NANYANG TECHNOLOGICAL UNIVERSITY

APPLICATION OF MOLECULAR DYNAMICS SIMULATIONS IN THE STUDY OF PROTEIN-LIGAND INTERACTIONS

DUAN RUI

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

2016

APPLICATION OF MOLECULAR DYNAMICS SIMULATIONS IN THE STUDY OF PROTEIN-LIGAND INTERACTIONS

DUAN RUI

DUAN RUI

School of Physical and Mathematical Sciences

A thesis submitted to the Nanyang Technological University

in fulfilment of the requirement for the degree of

Doctor of Philosophy of Science

2016

Acknowledgement

This thesis is written under the sincere guidance of my supervisor, Asst Prof. Zhang Dawei. I would like to express my greatest appreciation to him as being such a tremendous mentor and for all the continuous support throughout my PhD study, making me to grow as a computational chemist. I cannot imagine knowing a better advisor or mentor during my course of graduate study. The completion of this thesis will never be a success without his patience, encouragements and instructions in all ways. His advice and sharing are invaluable towards my research, career as well as the attitude of future life.

Also I would like to thank Prof. Mei Ye for his help and generous in providing the source code I used in my PhD study. My colleagues in Dr. Zhang's group, Cui Jinglan, Yang Hong, Yip Yew Mun, especially Dr. Wei Caiyi, Dr. Sun Tiedong and Dr. Raudah Lazim, thanks for their immense contributions towards both my personal and professional time at NTU. Thanks for the priceless discussions, joys and friendships have ever had with them.

Lastly, but never the least, a special thanks to my family. No words can express my gratefulness towards my mother, father and wife for all their sacrifices have been made for me. Their love, encouragement and support are what sustained me so far.

Contents

Abstract

Molecular dynamics (MD) simulation is a fundamental approach to allow researchers to study the molecular interactions of protein-ligand binding via the dynamics of protein-ligand complexes at atomic level. It can be observed precisely the motion of atoms and interactions involved in receptor and ligand binding using MD simulations. In this thesis, conventional MD simulations is applied to investigate the basis of mutation induced resistance HIV-1 integrase inhibitors, and positive agreement with experimental result has been observed and reasonable explanations are presented. Nevertheless, the deficiency of conventional molecular dynamics simulation is not negligible either. It lacks polarization effects in classical force fields utilized in traditional MD simulations, consequently the capability of describing the specific electrostatic properties of different biological systems is absent. Previously developed polarized protein-specific charge (PPC) allows electrostatic polarization effect to be introduced into molecular dynamics simulation as static charges. In this thesis, this method is employed in MD simulations to explore the mechanism of governing resistance to its inhibitor while HIV-1 protease Ile mutates to Val at position 50 in both chains; and the importance of electrostatic polarization effect and advantages compare to classical force field is demonstrated. The result shows polarization effect included simulations are more reliable comparing with experimental data. However, static PPC charges are calculated based on single conformation, and it causes bias of atomic charges, e.g. overestimate particular hydrogen bond strength resulting extra stability during simulation. As conformations of the systems varying dynamically would lead to electrostatic environment change, atomic charges in system should also

be re-calculated during MD simulations. Environmentally corresponding local polarized protein-specific charge (ECLPPC) scheme and environmentally corresponding hydrogen bond dependent polarized protein-specific charge (ECHBPPC) scheme have been newly developed and applied in protein-ligand MD simulations. They update the atomic charges of ligand and selected residues in certain criteria surrounding the ligand periodically to incorporate polarization into MD simulations. MD simulations are conducted by employing ECLPPC and ECHBPPC respectively to investigate the basis of avidin-biotin interaction and give reasonable results with agreement to experiment.

List of abbreviations

APV	Amprenavir
DTG	Dolutegravir
ECHBPPC	Environmentally Corresponding Hydrogen Bond Dependent Polarized Protein-Specific Charge
ECLPPC	Environmentally Corresponding Local Polarized Protein-Specific Charge
EIDC	Environment Independent Charge
IDV	Indinavir
INSTIs	Integrase Strand Transfer Inhibitors
LPPC	Local Polarized Protein-Specific Charge
MD	Molecular Dynamics
MFCC	Molecular Fragmentation with Conjugated Caps
MM	Molecular Mechanics
PB	PoissonBoltzmann
PIs	Protease Inhibitors
PPC	Polarized Protein-speci_c Charge
QM	Quantum Mechanics
RAL	Raltegravir
RESP	Restrained Electrostatic Potential

List of figures

List of tables

Chapter 1

Introduction

1.1 Overview

Molecular dynamics (MD) was initially introduced to theoretical physics and applied to studying the motion of liquids in the late 1950s to early 1960s.[1-2] Later on, in 1970s, the earliest papers on studying biomolecules by employing molecular dynamics (MD) simulations were conducted and published.[3-5] Though these were merely short simulations on small proteins, it demonstrated the possibility and capability of using MD simulations on exploring internal motion of molecules in biological systems.[3-5] Due to the complexity of biological systems and expediency of simulation method, computer-aided modelling technique has grown into an indispensable tool in sciences. Along with the increasing of computing power, the motional phenomena of larger and more complex systems can be explored to solve meaningful problems using MD simulations.[6-9] More importantly, theoretical study on biological systems possesses great advantages on providing very detailed information of molecular motion that experiments on the actual systems may easily failed to have. As realizing its features and power, a great variety of studies on motional phenomena of biological systems, for instance nucleic acids and proteins, were explored by MD simulations.[10-15]

Although MD simulations is powerful and efficient in general, it still has deficiency on

investigating specific systems and particular questions. Therefore, developing efficient and more proper simulation methodology for different objectives is continued alongside its applications.[16-18] One of the most widely known development is the hybrid QM/MM approach established by Martin Karplus, Michael Levitt and Arieh Warshel, leading them won the 2013 chemistry Nobel Prize. It combines the classical physics used in MD simulations and quantum physics for researching chemical processes in proteins. By applying this approach, users are permitted to implement accurate calculations via quantum mechanics on particles inside the binding site of protein which interact with the ligand; and the rest atoms are performed fast and simple calculations using molecular mechanics.[19-21] These achievements being appreciated and widely recognized by the world encourage computational chemists and biologists greatly to make more contributions on this modeling field. It also indicates the realization that 'Today the computer is just as important a tool for chemists as the test tube. Simulations are so realistic that they predict the outcome of traditional experiments'.[22]

With the continuous development on molecular modeling and simulation methodology, it allows researchers to choose an apt approach to study their own interested systems. In this thesis, MD simulations were employed to explore the interactions of protein-ligand complex. In chapter 2 and 3, the mechanisms of mutation induced drug resistance to anti-HIV drugs will be investigated using molecular dynamics simulations. Furthermore, Chapter 3 will highlight the significance of electrostatic polarization effect in protein-ligand binding studies. A modified MD simulations method, to reduce the defect of previous developed polarized force field in exploring protein-ligand interactions, will be discussed and employed to explore the basis of avidin-biotin binding in

Chapter 4. In the following of this chapter, the basis of molecular dynamics and fundamental theories involved in the works of this thesis will be concisely described.

1.2 Principles of molecular dynamics simulations

Molecular dynamics (MD) simulations is a computational approach for simulating the evolution in time of systems composed by interacting particles (e.g. molecules) at atomic level. In the classical MD simulations, atoms and bonds within models of biological systems are considered as classical objects, which follows the laws of classical mechanics. The Newton's second law of motion determines the dynamics of the systems. [23-25]

$$F_i = m_i \frac{d^2 r_i(t)}{dt^2} \tag{1.1}$$

$$F_i = -\nabla_i U(r_1, r_2, r_3, \dots, r_N) \tag{1.2}$$

where F_i represents the force acting on the i^{th} atom at the given time t, m_i defines the mass of the i^{th} atom and $r_i(t) = (x_i(t), y_i(t), z_i(t))$ stands for the position of the i^{th} atom. And the force exerted on the i^{th} atom at time t can be described as the gradient of the potential energy U(r_1, r_2,…, r_N) which is defined as a function of the atoms' coordinates in the system. (Eqn. 1.2) And more details about the important role of the potential energy U in MD simulations will be discussed in the late part of this section.

With the fundamental theory described above, the motion of all atoms is able to be simulated by numerically solving a set of the equations of motion. The trajectories as the result of MD simulations are composed of coordinates and velocities of atoms evolving over time. To propagate these vectors, integration algorithm, Verlet algorithm (Eqn. 1.3) [26], one commonly used algorithm is applied with the predefined initial positions and velocities. It allows the position of the i^{th} particle $r_i(t + \Delta t)$ at time $t + \Delta t$ to be located with its previous position at time *t*.

$$r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + a_i(t)\Delta t^2 + O(\Delta t^4) \quad (1.3)$$

$$a_i(t) = -(1/m)\nabla U(r(t)) \quad (1.4)$$

$$v_i(t) = \frac{r_i(t+\Delta t) - r_i(t-\Delta t)}{2\Delta t} + O(\Delta t^2) \quad (1.5)$$

As seen, the error of this algorithm is of order of Δt^4, as it is derived from approximated Taylor expansions. Though Verlet algorithm is straightforward and simple, it has limitation that the velocity is not explicitly generated which is essential for computing other physical quantities. Hence, the velocities of system are need to be calculated from the positions. (Eqn. 1.5) Nevertheless, the error is increased in this expression. To remedy this defect, improvement in integration algorithms has been made, such as leap-frog algorithm and velocity Verlet algorithm propagating velocities directly. In velocity Verlet algorithm[27], both position and velocity vectors at the next step can be derived by the same quantities at the current time. And this algorithm can be expressed in the following equations:

$$r_i(t + \Delta t) = r_i(t) + v_i(t)\Delta t + \frac{1}{2}a_i(t)\Delta t^2 \quad (1.6)$$

$$v_i(t + \Delta t) = v_i(t) + \frac{a_i(t)+a_i(t+\Delta t)}{2}\Delta t \quad (1.7)$$

The acceleration of the i^{th} atom at time $t + \Delta t$ is derived from the potential energy U which can be described as a function of the positions as mentioned above. (Eqn. 1.2) The potential function describes all the interactions in the system. Thus, determining a proper potential energy to mimic the real energy surface is significant in MD simulations.

In MD, empirical potentials are applied to describe physical and chemical features of the system, to simplify the calculation. The parameter sets are based on the experimental data or the derived results from quantum mechanical calculation. Force field, so called in MD simulations as the empirical potential function, contains the necessary functional form and parameters for imitating biological systems; and the function form of potential energy inside can be divided to bonded and non-bonded terms. (Eqn. 1.8) In a classical force field, such as an Amber force field, the total energy form is expressed in Eqn. 1.9, and each term in the equation are demonstrated in Figure 1.1.[23-24, 28-29]

$$E_{Total} = E_{bonded} + E_{nonbonded} \quad (1.8)$$

$$E_{Total} = \sum_{bonds} k_r(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2}[1 + \cos(n\phi - \Upsilon)]$$
$$+ \sum_{i<j} 4\varepsilon[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}] + \sum_{i<j} k\frac{q_i q_j}{r_{ij}} \quad (1.9)$$

In this expression, the bonded energy is represented by the first three terms while the last two terms compose the non-bonded energy term. The first term, stretching energy equation, represents the ability of bonds to stretch, based on Hooke's Law. In where, k_r is the force constant and r_0 is the equilibrium bond length, which are derived from experimental data or quantum mechanics calculations; and both of them are assigned distinct values according to the bonded atoms types, such as C-N, C-O, N-H, etc. The bending energy is defined in the second term, and as same as the stretching energy, it is adapted from Hooke's Law as well. k_θ is controlling the stiffness of the angle spring and θ_0 states the equilibrium bond angle. They are also acquired from empirical data or *ab initio* computing. Next, an equation with periodicity defines the torsion energy term, mainly for rectifying the rest terms to adjust the total energy in accordance with the experiment or *ab initio* calculation for different types of dihedral angle.

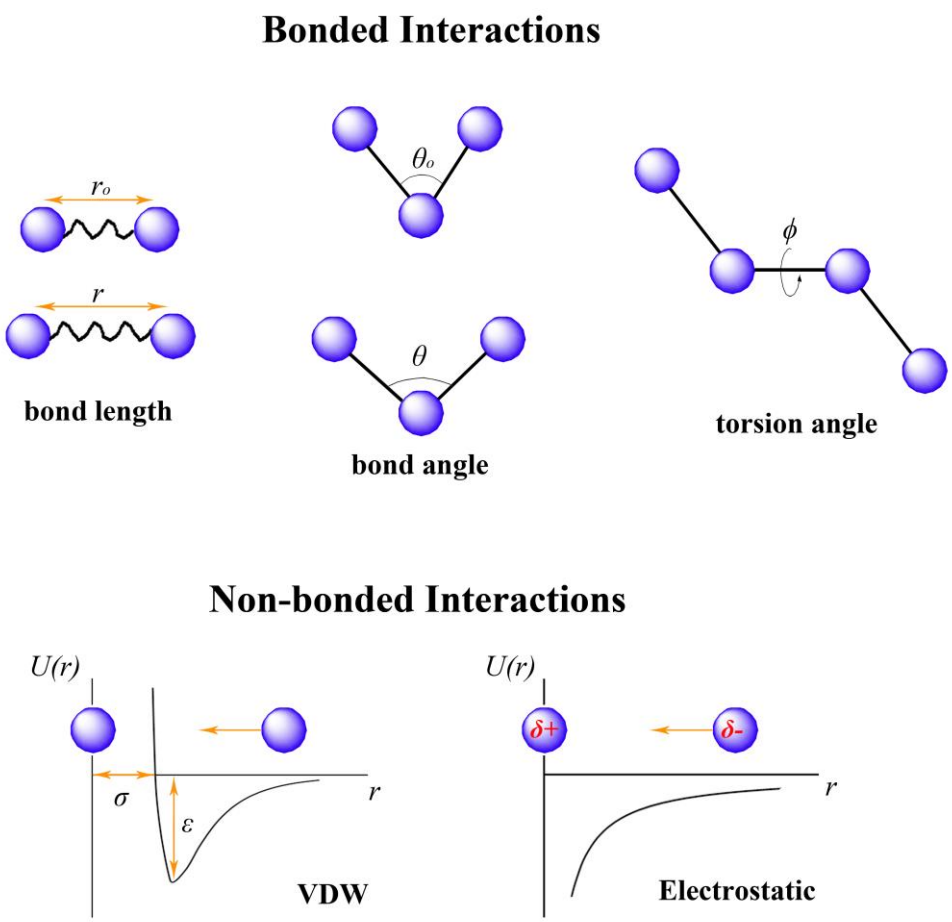


Figure 1.1 Diagrammatic illustration of terms in the potential energy function of force field listed in Eqn. 1.9. (adapted from Chapter 2 of Introduction of Computational Chemistry[30])

The following, van der Waals interactions and electrostatic interactions between two non-bonded atoms represent in the last two terms. The Lennard-Jones 6-12 equation[31] is applied to describe the repulsion/attraction between pairs of atoms (r_i, r_j) related to their distance r_{ij} $(r_{ij} = r_i - r_j)$; and ε controls the depth of the potential energy well. The electrostatic energy is modeled via Coulomb equation. q_i and q_j in the equation are the partial charges of i^{th} and j^{th} atom respectively, based on atom types, which are pre-assigned in the force field.

With the basic theories explained above, the processing path way in the classical MD simulations

program can be demonstrated in the following Schematic diagram. (Figure 1.2)

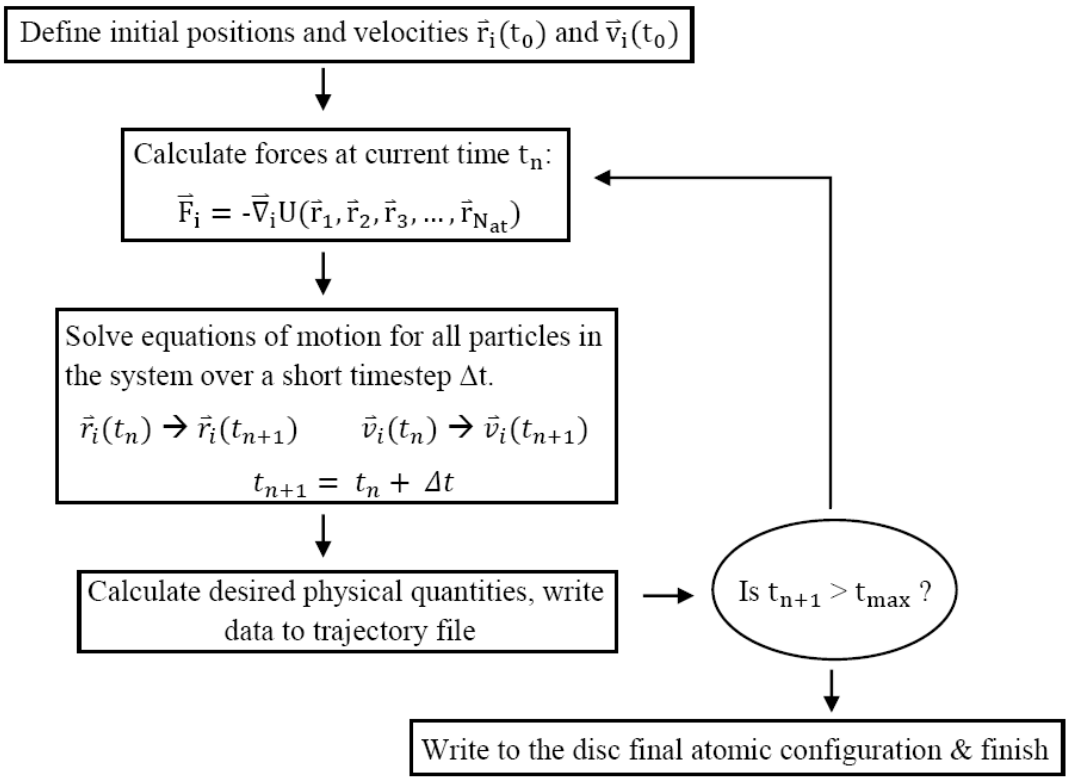


Figure 1.2 Schematic diagram illustration of basic MD procedure. (adapted from Ref.32)

1.3 overview of electrostatic polarization effect in MD simulations

As the potential energy equation (Eqn. 1.9) shown in the previous section, energy of electrostatic interaction is represented by Coulomb equation. The partial charges of the i^{th} and j^{th} atom (q_i and q_j) have significant influence on the value of electrostatic energy term. The charge of each atom is based on their amino acid types in the classical force field. For example, in Amber 03 force field, the atomic charges were acquired via fitting quantum mechanically calculated electrostatic

potential of dipeptide of each type of amino acid.[29] Due to the method applied, the atomic charges contained in the force field are amino acid-specific charges, ignoring the polarization effect between atoms; and consequently, distribution of the charge of amino acid is not permitted to vary correspondingly to its environment. Thus, it is unlikely to precisely describe the electrostatic environment of residues within various proteins.

To overcome this deficiency of the classical force field, researchers have developed a variety of polarizable or polarized force fields to include electrostatic polarization effect into MD simulations. Some of these polarizable force fields have been employed in the current available MD simulations packages, such as CHARMM, AMBER, and GROMACS. Nowadays, the methods comprising polarization effect in the force field commonly used include fluctuating charges model, Drude Oscillators, inducible dipoles and multipole moments.[33] In fluctuating charges model, the magnitude of atomic charge can vary along with the simulation; and the FQ/CHEQ force field developed based on this model implemented in the CHARMM program has been widely employed to simulate biological molecules.[33-36] Instead of applying point charge model, in Drude Oscillator model, a pair of charged particles is employed to mimic the polarizable atoms. One of the charge particles is on the centers of the polarizable atoms, while the other one is considered as a massless particle (so called, 'Drude particle') linked to the first one via a spring. And the polarization effect is accounted from the changes of dipole moment caused by the external environment which induces dislocation of the Drude particle.[33, 37-39] For the induced dipole method to include the polarization effect, the atomic charges are kept from a classical force field, at meanwhile, inducible point dipoles are introduced response to the electrostatic environment.[33,

[40-42] With usage of this method, the electrostatic energy term in classical force field is not applicable any more. Additional terms need to be considered for the newly generated interactions due to the presence of dipoles.[33, 40-42] Similar with induced dipole, in multipoles moments, higher order terms are used rather than monopoles to represent the polarization. AMBER ff02 and AMOEBA force field are developed based on these two models; and have been widely used in simulating biomolecules.[43-44]

Except for the models mentioned previously, many other algorithms have been developed to accurately calculate large systems.[45] Among these protocols, a method, namely polarized protein-specific charge (PPC) scheme to allow the polarization effect accounted into the atomic partial charges, has been developed by Zhang and his co-workers in 2008.[46] PPC scheme allows the electrostatic potential of specific protein to be calculated on quantum mechanics (QM) level by cutting proteins into individual amino acids through implementing molecular fractionation with conjugate caps (MFCC) approach.[47] Polarized protein specific charges can be fitted using the QM generated electrostatic potential. This method has been applied in a number of projects, including two projects in this thesis, to study the dynamics of protein systems and give-promising results.[47-53]

1.4 Principles in MFCC and PPC

Like aforementioned, in PPC scheme, MFCC approach is the foundation to realize quantum mechanical calculation of large protein systems. Due to the computational cost is exponential

scaling with the number of degrees of freedom of the entire system, full quantum mechanical computation on large proteins is hardly to achieve based on the current computing power and techniques. An alternative way, MFCC, has been developed. As one of the fragmentation methods, MFCC is able to reduce the computational cost of calculating large protein systems at QM level. It divides the protein into small fragments and computes each fragment sequentially, consequently, the cost scale linearly with the number of fragments.[54-58] The protocol of the MFCC approach is demonstrated in Figure 1.3.

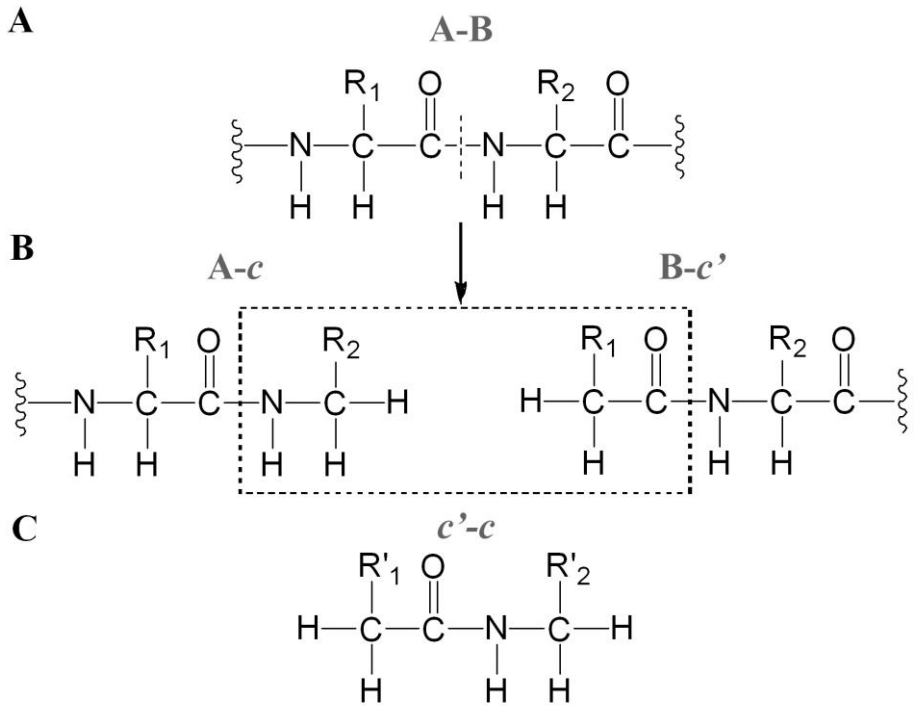


Figure 1.3 Molecular fractionation with conjugate caps (MFCC) approach. (A) peptide bond is cut between A and B; (B) conjugated caps are added forming fragment A-*c* and B-*c'*; (C) the "concap" *c*-*c'*. (from Gao et al.)[59]

In which, AB, representing a molecule, typically peptide or protein, is fractioned to individual residues (A and B) by clipping peptide bonds, and these bonds are capped with conjugate caps *c* or

c'. With using caps, the local environment of the peptide bond being cut is preserved, which acts as a similar role as the link atom in QM/MM method.[57-61] After adding caps at the cutting point, individual fragments, A-*c* and B-*c'*, are formed, then each fragment is calculated quantum mechanically, including the 'concaps' *c*-*c'*. The electron density of the original molecule is obtained by the following equation,

$$\rho_{AB} = \rho_{A-c} + \rho_{B-c'} - \rho_{c-c'} \tag{1.10}$$

where ρ_{A-c} and $\rho_{B-c'}$ describes the electron density of A-*c* and B-*c'* respectively, while $\rho_{c-c'}$ is the electron density of the concaps. This idea is generalized to the case of protein with N amino acids and the above equation can be promoted to Eqn. 1.11.

$$\rho = \sum_{i=1}^{N} \rho_i - \sum_{i=1}^{N-1} \rho_i^{cc'} \tag{1.11}$$

$$\mu = \sum_{i=1}^{N} \mu_i - \sum_{i=1}^{N-1} \mu_i^{cc'} \tag{1.12}$$

In which, the first term is the total density of all separated residues, while the sum of the density of all added concaps is represented in the second term. The electrostatic potential can be obtained in a similar way. (Eqn. 1.12)

Proteins is able to be computed on QM level without massive computational resources cost using MFCC approach. PPC scheme is realized via employing this idea with continuum-solvent model. In

this model, a charge distribution is utilized to describe the solute (protein), which is implanted in the cavity within the solvent, the polarizable dielectric medium.[62] At the time of the solvent being polarized by the solute, the solute is polarized by the reaction field which is formed by the discretization of induced charges on the cavity surface, until the system is equilibrated. The self-consistent reaction field is acquired by solving the Poisson-Boltzmann (PB) equation numerically. The procedure for generating PPC of a specific protein is demonstrated in Figure 1.4.[46]

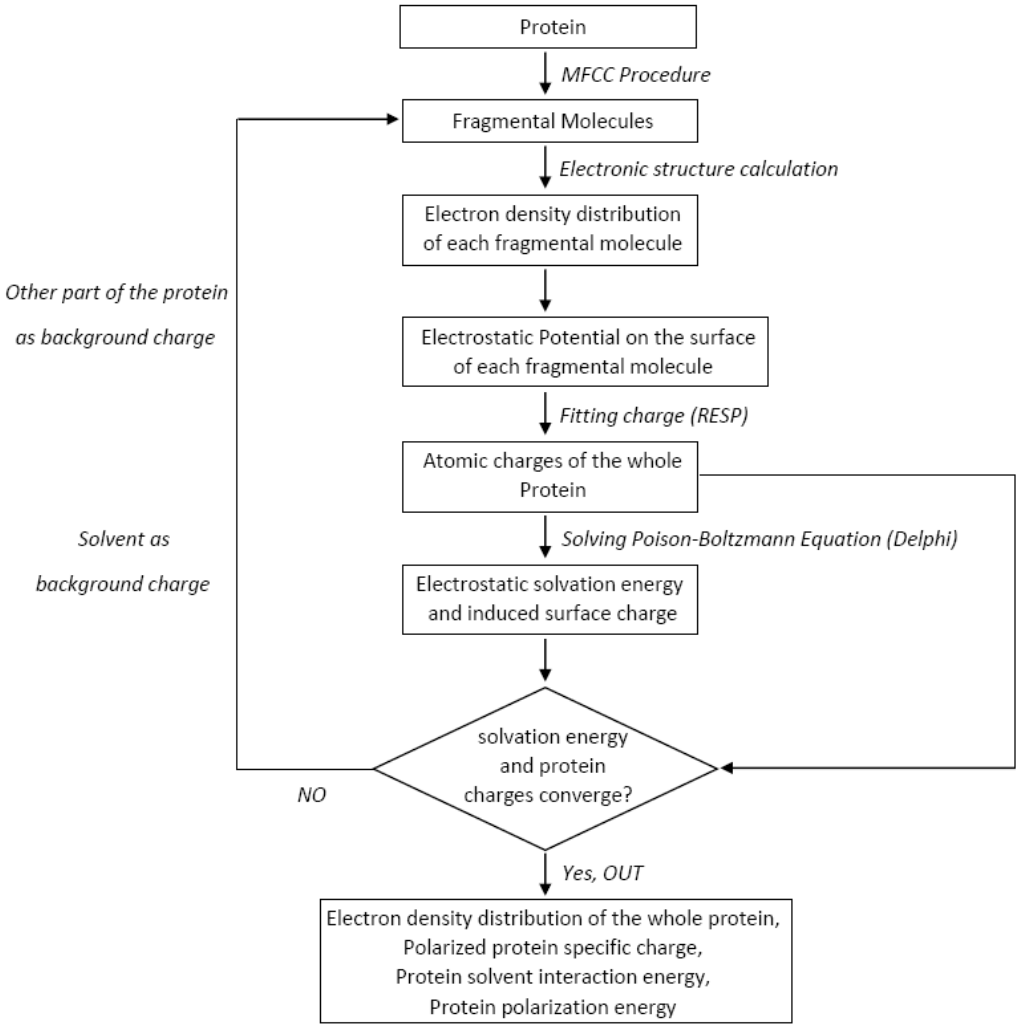


Figure 1.4 The common procedure of generating PPC for a specific protein. Figure is directly from Ref.46.

In which, the first step is the protein fragmentation using MFCC approach. Calculation on each capped residue from the last step is conducted on QM level in gas phase; and the electrostatic potential of the fragments can be obtained from the calculation. In the next step, Restrained Electrostatic Potential (RESP) method is applied to fit the atomic partial charge for each amino acid within the protein.[63-64] After the atomic charges are derived for the entire protein, the reaction field is derived by solving the Poisson-Boltzmann equation numerically using DELPHI.[65] And the induced surface charges on the boundary are generated to model the solvation effects. The discrete solvent surface charges and the charges of the other part of protein are treated as background charges in the quantum mechanical calculation of each fragment in the following circle and repeat the procedure mentioned above. The circulation will terminate until the solvation energy and charges of protein are converged.

While performing MD simulations, the PPC of protein is able to introduce the polarization effect force into the conventional field with direct replacement of the charges included in AMBER force field, at meantime, remaining the rest parameters, since in the PPC scheme, the atomic charges are derived consistently with the method implemented in AMBER force field. In the following chapters, both conventional force field and polarized force field will be involved to analyze protein-ligand interactions.

1.5 Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA)

MM-PBSA is one of the most popular approaches to calculate binding free energies of protein-ligand complexes using frames obtained from MD simulations. In this approach, continuum solvent model is applied and the free energy of a complex can be defined as following:

$$G = E_{MM} + G_{PBSA} - TS_{MM} \quad (1.13)$$

in which E_{MM} contains terms from conventional force fields, e.g. bond, angle, electrostatic, torsion angle and Van der Waals. And for estimating the non-bonded interactions, no cut-off is applied. Poisson-Boltzman equation is solved numerically to evaluate the solvation free energies. [66-68] And nonpolar free energy is estimated based on surface area calculation.[69] The solvation free energy and nonpolar free energy build up the G_{PBSA} term. Both E_{MM} and G_{PBSA} are derived based on a set of snapshots extracted from trajectories of MD simulations. The binding free energy of a ligand to a protein can be calculated as following:

$$\Delta G = G_{comp} - G_{rec} - G_{lig} \quad (1.14)$$

each term in this equation is derived from the simulation of the protein-ligand complex only. The G_{rec} and G_{lig} term are obtained by extracting protein's coordinates and ligand's coordinates from the trajectory respectively. In this approach, the conformations of protein and ligand is assumed un-changed upon binding. Even so, lots of remarkable results have been achieved by using MM-PBSA approach.[70] It possesses some attractive features comparing with FEP and TI method, especially while the ligands are significantly different with each other in structure. And its force field like equation allows us introducing PPC into the binding free energies calculation conveniently. Therefore, in this thesis, all binding free energies calculations are performed by MM-PBSA approach.

Chapter 2

Understanding the basis of G140S-Q148H double mutation induced drug resistance in HIV-1 integrase using molecular dynamics simulations

As introduced in Chapter 1, molecular dynamics (MD) simulation is making big differences in the field of biological systems and being advantageously informative on molecular motion. When analyzing protein-ligand interactions, applying MD simulations can provide information that is not approachable by stationary structures. Binding pocket is easily to be examined using MD method, where particulars with respect to the ligand binding and release procedure, and details on dynamic protein-ligand interactions will be produced. Moreover, it helps to compare different ligands through the relative binding free energy calculations derived from MD simulations, which is able to obtain insights on the binding mechanism. MD method has been extensively applied in studying receptor-ligand contacts and playing a significant role in new drugs design. In this and the next chapter, MD simulations are carried out to explore the mechanism of mutation induced drug resistance in HIV-1 integrase and HIV-1 protease, respectively. Conventional MD is employed in this chapter to shed some light on how the drug resistance is produced by G140S-Q148H double mutation.

2.1 Introduction

Acquired immunodeficiency syndrome (AIDS) is one of the human immune system disease led by infection with human immunodeficiency virus (HIV).[71-72] About 35.0 million people were suffering with HIV/AIDS globally according to the World Health Organization (WHO), and this number is still increasing.[73-74] Approximately 1.5 million people died of AIDS related illness globally in 2013, where the number was decreased by 35% comparing with the number in 2005.[73-74] It has been known, to a great extent, as a severe threat to human immune system, an essential protection for our bodies to against other diseases. Though so far there is no vaccination or cure available, fortunately, the life time of the sick can be prolonged and the deterioration of the disease can be slowed down by employing the antiretroviral treatment. Among the approved FDA antiretroviral drugs, one can observe inhibitions mainly towards the function of reverse transcriptase (RT), integrase (IN), protease (PR) and etc... We will concentrate on IN in the following of this project.

To begin with, let's take a look at the replication process of HIV-1 to know better how integrase works in the life cycle of HIV-1 before studying its inhibitors.[75-76] In Figure 2.1, it shows that at the beginning of the life cycle of HIV-1,[77] HIV-1 acts together with cells by the bound of the glycoproteins gp120/gp41 of virus with CD4 and CCR5 or CXCR4 co-receptors of the aimed cell.[75, 78-79] Then viral and cellular membranes fused with each other facilitated by gp41, soon after the viral core is released into the cytoplasm of the target cell. Viral RT inversely transcribes the viral RNA, which is two copies of single-stranded RNA, into a double-stranded DNA.[75] Next, the viral DNA is processed by integrase, combining together within PIC, known as pre-integration complex

being a large nucleoprotein complex. Following, PIC is transferred into the host nucleus and IN is there to catalyze the integration of the DNA into the host genome (strand transfer, ST).[75, 80] Genomic RNA or messenger RNAs (mRNAs) are exported after transcribing the viral DNA. The mRNAs are then translated into viral polyproteins and cleaved to generate the mature protein components of an infective HIV virion by the HIV-1 protease.[75, 81] The generated viral protein and RNA are assembled at the membrane of the host cell and the host cell releases the newly created infective virus by budding. As explained above, in the process of replication of HIV-1, integrase participates in several steps among the replicative process. Hence, it is essential to recognize at which step of the process the integrase inhibitor causes the function of integrase to be interrupted.

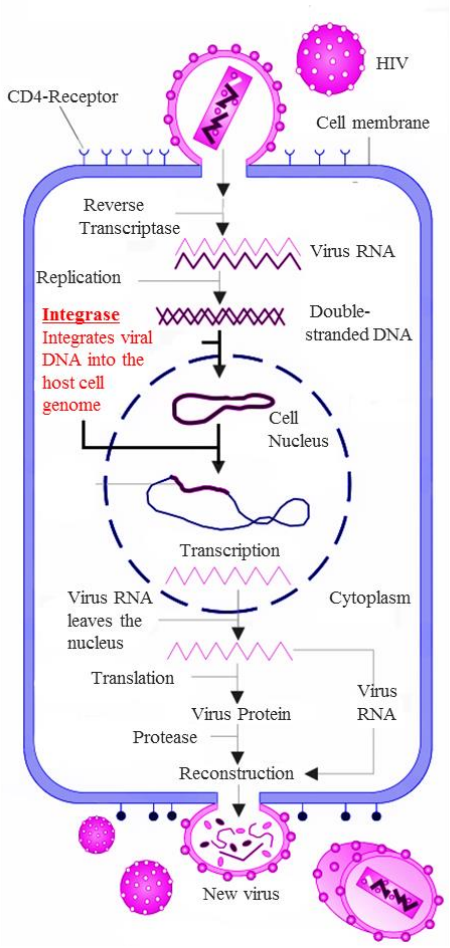


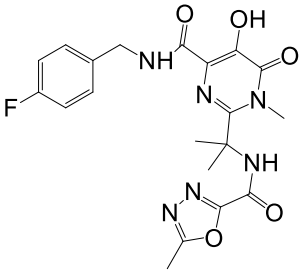
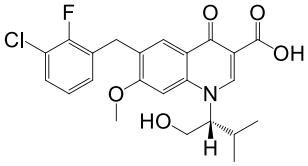
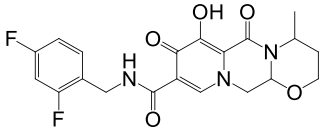
Figure 2.1. HIV-1 life cycle. The strand transfer step is highlighted with red color. (adapted from Ref.77)

Next, let's take a closer look at the integration process, among which integrase is enrolled in two steps of the process, 3'-processing and ST (strand transfer). In the cytoplasm of the host cell, when the reverse transcription is done, viral integrase cleaves the dinucleotide GT with help of a water molecule at two 3' ends of the newly generated double strand viral DNA through 3'-processing.[76, 82-83] The viral DNA is processed within the PIC, composing by viral integrase, viral co-factors, reverse transcribed viral DNA and co-factors of the host cell.[84-86] Then PIC is transferred into nucleus followed by the second step, strand transfer (ST) reaction. In the ST reaction, the cleaved reverse transcribed viral DNA is integrated into the host DNA. The PIC is directed by co-factors (e.g. LEDGF/p75) of the host cell to target the cellular DNA.[87-89] And the free 3'-OH end, generated in 3'-processing step, of the viral DNA is treated as the nucleophile to attack the host DNA.[75, 84]

Currently, all IN drugs approved by FDA have a common inhibition mechanism, which is blocking binding site for the host DNA to halt the strand transfer reaction, so called integrase strand transfer inhibitors (INSTIs).[75] There are only three drugs have been approved by FDA, including the first generation drug (raltegravir, also called RAL and elvitegravir, also known as EVG) and the second generation drug (dolutegravir, also called DTG), for their usage in the recommended treatment for HIV infection, the antiretroviral therapy. (Table 2.1) Unfortunately, drug resistance towards RAL and EVG are observed along with their usage in the antiretroviral therapy, as mutations appeared in HIV-1 IN.[90-94] Among these mutations, double mutation at residue 140 and 148 is the most popular combination. The G140S-Q148H double mutant generates up to 400 fold decrease on activity of the first generation INSTIs. On the other hand, DTG, as the second generation INSTI, shows its ability in evading the drug resistance caused by this double mutation.[90-94] Hence, two questions raised up

and need to be solved using molecular dynamics simulations in this study: (1) how G140S-Q148H mutated HIV-1 IN obtains drug resistance to the first generation drug; and (2) why DTG can combat this resistance. Thereafter, understanding the mechanism of drug resistance will be of great help in developing new INSTIs to overcome this drug resistance. In this study, RAL as the first generation inhibitor and DTG as the second generation INSTI will be examined.

Table 2.1. Structure of FDA approved INSTIs

Name	Company	Structure
Raltegravir (RAL) **MK-0518**	Merck & Co.	
Elvitegravir (EVG) **GS-9137** **JTK-303**	Gilead Sciences	
Dolutegravir (DTG) **S/GSK-1349572**	Shionogi-GlaxoSmithKline Pharmaceuticals, LLC (USA)	

To examine drug resistance via molecular dynamics simulations, it is necessary to have the structure of complexes of HIV-1 integrase binding with INSTIs. Unfortunately, the crystal structure of HIV-1 IN binding with DNA and its inhibitors has not been resolved yet in the present. Five years ago, the crystal structure of the prototype foamy virus (PFV) IN interacting with INSTIs and a short

oligonucleotide substrate which mimics the viral DNA end after 3'-P has been resolved by Dr. Hare's group from Imperial College London.[95] It furnishes the first three-dimensional structure of an active IN binding site, and it becomes possible to construct the structure of HIV-1 IN-INSTIs complexes using homology modelling approach based on the similarity of these two integrases.[75] (Figure 2.2) By comparing, IN of PFV and HIV-1 is found different as PFV IN has one more domain, the N-terminal extension domain (NED), but the good part is this domain has no direct affection on the active site of integrase.[82, 96] And rests of the PFV and HIV-1 integrase protein are similar with each other, besides, the fact that FDA approved INSTIs are found to be able to inhibit PFV integrase *in vitro* which supports the similarity of the integrase.[75] Crystal structures of PFV IN bound with INSTIs through two divalent metal ions together with viral DNA are able to be found in Protein Data Bank (PDB). By investigating on the available structures, the flexible loop is playing an important role in binding with RAL, EVG, DTG and other potential inhibitors.[75] And since the double mutation G140S-Q148H is locating at two end of the flexible loop, it is so called the "140s' loop".[97] Through aligning the sequence of the 140s' loop section with several retroviral IN, PFV IN appears 60% similarity on amino acids with integrase of HIV-1.[75] (Figure 2.2B) The research via applying the HIV-1 IN core domain illustrates that the flexible loop of HIV-1 and PFV IN are highly similar with each other.[75] It proves that INSTIs-complexed PFV IN is a proper template for building the model of HIV-1 IN-INSTIs complexes.[75] As such, in present, the model of HIV-1 IN binding to its inhibitor with viral DNA in presence is constructed using the crystal structure of PFV IN-INSTI complex as template in literature. (Figure 2.2C)

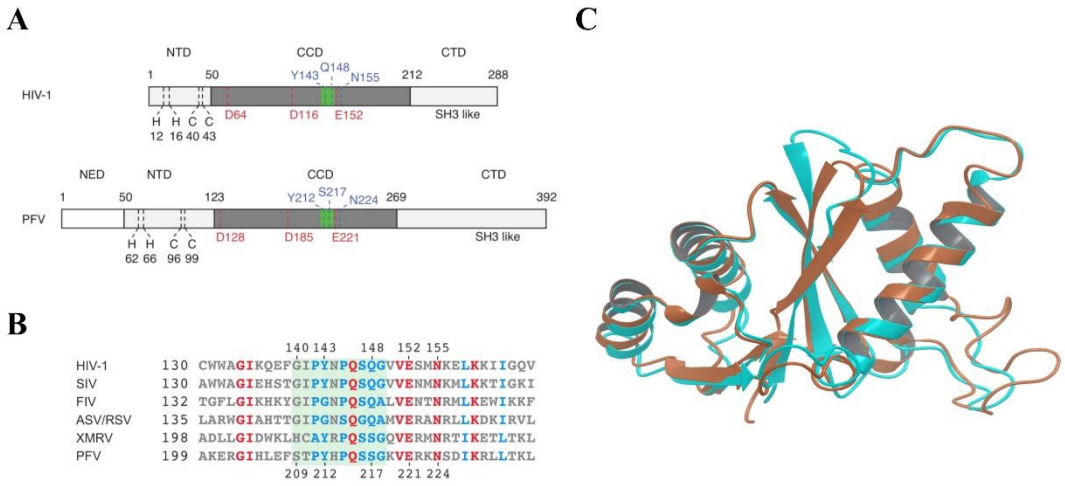


Figure 2.2. (A) Illustration of comparison of HIV-1 and PFV IN's primary structures; (B) Amino acid sequence alignment of the loop part of different retroviruses; (C) Alignment of catalytic core domain (CCD) in HIV-1 IN model (brown) and PFV IN crystal structure (aquamarine) (A and B extracted from Ref.75)

From the crystal structure of PFV IN-INSTIs complex, it is worth noting that there are two divalent metal ions involving in the binding of integrase and INSTIs.[95] These two metal ions form coordinate bonds with the DDE motif of Asp128, Asp185 and Glu221, and also with the inhibitors. Together with three water molecules, chelate forms, in which the metal ions acting as the bridge in between of integrase and its inhibitor. As same as in PFV IN, the divalent metal ions have the same function in HIV-1 IN-INSTIs complex. (Figure 2.3)

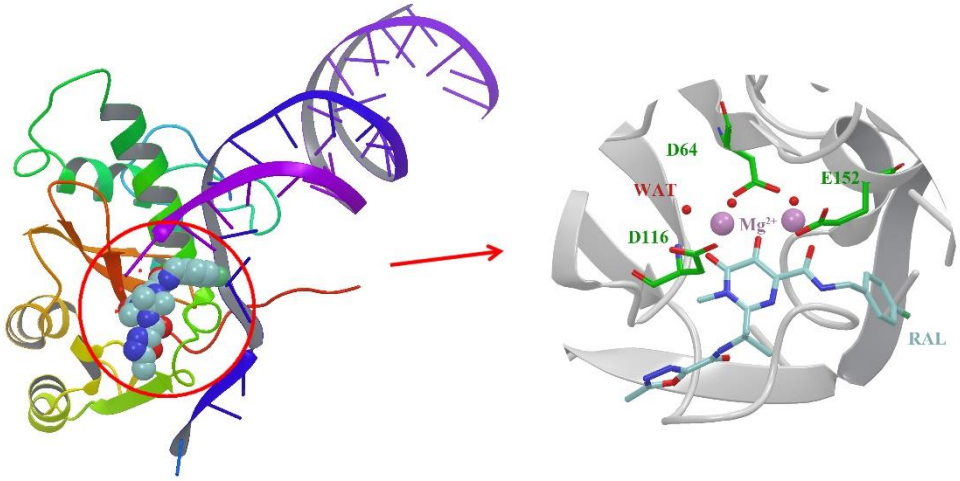


Figure 2.3. Model of CCD of HIV-1 IN binding with RAL constructed by homology modeling method with crystal structure of PFV IN-RAL complex as template (left); Active site of HIV-1 IN binding with RAL (right).

Magnesium ions are applied in the model of HIV-1 IN as the same with in crystal structure of PFV IN. The character of these two catalytic Mg^{2+} ions has been discussed in previous studies, which is important in both integration and inhibition process.[98] In the integration process, the catalytic metal ions bind to viral DNA to allow the 3'-processing step to be executed and hold the host DNA in the strand transfer reaction to make viral DNA integrating into the host DNA.[98] In the proposed mechanism of inhibition process, INSTIs interact with the two metal ions after the 3'-processing step to hinder the host DNA approaching leading to their inhibition function.[98] (Figure 2.4) Since the metal ions plays an essential role in the activities of integrase, it is important to maintain them at their own site to retain functioning throughout the MD simulations. It should be dealt with carefully instead it wasn't in previous research.[99]

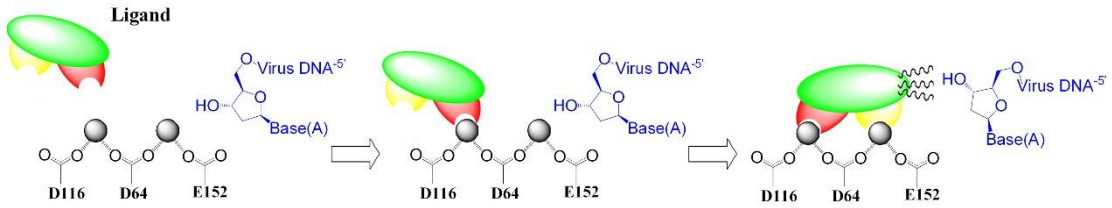


Figure 2.4. One of the proposed mechanistic model of INSTIs, in which, inhibitor binds to integrase of HIV-1 through interacting with the divalent metal ions to block the binding site of the host DNA. (adapted from Ref.98)

Another point need to be noted, as reported, it has possibility that RAL possesses a negative charge by deprotonation of the hydroxyl group while binding with HIV-1 IN.[100] And DTG may share the same situation, as both of RAL and DTG are provided in salt format.[101] It is worth to discuss which protonation state should be selected for INSTIs, protonated or deprotonated. Test run of simulations were conducted to confirm the protonation state of INSTIs before studying the mechanism of the double mutation induced drug resistance. The Mg^{2+} ions were dislocated caused by the repulsion between Mg^{2+} ions and the proton of the hydroxyl group of RAL, while it was maintained well in deprotonated state RAL-complex. These data will be further discussed later in the result section. As the negative charge of RAL is not only able to stabilize the Mg^{2+} ions, but also increase the solubility of RAL, the deprotonated RAL is employed in this research.[100] And the same arrangement is made for DTG, deprotonated state of DTG is applied in the following simulations.

2.2 Method

2.2.1 Initial structures

CCD of WT HIV-1 IN model is established utilizing homology modelling method. Firstly, MODELLER 9.9 program is employed to build HIV-1 IN structures with PFV structures as templates.[95, 102] The PFV structures combined with RAL and DTG (PDB ID: 3OYA[95] and 3S3M[93], respectively) are used to generate the structures of HIV-1 IN RAL and DTG complexes respectively. Subsequently, the INSTIs, Mg^{2+} ions and viral DNA, are introduced to the binding pocket of HIV-1 from the corresponding complexes of PFV using MAESTRO by locating residues position of Asp64, Asp116, and Glu152 in the active site. And the structures of G140S/Q148H double mutant of HIV-1 IN are produced by substituting residues 140 and 148 from previous obtained WT HIV-1 IN models.

2.2.2 Molecular dynamics simulations

HIV-1 IN-RAL and HIV-1 IN-DTG are the models of complexes obtained from previous step, where viral DNA and two Mg^{2+} ions, the parameters of the integrase, are described with AMBER03 force field[29]. The structures of RAL and DTG are optimized at AM1 level and the electrostatic potential are calculated at B3LYP/6-31G* level using Gaussian09 suit[103]. The atomic charges of

RAL and DTG are fitted using restrained electrostatic potential (RESP) with Antechamber program in Amber10 package[25]. General Amber force field (GAFF)[104] is applied to obtain the parameters of RAL and DTG in simulation. With using Leap module, the models of complexes are added with hydrogen atoms and solvated in the periodic box containing approximately 13,500 TIP3P water molecules[105] with a minimum distance of 10 Å between the complex and the surface of box. Sodium counter ions are introduced to the system to maintain the neutrality. Thereafter, by conducting the method employed in Xue and co-worker's study[99], minimization is implemented to the system. Step one is to conduct energy minimization for 1000 cycles through steepest descent method, then for 3000 cycles via conjugate gradient method with restraint of 500 kcal/mol/Å^2 on all heavy atoms of the complex molecule. Second, energy minimization for 1000 steps followed by a 50 ps heating procedure from 0 to 300 K is conducted by restraining all heavy atoms of the complex with less restraint weight at 5.0 kcal/mol/Å^2. Extra energy minimization phases are following with three steps, in which decreasing restraints from 2.0, to 0.1, to 0.05 kcal/mol/Å^2 are employed for 1000 cycles. Following, still with decreasing restraints weight from 1.0 to 0.5 then to 0.1 kcal/mol/Å^2, three short MD equilibrations are applied for 50 ps respectively. Subsequently, NPT simulation is executed for 5ns with a time step of 2fs, and the temperature of the system is regulated by Langevin[106] dynamics with 4.0 ps^{-1} collision frequency. The long-range electrostatic interaction is treated utilizing the Particle Mesh Ewald (PME) method[107], meanwhile, a 12 Å cutoff is applied to treat non-bonded interaction. Bond involving hydrogen atom is restrained by SHAKE algorithm[108]. Three MD simulations for each complex are performed by using different ig number to ensure obtaining different trajectories via employing SANDER program in AMBER10 package.

2.2.3 Binding free energy calculation

The binding free energy of the HIV-1 integrase-INSTIs complex is represented through the Eqn. 2.1.

$$\Delta G_{bind} = G_{complex} - G_{receptor} - G_{ligand} \quad (2.1)$$

where $G_{complex}$, $G_{receptor}$, G_{ligand} stands for, respectively, the free energy of complex, receptor and ligand. ΔG_{bind} was obtained by implementing MM/PBSA method. The following Eqn. 2.2 illustrates the calculation of the binding free energy.

$$\begin{aligned} \Delta G_{bind} &= \Delta G_{pbsa} - T\Delta S_{solute} \\ &= \Delta E_{MM} + \Delta G_{sol} - T\Delta S_{solute} \\ &= \Delta E_{ele} + \Delta E_{vdW} + \Delta G_{pol} + \Delta G_{nonpol} - \mathrm{T}\Delta S_{solute} \end{aligned} \quad (2.2)$$

where ΔG_{sol} composed by ΔG_{pol} and ΔG_{nonpol}, describing the polar and non-polar solvation energy of the system. The entropy of solute is defined by $T\Delta S_{solute}$. ΔE_{vdW} and ΔE_{ele} represents Van der Waals and electrostatic interaction energies respectively.

The following Eqn. 2.3 defines the difference of binding free energy among WT and double mutant HIV-1 integrase.

$$\Delta\Delta G = \Delta G_{WT} - \Delta G_{G140S/Q148H} \quad (2.3)$$

Where ΔG_{WT} and $\Delta G_{G140S/Q148H}$ describes the free energy of binding INSTIs to the WT and double mutated HIV-1 IN respectively. ΔΔG represents the difference of binding free energy between wild type-INSTIs complex and the double mutant-INSTIs complex.

To perform the calculations, 5000 snapshots are captured in the 5 ns simulations at 1ps intervals. The solvation energy is obtained through the PBSA module with setting the grid spacing to 0.25 Å. Equation $G_{nonpol} = \gamma SAS + \beta$ gives the nonpolar solvation energy, where γ = 0.00542 kcal/(mol ·Å^2) and β= 0.92 kcal/mol.[126] Entropy loss during protein-ligand binding is obtained by employing the covariance matrix method[109-113] with the PTRAJ module in Amber software package[114]. For each system, 5000 snapshots captured at an 1 ps interval from the 5ns of the trajectory are taken for calculation of entropy loss.

2.3 Results and discussion

2.3.1 Location of the catalytic magnesium ions in the simulated structure and stability of system

As discussed in introduction section, magnesium ions are playing an important character in the integration and inhibition processes. It is necessary to maintain the position of the catalytic Mg^{2+}

ions at their own site throughout the MD simulations. A 5 ns test run has been conducted using both deprotonated and protonated RAL to confirm the protonation state of INSTIs in this research. In order to evaluate the stability of the Mg^{2+} ions in both systems, distance between these two metal ions are plotted in Figure 2.5. It is clear that by applying protonated INSTIs position of the Mg^{2+} ions is hard to be maintained, while it is retained well in deprotonated RAL bound with HIV-1 IN (Mg^{2+} ions distance in crystal structure: 3.68 Å). The repulsion between the proton of hydroxyl group of RAL and Mg^{2+} is suspected as the reason of the separation of the catalytic metal ions. Therefore, deprotonated INSTIs are employed in the simulations to avoid the occurrence of the separation of the magnesium ions to keep their bridging function in inhibition process.

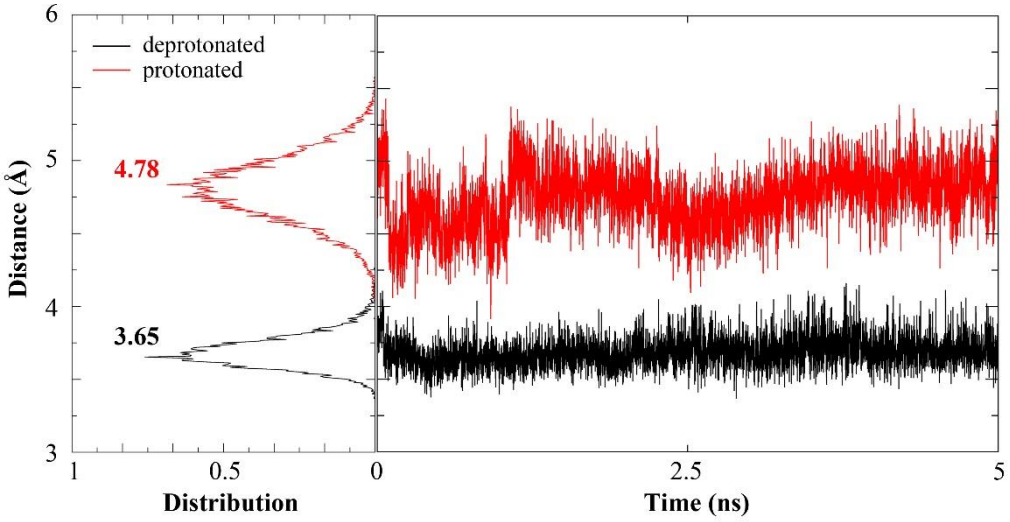


Figure 2.5. Distance between two catalytic Mg^{2+} ions with protonated and deprotonated RAL in present respectively. Right side is the change of distances during simulation; left side is the distribution of distance in the test simulation while binding with different protonation state RAL.

By applying deprotonated RAL and DTG, the simulated complexes models of WT HIV-1 IN-RAL and -DTG are extracted to compare with the crystal structures of WT PFV IN-RAL and -DTG complexes respectively to confirm the position of Mg^{2+} ions after simulation. (Figure 2.6) As shown, the position of Mg^{2+} ions in the simulated structures (Figure 2.6B) have no significant change

according to the crystal structures (Figure 2.6A). In order to obtain accurate and concise information about the position of the ions, distributions of the length of representative coordinate bonds between the ions and IN or INSTIs are plotted in Figure 2.6C. In Figure 2.6, it demonstrates that the length of these bonds are kept steady around the experimental data. In addition, the proper distances between the Mg^{2+} ions in WT IN-RAL and WT IN-DTG are preserved at 3.680 Å and 3.870 Å while the experimental data are 3.825 Å and 3.741 Å respectively. Since these simulations are able to maintain the position of magnesium ions to retain their function as the bridge of HIV-1 IN and INSTIs, analysis of mechanism of the double mutation induced drug resistance on the INSTIs is conducted based on the obtained trajectories.

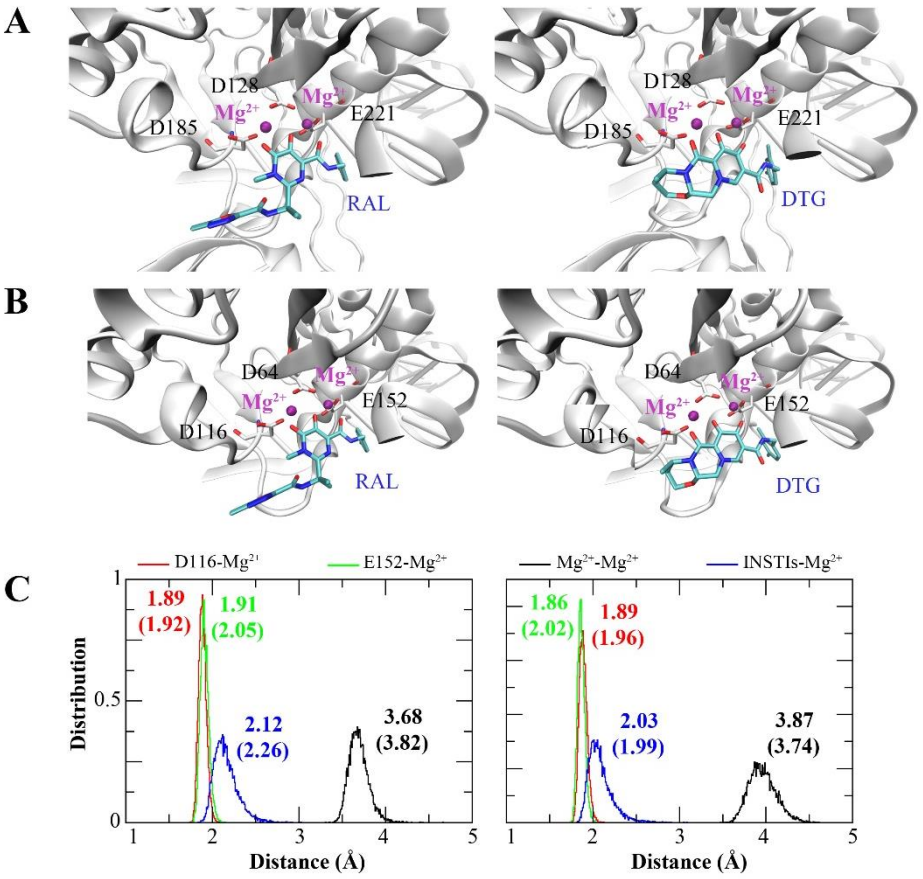


Figure 2.6. (A) crystal structures of active site of PFV IN bound with RAL and DTG respectively; (B) simulated structures of active site of HIV-1 IN model bound with RAL and DTG respectively; (C) distribution of selected distances related to position of the catalytical magnesium ions.

In order to evaluate the stability of the complexes of WT HIV-1 IN-RAL, -DTG and G140S/Q148H HIV-1 IN-RAL,-DTG, in each trajectory, value of root-mean-square deviations (RMSD) is calculated for each snapshot of the three production phase of MD simulations in each system. RMSD of HIV-1 IN CCD, viral DNA and INSTIs are plotted as a function of time in Figure 2.7. Based on the plotted RMSD, all trajectories remains stable during the simulations indicating that the systems have reached stability, also no significant conformational change is observed in the trajectories. The 5 ns production phase of each simulation is utilized for the following energetic and structural analysis.

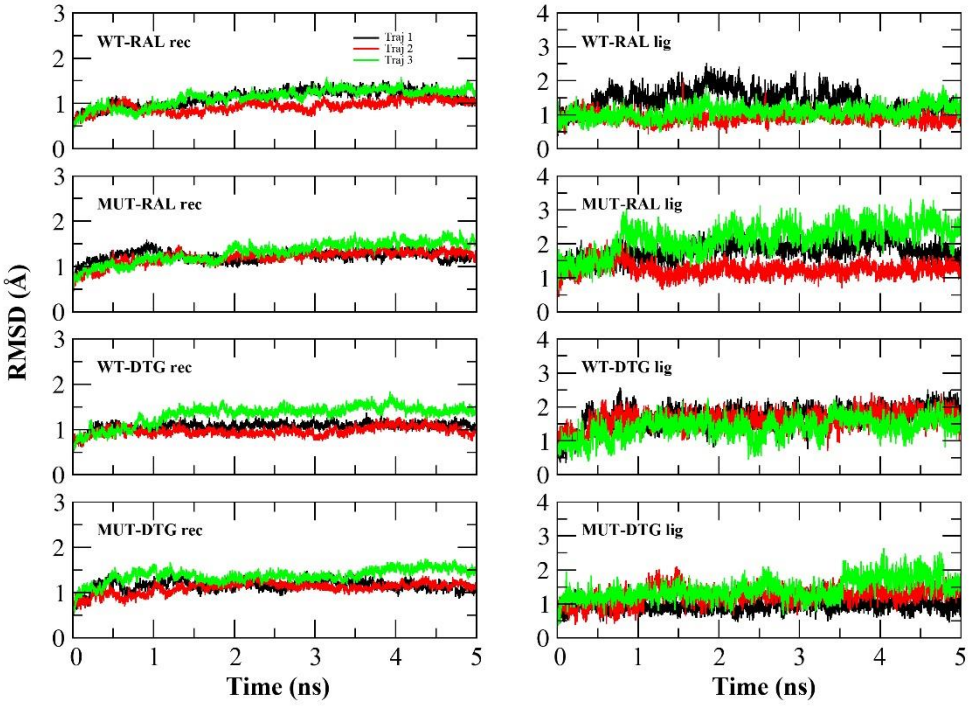


Figure 2.7. RMSD of simulated systems in multiple trajectories. Left side shows the backbone carbon RMSD of integrase CCD, and right side shows the RMSD of INSTIs.

2.3.2 Thermodynamic analysis

The trajectories extracted from simulations of each system are utilized to calculate binding free energies with MM-PBSA method. Since three trajectories are conducted for each complex, the binding affinities of RAL and DTG to integrase are calculated by averaging the binding free energies obtained from the multiple trajectories. In the MM-PBSA calculation, integrase, the Mg^{2+} ions and the viral DNA, are together treated as receptor while the INSTIs as the ligand. As expected, the binding affinities of RAL and DTG to the integrase are decreased in the double mutants than in WT, which is calculated using Eqn 2-3, tabulated in Table 2.2. For RAL, the theoretical value of the difference in binding energy ($\Delta\Delta G$) between bound with WT HIV-1 IN and G140S-Q148H HIV-1 IN is 3.91 kcal/mol, which is close to the experimental obtained value of 3.6 kcal/mol. As for the second generation drug, the calculated value (1.24 kcal/mol) shows less affection by the G140S-Q148H double mutation of HIV-1 IN, agreed with the experimental value of $\Delta\Delta G$ (1.74 kcal/mol). Since the average values reach good agreement with the experimental data, all analyses in the following will be based on the average data to explore the molecular details of how the G140S-Q148H double mutation of the integrase leads to its resistance to RAL and how DTG combat with this resistance.

Table 2.2. Theoretical and experimental obtained binding free energy of RAL and DTG binding to the WT and G140S-Q148H HIV-1 IN (Number in parentheses are standard error.)

RAL		WT			G140S-Q148H		ΔΔG
(Kcal/mol)	ΔH	-TΔS	ΔG	ΔH	-TΔS	ΔG	
Traj 1	-106.95	49.11	-57.84	-99.34	48.88	-50.46	
	(0.10)	(0.16)	(0.19)	(0.10)	(0.20)	(0.22)	
Traj 2	-104.04	48.02	-56.02	-100.45	47.8	-52.65	
	(0.10)	(0.33)	(0.34)	(0.10)	(0.16)	(0.19)	
Traj 3	-102.69	48.24	-54.45	-101.49	48.04	-53.45	
	(0.12)	(0.23)	(0.26)	(0.10)	(0.19)	(0.21)	
Ave.			-56.10			-52.19	**3.91**
			(0.47)			(0.36)	**0.59**
Exp.							**3.6**
DTG		WT			G140S-Q148H		ΔΔG
(Kcal/mol)	ΔH	-TΔS	ΔG	ΔH	-TΔS	ΔG	
Traj 1	-65.36	42.88	-22.48	-69.53	42.2	-27.33	
	(0.10)	(0.19)	(0.21)	(0.08)	(0.16)	(0.18)	
Traj 2	-74.87	42.48	-32.39	-67.93	42.9	-25.03	
	(0.08)	(0.20)	(0.22)	(0.10)	(0.16)	(0.19)	
Traj 3	-69.67	42.83	-26.84	-69.07	43.42	-25.65	
	(0.09)	(0.21)	(0.23)	(0.09)	(0.18)	(0.20)	
Ave.			-27.24			-26.00	**1.24**
			(0.38)			(0.33)	**0.50**
Exp.							**1.74**

In order to acquire detailed information on the contribution of each residue to the binding affinity, the overall binding free energy is decomposed into individual residues and the difference between WT and double mutant ($\Delta\Delta G_{per\ res} = \Delta G_{G140S-Q148H\ per\ res} - \Delta G_{WT\ pre\ res}$) are depicted in

Figure 2.8. As shown, the strength of interactions between RAL and residues of the 140s' loop, especially residues Tyr143 and Pro145, are decreased in the RAL-complexed IN double mutant. In particular, the contributions from Tyr143 and Pro145 decreases by 0.64 kcal/mol and 0.96 kcal/mol respectively. On the other hand, weakened interactions between DTG and these residues are unapparent suggesting DTG is able to against with the drug resistance caused by the double mutation comparing with RAL. Besides the difference in relative values in RAL and DTG systems, the absolute value of the decomposed binding free energies for Tyr143 and Pro145, are much lower while binding with RAL than DTG. For RAL-complexed WT, the contribution made by Tyr143 and Pro145 is -3.171 kcal/mol and -2.984 kcal/mol respectively, while it is -0.617 kcal/mol and -1.826 kcal/mol in DTG bound WT. It indicates that RAL possesses stronger contacts with the 140s' loop than DTG and the dynamical changes of the loop would exert more serious effects on RAL's binding.

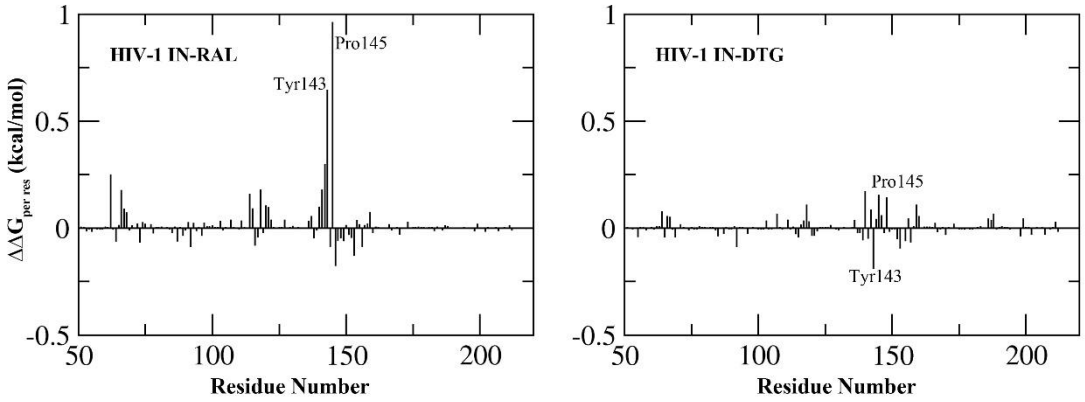


Figure 2.8. Difference of binding free energy of RAL and DTG to individual residues of HIV-1 IN variants. ($\Delta\Delta G_{per\ res} = \Delta G_{G140S-Q148H\ per\ res} - \Delta G_{WT\ pre\ res}$)

2.3.3 Changes in the binding site of HIV-1 IN upon G140S-Q148H

To characterize the effect of G140S-Q148H mutation on the conformational change of CCD of integrase, the atomic fluctuation of individual residue of WT and double mutated integrase are represented by the isotropic temperature factor (B-factor) of the protein at residual level. (Figure 2.9) It is observed that the fluctuation of residues does not vary dramatically in the double mutated HIV-1 IN comparing with in WT HIV-1 IN in both RAL- and DTG-complexed integrase except at the part of 140s' loop which composes active site of HIV-1 IN, while bound to RAL. The B-factor of the loop increases when RAL bound to G140S-Q148H HIV-1 IN while the B-factor of the loop remained the same in DTG-complexed integrase. A significant increase of B-factor can be observed for Tyr143 and Pro145 in RAL-complexed integrase double mutant, which makes major contribution in the binding affinity. Higher flexibility of these residues will produce negative effect on binding with RAL, being consistent with the observation in loss of binding affinity in the last section.

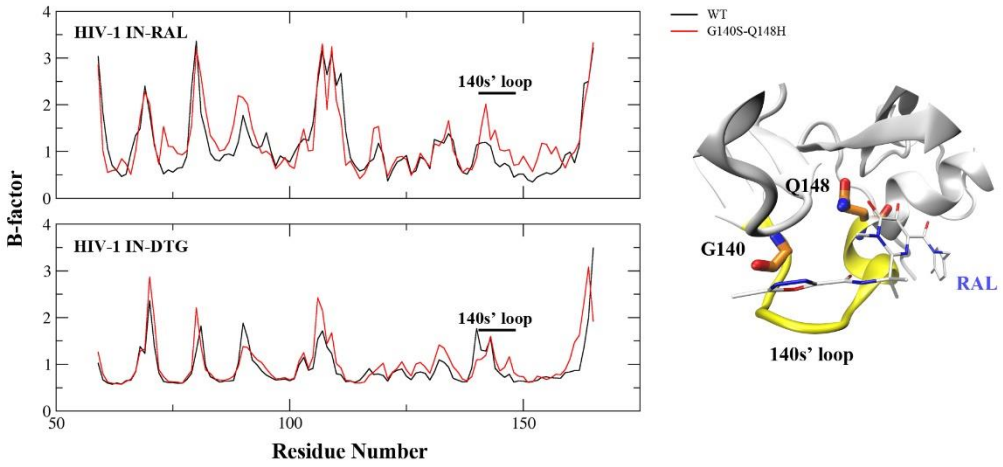


Figure 2.9. Left: (top) Fluctuation of the Cα atoms of integrase of RAL-complexed wild type HIV-1 IN (black) and G140S-Q148H mutated HIV-1 IN (red). (bottom) Fluctuation of the Cα atoms of integrase of DTG-complexed wild type HIV-1 IN (black) and G140S-Q148H mutated HIV-1 IN (red); Right: Position of G140, Q148 (orange) and the 140s' loop (yellow) in active site of wild type HIV-1 IN bound to RAL.

To understand the reason of the higher fluctuation, caused by the G140S-Q148H substitution, of the 140s' loop when binding with RAL, and why this is not observed in G140S-Q148H HIV-1 IN-DTG complex, hydrogen bond network of the HIV-1 IN binding pocket is analyzed. A new hydrogen bond formation is observed between residue 140 and 148 in RAL-complexed double mutant, and it is illustrated in Figure 2.10A and B. Besides the formation of hydrogen bond, the disruption of hydrogen bond in the 140s' loop is also observed in the complex of G140S-Q148H double mutant with RAL. Figure 2.10C shows distributions of the hydrogen bond length formed between residue 145 and 148 in RAL- and DTG-complexes respectively. It is clear that the hydrogen bond between Pro145 and Gln148 is retained well in WT HIV-1 IN-RAL complex, while the length of the hydrogen bond increased from 2.02 Å to 2.67 Å in the double mutated integrase indicating the broken of hydrogen bond during the simulation. The snapped interaction between Pro145 and

His148 causes the denaturation of helix at the end of the 140s' loop. And along with the newly formed hydrogen bond between Ser140 and His148, the loop part gains more flexibility than in WT HIV-1 IN, which is consistent with the previous observation in the analysis of the atomic fluctuation. The flexibility change of the 140s' loop can be accounted for the loss of binding affinity between RAL and HIV-1 IN with the G140S-Q148H mutation. The observation of this rearrangement of hydrogen bond network in the 140s' loop of RAL-complexed double mutated integrase is in agreement with the results obtained by Chen's group.[92] Similar with complex of double mutated HIV-1IN-RAL, the formation of hydrogen bond has also been detected when the G140S-Q148H HIV-1 IN bound with DTG. (Figure 2.10C) However, there is no significant change of the Pro145-H148 hydrogen bond interaction being noticed in the DTG-complexed double mutant (WT HIV-1 IN-DTG: 2.25 Å, G140S-Q148H HIV-1 IN-DTG：2.40 Å) during simulation. It explains the result of B-factor discussed above in which the 140s' loop possesses similar flexibility in the DTG-complexed WT and double mutated integrase. These conformational differentiations elucidate why and how G140S-Q148H mutation leads to higher resistance towards RAL than that towards DTG. The reason of this occurrence of differentiation when G140S-Q148H HIV-1 IN binding with RAL and DTG is explored and discussed in the following section.

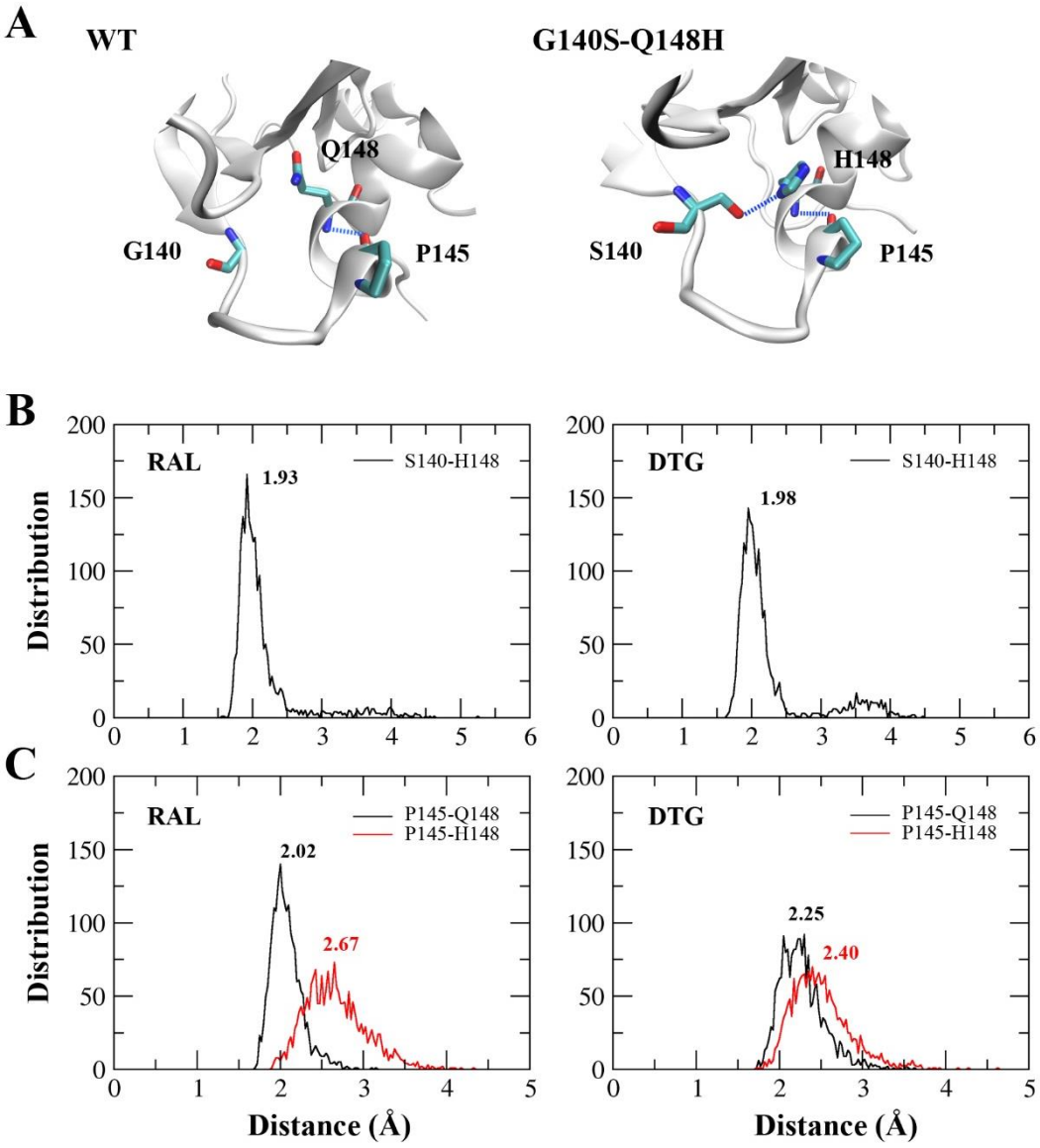


Figure 2.10. Changes in hydrogen bonds network of wild type and G140S-Q148H mutated HIV-1 IN binding to INSTIs. (A) Hydrogen bond between P145 and Q148 in wild type (left); hydrogen bond between P145 and Q148, and newly formed hydrogen bond between S140 and H148 in G140S-Q148H mutated HIV-1 IN. (B) Distribution of length of newly formed S140-H148 hydrogen bond while binding to RAL and DTG respectively. (C) Distribution of length of hydrogen bond between residues 145 and 148 in WT (black) and G140S-Q148H (red) double mutant while binding to RAL and DTG respectively.

2.3.4 Changes in the binding site of G140S-Q148H HIV-1 IN upon binding with RAL and DTG

To explain why the 140s' loop in the double mutated integrase performs differently when binding with different INSTIs, changes during simulation are explored for the binding pocket of RAL- and DTG-complexed G140S-Q148H HIV-1 IN. Figure 2.11B shows distributions of distance between Ser140 and Tyr143, and between Tyr143 and His148. It illustrates that in RAL bounded integrase these distances are longer than in DTG complex. As Tyr143 is located at the middle of the 140s' loop, these greater distance values indicate the loop range is enlarged by combining with RAL. Comparing with DTG, RAL possesses a long and flexible linker between pyrimidine group, which is the central part interacting with the catalytic Mg^{2+} ions directly, and oxadiazol group; at the same time, the oxadiazol group forms a face-to-face π–π stacking with Tyr143, which makes it possible that RAL provides a pressure pushing the loop away. And along with the Ser140-His148 hydrogen bond formation pulling the two end of the 140s' loop together, the loop is stretched causing the broken of hydrogen bond between Pro145 and His148 in RAL-complexed double mutated integrase.(Figure 2.11) Unlike RAL, in DTG, pyrido-pyrazino-oxazine group instead of the oxadiazol-linker-pyrimidine conformation, forms weaker interaction with Tyr143, as a result of the retaining of the Pro145-His148 hydrogen bond when double mutated integrase combining with DTG, which is conducive to combat the resistance. As such, a moiety possessing short and stable linkage with the central part of INSTIs and forming weaker interaction with Tyr143 may become an achievable strategy for designing the next generation of drug molecule to against drug resistance induced by G140S-Q148H double mutation.

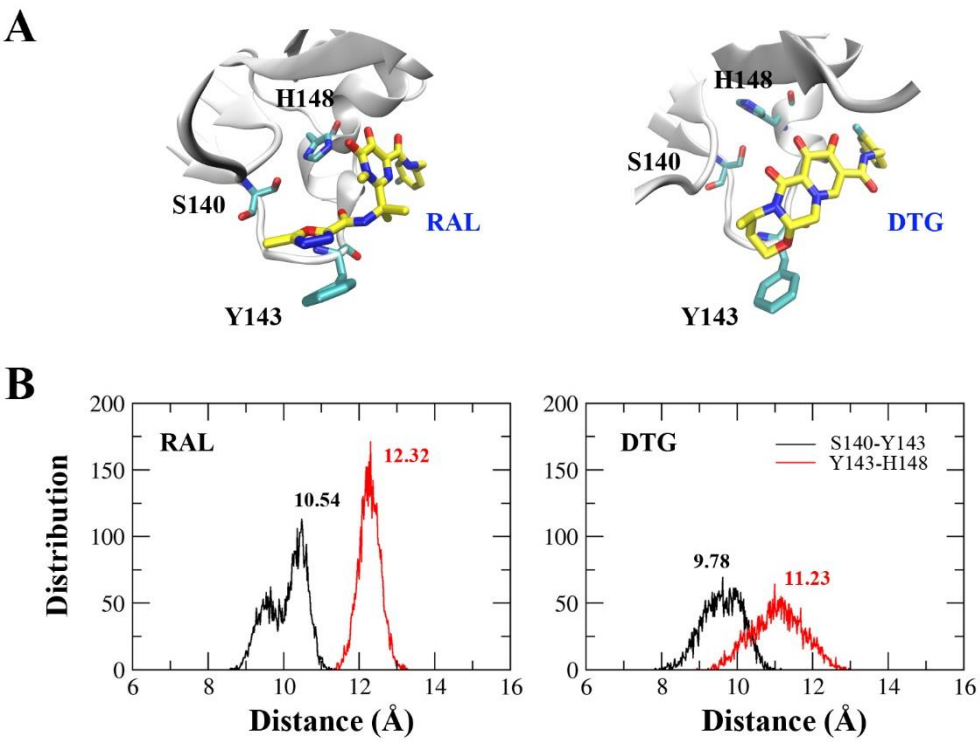


Figure 2.11. Differences in the active site of G140S-Q148H HIV-1 IN mutant binding to RAL and DTG respectively. (A) Active site of G140S-Q148H mutated HIV-1 IN in binding with RAL and DTG respectively. (B) Distribution of distances between Ser140 and Tyr143 (black), and between Tyr143 and His148 (red).

2.4 Conclusion

In this project, the basis of G140S-Q148H double mutation caused drug resistance to RAL and DTG is explored using MD simulations. By using deprotonated INSTIs, the catalytic metal ions, which play an important role in inhibition process, are kept at their own position ensuring them functioning. Through analyzing the obtained trajectories, the experimental observations are reproduced which show that activities of RAL and DTG decrease greatly and slightly respectively, in the presence of the double mutation in HIV-1 IN. A greater increase in the fluctuation of the 140s'

loop is observed in RAL-complexed HIV-1 IN double mutant suggesting that the stability of binding pocket is decreased. It is caused by the break of hydrogen bond between residues 145 and 148, which is induced by the new hydrogen bond formation in between S140 and H148 and together with the factor of RAL's structure forming stretch force on the 145-148 hydrogen bond. The instability of the 140s' contributes to the decreased binding affinity of RAL to G140S-Q148H double mutant. In contrary, DTG provides less pushing force on the loop leading to a slighter change on fluctuation of the 140s' loop, which subsequently generates the resistance combating effect of DTG observed. Through this project, helpful insight on how G140S-Q148H mutation affect the efficiency of RAL in integrase and how DTG is against this resistance are gained; and an idea for the designing a more effective and potent INSTI is displayed.

Chapter 3

Understand the mechanism of I50V caused drug resistance of HIV-1 protease via polarized force field applied molecular dynamics simulations

As mentioned in Chapter 2, in addition to integrase, HIV-1 protease is also an important drug target in antiretroviral (ARV) therapy. Several drugs have been developed and approved by FDA targeting protease as its inhibitors. However, similar as INSTIs, HIV-1 protease inhibitors (PIs) are also suffering from mutation caused drug resistance. In this study, polarized force field is employed to introduce polarization effect in MD simulations to study the mechanism of I50V mutation induced drug resistance of HIV-1 protease (HIV-1 PR). Global calculations of polarized protein specific charge (PPC) for the complexes of HIV-1 PR-PIs are conducted using the method introduced in Chapter 1. Good agreements have been reached with experimental data; and a rational explanation about how the mutation causes the affinity decreasing has been given. In the meanwhile, conventional force field is also applied in simulation, nevertheless it provides incompatible results with the data gained in experiments; therefore, the analysis are conducted based on the trajectories obtained using polarized force field.

3.1 Introduction

HIV virus was first discovered in 1983.[115] Until today there is still no vaccine or permanent cure for its infection. Fortunately there are drugs approved by the U.S. Food and Drug Administration (FDA) effective in prolonging the lives of HIV infected patients, operating through antiretroviral therapy. This therapy, involving the use of a combination of drugs, is nowadays the most widely used therapy in clinic. It includes reverse transcriptase inhibitors, HIV protease inhibitors and/or an IN inhibitor, also known as HAART (highly active antiretroviral therapy). Among those drugs, HIV protease inhibitor is one of the most essential elements, since protease is playing an important role in the life cycle of HIV and inhibiting its function to restraint the growth of the virus.

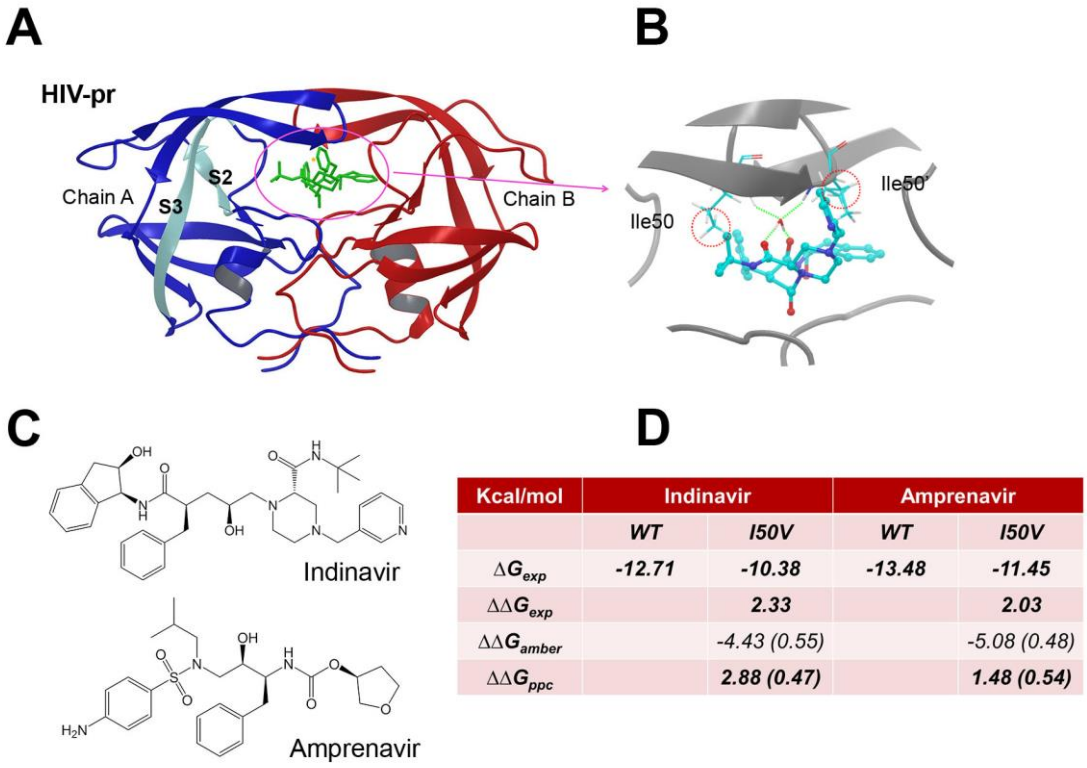
The HIV protease is a symmetric homodimer, and each chain contains 99 amino acid residues with an Asp at position 25. It performs for virus maturation by cleaving the Gag and Gag-Pol polyprotein precursor in the late step of replication process. At first, the Pol polyproteins is cut off from the Gag-Pol polyproteins; later on it is categorized into different components, protease, RNase H (p15), reverse transcriptase (p51), and IN. The polyprotein binding site is found not completely exposed. Instead it is covered by two ‘gate-like’ flaps. When the substrates try to enter into the binding site, the flexible flaps is in an open conformation.[116] As such, blocking the binding site of the protease is able to inhibit the HIV-1 protease enzyme activity.[117]

The significant role of HIV protease in the replication process makes it an important target for

anti-HIV drug design. And currently eight protease inhibitors have been approved by FDA for usage in antiretroviral therapy. However, the numerous HIV-1 PR mutants observed in clinic treatment influence the effectiveness of PIs, and then lead to the failure of the protease inhibitor therapy.[118] These drug resistance mutations arise within or distal from the binding pocket of the protease. Mutations arising inside the binding site, also called the primary mutations, straightly influence their inhibitory activities, being one of the major reasons of resistance in protease inhibitor therapy.[119-121] Among these primary mutations, I50V is recognized as one of the most important key mutations of the HIV-1 PR conferring resistance to several PIs in clinical therapy. In this mutation, the Ile is substituted to Val at residue 50 of both monomer of protease, which is located on the two flaps inside the active site (Figure 3.1B).[119-120]

Since it is frequently observed in clinical studies, I50V mutation has been investigated for its effect on protease binding with several PIs by employing standard force field (e.g. CHARMM and AMBER) in MD simulations.[122-124] Nevertheless, it may be necessary to include polarization effects in the force field to suggest a better description on the interaction of protease inhibitors with the HIV-1 PR binding site that contains polar residues, for example Asp at position 25, 29, and 30 in both chains. As a result of lacking polarization effects, conventional force fields is easy to fail in correctly describing the local electrostatic environment inside the active site as well as the deviation in the environment of the binding cavity when bind with distinct protease inhibitors. Hence, in this study, the free energies of binding protease inhibitors to HIV-PR and I50V mutant are investigated by employing polarization effect included MD simulations. It is realized by implementing polarized protein-specific charge (PPC) model proposed by Zhang and his

coworkers.[46] Unlike the AMBER force field, atomic charges of the whole protein in the PPC method are calculated iteratively based on quantum mechanical (QM) calculation with implementation of the molecular fractionation with conjugate caps (MFCC) method with existing of protein surface charges produced via solving Poisson-Boltzmann (PB) equation for solvent effect. The partial charges are fitted utilizing recently developed conformation-dependent RESP (delta-RESP) approach for obtaining better described charges for buried atoms. Thus, by applying the PPC model, the effect of intra-protein polarization and polarization between protein and ligand could be included in the atomic charges. Consequently, amino acids having the identical type will possess distinct atomic charges based on the particular conformation and electrostatic environment. PPC has been applied in a number of MD studies and provided promising results.[47, 50, 52, 125-127] Thus we believe that usage of PPC will give a correct binding mode of PIs in HIV-1 PR and suggest some insights on the basis of the binding affinity decrease along with I50V mutation of HIV-1 PR. In this study, two protease inhibitors are involved, indinavir (IDV) and amprenavir (APV) (Figure 3.1C), to explain the mechanism leading to the decrease of binding affinity between protease inhibitors and I50V mutated HIV-1 protease.



Kcal/mol	Indinavir		Amprenavir	
	WT	*I50V*	*WT*	*I50V*
ΔG_{exp}	*-12.71*	*-10.38*	*-13.48*	*-11.45*
$\Delta\Delta G_{exp}$		*2.33*		*2.03*
$\Delta\Delta G_{amber}$		*-4.43 (0.55)*		*-5.08 (0.48)*
$\Delta\Delta G_{ppc}$		*2.88 (0.47)*		*1.48 (0.54)*

Figure 3.1. (A) Schematic description of WT HIV-1 PR bound to IDV (green) with water (orange) molecules binding in the active site (PDB id: 1SDT[123]). (B) Schematic description of isoleucine at position 50/50' with the methyl groups that are eliminated in I50V mutated protease labeled with the red circle. (C) 2D structure of IDV and APV. (D) Summarization of the changes in the free energy of binding PIs to WT and I50V mutated HIV-1 protease in both PPC and Amber03 charge applied simulations.

3.2 Methodology

3.2.1 Delta RESP applied polarized protein-specific charge (PPC)

In earlier researches, PPC is derived using traditional Restrained Electrostatic Potential (RESP)

approach in continuum solvent, applying MFCC with Poisson-Boltzmann (MFCC-PB) approach.[57-59, 129] Nevertheless, for the buried atoms which are located at the central of the molecule, it is unable to determine their charges properly by employing RESP, since the grid points for RESP calculation is far away from these buried atoms and the atoms on the surface make major contributions to the RESP-fitted charges.[130-131] To conquer this difficulty, Mei and his group proposed a novel approach to calculate the conformation-dependent RESP charge (ΔRESP).[132] Through this method, atomic charge is considered as two components. First, the dominant part is using a predefined fixed charge, for example the charge from standard force fields. And for the residual charge part, it is varied according to the local electrostatic environment and conformation change.[132] Because the variable charge is much smaller than the dominant charge part, deficiency will be reduced if RESP is only used to fit the charge variation instead of the overall charge of the particular atom.[132] This project uses ΔRESP charge fitting approach to generate the PPC of the studied complex rather than the traditional RESP charge fitting approach. The rest part of the detailed PPC calculation could be found in Ref.46. All quantum mechanical calculations are carried out at B3LYP/6-31G* level in Gaussian09[103] software.

3.2.2 Initial structures

The crystal structures of HIV-1 PR WT-IDV, I50V-IDV, WT-APV and I50V-APV complexes are acquired from the Protein Data Bank (PDB ID: 1SDT[128], 2AVS[128], 3NU3[133] and 3NU5[133]). In the previous researches, the protonation state of Asp25 in both chains of protease is recognized to

be important in the binding of PIs.[134] In which Asp25 of chain A is deprotonated and Asp25' of chain B is protonated. Hence, during the preparation of initial structures, a proton is attached to atom OD2 of Asp25'. In the meantime, all water molecules appeared in the PDB file are preserved. Atomic charges of PIs are fitted using the AM1-bcc method after minimization at the AM1 semi-empirical level. The atomic partial charges are then assigned to the PIs using the antechamber module in the AMBER10 package[25]. The AMBER03 force field[29] and GAFF, the Generalized Amber force field[104], are applied to the receptor and ligand respectively with using the Leap module. With using Leap module, hydrogen atoms are added to HIV-1 PR-APV and HIV-1 PR-IDV complexes, which are added into a periodic box of TIP3P water molecules with the minimum distance at 10 Å between the surface of the box and protein.[105] Besides, counter ions are added to neutralize the whole system. The Particle Mesh Ewald (PME) approach[107] is employed to deal with long-range electrostatic interactions and a cutoff of 12 Å is applied for every non-bonded van der Waals interaction appeared in the systems. The complexes are optimized and followed by the PPC calculations.

3.2.3 Molecular dynamics simulations

The complexes are optimized with two procedures. At first, the steepest descent approach is applied to minimize the solvent molecules for 1000 steps, followed by 3000 steps minimization with using conjugate gradient approach, with a 500 kcal/mol/$Å^2$ restraint placed on the complexes. Secondly, the whole system is minimized for 10000 steps with steepest descent approach, then

with conjugate gradient till a convergence of 0.01 kcal/mol/Å for the RMS of the energy gradient is reached. Following, the heating phase is conducted for the systems and the temperature is raised from 0 to 300K in 100ps, then NPT simulation is performed for 20ns at a time step of 2 fs. To regulate the system temperature in MD simulations, Langevin[106] dynamics is employed with a collision frequency of 4.0 ps^{-1}. The SHAKE algorithm[108] is applied to restrain hydrogen atoms involved bonds. Both AMBER03 charge and PPC are implemented in MD simulations of this study. Comparing with simulations using AMBER03 force field, only the atomic charges are substituted by PPC where rest of the parameters is kept intact in the simulations using PPC.

3.2.4 Molecular Mechanics-Poisson Boltzmann Surface Area (MM/PBSA)

The following equation represents the binding free energy, ΔG_{bind}, of the PI complex.

$$G_{bind} = G_{complex} - G_{receptor} - G_{ligand} \tag{3.1}$$

in which $G_{complex}$, $G_{receptor}$ and G_{ligand} defines the free energy of the complex, receptor and ligand respectively. ΔG_{bind} is obtained from the following equation under MM/PBSA method:

$$\Delta G_{bind} = \Delta G_{PBSA} - T\Delta S_{solute}$$

$$= \Delta E_{MM} + \Delta G_{sol} - T\Delta S_{solute}$$

$$= \Delta E_{ele} + \Delta E_{vdW} + \Delta G_{pol} + \Delta G_{nonpol} - T\Delta S_{solute} \quad (3.2)$$

in which ΔG_{sol} is composed of the non-polar solvation (ΔG_{nonpol}) and polar solvation (ΔG_{pol}) energy of the system and $T\Delta S_{solute}$ is the entropy of solute. The system total molecular mechanical energy is represented by ΔE_{MM}, constituting ΔE_{ele}, the electrostatic interaction energies, and ΔE_{vdW}, van der Waals. ΔΔG, the relative biding free energy between I50V and WT mutant is acquired from Eqn. 3.3 below.

$$\Delta\Delta G = \Delta G_{I50V} - \Delta G_{WT} \quad (3.3)$$

in which ΔG_{I50V} is the binding free energy of the inhibitors to I50V mutated HIV-1 PR and ΔG_{WT} is that to WT.

In this study, molecular dynamics simulations are conducted for HIV-1 PI complexes applying both AMEAR03[29, 135] charges and PPC. 500 snapshots are taken from the last 5 ns at 10 ps intervals for each system, where the last 5 ns is the most stable part of the 20 ns MD simulations. The solvation energy ΔG_{sol} is obtained from the PBSA module of the AMBER10 package using the grid spacing at 0.25 Å. The nonpolar solvation energy is obtained according to equation $G_{nonpol} = \gamma SAS + \beta$ gives, where γ = 0.00542 kcal/(mol ·Å^2) and β= 0.92 kcal/mol.[126] Nmode[136] module in AMBER10 is applied for deriving the loss of entropy while protein-ligand binding. 50 snapshots are obtained from the last 5 ns with 100 ps interval for each system to calculate entropy. Every snapshot is minimized under a distance depending dielectric function until the energy gradient's root-mean-square falls below 10^{-5} kcal/mol/Å^2.

3.3 Results and discussion

3.3.1 Binding free energy

Free energies of binding protease to PIs are obtained based on the four simulations under MM-PBSA method. Table 3.1 lists the contributions of each energy component, which are contained in Eqn. 3.2, to the binding free energies of each protease-inhibitor complex under both AMBER03 charges and PPC. As a result of the drug resistance to protease inhibitors caused by the I50V mutation of HIV-1 protease, it is expected that the binding affinities of APV and IDV to the PR are stronger in WT than in the mutants. This is also tabulated in Figure 3.1D, by the $\Delta\Delta G_{exp}$ values. By comparison, PPC is able to provide a closer agreement with the relative binding free energy obtained from experiments. Henceforth, all analyses will be carried out according to the simulations based on PPC, to examine how the I50V mutation causes its resistance to protease inhibitors.

From Table 3.1, electrostatic interaction energy is increased in the presence of I50V mutation, which makes significant contributions towards the drop of binding affinity. This aligns with the conclusions made by Chen's[124] and Meher's[123] group that protease inhibitors resistance led by the I50V mutation is greatly credited to the growth of electrostatic energy. In this study, it also observes a growth in the van der Waals interaction energy between protease inhibitors and HIV-1 protease, which to some extent agrees with Kar et al.'s theoretical study who believes the growth

of vdW interaction energy between APV and I50V mutant makes major contributions in the binding affinity loss.[122] Nevertheless, in this study, it suggests the vdw interaction energy and electrostatic energy increasing together leads to the drop of binding affinity in I50V mutant-PIs complexes.

Other than calculating the overall binding free energy, the contribution of each residue to the binding affinity is also studied. (Figure 3.2) In the figure, it shows residues Ala28/28', Ile50/50' and Ile84/84' make major contributions in both HIV-1 protease-IDV and –APV complexes. (Figure 3.2a and b) And the relative binding free energy of each residue ($\Delta\Delta G_{per\ res}$) between the WT and mutant is obtained according to the equation below. (Figure 3.2c)

$$\Delta\Delta G_{per\ res} = \Delta G_{I50V.per\ res} - \Delta G_{WT.per\ res} \tag{3.4}$$

From the $\Delta\Delta G_{per\ res}$, residues Arg8', Asp29' and Val50' contributes primarily to the drop of binding affinity of IDV to mutated HIV-1 protease; at the meantime, residues Asp25, Gly48 and Val50/50' make major contributions to the drop of binding affinity of APV to the I50V mutated HIV-1 protease protein. It is resulted from the loss of a methyl group during the replacement from Ile to Val, making the hydrophobic interaction loss between AVP/IDV and the side chain of Res50/50'. These results agree with the results achieved in other studies.

Table 3.1. Free energy components for binding PIs to HIV-1 PR in both PPC and Amber03 charge applied simulations. (Number in parentheses are standard error.)

	AMBER(ff03)				**PPC**			
	IDV		**APV**		**IDV**		**APV**	
	WT	**I50V**	**WT**	**I50V**	**WT**	**I50V**	**WT**	**I50V**
ΔE_{ele}	-65.73	-66.89	-44.49	-52.93	-118.59	-114.82	-98.65	-90.09
	-0.01	-0.01	-0.28	-0.23	-0.4	-0.39	-0.29	-0.28
ΔE_{vdw}	-71.97	-70.32	-60.43	-57.05	-66.32	-65.78	-59.31	-58.19
	-0.18	-0.19	-0.15	-0.16	-0.21	-0.21	-0.19	-0.19
ΔG_{nopol}	-5.34	-5.25	-4.33	-4.30	-5.44	-5.40	-4.52	-4.48
	-0.01	-0.01	-0.01	-0.01	0	-0.01	0	0
ΔG_{pb}	92.43	90.78	73.01	73.06	128.46	125.96	109.47	102.81
	-0.27	-0.23	-0.18	-0.17	-0.25	-0.27	-0.21	-0.19
ΔG_{pol}	26.7	23.89	28.52	20.13	9.88	11.14	10.82	12.73
	-0.25	-0.23	-0.26	-0.18	-0.27	-0.27	-0.22	-0.22
ΔG_{total}	-53.58	-54.62	-38.89	-43.84	-64.9	-63.03	-55.72	-52.64
	-0.24	-0.2	-0.25	-0.18	-0.2	-0.2	-0.18	-0.22
$-T\Delta S$	29.51	25.88	24.99	24.87	24.28	25.3	22.67	21.08
	-0.26	-0.36	-0.24	-0.27	-0.26	-0.26	-0.3	-0.34
ΔG	-21.37	-25.80	-11.25	-16.34	-37.61	-34.74	-30.34	-28.86
	-0.36	-0.41	-0.34	-0.33	-0.33	-0.33	-0.35	-0.41
$\Delta\Delta G$		**-4.43**		**-5.08**		**2.88**		**1.48**
		-0.55		**-0.48**		**-0.47**		**-0.54**
ΔG_{exp}	-12.71	-10.38	-13.48	-11.45	-12.71	-10.38	-13.48	-11.45
$\Delta\Delta G_{exp}$		**2.33**		**2.03**		**2.33**		**2.03**

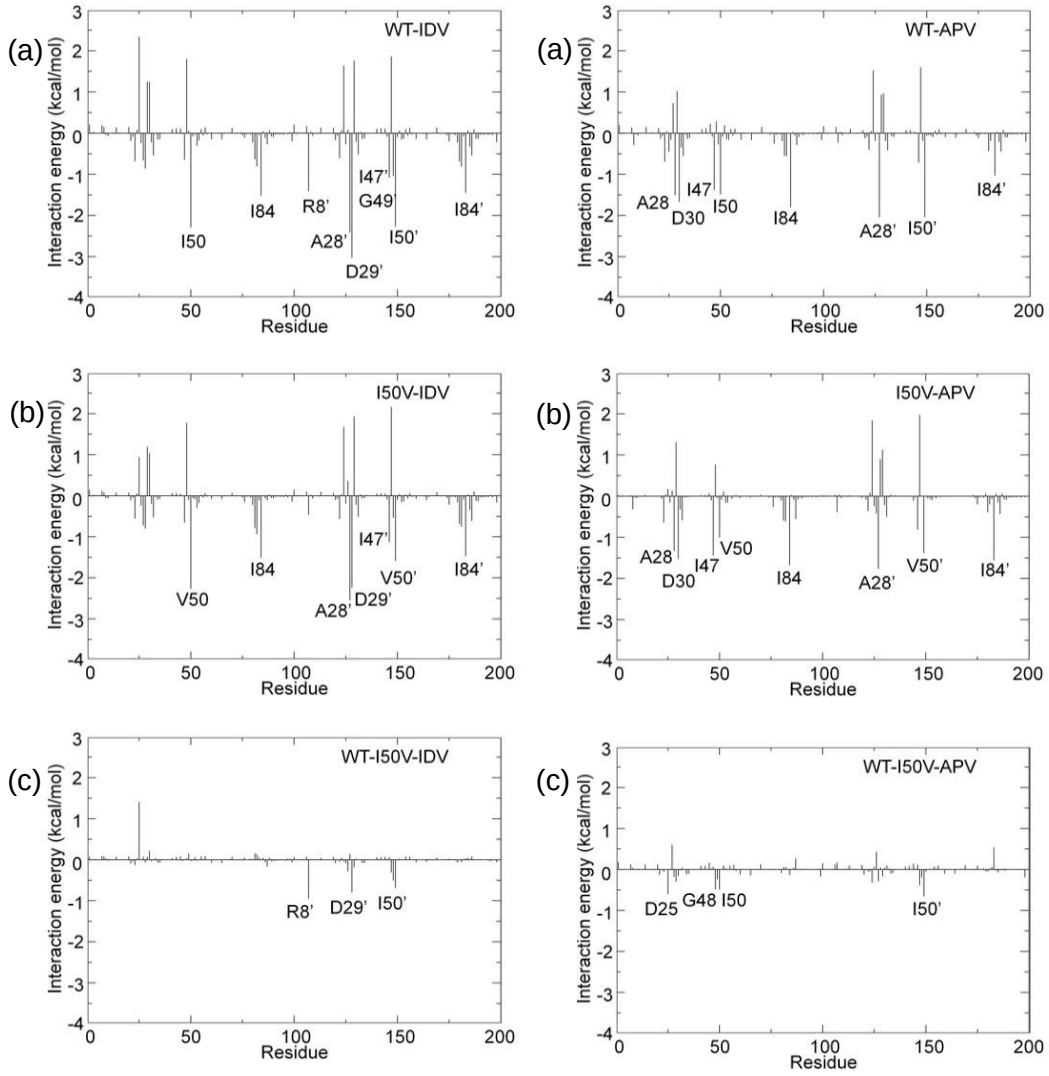


Figure 3.2. Per-residue decomposition of binding free energy of IDV and APV to HIV-1 PR variants. HIV-1 PR-IDV complexes are showed in the left column; and HIV-1 PR-APV complexes are showed in right column. (a) WT, (b) I50V, (c) $\Delta\Delta G_{per\ res}$($\Delta\Delta G_{per\ res} = \Delta G_{WT.per\ res} - \Delta G_{I50V.per\ res}$).

3.3.2 Changes at the active site of HIV-1 PR with I50V mutation in present

It is important to thoroughly examine the 3D conformation of the active site of the protein in order to study the mechanism regulating the changes in the binding affinity of protease inhibitors

mutated HIV-1 protease, as the I50V mutation forms inside the active site and the residue involved in the binding of inhibitors directly. Fewer waters are found in the binding pocket of the wild type than that of the mutant by observing the amount of water molecules existed in the active sites of the crystal structures of WT HIV-1 PR-IDV and I50V HIV-1 PR-IDV complexes. (Figure 3.3) This provides us a hint to investigate two factors that may possibly conduce to the resistance of HIV-1 protease to APV and IDV with I50V mutation, which are the active site volume change with I50V mutation and the flexibility variation of residues located at the edge of the binding pocket of the protease inhibitors with mutation. Both of the factors are able to cause the stated growth in number of water molecules inside the active site.

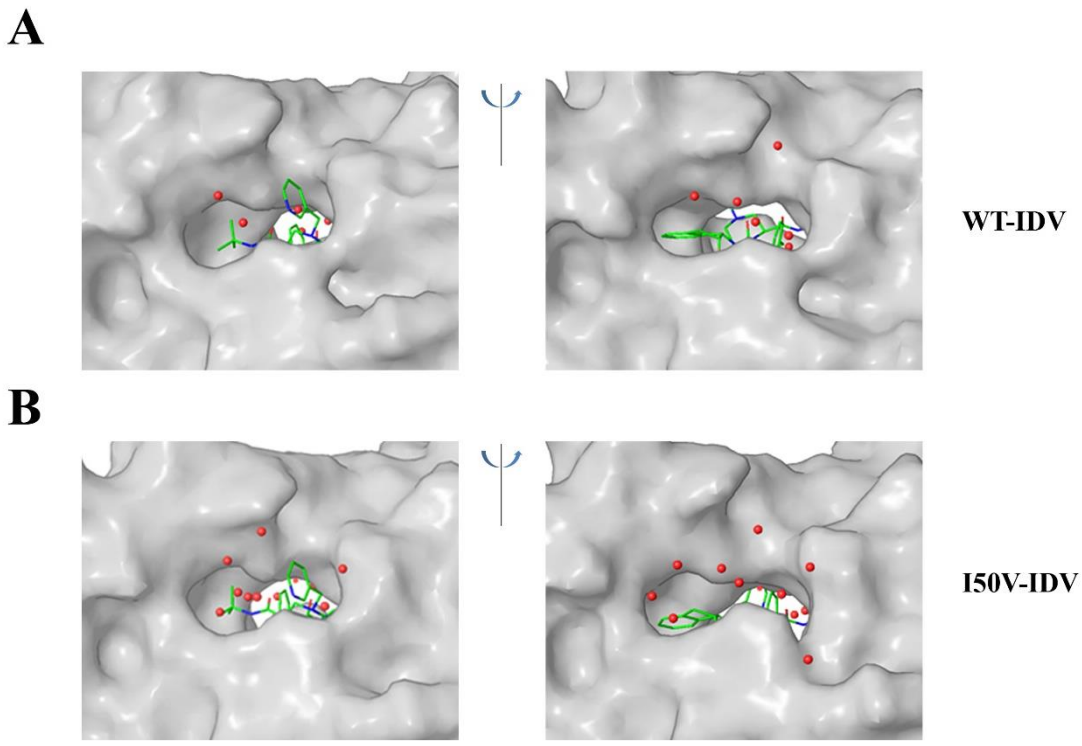


Figure 3.3. (A) Surface description of binding site of wild type HIV-1 protease (grey) interacted with indinavir (green) with the crystallized water molecules (red) located within 5 Å of indinavir showed. (B) Surface description of binding site of I50V mutated HIV-1 protease (grey) interacted with indinavir (green) with the crystallized water molecules (red) located within 5 Å of indinavir showed. Both the front (left) and the back (right) view of the protein are displayed.

POVME[137] is introduced to compute the volume of the binding site of the HIV-1 protease structures obtained in MD simulations in order to explore the fluctuations in the active site volume. Figure 3.4A plots the frequency distribution of the computed volume of the binding pocket of APV- and IDV-complexed HIV-protease variants, from where we can see a bigger active site volume is possessed by the mutated proteins. In particular, volume of a binding pocket in WT-IDV is 380 $Å^3$, while in I50V-IDV is 390 $Å^3$ which is 10 $Å^3$ more, whereas in WT-APV is 365 $Å^3$ and in I50V-APV is 375 $Å^3$, showing a same 10 $Å^3$ difference. In WT-APV, I50V-APV, WT-IDV and I50V-IDV complexes, the volumes of the active sites are 367, 379, 378 and 395 $Å^3$ on average. This observation verified the growth in number of water molecules contained in the binding site of I50V mutated protease.

After examining the volume change of the binding pocket, the atomic fluctuation of I50V mutated and WT HIV-1 protease at residue level are explored via computing the B-factor, or known as the isotropic temperature factor. (Figure 3.4B) The B-factor of each residue of the protease is not observed to be largely different between WT and mutated HIV-1 protease in APV- and IDV-complexed protease, except at residue Pro81' in both APV- and IDV-complexed PR and in the regions of flaps of HIV-1 PR-APV complex. HIV-1 protease as a homodimer, its Chain A is symmetric with Chain B which should share similar characteristics. Nonetheless, Pro81 of Chain A and Pro81' of Chain B behave differently, where the B-factor of Pro81 stay unchanged on I50V mutation while the B-factor of Pro81' rises upon inhibitor bound to mutated HIV-1 protease. Particularly, the B-factor of Pro81' in I50V-IDV is 24.35 $Å^2$, while in wild type is 9.92 $Å^2$; meanwhile in APV-complexed WT is 12.43 $Å^2$ and in I50V-APV is 17.61 $Å^2$. (Figure 3.4B) This

phenomenon is also stated by Shen et al.[133] who detected the higher flexibility of Pro81' in I50V HIV-1 protease. As the structures of inhibitors in binding with PR variants are asymmetric, contrasting the symmetric property of the homodimer, the fluctuation of the ligands is investigated to explore the reason of the distinction between the fluctuation of Pro81 of Chain A and Pro81' of Chain B in the I50V mutated HIV-1 protease. The fluctuation of each atom of indinavir and amprenavir in binding with HIV-1 PR variants are explored and described in the following section.

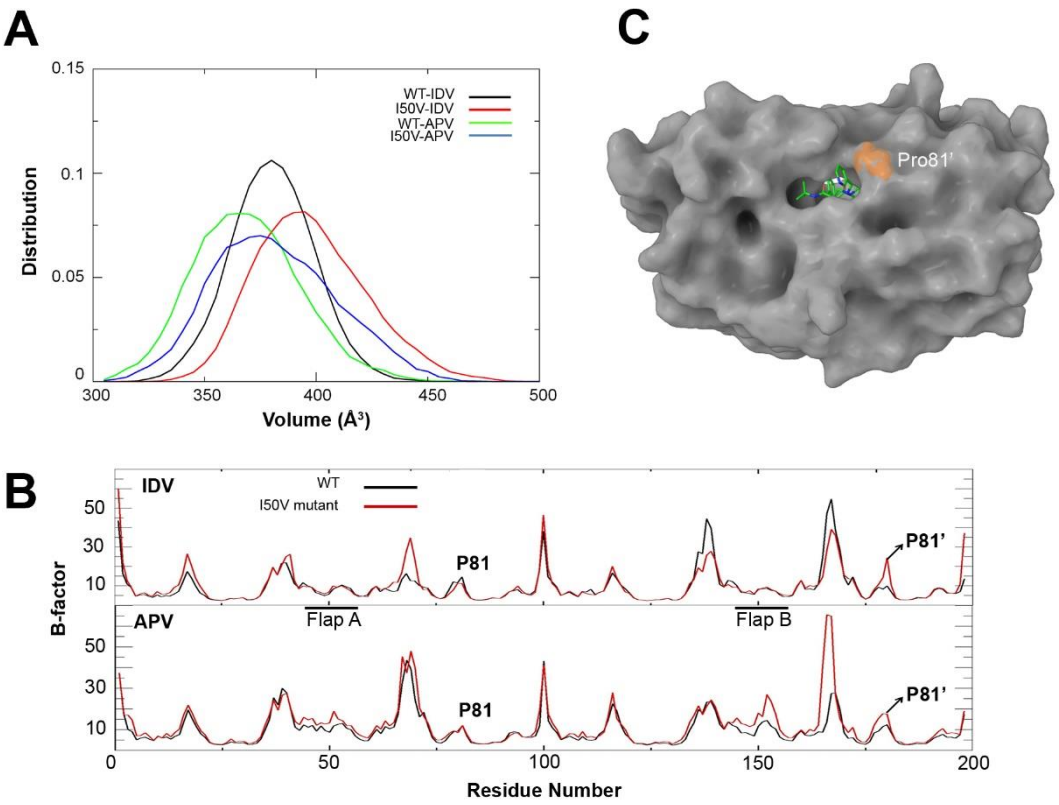


Figure 3.4. (A) Probability of the volume of the active site of indinavir and amprenavir interacted HIV-1 protease structures. (B) (top) Fluctuation of the Cα atom of HIV-1 PR in PPC applied simulations of wild type-IDV complex (black) and I50V mutant-IDV complex (red). (bottom) Fluctuation of the Cα atom of HIV-1 PR in PPC applied MD simulations of wild type-APV complex (black) and I50V mutant-APV complex (red). (C) Surface description of HIV-1 protease bound to indinavir with the position of residue Pro81' highlighted (orange).

3.3.3 Changes of protease inhibitors' stability in the binding site of HIV-1 PR with I50V mutation in present

The mutation causes larger binding pocket and the solvent accessibility rises which may cause the interactions break between the protease and PIs in the active site. These events lead to high flexibility of the inhibitor in binding pocket. The RMSD and inhibitors B-factor for APV- and IDV-complexed HIV-1 protease variants are computed to examine the I50V mutation's effect on the on the performance of protease inhibitors in the binding site. (Figure 3.5) The RMSD of IDV in binding with the I50V mutant rises from 0.90 Å to 1.42 Å between 5 to 10 ns during the simulation. (Figure 3.5A) Moreover, the RMSD of ligand in the I50V mutated HIV-1 protease-IDV complex varies in the range of 0.50 to 1.50 Å. (Figure 3.5A) Likewise for APV, a smaller RMSD range is detected in WT than in the I50V mutant. (Figure 3.5B) This concludes that the I50V mutation leads to a large fluctuation of the protease inhibitors in the HIV-1 protease active site.

Figure 3.5C and 3.5D plot the B-factors for protease inhibitors in binding with HIV-1 PR. Similar with the RMSD results, a larger B-factor of PIs (APV: Atoms 18-25, 39-49, 58-69, IDV: Atoms 24-34, 55-64) is observed, suggesting the higher flexibility of inhibitor bound to I50V mutated PR. Additionally, the degree of the flexibility differs in each part of the protease inhibitors. (Figure 3.5C and 5D) The phenyl ring in both structures of APV and IDV has direct interaction with Pro81 of the protease; and according to the fluctuation of each residue in protease (Figure 3.4B), Pro81 shows little changes in fluctuation in I50V mutant. Phenyl ring may protect Pro81 from the

nearby solvent molecule due to its hydrophobic feature, as a result to keep the fluctuation of Pro81 in the I50V mutant similar to in WT PR. In the contrary, the propyl group of APV (Figure 3.5D) and the pyridine ring of IDV (Figure 3.5C) changes more suggestively when in binding with HIV-1 PR I50V mutant than to WT. Those functional groups interact with protease via Pro81' and Pro81' shows larger flexibility within the I50V mutant of protease. (Figure 3.4B) By comparing with the phenyl ring, the pyridine group of IDV and the propyl group possess lower hydrophobicity. Hence these might be weaker than the phenyl group in shielding the side chains of residue Pro81' from the growth of solvent effect caused by the higher accessibility of solvent of the binding site suggested in Figure 3.4A by the growing volume of the binding pocket. As such, to overcome the I50V mutation's drug resistance, it may be a practical strategy to use a higher hydrophobic moiety replacing the pyridine group of IDV and the propyl group of APV for the designing of the new protease inhibitors.

Except for the interaction between the phenyl group and Pro81, the loss of interactions between residue 50 and the phenyl ring of both APV and IDV is also observed in I50V mutant, due to one methyl group removed during the substitution from Ile to Val. Same observation is found for the atoms 58-69 which is the THF group that initially interacting with residue Ile50. These phenomena are agreed by Shen et al.[133] in their study. According to the B-factor and RMSD of PIs in binding with HIV-1 PR variants, both IDV and APV have a higher flexibility in the I50V mutant-PIs complexes. This higher flexibility is caused by enlarge in the binding pocket volume allowing the PIs having a higher degree of freedom when binding to the I50V mutant. At the meantime, the interaction loss between protease and its inhibitors at the active site also contributes

to the higher degree of freedom of PIs. These events should be mainly responsible for the binding affinity decrease when PIs binds to the I50V mutated protease.

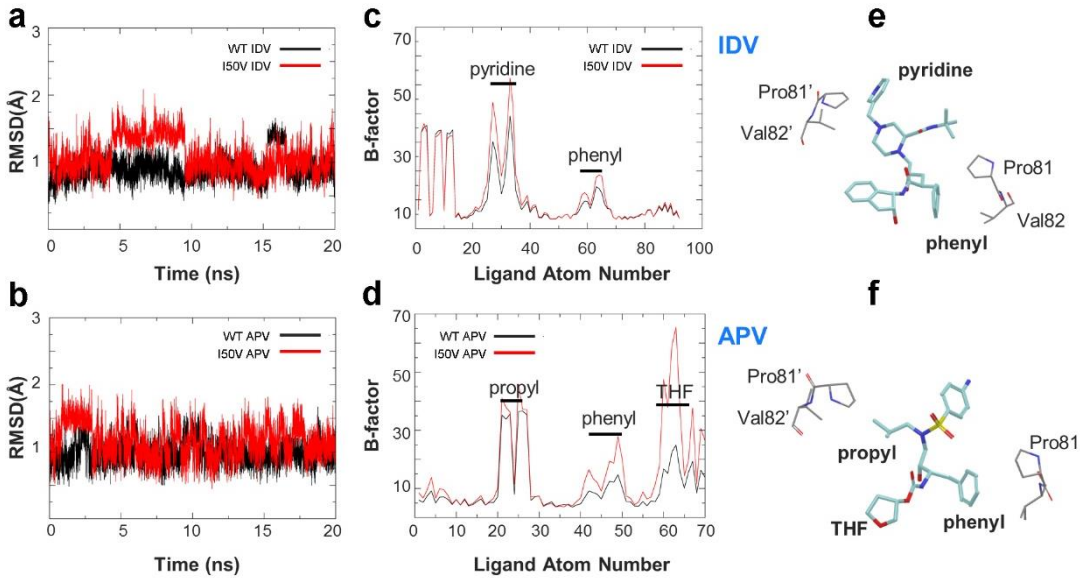


Figure 3.5. (A) Fluctuation in RMSD with simulation of IDV-complexed wild type (black) and I50V mutated HIV-1 protease (red). (B) Fluctuation in RMSD with simulation of APV-complexed wild type (black) and I50V mutated HIV-1 protease (red). (C) Fluctuation of atoms of indinavir bound to wild type (black) and I50V mutated HIV-1 protease (red). (D) Fluctuation of atoms of amprenavir bound to wild type (black) and I50V mutated HIV-1 protease (red). (E) Schematic representation of an indinavir molecule binding with adjacent residues of HIV-1 protease. (F) Schematic representation of an amprenavir molecule binding with adjacent residues of HIV-1 protease. Key functional groups are labeled accordingly.

3.3.4 Correlation matrices of WT and I50V mutated HIV-1 PR

The interactions between two flaps will be analyzed in this study, which are acting as the gate of the binding site, when the inhibitor binds to the protease the flaps can trap it inside. As shown above (Figure 3.4B), the fluctuations of the flaps residues are comparable in WT and I50V mutant while bound with IDV; on the other hand, it is found that the fluctuation is a little higher in both

flaps of APV complexed mutant. The correlation matrices of I50V mutant and WT bound to APV and IDV disclose the distinction in interaction of Flap A and B. As shown in Figure 3.6, less positive correlation is spotted for I50V mutant of HIV-1 protease binding with APV and IDV than that for WT. It suggests the interaction between two flaps is weakened in I50V mutant than in WT protease while binding with APV and IDV respectively. Moreover, in I50V mutant-APV complex, the interaction between the flaps is weaker than that I50V mutated protease bound to IDV. This observation effectively mirrors the findings obtained in the analysis of fluctuation of protease in Figure 3.4B which both flaps of APV-complexed I50V mutant possess higher flexibility than that of WT. The weakened interaction between Flap A and B of the I50V mutant may cause the decrease of the HIV-1 PR's ability on stabilizing the inhibitor in the binding pocket by closing the entry of binding site. The higher freedom of movement of APV (Figure 3.5D) associated with the growth in flexibility of Flap A and B (Figure 3.4B) and higher solvent accessibility of the active site weaken the interactions between PIs and protease (specifically the corresponding regions at Residue 50/50' in Figure 3.1B) as a result to induce the drug resistance to PIs upon I50V mutation.

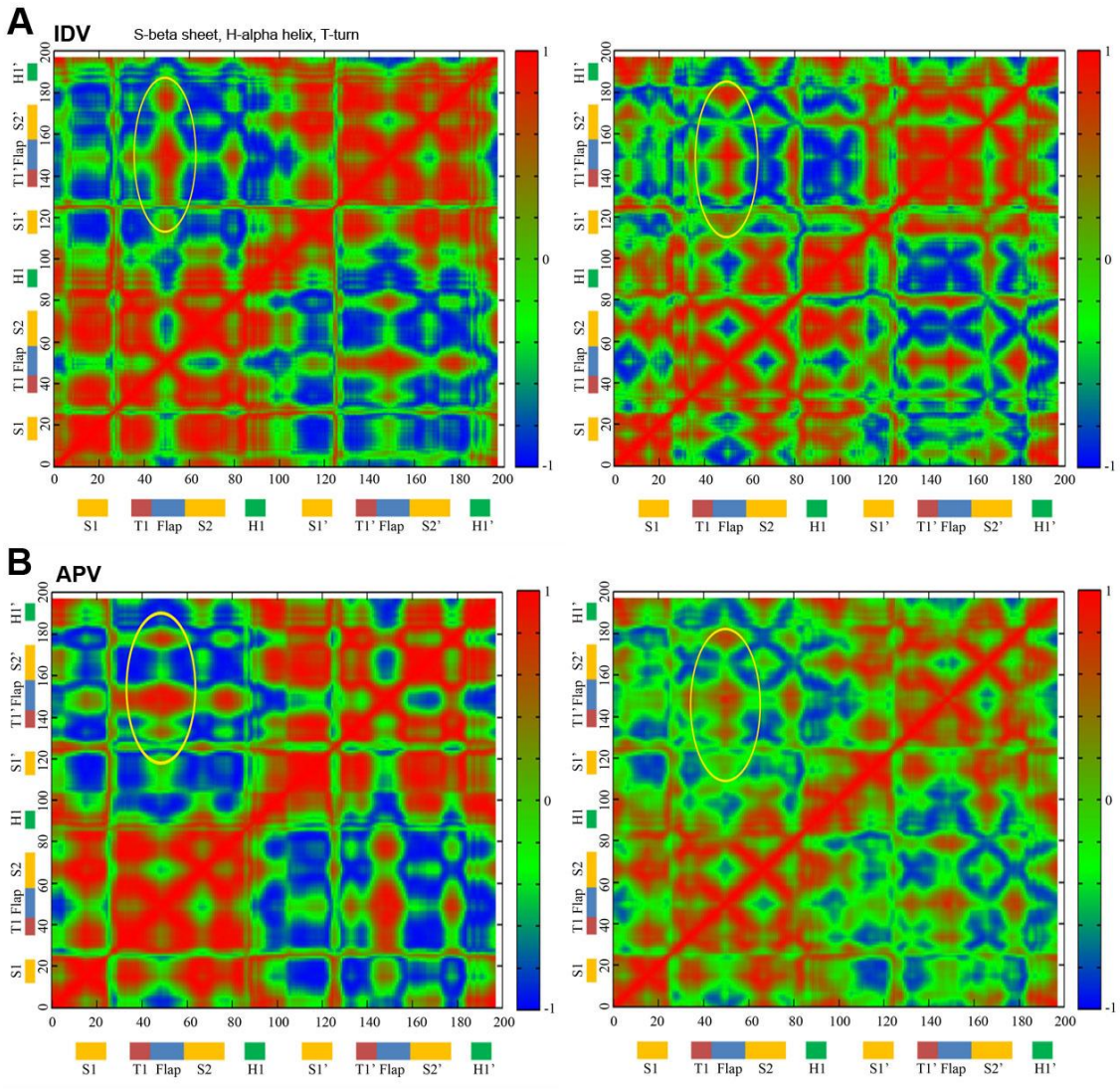


Figure 3.6. (A) Cross correlation matrices of IDV-complexed WT HIV-1 protease (left) and I50V mutated HIV-1 protease (right). (B) Cross correlation matrices of APV-complexed WT HIV-1 protease (left) and I50V mutated HIV-1 protease (right). The figure shows correlation between amino acids with indices displayed on x and y axes, blue color representing negative correlation and red color representing positive, with the range displayed in the color bars. Secondary-structure elements are labeled in sequence order. Area encircled in yellow circle illustrations the correlation between two flaps.

3.4 Conclusion

In this study, experimental observations are attempted to be reproduced, in which the binding

affinity of PIs are decreased in binding with I50V mutated HIV-1 protease. By introducing PPC into MD simulations, the results agree well with the experiment obtained data. This emphasizes the significance of including polarization effect into MD simulations. A study on I50V mutation induced binding affinity decrease when protease binds to IDV and APV is also established by Leonis et al.[138] via using the conventional AMBER force field. However, in this study, the relative free energy of binding PIs to I50V mutated and WT protease is calculated more precise than that given by Leonis et al.[138], suggesting the importance of the PPC scheme with applying Delta-RESP on giving accurate relative binding affinities. In addition, it enables us in reproducing the more weakening of the interactions of I50V mutant with IDV than with APV.

The analysis of PPC applied simulations rationally explains the question on how I50V mutation leads to a drop in binding affinity. By studying the trajectories obtained from MD simulations, an observation of the mutation caused volume growth of the binding site in I50V mutant suggests a raise in the accessibility of solvent of the active site. These changes make the interaction between protease inhibitors and protease become weaker in the presence of the I50V mutation; meanwhile, these contribute to the increase of the degree of freedom of inhibitors in the binding pocket which then cause the binding affinity decrease in PIs-complexed I50V mutant. Furthermore, the interaction between two flaps of protease is weakened by this particular mutation, as well as the correlation between the secondary structures. In this project, explanations on how mutation influence PIs' efficacy in inhibiting protease function are gained and at meantime an idea on designing more potent PIs have been given.

Chapter 4

Molecular dynamics simulations study on the basis of Avidin-Biotin binding using polarized force field with charge updating schemes

In the previous chapter, PPC is calculated globally for the protein-ligand complex based on single conformation before conducting MD simulations. Since PPC is based on the environment of each fragment, it may lead to bias while using the PPC obtained based on a single structure during the whole simulation. Periodical charge updating schemes, environmentally corresponding local PPC (ECLPPC) scheme and environmentally corresponding hydrogen bond dependent PPC scheme (ECHBPPC), have been proposed to improve the accuracy regarding protein-ligand interactions analysis in MD simulations. In this chapter, these charge updating schemes are applied to study the basis of avidin-biotin interactions, and the theoretical results agree well with data obtained in experiment.

4.1 Introduction

Previously, MD simulations has shown its power and advantages in studying biological molecules, and it has gained many achievements. However, with using the standard force field, such as AMBER03 force field,[29] the electrostatic interaction is hardly to be described properly, since the

effect of electrostatic polarization is failed to be taken into account. And the polarization effect has been proved playing an important role in providing more accurate and reliable results in simulating biomolecules in many studies.[126, 139-140] To overcome this deficiency of the standard force field, efforts have been made in improving force field to obtain a better description on electrostatic interaction over the past years. For this purpose, polarized protein-specific charge (PPC) is developed by Zhang's group[46]. In PPC, all the atomic charges in a specific protein are calculated based on quantum mechanical (QM) calculation which is realized by utilizing molecular fragmentation with conjugated caps (MFCC) method.[57-58] During calculation, the system is separated into individual fragment (amino acid residue) using MFCC method.[59] And the atomic charges of each fragment are derived under their electrostatic environment to incorporate polarization effect in. In which, the environment changes, then the derived charges vary accordingly. As such, applying atomic charges derived under its own environment should provide better description in electrostatic interaction than using the atomic charge derived without considering its environment. And currently, this charge model has been implemented in many studies effectively, from protein-ligand interaction to protein folding, as expected.[47, 50, 52, 125-127]

Although it improves the description on electrostatic interaction in MD simulations, there still is deficiency in this charge model. According to the previous studies and tests using PPC, it is noticed that protein dynamics is highly sensitive to the atomic charges used in the force field. In the meantime, it has been known that PPC is structural dependent and sensitive to the conformational changes. Therefore, it may be problematic using a set of PPC obtained based on a single conformation of a specific protein disregarding the conformational variation during the entire MD

simulations. To reduce the negative effects caused by this problem, the PPC updating scheme has been developed and examined in recent years, including studies of hydrogen bond stability in the structure of staphylococcal nuclease,[52] and the binding site of DFsc, a de novo designed dizinc metalloprotein.[141] In these studies, PPC update scheme applied simulations have accomplished promising results. Hence, it has been considered that it is necessary to employ PPC updating scheme in study of protein-ligand interactions, since displacement of ligand and residues, which comprise the binding pocket, from their initial position usually happens in MD simulations of protein-ligand complex, due to the instability of non-bonded interactions between them. By using PPC updating scheme in MD simulations, the conformational change caused by displacement would be able to reflect to the variation of atomic charges.

In this work, two PPC updating schemes are proposed, environmentally corresponding local polarized protein-specific charge (ECLPPC) scheme and environmentally corresponding hydrogen bond dependent PPC (ECHBPPC) scheme. Inspired by QM/MM method, in both schemes, the charge updating is only focused on the central domain of protein-ligand complex, which is the ligand and selected residues involved in comprising the binding pocket instead of calculating PPC for the complex globally. And the rest residues, which are not involved in interacting with ligand directly, would not affect the ligand binding significantly. In this way, the simulation cost can also be reduced effectively compare to calculating global PPC for the system. Particularly, under ECLPPC scheme, ligand and all residues within a certain radius surrounding the ligand are selected dynamically to periodically conduct LPPC calculation to update their atomic charges during simulation. And in this study, the radius is set to 4 Å, meaning that residues, which possessing at

least one atom has a distance less than 4 Å with any atom of the ligand, will be performed LPPC calculation at each time interval. (Figure 4.1A) By applying this residue selection criterion, the involved binding pocket residues can be selected according to the present conformation of ligand dynamically. In another scheme, ECHBPPC, the selected residues are narrowed down to those form hydrogen bonds with the ligand. (Figure 4.1B) The updating involved residues are preset according to the native structure and fixed during simulation to avoid weak hydrogen bond being ignored in calculation. Although the cost for ECHBPPC applied simulation is smaller than using ECLPPC scheme, due to the less fragments performed in charge updating calculation, ECHBPPC scheme is only valid for studying specific systems in which hydrogen bonding interactions play primary role in the binding. From this aspect, ECLPPC scheme is more practical to be applied in common.

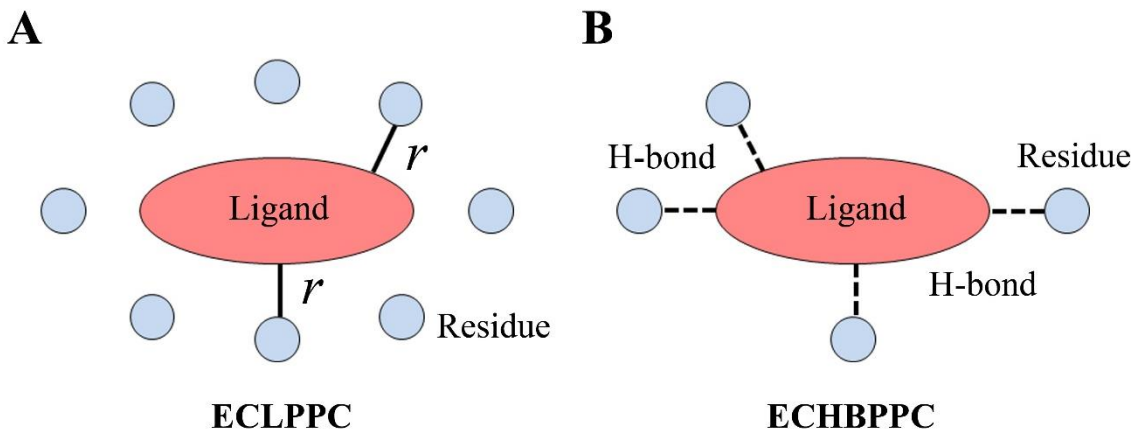
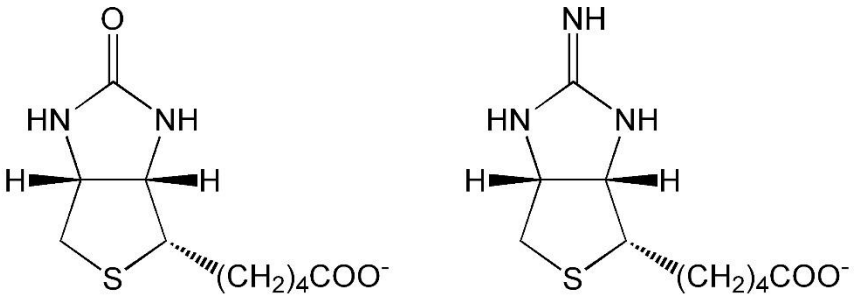


Figure 4.1. Diagrammatic illustration for ECLPPC (A) and ECHBPPC (B).

In this project, the newly proposed protocols are applied to analyze the mechanism of avidin-biotin binding system. Avidin was discovered by Snell and coworkers, which is responsible for biotin binding in egg white causing 'egg white injury'.[142] As a high binding affinity (20 kcal/mol) protein-ligand binding system, it draws researchers' attentions on studying its binding free energy

and it has been used as a benchmark test system.[126, 143-144] Tong et al. stated electrostatic interactions in between avidin and biotin (BTN1) make major contribution to the binding free energy and the loss of the hydrogen bond between Tyr33 and iminobiotin (BTN2) in binding with avidin is the major reason for the binding affinity decrease (6.1 kcal/mol) compare with avidin-BTN1 binding.[126] (Figure 4.2) The result suggested it is critical to describe polarization effect accurately during MD simulations while analyzing the binding of avidin-biotin systems.[126] Therefore, avidin-biotin systems are selected to test the proposed schemes, since the advantage of these two schemes is that they are able to describe electrostatic interaction better while the conformation changes dynamically. The result agrees with the experimental result of relative binding affinity and it is the best agreement so far in the literature obtained by a similar method. Some new insights on the mechanism of avidin-biotin binding have also been obtained and presented in the following. In addition, simulations are conducted using other protocols or force fields for comparison purpose, including AMBER03 force field (standard force field), local PPC without updating (polarized force field) and individual fragment based charge updating scheme. The details of applied protocols in this study are described in the following section.



Biotin (BTN1) **2'-iminobiotin (BTN2)**

Figure 4.2. Structure of biotin and 2'-iminobiotin.

4.2 Materials and methods

4.2.1 Initial structure preparation

The initial structures of avidin-BTN1 and –BTN2 complexes are acquired using the similar method as Tong et al. reported.[126] The crystal structure of BTN1 bound to avidin is obtained from the Protein Data Bank (PDB ID: 1AVD[145]). For BTN2, as the crystal structure is in absent, the complex of avidin-BTN2 is generated by mutating BTN1.[126, 143-144] The neutral form of guanidinium group in BTN2 is applied, since it has been proved that neutral form is shown while binds with protein.[146] Structure optimization is conducted on BTN1 and BTN2 respectively at HF/6-31G** level and the electrostatic potential are calculated at B3LYP/cc-PVTZ level using Gaussian03 suit.[147] The initial atomic charges of BTN1 and BTN2 are fitted through restrained electrostatic potential (RESP)[63-64] with Antechamber program in AMBER 10 suit.[25] Leap module is employed to add hydrogen atoms for the complexes. In the complexes, the parameters of avidin are described using AMBER03 force field, while General Amber force field (GAFF)[104] is employed to obtain the parameters for the ligands (BTN1 and BTN2). Approximately 8,500 TIP3P water molecules are added to the complexes in the periodic box with the minimum distance of 10 Å between the surface of box and atom of the complex.[105] The systems are neutralized with addition of counterions. Thereafter, the systems are minimized with treating long-range electrostatic interactions using Particle Mesh Ewald (PME)[107] method, in the meantime a cutoff of 12 Å is used for the non-bonded interactions. All atoms in both avidin-BTN1 and –BTN2 complex systems are optimized to obtain the initial

structures.

4.2.2 Molecular dynamics protocols

The starting structures are obtained using the previous described method. The whole systems are minimized with two steps. Firstly, all atoms of the avidin-biotin complex are restrained while 1000 cycles of steepest descent minimization followed by 3000 cycles conjugated gradient minimization are conducted. Secondly, the whole system is optimized for 10000 steps through steepest descent followed by conjugate gradient minimization until the convergence of energy gradient reached 0.01 kcal/mol/Å. Thereafter, the system is heated to 300 K from 0 K over 100 ps, with a time step of 2fs applied 3 ns NPT simulation following. Langevin[106] dynamics is employed to regulate the temperature of the systems in simulations, using a 4.0 ps^{-1} collision frequency. The SHAKE algorithm[108] is used to restrain all hydrogen atoms involved bonds. Three simulations are conducted for each system using different methods or force field. Simulation conducted with employing the standard force field (AMBER03 force field[29]) is using the protocol described above, and other applied schemes are described as follows. Preparation varies while applying different schemes.

(a) ECLPPC scheme

Under this scheme, LPPC is updated at every 50 ps during the simulation. LPPC of selected residues and ligand in avidin-BTN1 and –BTN2 complexes are obtained using delta RESP approach[127] applied PPC method[46] based on the last simulated structure of each 50 ps. The details of

PPC calculation can be found in Ref.46. LPPC calculation is only performed on the ligand and the residues involved in 4 Å radius surrounding the ligand. (Figure 4.3A) Approximately 20 residues are included in the updating process. These selected residues are changing dynamically during simulation according to the position shift of the ligand. Atomic charges of the rest protein are kept at AMBER03 charges which are treated as point charges in background during LPPC calculation. Each selected residues and ligand is calculated at B3LYP/6-31G* level with Gaussian03 software[147]. The partial charges are substituted with LPPC meanwhile other parameters and conditions are kept intact during the next 50 ps simulation.

(b) ECHBPPC scheme

Similar with ECLPPC scheme, the PPC calculation are conducted for selected residues and ligand at every 50 ps during MD simulations. Besides, only the residues involved in hydrogen bonding interactions with the ligand and the ligand itself are selected. And the selection of residues are preset at the beginning of the simulation according to the hydrogen bonds presented in initial structures and remain fixed during simulation instead of dynamical selection in ECLPPC scheme. The rest settings and approaches of this protocol are the same as in ECLPPC. (Figure 4.3B)

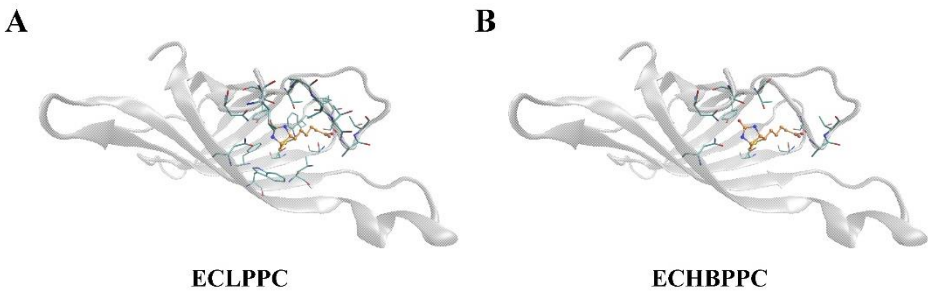


Figure 4.3. Selected residues in updating schemes. (A) ECLPPC includes residues (*cyan*) surrounding biotin (*orange*) within 4 Å; (B) ECHBPPC includes residues (*cyan*) involved in hydrogen bonding interactions with biotin (*orange*).

(c) LPPC

Same as in ECLPPC scheme, LPPC of avidin-BTN1 and –BTN2 complexes are obtained using delta RESP applied PPC method based on the initial structures. The ligand and residues within 4 Å radius of the ligand in the initial structures are selected for LPPC calculation; in the meantime, the rest atoms are treated as background charges in the calculation. QM calculation is also performed at B3LYP/6-31G* level using Gaussian03 software. The atomic charges are replaced by LPPC at the starting point of simulation with other parameters retained.

(d) Environment independent charge updating scheme

In this scheme, atomic charges of the ligand and residues within 4 Å radius of the ligand are derived by fitting the electrostatic potential obtained from quantum mechanical calculation based on individual fragment without treating the rest atoms as background charges. In this case, atomic charges of the selected fragments are determined by their current conformation regardless the local environment. The QM calculation is still performed at B3LYP/6-31G* level with Gaussian03 software.

4.2.3 Binding free energy calculation

Same method as in the previous chapters, the calculation of binding free energies of the avidin-BTN1 and -BTN2 complexes are described through the Eqn. 4.1.

$$\Delta G_{bind} = G_{avidin-BTN1,2} - G_{avidin} - G_{BTN1,2} \quad (4.1)$$

where $G_{avidin-BTN1,2}$, G_{avidin}, $G_{BTN1,2}$ is the free energy of complex, receptor and ligand respectively. ΔG_{bind} is calculated using MM/PBSA method. In Eqn. 4.2, it describes the calculation for the binding free energy.

$$\begin{aligned} \Delta G_{bind} &= \Delta G_{pbsa} - T\Delta S_{solute} \\ &= \Delta E_{MM} + \Delta G_{sol} - T\Delta S_{solute} \\ &= \Delta E_{ele} + \Delta E_{vdW} + \Delta G_{pol} + \Delta G_{nonpol} - \mathrm{T}\Delta S_{solute} \end{aligned} \quad (4.2)$$

where ΔG_{sol} contains ΔG_{pol} and ΔG_{nonpol}, representing the polar and non-polar solvation energy of the system. $T\Delta S_{solute}$ describes the entropy of the solute. Van der Waals and electrostatic interaction energies are described by ΔE_{vdW} and ΔE_{ele} respectively.

The binding affinity decrease from BTN1 to BTN2 in binding with avidin is acquired according to the following equation.

$$\Delta\Delta\mathrm{G} = \Delta G_{BTN2} - \Delta G_{BTN1} \quad (4.3)$$

where ΔG_{BTN1} and ΔG_{BTN2} represent the free binding energy of biotin and iminobiotin to avidin respectively.

1500 snapshots are extracted from the last 1.5 ns simulations at 1 ps intervals to conduct the calculations. The solvation energy is calculated using PBSA module. The grid spacing was set to 0.25 Å. The nonpolar solvation energy is calculated according to the equation G_nonpol= γSAS+ β, where γ = 0.00542 kcal/(mol ·Å2) and β= 0.92 kcal/mol.[126] Entropy loss is obtained using the Nmode[136] module in AMBER 10 package. For each system, 150 snapshots taken at 10 ps interval from the second half of the MD trajectory are calculated for the loss of entropy. Structure of complex in each snapshot is optimized until the rms of the energy gradient reaches 10^{-5} kcal/mol/$Å^2$ with a distance-dependent dielectric function. The corresponding atomic charges of each duration of simulation are applied in both enthalpy and entropy calculation.

4.3. Results and discussion

4.3.1 Atomic charge fluctuating in updating schemes

To evaluate the charge variation during MD simulations using updating schemes, the charge fluctuation of the selected atoms in ECLPPC are plotted in Figure 4.4. It shows the charge fluctuation of O3 in BTN1 and N2 in BTN2 evolves with time during simulation. Both atomic charges of O3 and N2 are dramatically decreased at the first updating point. Since the initial atomic charges of the ligands are utilized at the beginning of the simulation, which are obtained without considering its electrostatic environment in binding with avidin. In the LPPC calculation, polarization effect from protein, especially from the formation of hydrogen bonds, is included

leading to a more negative refitted charge for atom O3 of BTN1 and N2 of BTN2 respectively. In both fluctuation curves, it has been observed that the atomic charge tends to vary gently along with simulation. Particularly, the charge of O3 in BTN1 varies around -0.85*e* in the last half of the simulation, while it is -1.30*e* for the charge of N2 in BTN2. It suggests that the binding mode of ligand is gradually getting stable. The charge fluctuation in ECHBPPC scheme applied simulation is similar with the one using ECLPPC scheme.

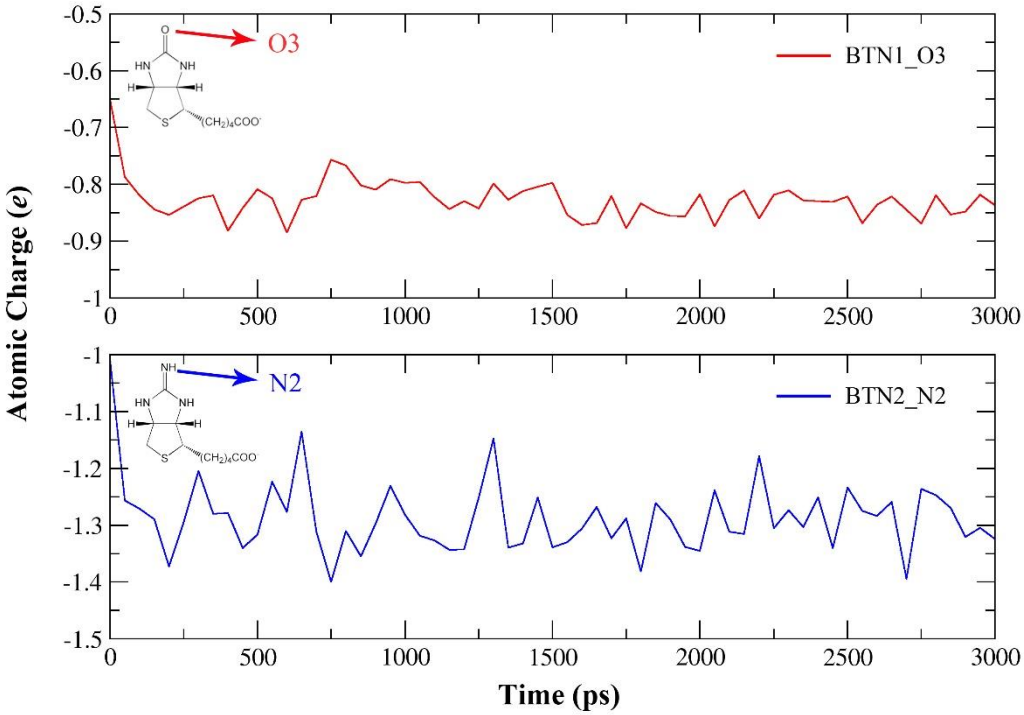


Figure 4.4. Charge variation of atom O3 in BTN1 (*top*) and atom N2 in BTN2 (*bottom*) during MD simulations using ECLPPC scheme.

To evaluate the rationality of the updated charges, four conformations of avidin-BTN1 complex are extracted randomly from ECLPPC scheme used simulation to investigate the relation between the refitted charge of atom O3 and current conformation. From the structure in Figure 4.5A, it shows that atom O3 of BTN1 forms three hydrogen bonds with three residues Asn12, Ser16 and Tyr33

respectively. According to the rationale of PPC, the strength of these hydrogen bonds should make primary contributions on the variation of O3's charge in LPPC calculation, since hydrogen bond is treated as a strong electrostatic interaction. In which, O3 of BTN1 possessing the strongest hydrogen bonds in total within these four snapshots will have the most negative atomic charge. The total strength of these three hydrogen bonds can be represented by the sum of electrostatic interaction energies of Asn12, Ser16 and Tyr33 to BTN1. The electrostatic interaction energies are calculated based on the initial charges (AMBER03 charges) for all the selected snapshots to eliminate the different refitted charges caused effects on the energy in each snapshot. In Figure 4.5B, the conformation with the lowest total electrostatic interaction energy (-16.302 kcal/mol) possesses the lowest atomic charge of O3 (-0.878*e*) as expected, indicating that the atomic charge of O3 is polarized properly according to the strength of formed hydrogen bonds in the current conformation which subsequently suggests the charge updating is reasonable based on the current electrostatic environment. Furthermore, with using charge updating schemes, the atomic charges of each updated fragment are equivalent to the average LPPC obtained from multiple conformations.

For comparison, the fluctuations of atomic charges of O3 in BTN1 and N2 in BTN2 under environment independent charge updating scheme are plotted in Figure 4.6. Unlike the charge variation observed in ECLPPC scheme, the atomic charges only vary slightly from their initial charges regardless to the formation of hydrogen bonds, due to lack of polarization effect. It proves that it is essential to consider the local electrostatic environment while fitting the atomic charges in a specific protein from another point of view.

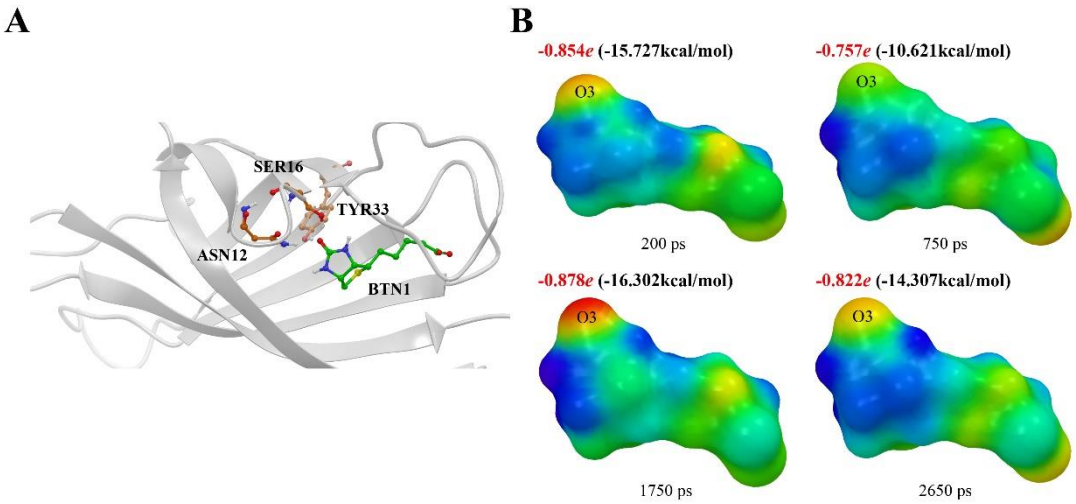


Figure 4.5. (A) Hydrogen bonding interactions between atom O3 of BTN1 and avidin in binding pocket. (B) Relationship between atomic charge and electrostatic environment. The atomic charge of O3 in BTN1 (*red*) is positively correlated to the total electrostatic interaction energy of residues Asn12, Ser16 and Tyr33 (*black*). BTN1 is shown electrostatic potential surface at different updating point (200ps, 750ps, 1750ps and 2650ps). The ESP surface is generated according to the difference of atomic charges of BTN1 between the current and initial charges ($\Delta Chg = Chg_{current} - Chg_{initial}$). The current atomic charges are obtained by LPPC calculation based on the selected snapshots, and the initial charges are described in the method section for initial structure preparation.

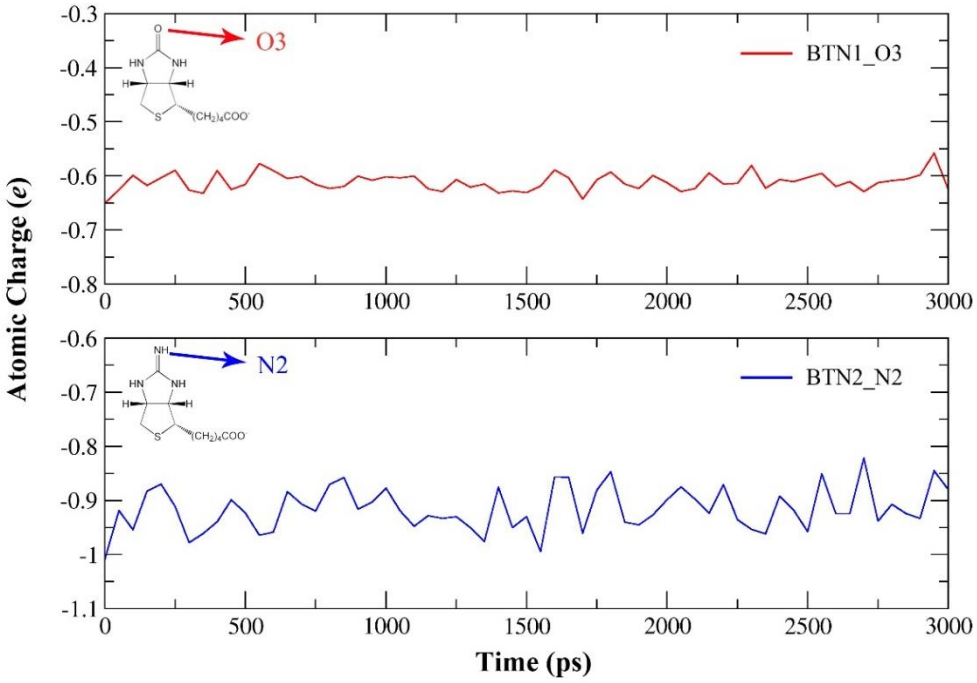


Figure 4.6. Charge variation of atom O3 in BTN1 (*top*) and atom N2 in BTN2 (*bottom*) during MD simulations under environment independent charge updating scheme.

4.3.2 Thermodynamic analysis

The three trajectories for each system are used to estimate binding free energies with MM/PBSA method. Since it is known that the difference in binding free energies of avidin-BTN1 and -BTN2 complexes is 6.1 kcal/mol in the experimentally measured result, the performances of the applied protocols are compared based on the experimental value. All the binding free energies results of avidin-BTN1 and -BTN2 complexes using different approaches are tabulated in Table 4.1, and ΔΔG values obtained based on the average values of ΔG_{bind} from multiple trajectories are plotted in Figure 4.7 to illustrate straightforwardly.

Under AMBER03 force field, the calculated relative binding free energy, 2.95 kcal/mol, is observed to be failed to reveal the large difference showed in experimental value. In addition, the standard deviation is reached 4.34 kcal/mol in binding free energies of avidin-BTN1 which is 19.69% of the average energy value, indicating the instability in performance by applying AMBER03 charges for this system during simulation. Using the LPPC obtained based on a single conformation, the difference in free energy of binding avidin to BTN1 and BTN2 is 3.91 kcal/mol. Comparing with using AMBER03 charge, it presents better agreement with experimental data. And at the same time, the standard deviations of the average energies are smaller suggesting it is more consistent in performance in different trajectories while using LPPC in MD simulations. The importance of including polarization effect is emerged.

Table 4.1. Theoretical and experimental obtained binding free energy of avidin-BTN1 and -BTN2 complexes under different protocols. (Number in parentheses are standard error.)

kcal/mol	BTN1	ΔH	ΔST	ΔG	BTN2	ΔH	ΔST	ΔG	ΔΔG
AMBER03	1	-36.58	-19.30	-17.28	1	-38.25	-17.80	-20.45	
		(0.16)	(0.14)	(0.21)		(0.12)	(0.19)	(0.23)	
	2	-44.68	-18.90	-25.78	2	-37.31	-19.20	-18.11	
		(0.18)	(0.11)	(0.21)		(0.22)	(0.13)	(0.25)	
	3	-43.55	-20.50	-23.05	3	-38.51	-19.80	-18.71	
		(0.11)	(0.14)	(0.18)		(0.12)	(0.15)	(0.19)	
	Average			**-22.04**	Average			**-19.09**	**2.95**
	SE			(0.35)	SE			(0.39)	(0.53)
	Stdev			4.34	Stdev			1.22	
LPPC	1	-60.70	-18.39	-42.31	1	-58.31	-20.51	-37.80	
		(0.12)	(0.16)	(0.20)		(0.08)	(0.12)	(0.14)	
	2	-62.95	-20.66	-42.29	2	-60.42	-20.26	-40.16	
		(0.10)	(0.11)	(0.15)		(0.11)	(0.17)	(0.20)	
	3	-62.16	-19.26	-42.90	3	-58.59	-20.77	-37.82	
		(0.17)	(0.13)	(0.21)		(0.08)	(0.08)	(0.11)	
	Average			**-42.50**	Average			**-38.59**	**3.91**
	SE			(0.33)	SE			(0.27)	(0.43)
	Stdev			0.35	Stdev			1.36	
ECLPPC	1	-63.21	-17.57	-45.64	1	-60.04	-21.70	-38.34	
		(0.11)	(0.15)	(0.19)		(0.16)	(0.19)	(0.25)	
	2	-63.08	-19.44	-43.64	2	-57.55	-20.74	-36.81	
		(0.11)	(0.10)	(0.15)		(0.11)	(0.05)	(0.12)	
	3	-62.69	-19.79	-42.90	3	-58.64	-21.39	-37.25	
		(0.12)	(0.08)	(0.14)		(0.12)	(0.13)	(0.18)	
	Average			**-44.06**	Average			**-37.47**	**6.59**
	SE			(0.28)	SE			(0.33)	(0.43)
	Stdev			1.42	Stdev			0.79	

ECHBPPC	1	-61.75	-19.70	-42.05	1	-58.84	-23.73	-35.11	
		(0.14)	(0.09)	(0.16)		(0.12)	(0.17)	(0.21)	
	2	-59.68	-19.50	-40.18	2	-57.95	-20.51	-37.44	
		(0.12)	(0.12)	(0.17)		(0.14)	(0.11)	(0.18)	
	3	-61.05	-21.64	-39.41	3	-52.66	-21.45	-31.21	
		(0.12)	(0.10)	(0.16)		(0.11)	(0.17)	(0.20)	
	Average			**-40.55**	Average			**-34.59**	**5.96**
	SE			(0.28)	SE			(0.34)	(0.45)
	Stdev			1.36	Stdev			3.15	
EIDC	1	-31.37	-17.14	-14.23	1	-32.90	-15.36	-17.54	
		(0.11)	(0.08)	(0.13)		(0.10)	(0.11)	(0.15)	
	2	-36.37	-18.75	-17.62	2	-36.39	-18.64	-17.75	
		(0.09)	(0.08)	(0.13)		(0.13)	(0.05)	(0.14)	
	3	-36.60	-17.76	-18.84	3	-34.11	-18.76	-15.35	
		(0.13)	(0.06)	(0.14)		(0.11)	(0.20)	(0.23)	
	Average			**-16.90**	Average			**-16.88**	**0.02**
	SE			(0.23)	SE			(0.30)	(0.38)
	Stdev			2.39	Stdev			1.33	
Expt.				**-20.4**				**-14.3**	**6.1**

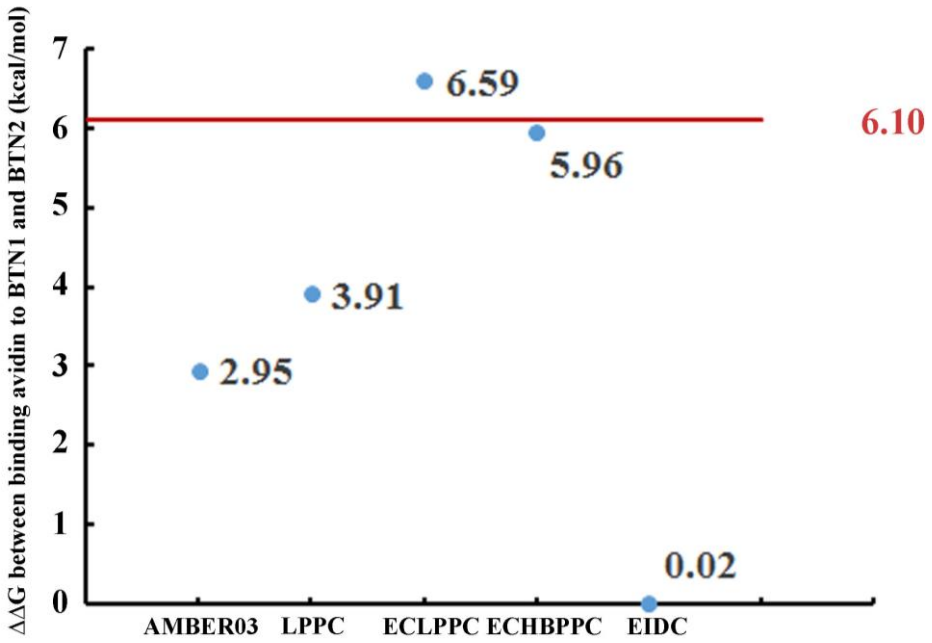


Figure 4.7. Difference between binding free energy of avidin-BTN1 and -BTN2 complexes under different protocols. Experimental data is shown as the red line.

However, there is still a gap between the theoretically obtained relative binding free energy and the experimental data. This gap is successfully filled with using ECLPPC and ECHBPPC schemes. The theoretically obtained relative binding free energy of avidin-BTN1 and -BTN2 systems are 6.59 kcal/mol and 5.96 kcal/mol under ECLPPC and ECHBPPC respectively. The experimentally obtained value is well reproduced by using both updating schemes. And the standard deviations of the average free energies are also acceptable. On the other hand, a bad performance is observed using the environment independent charge updating scheme. Though the atomic charges are updated in simulation, the refitted charges are derived only based on the current conformation without considering the electrostatic environment in the complex. In addition, the current conformation of each individual fragment is not optimized before QM calculations and to reduce the cost of QM calculations, an even smaller basis set (6-31G*) is utilized comparing with the basis set (ccpVTZ) used while generating AMBER03 charges[29]. Therefore, it is predictable that it gives a

worse performance than using AMBER03 charges. At the meantime, it is a further proof of the necessity of polarization effect in MD simulations. The reason is discussed in the following for ECLPPC and ECHBPPC successfully reproducing the relative free energy of binding avidin to BTN1 and BTN2 while using other protocols are failed. And the factor governing the large relative binding affinity is explored in the next section as well.

4.3.3 Hydrogen bond network

To investigate the basis of the significant binding affinity decrease in avidin-BTN2 binding, the hydrogen bonding interactions come into our attention because of the following points. Firstly, by comparing the structures of BTN1 and BTN2, the only change is the carboxyl group of BTN1 being substituted by imino group in BTN2 in the head part, which makes BTN2 less favorable in forming hydrogen bond. Secondly, it has been pointed out that the loss of hydrogen bond in avidin-BTN2 is mainly responsible for the binding affinity difference. In addition, the result obtained from ECHBPPC scheme applied simulation reaches good agreement with the experimental data, suggesting the rearrangement of hydrogen bond interactions between avidin and BTN2 could be primarily reason for the significant decrease in binding affinity. Therefore, the hydrogen bonds between avidin and BTN1 and BTN2 are analyzed. The occupancy of hydrogen bonds during simulations applying all different protocols are displayed in Figure 4.8.

At first, ECLPPC and ECHBPPC scheme applied simulations are utilized to interpret the basis of

the loss in binding affinity, since the results in relative binding free energy calculation are consistent with experimental data. In ECLPPC applied simulations (Figure 4.8B), the hydrogen bond formed between Asn12 and BTN1 (76.07%) is obviously stronger than that in avidin-BTN2 complex (15.37%). Besides this hydrogen bond, the strength of hydrogen bonds of Asn118-BTN2, Ser16-BTN2 and Thr35-BTN2 are decreased comparing with the corresponding hydrogen bonds in avdin-BTN1 complex. Meanwhile, the rest hydrogen bonds are maintained well in both complexes. The weakened hydrogen bonds between BTN2 and Asn12, Asn118, Ser16 and Thr35 respectively should altogether be responsible for the decreased free energy of binding BTN2 to avidin, which conflicts the statement made by Tong et al.[126] that the missing hydrogen bond between Tyr33 and BTN2 is primarily responsible. The occupancy of hydrogen bond obtained in ECHBPPC applied simulations is slightly discriminative with using ECLPPC scheme. (Figure 4.8C) The strength of hydrogen bonds of Asn12-BTN2 and Thr35-BTN2 shows obvious decreases, while the hydrogen bonding interactions of Asn118-BTN2 and Ser16-BTN2 are retained well. Even so the weakened hydrogen bonding interactions caused by the imino group replacement in the head group of BTN2 should still be considered as main the reason for the binding affinity loss.

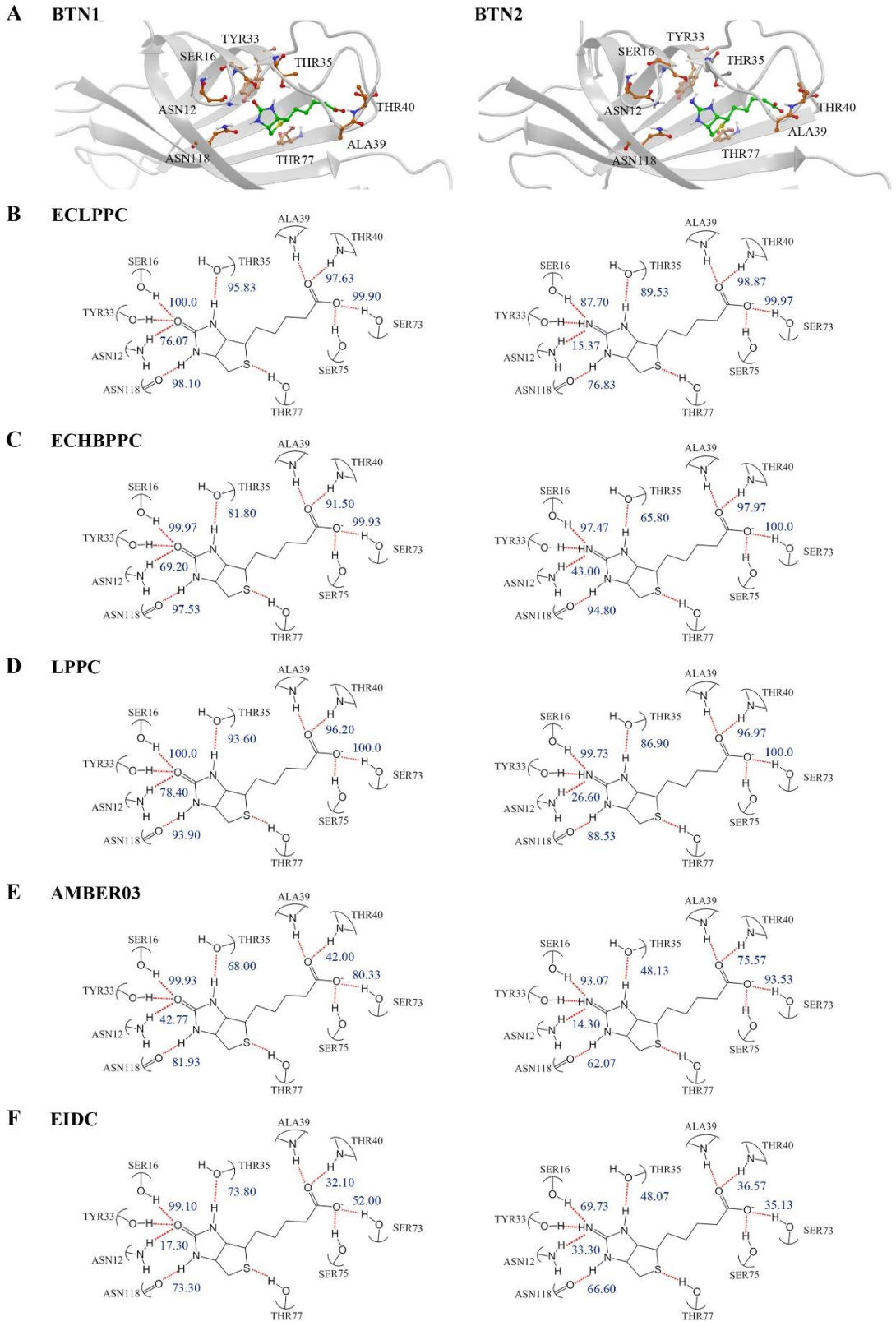


Figure 4.8. Hydrogen bond network in binding pocket of BTN1- (left) and BTN2-complexed avidin (right) and occupancy in simulations using different protocols. (A) 3D conformation; (B) ECLPPC; (C) ECHBPPC; (D) LPPC; (E) AMBER03; (F) EIDC (environment independent charge updating scheme. Hydrogen bonds are shown in red dash line with corresponding occupancy labaled.

Reason of the dissatisfied results when using other protocols is illustrated in the following. In LPPC applied simulations, the occupancy of hydrogen bonds are similar with the value obtained with ECLPPC scheme, except the hydrogen bonds between BTN2 and Asn12, Ser16 and Asn118 which are more stable than in ECLPPC applied simulation. (Figure 4.8D) In particular, the occupancy for these hydrogen bonds are increased over 10% using the non-updating scheme. It may be caused by using LPPC calculated on a single structure which would lead to over stabilizing on unfavorable hydrogen bonds. Thus, in the binding free energy calculation, this scheme fails to reproduce the large binding affinity difference between avdin binding to BTN1 and BTN2. Though the relative binding free energy obtained from AMBER03 charge applied simulations is slightly lower than that obtained from LPPC applied simulations, the hydrogen bond network is highly distinguishing with each other. In the previous discussed hydrogen bond networks, the dissociated carboxyl group of biotin and iminobiotin forms stable hydrogen bonding interactions with residues Ala39, Thr40, Ser73 and Ser75, due to the negative charges possessed by the oxygen atoms making them favorable in forming hydrogen bonds. And the weakened hydrogen bonds formed by the headgroup in avidin-BTN2 complex make the major contributions in the binding affinity loss. Nevertheless, in the simulations using AMBER03 charges, it shows the hydrogen bonds formed by the carboxyl group are unstable in both BTN1 and BTN2 bounded complexes. (Figure 4.8E) It explains why the difference in binding free energy of avidin-BTN1 and –BTN2 complexes is fail to be reproduced using conventional force field, though the occupancy of hydrogen bonds between BTN2 and Asn12, Asn118 and Thr35 respectively is decreased comparing with in avidin-BTN1 complex. Even less stable hydrogen bonds involving the carboxyl group of biotin and its analog are found in simulations using environment independent charge updating scheme. (Figure 4.8F) By comparing

LPPC involved schemes with non-PPC applied simulations, it is notable that LPPC is conducive to stabilize hydrogen bonds and describe electrostatic interactions correctly in MD simulations, accountable to the inclusion of polarization effect.

In addition, the secondary structure of avidin is also better preserved in LPPC involved simulations, especially those involved in forming the binding pocket. Although the denaturation of local structure at ASN118-ILE119, which is mentioned in Tong's work[126], is not observed, the unstable hydrogen bond interactions under AMBER03 force field have also been captured between Ser75 and Ser101. (Figure 4.9) This proves introduce LPPC into MD simulations is better for retaining the secondary structure of binding pocket and provides more reliable information in protein dynamics.

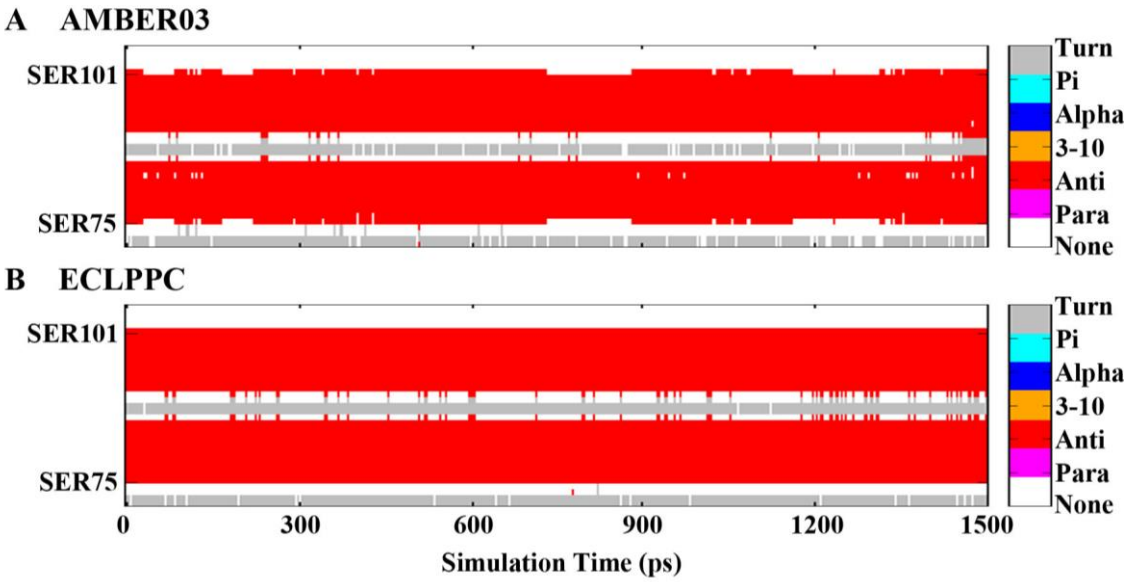


Figure 4.9. Secondary structure of avidin in the selected region during simulation applying AMBER03 charge (A) and ECLPPC scheme (B).

4.4 Conclusion

In this study, PPC updating scheme shows its ability in varying the atomic charges accurately and effectively based on the real-time conformation during MD simulations. Both ECLPPC and ECHBPPC schemes provide accurate relative binding free energies between avidin-biotin and avidin-iminobiotin complexes and give rational explanation on the large difference in binding affinity of these two systems. The hydrogen bonds formed between residues within the active site of avidin and the ureido group of biotin are weakened while avidin binds to iminobiotin. In particular, Asn12, Ser16, Thr35 and Asn118 involved hydrogen bonds are suspected being responsible for the binding affinity decrease in avidin-iminobiotin complex. In addition, with comparison, applying LPPC acquired based on single conformation in MD simulations may cause overestimation for the strength of undesirable hydrogen bonds, and using updating scheme should be able to minimize this deficiency. Moreover, the absence of polarization effect in standard force field leads to failure in describing hydrogen bonding interactions and subsequently causing the inaccurate estimation in binding free energies. Furthermore, it needs to be noted that since hydrogen bond interactions plays a crucial role in this benchmark system, ECHBPPC scheme, which only updates the residues involved in forming hydrogen bonds with ligand, reaches good agreement with the experimental data; however, this case is not common in protein-ligand binding systems, which making ECLPPC scheme a more practical method in analyzing protein-ligand interactions.

Chapter 5

Summary

In this thesis, protein-ligand interactions in several systems, including HIV-1 integrase-INSTIs, HIV-1 protease-PIs and avidin-biotin interactions, are investigated by using molecular dynamics simulation with different levels of polarization.

In the beginning, the mechanism of HIV-1 integrase G140S-Q148H double mutation induced drug resistance to RAL and DTG is researched using conventional MD simulations. Firstly, role of the catalytic metal ions in the active site and INSTIs protonation state is discussed. The deprotonated INSTIs are applied due to their ability in keeping the magnesium ions steady during simulation. The relative binding free energies obtained from simulations reach good agreement with experimental value in both RAL- and DTG-complexed HIV-1 integrase variants. In which, DTG shows its capability against the drug resistance caused by double mutation, while the binding affinity significantly decreases in the double mutant-RAL complex. It is caused by the larger increase in fluctuation of the 140s' loop in RAL-complexed mutant. Since the loop is one of the most important components of the binding pocket, the deceased stability of it will directly affect the interaction between the 140s' loop and RAL. The broken of hydrogen bond between residues 145 and 148 is mainly responsible for the higher fluctuation of the loop. And it is induced by the new hydrogen bond formation in between S140 and H148 with stretching force from RAL on the 145-148 hydrogen bond. On the other hand, DTG induces little stretching force on the 145-148 hydrogen bond leading to a stable loop while binding. Subsequently, DTG is

observed possessing the ability of combating drug resistance. In this study, new insights are provided on the basis of G140S-Q148H mutation causing significant resistance on RAL and the reason for DTG to be able to against it. Insights on design of next generation INSTIs have also been provided.

In the second study, the basis of drug resistance on indinavir and amprenavir induced by I50V mutation in HIV-1 PR is explored by applying PPC in MD simulations. Comparing with conventional force field, PPC provides better results in fitting with the experimental data suggesting the significance of including polarization effect in simulations. By analyzing the obtained trajectories, the binding pocket in I50V HIV-1 protease is found enlarged indicating the binding pocket is more solvent accessible. The interactions between protease and PIs are directly weakened by the mutation and the degree of freedom of PIs in the active site has also been increased. Consequently, the binding affinity of PIs to mutated HIV-1 PR is decreased. Through this study, a rational explanation of how I50V mutation causing drug resistance to PIs is provided and an idea on designing more potent PIs have been given.

Lastly, PPC updating scheme is applied into protein-ligand interaction study. ECLPPC and ECHBPPC schemes are proposed and utilized in analyzing avidin-biotin interactions. Both schemes provide consistent relative binding free energy with the experimental data and provide a rational explanation on the significant difference in free energy of binding avidin to biotin and 2'-iminobiotin. The weakened hydrogen bonds in avidin-iminobiotin complex are suspected as major reason for the binding affinity loss. By comparing with other applied protocols, it has been

found that it is important to incorporate polarization effect into MD simulations and at the same time it is necessary to use the 'on-the-fly' charge or PPC from multiple conformations due to the conformational change during simulation.

Though MD simulations has shown its power in study protein-ligand interactions and current methods to introduce polarization effect have presented an improvement in MD simulations studies, these methods still can be improved to provide more accurate description of biological systems. With development of calculation power or data processing techniques, a more accurate and efficient method compare to the current method for generating PPC can be developed; also the number of updated atoms and updating frequency during MD simulations can be increased. The problem of PPC has also been noticed during studying mutation induced drug resistance in HIV-1 integrase. PPC was also implemented in this project, however, it overestimated the binding affinity difference between WT-INSTIs and mutant-INSTIs complexes. This issue of PPC needs to be solved in the future. Besides incorporating polarization effect in molecular dynamics simulations, other parameters in current force field can also be improved by using different QM based fitting method. In summary, MD simulations is a powerful tool that can predict the dynamics and properties of biomolecules and molecule from other systems, however, it still can be improved in many aspect to enhance its accuracy and efficiency in the future.

List of Publications

i. Duan, R.; Lazim, R.; Zhang, D. Understanding the basis of I50V-induced affinity decrease in HIV-1 protease via molecular dynamics simulations using polarized force field. J. Comput. Chem. 2015, 36, 1885-1892. (selected as front cover of the issue)

ii. Duan, R.; Zhang, D. Theoretical study of avidin-biotin binding via molecular dynamics simulations using polarized force field. (submitted)

iii. Duan, R.; Zhang, D. Theoretical study of impact of G140S/Q148H double mutation on inhibitor binding to HIV-1 integrase. (submitted)

References

1. Rahman, A., Correlations in the motion of atoms in liquid argon. *Physical Review* **1964,** *136*, 405.

2. Alder, B. J.; Wainwright, T. E., Studies in molecular dynamics .1. general method. *J. Chem. Phys.* **1959,** *31*, 459.

3. Levitt, M.; Warshel, A., Computer simulation of protein folding. *Nature* **1975,** *253*, 694-698.

4. Warshel, A., Bicycle-pedal model for the first step in the vision process. *Nature* **1976,** *260*, 679.

5. McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins. *Nature* **1977,** *267*, 585.

6. Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J.; Mackenzie, K. M.; Bank, J.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J., In Millisecond-scale molecular dynamics simulations on Anton, *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* **2009,** 1.

7. Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W., Atomic-level characterization of the structural dynamics of proteins. *Science* **2010,** *330*, 341.

8. Sobieraj, M.; Krzysko, K. A.; Jarmula, A.; Kalinowski, M. W.; Lesyng, B.; Prokopowicz, M.; Ciesla, J.; Gojdz, A.; Kierdaszuk, B., A QM-MD simulations approach to the analysis of FRET processes in (bio)molecular systems. A case study: complexes of E. coli purine nucleoside phosphorylase and its mutants with formycin A. *Journal of molecular modeling* **2015,** *21*, 75.

9. Raczynski, P.; Dawid, A.; Gburski, Z., Molecular dynamics (MD) in homocysteine nanosystems - computer simulation. *Biomolecular engineering* **2007,** *24*, 577.

10. Abhilash, J.; Haridas, M., Metal ion coordination essential for specific molecular interactions of butea monosperma lectin: ITC and MD simulation studies. *Applied biochemistry and biotechnology* **2015,** *176*, 277.

11. Abramavicius, D.; Jiang, J.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S., Simulation study of chiral two-dimensional ultraviolet spectroscopy of the protein backbone. *Journal of the American Chemical Society* **2010,** *132*, 7769.

12. Abseher, R.; Schreiber, H.; Steinhauser, O., The influence of a protein on water dynamics in its vicinity investigated by molecular dynamics simulation. *Proteins* **1996,** *25*, 366.

13. Ahmed, M.; Sadek, M. M.; Serrya, R. A.; Kafafy, A. H.; Abouzid, K. A.; Wang, F., Assessment of new anti-HER2 ligands using combined docking, QM/MM scoring and MD simulations. *Journal of molecular graphics & modelling* **2013,** *40*, 91.

14. Alberga, D.; Mangiatordi, G. F.; Motta, A.; Nicolotti, O.; Lattanzi, G., Effects of different self-assembled monolayers on thin-film morphology: a combined DFT/MD simulations protocol. *Langmuir : the ACS journal of surfaces and colloids* **2015,** *31*, 10693.

15. Alder, N. N.; Sutherland, J.; Buhring, A. I.; Jensen, R. E.; Johnson, A. E., Quaternary structure of the mitochondrial TIM23 complex reveals dynamic association between Tim23p and other subunits. *Molecular biology of the cell* **2008,** *19*, 159.

16. van Gunsteren, W. F.; Weiner, P. K.; Wilkinson, A. J., Computer simulation of biomolecular systems: theoretical and experimental applications. *Springer Science & Business Media* 2013, *3*.

17. Becker, O. M.; MacKerell Jr, A. D.; Roux, B.; Watanabe, M., Computational biochemistry and

biophysics. *CRC Press* **2001.**

18. Berendsen, H. J.; Hayward, S., Collective protein dynamics in relation to function. *Current opinion in structural biology* **2000,** *10*, 165.

19. Warshel, A.; Levitt, M., Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology* **1976,** *103*, 227.

20. Polymeropoulos, E.; Warshel, A., Computer Modeling of Chemical Reactions in Enzymes and Solutions, *Berichte der Bunsengesellschaft für physikalische Chemie* **1992,** *96*, 1323.

21. Field, M. J.; Bash, P. A.; Karplus, M., A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry* **1990,** *11*, 700.

22. Nobelprize.org, The Nobel Prize in Chemistry 2013 - Press Release. *Nobel Media AB 2014. web* **2014**.

23. Petrenko, R.; Meller, J., Molecular dynamics. *eLS* **2010**.

24. Lindahl, E. R., Molecular dynamics simulations. *Molecular modeling of proteins, Springer:* **2008,** 3.

25. Case, D.; Darden, T.; Cheatham, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W. Amber 10. *University of California* **2008**.

26. Verlet, L., Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review* **1967,** *159*, 98.

27. Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R., A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules:

Application to small water clusters. *The Journal of Chemical Physics* **1982,** *76*, 637.

28. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **1995,** *117*, 5179.

29. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T., A point‐charge force field for molecular mechanics simulations of proteins based on condensed‐phase quantum mechanical calculations. *Journal of computational chemistry* **2003,** *24*, 1999.

30. Jensen, F., Introduction to computational chemistry. *John Wiley & Sons* **2013**.

31. Jones, J. E., On the determination of molecular fields. ii. from the equation of state of a gas. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **1924,** *106*, 463.

32. Zhigilei, L., Molecular Dynamics. *University of Virginia* **2015**.

33. Baker, C. M., Polarizable force fields for molecular dynamics simulations of biomolecules. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2015,** *5*, 241.

34. Mortier, W. J.; Ghosh, S. K.; Shankar, S., Electronegativity-equalization method for the calculation of atomic charges in molecules. *Journal of the American Chemical Society* **1986,** *108*, 4315.

35. Bauer, B. A.; Patel, S., Recent applications and developments of charge equilibration force fields for modeling dynamical charges in classical molecular dynamics simulations. *Theoretical Chemistry Accounts* **2012,** *131*, 1.

36. Lucas, T. R.; Bauer, B. A.; Patel, S., Charge equilibration force fields for molecular dynamics simulations of lipids, bilayers, and integral membrane protein systems. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2012,** *1818*, 318.

37. Lamoureux, G.; Roux, B. t., Modeling induced polarization with classical drude oscillators: Theory and molecular dynamics simulation algorithm. *The Journal of Chemical Physics* **2003,** *119*, 3025.

38. Lamoureux, G.; MacKerell Jr, A. D.; Roux, B., A simple polarizable model of water based on classical drude oscillators. *The Journal of chemical physics* **2003,** *119*, 5185.

39. Yu, H.; van Gunsteren, W. F., Accounting for polarization in molecular simulation. *Computer physics communications* **2005,** *172*, 69.

40. Davis, M. E.; McCammon, J. A., Electrostatics in biomolecular structure and dynamics. *Chemical Reviews* **1990,** *90*, 509.

41. Maple, J. R.; Cao, Y.; Damm, W.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A., A polarizable force field and continuum solvation methodology for modeling of protein-ligand interactions. *Journal of Chemical Theory and Computation* **2005,** *1*, 694.

42. Ma, B.; Lii, J. H.; Allinger, N. L., Molecular polarizabilities and induced dipole moments in molecular mechanics. *Journal of Computational Chemistry* **2000,** *21*, 813.

43. Ren, P.; Ponder, J. W., Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B* **2003,** *107*, 5933.

44. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio Jr, R. A., Current status of the AMOEBA polarizable force field. *The journal of physical chemistry B* **2010,** *114*, 2549.

45. Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V., Fragmentation methods: a route to accurate calculations on large systems. *Chem. Rev* **2012,** *112*, 632.

46. Ji, C.; Mei, Y.; Zhang, J. Z., Developing polarized protein-specific charges for protein dynamics: MD free energy calculation of pK a shifts for Asp 26/Asp 20 in Thioredoxin. *Biophysical journal* **2008,** *95*, 1080.

47. Ji, C.; Zhang, J., Protein polarization is critical to stabilizing AF-2 and helix-2′ domains in ligand binding to PPAR-γ. *Journal of the American Chemical Society* **2008,** *130*, 17129.

48. Tong, Y.; Ji, C. G.; Mei, Y.; Zhang, J. Z., Simulation of NMR data reveals that proteins' local structures are stabilized by electronic polarization. *Journal of the American Chemical Society* **2009,** *131*, 8636.

49. Wei, C.; Lazim, R.; Zhang, D., Importance of polarization effect in the study of metalloproteins: Application of polarized protein specific charge scheme in predicting the reduction potential of azurin. *Proteins: Structure, Function, and Bioinformatics* **2014,** *82*, 2209.

50. Wei, C.; Mei, Y.; Zhang, D., Theoretical study on the HIV-1 integrase–5CITEP complex based on polarized force fields. *Chemical Physics Letters* **2010,** *495*, 121.

51. Mei, Y.; Li, Y. L.; Zeng, J.; Zhang, J. Z., Electrostatic polarization is critical for the strong binding in streptavidin‐biotin system. *Journal of computational chemistry* **2012,** *33*, 1374.

52. Xu, Z.; Mei, Y.; Duan, L.; Zhang, D., Hydrogen bonds rebuilt by polarized protein-specific charges. *Chemical Physics Letters* **2010,** *495*, 151.

53. Duan, R.; Lazim, R.; Zhang, D., Understanding the basis of I50V‐induced affinity decrease in HIV‐1 protease via molecular dynamics simulations using polarized force field. *Journal of computational chemistry* **2015,** *36*, 1885.

54. Manthe, U.; Köppel, H., New method for calculating wave packet dynamics: Strongly coupled surfaces and the adiabatic basis. *The Journal of Chemical Physics* **1990,** *93*, 345.

55. Goedecker, S., Linear scaling electronic structure methods. *Reviews of Modern Physics* **1999,** *71*, 1085.

56. Junquera, J.; Paz, Ó.; Sánchez-Portal, D.; Artacho, E., Numerical atomic orbitals for linear-scaling calculations. *Physical Review B* **2001,** *64*, 235111.

57. Zhang, D. W.; Zhang, J., Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy. *The Journal of chemical physics* **2003,** *119*, 3599.

58. Zhang, D.; Chen, X.; Zhang, J., Molecular caps for full quantum mechanical computation of peptide–water interaction energy. *Journal of computational chemistry* **2003,** *24*, 1846.

59. Gao, A. M.; Zhang, D. W.; Zhang, J. Z.; Zhang, Y., An efficient linear scaling method for ab initio calculation of electron density of proteins. *Chemical physics letters* **2004,** *394*, 293.

60. Singh, U. C.; Kollman, P. A., A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the $CH_3Cl + Cl^-$ exchange reaction and gas phase protonation of polyethers. *Journal of Computational Chemistry* **1986,** *7*, 718.

61. Maseras, F.; Morokuma, K., IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *Journal of Computational Chemistry* **1995,** *16*, 1170.

62. Fedorov, D. G.; Kitaura, K.; Li, H.; Jensen, J. H.; Gordon, M. S., The polarizable continuum model (PCM) interfaced with the fragment molecular orbital method (FMO). *Journal of*

computational chemistry **2006,** *27,* 976.

63. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **1993,** *97,* 10269.

64. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A., Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *Journal of the American Chemical Society* **1993,** *115,* 9620.

65. Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B., Rapid grid‐based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *Journal of computational chemistry* **2002,** *23,* 128.

66. Warwicker, J.; Watson, H. C., Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. *J. Mol. Biol.* **1982,** *157,* 671.

67. Gilson, M. K.; Honig, B., Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins.* **1988,** *4,* 7.

68. Honig, B.; Nicholls, A., Classical electrostatics in biology and chemistry. *Science* **1995,** *268,* 1144.

69. Sitkoff, D.; Sharp, K. A.; Honig, B., Accurate calculation of hydration free energies using macroscopic solvent models. J. Phys. Chem. **1994,** *98,* 1978.

70. Kuhn, B.; Kollman, P. A., Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J Med Chem* **2000,** *43,* 3786.

71. Weiss, R. A., How does HIV cause AIDS? *Science* **1993,** *260*, 1273.

72. Douek, D. C.; Roederer, M.; Koup, R. A., Emerging concepts in the immunopathogenesis of AIDS. *Annual review of medicine* **2009,** *60*, 471.

73. Aids.Gov, Global HIV/AIDS Overview. *https://www.aids.gov* **2015**.

74. Organization, W. H., Global Health Observatory (GHO) data. *http://www.who.int* **2015**.

75. Métifiot, M.; Marchand, C.; Maddali, K.; Pommier, Y., Resistance to integrase inhibitors. *Viruses* **2010,** *2*, 1347.

76. Pommier, Y.; Johnson, A. A.; Marchand, C., Integrase inhibitors to treat HIV/AIDS. *Nature Reviews Drug Discovery* **2005,** *4*, 236.

77. Barre-Sinoussi, F.; Ross, A. L.; Delfraissy, J.-F., Past, present and future: 30 years of HIV research. *Nat Rev Micro* **2013,** *11*, 877.

78. Chan, D. C.; Kim, P. S., HIV entry and its inhibition. *Cell* **1998,** *93*, 681.

79. Wyatt, R.; Sodroski, J., The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* **1998,** *280*, 1884.

80. Hu, W.-S.; Temin, H. M., Retroviral recombination and reverse transcription. *Science* **1990,** *250*, 1227.

81. Hallenberger, S.; Bosch, V.; Angliker, H.; Shaw, E.; Klenk, H.-D.; Garten, W., Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gpl60. *Nature* **1992,** *360*, 358..

82. Jaskolski, M.; Alexandratos, J. N.; Bujacz, G.; Wlodawer, A., Piecing together the structure of retroviral integrase, an important target in AIDS therapy. *Febs Journal* **2009,** *276*, 2926.

83. Chiu, T. K.; Davies, D. R., Structure and function of HIV-1 integrase. *Current topics in medicinal chemistry* **2004,** *4*, 965.

84. Faure, A.; Calmels, C.; Desjobert, C.; Castroviejo, M.; Caumont-Sarcos, A.; Tarrago-Litvak, L.; Litvak, S.; Parissi, V., HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic acids research* **2005,** *33*, 977.

85. Van Maele, B.; Debyser, Z., HIV-1 integration: an interplay between HIV-1 integrase, cellular and viral proteins. *AIDS rev* **2005,** *7*, 26.

86. Van Maele, B.; Busschots, K.; Vandekerckhove, L.; Christ, F.; Debyser, Z., Cellular co-factors of HIV-1 integration. *Trends in biochemical sciences* **2006,** *31*, 98.

87. Ferris, A. L.; Wu, X.; Hughes, C. M.; Stewart, C.; Smith, S. J.; Milne, T. A.; Wang, G. G.; Shun, M.-C.; Allis, C. D.; Engelman, A., Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. *Proceedings of the National Academy of Sciences* **2010,** *107*, 3135.

88. Engelman, A.; Cherepanov, P., The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. *PLoS Pathog* **2008,** *4*, e1000046.

89. Craigie, R., Targeting HIV-1 DNA integration by swapping tethers. *Proceedings of the National Academy of Sciences* **2010,** *107*, 2735.

90. Kobayashi, M.; Yoshinaga, T.; Seki, T.; Wakasa-Morimoto, C.; Brown, K. W.; Ferris, R.; Foster, S. A.; Hazen, R. J.; Miki, S.; Suyama-Kagitani, A., In vitro antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. *Antimicrobial agents and chemotherapy* **2011,** *55*, 813.

91. Fransen, S.; Gupta, S.; Danovich, R.; Hazuda, D.; Miller, M.; Witmer, M.; Petropoulos, C. J.; Huang, W., Loss of raltegravir susceptibility by human immunodeficiency virus type 1 is conferred via multiple nonoverlapping genetic pathways. *Journal of virology* **2009,** *83*, 11440.

92. Jones, G. S.; Yu, F.; Zeynalzadegan, A.; Hesselgesser, J.; Chen, X.; Chen, J.; Jin, H.; Kim, C.

U.; Wright, M.; Geleziunas, R., Preclinical evaluation of GS-9160, a novel inhibitor of human immunodeficiency virus type 1 integrase. *Antimicrobial agents and chemotherapy* **2009,** *53*, 1194.

93. Da Silva, D.; Van Wesenbeeck, L.; Breilh, D.; Reigadas, S.; Anies, G.; Van Baelen, K.; Morlat, P.; Neau, D.; Dupon, M.; Wittkop, L., HIV-1 resistance patterns to integrase inhibitors in antiretroviral-experienced patients with virological failure on raltegravir-containing regimens. *Journal of Antimicrobial Chemotherapy* **2010**, *65*, 1262.

94. Hare, S.; Smith, S. J.; Métifiot, M.; Jaxa-Chamiec, A.; Pommier, Y.; Hughes, S. H.; Cherepanov, P., Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572). *Molecular pharmacology* **2011,** *80*, 565.

95. Hare, S.; Gupta, S. S.; Valkov, E.; Engelman, A.; Cherepanov, P., Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **2010,** *464*, 232.

96. Valkov, E.; Gupta, S. S.; Hare, S.; Helander, A.; Roversi, P.; McClure, M.; Cherepanov, P., Functional and structural characterization of the integrase from the prototype foamy virus. *Nucleic acids research* **2009,** *37*, 243.

97. Chen, Q.; Buolamwini, J. K.; Smith, J. C.; Li, A.; Xu, Q.; Cheng, X.; Wei, D., Impact of resistance mutations on inhibitor binding to HIV-1 integrase. *Journal of chemical information and modeling* **2013,** *53*, 3297.

98. Bacchi, A.; Carcelli, M.; Compari, C.; Fisicaro, E.; Pala, N.; Rispoli, G.; Rogolino, D.; Sanchez, T. W.; Sechi, M.; Sinisi, V., Investigating the role of metal chelation in HIV-1 integrase strand transfer inhibitors. *Journal of medicinal chemistry* **2011,** *54*, 8407.

99. Xue, W.; Jin, X.; Ning, L.; Wang, M.; Liu, H.; Yao, X., Exploring the molecular mechanism of cross-resistance to HIV-1 integrase strand transfer inhibitors by molecular dynamics simulation and

residue interaction network analysis. *Journal of chemical information and modeling* **2012,** *53*, 210.

100. Moss, D. M.; Siccardi, M.; Murphy, M.; Piperakis, M. M.; Khoo, S. H.; Back, D. J.; Owen, A., Divalent metals and pH alter raltegravir disposition in vitro. *Antimicrobial agents and chemotherapy* **2012,** *56*, 3020.

101. Johns, B. A.; Kawasuji, T.; Weatherhead, J. G.; Taishi, T.; Temelkoff, D. P.; Yoshida, H.; Akiyama, T.; Taoda, Y.; Murai, H.; Kiyama, R., Carbamoyl pyridone HIV-1 integrase inhibitors 3. A diastereomeric approach to chiral nonracemic tricyclic ring systems and the discovery of dolutegravir (S/GSK1349572) and (S/GSK1265744). *Journal of medicinal chemistry* **2013,** *56*, 5901.

102. Šali, A.; Blundell, T. L., Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology* **1993,** *234*, 779.

103. Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09, Revision D.01. *Gaussian, Inc.*,

Wallingford CT **2009**.

104. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of computational chemistry* **2004,** *25*, 1157.

105. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983,** *79*, 926.

106. Pastor, R. W.; Brooks, B. R.; Szabo, A., An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Molecular Physics* **1988,** *65*, 1409.

107. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics* **1993,** *98*, 10089.

108. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977,** *23*, 327.

109. Schlitter, J., Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chemical Physics Letters* **1993,** *215*, 617.

110. Karplus, M.; Kushick, J. N., Method for estimating the configurational entropy of macromolecules. *Macromolecules* **1981,** *14*, 325.

111. Andricioaei, I.; Karplus, M., On the calculation of entropy from covariance matrices of the atomic fluctuations. *The Journal of Chemical Physics* **2001,** *115*, 6289.

112. Kabsch, W., A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **1976,** *32*, 922.

113. Bao, J.; Dong, X. Y.; Zhang, J. Z.; Arora, P. S., Dynamical binding of hydrogen-bond surrogate derived bak helices to antiapoptotic protein bcl-xl. *The Journal of Physical Chemistry B* **2009,** *113*, 3565.

114. Case, D. A.; Darden, T.; Cheatham III, T. E.; Simmerling, C.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R.; Zhang, W., AMBER 10 users' manual. *University of California* **2008**.

115. Gallo, R. C.; Sarin, P. S.; Gelmann, E.; Robert-Guroff, M.; Richardson, E.; Kalyanaraman, V.; Mann, D.; Sidhu, G. D.; Stahl, R. E.; Zolla-Pazner, S., Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* **1983,** *220*, 865.

116. Lv, Z.; Chu, Y.; Wang, Y., HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV/AIDS – Research and Palliative Care* **2015,** *7*, 95.

117. Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R.; Scolnick, E. M.; Sigal, I. S., Active human immunodeficiency virus protease is required for viral infectivity. *Proceedings of the National Academy of Sciences* **1988,** *85*, 4686.

118. 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., Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. *Journal of virology* **2003,** *77*, 4836.

119. Gulnik, S. V.; Suvorov, L. I.; Liu, B.; Yu, B.; Anderson, B.; Mitsuya, H.; Erickson, J. W., Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry* **1995,** *34*, 9282.

120. Weber, I. T.; Agniswamy, J., HIV-1 protease: structural perspectives on drug resistance. *Viruses* **2009,** *1*, 1110.

121. Ali, A.; Bandaranayake, R. M.; Cai, Y.; King, N. M.; Kolli, M.; Mittal, S.; Murzycki, J. F.; Nalam, M. N.; Nalivaika, E. A.; Özen, A., Molecular basis for drug resistance in HIV-1 protease. *Viruses* **2010,** *2*, 2509.

122. Kar, P.; Knecht, V., Energetic basis for drug resistance of HIV-1 protease mutants against amprenavir. *Journal of computer-aided molecular design* **2012,** *26*, 215.

123. Meher, B. R.; Wang, Y., Interaction of I50V mutant and I50L/A71V double mutant HIV-protease with inhibitor TMC114 (darunavir): molecular dynamics simulation and binding free energy studies. *The Journal of Physical Chemistry B* **2012,** *116*, 1884.

124. Chen, J.; Zhang, S.; Liu, X.; Zhang, Q., Insights into drug resistance of mutations D30N and I50V to HIV-1 protease inhibitor TMC-114: Free energy calculation and molecular dynamic simulation. *Journal of molecular modeling* **2010,** *16*, 459.

125. Duan, L. L.; Mei, Y.; Zhang, Q. G.; Zhang, J. Z., Intra-protein hydrogen bonding is dynamically stabilized by electronic polarization. *The Journal of chemical physics* **2009,** *130*, 115102.

126. Tong, Y.; Mei, Y.; Li, Y. L.; Ji, C. G.; Zhang, J. Z., Electrostatic polarization makes a substantial contribution to the free energy of avidin-biotin binding. *Journal of the American Chemical Society* **2010,** *132*, 5137.

127. Mei, Y.; Wei, C.; Yip, Y. M.; Ho, C. Y.; Zhang, J. Z.; Zhang, D., Folding and thermodynamic studies of Trp-cage based on polarized force field. *Theoretical Chemistry Accounts* **2012,** *131*, 1.

128. Liu, F.; Boross, P. I.; Wang, Y.-F.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T., Kinetic, stability, and structural changes in high-resolution crystal structures of HIV-1 protease with drug-resistant mutations L24I, I50V, and G73S. *Journal of molecular biology* **2005,** *354*, 789.

129. Mei, Y.; Zhang, D. W.; Zhang, J., New method for direct linear-scaling calculation of electron density of proteins. *The Journal of Physical Chemistry A* **2005,** *109*, 2.

130. Stouch, T.; Williams, D. E., Conformational dependence of electrostatic potential derived charges of a lipid headgroup: glycerylphosphorylcholine. *Journal of computational chemistry* **1992,** *13*, 622.

131. Stouch, T. R.; Williams, D. E., Conformational dependence of electrostatic potential-derived charges: Studies of the fitting procedure. *Journal of computational chemistry* **1993,** *14*, 858.

132. Zeng, J.; Duan, L.; Zhang, J. Z.; Mei, Y., A numerically stable restrained electrostatic potential charge fitting method. *Journal of computational chemistry* **2013,** *34*, 847.

133. Shen, C. H.; Wang, Y. F.; Kovalevsky, A. Y.; Harrison, R. W.; Weber, I. T., Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters. *Febs Journal* **2010,** *277*, 3699.

134. Zoete, V.; Michielin, O.; Karplus, M., Protein-ligand binding free energy estimation using molecular mechanics and continuum electrostatics. Application to HIV-1 protease inhibitors. *Journal of computer-aided molecular design* **2003,** *17*, 861.

135. Lee, M. C.; Duan, Y., Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. *Proteins: Structure, Function, and Bioinformatics* **2004,** *55*, 620.

136. Nguyen, D. T.; Case, D. A., On finding stationary states on large-molecule potential energy surfaces. *The Journal of Physical Chemistry* **1985,** *89*, 4020.

137. Durrant, J. D.; de Oliveira, C. A. F.; McCammon, J. A., POVME: an algorithm for measuring binding-pocket volumes. *Journal of Molecular Graphics and Modelling* **2011,** *29*, 773.

138. Leonis, G.; Steinbrecher, T.; Papadopoulos, M. G., A contribution to the drug resistance mechanism of Darunavir, Amprenavir, Indinavir, and Saquinavir complexes with HIV-1 protease due to flap mutation I50V: A systematic MM-PBSA and thermodynamic integration study. *Journal of chemical information and modeling* **2013,** *53*, 2141.

139. Noskov, S. Y.; Lamoureux, G.; Roux, B., Molecular dynamics study of hydration in ethanol-water mixtures using a polarizable force field. *The Journal of Physical Chemistry B* **2005,** *109*, 6705.

140. Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell Jr, A. D.; Roux, B., Simulating monovalent and divalent ions in aqueous solution using a Drude polarizable force field. *Journal of chemical theory and computation* **2010,** *6*, 774.

141. Li, Y. L.; Mei, Y.; Zhang, D. W.; Xie, D. Q.; Zhang, J. Z., Structure and dynamics of a dizinc metalloprotein: Effect of charge transfer and polarization. *The Journal of Physical Chemistry B* **2011,** *115*, 10154.

142. Snell, E. E.; Eakin, R. E.; Williams, R. J., A quantitative test for biotin and observations regarding its occurrence and properties. *Journal of the American Chemical Society* **1940,** *62*, 175.

143. Kuhn, B.; Kollman, P. A., Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *Journal of medicinal chemistry* **2000,** *43*, 3786.

144. Kuhn, B.; Kollman, P. A., A ligand that is predicted to bind better to avidin than biotin: insights from computational fluorine scanning. *Journal of the American Chemical Society* **2000,** *122*, 3909.

145. Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M., Three-dimensional structure of the tetragonal crystal form of egg-white avidin in its functional complex with biotin at 2· 7 Å resolution.

Journal of molecular biology **1993,** *231*, 698.

146. Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U., Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field. *Journal of medicinal chemistry* **2006,** *49*, 6596.

147.Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Montgomery Jr, J.; Vreven, T.; Kudin, K.; Burant, J., Gaussian 03, revision D. 01. *Gaussian Inc., Wallingford, CT* **2004,** *26*.