Computer-aided Drug Design Targeting Aspartic Proteases

A Dissertation Presented

by

Yi Shang

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

May 2013

Stony Brook University

The Graduate School

Yi Shang

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Carlos L. Simmerling, Ph.D. – Dissertation Advisor
Professor, Department of Chemistry

Erwin London, Ph.D. - Chairperson of Defense
Professor, Department of Biochemistry and Cell Biology

Steven O. Smith, Ph.D. – Third member
Professor, Department of Biochemistry and Cell Biology

Robert C. Rizzo, Ph.D. – Outside member
Associate Professor, Department of Applied Mathematics and Statistics

This dissertation is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Dissertation

Computer-aided Drug Design Targeting Aspartic Proteases

by

Yi Shang

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

2013

Aspartic proteases are a family of protease enzymes that utilize two aspartate residues to catalyze the hydrolysis of their peptide substrates. Many aspartic-protease-family members are targets in drug discovery. For example, HIV-1 protease (HIVPR) is essential in HIV infection, human T-lymphotropic virus protease (HTLVPR) plays important role in adult T-cell leukemia, and β-secretase is an effective target in Alzheimer's disease. HIVPR has been a major success in structure-guided drug development, largely due to progress in crystallography and computational modeling. To date, there are nine FDA approved drugs acting on HIVPR, while there are currently no drugs targeting other aspartic proteases on the market.

The work herein seeks first to strengthen our understanding of aspartic protease dynamics and function. Armed with new understanding, this work seeks to improve HIVPR inhibitors for drug-resistant strains and non-B subtypes, and develop new inhibitors targeting other aspartic proteases. We used computer modeling to approach these goals.

Firstly, we improved the description of salt bridge strength and geometry in an implicit solvent model, which enabled more efficient sampling of HIVPR dynamics through molecular dynamics (MD) simulations. Secondly, based on the improved implicit solvent model, we studied protease dynamics from two HIV subtypes. We confirmed that the sequence-induced conformational shift suggested previously by our experimental collaborators through EPR, pinpointed key residues responsible for the difference through energy decomposition, and proposed enlarging binding pocket size as a drug-resistance mechanism in HIVPR. Thirdly, we compared the binding affinity change due to drug resistance mutations in HIVPR. We studied the change in protease dynamics upon binding first- and second- generation HIVPR drugs, and gave structural explanation. At last, we combined all existing sequence and structural information of aspartic proteases to give a systematic comparison among aspartic protease family members. Conserved and also distinguishing residues were picked out based on sequence/structural alignment, full-length models of several aspartic proteases were built, and we compared the active site gating mechanism between homodimeric and bilobed aspartic proteases based on MD simulations, providing insights into their inhibitor binding and design.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations

AIDS	acquired immunodeficiency syndrome
APV	Amprenavir (HIV-1 protease inhibitor)
ATV	Atazanavir (HIV-1 protease inhibitor)
BACE	β-secretase 1
Ddi1	DNA damage inducible protein 1
DRV	Darunavir (TMC-114, HIV-1 protease inhibitor)
EIAV	Equine Infectious Anemia Virus
EPR	electron paramagnetic resonance
EXP	explicit solvent
FIV	Feline Immunodeficiency Virus
HAART	highly active antiretroviral therapy
HIV	Human Immunodeficiency Virus
HIVPR	HIV-1 protease
HTLV	Human T-Lymphotropic Virus
IDV	Indinavir (HIV-1 protease inhibitor)
MD	molecular dynamics
MDR	multi-drug resistant
MMPBSA	molecular mechanics Poisson-Boltzmann surface area
MM	molecular mechanics
MUT	mutant
NFV	Nelfinavir (HIV-1 protease inhibitor)
PDB	Protein Data Bank
QM	quantum mechanics
QSAR	quantitative structure-activity relationship
REMD	replica exchange molecular dynamics
RSV	Rous Sarcoma Virus
RTV	Ritonavir (HIV-1 protease inhibitor)
SEM	standard error of the mean
SIV	Simian Immunodeficiency Virus
SQV	Saquinavir (HIV-1 protease inhibitor)
T12	TMC-126 (HIV-1 protease inhibitor)
TI	thermodynamic integration
TPV	Tipranavir (HIV-1 protease inhibitor)
WT	wild type

Acknowledgments

I would like to thank my advisor Professor Carlos Simmerling for giving me the opportunity to work in his group, and for his support and guidance throughout my Ph.D. research. I would like to thank my committee members including Professor Robert Rizzo, Erwin London, Steven Smith, and David Green for their thoughtful suggestions and comments during my committee meetings.

I would like to thank former and present members in the Simmerling lab. The lab is like my family in the US, and the enjoyable atmosphere makes discussions efficient and productive. I would like to thank former members Lauren Wickstrom, Fangyu Ding, Arthur Campbell, Christina Bergonzo for being great examples of hard working and caring senior students. I would not forget the early years in the lab how Carmenza Martinez, Hai Nguyen and I, all in the same year, would support each other and even thought about a new force field named CHM. I would also thank Amber Carr, Cheng-Tsung Lai, Kevin Hauser, and Haoquan Li for their help during lab meetings and presentations. Special thanks go to James Maier who kindly helped me revising manuscripts.

I would like to thank colleague Yulin Huang, who has been my high school friend, graduate school roommate, and who first introduced me to computational biology.

I would also like to thank my father Yunrui Shang, my mother Aiying Gu, and my husband Xiaoyang Gong. My parents have always been supportive in my decision to study abroad, and I am grateful that I grew up in such a loving family. Gong has been my number one listener for all my happiness and sadness. Whenever I get frustrated, he always managed to comfort me and point out the positive side of things. I would not have gone through the Ph.D. easily without the support from my family.

Publications

Shang, Y., Nguyen, H., Wickstrom, L., Okur, A., Simmerling, C. Improving the description of salt bridge strength and geometry in a Generalized Born model. *J Mol Graph Model.* 2011, 29, 676-84.

Shang, Y., Simmerling, C. Molecular dynamics applied in drug discovery: the case of HIV-1 protease. *Methods Mol Biol. 2012,* 819, 527-49.

Chapter 1 Introduction

1.1 Aspartic proteases

1.1.1 HIV-1 protease

HIV-1 protease (HIVPR) is an enzyme of HIV, which causes acquired immunodeficiency syndrome (AIDS). Like other retroviral proteases, HIVPR is an aspartic protease that utilizes two aspartic residues to hydrolyze its natural substrate. It is essential in HIV life cycle. After a HIV-infected host cell has synthesized viral RNA and protein, new virions are assembled and bud from the host cell surface. Viral proteins are synthesized on a single polypeptide chain called Gag. After budding, Gag polypeptides associate with each other and attach to the viral membrane to form the immature virion. Then, HIVPR becomes active and cleaves the Gag polypeptide at specific locations to release viral proteins. These proteins then fold independently and go to designated locations. This rearrangement leads to a morphology change inside the virion that is visible by microscopy, during which the virion mature. Without functional HIVPR, the virion is non-infectious. Therefore, HIVPR has been an effective target in combating HIV infection and AIDS. It is worth noting that HIVPR itself is synthesized through a delicate frameshift control. At a much lower rate than the Gag polypeptide synthesis, a longer polypeptide (Gag-pol) is synthesized due to a frameshift event during viral reverse transcription. The Gag-pol polypeptide contains extra sequence information for viral enzymes. This control ensures that no more enzymes are produced than necessary, leaving majority of resources for other massively needed proteins like the matrix proteins.

The structure of HIVPR is shown in Figure 1-1 (all the molecular graphics presented here are rendered using VMD package [1]). It is a C2-symmetric homodimer, with each monomer having 99 residues. The active site is covered by two flexible β hairpins, one on each monomer, that are called the “flaps”. Other structural elements on each monomer include another two β hairpins (fulcrum and cantilever), a helix, and the termini region. The helix is conserved in other aspartic proteases and has been shown to act as the folding nucleus during HIVPR monomer folding event[2]. The termini are where the two monomers form extensive contact with each other, and thus contribute to the dimer stability.

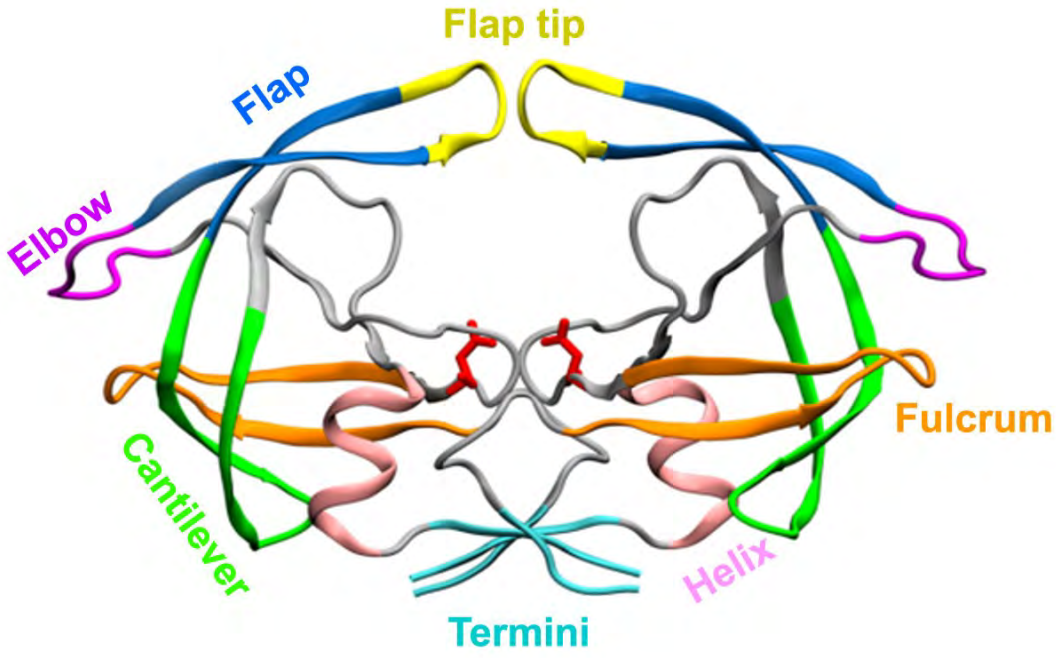

Figure 1-1 HIV-1 protease structure. Illustration generated using crystal structure 1HHP. Heavy atoms of catalytic aspartate residues are shown in red licorice representation. Structural elements are indicated with different colors: flap in blue (residue 43 to 58), flap tip in yellow (residue 49 to 52), flap elbow in magenta (residue 37 to 42), cantilever in green (residue 59 to 75), fulcrum in orange (residue 10 to 23), helix in pink (residue 87 to 94), and termini in cyan (residue 1 to 4 and 96 to 99). All the molecular graphics presented here are rendered using Visual Molecular Dynamics (VMD) package.

Being the most studied aspartic protease so far, HIV-1 protease is a major success in structure-guided drug development. Following the emergence of AIDS in 1981, anti-HIV drug development largely depended on the progress of crystalizing HIV enzymes[3]. HIVPR was the first HIV enzyme to be crystalized in 1989 [4-6] because of its relatively amenable biochemical properties[3, 7]. To date, there are more than 600 crystals of HIVPR deposited in Protein Data Bank (PDB), and there are nine FDA-approved HIVPR drugs on the market.

1.1.2 Other retroviral proteases

Apart from HIV, other retroviruses also depend on proteases for their life cycle. Existing structural information of non-HIV retroviral proteases, except for recently crystalized murine leukemia virus protease as discussed in the last chapter in more detail, are summarized in Figure 1-2. These retroviral proteases share considerable similarity with HIVPR: they are homodimeric, the catalytic site is covered by the flaps, and the termini from each monomer meet to stabilize the dimer. However, there are also noticeable differences from HIVPR: loops are introduced at regions like cantilever tip and fulcrum in RSVPR and HTLVPR, the flap tip and flap elbow are linked by less structured region instead of canonical β-strand in HTLVPR, and helical structures are present near the flap elbow in FIVPR, EIAVPR, and HTLVPR.

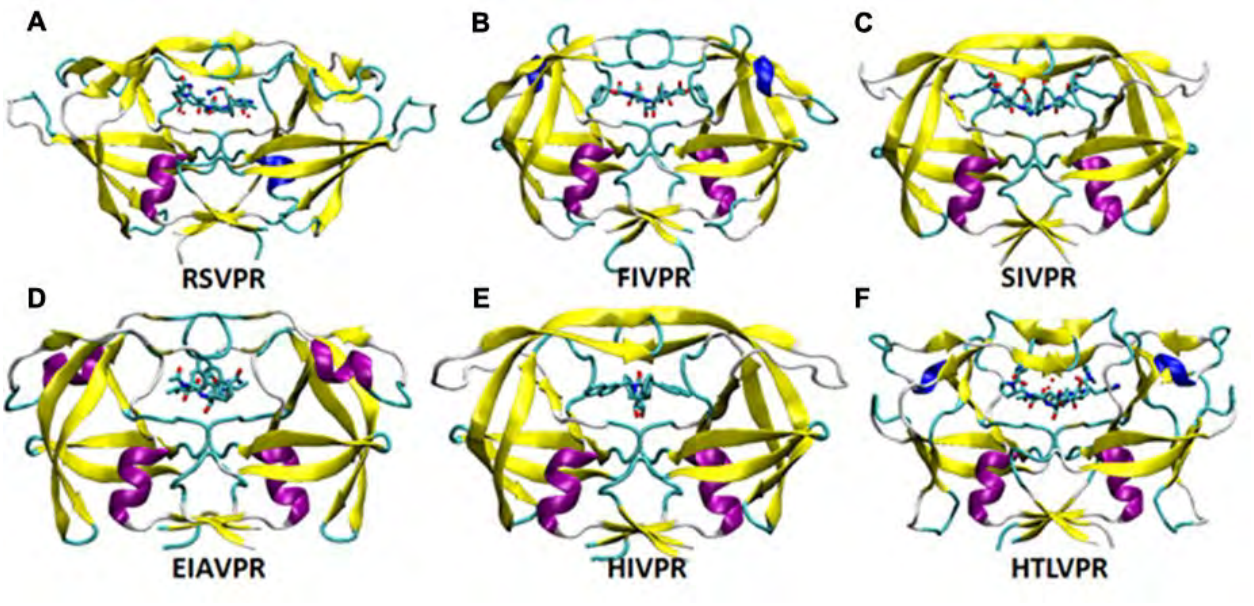

Figure 1-2 Holo crystal structures of proteases from A) Rous Sarcoma Virus (RSV, PDB ID 1BAI), B) Feline Immunodeficiency Virus (FIV, PDB ID 3FIV), C) Simian Immunodeficiency Virus (SIV, PDB ID 1YTI), D) Equine Infectious Anemia Virus (EIAV, PDB ID 1FMB), E) Human Immunodeficiency Virus (HIV, PDB ID 2P3D), and F) Human T-Lymphotropic Virus (HTLV, PDB ID 2B7F). The proteases are colored based on their secondary structure. The heavy atoms of inhibitors are shown in licorice representation.

Among these structures, HTLVPR has the largest deviation from HIVPR and is linked to human adult T-cell leukemia disease. The HTLVPR was crystalized in search of an effective inhibitor of the protease[8, 9]. Other proteases were crystalized primarily for the purpose of comparing HIV to retroviruses infecting other species. Retroviral proteases have been reviewed in detail by Dunn et al. [10].

1.1.3 Non-retroviral aspartic proteases

Retroviral proteases are not the only members of aspartic protease family. Actually, aspartic proteases are common to prokaryotes and eukaryotes as well. Unlike the homodimeric retroviral proteases, however, the non-retroviral aspartic proteases are mostly bilobed monomers known as “pepsin-like aspartic proteases” [11].

The structures of several pepsin-like aspartic proteases are shown in Figure 1-3. They are called bilobed because the active form is one molecule with two lobes resembling the two monomers found in retroviral proteases. Comparing their structural elements (Figure 1-3) to those of HIVPR (Figure 1-1), the biggest difference is the asymmetry of two lobes in pepsin-like aspartic proteases. More specifically, the left lobe preserves more structural motifs common to retroviral proteases, like the flap, helix, fulcrum and cantilever, while the right lobe only maintains the helix at a similar location, although the right lobe helix is abbreviated. The three β-hairpins (flap, fulcrum and cantilever) are not conserved on the right lobe due to the insertion of many variable regions.

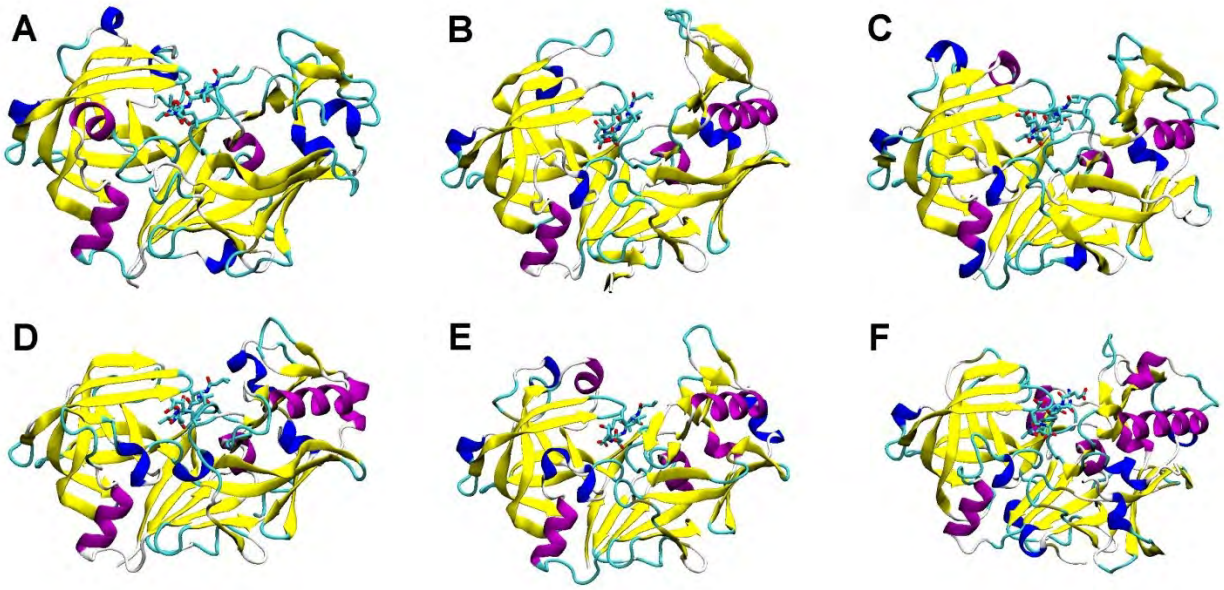

Figure 1-3 Holo crystal structures of proteases from A) Rhizomucor miehei (PDB ID 2RMP), B) Candida albicans (PDB ID 2QZX), C) Hypocrea jecorina (PDB ID 2EMY), D) Irpex lacteus (PDB ID 1WKR), E) Homo sapiens (pepsin, PDB ID 1PSO), and F) Homo sapiens (BACE1, PDB ID 1FKN). The proteases are colored based on their secondary structure. The heavy atoms of inhibitors are shown in licorice representation.

Many pepsin-like aspartic proteases are also linked to human diseases. For example, renin levels are related to hypertension [12]; β-secretase 1 is essential for generating β-amyloid [13-15], the precursor of Alzheimer's disease; and plasmepsin is produced by parasites causing malaria [16, 17]. Currently, the only FDA approved drug targeting pepsin-like aspartic proteases is renin inhibitor [12, 18].

1.2 Computer-aided drug design

Computers can facilitate the process of drug design. In so-called "ligand-based drug design", the structure of the drug target/receptor is unknown. Therefore drug design is based on the knowledge of ligands that bind to the target. A pharmacophore model is derived that defines the minimum necessary characteristics a ligand needs to possess in order to bind the target, and then this model could be used as a filter of tentative drugs. Also a quantitative structure-activity relationship (QSAR) can be derived, which represents the correlation between calculated properties of molecules and their experimentally measured biological activity [19]. The insights from pharmacophore model and QSAR can effectively inform drug design efforts. In so-called "structure-based drug design", the structure information of the drug target/receptor is known (usually from crystallography or NMR). In this case the drug design could be based on the knowledge of the target. It is worth noting that because of the fast development of experimental techniques, structure-based drug design has become more and more popular. Even if the target structure is unknown, a homology model could usually be built if the target is homologous to some known structures. Structure-based drug design usually involves docking proposed ligands into the binding pocket on the target, and then ranking the binding of different ligands based on a scoring function. A pharmacophore model can also be developed based on the binding pocket to

facilitate the docking. Methods like molecular dynamics simulation can be utilized to account for the flexibility of the ligand and target [20].

Computers can greatly facilitate the process of drug discovery in many other ways. For example, computers can also be used to design filters to eliminate compounds with undesirable ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties and select the most promising candidates [21, 22].

The work presented here has focused on structure-based drug design of aspartic protease inhibitors, with an emphasis on using molecular dynamics (MD) simulations to study the protease active site gating mechanism and the process of ligand binding. Below, the underlying theory and known limitations of molecular dynamics simulations are introduced, along with some previous applications of this method to study aspartic proteases.

1.2.1 Theory of MD simulations

Conventional MD simulations are based on molecular mechanics (MM), which is using classical mechanics to model molecular systems. Below, a more detailed introduction is given in terms of how the molecular system of interest is described and how the system evolves based on classical mechanics.

1.2.1.1 Representation of the system

Quantum effects are only implicitly considered in conventional MM-based MD simulations and the smallest unit modeled is the atom. Generally, spheres with mass, charge, and radius are used to represent atoms. These spheres are linked to represent molecules from solvent and small ligands to proteins and membranes. While the starting positions of solute atoms usually come from crystallography or NMR, the solvent molecules are usually added by the computer and equilibrated to the desired condition (pressure, temperature), as discussed below.

1.2.1.2 Force field

After defining and setting up the system, the next step is to let the system evolve under the laws of classical mechanics. This is achieved by defining a so-called “force field” that calculates the potential function of the system. In the AMBER simulation package, which is used in this study [23], the force field is composed of bonded energies (bond, angle, and dihedral torsion energies) and non-bonded energies (vdW and electrostatic energies, Equation 1-1). These energy terms act directly on the system and determine the force on each atom, which in turn adjusts each atom’s velocity. The coordinates of all atoms are updated simultaneously by integrating the velocity, and then the force field is used again to determine the next position of each atom.

$$V(r^N) = \sum_{bonds} k_b(l - l_0)^2 + \sum_{angles} k_a(\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2}V_n[1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1}\sum_{i=j+1}^{N}\left\{\epsilon_{i,j}\left[\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{0ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}\right\}$$

Equation 1-1 AMBER force field function [24]. First term is bond energy. Second term is angle energy. Third term is dihedral torsion energy. Fourth term is non-bonded energy including vdW and electrostatic energies.

An assumption inherent to this integration is that the forces and velocities will not change significantly from one integration step to the next. Therefore, the time step taken to evolve the system must be very small. Indeed, conventional MD simulations usually use time step around one femtosecond (fs), which is on the time scale of bond vibration.

1.2.1.3 Minimization, Equilibration, and Temperature/Pressure Control

Even with the physics defined, it remains a challenge to start a dynamic simulation from a static starting structure (usually from crystallography). Careful minimization and equilibration are needed to optimize the static starting structure to which solute hydrogen atoms and solvent may have been added, and heat up the system to the production temperature. Usually positional restraints maintain part of the system to the starting structure while the remainder is relaxed.

Temperature control is needed throughout MD simulations in order to account for the shift in system energy due to cutoffs and cumulative errors. This can be achieved by coupling the system to an external thermostat and adjusting velocities accordingly [25]. Pressure can be regulated through adjusting the system volume.

1.2.2 Limitations of MD simulations and possible solutions

There are two major limitations of MD simulations. First, the traditional molecular mechanics theory omits quantum mechanics (QM) effects such as charge polarization and bond forming/breaking, which means many interesting events such as protonation and enzyme catalysis are hard to model using MD. However, QM is not viable either since it is too expensive to simulate biological systems extensively. Second, the time scale routinely achievable by MM (~ microsecond), while much greater than QM, is still shorter than many interesting biological events (milliseconds to hours).

In order to account for the lack of QM in MD simulations, there are mainly three types of approaches. First, one can combine MM and QM. One can apply different theories to different parts of the system in the so-called QM/MM approach, in which usually the active site is described by QM [26]. Or, one can use different theories to simulate different stages of a process, though typically the QM is still only performed on a small subsystem, usually just the active site. Second, one can improve MM methods to mimic the QM effects (polarizable water model [27], etc.). Third, one can use a simple, often semi-empirical QM method to still allow some quantum effects at a speed such that QM simulations are possible. Of the three, the third approach has seen the least progress.

In order to overcome the time scale issue, there are also mainly three approaches. Firstly, approximation could be introduced into the simulation to speed up the computation, at the expense of accuracy. The approximation could be simplifying calculation of certain energy term is done. For example, SHAKE algorithm [28] reduces the calculation of bond vibration involving hydrogen atoms, distance cutoff reduces calculation of non-bonded interactions for distal atom pairs, and Particle-Mesh Ewald (PME) [29-32] method simplifies the electrostatic energy calculation. The approximation could also be reducing the number of atoms included in the system (and thus need to be simulated). For example, periodic boundary conditions are used to define boundaries of modeled space so that less solvent molecules need to be modeled, in implicit solvent model the solvent is approximated by a dielectric continuum, and in coarse-grained model a sphere is used to represent more than one atom. Secondly, since rare, high energy conformations require the most simulation time to adequately sample, speedup can be achieved by modifying the Hamiltonian to enhance sampling of rare states. Such methods depend upon the recovery of the unbiased free energy afterwards. So-called advanced sampling methods include those methods that reconstruct free energy along a few predefined collective variables (umbrella sampling [33], thermodynamic integration [34], steered MD [35], etc.), those methods that explore the transition mechanism (nudge elastic band [36], etc.), those methods that explore the potential energy surface and localize transition states (multi-time scale accelerated molecular dynamics [37], etc.), and those that explore phase space simultaneously at different conditions, conventionally different temperatures (replica exchange molecular dynamics [38], etc.). Thirdly, the computation speed has also been boosted greatly due to hardware and software improvement in recent years. On the one hand specialized super-computers have made millisecond long simulations possible [39]. On the other hand GPU computing has made super-computing available to more researchers who do not have access to super computers [40]. The increase in the time scale reachable by conventional MD simulations is significant, compared with the only tens-of-picosecond simulations possible in the 1980's [41].

1.2.3 Application of MD simulations to the study of aspartic proteases

The largest deviation among available crystal structures of HIVPR is around the β-hairpin flap region (Figure 1-1). Our group previously demonstrated how the two flaps gate the access of ligand to the HIVPR active site, through MD simulations sampling reversible and reproducible transitions among different flap conformations both observed in crystal structures and novel [42-44]. These simulations linked experimental observations and provided new insight into HIVPR inhibition.

MD simulations have been applied to HIVPR to understand more than ligand gating. For example, with MD simulations researchers have decomposed the impact of drug resistance mutations on inhibitor-protease interactions through binding affinity calculations [45], evaluated the influence of physiological factors such as salt concentration and molecular crowding on HIVPR dynamics [46, 47], complemented docking and QSAR in ranking inhibitors [48-50], and explored possible ligand binding pathways [51, 52].

Although HIVPR is the principal aspartic protease studied by previous MD simulations, there have also been MD studies on other aspartic proteases (page 2). Generally, the number of simulation studies of an aspartic protease is determined by the amount of available structure information, and the relevance of the aspartic protease to human diseases. Unsurprisingly, there

have been studies on HTLV [53], β-secretase [54, 55], plasmepsin [56], and renin [57]. These studies mostly focused on protease-ligand interactions.

1.3 Outline of research projects

In continuation of our lab's previous work on HIVPR, my research has focused on using MD simulations to study the dynamics of HIVPR and to improve existing HIVPR inhibitors. Moreover, we extended the study to other aspartic proteases, to transfer our knowledge of HIVPR dynamics to the design of inhibitors targeting other disease-causing aspartic proteases.

1.3.1 Improving the description of salt bridge strength and geometry in an AMBER implicit solvent model

The Generalized Born (GB) implicit solvent model [58-60] is used widely in MD simulation for several reasons. The elimination of explicit presentation of solvent molecules reduces the complexity of computation, and the lower viscosity enables faster conformational sampling compared to explicit solvent simulations. However, there is a speed-accuracy tradeoff. Previously, great efforts have been made in order to improve the precision and accuracy of GB models. The modification of intrinsic radii, which define the dielectric boundary between the solute and the solvent, has been shown to be effective [61]. Here, we attempted similar modifications to a widely used GB model in the AMBER simulation package [62], GB-OBC [63], to improve the description of salt bridge strength and geometry, which were found to be problematic. Potential of mean force and cluster analysis for small peptide replica exchange molecular dynamics simulations suggested that GB-OBC with the new radii set, combined with the ff99SB force field [64], corrected salt bridge strength and achieved geometries significantly more similar to those of TIP3P explicit solvent simulations [65]. The results were validated in 60ns GB simulation of HIVPR. Moreover, comparison among GB, Poisson-Boltzmann (PB [66]), and TIP3P suggests that accuracy of PB calculations can also benefit from radii modification.

1.3.2 Spin-labeled HIV-1 subtype protease simulations explain EPR spectrum shift and AIDS drug resistance mechanism

Because of the extremely heterogeneous HIV genome, HIV strains are classified into different subtypes. All current AIDS drugs are tested and developed towards HIV subtype B, which is primarily found in developed countries. HIV-1 protease (HIVPR) has been an effective target in AIDS treatment because of its essential role in HIV maturation. The two β-hairpin flaps of HIVPR govern the ligand access to the active site. Previous EPR experiments suggested that HIV-1 subtype C protease has a greater fractional occupancy of the open flap conformation than subtype B [67], which may explain its less favorable binding to current drugs. However, the nine polymorphisms are scattered throughout the protease structure, so it is not straightforward to conjecture their mechanistic role in differentiating flap dynamics.

To explain the EPR spectrum in atomic detail and aid next generation inhibitor design for subtype C, we performed MD simulations of spin-labeled HIV-1 subtype B and C proteases. We utilized batch simulation (each system has 50 runs totaling 1 microsecond) with an implicit solvent model to speed up the convergence of conformational sampling. We compared our simulations to the EPR spectrum and studied the possible influence from spin label dynamics.

Moreover, we used 2D measurement to correlate population with structure, and through energy decomposition we found three polymorphisms (M36I, S37A, and H69K) having strongest contribution in differentiating two subtypes, which is validated by additional simulations of mutants. Finally, we provided structural explanation of the subtype difference, which is closely related to HIVPR drug resistance mechanism.

1.3.3 HIV-1 protease multi-drug resistance studied by binding affinity calculations and communication network analysis

Because of the high mutation rate of the HIV virus, AIDS patients are usually treated with more than one antiretroviral drug at a time, to suppress drug resistant mutants as they evolve. This treatment is often called highly active antiretroviral therapy (HAART), or "cocktail" therapy. However, under HAART treatment the virus is still able to develop multi-drug resistance (MDR) that reduces the efficacy of all drugs used in the treatment [68]. When this happens, the treatment options become rather limited and the prognosis may deteriorate. Therefore, the design of the next generation HIVPR inhibitors should take MDR into account.

We aimed to improve HIVPR inhibitor design by studying the change in protease-inhibitor binding in response to drug resistance mutations. Thermodynamics integration (TI) is a method to compare the free energy of two states, and a thermodynamic cycle can be employed where TI calculations achieve accurate free energy calculations by following alchemical pathways. We performed TI calculation on two experimental inhibitors to study their loss of binding affinity due to active site MDR mutations.

Apart from studying MDR mutations that are near the active site, we were more interested in MDR mutations distal from the binding pocket and how these mutations can influence protease-inhibitor binding. We hypothesized that the influence is achieved through changing the coupling of different domains of the protease. We compared first- and second-generation HIVPR inhibitors, which have different levels of susceptibility to drug resistance mutations. We performed molecular mechanics Poisson-Boltzmann surface area (MMPBSA) binding affinity calculations, and also communication network analysis that accounts for coupling of distal residues. Our results are consistent with the distal-coupling hypothesis about distal MDR mutations.

1.3.4 Comparative study of aspartic protease family and modeling the active site gating mechanism in non-HIV aspartic proteases

HIVPR belongs to the aspartic protease family, which is a large protein family found in viruses, prokaryotes, and eukaryotes. Results from recent studies have added complexity to the family. To give an updated overview of the very diverse and complex aspartic protease family, we constructed an evolutionary profile using existing sequence and structural information. Based on the evolutionary profile and recent findings, we hypothesized the evolutionary path of the aspartic protease family. More importantly, through the sequence and structure comparison, we identified both conserved and distinguishing sequences belonging to individual branches, which may be used to design inhibitors targeting non-HIV aspartic proteases.

Apart from sequence and structure comparison, we were also interested in applying MD simulations to study the active site gating mechanism in non-HIV aspartic proteases, to help

adapt our knowledge of and experience with HIVPR to the design of inhibitors targeting other human diseases. Because of incomplete crystal structures and sampling issue, we combined homology modeling with implicit and explicit solvent simulations to model apo retroviral protease dynamics of Human T-lymphotropic virus (HTLV), Simian immunodeficiency virus (SIV), and Murine leukemia virus (MLV). The reproducible and reversible sampling of different flap conformations in apo–protease simulations indicates a shared active site gating mechanism. We also showed large conformational rearrangement upon drug binding in some of these proteases that would not be predicted from knowledge of HIVPR dynamics. For pepsin-like aspartic proteases, we proposed an active-site-gating mechanism based on sequence conservation, and we chose Alzheimer's-implicated β-secretase 1 (BACE) as a representative to test our hypothesis. The apo and holo BACE simulations provided, for the first time, insights into its ligand binding process.

Chapter 2 Improving the description of salt bridge strength and geometry in an AMBER implicit solvent model

2.1 Introduction

Since Generalized Born (GB) implicit solvent model [58-60] was first introduced in 1980's, it has provided molecular dynamics (MD) simulations [69, 70] an alternative way to represent the electrostatic effects from the bulk solvent. In contrast to explicit solvent simulations, in which the solvent molecules are modeled explicitly [71], the GB model uses the Born equation (Equation 2-1) to approximate the solvent electrostatic effect. This approximation has several advantages: 1) most of the time we focus on the solute's dynamics only so the explicit representation of the solvent is not essential, 2) exclusion of solvent molecules largely reduces the complexity of the computation and thus achieves significant speed up, and 3) the lack of viscosity during simulations results in much faster conformational sampling compared to explicit solvent simulations. GB methods have been reviewed in detail previously [72-75].

Apart from its advantages, GB model also suffers from the speed-accuracy tradeoff. Because of the low viscosity in implicit solvent simulations, any force field defects would also show up much sooner and be amplified. Therefore, errors from either the force field or the solvent model can lead to serious problems in implicit solvent simulations. Over the past two decades, several generations of force fields and GB solvent models have been developed [76-78], and there have been studies comparing the accuracy and efficiency of different force fields or solvent models [79-82]. Unfortunately, a "gold standard" or a consensus force field/solvent model combination that provides a satisfying balance between accuracy and speed has not been reached, so simulation results are likely to continue depending on which force field or solvent model is chosen in the near future. Under the circumstances it would be necessary to perform specialized optimization for force field/solvent model combination, at times with cancellation of errors from the solvent and solute models. Here we present our improvement in MD simulations with ff99SB force field [64] and GB-OBC [63] implicit solvent model in the AMBER simulation package [62]. The ff99SB represents a modified version of AMBER ff99 force field [83], which improved the backbone dihedral parameters in ff99 through re-parameterization of AMBER ff94 force field [84]. The GB-OBC represents an AMBER GB implicit solvent model that was shown to outperform GB-HCT [85] and GB-NECK [86].

Although ff99SB has been regarded as one of the best performing force fields and has been applied to many MD simulations [87, 88], it was recently shown to marginally destabilize helical secondary structures in some systems [89]. In contrast, GB-OBC solvent model, which has been used frequently with ff99SB, was shown to slightly over-stabilize helical content, and to produce erroneous salt bridge strength and geometry [90-92]. Although coupling the forced field and the solvent model optimizations has been explored previously by CHARMM simulation package developers [75], changing backbone parameter is beyond the scope of this study. Instead, we constrained the backbone conformation while improving the agreement between GB-OBC and TIP3P explicit solvent model.

We chose to improve GB-OBC model by correcting its intrinsic radii due to the simplicity of implementation. In GB models, the solvent electrostatic effect is represented by screening/damping the electrostatics of the solute. How much certain atom would be screened

depends generally on how solvent-accessible it is. The solvent-exposed atoms would be screened more, while the buried atoms would be screened less because they are surrounded by low dielectric solute atoms, compared to the solvent atoms with high dielectric constant. This property is also called one atom descreening another. Mathematical expression of the Born equation can be found in Equation 2-1 and Equation 2-2. In the Born equation, the effective radii are used to represent the different degrees of solvent-accessibility of different atoms. The atom with bigger effective radius is less solvent-accessible, and thus would be screened less. The intrinsic radii, which define the size of each solute atom and also the dielectric boundary between the solute and the solvent, are used to calculate the effective radii, and thus serve as the foundation of GB models.

$$\Delta G_{GB} = -\frac{1}{2}\sum_{ij}\frac{q_i q_j}{f^{GB}(r_{ij}, R_i, R_j)}\left(1 - \frac{e^{-\kappa f_{ij}^{GB}}}{\epsilon_w}\right)$$

Equation 2-1 GB equation, where r_{ij} is the distance between atoms i and j, Ri and Rj are the effective Born radii of i and j, respectively, and f^{GB} is a smooth function.

$$f^{GB} = \left[r_{ij}^2 + R_i R_j \exp\left(-r_{ij}^2/4R_i R_j\right)\right]^{1/2}$$

Equation 2-2 A common choice of f^{GB}, where r_{ij} is the distance between atoms i and j. Ri and Rj are the effective Born radii of i and j, respectively.

Although intrinsic radii are essential for GB models and they influence many important properties of MD simulations such as hydrogen bonds and salt bridges, the definition of GB intrinsic radii has been empirical because the atomic spheres only approximate the molecular surface used in more accurate models. Moreover, defining intrinsic radii for hydrogen atoms is even harder since their electron density is more sensitive to the varying electronegativity of neighboring atoms [61]. As a result, various groups have attempted to optimize intrinsic radii parameters previously and it was shown to improve GB simulations with CHARMM force fields [93], OPLS force fields [94], and AMBER force fields. Reducing the intrinsic radii of hydrogen atoms linked to charged nitrogen was attempted by Geney et al. on MD simulations using ff99SB/GB-HCT, which improved the agreement to explicit solvent simulations and also to the experimental melting curve of a mini protein [61]. GB-OBC is a modified version of GB-HCT that improves the description of interior residues. Due to the close relationship between GB-HCT and GB-OBC implicit solvent models, we evaluated similar correction in our GB simulations with ff99SB/GB-OBC, although the radii sets we used are not identical.

We began the intrinsic radii correction on a small salt-bridge model peptide (Ace-Arg-Ala-Ala-Glu-NH2). It was been shown by Okur et al. previously [91] that, in replica exchange molecular dynamics (REMD) simulations with GB implicit solvent using ff99SB/GB-OBC combination, this peptide has too strong salt bridge strength compared to explicit solvent (EXP) REMD simulations. Moreover, the most populated geometry of the salt bridge is also different from that found in explicit solvent REMD (Figure 2-1). To test whether these differences are caused by the intrinsic radii setting, we carried out several GB REMD simulations using the same parameters as those used by Okur et al. except for the fact that each simulation had a unique setting of H^n radii (H^n denotes Hε and Hη on Arginine residue side chain). We then compared these GB simulations to EXP REMD simulations. Our results suggest that reducing H^n

radii could improve both the strength and also the geometry description of Arg-Glu salt bridges in GB simulations.

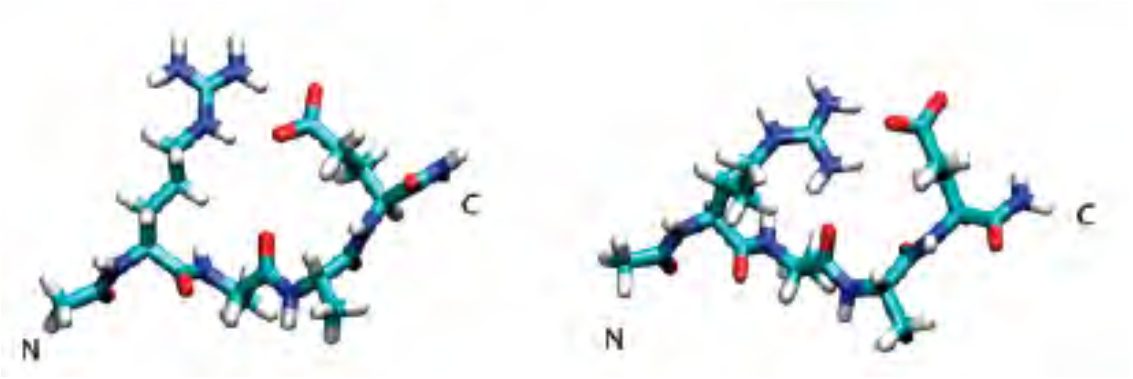

Figure 2-1 Representative structures of the most populated salt bridge geometries in standard REMD simulations using explicit solvent (TIP3P, left) and implicit solvent (GB-OBC, right) models.

Besides experimental observables and explicit solvent simulations, Poisson-Boltzmann (PB) implicit solvent method [66] has also been used as a standard for GB method optimization, because it models the electrostatic effects of solvent by solving the Poisson equation numerically. GB methods, on the other hand, usually use analytical or even pairwise approximations in modeling [63, 95], so they gain speed-up but lose accuracy compared to PB methods. Nevertheless, both GB and PB methods rely on the choice of intrinsic radii. Therefore, we were interested in comparing the dependence of both GB and PB method on intrinsic radii selection. However, since PB REMD would be too time-consuming to be feasible at present, we simplified the GB-PB comparison by only focusing on energetically preferred salt bridge geometries in each method, given a set of structures. We compared GB and PB energies for structures sampled in model peptide explicit solvent REMD simulations using either original or modified H^n radii. Results predicted that PB method would also produce inaccurate salt bridge strength and geometry given the original H^n radii. Therefore, caution needs to be taken when using PB as a standard for evaluating GB quality, since the errors originating from intrinsic radii selection are largely ignored.

After validating the effectiveness of radii modification in small model peptide, we also evaluated the same modification in macromolecule HIV-1 protease simulations. HIV-1 protease (HIVPR) has been an effective target in AIDS treatment because of its essential role in HIV maturation. Lengthy and stable HIVPR GB simulations are desirable because at present explicit solvent simulations of HIVPR can only reach sub-microsecond time scale routinely, while the interesting HIVPR dynamics to researchers, such as flap opening and ligand binding, all happen on or above microsecond time scale. Electron paramagnetic resonance (EPR) experiments have been done previously to probe dynamics of HIVPR flaps, in which spin labels were attached to flaps of HIVPR and echo intensity between the two spin labels were measured [96]. Our lab performed spin-labeled explicit solvent MD simulations of HIVPR and compared results to experimental data [96, 97]. These comparisons would be more convincing if quantitative GB simulations were possible, because at present only GB simulations could provide enough statistics on flap-conformation transitions to study the equilibrium among different flap conformations.

However, although our lab previously performed GB simulations of HIVPR and achieved good agreement with explicit solvent simulations and crystal structures [42, 43], there are two fundamental problems that hinder lengthy and converged GB simulations. First, HIVPR termini form an interleaved β-sheet formed by N- and C- termini from both monomers, and these termini, especially the N- termini, tend to form helical content during the simulation (Figure 2-2). Second, flap opening events are sampled too often in GB simulations. Although these open flap conformations are qualitatively comparable to those sampled in explicit solvent simulations, they sometimes lead to distorted flap conformations, which may result from forming helical content when two flaps are not in contact (Figure 2-2). The GB-OBC implicit solvent model we used was proven to have bias towards helical structures [90]. This explains the first problem. Since we did not aim to fix backbone preference, we applied termini restraints to maintain termini secondary structure (Figure 2-3, see the methods section for details). The reason for the second question is unclear. There are two salt bridges involving Arg on each flap elbow region (Figure 2-4) that have been proposed to have allosteric coupling to flap opening [42]. Based on our salt-bridge comparison in model peptide simulations, we hypothesized that the strong salt-bridge in HIVPR GB simulations may be responsible for the increased frequency of flap opening. To test this hypothesis and determine whether more accurate dynamics of macromolecules can be obtained with the simple GB model, we carried out HIVPR simulations with GB-OBC solvent model and optimized H^n radii.

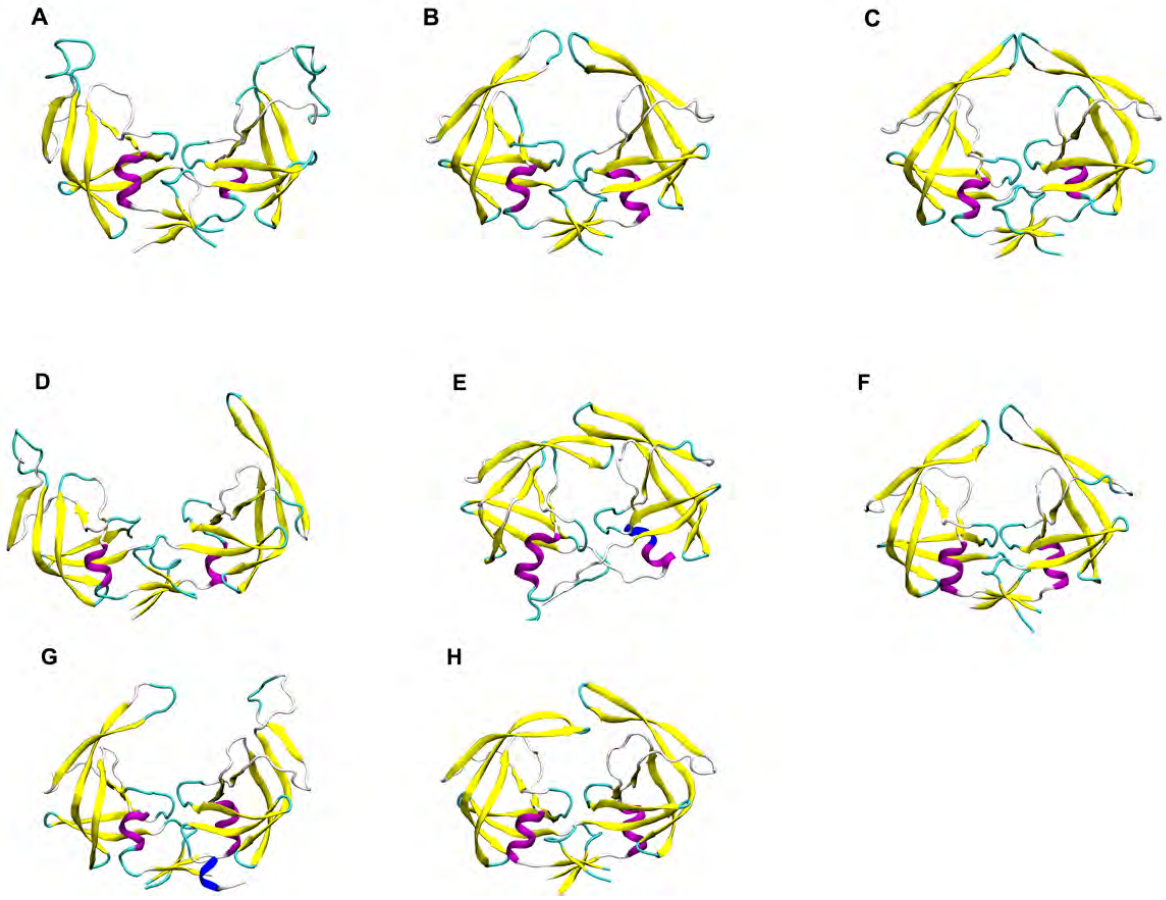

Figure 2-2 The last frames of eight HIVPR D25N mutant GB simulations. They all used the same parameter set as 1.3Å H^n (default value) radii GB simulation discussed in result section, except that they were unrestrained. Each simulation was 20 ns simulation length and had different velocity seeds to unsynchronize. Four simulations (B, C, F, and H) sampled reasonable structures, while the other four sampled deformation either in the flap region (A, D, and G) or the termini region (E and G).

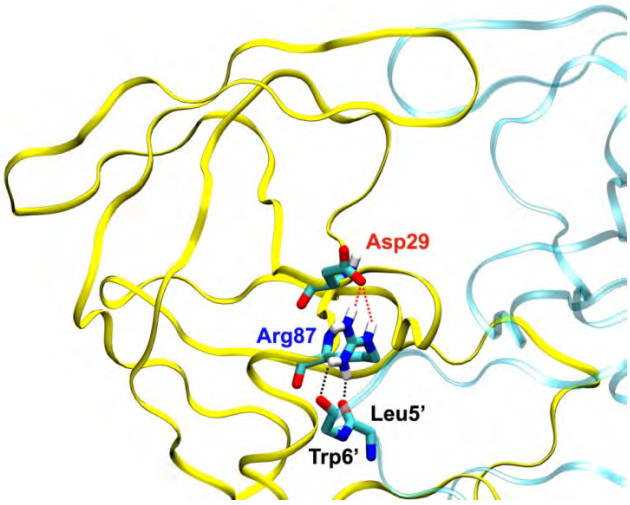

Figure 2-3 Hydrogen bonds and salt bridges formed by Arg87 at the dimer interface. The Arg87 residue forms two intra-monomer salt bridges with Asp29 (red dotted lines), and two inter-monomer hydrogen bonds with backbone oxygen atoms from Leu5 and Trp6 on the other monomer (black dotted lines). Harmonic distance restraints (see the methods section for details) on these four interactions was shown to stabilize the dimer interface as well as the termini secondary structure.

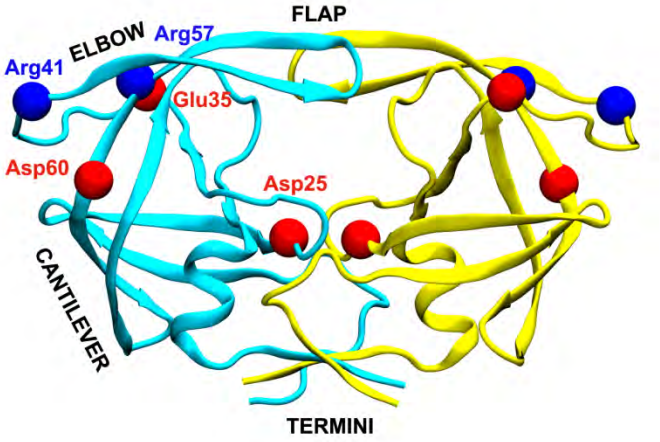

Figure 2-4 Salt bridges on the elbow region of HIV-1 protease (HIVPR). The elbow, cantilever, and termini regions are labeled with black text. Two monomers of HIVPR are colored cyan and yellow. Catalytic aspartate residues and residues involved in salt bridges on the elbow region are represented by balls and colored by charge. The figure was generated with crystal structure 1HVR.

2.2 Methods

2.2.1 Small peptide REMD preparation

A small peptide Arg-Ala-Ala-Glu capable of forming a salt bridge was simulated, with its N terminal acetylated and C terminal amidated. It was shown previously that GB REMD

simulation with this peptide gave inaccurate salt bridge strength and geometry compared to EXP REMD [91].

2.2.2 Small peptide EXP REMD

Simulation trajectories used here are from Okur et al. [91] and also an additional run they did subsequently. The simulation parameters and procedure they used were as follow. The AMBER simulation package [62] version 9 and ff99SB force field [64] were used. The peptide was solvated in a 16 Å truncated-octahedral TIP3P [65] water box, containing 2286 water molecules. Minimization and equilibration were done at 300 K for 65 ps, with reducing harmonic positional restraints on the solute atoms. They used 46 replicas spanning temperatures from 296 K to 584 K, which gave a uniform exchange acceptance ratio of about 25%. Time step was 2 fs, and the exchange between neighboring temperature-replicas was attempted every 1 ps. The REMD was run for 30 ns, which means there were 30,000 exchange attempts for each replica. All bonds involving hydrogen atoms were constrained using SHAKE [28] with geometry tolerance of 10^{-7}. Particle-Mesh Ewald (PME) [29-32] was used to calculate long range electrostatic interactions, and a 7 Å cutoff was used for vdW interactions. REMD was run in NVT ensemble with Berendsen temperature control [25]. Backbone of the peptide was restrained to the most populated structure of unrestrained TIP3P REMD using a harmonic potential (1 kcal/mol·Å^2). They also performed another REMD run with different velocity seeds to check for convergence. The first 5 ns trajectory structures from simulations were discarded to avoid bias from the initial structure, which gave a 50 ns combined trajectory. The last 36 ns of the combined trajectory at 300 K was used for PMF plot and cluster analysis. All 50 ns combined trajectory at 300 K was used for the lowest energy plot (see below).

2.2.3 Small peptide GB REMD

We performed GB REMD simulations with the sander module in AMBER simulation package version 10 [24]. The ff99SB force field [64] and GB-OBC implicit solvent model [63] were used (igb = 5 in AMBER). The mbondi2 intrinsic radii set [63, 98, 99] was used with the H^n radii modifications (H^n denotes Hε and Hη on Arginine residue side chain). In total 3 sets of GB REMD simulations were carried out, which only differ in their H^n intrinsic radii: 1.3 Å (standard mbondi2), 1.2 Å and 1.1 Å were used, respectively. For each REMD simulation, 6 replicas were chosen spanning temperatures from 300 K to 636 K, and no cutoff on non-bonded interactions was applied. The simulations were about 50 ns each, and the last 36 ns of each 300 K temperature trajectory was used for PMF plot and cluster analysis.

2.2.4 HIVPR MD preparation

The starting structure of HIVPR simulations were generated by deleting the inhibitor from the holo crystal structure 1HVR [100] and introducing active site mutation D25N using tleap module in AMBER package. The D25N inactivating mutation is widely used in crystallography and NMR studies because the mutant retains native-like fold and binds to its natural substrate without initiating catalytic reaction, so that its interaction with the natural substrate could be analyzed [101, 102]. The tleap module was then used to add hydrogen atoms, evaluate the crystal structure in terms of bond, angle, and clashes, and add water molecules for explicit solvent simulations. A truncated-octahedron TIP3P water box was added with 8 Å minimum clearance from the boundaries, resulting in adding 7,219 water molecules. For both

GB and explicit solvent MD simulations, AMBER simulation package version 10 and ff99SB force field were used. Following energy minimization, the system was heated gradually from 100 K to the desired temperature, with harmonic restraints added first on heavy atoms and then on backbone atoms only. The restraint force constant decreased from 100 to 0.1 kcal/mol·Å^2 during the equilibration. The equilibration was 750 ps for explicit solvent and 125 ps for GB simulations.

2.2.5 HIVPR EXP MD

We performed 400 ns unrestrained MD simulations of HIVPR inTIP3P explicit solvent. Time step was 2 fs. All bonds involving hydrogen atoms were constrained using SHAKE with geometry tolerance of 10^{-5}. Particle-Mesh Ewald (PME) was used for long range electrostatic interactions and 8 Å cutoff was applied to vdW interactions. Berendsen temperature and pressure control was used to maintain the system at 325 K and 1 atm.

2.2.6 HIVPR GB MD

We performed two 60 ns HIVPR simulations with GB-OBC implicit solvent model. The mbondi2 radii set was used besides the H^n modifications. The two runs had H^n set to 1.3 Å (standard mbondi2) and 1.1 Å, respectively. Time step was 1 fs. All bonds involving hydrogen atoms were constrained using SHAKE with geometry tolerance of 10^{-6}. No cutoff for long range electrostatics or vdW interactions was applied. But there was a 25 Å distance cutoff on pair interactions involved in effective radii calculation. Forces related to effective radii calculation, along with pair interactions whose distance was longer than 15 Å, were evaluated every 4 steps. Langevin temperature control (collision frequency 1 ps^{-1}) was used to maintain the system at 325 K. Harmonic restraints were applied to 4 hydrogen bonds (atom pairs Asp29_Cγ to Arg87_Hη21, Asp29_Cγ to Arg87_Hε, Arg87_Hη11 to Leu5'_O, and Arg87_Hη12 to Trp6'_O, where Leu5' denotes Leu5 on the other monomer) at the dimer interface near the termini, to prevent formation of helical structures in the termini region (Figure 2-3). The Cγ atom on Asp29 was chosen to account for carboxyl rotamers. For each atom pair, a distance cutoff was chosen based on its average distance measured from explicit solvent simulations, which were 4 Å for pairs involving Asp Cγ and 3 Å for the rest. An atom pair was only restrained when the distance went beyond the cutoff, and the restraint force constant was 10 kcal/mol·Å^2. The velocity seeds were changed during the simulations to unsynchronize Langevin dynamics [103].

2.2.7 Distance and root mean square deviation (RMSD) measurements

Ptraj module in AMBER package was used for distance and RMSD measurements. HIVPR flap RMSD was calculated using the Cα atoms of residues 46-55 in both monomers (46-55 and 46'-55') as mask. Crystal structures with the closed (PDB ID 1HVR [100]) or semiopen (1HHP [104]) flap conformation were used as references.

2.2.8 Potential of mean force (PMF)

PMF as a function of atom pair distance was used to represent the salt bridge strength of residue pair Arg-Glu and Arg-Asp. Cγ atom from Asp, Cδ from Glu and Cζ from Arg were chosen for atom pair distance measurement, to account for rotamers. The 300 K temperature trajectory was used to histogram pair distances. Then the histogram was converted to PMF using

Equation 2-3. Energies zeroed at most populated bin. The first and second half of each data set were compared to generate error bars for PMF plots.

$$\Delta G = -RTln\,(P/P_{ref})$$

Equation 2-3 Conversion from equilibrium population to free energy difference, where P is the population of a certain conformation (bin), and P_{ref} is population of the reference conformation (bin). P_{ref} usually is the most populated conformation (bin).

2.2.9 Cluster analysis

Salt bridge geometry was investigated by clustering the small peptide 300 K temperature REMD trajectories based on a RMSD similarity cutoff. The clustering was performed using Moil-View [105] using the following heavy atoms as the similarity criterion: Arg_Cδ, Arg_Nε, Arg_Nη1, Arg_Nη2, Arg_Cζ, Glu_Cδ, Glu_Oε1, Glu_Oε2. Every 2nd structure from TIP3P and GB trajectories was combined, and then the resulting combined trajectory was subjected to clustering using bottom-up approach and 1.3 Å similarity cutoff. Each structure was initially assigned into a unique cluster, and then clusters with the smallest RMSD were merged, until the smallest cluster RMSD was greater than the similarity cutoff. After clustering, the combined trajectory was again sorted into different simulations (explicit solvent simulation, or GB simulations with certain H^n radii). The percentage of individual simulation's structures within each cluster was plotted, and by comparing the distribution of each simulation's structures in different clusters, we determined the geometry similarity of conformations sampled by different simulations.

2.2.10 Lowest energy profile

Because simulation with PB is too slow to be feasible at present, we used lowest energy profile to focus the GB-PB comparison on those energetically favored conformations, given a reservoir of structures [64]. The combined 300 K temperature trajectory of small model peptide explicit solvent simulation (50 ns, 50 000 structures) was used as the structure reservoir. Four sets of energy calculations were carried out: two GB calculations which only differ in H^n intrinsic radii (1.1 or 1.3 Å), and two PB calculations which only differ in H^n intrinsic radii (1.1 or 1.3 Å). Non-H^n atoms in both GB and PB calculations used mbondi2 intrinsic radii set. The GB calculations were performed using AMBER 10 package, GB-OBC solvent model, and no non-bonded cutoff was applied. The calculations gave the potential energy which includes bond/angle/torsion energies, electrostatic energy, vdW energy, and polar solvation (GB) energy. The PB calculations were performed using DelPhi [106] package with 0.25 Å grid spacing. The internal and external dielectric constants were set to 1 and 78.5, respectively, to be consistent with AMBER calculations. The polar electrostatic energy calculated by Delphi was added to the non-solvation energies calculated by AMBER to get potential energy of PB calculations. After potential energy calculation, structures were histogramed according to the salt bridge distance (bin size 1 Å). The average potential energy of the 20 lowest-energy structures in each bin was plotted against the salt bridge distance to get "lowest energy profile" [64]. Each curve was zeroed at its minimum for easier comparison. The error bar was generated by comparing the first and second half of data.

2.3 Results

2.3.1 Test in a small model peptide system

We performed REMD simulations of salt-bridge model peptide (Ace-Arg-Ala-Ala-Glu-NH2) using GB-OBC implicit solvent model, and compared results to TIP3P explicit solvent REMD simulations.

Potential of mean force (PMF) profiles are plotted as a function of the salt bridge distance between Arg Cγ atom and Glu Cδ atoms (Figure 2-5). The depth of the global minimum in the PMF profile represents the strength of the salt bridge. For example, in the TIP3P explicit solvent simulation (blue curve in Figure 2-5), the global minimum is stabilized by a 2 kcal/mol energy barrier at around 5.2 Å salt bridge distance. This minimum represents the direct hydrogen bonding between the Arg and Glu residues (Figure 2-1).When this energy barrier is overcome, there is a local minimum at around 6.3 Å salt bridge distance stabilized by a 0.5 kcal/mol energy barrier. This minimum represents the water mediated hydrogen bonding between the Arg and Glu residues (Figure 2-6). For GB-OBC simulations with mbondi2 radii (H^n denotes Hε and Hη on Arginine residue side chain, H^n atoms in mbondi2 have default value 1.3 Å), however, the salt bridge has an energy barrier of about 4 kcal/mol, which is 2 kcal/mol over-stabilization compared to explicit solvent simulations. Reducing intrinsic radii of H^n reduces the energy barrier. This is expected because reducing intrinsic radii of hydrogen is equivalent to reducing its polarity, thus weakening the salt bridge. The best match to explicit solvent simulations is achieved when the H^n atoms have 1.1 radii. However, we also noticed that all the GB-OBC curves are missing the second minimum at around 6 Å salt bridge distance, and their global minima fall to the right of the one from explicit solvent simulation. Missing the water-mediated hydrogen bond in GB-OBC is understandable since that needs explicit modeling of water molecules. We compared salt bridge geometry to answer the second question.

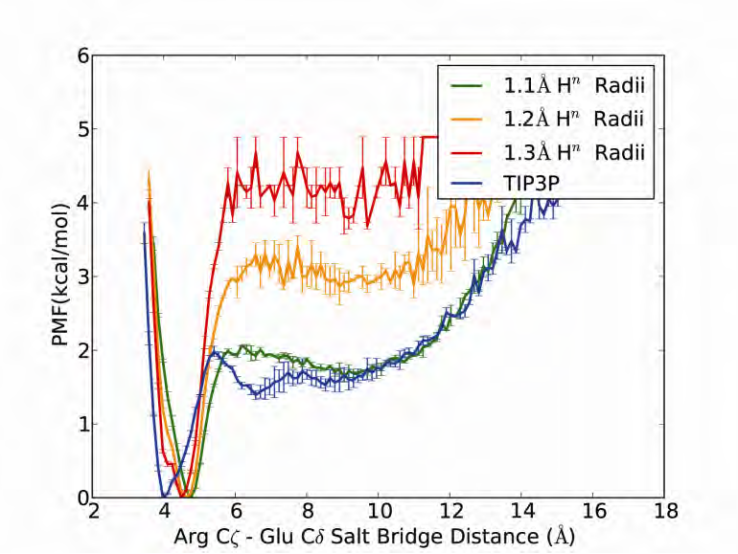

Figure 2-5 Potential of mean force (PMF) for salt bridge distance in small peptide. Curves were calculated from 300 K temperature REMD trajectory with different solvent models (GB-OBC implicit solvent with different H^n radii, and TIP3P explicit solvent).

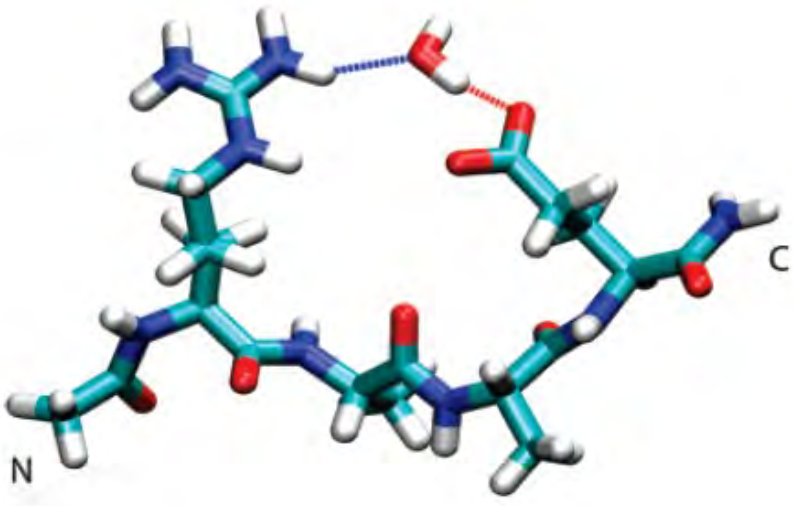

Figure 2-6 Water mediated salt bridge between Arg and Glu residues.

To compare popular structures sampled in different types of simulations, we performed cluster analysis on a trajectory combining structures from explicit solvent REMD simulations (of the small model peptide) as well as implicit solvent REMD simulations with different H^n intrinsic radii. Cluster analysis on the combined trajectory enabled us to compare these simulations under the same criterion. A structure-based similarity cutoff was used for clustering (see the methods section for details), and similar structures, from either the same or different REMD simulations, were put in the same cluster. Then we evaluated different simulations' structural preference, by comparing the distribution of simulation structures in different clusters (Figure 2-7). The representative structures of the four largest clusters are shown in Figure 2-7A. Comparing Figure 2-7 with Figure 2-1, it is clear that cluster 1 represents the salt bridge geometry most populated in explicit solvent simulations (Figure 2-1 left), while cluster 3

represents the salt bridge geometry most populated in GB-OBC simulations with original mbondi2 intrinsic radii (Figure 2-1 right), which indicates that our results are consistent with the previous study [91]. The population distribution of simulation structures in different clusters is shown in Figure 2-7 B-D. Overall, when the H^n intrinsic radii is decreased from 1.3 Å to 1.1 Å, the correlation coefficient between GB-OBC model and TIP3P explicit solvent increased from 0.36 to 0.90. The biggest improvement is the decreased population of cluster 3 structures, and the increased population of cluster 1 structures in GB-OBC model, which adapt GB-OBC structural preference to that of TIP3P water model. The small clusters are also in better agreement with TIP3P model when the H^n is 1.1 Å. The cluster 4 and 24 have worse correlation in 1.1 Å H^n radii than original 1.3 Å H^n radii, but the difference is much less significant compared to the improvement in cluster 1 and 3. Therefore, we concluded that decreasing H^n improves the agreement with TIP3P explicit solvent model. This also stresses the importance of having a multi-dimensional geometry comparison: the GB-OBC implicit solvent model and TIP3P explicit solvent model have big influence on salt bridge geometry, but they have only small difference in more simplistic 1D distance measurement.

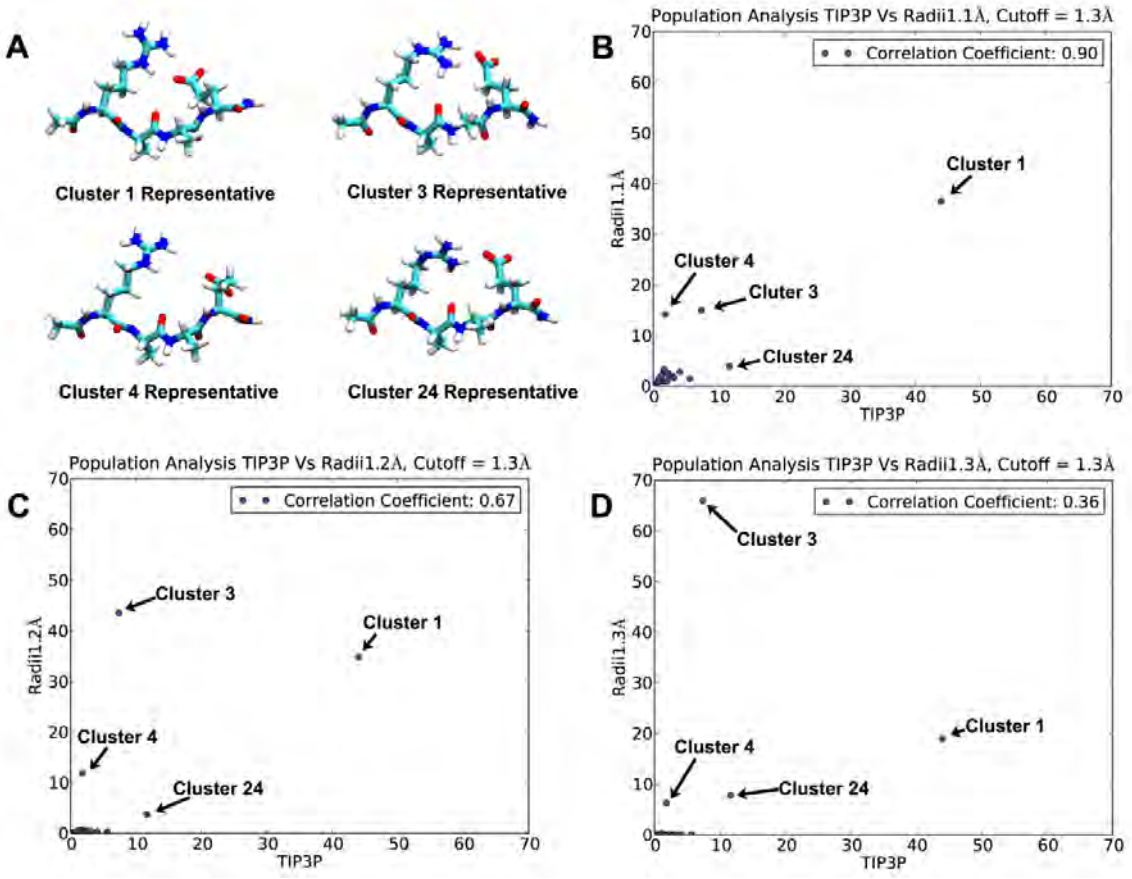

Figure 2-7 Cluster analysis of model peptide REMD simulations. Trajectories from GB and explicit solvent simulations were combined for clustering. A) The representative structures of four largest clusters. B-D) Comparing population distribution of simulation structures in different clusters. X and Y axis are the percentage of simulation structures in certain cluster. Ideally, if a GB model is

totally correlated with explicit solvent, then they should have exactly the same distribution in different clusters so that clusters would line up on the diagonal.

Poisson Boltzmann (PB) has been another way to calculate the solvent electrostatic effect, and serves as the standard for GB model optimization in many studies [99, 107].Therefore, we wanted to evaluate the dependence of PB on intrinsic radii choice. Since simulations with PB are too computationally expensive to be feasible, we compared the salt bridge strength of PB and GB methods on a structure reservoir, focusing on the energetically favored structures. The "lowest energy profile" method was proved useful previously in comparing different simulation settings [64]. The method compares the geometry preference of different simulation parameter sets by calculating the energy of a reservoir of structures using each parameter set. We plotted the lowest energy profile of salt bridge strength in Figure 2-8. Comparing it with the PMF profile in Figure 2-5, the lowest energy profile shows bigger error bars since much fewer structures were considered. But the trend is similar between the two figures: reducing H^n radii systematically decreases the salt bridge strength. If we set GB-OBC model with 1.1 Å H^n radii as the standard, since it has the closest agreement with the TIP3P water model, then the PB method also needs the intrinsic radii change to get better agreement with explicit solvent model. This result suggests caution in using PB as a standard in GB model optimization. At least using PB to calibrate salt bridge strength is insufficient, since PB is itself highly sensitive to intrinsic radii choice that defines the molecular surface. This result is consistent with a previous finding that the best GB method in reproducing PB results is not the one that best matches folding experiment [108].

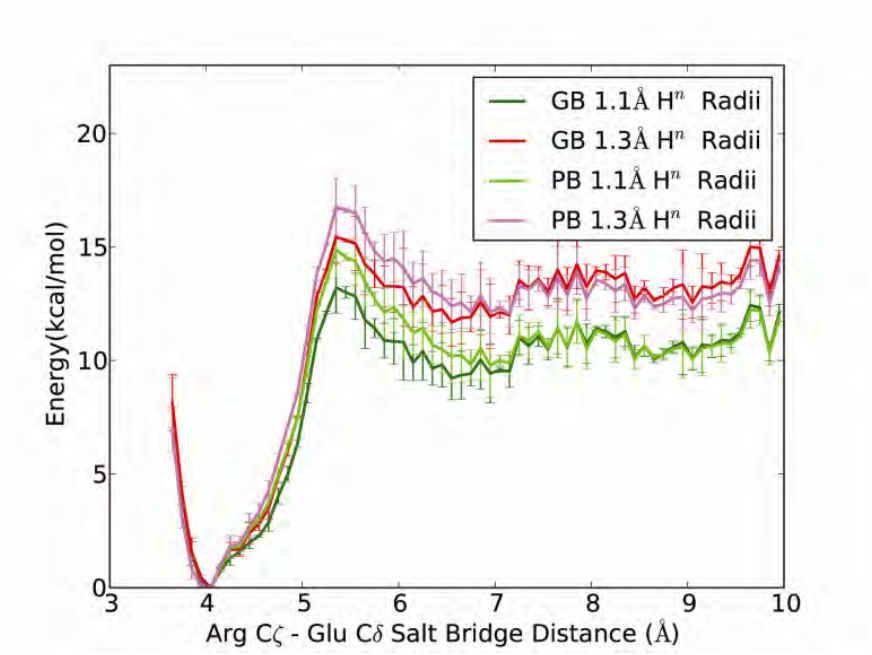

Figure 2-8 Lowest energy profile of GB or PB implicit solvent models using 1.1 Å or 1.3 Å H^n radii. Each point on the curve represents the average energy of 20 lowest energy structures with salt bridge distance within corresponding distance range (bin size 0.1 Å). Curves are zeroed at their global minimum for easier comparison. Error bar was generated using the first and second half of data.

2.3.2 Validation in HIV-1 protease (HIVPR) system

After validating the intrinsic radii modification in a small peptide system, we wanted to validate it in a macromolecule system. We chose HIVPR because our lab has previously carried out extensive studies on HIVPR using both GB and explicit solvent simulations [42, 43, 96, 97, 109]. From our experience, flap deformation occurs in GB simulations of HIVPR, especially for the D25N active site mutant. The deformation happens on nanosecond time scale. The D25N mutant HIVPR is widely used in experiments since it retains native fold and is able to bind natural substrates without initiating catalytic reaction. Although it has been found that D25N decreases protease dimer stability [110], we suspect that the flap deformation on such a small time scale is caused by simulation artifacts, probably the strong salt bridge near the flap elbow region. Therefore, we compared HIVPR GB simulations with 1.3 Å or 1.1 Å H^n radii, 60 ns each, to 400 ns EXP simulation.

There are four salt bridges involving Arginine near the elbow region, two on each monomer (Figure 2-4), and we calculated the PMF of them (Figure 2-9). For the salt bridge between Arg57 and Glu35, reducing H^n radii from 1.3 Å to 1.1 Å clearly improved the agreement between GB and TIP3P model (Figure 2-9 A-B). When H^n radii were set to 1.3 Å, salt bridge distance was always below 6 Å, which means the salt bridge never breaks. After decreasing H^n radii to 1.1 Å, the salt bridge strength in GB was closer to TIP3P model. The salt bridge between Arg41 and Asp60 (Figure 2-9 C-D) breaks more often than the Arg57-Glu35 salt bridge, which is indicated by the lower relative energy at longer salt bridge distance (more populated). Interestingly, although the 1.1 Å H^n radii profiles are consistent between two monomers, there is a 0.5 kcal/mol shift for both 1.3 Å H^n radii profiles and TIP3P model profiles, indicating inadequate conformational sampling in both simulations, which is a common problem faced in macromolecule simulations. Overall, although the agreement between GB-OBC-with-reduced-H^n-radii and TIP3P is less pronounced in Arg41-Asp60 salt bridge than in Arg57-Glu35 salt bridge, the 1.1 Å radii simulations still overlap with TIP3P better than 1.3 Å radii simulations. Therefore, we concluded that reducing H^n radii gives an overall better agreement between GB and explicit solvent on HIVPR elbow salt bridge strength.

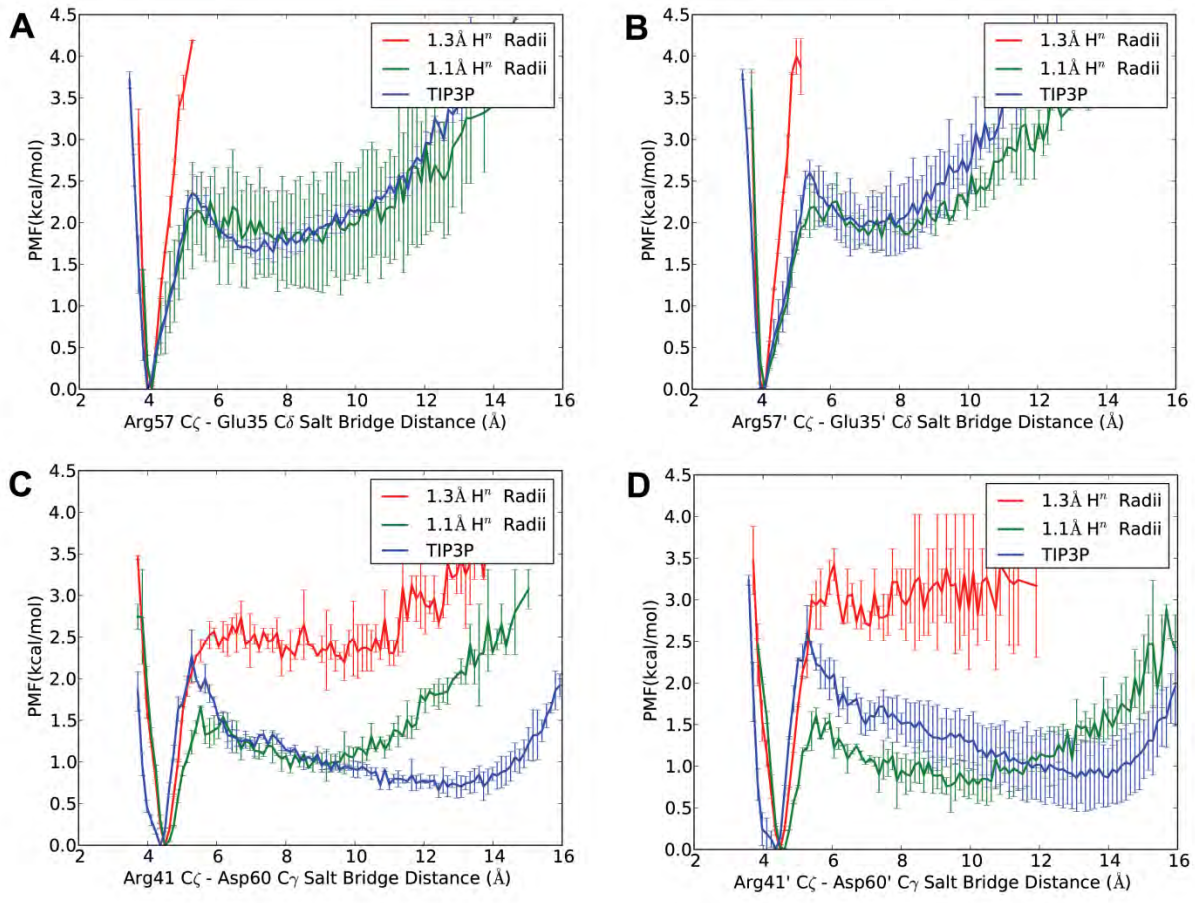

Figure 2-9 Potential of mean force (PMF) profile of salt bridges on the flap elbow of HIVPR. A-B) Salt bridge distances between Arg57_Cζ and Glu35_Cδ on each monomer, C-D) salt bridge between Arg41_Cζ and Asp60_Cγ on each monomer. Error bars in each set were obtained by comparing the first and second half of data.

We then evaluate if the improvement in salt bridge description would indeed improve the behavior of HIVPR GB simulation. We looked at the time dependence of salt bridge distances as well as flap RMSD (Figure 2-10). Looking at the salt bridge distances on the left, the TIP3P explicit solvent simulation have all four salt bridges break and form over 400 ns simulation time, although the salt bridge between Arg57 and Glu35 (on both monomers) is stronger than Arg41-Asp60 salt bridge (Figure 2-10 C, the salt bridge is formed when the pair distance maintains at lower distances). However, the Arg57-Glu35 salt bridge in the GB simulation with 1.3 Å radii almost never breaks (Figure 2-10 A), and the Arg41-Asp60 salt bridge is also much stronger compared to TIP3P simulations. Comparing their flap dynamics, in the TIP3P explicit solvent the HIVPR flaps maintained closed conformation throughout the 400 ns simulation (Figure 2-10 D, the flap RMSD to closed flap conformation is always small and lower than the RMSD to semiopen flap conformation), while in the GB implicit solvent with1.3 Å H^n radii the HIVPR flaps stayed open flap conformations (flap RMSD values to closed and semiopen flap conformation are both above 9 Å) more often and ended at very high flap RMSD which is an indication of serious flap deformation (Figure 2-10 B). The last snapshots of both simulations

(Figure 2-10 B and D) are consistent with the flap RMSD: the TIP3P simulation stayed at closed flap conformation while the GB simulation with original H^n radii suffered from flap deformation.

In contrast, when the H^n radii was reduced to 1.1 Å, the elbow salt bridges broke more often than the original radii GB simulations (Figure 2-10 E). The ratio of broken/formed salt bridges is comparable to TIP3P solvent simulation (Figure 2-10 C), although the salt bridge break/form time scales are not the same due to the viscosity difference between the two solvent models. The flap RMSD with reduced radii (Figure 2-10 F) indicates multiple flap handedness switching events, where the lowest RMSD switches between the black and red curves (meaning the flaps go from one conformation to the other conformation). Also there were multiple transient flap-opening events, during which the RMSD steadily increases to more than 9 Å and then decreases back to low RMSD to either reference structure. The last snapshot of the 1.1 Å radii simulation shows the HIVPR having one flap open and the other flap closed, which is a common transition state in HIVPR simulations [42]. Overall, reducing H^n radii corrected the salt bridge strength and prevented flap deformation in HIVPR simulations in GB-OBC implicit solvent model.

We also noticed that the flap RMSD does not seem to correlate with salt bridge distances directly, since salt bridge forming and flap opening do not happen at the same time. Therefore we hypothesize that, through an allosteric control, the salt bridges at the elbow region would increase/decrease the chance of flap opening upon formation/breaking, respectively.

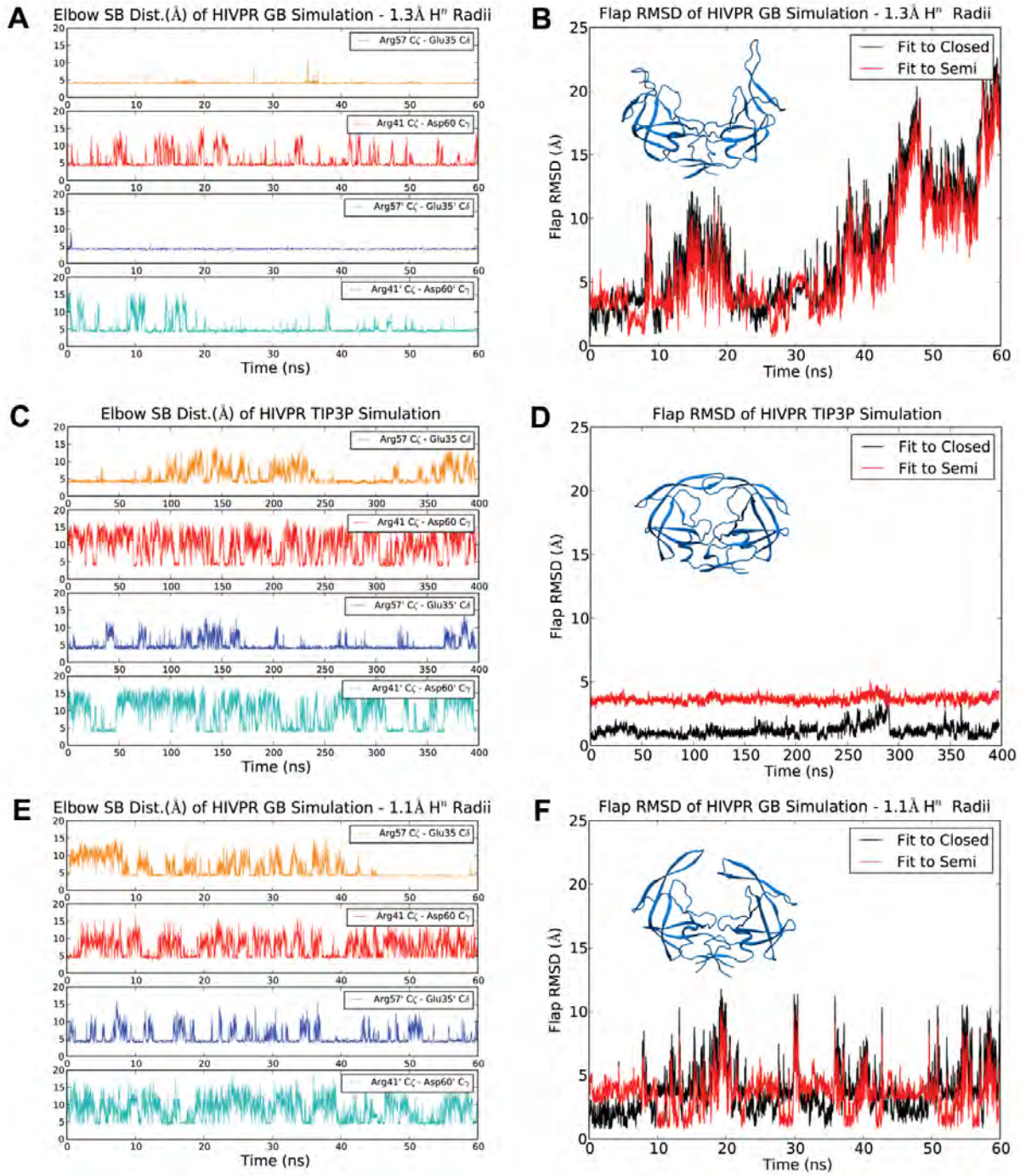

Figure 2-10 Salt bridge distance (A, C, and E) and flap RMSD (B, D, and F) calculated from HIVPR simulations. A-B) GB simulation with 1.3 Å H^n radii. C-D) TIP3P EXP simulation. E-F) GB simulation with 1.1 Å H^n radii. The salt bridges between atom pair Arg57_Cζ to Glu35_Cδ, Arg41_ Cζ to Asp60_Cγ, Arg57'_Cζ to Glu35'_Cδ, and Arg41'_ Cζ to Asp60'_Cγ are plotted in orange, red, blue and cyan, respectively. Flap RMSD to the closed and semiopen crystal structures are plotted in black and red, respectively. The final frame of each simulation is shown above the RMSD curves.

2.4 Conclusions

In this study, we improved salt bridge strength and description in simulations with ff99SB force field and GB-OBC implicit solvent by decreasing the H^n radii (H^n denotes Hε and Hη on Arginine residue side chain). It was suggested previously that the radius of a hydrogen atom should be smaller when the electronegativity of its bonding partner is greater [99]. We attempted H^n radii modification, and the simulations from both small model peptide and macromolecule HIVPR verified the approach of reducing H^n radii. Therefore, differentiating between H^n atoms and backbone hydrogen atoms when assigning intrinsic radii may be necessary to achieve better GB simulation results, and we suggest considering 1.1 Å H^n radii for peptide or protein simulations with ff99SB/GB-OBC.

Apart from comparing GB simulations with different H^n radii against TIP3P explicit solvent simulations, we also evaluated the intrinsic radii dependency of PB method, which is also widely used as a standard for GB model optimization. Our results demonstrated that PB method itself is also strongly dependent on intrinsic radii selection, and PB calculations with original H^n radii is equally erroneous as GB methods, if not worse. Therefore, optimizing intrinsic radii might be needed for all implicit solvent methods. We recommend using explicit solvent simulations as the standard when optimizing intrinsic radii parameters, as performed previously [111].

Importantly, the optimized HIVPR GB simulation protocol developed here would serve as the foundation for the study of HIVPR flap population distribution (chapter 3), which is too time-consuming using explicit solvent simulations.

Chapter 3 Spin-labeled HIV-1 subtype protease simulations explain EPR spectrum shift and AIDS drug resistance mechanism

3.1 Introduction

Due to development of antiretroviral therapy, the AIDS related death rate is decreasing steadily, although it is still on the order of million per year globally. One of the biggest obstacles in HIV inhibition is its extreme genetic heterogeneity, which results from the rapid viral turnover, the high virus burden, and the lack of proofreading machinery in error-prone reverse-transcription [112, 113]. Naturally occurring polymorphisms and drug-induced mutations can lead to changes in anti-HIV drug targets, so that drugs designed toward one sequence may not work well for other sequences.

HIV sequences have been classified into types and subtypes according to their phylogenetic relationship. There are two types, HIV type 1 (HIV-1) being more infectious than HIV type 2 (HIV-2). Most studies have been focused on type 1, which is then divided into three groups: M (major), O (outlier), and N (neither M nor O) group. The M group is further separated into subtypes and circulating recombinant forms (CRF). Most experimental and clinical data were collected for subtype B (responsible for 10% of global infections [114], Figure 3-1) because of its prevalence in developed countries. In contrast, relatively few data are available for non-B subtypes, such as subtype C (responsible for 50% of global infections [114] and is prevalent in Africa, Figure 3-1). There is evidence that polymorphisms often worsen drug binding [115, 116], and many polymorphisms found in non-B subtypes coincide with drug resistant mutations (Figure 3-2) in subtype B [117]. Therefore, studying non-B subtypes would contribute to AIDS treatment in developing countries, prepare developed countries for rapid globalization, and enrich existing knowledge on subtype B drug resistance.

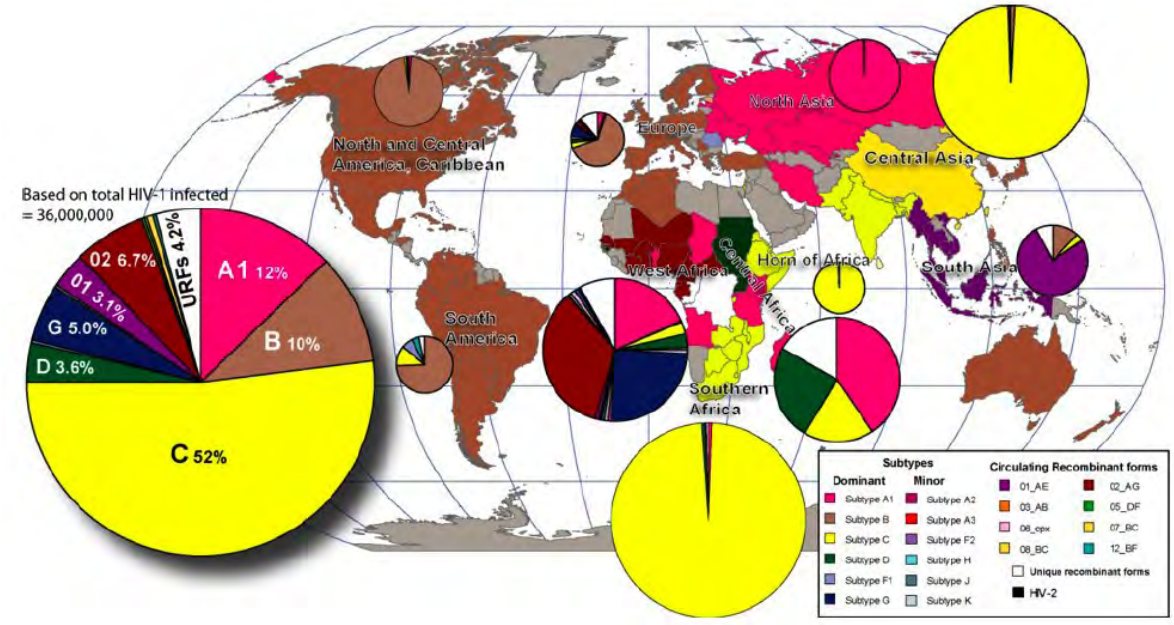

Figure 3-1 Subtype epidemic of HIV (www.hivviralload.com). Leftmost labeled pie plot shows the composition of global infections in terms of different subtypes. The other unlabeled pie plots

demonstrate the composition in different regions and are proportional in size to the number of regional infections.

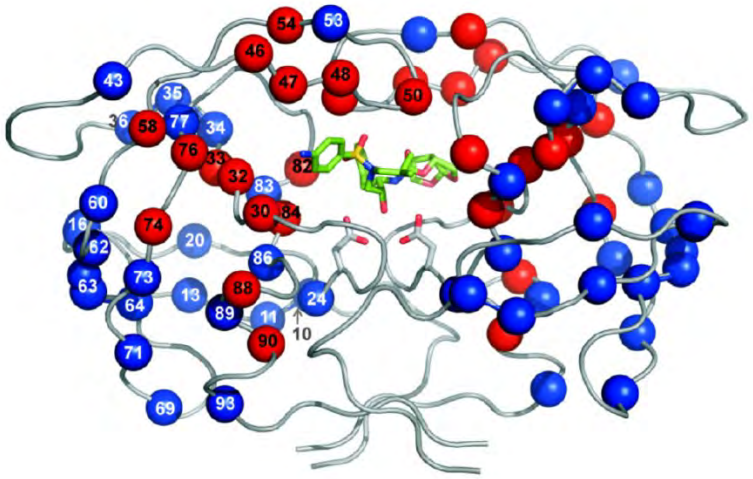

Figure 3-2 Primary and secondary mutation sites in HIV-1 protease. Residue numbers are labels on the monomer on the left, and the right monomer has the same mutation sites because of the homodimer symmetry. Primary mutation sites are shown as red balls, and secondary mutation sides are shown as blue balls. The inhibitor and catalytic site residues are shown in licorice representation.

One popular drug target in AIDS treatment is HIV-1 protease (HIVPR), which is crucial for viral maturation [118]. The functional form of HIVPR is a homodimer. Two catalytic aspartate residues, one on each monomer, reside on the dimer interface and are shielded by two flexible β-hairpin flaps (Figure 3-3). Because the two flaps govern the ligand access to the active site, many studies have been done to characterize their structure and dynamics. All existing crystal structures demonstrate that the ligand bound form of HIVPR has closed flap conformation, where the two flaps form extensive contact with the ligand [3], while the apo form can adopt semiopen and open forms besides the closed conformation (Figure 3-4). Semiopen flap conformation [4, 104] has the opposite flap handedness (the right flap is always in front of the left flap when viewed in the flap-up orientation shown in Figure 3-4 VII) and larger flap-tip-to-active-site distance, compared to closed conformation. Open flap conformation [119, 120] features much larger distance between flap tips which permits the entry of substrate peptide. However, there are limitations in crystallography that need to be considered when interpreting these structures: inter-flap interactions are missing in semiopen crystals because they are solved as monomers, apo HIVPR crystals with closed flap conformation are tethered in the termini region (Figure 3-4 II and VI, although the linker region is not resolved in the crystal structure) [121, 122], and the capture of transient and relatively high energy open flap conformation in crystals is largely due to crystal packing [109] etc. As Freedberg et al. concluded in their solution NMR study, apo HIVPR likely adopts various conformations from fully closed to wide open prior to ligand binding [123]. Their data suggest that flap tips fluctuate on the sub-nanosecond time scale, and the flaps open up on the microsecond to millisecond time scale [123-125]. Besides structures from crystallography and rates from NMR, molecular dynamics (MD)

simulations use physical laws to animate proteins and model their dynamics in real life. Hornak et al. showed for the first time the reversible interconversion of closed and semiopen forms, as well as the reversible and transient flap opening events [42]. Moreover, the flap-ligand interaction and ligand binding pathway have also been explored by different groups [51, 52].

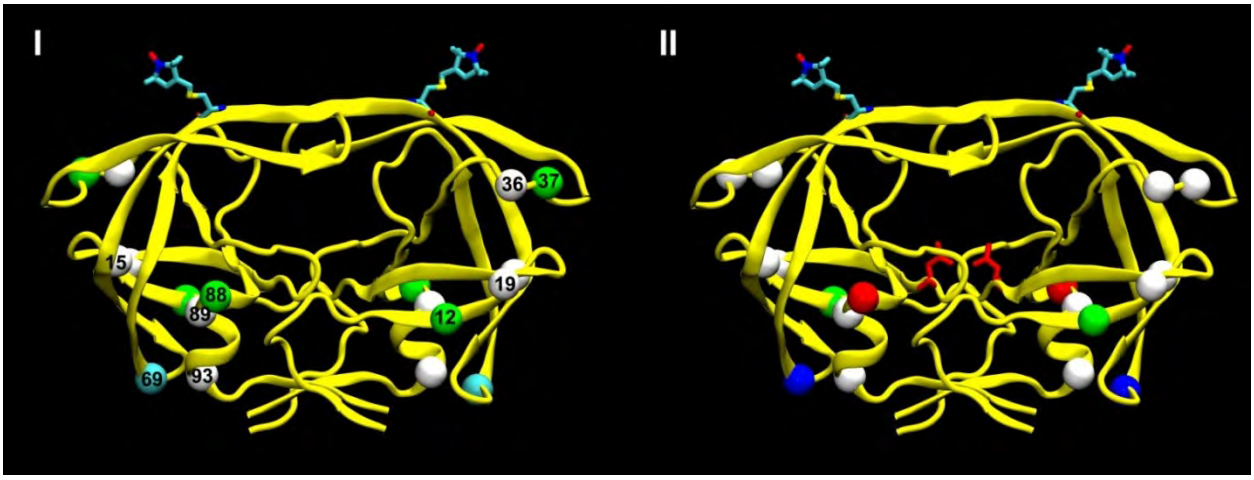

Figure 3-3 HIV-1 subtype B (I) and C (II) protease simulation starting structures built from a crystal structure of holo subtype C (PDB ID: 2R5P). Protein backbone is shown in yellow. MTSL-attached Cys55 on both monomers are shown in licorice representation. Polymorphisms are shown in ball representation, colored by their residue type, and labeled with residue number in panel I. Catalytic residues are colored red and shown in licorice representation in panel II.

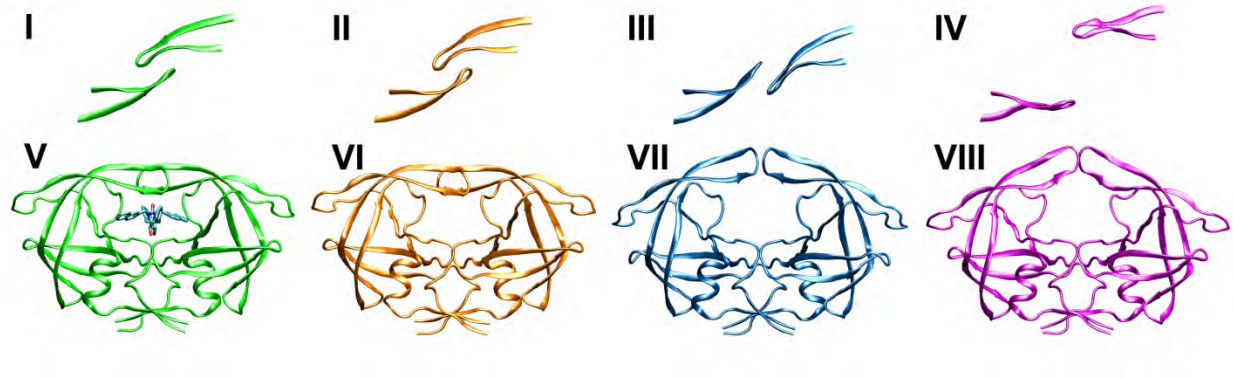

Figure 3-4 Crystal structures of holo and apo HIVPR. Panel I to IV are the top views of the flaps, while panel V to VIII are the front views of the whole molecule. From left to right: holo HIVPR is shown in green (PDB ID: 1HVR); apo HIVPR with closed flap conformation is shown in orange (PDB ID: 1G6L, in which the two protease monomers are tethered at the termini region, although the linker region is disordered in the crystal structure); apo HIVPR with semiopen flap conformation is shown in blue (PDB ID: 1HHP, the biological unit is used for rendering); and apo HIVPR with open flap conformation is shown in magenta (PDB ID: 1TW7).

Although apo HIVPR can adopt different flap conformations, the open form is needed for substrate entry and the closed form is needed to optimize the substrate orientation for catalytic

reaction. Previous simulations suggest that the substrate binds to the apo HIVPR with open flaps, induces the flap closure, and confines the holo protease flaps to the closed state until the substrate is cleaved [43, 44]. The holo structure is stabilized by the favorable interactions between the ligand and the enzyme, which offset the strain caused by the loss of conformational entropy, and also the unfavorable structural change, if any, to accommodate the binding. With the implication from crystallography and MD simulations, it is desirable to obtain the exact equilibrium information of HIVPR before and after drug binding, because the shift in equilibrium is directly related to the strain penalty upon binding [126]. In recent years, electron paramagnetic resonance (EPR) spectroscopy method has been successfully applied to HIVPR to help answer the flap population question. In so called site-directed spin labeling (SDSL) double electron-electron resonance (DEER) experiment, which is also referred to as pulsed electron double resonance (PELDOR), MTSL spin labels are attached to cysteine residues in the HIVPR flaps (one on each flap, Figure 3-3). With pulsed EPR, the modulation of the echo intensity of the nitroxide signal is measured and converted to distance profile [127]. However, although EPR provides dynamics information which is missing in crystal structures, EPR itself also has limitations: spin label local interactions as well as the liquid nitrogen environment may alter the EPR spectrum, curve fitting is performed to generate the distance profile, and additional methods are needed to connect the spectrum to structure. Previous collaboration between EPR and MD enabled assigning different flap conformations to different EPR signals [96, 97, 128, 129].

Kear et al. [67] used SDSL DEER to probe the apo flap conformations of HIVPR from different subtypes. The most significant difference is found between subtype B and C (Figure 3-5). Their EPR data, combined with previous MD results, suggests that HIV-1 subtype C protease has larger fractional occupancy of open flap conformation than subtype B. However, it remains unclear how the polymorphisms, mostly distal from the flap region (Figure 3-3), would affect the flap conformation. In order to explain the EPR data and pinpoint polymorphisms that are most important in differentiating two subtypes, we performed MD simulations of spin-labeled subtype B and C proteases (denoted as “sub.B” and “sub.C” later on). Because the flap opening happens in microsecond to millisecond time scale, we used batch simulations and implicit solvent model, both of which were shown to speed up the convergence of conformational sampling [71, 130]. After three most important polymorphisms were identified (M36I/S37A/H69K), we validated the finding by performing batch simulations on sub.C with the three polymorphisms back-mutated to sub.B sequence (denoted as “mutant” later on, which is expected to behave just like subtype B).

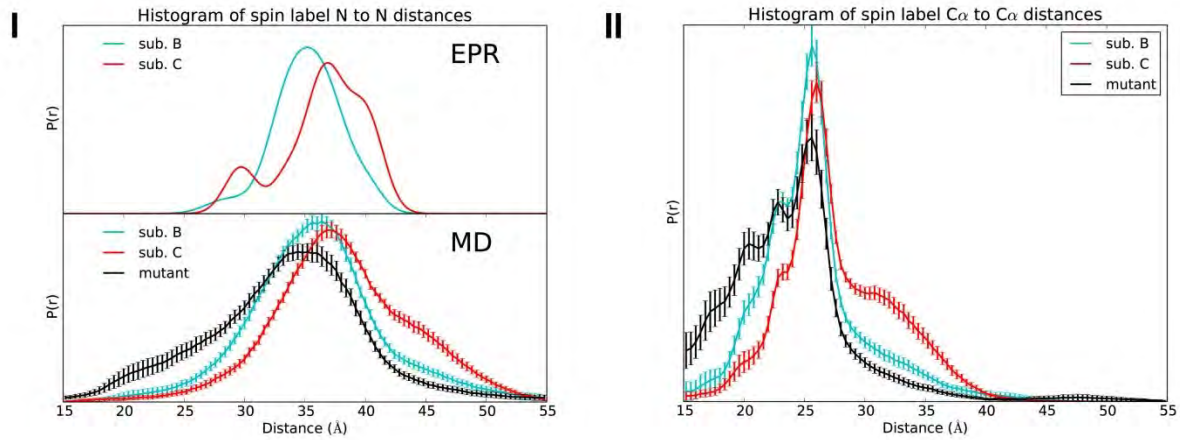

Figure 3-5 Normalized histogram of inter-label nitroxide nitrogen distances compared to EPR data from Kear et al. 2009 JACS paper (I) and inter-label Cys-MTSL CαCα distances (II) measured from simulations. Error bars were calculated as standard error of the mean of independent runs.

3.2 Methods

3.2.1 Simulation setup

The crystal structure of HIV-1 subtype C protease complexed with Indinavir (PDB ID: 2R5P [131]) was used to build the starting structures for both subtype B and C, to eliminate influence from different crystallization conditions. Virtual mutations were performed using SwissPdbViewer [132] to match subtype sequences to those used in the EPR experiment [67]. All simulated sequences contain mutations for EPR experiments (Q7K/L33I/L63I/C67A/C95A/D25N and K55C for MTSL attachment). Apart from these mutations, subtype B sequence follows consensus LAI sequence [67], while subtype C contains nine polymorphisms: T12S/I15V/L19I/M36I/S37A/H69K/N88D/L89M/I93L. Later we also carried out subtype C mutant simulations with sequence T12S/I15V/L19I/ N88D/L89M/I93L, and subtype B simulations with protonated Histidine at position 69. Protein parameters were from ff99SB [64]. Spin label was modeled using charge parameters from Haworth group, to be consistent with our previous study [97]. and non-charge parameters from generalized AMBER force field (GAFF [133, 134]). GB-OBC implicit solvent model [63] was used, with reduced Arg polar hydrogen radii to improve the salt bridge strength description [130]. We left out the surface area term, which accounts for hydrophobicity, because of the limitation in current implementations [135-138].

3.2.2 Simulation

All simulations used the AMBER 11 simulation package [23]. Subtype B and C starting structures were subjected to energy minimization and equilibration prior to the production runs. All MD equilibrations/simulations used 1 fs time step, SHAKE [28] constraints on bonds involving hydrogen atoms (tolerance 10^{-6}), Langevin temperature control (collision frequency 1 ps^{-1}), and no nonbonded cutoff. Forces related to effective radii calculation, along with each

interaction pair greater than 15 Å, were updated every 2 steps in equilibrations and every 4 steps in production. Firstly, hydrogen atoms were minimized for 10,000 steps (first 1000 steps being steepest descent gradient, and the rest being conjugate gradient) with harmonic positional restraints on all other atoms (force constant 100 kcal/mol·Å^2) to maintain their starting positions. Secondly, the system was heated from 100 K to 300 K during 250 ps and then maintained at 300 K for 250 ps, while keeping the same harmonic restraints on non-hydrogen atoms. Thirdly, hydrogen atoms were equilibrated for 500 ps, followed by a 10,000 step energy minimization on sidechain atoms while restraining backbone heavy atoms (force constant 100 kcal/mol· Å^2). At last, we again heated and equilibrated at 300 K but only backbone heavy atoms were restrained. In the last step, we adopted two slightly different equilibration protocols in order to generate two independent equilibrated structures for each subtype. In one protocol, backbone heavy atom restraints were added on residues that did not undergo virtual mutation, with force constant decreasing from 10 kcal/mol·Å^2 to 0.1 kcal/mol·Å^2 over 1.5 ns. In the other protocol, each virtual mutation (one on each monomer) was relaxed separately, with restraints (force constant 10 kcal/mol·Å^2) added on heavy atoms of residues more than 5 Å away from the mutated residue [139], for 50 ps prior to the 1.5 ns sidechain equilibration. Multiple independent production runs (28 runs for sub.B and sub.C, and 24 runs for mutant and sub.B within protonated H69, with randomized velocity) were carried out from each of the two equilibrated structures. Production runs were kept at 325 K. Weak distance restraints were applied to the termini portion of the dimer interface during production runs, as described in our previous study [130], to prevent flexible N terminus from forming helical structure. The simulation length for each independent run was about 20 ns, totaling about 1 us for each system. Simulation trajectories were recorded at 1 ps interval and used for distance measurement and energy decomposition.

3.2.3 Inter-label distance histograms

Ptraj tool in AMBER 11 simulation package was used for distance measurement. Distance histograms were generated for each independent run using the same binning (15 Å to 55 Å, 100 bins) and then the average population of each bin was plotted, with the standard error of the mean (SEM) as the error bar. All histograms were normalized.

3.2.4 2D-coordinate system for flap opening

Inter-label CαCα distance and handedness dihedral were combined to probe global motion of HIVPR. Handedness dihedral was defined by the center-of-mass positions of four groups: Cα atoms of residue 48/49/52/53 on monomer A as group one, Cα atoms of residue 87 to 92 on monomer A as group two, Cα atoms of residue 87 to 92 on monomer B as group three, and Cα atoms of residue 48/49/52/53 on monomer B as group four (Figure 3-6).

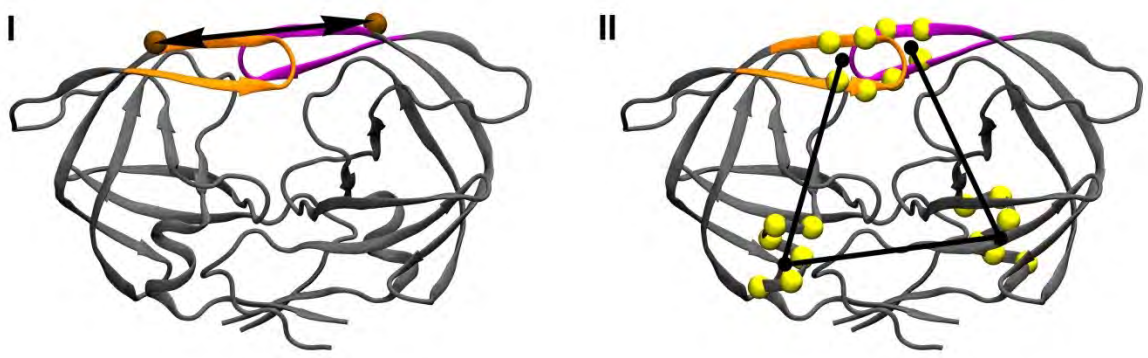

Figure 3-6 HIVPR flap conformation described by a 2D-coordinate system. Structures took from simulations. I) Inter-label Cys-MTSL CαCα distance, with two Cys55 Cα atoms highlighted. II) Flap opening dihedral. The Cα atoms involved (48/49/52/53/87/88/89/90/91/92) are shown as vdW spheres, and four centers of mass are highlighted to illustrate the opening dihedral.

3.2.5 Free energy profile

After 2D binning using inter-label CαCα distance and handedness dihedral as criteria (Binning limits were 15 Å to 45 Å in CαCα distance, and -40 ° to 40 ° in handedness dihedral, with 15 and 20 bins on each coordinate, respectively), the population of structures in each bin was converted to PMF using Equation 2-3. Only bins with more than 100 structures are shown.

3.2.6 Convergence check

The convergence of conformational sampling was evaluated by comparing free energy profiles of independent runs (Figure 3-7, Figure 3-8, and Figure 3-9).

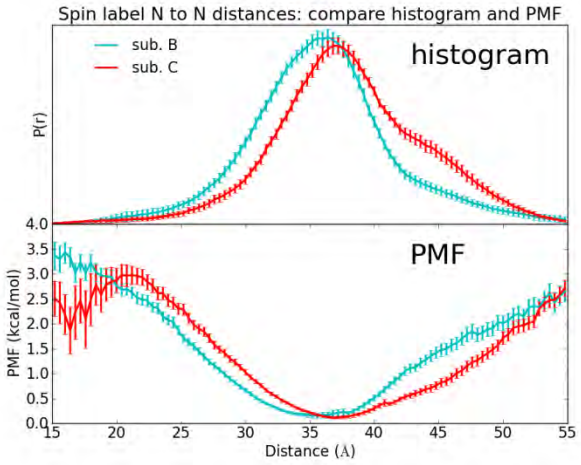

Figure 3-7 PMF of sub.B and sub.C simulations (II) compared with inter-label nitroxide nitrogen distance histogram (I). Error bars are calculated as SEM of independent runs.

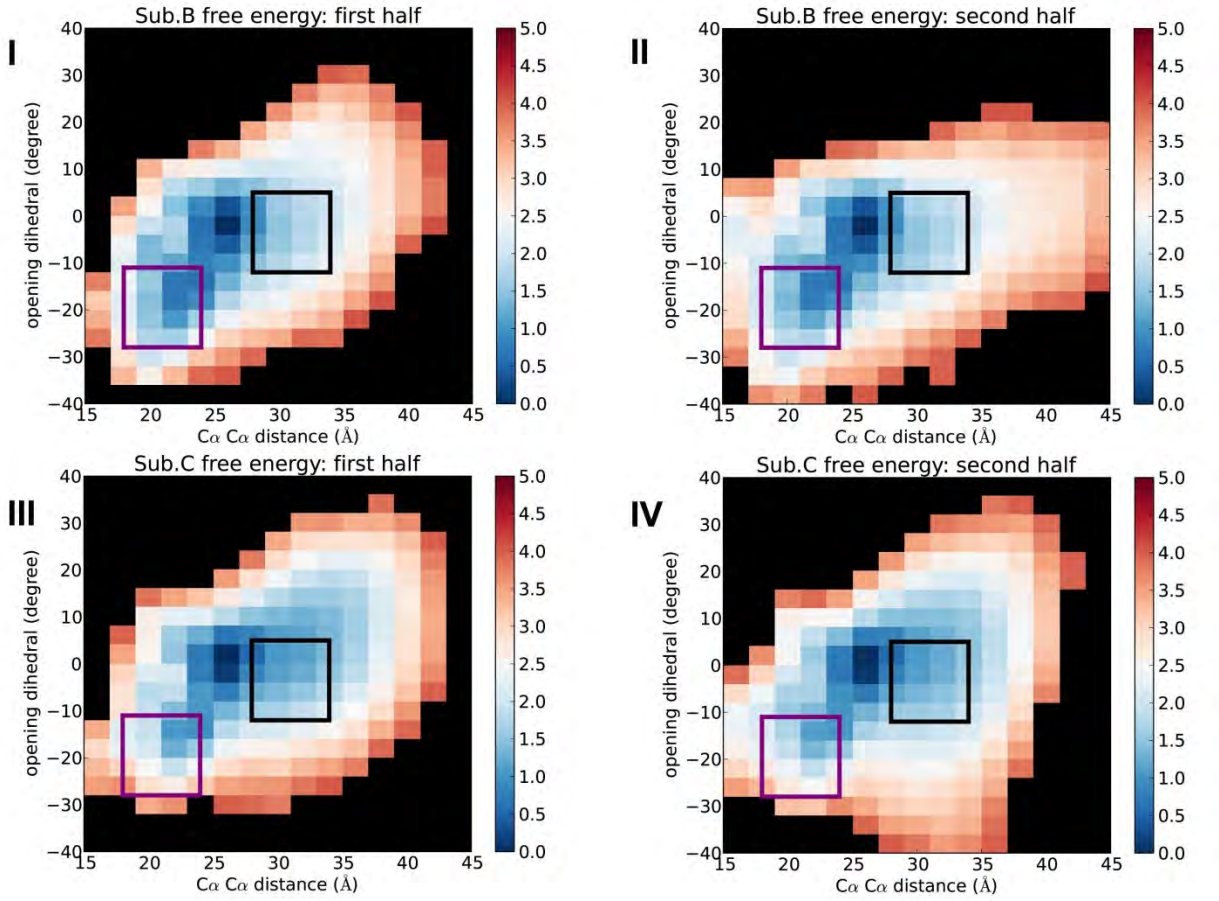

Figure 3-8 Convergence check of sub.B (I and II) and sub.C (III and IV) free energy profiles – part 1 of 2. Simulations trajectories were concatenated and then the first and second half data were plotted separately. Closed and open flap conformations are indicated by purple and black squares, respectively.

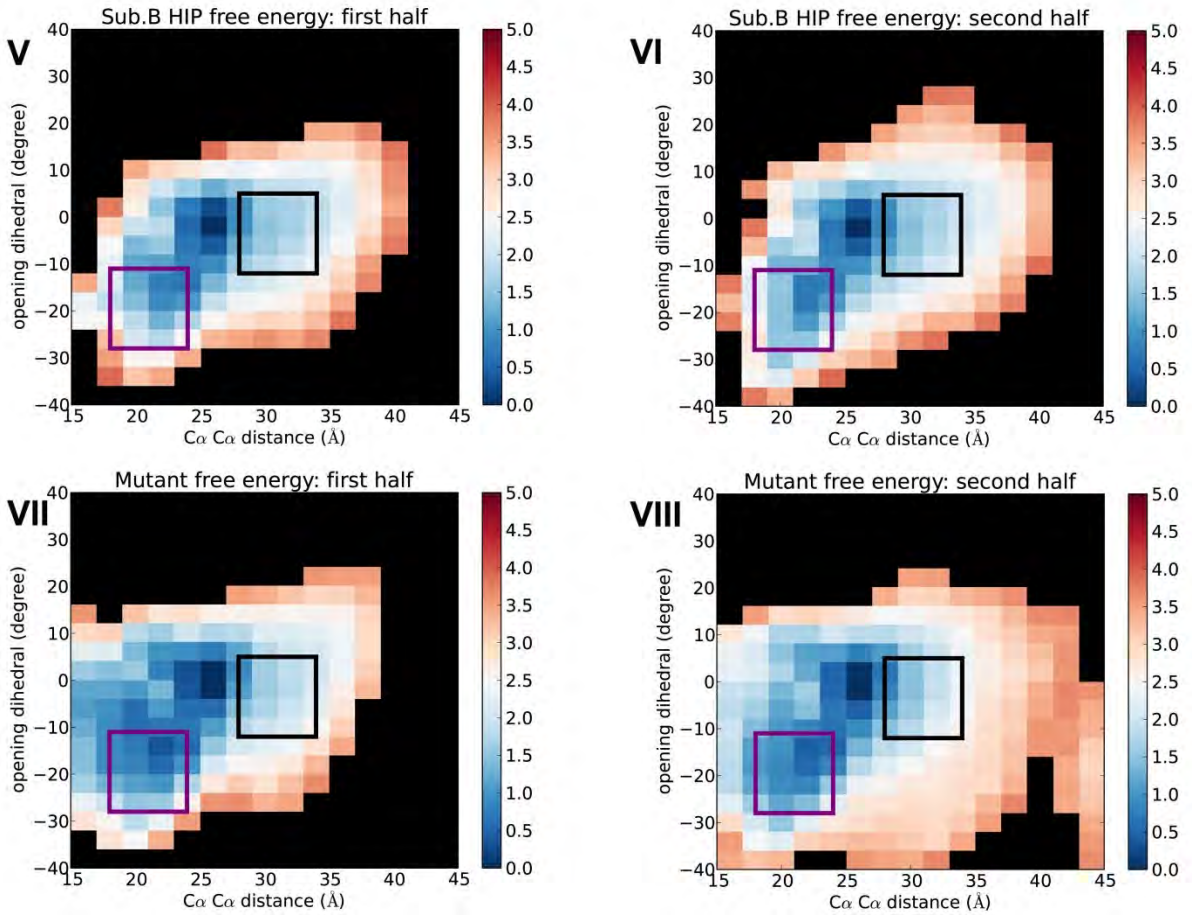

Figure 3-9 Convergence check of sub.B HIP (I and II) and mutant (III and IV) free energy profiles – part 2 of 2. Simulations trajectories were concatenated and then the first and second half data were plotted separately. Closed and open flap conformations are indicated by purple and black squares, respectively.

3.2.7 Potential energy profile

The same 2D-coordinate system as the free energy profile was used, and the average potential energy of each bin was plotted. Energies were zeroed at most populated bin. Only bins with relative energy between -40 kcal/mol to 40 kcal/mol were shown, and bins with relative energy out of bounds or population less than 100 were shown in black.

3.2.8 Energy decomposition

Simulation trajectories were combined and subjected to Molecular Mechanics Generalized Born Surface Area (MMGBSA) energy decomposition using sander module in AMBER, during which the potential energy of each structure was calculated and decomposed into per-residue and per-residue-pair components [140, 141]. In per-residue energy decomposition, a particular residue's energy is the sum of any atoms in this residue. In per-residue-pair decomposition, only the energy terms that are between one atom from the first

residue and another atom from the second residue are included. For GB energy, the effective radii are determined from all the atoms in the system, and then the pairwise energy terms are calculated. We included 1-4 vdW and 1-4 electrostatic energies into vdW and electrostatic energy, respectively. We used exactly the same parameters as the GB simulations, therefore the SA term was left out. Although we included termini restraints during the simulation, which were not considered in the energy decomposition calculation, the restraint energy turned out to be small (less than 0.5 kcal/mol) 91% of the time and had minimal influence on the simulation.

3.2.9 Select residues for per-residue energy ranking

We wanted to cut down the number of residues being ranked by searching polymorphisms and their neighbors, whose potential energy sum could give a decent match to the free energy profile. Polymorphisms are all included in the energy ranking. Neighbors of polymorphisms are differentiated based on their distance to the nearest polymorphism, and we tested different distance criteria to determine how many neighboring residues to include in the ranking. Different protease structures (Figure 3-10) were considered to account for protein flexibility: a neighbor-polymorphism distance may be smaller or larger than the distance criterion depending on the protease conformation, and we took the minimum distance found in all structures. We plotted out energy sum (vdW and electrostatic) of residues within certain distance cutoff from the polymorphisms (2 Å to 6 Å, see Figure 3-11, Figure 3-12, Figure 3-13, and Figure 3-14) and compared to the total molecular energy profiles (Figure 3-15 and Figure 3-16, V-VIII). The vdW energy profiles are essentially the same between two subtypes for the residues within distance cutoffs we analyzed (Figure 3-11 and Figure 3-12), so the relative vdW energy difference found in Figure 3-15 V-VI should come from residues distal from polymorphisms, most likely the flap region. It is possible that changes in protease core region have effect on vdW interactions elsewhere, but overall the vdW difference is much more subtle than the electrostatic difference. Based on electrostatic energy profiles (Figure 3-13 and Figure 3-14), we concluded that distance cutoff 2 Å is enough to give similar trend without involving too many residues.

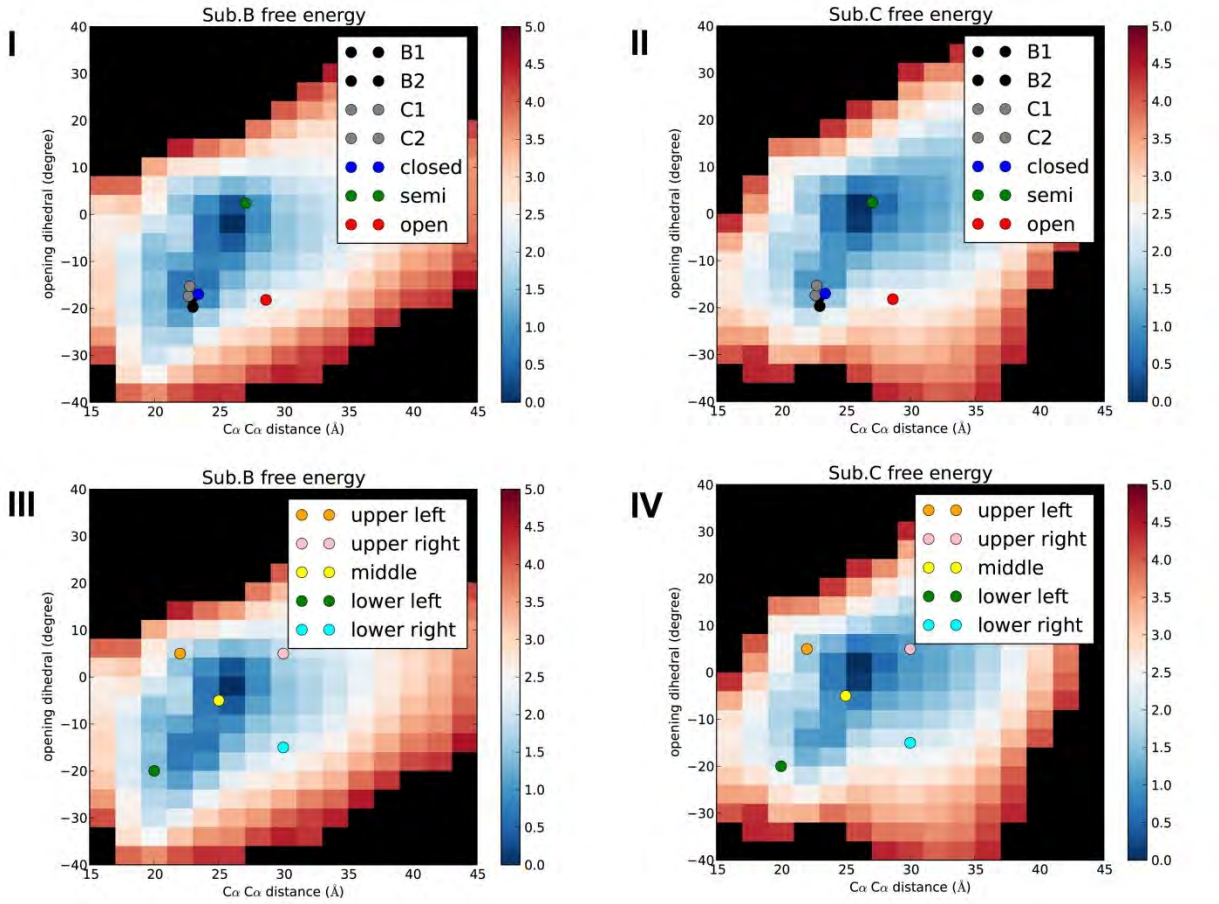

Figure 3-10 I-II) Simulations starting structures and crystal structures are labeled on the free energy surface. B1 and B2 denote structures after two independent equilibrations of subtype B. C1 and C2 denote structures after two independent equilibrations of subtype C. Three crystal structures are mapped: 2R5P as closed, 1HHP as semiopen, and 1TW7 as open flap conformation. III-IV) Five structure clusters used for determining nearby-residue mask and filtering per-residue energy contribution are labeled on the free energy surface.

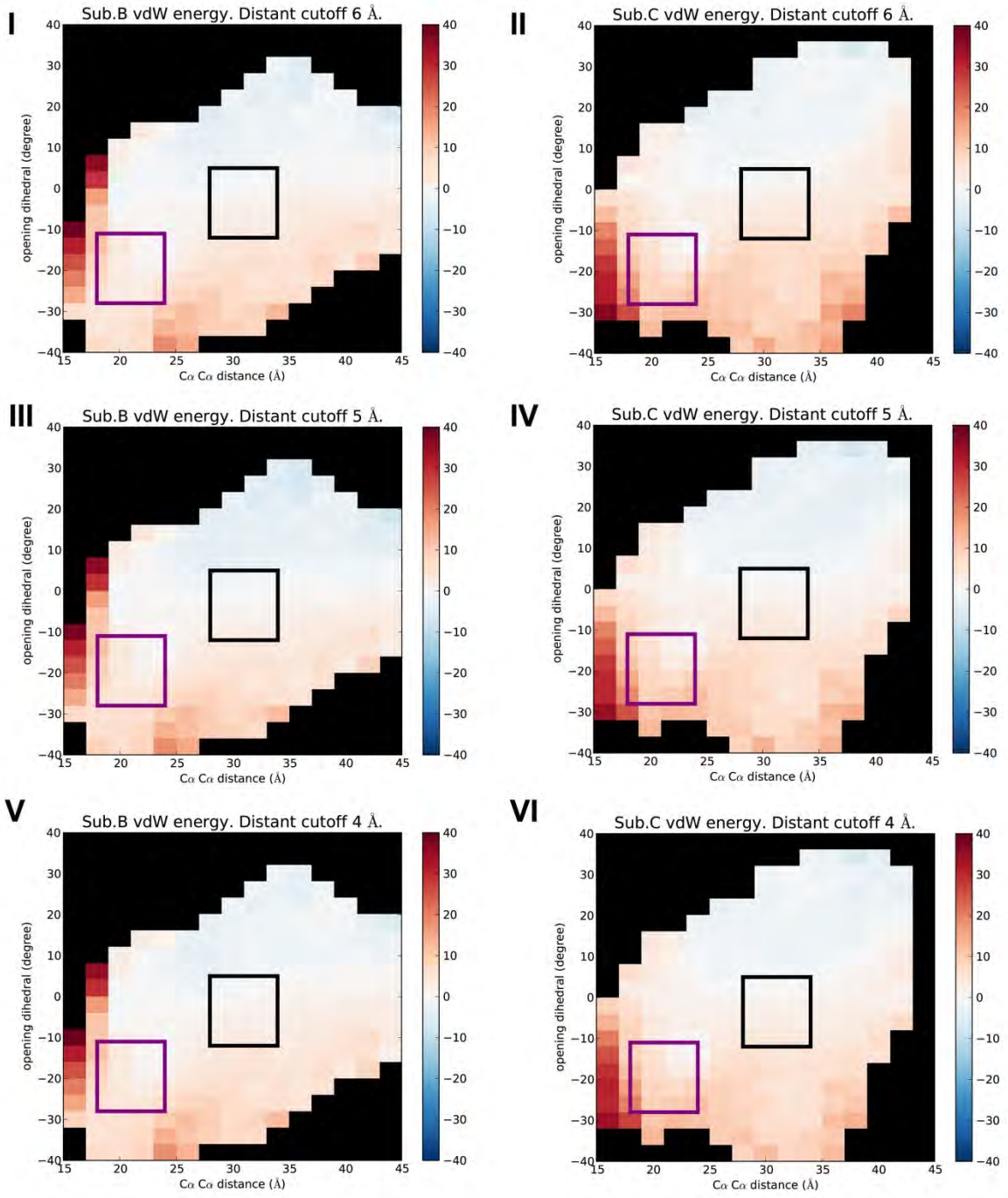

Figure 3-11 Sum of per-residue vdW energy, including polymorphisms and nearby residues within certain distance cutoff – part 1 of 2. Closed and open flap conformations are indicated by purple and black squares, respectively.

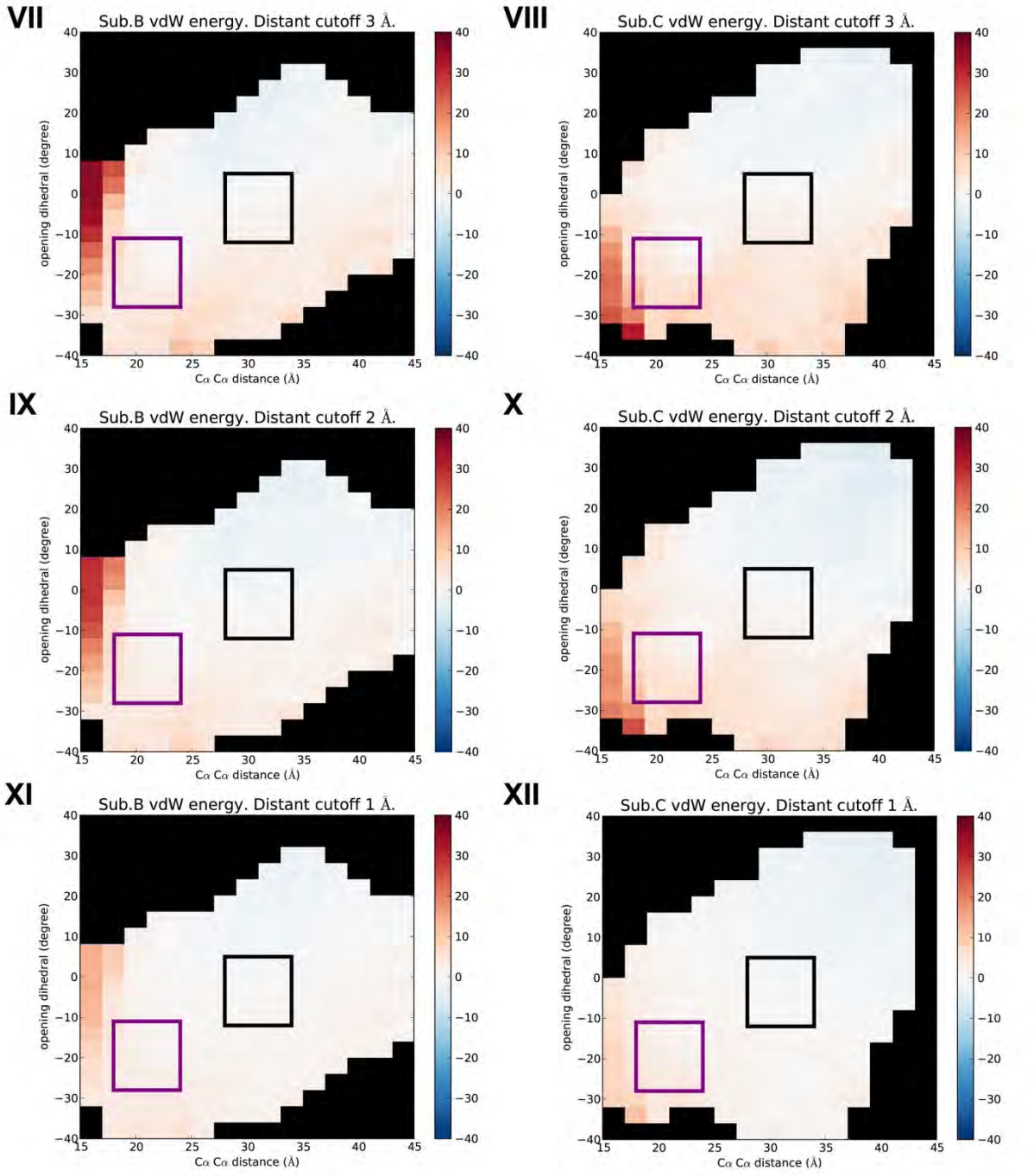

Figure 3-12 Sum of per-residue vdW energy, including polymorphisms and nearby residues within certain distance cutoff – part 2 of 2. Closed and open flap conformations are indicated by purple and black squares, respectively.

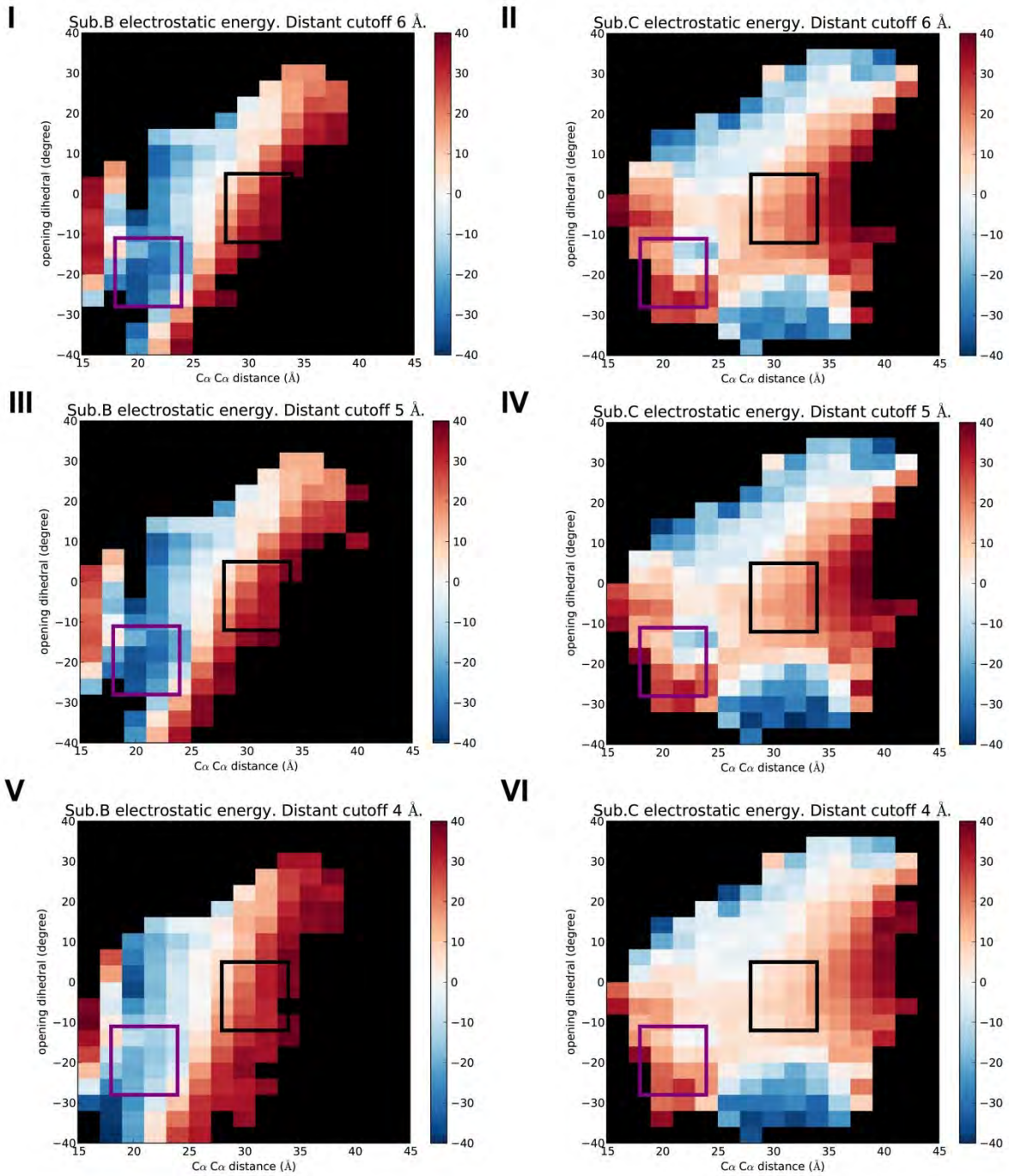

Figure 3-13 Sum of per-residue electrostatic energy, including polymorphisms and nearby residues within certain distance cutoff – part 1 of 2. Closed and open flap conformations are indicated by purple and black squares, respectively.

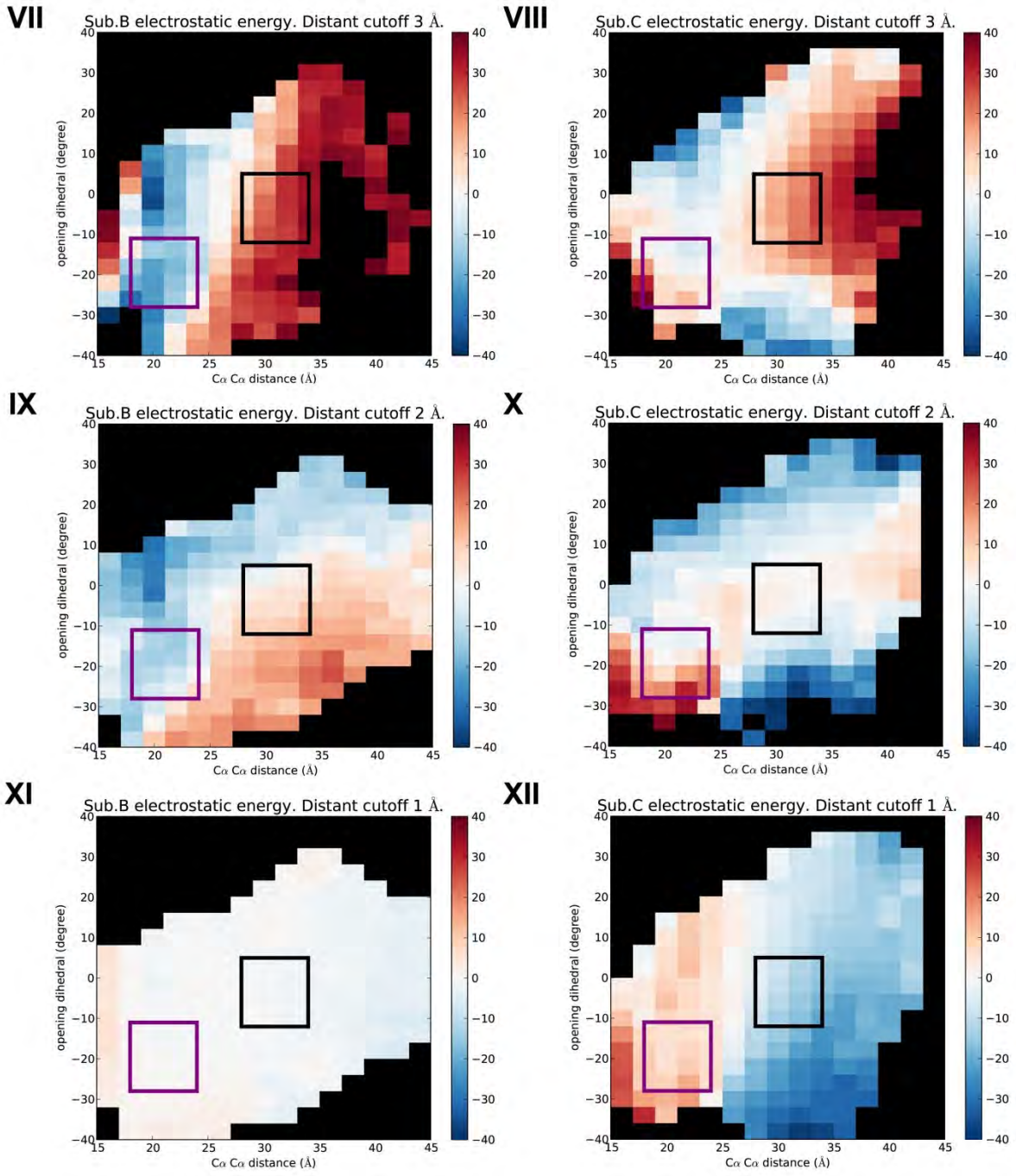

Figure 3-14 Sum of per-residue electrostatic energy, including polymorphisms and nearby residues within certain distance cutoff – part 2 of 2. Closed and open flap conformations are indicated by purple and black squares, respectively.

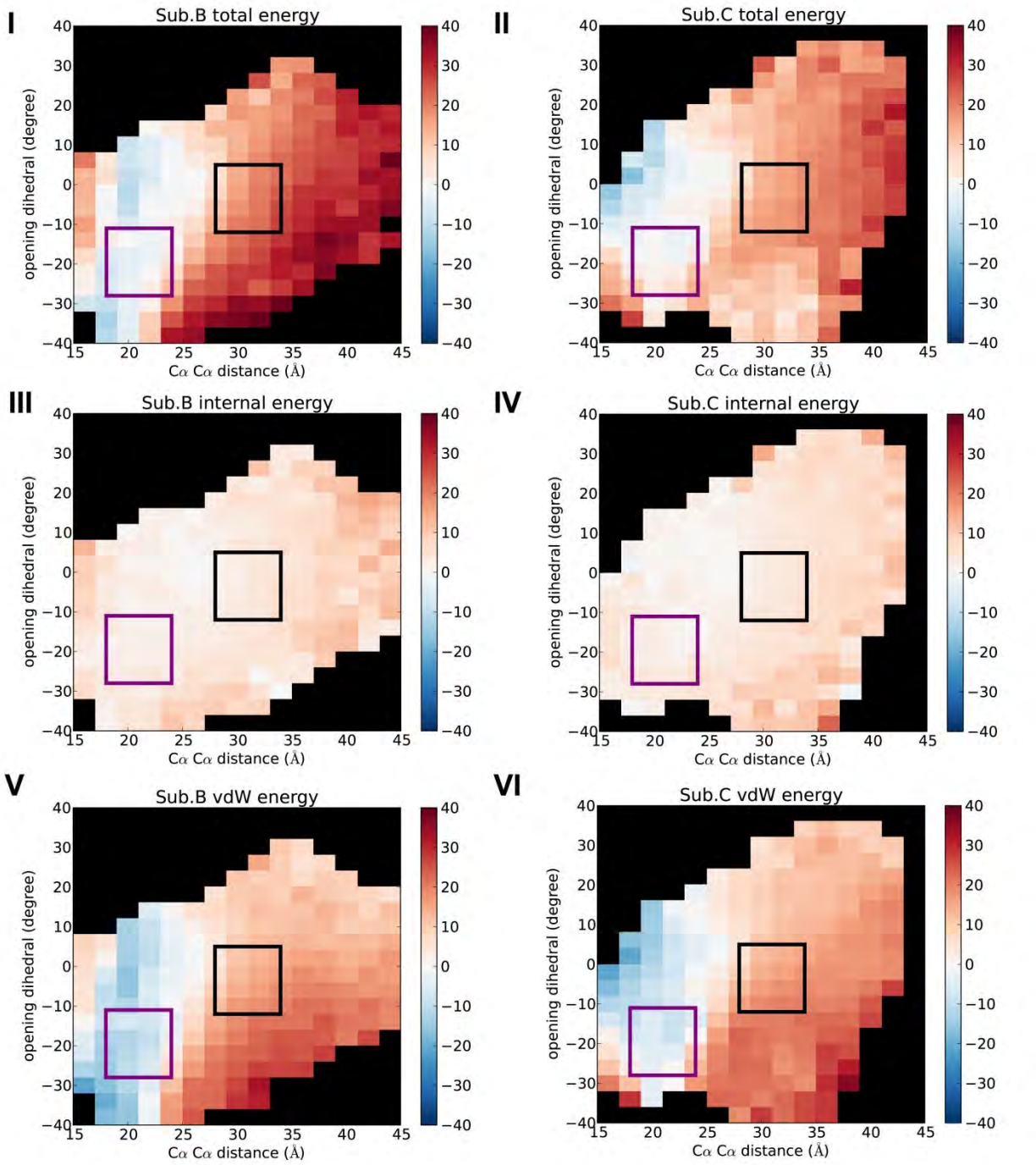

Figure 3-15 Energy decomposition of sub.B and sub.C simulation snapshots – part 1 of 2. Closed and open flap conformations are indicated by purple and black squares, respectively.

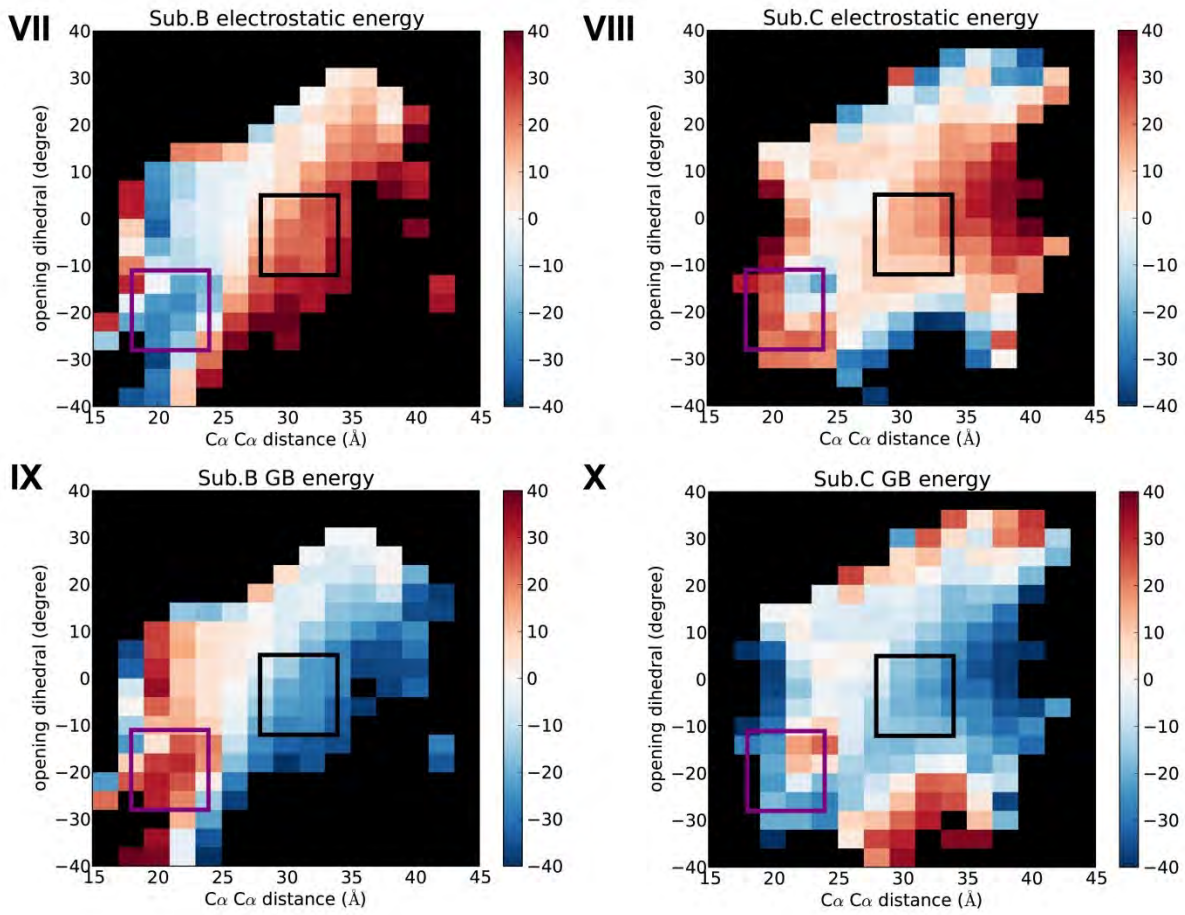

Figure 3-16 Energy decomposition of sub.B and sub.C simulation snapshots – part 2 of 2. Closed and open flap conformations are indicated by purple and black squares, respectively.

3.2.10 Plotting and structural rendering

All data plots were generated using Matplotlib module in python [142]. All structure images in this article were generated with VMD [1].

3.3 Results

3.3.1 Population shift studied by 1D inter-label distance measurement

After performing MD simulations of spin-labeled subtype B and C proteases, we compared simulation results against EPR experimental data to validate our modeling. In both EPR experiment and our MD simulations, MTSL spin labels were added at Cys55 on both monomers (Figure 3-10). For a direct comparison to experimental inter-label distance population, we histogramed inter-label NN distance (N as nitrogen in the nitroxide group) for sub.B and sub.C simulations (Figure 3-17). Consistent with EPR data, the sub.C has population shifted from below 35 Å to above, compared to sub.B. The peak around 40 Å in EPR data was suggested to represent the open flap conformations [67]. Therefore, both EPR and simulation data suggest that subtype C has a larger population of open flap conformation ensemble. We also

noticed that the leftmost peak around 30 Å is not reproduced in the simulation, and simulations have larger population at long distance (around 45 Å) compared to EPR experiment. These could be due to the lack of hydrophobic effects in the implicit solvent model we used. Hydrophobicity, if modeled correctly, would promote the curling of flaps, which is suggested to form the leftmost peak [67], and discourage the more open/extended flap conformations, which correspond to long NN distances. However, we left out the hydrophobic term because of the limitation in current implementations [135-138]. Overall, our modeling reproduced the major trend in experimental data, and we followed up with more analysis below to elucidate the cause of the difference between the two subtypes.

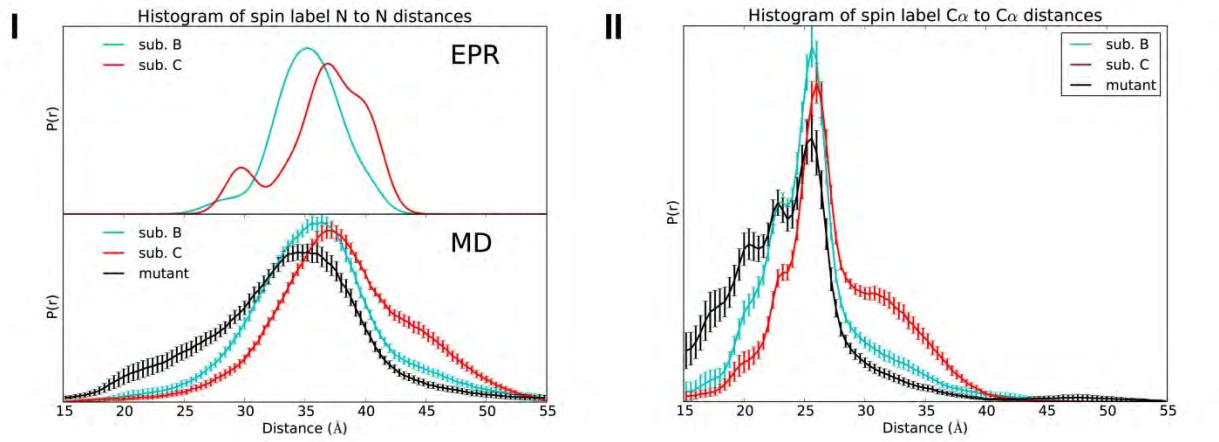

Figure 3-17 Normalized histogram of inter-label nitroxide nitrogen distances compared to EPR data from Kear et al. 2009 JACS paper (I) and inter-label Cys-MTSL CαCα distances (II) measured from simulations.

EPR distance populations are converted from spin-label echo intensity. One advantage of all-atom MD simulation is that coordinates of every atom are recorded throughout the simulation, which can then be processed to retrieve information (distance, torsion, energy, etc.) that is not restricted to part of the structure (spin labels in the case of EPR). We measured inter-label CαCα distance (Cα comes from Cys55 to which the MTSL label is attached, Figure 3-17 II and Figure 3-6 I) to eliminate possible influence from the spin label movement. Compared to NN distance, CαCα measurement demonstrates more fine details. In other words, NN distance distribution curves are smoothed out.

3.3.2 Spin label dynamics

Apart from the smoothing effect, using site-directed spin labeling to study HIVPR flap dynamics relies on the assumption that spin label does not interact with its surroundings, or does so in a consistent way [143], otherwise the population measured would be biased. Here we test this assumption by examining the spin label dynamics during the simulation. The sidechain dihedral histogram of spin labels are shown in Figure 3-18, Figure 3-19, and Figure 3-20. Comparisons were made between different subtypes, different monomers, and different flap conformations. Strikingly, the biggest outlier was found to be the X3 angle of spin labels on subtype C (column three in Figure 3-18, Figure 3-19, and Figure 3-20, compared row three to the other rows), which resulted from sampling of rare X3 dihedral change during the equilibration (before production runs). The time dependence of inter-label NN distance, as well as dihedrals

further confirmed that sampling of X3 is the slowest of all five sidechain dihedrals (Figure 3-21). We then asked if the X3 difference would lead to difference in distance measurement. The NN distances for subtype C protease with semiopen flap conformations are 36.3±3.0 and 36.8±2.9 Å for negative and positive X3 dihedrals, respectively, which are very close on the distance histogram (Figure 3-17 I). Therefore, we concluded that dihedral angle preference of the spin label is not affected by different flap conformations.

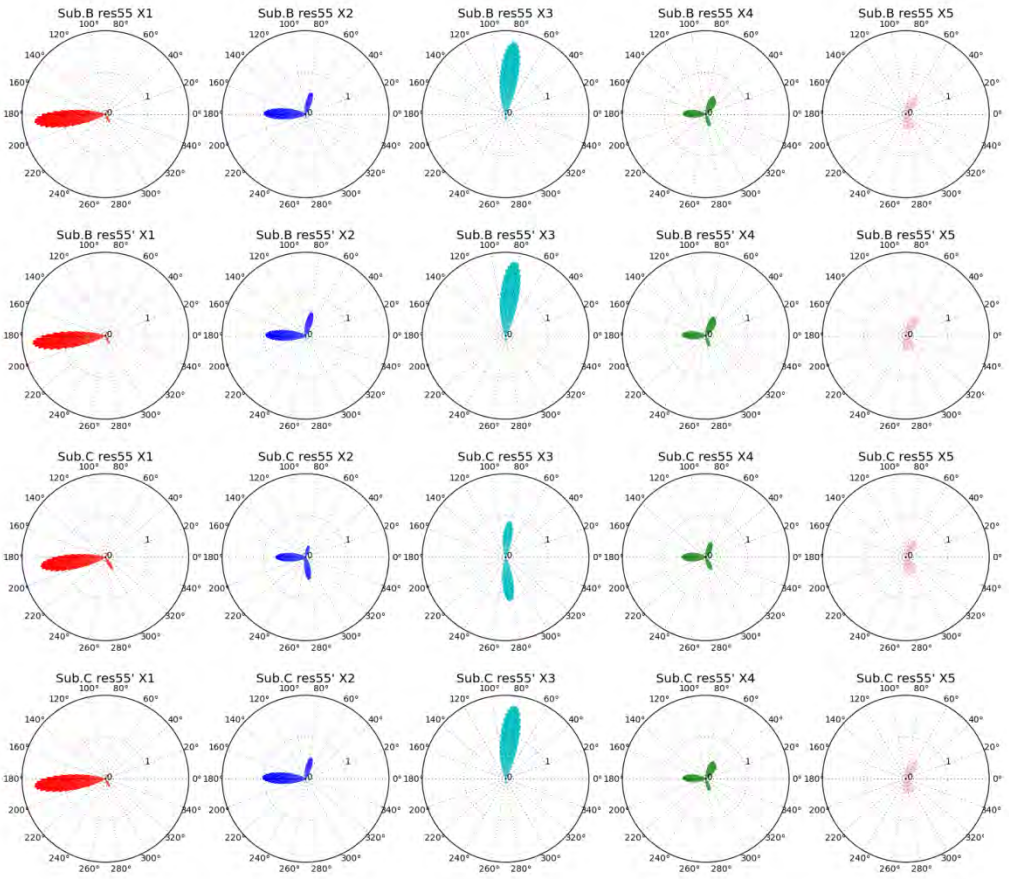

Figure 3-18 Polar plot of spin label sidechain dihedral during subtype B and subtype C simulations. All simulation structures were included in the calculation.

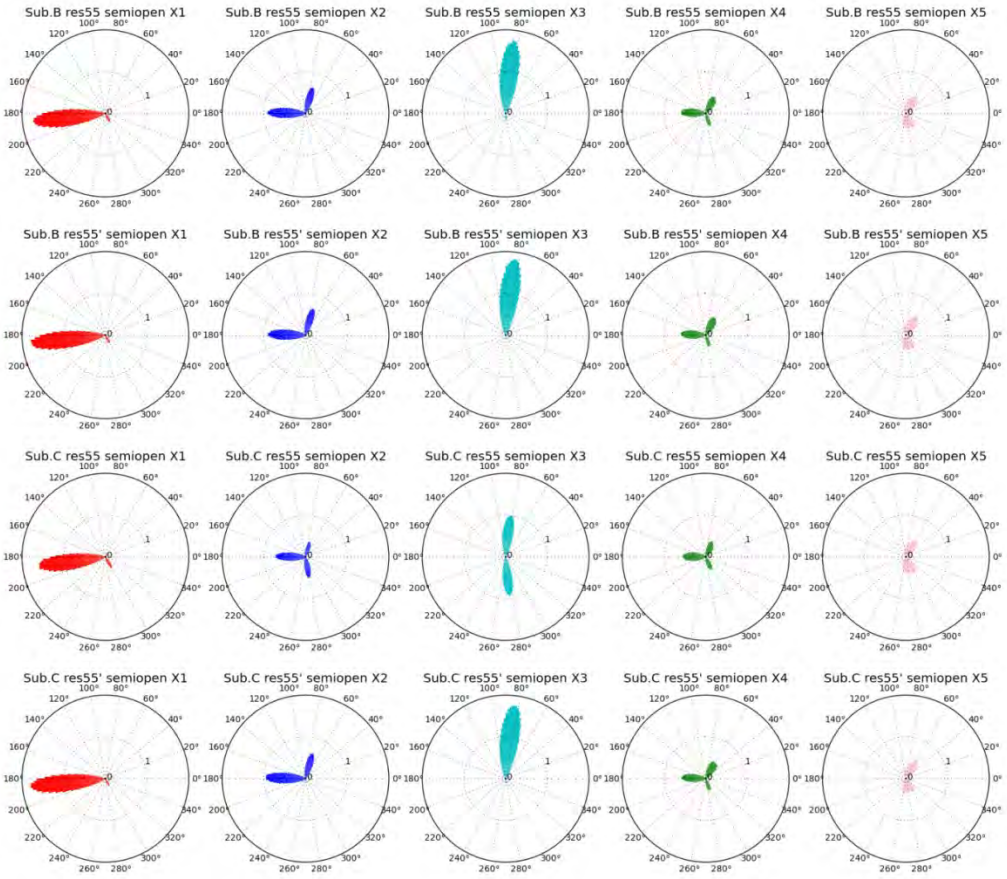

Figure 3-19 Polar plot of spin label sidechain dihedral during subtype B and subtype C simulations. Only simulation structures with semiopen flap conformations were included in the calculation.

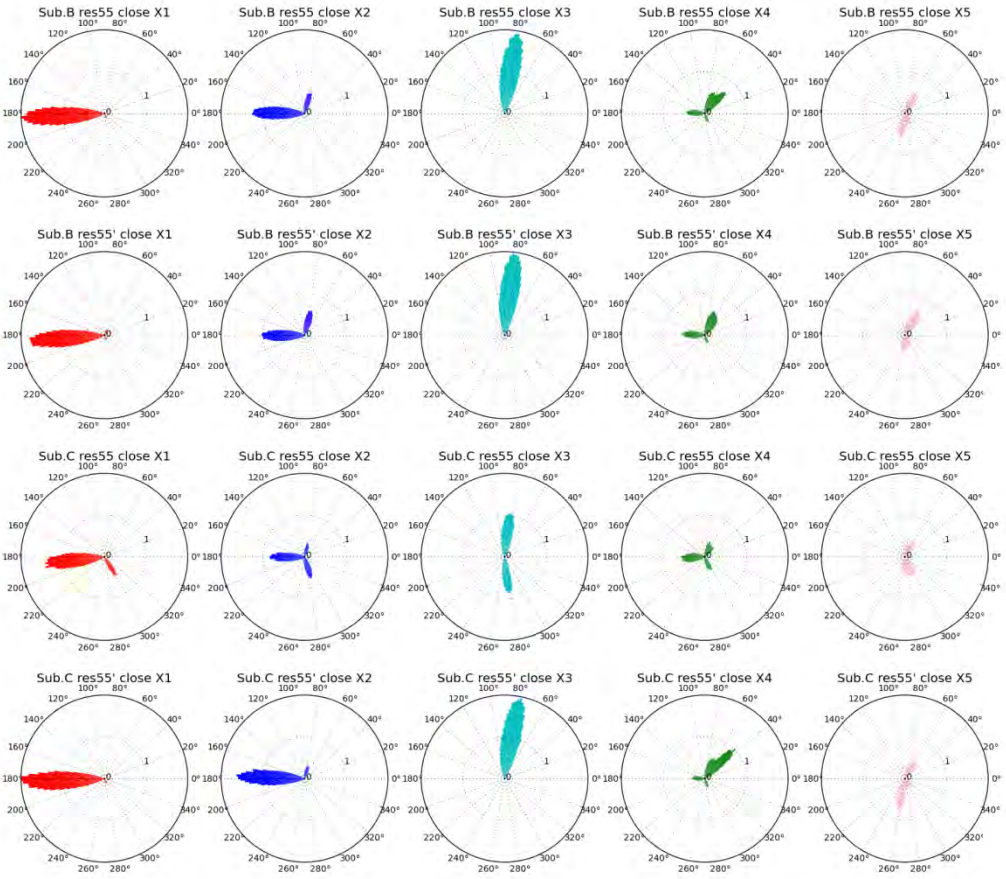

Figure 3-20 Polar plot of spin label sidechain dihedral during subtype B and subtype C simulations. Only snapshots with closed flap conformations were included in the calculation.

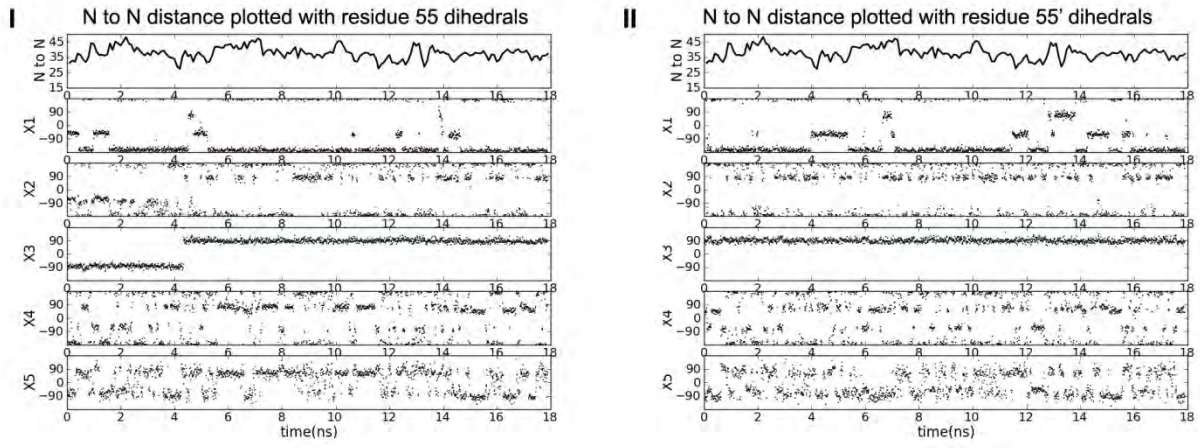

Figure 3-21 For subtype C run 11, the time dependence of NN distance as well as different spin label side chain dihedral angles are plotted out.

However, although the sidechain dihedral distributions are independent of flap conformations, the distribution is less symmetric compared to the MTSL rotamer distribution published earlier, where Polyhach et al. fixed position of the protein backbone and spin label Cβ atom and sampled MTSL sidechain conformations by heating and annealing [144]. For example, their MTSL X1 have an even distribution at 60, 180, and 300 degrees, while X1 prefers 180 degrees in our simulation (Figure 3-18 column one). This could be due to specific local interactions that alter spin label rotamer preference. We calculated the density of the oxygen in nitroxide group, to inspect local interactions (Figure 3-22). Although the space visited by the nitroxide group is large (Figure 3-22 I), there are two constricted locations with significantly higher density, which correspond to spin-label electrostatic interaction with Arg57 (Figure 3-22 II) and vdW interaction with Pro44/Met46 (Figure 3-22 III and IV). These two interactions can form with various χ2/χ3 combinations, but χ1 needs to be near 180 degrees, so local interactions likely result in the χ1 preference observed here. Using 5 Å as cutoff for the spin-label-nitroxide oxygen to Arg57 Cζ atom distance measurement, the hydrogen bond interaction between the two residues is formed about 20% of the time throughout the simulation, for both subtype B and C. Therefore in our case the local interactions do not seem to affect subtype comparison.

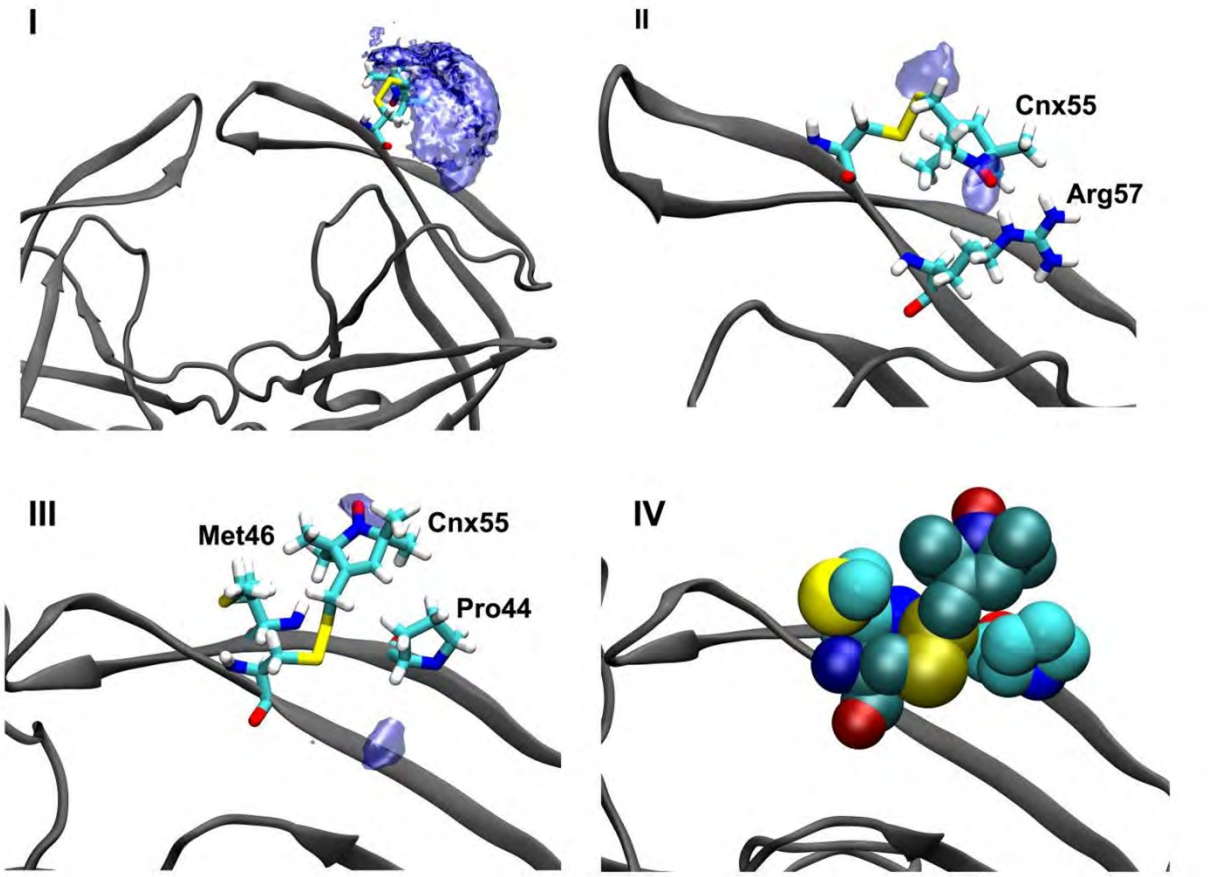

Figure 3-22 Local specific interactions involving spin label residues. Contour surfaces of spin label (nitroxide) oxygen density are shown in transparent blue (low density contour in panel I and high density contour in panel II and III). The two highest density regions correspond to spin-label electrostatic interaction with Arg57 (shown in licorice representation in panel II) and vdW

interaction with Pro44 and Met46 (shown in licorice representation in panel III and in vdW sphere representation in panel IV). Spin label residue is labeled as Cnx55.

3.3.3 Population shift studied by 2D-coordinate system

In order to more unambiguously delineate population peaks and correlate them with flap structure, we defined opening dihedral measurement, in which the flap tip and helix from both monomers (Figure 3-6 II) were used to detect the flap handedness switch. Then, a 2D-coordinate system involving spin label Cα to Cα distance and opening dihedral measurement was used to disperse the 1D distance profile over a 2nd dimension. The population dispersion was then converted to free energy profile for comparison. To give a concrete example of the 2D dispersion, we also mapped simulation starting structures and crystal structures on the map (Figure 3-10 I-II). There are many small differences between sub.B and sub.C free energy profiles (Figure 3-23 I-II), but we focused on the central region since the difference around the edges (less populated) may result from insufficient sampling. More specifically, since the relative energy of semiopen flap conformation (energies are zeroed at the most populated bin, see method section for details) is nearly the same in two subtypes, we focused on the difference in relative population/energy of closed and open flap conformations (in squares on the free energy profile). Experimental 1D distance profile suggests that subtype C disfavors closed flap conformation and favors open flap conformation compared to subtype B, our 2D dispersion profile confirmed this assumption (Figure 3-23 I-V).

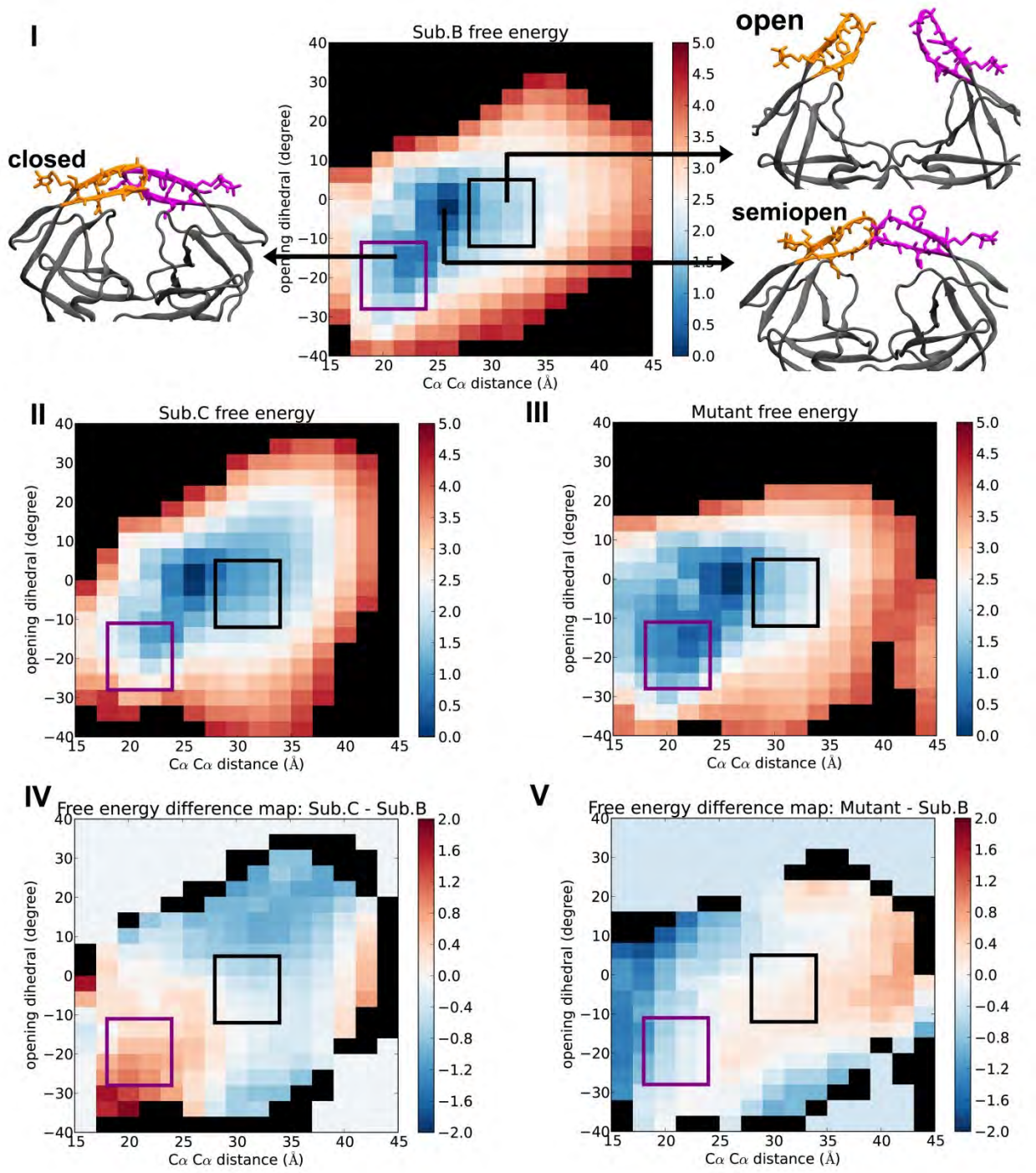

Figure 3-23 Free energy profile of sub.B (I), sub.C (II), and mutant (III), and free energy difference map of sub.C (IV) and sub.C I36M/A37S/K69H triple mutant (V). In sub.B profile (I), the areas corresponding to ensembles of closed, semiopen, and open flap conformations are labeled on the 2D map. Two flaps from monomer A and monomer B are colored in orange and magenta, respectively, and their heavy atoms are shown in licorice. The regions corresponding to closed (purple square) and open (black square) flap conformation peaks are shown in all maps. The regions corresponding to semiopen flap conformation are not squared because their relative

energies within each subtype are the same. The free energy difference of sub.C and sub.B, and the free energy difference of mutant and sub.B, are plotted in panel IV and V, respectively.

3.3.4 Population shift explained by energy decomposition

After demonstrating the population shift on a 2D map, we wanted to pinpoint key residues responsible for the shift. We expected the changes to come from the polymorphisms, since sub.B and sub.C simulations only differ in these locations (see method section for details). But we included all residues in our analysis for a more systematic approach. We analyzed each residue's contribution by performing energy decomposition on simulation, and then identified key residues through filtering out less relevant ones. It's worth noting that, due to noise from thermal fluctuation and unconverged conformational sampling, using energy decomposition to identify key residues has always been challenging [141, 145]. Below, we would first validate that the energy profile from energy decomposition reproduces the trend shown in free energy profile (converted from and thus equivalent to population profile), and then pinpoint residues contributing most to the profile difference between two subtypes.

We performed energy calculation on snapshots of sub.B and sub.C simulations (about 1 million structures each) and mapped them onto the same 2D coordinates used for free energy profiles (Figure 3-15 and Figure 3-16). The total energy terms calculated include internal (bond, angle, and dihedral), vdW, electrostatic, and GB solvation energy. The mapped total energy profiles (Figure 3-15 I-II) match well with the free energy profiles (Figure 3-23 I-II): subtype B has lower relative energy in closed flap structures, which are more populated, and higher relative energy in open flap structures, which are less populated. The total energy trend will not match the free energy trend exactly, because factors like entropy were not included in the calculation. Generally, structures with small Cα to Cα distance indicates collapsed flaps and less entropy, while structures with large Cα to Cα distance means detached flaps and more entropy. But here we assume entropy acts similarly on different systems since they share the same flap sequence. Looking at energy components (Figure 3-15 and Figure 3-16 III-X), relative internal energies are essentially the same for the two subtypes, relative vdW energies are the same for open flap conformations but have some difference in closed flap conformations, and relative electrostatic energies match very well with the total energy trend although they are offset by GB solvation energies. Therefore, we decided to focus on vdW and electrostatic energies for further analysis.

The good match between the molecular potential energy profile and the free energy profile ensures that we could find key residues, which are responsible for the population shift, by decomposing the potential energy into per-residue contribution. We asked whether polymorphisms and their neighbors cause most of the population shift. By filtering per-residue energies with different distance criteria, we confirmed that the energy sum of polymorphisms and their neighbors within 2Å, which includes 24 residues out of 198 residues, gives a decent match to the free energy profile (Figure 3-14 IX and X compared to Figure 3-23 I and II, see the methods section for details). These 24 residues were ranked (Table 3-1 and Table 3-2) which showed that subtype B has favorable energy for closed structures mainly due to residue 69 and 35, contributing -3.7 and -3.4 kcal/mol, respectively, and subtype B has less favorable energy for open structures mainly due to residue 69 and 88, contributing 5.6 and 4.6 kcal/mol, respectively. We then investigated these three residues' contribution by ranking per-residue-pair energies among the three polymorphisms (35/69/88) and their neighbors (within 5Å distance cutoff) in subtype C simulations. From the 2D map of pairs picked out, interactions near residue 35 and 69

led to big change in the relative energy of closed and open flap conformations, interactions near residue 88 are not as distinguishing (Figure 3-24). Therefore, we examined simulation trajectories to determine the mechanism through which change at residue 35 and 69 can lead to difference in flap dynamics.

Residue	Subtype B	Subtype C	Subtype B – subtype C
69	**0.1 +/- 0.0**	**3.8 +/- 0.3**	**-3.7**
35	**-13.9 +/- 0.7**	**-10.5 +/- 1.9**	**-3.4**
88	-0.6 +/- 0.1	1.2 +/- 0.4	-1.8
14	-1.1 +/- 0.6	0.2 +/- 0.6	-1.3
15	0.0 +/- 0.1	0.2 +/- 0.1	-0.2
20	3.2 +/- 0.5	3.3 +/- 0.7	-0.1
37	-0.1 +/- 0.1	-0.0 +/- 0.0	-0.1
11	-0.1 +/- 0.0	-0.1 +/- 0.1	0.0
13	0.0 +/- 0.0	-0.0 +/- 0.1	0.0
16	0.0 +/- 0.0	0.0 +/- 0.0	0.0
19	-0.0 +/- 0.0	-0.1 +/- 0.1	0.0
38	-0.1 +/- 0.0	-0.1 +/- 0.1	-0.0
66	-0.0 +/- 0.0	0.0 +/- 0.1	-0.0
68	0.0 +/- 0.0	0.0 +/- 0.0	0.0
90	0.0 +/- 0.1	-0.0 +/- 0.0	0.0
94	0.0 +/- 0.0	0.0 +/- 0.0	0.0
12	0.1 +/- 0.1	-0.1 +/- 0.0	0.1
36	0.1 +/- 0.1	-0.0 +/- 0.0	0.1
89	0.0 +/- 0.0	-0.1 +/- 0.2	0.1
93	0.2 +/- 0.0	-0.0 +/- 0.1	0.2
18	0.1 +/- 0.1	-0.2 +/- 0.1	0.3
92	0.4 +/- 0.5	-0.2 +/- 0.2	0.6
70	0.1 +/- 0.1	-0.9 +/- 0.6	1.0
87	1.4 +/- 0.2	-0.1 +/- 1.0	1.5

Table 3-1 Per-residue (relative) electrostatic energy. From left to right: residue number, relative energy of subtype B protease dimer with closed flap conformation (square [18, 24, -28, -11]) relative to semiopen conformation (square [25, 27, -10, 5]), relative energy of subtype C closed conformation relative to semiopen conformation, subtype B relative energy minus subtype C relative energy. Error bars calculated as the difference between two monomers. Residues are ranked by the last column. Important residues picked out are shown in bold.

Residue	Subtype B	Subtype C	Subtype B – subtype C
87	-4.6 +/- 0.6	-2.7 +/- 0.4	-1.9
92	2.1 +/- 0.2	3.1 +/- 0.7	-0.9
35	7.4 +/- 2.2	8.0 +/- 0.6	-0.6
14	1.1 +/- 0.2	1.5 +/- 0.1	-0.4
89	-0.0 +/- 0.0	0.2 +/- 0.1	-0.3
18	0.2 +/- 0.0	0.3 +/- 0.0	-0.2
37	-0.0 +/- 0.1	0.0 +/- 0.0	-0.1
38	0.1 +/- 0.0	0.2 +/- 0.0	-0.1
90	0.0 +/- 0.0	0.1 +/- 0.0	-0.1
16	0.0 +/- 0.0	0.0 +/- 0.0	0.0
36	-0.1 +/- 0.0	-0.0 +/- 0.1	-0.0
66	0.2 +/- 0.0	0.3 +/- 0.0	-0.0
68	0.0 +/- 0.0	0.0 +/- 0.0	0.0
93	0.1 +/- 0.0	0.1 +/- 0.0	0.0
94	0.0 +/- 0.0	0.0 +/- 0.0	0.0
11	0.2 +/- 0.1	0.2 +/- 0.0	0.1
12	0.0 +/- 0.0	-0.1 +/- 0.1	0.1
13	0.2 +/- 0.1	0.1 +/- 0.0	0.1
15	0.1 +/- 0.1	-0.0 +/- 0.1	0.1
19	0.1 +/- 0.0	0.1 +/- 0.0	0.1
20	-0.0 +/- 0.7	-0.5 +/- 0.2	0.5
70	3.0 +/- 0.6	1.4 +/- 0.2	1.7
88	**0.0 +/- 0.2**	**-4.7 +/- 2.0**	**4.7**
69	**-0.1 +/- 0.1**	**-5.6 +/- 0.3**	**5.6**

Table 3-2 Per-residue (relative) electrostatic energy. From left to right: residue number, relative energy of subtype B protease dimer with open flap conformation (square [28, 34, -12, -5]) relative to semiopen conformation (square [25, 27, -10, -5]), relative energy of subtype C closed conformation relative to semiopen conformation, subtype B relative energy minus subtype C relative energy. Error bars calculated as the difference between two monomers. Residues are ranked by the last column. Important residues picked out are shown in bold.

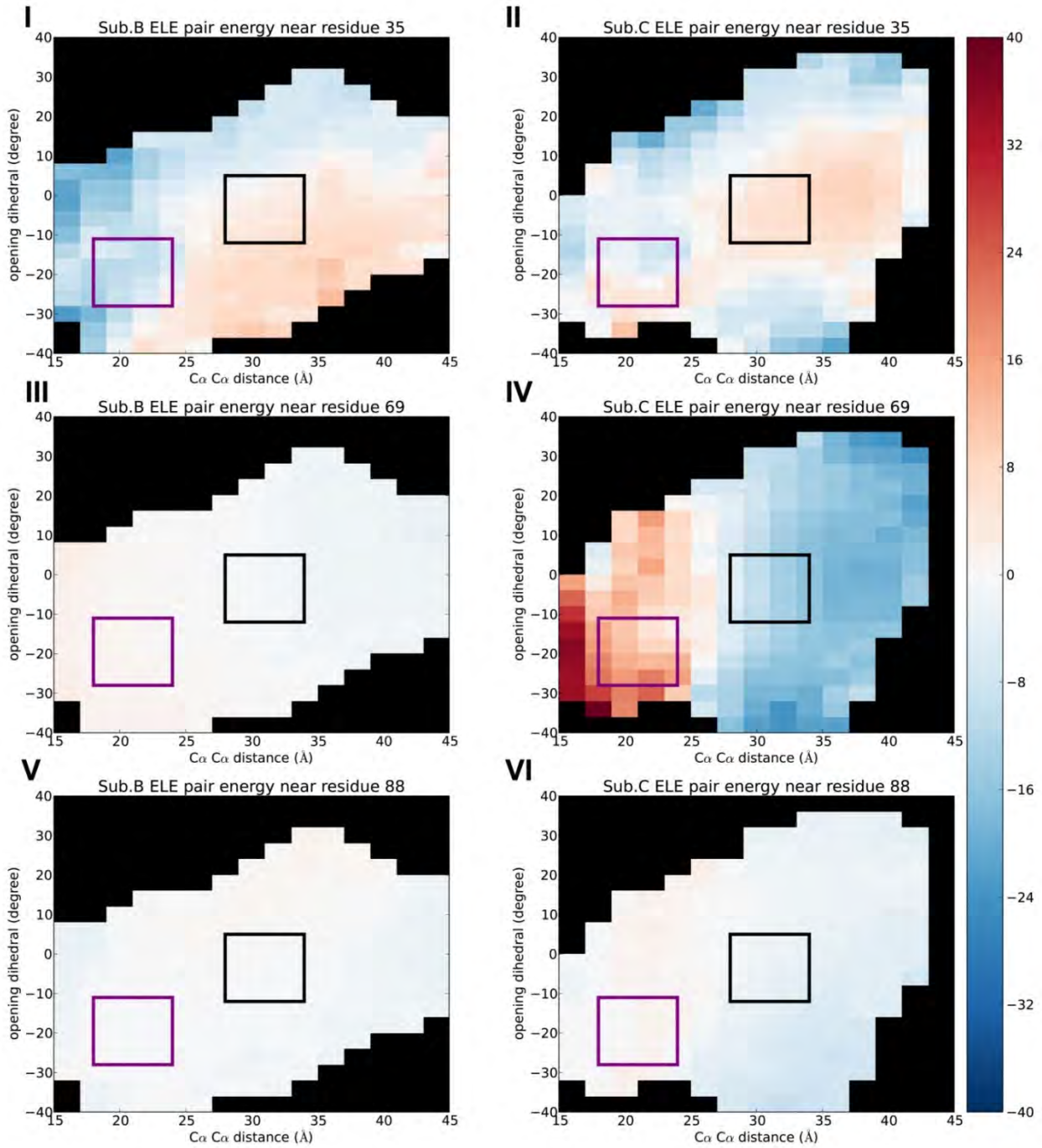

Figure 3-24 Electrostatic pair energy near residue 35, 69, and 88. I-II) Electrostatic energy of pair 35-57. III-IV) Electrostatic energy of pair 69-93 and 69-99'. V-VI) Electrostatic energy of pair 30-88 and 31-88. Closed and open flap conformations are indicated by purple and black squares, respectively.

Comparing subtype B and C with closed flap conformation we found that M36I and I15V polymorphisms likely cause the disruption of vdW packing among residue 36, fulcrum region (residue 10 to 23), and other branched residues in the back (Ile33, Leu38, etc. in Figure 3-25 I-II). This disruption, along with the S37A polymorphism, worsens the hydrogen bonding network involving Arg57 and Glu35 (Figure 3-25 I-II). The decoupling between fulcrum and elbow (residue 37 to 42) due to vdW packing disruption, and the decoupling between elbow and flap region due to hydrogen bond disturbance, likely promotes the flap motion and transition to semiopen and open flap conformations. The open conformations feature tighter packing between the fulcrum and elbow, and have more favorable entropy at flap tips.

Comparing subtype B and C with open flap conformation (Figure 3-25 III-IV), the superior ability of subtype C protease to form salt bridge between residue 69 and termini residues is quite obvious. Because of the H69K polymorphism, Lys69 in subtype C protease has longer and charged sidechain to optimize its hydrogen bonding with terminal residue Phe99 on the opposite monomer, and Leu83 on the same monomer. These hydrogen bonds in turn bring the cantilever (residue 59 to 75) closer to the termini region, which stabilizes the open flap conformation. In contrast, in our simulations of subtype B with either neutral or protonated His69, the hydrogen bond is much weaker.

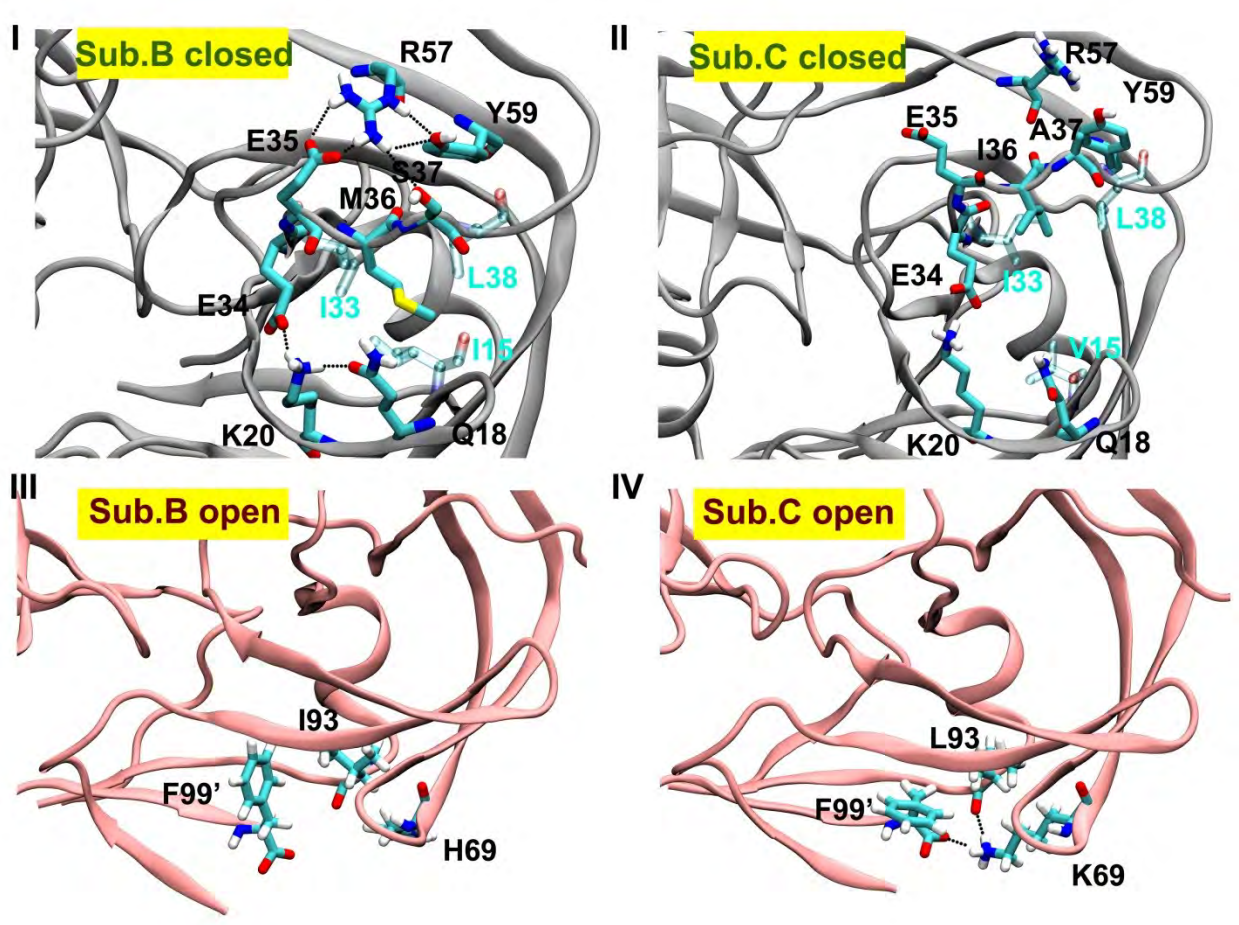

Figure 3-25 Simulation snapshots of subtype B (I and III) and subtype C (II and IV) with either closed flap conformation (protein backbone in grey, I and II) or open flap conformation (protein

backbone in pink, III and IV). Residues are shown in licorice representation and labeled with black text. Residues in the back are rendered as transparent with cyan label to facilitate visualization. Hydrogen bonds are indicated with black dotted lines.

Histidine has a pKa value near physiological pH. We had modeled His69 (the only Histidine in sub.B sequence) as neutral. However, H69 may become protonated when it comes closer to the termini region. Therefore, we also performed subtype B simulation with protonated His69 to investigate the influence of protonation/size. Both 1D distance measurement (Figure 3-26 I and II) and 2D free energy map (Figure 3-26 III and IV) show that the influence is negligible.

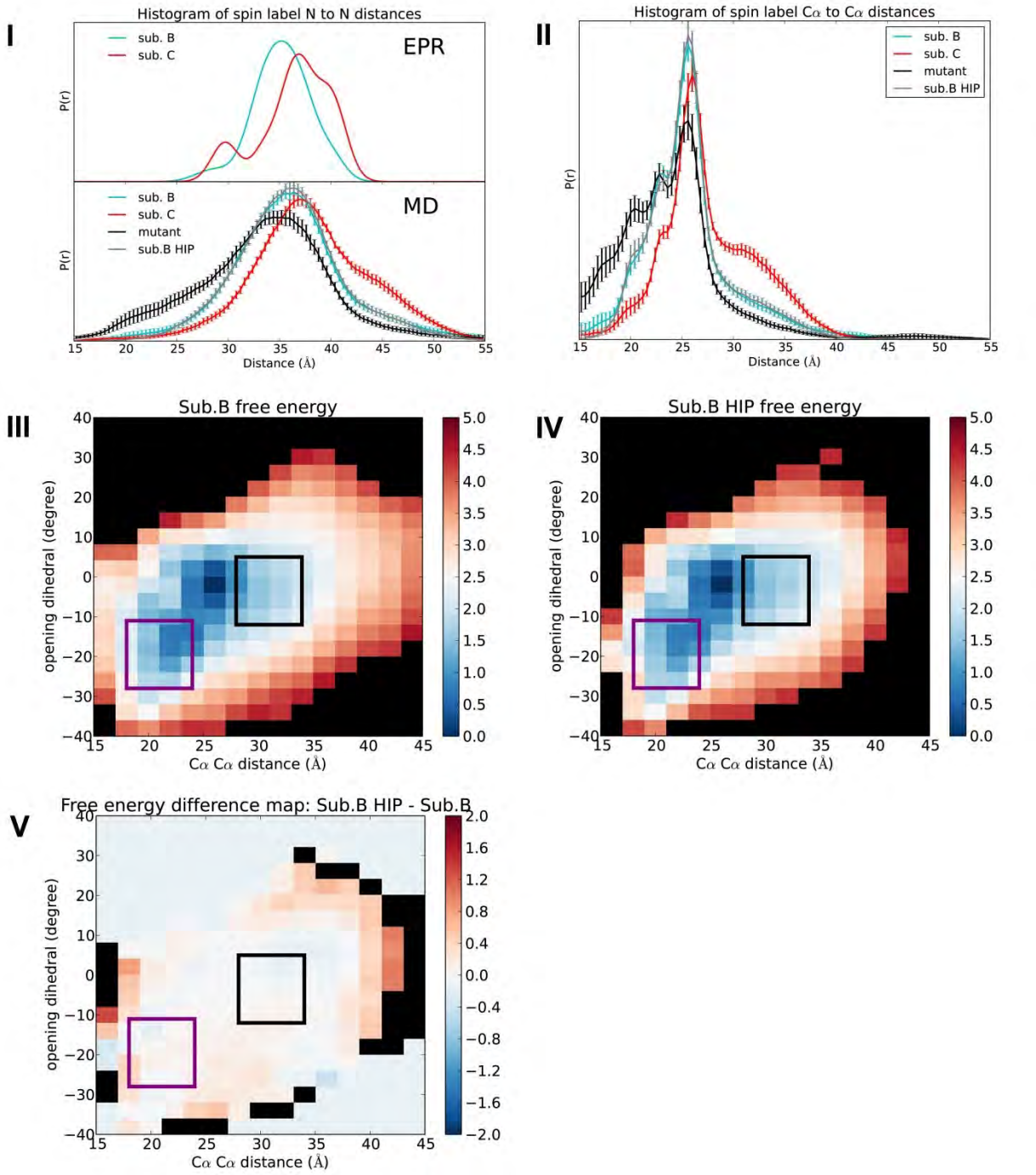

Figure 3-26 Normalized histogram of inter-label nitroxide nitrogen distances compared to EPR data from Kear et al. 2009 JACS paper (I) and inter-label Cys-MTSL CαCα distances (II) measured from simulations. Error bars were calculated as standard error of the mean (SEM) of independent runs. Free energy profile of sub.B HIP (IV) compared to sub.B (III). Closed and open flap conformations are indicated by purple and black squares, respectively. The free energy difference of sub.B HIP and sub.B is plotted in panel V.

3.3.5 Key polymorphisms validated and linked to drug resistant mutations

Since we included assumptions and simplifications in our study (potential energy profile does not match free energy profile perfectly, we did not study solvation energy in details, etc.), it is possible we have left out important information. In order to validate our hypothesis, we simulated subtype C mutant where three polymorphisms were back-mutated to subtype B sequences: M36/S37/I15. We expected the mutant simulation to give similar profile as subtype B, since the three residues were proposed to cause the predominant difference in flap conformations. Indeed, both the 1D distance measurement (Figure 3-17) and the 2D free energy map (Figure 3-23) show that the mutant has more population with closed flap conformation and less population with open flap conformation, just like subtype B.

Both M36 and H69 are susceptible to drug resistance associated mutation. Commonly found mutations are M36I, M36L, M36V, H69K and H69R. The resistance strategy at each location seems well defined: 1), all three mutations at position 36 are transforming methionine into a shorter, branched, and non-polar residue, so that the packing with surrounding branched residues is hampered, and 2), both mutations at position 69 are transforming Histidine into a base with a longer sidechain, so that the hydrogen bonding with the termini is facilitated. Putting it all together, these mutations work synergically to shift the flap conformation population from closed to semiopen and open. Since existing drugs inhibiting HIVPR are all targeting the closed conformation, it is conceivable that a protease less prone to close the flaps is less likely to bind drug well (it needs to overcome greater strain energy to accommodate the binding). Moreover, like the binding pathway studies suggested [51, 52]. due to the size difference between commercial drugs and polypeptide chains (as natural substrate of retroviral proteases), the fast tumbling of drugs could expediate their release when the flaps are not properly closed, while the long substrate could be kept in as long as the flaps are not fully open.

3.4 Conclusions

In this study, we performed MD simulations of spin-labeled HIVPR subtype B and C, to provide atomic explanation of the flap-conformation preference change previously suggested by EPR data. We combined batch simulation technique and implicit solvent model (previously validated on HIVPR [130]) to speed up conformational sampling, and used energy decomposition to propose key polymorphisms (M36I, S37A, and H69K) for the preference change. Structural analysis revealed that these polymorphisms work synergically to shift the flap conformation population from closed to open, which seems tightly related to the drug resistance strategy at position 36 and 69: the equilibrium shift results in an expanded binding pocket that may loosen the binding of small drugs but retain long substrate chain. We validated our hypothesis by additional batch simulations of mutant strains, and predicted EPR curves suitable for future experimental comparison. We also examined the local interactions of spin labels with HIVPR residues, which is hard to probe with experiments. Our results suggest that spin label sidechain rotation smoothes out EPR curves, and spin labels do have specific non-bonded interactions with nearby residues. However, their dihedral preference is not affected by different flap conformations, nor do the subtype spectral differences arise from conformational differences in the labels. Overall, by collaborating with experimentalists and providing insights into the dynamics and energetics of polymorphisms/drug resistance, we plan to design next generation

HIVPR inhibitors with higher efficacy towards non-B subtypes and multi-drug resistant strains in the future.

Chapter 4 HIV-1 protease multi-drug resistance studied by binding affinity calculations and communication network analysis

4.1 Introduction

Current anti-AIDS drugs include those inhibiting viral entry and those inhibiting key enzymes in HIV life cycle. The protease inhibitors were among the first anti-AIDS drugs to be marketed, and helped to turn AIDS from a death sentence to a chronic disease. However, soon after the anti-AIDS drugs were released to the market, the drug resistance problem emerged. Drugs were no longer as effective as when they were first taken, which called for continuous development of new AIDS inhibitors. The drug resistance problem is particularly obstinate in the case of HIV because of its heterogeneous genome.

To date, nine drugs targeting HIV protease have been approved by FDA (Table 4-1). Their inhibitors are larger than average drugs on the market, each having around 100 atoms and 15 rotatable bonds (Figure 4-1). They also differ in the entropy-enthalpy composition of their binding free energies to the protease. The binding of earlier protease inhibitors are usually entropy driven, while the binding of more recent inhibitors have significant gain in the enthalpy component.

Brand Name	Generic Name	Manufacturer Name	Approval Date
Invirase	Saquinavir(SQV) mesylate	Hoffmann-La Roche	6-Dec-95
Norvir	Ritonavir (RTV)	Abbott Laboratories	1-Mar-96
Crixivan	Indinavir (IDV)	Merck	13-Mar-96
Viracept	Nelfinavir (NFV) mesylate	Agouron Pharmaceuticals	14-Mar-97
Fortovase	Saquinavir (no longer marketed)	Hoffmann-La Roche	7-Nov-97
Agenerase	Amprenavir (APV)	GlaxoSmithKline	15-Apr-99
Kaletra	Lopinavir and ritonavir (LPV/RTV)	Abbott Laboratories	15-Sep-00
Reyataz	Atazanavir (ATV) sulfate	Bristol-Myers Squibb	20-Jun-03
Lexiva	Fosamprenavir (FOS-APV) Calcium	GlaxoSmithKline	20-Oct-03
Aptivus	Tipranavir (TPV)	Boehringer Ingelheim	22-Jun-05
Prezista	Darunavir (DRV)	Tibotec, Inc.	23-Jun-06

Table 4-1 Information on HIV protease drugs proved by FDA.

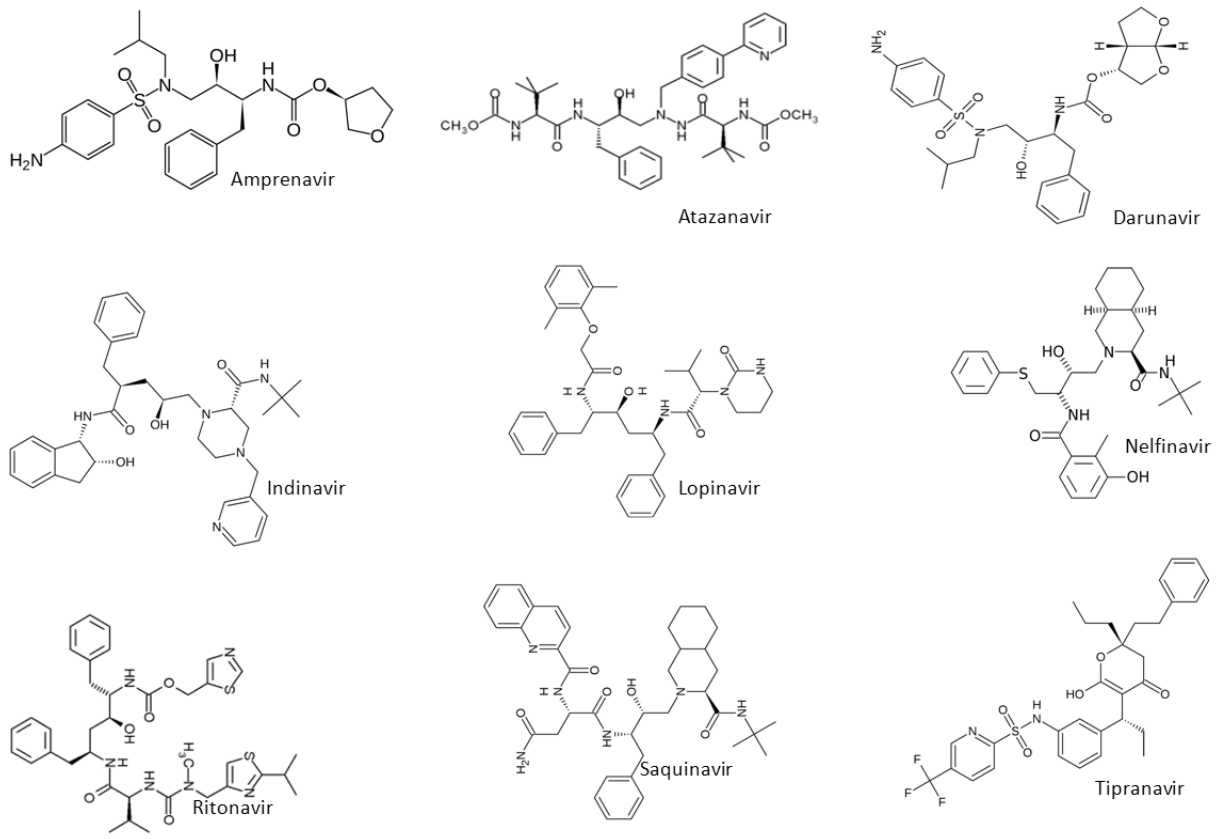

Figure 4-1 2D representation of HIV protease inhibitors.

To design new protease inhibitors that remain efficient upon drug resistant mutations, we aimed to understand the mechanisms of drug resistance. We are mainly interested in two problems: 1) how does certain HIVPR drug respond to drug resistance mutations, and 2) how the responses differ in different protease drugs. All-atom MD simulation is a good method for answering these questions, because the method has been improved over several decades through tuning parameters against experiments [64, 70, 77, 84], and was applied to study protein-ligand interactions over years due to its high resolution in time (femtosecond time scale) and space (atomic resolution) [146-148]. Using MD simulation as a tool, we studied the binding free energy change of protease-inhibitor complex due to drug resistance mutations. The drug resistance mutations involve those near the active site, and those distal from the active site (Figure 3-2). We first used thermodynamic integration (TI) method to study the binding free energy change upon active-site drug-resistance mutation. Then, since we are more interested in the drug resistance mutations distal from the active site, whose influence to the binding is less clear, we compared the response of first generation HIVPR drug, second generation inhibitor drug, and natural substrate upon multi-drug-resistance mutations. It was suggested previously that second generation drugs perform better upon multi-drug-resistance mutations than the first generation drugs [149], we expected to see that first generation drug loses more binding affinity than the second generation drug, while the natural substrate is less or equally affected by the mutations compared to the second generation drug. The goal of this study is to elucidate the role of distal mutations from the active site that cause drug resistance, and then suggest ways to

improve the efficacy of current inhibitors accordingly. Such information is not readily available from experimental data.

4.2 Methods

4.2.1 Thermodynamic integration

Thermodynamic integration (TI) is a way to calculate the free energy difference between two states by sampling configurations of the intermediate states, and MD simulation is usually used for the configuration sampling. Given two states A and B, with potential energy U_A and U_B, we could create a new potential energy function defined as:

$$U(\lambda) = U_A + \lambda\,(U_B - U_A)$$

Equation 4-1 Potential energy function in thermodynamic integration.

Here, λ is a coupling factor between 0 and 1, and thus the potential energy is a function of λ. When λ is 0 the potential energy is the same as state A, when λ is 1 the potential energy is the same as state B, and when λ is a value between 0 and 1 the potential energy becomes an intermediate as well.

In canonical ensemble, where N, V, and T are constant, the partition function of the system can be written as:

$$Q(N,V,T,\lambda) = \sum_s exp[-U_s(\lambda)/kT]$$

Equation 4-2 Partition function in thermodynamic integration.

Then we could derive the change in free energy by taking its derivative with respect to λ and we get:

$$\Delta F(A \rightarrow B) = \int_0^1 d\lambda\, \frac{\partial F(\lambda)}{\partial \lambda} = -\int_0^1 d\lambda\, \frac{k_B T}{Q}\frac{\partial Q}{\partial \lambda}$$
$$= \int_0^1 d\lambda\, \frac{k_B T}{Q}\sum_s \frac{1}{k_B T} e^{-\frac{U_s(\lambda)}{k_B T}}\, \frac{\partial U(\lambda)}{\partial \lambda} = \int_0^1 d\lambda\, \langle \frac{\partial U(\lambda)}{\partial \lambda} \rangle_\lambda$$

Equation 4-3 Derivation of thermodynamic integration.

So the change in free energy between state A and B can be computed by integrating the ensemble average for the change in potential energy as a function of parameter λ. In practice, a potential energy function linking state A and B is defined first. Then a series of MD simulations is performed, each sampling structures of one particular point on the path between A and B. Then the derivative of the potential energy in respect to parameter λ is calculated for each point along the path. At last the integration is performed to obtain the free energy change between the two states.

Commonly, TI is used in combination with a thermodynamic cycle. For example, when calculating the binding free energy difference between wild type protease and mutant protease to the same inhibitor, we could construct a thermodynamic cycle linking the states of apo and bound proteases (Figure 4-2). The advantage of such thermodynamic cycle is that using the Equation 4-4, we could calculate the binding free energy difference using the alchemical/imaginary pathways (blue arrows in Figure 4-2), which are only achievable through simulations but gives the same energy difference as if we simulated the more difficult binding processes (green arrows in Figure 4-2 that require simulating the inhibitor coming from extremely far away).

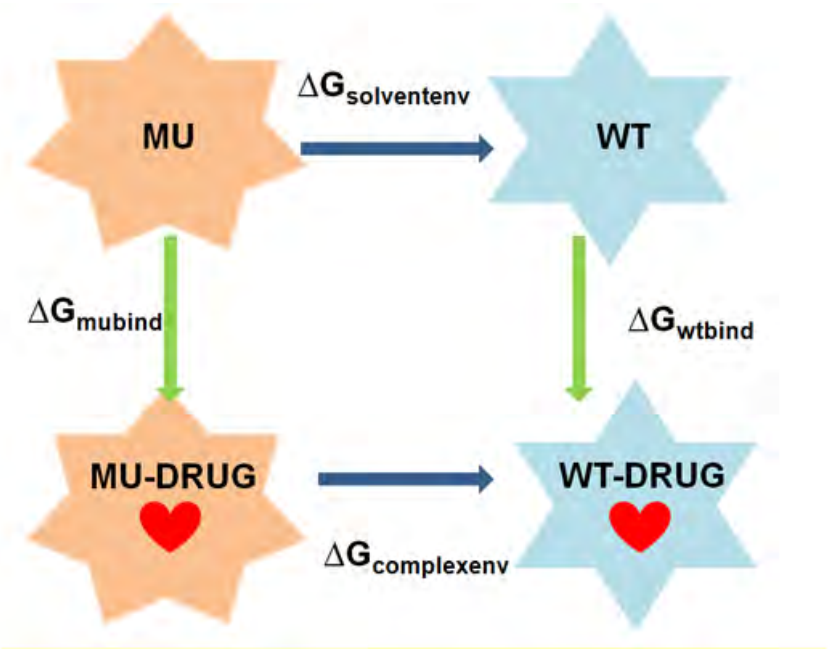

Figure 4-2 A diagram illustrating the thermodynamic cycle of thermodynamic integration for protease-inhibitor binding. MU: mutant protease. WT: wild type protease. MU-DRUG: mutant protease bound to inhibitor. WT-DRUG: wild type protease bound to the same inhibitor. ΔGsolventenv: free energy difference between MU and WT in solvent environment where the active site is exposed to solvent. ΔGcomplexenv: free energy difference between MU-DRUG and WT-DRUG in complex environment where the active site is bound with the ligand (complex solvated in solvent). ΔGmubind: free energy difference between apo and holo mutant protease in solvent. ΔGwtbind: free energy difference between apo and holo wild type protease in solvent.

$$\Delta G_{complexenv} - \Delta G_{solventenv} = \Delta G_{wtbind} - \Delta G_{mubind}$$

Equation 4-4 Equation of the thermodynamic cycle of thermodynamic integration for protease-inhibitor binding.

4.2.2 Molecular Mechanics Poisson-Boltzmann surface area (MMPBSA)

Apart from TI, molecular mechanics Poisson-Boltzmann surface area (MMPBSA) is another way to calculate binding affinity. Similar to TI, the theory of MMPBSA could also be illustrated by a thermodynamic cycle, but the cycle would involve binding in two phases: solvent environment and vacuum environment (Figure 4-3). According to the thermodynamic cycle constructed, we could calculate the binding free energy of the ligand and the protein in solvent environment using Equation 4-5:

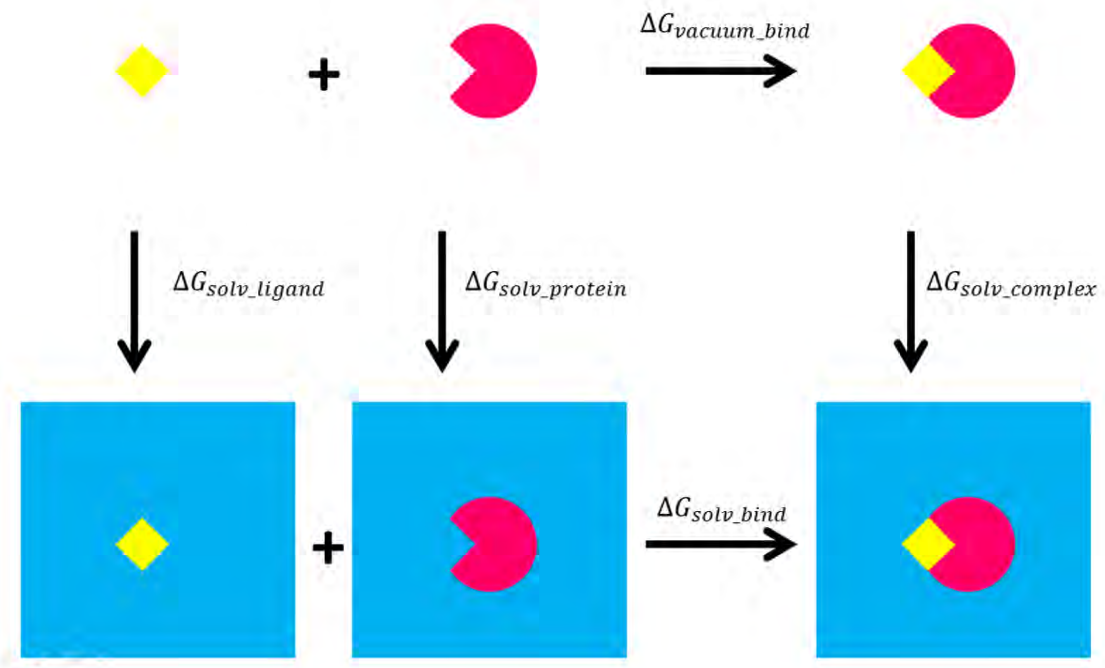

Figure 4-3 Diagram illustrating the thermodynamic cycle of MMPBSA for protein-ligand binding. Yellow square represents the ligand, while the pink pie represents the protein. ΔGvacuum_bind: binding free energy in vacuum. ΔGvacuum_bind: binding free energy in vacuum. ΔGsolv_bind: binding free energy in solvent. ΔGsolv_ligand: solvation free energy of the ligand. ΔGsolv_protein: solvation free energy of the protein. ΔGsolv_complex: solvation free energy of the complex.

$$\Delta G_{solv\_bind} = \Delta G_{vacuum\_bind} + \Delta G_{solv\_complex} - \left(\Delta G_{solv\_ligand} + \Delta G_{solv\_protein}\right)$$

Equation 4-5 Equation of the thermodynamic cycle of MMPBSA for protease-ligand binding.

The terms on the right side of the Equation 4-5 could be calculated accordingly using MMPBSA. More specifically, ΔGvacuum_bind is calculated by molecular mechanics (MM), while the polar and nonpolar components of the solvation energies (ΔGsolv_ligand, ΔGsolv_protein, and

ΔGsolv_complex) are calculated by Poison-Boltzmann (PB) and surface area (SA) terms, respectively. Since these calculations don't involve entropy components, an entropy term could be added using methods such quasi-harmonic analysis [150], if necessary.

The binding free energy examined from MMPBSA is the absolute energy of protein-ligand binding, rather than the free energy change calculated by TI. There are generally two ways to perform MMPBSA calculation on protein-ligand binding: the one-trajectory method and the three-trajectory method (Figure 4-4). In one-trajectory method only the protein-ligand complex is simulated, and then the coordinates of the protein and the ligand are extracted from the complex trajectory, to represent their dynamics before the binding. In three-trajectory method, three independent simulations are carried out for the ligand, the protein, and the complex, respectively. The three-trajectory method takes care of the dynamics change of the partners before and after the binding. However, because it also introduces noise that may take much longer time to converge, the one-trajectory method is used more often instead [139, 141, 151].

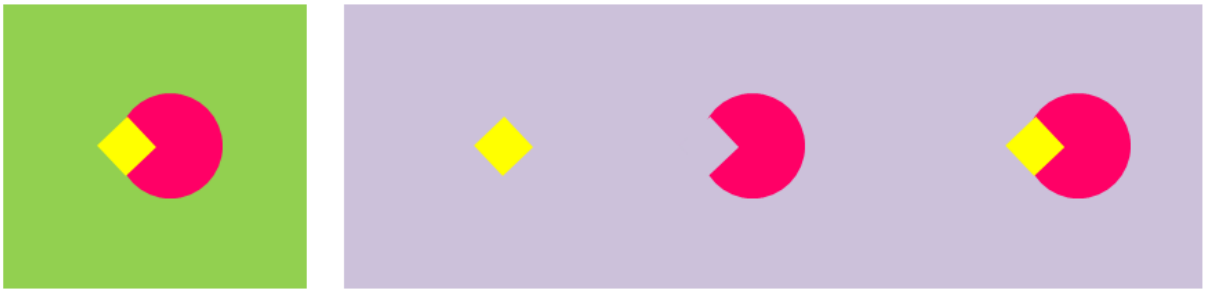

Figure 4-4 Comparison between one-trajectory (green background) and three-trajectory (purple background) MMPBSA method. Yellow square represents the ligand, while the pink pie represents the protein. In one-trajectory method, only the protein-ligand complex simulation is performed. In three-trajectory method, three simulations are carried out, corresponding to the ligand, the protein, and the complex, respectively.

4.2.3 Communication network analysis

Communication network analysis could be done on structures generated through MD simulations or elastic network models to understand the signal propagation in protein dynamics and function [152, 153]. We utilized such method to test the hypothesis that those drug resistance mutations distal from the active site may affect the binding through altering their communication with the active site. More specifically, we defined the communication propensity (CP) of each residue pair as the variance of the pair distance, as defined in Equation 4-6 [153]:

$$CP = \langle (d_{ij} - d_{ij,ave})^2 \rangle$$

Equation 4-6 Communication propensity (CP) of residue i and j, where d_{ij} is the distance between the Cα atoms of residue i and j, and $d_{ij,ave}$ is the average distance between the Cα atoms of residue i and j.

Under such definition, the two residues whose distance has small amplitude fluctuations (small CP values) are regarded as communicating efficiently since the change in position of one residue

would influence/reflect on the other residue, while the two residues whose distance has large amplitude fluctuations (large CP values) are regarded as communicating poorly since the change in one residue's position may not affect its partner because of the intrinsic amplitude of distance fluctuation.

4.2.4 HIVPR-inhibitor binding free energy change upon active site mutations

4.2.4.1 Simulation setup

Four crystal structures were used in this study, with PDB ID 1MRW, 1MRX, 1MSM, and 1MSN [154]. 1MRW is the wild type protease bound to experimental inhibitor KNI577 (WT-577), 1MRX is the active site mutant (V82F/I84V) bound to experimental inhibitor KNI577 (MUT-577), 1MSM is the wild type protease bound to experimental inhibitor KNI764 (WT-764), and 1MSN is the active site mutant (V82F/I84V) bound to experimental inhibitor KNI764 (MUT-764). All four crystals contain triple mutation Q7K/L33I/L63I to prevent the HIVPR auto-cleavage event during experiments. The non-charge parameters of the inhibitors (KNI577 and KNI764) were obtained from generalized AMBER force field (GAFF [133, 134]). The AM1BCC method [155] was used for their charge parameters.

Binding free energy change was calculated for three systems: L19A mutation (alchemical pathway) when the protease is either apo or bound to KNI577 inhibitor, V82F/I84V active site mutation when the protein is either apo or bound to KNI577 inhibitor, V82F/I84V active site mutation when the protein is either apo or bound to KNI764 inhibitor (Figure 4-5 and Figure 4-6).

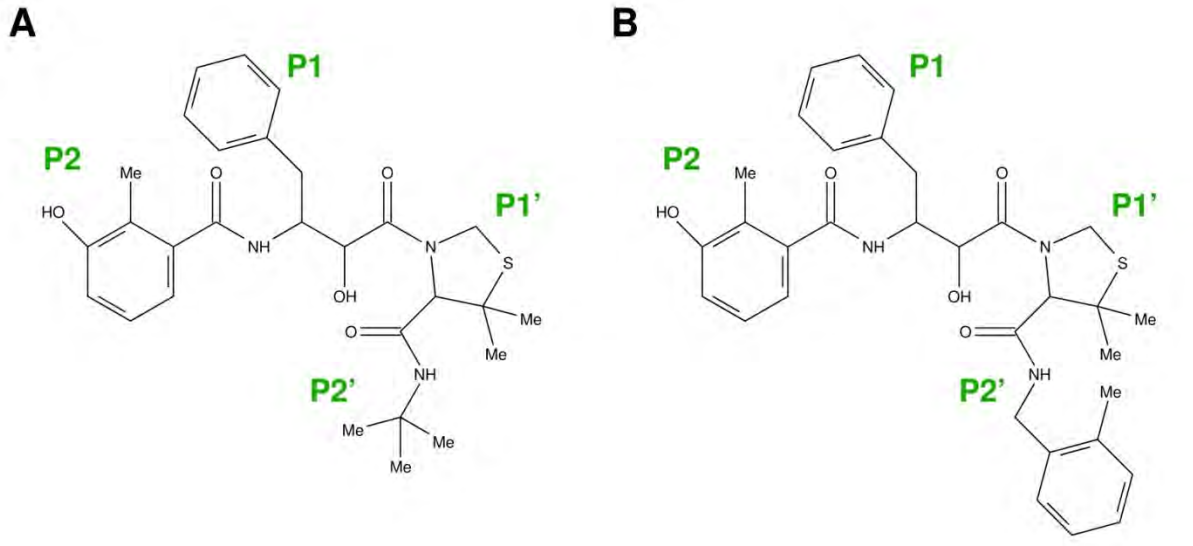

Figure 4-5 The chemical structure of KNI577 (A) and KNI764 (B). Both inhibitors share the same scaffold and functional groups, the only exception is at P2' position.

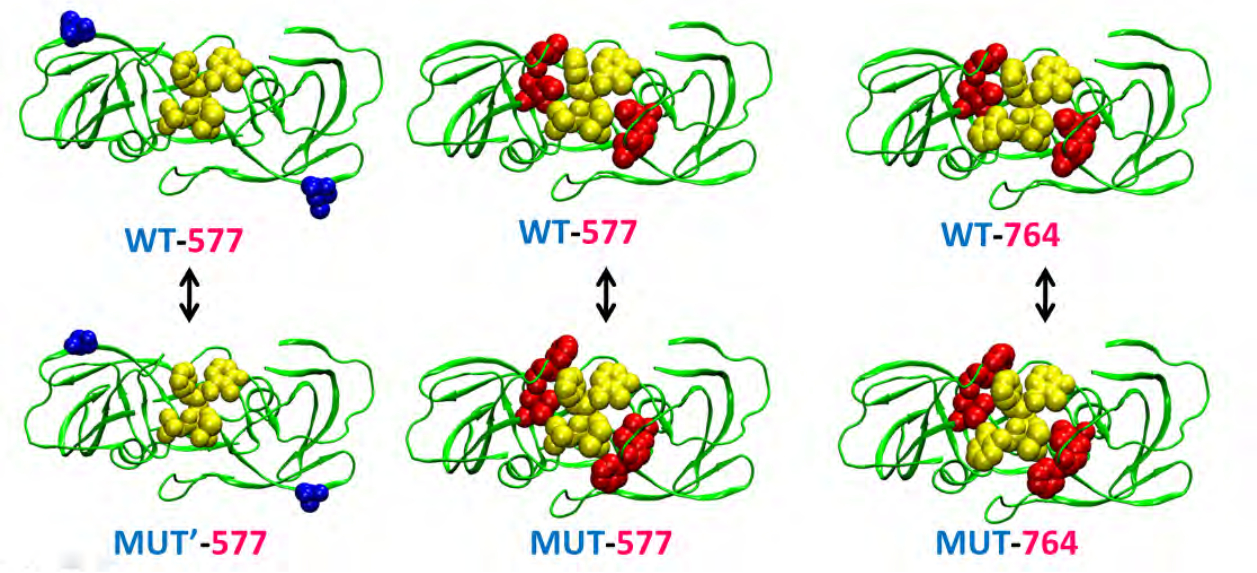

Figure 4-6 Binding free energy change were calculated for three systems using thermodynamic integration. Protein backbone shown as green ribbon, with the flaps removed to facilitate visualization. The inhibitor heavy atoms are shown as yellow vdW spheres. Left: control experiment where we mutated a surface residue distal from the active site while the protease is bound to inhibitor KNI577. The mutation sites on both monomers are shown as blue vdW spheres. Middle: mutation at two residues near the active site while the protease is bound to inhibitor KNI577. The mutation sites on both monomers are shown as red vdW spheres. Right: mutation at two residues near the active site (same mutations as the middle image) while the protease is bound to inhibitor KNI764. The mutation sites on both monomers are shown as red vdW spheres.

A truncated-octahedron TIP3P water box was added, with 8 Å minimum clearance from the boundaries, to each apo or holo structure prior to TI simulations, resulting in adding about 7,200 water molecules.

4.2.4.2 TI Simulation

With thermodynamic integration method implemented in AMBER simulation package [24], we calculated the free energy change upon mutation of specific protease residue(s) (Figure 4-2). The alchemical pathway of mutating protease residues is broken down into three steps: disappearing the charge of mutated residue(s), vdW volume transformation, and appearing the charge of the new residue(s). Breaking down the transformation as well as using a "softcore" Lennard-Jones scaling ([156]) help to improve simulation stability when large changes are simulated. In total six sets of simulations were carried out for each TI cycle (to get binding free energy change for one system in Figure 4-6): three-step-transformation when the apo protease is in the solvent and three-step-transformation when the holo protease is in the solvent.

For the free energy change upon L19A mutation, which is a surface residue, only 19 windows each transformation step were needed to get smooth energy derivative vs. lambda curve for the integration. For free energy change upon V82F/I84V double mutation, which involves residues near the active site and presumably have more influence on the binding, 39 windows were needed for each transformation step to get smooth energy derivative vs. lambda curve for the integration.

For each window, the solvated initial structure was first subjected to a 5000-step energy minimization. Then, the system was heated from 100 K to 298 K temperature over 100 ps while positional restraints are added on the heavy atoms (force constant 100 kcal/mol•Å^2), after which the system was equilibrated at 298 K for 300 ps with the positional restraints force constant 10 kcal/mol•Å^2. Then sidechains were equilibrated at 298 K for a total of 500 ps: in the first 250 ps, positional restraints were added on backbone heavy atoms with force constant 10 kcal/mol•Å^2, and in the second 250 ps, the force constant was decreased to 1 kcal/mol•Å^2. At last, 1 ns unrestrained constant pressure TI simulation was performed for each window, from which the energy derivatives were calculated. Time step was 2 fs. SHAKE Berendsen temperature and pressure control [25] were used. All bonds involving hydrogen atoms were constrained using SHAKE [28] with geometry tolerance of 10^{-5}. Particle-Mesh Ewald (PME) [29-32] were used to calculate long range electrostatic interactions, and a 8 Å cutoff was used for vdW interactions. For the TI transformation mask, only the sidechain atoms of mutating residues were involved.

4.2.4.3 MMPBSA

Conventional explicit solvent simulations of protease-inhibitor complexes were carried out for WT-577, MU-577, WT-764, and MU-764. Simulation trajectories were then subjected to MMPBSA calculations to pinpoint key residues responsible for the binding free energy change.

Since the HIVPR with deprotonated active site, where both Asp residues are deprotonated, would cause the inhibitor to significantly deviate from its crystal position during the protease-inhibitor complex simulation, we protonated Asp25 on monomer A. The mono-protonated active site maintained the inhibitor at an orientation closer to the crystal structure. The His69 were treated as protonated on both monomers. A truncated-octahedron water box was added with 8 Å clearance from the edges. After energy minimization and equilibration, which follow the same protocol as discussed in page 17, three independent production runs, 20 ns each, were carried out for each protease-inhibitor complex. Coordinates of the production runs were recorded every 2 ps and used for MMPBSA calculation.

The sander module in AMBER was used for MMPBSA calculation. The solvent probe radius probe in Poisson-Boltzmann (PB) energy calculation was 1.6 Å (the default value) and the surface tension value in surface area (SA) calculation was 0.0072 (the default value). Energy decomposition used the same principles as previously introduced in MMGBSA decomposition (page 37).

4.2.4.4 ACCENT

The configurationally entropy of the ligand before or after binding was estimated using ACCENT program [157, 158].

4.2.5 HIVPR-inhibitor binding free energy change upon multi-drug resistance mutations

4.2.5.1 Simulation setup

The charge parameters of inhibitor IDV (Table 4-1 and Figure 4-1) and TMC-126 (T12, Figure 4-7) were generated through multi-conformational RESP charge fitting. First, the coordinates of IDV and T12 were obtained from crystal structures 1SDT and 2I4V, respectively. Second, hydrogen atoms were added by tleap module in AMBER 11 simulation package [23],

and antechamber was used to assign initial charge using AM1-BCC model. Third, each inhibitor was then solvated in a TIP3P water box with about 500 water molecules, and went through energy minimization, equilibration, and 100 ns unrestrained MD simulation. Fourth, structural clustering was performed on simulation trajectories. At last, the representative structures of three largest clusters were subjected to geometry optimization (HF/6-31* level) and multi-conformational RESP charge fitting (MP2 level) using Gaussian98 [159] and R.E.D server [160]. The non-charge parameters of IDV and T12 were generated using antechamber module and generalized AMBER force field [133] in AMBER.

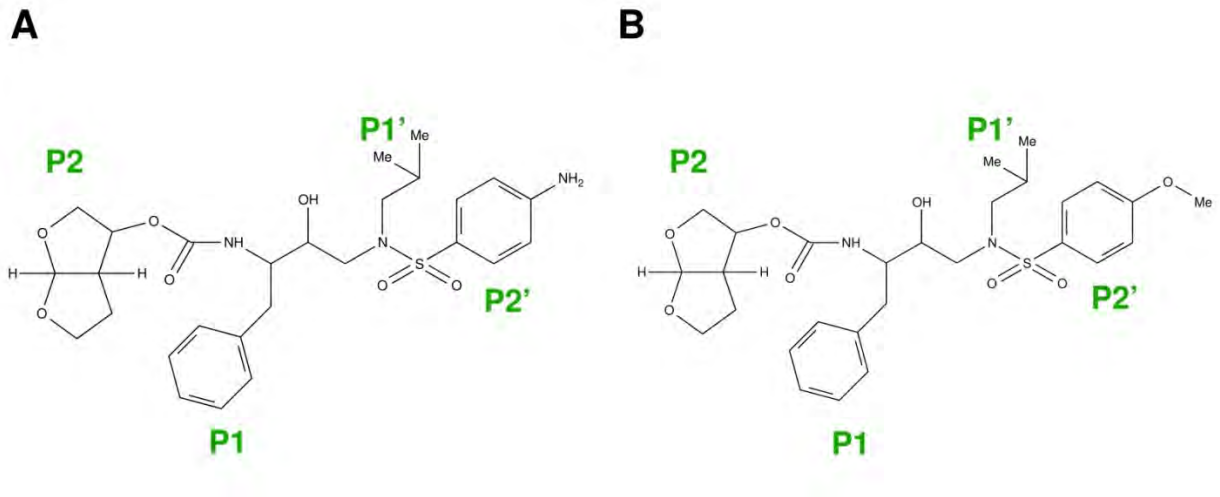

Figure 4-7 Comparison of the 2D-scheme between Darunavir (left, also named as DRV or TMC-114) and TMC-126 (right). The only structure difference is in P2'.

Six sets of protease-ligand simulations were prepared: wild type HIVPR bound to IDV, MDR mutant (L10I/M46I/I54V/V82A/I84V/L90M) bound to IDV, wild type HIVPR bound to T12, MDR mutant bound to T12, wild type HIVPR bound to substrate (RT-RH, with sequence Ala-Glu-Thr-Phe-Tyr-Val-Asp-Gly-Ala), MDR mutant bound to substrate. Three crystal structures were used:1SDT, 2I4V, and 1KJG, corresponding to the protease bound to IDV, T12, and RT-RH substrate, correspondingly. Virtual mutations were performed with swissPDB to match sequences to ones used in this study. Water molecules found in crystal structures that do not introduce clashes with mutated residues are kept. For protease-inhibitor simulations, the protease is protonated at one catalytic residue Asp25 (deprotonated on the other catalytic Asp25). For protease-natural substrate simulations the protease is diprotonated at the active site. Molprobity website (molprobity.biochem.duke.edu) was used to check crystal geometry and add hydrogen atoms. The starting structures were then solvated with truncated-octahedron TIP3P water box (about 7,200 water molecules added). Counter ions, either sodium or chloride ions, were added to neutralize the system.

4.2.5.2 Simulation

After energy minimization and equilibration, which followed the same protocol as introduced earlier on page 18, we performed batch simulations with random starting velocities, to speed up the convergence of conformation/energy samplings [71]. In total 50 runs were performed for each protein-ligand set, and each run was 4 ns. Berendsen temperature and

pressure control [25] was used to maintain the system at 298 K and 1 atm. Other simulation parameters were the same as introduced earlier (page 18).

4.2.5.3 MMPBSA

Simulation structures at 200 ps interval were subjected to MMPBSA calculation (about 10,000 structures for each protease-ligand set) using the sander module in AMBER. The solvent probe radius probe in Poisson-Boltzmann (PB) energy calculation was 1.6 Å (the default value) and the surface tension value in surface area (SA) calculation was 0.0072 (the default value). Energy decomposition followed the same principles as previously introduced in MMGBSA decomposition (page 37).

4.3 Results

4.3.1 HIVPR-inhibitor binding free energy change upon active site mutations

4.3.1.1 Free energy change studied by thermodynamic integration (TI)

Previously, Vega et al. performed isothermal titration calorimetry experiments to characterize the binding affinity of KNI577 and KNI764 before and after active site double mutation (V82F/I84V). They found that KNI764 is more tolerant to drug resistance mutations [154]. To assess the ability of thermodynamic integration (TI) method to reproduce experimentally measured binding free energy change, we carried out TI simulations of the protease bound to KNI577 or KNI764. Additional simulations of protease bound to KNI577 were performed to calculate binding free energy change upon surface residue mutation (L19A), which presumably has little effect on the binding, as a control (Figure 4-5 and Figure 4-6).

The comparison between computed and experimental binding free energy change is shown below (Table 4-2). As expected the surface mutation L19A has much less effect than active site mutations V82F/I84V, while the KNI764 is more tolerant to the drug resistant active site mutations. Although no experimental data are available for L19A mutation, the experimental data for the active site double mutation V82F/I84V are close to those calculated from TI.

inhibitor	mutation	Computed by TI	Experimental data
KNI577	L19A	-0.3	N/A
KNI577	V82F/I84V	2.7	3.3 ± 0.2
KNI764	V82F/I84V	1.2	1.9 ± 0.2

Table 4-2 Inhibitor binding free energy change upon mutation(s) in HIVPR. Experimental data were taken from Vega et al. 2004 Proteins paper. The TI error bars are calculated as the sum of standard error of the mean (SEM) of six transformation steps in each protein-inhibitor calculation.

As a way to check the convergence of TI calculations, we plotted out the integral versus time information, for each transformation step (Figure 4-8 and Figure 4-9). TI simulation is assumed to be converged when the integral does not change with time. Notably, vdW transformation integral has relatively higher amplitude of fluctuation in both protease-KNI577 and protease-KNI764 simulations (step2, panel C in both Figure 4-8 and Figure 4-9). However, although the standard deviation of the data is high, the data average is smooth and comparable to step 1 and 3 (panel D in both Figure 4-8 and Figure 4-9). Overall, the integral of all steps reached

plateau after 1 ns of production run. Therefore, we concluded that our TI simulations are converged, and this method could be used in the future to compare the free energy change among different protease-inhibitor complexes.

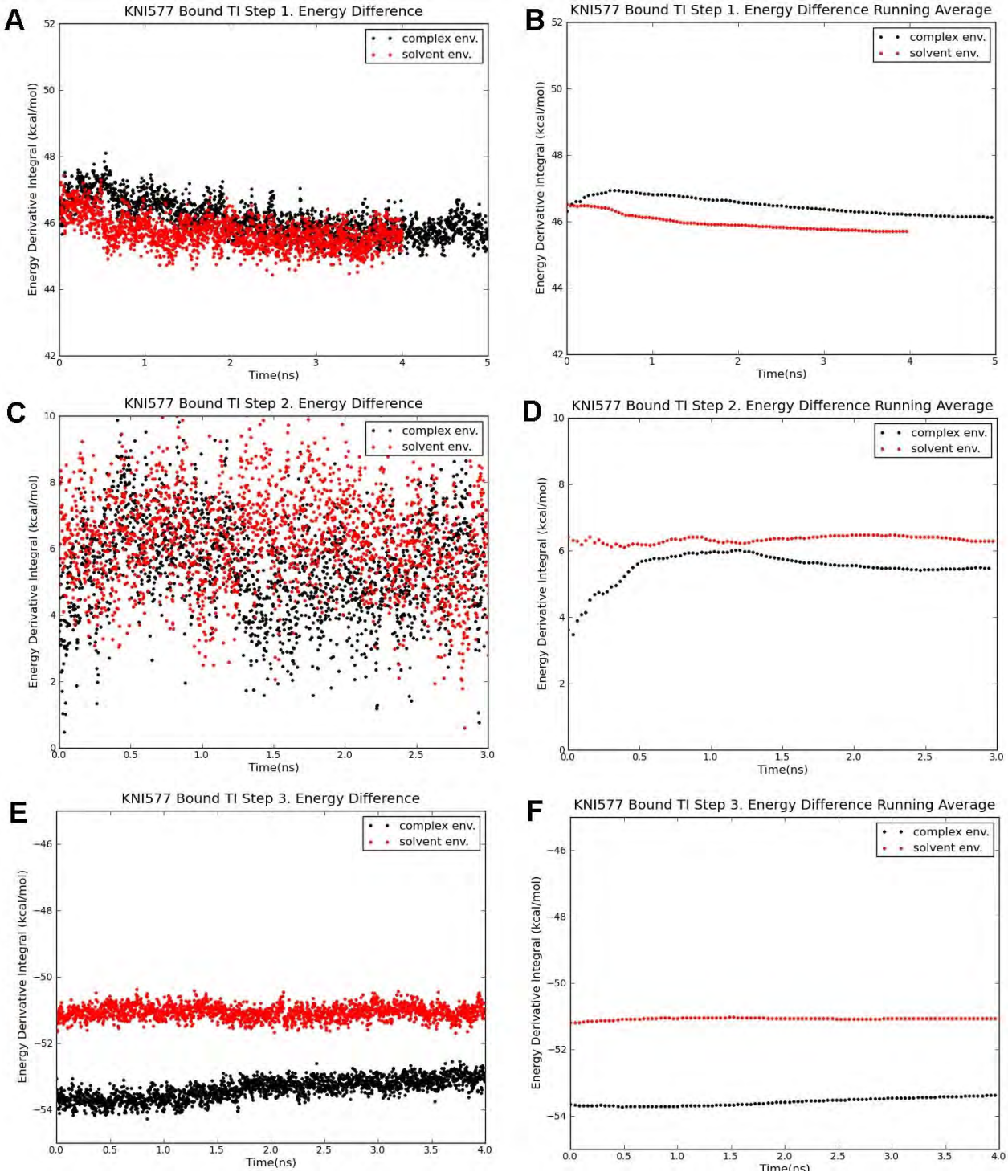

Figure 4-8 Integral vs. time curves for thermodynamic integration simulations of protease bound to inhibitor KNI577. Double mutation V82F/I84V was simulated. Complex environment (ligand

bound) simulation results are in black, and solvent environment (unbound) simulation results are in red. Panels B, D, and F are averaging every 10 data points on panel A, C, and E, respectively.

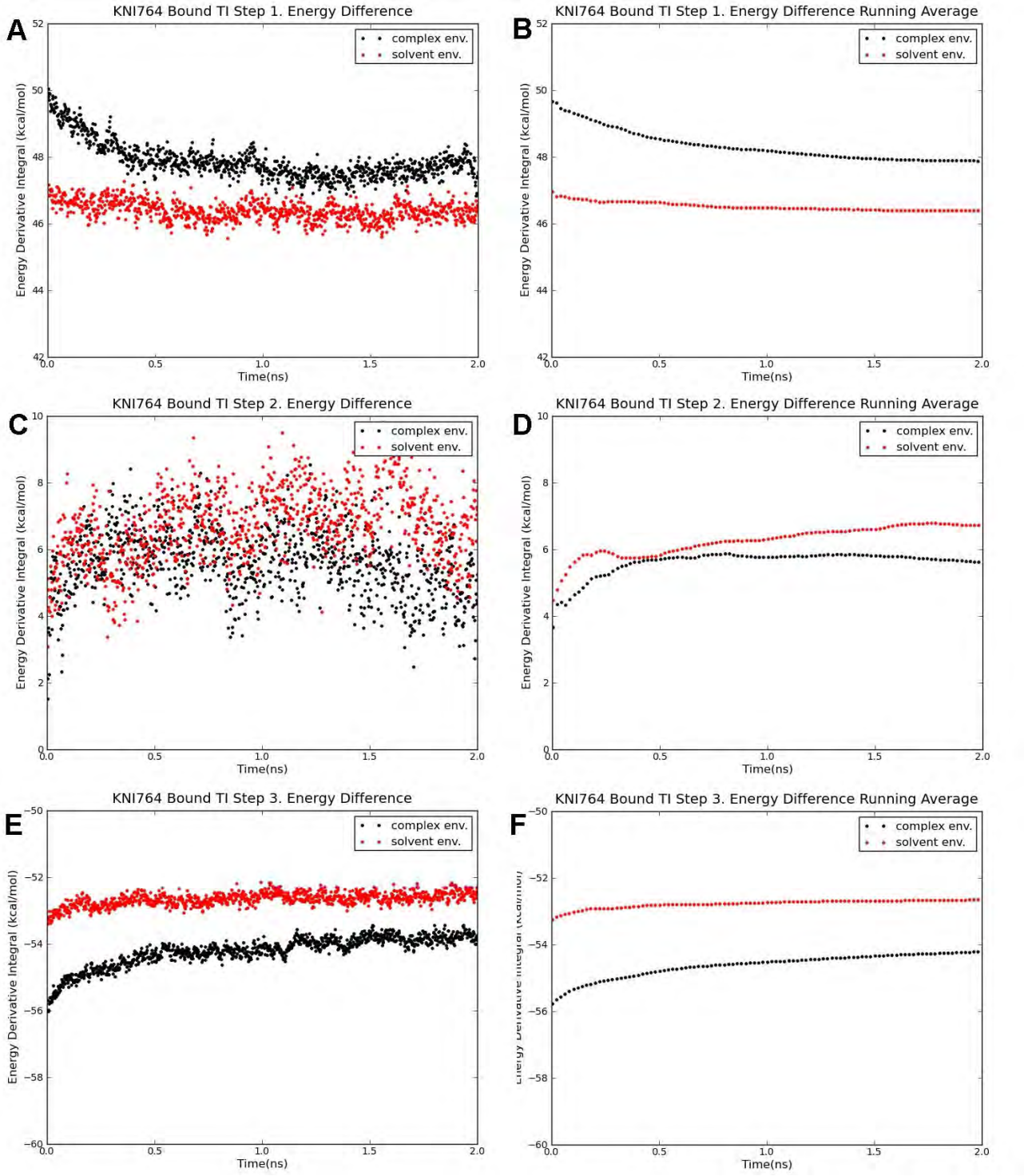

Figure 4-9 Integral vs. time curves for thermodynamic integration simulations of protease bound to inhibitor KNI764. Double mutation V82F/I84V was simulated. Complex environment (ligand bound) simulation results are in black, and solvent environment (unbound) simulation results are in red. Panels B, D, and F are averaging every 10 data points on panel A, C, and E, respectively.

4.3.1.2 Free energy change studied by MMPBSA and ACCENT

Although TI is good at predicting the relative binding energy, it is less informative for understanding where the energy difference comes from. So we performed molecular mechanics Poisson-Boltzmann surface area (MMPBSA) calculations to study the protease-inhibitor interactions.

Firstly, we compared the calculated free energy change and free energy components (Table 4-3) to experiments (Table 4-4). Note that although in experiments the binding enthalpy is measured separately from the entropy change, the calculated enthalpy in MMPBSA (third column in Table 4-3) already contained the solvent entropy components. Therefore the absolute values of energy components calculated are not comparable directly to the experimental data. However, if we assume that solvent entropy change is similar between KNI764 and KNI577 binding, since the two inhibitors share the same scaffold and water mediated interactions, then the relative energy within each energy components (relative energy of the four rows of data within column three of Table 4-3, etc.) should be comparable. Similar to experimental data, our calculation demonstrated that there is enthalpy loss upon active site mutation when the protease is bound to either KNI764 or KNI577, and there is favorable energy gain in the entropy term upon mutation when KNI764 is bound to the protease, rather than the unfavorable entropy experienced by KNI577 upon active site mutations. The different entropy response between KNI577 and KNI764 is likely because KNI764 has one more rotatable bond than KNI577, which enables it to better accommodate the drug-resistant mutations. However, probably because of an underestimation of the mutant-KNI764 entropy (21.5 kcal/mol), the 3.5 kcal/mol entropy gain is much larger the 0.4 kcal/mol entropy gain in experimental data, which affects the free energy comparison. Overall, the free energy we calculated is comparable to experimental data, except for mutant-KNI764 complex.

inhibitor	Protease sequence	ΔH – TΔSsolv	-ΔSconf	ΔG
KNI764	V82F/I84V	-39.2 ± 0.0	21.5 ± 0.0	-17.8 ± 0.0
KNI764	Wild type	-41.9 ± 0.0	25.0 ± 0.0	-16.9 ± 0.0
KNI577	V82F/I84V	-29.5 ± 0.0	18.4 ± 0.0	-11.1 ± 0.0
KNI577	Wild type	-32.6 ± 0.0	17.2 ± 0.0	-15.4 ± 0.0

Table 4-3 Absolute energy (unit in kcal/mol) of protease-inhibitor binding calculated using MMPBSA and ACCENT. Note that solvent entropy change (ΔSsolv) is implicitly accounted for in the solvation energy calculation in MMPBSA, while the solute configurational entropy (ΔSconf) is calculated by ACCENT separately.

inhibitor	Protease sequence	ΔH	-TΔSsolv -ΔSconf	ΔG
KNI764	V82F/I84V	-5.3 ± 0.3	-7.1 ± 0.3	-12.4 ± 0.1
KNI764	Wild type	-7.6 ± 0.2	-6.7 ± 0.2	-14.3 ± 0.1
KNI577	V82F/I84V	-2.1 ± 0.2	-7.8 ± 0.2	-9.9 ± 0.1
KNI577	Wild type	-4.7 ± 0.2	-8.5 ± 0.2	-13.2 ± 0.1

Table 4-4 Absolute energy (unit in kcal/mol) of protease-inhibitor binding from experiments presented in Vega et al. 2004 Proteins paper.

Secondly, we compared non-bonded energy components of MMPBSA among different protease-inhibitor complexes (Table 4-5). It is clear that the vdW energy is mainly responsible for the energy difference between KNI577 and KNI764, while the electrostatic energy is mainly responsible for the energy difference between wild type protease and mutant protease.

inhibitor	Protease sequence	Evdw	Eele + Epol_solv	Enon_pol_sol
KNI764	V82F/I84V	-70.0 ± 0.0	-40.2 ± 0.0	-9.4 ± 0.0
KNI764	Wild type	-70.6 ± 0.0	-37.9 ± 0.0	-9.2 ± 0.0
KNI577	V82F/I84V	-60.9 ± 0.0	-40.2 ± 0.0	-8.7 ± 0.0
KNI577	Wild type	-61.9 ± 0.0	-37.9 ± 0.0	-8.6 ± 0.0

Table 4-5 Non-bonded energy components of MMPBSA calculations. Evdw: vdW energy. Eele: electrostatic energy. Epol_sol: polar solvation energy. Enon_pol_sol: non-polar solvation energy.

Thirdly, we decomposed the binding energies calculated by MMPBSA into per-residue level (Figure 4-10), and looked at those residues with large energy change upon binding (Figure 4-11). As expected, the residues with large energy change upon binding are all clustered around the active site (Figure 4-12): many favorable interactions between the protease and the inhibitor are formed upon the binding. There are also some residues, however, that have unfavorable energy change upon binding. For example, Arg8 on both monomers. This residue Arg8 can form salt bridge interaction with Asp29 (Figure 2-3), however this salt bridge is destabilized when the inhibitor is bound.

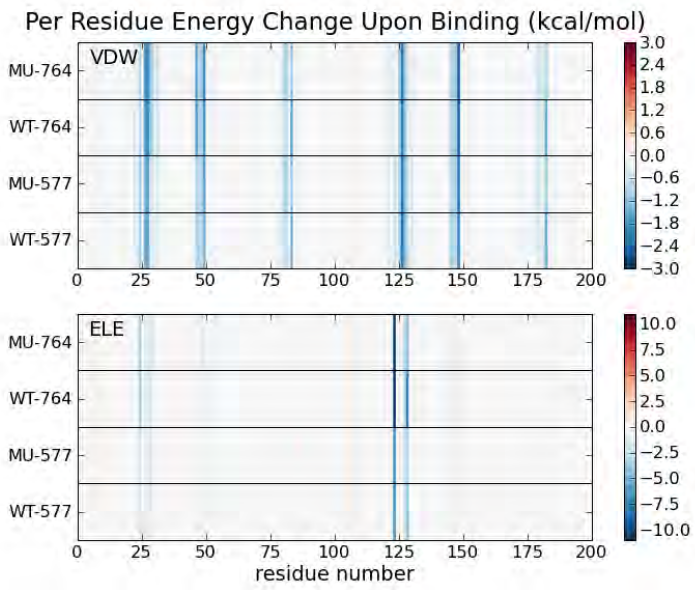

Figure 4-10 Decomposition of the vdW (VDW) and electrostatic (ELE) interaction energies between the protease and the inhibitor into per-residue basis. Different color scales were applied to vdW and electrostatic energy panel to facilitate visualization. The white color represents no change in residue energy upon binding, the red color represents unfavorable energy upon binding, and the blue color represents favorable energy upon binding. HIVPR has 99 residues in each monomer, and residues on monomer B are numbers from 100.

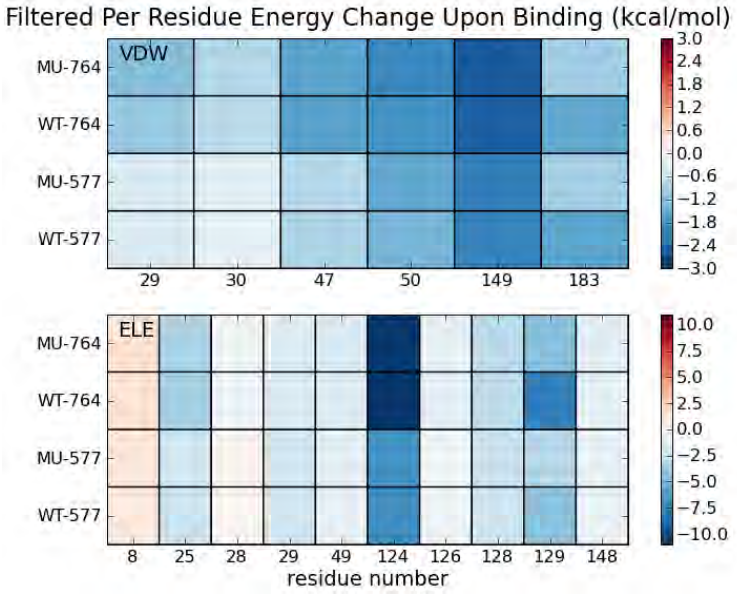

Figure 4-11 Decomposition of the vdW (VDW) and electrostatic (ELE) interaction energies between the protease and the inhibitor into per-residue basis. Only the residues with more than 5 kcal/mol energy change upon binding are shown. Different color scales were applied to vdW and electrostatic energy panel to facilitate visualization. The white color represents no change in residue energy upon binding, the red color represents unfavorable energy upon binding, and the blue color represents favorable energy upon binding. HIVPR has 99 residues in each monomer, and residues on monomer B are numbers from 100.

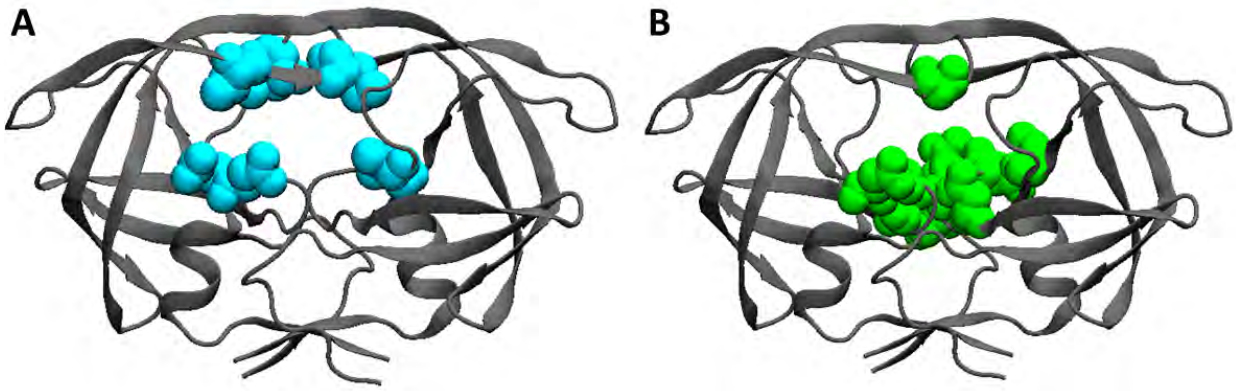

Figure 4-12 HIVPR residues that have more than 5 kcal/mol vdW energy (A) or electrostatic energy (B) change upon inhibitor (KNI577 or KNI764) binding.

Finally, we compared the average structures of different protease-inhibitor simulations, to identify the difference in protease-inhibitor interactions between KNI577 and KNI764 inhibitors

(Figure 4-13). The favorable vdW energy of KNI764 (compared to KNI577, Table 4-5) is between its toluene group and the flap residue, which is indicated by the closer distance between the flap residue and the toluene group in Figure 4-13 panel C. The favorable electrostatic interaction between wild type protease and the inhibitor (compared to mutant-inhibitor, Table 4-5) is likely represented by the better hydrogen bond orientation between the hydroxyl group on P2 group and Asp29 (on the left side of Figure 4-13 panel E). This better orientation of wild type protease is attributable to the V82F mutation, which disrupts the vdW packing at P1 group (on the right side of Figure 4-13 panel E) and in turn influences the interactions at the P2 site nearby.

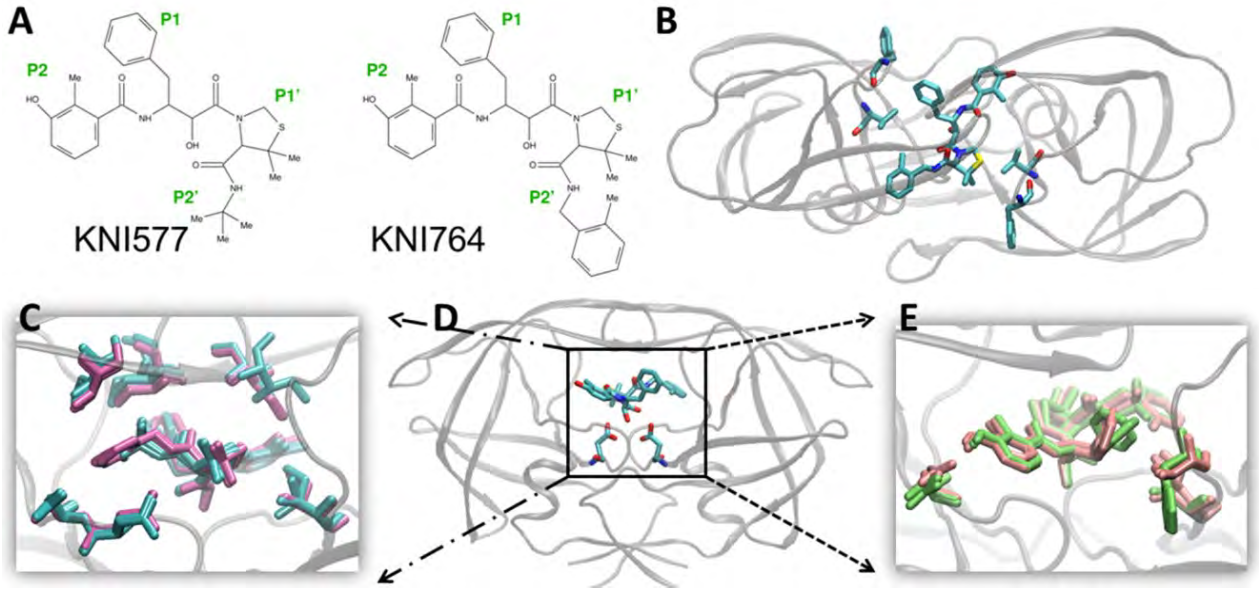

Figure 4-13 Comparison of protease-inhibitor interactions. A) The chemical structure of KNI577 and KNI764. Both inhibitors share the same scaffold and functional groups, with the only exception at P2' position. B) Top view of the double mutant protease (V82F/I84V) bound to KNI764. Heavy atoms from inhibitor and double mutations are shown in licorice representation. C) Front view of the average structure of the protease bound to KNI577 (blue) and KNI764 (red). D) Back view of the average structure of protease bound to KNI764. Heavy atoms from inhibitor and active site residues are shown in licorice representation. E) Back view of the average structure of the wild type protease (red) and double mutant protease (green) bound to KNI764.

4.3.2 HIVPR-inhibitor binding free energy change upon multi-drug resistance mutations

Multi-drug resistance (MDR) has been a major problem in AIDS treatment. Once a HIV strain develops resistance against multiple drugs it has been treated with, the virus can start replicate again and the infection is likely to worsen. An intriguing question in understanding HIVPR drug resistance is how the mutations distal from the binding site could affect the binding affinity and influence drug resistance [68].

Ohtaka et al. previously studied the MDR strain containing six amino acid mutations (L10I/M46I/I54V/V82A/I84V/L90M) located near the flap, active site, and also the termini region (Figure 4-14). They found that this mutant lowers the affinity of all inhibitors by 2-3 orders of magnitude, which is more than the sum of individual mutations at corresponding flap,

active site, and the termini regions. This indicates the existence of cooperative effects. Accordingly, there also have been studies trying to explain the effect of drug-resistance mutations outside the active site [161-163]. However, these studies were limited by the technique and computational resources available then, and could not provide a clear picture of the mechanism behind distal coupling.

Previous studies also found that different inhibitors have different responses to MDR mutations. More specifically, the second generation drugs (ATV, APV, LPV, TPV, and DRV) generally have better tolerance against MDR mutations than the first generation drugs (IDV, RTV, NFV, and SQV, see Table 4-1 for inhibitor details) [68, 164-166]. An extensive review on HIVPR inhibitors and drug resistance has been provided by Ali et al. [68].

To investigate the effect of distal drug resistant mutations, and examine whether the long-range coupling between residues could play a role in drug resistance, we studied the binding of wild type and multi-drug-resistance strain of HIV-1 protease (with sequence L10I/M46I/I54V/V82A/I84V/L90M, which is the same as the sequence used by Ohtaka et al.) to different types of ligands. The ligands we studied include the first generation protease inhibitor indinavir (IDV, Figure 4-1), the second generation protease inhibitor TMC-126 which is an analog to the most recent HIVPR drug darunavir (DRV, see Figure 4-7 for a comparison between DRV and TMC-126), and natural substrate RT-RH (the cleavage site between reverse-transcriptase and RNase H with the sequence Ala-Glu-Thr-Phe*Tyr-Val-Asp-Gly-Ala where the cleavage site is indicated with *). We expect that the multi-drug-resistant mutations would have more dramatic effect on the first generation inhibitor IDV than on the second generation drug TMC-126 (T12) or the natural substrate RT-RH.

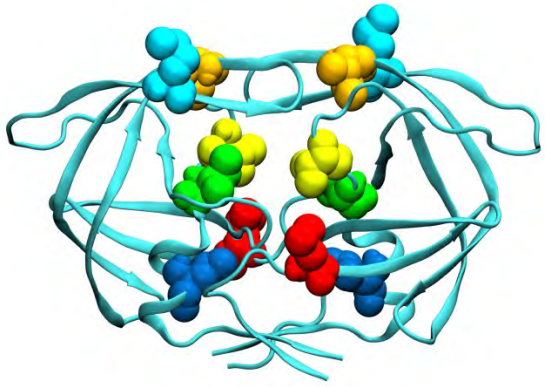

Figure 4-14 Structure of HIV-1 protease showing the location of mutations associated with multi-drug resistance: M46I in light blue, I54V in orange, V82A in yellow, I84V in green, L10I in red, and L90M in blue.

4.3.2.1 Protease-ligand interactions studied by MMPBSA calculation

To ensure the convergence of energy/structure sampling, we performed batch simulations of protease-inhibitor complexes (50 runs, 4 ns each, totaling about 200 ns for each simulated complex). Six complexes were simulated: WT protease bound to IDV, MDR protease bound to IDV, WT protease bound to T12, MDR protease bound to T12, WT protease bound to RT-RH, and MDR protease bound to RT-RH. Simulation trajectories were post processed to calculate the

binding energy using MMPBSA method. The good fit of our MMPBSA results to Gaussian indicates convergence of the data (Figure 4-15). The absolute binding energy difference among the ligands is unreasonable, since the T12 in reality is a much better binder than IDV. It is worth noting that because of the large variation among HIVPR inhibitors, ranking absolute binding free energies of different HIVPR inhibitors computationally has been a big challenge. However, although the absolute energies are unreasonable, the relative energy difference among three ligands still seem reasonable: the first generation inhibitor IDV lost 7 kcal/mol binding energy upon MDR mutations, while both natural substrate and the second generation drug T12 are more tolerant to the mutations (about 4 kcal/mol energy loss each).

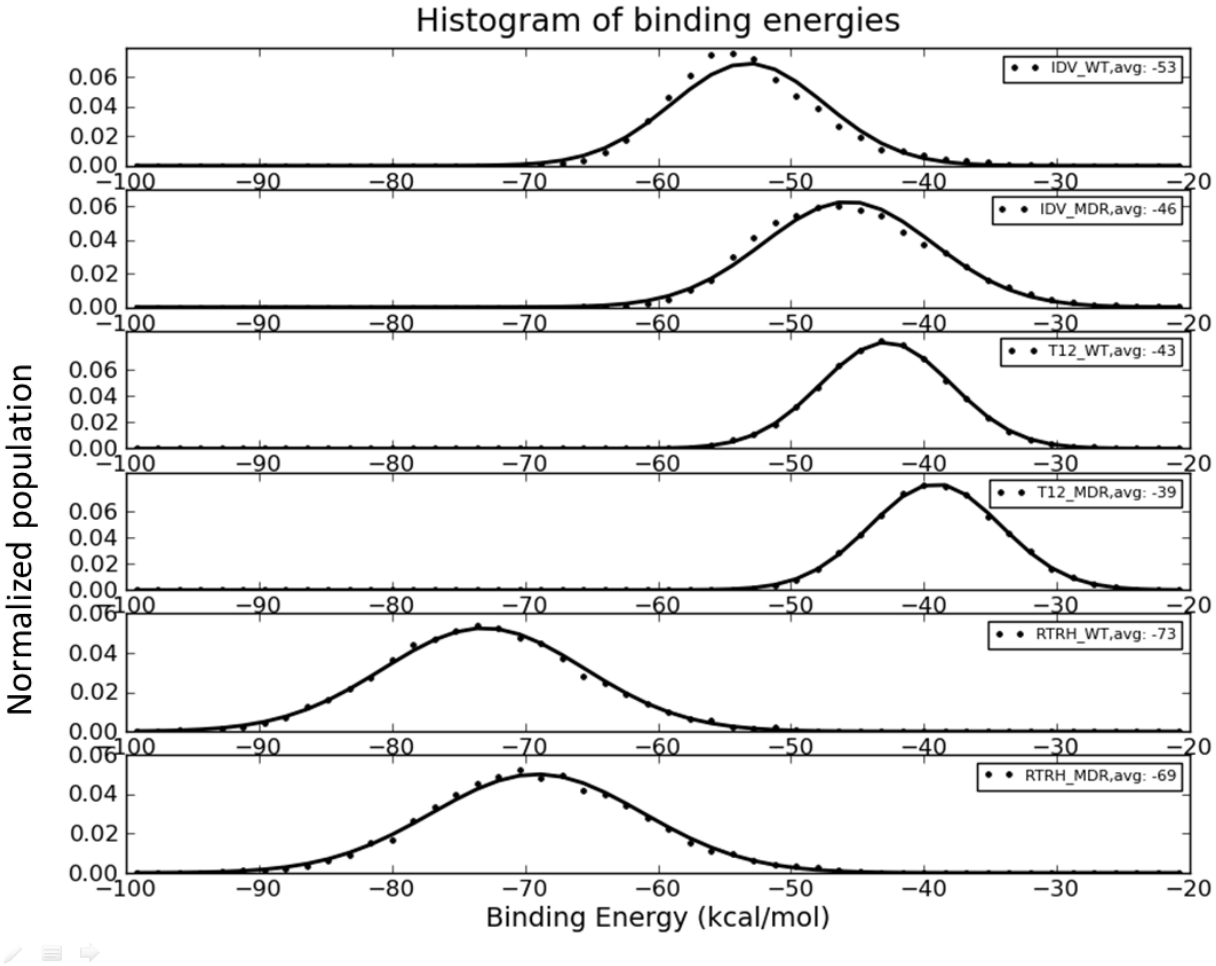

Figure 4-15 Gaussian fitting for protease-ligand binding energies calculated using MMPBSA method. The histogram of MMPBSA data is shown in black dots, and the fitted Gaussian curves are shown as black lines. The average energy obtained from each Gaussian curve is listed in the legend box at the upper-right corner.

We then looked at the binding energy components of each protease-ligand complex (Table 4-6). Interestingly, all three WT-ligand complexes experienced vdW energy loss upon MDR mutations, while only the WT-IDV complex has experienced significant loss in electrostatic energy component, which is partly offset by polar solvation energy. This difference

in electrostatic component also distinguishes IDV from the other two ligands in the total binding energy loss.

Complex	Total energy	vdW	Electrostatic + Solv_pol	Solv_nonpol
IDV_WT	-53.2±0.1	-74.3±0.0	31.0±0.1	-10.0±0.0
IDV_MDR	-45.7±0.1	-71.7±0.0	36.1±0.1	-10.1±0.0
T12_WT	-42.8±0.0	-67.5±0.0	33.0±0.0	-8.4±0.0
T12_MDR	-39.1±0.0	-63.7±0.0	33.2±0.0	-8.6±0.0
RTRH_WT	-73.1±0.1	-84.7±0.1	24.0±0.1	-12.4±0.0
RTRH_MDR	-69.1±0.1	-80.6±0.1	24.1±0.1	-12.5±0.0

Table 4-6 MMPBSA decomposition of protease-ligand binding. From left to right: the protease-ligand complex studied, total energy, vdW energy, electrostatic and polar solvation energies, non-polar solvation energy.

We then decomposed the binding energy into per-residue basis in order to compare the contribution of different residues and pick out those that are important (Figure 4-16). We found that the decrease in IDV binding affinity almost comes exclusively from the electrostatic energy loss of the inhibitor IDV (Figure 4-16 lower panel IDV_WT, residue number 199).

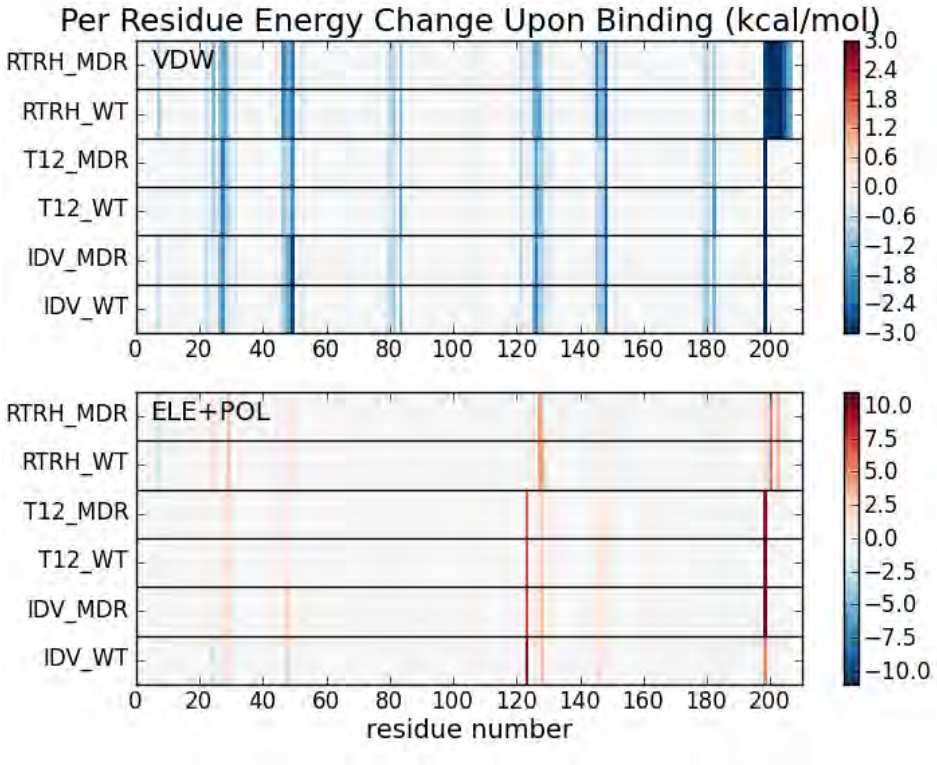

Figure 4-16 Decomposition of the vdW (VDW) and electrostatic (ELE) plus polar solvation energy (POL) into per-residue basis. Different color scales were applied to two panels to facilitate visualization. The white color represents no change in residue energy upon binding, the red color represents unfavorable energy upon binding, and the blue color represents favorable energy upon binding.

4.3.2.2 Distal coupling within the protease studied by communication network analysis

In the one-trajectory MMPBSA method we used (Figure 4-4), the same trajectory is used to represent both the bound and unbound state of the protease, so the residues distal from the active site would hardly experience any energy change upon binding. This is why the method always identifies residues near the active site as important. Therefore, although one-trajectory MMPBSA method could be useful in studying protease-ligand interaction, it is not helpful in explaining drug-resistant mutations distal from the active site.

We examined possible long-range communication among protease residues by means of communication network analysis ([153], see the methods section for details). We measured the fluctuation of residue-residue distance. The residues whose distance has small fluctuation/variance are regarded as communicating efficiently, while those residues whose distance has large fluctuation are considered as communicating poorly. We are especially interested in those residue pairs with long distance and small distance fluctuation. These residues are regarded as long-range communicators and they may explain how distal mutations affect the binding at the active site.

The residue pair distance matrix (Figure 4-17) reasonably reflects the structure of the protease. First of all, the distance matrices shown below are all symmetric along the diagonal, because of the pair-wise nature of the data. The blue strokes perpendicular to the diagonal are generally representing the β-hairpins, where the β-hairpin tip residue i is on the diagonal, and residues i+1, i+2, i+3 are close to residues i-1, i-2, i-3, etc. The six darkest strokes represent the three β-hairpins (fulcrum, flap, and cantilever, with the hairpin tip as residue 17, 50, and 67, respectively, see Figure 1-1 for details) on each monomer (Figure 4-17 A black square). The less dark strokes represent loop regions, for example, the active site loop. The helix is represented as a square alone the diagonal (Figure 4-17 A grey square). The lower right corner represents the inter-monomer interactions. Generally this region has red color, representing the larger distance between residues from different monomers, with the exception of some blue regions, representing the dimer interface (Figure 4-17 A green square). Overall, the distance plots for wild type and mutant protease bound to the same ligand are almost identical, which means the bound structure remains the same upon drug resistance mutations. This is as expected, since the ligand binding site is the same and the binding modes are similar. We hypothesized that the communication between distal sites may be hampered by the drug-resistance mutations, but the average residue-pair distance may not be affected.

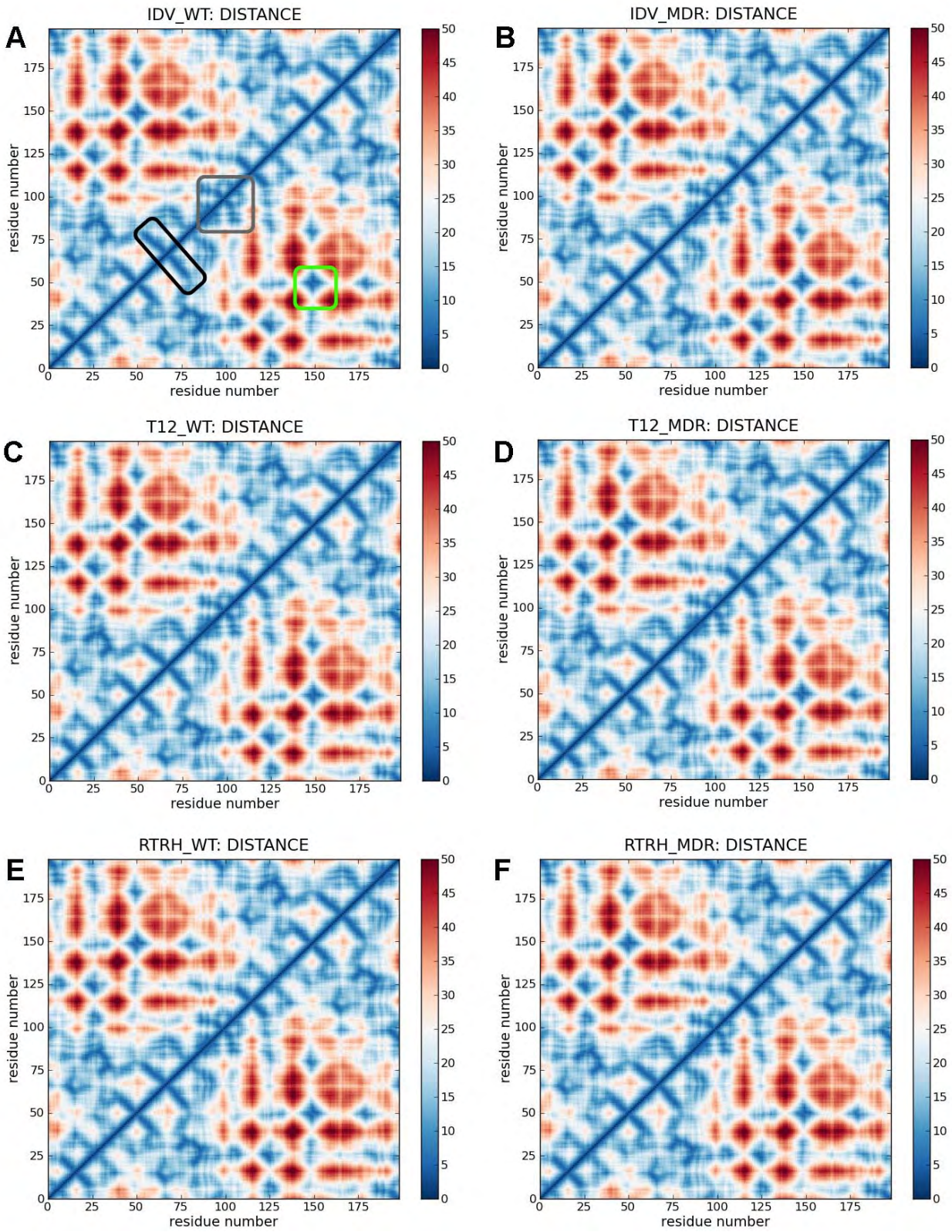

Figure 4-17 Distance matrix of protease-ligand complexes. Red blue color scheme is used: red color means longer distance, and blue color means shorter distance. Residues from the first monomer are

numbered as residue 1 to 99, and residues from the second monomer are numbered as residue 100 to 198. Flap tips are around residue 50 and 149, and the active site is composed of residue 25 and 124. Because of the pair-wise relationship, the matrix is symmetric along the diagonal. Elements are labeled in panel A: β-hairpin around residue 67 (the cantilever β-hairpin) in the black square, the helix and termini residue pairs in the grey square, and the inter-monomer residue pairs near the flap tip region in the green square.

The variance matrix (Figure 4-18) is not a one-by-one translation from the distance matrix otherwise the two matrices would have had the same pattern. Distance matrix pattern contains many diagonal elements, which means a residue is closest to its bonded neighbors. In contrast, variance matrix pattern is composed by many squares, meaning that the residues nearby are stabilized as a whole, although they don't have the same residue-pair distances. Also, it is not hard to spot that the variance matrices from the wild type and mutant protease differ, even if the same ligand is bound, which means that the variance is more sensitive to the protease sequences or ligand structures.

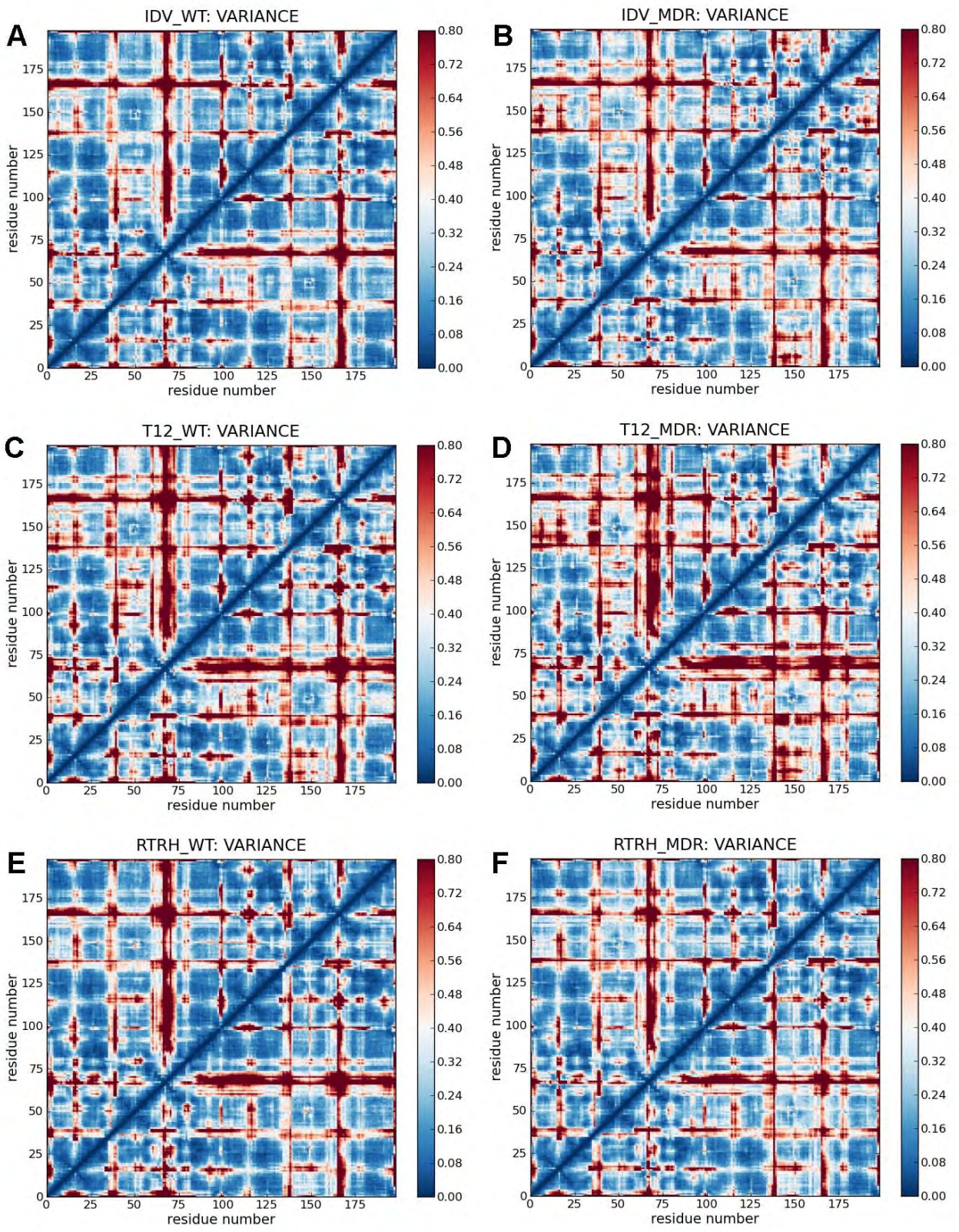

Figure 4-18 Variance matrix of protease-ligand complexes. Red blue color scheme is used: red color means larger variance, and blue color means smaller variance. Residues from the first

monomer are numbered as residue 1 to 99, and residues from the second monomer are numbered as residue 100 to 198. Flap tips are around residue 50 and 149, and the active site is composed of residue 25 and 124. Because of the pair-wise relationship, the matrix is symmetric along the diagonal.

Since we are interested in the change of protease communication ability upon drug-resistance mutations, we calculated the difference of variance matrixes, and included the results in Figure 4-19. Note that in Figure 4-19 we filtered out residue pairs with pair distance less than 10 Å to focus on long-range communications, and we also filtered out those residues pairs with variance bigger than 0.2 before and after the drug-resistance mutations, to focus on those efficient communicators and pick up those residue pairs that gain or lose communication upon mutations.

The filtered variance-difference matrices are shown in Figure 4-19. It is quite obvious that many of the communications in IDV become weaker upon drug resistant mutations, which is represented by the red color on the variance-difference matrices, while the change for T12 and RTRH are much smaller, which is reflected by less red dots on the variance difference matrices. Those residues being mutated in the MDR strain (L10I/M46I/I54V/V82A/I84V/L90M) are indicated with black lines in Figure 4-19, and we noticed that not all the variance changes occur near the mutated residues. For example, IDV has variance increase involving residue 175 (residue 175 is equivalent to residue 74 on the other monomer, see Figure 4-19 A around coordinate [175, 50]), and RTRH has variance decrease around residue 60 (see Figure 4-19 C near coordinate [60, 17]).

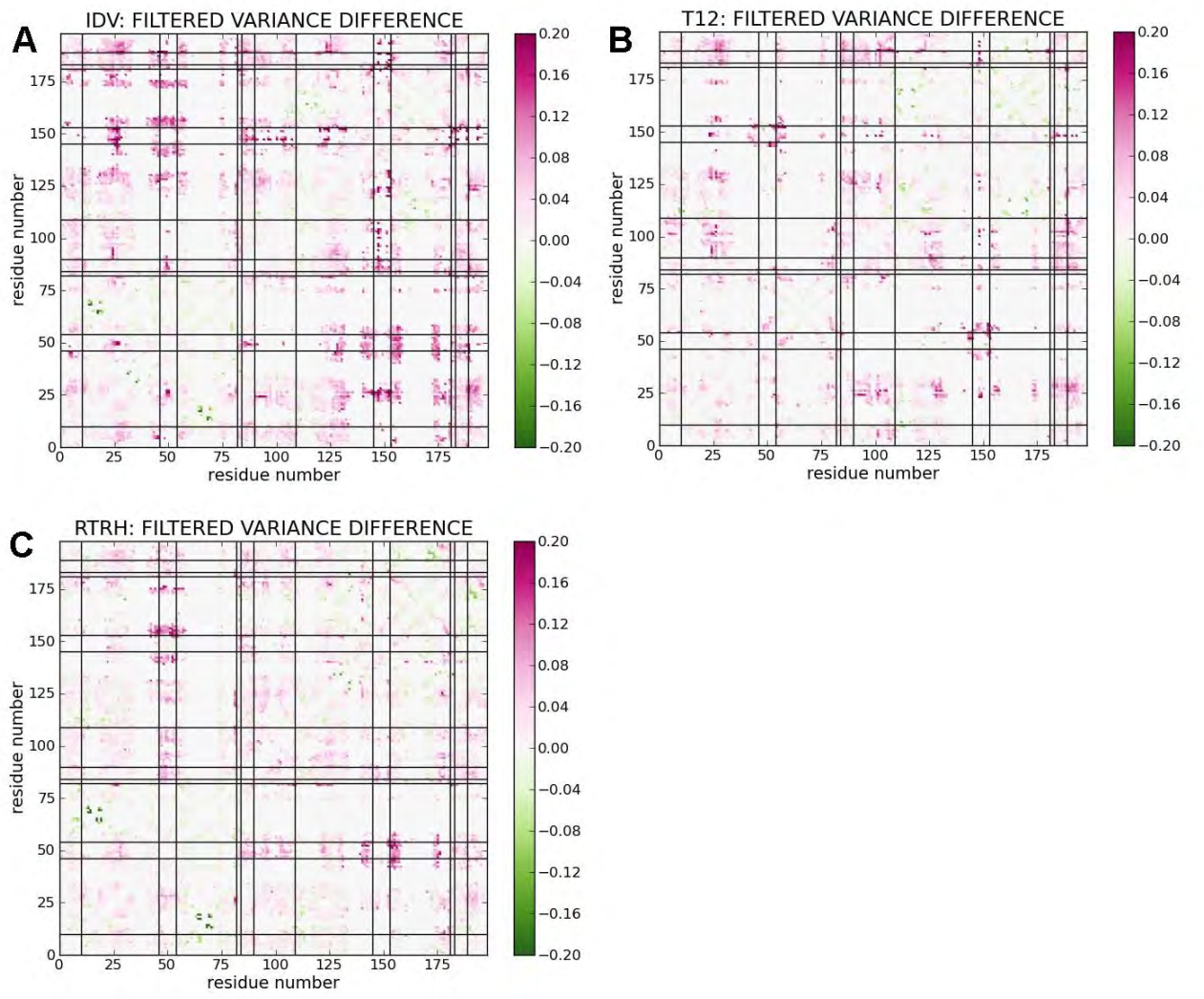

Figure 4-19 Filtered variance difference matrix. Only those long-range efficient communications are shown: 1) residue pairs with pair distance less than 10 Å are filtered out, and 2) residue pairs with distance variance larger than 0.2 before and after multi-drug resistant mutations are filtered out. Mutated residues in the MDR strain (L10I/M46I/I54V/V82A/I84V/L90M) are indicated as black lines over the matrices to help visualization.

To better assess the variance change, we re-processed the variance change shown in Figure 4-19 and listed those large variance changes (absolute value bigger than 0.2) of each system in Table 4-7. Consistent with the finding above, in the variance change listed IDV weakened much more residue-pair communications compared to T12 and RTRH upon drug resistant mutations, while RTRH strengthened the largest number of residue-pair communication upon drug resistant mutations compared to the other two ligands.

Weaker communication upon MDR mutations			Stronger communication upon MDR mutations		
IDV	**T12**	**RTRH**	**IDV**	**T12**	**RTRH**
49 - 26	3' - 26	43' - 52	64 - 17		64 - 17
50 - 26	24' - 2'	77' - 48	68 - 13		64 - 19
87 - 49	44' - 51	78' - 26'			65 - 17
92 - 24	49' - 23				68 - 13
46' - 26	50' - 23				68 - 14
47' - 26	86' - 49'				69 - 14
48' - 25	89' - 49'				69 - 18
48' - 26	90' - 2'				70 - 13
48' - 27	92' - 49'				
48' - 94	94' - 49'				
48' - 96					
48' - 97					
48' - 1'					
48' - 3'					
48' - 4'					
48' - 5'					
48' - 8'					
48' - 24'					
48' - 25'					
52' - 26					
52' - 25'					
53' - 25					
53' - 26					
53' - 21'					
53' - 25'					
53' - 32'					
53' - 33'					
55' - 3					
83' - 53'					
84' - 53'					
85' - 48'					
86' - 53'					
89' - 48'					
89' - 52'					
89' - 53'					
90' - 45					
90' - 46					
90' - 47					
94' - 48'					
94' - 53'					

Table 4-7 Change in protease residue communication efficiency upon MDR mutations. Only those long-range efficient communications are shown: 1) residue pairs with pair distance less than 10 Å are filtered out, and 2) residue pairs with distance variance larger than 0.2 before and after multi-drug resistant mutations are filtered out. Moreover, only those variance changes bigger than 0.2 (absolute value) are shown. From left to right: residue pairs that have their distance variance increased more than 0.2, and residue pairs that have their distance variance decreased more than 0.2. Residues on the second monomer are indicated with a prime following the residue number. Residue pair partners are separated with a hyphen.

We mapped the residue pairs listed in Table 4-7 onto the HIVPR structure (Figure 4-20 A-C), and we found that the loss of communication upon drug resistance mutations is mainly distributed along the dimer interface, namely the communication among the flap, the catalytic site, and the termini region. The trend is even more obvious if we include not only those variance changes larger than 0.2, but also those changes less than 0.2 but larger than 0.15 (Figure 4-20 D-F). The loss of communication is most common in IDV, which is followed by T12, and the natural substrate has the least residue pairs lose communication upon the MDR mutations. On the other hand, the gain of communication upon MDR mutations usually happens on the sides of the protease, among the fulcrum and the cantilever. The gain of communication is less symmetric than the loss of communication (changes happening only on one monomer but not on the other monomer at similar locations), and we did not observe any gain of communication for T12.

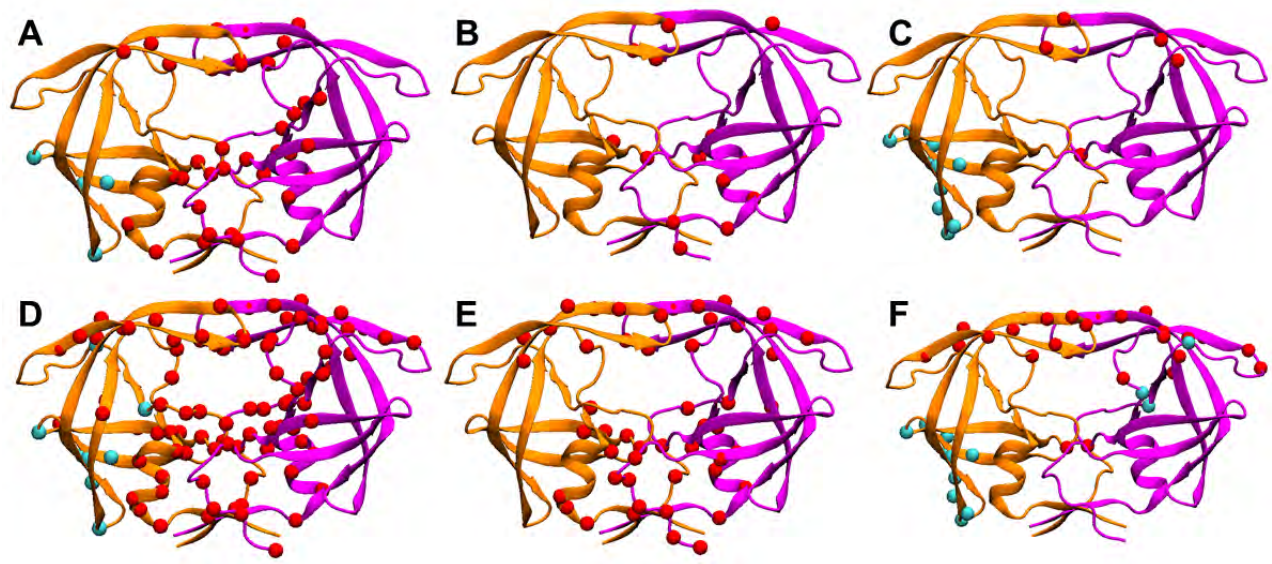

Figure 4-20 Communication analysis results mapped onto HIV-1 protease structure. The backbones of protease monomer A and B are shown in cartoon representation, colored orange and magenta, respectively. The Cα atoms of residue pairs with weaker communication upon MDR mutations are shown as red beads, and the Cα atoms of residue pairs with stronger communication upon MDR mutations are shown as cyan beads. A-C) The residue pairs with variance difference larger than 0.2 upon MDR mutations are shown for IDV (A), T12 (B), and RTRH (C). D-F) The residue pairs with variance difference larger than 0.15 upon MDR mutations are shown for IDV (D), T12 (E), and RTRH (F).

We then looked more closely at the protease-ligand interactions between the multi-drug resistant HIVPR strain and the three ligands, based on simulation trajectory snapshots. The 2D

diagrams (Figure 4-21, Figure 4-22, and Figure 4-23) show a simplified view of protease-ligand interactions. We could see that upon mutations, all three ligands are still able to remain hydrogen bonds to the bridging water, which bridges the flap tip residues Ile50, Ile149 (the Ile50 on the other monomer) and two oxygen atoms flanking the center of the ligand. They also remain hydrogen bonded to the active site, which is composed of Asp25 and Asp124, although the interaction between RTRH and protonated Asp124 (ASH124) did not show up on the 2D map (because the software does not contain parameters for protonated Aspartate). Apart from these conserved interactions, the native substrate clearly shows superior hydrogen bonding network to the protease backbone and sidechain atoms even in the presence of multi-drug mutations (Figure 4-23), while IDV and T12 are only forming a few interactions with the protease sidechain atoms. The 3D structure rendering (Figure 4-24, Figure 4-25, and Figure 4-26) confirmed the 2D diagram comparisons.

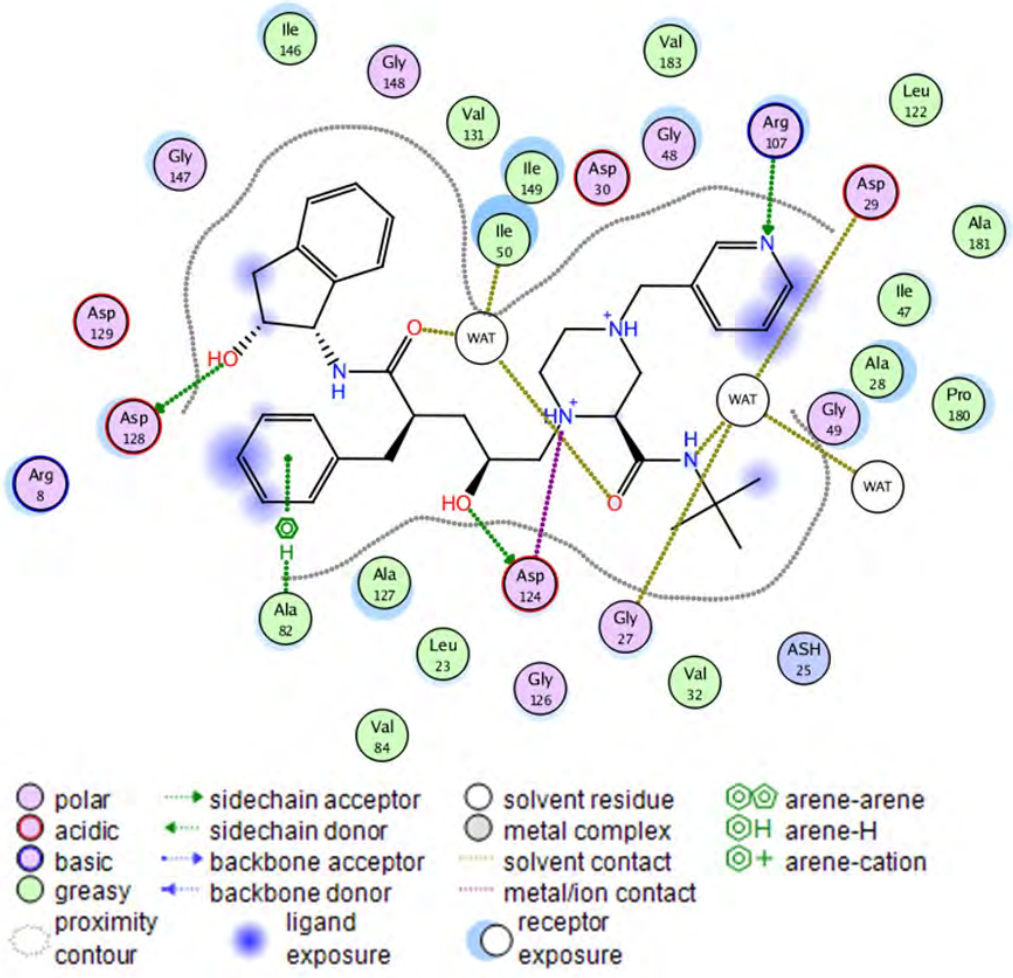

Figure 4-21 2D diagram of interactions between HIVPR multi-drug resistant strain and IDV. This and the following 2D diagrams of protein-ligand interactions were generated using MOE program. Analysis based on the structure extracted from the last frame of run 50.

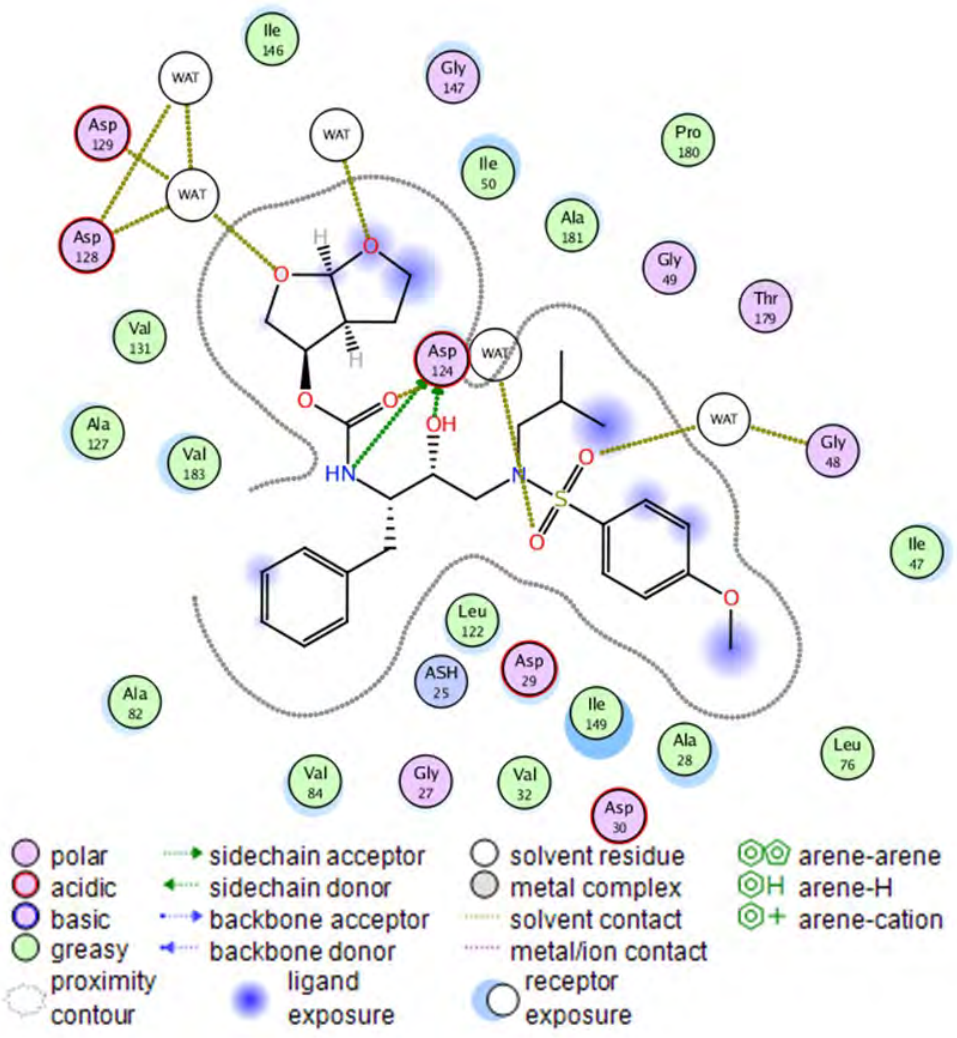

Figure 4-22 2D diagram of interactions between HIVPR multi-drug resistant strain and T12. Analysis based on the structure extracted from the last frame of run 50.

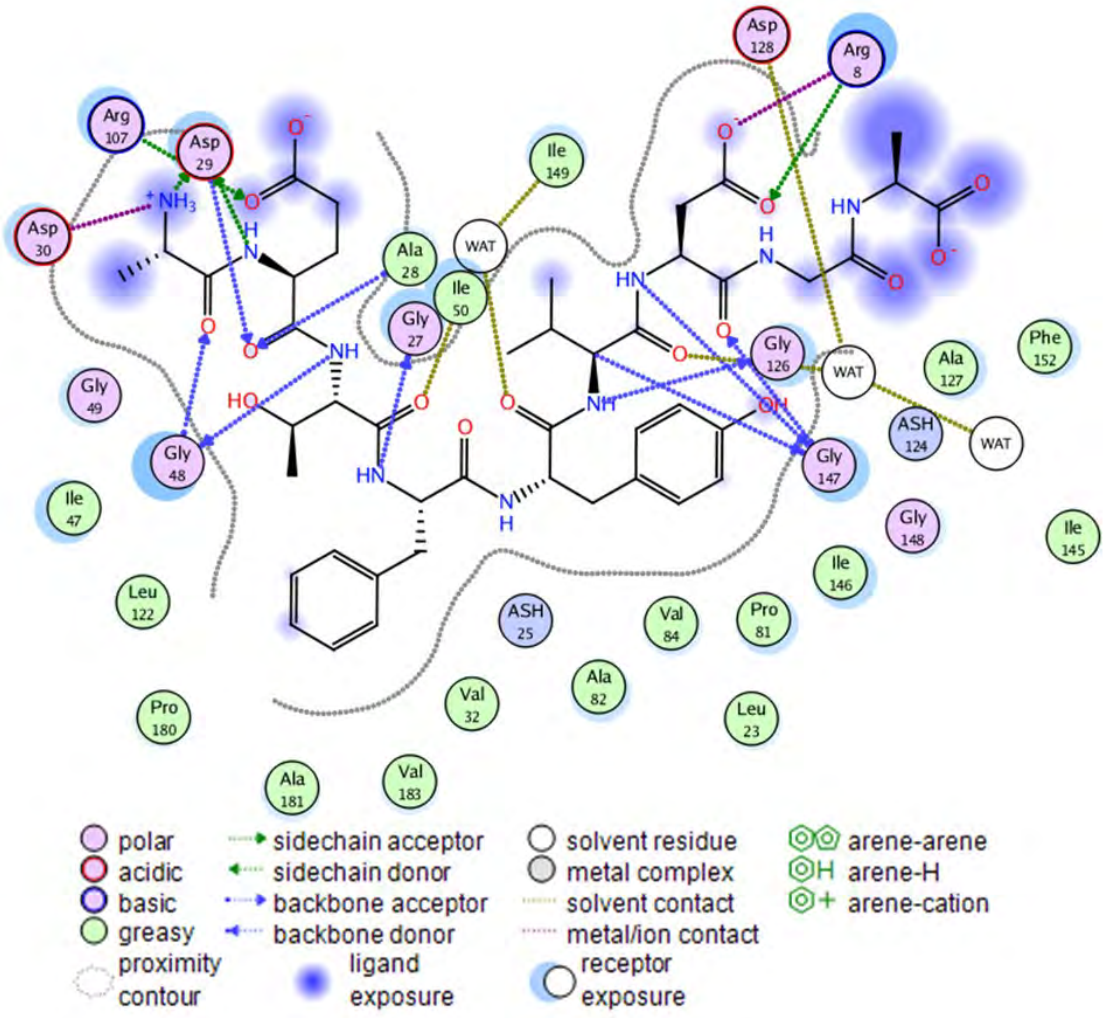

Figure 4-23 2D diagram of interactions between HIVPR multi-drug resistant strain and RTRH. Analysis based on the structure extracted from the last frame of run 50.

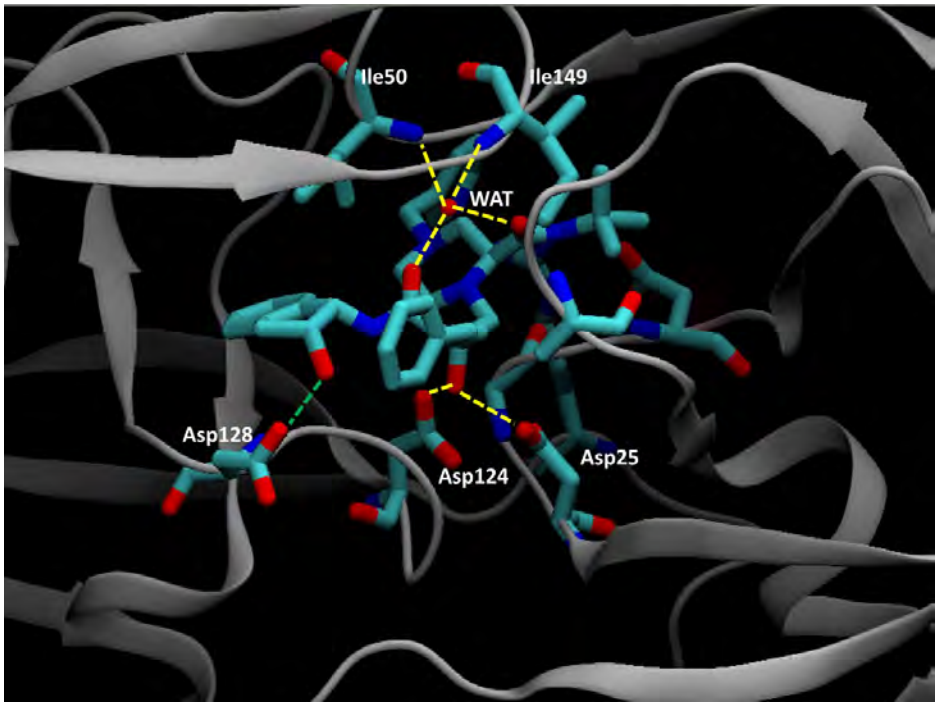

Figure 4-24 Structure of IDV bound to HIVPR multi-drug resistant strain. Structure extracted from the last frame of run 50. The hydrogen bonded residues are linked with dotted lines. To facilitate visualization the protease-ligand interactions in the back are not illustrated. Those hydrogen bonds between IDV and protease backbone, conserved water, or catalytic residues are shown in yellow lines. Those hydrogen bonds between IDV and protease sidechains are shown in green lines.

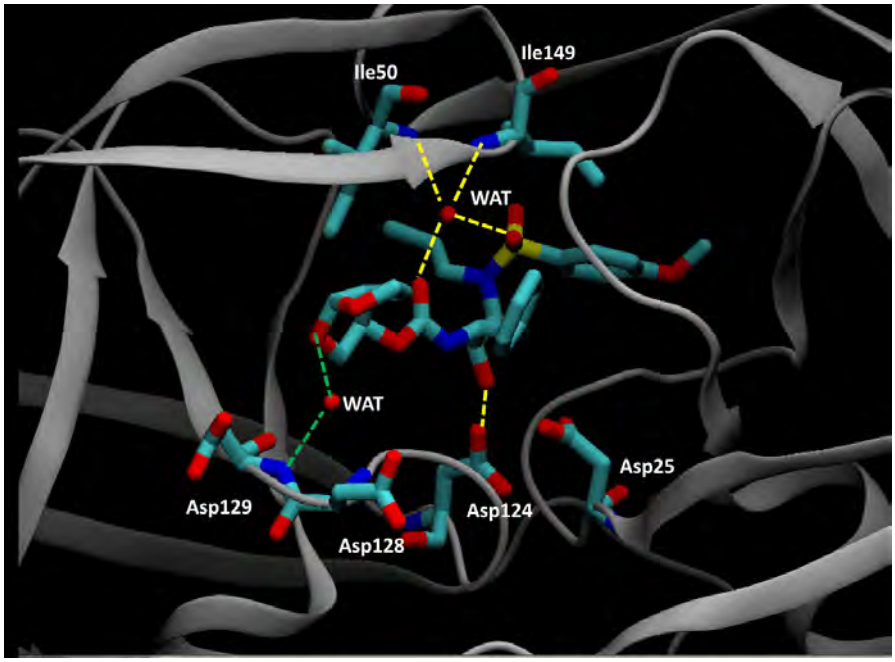

Figure 4-25 Structure of T12 bound to HIVPR multi-drug resistant strain. Structure extracted from the last frame of run 50. The hydrogen bonded residues are linked with dotted lines. To

facilitate visualization the protease-ligand interactions in the back are not illustrated. Those hydrogen bonds between T12 and protease backbone, conserved water, or catalytic residues are shown in yellow lines. Those hydrogen bonds between T12 and protease sidechains are shown in green lines.

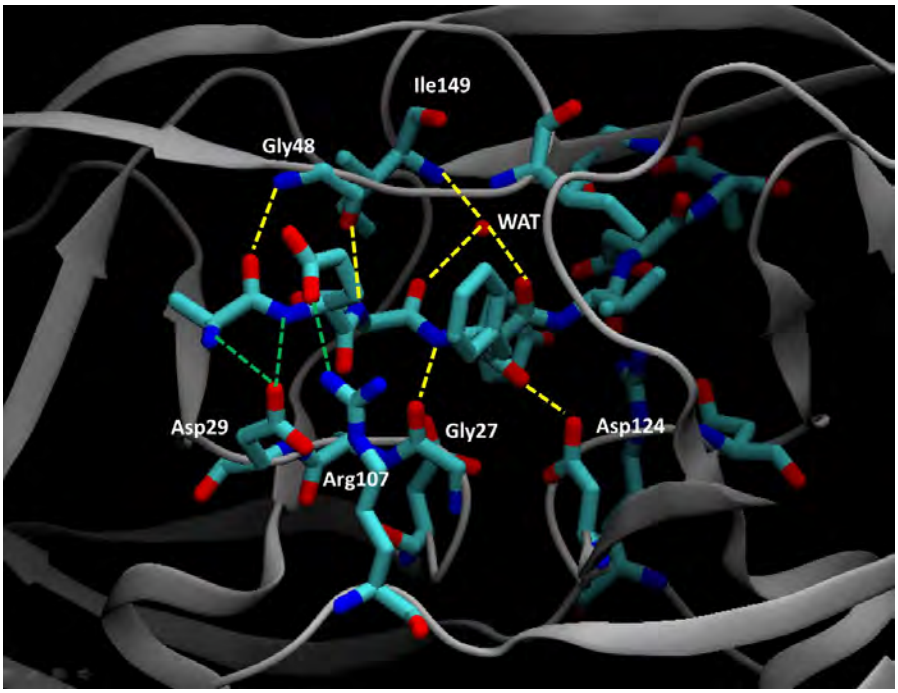

Figure 4-26 Structure of RTRH bound to HIVPR multi-drug resistant strain. Structure extracted from the last frame of run 50. The hydrogen bonded residues are linked with dotted lines. To facilitate visualization the protease-ligand interactions in the back are not illustrated. Those hydrogen bonds between RTRH and protease backbone, conserved water, or catalytic residues are shown in yellow lines. Those hydrogen bonds between RTRH and protease sidechains are shown in green lines.

It is interesting how the communication efficiency loss is correlated with the binding affinity change upon drug resistant mutations (Figure 4-15 and Figure 4-20, the ligand with more binding affinity loss also has more communication loss). Based on above comparison, we hypothesized that the second generation ligand T12 loss fewer communications than the first generation drug IDV mainly because it has relatively smaller size (77 atoms in T12 compared to 92 atoms in IDV) and thus greater flexibility to adapt for the change in the binding site upon drug resistance mutations. However, compared to the natural substrate RTRH, both T12 and IDV failed to make any additional hydrogen bonds to the protease backbone atoms, other than the water-bridged hydrogen bond to the flap tip backbone. Although the bis-tetrahydrofuranyl group in T12 was designed to make hydrogen bonds to backbone of Asp29 and Asp30 (Asp128 and Asp129 in Figure 4-25), the interactions were proved unstable in our simulations. The lack of hydrogen bonding to the protease backbone may also explain why the multidrug-resistant mutations around the termini region have effect on long-range communication in IDV and T12, but not on the natural substrate RTRH.

Overall, the comparison among simulation snapshots demonstrated the superior ability of the natural substrate RTRH to form hydrogen bonding network with the protease backbone and

sidechain atoms, while the protease drugs have rather a limited capability to make these hydrogen bonds. The analysis of long-range communication among protease residues suggests that the hydrogen bonding to protease backbone may help reduce the impact of drug-resistant mutations on protease dynamics. Therefore, the design of future inhibitors targeting protease needs to focus on creating more hydrogen bonds to the protease backbone atoms. The potential hydrogen bonding partners Gly27 and Gly48 (forming interactions with the natural substrate) might be better candidates than the backbone of Asp29 and Asp30 (targeted by inhibitor DRV and T12) because the former two residues are closer to the catalytic residues and thus could benefit more from the anchoring effect in the central region (conserved hydrogen bonds among the ligand, the protease flap, and the active site, Figure 4-27).

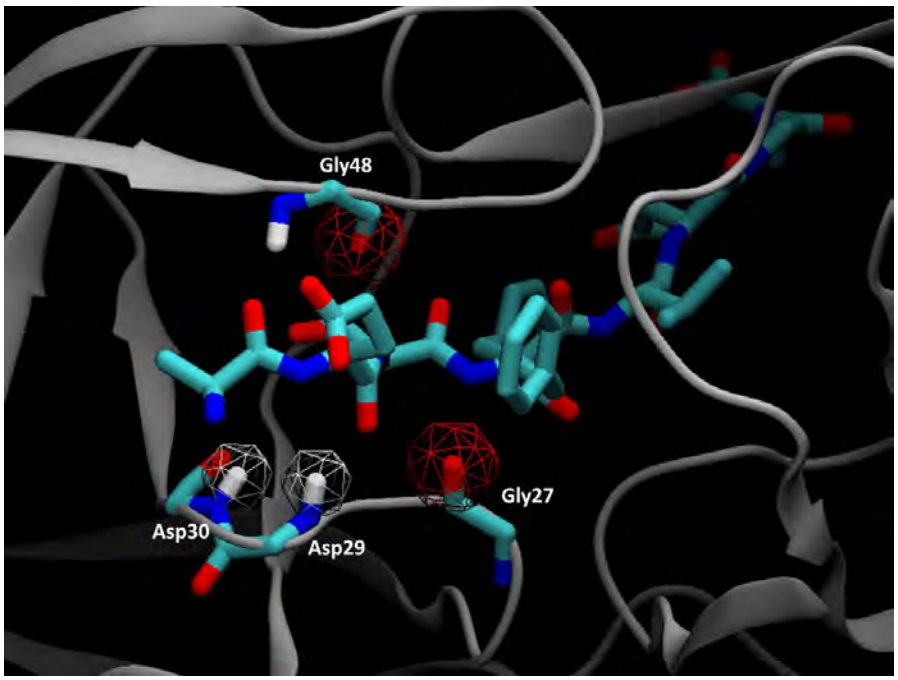

Figure 4-27 Protease backbone atoms as potential targets in protease inhibitor design. Backbone atoms of interest are illustrated as wireframe surface. Asp29 and Asp30 backbone hydrogen atoms were targeted by second generation drug DRV and T12, and Gly27 and Gly48 backbone oxygen atoms were proposed in this study.

4.4 Conclusions

In this study, we examined the interactions between the HIVPR protease and its ligands, to elucidate the reason for binding affinity loss upon drug resistant mutations. The findings here could serve as a guide for the design of next generation HIVPR inhibitors that have a broader spectrum against drug-resistant strains.

We first evaluated thermodynamic integration method in calculating the binding free energy change upon HIVPR drug-resistant mutations. The results were compared to available experimental data. The comparison confirmed that the thermodynamic integration method could be applied to accurately predict the binding free energy change in the HIVPR-ligand system. Therefore, the protocol used here could serve as a guideline for future studies on HIVPR binding free energy.

Apart from calculating binding free energy change using thermodynamic integration, we further performed explicit solvent MD simulations of protease-ligand complexes and decomposed the binding energy into per-residue basis using MMPBSA method. Energy decomposition was combined with structural inspection to explain the difference between KNI577 and KNI764 responses to active-site drug-resistant mutations. Results suggest that the favorable vdW binding between KNI764 and protease, compared to KNI577-protease, is due to the interaction between the toluene group of KNI764 and the protease flap residues. We found the loss in electrostatic energy upon active site mutations is mainly due to the V82F mutation at the active site, which disrupted its packing with the inhibitor and in turn hampered the hydrogen bonding at the P2 site nearby.

Since not all drug-resistant mutations occur near the active site, and previous studies have not been able to elucidate the contribution of mutations outside the active site, here we studied a multi-drug resistant strain that contains active-site and non-active-site mutations, and examined possible long range coupling between active-site and non-active-site residues by comparing the dynamics of the mutant to the wild type protease. We studied the binding of the protease to first generation inhibitor IDV, second generation inhibitor T12, as well as natural substrate RTRH. MMPBSA and energy decomposition were performed first to illustrate the impact of mutations on protease-ligand interactions. Then communication network analysis was conducted to detect any long range communications that are changed upon drug resistant mutations. Our results suggested that: 1) second generation drugs have better tolerance to drug resistant mutations than first generation drugs mainly because their relatively smaller size and better flexibility; 2) the natural substrate maintain superior communication profile upon drug resistant mutations, compared to protease inhibitors, mainly because its ability to form extensive hydrogen bonding network with the protease backbone and sidechain atoms. Based on structural comparison of protease-ligand complexes, we suggest designing new inhibitors to form hydrogen bonds with the backbone of G27 and G48, mimicking the natural substrate, which may be a better choice than the current selection of Asp29 and Asp39 backbone because of the better structural stability at the catalytic site.

Chapter 5 Comparative study of aspartic protease family and modeling the active site gating mechanism in non-HIV aspartic proteases

5.1 Introduction

Aspartic proteases are a family of enzymes that use two aspartate residues to catalyze the hydrolyzation of peptide bonds. The studies of aspartic proteases are largely driven by their connection to human diseases and cancers. HIV can cause acquired immune deficiency syndrome [167], and Human T-lymphotropic virus can cause adult T-cell leukemia [168], both of which need their proteases for viral infectivity. In the human body, β-secretases produce amyloid β that can cause Alzheimer's disease [169-171], renin levels can be modulated to alleviate high blood pressure [12, 18], and plasmepsins in malaria parasites are considered ideal targets fighting malaria [17, 56]. Structural information, mainly from crystallography and NMR, plays important role in aspartic protease inhibitor development [3, 18]. The most successful example is HIV protease (nine drugs approved by FDA, the latest being Darunavir approved in 2006), followed by renin (one drug named Aliskiren approved by FDA in 2007). Now the questions left are mainly how to improve the potency of existing drugs, and how to develop drugs for other aspartic protease targets.

A conceivable strategy to develop drugs for new targets is to utilize, as much as possible, the knowledge on existing successful targets. However, the knowledge translation between aspartic proteases may be hindered by the large sequence/structure variations, even though they belong to the same enzyme family.

Traditionally, members from aspartic protease family with known structures are broadly divided into two groups: HIVPR-like and pepsin-like aspartic proteases [11, 172]. The representative structures of these classes are shown in Figure 5-1 A and D, respectively. HIVPR-like aspartic proteases are found in retrovirus and share significant structural similarity to that of HIVPR. They are homodimeric, composed of two monomers. The structural elements of each monomer, can be simplified as the N-terminal, fulcrum, catalytic site, flap elbow, flap, cantilever, C-terminal helix, and C-terminal (Figure 5-1 A [42]). Two β-hairpin flaps, one from each monomer, cover the active site. The N and C terminals from both monomers are interleaved and form a β-sheet (Figure 5-1 E). Pepsin-like aspartic proteases are mostly found in eukaryotes (including animals, plants, and fungi) and share significant structural similarity to that of pepsin (Figure 5-1 D). They are bilobed, formed by two domains. The N-terminal domain has topology that more closely resembles homodimeric protease monomer than the C-terminal domain, but two domains align well in the core region and they differ mainly due to insertions/deletions on the protein surface (Figure 5-2). Only one β-hairpin flap from the N-terminal covers the active site. The termini of both domains make up a β-sheet (Figure 5-1 H). Note that the terminal arrangement of pepsin is significantly different from HIVPR: termini from two domains are well separated instead of interleaved, and the β-sheet has two more β strands.

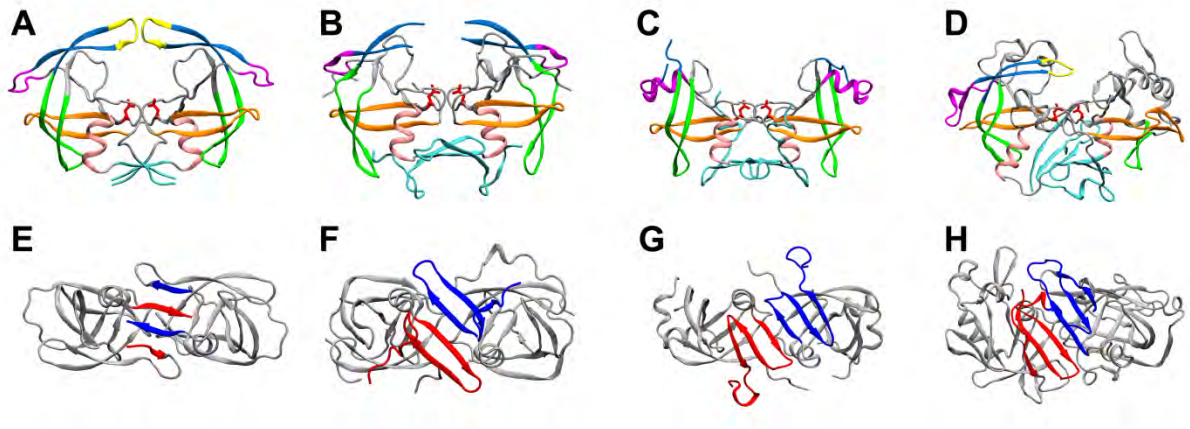

Figure 5-1 Front view (A-D) and bottom view (E-H) of apo crystal structures from HIVPR (AE, PDB ID: 1HHP), MLVPR (BF, PDB ID: 3NR6), Ddi1 central RVP domain (CG, PDB ID: 2I1A), and pepsin (DH, PDB ID: 1PSN). Proteins are shown in cartoon representation. Front view (upper panel): catalytic residues are shown in licorice representation. Color codes and naming of protein segments referenced those for HIVPR defined by Hornak et al. (fulcrum in orange, elbow in magenta, flap in blue, flap tip in yellow, cantilever in green, C-terminal helix in pink, and N/C terminals in cyan). The N terminal domain of pepsin is shown on the left in figure D. Bottom view (lower panel): the terminals from A and B monomers are colored red and blue, respectively. For pepsin, terminals from N and C domains are colored red and blue, respectively. Note that the N-terminals of MLVPR and Ddi1 RVP domain are not involved in forming the terminal β-sheet, so they are not colored.

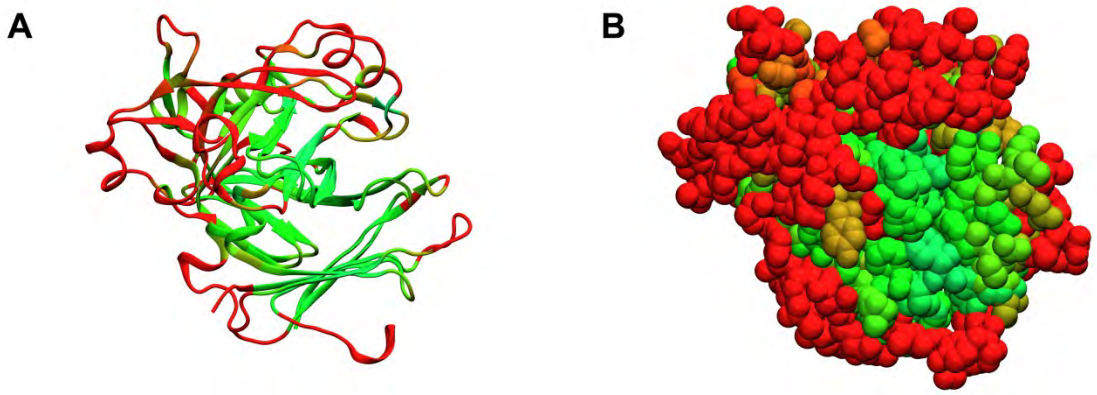

Figure 5-2 Cartoon (A) and vdW (B) representations of the alignment of two pepsin domains (PDB ID: 1PSN). RGB color coding is used to reflect the alignment score: the worst alignment is shown in red.

Because of the large structural difference between HIVPR-like and pepsin-like aspartic proteases, previous comparative studies on aspartic proteases have been done on each group separately rather than across all members [10, 11, 172, 173]. However, several recent studies suggest a closer connection between HIVPR-like and pepsin-like aspartic proteases, which makes the structural grouping much more difficult than previously. On the one hand, among

retroviral proteases, newly crystallized murine leukemia virus protease (MLVPR, Figure 5-1 BF) has termini topology resembling pepsin-like proteases. The structure was solved recently to seek inhibitors of xenotropic murine leukemia virus-related virus (XMRV) [174]. Although XMRV has later been identified as mouse DNA contamination and the linkage of XMRV to human diseases has been completely disproved [175, 176], the crystal has the same sequence as MLVPR and becomes the first structure on the gammaretrovirus branch. Therefore, we refer the structure as MLVPR and compared it to other aspartic proteases. Surprisingly, MLVPR has unique terminal arrangement (Figure 5-1 EFH and Table 5-1) resembling pepsin-like aspartic proteases. On the other hand, eukaryotic protein Ddi1 (DNA damage inducible protein 1), which has HIVPR-like fold and pepsin-like termini (Figure 5-1 lower panel and Table 5-1), is suggested to be an aspartic protease. Ddi1 is composed of N-terminal ubiquitin-like domain, C-terminal ubiquitin-associated domain, and a central retroviral protease-like (RVP) domain (Figure 5-1 C) [177-179]. It is involved in protein targeting to the proteasome, cell cycle control, and protein secretion [180]. The latter two functions have been recently suggested to be dependent on its catalytic activity [180, 181].

Protein name	HIVPR	MLVPR	Ddi1	pepsin
homodimeric	yes	yes	yes	no
N-ter involved in β sheet	yes	no	no	yes
Interleaved β sheet	yes	no	no	no
Number of β strand	4	4	6	6

Table 5-1 Structural comparison among aspartic proteases including HIVPR, MLVPR, Ddi1, and pepsin. Properties compared include whether they are homodimeric or bilobed, whether the N termini are involved in forming termini β sheet, whether the termini β sheet is interleaved, and the number of β strand involved in the termini β sheet.

To renew and organize the knowledge on aspartic protease family, so as to translate knowledge from successful drug targets to others, a more systematic comparison is needed. Fortunately, with increasing number of structures solved and developments in structural bioinformatics [182], previously unmanageable sequence alignment due to low sequence identity, is now available through structural similarity search [183]. The alignment of protein families can shed light on evolution, and provide sequence signatures that can be used to guide drug development [184, 185]. Accordingly, in this study we generated evolutionary profiles for aspartic protease family, using a non-redundant set of representative structures [186]. We created a maximum-likelihood phylogenetic tree, proposed a hypothesis on aspartic protease evolution, and discussed conserved residues of different branches.

Apart from sequence and structure comparison, since the ligand binding usually involves dynamic enzyme-ligand interactions, the comparison of enzyme dynamics may be more fundamental in determining their similarity for drug development. For example, through molecular dynamics simulations starting from crystal structures, it was proposed that HIVPR has mainly three flap conformations: closed, semiopen, and wide-open structures (Figure 5-3) [42]. The former two conformations are captured by crystal structures [100, 104], while the wide-open structure is validated through site-directed spin labeling EPR experiments [67, 97]. The relationship of these three states was also hypothesized: in the unbound form, the protease flaps are mostly in the closed or semiopen form, with the semiopen form easier to transit into the

wide-open form. The transient flap opening enables the substrate to enter, after which the flaps reclose and the catalytic reaction occurs (Figure 5-3 DH) [42, 43]. Flap control is considered the most important ligand gating mechanism in HIVPR [71]. Comparing flap dynamics among different proteases would not only strengthen our understanding of protease-inhibitor binding, but also help verify existing model if dynamics from different aspartic proteases are homologous. Therefore, we modeled MLVPR and studied its dynamics using MD simulations, because of its difference from HIVPR and smaller disordered regions compared to Ddi1. We incorporated both the apo crystal structure 3NR6 (Figure 5-4 [174]), which has disordered flaps but was the only structure available at the beginning of our study, and a holo crystal structure 3SM2 [187] into our study. We also performed MD simulations of HTLVPR and SIVPR, with HIVPR simulations as a control. We combined implicit and explicit solvent simulations to get sufficient yet accurate sampling of flap dynamics within appropriate time frame. Overall, the flap dynamics and active site gating in MLVPR, HTLVPR, and SIVPR show significant similarity to those of HIVPR, but the specific inter-flap interactions vary considerably among different retroviruses. Apart from modeling retroviral proteases, we also investigated the active site gating mechanism in pepsin-like aspartic proteases, which is far less well understood than that of HIVPR [55, 188-190]. We used β-secretase 1 (BACE) as a model system to test our hypothesis of active site gating based on sequence conservation. The apo and holo BACE simulations provided, for the first time, insights into its ligand binding process.

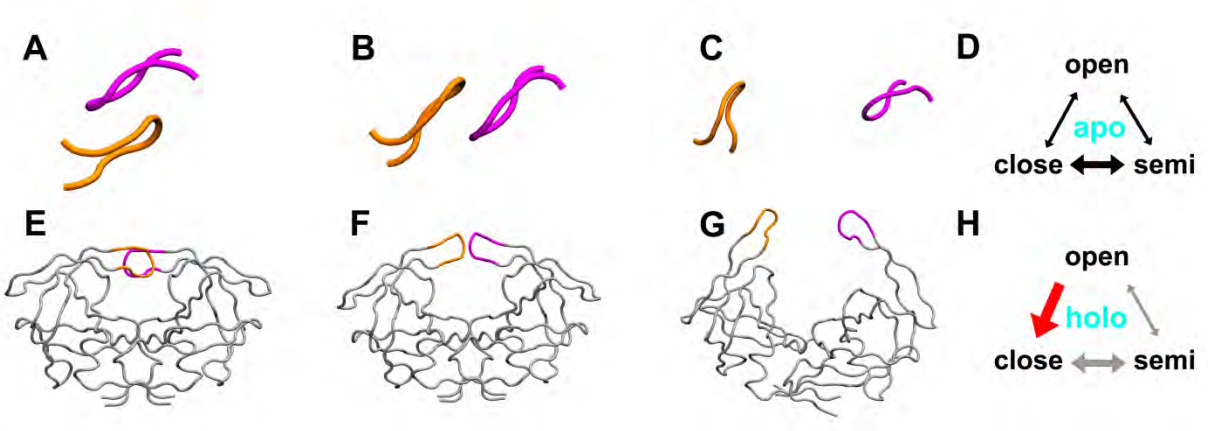

Figure 5-3 Top view of flap tips (A-C), and front view (E-G) of the whole HIVPR dimer. Closed (AE, PDB ID: 1HVR), semiopen (BF, PDB ID: 1HHP), and wide-open (CG, snapshot taken from Shang et al. 2011 JMGM paper) flap conformations are shown. Flap tips from two monomers are colored orange and magenta, respectively. Figure D: in apo state, HIVPR flaps are mainly in closed or semiopen conformation. Figure H: ligand binding shifts the equilibrium to force flap closing, which facilitates catalytic reaction.

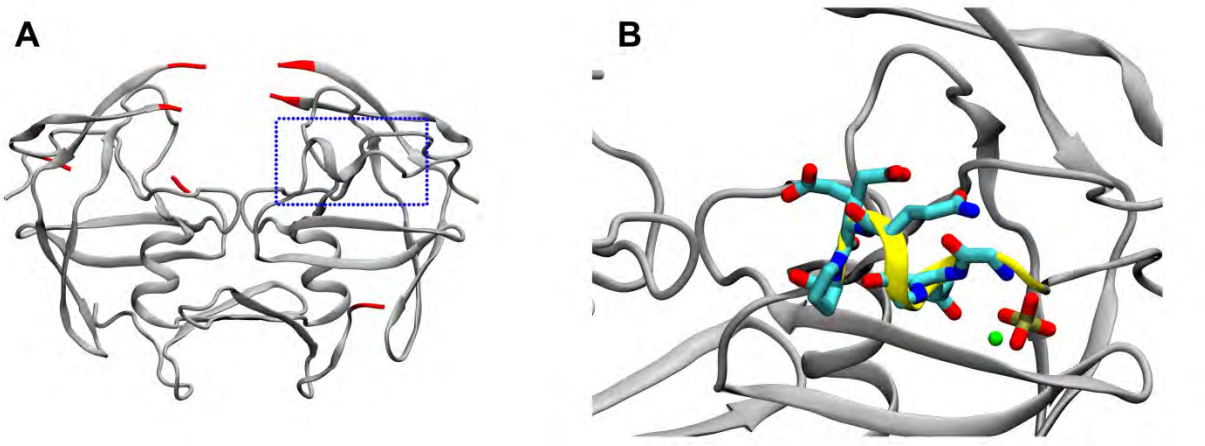

Figure 5-4 MLVPR crystal 3NR6. A) Disordered regions in 3NR6 crystal, residues adjacent to disordered regions are colored red. B) A close up view of the stabilized N-terminal helix (blue box in figure A). Co-crystallized ions near N- terminal helix in crystal 3NR6 are shown: chloride ion in green and phosphate in tan and red.

5.2 Methods

5.2.1 Evolutionary profile of aspartic proteases

Most steps below used multiseq[191] plugin in VMD [1], unless otherwise noted.

Generate non-redundant structure representatives of aspartic protease family. Because of the large difference in sequence length, it is not possible to retrieve both homodimeric and bilobed aspartic proteases using one sequence as BLAST seeding. Therefore we used one homodimeric (PDB ID: 1FMB [192] and one bilobed (PDB ID: 1PSN [193]) aspartic protease. For each seeding, we BLAST searched PDB database with E value criteria set to 10. QR factorization [186] was performed to filter out redundant hits and limit number of hits to around 100, which was followed by sequence alignment using ClustalW [194]. Then QR factorization was performed again to retain only hits with sequence percentage identity below 50. Zymogens (eukaryotic aspartic protease precursor) and HIVPR-like aspartic protease dimers tethered by chemical bonds were discarded. In the end, 8 hits retained from BLAST search using 1FMB (1FMB, 2I1A [178], 3NR6 [174], 2B7F [8], 3FIV [195], 1BAI [196], 1IVP [197], and 2P3D [198]), and 7 hits retained from BLAST search using 1PSN (1PSN, 3EXO [199], 2RMP [200], 3LIZ [201], 2QZX [202],1WKR [203], and 3C9X [204]).

Structure-guided sequence alignment. Chain A of each PDB structure obtained in the previous step were aligned using STAMP [205] method implemented in multiseq. To ensure the alignment efficiency, homodimeric and bilobed aspartic proteases were aligned separately first, and then all the structures were aligned. Because sequence obtained from PDB files may be incomplete at disordered regions, we retrieved FASTA sequence for each entry from PDB website instead. Expression tags, if found, were removed. FASTA sequences were then aligned using ClustalW, using the structure alignment profile as guide. The resulting minimum, average,

and maximum sequence percentage identity among all 15 sequences were 2, 11, and 47, respectively. The alignment is illustrated using software Jalview [206] in Figure 5-5, Figure 5-6, and Figure 5-7.

Maximum-likelihood phylogenetic tree and bootstrapping were generated using RAxML [207]. For maximum-likelihood phylogenetic tree, hill climbing algorithm and PROTMIXWAG model were used to generate 1000 parsimony trees. The bootstrapping values were then calculated for the best likelihood tree, based on 1000 bootstrap analyses, using random seed 12345. At last the bootstrapping values were mapped onto the maximum-likelihood phylogenetic tree.

Sequence conservation was based on the sequence alignment result, where conserved residues are shaded blue (Figure 5-5, Figure 5-6, and Figure 5-7). Since only 15 sequences were considered during structure-guided sequence alignment, we validated these conserved residues using sequence conservation data. Each of the 15 PDB IDs was fed into ConSurf server [208-210] for its sequence conservation information. Because only chain A from different molecules were aligned, the C-terminal domain of bilobed aspartic proteases were excluded since they would not aligned to homodimeric aspartic proteases. Not all searches found enough (at least 6) homologous sequences to calculation conservation score, the successful searches are summarized in Table 5-2. Only putative conserved residues having high conservation score (at least 8 out of 9) in all available conservation data are considered conserved. The parameters for ConSurf search are: MAFFT method [211] for sequence alignment, BLAST search in UNIREF90 database, PSI-BLAST E value as 10^{-8}, iterate only once, maximum percentage of sequence identity 90, minimum percentage of sequence identity 70, and maximum number of homologues 150. We set high criteria for E value and sequence identity, to focus on sequences closely related to the seed. The conservation score calculation used default parameters.

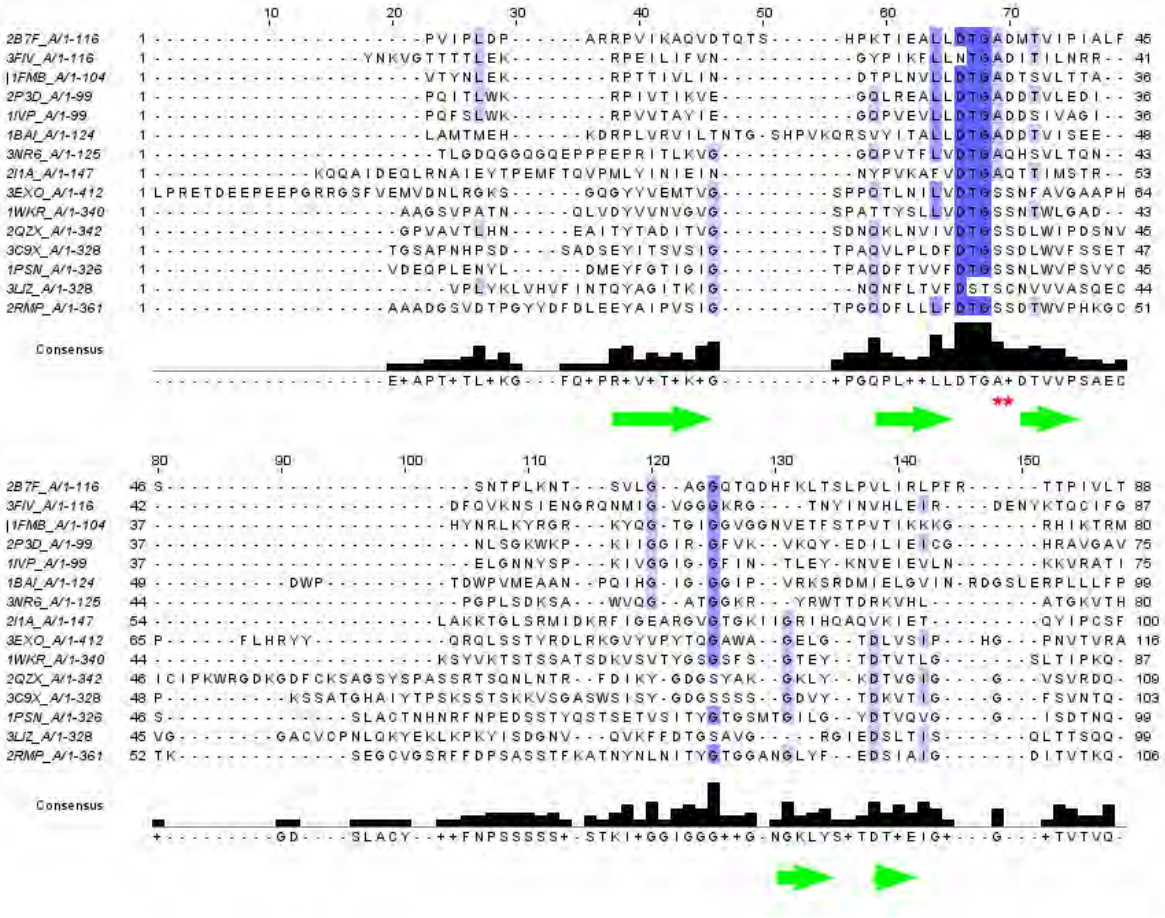

Figure 5-5 Sequence alignment of aspartic protease family representatives – part 1 of 3. PDB ID, chain ID and sequence length are shown on the left. Conserved residues are in blue background in the alignment. Consensus sequence is shown as a separate row below the alignment. Secondary Structures that aligned well are indicated by green arrow (β strand) and pink curve (helix) below the consensus sequence. Sequence signatures are indicated by red stars.

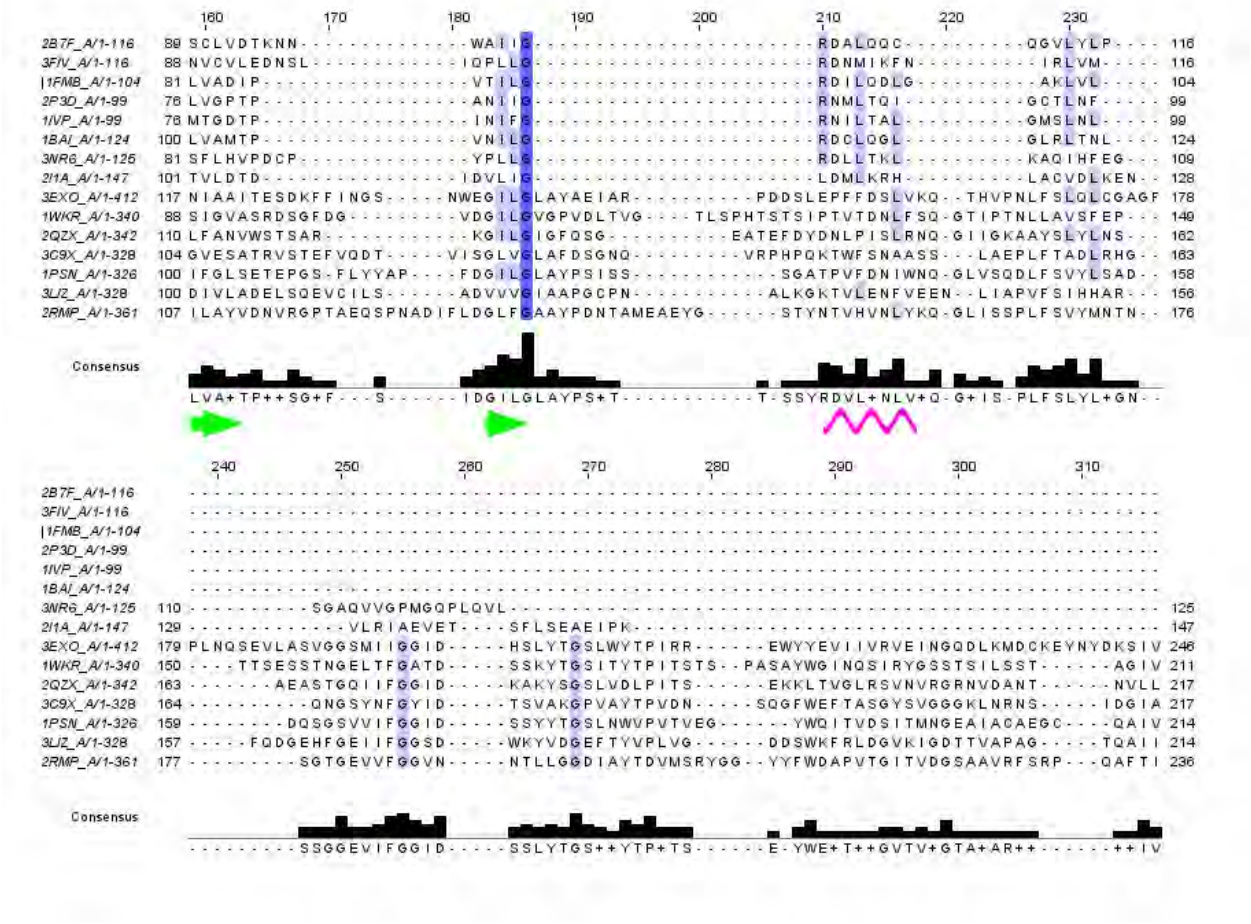

Figure 5-6 Sequence alignment of aspartic protease family representatives – part 2 of 3. PDB ID, chain ID and sequence length are shown on the left. Conserved residues are in blue background in the alignment. Consensus sequence is shown as a separate row below the alignment. Secondary Structures that aligned well are indicated by green arrow (β strand) and pink curve (helix) below the consensus sequence.

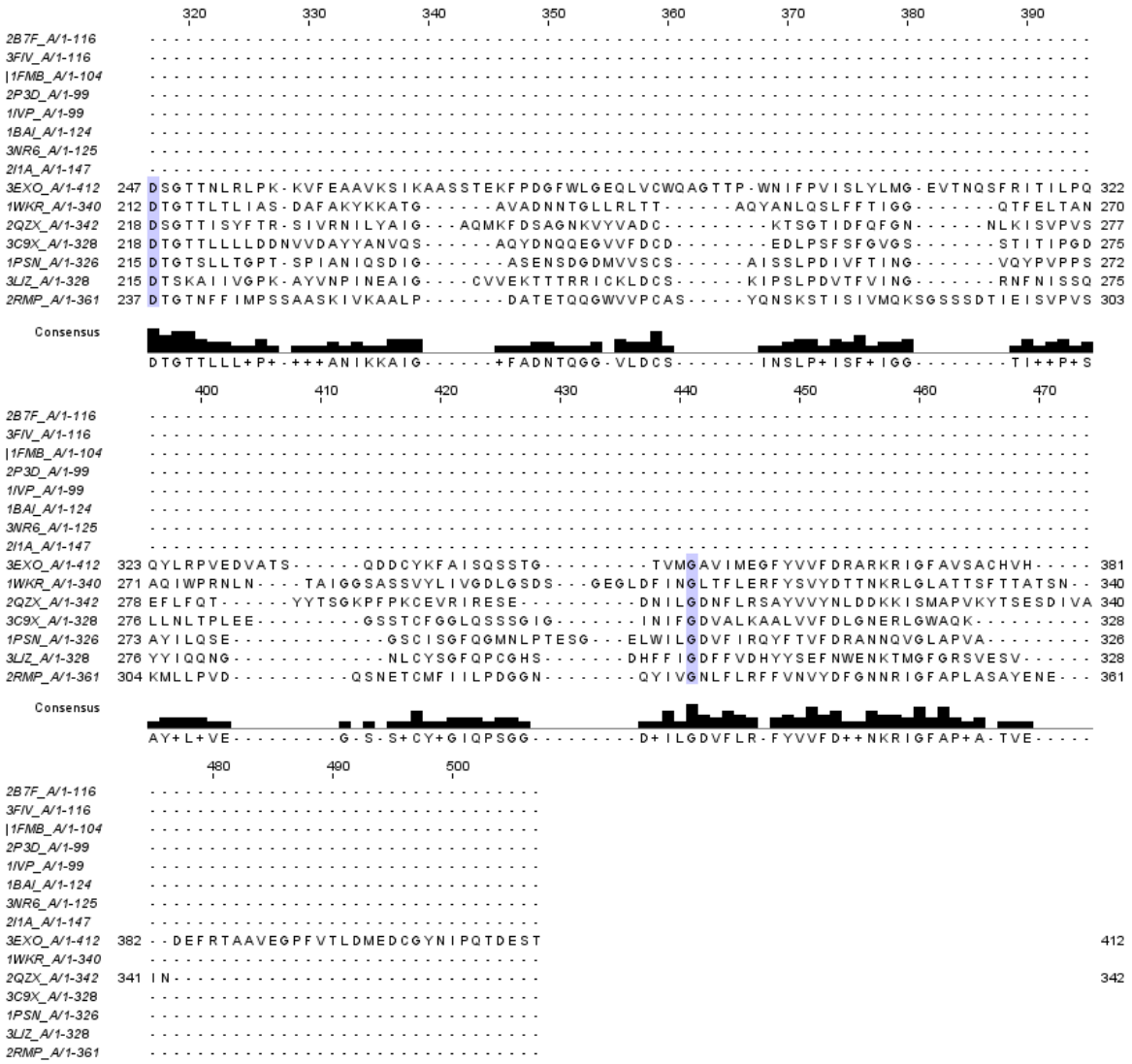

Figure 5-7 Sequence alignment of aspartic protease family representatives – part 3 of 3. PDB ID, chain ID and sequence length are shown on the left. Conserved residues are in blue background in the alignment. Consensus sequence is shown as a separate row below the alignment.

Category	PDB ID (query)	Query hits used for conservation calculation
HIVPR-like aspartic protease	2P3D (HIV-1 PR)	150
HIVPR-like aspartic protease	1IVP (HIV-2 PR)	59
MLVPR or Ddi1	2I1A (Ddi1)	6
pepsin-like aspartic protease	1PSN (pepsin)	20
pepsin-like aspartic protease	3EXO (BACE1)	16

Table 5-2 Statistics of sequence conservation calculation on ConSurf server.

5.2.2 Simulations of apo MLVPR

5.2.2.1 Starting structures

Simulation starting structures were built using two MLVPR crystals: apo MLVPR (PDB ID:3NR6 [174]) and MLVPR bound to HIVPR-inhibitor amprenavir (APV, PDB ID: 3SM2 [187]). MolProbity [212] server was used to add hydrogen atoms and flip Asn/Gln/His when necessary. The crystal 3NR6 has missing density in flap tips as well as N/C terminals. Because of the low sequence identity to known structures, it is hard to build the flap tips using homology modeling. Instead, we grafted HIVPR flap tip coordinates to MLVPR using backbone RMS fit. HIVPR flap sequence is MIGGIGGFIK, and MLVPR flap sequence is WV**QGATG**GKR (disordered regions in bold, although in monomer A only the four residues **GATG** were disordered). During grafting, backbone atoms of residues flanking the disordered flap segment, two residues on each side, were aligned between MLVPR and HIVPR, and then residue coordinates of HIVPR were used to fill MLVPR disordered region. Virtual mutations using swissPDB [132] were done to match MLVPR sequence. Different HIVPR crystals, namely 1HHP, 1NH0, and 1G6L, were used to model the flaps, and model built with 1G6L [121] was retained because it has least atom clashes (all among hydrogen atoms). Disordered regions in N/C terminals were modeled using crystal symmetry (monomer B N-terminal used monomer A as the template, etc.). The resulting full length model of MLVPR from crystal 3NR6 is shown in Figure 5-8. We also built full length MLVPR from holo crystal 3SM2. In this bound structure, flaps are resolved, but N/C terminals have missing density in both monomers. Modeling flexible N-terminal without introducing clashes to the rest residues in 3SM2 is challenging. We used backbone of residue 14 and 15 for fitting, which were relatively immobile in simulations from 3NR6 model, and utilized 3NR6 simulation snapshots besides the crystal to find a terminal conformation that doesn't introduce clashes. Finally, cluster analysis representative structure with closed flap conformation (see below for cluster analysis description) was used to model N and C terminals of 3SM2 crystal, which introduced no clashes. Apo MLVPR simulations from 3SM2 structure had active site residues virtually mutated to Asn, to be consistent with previous simulations [130]. Apo MLVPR simulations from 3SM2 structure had active site modeled as diprotonated, which was shown to describe HIVPR-APV interaction closest to the crystal structure [151].

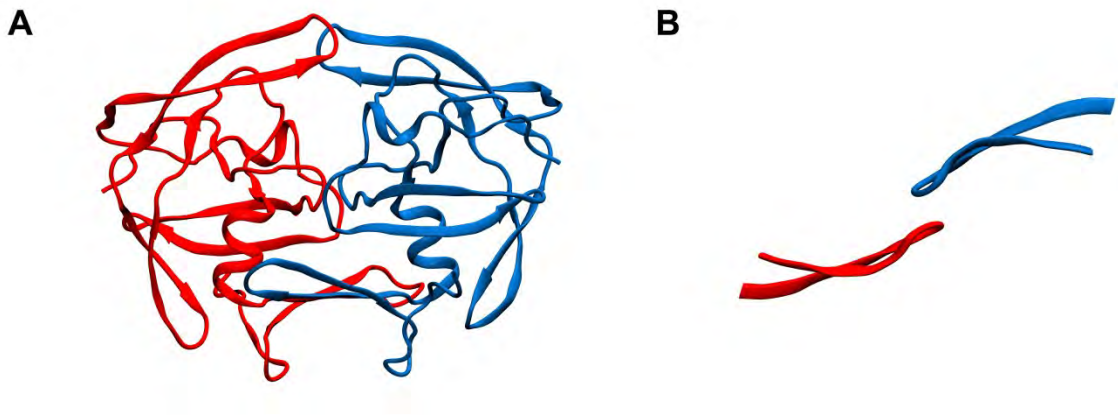

Figure 5-8 Modeled full length MLVPR from crystal 3NR6. A) Front view. B) Flap tip top view. Two monomers are colored red and blue, respectively.

5.2.2.2 Simulation system setup

Tleap program in Amber 11 was used to generate topology and coordinate files for simulations. All simulations used ff99SB protein force field [64]. For generalized Born (GB) [58-60] simulations, mbondi2 intrinsic radii set [63, 98, 99] with modifications to arginine polar hydrogen atoms[130] was used. Water molecules in the crystal structure were retained, then a truncated octahedron TIP3P [65] water box was used to solvate the solute with 8 Å minimum clearance from the box edges, adding 6941 and 9539 water molecules for 3NR6 and 3SM2 systems, respectively. Note that many more water molecules were added for 3SM2 runs because N-terminals in 3SM2 model were not as compact as those in 3NR6 model. Chloride ions were added to neutralize the system [213]. Missing hydrogen atoms were added by tleap.

5.2.2.3 Simulation parameters

For GB simulations, GB-OBC [63] implicit solvent model (igb=5 in AMBER) was used with 1fs time step. All bonds involving hydrogen atoms had SHAKE length constraint with geometry tolerance of 10^{-5}. No cutoff for long range electrostatic/vdW interactions was applied. Pair interactions that were involved in effective radii calculation had 25 Å distance cutoff. Forces related to effective radii calculation, along with each pair interaction whose distance was greater than 15 Å, were updated every 4 steps. Langevin temperature control at 300 K with collision frequency of 1 ps^{-1} was used. Different initial velocity seeds were used to unsynchronize Langevin dynamics [103]. Surface area was computed using LCPO[214] model (gbsa=1). Salt concentration was set to 1 M. For explicit solvent (EXP) simulations, the same SHAKE constraint as GB simulations was used, with 2 fs time step. Particle-Mesh Ewald (PME) [29-32] was used for long range electrostatic interactions. 8 Å cutoff was applied to vdW interactions. Berendsen temperature and pressure control [25] were applied to maintain the system at 300 K and 1 atm.

5.2.2.4 Equilibration

Energy minimizations and restrained simulations were performed before each GB or EXP unrestrained simulations, to optimize starting structure, heat the system to desired temperature, and equilibrate solvent and solute residues that needed to be modeled. EXP simulations were initiated from 3NR6 and 3SM2 solvated structures described above. Each system went through

10,000 step energy minimization, heating from 50 K to 300 K during 100 ps, and 600 ps restrained simulations with decreasing positional restraints (from 4 to 1 kcal/mol•Å^2). Restraints were added first on solute heavy atoms then on backbone atoms only. GB simulations started from the representative structure (with closed flap conformation, Figure 5-15 CF) of 3NR6 EXP simulation. Since it is a simulation snapshot, no heating was needed. A 500 ps restrained simulation with 0.1 kcal/mol•Å^2 positional restraints on backbone atoms were performed prior to unrestrained GB simulations [215].

5.2.2.5 Simulations

For explicit solvent simulations, three independent runs (with different velocity seedings) were carried out for 3NR6 system, and two independent runs were carried out for 3SM2 system. Simulation lengths for two systems were 400 ns and 100 ns, respectively. For GB simulations, five independent runs (with different velocity seedings) were initiated from the cluster representative structure (with closed flaps, Figure 5-15 CF). The first run was ~15 ns, while the remaining four runs were ~30 ns each. See Table 5-3 for a summary of MLVPR simulations performed in this study.

Solvent model	Starting structure	Simulation length in ns	Number of independent runs
EXP	model built from 3NR6	400	3
GB	cluster representative structure(close1)	15-30	5
EXP	model built from 3SM2	100	2

Table 5-3 Summary of MLVPR simulations.

5.2.3 Simulations of apo HIVPR, apo HTLVPR, and apo SIVPR

Since holo crystal structures for these retroviral proteases are available, which do not contain disordered region, the simulations were started from the crystal structures without the need of modeling any backbone atoms. Starting structures of HIVPR, HTLVPR, and SIVPR simulations were built from crystal structures 1NH0, 2B7F, and 1YTI, respectively. Active site was modeled as D25N mutant.

For each protease, four independent GB simulations, ~15ns each, were carried out first. Equilibration and simulation parameters were the same as MLVPR simulations. Then cluster analysis was performed on resulting GB simulation trajectories to select the cluster with semiopen flap conformation. The representative structure of that cluster was used as the starting structure of explicit solvent simulations. Two independent simulations in explicit solvent were carried out (~ 1 us each). The simulation parameters were the same as MLVPR simulations.

5.2.4 Simulations of apo and holo β-secretase 1

Full length model for β-secretase 1 (BACE) was built from apo BACE crystal structure 1W50 (Figure 5-9) [216], the disordered loop region was modeled using holo BACE crystal structure 1SGZ [188]. Explicit solvent simulations were carried out for apo and holo BACE.

For apo BACE explicit solvent simulations, wild type BACE simulations and double mutant (S35A/D83A) BACE simulations were performed, two independent runs for each

sequence. Equilibration and simulation parameters were the same as those in MLVPR explicit solvent simulations (page 108).

For holo BACE explicit solvent simulations, the snapshots at 78 ns of two wild type apo BACE simulations were used as the two protease starting structures. Using apo simulation snapshots as starting structures ensures that these structures were well equilibrated. The natural substrate (Glu-Val-Asn-Leu-Ala-Ala-Glu-Phe) coordinates were extracted from crystal structure 1FKN [217]. For each protease starting structure, the natural substrate was docked to the protease using MOE software, and two docked poses were retained. We selected the poses with ligand outside the active site, to evaluate the process of ligand binding in BACE. In total four holo BACE simulations were carried out: wild type protease simulated with natural substrate, two protease starting structures, and two substrate poses for each protease starting structure. The substrate Cα atoms were restrained to the Cα atom of Asp32 (one catalytic aspartate) when their distance goes beyond 45 Å (force constant 10 kcal/mol•Å^2), to prevent the substrate from drifting too far away from the proteases. Equilibration and simulation parameters were the same as those in MLVPR explicit solvent simulations (page 108).

Figure 5-9 Crystal structure 1W50. Heavy atoms near Tyr71 are shown in licorice representation. Structured water molecules are retained as red dots.

5.2.5 Analysis

5.2.5.1 Cluster analysis

clusters were formed with bottom-up approach using similarity (RMSD) cutoff: each structure was initially assigned to a distinct cluster, followed by calculation of averaged average RMSD between all cluster pairs, then cluster pair with the smallest RMSD was merged until the most similar cluster pair exceeded the similarity cutoff.

MLVPR cluster analysis was done on 3NR6 runs using RMSD cutoff 2.0 Å. Run 1 has different flap conformation sampled compared to run 2 and 3, so run1 were clusters separately, and then run2 and run3 were clustered together. Later, flap conformations sampled in 3NR6 run2

and 3 were validated by clustering them together with two runs from 3SM2 system (3SM2 has complete flap coordinates), and 3SM2 crystal structure. RMSD cutoff was set to 1.5 Å (Table 5-4). All RMSD were based on coordinates of Cα atoms of residue 53 to 62 on both monomers.

For HIVPR, HTLVPR, and SIVPR GB simulations, cluster analysis was done on flap tip Cα atoms (10 residues on each flap).

5.2.5.2 RMSD and distance calculation

Flap RMSD and atom distance were calculated using ptraj program in AMBER.

MLVPR flap RMSD used Cα atoms of residue 53 to 62 on both monomers, with 3SM2 crystal and semi2 structure (from cluster analysis, see the results section) as closed and semiopen flap conformation references, respectively.

HIVPR, HTLVPR, and SIVPR flap RMSD calculations also used also Cα atoms on flap tips, with 10 residues on each flap. Crystal structures were used for closed flap conformation reference. Semiopen flap structures obtained from clustering GB simulations were used as semiopen flap conformation reference.

5.3 Results

5.3.1 Evolutionary profile of aspartic proteases

Members of aspartic protease family have considerable sequence/structure differences that hinder direct sequence alignment. Therefore we attempted to retrieve non-redundant structure representatives from PDB, and then use structure alignment profile to guide the otherwise error-prone sequence alignment of representatives. Because of the sequence length difference, two BLAST searches were performed on PDB database, using 1FMB and 1PSN as seeding, to retrieve representatives for homodimeric as well as bilobed members. In total 15 structures were retrieved, and their structural alignment is shown in Figure 5-10 A. The N-terminal domains of bilobed aspartic proteases align well with homodimeric aspartic proteases in the core region (blue region in Figure 5-10 A). Structure-guided sequence alignment of these 15 molecules (Figure 5-5, Figure 5-6, and Figure 5-7) was used to generate maximum-likelihood phylogenetic tree, which is shown in Figure 5-10 B. Although most bootstrap values are not high, the phylogenetic tree clearly separates bilobed aspartic proteases (lower 7 molecules) from homodimeric aspartic proteases (upper 8 molecules), and nicely separates eukaryotic molecules (names in blue) from retroviral aspartic proteases (names in yellow). The tree also suggests that MLVPR and Ddi1 RVP domain may be in evolutionary branches distinct from other aspartic proteases.

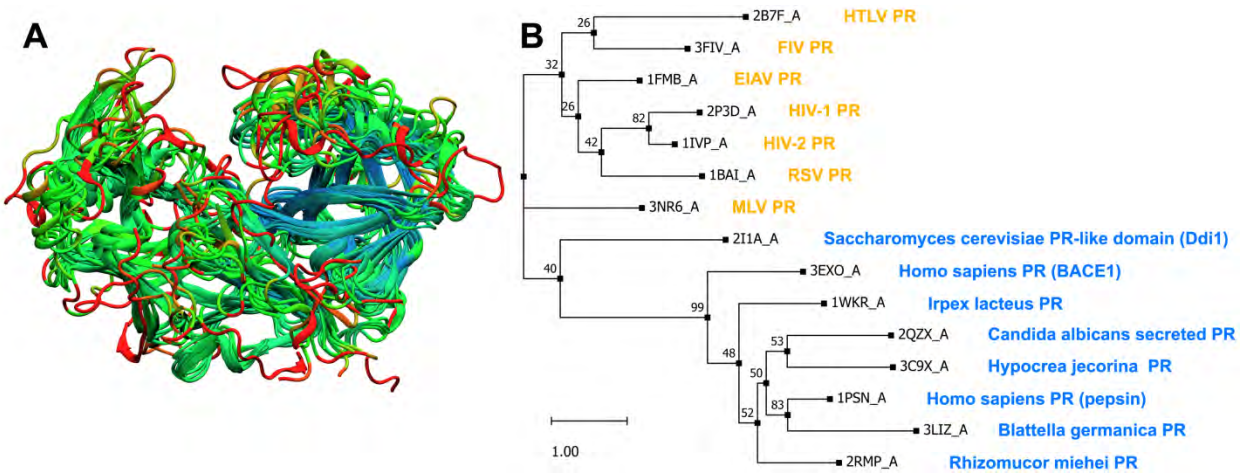

Figure 5-10 A) Structural alignment of aspartic protease representatives. Chain As of each crystal structure are aligned. RGB color scheme is used to indicate alignment score. Red color indicates the worst aligned regions. B) Maximum-likelihood phylogenetic tree. For each leaf node, its PDB ID, chain ID, and name are listed out. Names of retroviral proteases are in yellow, and names of eukaryotic molecules are in blue.

There have been different hypotheses on the evolution of aspartic protease family [218]. Based on the phylogenetic tree, and the apparent similarity between MLVPR and Ddi1 central RVP domain, we hypothesize that MLVPR and Ddi1 central domain may share one ancient ancestor: MLVPR later evolved to form other retroviral proteases, and Ddi1 central domain underwent gene duplication to generate bilobed aspartic proteases. The hypothesis is supported by several facts. Firstly, MLV has been a prototype in retrovirus study because its genome composition is much simpler than other retroviruses [219, 220], so it is conceivable that other virus may have branched out from MLV. Secondly, to our knowledge, homologues of Ddi1[177] and pepsin [221], but not retroviral proteases, are found in prokaryotes, which supports that gene duplication from Ddi1 RVP domain generated bilobed aspartic protease, before life was divided into three domains. Thirdly, besides the overall topology similarity between MLVPR and Ddi1 RVP domain, our finding that they share the same conserved sequence near the active site (see below) also supports grouping them together.

We examined conserved residues based on the structural alignment as well as sequence conservation (Table 5-2, see the methods section for details), and mapped these residues onto pepsin structure (Figure 5-11). Residues conserved across aspartic proteases include the DTG catalytic motif and the Gly residue equivalent to pepsin G122. The G122 is in the center of so-called "ψ-loop" secondary structure, which is two anti-parallel β strands with one extra β strand in between. The ψ-loop was suggested to have evolved from double-ψ β-barrel domain, which is shared by several protein families [218, 222]. Moreover, the ψ-loop in HIVPR was found to act as folding nucleus [2]. Double-ψ β-barrel domain can be viewed as homodimeric aspartic protease without the flap and terminal regions (Figure 5-12). It is possible that during evolution, the flaps were created to facilitate catalytic reactions, and terminal topology was varied depending on the need to dimerize or link to other protein domains.

We also examined residues that are conserved within homodimeric or bilobed aspartic proteases. Interestingly, besides four Gly residues at corresponding positions of pepsin ψ-loop

(G122 and G302) and termini (G169 and G177), the rest conserved residues (S35 and D87 in pepsin and A28 in HIVPR) occur near the flap region and may distinguish the two-flap versus one-flap topology. More specifically, Ser35 is involved in a hydrogen bond network anchoring the pepsin flap tip (Figure 5-13 A) [223], while the hydrogen bonding network around pepsin Asp87, which to our knowledge has been neglected by previous studies, seems to anchor the flap elbow to the cantilever (Figure 5-13 B). We thus hypothesized that S35 and Asp87 may be essential for the active site gating in bilobed aspartic proteases, due to the lack of inter-flap interactions (only available in homodimeric aspartic proteases). We later tested this hypothesis by simulations of apo and holo β-secretase 1.

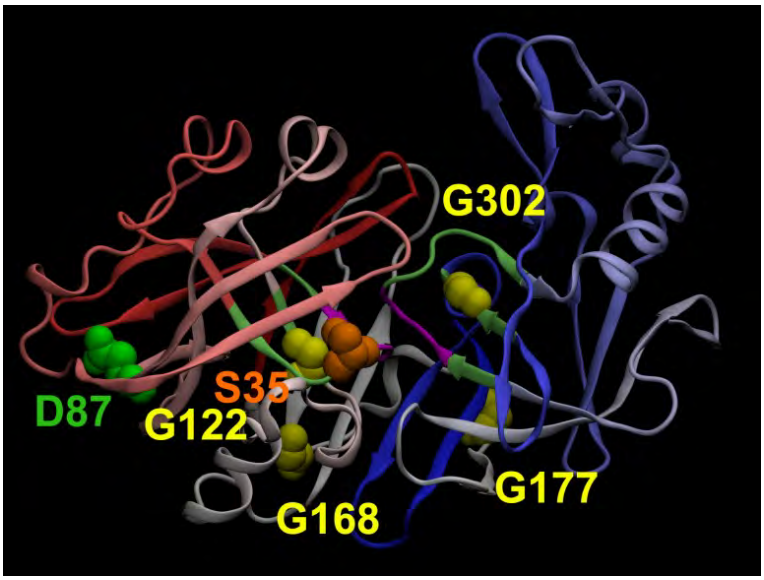

Figure 5-11 Conserved residues within bilobed aspartic proteases, mapped on pepsin structure. Catalytic residues are omitted to facilitate visualization. Crystal structure 1PSN is shown in secondary structure representation. Residues are colored according to their index number, with the N-terminal in red and the C-terminal in Blue. The catalytic site is colored in magenta. The ψ-loops, one on each domain including the catalytic site, are colored lime. Heavy atoms of conserved residues are shown as colored vdW spheres (Asp in green, Ser in orange, and Gly in yellow), and their residue numbers are given in the same color.

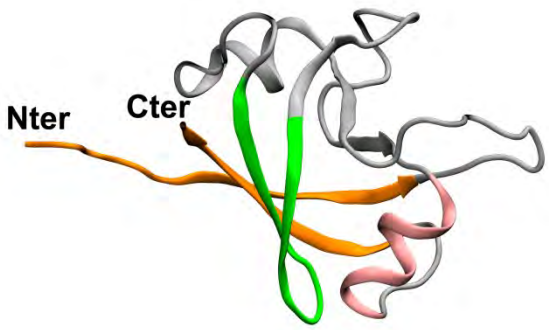

Figure 5-12 N-terminal double-ψ β-barrel domain of VAT (PDB ID: 1CZ4). Color coding is consistent with Figure 5-1. N and C terminals are labeled.

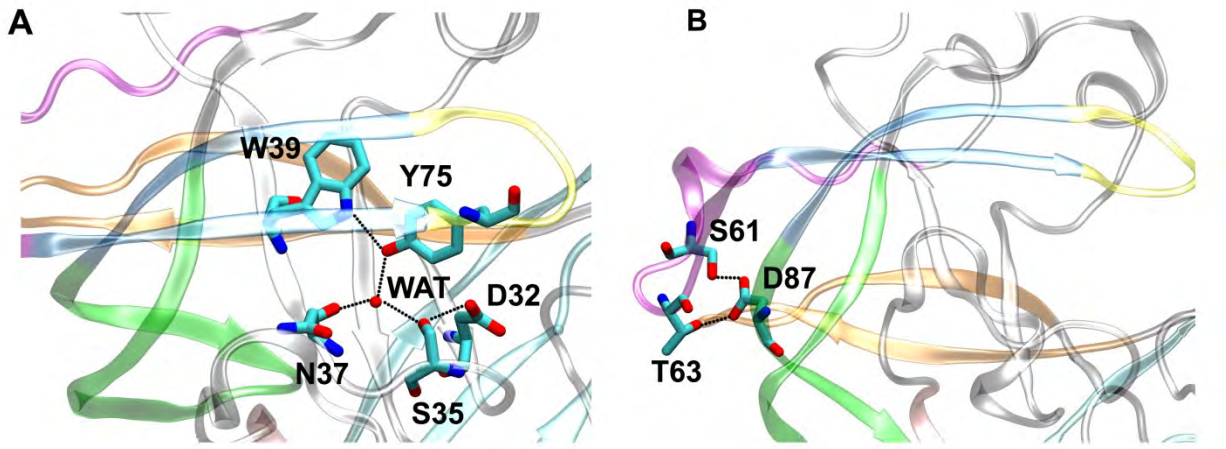

Figure 5-13 A close up view of surrounding hydrogen bond network near residue Ser35 (A) and Asp87 (B) in crystal structure 1PSN. The protein backbone is in the same color scheme as Figure 5-1. Hydrogen bond partners are linked using black dotted lines. The conserved structural water is labeled as "WAT" in panel A.

Apart from conserved residues, we also searched for sequence signatures: residues that are at the same 3-D location, conserved within their own phylogenetic branch but different across different branches (starred in Figure 5-5). The signature distinguishing homodimeric and bilobed aspartic proteases occurs at pepsin S35 position. As discussed above, the S35 in bilobed aspartic proteases is maintaining the flap conformation. In the meantime, it also forms hydrogen bond with catalytic Asp32, which results in limited mobility of Asp32 (Figure 5-13 A). Interestingly, this sequence signature also seems to determine substrate specificity at P2' site across different bilobed aspartic proteases. This site is the closest substrate site to the S35-involved hydrogen-bond network and selects Ala or Val, while various residues are found at other sites [170, 224, 225]. In contrast, all homodimeric aspartic protease has Ala at the same position instead of Ser. It was suggested that the Ala is important for the flexibility of catalytic

residues in retroviral proteases, which may explain their unique ability to process rigid Pro residue at P1' site [226].

The sequence signature distinguishing HIVPR-like aspartic protease from MLVPR and Ddi1 RVP domain is at the position equivalent to HIVPR D29 (pepsin Ser36). D29 in HIVPR forms conserved hydrogen bond network with residues on the N-terminal of the other monomer, which stabilizes the dimer. In the bound state, additional hydrogen bond can be formed with the substrate backbone (Figure 5-14). A Glu residue is found at equivalent position in MLVPR and Ddi1 RVP domain, which may explain the preference of MLVPR for Asn at P3' site because of the favorable interactions between P3'Asn and Gln.

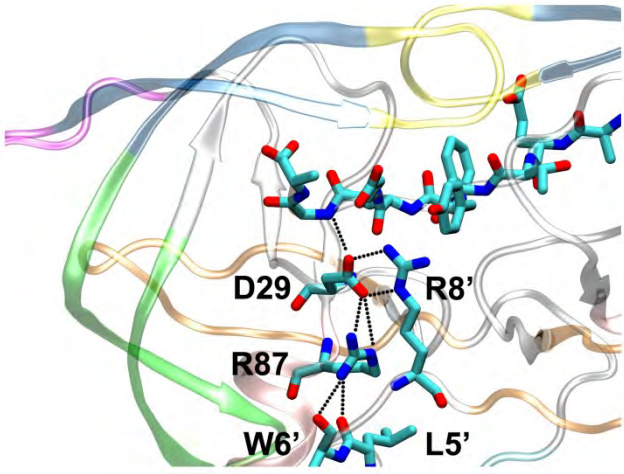

Figure 5-14 A close-up view of surrounding hydrogen bond network near residue Asp29 in HIVPR crystal structure 1KJG. The protein backbone is in the same color scheme as Figure 5-1. Hydrogen bond partner are linked using black dotted lines.

5.3.2 Protease dynamics

Although the dynamics of HIVPR has been studied extensively by MD simulations, the dynamics of other retroviral proteases and bilobed aspartic proteases are not well understood. Since the active site gating is a dynamic process that influences the drug binding, we used molecular dynamic simulation to compare the active site gating mechanisms in different aspartic proteases, in order to help design inhibitors towards non-HIV aspartic proteases.

5.3.2.1 Dynamics of apo MLVPR

5.3.2.1.1 Flap dynamics

We built our MLVPR model based on apo structure 3NR6 with missing densities in the flap and terminal region (PDB ID 3NR6[174]). We used HIVPR crystal structure 1G6L as the template to model MLVPR flaps, and we modeled terminals using crystal symmetry (Figure 5-8, also see the methods section for details). We then performed explicit solvent molecular dynamics simulations from the apo MLVPR model we built. Three independent runs, about 400 ns each, were performed. See Table 5-3 for a summary of all MLVPR simulations performed in this

study. Interestingly, the three runs sampled two different flap conformations, which are similar to HIVPR semiopen (run1) and closed flap conformation (run2 and run3), respectively. We used structural clustering to pick out the most sampled flap conformations. Clustering run1 trajectory produced two large clusters, and we named their representative structures as semi1 and semi2, respectively. Clustering run 2 and 3 structures produced one large cluster, and we named its representative structure as close1.

Since we obtained three relatively stable flap conformations through clustering, we were interested if any of these three structures (semi1, semi2, and close1) can match the crystal structure the simulations started from, which has flap tips and termini disordered. However, none of the three matches perfectly to the resolved flap region from the crystal structure (Figure 5-15). As residues adjacent to the disordered region, residue 54 and 60 on both flaps have average backbone RMSD of 4.0, 3.2 and 3.8 in semi1, semi2, and close1, respectively. This may because the apo MLVPR crystal is an average over different flap conformations. Since the space group of 3NR6 (P422) is unique from all other available crystals of aspartic proteases, it is also possible that crystal contacts play role in determining the flap conformation [109].

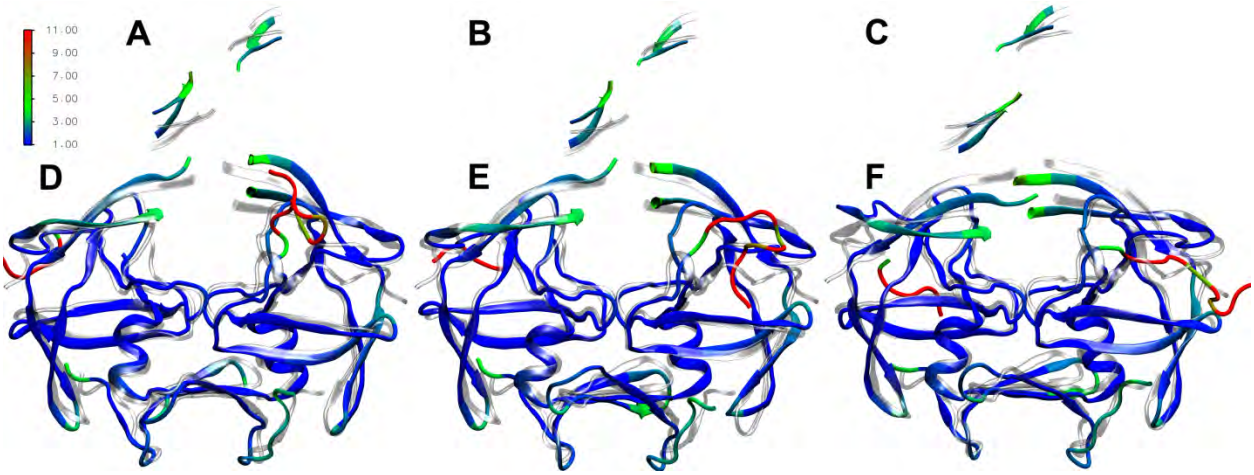

Figure 5-15 Flap tip top view (A-C) and front view (D-F) showing the alignment of cluster-analysis representative structures (colored by RMSD) to 3NR6 crystal (colored in transparent silver). The semi1 structure shown in AD, semi2 structure shown in BE, and close1 structure shown in CF. The core region, excluding flaps and termini, of each structure was fitted to the 3NR6 structure first. Then the backbone RMSD of each residue to the crystal is measured and used for coloring.

During manuscript preparation, three holo structures of MLVPR were published [187]. We used crystal structure 3SM2, which is MLVPR bound with APV, to validate our modeling. Two independent explicit solvent (EXP) simulations from 3SM2, 100 ns each, were performed (see the method section for details). Then, simulation trajectories from 3NR6 run2-3, 3SM2 run1-2, and 3SM2 crystal structure were combined and subjected to structural clustering of flap region using 1.5 Å RMSD cutoff (Table 5-4). 3SM2 crystal structure falls into cluster 3. Cluster 3 also includes all 3SM2 simulation structures, which means the flaps didn't move much in the 3SM2 EXP simulation. The slowness of conformational sampling is a known disadvantage of EXP simulations [130]. Structures of 3NR6 simulations instead sampled many other flap conformations, which may result from different starting structure and longer simulation length, but most structure still fall into cluster 3. Therefore, we concluded that the two sets of

simulation, starting from two different flap conformations, have reached convergence into the same closed flap conformation. The flap RMSD of 3NR6 EXP run 2-3 and 3SM2 EXP runs are presented in Figure 5-16, and 3SM2 and semi2 were used as closed and semiopen flap conformation reference, respectively.

Index of significant clusters	Percentage in 3NR6 MD	Percentage in 3SM2 MD
1	5%	0%
2	16%	0%
3	56%	100%
4	16%	0%
5	3%	0%
6	1%	0%
11	1%	0%

Table 5-4 Statistics of cluster analysis on combined trajectory including 3NR6 run2-3, 3SM2 run1-2, and 3SM2 crystal structure. Structures within each cluster were then separated into 3NR6 run structures and 3SM2 run structures. Percentage population is calculated as the number of structures in certain cluster divided by the total number of structures. Only clusters with more than 0.5% of either 3NR6 run structures or 3SM2 run structures are shown.

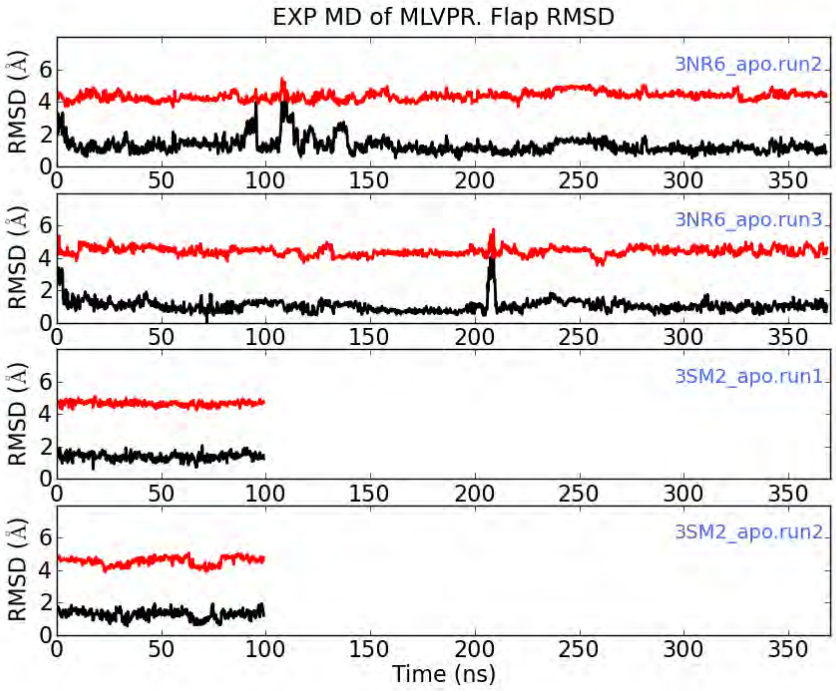

Figure 5-16 Flap RMSD of MLVPR simulations, including run 2-3 from EXP simulation of 3NR6 model and run1-2 from EXP simulation of 3SM2 model. The black curve is RMSD to closed flap conformation, and the red curve is RMSD to semiopen flap conformation sampled in simulation.

We next examined inter-flap interactions from EXP simulation trajectories. 3NR6_run1 first sampled a closed flap conformation with low RMSD to the 3SM2 crystal structure (Figure 5-17 A, around 5 ns), and then switched to and stayed at semiopen structure, which is indicated by low RMSD to semi2 reference structure. Notably, there are two most sampled semiopen conformations in run 1, having RMSD around 2 Å and 1 Å to semi2, respectively (Figure 5-17 A). The two conformations are interconverting during the simulation, and they are stabilized by backbone-sidechain (Figure 5-17 B) and backbone-backbone (Figure 5-17 C) inter-flap interactions, respectively. This is different from HIVPR, which does not have polar residues at flap tips, rely solely on backbone-backbone hydrogen bond for inter-flap interactions, and only has one most populated semiopen structure [227]. Another difference from HIVPR is that, HIVPR usually only has one inter-flap hydrogen bond formed in the semiopen flap conformation [227], while MLVPR can form four hydrogen bonds simultaneously (Figure 5-17 C and Figure 5-18), which could greatly increase its entropy penalty. Different from run1, the other two runs (run 2 and 3) sampled and stayed at closed flap conformations, maintaining less than 2Å RMSD to 3SM2 crystal structure without much deviation (Figure 5-16). Different from HIVPR, which maintains its closed flap conformation using inter-flap hydrogen bond, MLVPR again utilizes sidechain interaction to stabilize this conformation (Figure 5-19 B). Interestingly, the hydrogen bonds in both directions, between Gly59 backbone oxygen and Hγ atom from Thr58 on the opposite flap, rarely form at the same time. Instead, they are competing with each other. The equilibrium of breaking/forming hydrogen bond may help maintain the flap conformation while

preventing the entropy penalty due to rigidity. The competition between two hydrogen bonds is observed in both 3NR6 simulations and 3SM2 simulations (Figure 5-19 A).

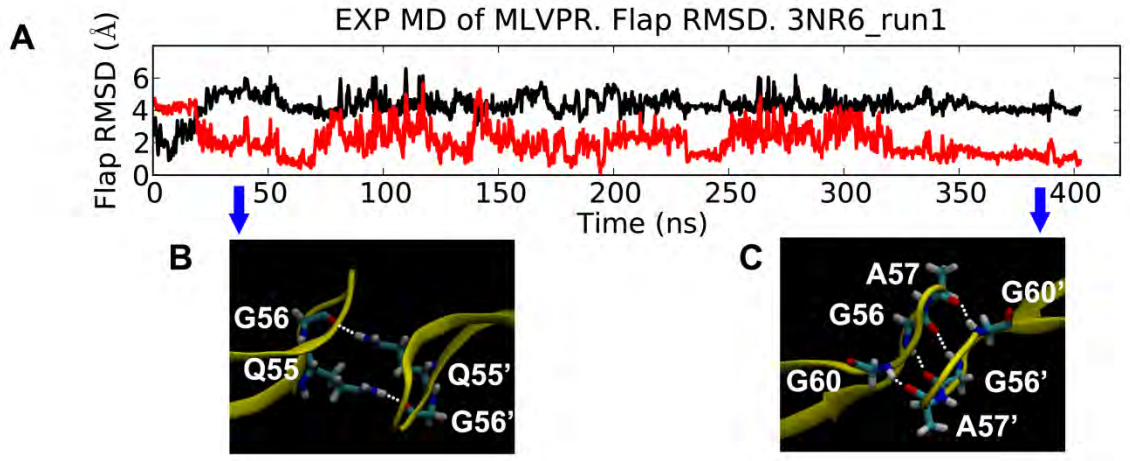

Figure 5-17 Flap RMSD (A) and flap snapshots (B-C) from 3NR6_run 1 simulation. B-C: flap tip top view of simulation snapshots. Inter-flap hydrogen bonds are shown as dotted lines. Residues participating in the hydrogen bonding are labeled, and residues on monomer B are indicated with a prime. The black curve is RMSD to closed flap conformation, and the red curve is RMSD to semiopen flap conformation.

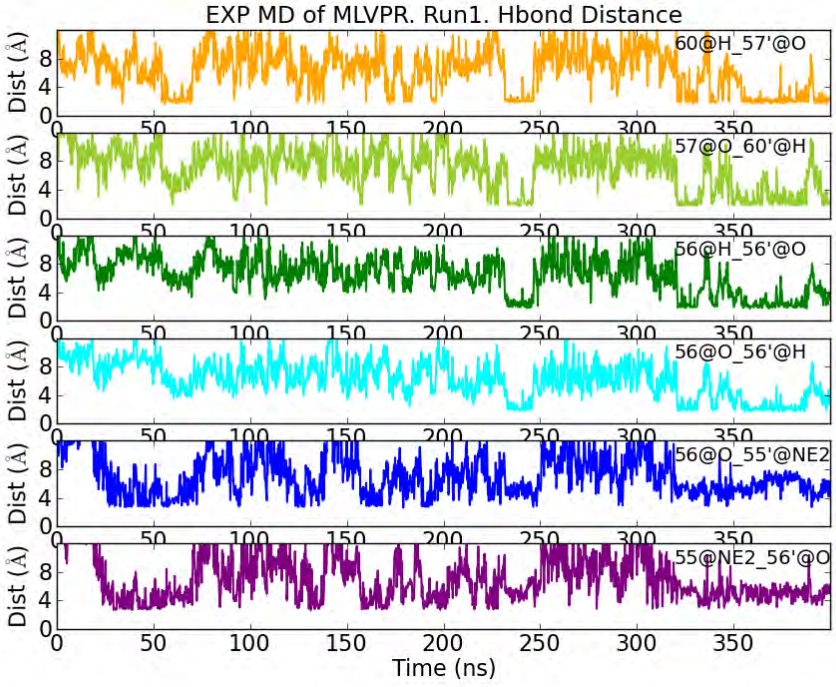

Figure 5-18 Hydrogen bond Vs time for 3NR6 run1. Hydrogen bond distances are calculated between the atom-pairs listed on top of each curve. The residues on monomer B have a prime after their residue number.

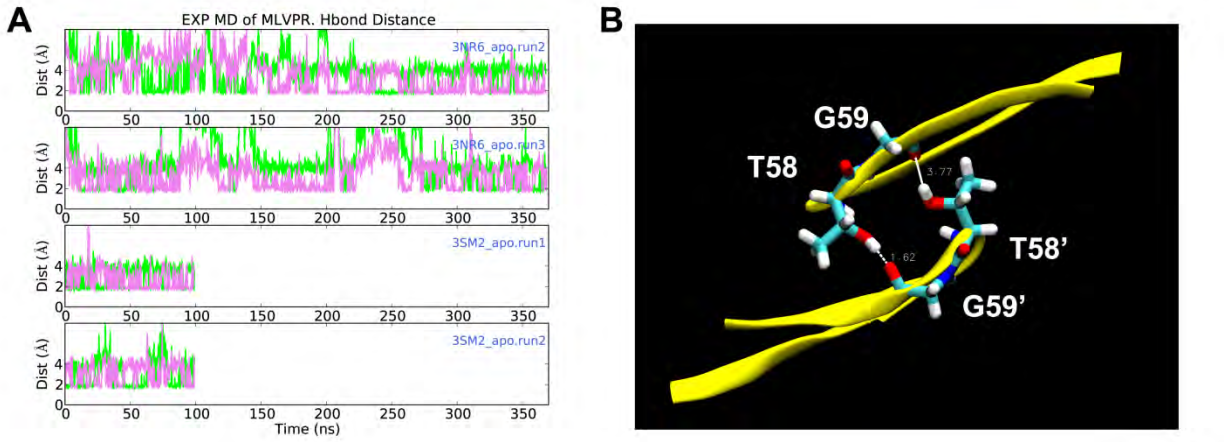

Figure 5-19 Hydrogen bond (A) and flap snapshot (B) for EXP simulations that sampled closed flap conformations. A) Hydrogen bonds bewteen G59@O and T58'@Hγ1 in both directions are shown as pink and green curves. B) Flap snapshot showing the inter-flap hydrogen bond partners, along with the atom-pair distance measurement.

Previously we showed that GB simulation of HIVPR outperforms EXP simulation in sampling transitions of flap conformations[130]. Here we also carried out GB simulations from close1 structure, to sample transitions among different MLVPR flap conformations. Five independent runs were performed, and two of them sampled transitions among closed, semiopen and open structures. The trajectory snapshots along with flap RMSD of run 3 are shown in Figure 5-20, which suggests similar flap dynamics as HIVPR (Figure 5-3).

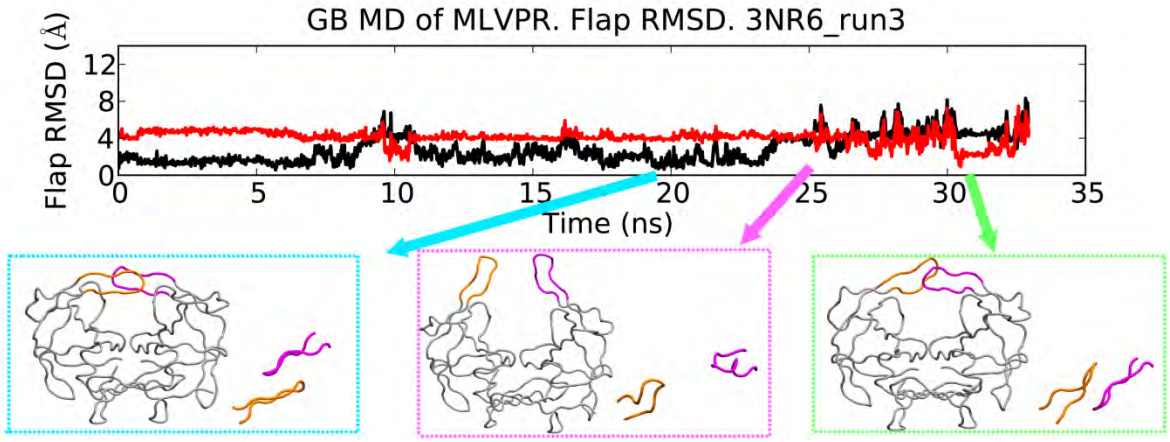

Figure 5-20 Flap RMSD and trajectory snapshots of 3NR6 GB simulation run3. The black curve is RMSD to closed flap conformation, and the red curve is RMSD to semiopen flap conformation. The front view and top view at three time points, having closed, open, and semiopen flap conformations, are included in cyan, pink, and green box, respectively. N-termini are omitted to facilitate visualization.

5.3.2.1.2 Termini dynamics

Since retroviral proteases are expressed in the same polypeptide chain as other viral proteins, they need to cleave themselves to free N and C terminals before they can function

properly. This makes autoprocessing an attractive target for protease inhibition. Because of the large difference in terminal topology between MLVPR and HIVPR, we examined the dynamics of MLVPR termini during simulations. The N/C termini snapshots of 3NR6 run2 are shown in Figure 5-21. N-terminal is very mobile during the simulation (Figure 5-21 A), while C-terminal remains stable (Figure 5-21 B). During the simulation, N-terminal of one monomer occupied positions over the flaps that can be seen as suitable for cleavage once the flaps open (Figure 5-21 CD). This is consistent with the HIVPR autoprocessing model proposed by Louis et al.: N-terminal autoprocessing occurs first, which is followed by C-terminal cleavage [228, 229]. Although all existing HIVPR simulations have stable N/C terminals, which renders the model hard to visualize, our MLVPR simulations give a vivid picture how N-terminal can extend into the catalytic site to get processed. Our simulation results suggest N-terminal of MLVPR is intrinsically flexible. The fact that one monomer in 3NR6 crystal has ordered N-terminal is likely due to nearby co-crystallized ions (Figure 5-4 B). MLV has served as a prototype for retroviral studies because of its simple genome, and here our results suggest that MLVPR may also be a better model for studying retroviral autoprocessing since its N termini dynamics are readily observed in relatively short simulations.

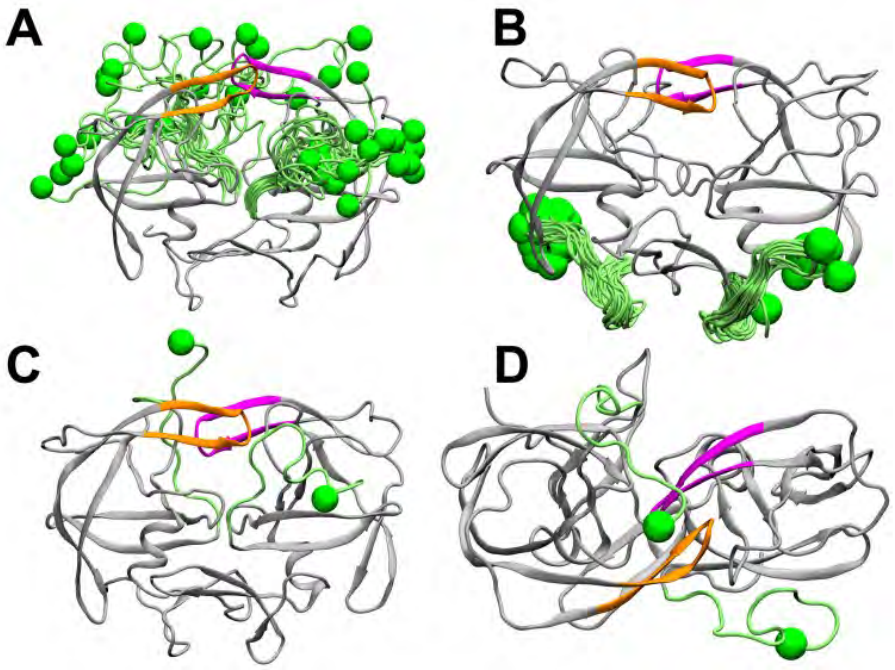

Figure 5-21 Snapshots of N and C termini during the simulation 3NR6 run2. The terminal tip residues are shown in vdW spheres, and the termini are highlighted in green. A) Snapshots of N termini saved every 20 ns, the coordinates of the rest part of the protein come from the EXP simulation starting structure. B) Snapshots of C termini saved every 20 ns, the coordinates of the rest part of the protein come from the EXP simulation ending structure. C-D) Front view and top view of the structure at 240 ns.

5.3.2.2 Dynamics of apo HIVPR, apo SIVPR and apo HTLVPR

In addition to MLVPR, we also investigated the dynamics of SIVPR and HTLVPR in the unbound state using MD simulations. Additional simulations of HIVPR were performed as control. Since the transition from the closed flap conformation to other configurations in explicit solvent is too slow (beyond the time scale of computation currently available, see Figure 2-10), we performed GB simulations to sample semiopen flap conformations prior to explicit solvent simulations (see the methods section for details). In our experience simulations starting from semiopen flap conformation have faster transition to other flap conformations, likely due to the greater entropy of the inter-flap interactions in semiopen flap conformation.

For each protease, two independent simulations (1 us each) were performed. The simulation structures from HIV, SIV, and HTLV protease simulations are shown in Figure 5-22, Figure 5-23, and Figure 5-24, respectively. The overall flap conformations sampled by these retroviral proteases, besides the MLVPR flap conformations introduced earlier (Figure 5-20), are very similar and are composed of mainly closed, semiopen, and open flap conformations. However, the specific inter-flap hydrogen bonds, which are found in closed and semiopen flap conformations, are different among these retroviral proteases. The inter-flap hydrogen bonds in HIVPR are between backbone atoms (Figure 5-25), which is very similar to SIVPR simulations. HTLVPR inter-flap hydrogen bonds are also between backbone atoms, although usually only one hydrogen bond is stable in semiopen flap conformation, different from HIVPR and SIVPR, which have two inter-flap hydrogen bonds in the same conformation. In contrast, the inter-flap hydrogen bonds in MLVPR involve sidechain atoms (Figure 5-17 and Figure 5-19), and at most four hydrogen bonds could form in its semiopen flap conformation (Figure 5-18).

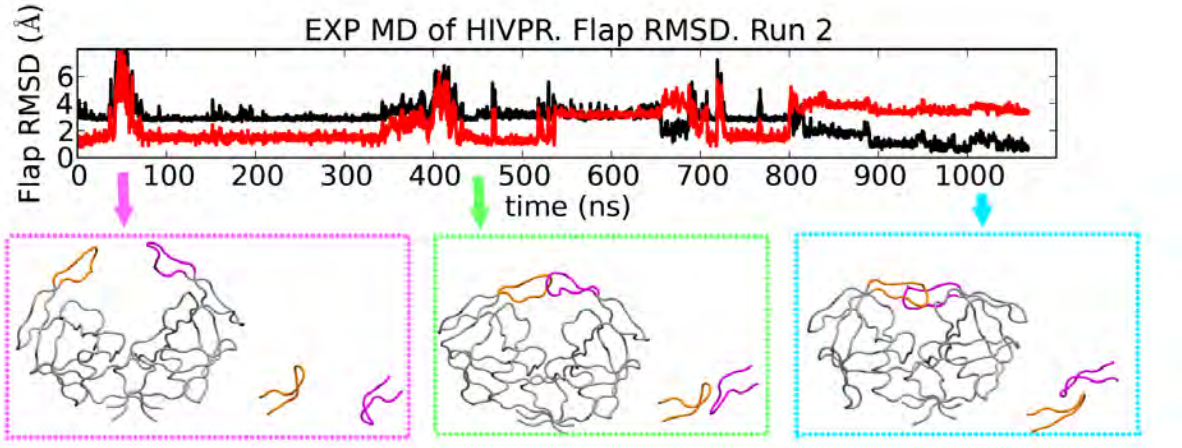

Figure 5-22 Flap RMSD and trajectory snapshots of HIVPR explicit simulation run 2. The black curve is RMSD to closed flap conformation, and the red curve is RMSD to semiopen flap conformation. The front view and top view at three time points, having closed, open, and semiopen flap conformations, are included in cyan, pink, and green box, respectively.

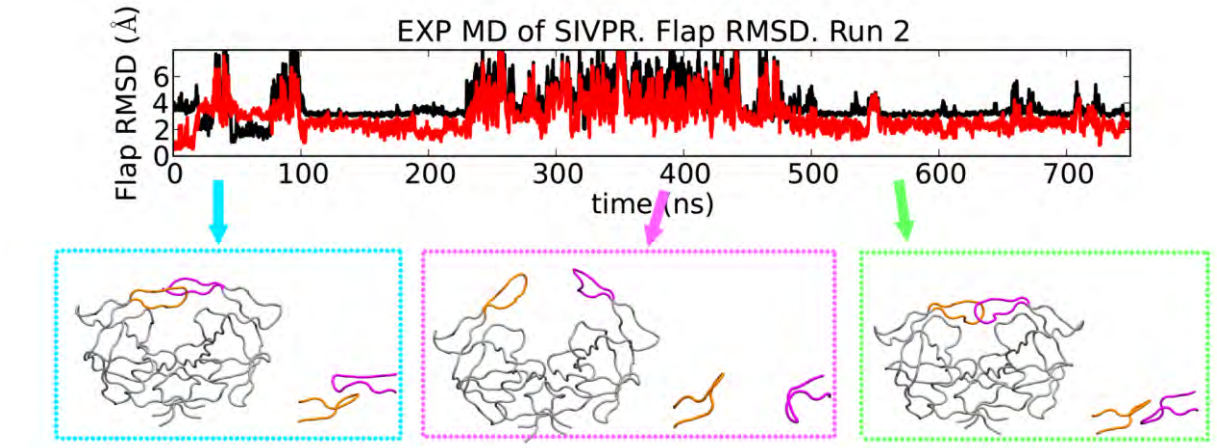

Figure 5-23 Flap RMSD and trajectory snapshots of SIVPR explicit simulation run 2. The black curve is RMSD to closed flap conformation, and the red curve is RMSD to semiopen flap conformation. The front view and top view at three time points, having closed, open, and semiopen flap conformations, are included in cyan, pink, and green box, respectively.

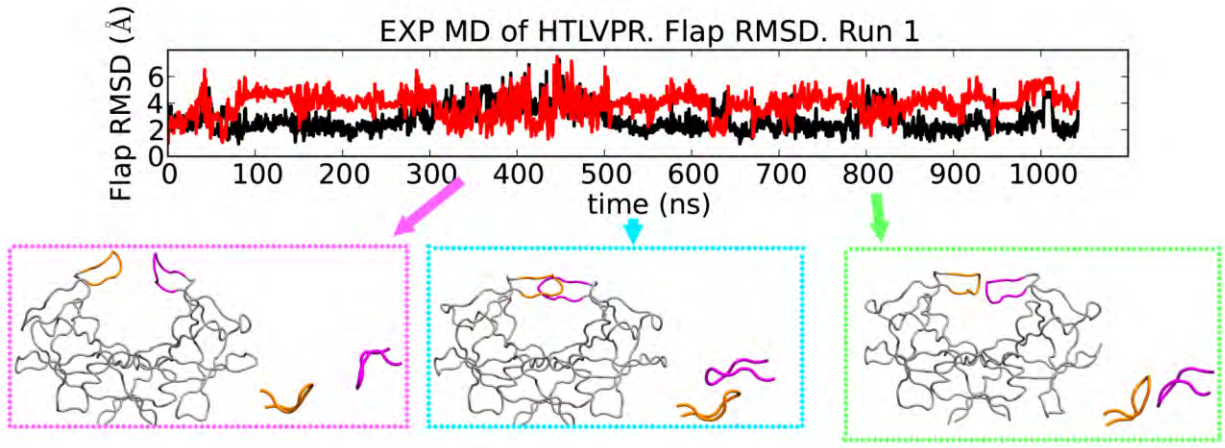

Figure 5-24 Flap RMSD and trajectory snapshots of HTLVPR explicit simulation run 1. The black curve is RMSD to closed flap conformation, and the red curve is RMSD to semiopen flap conformation. The front view and top view at three time points, having closed, open, and semiopen flap conformations, are included in cyan, pink, and green box, respectively.

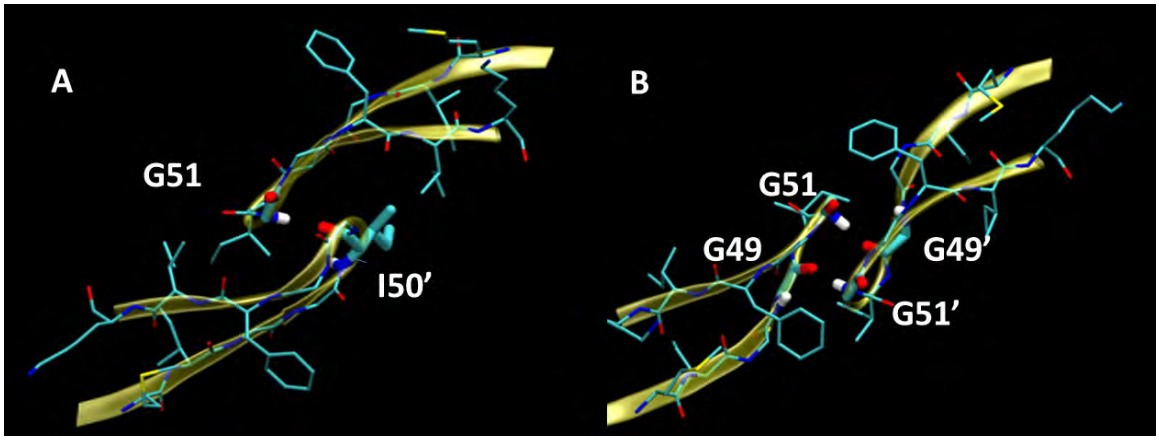

Figure 5-25 Inter-flap hydrogen bonding observed in HIVPR simulations. Residues participating in the hydrogen bonding are shown in licorice representation, and nearby residues are shown in line representation. Residues on monomer B are indicated with a prime.

Another significant difference is found in HTLVPR simulations. Unlike apo simulations of other retroviral proteases (HIVPR, SIVPR, and MLVPR), which sampled a closed flap conformation similar to their bound crystal structures (Figure 3-4), apo HTLVPR simulations suggest that its flaps would collapse further down when the ligand is bound to the protease (Figure 5-26). The difference between the unbound and bound HTLVPR flap conformations means that the protease needs to undergo significant conformational changes upon ligand binding, and the protease is under much constraint when bound to the peptidic inhibitor in the crystal structure (2B7F, Figure 1-2 E). Therefore, the apo HTLVPR structures sampled in this study may provide a better target to design HTLVPR inhibitors than the holo crystal structures, since the large active site cavity in the apo structure would allow both the ligand and the protease to be fully relaxed prior to binding.

Figure 5-26 Comparison between closed flap conformation HTLVPR sampled in the simulation (solid color), and the holo crystal structure 2B7F (transparent color). Monomer A and B in both structures are colored orange and magenta, respectively.

5.3.2.3 Dynamics of apo and holo β-secretase 1

Although little is known how the active site gating is achieved in pepsin-like aspartic proteases, from our evolutionary profile study we found there are two residues conserved in all bilobed aspartic proteases and might be related to their flap control (Figure 5-11): a Ser residue near the active site, and an Asp residue on the cantilever. We hypothesized that these two residues may be needed to stabilize the flap orientation by controlling the flap tip through water-mediated hydrogen bonding to the Ser (Figure 5-13 A) and by controlling the flap elbow through hydrogen bonding to the Asp (Figure 5-13 B). We tested this hypothesis by simulating apo BACE with either wild type sequence, or with two conserved residues mutated (S35A/D83A).

Starting simulations of apo β-secretase 1 (BACE) from apo crystal structure 1W50 (Figure 5-9), which does not have the conserved hydrogen network near the Ser (Figure 5-13 A), ensures that there is no memory of the hydrogen network in the starting structure. For both wild type and mutant sequences, two independent simulations were performed in explicit solvent (1.2 us each), and the last frame of each simulation is shown in Figure 5-27. Remarkably, the 3-10 helix (blue) near the elbow region (left lobe) is preserved in the wild type simulations but not in the mutant simulations (Figure 5-27 compared to Figure 5-9). However, in all four snapshots the flap backbones look similar and the flap orientation do not look so different. Therefore, we compared the hydrogen bonding patterns at the active site and the elbow region in more detail below.

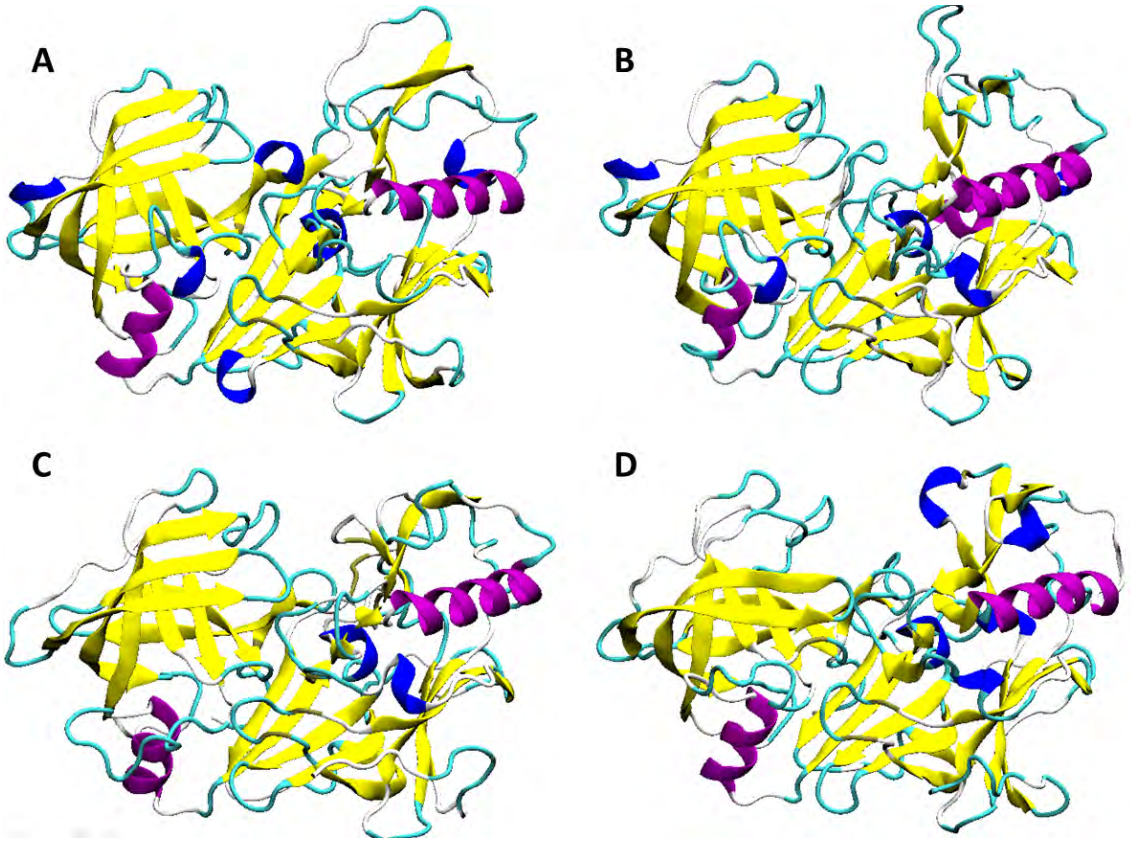

Figure 5-27 Last frame snapshots of apo BACE simulations. A) Wild type sequence run 1. B) Wild type sequence run 2. C) Mutant sequence run 1. D) Mutant sequence run 2.

Since we are most interested in how the Ser35 in BACE is stabilizing the Tyr71 at its flap tip, and how that would affect the active site, we created two dihedral measurements to assess the orientation of the Tyr71 as well as the active site (Figure 5-28). The time series of Tyr71 dihedrals are plotted in Figure 5-29. Generally, the wild type stayed relatively longer at dihedral of -100 degree, which corresponds to the conserved hydrogen bonding network observed in evolutionary study (Figure 5-13 A and Figure 5-28), while the mutant with Ser35 mutated to Ala can no longer maintain the hydrogen bonding and goes to other Tyr71 dihedrals at 150 degree (interaction with the catalytic residues, Figure 5-30 A) and 50 degree (π stacking with Trp76, Figure 5-30 B). Wild type BACE run 1 simulation also sampled Try71 orientations other than the conserved hydrogen bonding after about 800 ns, but the new dihedral at -150 degree turned out to resemble the orientation of mutant run1 simulation at 150 degree. This is likely due to the symmetry of Tyrosine side chain phenyl ring. It is worth noting that wild type BACE run1 Tyr71 interacts with the solvent instead of the catalytic residues when the dihedral is around -150 degree.

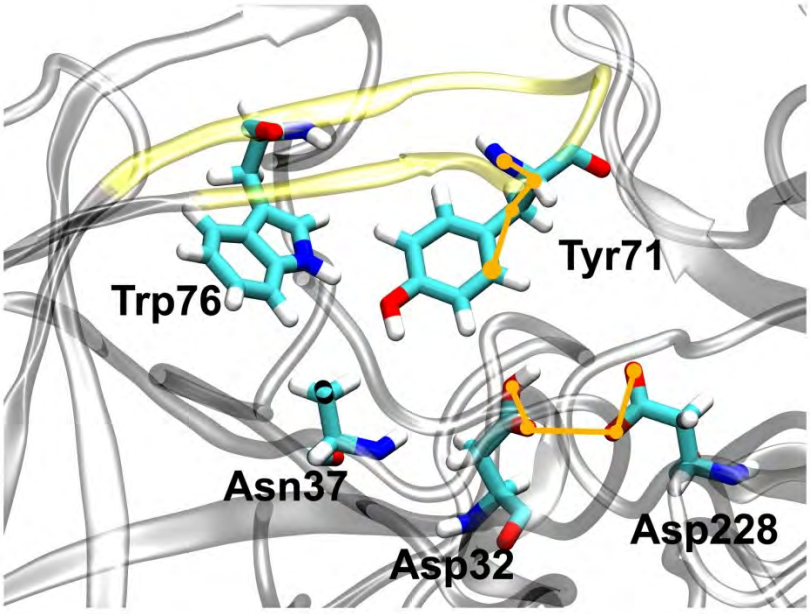

Figure 5-28 Dihedral measurements at the BACE active site. Structure is taken from wild type run2 simulation at 1.2 us. The Tyr71 orientation is characterized using the dihedral N-Cα-Cβ-Cδ1. The active site orientation is measured as the dihedral Asp32_Oδ2-Asp32_Oδ1-Asp228_Oδ1-Asp228_Oδ2. But when Asp228_Oδ2 is closer to Asp228_Cα than Asp228_Oδ1, the active site orientation is measured as the dihedral Asp32_Oδ2-Asp32_Oδ1-Asp228_Oδ1-Asp228_Oδ2 instead, to account for swapping of Asp228 two carboxyl oxygen atoms. Asp32 is protonated at Asp32_Oδ2.

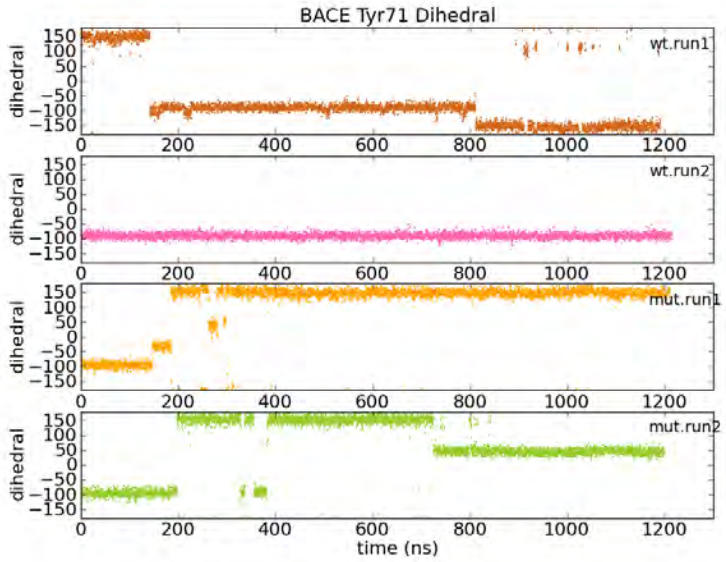

Figure 5-29 Dihedral of Tyr71 in BACE simulations. The Tyr71 orientation is characterized using the dihedral N-Cα-Cβ-Cδ1.

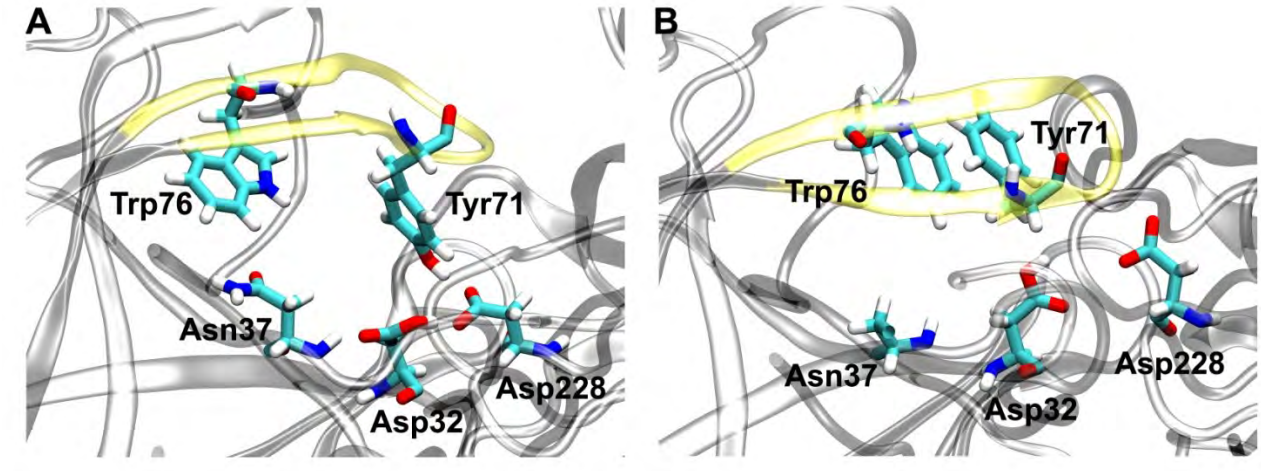

Figure 5-30 Tyr orientation observed in simulations. A) Mutant run 1 at 1.2 us. Dihedral of Tyr 155.6 degree. B) Mutant run2 at 1.2 us. Dihedral of Tyr 49.4 degree.

We then examined the active site geometry in the wild type and mutant BACE simulations. The dihedral measurement shows that the wild type has a stable active site orientation while the mutant has big fluctuation in the relative position of two catalytic residues (Figure 5-31). This is expected because the mutation S35A eliminates the hydrogen bonding between Ser35 and Asp32, which constrains the orientation of Asp32 and in turn stabilize the catalytic site orientation. In the mutant simulations, the catalytic residues have more freedom without the hydrogen bonding constraint.

Finally, we measured the hydrogen bonding near Asp83, and found that the hydrogen bonding network here is extremely stable (Figure 5-32 and Figure 5-33). This also explains the well preserved 3-10 helix in the wild type simulations (Figure 5-27). However, in the mutant simulations, because of the D83A mutation, the hydrogen bonds are all disrupted, the elbow region no longer attaches to the cantilever, and the 3-10 helix is not preserved (Figure 5-27).

Overall, the results of apo BACE simulations are consistent with our hypothesis: Ser35 is needed to control the flap tip (through maintaining Tyr71 at a backward orientation found in conserved hydrogen network) and also to expose the active site to the solvent, and Asp83 is needed to maintain the contact between the flap elbow and the cantilever. However, ligand needs to be included in the simulation to elucidate how these two hydrogen bonding networks would affect the ligand binding process.

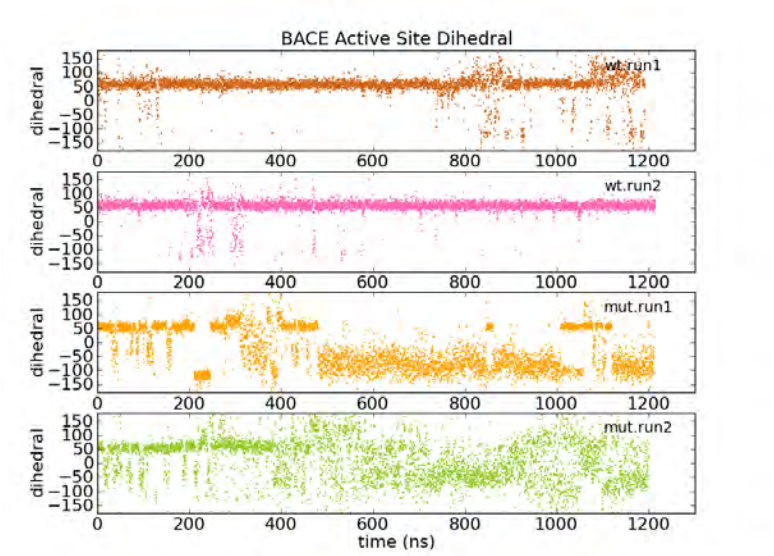

Figure 5-31 Dihedral of the catalytic residues in BACE simulations. The active site orientation is measured as the dihedral Asp32_Oδ2-Asp32_Oδ1-Asp228_Oδ1-Asp228_Oδ2. But when Asp228_Oδ2 is closer to Asp228_Cα than Asp228_Oδ1, the active site orientation is measured as the dihedral Asp32_Oδ2-Asp32_Oδ1-Asp228_Oδ1-Asp228_Oδ2 instead, to account for swapping of Asp228 two carboxyl oxygen atoms.

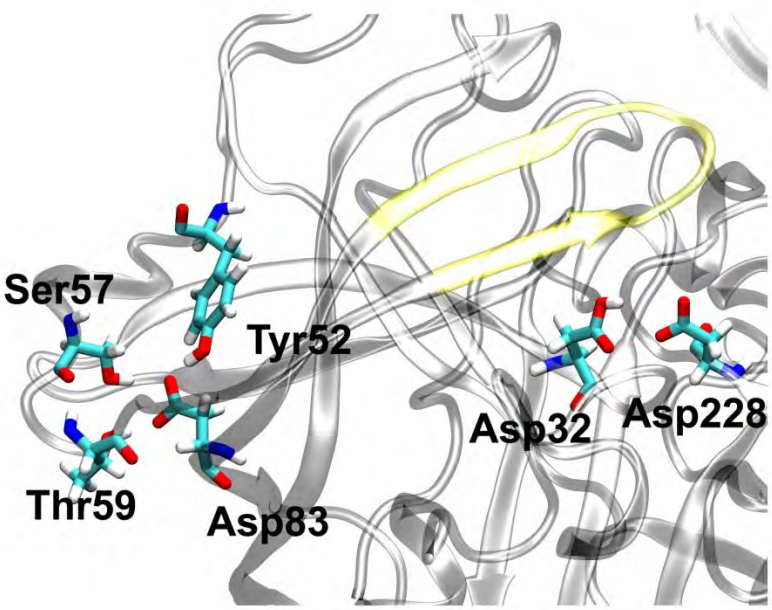

Figure 5-32 The hydrogen bonding partners near Asp83 in BACE (left). The catalytic residues are shown on the right.

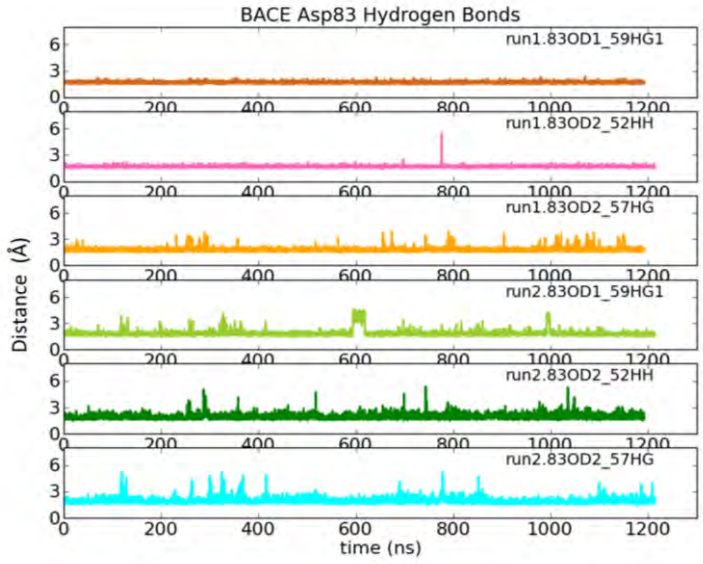

Figure 5-33 The hydrogen bonds near Asp83 in BACE wild type simulations.

Four independent simulations of holo BACE, two starting structures with two binding poses each, were performed to further explore the active site gating mechanism in pepsin-like aspartic proteases. In the starting structures, the natural substrate was docked onto the BACE surface using MOE software [230], instead of directly into the active site, so that the substrate would explore different binding locations, and the flap opening may be needed in order to let in the substrate.

The substrate to active site distance for each run, calculated as the CαCα distance between substrate central Ile to BACE active site residue Asp228, is plotted below (Figure 5-34). The same distance measured from the crystal structure 1FKN is 8Å. All four runs had the substrate exploring different binding positions over the BACE surface (Figure 5-35), which is indicated by CαCα distance larger than 20 Å (Figure 5-34). However, one run out of four (run2) sampled the substrate recognition process, which is indicated by the CαCα distance maintained below 10 Å.

The substrate was bound at a wrong orientation, with its N and C termini locations are swapped. The protease took quite a long time (about 200 ns) to first let in the substrate, and then to try to put the sidechains in to the right pocket, before it finally gave up. During this recognition process, we found there is a flap tip upward movement necessary to make room for the substrate to enter the active site, which is achieved by swinging in Tyr71 to form the conserved hydrogen bond with Ser35, and flipping up Gln73 side chain (Figure 5-36). During the recognition process, there is a relative movement of two lobes of BACE, however, the flap elbow position, which is stabilized by Asp83, remain unchanged. Therefore, we hypothesize that the active site gating in pepsin-like aspartic proteases is featured by modulating the flap tip while maintaining the secondary structure at flap elbow region.

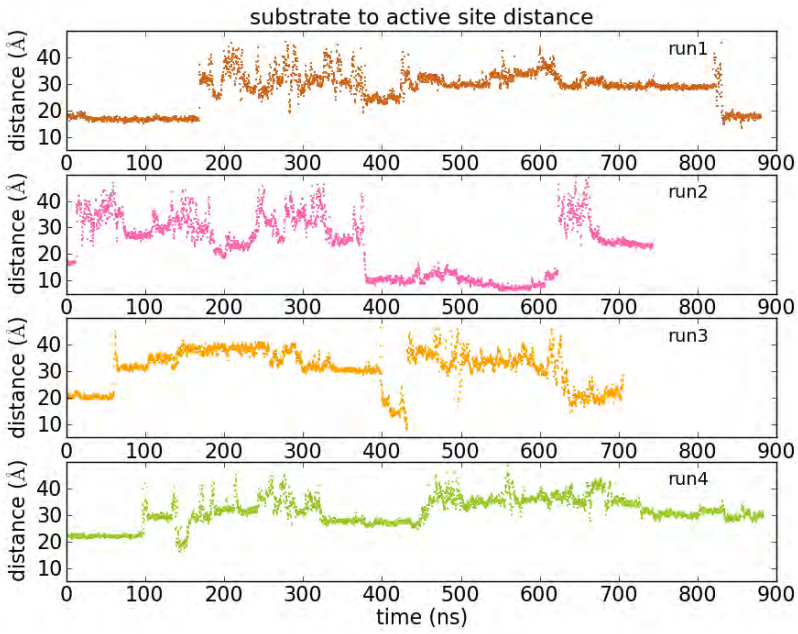

Figure 5-34 Substrate-to-active-site distance in BACE holo simulations, calculated as the CαCα distance between substrate center Ile to BACE active site residue Asp228.

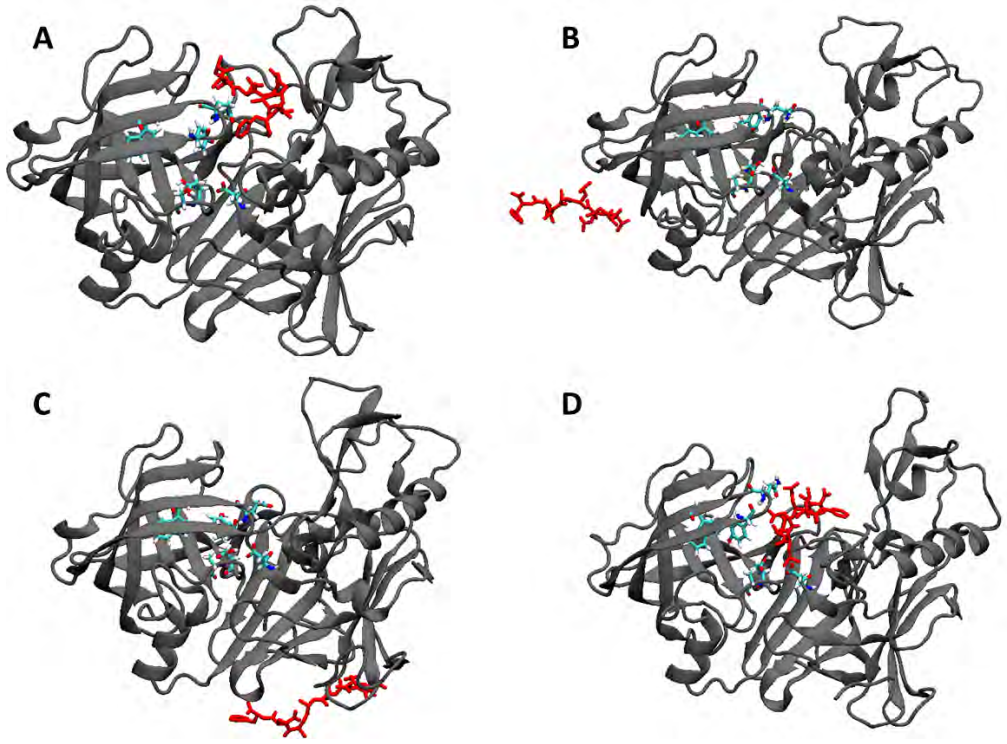

Figure 5-35 Simulation snapshots of holo BACE run2 at 0 ns (A), 120 ns (B), 350 ns (C), and 520 ns (D).

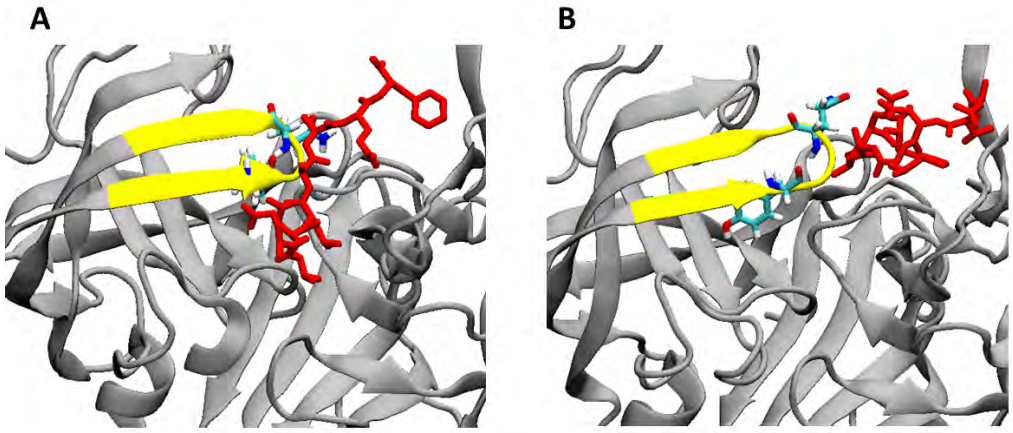

Figure 5-36 Simulation snapshots of holo BACE run2 illustrating the closed (A) and elevated (B) flap conformation.

5.4 Conclusions

In this study, we created evolutionary profile for aspartic protease family based on systematic comparison of non-redundant representatives from the whole family. Based upon structure-guided sequence alignment, the maximum-likelihood phylogenetic tree suggests that MLVPR and Ddi1 RVP domain occupy evolutionary branches distinct from HIVPR-like aspartic

proteases and pepsin-like aspartic proteases. We formulated a hypothesis for the evolution of aspartic proteases, and discussed evidences supporting our hypothesis. We also pointed out conserved residues within each evolutionary branch, and linked sequence signatures to their influence on substrate specificity.

Because the ligand-binding process is not static, comparison of protease dynamics, especially the comparison among active site gating mechanisms, is important in determining similarity among different proteases and designing new inhibitors. Previous studies already provided models for HIVPR active site gating by modulating its two flaps. Here we performed MD simulations of a series of retroviral proteases (MLVPR, SIVPR, and HTLVPR) along with HIVPR, to compare their active site gating mechanisms. Our results suggest that the active site gating is conserved among these retroviral proteases (all retroviral proteases studied sampled closed, semiopen and open conformations in the unbound state), although the specific inter-flap hydrogen bonding pattern varies among them, in a fashion which seems correlated with the evolutionary profile. Moreover, our MD simulations of HTLVPR demonstrated that the protease active site undergoes a large conformational change upon ligand binding, which is remarkably different from other retroviral proteases studied. Therefore, the unrestrained apo form HTLVPR sampled here may provide a better model for designing inhibitors specific to HTLVPR.

We examined the dynamics of MLVPR terminals, which have topology significantly different from HIVPR. Interestingly, the MLVPF N-terminal is mobile during simulation while C-terminal is stable, which is consistent with existing models for the HIVPR self-cleavage mechanism. Our simulations provide a vivid picture how N-terminal of the protease can extend into the active site to get self-cleaved once two monomers dimerize.

Finally, we also studied the dynamics of β-secretase 1 (BACE), which belongs to the bilobed aspartic proteases. Unlike HIVPR, the active site gating mechanism in BACE is not well understood. Through evolutionary study we identified two conserved residues that may be responsible for active site gating in bilobed aspartic proteases, and we tested our hypothesis by performing explicit solvent simulations of wild type and mutant BACE. Our preliminary data support our hypothesis. Further simulations of holo BACE captured the substrate recognition process, and provided insights into active site gating mechanism in bilobed aspartic proteases.

Reference

[1] Humphrey, W., Dalke, A., Schulten, K. VMD: Visual molecular dynamics. Journal of Molecular Graphics. 1996, 14, 33-8.
[2] Bonomi, M., Gervasio, F.L., Tiana, G., Provasi, D., Broglia, R.A., Parrinello, M. Insight into the folding inhibition of the HIV-1 protease by a small peptide. Biophysical Journal. 2007, 93, 2813-21.
[3] Miller, M. The early years of retroviral protease crystal structures. Biopolymers. 2010, 94, 521-9.
[4] Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., et al. X-ray analysis of HIV-1 proteinase at 2.7 A resolution confirms structural homology among retroviral enzymes. Nature. 1989, 342, 299-302.
[5] Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., et al. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. Science. 1989, 246, 1149-52.
[6] Navia, M.A., Fitzgerald, P.M.D., McKeever, B.M., Leu, C.-T., Heimbach, J.C., Herber, W.K., et al. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature. 1989, 337, 615-20.
[7] Wlodawer, A., Vondrasek, J. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. Annu Rev Biophys Biomol Struct. 1998, 27, 249-84.
[8] Li, M., Laco, G.S., Jaskolski, M., Rozycki, J., Alexandratos, J., Wlodawer, A., et al. Crystal structure of human T cell leukemia virus protease, a novel target for anticancer drug design. Proceedings of the National Academy of Sciences of the United States of America. 2005, 102, 18332-7.
[9] Satoh, T., Li, M., Nguyen, J.T., Kiso, Y., Gustchina, A., Wlodawer, A. Crystal structures of inhibitor complexes of human T-cell leukemia virus (HTLV-1) protease. J Mol Biol. 2010, 401, 626-41.
[10] Dunn, B.M., Goodenow, M.M., Gustchina, A., Wlodawer, A. Retroviral proteases. Genome Biol. 2002, 3, REVIEWS3006.
[11] Dunn, B.M. Structure and mechanism of the pepsin-like family of aspartic peptidases. Chem Rev. 2002, 102, 4431-58.
[12] Chen, A., Campeau, L.C., Cauchon, E., Chefson, A., Ducharme, Y., Dube, D., et al. Renin inhibitors for the treatment of hypertension: design and optimization of a novel series of pyridone-substituted piperidines. Bioorg Med Chem Lett. 2011, 21, 3970-5.
[13] Ghosh, A.K., Gemma, S., Tang, J. beta-Secretase as a therapeutic target for Alzheimer's disease. Neurotherapeutics. 2008, 5, 399-408.
[14] Vassar, R., Kovacs, D.M., Yan, R., Wong, P.C. The beta-secretase enzyme BACE in health and Alzheimer's disease: regulation, cell biology, function, and therapeutic potential. J Neurosci. 2009, 29, 12787-94.
[15] Kandalepas, P.C., Vassar, R. Identification and biology of beta-secretase. J Neurochem. 2012, 120 Suppl 1, 55-61.
[16] Nezami, A., Luque, I., Kimura, T., Kiso, Y., Freire, E. Identification and characterization of allophenylnorstatine-based inhibitors of plasmepsin II, an antimalarial target. Biochemistry. 2002, 41, 2273-80.

[17] Luksch, T., Chan, N.S., Brass, S., Sotriffer, C.A., Klebe, G., Diederich, W.E. Computer-aided design and synthesis of nonpeptidic plasmepsin II and IV inhibitors. ChemMedChem. 2008, 3, 1323-36.
[18] Fisher, N.D., Meagher, E.A. Renin inhibitors. J Clin Hypertens (Greenwich). 2011, 13, 662-6.
[19] Fernandez, M., Caballero, J., Fernandez, L., Sarai, A. Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM). Mol Divers. 2011, 15, 269-89.
[20] Durrant, J.D., McCammon, J.A. Computer-aided drug-discovery techniques that account for receptor flexibility. Curr Opin Pharmacol. 2010.
[21] Mohan, C.G., Gandhi, T., Garg, D., Shinde, R. Computer-assisted methods in chemical toxicity prediction. Mini Rev Med Chem. 2007, 7, 499-507.
[22] Hou, T., Li, Y., Zhang, W., Wang, J. Recent developments of in silico predictions of intestinal absorption and oral bioavailability. Comb Chem High Throughput Screen. 2009, 12, 497-506.
[23] Case, D.A., Darden, T.A., Cheatham, T.E., Simmerling, C.L., Wang, J., Duke, R.E., et al. AMBER 11. 2010.
[24] Case, D.A., Darden, T.A., Cheatham, T.E., Simmerling, C.L., Wang, J., Duke, R.E., et al. AMBER 10. 2008.
[25] Berendsen, H.J.C., Postma, J.P.M., Vangunsteren, W.F., Dinola, A., Haak, J.R. Molecular-Dynamics with Coupling to an External Bath. Journal of Chemical Physics. 1984, 81, 3684-90.
[26] Walker, R.C., Crowley, M.F., Case, D.A. The implementation of a fast and accurate QM/MM potential method in Amber. Journal of Computational Chemistry. 2008, 29, 1019-31.
[27] Donchev, A.G., Ozrin, V.D., Subbotin, M.V., Tarasov, O.V., Tarasov, V.I. A quantum mechanical polarizable force field for biomolecular interactions. Proc Natl Acad Sci U S A. 2005, 102, 7829-34.
[28] Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. Journal of Computational Physics. 1977, 23, 327-41.
[29] Darden, T., York, D., Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. Journal of Chemical Physics. 1993, 98, 10089-92.
[30] Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G. A Smooth Particle Mesh Ewald Method. Journal of Chemical Physics. 1995, 103, 8577-93.
[31] Crowley, M.F., Darden, T.A., Cheatham, T.E., Deerfield, D.W. Adventures in improving the scaling and accuracy of a parallel molecular dynamics program. Journal of Supercomputing. 1997, 11, 255-78.
[32] Toukmaji, A., Sagui, C., Board, J., Darden, T. Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. Journal of Chemical Physics. 2000, 113, 10913-27.
[33] Torrie, G.M., Valleau, J.P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. Journal of Computational Physics. 1977, 23, 187-99.
[34] Steinbrecher, T., Case, D.A., Labahn, A. A multistep approach to structure-based drug design: studying ligand binding at the human neutrophil elastase. J Med Chem. 2006, 49, 1837-44.

[35] Gullingsrud, J.R., Braun, R., Schulten, K. Reconstructing Potentials of Mean Force through Time Series Analysis of Steered Molecular Dynamics Simulations. Journal of Computational Physics. 1999, 151, 190-211.
[36] Mills, G., Jonsson, H. Quantum and thermal effects in H2 dissociative adsorption: Evaluation of free energy barriers in multidimensional quantum systems. Phys Rev Lett. 1994, 72, 1124-7.
[37] Miron, R.A., Fichthorn, K.A. Multiple-time scale accelerated molecular dynamics: addressing the small-barrier problem. Phys Rev Lett. 2004, 93, 128301.
[38] Sugita, Y., Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. Chemical Physics Letters. 1999, 314, 141-51.
[39] Shaw, D.E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R.O., Eastwood, M.P., et al. Atomic-level characterization of the structural dynamics of proteins. Science. 2010, 330, 341-6.
[40] Pierce, L.C., Salomon-Ferrer, R., Augusto, F.d.O.C., McCammon, J.A., Walker, R.C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. J Chem Theory Comput. 2012, 8, 2997-3002.
[41] Northrup, S.H., Pear, M.R., McCammon, J.A., Karplus, M. Molecular dynamics of ferrocytochrome c. Nature. 1980, 286, 304-5.
[42] Hornak, V., Okur, A., Rizzo, R.C., Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. Proceedings of the National Academy of Sciences of the United States of America. 2006, 103, 915-20.
[43] Hornak, V., Okur, A., Rizzo, R.C., Simmerling, C. HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. J Am Chem Soc. 2006, 128, 2812-3.
[44] Hornak, V., Simmerling, C. Targeting structural flexibility in HIV-1 protease inhibitor binding. Drug Discovery Today. 2007, 12, 132-8.
[45] Cai, Y., Yilmaz, N.K., Myint, W., Ishima, R., Schiffer, C.A. Differential Flap Dynamics in Wild-type and a Drug Resistant Variant of HIV-1 Protease Revealed by Molecular Dynamics and NMR Relaxation. J Chem Theory Comput. 2012, 8, 3452-62.
[46] Heyda, J., Pokorna, J., Vrbka, L., Vacha, R., Jagoda-Cwiklik, B., Konvalinka, J., et al. Ion specific effects of sodium and potassium on the catalytic activity of HIV-1 protease. Phys Chem Chem Phys. 2009, 11, 7599-604.
[47] Minh, D.D., Chang, C.E., Trylska, J., Tozzini, V., McCammon, J.A. The influence of macromolecular crowding on HIV-1 protease internal dynamics. J Am Chem Soc. 2006, 128, 6006-7.
[48] Amaro, R.E., Baron, R., McCammon, J.A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. J Comput Aided Mol Des. 2008, 22, 693-705.
[49] Durdagi, S., Mavromoustakos, T., Chronakis, N., Papadopoulos, M.G. Computational design of novel fullerene analogues as potential HIV-1 PR inhibitors: Analysis of the binding interactions between fullerene inhibitors and HIV-1 PR residues using 3D QSAR, molecular docking and molecular dynamics simulations. Bioorg Med Chem. 2008, 16, 9957-74.
[50] Okimoto, N., Futatsugi, N., Fuji, H., Suenaga, A., Morimoto, G., Yanai, R., et al. High-performance drug discovery: computational screening by combining docking and molecular dynamics simulations. PLoS Comput Biol. 2009, 5, e1000528.

[51] Chang, C.E.A., Trylska, J., Tozzini, V., McCammon, J.A. Binding pathways of ligands to HIV-1 protease: Coarse-grained and atomistic simulations. Chemical Biology & Drug Design. 2007, 69, 5-13.
[52] Pietrucci, F., Marinelli, F., Carloni, P., Laio, A. Substrate binding mechanism of HIV-1 protease from explicit-solvent atomistic simulations. J Am Chem Soc. 2009, 131, 11811-8.
[53] Rucker, P., Horn, A.H., Meiselbach, H., Sticht, H. A comparative study of HIV-1 and HTLV-I protease structure and dynamics reveals a conserved residue interaction network. J Mol Model. 2011.
[54] Park, H., Lee, S. Determination of the active site protonation state of beta-secretase from molecular dynamics simulation and docking experiment: implications for structure-based inhibitor design. J Am Chem Soc. 2003, 125, 16416-22.
[55] Mishra, S., Caflisch, A. Dynamics in the active site of beta-secretase: a network analysis of atomistic simulations. Biochemistry. 2011, 50, 9328-39.
[56] Bjelic, S., Nervall, M., Gutierrez-de-Teran, H., Ersmark, K., Hallberg, A., Aqvist, J. Computational inhibitor design against malaria plasmepsins. Cell Mol Life Sci. 2007, 64, 2285-305.
[57] Deleon, K.Y., Patel, A.P., Kuczera, K., Johnson, C.K., Jas, G.S. Structure and reorientational dynamics of angiotensin I and II: a microscopic physical insight. J Biomol Struct Dyn. 2012, 29, 671-90.
[58] Constanciel, R., Contreras, R. Self-Consistent Field-Theory of Solvent Effects Representation by Continuum Models - Introduction of Desolvation Contribution. Theor Chim Acta. 1984, 65, 1-11.
[59] Still, W.C., Tempczyk, A., Hawley, R.C., Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. J Am Chem Soc. 1990, 112, 6127-9.
[60] Qiu, D., Shenkin, P.S., Hollinger, F.P., Still, W.C. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. J Phys Chem A. 1997, 101, 3005-14.
[61] Geney, R., Layten, M., Gomperts, R., Hornak, V., Simmerling, C. Investigation of salt bridge stability in a generalized born solvent model. Journal of Chemical Theory and Computation. 2006, 2, 115-27.
[62] Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., et al. The Amber biomolecular simulation programs. Journal of Computational Chemistry. 2005, 26, 1668-88.
[63] Onufriev, A., Bashford, D., Case, D.A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins-Structure Function and Bioinformatics. 2004, 55, 383-94.
[64] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins-Structure Function and Bioinformatics. 2006, 65, 712-25.
[65] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L. Comparison of Simple Potential Functions for Simulating Liquid Water. Journal of Chemical Physics. 1983, 79, 926-35.
[66] Baker, N.A. Improving implicit solvent simulations: a Poisson-centric view. Curr Opin Struc Biol. 2005, 15, 137-43.
[67] Kear, J.L., Blackburn, M.E., Veloro, A.M., Dunn, B.M., Fanucci, G.E. Subtype Polymorphisms Among HIV-1 Protease Variants Confer Altered Flap Conformations and Flexibility. J Am Chem Soc. 2009, 131, 14650-1.

[68] Ali, A., Bandaranayake, R.M., Cai, Y., King, N.M., Kolli, M., Mittal, S., et al. Molecular Basis for Drug Resistance in HIV-1 Protease. Viruses. 2010, 2, 2509-35.
[69] Alder, B.J., Wainwright, T.E. Studies in Molecular Dynamics .1. General Method. Journal of Chemical Physics. 1959, 31, 459-66.
[70] Rahman, A. Correlations in Motion of Atoms in Liquid Argon. Physical Review A - General Physics. 1964, 136, 405-11.
[71] Sadiq, S.K., De Fabritiis, G. Explicit solvent dynamics and energetics of HIV-1 protease flap opening and closing. Proteins. 2010, 78, 2873-85.
[72] Bashford, D., Case, D.A. Generalized Born Models of Macromolecular Solvation Effects. Annu. Rev. Phys. Chem. 2000, 51, 129-52.
[73] Vickie Tsui, David A. Case. Theory and applications of the generalized born solvation model in macromolecular simulations. Biopolymers. 2000, 56, 275-91.
[74] Feig, M., Im, W., Brooks, C.L. Implicit solvation based on generalized Born theory in different dielectric environments. Journal of Chemical Physics. 2004, 120, 903-11.
[75] Chen, J.H., Im, W.P., Brooks, C.L. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. J Am Chem Soc. 2006, 128, 3728-36.
[76] Chen, J., Brooks III, C.L., Khandogin, J. Recent advances in implicit solvent-based methods for biomolecular simulations. Curr Opin Struc Biol. 2008, 18, 140-8.
[77] Ponder, J.W., Case, D.A. Force fields for protein simulations. Protein Simulations. 2003, 66, 27-85.
[78] Guvench, O., MacKerell, A.D. Comparison of Protein Force Fields for Molecular Dynamics Simulations. In: Methods in Molecular Biology, 2008, Vol. 443, pp. 63-88.
[79] Hess, B., van der Vegt, N.F.A. Hydration thermodynamic properties of amino acid analogues: A systematic comparison of biomolecular force fields and water models. J Phys Chem B. 2006, 110, 17616-26.
[80] Shell, M.S., Ritterson, R., Dill, K.A. A test on peptide stability of AMBER force fields with implicit solvation. J Phys Chem B. 2008, 112, 6878-86.
[81] Penev, E., Ireta, J., Shea, J.E. Energetics of infinite homopolypeptide chains: A new look at commonly used force fields. J Phys Chem B. 2008, 112, 6872-7.
[82] Matthes, D., de Groot, B.L. Secondary Structure Propensities in Peptide Folding Simulations: A Systematic Comparison of Molecular Mechanics Interaction Schemes. Biophysical Journal. 2009, 97, 599-608.
[83] Wang, J.M., Cieplak, P., Kollman, P.A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? Journal of Computational Chemistry. 2000, 21, 1049-74.
[84] Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules J Am Chem Soc. 1996, 118, 2309-.
[85] Terada, T., Shimizu, K. A comparison of generalized Born methods in folding simulations. Chemical Physics Letters. 2008, 460, 295-9.
[86] Hawkins, G.D., Cramer, C.J., Truhlar, D.G. Pairwise solute descreening of solute charges from a dielectric medium. Chemical Physics Letters. 1995, 246, 122-9.
[87] Luca, S., Yau, W.M., Leapman, R., Tycko, R. Peptide conformation and supramolecular organization in amylin fibrils: Constraints from solid-state NMR. Biochemistry. 2007, 46, 13505-22.

[88] Markwick, P.R.L., Bouvignies, G., Blackledge, M. Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy. J Am Chem Soc. 2007, 129, 4724-30.
[89] Best, R.B., Hummert, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. J Phys Chem B. 2009, 113, 9004-15.
[90] Roe, D.R., Okur, A., Wickstrom, L., Hornak, V., Simmerling, C. Secondary structure bias in generalized born solvent models: Comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation. J Phys Chem B. 2007, 111, 1846-57.
[91] Okur, A., Wickstrom, L., Simmerling, C. Evaluation of Salt Bridge Structure and Energetics in Peptides Using Explicit, Implicit, and Hybrid Solvation Models. Journal of Chemical Theory and Computation. 2008, 4, 488-98.
[92] Amaro, R.E., Cheng, X.L., Ivanov, I., Xu, D., McCammon, J.A. Characterizing Loop Dynamics and Ligand Recognition in Human- and Avian-Type Influenza Neuraminidases via Generalized Born Molecular Dynamics and End-Point Free Energy Calculations. J Am Chem Soc. 2009, 131, 4702-9.
[93] Chocholousova, J., Feig, M. Implicit solvent simulations of DNA and DNA-protein complexes: Agreement with explicit solvent vs experiment. J Phys Chem B. 2006, 110, 17240-51.
[94] Felts, A.K., Gallicchio, E., Chekmarev, D., Paris, K.A., Friesner, R.A., Levy, R.M. Prediction of protein loop conformations using the AGBNP implicit solvent model and torsion angle sampling. Journal of Chemical Theory and Computation. 2008, 4, 855-68.
[95] Mongan, J., Simmerling, C., McCammon, J.A., Case, D.A., Onufriev, A. Generalized Born model with a simple, robust molecular volume correction. Journal of Chemical Theory and Computation. 2007, 3, 156-69.
[96] Galiano, L., Ding, F., Veloro, A.M., Blackburn, M.E., Simmerling, C., Fanucci, G.E. Drug Pressure Selected Mutations in HIV-1 Protease Alter Flap Conformations. J Am Chem Soc. 2009, 131, 430-1.
[97] Ding, F., Layten, M., Simmerling, C. Solution structure of HIV-1 protease flaps probed by comparison of molecular dynamics simulation ensembles and EPR experiments. J Am Chem Soc. 2008, 130, 7184-5.
[98] Bondi, A. Van Der Waals Volumes and Radii. Journal of Physical Chemistry. 1964, 68, 441-51.
[99] Tsui, V., Case, D.A. Molecular dynamics simulations of nucleic acids with a generalized born solvation model. J Am Chem Soc. 2000, 122, 2489-98.
[100] Lam, P.Y.S., Jadhav, P.K., Eyermann, C.J., Hodge, C.N., Ru, Y., Bacheler, L.T., et al. Rational Design of Potent, Bioavailable, Nonpeptide Cyclic Ureas as Hiv Protease Inhibitors. Science. 1994, 263, 380-4.
[101] Tyndall, J.D., Pattenden, L.K., Reid, R.C., Hu, S.H., Alewood, D., Alewood, P.F., et al. Crystal structures of highly constrained substrate and hydrolysis products bound to HIV-1 protease. Implications for the catalytic mechanism. Biochemistry. 2008, 47, 3736-44.
[102] Altman, M.D., Nalivaika, E.A., Prabu-Jeyabalan, M., Schiffer, C.A., Tidor, B. Computational design and experimental study of tighter binding peptides to an inactivated mutant of HIV-1 protease. Proteins. 2008, 70, 678-94.
[103] Ciesla, M., Dias, S.P., Longa, L., Oliveira, F.A. Synchronization induced by Langevin dynamics. Physical Review E. 2001, 6306.

[104] Spinelli, S., Liu, Q.Z., Alzari, P.M., H., H.P., Poljak, R.J. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. Biochimie. 1991, 73, 1391-6.
[105] Simmerling, C., Elber, R., Zhang, J. MOIL-View - A Program for Visualization of Structure and Dynamics of Biomolecules and STO - A Program for Computing Stochastic Paths. Modelling of Biomolecular Structures and Mechanisms. 1995, 241-465.
[106] Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., Honig, B. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. Journal of Computational Chemistry. 2002, 23, 128-37.
[107] Onufriev, A., Case, D.A., Bashford, D. Effective Born radii in the generalized Born approximation: The importance of being perfect. Journal of Computational Chemistry. 2002, 23, 1297-304.
[108] Zhu, J., Alexov, E., Honig, B. Comparative study of generalized Born models: Born radii and peptide folding. J Phys Chem B. 2005, 109, 3008-22.
[109] Layten, M., Hornak, V., Simmerling, C. The open structure of a multi-drug-resistant HIV-1 protease is stabilized by crystal packing contacts. J Am Chem Soc. 2006, 128, 13360-1.
[110] Sayer, J.M., Liu, F., Ishima, R., Weber, I.T., Louis, J.M. Effect of the active site D25N mutation on the structure, stability, and ligand binding of the mature HIV-1 protease. J Biol Chem. 2008, 283, 13459-70.
[111] Swanson, J.M.J., Adcock, S.A., McCammon, J.A. Optimized radii for Poisson-Boltzmann calculations with the AMBER force field. Journal of Chemical Theory and Computation. 2005, 1, 484-93.
[112] Preston, B.D., Dougherty, J.P. Mechanisms of retroviral mutation. Trends Microbiol. 1996, 4, 16-21.
[113] Tebit, D.M., Nankya, I., Arts, E.J., Gao, Y. HIV Diversity, Recombination and Disease Progression: How Does Fitness "Fit" Into the Puzzle? AIDS Reviews. 2007, 9, 75-87.
[114] Hemelaar, J., Gouws, E., Ghys, P.D., Osmanov, S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. Aids. 2006, 20, W13-23.
[115] Velazquez-Campoy, A., Vega, S., Freire, E. Amplification of the effects of drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes. Biochemistry. 2002, 41, 8613-9.
[116] Lisovsky, I., Schader, S.M., Martinez-Cajas, J.L., Oliveira, M., Moisi, D., Wainberg, M.A. HIV-1 protease codon 36 polymorphisms and differential development of resistance to nelfinavir, lopinavir, and atazanavir in different HIV-1 subtypes. Antimicrob Agents Chemother. 2010, 54, 2878-85.
[117] Johnson, V.A., Brun-Vezinet, F., Clotet, B., Gunthard, H.F., Kuritzkes, D.R., Pillay, D., et al. Update of the Drug Resistance Mutations in HIV-1: December 2010. Top HIV Med. 2010, 18, 156-63.
[118] Ganser-Pornillos, B.K., Yeager, M., Sundquist, W.I. The structural biology of HIV assembly. Curr Opin Struct Biol. 2008, 18, 203-17.
[119] Martin, P., Vickrey, J.F., Proteasa, G., Jimenez, Y.L., Wawrzak, Z., Winters, M.A., et al. "Wide-open" 1.3 angstrom structure of a multidrug-resistant HIV-1 protease as a drug target. Structure. 2005, 13, 1887-95.
[120] Heaslet, H., Rosenfeld, R., Giffin, M., Lin, Y.C., Tam, K., Torbett, B.E., et al. Conformational flexibility in the flap domains of ligand-free HIV protease. Acta Crystallogr D Biol Crystallogr. 2007, 63, 866-75.

[121] Pillai, B., Kannan, K.K., Hosur, M.V. 1.9 A x-ray study shows closed flap conformation in crystals of tethered HIV-1 PR. Proteins. 2001, 43, 57-64.
[122] Kumar, M., Kannan, K.K., Hosur, M.V., Bhavesh, N.S., Chatterjee, A., Mittal, R., et al. Effects of remote mutation on the autolysis of HIV-1 PR: X-ray and NMR investigations. Biochemical and biophysical research communications. 2002, 294, 395-401.
[123] Freedberg, D.I., Ishima, R., Jacob, J., Wang, Y.X., Kustanovich, I., Louis, J.M., et al. Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations. Protein Sci. 2002, 11, 221-32.
[124] Ishima, R., Freedberg, D.I., Wang, Y.X., Louis, J.M., Torchia, D.A. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. Structure. 1999, 7, 1047-55.
[125] Ishima, R., Torchia, D.A. Protein dynamics from NMR. Nat Struct Biol. 2000, 7, 740-3.
[126] Csermely, P., Palotai, R., Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci. 2010, 35, 539-46.
[127] Galiano, L., Bonora, M., Fanucci, G.E. Interflap distances in HIV-1 protease determined by pulsed EPR measurements. J Am Chem Soc. 2007, 129, 11004-+.
[128] Torbeev, V.Y., Raghuraman, H., Mandal, K., Senapati, S., Perozo, E., Kent, S.B.H. Dynamics of "Flap" Structures in Three HIV-1 Protease/Inhibitor Complexes Probed by Total Chemical Synthesis and Pulse-EPR Spectroscopy. J Am Chem Soc. 2009, 131, 884-+.
[129] Torbeev, V.Y., Raghuraman, H., Hamelberg, D., Tonelli, M., Westler, W.M., Perozo, E., et al. Protein conformational dynamics in the mechanism of HIV-1 protease catalysis. Proc Natl Acad Sci U S A. 2011, 108, 20982-7.
[130] Shang, Y., Nguyen, H., Wickstrom, L., Okur, A., Simmerling, C. Improving the description of salt bridge strength and geometry in a Generalized Born model. J Mol Graph Model. 2011, 29, 676-84.
[131] Coman, R.M., Robbins, A.H., Fernandez, M.A., Gilliland, C.T., Sochet, A.A., Goodenow, M.M., et al. The contribution of naturally occurring polymorphisms in altering the biochemical and structural characteristics of HIV-1 subtype C protease. Biochemistry. 2008, 47, 731-43.
[132] Guex, N., Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis. 1997, 18, 2714-23.
[133] Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., Case, D.A. Development and testing of a general amber force field. J Comput Chem. 2004, 25, 1157-74.
[134] Wang, J., Wang, W., Kollman, P.A., Case, D.A. Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model. 2006, 25, 247-60.
[135] Levy, R.M., Zhang, L.Y., Gallicchio, E., Felts, A.K. On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. J Am Chem Soc. 2003, 125, 9523-30.
[136] Wagoner, J.A., Baker, N.A. Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. Proc Natl Acad Sci U S A. 2006, 103, 8331-6.
[137] Chen, J., Brooks, C.L., 3rd. Critical importance of length-scale dependence in implicit modeling of hydrophobic interactions. J Am Chem Soc. 2007, 129, 2444-5.
[138] Chen, J., Brooks, C.L. Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions. Physical Chemistry Chemical Physics. 2008, 10, 471-81.

[139] Sadiq, S.K., Wright, D., Watson, S.J., Zasada, S.J., Stoica, I., Coveney, P.V. Automated Molecular Simulation Based Binding Affinity Calculator for Ligand-Bound HIV-1 Proteases. Journal of Chemical Information and Modeling. 2008, 48, 1909-19.
[140] Gilson, M.K., Zhou, H.X. Calculation of protein-ligand binding affinities. Annual Review of Biophysics and Biomolecular Structure. 2007, 36, 21-42.
[141] Cai, Y., Schiffer, C.A. Decomposing the Energetic Impact of Drug Resistant Mutations in HIV-1 Protease on Binding DRV. Journal of Chemical Theory and Computation. 2010, 6, 1358-68.
[142] Hunter, J.D. Matplotlib: A 2D graphics environment. Computing in Science & Engineering. 2007, 9, 90-5.
[143] Fanucci, G.E., Cafiso, D.S. Recent advances and applications of site-directed spin labeling. Curr Opin Struct Biol. 2006, 16, 644-53.
[144] Polyhach, Y., Bordignon, E., Jeschke, G. Rotamer libraries of spin labelled cysteines for protein studies. Phys Chem Chem Phys. 2011, 13, 2356-66.
[145] Meher, B.R., Wang, Y. Interaction of I50V mutant and I50L/A71V double mutant HIV-protease with inhibitor TMC114 (darunavir): molecular dynamics simulation and binding free energy studies. J Phys Chem B. 2012, 116, 1884-900.
[146] Dahl, S.G., Edvardsen, O., Sylte, I. Molecular dynamics of dopamine at the D2 receptor. Proc Natl Acad Sci U S A. 1991, 88, 8111-5.
[147] Kurt, N., Scott, W.R., Schiffer, C.A., Haliloglu, T. Cooperative fluctuations of unliganded and substrate-bound HIV-1 protease: a structure-based analysis on a variety of conformations from crystallography and molecular dynamics simulations. Proteins. 2003, 51, 409-22.
[148] Arkhipov, A., Shan, Y., Das, R., Endres, N.F., Eastwood, M.P., Wemmer, D.E., et al. Architecture and membrane interactions of the EGF receptor. Cell. 2013, 152, 557-69.
[149] McKeage, K., Perry, C.M., Keam, S.J. Darunavir: a review of its use in the management of HIV infection in adults. Drugs. 2009, 69, 477-503.
[150] Chang, C.-E., Chen, W., Gilson, M.K. Evaluating the Accuracy of the Quasiharmonic Approximation. Journal of Chemical Theory and Computation. 2005, 1, 1017-28.
[151] Wittayanarakul, K., Hannongbua, S., Feig, M. Accurate prediction of protonation state as a prerequisite for reliable MM-PB(GB)SA binding free energy calculations of HIV-1 protease inhibitors. J Comput Chem. 2008, 29, 673-85.
[152] Chennubhotla, C., Bahar, I. Signal propagation in proteins and relation to equilibrium fluctuations. PLoS Comput Biol. 2007, 3, 1716-26.
[153] Morra, G., Verkhivker, G., Colombo, G. Modeling signal propagation mechanisms and ligand-based conformational dynamics of the Hsp90 molecular chaperone full-length dimer. PLoS Comput Biol. 2009, 5, e1000323.
[154] Vega, S., Kang, L.-W., Velazquez-Campoy, A., Kiso, Y., Amzel, L.M., Freire, E. A structural and thermodynamic escape mechanism from a drug resistant mutation of the HIV-1 protease. Proteins: Structure, Function, and Bioinformatics. 2004, 55, 594-602.
[155] Jakalian, A., Jack, D.B., Bayly, C.I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem. 2002, 23, 1623-41.
[156] Steinbrecher, T., Mobley, D.L., Case, D.A. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. Journal of Chemical Physics. 2007, 127, -.

[157] Killian, B.J., Kravitz, J.Y., Gilson, M.K. Extraction of configurational entropy from molecular simulations via an expansion approximation. Journal of Chemical Physics. 2007, 127, 16.
[158] Killian, B.J., Kravitz, J.Y., Somani, S., Dasgupta, P., Pang, Y.P., Gilson, M.K. Configurational Entropy in Protein-Peptide Binding: Computational Study of Tsg101 Ubiquitin E2 Variant Domain with an HIV-Derived PTAP Nonapeptide. Journal of Molecular Biology. 2009, 389, 315-35.
[159] Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., et al. Gaussian 98 (Revision A.1x). 2001.
[160] Dupradeau, F.Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., et al. The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. Phys Chem Chem Phys. 2010, 12, 7821-39.
[161] Piana, S., Carloni, P., Rothlisberger, U. Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations. Protein Sci. 2002, 11, 2393-402.
[162] Clemente, J.C., Moose, R.E., Hemrajani, R., Whitford, L.R.S., Govindasamy, L., Reutzel, R., et al. Comparing the accumulation of active- and nonactive-site mutations in the HIV-1 protease. Biochemistry. 2004, 43, 12141-51.
[163] Foulkes-Murzycki, J.E., Scott, W.R., Schiffer, C.A. Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. Structure. 2007, 15, 225-33.
[164] De Meyer, S., Azijn, H., Surleraux, D., Jochmans, D., Tahri, A., Pauwels, R., et al. TMC114, a novel human immunodeficiency virus type 1 protease inhibitor active against protease inhibitor-resistant viruses, including a broad range of clinical isolates. Antimicrob Agents Chemother. 2005, 49, 2314-21.
[165] Surleraux, D.L., Tahri, A., Verschueren, W.G., Pille, G.M., de Kock, H.A., Jonckers, T.H., et al. Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. J Med Chem. 2005, 48, 1813-22.
[166] Cihlar, T., He, G.-X., Liu, X., Chen, J.M., Hatada, M., Swaminathan, S., et al. Suppression of HIV-1 Protease Inhibitor Resistance by Phosphonate-mediated Solvent Anchoring. Journal of Molecular Biology. 2006, 363, 635-47.
[167] Peters, B.S., Conway, K. Therapy for HIV: past, present, and future. Adv Dent Res. 2011, 23, 23-7.
[168] Boross, P., Bagossi, P., Weber, I.T., Tozser, J. Drug targets in human T-lymphotropic virus type 1 (HTLV-1) infection. Infect Disord Drug Targets. 2009, 9, 159-71.
[169] Roggo, S. Inhibition of BACE, a promising approach to Alzheimer's disease therapy. Curr Top Med Chem. 2002, 2, 359-70.
[170] Gruninger-Leitch, F., Schlatter, D., Kung, E., Nelbock, P., Dobeli, H. Substrate and inhibitor profile of BACE (beta-secretase) and comparison with other mammalian aspartic proteases. J Biol Chem. 2002, 277, 4687-93.
[171] Barman, A., Schurer, S., Prabhakar, R. Computational modeling of substrate specificity and catalysis of the beta-secretase (BACE1) enzyme. Biochemistry. 2011, 50, 4337-49.
[172] Cascella, M., Micheletti, C., Rothlisberger, U., Carloni, P. Evolutionarily conserved functional mechanics across pepsin-like and retroviral aspartic proteases. J Am Chem Soc. 2005, 127, 3734-42.
[173] Wlodawer, A., Gustchina, A. Structural and biochemical studies of retroviral proteases. Biochim Biophys Acta. 2000, 1477, 16-34.

[174] Li, M., Dimaio, F., Zhou, D., Gustchina, A., Lubkowski, J., Dauter, Z., et al. Crystal structure of XMRV protease differs from the structures of other retropepsins. Nat Struct Mol Biol. 2011, 18, 227-9.
[175] Paprotka, T., Delviks-Frankenberry, K.A., Cingoz, O., Martinez, A., Kung, H.J., Tepper, C.G., et al. Recombinant origin of the retrovirus XMRV. Science. 2011, 333, 97-101.
[176] Cingoz, O., Paprotka, T., Delviks-Frankenberry, K.A., Wildt, S., Hu, W.S., Pathak, V.K., et al. Characterization, Mapping and Distribution of the Two XMRV Parental Proviruses. J Virol. 2011.
[177] Krylov, D.M., Koonin, E.V. A novel family of predicted retroviral-like aspartyl proteases with a possible key role in eukaryotic cell cycle control. Curr Biol. 2001, 11, R584-7.
[178] Sirkis, R., Gerst, J.E., Fass, D. Ddi1, a eukaryotic protein with the retroviral protease fold. J Mol Biol. 2006, 364, 376-87.
[179] Su, V., Lau, A.F. Ubiquitin-like and ubiquitin-associated domain proteins: significance in proteasomal degradation. Cell Mol Life Sci. 2009, 66, 2819-33.
[180] White, R.E., Dickinson, J.R., Semple, C.A., Powell, D.J., Berry, C. The retroviral proteinase active site and the N-terminus of Ddi1 are required for repression of protein secretion. FEBS Lett. 2011, 585, 139-42.
[181] Gabriely, G., Kama, R., Gelin-Licht, R., Gerst, J.E. Different domains of the UBL-UBA ubiquitin receptor, Ddi1/Vsm1, are involved in its multiple cellular roles. Mol Biol Cell. 2008, 19, 3625-37.
[182] Chandra, N., Anand, P., Yeturu, K. Structural bioinformatics: deriving biological insights from protein structures. Interdiscip Sci. 2010, 2, 347-66.
[183] Hill, E.E., Morea, V., Chothia, C. Sequence conservation in families whose members have little or no sequence similarity: the four-helical cytokines and cytochromes. J Mol Biol. 2002, 322, 205-33.
[184] Eargle, J., Black, A.A., Sethi, A., Trabuco, L.G., Luthey-Schulten, Z. Dynamics of Recognition between tRNA and elongation factor Tu. J Mol Biol. 2008, 377, 1382-405.
[185] Roberts, E., Sethi, A., Montoya, J., Woese, C.R., Luthey-Schulten, Z. Molecular signatures of ribosomal evolution. Proc Natl Acad Sci U S A. 2008, 105, 13953-8.
[186] O'Donoghue, P., Luthey-Schulten, Z. Evolutionary profiles derived from the QR factorization of multiple structural alignments gives an economy of information. J Mol Biol. 2005, 346, 875-94.
[187] Li, M., Gustchina, A., Matuz, K., Tozser, J., Namwong, S., Goldfarb, N.E., et al. Structural and biochemical characterization of the inhibitor complexes of xenotropic murine leukemia virus-related virus protease. FEBS J. 2011, 278, 4413-24.
[188] Hong, L., Tang, J. Flap position of free memapsin 2 (beta-secretase), a model for flap opening in aspartic protease catalysis. Biochemistry. 2004, 43, 4689-95.
[189] Gorfe, A.A., Caflisch, A. Functional plasticity in the substrate binding site of beta-secretase. Structure. 2005, 13, 1487-98.
[190] Spronk, S.A., Carlson, H.A. The role of tyrosine 71 in modulating the flap conformations of BACE1. Proteins. 2011, 79, 2247-59.
[191] Roberts, E., Eargle, J., Wright, D., Luthey-Schulten, Z. MultiSeq: unifying sequence and structure data for evolutionary analysis. BMC bioinformatics. 2006, 7, 382.
[192] Gustchina, A., Kervinen, J., Powell, D.J., Zdanov, A., Kay, J., Wlodawer, A. Structure of equine infectious anemia virus proteinase complexed with an inhibitor. Protein Science. 1996, 5, 1453-65.

[193] Fujinaga, M., Chernaia, M.M., Tarasova, N.I., Mosimann, S.C., James, M.N. Crystal structure of human pepsin and its complex with pepstatin. Protein Sci. 1995, 4, 960-72.
[194] Thompson, J.D., Higgins, D.G., Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994, 22, 4673-80.
[195] Laco, G.S., Schalk-Hihi, C., Lubkowski, J., Morris, G., Zdanov, A., Olson, A., et al. Crystal structures of the inactive D30N mutant of feline immunodeficiency virus protease complexed with a substrate and an inhibitor. Biochemistry. 1997, 36, 10696-708.
[196] Wu, J., Adomat, J.M., Ridky, T.W., Louis, J.M., Leis, J., Harrison, R.W., et al. Structural Basis for Specificity of Retroviral Proteases. Biochemistry. 1998, 37, 4518-26.
[197] Mulichak, A.M., Hui, J.O., Tomasselli, A.G., Heinrikson, R.L., Curry, K.A., Tomich, C.S., et al. The crystallographic structure of the protease from human immunodeficiency virus type 2 with two synthetic peptidic transition state analog inhibitors. J Biol Chem. 1993, 268, 13103-9.
[198] Sanches, M., Krauchenco, S., Martins, N.H., Gustchina, A., Wlodawer, A., Polikarpov, I. Structural characterization of B and non-B subtypes of HIV-protease: insights into the natural susceptibility to drug resistance development. J Mol Biol. 2007, 369, 1029-40.
[199] Steele, T.G., Hills, I.D., Nomland, A.A., de Leon, P., Allison, T., McGaughey, G., et al. Identification of a small molecule beta-secretase inhibitor that binds without catalytic aspartate engagement. Bioorg Med Chem Lett. 2009, 19, 17-20.
[200] Yang, J., Quail, J.W. Structure of the Rhizomucor miehei aspartic proteinase complexed with the inhibitor pepstatin A at 2.7 A resolution. Acta Crystallogr D Biol Crystallogr. 1999, 55, 625-30.
[201] Li, M., Gustchina, A., Glesner, J., Wunschmann, S., Vailes, L.D., Chapman, M.D., et al. Carbohydrates contribute to the interactions between cockroach allergen Bla g 2 and a monoclonal antibody. J Immunol. 2011, 186, 333-40.
[202] Borelli, C., Ruge, E., Lee, J.H., Schaller, M., Vogelsang, A., Monod, M., et al. X-ray structures of Sap1 and Sap5: structural comparison of the secreted aspartic proteinases from Candida albicans. Proteins. 2008, 72, 1308-19.
[203] Fujimoto, Z., Fujii, Y., Kaneko, S., Kobayashi, H., Mizuno, H. Crystal structure of aspartic proteinase from Irpex lacteus in complex with inhibitor pepstatin. J Mol Biol. 2004, 341, 1227-35.
[204] Nascimento, A.S., Krauchenco, S., Golubev, A.M., Gustchina, A., Wlodawer, A., Polikarpov, I. Statistical coupling analysis of aspartic proteinases based on crystal structures of the Trichoderma reesei enzyme and its complex with pepstatin A. J Mol Biol. 2008, 382, 763-78.
[205] Russell, R.B., Barton, G.J. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. Proteins. 1992, 14, 309-23.
[206] Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., Barton, G.J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009, 25, 1189-91.
[207] Stamatakis, A., Hoover, P., Rougemont, J. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol. 2008, 57, 758-71.
[208] Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E., et al. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics. 2003, 19, 163-4.

[209] Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res. 2005, 33, W299-302.
[210] Ashkenazy, H., Erez, E., Martz, E., Pupko, T., Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 2010, 38, W529-33.
[211] Katoh, K., Misawa, K., Kuma, K., Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002, 30, 3059-66.
[212] Davis, I.W., Murray, L.W., Richardson, J.S., Richardson, D.C. MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Research. 2004, 32, W615-W9.
[213] Dang, L.X. Mechanism and Thermodynamics of Ion Selectivity in Aqueous-Solutions of 18-Crown-6 Ether - a Molecular-Dynamics Study. J Am Chem Soc. 1995, 117, 6954-60.
[214] Weiser, J., Shenkin, P.S., Still, W.C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). Journal of Computational Chemistry. 1999, 20, 217-30.
[215] Shang, Y., Simmerling, C. Molecular dynamics applied in drug discovery: the case of HIV-1 protease. Methods Mol Biol. 2012, 819, 527-49.
[216] Patel, S., Vuillard, L., Cleasby, A., Murray, C.W., Yon, J. Apo and inhibitor complex structures of BACE (beta-secretase). J Mol Biol. 2004, 343, 407-16.
[217] Hong, L., Koelsch, G., Lin, X., Wu, S., Terzyan, S., Ghosh, A.K., et al. Structure of the protease domain of memapsin 2 (beta-secretase) complexed with inhibitor. Science. 2000, 290, 150-3.
[218] Hulko, M., Lupas, A.N., Martin, J. Inherent chaperone-like activity of aspartic proteases reveals a distant evolutionary relation to double-psi barrel domains of AAA-ATPases. Protein Sci. 2007, 16, 644-53.
[219] Weiss, R.A. Retrovirus classification and cell interactions. Journal of Antimicrobial Chemotherapy. 1996, 37, 1-11.
[220] Tang, H.L., Kuhen, K.L., Wong-Staal, F. Lentivirus replication and regulation. Annual Review of Genetics. 1999, 33, 133-70.
[221] Rawlings, N.D., Bateman, A. Pepsin homologues in bacteria. BMC Genomics. 2009, 10, 437.
[222] Castillo, R.M., Mizuguchi, K., Dhanaraj, V., Albert, A., Blundell, T.L., Murzin, A.G. A six-stranded double-psi beta barrel is shared by several protein superfamilies. Structure. 1999, 7, 227-36.
[223] Andreeva, N.S., Rumsh, L.D. Analysis of crystal structures of aspartic proteinases: on the role of amino acid residues adjacent to the catalytic site of pepsin-like enzymes. Protein Sci. 2001, 10, 2439-50.
[224] Turner, R.T., 3rd, Koelsch, G., Hong, L., Castanheira, P., Ermolieff, J., Ghosh, A.K., et al. Subsite specificity of memapsin 2 (beta-secretase): implications for inhibitor design. Biochemistry. 2001, 40, 10001-6.
[225] Turner, R.T., 3rd, Loy, J.A., Nguyen, C., Devasamudram, T., Ghosh, A.K., Koelsch, G., et al. Specificity of memapsin 1 and its implications on the design of memapsin 2 (beta-secretase) inhibitor selectivity. Biochemistry. 2002, 41, 8742-6.

[226] Hong, L., Hartsuck, J.A., Foundling, S., Ermolieff, J., Tang, J. Active-site mobility in human immunodeficiency virus, type 1, protease as demonstrated by crystal structure of A28S mutant. Protein Sci. 1998, 7, 300-5.
[227] Li, D., Ji, B., Hwang, K., Huang, Y. Crucial roles of the subnanosecond local dynamics of the flap tips in the global conformational changes of HIV-1 protease. J Phys Chem B. 2010, 114, 3060-9.
[228] Louis, J.M., Nashed, N.T., Parris, K.D., Kimmel, A.R., Jerina, D.M. Kinetics and mechanism of autoprocessing of human immunodeficiency virus type 1 protease from an analog of the Gag-Pol polyprotein. Proc Natl Acad Sci U S A. 1994, 91, 7970-4.
[229] Wondrak, E.M., Nashed, N.T., Haber, M.T., Jerina, D.M., Louis, J.M. A transient precursor of the HIV-1 protease. Isolation, characterization, and kinetics of maturation. J Biol Chem. 1996, 271, 4477-81.
[230] Chemical Computing Group, I.M., Quebec, Canada. Molecular Operating Environment (MOE). 2011, Vol. Version 2011.10.