2009-11-02

Energetic and Dynamic Analysis of Inhibitor Binding to Drug-Resistant HIV-1 Proteases: A Dissertation

Yufeng Cai
University of Massachusetts Medical School

ENERGETIC AND DYNAMIC ANALYSIS OF INHIBITOR BINDING TO DRUG-RESISTANT HIV-1 PROTEASES

A Dissertation Presented

By

YUFENG CAI

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

NOVEMBER 2, 2009

BIOCHEMISTRY AND MOLECULAR PHARMACOLOGY

ENERGETIC AND DYNAMIC ANALYSIS OF INHIBITOR BINDING
TO DRUG-RESISTANT HIV-1 PROTEASE

A Dissertation Presented
By

YUFENG CAI

The signatures of the Dissertation Defense Committee signifies
completion and approval as to style and content of the Dissertation

Celia A. Schiffer, Ph.D., Thesis Advisor

Dan Bolon, Ph.D., Member of Committee

Stephen Miller, Ph.D., Member of Committee

David Lambright, Ph.D., Member of Committee

Michael Gilson, M.D., Ph.D., Member of Committee

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

C. Robert Matthews, Ph.D., Chair of Committee

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

Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

Program in Biochemistry and Molecular Pharmacology

November 2, 2009

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Celia Schiffer. Celia has been a wonderful mentor, always available for discussion and advice but never trying to micromanage. Without her continuous support and encouragement, I would not be able to finish my Ph.D. study. I would also thank all the former and current lab members from Schiffer's lab especially Ellen Nalivaika and Jennifer Foulkes-Murzycki. I appreciate Jeff Peng, Francesca Massi and Ramakrishna Vadrevu for their help in collecting the NMR data. I thank my research committee members C. Robert Matthews, Dan Bolon, Stephen Miller, David Lambright and Mike Gilson for their guidance on my research project. Most of all I would like to thank my family, from my grandparents to my sister, who have helped me throughout this process with their love.

ABSTRACT

HIV-1 protease is a very important drug target for AIDS therapy. Nine protease inhibitors have been proved by FDA and used in AIDS treatment. Due to the high replication rate and the lack of fidelity of the HIV-1 reverse transcriptase, HIV-1 virus developed various drug-resistant variants. Although experimental methods such as crystallography and isothermal titration calorimetry provide structural and thermodynamic data on drug-resistant variants, they are unable to discern the mechanism by which the mutations confer resistance to inhibitors. Understanding the drug-resistance mechanism is crucial for developing new inhibitors more tolerant to the drug-resistant mutations. Computational methods such as free energy calculations and molecular dynamic simulations can provide insights to the drug resistance mechanism at an atomic level. In this thesis, I have focused on the elucidation of the energetic and dynamics of key drug-resistant variants of HIV-1 protease.

Two multi-drug resistant variants, in comparison with wild-type HIV-1 protease were used for the comparisons: Flap+ (L10I, G48V, I54V, and V82A) which contains a combination of flap and active site mutations and ACT (V82T, I84V) that only contains active site mutations. In Chapter II, I applied free energy simulations and decomposition methods to study the differential mechanism of resistance to the two variants, Flap+ and ACT, to the recently FDA-approved protease inhibitor darunavir (DRV). In this study, the absolute and relative binding free energies of DRV with wild-type protease and the two protease variants were calculated with MM-PB/GBSA and thermodynamic integration methods, respectively. And the

predicted results are in good agreement with the ITC experimental results. Free energy decomposition elucidates the mutations alter not only its own interaction with DRV but also other residues by changing the geometry of binding pocket. And the VdW interactions between the bis-THF group of DRV is predominant even in the drug-resistant variants. At the end of this chapter, I offer suggestions on developing new inhibitors that are based on DRV but might be less susceptible to drug-resistant mutations.

In Chapter III, 20-ns MD simulations of the apo wildtype protease and the apo drug-resistant protease variant Flap+ are analyzed and compared. In these studies, these mutations have been found to decrease the protease flexibility in the apo form but increase the mobility when the protease is binding with inhibitor.

In Chapter IV, more details of the free energy simulation and decomposition are discussed. NMR relaxation experiments were set up as a control for the MD simulation study of the dynamics of the Flap+ variant. The difficulty of finishing the NMR experiment is discussed and the solution and some preliminary results are shown.

In summary, the scope of this thesis was to use computational methods to study drug-resistant protease variants' thermodynamic and dynamic properties to illuminate the mechanism of protease drug resistance. This knowledge will contribute to rational design of new protease inhibitors which bind more tightly to the protease and hinder the development of drug-resistant mutations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

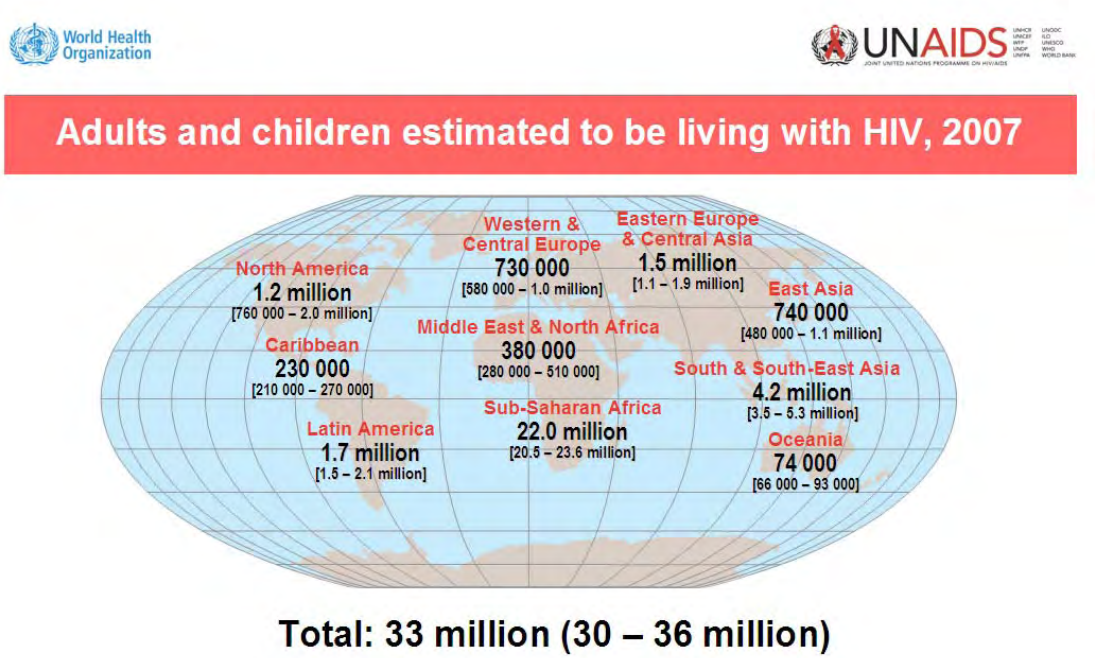
Chapter I

Introduction

Global AIDS Epidemic

The first patient diagnosed with the acquired immunodeficiency syndrome (AIDS) was in the United States in 1981. Since then, AIDS has come to attention of the medical community and quickly emerged as a public health threat. More than 25 million people over the world have died of the disease. The immune system of patients with the AIDS is weakened by the human immunodeficiency virus (HIV), leading to these patients becoming more vulnerable to different kinds of opportunistic infections. HIV is transmitted by three major routes: (1) sexual contact, (2) exposure to infected bodily fluids, including intravenous drugs, and (3) mother-to-child transmission. According to the Joint United Nations Program on HIV/AIDS and the World Health Organization AIDS epidemic statistics updated in July 2008 (referring to the end of 2007) approximately 33 million people globally are living with HIV infection (Figure 1.1). Of the 2.7 million people newly infected with HIV in 2007, more than half were people under 25 years old (Table 1.1). These devastating statistics shows that AIDS is one of the most urgent public health crises of modern society.

Figure 1.1: Global Prevalence of HIV-1 Infections



(http://www.avert.org/worldstats.htm)

Table 1.1: Global HIV/AIDS estimates by the end of 2007

(http://www.avert.org/worldstats.htm)

	Estimate	Range
People living with HIV/AIDS in 2007	33.0 million	30.3-36.1 million
Adults living with HIV/AIDS in 2007	30.8 million	28.2-34.0 million
Women living with HIV/AIDS in 2007	15.5 million	14.2-16.9 million
Children living with HIV/AIDS in 2007	2.0 million	1.9-2.3 million
People newly infected with HIV in 2007	2.7 million	2.2-3.2 million
Children newly infected with HIV in 2007	0.37 million	0.33-0.41 million
AIDS deaths in 2007	2.0 million	1.8-2.3 million
Child AIDS deaths in 2007	0.27 million	0.25-0.29 million

HIV life cycle

HIV, a retrovirus belonging to the *Retroviridae* family, impairs the human immune system by infecting immune cells such as CD4+ cells, macrophages and dendritic cells. The virus particle is roughly a sphere with diameter of 120 nm. Unlike most bacteria, HIV particles are much too small to be seen through an ordinary microscope. However they can be seen clearly with an electron microscope.

HIV has just nine genes (compared to more than 5,000 genes in a bacterium, and 20,000-25,000 in a human), which are encoded by two copies of single-stranded RNA which is 9749 nucleotides long. The *gag*, *pol,* and *env* genes carry information for the structural proteins (matrix [MA], capsid [CA], P6, and nucleocapsid [NC]), two envelope proteins (transmembrane [TM] and surface [SU]), and three enzymes (protease [PR], integrase [IN], and reverse transcriptase [RT]), respectively. The *tat*, *rev*, *nef*, *vif*, *vpr* and *vpu* genes code for 6 regulatory proteins and enzymes. The viral life cycle includes these stages: (1) entry, (2) reverse transcription, (3) integration and transcription, (4) translation, (5) budding, and (6) maturation (Figure 1.2).[1]

(1) Entry: HIV, an enveloped virus, enters cells by fusing the viral and host cell plasma membranes. This process is carried out by the viral envelope proteins SU (gp120) and TM (gp41). The SU protein binds to the N-terminal immunoglobulin domain of the CD4+ receptor, followed by interactions between the env protein and chemokine co-receptor, typically CCR5 or CXCR4. These interactions change

the conformation of the TM protein to expose it, facilitating fusion of the viral and host cell membranes. Finally, the viral genome is released into the host cell.

(2) Reverse transcription: After the virus enters the host cell, RT uses the host cell machinery to reverse transcribe the viral RNA to generate double-stranded DNA. This process is extremely error prone due to HIV RT lacking a proofreading mechanism. Thus, this step introduces many mutations into the viral genetic code.

(3) Integration and transcription: The synthesized double-stranded DNA is transported into the host cell nucleus with other viral proteins such as RT and IN. In the host cell nucleus, IN inserts viral cDNA into the host cell chromosomes. After transcription, the viral mRNA is exported to the cytoplasm.

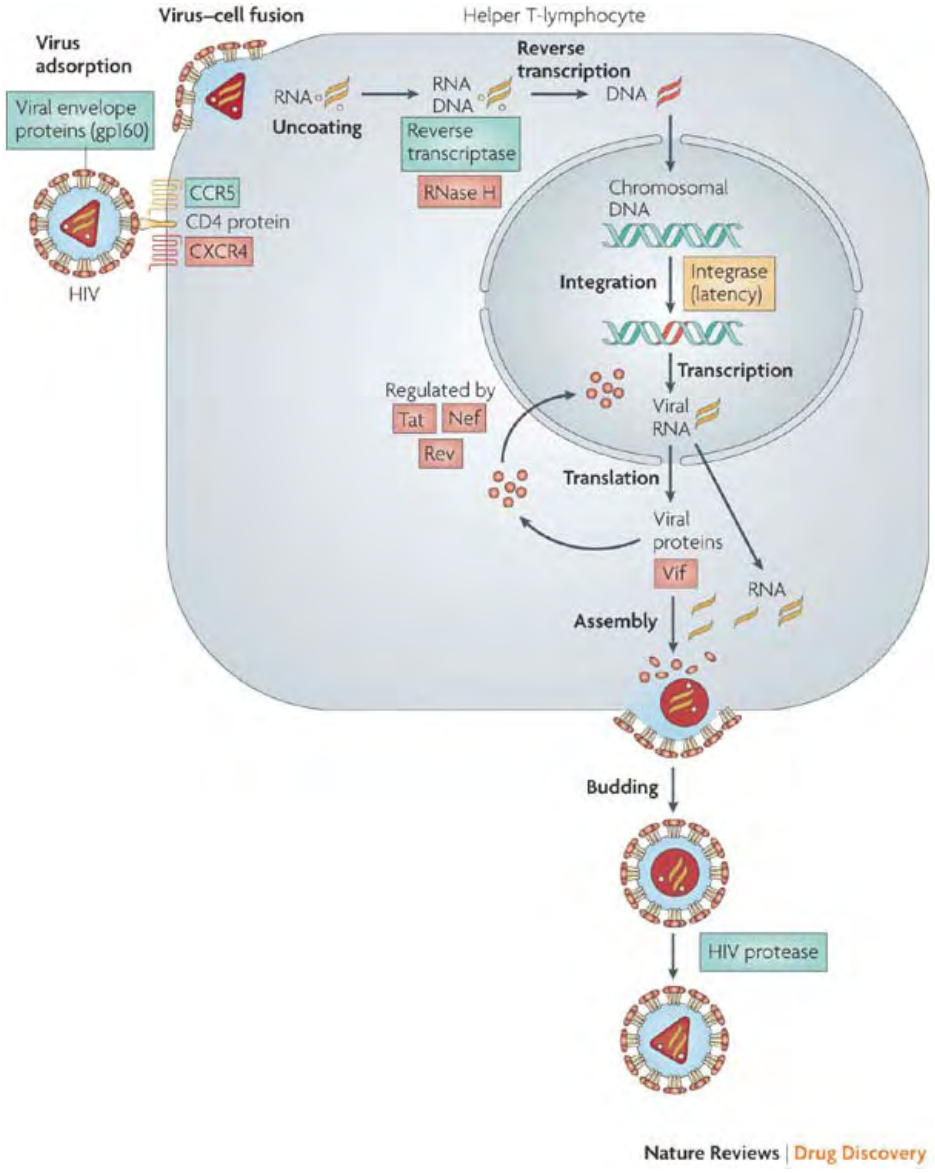
(4) Translation: The regulatory proteins Rev and Tat are first translated from sliced piece of the viral mRNA. They allow the full-length viral RNA to be transported into the cytoplasm. The other viral proteins are expressed from the unspliced RNA.

(5) Budding: Two copies of full-length viral RNA along with structural proteins and enzymes in the form of poly-proteins are transported to the host cell plasma membrane where TM anchors the SU protein. The forming virion begins to bud from the host cell to become an immature noninfectious virion.

(6) Maturation: PR cleaves the gal and gal-pol polyproteins into individual functional HIV proteins and enzymes. The structural components then assemble to produce a mature HIV virion which is able to attack another cell. This cleavage step can be

inhibited by protease inhibitors.

Figure 1.2: Replication cycle of HIV with current and possible targets for antiviral intervention. [1]



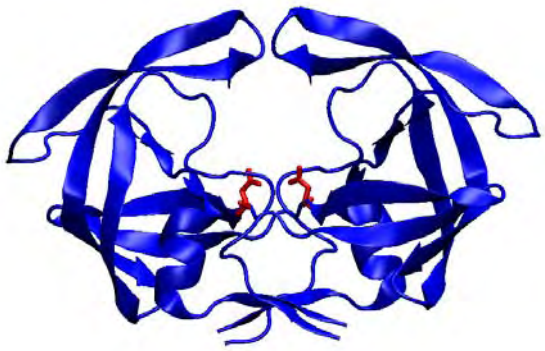
HIV protease and HIV protease inhibitor

HIV-1 protease is responsible for the post-translational processing of the viral Gag and Gag-Pol polyproteins to yield the structural proteins and enzymes of the virus. The enzyme is an aspartic protease composed of two noncovalently-associated, structurally identical monomers, 99 amino acids in length. The protease dimer is about 22 kDa. Its active site resembles that of other aspartic proteases and contains the conserved triad, Asp-Thr-Gly, at positions 25-27 (Figure 1.3). The hydrophobic substrate cleft recognizes and cleaves 10 different peptide sequences to produce the matrix (MA), capsid (CA), nucleocapsid (NC), and p6 proteins from the Gag polyprotein and the protease (PR), reverse transcriptase (RT), and integrase (IN) from the Gag-Pol polyprotein.

HIV-1 protease contains a flexible flap region that closes down on the active site upon substrate binding. Due to its importance for viral replication, HIV-1 protease has been a prime target of drugs to combat AIDS. There are ten FDA-approved protease inhibitors (PIs): amprenavir, atazanavir, fosamprenavir (slow release version of amprenavir), indinavir, lopinavir (marketed in combination with ritonavir), nelfinavir, ritonavir, saquinavir, tipranavir, and the recently approved darunavir. Protease inhibitors have been used in combination with drugs targeting other HIV enzymes. This combined treatment has been credited with an approximately 3-fold drop in the death rate from AIDS since its introduction in 1995. The development of

HIV-1 protease inhibitors is regarded as a major success of structure-based drug design. Indeed, PIs are considered the most potent medications currently available for the treatment of AIDS.

Figure 1.3: HIV-1 protease structure. The catalytic residues Asp25 and Asp25' are colored in red



Drug-resistant HIV protease mutations

Despite the remarkable success of PIs in treating AIDS, this success is diminished largely by the emergence of HIV mutants that resist current therapy. In fact, viral resistance is regarded as a critical factor in clinical failure of antiviral therapy. The infidelity of HIV-1 reverse transcriptase and the rapid turnover rate of the virus leads to a high rate mutation in the viral genome. As a result, some viral genes acquire mutations that give the virus a replicative advantage under the selective pressure of inhibitors. These mutations result in amino acid substitutions in the protease that reduce its binding affinity for the inhibitors, yet maintain sufficient substrate binding and catalysis for viral maturation and propagation, leading to drug resistance.

The study of drug-resistant protease mutations has been critical to HIV therapy involving application of PIs. Some protease mutations directly change interactions between the protease and its inhibitors. For example, mutations in the protease substrate cleft, e.g., at D30, V82 and I84, tend to cause resistance by reducing the binding affinity between the mutant protease and inhibitors. Mutations at M46, I47, G48, I50 and I54 in protease affect its dynamic properties and influence its ability to bind with substrate and inhibitors. Mutations elsewhere in the enzyme either compensate for its decreased kinetics due to active-site mutations or also cause resistance by altering inhibitor-binding kinetics, or by reshaping the active site

through long-range structural perturbations. Most substrate-cleft mutations reduce the *in vitro* susceptibility to one or more PIs by 2- to 5-fold. However, additional mutations are usually required in the enzyme flap and in other parts of the molecule for the emergence of *in vivo* resistance.

V82A/T mutations occur predominantly in HIV-1 isolates from patients receiving treatment with indinavir or ritonavir.[2,3] V82A also occurs in isolates from patients receiving prolonged saquinavir therapy following development of the mutation G48V. By themselves, mutations at codon 82 confer reduced susceptibility *in vitro* to indinavir, ritonavir, and lopinavir, but not to nelfinavir, saquinavir, or amprenavir. However, when present with other PI-induced mutations, the V82A/T mutations contribute phenotypic and clinical resistance to each of the PIs. V82A is the most common mutation at position 82 (V82A/V82T is 7/1).[4] The I84V mutation, which has been reported in patients receiving indinavir, ritonavir, saquinavir, and amprenavir as their sole PI,[2,3] causes phenotypic and/or clinical resistance to each PI. The G48V mutation occurs primarily in patients receiving saquinavir and rarely in patients receiving indinavir. [5] This mutation causes 10-fold resistance to saquinavir and about 3-fold resistance to indinavir, ritonavir, and nelfinavir. G48V has been reported to cause low-level biochemical resistance to amprenavir when present in site-directed mutants, but to interfere with amprenavir resistance when present with more typical amprenavir-resistant mutations such as M46I, I47V, and I50V.[5] Its effect on lopinavir and atazanavir is not known. G48V usually occurs with mutations at

positions 54 and 82.[6] The protease flaps (residues 33-62, which consist of a short anti-parallel β-sheet with a turn, extend over the substrate-binding cleft and must be flexible to allow entry and exit of the polypeptide substrates and products.[7] The flap tips (residues 46-54) are particularly mobile and are the site of many drug-resistant mutations. In addition to mutations at positions 48 and 50, which extend into the substrate cleft, mutations at positions 46, 47, 53, and 54 make important contributions to drug resistance. Mutations at position 54 (generally I54V, less commonly I54T/L/M/S) contribute resistance to each of the approved PIs.[2] I54L and I54M are particularly common in persons receiving amprenavir and have a greater effect on amprenavir than does I54V.[8]

Thermodynamics of protease-inhibitor binding

To elucidate the mechanisms of these drug-resistant mutations of HIV protease, over 200 crystal structures of HIV-1 protease variants have been solved. Changes in affinity due to drug-resistant mutations and thus the thermodynamics of binding can be measured by ITC (isothermal titration calorimetry). Comparison of the structures of wild-type and drug-resistant variant protease in complex with inhibitors partially rationalizes how specific protease mutations decrease protease-inhibitor binding affinity. However, it is still a challenge to illuminate critical changes in binding affinity in a quantitative way from the structural data. Free-energy simulations, in principle, can aid in elucidating these components of the binding affinities to particular atomic interactions. Among all the methodologies of free energy prediction, free-energy perturbation[4,9-11] and MM-PB/GBSA methods have been broadly applied to a wide range of biomolecular systems to study protein-protein, protein-DNA/RNA, and protein-ligand interactions.

Thermodynamic integration method

The thermodynamic integration (TI) method is a free energy perturbation method that is mostly employed to study relative binding free energy changes. If one takes the derivative of the equation $G_{NVT} = -RT \ln Q$ with respect to λ, where the Hamiltonian H has been replaced by the λ-dependent potential function V (λ, x), one

obtains:

$$\frac{dG}{d\lambda} = -\left[\frac{RT}{Q}\right]\left[\frac{dQ}{d\lambda}\right]$$

By reference to the partition function defined as

$Q_{NVT} = C\int\int dxdp \exp(-H(x,p)/RT)$, differentiation and substitution leads to

$$\frac{dG}{d\lambda} = \frac{\int\left(\frac{dV(\lambda,x)}{d\lambda}\right)e^{-V(\lambda,x)/RT}dx}{\int e^{-V(\lambda,x)/RT}dx} = \left\langle\frac{dV(\lambda,x)}{d\lambda}\right\rangle_{\lambda}.$$

Integration gives:

$$\Delta G = \int_0^1 \left\langle\frac{dV(\lambda,x)}{d\lambda}\right\rangle_{\lambda} d\lambda$$

This is the master equation for the free energy method known as TI. In the master equation equation, ΔG refers to the change in free energy, the integrand is evaluated at a series of discrete points and the integral is approximated from these points using a numerical integration method. For example, the simple trapezoidal integration method is frequently used. When this method is applied to the previous equation, one gets

$$\Delta G \approx \sum_{1}^{N-1}\left(\left\langle \frac{\partial V(x,\lambda)}{\partial \lambda}\right\rangle_{\lambda(i)} + \left\langle \frac{\partial V(x,\lambda)}{\partial \lambda}\right\rangle_{\lambda(i+1)}\right)\times\left(\frac{\lambda(i+1)-\lambda(i)}{2}\right)$$

For TI, the main concern with respect to the λ pathway is that enough λ points are selected so that the numerical integration over these points is reasonably accurate. To obtain ΔG versus λ curves that vary smoothly and slowly, a modest number of points will usually suffice. Fortunately, such a curve is characteristic of most free energy calculations.

MM-PB/GBSA method

The MM-PB/GBSA method combines molecular mechanics and the continuum solvation model to calculate the binding free energy of biomolecular interactions.[12] MM refers to molecular mechanics, PB refers to the Poisson-Boltzmann equations, GB refers to the general Born model which approximates the PB model, and SA refers to the solvent-accessible surface area.

In applying the MM-PB/GBSA model to calculating the binding affinity between the protease and inhibitor, the binding free energy change $\Delta G_{binding}$ in the process:

$$\text{Protease} + \text{Inhibitor} \xrightarrow{\triangle G_{binding}} \text{Complex}$$

is represented by equation 1:

$$\Delta G_{binding} = \Delta G_{MM} - T\Delta S + \Delta G_{PB/GB} + \Delta G_{NP} \qquad (1)$$

ΔG_{MM} is the molecular mechanics energy change, here described and calculated by the AMBER force field functions and parameters as in equation 2:

$$\Delta G_{MM} = \Delta G_{bond} + \Delta G_{angle} + \Delta G_{dihedral} + \Delta G_{vdw} + \Delta G_{ele} \qquad (2)$$

The bond stretching energy change (ΔG_{bond}), angle vibrational energy change (ΔG_{angle}), and dihedral angle torsion energy change ($\Delta G_{dihedral}$) are the bonded energy change terms. The van der Waals interaction energy change (ΔG_{vdW}) and electrostatic interaction energy change (ΔG_{ele}) are the non-bonded energy change terms, which are of specific interest in my study of protease-inhibitor binding.

In equation 1, the entropy was calculated by normal-mode analysis using the AMBER 8 NMODE module.

$$\Delta S = \Delta S_{translational} + \Delta S_{rotational} + \Delta S_{vibrational}$$

$$s = k \ln Q - \left(\frac{\partial \ln Q}{\partial k_B T} \right)_v$$

$$Q_{trans} = \frac{V}{h(k_B T / 2\pi m)^{3/2}}$$

$$Q_{rotational} = \left(\frac{\pi^{1/2}}{\sigma} \right) \left(\frac{1}{hck_B T} \right)^{3/2} \left(\frac{1}{I_A I_B I_C} \right)^{1/2}$$

The polar component of the solvation free energy change, represented by $\Delta G_{PB/GB}$ is calculated as:

$$\Delta G_{PB/GB} = \Delta G_{PB/GB}^{Protease} + \Delta G_{PB/GB}^{Inhibitor} - \Delta G_{PB/GB}^{Complex}$$

For each species (protease, inhibitor, and complex), the polar solvation changes are calculated by the PB or GB method. In the PB method, the electrostatic

potential produced by a molecular charge distribution is determined by solving the Poisson-Boltzmann equation.

The Poisson function

$$-\nabla \cdot \nabla \phi(\mathrm{r}) = \frac{\rho(\mathrm{r})}{\varepsilon_0}$$

$$\mathrm{E}(\mathrm{r}) = -\nabla \phi(\mathrm{r})$$

defines the relation between the electrostatic potential φ(r), electrostatic field E(r), and the charge density ρ(r). The charge density is not isotropic because interactions between ions create the atomic atmosphere. To account for this effect, the Debye-Huckel theory employs the Boltzmann factor of dissolved ions in the local electrostatic potential ($e^{-\beta\phi(\mathrm{r})q_i}$) to estimate changes in the local ion concentration C(r) relative to their bulk concentration C_i^0:

$$\mathrm{C}(\mathrm{r}) = \mathrm{C}_i^0 e^{-\beta\phi(\mathrm{r})q_i}$$

In a system where the electrolyte has N types of ions with charge q_i and concentration $C_{i.,}$ the Poisson-Boltzmann equation is:

$$-\nabla \cdot \left[\varepsilon(\mathrm{r})\nabla\phi(\mathrm{r})\right] = \rho(\mathrm{r}) + \sum_{i=1}^{N} q_i C_i^0 e^{-\beta\phi(\mathrm{r})q_i}$$

This analytical approach was first developed by Kirkwood[13] and later revised to the Tanford-Kirkwood model.[14] In this approach, biomolecules are considered as symmetric geometric bodies to allow the analytical solution of the PB equation. With increases in asymmetry, it becomes more difficult to solve the equation analytically. One must then resort to numerical methods.

One numerical approach that has achieved great success in the study of electrostatic interactions among biomolecules is the Finite Difference Solution to the PB equation (FDPB) method.[15-18] A typical FDPB process[19] includes the following steps: assigning atomic changes to the finite difference grid points, determining electrostatic potentials on the boundary surfaces of the finite difference grid, and introducing a dielectric model to define the boundary between a high dielectric region (usually water) and low dielectric region (biomolecule interior). Then the partial differential equation can be converted to a linear system and the numerical solution of the PB equation is calculated to convergence.

In the GB method, the generalized Born equation is used to calculate the polar solvation energy change. The electrostatic free energy of a system comprising particles with radii L_i and charge q_i in a solution with dielectric constant ε can be represented as:

$$G_{ele} = \sum_i^N \sum_j^N \frac{q_i q_j}{\varepsilon r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_i^N \frac{q_i^2}{L_i}$$

where r_{ij} is the distance between atom i and atom j. The electrostatic free energy difference between the system *in vacuo* and in solution can be calculated by the generalized Born equation:

$$\Delta G_{ele} = -\left(1 - \frac{1}{\varepsilon}\right)\sum_i^N \sum_j^N \frac{q_i q_j}{r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_i^N \frac{q_i^2}{L_i}$$

The GB equation can be expressed in one term[20]

$$\Delta G_{ele} \approx \Delta G_{GB} = -\frac{1}{2}\left(1-\frac{1}{\varepsilon}\right)\sum_{i}^{N}\sum_{j}^{N}\frac{q_i q_j}{f\left(r_{ij}, R_i, R_j\right)}$$

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

where R_i is the effective Born radius of atom i, which reflects the degree of its burial inside the low dielectric region defined by the solvent boundary. In the OBC (Onuvriev, Bashfod and Case) GB method[21], R_i was modified to be suitable for macromolecules.

The nonpolar contribution of the solvation free energy change is represented by ΔG_{NP}, which is linearly dependent on the solvent-accessible surface area A: $\Delta G_{NP}=\gamma A+b$.[22] This dependence is due to ΔG_{NP} being mainly determined by interactions between the solute and solvent molecules in the first solvation shell, where the number of solvent molecules is approximately proportional to the accessible surface area. The surface area can be calculated by Sanner's MSMS software[23] using a water-sized probe. The MM-PB/GBSA method has been widely used to calculate absolute binding free energy changes associated with biomolecular recognition and relative free energies of different conformations.[24-28]

Protease dynamics

Protein dynamics, structure, and function are highly correlated. In the case of HIV protease, the dynamics of conformational changes are extremely critical for enzyme function. The first step in the HIV-1 protease substrate-cleavage reaction is opening of the protease flaps to expose the active site pocket to the substrate. This conformational change is highly related to the flexibility of the flap region and hydrophobic core movements of the protease.[29] The motions of the flap tips in unbound HIV-1 protease have been investigated using nuclear magnetic resonance (NMR)[30,31] and molecular dynamics (MD) simulations.[29] These studies indicate that the flaps are highly mobile and can adopt a large number of conformations on the nanosecond to microsecond timescale, suggesting that mutations in the flaps can be very important for the enzyme's kinetics and inhibitor binding.

NMR spectroscopy can determine the time scales of atomic fluctuations and relative atomic positions of proteins, but the results are based on an ensemble of molecules in the solution. Thus, NMR cannot provide enough detail about the dynamic properties in a single molecule. One of the principal tools in the theoretical study of biological molecules is the method of MD simulation. This computational method calculates the time-dependent behavior of a molecular system. It is a deterministic method, which means that the state of the system at any future time can be predicted from its current state. MD simulations provide detailed information on

the fluctuations and conformational changes of proteins and nucleic acids at the atomic level. In MD simulation, Newton's equation of motion

$$\vec{F} = m\vec{a} = \frac{\partial V}{\partial \vec{r}}$$

is solved for a given potential $V(\vec{r})$, to represent a protein and its environment, where the potential of this system is described by a force field function:

$$V(\vec{r}) = \sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \sum_{dihederals} K_\phi(1+\cos(n\phi+\varphi))$$
$$+\sum_{VDW}\left[\frac{A}{r^{12}}-\frac{B}{r^6}\right]+\sum_{electrostatic}\frac{q_1q_2}{\varepsilon_0 r}$$

Sets of atomic positions are derived in sequence by applying Newton's equation. The displacement of each atom ($\triangle\vec{r}_i(t)$) is integrated by breaking the calculation into a series of very short time steps, typically between 1 and 10 femtoseconds. At each step, the forces on the atoms are computed and combined with their current positions and velocities to generate new positions and velocities a short time ahead. The force acting on each atom is assumed constant during the time interval. The atoms are then moved to their new positions, an updated set of forces is computed, and the process is repeated. In this way a molecular dynamics simulation generates a trajectory that describes how the dynamic variables change with time. MD simulations are typically run for several to a hundred nanoseconds. The MD method is

now routinely used to investigate the structure, dynamics, and thermodynamics of biological molecules and their complexes.

Summary

The scope of this thesis is to use computational methods to investigate the mechanism of HIV-1 protease drug resistance on an atomic level. Free energy simulation methods were used to calculate the free energy change of protease-inhibitor binding and the results were compared with ITC experimental data. Free energy decomposition analysis was performed on the binding of drug resistant variants and WT with inhibitors to find out the major reason of binding affinity loss. Molecular dynamic simulation of proteases was performed to compare the difference in dynamic properties between WT and drug resistant variants. Improved understanding of the thermodynamic and dynamic properties of the protease variants will contribute to rational design of new protease inhibitors which bind more tightly to the protease and hinder the development of drug-resistant mutations.

Chapter II

Decomposing the energetic impact of drug resistant mutations in HIV-1 protease on binding DRV

Abstract

Darunavir (DRV) is a high affinity (K_d=4.5X10^{-12}M, ΔG=-15.2 kcal/mol) HIV-1 protease inhibitor. Two drug-resistant protease variants, FLAP+ (L10I, G48V, I54V, V82A) and ACT (V82T, I84V), decrease the binding affinity of DRV by 1.0 kcal/mol and 1.5 kcal/mol respectively. In this study, the absolute and relative binding free energies of DRV with wild-type protease, FLAP+ and ACT were calculated with MM-PB/GBSA and thermodynamic integration methods, respectively. Free energy decomposition elucidates that the mutations confer resistance by distorting the active site of HIV-1 protease so that the residues that lose binding free energy are not limited to the sites of mutation. Specifically, the bis-tetrahydrofuranylurethane moiety of DRV maintains interactions with the FLAP+ and ACT variants, whereas the 4–aminophenyl group is more flexible in the FLAP+ and ACT complexes than in the wild-type protease which could account for the majority of the loss in binding free energy. This suggests that replacement of the 4–aminophenyl group may generate new inhibitors less susceptible to the drug resistant mutations.

Introduction

The human immunodeficiency virus type 1 (HIV-1) protease is a homodimeric aspartyl enzyme with 99 residues in each chain. The two HIV-1 monomers are bound by non-bonded interactions, with the active site at the interface between the two monomers.[32] The protease processes the viral Gag-Pol polyprotein, yielding the structural proteins and enzymes critical for the maturation of infectious viral particles.[33] Thus, HIV-1 protease has been a major target for structure-based drug design. Nine protease inhibitors have been approved by the Food and Drug Administration (FDA) for HIV therapy, effectively decreasing the mortality rate of HIV/AIDS patients. These FDA-approved HIV-1 protease inhibitors, were developed at least in part, using structure based drug design, are competitive inhibitors.[33] Unfortunately, exposure to protease inhibitors leads to the virus developing drug resistance mutations in the protease due to the high replication rate of HIV-1 and to lack of a proofreading mechanism in its reverse transcriptase. These drug-resistant protease variants lose their high binding affinity to the inhibitors, while maintaining enough enzyme activity for the virus to propagate.[34]

To understand the basis for these changes in drug-resistant proteases, over 200 crystal structures of HIV-1 protease variants have been solved in the past 25 years. Changes in affinity due to drug resistant mutations and thus the thermodynamics of binding can be measured by isothermal titration calorimetry.[35,36] Comparison between

the structures of wild-type and drug-resistant variant proteases in complex with inhibitors partially elucidates how specific protease mutations decrease protease-inhibitor binding affinity.[37,38] However, elucidating the critical components of the binding affinity quantitatively from the structural data still remains a challenge. Free-energy simulations[39-44], in principle, can aid in elucidating these components of the binding affinities to particular atomic interactions.

Among these computational methods, free-energy perturbation (FEP) and thermodynamic integration (TI) methods, which are derived from statistical mechanics[11,42,45-48], are mostly used with the thermodynamic cycle to calculate relative binding free energy changes in similar systems. The Molecular-mechanics Poisson-Boltzmann surface area (MM-PBSA) method combines molecular mechanics and the continuum solvation model.[12,43,49-51] Solvation properties can be described by the Poisson-Boltzmann (PB) or generalized Born (GB) equation. This method is reliable and applicable to calculate absolute binding free energy change associated with biomolecular recognition. To achieve a better match with experimental data, MM-PB/GBSA method is usually supplemented by entropy estimation. Free-energy calculation methods provide a way to estimate the binding free energy of inhibitors with different protease variants, allowing computational screening of lead compounds in rational drug design. Furthermore, the calculation results can be further analyzed, e.g., for free energy decomposition, to provide information about affinity changes due

to specific kinds of interactions on an atomic level, which can not be determined by experimental methods.[12,43,52,53]

The HIV-1 protease inhibitor, Darunavir (DRV, formerly known as TMC114) (Figure 2.1A) has recently been approved by the FDA.[54] This second-generation protease inhibitor, which was developed after extensive effort in rational drug design[55], has the lowest dissociation constant of all reported HIV protease inhibitors (K_d =4.5X10^{-12}M).[35] Nonetheless, DRV still loses affinity to drug resistant variants of HIV-1 protease[35]. In this study, the binding of DRV was investigated with wild-type HIV-1 protease and two drug-resistant variants: FLAP+ (Figure 2.1B) with L10I, G48V, I54V, V82A which are a combination of flap and active site mutations, and ACT (Figure 2.1C) with V82T, I84V which are active site mutations. Each of these three systems was analyzed in three parallel 20 ns molecular dynamics (MD) simulations using initial coordinates from their crystal structures. In these MD simulation trajectories, the MM/PBSA and MM/GBSA methods were applied to calculate changes in binding free energy, which were compared with ITC results. The classical TI method was also used to calculate and compare differences in binding free energy of the DRV-ACT and DRV-WT complexes. The accuracy, convergence and reproducibility of the calculated results have been compared and discussed. The MM-PB/GBSA and TI calculation results are in good agreement with the experimental data. Free energy component analysis was performed to elucidate the mechanism for resistance of FLAP+ and ACT to DRV. The free energy

decomposition study results show that the *bis*-THF group of DRV maintains its favorable van der Waals (vdW) contact with the protease even in the drug resistant variants. Understanding how the protease mutates to decrease its binding affinity with a very high affinity inhibitor will contribute to developing better strategies to design protease inhibitors.

Abbreviations:

Bis-THF: bis-tetrahydrofuranyl

ITC: isothermal titration calorimetry

TI: thermodynamic integration

MM/PBSA: molecular mechanics/Poisson-Boltzmann surface area

GB: generalized Born

DRV: Darunavir

ACT: HIV-1 protease variant V82T, I84V

FLAP+: HIV-1 protease variant L10I, G48V, I54V, V82A

vdW: Van der Waals

MD: Molecular Dynamics

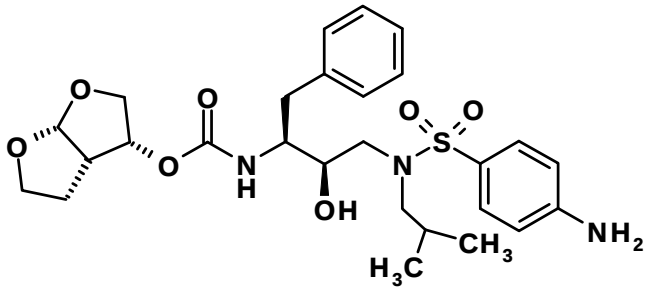
ns: nanosecond

ps: picosecond

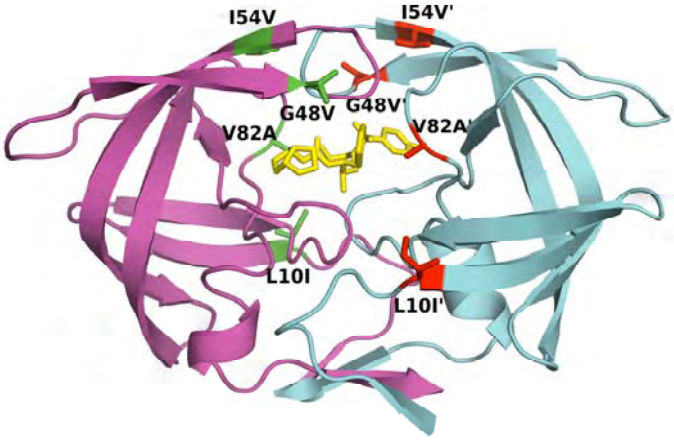
fs: femtosecond

Figure 2.1: (A) Chemical structure of DRV. (B) Structure of protease variant FLAP+-DRV complex, DRV is colored yellow. The side chains of the mutated residues Ile10, Val48, Val54, and Ala82 are displayed and colored red or green. (C) Structure of protease variant ACT-DRV complex. DRV is colored yellow. The side chains of the mutated residues Thr82 and Val84 are displayed and colored red or green.

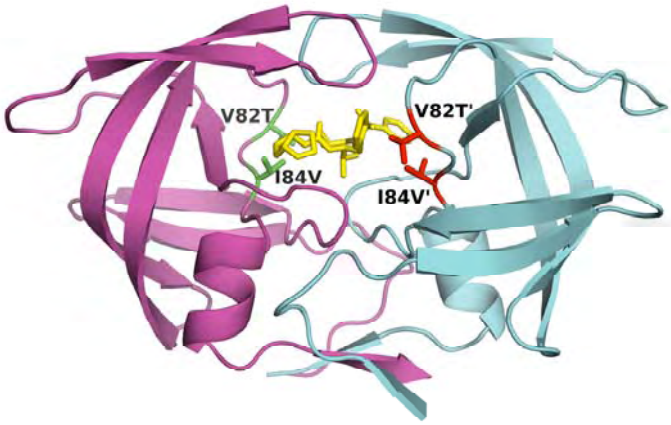
(A)


O
O
O
O
N
H
OH
N
O
S
O
NH2
CH3
H3C

(B)


I54V
I54V'
G48V
G48V'
V82A
V82A'
L10I
L10I'

(C)


V82T
V82T'
I84V
I84V'

Methods

MD simulation with the program Sander in the AMBER 8 package

The initial coordinates of the DRV-WT protease complex are taken from its crystal structure 1T3R.[35] Similarly, the DRV-FLAP+ protease complex coordinates are from its crystal structure 3EKT and the DRV-ACT protease complex coordinates are from its crystal structure 1T7I.[35]

Molecular dynamics simulations were performed using the program Sander in the MD simulation package AMBER 8.[56] For the standard protease residues, the atomic partial charges, van der Waals parameters, equilibrium bond lengths, bond angles, dihedral angles, and their relative force constants were taken from the AMBER database (ff03).[57] For DRV parameters, the van der Waals parameters, equilibrium bond lengths, bond angles, dihedral angles, and force constants were taken from the General AMBER Force Field database.[58] The partial charges of inhibitor atoms were obtained as follows. First, the coordinates of the DRV atoms were taken from the 1T3R crystal structure, and the missing hydrogen atoms added by the program Quanta. Second, the geometry of the resulting structure was optimized with the (HF)/6-31G* basis set by Gaussian 03 package.[59] Finally, the resulting electrostatic potential was used in the RESP[60] module of the AMBER 8 package to derive the atomic partial charges of the inhibitor.

The explicit solvent model was applied to all systems. Each structure was solvated with the TIP3P water cubic box to allow for at least 8 Å of solvent on each face of the protease. The vdW dimensions for the protease are 44 by 35 by 59 Å. The dimensions of the final periodic box are 63 by 55 by 78 Å. The simulation system had approximately 7000 water molecules, and six Cl- counterions were added to balance the charge of the system.

A three-step energy minimization process with the steepest descent method was used to allow the system to reach an energetically favorable conformation. In the first energy minimization step, all the heavy atoms of the protease were restrained with a harmonic force constant of 10 kcal $mol^{-1}Å^{-2}$. In the second step, only the backbone nitrogen, oxygen, and carbon atoms were restrained. In the third step, all atoms were allowed to move. Each of the three steps had 2000 cycles. The temperature of the energy-minimized system was then gradually raised from 50ºK to 300ºK in the NVT ensemble. Initial velocities were assigned according to the Maxwellian distribution, and random seeds were assigned with three different values to generate nine simulations, three parallel simulations for each of the WT-DRV, FLAP+-DRV, and ACT-DRV systems. In the thermalization process, heavy atoms were restrained with a harmonic force constant of 10 kcal mol^{-1} $Å^{-2}$. The whole process was 50 picoseconds (50,000 steps, each of which was 1 femtosecond). A 50 picoseconds equilibration was then performed in the NPT ensemble without restraining heavy atoms. In the subsequent sampling MD simulations, each step was 2fs, and the total simulation was

20 ns. For the thermalization, equilibration, and sampling simulations, the SHAKE algorithm[61] was applied to constrain all hydrogen atoms.

MM-PB/GBSA method

For the protease–ligand system, the binding free energy change is represented by:

$$\text{Protease} + \text{Inhibitor} \xrightarrow{\triangle G_{binding}} \text{Complex}$$

and

$$\Delta G_{binding} = \Delta G_{MM} - T\Delta S + \Delta G_{PB/GB} + \Delta G_{NP}$$

where

$$\Delta G_{MM} = \Delta G_{bond} + \Delta G_{angle} + \Delta G_{dihedral} + \Delta G_{vdw} + \Delta G_{ele}$$

$$\Delta S = \Delta S_{translational} + \Delta S_{rotational} + \Delta S_{vibrational}$$

The molecular mechanical energy ΔG_{MM} is the estimated free energy change associated with the binding process in gas phase. ΔG_{MM} is calculated by standard force field functions and parameters. Depending on the type of interaction, ΔG_{MM} has two kinds of energetic terms: bonded and non-bonded. The bonded term includes terms representing bond stretching energy (ΔG_{bond}), angle vibrational energy (ΔG_{angle}), and dihedral angle torsion energy ($\Delta G_{dihedral}$). The non-bonded term includes terms

representing the van der Waals interaction energy (ΔG_{vdW}) and electrostatic interaction energy (ΔG_{ele})

The polar component of the solvation free energy, represented by $\Delta G_{PB/GB}$, can be calculated either by solving the Poisson-Boltzmann equation (PB method) or the generalized Born equation (GB method). The nonpolar component of the solvation free energy is represented by ΔG_{NP}. The sum of $\Delta G_{PB/GB}$ and ΔG_{NP} estimates the free energy change associated with molecules entering solvation from the gas phase. The GB calculation was done using the model developed by Onufriev et al.[62,63] The PB calculation was done with the AMBER 8 numerical PB solver.[19] The solute dielectric constant was 1.0, and solvent dielectric constant was 80.0. ΔG_{NP} was calculated by the LCPO (linear combinations of pairwise overlaps) method, which is linearly dependent on the solvent access surface area: ΔG_{NP}=0.0072 X SASA.[22] The entropy was calculated by normal-mode analysis using the AMBER 8 NMODE module. For every 20ps of the 20 ns MD simulation trajectory, a snapshot of the protease and inhibitor was taken removing the solvent and counter ions. Total number of the atoms for each of the three systems: DRV-WT, DRV-FLAP+, DRV-ACT are 3203, 3209, 3203 respectively. Altogether, 1000 frames were used for the MM-PB/GBSA calculations.

Thermodynamic integration method

When studying drug-resistant protease mutants, the binding free energy relative to wild-type protease is even more important than the absolute binding free energy.

Thermodynamic integration method[4,64] was applied to the protease-inhibitor system to compute the free-energy difference between different states of the system. From statistical mechanics, the Gibbs free energy (G) can be calculated from the partition function Q as follows:

$$G = -RT \ln Q \quad (1)$$

The partition function can be expressed as the integral of the system's Hamiltonian function H(r,p). After introduced a coupling parameter, lambda, into the Hamiltonian, Q can be expressed as:

$$Q = \int\int dr dp \exp(-H(r, p, \lambda) / RT) \quad (2)$$

From equation (1) and (2), the derivative of G with respect to lambda is:

$$\frac{dG}{d\lambda} = \frac{\int\int \frac{dH(r, p, \lambda)}{d\lambda} e^{-H(r,p,\lambda)/RT} dr dp}{\int\int e^{-H(r,p,\lambda)/RT} dr dp} = \left\langle \frac{dH(r, p, \lambda)}{d\lambda} \right\rangle_{\lambda} \quad (3)$$

and

$$\Delta G = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda$$

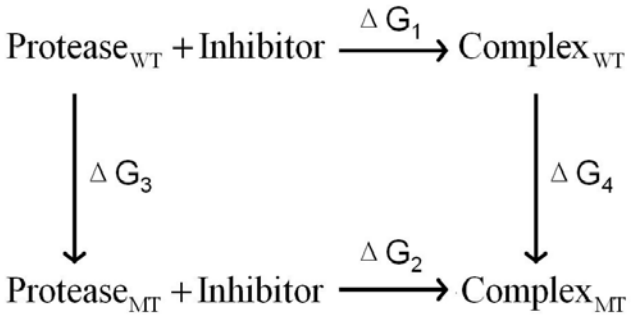
(4)

Equation 4 is the master equation of the thermodynamic integration method. When applying this equation to the protein–ligand system, the kinetic component of the Hamiltonian can be neglected. Thus, the lambda-coupling force field function V (λ, r) is used to replace the Hamiltonian. The lambda is chosen such that when it equals zero, the force field function V(0) and its relative parameters are correlated with the wild-type protease. When λ=1, V(1) and its parameters are correlated with the mutant protease. The numerical estimation of equation 4 is:

$$\Delta G \approx \sum_{i=1}^{n} w_i \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda i} \qquad (5)$$

The lambda values and their relative weights were assigned from the Gaussian quadratic formula.

Directly calculating thermodynamic integration from the unbound to the bound state is not feasible. The thermodynamic cycle below is used since G represents a state function and is independent of the path.



As shown above, instead of calculating the free energy changes ΔG_1 and ΔG_2 associated with the chemical reaction path, the ΔG_3 and ΔG_4 through the "alchemical" path [65,66] are calculated.

Thus, the drug-resistant mutant's loss of binding free energy compared to the wild-type protease is represented by:

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$$

(6)

The thermodynamic integrations were carried out in the Sander module of the AMBER 8 package.[67,68] The wild-type and mutant proteases have different numbers of side chain atoms on the mutated residues. To keep the same number of atoms in the initial and final states, I perturbed the extra atoms to dummy atoms, which had no nonbonding interactions with the rest of the system. When carrying out thermodynamic integration calculations, the Sander module in AMBER 8 only allowed dummy atoms to appear in the perturbed Hamiltonian.[67,68] It was not feasible for Sander to perform the thermodynamic integration calculations for the FLAP+

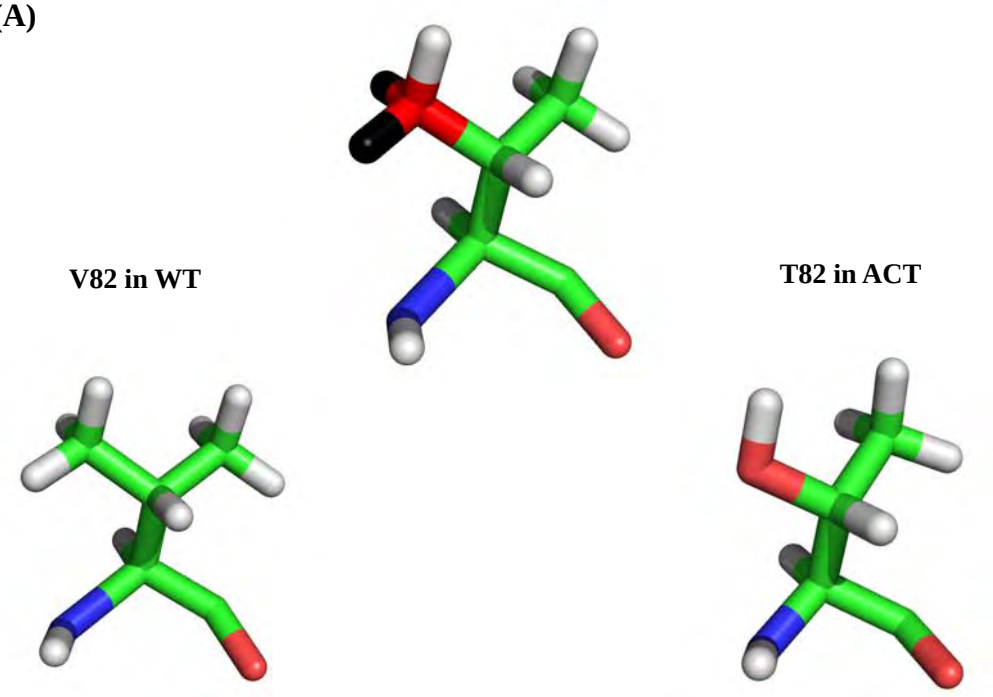
mutant (L10I, G48V, I54V and V82A), since Leu10 in the WT and Ile10 in the FLAP+ mutant have the same number of side-chain atoms but different tree structures, Gly48 has fewer atoms than Val48, and Ile54 and Val82 in the WT have more atoms than Val54 and Ala82 in FLAP+. For the ACT mutant, both mutated residues (V82T, I84V) have fewer atoms than the wild type. Thus the perturbation is done from WT to ACT. (Figure 2.2)

The DRV-WT crystal structure 1T3R was used to generate the perturbed topology file and the coordinates file for the calculation of the ΔG_4. For the calculation of ΔG_3, two sets of topology and coordinates files were created. One is from the unbound wildtype protease crystal structure 1HHP and one is the protease atoms coordinates extracted from the WT-DRV complex crystal structure 1T3R. The three steps energy minimization was performed as described before. The structure was then thermalized and pre-equilibrated with a harmonic restrained force and the lambda value was 0.5. During the thermalization, different random seed values were assigned to create parallel calculations as control. The pre-equilibrated structure was then started to sampling at 12 lambda values. The time steps were 1fs and the time for the calculation at each lambda values was 2 ns. Thus the total sampling time for each alchemical free energy change calculation is 24 ns. The expected error in the free energy calculations is the root mean square deviation in the energies of the sample in production period divided by the square root of the number of independent samples in the production period.[68,69]

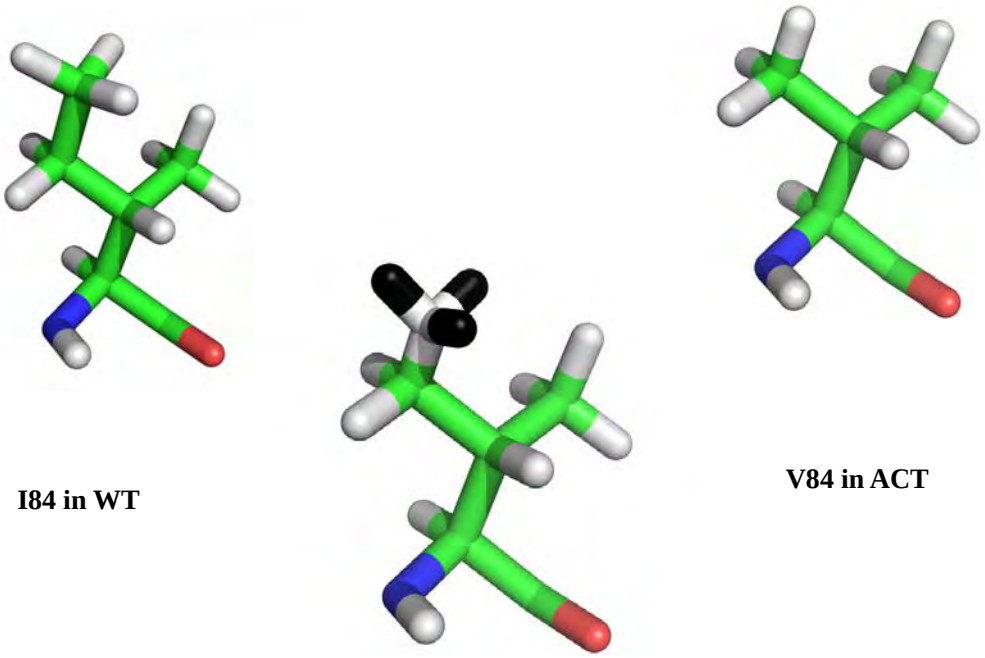
$$\text{expected-error} = \frac{\text{sample-rms}}{\sqrt{\text{number of independent samples}}} \quad \textbf{(7)}$$

Figure 2.2: Perturbation of Val82 to Thr and Ile84 to Val. Hydrogen atoms are colored white, oxygen atoms are colored red, nitrogen atoms are colored blue, carbon atoms are colored green and dummy atoms are colored black. Left: residue in the wild-type protease as the initial state; Middle: the hybrid residue in the calculation process; Right: the mutated residue as end state. (A) The perturbation of Val82 to Thr82. (B) The perturbation of Ile84 to Val84.

(A)



(B)



Results and Discussion

Comparison between predicted binding affinity and ITC data

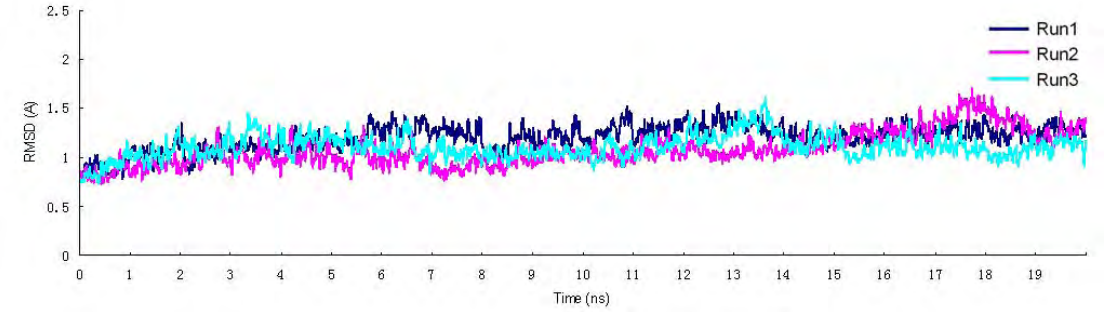
Calculations of absolute binding free energy by MM/PBSA and MM/GBSA methods

To evaluate the reproducibility and convergence of our free-energy calculation results, the same MM/GBSA protocol was applied to three independent 20 ns MD simulation runs of each of the WT-DRV, FLAP+-DRV and ACT-DRV systems starting from each of their corresponding crystal structure (see Methods). To study the structural stability of the systems, the root mean square displacement (RMSD) of the Cα atoms of the simulated proteins were plotted over time with respect to their corresponding crystal structures. (Figure 2.3) For all the DRV-protease systems after 2 ns of MD simulations the RMSD values are approximately 1.5 Å. As the calculations all require extensive equilibration, the averages of potential production periods were evaluated. After 10 ns simulations the calculated binding free energy for DRV-WT stabilizes (Figure 2.4) for all three parallel simulations. Each of the triplicates of DRV-FLAP+ and DRV-ACT stabilize within 6 and 9 ns respectively (Figure 2.4). Thus the first 10 ns is used as the equilibration period, as the free energy

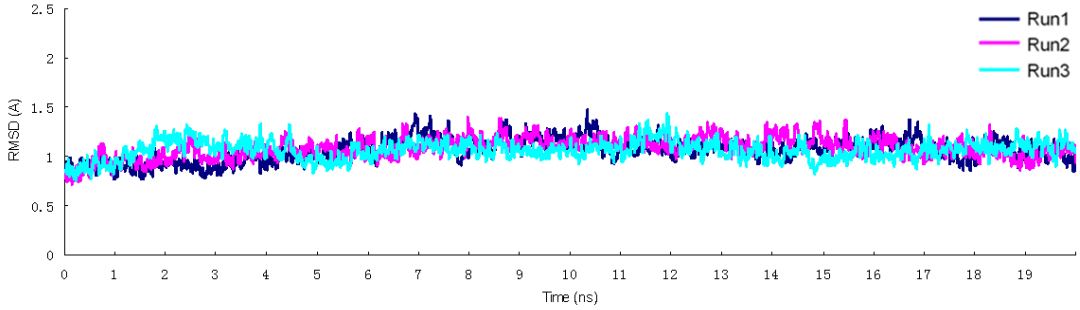
generally does not converge between the runs (Table 2.1), while the second 10 ns is used as the production period, since generally the runs are converged.

Figure 2.3: RMSD of Cα atoms of protease with respect to their corresponding crystal structures during the 20 ns MD simulations. (A) RMSD of Cα atoms of DRV-WT simulations. (B) RMSD of Cα atoms of DRV-Flap+ simulations. (C) RMSD of Cα atoms of DRV-ACT simulations.

(A)


RMSD (A)
Time (ns)
Run1
Run2
Run3

(B)


RMSD (A)
Time (ns)
Run1
Run2
Run3

(C)

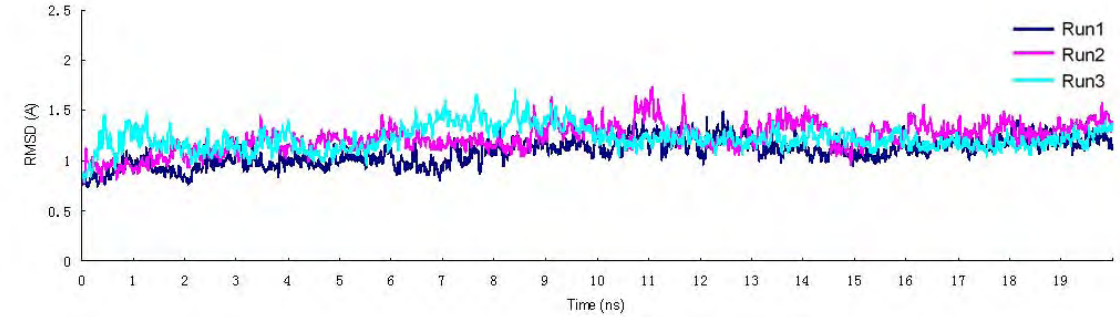
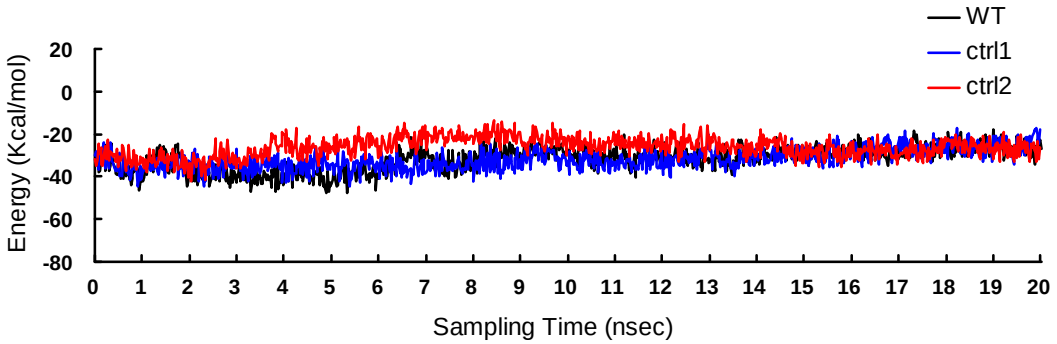
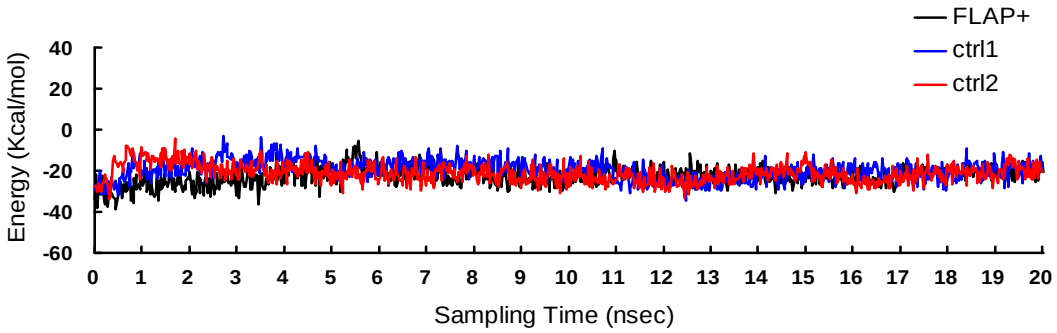

RMSD (A)
Time (ns)
Run1
Run2
Run3

Figure 2.4: MM/GBSA calculated results of DRV-protease binding free energy with respect to the time. The three curves represent calculation results based on the three independent MD trajectories. (A) DRV-WT (B) DRV-FLAP+ (C) DRV-ACT

(A)


WT
ctrl1
ctrl2
Energy (Kcal/mol)
20
0
-20
-40
-60
-80
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Sampling Time (nsec)

(B)


FLAP+
ctrl1
ctrl2
Energy (Kcal/mol)
40
20
0
-20
-40
-60
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Sampling Time (nsec)

(C)

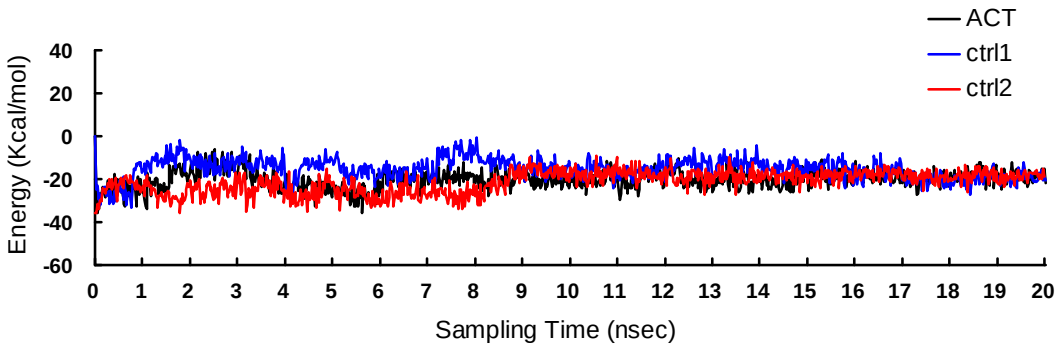

ACT
ctrl1
ctrl2
Energy (Kcal/mol)
40
20
0
-20
-40
-60
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Sampling Time (nsec)

Table 2.1: Results of MM/GBSA Calculation for Absolute Binding Free Energy (kcal/mol) of DRV-Protease Based on Equilibration (1-10 ns) and Production (11-20 ns) Periods.

Protease	Sampling time (ns)	Run 1	Run 2	Run 3	Average
WT	1-10	-33.4	-32.8	-24.9	-30.4
	11-20	-27.3	-27.0	-25.7	-26.7
Flap+	1-10	-23.4	-17.7	-20.1	-20.4
	11-20	-21.1	-20.8	-21.1	-21.0
ACT	1-10	-20.3	-13.9	-25.5	-19.9
	11-20	-17.8	-17.6	-19.8	-18.4

The average predicted binding free energy of WT-DRV is -26.7 kcal/mol, FLAP+-DRV is -21.0 kcal/mol, and ACT-DRV is -18.4 kcal/mol (Table 2.1). Although these values for each system deviate from the ITC experimental values (-15.2 kcal/mol for WT-DRV, -14.2 kcal/mol for FLAP+-DRV, and -13.6 kcal/mol for ACT-DRV), they rank in the correct order for the three protease variants' binding free energy: WT > FLAP+ > ACT. The more rigorous and time-consuming PB method was also tried to calculate the polar solvation free energy. With this method, the predicted results were in better agreement with the ITC experimental data: -15.1 kcal/mol for WT-DRV, -11.6 kcal/mol for FLAP+-DRV, -10.5 kcal/mol for ACT-DRV (Table 2.2). Comparison of the predicted polar solvation free energy difference calculated using the GB and PB models shows that the GB model has underestimated the polar solvation free energy of all three systems. This difference in estimates of polar solvation free energy by the GB and PB models has been reported and discussed in several studies involving different protein-ligand systems.[70-73] Such bias did not affect the ranking of binding energies for a given receptor with different ligands or for receptor variants with a specific ligand. Consistent with other groups[71], the MM/GBSA and MM/PBSA methods provided the same rank order of binding energies although the absolute values were different, for complexes of HIV-1 protease with DRV. (Table 2.2)

Table 2.2: Binding free energy change calculated by the MM/GBSA and MM/PBSA methods (kcal/mol)

Protease	$\Delta G_{CAL\text{-}GB}$	$\Delta G_{CAL\text{-}PB}$	ΔG_{EXP*}
WT	-26.7±1.8	-15.1±1.8	-15.2 ± 0.8
FLAP+	-21.0±1.5	-11.6±1.5	-14.2 ± 0.6
ACT	-18.4±1.7	-10.5±1.7	-13.6 ± 0.1

*Experimental binding free energy data were obtained by ITC (King et al. 2004)

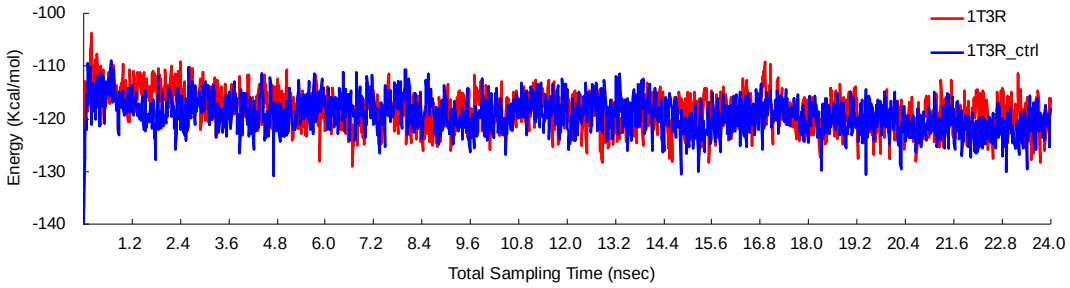
Calculation of relative binding free energy

For the ACT double mutant V82I, I84V, I also calculated relative binding free energy by the more rigorous and computationally more intensive thermodynamic integration method. This method has proven to be a powerful tool for studying binding free energy differences in a receptor-ligand system as statistical mechanics is its theoretical framework.[11,74] As described in the Methods section, thermodynamic integration calculates the binding free energy change from WT-DRV to ACT-DRV. The free energy changes associated with the alchemical pathways ΔG_3 and ΔG_4, which are the sum of 12 weighted $dV/d\lambda$ values, are plotted versus time for the study of calculation convergence (Figure 2.5). For thermodynamic integration calculations, their reproducibility and internal consistency were studied by setting up two sets of independent simulations. Comparison of the two calculations of ΔG_4, which were started from the DRV-WT complex crystal structure coordinates, resulted in using the first 0.5 ns of each of the 12 lambda as the equilibration period and the second 1.5 ns of each of the 12 lambda as the production period. Total time for the equilibration period and production period are 6 ns and 18 ns, respectively. The two ΔG_4 values are -119.3 kcal/mol for run 1 and -119.9 kcal/mol for run 2 (Table 2.3). The ΔG_3 values calculated from the 1HHP and 1T3R crystal structure coordinates are -121.4 and -121.5 kcal/mol, respectively (Table 2.3). The relative binding free energy between DRV-WT and DRV-ACT is 1.8 kcal/mol. This result is a better match with the experimental relative binding free energy of 1.6 kcal/mol than 4.6 kcal/mol and 8.3

kcal/mol, which were calculated from the MM/PBSA and MM/GBSA methods, respectively (Table 2.4).

Figure 2.5: Thermodynamic integration results over total sampling time shows calculations are stable. (A)ΔG_4 from two independent calculations of the coordinates of DRV-WT crystal structure (1T3R). (B)ΔG_3 from two independent starting calculations from different starting structures: DRV-WT crystal structure is colored magenta, with the inhibitor removed, and the apo protease crystal structure (1HHP) is colored cyan.

(A)



(B)

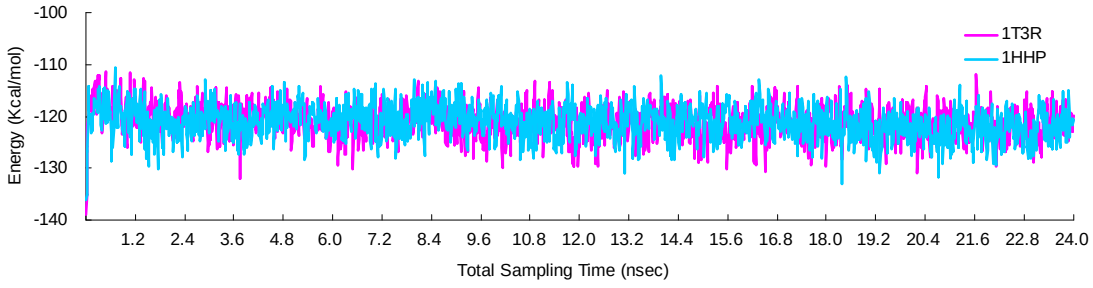


Table 2.3: Thermodynamic Integration Calculation over 12 lamda on the Equilibration and Production Periods (kcal/mol)

Period	ΔG_4		ΔG_3		$\Delta\Delta G^*$
	1st Run	2nd Run	1HHP	1T3R	
equilibration period**	-116.1 ± 0.2	-118.1 ± 0.2	-120.7 ± 0.2	-119.9 ± 0.2	3.2 ± 0.4
production period**	-119.3 ± 0.1	-119.9 ± 0.1	-121.4 ± 0.1	-121.5 ± 0.1	1.8 ± 0.2

***:ΔΔG=Mean (ΔG_4-ΔG_3)**

**** Note equilibration period is the first 0.5 ns of each of the 12 lamda and production period is the second 1.5 ns of each of the 12 lamda, of the entire calculation. Total equilibration time is 6 ns and production time is 18 ns.**

Table 2.4: Relative Binding free energy (kcal/mol) of ACT and WT HIV-1 Protease Calculated by Thermodynamic Integration, MM/GBSA and MM/PBSA Methods vs ITC Data

	Thermodynamic integration	MM/GBSA	MM/PBSA	ITC
ΔΔG	1.8	8.3	4.6	1.6

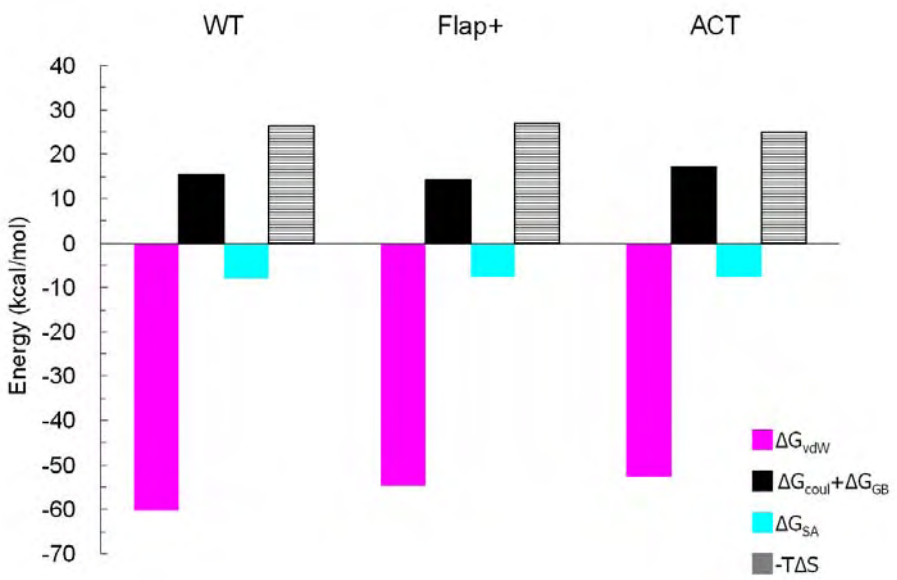
Free-energy decomposition analysis

Analysis of contributions from different energy components

Free energy component analysis is performed to elucidate the mechanism for resistance to DRV of FLAP+ and ACT. The different energy components in the MM-PB/GBSA model (see Methods) are shown in Figure 2.6 and tabulated in Table 2.5 in more detail. Both translational entropy ($-T\Delta S_{translational}$) change and rotational entropy change ($-T\Delta S_{rotational}$) are close in value for DRV binding in the three protease variants. They represent at least 90% of the change in entropy. The remaining vibration entropy change ($-T\Delta S_{vibrational}$) is 1.5 kcal/mol for DRV-WT binding, 2.2 kcal/mol for DRV-Flap+ and 0.2 kcal/mol for DRV-ACT. Further free-energy component analysis reveals that the favorable electrostatic interaction energy term (ΔG_{ELE}) from the molecular mechanical energy (ΔG_{MM}) has been canceled by the unfavorable polar solvation energy (ΔG_{GB}) penalty. This result is in agreement with other MM-PB/GBSA studies.[50,75-78]. The total electrostatic interaction energy ($\Delta G_{ELE}+ \Delta G_{GB}$) for DRV-WT is 15.4 kcal/mol, for DRV-FLAP+ is 14.1 kcal/mol, and for DRV-ACT is 17.0 kcal/mol. The vdW interaction energy is -60.3 kcal/mol for DRV-WT, -54.5 kcal/mol for DRV-FLAP+, and -52.8 kcal/mol for DRV-ACT. vdW interactions have the largest contribution to protease-inhibitor binding (Figure 2.6A) and sustain the largest energy loss in both the FLAP+ and ACT drug-resistant mutants (Figure 2.6B).

Figure 2.6: (A) Binding free energy components of DRV-WT, DRV-FLAP+ and DRV-ACT. (B) The loss of binding free energy components with DRV of FLAP+ and ACT compared to the WT protease.

(A)



(B)

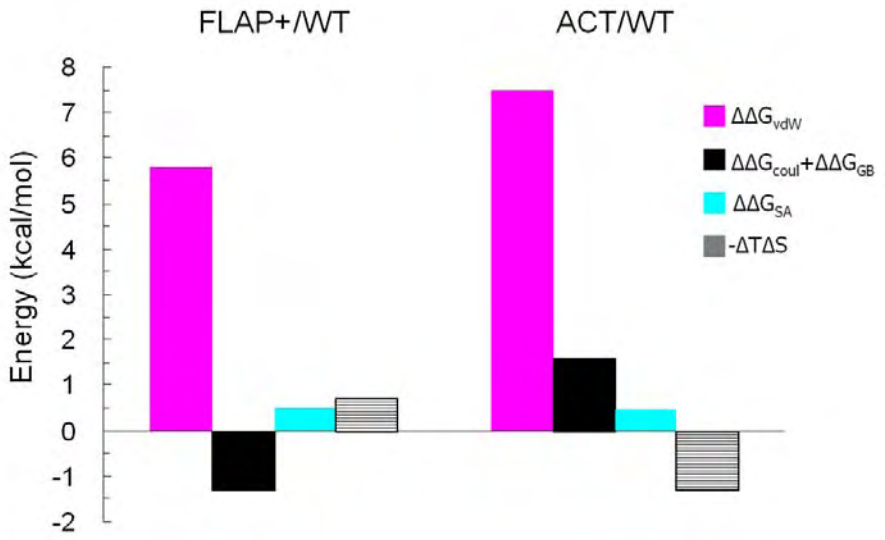


Table 2.5. Binding Free Energy Components (kcal/mol) of DRV with WT, FLAP+, and ACT Protease

Protease	ΔG_{vdW}	ΔG_{ELE}	ΔG_{GB}	ΔG_{SA}	$-T\Delta S_{tran}$	$-T\Delta S_{rot}$	$-T\Delta S_{vib}$	ΔG_{PB}
WT	-60.3	-37.4	52.8	-8.1	13.4	11.4	1.5	64.4
	±0.2	±0.2	±0.2	±0.0	±0.0	±0.0	±1.8	±0.4
FLAP+	-54.5	-39.8	53.9	-7.6	13.3	11.5	2.2	63.3
	±0.1	±0.2	±0.2	±0.0	±0.0	±0.0	±1.5	±0.3
ACT	-52.8	-35.3	52.3	-7.6	13.3	11.5	0.2	60.2
	±0.2	±0.2	±0.2	±0.0	±0.0	±0.0	±1.7	±0.4

Free energy projected to each residue of HIV-1 protease

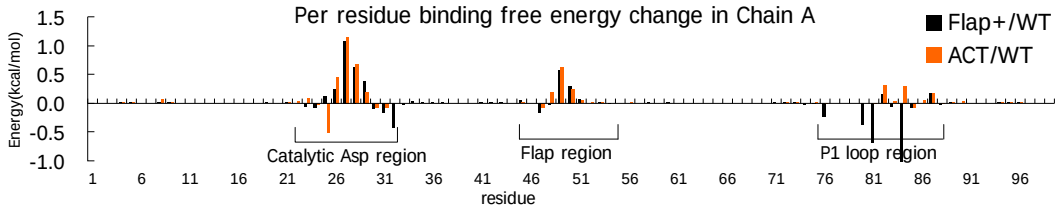
In order to gain extra insight into the mechanisms of protease-inhibitor binding and drug resistance, the binding free energy calculated from the MM-GBSA method has been broken down to individual protease residues. The energy difference was investigated between the WT-DRV complex and the two drug-resistant mutant protease-DRV complexes for each residue (Figure 2.7). The residues with energy changes were mainly located in three areas (Figure 2.8), the catalytic region (residues 22 to 33), the flap region (residues 45 to 55), and the P1 loop region (residues 79 to 87) on both monomers of the protease. These energy changes vary asymmetrically in the two protease monomers. Many residues (G27, A28, I50, R87, K8', A28', D29', I47', L76') structurally adjacent to DRV other than those that mutate (L10I, G48V, I54V and V82A for FLAP+, V82T and I84V for ACT) respond to the mutations (Figure 2.7) as has been previously observed [44,53]. The sites of mutation not only impact their own binding free energy interactions with inhibitors, but also influence the interaction of other residues with inhibitor by inducing alterations in the geometry of the binding site.

Favorable electrostatic interactions opposed by the polar solvation energy penalty also apply to individual residues. A change in electrostatic energy ($\Delta\Delta G_{ELE}$) of any residue is always associated an equal but opposite compensation in solvation energy ($\Delta\Delta G_{GB}$) of very similar amplitude but in a different direction. The correlation coefficient for the $\Delta\Delta G_{ELE}$ and $\Delta\Delta G_{GB}$ of FLAP+ is -0.94 and for the ACT is -0.91

(Figure 2.9A for FLAP+, Figure 2.9B for ACT). This high correlation of $\Delta\Delta G_{ELE}$ and $\Delta\Delta G_{GB}$ makes the change of vdW energy the largest factor in the loss of binding free energy between DRV and FLAP+/ACT. These residues (in the catalytic, flap, and P1 loop regions) also have the largest change in vdW interaction energy (Figure 2.10). To highlight those residues with a significant difference between the WT and the two drug-resistant mutants, a cutoff of 0.05 kcal/mol of vdW energy change was set. The residues in FLAP+ and ACT with a loss of vdW energy greater than the cutoff plotted in Figure 2.10C. In chain A these residues include T26, G27, A28, I47, G49, and I50; in chain B, these residues are K8', D25', G27'-T31', I47'-G49', G51', G52', I54', L76', V82', and G86'. (Figure 2.8D). The loss in vdW interaction energy of chain B is significantly larger than that of chain A.

Figure 2.7: Per residues energy of MM/GBSA calculated results. (A) Energy difference between Wild-type protease and FLAP+ variant. (B) Energy difference between Wild-type protease and ACT variant.

(A)


Per residue binding free energy change in Chain A
Flap+/WT
ACT/WT
Energy(kcal/mol)
1.5
1.0
0.5
0.0
-0.5
-1.0
Catalytic Asp region
Flap region
P1 loop region
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96
residue

(B)

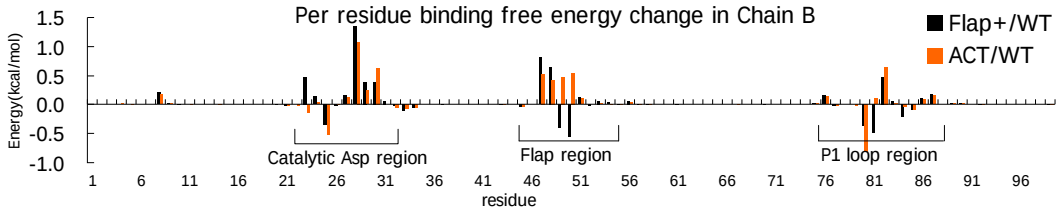
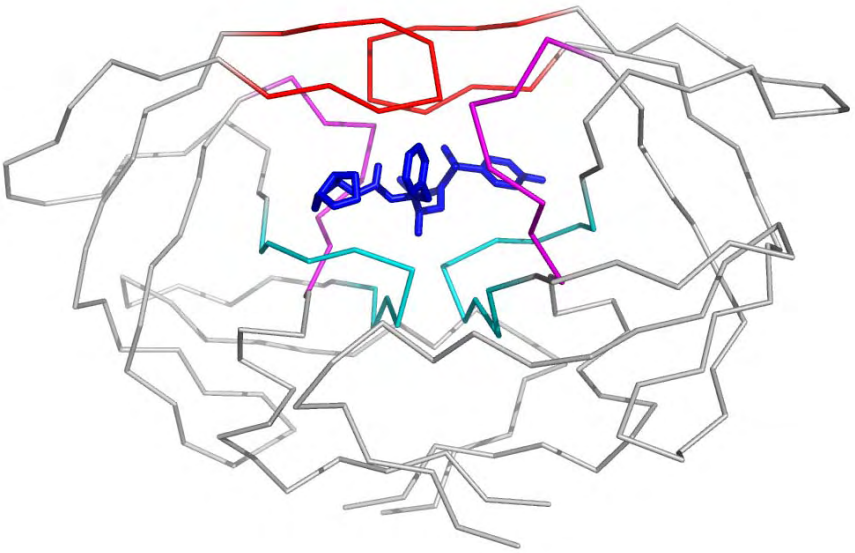
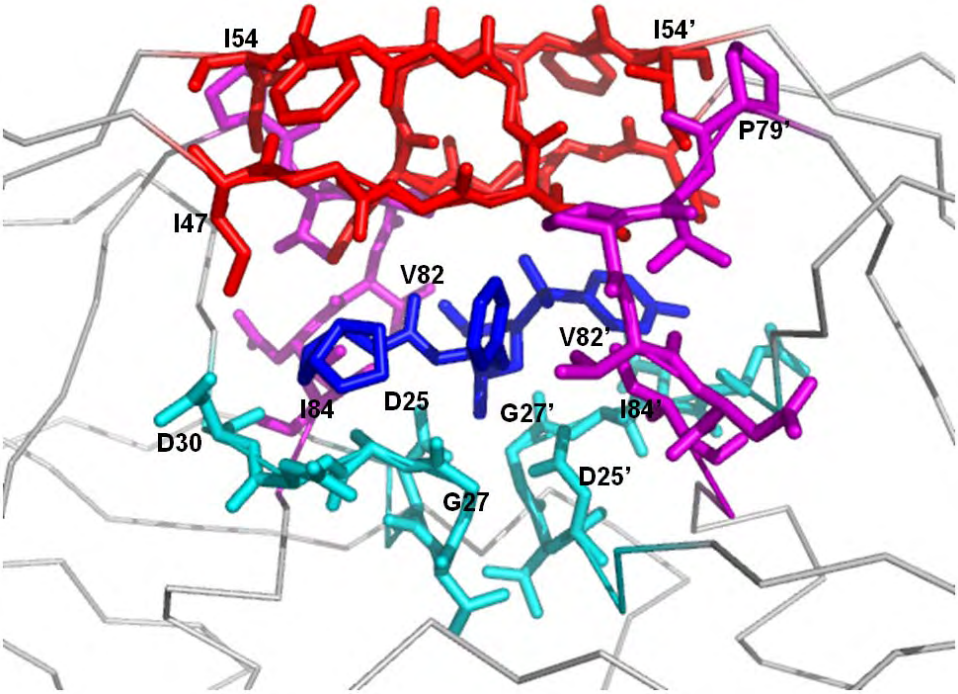

Per residue binding free energy change in Chain B
Flap+/WT
ACT/WT
Energy(kcal/mol)
1.5
1.0
0.5
0.0
-0.5
-1.0
Catalytic Asp region
Flap region
P1 loop region
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96
residue

Figure 2.8: (A) Darunavir in the wild-type protease binding site, highlighting the flap region(residue 47-54 of both monomers in red), catalytic region(residue 25-30 in magenta) and P1 loop region(residue 79-84 in cyan). (B&C) Atomic details of the protease binding pocket viewed from the front(B) and top(C). (D) Residues which have significant lost of vdW interactions energy (larger than 0.2 kcal/M) in the drug resistant variants.

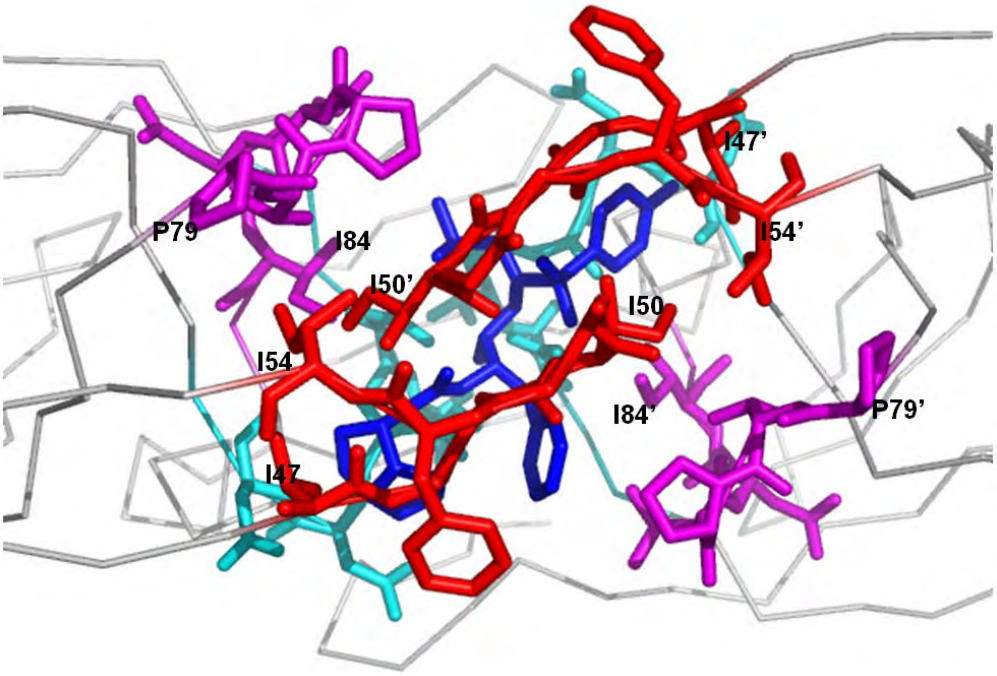
(A)



(B)



(C)


I47'
P79
I84
I54'
I50'
I50
I54
I84'
P79'
I47

(D)

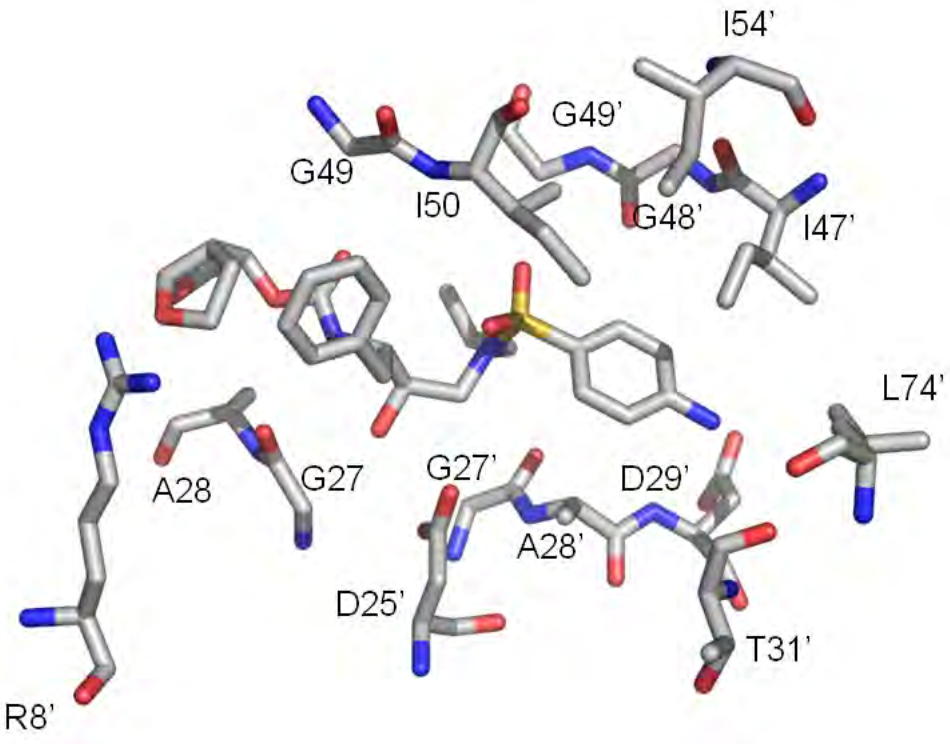
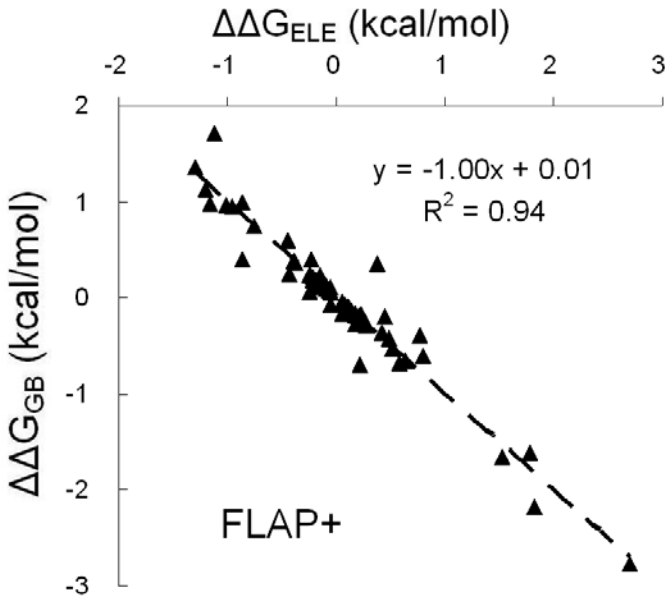

I54'
G49'
G49
I50
G48'
I47'
L74'
A28
G27
G27'
D29'
A28'
D25'
T31'
R8'

Figure 2.9: Correlation between $\Delta\Delta G_{ELE}$ and $\Delta\Delta G_{GB}$ of each residue. (A) Energy difference between FLAP+ and WT. (B) Energy difference between ACT and WT.

(A)



(B)

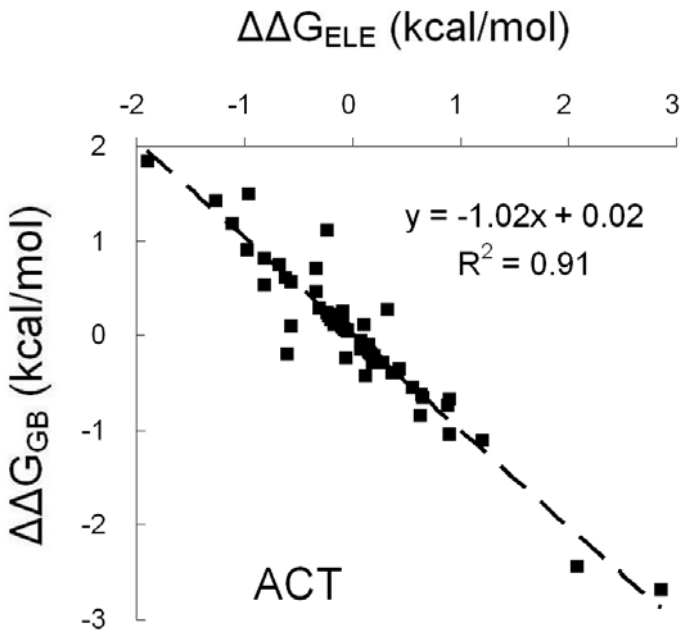
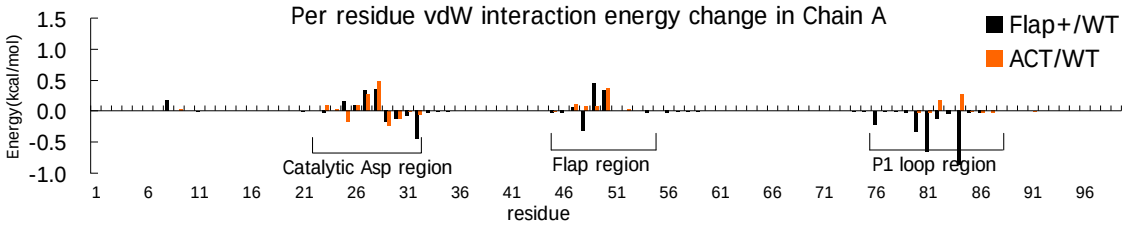
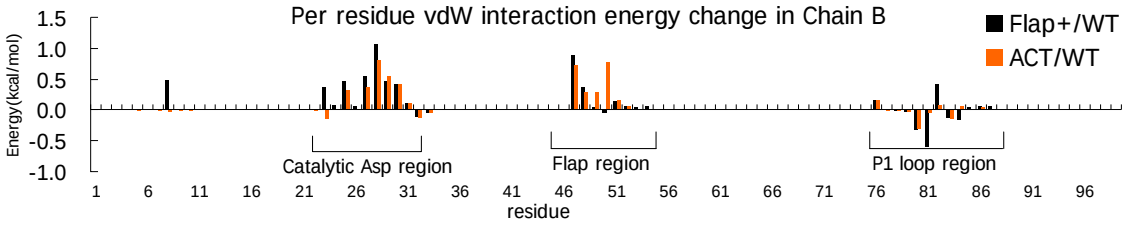


Figure 2.10: (A) Per residues vdW energy loss between FLAP+ and WT protease. (B) Per residues vdW energy loss between ACT and WT protease. (C) Residues with a vdW energy loss larger than 0.05 kcal/mol.

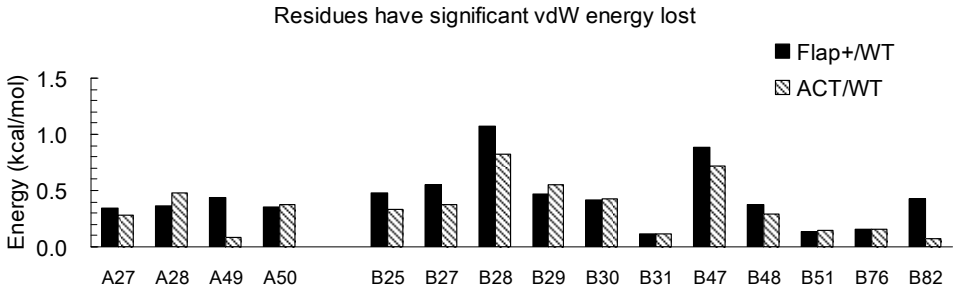
(A)


Per residue vdW interaction energy change in Chain A
Flap+/WT
ACT/WT
Energy(kcal/mol)
1.5
1.0
0.5
0.0
-0.5
-1.0
Catalytic Asp region
Flap region
P1 loop region
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96
residue

(B)


Per residue vdW interaction energy change in Chain B
Flap+/WT
ACT/WT
Energy(kcal/mol)
1.5
1.0
0.5
0.0
-0.5
-1.0
Catalytic Asp region
Flap region
P1 loop region
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96
residue

(C)


Residues have significant vdW energy lost
Flap+/WT
ACT/WT
Energy (kcal/mol)
1.5
1.0
0.5
0.0
A27 A28 A49 A50 B25 B27 B28 B29 B30 B31 B47 B48 B51 B76 B82

vdW energy contribution from each DRV atom

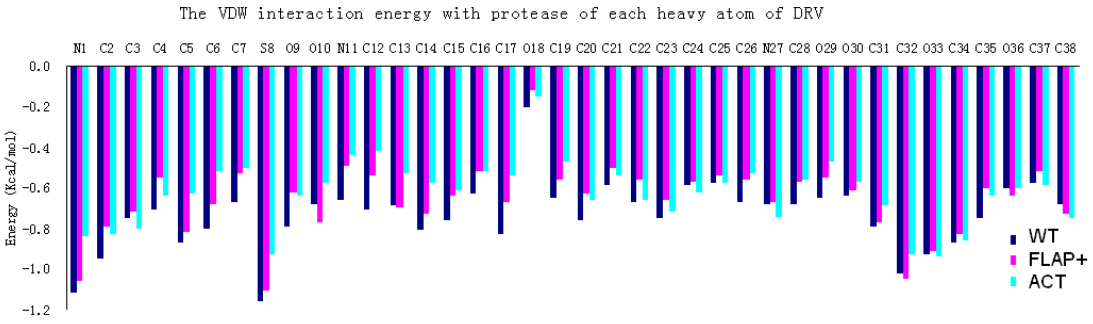
To explore the mechanism of binding free energy loss between DRV and the drug-resistant mutants, the vdW energy contributions were calculated for each DRV atom and compared between complexes with the WT, FLAP+ and ACT mutant proteases. DRV has 75 atoms, of which 37 are hydrogen atoms with very limited contribution to the vdW interaction energy. Thus, data are presented for only the 38 heavy atoms in DRV (Figure 2.11A). Structurally, DRV can be considered formed by four major moieties: A) 4–aminophenyl group, B) isopropyl group, C) benzyl ring, and D) *bis*-tetrahydrofuranylurethane (THF) (Figure 2.11B). To compare the energy change between these 4 moieties, I define DV (loss of vdW interaction energy ratio) as:

$$DV = \frac{\sum_i (E_i^{Flap+} - E_i^{WT}) + \sum_i (E_i^{ACT} - E_i^{WT})}{2 \times \sum_i E_i^{WT}} \times 100\%$$

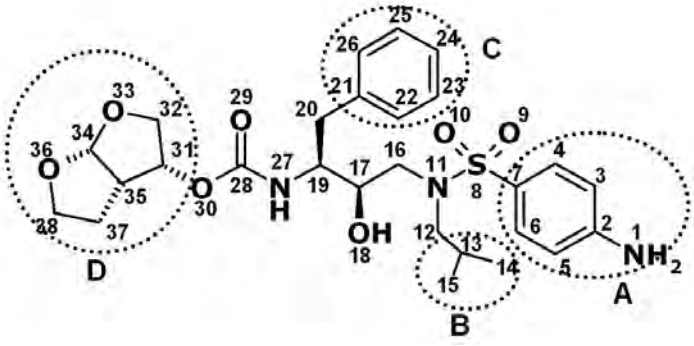
where *i* is the atom within a specific moiety. The *bis*-THF moiety and the benzyl ring have relatively low DV of 3.1% and 8.5%, respectively. The 4–aminophenyl and isopropyl groups have significantly higher DV of 17.0% and 19.2%, respectively (Table 2.6).

Figure 2.11: (A) vdW interaction energy of each heavy atoms of DRV with protease. Energy of DRV-WT is colored dark blue, energy of DRV-FLAP+ is colored magenta and energy of DRV-ACT is colored cyan. (B) The definition of four moieties of DRV. (C) The chemical structure of APV.

(A)



(B)



(C)

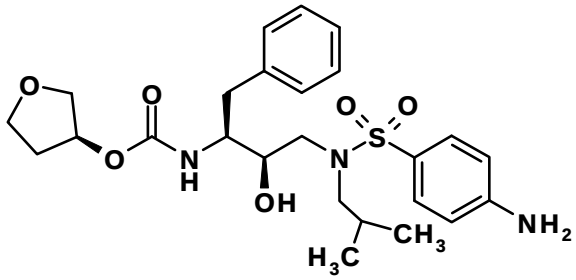


Table 2.6: Loss of Van der Waals' Interaction Energy (DV) for Different DRV Moieties

	4 – Amino Phenyl Group	Isopropyl Group	Benzyl Ring	*bis*-Tetrahydrofuranyl
DV (%)	17	19	9	3
Number of heavy atoms	7	4	7	8

The only difference between DRV and a previous generation protease inhibitor, amprenavir (APV) (Figure 2.11C), is that DRV has a second tetrahydrofuran ring, which is part of its *bis*-THF moiety. Nonetheless, DRV has been shown by ITC experiments[35] to bind more tightly than APV with the protease, with a 2.6 kcal/mol larger binding affinity. In the DRV-WT protease structure (1T3R) the *bis*-THF moiety is surrounded by the protease chain A residues Ala28, Asp29, Asp30, Ile47, Gly48, and Gly49, which form a cluster of vdW contacts (Figures 2.12A and 2.12B). This packing can be also observed from the crystal structures of the DRV-FLAP+ and DRV-ACT complexes[35]. Examination of the MD simulation structures of DRV in complex with the WT, FLAP+ and ACT proteases show that these residues and the *bis*-THF moiety maintain a relatively stable structure compared to other parts of the inhibitor. This stability leads to the small ratio of the *bis*-THF group's vdW energy loss (Table 2.6).

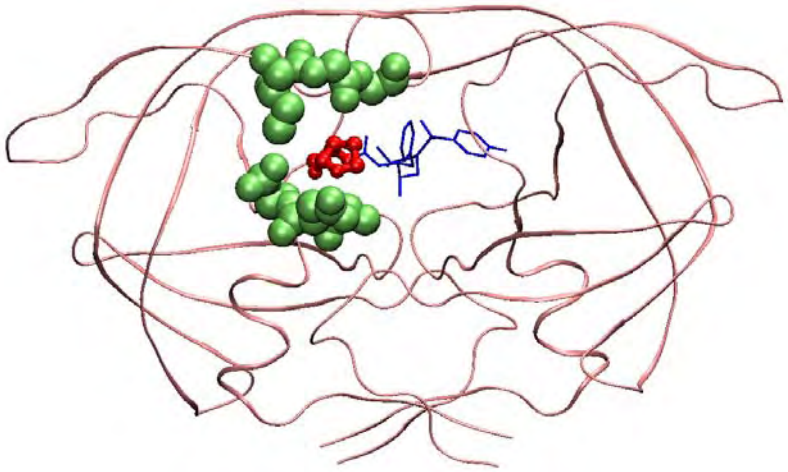
Similar to the *bis*-THF group, the benzyl ring maintains its vdW interactions with protease residues in chain A (Figure 2.12C) in most conformations sampled by the MD simulations. This stability in DRV interactions with chain A explains the asymmetric vdW energy losses of the protease's two chains. Chain B is shown by free-energy decomposition of protease residues to have more residues with significant energy loss than chain A (Figure 2.9C). Unlike the *bis*-THF group and the benzyl ring, whose vdW interactions were only slightly influenced by the drug-resistant mutations, the 4–aminophenyl and isopropyl groups of DRV in complex with FLAP+ and ACT

lost approximately 20% of the vdW interaction energy. Comparison of the MD simulation structure of DRV-WT with those of the two drug-resistant mutants shows that the 4 – amino phenyl and isopropyl groups of DRV in the DRV-FLAP+ and DRV-ACT complexes undergo significant geometry changes (Figure 2.13) that lead to these groups losing their vdW contacts with the drug-resistant proteases.

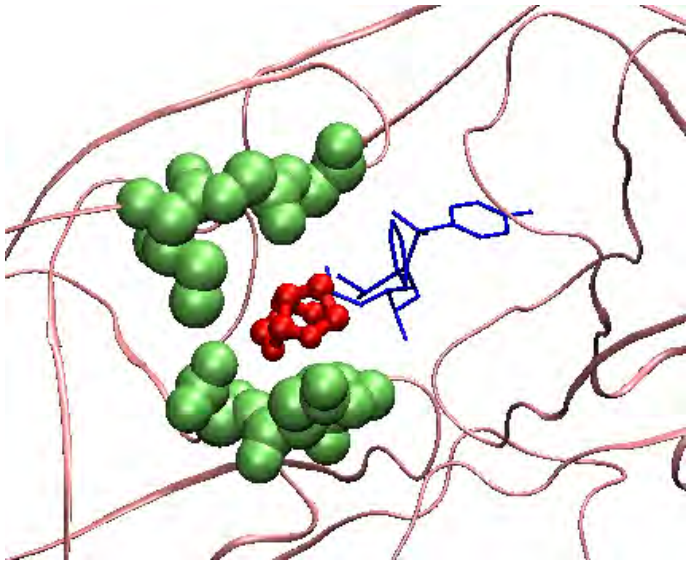
The free-energy decomposition by residue shows that the mutations have induced changes in the shape of the binding pocket as evidenced by the predominant changes occurring in the vdW interactions energy. Overall there is a decrease in the vdW interaction energy between the protease and DRV, mostly on the 4- amino phenyl side, as the volume of the binding pocket is effectively enlarged as the mutations within the active site are to smaller residues (V82A in Flap+ and I84V in ACT). Expansion of the active site permits, other residues to interact to varying degrees with the inhibitor. In this way, the FLAP+ and ACT mutant proteases could develop drug resistance.

Figure 2.12: (A) and (B) Cluster of vdW contacts formed by the *bis*-THF group and the protease residues Ala28, Asp29, Asp30, Ile47, Gly48, and Gly49 of chain A. The atoms of above residues are displayed and colored green. The atoms of bis-THF group are colored red while the remain of DRV is colored blue. (C) Relative position of DRV's four moieties (colored yellow) to Chain A (colored cyan) and Chain B (colored purple) of protease.

(A)



(B)



(C)

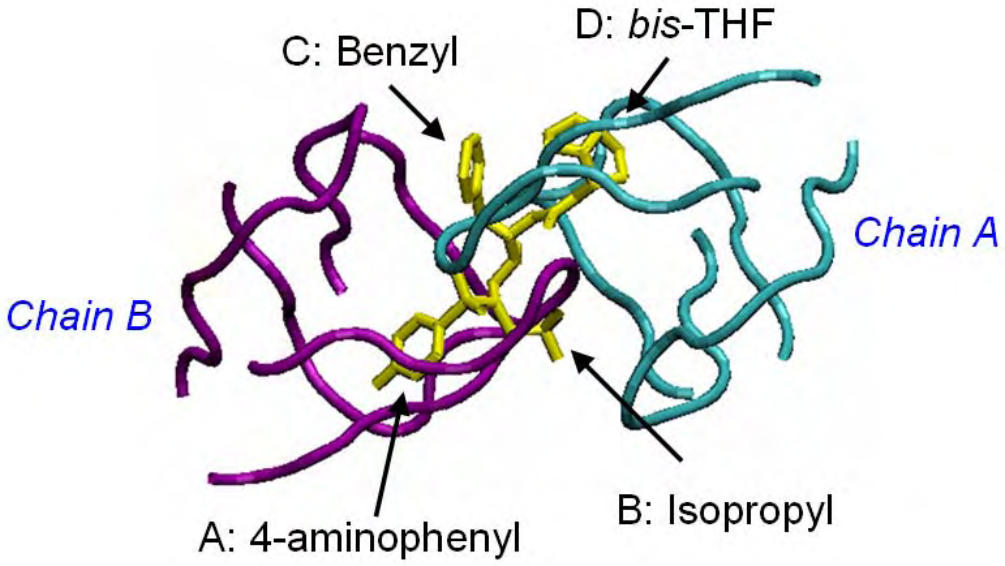
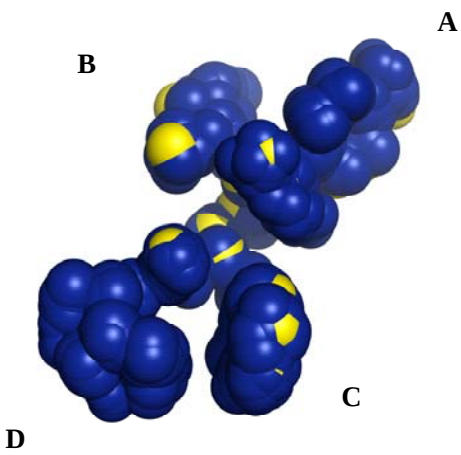
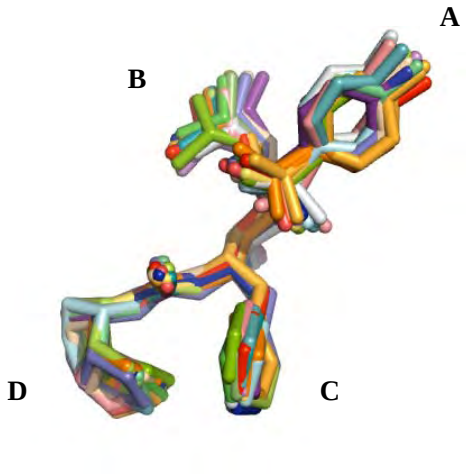


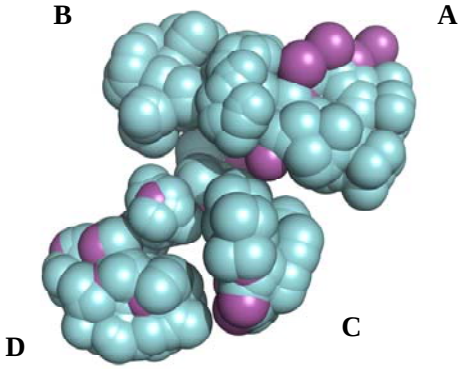
Figure 2.13: (A) Conformational space of DRV sampled in DRV-WT complex simulations. Left: DRV ensemble is shown with atoms vdW radii. The original conformation as in the crystal structure is colored yellow. The sampled conformations ensemble from MD simulation is colored blue. Right: 20 snapshots of DRV conformations taken every 1ns from MD simulations. (B) Conformational space of DRV sampled in DRV-FLAP+ complex simulations. Left: DRV ensemble is shown with atoms vdW radii. The original conformation as in the crystal structure is colored purple. The sampled conformations ensemble from MD simulation is colored cyan. Right: 20 snapshots of DRV conformations taken every 1ns from MD simulations. (C) Conformational space of DRV sampled in DRV-ACT complex simulations. Left: DRV ensemble is shown with atoms vdW radii. The original conformation as in the crystal structure is colored orange. The sampled conformations ensemble from MD simulation is colored green. Right: 20 snapshots of DRV conformations taken every 1ns from MD simulations.

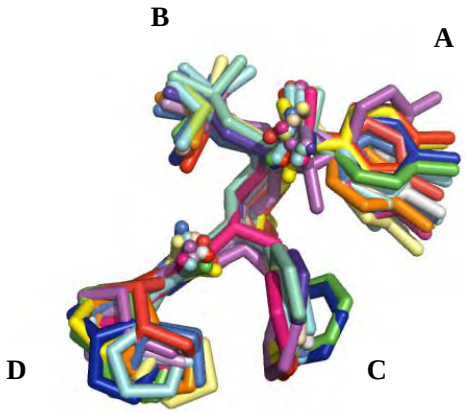
(A)



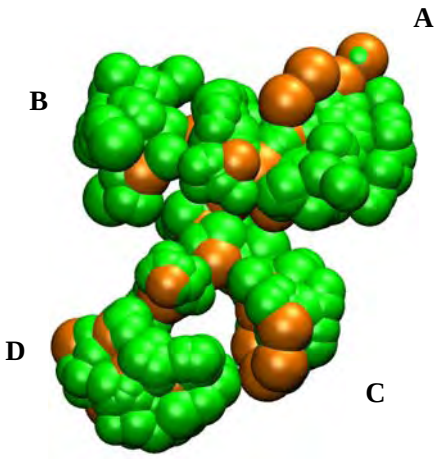


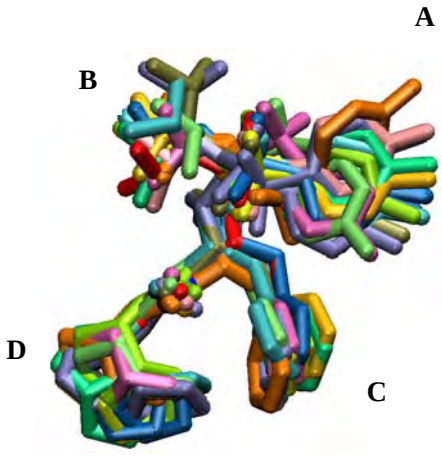
(B)





(C)





Conclusion

With the appearance of drug-resistant HIV-1 protease variants becoming the major challenge to current AIDS therapy, understanding the mechanism of drug resistance is critical. This goal is best addressed by cross-analyzing the data on protease mutants from different experimental methods such as crystallography and isothermal titration calorimetry. However, more details about inhibitor-protease binding can be provided by free energy calculations, which start from structural coordinates and yield thermodynamic data. In this study I performed MM-PB/GBSA calculations and free-energy component analysis of DRV-WT, DRV-FLAP+ (L10I, G48V, I54V, V82A) and DRV-ACT (V82T, I84V). By running three independent 20 ns simulations for each of these systems, I not only identified the convergence and consistency of our calculations, but also predicted binding energies in good agreement with ITC data. Moreover, the relative binding free energy between DRV-WT and DRV-ACT using MM-PB/GBSA and thermodynamic integration (TI) methods was calculated. The accuracy of these result had the rank order TI>MM/PBSA>MM/GBSA, which is the same order of the computational times required for these methods. The TI method is more suitable for comparing the energy difference between two similar systems. In the case of the ACT (V82T, I84V) mutant here, the TI method not only gave the more accurate predicted energy than the MM-PB/GBSA method, but also had better reproducibility and faster convergence

(Figure 2.4, Figure 2.5). The free-energy decomposition analysis indicated that mutations in the protease induced conformational changes in its active site. The *bis*-THF group and benzyl ring of DRV sustain their vdW interactions with the drug-resistant protease variants and contribute most to the inhibitor-protease binding, while DRV's 4 – amino phenyl and isopropyl groups are susceptible to changes in the protease's binding pocket and adopt conformations that lose vdW interaction with drug-resistant variants. (Table 2.6)

These findings suggest that the design of new protease inhibitors based on the DRV scaffold should consider replacing the 4–aminophenyl and isopropyl groups since these regions do not maintain their interactions with drug resistant protease variants as much as the *bis*-THF group. Such new inhibitors would likely bind more tightly to HIV protease and may be less susceptible to drug resistance.

Chapter III

Molecular dynamics simulation of HIV-1 protease: Comparison of the dynamic properties of WT protease and a drug-resistant variant

Abstract

HIV-1 protease is an aspartyl enzyme that processes the viral Gag-Pol polyprotein, yielding the structural proteins and enzymes necessary for the maturation of infectious viral particles. The protease is a homodimer with each chain having 99 residues, 40 of which are hydrophobic. The two monomers bind together by nonbonded interactions, and the active site is located at the interface between the two monomers. HIV-1 infected individuals are currently treated with 9 Federal Drug Administration-approved protease inhibitors. Due to the high HIV-1 replication rate and the lack of proofreading ability of its reverse transcriptase, exposure to protease inhibitors leads to the protease developing drug-resistant mutations. Understanding drug-resistant mechanisms on a molecular level is crucial for developing more efficient inhibitors. To compare the dynamic properties of the wild-type protease and drug-resistant variants, I performed molecular dynamics simulations of the wild-type and Flap+ (L10I, G48V, I54V, V82A) variant using the AMBER 8 package. The computational analysis shows that this variant has decreased the flexibility in the apo form, which might lead to more favorable entropy change upon binding with inhibitors compared to the wild-type protease.

Introduction

Acquired immunodeficiency syndrome (AIDS) has become a worldwide, public health threat since 1980. The AIDS patient's immune system is weakened by infection with the human immunodeficiency virus (HIV-1). Maturation of HIV-1 virus requires HIV-1 protease to cleave the Gag and Gag-Pol polyproteins to release the structural proteins MA, CA, NC and p6, and the enzymes RT, IN and PR. The important role of protease in the life cycle of HIV-1 has made it a crucial target for drug design in AIDS therapy.

The 22 kDa HIV-1 protease is a homodimer protein with 99 amino acids on each subunit. HIV-1 belongs to the aspartyl enzyme family, with both monomers having the conserved active triad residues: Asp25, Thr26, and Gly27. The two protease monomers are joined by nonbonded interactions at the N- and C-terminal regions. Each monomer has a glycine rich flap region: K45-M-I-G-G-I-G-G-F-I-K55. This flap region folds as an anti-parallel β strand that covers the active site. Since the flap regions control access to the buried active site, the flexibility of the flap region is crucial for enzyme activity and active-site inhibitor binding. Unlike the fully covered status in complexes of the protease with inhibitor/substrate, the flap regions in the apo protease adopt a "semi-open" status [79-84]. However, entrance of the substrate or inhibitor requires further conformational changes of the flap regions. The flexibility of the flaps in unliganded protease has been studied by fluorescence experiments [85,86].

NMR relaxation experiments have shown that the flap region is mobile on a microsecond time scale [30]. An even higher degree of mobility, on a sub-nanosecond time scale, has been shown for the flap tip 48G-G-I-G-G52 [30]. NMR relaxation experiments can detect the overall rate of regional movements of the protease, but this approach cannot provide details about the role of each atom in these motions. Such information is experimentally difficult to attain.

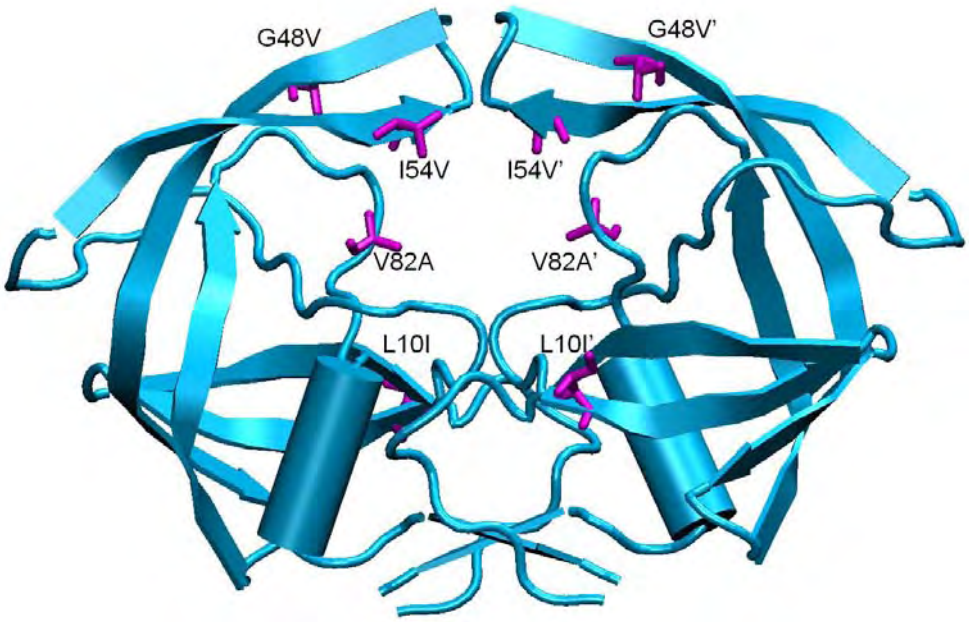
Since HIV-1 protease is an important drug target for HIV-1 therapy, the development of protease inhibitor drugs has been well studied. Nine protease inhibitors have been approved by the Federal Drug Administration for clinical use in fighting AIDS, leading to a significant decrease in the death rate due to AIDS. However, due to the high replication rate of the virus and lack of proofreading mechanism of its reverse transcriptase, drugs exert a selection pressure on the virus, which quickly develops many drug-resistant protease variants.

The structural basis for these changes in drug-resistant proteases has been studied by solving the crystal structures of HIV-1 protease variants. Comparison between the structures of wild-type and drug-resistant variant proteases in complex with inhibitors partially elucidated how specific protease mutations decrease protease-inhibitor binding affinity on the atomic level [37,38]. However, static crystal structures cannot provide insights into the mechanisms by which mutations change the dynamic properties of the protease. Neither can it provide an illustration on the enthalpy-entropy compensation observed from the ITC experiments. To fill in the

blanks of information gained by experimental methods such as NMR and crystallography, computational methods, such as molecular dynamics (MD) simulations are used.

Flap+ is a multiple drug-resistant protease variant with mutations L10I, G48V, I54V, and V82A (Figure 3.1). The dissociation constant K_d of DRV-Flap+ binding measured by isothermal titration calorimetry experiment is $2.58X10^{-11}$ M, for DRV-WT is $4.48X10^{-12}$ M.[87] The Gibbs free energy change for DRV-WT is -15.2kcal/mol, for DRV-Flap+ is -14.2kcal/mol. The ΔΔG is 1 kcal/mol. Both the enthalpy and entropy change are remarkably large. The ΔΔH is 14.1 kcal/mol and the Δ(-TΔS) is -13.1 kcal/mol. Similar scale of enthalpy-entropy change is observed from binding of Flap+ with other inhibitors.[87] In order to investigate the molecular mechanism that causes these unusual thermodynamics, 20 ns MD simulations were performed and analyzed on the apo wild-type protease and apo Flap+ protease. The results of comparative analysis show that the apo Flap+ variant is less flexible than the wild-type protease probably due to the flap mutations G48V and I54V decreasing the flexibility of the protease flap region, but upon complex formation the Flap+ variant become more flexible.

Figure 3.1: Drug-resistant HIV-1 protease variant Flap+. Mutations are highlighted in magenta.



Methods

MD simulations were performed using the program Sander in the AMBER 8 (Assisted Model Building with Energy Refinement) package [67]. The initial coordinates for the MD simulation of the wild-type protease crystal structure were provided by 2HB4 [81], an apo wild-type HIV-1 protease structure. The initial coordinates for the MD simulation of the Flap+ variant were modeled from 2HB4 using geometry in the AMBER package. Another apo HIV-1 protease crystal structure, 1HHP, was not used in this study because of its poorer resolution. The first conformation of multiple occupancy structures was used in the simulations. Any missing side-chain atoms were built back into the structures using the Leap program and default geometry in AMBER. All ionizable residues were left in their standard states at pH 7.

For the standard protease residues, atomic partial charges, van der Waals parameters, equilibrium bond lengths, bond angles, dihedral angles, and their relative force constants were taken from the AMBER database (ff03)[57]. The explicit solvent model was applied to all systems. Each structure was solvated with the TIP3P water cubic box to allow for at least 8 Å of solvent on each face of the protease. The vdW dimensions for the protease are 44 by 35 by 59 Å. The dimensions of the final periodic box are 63 by 55 by 78 Å. The simulation system had approximately 7000 water molecules, and six Cl- counterions were added to balance the charge of the

system. A three-step energy minimization process with the steepest descent method was used to allow the system to reach an energetically favorable conformation.

In the first energy minimization step, all the heavy atoms of the protease were restrained with a harmonic force constant of 10 kcal $mol^{-1}Å^{-2}$. In the second step, only the backbone nitrogen, oxygen, and carbon atoms were restrained. In the third step, all atoms were allowed to move. Each of the three steps had 2000 cycles. The temperature of the energy-minimized system was then gradually raised from 50ºK to 300ºK in the NVT ensemble. Initial velocities were assigned according to the Maxwellian distribution with a random seed. In the thermalization process, heavy atoms were restrained with a harmonic force constant of 10 kcal mol^{-1} $Å^{-2}$. The whole process was 50 picoseconds (50,000 steps, each 1 femtosecond [fs]). A 50 picosecond equilibration was then performed in the NPT ensemble without restraining heavy atoms. In the subsequent sampling MD simulations, each step was 2 fs, and the trajectory was recorded every 100 fs. The total simulation time was 20 ns. For the thermalization, equilibration, and sampling simulations, the SHAKE algorithm [61] was applied to constrain all hydrogen atoms. At every 1 ps, a snapshot was taken to be analyzed for the production phase.

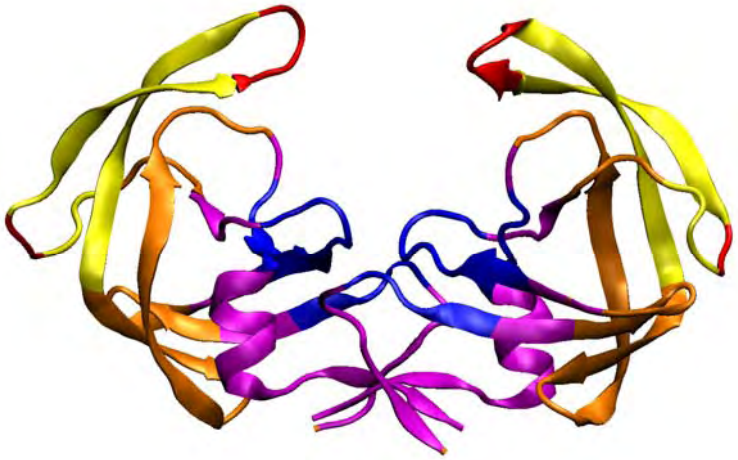
Results and Discussion

RMSF values of Cα atoms

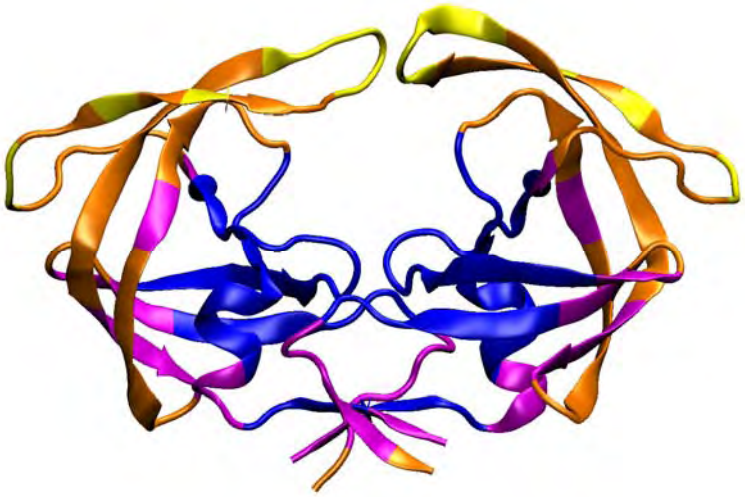
One approach to investigating the dynamic features of a system during MD simulation is the root-mean-square fluctuation (RMSF) values of Cα atoms for each residue, which reflects the backbone mobility of a residue around its average position. For both wild-type protease and the Flap+ variant, the regions with the highest RMSF values are the flap area and their terminal loops (residues 37-59) (Figure 3.2). The most stable areas with the lowest RMSF values in both wild-type protease and the Flap+ variant are the nearby active-triad regions (residues 23-28) and the dimer interface (residues 3-9, 90-97). These findings are consistent with NMR relaxation experiments.[30] The flap region residues have the lowest N-H S^2 order parameters, and the active-triad and dimer-interface regions residues have the highest N-H S^2 order parameters [30,31]. Comparison of the RMSF values between the wild-type protease and the Flap+ variant shows that the wild-type protease has overall higher RMSF values than the Flap+ variant, which indicates that the wild-type protease is more dynamic than the Flap+ variant in solution.

Figure 3.2: RMSF values of Cα atoms for each residue highlighted in different colors. (A) Wild-type protease. (B) Drug-resistant variant Flap+. Red: >4Å, Yellow: >3Å, Orange: >2Å, Magenta: >1.5Å, Blue: <1.5Å.

(A)



(B)



Nanosecond snapshots comparison

To observe conformational changes with respect to time, a snapshot was taken every nanosecond and compared with the crystal and pre-equilibrated structures (Figure 3.3). From 1 to 6 ns, the wild-type protease backbone conformation has similar changes to those of the Flap+ variant. The most significant conformational changes can be observed from the diagram at the flap regions, especially at the flap tips and external loops where the residues have the highest RMSF values and lowest S^2 order parameters. From 7 to 17 ns, the Flap+ variant's flaps on both monomers moved slightly towards each other, closer to the active triads than in the crystal structure.

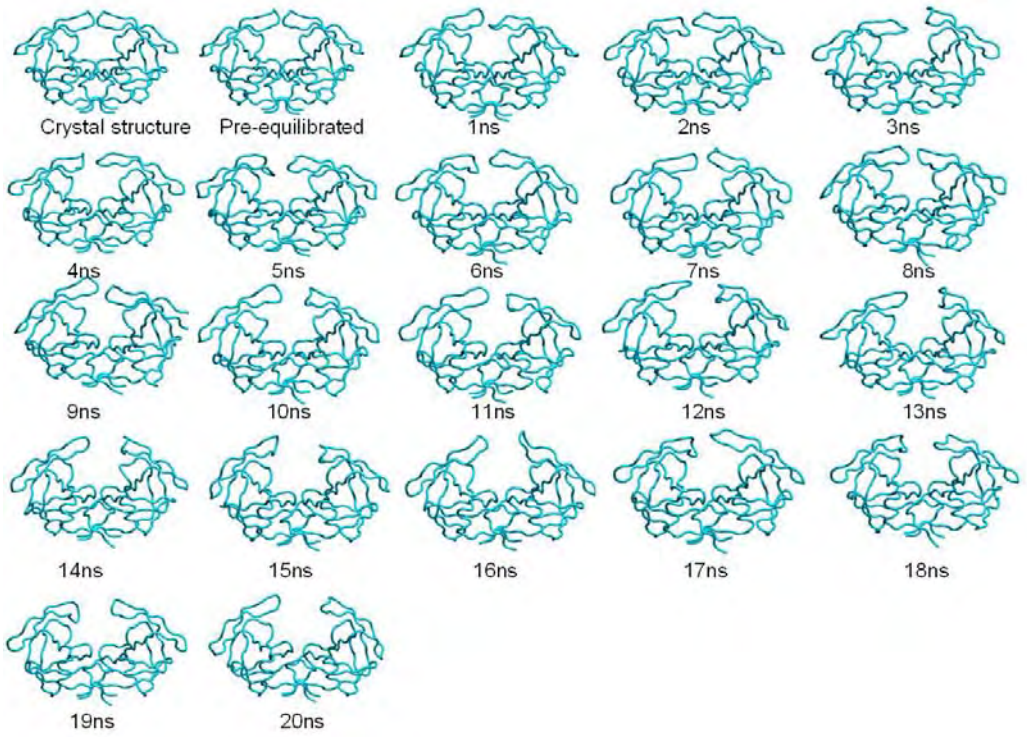
One major conformational difference between the apo protease and liganded protease is that the flap regions of the apo protease are in a "semi-open" conformation while the liganded protease flaps are fully closed [80-83]. Both NMR relaxation experiments [88] and free-energy simulations [89] demonstrated that in the apo protease, the closed conformation is less favorable for the flaps to adopt than the semi-open conformation. The ability of the Flap+ variant to sample conformations more similar to the closed state indicates that the G48V and I54V flap mutations have stabilized the flap structures.

After 17ns, the distance between the flap tips of the Flap+ variant started to increase again, and the conformations are more similar to those sampled in the first 6 ns. For the wild-type protease, its flaps gradually opened and exposed the active-site

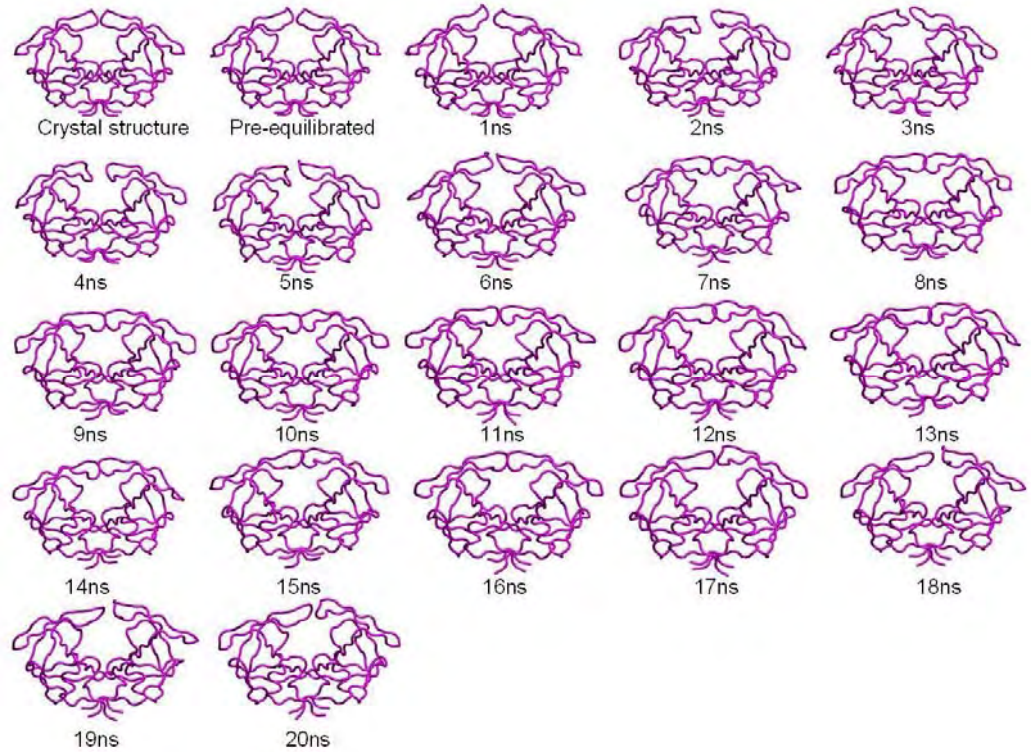
pocket. During this process, between 17 to 18 ns, the wild-type protease flaps once again moved towards each other. After 18 ns, the opening of the flaps resumed. The enzyme activity of HIV-1 protease is highly influenced by the residues in the flap regions [7,90]. Since the substrate and inhibitor entrance to the active site requires the flaps of the protease to open, the lower flap opening rates of the Flap+ variant compared to the wild-type protease observed here could be one reason that the Flap+ variant has lower enzyme activity and less binding affinity with inhibitors.

Figure 3.3: Snapshots of the simulated structures of wild-type protease and the Flap+ variant. (A) Wild-type protease conformations are in cyan. (B) Flap+ variant conformations are in magenta.

(A)


Crystal structure
Pre-equilibrated
1ns
2ns
3ns
4ns
5ns
6ns
7ns
8ns
9ns
10ns
11ns
12ns
13ns
14ns
15ns
16ns
17ns
18ns
19ns
20ns

(B)

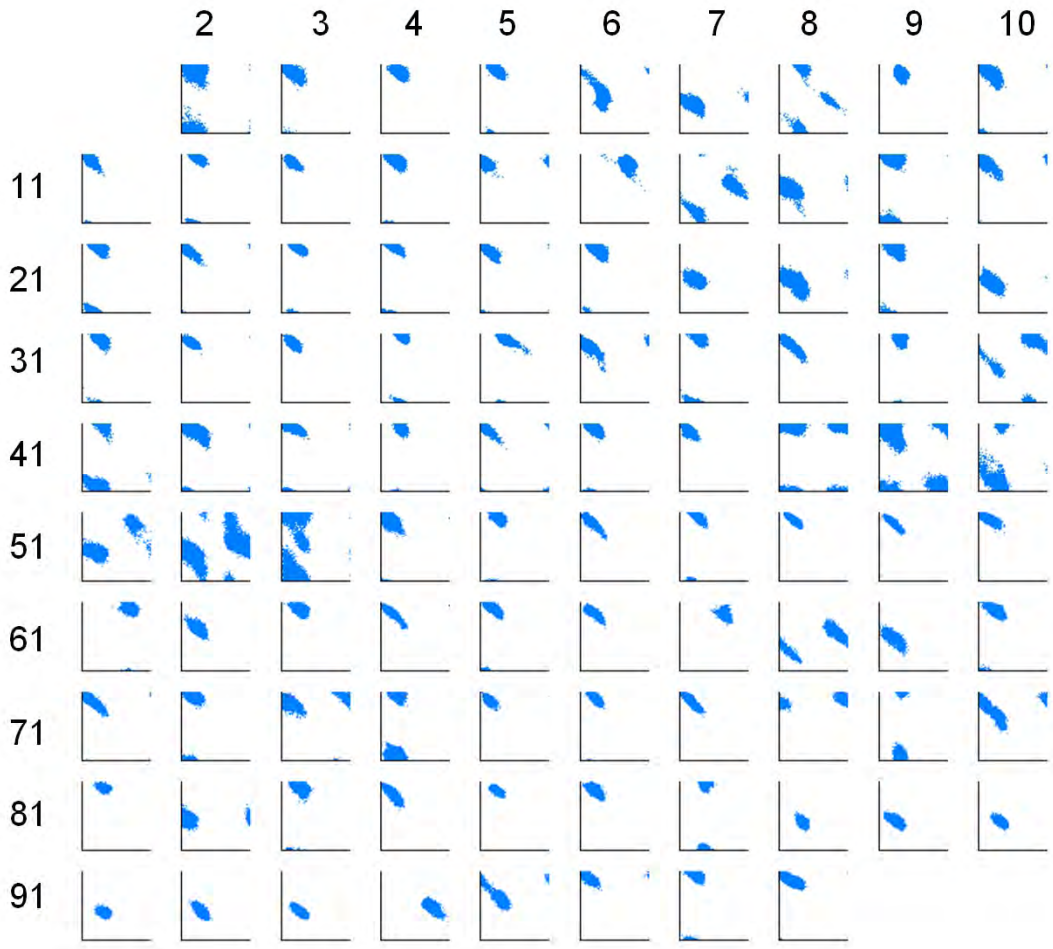

Crystal structure
Pre-equilibrated
1ns
2ns
3ns
4ns
5ns
6ns
7ns
8ns
9ns
10ns
11ns
12ns
13ns
14ns
15ns
16ns
17ns
18ns
19ns
20ns

Phi-Psi dihedral angle distribution

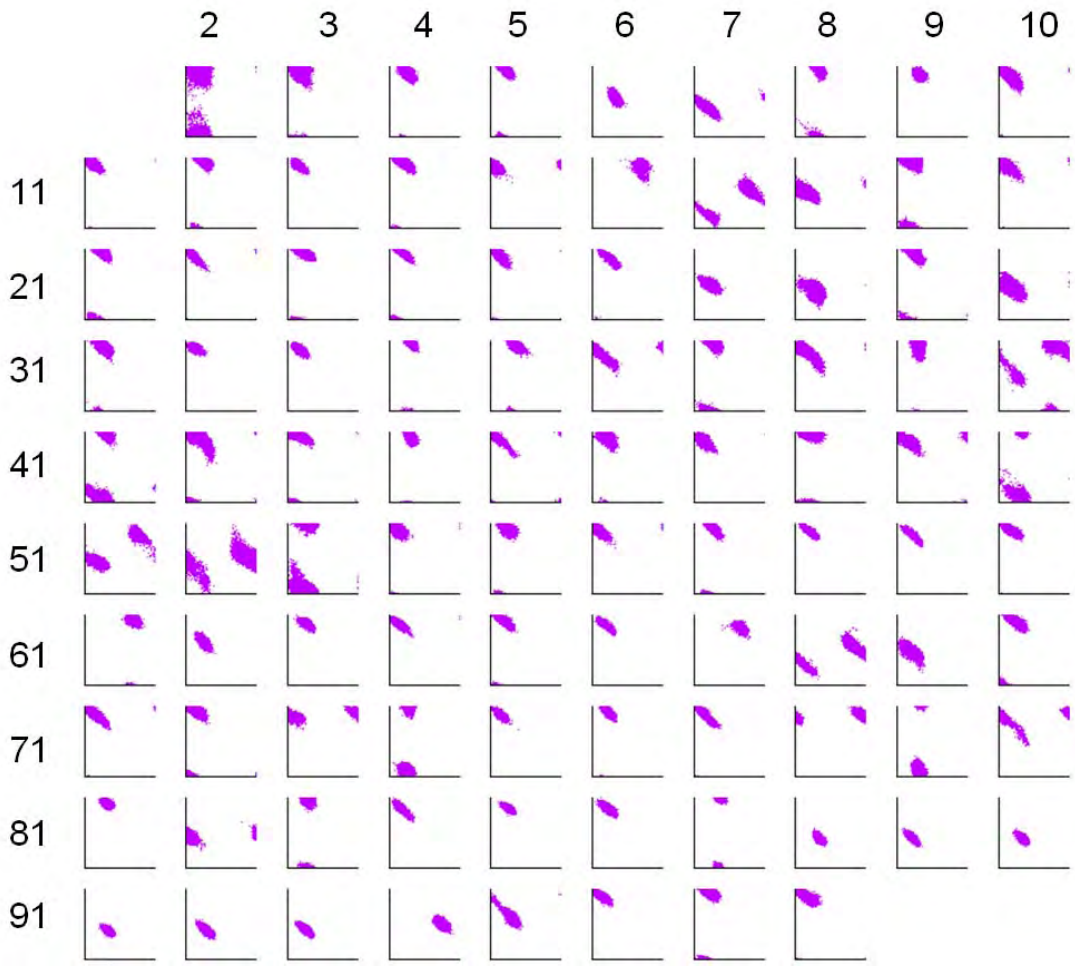
In order to compare backbone flexibility of the wild-type protease and Flap+ variant, the phi-psi dihedral angles were plotted for each residue of the proteases for each pico-second over the course of the 20 nano-second trajectories. (Figure 3.4A, B) As the calculation is on the symmetric apo enzyme, data from each monomer was superimposed. Although most of the phi-psi pairs exhibit very similar ensembles between the two trajectories, residues W6, R8, G48V, I49, G52 and F53 exhibit significant differences (Figure 3.4C). All of these residues, mainly located in the flap, sampled larger ranges of phi-psi space in the trajectory of WT compared to Flap+. This result possibly indicates that the mutations at G48V and I54V are restricting the flexibility of the flap in Flap+.

Figure 3.4: Backbone dihedral angles of protease residues 2 to 98. Phi angles on the x-axis range from -180 to 180 degrees. Psi angles on the y-axis range from -180 to 180 degrees. (A) Wild-type protease. (B) Flap+ variant. (C) Residues have significant difference between WT and Flap+.

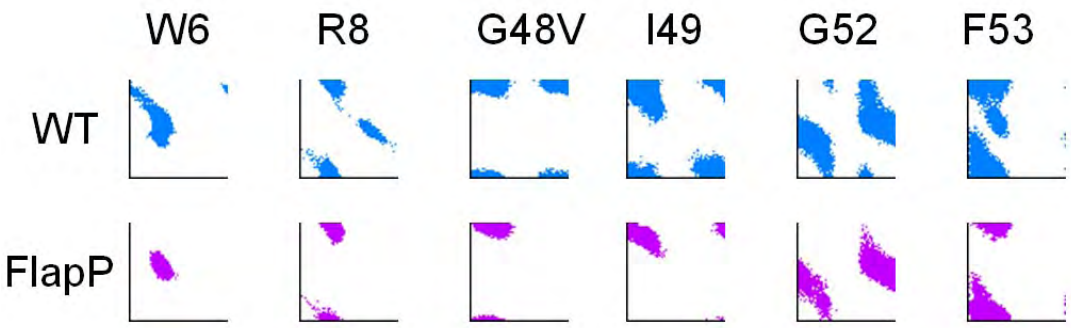
(A)



(B)



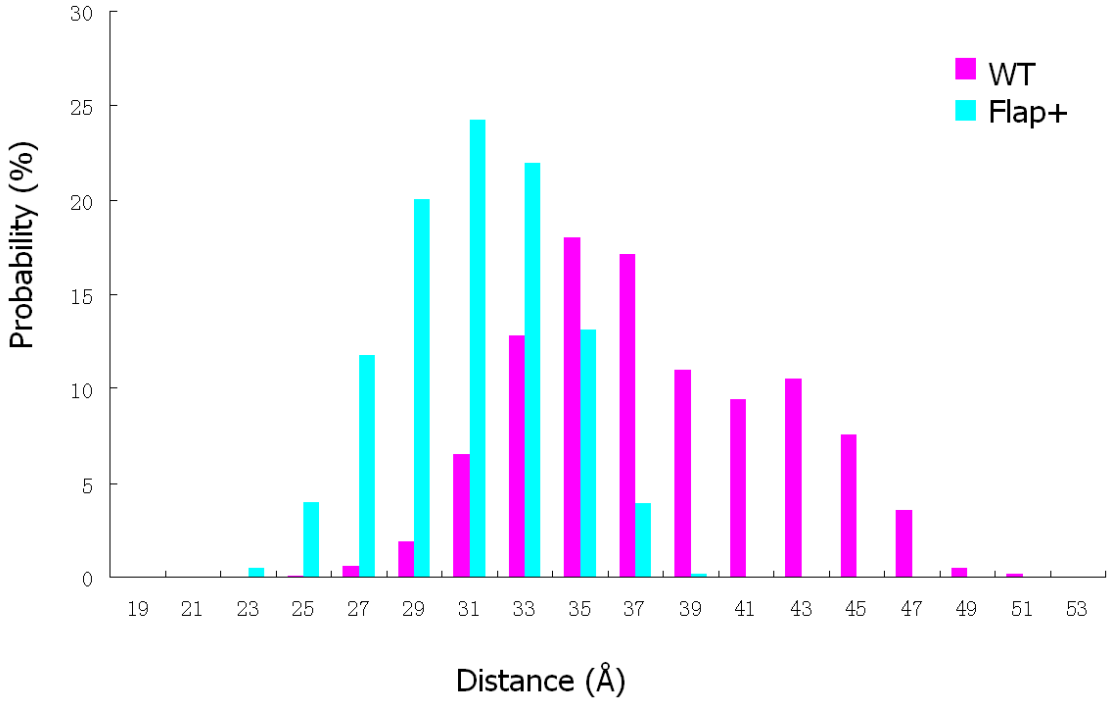
(C)



Distribution of distance between N-K55 and N-K55’

The distance between the Cα atom of residue 55 and the labeled nitrogen atom in previous studies [21,22] is similar to the distance between the nitrogen atom of the amino group in the original K55 side chain and its Cα atom, which is about 6 Å. The distance distribution of the two nitrogen atoms in the K55 side chain from the MD simulation can be compared with the results of previous studies, in which the inter-flap distance of apo HIV-1 protease was measured by site-directed spin labeling pulsed double electron-electron resonance (DEER) electron paramagnetic resonance spectroscopy[91,92]. In those studies, the protease was engineered with the Q7K/L33I/L63I mutations to decrease autoproteolysis and C67A/C95A to prevent inter-molecule disulfide bond formation. K55 was mutated to cysteine with a nitroxide spin-labeled moiety. Their results show that the distance between the two labeled nitrogen atoms ranges from 26 to 48 Å. Similarly, the distance of the two nitrogen atoms in the apo wild-type protease simulation ranges from 26 to 49 Å (Figure 3.5), and for the Flap+ variant, this distance ranges from 24 to 40 Å (Figure 3.5). The larger range for the wild-type protease indicates that its conformation ensemble in solution is more flap-open than that of the Flap+.

Figure 3.5: Distribution of distances between the nitrogen atoms in the amino group of the K55 side chain and N-K55' side chain. Distribution data from the wild-type protease simulation are in magenta, and from the Flap+ variant are in cyan.

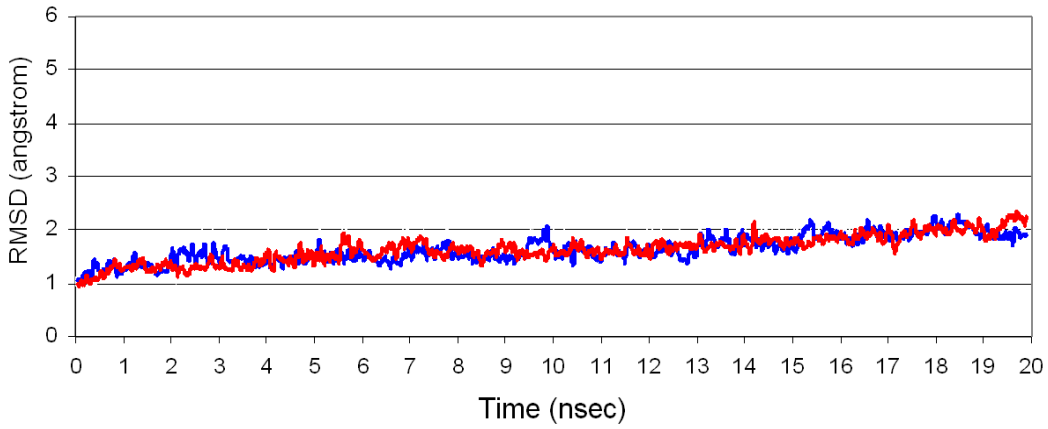


Dynamic properties of wild-type protease and Flap+ when binding with inhibitor

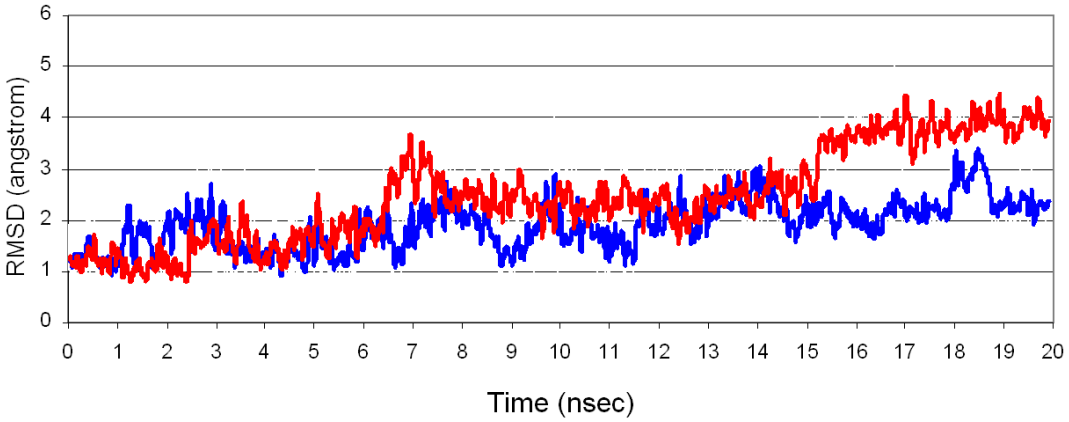
To understand the influence of binding with inhibitor on the dynamic properties of the wild-type and Flap+ variants, MD simulations of both proteins bound to the inhibitor, DRV, were performed and analyzed. When binding with DRV, both the wild-type protease and Flap+ are stable, with the same 1.64 Å average RMSD for all non-hydrogen atoms. (Figure 3.6) It is very interesting that the flap region of the DRV-Flap+ complex is a bit more mobile than the DRV-WT complex. The average RMSD of residues 45 to 55 for the WT protease is 1.91 Å and for the Flap+ is 2.45 Å, with average RMSD of residues 45' to 55' of 1.85 Å and for Flap+ of 2.28 Å. For the apo protease, the flap region of the WT protease is more flexible and mobile than the Flap+ according to the analysis above. The RMSF of the Cα atoms of the protease also supports this conclusion. In the apo form, the Cα atoms of the WT in the flap region are more fluctuated than the Flap+ variant (Figure 3.7A, B). When binding with DRV, the Cα atoms in the flap region are significantly decreased compared to their apo form. The atoms in the WT are slightly less fluctuated than the Flap+ variant. (Figure 3.7C, D) The simulated structure at the end of 20ns simulations of the DRV-Flap+ complex has deviated more extensively from its crystal structure than the DRV-WT complex. (Figure 3.8) The mutations of Flap+ affect protease dynamics destabilizing the DRV complex.

Figure 3.6: RMSD of protease atoms with respect to their corresponding crystal structures during the 20 ns MD simulations. Data from the DRV-WT simulation is in blue, and data from the DRV-Flap+ simulation is in red. (A) RMSD of all protease atoms except hydrogen atoms. (B) RMSD of atoms in residues 45 to 55. (C) RMSD of atoms in residues 45' to 55'.

(A)


RMSD (angstrom)
0
1
2
3
4
5
6
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Time (nsec)

(B)


RMSD (angstrom)
0
1
2
3
4
5
6
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Time (nsec)

(C)

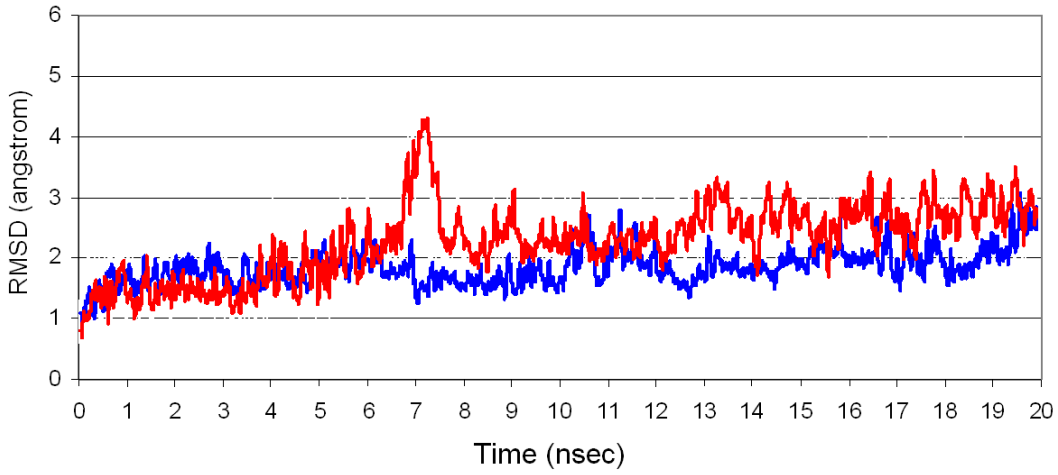
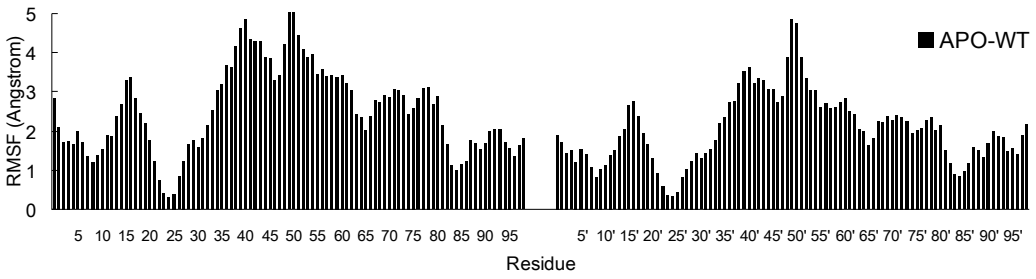
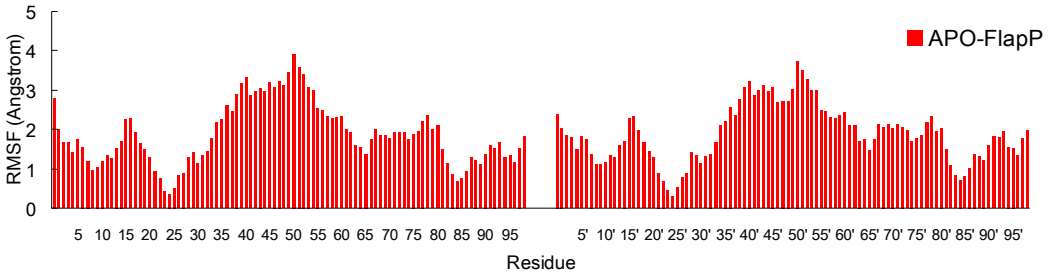

RMSD (angstrom)
0
1
2
3
4
5
6
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Time (nsec)

Figure 3.7: RMSF values of the Cα atoms for each residue in the MD simulation. (A) APO wildtype (B) APO Flap+ variant (C) DRV-WT complex (D) DRV-Flap+ complex

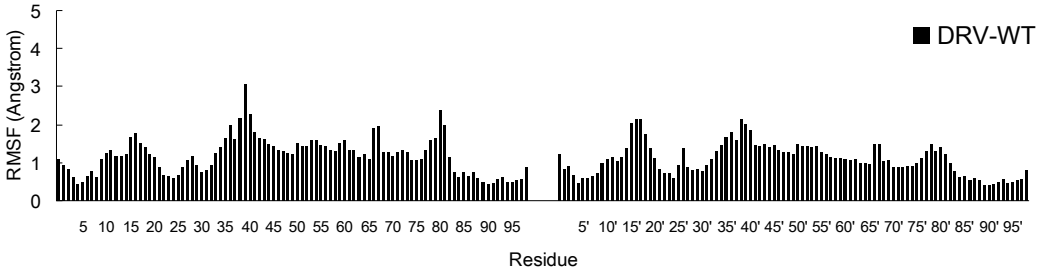
(A)


RMSF (Angstrom)
0 1 2 3 4 5
APO-WT
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 5' 10' 15' 20' 25' 30' 35' 40' 45' 50' 55' 60' 65' 70' 75' 80' 85' 90' 95'
Residue

(B)


RMSF (Angstrom)
0 1 2 3 4 5
APO-FlapP
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 5' 10' 15' 20' 25' 30' 35' 40' 45' 50' 55' 60' 65' 70' 75' 80' 85' 90' 95'
Residue

(C)


RMSF (Angstrom)
0 1 2 3 4 5
DRV-WT
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 5' 10' 15' 20' 25' 30' 35' 40' 45' 50' 55' 60' 65' 70' 75' 80' 85' 90' 95'
Residue

(D)

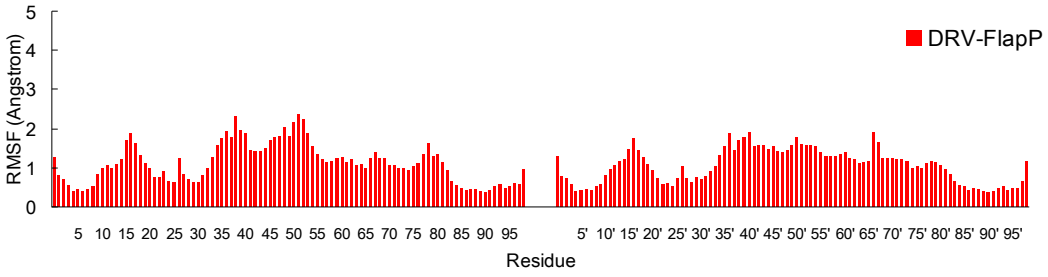
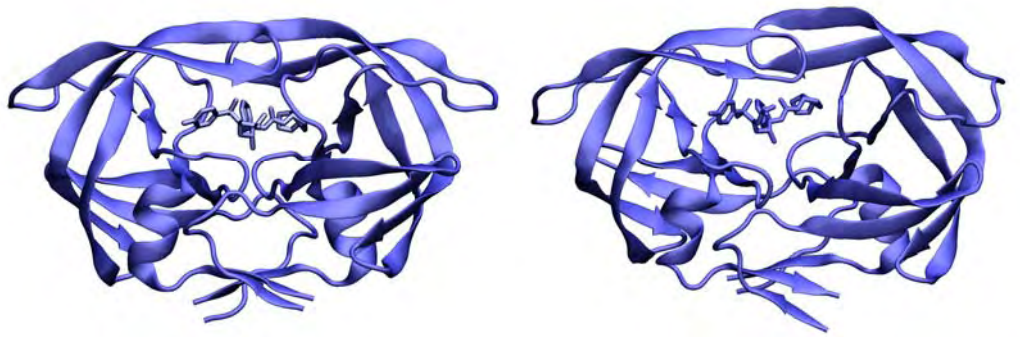
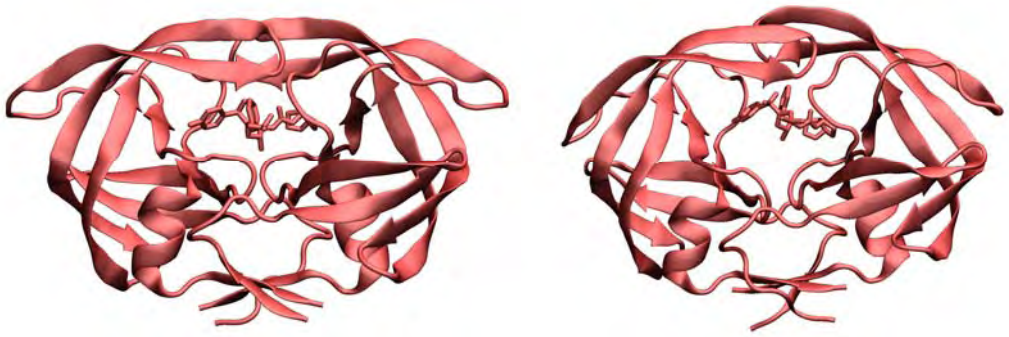

RMSF (Angstrom)
0 1 2 3 4 5
DRV-FlapP
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 5' 10' 15' 20' 25' 30' 35' 40' 45' 50' 55' 60' 65' 70' 75' 80' 85' 90' 95'
Residue

Figure 3.8: Crystal structures (left) of DRV-WT (A) and DRV-Flap+ (B) and their structures after 20ns MD simulations (right).

(A)



(B)



Summary

In this study, I performed molecular dynamics simulations to investigate and compare dynamic changes in the drug-resistant protease variant Flap+ to those of the wild-type protease. The flexibility of the protease flaps is important to enzyme activity since the flaps control access of the substrate to the protease active site. The results indicate that the wild-type apo-protease is more flexible than the Flap+ variant, especially in the flap region. However, when wild-type and the Flap+ protease bind with the inhibitor DRV, the DRV is more stable in the wild-type active site, despite the similar overall flexibility of the backbones of the two enzymes. Yet, when the flap dynamics are compared, the flaps are more flexible in the complex of Flap+ than in the wild-type complex. Although the binding free energy measured by calorimetry experiments shows that DRV binding with a higher affinity to the WT, the Flap+ exhibits a more favorable change in entropy. The value of $\Delta\Delta S$, the entropy change of Flap-DRV binding with respect to the entropy change of WT-DRV, is 45 cal/(mol·T). The MD simulation results indicate that the Flap+ variant binding with DRV compared to WT is likely the case of a less flexible apo protease binding with DRV and forming a more flexible, less optimal, complex. Assuming the level of disorder of a system is positively correlated with its entropy, the predicted entropy change of Flap+-DRV binding compared to WT-DRV binding from MD simulation result is larger in agreement with calorimetric results. Thus MD simulation provides

a rationalization for the experimentally observed thermodynamic data.

Chapter IV

Discussion

Abstract

In this chapter, the following topics, either as further analysis or future directions based on the Chapter II and Chapter III, are being discussed. A summary on this thesis is given at the end of this chapter.

Further discussion of calculations and decompositions of changes in protease-inhibitor binding free energy

1. Free-energy component analysis of APV binding with wild-type and drug-resistant variants and comparison with those of DRV.
2. Analysis of a representative conformer to elucidate changes in free-energy components of protease-DRV binding.
3. Comparison with the published results of free energy calculations on HIV-1 protease binding with protease inhibitors

NMR relaxation experiments on apo protease dynamics

1. Significance of doing NMR relaxation experiments to compare dynamic properties of apo wild-type protease and apo Flap+ protease variant
2. Necessity of introducing mutations L33I, L63I, C67A and C95A to prepare stable protease samples for NMR experiments – HSQC of original wild-type
3. HSQC of new construct WT_{NMR}; ITC experimental results of WT_{NMR} and $Flap+_{NMR}$

4. Preliminary results of NMR relaxation experiments on Flap+$_{NMR}$ and comparison with the published WT$_{NMR}$ data

Further discussion of calculations and decomposition of change in protease-inhibitor binding free energy

Free-energy component analysis of APV binding with wild-type and drug-resistant variants and comparison with those of DRV

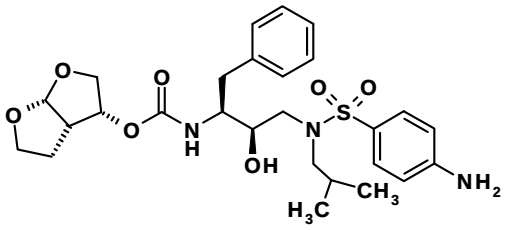
APV (amprenavir), which is the basis for the design of DRV (darunavir), differs from DRV on the single-ring tetrahydrofuranyl (THF) moiety with stereochemical change instead of bis-THF moiety on DRV (Figure 4.1). The bis-THF group of DRV forms strong interactions with protease residues D29 and D30 in the crystal structure of the DRV-protease complex [93], and these interactions are preserved in the Flap+ and ACT variants (Chapter II). The binding free-energy change between protease and APV is -12.6 kcal/mol for wild-type, -11.90 kcal/mol for Flap+, and -11.58 kcal/mol for ACT [93]. Crystallographic analysis [93] shows, for the ACT (V82I, I84V) variant, both APV and DRV lose favorable vdW contacts with residue 84, which might be the major reason of the loss in binding free energy with protease compared with wild-type. Another crystallographic study[87] comparing the crystal structure of DRV/APV complexed with the wild-type and Flap+ proteases shows that the Val to Ala mutation on residue 82 decreases the favorable vdW interaction between the protease and inhibitors. Comparing the binding of APV and DRV with protease will elucidate how these similar inhibitors respond to drug-resistant mutations, thus providing insights to benefit the design of new inhibitors with higher binding affinity and less susceptibility to drug-resistant mutations. Here I applied the same free-energy calculation and decomposition scheme described in Chapter II to APV-protease binding and compared the results with DRV-protease binding.

The calculated absolute binding free-energy changes of APV binding with WT, Flap+ and ACT are in good agreement with the experimental data. The predicted/experimental binding free energy change is -15.5/-12.6 kcal/mol for APV-WT, -12.5/-11.9 kcal/mol for APV-Flap+ and -12.1/-11.6 kcal/mol for APV-ACT. The contributions from different kinds of interactions to the absolute binding free-energy change and the relative binding free-energy change are similar to the distributions of the DRV-protease binding free-energy changes (Figure 4.2A, B). The contribution from vdW interactions is dominant in the total free-energy change (Figure 4.2C, D). The favorable Coulombic interaction energy is largely cancelled by the polar solvation energy.

In terms of each atom's contribution to the binding free-energy change, the DRV binding energy profiles with the two drug-resistant variants Flap+ and ACT are similar (Figure 4.3A), whereas APV has different profiles with the two variants (Figure 4.3B). Changes in vdW interaction energy for each moiety of the two inhibitors, their percentage loss with respect to the same atoms binding with WT (Chapter II), and their percentage loss with respect to all inhibitor atoms binding with WT are shown in Table 4.1. In the case of DRV, the loss of favorable vdW interaction energy for both drug-resistant variants is mainly on the 4-amino phenyl and isopropyl groups. In the case of APV binding with ACT, the energy profile is similar to DRV binding. In the case of APV binding with Flap+, the tetrahydrofuranyl and isopropyl groups have the most significant loss in vdW interaction energy.

Figure 4.1: DRV (A) and APV (B) chemical structures

(A)



(B)

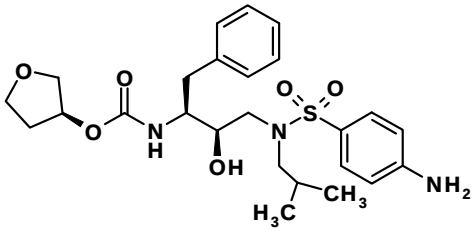
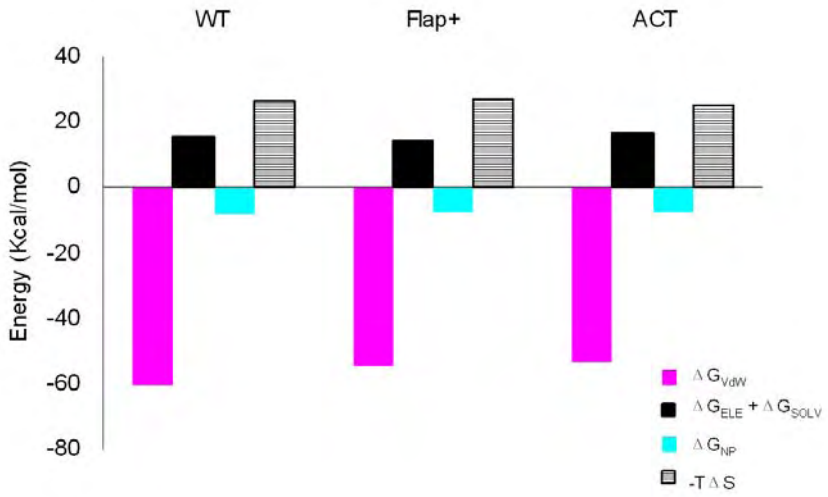
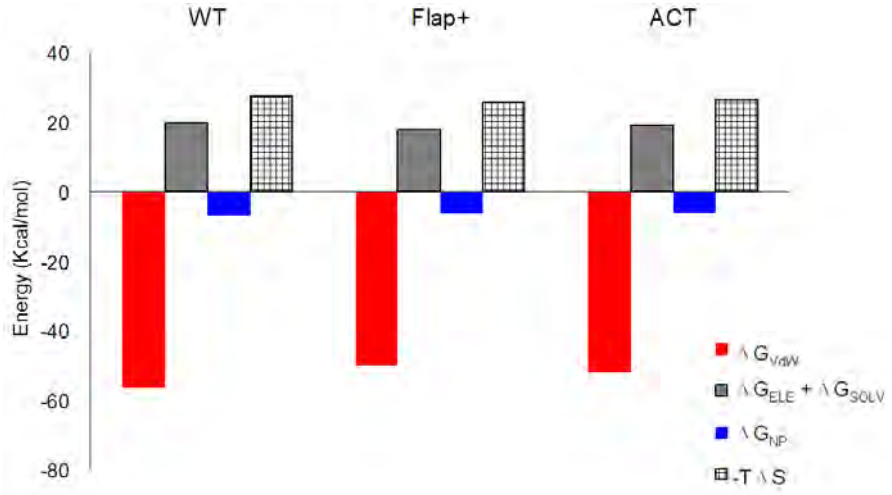


Figure 4.2: (A) Binding free-energy components of DRV-WT, DRV-FLAP+ and DRV-ACT. (B) Binding free-energy components of APV-WT, APV-FLAP+ and APV-ACT. (C) Loss of binding free-energy components of DRV with FLAP+ and ACT compared to that with WT protease. (D) Loss of binding free-energy components of APV with FLAP+ and ACT compared to that with WT protease.

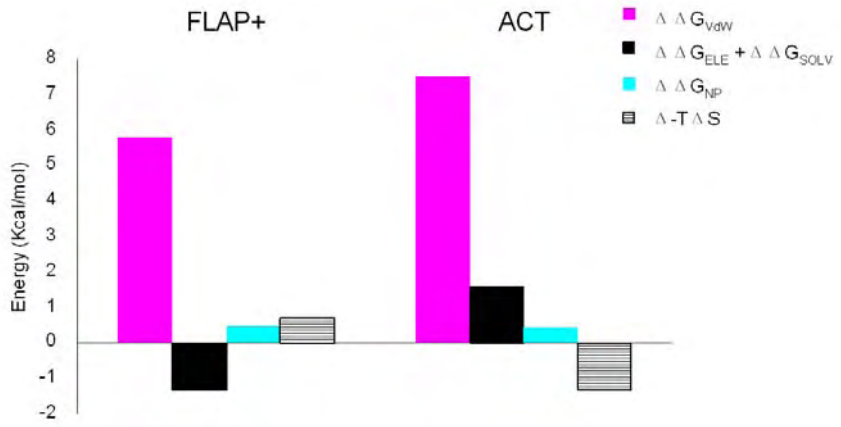
(A)



(B)



(C)


FLAP+
ACT
Δ Δ G VdW
Δ Δ G ELE + Δ Δ G SOLV
Δ Δ G NP
Δ -T Δ S
Energy (Kcal/mol)
8
7
6
5
4
3
2
1
0
-1
-2

(D)

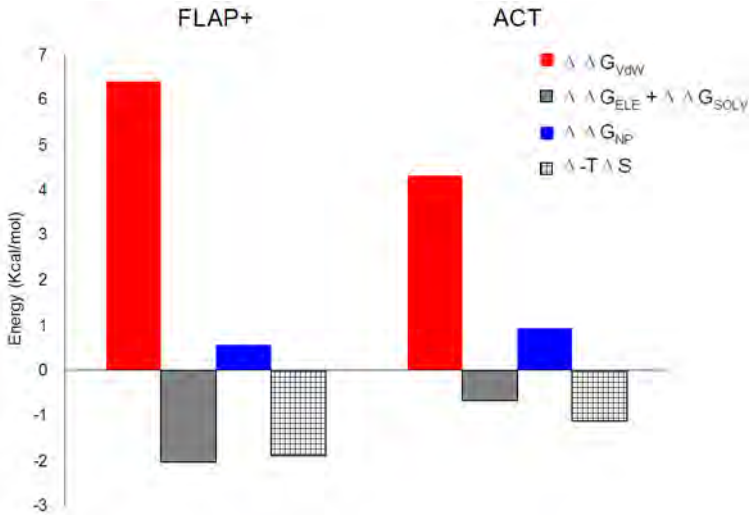
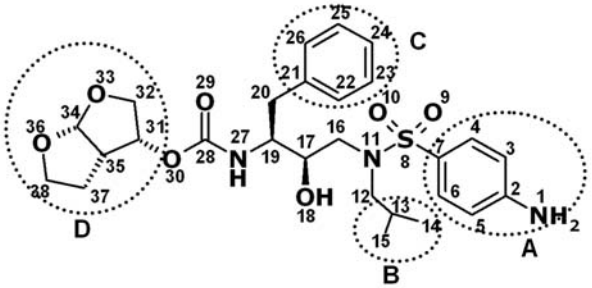
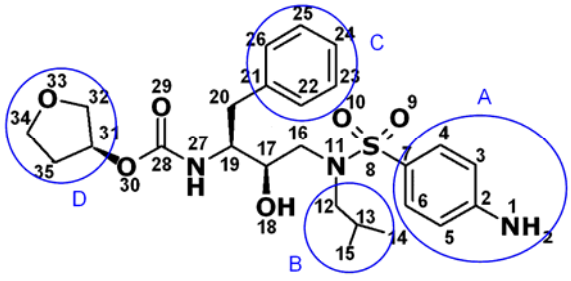

FLAP+
ACT
Δ Δ G VdW
Δ Δ G ELE + Δ Δ G SOLV
Δ Δ G NP
Δ -T Δ S
Energy (Kcal/mol)
7
6
5
4
3
2
1
0
-1
-2
-3

Figure 4.3: (A) The four moieties of DRV. (B) The four moieties of APV. (C) vdW interaction energy for each heavy atom of DRV with protease. Energy changes of DRV-FLAP+ with respect to the WT are shown in magenta and energy changes of DRV-ACT with respect to the WT are shown in cyan. (D) vdW interaction energy for each heavy atom of APV with protease. Energy changes of APV-FLAP+ with respect to the WT are shown in red and energy changes of APV-ACT with respect to the WT are shown in blue.

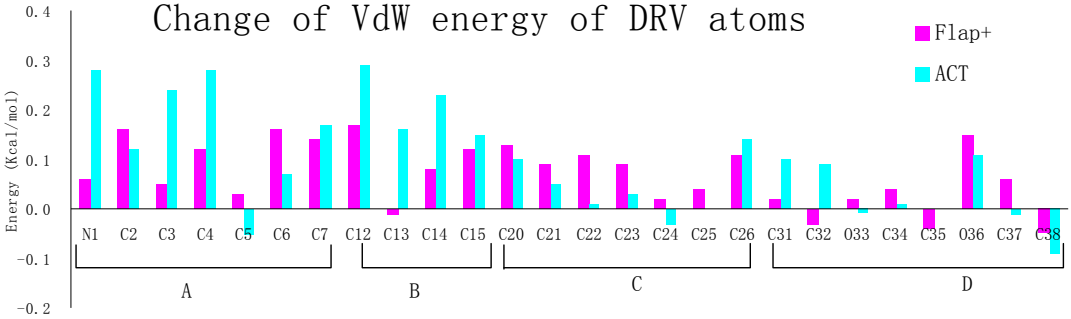
(A)



(B)



(C)


Change of VdW energy of DRV atoms
Flap+
ACT
Energy (Kcal/mol)
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
N1 C2 C3 C4 C5 C6 C7 C12 C13 C14 C15 C20 C21 C22 C23 C24 C25 C26 C31 C32 O33 C34 C35 O36 C37 C38
A
B
C
D

(D)

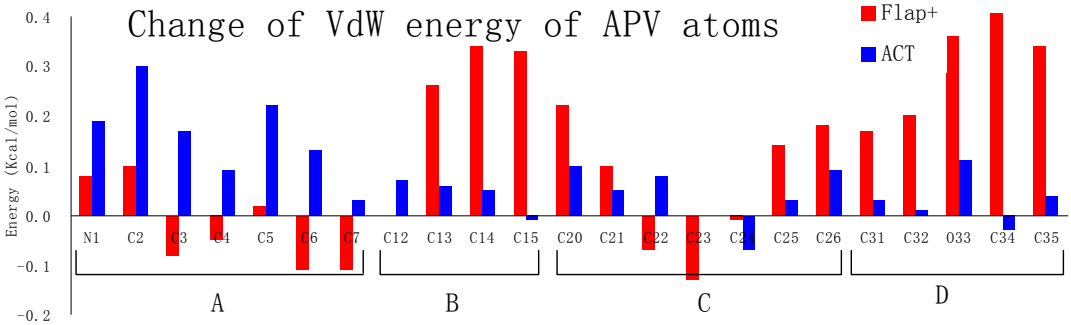

Change of VdW energy of APV atoms
Flap+
ACT
Energy (Kcal/mol)
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
N1 C2 C3 C4 C5 C6 C7 C12 C13 C14 C15 C20 C21 C22 C23 C24 C25 C26 C31 C32 O33 C34 C35
A
B
C
D

Table 4.1: Loss of Van der Waals' Interaction Energy for Different Moieties of DRV and APV

DRV		4–Amino Phenyl Group	Isopropyl Group	Benzyl Ring	*bis*-Tetrahydrofuranyl
DRV-Flap+	kcal/mol*	0.72	0.36	0.59	0.17
	%**	12.3	12.1	12.8	2.7
	%WT***	3.7	1.8	3.0	0.9
DRV-ACT	kcal/mol	1.11	0.83	0.30	0.20
	%	18.9	28.0	6.5	3.2
	%WT	5.6	4.2	1.5	1.0
APV		4–Amino Phenyl Group	Isopropyl Group	Benzyl Ring	Tetrahydrofuranyl
APV-Flap+	kcal/mol	-0.15	0.93	0.43	1.52
	%	-3.2	33.2	9.2	31.7
	%WT	-0.9	5.5	2.5	8.9
APV-ACT	kcal/mol	1.13	0.17	0.28	0.16
	%	23.9	6.1	6.0	3.3
	%WT	6.6	1.0	1.7	0.9

* The energy change compared to the same atoms binding with WT

** Percentage of energy loss with respect to atoms of the same moiety binding with WT

*** Percentage of energy loss with respect to all inhibitor atoms binding with WT

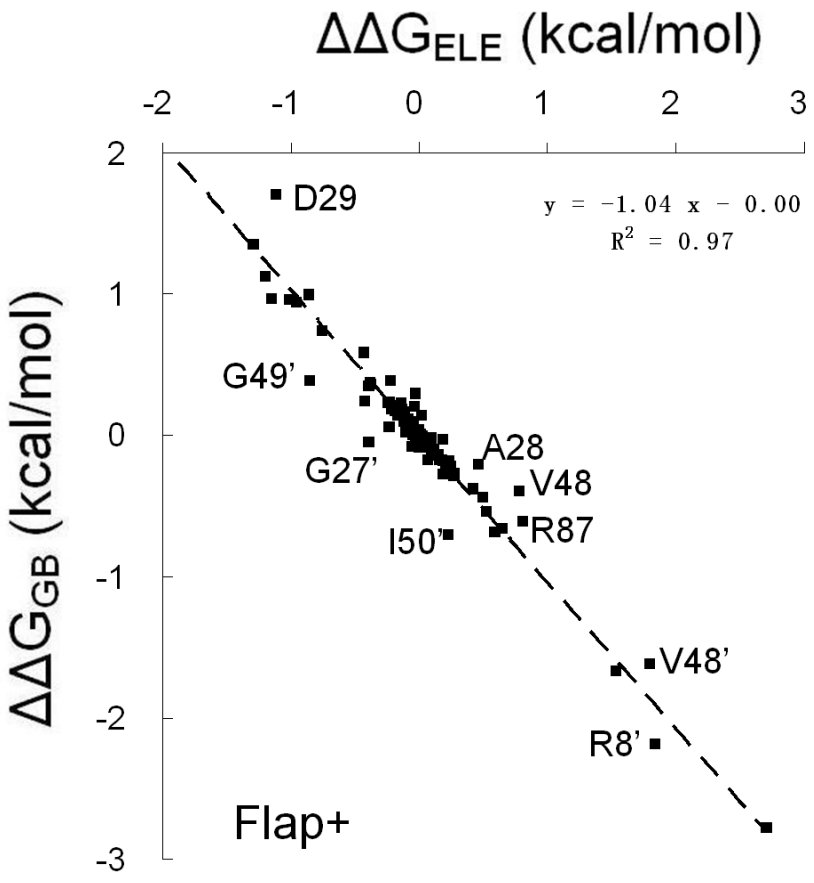
The contribution of charge interactions to APV-protease binding and DRV-protease binding

In Chapter II I showed that, compared to WT protease binding with DRV, for each residue of the drug-resistant variants, changes in Coulomb interaction energy ($\Delta\Delta G_{ELE}$) and changes in solvation energy ($\Delta\Delta G_{GB}$) are highly correlated (Figure 4.4A, B). The linear regression analysis of $\Delta\Delta G_{GB}$ with respect to $\Delta\Delta G_{ELE}$ shows their correlation coefficient r is -0.98 for Flap+ and -0.97 for ACT. The same high correlation applies to the binding of these two variants with APV. (Figure 4.4C, D). The correlation coefficient r is -0.95 for Flap+ and -0.94 for ACT. The high negative correlation between the $\Delta\Delta G_{GB}$ and $\Delta\Delta G_{ELE}$ partially explains the how the change of total charge interaction energy become a minor factor in the total change of binding free energy compared to the vdW interactions. Interestingly, a previous free energy component analysis shows that, compared to the total predicted binding free energy, the predicted contribution from electrostatic interaction has a higher correlation with the experimental binding free energy.[94] Recently a similar analysis from the same group on interactions between HIV protease and inhibitors yields the conclusions that the total theoretical binding energy is in better agreement with the experimental data, although the free energy component from charged interactions itself is also in good correlation with experimental binding free energy.[95] This difference might have resulted from, as the author pointed out, the different environments of the two systems.[95] The former calculation is on the large solvated protein surface while the

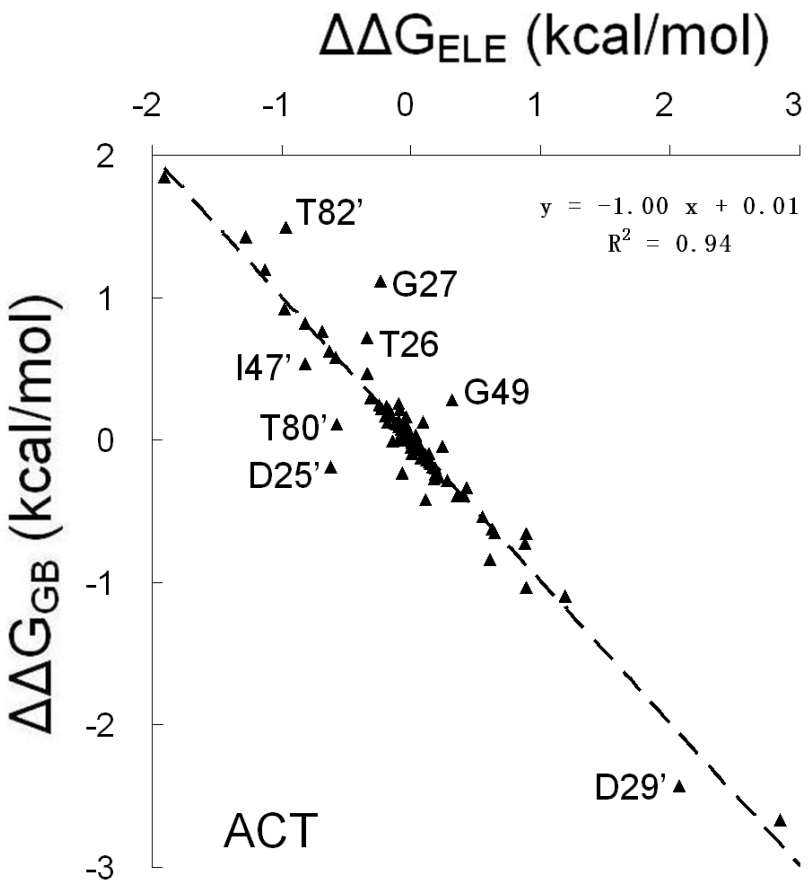
latter calculation on HIV protease is on a relatively small and buried binding pocket. In attempt to find those residues which have a significant charge interaction energy change, a cutoff of 0.2 kcal/mol is set for the change in total charge interaction energy (sum of $\Delta\Delta G_{GB}$ and $\Delta\Delta G_{ELE}$). Residues with a change in total charge interaction energy over 0.2 kcal/mol are labeled in Figure 4.4. The change of the charge interaction energy differs in detail for the binding of DRV and APV with Flap+ and ACT (Table 4.2). A further investigation on how these residues changed their coulombic interactions with the inhibitors or how they changed the solvation energy upon binding with inhibitors will be valuable to improving the design of inhibitors with optimal charge distribution to maximize the binding energy of charge interactions with protease.

Figure 4.4: Correlation between $\Delta\Delta G_{ELE}$ and $\Delta\Delta G_{GB}$ of each residue. The highlighted residues are with total charge energy change ($\Delta\Delta G_{ELE}+\Delta\Delta G_{GB}$) larger than 0.2kcal/mol or less than -0.2kcal/mol. (A) Energy difference between DRV-FLAP+ and DRV-WT. (B) Energy difference between DRV-ACT and DRV-WT. (C) Energy difference between DRV-FLAP+ and DRV-WT. (D) Energy difference between DRV-ACT and DRV-WT.

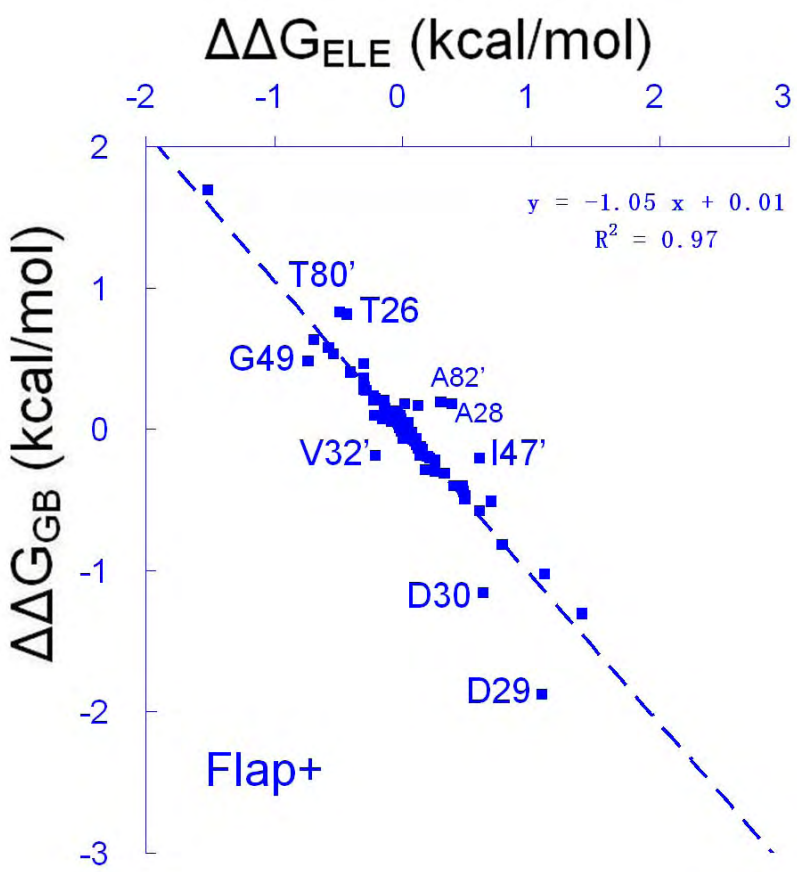
(A)



(B)


ΔΔG_ELE (kcal/mol)
-2
-1
0
1
2
3
2
1
0
-1
-2
-3
ΔΔG_GB (kcal/mol)
y = −1.00 x + 0.01
R2 = 0.94
T82'
G27
T26
I47'
G49
T80'
D25'
D29'
ACT

(C)


ΔΔG_ELE (kcal/mol)
-2
-1
0
1
2
3
2
1
0
-1
-2
-3
ΔΔG_GB (kcal/mol)
y = −1.05 x + 0.01
R^2 = 0.97
T80'
T26
G49
A82'
A28
V32'
I47'
D30
D29
Flap+

(D)

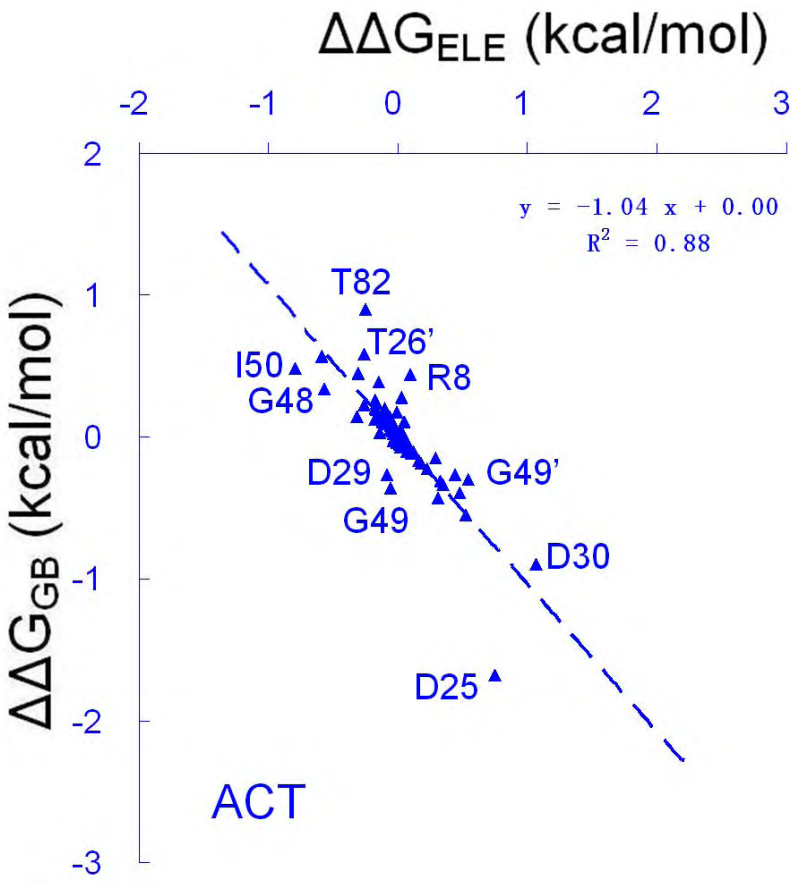

ΔΔG_ELE (kcal/mol)
-2
-1
0
1
2
3
2
1
0
-1
-2
-3
ΔΔG_GB (kcal/mol)
y = −1.04 x + 0.00
R2 = 0.88
T82
T26'
I50
R8
G48
D29
G49'
G49
D30
D25
ACT

Table 4.2: Residues with change in charge interaction energy larger than 0.2 kcal/mol in drug-resistant variants Flap+ and ACT

	$\Delta\Delta G_{Charge} > 0.2$			$\Delta\Delta G_{Charge} < 0.2$		
	$\Delta\Delta G_{ELE} > 0$ $\Delta\Delta G_{GB} > 0$	$\Delta\Delta G_{ELE} > 0$ $\Delta\Delta G_{GB} < 0$	$\Delta\Delta G_{ELE} < 0$ $\Delta\Delta G_{GB} > 0$	$\Delta\Delta G_{ELE} < 0$ $\Delta\Delta G_{GB} < 0$	$\Delta\Delta G_{ELE} > 0$ $\Delta\Delta G_{GB} < 0$	$\Delta\Delta G_{ELE} < 0$ $\Delta\Delta G_{GB} > 0$
DRV-Flap+		A28, V48, R87, V48’	D29	G27’	I50’, R8’	G49’
APV-Flap+	A28, A82’	I47’	T80’, T26’	V32’	D29, D30	G49
DRV-ACT	G49		T82’, G27, T26	D25’	D29’	I47’, T80’
APV-ACT	R8	G49’, D30	T82, T26’,	D29, G49	D25	I50, G48

Using a representative conformer to elucidate the free-energy change of Flap+ and ACT binding with DRV

The MM-GBSA free-energy simulation method and the following free-energy components analysis are based on conformers of the protease-inhibitor complex sampled during molecular dynamic simulations. The calculated result represents an ensemble average of the thermodynamic properties of the system. The average structure of the conformers from MD simulations reflects the most frequently sampled positions of each atom of the system. From the free-energy decomposition by protease residue, one can observe that most of the protease residues that contribute to binding with inhibitors, and have a significant free-energy change on drug-resistant variants, are structurally close to the inhibitor binding site. (Chapter II) A representative conformer which can be used to elucidate the calculated energetic properties should be able to mostly represent the ensemble of sampled positions of these protease atoms. For each trajectory of the MD simulation of protease-inhibitors complex, an average structure was generated. Each frame of the trajectory is superimposed on the average structure by the proteases atoms within 7 Å of any inhibitor atom. The frame with the lowest RMSD was chosen as a representative structure to illuminate changes in binding free energy between the drug-resistant variants and wild-type protease. In Figure 4.5, the protease atoms which were within 7 Å of atoms of certain inhibitor moieties are shown.

For the WT-DRV complex, the 4-amino phenyl group of DRV is surrounded by protease residues G27’ to V32’, I47’ to G49’, I84’ and I50. (Figure 4.5A) Some side-chain atoms from V82’, V82, L76’ and I85’ also significantly contribute to the vdW interactions with this group. In the Flap+ and ACT cases, the vdW interactions are maintained between this group and the protease residues located in the flap region. Furthermore, the larger side chain from mutation G48’V in the Flap+ variant is not involved in interactions with this part of DRV. In the DRV-Flap+ complex, I84’ maintains vdW contacts while A82 and A82’ do not. The loss of vdW interaction energy is mainly in the G27’ to V32’ region of Flap+. In the DRV-ACT complex, the G27’ to V32’ region maintains its vdW interactions with the 4-amino phenyl group of DRV. The major loss is at position 84’, which is Ile in the WT protease and Val in the ACT variant.

In the DRV-WT complex, the isopropyl group of DRV is flanked by the two catalytic aspartic acids D25/D25’ and the adjacent residues from T26’ to D29’, flap residues G48’ to I50’, and the P1 loop residues V82 and I84. Mutation V82A in Flap+ and I84V in ACT have a similar impact on this isopropyl group. A smaller side chain leads to the isopropyl group sliding towards the P1 group, and losing its vdW interactions with D25’ to D29’.

The major difference between the benzyl ring (group C) in DRV-WT and DRV-Flap+ is at position 82. The loss of favorable vdW contact due to Ala in the Flap+ being a smaller side chain than Val in the WT protease cannot be compensated

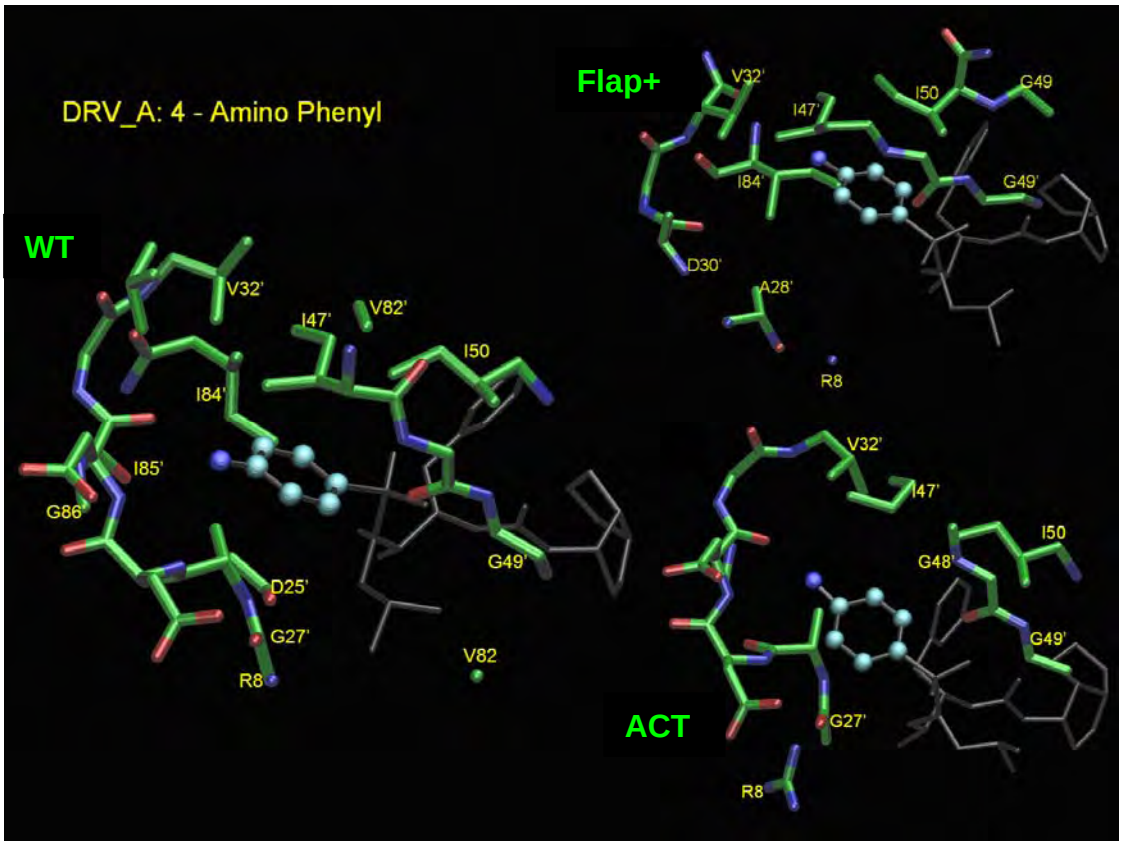
by adding a vdW contact from G48V in the DRV-Flap+ complex.

Among the 4 moieties of DRV, the bis-THF group is the one experiencing the least vdW interaction energy loss. This group is well packed with the residues R8', I50', I/V84, G27 to V32 and I47 to G49. These key residues maintain their interactions with DRV in the drug-resistant variants Flap+ and ACT.

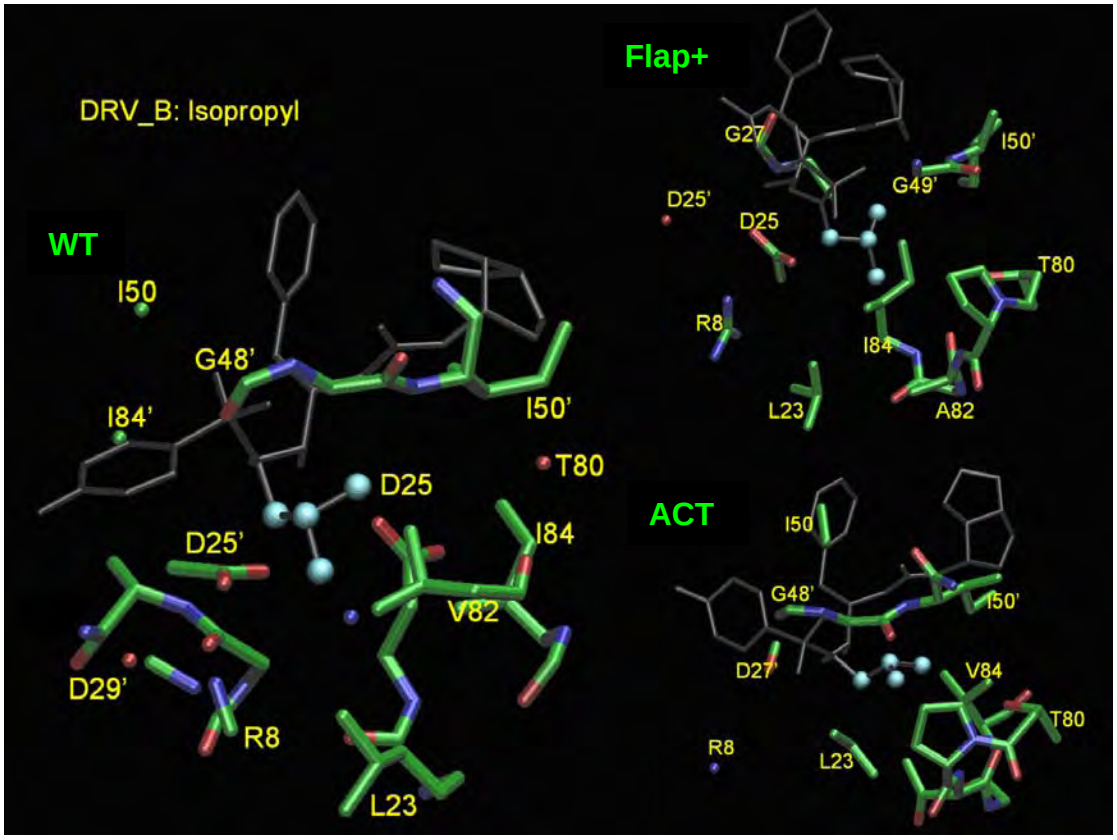
From these observations and the free-energy decomposition data shown in Table 4.1, one can conclude that the 4–amino phenyl and isopropyl groups are susceptible to P1 loop mutations (V82A/T, I84V), and the benzyl ring is more sensitive to flap mutations. The *bis*-tetrahydrofuranyl group is the least susceptible part of DRV to drug-resistant mutations.

Figure 4.5: Protease atoms making major contributions to vdW interactions with different moieties of DRV. (A) The 4-amino phenyl group. (B) The isopropyl group. (C) the benzyl ring. (D) the bis-tetrahydrofuranyl ring.

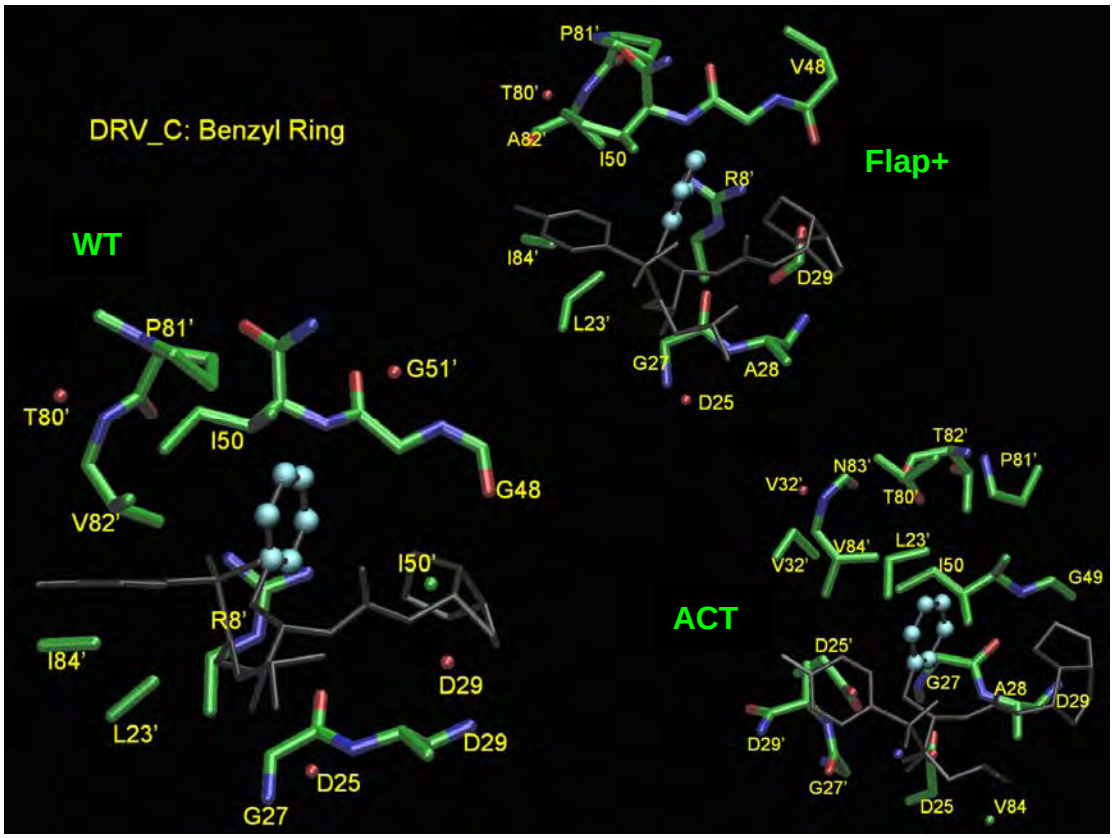
(A)



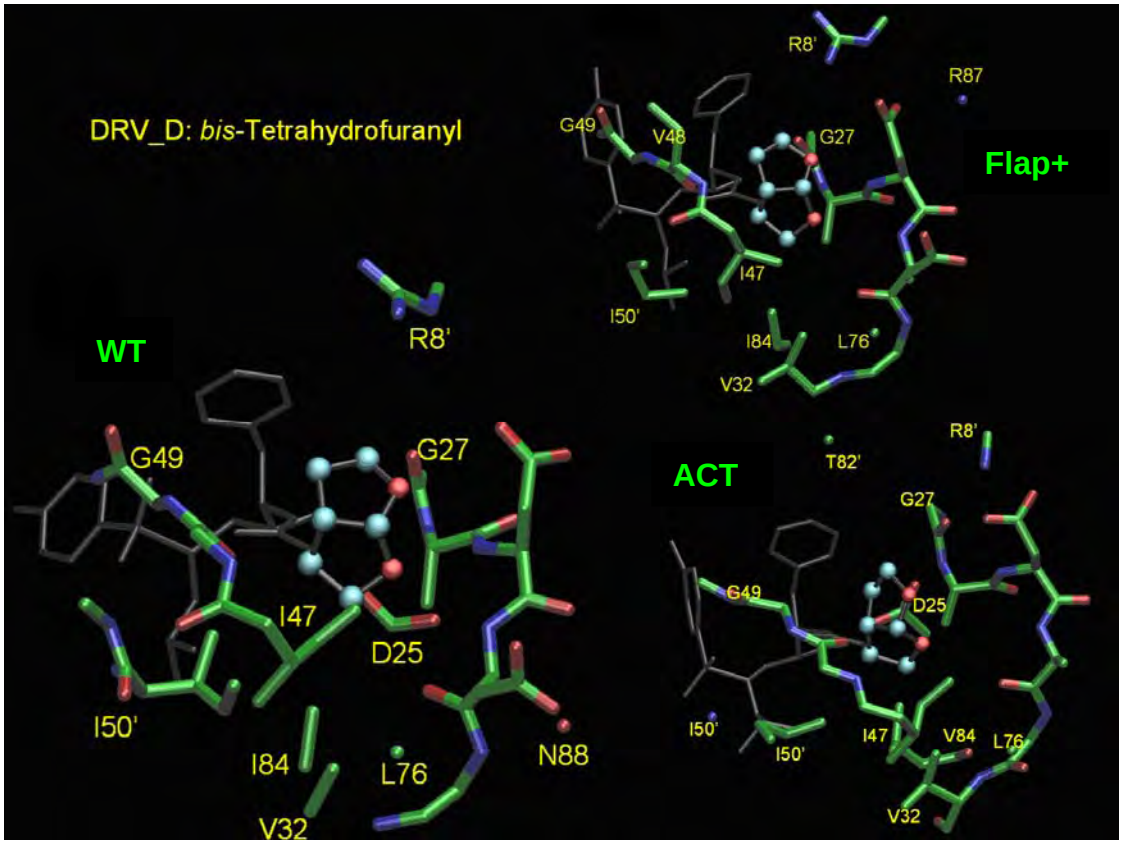
(B)


DRV_B: Isopropyl
WT
I50
G48'
I84'
I50'
T80
D25
D25'
I84
V82
D29'
R8
L23
Flap+
G27
I50'
G49'
D25'
D25
T80
R8
I84
L23
A82
ACT
I50
G48'
I50'
D27'
V84
T80
R8
L23

(C)


DRV_C: Benzyl Ring
WT
Flap+
ACT
P81'
T80'
A82'
I50
V48
R8'
I84'
D29
L23'
G27
A28
D25
G51'
G48
V82'
I50'
T82'
N83'
V32'
T80'
V84'
G49
D25'
D29'
G27'
V84

(D)


DRV_D: bis-Tetrahydrofuranyl
R8'
R87
G49
V48
G27
Flap+
I47
I50'
R8'
I84
L76
V32
WT
G49
G27
I47
D25
I50'
I84
L76
N88
V32
ACT
T82'
R8'
G27
G49
D25
I50'
I47
I50'
V84
L76
V32

NMR relaxation experiments on apo protease dynamics

Significance of doing NMR relaxation experiments to compare dynamic properties of apo wild-type protease with the apo Flap+ protease variant

The interesting high entropy-enthalpy compensation of the Flap+ protease variant binding with DRV and APV[87] could not be elucidated by crystallographic studies[87] or free-energy simulation methods. Flap dynamics is critical to protease function as shown in NMR relaxation experiments [30,31,88], molecular dynamics simulations[29,96], and site-directed spin labeling experiments[91,92,97]. The flap mutations G48V and I54V were shown by theoretical studies (Chapter III) to greatly influence flap and even protease dynamics. No experiments have directly shown how much influence these mutations have on protease dynamic properties. In order to understand the mechanism of the dynamic properties difference between the WT protease and the Flap+ protease, NMR relaxation experiments were employed to study the apo Flap+ protease dynamics behavior in solution.

^{15}N labeled protease expression, purification and refolding

HIV-1 protease was expressed in *Escherichia coli* strain *BL21-Gold*(*DE3*)*pLysS* competent cell using a T7 expression system. The bacteria were glown in LB or TB(Table 4.3) at 37°C. When OD_{600} value of the culture reached 0.4, the bacteria were separated from the LB culture by centrifugation. The pellet was resuspended in 1L cold (4°C-10°C) wash buffer (Table 4.4). The bacteria were separated from the wash buffer by centrifugation and resuspended with 250 mL ^{15}N minimal medium (Table 4.5). The M9 bacterial culture was glown at 37°C for 10 to 20 min. Afterwards protease expression is induced with 2 mM IPTG for 4 hours. (Figure 4.6) After induction, the culture was centrifuged and the bacteria pellet was separated and stocked at -80°C. The protease was contained in the inclusion body in the bacteria pellet. After the bacteria was lysed, the inclusion bodies were isolated by centrifugation and the pellet was dissolved in 50% acetic acid to extract protease. Proteins with different molecular weight were separated by size exclusion chromatography on a 2.1 L Sephadex G-75 superfine column. Refolding was accomplished by diluting the protease solution into a 100-fold excess of refolding buffer. Excess acetic acid was removed through dialysis.

Table 4.3 Recipe for 1 L LB medium (autoclaved)

Tryptone	10 g
Yeast Extract	5 g
NaCl	5 g
1M NaOH	1 ml

Table 4.4 Recipe for 1 L of wash buffer (autoclaved)

KH_2PO_4	3.0 g
$Na_2HPO_4 \cdot 7H_2O$	12.8 g
NaCl	0.5 g
1 M $MgSO_4$(autoclaved)	2.0 ml
1 M $CaCl_2$(autoclaved)	0.1 ml
pH adjusted to 7.2 with NaOH	

Table 4.5 Recipe for 1 L of minimal medium and 1 L of 5 X M9 Salts (autoclaved)

Minimal medium

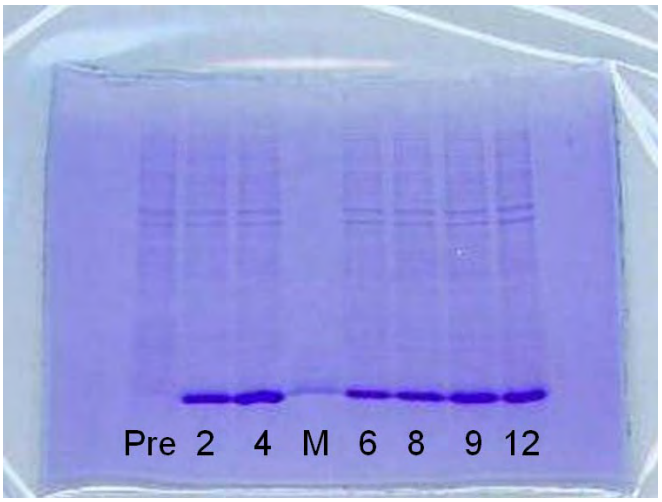
5 X M9 Salts	200.0 ml
D-glucose Stock (20 g/100 ml)	20.0 ml
1 M $MgSO_4$(autoclaved)	2.0 ml
1 M $CaCl_2$(autoclaved)	0.1 ml

5 X M9 Salts (autoclaved)

NH_4Cl	5.0 g
KH_2PO_4	15.0 g
$Na_2HPO_4 \cdot 7H_2O$	64.0 g
NaCl	2.5 g

pH adjusted to 7.2 with NaOH

Figure 4.6 ^{15}N labeled protease expression samples on 16% SDS PAGE. Pre: sample before IPTG induction. M: marker for HIV-1 protease. 2, 4, 6, 8, 9, 12: expression samples after different hours IPTG induction.

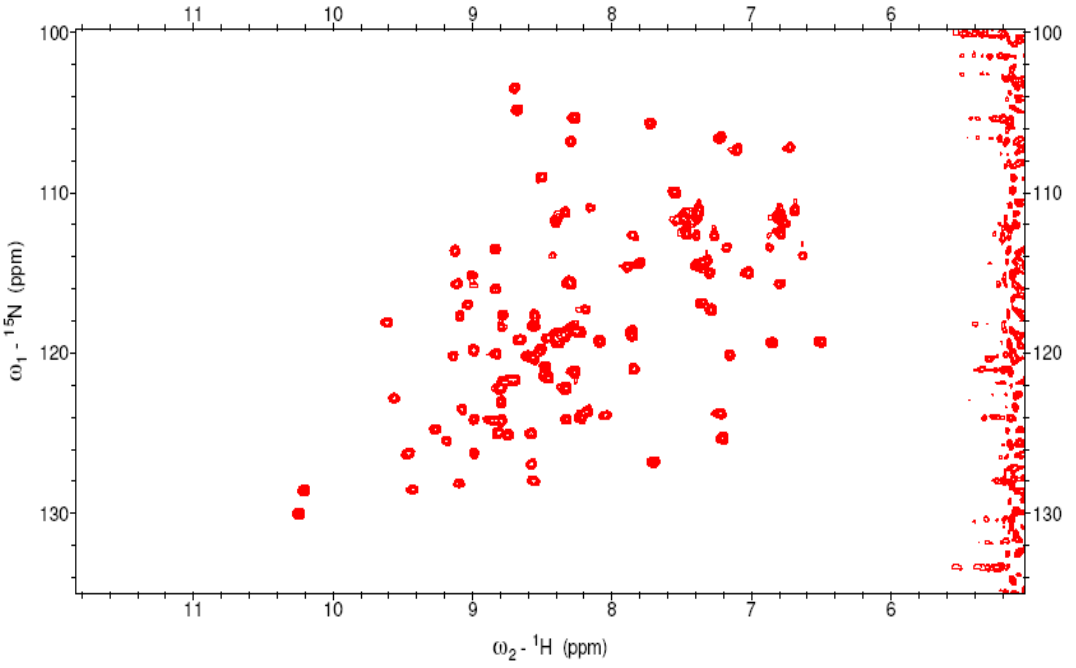


Necessity of introducing mutations to prepare stable protease samples for NMR experiments – HSQC of original wild-type

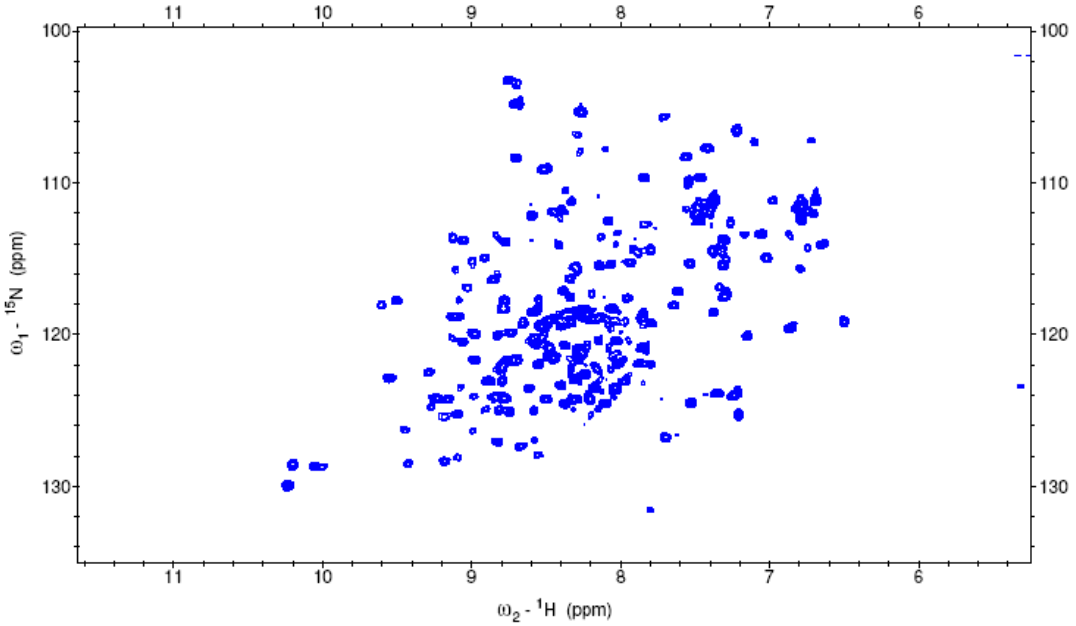
In studies of the active form of HIV-1 protease, a critical problem that has to be solved in preparing protease samples for NMR is the autoproteolysis. HIV protease with the Q7K mutation, which is supposed to decrease self-cleavage[98], still shows high levels of autoproteolysis in ^{15}N-^{1}H HSQC spectra (Figures 4.7A and 4.7B). The sample contained 67 μM Q7K, 5% D2O, 5mM DTT in 20mM sodium phosphate buffer.[30,31,99,100] The pH was adjusted to 5.8.[30,31,99,100] The HSQC spectrum was obtained using a 700 MHz magnet at 293 K. The first spectrum was taken on August 1, 2005 and the second on August 4, 2005 by J. Peng in University of Notre Dame.

Figure 4.7: ^{15}N-^{1}H HSQC spectra of Q7K protease. (A) Spectra recorded on Aug 1st, 2005. (B) Spectra recorded on Aug 4th, 2005.

(A)



(B)

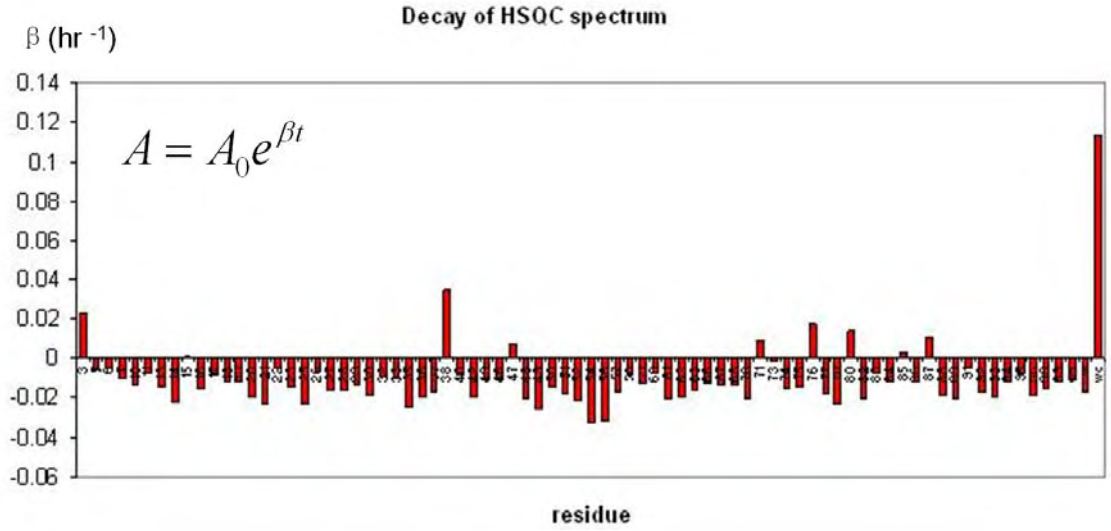


To evaluate the relative velocity of degradation of the signals, HSQC spectra were recorded from August 5 to August 8. The amplitudes of the major peaks from the HSQC spectrum were integrated and simulated the decay with respect to time using the following formula:

$$A = A_0 e^{\beta t}$$

A is the integrated amplitude of a potential peak for a residue (based on the chemical shift). Setting the time that the first spectrum is taken as 0 h, the other time points are 75, 97, 119, 145, and 165 h.

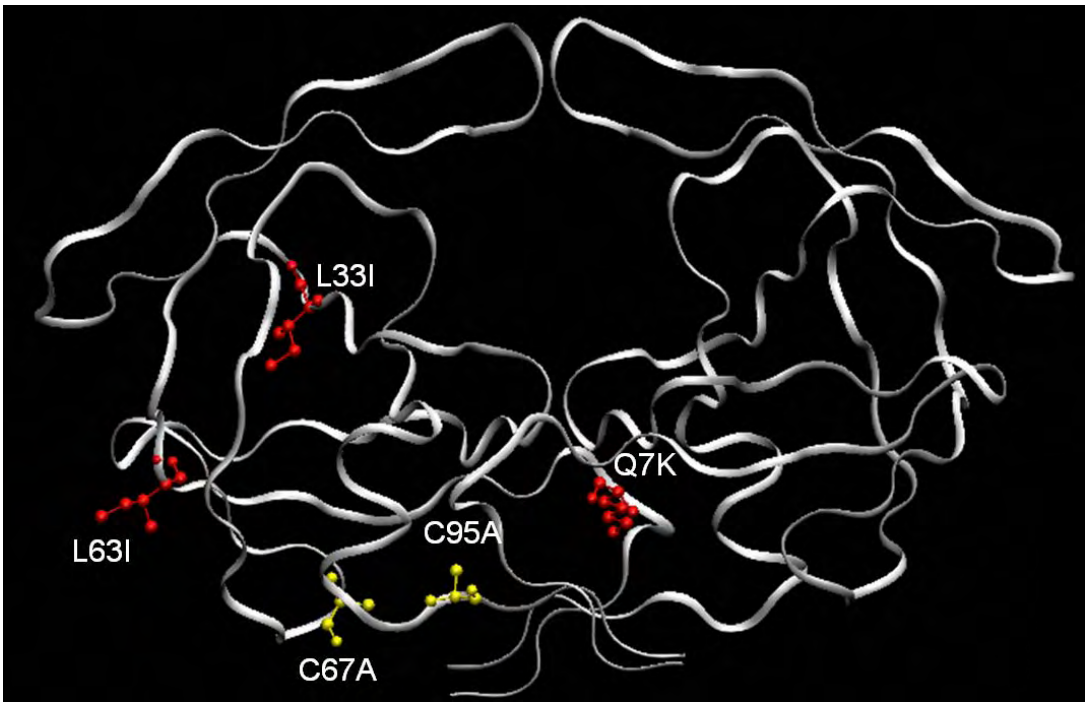
Figure 4.8: The simulated decay of HSQC spectrum of Q7K.



Most of the peaks have degrading amplitudes ($\beta<0$). (Figure 4.8) To improve stability of the protein, 2 mutations were introduced: L33I and L63I. These two mutations have been shown to significantly increase protease stability and retain its enzyme activity [98,101]. Two cysteine residues of the protease at positions 67 and 95 were mutated to alanines to prevent potential formation of intermolecular disulfide bonds. These mutations have frequently been employed in NMR research to study protease dynamics [30,31,88,99,100].

In the following discussion, variant WT_{NMR} refers to protease Q7K, L33I, L63I, C67A and C95A (Figure 4.9).Variant $Flap+_{NMR}$ refers to protease Q7K, L33I, L63I, C67A, C95A, L10I, G48V, I54V and V82A. The plasmid of WT_{NMR} was prepared by Dr. Jennifer Foulkes-Murzycki and the plasmid of $Flap+_{NMR}$ was prepared by Ellen Nalivaika.

Figure 4.9: Structure of WT_{NMR} protease variant with the new mutations highlighted



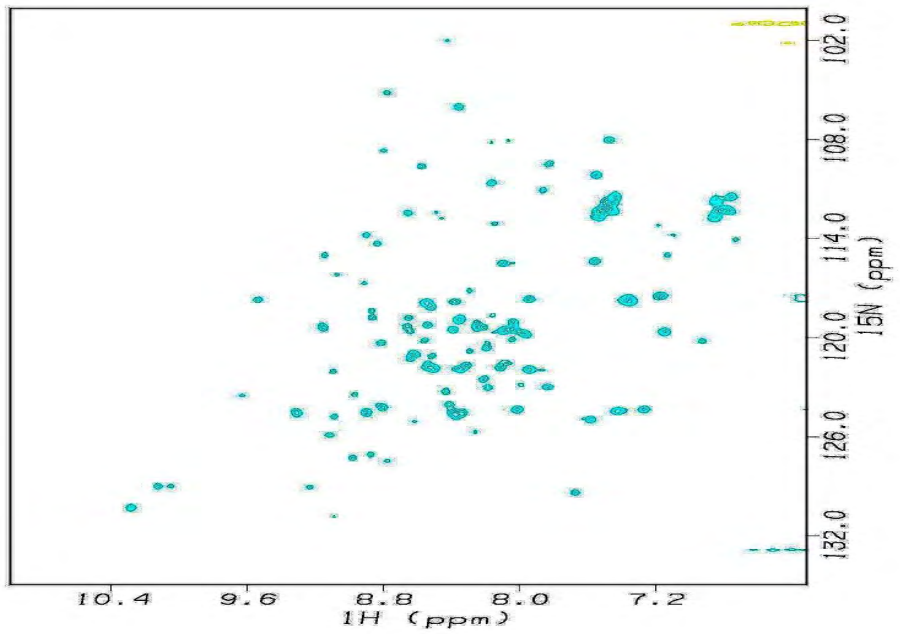
HSQC experiments on the newly constructed protease WT_{NMR}

The HSQC spectra for a sample of WT_{NMR} are shown in Figures 4.10. The sample contained 100 μM WT_{NMR} and 5% D2O in 20 mM sodium phosphate buffer at pH 5.8.[30,31,99,100] The HSQC spectrum was measured on a 600 MHz magnetic at 293 K. The first spectrum was obtained on Feb 15, 2006 and the second on Feb 21, 2006 by R. Vadrevu in University of Massachusetts, medical school in Worcester.

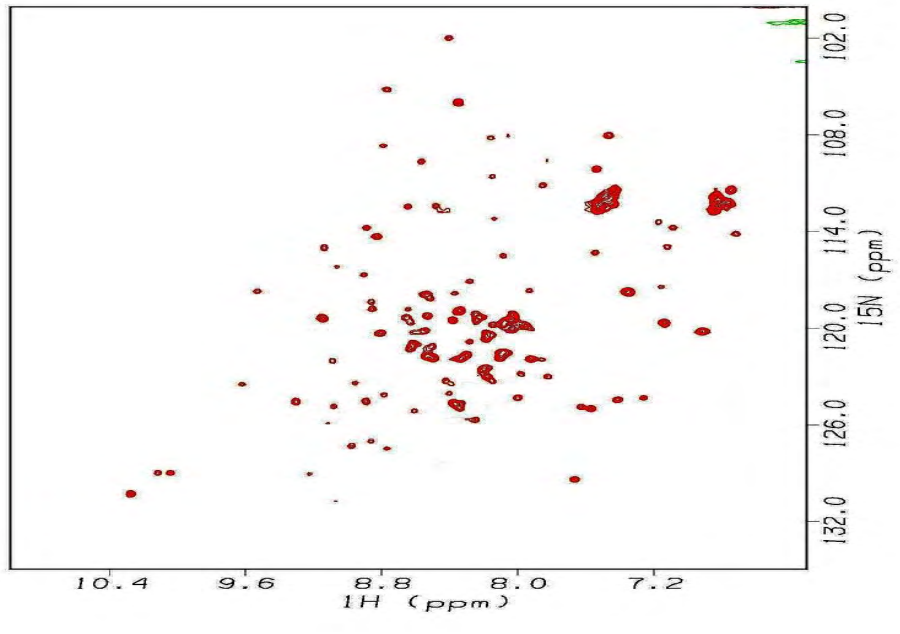
Compared with the HSQC spectrum of Q7K, the new construct significantly increases protease stability.

Figure 4.10: HSQC spectra of a WT_{NMR} sample. (A) Spectra recorded on Feb 15th, 2006. (B) Spectra recorded on Feb 21st, 2006.

(A)



(B)



ITC experimental results of WT_{NMR} and $Flap+_{NMR}$ proteases binding with DRV

Isothermal titration calorimetry displacement was used to measure the binding affinity of DRV with WT_{NMR} and $Flap+_{NMR}$. The results are compared with the ITC data for WT and Flap+ variants. All the conditions and steps of the WT_{NMR} and $Flap+_{NMR}$ ITC experiments followed those used in the WT and Flap+ ITC experiments[87,93] for parallel comparison. Pepstatin was used as a binding competitor of DRV to protease.

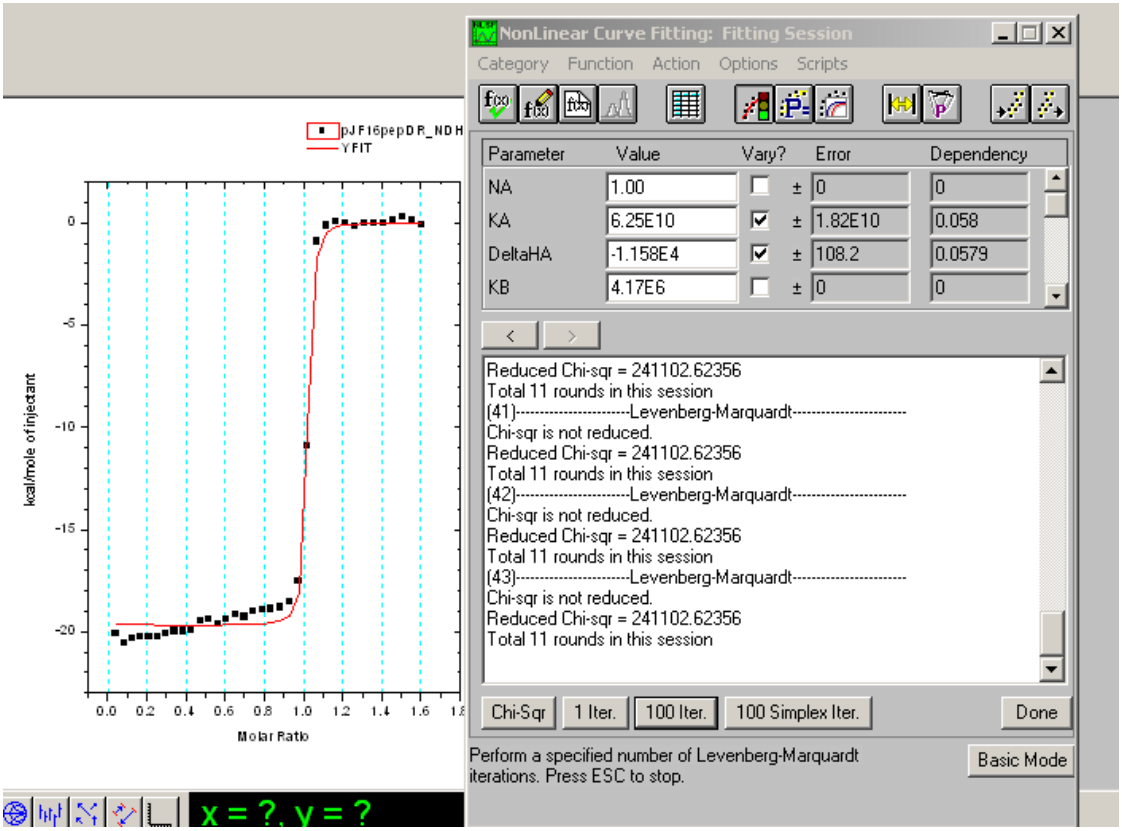
The binding affinity and enthalpy change of DRV with WT_{NMR} and $Flap+_{NMR}$ were measured using competitive displacement ITC experiment on a VP- isothermal titration calorimeter at temperature 293 K. The measurements were done in buffer at pH 5.0 consisted of 10 mM sodium acetate, 2.0% dimethyl sulfoxide, and 2 mM TCEP.[93] 25 µM protease solution was saturated with twenty-nine 10 µl injections of 0.3 mM pepstatin. Pepstatin was then displaced from protease by adding twenty-nine 10 µl injections of 0.25 mM DRV. Software Origin7 (Figure 4.11) was used to process the data to get the binding affinity and enthalpy of the interaction between the protease and DRV.

The ITC experiments results for WT, Flap+, ACT, WT_{NMR} and $Flap+_{NMR}$ are shown in Table 4.6. The introduced mutations did change the binding affinities which were 0.32 kcal/mol for the WT and 0.55 kcal/mol for the Flap+. The entropy and enthalpy changes between WT_{NMR} and $Flap+_{NMR}$ are in agreement with the profiles of

WT and Flap+ variants. (Figure 4.12) The very high enthalpy-entropy compensation was observed from the ITC results.

Figure 4.11: ITC displacement experiments (A) and (B) of WT_{NMR} binding with DRV using pepstatin as competitor protease binder.

(A)



(B)

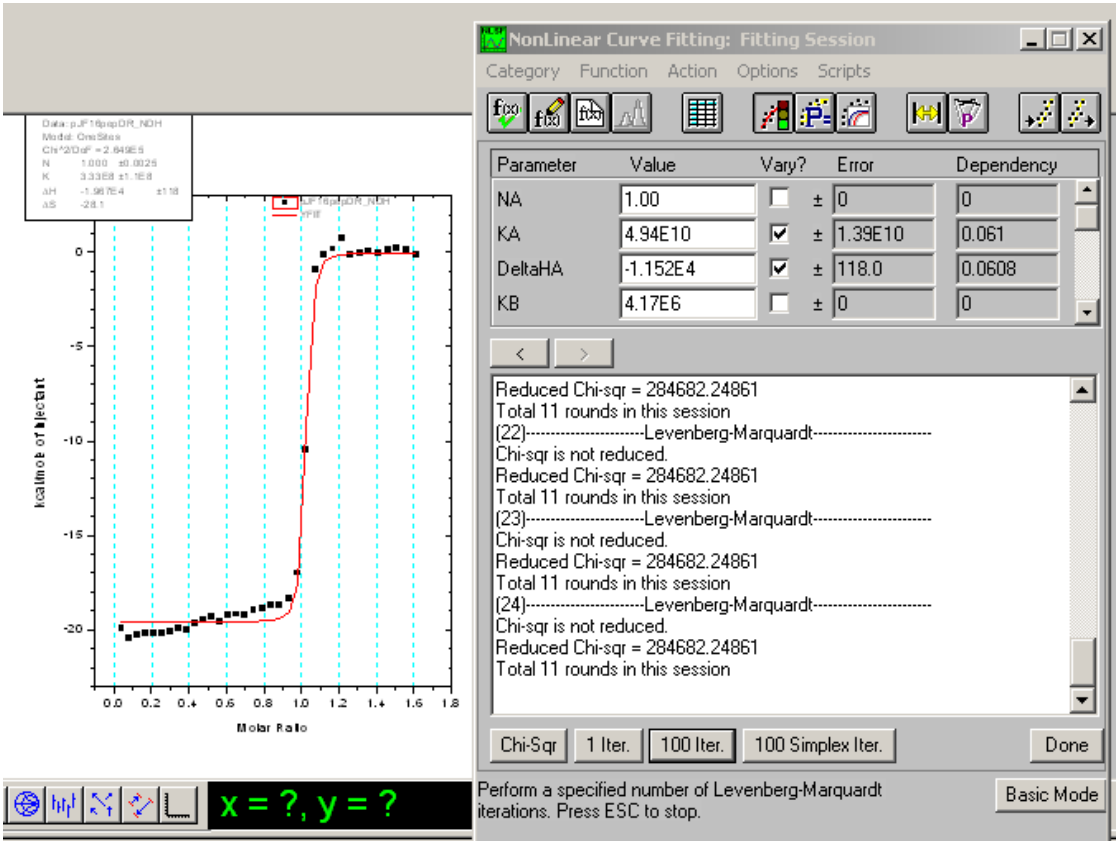
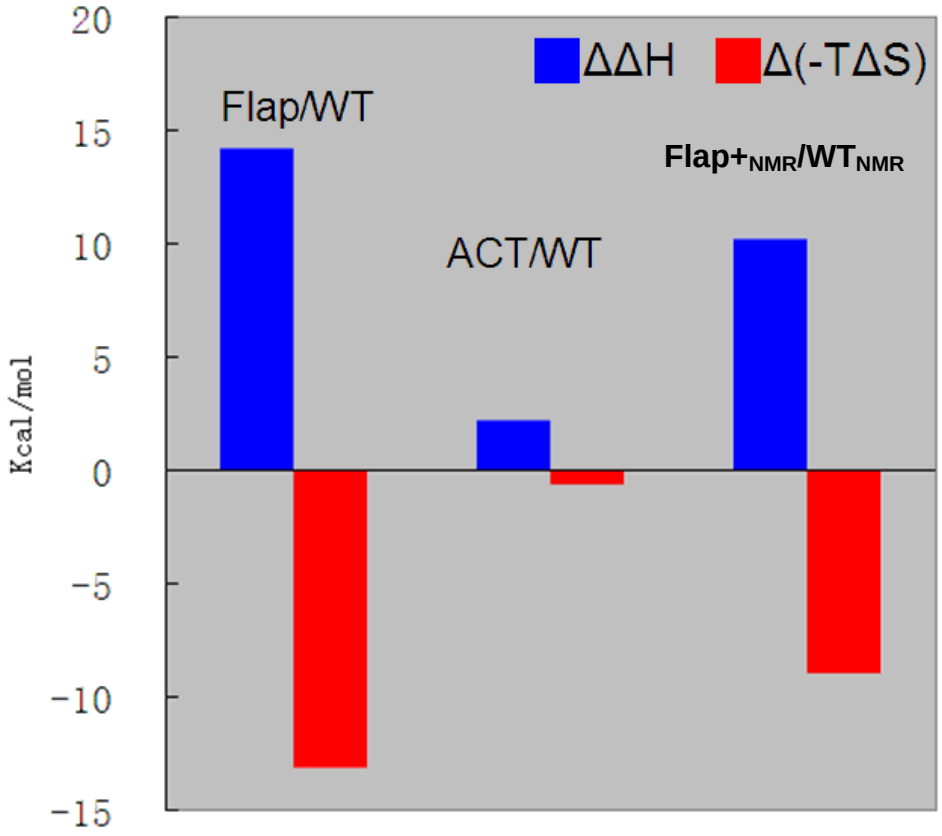

NonLinear Curve Fitting: Fitting Session
Category Function Action Options Scripts
Parameter Value Vary? Error Dependency
NA 1.00 0 0
KA 4.94E10 1.39E10 0.061
DeltaHA -1.152E4 118.0 0.0608
KB 4.17E6 0 0
Reduced Chi-sqr = 284682.24861
Total 11 rounds in this session
(22)----------Levenberg-Marquardt----------
Chi-sqr is not reduced.
Reduced Chi-sqr = 284682.24861
Total 11 rounds in this session
(23)----------Levenberg-Marquardt----------
Chi-sqr is not reduced.
Reduced Chi-sqr = 284682.24861
Total 11 rounds in this session
(24)----------Levenberg-Marquardt----------
Chi-sqr is not reduced.
Reduced Chi-sqr = 284682.24861
Total 11 rounds in this session
Chi-Sqr 1 Iter. 100 Iter. 100 Simplex Iter. Done
Perform a specified number of Levenberg-Marquardt iterations. Press ESC to stop.
Basic Mode
kcal/mole of Injectant
Molar Ratio
x = ?, y = ?

Table 4.6: Thermodynamic data of WT_{NMR} and $Flap+_{NMR}$ binding with DRV measured by displacement isothermal titration calorimetry experiments.

		ΔH (kcal/mol)	**ΔG (kcal/mol)**	**-TΔS (kcal/mol)**	**ΔS (cal/[mol·T])**
DRV-WT		**-12.1±0.8**	**-15.2±0.1**	**-3.1±0.9**	**10.6±2.7**
DRV-Flap+		**2.0±0.6**	**-14.2±0.2**	**-16.2±0.8**	**55.3±2.5**
DRV-WT_{NMR}	Exp1	-11.6±0.1	-15.0±0.1	-3.4±0.2	11.6±0.5
	Exp2	-11.6±0.1	-14.8±0.1	-3.2±0.2	11.0±0.5
	Avg	**-11.6±0.1**	**-14.9±0.1**	**-3.3±0.2**	**11.3±**0.5
DRV-$Flap+_{NMR}$	Exp1	-1.9±0.2	-13.8±0.1	-11.9±0.3	40.6±1.0
	Exp2	-1.6±0.2	-13.5±0.1	-11.9±0.3	40.6±1.0
	Avg	**-1.8±0.3**	**-13.6±0.1**	**-11.9±0.3**	**40.6±1.0**
DRV-ACT		**-10.0±0.1**	**-13.6±0.1**	**-3.6±0.2**	**12.5±0.7**

Figure 4.12. The WT_{NMR} and $Flap^{+}_{NMR}$ also have large scale of enthalpy and entropy change.



Preliminary results of NMR relaxation experiments on Flap+$_{NMR}$ and comparison with published wild-type data

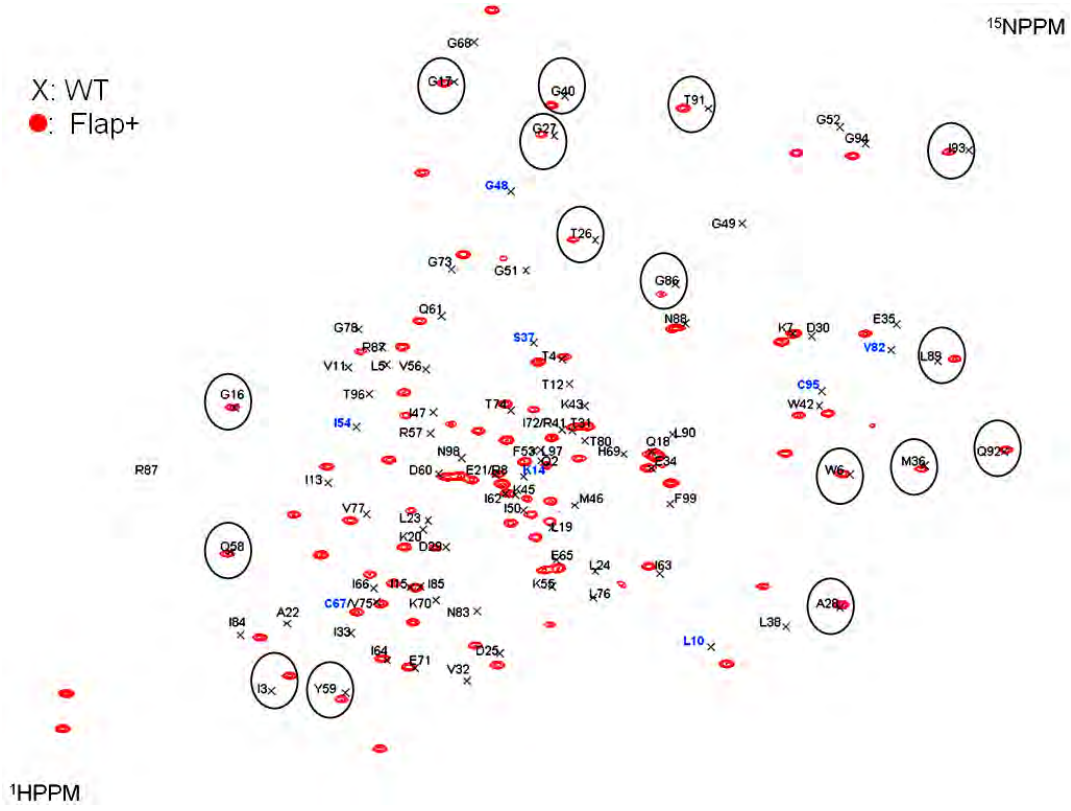
To date, the T1, T2 and NOE data of apo Flap+$_{NMR}$ backbone amide were collected from an 800 MHz magnet by J. Peng in University of Notre Dame. The Flap+$_{NMR}$ protease sample was 250 μL solutions in Shigemi microcells at dimer concentrations 400 μM in 20 mM sodium phosphate with 2% D_2O. Spectra were recorded at temperature 293K. ^{15}N T_1's were measured using relaxation delays of 32, 56, 88, 320, 640, and 960 ms and 8, 16, 128, 256, 384, 512, and 640 ms. T_2 were measured with relaxation delays of 8, 16 24, 32, 56, 88, and 120 ms, and of 6, 12, 18, 24, 36, 48, and 60 ms, respectively. ^{15}N–^{1}H NOE experiments were performed using a water flip-back sequence.[31] NOE values were measured by taking the ratio of peak intensities from experiments performed with and without ^{1}H presaturation. All data were processed using the TOPSpin software package[30] and peak heights measured in the processed spectra were fitted with a two-parameter exponential function to extract relaxation rates. Errors in T_1 and T_2 (or $T_{1\rho}$) were determined by Monte-Carlo simulations[31].

As a preliminary comparison, I superimposed the HSQC spectrum of Flap+$_{NMR}$ onto the HSQC spectrum of WT$_{NMR}$ with residues assignment publised[31]. Most of the residues have chemical shifts. Several residues have relatively isolated peaks on the spectrum and might be identified. (Figure 4.13A) Their T1 and NOE values are

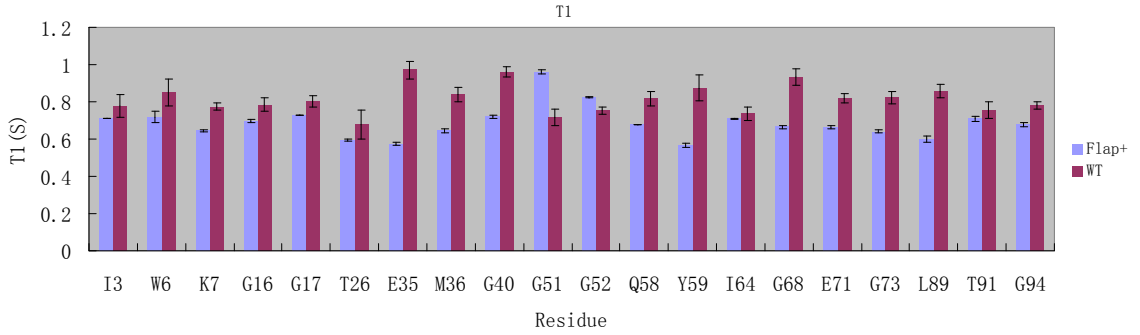
compared from the WT_{NMR} and $Flap^{+}_{NMR}$. (Figure 4.13B, C) The T1 and NOE show different values indicating change of the protease dynamics. Residues have large chemical shift might have larger difference in dynamics.

Figure 4.13. (A) Superimposition of the HSQC spectra of Flap+$_{NMR}$ on the WT$_{NMR}$ spectra with residues assignments. The cycled residues are assumed to be the same residue in Flap+$_{NMR}$ and WT$_{NMR}$. (B) T1 values of residue of the Flap+$_{NMR}$ and WT$_{NMR}$. (C) NOE values of residue of the Flap+$_{NMR}$ and WT$_{NMR}$.

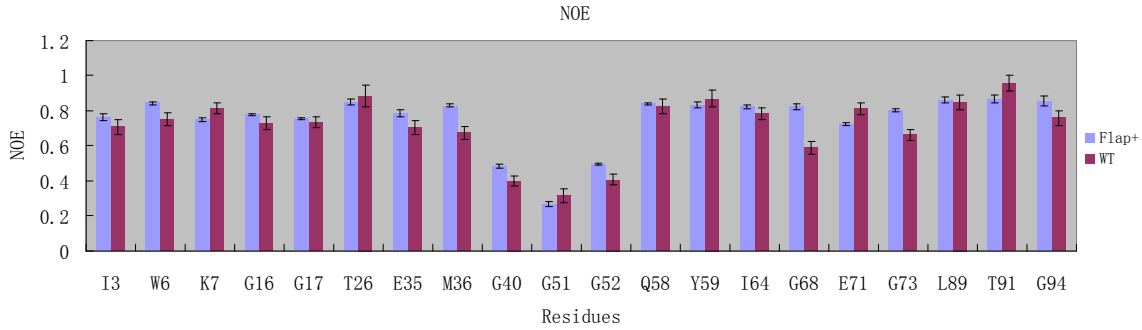
(A)



(B)


T1
T1(S)
1.2
1
0.8
0.6
0.4
0.2
0
I3 W6 K7 G16 G17 T26 E35 M36 G40 G51 G52 Q58 Y59 I64 G68 E71 G73 L89 T91 G94
Residue
Flap+
WT

(C)


NOE
NOE
1.2
1
0.8
0.6
0.4
0.2
0
I3 W6 K7 G16 G17 T26 E35 M36 G40 G51 G52 Q58 Y59 I64 G68 E71 G73 L89 T91 G94
Residues
Flap+
WT

Summary

The scope of this study was to understand how drug-resistant mutations change the thermodynamic and dynamic properties of HIV-1 protease. Thereby gaining knowledge which could be applied in the structure-based drug design of HIV-1 protease inhibitors and make them more effective against drug-resistance. Comparing the crystal structure of APV bound to wild-type protease and a drug-resistant protease variant, King et al. found that the mutation I84V has decreased the vdW interaction between APV and the drug-resistant variant, which might account for the loss of binding affinity between APV and the drug-resistant variant. By analyzing the ITC experiments results, Irene et al. suggested that the drug-resistant mutations change the shape of the active site. The very flexible substrates are less susceptible to the change than the synthetic inhibitors[102], which might enable the drug-resistant protease variant to still recognize the substrate while having less binding affinity with the synthetic inhibitors. Comparing the trajectories from MD simulations on the wild-type protease and the V82F/I84V protease variant, Perryman et al. suggested that the mutations change the equilibrium between the flap-semiopen and closed conformations could be one aspects of the protease drug-resistant mechanism[96]. To study the changes in thermodynamic properties of inhibitor binding with drug-resistant mutations, I performed free-energy calculations and decomposition using the crystal structure atomic coordinates as input and thermodynamic data from isothermal titration

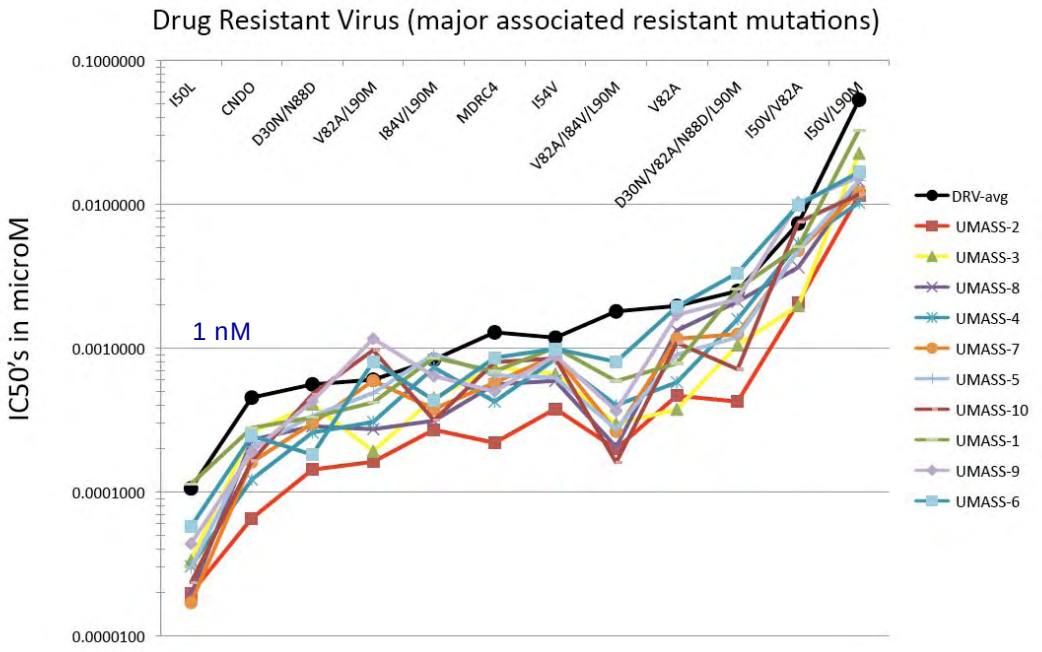
calorimetry experiments as control. The predicted binding free energy was in good agreement with the ITC experimental data.

The thermodynamic integration method, developed from statistical mechanics, was shown to be advantageous in precisely predicting the relative binding free-energy change (Table 2.4). Applying free-energy decomposition to classical energy simulation methods such free energy perturbation and thermodynamic integration has been questioned. The application of such methods usually calculates through the thermodynamic cycle (Chapter II, Methods) which requires the quantity to be a state function such as Gibbs free energy. The free-energy components defined by either different types of molecular interactions (ie. vdW, electrostatic) or contributions from different protein residues or ligand atoms, are no longer state functions which are not path dependent [103-105]. The MM-PB/GBSA method is a faster but less accurate alternative to traditional perturbation methods. This method has increased in popularity to calculate absolute binding free-energy changes associated with biomolecular recognition and relative free energies of different conformations [24-28]. The MM-PB/GBSA method is also convenient for carrying out free-energy component analysis. The free energy analysis of protease-inhibitor binding in Chapter II and the further discussions here reveal that a drug-resistant mutation at a residue not only changes its interaction with an inhibitor, but also propagates changes throughout the active site to other residues interactions with the inhibitor.

The design of DRV from APV, which replaced the tetrahydrofuranyl group with

the bis-tetrahydrofuranyl group, successfully increased the binding affinity of the drug with the protease.[54,55] Although DRV lost binding affinity with the drug-resistant variants Flap+ and ACT, it still binds better than APV with these variants.[87,93] From energy decomposition of the different moieties of DRV, the bis-tetrahydrofuranyl group maintains its vdW interaction energy even in the drug-resistant protease variants. The design of new inhibitors from the DRV structure should consider modifying the 4-amino phenyl and isopropyl groups to reduce DRV susceptibility to mutations in the P1 loop. Recently, several new DRV-based compounds have been synthesized by chemists in Schiffer's group. These compounds have been shown to have overall better binding affinity with several typical drug-resistant variants (Figure 4.14). All these compounds are larger than DRV at the 4-amino phenyl and isopropyl groups, which generate more vdW interactions with the protease. These changes could be the molecular basis for these analogues being more tolerant to drug-resistant mutations than DRV.

Figure 4.14: Antiviral experiments performed by Monogram show the new DRV analogues have better binding affinity than DRV to drug-resistant protease variants



The Gibbs free energy calculations results were highly correlated with the experimental result. (Chapter II) The calculation of entropy change remained as a big challenge. The calculations of the change of Gibbs free energy by TI or FEP methods only rely on the sampling of the low energy regions of the phase space. Using TI and FEP to calculate the change of entropy required the sampling of all the configurational space which is hard to achieve with current hardware techniques. In MM-PB/GBSA methods, some of the entropy changes (ΔSrot, ΔStran, ΔSvib) of the protease-inhibitor binding were calculated explicitly. These terms can not be compared with the experimental binding entropy change directly. Part of the entropy change was coupling with the solvation model and calculated implicitly associating with the polar (PB or GB) solvation energy change or the non-polar solvation energy change.

In Chapter III the dynamics of two apo proteases, the wild-type and a drug-resistant variant Flap+, were compared *in silico.* 20 ns molecular dynamic simulations were performed and analyzed. The molecular mechanics properties indicate that the apo Flap+ protease variant is less dynamic than the wild-type protease. Molecular dynamic simulations can elucidate protease dynamics on an atomic level, which is difficult to access by any experimental methods. NMR experiments can provide dynamic information about the protease on a longer time scale, e.g., microseconds, which is hard to accomplish in standard MD simulation with explicit solvent. NMR relaxation experiments require high concentration

protease samples, 300 to 500 μM of dimer concentrations are preferable. Under these high concentration conditions, protease, which is its own cleavage substrate, undergoes severe autoproteolysis. Thus, new constructs (e.g., Flap+$_{NMR}$) were synthesized to prepare stable NMR samples. The thermodynamic properties of these new constructs in binding with DRV were measured by ITC experiments and compared to the original wild-type and Flap+ variant. R1, R2 and NOE data of Flap+$_{NMR}$ were collected from an 800 MHz magnet. To identify all peaks, assignment experiments were needed. Thus, all the relaxation data will be compared to the published wild-type data[30,31,99] to understand how the drug-resistant mutations changed the apo protease dynamics.

In summary, the molecular mechanism of HIV-1 protease drug resistance is complicated and comprehensive investigation is needed. Computational methods such as free energy analysis and MD simulation can provide thermodynamic and dynamic insights about protease variants on an atomic level. These computational data complement the experimental calorimetry and NMR, and by obtaining a series of datasets from a series of parallel experiments the details of how inhibitor binding is altered by drug resistance can be elucidated. The knowledge from correct modeling will enrich understanding of how drug-resistant mutations change the energetic and dynamic characteristics of HIV-1 protease and help to develop protease inhibitors that bind more tightly to the protease and are more tolerant of drug-resistant mutations.

References

(1) Flexner, C. *Nat Rev Drug Discov* **2007**, *6*, 959-66.

(2) Condra, J. H.; Holder, D. J.; Schleif, W. A.; Blahy, O. M.; Danovich, R. M.; Gabryelski, L. J.; Graham, D. J.; Laird, D.; Quintero, J. C.; Rhodes, A.; Robbins, H. L.; Roth, E.; Shivaprakash, M.; Yang, T.; Chodakewitz, J. A.; Deutsch, P. J.; Leavitt, R. Y.; Massari, F. E.; Mellors, J. W.; Squires, K. E.; Steigbigel, R. T.; Teppler, H.; Emini, E. A. *J Virol* **1996**, *70*, 8270-6.

(3) Molla, A.; Korneyeva, M.; Gao, Q.; Vasavanonda, S.; Schipper, P. J.; Mo, H. M.; Markowitz, M.; Chernyavskiy, T.; Niu, P.; Lyons, N.; Hsu, A.; Granneman, G. R.; Ho, D. D.; Boucher, C. A.; Leonard, J. M.; Norbeck, D. W.; Kempf, D. J. *Nat Med* **1996**, *2*, 760-6.

(4) Kollman, P. A. *Chem . Rev.* **1993**, *93*, 2395-2417.

(5) Markland, W.; Rao, B. G.; Parsons, J. D.; Black, J.; Zuchowski, L.; Tisdale, M.; Tung, R. *J Virol* **2000**, *74*, 7636-41.

(6) Wu, T. D.; Schiffer, C. A.; Gonzales, M. J.; Taylor, J.; Kantor, R.; Chou, S.; Israelski, D.; Zolopa, A. R.; Fessel, W. J.; Shafer, R. W. *J Virol* **2003**, *77*, 4836-47.

(7) Shao, W.; Everitt, L.; Manchester, M.; Loeb, D. D.; Hutchison, C. A., 3rd; Swanstrom, R. *Proc Natl Acad Sci U S A* **1997**, *94*, 2243-8.

(8) Maguire, M.; Snowden, W. *Antimicrob Agents and Chemother* **2002**, *46*, 731-738.

(9) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *287*, 564-567.

(10) Singh, U. C.; Benkovic, S. J. *Proceedings of the National Academy of Sciences of the United States of America* **1988**, *85*, 9519-23.

(11) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. *Science* **1989**, *244*, 1069-72.

(12) Massova, I.; Kollman, P. A. *Perspectives in Drug Discovery and Design* **1999**, *18*, 113-135.

(13) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351–361.

(14) Tanford, C.; Kirkwood, J. G. *J. Am. Chem. Soc* **1957**, *79*, 5333-5339.

(15) Warwicker, J.; Watson, H. C. *J Mol Biol* **1982**, *157*, 671-9.

(16) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins* **1986**, *1*, 47-59.

(17) Davis, M. E.; McCammon, J. A. *J Comput Chem* **1989**, *10*, 386-391.

(18) Nicholls, A.; Honig, B. *J Comput Chem* **1991**, *12*, 435-445.

(19) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244-53.

(20) Still *J. Am. Chem. Soc* **1990**, *112*, 6127.

(21) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B.* **2000**, *104*, 3712-3720.

(22) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978-1988.

(23) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305-20.

(24) Smith, K. C.; Honig, B. *Proteins* **1994**, *18*, 119-32.

(25) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc* **1998**, *120*, 9401.

(26) Yang, A. S.; Honig, B. *J Mol Biol* **1995**, *252*, 366-76.

(27) Yang, A. S.; Honig, B. *J Mol Biol* **1995**, *252*, 351-65.

(28) Yang, A. S.; Hitz, B.; Honig, B. *J Mol Biol* **1996**, *259*, 873-82.

(29) Scott, W. R.; Schiffer, C. A. *Structure* **2000**, *8*, 1259-65.

(30) Ishima, R.; Freedberg, D. I.; Wang, Y. X.; Louis, J. M.; Torchia, D. A. *Structure* **1999**, *7*, 1047-55.

(31) Freedberg, D. I.; Ishima, R.; Jacob, J.; Wang, Y. X.; Kustanovich, I.; Louis, J. M.; Torchia, D. A. *Protein Sci* **2002**, *11*, 221-32.

(32) Debouck, C. *AIDS Res. Hum. Retroviruses* **1992**, *8*, 153-164.

(33) Wlodawer, A.; Erickson, J. W. *Annu. Rev. Biochem.* **1993**, *62*, 543-585.

(34) Schinazi, R. F.; Larder, B. A.; Mellors, J. W. *Internatl Antiviral News* **1997**, *5*, 129-142.

(35) King, N. M.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Wigerinck, P.; de Bethune, M. P.; Schiffer, C. A. *J. Virol.* **2004**, *78*, 12012-12021.

(36) Todd, M. J.; Luque, I.; Velazquez-Campoy, A.; Freire, E. *Biochemistry* **2000**, *39*, 11876-11883.

(37) King, N. M.; Melnick, L.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Yang, S. S.; Gao, Y.; Nie, X.; Zepp, C.; Heefner, D. L.; Schiffer, C. A. *Protein Sci.* **2002**, *11*, 418-429.

(38) Prabu-Jeyabalan, M.; Nalivaika, E. A.; King, N. M.; Schiffer, C. A. *J. Virol.* **2003**, *77*,

1306-15.

(39) Talhout, R.; Villa, A.; Mark, A. E.; Engberts, J. B. *J. Am. Chem. Soc.* **2003**, *125*, 10570-9.

(40) Huang, N.; Jacobson, M. P. *Current opinion in drug discovery and development* **2007**, *10*, 325-331.

(41) Jorgensen, W. L. *Science* **2004**, *303*, 1813-1818.

(42) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* **1987**, *235*, 574-6.

(43) Wang, W.; Kollman, P. A. *Proc Natl Acad Sci U S A* **2001**, *98*, 14937-42.

(44) Wittayanarakul, K.; Aruksakunwong, O.; Sompornpisut, P.; Sanghiran-Lee, V.; Parasuk, V.; Pinitglang, S.; Hannongbua, S. *J Chem Inf Model* **2005**, *45*, 300-8.

(45) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420-1426.

(46) Michielin, O.; Karplus, M. *Journal of Molecular Biology* **2002**, *324*, 547–569.

(47) Archontis, G.; Simonson, T.; Moras, D.; Karplus, M. *J. Mol. Biol.* **1998**, *275*, 823-46.

(48) Singh, U. C.; Benkovic, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 9519-23.

(49) Xu, Y.; Wang, R. *Proteins* **2006**, *64*, 1058-68.

(50) Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891-913.

(51) Archontis, G.; Simonson, T.; Karplus, M. *J. Mol. Biol.* **2001**, *306*, 307-27.

(52) Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1999**, *8*, 1381-92.

(53) Hou, T.; Yu, R. *J. Med. Chem.* **2007**, *50*, 1177-88.

(54) Surleraux, D. L.; Tahri, A.; Verschueren, W. G.; Pille, G. M.; de Kock, H. A.; Jonckers, T. H.; Peeters, A.; De Meyer, S.; Azijn, H.; Pauwels, R.; de Bethune, M. P.; King, N. M.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Wigerinck, P. B. *J. Med. Chem.* **2005**, *48*, 1813-22.

(55) Surleraux, D. L.; de Kock, H. A.; Verschueren, W. G.; Pille, G. M.; Maes, L. J.; Peeters, A.; Vendeville, S.; De Meyer, S.; Azijn, H.; Pauwels, R.; de Bethune, M. P.; King, N. M.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Wigerinck, P. B. *J Med Chem* **2005**, *48*, 1965-73.

(56) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668-88.

(57) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.;

Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999-2012.

(58) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-74.

(59) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb,; M. A.; Cheeseman, J. R. Z., V. G.; Montgomery, J. A. J.;; Stratmann, R. E. B., J. C.; Dapprich, S.; Millam, J. M.; Daniels,; A. D.; Kudin, K. N. S., M. C.; Farkas, O.; Tomasi, J.; Barone,; V.; Cossi, M. C., R.; Mennucci, B.; Pomelli, C.; Adamo, C.;; Clifford, S. C., J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.;; Morokuma, K. M., D. K.; Rabuck, A. D.; Raghavachari, K.;; Foresman, J. B. C., J.; Ortiz, J. V.; Baboul, A. G.; Stefanov,; B. B.; Liu, G. L., A.; Piskorz, P.; Komaromi, I.; Gomperts,; R.; Martin, R. L. F., D. J.; Keith, T.; Al-Laham, M. A.; Peng, C.; Y.; Nanayakkara, A. G., C.; Challacombe, M.; Gill, P. M.; W.; Johnson, B. C., W.; Wong, M. W.; Andres, J. L.; Gonzalez,; C.; Head-Gordon, M. R., E. S.; Pople, J. A. Gaussian Inc. **2003**.

(60) Bayly, C. I. C., P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.

(61) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.

(62) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297-304.

(63) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383-94.

(64) Radmer, R. J.; Kollman, P. A. *J. Comput. Aided Mol. Des.* **1998**, *12*, 215-27.

(65) Lau, F. T.; Karplus, M. *Journal of Molecular Biology* **1994**, *236*, 1049-1066.

(66) Blondel, A. *Journal of Computational Chemistry* **2004**, *25*, 985-993.

(67) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *Journal of Computational Chemistry* **2005**, *26*, 1668-88.

(68) Case, D. A. 2005.

(69) Bishop, M.; Frinks, S. *Journal of Chemical Physics* **1987**, *87*, 3675-3676.

(70) Bea, I.; Gotsev, M. G.; Ivanov, P. M.; Jaime, C.; Kollman, P. A. *J. Org. Chem.* **2006**, *71*, 2056-63.

(71) Wittayanarakul, K.; Hannongbua, S.; Feig, M. *J. Comput. Chem.* **2008**, *29*, 673-85.

(72) Gohlke, H.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 238-50.

(73) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., 3rd *J. Med. Chem.* **2004**, *47*,

3032-47.

(74) Lee, T.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 4385.

(75) Stoica, I.; Sadiq, S. K.; Coveney, P. V. *J. Am. Chem. Soc.* **2008**, *130*, 2639-48.

(76) Massova, I.; Kollman, P. A. *J. Am. Chem. Soc* **1999**, *11*, 8133-8143.

(77) Wang, W.; Kollman, P. A. *J. Mol. Biol.* **2000**, *303*, 567-82.

(78) Huo, S.; Massova, I.; Kollman, P. A. *J. Comput. Chem.* **2002**, *23*, 15-27.

(79) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. *Science* **1989**, *245*, 616-21.

(80) Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P. H.; Poljak, R. J. *Biochimie* **1991**, *73*, 1391-6.

(81) Heaslet, H.; Rosenfeld, R.; Giffin, M.; Lin, Y. C.; Tam, K.; Torbett, B. E.; Elder, J. H.; McRee, D. E.; Stout, C. D. *Acta Crystallogr D Biol Crystallogr* **2007**, *63*, 866-75.

(82) Wlodawer, A.; Erickson, J. W. *Annual Review of Biochemistry* **1993**, *62*, 543-585.

(83) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. *J Mol Biol* **2000**, *301*, 1207-20.

(84) Lapatto, R.; Blundell, T.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S.; Merson, J. R.; Whittle, P. J.; Danley, D. E.; Geoghegan, K. F.; et al. *Nature* **1989**, *342*, 299-302.

(85) Furfine, E. S.; D'Souza, E.; Ingold, K. J.; Leban, J. J.; Spector, T.; Porter, D. J. *Biochemistry* **1992**, *31*, 7886-91.

(86) Rodriguez, E. J.; Debouck, C.; Deckman, I. C.; Abu-Soud, H.; Raushel, F. M.; Meek, T. D. *Biochemistry* **1993**, *32*, 3557-63.

(87) King, N. M.; Prabu-Jeyabalan, M.; Bandaranayake, R. M.; Nalam, M. N.; Ozen, A.; Haliloglu, T.; Schiffer, C. *In Preparation* **2009**.

(88) Ishima, R.; Louis, J. M. *Proteins* **2008**, *70*, 1408-15.

(89) Rick, S. W.; Erickson, J. W.; Burt, S. K. *Proteins* **1998**, *32*, 7-16.

(90) Moody, M. D.; Pettit, S. C.; Shao, W.; Everitt, L.; Loeb, D. D.; Hutchison, C. A., 3rd; Swanstrom, R. *Virology* **1995**, *207*, 475-85.

(91) Galiano, L.; Blackburn, M. E.; Veloro, A. M.; Bonora, M.; Fanucci, G. E. *J Phys Chem B* **2009**, *113*, 1673-80.

(92) Galiano, L.; Ding, F.; Veloro, A. M.; Blackburn, M. E.; Simmerling, C.; Fanucci, G. E. *J Am*

Chem Soc **2009**, *131*, 430-1.

(93) King, N. M.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Wigerinck, P.; de Bethune, M. P.; Schiffer, C. A. *Journal of Virology* **2004**, *78*, 12012-12021.

(94) Lippow, S. M.; Wittrup, K. D.; Tidor, B. *Nat Biotechnol* **2007**, *25*, 1171-6.

(95) Huggins, D. J.; Altman, M. D.; Tidor, B. *Proteins* **2009**, *75*, 168-86.

(96) Perryman, A. L.; Lin, J. H.; McCammon, J. A. *Protein Sci* **2004**, *13*, 1108-23.

(97) Galiano, L.; Bonora, M.; Fanucci, G. E. *J Am Chem Soc* **2007**, *129*, 11004-5.

(98) Mildner, A. M.; Rothrock, D. J.; Leone, J. W.; Bannow, C. A.; Lull, J. M.; Reardon, I. M.; Sarcich, J. L.; Howe, W. J.; Tomich, C. S.; Smith, C. W.; et al. *Biochemistry* **1994**, *33*, 9405-13.

(99) Ishima, R.; Ghirlando, R.; Tozser, J.; Gronenborn, A. M.; Torchia, D. A.; Louis, J. M. *J Biol Chem* **2001**, *276*, 49110-6.

(100) Louis, J. M.; Ishima, R.; Nesheiwat, I.; Pannell, L. K.; Lynch, S. M.; Torchia, D. A.; Gronenborn, A. M. *J Biol Chem* **2003**, *278*, 6085-92.

(101) Szeltner, Z.; Polgar, L. *J Biol Chem* **1996**, *271*, 5458-63.

(102) Luque, I.; Todd, M. J.; Gomez, J.; Semo, N.; Freire, E. *Biochemistry* **1998**, *37*, 5791-7.

(103) Boresch, S.; Karplus, M. *J Mol Biol* **1995**, *254*, 801-7.

(104) Brady, G. P.; Sharp, K. A. *J Mol Biol* **1995**, *254*, 77-85.

(105) Mark, A. E.; van Gunsteren, W. F. *J Mol Biol* **1994**, *240*, 167-76.