CHAPTER V

Hybrid Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of HIV-1 Integrase/Inhibitor Complexes

This chapter gives detailed discussions of the investigation of structural and dynamical behaviors of HIV-1 IN complexed with DKA analogues by using classical and hybrid QM/MM MD simulations. As stated in Chapter II, MD simulations of either HIV-1 IN complexed with inhibitor or the free enzyme have been carried out [66, 67, 69], however, to the best of our knowledge, this is the first time combined QM/MM MD simulations have been reported for an HIV-1 IN-inhibitor complex. The main aim of this chapter is to investigate structural and dynamical properties of HIV-1 IN complexed with 5CITEP and with DKA. An effort was made to explain the difference in inhibitory potency between these two compounds. Moreover, to provide a point of comparison with conventional MD, classical MD simulations of HIV-1 IN-5CITEP and HIV-1 IN-DKA were also performed and the results obtained from the two levels of theory (conventional and QM/MM) were compared.

In order to make it clear and easier to understand, this chapter is separated into two main parts i.e. initial structure of protein-ligand complex and MD simulations. In the former part, computational method is described altogether with results and discussion because the obtained information from any preliminary step must be used in the subsequent calculations. In the latter part, the computational method is separated from results and discussion.

5.1 Initial structure of protein-ligand complex

In this work, two DKA derivatives i.e. 5CITEP and DKA were studied because of their high potency and high selectivity against HIV-1 IN. Although, the general structure of 5CITEP is very similar to that of DKA, except the tetrazole ring is replaced by a carboxylic acid group in DKA, these two compounds exhibit different inhibition activity. In the presence of Mg^{2+}, the most catalytic metal ion found in human body, DKA is about sixteen-fold and fifty-fold more active than 5CITEP for 3'-processing and ST reactions, respectively [131]. In addition, the inhibitory activities of these two compounds are metal-dependent. Therefore, it is interesting to investigate the structural and

dynamical behaviors of HIV-1 IN-5CITEP and HIV-1 IN-DKA by means of classical or hybrid QM/MM MD simulations and compare the results in order to provide the information supporting the difference of inhibitory potency of the two compounds.
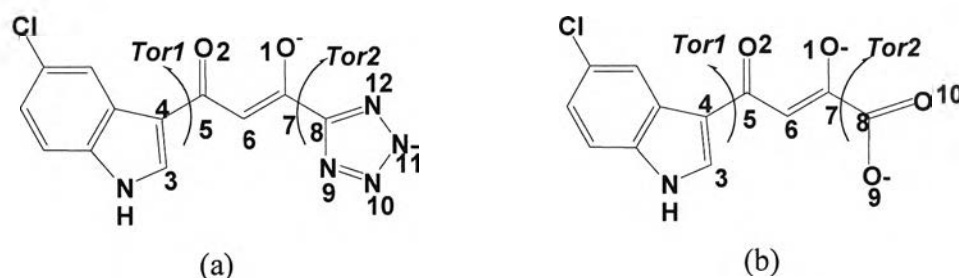

Figure 5.1 Chemical structure and atomic definition of (a) 5CITEP and (b) DKA.

Before performing a conventional or QM/MM MD simulation, the initial structure of the protein-ligand complex must be created. This section gives details how to obtain the starting structures.

Although the co-crystal structure of HIV-1 IN-5CITEP is available, the orientation of inhibitor in the binding pocket and its mechanism of action are still doubted. In the co-crystal structure, 5CITEP located in the center of the active site, lying between Asp64, Asp116 and Glu152 (Figure 1.5). The keto-enol motif of 5CITEP forms a hydrogen-bond with Glu152 while all four nitrogen atoms in the tetrazole ring form hydrogen-bonds with Asn155, Thr66, Lys159 and Lys156. The indole ring of the ligand is pointed toward Gln148. There is no direct interaction between ligand and Mg^{2+} ion. Four water molecules and the side chain oxygen atoms of Asp64 and Asp116 were found to be octahedrally coordinated with the divalent metal ion. However, several docking calculations of 5CITEP to the active site of HIV-1 IN revealed that the bound conformation of 5CITEP in the co-crystal structure could not be reproduced [66, 68, 78, 132-134]. This is possibly owing to the crystal packing effect of the close contact between both subunits A of the dimeric form of HIV-1 IN. In addition, two 5CITEP molecules form hydrogen bonds with one another as they bind to two active sites. This

means that 5CITEP orientation existing in the X-ray structure may not be the actual bioactive form. Therefore, several computational methodologies, including molecular docking and pharmacophore modeling have been employed to elucidate the binding mode of 5CITEP and the other DKA compounds [35, 48, 78, 130, 135-138]. Many of them reveal that the mechanism of the inhibition of this class of compound is due to the chelation of the keto-enol group with the divalent metal ion. As the mechanism of action of DKAs may involve the metal ion, different HIV-1 IN-inhibitor complexes when ligand chelating the Mg^{2+} were herein proposed. Detailed calculations of this part are divided into 4 steps. To determine which part of ligand interacts better with the Mg^{2+}, binding models between metal and ligand were computed in the first step. In the second step, the binding between metal and amino acids were calculated to find the active protonation state of amino acids. In the third step, models combining metal, ligand and amino acids were calculated. In the last step, pair and total interaction energies were calculated to determine the most energetically favorable model.

### 5.1.1. Metal-ligand system
5.1.1.1 Computational Methods

There are many different possibilities that ligand can directly coordinate with the Mg^{2+} ion. Normally the metal complex exists in both tetrahedral and octahedral forms. To find out which part of ligand can better interact with the catalytic metal ion and between tetrahedral and octahedral complexes which produces the lower energy, seven possible tetrahedral and octahedral configurations of metal-ligand chelation were created, as illustrated in Figure 5.2. To reduce computational time and to make it easier to include several possibilities of the models, only fragments of ligands were used. The 1, 3-propanedione, 1-hydroxy-3-propenone and its ionized form represent the diketone, the keto-enol, and the keto-enolate moieties, respectively, of the two inhibitors while the tetrazolate and acetate moieties represent the tetrazolate anion and carboxylate groups of 5CITEP and DKA, respectively. For the keto-enol or the acetate fragment, the Mg^{2+} was manually placed between their two oxygen atoms (model no. 1ml-3ml, no. 7ml) whereas the divalent metal ion was positioned between N9 and N10 (model no. 4ml) or between N10 and N11 (model no. 5ml) or above the plane of tetrazole ring (model no. 6ml). Each

model was geometry optimized using B3LYP/6-31++G (d,p). Interaction energies (ΔE_{int}) were calculated according to the following equation.

$$\Delta E_{int} = E_{cpx} - \sum_i E_i \quad (5.1)$$

where E_{cpx} is the total energy of the system and E_i is the energy of each isolated species. The summation is up to $i = 4$ for the tetrahedral structure and $i = 6$ for the octahedral structure.

Figure 5.2 Proposed models of metal-ligand chelation

4ml-t

4ml-o

5ml-t

5ml-o

6ml-t

6ml-o

7ml-t

7ml-o
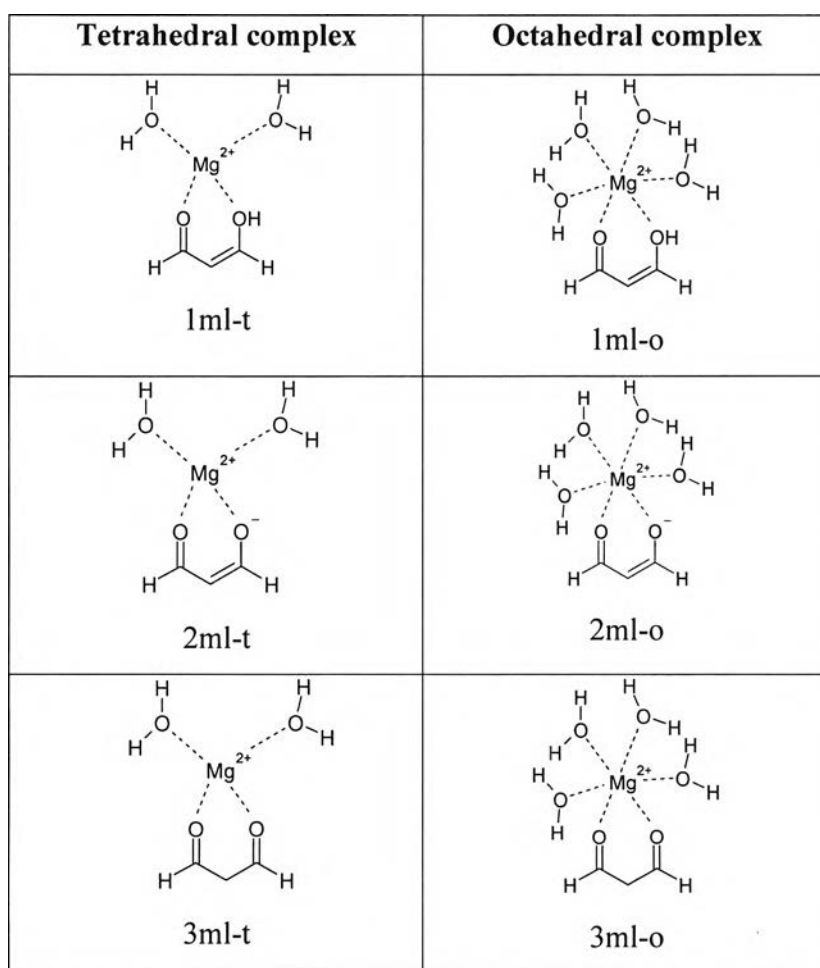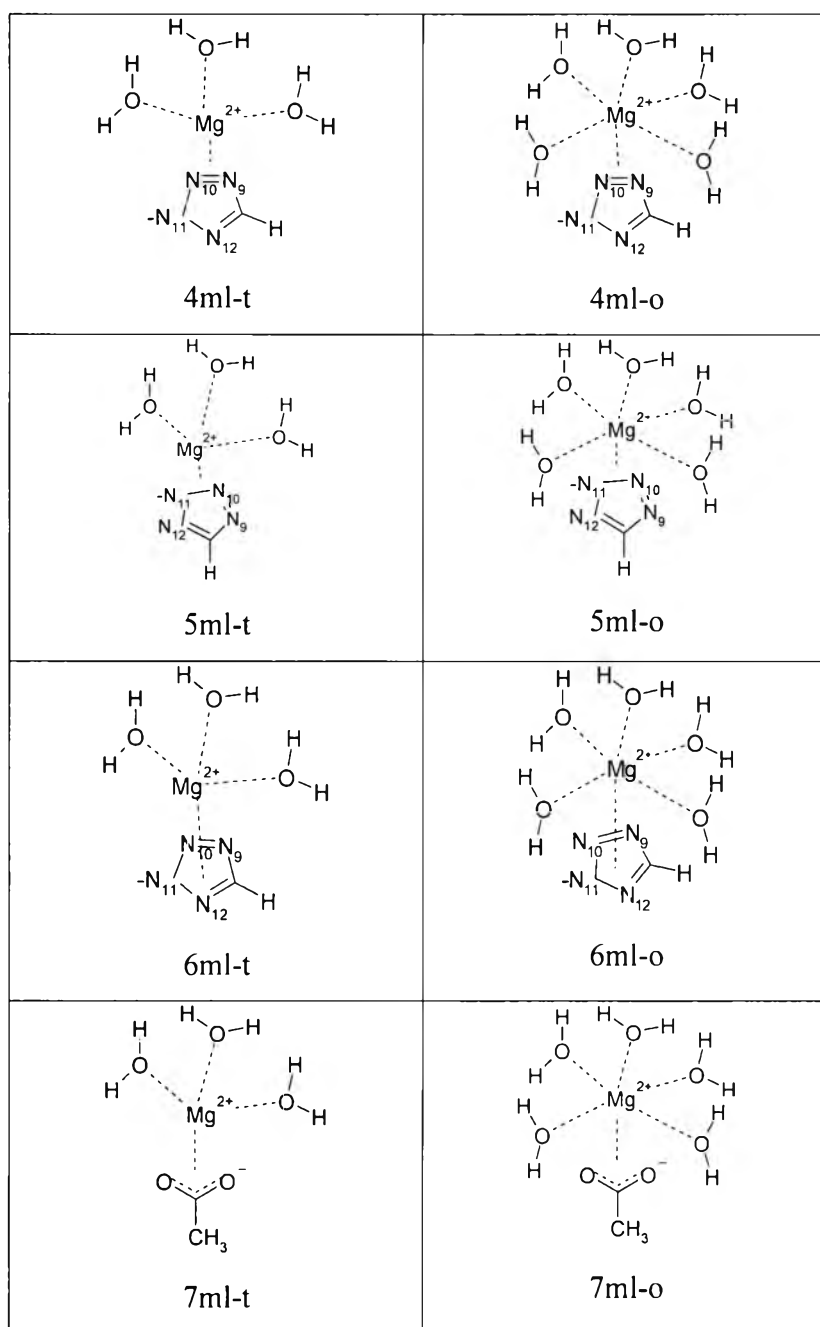
5.1.1.2 Results and Discussion

Interaction energies for each model in both tetrahedral and octahedral complexes are given in Table 5.1. The followings can be summarized from the calculations.

i) Octahedral complexes have lower interaction energies than tetrahedral complexes for almost all models in exception of model no. 4ml. Thus, the six-fold coordination is more stable than the four-fold coordination. Therefore, only the octahedral complex is implemented in our further calculations.

ii) For ligand ionization of 0, the enol form, 1-hydroxy-3-propenone, (Figure 5.2, model no.1ml) posses lower interaction energy than the diketone form, 1, 3-propanedione (model no. 2ml). This supports the predominant of enol tautomer [139]. Only the enol conformer is considered in our subsequent calculations.

iii) For ligand ionization of -1, models no. 2ml, 5ml and 7ml show comparable interaction energy suggesting that the keto-enolate or the tetrazolate or acetate moieties can directly chelate with the metal ion.

iv) When Mg^{2+} forms coordination complex with tetrazolate ring (5CITEP), it will be located between N10 and N11 (model no. 5ml) as indicated by the lower interaction energy in comparison with models no. 4 and 6.

v) For DKA, metal ion can chelate only with the oxygen atoms of acetate (model no. 7ml) which represents the carboxylate moiety of DKA.

Table 5.1 Interaction energies of complexed Mg^{2+}-ligand models

Model No.	Ligand Fragment	Ligand ionization state	ΔE(kcal/mol)	
			Tetrahedral complex	Octahedral complex
1	1-hydroxy-3-propenone	0	-263.38	**-325.05**
2	Ion form of 1-hydroxy-3-propenone	-1	-473.72	**-516.91**
3	1,3-propanedione	0	-254.65	-314.99
4	Tetrazole (N9-N10)*	-1	-452.78	-503.64
5	Tetrazole (N10-N11)**	-1	-458.04	**-507.23**
6	Tetrazole (top)***	-1	-454.67	-448.72
7	Acetate	-1	-491.16	**-526.16**

* = Mg^{2+} is placed between N9 and N10 of tetrazole

** = Mg^{2+} is placed between N10 and N11 of tetrazole

*** = Mg^{2+} is placed in the top of the plane of tetrazole ring

5.1.2 Metal-amino acid system

5.1.2.1 Computational Methods

The X-ray HIV-1 IN-5CITEP indicates that the Mg^{2+} is hexa-coordinated with oxygens of Asp64, Asp116 and four water molecules [39]. The protonation states of key residues are often important for the function of a protein. However due to limitations of experimental techniques, hydrogen atoms are generally not given in the X-ray structure and the ionization state of amino acids is unclear. Thus, different protonation states of Asp64 and Asp116 which are octahdrally complexed with the metal were modeled, as shown in Figure 5.3. To reduce computational cost, the acetate was used to represent the aspartic acid residues. Each model was subjected to geometry optimize using the B3LYP/6-31++G (d, p) and interaction energies were calculated according to Eq. 5.1.
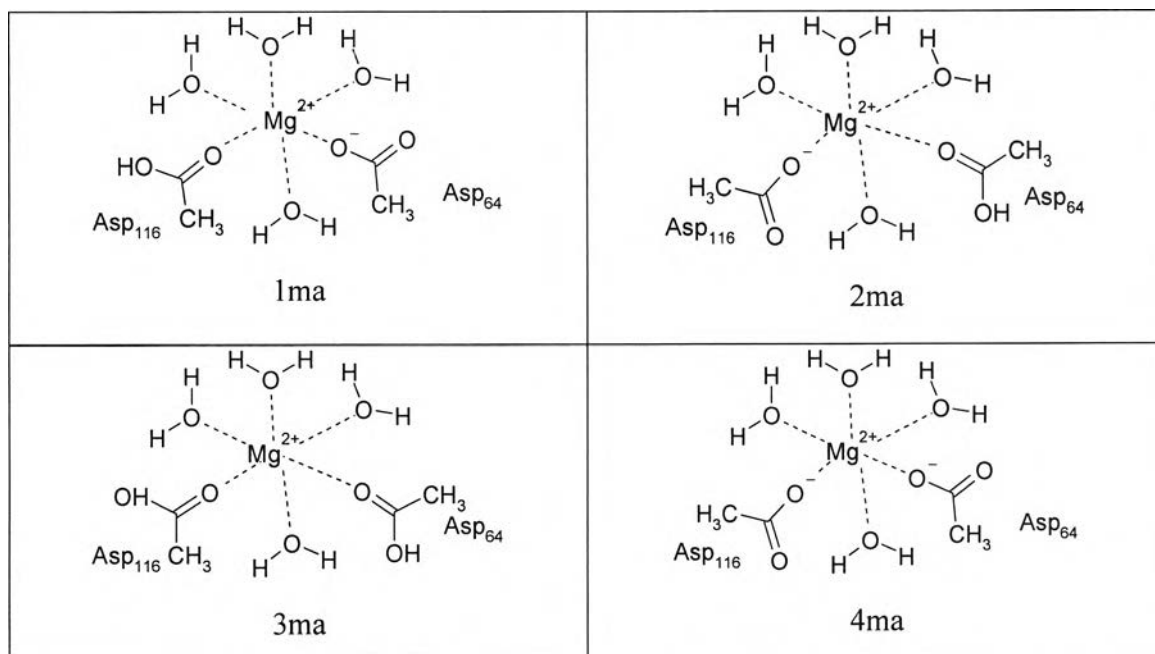
Figure 5.3 Proposed models of metal-amino acids chelation.

5.1.2.2 Results and discussion

Interaction energies given in Table 5.2 indicate that the deprotonated form of both Asp64 and Asp116 (model no. 4ma) is the most energetically favorable model. The monoprotonated form at either Asp64 (-533.05 kcal/mol) or Asp116 (-531.36 kcal/mol) shows no favorable energy significant difference. However, the diprotonated form posses the highest interaction energy. Therefore, the ionization state of -1 for both Asp64 and Asp116 was employed in further calculations.

Table 5.2 Interaction energies of Mg^{2+}- amino acids models.

Model No.	Ionization state		ΔE(kcal/mol)
	Asp64	Asp116	
1ma	-1	0	-533.05
2ma	0	-1	-531.36
3ma	0	0	-355.32
4ma	**-1**	**-1**	**-643.80**

5.1.3 Hybrid of metal-ligand-amino acid system

5.1.3.1 Computational Methods

The coordination complexes of Mg^{2+} with either ligand or amino acids were proposed in 5.1.1 and 5.1.2, respectively. The next step is to combine Mg^{2+}, ligand, and amino acids altogether. From the results of 5.1.1 (Table 5.1), we know which part of ligand better interacts with the Mg^{2+} and between the tetrahedral and octahedral complexes, which gives the lower energy. In addition, the results of 5.1.2 suggest the active protonation state of Asp64 and Asp116. Based on the information obtained from these previous two sections, four hexa-coordination complexes of Mg^{2+} with both ligand and amino acids were proposed as shown in Figure 5.4. Geometry optimization of each model was consequently performed at the B3LYP/6-31++G (d,p) and the interaction energies were calculated based on Eq. 5.1.
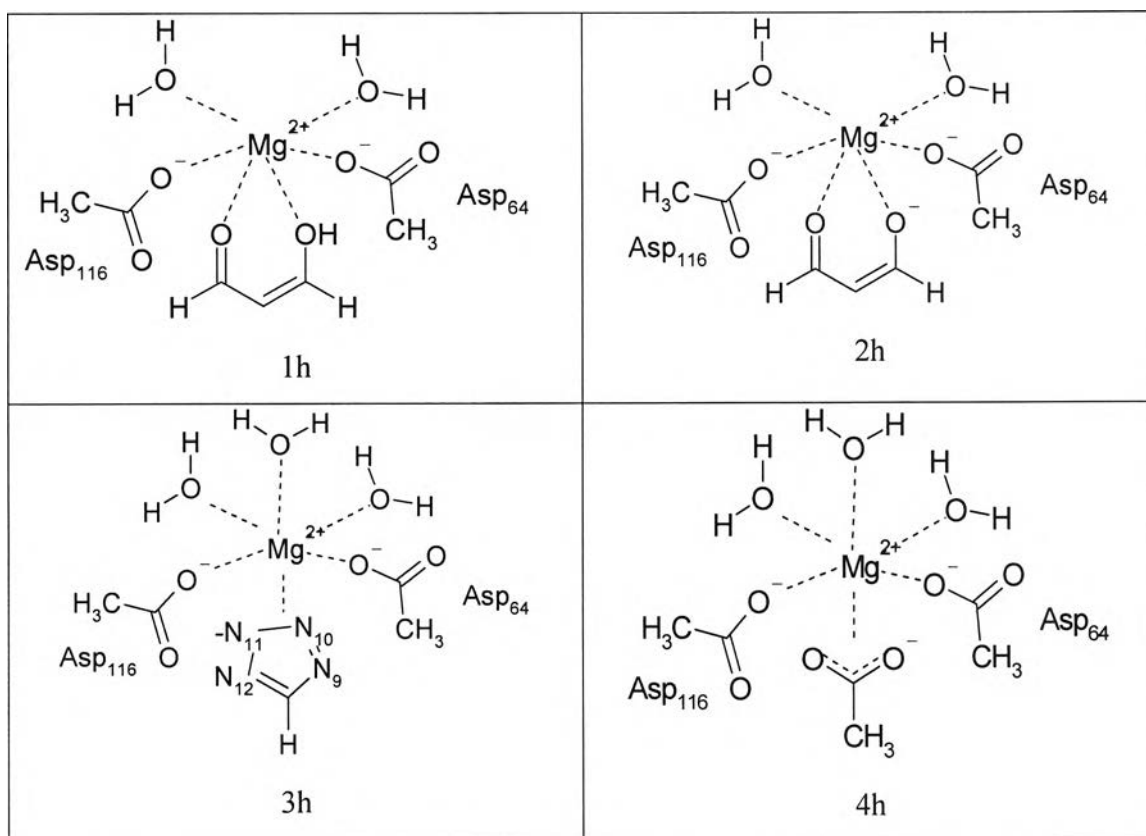
Figure 5.4 Hybrid chelation of metal, ligand and amino acids.

5.1.3.2 Results and Discussion

It can be clearly seen from Table 5.1 that the enol tautomer is energetically more favorable than the diketone. Moreover, according to the interaction energy, either the keto-enolate or the tetrazolate or the acetate groups can form coordination with the Mg^{2+}. However, the protein environment was not taken into account. Therefore, the calculations were extended by including the Mg^{2+}, ligand and amino acids altogether. This combination complex of metal, ligand, and amino acids, was termed "hybrid" system. Interaction energy of each hybrid complex is shown in Table 5.3. The interaction energy of the keto-enolate conformer (model no. 2h) compared to that of the keto-enol one (model no.1h) was lower by 15 kcal/mol. This implies that the ligand state of -1 is energetically more favorable than that of state of 0. For ligand state of -1, it is apparent that both keto-enolate (model no. 2h) and tetrazolate (model no.3h) of 5CITEP can

coordinate with the divalent metal ion as these two models gave almost the same values of interaction energies. For DKA, model no. 4h in which acetate moiety chelates with the Mg^{2+} gave lower interaction energy (~ 15kcal/mol) than that of model no. 2h in which the keto-enolate coordinates with the divalent metal ion.

Because only fragments of ligand or a few amino acids were considered during the calculations, the question of which part of ligand could directly chelate with the metal ion does not absolutely resolved. However, the results in this section provide a guidance of possible binding between metal and ligand.

Table 5.3 Interaction energies of the hybrid system.

Model No.	Ligand fragments	Ionization State					ΔE (kcal/mol)
		Total	ligand	Mg^{2+}	Asp64	Asp116	
1h	1-hydroxy-3-propenone	0	0	2	-1	-1	-655.93
2h	Ion form of 1-hydroxy-3-propenne	-1	-1	2	-1	-1	-670.62
3h	Tetrazolate	-1	-1	2	-1	-1	-670.26
4h	Acetate	-1	-1	2	-1	-1	-684.36

5.1.4 Pair and total interaction energies calculations
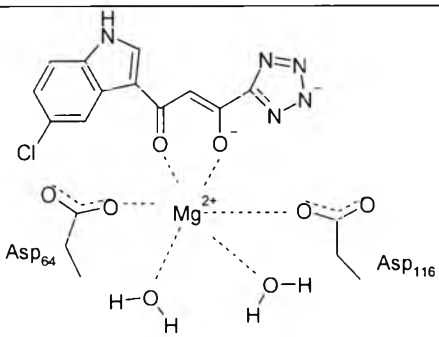
5.1.4.1 Computational Methods

Based on the information obtained from section 5.1.3, seven different possible HIV-1 IN-inhibitor configurations (models 1p-7p) for each inhibitor i.e. 5CITEP and DKA, were created (Table 5.4). The keto-enol and the keto-enolate moieties of the complete structure of 5CITEP and DKA were superimposed to the ion form of 1-hydroxy-3-propenone. In addition, the tetrazole of 5CITEP and the carboxylate of DKA
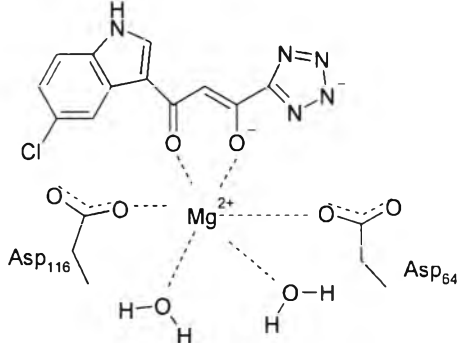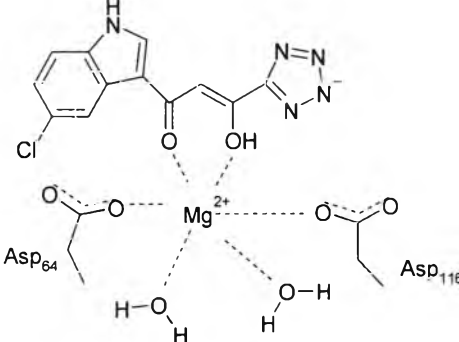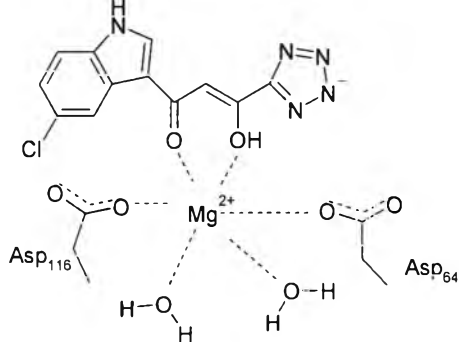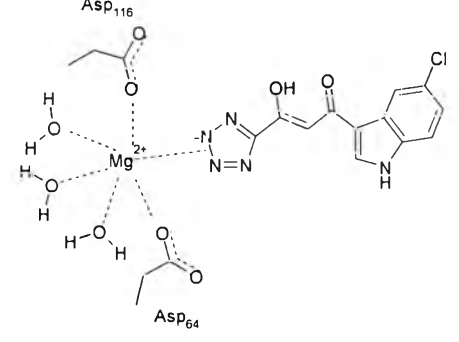
were overlaid to the tetrazolate ring and the acetate moieties, respectively. In the first two models (no.1p and 2p), ligand is in dianionic form i.e. the keto-enol and the tetrazole moieties were ionized and the keto-enolate was directly chelated with the Mg^{2+}. Model 1p and 2p differ in the orientation of the chloroindole ring of inhibitor. For the remaining models no. 3p-6p, ligand is in monoanionic form with the deprotonated tetrazole ring.
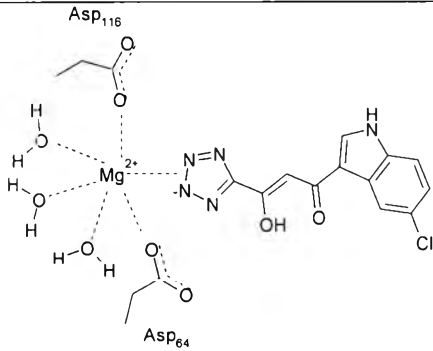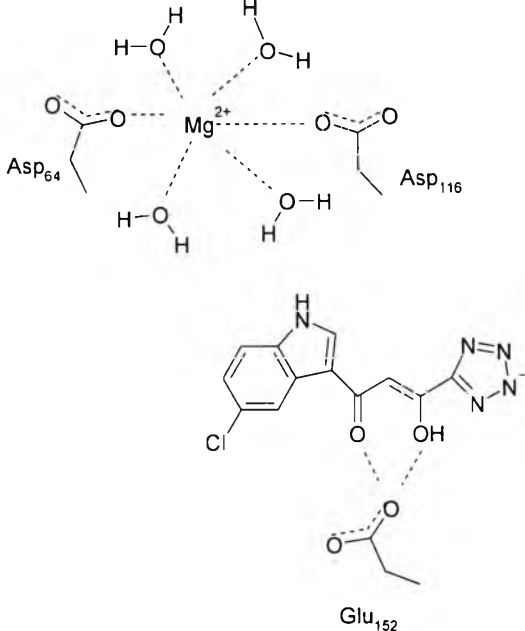
For models no. 3p and 4p, the keto-enol moiety of ligand forms coordination with Mg^{2+} while the chloroindole ring of these two models is oriented in the opposite direction. In model no. 5p, the tetrazolate anion chelates with the divalent metal ion and the keto-enol group points inward to enzyme. Model no. 6p is similar to model no. 5p but the keto-enol function points toward solvent. In addition to the proposed protein-ligand complexes, HIV-1 IN-inhibitor configuration corresponding to the X-ray co-crystal structure of 5CITEP (1QS4) was also included (X-ray model, no.7p). In this model, the keto-enol of ligand interacts with Glu152 and no direct interaction between ligand and metal ion.

For every model, the HIV-1 IN-ligand configurations were geometry minimized using the AMBER program. Pair interaction energies (see Eq. 5.1) between inhibitor and Mg^{2+} and between inhibitor and amino acids whose locate within 8 Å from inhibitor (Table 5.4) were calculated at B3LYP/6-31++G (d,p). Subsequently, total interaction energies were determined from the summation of the pair energies.

Table 5.4 Description of models and list of amino acids within 8 Å from inhibitor.

Model	Ligand state	Description	Residues within 8 Å from inhibitor
 1p	-2	Keto-enolate coordinates with Mg^{2+} and tetrazolate ring is near to Asp116	Asp64, Cys65, Thr66, His67, Asp116, Asn117, Gly118, Ser119, Glu152, Asn155, Lys156, Lys159,

2p	-2	Similar to model 1p but the tetrazolate ring is flipped (close to Asp64).	Asp64, Cys65, Thr66, His67, Val75, Asp116, Asn117, Gly118, Ile151, Glu152, Asn155, Lys156, Lys159
3p	-1	Similar to model 1p but enolate is replaced by enol group	Asp64, Cys65, Thr66, His67, Asp116, Asn117, Gly118, Ser119, Asn120, Phe121, Glu152, Asn155, Lys156, Lys159
4p	-1	Similar to model 2p but enolate is replaced by enol group	Asp64, Cys65, Thr66, His67, Val75, Asp116, Asn117, Gly118, Ile151, Glu152, Asn155, Lys156, Lys159
	-1	Tetrazolate ring coordinates with Mg^{2+}, the keto-enol points inward to enzyme	Asp64, Cys65, Thr66, Hid67, Val75, Asp116, Ile151, Glu152, Ser153, Asn155, Lys156, Lys159, Lys160

5p			
Asp₁₁₆ ... Mg²⁺ ... Asp₆₄ structure (keto-enol indole tetrazole, Cl) **6p**	-1	Similar to model 5p but the keto-enol is flipped i.e. it points toward solvent	Asp64, Cys65, Thr66, His67, Leu68, Glu92, Asp116, Gly118, Glu152, Asn155, Lys156, Lys159
Asp₆₄ ... Mg²⁺ ... Asp₁₁₆ structure with Glu₁₅₂ (Cl indole, tetrazole) **7p**	-1	Based on the X-ray co-crystal structure of HIV-1 IN-5CITEP	Gln62, Asp64, Cys65, Thr66, His67, Thr115, Asp116,Asn117, Gln148, Gly149, Ile151, Glu152, Ser153, Asn155, Lys156, Lys159

5.1.4.2 Results and Discussion

To further determine the lowest energetically favorable model of protein-ligand complex, pair interaction energies between ligand and its surrounding amino acids within 8 Å of seven models for each inhibitor were calculated and the results are given in Table 5.5 and 5.6 for HIV-1 IN-5CITEP and HIV-1 IN-DKA, respectively. From the list of pair interaction energies, it is likely that the interaction between ligand and charge amino

acids gives high contribution to total interaction energy. Therefore, it is expected that if there is no charge residues, the total interaction energy may not significantly increase. Model no.1p is given as an example. Only one charge residue i.e. Glu157 was found in the region between 8 Å and 12 Å from inhibitor. The total interaction energy of this model based on the cutoff 12 Å (-340.78 kcal/mol) slightly differs from than that of 8 Å (~ 30 kcal/mol). Therefore, the use of cutoff at 8 Å should sufficient enough to estimate the total interaction energy of each model.

From both tables, it is clear that for ligand state = -2, model no. 2p posses the lowest total interaction energy for each inhibitor. For the ligand state =-1, the X-ray model (model no. 7p) of each inhibitors is the most stable configuration.

Table 5.5 Pair and total interaction energies (kcal/mol) of HIV-1 IN-5CITEP.

No.	Amino/water	Model						
		1p	2p	3p	4p	5p	6p	7p (X-ray)
1	Gln62	-	-	-	-	-	-	-2.02
2	Asp64	108.26	121.33	56.35	58.89	66.18	63.91	39.48
3	Cys65	0.69	7.80	0.86	-0.79	4.81	1.01	1.98
4	Thr66	-6.00	-21.67	-0.51	-10.44	-8.07	-1.99	-9.51
5	Hid67	3.01	2.56	2.75	4.03	2.46	-4.87	2.17
6	Leu68	-	-	-	-	-	-0.28	-
7	Val75	-	-1.54	-	0.01	0.02	-	-
8	Glu92	-	-	-	-	-	49.18	-
9	Thr115	-	-	-	-	-	-	-0.58
10	Asp116	115.92	106.27	51.36	52.26	50.96	53.16	23.88
11	Asn117	-16.33	3.17	-8.26	5.38	-	-	-5.47
12	Gly118	-5.67	-2.93	-1.47	-1.41	-	-1.72	-
13	Ser119	-1.05	-	4.18	-	-	-	-
14	Asn120	-	-	0.33	-	-	-	-
15	Phe121	-	-	-3.26	-	-	-	-
16	Gln148	-	-	-	-	-	-	3.10
17	Gly149	-	-	-	-	-	-	1.31
18	Ile151	-		-	2.22	0.95	-	1.16
19	Glu152	61.94	81.78	28.33	41.46	46.26	37.62	56.56
20	Ser153	-	-	-	-	-	-	0.16
21	Asn155	-4.32	-27.96	-3.97	-19.09	-13.34	-4.66	-10.37
22	Lys156	-51.22	-71.44	-25.42	-35.78	-54.01	-30.81	-94.03
23	Lys159	-78.54	-138.15	-38.79	-82.11	-76.62	-45.16	-91.07
24	Lys160	-	-	-	-	-22.68	-	-
25	Mg^{2+}	-519.67	-522.58	-322.81	-315.03	-310.74	-287.67	-250.26
26	Water#1	8.68	10.50	7.60	7.41	3.99	3.39	1.12
27	Water#2	12.15	10.84	8.60	6.33	2.39	4.57	1.29
28	Water#3					5.64	6.62	-0.55
29	Water#4							0.94
TOTAL		-372.15	-442.02	-244.13	-286.66	-301.80	-157.70	-331.65

Table 5.6 Pair and total interaction energies (kcal/mol) of HIV-1 IN-DKA.

No.	amino	Model						
		1p	2p	3p	4p	5p	6p	7p (X-ray)
1	Gln62	-	-	-	-	-	-	0.55
2	Asp64	115.25	121.24	58.74	66.07	63.53	60.78	37.13
3	Cys65	0.35	7.41	0.79	6.44	5.16	4.07	1.97
4	Thr66	-6.55	-35.81	-0.60	-19.40	-10.88	-1.60	-9.88
5	Hid67	4.01	-0.85	2.54	2.28	2.57	3.16	1.99
6	Leu68	-	-	-	-	-	-0.33	-
7	Val75	-	-3.23	-	0.04	0.000	-	-
8	Thr115	-	-	-	-	-	-	-0.44
9	Asp116	127.24	107.95	61.69	53.13	39.71	49.74	33.03
10	Asn117	-32.11	8.03	-11.38	2.70	-	-	-
11	Gly118	-8.01	-2.91	-2.22	-5.39	-	-	-
12	Ser119	-2.16	-	-1.32	-	-	-	-
13	Asn120	-	-	0.43	-	-	-	-
14	Phe121	-7.83	-	-4.62	-	-	-	-
15	Pro142	-	-	-	-	-	-	-0.07
16	Gln148	-	-	-	-	-	-	1.50
17	Gly149	-	-	-	-	-	-	3.25
18	Ile151	-	0.44	0.12	1.09	-2.16	-	-
19	Glu152	56.95	80.61	28.52	40.10	52.27	36.65	58.78
20	Ser153	-	-	-	-	-	-	-0.09
21	Asn155	-11.21	-33.53	-10.62	-5.58	-16.40	-5.83	-22.03
22	Lys156	-54.22	-70.44	-24.70	-33.82	-49.74	-36.85	-100.27
23	Lys159	-77.02	-162.51	-37.58	-71.87	-76.67	-51.23	-112.25
24	Mg^{2+}	-532.76	-525.22	-330.54	-312.86	-273.47	-276.51	-255.12
25	Water#1	7.17	11.79	8.12	7.13	-11.96	-8.92	0.87
26	Water#2	11.79	10.04	8.96	6.19	-13.62	-12.80	0.71
27	Water#3					4.11	4.43	-1.80
28	Water#4							-2.62
TOTAL		-409.11	-486.99	-253.67	-263.75	-287.55	-235.24	-364.79

Recently, the cavities for 3'-processing and ST have been proposed (Figure 5.5) [135]. The area which starts from Lys159 and extends to the region surrounded by residues Asp64, Asp116, Asn120, Glu92, His67, Thr66, and Cys65, refers to 3'-processing cavity. Model no. 2p which was based on the assumption that (i) ligands (5CITEP and DKA) are in their dianioic form according to the experimental studied by Maurin and coworkers [140] (ii) only one Mg^{2+} is modeled in the active site since the DNA substrates are not included in our structures and (iii) the keto-enolate group of the ligand directly forms a coordination complex with Mg^{2+}, could correspond to the binding at this site because the inhibitor coordinates with the metal ion and located in the 3'-

processing cavity which requires divalent metal ion. On the other hand, the ST cavity is covered by a part of flexible loop (Gly140-Gly149) and spreads into the region near to α4-helix. The X-ray model could correspond to the binding at this site because inhibitor is located near to the flexible loop region and the metal ion chelates with only amino acids. However in this study, our attention was paid on the binding between ligand and Mg^{2+} existing in HIV-1 IN active site, therefore, only Model no. 2p was focused and consequently used as the initial structure to perform classical or QM/MM MD simulations.
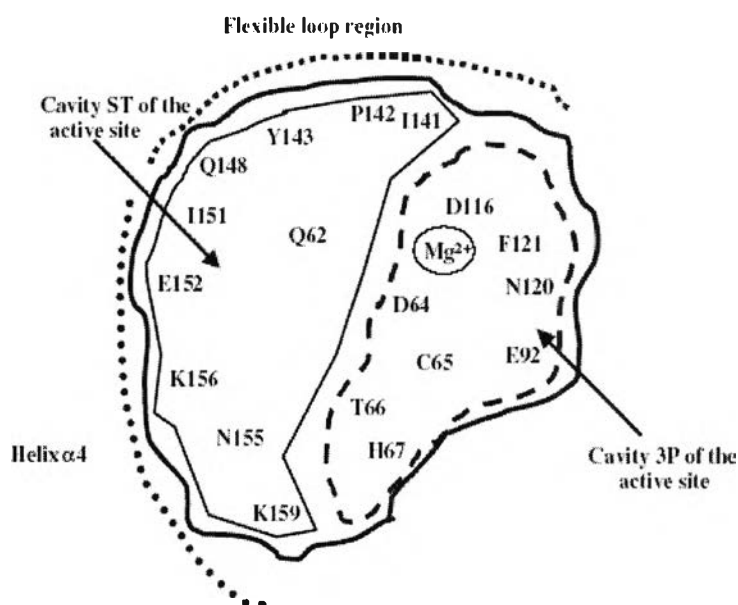
Figure 5.5 Schematic representation of the active site of HIV-1 IN and the amino acid residues involved in 3'-processing and strand transfer reactions [135].

5.2 MD simulation

5.2.1 Computational methods

5.2.1.1 System set-up and Conventional MD simulation

Initial structures of the two systems of HIV-1 IN-inhibitor complexes were taken from the lowest total interaction energy of ligand state =-2 (model 2p, see 5.1.4.2) A schematic representation of the ligand binding to the HIV-1 IN active site is illustrated in

Figure 5.6. For the first system, HIV-1 IN-5CITEP, HIV-1 IN complexes with 5CITEP and for the second system, HIV-1 IN-DKA, the protein complexes with DKA. For both systems, Mg^{2+} is octahedrally coordinated with the OD1 atoms of residues Asp64, Aps116, O atoms of water molecules (water #1 and water #2) and O atoms belonging to the keto-enolate moiety of the ligands whose dianionic form was used (Figure 5.6). No unusual amino acid ionization states were identified through application of the ionization state prediction methodology in the UHBD program [141] (i.e., Lys and Arg = +1, Asp and Glu = -1, His = 0).
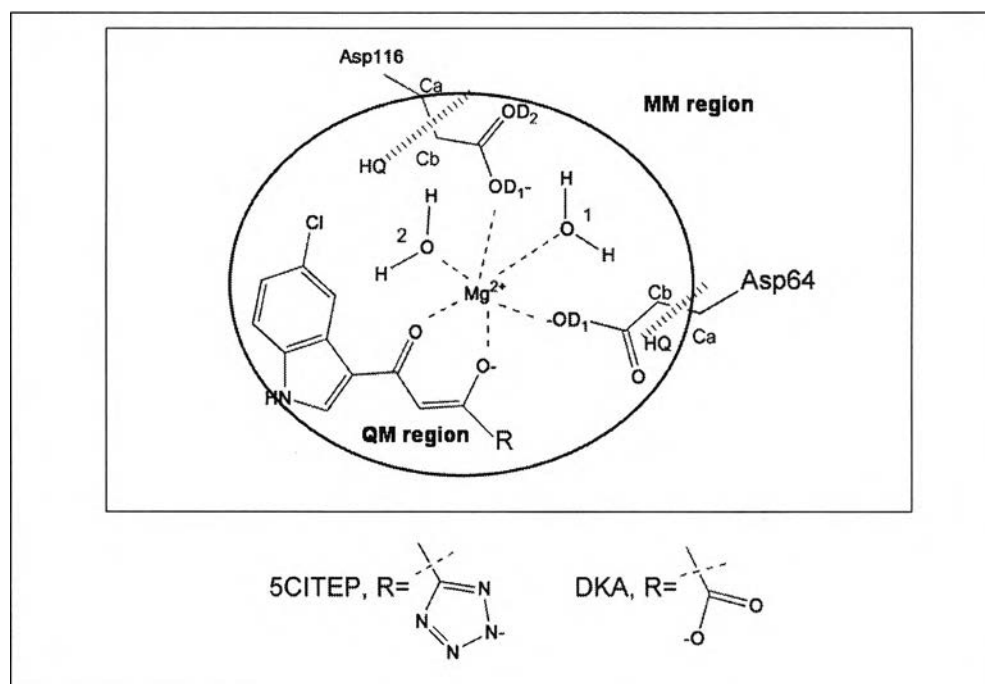
Figure 5.6 Schematic drawing of proposed metal-ligand binding of HIV-1 IN-inhibitor complexes. A QM regions used in QM/MM calculation are encircled. HQ refers to hydrogen link atom HQ type which is added between the alpha carbon and beta carbon of both Asp64 and Asp116.

In each system, the complex was solvated with a 28 Å radius sphere of TIP3P water centered on the Mg^{2+} ion in the active site. Overlapping water molecules whose their oxygen atoms were closer than 2.8 Å to any other heavy atom were removed. Subsequently, the added water molecules were energy minimized using the two series of 300 steps of steepest descent (SD) following by two series of 500 steps of adopted basis Newton Raphson (ABNR) to relieve steric clashes within the system. The solvent sphere was equilibrated (at 300 K) for 25 ps with protein and ligand fixed. These solvation procedures (sphere addition, water deletion and equilibration) were repeated several times to avoid solvent cavities. As a result, the IN-5CITEP system contained 2,406 solute atoms (protein, Mg^{2+}, 5CITEP, two "catalytic" water molecules) and 7,371 water atoms. For the IN-DKA system, the numbers of solute (protein, Mg^{2+}, DKA, two catalytic water molecules) and solvent atoms were 2,404 and 7,365, respectively.

In each case, the enzyme-inhibitor complex was energetically minimized using the two series of 300 steps SD and 1,000 steps ABNR. The system was further slowly heated from 10 to 300 K in three 10 ps intervals, 10-100 K, 100-200 K, and 200-300 K. Afterward, the complex was equilibrated for 300 ps and the resulting structure was consequently used as the starting configuration for QM/MM MD simulation. For the conventional MD simulation, the system was further equilibrated with stochastic boundary conditions [142] over 1,000 ps and the production run was collected from 400-1,000 ps.

All calculations were conducted using the CHARMM suite of programs, version 29b2 [82]. A stochastic boundary molecular dynamics [142] simulation was employed. The 28 Å sphere was subdivided into the reaction zone (0-22 Å) treated by Newtonian dynamics and a buffer zone (22-26 Å) treated by Langevin dynamics. Protein atoms within the buffer region were harmonically restrained to their minimized positions. Friction coefficients of 250/ps and of 62/ps [143] were used for non hydrogen protein atom and water oxygen atoms, respectively. Amino acid residues that were more than 26 Å away from the center of the sphere were restrained during the calculations. To prevent escape of water molecules, a deformable boundary potential [87, 144] at 28 Å was employed. The SHAKE [145] algorithm was employed to constrain all bond containing hydrogen atoms. The time step was 1 fs. A non-bonded cutoff with an atom switching

function in the region from 8 to 13 Å was used to smoothly scale down the interaction to zero at 13 Å.

5.2.1.2 QM/MM set up

The QM region contains the acetate side chains of Asp64 and of Asp116, two "catalytic" water molecules, an Mg^{2+} ion, ligand and two hydrogen link atoms totaling 47 and 45 atoms in the IN-5CITEP and IN-DKA systems, respectively (Figure 5.5). To ensure that the MM charges were treated consistently in the Hamiltonian for the QM atoms, all QM atoms were defined as a single group. The hydrogen link atoms, HQ type,[93] were added between the Cα and the Cβ of aspartic acids to saturate the QM and MM boundary. The overall charge in the QM region was -2e for both systems. The QM atoms were described by the semi-empirical PM3 method [146-148] while the other atoms were treated by a classical CHARMM22 force field [149]. QM/MM energy minimization (250 steps of SD followed by 750 steps of ABNR) was performed for each system. Afterward, the QM/MM MD simulation was conducted using the same conditions as the classical MM simulation and was run for 1200 ps, with the final 600 ps taken as the production data.

5.2.2 Results and Discussion

A. Comparison of HIV-1 IN-5CITEP using MM and QM/MM force fields

In this section, comparison of results between MM and QM/MM calculations of the IN-5CITEP and IN-DKA systems was made in order to elucidate the differences between the two levels of theories.

A.1 IN-5CITEP

A.1.1 Structural dynamics of the HIV-1 IN in HIV-1 IN-5CITEP

To compare the use of MM and QM/MM parameters, MD simulations of HIV-1 IN-5CITEP based on classical and QM/MM parameters were examined. Trajectory analyses were carried out on the production runs, i.e. 400-1000 ps and 600-1200 ps for the conventional MD and hybrid QM/MM MD simulations, respectively. To assess the stability of the MM and QM/MM trajectories, the root mean square deviations (RMSD)

for all heavy atoms were plotted versus simulation time. Figure 5.7 shows RMSD with respect to the average structures of the two simulations. It can be clearly seen that both simulations produced stable trajectories as indicated by their small RMSDs (~0.6 Å). The small RMSD fluctuation of the two trajectories during the course of dynamics runs confirms the reasonable stabilities of the systems.
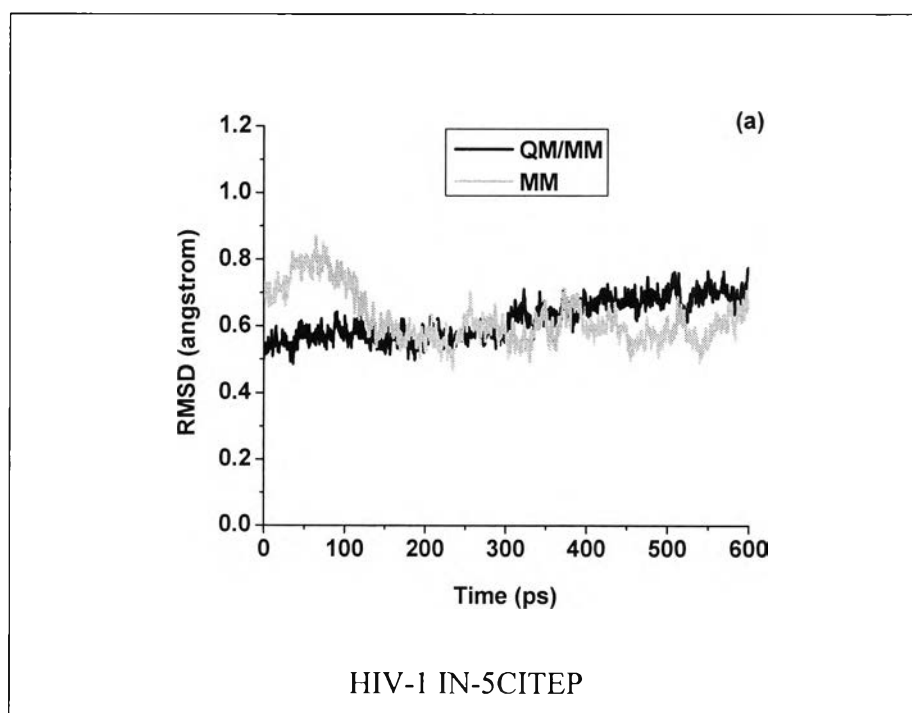

HIV-1 IN-5CITEP

Figure 5.7 Time-dependent heavy atom RMSDs with respect to the average structure versus simulation time of HIV-1 IN-5CITEP which were obtained from conventional MD and QM/MM MD simulations.

To find out which part of protein shows high flexibility, the root mean square fluctuation (RMSF) of backbone atoms for each residue during the MD simulations was monitored, as depicted in Figure 5.8. The MM results revealed that the largest fluctuations of the protein are seen in the two areas corresponding to residues 116-119 and residues 140-149. Meanwhile, the QM/MM calculation showed that only the loop

region 140-149 was found to be flexible whereas residues 116-119 were quite rigid throughout the simulation.
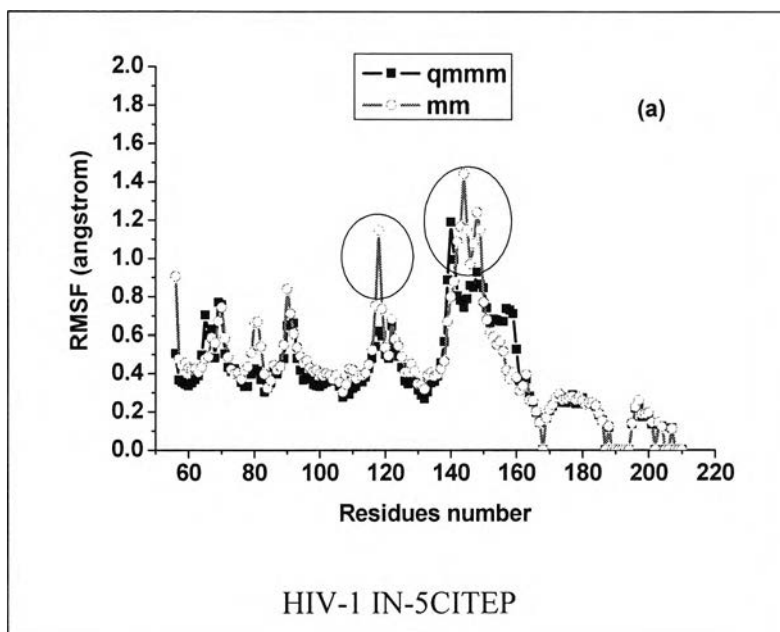

Figure 5.8 Average RMSF of protein backbone atoms as a function of residue number of HIV-1 IN-5CITEP. The protein structures were taken from classical and hybrid QM/MM MD simulations.

To further investigate the structural difference of the protein-ligand complex between the two levels of calculations, the average structures of the protein taken from the two simulations is superimposed as illustrated in Figure. 5.9 (a). The RMSD of all heavy atoms in the protein between the two simulations is 1.36 Å. As can be clearly seen, two significant differences in protein structure between the two levels of theory are the conformations in the area involving residues 116-119 and residues 140-149. The RMSD (all heavy atoms) of amino acid residues 116-119 and 140-149 of the average structures between MM and QM/MM simulations are 1.24 Å and 1.77 Å, respectively. The conformation of Asp64 using MM parameters was somewhat similar to that of using QM/MM while Asp116 adopted a different orientation in the two levels of calculations. Due to the adjustment of Mg^{2+} and 5CITEP to the new equilibrated state of the system,

Asp116 adopted its conformation to preserve the interaction with this catalytic metal ion. This leads to the conformational difference of the region containing residues 116-119. In this area, we note that Gly118 exhibited the highest fluctuation. In the conventional MM simulation, this loop adopted a conformation away from residues Phe121 and Thr122, making it difficult for Gly118 to form hydrogen-bonds with the above-mentioned two amino acids (Figure 5.9 (b)). An analysis of hydrogen-bonding shows that hydrogen-bonds occur through 20% of the simulation between Gly118 and Phe121, and 30% of the time between Gly118 and Thr122. On the contrary, in the QM/MM MD simulation, Gly118 is located close to Phe121 and Thr122. The carbonyl oxygen of Gly118 is positioned between the NH's of Phe121 and Thr122, thus, it is easier for Gly118 to establish such hydrogen-bonds (Figure 5.9 (c)). These hydrogen-bonds are almost always present during the dynamics simulation, indicated by 83% and 97% of an occurrence of hydrogen-bond between Gly118 and Phe121 and between Gly118 and Thr122, respectively. Apparently, the rigidity of Gly118 observed in QM/MM is due to hydrogen-bond between its carbonyl oxygen and the NH belonging to Phe121 and Thr122.
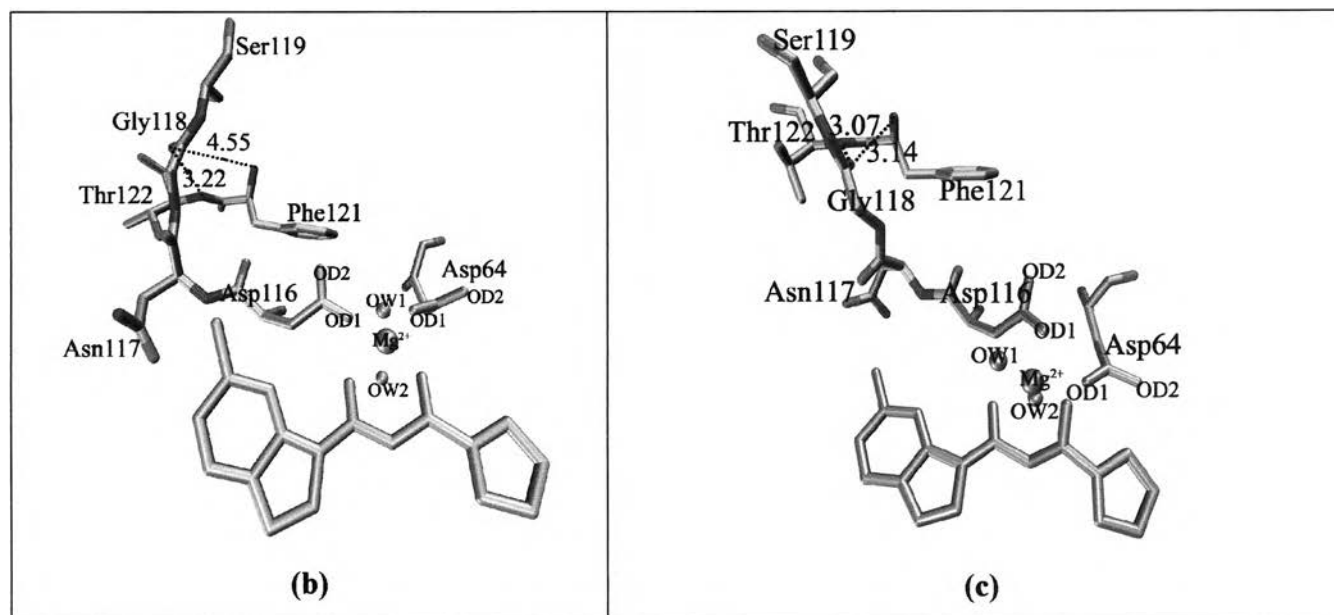

(a)
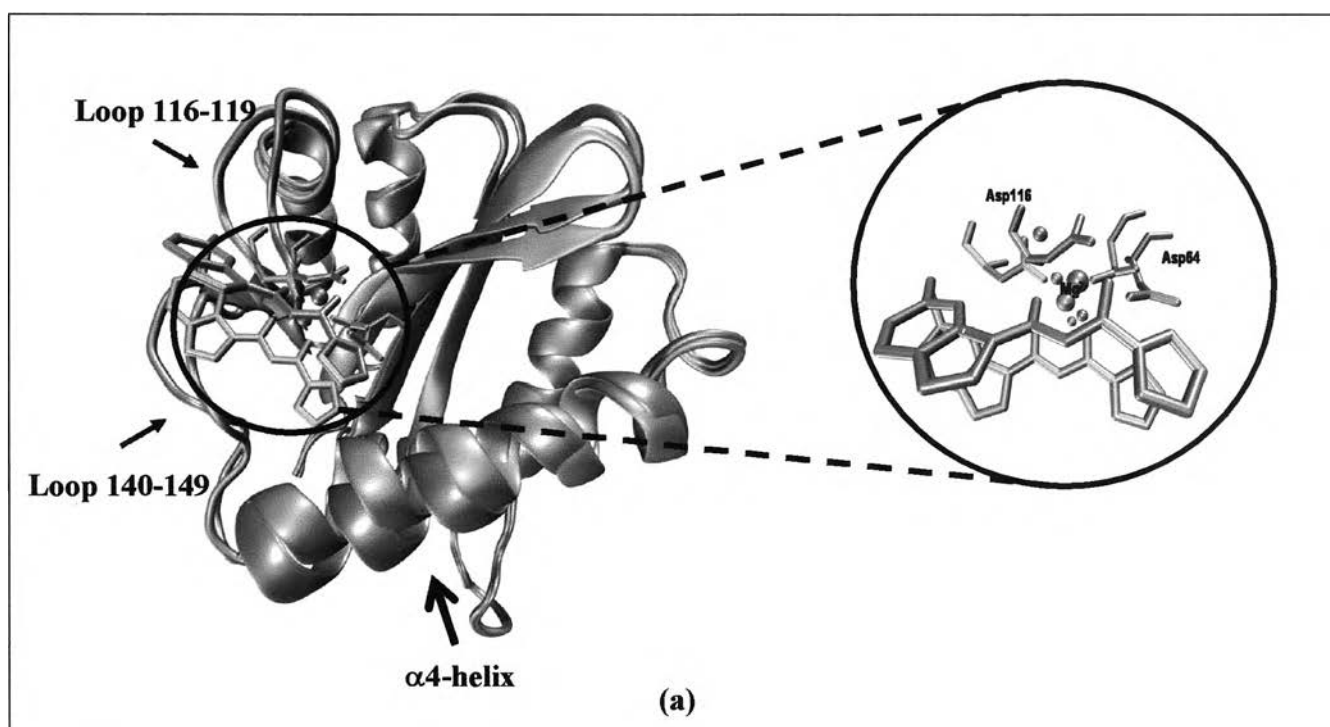
Figure 5.9 (a) Superimposition of the protein conformation of the average structure obtained from conventional MD (pink,▬▬▬) and combined QM/MM MD calculation (ice blue,▬▬▬▬). Residues Asp64, Asp116 and inhibitors are shown in stick models while Mg^{2+} ion and the catalytic water molecules are displayed in balls. (b) and (c) display hydrogen-bonds between amino acids located in 116-119 obtained from conventional and QM/MM MD simulations of HIV-1 IN-5CITEP complex, respectively.

We now turn our attention to the loop involving residues 140-149. The large flexibility of this loop is not surprising since it is located at the surface of the protein and is flanked by Gly residues. Moreover, the observed high mobility of this loop is in agreement with the previous MD simulations of HIV-1 IN [66, 67, 69] and with crystal structures (i.e., this loop is typically not resolved). It is worth mentioning that although loop 140-149 is very flexible, it does not undergo the same conformational changes in the two simulations (Figure 5.10). The MM results show larger fluctuations of this loop than that observed in the QM/MM calculation. The pink and blue ribbons represent the MM and QM/MM results, respectively, with key residues rendered as stick models. As can be seen, the different conformation of this particular flexible loop resulted in a variation of the conformation of the α4-helix of the region consisting of residues 148-166. Amino acids located in this α4-helix take on very different conformations which leads to

different interactions between the ligand and the protein, especially in salt bridge interactions, as discussed in the next section.
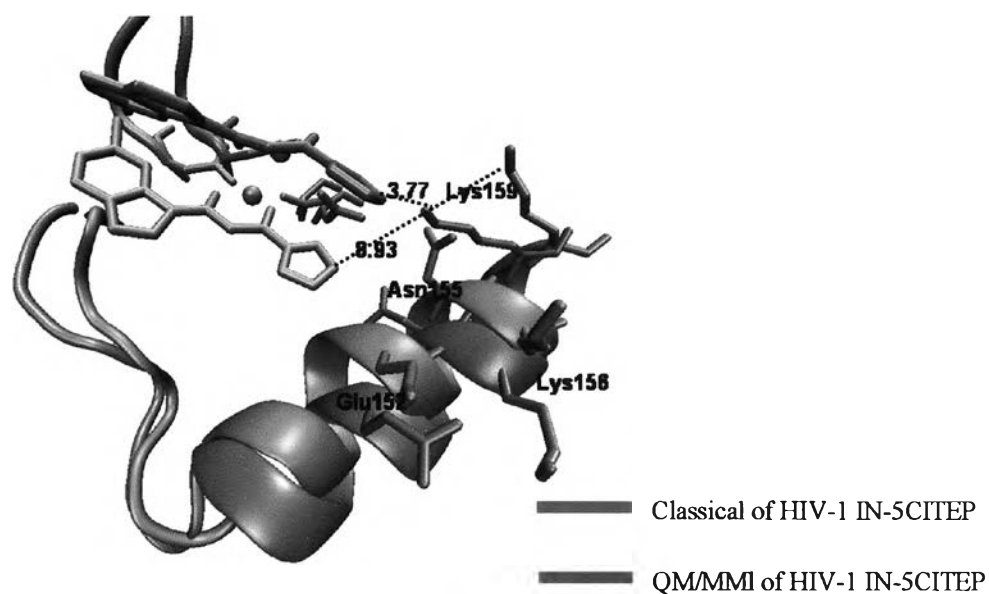

| | Classical of HIV-1 IN-5CITEP |
| | QM/MMl of HIV-1 IN-5CITEP |

Figure 5.10 Stereo views of the flexible loop region (residues 140-149) and the α4-helix. Selected residues and inhibitors are displayed in stick models. The distances between ligand and Lys159 are shown in dotted lines.

A.1.2 Mg^{2+} coordination and 5CITEP binding mode

One significant difference observed between the two levels of theory is the coordinating distances between Mg^{2+} and its chelated oxygen atoms. The key distances between the catalytic metal ion and the coordinated oxygen atoms were plotted as a function of simulation time, as illustrated in Figure 5.11. For the classical MD simulation, the Mg^{2+} was octahedrally coordinated by the two aspartate residues (Asp64 and Asp116) via their OD1 atoms, the oxygen atoms (O1 and O2) of the keto-enolate moiety of 5CITEP, and the two oxygen atoms of the catalytic water molecules (OW1 of water #1 and OW2 of water #2). The chelate distances between Mg^{2+} and the aforementioned

oxygen atoms fluctuate in the range of 1.8 - 2.0 Å (Figure 5.11 (a)) which are close to a perfect octahedral geometry. However, in the QM/MM simulation, the Mg^{2+} formed a distorted octahedral complex with the surrounding oxygen atoms. The distances between Mg^{2+} and four oxygen atoms (O1, O2 of 5CITEP, OD1---Asp116 and OW2) are approximately 1.8 Å while the other two (Mg^{2+}--OD1-Asp64 and Mg^{2+}--OW1) are ca. 2.5 Å, as depicted in Figure 5.11 (b). The different coordination geometry of Mg^{2+} produced by MM and QM/MM leads to different conformations of Asp116 and induces a different conformation of loop 116-119, as mentioned earlier.
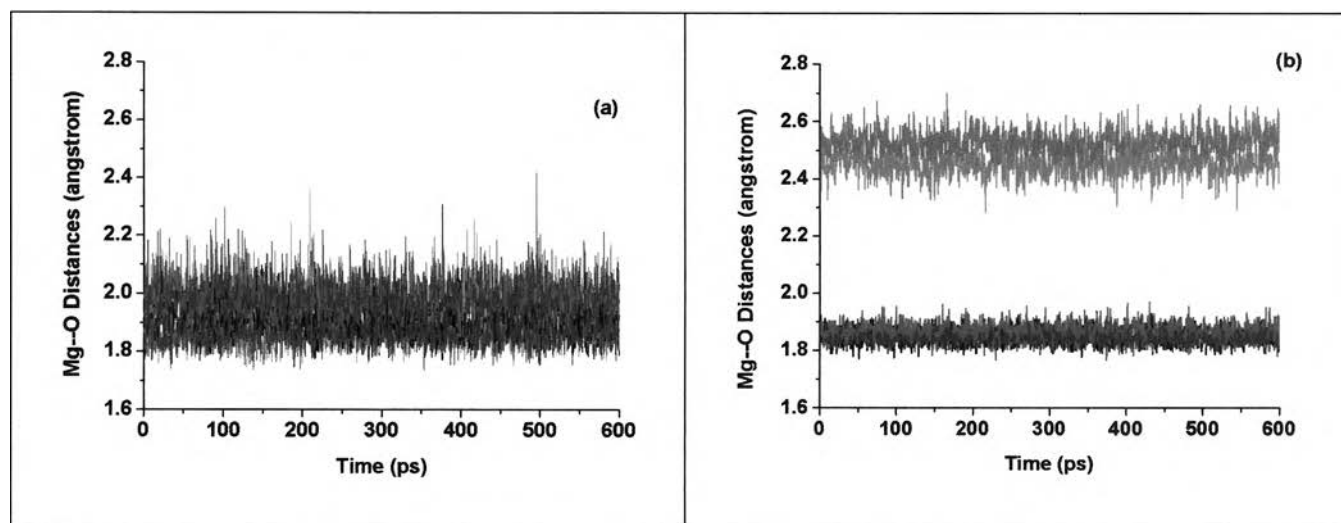

Figure 5.11 Time evolution of the inter-atomic distances between Mg^{2+} and its coordinating atoms obtained from (a) conventional MD of HIV-1 IN-5CITEP and (b) QM/MM of HIV-1 IN-5CITEP. The color coding is as follows: Mg^{2+}---O1, black, Mg^{2+}---O2, red, Mg^{2+}---OD1 (Asp64) green, Mg^{2+}---OD1 (Asp116), blue, Mg^{2+}---OW (water #1), cyan, Mg^{2+}---OW (water #2), pink.

We next turn to an analysis of inhibitor binding mode. Both MM and QM/MM results showed that 5CITEP occupied the active site region of HIV-1 IN, although with a somewhat different position. Generally, 5CITEP shared common features of binding in which the keto-enolate motif directly interacted with the Mg^{2+} ion with the tetrazolate ring pointed toward the region involving residues Asn155, Lys156 and Lys159 (Figure 5.10. The salt bridge between the ligand and the enzyme was monitored during the course of the MD trajectory. Surprisingly, no salt link was observed between the protein and 5CITEP in the classical MD simulation. This is because Lys159 is located far from 5CITEP (ca. 6.8 Å - 9 Å, Figure 5.12 (a)), hence, the tetrazolate motif of the ligand is not able to form a salt bridge with this particular residue. On the other hand, in the QM/MM trajectory, the ligand adopted an orientation allowing it to make this type of interaction with Lys159. This is supported by a plot of distance versus simulation time between the tetrazolate anion and the ammonium group of Lys159, as demonstrated in Figure 5.12 (b).
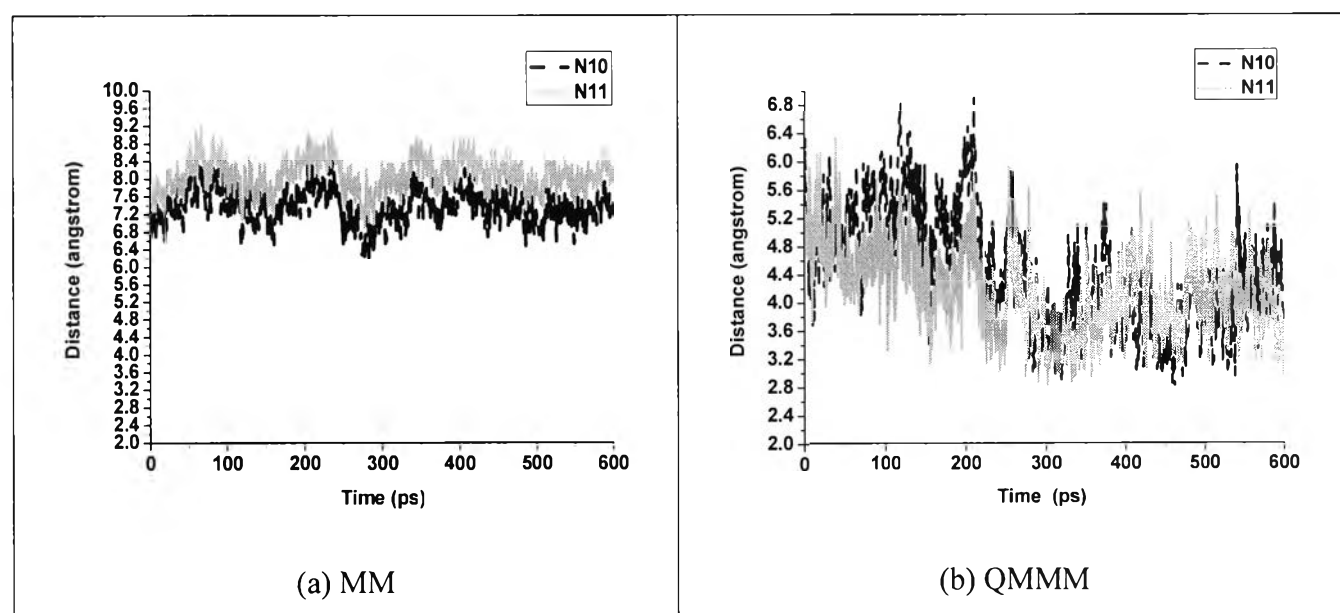

(a) MM   (b) QMMM

Figure 5.12 Plots of the salt link distances between the negatively charged group of ligand and the ammonium group of Lys59. The distance between tetrazolate ring of 5CITEP and the ammonium nitrogen of Lys159 taken from the conventional MD is shown in panel (a) while that from QM/MM is given in panel (b).

To further examine the mobility of 5CITEP, a detailed discussion is focused on the conformational changes of torsional angles *Tor1* (C3-C4-C5-C6) and *Tor2* (C6-C7-C8-N9) (defined in Figure 5.1). Remarkable differences were observed for the angle *Tor2* while no significant difference was found for the angle *Tor1* (Figure 5.13). For angle *Tor1*, both the MM and QM/MM simulations show only one sharp peak at about 0 degrees suggesting the single preferential conformation of the indole ring of 5CITEP. During the simulation period, the indole ring of the ligand always maintains a co-planar conformation relative to the keto-enolate motif. Additionally, the MM results indicated again a single preferential conformation of the tetrazolate ring of 5CITEP, as evidenced by the angle *Tor2* at about ±180 degrees. On the other hand, the QM/MM MD simulation produced two favorable conformations of the tetrazolate ring. The first conformation corresponds to *Tor2* at about -180 to -100 degrees and another one corresponds to 0 to 90 degrees. The broader and higher peak of *Tor2* at 0 to 90 degrees indicated higher mobility and higher probability of finding this conformation. The observation of double peaks in QM/MM might be due the smaller repulsion between the negatively charged group of 5CITEP and the surrounding amino acids such as Glu152 (tetrazolate---Glu152 ~9.0 Å). It is apparent that atomic charges are better distributed in QM/MM, so, it shows the smaller repulsion. In contrast, MM has more localized charges, hence, the repulsion between the tetrazolate anion and neighboring amino residues (tetrazolate---Glu152~ 4.5 Å) is stronger. The planar orientation presumably has the lowest repulsion. The tetrazolate ring therefore remains at this position for the whole period of simulation. This leads to the finding of a single sharp peak of *Tor2* in the classical MD simulation.
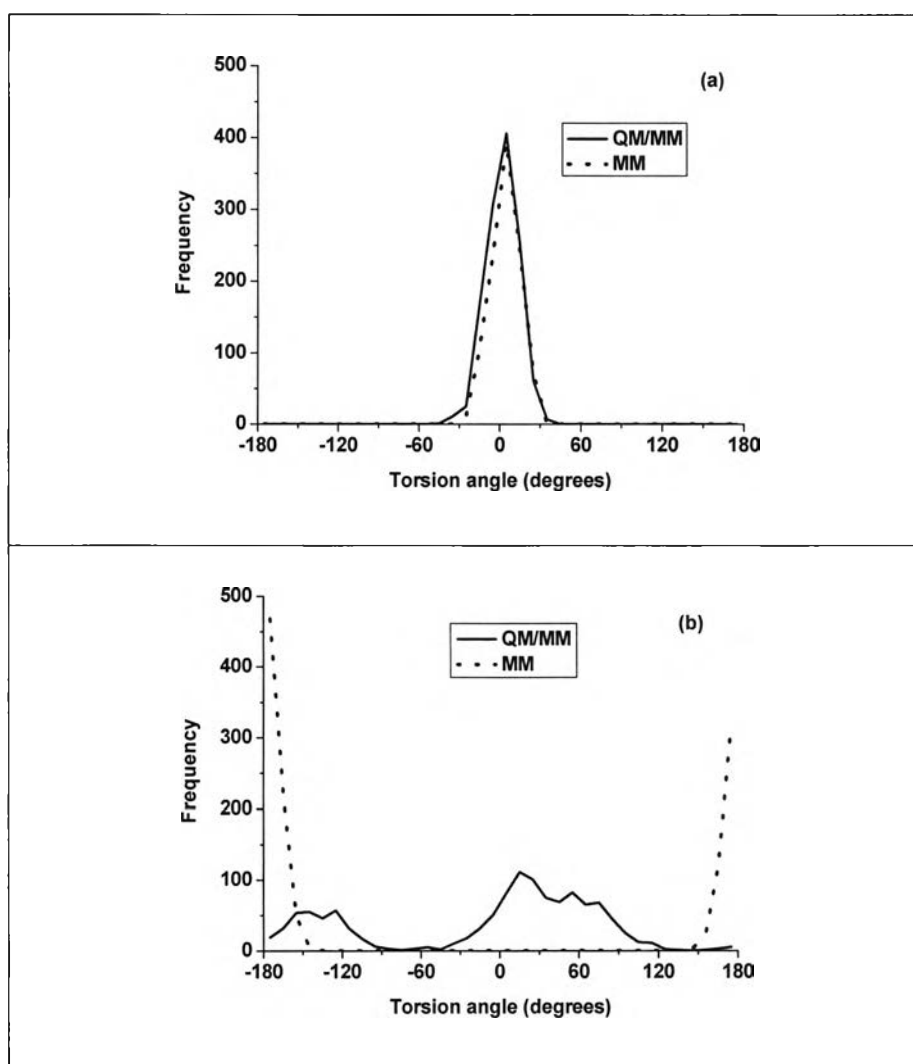
Figure 5.13 Comparison of the distributions of torsional angles (a) *Tor1* and (b) *Tor2* of 5CITEP using MM and QM/MM.

A.2 HIV-1 IN-DKA

A.2.1 Structural dynamics of HIV-1 IN-DKA

All heavy atoms RMSD relative the average structure of the simulated HIV-1 IN-DKA using classical or QM/MM is shown in Figure 5.14. The plot illustrates that structural fluctuations for each simulation are well behaved (RMSD ~ 0.6 Å) and consistent over the entire trajectory.
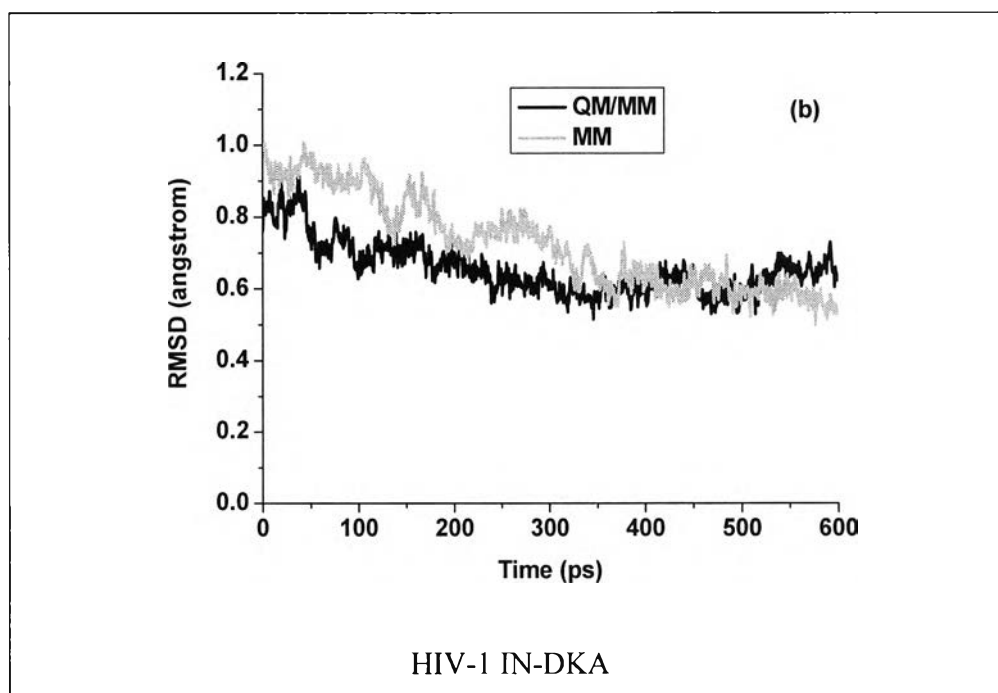
Figure 5.14 Time-dependent heavy atom RMSDs with respect to the average structure versus simulation time of HIV-1 IN-DKA which were obtained from conventional MD and QM/MM MD simulations.

The RMSF of backbone atoms of individual amino acids is depicted in Figure 5.15. The MM results show that the largest fluctuations were observed at residue Ala91 as well as the area involving residues 140-149, a catalytic loop. In classical MD simulation, there is only one intra-molecular hydrogen bond between NH and O of Ala91. However besides the intra hydrogen bonding interaction observed in QMMM simulation, O of Ala91 also formed inter-molecular hydrogen bond with NH of Thr93 and this leads to the rigidity of Ala91. For QM/MM MD simulation, it revealed that residues 116-119 and residues 140-149 were noticed to be flexible. In the former case, residues Asp116 and Gly118 in QM/MM MD simulation exhibit larger fluctuation than that observed in conventional MD simulation. This might be the reason that the percentage of hydrogen bonding interaction between O of Asp64 and NH of Gly118 found in QM/MM (30%) is less than that found in MM MD (89%). Moreover, it is worth

mentioning that the catalytic loop region between the two levels of simulations exhibit somewhat different conformation.
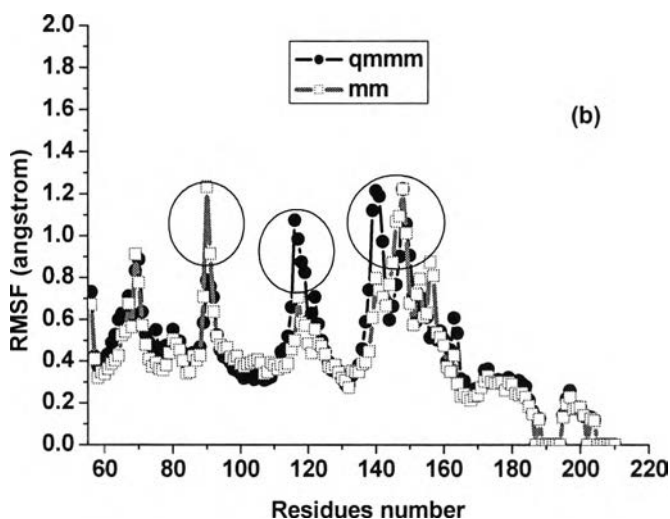

Figure 5.15 Average RMSF of protein backbone atoms as a function of residue number of IN-DKA. The protein structures were taken from classical and hybrid QM/MM MD simulations.

Figure 5.16 (a) shows the superimposition of the average structure of the IN-DKA complexes taken from conventional and QM/MM MD simulations (all heavy atoms RMSD ~ 1.36 Å). It is clear that three substantial differences between the two simulations are noticed. The first region is residues 64-69 (all heavy atoms RMSD = 1.47 Å) which is located in β1-sheet. In conventional MD simulation, the length of the loop starts from residues 60-68 while that of in QM/MM is shorter (residues 60-66). This is probably because in QM/MM, Asp64 adopted its oxygen atoms (OD1 and OD2) to bidentately coordinate with the Mg^{2+} ion and this affects the length of this β1-sheet. Among amino acids located in the area, His67 exhibits the most different orientation between the two simulations (Figure 5.16 (b)).

(a)

(b)
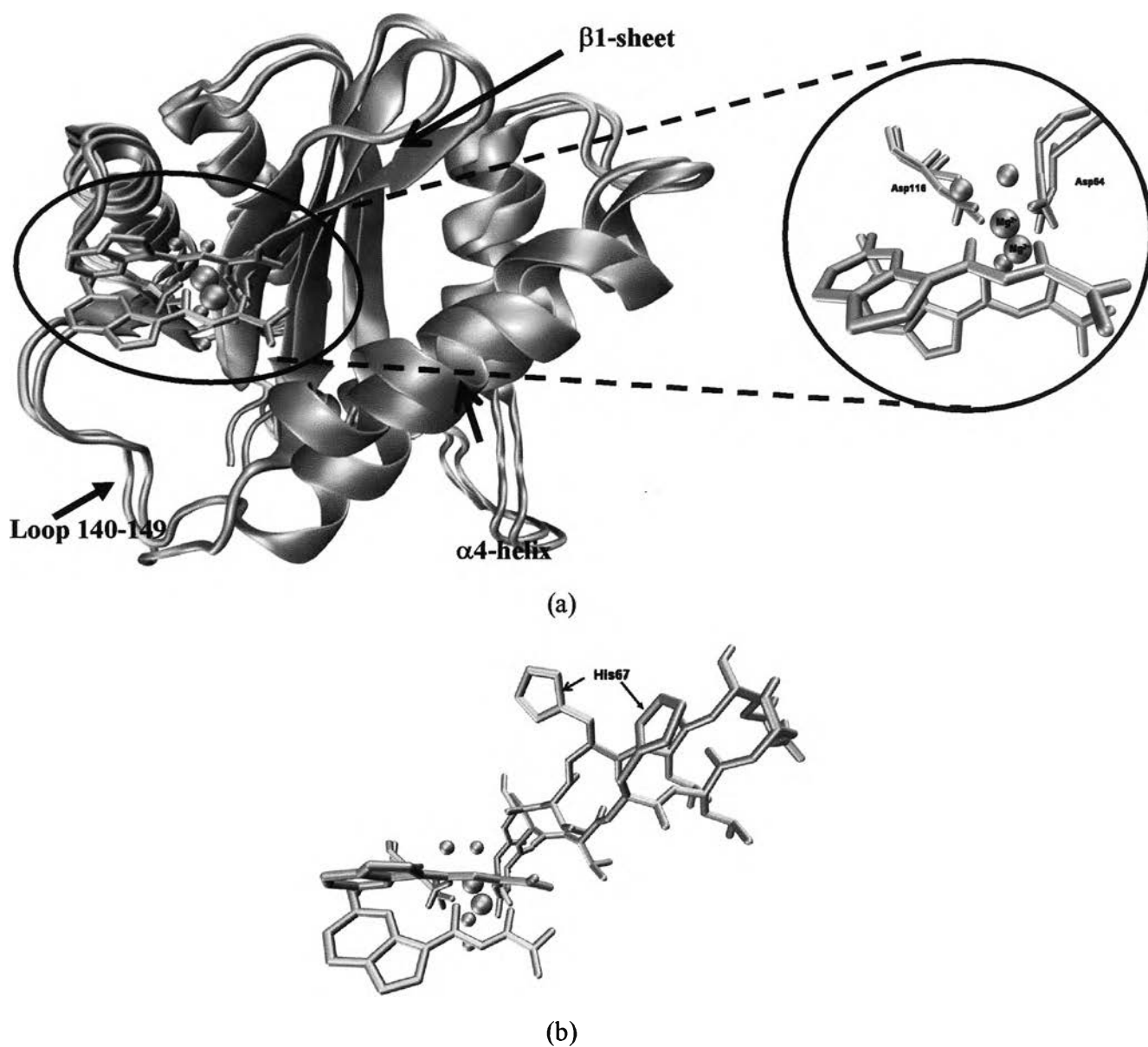
Figure 5.16 (a) Superimposition of the protein conformation of the average structure obtained from conventional MD (pink, ▀▀▀▀) and taken from a combined QM/MM MD calculation (cyan, ▀▀▀▀). Residues Asp64, Asp116 and inhibitors are shown as stick models while Mg²⁺ ion and the catalytic water molecules are displayed in balls. (b) displays the orientations of amino acid residues 64-69 obtained from conventional and QM/MM MD simulations of IN-DKA complex.

The next two different areas which were observed between the two levels of calculation are the catalytic loop residues 140-149 and residues 150-166, a α4-helix (Figure 5.17). The RMSD (all heavy atoms) of amino acid residues 140-149 and 150-166 of the average structures between MM and QM/MM simulations are 2.45 Å and 1.36 Å, respectively. As in the case of comparison between MM and QM/MM results of IN-5CITEP system, the former loop can be different conformation as it is flexible in nature and exhibits different conformation in different HIV-1 IN structure [135]. In the loop residues 140-149, we note that Gln146 (1.29 Å) displayed the most substantial difference between the two simulations. In conventional MD simulation, Gln146 points outward the active site of HIV-1 IN while it points toward the protein active site as noticed in hybrid QM/MM MD simulation. As aforementioned, the different conformation of this catalytic loop region leads to various conformations of the α4-helix. As a consequent, the orientations of amino acids in this helix are quite different which resulted in different salt link interaction between HIV-1 IN and ligand, further discussed in the next section. Moreover, it should be pointed out that the α4-helix observed in QM/MM starts from residues 149-166 which is shorter than that of conventional MD simulation (residues 150-166).
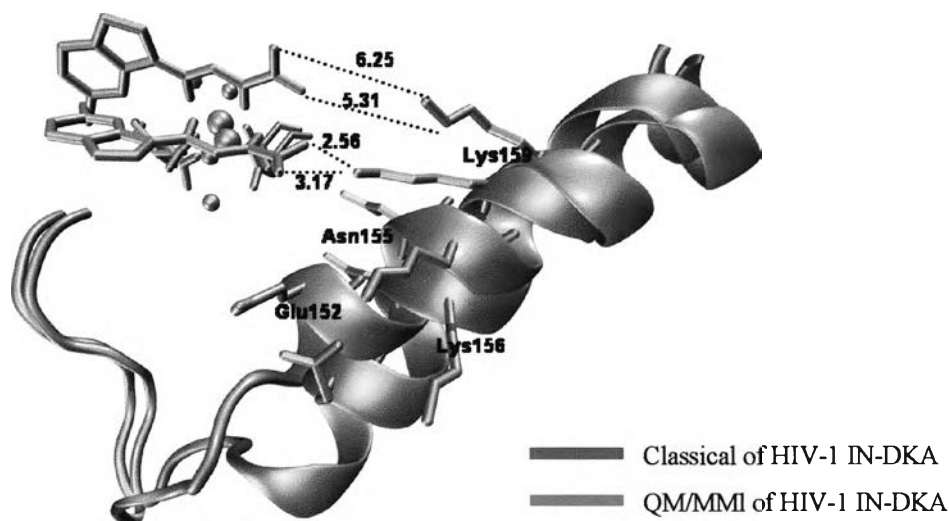

Figure 5.17 Stereo views of the flexible loop region (residues 140-149) and the α4-helix of HIV-1 IN-DKA. Selected residues and inhibitors are displayed in stick models. The distances between ligand and Lys159 are shown in dotted lines.

A.2.2 Mg^{2+} coordination and DKA binding mode

The distances between Mg^{2+} ion and the chelated atoms of the two calculations are given in Figure 5.18. The coordination distances of Mg^{2+} observed in conventional MD simulation of IN-DKA are similar to that of the classical MD simulation of IN-5CITEP. There are six oxygen atoms coordinated with the Mg^{2+} which is found to be a near perfect octahedral complex as indicated by the distances between Mg^{2+} and its surrounding oxygen atoms of about around 1.8 - 2.1 Å. Unlike conventional MD simulation, a distorted octahedral geometry of Mg^{2+} was observed in the combined QM/MM MD simulation. The Mg^{2+} coordinated with four oxygen atoms (O1, O2 of DKA, OD1-Asp64 and OW2) with the distances of about 1.8 Å while the remaining three (Mg^{2+}--OD2-Asp64, Mg^2+---OD1-Asp116, and Mg^{2+}--OW1) are about 2.5 Å, as illustrated in Figure 5.18 (b). In QM/MM, it is clear that the OD1 and OD2 of Aps64 form a bidentate coordination with the divalent metal ion.
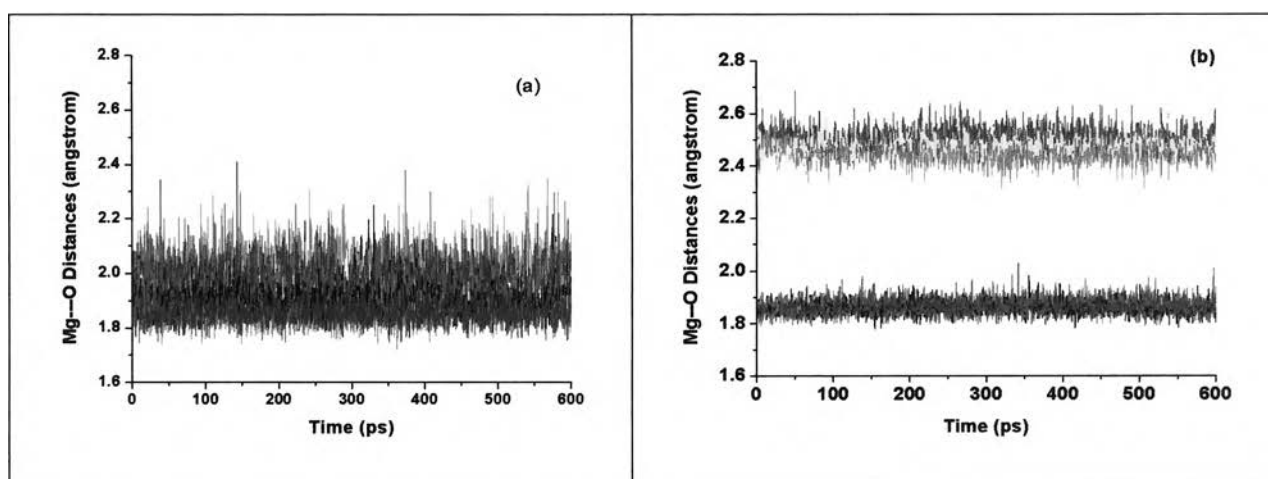

Figure 5.18 Time evolution of the inter-atomic distances between Mg^{2+} and its coordinating atoms obtained from (a) conventional MD of HIV-1 IN-DKA and (b) QM/MM of HIV-1 IN-DKA. The color coding is as follows: Mg^{2+}---O1, black, Mg^{2+}---O2, red, Mg^{2+}---OD1 (Asp64) green, Mg^{2+}---OD1 (Asp116), blue, Mg^{2+}---OW (water #1), cyan, Mg^{2+}---OW (water #2), pink.

We further investigate the inhibitor binding orientations. Although the positions of DKA found in both classical and hybrid QM/MM MD simulations are somewhat different, they occupied the same binding site area. Similar to the case of IN-5CITEP, the DKA of the two simulations shared a common binding pattern in which its keto-enolate moiety directly forms coordination with Mg^{2+} ion while the carboxylate group points toward the area containing residues Asn155, Lys156 and Lys159 (Figure 5.17). The salt link distances between oxygen atoms (O9 and O10) of DKA and the ammonium nitrogen (NZ) of Ly159 were detected and plotted in Figure 5.19. For conventional MD simulation, the salt link distances were stable at about 4.4 – 4.8 Å during the first 300 ps of the production run and after that (300-600 ps) these distances were increased (~ 5.2 - 7.0 Å for O9---NZ and ~ 6.0 -9.4 Å for O10---NZ). It is likely that the salt-link interaction between the negatively charged of carboxylate of DKA and the positively charged of ammonium group of Lys159 is become weaker in the last of simulation. This is possibly conformational changes of the α4-helix and the side chain of Lys159 observed in conventional MD simulation undergoes larger amplitude movements than that found in combined QM/MM MD.
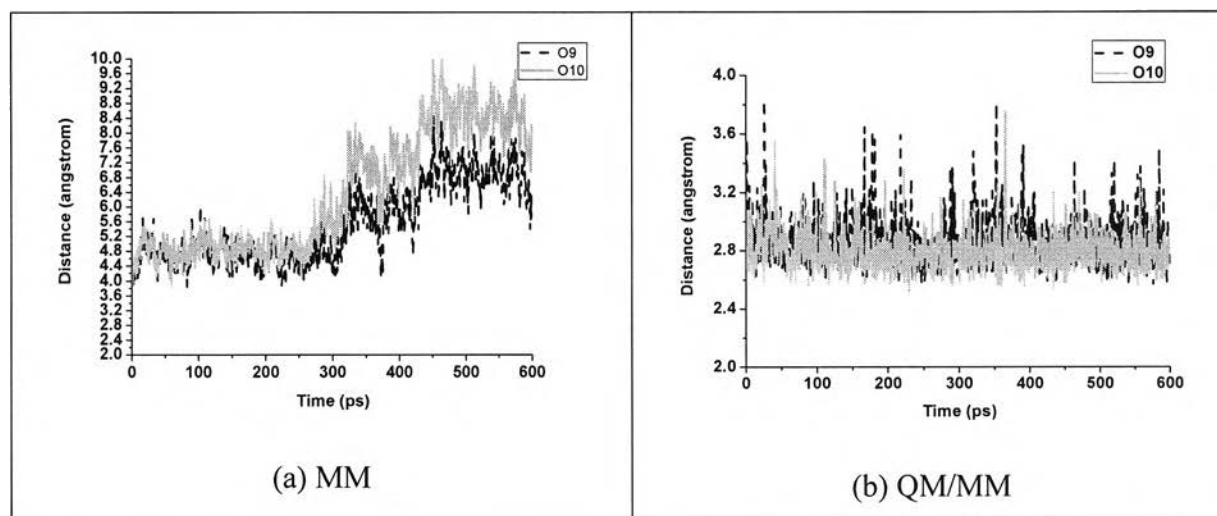

(a) MM

(b) QM/MM

Figure 5.19 Plot of salt link distances between the negatively charged group of DKA and the ammonium group of Lys59. The distance between carboxylate group of DKA and the ammonium nitrogen of Lys159 taken from the conventional MD is shown in panel (a) while that from QM/MM is given in panel (b).

The mobility of DKA was expressed by monitoring the changes of its two torsion angles, *Tor1* (C4-C4-C5-C6) and *Tor2* (C6-c7-C8-O9), defined in Figure 5.1. The distribution plots of both torsional angles are given in Figure 5.20. Similar to IN-5CITEP system, there is no substantial difference of *Tor1* between the two levels of simulations. This is supported by an existing of only one peak of *Tor1* at about 0 degree. The indole ring of DKA keeps co-planar conformations with the plan of its keto-enolate moiety. Noticeable difference between the two simulations is the angle *Tor2*. Although either MM or QM/MM produced a single preferential conformation of the carboxylate group of DKA, it does not the same conformation. The classical MD showed a peak of *Tor2* at about ±180 degrees while the hybrid QM/MM give a *Tor2* at about -120 degrees. It is likely that the carboxylate moiety of DKA observed in MM maintains a co-planar conformation relative to the plane of keto-enolate while this carboxylate group found in QM/MM adapted different conformations from the co-planar one. In QM/MM, the adopted conformation of carboxylate function makes it easier of DKA to keep a salt link interaction with the ammonium group of Lys159. On the other hand, the MM has more localized charges than QM/MM as stated before. Therefore, the co-planar conformation of DKA might give the lowest repulsion leading to an existing of this conformation of DKA for the entire course of simulation in the classical MD.
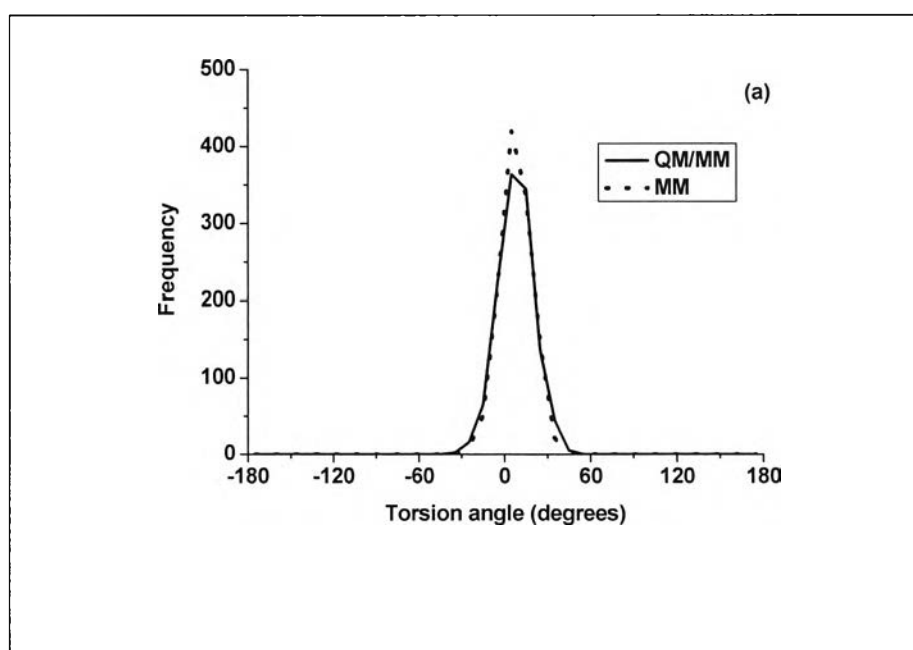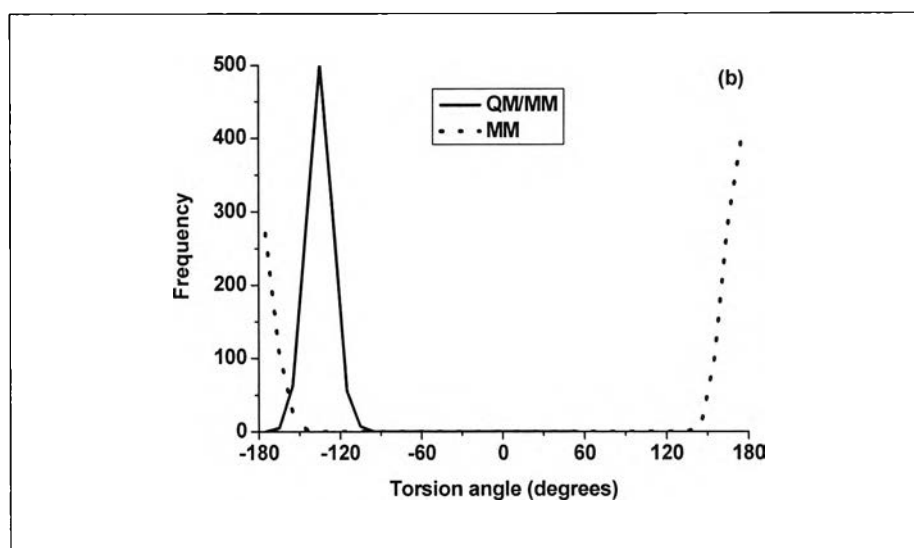
Figure 5.20 Distribution of torsional angles (a) *Tor1*, (b) *Tor2* of DKA using MM and QM/MM.

The comparisons between classical and hybrid QM/MM MD simulations of both HIV-1 IN-5CITEP and HIV-1 IN-DKA are clearly demonstrated somewhat different structural and dynamical behaviors of the protein-ligand complex. In classical MD simulations, Mg^{2+} coordination geometry is forced to be a perfect octahedral whereas this geometry can distort in QM/MM MD simulation. It is apparent that the classical MD simulations cannot explain the difference of the inhibitory potency between 5CITEP and DKA. Moreover as aforementioned in Chapter II, the mixed QM/MM studies of protein-ligand system were reported to exhibit advantages over classical force fields [150, 151] in such a way that QM/MM yields the results much more close to experimental data than those of MM force fields. Thus in this work, combined QM/MM approach was used to explain the inhibitory potency difference between 5CITEP and DKA as discussed in depth below.

B. Comparison of QM/MM results between HIV-1 IN-5CITEP and HIV-1 IN-DKA

B.1 Stability of MD trajectories

The stability of the MD trajectories of the two systems was evaluated by plotting of RMSD values to their average structures (Figure 5.21). The mean RMSD value of both systems exhibits convergence to about 0.6 Å. The steady and small magnitude of RMSD fluctuation suggests that the overall structure of the protein remained stable over the course of the MD simulations. Therefore, our MD trajectories should provide a suitable basis for more detailed analyses.
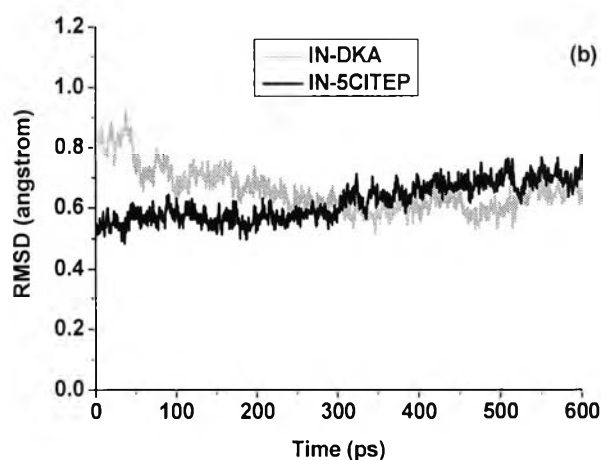
Figure 5.21 Heavy atom RMSDs between IN-DKA and IN-5CITEP taken from a hybrid QM/MM MD simulations.

B. 2 Residue fluctuation

For each system, a detailed analysis of RMSF of protein backbone atoms for each residue is illustrated in Figure 5.22. The remarkable differences between IN-DKA and IN-5CITEP were again observed in residues 116-119, which are located in the 3'-processing cavity, and the loop residues 140-149. In the former loop, Asp116 and Asn117

of the HIV-1 IN-DKA complex undergo large amplitude movement during the trajectories, while this motion is small in the HIV-1 IN-5CITEP complex. Regarding loop residues 140-149, although it was found to be very flexible in the two systems, this catalytic loop region exhibited different conformational changes in the two complexes. The HIV-1 IN-DKA complex exhibited a greater fluctuation of residue Gln148 compared to that in the HIV-1 IN-5CITEP system.
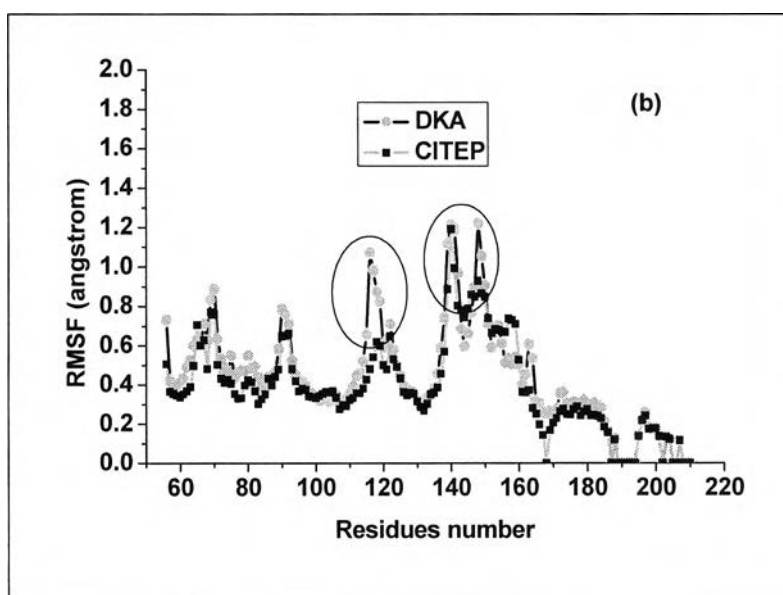
Figure 5.22 RMSF of protein backbones of HIV-1 IN-DKA and HIV-1 IN-5CITEP complexes obtained from QM/MM MD simulations.

B.3 Structural difference between HIV-1 IN-DKA and HIV-1 IN-5CITEP

The superimposition of the average structure between HIV-1 IN-DKA and HIV-1 IN-5CITEP is given in Figure 5.23. The RMSD of heavy atoms in the protein between the two complexes is 1.34 Å suggesting a significant difference of the protein structure. As can be seen in Figure 5.23, three apparent differences in the HIV-1 IN structure between the HIV-1 IN-DKA and HIV-1 IN-5CITEP complexes are noticeable around regions formed by residues 60-68, 116-119 and 140-149. In the HIV-1 IN-DKA complex, the side chains of both Asp64 and Asp116 are almost perpendicular to the plane defined by the keto-enolate motif of DKA whereas these two residues are mostly horizontal to the

plane of the ligand in HIV-1 IN-5CITEP. In HIV-1 IN-DKA, the adapted side chain of Asp64 allows its OD1 and OD2 atoms to coordinate the divalent metal ion (see Figure 5.23 (b)). Unlike HIV-1 IN-DKA, in HIV-1 IN-5CITEP, only the OD1 atom of Asp64 chelates with the Mg^{2+} ion (see Figure 5.23 (c)). The conformational difference of the Asp64 side chain in the two complexes affects the length of the β1-sheet, which consists of seven residues (60-66) in IN-DKA and nine residues (60-68) in IN-5CITEP.
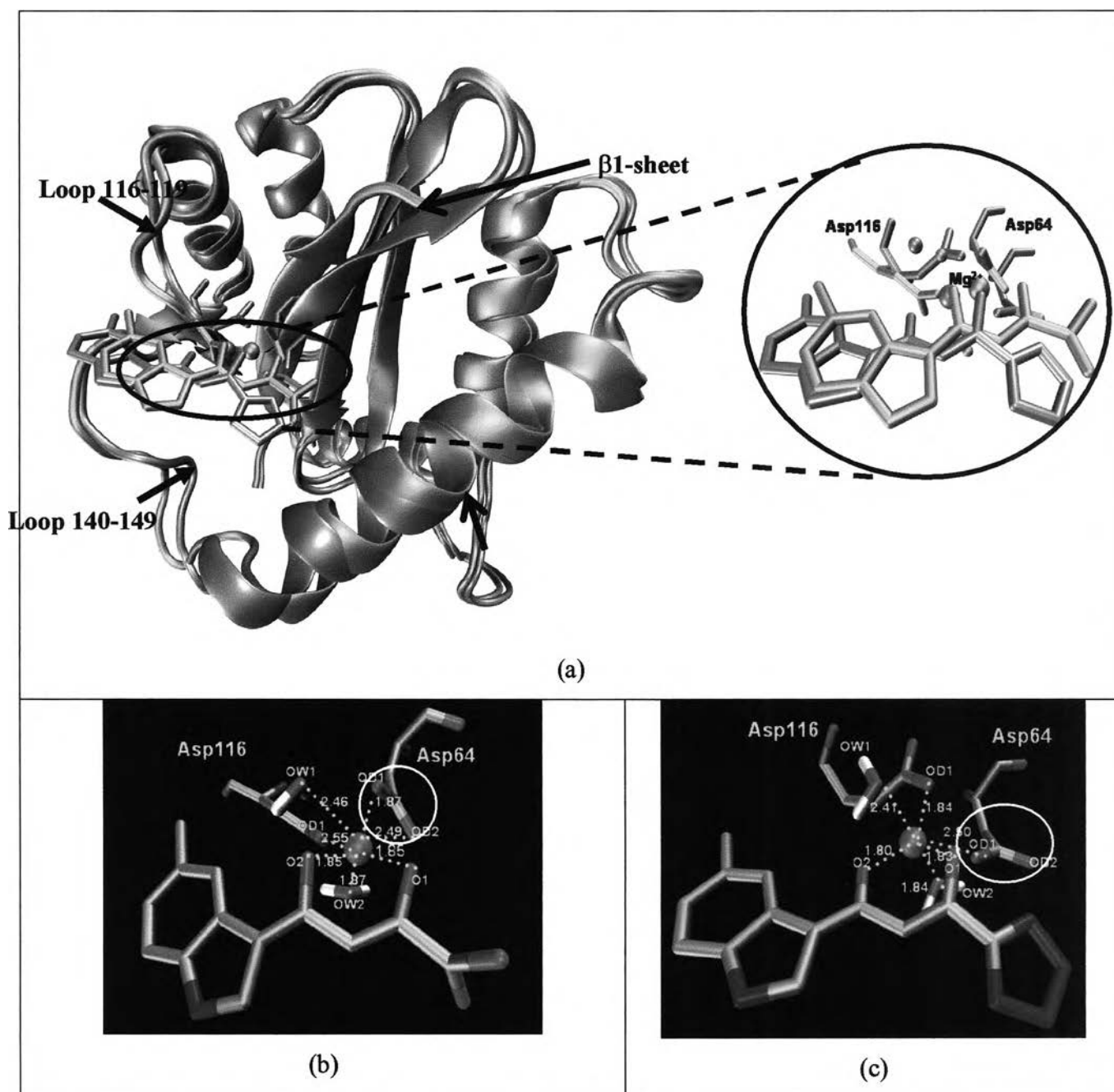
Figure 5.23 (a) Superimposition of average structure obtained from QM/MM MD simulations between HIV-1 IN-DKA (cyan, ▬▬) and HIV-1 IN-5CITEP (ice blue, ▬▬). Binding modes of DKA (b) and 5CITEP (c) in the HIV-1 IN active site. Mg^{2+} ion is displayed as a green ball. The coordinating distances between Mg^{2+} and its chelating atoms are represented in dotted lines.

In addition, Asp116 in the IN-DKA complex adapted its side chain conformation to allow a hydrogen-bonding network to form between their OD1 and OD2 atoms and OW2 of water #2. In contrast, in the IN-5CITEP complex only the OD1 of Asp116 was hydrogen-bonded to OW2. Thus, the HIV-1 IN-DKA complex was stabilized by a larger number of hydrogen-bonds with water than those in the IN-5CITEP system. Furthermore, the conformational difference of Asp116 results in conformational changes in residues 117-119 in the IN-DKA complex.

The final substantial difference was detected in the catalytic loop region involving residues 140-149. In fact, this loop is very flexible in nature (residues 140 and 149 are glycines) and different crystal structures of HIV-1 IN exhibit various conformations of this particular loop, and in many cases, absence of coordinates as it as too flexible. The heavy atom RMSD value of amino acids 140-149 between the average structure of IN-DKA and IN-5CITEP was 3.08 Å indicating a large difference of the catalytic loop between the two complexes, especially Gln148 (RMSD = 0.71 Å). It has been hypothesized that Gln148 is involved in the position of viral DNA binding before strand transfer into host DNA thus, the behavior of this residue might be crucial for DNA binding. For IN-5CITEP, the rigidity of Gln148 occurs because the O of Gln148 is capable of forming a hydrogen-bond with the NH of Gln146 (20%) and with the NH of Glu152 (80%) more than that found in IN-DKA (O (Gln148)---NH (Glu152) = 70%). This hydrogen bond formation was also observed in the previously reported conventional MD simulations. One remarkable observation between the two complexes is the length of the α4-helix residues which probably arose from the conformational change of Gln148. This helix in IN-DKA spans residues 149-166 which is one residue shorter than that in IN-5CITEP (residues 148-166). In the X-ray structure (1QS4) as well as in the wild-type model of the conventional MD simulation of the X-ray complex structure of HIV-1 IN-5CITEP, this helix involves residues 150-166 and 149-165, respectively. Moreover, the instability of this helical behavior was also observed in the conventional MD simulation of the apo HIV-1 IN catalytic core domain containing one Mg^{2+}.

B.4 Inhibitor binding mode/structural features of the inhibitor

It is clear from Figure 5.23 (a) that both DKA and 5CITEP occupied nearly the same region in the active site of HIV-1 IN and they showed a similar pattern of inhibitor binding orientation. For each system, the keto-enolate moiety of the ligand chelated the Mg^{2+} ion whereas the chloroindole ring was located close to residues Thr115, Asn117 and Gly118. An electrostatic interaction was also observed between the chlorine atom in the indole ring and the amide group of Asn117. The NH of the indole ring of the two ligands is exposed to solvent and does not make contact with any residue. Furthermore, the carboxylate group of DKA was pointed toward Lys159 while 5CITEP oriented its tetrazolate anion toward this residue. The salt bridges were observed between the ammonium group of Lys159 and the carboxylate (DKA) or the tetrazolate ring (5CITEP) which will be discussed in salt bridge analysis.

To further investigate the structural changes of ligands, the torsional angles *Tor1* (C3-C4-C5-C6) and *Tor2* (C6-C7-C8-O9, C6-C7-C8-N9), denoted in Figure 5.1, of DKA and 5CITEP were statistically analyzed as a population frequency, (Figure 5.24). The *Tor1* distribution peaks of the DKA and 5CITEP mostly overlap while the *Tor2* angles do not. Both DKA and 5CITEP produced a single preferential co-planar conformation of indole ring, as evidenced by the only one peak for *Tor1* at about 0 degrees (Figure 5.24 (a)). For the torsional angle *Tor2*, a single preferential conformation of the carboxylate moiety of DKA was represented by a single sharp peak at about -120 degrees (Figure 5.24 (b)). This sharp and pronounced peak implied a low flexibility of the carboxylate functional group in a narrow range. Meanwhile, 5CITEP yielded two preferential conformations of the tetrazole ring, as mentioned in **A.2.2** i.e. two broad peaks at *Tor2* = -180 to -100 degrees and at 0 to 90 degrees. At *Tor2* = -180 to -100 degrees, the N10 of the tetrazole ring was close to Lys159 while at *Tor2* = 0 to 60 degrees, the N11 was close to the same residue. This suggests that the tetrazolate ring of 5CITEP might alternate between two conformations in the binding pocket and also supports the finding that either N10 or N11 of 5CITEP can make a salt link with Lys159. The higher peak at *Tor2* = 0 to 90 degrees indicates a higher probability of finding the N11 in a salt bridge with Lys159 as compared to N10. According to the observations above, it is worth noting that the rigidity of *Tor2* in DKA is probably due to the existence of a stronger salt bridge to

Lys159 for the entire simulation. In contrast, the tetrazolate moiety of 5CITEP formed a weaker salt link with Lys159.
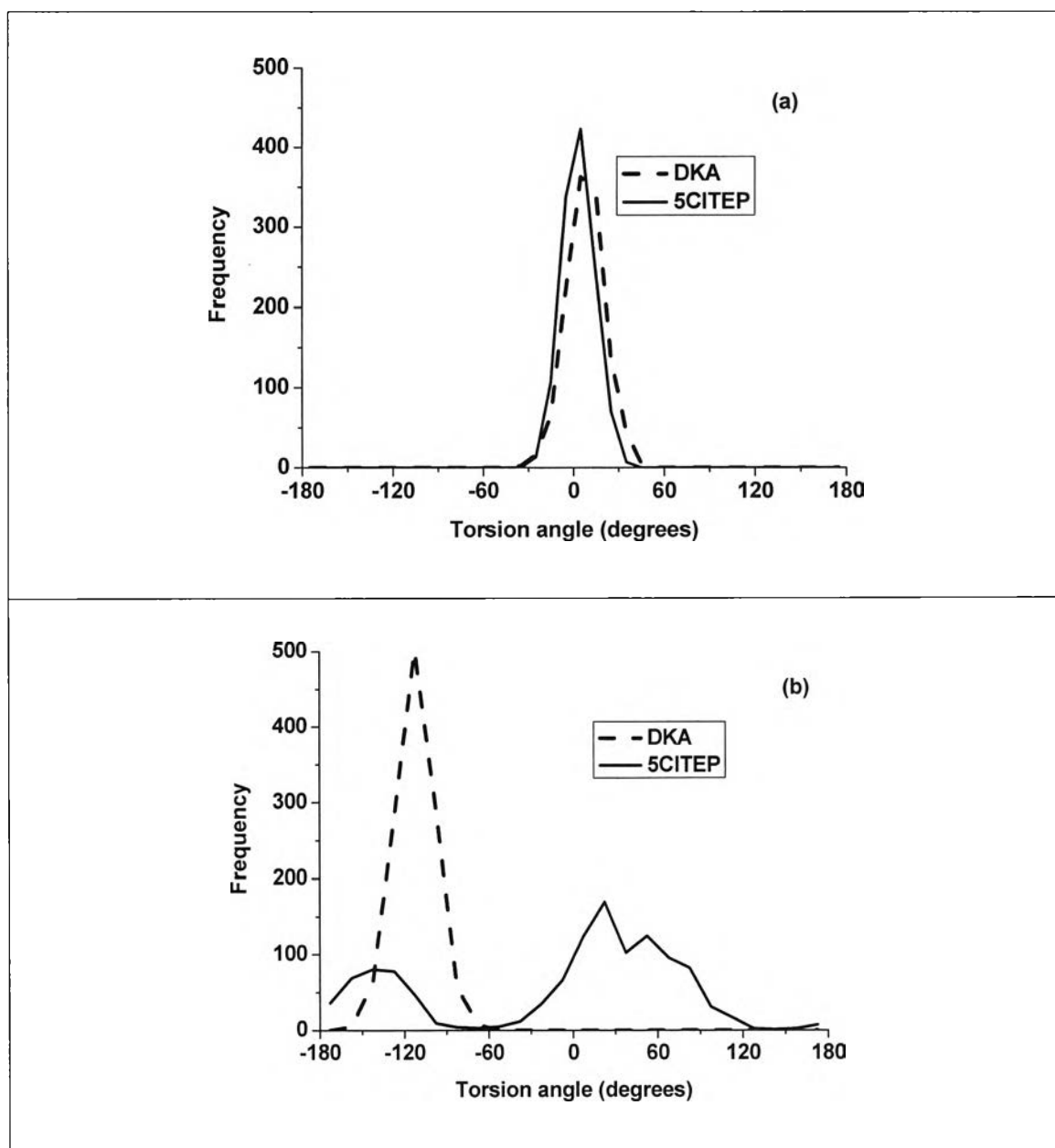

Figure 5.24 Comparison of the distributions of torsional angles (a) *Tor1* and (b) *Tor2* of both DKA and 5CITEP taken from QM/MM MD simulation.

B.5 Mg^{2+} coordination

To explore the dynamical properties of the divalent cation and its coordination state, the inter-atomic distances between Mg^{2+} and the chelating oxygen atoms are plotted as a function of simulation time. For IN-DKA (Figure 5.18 (b)), there are seven oxygen atoms (OD1-Asp64, OD2-Asp64, OD1-Asp116, OW1, OW2, O1 and O2 of DKA) whose distances are able to coordinate with the metal ion. The distances between Mg^{2+} and OD1-Asp64, OW2, O1 and O2 were observed at about 1.8 Å while the other three (Mg^{2+}--OD2-Asp64, Mg^{2+}---OD1-Asp116 and Mg^{2+}---OW1) were found to be 2.5 Å (Figure 5.18 (b)). These coordination distances were stable along the whole trajectory. It appears that the Mg^{2+} bound tightly with the OD1-Asp64, OW2, O1 and O2 of DKA, as indicated by their shorter distances (1.8 Å). The Mg^{2+} forms a distorted octahedral complex in which the carboxylate group of Asp64 is twisted toward the catalytic metal ion, allowing both OD1 and OD2 to chelate the Mg^{2+} in a monodentate configuration (Fig. 5.23 (b)).

For IN-5CITEP (Figure 5.11 (b)), Mg^{2+} yields a somewhat similar pattern of Mg^{2+} coordination state as found in IN-DKA. As mentioned above (section A), Mg^{2+} is chelated by six oxygen atoms (O1-Asp64, OD1-Asp116, OW1, OW2, O1 and O2 of 5CITEP). In IN-5CITEP, Asp64 uses only its OD1 atom to chelate the Mg^{2+} (Mg^{2+}---OD2-Asp64 = 4.5 Å) while both OD1 and OD2 were seen to coordinate the divalent cation in IN-DKA. The coordination distances were 1.8 Å (Mg^{2+}---OD1-Asp116, Mg^{2+}---OW2, Mg^{2+}---O1 and Mg^{2+}---O2) and ca. 2.5 Å (Mg^{2+}---OD1-Asp64 and Mg^{2+}---OW1) (Figure 5.11 (b)). In this system, the Mg^{2+} is tightly bound with the OD1 atom of Asp116 rather than OD1 of Asp64 (found in IN-DKA). This distorted octahedral geometry of Mg^{2+} coordination persists for the entire course of the simulation.

It is worth noting that besides its catalytic role, the Mg^{2+} also plays a structural role. In the HIV-1 IN active site, two Asp residues (Asp64 and Asp116) and the keto-enolate motif of the inhibitor are close to each other and form a carboxylate cluster, thereby generating an excess of negative charge in the binding pocket. The positive charge of Mg^{2+} is necessary to stabilize the carboxylate cluster by diminishing the electrostatic repulsion between these oxygen atoms.

B.6 Salt bridge analysis

Ionic interactions play a critical role in biological systems, therefore, a detailed analysis of salt bridge formation between HIV-1 IN and ligands was carried out for the dynamics trajectories. The distances between the negatively charged group (carboxylate or tetrazolate) of ligand and the ammonium group of Lys159 as a function of simulation time were plotted in Figure 5.12 (IN-5CITEP) and 5.19 (IN-DKA), respectively. In IN-DKA, the average distances between the negatively charged group of the ligand (O9 and O10) and the ammonium nitrogen (NZ) of Lys159 is ca. 2.8 ± 0.17 Å (Figure 5.19 (b)) implying that this strong salt bridge finding was observed throughout the simulation period. In contrast, there is a weaker ionic interaction between the tetrazolate ring of 5CITEP and the Lys159 during the first 200 ps of the simulation (Figure 5.12 (b)). Within this period, the average distances between the tetrazolate ring and Lys159 (NZ) are longer than ca. 4 Å (N10---NZ = 5.21 ± 0.57 and N11---NZ = 4.59 ± 0.52). For the remainder of the simulation, the tetrazolate ring of 5CITEP moved closer to Lys59 (N10---NZ = 4.07 ± 0.69 and N11---NZ = 3.97 ± 0.57), thus, the ionic interaction becomes stronger as compared to that found in the early stage of the production phase. It should be mentioned that although an ionic interaction between 5CITEP and Lys159 can occur, it is weaker than those observed in the IN-DKA complex.

To clearly demonstrate the dynamic nature of the salt links between HIV-1 IN and the two inhibitors, snapshots were taken from trajectories of IN-DKA and IN-5CITEP complexes and are given in Figure 5.25. In both complexes, only residue Lys159 was found to form a salt link with ligands. In IN-DKA, the ammonium group of Lys159 interacts with the carboxylate moiety of DKA (Figure 5.25 (a)) while that of the residue interacts with the tetrazolate ring as found in IN-5CITEP (Figure 5.25 (b)). A hydrogen-bond was observed between bulk water and the NH of indole ring of inhibitors in the two systems. In addition, the ligand also formed hydrogen bonds with the bridging water in the active site of HIV-1 IN, as discussed in the next section. It is worth mentioning that the strong salt link between Asp64 and Lys159 was detected in the IN-5CITEP complex (Figure 5.25 (b)). Meanwhile, the salt link between these two amino acids was not observed in the IN-DKA complex because of the long distance (longer than 6.5 Å) between these two residues.

According to the existence of a strong salt link between DKA and HIV-1 IN throughout the simulation, ionic interaction seems to be a major reason why DKA is a more potent inhibitor than 5CITEP.

(a)

(b)
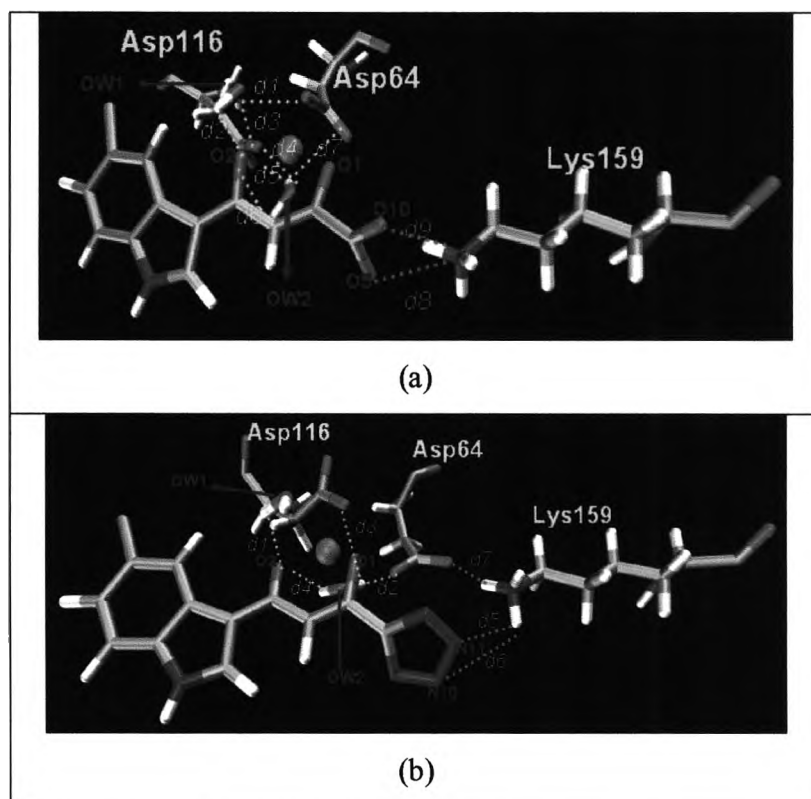
IN-DKA			IN-5CITEP	
d1 = OH (W1)—OD1 (Asp64)	2.39		d1 = OH (W1)—O2	1.80
d2= OH (W1)—O2	1.68		d2 = OH (W2)—OD1 (Asp64)	1.72
d3= OH (W1)—OD1 (Asp116)	1.85		d3 = OH (W2)—OD1 (Asp116)	3.04
d4= OH (W2)—OD1 (Asp116)	1.75		d4 = OH (W2)—O2	2.65
d5= OH (W2)—OD2 (Asp116)	2.55		d5 = N10—HZNZ (Lys159)	3.56
d6= OH (W2)—O2	3.13		d6 = N11—HZNZ (Lys159)	2.92
d7 = OH (W2)—OD2 (Asp64)	2.84		d7 = OD1 (Asp64)—HZNZ (Lys159)	1.74
d8 = O9—HZNZ (Lys159)	1.91			
d9 = O10—HZNZ (Lys159)	1.95			

Figure 5.25 Snapshots of the active site of HIV-1 IN-DKA complex (a) and HIV-1 IN-5CITEP complex (b). Hydrogen-bonds and salt bridges are indicated by dotted lines. All distances are in Angstrom unit.

B.7 Role of water molecules

It is well known that water molecules are important in biological systems, including in catalysis and in mediating ligand binding. The X-ray structure of the HIV-1 IN-5CITEP complex [39] revealed that four interfacial water molecules were located close to Asp64, Asp116 and 5CITEP. These water molecules as well as the side chain oxygen atoms of both aspartic residues were found to octahedrally coordinate the Mg^{2+} ion. The importance of these water molecules in HIV-1 IN is uncertain. Based on an MD simulation study [68] of the X-ray structure of the HIV-1 IN-5CITEP complex, it was shown that water molecules present in the active site may be crucial for bridging between the NH of indole ring of ligand and carbonyl oxygen of Asp64 and OD of Asn155, through hydrogen-bonding interactions.

In the present study, our attention was focused on the two "catalytic" water molecules (water #1 and water #2) coordinated to the Mg^{2+}. The positions of both water molecules were taken from the X-ray co-crystal structure of HIV-1 IN/5CITEP (1QS4) i.e. water #226 and water #443. To study the role of these water molecules, the hydrogen-bond bridges between these water and protein as well as ligand were examined along the MD trajectories. The two systems exhibited a somewhat similar pattern of bridging water molecules between Asp64, Asp116 and the keto-enolate motif of the ligands (Figure 5.22). These coordinated water molecules did not significantly move and did not exchange with the bulk water. The hydrogen-bond bridges formed by water molecules between the ligand and aspartic acids can therefore stabilize the binding between inhibitors and HIV-1 IN. It should be pointed out that a number of hydrogen-bonds between the bridging water molecules and the protein, as well as the ligand in IN-DKA, were larger that those seem in the IN-5CITEP complex. In IN-DKA, both OD1 and OD2 atoms of Asp64 and Asp116 were involved in a hydrogen-bond network with waters while only the OD1 atoms of these two residues were hydrogen-bonded to the bridging water in IN-5CITEP. Besides the bridging between protein and ligand, these two water molecules sometimes mediated hydrogen-bond bridges between the Asp64 and Asp116.

B.8 Residue contributions

To elucidate the role of individual MM residues in stabilizing or destabilizing the QM subsystem, an amino acid decomposition analysis [152, 153] was carried out as described below. The QM/MM interaction energy, which is consisted of electrostatic and of van der Waals terms, of each frame was computed. Afterward, each MM residue was removed, one by one, and the QM/MM interaction was recalculated. The contribution of each individual amino acid is defined in terms of the energy difference resulting from the residue deletion. Although, the average contributions were examined from a set of 120 snapshots (at every 5 ps from 600-1200 ps), the results should qualitatively estimate which amino acids significantly stabilize (positive) or destabilize (negative) the QM subsystem.

Figure 5.26 shows the contributions of each MM residue in stabilizing or destabilizing the QM part of the IN-DKA and IN-5CITEP complexes. The standard deviations of each residue contribution are given in Table 5.7. It should be pointed out that the contributions of the stabilizing amino acids are much more important than those of the destabilizing ones. For both complexes, among all residues, Lys159 exhibited the greatest contribution in stabilizing the QM subsystem, as indicated by it high interaction energy (ca. 180 kcal/mol for both systems). This stabilizing effect is mainly due to a strong ionic interaction between the QM part and Lys159, i.e. the negatively charged group of ligand (carboxylate or tetrazolate) and the positively charged of ammonium group of Lys159. The remarkable differences between the two complexes are the roles of Gln62 and Leu63. Gln62 neither stabilizes nor destabilizes the QM part in IN-DKA complex however in IN-5CITEP, it plays a stabilizing role. Moreover, Leu63 provides a major destabilizing effect on the IN-DKA complex while it yields a stabilizing effect on the IN-5CITEP complex. This arises from the differential ability for forming hydrogen-bonds between the O of Leu63 and the NH of Asp116 between the two complexes. Furthermore, the QM part of IN-DKA is stabilized by Thr66 with a higher magnitude than that observed in IN-5CITEP. Residues Thr115, Asn117 and Phe121 play a critical role in stabilizing the QM subsystem in both complexes. The roles of Cys65, Gly118 and Asn155 in terms of stabilizing or destabilizing are unclear due to the high values of their standard deviations.
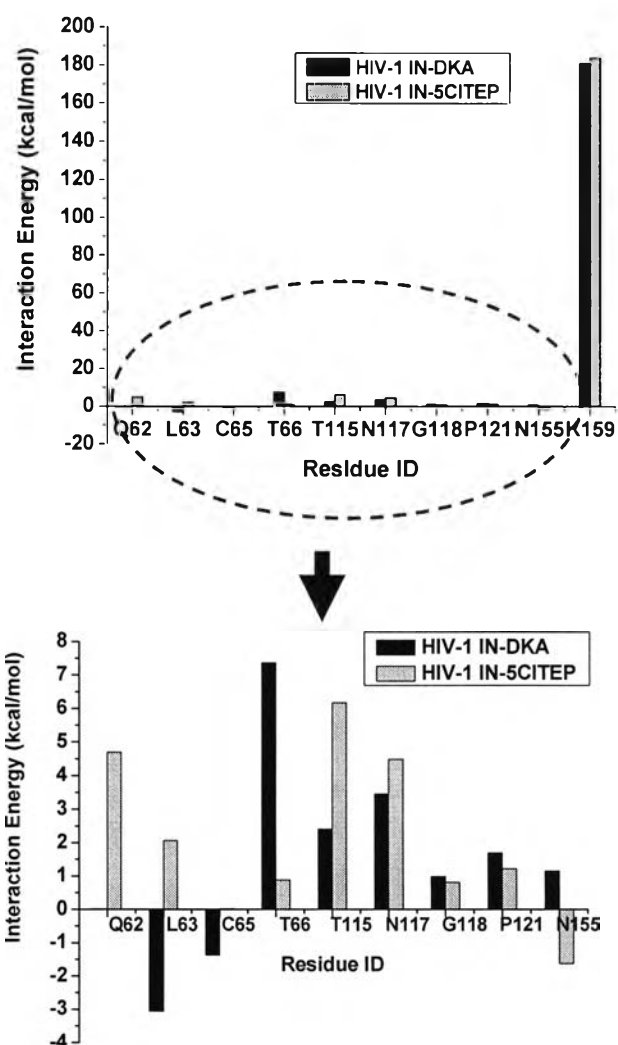
Figure 5.26 Amino acid decomposition analysis showing contributions of each individual MM amino acid to the QM subsystem. Stabilizing and destabilizing interactions are represented as positive and negative energy values, respectively. Only residues exhibiting an energy greater than 1 kcal/mol are shown.

Table 5.7 Standard deviation of amino acid decomposition analysis.

Amino Acid	Standard deviation (kcal/mol)	
	IN-DKA	IN-5CITEP
Gln62	0.82	0.87
Leu63	1.70	0.98
Cys65	2.30	2.41
Thr66	2.38	2.09
Thr115	1.42	0.85
Asn117	1.97	1.24
Gly118	1.13	0.94
Phe121	0.33	0.51
Asn155	2.40	1.14
Lys159	4.70	5.92

B.9 Interaction Energy Decomposition

The total interaction energy between a QM subsystem and MM protein consists of contributions from both the van der Waals and electrostatic interaction. For the investigation of QM part electronic polarization effects, the electrostatic interaction energy is of specific interest, and it is given below:

$$\Delta E_{el} = < \Psi \mid H_{QM}^0 + H_{QM/MM} \mid \Psi > - < \Psi^0 \mid H_{QM}^0 \mid \Psi^0 > \qquad (5.2)$$

$$= E_{QM} + E_{QM/MM} - E_{QM}^0 \qquad (5.3)$$

where Ψ and Ψ^0 refer to the wave function of the QM subsystem in protein environment and in the gas phase, respectively.

To explore the electronic polarization of QM subsystem which is affected by MM protein environment, the total electrostatic interaction energy (ΔE_{el}) and its components were computed [152-155]. In this concept, the total electrostatic interaction is decomposed into a permanent interaction energy (ΔE_{perm}), corresponding to the

interaction energy between the "unpolarized" QM subsystem and the protein, and a polarization energy (ΔE_{pol}) involves changes in the molecular wave function in the environment of MM residues. Therefore, the total electrostatic interaction energy between QM part and environment is:

$$\Delta E_{el} = \Delta E_{perm} + \Delta E_{pol} \qquad (5.4)$$

In Eq. 5.4, the permanent interaction energy is computed using the wave function of the QM subsystem in the gas phase through the QM/MM electrostatic interaction Hamiltonian:

$$\Delta E_{perm} = < \Psi^0 \mid H_{QM/MM} \mid \Psi^0 > \qquad (5.5)$$

The polarization energy can be further divided into a polarization stabilization term (ΔE_{stab}) and a QM subsystem electronic distortion term (ΔE_{dist}):

$$\Delta E_{pol} = \Delta E_{stab} + \Delta E_{dist} \qquad (5.6)$$

The polarization stabilization energy (ΔE_{stab}) indicates the increase in interaction energy of QM subsystem due to its new charge distribution in the MM environment. The electronic distorsion energy (ΔE_{dist}) refers to the penalty for reorganizing the electron distribution of the QM subsystem in MM environment.

The polarization energies converge very quickly and they do not fluctuate significantly with longer dynamics simulation. Therefore, one typically does not need to run extensive averages for the electronic decomposition energy. Actually in this case, the interaction energy decomposition calculations on the fly during dynamics run were tried but without success. This is probably the initial guesses for either the gas phase or condensed phase wave functions may not be good enough and consequently the gas phase wave function is not converged, as suggested by Dr. Jaili Gao. Hence in this work, the average values of total electrostatic interaction and its components (Table 5.7) for each system were taken from a set of 120 snapshots (at every 5 ps, from 600-1200 ps). Our attention is to elucidate how much polarization effect contributes to electrostatic energy.

Table 5.8 Electrostatic interaction energies and their components (kcal/mol) of QM subsystem of HIV-1 IN-DKA and HIV-1 IN-5CITEP.

system	ΔE_{el}	ΔE_{perm}	ΔE_{pol}	ΔE_{stab}	ΔE_{dist}	%polarize
IN-DKA	-319.4± 24.39	-282.4 ±24.03	-37.1 ±2.79	-73.4 ±5.41	36.3 ±2.65	11.60
IN-5CITEP	-351.4 ± 16.07	-303.3 ± 15.37	-42.8 ±3.16	-84.9 ± 6.22	42.1±3.08	12.16

From the above table, the electrostatic interaction energy of IN-DKA is less than that of HIV-1 IN-5CITEP. Of course, the computed interaction energies cannot be directly compared with experimental binding constants, which relates to the change in free energy for the transfer of ligand molecule from water into the enzyme active site. Moreover the computed electrostatic interaction energies in this case were calculated from the QM subsystem not for a pure ligand. It is apparent that both systems have a very similar electronic polarization contribution i.e. ~ 11-12 % which is in the same range of the reported of molecular polarization of small molecules (10 -30 %) [154]. This is probably the tetrazolate ring has a similar acidity to the carboxylate group and they are almost isosteric which can be seen by the replacement of tetrazole by carboxylic acid in biological active molecules. Moreover, it is interesting to note that the absolute value of electronic distortion energy (ΔE_{dist}) is one-half of the absolute value polarization stabilization energy (ΔE_{stab}). This is consistent with the linear response theory, which provides further validation of the computational procedure for the interaction energy decomposition [152, 153, 155].

B.10 Atomic charge polarization

The polarization was also analyzed by computing the partial atomic charges through Mülliken population analysis. Although the calculation of atom charges is not unique and there are many different algorithms to derive the atomic charges, the main aim of this calculation is to show the variation of atomic charges due to molecular polarization. The previous studies showed that the relative change due to charge polarization obtained by Mülliken population analysis can still provide meaningful and valuable insight on the molecular polarization [152, 153, 155]. Therefore, the atomic charges of the QM subsystem for a set of 120 snapshots (at every 5 ps, from 600-1200 ps)

were computed in the environment of MM residues (condense phase) and in gas phase were computed. To focus the discussion, the average values of partial atomic charges of some selected QM atoms which are relevant to binding interactions are given in Table 5.8. The most polarized atoms of DKA are O9 and O10 which make salt-link interactions with the ammonium nitrogen of Lys159. For IN-5CITEP, significant polarization occurred at the OD2 of Asp64, O1, C4, C8, N10, and N11. The OD2 of Asp64, N10 and N11 of 5CITEP formed salt-bridge interaction with Lys159 while O1 chelated with the Mg^{2+} ion. Considering only ligand's structure, the polarization of carboxylate is somewhat greater than that of tetrazolate of 5CITEP and this might support the higher inhibitory potency of DKA.

Table 5.9 Average partial atomic charges and polarizations. Polarization is defined as the difference in averaged charged in atomic units on a given atom in condense phase and in gas phase.

QM atoms	Atomic charge				Polarization	
	Condense phase		Gas phase		IN-DKA	IN-5CITEP
	IN-DKA	IN-5CITEP	IN-DKA	IN-5CITEP		
OD1-Asp64	-0.426	-0.549	-0.428	-0.571	0.002	0.022
OD2-Asp64	-0.535	-0.773	-0.485	-0.580	-0.050	**-0.193**
OD1-Asp116	-0.574	-0.428	-0.554	-0.358	-0.020	-0.070
OD2-Asp116	-0.611	-0.584	-0.576	-0.536	-0.035	-0.048
Mg^{2+}	0.537	0.498	0.514	0.451	0.023	0.047
O-water#1	-0.375	-0.367	-0.355	-0.330	-0.020	-0.037
O-water#2	-0.308	-0.283	-0.287	-0.280	-0.021	-0.003
O1	-0.341	-0.332	-0.283	-0.217	-0.057	**-0.115**
O2	-0.320	-0.358	-0.359	-0.403	0.039	0.045
C3	-0.193	-0.185	-0.212	-0.201	0.019	0.016
C4	-0.224	-0.255	-0.128	-0.141	-0.096	**-0.114**
C5	0.442	0.468	0.363	0.372	0.079	0.096
C6	-0.576	-0.546	-0.567	-0.536	-0.009	-0.010
C7	0.300	0.407	0.343	0.479	-0.043	-0.072
C8	0.482	-0.126	0.388	-0.269	0.094	**0.143**
O9	-0.777	-	-0.619	-	**-0.158**	-
O10	-0.793	-	-0.628	-	**-0.165**	-
N9	-	-0.184	-	-0.181	-	-0.003
N10	-	-0.305	-	-0.199	-	**-0.106**
N11	-	-0.317	-	-0.201	-	**-0.116**
N12	-	-0.205	-	-0.156	-	-0.049

B.11 Protein-ligand binding free energies

For every snapshot taken from QM/MM MD simulations, the free energy of binding was computed according to the MM/GBSA methodology. As it was reported that the results of MM/PBSA approach are comparable to that of GB-based scoring function [99], the GB approach, a cheaper computing time, was used in this study. The atomic charges of QM atoms were taken from the QM/MM MD calculations while that of MM atoms were taken from Charmm22 force field. The entropy of the system was estimated according to normal mode calculation.

Table 5.9 shows the average values of binding free energy of HIV-1 IN-DKA and HIV-1 IN-5CITEP. It is interestingly to note that the calculated and the estimated experimental free energies of binding are in the same trend. The predicted free energies for DKA or 5CITEP bound to the active site of HIV-1 IN are -11.01 kcal/mol and -5.59 kcal/mol, respectively, indicating that DKA has stronger interaction with IN than 5CITEP does. This is in accordance with the fact that DKA is more potent than 5CITEP.

Table 5.10 Binding free energy (kcal/mol) of HIV-1 IN-DKA and HIV-1 IN-5CITEP, calculated using the MM/GBSA approach.

system	3'-processing IC_{50} (μM)	ΔG (kcal/mol)	
		Experimental*	Calculated
HIV-1 IN-DKA	25, 30	-6.5 ± 0.1	-11.01
HIV-1 IN-5CITEP	400, 333	-4.8 ± 0.1	-5.59

* the experimental free energy of binding is estimated by IC_{50} using the following relationship

$$\Delta G_{binding} = RT \ln K_{dissociated}$$
$$= RT \ln (IC_{50} + 0.5Cenzyme)$$
$$\sim RT \ln IC_{50} \qquad (5.7)$$

where R is ideal gas constant, T is temperature in K (310 in this case), C_{enzyme} is the concentration of enzyme, which is a very small number after equilibration and can be omitted in most cases.

5.3 Conclusion

The starting structures of HIV-1 IN complexed with both 5CITEP and DKA were preliminary proposed based on different possibilities that ligand can chelate with the catalytic metal ion. Total interaction energies approximated by pair interaction energies between the amino acid residues (within 8 Å from the inhibitor) and the inhibitor (5CITEP or DKA) of the each model were determined at B3LYP/6-31++G (d,p) level of theory. The results indicated that for ligand state = -2, model no. 2p which the keto-enolate group of ligand directly coordinated with the Mg^{2+} and the tetrazolate/carboxylate points to the same direction of Asp64 was the most stable for both inhibitors. However, for ligand state = -1, the X-ray orientation (Model no. 7p) is the most stable. As we were interesting in the mechanism of action of the metal-ligand binding, only model no. 2p of both inhibitors was thereby studied by means of MD simulation.

Both convention and hybrid QM/MM approaches were employed to perform MD simulation. Therefore, there are two important aspects of the analysis of our results. The first one is the comparison of the results obtained from conventional and mixed QM/MM MD simulations of HIV-1 IN-5CITEP and HIV-1 IN-DKA to determine whether classical MD is sufficient and suitable for the studied systems. Another aspect is how HIV-1 IN-5CITEP and HIV-1 IN-DKA complexes differed in the QM/MM MD simulations.

Comparison of HIV-1 IN structures taken from classical and QM/MM MD simulations of HIV-1 IN-5CITEP reveals two notable differences around residues 116-119 and 140-149. In the conventional MD simulation, residues 116-119 show high mobility, however, this loop was quite rigid in the QM/MM MD simulation. This is probably because in conventional MD, Gly118 occasionally forms a hydrogen-bond with Phe121 and Thr122 while in the QM/MM study, it is always involved in a hydrogen-bond. Moreover, differences in the dynamics of loop residues 140-149 lead to

substantially different orientations of key amino acids located in the α4-helix, particularly Lys159, between the two calculations. Another significant difference between the two levels of calculation is the coordination geometry of Mg^{2+} which is found to be octahedral and distorted octahedral for the MM and QM/MM simulations, respectively. In addition, the bound positions of 5CITEP in the classical and QM/MM MD simulations are quite different. In the MM study, the ligand is positioned far from Lys159, thus, it cannot form a salt-link with this residue. On the contrary, a salt-bridge between this amino acid and 5CITEP was detected in the QM/MM study.

For HIV-1 IN-DKA, the differences of protein structure between conventional and QM/MM MD results were observed at the regions involving residues 64-69, residues 140-149 (a catalytic loop), and residues 150-166 (α4-helix). The conformational difference of the first region is probably arisen from the different pattern of metal coordination geometry. In classical MD simulation, Mg^{2+} ion was a near perfect octahedral complex whereas a distorted octahedral geometry was observed in QM/MM. Only OD1 of Asp64 found in classical MD simulation chelated with the Mg^{2+} while both OD1 and OD2 of the same residue coordinated this divalent metal ion as found in QM/MM. Similar to the case of HIV-1 IN-5CITEP, the difference of the loop residues 140-149 affects the orientation of amino acid residues located in the α4-helix. This leads to a dissimilar salt-link interaction between Lys159 and ligand.

In addition, detailed structural analyses of the HIV-1 IN complexed with DKA and with 5CITEP indicate that the protein and inhibitor in the two complexes display a different dynamical behavior. Substantial structural differences were found at (ii) residues 60-68, (ii) residues 116-119, and (iii) residues 140-149 (a flexible loop). The IN-DKA complex exhibits a shorter β1-sheet (60-66) than in the IN-5CITEP complex (residues 60-68). Furthermore, loop residues 116-119 in HIV-1 IN-DKA is much more curved than that seen in HIV-1 IN-5CITEP. Moreover, the side chains of Asp64 and Asp116 in the two systems are significantly different resulting somewhat different pattern of the distorted octahedral geometry of Mg^{2+} between HIV-1 IN-DKA and HIV-1 IN-5CITEP systems. In the HIV-1 IN-DKA complex, Asp64 oriented its OD1 and OD2 atoms to chelate the Mg^{2+} in a bidentate configuration while only the OD1 is used to coordinate the divalent metal ion in the HIV-1 IN-5CITEP complex. Moreover, both

OD1 and OD2 belonging to Asp116 in HIV-1 IN-DKA are involved in water-mediated hydrogen-bonding while only the OD1 of IN-5CITEP can form a hydrogen-bond with a bridging water molecule.

The bound position of DKA and 5CITEP is somewhat similar. The keto-enolate moiety directly interacts with the Mg^{2+} while the carboxylate (for DKA) or tetrazolate (for 5CITEP) is oriented toward residue Lys159. Moreover, the keto-enolate motif of both DKA and 5CITEP is indirectly hydrogen-bonded to Asp64 and Asp116 via bridging water molecules. An analysis of the conformational dynamics of the ligand reveals that the DKA conformation is quite rigid across the MD simulation while the 5CITEP conformation exhibits high mobility. Furthermore, a strong salt link between DKA and HIV-1 IN existed for the entire course of the simulation whereas that between 5CITEP and Lys159 was weaker. This means that ionic interactions were stronger in the HIV-1 IN-DKA complex as compared with the HIV-1 IN-5CITEP system. This supports experimentally observed the higher potency of DKA as compared to 5CITEP. Amino acid energy decomposition is also in support of the important role of Lys159 in the mechanism of action. The combined QM/MM energy decomposition calculations indicated that the polarization contribution, while subtle in terms of changes in the charge density, is energetically significant, accounting for about 11-12% of the total electrostatic interaction energy. Therefore, it is essential to consider electronic polarization of ligand in order to understand protein-ligand interactions. The atomic polarization analysis indicated that the O9 and O10 of DKA and N10 and N11 of 5CITEP have a high degree of polarize. These mentioned atoms established salt-link interaction with the ammonium nitrogen of Lys159. In addition, the polarization of acetate oxygens (OD1 and OD2) of Asp64 in the case of HIV-1 IN-DKA was lower than that of HIV-1 IN-5CITEP. The binding free energies calculated by MM/GBSA approach were -11.01 kcal/mol and -5.59 kcal/mol for HIV-1 IN-DKA and HIV-1 IN-5CITEP, respectively. In general, our study provides insights into the detailed mechanism of action of DKA and 5CITEP and supports the higher inhibitory potency of DKA in comparison with 5CITEP