

INVESTIGATING THE CONFORMATIONAL DYNAMICS OF HIV-1 PROTEASE  
THROUGH MOLECULAR DYNAMICS

By

TERRY DWIGHT MCGEE JR.

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2013

© 2013 Terry Dwight McGee Jr.

To J.P. and Nora Williams

## ACKNOWLEDGMENTS

It has been through the love and sacrifice of others that I have been able to obtain so much in my life. I would like to commend Dr. Roitberg for his continuous efforts to make sure that I was equipped with the necessary tools to succeed. I appreciate most of all the patience and understanding that Dr. Roitberg has shown me, which has greatly helped to nurture my development.

I would like to acknowledge my committee who has not only been instrumental in my development as a scientist, but as a person as well. I appreciate your patience and thoughtful advice. I am grateful to Dr. Dunn for his guidance and the opportunity to work on a project of such magnitude. I appreciate Dr. Fanucci for your advice and positive reassurance of my work and progress. I would also like to extend my thanks to Dr. Bowers for challenging me in and out of the classroom and for your willingness to see me succeed. Dr. Horenstein, I appreciate how your classes strengthen my knowledge of biochemistry, which has had a profound impact on my research. I am indebted to Dr. Edwards for his invaluable advice and inspirational messages. You have been a valuable mentor.

I would like to acknowledge my parents Terry and Sylvia McGee who have been unbelievably understanding and encouraging in my pursuit of a higher education. Throughout my life, my parents have been a continuous stream of motivation and wisdom. I would like to thank the Roitberg Group for all their support, especially Natali Di Russo. Finally, I would like to acknowledge my wife Danielle McGee who has been nothing but supportive throughout this entire process. Thank you for being so understanding on the demands, which graduate school places on my time and life.

## TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS.....	4
LIST OF TABLES.....	9
LIST OF FIGURES.....	10
LIST OF ABBREVIATIONS.....	14
ABSTRACT.....	16
CHAPTER	
1 INTRODUCTION.....	17
1.1 HIV Worldwide Epidemic.....	17
1.2 HIV-1 Genome.....	17
1.3 Genetic Variability.....	19
1.4 HIV Life Cycle.....	20
1.5 HIV-1 Protease.....	22
1.6 HIV Protease as a Potential Drug Target.....	27
1.7 Protease Inhibitors.....	28
1.8 Drug Resistance.....	31
1.9 General Overview.....	34
2 METHODS.....	36
2.1 Molecular Dynamics.....	36
2.1.1 Molecular Dynamics Integrator.....	37
2.1.2 Langevin Dynamics.....	38
2.2 Molecular Mechanics.....	39
2.3 Solvent Models.....	40
2.3.1 Explicit Solvent.....	41
2.3.2 Implicit Solvent Models.....	42
2.3.2.1 Poisson-Boltzmann solvent model.....	42
2.3.2.2 Generalized Born solvent model.....	43
2.4 Steered Molecular Dynamics.....	44
2.5 Constant pH Methods.....	45
2.5.1 Constant pH Molecular Dynamics.....	45
2.5.2 Constant pH Replica Exchange Molecular Dynamics.....	46
2.6 Accelerated Molecular Dynamics.....	47
2.7 Docking.....	49
2.7 Molecular Mechanics Poisson Boltzmann Surface Area.....	50

3	MOLECULAR DYNAMICS SIMULATIONS INDICATE THAT THE CATALYTIC-DYAD OF HIV-1 PROTEASE IS MONO-PROTONATED.....	52
3.1	Introduction .....	52
3.2	Methods .....	54
3.2.1	General Setup.....	54
3.2.2	Constant pH Molecular Dynamics Protocol.....	55
3.2.3	Constant pH Replica Exchange Molecular Dynamics Protocol .....	56
3.2.4	General Analysis.....	57
3.3	Results.....	57
3.3.1	Constant pH Molecular Dynamics.....	57
3.3.2	Constant pH Replica Exchange Molecular Dynamics .....	58
3.3.4	Location of the proton in the mono-protonated state.....	59
3.4	Discussion.....	61
3.4.1	Apo (Free enzyme) .....	61
3.4.2	Cyclic Urea Inhibitors .....	61
3.4.3	Neutral Asymmetric Inhibitors .....	62
3.4.4	Cationic Inhibitor .....	63
3.4.5	Low-Barrier Hydrogen Bond.....	64
3.5	Conclusions.....	64
4	UNDERSTANDING HOW THE G48T/L89M MUTATIONS REDUCE THE EFFICACY OF THE HIV-1 PROTEASE INHIBITOR SAQUINAVIR .....	66
4.1	Introduction .....	66
4.2	Methods .....	68
4.2.1	General Setup.....	68
4.2.2	Steered Molecular Dynamics .....	68
4.2.3	Molecular Mechanics Generalized-Born Surface Area Calculations ...	69
4.2.4	General Analysis.....	69
4.3	Results and Discussions.....	69
4.3.1	Inhibitor Simulations.....	69
4.3.1.1	Flexibility.....	69
4.3.1.2	Hydrogen bonding analysis.....	73
4.3.1.3	MMGBSA calculations.....	74
4.3.1.4	Steered Molecular Dynamics.....	76
4.3.2	Apo Simulations .....	78
4.3.3	Drug Resistant Mechanism.....	80
4.4	Conclusions.....	82
5	EVALUATING THE EFFECTIVENESS OF SAQUINAVIR AND TIPRANAVIR ON HIV-1 SUBTYPE C WILD-TYPE AND THE DRUG RESISTANT N88D/L90M MUTANT PROTEASE .....	84
5.1	Introduction .....	84
5.2	Methods .....	86
5.2.1	General Setup.....	86

5.2.3	Molecular Dynamics.....	86
5.2.3	Molecular Mechanics Generalized-Born Surface Area Calculations ...	87
5.3	Results and Discussion.....	87
5.3.1	Saquinavir-bound MD Simulations.....	87
5.3.1.1	Flexibility.....	87
5.3.1.2	Hydrogen bonding analysis.....	90
5.3.1.3	MMGBSA calculations.....	91
5.3.2	Tipranavir-bound MD Simulations.....	93
5.3.2.1	Flexibility.....	93
5.3.2.2	Hydrogen bonding analysis.....	96
5.3.2.3	MMGBSA calculations.....	97
5.3.3	Saquinavir vs. Tipranavir.....	99
5.3.4	Drug Resistant Mechanism.....	99
5.5	Conclusions.....	101
6	INVESTIGATING WHAT EFFECTS THE STABILIZING MUTATIONS HAVE ON THE EPR MEASURED DISTANCE PROFILES.....	103
6.1	Introduction.....	103
6.2	Methods.....	105
6.2.1	General Setup.....	105
6.2.2	MTSL Parameters.....	106
6.2.3	Molecular Dynamics.....	107
6.3	Results and Discussions.....	107
6.3.1	Inhibitor-bound MD simulations.....	107
6.3.2	Apo MD Simulations.....	110
6.3.3	Comparing the N-N and Ca-Ca distances.....	113
6.4	Conclusions.....	114
7	PRELIMINARY STUDIES: EVALUATING THE EFFECTIVENESS OF ANTIRETROVIRAL PROTEASE INHIBITORS ON XMRV.....	116
7.1	Introduction.....	116
7.2	Methods.....	118
7.2.1	Completing XMRV.....	118
7.2.1	General Setup.....	118
7.2.2	Molecular Dynamics.....	119
7.2.3	Molecular Mechanics Poisson Boltzman Surface Area Calculations	119
7.3	Results and Discussion.....	120
7.3.1	Flexibility.....	120
7.3.2	Hydrogen Bonding Analysis.....	122
7.3.3	Molecular Mechanics Poisson Boltzmann Surface Area Calculations.....	124
7.3.4	Potential Inhibitor Interactions.....	125
7.4	Conclusions and Future Work.....	126
8	CONCLUSIONS AND FUTURE WORK.....	128

APPENDIX: HIV-1 PROTEASE SEQUENCES .....	130
LIST OF REFERENCES .....	131
BIOGRAPHICAL SKETCH .....	146

## LIST OF TABLES

<u>Table</u>	<u>page</u>
1-1	The twelve different substrates that HIV-1 protease must cleave in order to produce an infectious virion..... 27
4-1	Hydrogen bonding analysis of the wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir. .... 73
4-2	Results of MMGBSA calculations for the wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir. .... 75
5-1	Hydrogen bonding analysis of the wild-type and the mutant (N88D/L90M), subtype C, bound to saquinavir ..... 91
5-2	Results of the MMGBSA calculations for the wild-type and the mutant (N88D/L90M), subtype C, bound to saquinavir. All units are in kcal/mol ..... 92
5-3	Hydrogen bonding analysis of the wild-type and the mutant (N88D/L90M), subtype C, bound to tipranavir..... 96
5-4	Results of the MMGBSA calculations for the wild-type and the mutant (N88D/L90M), subtype C, bound to tipranavir. .... 97
6-1	Distance profiles of inhibitor-bound simulations..... 109
6-2	Distance profiles of the apo MD simulations..... 112
7-1	Hydrogen bond analysis XMRV (amprenavir)..... 123
7-2	Hydrogen bond analysis XMRV (pepstatin A) ..... 123
7-3	Hydrogen bond analysis XMRV (TL-3)..... 123
7-4	MMPBSA results for the inhibitors: amprenavir, pepstatin A, and TL-3..... 125
A-1	HIV-1 Protease Subtype B Sequence (Wild-type). .... 130
A-2	HIV-1 Protease Subtype B Sequence (LAI')..... 130
A-3	HIV-1 Protease Subtype C Sequence (Wild-type). .... 130
A-4	HIV-1 Protease MDR769 Sequence (Drug Resistant Construct)..... 130
A-5	HIV-1 Protease V6 Sequence (Drug Reistant Construct). .... 130

## LIST OF FIGURES

<u>Figure</u>	<u>page</u>
1-1 Structure of mature HIV virion.. .....	18
1-2 Schematic diagram of the HIV-1 Genome. ....	19
1-3 Genetic classification of HIV. HIV is classified in types, groups, subtypes/clades and circulating recombinant forms. ....	20
1-4 Schematic of the HIV viral life cycle.....	21
1-5 Schematic representation of HIV-1 protease.....	22
1-6 “Fireman’s grip”, a hydrogen bond network located in the active site of HIV-1 protease.. .....	23
1-7 Flap conformations of HIV-1 protease.....	24
1-8 General Acid/Base mechanism of HIV-1 protease .....	26
1-9 Transition state mimetics.....	29
1-10 Regions targeted by non-active site inhibitors .....	31
1-11 Per-residue mutation frequency of protease inhibitor naïve patients and patients who have been exposed to protease inhibitors (subtypes A, B, C, D, G, AE and AG.).....	32
1-12 Schematic of HIV-1 protease illustrating the location of the primary and secondary mutations. ....	33
2-1 Schematic representation of a TIP3P water molecule.....	42
2-3 Schematic representation of REMD simulations.....	47
2-4 Schematic representation of the normal and biased potential of AMD simulations.....	49
2-5 Thermodynamic cycle of MMPBSA.. .....	51
3-1 2D schematic of the protease inhibitors used in the Constant pH Molecular Dynamic Simulations. ....	53
3-2 Plot of Average total molecular protonation versus time for HIVPR complex KNI-272. CpHMD performed at pH 5.....	58
3-3 Schematic of the four different possible mono-protonated protonation states ....	60

4-1	HIV-1 protease, the colored spheres indicate locations on the protease the where wild-type and mutant differ. Position 48 (G48T) is colored yellow and position 89 (L89M) is colored orange. The catalytic aspartates are also depicted.....	66
4-2	Schematic representation of the HIV-1 protease inhibitor saquinavir. ....	67
4-3	Atomic Fluctuations of the wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir.. ....	70
4-4	Histogram of the RMSD of the C $\alpha$ atoms of the flap residues (43-58) of each monomer . MD simulations of the wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir. ....	71
4-5	Histogram of the RMSD of saquinavir. MD simulations of the wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir. ....	71
4-6	Histograms of the different distances in the active site of the wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir .....	72
4-7	Per-residue energy decomposition of the binding affinity of the wild-type and the mutant (G48T/L89M), subtype B bound to saquinavir. ....	75
4-8	Work Profiles from the SMD simulations for wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir. ....	77
4-9	Free energy profile of removing the protease inhibitor saquinavir from the active site cavity of the wild-type and mutant (G48T/L89M), subtype B. ....	78
4-10	Plot of the RMSD of wild-type subtype B, flap residues, 43-58, versus the closed, semi-open and wide-open conformations. MD simulations performed in the apo form.....	78
4-11	Plot of the RMSD of mutant (G48T/L89M) subtype B, flap residues, 43-58, versus the closed, semi-open and wide-open conformations. MD simulations performed in the apo form. ....	79
4-12	Atomic Fluctuations of the wild-type and the mutant (G48T/L89M), subtype B, apo.....	80
4-13	Schematic of the Phe53 pi-pi stacking interaction with the quinoline ring of saquinavir.. ....	81
5-1	Schematic of HIVPR subtype C.....	84
5-2	Schematic of the protease inhibitors saquinavir and tipranavir.....	85

5-3	Atomic Fluctuations of the wild-type and the mutant (N88D/L90M), subtype C, bound to saquinavir.....	88
5-4	Histogram of the RMSD of saquinavir. MD simulations of the wild-type and the mutant (N88D/L90M), subtype C, bound to saquinavir.....	89
5-5	Histogram of the RMSD of the flap residues (43-58) of each monomer. MD simulations of the wild-type and the mutant (N88D/L90M), subtype C, bound to saquinavir.....	89
5-6	Histograms of the different distances in the active site of the wild-type and the mutant (N88D/L90M), subtype C, bound to saquinavir.....	90
5-7	Per-residue energy decomposition of the binding affinity of the wild-type and the mutant (N88D/L90M), subtype C bound to saquinavir.....	92
5-8	Atomic Fluctuations of the wild-type and the mutant (N88D/L90M), subtype C, bound to tipranavir.....	94
5-9	Histogram of the RMSD of tipranavir. MD simulations of the wild-type and mutant (N88D/L90M), subtype C, bound to tipranavir.....	94
5-10	Histogram of the RMSD of the flap residue (43-58) of each monomer. MD simulations of the wild-type and mutant (N88D/L90M), subtype C, bound to tipranavir.....	95
5-11	Histograms of the different distances in the active site of the wild-type and mutant (N88D/L90M), subtype C, bound to tipranavir.....	95
5-12	Per-residue energy decomposition of the binding affinity of the wild-type and the mutant (N88D/L90M), subtype C bound to tipranavir.....	98
5-13	Schematic diagram of the Asp88/Asn88 interactions with Thr31 and Thr74 A) wild-type B) mutant (N88D/L90M).....	100
6-1	Schematic of HIV-1 protease. Highlighted are the spin labels at position 55 of each monomer.....	104
6-2	Schematic of the five torsional angles of the spin label MTSL.....	106
6-3	Histograms of the RMSD of HIV-1 protease subtype B and the protease inhibitor bound.....	108
6-4	Histogram of N—N distance distribution profiles of apo form HIVPR subtype B compared to when it is bound to protease inhibitors.....	110
6-5	Plot of the RMSF of the wild-type, subtype B (black), MDR769 (red), V6 (blue).....	111

6-6	Histogram of the N—N distance profiles. Wild-type subtype B, (black), MDR769 (red), V6 (blue) .....	112
6-7	Schematic of the average structures of HIVPR: wild-type subtype B (black), MDR769 (red), and V6 (blue) from spin-labeled MD simulations.....	113
6-8	Histogram of the 55-55 C $\alpha$ distance profile, wild-type subtype B (black), MDR769 (red), V6 (blue).....	114
7-1	Schematic of XMRV protease. Highlighted are the catalytic aspartates .....	117
7-2	Protease inhibitors: Amprenavir (HIV-1 protease), Pepstatin A (generic inhibitor of aspartic proteases), TL-3 (FIV protease) .....	117
7-3	Plot of the RMSF of XMRV bound to amprenavir, pepstatin A .....	120
7-4	Histogram of the RMSD of the protease inhibitors: amprenavir, pepstatin A and TL-3. ....	121
7-5	Histogram of the RMSD of the flap residues 48-66 of each monomer.....	122
7-6	Schematic representation of XMRV protease. Represented by spheres are the residues in the active site that potential inhibitors can target.....	126

## LIST OF ABBREVIATIONS

AIDS	Acquired Immunodeficiency Syndrome
aMD	Accelerated Molecular Dynamics
APV	Amprenavir
ASN	Asparagine
ASP	Aspartic Acid
ATZ	Atazanavir
cpHMD	Constant pH Molecular Dynamics
CRF	Circular Recombinant Form
DEER	Double Electron-Electron Resonance
DRV	Darunavir
EPR	Electron Paramagnetic Resonance
GB	Generalized Born
GLY	Glycine
FDA	Federal Drug Administration
FIV	Feline Immunodeficiency Virus
HIV	Human Immunodeficiency Virus
HIVPR	Human Immunodeficiency Virus Protease
IDV	Indinavir
ILE	Isoleucine
LPV	Lopinavir
MD	Molecular Dynamics
MET	Methionine
MM	Molecular Mechanics
MMPBSA	Molecular Mechanics Poisson Boltzmann Surface Area

MMGBSA	Molecular Mechanics Generalized Born Surface Area
MTSL	(1-Oxyl-2,2,5,5-Tetramethyl- $\Delta^3$ -Pyrroline-3-Methyl) Methanethiosulfonate
NFV	Nelfinavir
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
PB	Poisson Boltzmann
pH-REMD	pH Replica Exchange Molecular Dynamics
RMSD	Root Mean Squared Deviation
RMSF	Root Mean Squared Fluctuations
RTV	Ritonavir
SASA	Solvent Accessible Surface Area
SDSL	Site-Directed Spin Label
SQV	Saquinavir
THR	Threonine
TPV	Tipranavir
WHO	World Health Organization
UNAIDS	Joint United Nations on HIV/AIDS
VAL	Valine
XMRV	Xenotropic Murine Leukemia Virus-Related Virus

Abstract of Dissertation Presented to the Graduate School  
of the University of Florida in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy

INVESTIGATING THE CONFORMATIONAL DYNAMICS OF HIV-1 PROTEASE  
THROUGH MOLECULAR DYNAMICS

By

Terry Dwight McGee Jr.

December 2013

Chair: Adrian E. Roitberg  
Major: Chemistry

Human Immunodeficiency Virus (HIV), the virus that causes Acquired Immunodeficiency Syndrome (AIDS), is still one of the most prominent diseases in the world. One of the main targets in anti-retroviral therapy is the protease because of its role in the life cycle of the virus and inhibition would prevent the maturation and the spread of the virus to neighboring cells.

HIV-1 Protease (HIVPR) has two aspartates in the active site that are responsible for its catalytic ability. The protonation of the aspartates in the apo and bound forms of the protease has been topic of much debate. Using Constant pH Molecular Dynamics we investigate the protonation of the catalytic aspartates.

The mechanism of how drug pressure selected mutations, especially those not located in the active site, confer resistance is not well understood. Through our results we provide an explanation as to how they reduce the efficacy of protease inhibitors. The results of this work could offer valuable insight on ways to improved protease inhibitors.

## CHAPTER 1 INTRODUCTION

### 1.1 HIV Worldwide Epidemic

Human Immunodeficiency Virus (HIV) is the causative agent of Acquired Immunodeficiency Disease Syndrome (AIDS).<sup>1</sup> As of 2011, the Joint United Nations Program on HIV/AIDS (UNAIDS) has projected that there are approximately 34 million people living with HIV. To date, there is no cure or vaccine for HIV, however recent advances in antiretroviral therapy have improved the life expectancy of those infected. Consequently, the number of AIDS-related deaths has decreased from 2.3 million in 2005 to 1.7 million in 2011 according to the World Health Organization (WHO). A majority of those deaths were in Sub-Saharan Africa—this startling fact can be attributed to the inabilities of these countries to afford antiretroviral treatment, social stigmatism, etc. There has been a decline in new HIV cases because only 2.5 million people were newly infected with HIV in 2011, which is down from 3.2 million in 2001 according to data provided by to the UNAIDS.

### 1.2 HIV-1 Genome

HIV is a retrovirus and is further classified as a lentivirus, due to its long incubation period. HIV is approximately 120 nm in diameter and possesses a spherical shape, Figure 1-1.<sup>2</sup> It is comprised of two copies of single stranded RNA that is encapsulated by the viral protein p24. Its genome, Figure 1-2, is roughly 9.4 Kb in length. The group-specific antigen (*gag*) gene encodes the structural proteins matrix (MA), capsid (CA), and nucleocapsid (NC).<sup>3</sup> The envelope proteins g41 and gp120, are responsible for the viral tropism and are encoded by the envelope (*env*) gene. The polymerase (*pol*) gene encodes for the three enzymes: reverse transcriptase, integrase

and protease. The other genes *rev*, *nef*, *vpr*, *vif*, *vpu*, and *tat* encode for the six accessory proteins. Nef has two roles: decreasing the number of CD4 receptors located on the surface of infected lymphocytes and enhancing viral replication.<sup>4,5</sup> Rev is responsible for controlling the expression of HIV proteins by governing the rate at which mRNAs are exported from the nucleus to cytoplasm.<sup>6</sup> The Tat enhances transcription of integrated proviral DNA.<sup>7</sup> Vpr plays a critical regulatory role during replication in the early stages of the life cycle of the virus.<sup>8,9</sup> The exact function of Vif is not known but has been implicated to be a factor in the viral replication.<sup>10,11</sup> HIV strains that have mutations in Vif do not replicate as efficiently and have been linked to low viral loads when compared to the wild-type strains.<sup>10,12</sup> Vpu governs the release of the virus from cell surface of the infected host and degradation of CD4 in the endoplasmic reticulum.<sup>10,13</sup>

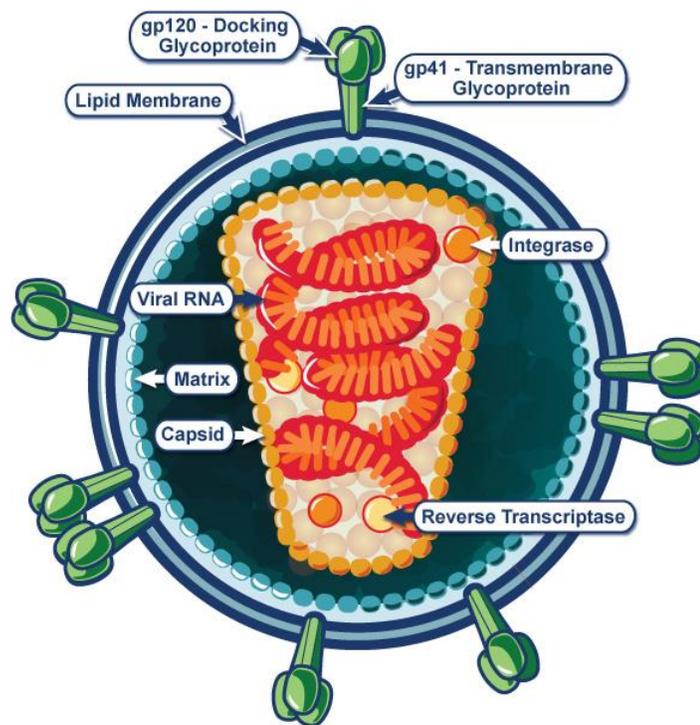


Figure 1-1. Structure of mature HIV virion. Image courtesy of National Institute of Allergy and Infectious Diseases.

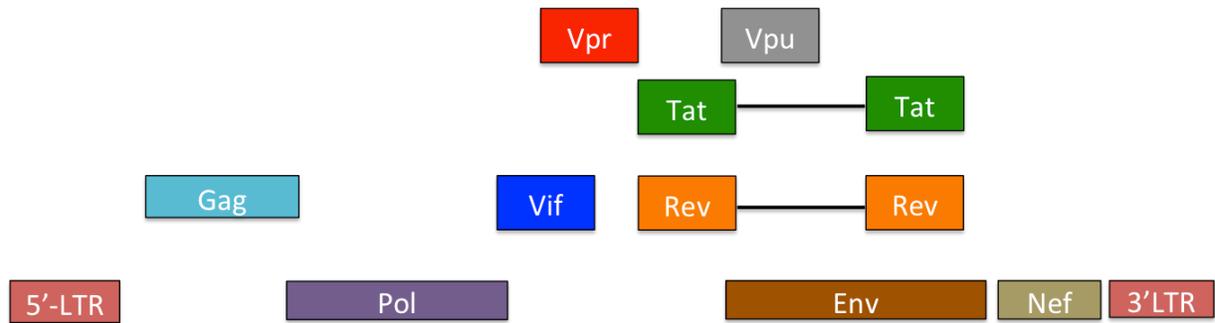


Figure 1-2. Schematic diagram of the HIV-1 Genome.

### 1.3 Genetic Variability

HIV has a high genetic variability, Figure 1-3, and it can be divided into two distinct types: HIV-1 and HIV-2. HIV-1, which accounts for the majority of infections, can be divided into groups, M “main”, N “non-M/non-O”, and O “outlier” and the newly discovered group, P.<sup>14-17</sup> Group M is the largest and is composed of subtypes or clades, A-D, F-H, J, K and circulating recombinant forms. The subtype relationships and classifications were constructed by performing sequence alignments of the gag, and env regions.<sup>14,18,19</sup>

Subtype A is predominantly in West and Central Africa. Subtype B is generally found in North and South Americas, Europe and Australia. Subtype C is essentially in Southern and East Africa and throughout Eastern Asia. Subtype D is limited mainly to Central Africa. Subtype F is primarily in South America. Subtype H is mostly in Central Africa and subtype J in Central America. Lastly, subtype K is mainly in the countries Congo and Cameroon.

Persons infected with HIV are generally infected with subtypes A-D. Moreover, subtype C accounts for the highest percentages of those with HIV.<sup>20</sup> However, Subtype D has been credited with being the deadliest and those infected with this subtype

develop AIDS at a much faster rate and have a higher mortality rate when compared to the other subtypes.<sup>21</sup>

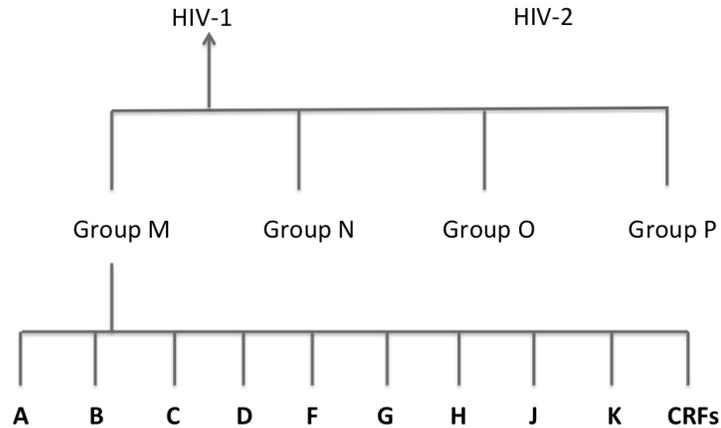


Figure 1-3. Genetic classification of HIV. HIV is classified in types, groups, subtypes/clades and circulating recombinant forms.

#### 1.4 HIV Life Cycle

In general, the surface glycoproteins of enveloped viruses play a critical role in viral invasion of host cells. The HIV envelope glycoprotein (gp) consists of two non-covalent subunits gp120 and gp41. First, gp120 is responsible for viral tropism by binding to target-cell receptors and gp41 directs fusion between cellular and viral membranes.<sup>22,23</sup> Gp120 binds to the CD4 receptor and the CCR5 or CXCR4 co-receptors, which are members of the chemokine family.<sup>22,24</sup> Macrophage (M-tropic) strains of HIV-1 use the CD4 and CCR5 receptors to invade the host cell.<sup>25</sup> In contrast, T-tropic strains use CD4 and CXCR4 receptors to gain entry into the cell.<sup>25</sup> In addition to CCR5 and CXCR4, other co-receptors have been identified that the virus may use to invade a host cell.<sup>26,27</sup> These receptors are present in such cells as T lymphocytes and macrophages.

Gp41 undergoes a conformational change to a fusion-active state, when gp120 binds to the host cell receptor.<sup>28</sup> This conformational change allows the virus to inject its

viral genome into the host through fusion.<sup>28</sup> Soon after entry, the process of reverse transcription occurs—the enzyme reverse transcriptase transcribes a complementary, double-stranded DNA from a single-stranded RNA genome. The process of reverse transcription is prone to error. The high frequency of mutations is a direct result of having an average error of  $3 \times 10^{-5}$  per nucleotide base per cycle.<sup>29</sup>

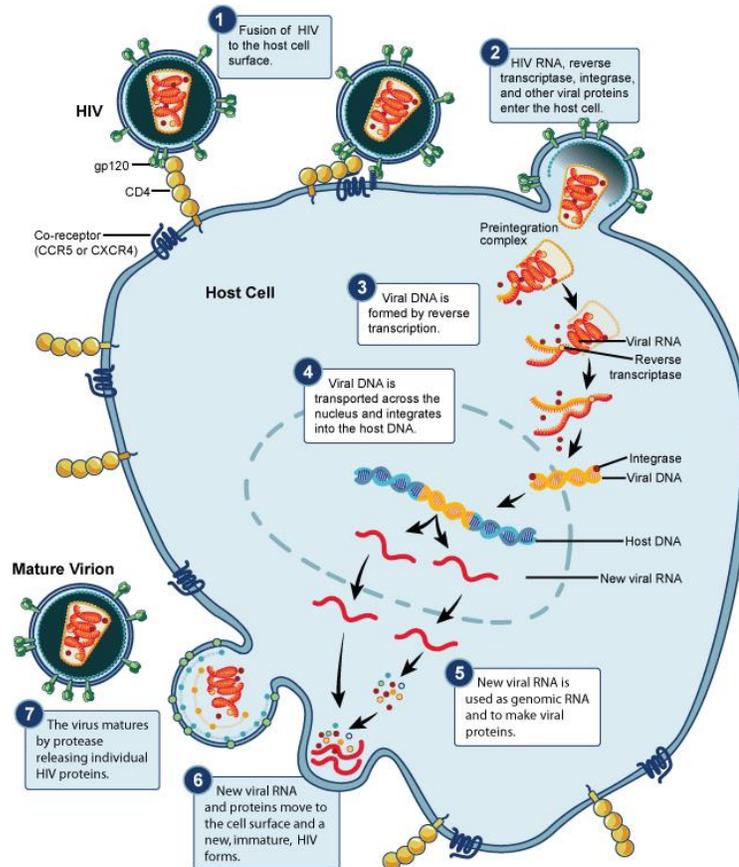


Figure 1-4. Schematic of the HIV viral life cycle. Image courtesy of National Institute of Allergy and Infectious Diseases.

The newly formed DNA is transported to the nucleus of the cell where it is integrated into the host DNA by the enzyme, integrase. Inclusion of the virus genetic material into the host cell genome results in the viral DNA becoming a provirus.<sup>30</sup> The virus can remain dormant for years in this state.<sup>31</sup> Upon activation<sup>32</sup>, the provirus uses the host enzyme RNA polymerase to transcribe the viral mRNA. The mRNA must be

exported from the nucleus and remain intact—the rev protein binds to the un-spliced mRNA and transports it from the nucleus.<sup>30,33</sup> The last and final steps are assembly and budding. The protease enzyme is responsible for cleaving the polypeptide chain into individual proteins, which promotes assembly and is vital for virus maturation.<sup>34,35</sup> The newly formed virus buds from the host cell and can now infect a new cell.

### 1.5 HIV-1 Protease

A protease is an enzyme that directs proteolysis and cleaves the peptide bonds that link amino acids to form polypeptide chains. Proteases are divided into several groups: serine, threonine, aspartyl, cysteine, metallo and glutamic.<sup>10</sup> HIV protease, Figure 1-5, is classified as an aspartyl protease because it has two aspartic residues in the active site and substitution of these aspartic residues leads to a loss of catalytic ability.<sup>36</sup>

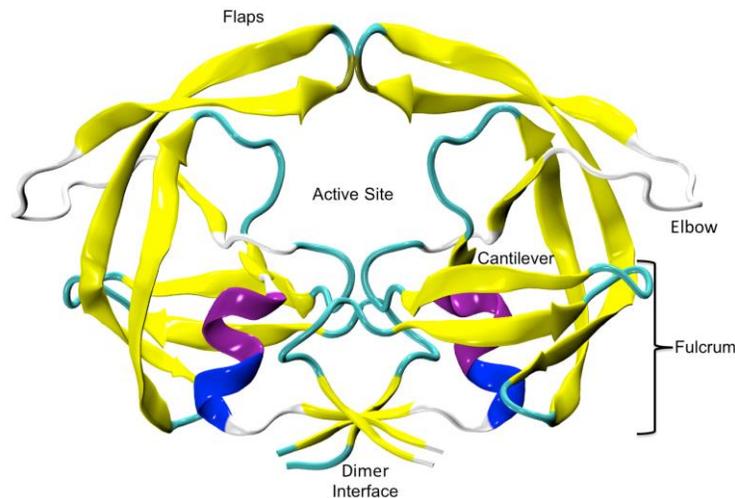


Figure 1-5. Schematic representation of HIV-1 protease, PDB 1HHP. Flap residues 43-58, flap elbow residues 37-42, active site residues 25-27, cantilever 59-75, fulcrum residues 10-23, dimer interface residues 1-4, and 96-99

The three-dimensional structure of HIV protease Protein Data Bank (PDB ID 2HVP<sup>37</sup>, 3HVP<sup>38</sup>, 4HVP<sup>39</sup>) was not determined until 1989. To date, hundreds of crystal

structures of HIVPR in various forms have been solved. HIV-1 Protease (HIVPR) is composed of two identical monomers each containing 99 amino acid residues that give rise to a symmetrical homodimer. Each monomer has a small  $\alpha$  helix and four-stranded  $\beta$  sheet that are formed by the N and C terminus strands.<sup>38</sup>

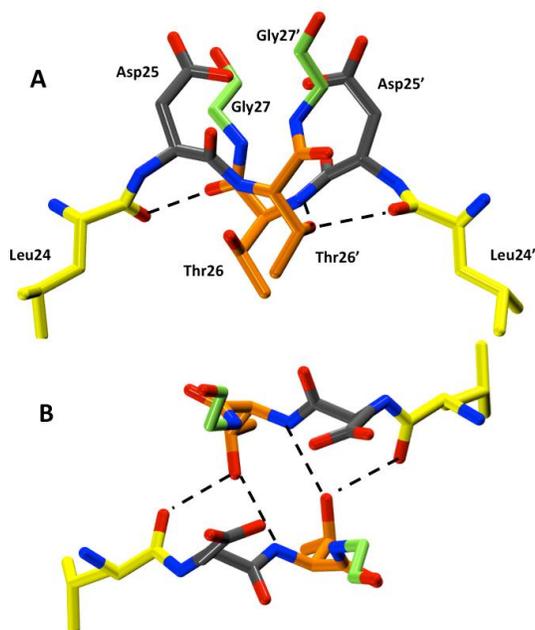


Figure 1-6. “Fireman’s grip”, a hydrogen bond network located in the active site of HIV-1 protease. A) Front view B) Side view. Leu24 colored yellow, Asp25 colored gray, Thr26 colored orange, Gly27 colored green.

The active site of HIVPR is highly conserved—it is located at the interface of the two monomers where each monomer contributes the sequence Asp25-Thr26-Gly27. The three-residue motif is referred to as the catalytic triad. The active site is held together, through a hydrogen-bond network, known as the “fireman’s grip”, Figure 1-6.<sup>40</sup> The amide group of Thr26 donates a hydrogen bond to the Thr26 side-chain hydroxyl group located on the opposing monomer—Thr26 donates another hydrogen bond to the carbonyl group of Leu24 of the opposite monomer. Previous studies have shown that

Thr26 has an important role in maintaining the dimer complex—the mutation of Thr26 would result in the monomer being favored as opposed to the dimer.<sup>41,42</sup>

The active site is covered by two  $\beta$ -hairpins referred to as flaps (residues 43-58). The flaps undergo large conformational changes of backbone phi and psi angles in order for a substrate or inhibitor to access the active site.<sup>43</sup> The flaps are in a closed conformation when an inhibitor/substrate is bound—the flaps are stabilized by a water molecule (WAT301) that forms a hydrogen bond with the substrate/inhibitor and amide group of Ile50. The reaction pathways of substrate binding and product released have been investigated using umbrella sampling<sup>44</sup> and coarse-grained molecular dynamics<sup>45</sup>. Trylska *et al.*<sup>45</sup> suggested that the flaps do not need to open in order for the products to be released—the products can instead slide out from the binding cleft to the sides of the enzyme.

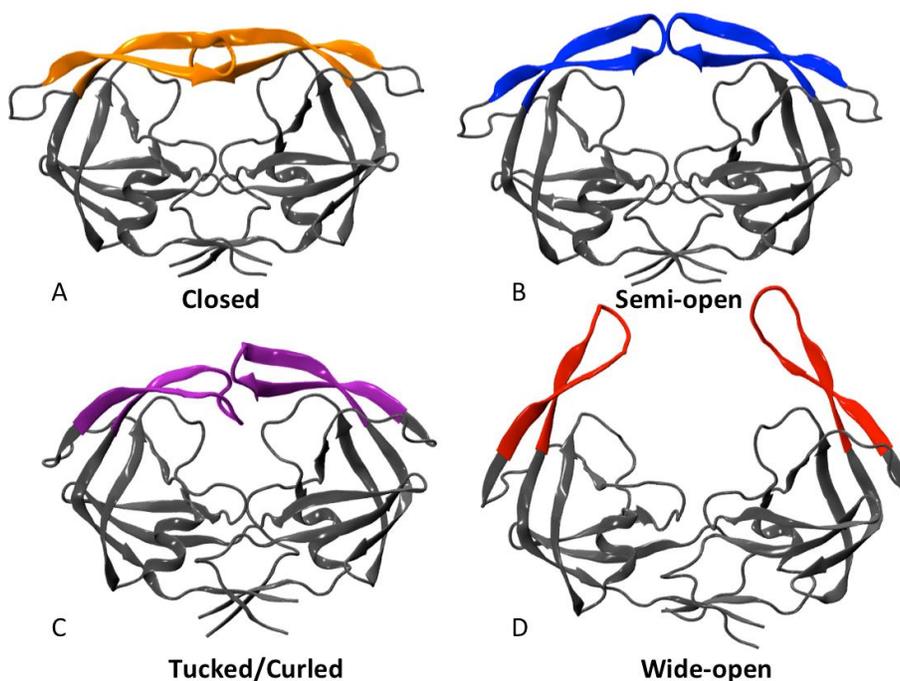


Figure 1-7. Flap conformations of HIV-1 protease. A) Closed conformation (orange) B) Semi-open conformation (blue), C) Tucked/Curled conformation (purple) D) Wide-open conformation (red)

In the apo form of HIVPR, the flaps are either in the closed, wide-open, tucked/curled, or semi-open conformation, Figure 1-7. Ishima *et al.*<sup>43</sup> presented a working model of the flap-opening mechanism for the free HIVPR that requires the flaps to transition from semi-open conformation in order to reach the wide-open conformation. Access to the active site even in the semi-open conformation is still blocked by the flaps and does not allow adequate space for a substrate/ligand to bind. Flap openings as large as 32 angstroms, measuring distance between the Ile50 residue of each monomer, have been observed in computational studies performed on HIVPR.<sup>46</sup> NMR as well as MD studies have concluded that the flaps are predominantly in a semi-open conformation in the apo form HIVPR<sup>46,47</sup> The “handedness” of the flaps are reversed when flaps transition between the semi-open to closed conformations.

NMR studies have indicated that the entire flaps of the apo protease move on the 100  $\mu$ s time scale—the flap tips (residues 49-52) undergo many rapid fluctuations on the sub nanosecond time scale.<sup>43</sup> The increased flexibility of the flap tips when compared to the entire flap is attributed to the flap tips glycine-rich sequence. A 10 ns MD study performed by Scott and Schiffer<sup>48</sup> on HIVPR, suggested that the curling of the flap tips elicits the flaps to “fully” open. However, Meagher *et al.*<sup>49</sup> claim the conclusions of Scott and Schiffer<sup>48</sup> could be a result of improper equilibration of the flaps. Li *et al.*<sup>50</sup> concluded that the sub nanosecond local motion of the flap tips were essential to the global conformational changes of the flap tips, thus confirming Scott and Schifer<sup>48</sup> previous notion.

The HIV protease cleaves its substrate through a general acid/base mechanism. HIVPR must perform twelve proteolytic reactions in order to produce a mature virion.<sup>51</sup>

Kinetic studies performed on HIVPR with its substrates observed a bell-shape curve when the rate as a function of pH was measured, indicating that one aspartate acts as an acid and the other as a base.<sup>52,53</sup> The mechanism proposed by Meeks *et al.*<sup>52</sup>, Figure 1-8, is that Asp25' residue accepts a proton (acting as a base) from a water molecule. The newly formed nucleophile is very potent and attacks the carbonyl of the substrate. The carbonyl oxygen of the substrate is stabilized by accepting a proton from the Asp25 residue (acting as an acid). The Asp25 accepts proton from the oxygen as it begins to reform the double bond with the carbon. Simultaneously, the carbonyl is reformed as the N-C bond begins to elongate and eventually break as the nitrogen accepts a proton from Asp25'. The catalyst is regenerated and the product exits the active site.

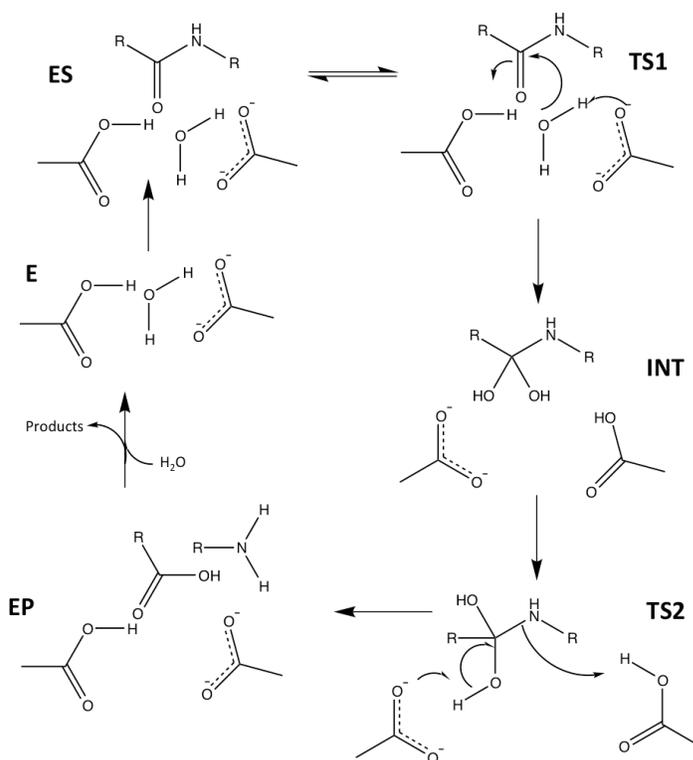


Figure 1-8. General Acid/Base mechanism of HIV-1 protease

Table 1-1. The twelve different substrates that HIV-1 protease must cleave in order to produce an infectious virion. The “ / “ is indicative of where HIVPR cleaves the substrate. HXB2 reference

Location of Proteolytic Reactions	Amino Acid Sequence
gag polyprotein cleavage sites	
MA/CA	VSQNY / PIVQN
CA/p2	KARVL / AEAMS
p2/NC	TSAIM / MQRGN
NC/p1	ERQAN / FLGKI
p1/p6 <sup>gag</sup>	RPGNF / LQSRP
gag-pol polyprotein cleavage sites	
NC/TFP	ERQAN / FLREN
TFP/p6 <sup>pol</sup>	ENLAF / LQGKA
p6 <sup>pol</sup> /PR	VSFNF / LQGKA
PR/RTp51	CTLNF / PISPI
RT/RT/p66	GAETF / YVDGA
RTp66/INT	IRKVL / FLDGI
Nef protein cleavage sites	
Nef	AACAW / LEAQE

## 1.6 HIV Protease as a Potential Drug Target

First generation antiretroviral drug therapy involved developing inhibitors for reverse transcriptase. The efficacy of reverse transcriptase inhibitors was proven to be limited due to toxicities caused by interference with the human cell metabolism.<sup>54</sup> This ignited the need to search for other alternative drug targets.

It was not until the mid 1980s that momentum for the development of HIVPR inhibitors began to take form. Several landmark studies, revealed vital information that helped solidify the protease as a promising drug target. When the complete nucleotide sequence was determined, the results suggest that HIV encoded a protease<sup>55</sup>—a

mechanistic relationship of HIVPR with other aspartic proteases had been proposed.<sup>56</sup> Kramer *et al.*<sup>57</sup> were the first to show that protease was essential in cleaving the gag and gag-pol polyproteins for the HIV virus specifically. Sanchez-Pescador *et al.*<sup>58</sup> determined two cleavage sites of HIVPR in the gag protein (Met/Met and Tyr/Pro). Because HIVPR cleaves bonds between Tyr/Phe and Pro residues and no other human enzymes show this specificity—this affords an opportunity to develop protease inhibitors that are very selective and nontoxic to the host cell.

## 1.7 Protease Inhibitors

Since HIVPR offered a promising new drug target in the antiretroviral drug therapy, the next issue that had to be addressed was how to design an inhibitor. Only competitive inhibitors were considered—the inhibitor competes with the substrate for binding to the active site. It was important that inhibitors be highly selective to HIVPR and have minimum affinity to the other human aspartic proteases such as pepsin, renin, gastricsin and cathepsins D and E. Early inhibitors were developed using the “transition-state mimetic concept”<sup>59</sup>, which had much success in the design of other aspartic proteases such as renin.<sup>54</sup> The “transition state mimetic concept”<sup>59</sup>, when applied to aspartic proteases, is where the scissile carbonyl of an oligopeptide substrate is replaced by a nonhydrolyzable analogue with tetrahedral geometry.<sup>60</sup> Many different transition state mimetics were proposed based on the Phe-Pro and Tyr-Pro cleavage sites in order to ensure that the potential inhibitor would be potent as well as selective. Transition state mimetics such as: reduced amide, ketomethylamine, phosphinate, hydroxyethylene and hydroxyethylamine, were evaluated as possible inhibitors of HIVPR.<sup>60–62</sup> Kinetic studies revealed that the hydroxyethylamine and hydroxyethylene mimetics were the best candidates for being potential HIVPR inhibitors.<sup>54,62</sup>

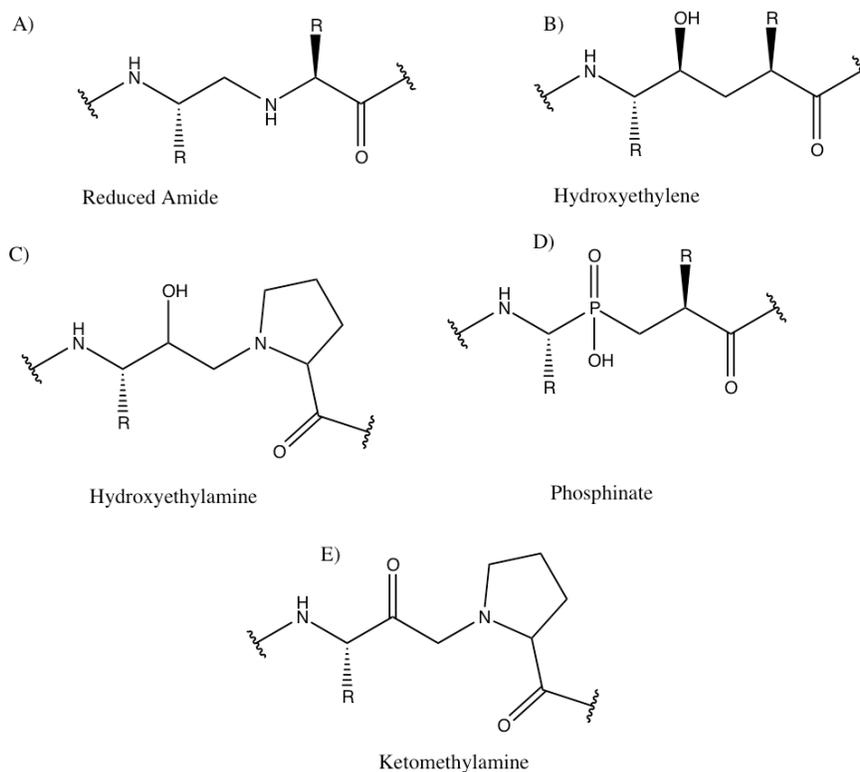


Figure 1-9. Transition state mimetics. A) Reduced Amide B) Hydroxyethylene C) Hydroxyethylamine D) Phosphinate E) Ketomethylamine

In December of 1995, saquinavir, marketed as Invirase, became the first FDA approved HIV protease inhibitor.<sup>54</sup> The structure of saquinavir is a derivative of the hydroxyethylamine mimetic. To date, there are ten FDA-approved protease inhibitors and they are a mixture of peptidomimetic and non-peptidomimetic drugs.

The potency of an inhibitor is measured by the free energy of binding, which is composed of both enthalpic and entropic contributions. The earlier generation HIVPR inhibitors were designed to be rigid—the dominant driving force for binding was entropically driven with unfavorable or small favorable change enthalpy change.<sup>63,64</sup> The large entropic gains were a result of desolvation of the non-polar groups, which releases water to the bulk solvent, increasing the solvation entropy and ligand being pre-shaped

to geometry of the active site, minimizing the loss of conformational entropy.<sup>63</sup> Newer developed inhibitors, such as darunavir, were found to be instead enthalpically driven and are less susceptible to drug resistance.<sup>65</sup>

As mentioned earlier all FDA approved HIVPR inhibitors bind to the active site of HIVPR. These inhibitors are susceptible to drug resistance due to drug pressure selected mutations on HIVPR. In order to rectify this problem, non-competitive inhibitors have been proposed as a means to increase the efficacy of the current HIVPR inhibitors. Two different types of non-competitive inhibitors have been proposed: inhibitors that prevent dimerization<sup>66,67</sup> and inhibitors that perturb the dynamics of the protein, more specifically, the dynamics of the flaps.<sup>68-70</sup>

The protease dimer in addition to the “fireman’s grip” is stabilized by a four stranded antiparallel  $\beta$ -sheet that comprises both N and C termini of each monomer. Residues 1-4 and 96-99 of the terminal residues account for over half the hydrogen bonds along the dimer interface<sup>71,72</sup>—75% of the dimerization free energy is attributed to the contacts in N and C termini.<sup>73</sup> The inhibitors that bind and prevent dimerization, target the N and C terminal strands of each monomer residues 1-4 and 96-99. This region is highly conserved thus making drug resistance minimum.<sup>74</sup> Furthermore, the monomer form the HIVPR is catalytically inactive.<sup>66</sup> The first dimerization inhibitors were derived from the amino acid sequences of N and C terminus strands.<sup>75,76</sup> Dimerization inhibitors have since evolved to N-terminally cross-linked, interfacial protease peptides that were devised to mimic the protease dimerization interface.<sup>77</sup>

Computational studies have identified two druggable regions: “exo site”, and the “outside-top/ of the flap” of the protease that could accommodate an allosteric

inhibitor.<sup>69,70</sup> The “exo site” includes residues Gly16, Gly17, Gln18, Lys14 and Leu63 and “outside-top/ of the flap” is composed of Trp42, Pro44, Met 46, Lys55, Val56 and Arg57.<sup>69</sup> The non-active site allosteric inhibitors can bind to the apo form of HIVPR trapping the flaps in conformations such as wide-open or closed thus rendering the protease catalytically inactive. These non-active site allosteric inhibitors can also bind the protease-active-site-inhibitor complex and synergistically improve the binding affinity of an active site protease inhibitor.<sup>69</sup>

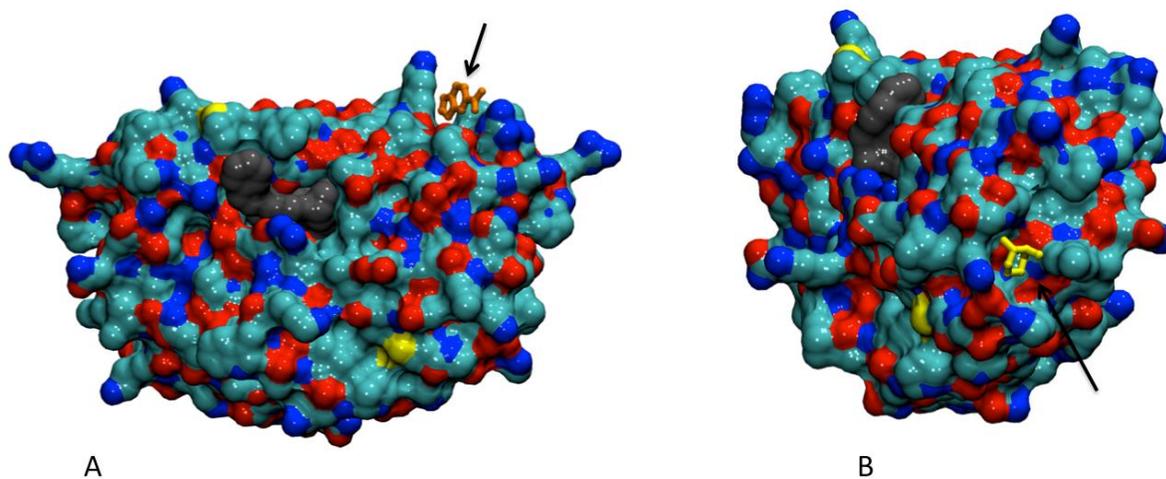


Figure 1-10. Regions targeted by non-active site inhibitors A) “outside/top of the flaps” fragment 1F1 bound colored orange PDB 3KFR B) “exo-site” PDB 3KFN fragment 4DX bound colored yellow

## 1.8 Drug Resistance

Drug resistance has become a dire problem and reduces the efficacy of protease inhibitors. As discussed previously, the replication process of HIV is very error prone, which provides the foundation for drug resistance. Without exposure to any protease inhibitors, patients will naturally develop polymorphisms in HIVPR. However, when patients are exposed protease inhibitors, drug-pressured selected mutations will

develop. When comparing the rates of mutations versus naïve-patients, (have not been exposed to any protease inhibitors) and protease-inhibitor-treated patients (have been exposed to protease inhibitors) residues in the N and C termini, 1-10, 96-99, active site 25-27 and the flaps tips 49-52 are conserved, Figure 1-11.<sup>78-80</sup>

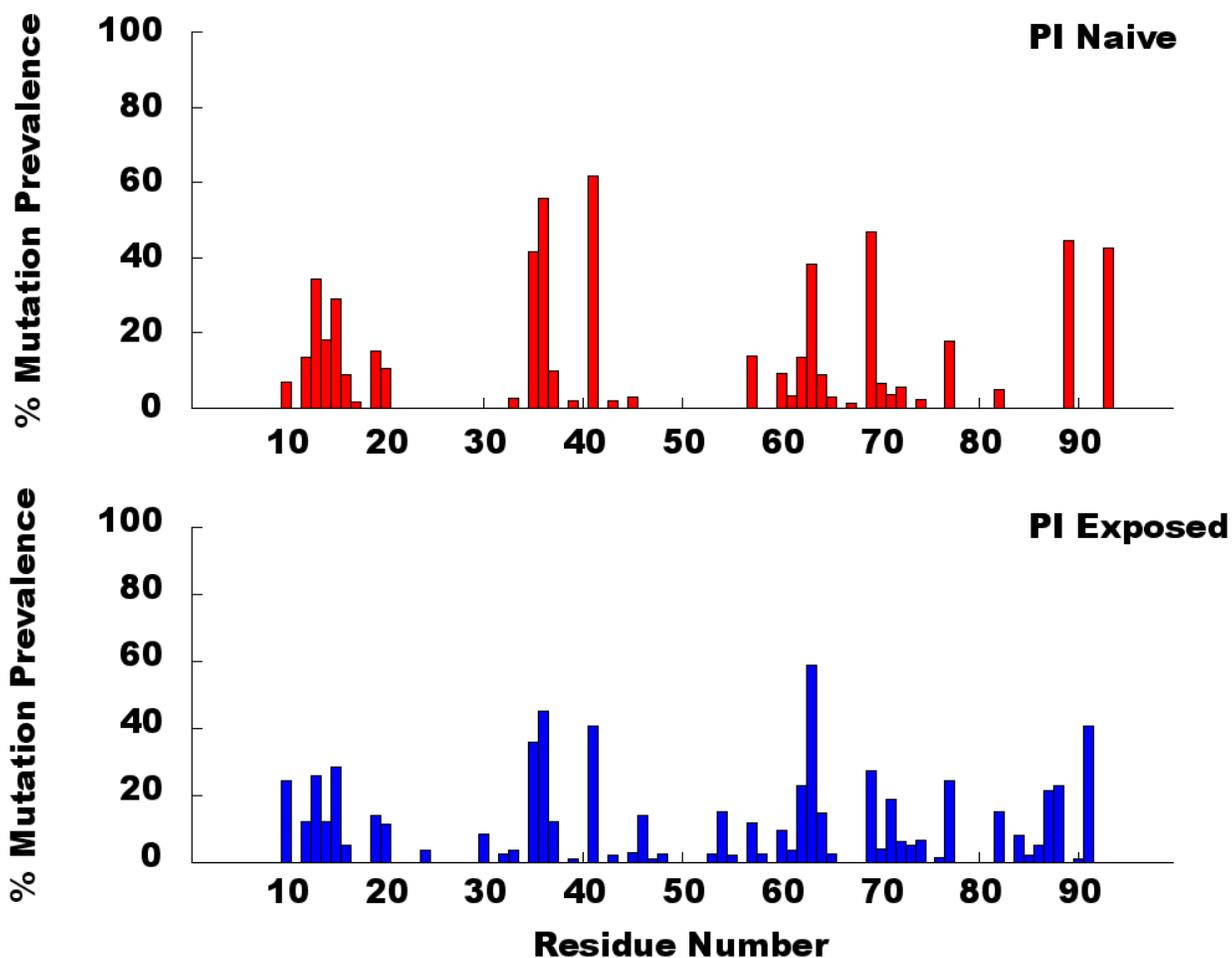


Figure 1-11. Per-residue mutation frequency of protease inhibitor naïve patients and patients who have been exposed to protease inhibitors (subtypes A, B, C, D, G, AE and AG.)

The drug-pressure selected mutations are the result of viral replication not being fully suppressed in the presence of an inhibitor. Drug resistance can occur not only from the presence of protease inhibitors but from mutations in the cleavage sites of the

protease.<sup>81,82</sup> The mutations of HIVPR are classified as either primary or secondary mutations as depicted in Figure 1-12. Primary mutations are often residues in the active site that come into direct contact with the inhibitor. These mutations elicit resistance by reducing the number of interactions with competitive inhibitors by altering the geometry of the active site.<sup>83</sup> Consequently, primary mutations reduce the catalytic efficiency of the enzyme.<sup>84</sup> Secondary mutations, which are not always a result of protease inhibitor therapy<sup>85</sup>, are thought to restore the catalytic efficiency of the enzyme while still maintaining resistance to the inhibitor.<sup>86</sup> Furthermore, these mutations often do not directly interact with the inhibitor yet the drug-resistance is still maintained. As of now, fifteen sites for major mutations and nineteen for minor mutations have been determined for all the currently approved FDA protease inhibitors.<sup>87</sup>

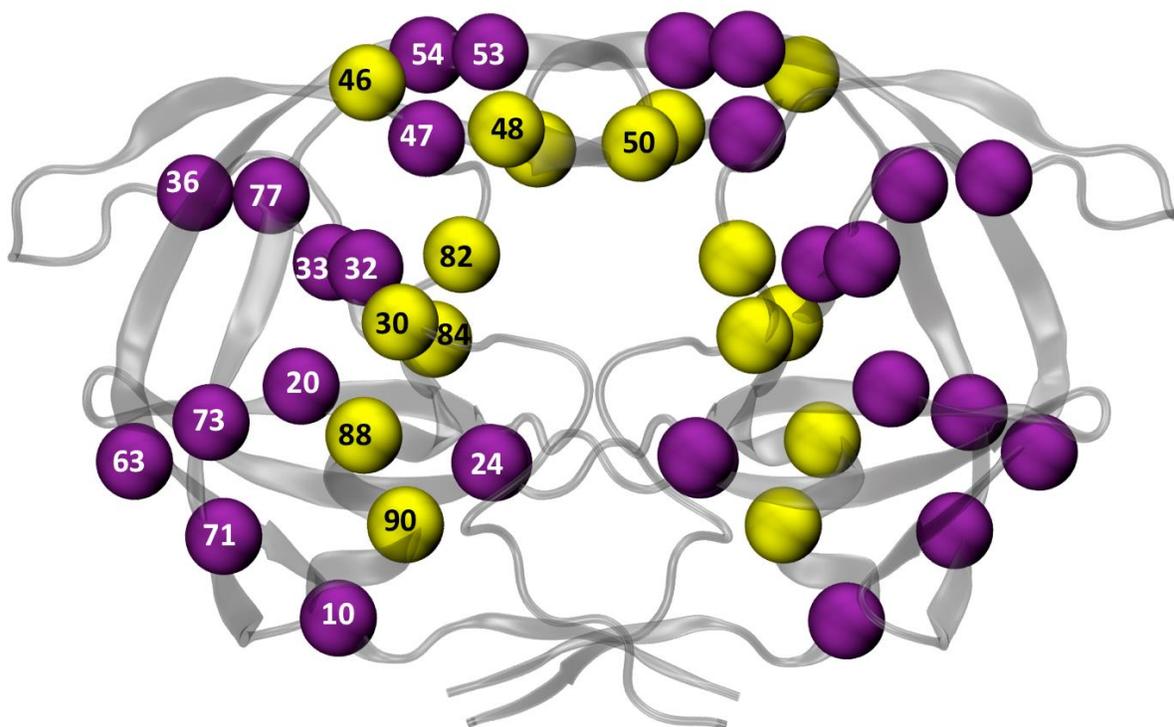


Figure 1-12. Schematic of HIV-1 protease illustrating the location of the primary (yellow) and secondary (purple) mutations.

King and Schiffer *et al.*<sup>88</sup> proposed a mechanism describing how the primary drug-pressured selected mutations evolve in the protease. The mechanism entails that any inhibitor that protrudes beyond the substrate envelope causes mutations in the specific residues that it directly contacts.<sup>88</sup> The residues beyond the substrate envelope that the inhibitor interacts with directly are more important for the inhibitor binding than the substrate binding and this explains why HIVPR can tolerate substitutions at those positions. The residues in the active site that directly interact with inhibitor(s) and mutations at these sites often result in drug resistance for one or more protease inhibitors are D30, I47, G48 and I50, V82 and I84.

Although several mechanisms<sup>83,89,90</sup> have suggested how non-active site mutations confer resistance, it is still not well defined. The hydrophobic sliding mechanism, which was developed through MD simulations of the apo form HIVPR, involves nineteen non-active site hydrophobic residues. Each of these residues in the hydrophobic core slide by each other exchanging, one hydrophobic van der Waal contact for another, with minimum energy penalty, while maintaining hydrogen bonds that are important to the stability of the protease.<sup>89</sup> Non-active site mutations have been implicated in perturbing dynamics of the protease by decreasing the rate of flap closure causing the flaps of the protease to be more open and bind the substrate over the inhibitor.<sup>83</sup> The last mechanism suggests non-active mutations alter the fractional occupancy of the equilibrium conformational sampling.<sup>90</sup>

## **1.9 General Overview**

This body of work involves using MD as a means to study HIVPR. The aims of this work are: determine the protonation of the catalytic aspartates, explain the mechanism of active/non-active site mutations, determine what effects if any the

stabilizing mutations have the on EPR measured distance profiles and make suggestions on developing an inhibitor for XMRV protease.

## CHAPTER 2 METHODS

### 2.1 Molecular Dynamics

The first MD simulation was published by Alder and Wainwright, in which they reported on the interactions of hard spheres.<sup>91</sup> Bovine pancreatic trypsin inhibitor became the first MD simulation involving a protein, and it was instrumental in paving the way for the application of MD to biological systems. Since then Molecular Dynamics has been beneficial in providing description on the atomic level, validating experimental results, structure refinement, drug-design, enzymatic reactions, etc. Currently with the increase of computing power, MD simulations can reach time scales of microseconds.

MD is a deterministic method meaning that the state of the system at any future time can be projected from its current state. Molecular Dynamics is based on Newton's laws of motion, in particular his second law that states  $F=ma$ , where  $F$  is the force exerted on the particle,  $m$  is its mass and  $a$  is its acceleration.

$$\frac{d^2x_i}{dt^2} = \frac{Fx_i}{m_i} \quad (2-1)$$

Equation 2-1 describes the motion of particle of mass  $m_i$  along a single coordinate  $x_i$ , with the force  $Fx_i$  acting on the particle in that direction—solving the differential Equation 2-1 yields a trajectory.<sup>92</sup> In order to calculate a trajectory the initial positions, velocities and accelerations must be defined. For example, the initial positions can be from an x-ray crystal or NMR structure, and the velocities and the acceleration are calculated by the gradient of the potential energy function.<sup>92,93</sup>

MD calculates the dynamics of system as it evolves with time, from which time average properties can be calculated. If a system is simulated for infinity then it will eventually cover the entire phase space of system. The ergodic hypothesis states that

the time average is equal to the ensemble average only if enough phase space has been sampled. The ergodic hypothesis allows MD simulations to calculate macroscopic variables using only a microscopic representation of a system.

### 2.1.1 Molecular Dynamics Integrator

The integration of equations of motion can be solved analytically for systems like one-dimensional harmonic oscillator. On the contrary, this is not feasible for such complex systems as proteins and these equations must be solved numerically.<sup>94</sup> Many numerical algorithms have been developed in order to accomplish this, for example, Verlet<sup>95</sup>, leap-frog<sup>96</sup>, velocity-Verlet<sup>97</sup>, and Beeman<sup>98</sup>. All of these algorithms use a Taylor series expansion to make approximations about the positions, velocities and acceleration. No matter what algorithm is employed during a MD simulation it must adhere to the following: conserve energy and momentum, be time-reversible and allow a long time step  $\Delta t$ .<sup>92</sup> The size of the time step is extremely important because if it is too small then time is wasted sampling the same conformation. On the other hand if the time step is too big then instabilities such as steric clashes occur and conservation of energy becomes difficult. The leap-frog algorithm was used in the MD simulations for this body of work and is defined as:

$$\vec{q}(t + \Delta t) = \vec{q}(t) + \vec{v}\left(t + \frac{1}{2}\Delta t\right)\Delta t \quad (2-2)$$

$$\vec{v}\left(t + \frac{1}{2}\Delta t\right) = \vec{v}\left(t - \frac{1}{2}\Delta t\right) + \vec{a}(t)\Delta t \quad (2-3)$$

where  $q$ ,  $v$  and  $a$  in Equations 2-2 and 2-3 correspond to the coordinate, velocity and acceleration vectors of all the atoms.<sup>96</sup> In the leap-frog algorithm the velocities ( $t + \frac{1}{2}\Delta t$ ) are initially calculated from the velocities at time ( $t - \frac{1}{2}\Delta t$ ) and the accelerations at

time  $t$ .<sup>92</sup> The positions  $q(t + \Delta t)$  are calculated from the velocities along with the positions using Equation 2-2. The velocities at time  $t$  are determined from:

$$v(t) = \frac{1}{2} \left[ v \left( t + \frac{1}{2} \Delta t \right) + v \left( t - \frac{1}{2} \Delta t \right) \right] \quad (2-4)$$

Thus, the velocities leap over the positions and vice versa. The advantage of this method is the velocities are explicitly calculated but the drawback of this method is that calculation of the positions and velocities are not deduced in the same step.

### 2.1.2 Langevin Dynamics

Langevin Dynamics are often used in conjunction with implicit solvation models or as a thermostat for temperature.<sup>92</sup> The Langevin equation of motion is a stochastic differential equation where two additional force terms have added to Newton's second law of motion in order to account for the omitted degrees of freedom. In the description of stochastic dynamics, the force exerted on particle is attributed to three different factors. The first force is owed to the interaction of a particles with other particles—the magnitude of the force acting on a particle depends on its position relative to the other particles.<sup>92</sup> The second force is described as the force felt by a particle as it moves through the solvent and is equated to the frictional drag on the particle due to the solvent.<sup>92</sup> The last force exerted on a particle is due to the random fluctuations that are a result of interactions with solvent molecules.<sup>92</sup> The Langevin Equation can be expressed as:

$$m_i \frac{d^2 x_i(t)}{dt^2} = F_i \{ x_i(t) \} - \gamma \frac{dx_i(t)}{dt} m_i + R_i(t) \quad (2-5)$$

where  $m$  is the mass,  $x$  is the position,  $F$  is an interaction force between a particle and other particles,  $\gamma$  represents the collision frequency, and  $R(t)$  is the force exerted on a particle due to random fluctuations with solvent molecules.

## 2.2 Molecular Mechanics

Molecular Mechanics is possible because of three main assumptions. First, the Born-Oppenheimer approximation, which allows the electronic and nuclear motion in molecules to be separated—the nuclear motion is neglected thus simplifying the construction of the Hamiltonian.<sup>92</sup> Second, Molecular Mechanics assumes that such processes as bond stretching, bond angles and rotations about the bond can describe the interactions within a system.<sup>94</sup> Lastly, the parameters calculated for the small molecules can be translated to describe larger systems such as proteins.<sup>94</sup> Force fields involve the following parameters:

$$\begin{aligned}
 V(r^N) = & \sum_{\text{Bonds}} \frac{k_r}{2} (r - r_{eq})^2 + \sum_{\text{Angles}} \frac{k_\theta}{2} (\theta - \theta_{eq})^2 + \sum_{\text{Torsions}} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \\
 & + \sum_i \sum_j \left( 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right) + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (2-6)
 \end{aligned}$$

where the first three terms describe the covalent interactions and the last two terms describe the non-covalent interactions.  $V$  denotes the potential energy, which is a function of the positions ( $r$ ) of  $N$  atoms. The first two terms in Equation 2-6, modeled as a harmonic potential due to the second order truncation of the Taylor series expansion, account for the energy change that occurs when the bonds and angles deviate from their reference value— $k$  is representative of the force constants for bonds and angles. The third term in Equation 2-6 defines the torsional angles. The torsional term is expressed in a Fourier series where  $V_n$  is the dihedral force constant,  $n$  is the dihedral periodicity,  $\omega$  is the torsional angle and  $\gamma$  is a phase of the dihedral angle  $\theta$ .<sup>99</sup> There are instances where additional terms called improper torsions are included in the force field to adequately describe such motions as out-of-plane bending. The incorporation of

improper torsions is important when atoms must remain in the same plane such as the case for a benzene ring. The fourth and fifth terms of Equation 2-6 are the Lennard-Jones and Coulomb potentials. The Lennard-Jones potential is used to describe the van der Waals interactions whereby  $\epsilon$  is the depth of the potential well,  $\sigma$  is the distance at which the inter-particle potential is zero,  $r^{-12}$  is the repulsive term and  $r^{-6}$  is the attractive term. No theoretical evidence has been provided to justify why the repulsive term is  $r^{-12}$ —many force fields implement the 6-12 Lennard-Jones potential because the repulsive term can be calculated by squaring the  $r^{-6}$  term thus reducing the computational cost.<sup>92</sup> The Coulomb potential is used to describe electrostatic interactions where  $q_1$  and  $q_2$  denote the charges of atoms,  $r$  represents the inter-particle distance and  $\epsilon$  is the electric constant.

In general, most of the force fields employed for describing protein have the same functional form as Equation 2-6. However, they differ greatly by what parameters are used in Equation 2-6, for example, the value of the forces constants of the bonds, angles and torsions. The AMBER ff99SB<sup>99</sup>, CHARMM22<sup>100</sup> and OPLS<sup>101</sup> are examples of force fields commonly employed in MD studies of biological systems. The force field AMBER ff99SB was used for the parameters of the MD simulations performed in this study.

### 2.3 Solvent Models

Solvent models are often employed to model the environment that proteins or any other system of interest exist in. The solvent is typically represented using two different approaches: explicit and implicit solvent. In an explicit solvent model, the

solvent is modeled with real atoms that interact with the solute and in the implicit solvent model the solvent is represented using a dielectric continuum. In most cases, the computational cost of MD simulations is increased when the solvent is explicitly modeled as opposed to implicitly.

### 2.3.1 Explicit Solvent

There are many different explicit water models that are utilized in molecular dynamics simulations: SPC/E,<sup>102</sup> TIP3P,<sup>103</sup> TIP4P<sup>103</sup>, and TIP5P<sup>104</sup>. The water models have been parameterized by fitting to the bulk properties of water such as: heat of vaporization, density and dipole moment.<sup>103</sup> The water models listed above differ mainly by the number of interaction sites they include. TIP3P and SPC/E have three interaction sites, TIP4P has four sites and TIP5P has five sites. The computational demand increases with the number of sites in the water model—the central processing unit (CPU) time is approximately proportional to the number of interatomic distances that must be computed.

The TIP3P model has three sites for electrostatic interactions—a negative charge (-8.314 e) is placed on the oxygen atom and a partial positive charge (0.417 e) is placed on each hydrogen atom. The van der Waals interactions between two waters molecules or atoms of solute are only computed with the oxygen atom. The geometry of TIP3P water is identical to the experimental geometry of a water molecule—the OH bond length is 0.9572 Å and the HOH angle of 104.52°. Unlike TIP5P, TIP3P does not accurately reproduce the density of water that exhibits a maximum at 4° C.<sup>104</sup> The TIP3P water model was used in all of the explicit solvent simulations described in this study.

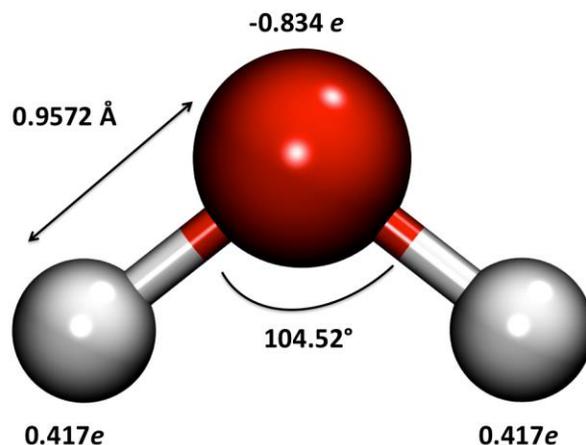


Figure 2-1. Schematic representation of a TIP3P water molecule

### 2.3.2 Implicit Solvent Models

Due to the cheaper computational cost and lower viscosity that enhances conformation sampling, the solvent is sometimes modeled implicitly. The methods are particularly helpful when the main focus of a study is the solute and the role of solvent is miniscule. Two main approaches are used to model the solvent-solute electrostatic interactions: Poisson-Boltzmann (PB) and Generalized-Born (GB) methods. The GB method is an approximation to the PB implicit solvent model and is less computationally demanding. The free energy of solvation is composed of an electrostatic and a non-electrostatic component. The PB and GB equations are used to calculate the electrostatic term while algorithms like LCPO<sup>105</sup> determine the non-electrostatic term by calculating the solvent-accessible surface.

#### 2.3.2.1 Poisson-Boltzmann solvent model

The PB models have shown that they can replicate conformations and energetics comparable to explicit solvent simulations as well as experimental measurements for a variety of systems.<sup>106</sup> The Poisson equation is the foundation of all

continuum solvent models. However, The Poisson equation, Equation 2-7 is only valid when the ionic strength of the system is zero.

$$\nabla\epsilon(r) \cdot \nabla\phi(r) = -4\pi\rho(r) \quad (2-7)$$

The Poisson equation must be amended in order to account for the presence of ions thus the electrostatic energy is computed by solving Poisson-Boltzmann equation:

$$\nabla \cdot [\epsilon(r)\nabla\phi(r)] = -4\pi\rho(r) - 4\pi\lambda(r) \sum_i z_i c_i \exp\left(\frac{-z_i\phi(r)}{k_B T}\right) \quad (2-8)$$

where  $\epsilon(r)$  is the dielectric constant,  $\phi(r)$  is the electrostatic potential, and  $\rho(r)$  is the solute charge,  $\lambda(r)$  is the Stern layer masking function,  $z_i$  is the charge of the ion type  $i$ ,  $c_i$  is the bulk number density of ion type  $i$ ,  $k_B$  is the Boltzmann constant and  $T$  is the temperature.

### 2.3.2.2 Generalized Born solvent model

The Generalized Born model is often used to represent the solvent in protein simulations due to a cheaper computational cost when compared to other implicit solvents. As discussed earlier the GB is an approximation to PB and when given the same parameters such as atomic radii, the results of PB and GB are the same for small to medium-sized molecules.<sup>107</sup> The basis of the GB method is from the equation for the polarization energy of monoatomic ion in atomic units, often referred to as *so-called Born equation*, Equation 2-9.

$$G_p = -\frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \frac{q^2}{a} \quad (2-9)$$

An expansion of Equation 2-9 to apply to polyatomic molecules yields the GB equation that can be expressed as:

$$\Delta G_{GB} = -\frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{f(r_{ij}, a_{ij})} \quad (2-10)$$

where  $\epsilon$  is the dielectric constant,  $q$  is the charge of an atom,  $r_{ij}$  is the inter-particle distance, and  $a_{ij}$  is the Born radii. The  $f(r_{ij}, a)$  which is an arbitrary function developed by Still *et al.*<sup>108</sup> that can be further simplified to:

$$f(r_{ij}, a_{ij}) = \sqrt{r_{ij}^2 + a_{ij}^2 \exp\left(\frac{r_{ij}^2}{4a_{ij}^2}\right)} \quad (2-11)$$

where  $r_{ij}$ ,  $a_{ij}$  correspond to the inter-particle distance and the Born radii. Equation 2-11 has several advantages but the main advantage is that the expression can be differentiated analytically, allowing the solvation term to be incorporated in MD simulations and gradient-based optimization methods.<sup>92</sup>

## 2.4 Steered Molecular Dynamics

Steered Molecular Dynamics involves applying an external force onto a system in order to drive changes in its coordinates within a certain time. SMD can be applied to study such problems as ligand dissociation, ion diffusion, and conformation changes in biological systems on time scales accessible to MD simulations. SMD is similar to umbrella sampling except that the center of restraint is time-dependent.

The second law of thermodynamics mandates that the ensemble average of work performed on a system by an external force be either larger than or equal to free energy difference between two states—the dissipated work is just the difference between  $\overline{W} - \Delta F$ .

$$\overline{W} = \Delta F_{A \rightarrow B} \quad (2-12)$$

The equality in Equation 2-12 is only valid for quasistatic processes, which means that the change from A to B must occur infinitely slowly. However, using the so-called Jarzynski relationship equality<sup>109</sup>, Equation 2-13 allows work calculated from going from

state A to B for example, to be related to equilibrium free energies from non-equilibrium processes.

$$\exp(-\beta\Delta F)_{A\rightarrow B} = \overline{\exp(-\beta W)_{A\rightarrow B}} \quad (2-13)$$

Because the average work must be calculated, many MD simulations must be run, each simulation starting from different configurations of the initial state (in this case state A).

## 2.5 Constant pH Methods

The structure of a protein and its function is contingent on the pH of the solvent. The pK<sub>a</sub> dictates the protonation of a titratable residue at a particular pH. The pK<sub>a</sub> values of titratable residues are governed by its surrounding electrostatic environment. In traditional MD simulations, protonation states of titratable residues are assigned prior to the simulations. This approach has two issues. First, the assignment of the protonation state is usually taken from the protonation state of the residue in free solution.<sup>110</sup> The pK<sub>a</sub> of a residue in free solution and its pK<sub>a</sub> in the folded protein can differ significantly. Lastly, when the pK<sub>a</sub> values of titratable residues of a system are near solution pH, no single protonation state properly describes the ensemble at that pH.

### 2.5.1 Constant pH Molecular Dynamics

Constant pH MD alleviates the problems listed above by allowing proper sampling of coupled conformational and protonation state distributions corresponding to the semi-grand canonical ensemble.<sup>111,112</sup> The different CpHMD methods available can be categorized in two main approaches: discrete protonation states<sup>112</sup> and continuous protonation states.<sup>113</sup> In discrete protonation state models, Monte Carlo exchanges between different protonation states are attempted during the simulation. Continuous

protonation state models use a titration coordinate  $\lambda_i$  ( $0 \leq \lambda \leq 1$ ) to control the process of protonation and deprotonation of a group labeled  $i$ .

The CpHMD algorithm<sup>114</sup> implemented in AMBER treats the solvent implicitly using the Generalized Born method. At each Monte Carlo step, a change in protonation state of either one or two closely interacting titratable residues is attempted. The transition free energy, Equation 2-14, between the present and proposed protonation state is calculated applying the Metropolis criterion<sup>115</sup> to determine if the proposed state is accepted or rejected.

$$\Delta G = k_B T (pH - pK_{a,ref}) \ln 10 + \Delta G_{elec} - \Delta G_{elec,ref} \quad (2-14)$$

In Equation 2-14  $k_B$  is the Boltzmann constant,  $T$  is the temperature,  $pH$  is the current  $pH$  of the solvent,  $pK_{a,ref}$  is the  $pK_a$  of the appropriate reference compound,  $\Delta G_{elec}$  is the electrostatic component of the free energy calculated for the titratable group in the protein,  $\Delta G_{elec,ref}$  is the electrostatic component of the transition free energy for the reference compound, a free dipeptide amino acid.<sup>114</sup>

## 2.5.2 Constant pH Replica Exchange Molecular Dynamics

Before introducing Constant pH Replica Exchange Dynamics, a brief introduction to Replica Exchange will be given. Replica Exchange Molecular Dynamics is the MD version of parallel tempering developed by Sugita and Okamoto.<sup>116</sup> REMD is a method where multiple copies or replicas of the same system are simulated simultaneously with different parameters i.e. temperature. Periodically, the MD simulation is halted and information, i.e. conformations are swapped between adjacent replicas based on the Metropolis criterion. A schematic representation of REMD is shown in Figure 2-2.

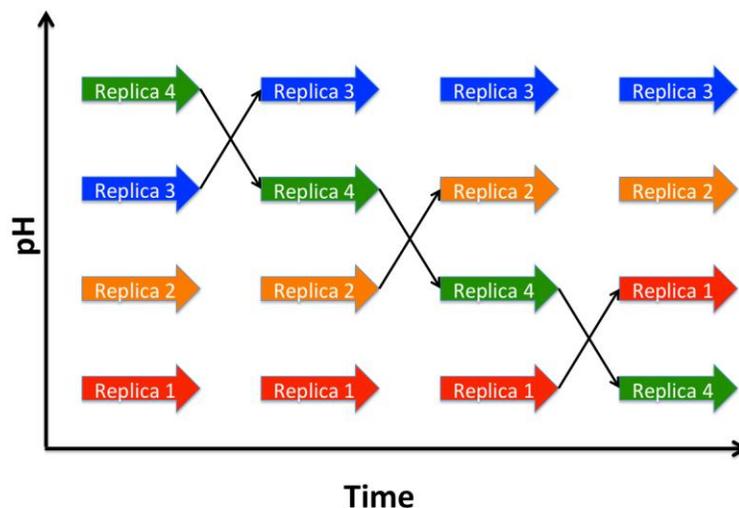


Figure 2-3. Schematic representation of REMD simulations. Each arrow represents an independent MD simulation—MD is halted and MC exchange is attempted, which can be accepted or rejected. If the exchange attempt is successful then replicas are swapped between adjacent neighbors if not the MD continues.

The concept of coupling CpHMD and Replica Exchange Molecular Dynamics, which is referred to as pH Replica Exchange Molecular Dynamics (pH-REMD), was first introduced by Itoh et al.<sup>117</sup> and implemented into AMBER by Swails et al.<sup>110</sup> In pH-REMD each replica is simulated at a different pH using CpHMD as described previously, and the solution pH is swapped between neighboring replicas. This method has been shown to enhance conformational sampling as well as produce  $pK_a$  values that converge faster and achieve better precision when compared to just employing CpHMD.<sup>117,110</sup>

## 2.6 Accelerated Molecular Dynamics

The potential energy surface of biological systems is composed of multiple minima separated by large barriers. Correct sampling of the PES can be difficult using conventional MD because the time scales of MD simulations are limited to the nanosecond-microsecond ranges. Consequently, MD simulations become trapped in

local minima for long periods of time—ensemble properties calculated from the MD simulations become erroneous due to incomplete sampling. As stated previously, the ergodic hypothesis states that the time average is equal to the ensemble average only if enough phase space has been sampled.

Hamelberg *et al.*<sup>118</sup> introduced a bias potential method named Accelerated Molecular Dynamics that has been shown to circumvent the problem of kinetic-trapping. AMD lowers the barrier height therefore decreasing the amount of time it takes to properly sample the conformational space.<sup>119</sup> The boosted energy surface from AMD maintains the integrity of the underlying shape of the real one—this ensures that original distribution can be recovered by simply reweighing the distribution of structures from the AMD simulation. The criteria for when the potential boost is added during the simulation are given by Equation 2-15 and AMD modifies the potential using the following equation:

$$V(r)^* \begin{cases} V(r) & V(r) \geq E \\ V(r) + \Delta V(r) & V(r) < E \end{cases} \quad (2-15)$$

$$\Delta V(r) = \frac{(E-V(r))^2}{\alpha+(E-V(r))} \quad (2-16)$$

where  $V(r)$  is then normal potential,  $\Delta V$  is the boosted potential,  $E$  dictates the energy below which the system is accelerated,  $\alpha$ , the tuning-parameter affects the smoothness boosted energy potential  $\Delta V$ . There has been no theoretical justification as to how to choose the parameters  $E$  and  $\alpha$ —parameters used in the AMD must be chosen carefully if the parameters are too small then little enhancement will be seen in the sampling and if they are too large then simulation will become unstable.

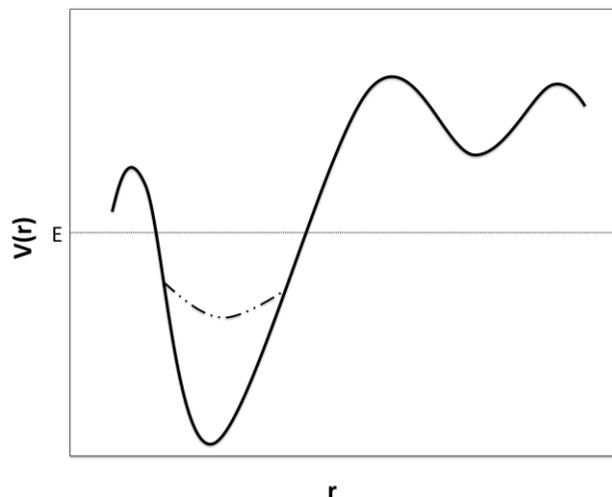


Figure 2-4. Schematic representation of the normal and biased potential of AMD simulations. (Solid line) 1-D potential energy landscape, unmodified. (Dashed line) modified potential energy landscape. E is the threshold boost energy.

## 2.7 Docking

The concept of docking derives from the proposed “lock and key” model that states only the key (substrate) with the correct size/shape will fit into the lock (active site of protein). Docking is a method that predicts the orientation of a ligand when bound to a protein. Docking is often used in conjunction with drug design because large libraries of compounds can be rapidly screened for potential inhibitors as opposed to experimentally determining the binding affinities of the compounds, which would be expensive and time-consuming.

Docking calculations proceed in the following fashion: generating a grid around the target receptor, conformational sampling of the ligand, and scoring and ranking each docked pose. Each grid point within the grid encompassing the target receptor is used to calculate the interaction potential between the ligand and target receptor. The larger the grid the more expensive the docking calculation becomes, therefore, only the region (active site) of the target receptor where the ligand will bind is generally included in the

grid. The ideal scenario would be to generate every possible conformation of ligand and receptor bound together—this is only plausible if considering only a few ligands and as the number ligands evaluated increases so does the computational expense. Some docking programs instead only allow certain residues in the protein deemed important to be flexible while the rest of protein is rigid. Generally, most docking calculations are performed keeping the receptor rigid and only grant flexibility to the ligand. Different poses of the ligand are generated by an exhaustive search of all angle-torsion space for the rotatable bonds. Each pose is scored using an empirically derived scoring function similar in form to Equation 2-6.

## 2.7 Molecular Mechanics Poisson Boltzmann Surface Area

Free energy calculations have proven to be great resource in computational biology, especially as it pertains to drug-design. Thermodynamic Integration<sup>120</sup> (TI) and Free Energy Perturbation<sup>121</sup> (FEP) are examples of methods available to calculate free energies. These methods, though theoretically rigorous, become computationally expensive as the size of system increases. The TI and FEP divide the reaction coordinate into intermediate states to calculate the free energy between two states. Molecular Mechanics Poisson-Boltzmann Surface Area<sup>122</sup> is an end-state method that eliminates the need for these intermediate states therefore reducing the computational cost. MMPBSA is a post-processing end-state method that calculates the free energies using the thermodynamic cycle in Figure 2-5.

The binding energies in MMPBSA can be calculated in the following fashion.

$$\Delta G_{Binding,solvated} = G_{Complex,solvated} - (G_{Receptor,solvated} + G_{Ligand,solvated}) \quad (2-17)$$

The terms on the right-hand side of the equation can further simplified to:

$$\Delta G_{solvated} = H_{gas} - TS_{solute} + G_{solvation} \quad (2-18)$$

where,  $H_{gas}$  is the gas-phase energies,  $TS_{solute}$  is the entropy of the solute, and  $\Delta G_{solvation}$  is the free energy of solvation.  $H_{gas}$  is calculated from the MM force field, Equation 2-6. The entropy of the solute can be determined using either the normal mode or quasi-harmonic approximation. The free energy of solvation is composed of electrostatic and non-electrostatic contributions, where the electrostatic contributions are determined from the implicit solvent model i.e. PB/GB and the non-polar contribution is calculated from the SASA.

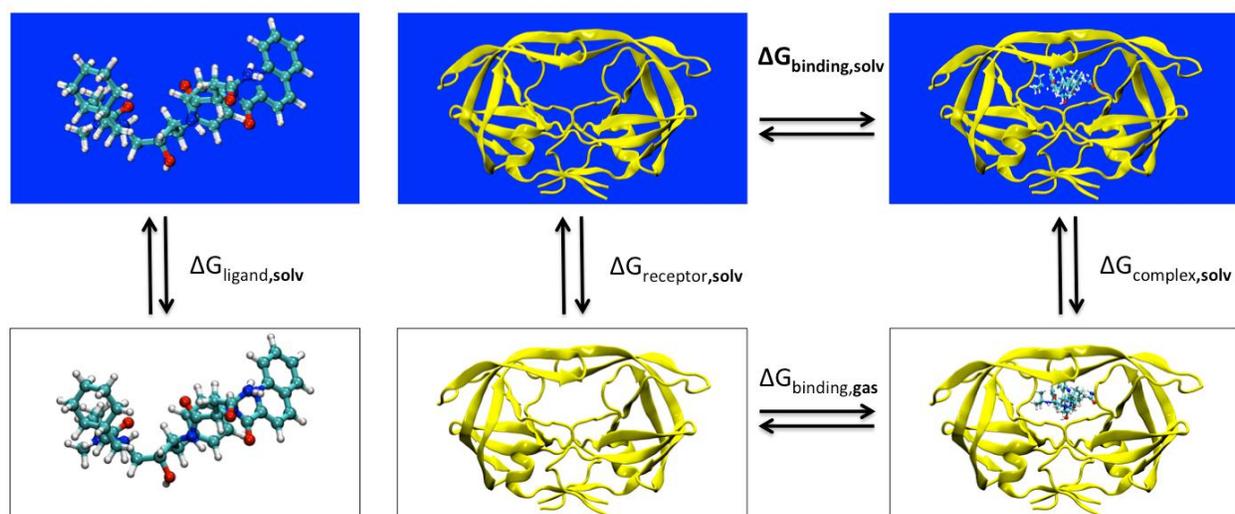


Figure 2-5. Thermodynamic cycle of MMPBSA. HIV- protease and saquinavir, PDB 3OXC. The blue background represents the aqueous form of the ligand, receptor and complex. The goal of the MMPBSA is to calculate the free energy difference between two states, which is frequently represented by the bound and unbound state of two molecules that are solvated (colored blue).

## CHAPTER 3 MOLECULAR DYNAMICS SIMULATIONS INDICATE THAT THE CATALYTIC-DYAD OF HIV-1 PROTEASE IS MONO-PROTONATED

### 3.1 Introduction

The catalytic-dyad has a significant role in protein-ligand complexes as well as on the dynamics of the flaps in the apo form.<sup>43,123</sup> In computational drug design studies, it is imperative to assign protonation states correctly in order to achieve accurate binding affinities and conformational sampling.<sup>124,125</sup> The catalytic-dyad of HIVPR has been modeled using molecular dynamics in the deprotonated, di-protonated, and mono-protonated ionization states.<sup>50,99,126</sup> Constant pH Molecular Dynamics (CpHMD) is an ideal technique to resolve this problem due to the uncertainty of assigning the correct protonation states of the catalytic-dyad.

We performed CpHMD and pH-REMD simulations on the protease in its apo form and bound to thirteen protease inhibitors, which are shown in Figure 3-1. The inhibitors used in this study can be divided into three main groups: cationic, neutral asymmetric and cyclic urea inhibitors. The neutral asymmetric group is chiefly composed of FDA approved inhibitors that are currently being used to treat patients infected with HIV.

CpHMD was performed at pH 5 (HIVPR is most catalytically active in a pH range of 4 – 6)<sup>127</sup> and pH 7 (physiological pH). pH-REMD simulations were performed at pH 0 – 11. This study shows evidence that the protonation state of the catalytic-dyad is mono-protonated for both the apo and bound forms of HIVPR over a wide pH range. In addition, we determine which of catalytic aspartates D25 or D25' defined in later sections, should be protonated and which of the oxygen atoms, OD1 or OD2, possesses the proton when the aspartate is protonated in the bound form of HIVPR.

These results will be compared to previous theoretical and experimental results in which the protonation state of the catalytic dyad have been investigated.

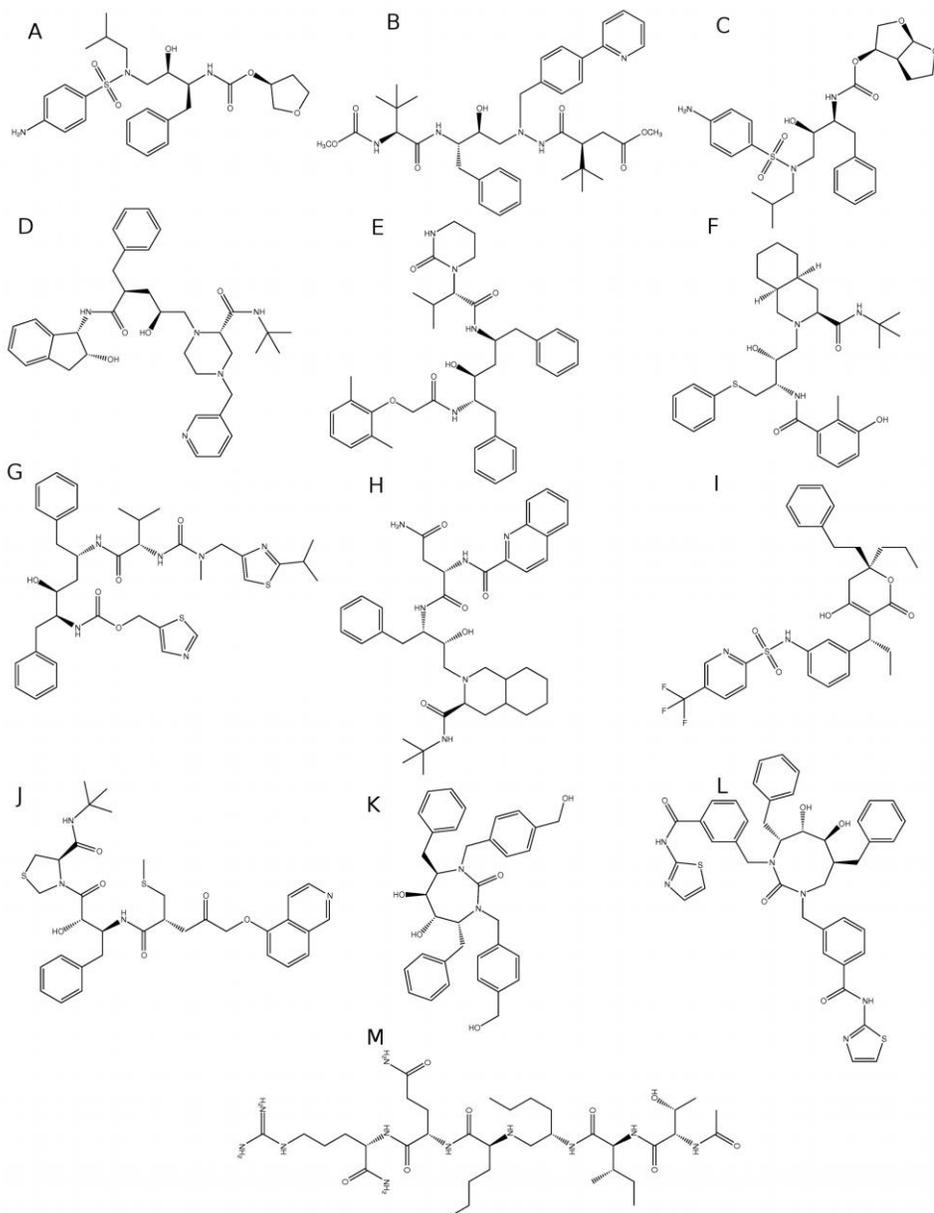


Figure 3-1. 2D schematic of the protease inhibitors used in the Constant pH Molecular Dynamic Simulations. A) Amprenavir, B) Atazanavir, C) Darunavir, D) Indinavir, E) Lopinavir, F) Nelfinavir, G) Ritonavir, H) Saquinavir, I) Tipranavir, J) KNI-272, K) DMP-323, L) XV-638, M) MVT-101. Inhibitors A – J are the neutral asymmetric inhibitors, K – L are the cyclic urea inhibitors, and M belongs to the cationic group.

## 3.2 Methods

### 3.2.1 General Setup

Fourteen different forms of HIVPR were simulated in which the initial coordinates were X-ray crystal structures from the Protein Data Bank, PDB codes: 1HHP – apo<sup>128</sup>, 3NU3 – Amprenavir<sup>129</sup>, 2O4K – Atazanavir<sup>130</sup>, 3TKW – Darunavir<sup>131</sup>, 2BPX – Indinavir<sup>132</sup>, 2O4S – Lopinavir<sup>130</sup>, 1OHR – Nelfinavir<sup>133</sup>, 1HXW – Ritonavir<sup>134</sup>, 2O4P – Tipranavir<sup>130</sup>, 1QBS – DMP -323<sup>135</sup>, 1QBR – XV-638<sup>135</sup>, 4HVP – MVT-101<sup>39</sup> and 3FX5 – KNI-272.<sup>136</sup> In order to be consistent, all the crystal structure sequences were mutated to match that of 1HHP, Table A-1. Having identical sequences is important because previous studies have shown that active site mutations and non-active site mutations, even though not close in proximity to the active site, can have an impact on the dynamics of the flaps as well as how residues in the active site interact with the ligand.<sup>137,131</sup> Missing hydrogen and heavy atoms were added using the *LEaP* module in AmberTools 12.<sup>138</sup> The AMBER *ff99SB*<sup>99</sup> force field was used for the protein and the parameters for the ligands were obtained using the *antechamber* module and GAFF force field<sup>139</sup> with AM1-BCC charges.<sup>140</sup>

Each HIVPR complex began from three different protonation states: deprotonated, di-protonated and mono-protonated. In the case of the simulations that began in the mono-protonated state, the hydrogen was initially placed on the OD2 atom of D25 and in the simulations that started in the di-protonated state, a hydrogen was placed on the OD2 atom of D25 and the OD1 atom D25'.

The effects of the solvent were accounted for by using the Generalized Born<sup>108</sup> implicit solvent model, GB<sup>OBC</sup> model<sup>141</sup> (*igb=5* in AMBER) in conjunction with the *mbondi2* intrinsic radii. The internal dielectric was set to 1 and the external dielectric to

78.5. The cutoff for the van der Waals and electrostatic interactions was set to 30 Å and the salt concentration, modeled as Debye-Hückel screening parameter was set to 0.1 M. Langevin Dynamics was utilized in order to maintain the temperature at 300 K and the collision frequency was set to 50 ps<sup>-1</sup> to mimic the viscosity of liquid water. Different initial seeds were applied to each simulation to avoid synchronization artifacts.<sup>142</sup> The SHAKE algorithm<sup>143</sup> was used to constrain bonds containing hydrogen so that a 2 fs time step could be used.

### 3.2.2 Constant pH Molecular Dynamics Protocol

An equilibration procedure was employed prior to performing CpHMD on the various HIVPR complexes. First, the hydrogen atoms were minimized using 500 steps of steepest descent followed by 7000 steps of conjugate gradient. Next the entire system was minimized using 500 steps of steepest descent and 10000 steps of conjugate gradient. All systems were heated from 100 K to 300 K for 1 ns where 100 kcal/mol·Å<sup>2</sup> positional restraints were used on the backbone atoms. Following heating, four 200 ps simulations were performed where the positional restraint was reduced from 100 to 0.1 kcal/mol·Å<sup>2</sup> for the backbone atoms. Finally, in the last step of equilibration all restraints were removed and a 200 ps simulation was performed. No protonation state changes were attempted during the equilibration process.

The production phase consisted of 12 ns of unrestrained CpHMD of which the last 10 ns were used for analysis. A MC step was performed every 10 fs in order to determine if a change in protonation should occur. Only D25 and D25' were allowed to change protonation states. All minimizations and CpHMD simulations were performed using the *sander* module in AMBER 12.<sup>138</sup>

### 3.2.3 Constant pH Replica Exchange Molecular Dynamics Protocol

Two different sets of pH-REMD simulations were performed. The first set of pH-REMD simulations included all 14 HIVPR complexes—where 12 equally spaced replicas spanning from pH 0 – 11 were used. Each simulation was begun in a mono-protonated state as described in the previous section. Only D25 and D25' residues were allowed to change protonation states.

The second set of pH-REMD simulations included 12 replicas spanning from pH 0-11. All of the titratable residues of the protein, excluding the arginine amino acids, N and C termini of the HIVPR complexes were allowed to change protonation states. The arginine amino acid residues, N and C termini, were excluded because parameters are not available. All of the aspartates and glutamates began deprotonated except for one of the catalytic aspartates, D25'. The cysteine, tyrosine and histidine (the hydrogen was placed in the epsilon position) amino acid residues began each simulation in the neutral form. The lysine amino acid residues started each simulation protonated.

The starting structures for the pH-REMD simulations were acquired from employing an equilibration procedure for each pH, similar as to what was done for the CpHMD simulations. For the first set of simulations, a protonation state change was attempted every 10 fs and for the second set of simulations a protonation state change was attempted every 4 fs to determine if a change in protonation should occur. The exchange attempt frequency (EAF) of  $100 \text{ ps}^{-1}$  was used for the first set of pH-REMD simulations. The impact of a larger EAF would be negligible since only two residues protonation states were probed. The EAF was  $10 \text{ ps}^{-1}$  for the second the set of pH-REMD simulations. A larger EAF is warranted given that the number of titratable residues to consider is substantially more, 36 in all, when compared to the first set of

pH-REMD simulations. Exchanging more frequently will allow simulations to transverse the replica ladder more rapidly, which improves the conformational sampling.<sup>110,144</sup> The production phase for the first set of pH-REMD was 30 ns and the production phase for the second set of pH-REMD simulations was 50 ns.

### 3.2.4 General Analysis

The *cpptraj*<sup>145</sup> module of AmberTools was used to analyze the trajectories from the CpHMD and pH-REMD simulations. The program *calcpka* in AmberTools was utilized to calculate the resulting protonation state distributions from the CpHMD and pH-REMD simulations.

## 3.3 Results

This section is organized as follows: first, the results from the CpHMD, and secondly, the pH-REMD simulations are presented. Lastly, we offer results to determine if D25 or D25' of the catalytic-dyad is protonated in conjunction with which of the oxygen atoms, OD1 or OD2, should possess the proton. Only the results from the CpHMD and pH-REMD simulations in the pH range pH 4 – 7 will be considered, since this is the most experimentally relevant range.

### 3.3.1 Constant pH Molecular Dynamics

All HIVPR CpHMD simulations executed at pH 5 converge to the mono-protonated protonation state regardless of their initial protonation state. Furthermore, the results indicate that 10 ns is more than ample time due mainly in part to each of the different initial protonation states converging to the same protonation state as shown in Figure 3-2. As mentioned in earlier sections, the mono-protonated simulations were run with the D25 being initially protonated. In order to ensure that the results of the mono-protonated simulations were not biased we ran three HIVPR complexes amprenavir,

DMP-323, and MVT-101 in which D25' was instead protonated and the results remained consistent with those obtained from the D25 simulations.

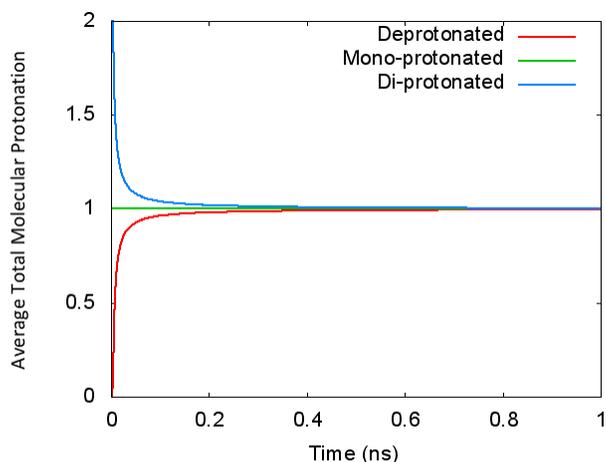


Figure 3-2. Plot of Average total molecular protonation versus time for HIVPR complex KNI-272. CpHMD performed at pH 5, where a total molecular protonation of 0.00 indicates that neither aspartate is protonated, 1.00 that only one aspartate is protonated and 2.00 that both are protonated.

Because each system converged to the same result at pH 5, only the apo structure, one neutral inhibitor, (nelfinavir), one cationic inhibitor, (MVT-101) and one cyclic urea inhibitor, (DMP-323) were each simulated at pH 7. The simulations were initially begun in the same manner as described earlier. Each system was simulated at three different initial protonation states in order to determine if they will too converge to the same protonation state. All four HIVPR complex simulations converged to the mono-protonated state regardless of its initial protonation similar to the CpHMD simulations at pH 5.

### 3.3.2 Constant pH Replica Exchange Molecular Dynamics

In order to ensure that the results obtained from the CpHMD simulations were not a result of conformational trapping, pH-REMD was employed. The first set of pH-REMD simulations included all 14 HIVPR complexes. The pH values of the pH-REMD

simulations spanned from 0 – 11 and only the catalytic aspartates were allowed to change protonation. The results of the pH-REMD simulations reveal that the catalytic aspartates of HIVPR are mono-protonated between pH 4 – 7.

The second set of pH-REMD simulations consisted of allowing all titratable residues in the protein, excluding those listed in the methods section, to change protonation. The  $pK_a$  is governed by local electrostatic interactions, which is dictated by conformation of the protein and protonation state of titratable residues—because of the tight coupling of protein conformation and protonation state we chose to include all possible titratable residues.<sup>114</sup> Each replica was initiated in the same fashion as described previously. As was done in the CpHMD simulations, only the apo structure, as well as one neutral inhibitor, (KNI-272), one cationic inhibitor, (MVT-101), and one cyclic urea inhibitor (DMP-323) were used in the pH-REMD simulations. The catalytic aspartates for all four different HIVPR complexes remained mono-protonated between pH 4 – 7.

### **3.3.4 Location of the proton in the mono-protonated state**

Our results have shown that the catalytic aspartates are mono-protonated in the apo and holo forms with the various inhibitors used in the study. The proton could exist on D25 or D25'. However, when D25 or D25' is protonated which of the oxygen atoms, OD1 or OD2 possesses the proton?

In the apo form of the enzyme the D25 and D25' are chemically indistinguishable. On the other hand, in the bound form of the enzyme D25 and D25' become distinguishable. In the bound form D25 and D25' have been labeled based on their state at  $t=0$ , where  $t$  is time. D25 is the aspartate that is protonated and D25' is the aspartate that is deprotonated as shown in Figure 3-3.

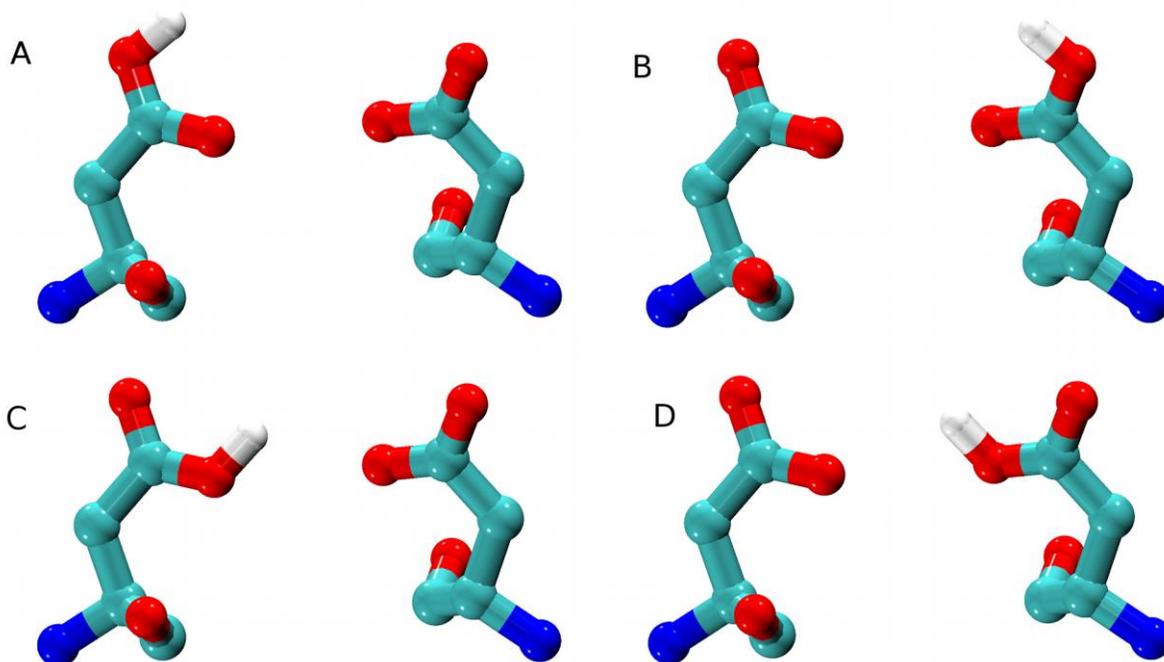


Figure 3-3. Schematic of the four different possible mono-protonated protonation states. A) Asp25 OD2 protonated, B) Asp25' OD2 protonated, C) Asp25 OD1 protonated, D) Asp25' OD1 protonated

In the holo simulations, D25' does become protonated while D25 simultaneously deprotonates. This illustrates that either of the catalytic aspartates can be protonated in the neutral asymmetric, cyclic-urea and cationic inhibitor simulations. In the simulations involving the neutral asymmetric inhibitors, the central hydroxyl group of the inhibitor rotates to interact with the catalytic aspartate that is deprotonated when a change in protonation occurs between D25 and D25'.

The next issue is which of the oxygen atoms possesses the proton when one of the catalytic aspartates is protonated, OD1 (inner) or OD2 (outer) oxygen atoms. The oxygen atoms OD1 and OD2 have been defined based on their orientation at  $t=0$  as shown in Figure 2-3 A. The four different mono-protonated arrangements are shown in Figure 2-3. It is not clear from the simulations which oxygen atom OD2 or OD1 should

possess the proton. However, the proton is never seen in the *anti* position but only in *syn* position relative to the oxygen atoms.

### 3.4 Discussion

#### 3.4.1 Apo (Free enzyme)

Our results suggest that the aspartyl dyad of the apo form of HIVPR is mono-protonated, which corroborates with previous kinetic and theoretical studies.<sup>53,127,146–148</sup> In experimental kinetic studies done by Ido *et al.*<sup>127</sup> and Hyland *et al.*<sup>53</sup> their results show that the aspartyl dyad of free protease has two apparent pK<sub>a</sub>s ~3 and ~6. In theoretical studies done by Trylska *et al.*<sup>146</sup> and Czodrowski *et al.*<sup>147</sup> their results yielded pK<sub>a</sub>s comparable to those found in the kinetic studies.

However, a previous NMR study, conducted by Smith *et al.*<sup>149</sup> determined that the aspartyl dyad is deprotonated at pH 6.0 in the absence of an inhibitor, which our results contradict. Trylska *et al.*<sup>146</sup> have already addressed explanations as to why the results of Smith *et al.*<sup>149</sup> may not agree with the general consensus.

#### 3.4.2 Cyclic Urea Inhibitors

When cyclic urea inhibitors, DMP-323 and XV-638 are bound to HIVPR the catalytic aspartates remain mono-protonated between pH 4 – 7. This result is in agreement with the findings of Mardis *et al.*<sup>150</sup> –they sought to explain the observed U shape of binding free energy as function of length for series of cyclic urea inhibitors. The study revealed that the mono-protonated model of the catalytic-dyad yields the best agreement with experimental trends, while the di-protonated model was unsuccessful in reproducing the characteristic U shape of measured binding free energy chain length.<sup>150</sup> However, the author did note that the calculated absolute binding free energies suggest that the cyclic urea inhibitors bind to the protein in the di-protonated state. Trylska *et*

*et al.*<sup>146</sup> performed two sets of calculations using the full-charge/4 model<sup>151</sup> and the single-site/20 model.<sup>152</sup> The results from the single-site/20 model yielded pK<sub>a</sub>s of ~3 and ~6 for the catalytic aspartates that coincide with our results.

On the contrary, a previous NMR study, found that when DMP-323 is bound to the protease that the catalytic-dyad exists in the di-protonated state over the pH range 2 – 7.<sup>153</sup> Trylska *et al.*<sup>146</sup> full-charge/4 model results and Czodrowski *et al.*<sup>147</sup> found that their theoretical results suggest that the aspartyl dyad is fully protonated. However, Trylska *et al.*<sup>146</sup> did experimentally determine that the inhibition constant for DMP-323 remained constant from a pH range of 5.5 – 7.<sup>146</sup> If HIVPR when bound DMP-323 is di-protonated then the binding affinity should decrease with increasing pH.

Reasons as to why there are some discrepancies between our proposed results and Trylska *et al.*<sup>146</sup> and Czodrowski *et al.*<sup>147</sup> might be that their calculations only incorporated static structures, and did not account for the flexibility of the protein-ligand complex. Their solvent was modeled implicitly using the Poisson Boltzmann method while we instead used the Generalized Born method. Lastly, different dielectrics were used which can have a profound impact on the calculated pK<sub>a</sub> value.<sup>154</sup>

### **3.4.3 Neutral Asymmetric Inhibitors**

The catalytic aspartates of HIVPR are mono-protonated from pH 4–7 when bound to the neutral asymmetric inhibitors employed in this study. Our results obtained from the CpHMD and pH-REMD simulations are found to be in agreement with other previous experimental and theoretical studies suggesting that the protease is mono-protonated for neutral asymmetric inhibitors.<sup>146,147,155,156</sup> Unfortunately not all the inhibitors used in this study have experimental and theoretical pK<sub>a</sub> values that we can compare to but we will compare when possible.

Wang *et al.*<sup>155</sup> experimentally determined that the aspartyl dyad of HIVPR is mono-protonated from a pH range of 2.5 – 6.2 for the inhibitor KNI-272 and were unable to extend beyond this range. The theoretical calculations of Trylska *et al.*<sup>146</sup> and Czodrowski *et al.*<sup>147</sup> does show that the catalytic-dyad is mono-protonated when KNI-272 is bound, between pH 4 – 7 with pK<sub>as</sub> < 4 and ~7 for D25 and D25'. Our results for amprenavir are found to be in good agreement with those of Czodrowski *et al.*<sup>147</sup>

#### 3.4.4 Cationic Inhibitor

The catalytic aspartates remained mono-protonated between pH 4 to 7 when the inhibitor MVT-101 is bound. The pK<sub>as</sub> of D25 and D25' have not been experimentally determined for this protein-ligand complex. However, our findings agree with previous theoretical studies of Geller *et al.*<sup>157</sup> and the results from full-charge/4 model of Trylska *et al.*<sup>146</sup>

A MD study performed by Harte *et al.*<sup>158</sup> proposed that the protonation state of the aspartyl dyad is deprotonated when cationic protease inhibitor MVT-101 is bound.<sup>158</sup> However, as previously noted, Trylska *et al.*<sup>146</sup> performed two sets of calculations and the single-site/20 calculation suggested that the aspartyl dyad is deprotonated from pH 5 – 7.<sup>146</sup>

Here we offer an explanation as to why Harte *et al.*<sup>158</sup> and our results do not coincide—explanations as to why the results from Trylska *et al.*<sup>146</sup> do not possibly corroborate with ours have been addressed previously. Harte *et al.*<sup>158</sup> performed 100 ps MD simulations using different ionization states of the protein: mono-protonated, di-protonated, and deprotonated. Their assignment of the protonation state of the D25 and D25' were based merely on comparing the backbone average structure of the ligand from the MD simulations to the orientation of the ligand in the crystal structure and

deprotonated model of the catalytic dyad yielded the lowest deviations of 0.5 Å. The lengths of their simulations might not be sufficient and longer time scales are needed—the correlation between ligand orientation and assigning protonation of a residue is not obvious.

### **3.4.5 Low-Barrier Hydrogen Bond**

As mentioned above, either D25 or D25' can be protonated in the bound form of HIVPR. Furthermore, the proton can exist on either the OD1 or OD2 when the aspartate is protonated. This topic is still under investigation but has been addressed by previous studies.<sup>156,159,160</sup>

One explanation as to why the proton can be on either oxygen could be attributed to the low-barrier hydrogen bond (LBHB).<sup>148,161</sup> The LBHB is believed to be a compensation for the strong repulsion between the two catalytic aspartates.<sup>148</sup>

## **3.5 Conclusions**

Assigning the correct protonation state to the catalytic-dyad is important in computational drug design. We have shown through the use of the CpHMD and pH-REMD simulations that the protonation state of the protease is mono-protonated in the apo form as well as with the various inhibitors included in this study. This is illustrated by the convergence of the HIVPR complexes to the mono-protonated state regardless of the initial protonation state. We conclude that the binding of an inhibitor does not induce a change in the protonation state when transitioning from the apo to bound form of HIVPR for the inhibitors used in this investigation.

There is still some ambiguity concerning where the proton is located in the mono-protonated state and further investigation is needed. Possibly the position of the proton

in the mono-protonated state cannot be defined by a single state, but instead by an ensemble of arrangements in the mono-protonated bound form of HIVPR.

CHAPTER 4  
UNDERSTANDING HOW THE G48T/L89M MUTATIONS REDUCE THE EFFICACY  
OF THE HIV-1 PROTEASE INHIBITOR SAQUINAVIR

4.1 Introduction

The lifetime of HIV patients has been extended due to the success of highly active antiretroviral therapy (HAART). The HIV protease has remained an important for antiretroviral drug therapy target due to its role in the life cycle of the virus.<sup>57</sup> Inhibition of the protease would lead to an immature virion that is non-infectious. The efficacy of protease inhibitors is reduced as a result of drug pressure selected mutations and mutations cleavage sites of the protease.<sup>81,82</sup>

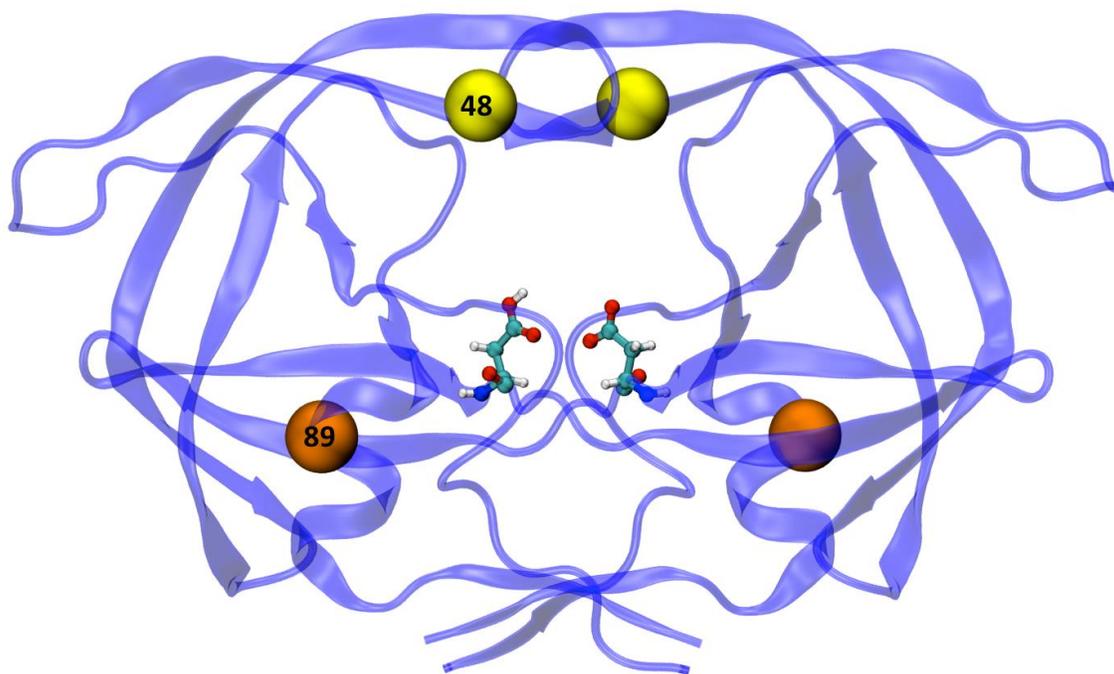


Figure 4-1. HIV-1 protease, the colored spheres indicate locations on the protease the where wild-type and mutant differ. Position 48 (G48T) is colored yellow and position 89 (L89M) is colored orange. The catalytic aspartates are also depicted.

Mutations that cause resistance to the protease inhibitors are not just restricted to residues in the active site but can include residues in other regions of the protein.<sup>86</sup>

The mechanism of how an active site mutation elicits resistance is caused by reduced interactions with the inhibitor and altering the size of the active site cavity.<sup>83</sup> Non-active site mutations are defined as sites that do not directly interact with inhibitors yet confer resistance—the mechanism of drug resistance for non-active site mutations is not well understood although several mechanisms have been proposed.<sup>83,89,90</sup>

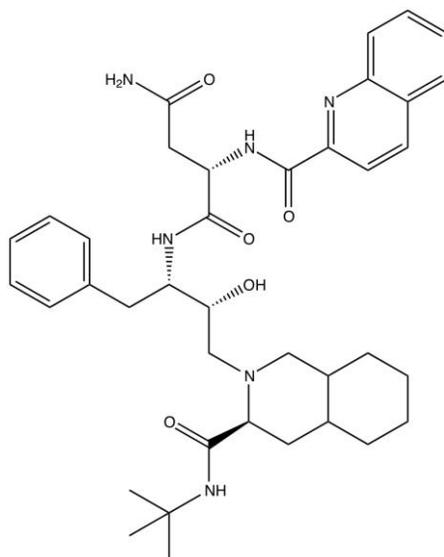


Figure 4-2. Schematic representation of the HIV-1 protease inhibitor saquinavir.

Forty-two patients failing treatment with saquinavir and ritonavir were discovered to possess the drug resistant mutations G48T and L89M. In this study we describe how the two mutations, G48T (active site) located in the flap tips and L89M (non-active-site) located in the  $\alpha$ -helix, reduce the efficacy of the protease inhibitor saquinavir, shown in Figure 4-1. Saquinavir, Figure 4-2, is a transition state analog inhibitor and was the first FDA approved HIV-1 protease inhibitor. Mutations at position 48 frequently occur as a result of treatment with saquinavir.<sup>162</sup>

In order to evaluate the mechanism of how these mutations diminish the efficacy of saquinavir we performed: 1  $\mu$ s of MD with saquinavir bound to both the wild-type and

mutant, 1  $\mu$ s of Steered Molecular Dynamics to investigate the disassociation pathway of saquinavir, and 1.8  $\mu$ s of MD of the wild-type and mutant in the apo form.

## 4.2 Methods

### 4.2.1 General Setup

The initial coordinates and sequence for the wild-type and the mutant were obtained from the x-ray crystal structure PDB ID 3OXC<sup>163</sup> (holo simulations) and PDB ID 1HHP<sup>128</sup> (apo simulations). An in-house python script was used to make the mutations to the protease sequence. The stabilizing mutations<sup>164,165</sup> (Q7K, L33I, L63I, C67A C95A) were removed from the protease sequence, Table A-1. The missing atoms from x-ray structure were added using the LEaP module in AmberTools13. The AMBER ff99SB<sup>99</sup> force field was used for the protein and the parameters for saquinavir were generated using the *antechamber* module and GAFF<sup>139</sup> force field with AM1-BCC<sup>140</sup> charges. Only one of the catalytic aspartates was protonated.

The explicit solvent was modeled using a 15 Å TIP3P<sup>103</sup> solvent buffer encapsulating the protein in a truncated octahedron box and chloride ions were added to force neutrality. The long-range electrostatics were treated with the particle-mesh Ewald method (PME)<sup>166</sup>, using direct space and van der Waals cutoff of 9 Å. The temperature was maintained at 310 K using Langevin Dynamics with a collision frequency of 1.0 ps<sup>-1</sup>. The SHAKE<sup>143</sup> algorithm was employed so that a 2 fs time step could be utilized.

### 4.2.2 Steered Molecular Dynamics

The equilibration procedure for the Steered Molecular Dynamics simulations was begun in the same fashion as previously described in the MD subsection. A 15 ns

explicit solvent MD simulation was performed in order to extract the snapshots necessary for the SMD simulations. The snapshots were extracted every 100 ps.

During the production phase the solvent was modeled implicitly using the Generalized Born (igb=8)<sup>167</sup> solvation model in AMBER in conjunction with the *mbondi3* intrinsic radii. No cutoff was used for the non-bonded interactions. The temperature was maintained at 310 K using the weak-coupling algorithm<sup>168</sup> (ntt=1) in AMBER. The reaction coordinate was the distance between the center of mass of the inhibitor and the protease (residues in the flaps 43-58 of each monomer were excluded). The force constant was 400 kcal/mol•Å<sup>-2</sup> and the pulling rate was 0.8 Å/ns.

#### **4.2.3 Molecular Mechanics Generalized-Born Surface Area Calculations**

The MMGBSA calculations were performed using MMPBSA.py<sup>122</sup> in AmberTools13. The polar solvation energy was calculated using the Generalized Born Method (igb=8) in conjunction with the *mbondi3* intrinsic radii. The non-polar solvation energy was calculated using the LCPO<sup>105</sup> method. The quasi-harmonic approximation was used to account for the entropic contributions.

#### **4.2.4 General Analysis**

Analyses of the trajectories were performed using the *cpptraj* module of AmberTools13 and VMD.<sup>169</sup> All MD simulations were performed using *pmemd.cuda*<sup>119</sup> in AMBER 12.<sup>170</sup>

### **4.3 Results and Discussions**

#### **4.3.1 Inhibitor Simulations**

##### **4.3.1.1 Flexibility**

To evaluate the differences in flexibility of the wild-type and mutant the root mean squared fluctuations of the C $\alpha$  atoms were measured and the difference between the

wild-type and mutant (positive value indicates that the wild-type residue fluctuates more and a negative value indicates that the mutant residue fluctuates more), Figure 4-3. The largest differences between the wild-type and mutant are in the 70's strand and the flap tip residues (48-53). The 70's strand of the mutant is not as flexible when compared to the wild-type because of the larger side-chain Met89 in the mutant increases the number of van der Waals contact with residues belonging to the 70's strand thus reducing its flexibility. The flap tips of the mutant exhibit more flexibility than the wild-type, which will be explained in more detail later.

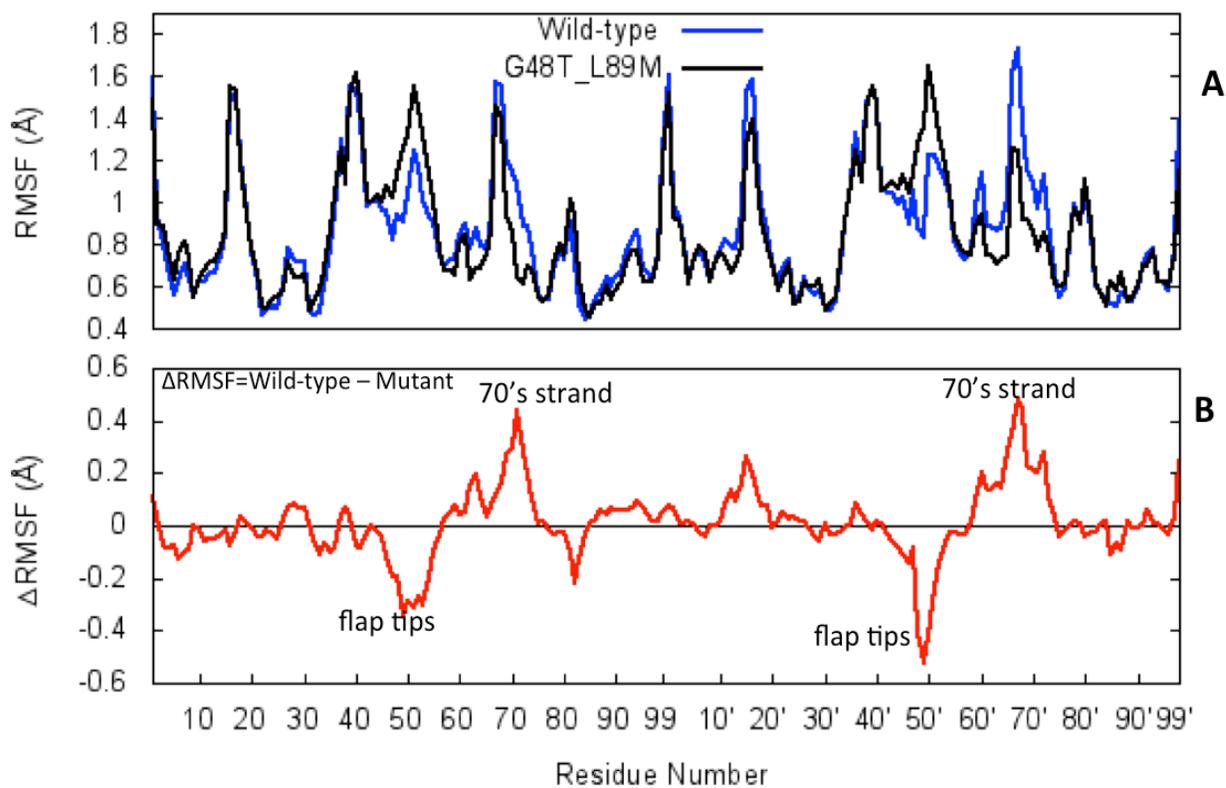


Figure 4-3. Atomic Fluctuations of the wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir. A) Plot of the RMSF. B) Plot of the difference between the wild-type and mutant (G48T/L89M). Positive value signifies the wild-type fluctuates more and a negative value indicates the mutant fluctuates more. Monomer A (1-99) Monomer B (1'-99').

To probe the conformation of the flaps, we plotted the values of the root mean squared deviations of the C $\alpha$  in the flaps (43-58) versus the minimized crystal structure, Figure 4-4. The maximum peak of the wild-type is shifted to the left of the maximum peak of the mutant, and the width of the distribution is broader for the mutant, indicating that the flaps of the wild-type are less flexible when compared to the mutant.

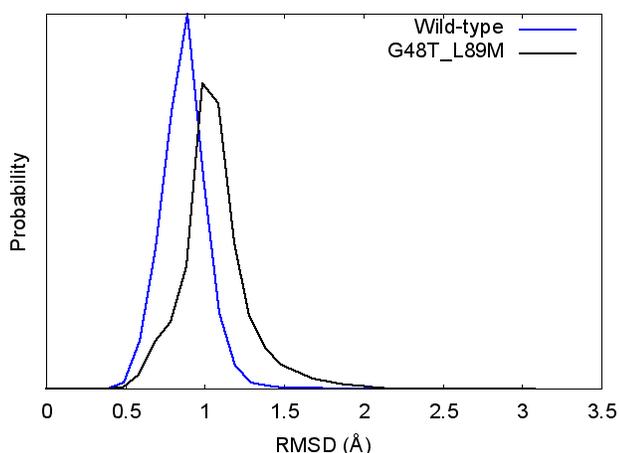


Figure 4-4. Histogram of the RMSD of the C $\alpha$  atoms of the flap residues (43-58) of each monomer. MD simulations of the wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir.

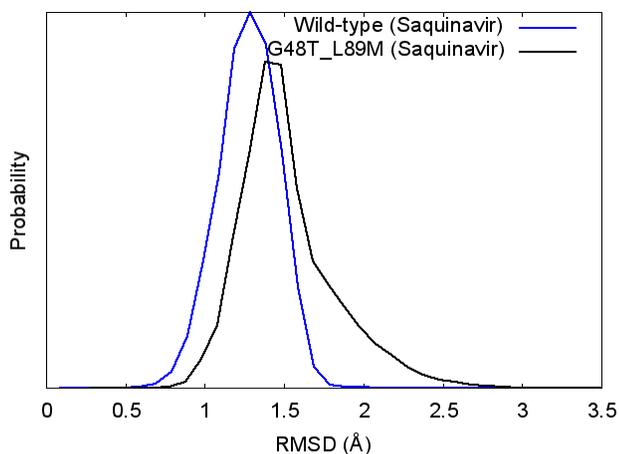


Figure 4-5. Histogram of the RMSD of saquinavir. MD simulations of the wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir.

To assess the flexibility of saquinavir, the RMSD of saquinavir versus the minimized crystal structures was calculated, and to make the distinction between the two clearer, the data was histogrammed and plotted in Figure 4-5. The RMSD distribution in Figure 4-5 is much broader for saquinavir in the mutant, which is indicative of saquinavir possessing more flexibility in the active site of the mutant relative to the wild-type.

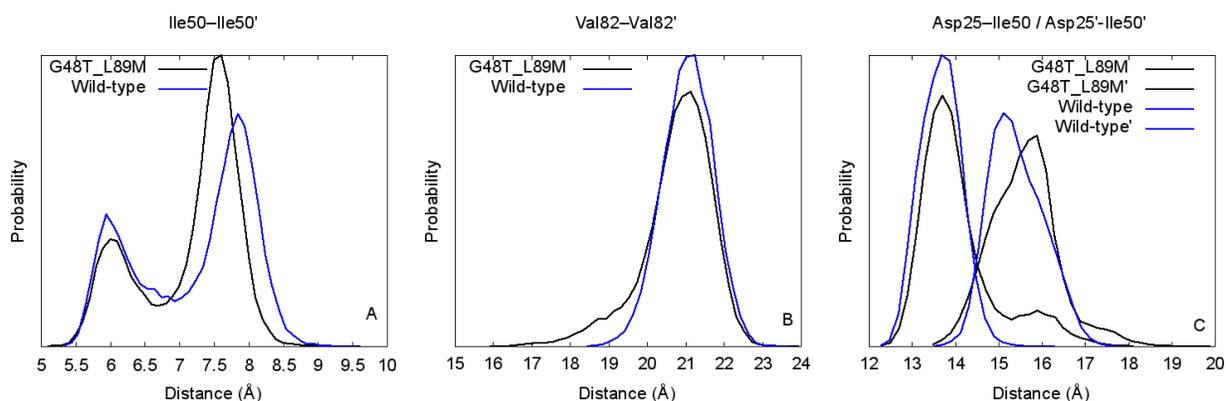


Figure 4-6. Histograms of the different distances in the active site of the wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir. A) Ile50-Ile50' distance B) Val82-Val82' distance C) Asp25-Ile50/Asp25'-Ile50' distance

The distance between Ile50 C $\alpha$  of each monomer for the wild-type and mutant were measured and plotted in order to quantify the flap tip separation, Figure 4-6 A. This distance provides insight as to the openness of the flaps of HIVPR. The wild-type and the mutant exhibit a peak at 6 Å. However, the maximum peak of the wild-type is 8 Å and the 7.5 Å for the mutant, suggesting that the flap tips separation in the wild-type is slightly larger. In Figure 4-6 B, the distance between the active site residue Val82 of each monomer was measured, to assess the width of the active site cavity. This distance fluctuates less in the wild-type because the width of distribution is smaller than the mutant. There is a small tail seen between 17-19 Å in the mutant that is not present

in the wild-type. This signifies that occasionally the distance between Val82 and Val82' can become smaller in the mutant than the wild-type. Lastly, the distance between the catalytic Asp25 and Ile50 of monomer was measured and plotted, Figure 4-6 C. This distance provides insight as to how large the active site is vertically. The distance distribution in the mutant is broader than that of the wild-type. The flap tips of the wild-type point downward towards the active site thus increasing the flap tip separation while decreasing the distance between the Asp25 and Ile50.

#### 4.3.1.2 Hydrogen bonding analysis

Table 4-1. Hydrogen bonding analysis of the wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir.

Acceptor	Donor	% Occupancy
Wild-type		
ASP'—O	SQV199 O—H	97
GLY48—O	SQV199 N—H	57
SQV199—O	ASP25 O—H	39
Water Mediated		
SQV199—Asp29'		18
G48T_L89M		
ASP25'—O	SQV199 O—H	90
THR48—O	SQV199 N—H	23
Water Mediated		
SQV199—ASP29'		11

The direct and water-mediated hydrogen bonds that saquinavir forms with residues in the active site were investigated to determine if the mutations cause any differences, Table 4-1. The catalytic aspartates of the wild-type, Asp25 (protonated) and Asp25' (deprotonated) interact with the central hydroxyl group of saquinavir more in the wild-type when compared to the mutant. Furthermore, the Gly48 backbone carbonyl of the wild-type forms a hydrogen bond more frequently with saquinavir than Thr48 of the mutant. Both wild-type and mutant form a water-mediated hydrogen bond between saquinavir and Asp29.

Blackburn *et al.*<sup>171</sup> demonstrated that an inhibitor's ability to close the flaps correlates with the number of non-water mediated hydrogen bonds between the inhibitor and protease and our data corroborates this story as shown in Table 4-1. The flaps are more closed-like in the wild-type and the ligand is more rigid, therefore the saquinavir is forming more interactions with the protein in the wild-type than in the mutant. The G48T mutation appears to reduce the number of interactions with the flaps therefore increasing the flexibility of the flaps.

#### **4.3.1.3 MMGBSA calculations**

The MMGBSA calculations in Table 4-2 reveal that there is approximately a 2 kcal/mol difference between the binding affinities of the wild-type and the mutant. The enthalpic contributions is ~7 kcal/mol higher for the wild-type than the mutant. This is indicative of the residues in the active site of the wild-type forming more favorable interactions (electrostatic, van der Waals) with the ligand. The solvation free energy is ~4 kcal/mol lower in the mutant. The difference between the entropic contributions for the wild-type and mutant are negligible.

Table 4-2. Results of MMGBSA calculations for the wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir. All units in kcal/mol

System	$\Delta H_{\text{gas}}$	$\Delta G_{\text{solvation}}$	$T\Delta S_{\text{quasi-harm}}$	$\Delta G_{\text{bind}}$
Wild-type	$-106.79 \pm 0.05$	$51.94 \pm 0.04$	-23.05	$-31.79 \pm 0.02$
G48T_L89M	$-99.22 \pm 0.05$	$47.72 \pm 0.03$	-22.38	$-29.12 \pm 0.02$

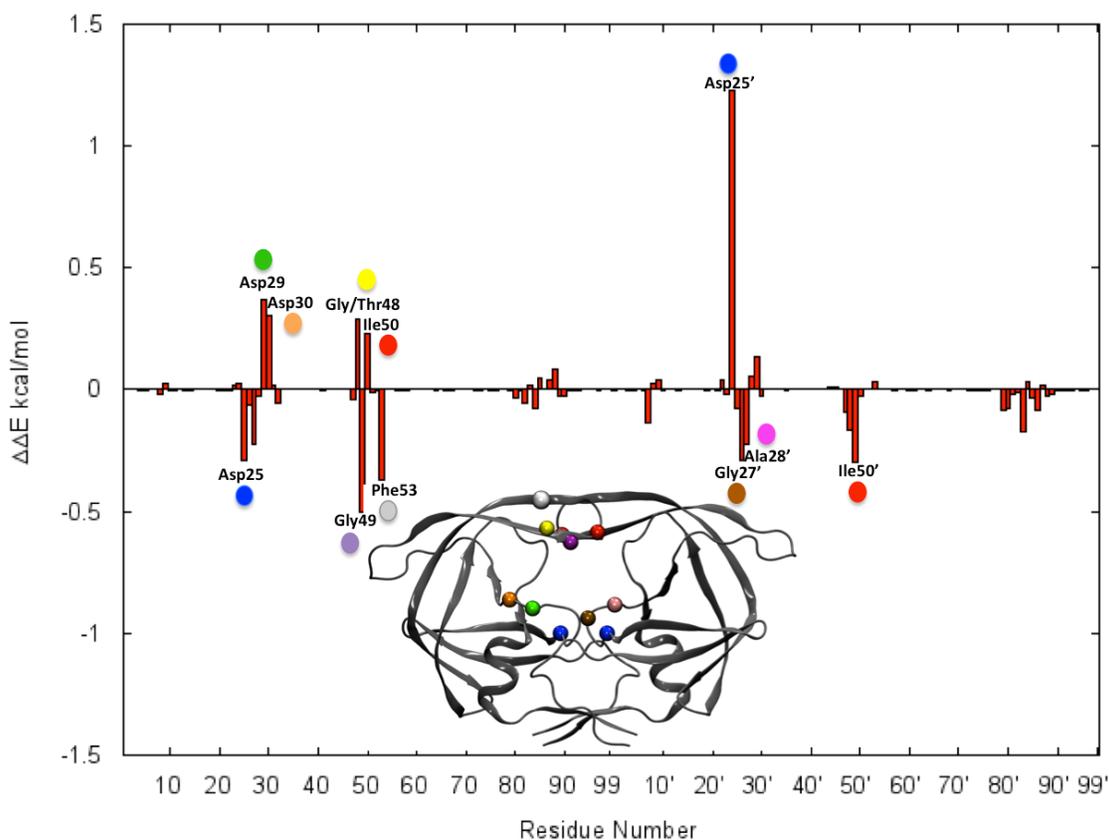


Figure 4-7. Per-residue energy decomposition of the binding affinity of the wild-type and the mutant (G48T/L89M), subtype B bound to saquinavir. The difference between the wild-type and mutant is shown, where a positive  $\Delta\Delta$  binding free energy indicates that the mutant (G48T/L89M) has a more favorable contribution and vice-versa for the wild-type. Monomer A (1-99), Monomer B (1'-99'). The residues in the plot are highlighted on HIVPR Asp25/Asp25' (blue), Asp30 (green), Asp29 (orange) Gly/Thr48 (yellow), Gly49 (purple), Ile50/Ile50' (red), Gly27' (brown), Ala28' (pink)

The binding free energy was decomposed into contributions from the individual residues of the protein to gain insight into the role of particular residues in the binding

mechanism of saquinavir. The difference between the wild-type and mutant is plotted in Figure 4-7, to make the comparison clearer. One of the main differences between the wild-type and mutant is the energetic contributions of the catalytic aspartates. Asp25 in the wild-type contributes more to the binding free energy because it forms more favorable electrostatic interactions with saquinavir. On the other hand, Asp25' in the mutant has a higher contribution to the binding free energy than Asp25' in the wild-type.

Asp25' in the mutant interacts less with the hydroxyl group of saquinavir therefore reducing its desolvation penalty. Asp29 of the mutant forms more electrostatic interactions with saquinavir via a hydrogen bond between an oxygen atom of saquinavir and backbone amide group—Asp30 in the mutant forms more favorable van der Waals contacts with saquinavir. Residue Gly27' and Ala28' of the wild-type forms more van der Waals interactions with saquinavir than the mutant. The interactions between saquinavir and flap residues Gly49, Phe53, Gly48', Gly49' Ile50' of the wild-type contribute more to the binding energy. Phe53 interacts strongly with saquinavir in the wild-type, because of pi-stacking and this same interaction is not formed in the mutant. The difference in the energetic contribution of Gly49, Gly48', Gly49' and Ile50' is due to an increased van der Waals and electrostatic interactions with saquinavir. Thr48 and Ile50 of the mutant have a larger energetic contribution. The larger contributions are a result of the orientation of the quinoline ring of saquinavir.

#### **4.3.1.4 Steered Molecular Dynamics**

Steered Molecular Dynamics simulations were performed to investigate what effect the G48T mutation in the flaps had on the escape pathway of saquinavir. The work profiles of the various SMD simulations are shown in Figure 4-8. The binding free energy profiles of the SMD simulations are plotted in Figure 4-8. The free energy

difference between the mutant and wild-type is approximately 2 kcal/mol, which is comparable to the difference that the MMPBSA results yielded. Figure 4-9 shows that it takes more energy to remove saquinavir from the active site of the wild-type because saquinavir forms more favorable interactions with the residues in the active site of the wild-type than the mutant. These results agree very well from those obtained in MMGBSA that demonstrated that the G48T and L89M reduces the affinity for saquinavir, therefore, the drug is easier to remove from the active site.

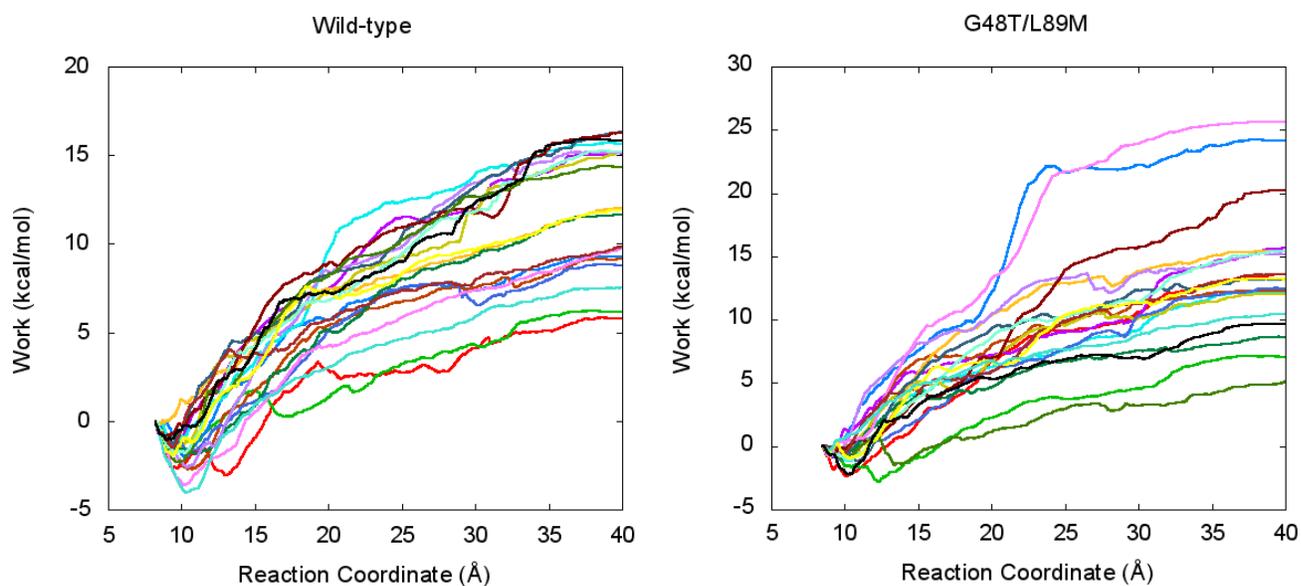


Figure 4-8. Work Profiles from the SMD simulations for wild-type and the mutant (G48T/L89M), subtype B, bound to saquinavir.

All of the ligands escape the active site by the flaps opening; no lateral escapes were seen, in which the flaps do not have to fully open. After the ligand exits the active site the flaps remain predominantly in the semi-open conformation. The SMD simulations revealed that the flaps need to open at least 35 Å (distance between Ile50 and Ile50') for saquinavir to escape the active site.

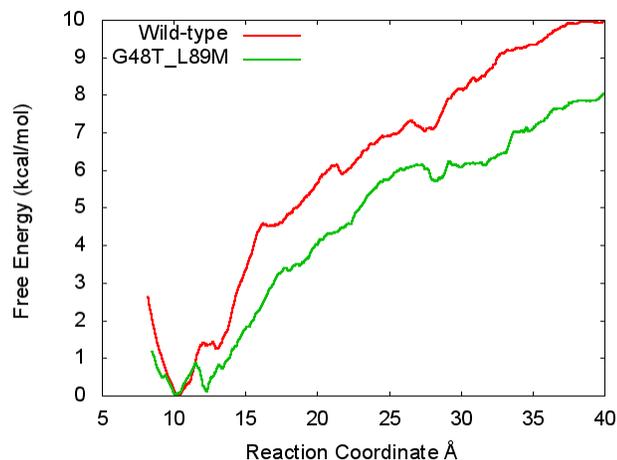


Figure 4-9. Free energy profile of removing the protease inhibitor saquinavir from the active site cavity of the wild-type and the mutant (G48T/L89M), subtype B.

### 4.3.2 Apo Simulations

Simulations of the apo (unbound) conformations were investigated in order to determine what effects the G48T and L89M had on the conformation of the flaps and other regions of the protease. Previously studies have demonstrated that mutations can alter the equilibrium between the open and closed conformations of the flaps.<sup>90,137,172</sup>

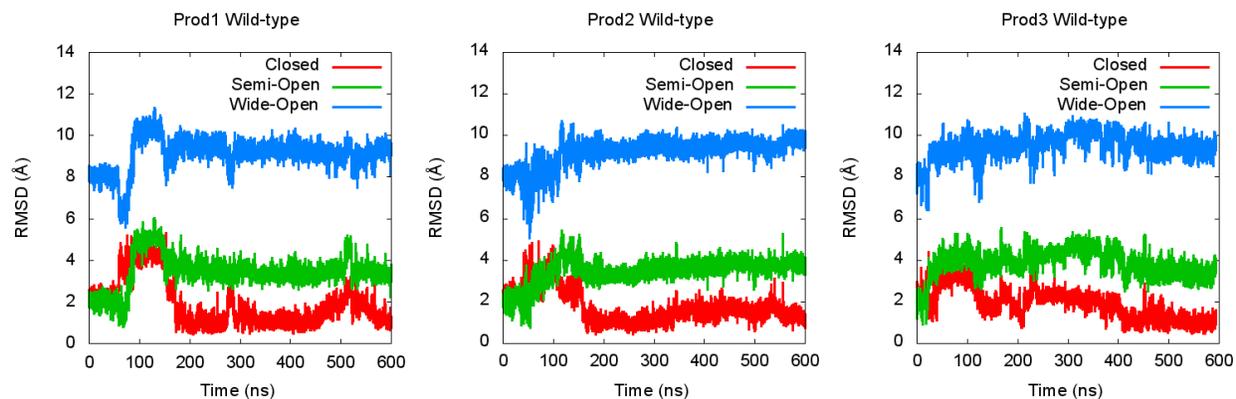


Figure 4-10. Plot of the RMSD of wild-type subtype B, flap residues, 43-58, versus the closed, semi-open and wide-open conformations. MD simulations performed in the apo form.

The rmsd of the dynamics of the flaps, residues 43-58 of each monomer of the wild-type and mutant versus the closed (PDB ID 3OXC), semi-open (PDB ID 1HHP),

and wide-open (obtained from a previous MD simulations unpublished) conformations are shown in Figures 4-10 and 4-11. A schematic of the three reference structures is shown in Figure 1-7. The lower the rmsd signifies that the flaps favor that particular structure as opposed to the others. The dynamics of the flaps for the wild-type began in semi-open conformation but as the simulation progresses the flaps become closed and no other flap conformation is visited. There are instances where the rmsd of both the closed, semi-open, wide-open conformations are higher than 4 Å and in such cases the flaps were in tucked/curled conformation. However, the flap dynamics of the mutant favor predominantly the semi-open conformation in the early stages of simulation. However as the simulation progresses to ~400 ns the flaps in all three simulations favor the tucked/curled conformation. The wide-open conformation is never seen in either the

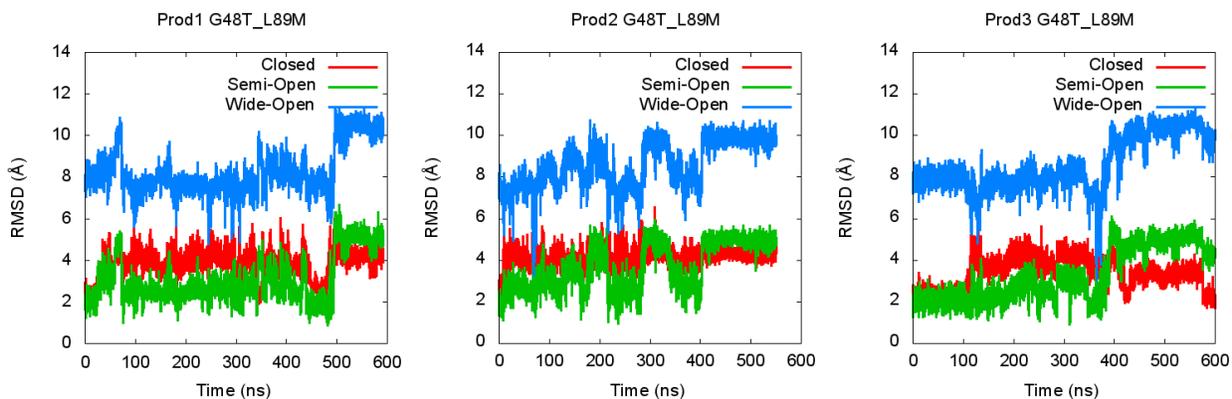


Figure 4-11. Plot of the RMSD of mutant (G48T/L89M) subtype B, flap residues, 43-58, versus the closed, semi-open and wide-open conformations. MD simulations performed in the apo form.

wild-type or the mutant simulations. The flaps of wild-type transition from the semi-open to the closed conformation and remain there for the entirety of the simulation, which could be a result of kinetic trapping.

The values of the RMSF of the C $\alpha$  atoms for the wild-type and mutant, and the difference between them are plotted in Figure 4-12. The largest deviations are in the residues 43-58 of each monomer, which are the flap tips. This difference is due to the of the flaps of the wild-type remaining in the closed conformation for the majority of the simulations. The largest differences between the wild-type and mutant are in the flap tips and the 70's strand. Figure 4-12 exhibits the same trend as the saquinavir bound MD simulations.

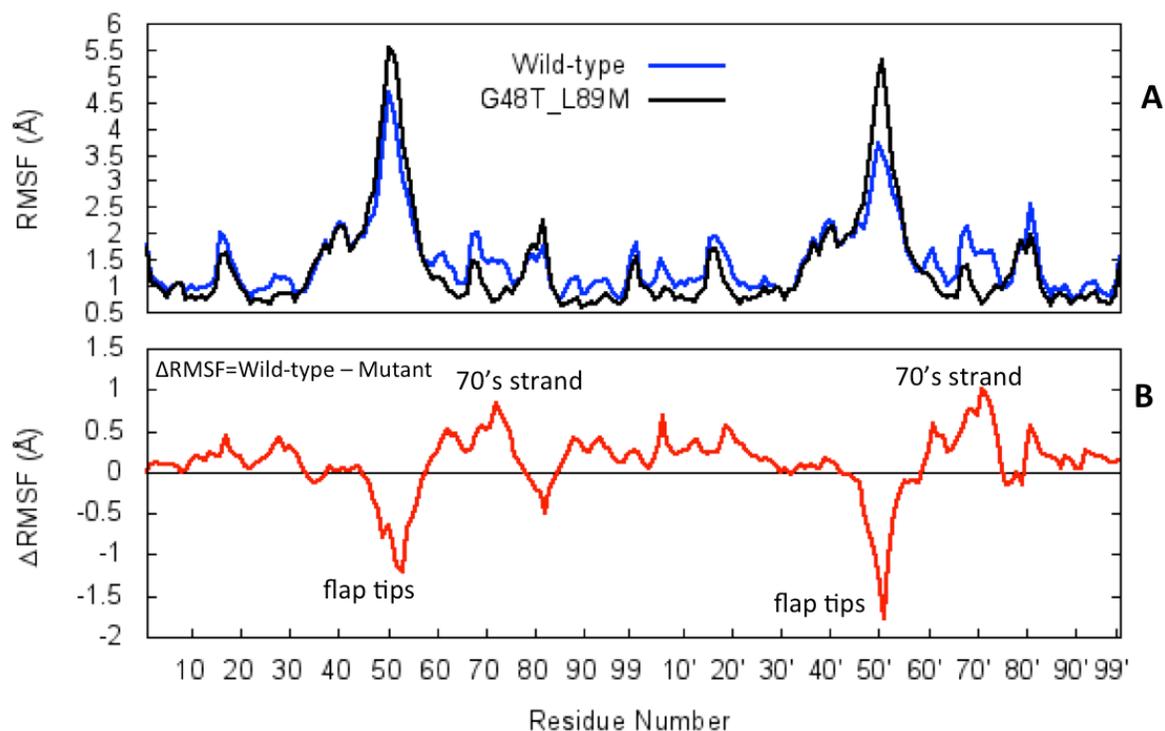


Figure 4-12. Atomic Fluctuations of the wild-type and the mutant (G48T/L89M), subtype B, apo. A) Plot of the RMSF. B) Plot of the difference between the wild-type and mutant (G48T/L89M). Positive value signifies the wild-type fluctuates more and a negative value indicates the mutant fluctuates more. Monomer A (1-99) Monomer B (1'-99').

### 4.3.3 Drug Resistant Mechanism

The G48T mutation causes a conformational change in the quinoline ring of saquinavir. This conformational change reduces the ability of the quinoline ring in the

mutant to form pi-stacking interactions with Phe53, Figure 4-13. This same observation has been made as a result of the G48V mutation and has been well documented.<sup>173,174</sup> The side-chain hydroxyl group of Thr48 possesses the ability to form a water-mediated hydrogen bond with saquinavir that Val48 is unable to form. The backbone oxygen of the Gly48 residue in the wild-type forms a hydrogen bond with saquinavir approximately 20% more than the Thr48 residue of the mutant.

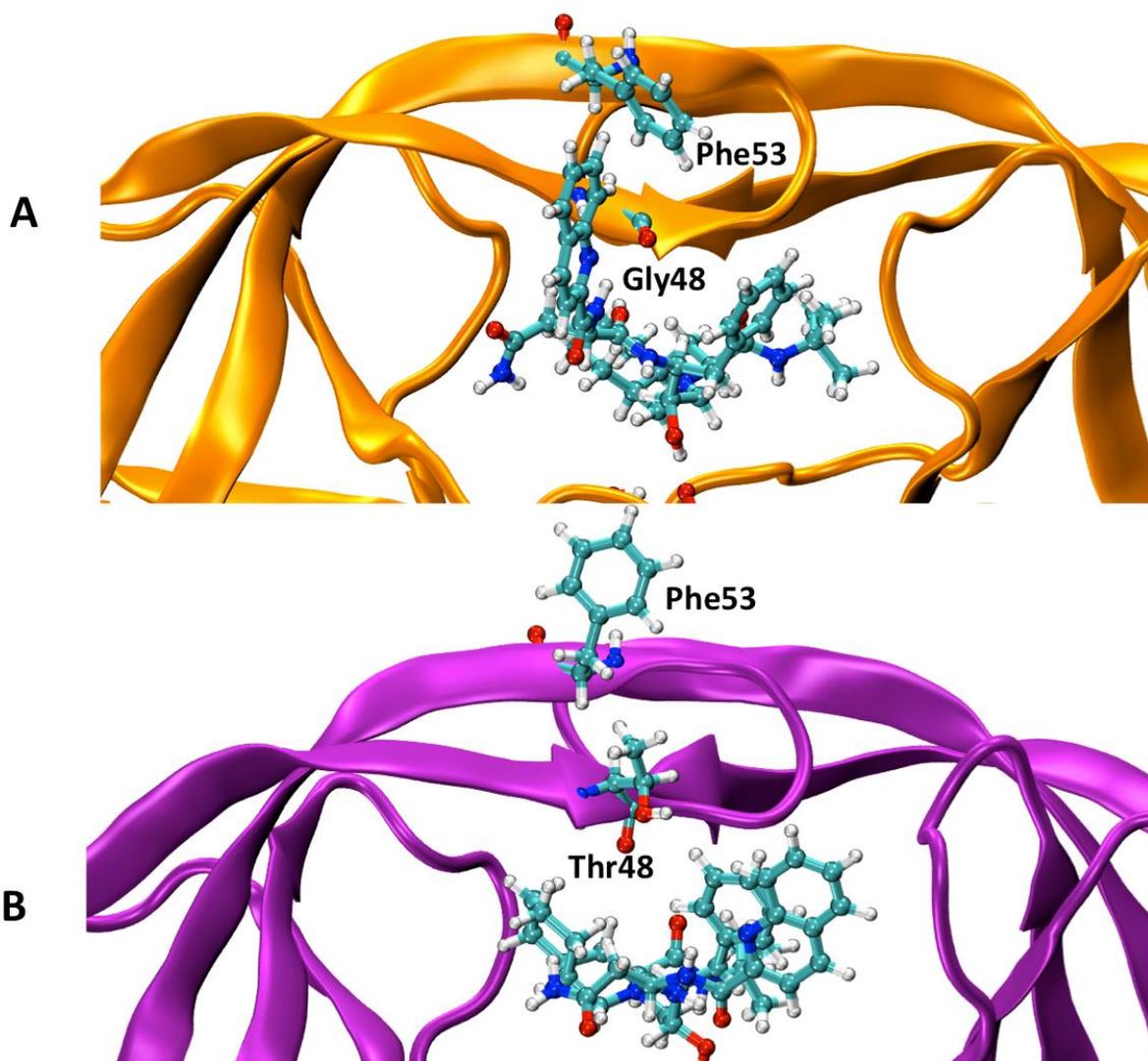


Figure 4-13. Schematic of the Phe53 pi-pi stacking interaction with the quinoline ring of saquinavir. A wild-type and the B) mutant. Residue Thr48 of the mutant reduces the interaction between Phe53 and the quinoline ring of saquinavir.

The interaction of the central hydroxyl groups of saquinavir with the catalytic aspartates is diminished in the mutant by almost 10%. The diminished interactions between the catalytic aspartates and the hydroxyl group of saquinavir might be attributed to the L89M mutation although it is not obvious. Sanches *et al.*<sup>175</sup> proposed that the L89M mutation causes the L90 residue to shift and form more van der Waals interactions with residues close in vicinity to Asp25. No shift of the L90 was observed during the course of our MD simulations—pair-wise energy decomposition calculations revealed that there were no differences in van der Waals interactions of Leu90 in wild-type or mutant. The side chain of the larger Met residue does form more van der Waals contacts with the 70's strand (Ala71, Ile72, Gly73, Thr74, Val75). The flexibility of the 70's strand in mutant was reduced as a result of these increased interactions.

#### **4.4 Conclusions**

Through the use molecular dynamics we have described how the G48T and L89M mutations reduce the efficacy of saquinavir on an atomic level. The MMGBSA calculations and SMD simulations demonstrate that the binding affinity of the mutant is lowered and it takes less energy for saquinavir to be removed from the active site of the mutant. The major effects of the mutations are: the G48T mutation causes the reorientation of the quinoline ring that reduces the number of interactions between saquinavir and residues in the flaps, specifically Phe53 and the L89M mutation has been implicated in causing the reduced interactions of the central hydroxyl group of saquinavir with the catalytic aspartates, though further investigation is needed. The flaps are more closed-like in the wild-type indicating that the saquinavir makes more direct hydrogen bonds with residues in the active site.

Saquinavir when bound to the mutant or the wild-type forms a water-mediated hydrogen bond with Asp29' suggesting that a modification such that saquinavir forms a direct bond with Asp29' might increase its potency.<sup>176</sup> These results can aid in the design of new inhibitors to become more potent against drug-resistant proteases.

CHAPTER 5  
EVALUATING THE EFFECTIVENESS OF SAQUINAVIR AND TIPRANA VIR ON HIV-1  
SUBTYPE C WILD-TYPE AND THE DRUG RESISTANT N88D/L90M MUTANT  
PROTEASE

**5.1 Introduction**

Subtype C accounts for the majority of all the HIV infections worldwide.<sup>20</sup> The current anti-retroviral drugs used to combat HIV, more specifically the FDA-approved protease inhibitors have been developed using subtype B. Fortunately these protease inhibitors have been proven to still inhibit subtype C, although the efficacy of these protease inhibitors when compared to subtype B is diminished.<sup>177-179</sup>

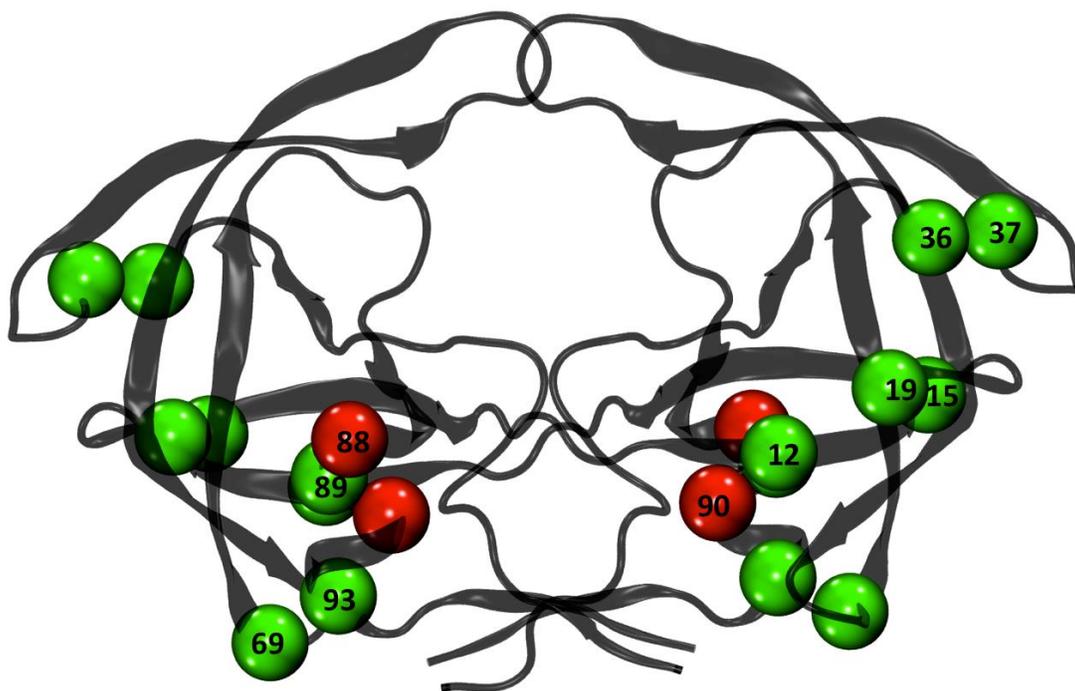


Figure 5-1. Schematic of HIVPR subtype C. Green spheres indicate where the sequences of subtypes B and C differ, (T12S, I15V, L19V, M36I, S37A, H69K, L89M, I93L). Red spheres illustrate the positions of the drug resistant mutations N88D and L90M

The first crystal structure of Subtype C protease was not determined until 2008.<sup>177,180</sup> The sequences of subtype C and B differ by only eight residues, T12S,

I15V, L19V, M36I, S37A, H69K, L89M, I93L. These are non-active site mutations that are located in the elbow region of the flaps, and the hydrophobic core, Figure 5-1.

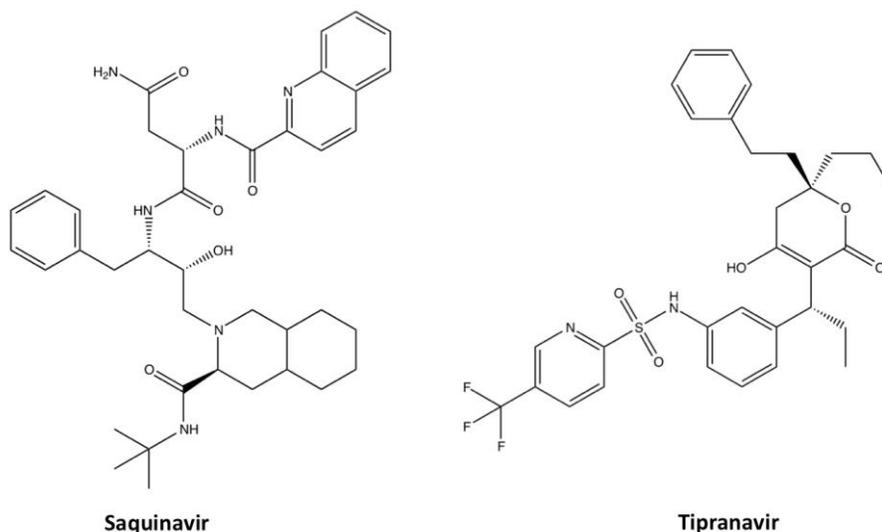


Figure 5-2. Schematic of the protease inhibitors saquinavir and tipranavir.

As mentioned in previous chapters the introduction of protease inhibitors causes drug selected pressure mutations that reduce the efficacy of the protease inhibitors. In this study we investigate how the non-active site mutations N88D, and L90M, diminish the effectiveness of saquinavir, Figure 5-2. The L90M mutation reduces the efficacy of all the protease inhibitors except darunavir and tipranavir.<sup>87</sup> The N88D mutation is a facilitator in the co-occurrence of D30N and L90M mutations, causing multidrug resistance.<sup>181</sup> The single N88D mutant increases the dimer stability.<sup>182</sup> The single L90M mutant increases the catalytic ability<sup>129,131,183</sup> of the protease even though it destabilizes the stability of the dimer.<sup>182</sup> The stability of the double mutant N88D/L90M is moderate when compared to the wild-type and demonstrated a higher catalytic ability than the wild-type for certain substrates.<sup>177,182</sup>

Coman *et al.*<sup>177</sup> determined the  $K_i$  values for the wild-type and several drug resistant mutants for subtype C—the N88D/L90M double mutant exhibited 42-fold

increase in  $K_i$  against saquinavir, while same double mutant displayed only 1.3-fold increase against tipranavir. The aim of this study is to explain how the N88D and L90M mutations reduce the efficacy of the protease inhibitors saquinavir and tipranavir.

## 5.2 Methods

### 5.2.1 General Setup

The initial coordinates and sequence for the wild-type and mutant were derived from the x-ray crystal structure PDB 2R5Q.<sup>177</sup> The stabilizing mutations were removed from the sequences prior to performing any MD simulations, Table A-3. The missing atoms from the x-ray structure were added using the Leap module in AmberTools13. The AMBER ff99SB<sup>99</sup> force field was used for the protein. The parameters for saquinavir and tipranavir were constructed using the AM1-BCC<sup>140</sup> charges generated from the antechamber module and the GAFF<sup>139</sup> force field. The mono-protonated state was chosen for the catalytic aspartates.

The explicit solvent was modeled using a 15 Å TIP3P<sup>103</sup> solvent buffer encapsulating the protein in a truncated octahedron box and chloride ions were added to force neutrality. The long-range electrostatics were treated with the PME method<sup>166</sup>, using direct space and van der Waals cut off of 9 Å. Langevin Dynamics was utilized to maintain the temperature at 310 K—the collision frequency was 1.0 ps<sup>-1</sup>. In order to achieve a 2 fs time step all atoms bonded to hydrogens were constrained using the SHAKE<sup>143</sup> algorithm.

### 5.2.3 Molecular Dynamics

The hydrogen atoms were minimized followed by the entire system. Each minimization process consisted of 500 steps of steepest descent followed and 10,000 steps of conjugate gradient. The systems were heated linearly from 100 K to 310 K over

a 1 ns MD simulation, where a 5 kcal/mol·Å<sup>2</sup> positional restraint was placed on the solute. Following heating, the density was equilibrated for 500 ps. Next, five simulations were performed where the positional restraint was reduced from 5 kcal to .1 kcal/mol·Å<sup>2</sup>. In the last step of equilibration all restraints were removed and 5 ns of unrestrained MD were performed.

The production phase consisted of running 20 x 51 ns simulations, but only the last 50 ns was used for analysis. Snapshots of the MD trajectory were saved every 30 ps.

### **5.2.3 Molecular Mechanics Generalized-Born Surface Area Calculations**

The MMGBSA calculations were performed using *MMPBSA.py*<sup>122</sup> in AmberTools13. The polar solvation energy was calculated using the Generalized Born Method (igb=8)<sup>167</sup> in conjunction with the *mbondi3* intrinsic radii. The non-polar solvation energy was calculated using the LCPO<sup>105</sup> method. The quasi-harmonic approximation was used to account for the entropic contributions.

## **5.3 Results and Discussion**

### **5.3.1 Saquinavir-bound MD Simulations**

#### **5.3.1.1 Flexibility**

The RMSF of the carbon  $\alpha$  atoms and the difference between wild-type and mutant is plotted in Figure 5-3. The mutant exudes a greater flexibility especially in the flap tips residues 48-53 of monomer A (residues 1-99). There are differences between the two in the  $\alpha$  helix residues 86-92 of monomer B (residues 100-198). The active site residues Ala28' and Asp29' display more fluctuations in the mutant than the wild-type.

A histogram of the RMSD of saquinavir versus the minimized crystal structure is shown in Figure 5-4. The RMSD distribution of saquinavir in the mutant is much broader

than that of the saquinavir in the wild-type. This suggests that saquinavir in the mutant is not interacting in a similar fashion as when it is bound to the wild-type, thus the flexibility of the ligand is increased.

Plotted in Figure 5-5 is a histogram of the RMSD of the flaps (43-58) of each monomer versus the minimized crystal structure (closed). The flaps of the wild-type are more “closed-like” than the mutant. In the mutant there is a small tail between 1.5 and 2 Å that is not present in the wild-type. Furthermore, this plot reaffirms the notion that the mutant exhibits more flexibility as was demonstrated in Figure 5-3.

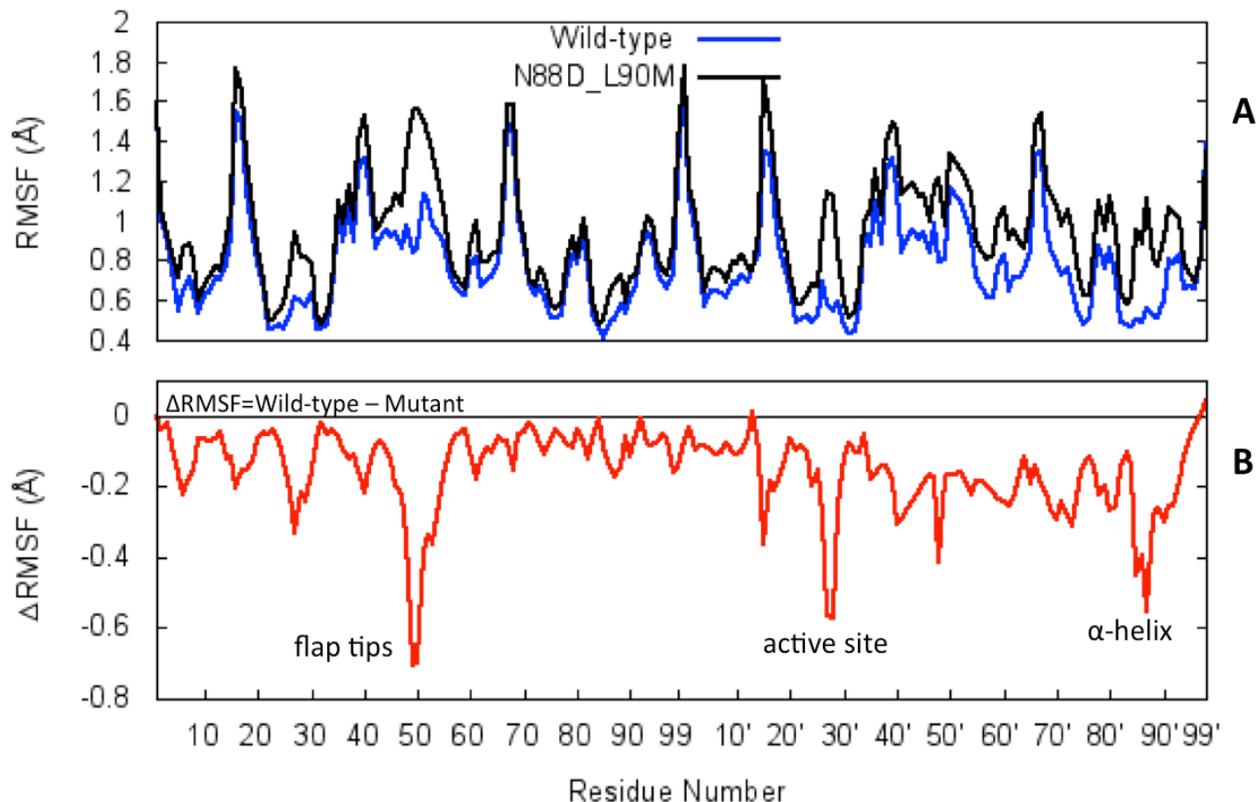


Figure 5-3. Atomic Fluctuations of the wild-type and the mutant (N88D/L90M), subtype C, bound to saquinavir. A) Plot of the RMSF. B) Plot of the difference between the wild-type and mutant (N88D/L90M). Positive value signifies the wild-type fluctuates more and a negative value indicates the mutant fluctuates more. Monomer A (1-99) Monomer B (1'-99').

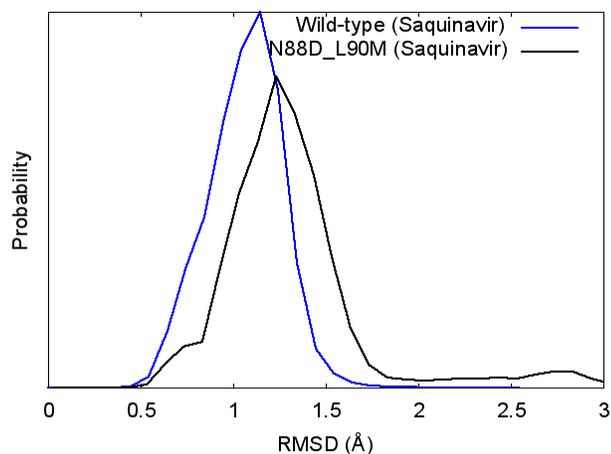


Figure 5-4. Histogram of the RMSD of saquinavir. MD simulations of the wild-type and mutant (N88D/L90M), subtype C, bound to saquinavir.

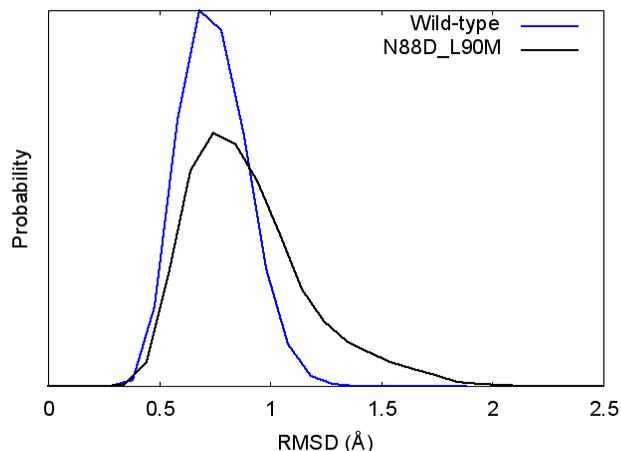


Figure 5-5. Histogram of the RMSD of the flap residues (43-58) of each monomer. MD simulations of the wild-type and the mutant (N88D/L90M), subtype C, bound to saquinavir.

The distance between the carbon  $\alpha$  atoms of Ile50 of each monomer was measured and plotted to determine the flap tip separation, Figure 5-6 A. The Ile50 distance for the wild-type has one large peak at 6 Å and smaller peak at 8 Å. The Ile50 distance for the mutant has a maximum peak at 6 Å and second peak that ranges from 7-8 Å. In order to gain further insight on the size of the active site, the distances between the residues Val82 and Val82', Figure 5-6 B, and Asp25-Ile50, of each

monomer, Figure 5-6 C were measured. In Figure 5-6 B, the maximum peak of the mutant is shifted to left of the maximum peak of the wild-type—this is indicative that the distance in between in Val82 and Val82' is shorter in the mutant. Furthermore, there is a tail between 17-20 Å that is located in the mutant but is not as prevalent in the wild-type. In Figure 5-6 C, both the wild-type and mutant exhibit peaks at 13 and 15 Å but the wild-type peaks are narrower than those of the mutant. The reason why Asp25-Ile50 and Asp25'-Ile50' distance distributions are not the same is due to the asymmetry of the ligand.

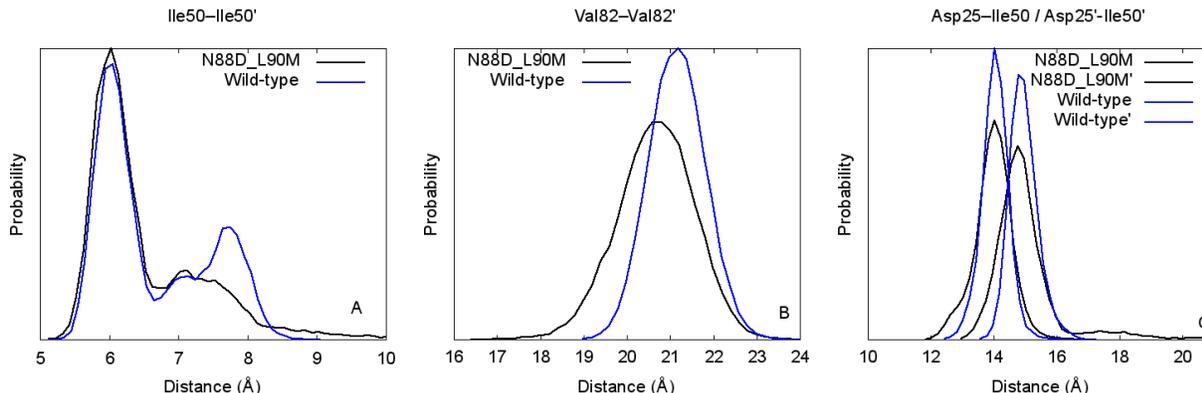


Figure 5-6. Histograms of the different distances in the active site of the wild-type and the mutant (N88D/L90M), subtype C, bound to saquinavir. A) Ile50-Ile50' distance B) Val82-Val82' distance C) Asp25-Ile50/Asp25'-Ile50' distance

### 5.3.1.2 Hydrogen bonding analysis

In order to assess how the mutations N88D and L90M affected the interactions of saquinavir with residues in the active site the direct and the water-mediated hydrogen bonds were calculated, Table 5-1. The central hydroxyl group of saquinavir in the wild-type interacts very strongly with the catalytic aspartates (Asp25/Asp25') while the mutant catalytic aspartates interacts with the central hydroxyl group of saquinavir significantly less.

Table 5-1. Hydrogen bonding analysis of the wild-type and the mutant (N88D/L90M), subtype C, bound to saquinavir

Acceptor	Donor	%Occupancy	
		Wild-type	N88D/L90M
Asp25'—O	SQV O—H	98	86
SQV—O	Asp25 O—H	81	41
Gly48—O	SQV N—H	52	38
Gly27—O	SQV N—H	26	21
Water Mediated			
SQV199—Asp29'		18	11

The most noticeable difference is the hydrogen bond between Asp25 (protonated) and saquinavir, which in the mutant is formed ~50% less. The difference between the frequency of hydrogen bonding between Asp25' (deprotonated) is not as extreme and the wild-type forms this hydrogen approximately 12% more. The backbone carbonyl oxygen of Gly48, located in the flaps, interacts less with saquinavir in the mutant than in the wild-type. Lastly, Gly27, located at the base of the active site, interacts only slightly more with saquinavir than the mutant. Residue Asp29' in both the wild-type and mutant forms a hydrogen bond with saquinavir via a water molecule.

### 5.3.1.3 MMGBSA calculations

Saquinavir binds more favorably to the wild-type and there is roughly a 4 kcal/mol difference between the two, Table 5-2. The  $\Delta H_{\text{Gas}}$  term of the wild-type is roughly 10 kcal/mol larger and that is indicative of it forming stronger interactions with saquinavir than the mutant. The desolvation penalty is much lower in the mutant—this is due to the

mutant not displacing as many waters as wild-type in the active site. The difference in the entropy between the two systems is negligible.

Table 5-2. Results of the MMGBSA calculations for the wild-type and the mutant (N88D/L90M), subtype C, bound to saquinavir. All units are in kcal/mol

System	$\Delta H_{\text{gas}}$	$\Delta G_{\text{solvation}}$	$T\Delta S_{\text{quasi-harm}}$	$\Delta G_{\text{bind}}$
Wild-type	$-125.60 \pm 0.08$	$65.27 \pm 0.05$	-23.88	$-36.43 \pm 0.03$
N88D_L90M	$-114.87 \pm 0.08$	$58.09 \pm 0.05$	-24.52	$-32.24 \pm 0.04$

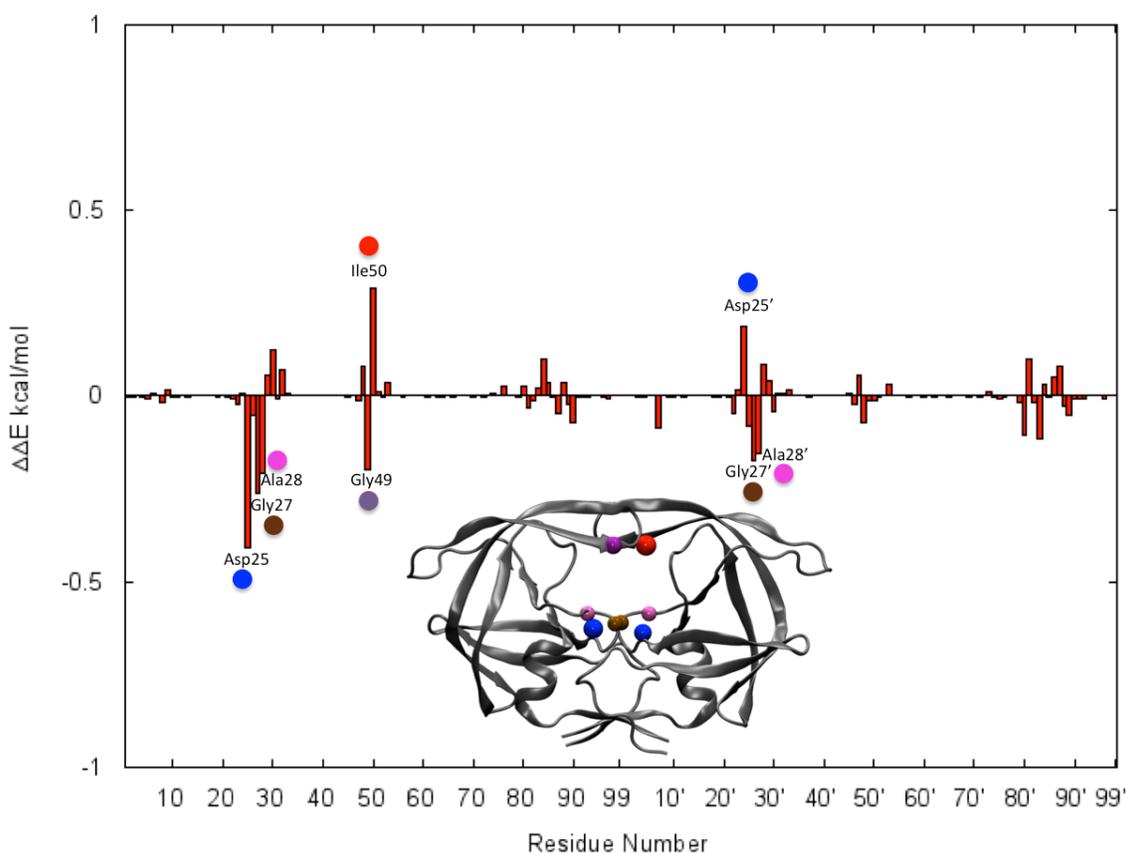


Figure 5-7. Per-residue energy decomposition of the binding affinity of the wild-type and the mutant (N88D/L90M), subtype C bound to saquinavir. The difference between the wild-type and mutant is shown, where a positive  $\Delta\Delta$  binding free energy indicates that the mutant (N88D/L90M) has a more favorable contribution and vice-versa for the wild-type. Monomer A (1-99), Monomer B (100-198). The residues in the plot are highlighted on HIVPR Asp25/Asp25' (blue), Gly49 (purple), Ile50 (red), Gly27/Gly27' (brown), Ala28/Ala28' (pink)

In order to gain further insight the binding affinity was decomposed into individual residue contributions and the difference between the wild-type and mutant is plotted in Figure 5-7. If the number is negative than the residue of the wild-type contributes more to the total binding affinity and vice-versa if it is positive. The difference in the contribution of Asp25 between the two is attributed to the Asp25 of the wild-type making more favorable electrostatic interactions (hydrogen-bonding) with saquinavir. The desolvation penalty of Asp25' in the mutant is much lower which causes its energetic contribution to be higher than the mutant. The Gly27 and Ala28 residues of the wild-type make for favorable van der Waals and electrostatic interactions with saquinavir, resulting in a larger contribution to the total binding affinity. A slight difference in the orientation of the quinoline ring causes the energetic contributions of the Gly49 and Ile50 to be different between the mutant and the wild-type.

### **5.3.2 Tipranavir-bound MD Simulations**

#### **5.3.2.1 Flexibility**

The RMSF of the carbon  $\alpha$  atoms and the difference between them are plotted in Figure 5-8. The mutant shows a greater flexibility than the wild-type. The most noticeable differences are Asp35, flap elbow residue, Gln61 and the 70's strand, both located in the hydrophobic core, of monomer A (residues 1-99). Asp88',  $\alpha$ -helical residue, of monomer B (residue 100-198) fluctuates more in the mutant. The difference between the flap residues (43-58 of each monomer) of the wild-type and mutant is insignificant,  $<.1 \text{ \AA}$ .

The RMSD of tipranavir versus the minimized crystal structure is plotted in Figure 5-9. The maximum peaks of both the wild-type and mutant are centered at  $1 \text{ \AA}$  but the

RMSD distribution of the mutant is slightly larger—the mutant has a tail between 1.5 and 2 Å that is not present in wild-type.

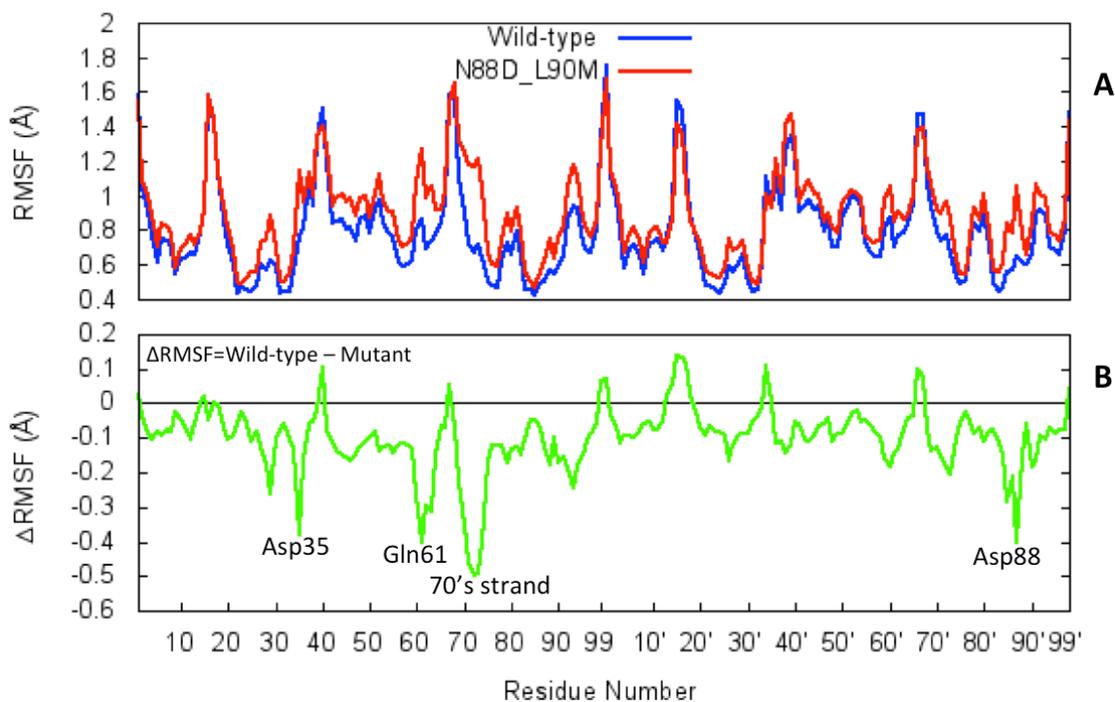


Figure 5-8. Atomic Fluctuations of the wild-type and the mutant (N88D/L90M), subtype C, bound to tipranavir. A) Plot of the RMSF. B) Plot of the difference between the wild-type and mutant (N88D/L90M). Positive value signifies the wild-type fluctuates more and a negative value indicates the mutant fluctuates more. Monomer A (1-99) Monomer B (1'-99').

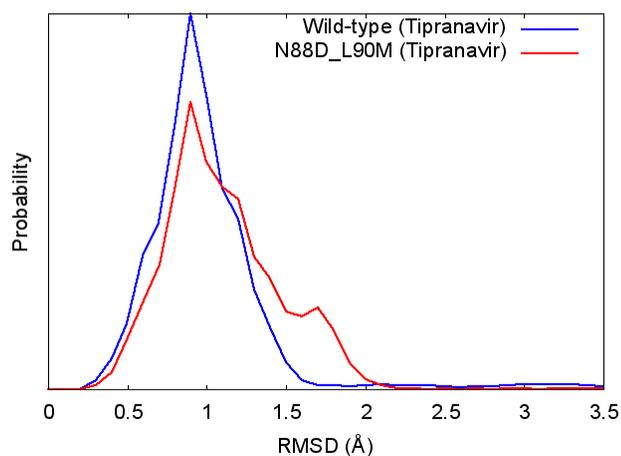


Figure 5-9. Histogram of the RMSD of tipranavir. MD simulations of the wild-type and the mutant (N88D/L90M), subtype C, bound to tipranavir.

The RMSD of the flap residues versus the minimized crystal structure is shown in Figure 5-10. Both systems exhibit maximum peaks at approximately .7 Å. There is no obvious difference between the two although the flaps favor the closed structure slightly more for the wild-type because its maximum peak is larger.

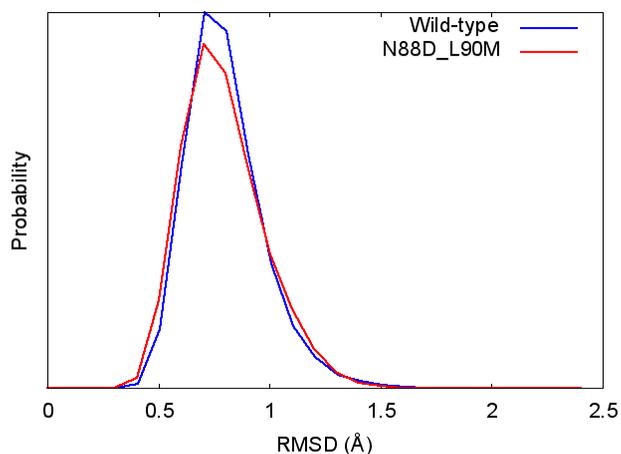


Figure 5-10. Histogram of the RMSD of the flap residue (43-58) of each monomer. MD simulations of the wild-type and the mutant (N88D/L90M), subtype C, bound to tipranavir.

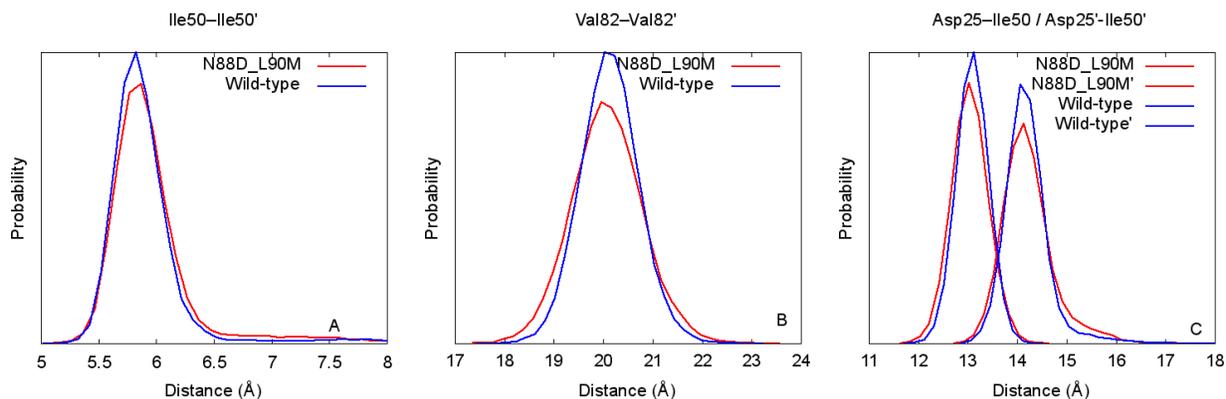


Figure 5-11. Histograms of the different distances in the active site of the wild-type and the mutant (N88D/L90M), subtype C, bound to tipranavir. A) Ile50-Ile50' distance B) Val82-Val82' distance C) Asp25-Ile50/Asp25'-Ile50 distance

The distance between Ile50 of each monomer was determined and plotted in Figure 5-11 A, in order to assess the flap tip separation. A maximum peak at roughly 6 Å is

exhibited by both systems. The measured distance between Val82 of each monomer is plotted in Figure 5-11 B. There is no major difference between the two and both wild-type and mutant display peaks at 20 Å. The distance between residue Asp25 and Ile50, was computed and plotted in Figure 5-11 C. The wild-type and mutant have maximum peaks at 13 and 14.5 Å.

### 5.3.2.2 Hydrogen bonding analysis

To gauge the impact the mutations had with tipranavir and residues in the active site, a hydrogen bond analysis was performed, Table 5-3. The mutations does not affect how the deprotonated Asp25' interacts with the central hydroxyl group of tipranavir. However, the protonated Asp25 of both the wild-type and mutant do not interact strongly with the hydroxyl group of tipranavir. The hydrogen bond interaction between backbone carbonyl oxygen Gly48 and tipranavir is slightly stronger in the wild-type. The wild-type and the mutant flap residues Gly48, Ile49, and Ile49' all form backbone hydrogen bonds with tipranavir via the amide group.

Table 5-3. Hydrogen bonding analysis of the wild-type and the mutant (N88D/L90M), subtype C, bound to tipranavir.

Acceptor	Donor	%Occupancy	
		Wild-type	N88D/L90M
Asp25'—O	TPV O—H	99	99
Gly48—O	TPV N—H	67	52
TPV—O	Ile50' N—H	43	43
TPV—O	Gly48 N—H	31	33
TPV—O	Ile50 N—H	26	31

### 5.3.2.3 MMGBSA calculations

Table 5-4, lists the results of the MMPBSA calculations performed on the wild-type and mutant. The difference in binding affinity between the wild-type and mutant is approximately 4 kcal/mol. The differences in the free energy of solvation and the entropy of the wild-type and mutant are ~1 kcal/mol. The wild-type makes more van der Waals interactions with tipranavir and explains why the  $\Delta H_{\text{Gas}}$  term is higher—the difference between the electrostatic interactions of the wild-type and mutant is minute.

Table 5-4. Results of the MMGBSA calculations for the wild-type and the mutant (N88D/L90M), subtype C, bound to tipranavir. All units in kcal/mol

System	$\Delta H_{\text{gas}}$	$\Delta G_{\text{solvation}}$	$T\Delta S_{\text{quasi-harm}}$	$\Delta G_{\text{bind}}$
Wild-type	$-120.60 \pm 0.03$	$53.87 \pm 0.03$	-22.27	$-44.44 \pm 0.02$
N88D_L90M	$-118.13 \pm 0.04$	$54.33 \pm 0.03$	-23.29	$-40.5 \pm 0.02$

The binding free energy was decomposed into individual contributions and the differences between the wild-type and mutant are plotted in Figure 5-12. The difference between the wild-type's and mutant's energetic contribution of catalytic Asp25 (protonated) is negligible. On the other hand, the difference of the energetic contribution of Asp25' (deprotonated) is not so subtle and is attributed to the wild-type having a lower desolvation penalty. The hydrogen bonding analysis, Table 5-3, demonstrated that the central hydroxyl group of tipranavir interacted with the Asp25' of the mutant and the wild-type equally and further investigation is need to explain why the desolvation penalty is lower in the wild-type. Ala28 of wild-type contributes more to the total binding free energy as a result of forming more favorable van der Waals and electrostatic interactions. On the other hand, Asp29 of the mutant contributes more by the same token. The only noticeable difference between the wild-type and mutant flap residues

interaction with tipranavir is Gly48. The difference between them is a direct result of the electrostatic interactions between tipranavir and Gly48 being more favorable in the wild-type. The most interesting difference between the wild-type and mutant is Arg87. Arg87 is not located in the active site but does occasionally form hydrogen bonds with the trifluoromethyl group of tipranavir (<3%). The lower desolvation penalty of Arg87 accounts for the majority of the difference between the wild-type and mutant.

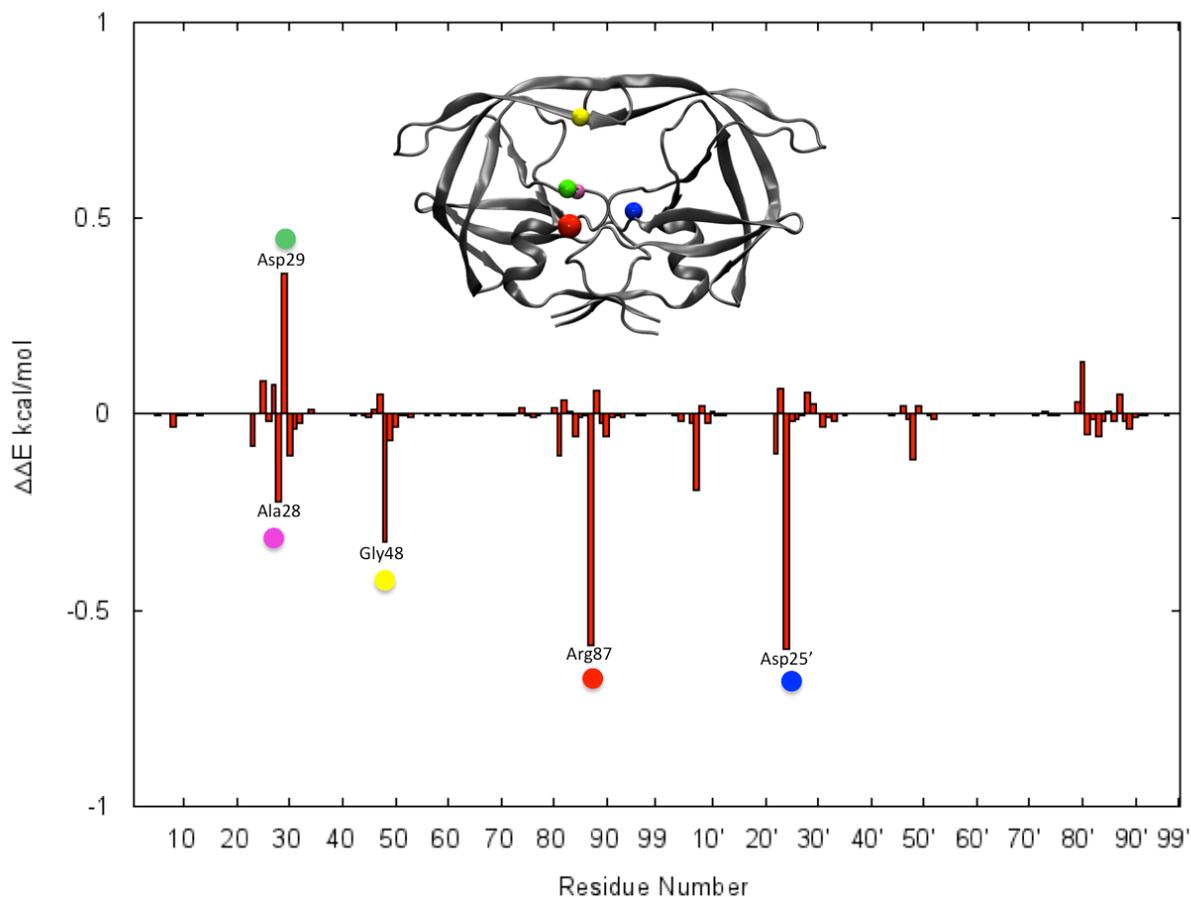


Figure 5-12. Per-residue energy decomposition of the binding affinity of the wild-type and mutant (N88D/L90M), subtype C bound to tipranavir. The difference between the wild-type and mutant is shown, where a positive  $\Delta\Delta$  binding free energy indicates that the mutant has a more favorable contribution and vice-versa for the wild-type. Monomer A (1-99), Monomer B (100-198). The residues in the plot are highlighted on HIVPR Asp25' (blue), Gly48 (yellow), Asp29 (green), Ala28 (pink), Arg87 (red).

### 5.3.3 Saquinavir vs. Tipranavir

Tipranavir is not as susceptible to the L90M/N88D mutations as saquinavir, which is consistent with kinetic studies of Coman *et al.*<sup>177</sup> There is approximately a 4 kcal/mol difference in the binding affinity between wild-type and mutant when either tipranavir or saquinavir is bound. The reason why binding of tipranavir is decreased in the mutant could be attributed to the fact the subtype C possesses the polymorphisms 36I, 69K and 89M that are secondary mutations, not located in the active site, associated with tipranavir<sup>87</sup>.

However the binding affinity of mutant-tipranavir is still higher than the binding affinity of wild-type-saquinavir, Tables 5-2 and 5-4. One explanation as to why tipranavir is more effective than saquinavir is because it forms more direct hydrogen bonds with residues in the active site especially in the flaps. The flap hydrogen bonds help to maintain the rigidity of tipranavir allowing it still interact strongly with the catalytic aspartates even though the longer side chain of Met90 disrupts the dimer interface interactions.

### 5.3.4 Drug Resistant Mechanism

The Asn88 residue in the wild-type forms direct hydrogen bonds with Thr74, Thr31, and Asp29 through the following interactions: the side-chain carbonyl of N88 forms a hydrogen bond with the side chain hydroxyl group of Thr31, the backbone carbonyl of Thr74 forms a hydrogen bond with the side chain amine of Asn88, the backbone amide of Asn88 forms a hydrogen bond to the backbone carbonyl of Asp29, Figure 5-13. Furthermore Asn88 hydrogen bonds with Thr74 via a water molecule. The interactions the Asn88 residue makes within the protease are consistent with previous studies.<sup>178,184</sup> Asp88 does not form any significant interactions Thr74 or Thr31 via—

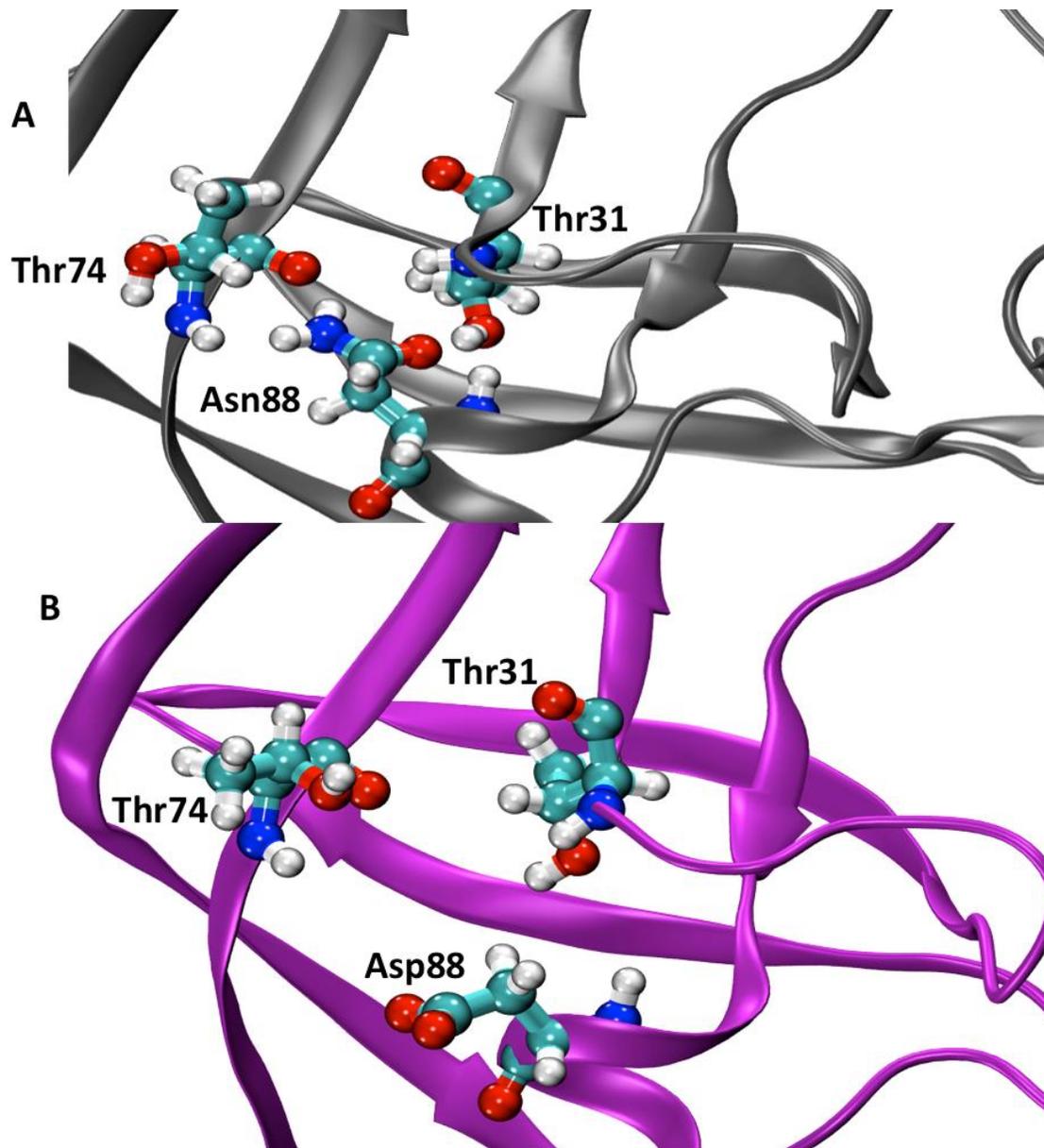


Figure 5-13. Schematic diagram of the Asp88/Asn88 interactions with Thr31 and Thr74  
 A) wild-type B) mutant (N88D/L90M). Asp88 (purple) is unable to hydrogen bond to Thr31 and Thr74 residues but Asn88 (grey) residue can. direct or water mediated hydrogen bonds with Thr74 or Thr31. Asp88 does form a hydrogen bond with Asp29 via backbone, amide and carbonyl group. Asp88 inability to form a hydrogen bond with Thr74 might affect the dynamics of the Pro79-Val82 loop in

the active site and explains why the distance between Val82 and Val82' is shorter and fluctuates more when compared to the wild-type, (saquinavir bound MD simulations).

A pairwise energy decomposition calculation, results not shown, reveals that the longer side-chain of the Met90 residue does form van der Waals interactions with residues Leu24, Asp25 and Thr26 than Leu90. This increased interaction causes a shift in the conformation of Asp25 and Asp25' that affects how the aspartates interact with the central hydroxyl group of saquinavir. Met90 mutation is responsible for the decreased interaction of Ala28 and Gly27 with both saquinavir and tipranavir due to the weakened interactions at the dimer interface.<sup>173,175</sup>

## 5.5 Conclusions

Our MD studies have shown how the non-active site mutations reduce the efficacy of the protease inhibitor saquinavir. The Asp88 mutation affects the dynamics of the Pro79-Val82 loop in the active site by making it more flexible because of the lost hydrogen bond between Thr74 and Asp88. This effect is more pronounced when saquinavir is bound. The larger Met90 mutations impacts how the catalytic aspartates interact with the central hydroxyl group of saquinavir—the interactions between the central hydroxyl group and the catalytic aspartates are unaffected by the L90M mutation. Tipranavir is able to overcome these mutations by forming direct hydrogen bonds with the protein especially in the flap region.

The binding affinities of saquinavir and tipranavir when bound to the mutant are reduced when compared to the wild-type. The reduced binding affinity of saquinavir is primarily due to the L90M mutation. The naturally occurring polymorphisms of subtype C 36I, 69K and 89M might be the reason the binding affinity of tipranavir is lower. The binding affinity of the mutant-tipranavir is still higher than the wild-type-saquinavir

suggesting that tipranavir might still be an effective inhibitor in spite of the lower binding affinity when compared to the wild-type. In addition, darunavir and tipranavir are the only inhibitors not susceptible to the L90M mutations.

Although these mutations do not form direct contact with the inhibitor they confer resistance by altering the geometry of the active site, which in turn affects how the active site residues interact with ligand. These results offer insight as to how new inhibitors can be designed to overcome drug resistant strains of HIV.

## CHAPTER 6 INVESTIGATING WHAT EFFECTS THE STABILIZING MUTATIONS HAVE ON THE EPR MEASURED DISTANCE PROFILES

### 6.1 Introduction

A variety of biomolecular conformational changes and structures have been studied using double electron-electron resonance (DEER).<sup>185</sup> Distance measurements by site-directed spin labeling (SDSL) EPR are based on the degree of the magnetic dipolar coupling of the unpaired nitroxide electrons that scales as  $1/r^3$ , where  $r$  denotes the distance between unpaired spins.<sup>186</sup>

HIV-1 protease has two flaps that govern access to the active site. The dynamics of the flaps have been studied using a variety of different methods.<sup>43,46,47,50</sup> In particular, Galiano *et al.*<sup>186</sup>, Kear *et al.*<sup>137</sup>, and Blackburn *et al.*<sup>172</sup> showed that DEER EPR spectroscopy could discern between the different flap conformations of HIVPR: closed, semi-open, wide-open and tucked/curled. In these DEER experiments the distance between K55SL-(spin label) of each monomer is measured, Figure 6-1. The K55 residue is mutated to a Cys and then the label (1-Oxyl-2,2,5,5-Tetramethyl- $\Delta$ 3-Pyrroline-3-Methyl) Methanethiosulfonate (MTSL) is attached. Lys55 was chosen because it was one of the few positions in the flap region that could tolerate an amino acid substitution while still retaining activity comparable to the wild-type.<sup>186</sup> Furthermore, the distances of the flaps span from 26 – 48 Å<sup>186</sup>, which is well within the sensitivity of the methodology.<sup>187,188</sup>

In order to perform the DEER experiments several mutations were made besides the K55SL: the mutations Q7K, L63I, L33I, and D25N prevent auto-proteolysis<sup>164</sup>, C67A and C95A are to ensure the spin label attaches exclusively to C55. This sequence is referred to as LAI. MD studies have shown that the EPR measured distances can be

reproduced.<sup>172,189</sup> The question remains, do the mutations (Q7K, L33I, D25N, L63I, C67A, C95A) affect the distance profiles and trends observed in the previous EPR studies<sup>137,171,172,186</sup> performed on HIVPR?

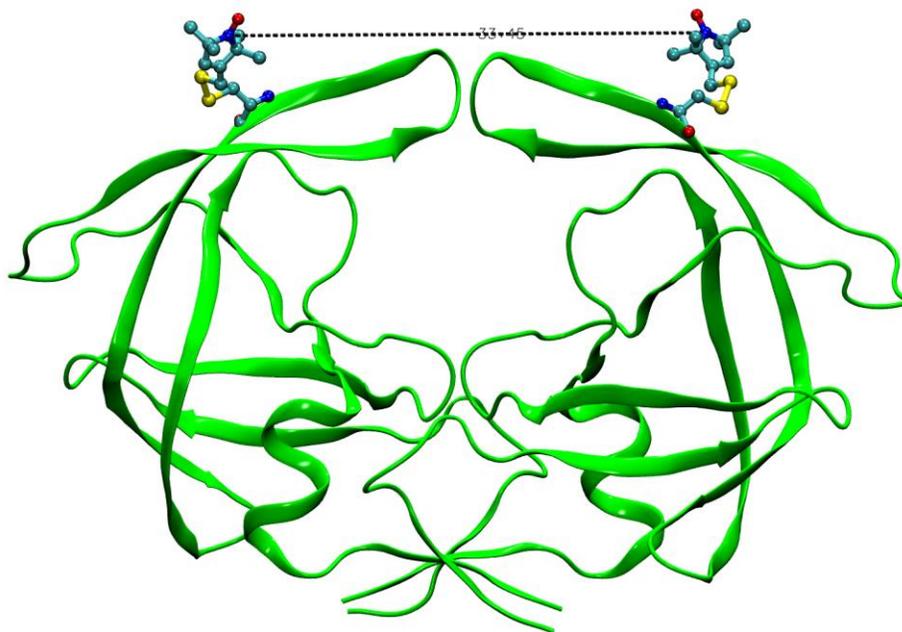


Figure 6-1. Schematic of HIV-1 protease. Highlighted are the spin labels at position 55 of each monomer. The dashed line is represents the distance that is measured.

A total of twelve MD simulations containing the spin label MTSL were performed. Nine MD simulations consisted of HIVPR, subtype B, bound to all nine FDA protease inhibitors. The others three simulations included the apo form of the wild-type (subtype B), MDR769<sup>190</sup> and V6<sup>191</sup>, which are drug-resistant constructs. The sequence of MDR769 when compared to the wild-type differs at positions L10I, M36V, S37N, M46L, I54V, I62V, L63P, A71V, V82A, I84V, L90M and the sequence of V6 compared to the wild-type differs at K20R, V32I, L33F, M36I, L63P, A71V, V82A, L90M. The sequences of wild-type (subtype B), LAI', MDR769 and V6 are given in Appendix A.

## 6.2 Methods

### 6.2.1 General Setup

Twelve different forms of HIVPR were simulated and the initial coordinates were X-ray crystal structures from the PDB, PDB Codes: 1HHP – wild-type (apo)<sup>128</sup>, 1HPV – Amprenavir<sup>192</sup>, 2O4K – Atazanavir<sup>130</sup>, 3TKW – Darunavir<sup>131</sup>, 2BPX – Indinavir<sup>132</sup>, 2O4S – Lopinavir<sup>130</sup>, 1OHR – Nelfinavir<sup>133</sup>, 1HXW – Ritonavir<sup>134</sup>, 2O4P – Tipranavir<sup>130</sup>. If any of the sequences of the crystal structures possess stabilizing mutations, they were removed prior to running any MD simulations.

The initial coordinates for the drug resistant constructs MDR769 and V6 were from the PDB 1HHP. Due to the large amount of mutations made to 1HHP to match the sequences of MDR769 and V6, a special equilibration procedure was employed prior to solvating the structures and performing any MD simulations. The structures were minimized using 500 steps of steepest descent followed by 15000 steps of conjugate gradient, where only the side chains were allowed to move. Next, the structures were heated linearly from 100 to 300 K for 500 ps and a 50 kcal/mol-Å<sup>2</sup> positional restraint was used for backbone atoms. The solvent was modeled implicitly using the Generalized Born Method (igb=8)<sup>167</sup> in AMBER in conjunction with *mbondi3* intrinsic radii.

The missing hydrogen and heavy atoms were added using the LEaP module in AmberTools13. The AMBER ff99SB<sup>99</sup> force field was used for the protein and the parameters for protease inhibitors were generated using the *antechamber* module and GAFF<sup>139</sup> force field with AM1-BCC<sup>140</sup> charges. Only one of the catalytic aspartates was protonated.

The explicit solvent was modeled using a truncated octahedron box of TIP3P<sup>103</sup> waters extended 15 Å from the edge of the solute—chloride ions were added to force neutrality. Long range electrostatics were treated with PME method<sup>166</sup> using a direct space and a van der Waals interactions cutoff of 9 Å. The temperature was maintained at 300 K using Langevin Dynamics with a collision frequency of 1.0 ps<sup>-1</sup>. The SHAKE<sup>143</sup> algorithm was employed so that a 2 fs time step could be utilized.

### 6.2.2 MTSL Parameters

The MTSL label was built using the visualization software *Avogadro*.<sup>193</sup> Gaussian 09<sup>194</sup> was used to perform the geometry optimization of the MTSL in conjunction with functional M062X and 6-31+G\* basis set. The antechamber module was employed to calculate the Restrained Electrostatic Potential (RESP) charges for the label.

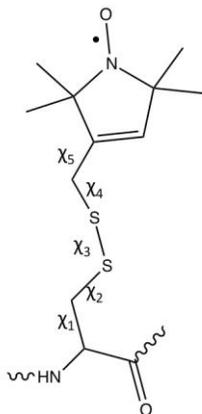


Figure 6-2. Schematic of the five torsional angles of the spin label MTSL

The last step in the building of MTSL was parameterizing the torsions  $\chi_1$ ,  $\chi_2$ ,  $\chi_3$ ,  $\chi_4$ ,  $\chi_5$ , Figure 6-2. The  $\chi_4/\chi_5$  model<sup>195</sup> suggests that the last two dihedrals,  $\chi_4$ , and  $\chi_5$  are largely responsible for the internal motion of MTSL and that the other dihedrals,  $\chi_1$ ,  $\chi_2$ , and  $\chi_3$  are essentially rigid. Because of this model, only the last two torsions have been previously deemed important enough to parameterize.<sup>196</sup> Furthermore, torsional fittings

are generally performed by fitting each dihedral independently of each other because the assumption is that the coupling between torsions is negligible. We have determined that not only are these assumptions incorrect but that all five torsions of the spin label are coupled and need to be fitted simultaneously. The program *Torsion\_Fit* written by Chad Hopkins (in review), was used to generate the parameters for all five torsions.

### **6.2.3 Molecular Dynamics**

The hydrogen atoms were minimized first, followed by minimization of the entire system. Each minimization process consisted of 500 steps of steepest descent followed by 10000 steps of conjugate gradient. The systems were heated linearly from 100 K to 310 K for 1 ns where a 5 kcal/mol·Å<sup>2</sup> positional restraint was used on the solute. Following heating, the density was equilibrated for 500 ps. Next, five simulations were performed where the positional restraint was reduced from 5 kcal to .1 kcal/mol·Å<sup>2</sup>. In the last step of equilibration all restraints were removed and 2 ns of unrestrained MD simulation was performed.

The production phase for the ligand-bound simulations consisted of running 4 x 50.5 ns simulations but only the last 50 ns was used for analysis. The apo simulations consisted of running 10 x 30.5 ns simulations. Snapshots of the MD trajectory were saved every 30 ps.

## **6.3 Results and Discussions**

### **6.3.1 Inhibitor-bound MD simulations**

A histogram of the RMSD distribution of the backbone atoms of the protein and the heavy atoms of the inhibitor for each of the inhibitor-bound MD simulations was plotted in Figure 6-3 to demonstrate that the simulations were stable. The reference structure for RMSD calculations was the minimized crystal structure of each.

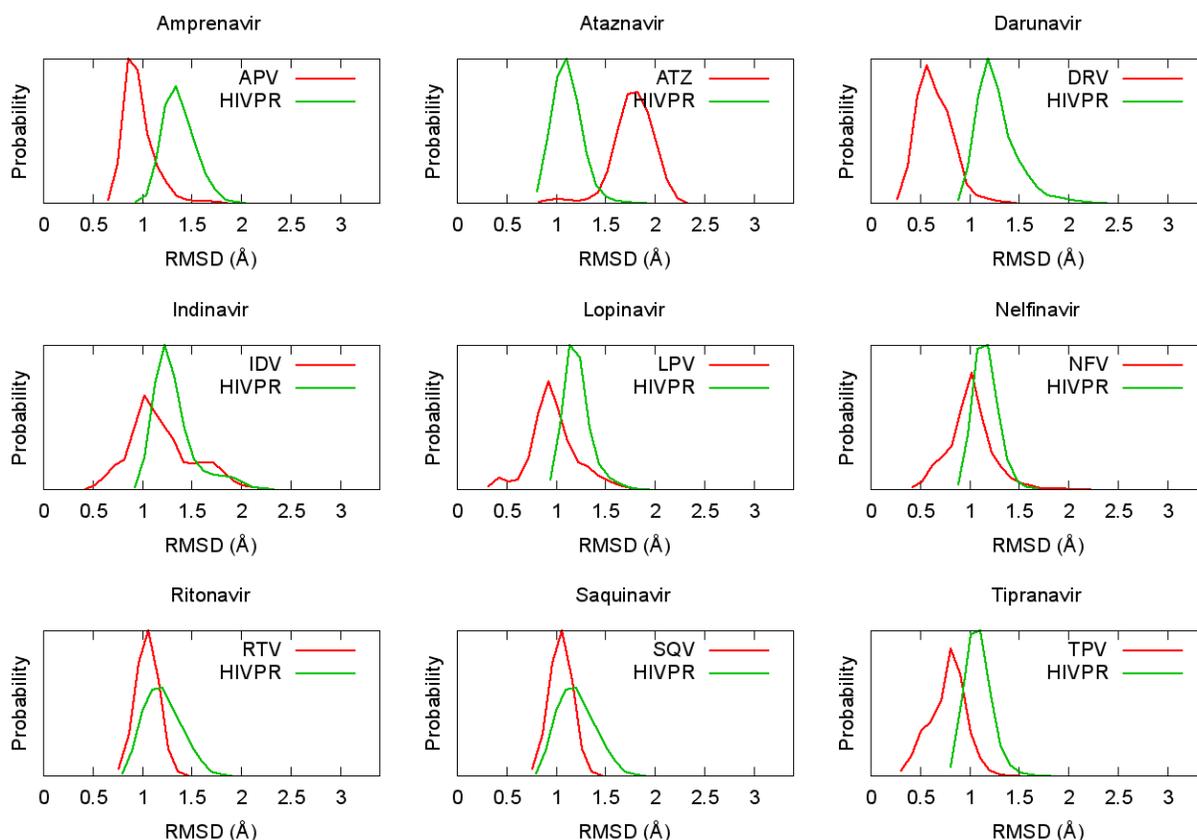


Figure 6-3. Histograms of the RMSD of HIV-1 protease subtype B and the protease inhibitor bound. HIVPR refers to only the protein residues 1-198, and the protease inhibitors.

The measured 55N-55N distance profiles of each the apo and inhibitor-bound forms of HIVPR are plotted in Figure 6-4. The interflap distances of the inhibitor bound form of HIVPR are shorter when compared to the apo, and this trend is consistent with previous studies.<sup>171,186,189</sup> The most probable, average, and range span of the interflap distances of all the inhibitors are approximately the same, Table 6-1. Unlike previous studies<sup>171,186,189,197</sup>, the apo distance distribution profile is narrower than the ligand-bound distance distribution in certain instances; possibly as a result of kinetic trapping. The distribution profile of the apo is expected to be broader because the flaps sample

more conformations in the apo form of the enzyme. The apo distance distribution will be discussed in later sections.

Blackburn *et al.*<sup>171</sup> had two main conclusions from their experimental DEER study on HIVPR: that of the nine FDA approved inhibitors, nelfinavir, indinavir, and atazanavir were the least effective at closing the flaps, and that the percentage of time the flaps are closed correlates to the number of direct hydrogen bonds made between the active site residues and the inhibitor observed in the crystal structure (excluding hydrogen bonds between D25).

However our results do not reflect the trend in Blackburn *et al.*<sup>171</sup> and we find that all of the inhibitors are effective in closing the flaps. The most probable distance for the inhibitors are at ~32 Å, Table 6-1, which is more or less what the results of Blackburn *et al.*<sup>171</sup> show (excluding atazanavir, indinavir, nelfinavir). The relatively good agreement for the other inhibitors is no surprise because a similar EPR study performed by Torbeev *et al.*<sup>197</sup> measured the interflap distances of the active (D25) and inactive (N25) HIVPR using the protease inhibitors MVT-101, JG-365, and KV201, and there were only slight differences between two distance profiles.

Table 6-1. Distance profiles of inhibitor-bound simulations

<b>Inhibitor</b>	<b>Average Distance</b>	<b>Most Probable</b>	<b>Range Span</b>
Amprenavir	31.9 ± 0.6 Å	32.7 ± 0.5 Å	21.0 – 36.7 Å
Atazanavir	32.0 ± 0.6 Å	32.7 ± 0.5 Å	21.8 – 35.7 Å
Darunavir	31.7 ± 0.6 Å	32.7 ± 0.5 Å	20.3 – 35.7 Å
Indinavir	31.5 ± 0.6 Å	32.3 ± 0.5 Å	21.5 – 36.1 Å
Lopinavir	31.3 ± 0.6 Å	32.4 ± 0.5 Å	21.0 – 35.7 Å
Nelfinavir	31.7 ± 0.6 Å	32.9 ± 0.5 Å	21.4 – 35.9 Å
Ritonavir	31.8 ± 0.5 Å	32.6 ± 0.4 Å	21.8 – 36.5 Å
Saquinavir	31.8 ± 0.7 Å	32.8 ± 0.6 Å	19.7 – 36.3 Å
Tipranavir	30.8 ± 1.0 Å	32.7 ± 0.8 Å	18.0 – 35.6 Å

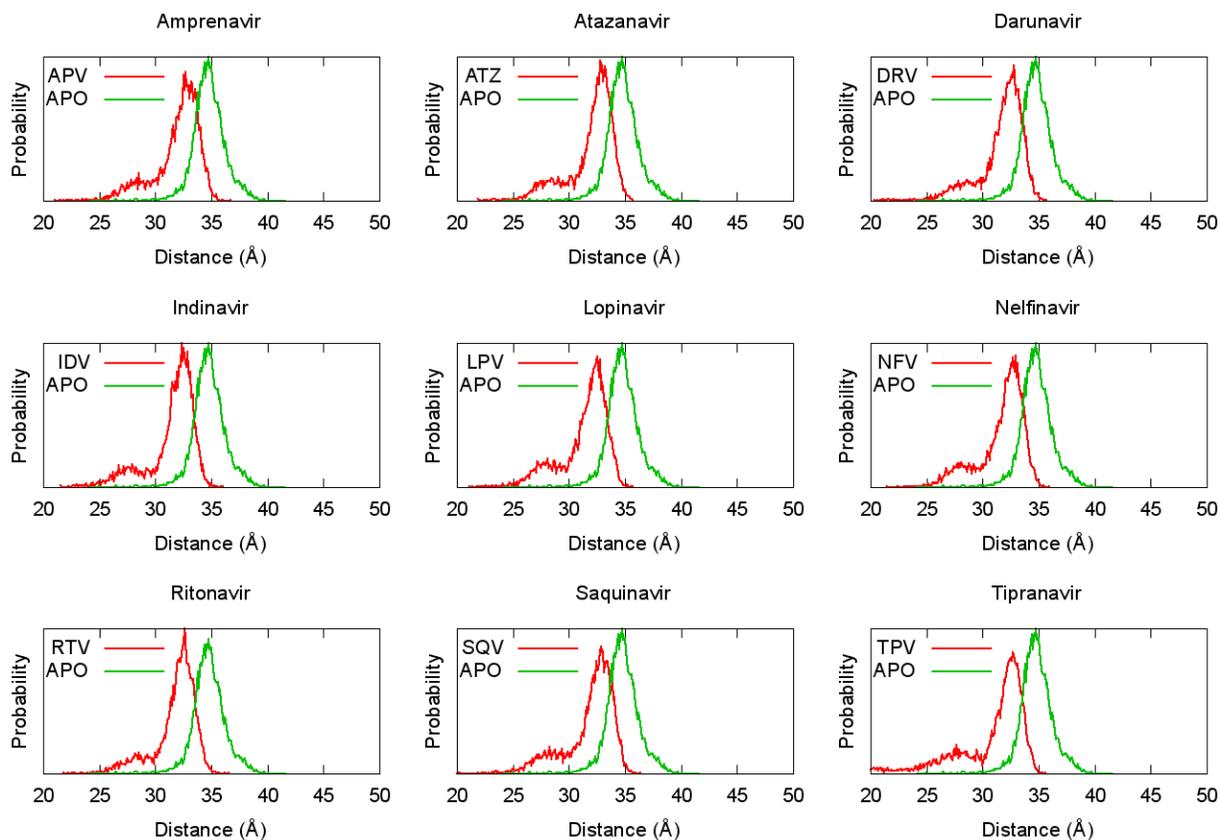


Figure 6-4. Histogram of N—N distance distribution profiles of apo form HIVPR subtype B compared to when it is bound to protease inhibitors.

One explanation as to why the inhibitors nelfinavir, atazanavir and indinavir do not close the flaps in the study performed by Blackburn *et al.*<sup>171</sup> could be that the inhibitors are more susceptible to the D25N mutation than the other inhibitors because it is the only stabilizing mutation located in the active site. However previous chapters have demonstrated that mutations outside of the active site can alter the interactions of the inhibitor with residues in the active site thus changing the dynamics of the inhibitor.

### 6.3.2 Apo MD Simulations

In order to investigate the dynamics of the backbone atoms a RSMF calculation was performed, Figure 6-5. The flaps (residues 43-58) of the drug resistant constructs MDR769, and V6 exhibit a greater flexibility than the wild-type. The differences between

MDR769 and the wild-type residues 1-30 and 60-99 is negligible. However, this is not the case for V6, which fluctuates more than the wild-type in these regions.

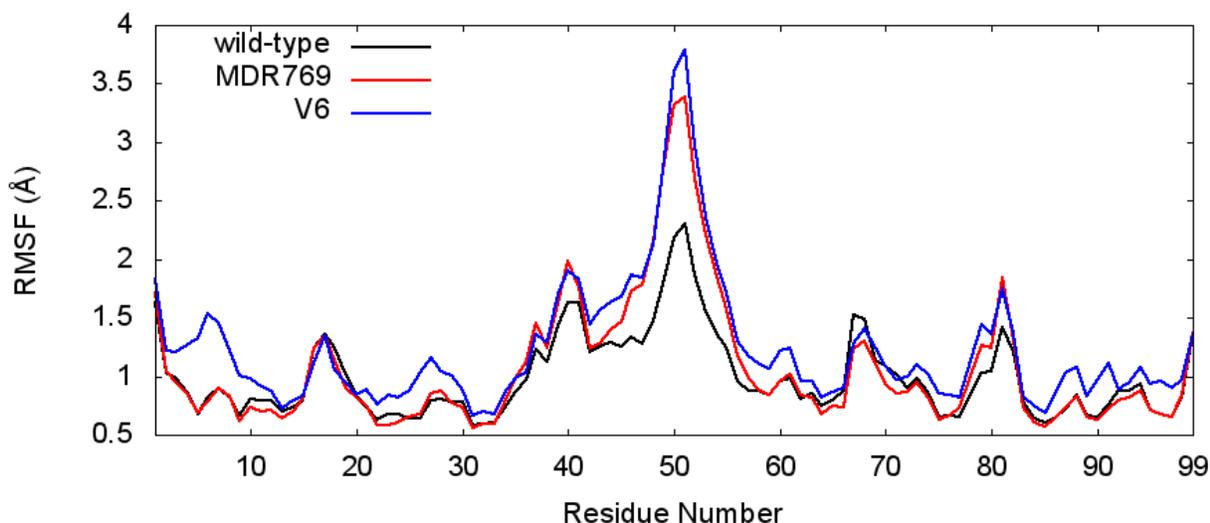


Figure 6-5. Plot of the RMSF of the wild-type, subtype B (black), MDR769 (red), V6 (blue). The RMSF of the two monomers have been averaged together.

Figure 6-6, illustrates the interflap distance profiles for the three different constructs. The most probable distance of the wild-type and V6 are approximately the same, 34.7 Å, Table 6-2. The distance distribution profile breadth of V6 is broader than the wild-type. The most probable distance of MDR769 is 35.7 Å, which is larger than the wild-type. Similar to V6, the distance profile breadth of MDR769 is also larger than the wild-type. The order of the distance profile breadths for the three constructs: V6 > MDR769 > wild-type, agrees well with the trend in Figure 6-5, where V6 was the most flexible and the wild-type was the least flexible. However, there is no direct correlation between the most probable distance and the range span of the distance profiles.

The average structures of the three constructs from MD simulations are depicted in Figure 6-7. The flaps of all three are in a semi-open conformation. The Ile50-Ile50' is slightly larger in MDR769 when compared to the other two.

Table 6-2. Distance profiles of the apo MD simulations

Construct	Average Distance	Most Probable	Range Span
Wild-type	$34.9 \pm 0.5 \text{ \AA}$	$34.7 \pm 0.4 \text{ \AA}$	24.1 – 41.5 $\text{\AA}$
V6	$35.7 \pm 0.7 \text{ \AA}$	$34.7 \pm 0.6 \text{ \AA}$	23.8 – 48.8 $\text{\AA}$
MDR769	$35.7 \pm 0.6 \text{ \AA}$	$35.3 \pm 0.5 \text{ \AA}$	28.0 – 50.0 $\text{\AA}$

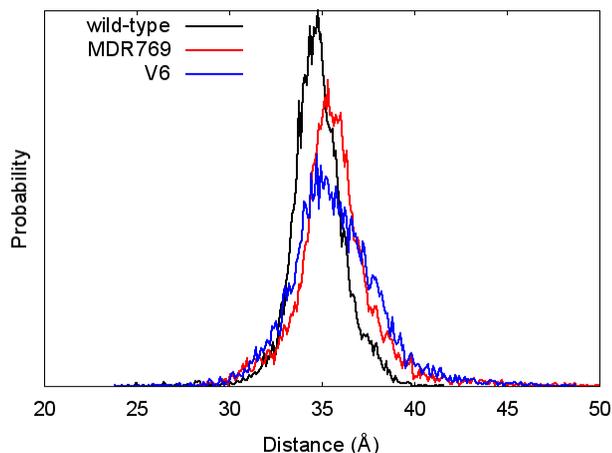


Figure 6-6. Histogram of the N—N distance profiles. Wild-type subtype B, (black), MDR769 (red), V6 (blue)

The results obtained from our MD simulations and those experimentally and computationally determined by Galiano *et al.*, exhibit different trends for the three constructs. The LAI, which in this case is equivalent to the wild-type, has the broadest distance distribution profile, followed by V6 and MDR769 had the narrowest distribution profile. Galiano *et al.*<sup>172</sup> determined that the most probable distance for V6 is  $34.9 \pm 0.7 \text{ \AA}$ , LAI is  $35.5 \pm 1 \text{ \AA}$ , and MDR769  $36.4 \pm 0.5 \text{ \AA}$ .

Both studies find that MDR769 is more open than the wild-type/LAI, which agrees with the apo crystal structure of MDR769 PDB ID 1TW7<sup>190</sup> when compared to apo structure of the wild-type PDB ID 1HHP<sup>128</sup> There is no apo crystal structure for the construct V6 for comparison, only inhibitor bound forms. The most probable distance calculated for V6 by both studies are found to be in a good agreement,  $\sim 34 \text{ \AA}$ .

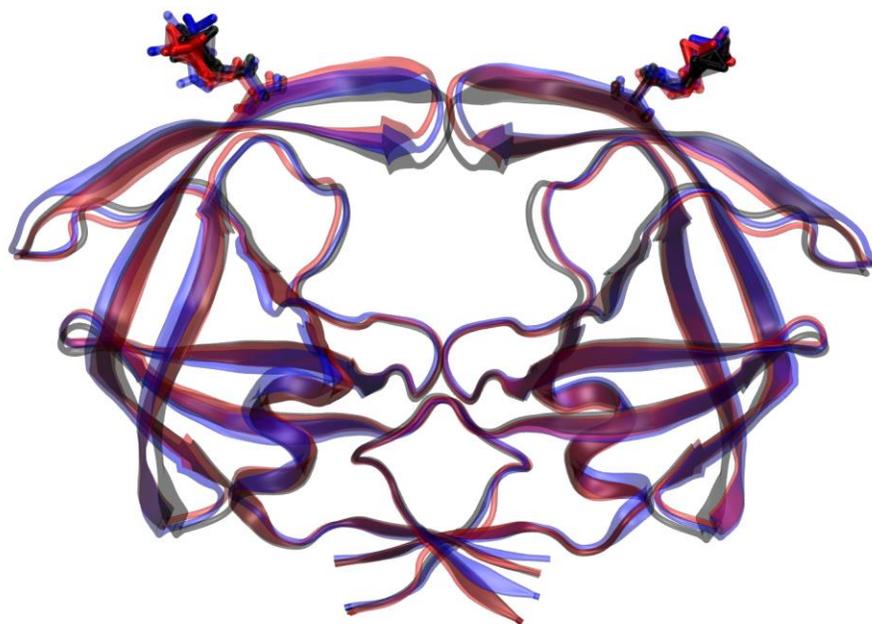


Figure 6-7. Schematic of the average structures of HIVPR: wild-type subtype B (black), MDR769 (red), and V6 (blue) from spin-labeled MD simulations.

### 6.3.3 Comparing the N-N and C $\alpha$ -C $\alpha$ distances

A plot of the distance between residues 55-55 C $\alpha$  is given in Figure 6-8. The plot was to determine how accurately the spin label, could reproduce the trends seen in the backbone dynamics of the protease since that is essentially its aim. The flexibility of the spin label could be altered by its interactions with the neighboring side chains of other residues. The trend in Figure 6-8 is in good agreement with that depicted in Figure 6-6. The wild-type has the narrowest distance profile width and the V6 has the widest distance profile. However, there is a faint peak at 27 Å exhibited by V6 in Figure 6-8, indicating another conformation of the flaps that is not present in Figure 6-6. Nonetheless, this confirms that the EPR measured profiles are dominated by the dynamics of the backbone and not the “floppiness” of the spin label, which agrees with computational studies done by Galiano *et al.*<sup>172</sup>

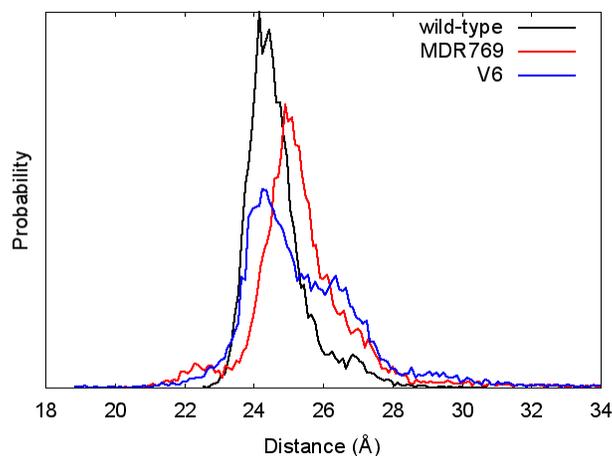


Figure 6-8. Histogram of the 55-55 C $\alpha$  distance profile, wild-type subtype B (black), MDR769 (red), V6 (blue)

Our results indicate that the drug resistance of V6 and MDR769 is due to the flaps showing a greater flexibility than the wild-type, therefore allowing an escape mechanism for an inhibitor. In addition, a larger degree of flexibility could result in a larger entropic penalty for closing the flaps upon the binding of an inhibitor.

#### 6.4 Conclusions

Our MD simulations show that the effect of the stabilizing mutations (Q7K, I33L, I63L, D25N, C67A, C95A) on the EPR measured interflap distances are minimal for the inhibitor bound forms of HIVPR excluding the inhibitors indinavir, nelfinavir, and atazanavir. The average most probable distances determined for the inhibitor bound forms of the protease are within  $\sim 0.5$  Å of each other. Maintaining D25, the active form of the HIVPR, causes all the inhibitors to close the flaps tightly.

The biggest difference between was between LAI' and the wild-type. Our measured spin-label distance profile is much more narrower than that of the LAI'. The most probable distances differ by more  $\sim 1$  Å of each other. Further investigation is needed to determine if our distance profile is a result of not sampling enough or if it is

caused by the parameters of the spin label etc. However, both studies concluded that MDR769 is more open than the wild-type/LAI' and the dynamics of the flaps could explain its drug resistance.

Lastly, we showed that the N-N distances are a good reflection of the backbone dynamics of the protease and that the difference between them is minimal for the constructs studied in this investigation.

CHAPTER 7  
PRELIMINARY STUDIES: EVALUATING THE EFFECTIVENESS OF  
ANTIRETROVIRAL PROTEASE INHIBITORS ON XMRV

**7.1 Introduction**

The newly discovered human retrovirus xenotrophic murine leukemia virus-Related virus (XMRV), has been detected in patients with prostate cancer and chronic fatigue syndrome.<sup>198–203</sup> Furthermore, it is the first gammaretrovirus to be linked with human diseases. However, subsequent studies have failed to detect the virus in patients with prostate cancer or chronic fatigue syndrome despite earlier reports.<sup>204–213</sup> In addition, neither the origin nor the mode of transmission have been established—this has led to much debate and is an on-going investigation in the scientific community.

Because XMRV is a retrovirus much like HIV, the main targets of inhibition are its main three enzymes: reverse transcriptase, integrase and the protease. Targeting these enzymes has had much success in HIV antiretroviral therapy. Various retroviral inhibitors have been tested on XMRV and raltegravir, an integrase inhibitor designed for HIV, has been shown to be a potent inhibitor of XMRV.<sup>214,215</sup>

The first x-ray crystal structure of XMRV protease (apo) was solved in 2011.<sup>216</sup> Since then, there have been several crystal structures of XMRV complexed with various antiretroviral protease inhibitors.<sup>217,218</sup> The crystal structures revealed that XMRV is a homodimer, composed of 250 residues (125 residues per monomer), Figure 7-1.<sup>216</sup> Two aspartates lie at the bottom of the active site that are responsible for its catalytic ability. The N and C termini when compared to other retropepsins are longer.<sup>216</sup> Unlike other retroviral proteases, the dimer interface of XMRV is not composed of N and C termini that are interdigitated, but is instead composed of hairpins formed by strands close in proximity to the C-termini.<sup>216</sup>

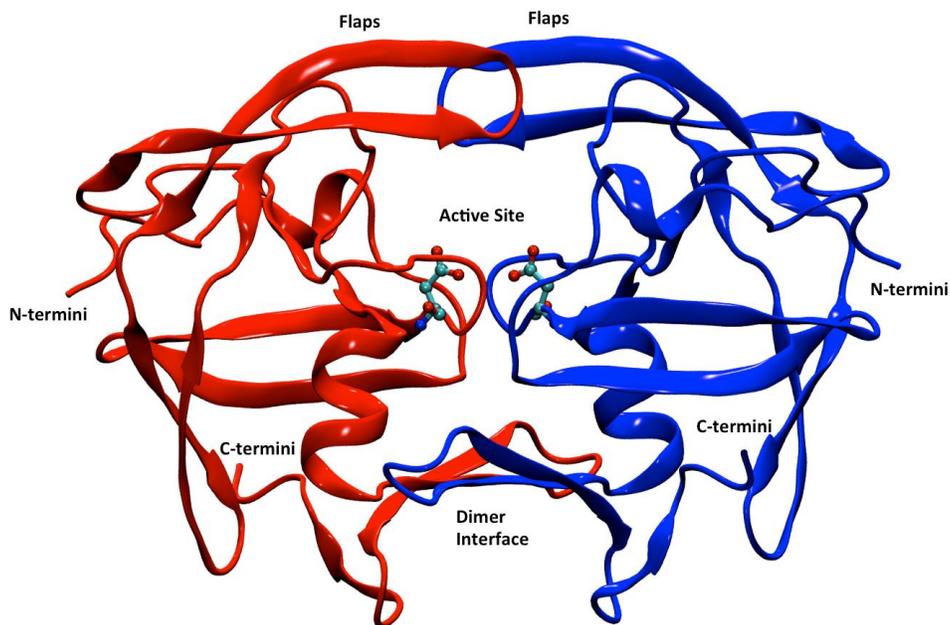


Figure 7-1. Schematic of XMRV protease. Highlighted are the catalytic aspartates

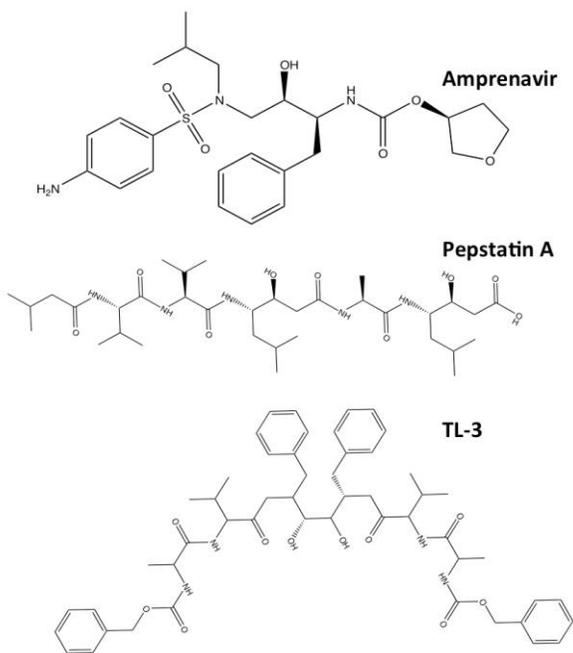


Figure 7-2. Protease inhibitors: Amprenavir (HIV-1 protease), Pepstatin A (generic inhibitor of aspartic proteases), TL-3 (FIV protease)

A variety of protease inhibitors have been tested on XMRV but with limited success.<sup>214,217</sup> In this study we perform MD with XMRV bound to the protease inhibitors:

amprenavir, pepstatin A, and TL-3, Figure 7-2. Amprenavir is a FDA approved protease inhibitor for HIV. TL-3 is a protease inhibitor that was developed to inhibit the FIV protease and it has been demonstrated to be effective against other retroviral proteases.<sup>219</sup> Pepstatin A is a nonspecific inhibitor, capable of inhibiting aspartic proteases.<sup>220</sup> The goal of this study is to analyze the interactions each inhibitor makes in an to effort develop a potent inhibitor of XMRV protease.

## 7.2 Methods

### 7.2.1 Completing XMRV

The coordinates for the wild-type and mutant were derived from the x-ray crystal structures PDB ID: 3SLZ – TL-3, 3SM1 – pepstatin A, 3SM2 – amprenavir. In order to model the missing residues from the N and C-terminus the crystal structure 3NR6 (apo) was utilized. A description of the equilibration procedure for these residues will be discussed in later sections. Pepstatin A was disordered in the crystal structure 3SM1. Pepstatin A was constructed using the visualization software *Avogadro*.<sup>193</sup> Gaussian 09<sup>194</sup> was used to perform the geometry optimization of the pepstatin A in conjunction with the density functional M062X and 6-31+G\* basis set. The pepstatin A inhibitor was then docked into the crystal structure 3SM1 using the GLIDE<sup>221</sup> module of the Schrödinger Suite.

### 7.2.1 General Setup

The missing hydrogen and heavy atoms were added using the Leap module in AmberTools13. The AMBER ff99SB<sup>99</sup> force field was used for the protein. The parameters for amprenavir, TL-3 and pepstatin A were built by using RESP charges generated from the antechamber module and the GAFF<sup>139</sup> force field. The catalytic aspartates were modeled using a mono-protonated state.

The explicit solvent was modeled using a 15 Å TIP3P<sup>103</sup> solvent buffer encapsulating the protein in a truncated octahedron box and chloride ions were added to force neutrality. Long-range electrostatics were accounted for using the PME<sup>166</sup> method utilizing a direct space and a van der Waals cutoff of 9 Å. Langevin Dynamics was utilized to maintain the temperature at 300 K—the collision frequency was 1.0 ps<sup>-1</sup>. In order to achieve a 2 fs time all atoms bonded to hydrogen atoms were constrained using the SHAKE<sup>143</sup> algorithm.

### 7.2.2 Molecular Dynamics

The first step of the equilibration procedure consisted of minimizing the missing N and C-terminus residues and the solvent, followed by a 500 ps MD simulation, while placing a positional restraint weight of 10 kcal/mol·Å<sup>2</sup> on the rest of the solute. The hydrogen atoms were minimized followed by the entire system. Each minimization process consisted of 500 steps of steepest descent followed and 10000 steps of conjugate gradient. The systems were heated linearly from 100 K to 300 K for 1 ns where a 5 kcal/mol·Å<sup>2</sup> positional restraint was used on the solute. Following heating, the density was equilibrated for 500 ps. Next, five simulations were performed where the positional restraint was reduced from 5 kcal to .1 kcal/mol·Å<sup>2</sup>. In last step of equilibration all restraints were removed and 5 ns of unrestrained MD was performed.

The production phase consisted of running a 20.5 ns simulation but only the last 20 ns were used for analysis. Snapshots of the MD trajectory were saved every 15 ps.

### 7.2.3 Molecular Mechanics Poisson Boltzmann Surface Area Calculations

The MMPBSA calculations were performed using *MMPBSA.py*<sup>122</sup> in AmberTools13. The polar solvation energy was calculated using the Poisson Boltzmann Method in conjunction with the *mbondi3* intrinsic radii. The non-polar solvation energy

was calculated using the LCPO<sup>105</sup> method. The normal mode approximation was used to account for the entropic contributions. Due to the high computational cost of normal mode calculations, only 65 frames spaced every 975 ps were used.

## 7.3 Results and Discussion

### 7.3.1 Flexibility

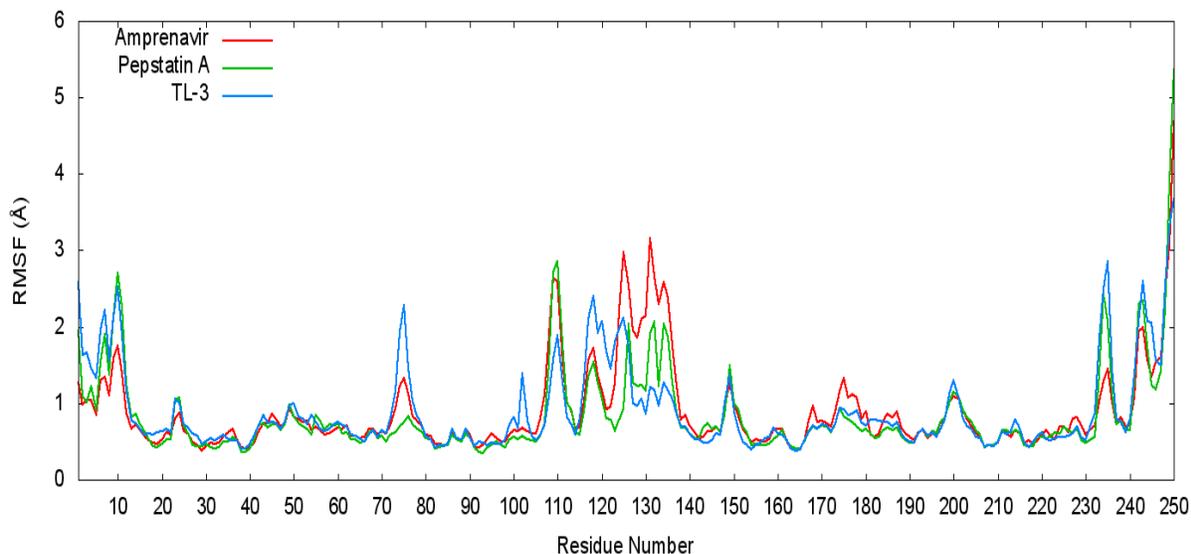


Figure 7-3 Plot of the RMSF of XMRV bound to amprenavir, pepstatin A, TL-3. Monomer A residues (1-125) Monomer B residues (126-250)

The RMSF of XMRV bound to the three different ligands is plotted in Figure 7-3. All three systems exhibit high fluctuations at the N and C termini, residues 1-10, 115-125 of each monomer. The flaps, residues 48-66 of each monomer remain fairly rigid throughout the MD simulation. The fluctuation of residue Thr75 is more pronounced in the TL-3 simulation. The increase in flexibility of Thr75 is attributed to the hydrogen bond between its side-chain hydroxyl group and the backbone carbonyl of Lys100 being broken due to a small conformational change of residues 73-76 of monomer A in TL-3. A histogram of the RMSD of the inhibitors: amprenavir, pepstatin A and TL-3 is plotted in Figure 7-4. The minimized crystal structure was used as the reference. In the case of

pepstatin A the minimized docked structure was the reference. Amprenavir is the most rigid of the three inhibitors and has two peaks at 0.8 and 1.2 Å. The two peaks are caused by a change in the orientation of the benzyl ring that is solvent exposed and protrudes from the active site. TL-3 has two benzyl rings that protrude from both sides of the active site and do not form any interactions with residues in the active site, consequently accounting for most of its flexibility. The flexibility of pepstatin A is due to it being a hexa-peptide, which contains many more rotatable bonds than the other two inhibitors.

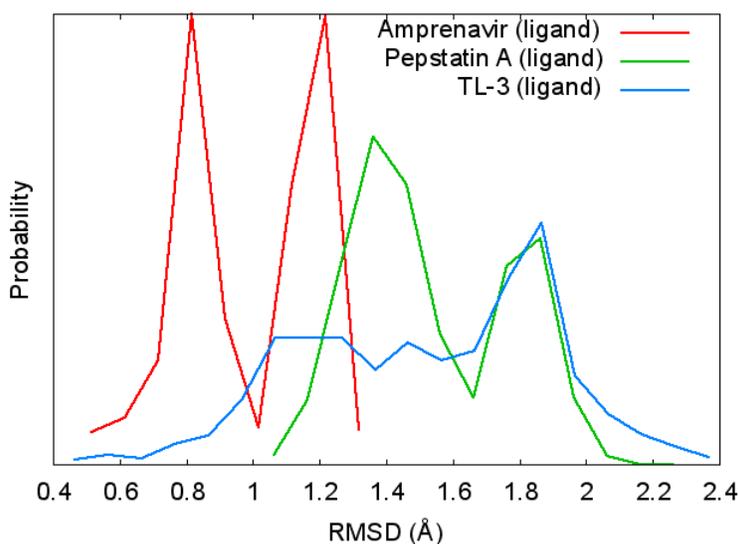


Figure 7-4. Histogram of the RMSD of the protease inhibitors: amprenavir, pepstatin A and TL-3.

Figure 7-5 is a histogram of the RMSD distribution of the flap residues 48-66 of each monomer. The minimized crystal structures were used as the reference. The flaps of the XMRV (pepstatin A) are the most flexible of the three inhibitors. Its RMSD distribution is fairly broad and spans from approximately .7 -1.6 Å. The flaps of XMRV (amprenavir) and (TL-3) exhibit a unique trend; XMRV (amprenavir) has the narrowest

RMSD distribution but its maximum peak is shifted to the right of the maximum peak of XMRV (TL3) indicating that XMRV (amprenavir) is farther from its reference.

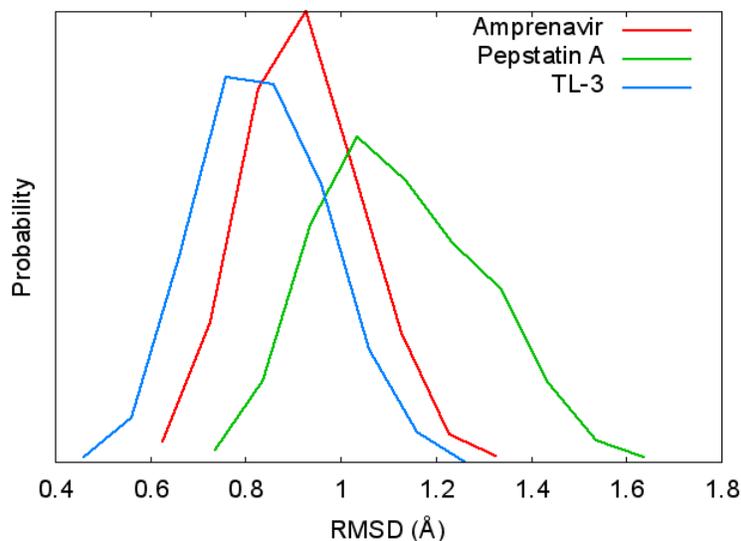


Figure 7-5. Histogram of the RMSD of the flap residues 48-66 of each monomer.

### 7.3.2 Hydrogen Bonding Analysis

Because the ligands bind non-covalently, the direct hydrogen bonds and water-mediated hydrogen bonds between the ligand and protein were calculated, Tables 7-1, 7-2 and 7-3. TL-3 makes the most hydrogen bonds with residues in the active site. None of the ligands formed any water mediated hydrogen bonds with the active site residues >5%. All of the ligands central hydroxyl groups(s) interact strongly with the catalytic aspartates Asp32 (protonated) and Asp32' (deprotonated). TL-3 and pepstatin A form hydrogen bonds with the flap residues Gln55 of each monomer. Amprenavir does not form any hydrogen bonds with Gln55 but instead hydrogen bonds to the flap residue Ala57. Pepstatin A is the only inhibitor to make any substantial interactions with residues located at the base of the active site, Gly34. TL-3 hydrogen bonds to Gln36

and Amprenavir hydrogen bonds to His37, which are residues located on the side of the active site.

Table 7-1. Hydrogen bond analysis XMRV (amprenavir)

Amprenavir		
Acceptor	Donor	%Occupancy
Asp32'—O	APV O—H	97
APV—O	Asp32 O—H	79
APV—O	ALA57 N—H	42
HIS37'—O	APV N—H	34

Table 7-2. Hydrogen bond analysis XMRV (pepstatin A)

Pepstatin A		
Acceptor	Donor	%Occupancy
Asp32'—O	Pep O—H	99
Pep—O	Asp32 O—H	94
Pep—O	Gln55' N—H	63
Gln55'—O	Pep N—H	51
Gly34—O	Pep N—H	45
Pep—O	Gln55 N—H	30

Table 7-3. Hydrogen bond analysis XMRV (TL-3)

TL-3		
Acceptor	Donor	%Occupancy
Asp32'—O	TL3 O—H	97
TL3—O	Gln36' N—H	82

Table 7-3. Continued

Acceptor	Donor	%Occupancy
Gln55—O	TL3 N—H	76
TL3—O	Gln36 N—H	74
Gln55'—O	TL3 N—H	71
TL3—O	Asp32 O—H	54
TL3—O	Gln55 N—H	53
TL3—O	Gln55' N—H	50

### 7.3.3 Molecular Mechanics Poisson Boltzmann Surface Area Calculations

TL-3 had the most favorable binding affinity of the three inhibitors, Table 7-4. The inhibitor TL-3 made the most hydrogen bond contacts with XMRV and as a result it has the highest  $\Delta H_{\text{gas}}$  term. Amprenavir has the lowest entropic and desolvation penalty. The computed binding affinities in Table 7-4 do not coincide with  $K_i$  values published in a recent kinetic study.<sup>217</sup> Li *et al.*<sup>217</sup> found that amprenavir had the lowest  $K_i$  value of the inhibitors used in their study. However our results do agree with Li *et al.* in that pepstatin A performs the worst of three inhibitors. The reason why results do not corroborate those published by Li *et al.* could be due to the following: inadequate sampling and the MD simulations need to be extended to a longer time scale, MMPBSA is an end-state method and assumes that there is little difference in the apo and bound forms of the enzyme or in spite of TL-3 and pepstatin A forming lots of interactions with the XMRV, the flaps of the protein are not flexible enough to accommodate the size of them and explains why the smaller inhibitor amprenavir prevails.

Table 7-4. MMPBSA results for the inhibitors: amprenavir, pepstatin A, and TL-3. All units are in kcal/mol

System	$\Delta H_{\text{gas}}$	$\Delta G_{\text{solvation}}$	$T\Delta S_{\text{nmode}}$	$\Delta G_{\text{bind}}$
Amprenavir	$-113.22 \pm 0.22$	$56.35 \pm 0.14$	$-28.26 \pm 1.12$	$-28.61 \pm 1.01$
Pepstatin A	$-157.11 \pm 0.18$	$94.71 \pm 0.15$	$-39.78 \pm 0.74$	$-22.62 \pm 0.68$
TL-3	$-195.21 \pm 0.27$	$108.38 \pm 0.16$	$-39.50 \pm 2.00$	$-47.33 \pm 1.87$

### 7.3.4 Potential Inhibitor Interactions

Residues in the active site that a potential protease inhibitor of XMRV could interact with are illustrated in Figure 7-6. Although pepstatin A and TL-3 forms two hydrogen bonds with residues in the flaps, the flaps of XMRV (pepstatin A) when compared to XMRV (amprenavir) are more flexible and the difference between XMRV (amprenavir/TL-3) is minute. An inhibitor that targets either one or both residues Gln55 and Ala57 will be sufficient enough to ensure the closure of the flaps. The positioning of the backbone carbonyl of Gly56 could potentially make it a good target but it is solvent exposed.

The scaffold HEA (amprenavir) and statine (pepstatin A) interacted the best with both catalytic aspartates—the scaffold HEA might be a better choice to pursue to increase its selectivity therefore diminishing its toxicity, but further investigations are warranted. The interactions of the central hydroxyl groups of TL-3 with Asp32' (deprotonated) are comparable for the two inhibitors but TL-3 interacted the least with Asp32 (protonated).

None of the inhibitors formed a hydrogen bond directly or through a water molecule with the residue Cys88, which is located in the 86-90 loop of the active site.

The side chain thiol group could act as either a hydrogen bond donor or hydrogen bond acceptor.

Gln36 and His37 are residues that inhibitors could form interactions with either their backbone or side chains atoms. The backbone atoms of Gln36 and His37 residues might serve as a more attractive target because the side chains of both residues protrude from the active site and are solvent exposed. Although amprenavir and TL-3 form interactions with either Gln36 or His37, none of the inhibitors form interactions with both these residues. The potency of the inhibitor might be increased if it formed interactions with both residues.

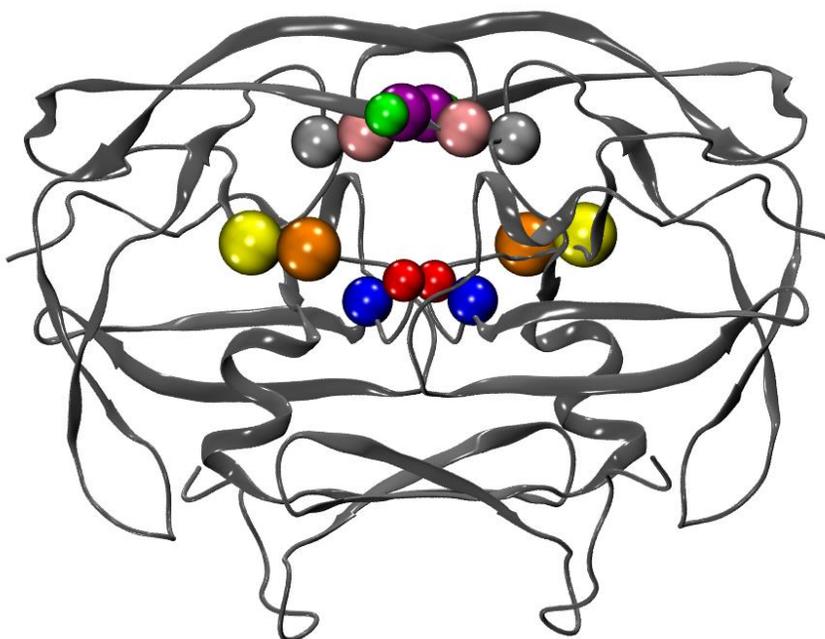


Figure 7-6. Schematic representation of XMRV protease. Represented by spheres are the residues in the active site that potential inhibitors can target. Asp32 (blue), Gly34 (red), Gln36 (orange), His37 (yellow), Gln55 (green), Gly56 (purple), Ala57 (pink) and (Cys88 (silver)

#### 7.4 Conclusions and Future Work

Our MD simulations provide preliminary information as to how to develop a potent inhibitor for XMRV. The HEA or the statine are good scaffolds that interact

strongly with the catalytic aspartates. The flap residues Gln55 and Ala57 are good targets for inhibitors, Trp58 and Gly57 might be new considerations. Neither of the inhibitors targeted these residues. Although pepstatin A formed interactions with the backbone carbonyl of Gly34, located at the base in of the active site, it was only for approximately 30% of the time therefore there is room for improvement. Residues that are located in close proximity to the catalytic aspartates generally have a lower mutational rate and therefore should be targeted.

While the inhibitors TL-3 and pepstatin A make more hydrogen bonds with residues in the active site than amprenavir, the  $K_i$  of the amprenavir was determined to be lower. One possibility might be that flaps of XMRV are flexible enough to accommodate the larger inhibitors TL-3 and pepstatin A and explains why they are not as effective inhibiting the virus.

Future work on XRMV protease will include: extending the time scale of the inhibitor bound simulations, using SMD simulations to gain insight on the dissociation pathway of an inhibitor, and performing long scale simulations of the apo form of the XMRV protease to in order to study the dynamics of the flaps.

## CHAPTER 8 CONCLUSIONS AND FUTURE WORK

Through the use of CpHMD simulations we determined that the two catalytic aspartates of the protease are mono-protonated in apo as well as the bound form for the inhibitors used in this study. Furthermore, either of the catalytic aspartates can be protonated and the proton is easily transferred due to a low-barrier hydrogen bond. Future work will include running these simulations where the solvent is instead modeled explicitly.

We explained how the active site and non-active site mutations G48T/L89M and N88D/L90M confer drug resistance against the protease inhibitor saquinavir in subtype B and C. In both cases the interactions of saquinavir with other residues in the active site were altered. Next simulations running only one mutation will provide insight as how much each mutation contributes to the overall resistance as well as its impact on the dynamics of the protease. More rigorous methods such as TI will be used to calculate the binding affinities.

The effects of the stabilizing mutations were investigated on the EPR distance measured profiles. Restoring the HIVPR to its active form (N25D) caused all the inhibitors to close the flaps. The most probable distances of the inhibitor bound form of HIVPR were found to be in good agreement with those obtained from previous EPR studies where the stabilizing mutations were included. The wild-type and LAI distance distributions did not correlate, and further investigation is warranted.

XMRV is a newly discovered retrovirus that has been linked to prostate cancer and chronic fatigue syndrome although it is still controversial. MD simulations with XMRV bound to protease inhibitors amprenavir, TL-3 and pepstatin A provide

information on the how to design a potent inhibitor. With information obtained from the study a pharmacophore can be developed and used to screen possible inhibitors.

APPENDIX  
HIV-1 PROTEASE SEQUENCES

Table A-1. HIV-1 Protease Subtype B Sequence (Wild-type). Residues 1 – 50 (first row) 51 – 99 (second row)

PQITLWQRPL	VTIKIGGQLK	EALLDTGADD	TVLEEMSLPG	RWKPKMIGGI
GGFIKVRQYD	QILIEICGHK	AIGTVLVGPT	PVNIIGRNLL	TQIGCTLNF

Table A-2. HIV-1 Protease Subtype B Sequence (LAI'). Residues 1 – 50 (first row) 51 – 99 (second row)

PQITLWKRPL	VTIKIGGQLK	EALLNTGADD	TVIEEMSLPG	RWKPKMIGGI
GGFIKVRQYD	QIIIEIAGHK	AIGTVLVGPT	PVNIIGRNLL	TQIGATLNF

Table A-3. HIV-1 Protease Subtype C Sequence (Wild-type). Residues 1 – 50 (first row) 51 – 99 (second row)

PQITLWQRPL	VSIKVGQIK	EALLDTGADD	TVLEEIALPG	RWKPKMIGGI
GGFIKVRQYD	QILIEICGKK	AIGTVLVGPT	PVNIIGRNML	TQLGCTLNF

Table A-4. HIV-1 Protease MDR769 Sequence (Drug Resistant Construct). Residues 1 – 50 (first row) 51 – 99 (second row)

PQITLWQRPI	VTIKIGGQLK	EALLDTGADD	TVLEEVNLPG	RWKPKLIGGI
GGFVKVRQYD	QVPIEICGHK	VIGTVLVGPT	PANVIGRNLM	TQIGCTLNF

Table A-5. HIV-1 Protease V6 Sequence (Drug Resistant Construct). Residues 1 – 50 (first row) 51 – 99 (second row)

PQITLWQRPL	VTIKIGGQLR	EALLDTGADD	TIFEEISLPG	RWKPKMIGGI
GGFIKVRQYD	QIPIEICGHK	VIGTVLVGPT	PANIIGRNLM	TQIGCTLNF

## LIST OF REFERENCES

- (1) Weiss, A.; Hollander, H.; Stobo, J. *Annu. Rev. Med.* **1985**, *36*, 545–562.
- (2) Kuznetsov, Y. G.; Victoria, J. G.; Robinson, W. E.; McPherson, A. *J. Virol.* **2003**, *77*, 11896–11909.
- (3) King, S. R. *Ann. Emerg. Med.* **1994**, *24*, 443–449.
- (4) Lee, C. H.; Saksela, K.; Mirza, U. A.; Chait, B. T.; Kuriyan, J. *Cell* **1996**, *85*, 931–942.
- (5) Collins, K. L.; Chen, B. K.; Kalams, S. A.; Walker, B. D.; Baltimore, D. *Nature* **1998**, *391*, 397–401.
- (6) Zapp, M. L.; Green, M. R. *Nature* **1989**, *342*, 714–716.
- (7) Ruben, S.; Perkins, A.; Purcell, R.; Joung, K.; Sia, R.; Burghoff, R.; Haseltine, W. A.; Rosen, C. A. *J. Virol.* **1989**, *63*, 1–8.
- (8) Moon, H. S.; Yang, J.-S. *Mol. Cells* **2006**, *21*, 7–20.
- (9) Rogel, M. E.; Wu, L. I.; Emerman, M. *J. Virol.* **1995**, *69*, 882–888.
- (10) Whitford, D. *Proteins: structure and function*; J. Wiley & Sons, 2005.
- (11) Mariani, R.; Chen, D.; Schröfelbauer, B.; Navarro, F.; König, R.; Bollman, B.; Münk, C.; Nymark-McMahon, H.; Landau, N. R. *Cell* **2003**, *114*, 21–31.
- (12) Conticello, S. G.; Harris, R. S.; Neuberger, M. S. *Curr. Biol.* **2003**, *13*, 2009–2013.
- (13) Fujita, K.; Omura, S.; Silver, J. *J. Gen. Virol.* **1997**, *78*, 619–625.
- (14) Robertson, D. L.; Anderson, J. P.; Bradac, J. A.; Carr, J. K.; Foley, B.; Funkhouser, R. K.; Gao, F.; Hahn, B. H.; Kalish, M. L.; Kuiken, C.; Learn, G. H.; Leitner, T.; McCutchan, F.; Osmanov, S.; Peeters, M.; Pieniazek, D.; Salminen, M.; Sharp, P. M.; Wolinsky, S.; Korber, B. *Science* **2000**, *288*, 55d.
- (15) Gürtler, L. G.; Hauser, P. H.; Eberle, J.; von Brunn, A.; Knapp, S.; Zekeng, L.; Tsague, J. M.; Kaptue, L. *J. Virol.* **1994**, *68*, 1581–1585.
- (16) Simon, F.; Maucière, P.; Roques, P.; Loussert-Ajaka, I.; Müller-Trutwin, M. C.; Saragosti, S.; Georges-Courbot, M. C.; Barré-Sinoussi, F.; Brun-Vézinet, F. *Nat. Med.* **1998**, *4*, 1032–1037.
- (17) Plantier, J.-C.; Leoz, M.; Dickerson, J. E.; De Oliveira, F.; Cordonnier, F.; Lemée, V.; Damond, F.; Robertson, D. L.; Simon, F. *Nat. Med.* **2009**, *15*, 871–872.

- (18) Penny, M. A.; Thomas, S. J.; Douglas, N. W.; Ranjbar, S.; Holmes, H.; Daniels, R. S. *AIDS Res. Hum. Retroviruses* **1996**, *12*, 741–747.
- (19) Archer, J.; Robertson, D. L. *AIDS* **2007**, *21*, 1693–1700.
- (20) Hemelaar, J.; Gouws, E.; Ghys, P. D.; Osmanov, S. *AIDS Lond. Engl.* **2006**, *20*, W13–23.
- (21) Baeten, J. M.; Chohan, B.; Lavreys, L.; Chohan, V.; McClelland, R. S.; Certain, L.; Mandaliya, K.; Jaoko, W.; Overbaugh, J. *J. Infect. Dis.* **2007**, *195*, 1177–1180.
- (22) Chan, D. C.; Fass, D.; Berger, J. M.; Kim, P. S. *Cell* **1997**, *89*, 263–273.
- (23) Pantophlet, R.; Burton, D. R. *Annu. Rev. Immunol.* **2006**, *24*, 739–769.
- (24) Dalgleish, A. G.; Beverley, P. C. L.; Clapham, P. R.; Crawford, D. H.; Greaves, M. F.; Weiss, R. A. *Nature* **1984**, *312*, 763–767.
- (25) Coakley, E.; Petropoulos, C. J.; Whitcomb, J. M. *Curr. Opin. Infect. Dis.* **2005**, *18*, 9–15.
- (26) Deng, H. K.; Unutmaz, D.; KewalRamani, V. N.; Littman, D. R. *Nature* **1997**, *388*, 296–300.
- (27) Liao, F.; Alkhatib, G.; Peden, K. W. C.; Sharma, G.; Berger, E. A.; Farber, J. M. *J. Exp. Med.* **1997**, *185*, 2015–2023.
- (28) Chan, D. C.; Kim, P. S. *Cell* **1998**, *93*, 681–684.
- (29) Robertson, D. L.; Hahn, B. H.; Sharp, P. M. *J. Mol. Evol.* **1995**, *40*, 249–259.
- (30) Pollard, V. W.; Malim, M. H. *Annu. Rev. Microbiol.* **1998**, *52*, 491–532.
- (31) Zheng, Y.-H.; Lovsin, N.; Peterlin, B. M. *Immunol. Lett.* **2005**, *97*, 225–234.
- (32) Hiscott, J.; Kwon, H.; Génin, P. *J. Clin. Invest.* **2001**, *107*, 143–151.
- (33) Bray, M.; Prasad, S.; Dubay, J. W.; Hunter, E.; Jeang, K. T.; Rekosh, D.; Hammarskjöld, M. L. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 1256–1260.
- (34) Witte, O. N.; Baltimore, D. *J. Virol.* **1978**, *26*, 750–761.
- (35) Crawford, S.; Goff, S. P. *J. Virol.* **1985**, *53*, 899–907.

- (36) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A.; Scolnick, E. M.; Sigal, I. S. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 4686–4690.
- (37) Navia, M. A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C.-T.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. *Nature* **1989**, *337*, 615–620.
- (38) Wlodawer, A.; Miller, M.; Jaskólski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. *Science* **1989**, *245*, 616–621.
- (39) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B.; Wlodawer, A. *Science* **1989**, *246*, 1149–1152.
- (40) Pearl, L.; Blundell, T. *FEBS Lett.* **1984**, *174*, 96–101.
- (41) Strisovsky, K.; Tessmer, U.; Langner, J.; Konvalinka, J.; Kräusslich, H. G. *Protein Sci. Publ. Protein Soc.* **2000**, *9*, 1631–1641.
- (42) Ingr, M.; Uhlíková, T.; Strisovský, K.; Majerová, E.; Konvalinka, J. *Protein Sci. Publ. Protein Soc.* **2003**, *12*, 2173–2182.
- (43) Ishima, R.; Freedberg, D. I.; Wang, Y. X.; Louis, J. M.; Torchia, D. A. *Struct. Lond. Engl. 1993* **1999**, *7*, 1047–1055.
- (44) Rick, S. W.; Erickson, J. W.; Burt, S. K. *Proteins* **1998**, *32*, 7–16.
- (45) Trylska, J.; Tozzini, V.; Chang, C. A.; McCammon, J. A. *Biophys. J.* **2007**, *92*, 4179–4187.
- (46) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 915–920.
- (47) Freedberg, D. I.; Ishima, R.; Jacob, J.; Wang, Y.-X.; Kustanovich, I.; Louis, J. M.; Torchia, D. A. *Protein Sci.* **2002**, *11*, 221–232.
- (48) Scott, W. R.; Schiffer, C. A. *Struct. Lond. Engl. 1993* **2000**, *8*, 1259–1265.
- (49) Meagher, K. L.; Carlson, H. A. *Proteins Struct. Funct. Bioinforma.* **2005**, *58*, 119–125.
- (50) Li, D.; Ji, B.; Hwang, K.; Huang, Y. *J. Phys. Chem. B* **2010**, *114*, 3060–3069.

- (51) De Oliveira, Tulio, T.; Engelbrecht, S.; Janse van Rensburg, E.; Gordon, M.; Bishop, K.; zur Megede, J.; Barnett, S. W.; Cassol, S. *J Virol* **2003**, *77*, 9422–9430.
- (52) Meek, T. D.; Rodriguez, E. J.; Angeles, T. S. *Methods Enzymol.* **1994**, *241*, 127–156.
- (53) Hyland, L. J.; Tomaszek, T. A.; Meek, T. D. *Biochemistry (Mosc.)* **1991**, *30*, 8454–8463.
- (54) Ogden, R. C.; Flexner, C. W. *Protease Inhibitors in AIDS Therapy*; Taylor & Francis, 2001.
- (55) Ratner, L.; Haseltine, W.; Patarca, R.; Livak, K. J.; Starcich, B.; Josephs, S. F.; Doran, E. R.; Rafalski, J. A.; Whitehorn, E. A.; Baumeister, K.; Ivanoff, L.; Petteway, S. R.; Pearson, M. L.; Lautenberger, J. A.; Papas, T. S.; Ghrayeb, J.; Chang, N. T.; Gallo, R. C.; Wong-Staal, F. *Nature* **1985**, *313*, 277–284.
- (56) Toh, H.; Ono, M.; Saigo, K.; Miyata, T. *Nature* **1985**, *315*, 691–691.
- (57) Kramer, R. A.; Schaber, M. D.; Skalka, A. M.; Ganguly, K.; Wong-Staal, F.; Reddy, E. P. *Science* **1986**, *231*, 1580–1584.
- (58) Sanchez-Pescador, R.; Power, M. D.; Barr, P. J.; Steimer, K. S.; Stempien, M. M.; Brown-Shimer, S. L.; Gee, W. W.; Renard, A.; Randolph, A.; Levy, J. A.; Et, A. *Science* **1985**, *227*, 484–492.
- (59) Wolfenden, R. *Annu. Rev. Biophys. Bioeng.* **1976**, *5*, 271–306.
- (60) Dreyer, G. B.; Metcalf, B. W.; Tomaszek, T. A.; Carr, T. J.; Chandler, A. C.; Hyland, L.; Fakhoury, S. A.; Magaard, V. W.; Moore, M. L.; Strickler, J. E. *Proc. Natl. Acad. Sci.* **1989**, *86*, 9752–9756.
- (61) Moore, M. L.; Bryan, W. M.; Fakhoury, S. A.; Magaard, V. W.; Huffman, W. F.; Dayton, B. D.; Meek, T. D.; Hyland, L.; Dreyer, G. B.; Metcalf, B. W.; Strickler, J. E.; Gorniak, J. G.; Debouck, C. *Biochem. Biophys. Res. Commun.* **1989**, *159*, 420–425.
- (62) Tomasselli, A. G.; Olsen, M. K.; Hui, J. O.; Staples, D. J.; Sawyer, T. K.; Henrikson, R. L.; Tomich, C. S. C. *Biochemistry (Mosc.)* **1990**, *29*, 264–269.
- (63) Velazquez-Campoy, A.; Todd, M. J.; Freire, E. *Biochemistry (Mosc.)* **2000**, *39*, 2201–2207.
- (64) Velazquez-Campoy A.; Kiso Y.; Freire E. *Arch. Biochem. Biophys.* **2001**, *390*, 169–175.

- (65) King, N. M.; Prabu-Jeyabalan, M.; Bandaranayake, R. M.; Nalam, M. N. L.; Nalivaika, E. A.; Özen, A.; Haliloğlu, T.; Yilmaz, N. K.; Schiffer, C. A. *ACS Chem. Biol.* **2012**, *7*, 1536–1546.
- (66) Boggetto, N.; Reboud-Ravaux, M. *Biol. Chem.* **2002**, *383*, 1321–1324.
- (67) Breccia, P.; Boggetto, N.; Pérez-Fernández, R.; Van Gool, M.; Takahashi, M.; René, L.; Prados, P.; Badet, B.; Reboud-Ravaux, M.; de Mendoza, J. *J. Med. Chem.* **2003**, *46*, 5196–5207.
- (68) Perryman, A. L.; Lin, J.; McCammon, J. A. *Biopolymers* **2006**, *82*, 272–284.
- (69) Perryman, A. L.; Zhang, Q.; Soutter, H. H.; Rosenfeld, R.; McRee, D. E.; Olson, A. J.; Elder, J. E.; David Stout, C. *Chem. Biol. Drug Des.* **2010**, *75*, 257–268.
- (70) Genoni, A.; Morra, G.; Merz, K. M.; Colombo, G. *Biochemistry (Mosc.)* **2010**, *49*, 4283–4295.
- (71) Ishima, R.; Ghirlando, R.; Tözsér, J.; Gronenborn, A. M.; Torchia, D. A.; Louis, J. M. *J. Biol. Chem.* **2001**, *276*, 49110–49116.
- (72) Weber, I. T. *J. Biol. Chem.* **1990**, *265*, 10492–10496.
- (73) Todd, M. J.; Semo, N.; Freire, E. *J. Mol. Biol.* **1998**, *283*, 475–488.
- (74) Gustchina, A.; Weber, I. T. *Proteins Struct. Funct. Bioinforma.* **1991**, *10*, 325–339.
- (75) Zhang, Z. Y.; Poorman, R. A.; Maggiora, L. L.; Heinrikson, R. L.; Kézdy, F. J. *J. Biol. Chem.* **1991**, *266*, 15591–15594.
- (76) Schramm, H. J.; Nakashima, H.; Schramm, W.; Wakayama, H.; Yamamoto, N. *Biochem. Biophys. Res. Commun.* **1991**, *179*, 847–851.
- (77) Zutshi, R.; Franciskovich, J.; Shultz, M.; Schweitzer, B.; Bishop, P.; Wilson, M.; Chmielewski, J. *J. Am. Chem. Soc.* **1997**, *119*, 4841–4845.
- (78) Rhee, S.-Y.; Gonzales, M. J.; Kantor, R.; Betts, B. J.; Ravela, J.; Shafer, R. W. *Nucleic Acids Res.* **2003**, *31*, 298–303.
- (79) Rhee, S.-Y.; Kantor, R.; Katzenstein, D. A.; Camacho, R.; Morris, L.; Sirivichayakul, S.; Jorgensen, L.; Brigido, L. F.; Schapiro, J. M.; Shafer, R. W. *AIDS Lond. Engl.* **2006**, *20*, 643–651.
- (80) Shafer, R. W. *J. Infect. Dis.* **2006**, *194*, S51–S58.

- (81) Maguire, M. F.; Guinea, R.; Griffin, P.; Macmanus, S.; Elston, R. C.; Wolfram, J.; Richards, N.; Hanlon, M. H.; Porter, D. J. T.; Wrin, T.; Parkin, N.; Tisdale, M.; Furfine, E.; Petropoulos, C.; Snowden, B. W.; Kleim, J.-P. *J. Virol.* **2002**, *76*, 7398–7406.
- (82) Ho, S. K.; Coman, R. M.; Bunger, J. C.; Rose, S. L.; O'Brien, P.; Munoz, I.; Dunn, B. M.; Sleasman, J. W.; Goodenow, M. M. *Virology* **2008**, *378*, 272–281.
- (83) Rose, R. B.; Craik, C. S.; Stroud, R. M. *Biochemistry (Mosc.)* **1998**, *37*, 2607–2621.
- (84) Wilson, S. I.; Phylip, L. H.; Mills, J. S.; Gulnik, S. V.; Erickson, J. W.; Dunn, B. M.; Kay, J. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzym.* **1997**, *1339*, 113–125.
- (85) Lech, W. J.; Wang, G.; Yang, Y. L.; Chee, Y.; Dorman, K.; McCrae, D.; Lazzeroni, L. C.; Erickson, J. W.; Sinsheimer, J. S.; Kaplan, A. H. *J. Virol.* **1996**, *70*, 2038–2043.
- (86) Weber, I. T.; Agniswamy, J. *Viruses* **2009**, *1*, 1110–1136.
- (87) Johnson, V. A.; Calvez, V.; Gunthard, H. F.; Paredes, R.; Pillay, D.; Shafer, R. W.; Wensing, A. M.; Richman, D. D. *Top. Antivir. Med.* **2013**, *21*, 6–14.
- (88) King, N. M.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Schiffer, C. A. *Chem. Biol.* **2004**, *11*, 1333–1338.
- (89) Foulkes-Murzycki, J. E.; Scott, W. R. P.; Schiffer, C. A. *Structure* **2007**, *15*, 225–233.
- (90) De Vera, I. M. S.; Smith, A. N.; Dancel, M. C. A.; Huang, X.; Dunn, B. M.; Fanucci, G. E. *Biochemistry (Mosc.)* **2013**, *52*, 3278–3288.
- (91) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, *27*, 1208.
- (92) Leach, A. *Molecular Modelling: Principles and Applications*; 2nd ed.; Prentice Hall, 2001.
- (93) Frenkel, D.; Smit, B. *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications*; 2nd ed.; Academic Press, 2001.
- (94) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; 2nd ed.; Wiley, 2004.
- (95) Verlet, L. *Phys. Rev.* **1967**, *159*, 98.
- (96) Hockney, R. *Methods Comput Phys* **1970**, *9*, 136–211.

- (97) Swope, W.; Andersen, H.; Berens, P.; Wilson, K. *J. Chem. Phys.* **1982**, *76*, 649, 637.
- (98) Beeman, D. *J. Comput. Phys.* **1976**, *20*, 130–139.
- (99) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.
- (100) MacKerell; Bashford, D.; Bellott; Dunbrack; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (101) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
- (102) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- (103) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (104) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910–8922.
- (105) Weiser, J.; Shenkin, P. S.; Still, W. C. *J. Comput. Chem.* **1999**, *20*, 217–230.
- (106) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144–1149.
- (107) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.
- (108) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (109) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
- (110) Swails, J. M.; Roitberg, A. E. *J. Chem. Theory Comput.* **2012**, *8*, 4393–4404.
- (111) Baptista, A. M.; Martel, P. J.; Petersen, S. B. *Proteins Struct. Funct. Bioinforma.* **1997**, *27*, 523–544.
- (112) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. *J. Chem. Phys.* **2002**, *117*, 4184–4200.
- (113) Lee, A. C.; Crippen, G. M. *J. Chem. Inf. Model.* **2009**, *49*, 2013–2033.

- (114) Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comput. Chem.* **2004**, *25*, 2038–2048.
- (115) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (116) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (117) Itoh, S. G.; Damjanović, A.; Brooks, B. R. *Proteins Struct. Funct. Bioinforma.* **2011**, *79*, 3420–3436.
- (118) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919–11929.
- (119) Pierce, L. C. T.; Salomon-Ferrer, R.; Augusto F. de Oliveira, C.; McCammon, J. A.; Walker, R. C. *J. Chem. Theory Comput.* **2012**, *8*, 2997–3002.
- (120) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (121) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.
- (122) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.
- (123) Nicholson, L. K.; Yamazaki, T.; Torchia, D. A.; Grzesiek, S.; Bax, A.; Stahl, S. J.; Kaufman, J. D.; Wingfield, P. T.; Lam, P. Y.; Jadhav, P. K. *Nat. Struct. Biol.* **1995**, *2*, 274–280.
- (124) Chen, X.; Tropsha, A. *J. Med. Chem.* **1995**, *38*, 42–48.
- (125) Xie, D.; Gulnik, S.; Collins, L.; Gustchina, E.; Suvorov, L.; Erickson, J. W. *Biochemistry (Mosc.)* **1997**, *36*, 16166–16172.
- (126) Wang, W.; Kollman, P. A. *J. Mol. Biol.* **2000**, *303*, 567–582.
- (127) Ido, E.; Han, H. P.; Kezdy, F. J.; Tang, J. *J. Biol. Chem.* **1991**, *266*, 24359–24366.
- (128) Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P. H.; Poljak, R. J. *Biochimie* **1991**, *73*, 1391–1396.
- (129) Shen, C.-H.; Wang, Y.-F.; Kovalevsky, A. Y.; Harrison, R. W.; Weber, I. T. *FEBS J.* **2010**, *277*, 3699–3714.
- (130) Muzammil, S.; Armstrong, A. A.; Kang, L. W.; Jakalian, A.; Bonneau, P. R.; Schmelmer, V.; Amzel, L. M.; Freire, E. *J Virol* **2007**, *81*, 5144–5154.

- (131) Agniswamy, J.; Shen, C.-H.; Aniana, A.; Sayer, J. M.; Louis, J. M.; Weber, I. T. *Biochemistry (Mosc.)* **2012**, *51*, 2819–2828.
- (132) Munshi, S.; Chen, Z.; Li, Y.; Olsen, D. B.; Fraley, M. E.; Hungate, R. W.; Kuo, L. C. *Acta Crystallogr. D Biol. Crystallogr.* **1998**, *54*, 1053–1060.
- (133) Kaldor, S. W.; Kalish, V. J.; Davies, J. F.; Shetty, B. V.; Fritz, J. E.; Appelt, K.; Burgess, J. A.; Campanale, K. M.; Chirgadze, N. Y.; Clawson, D. K.; Dressman, B. A.; Hatch, S. D.; Khalil, D. A.; Kosa, M. B.; Lubbehusen, P. P.; Muesing, M. A.; Patick, A. K.; Reich, S. H.; Su, K. S.; Tatlock, J. H. *J. Med. Chem.* **1997**, *40*, 3979–3985.
- (134) Kempf, D. J.; Marsh, K. C.; Denissen, J. F.; McDonald, E.; Vasavanonda, S.; Flentge, C. A.; Green, B. E.; Fino, L.; Park, C. H.; Kong, X. P. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 2484–2488.
- (135) Lam, P. Y. S.; Ru, Y.; Jadhav, P. K.; Aldrich, P. E.; DeLucca, G. V.; Eyermann, C. J.; Chang, C.-H.; Emmett, G.; Holler, E. R.; Daneker, W. F.; Li, L.; Confalone, P. N.; McHugh, R. J.; Han, Q.; Li, R.; Markwalder, J. A.; Seitz, S. P.; Sharpe, T. R.; Bachelier, L. T.; Rayner, M. M.; Klabe, R. M.; Shum, L.; Winslow, D. L.; Kornhauser, D. M.; Jackson, D. A.; Erickson-Viitanen, S.; Hodge, C. N. *J. Med. Chem.* **1996**, *39*, 3514–3525.
- (136) Adachi, M.; Ohhara, T.; Kurihara, K.; Tamada, T.; Honjo, E.; Okazaki, N.; Arai, S.; Shoyama, Y.; Kimura, K.; Matsumura, H.; Sugiyama, S.; Adachi, H.; Takano, K.; Mori, Y.; Hidaka, K.; Kimura, T.; Hayashi, Y.; Kiso, Y.; Kuroki, R. *Proc. Natl. Acad. Sci.* **2009**, *106*, 4641–4646.
- (137) Kear, J. L.; Blackburn, M. E.; Veloro, A. M.; Dunn, B. M.; Fanucci, G. E. *J. Am. Chem. Soc.* **2009**, *131*, 14650–14651.
- (138) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (139) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (140) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623–1641.
- (141) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383–394.
- (142) Sindhikara, D. J.; Kim, S.; Voter, A. F.; Roitberg, A. E. *J. Chem. Theory Comput.* **2009**, *5*, 1624–1631.

- (143) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (144) Sindhikara, D.; Meng, Y.; Roitberg, A. E. *J. Chem. Phys.* **2008**, *128*, 024103–024103–10.
- (145) Roe, D. R.; Cheatham, T. E. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.
- (146) Trylska, J.; Antosiewicz, J.; Geller, M.; Hodge, C. N.; Klabe, R. M.; Head, M. S.; Gilson, M. K. *Protein Sci. Publ. Protein Soc.* **1999**, *8*, 180–195.
- (147) Czodrowski, P.; Sotriffer, C. A.; Klebe, G. *J. Chem. Inf. Model.* **2007**, *47*, 1590–1598.
- (148) Piana, S.; Carloni, P. *Proteins Struct. Funct. Bioinforma.* **2000**, *39*, 26–36.
- (149) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. *Nat. Struct. Biol.* **1996**, *3*, 946–950.
- (150) Mardis, K. L.; Luo, R.; Gilson, M. K. *J. Mol. Biol.* **2001**, *309*, 507–517.
- (151) Antosiewicz, J.; Briggs, J. M.; Elcock, A. H.; Gilson, M. K.; McCammon, J. A. *J. Comput. Chem.* **1996**, *17*, 1633–1644.
- (152) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415–436.
- (153) Yamazaki, T.; Nicholson, Linda; Torchia, Dennis; Wingfield, Paul; Stahl, Stephen; Kaufman, Joshua; Eyermann, Charles; Hodge, C. Nicholas; Lam, Patrick; Ru, Yu; Jadhav, Prabhakar; Chang, Chong-Hwan; Weber, Patricia *J Am Chem Soc* **1994**, *116*, 10791–10792.
- (154) Schutz, C. N.; Warshel, A. *Proteins* **2001**, *44*, 400–417.
- (155) Wang, Y.-X.; Freedberg, D. I.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Kiso, Y.; Torchia, D. A. *Biochemistry (Mosc.)* **1996**, *35*, 9945–9950.
- (156) Wittayanarakul, K.; Hannongbua, S.; Feig, M. *J. Comput. Chem.* **2008**, *29*, 673–685.
- (157) Geller, M.; Miller, M.; Swanson, S. M.; Maizel, J. *Proteins Struct. Funct. Bioinforma.* **1997**, *27*, 195–203.
- (158) Harte, W. E.; Beveridge, D. L. *J. Am. Chem. Soc.* **1993**, *115*, 3883–3886.

- (159) Ky-Youb Nam; Chang, B. H.; Han, C. K.; Ahn, S. K.; No, K. T. *Bull.-Korean Chem. Soc.* **2003**, *24*, 817–823.
- (160) Das, A.; Mahale, S.; Prashar, V.; Bihani, S.; Ferrer, J.-L.; Hosur, M. V. *J. Am. Chem. Soc.* **2010**, *132*, 6366–6373.
- (161) Northrop, D. B. *Accounts Chem. Res.* **2001**, *34*, 790–797.
- (162) Jacobsen, H.; Hänggi, M.; Ott, M.; Duncan, I. B.; Owen, S.; Andreoni, M.; Vella, S.; Mous, J. *J. Infect. Dis.* **1996**, *173*, 1379–1387.
- (163) Tie, Y.; Kovalevsky, A. Y.; Boross, P.; Wang, Y.-F.; Ghosh, A. K.; Tozser, J.; Harrison, R. W.; Weber, I. T. *Proteins Struct. Funct. Bioinforma.* **2007**, *67*, 232–242.
- (164) Rosé, J. R.; Salto, R.; Craik, C. S. *J. Biol. Chem.* **1993**, *268*, 11939–11945.
- (165) Mildner, A. M.; Rothrock, D. J.; Leone, J. W.; Bannow, C. A.; Lull, J. M.; Reardon, I. M.; Sarcich, J. L.; Howe, W. J.; Tomich, C. S.; Smith, C. W. *Biochemistry (Mosc.)* **1994**, *33*, 9405–9413.
- (166) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (167) Nguyen, H.; Roe, D. R.; Simmerling, C. *J. Chem. Theory Comput.* **2013**, *9*, 2020–2034.
- (168) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (169) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38, 27–28.
- (170) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A. E.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; AMBER; University of California, San Francisco, 2012.
- (171) Blackburn, M. E.; Veloro, A. M.; Fanucci, G. E. *Biochemistry (Mosc.)* **2009**, *48*, 8765–8767.
- (172) Galiano, L.; Ding, F.; Veloro, A. M.; Blackburn, M. E.; Simmerling, C.; Fanucci, G. E. *J. Am. Chem. Soc.* **2009**, *131*, 430–431.

- (173) Hong, L.; Zhang, X. C.; Hartsuck, J. A.; Tang, J. *Protein Sci.* **2000**, *9*, 1898–1904.
- (174) Stoica, I.; Sadiq, S. K.; Coveney, P. V. *J. Am. Chem. Soc.* **2008**, *130*, 2639–2648.
- (175) Sanches, M.; Krauchenco, S.; Martins, N. H.; Gustchina, A.; Wlodawer, A.; Polikarpov, I. *J. Mol. Biol.* **2007**, *369*, 1029–1040.
- (176) Hou, T.; McLaughlin, W. A.; Wang, W. *Proteins* **2008**, *71*, 1163–1174.
- (177) Coman, R. M.; Robbins, A. H.; Fernandez, M. A.; Gilliland, C. T.; Sochet, A. A.; Goodenow, M. M.; McKenna, R.; Dunn, B. M. *Biochemistry (Mosc.)* **2008**, *47*, 731–743.
- (178) Matsuyama, S.; Aydan, A.; Ode, H.; Hata, M.; Sugiura, W.; Hoshino, T. *J. Phys. Chem. B* **2010**, *114*, 521–530.
- (179) Liu, X.; Dai, Q.; Li, L.; Xiu, Z. *J. Mol. Struct.* **2011**, *986*, 30–38.
- (180) Coman, R. M.; Robbins, A. H.; Goodenow, M. M.; Dunn, B. M.; McKenna, R. *Acta Crystallogr. D Biol. Crystallogr.* **2008**, *D64*, 754–763.
- (181) Mitsuya, Y.; Winters, M. A.; Fessel, W. J.; Rhee, S.-Y.; Hurley, L.; Horberg, M.; Schiffer, C. A.; Zolopa, A. R.; Shafer, R. W. *AIDS Res. Hum. Retroviruses* **2006**, *22*, 1300–1305.
- (182) Mahalingam, B.; Boross, P.; Wang, Y.; Louis, J. M.; Fischer, C. C.; Tozser, J.; Harrison, R. W.; Weber, I. T. *Proteins Struct. Funct. Bioinforma.* **2002**, *48*, 107–116.
- (183) Kovalevsky, A. Y.; Tie, Y.; Liu, F.; Boross, P. I.; Wang, Y.-F.; Leshchenko, S.; Ghosh, A. K.; Harrison, R. W.; Weber, I. T. *J. Med. Chem.* **2006**, *49*, 1379–1387.
- (184) Bihani, S. C.; Das, A.; Prashar, V.; Ferrer, J.-L.; Hosur, M. V. *Biochem. Biophys. Res. Commun.* **2009**, *389*, 295–300.
- (185) Fanucci, G. E.; Cafiso, D. S. *Curr. Opin. Struct. Biol.* **2006**, *16*, 644–653.
- (186) Galiano, L.; Bonora, M.; Fanucci, G. E. *J. Am. Chem. Soc.* **2007**, *129*, 11004–11005.
- (187) Pannier, M.; Veit, S.; Godt, A.; Jeschke, G.; Spiess, H. W. *J. Magn. Reson. San Diego Calif 1997* **2000**, *142*, 331–340.

- (188) Jeschke, G. *Chemphyschem Eur. J. Chem. Phys. Phys. Chem.* **2002**, *3*, 927–932.
- (189) Ding, F.; Layten, M.; Simmerling, C. *J. Am. Chem. Soc.* **2008**, *130*, 7184–7185.
- (190) Martin, P.; Vickrey, J. F.; Proteasa, G.; Jimenez, Y. L.; Wawrzak, Z.; Winters, M. A.; Merigan, T. C.; Kovari, L. C. *Struct. Lond. Engl. 1993* **2005**, *13*, 1887–1895.
- (191) Clemente, J. C.; Moose, R. E.; Hemrajani, R.; Whitford, L. R. S.; Govindasamy, L.; Reutzel, R.; McKenna, R.; Agbandje-McKenna, M.; Goodenow, M. M.; Dunn, B. M. *Biochemistry (Mosc.)* **2004**, *43*, 12141–12151.
- (192) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. *J. Am. Chem. Soc.* **1995**, *117*, 1181–1182.
- (193) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *J. Cheminformatics* **2012**, *4*, 17.
- (194) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A.F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian09*.
- (195) Columbus, L.; Hubbell, W. L. *Trends Biochem. Sci.* **2002**, *27*, 288–295.
- (196) Sezer, D.; Freed, J. H.; Roux, B. *J. Am. Chem. Soc.* **2009**, *131*, 2597–2605.
- (197) Torbeev, V. Y.; Raghuraman, H.; Mandal, K.; Senapati, S.; Perozo, E.; Kent, S. B. H. *J. Am. Chem. Soc.* **2009**, *131*, 884–885.
- (198) Urisman, A.; Molinaro, R. J.; Fischer, N.; Plummer, S. J.; Casey, G.; Klein, E. A.; Malathi, K.; Magi-Galluzzi, C.; Tubbs, R. R.; Ganem, D.; Silverman, R. H.; DeRisi, J. L. *PLoS Pathog* **2006**, *2*, e25.
- (199) Schlaberg, R.; Choe, D. J.; Brown, K. R.; Thaker, H. M.; Singh, I. R. *Proc Natl Acad Sci U S A* **2009**, *106*, 16351–16356.

- (200) Lombardi, V. C.; Ruscetti, F. W.; Das Gupta, J.; Pfof, M. A.; Hagen, K. S.; Peterson, D. L.; Ruscetti, S. K.; Bagni, R. K.; Petrow-Sadowski, C.; Gold, B.; Dean, M.; Silverman, R. H.; Mikovits, J. A. *Science* **2009**, *326*, 585–589.
- (201) Arnold, R. S.; Makarova, N. V.; Osunkoya, A. O.; Suppiah, S.; Scott, T. A.; Johnson, N. A.; Bhosle, S. M.; Liotta, D.; Hunter, E.; Marshall, F. F.; Ly, H.; Molinaro, R. J.; Blackwell, J. L.; Petros, J. A. *Urology* **2010**, *75*, 755–761.
- (202) Danielson, B. P.; Ayala, G. E.; Kimata, J. T. *J. Infect. Dis.* **2010**, *202*, 1470–1477.
- (203) Fischer, N.; Schulz, C.; Stieler, K.; Hohn, O.; Lange, C.; Drosten, C.; Aepfelbacher, M. *Emerg. Infect. Dis.* **2010**, *16*, 1000–1002.
- (204) Aloia, A. L.; Sfanos, K. S.; Isaacs, W. B.; Zheng, Q.; Maldarelli, F.; Marzo, A. M. D.; Rein, A. *Cancer Res.* **2010**, *70*, 10028–10033.
- (205) Barnes, E.; Flanagan, P.; Brown, A.; Robinson, N.; Brown, H.; McClure, M.; Oxenius, A.; Collier, J.; Weber, J.; Günthard, H. F.; Hirschel, B.; Fidler, S.; Phillips, R.; Frater, J. *J. Infect. Dis.* **2010**, *202*, 1482–1485.
- (206) Cornelissen, M.; Zorgdrager, F.; Blom, P.; Jurriaans, S.; Repping, S.; van Leeuwen, E.; Bakker, M.; Berkhout, B.; van der Kuyl, A. C. *PLoS ONE* **2010**, *5*, e12040.
- (207) Erlwein, O.; Kaye, S.; McClure, M. O.; Weber, J.; Wills, G.; Collier, D.; Wessely, S.; Cleare, A. *PloS One* **2010**, *5*, e8519.
- (208) Groom, H. C. T.; Boucherit, V. C.; Makinson, K.; Randal, E.; Baptista, S.; Hagan, S.; Gow, J. W.; Mattes, F. M.; Breuer, J.; Kerr, J. R.; Stoye, J. P.; Bishop, K. N. *Retrovirology* **2010**, *7*, 10.
- (209) Hohn, O.; Krause, H.; Barbarotto, P.; Niederstadt, L.; Beimforde, N.; Denner, J.; Miller, K.; Kurth, R.; Bannert, N. *Retrovirology* **2009**, *6*, 92.
- (210) Hong, P.; Li, J.; Li, Y. *Virol. J.* **2010**, *7*, 224.
- (211) Jeziorski, E.; Foulongne, V.; Ludwig, C.; Louhaem, D.; Chiocchia, G.; Segondy, M.; Rodière, M.; Sitbon, M.; Courgnaud, V. *Retrovirology* **2010**, *7*, 63.
- (212) Switzer, W. M.; Jia, H.; Hohn, O.; Zheng, H.; Tang, S.; Shankar, A.; Bannert, N.; Simmons, G.; Hendry, R. M.; Falkenberg, V. R.; Reeves, W. C.; Heneine, W. *Retrovirology* **2010**, *7*, 57.
- (213) Kim, S.; Kim, N.; Dong, B.; Boren, D.; Lee, S. A.; Das Gupta, J.; Gaughan, C.; Klein, E. A.; Lee, C.; Silverman, R. H.; Chow, S. A. *J. Virol.* **2008**, *82*, 9964–9977.

- (214) Smith, R.; Gottlieb, G.; Miller, A. D. *Retrovirology* **2010**, *7*, 70.
- (215) Singh, I. R.; Gorzynski, J. E.; Drobysheva, D.; Bassit, L.; Schinazi, R. F. *PLoS ONE* **2010**, *5*, e9948.
- (216) Li, M.; Dimaio, F.; Zhou, D.; Gustchina, A.; Lubkowski, J.; Dauter, Z.; Baker, D.; Wlodawer, A. *Nat. Struct. Mol. Biol.* **2011**, *18*, 227–229.
- (217) Li, M.; Gustchina, A.; Matúz, K.; Tözsér, J.; Namwong, S.; Goldfarb, N. E.; Dunn, B. M.; Wlodawer, A. *FEBS J.* **2011**, *278*, 4413–4424.
- (218) Matúz, K.; Mótyán, J.; Li, M.; Wlodawer, A.; Tözsér, J. *FEBS J.* **2012**, *279*, 3276–3286.
- (219) Slee, D. H.; Laslo, K. L.; Elder, J. H.; Ollmann, I. R.; Gustchina, A.; Kervinen, J.; Zdanov, A.; Wlodawer, A.; Wong, C.-H. *J. Am. Chem. Soc.* **1995**, *117*, 11867–11878.
- (220) Umezawa, H.; Aoyagi, T.; Morishima, H.; Matsuzaki, M.; Hamada, M. *J. Antibiot. (Tokyo)* **1970**, *23*, 259–262.
- (221) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739–1749.

## BIOGRAPHICAL SKETCH

Terry Dwight McGee Jr. was born in Macon, Georgia. He received his bachelor's degree in Chemistry from Florida Agricultural and Mechanical University in 2005. In the fall of 2008 he entered the University of Florida chemistry program to study physical chemistry with a specialization in computational chemistry. Upon entering the graduate program, he joined the lab of Prof. Adrian E. Roitberg. During his tenure at the University of Florida, he married Danielle McGee and they had three beautiful children: Brendan, Cameran, and Logan McGee.