

Exploring the Structure and Dynamics of HIV-1 PR by

MD Simulations

A Dissertation Presented

by

Fang Yu Ding

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Chemistry

Stony Brook University

August 2010

Stony Brook University

The Graduate School

Fang Yu Ding

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Carlos L. Simmerling - Thesis Advisor

Professor, Department of Chemistry

David F. Green – Chairperson of Defense

Assistant Professor, Department of Applied Mathematics and Statistics

Robert C. Rizzo – Committee Member

Assistant Professor, Department of Applied Mathematics and Statistics

Carol A. Carter – Outside Member

Professor, Department of Molecular Genetics and Microbiology,

Stony Brook University

This dissertation is accepted by the Graduate School

Lawrence Martin

Dean of the Graduate School

Abstract of the Dissertation

Exploring the Structure and Dynamics of HIV-1 PR by MD Simulations

by

Fang Yu Ding

Doctor of Philosophy

in

Chemistry

Stony Brook University

2010

Proteins are known to be dynamic molecules that undergo conformational fluctuations. A fundamental issue that remains to be clarified is whether there is a linkage between the dynamic nature of proteins and their catalytic function. The structural changes observed between ligand-bound and ligand-free forms of a given enzyme have classically been described as following an induced fit model. Rapidly accumulating experimental and computational data, however, has led to the emergence of novel theories, such as intrinsic ability of proteins to undergo conformational changes along directions that enable their function. Of particular interest are those conformations accessible near global free energy minima which are in equilibrium and separated by low energy barrier, also called substates. Because proteins only function in their native state, interconversions between these substates are important. As a result, a complete understanding of the mechanisms governing the interconversions between these substates

not only sheds light on how the enzyme works, but also has profound and practical implications for revealing new approaches to drug design.

Experiments permit us to visualize the structural flexibility and heterogeneity of proteins and assess their relevance to catalysis. On the theoretical side, molecular dynamics (MD) simulations at the atomic level have reached at a state where they can provide realistic models of biomolecular processes such as long timescale conformational transitions associated with protein functions. In addition, MD simulations can be used to interpret experimental data and determine relevant information concerning structural, dynamic and thermodynamic properties of targeted proteins. In turn, simulations rely on experimental observables for validation of a particular model or method.

In this work, both tools have been employed collaboratively to explore the structural and dynamic features of HIV-1 protease, a primary target for anti-retroviral agents. In a number of recent studies, X-ray crystallography, solution NMR, electron paramagnetic resonance (EPR) spectroscopy and theoretical data have provided evidence for the existence of a pre-existing equilibrium between different conformations of the enzyme in the absence of ligand, and have suggested that the protein dynamics are crucial for its catalytic function. Therefore, a complete understanding of conformational transitions of the enzyme may open new avenues in the design of more effective treatment regimes. Here, we present a hypothesis describing how the twisting of the backbone of the flap tips transforms the β -hairpin structure of each flap from the ‘closed’ conformation to the ‘semi-open’. In addition, we suggest that it is the various binding interactions within the protease dimer interface that govern the gating properties of the flaps; the opening of the flaps most likely results from the concerted partial dissociation of the dimer interface

facilitated by water dynamics. Significantly, a novel target for allosteric inhibition of the viral protease has also been predicted from our working model. This has a great potential in rational design of more effective treatment regimes.

Moreover, to explore how resistance caused by protease mutations arises, we collaborated with EPR experimentalists and performed a series of MD simulations on the spin-labeled wild-type and multi-drug resistant proteases. A combined analysis of the MD simulations and the EPR spectra suggests that the semi-open form is most likely the dominant configuration for the ligand-free wild-type HIV-PR; mutations conferring drug resistance may alter either the conformation of the flaps or the mobility of the flaps, or both.

Table of Contents

List of Figures	xi
List of Tables	xxiii
Acknowledgements	xxiv
Chapter 1	1
Introduction	1
1.1 Simulation Methodology	1
1.1.1 Force Fields	2
1.1.2 Solvation Effects.....	4
1.1.3 Advanced MD Simulation Methods	8
1.1.4 Challenges.....	14
1.2 Model System	16
1.2.1 Crystal Structures of HIV-1 Protease	22
1.2.2 Dynamics of HIV-1 PR: insight from experimental and MD simulations	27
1.2.3 Current HIV-1 PR Inhibitors	33
1.2.4 Alternative Strategies.....	39
1.3 Overview of My Research Projects	44
1.3.1 Exploring Rearrangements between the Closed and Semi-open Conformations of HIV-1 PR by MD simulations.....	44
1.3.2 Investigating the Gating Dynamics of the Flaps	45

1.3.3	Solution Structure of HIV-1 Protease Flaps Probed by Comparison of Molecular Dynamics Simulation Ensembles and Electron Paramagnetic Resonance (EPR) Spectra	46
1.3.4	Exploring Drug Resistance Mechanism by the Combination of MD Simulations and EPR Spectroscopy.....	47
Chapter 2 Exploring Rearrangements between the Closed and Semi-open Conformations of apo HIV-1 PR by MD simulations		
49		
2.1	Introduction.....	50
2.2	Methodology and Model Systems	53
2.2.1	Initial Preparation	53
2.2.2	Minimization and Equilibration Protocols.....	55
2.2.3	Production Runs	56
2.2.4	Data Analysis.....	57
2.3	Results and Discussion	57
2.3.1	Structural Differences in the Closed and Semi-open Conformations.....	57
2.3.2	Temperature Dependence of Atomic Fluctuations	61
2.3.3	Flap Conformations Observed During the Trajectories.....	63
2.3.4	Local Dynamics of the Flap Tips.....	68
2.4	Conclusions	74
2.4.1	Twisting of the Flap Tips Initiating the Rearrangements of the Flaps	74
2.4.2	Swapping Ile50 between the Hydrophobic Clusters.....	81
2.4.3	Significance of the Flap Tip Ile50 Residue in the Transition Dynamics..	85
2.5	Conclusions	89

Chapter 3	91
Microsecond Timescale MD simulation Suggests that Partial Dimer Dissociation is the Flap Opening Mechanism of HIV-1 PR	91
3.1 Introduction.....	92
3.2 Methods	97
3.2.1 Molecular Dynamics Simulations of apo HIV-1 PR	97
3.2.2 Data Analysis.....	97
3.2.3 Free Energy Estimation Using the MMPBSA/GBSA Approaches	97
3.3 Results and Discussion	100
3.3.1 Extent of Flap Motions Possible on the Microsecond Timescale.....	100
3.3.2 Partial Dissociation of the Dimer upon Flap Opening.....	104
3.3.3 Dimer Dissociation upon Flap Opening	108
3.3.4 Binding Free Energy Calculation on Dimer Stability of the HIV-PR	111
3.3.5 Per-residue Free Energy Decomposition	114
3.3.6 Solvation of the Dimer Interface	121
3.3.7 Further Theoretical and Experimental Evidence	125
3.3.8 Biological Implication: A Potential Allosteric Site	129
3.4 Conclusions	130
Chapter 4	132
Solution Structure of HIV-1 PR Flaps Probed by Comparison of Molecular Dynamics Simulation Ensembles and Electron Paramagnetic Resonance (EPR) Spectra	132
4.1 Introduction.....	133

4.2	Methods	138
4.2.1	Construction of the Modeled Structures	138
4.2.2	Minimization and Equilibration.....	139
4.2.3	Production Runs	141
4.2.4	Data Analysis.....	142
4.3	Results and Discussion	143
4.3.1	Internitroxide Distances from MD Simulations and the EPR Parameters 143	
4.4	Conclusions	149
Chapter 5		151
Drug Pressure Selected Mutations in HIV-1 PR Alter Flap Conformations as well as its Dynamics		151
5.1	Introduction.....	152
5.2	Methods	160
5.2.1	Construction of the Modeled Structures	160
5.2.2	Minimization and Equilibration.....	160
5.2.3	Production Runs	162
5.2.4	Data Analysis.....	163
5.3	Results and Discussion	163
5.3.1	Overall Structural Variation	163
5.3.2	Comparison of the Interspin Distances from EPR and MD Simulations	165
5.3.3	Comparison of Average Conformations.....	171
5.4	Conclusions	174

Chapter 6	176
Concluding Remarks	176
References.....	181
Appendix 1- Parameters of Ritonavir.....	208
Appendix 2- Structure and the AMBER force field parameters of the phosphorothioate-substituted nitroxide spin label	213

List of Figures

- Figure 1-1. Schematic illustrates the energy fluctuations for simulations at two temperatures for neighboring replicas. In order to obtain high exchange probabilities, the energy fluctuations δE_n in each simulation should close to the mean energy difference ΔE 12
- Figure 1-2. Global estimates of HIV infection from the World Health Organization (WHO) and the Joint United Nations Programme on HIV/AIDS (UNAIDS). 17
- Figure 1-3. Schematic representation of an HIV virion (<http://commons.wikimedia.org>) 18
- Figure 1-4. Genetic organization of HIV-1 and cleavage sites of HIV PR at Gag and Gag-pol polyproteins. Some of accessory proteins are omitted for clarity. MA for matrix, CA for capsid, NC for nucleocapsid, TF for transframe, RT for reverse transcriptase, RH for RNase H, IN for integrase, p1 and p2 are spacer peptides. p6 is peptide at 3' region of the Gag precursor and negatively regulates the PR activity. SU stands for surface unit glycoprotein and TM for transmembrane envelope..... 20
- Figure 1-5. The life cycle of HIV includes fusion, reverse transcription, integration, and assembly. (<http://en.wikipedia.org/wiki/HIV>)..... 21
- Figure 1-6. A structure of a homodimer of an apo HIV-PR (PDB code 1HHP). The flap regions are highlighted in yellow; the active site loops in red; N- and C- termini in ice blue; the single helix in each monomer in mauve; the turn (residues 5'-9') in

orange; two aspartic acids (Asp25) are rendered as van der Waals spheres and colored in red.	23
Figure 1-7. A detailed view of the 4-stranded antiparallel β -sheet at the N- and C-termini region of HIV-1PR.	24
Figure 1-8. A detailed view of the active site region (Leu24-Gly27) of HIV-1 PR forming the ‘fireman’s grip’. Note the network of hydrogen bonds between Thr26, Thr26’ and Leu24, Leu24’	25
Figure 1-9. A detailed view of the dimer interface region encompassing the helix (residues 86-94) and the β -turn (residues 5’-9’) stabilized by both intra- and inter-monomer salt bridges as well as an intra-monomer hydrogen bond between Arg87 and Leu5’	26
Figure 1-10. Cartoon draws of three distinct X-ray structures of HIV-1 protease: a) substrate-bound closed conformation (PDB code 1TSU), b) unbound semi-open conformation (PDB code 1HHP), and c) unbound wide-open state (PDB code 1TW7). Using the same color scheme as in Figure1-6 (The flap tips are highlighted in violet). Top views illustrate the distance between the flaps and the reversal of flap handedness in the three conformations.....	28
Figure 1-11. Schematic representation of simulated transitions between the three protease forms ¹²⁶ . The closed flap conformation converts to semi-open upon removal of ligand. Ligand induces the closure of the open form. Free protease exists in an ensemble of different conformers, closed, semi-open and open forms, which are in dynamic equilibrium.....	33
Figure 1-12. Chemical structures of 10 FDA approved HIV PR inhibitors.	34

Figure 2-1. a) The closed conformation (PDB code: 1TSU, with the substrate stripped out). b) The semi-open flap conformation (PDB code: 1HHP). The flap residues (44-55) are rendered by ‘Licorice’ in VMD. The flap tip residue Ile50 is rendered as ‘VDW’ and colored in orange in monomer A, in purple in monomer B. Asn25/25’ at the active site are rendered by licorice and colored in red. For clarity, only the side chains of hydrophobic residues on the loop region of each monomer are shown, including Val32/32’, Ile47/47’, Ile54/54’, Val56/56’, Pro79/79’, Pro81/81’ and Val82/82’, and rendered as both ‘Licorice’ and ‘Surf’ (residues on monomer A are colored in orange; residues on monomer B are colored in purple). It is worth noting that the relative orientation of the two flaps (the handednesses) is switched in the bound and unbound form..... 59

Figure 2-2. A detailed view of the ‘semi-open’ (PDB code 1HHP) crystal packing interactions around the flap region. The backbones of Ile50 and Lys54 (colored by atom type) on the flap of the central dimer hydrogen bond with the side chains of residues Gln61 and Gln92 (colored by atom type) from a neighboring dimer. 61

Figure 2-3. Comparison of the atomic fluctuations from simulations at 300K (purple line) and 375K (green line). Error bars reflect the difference between the two runs at the same temperature, starting from different conformations, i.e., closed and semi-open forms. 62

Figure 2-4. a) Time evolution of C α RMSD of the flap region. b) Time evolution of C α RMSD of the non-flap region with respect to the semi-open crystal structure. ... 64

Figure 2-5. Time evolution of C α RMSD of the flap region (a) and non-flap regions during the MDsemi-open simulation. the flap RMSD are colored using the same color code as in Figure 2-4a..... 65

Figure 2-6. The PMF profiles from the explicit solvent (black line) and the implicit solvent (red line) simulations, as a function of Flap C α RMSD with respect to the closed X-ray structure. Both simulations are started from the closed form (PDB code: 1TSU, without the substrate)..... 67

Figure 2-7. Time evolution of the intraflap hydrogen bonds within flapA (top) and flapB in the MDclosed simulation. Most hydrogen bonds are stable, and thus result in a stable β -hairpin structure with the exception of the tips (residues 49-52). Running averages over 100 data points (red line) are also shown. 69

Figure 2-8. Time evolution of the intraflap hydrogen bonds within flapA (top) and flapB in the MDsemi-open simulation. Consistent with MDclosed simulation, most hydrogen bonds are stable except the tip hydrogen bonds between Gly49 and Gly52. Running averages over 100 data points (red line) are also shown. 70

Figure 2-9. Backbone dihedral angles (Φ , green; Ψ purple) of the flap tip residues (G48-G52) during the MDclosed simulation 72

Figure 2-10. Backbone dihedral angles (Φ , green; Ψ purple) of the flap tip residues (G48-G52) during the MDsemiopen simulation. 73

Figure 2-11. Time evolution of flap C α RMSD with respect to the closed (blue) and semi-open (red) references (top); the torsion angles (Φ , green; Ψ purple) of Gly49/49' and Ile50/50' during the first (left) and second (right) transitional periods, respectively in MDclosed simulation..... 75

Figure 2-12. Time evolution of flap C α RMSD (top) with respect to the closed (blue) and semi-open (red) references; the torsion angles (Φ , green; Ψ purple) of Gly49'/49 and Ile50'/50 during the first (left) and second (right) transitional periods in MDsemi simulation.	76
Figure 2-13. Ramachandran plots of residues Gly48, Gly49 and Ile50 on both flaps in MDclosed simulation.....	79
Figure 2-14. Ramachandran plots of residues Gly48, Gly49 and Ile50 on both flaps in MDsemi-open simulation	80
Figure 2-15. Time evolution of flap RMSD (top), distance between the COM of residues 48-52 relative to Asn25/25' (middle) and Rg of intra- and inter- monomer hydrophobic clusters (bottom) formed between Ile50 and hydrophobic residues Val32, Ile47, Ile54, Val56, Pro79, Pro81 and Val82 from the same or its symmetry-related monomer in the course of the first transition in the free HIV-PR simulation started from a closed crystal structure. Representative snapshots during this first transition are shown on top. The backbone of HIV-1 PR is shown in NewCartoon representation with flapA colored in orange and flapB in purple. Residues Val32, Ile50, Ile47, Ile54, Val56, Ile79, Pro81 and V82 are shown in Surface and colored in orange in monomer A and colored in purple in monomer B.....	82
Figure 2-16. Time evolution of flap RMSD (top), distance between the COM of residues 48-52 relative to Asn25/25' (middle) and Rg of intra- and inter- monomer hydrophobic clusters in both monomers (bottom) in the course of the second	

transition in the free HIV-PR simulation started from a closed crystal structure.
 Representative snapshots during this first transition are shown on top..... 85

Figure 2-17. Time evolution of (a) the flap C α RMSD with respect to the two X-ray structures; (b) flap tips distance and (c) flap-Asn25/25' distances throughout the entire simulation of the I50A mutant system..... 87

Figure 2-18. For the I50W mutant system, time evolution of (a) flap C α RMSD with respect to the two X-ray structures, closed and semi-open (b) flap tips distance and (c) flap-Asn25/25' distances..... 87

Figure 2-19. A snapshot of the 'closed' conformation sampled in the I50W mutant system (rendered by New Cartoon in VMD⁸⁵ and flapA colored in orange, flapB in purple) overlapped on the closed crystal structure (gray, transparent). Trp50 is rendered by VdW and colored in yellow. For clarity, only the sidechains of hydrophobic residues on the flap and loop region from each mutant monomer are shown, including Val32/32', Ile47/47', Ile54/54', Val56/56', Pro79/79', Pro81/81' and Val82/82', and rendered by surface (residues on monomer A are colored in orange; residues on monomer B are colored in purple)..... 88

Figure 3-1. A detailed view of the free HIV-1 PR with the semi-open conformation (PDB code 1HHP¹⁹⁵). The residues are colored according to their conservation score computed by the ConSurf²¹⁹ program. The figure was generated with PyMol¹... 93

Figure 3-2. Time evolution of flap C α RMSDs (a), I50C α - I50'C α distances (b) and flapA- and flapB-Asn25/25' distances (c) during the MDclosed simulation. 101

Figure 3-3. Snapshots of HIV-1 PR over the course of the full flap opening at 0.5ns intervals in the simulation starting from the closed conformation (MDclosed).

Monomer A is colored in blue; monomer B in green. The active site Asn25/25' residues are highlighted in red; the flap tip Ile50 is highlighted in yellow. 103

Figure 3-4. Time evolution of all C α RMSD of each monomer (blue line, monomer A; green line, monomer B) when fitting to the corresponding monomer in the semi-open reference (top panel); when only fitting its symmetry-related monomer to the corresponding monomer in the semi-open reference (bottom panel). 106

Figure 3-5. A detailed view of the superimposition of the semi-open X-ray structure (gray, transparent cartoon) and a fully-open snapshot (monomer A in blue and monomer B in green) from the MD simulation based on the best fit using one monomer (blue). 106

Figure 3-6. Distance between the core domains (residues 32, 75-76 and 57-58) of the two monomers as a function of flap C α RMSD with respect to the fully open configuration captured in the simulation (gray cartoon). The C α atoms of the core domain residues represented as green spheres in the open reference structure. Note that distances lower than 28 Å correspond to the curled/tucked conformations. 108

Figure 3-7. a) An open snapshot from the MD simulation is shown to illustrate the definition of the open angle; a triangle between the two vectors, which are formed by the COM of the C α atoms of residues 32, 75-76 and 57-58, located in the core domain of each monomer, and the COM of residues 2-3, 96-97 from the N-, C-termini of both monomers, respectively (solid black lines). Carbon atoms of these residues are represented as red spheres. b) Contour plot of the free energy as a

function of the open angle and the flap C α RMSD with respect to the fully open configuration captured in the MD simulation.....	110
Figure 3-8. Binding free energies (ΔG) of the dimer as a function of the core domain distance.	112
Figure 3-9. Time evolution of the calculated binding energy (ΔG) for 10,000 snapshot structures, divided into nonpolar and polar energetic contributions: intermolecular vdW (E_{vdW}), intermolecular electrostatic (E_{ele}) plus change in reaction field (E_{RF}), and change in nonpolar solvation energy (E_{cav}) which is proportional to change in solvent accessible surface area (SASA) . Energies are in kcal/mol.	113
Figure 3-10. Intermonomer vdW interaction energies of the residues from monomer A (blue) and monomer B (green) averaged over the 10,000 snapshots from the 1 μ s simulation.	115
Figure 3-11. a) A semi-open structure colored by the magnitude of the correlation coefficient between the intermonomer vdW interaction energy of the corresponding residues with the total interaction vdW interaction energy of the dimer. Red and orange indicate negative correlations; blue and green indicate positive correlations. Actual values of the correlation coefficient is shown in (b) for residues from monomer A, (c) from monomer B.	118
Figure 3-12. a) A semi-open structure colored by the magnitude of the change in the intermonomer vdW interaction energy of the corresponding residues upon flap opening using RWG color code was used (red to green corresponding to -2.5 to 2 kcal/mol). The actual values of these changes are shown in (b) for residues from monomer A, (c) from monomer B.....	120

Figure 3-13. a) A semi-open structure colored by the magnitude of the change in the average SASA of the corresponding residues upon flap opening: blue indicates an increase in residue SASA ($\Delta\text{SASA} > 2 \text{ \AA}^2$); red indicates a decrease in residue SASA ($\Delta\text{SASA} < -2 \text{ \AA}^2$); and gray indicates no significant changes in SASA ($-2 \text{ \AA}^2 < \Delta\text{SASA} < 2 \text{ \AA}^2$). The exact values of the changes in the average SASA of residues on monomer A is shown in b, and monomer B in c. 122

Figure 3-14. Top view of the active site in a fully-open structure (monomer A colored in blue; monomer B in green) from the simulation. The Figure was drawn with VMD.²⁴¹ 123

Figure 3-15. A novel potential allosteric site identified in the fully-open form of HIV-1 PR captured in the MD simulation, defined by residues 4'-9' on the β -turn from one monomer (green), residues 22-29 on the active site loop and residues 87-91 on the α -helix from the other monomer (blue). Also shown are water molecules that enter this dimer interface region during the opening event. 124

Figure 3-16. Time evolution of a) flap C α RMSDs with respect to the three references, the closed, semi-open crystal structures, and a fully open configuration obtained from the previous MD simulation on the wild-type with the ITSU sequence; and b) core domain distances during the PR_{D29A/R87K} simulation starting from the closed state..... 127

Figure 3-17. Time evolution of a) flap C α RMSDs and b) core domain distances in the PR_{D29A/R87K} simulation starting from the semi-open state. 128

Figure 4-1. Three conformations of HIV PR during all-atom MD with EPR spin labels: a) closed Ritonavir bound; b) semi-open unbound; and c) fully open unbound. Top

views illustrate the reversal of handedness between the closed and semi-open form and the separation of open flaps.....	134
Figure 4-2. Structure of a nitroxide spin-label side chain with the distance from the C β atom to the spin label indicated.	137
Figure 4-3. MTSL spin label distances in the Ritonavir bound complex from EPR experiment and MD simulations, respectively.	144
Figure 4-4. MTSL spin label distances in the unbound LAI' from EPR experiment and MD simulations starting from the unbound closed form.....	145
Figure 4-5. The H-bonding pattern within the flap region in unbound closed simulations.	146
Figure 4-6. MTSL spin label distances in the unbound LAI' from EPR experiment and MD simulations starting from the unbound semi-open form.	147
Figure 4-7. Flap RMSD as a function of the interspin distance during the MD simulations starting from the semi-open conformation, compared with the closed, the semi-open crystal structures. While the structures are sampled that are similar to either the closed (1HXW), semi-open (1HHP) or "wide-open" (1TW7 ¹⁸²) structures, the simulated conformations with long label distances (> 40Å) match none of these crystal structures.	148
Figure 5-1. Structure of a dimeric HIV-1 protease with the C α positions of most common primary and secondary drug-resistance mutations shown as spheres. Primary mutations (30, 46, 48, 50, 82, 84, 88 and 90) are shown in red; secondary mutations (10, 20, 24, 32, 33, 36, 47, 53, 54, 63, 71, 73, 77 and 93) in blue. Both types of mutations are labeled in one of the subunits.	154

Figure 5-2. A Summary of sequence analysis showing the percentage of mutation prevalence of naturally occurring polymorphisms in HIV-1 PR subtype B for protease inhibitor (PI) naïve and exposed patient. Prevalence is defined as a measure of variability for each protease sequence vs. the subtype consensus (e.g. 0% prevalence corresponds to a residue that is conserved in all sequences for a particular subtype). 156

Figure 5-3. Ribbon diagrams of HIV-1PR in the semi-open conformation (1HHP) with the nitroxide spin probe, MTSL, appended at site K55C. Colored spheres represent the C α position of mutations relative to LAI' in MDR769' (top) and V6' (bottom) in the active site cavity, and the nonactive site region, and flaps/elbows are shown in red, blue, and green; respectively. Diagrams were rendered with VMD..... 159

Figure 5-4. Protein backbone RMSD with respect to semi-open crystal structure (2G69) throughout the three simulations. 164

Figure 5-5. Atomic fluctuations simulated for the three proteases, LAI' (black), MDR769' (red) and V6' (blue). Error bars reflect the difference between the two monomers 165

Figure 5-6. Normalized and background subtracted time domain dipolar evolution for the three constructs. The solid smooth lines represent the respective best solutions obtained from Tikhonov Regularization²⁷² for HIV-1PR samples labeled at site K55C with MTSL..... 166

Figure 5-7. Interspin distance distribution profiles from TKR of EPR data (top) and 55N–55'N distance profiles from MD simulations (bottom)..... 169

Figure 5-8. Distributions of distances between the carbon atoms (top) and the nitroxide nitrogen atoms (bottom) in the residues K55C and K55'C sampled during the MD simulations of LAI (black), MDR769 (red) and V6 (blue). 171

Figure 5-9. Comparison of averaged structures sampled during MD simulations of the three sequences. For clarity, only the backbone ribbon and Cys-MTSL side chain are shown. KEY: LAI "wild-type" (black); V6' (blue); MDR769' (red)..... 172

Figure 5-10. Histogram of distances sampled between the COM of 5 central residues on each flap (residues 48-52) and the COM of the two Asn25/25' residues sampled during the MD simulations of LAI (black), MDR769 (red) and V6 (blue)..... 173

List of Tables

Table 4-1. Timescale of All Spin-labeled Simulations.....	142
Table 5-1 Mutations in each mutant construct relative to the wild-type (LAI')	158
Table 5-2. Summary of distance parameters obtained from DEER distance profiles and MD simulations	168

Acknowledgements

I would like to express my sincere thanks to Professor Carlos Simmerling for offering me the greatest opportunity of my life by letting me to pursue my graduate studies in his laboratory. He suggested me to take on challenging projects, letting me explore my passion and work on that I love. Along this freedom, he has given me an incredible amount of support and directed me to the paths of thought that have taken my research to new levels in new directions. Out of all the people who have mentored me in my life, Professor Simmerling is one of the few who has changed my life for the better.

Sincere thanks also go to Professor Jin Wang and Professor Robert C. Rizzo for serving as members on my committee. I thank them for their candid, constructive criticism and valuable suggestions over the years. I would also like to thank Professor David F. Green and my outside member, Dr. Carol A. Carter, who gave me a lot of insightful comments and suggestions to improve my thesis work. I would also like to thank Gail E. Fanucci's group for fruitful collaborations, providing me with their experiment data for comparison, and giving me lots of valuable advice in my manuscripts.

I would also like to acknowledge the past and present members of the Simmerling lab. This lab has been like a second family over the years to me. Everyone has been part of my life in so many aspects. I would like to thank Melinda Layten for helping me immensely at the beginning of my graduate study, patiently answering my silly questions, and offering suggestions for new directions of my research. Special thanks go to Dr. Lauren Wickstrom for always being there when I needed advice. Thanks to Dr. Lin Fu for providing me the codes and programs. My deep appreciation goes to Arthur J.

Campbell and Amber Carr, for being greatest friends in this world. Thank you for opening your hearts, and for being there every step of the way. I love you guys! Your friendship means everything to me.

I would like to thank all my dear friends at Stony Brook, for always encouraging me and keeping my spirits up. Their friendships have helped me through many a rough time. I will never forget these great times we spent together.

Finally, there are not enough words to express my gratitude for the love and encouragement I have received from my family, especially my mother and my husband. Without your support, I could not have made it so far in my life.

Chapter 1

Introduction

1.1 Simulation Methodology

Proteins are dynamic polymers, and in most cases their functions are associated with their conformational flexibility. In structural biology, there are two commonly used tools for determining 3-dimensional structures of a biological macromolecule: X-ray crystallography and NMR Spectroscopy. Both experimental tools have provided invaluable structural information for proteins. These methods, however, usually provide only snapshots of the native state or time-averaged data. Moreover, studying protein dynamics through these experiments is not trivial.

To overcome these limitations, modern theoretical methods, such as molecular dynamics simulations, have been used to supplement experimental techniques. MD simulations can provide insight into structural as well as dynamic features of biomolecular systems at spatial and temporal scales that are difficult to access by experimental tools. MD simulations have several advantages when compared with conventional experimental methods. Firstly, unlike many experimental techniques, which only yield time-averaged results, MD simulations capture biological events occurring on timescales spanning 12 orders of magnitude, from femtosecond to millisecond. Secondly, MD simulations provide an approach to directly observe biological events at atomic

resolution. The time-sequence trajectories generated by MD simulations can be visualized by graphics software, such as PyMOL¹ and VMD². Thirdly, MD simulations can yield thermodynamics describing the driving force for biological events, such as protein folding³⁻⁵ and conformational changes.⁶

1.1.1 Force Fields

In molecular mechanics, a force field is the mathematical description of the potential energy of a system. Force field functions and parameter sets are derived from both experimental and high-level quantum mechanical calculations. "All-atom" force fields provide parameters for every atom in a system, including hydrogen, while "united-atom" force fields treat the hydrogens and carbon atoms in methyl and methylene groups as a single center. "Coarse-grained" force fields, which are frequently used in long-time simulations of proteins, provide even more abstract representations for increased computational efficiency.

Commonly used force fields include AMBER⁷, CHARMM,⁸ GROMOS⁹⁻¹⁰ and OPLS.¹¹ Although the specific decomposition of the terms depends on the force field, a general form for the total energy in an additive force field encapsulates both bonded terms relating to atoms that are linked by covalent bonds, and nonbonded (also called "noncovalent") terms describing the long-range electrostatic and van der Waals (VDW) forces.

As one of the most commonly used force fields, the AMBER ff99SB force field¹² has been used throughout my study. The potential energy function in the Amber force field is shown as Equation 1-1, incorporating both bonded and non-bonded terms.

$$\begin{aligned}
 U = & \sum_{\text{all bonds}} \frac{1}{2} K_b (b - b_0)^2 \\
 & + \sum_{\text{all angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \\
 & + \sum_{\text{all torsions}} K_\phi [1 - \cos(n\phi)] \\
 & + \sum_{i,j \text{ nonbonded}} \epsilon_{ij} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^6 \right] \\
 & + \sum_{i,j \text{ nonbonded}} \frac{q_i q_j}{4\pi\epsilon_0 \epsilon r_{ij}}
 \end{aligned}
 \tag{1-1}$$

The bonded terms apply to sets of two to four atoms that are covalently linked, and they serve to constrain bond lengths and angles near their equilibrium values, and also include the torsional potential that represents the energy for rotating a bond due to bond order and neighboring bonds or lone pairs of electrons. The non-bonded terms consist of Lennard-Jones (LJ) function for van der Waals interactions and Coulomb's law for electrostatic interactions. The parameters for these bonded and non-bonded terms are derived from quantum mechanical calculations and from thermodynamic, crystallographic and spectroscopic data on a wide range of systems.¹³ Usually the most important differences among various force fields are the treatments of electrostatics.¹⁴

It is worth noting that one of the most important qualities of a force field is that its functional form and parameters must be transferable. This means that the same set of parameters can be used to model a series of related molecules. However, careful

evaluation of the accuracy of the force field on the subject before further analyzing simulation results is always critical.

1.1.2 Solvation Effects

The critical role that solvent plays in dictating the stability, flexibility, and interactions of molecules necessitates a good description of solvent effects in theoretical approaches to chemical and biomolecular problems. There are two different approaches to include the solvent effects in molecular dynamics. One is the explicit solvent model and the other is the continuum solvent model, also known as the implicit solvent model. Explicit solvent methods offer a more detailed and accurate description of a macromolecular system; these involve extensive searches in the configuration space of the solute and solvent. A wide range of water models have been proposed, and these models can be divided into three types: simple, rigid, flexible models, and polarizable models.¹⁵ In the simple interaction-site models, each water molecule is maintained in a rigid geometry, and the interactions between molecules are described using pair wise Columbic and Lennard-Jones expressions. The most popular simple water model is TIP3P¹⁶ (transferable intermolecular potential 3P), which uses a total three sites for the electrostatic interactions; the partial positive charges on the hydrogen atoms are exactly balanced by an appropriate negative charge on the oxygen atom. The van der Waals interaction between two water molecules is computed using LJ function with just a single interaction point per molecule, centered on the oxygen atom; no van der Waals interactions involving the hydrogen atoms are calculated. There are several other water

models, such as TIPS¹⁷ and SPC¹⁸⁻¹⁹ (simple point charge) which use a similar strategy. In complex water models, there are more charge sites on dummy atoms to represent a more realistic electron distribution around the oxygen atom; these include TIP4P,¹⁶ TIP4Pew,²⁰ TIPS2,²¹ TIP5P.²² There are also studies for using flexible water models²³ and including polarization effect.²⁴

The number of nonbonded interactions scales with the square of the number of interaction sites; due to the large numbers of solvent atoms involved in explicit solvent models, nonbonded interactions usually require by far the greatest amount of computational effort. Thus, explicit solvent models are computationally demanding. In addition, explicit solvent models need more time to average over many solvent configurations to obtain meaningful thermodynamic data. Algorithms such as Particle Mesh Ewald (PME),²⁵ particle-particle/particle-mesh Ewald (P3M)²⁶ are developed to accelerate the simulations. However, certain periodic boundary methods, such as PME, evoke artificial real-virtual solute interactions that can be also problematic when determining thermodynamic quantities such as free energies²⁷.

An alternative approach is using implicit solvation (sometimes known as continuum solvent), a method of representing solvent as a continuous medium instead of as individual explicit solvent molecules. A clear goal of the continuum solvent methods has been to eliminate nonessential degrees of freedom or capture them in some approximate manner, and increase the investigation in the space and time domains of the biological system without compromising its structural, dynamic, and thermodynamic features. In an implicit solvent model, solvent is treated as a homogeneous isotropic dielectric continuum, and ΔG_{solv} is typically decomposed into a cavity term (G_{cav}), a

solute-solvent van der Waals term (G_{vdW}), and a solute-solvent electrostatic polarization term (G_{pol}):

$$G_{sol} = G_{cav} + G_{vdW} + G_{pol} \quad \mathbf{1-2}$$

The nonpolar terms, G_{cav} and G_{vdW} , account for the energy expense to form a cavity in the solvent to accommodate solute and the van der Waals interaction energy of the solute with solvent respectively and are included by a surface tension/area or free energy density/volume type terms.

$$G_{cav} + G_{vdW} = \sum \sigma_k SA_k \quad \mathbf{1-3}$$

where SA_k is the total solvent-accessible surface area of atoms of type k and σ_k is an empirical atomic solvation parameter.

For ΔG_{pol} , Poisson-Boltzmann (PB) and Generalized Born (GB) models can be used to describe the electrostatic interactions between any two charged sites. The PB equation appears to be the most accurate model for describing molecules in water,^{23-24, 28} where contributions from solvent polarization along with the asymmetric shapes of biological molecules are taken into account; however, it is computationally expensive to calculate.²⁸⁻³¹ The Born equation provides an analytical solution to the linear Poisson-Boltzmann equation for a spherical ion with a point charge in its center, while the GB is an approximation to account for the shape of the molecule.³² The GB model has the following functional form:

$$\Delta G_{pol} = -\frac{1}{2} \left(1 - \frac{1}{\epsilon_w}\right) \sum_{i,j} \frac{q_i q_j}{f_{GB}} \quad \mathbf{1-4}$$

where ϵ is dielectric constant, q is charge, and f_{GB} is a function of distance and radii. f_{GB} is commonly calculated using the formula:

$$f_{GB} = \sqrt{r_{ij}^2 + R_i R_j} \exp\left(\frac{-r_{ij}^2}{4R_i R_j}\right) \quad 1-5$$

were r_{ij} is the distance between atoms i and j , and R_i and R_j are the effective Born radii of atoms i and j .³² The efficiency of R_i is a critical issue, with different setting of R_i , derived several GB models, such as GB-HCT³³⁻³⁵, GB-OBC³⁶ and GB-neck.³⁷

Not including solvent atoms considerably reduces the size of a system, resulting in a significant decrease in the computational cost of a simulation. In addition, conformational sampling in an implicit solvent model is enhanced in two ways: 1) there is no need to equilibrate and average over the great number of solvent configurations in a simulation; 2) the low viscosity that water molecules impart by randomly colliding and impeding the motion of solutes through their van der Waals repulsion accelerates molecular motions.³⁸ As a result, implicit solvent simulations are typically faster and easier to interpret as the water degrees of freedom are absent. Implicit solvent models, however, are of lower resolution and have been known to blur the potential energy landscape of a protein, cause structural distortions, and over-stabilize salt bridges.^{3, 39-44} Moreover, although all implicit solvation models allow estimation of the mean electrostatic free energy, they do not account for the hydrophobic effect, a major factor in the folding process of globular proteins with hydrophobic cores. To augment this deficiency, the solvent accessible surface area (SASA) is taken into account as a proxy for the extent of the hydrophobic effect. Yet, this surface area pertains to the solute, while the hydrophobic effect is mostly entropic in nature and occurs on the side of the solvent.

To sum up, both explicit and implicit solvent models have certain strengths and weaknesses. Calculations using explicit solvent generally yield more accurate results. In

some cases, such as simulations involving water bridges, explicit water molecules are essential for the calculation.⁴⁵ However, systems using explicit solvent have many more atoms, demanding additional computing resources. Due to the friction force from the water molecules, the dynamics of the solute is also slower in explicit solvent. This may be useful for understanding the real timescale of a biological event. In other cases where rate is not a factor, but the results of the motion are important, the implicit solvent model will be more efficient.

1.1.3 Advanced MD Simulation Methods

MD simulations are increasingly demonstrating their practical value in the investigation of biological systems. The potential energy surface of complex biomolecular systems, however, can be extremely rugged, and so the trajectory can be easily trapped sampling within a high-energy local minimum⁴⁶. Several methods have been developed to enhance sampling, such as umbrella sampling method, which applies a biasing ‘umbrella’ potential to force sampling along a particular reaction coordinate.⁴⁷⁻⁴⁹ An unbiased free energy profile, also known as a potential of mean force (PMF), can be extracted by using a post-processing algorithm called weighted-histogram analysis method (WHAM).⁵⁰⁻⁵¹

Equation 1-6 shows the general formula for calculating the free energy difference between two states, where, $P(q_A)$ is the probability to find the system in state A at reaction coordinate q .

$$F(q_B) - F(q_A) = -kT \ln \frac{P(q_B)}{P(q_A)} \quad \mathbf{1-6}$$

where k_B is the Boltzmann constant and T is the temperature. How often the system samples a certain value of the reaction coordinate q can be analyzed via umbrella sampling with a biased potential $V(q)$. The unbiased potential of mean force (PMF) for the i th window then is:

$$F^{(i)}(q_B) = F(q_A) - kT \ln \frac{P(q_B)}{P(q_A)} - V^{(i)}(q_B) + K^{(i)} \quad \mathbf{1-7}$$

The underestimated constant $K^{(i)}$, called the free energy constants, are defined as:

$$e^{\frac{K^{(i)}}{kT}} = \langle e^{-\frac{V^{(i)}(q_B)}{kT}} \rangle \quad \mathbf{1-8}$$

where $K^{(i)}$ can be efficiently determined by WHAM analysis. The WHAM equation (Equation 1-9) expresses the optimal estimate for the unbiased distribution function as a reaction coordinate (q)-dependent weighted sum over the N_w (the number of biased window simulations) individual unbiased distribution functions:

$$P(q) = \sum_{i=1}^{N_w} n_i P(q)_{(i)} \times \left[\sum_{j=1}^{N_w} n_j e^{-\frac{[V^{(j)}(q) - K^{(j)}]}{kT}} \right]^{-1} \quad \mathbf{1-9}$$

where n_i is the number of independent data points used to construct the biased distribution function. The free energy constants $K^{(i)}$ needed in Equation 1-9 are determined from Equation 1-8 using the optimal estimate for the distribution function.

To accurately construct the free energy profile of a given system, three key parameters need to be chosen carefully in umbrella sampling simulations. The first parameter is the reaction coordinate, which should be able to represent the physiological

properties of interest. The last two parameters are the size of each window and the force constant for the biasing potential. WHAM analysis needs the adjacent windows to have the same free energy value for the overlapping region. Insufficient overlap between neighboring windows will introduce statistical errors in each individual estimate. Moreover, a weak force constant will not be sufficient to generate an ensemble in a desired region, while if force constant is too strong force constant that will limit the range of sampling. Thus, the success of umbrella sampling simulations relies heavily on the choice of these three parameters.

Another successful enhanced sampling method is replica exchange molecular dynamics (REMD).⁵² The replica exchange method was developed first in the physics community to improve sampling in glassy systems⁵³⁻⁵⁴, and has been recently applied to MD simulations of biomolecules.⁵⁵⁻⁵⁷ In this method, a number of simulations are performed at different temperatures in parallel, and exchanges of configurations are attempted periodically. Even if a trajectory is temporarily trapped in a local minimum, the simulation can escape from this minimum via an exchange with a higher temperature configuration. With this method, one can obtain various thermodynamic quantities as a function of temperature for a wide temperature range from a single simulation run. Moreover, because each replica can be simulated using its own computer processor, the REMD method is well suited for and very efficiently runs on parallel computers, which have become ubiquitous in recent years.

In a standard REMD simulation, the simulated system consists of M non-interacting copies (replicas) at M different temperatures. The positions, momenta and temperature

for each replica are denoted by $\{q^{[i]}, p^{[i]}, T_m\}$, $i = 1, \dots, M$. The equilibrium probability for this generalized ensemble is

$$W(p^{[i]}, q^{[i]}, T_m) = \exp\left\{-\sum_{i=1}^M \frac{1}{k_B T_m} H(p^{[i]}, q^{[i]})\right\} \quad \mathbf{1-10}$$

where, the Hamiltonian H is the sum of kinetic energy, K , and potential energy, E . For convenience, momenta $q^{[i]} p^{[i]}$ at temperature T_m are denoted by $x_m^{[i]}$, and X is defined as one state of the generalized ensemble. For example, X and X' represent two states of replicas i and j , which are at temperatures T_m and T_n respectively. In order to maintain the balance of the generalized system, microscopic reversibility has to be satisfied, thus giving:

$$W(X)\rho(X \rightarrow X') = W(X')\rho(X' \rightarrow X) \quad \mathbf{1-11}$$

where $\rho(X \rightarrow X')$ is the exchange probability between two states X and X' . With the canonical ensemble, the potential energy, E , rather than total Hamiltonian, H , is used simply because the momentum can be integrated out. Inserting equation 3 into equation 4, the following equation for the Metropolis exchange probability is obtained:

$$\rho = \min\left(1, \exp\left\{\left(\frac{1}{k_B T_m} - \frac{1}{k_B T_n}\right)(E(q^{[i]}) - E(q^{[j]}))\right\}\right) \quad \mathbf{1-12}$$

In practice, several replicas at different temperatures are simulated simultaneously and independently for a chosen number of MD steps. Exchange between a pair of replicas is attempted with a probability of success calculated from equation 1-12. If the exchange is accepted, the bath temperatures of these replicas will be swapped, and the velocities will be scaled accordingly. Otherwise, if the exchange is rejected, each replica will continue on its current trajectory with the same thermostat temperature.

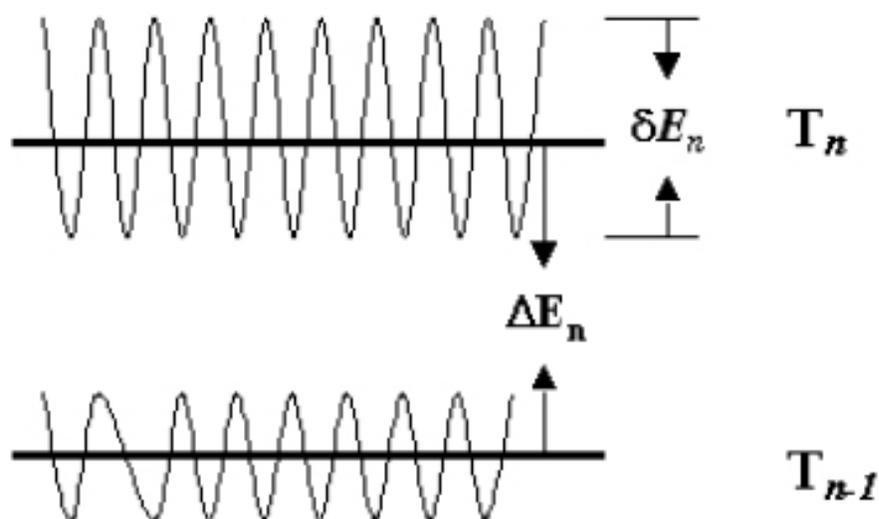


Figure 1-1. Schematic illustrates the energy fluctuations for simulations at two temperatures for neighboring replicas. In order to obtain high exchange probabilities, the energy fluctuations δE_n in each simulation should close to the mean energy difference ΔE .

In this manner, REMD is hampered to a lesser degree by the local minima problem, since simulations at low temperature can escape kinetic traps by “jumping” directly to alternate minima being sampled at higher temperatures. Likewise, the structures sampled at high temperatures can anneal by being transferred to successively lower temperatures. Moreover, the transition probability is constructed such that the canonical ensemble properties are maintained during each simulation, thus providing potentially useful information about conformational probabilities as a function of temperature.

Due to these advantages, REMD has been applied to studies of peptide and small protein folding.^{3, 58} To date, however, most of these studies have used REMD with continuum solvent models, since implicit solvent model reduces system size and replica requirements. REMD with explicit solvent model is greatly limited by the number of

replicas needed to span a temperature range, which grows proportionally to the square root of number of degrees of freedom in the simulated system. As explicit solvent simulations require a greater length of time to converge, a typical REMD simulation with explicit solvent on a large system is impractical, due to computational cost and slow conformational sampling. To our knowledge, converged REMD simulations in explicit solvent from independent starting conformations have been reported only for short helical or unstructured peptides.^{39, 43} Several promising techniques have been proposed to deal with this apparent disadvantage of REMD.

Furthermore, as mentioned earlier, MD simulations can yield detailed information on the thermodynamics and kinetics of dynamic processes that occur in biological systems, and thus provide a means to estimate free energy. The combination of molecular mechanics and the PB continuum solvent model to compute binding free energies was pioneered by Kollman and his co-workers,⁵⁹ termed MM/PBSA (Molecular Mechanics/Poisson–Boltzmann Surface Area). An MD simulation (typically in explicit solvent) is first carried out to yield a representative ensemble of structures; the average total free energy is then calculated based on existing snapshots from the MD simulations. After removing any solvent, the free energy of the system, G , is evaluated as:

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{PBSA} - TS_{MM} \quad \mathbf{1-13}$$

where E_{MM} is the internal energy and contains all intramolecular bonded (stretch, bend and torsion) and non-bonded (van der Waals and electrostatic) interactions, as shown in Equation 1-14:

$$\Delta E_{MM} = \Delta E_{bond} + \Delta E_{angle} + \Delta E_{tors} + \Delta E_{vdw} + \Delta E_{elec} \quad \mathbf{1-14}$$

G_{PBSA} denotes the solvation free energy, including both the polar and the non-polar contribution to the solvation free energy. The polar fraction is determined by solving the relatively sophisticated Poisson–Boltzmann equation^{23,60} or by applying the significantly cheaper GB model,⁶¹⁻⁶² while the non-polar part is calculated by an empirical formula based on the solvent accessible surface area (SASA). The last term in Equation 1-13, TS_{MM} , is the solute entropy and is usually estimated by a combination of classical statistical formulas⁶³ and normal mode analysis. In most cases, however, MM-PB/GBSA method⁵⁹ is used to estimate the relative energy between similar states, in which the entropic contribution will be canceled out.

The ability to accurately calculate ΔG for a given macromolecular system with various different conformations or structures presents a very important methodology to our computational arsenal. Expensive but rigorous free energy calculations by MM-PB/GBSA scoring have been applied extensively in the later stage of a lead optimization, facilitating the drug discovery process⁶⁴⁻⁶⁵.

1.1.4 Challenges

When planning and conducting MD simulations, there are three challenges that must be taken into account: force field, searching and sampling. A cornerstone for accurate simulations is the force field. One of the primary limitations of MD simulations is the deficiencies of the force fields, which arise from the assumptions of an additive force field, harmonic terms, and the use of fixed atomic charges that by their very nature cannot easily model the effects of electronic polarizability. In addition, the accuracy of the calculated energies may be limited by the common use of cut-offs for van der Waals

interactions and by approximate treatments of electrostatics. To improve the accuracy of classic force fields, such as CHARMM⁸ and AMBER⁶⁶ force fields, the ϕ/ψ dihedral terms in the potential energy functions have been modified,⁶⁷⁻⁶⁸ leading to improved agreement with experimental quantities. Moreover, one possibility is that the future development of force field may move beyond the current mechanics approach, by using quantum mechanics explicitly to construct the force field. To this end, a number of the “polarizable force fields” have been developed,⁶⁹⁻⁷¹ which is speculated to offer more consistent and balanced treatment of solute-solute and solute-solvent electrostatic components.

Another major challenge facing computational chemistry today is the timescales accessible by atomistic MD simulations. MD simulations must be discretized at the level of femtoseconds, which, when coupled with the computer time required to perform one step of a calculation, invariably limits the total length of the simulation. The first all-atom MD simulation of a small protein *in vacuo*, performed more than three decades ago, covered less than 10 ps⁷². Over the years, improvements in molecular dynamics algorithms, software, and computer hardware have allowed MD simulations to access longer timescales,⁷³ such that accurate all-atom simulations of more than a microsecond are now becoming practical. A number of studies with individual trajectories longer than one microsecond have been reported.⁷⁴⁻⁸⁰ The ability to efficiently perform simulations on the timescales over which many physiologically relevant processes take place expands substantially the set of problems for which the MD approach is tractable. In addition, a recent confluence of progress in both simulation and laboratory techniques has offered more opportunities to compare computational results to experimental data gathered on

similar timescales. These developments allow not only for a better understanding of the biomolecular systems of interest, but also for systematic validation of the models and methods underlying MD simulations.

In this work, MD simulations have been employed to study the dynamics of HIV-1 protease, one of the primary targets for anti-AIDS drug discovery. The introduction of HIV-1 PR and the significance of this study are presented in the following section.

1.2 Model System

It has been over two decades since HIV (Human Immunodeficiency Virus) was first characterized as the causative agent for AIDS (Acquired Immunodeficiency Deficiency Syndrome). Since then, the pathogenesis and treatment of AIDS has been extensively studied and great progress has been made. However, the pandemic of AIDS is still globally expanding, and the fight against this dreaded disease is long-lasting. According to the Joint Nations Programme on HIV/AIDS (UNAIDS) report, ~33.4 million people worldwide were living with HIV at the end of 2008, of which a half million were in the United States. In 2008 alone, 2.7 million individuals were newly infected with HIV, and 2 million people died of AIDS.

Adults and children estimated to be living with HIV, 2008

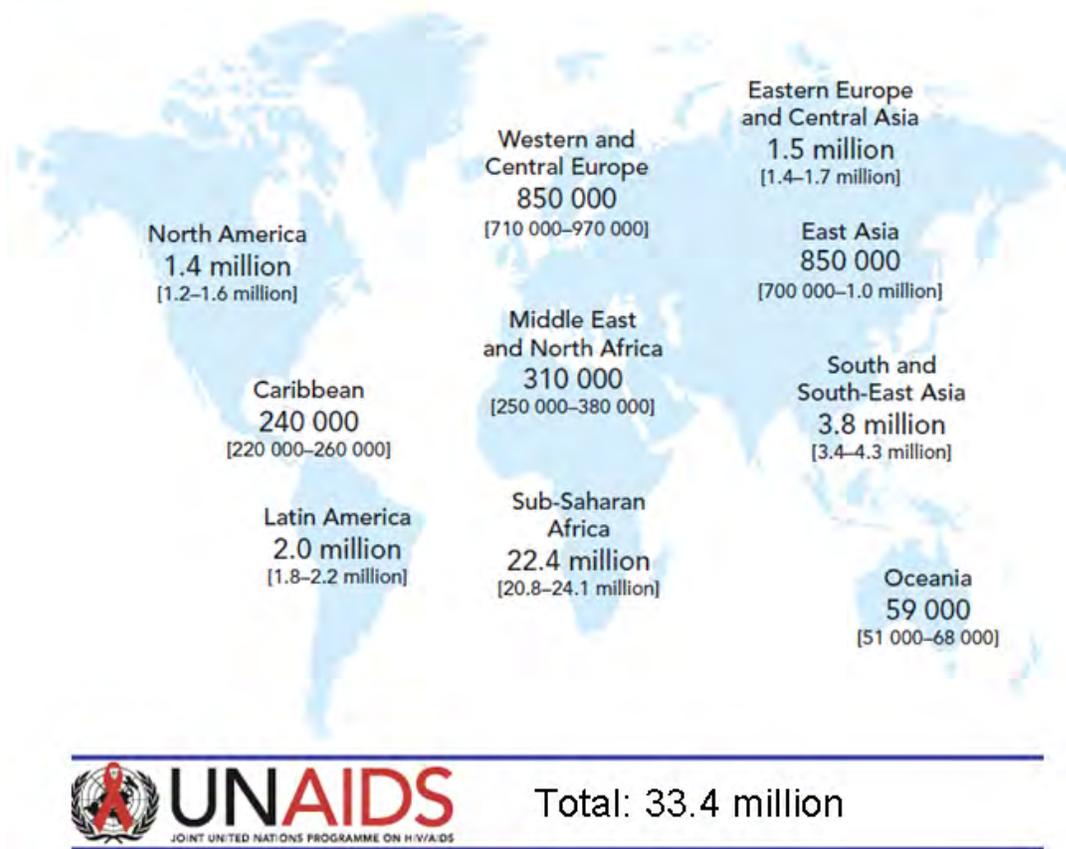


Figure 1-2. Global estimates of HIV infection from the World Health Organization (WHO) and the Joint United Nations Programme on HIV/AIDS (UNAIDS).

As HIV continues to spread around the world, an increasing amount of funding is being provided to fight the AIDS epidemic. Since 1996 funding for the response to AIDS in low- and middle-income countries rose from US\$300 million annually to US\$13.7 billion in 2008.⁸¹ Nevertheless, the budget is far below the need especially in the south of Africa, where has been hit the hardest by *AIDS* pandemic

AIDS is caused by Human Immuno-deficiency virus (HIV). The genome and proteins of HIV have been the subject of extensive studies since the discovery of the

virus in 1983.⁸² HIV is a retrovirus, containing Gag, Pol and Env genes as the basic components of a replicating genome. HIV is composed of two copies of single-stranded RNA (Figure 1-3).

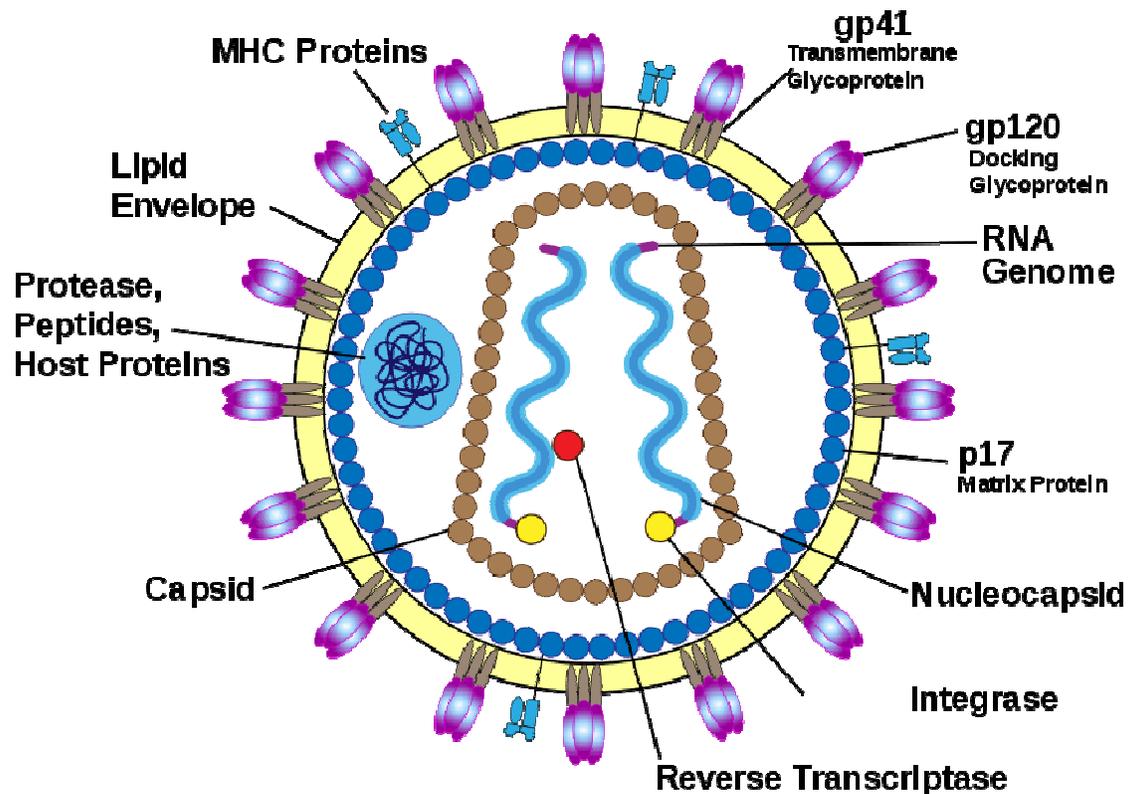


Figure 1-3. Schematic representation of an HIV virion (<http://commons.wikimedia.org>)

The viral RNA is tightly bound to nucleocapsid (NC) proteins, p7, and enzymes that are indispensable for the development of the virion, such as transcriptase and integrase. A matrix composed by the viral protein p17 surrounds the capsid (CA), ensuring the integrity of the virion particle. This is, in turn, surrounded by the viral envelope, which is formed when the capsid buds from the host cell. Embedded in the viral envelope are glycoproteins from the host cell, gp 120 and gp 41, which enable the virus to attach to

and fuse with target cells to initiate the infectious cycle. Also enclosed within the virion particle are Vif, Vpr, Nef and viral protease.

There are two types of HIV: HIV-1 and HIV-2, both were discovered jointly by Luc Montagnier and Robert Gallo and their associates⁸³⁻⁸⁵. Although both types are transmitted by sexual contact, through blood, and from mother to child, and both appear to cause clinically indistinguishable AIDS, HIV-2 is less easily transmitted. In addition, the period between initial infection and illness is longer in the case of HIV-2. Worldwide, the predominant virus is HIV-1; the relatively uncommon HIV-2 type is concentrated in West Africa and is rarely found elsewhere.

The infection of HIV begins with the recognition of viral envelope glycoprotein by the cell surface receptors CD4, a member of the immunoglobulin superfamily, and other coreceptors on the host cell. HIV-1 attaches to CD4 with its envelope glycoprotein gp120. The binding to CD4 induces conformational changes in gp120, allowing HIV-1 to bind to other surface coreceptors on the host cell. After the virus fuses with the host cell membrane, the genetic material (RNA) is released into the cytoplasm of the cell. The viral RNA is transcribed into DNA with the help of viral Reverse Transcriptase (RT); the DNA replicates into double strands, and then is integrated with the host cell's DNA with the assistance of viral Integrase. Thereafter, the viral genome is replicated with the host cell genome. The viral genome is translated into three primary polyproteins (Figure 1-4), Gag, Gag-pol and Env. Gag (group antigen) encodes internal structural components of the virion: matrix (MA), capsid (CA) and nucleocapsid (NC) proteins. Pol (polymerase) contains reverse transcriptase and integrase, two key enzymes in the viral replication. Another important enzyme is the protease (PR), which is located upstream of Pol in Gag-

pol polyprotein. Env encodes two exterior proteins, surface unit glycoprotein and transmembrane envelope. These exterior proteins recognize the surface receptors on the target cells in the early stage of infection. The components of the virus gather together (assemble) near the cell membrane and form by a ‘pinching’ action of the membrane. The new virus buds off from the host cell, and as it buds, maturation takes place where the long strands of protein precursors are cleaved into smaller functional fragments by HIV protease. When the uncleaved viral precursor polyproteins, viral RNA and other elements are packed into the viral particles and released from the infected cells; they are immature and noninfectious. The life cycle of HIV is depicted in Figure 1-5.

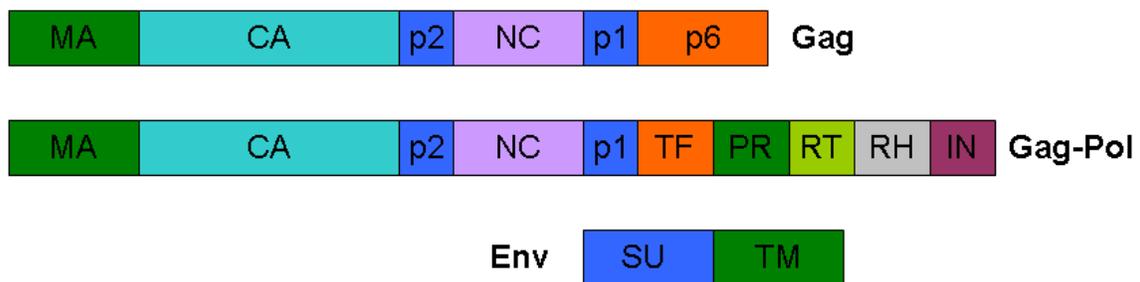


Figure 1-4. Genetic organization of HIV-1 and cleavage sites of HIV PR at Gag and Gag-pol polyproteins. Some of accessory proteins are omitted for clarity. MA for matrix, CA for capsid, NC for nucleocapsid, TF for transframe, RT for reverse transcriptase, RH for RNase H, IN for integrase, p1 and p2 are spacer peptides. p6 is peptide at 3' region of the Gag precursor and negatively regulates the PR activity. SU stands for surface unit glycoprotein and TM for transmembrane envelope.

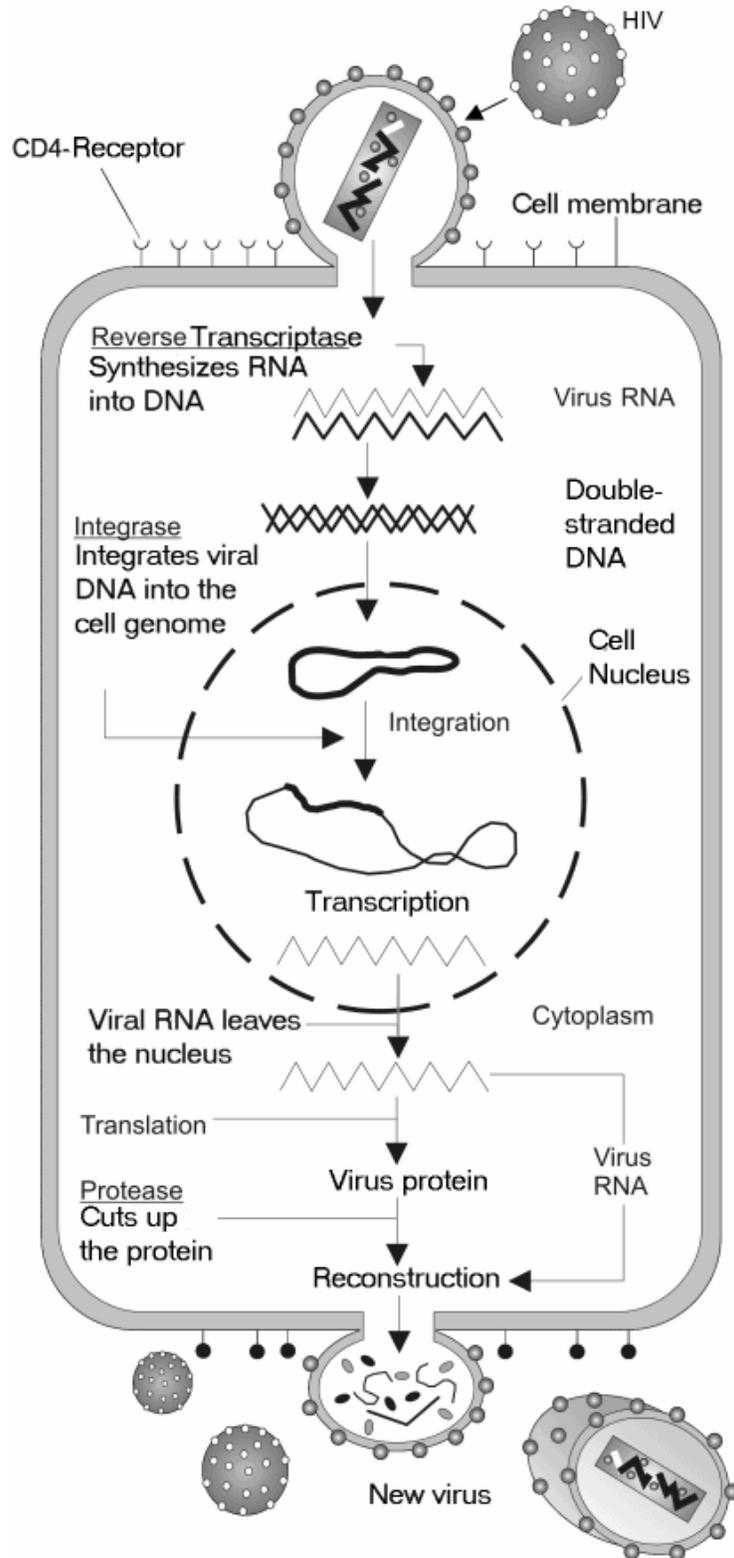


Figure 1-5. The life cycle of HIV includes fusion, reverse transcription, integration, and assembly. (<http://en.wikipedia.org/wiki/HIV>)

Since all the steps in the cycle are required for generation of new infectious virions, intervention can be conceived at each step involved in the life cycle of virus.⁸⁶ In particular, one of the most intense areas of research has been the effort to find effective inhibitors of the essential aspartic protease (PR), an endopeptidase that catalyzes the cleavage of Gag and Gag-pol polyproteins into mature proteins.⁸⁷ The active form of the PR is a homodimer with 99 amino acid residues in each subunit. For convenience, the residues in one subunit are numbered 1-99 and those from the other subunit are numbered 1'-99'. The two subunits form an active site cavity where the substrate binds and is hydrolyzed. When the PR is inactivated by an inhibitor or mutation of key residues, the cleavage of Gag and Gag-pol polyproteins is interrupted; as a result, the budding viral particles become noninfectious.⁸⁸ Therefore, PR has been identified as a major target for anti-AIDS drug discovery, owing to its indispensable role in viral replication and infection.⁸⁹

1.2.1 Crystal Structures of HIV-1 Protease

After recognizing the crucial role of PR in virus maturation, intensive efforts have been made to determine the three dimensional structures of PR. The first three dimensional (3D) structure of HIV- 1 PR was solved by X-ray crystallography in 1989.⁹⁰⁻⁹² Since then, 386 HIV PR structures (363 HIV-1, 16 HIV-2, 7 SIV) of either unliganded or complexed with various inhibitors or substrates have been deposited in the Protein Data Bank (PDB).⁹³ The HIV-1 PR is a homodimer with C2 symmetry in the absence of ligand, albeit this is lost with the binding of asymmetric ligands. The enzyme is an

aspartic protease that consists of two identical 99-residue subunits, with each subunit contributing a catalytic triad (25Asp-26Thr-27Gly) to form the active site. A structure of an unbound homodimeric HIV-1 PR (PDB code 1HHP) is shown as Figure 1-6, in which flap (residues 43–58), flap elbow (residues 35–42), fulcrum (residues 11–22), cantilever (residues 59–75), the N- and C- termini (residues 1-4, and 95-99), and the active site cleft are represented.⁹⁴

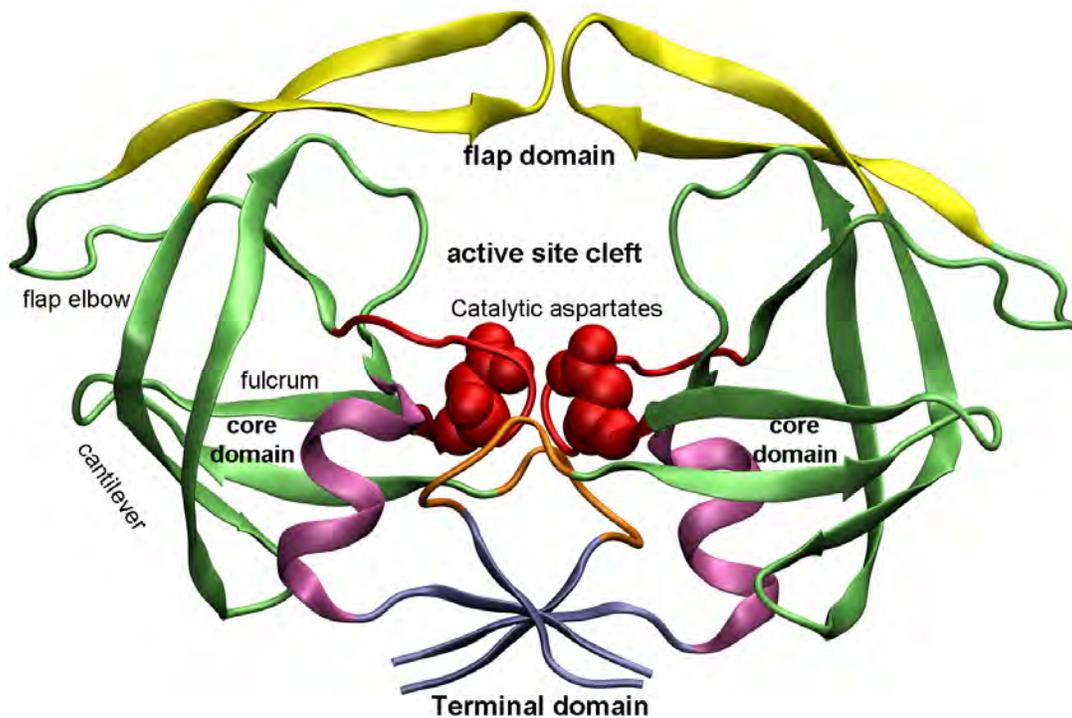


Figure 1-6. A structure of a homodimer of an apo HIV-PR (PDB code 1HHP). The flap regions are highlighted in yellow; the active site loops in red; N- and C- termini in ice blue; the single helix in each monomer in mauve; the turn (residues 5'-9') in orange; two aspartic acids (Asp25) are rendered as van der Waals spheres and colored in red.

The two monomers interact with each other at different regions; two of them are believed to contribute significantly to the dimer stability: the N- and C- termini region, forming a 4-stranded antiparallel β -sheet (Figure 1-7), and the active site region which is

stabilized by a hydrogen bond scaffold, termed as the 'fireman's grip' (Figure 1-8),⁹⁵⁻⁹⁶ involving hydrogen bonds formed between the oxygen of side chain of the active site Thr26 and the main-chain amide of the active site Thr26' on the opposite loop, as well as one hydrogen bond between the oxygen of side chain of the active site Thr26 and the main-chain carbonyl of the preceding Leu24' of the other molecule in the dimer.

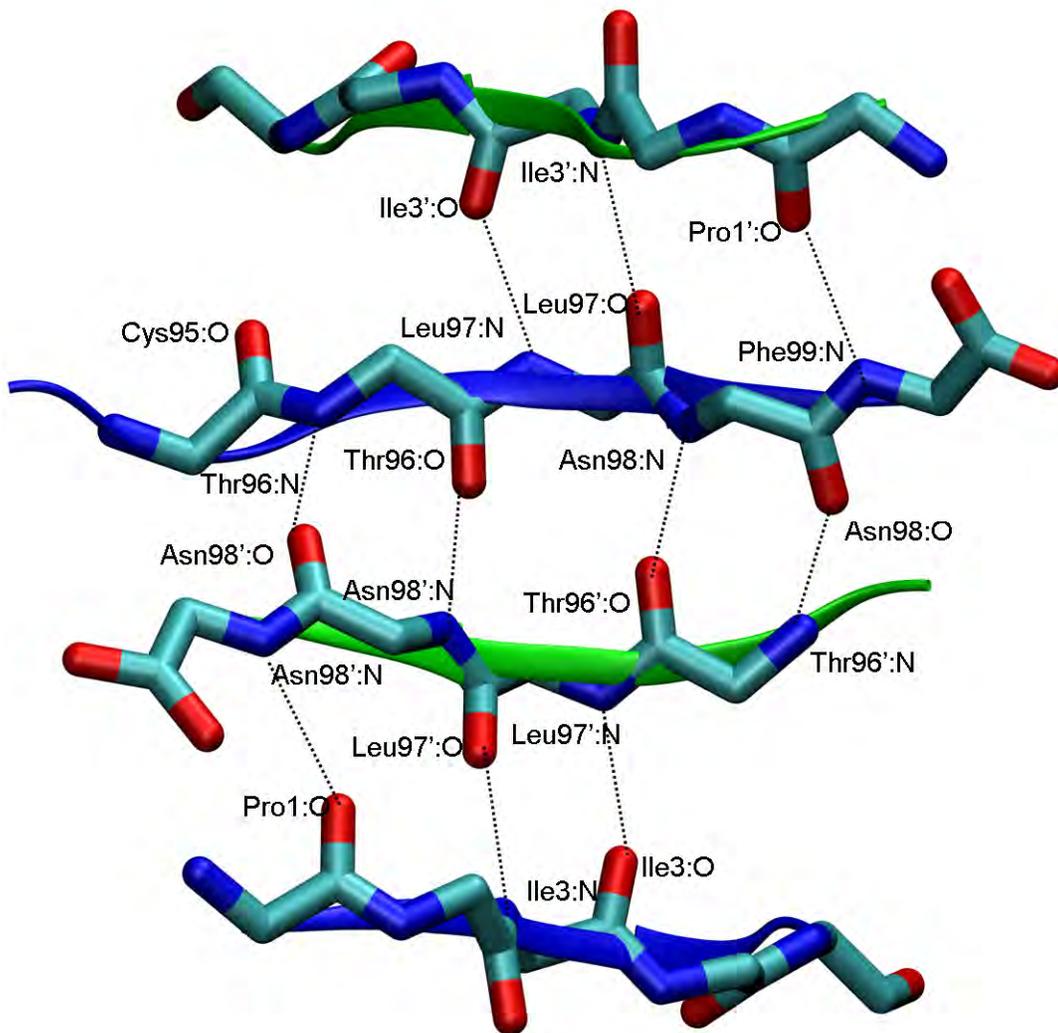


Figure 1-7. A detailed view of the 4-stranded antiparallel β -sheet at the N- and C-termini region of HIV-1PR.

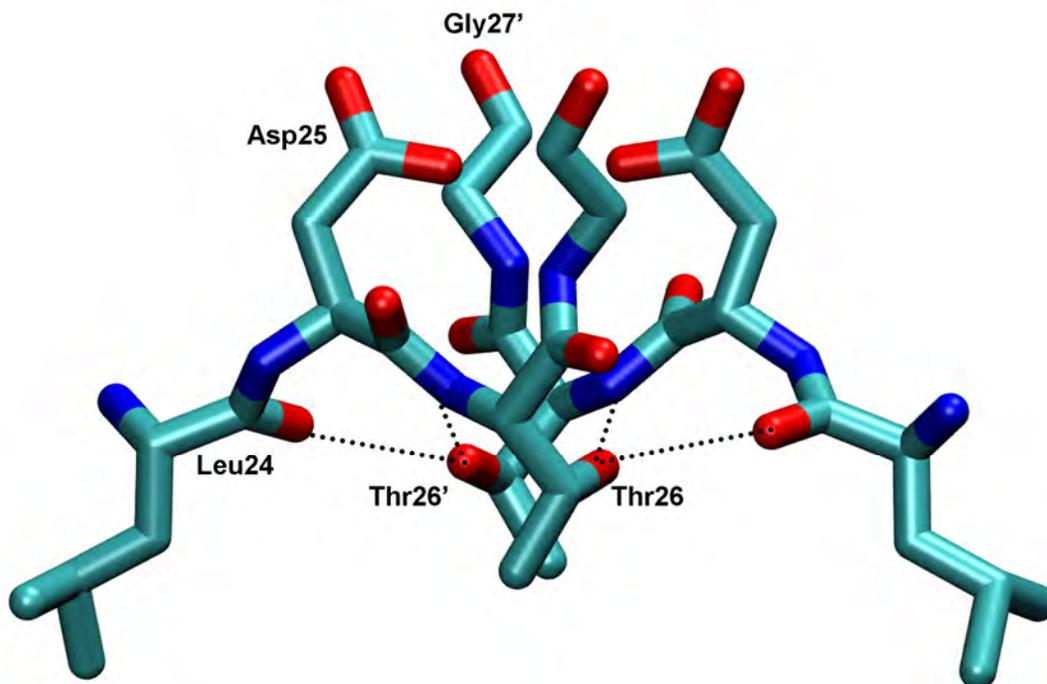


Figure 1-8. A detailed view of the active site region (Leu24-Gly27) of HIV-1 PR forming the ‘fireman’s grip’. Note the network of hydrogen bonds between Thr26, Thr26’ and Leu24, Leu24’.

In addition, the dimerization interactions also occur within the region encompassing the α -helix (residues 86-94) and the β -turn (residues 5’-9’, primes indicate residues from the symmetry-related monomer), including intramonomer salt bridges between Asp29 to Arg87, intermonomer salt bridges between Asp29 and Arg8’, and intermonomer hydrogen bond between Arg87 and Leu5’, as illustrated in Figure 1-9. These dimerization interactions have also been suggested to influence the dimerization significantly^{86,97}.

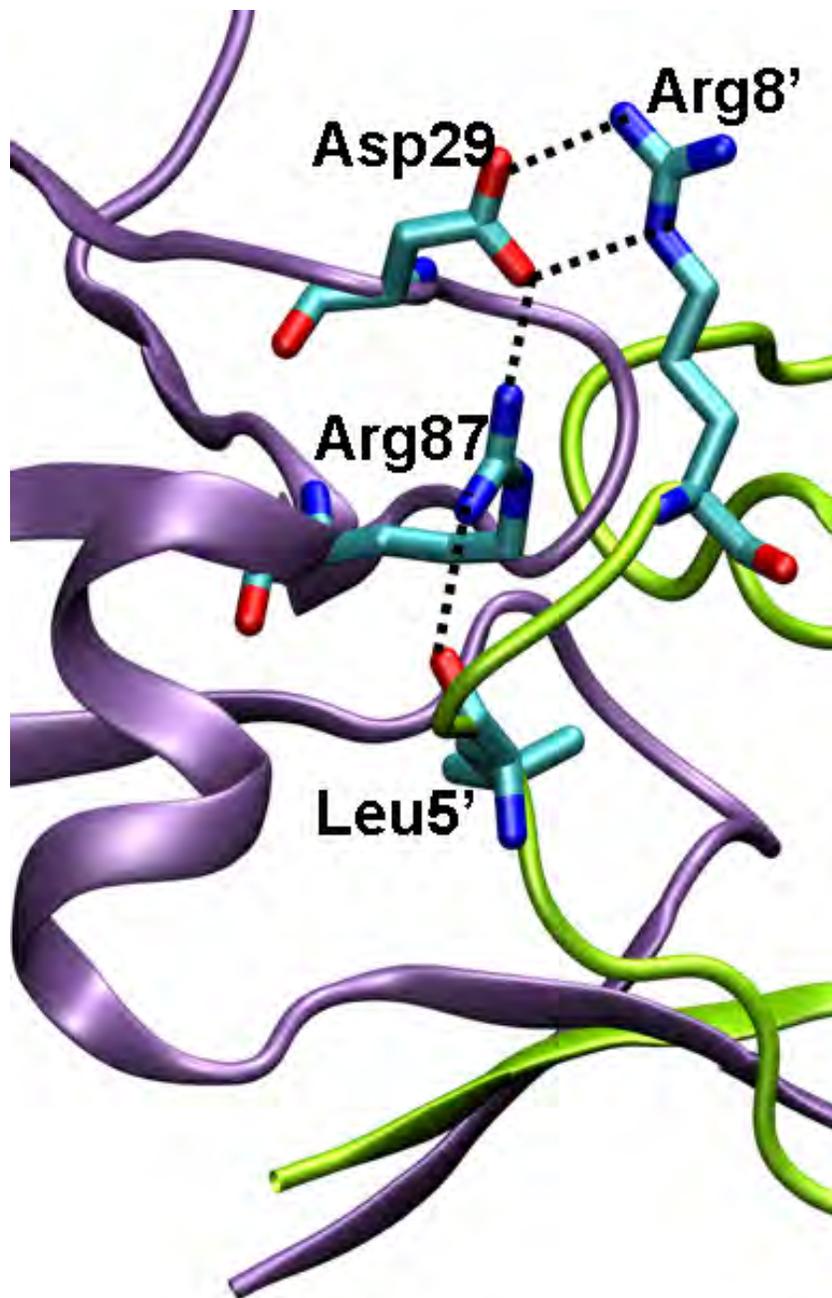


Figure 1-9. A detailed view of the dimer interface region encompassing the helix (residues 86-94) and the β -turn (residues 5'-9') stabilized by both intra- and inter-monomer salt bridges as well as an intra-monomer hydrogen bond between Arg87 and Leu5'.

1.2.2 Dynamics of HIV-1 PR: insight from experimental and MD simulations

As mentioned above, an extensive set of X-ray crystal structures of HIV-1 protease in both bound and unbound forms have been solved,⁹⁸ revealing a C2 symmetric homodimer with a large substrate binding pocket covered by two Gly-rich flaps^{90, 99-100}. Nearly all complexes have been solved in the ‘closed’ conformation, showing the two flexible glycine-rich β -hairpins, the so-called “flaps”, interacting with the ligand and completely blocking access to the active site (Figure 1-10a). Crystal structures of the ligand-free protease reported to date are more heterogeneous.¹⁰¹ Although most of them exhibit the “semi-open” form (Figure 1-7b), closed flaps were also seen in the crystal structure of two unbound tethered subtype B PRs (PDB code 1LV1 and 1G6L), and a subtype A unbound PR (PDB code 3IXO). Moreover, a “wide-open” form has been reported for three unbound PRs (PDB code 2PC0, 1TW7 and 2R8N). Although large-scale flap opening is presumably required to allow substrate entry since the active site access remains blocked in both the closed and semi-open forms, the ‘wide-open’ conformation of the flaps (1TW7) has been shown to be only transiently populated during the MD simulations without crystal packing contacts.¹⁰²⁻¹⁰³ In addition, earlier studies have suggested a role for crystal packing and crystallization conditions in the closed¹⁰⁴ and semi-open forms,¹⁰⁵ and other calculations have also suggested that the free energy difference between the different conformations of HIV-1 PR may be quite small,¹⁰⁶ implying that the equilibrium of different configurations of the flaps might be easily shifted by many factors such as mutations, ligand binding, and even crystal contacts. Thus, whether the “wide-open” crystal structure is relevant to the flap dynamics or drug

resistance, and what is the dominant structure of unbound HIV-1 PR in solution at room temperature remain open questions.

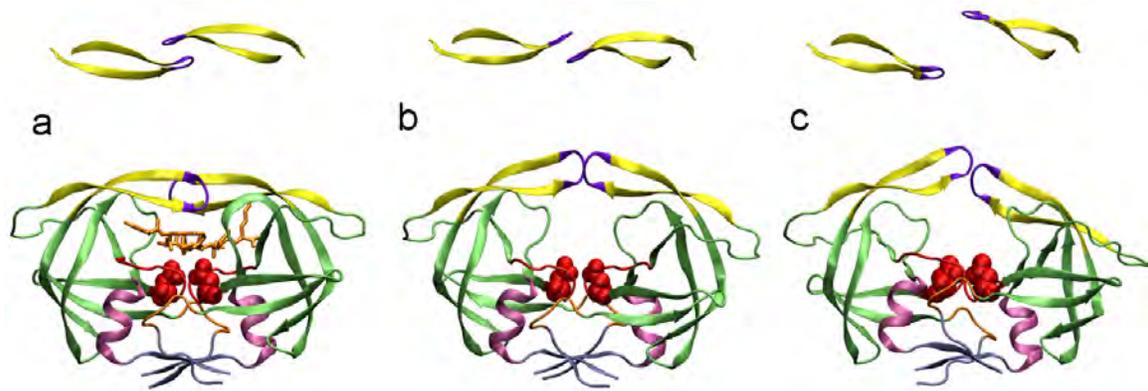


Figure 1-10. Cartoon draws of three distinct X-ray structures of HIV-1 protease: a) substrate-bound closed conformation (PDB code 1TSU), b) unbound semi-open conformation (PDB code 1HHP), and c) unbound wide-open state (PDB code 1TW7). Using the same color scheme as in Figure 1-6 (The flap tips are highlighted in violet). Top views illustrate the distance between the flaps and the reversal of flap handedness in the three conformations.

On the other hand, another plausible explanation for the variability of apo structures of HIV-1 PR may be the intrinsic flexibility of the protein, supported by an emerging concept that proteins sample an ensemble of conformations that meet functional requirements under equilibrium conditions.¹⁰⁷ This ability is structure-encoded, implying an evolutionary role in selecting/conserving structures based on their ability to achieve functional dynamics.¹⁰⁸ In fact, NMR experiments have revealed¹⁰⁹⁻¹¹⁰ that the flap region has a high degree of flexibility. Based on solution NMR data for the free protease, Torchia et al. have suggested¹¹⁰⁻¹¹² that the ensemble of unbound structures is dominated by the semi-open family with sub-nanosecond timescale fluctuations in the flap tips, and with the closed conformation possibly being a minor component of the ensemble. The

semi-open form is in slow equilibrium ($\sim 100\mu\text{s}$) with a less structured, open form that exposes the binding site cavity.

In addition, EPR spectroscopy was recently employed by Fanucci's group¹¹³ to investigate dipolar coupling of the unpaired nitroxide electrons in spin labels attached to K55C /K55'C on each flap. A different flexibility of the flaps in the bound and unbound forms was clearly identified, and the data suggested that the unbound flaps sampled a much larger degree of separation than those in the bound form.

Thus, these experimental data provide strong support for the hypothesis that the protease in the unbound state exists in a diverse ensemble of conformations fluctuating between semi-open, closed, and open, and exhibits considerable flexibility which allows substrate entry and product exit. Despite these findings, many aspects of both the structure and dynamics of HIV-PR in aqueous solution remain unresolved as the experiments provide only indirect evidence of protein structures in solution.

On the theoretical side, the experimental findings stimulate interest in exploring the relationship between protein dynamics and structural changes involved in function. Unfortunately, until recently, realistic simulations of HIV PR have been hampered by limitations in the model description and timescales that could be reached. Collins et al.¹¹⁴ reported flap opening resulting from MD simulations in the gas phase which involved forcing the atomic coordinates for non-flap regions of a closed structure to the semi-open state. Scott and Schiffer¹¹⁵ also observed irreversible flap opening, but the extent of flap opening was not quantitatively described. Instead the authors focused on the flap tip regions, which "curled" back into the protein structure during the opening event, burying several hydrophobic residues. This flap curling was hypothesized to provide a key

conformational trigger necessary for subsequent large-scale flap opening and therefore HIV-PR function. A more recent study¹¹⁶ highlighted the challenges in obtaining accurate simulation data by demonstrating that similar irreversible flap openings could arise from insufficient equilibration during system setup; these artifacts were not observed when more extensive solvent equilibration was performed. Later, Hamelberg and McCammon¹¹⁷ used activated dynamics to produce flap opening in HIV-PR. In this case, a *trans*→*cis* isomerization of the Gly-Gly peptide bond was hypothesized to trigger the flap opening. Perryman et al.¹¹⁸⁻¹¹⁹ reported dynamics of unbound wild-type and V82F/I84V mutant in which the closed form opened somewhat, but the authors did not report whether the flaps in these unbound protease simulations actually adopted the semi-open flap handedness observed in crystal structures. Nevertheless, the high flexibility of the flaps, particularly for the mutant, was demonstrated and used for active site inhibitor design for the drug-resistant mutant.¹²⁰ Notably, none of these prior computational studies of the free protease reported that the flaps were able to adopt the semi-open conformation from either the open structures that were sampled or in other cases from the initial closed conformation. Therefore it is unclear if such opening events are relevant to the true dynamics of the HIV protease or simply represent an inability of the simulation models to reproduce experimental observations.

Recently, several reports have been published where multiple and, most importantly, reversible opening of the protease flaps was observed. These serve as a testimony that simulation methods have finally reached a state where they can provide valuable insights into enzyme function on biologically relevant timescales. McCammon and coworkers¹²¹⁻¹²² developed a coarse grain model of HIV protease in which each residue is modeled

using a single bead at the position of the C α carbon. This treatment substantially reduces the complexity of the system, permitting the simulations to model behavior on the microsecond timescale. Numerous opening and closing events were seen; these were realized primarily by large lateral movements of the flaps that exposed the binding cavity. With the current coarse grain model, however, the long timescales are enabled yet at the cost of atomic detail, and there is also no straightforward way to determine how flap behavior is influenced by dynamics on the atomic level in terms of specific side chain interactions, or to gain an understanding of how solvation is coupled to dynamics.

Most recently, our group applied a multi-scale model to HIV PR dynamics in which full atomic detail was maintained for the protease, and aqueous solvent was modeled using a continuum approach.⁴⁵ These simulations showed spontaneous conversions between the bound and unbound crystal forms upon removal of an inhibitor, and reversible opening of the flaps. The simulations of the inhibitor bound form were very stable with no substantial conformational changes, thus providing additional support for experimental results that the flaps participate in stabilizing interactions with the ligand in the bound complex. In contrast, the behavior of the system changed dramatically in the absence of ligand; the closed flap rearranged to semi-open form, similar to what was observed in ligand-free protease crystal structures. More importantly, when these simulations were extended to longer times, flexibility of the flaps produced transient openings with large-scale rearrangements of the flaps and flap tip distances over 20Å. These fully open conformations were only transiently populated, and reproducibly returned to the semi-open state, indicating that the opening events were not likely artifacts caused by instability of the system or a poor quality model.

Even though the direct observation of the fully open structure and of conversions between three different flap conformations in atomic detail simulations was very encouraging, the question of how relevant these open states were for ligand binding remained unclear. To address this question, several studies¹²³⁻¹²⁴ were conducted by performing MD simulations following manual placement of either an inhibitor or a substrate into the active site of HIV-1 PR protease with an open conformation. In those simulations, the ligand induced the closing of the flaps in the closed conformation in an asymmetrical way, as seen in all inhibitor bound HIV PR crystal structures. Significantly, the asynchronous closing of two flaps observed in these MD simulations agree with a novel X-ray crystal structure solved in both the wild-type and drug-resistant variant complexes,¹²⁵ with one flap intermediate and the other flap closed. Hence, both experimental and theoretic studies suggest that a rearrangement of the ensemble of conformations sampled by the protease-binding pocket indeed occurs on ligand binding.

In summary, recent experimental and theoretical studies provide compelling evidence that HIV-1 protease assumes a well-defined ensemble of substates which are in a dynamic equilibrium (Figure 1-11), and the dynamic behavior of protein is most likely associated with its function. Thus, a thorough understanding of the fundamental mechanisms underlying the interconversions between different conformations may open new opportunities for developing protease inhibitors in which protease dynamics and flexibility are explicitly targeted in the inhibitor design process.¹²⁶

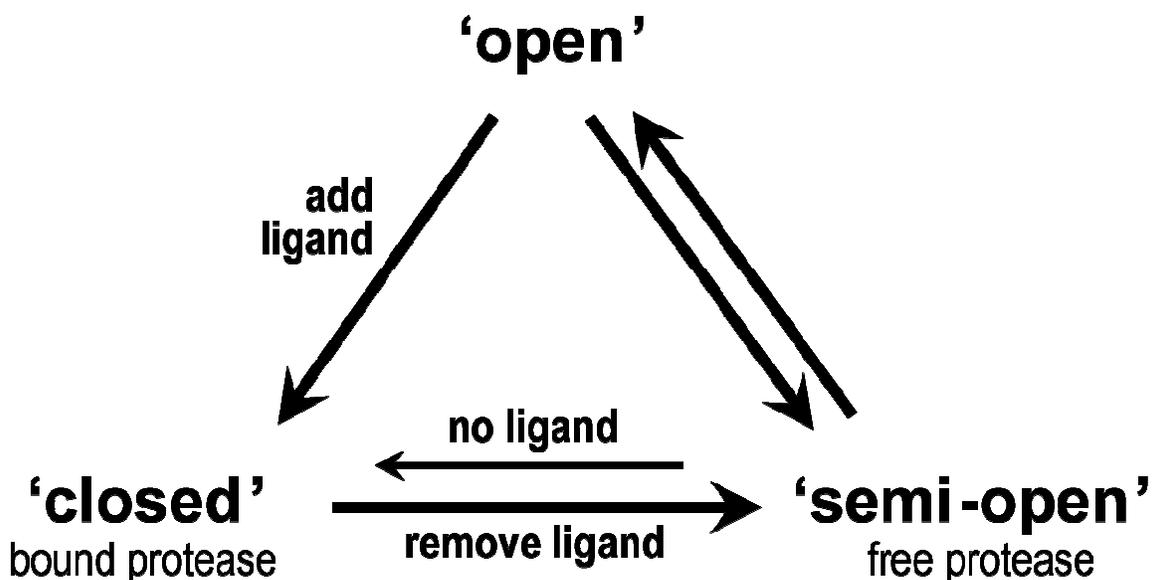


Figure 1-11. Schematic representation of simulated transitions between the three protease forms¹²⁶. The closed flap conformation converts to semi-open upon removal of ligand. Ligand induces the closure of the open form. Free protease exists in an ensemble of different conformers, closed, semi-open and open forms, which are in dynamic equilibrium.

1.2.3 Current HIV-1 PR Inhibitors

The understanding of the HIV life cycle was a major breakthrough in the discovery of the available HIV drugs. Although there is not yet a cure for HIV/AIDS, 25 anti-HIV drugs have been formally approved for clinical use and have greatly prolonged life by delaying the onset of AIDS.¹²⁷ Among these anti-HIV drugs, protease inhibitors (PIs) have emerged as potent antiretroviral agents available for the treatment of HIV infection, which were invented after the first three dimensional atomic structure was solved in 1989. The successful story of the clinical use of HIV protease inhibitors represents a remarkable achievement of structure-based drug design. Most of PIs are designed so that they imitate substrate binding at the active site of the PR in terms of hydrogen bonding

behavior, thus blocking the replication cycle of HIV. To date, there are 10 FDA approved drugs,¹²⁷ eight of them are peptide-like substrate analogues, including Saquinavir (SQV), Ritonavir (RIT), Indinavir (IDV), Nelfinavir (NFR), Amprenavir (APV), Lopinavir (LPV), Atazanavir (ATV) and Fosamprenavir (FPV). The two most recently approved drugs are nonpeptidic, Tipranavir (TPV) and Darunavir (DRV).

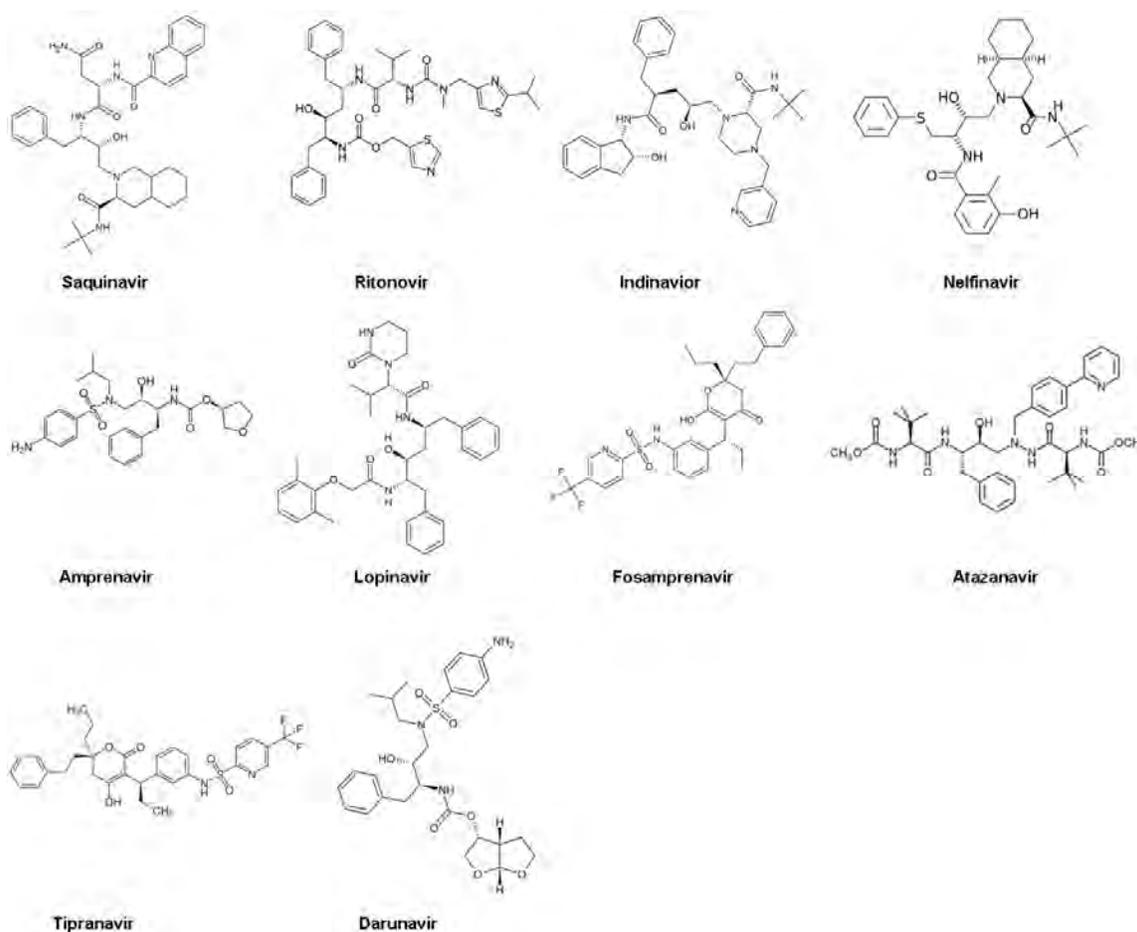


Figure 1-12. Chemical structures of 10 FDA approved HIV PR inhibitors.

In spite of the great success of the currently used competitive inhibitors in significantly reducing infection rates,¹²⁸ the efficacy of these drugs is limited due to the natural selection of protease variants that are still catalytically competent but have lower

affinity of the drugs than wild-type enzyme. Use of PIs for an extended period in the treatment of HIV results in resistance of HIV protease enzymes.¹²⁹ Hence, both academia and the industry are rigorously exploring the structure of PR aiming to come up with new ideas on designing new drugs that are more potent yet less prone to resistance.

One of the strategies that are being put in place is the modification of the already available PR drugs. To this end, Schiffer in her earlier MD simulation study¹¹⁵ proposed a model for overcoming resistance based on an observation of HIV-1 protease conformation with flaps “curled” such that they allow substrate access to the active site. In this conformation the hydrophobic tips of the flaps curl in and pack against the hydrophobic inside wall of the active site groove. The authors suggested that this ‘open’ conformation is crucial and the inhibitors should be designed to lock the flaps in their ‘open’ conformation, and believed that such inhibitors would be less susceptible to the development of drug-resistant variants. Later, to rationalize the design of new generation, the same group proposed that a ‘substrate-envelope’ within the HIV-1 PR binding pocket, defined by consensus volume occupied by the substrates, should represent a spatial constraint for the inhibitor design,¹³⁰ inhibitors that fit within the ‘substrate-envelope’ are less likely to be susceptible to drug-resistant mutations, since a mutation impacting such inhibitors would simultaneously impact the processing of substrates.

In addition, classical and *ab-initio* MD simulations reveal¹³¹⁻¹³² that protease flexibility modulates the activation free energy barrier of the enzymatic cleavage reaction. In drug-resistant mutants, the active site mutations are often associated with mutations that partially restore the enzymatic function (compensatory mutations) and frequently occur in regions distant from the active site. The mutations in these positions

may enhance the catalytic rate of the protease mutants by affecting the flexibility of the protein. While the authors provided a plausible explanation of how compensatory mutations work, they did not suggest how this understanding could be extended to the design of drugs that escape protease mutations.

Along this line, a crystal structure of the unbound HIV-1 PR for the multiple drug resistant (MDR) 769 isolate (PDB code 1TW7) in a ‘wide-open’ conformation was reported.¹³³ The authors indicated that drug resistance in this strain might arise at least partly from the changes in the flap conformation, and presumably the inability of current inhibitors to induce flap closing. However, as mentioned above, our previous MD simulations suggest that the wide-open structure observed for MDR 769 is not caused by the sequence variation, but instead is an artifact from crystal packing.¹⁰² Thus, this structure may not be directly relevant to studies of inhibitor entry or to the cause of HIV-PR drug resistance.

Another appealing explanation of resistance was provided by Freire based on thermodynamic differences observed between substrate and inhibitors.¹³⁴⁻¹³⁵ In solution, the peptide substrate has a higher flexibility than the synthetic inhibitors and therefore suffers a higher conformational entropy loss upon binding. On the other hand, due to its higher flexibility, the peptide substrate is more amenable to adapt to backbone rearrangements or subtle conformational changes induced by mutations in the protease. In contrast, the synthetic inhibitors are less flexible, and their capacity to adapt to changes in the geometry of the binding pocket is more restricted. This hypothesis is supported by the existing structural information for resistant mutants. For example, analysis of the crystallographic structure of the resistant mutant V82A revealed a widespread

rearrangement of the backbone around the binding pocket,¹³⁶ the loss in binding affinity by the synthetic inhibitors can be rationalized in terms of their inability to successfully accommodate to a distorted binding pocket. Thus, these observations provide a plausible explanation for the molecular origin of resistance. This finding should be helpful for the design of novel and more effective drugs, e.g., by developing flexible inhibitors that are capable of accommodating minor changes in the geometry of the binding site.

In addition, both previous computational and experimental studies have shown that there are differences in thermodynamic stability among the alternate protease forms that should be included when considering ligand binding affinity. Since the structure of the transient open form was only suggested in recent molecular dynamics studies, the description thus far has focused on thermodynamic differences between closed and semi-open forms. For example, the free energy change ΔG calculated by reaction path method estimated that the semi-open form is more favorable than closed, with stabilization contribution coming primarily from the entropic term.¹³⁷ This analysis is consistent with NMR relaxation data and is very reasonable given the high glycine content of the flap tips. As was shown by calorimetric experiments, a large favorable entropy change is also the major driving force for high binding affinity of current HIV-1 PR inhibitors.¹³⁸⁻¹³⁹ However, in this case it is the favorable solvation entropy associated with the burial of a large hydrophobic surface upon inhibitor binding. On the basis of thermodynamic analysis of wild-type and active site resistant mutant (V82F/I84V), Fieire et al.¹³⁸ suggested that the V82F/I84V active site mutation lowers the binding affinity of the inhibitors by affecting the binding enthalpy and to a lesser extent the binding entropy. From a structural point of view, two different effects account for the less favorable

binding thermodynamics: direct effects that alter the interactions between inhibitors and protein and indirect effects that alter the energetics of the conformational rearrangement of the protease upon inhibitor binding. Therefore, the free energy change associated with the conformational change of the protein and ligand has to be included in any accurate calculation of binding affinity.¹²⁶

These few examples of proposed mechanism of resistance raise several important points. To preserve the function of the mutant protease (i.e. still efficiently cleave the viral polyprotein) the enzyme can introduce alterations in the active site but the correct dynamics or flexibility must be preserved, and/or the active site changes must be compensated by the flexibility of the substrate. Since the competitive advantage of the synthetic inhibitors' strong binding likely arises from their rigidity, it has been rather challenging to design flexible inhibitors that bind stronger than the natural substrate while retaining the ability to adapt to a binding pocket that varies in shape. An example of such flexible inhibitor is KNI-764 (also known as JE-2147) which was shown to remain potent against MDR protease strains.¹⁴⁰ It was again demonstrated by calorimetric measurements¹⁴¹⁻¹⁴² that, in contrast to previous inhibitors, these second generation inhibitors bind strongly mainly due to favorable enthalpy change.

Another strategy to evade mutations in the active site is to design inhibitors that primarily form interactions with the backbone rather than side chains of the active site,¹⁴³⁻¹⁴⁴ such that mutations in the binding site may not effect the inhibitor binding. Moreover, based on MD simulation and free energy calculation,¹⁴⁵ Hou et al. proposed that when a mutation occurs to a not-well conserved residue, presumably unimportant for viral function, it impairs the binding of inhibitors more than substrates, such that causes drug

resistance. Therefore, they suggest that an ideal drug should only form strong interactions with the most conserved residues such as Asp25, Gly27, Ala28, Asp29, and Gly49.

1.2.4 Alternative Strategies

Sequence alignments have identified that there are five conserved domains in all examined HIV sequences derived from treatment-naïve patients, including residues 1-9 and 94-99 (N- and C- termini), 21-32 (active site core), 47-56 (flap region) and 78-88 (substrate-binding region).¹⁴⁶ Thus, it has been suggested that compounds binding conservative domains of the enzyme outside the active site might be ‘resistance-repellen’.¹²⁸ Moreover, inhibitors targeted to the domains outside the active cleft might show a synergistic effect to the conventional active-site targeted compounds. Finally, blocking an earlier event in the maturation pathway of the virus, such as HIV PR dimerization, could constitute an alternative strategy to the conventional targeting of the active site and lead to potent inhibitors for PR mutants.

As mentioned earlier, HIV PR is only active as a dimer, in which each of the two catalytic aspartates is contributed by one monomer. Experimental studies have shown that there is an equilibrium between the monomer and dimer of HIV-1 PR,¹⁴⁷⁻¹⁴⁸ although the dimer dissociation constant (K_D) varies over a range of 10^4 –fold, from 39pM to 0.4μM,¹⁴⁹ depending on assay conditions. Therefore, blocking the dimerization of the protease monomers could be an effective means for inactivating the enzyme.¹⁵⁰ As mentioned above, the two monomers interact with each other at different regions; one major dimer interface region is a four-stranded antiparallel β -sheet composed of the two

N-termini (residues 1-4) interdigitating the two C-termini (residues 96-99), which was suggested to contribute to 75% of the stabilizing energy.¹⁵¹ Thus, this highly conserved dimer interface region represents an attractive target for development of ligands preventing dimerization.¹⁵²⁻¹⁵⁴ Submicromolar inhibitors were obtained when using a flexible linker¹⁵⁵⁻¹⁵⁶ or more rigid scaffolds.¹⁵⁷⁻¹⁵⁸ In addition, novel interfacial peptides which are tethered through their side chain have shown more potency with a low nM inhibition constant.¹⁵⁹ Although designed agents have had good dimerization inhibitory activity against HIV-1 PR, they suffered from their high molecular complexities. To overcome this drawback, truncation and mutation studies were performed to find the minimal structure necessary for activity. Modification of the termini of an interfacial peptide by attachment of a lipophilic group and alkyl chains has shown to improve both the inhibition potency and the specificity.¹⁶⁰⁻¹⁶¹ A highly potent HIV-PR dimerization inhibitor is an alkyl tripeptide, palmitoyl-Leu-Glu-Tyr, with K_{id} of 0.3 nM. In addition, other interface peptides have been reported to inhibit the dimerization of HIV-1 PR by forming a disulfide bond with the Cys95 residue.¹⁶²⁻¹⁶³ Moreover, interface peptides with a cell permeable domain (CPD) derived from HIV-1 tat also exhibited dimerization inhibition, yet with K_i values in low micromolar range.¹⁶⁴ Also, a monoclonal antibody targeting N-terminal of HIV PR (residues 1–6) inhibits activity of both HIV-1 and HIV-2.¹⁶⁵⁻¹⁶⁶

Besides a significant effort to develop dimerization inhibitors of HIV PR and characterize their binding on a structural level,¹⁶⁷ there is also great interest to develop allosteric inhibitors that do not directly compete with substrate for the same binding site but indirectly change the flexibility of the protease such that the thermodynamic balance

of the closed, semi-open, and open ensembles is shifted.¹²⁶ The possibility of allosteric inhibitors of HIV protease was suggested previously,¹¹⁸⁻¹¹⁹ and it has been argued that the presence of allosteric site is very likely for all dynamic proteins¹⁶⁸ that exist as a population of conformational states. The allosteric inhibitors do not compete with natural substrate and thus their effect is not decreased by higher concentration of the substrate. As a result, considerable effort has been invested in the identification of allosteric sites in HIV-1 PR.

Based on molecular dynamics simulations that showed anticorrelated behavior between flap opening and compression of the allosteric site in the elbow region, McCammon et al. suggested targeting the protease elbow regions (see Figure 1-6) as an allosteric site.¹²² With the exception of an insertion in position 35, no resistance mutations are associated with this region. Thus, the flap elbow might represent a promising drug target. It is interesting to note that the experimentally determined structure with an open binding pocket¹³³ (the crystal structure of the MDR isolate discussed above) indeed has a crystal packing contact involving insertion of residues from a symmetry-related neighbor into the elbow region,¹⁰² thus providing evidence that this site may be a promising candidate for allosteric inhibition.

Recently, Carlson and co-workers presented a novel mode of action for HIV-1 PR inhibitors: modulating the conformation behavior of HIV-1 PR by targeting the flap-recognition site,¹⁶⁹ inspired by the observations of a 5-7 Å shift from the apo form when the flaps close over the active site, and an inward rotation of each monomer including the reversal of flap “handedness”.¹²³

Another potential target for allosteric inhibition is the dimer interface at the N-, C-termini region. NMR experiments measuring backbone amide chemical exchange transverse relaxation rates¹¹¹ indicated that the flexibility in the four-stranded β -sheet dimer interface increases upon inhibitor binding, suggesting a coupling between the binding site and the dimer interface. This coupling (even though in the opposite direction) is also observed in the crystal structure of a free HIV-1 protease in which the N- and C-termini of the two protease monomers were tethered.¹⁷⁰ Unlike all other crystal structures of the free protease, this ‘monomeric’ protease exhibits the closed flap conformation. In addition, an interesting report¹⁷¹ demonstrated that some of the inhibitors initially designed to prevent dimerization actually did not disrupt the dimer interface and yet showed substantial protease inhibition. The authors thus concluded that these compounds acted as allosteric inhibitors binding at the dimer interface, indirectly reducing the binding affinity of the substrate.

Further evidence that these sites may provide useful targets for allosteric inhibitors has been shown by, Rezacova et al.,¹⁷² who developed monoclonal antibodies with potent inhibition of the protease function. These targeted two non-binding site regions of the enzyme: one corresponds to residues 36-46 (flap elbow) and the other to residues 1-6 (N-terminal) at the dimer interface. The proposed inhibition mechanism based on the crystal structure of the antibody fragment in complex with the 36-46 epitope peptide postulates that antibody binding prevents flap closure over the active site. Moreover, another example of potentially exploitable allosteric inhibition was reported for beta-lactam compounds.¹⁷³ The authors demonstrated that the inhibitors are noncompetitive, and they only interact with ligand-bound enzyme. In addition, they suggested the mechanism of

inhibition through interaction of beta-lactam compounds with the closed flap region of the enzyme-substrate complex.

In light of the functional significance of the flaps which undergo conformational changes upon ligand binding, another alternative inhibitory mechanism previously proposed is to target the thermodynamic balance of different conformations of the flaps. One example of inhibitors against the flap region are Nb-containing polyoxometalates (POMs)¹⁷⁴. They bind to a cationic pocket on the outer surface of the flaps involving residues Lys41, Lys43, and Lys55, and exhibit both high selectivity and activity against HIV-1. The noncompetitive inhibition mode was further confirmed by the computational studies as well as the kinetics and binding studies. In addition, a recent fragment-based crystallographic screening against HIV-1 PR has also identified small molecules which bind to two novel sites outside the active site of the PR dimer in its inhibitor-bound, closed conformation¹⁷⁵, corresponding the 'exo site' adjacent to the Gly16, Gly17 and Gln18 loop and the 'outside/top of the flap' encompassing Trp42, Pro44, Met46, Lys55, Val56 and Arg57. Thus, both studies provide experimental evidence for the outer surface of the flap as a potential new target site for allosteric inhibitors.

Another example of inhibitors targeting the flap conformation is metallacarborane-based¹⁷⁶ compounds. The metallacarboranes bind to the hydrophobic pockets as in the flap-proximal region of the enzyme, above the site for conventional active site inhibitors. The authors proposed that these compounds block up flap closure in addition to filling the binding pocket as conventional PIs.

Thus, these studies provided experimental confirmation of the existence of allosteric binding sites that were predicted in previous simulations, and support the idea that

various conformations of the flaps can be specifically targeted to control PR activity, demonstrating the possibility for allosteric control of HIV protease. To this end, the conformation behavior of the flap region has been extensively studied theoretically.^{111, 126} However, sampling large magnitude conformational changes of HIV-1PR has been hampered by the long timescale suggested by NMR relaxation data¹¹¹.

1.3 Overview of My Research Projects

As a powerful tool to provide a detailed, atomic resolution model for time-dependent structural evolution, MD simulations can provide estimates of the energetics associated with different HIV PR states. Thus, in the present study, we employed MD simulations to investigate the structure and dynamics of HIV-PR. This dissertation contains four research projects aimed to obtain insight into the mechanism underlying these conformational changes of HIV-1 PR and drug resistance caused by mutations of viral proteins. Some of this work was conducted in close collaboration with the experimentalists in Gail Fanucci's lab at the University of Florida.

1.3.1 Exploring Rearrangements between the Closed and Semi-open Conformations of HIV-1 PR by MD simulations

To obtain insight into the mechanism underlying the conformational changes of HIV-1 PR, we performed microsecond-long simulations of an apo wild-type HIV-1 PR. In these simulations, multiple and reversible interconversions between different

conformations of HIV-PR were captured, thus providing a unique opportunity to investigate the mechanism underlying these transition dynamics. We first explored the transition mechanism between the closed and semi-open forms, the two most populated conformations in both experiments and MD simulations. Detailed structural analysis suggests that the rearrangement of flap from the closed to the semi-open conformation is likely induced by the twisting of the backbone of the flap tips, caused by the rotation of the backbone of the flap tips owing to intrinsic properties of the glycine residues on the flap region. The backbone rotation, in turn, disrupts the interflap interactions such as interflap hydrogen bonds between the two flap tips as well as the intermonomer hydrophobic contacts between the flap tip Ile50 residue and the hydrophobic residues from its symmetry-related monomer.

1.3.2 Investigating the Gating Dynamics of the Flaps

In one of our microsecond simulations, transient and reversible full opening of the flaps was captured. The detailed structural and energetic analyses reveal that it is the various binding interactions of the dimer interface that governs the gating properties of the flaps; the opening of the flaps results from the concerted partial dissociation of the dimer interface facilitated by water dynamics. The significance of the inter-subunit interactions along the dimer interface in the gating dynamics is further supported by the subsequent simulations on a double mutant system (R87K/D29A), in which the flaps opened more often and rapidly due to the decreased dimerization energies. More significantly, this working model offers a novel site for allosteric regulation of the gating dynamics of the flaps; targeting the highly conserved yet weak region of the dimer

interface encompassing the single α -helix (residues 86–94) and the β -turn (residues 4'-9') may affect the equilibrium of different conformational states, and thus inhibits its catalytic activity. It is worth noting that since all currently approved FDA PIs target the closed conformation, developing of inhibitors targeted to the open flap conformation with a different binding mode might be an alternative to circumvent the cross-resistance.

1.3.3 Solution Structure of HIV-1 Protease Flaps Probed by Comparison of Molecular Dynamics Simulation Ensembles and Electron Paramagnetic Resonance (EPR) Spectra

Previously, Fanucci's group performed site-directed spin labeling (SDSL) to derive conformational flexibility of the flaps in the absence and presence of inhibitor (Ritonavir), via electron paramagnetic resonance (EPR) spectroscopy measurements of dipolar coupling of the unpaired nitroxide electrons in spin labels attached to K55C /K55'C on the flaps of LAI consensus sequence.¹¹³ This work is particularly notable since for the first time, experiments characterized the extent of flap opening in an unbound form, and different conformations and flexibility of the flaps in the bound and unbound forms were able to be distinguished as well. Yet owing to the intrinsic experimental limitations, there is a need to establish the correlation between EPR-measured interspin distances and structural and dynamic features of the flaps. For this purpose, we performed a series of MD simulations in explicit solvent on the same LAI consensus sequence in the bound and unbound forms. The reconstructed distance distribution profiles from our MD simulations of the both bound and unbound protease agree quite

well with EPR measurement. Both experimental and theoretical studies characterized the restricted fluctuations in the presence of the inhibitor comparing with those in the absence of the inhibitor, providing further evidence that a rearrangement of the flap region undergoes upon ligand binding. Moreover, it is most likely the semi-open form is the dominant conformation of the unbound LAI consensus sequence of HIV-1 PR in solution. In addition, this work not only confirms the robustness of our protocol, but also that MD simulations have reached a stage where structural and thermodynamic properties of biological systems can now be reproduced and predicted. This work was published in Journal of the American Chemical Society (JACS) in 2008.

1.3.4 Exploring Drug Resistance Mechanism by the Combination of MD Simulations and EPR Spectroscopy

The emergence of multi-drug-resistant strains creates an urgent need to develop novel drugs. Deeper insight into the effect of resistance mutations on the structure and dynamic behavior of HIV-1 PR would greatly facilitate the design of inhibitors that could overcome resistance. We collaborated with Gail E. Fanucci and her colleagues to study the flap conformations of two drug-resistant HIV-1 protease constructs, MDR' and V6'. Again, MD simulations accurately regenerate the experimentally determined distance profiles and provide structural interpretations of the EPR data. The combined analyses show that the average conformation of the flaps, the range of flap opening and closing, and the flexibility of the flaps differ markedly in HIV-1 PR as multiple mutations arise in response to antiviral therapy. Thus, both experimental and theoretical studies provide

valuable insight into the coupling of drug resistance and protein backbone conformational flexibility. We suggest that the limited conformational opening of the flaps in V6' might alter the ability of the inhibitor, and possibly substrate, to enter into the active site cavity, whereas in MDR769', the longer average semi-open distance might increase the free energy cost for the flaps to close tightly accommodating inhibitor or substrate. This work was published in JACS in 2009.

Chapter 2

Exploring Rearrangements between the Closed and Semi-open Conformations of apo HIV-1 PR by MD simulations

Abstract

HIV-1 protease (PR) remains a prime target of anti-AIDS drugs. The flaps of HIV-1 PR are known to be highly flexible and undergo substantial conformational changes even in the absence of a ligand. A complete understanding of the detailed mechanism of the flap dynamics is crucial in rational design of more effective treatment regimes. Here, we present a hypothesis, based on microsecond molecular dynamics simulations of an apo protease, describing how the twisting of the backbone of the flap tips transforms the geometry of the β -hairpin structure of each flap from the 'closed' conformation to the 'semi-open' one, most likely owing to the intrinsic flexibility of the glycine residues. In addition, the twisting of the β -hairpin disrupts the interflap hydrogen bonds between the two flap tips, as well as van der Waals contacts between the flap tip Ile50 residue and hydrophobic residues from the symmetry-related monomer. The disruption of the inter-monomer interactions facilitates swapping the Ile50 residue between the two hydrophobic clusters within each monomer. To verify that the dynamics of the flap tip Ile50 residue is a key determinant of the conformational rearrangements, rather than just a measurement of the motion accompanied with the transition between the closed and semi-open forms,

two subsequent single mutant simulations (I50A and I50W) were carried out. Neither mutant system assumed the proper closed form seen in the wild-type simulations. Thus, our transition mechanism sheds insight into the dynamics of HIV-1 PR, implying the invariance of the flap tip residues Gly49 and Ile50, and highlighting the significance of maintaining a favorable hydrophobic environment within each monomer in protein dynamics.

Acknowledgments

This chapter contains direct excerpts from the manuscript (presented with minor modifications) written by Fangyu Ding with suggestions and revisions from Professor Carlos Simmerling. Additional assistance was provided by Amber C. Carr.

2.1 Introduction

HIV-1 PR is essential for the life cycle of the virus. It cleaves the Gag and Gag-Pol polyprotein precursors to produce the mature and functional Gag and Pol proteins; in the absence of HIV-1 PR activity, the viral particles are noninfectious.^{88,177} The introduction of multidrug HIV treatment regimens referred to as highly active antiretroviral therapy (HAART),¹⁷⁸ has dramatically extended the progression time between HIV infection and the development of AIDS.¹⁷⁹ However, the success of the treatment has been hindered by

the increasing drug-resistance observed in clinical HIV strains following long time treatment, caused by short life cycle and high error rate of viral replication, as well as the pressure of natural selection.¹⁸⁰

To develop novel inhibitors which are more potent yet less prone to resistance, continuing efforts have been made to elucidate the structure and dynamics of this protein and have yielded valuable information. X-ray crystallography has resolved diverse crystal structures for apo HIV-PR,¹⁸¹ including the ‘semi-open’, ‘wide-open’,¹⁸² and ‘closed’ forms.¹⁸³ The heterogeneity of the apo structures might reflect the intrinsic flexibility of the flap regions; however, it could also be an artifact of the crystallization conditions and/or the crystal packing.^{86, 105} Nonetheless, Nuclear Magnetic Resonance (NMR) relaxation studies^{110, 184-185} have also identified the flexibility of flaps in the apo state and predicted a slow (μ s timescale) equilibrium among different conformers. Recent pulsed Electron Paramagnetic Resonance (EPR) measurements^{113, 186-187} have also characterized the increased flexibility of the unbound form. Although these studies have provided invaluable information about the protein’s structure and dynamics, the determination of the details of coupling between the dynamics of individual residues, flap conformational changes, and drug resistance has not been experimentally accessible.

To this end, MD simulations have been employed as a powerful tool to explore the dynamics of the flaps associated with the enzymatic function of the protease. It has been well established that the flexible flaps govern the access of ligand to the active site. The flaps need to completely open in order to give a ligand access to the active site, and they must close once the ligand is positioned appropriately in the binding pocket to allow the subsequent cleavage event to occur. Other than these facts, no agreement regarding the

transition mechanism of the flaps has been reached, and various mechanisms have been proposed by different groups.^{115, 118, 122, 188-191} Therefore, there is still a need for a thorough understanding of the issues that govern HIV-PR flap dynamics, which could have profound implications for designing new therapeutic agents such as allosteric inhibitors, which prevent the flaps from changing conformation and thereby interfere with substrate binding and/or catalytic function.

Previously, we performed all-atom MD simulations on an apo HIV-1 PR with a continuum solvent representation⁴⁵, and reproducibly sampled transitions among different conformations of the apo state. However, it has been reported that this simplified model has provided results only in qualitative agreement with the data obtained with the explicit solvent or experimental observations, likely owing to the insufficiency of the implicit solvent model to accurately describe the interactions between hydrophobic groups or to the gross overestimation of electrostatic energy between the charged groups (overstabilizing salt bridges).^{44, 58, 192-194} Additionally, the use of an implicit solvent model did not provide an opportunity to probe the solvent dynamics, which are coupled with the protein dynamics.⁹² Therefore, exploring the dynamics of HIV-1 PR with explicit solvent is highly desirable to resolve several fundamental questions associated with the catalytic function.

In the present work, we carried out unrestrained all-atom MD simulations on an apo wild-type HIV-1 PR using an explicit solvent model, initiated from two distinct conformations, the closed and the semi-open. The implementation of a high temperature (375K) and the microsecond timescale allowed extensive sampling of conformational space in both simulations, and captured reversible and multiple interconversions among

different conformers. In this present work, we are aimed to elucidate the transition mechanism between the closed and semi-open form, the two major crystal structures of HIV-1 PR. These results might shed insight into the nanosecond timescale motions of the flaps of HIV-1 PR, providing important guidelines for design of novel potent inhibitors.

To this end, we propose that the rearrangements between these two forms are most likely induced by the twisting of the backbone of the highly flexible flap tips, with the sequence Gly48-Gly49-Ile50-Gly51-Gly52; this backbone twisting in turn disrupts the intra- and inter-flap hydrogen bonds as well as van der Waals (vdW) interactions between the flap tip residue Ile50 and the hydrophobic clusters within each monomer. In addition, to investigate the role of Ile50 in the protein dynamics, two single mutant simulations were carried out by substituting Ile50 to either Ala or Trp. In either mutant, the flaps cannot properly close. Thus, these simulations highlight that a hydrophobic residue with the correct size and hydrophobicity is required at position 50 to allow the tip of the flap to undergo the conformational change and bury itself in the hydrophobic core within each monomer. Moreover, this study also provides a theoretical justification of the inhibition mechanism of a novel inhibitor class,¹⁶⁹ which binds in the hydrophobic cluster within each monomer, thus preventing the flaps from assuming the proper closed conformation.

2.2 Methodology and Model Systems

2.2.1 Initial Preparation

MD simulations were performed on the same sequence of an apo wild-type HIV-1 protease (PDB ID code 1TSU¹²⁵) with both catalytic Asp residues modeled as Asn to be consistent with the crystallographic experiments, and initiated from two distinct conformations. The simulations are termed as MDclosed and MDsemi-open simulation, respectively, according to their initial conformation in the following context. The closed form with the coordinates obtained from a complex X-ray structure after removal the substrate, and the semi-open form with the coordinates from an unliganded crystal structures (PDB ID code 1HHP¹⁹⁵). The residues of the protease monomers labeled as A and B in the crystal structure were numbered 1-99 and 1'-99', respectively. Missing hydrogens were inserted by the LEaP module in the AMBER 9.0 software package.¹⁹⁶ The ff99SB¹⁹⁷ force field was used to describe the protein parameters.

For the additional two single mutant systems, the coordinates for the two starting structures, closed and semi-open forms, were also taken from the same crystal structures as in the wild-type simulation, 1TSU and 1HHP, respectively, with a single mutation (I50A or I50W) introduced by Swiss-PdbViewer.¹⁹⁸

Simulations with explicit solvent used TIP3P water model¹⁶ and each system was solvated in a truncated octahedron periodic box containing 6451 TIP3P¹⁶ water molecules. The distance between the edges of the water box and the closest atom of the solutes was at least 7Å. The time step for all simulations is 2fs. Bond lengths involving hydrogens were constrained using the SHAKE¹⁹⁹ algorithm. An atom-based cutoff of 8Å was used for non-bonded van der Waals interactions. The long-range electrostatics was calculated using the particle-mesh Ewald method (PME).²⁵

2.2.2 Minimization and Equilibration Protocols

Before MD simulations, the two wild-type systems were subject to energy minimization in three stages to remove bad contacts between the complex and the solvents molecules. Firstly, the water molecules were minimized by keeping all heavy atoms of the solutes restrained with a force constant of 50 kcal/ (mol·Å²). Secondly, the restraint force constant was then reduced to 10 kcal/ (mol·Å²). The entire systems were finally minimized without restraints. At each stage, the steepest descent minimization of 500 steps was performed followed by a conjugate gradient minimization of 2500 steps. The systems were then heated gradually from 100K to 300K in 100ps under *NVT* condition using the Berendsen algorithm²⁰⁰ with a coupling constant of 0.5ps, with all heavy atoms of protein restrained by 10 kcal/mol·Å². Three stages of equilibration each lasting 50ps were followed under *NPT* condition with a coupling constant of 1ps and a constant pressure of 1atm. During the first two stages, positional restraints were imposed firstly on all the heavy atoms, and then on the backbone with a force constant of 1.0 kcal/mol·Å². Finally, a short equilibration of 20 ps without any restraints was performed.

To alleviate steric clashes caused by a single mutation, especially the bulky residue Trp, the two single mutant systems were first subjected to a stepwise minimization and equilibration in the presence of the implicit solvent using a modified Generalized Born (GB) model.³⁶ It starts with one thousand steps of steepest descent minimization with positional restraints used first on the introduced mutated residue (I50A or I50W), then on all heavy atoms, and finally on only backbone atoms. Restraint force constants were decreased from 50, 10 to 1 kcal/mol·Å² in each stage. All atoms were finally permitted to

move freely. After initial minimization, the mutated systems were heated up from 100K to 300 K over 100ps. The temperature was controlled using Langevin dynamics with a reduced solvent viscosity (collision frequency of 1ps^{-1}) and a time step of 1fs. The systems were then subjected to a three-step equilibration over 150ps. Positional restraints were applied to the backbone atoms with force constants of 5 to 1, 0.1 kcal/ ($\text{mol}\cdot\text{\AA}^2$), respectively. Thereafter, the systems were solvated in a truncated octahedron periodic box containing 5953 and 7641 TIP3P¹⁶ water molecules for the two I50A system starting from closed and semi-open states, respectively; 5955 and 7626 TIP3P water molecules for the I50W systems starting from the closed and semi-open state, respectively. Then all mutated systems were subjected to stepwise equilibration following a procedure as described above for the wild-type system.

2.2.3 Production Runs

The configurations from the above equilibration stages were used as the starting points for the production runs. For the wild-type, two different temperatures were used for each configuration, 375K and 300K. At 375K, the production runs were extended to one microsecond for the MDclosed simulation, and 400ns for the MDsemi-open simulation. Each production run was only extended to 100 ns at 300 K. For most results reported in the following, the high temperature simulations were used due to the fact that the lower temperature simulations had little structural changes during the simulations. The high temperature simulations were done with the NVT ensemble with a coupling

constant of 1ps. The temperature was controlled by the Berendsen thermostat.²⁰⁰ The coordinate sets were saved at every 10 ps for subsequent analyses.

2.2.4 Data Analysis

Root-mean-squared deviation (RMSD), distance, radius of gyration (Rg) of hydrophobic cluster, torsion angles and atomic fluctuation were calculated using the PTRAJ module in the AMBER10 software package.¹² The atomic position fluctuations were computed for the backbone atoms (N, C α , and C), the initial closed structure was used as the reference. Molecular graphics were prepared by VMD,² PyMol¹ and Chimera.²⁰¹

2.3 Results and Discussion

2.3.1 Structural Differences in the Closed and Semi-open Conformations

MD simulations were performed on the same sequence of an apo wild-type HIV-1 protease (PDB ID code 1TSU¹²⁵) starting from two different structures, the closed and the semi-open conformations, corresponding to the bound and unbound form of HIV-1 PR, respectively. The preparation of the two models can be found in the Methods section. As revealed by hundreds of X-ray crystal structures of HIV-1 protease that have been resolved in bound and unbound forms, the active enzyme is a C₂ symmetric homodimer

with a binding pocket covered by two β -hairpins, or so called flaps. Large-scale mutagenesis has been done to identify the side chains required for PR activity at each residue in the flap region.²⁰² Met46, Phe53 and Lys55, whose side chains direct outward toward solvent, are the most tolerant to substitutions; Ile47, Ile50, Ile54 and Val56, with the side chains pointing inward toward the active site, only tolerate a few conservative substitutions; and the Gly-rich region, Gly48, Gly49, Gly51, Gly52, which easily accounts for the high degree of mobility demonstrated in this region by NMR, is highly conserved.²⁰² Moreover, a recent study using MD simulation and MM-GBSA calculation has also revealed that the residues in the flap region, Ile47, Gly48, Gly49, and Ile50, greatly contribute to inhibitor binding affinity.¹⁴⁵

In spite of the above common structural features, the conformation of the flap region clearly differs in the ‘closed’ and ‘semi-open’ form. As illustrated in Figures 2-1b, in the ‘closed’ form, both flaps are pulled in toward the bottom of the active site, making contacts with each other by forming an inter-flap hydrogen bond between Ile50 (acceptor) and Gly51’ (donor).⁴⁵ Additionally, Ile50 from one monomer is positioned into the hydrophobic cluster within the other monomer, encompassing the side chains of Val32’, Ile47’, Ile54’, Val56’, Pro79’, Pro81’ and Val82’, thus making close van der Waals (vdW) contacts with these hydrophobic residues, referred to as the ‘intermonomer hydrophobic cluster’ in this context.

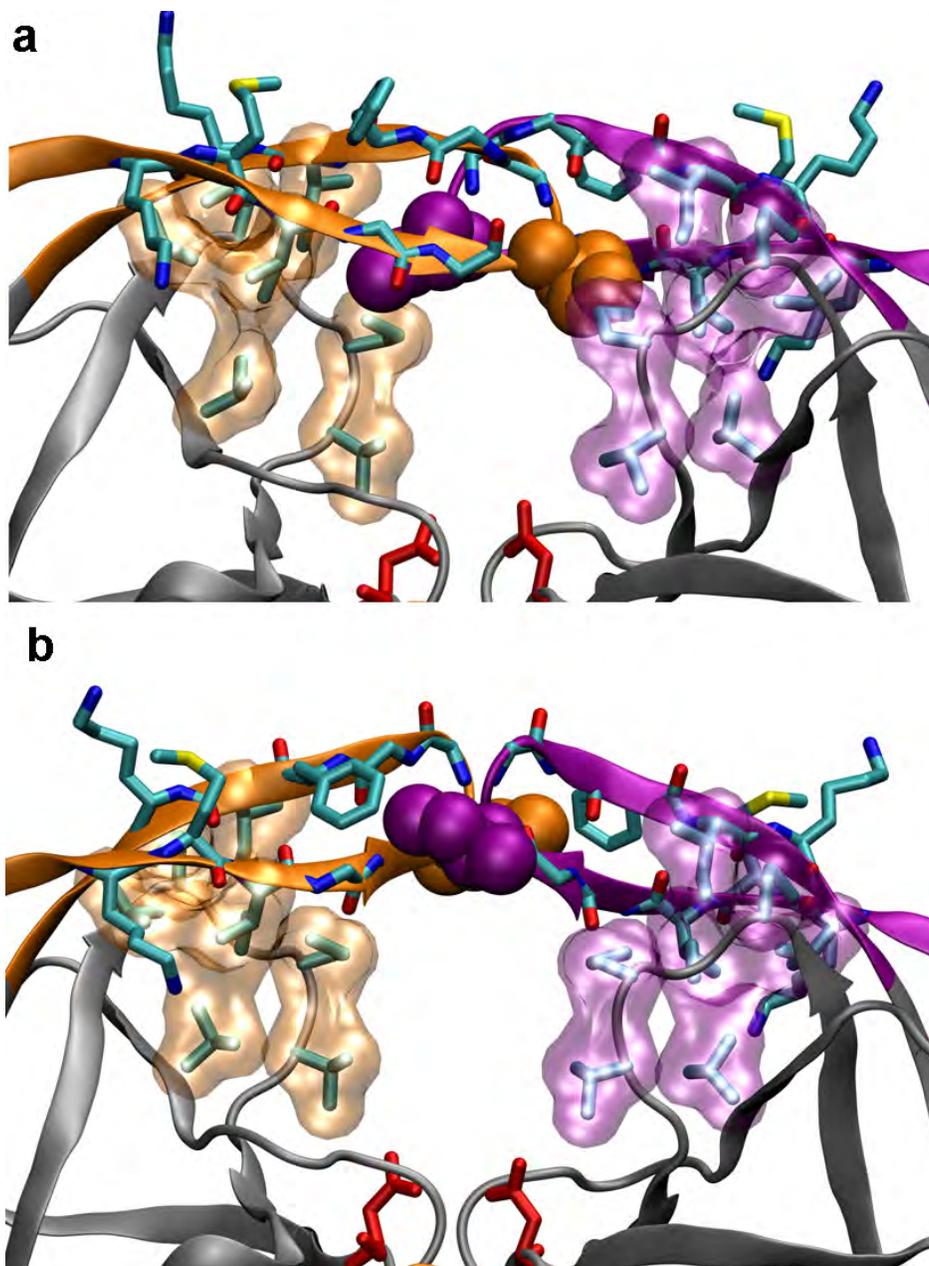


Figure 2-1. a) The closed conformation (PDB code: 1TSU, with the substrate stripped out). b) The semi-open flap conformation (PDB code: 1HHP). The flap residues (44-55) are rendered by 'Licorice' in VMD. The flap tip residue Ile50 is rendered as 'VDW' and colored in orange in monomer A, in purple in monomer B. Asn25/25' at the active site are rendered by licorice and colored in red. For clarity, only the side chains of hydrophobic residues on the loop region of each monomer are shown, including Val32/32', Ile47/47', Ile54/54', Val56/56', Pro79/79', Pro81/81' and Val82/82', and rendered as both 'Licorice' and 'Surf' (residues on monomer A are colored in orange; residues on monomer B are colored in purple). It is worth noting that the relative orientation of the two flaps (the handednesses) is switched in the bound and unbound form.

While in the semi-open structure (Figure 2-1b), the two flaps are no more than 7.7 Å apart, and no hydrogen bond exists between the tips of the two flaps. Both flaps are pulled up and shifted away from the active site, but still substantially cover the binding pocket. It is worth noting that relative orientation (the handedness) of the two β-hairpin flaps is reversed between the closed and semi-open states. In contrast to the closed state, the orientation of the flap tips in the semi-open form leads to the proximity of Ile50 to the residues on the 80s loop (residues 79-81) and Val32, Ile47, Ile54 and Val56 from the same monomer, thus forming van der Waals interactions with these hydrophobic residues, termed as the ‘intramonomer hydrophobic cluster’ in this context. Notably, the side chains of Ile50 and Phe53’ from the opposite flap form a stabilizing contact. In addition, the aromatic ring of Phe53 is positioned above the opposite strand of the same flap, likely forming aromatic-amide and CH-π interactions with the backbone amide and the H-Cα groups of Gly48/Gly49. Both types of weakly polar interactions might also contribute to the stability of the local structure.

It is worth noting that the crystal packing contacts may also contribute to the overall stability of the semi-open form, as suggested by earlier studies.^{86, 105, 203} Figure 2-2 illustrates the crystal packing interactions around the flap region in the semi-open X-ray structures. It is evident that the side chains of residues Gln61’ and Gln92’ from a monomer in a neighboring dimer hydrogen bond with the backbones of the Ile50 and Lys54 residues on the flap in the central dimer, respectively. Therefore, it is possible that the unbound enzyme might adopt a different conformation in the absence of the crystal environment.

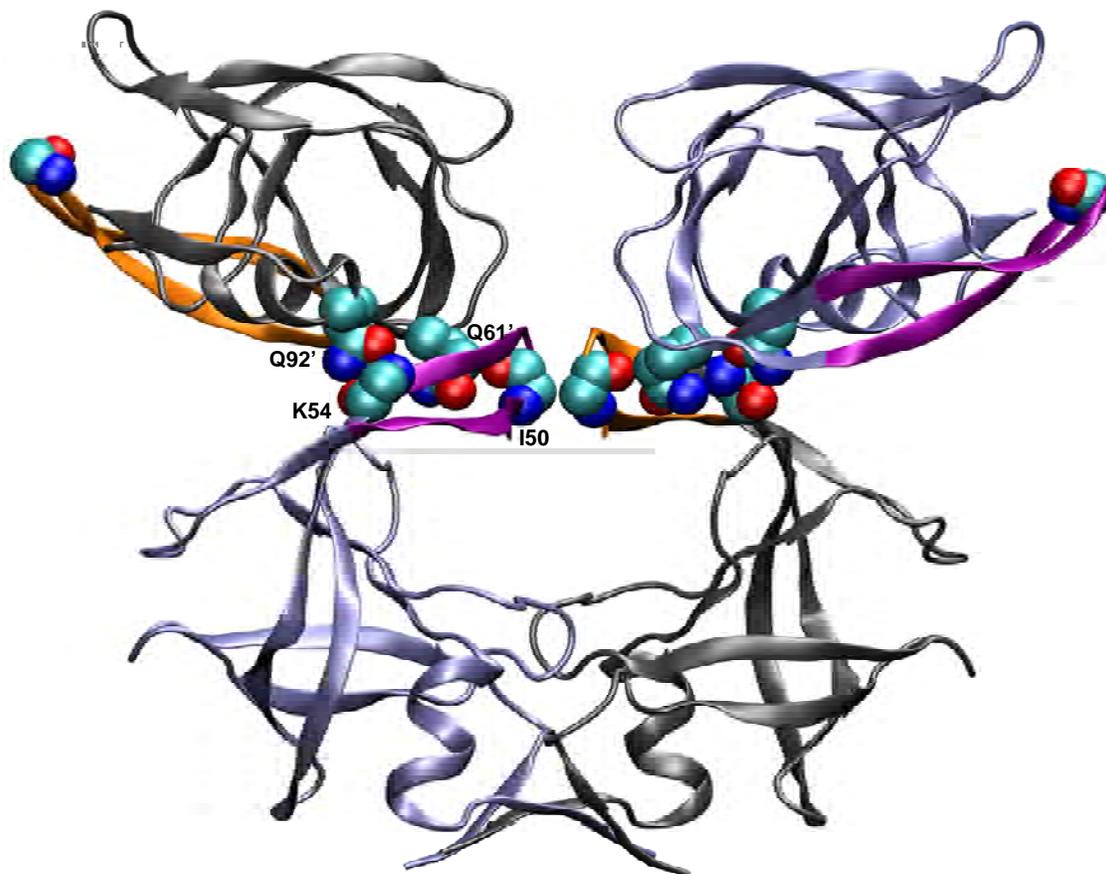


Figure 2-2. A detailed view of the ‘semi-open’ (PDB code 1HHP) crystal packing interactions around the flap region. The backbones of Ile50 and Lys54 (colored by atom type) on the flap of the central dimer hydrogen bond with the side chains of residues Gln61 and Gln92 (colored by atom type) from a neighboring dimer.

2.3.2 Temperature Dependence of Atomic Fluctuations

To explore the conformational space available to the protease in the simulations, MD simulations were performed at an elevated temperature (375K). We note that there may be significant differences between the high-temperature and physiological temperature free energy landscapes,²⁰⁴⁻²⁰⁵ in many cases, however, the nature of protein transitions

appears to be largely temperature independent.²⁰⁶⁻²⁰⁷ A comparison of the C α atomic fluctuations from the MD simulations at two different temperatures (Figure 2-3), 300K and 375K, further confirms that the dependence of the magnitude of fluctuations on the temperature is relatively small.

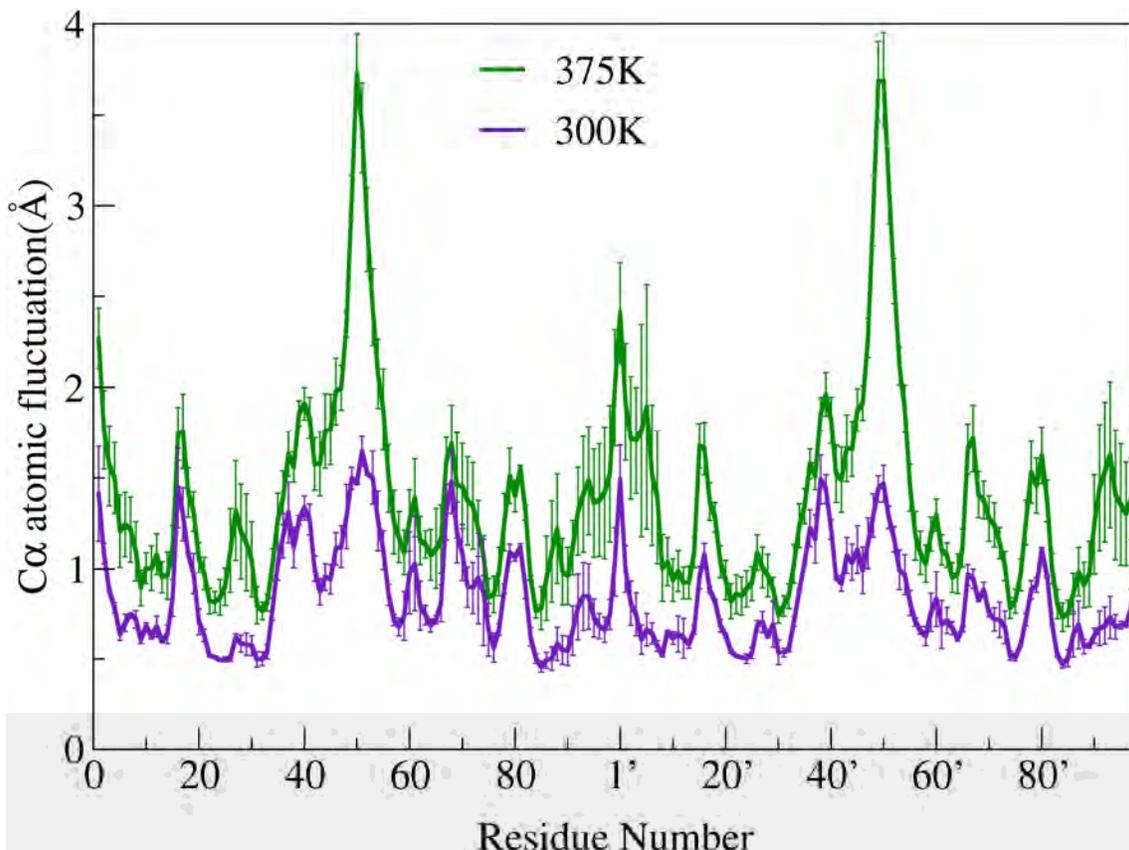


Figure 2-3. Comparison of the atomic fluctuations from simulations at 300K (purple line) and 375K (green line). Error bars reflect the difference between the two runs at the same temperature, starting from different conformations, i.e., closed and semi-open forms.

Although the magnitude of fluctuations was much higher at 375K, the overall protein dynamics was not greatly perturbed by the high temperature. At both temperatures, high atomic fluctuations occurred in the flap (residues 45–55) region, the N, C-termini region(residues 1-4, 96-99), the flap elbow (residues 37-42), the cantilever β -turn region

(residues 66-69) and short regions around Gly17 and Thr80. Whereas other regions in HIV-1 PR only exhibited small fluctuations ($< 1\text{\AA}$). Thus, this comparison suggests that the transitions observed in this work may reflect the actual dynamics of the flaps, and are not indicative of any instability of the system caused by the high simulation temperature.

However, we note that the application of the high temperature could lead to under-sampling of the low energy subsites since the high temperature enhances sampling near the transition state by increasing energies everywhere else. Thus, the conformational subsites sampled in the high temperature MD simulation may be different from those at a low temperature, as well as the distributions of these subsites. As a result, the use of high temperature precludes us from exploring the energy difference among various conformations at low biological temperature, associated with drug resistance.¹³¹

2.3.3 Flap Conformations Observed During the Trajectories

To monitor the conformation of the flaps in the molecular dynamics structures, we calculated flap $C\alpha$ RMSD by superimposing the simulated system with the two crystal structures, the closed (PDB ID code 1TSU) and the semi-open structure (PDB ID code 1HHP), by fitting the $C\alpha$ atoms of residues 46-55 of both monomers. This metric has been employed in our previous implicit solvent simulations.⁴⁵ We also monitored $C\alpha$ RMSD of the non-flap region between the molecular dynamics structure and the semi-open crystal structure. The results from the two simulations starting from different structures are presented in Figures 2-4 and 2-5, respectively.

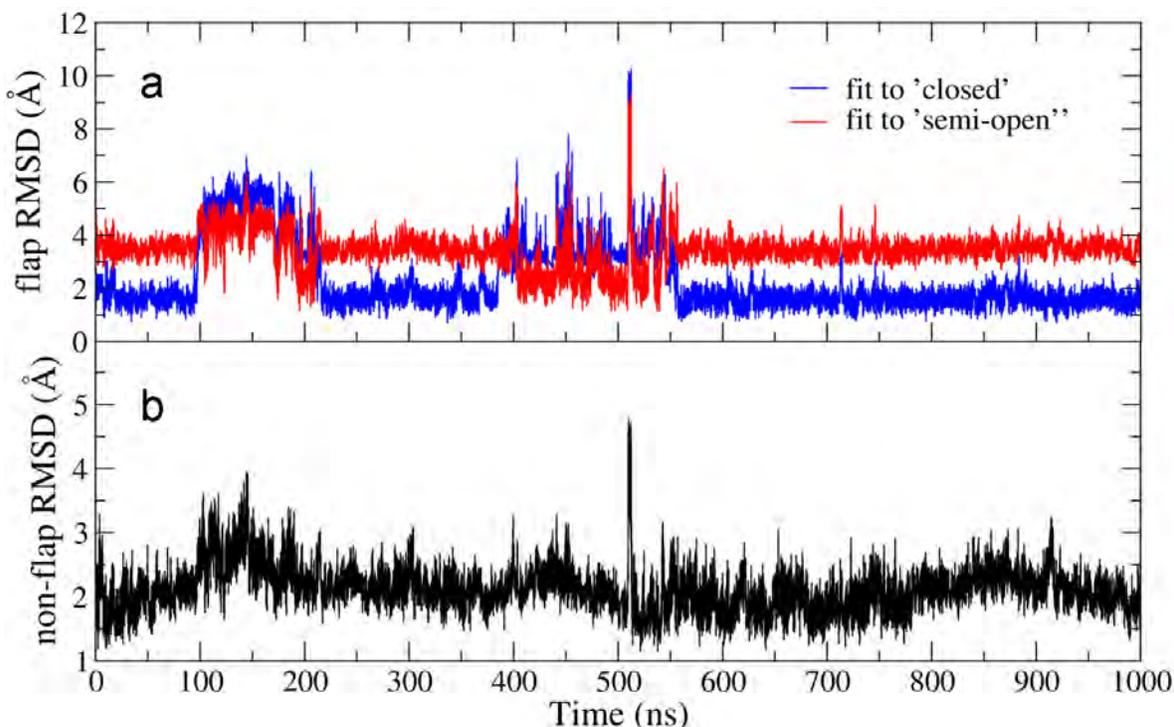


Figure 2-4. a) Time evolution of C α RMSD of the flap region. b) Time evolution of C α RMSD of the non-flap region with respect to the semi-open crystal structure.

As seen in Figure 2-4a, in the MD_{closed} simulation, the flaps underwent significant fluctuations during two periods, 97ns-216ns and 385ns-558ns. The initial large deviation from the original closed structure was observed after \sim 97ns. The two flaps then rearranged to the semi-open state at \sim 104ns prior to sampling a flexible ensemble of structures that are neither semi-open nor closed with both flap-C α RMSDs being $> 4\text{\AA}$ to the two references. (The structural features of this novel conformation will be discussed in detail in the following section). The semi-open flap conformation, however, was still sampled multiple times thereafter before the two flaps returned to the closed conformation (at \sim 216ns). Over the course of the second transitional period (385ns-558ns), the flaps converted from the closed state to a more flexible ensemble with the dominance of the semi-open form, as indicated by consistently low values of flap-C α

RMSD with respect to the semi-open reference (red line in Figure 2-4a). Significantly, the high flexibility of the semi-open conformation produced a full flap opening around 509ns, as manifested by flap-C α RMSD values to both references being $> 8 \text{ \AA}$. The flaps reverted back to the semi-open conformation after the transient opening event, and eventually returned to the closed conformation after 558ns. No large deviations were observed afterwards till the end of this one μsec simulation.

Whereas non-flap regions did not experience substantial changes (Figure 2-4b), with non-flap-C α RMSD of $\sim 2 \text{ \AA}$ even during the transitional periods; yet with the only exception of the wide-opening event (509ns~512ns), increasing up to 5 \AA , indicating that certain residues outside the flaps are involved in this large-scale conformational changes. (The mechanism underlying this opening event will be discussed in detail in Chapter 3.)

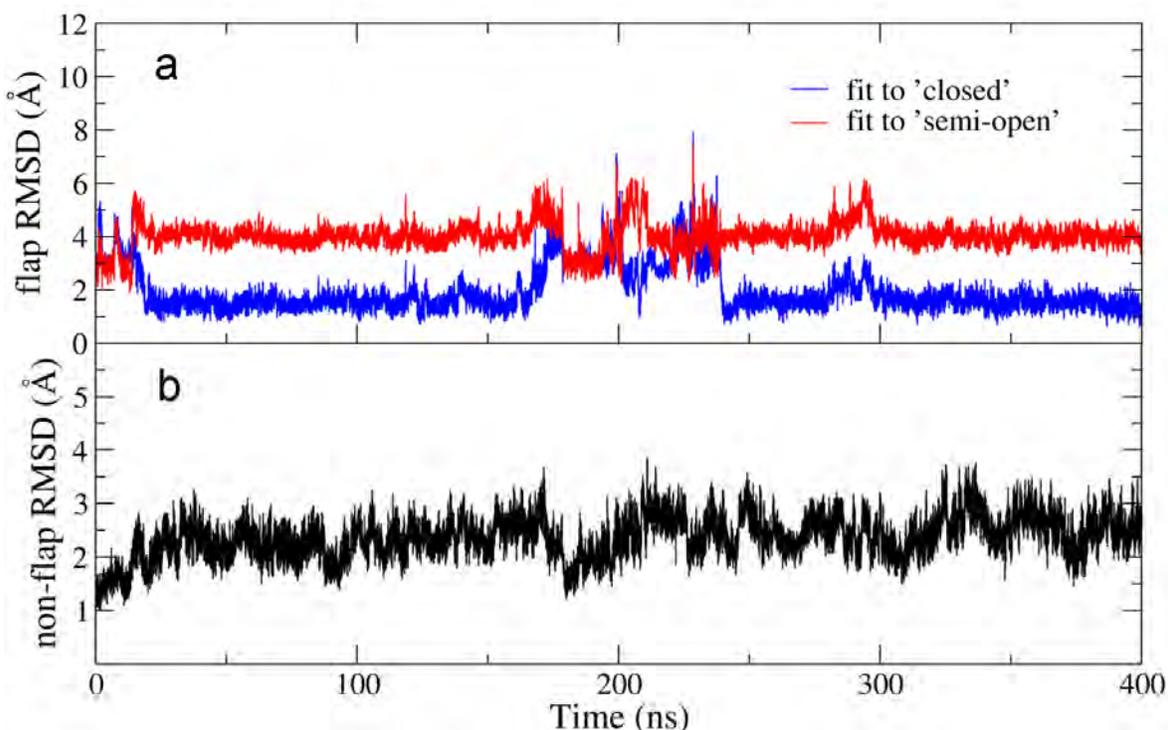


Figure 2-5. Time evolution of C α RMSD of the flap region (a) and non-flap regions during the MDsemi-open simulation. the flap RMSD are colored using the same color code as in Figure 2-4a.

As observed in the MD_{closed} simulation, the two flaps also exhibited substantial flexibility over the course of the simulation starting from the semi-open state (Figure 2-5a). The tips of the two flaps separated at the early stage (at ~20ns). Then the two flaps spontaneously rearranged to the closed form and maintained this structure for a long period of time (~160ns). Nonetheless, the semi-open state was still revisited not only from 178ns~202ns, but also from 221ns~237ns. Eventually, the flaps returned to the closed state after 230ns and stayed in this structure until the end of the simulation. Consistently, non-flap regions did not exhibit large conformational changes (Figure 2-5b), with non-flap-C α RMSD of ~2.5 Å along the entire simulation.

It is worth mentioning that in our previous MD simulations of apo HIV-1 PR by using a continuum generalized Born (GB) model, reversible and multiple conformational changes were also captured. The flaps, however, underwent much more rapid structural rearrangements, transitioning from the closed to semi-open form near the beginning of the simulation and reaching large-scale flap openings after 27ns. In addition, the ensemble of unbound structures was dominated by the semi-open conformation in the GB simulations. We realize that the greater plasticity of the flaps in the implicit solvent simulations may be due to the lack of friction from the water molecules. Additionally, the preference of the semi-open flap conformation in the GB simulations may be owing to the lack of the description of the hydrophobicity in implicit solvent model, which plays a significant role in the stability of HIV-1 PR,²⁰⁸⁻²⁰⁹ due to the fact that nearly 45% of the amino acid residues of HIV-1 protease are hydrophobic.

Nonetheless, please note that the free energy difference between the closed and semi-open states could be as little as < 1kcal/mol, as estimated by the potential of mean force

(PMF) profiles obtained in the explicit and implicit solvent simulations, respectively (Figure 2-6). In addition, a previous calculation of free energy along the reaction path connecting the semi-open and the closed conformations¹⁰⁶ also predicts that the semi-open conformation is approximately 7kcal/mol more stable than the closed conformation, and that entropic freedom of the semi-open state in solution is responsible for the free energy difference. Based on both results, it can be inferred that the errors between the two solvent models should be relatively small. Thus, although the quantitative disagreement exists between the implicit and explicit solvent simulations, both types of simulations provide solid evidence that the true ensemble of apo HIV-PR might cover the complete structural heterogeneity as observed in both crystal¹¹⁰ and MD simulations.^{45, 123}

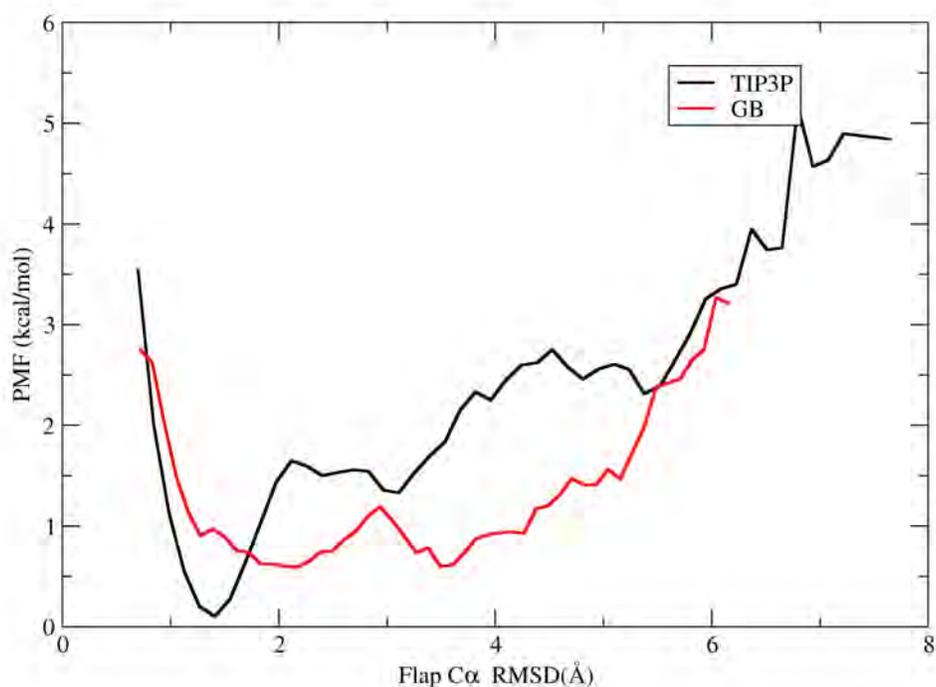


Figure 2-6. The PMF profiles from the explicit solvent (black line) and the implicit solvent (red line) simulations, as a function of Flap C α RMSD with respect to the closed X-ray structure. Both simulations are started from the closed form (PDB code: 1TSU, without the substrate).

2.3.4 Local Dynamics of the Flap Tips

To analyze the local dynamics of the flap tips, we calculated the intra-flap hydrogen bonds as well as the backbone dihedral angles of the flap tips, Gly48, Gly49, Ile50, Gly51 and Gly52. As shown in Figures 2-7 and 2-8, in both simulations, the flap β -hairpins were well maintained, and the intra-flap hydrogen bonds were always present with the exception of the tip hydrogen bonds between Gly49 and Gly52, respectively.

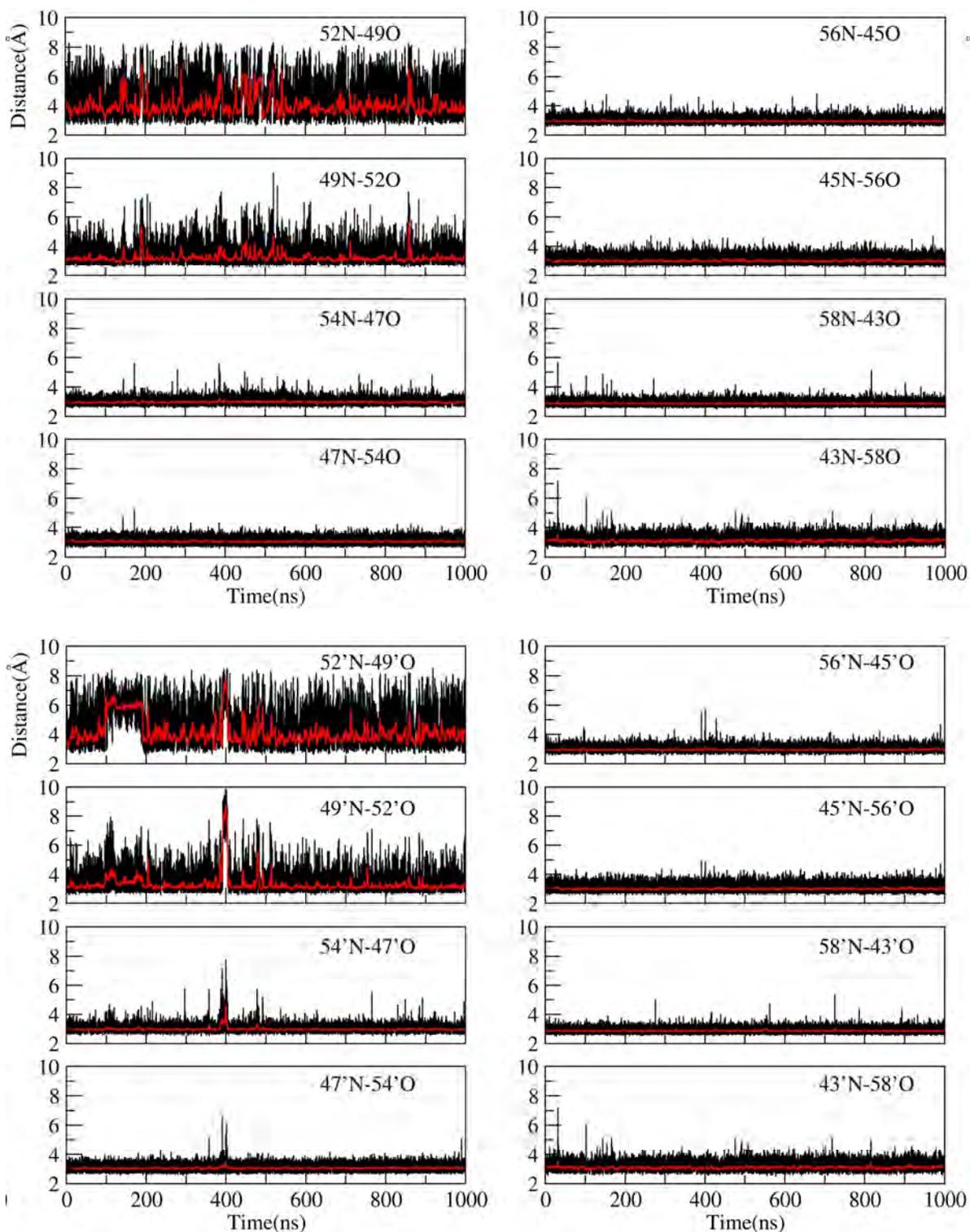


Figure 2-7. Time evolution of the intraflap hydrogen bonds within flapA (top) and flapB in the MDclosed simulation. Most hydrogen bonds are stable, and thus result in a stable β -hairpin structure with the exception of the tips (residues 49-52). Running averages over 100 data points (red line) are also shown.

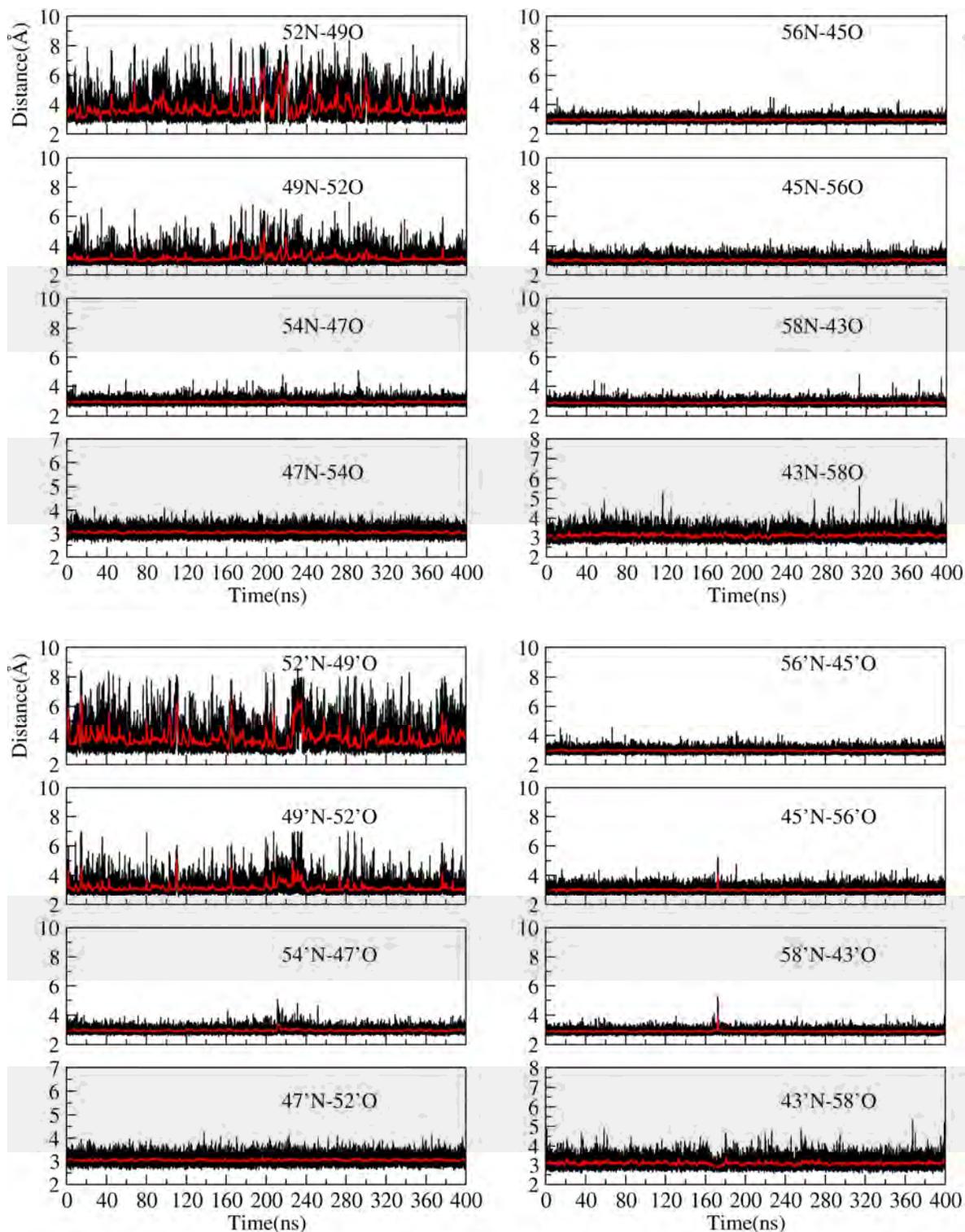


Figure 2-8. Time evolution of the intraflap hydrogen bonds within flapA (top) and flapB in the MDsemi-open simulation. Consistent with MDclosed simulation, most hydrogen bonds are stable except the tip hydrogen bonds between Gly49 and Gly52. Running averages over 100 data points (red line) are also shown.

The above observation is in consistent with NMR data^{28,110} which indicate that large amplitude angular fluctuations of flap residues are restricted to residues 49-53 at the tips of the flaps. The high flexibility of the tips of the flaps was also manifested by large fluctuations of the backbone torsions of the flap tip residues in both simulations, as shown in Figures 2-9 and 2-10.

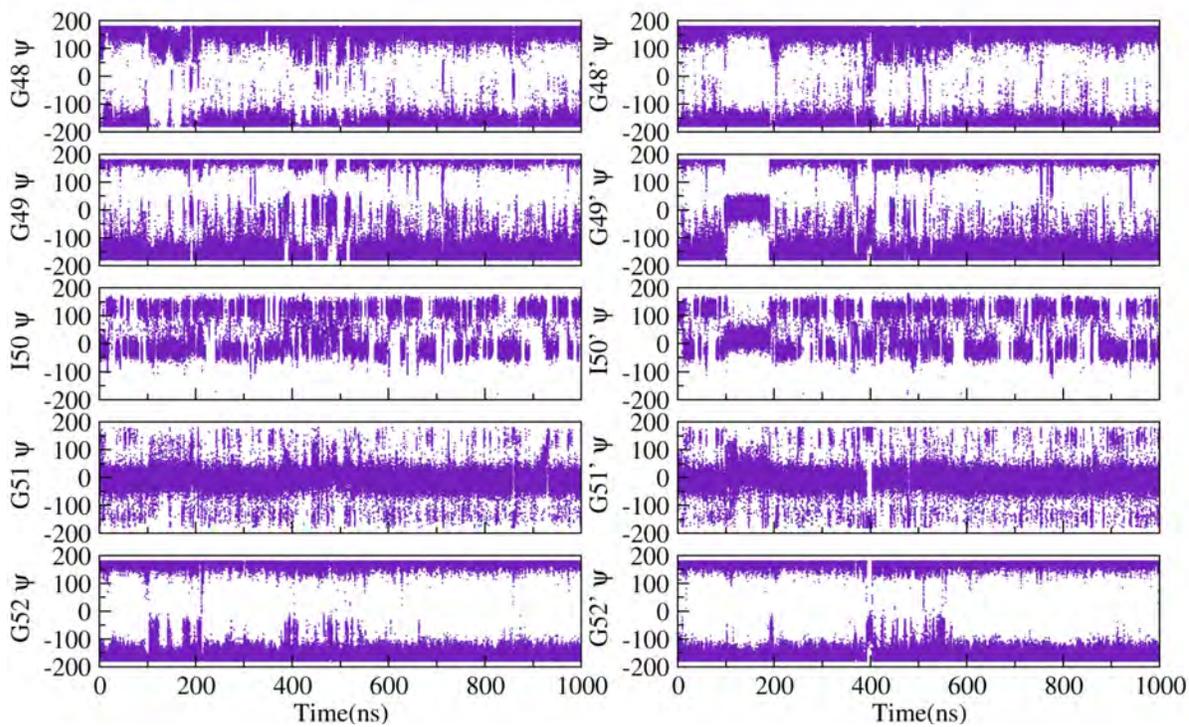
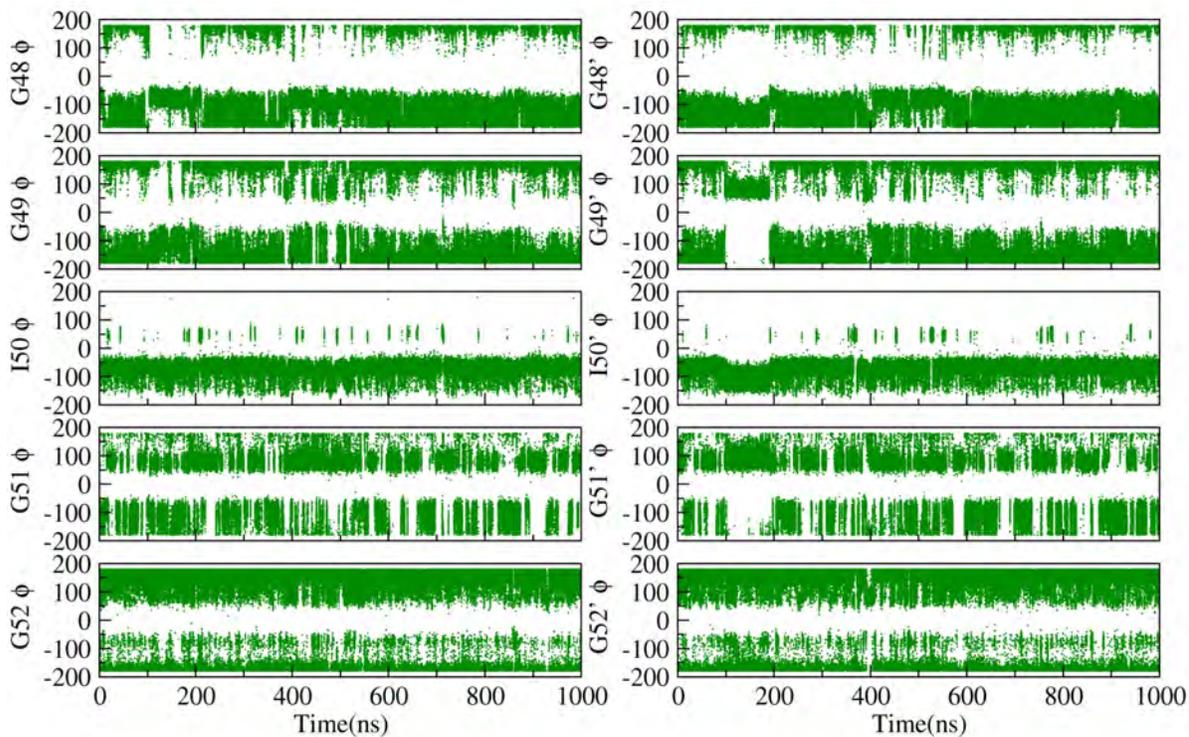


Figure 2-9. Backbone dihedral angles (Φ , green; Ψ purple) of the flap tip residues (G48-G52) during the MDclosed simulation

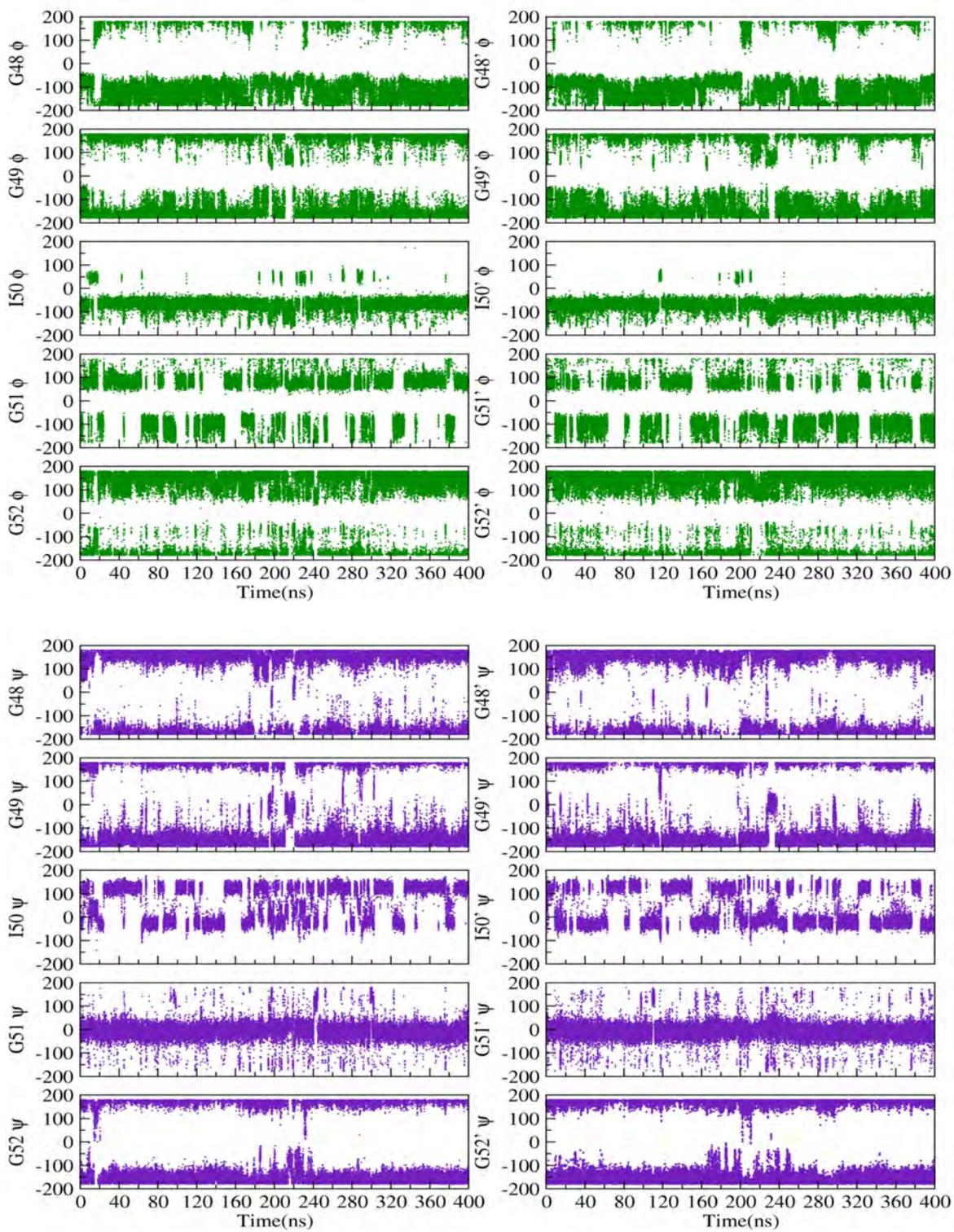


Figure 2-10. Backbone dihedral angles (Φ , green; Ψ purple) of the flap tip residues (G48-G52) during the MDsemiopen simulation.

2.4 Conclusions

2.4.1 Twisting of the Flap Tips Initiating the Rearrangements of the Flaps

A careful inspection of the torsion angles of the flap tip residues during each transitional period in the MD simulations (Figure 2-11) revealed that the backbone torsions of the flap tips residues underwent substantial changes preceding each rearrangement between the closed and semi-open form, particularly Gly49 and Ile50. As illustrated in Figure 2-11 (left), just before the initial rearrangement from the closed to semi-open form in the MD_{closed} simulation, the Φ angle of Gly49 rotated from -180° to -130° at ~ 96 ns, orienting the Gly48-Gly49 peptide bond normal to the antiparallel β -strand of flapA, thereby breaking the intra-flap hydrogen bonds between Gly49 and Gly52 (Figure 2-7, top). Simultaneously, the Ψ angle of Ile50 rotated abruptly from 120° to -10° , displacing its sidechain from being buried into the hydrophobic cluster of monomer B, as seen in the initial closed conformation. It appears that the disruption of the intermonomer hydrophobic interactions between the flap tip Ile50 and the hydrophobic cluster within monomer B greatly enhanced the flexibility of the flaps, initiating the rearrangement from the closed to semi-open state to take place.

Reversibly, when the semi-open flap conformation converted back to the closed form near the end of this transitional period (at ~ 215 ns), the backbone torsions of the flap tip residues were restored to the angles in the closed flap conformation.

Consistent with the observation for the first transition, the backbone of the tip of flapB underwent significant twisting, initiating the second transition from the closed to

the semi-open state in MDclosed simulation. As shown in Figure 2-11 (right), prior to this transition event (at ~385ns), the Φ angle of Gly49' rotated abruptly from -150° to -60° together with the rotation of the Ψ angle of Ile50' by 130° , thereby inverting its side chain from the hydrophobic cluster within monomer A. Again, the backbone torsions of these two flap tip residues were restored before the semi-open form reverted back to the closed form near the end of this transitional period.

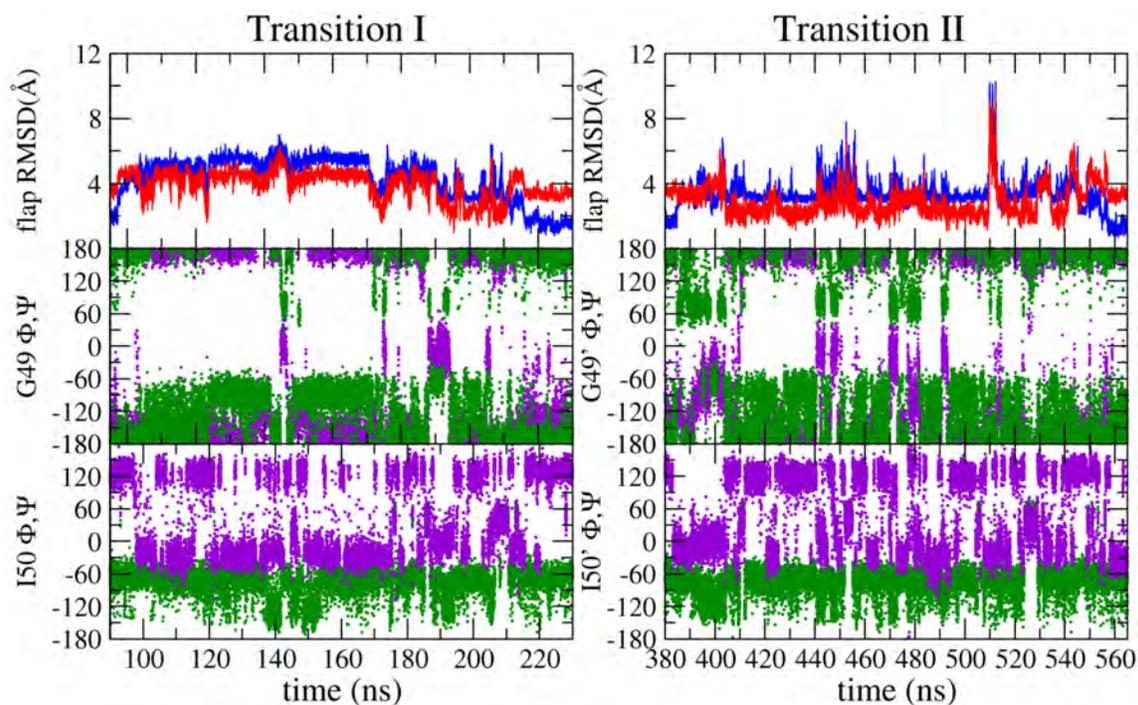


Figure 2-11. Time evolution of flap C α RMSD with respect to the closed (blue) and semi-open (red) references (top); the torsion angles (Φ , green; Ψ purple) of Gly49/49' and Ile50/50' during the first (left) and second (right) transitional periods, respectively in MDclosed simulation.

We subsequently investigated the coupling between the twisting of the backbone of the flap tips with the flap rearrangements in MDsemi-open simulation. As illustrated in

Figure 2-12, the transitions between the closed and semi-open forms were also preceded by the rotations of Φ , Ψ angles of residues Gly49 and Ile50.

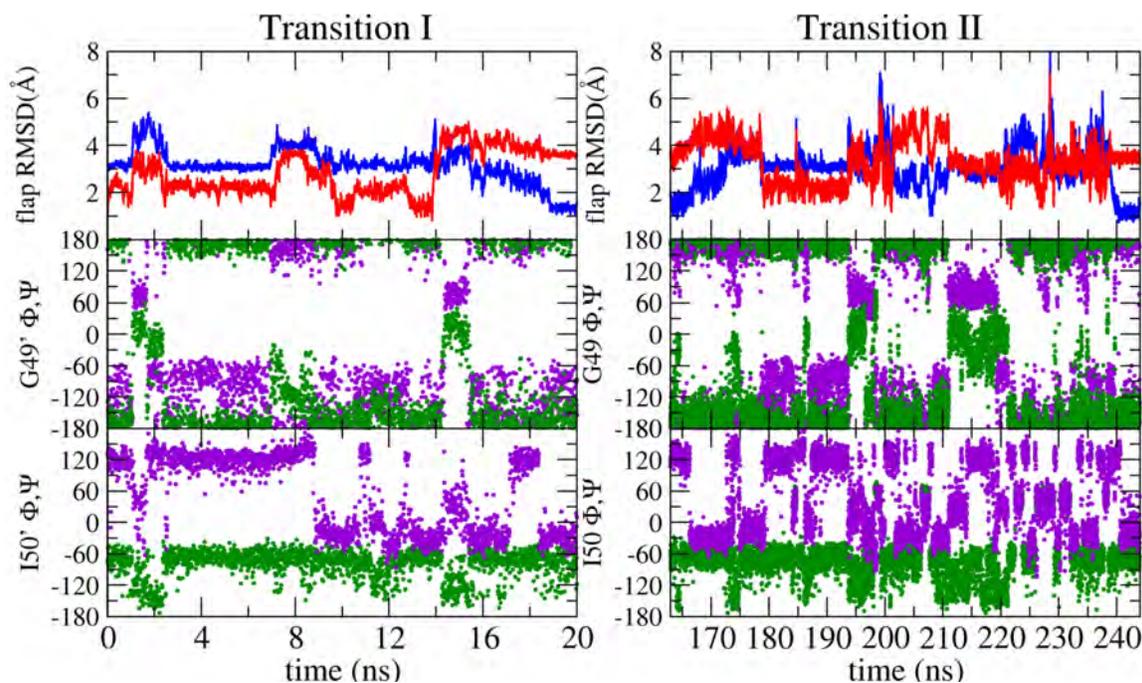


Figure 2-12. Time evolution of flap C α RMSD (top) with respect to the closed (blue) and semi-open (red) references; the torsion angles (Φ , green; Ψ purple) of Gly49'/49 and Ile50'/50 during the first (left) and second (right) transitional periods in MDsemi simulation.

From the above analyses, we suggest that the rearrangements of the flaps between the closed and the semi-open states are most likely induced by the Φ - Ψ rotations of the residues at the tips of the flaps, Gly48-Gly49-Ile50-Gly51-Gly52; the rotation of its backbone torsions flips the carbonyl of Gly49 out of the plane of the flap, which in turn breaks the intraflap hydrogen bonds between Gly49 and Gly52. The subsequent disruption of hydrophobic interactions between the side chain of the flap tip residue Ile50 and the hydrophobic cluster within each monomer greatly enhances flexibility of both

flaps, thus causing the rearrangements of the flaps to occur. Our proposed mechanism agrees well with a recent MD simulation of an unbound HIV-1 PR.¹⁹⁰

It is also worth mentioning that three of the four glycine residues in the flap (Gly49, 51 and 52) adopt conformations that would exclude a larger sidechain from existing at these positions. These three residues are completely invariant²¹⁰ and cannot be substituted in vitro without compromising activity.²⁰² Based on the simulations, we hypothesize that Gly49 is at a key hinge point in the flap movements, as reflected in its motions in Ramachandran space, which assumed conformations unfavorable for non-Gly residues in both simulations (Figures 2-14 and 2-15).

The fourth glycine, Gly-48, confers drug resistance when it mutates to a valine, precluding the binding of particular inhibitors through steric hindrance²¹¹. However, in vitro mutagenesis shows that a variety of substitutions, including polar and basic amino acids, gave full activity.^{202, 212} This is consistent with the conformations sampled in our MD simulation as this glycine remained in allowed Ramachandran space throughout both trajectories (Figures 2-13, 2-14), which is necessary for any other amino acid residue to be substituted at this position.

The central Ile-50 rarely mutates with the currently used inhibitors, but earlier drug trials showed this residue mutating to a valine, which conferred resistance²¹⁰. Thus, while a valine substitution would reduce some of the internal packing, it would not preclude continued interaction with substrate, consistent with the near wild-type level of activity that was observed with the Ile-50 → Val mutant. In vitro, activity is maintained in the enzyme with either leucine or valine at position 50.²⁰² These two residues have sidechains similar to Ile in terms of size and hydrophobicity, and thus they may likely maintain the

ability of the flap region, allowing for conformational changes to take place. The significance of Ile50 residue will be further explored in the next section.

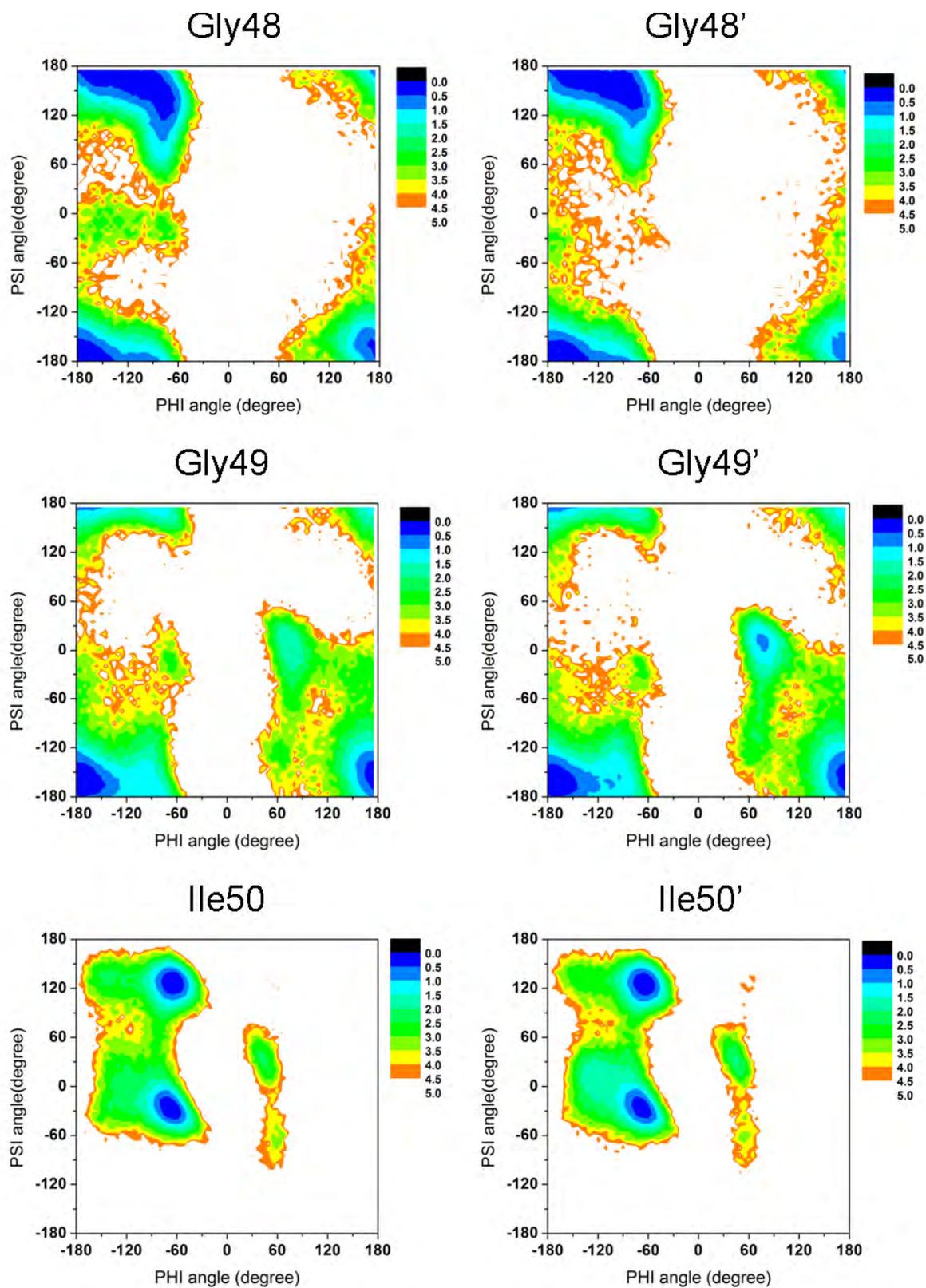


Figure 2-13. Ramachandran plots of residues Gly48, Gly49 and Ile50 on both flaps in MDclosed simulation.

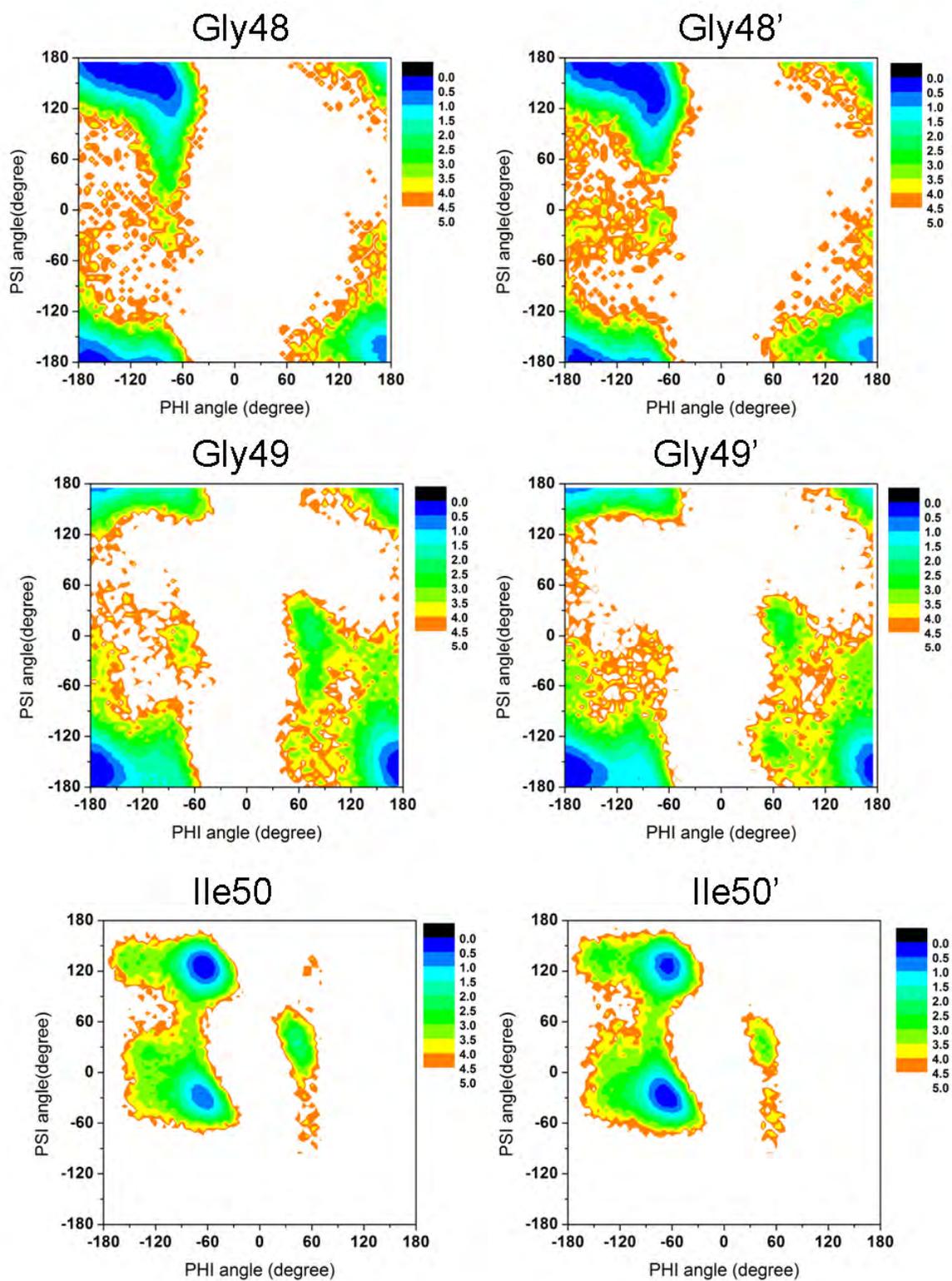


Figure 2-14. Ramachandran plots of residues Gly48, Gly49 and Ile50 on both flaps in MDsemi-open simulation

2.4.2 Swapping Ile50 between the Hydrophobic Clusters

As revealed by the above structural analysis (Figure 2-1), the flap tip Ile50 residue is packed into different hydrophobic clusters in the closed and semi-open conformations. Thus, the rearrangements of the flaps between these two forms should also involve swapping this flap tip residue between the two types of hydrophobic clusters within each monomer. To test our hypothesis, we calculated the radius of gyration (Rg) of the two types of hydrophobic clusters in each monomer, defined by the side chains of Ile50 and Val32/32', Ile47/47', Ile54/54', Val56/56', Pro79/79', Pro81/81' and Val82/82', from the same or its symmetry-related monomer. In addition, the curling motions of the flaps were monitored by the distance between the center of mass (COM) of five central residues (residues 48-52) of each flap relative to the COM of all heavy atoms of the two asparagines at the active site (Asn25/25'), termed as flapA-Asn25/25' and flapB-Asn25/25', respectively. The results during the two major transitional periods in MD_{closed} simulation are presented in Figures 2-15 and 2-16, respectively, with their representative snapshots shown on top.

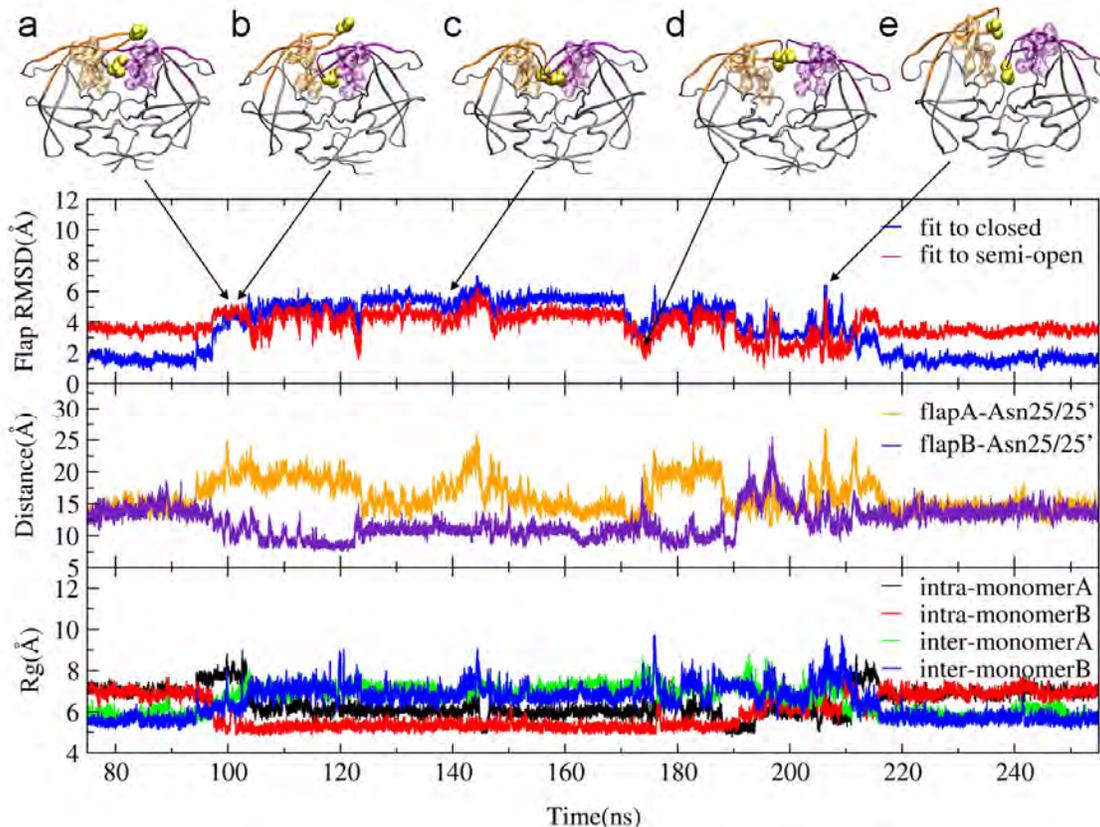


Figure 2-15. Time evolution of flap RMSD (top), distance between the COM of residues 48-52 relative to Asn25/25' (middle) and Rg of intra- and inter- monomer hydrophobic clusters (bottom) formed between Ile50 and hydrophobic residues Val32, Ile47, Ile54, Val56, Pro79, Pro81 and Val82 from the same or its symmetry-related monomer in the course of the first transition in the free HIV-PR simulation started from a closed crystal structure. Representative snapshots during this first transition are shown on top. The backbone of HIV-1 PR is shown in NewCartoon representation with flapA colored in orange and flapB in purple. Residues Val32, Ile50, Ile47, Ile54, Val56, Ile79, Pro81 and V82 are shown in Surface and colored in orange in monomer A and colored in purple in monomer B.

As shown in Figure 2-15, at ~94ns, the intermonomer hydrophobic cluster involving Ile50 was disrupted (bottom panel, blue lines), the tips of the two flaps were thereby separated. At ~ 100ns the tip of flapA curled up and simultaneously slid to the other side of flapB, thus switching the handedness of the two flaps (snapshot a, flap in orange). Meanwhile, the high degree of mobility of the flaps facilitated the position of Ile50' at the

tip to pack into a tight hydrophobic cluster within the same monomer (snapshot b, flap in purple), as indicated by the low values of R_g of the intramonomer (bottom panel, red lines). At 103ns, flapA slid down with its tip packed into the hydrophobic cluster from the same monomer (bottom panel, red lines); thus, the system reached a flexible ensemble that was dominated by the semi-open state, as evidenced by the flap RMSD values to the semi-open reference of $\sim 2 \text{ \AA}$ (top panel, red lines).

It is noticed that at approximately 124ns, flapA curled further down towards the active site (middle panel, orange lines); the system then sampled an ensemble of stable conformations with flap- $C\alpha$ RMSD values to both references being $> 4 \text{ \AA}$, termed as ‘tucked’ conformation. This conformation is unique, in which the tips of both flaps curl in towards the active site cavity, and thus bury themselves into the respective hydrophobic cluster within the same monomer with the active site still completely covered (Figure 2-15, snapshot c). Although there is no direct experimental evidence to prove whether the enzyme actually exists in the ‘tucked’ conformation predicted by our simulation, a recent electron paramagnetic resonance (EPR) study on apo HIV-1 PR by Fanucci et al.²¹³ also characterized a novel conformation with the average flap distances of $25\text{--}30 \text{ \AA}$, corresponding well with the conformation observed here. Thus, this EPR study may provide evidence that the ligand-free enzyme may assume such a compact conformation with the hydrophobic residues in the core. We realize that further exploration is needed to elucidate the biological significance of the ‘tucked’ conformation and its potential for future drug design.

Notably, no large scale opening was reached afterwards despite the large amplitude curling motions of the flaps, as predicted before by a 10ns solvated MD simulation of the

unbound HIV-1 protease.¹¹⁵ Instead, the system stayed in this novel ‘curled’ confirmation for >50ns. Thus, we hypothesize that curling seems to be primarily the consequence of inherent flap tip flexibility rather than a trigger for flap opening.

Both tightly packed intramonomer hydrophobic clusters disintegrated after 174ns, and the system reached again an ensemble of semi-open conformations (snapshot d) before reverting back to the closed state after 216ns. A snapshot depicting an intermediate state from the semi-open to the closed state is shown in Figure 2-15, snapshot e. As mentioned earlier, the rearrangement from the semi-open to the closed was also induced by the Φ - Ψ rotation of Gly49 and Ile50, facilitating the packing of the flap tip Ile50 residue into the hydrophobic cluster within the symmetry-related monomer. The formation of the intermonomer hydrophobic cluster within each monomer was demonstrated by their respective low values of R_g (bottom panel, green and blue lines).

Consistent with the first transition period, over the course of the second transitional period (Figure 2-16), the changes in flap conformations were coupled to the translocation of Ile50 between the hydrophobic clusters within each monomer. The intermonomer hydrophobic cluster involving the flap tip Ile50’ residue was disrupted at 385ns (bottom panel, green lines). FlapA pulled up at 392ns (snapshot a), and thus separated from the other flap. Thereafter both flaps exhibited a high degree of mobility (snapshots b-c), as illustrated by large fluctuations in both flapA-Asn25/25’ and flapB-Asn25/25’ distances (middle panel). Significantly, during this transitional period, the system reached a flexible ensemble dominated by the semi-open conformation, and even sampled a transient large-scale opening event (snapshot d). As expected, the fully open structure was not stable and converted to the semi-open form (snapshot e). After 549ns, the inherent mobility of the

flaps positioned Ile50 to approach residues 32 and 79-81 on the symmetry-related monomer, thus packing it into the intermonomer hydrophobic cluster within each monomer, thereby returning to the closed conformation.

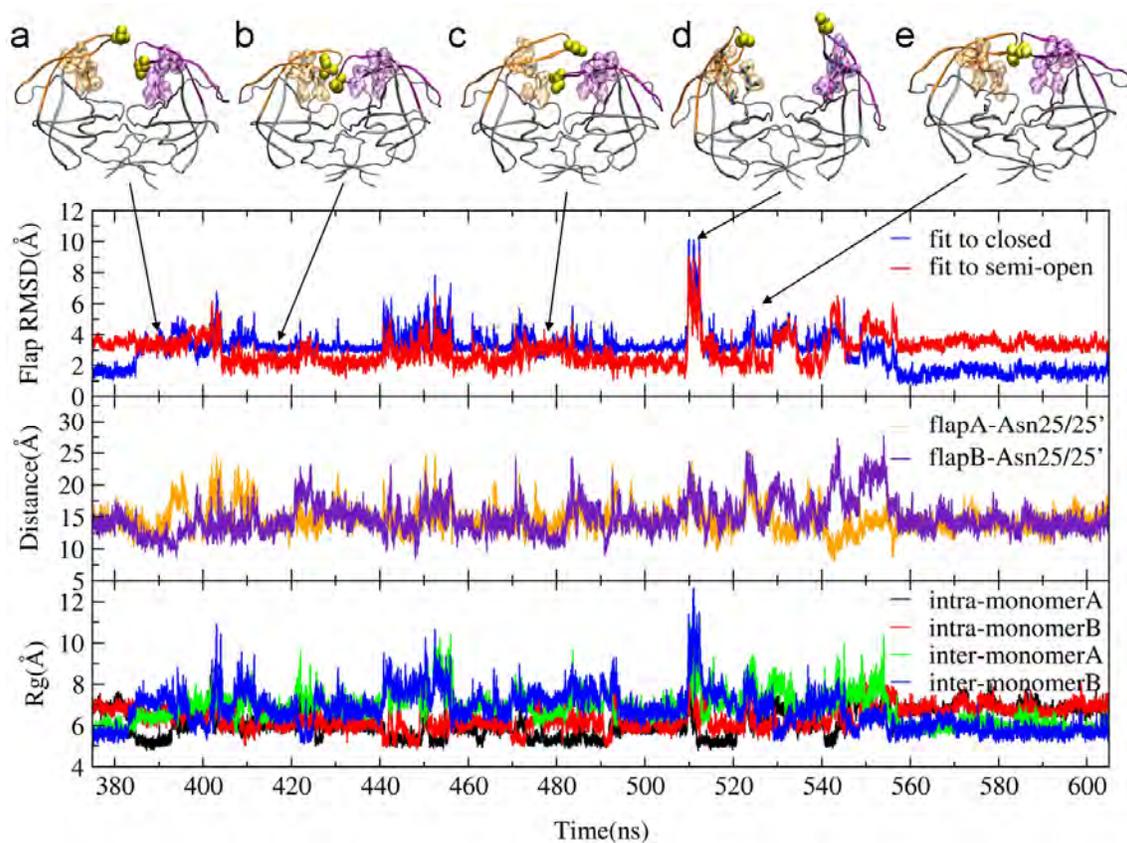


Figure 2-16. Time evolution of flap RMSD (top), distance between the COM of residues 48-52 relative to Asn25/25' (middle) and Rg of intra- and inter- monomer hydrophobic clusters in both monomers (bottom) in the course of the second transition in the free HIV-PR simulation started from a closed crystal structure. Representative snapshots during this first transition are shown on top.

2.4.3 Significance of the Flap Tip Ile50 Residue in the Transition Dynamics

To further clarify that the translocation of Ile50 residue is a key determinant of the protein dynamics rather than a measurement of the motion accompanied with the

transition between the two crystal structures, we carried out two additional short MD simulations with Ile50 substituted with two different types of mutations: I50A, with a shorter and less hydrophobic side chain than Ile, which would enhance the flexibility of the flaps, as well as accelerate the transition rate between the closed and semi-open forms; and I50W, with a bulky and more hydrophobic side chain, which are attempted to form stronger van der Waals interactions within the intermonomer hydrophobic cluster.

As expected, the I50A single mutant system exhibited higher flap flexibility and distinct dynamics from the wild-type (PR_{D25N}). The flaps converted from the initial closed form to a semi-open form around 13.5ns, ~10 times faster than the initial conversion captured in the wild-type simulation, and then sampled a broad range of opening events prior to return to the semi-open form (Figure 2-17).

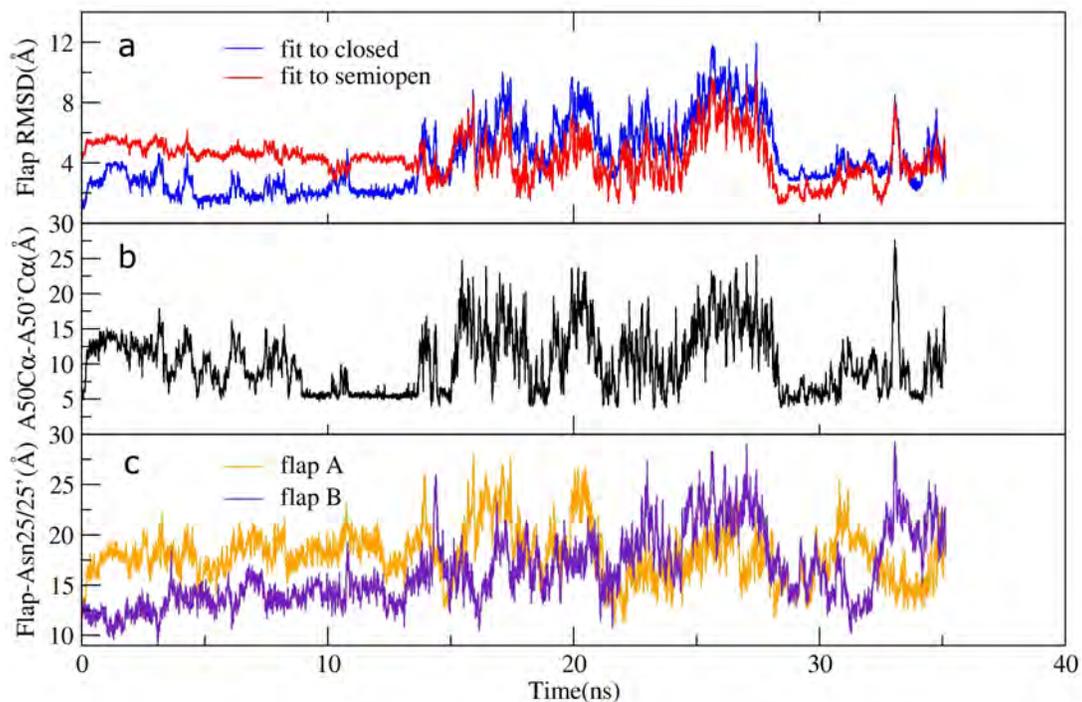


Figure 2-17. Time evolution of (a) the flap C α RMSD with respect to the two X-ray structures; (b) flap tips distance and (c) flap-Asn25/25' distances throughout the entire simulation of the I50A mutant system.

In contrast, the flaps in the initial closed structure of the I50W mutant were rather stable and no major conformational changes were observed over the course of the MD simulation (Figure 2-18).

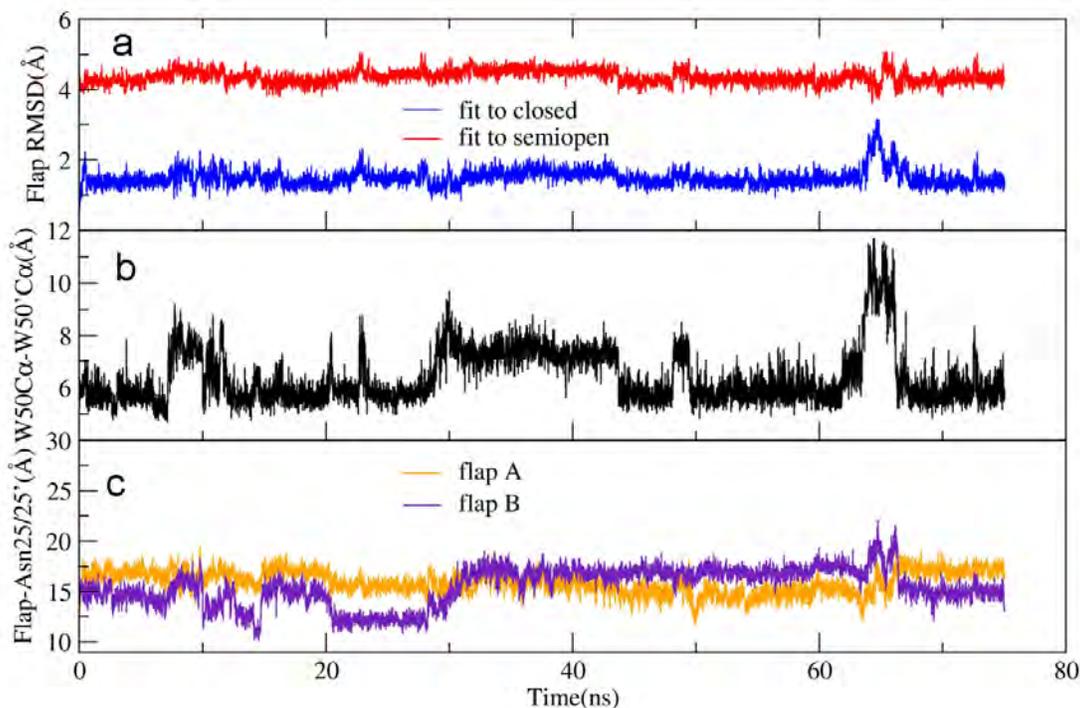


Figure 2-18. For the I50W mutant system, time evolution of (a) flap C α RMSD with respect to the two X-ray structures, closed and semi-open (b) flap tips distance and (c) flap-Asn25/25' distances.

As demonstrated in Figure 2-19, the I50W assumed a closed conformation which is similar to that observed in the wild-type enzyme (1TSU), yet with the exception of the flap regions. In the I50W mutant, the two flaps were pulled upward slightly, indicating that the bulky side chains from Trp50/50' interrupted the complementarity of the intermonomer hydrophobic cluster, which in turn prevented the appropriate closing of the

two flaps. Interestingly, a novel class of inhibitor has recently been reported,¹⁶⁹ which acts through a mechanism similar to the single I50W mutation, targeting the hydrophobic core of one monomer (or so-called flap-recognition pocket) might alter the conformational behavior of the flaps, and thus prevent the substrate's access to the active site or disrupt substrate cleavage due to inappropriate or incomplete flap closure.

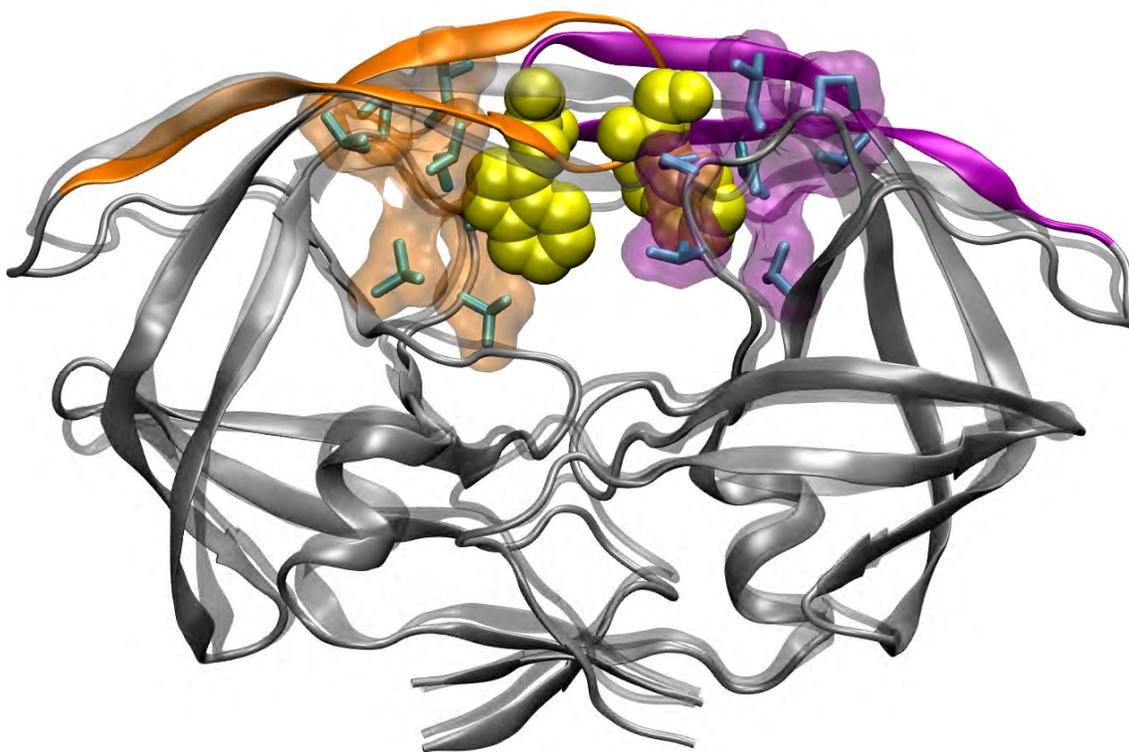


Figure 2-19. A snapshot of the ‘closed’ conformation sampled in the I50W mutant system (rendered by New Cartoon in VMD⁸⁵ and flapA colored in orange, flapB in purple) overlapped on the closed crystal structure (gray, transparent). Trp50 is rendered by VdW and colored in yellow. For clarity, only the sidechains of hydrophobic residues on the flap and loop region from each mutant monomer are shown, including Val32/32’, Ile47/47’, Ile54/54’, Val56/56’, Pro79/79’, Pro81/81’ and Val82/82’, and rendered by surface (residues on monomer A are colored in orange; residues on monomer B are colored in purple).

2.5 Conclusions

On the basis of the presented results, a number of conclusions can be drawn both from the methodological and from the biological point of view. First, we have shown that the present all-atom MD simulations in explicit solvent are able to simulate long timescale protein dynamics, and accurately reproduce experimentally determined structural and dynamical features of the HIV-1-PR. Furthermore, extensive sampling of multiple and reversible interconversions among different states of the flaps, ‘closed’, ‘semi-open’, ‘open’, and novel ‘curved’ conformations, in our microsecond time scale simulations of apo HIV-1 PR in explicit solvent, allowed a better understanding of the mechanisms underlying these transition dynamics. In this present work, we aimed to explore the conformational rearrangements between the closed and semi-open forms. On the basis of detailed structural analysis, we propose that the backbone of the highly conserved and flexible flap tip residues, particularly Gly49, undergoes significant twisting, initiating the large-scale rearrangements of the flaps between the closed and semi-open forms. In addition, the simultaneous rotation of the backbone torsion angles of the flap tip Ile50 residue displaces its side chain from being buried into the hydrophobic cluster with each monomer, facilitating swapping Ile50 between the hydrophobic clusters, encompassing Val32/32’, Ile47/47’, Ile54/54’, Val56/56’, Pro79/79’, Pro81/81’ and V82/82’, from the same or symmetry-related monomer. Thus, our proposed transition mechanism rationalizes the sequence requirements in the flap region and highlights the

importance of maintaining the hydrophobic interactions between Ile50 and the hydrophobic core region of each monomer in protein dynamics.

Chapter 3

Microsecond Timescale MD simulation Suggests that Partial Dimer Dissociation is the Flap Opening Mechanism of HIV-1 PR

Abstract

Human immunodeficiency virus type 1 protease (HIV-1 PR) is a major drug target in the fight against AIDS. Drug resistance mutations, however, are demanding the development of novel and more potent therapeutics. For a ligand to access the active site, the two flaps of the protease are presumably to open. Here, we present a working model of the flap-opening mechanism captured in a microsecond timescale simulation with explicit solvent model. The detailed structural and energetic analyses reveal that it is the various binding interactions of the dimer interface that govern the gating properties of the flaps; the opening of the flaps results from the concerted partial dissociation of the dimer interface facilitated by water dynamics. The significance of the inter-subunit interactions along the dimer interface in the gating dynamics is further supported by the subsequent simulations on a double mutant system (R87K/D29A), in which the flaps opened more readily due to decreased dimerization energies. In addition, our working model offers a novel site for allosteric regulation of the gating dynamics of the flaps; targeting the highly conserved yet weak region of the dimer interface encompassing the single α -helix

(residues 86–94) and the β -turn (residues 4'-9') may affect the equilibrium of different conformational states, and thus inhibit its catalytic activity.

Acknowledgments

This chapter contains direct excerpts from the manuscript written by Fangyu Ding with suggestions and revisions from Professor Carlos Simmerling. The code used in the free energy decomposition was written by Dr. Lin Fu. Additional assistance was provided by Amber C. Carr.

3.1 Introduction

HIV-1 PR is essential in post-processing viral polypeptide precursors to produce the infectious virion.¹⁷⁷ The enzyme is only active as a dimer, with each subunit made up of 99 amino acids, and there is an equilibrium between the monomeric and dimeric form of HIV-1 PR with a low dissociation constant ($K_D < 5\text{nM}$).²¹⁴ Analysis of data from the Stanford Drug Resistance Database²¹⁵ indicates that while polymorphisms in the sequence of HIV-1PR naturally occur, there are regions in the protein sequence that appear invariant under normal evolutionary pressures. These invariant regions coincide with the structural elements of the dimer interface^{150, 216-217} (Figure 3-1): i) the four-stranded anti-parallel β -sheet comprising the N-, C-termini of the two subunits (residues 1–4 and 96–99);^{90, 151} ii) the active-site region (residues 24–29) comprising the catalytic triads Asp25-Thr26-Gly27, and forming a hydrogen bond network, called the “fireman’s grip”,⁹⁵ iii) the Gly-rich flap tips (residues 49-52), which are believed to control substrate

access to the active site; and iv) a region encompassed by the single helix (residues 84–94) and the β -loop (residues 4'-9').^{97, 218} (The prime indicates a residue from the symmetry-related monomer in the following context).

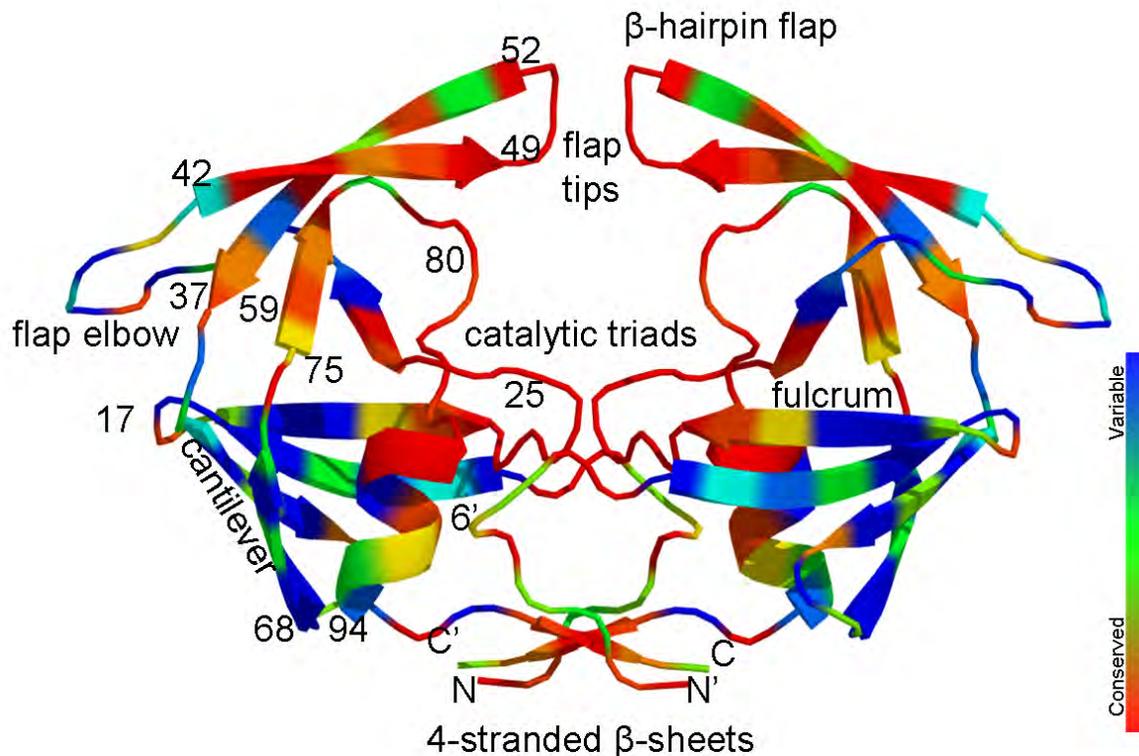


Figure 3-1. A detailed view of the free HIV-1 PR with the semi-open conformation (PDB code 1HHP¹⁹⁵). The residues are colored according to their conservation score computed by the ConSurf²¹⁹ program. The figure was generated with PyMol¹.

Due to its central role in processing viral polypeptide precursors, HIV-1 PR remains a primary target for anti-AIDS drugs. Knowledge of the substrate specificity and catalytic mechanism of this enzyme has led to a great success of the structure-based design of competitive inhibitors, which have resulted in a great reduction in HIV viral load when combined with other antiretroviral drugs (HAART).¹⁷⁵ Long-term treatment, however, has been hindered by the emergence of drug-resistant strains. Therefore, development of

new inhibitors acting by an alternative mode of inhibition is essential for successful treatment of HIV-positive patients.¹²⁶

One potential alternative mode of inhibition is to target the dimer interface because dimerization of HIV-1 PR subunits is indispensable for its proteolytic activity; each subunit contributes one of the two catalytic aspartic acid residues (Asp25) that form the active site.²²⁰⁻²²¹ As a result, the N- and C-termini which are intertwined in an antiparallel sheet and contribute most to the dimer stability of HIV-1 PR,¹⁵¹ have been employed as a dimerization inhibition target by several groups,^{163, 167, 171, 222} and certain compounds have been reported which exhibited potent activity against HIV-1 PR.

A second potential mode of inhibition targets the flap region of HIV-1 PR. This region is seen to exhibit substantial flexibility, as revealed by the heterogeneity of X-ray crystal structures of apo protease, including the semi-open (PDB code 1HHP¹⁹⁵), open (PDB code 1TW7¹⁸²), and closed (PDB code 1G6L,²²³ observed in a tethered dimer) conformations. NMR relaxation studies^{97, 111-112} have also identified motions on two different timescales occurring in the flap region of the free protease. Another attractive strategy to inhibit HIV-1 PR has been suggested based on the above observations, which is to target the thermodynamic balance of the semi-open, open and closed states of the free protease.¹²⁶ To this end, numerous computational studies have been carried out with an effort to elucidate the gating dynamics of the flaps, which will greatly aid in understanding its functional mechanism and substrate/inhibitor binding dynamics, and facilitating the design of novel inhibitors.

Scott and Schiffer¹¹⁵ performed a conventional MD simulation on a free protease using explicit solvent, in which the flap opened from the semi-open conformation after 3

ns and persisted until the end of the 10 ns simulation. Based on this observation, they proposed that a curling motion of the flap tips triggered the large opening event. However, the timescale of this large opening event is not consistent with NMR data from Torchia's group,¹¹¹ which indicates that the transition from the semi-open to the open conformation of the apo HIV-1 PR occurs on the 100 μ s timescale; thus, it is unclear if this fast opening dynamics is relevant to its biological function or indicative of instability in the model. Indeed, other possible causes for such irreversible opening motions have also been suggested, including the use of the GROMOS force field or an insufficient equilibration protocol.²²⁴

Owing to the μ s-ms timescale for large conformational changes¹¹¹, accelerated methods have been employed to model protein dynamics. Hamelberg et al.²²⁵ and Wiley et al.²²⁶ developed accelerated MD simulation methods to produce flap opening in HIV-1 PR, and the cis-trans isomerization of the Gly-Gly ω -bond in the flaps has been suggested to trigger the opening event. Tozzini and McCammon developed a coarse-grained model,^{122, 227} which enabled sampling of multiple opening and closing events on the microsecond timescale at the cost of atomic details, thus precluding an atomic understanding of the interactions between side chains which may have a major influence on protein dynamics, or of how sequence mutations affect dynamic behavior, and of how solvation is coupled to dynamics.

The first report about reversible opening events of the flaps using unstrained, all-atom MD simulations of HIV-1 PR was given by our group with an implicit solvent model.⁴⁵ The use of an artificially low viscosity allowed us to simulate a multi-scale model of HIV-1 PR dynamics in a 10 ns timescale MD simulation. The flap dynamics,

however, might have been biased due to limitations of the implicit solvent model, such as inadequate hydrophobic interactions and overstabilization of salt-bridges.^{58, 192, 194} Therefore, there is still a need to understand the internal dynamics of the protease, and especially the mobility of its flap region, on a much longer timescale with a more accurate explicit solvent model.

Over the years, improvements in molecular dynamics algorithms and computer hardware have allowed MD simulations to access longer timescales over which many physiologically relevant processes take place.¹⁹⁰ In this present study, we observed a spontaneous and reversible full flap opening event during a 1-microsecond solvated MD simulation starting from a closed crystal structure with the substrate deleted. Based on detailed structural and energetic analyses, we propose that it is the various binding strengths of the dimer interface that govern the gating properties of the flaps; the opening of the flaps results from the concerted partial dissociation of the dimer interface facilitated by water penetration. This opening mechanism is further supported by our subsequent simulations on a double mutant protease (R87K/D29A); the modified dimerization energy results in drastic changes in protein dynamics and much accelerated opening process. In addition, a potential allosteric site could be predicted by this working model; small molecules targeting the highly conserved yet weak region of the dimer interface could change the flexibility of the protease such that the balance of the three states is shifted.¹²⁶

3.2 Methods

3.2.1 Molecular Dynamics Simulations of apo HIV-1 PR

MD simulations were performed on an inactive (D25N) apo HIV-1 protease (PDB ID code 1TSU), and initiated from two distinct conformations, closed and semi-open forms, termed MD_{closed} and MD_{semiopen} simulation, accordingly. System preparation and simulation details have been described in the Methods section in Chapter 2. The simulations were extended for 1 μ s (MD_{closed}) and 400ns (MD_{semiopen}), respectively.

3.2.2 Data Analysis

Root-mean-squared deviation (RMSD), distance and angle were calculated using the PTRAJ module in the AMBER10 software package.¹² Solvent Accessible Surface Area (SASA) for each atom was computed using the SANDER module of AMBER 10¹² by post-processing the trajectory with a continuum solvent model³⁶ after removing water molecules from the trajectory. The SASA was estimated using the MSMS algorithm²²⁸ with a probe radius of 1.4 Å. Molecular graphics were prepared by VMD,² PyMol¹ and Chimera.²⁰¹

3.2.3 Free Energy Estimation Using the MMPBSA/GBSA Approaches

For the energetic analysis, 10,000 structures were taken as snapshots at intervals of 100ps from the 1 μ s-long trajectory of the MDclosed simulation after removing the water molecules. For absolute binding free energy of the two monomer, the SIETRAJ²²⁹⁻²³⁰ program, a collection of scripts for calculating binding free energies from Amber-generated molecular dynamics trajectories, was used to estimate the binding free energy of each dimer using a continuum solvent model³⁶ The solvated interaction energy (SIE) is calculated for each snapshot by rigid infinite separation of the target and ligand. In this approach, the binding free energy ΔG is approximated by²²⁹

$$\Delta G_{bind}(\rho, D_{in}, \alpha, \gamma, C) = \alpha \times [E_c(D_{in}) + \Delta G_{bind}^R(\rho, D_{in}) + E_{vdw} + \gamma \Delta MSA(\rho)] + C \quad 3-1$$

where E_C and E_{VDW} are the intermolecular Columbic and van der Waals interaction energies in the bound state, respectively; these values are calculated using the AMBER molecular mechanics force field (ff99SB) with an optimized dielectric constant. ΔG_{bind}^R , which is the change in the reaction field energy between the bound and free states, accounts for solvation electrostatic energy and is calculated by solving the Poisson equation with the boundary element method program, BRI BEM,²³¹⁻²³² with a molecular surface generated with a variable-radius solvent probe using a marching tetrahedra algorithm.²³³⁻²³⁴ The cavitation cost is taken to be proportional to the change in molecular surface area upon binding, ΔMSA . The following parameters are calibrated by fitting to the absolute binding free energies for a set of 99 protein—ligand complexes.²²⁹ the AMBER van der Waals radii linear scaling coefficient (ρ), the solute interior dielectric constant (D_{in}), the molecular surface area coefficient (γ), the global proportionality coefficient related to the loss of configurational entropy upon binding (α), and a constant

(C). The optimized values of these parameters are $\alpha = 0.1048$, $D_{in} = 2.25$, $\rho = 1.1$, $\gamma = 0.0129$ kcal/(mol Å²), and $C = -2.89$ kcal/mol.

The pair-wise, per-residue free energy decomposition was performed using the MM-GBSA method encoded in the AMBER 10 program.¹² The MM-GBSA method combines the molecular mechanical energies with the continuum solvent approaches to evaluate the contribution of each residue to the total binding free energies or absolute free energies of molecules in solution, as well as the contributions of its side chain and backbone. The molecular mechanical energies are determined with the SANDER module from the AMBER program and represent the internal energy (bond, angle and dihedral), and van der Waals and electrostatic interactions. An infinite cutoff for all interactions is used. The electrostatic contribution to the solvation free energy is calculated with the generalized Born (GB) method implemented in SANDER.⁶¹ The Energy Decomposition Analysis (EDA) using the MM-GBSA method in the AMBER package can be performed by setting “idecomp” >0 in the input to call this function. “idecomp” has four options for four types of calculations□

- idecomp = 0, do nothing (default).
- idecomp = 1, decompose energies on a per-residue basis; 1-4 EEL + 1-4 VDW are added to internal (bond, angle, dihedral) energies.
- idecomp = 2, decompose energies on a per-residue basis; 1-4 EEL + 1-4 VDW are added to EEL and VDW
- idecomp = 3, decompose energies on a pairwise per-residue basis; the rest is equal to "1".

- `idecomp = 4`, decompose energies on a pairwise per-residue basis; the rest is equal to "2".

In this study, “idecomp” was set to 4, which gives the interaction energy between the pairs. The energy contribution of single residues in one monomer was obtained by summing its interactions over all residues in the other monomer.

3.3 Results and Discussion

3.3.1 Extent of Flap Motions Possible on the Microsecond Timescale

To understand the dynamic behavior of the enzyme, we performed two unrestrained, solvated MD simulations for an unliganded “apo” form of HIV-1 PR (PDB code 1TSU), starting from the closed and semi-open forms, respectively. In both simulations, multiple interconversions were observed, and allowed for a better understanding of transitions between the closed and semi-open forms, as discussed in the previous Chapter. Of particular interest in this study is to explore the full opening dynamics of the flaps, captured in the microsecond timescale simulation.

Consistent with our previous study,⁴⁵ the flap conformation was monitored by RMSD of the $C\alpha$ atoms of the two flaps (residues 46-55, 46'-55') overlapped on the flaps of the initial closed state and of the apo X-ray crystal structure (PDB code: 1HHP¹⁹⁵), respectively. The closed state is defined by flap $C\alpha$ RMSD to the closed reference structure of $< 1.7 \text{ \AA}$, and the semi-open state is defined by flap $C\alpha$ RMSD $< 2.5 \text{ \AA}$ with respect to the semi-open reference structure. The fully open configuration has not yet

been observed experimentally; thus, it is somewhat arbitrary to define which structures qualify as open. To measure the extent of flap opening, two distinct distances were selected: the distance between the C α atoms of the two flap tip Ile50 residues²³⁵ (termed as Ile50C α -Ile50'C α), and the distance between the Center of Mass (COM) of five central residues (residues 48-52) of each flap relative to the COM of all heavy atoms of the two active site asparagine residues, Asn25/25' (termed as flapA-Asn25/25' and flapB-Asn25/25', respectively). These distance values from the simulations were compared to those values measured in the semi-open crystal structure. Thereby, snapshots are denoted as 'fully open' states with flap C α RMSDs to the two references (closed and semi-open states) of > 6.5 Å, flap tips distance (Ile50C α -Ile50'C α) of > 8 Å, and flapA- and flapB-Asn25/25' distances of > 19 Å.

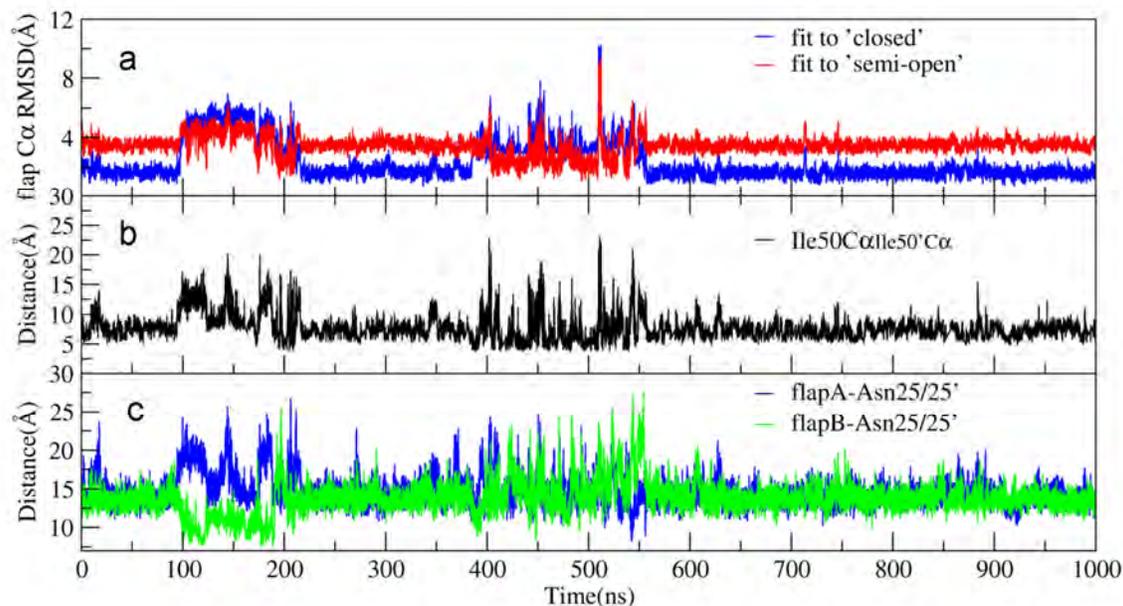


Figure 3-2. Time evolution of flap C α RMSDs (a), I50C α - I50'C α distances (b) and flapA- and flapB-Asn25/25' distances (c) during the MDclosed simulation.

Figure 3-2 displays the time evolution of the flap RMSD values, flap-tip distances and flap-Asn25/25' distances. It is obvious that a spontaneous and reversible full flap opening took place at ~509ns. This full flap opening occurred after the flaps had rearranged from the closed form to the semi-open form during the course of the second transitional period (385ns-558ns), as evidenced by flap RMSD values to both closed and semi-open references increasing to $>6.5 \text{ \AA}$. In these significantly open configurations, the two flaps were widely separated from each other with flap-tip distances reaching up to $\sim 23 \text{ \AA}$ (Figure 3-2b), and both flaps curled up and shifted away from the active site, with both flapA- and flapB-Asn25/25' being up to $\sim 27 \text{ \AA}$ (Figure 3-2c). Snapshots taken during this full flap opening event (509ns-513ns) are shown in Figure 3-3. It is apparent that the large separation of the two flaps leaves the catalytic residues (red) completely exposed to solvent, allowing a ligand to potentially access to the active site.

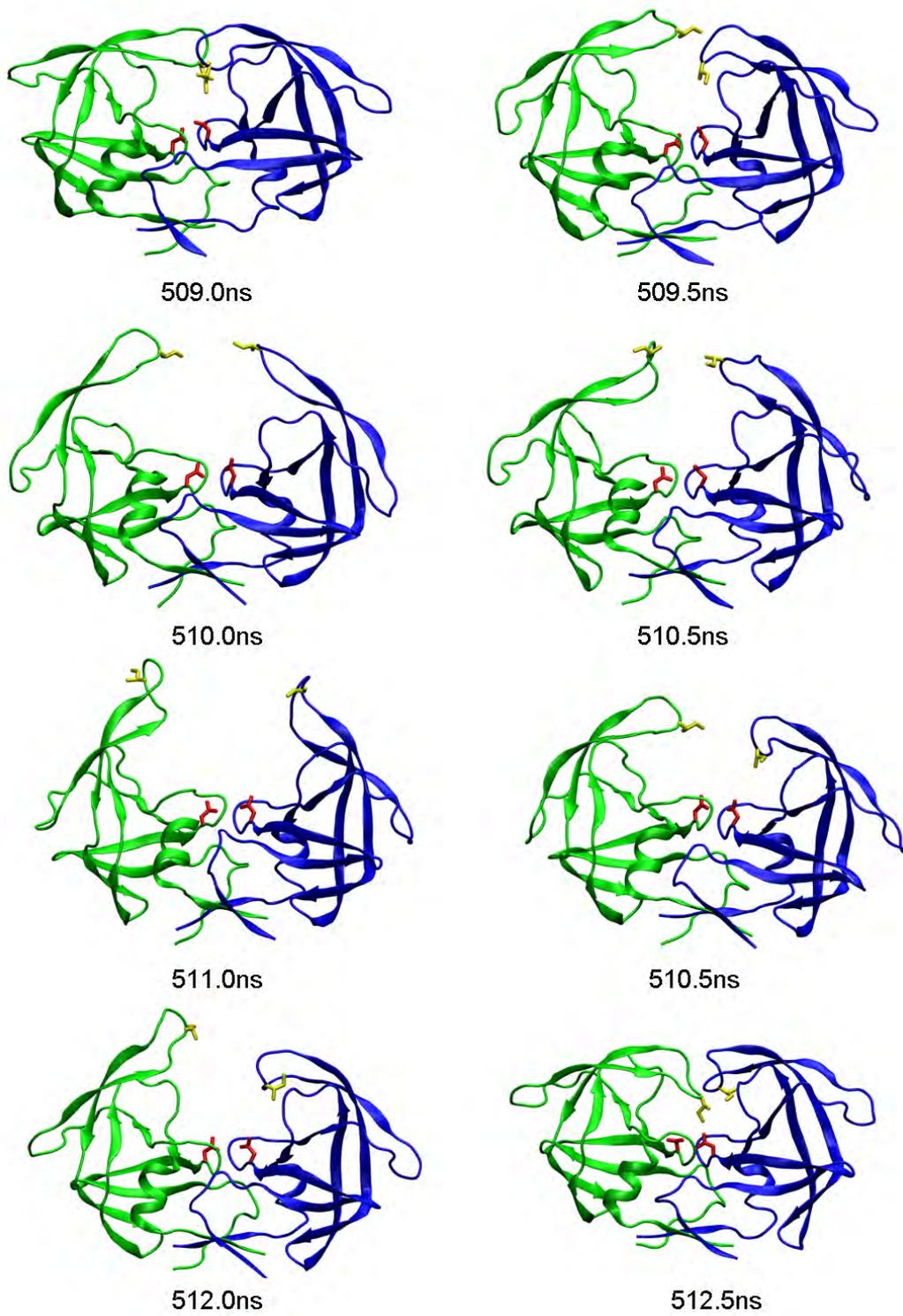


Figure 3-3. Snapshots of HIV-1 PR over the course of the full flap opening at 0.5ns intervals in the simulation starting from the closed conformation (MDclosed). Monomer A is colored in blue; monomer B in green. The active site Asn25/25' residues are highlighted in red; the flap tip Ile50 is highlighted in yellow.

Notably, the transiently sampled fully-open structures differ significantly from the wide-open crystal structure (PDB code 1TW7¹⁸²) which has the flap handedness characteristic of bound (closed) crystal structures, even though the two flaps are pulled upward and separated more widely than those in the semi-open structure. Thus, this simulation provides further evidence that the wide-open structure observed for MDR 769 does not arise from sequence variation, but instead is an artifact from crystal packing, as suggested previously by Langevin dynamics (LD) simulations.¹⁰²

It is also worth noting that the full opening of the flaps sampled in the MD simulation was reversible; the system returned to the semi-open form and eventually converted back to the closed form. To our knowledge, such a reversible, long-timescale and large-magnitude flap opening has not yet been observed in any unstrained, all-atom MD simulation with explicit solvent. Moreover, despite the reported differences in the exact extent of flap opening, the fully-open configurations which we obtained in our simulation are very similar to those reported in computational studies of the apo enzyme which have sampled an increased range of conformations by employing approaches to enhance conformational sampling.^{235,122, 236} Thus, it is likely that the full opening event captured in our simulation may be biologically relevant to dynamic processes in protein, and not indicative of instability in the model.

3.3.2 Partial Dissociation of the Dimer upon Flap Opening

To examine the changes in the internal structure of each monomer, we calculated RMSD values for the individual monomers with and without a best-fit to their

corresponding monomers in the semi-open reference structure. As shown in Figure 3-4a, the monomer structures showed good agreement with their corresponding monomers in the semi-open reference (1HHP pdb code), with the best-fit RMSD values of both monomers of being ~ 2 Å even during the full opening event. This indicated that no significant changes occurred in the internal structure of individual monomers despite the large scale rearrangements of the flap region. On the other hand, the no-fit RMSD values of each monomer (Figure 3-4b), which were calculated by best fitting the other monomer onto its corresponding monomer in the semi-open reference, exhibited considerable fluctuations, which reached up to ~ 17 Å upon flap opening, and then reduced to ~ 2 Å when the system returned to the semi-open state. Thus, these results revealed that the relative orientation of the two monomers in the full open structure significantly yet transiently drifted away from that in the semi-open structure without significant changes in their respective internal structures.

A graphic representation of these results can also be seen in Figure 3-5, where a best-fit superimposition is shown based on one monomer from each dimer: a fully-open snapshot from the MD simulation, and the semi-open X-ray structure. This superimposition confirms that the internal structure of each monomer remained unperturbed as the secondary structure elements of the best-fit monomer A (blue) are well aligned with those in the corresponding monomer in the semi-open reference structure. The noticeable exception to this observation is the flap region, which fluctuated considerably and became less ordered.

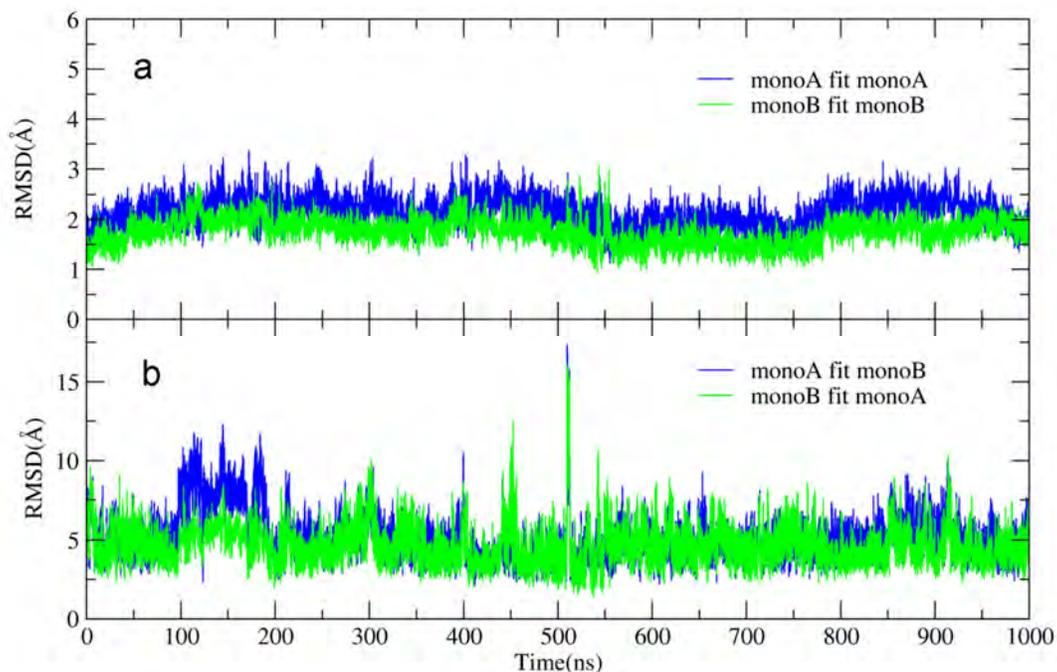


Figure 3-4. Time evolution of all C α RMSD of each monomer (blue line, monomer A; green line, monomer B) when fitting to the corresponding monomer in the semi-open reference (top panel); when only fitting its symmetry-related monomer to the corresponding monomer in the semi-open reference (bottom panel).



Figure 3-5. A detailed view of the superimposition of the semi-open X-ray structure (gray, transparent cartoon) and a fully-open snapshot (monomer A in blue and monomer B in green) from the MD simulation based on the best fit using one monomer (blue).

Moreover, it is apparent that the no-fit monomer in the fully-open structure (green) rotated by $\sim 40^\circ$ from its counterpart in the semi-open reference (gray cartoon), revealing a domain rotation of the two monomers relative to each other upon flap opening. Remarkably, we note that other than the two flaps, the dimer interface was partially dissociated, involving the active site region as well as the region encompassed by the single helix from one monomer (residues 84–94) and the β -loop from the other monomer (residues 4'-9').

Taking into account the experimental observations¹⁴⁷⁻¹⁴⁸ that HIV-1 PR exists in a monomer-dimer equilibrium, we hypothesize that a large-scale flap opening might be an intermediate upon association/dissociation of the dimer. Although native monomers are extremely unfavorable such that no monomeric intermediates have been detected experimentally,¹⁵¹ variants containing mutations that dramatically disrupt dimer interface contacts, PR_{R87K}, PR_{D29N}, PR_{T26A}, PR₅₋₉₉, and PR₁₋₉₅, have been reported to maintain a folded monomer in the absence of inhibitor.^{97, 218} The NMR structure of the HIV protease monomer reveals that that it exhibits a stable tertiary fold spanning the region of residues 10–90, which is essentially identical to the individual monomer of the dimer; and that its flap adopts an open conformation with significant disorder in its tip (residues 48-53). These structural features of the isolated monomer are consistent with the monomers in the fully open configuration captured in our MD simulation, suggesting that our working model may have reproduced correctly the structures in the intermediate steps of the dissociation of the dimer, i.e., the flap opening.

3.3.3 Dimer Dissociation upon Flap Opening

To monitor the lateral movement of the two monomers, which seems to be coupled with large-scale flap opening, we measured the distance between the center of mass (COM) of the C α atoms of residues 32, 75-76 and 57-58 located in the core domain of each monomer, termed the ‘core domain distance’. These residues were chosen because they are distant from the flap region, and did not exhibit conformational changes on the μ s-ms timescale in NMR relaxation measurements. Thus, this distance measurement is not likely affected by the flap fluctuations, but is sensitive to the relative movement of the two monomers. For comparison, the core domain distance is ~ 30 Å in closed form and ~ 31 Å in semi-open form.

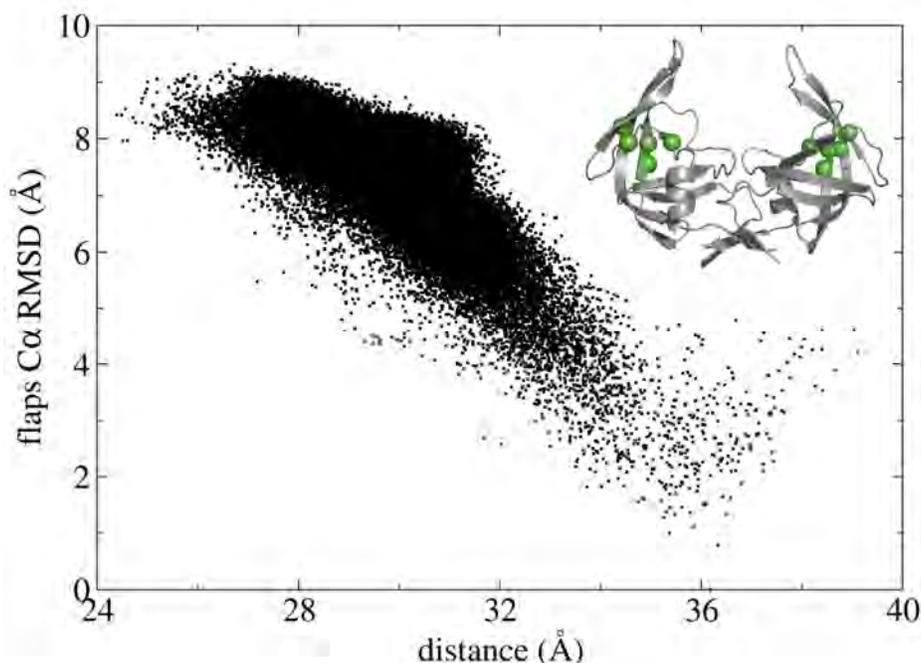


Figure 3-6. Distance between the core domains (residues 32, 75-76 and 57-58) of the two monomers as a function of flap C α RMSD with respect to the fully open configuration captured in the simulation (gray cartoon). The C α atoms of the core domain residues represented as green spheres in the open reference structure. Note that distances lower than 28 Å correspond to the curled/tucked conformations.

Figure 3-6 reports the core domain distance as a function of flap C α -RMSD with respect to a fully open configuration captured in the MD simulation. As it is shown, the core domain distance exhibited a strong negative correlation with the flap C α -RMSD to the fully open structure with the correlation coefficient < -0.6 . Remarkably, the two monomers were partially dissociated when the flaps adopted fully open conformations (with RMSD values to the open reference of < 4 Å), as evidenced by significant increases in the core domain distances up to ~ 39 Å. Therefore, this analysis provides further evidence that along with local large-scale rearrangements of the flap region, the dimer interface underwent dissociation during the full flap opening event.

In addition, it is worth noting that evaluating the free energy difference between different conformations of HIV-1 PR at biological temperatures was precluded, owing to the high temperature used in this simulation. A reaction coordinate for umbrella sampling simulations,⁴⁷⁻⁴⁹ however, can be suggested from this working model; an angle between the two vectors (Figure 3-7a), which are formed by the COM of the C α atoms of residues 32, 75-76 and 57-58, located in the core domain of each monomer, and the COM of residues 2-3, 96-97 from the N-, C- termini of both monomers, respectively. As illustrated in Figure 3-7b, a contour plot of the free energy as a function of the open angle and the flap C α RMSD with respect to the fully open configuration, this angle appears to be a dynamically meaningful measure of conformational changes of the protease, ranging from the closed, semi-open and open form. Hence, performing an umbrella sampling simulation along this reaction coordinate will most likely generate the potential of mean force (PMF) of the protease at low temperature of interest. This will help to gain valuable

insight into the mechanism underlying the interconversions between different conformers of HIV-1 PR.

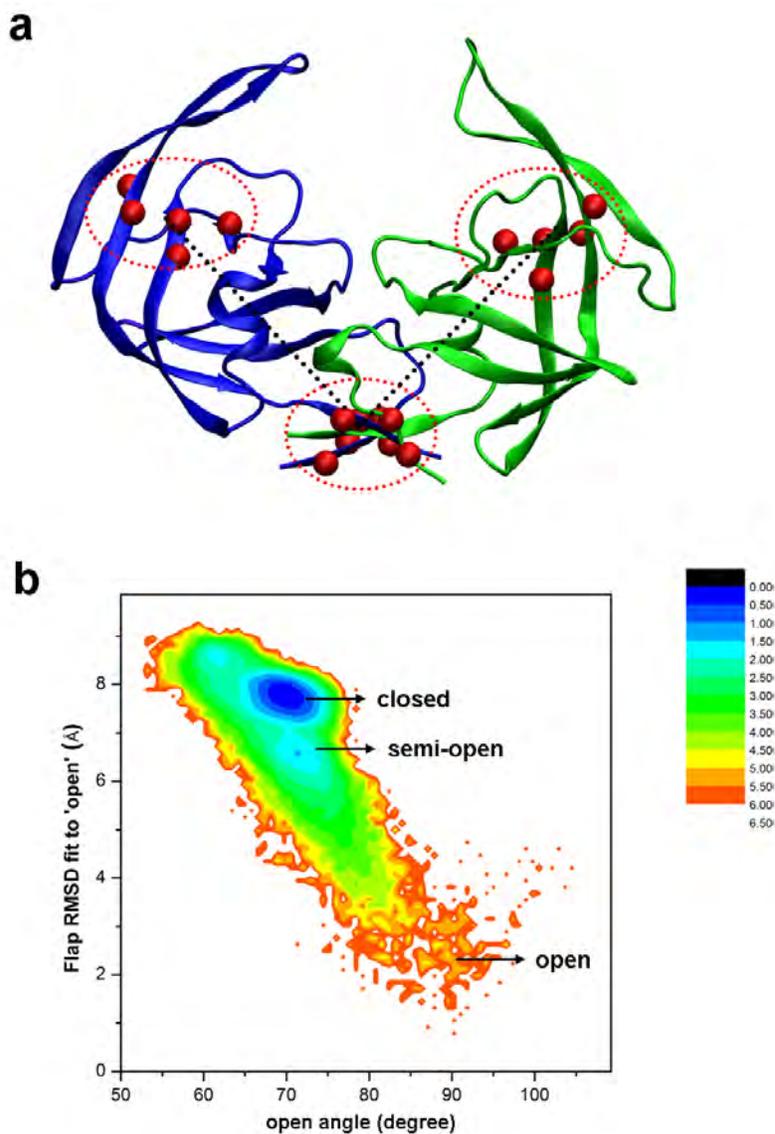


Figure 3-7. a) An open snapshot from the MD simulation is shown to illustrate the definition of the open angle; a triangle between the two vectors, which are formed by the COM of the C α atoms of residues 32, 75-76 and 57-58, located in the core domain of each monomer, and the COM of residues 2-3, 96-97 from the N-, C- termini of both monomers, respectively (solid black lines). Carbon atoms of these residues are represented as red spheres. b) Contour plot of the free energy as a function of the open angle and the flap C α RMSD with respect to the fully open configuration captured in the MD simulation.

3.3.4 Binding Free Energy Calculation on Dimer Stability of the HIV-PR

In light of the partial dissociation of the dimer interface upon flap opening, a significant decrease in the binding free energy (ΔG_{bind}) of the dimer is anticipated upon flap opening. To quantify the strength of the intermonomer interaction energy, which will shed light on the interpretation of the observed opening dynamics in terms of energetics, the binding energy (ΔG) of the two monomers in the snapshot structures and its different components were estimated using the solvated interaction energies method, SIETRAJ,²³⁰ an alternative to the MM-PBSA software provided by the AMBER distribution. Owing to the high computational demand of this program, snapshots taken every 100ps from the 1 μ s MD trajectory were evaluated for a total of 10,000 structures. The results are reported in Figure 3-8 as a function of the core domain distance. As shown, the estimated binding free energies (ΔG) of the dimer ranged from -19 to -31 kcal/mol. Significantly, the values of ΔG were weakened by > 6 kcal/mol when the two monomers were partially dissociated with their core domains separated by > 34 Å, as compared with the closed ($\Delta G \approx -27$ kcal/mol) and the semi-open forms ($\Delta G \approx -25$ kcal/mol). Thus, this energetic analysis provides strong support for the partial dissociation of the dimer during the large-scale flap opening event, which results in a striking decrease in ΔG .

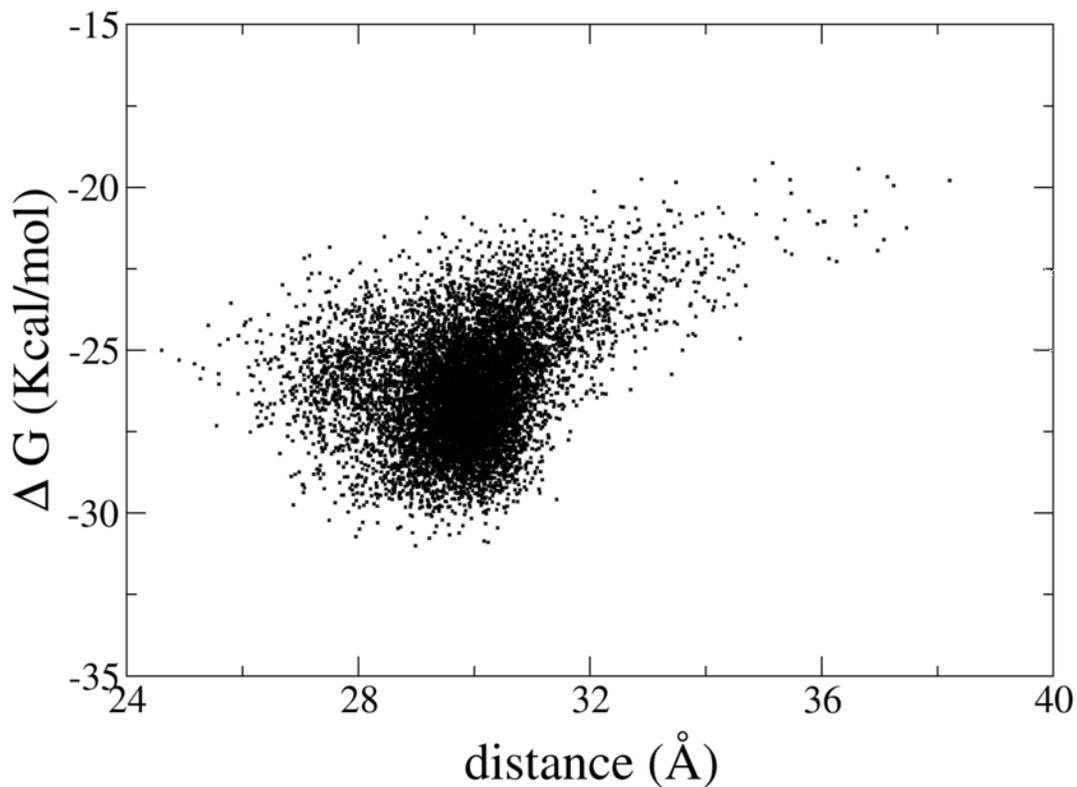


Figure 3-8. Binding free energies (ΔG) of the dimer as a function of the core domain distance.

It is worth pointing out that the binding free energy of the HIV-1 protease dimer has been measured experimentally and varies with experimental conditions, such as protein concentration, urea, pH and temperature.^{152, 237-239} At pH 7, the K_d was measured as 50 nM,²³⁷ which corresponds to a binding free energy of -10.0 kcal/mol. Thus, the order of magnitude of our results is consistent with the experimental data, but the absolute values of the binding free energies overestimate the strength of binding. The discrepancy between our results and the experimental data may be resulted from the caveats of implicit solvation and neglect of vibrational entropy upon dimerization in the free energy calculation.²²⁹ However, we are interested in relative binding free energies between different conformations, the errors of the free energy values are likely to cancel out.

When examining the correlation of ΔG with each of its components (Figure 3-9, we found that the intermonomer van der Waals interaction energy (E_{vdw}) is the dominant contribution to the calculated ΔG (with the correlation coefficient $R > 0.9$).

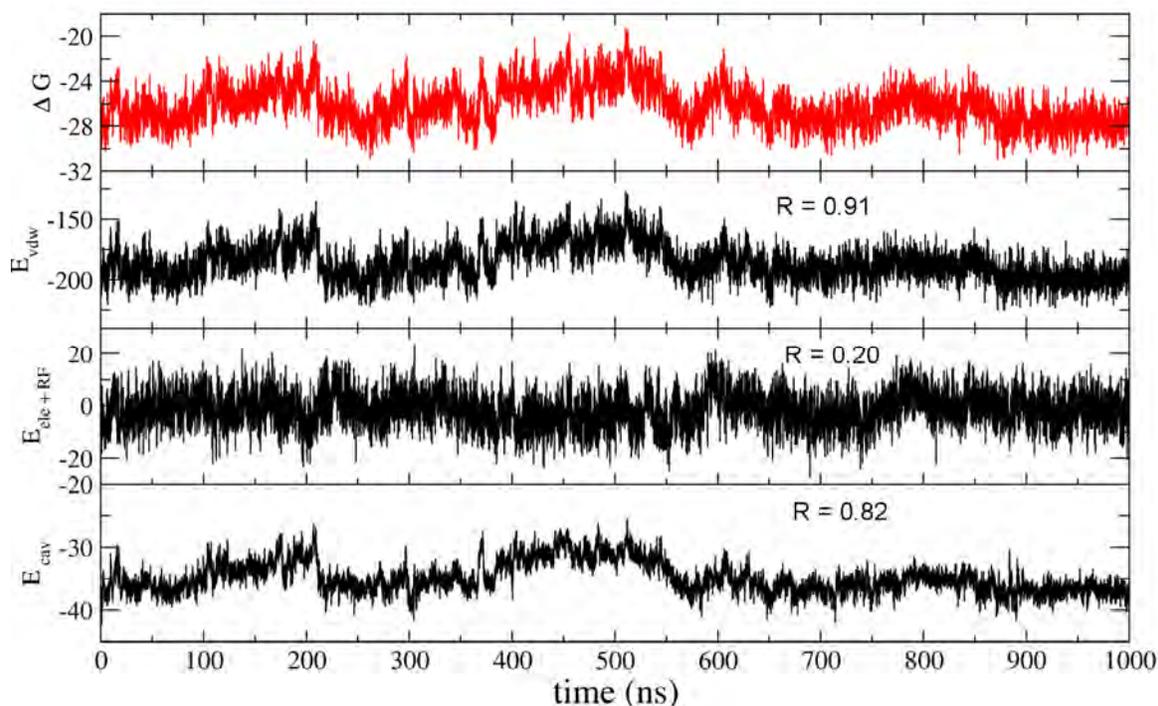


Figure 3-9. Time evolution of the calculated binding energy (ΔG) for 10,000 snapshot structures, divided into nonpolar and polar energetic contributions: intermolecular vdW (E_{vdw}), intermolecular electrostatic (E_{ele}) plus change in reaction field (E_{RF}), and change in nonpolar solvation energy (E_{cav}) which is proportional to change in solvent accessible surface area (SASA). Energies are in kcal/mol.

The dominance of the vdW interaction energy is not surprising, since upon folding and dimerization, 64% of the non-polar and 62% of the polar surface of the HIV-1 protease are buried into the solvent.¹⁵¹ An extensive hydrophobic core extends through the dimeric interface. The N- and C-termini (residues 1-3, 96-99) form one side of the central four-stranded β -sheet and pack onto Leu24 and Thr26, adjacent to the catalytic residues and the hydrophobic residues of the helix (residues 86-94). It is worth

mentioning that the dominance of the inter-molecular vdW interactions has also been reported by a previous binding free energy calculation on dimer stability of HIV PR by a MM-PBSA method.²⁰⁹

The second contribution to the binding free energy arises from the change in nonpolar solvation energy, which is proportional to the solvent accessible surface area. The significance of the nonpolar solvation energy is evidenced by the strong positive linear correlation between E_{cav} and ΔG (with $R > 0.8$). In addition, the favorable contribution of the electrostatic interactions between the two monomers (E_{ele}) is canceled out by the electrostatic desolvation upon dimerization. As a result, the sum of the electrostatic interaction energy and the reaction field energy ($E_{\text{ele+RF}}$) only exhibited a weak correlation with ΔG ($R > 0.2$). Notably, E_{vdw} significantly decreased upon opening event (509 ns – 512 ns), from -166 to -139 kcal/mol, as did E_{cav} , which decreased from -31 to -26 kcal/mol. These energies, however, were fully recovered when the flaps re-closed. Thus, these results further support that the dimer underwent a reversible dissociation upon large-scale flap opening.

3.3.5 Per-residue Free Energy Decomposition

To identify regions which contribute most to the association free energy, and thus are critical in the gating dynamics of the flaps, a pair-wise per-residue decomposition of the binding free energy has been performed using the MM-GBSA approach^{61, 240} on the 10,000 structures extracted from the entire trajectory. Figure 3-10 reports the average intermonomer vdW interaction energy of each residue.

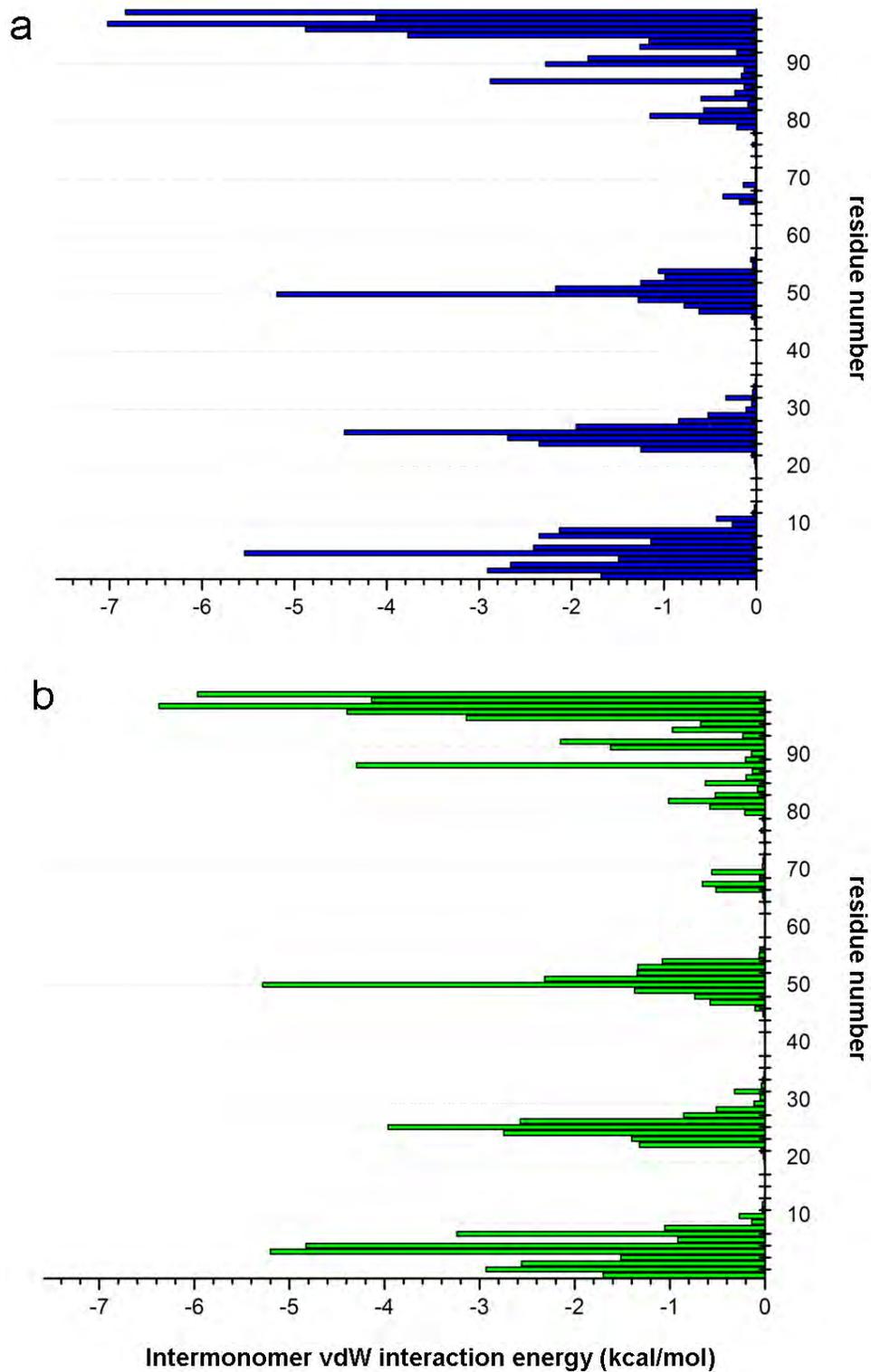


Figure 3-10. Intermonomer vdW interaction energies of the residues from monomer A (blue) and monomer B (green) averaged over the 10,000 snapshots from the 1 μ s simulation.

It is evident that the binding strength is not evenly distributed, with residues located at the N- and C- termini (1-4, 96-99) making the most significant contributions to the total vdW interaction energy of the two monomers. This is in good agreement with the structure-based thermodynamic analysis,¹³⁸ which has shown that the N- and C- termini contribute close to 75% of the total Gibbs free energy of stabilization. Notably, residues located at other dimer interface regions also make comparably favorable vdW interactions, in particular, Thr26 located at the base of the active site, Ile50 located at the tip of the flap region, Leu5 from the β -loop (residues 4-9), and Arg87, and Leu90 located in the α -helix (residues 86-94). This result implies that these residues play a crucial role in the stability of the dimer; substitutions at these residues may lead to the disruption of the interactions along the binding interface, resulting in a drastic decrease in the dimer stability. The structural significance of Leu5, Thr26, Ile50, Arg87 and Leu90 revealed by this energetic analysis is in good agreement with experimental studies, which have shown that the introduction of T26A, D29A, D29N and R87K to HIV-1 PR disrupts the dimer interface interactions and therefore depletes its catalytic activity.^{97, 218}

To explore the coupling between the distal regions of the protease,⁹⁴ we subsequently calculated the correlation coefficient of the intermonomer vdW interaction energy for each residue and the total vdW interaction energy of the dimer along the open trajectory (509ns -513ns). The result is reported in Figure 3-11, with the semi-open structure color coded according to the magnitude of the correlation coefficient of the corresponding residue. It is apparent that the dimer interface residues at the flap regions demonstrate strong and positive correlations with the total vdW interaction energy and to lesser extent, the active site residues (24-26), the 80s loop, and residues from the α -helix and β -loop.

Notably, the N- and C-termini exhibit weak correlations with the total vdW interaction energy. These results indicate that the decrease in the total intermonomer vdW interaction energy upon flap opening is likely due to the disruptions of the nonpolar interactions between the flap tips and the active site, and between the helix and the turn; the N- and C-termini β -sheet interactions remain largely unperturbed, preventing complete dissociation of the dimer during the large-scale opening event.

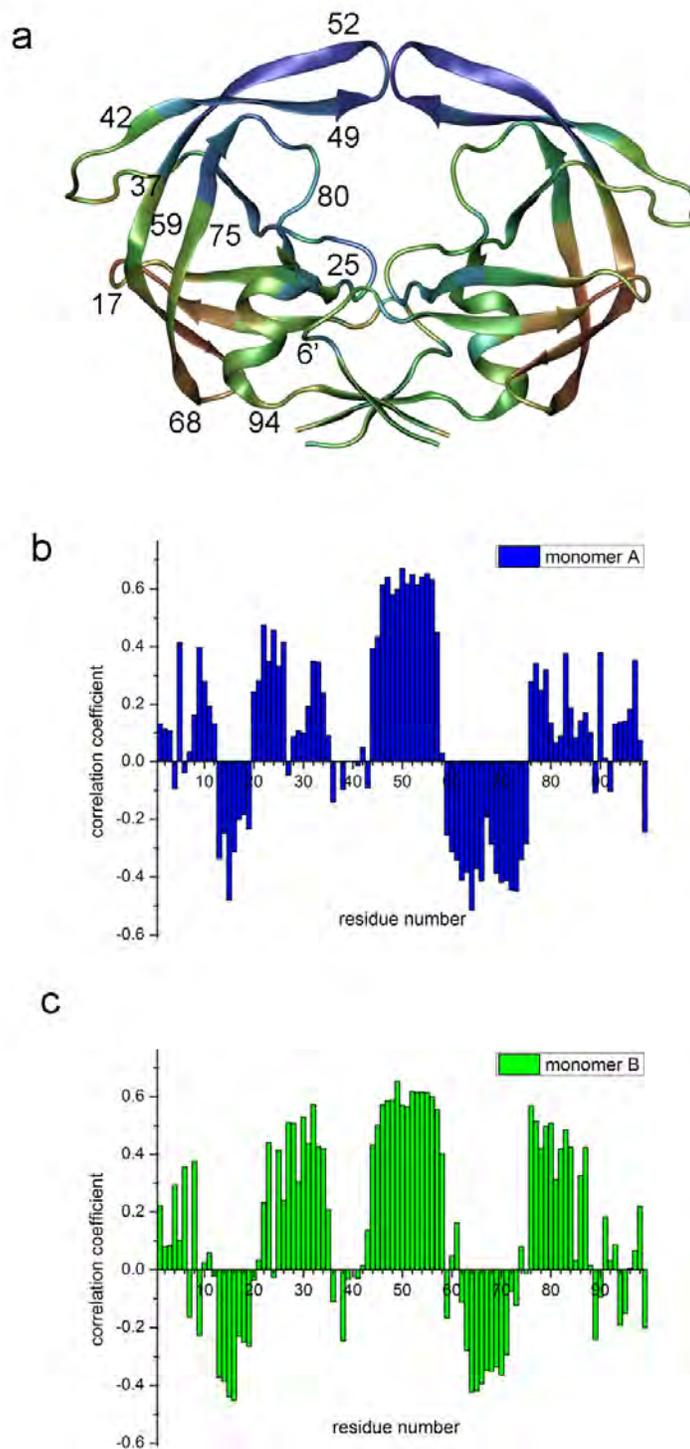


Figure 3-11. a) A semi-open structure colored by the magnitude of the correlation coefficient between the intermonomer vdW interaction energy of the corresponding residues with the total interaction vdW interaction energy of the dimer. Red and orange indicate negative correlations; blue and green indicate positive correlations. Actual values of the correlation coefficient is shown in (b) for residues from monomer A, (c) from monomer B.

The potential role of the dimer interface interactions in the flap opening dynamics, excluding the N- and C- termini region, is further confirmed by calculating the change in the absolute intermonomer vdW interaction energy of each residue upon flap opening (Figure 3-12). The difference in intermonomer interaction energy for each residue was obtained by subtracting the intermonomer vdW interaction energy averaged over the open trajectory from the intermonomer vdW interaction energy averaged over the semi-open trajectory ($\bar{E}_{\text{semi}} - \bar{E}_{\text{open}}$). A semi-open structure is shown in Figure 3-12a, and is color coded according to the magnitude of the change intermonomer vdW interaction energy of the corresponding residue. As shown, in addition to the flap tips, residues from the active site loops (21-31/21'-31'), Leu5, Leu90 from monomer A (Figure 3-12b), Trp6', Arg8', and Arg87' from the β -loop and the helix of monomer B (Figure 3-12c) also exhibit significant decreases in their respective intermonomer vdW interaction energies when the flaps open. Thus, this calculation suggests that the disruption of the intermonomer interactions involving the residues at the active site region, the helix and β -loop region (pink regions in Figure 3-12a), may play a significant role in the dissociation of the dimer, thus causing the flaps to open.

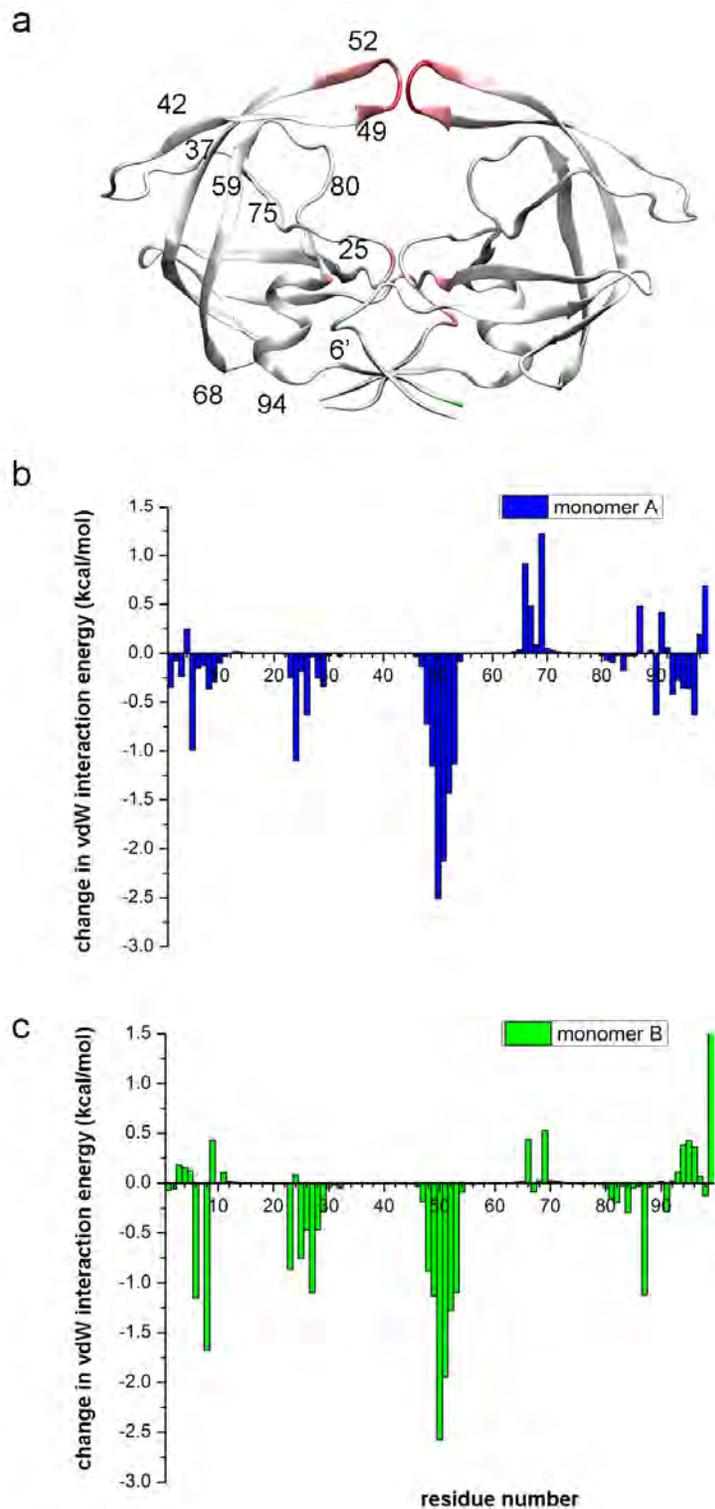


Figure 3-12. a) A semi-open structure colored by the magnitude of the change in the intermonomer vdW interaction energy of the corresponding residues upon flap opening using RWG color code was used (red to green corresponding to -2.5 to 2 kcal/mol). The actual values of these changes are shown in (b) for residues from monomer A, (c) from monomer B.

It is also noteworthy that the fulcrum (residues 10-20) and the cantilever (residues 60–75) regions demonstrate strong negative correlations with the total vdW interaction energy (Figure 3-10). This anticorrelation is also reflected in the ability of these regions to make more favorable vdW interactions when the flaps open (Figure 3-12a, green regions). This anticorrelated behavior can be explained by the compression of these regions caused by the hinged motion of the two monomers upon flap opening. A similar anticorrelated behavior between the cleft formed between a loop (residue 38-42) and a β -strand (residues 59-63) and the opening of the flaps has been previously reported in both our previous implicit simulations⁴⁵ and coarse-grained models developed by McCammon and coworkers.^{122, 236} Thus, this working model provides further support for the potential of these regions as allosteric sites; molecules making contacts with the flap elbow, cantilever and fulcrum may exert allosteric control of the flaps.

3.3.6 Solvation of the Dimer Interface

Taking advantage of the explicit solvent model, we subsequently investigated the water dynamics during the flap opening by calculating the differences in the averaged SASA per residue between the semi-open and the open trajectories. The results are presented in Figure 3-13 and mapped onto the semi-open X-ray structure in Figure 3-13a.

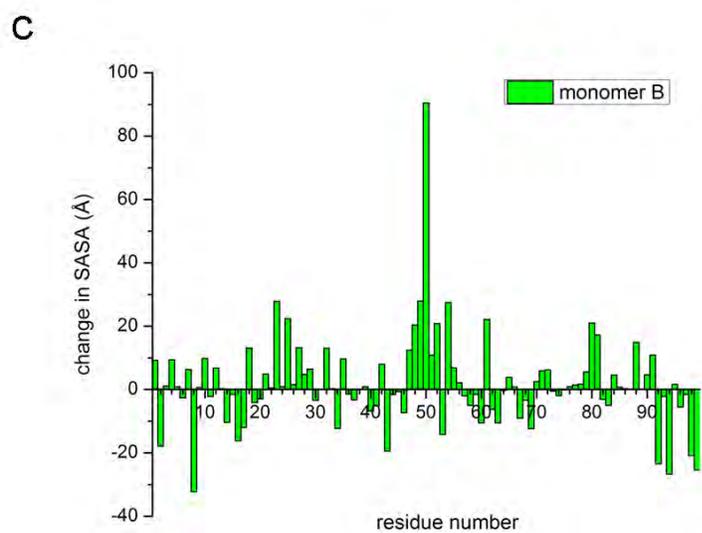
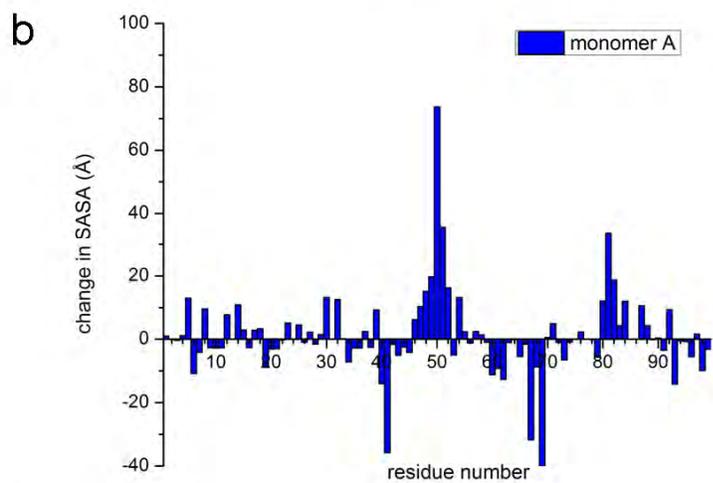
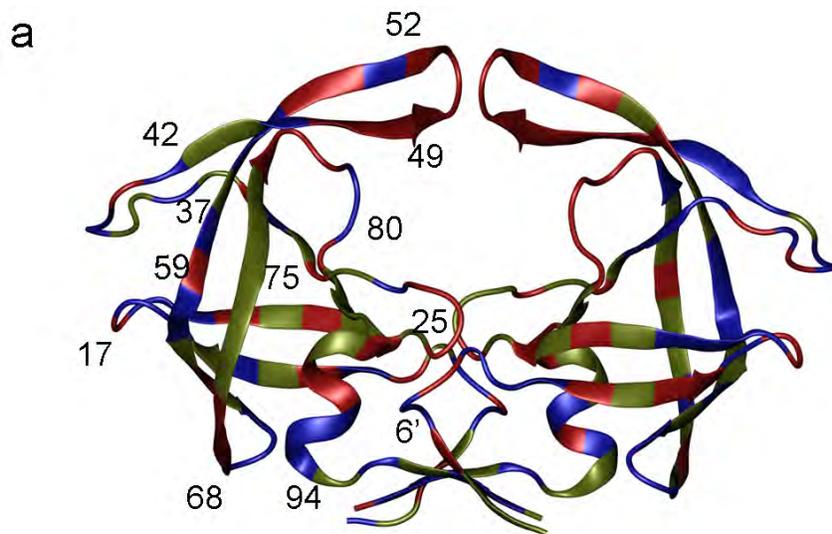


Figure 3-13. a) A semi-open structure colored by the magnitude of the change in the average SASA of the corresponding residues upon flap opening: blue indicates an

increase in residue SASA ($\Delta\text{SASA} > 2 \text{ \AA}^2$); red indicates a decrease in residue SASA ($\Delta\text{SASA} < -2 \text{ \AA}^2$); and gray indicates no significant changes in SASA ($-2 \text{ \AA}^2 < \Delta\text{SASA} < 2 \text{ \AA}^2$). The exact values of the changes in the average SASA of residues on monomer A is shown in b, and monomer B in c.

Taken together with the significant decrease in the intermolecular vdW interaction energy for the dimer interface residues when the flaps open (Figure 3-12), it is conceivable that water molecules may enter the dimer interface region during the opening event. Indeed, this hypothesis is supported by the visualization of the snapshots during the flap opening event (Figure 3-14, a top view from the active site). It is apparent that along with the opening of the flaps, the dimer interface partial dissociated, resulting in a large expansion of the binding cleft, thus making it more accessible to the solvent.

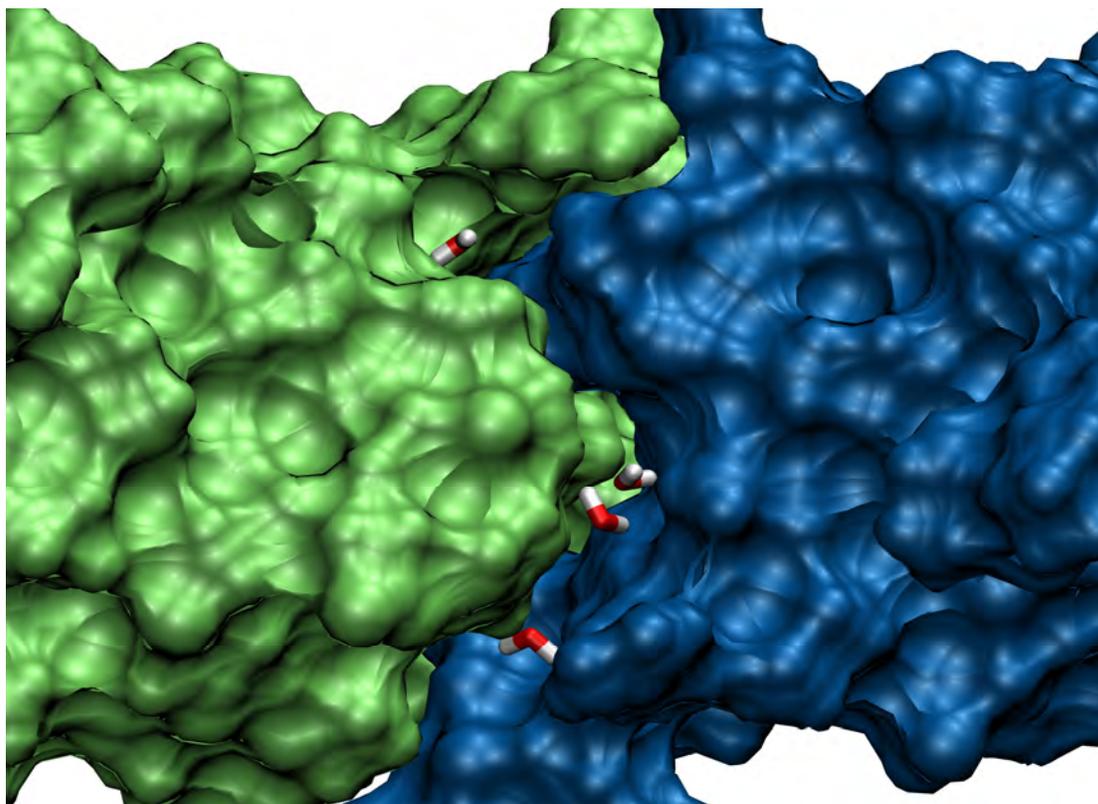


Figure 3-14. Top view of the active site in a fully-open structure (monomer A colored in blue; monomer B in green) from the simulation. The Figure was drawn with VMD.²⁴¹

It is apparent that other than the flap regions, the entire dimer interface partially disassociated, resulting in an expansion of the binding cleft between the two monomers. Particularly, the dimer interface formed between the highly conserved α -helix (residues 86-94) and the β -loop (residues 4'-9') was completely disrupted, thus allowing passage of water molecules, as illustrated in Figure 3-15.

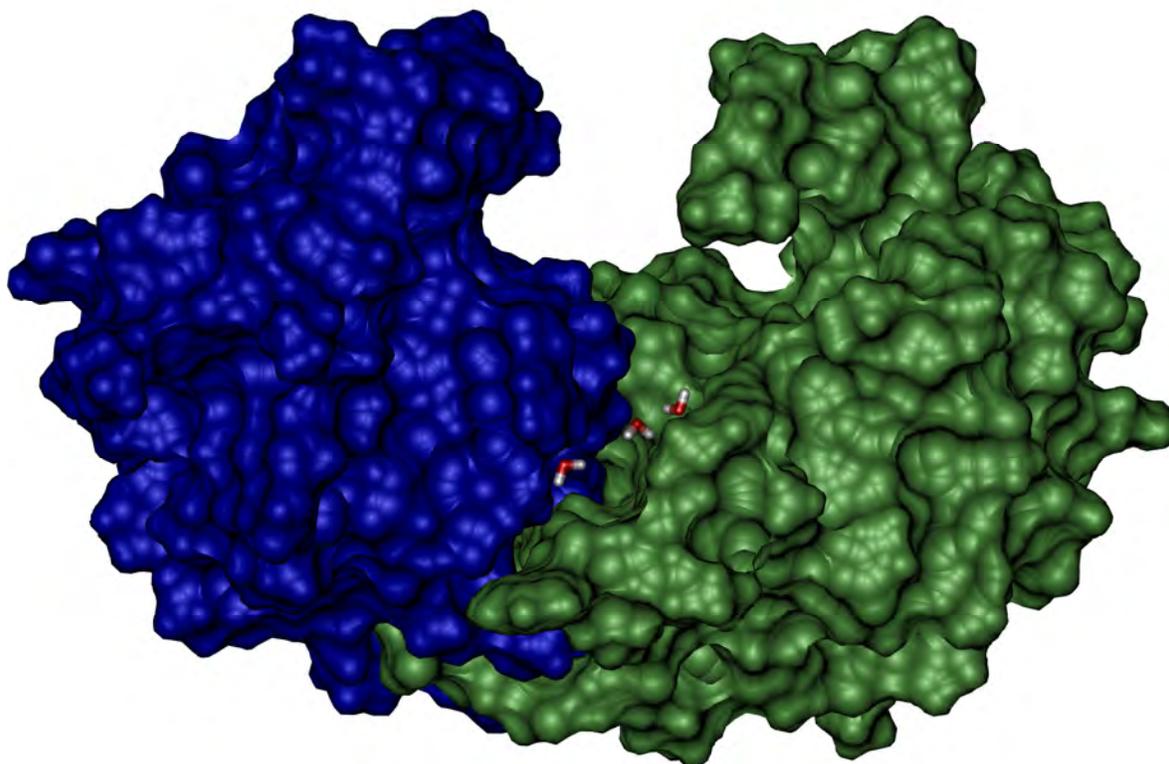


Figure 3-15. A novel potential allosteric site identified in the fully-open form of HIV-1 PR captured in the MD simulation, defined by residues 4'-9' on the β -turn from one monomer (green), residues 22-29 on the active site loop and residues 87-91 on the α -helix from the other monomer (blue). Also shown are water molecules that enter this dimer interface region during the opening event.

The weakness of the dimer interface at this region is likely due to the solvent exposed β -turn, which demonstrates conformational fluctuations on the μ s-ms timescale in NMR relaxation studies.¹¹¹ Taken together with our above energy decomposition analysis: the intermonomer vdW interaction energy of the residues located at this dimer

interface region demonstrates a strong correlation with the vdW interaction energy of the dimer; it experiences a significant loss of favorable vdW interactions when the flaps open, we suggest that the binding strength of the dimer interface encompassing the α -helix and the β -turn is likely a vulnerable region across the dimer interface, thus more prone to be disrupted.

3.3.7 Further Theoretical and Experimental Evidence

On the basis of the above analyses, we hypothesize that it is the intersubunit interactions that govern the gating dynamics of the flaps; the opening may result from the partial dissociation of the dimer. Therefore, it can be reasoned that changes in the strength of intersubunit interactions via mutations might not only affect the dimer stability, but also alter the frequency and the rate of flap opening, thereby changing the rate for association of HIV-1 PR with a ligand.

This hypothesis is supported by experimental observations that the active PR (with one of the aspartic acids protonated and the other deprotonated)²⁴² and inactive PR_{D25N}¹¹² demonstrated different structural stabilities and kinetics of ligand binding. In addition to reducing the binding affinity of the dimer in PR_{D25N}, the D25N mutation increases the association rate (k_{on}) of ligand by a factor >100-fold relative to the active PR_{D25}. Although the difference in the K_d and k_{on} between the active and inactive PRs could be caused by the different ligands, temperatures, and concentrations of protein and substrate used in these experimental studies, it is possible that distinct kinetics in the active and inactive

PRs are due to different strengths of the interactions, formed between the protonated Asp and the deprotonated Asp, and formed between the two Asn at the active site.

Moreover, several experiments have shown that the introduction of D29A and R87K to HIV-1 PR disrupts the dimer interface interactions and therefore depletes its catalytic activity.^{97, 218} As discussed in Chapter 1 (Figure 1-9), the interactions formed by the residues between the two monomer subunits involve an intra-molecular salt bridge formed between Asp29 and Arg87, an inter-molecular salt bridge formed between Asp29 and Arg8' and the intermonomer H-bonds between the guanidinium of Arg87 and the carbonyls of Leu5'/Trp6'.¹⁵⁰ To elucidate the effects of the mutations within this dimer interface region in the opening/gating dynamics of the flaps, we carried out two additional simulations by introducing two substitutions D29A and R87K to the wild-type sequence (PDB code 1TSU¹²⁵), such that both intra-/inter- monomer salt bridges are eliminated. The two double mutant simulations (PR_{D29A/R87K}) were initiated from the closed and semi-open structures, respectively, and prepared using the protocol described in the Methods section of Chapter 2 for the two single mutant systems (I50A and I50W). Consistent with the wild-type system, the flap dynamics was monitored by flap C α RMSD with respect to the two crystal structures, closed and semi-open, and the fully open configuration obtained from the wild-type simulations. The extent of the dissociation of the dimer was also measured by the distance between the core domains of the two monomers.

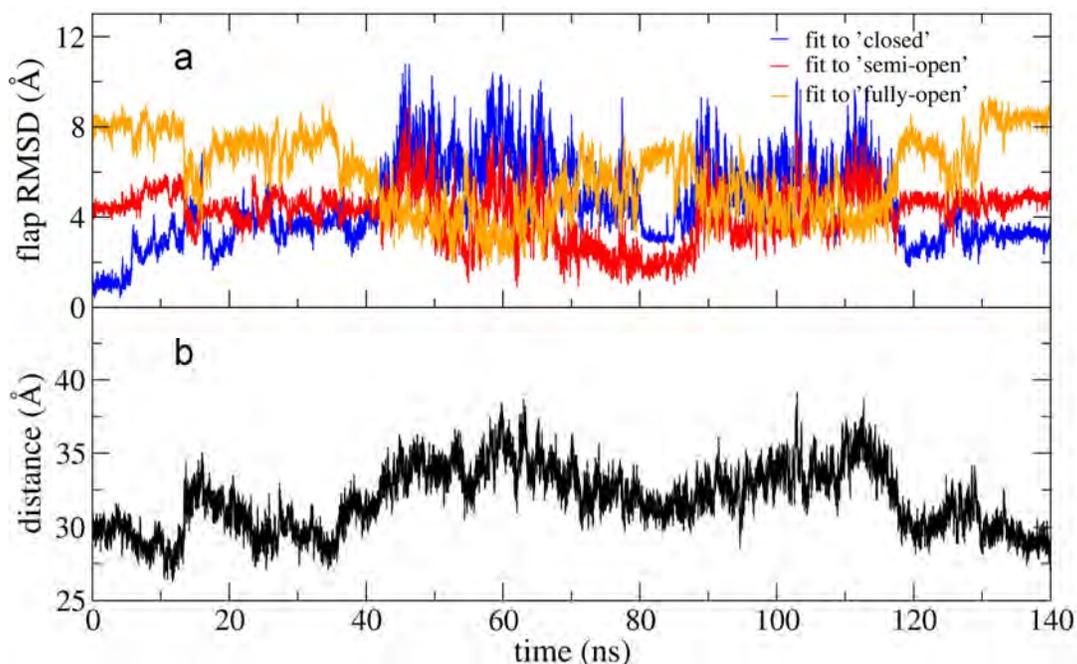


Figure 3-16. Time evolution of a) flap C α RMSDs with respect to the three references, the closed, semi-open crystal structures, and a fully open configuration obtained from the previous MD simulation on the wild-type with the ITSU sequence; and b) core domain distances during the PR_{D29A/R87K} simulation starting from the closed state.

As demonstrated in Figure 3-16a, the flaps deviated from the initial closed conformation at the very early stage of the simulation (~5ns), and rearranged to a flexible ensemble of fully open configurations after ~42ns, as evidenced by low flap C α RMSD values from the open reference obtained from the wild-type (ITSU) simulation (orange line). Note that the flaps reached an ensemble of semi-open states after ~68ns (red line) and eventually returned to the closed state after ~120ns (blue line). In the meantime, as demonstrated in Figure 3-16b, accompanied with the initial conformational changes of the two flaps, the core domain distance fluctuated between 26Å-32Å, and gradually increased to larger extent (~35Å) before the flaps rearranged the fully-open conformation at ~42ns. Notably, the two monomers approached to each other before the flaps re-closed

at ~120ns with the core domain distance decreased to the same value as that in the initial closed state.

In the simulation initiating from the semi-open flap conformation (Figure 3-17), the flaps rearranged to a flexible ensemble at ~6ns before reaching fully-open configurations around 23ns (orange line). Thereafter, the flaps demonstrated substantial flexibility, and revisited the semi-open state multiple times, with flap C α RMSD values to the semi-open reference fluctuating between 2Å and 7Å until the end of this 115ns simulation (red line). Meanwhile, the core domain distance also exhibited large-scale fluctuations. Again, the two monomers were separated very sharply from each other with their core domain distance increasing up to ~40Å before the flaps reached the fully-open state. The two monomers were brought closer together when the two flaps approached each other, and thereafter adopted the semi-open state.

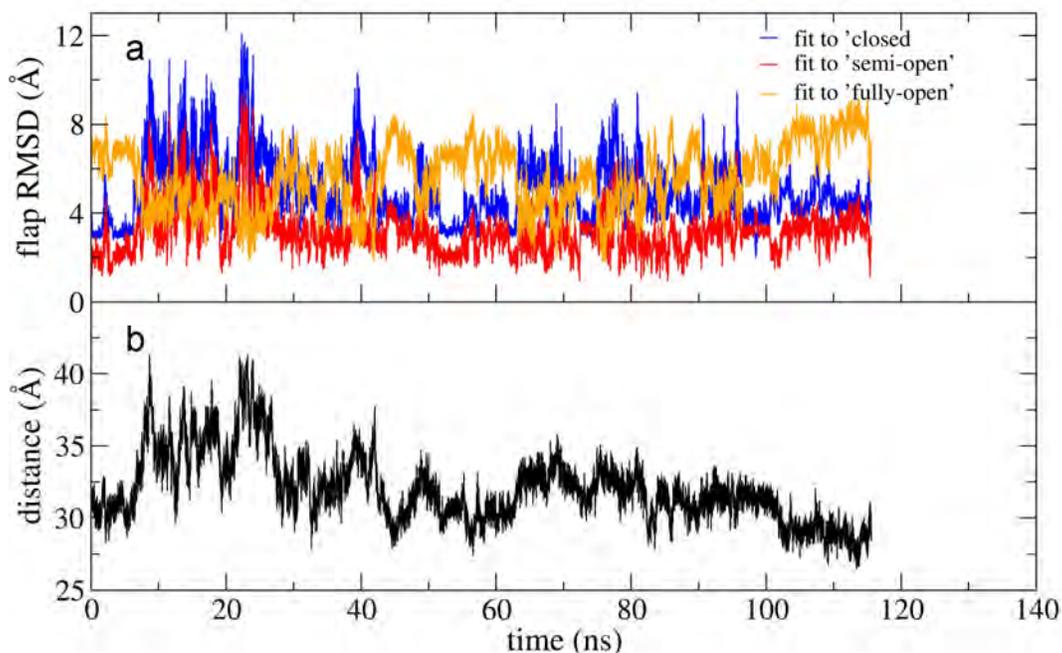


Figure 3-17. Time evolution of a) flap C α RMSDs and b) core domain distances in the PR_{D29A/R87K} simulation starting from the semi-open state.

Thus, both the wild-type and mutant simulations support that there is a strong correlation between the dimer stability and the flap opening dynamics; changes in the binding strength of the dimer interface via mutations affect the gating/opening dynamics of the flaps. In addition, the greatly enhanced flexibility and distinct dynamics of the PR_{D29A/R87K} mutant simulations suggest that the interactions involving these residues are critical in the dimer stability. Notably, this finding is also supported by the experimental data^{97, 218, 243}, which have showed that Arg87 is critical for the monomer-dimer equilibrium of the mature PR; mutations Leu5, Asp29, and Arg87 increase the dissociation constant of the dimer, resulting in drastically reduced catalytic activity. Therefore, both experimental and theoretical results reveal the significance of the interactions within the dimer interface formed between the highly-conserved α -helix (residues 86–94) and the solvent exposed β -loop (residues 4’-9’) in the protein structural stability and dynamics.

3.3.8 Biological Implication: A Potential Allosteric Site

The crucial role of HIV-1 PR in the viral life cycle has made it an important therapeutic target. One strategy is to design competitive inhibitors which bind at the active site to cause the HIV infected cell to release immature and non-infectious particles.²⁴⁴ However, the efficacy of these drugs is limited due to the emergence of drug-resistant HIV-1 variants.¹⁸⁰ Hence, recent drug discovery efforts have been changing focus from the active site and seeking other regions of the PR structure as anti-PR targets, such as the dimer interface at the N- and C- termini.^{171-172, 245-246} Another attractive

alternative approach is designing allosteric inhibitors, which do not compete for the same binding site with substrates, and would target the thermodynamic balance of the closed, semi-open, and open ensembles.¹²⁶ Such allosteric inhibitors would have the virtue of avoiding selective pressure for the PR active site to mutate,²⁴⁷ and in combination with active site inhibitors, these allosteric inhibitors would likely increase the number of PR mutations required for significant clinical resistance to HAART.¹⁷⁵

In our working model, the anticorrelated behavior between the flap opening and the compression of the elbow, fulcrum and cantilever regions has been well characterized, thus substantiating the potential of these regions as allosteric sites¹¹⁸⁻¹¹⁹ to modulate the flap dynamics. More importantly, a novel attractive allosteric site could also be suggested from our model; a small molecule binding against the highly conserved yet weak region of the dimer interface in the fully-open configuration, encompassing the highly-conserved α -helix (residues 86–94) and the solvent exposed β -loop (residues 4'-9'), would trap the enzyme in this 'inactive' conformation, and thus inhibit its catalytic activity.¹²⁶

3.4 Conclusions

In this study we present a novel, yet plausible, working model of the flap-opening mechanism captured in a microsecond timescale simulation with an explicit solvent model. On the basis of the above structural and energetic analysis, we suggest that the highly conserved dimer interface is a critical element not only structurally, but also functionally; the full flap opening event is likely an intermediate state along the path of

dissociation/association of the HIV-1 PR dimer, owing to the fact that the dimer is at equilibrium with the monomer. The significance of the inter-subunit interactions across the dimer interface to the gating dynamics of the flaps is not only supported by previous experimental observations, but also confirmed by our subsequent simulations on a double mutant dimer with reduced binding affinity at the α -helix and β -loop dimer interface, in which the flaps open much more readily. In addition, our model not only provides additional support for the potential of the flap elbow, fulcrum and cantilever regions as allosteric sites, but also predicts a novel attractive target for allosteric inhibition of HIV-1 PR; a molecule binds to the critical and highly conserved region of the dimer interface in the fully open form may trap the enzyme in an inactive conformation, thereby preventing substrate binding.

Chapter 4

Solution Structure of HIV-1 PR Flaps Probed by Comparison of Molecular Dynamics Simulation Ensembles and Electron Paramagnetic Resonance (EPR) Spectra

Abstract

The introduction of multidrug treatment regimens has dramatically prolonged the progression and survival of AIDS patients. However, the success of the long-term treatment has been hindered by strains of HIV that are increasingly resistant to inhibitors of targets such as HIV protease (HIV-1 PR). Therefore, the need for a thorough understanding of the structure and dynamics of HIV PR and how these are altered in resistant mutants is crucial for the design of more effective treatments. Crystal structures of unbound HIV PR show significant heterogeneity and often have extensive crystal packing interactions. Recent site-directed spin labeling (SDSL) and double electron-electron resonance (DEER) spectroscopy studies characterized flap conformations in HIV-1 protease in an inhibited and uninhibited form and distinguished the extent of flap opening in an unbound form. However, the correlation between EPR-measured interspin distances and structural/dynamic features of the flaps has not been established. In this report, we link EPR experiments and MD simulations to gain insight into the ensemble of HIV PR conformations sampled in solution, both in the presence and in the absence of an

FDA-approved HIV PR inhibitor. We find that the trends in the spin label distance distributions obtained from EPR data for bound and unbound HIV PR are only reproduced by a simulation model in which the protease significantly changes conformation upon binding. Furthermore, the longest spin label distances are only sampled by fully open HIV PR structures transiently observed during MD.

Acknowledgments

The material presented in this chapter has been published (Fangyu Ding, Melinda Layten and Carlos Simmerling, (2008) *J. Am. Chem. Soc.* 130(23):7184-7185). This chapter contains direct excerpts from the manuscript (presented with minor modifications) written by Fangyu Ding with suggestions and revisions from Professor Carlos Simmerling. Ritonavir parameters were generated by Melinda Layten using the antechamber module of AMBER with AM1/BCC charge assignment. We thank Gail E. Fanucci's group for helpful discussions and providing their experiment data for comparison, and Ian S. Haworth for offering the parameters of the nitroxide spin label.²⁴⁸

4.1 Introduction

The introduction of multidrug HIV treatment regimens has dramatically prolonged the progression and survival of patient. However, the success of the long-term treatment has been hindered by the increasing drug-resistant strains of HIV-PR; therefore, the need of complete understanding the structure and dynamics of HIV-PR is still crucial for the design of novel inhibitors of this enzyme. At present, diverse structural techniques have

provided valuable structural information about HIV-PR, including X-ray crystallography, nuclear magnetic resonance (NMR), electron paramagnetic resonance (EPR) and molecular dynamics (MD) simulations. Crystal structures of all ligand-bound proteases are homogeneous,^{90,249} showing the two flexible glycine-rich β -hairpins, the so-called “flaps”, interacting with the ligand and completely blocking access to the active site (Figure 4-1a).

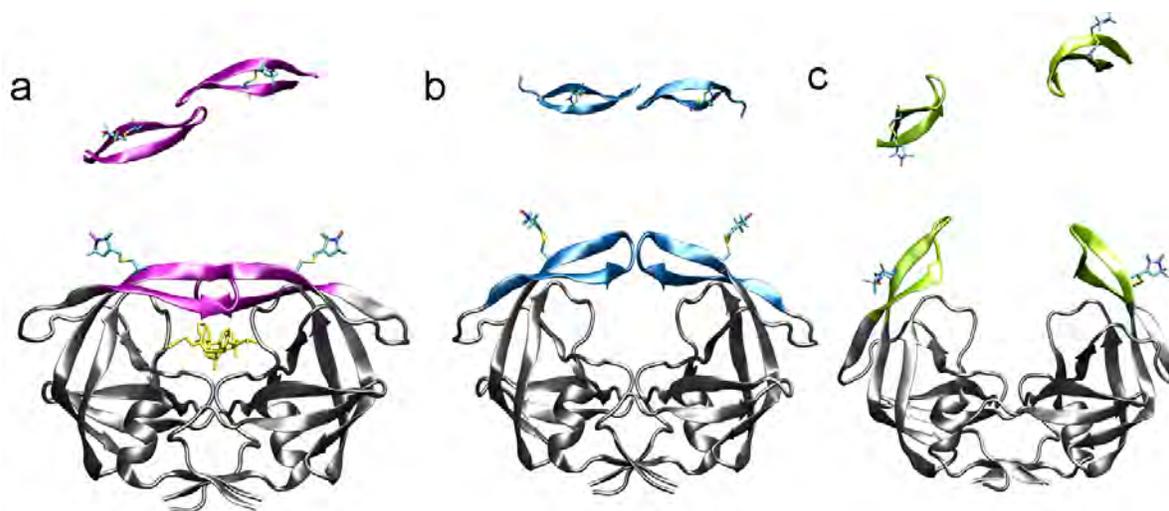


Figure 4-1. Three conformations of HIV PR during all-atom MD with EPR spin labels: a) closed Ritonavir bound; b) semi-open unbound; and c) fully open unbound. Top views illustrate the reversal of handedness between the closed and semi-open form and the separation of open flaps.

In contrast, crystal structures of apo HIV-1 PR reported to date are more heterogeneous,²⁵⁰ nearly all exhibit the “semi-open” form (Figure 4-1b), although “closed” and “wide-open” forms have also been reported. Interestingly, active site access remains blocked in both the closed and semi-open forms, thus large-scale flap opening is presumably required to allow substrate entry. However, we recently demonstrated that the crystallographic wide-open structure¹⁸² may be an artifact of the extensive

interactions between symmetry-related neighbors. Furthermore, this crystal structure differs substantially from the transient open form we observed in our previous¹²³ and the present study (Figure 4-1c). Earlier studies also suggested a role for crystal packing in the semi-open form.^{94, 105, 251} Other calculations have suggested that the free energy difference between the semi-open and closed conformations may be quite small,¹⁰⁶ implying that the equilibrium of different configurations of the flaps might be easily shifted by many factors such as mutations, ligand binding, and even crystal contacts.

Although the relationship between the conformational flexibility and catalytic activity is still unclear, it has been suggested that the mutation might affect the flexibility of the ligand-free enzyme, for example, M46I mutation appears to stabilize the closed form.²³⁵ To date, obtaining structural data on the ensemble of structures adopted by the flaps in solution is not readily accessible to experiment. Solution NMR studies on unbound HIV PR indicate that the flap tips experience rapid (nanoseconds) local fluctuations, while larger motions of the entire flaps occur on the microsecond timescale, suggesting that flaps stay in a dynamic equilibrium among different conformations.¹¹¹ Recently, Fanucci's group¹¹³ performed site-directed spin labeling (SDSL) to derive conformational flexibility of the flaps in the absence and presence of inhibitor (Ritonovir), via electron paramagnetic resonance (EPR) spectroscopy measurements of dipolar coupling of the unpaired nitroxide electrons in spin labels attached to K55C/K55'C on the flaps of LAI consensus sequence. This work is particularly notable since it revealed a markedly different extent of label flexibility in the bound and unbound forms, with an interspin distance distribution that is narrower and has a shorter average in the inhibitor-bound as compared to unbound protease.

The distance measured by SDSL is based on the dipolar coupling between two unpaired nitroxide electrons, which are located ~ 7 Å from the C_{α} atom of the protein backbone (Figure 4-2). Thus the information obtained from this technique reports only indirectly on the behavior of the flaps themselves. It is likely that the observed label distributions report on flap dynamics, rather than changes in the label as a result of inhibitor binding. The shift in distribution in the presence of inhibitor could reflect the rearrangement of the flaps from semi-open to closed handedness (Figures 4-1a and 1b, top) or could arise from decreased flap motion due to direct interactions between flaps and inhibitor (Figure 4-1a). However, the successful interpretation of SDSL-EPR data and potential application to drug-resistant HIV PR requires additional data concerning which specific flap conformations give rise to particular ranges of spin label distances, and how these ensembles are affected by inhibitor binding. Importantly, it is unclear whether the observed interspin distance distribution can be explained solely by an ensemble consisting of conformations seen in the various crystal structures. Therefore, establishing a correlation between EPR-measured interspin distances and structural dynamic features of the flaps is essential in the interpretation of the current and future EPR data for this system.

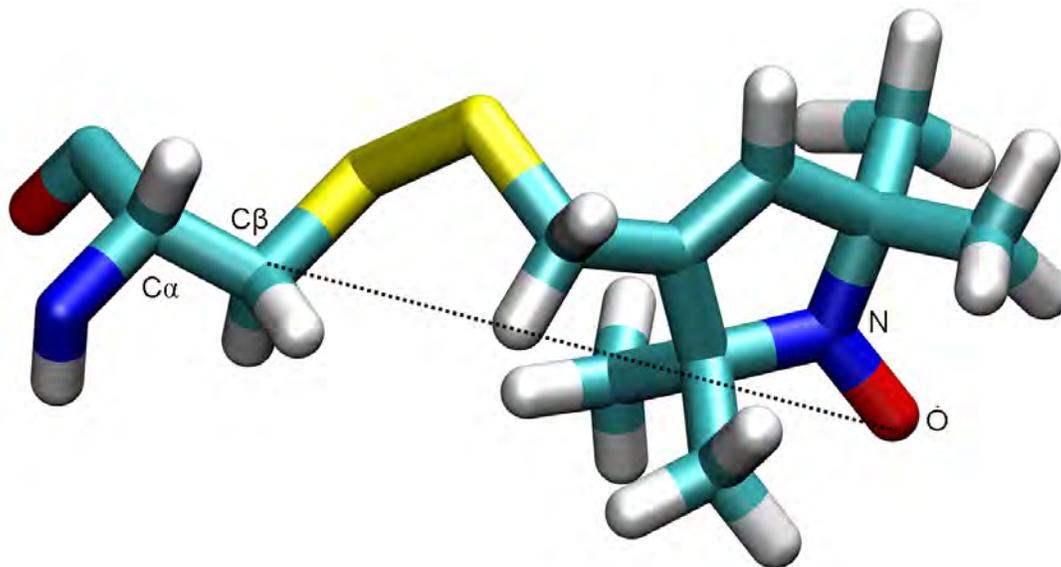


Figure 4-2. Structure of a nitroxide spin-label side chain with the distance from the C_{β} atom to the spin label indicated.

We have previously shown that our simulation model of HIV PR is able to accurately reproduce a spontaneous change between semi-open and closed handedness upon addition or removal of a cyclic urea inhibitor.^{45, 123} We employed this model, with addition of spin label probes²⁴⁸ to the simulated HIV PR, for comparison against EPR-based data in order to determine the ensemble of conformations that best agrees with EPR data.

In this present work, we performed a series of MD simulations, in fully explicit solvent with the ff99SB protein force field,⁶⁷ using the LAI consensus HIV PR sequence, which contains several mutations to match the EPR experiments (Q7K, L33I, L63I, 67A, C95A, D25N and K55C for MTSL attachment), and two methanethiosulfonate (MTSL) spin labels attached via disulfide bonds at C55 and C55', respectively. In the Ritonavir-bound simulations, the N–N distance exhibited similar restricted fluctuations, in agreement with EPR measurement of the bound HIV-1 PR. In unbound semi-open,

significant but transient opening and closing events were observed, giving further evidence for the considerable variability of this flap configuration. A significantly longer and wider N–N distance distribution was obtained, in very good agreement with EPR measurements of unbound HIV PR. Thus, our results strongly support the hypothesis that the flaps in unbound form exist in an ensemble of conformations between ‘semi-open’, ‘closed’ and ‘open’ conformations, and exhibit considerable flexibility to allow substrate entry and product exit. Moreover, the combination of our MD simulations and EPR spectra suggested that the semi-open form is likely the dominant configuration for this ligand-free HIV-PR in solution, and provided further evidence that a rearrangement of the flap region occurs upon ligand binding.

4.2 Methods

4.2.1 Construction of the Modeled Structures

The coordinates for Ritonovir (RIT)-bound simulation were obtained from a complex X-ray structure (PDB code: 1HXW²⁵²). In order to identify the dominant conformation of the flaps in this free enzyme which can exhibit substantial flexibility and reproduce the EPR-measured interspin distances, simulations for the unbound protein were initiated from two different conformations: closed and semi-open. The coordinates for semi-open simulations were obtained from the semi-open crystal structure (PDB code: 1HHP¹⁹⁵) for apo HIV-PR; the coordinates for closed simulations using the crystal structure with Ritonavir bound complex (1.8Å; PDB code: 1HXW²⁵²), with the inhibitor

stripped out prior to the simulation. Both catalytic Asp residues were modeled as Asn to be consistent with the EPR experiments, which was incorporated to prevent protein degradation during the time course of data collection. Additionally, each system contained the same mutations as introduced in the EPR experiments: three mutations that provide protection from autocatalytic cleavage, Q7K, L33I, L63I,²⁵³ and mutations C67A and C95A to avoid unspecific spin labeling as well as disulfide crosslinking. The construct was referred as PMPR in the EPR experiments.¹¹³ Site K55C was chosen to attach methanethiosulfonate (MTSL) spin labels via disulfide bonds. All mutations were modeled by Swiss-PdbViewer.¹⁹⁸ Ritonovir parameters were generated using the antechamber module of Amber with AM1/BCC charge assignment. The force field parameters of the nitroxide spin label was provided by Haworth's group.²⁴⁸ Amber parameter files are included with Appendix Information. Hydrogen atoms were added using the Leap module in the AMBER 9 software package.¹⁹⁶

4.2.2 Minimization and Equilibration

We performed minimizations and MD simulations in a manner similar to that described in the Methods section of Chapter 2. All calculations were performed using the Amber 9 program package¹⁹⁶ and the ff99SB modification⁶⁷ of the Amber ff99 force field.⁶⁶ Minimizations and MD simulations were carried out using the Sander module of the AMBER 9 package.¹⁹⁶ For all simulations, temperature was maintained at 300K using the Berendsen algorithm.²⁰⁰ The SHAKE algorithm was used to treat the bonds involving hydrogen.¹⁹⁹ The long-range electrostatic interactions were calculated by the particle

mesh Ewald (PME) method²⁵ with a dielectric constant 1.0. A cutoff of 8Å was used to calculate the direct space sum for PME.

To avoid steric clashes caused by the introduced mutations, the structures were first subject to a stepwise minimization and equilibration in the presence of the implicit solvent model (using the ‘mbondi2’ radii and ‘igb = 5’). Energy minimization was achieved in four steps. First, movement allowed only for the mutated residues; while the protein (i.e., the coordinates of those atoms are experimentally determined) were positionally constrained (not fixed) using a harmonic potential with a force constant of 50kcal/(mol•Å²). Next, all heavy atoms of the protein were restrained with forces of 10kcal/(mol•Å²). Then the position restraints were only imposed on the backbone with forces of 1kcal/(mol•Å²). In this step, steric collisions of the automatically generated residues were minimized, and favorable configurations of the side-chains of the mutated residues were obtained while the experimentally determined coordinates were maintained. Finally, all atoms were free to move. Next, the purpose of the subsequent MD simulation is to extensively relax the configuration of the placed mutations by adding the thermal fluctuation. For that purpose, the temperature was raised from 100K to 300K over 50 ps in 10-ps intervals using Langevin dynamics. Only the mutated residues were permitted to move freely in the energy minimization, while the experimentally determined coordinates were positionally constrained with a force constant of 50kcal/(mol•Å²). This was followed by an equilibration phase of 200 ps under constant pressure (NPT) at 300K using a coupling constant of 0.5 ps, with restraints on the experimentally determined coordinates, and gradually reduced from 10 to 1, 0.1 and 0 (mol•Å²) in 50-ps intervals.

Then each system was solvated in a truncated octahedron periodic box containing 6000–6500 TIP3P water molecules, with the box extending 8Å from the extremes of the solute. Energy minimization was achieved in five steps. The solvent molecules were firstly relaxed, while all heavy atoms in protein were restrained with forces of $50\text{kcalmol}^{-1}\text{Å}^{-2}$. Then, the systems were continually relaxed with the restraint force constant gradually reduced to 10 and $0.1\text{kcal}/(\text{mol}\cdot\text{Å}^2)$. Then only the backbone was restrained with forces of $1\text{kcalmol}^{-1}\text{Å}^{-2}$. Finally, all restraints were lifted and whole system was relaxed. In each step, energy minimization was executed by the steepest descent method for the first 10,000 cycles and the conjugated gradient method for the subsequent 10,000 cycles. After the relaxation, the systems were gradually heated to 300K during a 50 ps run under *NVT* condition using the Berendsen algorithm²⁰⁰ with a coupling constant of 0.5 ps, with all heavy atoms restrained by $5\text{kcal}/\text{mol}\cdot\text{Å}^2$. This was followed by a three-stage equilibration under *NPT* condition with a coupling constant of 1 ps and at the pressure of 1atm: i) a restrained MD simulation for 50 ps, while keeping the heavy atoms restrained with a force constant of $1\text{kcal}/\text{mol}\cdot\text{Å}^2$; ii) an additional 50 ps long MD with a restraint force constant $0.5\text{kcal}/\text{mol}\cdot\text{Å}^2$ only imposed on the backbone; iii) a short equilibration of 50 ps without any restraints was performed.

4.2.3 Production Runs

The configurations from the above equilibration stages were used as the starting configurations for the production runs. The time step for integration was set as 2 fs, and the coordinate sets were saved at every 10 ps for subsequent analyses. The production

phase of the MD simulations was ~150 ns for each of two simulations for each set. Simulation lengths are summarized in Table 4-1.

Table 4-1. Timescale of All Spin-labeled Simulations

Simulation	Bound		Unbound closed		Unbound semi-open	
	Run1	Run2	Run1	Run2	Run1	Run2
Timescale	114ns	160ns	140ns	125ns	168ns	125ns

4.2.4 Data Analysis

Distances and RMSD values were calculated using the PTRAJ module in the Amber 9 software package.¹⁹⁶ Histograms of distance distributions were constructed with intervals of 0.1 Å using the Xmgrace program. The flap conformation was monitored by RMSD of the C α atoms of the two flaps (residues 46-55, 46'-55') overlapped on the flaps of the initial closed state and of the apo X-ray crystal structure (PDB code: 1HHP¹⁹⁵), respectively. The extent of flap opening was gauged by the distance between the nitrogen atoms on the side chain of the two spin labels attached at position 55 in each subunit, termed as interspin distance. RMSD and distance were calculated from the combined set of structures from two simulations initiated in the semi-open structure (~30,000 frames). The populations in both EPR experiments and MD simulations were assigned to flap conformations of curled/tucked, closed, semi-open, and wide-open: populations with flap C α RMSDs to both closed and semi-open references of > 5Å and with the average interspin distance of 28 Å are in 'tucked' form; populations with flap C α RMSD to the closed reference of < 1.5Å and with the average interspin distance of 33Å are in 'closed'

form; populations with flap C α RMSD to the semi-open reference of $< 2.0\text{\AA}$ and with the average interspin distance of 36\AA are designated as ‘semi-open’ states; populations with flap C α RMSD to both references $> 10\text{\AA}$ and the average interspin distance of 40\AA are in ‘fully-open’ forms. These assignments are based upon extensive characterization of apo-HIV-1PR in our previous MD simulations and EPR experiments.²⁵⁴

4.3 Results and Discussion

4.3.1 Internitroxide Distances from MD Simulations and the EPR Parameters

We performed a series of MD simulations, in fully explicit solvent with the ff99SB protein force field, using the LAI consensus HIV PR sequence (with several mutations to match the EPR experiment) and methanethiosulfonate (MTSL) spin labels attached via disulfide bonds at C55 and C55'. Bound HIV PR simulations used closed coordinates. Simulations for unbound protease were initiated from both semi-open and closed coordinates. Two simulations of 150 ns were performed for each of the three systems (750 ns total).

In the labeled-bound simulations, the MTSL N–N distance was confined to $30\text{--}33\text{\AA}$, in near-quantitative agreement with EPR measurement of the bound HIV PR (Figure 4-3). Thus, both the EPR experiment and the simulations revealed that the flaps assumed the closed state with Ritonavir bound, in consistent with X-ray structures of HIV-1 PR

complexes. In addition, restricted amplitude of flap motions was observed, perhaps due to the inhibitor.

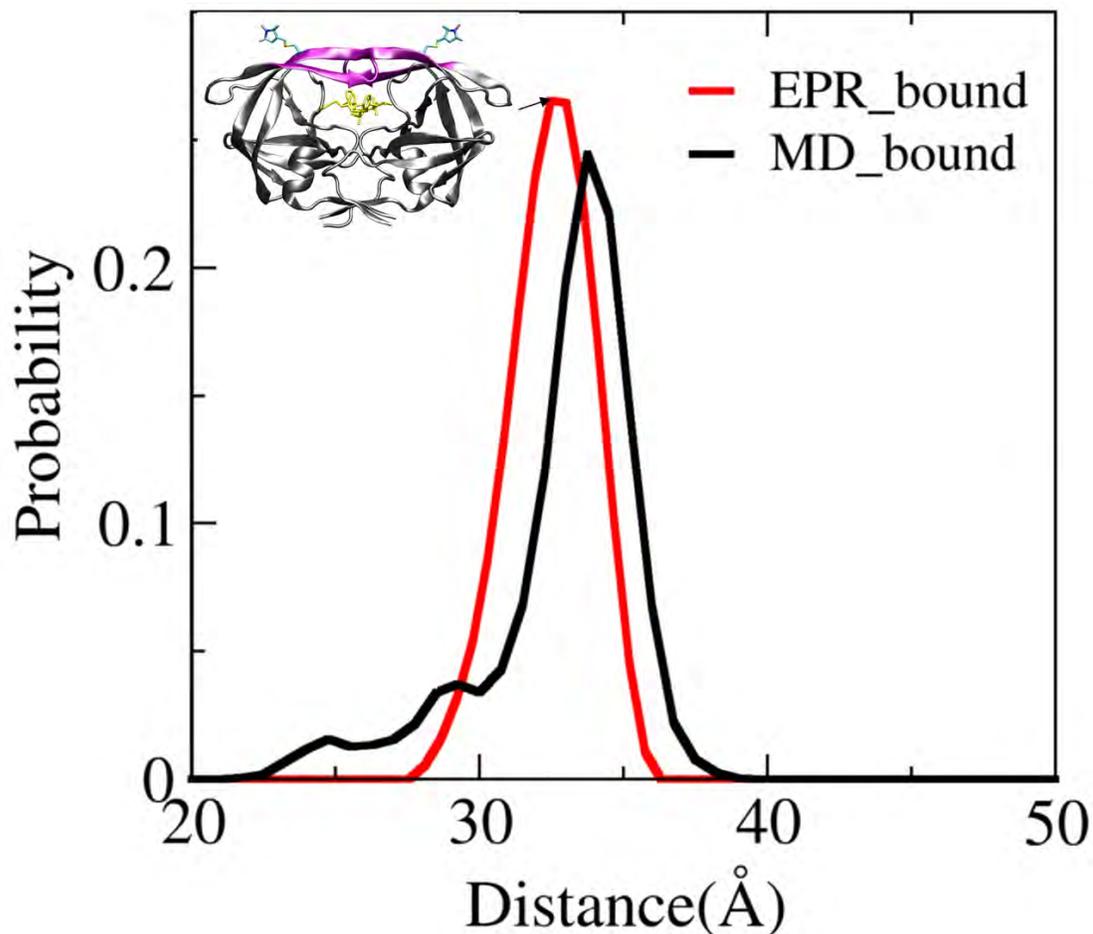


Figure 4-3. MTSL spin label distances in the Ritonavir bound complex from EPR experiment and MD simulations, respectively.

However, in the unbound closed simulations(Figure 4-4), the overall shape and breadth of the distance profile exhibited similar restricted fluctuations as in Ritonavir-bound simulations, with a similar average distance despite the loss of the flap-inhibitor interactions. This unbound closed model is in disagreement with the EPR data, suggesting that simple loss of the inhibitor is not enough to account for the observed

trends in both width and average of the nitroxide distance distribution. Interflap hydrogen bonding was observed in these simulations, contributing to the stability of the closed flaps on the >100 ns timescale of both simulations (Figure 4-5).

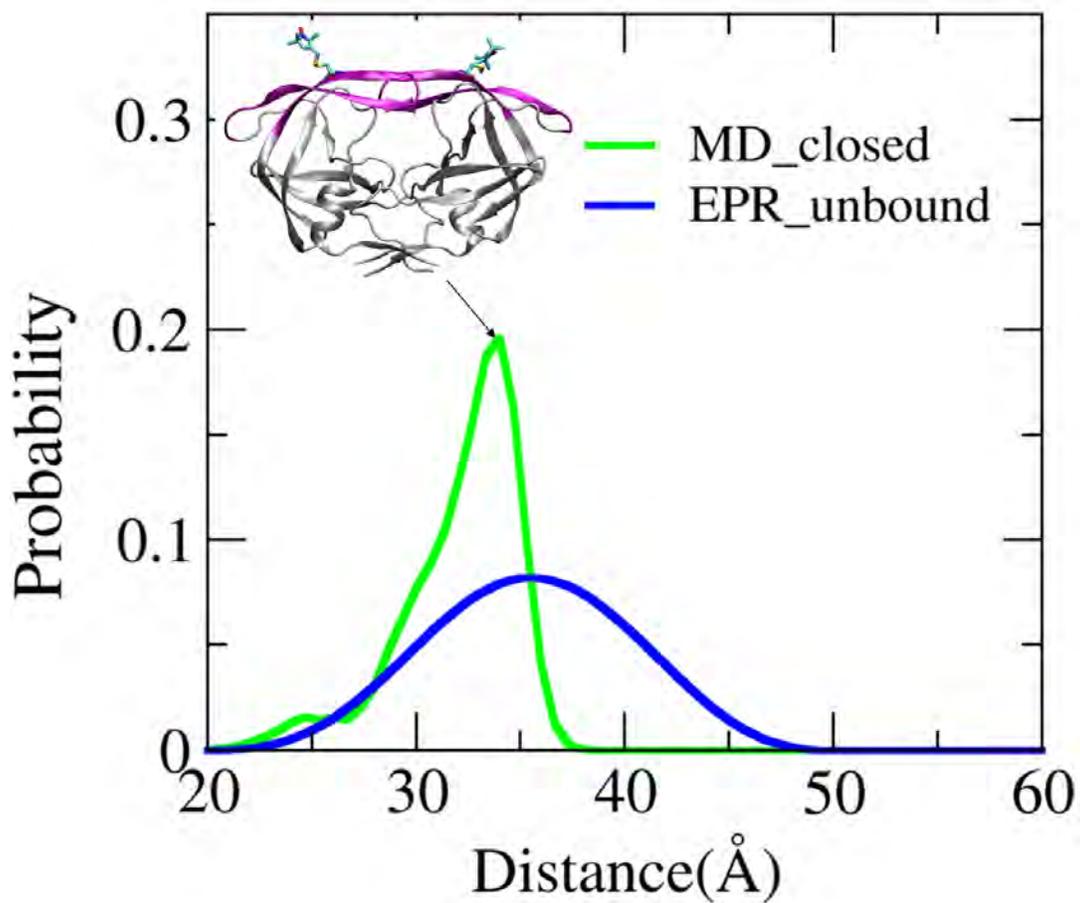


Figure 4-4. MTSL spin label distances in the unbound LAI' from EPR experiment and MD simulations starting from the unbound closed form.

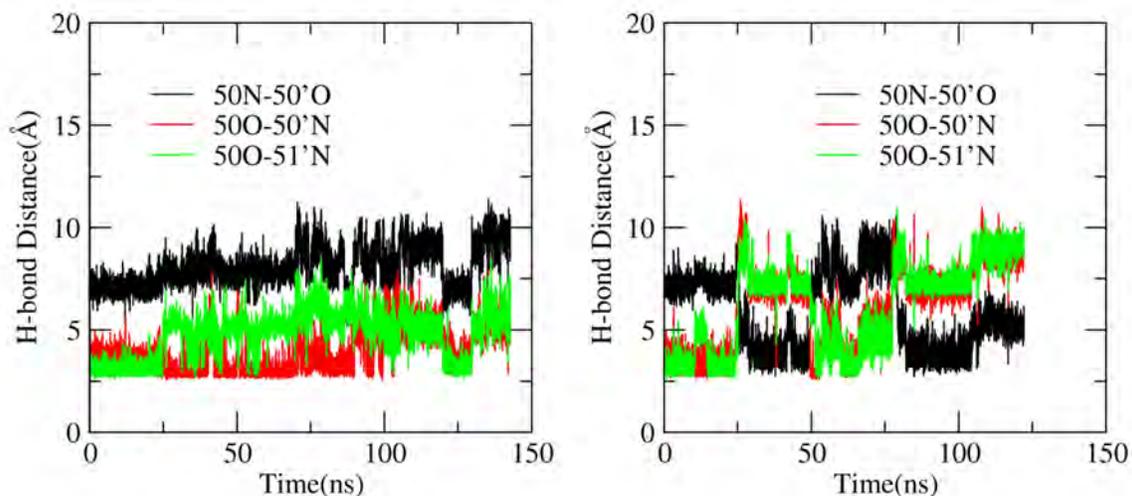


Figure 4-5. The H-bonding pattern within the flap region in unbound closed simulations.

This observation is consistent with the microsecond timescale of large-scale flap motion suggested by NMR, further indicating that nanosecond timescale flap tip motion is not responsible for the changes seen in the EPR data upon inhibitor binding.

In unbound semi-open simulations, significant but transient opening and closing events were observed, giving further evidence for the considerable variability of this flap configuration. The changes in flexibility are inferred from the breath of the distance profile. As indicated in Figure 4-7, a significantly longer and wider N–N distance distribution was obtained in the simulations, with each peak corresponding well to distinct flap conformers, ranging from the ‘tucked’, ‘closed’, ‘semi-open’ and ‘fully-open’ forms. It is worth noting that EPR spectra is sensitive to tau values; better agreement between the MD reconstructed interspin distance and the EPR measurement of unbound HIV PR was achieved when the EPR data was recollected using a tau of 1.8microsec and reanalyzed using a lambda value of 20. Note the distance profile differs slightly from the one reported before.¹⁸⁷

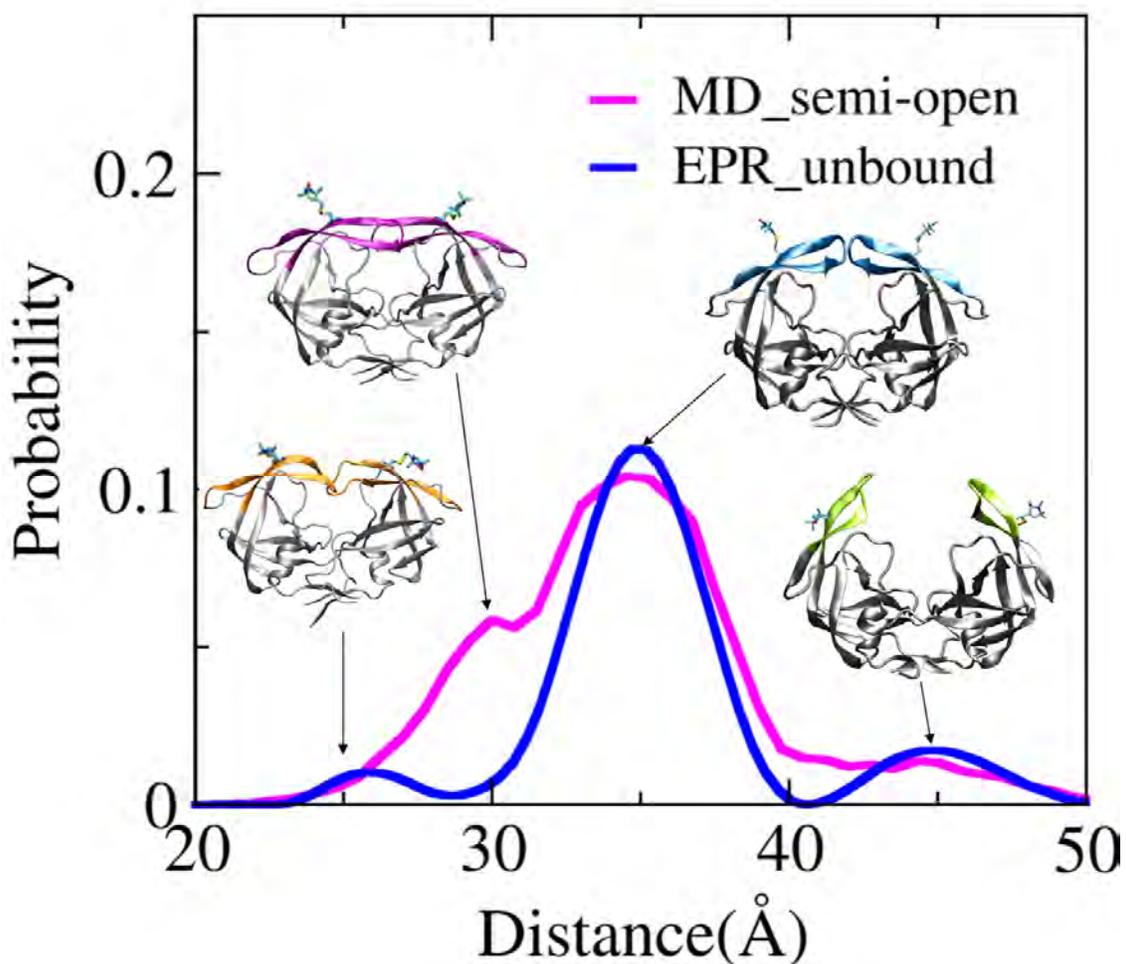


Figure 4-6. MTSL spin label distances in the unbound LAI' from EPR experiment and MD simulations starting from the unbound semi-open form.

Furthermore, we note that the motion of the nitroxide side chain also contributes to the distance distribution between the nitroxide spin labels since structures with flap backbone rmsd values of $< 1 \text{ \AA}$ still span a range of distances from 30 to 40 \AA (Figure 4-4). Importantly, MD structures with spin label distances greater than 40 \AA always had flap rmsd values of at least 3 \AA as compared to the closed, semi-open, and wide-open crystallographic forms. This suggests that the EPR-based ensemble includes flap

conformations that match *none* of the reported unbound crystal structures but can indeed be explained by full opening events as observed in MD simulations.

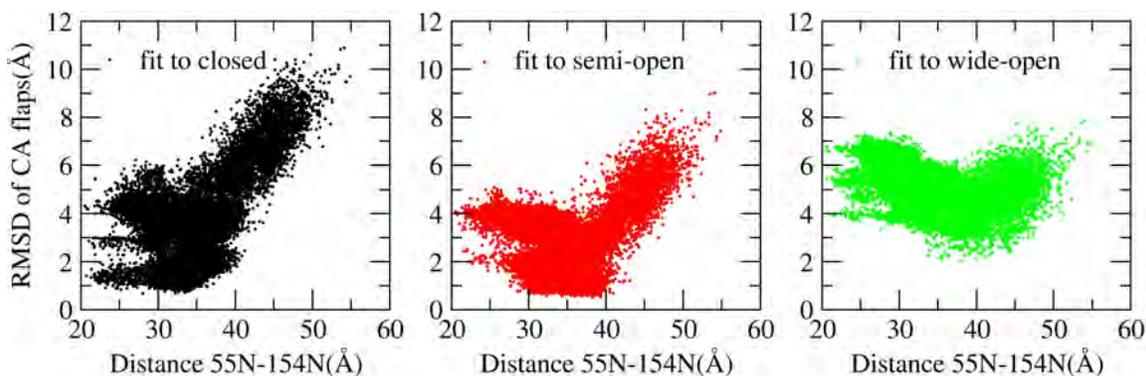


Figure 4-7. . Flap RMSD as a function of the interspin distance during the MD simulations starting from the semi-open conformation, compared with the closed, the semi-open crystal structures. While the structures are sampled that are similar to either the closed (1HXW), semi-open (1HHP) or “wide-open” (1TW7¹⁸²) structures, the simulated conformations with long label distances ($> 40\text{\AA}$) match none of these crystal structures.

Nonetheless, discrepancies in the distance distribution between MD simulations and EPR measurement still remain. First of all, structures with the average distance of 33\AA , corresponding to the closed states, are present in the simulations; whereas, there is no detectable population near that range in the EPR data. Secondly, a greater percentage of the unbound ensemble is seen in the ‘tucked’ conformations in the EPR experiment than in the MD simulation. A likely source of these uncertainties is that, even with >100 ns of simulation, the populations of different structures have not reached convergence. Noticeably, structures with long nitroxide distances ($40\text{--}45\text{\AA}$) are present in the simulations, but at a lower population than indicated by EPR data. The difference cannot be explained solely by changes in equilibrium among the crystal forms since none of those structures can sample long enough distances. Therefore, this uncertainty might be

explained by the insufficient sampling of fully-open structures in the MD population; on the other hand, the simulations modeled only the dimer, while the experiment likely also contains a population of monomeric HIV PR. It will also be important to determine whether the contribution of these “open” structures with long label distances changes as the glassing agents which are used to ensure the formation of a good glass (separate molecules better) in the EPR experiments are varied, typically ethylene glycol or glycerol. These glassing agents are used to break up the spin–spin communication in the aqueous that contributes to the relaxation time.

4.4 Conclusions

To establish the correlation between EPR-measured interspin distances and structural/dynamic features of the flaps, we performed a series of MD simulations in explicit water using the LAI consensus HIV PR sequence with methanethiosulfonate (MTSL) spin labels attached via disulfide bonds at C55 and C55'. This study provides insight into the ensemble of conformations sampled by HIV PR flaps in solution, both in the presence and in the absence of an FDA-approved HIV PR inhibitor. Overall, the interspin distances sampled in semi-open unbound simulations are in much better agreement with the EPR data than are the simulations initiated with the closed HIV PR structure, suggesting that the semi-open form is the dominant configuration for this ligand-free HIV PR in solution, at least under the conditions probed by EPR. In addition, both the simulations and the EPR spectroscopic data strongly support the hypothesis that the flaps in the unbound state exist in a diverse ensemble of conformations fluctuating

between semi-open, closed, and open, exhibiting considerable flexibility to allow substrate entry and product exit.

Chapter 5

Drug Pressure Selected Mutations in HIV-1 PR Alter Flap Conformations as well as its Dynamics

Abstract

Human immunodeficiency virus type 1 (HIV-1) protease plays a fundamental role in the maturation and life cycle of the retrovirus HIV-1, as it functions in regulating post-translational processing of the viral polyproteins *gag* and *gag-pol*; thus, it is a key target of AIDS antiviral therapy. Accessibility of substrate to the active site is mediated by two flaps, which must undergo a large conformational change from an open to a closed conformation during substrate binding and catalysis. In the present work, the flap conformations of two drug-resistant HIV-1 protease constructs were characterized by molecular dynamic (MD) simulations and distance measurements with the electron paramagnetic resonance (EPR) method of site-directed spin labeling (SDSL) with double electron-electron resonance (DEER) spectroscopy, and compared to wild-type HIV-1 PR. MD simulations accurately regenerate the experimentally determined distance profiles and provide structural interpretations of the EPR data. The combined analyses show that the average conformation of the flaps, the range of flap opening and closing, and the flexibility of the flaps differ markedly in HIV-1PR as multiple mutations arise in response to antiviral therapy, providing structural insights into the mechanism of inhibitor

resistance that can be understood as inducing shifts in the relative populations, or conformational sampling.

Acknowledgments

The material presented in this chapter has been published (Luis Galiano, Fangyu Ding, Angelo M. Veloro, Mandy E. Blackburn, Carlos Simmerling and Gail E. Fanucci (2009) *J. Am. Chem. Soc.* 131(2):430-431). This chapter contains excerpts from the manuscript (presented with minor modifications) written by Fangyu Ding and Luis Galiano with suggestions and revisions from Professor Carlos Simmerling and Professor Gail E. Fanucci. Fangyu Ding performed MD simulations and data analysis; Luis Galiano, Angelo M. Veloro and Mandy E. Blackburn conducted the experimental studies. We thank Gail E. Fanucci's group for helpful discussions and providing their experiment data for comparison, and Ian S. Haworth for offering the parameters of the nitroxide spin label.²⁴⁸ We thank Ben Dunn for helpful discussions and Ralph Weber for initial DEER data collection of V6' and MDR769'.

5.1 Introduction

The global spread of the human immunodeficiency virus (HIV) causing the acquired immune deficiency syndrome (AIDS) has evolved into an immense health problem with total estimated infection numbers ranging from 34 to 46 million people (UNAIDS/WHO 2008). Human Immunodeficiency Virus type 1 protease (HIV-1PR) is an enzyme

responsible for gag-pol processing, an essential step in viral maturation and the lifecycle of HIV-1. Inhibition of the activity of HIV-1PR results in immature virus particles that are noninfectious.²⁵⁵ The idea of inhibiting viral replication by disturbing the protease function has led to the development of a class of drugs known as protease inhibitors (PI).¹²⁸ Modern HIV combination therapies, referred to as “Highly Active Anti-Retroviral Therapy” (HAART), attack the virus with a combination of one protease inhibitor and two reverse transcriptase (RT) inhibitors.¹⁷⁸ The long-term efficacy of antiretroviral therapy, however, has led to the emergence of HIV-1 variants with reduced susceptibility to antiviral drugs.²⁵⁶ The frequency of one or more major resistance mutation in recently infected individuals increases from 3.8% to 10.2% while on treatment in a five-year period as reported from the West.²⁵⁷

Mutations at more than 20 positions in HIV-1 protease have been associated with resistance to currently available protease inhibitors (PIs). Ten of these mutations (D30N, V32I, L33F, M46IL, I47VA, G48V, I50VL, V82A/F/L/S/T, I84V and L90M) do not occur as natural polymorphisms in HIV-1 isolates from untreated persons and have been designated primary resistance mutations. Primary mutations directly confer resistance to one or more protease inhibitors, whereas secondary mutations reduce drug susceptibility or improve replicative fitness of the virus in conjunction with primary mutations or in synergistic form with other secondary mutations. Mutations can either occur at active-site or nonactive-site locations in HIV-1 protease (Figure 5-1), and can also confer different levels of resistance.

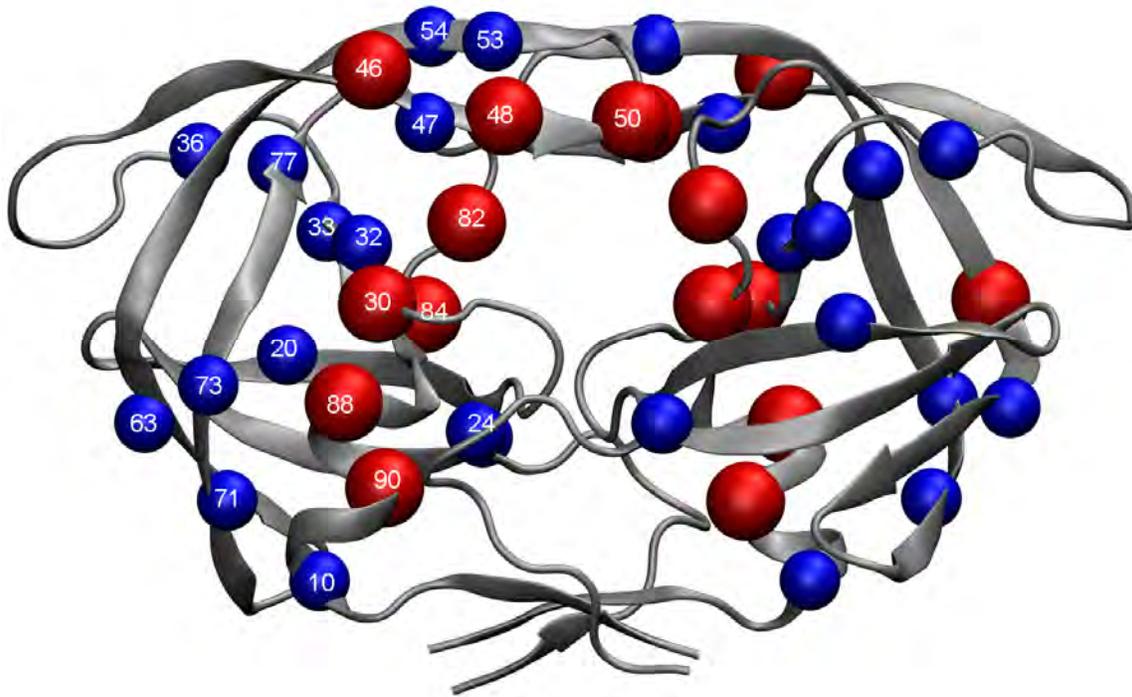


Figure 5-1. Structure of a dimeric HIV-1 protease with the C α positions of most common primary and secondary drug-resistance mutations shown as spheres. Primary mutations (30, 46, 48, 50, 82, 84, 88 and 90) are shown in red; secondary mutations (10, 20, 24, 32, 33, 36, 47, 53, 54, 63, 71, 73, 77 and 93) in blue. Both types of mutations are labeled in one of the subunits.

Active-site mutations are exclusively primary ones, but not all primary mutations are necessarily limited to the active-site (e.g. the nonactive site mutations at sequence positions 46, 88 and 90 can also directly confer resistance. It can be readily understood how mutations within the active site pocket reduce inhibitor effectiveness considering that many of the current PIs have been specifically designed to bind tightly to the shape of the active site cavity. However, the mechanism by which mutations that are *NOT* within the active site cavity modulate PI efficiency remains uncertain.

Analysis of data from the Stanford Drug Resistance Database²¹⁵ indicates that while polymorphisms in the sequence of HIV-1PR naturally occur there are regions in the

protein sequence that appear invariant under normal evolutionary pressures (Figure 5-2). These invariant regions coincide with the structural elements of the dimer interface, the active site floor, the P3–P3' substrate binding region, and the flaps. Strikingly, upon exposure to protease inhibitor (PI) cocktail treatment, numerous mutations develop, with high occurrences near residues 40–56 and 80–90, which correspond to the hairpin flaps and the P3–P3' substrate binding cleft. Amino acid substitutions arise in these regions of the protein from random mutations that alter the ability of a given inhibitor to bind as tightly to the active site pocket, allowing for effective protease function with subsequent viral maturation and proliferation of the mutation. Many of these mutations also alter the kinetics of the protease for the multiple polypeptide cleavage sequences in the gag-pol polypeptide.²⁵⁸⁻²⁶¹

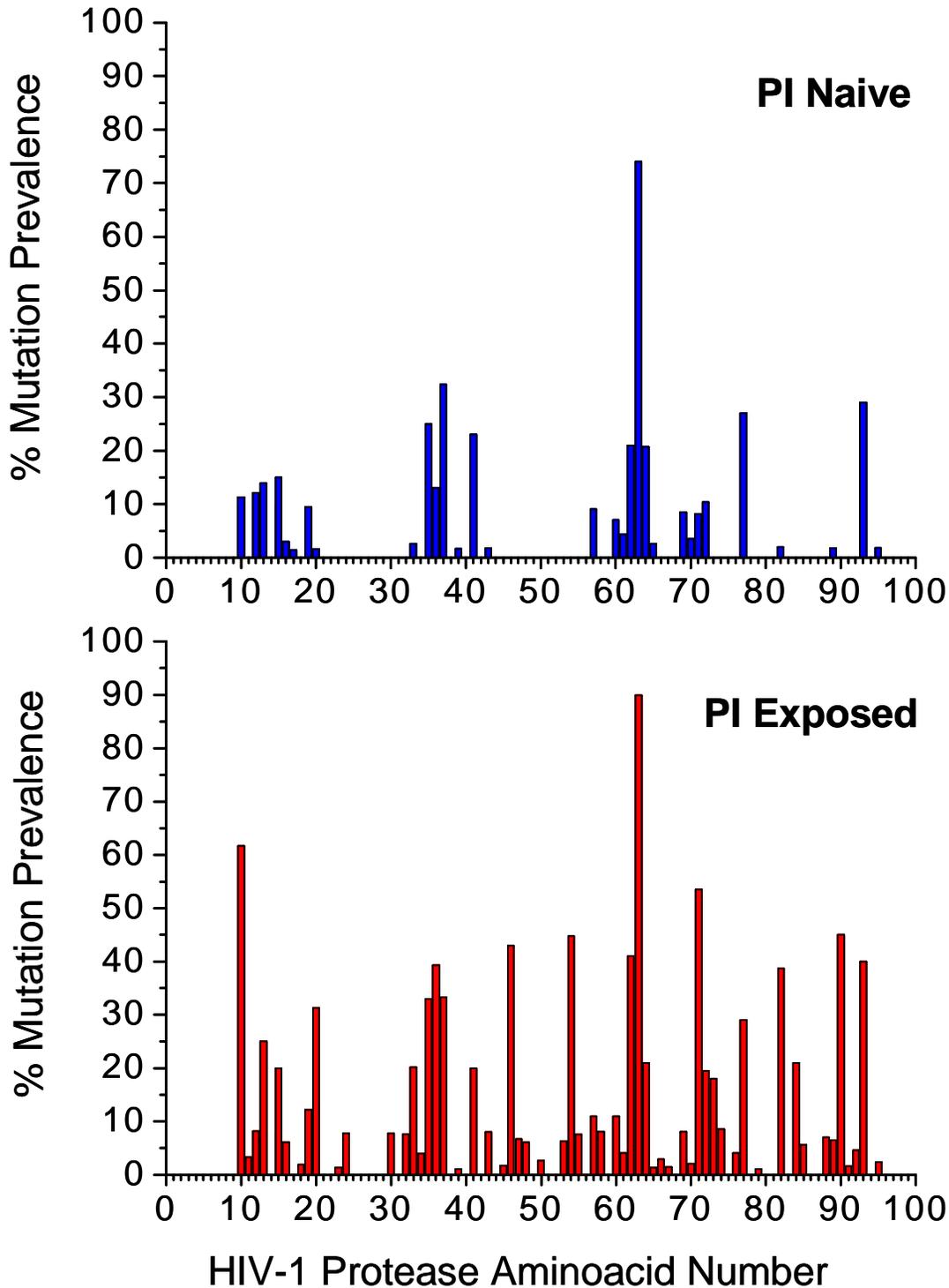


Figure 5-2. A Summary of sequence analysis showing the percentage of mutation prevalence of naturally occurring polymorphisms in HIV-1 PR subtype B for protease inhibitor (PI) naïve and exposed patient. Prevalence is defined as a measure of variability for each protease sequence vs. the subtype consensus (e.g. 0% prevalence corresponds to a residue that is conserved in all sequences for a particular subtype).

It has been hypothesized that these nonactive-site mutations are able to affect the enzymatic activity via mechanisms that have been related to differences in the conformational flexibility,²⁶²⁻²⁶³ and thus may also influence ligand binding affinity due to the presence of different amounts of “binding-competent” conformations.²⁶⁴⁻²⁶⁵ For example, mutations in the elbow and flap regions (residues 36–58) have been suggested to alter either the conformation of the flaps or closing or the mobility of the flaps, or both.^{258-259, 266} In addition, a previous MD simulation²⁶⁷ also suggested that different molecular mechanisms contribute to resistance in active-site and nonactive-site mutants, and a nonactive-site mutation (N88S) can actually shift the conformational equilibrium of the free protease owing to the alternative hydrogen bonding pattern at the site of mutation.

Previously, Fanucci’s group has shown that site-directed spin labeling (SDSL) and pulsed double electron-electron resonance (DEER) electron paramagnetic resonance (EPR) spectroscopy can distinguish between conformations of the flaps in the inhibitor bound “closed” state and the apo-state of HIV-1 PR.¹¹³ Our previous molecular dynamics (MD) simulations of flap motion in the LAI’ (PMPR + K55MTSL + D25N) sequence have also reproduced the DEER-based distance distribution profiles, and provide a necessary link that correlates the EPR distances to structural and dynamic features of the flaps.²⁶⁸ From the most probable distance and the distance distribution profiles, information about the ensemble flap conformations in solution is obtained.

In this present work, we show that mutations that arise in response to PI treatment alter the flap conformations in the apo-state, defined as the conformations sampled in the absence of substrate/inhibitor. Specifically, we investigated two drug resistant variants, V6,¹⁵⁷ isolated from a pediatric patient while on Ritonovir therapy, and MDR769^{182, 266}

that is known to confer drug resistance to seven U.S. FDA-approved protease inhibitors²⁶⁹ (nelfinavir, saquinavir, indinavir, ritonavir, amprenavir, lopinavir, Atazanavir). Each of the two constructs contains the following mutations relative to the LAI consensus sequence, as summarized in Table 5-1. The locations of these mutations are shown in Figure 5-3. As in our earlier work,^{113, 187} the two naturally occurring Cys residues (C67 and C95) were mutated to Ala and site K55C was chosen as the EPR reporter site and labeled with (1-Oxyl-2,2,5,5-Tetramethyl-3-Pyrroline-3-Methyl) methanethiosulfonate, (MTSL). In addition, in order to prevent protein degradation during the time course of data collection, the active site mutation, D25N was incorporated. These inactive constructs used for the EPR studies and MD simulations are referred to as LAI', V6' and MDR769'. Notably, the LAI' sequence also contained three mutations that provide protection from autocatalytic cleavage: Q7K, L33I, L63I.²⁵³ In an earlier work, the LAI' sequence was referred as PMPR.¹⁸⁷

Table 5-1 Mutations in each mutant construct relative to the wild-type (LAI')

Variants	mutations
LAI'	
MDR769'	L10I, M36V, S37N, M46L, I54V, I62V, L63P, A71V, V82A, I84V, L90M
V6'	K20R, V32I, L33F, M36I, L63P, A71V, V82A, L90M,

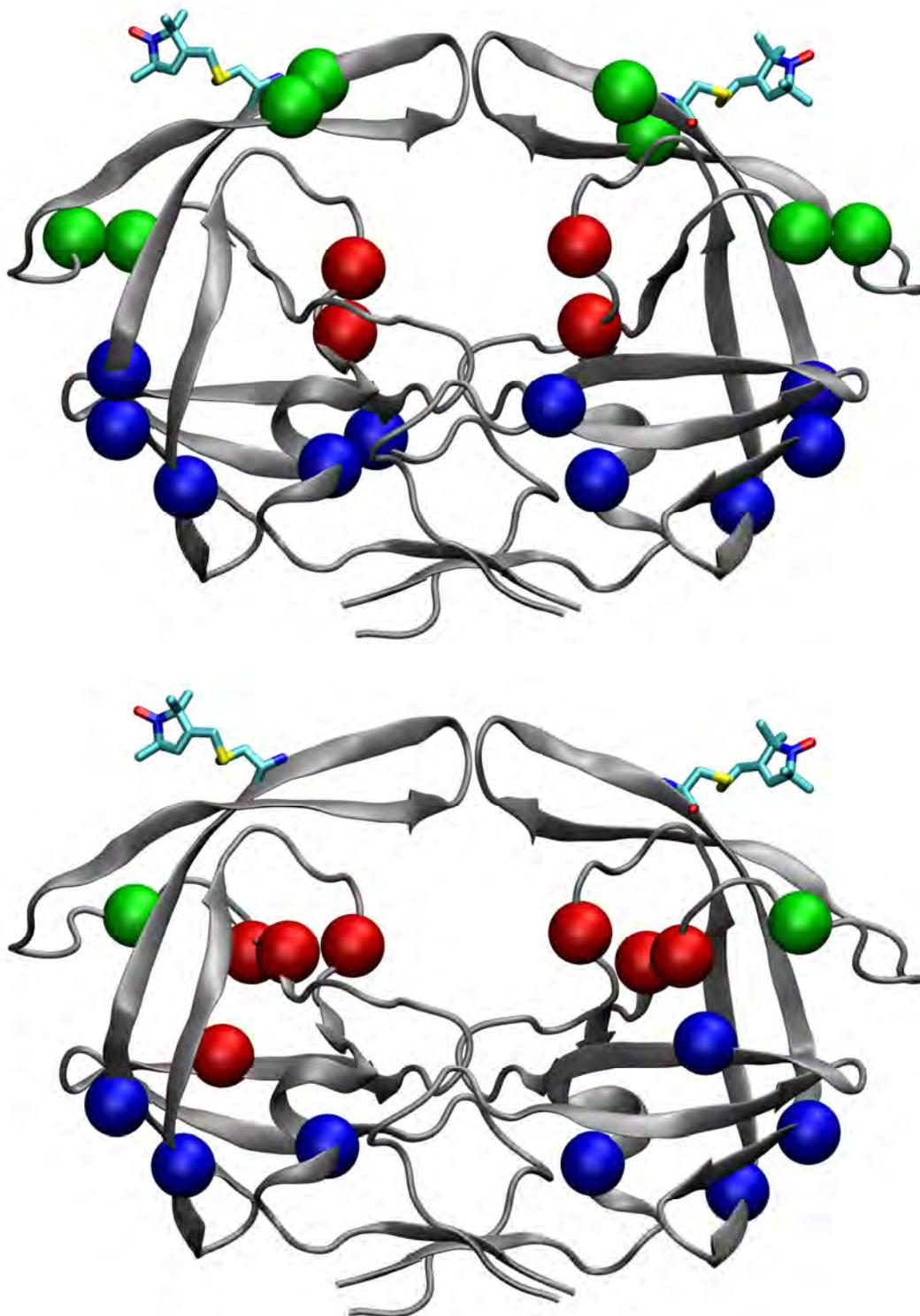


Figure 5-3. Ribbon diagrams of HIV-1PR in the semi-open conformation (1HHP) with the nitroxide spin probe, MTSL, appended at site K55C. Colored spheres represent the $C\alpha$ position of mutations relative to LAI' in MDR769' (top) and V6' (bottom) in the active site cavity, and the nonactive site region, and flaps/elbows are shown in red, blue, and green; respectively. Diagrams were rendered with VMD.

5.2 Methods

5.2.1 Construction of the Modeled Structures

The 2G69²⁶⁹ semi-open crystal structure was used as the starting point for our MD simulations. Both catalytic Asp25 residues were modeled as Asn25 to be consistent with the EPR experiments. Additionally, same mutations were introduced to the 2G69 sequence as in the EPR experiments (the sequences were the LAI', V6' and MDR769' described above). All calculations were performed using the Amber 9 program package¹⁹⁶ and the ff99SB modification⁶⁷ of the Amber ff99 force field.⁶⁶ All mutations were performed using Swiss-Pdb Viewer.¹⁹⁸ Based on Molprobit analysis,²⁷⁰ ASN88/88' amide groups were flipped in both mutants. Hydrogen atoms were added using the Leap module in the Amber9 software package.¹⁹⁶ The parameters for the Cys- MTSL spin label construct were previously published.²⁷¹

5.2.2 Minimization and Equilibration

We performed minimizations and MD simulations in a manner similar to that described in the Methodology section of Chapter 4. To avoid steric clashes caused by the introduced mutations, the structures were first subject to a stepwise minimization and equilibration in the presence of the implicit solvent model. Energy minimization was achieved in four steps. First, movement allowed only for the mutated residues; while the

protein (i.e., the coordinates of those atoms are experimentally determined) were positionally constrained (not fixed) using a harmonic potential with a force constant of $50\text{kcal}/(\text{mol}\cdot\text{\AA}^2)$. Next, all heavy atoms of the protein were restrained with forces of $10\text{kcal}/(\text{mol}\cdot\text{\AA}^2)$. Then the position restraints were only imposed on the backbone with forces of $1\text{kcal}/(\text{mol}\cdot\text{\AA}^2)$. In this step, steric collisions of the automatically generated residues were minimized, and favorable configurations of the side-chains of the mutated residues were obtained while the experimentally determined coordinates were maintained. Finally, all atoms were free to move. Next, the purpose of the subsequent MD simulation is to extensively relax the configuration of the placed mutations by adding the thermal fluctuation. For that purpose, the temperature was raised to 300 K over 50 ps in 10-ps intervals using Langevin dynamics under constant volume condition (NVT), and only the mutated residues were permitted to move freely as the first energy minimization, while the experimentally determined coordinates were positionally restrained with a force constant of $50\text{kcal}/(\text{mol}\cdot\text{\AA}^2)$. This was followed by an equilibration phase of 200ps under constant pressure (NPT) at 300K using a coupling constant of 0.5ps, with restraints on the experimentally determined coordinates, and gradually reduced from 10 to 1, 0.1 and 0 $\text{kcal}/(\text{mol}\cdot\text{\AA}^2)$ in 50-ps intervals.

Each system was then solvated using the tleap module in a truncated octahedron periodic box containing 7233 (MDR769) and 7264 (V6) TIP3P¹⁶ water molecules, with the box extending 8\AA from the extremes of the solute. Energy minimization was achieved in five steps. The solvent molecules were firstly relaxed, while all heavy atoms in protein were restrained with forces of $50\text{kcal}/(\text{mol}\cdot\text{\AA}^2)$. Then, the systems were continually relaxed with the restraint force constant gradually reduced to 10 and 0.1 $\text{kcal}/(\text{mol}\cdot\text{\AA}^2)$.

Then only the backbone was restrained with forces of $1\text{kcal}/(\text{mol}\cdot\text{\AA}^2)$. Finally, all restraints were lifted and whole system was relaxed. In each step, energy minimization was executed by the steepest descent method for the first 10,000 cycles and the conjugated gradient method for the subsequent 10,000 cycles. After the relaxation, the systems were gradually heated to 300 K during a 50 ps run under *NVT* condition using the Berendsen algorithm²⁰⁰ with a coupling constant of 0.5 ps, with all heavy atoms restrained by $5\text{ kcal}/(\text{mol}\cdot\text{\AA}^2)$. This was followed by a three-stage equilibration under *NPT* condition with a coupling constant of 1 ps and at the pressure of 1atm: i) a restrained MD simulation for 50 ps, while keeping the heavy atoms restrained with a force constant of $1\text{ kcal}/(\text{mol}\cdot\text{\AA}^2)$; ii) an additional 50 ps long MD with a restraint force constant $0.5\text{ kcal}/(\text{mol}\cdot\text{\AA}^2)$ only imposed on the backbone; iii) a short equilibration of 50 ps without any restraints was performed.

5.2.3 Production Runs

The configurations from the above equilibration stages were used as the starting configurations for the production runs. For all simulations, temperature was maintained at a temperature of 300K using the Berendsen algorithm²⁰⁰ and pressure of 1atm for subsequent production runs of ~ 30 ns. The SHAKE algorithm was used to treat the bonds involving hydrogen.¹⁹⁹ The long-range electrostatic interactions were calculated by the particle mesh Ewald (PME) method²⁵ with a dielectric constant 1.0. A cutoff of 8\AA was used to calculate the direct space sum for PME. The time step for integration was set as 2 fs, and the coordinate sets were saved at every 10 ps for subsequent analyses.

5.2.4 Data Analysis

Distances and RMSD values were calculated using the ptraj module in Amber 9. Histograms of distance distributions were constructed with intervals of 0.1 Å using the Xmgrace program. Average structures were obtained using ptraj, following best-fit of all frames to the non-flap region of the 2G69 crystal structure. Following calculation of average coordinates, conformations were energy minimized for 1000 steps to regularize the structure. The differences in the flap conformations among the mutants were not affected by the minimization.

5.3 Results and Discussion

5.3.1 Overall Structural Variation

A qualitative examination of the trajectories obtained from both the wild-type (LAI'/PMPR) and the two mutant simulations (V6' and MDR769') shows that all systems deviated to different extents from their starting structures, resulting in protein backbone RMSDs from the starting semi-open reference (PDB code 2G69) of approximately 1.2Å - 3.5Å after 20 ns. Notably, the magnitude of fluctuations varied among the three systems with the wild-type (LAI') demonstrating much larger fluctuations than the two mutant systems.

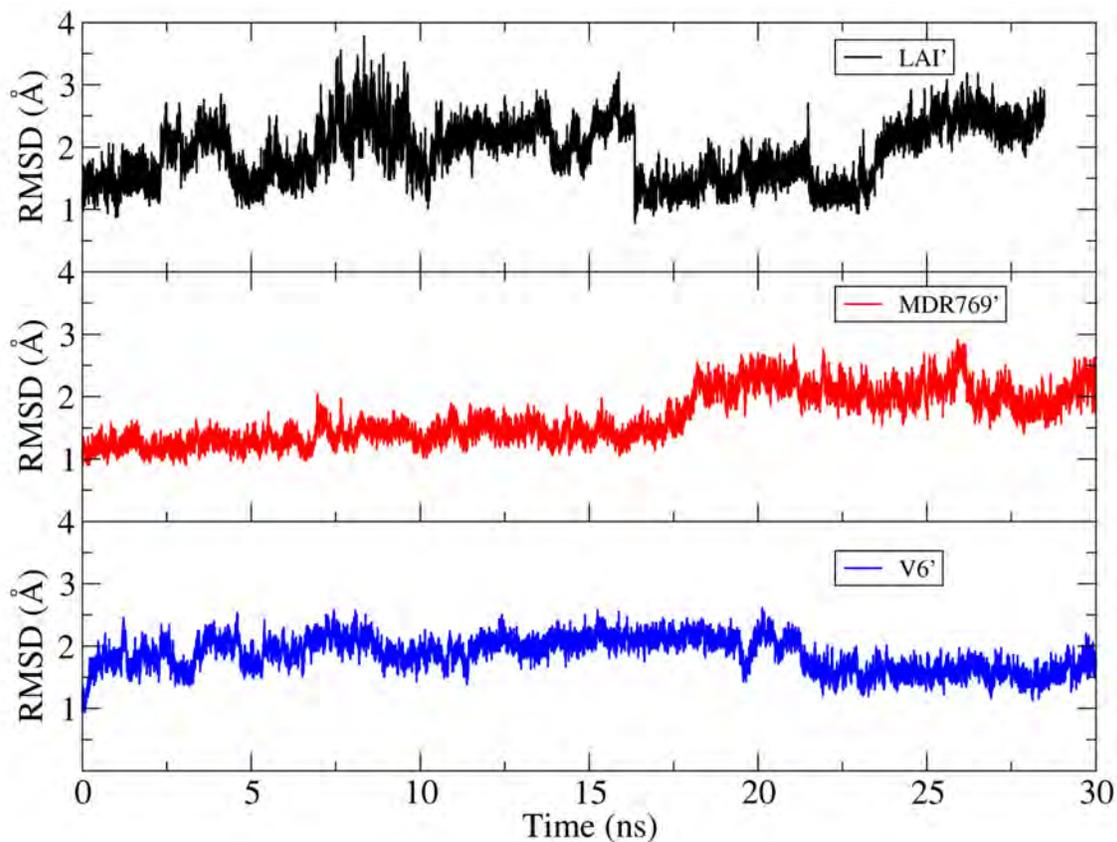


Figure 5-4. Protein backbone RMSD with respect to semi-open crystal structure (2G69) throughout the three simulations.

The structural variations were further assessed by calculating atomic fluctuations in proteins over the entire simulations. The atomic fluctuation per residues varied along the poly peptide chain in a similar manner for all simulations, as shown in Figure 5-5. The regions of low variation in all constructs were residues 5, 22-26, 31-33, 64, 75, 85-90 in both subunits. The largest variation was observed for surface turn around residues 16-18, the surface loop from 34-43, and flap residues 44-57 of both subunits. In particular, the flaps in the wild-type protease exhibited much larger fluctuations than those in the mutant systems, and no large-scale flap opening were observed for the V6' and MDR769' systems during the courses of simulations.

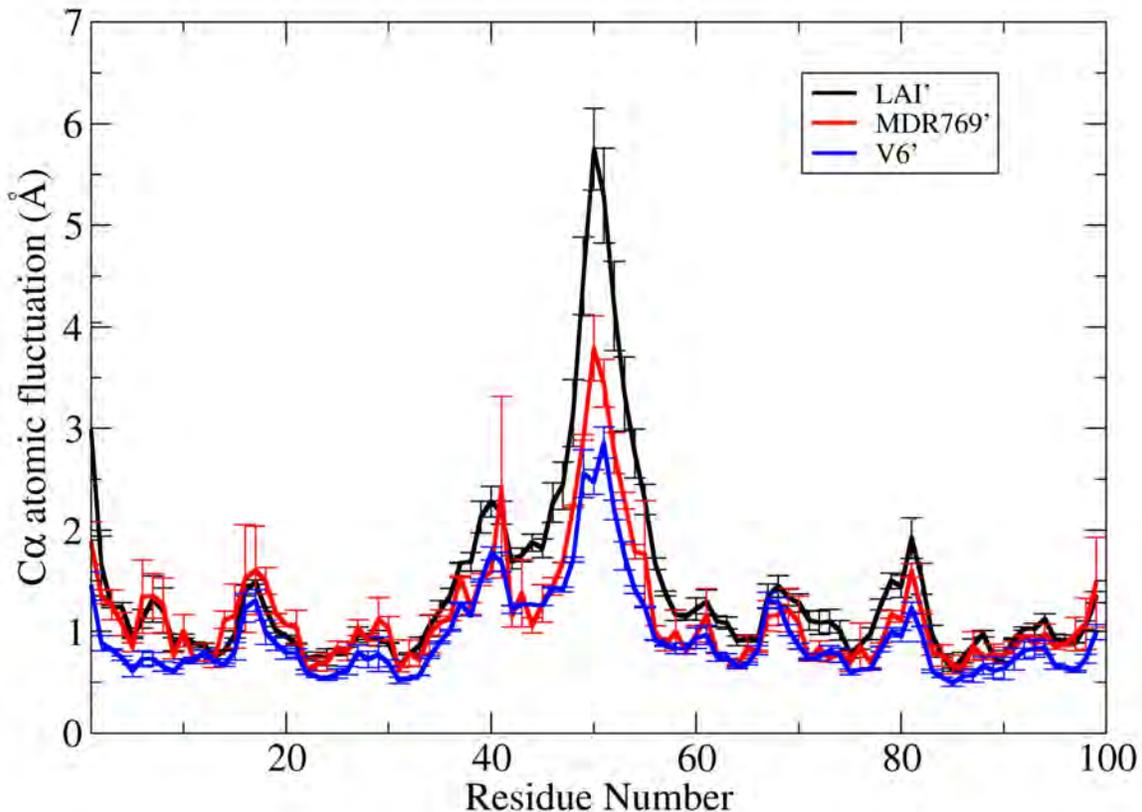


Figure 5-5. Atomic fluctuations simulated for the three proteases, LAI' (black), MDR769' (red) and V6' (blue). Error bars reflect the difference between the two monomers

5.3.2 Comparison of the Interspin Distances from EPR and MD Simulations

Because HIV-1 PR is a homodimer, generation of a single cysteine mutant for spin labeling provides a pair of spin labels for DEER measurements, where the magnitude of the magnetic dipolar coupling of the unpaired nitroxide electrons, which scales as $1/r^3$, is detected from analysis of the modulation of the spin echo amplitude.²⁷²⁻²⁷³ Shown in Figure 5-6 is the DEER echo curves for the three constructs, with their respective best

solutions obtained from Tikhonov Regularization. The corresponding distance distribution profiles for spin-labeled constructs of LAI', V6', and MRD769' were obtained by applying Tikhonov regularization, using the L curve as criterion for optimal parameter regularization.²⁷² A Gaussian fitting was then performed on these curves to enable the extraction of the calculated width at half-height. The results are presented in Figure 5-7 (top).

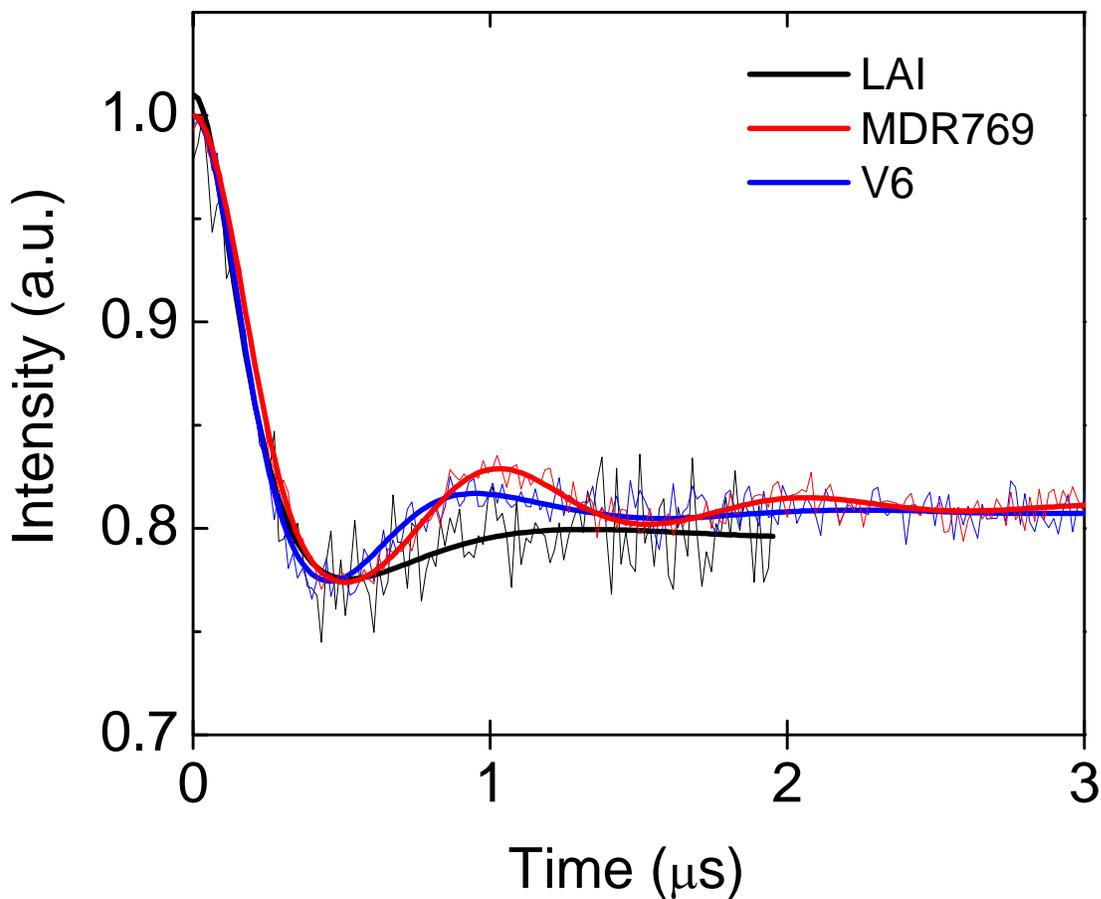


Figure 5-6. Normalized and background subtracted time domain dipolar evolution for the three constructs. The solid smooth lines represent the respective best solutions obtained from Tikhonov Regularization²⁷² for HIV-1PR samples labeled at site K55C with MTSL.

It is apparent from the distance distribution profiles (Figure 5-7, top) that the conformations of the flaps in V6' and MDR769' vary from those of LAI'. The most probable distance between spin labels in V6' is $34.9 \pm 0.7 \text{ \AA}$, which is slightly shorter than that determined for LAI' ($35.5 \pm 1 \text{ \AA}$). On the other hand, for MDR769', the most probable distance is slightly larger than in LAI' and is found to be $36.4 \pm 0.5 \text{ \AA}$. A more open structure for MDR769' than the wild-type LAI' agrees well with the reported crystal structure (PDB code: 1TW7),¹⁸² a “wide-open” structure representing an opening that is 8 \AA wider than the “semi-open” structure of apo HIV-1 PR. However, for the three constructs, a striking difference is observed in the breadth of the distance distribution profiles, which reflects the range of opening and the conformational flexibility of the flaps. In both V6' and MDR769', analysis of the DEER data shows that the flaps do not span the full range of distances seen for LAI'. For V6', the flaps span distances of $28\text{--}42 \text{ \AA}$, and for MDR769', a distance breadth of $31\text{--}42 \text{ \AA}$ is obtained. Both of these ranges are narrower than the $23\text{--}48 \text{ \AA}$ seen for LAI' (from the experimental signal/noise, errors for distance distribution profile breadths are estimated to be $\pm 2 \text{ \AA}$ for V6'/MRD769' and $\pm 4 \text{ \AA}$ for LAI').

In the MD simulations of the three sequences, overall, the trends of the distance distribution profiles are in excellent agreement with those derived from the DEER data, as summarized in Table 5-2 and shown in Figure 5-7. The MD reconstructed 55N–55'N distance distribution profiles for both V6' and MDR769' are narrower than for LAI', with both lacking long distances that correspond to a fully-open conformation of the flaps as seen in the wild-type sequence. Furthermore, the MD results predict the same shift of the most probable distance between the nitroxide spin labels seen in the EPR data for the

three constructs: V6' < LAI' < MDR769', implying different conformational flexibility of these proteases. Thus, a very good agreement between the distances measured experimentally or extracted from MD simulations verifies that our simulations could robustly reproduce the experimental data, as observed in our previous work.¹⁸⁶

Table 5-2. Summary of distance parameters obtained from DEER distance profiles and MD simulations

construct	avg. dist.		range (span)	
	EPR	MD	EPR	MD
LAI'	35.5±1.0Å	34.5±0.7Å	23 to 48Å	25 to 50Å
MDR'	36.4±0.5 Å	36.5±0.5 Å	31 to 42 Å	32 to 41 Å
V6'	34.9±0.7 Å	34.2 ±0.3Å	28 to 42 Å	28 to 43 Å

The estimation of error in the average spin label distance for each construct from EPR measurement was assessed by shifting the distance distribution to larger and smaller values that typically range from 0.5 to 5.0Å. The range of the distance distribution for each system from MD simulations was determined by histogram of the distances between the nitroxide nitrogen atoms attached to 55K on each flap.

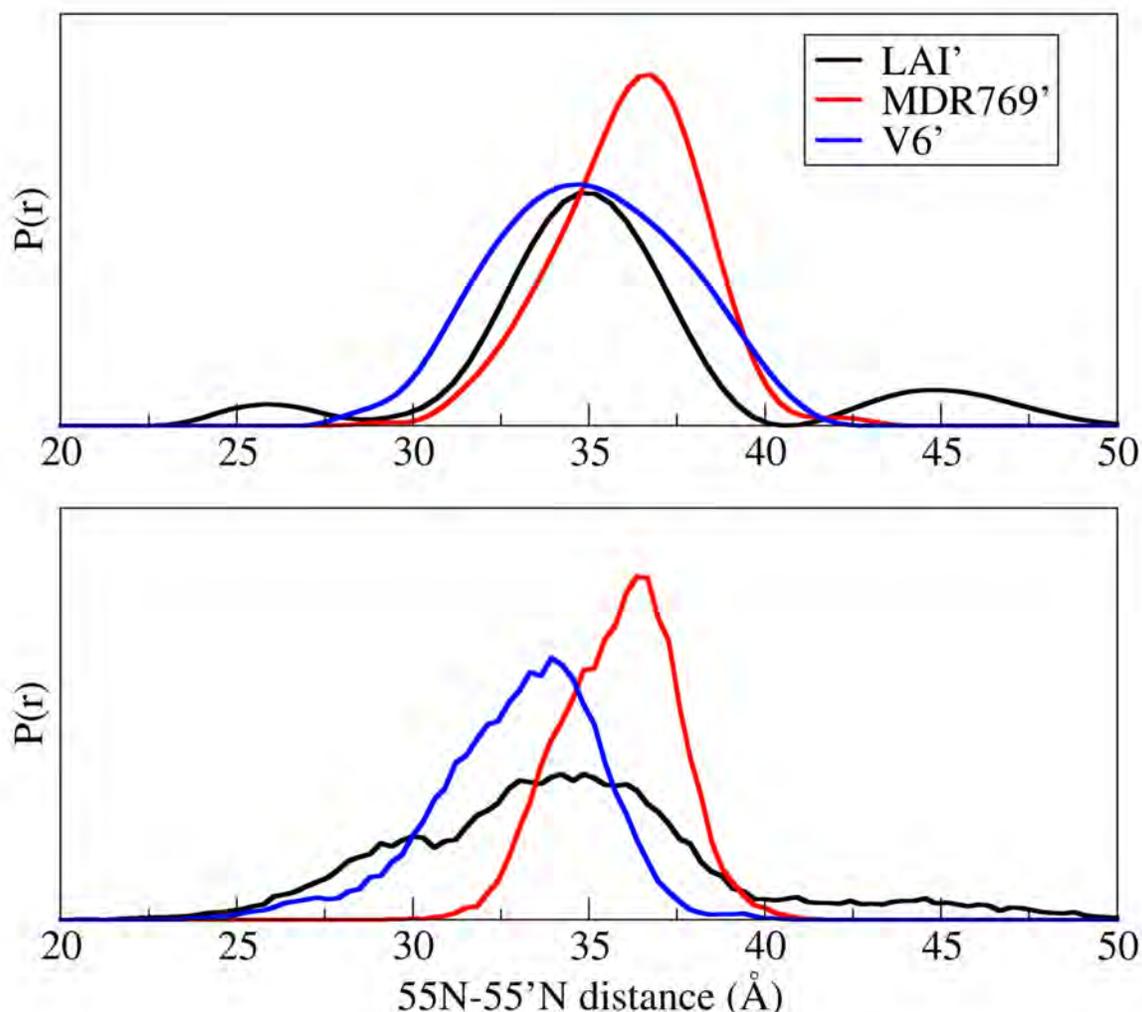


Figure 5-7. Interspin distance distribution profiles from TKR of EPR data (top) and 55N–55'N distance profiles from MD simulations (bottom).

Moreover, we note that the EPR-measured interspin distances might have a rather large uncertainty when translated into distances between α or β carbons, owing to the significant length of the spin label linking arm. In addition, spin-label conformations are likely stabilized by weak specific interactions with neighboring amino acid side chain or backbone atoms. To address the question of whether the spin label distances could represent the local motions of the protein backbone, we calculated the distances between the two α carbons on the K55C and K55'C residues, where the spin labels are attached,

termed as $55\text{C}\alpha$ - $55'\text{C}\alpha$ distance, and compared with 55N - $55'\text{N}$ distances. The results presented in Figure 5-8 reveal that the distance distributions between the $\text{C}\alpha$ pairs for the residues $55\text{C}/55'\text{C}$ have the same trend in most probable distances as those observed in the nitroxide nitrogen pairs, with the flap tips sampling shorter distances in $\text{V6}'$, whereas the flaps being on average more open in the $\text{MDR769}'$ mutant. However, there exists a discrepancy between the two measurements. The distance distributions between the $\text{C}\alpha$ pairs for the residues $55\text{C}/55'\text{C}$ has two peaks in both the LAI' and $\text{V6}'$ sequences, representing two conformations of the flaps, whereas there is only one peak in their corresponding distance distributions for the nitroxide nitrogen pairs. This apparent discrepancy can be explained by the flexibility of the spin labels which accommodates the differences in the backbone conformations. Thus, these measurements provide further evidence that the changes seen in the EPR data analysis for LAI' , $\text{V6}'$ and $\text{MDR769}'$ are dominated by changes in the protein backbone positions with minor contributions from the internal motions of the side chains of the spin labels.

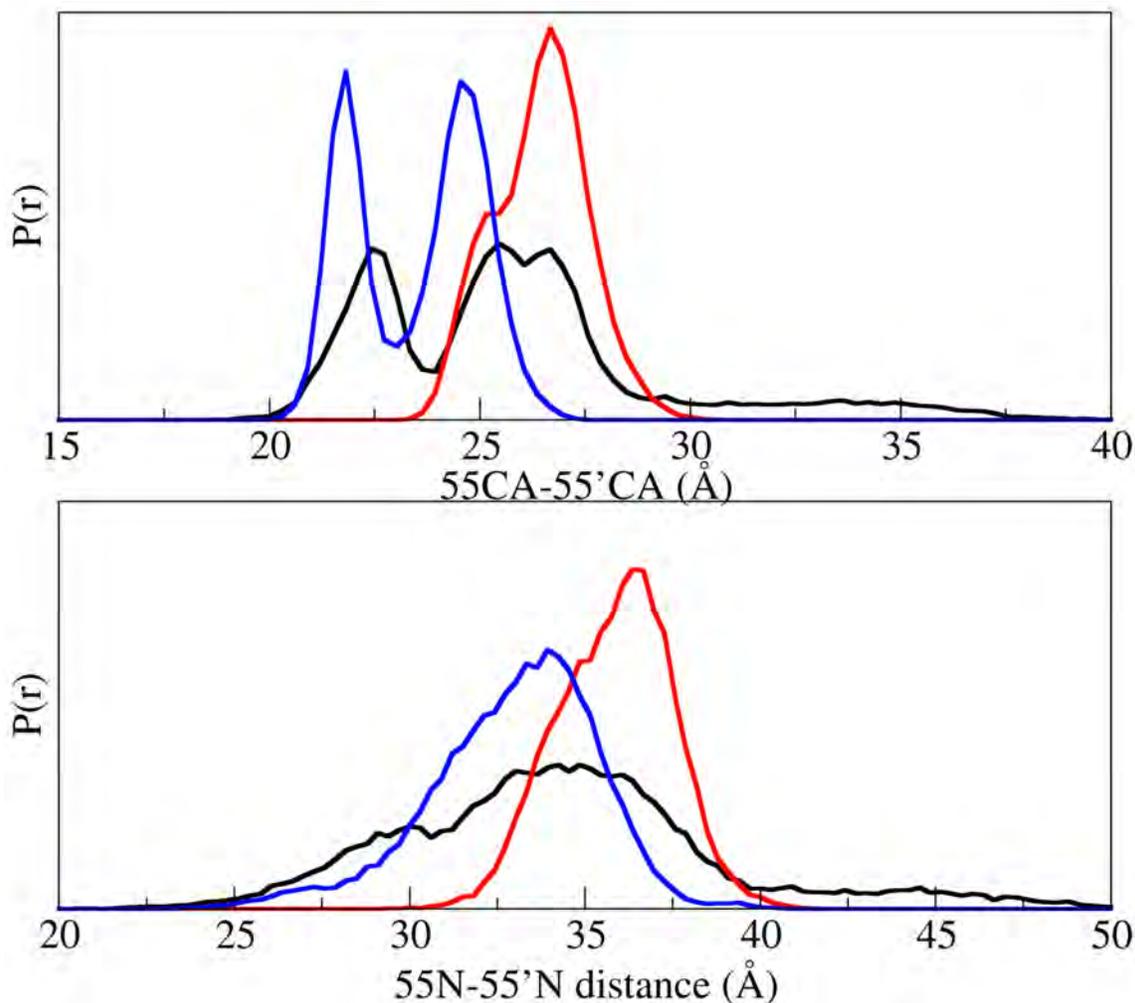


Figure 5-8. Distributions of distances between the carbon atoms (top) and the nitroxide nitrogen atoms (bottom) in the residues K55C and K55'C sampled during the MD simulations of LAI (black), MDR769 (red) and V6 (blue).

5.3.3 Comparison of Average Conformations

To obtain insight into structural conformations that correspond to the experimentally determined distance profiles, we calculated the average conformations sampled during the MD simulations (Figure 5-9). The flaps in the LAI' simulation adopted a degree of closure in excellent agreement with the semi-open crystal structure of apo HIV-PR (PDB

code: 2G69). In contrast, but also consistent with the EPR data analysis, the flaps of the MDR769' mutant assumed a more open conformation than those of LAI', while the flaps in V6' were more closed relative to LAI'. Note that all structures adopt the semi-open flap handedness⁴⁵.

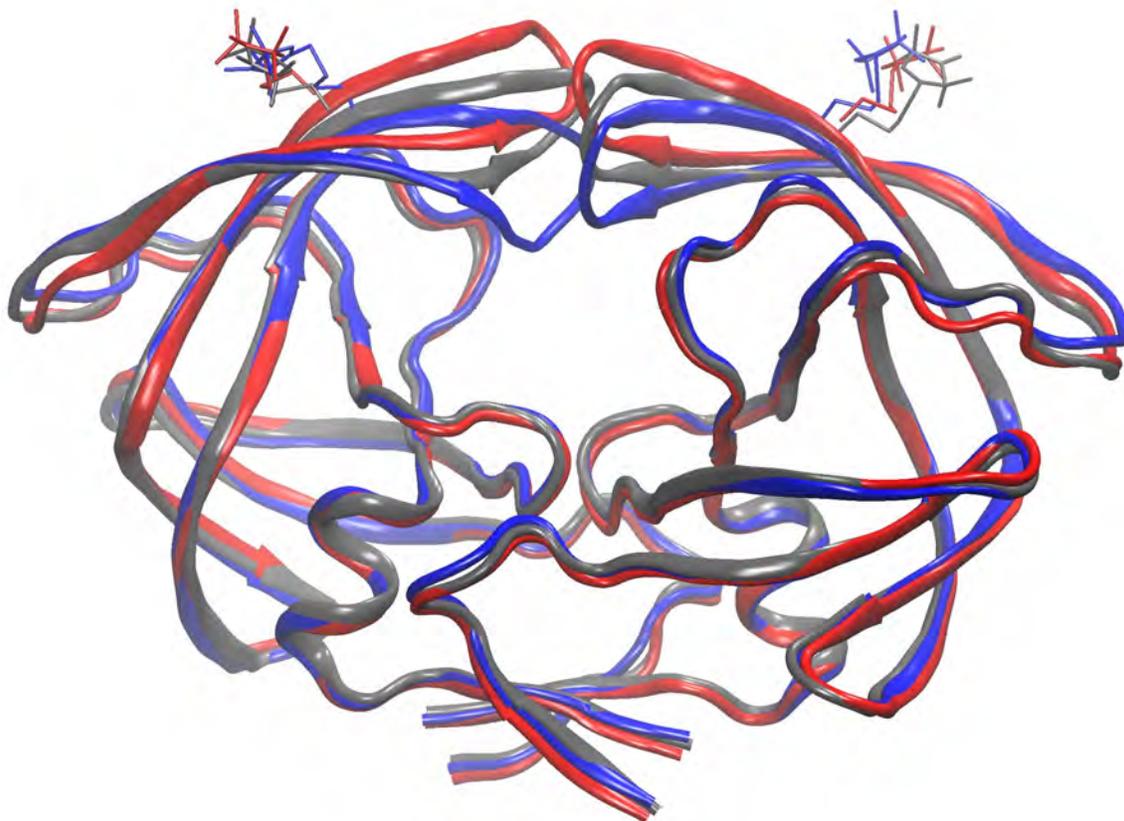


Figure 5-9. Comparison of averaged structures sampled during MD simulations of the three sequences. For clarity, only the backbone ribbon and Cys-MTSL side chain are shown. KEY: LAI "wild-type" (black); V6' (blue); MDR769' (red).

To quantitatively determine the extent of flap opening with respect to the active site in different sequences, we measured the distances sampled between the center of mass (COM) of 5 central residues on each flap (residues 48-52) and the COM of the two Asn25/25' residues during the simulations. The results confirm the conclusions obtained

from the average structures. As demonstrated in Figure 5-10, the flap tips are closer to the active site in V6' sequence, whereas the flaps are on average more open in the MDR769 mutant.

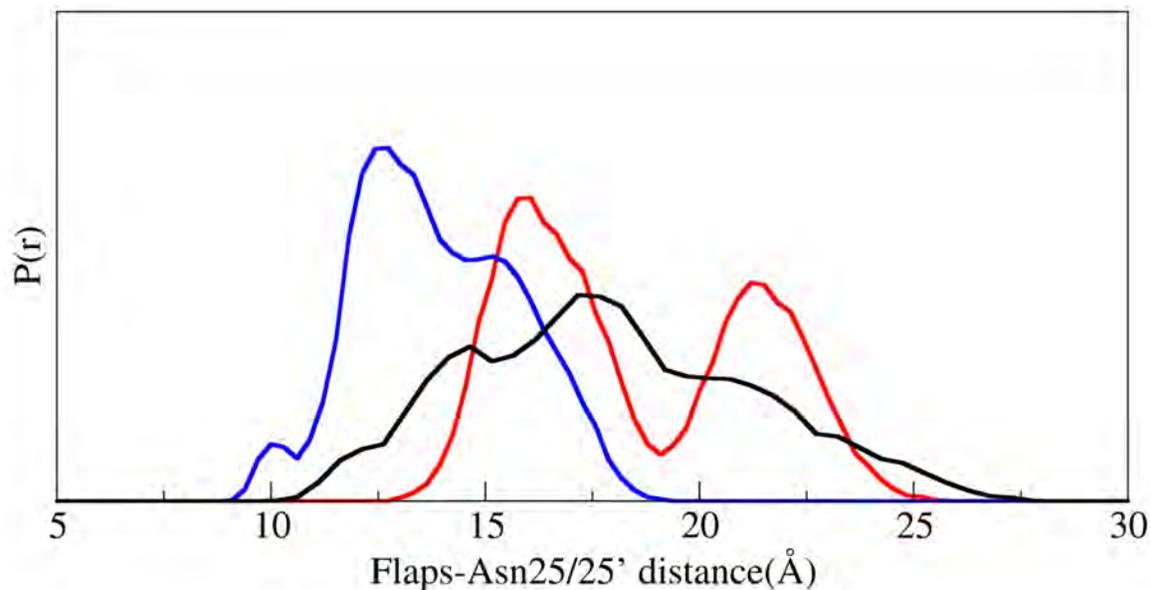


Figure 5-10. Histogram of distances sampled between the COM of 5 central residues on each flap (residues 48-52) and the COM of the two Asn25/25' residues sampled during the MD simulations of LAI (black), MDR769 (red) and V6 (blue).

To sum up, results from MD and DEER show unambiguously that mutations linked to function and inhibitor resistance can alter flap conformations in HIV-1PR. Upon the basis of the combined analysis, we see that both the breadth of the flap distance distribution profile and the average conformation are altered in the mutants, providing valuable insight into the coupling of drug resistance and protein backbone conformational flexibility. We hypothesize that the limited conformational opening of the flaps in V6' might alter the ability of the inhibitor, and possibly substrate, to enter into the active site cavity, whereas in MDR769', the longer average semi-open distance might increase the

free energy cost for the flaps closing tightly in the presence of inhibitor or substrate. This would be consistent with an inability to form stable interactions with the flaps.

Notably, our hypothesis is supported by previous experimental observations^{258, 266} that the effects of nonactive-site mutations, such as M46I, I54V, M36I and A71V, are not affecting the final bound conformation but might be affecting the mechanics of the flap opening and closing or the stability of the opened or closed conformation.

5.4 Conclusions

Both EPR experimental data and MD simulations converge to the same distance changes, demonstrating that the combination of these two methods is adapted to the study of the intrinsic dynamics of HIV-1 PR associated with its catalytic function. The DEER results show that mutations linked to function and inhibitor resistance can alter flap conformations in HIV-1PR. On the other hand, the MD simulations of the flap motion provide a structural interpretation of the EPR data. Upon the basis of the combined analysis, we see that both the breadth of the flap distance distribution profile and the average conformation are altered in the mutants, providing valuable insight into the coupling of drug resistance and protein backbone conformational flexibility. We postulate that changes in the dynamics of the flaps opening and closing would have a greater effect on the stable binding of rigid inhibitors than the transient binding of flexible substrates; perhaps the limited conformational opening of the flaps in V6' alters the ability of the inhibitor, and possibly substrate, to enter into the active site cavity,

whereas in MDR769', the longer average semi-open distance might increase the free energy cost for the flaps closing tightly in the presence of inhibitor or substrate.

Chapter 6

Concluding Remarks

In this work, we employed MD simulations in combination with experimental techniques to study the structural and dynamic features of HIV-1 protease. Our simulations provide solid evidence for the existence of pre-existing equilibrium between different conformations of this enzyme in the absence of the ligand, ranging from closed, curled, semi-open, and fully-open forms, which is in excellent agreement with the observations from X-ray crystallography, solution NMR and EPR spectroscopy. Based on the results from our μs MD simulations on apo HIV-PR with explicit solvent, we suggested that the rearrangements of the flaps between the closed and semi-open conformations is likely induced by the twisting of the backbone of the flap tips, mainly owing to the rotation of the Ψ angle of Gly49 and Φ angle of Ile50. Such twisting in turn disrupts the inter-monomer interactions between the two tips, especially the van der Waals contacts between the flap tip Ile50 residue and the hydrophobic cluster within the other monomer, thus causing the rearrangements of the flaps to take place.

Moreover, on the basis of energetic analyses, we speculate that the highly conserved dimer interface is a critical element not only structurally, but also functionally; the full flap opening event is likely an intermediate state along the path of dissociation/association of the HIV-1 PR dimer, owing to the fact that the dimer is at equilibrium with the monomer. In addition, residue-based energy decomposition analysis revealed that the intermonomer interaction energies are not evenly distributed along the

entire dimer interface, with the N- and C-termini making very favorable intermonomer interactions, as well as certain residues from the flaps (residues 45-54), the active site (residues 24-27), and the region encompassing the α -helix (residues 86-94) and the β -loop (residues 4'-9'). Importantly, residues Leu9 and Arg87 from the helix, as well as Ile5 and Trp6 from the solvent-exposed β -loop, exhibited significant decreases in their respective interaction energies when the flaps opened, implying that the disruption of the intermonomer interactions formed among these residues might be associated with the observed partial dissociation of the dimer, i.e., the opening of the flaps. The significance of the inter-subunit interactions within this dimer interface to the gating dynamics of the flaps was confirmed by double mutant simulations, which exhibit reduced binding affinity at the α -helix and β -loop dimer interface. Thus, we hypothesize that targeting the highly conserved region of the dimer interface formed between the helix and the β -loop in the fully open form may trap the enzyme in an inactive conformation, and thus may effectively interfere with the equilibrium between the different conformations of the protease associated with its function, thus creating new opportunities for inhibitor design. That this novel class of inhibitors is distinct from the dimerization inhibitors, which are designed to mimic the dimerization interface of the monomeric form of HIV-1 PR, thus blocking the assembly of the homodimer.

In addition, energy decomposition revealed that there is an anticorrelation between the interaction energies within the flap elbow, fulcrum, and cantilever regions with the total binding affinity; these three outer loops make more favorable intermonomer interactions when the flaps open. Thus, this result provides further solid support for the potential of these regions as allosteric sites to inhibit HIV-1 PR in terms of energy, not

just based on the structural anticorrelation between these regions, as reported before by different groups. Thus, it is most likely that if a molecule binds to the 'exo' site surrounded by the flap elbow, the fulcrum and cantilever regions will exert negative allosteric control on the motion of the flaps, thus inhibit its function.

It is worth noting that a high temperature (375K) was used in our MD simulations, enabling efficient sampling over the long period of one microsecond. Although the temperature dependence of the protein dynamics appears to be small, as the atomic fluctuations follow similar trends in both the high and the low temperature simulations, the sampling of conformational substates in the high temperature MD simulation is, in fact, different from that at a low temperature. As a result, the use of high temperature precludes us from exploring the energy difference among various conformations at low biological temperature, which may be associated with drug resistance.¹³¹ Nonetheless, a reaction coordinate for umbrella sampling was suggested from this high temperature MD simulation; an angle between the two vectors (Figure 3-7). Hence, performing an umbrella sampling simulation along this reaction coordinate will most likely generate the potential of mean force (PMF) of the protease at the low temperature of interest.

The second project aimed at gaining a better understanding of how resistance caused by protease mutations arises. In the simulations on the spin-labeled proteases, the quantitative agreement of the interspin distance profiles with EPR measurement were achieved, and the breadth of the flap distance distribution profile and the average conformations were observed to be altered in the mutants. Thus, these studies validate the robustness of current computational techniques that have reached the stage where they can reproduce experimental observations, and thus provide valuable insight into the

coupling of drug resistance and protein backbone conformational flexibility. We suggest that mutations that arise in response to PI treatment alter the flap conformations in the absence of substrate/inhibitor; the limited conformational opening of the flaps in V6' may alter the ability of the inhibitor, and possibly substrate, to enter into the active site cavity; whereas in MDR769', the longer average semi-open distance might increase the free energy cost for the flaps closing tightly in the presence of inhibitor or substrate.

While the coupling between mutations and conformational changes is recognized, quantitative assessment of the correlation between mutations and energetics remains to be clarified, which requires the potential of mean force (PMF) profiles for different proteases. Mutations might alter the energy difference between different conformations of HIV PR. To address this issue, we can carry out umbrella sampling simulations using the open angle as the reaction coordinate, as suggested in our high temperature MD simulations. A detailed knowledge of the energy differences between different conformers of variants is of importance for a better understanding of how drug resistance arises from mutations in the viral genome, which will help physicians design the best therapeutic regimen for a HIV-infected patient.

Finally, as discussed in Chapter 1, the performance of MD simulations is also limited by inaccuracies in the potential energy function. Specifically, a force field may bias the simulation toward incorrect conformations.⁷³ This may explain the unexpected stability of the protease under high pressure and high temperature. In particular, four residue types, isoleucine, leucine, aspartate, and asparagine, have been reported to exhibit particularly large deviations from the PDB distribution⁷³, suggesting that the ff99SB force field does not model these side chains accurately. The χ_1 torsion potentials for these

four residues also demonstrate considerable differences from MP2 calculations (this work is done by James Maier, also in Simmerling's lab). Notably, systematic revision of the χ_1 torsion potentials for amino acid side chains is in progress. To further verify the mechanisms revealed by current work, future study should be carried out employing the updated force field with modified torsion potentials.

References

1. DeLano W. L. (2002) <http://www.pymol.org>.
2. Humphrey W., Dalke A., & Schulten K. (1996) *Journal of Molecular Graphics* 14(1):33-&.
3. Simmerling C., Strockbine B., & Roitberg A. E. (2002) *Journal of the American Chemical Society* 124(38):11258-11259.
4. Ivankov D. N. & Finkelstein A. V. (2010) *The Journal of Physical Chemistry B* 114(23):7920-7929.
5. Daggett V. (2006) *Chemical Reviews* 106(5):1898-1916.
6. Reyes C. M. & Kollman P. A. (2000) *Journal of Molecular Biology* 297(5):1145-1158.
7. Weiner P. K. & Kollman P. A. (1981) *Journal of Computational Chemistry* 2(3):287-303.
8. Brooks B. R., Bruccoleri R. E., Olafson B. D., States D. J., Swaminathan S., & Martin Karplus (1983) *Journal of Computational Chemistry* 4(2):187-217.
9. Van Gunsteren W. F. (1987) *University of Groningen*.
10. Scott W. R. P., Hunenberger P. H., Tironi I. G., Mark A. E., Billeter S. R., Fennen J., Torda A. E., Huber T., Kruger P., & van Gunsteren W. F. (1999) *The Journal of Physical Chemistry A* 103(19):3596-3607.
11. Jorgensen W. L. & Tiradorives J. (1988) *Journal of the American Chemical Society* 110(6):1657-1666.
12. Case D. A., Darden, T. A. , Cheatham, III , T.E. , Simmerling, C.L., Wang, J. , Duke, R.E., Luo, R. , Crowley, M., Walker, Ross C. , Zhang, W., Merz, K.M.,

- Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K.F., Paesani, F., Vanicek, J., Wu, X., Brozell, S.R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D.H., Seetin, M.G., Sagui, C., Babin, V. and Kollman P.A. (2008) AMBER 10. in *University of California, San Francisco*.
13. Jorgensen W. L. & Tirado-Rives J. (2005) *Proceedings of the National Academy of Sciences of the United States of America* 102(19):6665-6670.
 14. Tan Y. H., Tan C. H., Wang J., & Luo R. (2008) *Journal of Physical Chemistry B* 112(25):7675-7688.
 15. Leach A. R. (*Molecular Modelling Principles and Applications*).
 16. Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W., & Klein M. L. (1983) *Journal of Chemical Physics* 79(2):926-935.
 17. Jorgensen W. L. (1981) *Journal of the American Chemical Society* 103(16):4721-4726.
 18. Berendsen H. J. C. P., J. P. M.; van Gunsteren, W. F.; Hermans, J. (1969) *Nature* 224:175-177.
 19. Berweger C. D., van Gunsteren W. F., & Müller-Plathe F. (1995) *Chemical Physics Letters* 232(5-6):429-436.
 20. Horn H. W., Swope W. C., Pitner J. W., Madura J. D., Dick T. J., Hura G. L., & Head-Gordon T. (2004) *Journal of Chemical Physics* 120(20):9665-9678.
 21. Jorgensen W. L. (1982) *Journal of Chemical Physics* 77(8):4156-4163.
 22. Mahoney M. W. & Jorgensen W. L. (2000) *Journal of Chemical Physics* 112(20):8910-8922.

23. Gilson M. K., Davis M. E., Luty B. A., & McCammon J. A. (1993) *The Journal of Physical Chemistry* 97(14):3591-3600.
24. Sitkoff D., Sharp K. A., & Honig B. (1994) *Journal of Physical Chemistry* 98(7):1978-1988.
25. Darden T., York D., & Pedersen L. (1993) *Journal of Chemical Physics* 98(12):10089-10092.
26. Hockney R. W., and Eastwood, J. W. (1988) *A Hilger, Bristol [England] ; Philadelphia.*
27. Lee M. S. & Olson M. (2004) *Abstracts of Papers of the American Chemical Society* 228:U507-U507.
28. Shestakov A. I., Milovich J. L., & Noy A. (2002) *Journal of Colloid and Interface Science* 247(1):62-79.
29. Harvey S. C. (1989) *Proteins-Structure Function and Genetics* 5(1):78-92.
30. Baker N. A., Sept D., Joseph S., Holst M. J., & McCammon J. A. (2001) *Proceedings of the National Academy of Sciences of the United States of America* 98(18):10037-10041.
31. Lu B. Z., Zhang D. Q., & McCammon J. A. (2005) *Journal of Chemical Physics* 122(21):7.
32. Still W. C., Tempczyk A., Hawley R. C., & Hendrickson T. (1990) *Journal of the American Chemical Society* 112(16):6127-6129.
33. Hawkins G. D., Cramer C. J., & Truhlar D. G. (1995) *Chemical Physics Letters* 246(1-2):122-129.

34. Tsui V. & Case D. A. (2000) *Journal of the American Chemical Society* 122(11):2489-2498.
35. Tsui V. & Case D. A. (2000) *Biopolymers* 56(4):275-291.
36. Onufriev A., Bashford D., & Case D. A. (2004) *Proteins: Structure, Function, and Bioinformatics* 55(2):383-394.
37. Mongan J., Simmerling C., McCammon J. A., Case D. A., & Onufriev A. (2006) *Journal of Chemical Theory and Computation* 3(1):156-169.
38. Zagrovic B. & Pande V. (2003) *Journal of Computational Chemistry* 24(12):1432-1436.
39. Nymeyer H. & Garcia A. E. (2003) *Proceedings of the National Academy of Sciences of the United States of America* 100(24):13934-13939.
40. Okur A., Wickstrom L., & Simmerling C. (2008) *Journal of Chemical Theory and Computation* 4(3):488-498.
41. Roe D. R., Okur A., Wickstrom L., Hornak V., & Simmerling C. (2007) *Journal of Physical Chemistry B* 111(7):1846-1857.
42. Zhou R. (2003) *Proteins* 53(2):148-161.
43. Zhou R. (2003) *Proceedings of the National Academy of Sciences of the United States of America* 100(23):13280-13285.
44. Zhou R. & Berne B. J. (2002) *Proceedings of the National Academy of Sciences of the United States of America* 99(20):12777-12782.
45. Hornak V., Okur A., Rizzo R. C., & Simmerling C. (2006) *Proceedings of the National Academy of Sciences of the United States of America* 103(4):915-920.

46. Freisner R. A. & Gunn J. R. (1996) *Annual Review of Biophysics and Biomolecular Structure* 25(1):315-342.
47. Torrie G. M. & Valleau J. P. (1977) *Journal of Computational Physics* 23(2):187-199.
48. Bowman A. L., Ridder L., Rietjens I. M. C. M., Vervoort J., & Mulholland A. J. (2007) *Biochemistry* 46(21):6353-6363.
49. Piccinini E., Ceccarelli M., Affinito F., Brunetti R., & Jacoboni C. (2007) *Journal of Chemical Theory and Computation* 4(1):173-183.
50. Kumar S., Rosenberg J. M., Bouzida D., Swendsen R. H., & Kollman P. A. (1992) *Journal of Computational Chemistry* 13(8):1011-1021.
51. Kumar S., Rosenberg J. M., Bouzida D., Swendsen R. H., & Kollman P. A. (1995) *Journal of Computational Chemistry* 16(11):1339-1350.
52. Hansmann U. H. E. (1997) *Chemical Physics Letters* 281(1-3):140-150.
53. Hukushima K. & Nemoto K. (1996) *Journal of the Physical Society of Japan* 65(6):1604-1608.
54. Shirakura T. & Matsubara F. (1996) *Journal of the Physical Society of Japan* 65(10):3138-3141.
55. Sugita Y. & Okamoto Y. (1999) *Chemical Physics Letters* 314(1-2):141-151.
56. García A. E. & Sanbonmatsu K. Y. (2001) *PROTEINS: Structure, Function, and Genetics* 42(3):345-354.
57. Sanbonmatsu K. Y. & García A. E. (2002) *PROTEINS: Structure, Function, and Genetics* 46(2):225-234.

58. Pitera J. W. & Swope W. (2003) *Proceedings of the National Academy of Sciences of the United States of America* 100(13):7587-7592.
59. Kuhn B. & Kollman P. A. (2000) *Journal of Medicinal Chemistry* 43(20):3786-3791.
60. Im W., Beglov D., & Roux B. (1998) *Computer Physics Communications* 111(1-3):59-75.
61. Srinivasan J., Cheatham T. E., Cieplak P., Kollman P. A., & Case D. A. (1998) *Journal of the American Chemical Society* 120(37):9401-9409.
62. Qiu D., Shenkin P. S., Hollinger F. P., & Still W. C. (1997) *The Journal of Physical Chemistry A* 101(16):3005-3014.
63. Trieb M., Rauch C., Wibowo F. R., Wellenzohn B., & Liedl K. R. (2004) *Nucleic Acids Research* 32(15):4696-4703.
64. Bonnet P. & Bryce R. A. (2005) *Journal of Molecular Graphics and Modelling* 24(2):147-156.
65. Pencheva T., Soumana O. S., Pajeva I., & Miteva M. A. (2010) *European Journal of Medicinal Chemistry* 45(6):2622-2628.
66. Cornell W. D., Cieplak P., Bayly C. I., Gould I. R., Merz K. M., Ferguson D. M., Spellmeyer D. C., Fox T., Caldwell J. W., & Kollman P. A. (1996) *Journal of the American Chemical Society* 118(9):2309-2309.
67. Hornak V., Abel R., Okur A., Strockbine B., Roitberg A., & Simmerling C. (2006) *Proteins: Structure, Function, and Bioinformatics* 65(3):712-725.
68. Jr. A. D. M., Feig M., & III C. L. B. (2004) *Journal of Computational Chemistry* 25(11):1400-1415.

69. Friesner R. A. (2005) *Advances in Protein Chemistry* 72:79-104.
70. Anisimov V. M., Vorobyov I. V., Roux B., & MacKerell A. D. (2007) *Journal of Chemical Theory and Computation* 3(6):1927-1946.
71. Schnieders M. J., Baker N. A., Ren P. Y., & Ponder J. W. (2007) *Journal of Chemical Physics* 126(12):124114.
72. Mccammon J. A., Gelin B. R., & Karplus M. (1977) *Nature* 267(5612):585-590.
73. Klepeis J. L., Lindorff-Larsen K., Dror R. O., & Shaw D. E. (2009) *Current Opinion in Structural Biology* 19(2):120-127.
74. Freddolino P. L., Liu F., Gruebele M., & Schulten K. (2008) *Biophysical Journal* 94(10):L75-L77.
75. Ensign D. L., Kasson P. M., & Pande V. S. (2007) *Journal of Molecular Biology* 374(3):806-816.
76. Pérez A., Luque F. J., & Orozco M. (2007) *Journal of the American Chemical Society* 129(47):14739-14745.
77. Seibert M. M., Patriksson A., Hess B., & van der Spoel D. (2005) *Journal of Molecular Biology* 354(1):173-183.
78. Maragakis P., Lindorff-Larsen K., Eastwood M. P., Dror R. O., Klepeis J. L., Arkin I. T., Jensen M. O., Xu H., Trbovic N., Friesner R. A., Palmer A. G., & Shaw D. E. (2008) *The Journal of Physical Chemistry B* 112(19):6155-6158.
79. Grossfield A., Pitman M. C., Feller S. E., Soubias O., & Gawrisch K. (2008) *Journal of Molecular Biology* 381(2):478-486.
80. Dror R. O., Arlow D. H., Borhani D. W., Jensen M. Ø., Piana S., & Shaw D. E. (2009) *Proceedings of the National Academy of Sciences* 106(12):4689-4694.

81. Soni A. & Gupta R. (2009) *Health Affairs* 28(6):1617-1628.
82. Barre-Sinoussi F., Chermann J. C., Rey F., Nugeyre M. T., Chamaret S., Gruest J., Dauguet C., Axler-Blin C., Vezinet-Brun F., Rouzioux C., Rozenbaum W., & Montagnier L. (1983) *Science* 220(4599):868-871.
83. Barresinoussi F., Chermann J. C., Rey F., Nugeyre M. T., Chamaret S., Gruest J., Dauguet C., Axlerblin C., Vezinetbrun F., Rouzioux C., Rozenbaum W., & Montagnier L. (1983) *Science* 220(4599):868-871.
84. Montagnier L., Chermann J. C., & Barresinoussi F. (1985) *New Scientist* 105(1447):44-44.
85. Vittecoq D., Ferchal F., Chamaret S., Benbunan M., Gerber M., Hirsch A., & Montagnier L. (1987) *Lancet* 1(8542):1150-1151.
86. Wlodawer A. & Vondrasek J. (2003) *Annual Review of Biophysics and Biomolecular Structure* 27(1):249-284.
87. Yoshinaka Y. & Luftig R. B. (1977) *Cell* 12(3):709-720.
88. Kohl N. E., Emini E. A., Schleif W. A., Davis L. J., Heimbach J. C., Dixon R. A. F., Scolnick E. M., & Sigal I. S. (1988) *Proceedings of the National Academy of Sciences of the United States of America* 85(13):4686-4690.
89. Riviere Y., Blank V., Kourilsky P., & Israel A. (1991) *Nature* 350(6319):625-626.
90. Wlodawer A., Miller M., Jaskolski M., Sathyanarayana B. K., Baldwin E., Weber I. T., Selk L. M., Clawson L., Schneider J., & Kent S. B. H. (1989) *Science* 245(4918):616-621.

91. Miller M., Schneider J., Sathyanarayana B. K., Toth M. V., Marshall G. R., Clawson L., Selk L., Kent S. B. H., & Wlodawer A. (1989) *Science* 246:1149.
92. 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. (1989) *Nature* 337(6208):615-620.
93. Sussman J. L., Lin D., Jiang J., Manning N. O., Prilusky J., Ritter O., & Abola E. E. (1998) *Acta crystallographica Section D, Biological crystallography* 54(Pt 6 Pt 1):1078-1084.
94. Harte W. E., Swaminathan S., Mansuri M. M., Martin J. C., Rosenberg I. E., & Beveridge D. L. (1990) *Proceedings of the National Academy of Sciences of the United States of America* 87(22):8864-8868.
95. Pearl L. & Blundell T. (1984) *Federal of European Biochemical Societies Letters* 174(1):96-101.
96. Pearl L. H. & Taylor W. R. (1987) *Nature* 329:351-354.
97. Ishima R., Torchia D. A., Lynch S. M., Gronenborn A. M., & Louis J. M. (2003) *Journal of Biological Chemistry* 278(44):43311-43319.
98. Vondrasek J. & Wlodawer A. (2002) *Proteins-Structure Function and Genetics* 49(4):429-431.
99. 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. (1989) *Nature* 337(6208):615-620.

100. Lapatto R., Blundell T., Hemmings A., Overington J., Wilderspin A., Wood S., Merson J. R., Whittle P. J., Danley D. E., Geoghegan K. F., Hawrylik S. J., Lee S. E., Scheld K. G., & Hobart P. M. (1989) *Nature* 342(6247):299-302.
101. Robbins A. H., Coman R. M., Bracho-Sanchez E., Fernandez M. A., Gilliland C. T., Li M., Agbandje-McKenna M., Wlodawer A., Dunn B. M., & McKenna R. (2010) *Acta Crystallographica Section D* 66(3):233-242.
102. Layten M., Hornak V., & Simmerling C. (2006) *Journal of the American Chemical Society* 128(41):13360-13361.
103. Lexa K. W., Damm K. L., Quintero J. J., Gestwicki J. E., & Carlson H. A. (2009) *Proteins: Structure, Function, and Bioinformatics* 74(4):872-880.
104. Kumar M., Kannan K. K., Hosur M. V., Bhavesh N. S., Chatterjee A., Mittal R., & Hosur R. V. (2002) *Biochemical and Biophysical Research Communications* 294(2):395-401.
105. Lange-Savage G., Berchtold H., Liesum A., Budt K. H., Peyman A., Knolle J., Sedlacek J., Fabry M., & Hilgenfeld R. (1997) *European Journal of Biochemistry* 248(2):313-322.
106. Rick S. W., Erickson J. W., & Burt S. K. (1998) *Proteins* 32(1):7-16.
107. Ma B. Y., Shatsky M., Wolfson H. J., & Nussinov R. (2002) *Protein Science* 11(2):184-197.
108. Bahar I., Chennubhotla C., & Tobi D. (2007) *Current Opinion in Structural Biology* 17(6):633-640.

109. Nicholson L. K., Yamazaki T., Torchia D. A., Grzesiek S., Bax A., Stahl S. J., Kaufman J. D., Wingfield P. T., Lam P. Y. S., Jadhav P. K., Hodge C. N., Domaille P. J., & Chang C. H. (1995) *Nature Structural Biology* 2(4):274-280.
110. Freedberg D. I., Ishima R., Jacob J., Wang Y. X., Kustanovich I., Louis J. M., & Torchia D. A. (2002) *Protein Science* 11(2):221-232.
111. Ishima R., Freedberg D. I., Wang Y. X., Louis J. M., & Torchia D. A. (1999) *Structure* 7(9):1047-1055.
112. Katoh E., Louis J. M., Yamazaki T., Gronenborn A. M., Torchia D. A., & Ishima R. (2003) *Protein Science* 12(7):1376-1385.
113. Galiano L., Bonora M., & Fanucci G. E. (2007) *Journal of the American Chemical Society* 129(36):11004-11005.
114. Collins J. R., Burt S. K., & Erickson J. W. (1995) *Nature Structural Biology* 2(4):334-338.
115. Scott W. R. P. & Schiffer C. A. (2000) *Structure* 8(12):1259-1265.
116. Meagher K. L. & Carlson H. A. (2005) *Proteins-Structure Function and Bioinformatics* 58(1):119-125.
117. Hamelberg D. & McCammon J. A. (2005) *Journal of the American Chemical Society* in press.
118. Perryman A. L., Lin J. H., & McCammon J. A. (2004) *Protein Science* 13(4):1108-1123.
119. Perryman A. L., Lin J. H., & McCammon J. A. (2006) *Biopolymers* 82(3):272-284.

120. Perryman A. L., Lin J. H., & McCammon J. A. (2006) *Chemical Biology & Drug Design* 67(5):336-345.
121. Chang C. E., Shen T., Trylska J., Tozzini V., & McCammon J. A. (2006) *Biophysical Journal* 90(11):3880-3885.
122. Tozzini V. & McCammon J. A. (2005) *Chemical Physics Letters* 413(1-3):123-128.
123. Hornak V., Okur A., Rizzo R. C., & Simmerling C. (2006) *Journal of the American Chemical Society* 128(9):2812-2813.
124. Toth G. & Borics A. (2006) *Biochemistry* 45(21):6606-6614.
125. Prabu-Jeyabalan M., Nalivaika E. A., Romano K., & Schiffer C. A. (2006) *The Journal of Virology* 80(7):3607-3616.
126. Hornak V. & Simmerling C. (2007) *Drug Discovery Today* 12(3-4):132-138.
127. De Clercq E. (2009) *International Journal of Antimicrobial Agents* 33(4):307-320.
128. Pokorná J., Machala L., Řezáčová P., & Konvalinka J. (2009) *Viruses* 1(3):1209-1239.
129. Bennett D. E., Camacho R. J., Otelea D., Kuritzkes D. R., Fleury H., Kiuchi M., Heneine W., Kantor R., Jordan M. R., Schapiro J. M., Vandamme A. M., Sandstrom P., Boucher C. A. B., van de Vijver D., Rhee S. Y., Liu T. F., Pillay D., & Shafer R. W. (2009) *PLoS One* 4(3):e4724.
130. Prabu-Jeyabalan M., Nalivaika E., & Schiffer C. A. (2002) *Structure* 10(3):369-381.

131. Piana S., Carloni P., & Rothlisberger U. (2002) *Protein Science* 11(10):2393-2402.
132. Piana S., Carloni P., & Parrinello M. (2002) *Journal of Molecular Biology* 319(2):567-583.
133. Martin P., Vickrey J. F., Proteasa G., Jimenez Y. L., Wawrzak Z., Winters M. A., Merigan T. C., & Kovari L. C. (2005) *Structure* 13(12):1887-1895.
134. Luque I., Todd M. J., Gomez J., Semo N., & Freire E. (1998) *Biochemistry* 37(17):5791-5797.
135. Vega S., Kang L. W., Velazquez-Campoy A., Kiso Y., Amzel L. M., & Freire E. (2004) *Proteins-Structure Function and Bioinformatics* 55(3):594-602.
136. Baldwin E. T., Bhat T. N., Liu B. S., Pattabiraman N., & Erickson J. W. (1995) *Nature Structural Biology* 2(3):244-249.
137. Rick S. W., Erickson J. W., & Burt S. K. (1998) *Proteins-Structure Function and Bioinformatics* 32(1):7-16.
138. Todd M. J., Luque I., Velazquez-Campoy A., & Freire E. (2000) *Biochemistry* 39(39):11876-11883.
139. Ohtaka H., Muzammil S., Schon A., Velazquez-Campoy A., Vega S., & Freire E. (2004) *International Journal of Biochemistry & Cell Biology* 36(9):1787-1799.
140. Yoshimura K., Kato R., Yusa K., Kavlick M. F., Maroun V., Nguyen A., Mimoto T., Ueno T., Shintani M., Falloon J., Masur H., Hayashi H., Erickson J., & Mitsuya H. (1999) *Proceedings of the National Academy of Sciences of the United States of America* 96(15):8675-8680.

141. Velazquez-Campoy A. & Freire E. (2001) *Journal of Cellular Biochemistry*:82-88.
142. Velazquez-Campoy A., Kiso Y., & Freire E. (2001) *Archives of Biochemistry and Biophysics* 390(2):169-175.
143. Ghosh A. K., Sridhar P. R., Leshchenko S., Hussain A. K., Li J. F., Kovalevsky A. Y., Walters D. E., Wedekind J. E., Grum-Tokars V., Das D., Koh Y., Maeda K., Gatanaga H., Weber I. T., & Mitsuya H. (2006) *Journal of Medicinal Chemistry* 49(17):5252-5261.
144. Tie Y. F., Boross P. I., Wang Y. F., Gaddis L., Hussain A. K., Leshchenko S., Ghoshl A. K., Louis J. M., Harrison R. W., & Weber I. T. (2004) *Journal of Molecular Biology* 338(2):341-352.
145. Hou T., McLaughlin W. A., & Wang W. (2008) *Proteins* 71(3):1163-1174.
146. Fontenot G., Johnston K., Cohen J. C., Gallaher W. R., Robinson J., & Luftig R. B. (1992) *Virology* 190(1):1-10.
147. Paul L. Darke D. L. H., Susan P. Jordan, Joan A. Zugay, Jules A. Shafer, Lawrence C. Kuo (1994) *Biochemistry* 33(1):98-105.
148. Kuzmic P. (1993) *Biochemical and Biophysical Research Communications* 191(3):998-1003.
149. Xie D., Gulnik S., Gustchina E., Yu B., Shao W., Qoronfleh W., Nathan A., & Erickson J. W. (1999) *Protein Sci* 8(8):1702-1707.
150. Weber I. T. (1990) *Journal of Biological Chemistry* 265(18):10492-10496.
151. Todd M. J., Semo N., & Freire E. (1998) *Journal of Molecular Biology* 283(2):475-488.

152. Zhang Z. Y., Poorman R. A., Maggiora L. L., Heinrikson R. L., & Kézdy F. J. (1991) *Journal of Biological Chemistry* 266(24):15591-15594.
153. Babé L. M., Rosé J., & Craik C. S. (1992) *Protein Science* 1(10):1244-1253.
154. Schramm H. J., Boetzel J., Büttner J., Fritsche E., Göhring W., Jaeger E., König S., Thumfart O., Wenger T., Nagel N. E., & Schramm W. (1996) *Antiviral Research* 30(2-3):155-170.
155. Zutshi R., Franciskovich J., Shultz M., Schweitzer B., Bishop P., Wilson M., & Chmielewski J. (1997) *Journal of the American Chemical Society* 119(21):4841-4845.
156. Zutshi R. & Chmielewski J. (2000) *Bioorganic & Medicinal Chemistry Letters* 10(17):1901-1903.
157. Bouras A., Boggetto N., Benatalah Z., de Rosny E., Sicsic S., & Reboud-Ravaux M. (1999) *Journal of Medicinal Chemistry* 42(6):957-962.
158. Bannwarth L., Kessler A., Pèthe S., Collinet B., Merabet N., Boggetto N., Sicsic S., Reboud-Ravaux M., & Ongerì S. (2006) *Journal of Medicinal Chemistry* 49(15):4657-4664.
159. Bowman M. J. & Chmielewski J. (2009) *Bioorganic & Medicinal Chemistry* 17(3):967-976.
160. Schramm H. J., Rosny E. D., Reboud-Ravaux M., Büttner J., Dick A., & Schramm W. (2005) *Biological Chemistry* 380(5):593-596.
161. Dumond J., Boggetto N., Schramm H. J., Schramm W., Takahashi M., & Reboud-Ravaux M. (2003) *Biochemical Pharmacology* 65(7):1097-1102.

162. Rezacova P., Brynda J., Fabry M., Horejsi M., Stouracova R., Lescar J., Chitarra V., Riottot M. M., Sedlacek J., & Bentley G. A. (2002) *Journal of Molecular Recognition* 15(5):272-276.
163. Bartonová V., Král V., Siegllová I., Brynda J., Fábry M., Horejsí M., Kozísek M., Sasková K. G., Konvalinka J., Sedláček J., & Rezacová P. (2008) *Antiviral Research* 78(3):275-277.
164. Davis D. A., Brown C. A., Singer K. E., Wang V., Kaufman J., Stahl S. J., Wingfield P., Maeda K., Harada S., Yoshimura K., Kosalaraksa P., Mitsuya H., & Yarchoan R. (2006) *Antiviral Research* 72(2):89-99.
165. Lescar J., Brynda J., Bentley G. A., Riottot M.-M., Chitarra V., Rezacova P., Stouracova R., Fabry M., Horejsi M., & Sedlacek J. (1999) *Protein Science* 8(12):2686-2696.
166. Rezacova P., Lescar J., Brynda J., Fabry M., Horejsi M., Sedlacek J., & Bentley G. A. (2001) *Structure* 9(10):887-895.
167. Frutos S., Rodriguez-Mias R. A., Madurga S., Collinet B., Reboud-Ravaux M., Ludevid D., & Giralt E. (2007) *Peptide Science* 88(2):164-173.
168. Gunasekaran K., Ma B. Y., & Nussinov R. (2004) *Proteins-Structure Function and Bioinformatics* 57(3):433-443.
169. Damm K. L., Ung P. M. U., Quintero J. J., Gestwicki J. E., & Carlson H. A. (2008) *Biopolymers* 89(8):643-652.
170. Pillai B., Kannan K. K., & Hosur M. V. (2001) *Proteins-Structure Function and Genetics* 43(1):57-64.

171. Bowman M. J., Byrne S., & Chmielewski J. (2005) *Chemistry & Biology* 12(4):439-444.
172. Rezacova P., Brynda J., Lescar J., Fabry M., Horejsi M., Sieglova I., Sedlacek J., & Bentley G. A. (2005) *Journal of Structural Biology* 149(3):332-337.
173. Sperka T., Pitlik J., Bagossi P., & Tozser J. (2005) *Bioorganic & Medicinal Chemistry Letters* 15(12):3086-3090.
174. Judd D. A., Nettles J. H., Nevins N., Snyder J. P., Liotta D. C., Tang J., Ermolieff J., Schinazi R. F., & Hill C. L. (2001) *Journal of the American Chemical Society* 123(5):886-897.
175. Perryman A. L., Zhang Q., Soutter H. H., Rosenfeld R., McRee D. E., Olson A. J., Elder J. E., & Stout C. D. (2010) *Chemical Biology & Drug Design* 75(3):257-268.
176. Řezáčová P. n., Pokorná J., Brynda J. i., Kozíšek M., Cígler P., Lepšík M., Fanfrlík J. i., Řezáč J., Grantz Šašková K. r., Sieglová I., Plešek J. r., Šícha V. c., Grüner B. r., Oberwinkler H., Sedláček J., Kräusslich H.-G., Hobza P., Král V. r., & Konvalinka J. (2009) *Journal of Medicinal Chemistry*.
177. Seelmeier S., Schmidt H., Turk V., & Vonderhelm K. (1988) *Proceedings of the National Academy of Sciences of the United States of America* 85(18):6612-6616.
178. Stephenson J. (1997) *Jama-Journal of the American Medical Association* 277(8):614-616.
179. Palella F. J., Delaney K. M., Moorman A. C., Loveless M. O., Fuhrer J., Satten G. A., Aschman D. J., Holmberg S. D., & Investigators H. I. V. O. S. (1998) *New England Journal of Medicine* 338(13):853-860.

180. Condra J. H., Schleif W. A., Blahy O. M., Gabryelski L. J., Graham D. J., Quintero J. C., Rhodes A., Robbins H. L., Roth E., Shivaprakash M., Titus D., Yang T., Teppler H., Squires K. E., Deutsch P. J., & Emini E. A. (1995) *Nature* 374(6522):569-571.
181. Heaslet H., Rosenfeld R., Giffin M., Lin Y.-C., Tam K., Torbett B. E., Elder J. H., McRee D. E., & Stout C. D. (2007) *Acta Crystallographica Section D* 63(8):866-875.
182. Martin P., Vickrey J. F., Proteasa G., Jimenez Y. L., Wawrzak Z., Winters M. A., Merigan T. C., & Kovari L. C. (2005) *Structure* 13(12):1887-1895.
183. Bindu P., Kannan K. K., & Hosur M. V. (2001) *Proteins* 43(1):57-64.
184. Nicholson L. K., Yamazaki T., Torchia D. A., Grzesiek S., Bax A., Stahl S. J., Kaufman J. D., Wingfield P. T., Lam P. Y. S., Jadhav P. K., Hodge C. N., Domaille P. J., & Chang C.-H. (1995) *Nature Structural Molecular Biology* 2(4):274-280.
185. Ishima R., Freedberg D. I., Wang Y.-X., Louis J. M., & Torchia D. A. (1999) *Structure* 7(9):1047-1055.
186. Ding F., Layten M., & Simmerling C. (2008) *Journal of the American Chemical Society* 130(23):7184-7185.
187. Galiano L., Ding F., Veloro A. M., Blackburn M. E., Simmerling C., & Fanucci G. E. (2008) *Journal of the American Chemical Society* 131(2):430-431.
188. Foulkes-Murzycki J. E., Scott W. R. P., & Schiffer C. A. (2007) *Structure* 15(2):225-233.

189. Kurt N., Scott W. R. P., Schiffer C. A., & Haliloglu T. (2003) *Proteins* 51(3):409-422.
190. Li D., Ji B., Hwang K., & Huang Y. (2010) *The Journal of Physical Chemistry B*.
191. Pietrucci F., Marinelli F., Carloni P., & Laio A. (2009) *Journal of the American Chemical Society* 131(33):11811-11818.
192. Levy R. M., Zhang L. Y., Gallicchio E., & Felts A. K. (2003) *Journal of the American Chemical Society* 125(31):9523-9530.
193. Geney R., Layten M., Gomperts R., Hornak V., & Simmerling C. (2006) *Journal of Chemical Theory and Computation* 2(1):115-127.
194. Zhou R. (2003) *Proteins: Structure, Function, and Bioinformatics* 53(2):148-161.
195. Spinelli S., Liu Q. Z., Alzari P. M., Hirel P. H., & Poljak R. J. (1991) *Biochimie* 73(11):1391-1396.
196. Case D. A., Thomas E. Cheatham III, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M. Merz Jr., Alexey Onufriev, Carlos Simmerling, Bing Wang, & Woods R. J. (2005) *Journal of Computational Chemistry* 26(16):1668-1688.
197. Hornak V., Abel R., Okur A., Strockbine B., Roitberg A., & Simmerling C. (2006) *Proteins* 65(3):712-725.
198. Nicolas G. & Manuel C. P. (1997) *Electrophoresis* 18:2714-2723.
199. Ryckaert J.-P., Ciccotti G., & Berendsen H. J. C. (1977) *Journal of Computational Physics* 23(3):327-341.
200. Berendsen H. J. C., Postma J. P. M., Vangunsteren W. F., Dinola A., & Haak J. R. (1984) *Journal of Chemical Physics* 81(8):3684-3690.

201. Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C., & Ferrin T. E. (2004) *Journal of Computational Chemistry* 25(13):1605-1612.
202. Shao W., Everitt L., Manchester M., Loeb D. D., Hutchison C. A., & Swanstrom R. (1997) *Proceedings of the National Academy of Sciences of the United States of America* 94(6):2243-2248.
203. Heaslet H., Rosenfeld R., Giffin M., Lin Y. C., Tam K., Torbett B. E., Elder J. H., McRee D. E., & Stout C. D. (2007) *Acta Crystallogr D Biol Crystallogr* 63(Pt 8):866-875.
204. Finkelstein A. V. (1997) *Protein Eng* 10(8):843-845.
205. Dinner A. R. & Karplus M. (1999) *Journal of Molecular Biology* 292(2):403-419.
206. Day R., Bennion B. J., Ham S., & Daggett V. (2002) *J Mol Biol* 322(1):189-203.
207. Mayor U., Guydosh N. R., Johnson C. M., Grossmann J. G., Sato S., Jas G. S., Freund S. M. V., Alonso D. O. V., Daggett V., & Fersht A. R. (2003) *Nature* 421(6925):863-867.
208. Ishima R., Louis J. M., & Torchia D. A. (2001) *Journal of Molecular Biology* 305(3):515-521.
209. Wang W. & Kollman P. A. (2000) *Journal of Molecular Biology* 303(4):567-582.
210. Shafer R. W., Stevenson D., & Chan B. (1999) *Nucleic Acids Research* 27(1):348-352.
211. Ermolieff J., Lin X., & Tang J. (1997) *Biochemistry* 36(40):12364-12370.

212. Lin Y. Z., Lin X. L., Hong L., Foundling S., Heinrikson R. L., Thaisrivongs S., Leelamanit W., Raterman D., Shah M., Dunn B. M., & Tang J. (1995) *Biochemistry* 34(4):1143-1152.
213. Veloro A. M., Blackburn M. E., Galiano L., & Fanucci G. E. (2009) *Biophysical Journal* 96(3, Supplement 1):70a-70a.
214. Louis J. M., Clore G. M., & Gronenborn A. M. (1999) *Nature Structural Biology* 6(9):868-875.
215. <http://hivdb.stanford.edu/>.
216. Loeb D. D., Swanstrom R., Everitt L., Manchester M., Stamper S. E., & Hutchison C. A. (1989) *Nature* 340(6232):397-400.
217. Wu T. D., Schiffer C. A., Gonzales M. J., Taylor J., Kantor R., Chou S., Israelski D., Zolopa A. R., Fessel W. J., & Shafer R. W. (2003) *The Journal of Virology* 77(8):4836-4847.
218. Ishima R., Ghirlando R., Tozser J., Gronenborn A. M., Torchia D. A., & Louis J. M. (2001) *Journal of Biological Chemistry* 276(52):49110-49116.
219. Glaser F., Pupko T., Paz I., Bell R. E., Bechor-Shental D., Martz E., & Ben-Tal N. (2003) *Bioinformatics* 19(1):163-164.
220. Oroszlan S. & Luftig R. B. (1990) *Current Topics in Microbiology and Immunology* 157:153-185.
221. Krausslich H. G. (1991) *Proceedings of the National Academy of Sciences of the United States of America* 88(8):3213-3217.
222. Levy Y., Caflish, Amedeo, Onuchic, Jose N., Wolynes, Peter G. (2004) *Journal of Molecular Biology* 340(1):67-79.

223. Bindu Pillai K. K. K., M. V. Hosur (2001) *PROTEINS: Structure, Function, and Genetics* 43(1):57-64.
224. Meagher K. L. & Carlson H. A. (2005) *Proteins* 58(1):119-125.
225. Hamelberg D. & McCammon J. A. (2005) *Journal of the American Chemical Society* 127(40):13778-13779.
226. Wiley A. P., Williams S. L., & Essex J. W. (2009) *Journal of Chemical Theory and Computation* (0).
227. Chang C.-E., Shen T., Trylska J., Tozzini V., & McCammon J. A. (2006) *Biophysical Journal* 90(11):3880-3885.
228. Sanner M. F., Olson A. J., & Spehner J.-C. (1996) *Biopolymers* 38(3):305-320.
229. Naïm M., Bhat S., Rankin K. N., Dennis S., Chowdhury S. F., Siddiqi I., Drabik P., Sulea T., Bayly C. I., Jakalian A., & Purisima E. O. (2007) *Journal of Chemical Information and Modeling* 47(1):122-133.
230. Cui Q., Sulea T., Schrag J. D., Munger C., Hung M.-N., Naïm M., Cygler M., & Purisima E. O. (2008) *Journal of Molecular Biology* 379(4):787-802.
231. Purisima E. O. & Nilar S. H. (1995) *Journal of Computational Chemistry* 16(6):681-689.
232. Purisima E. O. (1998) *Journal of Computational Chemistry* 19(13):1494-1504.
233. Chan S. L. & Purisima E. O. (1998) *Journal of Computational Chemistry* 19(11):1268-1277.
234. Bhat S. & Purisima E. O. (2006) *Proteins: Structure, Function, and Bioinformatics* 62(1):244-261.

235. Collins J. R., Burt S. K., & Erickson J. W. (1995) *Nat Struct Mol Biol* 2(4):334-338.
236. Tozzini V., Trylska J., Chang C.-e., & McCammon J. A. (2007) *Journal of Structural Biology* 157(3):606-615.
237. Cheng Y.-S. E., Yin F. H., Foundling S., Blomstrom D., & Kettner C. A. (1990) *Proceedings of the National Academy of Sciences of the United States of America* 87(24):9660-9664.
238. Grant S. K., Deckman I. C., Culp J. S., Minnich M. D., Brooks I. S., Hensley P., Debouck C., & Meek T. D. (1992) *Biochemistry* 31(39):9491-9501.
239. Jordan S. P., Zugay J., Darke P. L., & Kuo L. C. (1992) *Journal of Biological Chemistry* 267(28):20028-20032.
240. Kollman P. A., Massova I., Reyes C., Kuhn B., Huo S. H., Chong L., Lee M., Lee T., Duan Y., Wang W., Donini O., Cieplak P., Srinivasan J., Case D. A., & Cheatham T. E. (2000) *Accounts of Chemical Research* 33(12):889-897.
241. Humphrey W., Dalke A., & Schulten K. (1996) *J Mol Graphics Modell* 14(1):33-&.
242. Joanna Trylska J. A., Maciej Geller, C. Nicholas Hodge, Ronald M. Klabe, Martha S. Head, Michael K. Gilson (1999) *Protein Science* 8(1):180-195.
243. Koh Y., Matsumi S., Das D., Amano M., Davis D. A., Li J., Leschenko S., Baldrige A., Shioda T., Yarchoan R., Ghosh A. K., & Mitsuya H. (2007) *Journal of Biological Chemistry* 282(39):28709-28720.
244. Rhee S. Y., Gonzales M. J., Kantor R., Betts B. J., Ravela J., & Shafer R. W. (2003) *Nucleic Acids Research* 31(1):298-303.

245. Hwang Y. S. & Chmielewski J. (2004) *Journal of Medicinal Chemistry* 48(6):2239-2242.
246. Shultz M. D., Ham Y. W., Lee S. G., Davis D. A., Brown C., & Chmielewski J. (2004) *Journal of the American Chemical Society* 126(32):9886-9887.
247. Colman P. M. (2009) *Annual Review of Biochemistry* 78(1):95-118.
248. Price E. A., Sutch B. T., Cai Q., Qin P. Z., & Haworth I. S. (2007) *Biopolymers* 87(1):40-50.
249. Zoete V., Michielin O., & Karplus M. (2002) *Journal of Molecular Biology* 315(1):21-52.
250. Vondrasek J. & Wlodawer A. (2002) *Proteins: Structure, Function, and Bioinformatics* 49(4):429-431.
251. York D. M., Darden T. A., Pedersen L. G., & Anderson M. W. (1993) *Biochemistry* 32(6):1443-1453.
252. 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., Wideburg N. E., Saldivar A., Ruiz L., Kati W. M., Sham H. L., Robins T., Stewart K. D., Hsu A., Plattner J. J., Leonard J. M., & Norbeck D. W. (1995) *Proceedings of the National Academy of Sciences of the United States of America* 92(7):2484-2488.
253. 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. C. (1994) 33(32):9405-9413.
254. Kear J. L., Blackburn M. E., Veloro A. M., Dunn B. M., & Fanucci G. E. (2009) *Journal of the American Chemical Society* 131(41):14650-14651.

255. Ashorn P., McQuade T. J., Thaisrivongs S., Tomasselli A. G., Tarpley W. G., & Moss B. (1990) *Proceedings of the National Academy of Sciences of the United States of America* 87(19):7472-7476.
256. Walensky R., Paltiel A. D., Losina E., Mercincavage L., Schackman B., Sax P., Weinstein M., & Freedberg K. (2006) *The Journal of Infectious Diseases* 194(1):11-19.
257. Little S. J., Holte S., Routy J.-P., Daar E. S., Markowitz M., Collier A. C., Koup R. A., Mellors J. W., Connick E., Conway B., Kilby M., Wang L., Whitcomb J. M., Hellmann N. S., & Richman D. D. (2002) *The New England Journal of Medicine* 347(6):385-394.
258. Clemente J. C., Hemrajani R., Blum L. E., Goodenow M. M., & Dunn B. M. (2003) *Biochemistry* 42(51):15029-15035.
259. Clemente J. C., Coman R. M., Thiaville M. M., Janka L. K., Jeung J. A., Nukoolkarn S., Govindasamy L., Agbandje-McKenna M., McKenna R., Leelamanit W., Goodenow M. M., & Dunn B. M. (2006) *Biochemistry* 45(17):5468-5477.
260. Mahalingam B., Boross P., Wang Y.-F., Louis J. M., Fischer C. C., Tozser J., Harrison R. W., & Weber I. T. (2002) *PROTEINS: Structure, Function, and Genetics* 48(1):107-116.
261. 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. (2004) *Biochemistry* 43(38):12141-12151.

262. Tsai C.-J., Ma B., & Nussinov R. (1999) *Proceedings of the National Academy of Sciences of the United States of America* 96(18):9970-9972.
263. Luque I. & Freire E. (2000) *PROTEINS: Structure, Function, and Genetics* 41(S4):63-71.
264. Davis A. M. & Teague S. J. (1999) *Angewandte Chemie International Edition* 38(6):736-749.
265. Rauh D., Klebe G., & Stubbs M. T. (2004) *Journal of Molecular Biology* 335(5):1325-1341.
266. Clemente J. C., Moose R. E., Hemrajani R., Whitford L. R. S., Govindasamy L., Reutzler R., McKenna R., Agbandje-McKenna M., Goodenow M. M., & Dunn B. M. (2004) *Biochemistry* 43(38):12141-12151.
267. Wartha F., Horn A. H. C., Meiselbach H., & Sticht H. (2005) *Journal of Chemical Theory and Computation* 1(2):315-324.
268. Ding F., Layten M., & Simmerling C. (2008) *J Am Chem Soc* 130(23):7184-7185.
269. Liu F., Kovalevsky A. Y., Louis J. M., Boross P. I., Wang Y.-F., Harrison R. W., & Weber I. T. (2006) *Journal of Molecular Biology* 358(5):1191-1199.
270. Lovell S. C., Davis I. W., Adrendall W. B., de Bakker P. I. W., Word J. M., Prisant M. G., Richardson J. S., & Richardson D. C. (2003) *Proteins-Structure Function and Genetics* 50(3):437-450.
271. Cai Q., Kusnetzow A., Hubbell W., Haworth I., Gacho G., Van Eps N., Hideg K., Chambers E., & Qin P. (2007) *Biophysical Journal*:413a-413a.
272. Jeschke G., Chechik V., Ionita P., Godt A., Zimmermann H., Banham J., Timmel C. R., Hilger D., & Jung H. (2006) *Applied Magnetic Resonance* 30(3-4):473-498.

273. Chiang Y.-W., Borbat P. P., & Freed J. H. (2005) *Journal of Magnetic Resonance* 172(2):279-295.

Appendix 1- Parameters of Ritonavir

0 0 2

This is a remark line

molecule.res

RIT INT 0

CORRECT OMIT DU BEG

0.0000

1 DUMM DU M 0 -1 -2

0.000 .0 .0 .00000

2 DUMM DU M 1 0 -1

1.449 .0 .0 .00000

3 DUMM DU M 2 1 0 1.522

111.1 .0 .00000

4 C48 c3 M 3 2 1 1.540 111.208 180.000 -

0.05715

5 H86 hc E 4 3 2 1.071 54.216 -153.464

0.04140

6 H87 hc E 4 3 2 1.070 74.272 -24.375

0.05002

7 H88 hc E 4 3 2 1.070 84.422 87.632

0.04384

8 C47 c3 M 4 3 2 1.525 162.401 -129.281 -

0.15284

9 C49 c3 3 8 4 3 1.479 105.671 110.009 -

0.09652

10 H89 hc E 9 8 4 1.069 109.480 -144.480

0.06233

11 H90 hc E 9 8 4 1.070 109.413 -24.467

0.03902

12 H91 hc E 9 8 4 1.070 109.446 95.525

0.03551

13 H98 hc E 8 4 3 1.070 102.409 -140.775

0.06700

14 C45 cd M 8 4 3 1.380 132.652 -20.642

0.53884

15 N46 nc E 14 8 4 1.306 140.083 131.630 -

0.64913

16 S44 ss M 14 8 4 1.774 108.379 -49.361 -

0.31293

17 C43 cd M 16 14 8 1.783 89.130 -179.492 -

0.12988

18 H85 h4 E 17 16 14 1.070 126.769 -179.610

0.19145

19 C42 cc M 17 16 14 1.369 106.430 0.389

0.31818

20 C40 c3 M 19 17 16 1.472 118.181 178.651

0.06379

21 H83 h1 E 20 19 17 1.070 105.507 148.108

0.07281

22 H84 h1 E 20 19 17 1.071 105.479 20.131

0.08498

23	N39	n	M	20	19	17	1.474	115.434	-95.860	-
0.45512										
24	C50	c3	3	23	20	19	1.338	110.194	-80.145	
0.08766										
25	H92	h1	E	24	23	20	1.070	109.463	13.984	
0.06518										
26	H93	h1	E	24	23	20	1.070	109.444	133.998	
0.05233										
27	H97	h1	E	24	23	20	1.070	109.439	-106.011	
0.00945										
28	C17	c	M	23	20	19	1.381	120.951	99.706	
0.74959										
29	O41	o	E	28	23	20	1.259	124.917	0.259	-
0.66438										
30	N16	n	M	28	23	20	1.361	113.784	-179.710	-
0.49122										
31	H59	hn	E	30	28	23	1.000	103.550	-71.705	
0.27846										
32	C15	c3	M	30	28	23	1.485	124.132	171.289	
0.00414										
33	C36	c3	3	32	30	28	1.574	108.324	127.891	-
0.08662										
34	C37	c3	3	33	32	30	1.551	116.188	178.692	-
0.07909										
35	H52	hc	E	34	33	32	1.070	109.469	179.963	
0.04428										
36	H53	hc	E	34	33	32	1.070	109.477	60.011	
0.02956										
37	H54	hc	E	34	33	32	1.070	109.461	-60.003	
0.04868										
38	C38	c3	3	33	32	30	1.547	110.982	-58.094	-
0.09067										
39	H55	hc	E	38	33	32	1.070	109.449	-179.989	
0.04802										
40	H56	hc	E	38	33	32	1.070	109.478	59.957	
0.02498										
41	H57	hc	E	38	33	32	1.070	109.459	-59.969	
0.04475										
42	H51	hc	E	33	32	30	1.070	102.930	62.406	
0.07318										
43	H58	h1	E	32	30	28	1.071	112.688	10.304	
0.06153										
44	C14	c	M	32	30	28	1.537	108.639	-107.037	
0.66861										
45	O35	o	E	44	32	30	1.254	123.398	-90.568	-
0.60522										
46	N34	n	M	44	32	30	1.339	112.972	87.796	-
0.56036										
47	H96	hn	E	46	44	32	1.000	117.779	2.162	
0.31455										
48	C13	c3	M	46	44	32	1.452	124.462	-177.836	
0.10505										
49	C27	c3	3	48	46	44	1.511	109.089	126.857	-
0.04403										
50	C28	ca	S	49	48	46	1.456	109.877	-76.540	-
0.09659										

51	C29	ca	B	50	49	48	1.428	116.444	109.391	-
0.12547										
52	C30	ca	B	51	50	49	1.439	120.360	-179.977	-
0.13526										
53	C31	ca	B	52	51	50	1.416	118.743	-0.083	-
0.13145										
54	C32	ca	B	53	52	51	1.415	120.604	-0.022	-
0.12417										
55	C33	ca	S	54	53	52	1.422	120.961	0.030	-
0.11077										
56	H82	ha	E	55	54	53	1.070	120.392	-179.934	
0.13920										
57	H81	ha	E	54	53	52	1.070	119.531	-179.999	
0.12969										
58	H80	ha	E	53	52	51	1.070	119.708	179.960	
0.12593										
59	H79	ha	E	52	51	50	1.070	120.632	179.974	
0.13687										
60	H78	ha	E	51	50	49	1.070	119.810	0.025	
0.12839										
61	H76	hc	E	49	48	46	1.070	106.663	38.758	
0.05768										
62	H77	hc	E	49	48	46	1.070	106.663	168.131	
0.06890										
63	H67	h1	E	48	46	44	1.070	106.170	3.853	
0.10653										
64	C12	c3	M	48	46	44	1.540	112.819	-116.835	-
0.14162										
65	H65	hc	E	64	48	46	1.070	104.364	-64.477	
0.04450										
66	H66	hc	E	64	48	46	1.070	104.351	168.787	
0.07650										
67	C11	c3	M	64	48	46	1.506	120.562	52.172	
0.14333										
68	O26	oh	S	67	64	48	1.404	107.510	-162.550	-
0.60160										
69	H75	ho	E	68	67	64	0.960	109.490	64.225	
0.40678										
70	H64	h1	E	67	64	48	1.070	106.889	-39.881	
0.04072										
71	C10	c3	M	67	64	48	1.538	114.374	78.311	
0.13447										
72	C19	c3	3	71	67	64	1.550	109.414	-165.941	-
0.04502										
73	C20	ca	S	72	71	67	1.448	121.098	-178.922	-
0.09216										
74	C21	ca	B	73	72	71	1.428	110.201	-88.591	-
0.12618										
75	C22	ca	B	74	73	72	1.419	118.120	179.517	-
0.12823										
76	C23	ca	B	75	74	73	1.431	121.593	0.080	-
0.13011										
77	C24	ca	B	76	75	74	1.441	119.651	0.100	-
0.12953										
78	C25	ca	S	77	76	75	1.417	119.556	-0.206	-
0.11033										

79	H74	ha	E	78	77	76	1.070	120.145	-179.848	
0.13157										
80	H73	ha	E	77	76	75	1.071	120.232	179.788	
0.13146										
81	H72	ha	E	76	75	74	1.071	120.190	-179.870	
0.13120										
82	H71	ha	E	75	74	73	1.070	119.175	-179.914	
0.12751										
83	H70	ha	E	74	73	72	1.070	120.944	-0.498	
0.14210										
84	H68	hc	E	72	71	67	1.070	104.233	64.368	
0.06556										
85	H69	hc	E	72	71	67	1.070	104.229	-62.259	
0.06096										
86	H63	h1	E	71	67	64	1.070	107.134	-48.626	
0.08875										
87	N9	n	M	71	67	64	1.480	115.563	63.098	-
0.55350										
88	H95	hn	E	87	71	67	1.001	118.810	66.956	
0.33188										
89	C8	c	M	87	71	67	1.370	122.384	-113.015	
0.76223										
90	O18	o	E	89	87	71	1.240	125.533	-0.206	-
0.58657										
91	O7	os	M	89	87	71	1.403	112.141	179.301	-
0.47630										
92	C6	c3	M	91	89	87	1.414	126.036	-179.409	
0.25011										
93	H62	h1	E	92	91	89	1.070	110.268	-60.946	
0.05062										
94	H94	h1	E	92	91	89	1.070	110.263	61.133	
0.05162										
95	C2	cd	M	92	91	89	1.422	105.439	-179.911	-
0.15384										
96	C1	cc	M	95	92	91	1.363	138.076	-67.796	
0.40555										
97	H60	h4	E	96	95	92	1.070	122.771	-0.791	
0.04451										
98	N5	nc	M	96	95	92	1.278	114.507	179.187	-
0.60884										
99	C4	cd	M	98	96	95	1.285	107.875	-0.121	
0.44772										
100	H61	h5	E	99	98	96	1.070	118.217	-179.867	
0.02887										
101	S3	ss	M	99	98	96	1.777	123.581	0.104	-
0.22820										

LOOP

C42	N46		
C33	C28		
C25	C20		
S3	C2		

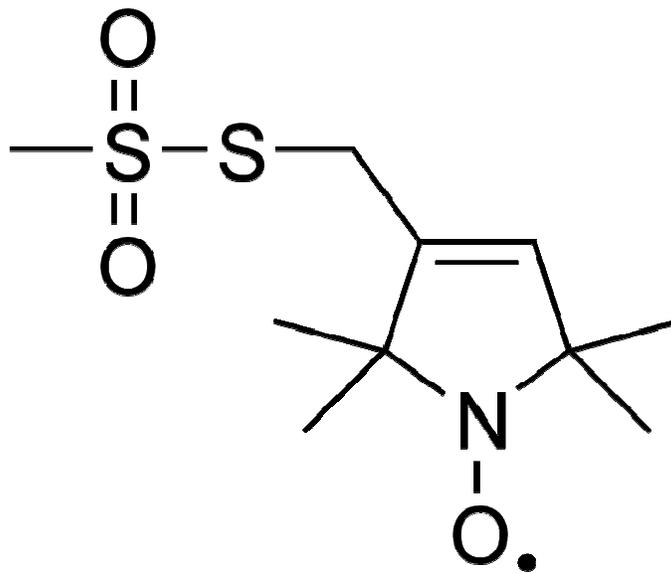
IMPROPER

C47	N46	C45	S44
C42	H85	C43	S44

C40	C43	C42	N46
N39	N16	C17	O41
C15	N34	C14	O35
C27	C29	C28	C33
C28	C30	C29	H78
C29	C31	C30	H79
C30	C32	C31	H80
C31	C33	C32	H81
C28	C32	C33	H82
C19	C21	C20	C25
C20	C22	C21	H70
C21	C23	C22	H71
C22	C24	C23	H72
C23	C25	C24	H73
C20	C24	C25	H74
N9	O18	C8	O7
C6	C1	C2	S3
C2	H60	C1	N5
H61	N5	C4	S3

DONE
STOP

Appendix 2- Structure and the AMBER force field parameters of the phosphorothioate-substituted nitroxide spin label



remark goes here
MASS

BOND

ANGLE

c3-cd-ss	71.372	118.380	Calculated with empirical approach
cc-c3-n	66.905	110.900	Calculated with empirical approach

DIHE

c3-cd-ss-cd	1	1.100	180.000	2.000	same as X -
c2-ss-X					
cd-ss-cd-h4	1	1.100	180.000	2.000	same as X -
c2-ss-X					
cd-ss-cd-cc	1	1.100	180.000	2.000	same as X -
c2-ss-X					
nc-cd-ss-cd	1	1.100	180.000	2.000	same as X -
c2-ss-X					
cd-ss-cd-h5	1	1.100	180.000	2.000	same as X -
c2-ss-X					

IMPROPER

NONBON