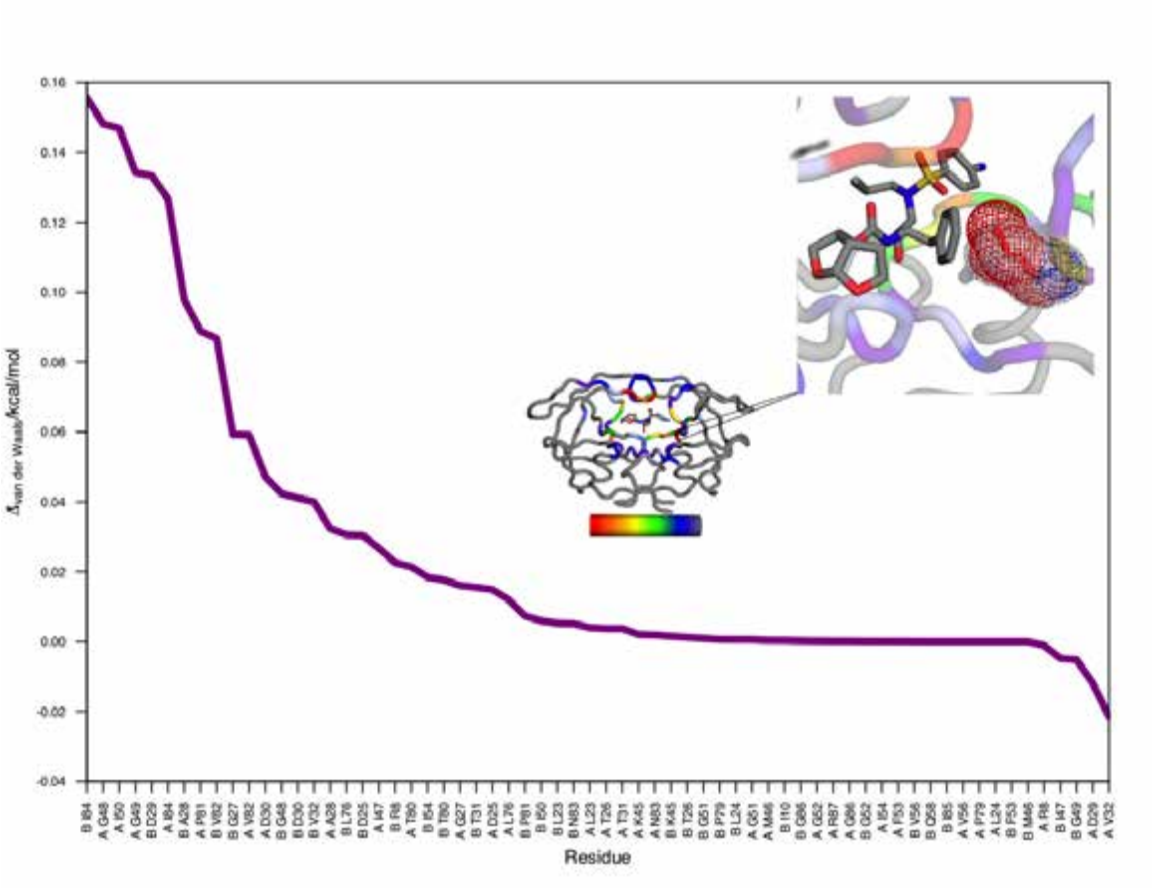



**Figure 3.15 Difference in departure from the mean mapped onto protease structure, colored from highest variability (red) to no variability (gray). Inset; view of I84' (red mesh) versus V84' (gray mesh) packing with DRV.**

## 3.5 Dynamic cross correlations may link changes in van der Waals contacts to clinical measures of resistance

In the vdW analyses above, the variants $KY_{26}$ and $SLK_{19}$ closely group with the single and double site mutants along PC1, which is in contrast to the other patient-derived variants. To examine whether the global dynamic network of these variants are also more similar to wild type protease, the community structures were evaluated along with the underlying dynamical cross correlation

networks and edge betweenness for all 15 variants in the panel. The trajectory frames from the MD simulations of the variants were used to conduct the dynamic protein structure network analysis using the Bio3D package for R [199]. Comparison of the SF-2 and NL4-3 wild-type proteases cross correlation matrices, community structures, and betweenness centrality values reveal that the differences at five amino acid positions between the two, reflect in their dynamic networks have a slightly different make up (Figure 3.16).

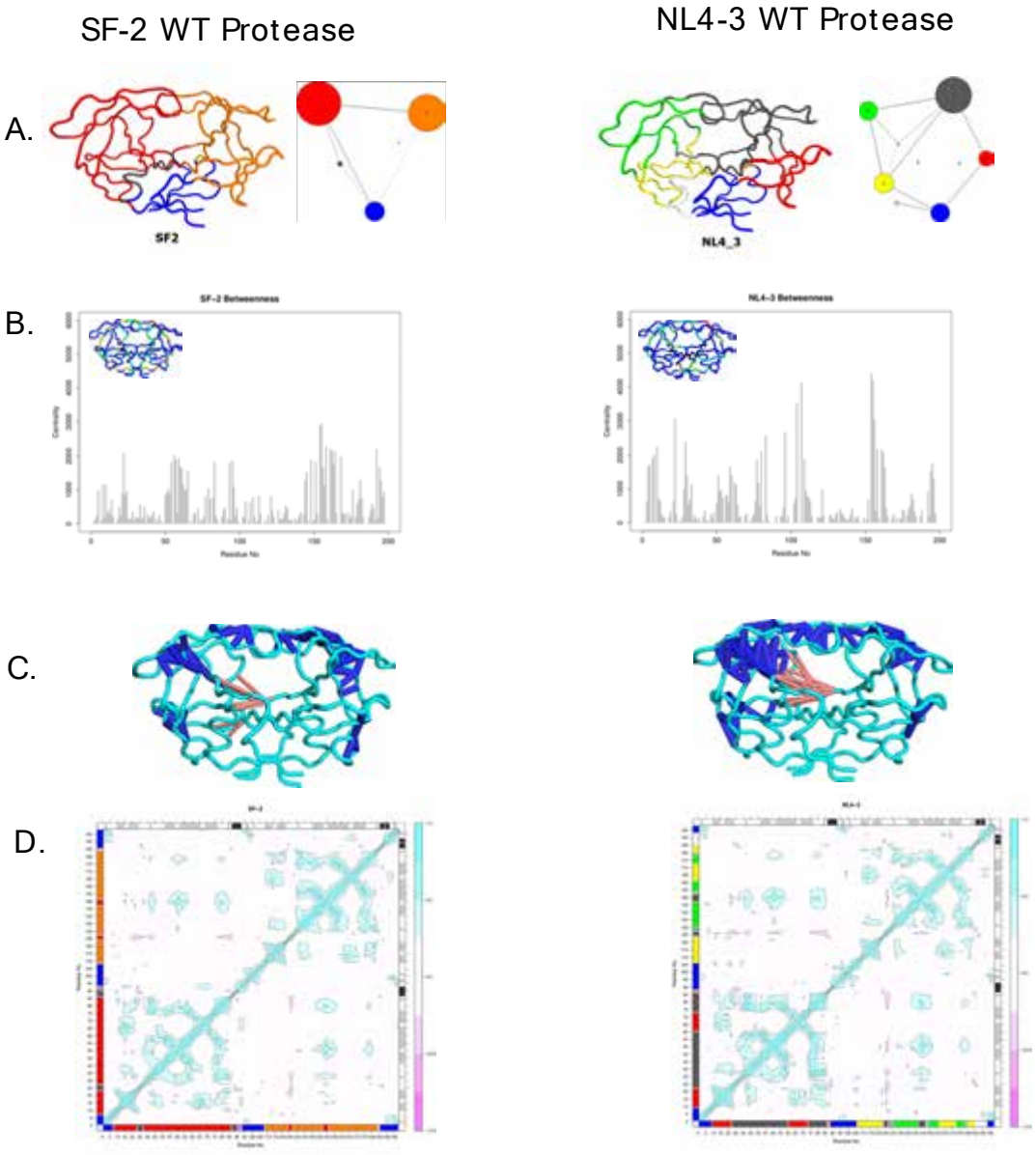



**Figure 3.16 Comparison of protein structure networks of SF-2 and NL4-3 WT proteases. A. Community structure networks shown in 3D mapped onto the protease (left) and simplified in 2D (right). B. Histograms of betweenness centrality values, with values mapped onto the protease structure shown in the inset of the graph. C. Dynamic cross-correlation network values mapped onto the protease structure. Positive correlations are blue and negative correlations are red. D. Dynamic cross correlation matrices, with the same coloring as in C. Annotations on the left side bar and bottom side bar are representative of community structures seen in A.**

### 3.5.1 Community Structures

One way to interpret data contained in a cross correlation matrix is to look at community structures. Overall, the residue communities that lie within each protease align with the lineages of origin (Figure 3.17). The exception to this trend comes along with the more highly mutated passaged and patient derived variants. While the ancestral or temporal treatment information of the patient-derived variants are unknown, Swan 8 and Swan 10 are from an NL4-3 background and resolve to have a community structure similar to those variants from the SF-2 lineage. $ATA_{21}$, $KY_{26}$, and $VEG_{23}$, which align closely in sequence, contain community structure networks similar to variants of the SF-2 lineage. Contrarily, variant $SLK_{19}$ has a similar community structure network to the NL4-3 wild-type along with I93L. A notable secession from this seemingly ancestral drift are the community structures of I84V and $VSL_{23,}$ which take on very distinct community structures but are more aligned with the NL4-3 structured variants (Figure 3.18). The community analysis provides insight into the protease's resiliency by displaying how, in the presence of combinations of mutations, the protease is able to alter its phenotypic composition to sustain resistance. Interestingly, the highly mutated patient variants $KY_{26}$ and $SLK_{19}$, which partitioned with the wild-type proteases in the vdW analysis, display community structures that are overall more similar to wild type than the other patient-derived variants.

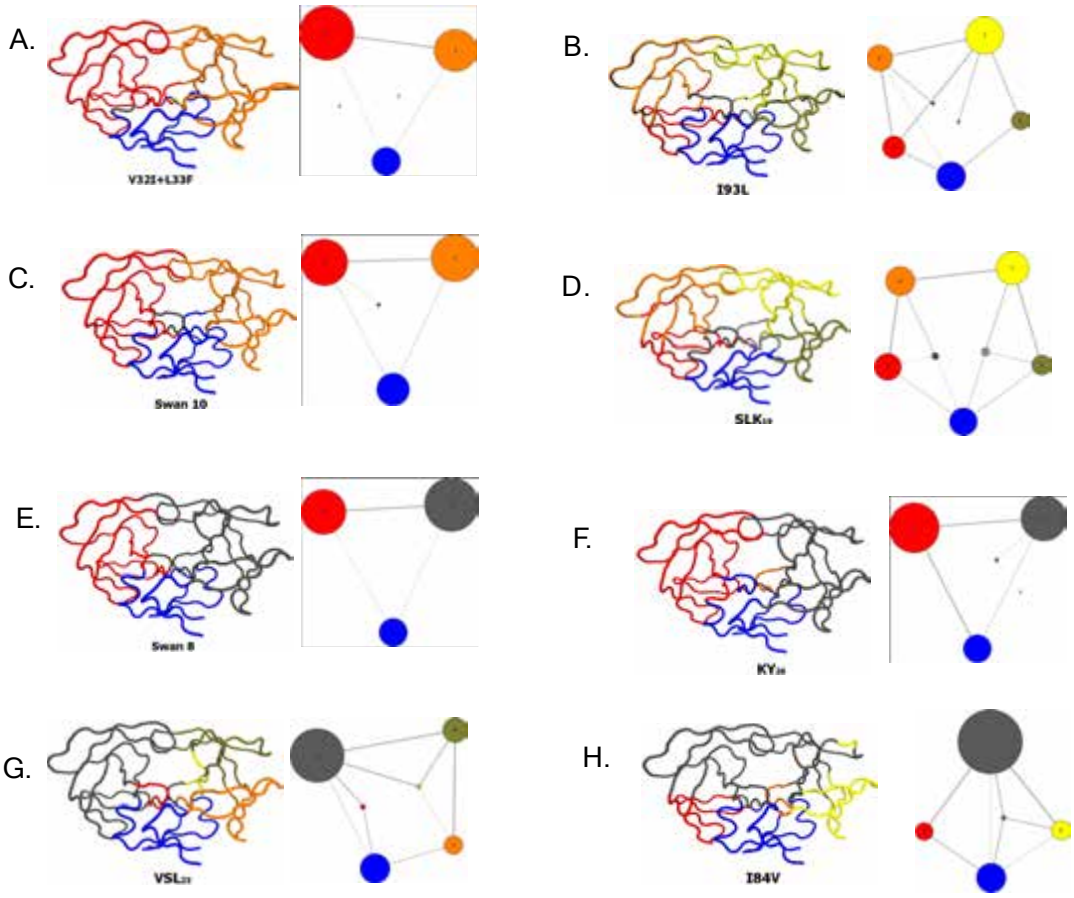



**Figure 3.17 Example community structure networks mapped on to the protease. A-H; Community structure networks for proteases V32I_L33F (DM), I93L, Swan 10, SLK19, Swan 8, KY26, VSL23 and I84V respectively. Both A and C have highly similar community networks to the SF-2 protease while B and D have networks similar to the NL4-3 protease. E and F have similar community networks as the SF-2 protease but there are some residues that have switched from the orange communities seen in A and C in to the gray community. G and H have the most highly dissimilar community structures compared to the two WT proteases.**

### 3.5.2 Girvan-Newman Edge Betweenness

In the community structures mentioned above, each residue is taken as a node or a vertex. Within and between communities these nodes or groups of

nodes respectively are connected through edges. Examining the connectivity of these edges within and between communities gives rise to betweenness centrality. Betweenness centrality is an indicator of a node's centrality (or importance) in a network [200]. Those nodes with higher betweeness centrality can be deemed as hubs essential for communication from one community to another. The inspection of betweenness centrality by residue plots or "betweenness spectra" shows that there is nearly an even split between those variants that make use of residue 83 as a hub and those that do not. This use of residue 83 is interesting because N83 is conserved and lies in the middle of the 80s loop, between two highly variable residues, V82 and I84. Hence, residue 83 may be important for communication to this region of the active site. There are many commonalities across the 15 proteases with respect to betweenness centrality, especially the use of residues 5, 8, 10 and 22 in one chain as hubs. The use of these residues is modified slightly in variants $KY_{26}$ and $SLK_{19}$ which make use of residues 5, 9, and 23 as hubs (Figure 3.19). The significance of this observation will be discussed below.

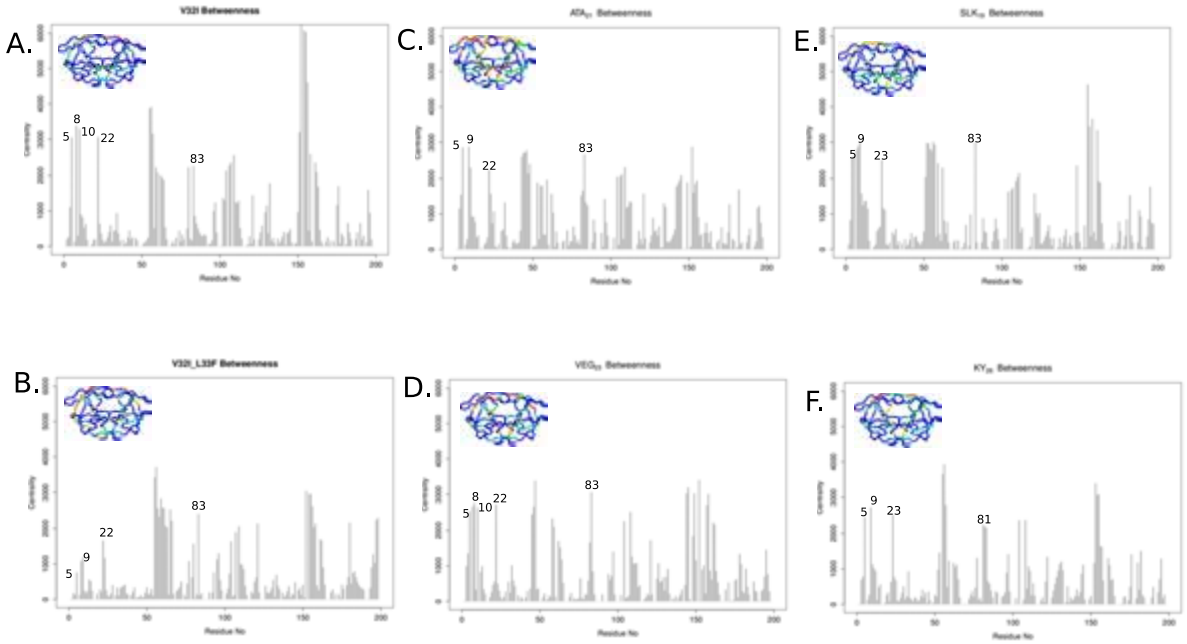



**Figure 3.18 Betweenness spectra. A-F; Histograms of betweenness centrality values for variants V32I, V32I_L33F(DM), ATA21, VEG23, SLK19, and KY26 respectively. The majority of variants in the panel use residues 5, 8, 10 and 22 for communication between the termini and the active site region of the protease while SLK19 and KY26 do not. Inset for all figures in the panel are the betweenness centrality values mapped on to the protease structure.**

### 3.5.3 Dynamic Cross Correlation Matrices

The dynamic cross correlation matrices are the basis of the network analyses presented above and direct examination of them can yield additional insight. Within these matrices, residues that move in the same direction have positive cross correlations and those that move in opposite directions have negative cross correlations. When the movement of two residues is not

correlated, the cross correlation value is zero. Domains of contiguous residues, such as in an alpha helix or beta strand, would give rise to significant positive correlations emanating from the diagonal of the matrix [201, 202]. Visual inspection of the matrices and corresponding mapping of cross correlation coefficients onto protease crystal structures shows drastic changes between certain variants (Figure 3.19). For example, variant L33F loses both positive and negative inter- and intra-monomeric cross correlation peaks while variant $VSL_{23}$ gains both positive and negative cross correlations relative to wild-type protease. There is also some overlap between Cα-Cα distances of certain residues and the corresponding increase or decrease in signal seen in the cross correlation matrices (Figure 3.19). The trend of either losing key positive cross correlation peaks or gaining negative cross correlation peaks can be seen throughout the variants except in patient variants $KY_{26}$ and $SLK_{19}$. Both $KY_{26}$ and $SLK_{19}$ are able to retain key positive inter-monomeric cross correlation peaks and seemingly exchange negative inter-monomeric cross correlation peaks in the active site. This suggests that $KY_{26}$ and $SLK_{19}$ contain combinations of mutations that are capable of rendering the enzyme resistant to the inhibitor but able to retain wild-type-like conformational dynamics and function in the presence of drug (Figure 3.20).

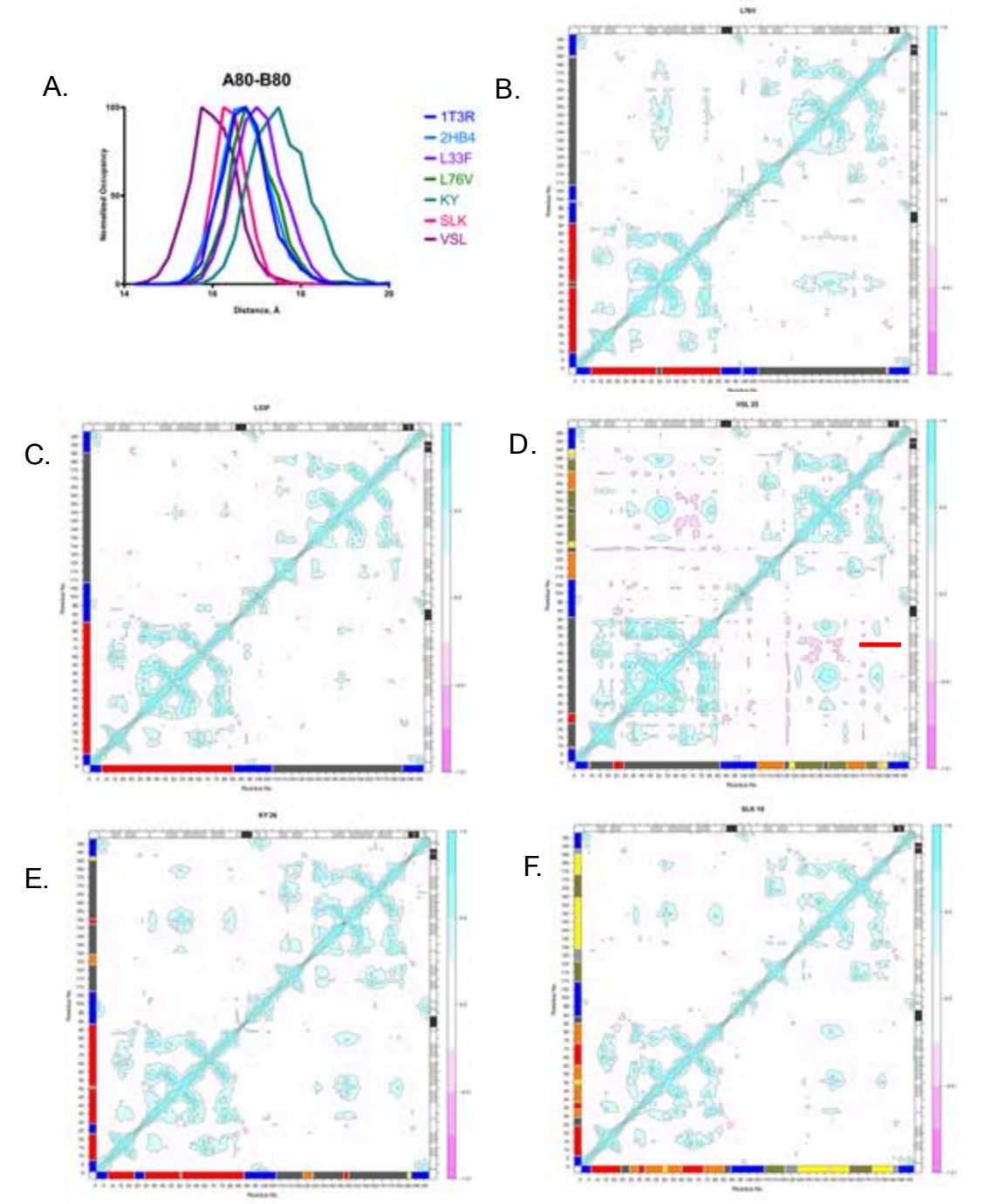



**Figure 3.19 Atomic distances between residues 80 and 80' and dynamic cross correlation matrices. B-F; Dynamic cross correlation matrices for variants L76V, L33F, VSL23, SLK19, and KY26 respectively. Regions containing peaks for correlations between residues 80 and 80' are underlined in red. In these regions peaks for residues 80 and 80' get smaller as the interatomic distances move further apart and larger as they get closer together as seen in A.**

## 3.6 Discussion

In this study, the effects of mutations on protease structure and dynamics have been investigated via hydrogen bonding, hydrophobic contacts with the inhibitor, and the overall networking within the protein. Specific residues have been identified that account for the variance in these properties, comparing multi-drug resistant variants to that of the lesser or non-mutated counterparts. While the active site mutations impede inhibitor binding in an intuitive manner, the constellation of mutations that arise throughout the rest of the protein, as is the case with DRV, have long been thought to aid in recovery of viral fitness only. To better understand the role of non-active site mutations in drug resistance, 15 protease variants have been chosen, which taken together contain substitutions at 50 of the 99 amino acid positions within the enzyme.

The accumulation of mutations within a drug target allows the balance of substrate processing versus inhibitor binding to tip in favor of the former. The HIV-1 protease is a very small protein but exemplifies a high level of resiliency under selective pressure. While the robustness of a mutated enzyme may seem less comparable with its more replication competent non-mutant counterparts, drug resistant protease variants are able to adapt and perpetuate viral escape and growth just as well given their new amino acid composition and environment.

We have shown previously that mutations compromise the hydrogen bonding and van der Waals contacts necessary for DRV binding to ensure that the mutations render the protein resistant while retaining its biological function

[169]. In this study, we used a novel combination of molecular dynamics simulations and unsupervised machine learning to identify and characterize those positions that are key to DRV binding. The resulting positions are highly conceivable given their prior examinations by other laboratories. For instance, position 71 was found to be a major contributor to the variance of the hydrogen bonds. The A71V mutation has been shown previously to be a key mutation in the re-stabilization of the enzyme in the presence of major mutations such as I50L/V, and has also been shown to propagate its effects from its resident position in the cantilever region of the protease to the active site via the hydrogen bond network of the protein [139, 141, 149, 203-205]. Because the mutations that were found to be highly predictive of changes in the dynamic hydrogen bonding within the protein are all distal to the active site, this suggests that the impact of mutations on the hydrogen bonding network is spread throughout the protein probably affecting the overall dynamic ensemble of the protein, consistent with our previous findings [169].

Mutation I84V is a mainstay in protease inhibitor cross-resistance. The change from the bulkier beta-branched isoleucine to the smaller valine has been the Achilles heel of PI treatment since its first emergence against saquinavir (SQV) treatment [206]. M46I was thought to only be a compensatory mutation [130] until early MD studies found it to be a key modulator of flap dynamics [207]. The pair of I84V and M46I has been long studied as a major/minor co-mutant pair in the midst of other compensatory mutations able to drive resistance to

early PIs [130, 208]. The other mutations that were atop the list of likely variants underlying the variance seen in the van der Waals contacts all lie outside of the active site (with the exception of the peripheral V32I) and have been explored thoroughly for their compensatory effects [188]. Figure 3.16 illustrates that the residues with the highest variability in vdW contact energies lie within the active site (G48, I50, I84, G27' and I84'); however the other mutations with high variability are either juxtaposed or distal to the active site. The calculation of the departure from the mean (σ*) explicitly demonstrates that mutations outside of the active site are in fact able to perpetuate changes that occur at generally non-mutating positions within the active site necessary to perturb vdW packing crucial to DRV binding.

The dynamic cross correlation matrices and variations of these matrices (i.e. network community analyses and betweenness centrality) show that the presence of combinations of mutations influence residue communication pathways within the protease. Visual inspection of the cross correlation matrices clearly display the degradation of residue communication caused by mutations such as the single mutations L33F and L76V, or more complex variants such as $VSL_{23}$ . Close observation of all the cross correlation matrices reveals that the majority of the mutants in the panel have not achieved the appropriate combinations of mutations to render them having wild-type-like dynamics (likely biologically functional) and inhibitor resistant at the same time. For most of the variants, there is either an accumulation of increasing negative cross correlations

or a reduction of positive cross correlations or loss of both simultaneously. However, in the case of variants $KY_{26}$ and $SLK_{19}$ the majority of positive intra and inter-monomeric cross correlations are retained while negative cross correlations involving active site residues are converted into positive correlations unique only to these two enzymes, relaying information about possible mechanisms they employ to confer resistance.

Some indication of why variants $KY_{26}$ and $SLK_{19}$ group closely with the lesser mutated variants and the wild-type proteasess may lie in protein network analyses. Although these variants do not have similar community structures and the correlations between their respective betweenness centrality profiles are low, they do share unique patterning of both positive cross correlations (Figure 3.21) and edge betweenness at least in one monomer.

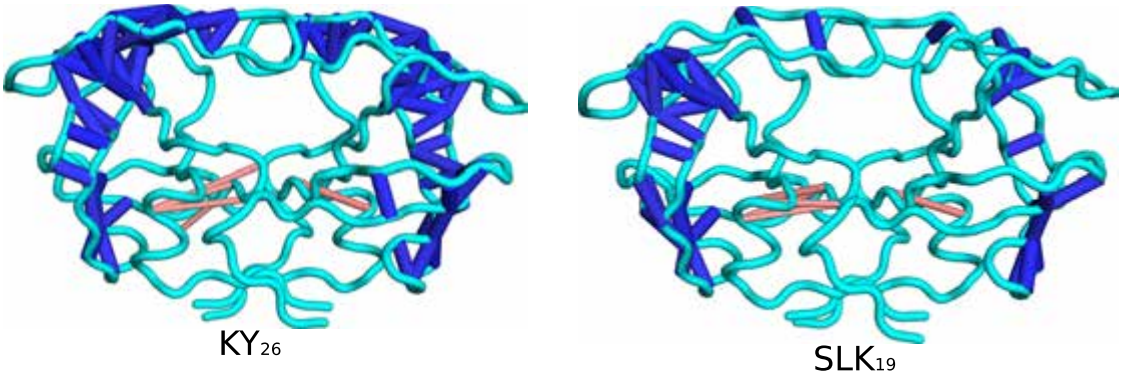



**Figure 3.20 Dynamic cross correlation peaks mapped on to structure of clinically derived variants $KY_{26}$ and $SLK_{19}$. Cylinder size is proportional to the cross correlation value. Blue cylinders denote most positively correlated (values 0.75-1) and red cyliners denote the most anti-correlated residues (values -0.5 - -0.4).**

This is evident in that $KY_{26}$ and $SLK_{19}$ make use of residues 5, 9 and 23 as hubs for communication in one monomer while all of the other proteases within the panel use a 5-X-10-22 pattern where X can be residue 7, 8, or 9. Perhaps the makeup of these variants has allowed them to use an alternative route for communication from the termini to the active site that does not involve residues that are prone to mutation and thus allows for inhibition evasion. In keeping with this hypothesis, both residues 10 and 22 are mutated in other variants within the panel; residue 10 also forms hydrogen bonds with residue 8, which could mean that a change at position 10 could be detrimentally propagated through to the active site. The use of residues 5, 9, and 23 appears to be more stable in that residue 9 hydrogen bonds to residue 24 and residue 23 hydrogen bonds both to the non-mutating residue 83 and residue 85 so that communication along this pathway is less likely to be disrupted. Interestingly, the variant $KY_{26}$ has the highest replication capacity among all the patient-derived variants, as assayed by Monogram Biosciences, and a medium-level resistance to DRV. In contrast, variant $SLK_{19}$ has the lowest replication capacity and the lowest level of resistance to DRV but its activity is enhanced in the presence of RNA [209]. Both of these variants contain WT NC-p1 and p1-p6 substrate sequences, which suggests that $KY_{26}$ has further accumulated the necessary mutations to cleave WT substrates and either $SLK_{19}$ has not yet achieved the necessary mutations either in the protease or in the substrate to render it fit in the absence of drug.

Taken together it appears that these two variants have different combinations of mutations that have allowed them to tip the balance of substrate turnover and inhibition in favor of the former by using similar mechanisms to drive resistance to DRV.

The analyses presented here show how mutations outside of the active site impact DRV targeting. Consistent with the findings of Appadurai et al. [153], we find here that mutations distal to the active site, in single positions or highly complex combinations, are able to perturb inhibitor binding through changes in certain key interactions between the enzyme and inhibitor. We also find that changes in overall protein dynamics may be indicative of resistance. Because the proportion of variance within each of the data sets examined can be largely accounted for by the first principal component, linear regression models can be used to probe further which non-active site mutations play a role in resistance. The individual methodologies used for the analyses are not novel individually; however, combining a diverse set of protease variants and using a combination of MD simulations, machine learning and statistics to characterize key interactions and infer how mutations in complex combinatorial backgrounds work together to drive resistance is unprecedented. These techniques could be used together to probe other panels of proteases and PIs or proteases and substrates as well as other protein systems where the interdependency of resistance is not well understood to tease out the impact of mutations both at and distal from the active site in complex mutational combinations.

## 3.7 Methods

### 3.7.1 Protease Panel & Nomenclature

The panel of 15 proteases used in this study consisted of the SF-2 and NL4-3 WT along with 13 mutant variants. The WT proteases served as controls for the variants with respect to their subtype B backgrounds and laboratory origin. The L76V, V32I and V32I/L33F (PDB accession codes 3OY4, 4Q1X and 4Q1Y respectively) variants were taken from the previous study [169]. To this group of single mutants, a L33F single mutant was added. Another set of protease sequences were obtained from HIV-1 cell culture passaging studies in the presence of DRV. These sequencing and passaging studies were carried out by Dr. Sook-Kyung Lee and Dr. Shuntai Zhou of the Swanstrom laboratory at UNC-Chapel Hill. Briefly, *in vitro* selections were carried out with DRV using an initial mixture of 26 variants, each containing a single resistance mutation. The selections were carried out with increasing inhibitor concentrations between 1000-10,000 fold over $IC_{50}$ measured in WT strains. Using the Primer ID-based paired-end MiSeq platform [210], mutations in the protease were analyzed at four time points at varying inhibitor concentrations. The variants in this set include models of single site I93L and I84V variants complexed with DRV along with a variant containing eight mutations (Swan 8) and a variant containing 10 mutations (Swan 10) in an NL4-3 background. The final group of variants in the panel was selected from a set of patient-derived proteases in the HIV-1 drug

resistance database. This group of proteins contains 19–26 mutations compared to the SF-2 WT protease.

Each patient variant is named based on which amino acid substitutions are unique to that variant and the number of mutations it contains compared to the SF-2 WT. For example, variant $KY_{26}$ is the only variant that contains substitutions H69K and C67Y and it has 26 mutations. All other variants are named for the mutations they contain (e.g. variant I84V only contains this mutations). The V32I+L33F double mutant is referred to as DM for double mutant. 1T3R and 2HB4 PDB accession codes are used to refer to the SF-2 and NL4-3 WT proteases respectively.

### 3.7.2 Homology Modeling & MD Simulations

The SF-2 WT, L76V, V32I and V32I/L33F had available crystal structures bound to DRV. The NL4-3 WT and remaining variants were all modeled based on the DRV bound 1T3R structure. The crystallographic waters, including the ever-important bridge water between the inhibitor and the protease flaps, were preserved in each model, as was the drug. Briefly, using the Prime Structure Prediction Wizard by Schrödinger (Release 2014-4, LLC, New York, NY [211, 212]), each of the 11 variant sequences was used as a query to search for homologs via BLAST [213]. Both chains of PDB structure 1T3R were selected as templates to build the homodimer containing the appropriate variant sequence. Once the model was prepared, the structure was built retaining the ligand from the template structure. Waters from the template structure were merged into the

newly built variant structure and a series of refinement and minimization steps were undertaken to complete the model building using the Prime Protein Prep Wizard in Maestro.

MD simulations were performed using the Desmond Molecular Dynamics System as previously described [138]. A set of three simulations totaling 300 ns was run for each of the thirteen variants and the two WT proteases for a total of 45 simulations. To calculate hydrogen bond occupancies, the Simulation Event Analysis Tool within Maestro was used to determine the occupancies of 143 inter and intra-main chain and side chain hydrogen bond pair as well as those bonds made with the ligand. In order to facilitate analysis, including computation of the mean protein-ligand van der Waals interaction energies, Visual Molecular Dynamics (VMD) version 1.9.2 [192] was used to translate the Desmond trajectories to PDB format.

### 3.7.3 Evaluation of Hydrogen Bonding and van der Waals Interactions

For each hydrogen bond pair, the donor heavy atom along with its hydrogen and the acceptor atom were specified for calculation of hydrogen bonding occupancy. For each frame, only pairs that satisfied the hydrogen bond geometric criteria as set forth by Schrödinger were chosen: the distance between hydrogen atom and acceptor atom must be at least 2.5 Å, the angle between donor heavy atom and its hydrogen and the acceptor must be at least 120°, and the angle between the hydrogen and acceptor heavy atom must be at least 90°.

The vdW contacts between the inhibitor and the protease were calculated using a simplified Lennard-Jones potential, following published protocols [191].

### 3.7.4 Principal Components Analysis and Statistical Testing

The hydrogen bond and van der Waals observations were combined into matrices of 15 x N, where N is the number of observations in each data set (fore example there were N=64 protein-ligand van der Waals energies per variant). In order to remove the variance between different factors (e.g. between highly favorable van der Waals interactions and less favorable interactions) within a single protease variant and focus on the variance between variants, a 15 x 15 correlation matrix, **C**, was computed. Principal components can be defined in terms of an eigenvalue problem: $\boldsymbol{C}u_i = \lambda_i u_i$, which can be solved by the diagonalization of **C**: **C = UC'U$^{-1}$** where and the diagonal elements of **C'** are ordered components of the variance (i.e. the eigenvalues $\lambda_i$). Because this transformation preserves the trace of matrix **C** (Tr**C** = Tr**C'**), and the proportion of total variance, Tr**C**, that is explained by eigenvector $u_i$ is defined as $\frac{\lambda_i}{Tr\boldsymbol{C}}$. Eigenvalues and eigenvectors were calculated using an R interface to the LAPACK (Linear Algebra Package) library. With the van der Waals and hydrogen bonding data, the explained variance was dominated by the first eigenvector (or principal component) and hypothesis testing was use to interpret the spread among the different protease variants. For each amino acid position, the following quantity was computed in order to measure the standard deviation

within each class of van der Waals energies or hydrogen bond occupancies (i.e. high or low values of $u_1$) of protease variants relative to the total mean:

$$\sigma^{*} = \Sigma_{j=1}^{N_{class}} \sqrt{(E_j - \mu^2)}$$

where $\mu = \Sigma_{i=1}^{15} E_{van\ der\ Waals}^{\ i}$ and $N_{class}$ is the total number of variants classified by sequence changes. This deviation allows us to identify important, or at least highly variable residue interactions. All other figures and calculations were determined using various R packages [214-222].

### 3.7.5 Cross Correlation and Dynamical Network Analysis

From each concatenated 300 ns trajectory, 150 frames were used to conduct a dynamic cross correlation network analysis. VMD 1.9.2 was used to convert the fully hydrated all atom Desmond trajectories into DCD binary trajectory file format containing only the Cα atoms of the protein. Once each trajectory file was converted, network analysis was completed using previously published methods for the Bio3D package version 2.2 [199, 223] was used to compute the cross correlation coefficients, $C_{ij}$,

$$C_{ij} = \frac{< \Delta r_i \cdot \Delta r_j >}{\sqrt{( < \Delta r_i^2 > < \Delta r_j^2 > )}}$$

for all Cα atoms pairs in R Studio version 3.1.3 [218], where $\Delta r_i$ is the displacement from the mean position of the *ith* atom determined from all configurations in the trajectory segment being analyzed. Using this data, dynamical networks were constructed following the method of Luthey-Schulten and coworkers [202]. Briefly, this approach uses a weighted graph where each residue represents a node. Edges between nodes *i* and *j* are weighted ($w_{ij}$) by their respective correlation value ($C_{ij}$*):*

$$w_{ij} = -\log(|C_{ij}|)$$

# Chapter IV

# Discussion

## 4.1. Conspectus

Drug resistance occurs as the result of many different mechanisms; although there is a clear tie between the ascent of compensatory mutations and inhibitor treatment their roles in drug resistance are vastly overlooked. With the exception of some cancer and virus studies, compensatory mutations in drug resistant targets are largely studied in an evolutionary manner. While understanding the evolution of mutations within a target may be the first step in predicting further target evolution, solely studying the evolution of the target ignores key parameters that gave rise to both major and minor mutations initially.

To understand the molecular mechanism by which resistance occurs, one must be able to break the resistance down into the mechanistic components and identify those inter-molecular interactions, be they bonded or non-bonded, that are vital for sustained inhibitor targeting. Elucidating the resistance mechanisms used by a target to compromise any or all of the necessary binding interactions is the first step of understanding the role of secondary mutations in resistance. Correspondingly, it is necessary to understand the relationships between secondary mutations alone and together with major or primary mutations. The phenomenon that secondary mutations restore fitness of the target in the presence and/or absence of drug is well documented however, to understand why certain combinations arise together to prompt further resistance goes back to the aforementioned point; one must decompose all parts of drug resistance in the system to elucidate mechanisms used by the target.

Identification and characterization of key interactions between inhibitor and target, in this thesis, DRV and HIV-1 protease, and subsequent methods used to determine mechanisms utilized by compensatory mutations to drive resistance should be broadly applicable to many systems. The specific methods may need to be customized for optimal understanding, but the general ideas of understanding the role of secondary mutations alone and in complex combinations and how they perturb key components of inhibitor binding is necessary to better prepare for resistance with the development of new inhibitors.

## 4.2. Implications for understanding the role of accessory mutations in drug resistance in other protein systems

### 4.2.1. Application of techniques in this thesis to other drug resistant targets

The ability of mutations regarded as compensatory to restore the fitness lost due to drug resistant mutations is well documented in a number of cancers and pathogenic systems including viruses, bacteria, fungi and protozoa [177, 224-226]. For example, compensatory mutations R222Q and V234M in influenza neuraminidase aid in loss of fitness driven by the oseltamivir resistant major mutation H274Y and increase surface expression of neuraminidase to cleave sialic acid moieities [227]. And while the elucidation of this combination of mutations was completed using a combination of experimental and phylogenetic tree analysis, the authors had previously used Bayesian inference to predict the

effect of the compensatory mutation R194G on increasing the stability of neuraminidase containing H274Y [228]. Similarly, single nucleotide substitutions mutations like c.9106C>G in the *BRCA2* gene have been found to restore function of the BRCA2 protein, as this base change removes the ORF truncation at Q2960 by creating Q2960E mutation. Like many studies geared toward the elucidation of mechanisms of resistance, this study was able to identify certain mechanisms (mutations and deletions) using experimental methods and some bioinformatic analysis [229]. It is apparent from studies such as these that there is increasing need to employ computational techniques in combination with experiments in order to fully understand the mechanisms used by mutations distal to the active site of drug resistant targets. In chapter II, I decided to expound upon my static structural vdW and hydrogen bond findings by running molecular dynamics simulations to monitor these potential residue interactions over time. Without making use of this computational technique, many of the overall findings in this thesis may have not been uncovered.

### 4.2.2. Residue “communication” should not be discounted in resistance

Understanding residue communication in mutable drug targets is essential for continued suppression of diseases. Residue communication is how a series of often neighboring residues impact each other in three-dimensional space, either through alterations in structure/packing or dynamics. For instance, residue communication and the perturbations of residue communication has been probed in kinesin motor proteins extensively using MD simulations and correlation

network analyses[223, 230]. They were able to elucidate that various regions of the protein, including those that bind nucleotide and inhibitor, are highly coupled. They were also able to identify, through mutational analyses, amino acid sites that are crucial for the dynamic coupling between various regions in the protein, such E166's role in coordination of the nucleotide and microtubule binding domains. Similarly, Young and coworkers [231] were able to determine that the connector region between the SH2 and SH3 domains of Src kinase becomes uncoupled in the presence of mutations that promote the lack of binding to a phosphorylated tyrosine, which would normally cause the SH2 domain to be coupled with the SH3 domain. The loss of binding and coupling of these two domains in the presence of mutations causes the Src kinase to be constitutively activated, which is detrimental to the regulation of the proteins' active site. Thus such residue communication has been studied in several large protein systems where inter-domain communication is crucial for function.

With respect to HIV-1 protease, the first report of domain communication was published in 1990[201]. In the decades since then the use of aforementioned methods to probe the changes in communication patterns in the presence of mutations and the impact that these changes have on inhibitor binding has been minimal. It hasn't been until recently that a thorough investigation of the proteases communication patterns in a complex mutant background has been completed[153]. In this instance, Appadurai and coworkers were able to elucidate that communication from the proteases allosteric sites to the active site and the

flaps travels along a more direct path while in the mutant proteases, messages are relayed via multiple intermediate steps along a path. These indirect communication pathways in turn lead to uncoupled motions of the flaps, which disrupts inhibitor binding causing resistance. Whether the relationships between residues are manifested in more limited settings such as the relationship between residues V32, I47 and L33 in chapter II or in more complex settings as seen in chapter III, examining the global communication patterns and their impediment in the presence of mutations is imperative for avoiding drug resistance in key therapeutic targets.

### 4.2.3. Incorporating information about accessory mutations into structure based drug design

Though the process of structure based drug design is highly rigorous and iterative by nature, the impact of potential drug resistance mutations against an inhibitor is often neglected. The DRV-resistant variants at the basis of chapter II were crystallized initially because they (with the exception of L90M) had been identified in the clinical POWER trials [186]. In these trials, clinically derived resistant isolates containing many cross-resistant mutations remained susceptible to DRV. However, during virological failure, DRV also selected for mutations outside the active site of the protease in both the POWER trials and clinical studies involving treatment naïve patients [198]. That this phenomenon was discovered in clinical trials and the possible mechanistic rationale that we were able to unearth in chapter II gives rise to the necessity of thoroughly

understanding and incorporating the roles of accessory mutations into drug design.

Identifying a lead compound in the structure based design process is an arduous task. Ensuring that the compound is able to withstand the trade-off between toxicity, convenience, adaptability and potency *in vitro* and *in vivo* is difficult. Because flexible inhibitors are less potent, the task of finding a rigid inhibitor that also has a high genetic barrier to resistance makes the entire process even more complex. However, time should be allotted for compounds identified as potential inhibitors of various drug targets to undergo more diverse analyses at certain points in the structure based drug design process upon the ascent of mutations. As is the case in chapter II we were not able to identify the influence of accessory mutations on the proteases active site until well after DRV was approved. If the analyses presented in both chapters II and III had been completed soon after the discovery of DRV RAMs in clinical trials and prior to its approval, perhaps DRV could have been improved upon to minimize the chance of virological failure. Because mutations are often inevitable, as is the case with HIV-1 drug targets, in the very least, integrating the types of analyses presented here would provide a comprehensive view of the causes behind potential drawbacks of otherwise highly potent compounds.

## 4.3. The importance of understanding dynamics in drug binding

### 4.3.1. Relating protein dynamics to experimental inhibition data

To understand the molecular basis for biological interactions such as protein-protein, protein-inhibitor, or protein folding interactions, the canonical procedure usually involves interpretation of some combination of structure, enzyme kinetics, and thermodynamic properties of the system under evaluation. These can involve many techniques including X-ray crystallography, NMR, various proteolysis and inhibition assays, circular dichroism, and isothermal titration calorimetry or differential scanning calorimetry to name a few. And while the data acquired from these techniques and others are invaluable to our understanding of many systems, newly discovered or thoroughly explored, they lack the ability to provide insight into protein flexibility and overall dynamics on an atomic scale. To explore atomic fluctuations in depth, molecular dynamics simulations are emerging as a way to study all molecular aspects of proteins mentioned above including binding kinetics and folding [232, 233].

In this thesis, I have combined both structural and dynamics data along with experimental data to try and capture as much information relating the structural and dynamic properties of the HIV-1 protease to experimental properties with regard to the roles of accessory mutations' impact on drug binding. In a collaborative study with Dr. Marc Potempa at UNC-Chapel Hill, it can be seen that for the clinically-derived variants in chapter III, those variants with WT like dynamic properties and higher susceptibility to DRV also perform

substrate turnover better than the variants that are more resistant to DRV in the presence and absence of RNA [209]. We see through these studies that variant $SLK_{19}$ has the highest fold change in initial velocity with the addition of RNA on processing the MA/CA substrate peptide. In chapter II, the levels of change in dynamics for the variants with one or two mutations directly aligned with the inhibitor binding constants measured. However, in the more complicated systems seen in chapter III, the connections between dynamics and experimental data are not so simple. This lack of correlation may partially be due the experimental difficulty in measuring the inhibition constant due to the high levels of resistance in the proteases. While I have been able to elucidate that $KY_{26}$ and $SLK_{19}$ are both dynamically similar to WT protease, aligning these findings with experimental and virological properties is not straightforward. Nevertheless, acquisition of various types of data provides a wealth of information about how much protein dynamics plays a role in the physical properties of protein binding and turnover.

### 4.3.2. IV.II.B. Incorporating protein dynamics into inhibitor design

To fully understand the molecular determinants of inhibitor binding, it is necessary to first elucidate the molecular determinants of protein molecular recognition, be it to substrates, other proteins, or small molecules. Previous studies in our lab have determined that the HIV-1 proteases recognition is not based on sequence, as the enzyme is responsible for cleaving twelve sites lacking sequence homology. Alignment of substrate sites led to the discovery

that the substrates occupied a conserved volume and that the protease recognized the shape of the substrates rather than specific sequences. This led to the substrate envelope hypothesis [133, 189]. Further, our lab was able to determine that if inhibitors were designed to fit within the substrate envelope, they would more effectively inhibit the protease [134]. Our lab was also able to probe further the molecular determinants of protease-substrate binding by elucidating and validating the proteases recognition pattern with the dynamic substrate envelope [191, 234].

With respect to inhibitor binding, the canonical *in silico* analysis of ligand binding may not be enough to ensure that an inhibitor is able to outcompete naturally coevolving substrates. Our lab has found that use of dynamic substrate envelope in inhibitor design for both the HIV-1 and HCV NS3/4A proteases provides a strong scaffold for ensuring the robustness of inhibitors as the dynamic substrate envelope is preserved even in the presence of mutations in the enzyme and its naturally coevolved substrates [235]. Certainly incorporation of the dynamic substrate envelope along with the dynamics based analyses presented in this thesis together would promote lasting susceptibility of a drug target to an inhibitor.

## 4.4. The value of machine learning for predicting probabilistic phenotypic drug resistance mechanisms

As mentioned previously, (statistical) machine learning is used heavily for prediction of mutational stabilities of proteins in evolutionary studies [228].

Machine learning techniques are widely employed in the clinic to guide physicians and physician-scientists in choice of viable treatment options for virological success in naïve patients or in the event of necessitated treatment change as a result of virological failure in various viral systems [82, 236, 237]. In the case of HIV, treatment failure is often associated with drug resistance and thus genotypic resistance testing (GRT) is used to predict virological success of subsequent treatment. For prediction of virological response, experts' rules (genotypic interpretation systems (GIS)) and (statistical) machine learning or a combination of the two are employed in several ways that typically fall under the GRT umbrella. The largest use of machine learning is incorporated into various treatment optimization systems (TOSs), such as EuResist, which is an integrated database that allows a user to input patient specific variables to determine the probability of virological success [238].

Machine learning has also proven quite useful for *in vitro* phenotypic susceptibility studies involving decision trees and neural networks [239]. Although various factors, including drug-drug interactions, virus-host interactions, and viral mutational patterns make the prediction of virological response *in vivo* highly complex, machine learning has been suitably employed for prediction *in vivo* response with the help of sequence and clinical databases. While there are some machine learning based methods that do not utilize genotypic data for resistance testing, the goal of predicting virological outcome remains the bottom line for how computation is employed[237]. In summary, machine learning is

largely used to predict, one drug at a time, phenotypic resistance or susceptibility from viral genotypic data.

In this thesis, I am not using machine learning to predict virological response, instead I am using these powerful methods to examine the enzyme mechanistic response itself. The use of machine learning to predict virologic success or failures generally result in a shuffling of medications in hopes of maintaining sustained virological response. However, in depth probing of the mechanisms underlying the observed resistance responses is studied mostly in the laboratory and largely not with machine learning techniques. This thesis examines clinically derived phenotypic responses to DRV, dissects perturbations of key interactions driving DRV resistance, and determines which and how mutations impact the response of the protease to the drug. Because the variants used in this study were essentially selected at random, the techniques used here can be applied to larger panels of proteins and in other systems much like the techniques above for GRT. The specific machine learning techniques used in this study may have to be modified in order for the user to make the most logical interpretations of their data. Similarly, the interactions most perturbed between the protein and the inhibitor at the onset of mutations will also have to be identified and probed for different drugs and different systems. Still, the methods used in this thesis allow for the recognition that interpretation of virological outcome is just as necessary as its predictions in order to prevent resistance.

## 4.5. Using prior knowledge and current techniques to predict the onset of mutations to combat resistance

Throughout this thesis, the importance of studying various biochemical properties both statically and dynamically has been demonstrated. However, to capture the power of dynamics in its totality with respect to elucidating drug resistance, one must account for global patterns and trends within the target and how these patterns relate to drug resistance. Perhaps not all distal mutations play a direct role in drug resistance but their presence is necessary for viral replication otherwise that particular viral species would not be able to outcompete its counterparts. This in and of itself necessitates the need to understand why mutations far from the active site are present and how they contribute to resistance.

In the process of structure based drug design, the methods used in this thesis to examine the roles of mutations outside of the active site can be subsequently incorporated after the “analyze structure of target and lead compound for interactions” step (Figure 4.1). Inclusion could consist of small time scale molecular dynamics simulations to monitor interactions in the identification of lead compounds, in addition to docking and other energy based design strategies. In chapter II, the MD simulations conducted were of a very short time scale but the wealth of information provided led to the network hypothesis where mutations are able to communicate global protein changes via hydrogen bonds. Without the use of dynamics this mechanism may have been overlooked.

These interactions can then be examined for perturbations by mutations founded through expert based rules or other means. The examination of structure networks within a target for the recognition of global communication patterns can be included in this analysis as well. In this regard, the full target network structure would need to be examined in the absence and presence of ligand to determine changes that occur throughout the protein when ligand is bound. In chapter III, we see that generally fixed (i.e. non-mutant) residues are used as “anchors” to communicate changes to the active site and flap regions of the protease (residues 5, 8, 9, 55, 83 etc) by examining the betweenness centrality, even in highly complex mutant backgrounds. A more in-depth analysis would reveal how the communication within each residue community changes between the proteases non-DRV and DRV bound states. Understanding the global protease patterning and perturbations of global communication is useful for gaining insight into what mutations a new compound may trigger that are distal to the active site (Figure 4.2).

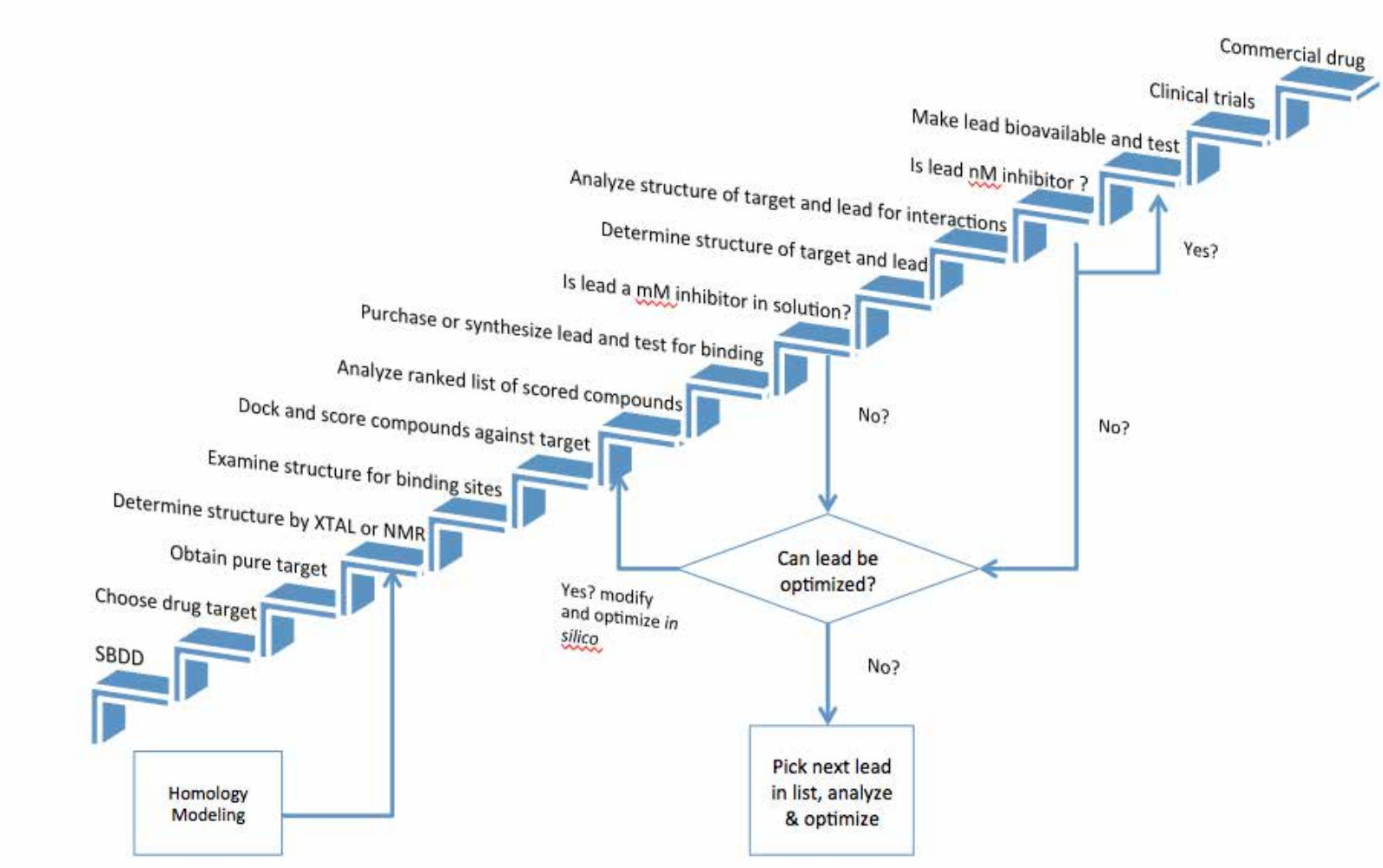



**Figure 4.1 The process of structure-based drug design (SBDD) adapted from [117].**

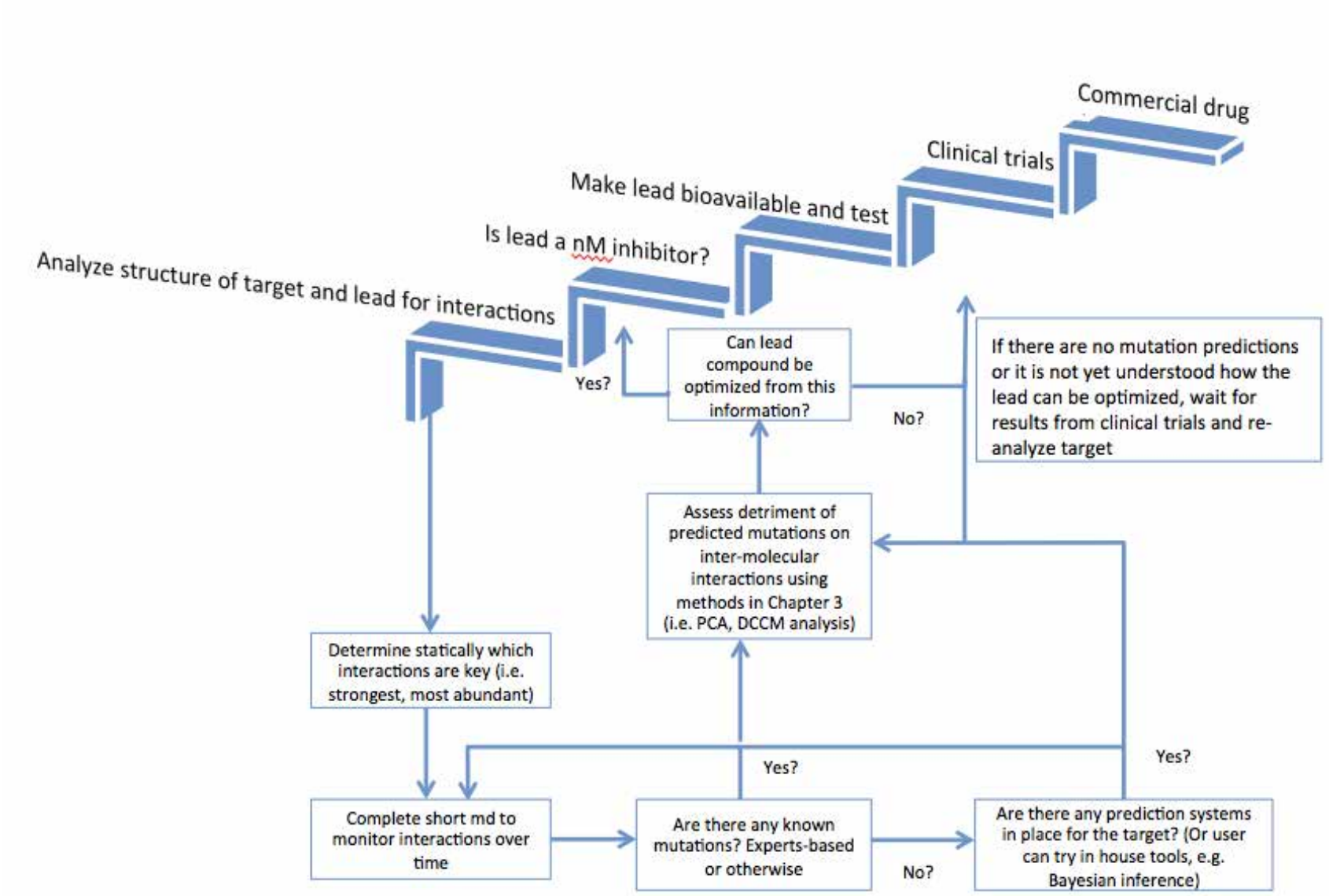



**Figure 4.2 Modification to Figure 4.1 based on the findings presented in this thesis.**

In chapter III, we also determine, from proteases of varied backgrounds, which specific mutations are most responsible for the changes in interactions necessary for DRV targeting. Knowing which mutations most heavily influence alterations in key interactions between a target and compound unlocks further implications behind the reason for the presence of mutations within and distal to the active site. For instance, determining that mutation I84V is going induce mutations at peripheral and distal sites such as 13, 32, 33 and 46 provides insight into why these mutations may have appeared in clinical trials with DRV. Previously, other groups have shown that non-active site mutations can appear before their active site counterparts to minimize the deleterious effects of active site mutations on viral replication [227]. DRV has been shown in multiple clinical trials to select for accessory mutations such as V32I and L33F without selecting for any active site mutations early on [186, 198]. Our findings substantiate that this inverse resistance mechanism may be employed by the protease against potent inhibitors such as DRV. Without the knowledge gained from our mutational studies, this information would not be confirmed.

In this work, we also looked at clinically derived variants and their resistance to DRV. Based on experts rules, DRV doesn't necessarily select for all of the mutations we found to impede on its binding. This gives further insight into residues' possible multiple roles in cross-resistance. Had this knowledge been implemented over the course of patient treatment to whichever approved inhibitor they were given at the time, a more informed decision could have been made for

their treatment optimization prior to accumulation of the many mutations seen. To aid clinicians and scientists in vigilance against drug resistance I propose that several steps be taken to comprehensively determine the best treatment options for patients. First, the collection of temporal resistance data must be acquired. This acquisition can be achieved in the laboratory such as those studies carried out by the scientists in the Swanstrom Lab at UNC-Chapel Hill from which we used several variants seen in chapter III. Or this acquisition could be attained from chronological patient sample collection and corresponding treatment history routinely and before treatment failure. Second, the dynamic interactions made between the target and inhibitor should be examined via molecular dynamics simulations and stored into a database. Finally, key interactions and their perturbations in the presence of mutations should be thoroughly investigated, including mutational impact on bonded and non-bonded interactions and the global networking of the protein.

Much like the computer-aided GIS and TOS databases used for GRT to predict virological outcomes, taking the above steps could lead to the generation of computer-aided tools that would predict the onset mutations to an otherwise predictably successful inhibitor. Combining experimental knowledge and that information which was explored in this thesis; mechanisms employed in cross-resistance, the interdependency of mutations within and away from the active site, and global communication patterns into a program could be complementary to the databases mentioned above. A program like this would be able to predict

likely mutations via the examination of interactions between protein and inhibitor and anticipate subsequent mutational patterns that may arise after the onset of mutations to a drug and allow for a more extensive prediction of treatment options that would best work for patients.

## 4.6. Concluding Remarks

In conclusion, this thesis is an attempt to elucidate the interdependency of mutations distal to the active site of the HIV-1 protease and their direct role in drug resistance to protease inhibitor darunavir (DRV). We employed various techniques to examine the molecular dynamics simulations of selected protease variants complexed with DRV containing single site accessory mutations and more highly complex combinations of mutations. From these studies we have been able to determine that accessory mutations alone drive low levels of resistance to this potent inhibitor but they are able to alter the means by which DRV binds to the protease via the dynamic network of hydrogen bonds within the protease. In addition, we have been able to identify which mutations most perturb the interactions necessary for DRV to inhibit the protease. We were also able to elucidate from this panel of proteases, how the global interactions change throughout the protein in the presence of accessory mutations. Future directions for this work would be to examine this panel of variants using the methods in this thesis both in the absence of ligand and with substrates to gain insight into how the presence of mutations generally impacts dynamics. My hope is that this work

will aid in the improvement of inhibitor design in the future and provide a more complete assessment of treatment optimization in the clinic.

# Bibliography


1. CDC, *Pneumocystis Pneumonia --- Los Angeles.* Morbidity and Mortality Weekly Report, 1981. **30**(21): p. 1-3.
2. CDC, *A Cluster of Kaposi's Sarcoma and Pneumocystis carinii Pneumonia among Homosexual Male Residents of Los Angeles and Orange Counties, California.* Morbidity and Mortality Weekly Report, 1982. **31**(23): p. 305-307.
3. CDC, *Current Trends Update on Acquired Immune Deficiency Syndrome(AIDS) -- United States.* Morbidity and Mortality Weekly Report, 1982. **31**(37): p. 507-508, 513-514.
4. Ratner, L., et al., *Complete nucleotide sequence of the AIDS virus, HTLV-III.* Nature, 1985. **313**(6000): p. 277-84.
5. Sanchez-Pescador, R., et al., *Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).* Science, 1985. **227**(4686): p. 484-92.
6. Alizon, M.M., L. , *Lymphodenopathy/AIDS virus: genetic organization and relationship to animal lentiviruses.* Anticancer Research, 1986. **6**(3): p. 403-411.
7. Barr, et al., *Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS).* Science, 1983. **220**(4599): p. 868-871.
8. Gallo , R.C. and L. Montagnier *The Discovery of HIV as the Cause of AIDS.* New England Journal of Medicine, 2003. **349**(24): p. 2283-2285.
9. Coffin, J., et al., *Human Immunodeficiency Viruses.* Science, 1986. **232**(4751): p. 697-697.
10. Clavel, F., et al., *Isolation of a new human retrovirus from West African patients with AIDS.* Science, 1986. **233**(4761): p. 343-6.
11. Chakrabarti, L., et al., *Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses.* Nature, 1987. **328**(6130): p. 543-7.
12. Quinn, T.C., et al., *AIDS in Africa: An Epidemiologic Paradigm.* Science, 1986. **234**(4779): p. 955-963.
13. Rudicell, R.S., et al., *Impact of simian immunodeficiency virus infection on chimpanzee population dynamics.* PLoS Pathog, 2010. **6**(9): p. e1001116.
14. Keele, B.F., et al., *Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz.* Nature, 2009. **460**(7254): p. 515-9.
15. Gray, R.H., et al., *Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda.* Lancet, 2001. **357**(9263): p. 1149-53.
16. Bailes, E., et al., *Hybrid origin of SIV in chimpanzees.* Science, 2003. **300**(5626): p. 1713.

17. Sharp, P.M. and B.H. Hahn, *Origins of HIV and the AIDS Pandemic.* Cold Spring Harbor Perspectives in Medicine:, 2011. **1**(1): p. a006841.
18. Keele, B.F., et al., *Chimpanzee reservoirs of pandemic and nonpandemic HIV-1.* Science, 2006. **313**(5786): p. 523-6.
19. Peeters, M., et al., *Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat.* Emerg Infect Dis, 2002. **8**(5): p. 451-7.
20. Subbarao, S. and G. Schochetman, *Genetic variability of HIV-1.* Aids, 1996. **10 Suppl A**: p. S13-23.
21. Simon, F., et al., *Identification of a new human immunodeficiency virus type 1 distinct from group M and group O.* Nat Med, 1998. **4**(9): p. 1032-7.
22. Vidal, N., et al., *Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa.* J Virol, 2000. **74**(22): p. 10498-507.
23. Worobey, M., et al., *Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960.* Nature, 2008. **455**(7213): p. 661-4.
24. Taylor, B.S. and S.M. Hammer, *The challenge of HIV-1 subtype diversity.* N Engl J Med, 2008. **359**(18): p. 1965-6.
25. Gilbert, M.T., et al., *The emergence of HIV/AIDS in the Americas and beyond.* Proc Natl Acad Sci U S A, 2007. **104**(47): p. 18566-70.
26. Sundquist, W.I. and H.G. Krausslich, *HIV-1 assembly, budding, and maturation.* Cold Spring Harb Perspect Med, 2012. **2**(7): p. a006924.
27. Turner, B.G. and M.F. Summers, *Structural biology of HIV.* J Mol Biol, 1999. **285**(1): p. 1-32.
28. Usami, Y., et al., *The ESCRT pathway and HIV-1 budding.* Biochem Soc Trans, 2009. **37**(Pt 1): p. 181-4.
29. Strack, B., et al., *AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding.* Cell, 2003. **114**(6): p. 689-99.
30. Garrus, J.E., et al., *Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding.* Cell, 2001. **107**(1): p. 55-65.
31. Martin-Serrano, J., T. Zang, and P.D. Bieniasz, *HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress.* Nat Med, 2001. **7**(12): p. 1313-9.
32. Weiss, R.A., *Gulliver's travels in HIVland.* Nature, 2001. **410**(6831): p. 963-7.
33. Wei, X., et al., *Viral dynamics in human immunodeficiency virus type 1 infection.* Nature, 1995. **373**(6510): p. 117-22.
34. Lemey, P., A. Rambaut, and O.G. Pybus, *HIV evolutionary dynamics within and among hosts.* AIDS Rev, 2006. **8**(3): p. 125-40.
35. Ho, D.D., et al., *Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection.* Nature, 1995. **373**(6510): p. 123-6.
36. Perelson, A.S., et al., *HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time.* Science, 1996. **271**(5255): p. 1582-6.

37. Mansky, L.M., *Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line.* AIDS Res Hum Retroviruses, 1996. **12**(4): p. 307-14.
38. Kelly, J.K., *Replication rate and evolution in the human immunodeficiency virus.* J Theor Biol, 1996. **180**(4): p. 359-64.
39. Mansky, L.M. and H.M. Temin, *Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase.* J Virol, 1995. **69**(8): p. 5087-94.
40. Eigen, M., *Viral Quasispecies*. 1993, Scientific American, Inc: UNITED STATES. p. 42-49.
41. Cichutek, K., et al., *Development of a quasispecies of human immunodeficiency virus type 1 in vivo.* Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7365-9.
42. Delwart, E.L., et al., *Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS.* J Virol, 1997. **71**(10): p. 7498-508.
43. McDonald, R.A., et al., *Evolution of human immunodeficiency virus type 1 env sequence variation in patients with diverse rates of disease progression and T-cell function.* J Virol, 1997. **71**(3): p. 1871-9.
44. Wahlberg, J., et al., *Dynamic changes in HIV-1 quasispecies from azidothymidine (AZT)-treated patients.* Faseb j, 1992. **6**(10): p. 2843-7.
45. Robertson, D.L., et al., *Recombination in HIV-1.* Nature, 1995. **374**(6518): p. 124-6.
46. Stahl, F.W., *Genetic recombination.* Sci Am, 1987. **256**(2): p. 90-101.
47. Takehisa, J., et al., *Human immunodeficiency virus type 1 intergroup (M/O) recombination in cameroon.* J Virol, 1999. **73**(8): p. 6810-20.
48. McCune, J.M., *The dynamics of CD4+ T-cell depletion in HIV disease.* Nature, 2001. **410**(6831): p. 974-9.
49. Rosenberg, Y.J., et al., *Decline in the CD4+ lymphocyte population in the blood of SIV-infected macaques is not reflected in lymph nodes.* AIDS Res Hum Retroviruses, 1993. **9**(7): p. 639-46.
50. Kirschner, D., G.F. Webb, and M. Cloyd, *Model of HIV-1 disease progression based on virus-induced lymph node homing and homing-induced apoptosis of CD4+ lymphocytes.* J Acquir Immune Defic Syndr, 2000. **24**(4): p. 352-62.
51. Hengel, R.L., et al., *Markers of lymphocyte homing distinguish CD4 T cell subsets that turn over in response to HIV-1 infection in humans.* J Immunol, 1999. **163**(6): p. 3539-48.
52. Hazenberg, M.D., et al., *T cell depletion in HIV-1 infection: how CD4+ T cells go out of stock.* Nat Immunol, 2000. **1**(4): p. 285-9.
53. Grossman, Z., M.B. Feinberg, and W.E. Paul, *Multiple modes of cellular activation and virus transmission in HIV infection: a role for chronically and latently infected cells in sustaining viral replication.* Proc Natl Acad Sci U S A, 1998. **95**(11): p. 6314-9.
54. Moses, A., J. Nelson, and G.C. Bagby, Jr., *The influence of human immunodeficiency virus-1 on hematopoiesis.* Blood, 1998. **91**(5): p. 1479-95.

55. Ventoso, I., et al., *Involvement of HIV-1 protease in virus-induced cell killing.* Antiviral Res, 2005. **66**(1): p. 47-55.
56. Doitsh, G., et al., *Abortive HIV infection mediates CD4 T cell depletion and inflammation in human lymphoid tissue.* Cell, 2010. **143**(5): p. 789-801.
57. Doitsh, G., et al., *Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection.* Nature, 2014. **505**(7484): p. 509-14.
58. Monroe, K.M., et al., *IFI16 DNA sensor is required for death of lymphoid CD4 T cells abortively infected with HIV.* Science, 2014. **343**(6169): p. 428-32.
59. Wain-Hobson, S., et al., *Nucleotide sequence of the AIDS virus, LAV.* Cell, 1985. **40**(1): p. 9-17.
60. Ohlmann, T., et al., *In vitro cleavage of eIF4GI but not eIF4GII by HIV-1 protease and its effects on translation in the rabbit reticulocyte lysate system.* J Mol Biol, 2002. **318**(1): p. 9-20.
61. Nie, Z., et al., *Human immunodeficiency virus type 1 protease cleaves procaspase 8 in vivo.* J Virol, 2007. **81**(13): p. 6947-56.
62. Towler, E.M., et al., *Functional characterization of the protease of human endogenous retrovirus, K10: can it complement HIV-1 protease?* Biochemistry, 1998. **37**(49): p. 17137-44.
63. Benes, P., V. Vetvicka, and M. Fusek, *Cathepsin D--many functions of one aspartic protease.* Crit Rev Oncol Hematol, 2008. **68**(1): p. 12-28.
64. Meek, T.D., et al., *Human immunodeficiency virus 1 protease expressed in Escherichia coli behaves as a dimeric aspartic protease.* Proc Natl Acad Sci U S A, 1989. **86**(6): p. 1841-5.
65. Pearl, L.H. and W.R. Taylor, *A structural model for the retroviral proteases.* Nature, 1987. **329**(6137): p. 351-4.
66. Ingr, M., et al., *Kinetics of the dimerization of retroviral proteases: the "fireman's grip" and dimerization.* Protein Sci, 2003. **12**(10): p. 2173-82.
67. Alvizo, O., et al., *Structural, kinetic, and thermodynamic studies of specificity designed HIV-1 protease.* Protein Sci, 2012. **21**(7): p. 1029-41.
68. Pearl, L.H. and W.R. Taylor, *Sequence specificity of retroviral proteases.* Nature, 1987. **328**(6130): p. 482.
69. Agniswamy, J., et al., *HIV-1 protease with 20 mutations exhibits extreme resistance to clinical inhibitors through coordinated structural rearrangements.* Biochemistry, 2012. **51**(13): p. 2819-28.
70. Surleraux, D.L., et al., *Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor.* J Med Chem, 2005. **48**(6): p. 1813-22.
71. Nelson, D.L.a.C., Michael M. , *Lehninger Principles of Biochemistry*. 5 ed. 2008, New York, NY: Sara Tenney. 1294.
72. Brik, A. and C.-H. Wong, *HIV-1 protease: mechanism and drug discovery.* Organic & Biomolecular Chemistry, 2003. **1**(1): p. 5-14.
73. Kramer, R.A., et al., *HTLV-III gag protein is processed in yeast cells by the virus pol-protease.* Science, 1986. **231**(4745): p. 1580-4.

74. Richards, A.D., et al., *Effective blocking of HIV-1 proteinase activity by characteristic inhibitors of aspartic proteinases.* FEBS Lett, 1989. **247**(1): p. 113-7.
75. Dreyer, G.B., et al., *Inhibition of human immunodeficiency virus 1 protease in vitro: rational design of substrate analogue inhibitors.* Proc Natl Acad Sci U S A, 1989. **86**(24): p. 9752-6.
76. Meek, T.D. and G.B. Dreyer, *HIV-1 protease as a potential target for anti-AIDS therapy.* Ann N Y Acad Sci, 1990. **616**: p. 41-53.
77. Roberts, N.A., et al., *Rational Design of Peptide-Based HIV Proteinase Inhibitors.* Science, 1990. **248**(4953): p. 358-361.
78. Debouck, C., *The HIV-1 protease as a therapeutic target for AIDS.* AIDS Res Hum Retroviruses, 1992. **8**(2): p. 153-64.
79. Richman, D.D., *HIV chemotherapy.* (0028-0836 (Print)).
80. Mascolini, M., *HAART, hubris, and humility. Decade of HAART. September 25-26, 2006--San Francisco.* IAPAC Mon, 2006. **12**(12): p. 424-35.
81. Miller, M., et al., *Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution.* Science, 1989. **246**(4934): p. 1149-52.
82. Erickson, J., et al., *Design, activity, and 2.8 A crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease.* Science, 1990. **249**(4968): p. 527-33.
83. Navia, M.A., et al., *Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1.* Nature, 1989. **337**(6208): p. 615-20.
84. Mellors, J.W., et al., *Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion.* Ann Intern Med, 1995. **122**(8): p. 573-9.
85. Mellors, J.W., et al., *Prognosis in HIV-1 infection predicted by the quantity of virus in plasma.* Science, 1996. **272**(5265): p. 1167-70.
86. Swanstrom, R. and J. Erona, *Human immunodeficiency virus type-1 protease inhibitors: therapeutic successes and failures, suppression and resistance.* Pharmacol Ther, 2000. **86**(2): p. 145-70.
87. Frost, S.D. and A.R. McLean, *Quasispecies dynamics and the emergence of drug resistance during zidovudine therapy of HIV infection.* Aids, 1994. **8**(3): p. 323-32.
88. Nowak, M.a.M., Robert, M. , *AIDS pathogenesis: mathematical models of HIV and SIV Infections.* AIDS 1993. **7**(Supplement 1): p. S3-S18.
89. Nowak, M.A., et al., *Anti-viral drug treatment: dynamics of resistance in free virus and infected cell populations.* J Theor Biol, 1997. **184**(2): p. 203-17.
90. Coffin, J.M., *HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy.* Science, 1995. **267**(5197): p. 483-9.
91. Larder, B.A., S.D. Kemp, and P.R. Harrigan, *Potential mechanism for sustained antiretroviral efficacy of AZT-3TC combination therapy.* Science, 1995. **269**(5224): p. 696-9.

92. Eron, J.J., et al., *Treatment with lamivudine, zidovudine, or both in HIV-positive patients with 200 to 500 CD4+ cells per cubic millimeter. North American HIV Working Party.* N Engl J Med, 1995. **333**(25): p. 1662-9.
93. Gulick, R.M., et al., *Simultaneous vs sequential initiation of therapy with indinavir, zidovudine, and lamivudine for HIV-1 infection: 100-week follow-up.* Jama, 1998. **280**(1): p. 35-41.
94. CDC, *Report of the NIH Panel to Define Principles of Therapy of HIV Infection and Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents.* Morbidity and Mortality Weekly Report Recommendations and Reports, 1998. **47**(No. RR-5).
95. Rerks-Ngarm, S., et al., *Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand.* N Engl J Med, 2009. **361**(23): p. 2209-20.
96. Liao, H.X., et al., *Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus.* Nature, 2013. **496**(7446): p. 469-76.
97. Hammer, S.M., et al., *Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine.* N Engl J Med, 2013. **369**(22): p. 2083-92.
98. Fauci, A.S. and H.D. Marston, *Ending AIDS--is an HIV vaccine necessary?* N Engl J Med, 2014. **370**(6): p. 495-8.
99. Cihlar, T. and M. Fordyce, *Current status and prospects of HIV treatment.* Curr Opin Virol, 2016. **18**: p. 50-6.
100. Palella, F.J., Jr., et al., *Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators.* N Engl J Med, 1998. **338**(13): p. 853-60.
101. Harrigan, P.R., et al., *Predictors of HIV drug-resistance mutations in a large antiretroviral-naive cohort initiating triple antiretroviral therapy.* J Infect Dis, 2005. **191**(3): p. 339-47.
102. Gardner, E.M., et al., *Antiretroviral medication adherence and the development of class-specific antiretroviral resistance.* Aids, 2009. **23**(9): p. 1035-46.
103. Havlir, D.V., et al., *Drug susceptibility in HIV infection after viral rebound in patients receiving indinavir-containing regimens.* Jama, 2000. **283**(2): p. 229-34.
104. Vanhove, G.F., et al., *Patient compliance and drug failure in protease inhibitor monotherapy.* Jama, 1996. **276**(24): p. 1955-6.
105. Maggiolo, F., et al., *Effect of adherence to HAART on virologic outcome and on the selection of resistance-conferring mutations in NNRTI- or PI-treated patients.* HIV Clin Trials, 2007. **8**(5): p. 282-92.
106. Arribas, J.R., et al., *The MONET trial: darunavir/ritonavir with or without nucleoside analogues, for patients with HIV RNA below 50 copies/ml.* Aids, 2010. **24**(2): p. 223-30.
107. Pulido, F., et al., *No evidence for evolution of protease inhibitor resistance from standard genotyping, after three years of treatment with darunavir/ritonavir, with or without nucleoside analogues.* AIDS Res Hum Retroviruses, 2012. **28**(10): p. 1167-9.

108. Taiwo, B., et al., *Efficacy of a nucleoside-sparing regimen of darunavir/ritonavir plus raltegravir in treatment-naive HIV-1-infected patients (ACTG A5262).* Aids, 2011. **25**(17): p. 2113-22.
109. Kempf, D.J., et al., *Incidence of resistance in a double-blind study comparing lopinavir/ritonavir plus stavudine and lamivudine to nelfinavir plus stavudine and lamivudine.* J Infect Dis, 2004. **189**(1): p. 51-60.
110. Rosenbloom, D.I., et al., *Antiretroviral dynamics determines HIV evolution and predicts therapy outcome.* Nat Med, 2012. **18**(9): p. 1378-85.
111. Moore, M.L., et al., *Peptide substrates and inhibitors of the HIV-1 protease.* Biochem Biophys Res Commun, 1989. **159**(2): p. 420-5.
112. Kolli, M., et al., *HIV-1 protease-substrate coevolution in nelfinavir resistance.* J Virol, 2014. **88**(13): p. 7145-54.
113. Clavel, F. and F. Mammano, *Role of Gag in HIV Resistance to Protease Inhibitors.* Viruses, 2010. **2**(7): p. 1411-26.
114. Shibata, J., et al., *Within-host co-evolution of Gag P453L and protease D30N/N88D demonstrates virological advantage in a highly protease inhibitor-exposed HIV-1 case.* Antiviral Res, 2011. **90**(1): p. 33-41.
115. Parry, C.M., et al., *Gag determinants of fitness and drug susceptibility in protease inhibitor-resistant human immunodeficiency virus type 1.* J Virol, 2009. **83**(18): p. 9094-101.
116. Mountain, V., *Astex, Structural Genomix, and Syrrx. I can see clearly now: structural biology and drug discovery.* Chem Biol, 2003. **10**(2): p. 95-8.
117. Anderson, A.C., *The process of structure-based drug design.* Chem Biol, 2003. **10**(9): p. 787-97.
118. Kuritzkes, D., S. Kar, and P. Kirkpatrick, *Fresh from the pipeline - Maraviroc.* NATURE REVIEWS DRUG DISCOVERY, 2008. **7**(1): p. 15-16.
119. Matthews, T., et al., *Enfuvirtide: the first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes.* (1474-1776 (Print)).
120. Kohlstaedt, L.A., et al., *Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor.* Science, 1992. **256**(5065): p. 1783-90.
121. Altman, M.D., et al., *HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants.* J Am Chem Soc, 2008. **130**(19): p. 6099-113.
122. Cihlar, T., et al., *Suppression of HIV-1 protease inhibitor resistance by phosphonate-mediated solvent anchoring.* J Mol Biol, 2006. **363**(3): p. 635-47.
123. Ghosh, A.K., Z.L. Dawson, and H. Mitsuya, *Darunavir, a conceptually new HIV-1 protease inhibitor for the treatment of drug-resistant HIV.* Bioorg Med Chem, 2007. **15**(24): p. 7576-80.
124. Clavel, F. and A.J. Hance, *HIV drug resistance.* N Engl J Med, 2004. **350**(10): p. 1023-35.
125. Boucher, C.A., et al., *High-level resistance to (-) enantiomeric 2'-deoxy-3'-thiacytidine in vitro is due to one amino acid substitution in the catalytic site of*

*human immunodeficiency virus type 1 reverse transcriptase.* Antimicrob Agents Chemother, 1993. **37**(10): p. 2231-4.
126. Esnouf, R.M., et al., *Unique features in the structure of the complex between HIV-1 reverse transcriptase and the bis(heteroaryl)piperazine (BHAP) U-90152 explain resistance mutations for this nonnucleoside inhibitor.* Proc Natl Acad Sci U S A, 1997. **94**(8): p. 3984-9.
127. Rimsky, L.T., D.C. Shugars, and T.J. Matthews, *Determinants of human immunodeficiency virus type 1 resistance to gp41-derived inhibitory peptides.* J Virol, 1998. **72**(2): p. 986-93.
128. Di Santo, R., *Inhibiting the HIV integration process: past, present, and the future.* J Med Chem, 2014. **57**(3): p. 539-66.
129. Mouscadet, J.F. and L. Tchertanov, *Raltegravir: molecular basis of its mechanism of action.* Eur J Med Res, 2009. **14 Suppl 3**: p. 5-16.
130. Condra, J.H., et al., *In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors.* Nature, 1995. **374**(6522): p. 569-71.
131. Leavitt, S. and E. Freire, *Direct measurement of protein binding energetics by isothermal titration calorimetry.* Curr Opin Struct Biol, 2001. **11**(5): p. 560-6.
132. Weber, I.T. and J. Agniswamy, *HIV-1 Protease: Structural Perspectives on Drug Resistance.* Viruses, 2009. **1**(3): p. 1110-36.
133. Prabu-Jeyabalan, M., E. Nalivaika, and C.A. Schiffer, *Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes.* Structure, 2002. **10**(3): p. 369-81.
134. King, N.M., et al., *Combating susceptibility to drug resistance: lessons from HIV-1 protease.* Chem Biol, 2004. **11**(10): p. 1333-8.
135. Wang, Y., et al., *The higher barrier of darunavir and tipranavir resistance for HIV-1 protease.* Biochem Biophys Res Commun, 2011. **412**(4): p. 737-42.
136. Jacobsen, H., et al., *Characterization of human immunodeficiency virus type 1 mutants with decreased sensitivity to proteinase inhibitor Ro 31-8959.* Virology, 1995. **206**(1): p. 527-34.
137. Molla, A., et al., *Ordered accumulation of mutations in HIV protease confers resistance to ritonavir.* Nat Med, 1996. **2**(7): p. 760-6.
138. Ozen, A., et al., *Structural basis and distal effects of Gag substrate coevolution in drug resistance to HIV-1 protease.* Proc Natl Acad Sci U S A, 2014. **111**(45): p. 15993-8.
139. Mittal, S., et al., *Structural and thermodynamic basis of amprenavir/darunavir and atazanavir resistance in HIV-1 protease with mutations at residue 50.* J Virol, 2013. **87**(8): p. 4176-84.
140. Martinez-Picado, J., et al., *Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1.* J Virol, 1999. **73**(5): p. 3744-52.
141. Nijhuis, M., et al., *Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy.* Aids, 1999. **13**(17): p. 2349-59.

142. Theys, K., et al., *Treatment-associated polymorphisms in protease are significantly associated with higher viral load and lower CD4 count in newly diagnosed drug-naive HIV-1 infected patients.* Retrovirology, 2012. **9**: p. 81.
143. Arts, E.J., *Commentary on the role of treatment-related HIV compensatory mutations on increasing virulence: new discoveries twenty years since the clinical testing of protease inhibitors to block HIV-1 replication.* BMC Med, 2012. **10**: p. 114.
144. Deforche, K., et al., *Estimation of an in vivo fitness landscape experienced by HIV-1 under drug selective pressure useful for prediction of drug resistance evolution during treatment.* Bioinformatics, 2008. **24**(1): p. 34-41.
145. Liu, Z., et al., *Nine crystal structures determine the substrate envelope of the MDR HIV-1 protease.* Protein J, 2011. **30**(3): p. 173-83.
146. Liu, Z., et al., *Conserved hydrogen bonds and water molecules in MDR HIV-1 protease substrate complexes.* Biochem Biophys Res Commun, 2013. **430**(3): p. 1022-7.
147. Saskova, K.G., et al., *Molecular characterization of clinical isolates of human immunodeficiency virus resistant to the protease inhibitor darunavir.* J Virol, 2009. **83**(17): p. 8810-8.
148. Mahalingam, B., et al., *Combining mutations in HIV-1 protease to understand mechanisms of resistance.* Proteins, 2002. **48**(1): p. 107-16.
149. Chang, M.W. and B.E. Torbett, *Accessory mutations maintain stability in drug-resistant HIV-1 protease.* J Mol Biol, 2011. **410**(4): p. 756-60.
150. Olsen, D.B., et al., *Non-active site changes elicit broad-based cross-resistance of the HIV-1 protease to inhibitors.* J Biol Chem, 1999. **274**(34): p. 23699-701.
151. Perno, C.F., et al., *Low prevalence of primary mutations associated with drug resistance in antiviral-naive patients at therapy initiation.* Aids, 2002. **16**(4): p. 619-24.
152. Muzammil, S., P. Ross, and E. Freire, *A major role for a set of non-active site mutations in the development of HIV-1 protease drug resistance.* Biochemistry, 2003. **42**(3): p. 631-8.
153. Appadurai, R. and S. Senapati, *Dynamical Network of HIV-1 Protease Mutants Reveals the Mechanism of Drug Resistance and Unhindered Activity.* Biochemistry, 2016. **55**(10): p. 1529-40.
154. Young, T.P., et al., *Prevalence, mutation patterns, and effects on protease inhibitor susceptibility of the L76V mutation in HIV-1 protease.* Antimicrob Agents Chemother, 2010. **54**(11): p. 4903-6.
155. Kuiper, B.D., et al.
156. Louis, J.M., et al., *The L76V drug resistance mutation decreases the dimer stability and rate of autoprocessing of HIV-1 protease by reducing internal hydrophobic contacts.* Biochemistry, 2011. **50**(21): p. 4786-95.
157. Foulkes-Murzycki, J.E., W.R. Scott, and C.A. Schiffer, *Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease.* Structure, 2007. **15**(2): p. 225-33.

158. Mittal, S., et al., *Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease.* J Am Chem Soc, 2012. **134**(9): p. 4163-8.
159. Nakahara, K., et al., *Secondary mutations in viruses resistant to HIV-1 integrase inhibitors that restore viral infectivity and replication kinetics.* Antiviral Res, 2009. **81**(2): p. 141-6.
160. Huang, L. and L. Fu, *Mechanisms of resistance to EGFR tyrosine kinase inhibitors.* Acta Pharm Sin B, 2015. **5**(5): p. 390-401.
161. Gazdar, A.F., *Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors.* Oncogene, 2009. **28 Suppl 1**: p. S24-31.
162. Levantino, M., et al., *Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules.* Curr Opin Struct Biol, 2015. **35**: p. 41-8.
163. Ishima, R., J.M. Louis, and D.A. Torchia, *Characterization of two hydrophobic methyl clusters in HIV-1 protease by NMR spin relaxation in solution.* J Mol Biol, 2001. **305**(3): p. 515-21.
164. Galiano, L., et al., *Solute effects on spin labels at an aqueous-exposed site in the flap region of HIV-1 protease.* J Phys Chem B, 2009. **113**(6): p. 1673-80.
165. Naicker, P., et al., *Amide hydrogen exchange in HIV-1 subtype B and C proteases--insights into reduced drug susceptibility and dimer stability.* Febs j, 2014. **281**(24): p. 5395-410.
166. Amaro, R.E., et al., *Mechanism of 150-cavity formation in influenza neuraminidase.* Nat Commun, 2011. **2**: p. 388.
167. Cai, Y., et al., *Differential Flap Dynamics in Wild-type and a Drug Resistant Variant of HIV-1 Protease Revealed by Molecular Dynamics and NMR Relaxation.* J Chem Theory Comput, 2012. **8**(10): p. 3452-3462.
168. Cai, Y., et al., *Drug Resistance Mutations Alter Dynamics of Inhibitor-Bound HIV-1 Protease.* J Chem Theory Comput, 2014. **10**(8): p. 3438-3448.
169. Ragland, D.A., et al., *Drug resistance conferred by mutations outside the active site through alterations in the dynamic and structural ensemble of HIV-1 protease.* J Am Chem Soc, 2014. **136**(34): p. 11956-63.
170. Nakashima, M., et al., *Unique Flap Conformation in an HIV-1 Protease with High-Level Darunavir Resistance.* Front Microbiol, 2016. **7**: p. 61.
171. N, N., et al., *Mechanism of artemisinin resistance for malaria PfATP6 L263 mutations and discovering potential antimalarials: An integrated computational approach.* Sci Rep, 2016. **6**: p. 30106.
172. Nalam, M.N., et al., *Substrate envelope-designed potent HIV-1 protease inhibitors to avoid drug resistance.* Chem Biol, 2013. **20**(9): p. 1116-24.
173. Shen, C., et al., *Automated prediction of HIV drug resistance from genotype data.* BMC Bioinformatics, 2016. **17 Suppl 8**: p. 278.
174. Lefebvre, E. and C.A. Schiffer, *Resilience to resistance of HIV-1 protease inhibitors: profile of darunavir.* AIDS Rev, 2008. **10**(3): p. 131-42.

175. Thompson, M.A., et al., *Antiretroviral treatment of adult HIV infection: 2010 recommendations of the International AIDS Society-USA panel.* Jama, 2010. **304**(3): p. 321-33.
176. Volberding, P.A. and S.G. Deeks, *Antiretroviral therapy and management of HIV infection.* Lancet, 2010. **376**(9734): p. 49-62.
177. Wu, T.D., et al., *Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments.* J Virol, 2003. **77**(8): p. 4836-47.
178. Saskova, K.G., et al., *Enzymatic and structural analysis of the I47A mutation contributing to the reduced susceptibility to HIV protease inhibitor lopinavir.* Protein Sci, 2008. **17**(9): p. 1555-64.
179. Kovalevsky, A.Y., et al., *Effectiveness of nonpeptide clinical inhibitor TMC-114 on HIV-1 protease with highly drug resistant mutations D30N, I50V, and L90M.* J Med Chem, 2006. **49**(4): p. 1379-87.
180. Ode, H., et al., *Computational simulations of HIV-1 proteases--multi-drug resistance due to nonactive site mutation L90M.* J Am Chem Soc, 2006. **128**(24): p. 7887-95.
181. Shen, C.H., et al., *Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters.* Febs j, 2010. **277**(18): p. 3699-714.
182. King, N.M., et al., *Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type 1 protease inhibitor.* J Virol, 2004. **78**(21): p. 12012-21.
183. Nalam, M.N., et al., *Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance.* J Virol, 2010. **84**(10): p. 5368-78.
184. Santos, J.R., et al., *Antiretroviral simplification with darunavir/ritonavir monotherapy in routine clinical practice: safety, effectiveness, and impact on lipid profile.* PLoS One, 2012. **7**(5): p. e37442.
185. Cai, Y. and C.A. Schiffer, *Decomposing the energetic impact of drug resistant mutations in HIV-1 protease on binding DRV.* J Chem Theory Comput, 2010. **6**(4): p. 1358-1368.
186. de Meyer, S., et al., *Resistance profile of darunavir: combined 24-week results from the POWER trials.* AIDS Res Hum Retroviruses, 2008. **24**(3): p. 379-88.
187. Wiesmann, F., et al., *The L76V mutation in HIV-1 protease is potentially associated with hypersusceptibility to protease inhibitors Atazanavir and Saquinavir: is there a clinical advantage?* AIDS Res Ther, 2011. **8**: p. 7.
188. Varghese, V., et al., *Prototypical Recombinant Multi-Protease Inhibitor Resistant Infectious Molecular Clones of Human Immunodeficiency Virus Type-1.* Antimicrob Agents Chemother, 2013.
189. Prabu-Jeyabalan, M., E. Nalivaika, and C.A. Schiffer, *How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease.* J Mol Biol, 2000. **301**(5): p. 1207-20.

190. Aydin, C., et al., *The interdomain interface in bifunctional enzyme protein 3/4A (NS3/4A) regulates protease and helicase activities.* Protein Sci, 2013. **22**(12): p. 1786-98.
191. Ozen, A., T. Haliloglu, and C.A. Schiffer, *Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope.* J Mol Biol, 2011. **410**(4): p. 726-44.
192. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.
193. Hsin, J., et al., *Using VMD: an introductory tutorial.* Curr Protoc Bioinformatics, 2008. **Chapter 5**: p. Unit 5.7.
194. Tenore, S.B. and P.R. Ferreira, *The Place of protease inhibitors in antiretroviral treatment.* (1678-4391 (Electronic)).
195. Bethell, R., et al., *Short communication: Phenotypic protease inhibitor resistance and cross-resistance in the clinic from 2006 to 2008 and mutational prevalences in HIV from patients with discordant tipranavir and darunavir susceptibility phenotypes.* AIDS Res Hum Retroviruses, 2012. **28**(9): p. 1019-24.
196. Ghosh, A.K., et al., *Design of HIV protease inhibitors targeting protein backbone: an effective strategy for combating drug resistance.* Acc Chem Res, 2008. **41**(1): p. 78-86.
197. Kozisek, M., et al., *Thermodynamic and structural analysis of HIV protease resistance to darunavir - analysis of heavily mutated patient-derived HIV-1 proteases.* Febs j, 2014. **281**(7): p. 1834-47.
198. *Prezista(R)[package insert]. Janssen Therapeutics, Division of Janssen Products, LP, Titusville, NJ 08560*. Available from: https://www.prezista.com/sites/default/files/pdf/us_package_insert.pdf.
199. Grant, B.J., et al., *Bio3d: an R package for the comparative analysis of protein structures.* Bioinformatics, 2006. **22**(21): p. 2695-6.
200. Girvan, M. and M.E. Newman, *Community structure in social and biological networks.* Proc Natl Acad Sci U S A, 2002. **99**(12): p. 7821-6.
201. Harte, W.E., Jr., et al., *Domain communication in the dynamical structure of human immunodeficiency virus 1 protease.* Proc Natl Acad Sci U S A, 1990. **87**(22): p. 8864-8.
202. Sethi, A., et al., *Dynamical networks in tRNA:protein complexes.* Proc Natl Acad Sci U S A, 2009. **106**(16): p. 6620-5.
203. Skalova, T., et al., *HIV-1 protease mutations and inhibitor modifications monitored on a series of complexes. Structural basis for the effect of the A71V mutation on the active site.* J Med Chem, 2006. **49**(19): p. 5777-84.
204. Meher, B.R. and Y. Wang, *Interaction of I50V mutant and I50L/A71V double mutant HIV-protease with inhibitor TMC114 (darunavir): molecular dynamics simulation and binding free energy studies.* J Phys Chem B, 2012. **116**(6): p. 1884-900.

205. Logsdon, B.C., et al., *Crystal structures of a multidrug-resistant human immunodeficiency virus type 1 protease reveal an expanded active-site cavity.* J Virol, 2004. **78**(6): p. 3123-32.
206. Gulnik, S.V., et al., *Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure.* Biochemistry, 1995. **34**(29): p. 9282-7.
207. Collins, J.R., S.K. Burt, and J.W. Erickson, *Flap opening in HIV-1 protease simulated by 'activated' molecular dynamics.* Nat Struct Biol, 1995. **2**(4): p. 334-8.
208. Weber, I.T. and R.W. Harrison, *Molecular mechanics analysis of drug-resistant mutants of HIV protease.* Protein Eng, 1999. **12**(6): p. 469-74.
209. Potempa, M., et al., *A Direct Interaction with RNA Dramatically Enhances the Catalytic Activity of the HIV-1 Protease In Vitro.* J Mol Biol, 2015. **427**(14): p. 2360-78.
210. Jabara, C.B., et al., *Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID.* Proc Natl Acad Sci U S A, 2011. **108**(50): p. 20166-71.
211. Jacobson, M.P., et al., *On the role of the crystal environment in determining protein side-chain conformations.* (0022-2836 (Print)).
212. Jacobson, M.P., et al., *A hierarchical approach to all-atom protein loop prediction.* (1097-0134 (Electronic)).
213. Altschul, S.F., et al., *Basic local alignment search tool.* (0022-2836 (Print)).
214. Auguie, B., *gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.2.1.* 2016.
215. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York, 2009.
216. Wickham, H., *Reshaping data with the reshape package.* Journal of Statistical Software
, 2007. **21**(12).
217. Warnes, G.R., Bolker, B., Lodewijk, B., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B. , *gplots: Various R Programming Tools for Plotting Data. R package version 2.17.0.* 2015.
218. Team, R.C., *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* 2015.
219. Plummer, M., Best, N., Cowles, K., and Vines, K., *CODA: Convergence Diagnosis and Output Analysis forMCMC.* R News, 2006. **6**: p. 7-11.
220. Dowle, M., Srinivasan, A., Short, T., Lianoglou with contributions from Saporta, S., Antonyan, E., , *data.table: Extension of Data.frame. R package version 1.9.6.* 2015

.

221. Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., *NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set.* Journal of Statistical Software, 2014. **61**(6): p. 1-36.
222. Csardi, G., Nepusz, T., *The igraph software package for complex network research.* InterJournal, Complex Systems, 2006
. **1695**.
223. Scarabelli, G. and B.J. Grant, *Mapping the structural and dynamical features of kinesin motor domains.* PLoS Comput Biol, 2013. **9**(11): p. e1003329.
224. Handel, A., R.R. Regoes, and R. Antia, *The role of compensatory mutations in the emergence of drug resistance.* PLoS Computational Biology, 2006. **2**(10): p. 1262-1270.
225. Schoustra, S.E., et al., *Reducing the cost of resistance; experimental evolution in the filamentous fungus Aspergillus nidulans.* Journal of Evolutionary Biology, 2006. **19**(4): p. 1115-1127.
226. Ma, C., et al., *Secondary Mutations Correct Fitness Defects in Toxoplasma gondii With Dinitroaniline Resistance Mutations.* Genetics, 2008. **180**(2): p. 845-856.
227. Bloom, J.D., L.I. Gong, and D. Baltimore, *Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance.* Science, 2010. **328**(5983): p. 1272-1275.
228. Bloom, J.D. and M.J. Glassman, *Inferring Stabilizing Mutations from Protein Phylogenies: Application to Influenza Hemagglutinin.* PLoS Computational Biology, 2009. **5**(4): p. e1000349.
229. Barber, L.J., et al., *Secondary mutations in BRCA2 associated with clinical resistance to a PARP inhibitor.* The Journal of Pathology, 2013. **229**(3): p. 422-429.
230. Scarabelli, G. and B.J. Grant, *Kinesin-5 allosteric inhibitors uncouple the dynamics of nucleotide, microtubule, and neck-linker binding sites.* Biophys J, 2014. **107**(9): p. 2204-13.
231. Young, M.A., et al., *Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation.* Cell, 2001. **105**(1): p. 115-26.
232. Cho, S.S., P. Weinkam, and P.G. Wolynes, *Origins of barriers and barrierless folding in BBL.* Proc Natl Acad Sci U S A, 2008. **105**(1): p. 118-23.
233. Pan, A.C., et al., *Molecular determinants of drug-receptor binding kinetics.* Drug Discov Today, 2013. **18**(13-14): p. 667-73.
234. Ozen, A., T. Haliloglu, and C.A. Schiffer, *HIV-1 Protease and Substrate Coevolution Validates the Substrate Envelope As the Substrate Recognition Pattern.* J Chem Theory Comput, 2012. **8**(2).
235. Ozen, A., W. Sherman, and C.A. Schiffer, *Improving the Resistance Profile of Hepatitis C NS3/4A Inhibitors: Dynamic Substrate Envelope Guided Design.* (1549-9626 (Electronic)).

236. Zazzi, M., A. Cozzi-Lepri, and M.C. Prosperi, *Computer-Aided Optimization of Combined Anti-Retroviral Therapy for HIV: New Drugs, New Drug Targets and Drug Resistance.* Curr HIV Res, 2016. **14**(2): p. 101-9.
237. Prosperi, M.C., et al., *Antiretroviral therapy optimisation without genotype resistance testing: a perspective on treatment history based models.* PLoS One, 2010. **5**(10): p. e13753.
238. Zazzi, M., et al., *Predicting response to antiretroviral treatment by machine learning: the EuResist project.* Intervirology, 2012. **55**(2): p. 123-7.
239. Wang, D., et al., *A comparison of three computational modelling methods for the prediction of virological response to combination HIV therapy.* Artif Intell Med, 2009. **47**(1): p. 63-74.