CHAPTER 1

INTRODUCTION

1.1 Research Rationale

Acquired immune deficiency syndrome (AIDS) caused by human immunodeficiency virus (HIV) infection has developed into a major epidemic in the world. Currently, about 32 million people died have been estimated from AIDS and the new infection is proliferating at an alarming rate[1]. The knowledge of identification for the HIV retrovirus accumulated about the role of different elements in HIV life cycle led researchers around the world to develop inhibitors that target different step in the life cycle of the virus. One promising target is HIV-1 protease (HIV-1 PR) playing the crucial role in virus maturation. The elongate treatment of available drugs, however, has led to the emergence of resistant virus variants. Therefore, the ability to predict the drug resistance of HIV-1 PR mutants may be helpful in developing more effective and longer lasting treatment regimens. The other critical issues is the identification of the protonation state for catalytic aspartate residues (Asp25/Asp25') of the HIV-1 PR enzyme complexed with the inhibitors because incorrect protonation state leads to dramatic changes in the predicted binding mode of the inhibitor.

The computational studies are greatly valuable and the challenging approach to investigate both of dynamical and structural properties at the molecular level. Thus, the molecular dynamics simulations have been chosen to apply for structural and energetic analysis for this work.

1.2 Acquired immunodeficiency syndrome (AIDS)

AIDS is a major worldwide epidemic spread primarily through contact with infected blood during sexual activity, drug injection, birth, and, rarely now, blood transfusion. In 1981, this syndrome was observed in USA[2] and became known. AIDS is caused by two variants of the human immunodeficiency virus, HIV-1 and HIV-2. The genomes of these retroviruses compose with three open reading frames (ORF), gag, pol, and env and encode only three enzymes, reverse transcriptase (RT), integrase, and protease (PR).

A distinct feature of HIV is that the virus infects the helper T-lymphocytes, which strive a central role in the regulating of the immune response[3]. The immune system of AIDS patients will be weakened because of the depletion of helper T-lymphocytes. Thus, the gradual destruction of these cells causes the patient increasingly susceptible to opportunistic infections of bacteria, viral or fungal origin and to certain cancers, which are the key features of the final stage of the HIV infection or AIDS[4].

1.3 Human Immunodeficiency Virus (HIV)

Human Immunodeficiency Virus (HIV) belongs to the class of viruses called retroviruses carrying information in the form of RNA. The retroviruses consists of duplicate copies of positive single-stranded RNA. The name retrovirus is derived from the unique event, which is completely opposite to the normal process where RNA is transcribed from DNA.

According to estimates from the UNAIDS/WHO[5] AIDS epidemic update (November, 2006), around 37.2 million adults and 2.3 million children were living

with HIV at the end of 2006. During 2006, some 4.3 million people became infected with the human immunodeficiency virus (HIV), which causes AIDS (see in Table 1.1). The number of people living with HIV has risen from around 8 million in 1990 to nearly 40 million today (Figure 1.1), and is still growing. Around 63% of people living with HIV are in sub-Saharan Africa.

Table 1.1 Global HIV/AIDS estimates, end of 2006[6].

	Estimate (million)
People living with HIV/AIDS in 2006	39.5
Adults living with HIV/AIDS in 2006	37.2
Women living with HIV/AIDS in 2006	17.7
Children living with HIV/AIDS in 2006	2.3
people newly infected with HIV in 2006	4.3
Adults newly infected with HIV in 2006	3.8
Children newly infected with HIV in 2006	0.53
AIDS deaths in 2006	2.9
Adults AIDS deaths in 2006	2.6
Child AIDS deaths in 2006	0.38

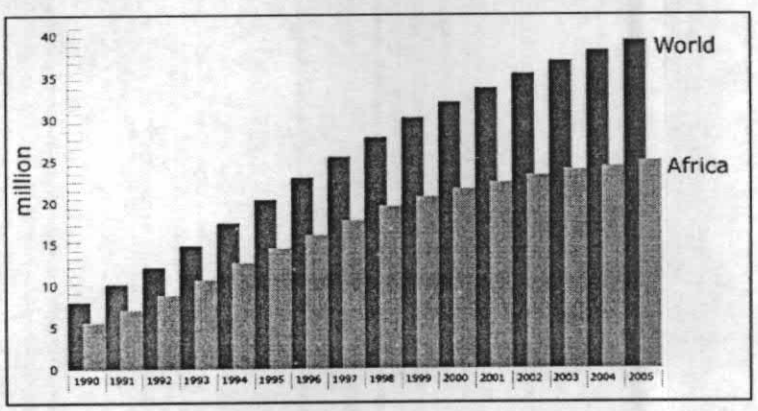

Figure 1.1 HIV/AIDS Global trends[6].

1.3.1 Structure of HIV virus

The viral envelope of HIV is composed of two layers of fatty molecules that are take from an infected cell when the new virus buds off. Underneath the fatty molecules is a layer of matrix protein called p17. The HIV virion (Figure 1.2) contains two important envelope proteins. Gp160 is a glycoprotein that is cleaved into gp120 and gp41, which are expressed as trimers to make up the viral envelope[7]. The gp120 moiety binds the virion and host via a CD4 receptor, whereas, gp41 is believed to induce fusion of the virus to the host cell membrane as well as anchor gp120[8]. On average 72 copies of the protein ENV, containing three to four gp120 and three to four gp41 molecules, exist on the virus surface[9]. The viral core of HIV is contained by 2000 copies of the viral protein, p24[10]. This capsid holds 2 single strands of HIV RNA that encode nine genes for the virus (see in Table 1.2).

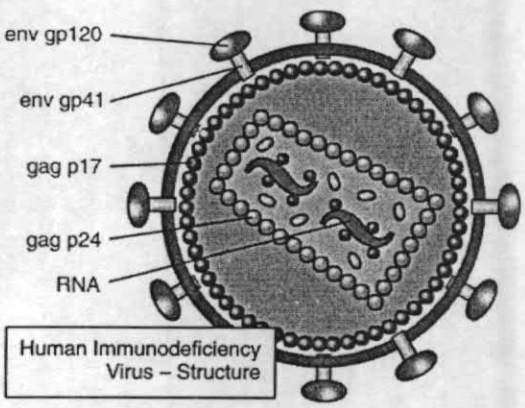

Figure 1.2 Structure of the HIV virus[11].

Table 1.2 HIV-gene function[7].

Gene	Gene Product and Function
gag (group-specific antigen)	Core proteins and matrix proteins
pol (polymerase)	reverse transcriptase, protease, and integrase enzymes
env (envelope)	transmembrane glycoproteins: gp120 binds to CD4 and CCR5; gp41 is required for virus internalization
tat (transactivator)	positive regulator of transcription
rev (regulator of viral expression)	allows export of unspliced transcripts from nucleus
vif (viral infectivity)	affects particle infectivity
vpr (viral protein R)	transport of DNA to nucleus, augments virion production, and controls cell cycle arrest
vpu (viral protein U)	unique to HIV-1; downregulates CD4
nef (negative-regualtion factor)	augments viral replication *in vivo* and *in vitro*; downregulates CD4

1.3.2 Replication cycle of HIV

A new virus will attack a cell, and use it to make new viruses. Just as your body is constantly making new skin cells, or new blood cells, each cell often makes new proteins in order to stay alive and to reproduce itself. Viruses hide their own DNA in the DNA of the cell, and then, when the cell tries to make new proteins, it accidentally makes new viruses as well. HIV mostly infects cells in the immune system. HIV life cycle (Figure 1.3) shows the following number of each step.

1.3.2.1 Binding and fusion

HIV begins its life cycle when it binds to a CD4 receptor and one of two co-receptor on the surface of a $CD4^+$ T-lymphocyte. The virus then fuses with the host cell. After fusion, the virus releases RNA, its genetic material, into the host cell.

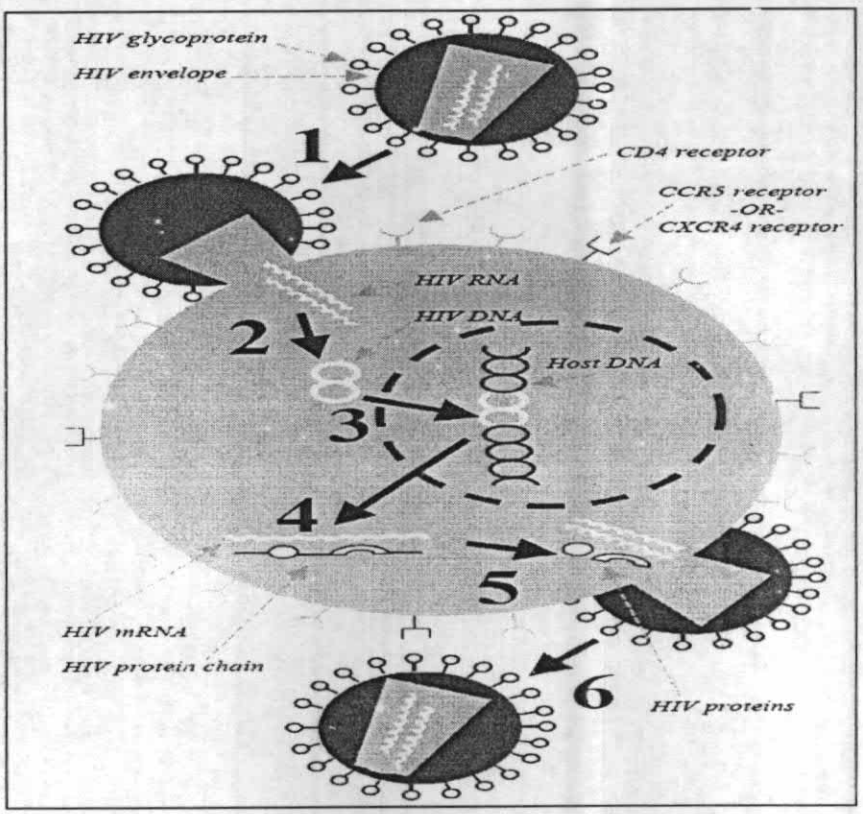

Figure 1.3 The HIV life cycle[12].

1.3.2.2 Reverse transcriptase

An HIV enzyme called reverse transcriptase converts the single-stranded HIV RNA to double-stranded HIV DNA.

1.3.2.3 Integration

The newly formed HIV DNA enters the host cell's nucleus, where an HIV enzyme called integrase "hides" the HIV DNA within the host cell's own DNA. The integrated HIV DNA is called provirus. The provirus may remain inactive for several years, producing few or no new copies of HIV.

1.3.2.4 Transcription

When the host cell receives a signal to become active, the provirus uses a host enzyme called RNA polymerase to create copies of the HIV genomic material, as well as shorter strands of RNA called messenger RNA (mRNA).The mRNA is used as a blueprint it make long chains of HIV proteins.

1.3.2.5 Assembly

An HIV enzyme called protease cuts the long chains of HIV proteins into smaller individual proteins. As the smaller HIV proteins come together with copies of HIV's RNA genetic material, anew virus particle is assembled.

1.3.2.6 Budding

The newly assembled virus pushes put ("buds") from the host cell. During budding, the new virus steals part of the cell's outer envelope. This envelope, which acts as a covering, is studded with protein/sugar combinations called HIV glycoproteins. These HIV glycoproteins are necessary for the virus to bind CD4 and co-receptors. The new copies of HIV can now move on to infect other cells.

1.3.3 Targets for Anti-HIV Chemotherapy

In principle, every step in the HIV replication cycle can be considered as a potential target for anti-viral chemotherapy. However, the number of practical targets for drug interventions is decreased due to the fact that the virus is an intracellular parasite, which relies on the metabolic pathways of the host cells. Hence, most agents blocking the replication of the virus are also lethal to the host cell. The key in selective anti-viral therapy is therefore to identify any process that is essential for the replication of the virus, but not for the survival of the cell[13]. The gained knowledge about the replicative cycle of the HIV-virus has led to extraction of virus-specific processes. Predominantly, scientists have focused their attentions on the following HIV life cycle process. So far, the two strategies reverse transcriptase and protease have been proven to the most successful in the search for drugs that can be used for treatment of AIDS[14]. Protease target will be studied in this work.

1.4 HIV -1 protease

The HIV-1 PR is essential for cleavage of the precursor polyproteins and it plays an important role in the release of mature forms of viral and enzymatic proteins.

1.4.1 HIV-1 PR structure

HIV-1 PR (Figure 1.4) is a homodimer consisting of 99 amino acids for each chain. The flap, an extended β-sheet region or a glycine-rich loop, constitutes in the part of substrate-binding site and plays an important role in the

substrate binding. The active site triad (Asp25-Thr26-Gly27) is located on the bottom of the cavity.

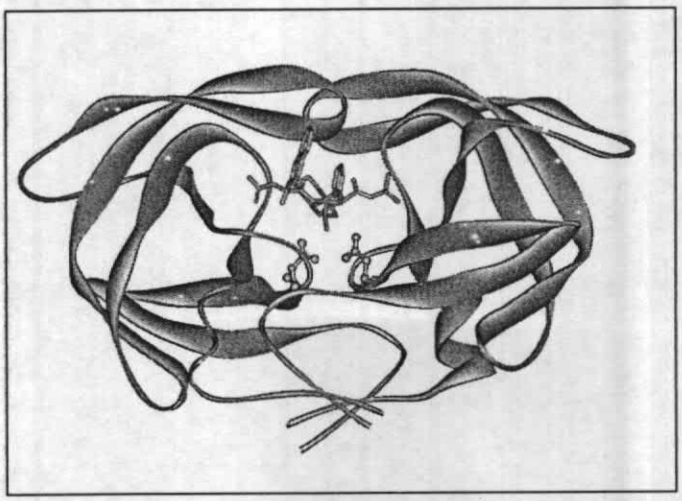

Figure 1.4 HIV-1 PR structure bound to the inhibitor/substrate (green stick). The green and the yellow ribbon represent chain A and chain B, respectively. The orange and purple ball and stick show the Asp25 of chain A and B, respectively.

1.4.2 Mechanism of the HIV-1 Protease

Proteases have been known to play essential roles in many biological processes. Two different mechanisms are divided into broad classes of protease enzyme. The fist class of enzymes is an attack of the activated water molecule to amide bond carbonyl of the subsite's scissile bond. In the second class, a nucleophilic atom of an amino acid side chain is used to initiate amide hydrolysis shown in Figure 1.5.

According to several studies, HIV-1 PR has been shown to belong to the class of the aspartic proteases. Asp-Thr-Gly of HIV-1 PR is the conserved residues among the aspartic protease[15]. The previous studies show that substituting Asp25 with Asn[16, 17], Thr[18], or Ala[19], leads to a protein without any

proteolytic activity. In addition, most of those studies are consistent with a general acid-base mechanisms. Because two active site aspartate residues play an essential general acid-base role to activate water molecule that acts a nucleophile and attack the carbonyl carbon of the scissile bond. The most widely accepted mechanism for an aspartic protease has been described by Suguna, et al. (Figure 1.5)[20]. The pH-rate profile of this enzyme implies that only one of the two Asp25 is unprotonated in the active pH range[21]. In the proposed mechanism of the Asp25 group is closer to the nucleophilic water molecule was assigned the negative charge (Figure 1.5). However, the identification of Asp25 group remains the controversy for the HIV-1 PR bound to the substrate/inhibitor. The corresponding suggestions can be concluded in Table 1.3.

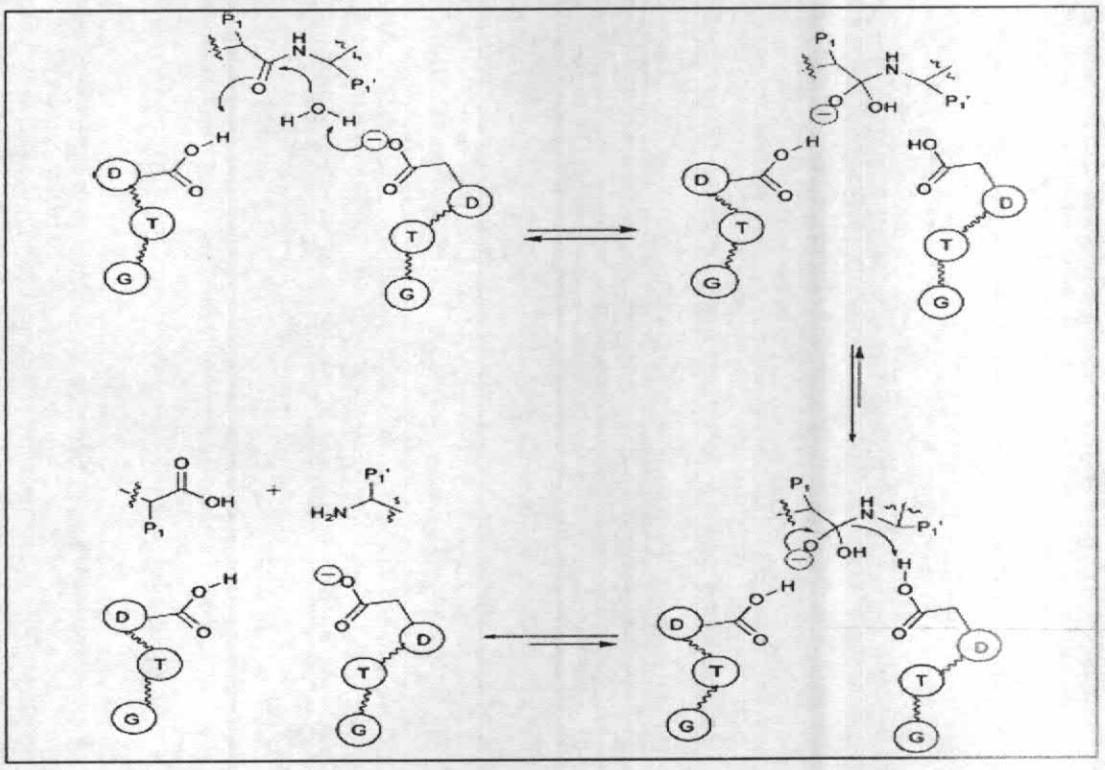

Figure 1.5 The proposed mechanism for aspartic protease[22].

Table 1.3 The prediction of protonation state of HIV-1 PR complexed with different inhibitors.

Inhibitor/Substrate	Method	Predicted state
U85548E (neutral charge)	Molecular dynamics	diprotonation[23]
MVT-101 (positive charge)	Molecular dynamics	unprotonation[23]
S- and R- isomers of U85548E (neutral charge)	Free energy pertubation	monoprotonation[24]
Pepstatin (neutral charge)	NMR	At least monoprotonation[25]

1.4.3 Substrate specificity

The gag and gag-pol polyproteins are the natural substrates for HIV-1 PR. Ten different sites of the polyproteins are cleavaged[26]. The substrate binds with the different amino acids side chains determined the specificity of the enzyme. Using standard nomenclature (Figure 1.6), the S1 and S1' (S2 and S2', etc.) subsites are structurally equivalent. The very hydrophobic regions are the two S1 subsites while, the S2 subsites are mostly hydrophobic except Asp29, Asp29', Asp30 and Asp30'. The S3 subsites are adjacent to S1 subsites and are also mostly hydrophobic.

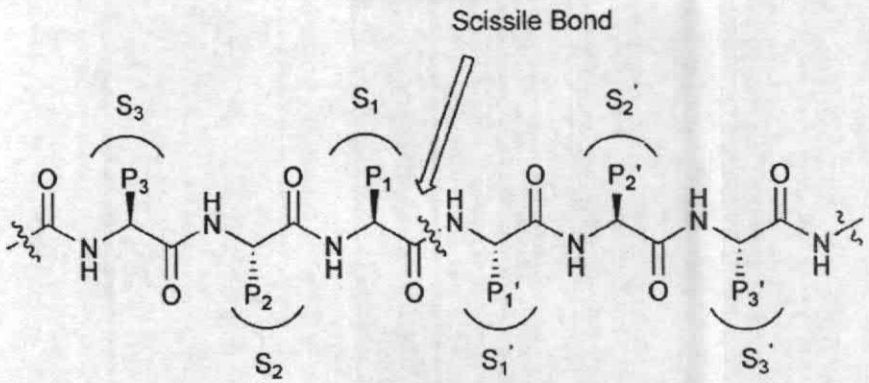

Figure 1.6 Amino acid residues of peptide substrate are assigned as $P_1 \cdots P_n$, $P'_1 \cdots P'_n$ corresponding with binding site on the protease refered as $S_1 \cdots S_n$, $S'_1 \cdots S'_n$ subsite[22].

1.5 HIV-1 protease inhibitors

There are ten FDA-approved protease inhibitors (PIs)[27] of United state of America, Amprenavir (APV), Indinavir (IDV), Lopinavir (LPV), Nelfinavir (NFV), Ritonavir (RTV), Saquinavir (SQV), Tipranavir (TPV), Fosamprenavir (FPV), Darunavir (DRV), and Atazanavir (ATV) (Figure 1.7).

1.5.1 Common structural features of the inhibitor

Binding of an inhibitor introduces substantial conformational changes to the enzymes. The overall movement of the subunits can be described as a rotation of up to about 2° around a hinge axis located in the subunit β-sheet interface. This motion, which slightly tightens the cavity of the active site, is also accompanied by a very large motion of the flap region as much as 7 Å for the tips of flaps[28]. However, the enzyme structure is well conserved among the different complexes, with rms

deviations between the Cα atoms seldom exceeding 0.6 Å. Such differences are well within the agreement range for protein structures refined independently or crystallized in different space groups[29].

Most of the inhibitors cocrystallized with HIV-1 PR, including all peptidomimetic inhibitors, are bound in the enzyme active site in an extended conformation so that when they are superimposed upon one another, their functional elements align quite well overall[30]. The contacts between the main chain of the peptidomimetic inhibitors and the protease are almost uniform for all the complexes (Figure 1.8). Following a similar pattern, the hydrogen bonds are made mostly between the main-chain atoms of both the enzyme and the inhibitor. The hydroxy group at the nonscissile junction, present in inhibitors other than those containing the reduced peptide bond isosteres, is positioned between the Asp25/Asp25' carboxyls of the protease, within hydrogen-bonding distance to least one carboxylate oxygen of each aspartate. A feature common to almost all complexes of HIV-1 PR is a buried water molecule that bridges the P2 and P1' CO groups of the inhibitor and Ile50 and Ile50 NH groups of the flaps. This water is approximately tetrahedrally coordinated and is completely seperated from the bulk solvent[31]. The functional substitution of this water has led to design of urea-based inhibitors (Figure 1.8).

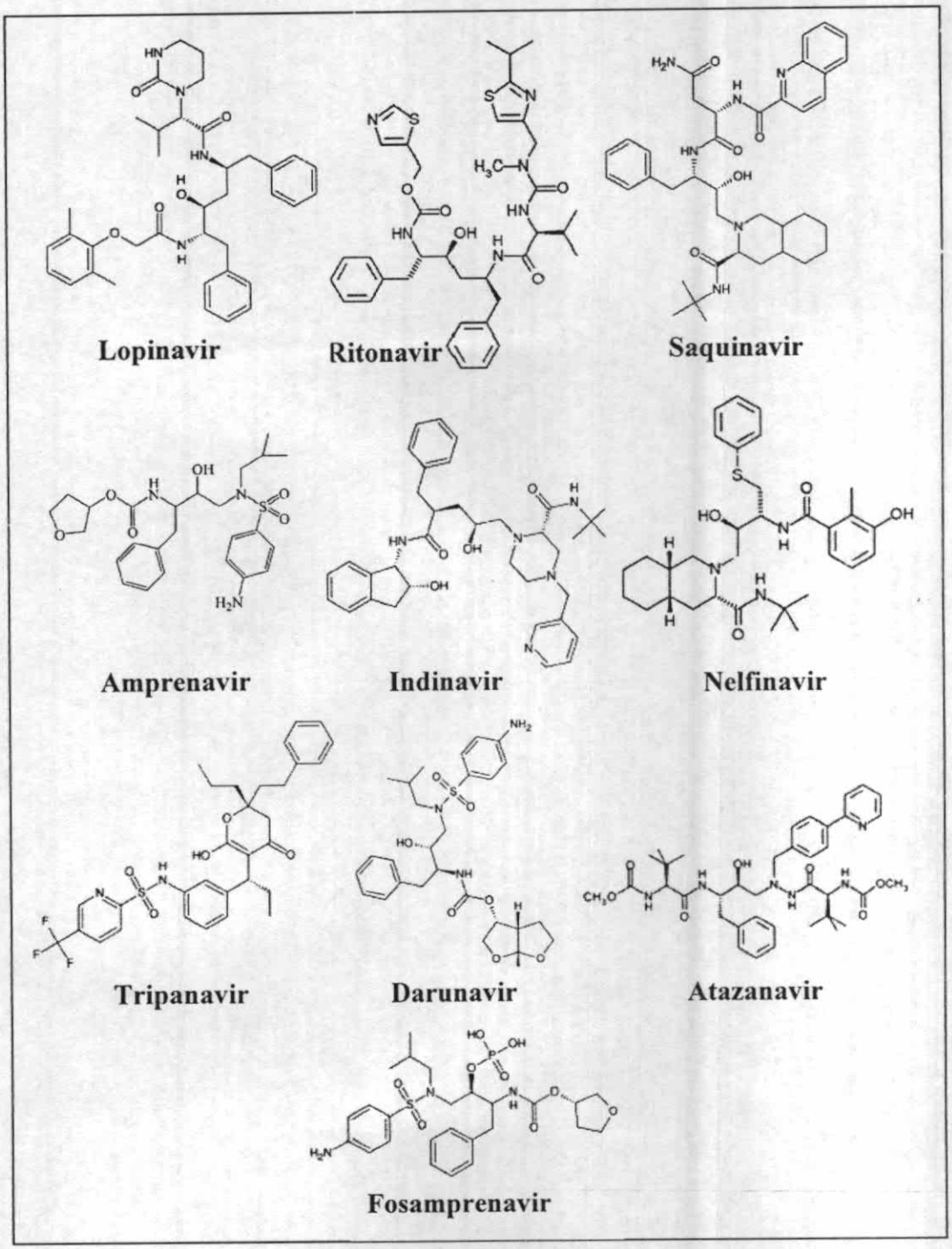

Figure 1.7 FDA approved drug PIs

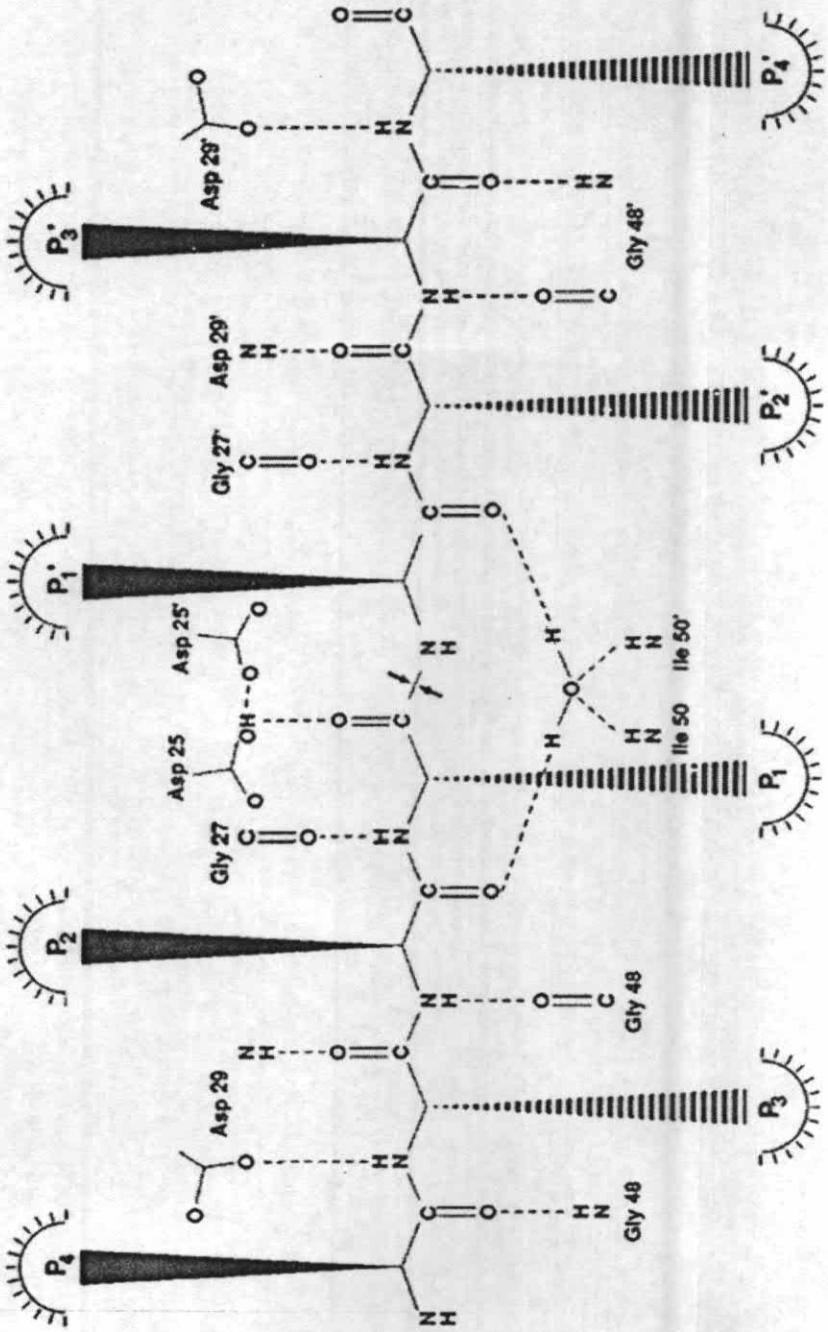

Figure 1.8 The nomenclature of subsites is that of Scechter and Berger modeled [26] reprinted from Wlodawer, A., *et al*[30].

1.6 HIV drug resistance

In the past two decades, researchers and clinicians have made vast efforts to identify drug resistance mutations in HIV-1 protease, a major target of anti-retroviral therapy. Discovery of a new drug resistance mutation typically needs a combination of clinical research, molecular biology, and biochemistry. Over the last 20 years, a lot of mutations and combinations of mutations in the protease have been reported to play role in drug resistance[32, 33]. Mutation in HIV-1 PR directly and substantially decrease druge efficacy (primary mutation) or favor resistance development caused by primary mutations, even indirectly affecting drug efficacy (secondary mutations)[34, 35]. Using combinations of drugs sufficiently can achieve the maximum reduction of viral replication. The combinations of antiretroviral drugs are now used for the treatment for AIDS called highly active antiretroviral therapy (HAART). Although, the use of combination is instrumental in controlling the development of resistance, some drug combinations have been shown to be antagonistic.

1.6.1 Development of resistance

The understanding of development for drug resistance has been described corresponding two concepts. First, HIV infection is characterized by high levels of virus production. In most untreated patients, the total number of productively infected cells in the lymphoid tissue has been estimated to be approximately 10^7 to 10^8 cells[36]. This number is relatively stable, reflecting the balance between the infection of new target cells and their clearance during the chronic phase of HIV infection. Because the half-life of infected cells is remarkably short (one or two days),

the maintenance of this steady state requires that HIV infect new target cells at a very high rate[37]. Second, the viral population in an infected person is highly heterogeneous[38].

The rapidity of this process depends on the level of the selective advantage conferred by the mutation, the prevalence of the mutant within the virus population, and the level of drug at the site of HIV replication. In some cases, substitutions of single amino acids can produce high levels of resistance. Since minority viral quasispecies carring any single mutation are believed to exist even before treatment is started[39], the emergence of these highly resistant single mutants can occur in a matter of weeks. In these cases, high levels of resistance or complete resistance requires the gradual accumulation of additional mutations. Combination therapy can block this selection process for two reasons. First multiple mechanisms (each requiring different mutations) are required for resistance to occur to all drugs in the regimen. Second, multiple drugs suppress viral replication more effectively than single agents[40].

1.6.2 Mechanism of resistance

A variety of mechanisms have been identified that differ both for different classes of drugs. The drug resistance of HIV-1 PR was focused for this study. Therefore, the observation of mutated residues and mechanism of HIV-1 PR mutation for protease inhibitors (PI) have been described in Table 1.4 and Table 1.5, respectively. The locations of drug-resistance mutations in protease are shown in Figure 1.9. In particular, more detailed about HIV-1 PR mutation are described below (section 1.6.3).

Table 1.4 Mutation involved in resistance of HIV to protease inhibitors.

Protease Mutations	Comments
L90M	Frequent resistance mutation, observed during failure of the therapy with the most protease inhibitors Mutation most frequently selected for by saquinavir therapy
V82A, V82T, V82F	Common resistance mutations can emerge early during failure of therapy with the most protease inhibitors Mutations most frequently selected for by ritonavir and indinavir therapy
D30N, N88D, N88S	Mutations most frequently selected for by nelfinavir therapy D30 always first
L10I, L10F, K20R, K20M, M36I, M46L, I54V, I54L, A71V, A71T, G73S, V77I, M93L	Mutation that can accumulate during failure of therapy with most protease inhibitors, causing gradual increases in the level of resistance
I84V	Frequently found after prolonged infective therapy with protease inhibitors Associated with high-level resistance to most protease inhibitors
G48V	Exclusively selected for by saquinavir therapy Associated with high-level resistance to saquinavir
L24I	Emerges occasionally during failure of indinavir therapy Also found with lopinavir therapy
I47V, I50V	Most often selected for by amprenavir therapy Also found with lopinavir therapy
V32I, F53L	Rare mutations Confer high high-level resistance to most protease inhibitors

Table 1.5 Antiretroviral agents used in the treatment of HIV infection have shown the mechanism of resistance.

Protease inhibitors	Mechanisms of action	Mechanisms of resistance
Saquinavir **Ritonavir** **Indinavir** **Nelfinavir** **Amprenavir** **Lopinavir**	Structure derived from natural peptidic substrates of the HIV type 1 protease. Bind the active site of the protease	Mutations reduce affinity of the inhibitors for the enzyme High-level resistance requires accumulation of mutations

1.6.3 Resistance to protease inhibitor

The HIV protease cleaves large polyprotein precursors at specific sites, releasing the structural proteins and enzymes necessary for the assembly of infectious viral particles. In the absence of a functional protease, viral particles are produced, but they are immature and are not infectious. The protease is a symmetrically assembled homodimer with a central, symmetric, substrate-binding cavity. Detailed knowledge of the structure of this domain and of the structure of the natural protein substrates of the enzyme has led to the design of specific inhibitors whose chemical structure mimics that of the viral peptides that are normally recognized and cleaved by the protease[41]. These compounds display a strong affinity for the active site of the HIV protease and inhibit the catalytic activity of the enzyme in a highly selective manner. Resistance to protease inhibitors is the consequence of amino acid substitutions that emerge either inside the substrate-binding domain of the enzyme or at distant sites[42] (Figure 1.9).

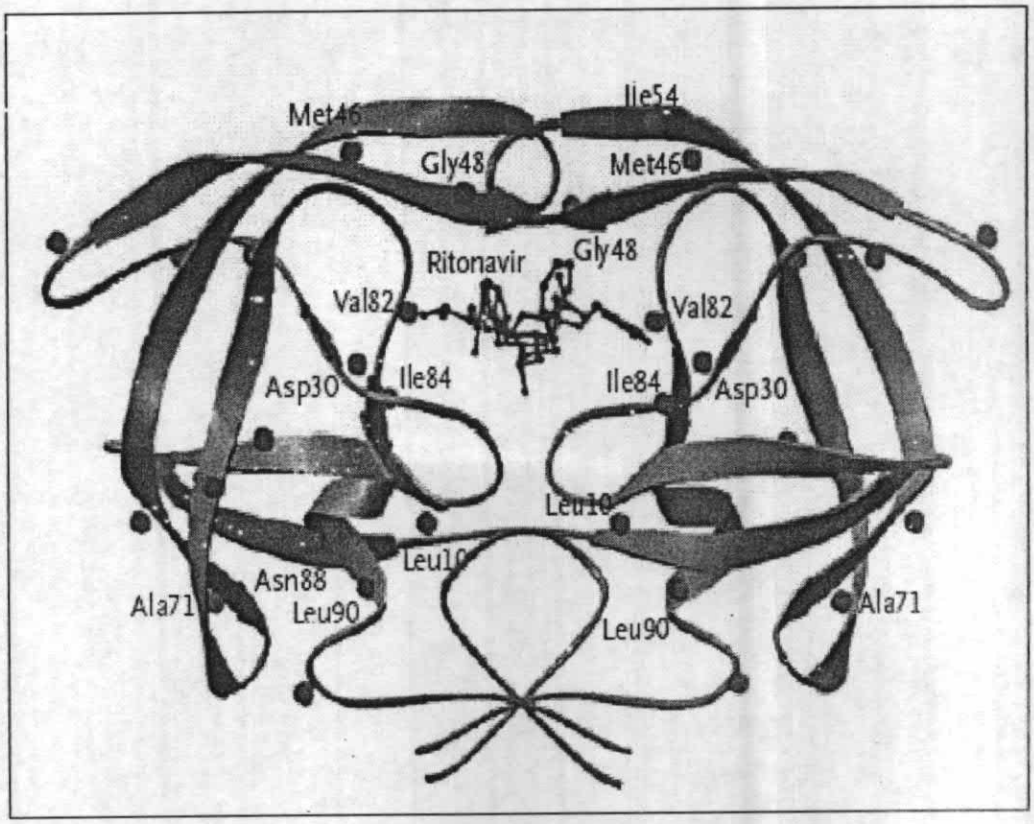

Figure 1.9 HIV-1 protease dimer binding with a protease inhibitor[43].

Directly or indirectly, these amino acid changes modify the number and the nature of the points of contact between the inhibitors and the protease, thereby reducing their affinity for the enzyme[44]. As an example, the common resistance mutation V82A reduces the size of an amino acid residue in the protease that is more important for binding most inhibitors than for binding the natural viral protein substrate (Figure 1.9). Protease inhibitors have been designed to bind the protease with maximal affinity and tend to occupy more space inside the active site cavity than do natural substrates. Unlike the inhibitors, the natural substrates of the protease have a variable, but generally less tight, interaction with the catalytic site, a phenomenon that promotes the ordered sequential cleavage of the polyproteins required for proper assembly of the viral particle. Resistance mutations in the protease, which result in an

overall enlargement of the catalytic site of enzyme, would thus be predicted to have greater effect on the binding of inhibitors than the natural templates.

Some mutations are selected for only by certain protease inhibitors (Table 1.4), reflecting particularities in the chemical structure of the inhibitors that influence their interaction with the substrate-binding domain of the enzyme. However, there is considerable overlap between the combinations of mutations in HIV strains that develop resistance to protease inhibitors. This overlap explains the wide cross-resistance that is generally observed within this drug class[45]. Remarkably, resistance to protease inhibitors can also be promoted by mutations in some of the natural viral substrates of the protease[46]. Characteristic substitutions of amino acids near cleavage sites of the gag polyprotein have been identified that can increase the level of resistance and the replicative capacity of the virus by faciliting cleavage under conditions in which the amount of active enzyme is suboptimal or improving the ability of proteases containing resistance mutations to interact with the substrate.

To furthermore information of mutation sites, the clinical data reports the mutation pattern for 7 FDA-approved drugs as shown in Figure 1.10. The frequent mutations are divided in order to protease substrate cleft (residue 82, 84, 48, 30, 50, 32, and 8), protease flap (residue 46, 47, 53, and 54), protease mutation at other conserved residues (residue 90, 73, 88, 24, and 33), and polymorphic sites contributing to resistance (residue 10, 20, 36, 71, 63, 77, and 93).

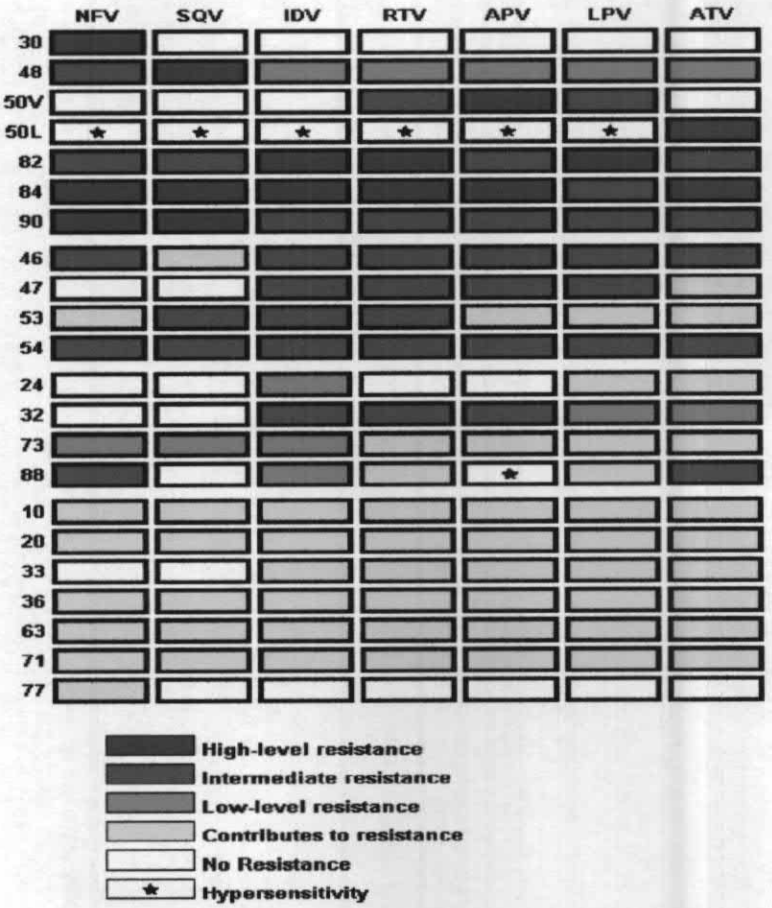

Figure 1.10 Protease inhibitors (PIs) and PI-resistance mutation[47].

1.6.4 Computational study: HIV-1 PR mutation

During the past decade, several laboratories have studied the source of HIV-1 PR due to PIs. For example, the different in conformational flexibility between wild type and mutants was studied to investigate how catalytic activity change[48, 49]. A classical molecular dynamics (MD) study on wild type and M46I variant free enzyme has indicated that the two systems differ in their flexibility and, in particular, that this mutation appears to stabilize the closed conformation of the flaps[50]. However, the M46I mutation appears to have little effect on the HIV-1 PR

affinity for most of the inhibitors, thus indicating that other factors, besides the binding affinity must play role in this kind of mutation. The mutant variants are found that these are able to evolve resistance by complementary strategies, several active-site (Figure 1.11, red) and nonactive-site (Figure 1.11, green) mutations reduce the affinity for both drugs and substrates. This may be achieved by disrupting favorable drug-protease interactions[51, 52].

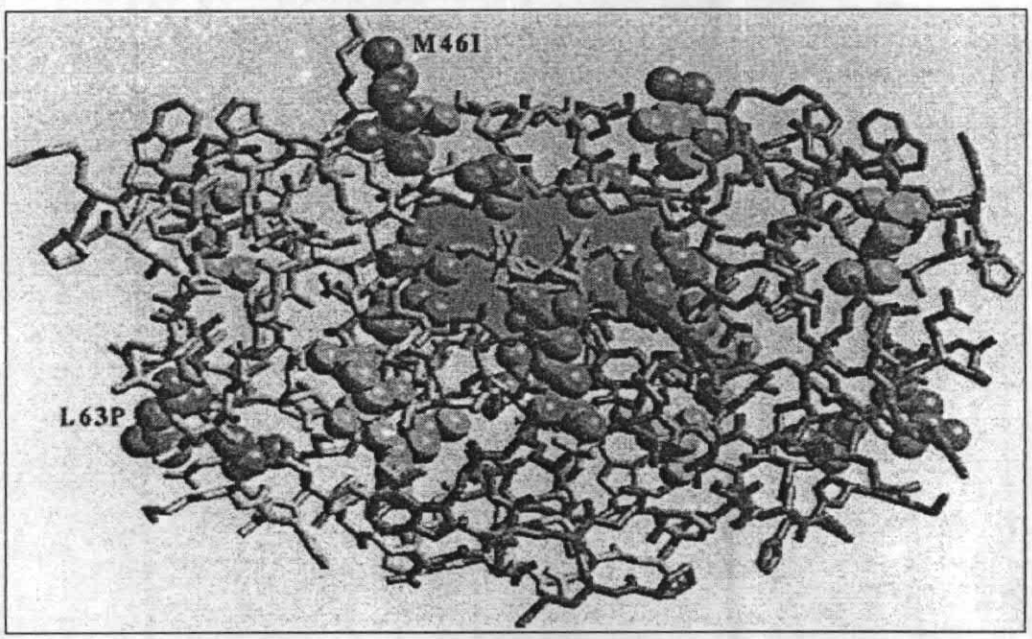

Figure 1.11 The mutation regions located at the active-site (red) and nonactive-site (green) of HIV-1 PR[53].

In contrast, compensatory mutations (Figure 1.11, blue) are able to enhance the protease activity that is decreased by the active-site mutations[49]. Based on structural features[54], lists of contacts between the inhibitor and the protease, are considered to predict mutated residues for two current protease inhibitors, IDV and SQV. The using of structural information correctly classified previously unseen mutants with an accuracy of between 60-70%. In addition, the drug resistance was also studied in of various mutations[55, 56] and attempts to classify and predict drug

resistance based on genotypic information. One of the most successful attempt was realized at UC Irvine[57].

1.7 Research objective

HIV-1 PR inhibitor research has become one of the major success stories of developing drugs using structure-based drug design, including a combination of X-ray crystallography and computational technologies. Molecular dynamics similar simulations and free energy studies have given valuable information regarding molecular basis of drug resistance and may provide the basis for a rational design of improved AIDS treatments

In this study, we used molecular dynamics simulations and free energy calculations a tool to specifically address an unknown protonation state in the HIV-1 PR binding site and analyze per-residue decomposition energy to correlate between the clinical data and drug resistance mutations. The former is one of the fundamental issues in understanding the binding mode of HIV-1 PR. The latter should give a recognition pattern of the inhibitors response to HIV-1 PR. This may provide an accurate prediction of drug resistance mutations. The ability to predict the drug resistance of HIV protease mutants may be useful in developing more effective and longer lasting treatment regimens.

The following steps have been established to achieve the goals of the study.

1. The explicit molecular dynamics simulations were performed to obtain structure, dynamics salvation data and to sample conformational space for free energy calculations.

2. A solvent continuum method so-called molecular mechanics Poisson Boltzmann and solvent accessible surface area (MMPB/SA) with standard and hybrid implicit/explicit solvent models such as MMPB/SA, MMGB/SA, MMGBMV/SA (GB = Generalized Born, GBMV = Generalized Born Molecular Volume) was used to estimate the binding free energies of those complexes.

3. Protonation state of the HIV-1 PR complexed with six FDA-approved drugs, including saquinavir (SQV), ritonavir (RTV), indinavir (IDV), amprenavie (APV), and nelfinavir (NFV), was examined using computed binding free energy and structural analysis.

4. The clinical mutation residues of HIV-1 PR complexed with 6 drugs was predicted using decomposition free energy.

5. Additional MD studies were carried out for the HIV-1 PR mutants complexed with SQV.