CHAPTER II
LITERATURE REVIEW ON MOLECULAR MODELING OF HIV-1 IN

2.1 Molecular Modeling

The process of designing a new drug and bringing it to market is very complex and take a long time. Conventional drug discovery methods are often based on guess, serendipity and much, sometimes inefficient, synthetic work. This leads to time consuming and high costs. It typically takes 12 years and 500 million dollars in average for new drugs to go from the research laboratory to patient use. This enormous pressure that the pharmaceutical companies are facing has created the need to apply all available techniques to decrease costs and the time of drug discovery to the market. Currently, one of the most widely applied techniques in drug discovery is computational chemistry and molecular modeling. This branch of science is centered on applying the fundamental laws of physics and chemistry to the studied molecules. In case of drug discovery, the molecules under study are those directly or indirectly involved in human disease. The ultimate aim is to create models and simulations which can help in the different stages of a discovery pipeline of molecules and their interactions with certain targets, thereby allowing a more rational approach to drug development. The greatest success of computational chemistry in drug design to date is the development of HIV-1 PR inhibitors that have been approved by the FDA and commercialized in the market [70]. Other examples of commercial drugs that were developed using computer assisted drug design include Captopril (anti hypertensive), Crixican (anti HIV) [71], Teveten (anti hypertensive) [72], Aricept (for Alzheimers disease), Trusopt (for Glaucoma) and Zomig (for migraine) [73].

2.2 Molecular Modeling of HIV-1 IN

As aforementioned, the rapid advances in computational technologies have become progressively more important in modern drug discovery. During the last few years, different molecular modeling approaches such as QSAR and MD simulations have been applied to study HIV-1 IN and its inhibitors. Some of very interesting works are described in the following part.

2.2.1 Two/Three Dimensional Quantitative Structure Activity Relationship (2/3D-QSAR) of HIV-1 IN inhibitors.

In the classical QSAR studied by Yuan and Parrill [60], two structural classes of compounds namely salicyhydrazines and tyrophostins were investigated. Five descriptors i.e. the square log of the octanol/water partition coefficient, the sum of the van der Waals surface, the area of van der Waals surface area of atoms whose partial charges are negative, Kier1 and KierA1 (the Kier Kappa Shape Indices), were chosen to build QSAR models for either salicylhydrazines or tyrphostins (see Figure 2.1). However, the two QSAR equations exhibited different signs and coefficients. Therefore, the authors proposed that salicylhydrazines and tyrophostins might bind to HIV-1 IN at different site. To confirm this hypothesis, a single QSAR model describing the activities of both classes of compounds was determined using the Genetic Function Approximation (GFA). By using the same descriptors as in the former individual single model, the QSAR model of salicylhydrazines with reasonable correlation (r^2 = 0.85) was obtained while that of tyrophostins yielded a poor correlation (r^2 < 0.6). This finding clearly supported the different binding modes or sites of these groups of inhibitors.

Makhija and Kulkarni [62] utilized 3D-QSAR using the Comparative Molecular Similarity Indices Analysis (CoMSIA) methodology to a set of 41 compounds belonging to five structurally diverse groups i.e. coumarins, quinines, hydrazides, dioxepinones and salicylpyrazolinones (Figure 2.1). The derived CoMSIA models showed good predictive ability as evidenced by a cross-validated r^2 of 0.76 - 0.8. The analyses of CoMSIA models demonstrated that the requirements of steric, electrostatic, hydrogen donor and hydrogen bond acceptor for the 3'-processing reaction were similar to those for the ST process while their hydrophobic requirement was different. The presence of hydrogen bond donor character such as OH and hydrogen bond acceptor such as NO_2 were required to enhance their HIV-1 IN inhibition.

Another conventional QSAR and 3D-QSAR employed to various structural classes of HIV-1 IN inhibitors were recently investigated by Yuan and Parrill [59, 61]. The authors found that the activities of these structurally diverse compounds could not be described by a single QSAR model. Therefore, a hierarchical cluster analysis was utilized to classify 11 different classes of HIV-1 IN inhibitors into two subsets and the QSAR

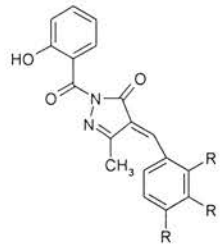
model was derived for each cluster separately. The first cluster contained coumarins, aromatic sulfonamides, chicoric acids, tetracyclines and tyrophostins while the second cluster included arylamides, thiazolothazepines, curcumins, salicylhydrazides, styrylquinolines, and lichen acids (see Figure 2.1). Consequently, the authors proposed that these two classes of inhibitors could interact with HIV-1 IN either at different binding regions or an overlapping area with different sets of amino acids. This work supported the previously reported QSAR study [60] of salicylhydrazines and tyrophostins classes which belong to different clusters.

Another classical QSAR was focused on several classes of HIV-1 IN inhibitors [63]. Up to 80 compounds were separated into catechols and non-catechol. Like the previous QSAR studies [59, 61], the authors classified thiazolothaizepines, salicylhydrazides and lichen acids (see Figure 2.1) in the same group whereas coumarins and curcumins were separated into different families. For catechol compounds, the electrostatic, shape related and thermodynamics descriptors were suggested to play a major role for inhibitory potency while spatial, structural and thermodynamics features were required for non-catechol compounds. Furthermore, researchers found that hydrogen bond donating groups such as OH, COOH and SO_3H were required for both catechol and non-catechol compounds.

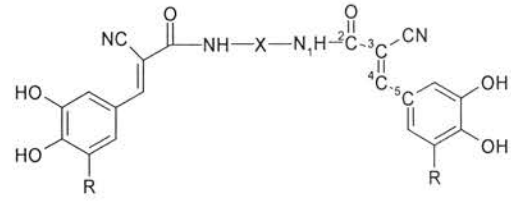
Moreover, Makhija and Kulkarni employed both CoMFA and CoMSIA strategies to 27 thiazolothiazepines HIV-1 IN inhibitors [64]. Their constructed 3D-QSAR models were evaluated by using two different test sets. Both models were able to predict the activities of additional thiazolothiazepines while they exhibited a poor predictive capability with the coumarin test set. This is probably because of the different apparent binding sites of thiazolothiazepines and coumarin compounds, as proposed by Yuan and Parrill [61].

From the previous 2D/3D-QSAR studies of HIV-1 IN, it is apparent that for the calculations including either one or a few classes of HIV-1 IN [62, 63, 74], a single 3D-QSAR model could be achieved. However, in a case of various classes of IN inhibitors [59, 61], a single QSAR model describing the activities of all structural diversity inhibitors couldnot be generated and all the compounds must be clustered to obtain acceptable models. This means that in order to predict the activities of any given

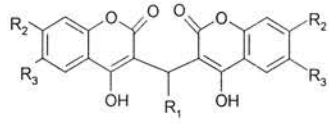
compounds, we have to find a suitable or the appropriate model because the prediction depends on structures of ligands. This is inconvenience in case there are many unknown compounds. Therefore in this thesis an attempt was made to create a single 3D-QSAR model for several classes of HIV-1 IN inhibitors.



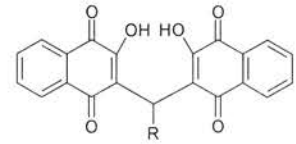
salicylhydrazine



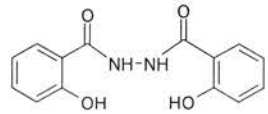
tyrophostin



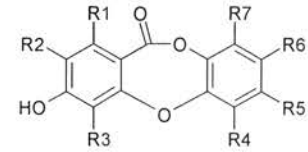
coumarin



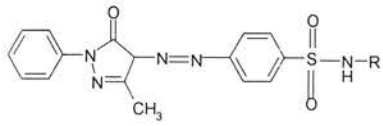
quinone



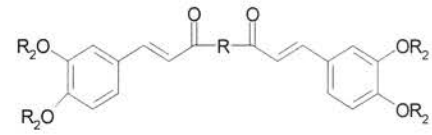
hydrazide



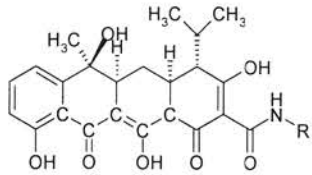
dioxepinones/lichen acid



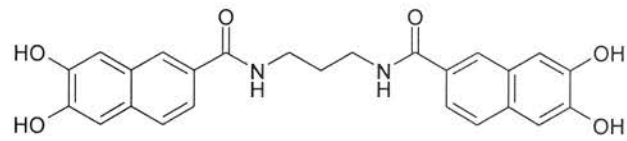
aromatic sulfonamide



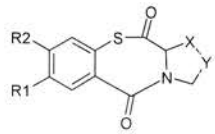
chicoric acid



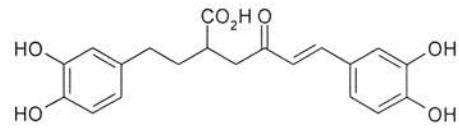
tetracycline



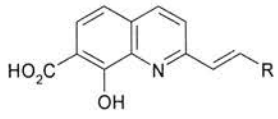
arylamide



thiazolothiazepine



curcumin



styrylquinoline

Figure 2.1 Chemical structures of compounds used in references [59-63].

2.2.2 Molecular Dynamics (MD) Simulation

The main goal of MD simulation (further discussed in CHARPTER III) is to investigate the dynamical behavior of the proteins and complexes in solution as its advantage is the inclusion of the explicit water molecules (or other solvents) in calculations.

The first MD simulation of HIV-1 IN was reported in 1999 by Lins et al [67]. They performed two MD simulations of the catalytic core domain of HIV-1 IN containing no metal and one metal ion in the active site. The main goal of that work was to investigate the dynamical behavior of HIV-1 IN with and without metal ion binding in the active site. The MD analyses revealed a high mobility of the loops containing residues 141-148 and 190-192. This observation was consistent with experimental data that this region has high B-factors and/or was not identifiable by X-ray crystallography [25, 39]. Significant conformational changes of Tyr143, Gln148, Asn117 and Asp116 were observed in the simulation without metal. It is apparent that the presence of Mg^{2+} in the active site stabilized the loop formed by residues 141-148 (Figure 2.2). Hydrogen bond formation between Gln62 and Gln148 as well as a solvent bridge between Tyr143 and Glu152 played a major role in stabilizing the protein structure in the absence of Mg^{2+}. Without metal ion, water molecules in the HIV-1 IN active site were proposed to interact with Gln148 or the three catalytic residues while these waters were shown to form an octahedral complex with Mg^{2+} in the one metal ion system. The authors proposed that the metal ion was required for catalytic activity.

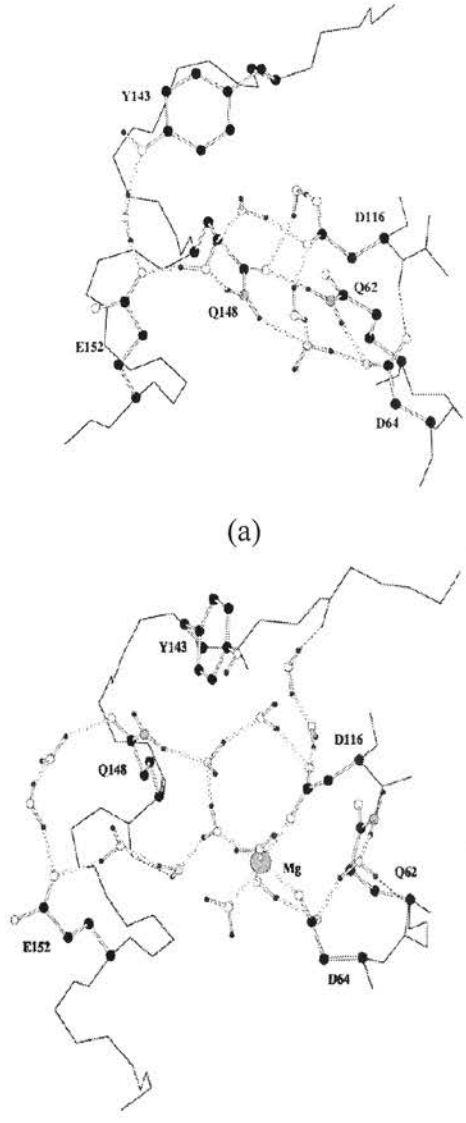


Figure 2.2 Representation of protein-protein and protein-solvent interactions seen in HIV-1 IN active site (a) without Mg^{2+} and (b) with Mg^{2+}.

The above MD simulations [67] focused on only the apo protein, i. e.it lacks information about inhibitor binding. Once the co-crystallized structure of the HIV-1 IN-5CITEP complex became available [39], two MD simulations investigating the

interactions between 5CITEP and the HIV-1 IN catalytic core domain were reported [66, 68]. One of the recent publications involves MD studies of wild-type and the double mutant (T66I/M154I) IN [66]. In this work, both IN were complexed with 5CITEP. This work shed some light on the mechanism for inhibition and drug resistance. The results showed that the binding of 5CITEP to the wild-type IN resulted in conformational changes, especially in the loop 138–149 near the catalytic center. However, the double mutant IN (T66I/M154I) complex demonstrated a different conformational change in loop 138–149. Especially in the orientation of the Tyr143 side chain. For the double mutant with inhibitor, this residue completely pointed towards the active site, while in the wild-type trajectory, it was observed directing away from the active site. Furthermore, the binding mode of the inhibitor was quite different in the two systems. The MD simulations provided more insights into the detailed mechanism of action of the 5CITEP at an atomic level, and suggested the mechanistic cause of IN inhibition and drug resistance.

Moreover, MD simulations of the wild-type and the double mutant HIV-1 IN catalytic core domain complexed with L-731,988, another DKA compound, were conducted [69]. To take into account protein flexibility on complex formation, the Merck inhibitor L-731,988 was docked into different IN conformations. It was discovered that conformation (bound, native, Figure 2.3 (a)) in which the ligand adopted an orientation somewhat similar to the bound 5CITEP conformation in the crystal structure was the most frequently identified orientation for the wild-type complex. A hydrogen bond formed between the ligand and Gln148 played a significant role in stabilizing the complex. For the double mutant complex, two major bound ligand conformations were observed. In the first conformation, (ion bound, double mutant, Figure 2.3 (b)), the ligand interacted directly with Mg^{2+} via its keto-enol moiety while in another conformation (bound, double mutant, Figure 2.4 (b)), the best clustering showed an orientation similar to the conformation of 5CITEP in the crystal structure with the fluoro-benzyl ring interacting with the side chain of Asp116 . Contrary to the wild-type complex, the authors suggested that Gln62 plays an essential role in the interactions in the double mutant complex.

The referred conformations (bound, native), (ion bound, double mutant) and (bound, double mutant) which were observed from molecular docking were subsequently

used as the initial structures to perform MD simulations. The trajectory analyses demonstrated significant differences in the dynamical behavior in the flexible loop (residues 138-149) of these three systems. The catalytic loop displayed a very low mobility for the wild-type complex with inhibitor, whereas this loop was highly flexible in the double mutant complexes (i.e., like that in the unliganded wild-type). Moreover, the conformational change of the catalytic loop of the double mutant complex induced a change in the Gln148 side chain orientation preventing this residue from forming a hydrogen bond with the inhibitor. From these findings, the authors hypothesized that L-731,988 inhibited the wild-type IN by stabilizing the catalytic loop in a non-active conformation as well as potentially preventing DNA binding, and that drug resistance was conferred by the double mutation that restored wild-type flexibility of the catalytic loop and the inhibitor sunk deeper into the active site such that it no longer could prevent substrate DNA binding.

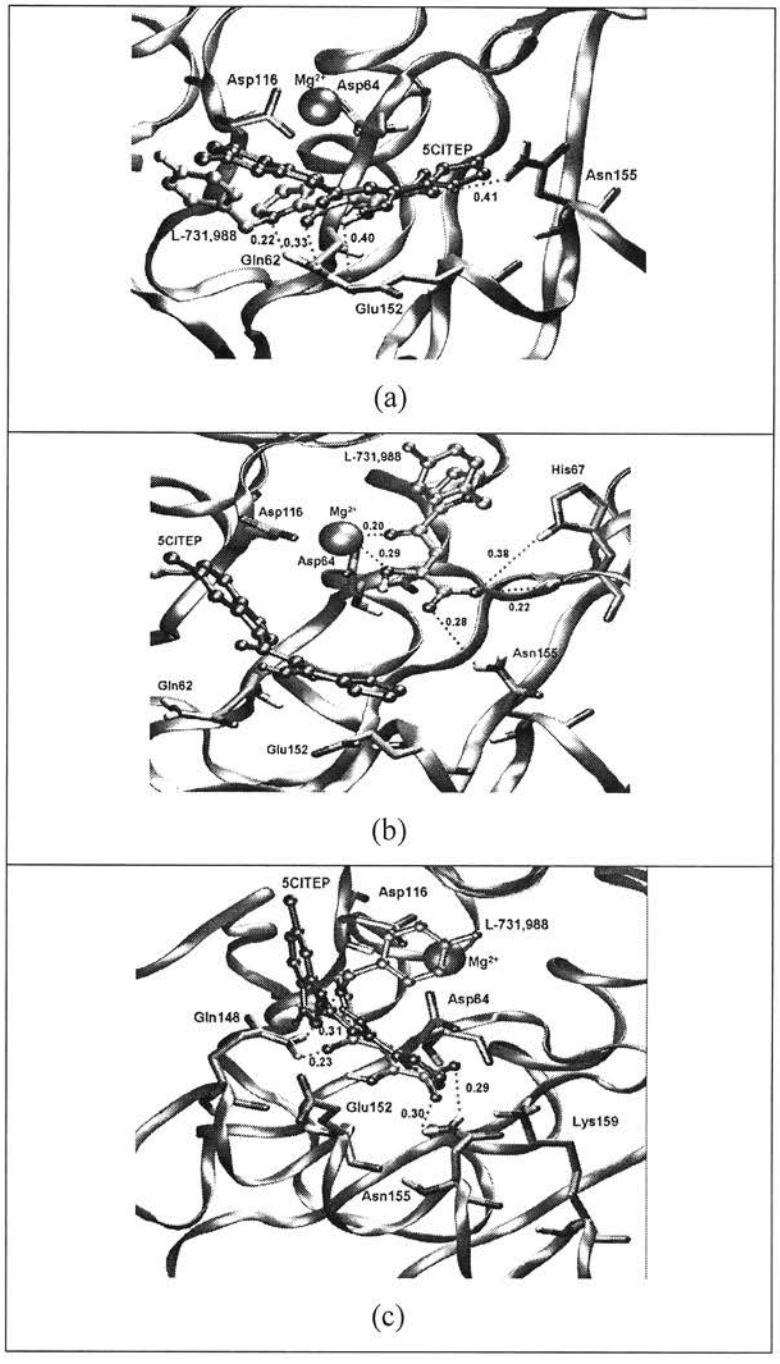


Figure 2.3 Orientation of L-731, 988 in the (a) bound native, (b) ion bound double mutant and (c) the bound double mutant. 5CITEP is shown by dark grey ball and stick model.

Recently, Lee et al. [75] have carried out a series of extended-time MD simulations to investigate the conformational dynamics of the catalytic core domain,

especially the catalytic loop. A set of 44 independent simulations (11 for each model) of the four molecular systems containing wild-type IN, the G140A mutant, the G149A mutant, and the G140A/G149A double mutant were performed. The catalytic loop of the resulting multiple trajectories were subsequently clustered using a heuristic algorithm to highlight the conformational transitions of the loop. The main finding was a major conformational change (>20 Å) in the catalytic loop which revealed a gating-like dynamics, and a transient intra-loop structure which provided a rationale for the mutational effects of several residues on the loop including Gln148, Pro145, and Tyr143. The wild-type catalytic loop has a slow mode of motion that closes and opens the space around the active site, and the transient formation of a 3_{10}-helical structure by the residues around Pro145 appears to be a major influence in the dynamics of the loop. Cluster analyses showed that several types of open conformations were the majority of loop conformations. Moreover, the hinge mutations sterically hinder the loop from closing, therefore, the associated loss of activity strongly suggests that this closing-opening conformational change is functionally important.

In summary, although MD simulations of either HIV-1 IN complexed with inhibitor or free enzyme have been carried out, all of them were based on molecular mechanics theory only. In this thesis, both molecular mechanic and hybrid QM/MM were employed to study the protein-ligand complexes in order to elucidate the difference between the two levels of theory. In addition, the structural and dynamical behaviors of the complexes taken from QM/MM, a higher accurate method, were carefully investigated to describe the difference of inhibitory potency of the two inhibitors, 5CITEP and DKA.

Other computational approaches such as molecular docking, *de novo* design, pharmacophore modeling and database searching of HIV-1 IN inhibitors were also reported [76-78]. This leads to the more progress of HIV-1 IN inhibitors design.