CHAPTER 3
COMPUTATIONAL DETAILS

3.1 Molecular modeling

It should be noted that 3.1.1 is discussed here about the experimental structures of the HIV-1 IN. Detailed analysis in term of theoretical point of view helps in the modeling studies of the full-length system.

3.1.1 Experimental structures: A theoretical point of view

The structures of HIV-1 IN were experimentally resolved using X-ray crystallography and NMR techniques. All of the core domain and the two two-domain fragment structures, core connected to N-terminal (CORE-N) and core connected to C-terminal (CORE-C), were solved by X-ray crystallographic technique whereas the two terminal end domains, the N-terminal and the C-terminal domain, were obtained from NMR studies. Up to date, there are 25 HIV-1 structures available at the Protein Data Bank (PDB) **(78)**, 6 for the N-terminal domain, 2 for the C-terminal domain, 15 for the core domain, 2 for the two-domain fragment. The full-length structure composing of three domains is still experimentally unsolvable. Among those structures, there is only one complex structure with an inhibitor targeted at the core domain. The detailed information extracted from the experimental structures were summarizes in Table 3.1

3.1.1.1 The one-domain structures

3.1.1.1.1 N-terminal domain

The N-terminal domain contains a conserved two histidine and two cysteine residues, known as HHCC, as a zinc binding residues (see Figure 3.1) **(29)**. The two histidine, His12 and His16, and two cysteine residues, Cys40 and Cys43 form a pocket and bind to Zn^{2+}. The determination study revealed that there are three forms in which the orientation of these four key residues are different. The N-terminal domain exists in two forms, which are D and E forms, differing in an orientation of His12 and His16 binding to zinc (Figure 3.2). The existing of these two

forms depends on the experimental temperature, *i.e.* the E form is dominated below ~300K while the D form is the major one of temperature above ~300K **(31)**. There is not any evidence suggesting which form is active in the nature.

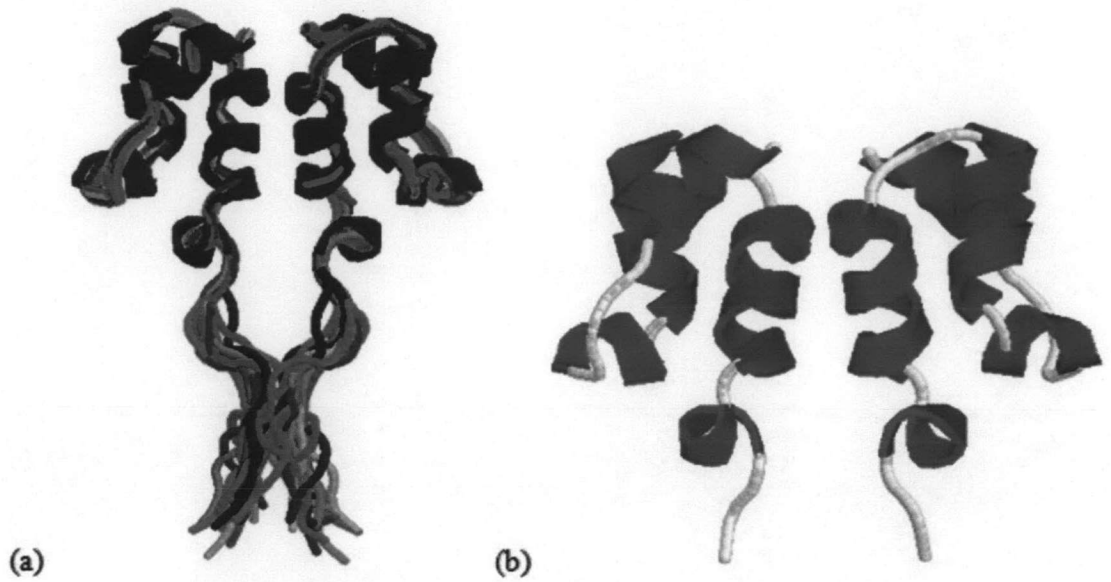


Figure 3.1 (a) NMR solved solution structures of the N-terminal domain (residues 1 – 55) and (b) their regularized mean structure (residues 1 – 47) **(31)**.

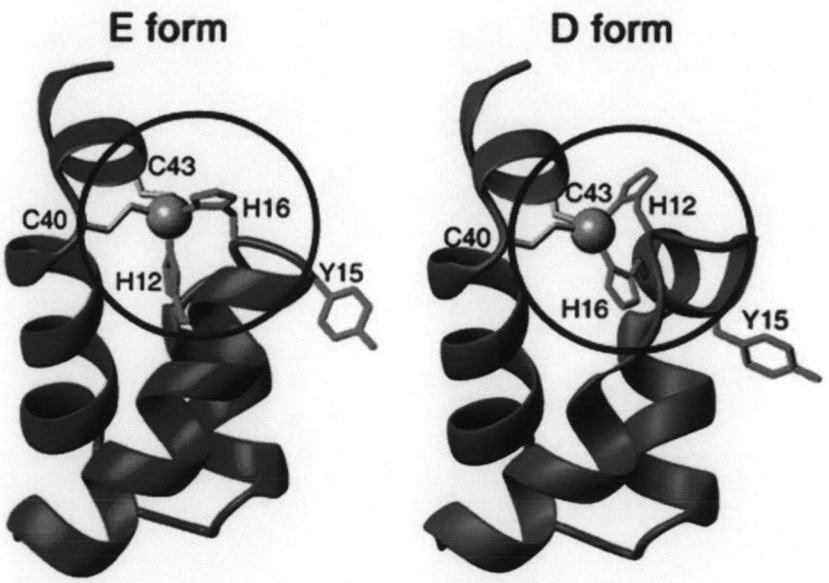


Figure 3.2 Differing in the nature of the metal coordination by the two histidine residues (His12 and His16), designated as D and E forms.

3.1.1.1.2 Catalytic core domain

The crystallographic studies illustrate that the core domains form monomer, dimer and trimer **(79-81)** (Figure 3.3). Only several structures (PDB code 1BIS, 1BL3 and 1EXQ) were found to bind to divalent metal ions (See also Table 3.1). Almost all structures contain some missing residues except for some chains (Table 3.1). The structure of the central core domain can be grouped into two forms, cacodylate form and non-cacodylate forms. There are monomeric, dimeric and trimeric forms of the core domain available in the database (Figure 3.3). All structures are differing in the modification of the residue by cacodylate ion and derivatives. This cacodylate modification has an effect on the orientation of the Asp116. The difference in an orientation of the Asp116 in both forms can be clearly observed (Figure 3.4). The modification were made to the structure to help the experimentally resolve process **(79)**. The root mean square deviation (RMSD) of the 15 available X-ray structures was shown in Figure 3.4 c. It can be seen that the RMSD in residue Asp116 and nearby residues are larger than the others. The two regions, residues 140 – 148 and 189 – 192, were considering missing in almost all structures. The RMSD in these two regions are supposed to be high. The crystal structure of 1BL3 was selected and named CORE in this study.

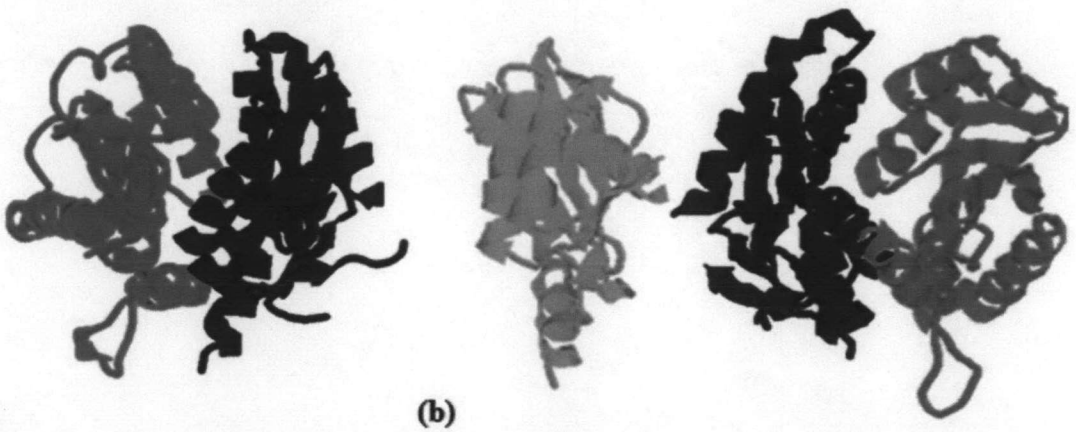


Figure 3.3 Crystal structure of the core only domain in (a) a dimeric form and (b) a trimeric form. The structure was colored by specified chain.

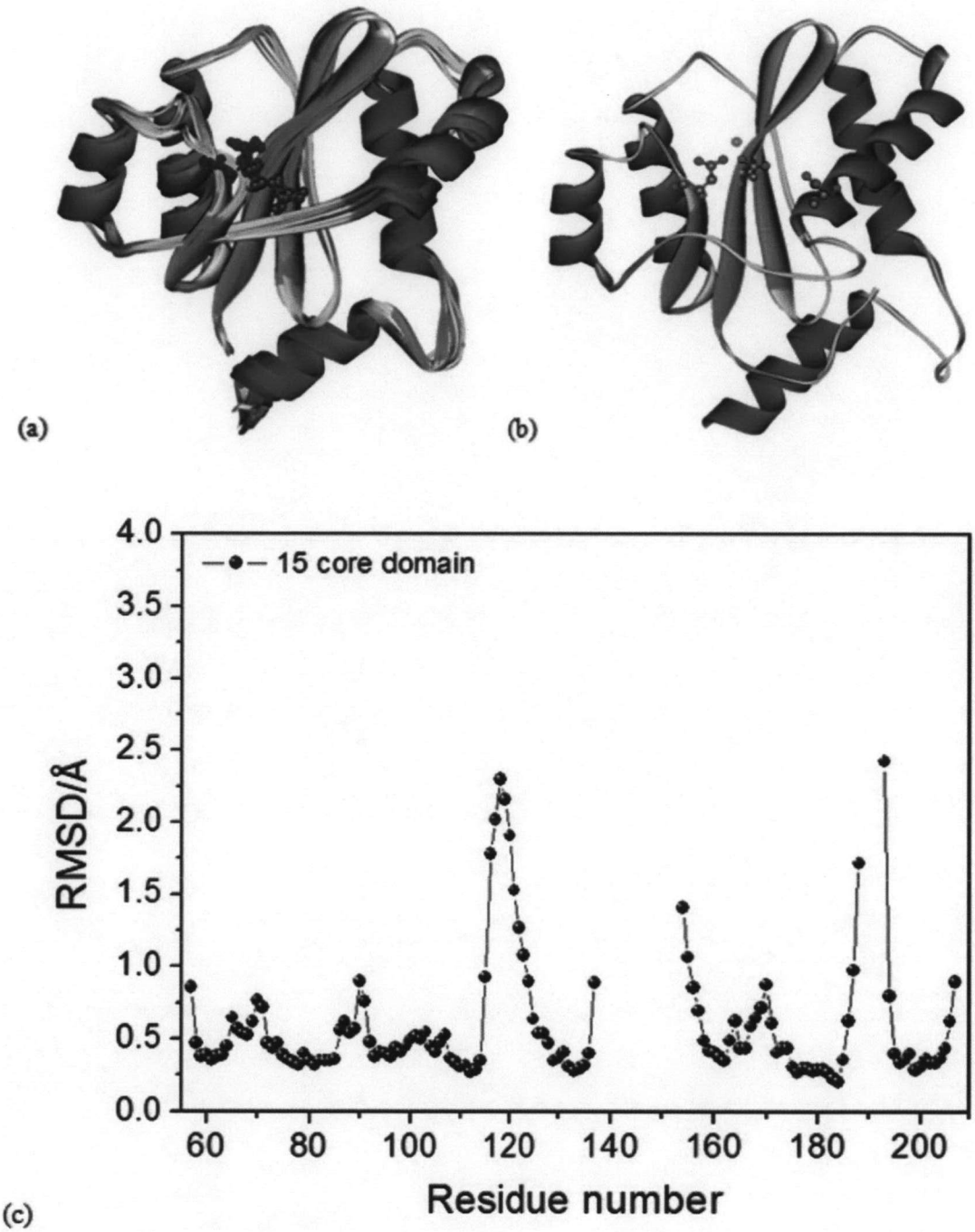


Figure 3.4 (a) Superimposition of 15 core domain crystal structures, both orientations of Asp116 are highlighted in blue and green, respectively, as for non-cacodylate and cacodylate forms, (b) the orientation of the three acidic residues in the active site region and a Mg^{2+}, and (c) the correspondence RMSD values for 15 core structures.

3.1.1.1.3 C-terminal domain

The C-terminal domain structure was considerably less conserve among the three domains **(29)**. The solution structure obtained from NMR

study show that the dimer was formed (Figure 3.5). The dimer structure has a groove between the two monomer chains. It was proposed by many studies that the structure is suitable for non-specific binding of DNA **(80-82)**.

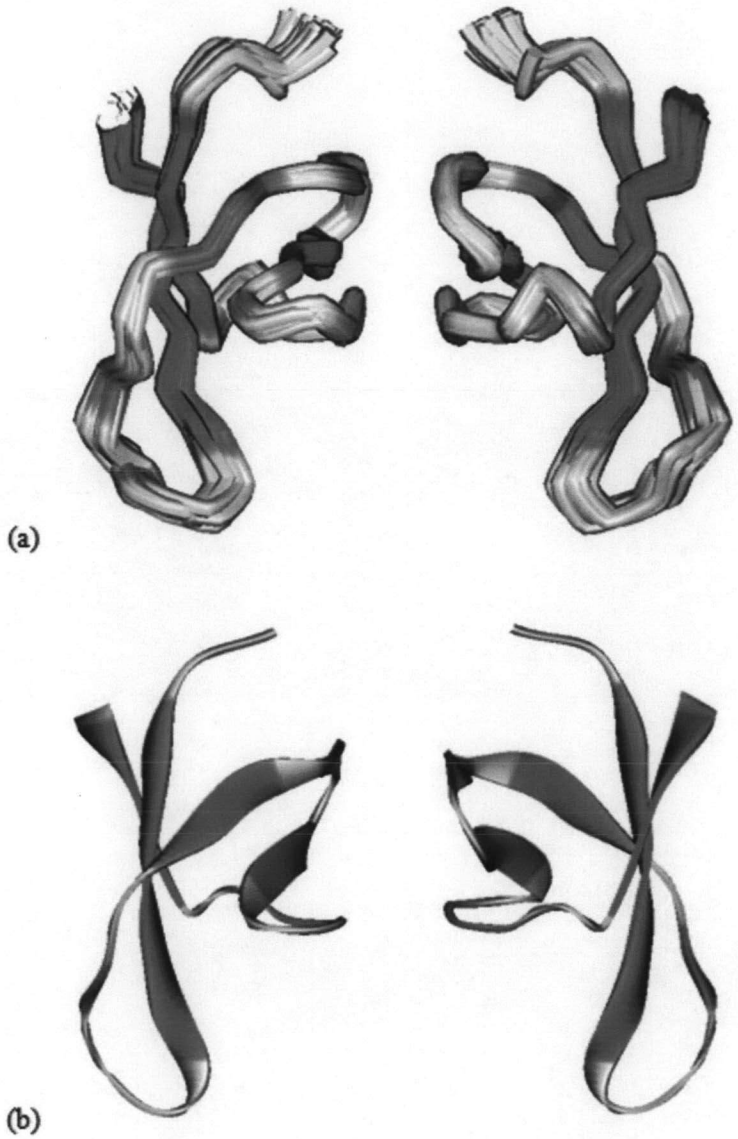


Figure 3.5 (a) Set of structures resolved from NMR technique of the C-terminal domain and (b) the mean structure **(83-85)**.

3.1.1.2 The two-domain fragment structures

3.1.1.2.1 Core connected to N-terminal domain

The crystal structure of the two-domain fragment for the core domain connected to the N-terminal domain **(86)** was name, for simplicity, CORE-N in our study. The structure composes of residues 1 – 212 and form tetramer with

bound Zn ion in the N-terminal domain (Figure 3.6). There are two missing regions in the connection between N-terminal and core domain (residues 47 – 55) and in the core part (residues 140 – 148). The structure contains three-site mutations, W131D, F139D and F185K for improving protein solubility and facilitating the crystallization process **(76, 84)**. Previous biochemical and structural studies suggested that the active form of the HIV-1 IN is likely to be at lease a tetramer **(87)**. The crystal structure of the CORE-N shows the tetramer structure (chain A, B, C and D) forming by two dimer in which relates by a non-crystallographic two fold axis. The Zn ion bound to chain B can interact with the residues in the core domain of the chain D **(86)**.

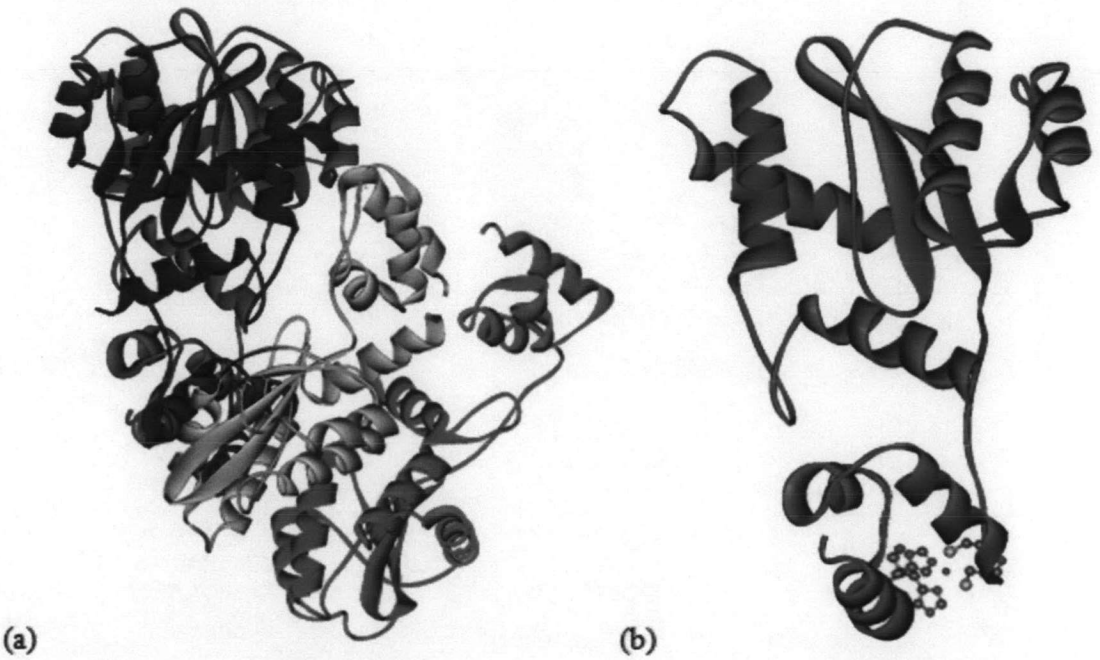


Figure 3.6 Crystal structure of the CORE-N (PDB code 1K6Y, **86**) (a) in a tetramer form and (b) chain A of the CORE-N was selected in our study, the coordination of Zn^{2+} and the two histidine and two cysteine residues in the N-terminal domain were shown in ball and stick model.

3.1.1.2.2 Core connected to C-terminal domain

The structure of the core connected to the C-terminal, CORE-C, composes of residues 56 – 270. The structure forms a symmetric dimer in which each catalytic core monomer was linked to the C-terminal domain by residues 195 – 220 **(88)** (Figure 3.7 a). Comparing to the CORE-N structure which is rather compact, the CORE-C is longer in shape, *i. e.,* the orientations of the C-terminal domain were

extended from the central core domain. Heteroatoms, named CHAPS, were found in the C-terminal region to mediate the packing of the hydrophobic interface in the C-terminal region. Chain A of the dimer form was taken for our study (Figure 4.7 b). The CORE-C structure contains five-mutation; C56S, W131D, F139D, F185K and C280S, without altering the structure of the core region **(88)**. There is one missing region in each subunit, residues 142 – 145 in chain A and residues 138 – 149 in chain B.

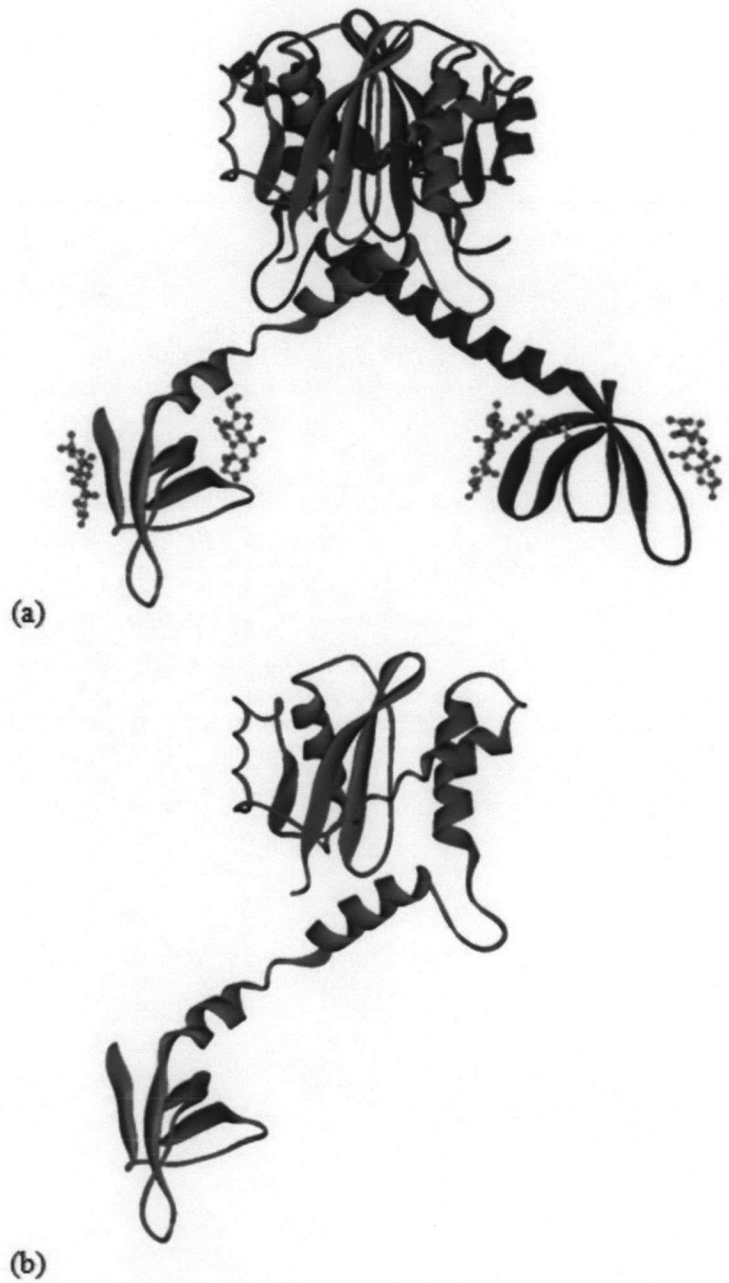


Figure 3.7 Crystal structure of the CORE-C (PDB code 1EX4, **88**) (a) in a dimer form, CHAPS heteroatoms were shown in ball and stick model, chain A of the structure was selected in this study (b).

Table 3.1 Detailed information of the experimental structure of each domain in the HIV-1 IN.

Domain	ID		No. of				Residue	Missing	Mutation	Ref.
	PDB	Chain	Residue	Water	Ion	Others				
N-terminal	1WJA*	A	47	-	Zn	-	1 – 47	-	-	31
		B	47		Zn		1 – 47	-		
	1WJB**	A	55	-	Zn	-	1 – 55	-	-	
		B	55		Zn		1 – 55	-		
	1WJC*	A	47	-	Zn	-	1 – 47	-	-	
		B	47		Zn		1 – 47	-		
	1WJD**	A	55	-	Zn	-	1 – 55	-	-	
		B	55		Zn		1 – 55	-		
	1WJE*	A	46	-	Cd	-	1 – 46	-	H12C	
		B	46		Cd		1 – 46	-		
	1WJF**	A	55	-	Cd	-	1 – 55	-	H12C	
		B	55		Cd		1 – 55	-		
C-terminal	1IHV*	A	52	-	-	-	219 – 270	-	-	38
		B	52				219 – 270			
	1IHW**	A	52	-	-	-	219 – 270	-	-	
		B	52				219 – 270			

Table 3.1 (cont.)

Domain	ID		No. of				Residue	Missing	Mutation	Ref.
	PDB	Chain	Residue	Water	Ion	Others				
Core	1B92	A	144	100	2(As), 2(SO_4)	2(CAC)	56 – 210	141 – 148, 190 – 192	F185K, G149A	91
	1B9D	A	143	110	2(As), 2(SO_4)	2(CAC)	57 – 210	141 – 148, 190 – 192	F185K	
	1B9F	A	151	69	2(As), 2(SO_4)	2(CAC)	56 – 210	189 – 192	F185K, G140A, G149A	
	1BHL	N/A	135	53	2(As)	2(CAS)	57 – 207	138 - 153	F185H	60
	1BI4	A	142	150	-	-	57 – 208	141 -150	F185H	
		B	143				57 – 209	140 – 149		
		C	160				50 – 209	-		
	1BIS	A	146	152	-	-	56 – 209	143 – 150	W131E, F185K	
		B	154				56 – 209	-		
	1BIU	A	146	274	Mg	-	56 – 209	141 -147	W131E, F185K	35
		B	146		Mg	-	56 – 209	141 – 147		
		C	146		Mg	-	56 – 209	141 – 147		
	1BIZ	A	147	212	-	2(CAC)	54 – 208	141 -147	C56S, F185K	
		B	143			2(CAC)	58 – 212	140 – 152		

Table 3.1 (cont.)

Domain	ID		No. of				Residue	Missing	Mutation	Ref.
	PDB	Chain	Residue	Water	Ion	Others				
Core	1BL3	A	142	150	Mg	-	57 – 208	141 -150	F185H	60
		B	143		Mg	-	57 – 209	140 – 149		
		C	160		Mg	-	50 – 209	-		
	1EXQ	A	147	99	4(Cd), 4(SO_4), 2(Cl)	-	56 – 209	144 -150	W131D, F185K	89
		B	145				56 – 209	142– 150		
	1HYV	A	144	98	SO_4, Cl, 3(As)	2(CAF) TTA	57 – 210	141 -147	F185K	92
	1HYZ	A	144	81	SO_4, CL, 4(AS)	2(CAF) TTO	57 – 210	141 -147, 190 -192	F185K	
	1ITG	N/A	142	65	-	2(CAC)	55 – 209	141 – 152	F185K	91
	1QS4	A	150	544	Mg	5CITEP	56 – 209	141 -144	W131E, F185K	81
		B	152		Mg	-	56 – 209	141 – 142		
		C	151		Mg	-	56 – 209	141 – 143		
	2ITG	N/A	160	58	-	-	51 - 210	-	F185H	79

Table 3.1 (cont.)

Domain	ID		No. of				Residue	Missing	Mutation	Ref.
	PDB	Chain	Residue	Water	Ion	Others				
Core-N***	1K6Y	A	192	176	$4(PO_4)$ 4(K) 4(Zn)	-	1 – 210	47 – 55, 140 – 148	W131D, F139D, F185K	**86**
		B	194				1 – 212	47 – 55, 140 – 148		
		C	191				1 – 209	47 – 55, 140 – 148		
		D	189				1 – 211	47 – 55, 140 – 152		
Core-C***	1EX4	A	212	79	-	4(CHAPS)	56 – 270	142 -145	C56S, W131D, F139D, F185K, C280S	**88**
		B	204				56 – 270	138– 149		

* NMR regularized mean structure
** NMR, 38 – 40 structures
*** Two-domain fragment structure

SO_4: sulfate ion
PO_4: phosphate ion
CAC: cacodylate ion
CAS: s-(dimethylarsenic)-cysteine
CAF: s-dimethylarsinoyl-cysteine
CHAPS: 3-[(3-chloramidopropyl)dimethylammonia]-1-propanesulfonate
TTA: tetraphenyl-arsonium
TTO: (3,4-dihydroxy-phenyl)-triphenyl-arsonium

3.1.1.3 Comparison of HIV-1 IN domain structures

In order to study the structure and dynamical behavior of the HIV-1 IN, set of structures was chosen as follow; core domain (PDB code 1BL3, **60**), which will be named CORE in order to distinguish between the core domain and the core system of study, CORE-N (PDB code 1K6Y, **86**) and CORE-C (PDB code 1EX4, **88**). The structures of the one-domain end terminals, N-terminal domain (PDB Code 1WJA, **31**) and C-terminal domain (PDB code 1IHV, **38**) were also considered, namely, for simplicity NTER and CTER, respectively. By considering only the monomeric form, chain A of each structure was obtained, except for the CORE. We selected chain C from the CORE since it was a complete chain, *i. e.*, no missing residues present.

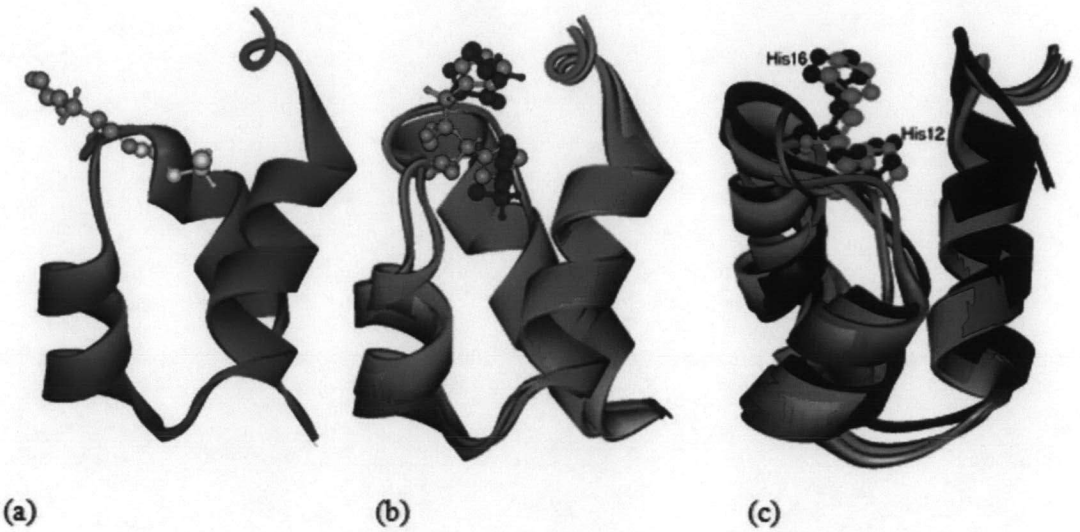
Comparisons were made to each domain region separately. The crystal structures of the one-domain (N-terminal domain, CORE and C-terminal domain) were superimposed with the same domain taken from the two-domain structures, *i. e.*, N-terminal and core domain of the CORE-N and core and C-terminal domain of the CORE-C. The detailed of studies on each domain region were described as follows.

3.1.1.3.1 N-terminal domain

N-terminal region of the N-terminal structure in various form, D form (PDB code 1WJA), E form (PDB code 1WJC) and the mutant form (PDB code 1WJE), and the CORE-N were superimposed in order to observe the difference in the structure. The four structures were superimposed using MOLMOL program and their correspondence RMSD values were calculated and plotted (Figure 3.8).

The superimposition study reveals that the orientation of the two histidine residues are different as mentioned above. Interestingly, the H12C mutation (PDB code 1WJE) shows that the His16 (orange, Figure 3.8a) is in the configuration different from the H16. (His12 (blue) and His16 (orange), (Figure 3.8b). Moreover, the superimposition between these four regions show that the N-terminal domain of the two-domain CORE-N is in E form (Figure 3.8c). The results can be

clearly seen in Figure 3.8d in which the RMSD values were plotted against the residue number. There is not much different between the two histidine regions of the E-form and the CORE-N (red line) while the other two lines show larger RMSD values.



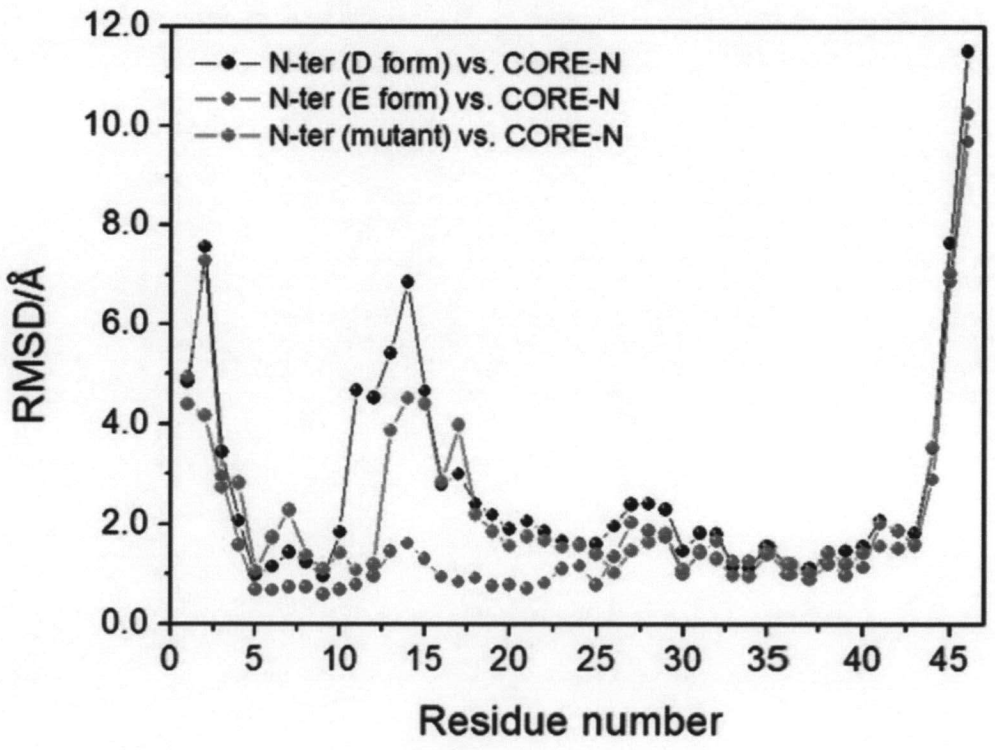


Figure 3.8 Superimposition of the four N-terminal regions, (a) only mutant form shown (red), (b) D (violet) and E (cyan) form, note that His12 are colored blue while His16 are colored orange, (c) inclusion of the N-terminal region from the CORE-N

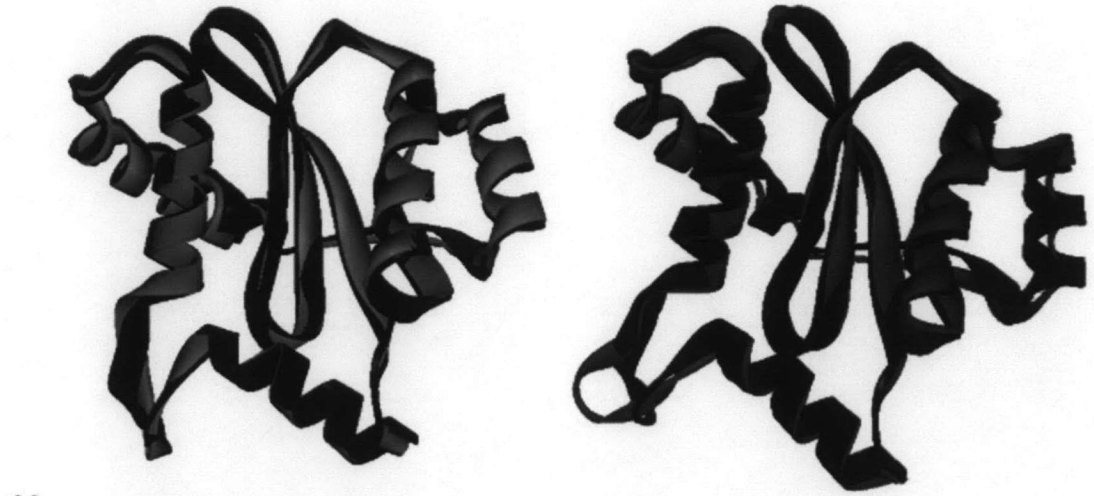
(black), and (d) their correspondence RMSD values. The two histidine residues of the structures were shown in ball and stick model.

3.1.1.3.2 Catalytic core domain

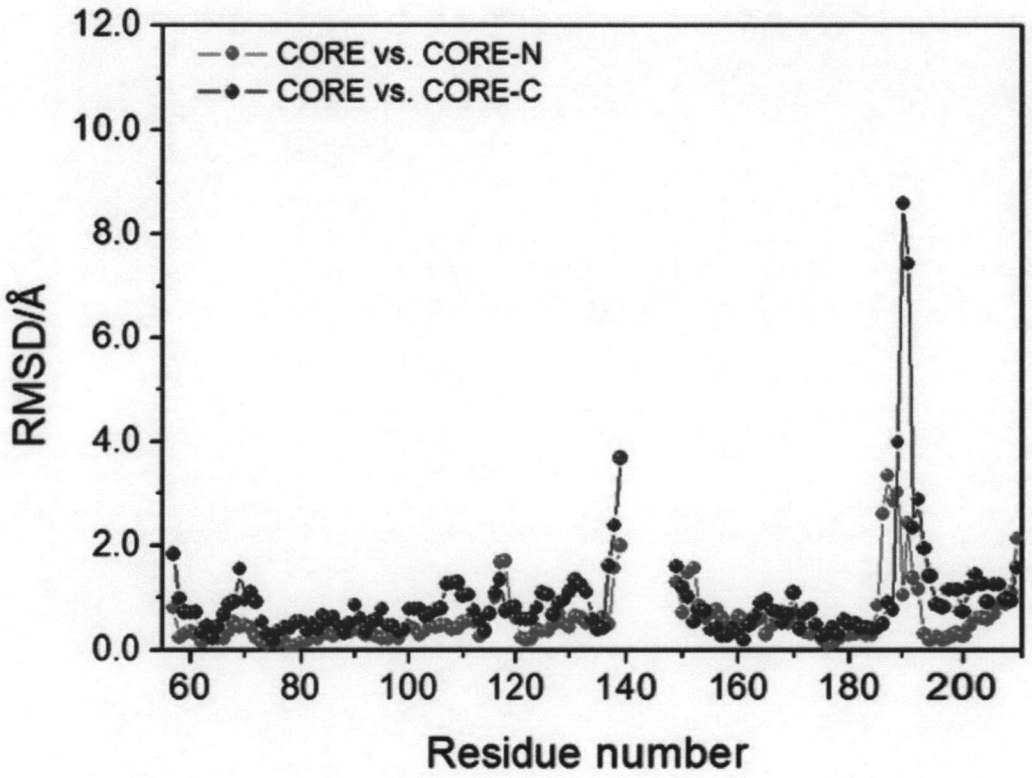
The catalytic core domain considering in this topic spans from residue 56 to 209. The core region from the two-domain structure; CORE-N and CORE-C were taken and compared to the CORE domain. Residues 140 – 148 were considered in all structures as a missing region. The result from superimposition of the core domain region from three structures shows that the whole structure is likely to be the same except for the missing region (residues 140 – 148 and residues 188 – 192). The second region (residues 188 – 192) is highly flexible and found to be missed in some crystal structures (Figure 3.9). Unlike the rest of the structure, residues 188 – 192 shows high different between the core only domain and the core domain from the two-domain structure especially the CORE-C. Superimposed structure between the CORE and CORE-C (Figure 3.9a, right) shows that this region in both structures is different way. From the result, it is clearly seen that the three dimensional structure of the core domain region is rather conserved with the RMSD fluctuate in the range of 0.1 – 1.9 Å (Figure 3.9b).

3.1.1.3.3 C-terminal domain

The structure of the C-terminal region taken from the CORE-C was superimposed to the C-terminal structure (Figure 3.10). By ignoring the orientation of the C-terminal domain prior to the core structure, we observed that the structures of the residues at both terminal ends (residues 220 – 223 and 269 – 270) and residues 231 – 233, as for the RMSD > 2.0, are greatly different compare to the entire C-terminal domain. It is fairly different for the rest of the structure.

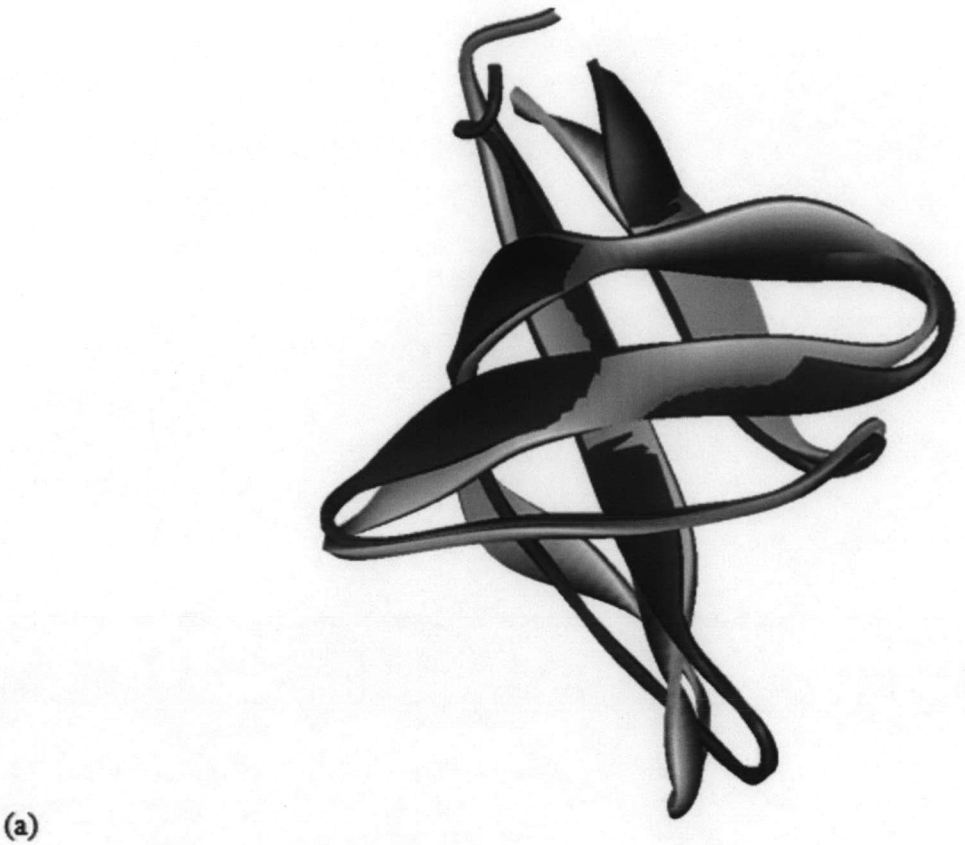


(a)

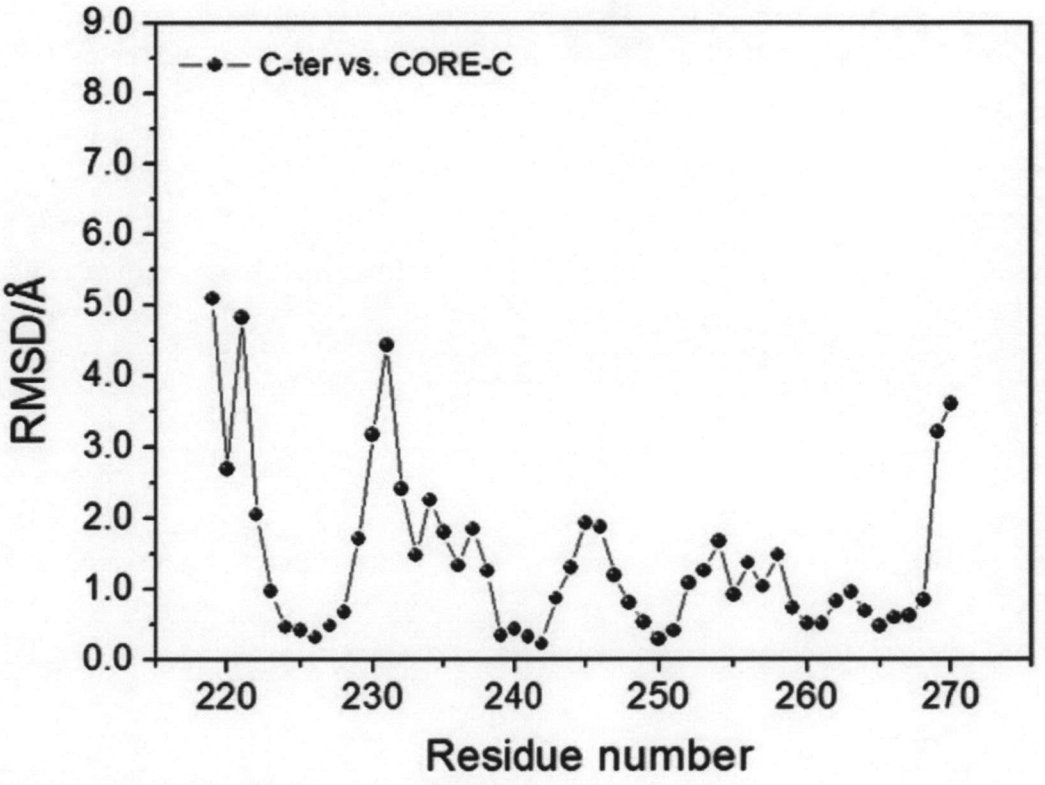


(b)

Figure 3.9 (a) Superimposition of the core region of the CORE (black), the CORE-N (red) and the CORE-C (blue) and (b) the RMSD values.



(a)



(b)

Figure 3.10 (a) Superimposed structure between the C-terminal domain (CTER, yellow) and the C-terminal region in the CORE-C (green) and (b) the correspondence RMSD values.

3.1.2 Homology modeling step

All missing residues in the four structures were modeled using homology modeling module in the Insight II program (**94**).

There are several databases which relate the sequence and structure available base upon the fact that the homologous amino acid sequence at a given level leads to the similar 3D structures of protein. By comparison to the data extracted from the homologous sequences, the model building of a target structure can be achieved.

Figure 3.11 A homology module in the Insight II program, the main window shows the molecule and the amino acid sequence window demonstrates the homologous sequences **(94)**.

With the Insight II program **(94)**, the two crystal structures; CORE-N and CORE-C were used as starting structures. The homology module was, then, loaded up. The complete amino acid sequences of each structure are also needed. The steps for modeling the missing residues are as follow:

Step 1 Load the starting structure and extract its sequence

The structure was displayed in the main window. After extracting its sequence, the one letter representing the amino acid residue in the structure will appear in another window. (Figure 3.11)

Step 2 Load the complete sequence and align

The complete sequence of the structure was shown together with the extracted one. Both sequences are aligned and leave the missing region blank.

Step 3 Propose structures for the loops or missing regions

The program will search protein databases for the structure similar to the model protein. Ten loop structures were proposed. The first proposed loop of having highest score is normally selected.

Step 4 Copy coordinates proposed to the model structure and refine

The coordinates proposed from the program were copied to the model structure. The newly built structure was refined using the molecular mechanics facilities (Discover module) in the Insight II program. The complete structure was then used as a starting structure for the simulation step.

3.1.3 Modeling of the full-length structure

3.1.3.1 Full-length monomer

The full-length structure was built using the two experimental two-domain fragments, CORE-N, PDB code 1K6Y and CORE-C, PDB code 1EX4 **(86, 88)**. The structures of CORE-N and CORE-C were superimposed using the core part as a common region. Then, residues 56 – 210 of the core domain of the CORE-C were removed (see Figure 3.12). The peptide linkage between the residues Thr210 of the CORE-N and the Lys211 of the C-terminal was created and refined using Insight II. The reason for removing the core domain from the CORE-C rather from the CORE-N fragment in the superimposed structure is that the CORE-C contains more mutate residues. The obtained structure containing all three domains is referred to hereafter as FULL (Residues 1 – 270). Note that the residues 271 – 288 were not

included in our full-length structure because they are still not resolved experimentally. It was reported that these residues do not play role in the DNA binding **(39)**. The missing residues regions of the structure were modeled and refined using homology modeling module in the Insight II 1999 package **(94)**. The HIV-1 IN full-length model herein was displayed in Figure 3.13. In order to investigate the effect of divalent metal ions in the active site, two divalent metal ions, Zn^{2+} and Mg^{2+}, were placed in the FULL model at the same position as found in the crystal structures **(60, 81)**. The Zn^{2+} coordinated with His12, His16, Cys40 and Cys43 in the N-terminal region while the Mg^{2+} coordinates with Asp64, Asp116 and Glu152 in the catalytic site in the core domain region. The full-length structure is, then, named FULL+ION in this study.

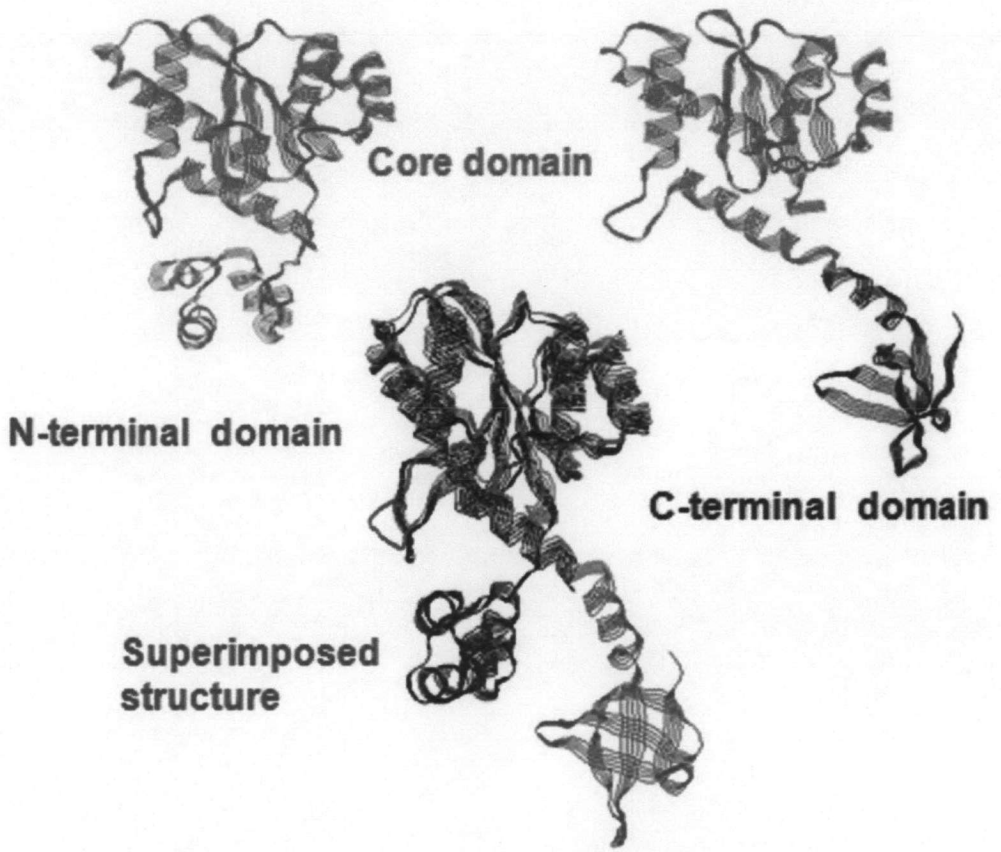


Figure 3.12 The CORE-N, CORE-C and the superimposed structures.

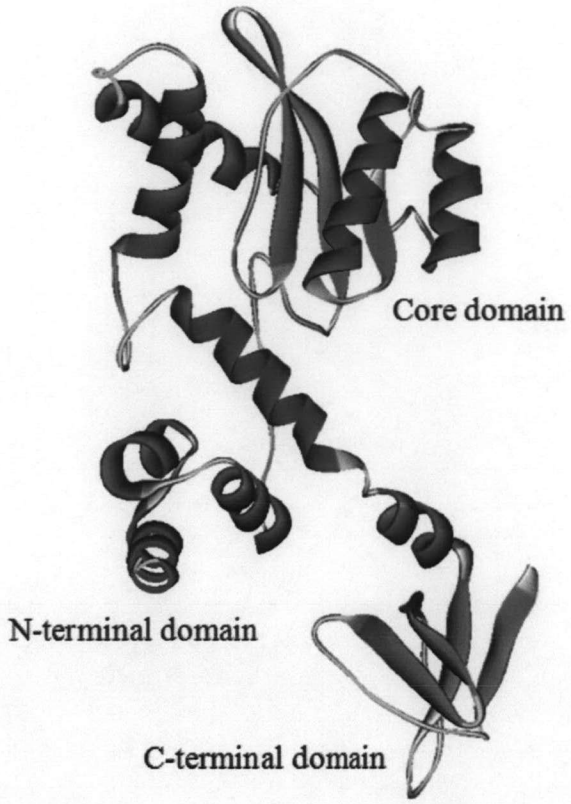


Figure 3.13 The full-length model proposed in this study. Each domain was labeled.

3.1.3.2 Full-length dimer and tetramer

The dimer and tetramer forms of the HIV-1 IN were built based on the two-domain fragment structures, CORE-N and CORE-C, in the same manner. All missing residues in the structure were homology modeled and refined. Table 3.2 summarizes the missing residues in the CORE-N, CORE-C and the full-length structures.

Table 3.2 Missing residues in the two-domain structures, CORE-N and CORE-C, and the full-length (FULL).

Structure	Missing residues
CORE-N	47 – 55, 140 – 148
CORE-C	141 – 144
FULL	47 – 55, 140 – 148

3.2 MD simulations

All MD calculations and trajectories analysis were performed using the AMBER7, a suite of programs allowed to carry out the MD simulations, particularly for biomolecules **(95)**. Information files required for the simulation programs are *(1)* coordinate file (*prmcrd*) in the term of Cartesian coordinates for each atom in the system, which is usually taken from crystal structure, NMR structure or even a model building structure, *(2)* topology file (*prmtop*) containing names and types of atoms, residue names, connectivity of atoms and charges. The information can be obtained from the AMBER database for standard amino acid, *(3)* force field parameters files containing the data for bonds, angles, dihedral angles and atom types in the system which is available in the AMBER program, and *(4)* input command file containing specified options and parameter desired. The basic information flows and the modules in AMBER7 were displayed in Figure 3.14.

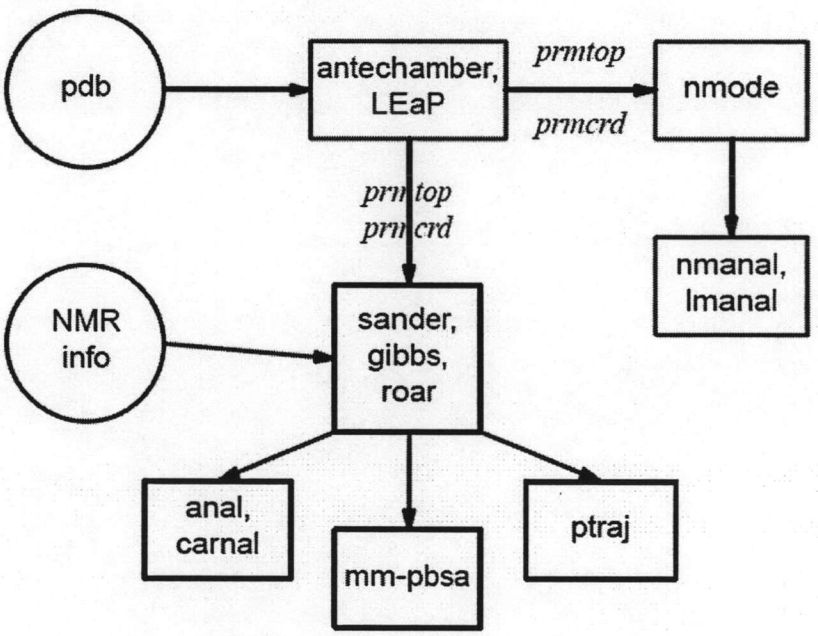


Figure 3.14 Basic information flows for module in AMBER7 **(95)**.

3.2.1 Basic steps in the AMBER7

Basic procedure for MD simulations using AMBER7 program consists of three steps as follow **(95)**:

Preparation step (LEaP)

In the first step, starting coordinates were obtained, normally, from the Protein Data Bank. The coordinates file need to be modified prior to be used by AMBER. The desired molecules were kept whereas others were removed. The names of the amino acid residues need to be selected in this step depending on which form the simulation is going to be carried out. For example, the name for histidine is HIP for the protonate form or neutral form, *i. e.*, which position hydrogen atom is at delta (HID) or epsilon (HIE). All hydrogen atoms and any other missing atoms in the amino acid as well as water molecules and counter ion have to be added. The *leap* module is used for all tasks. The parameter and topology files are generated for the whole system and will be used in the next steps.

Calculation step (sander)

In this step, all energies calculations (minimizations and MD simulations) were performed using the starting parameters, topology and coordinates generated from the first step. Parameters for running either energy minimization or MD simulations such as temperature, pressure, time step, *etc.* can be set prior to the calculation. A *sander* is the module which carried out such calculations. For MD simulations, the history files, called trajectory containing coordinates and velocities for each frame, are written and used for all analysis in the last step.

Analysis step (carnal and ptraj)

In the last step, the trajectory files storing the coordinates and velocities at each time step determined in the calculation step were used for static and dynamics analysis. The dynamical properties over the simulation time of the system such as energies, temperature, root mean square deviation, distances, angles, *etc.* were extracted out. *Carnal* and *ptraj* modules are used.

3.2.2 Set up the MD simulations for the five systems

In this study, the MD simulations were carried out for five main systems, CORE, CORE-N, CORE-C and full-length with (FULL+ION) and without (FULL) ions. All five systems run aim to explore the dynamical behavior of each one-

domain, two-domain and the full-length structures. The last two systems were carried out in order to investigate the effect of the ion on both structural and dynamical behaviors of the enzyme. Figure 3.15 shows the solvation of the FULL+ION system. Characteristics of the simulated systems are summarized in Table 3.3. Following detailed calculations were set up.

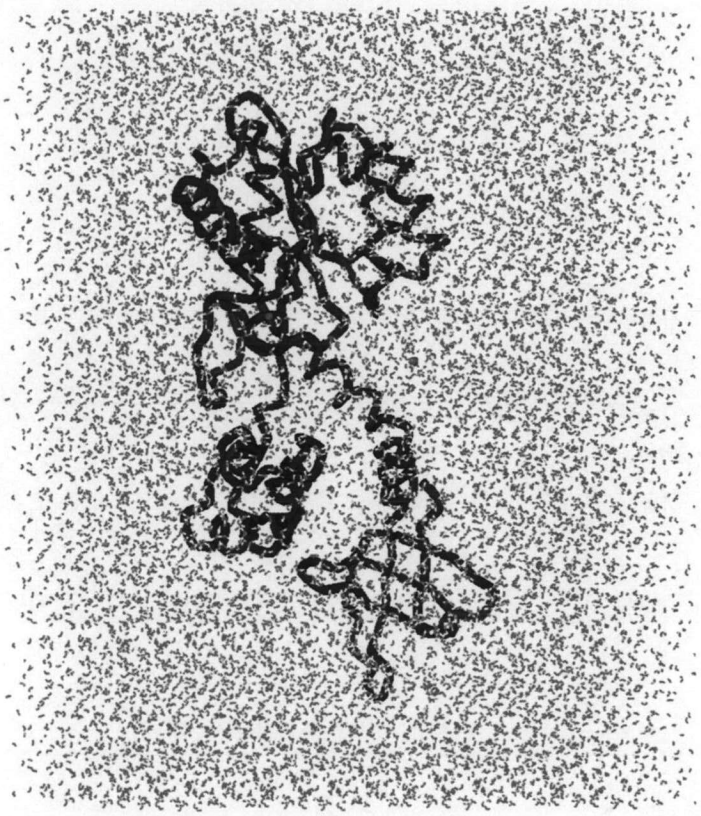


Figure 3.15 The FULL+ION in a solvate box.

Step 1: Addition and energy minimization of hydrogen and missing atoms

The CORE and four homology modeled structures, CORE-N, CORE-C, FULL+ION and FULL (detail as described above) were taken to be a starting structures for MD simulations. All hydrogen atoms and some side chain atoms in amino acid residues which are not available in the crystal structure were added according to the residue templates. In addition, the ionization state of the ionizable residues (*i. e.*, Asp, Glu, Arg, Lys and His), and the charges at N-terminal residue, Phe1, and C-terminal residue, Asp270, were taken into consideration. The new modified structures as well as the topology files generated were then saved for the whole calculations.

Step 2: Addition and energy minimization of water molecules and counter ions

All structures were solvated in rectangular boxes of water. Numbers of Cl^- ions were added in order to keep the overall system neutral. Characteristics of the simulated systems are summarized in Table 3.3. The following partial relaxations were applied while keeping the remaining structure fixed: first the hydrogen atoms, the newly added residues, and then the solvent molecules in order to reduce bad atomic contact.

Step 3: Minimization of the system

The energy of the system was minimized for 100 steps using the steepest descent method followed by another 9900 steps using the conjugate gradient method to relax the systems prior to the next step.

Step 4: Thermalization, Equilibration and Long run MD simulation

The systems were thermalized to the temperature of 300 K for 500 ps and then equilibrated for 1 ns. The long-range electrostatic interaction beyond 10 Å was cut off. 2 ns MD simulations were carried out using the AMBER7 package with the new force field of Carlos Simmerling *et al.* **(96)**. The MD simulations were carried out in an isobaric-isothermal ensemble (NPT) with a constant pressure of 1 atm and the temperature of 300 K. The external coupling bath **(97)** with coupling constant of 0.2 ps was employed in order to keep the pressure constant throughout the simulations. A 2-fs time step was applied with SHAKE to constraint all bonds involving hydrogen atoms. Snapshots of coordinates of enzyme structure and water molecules were collected every 10 ps along the 2 ns MD simulations.

Step 5: Analysis of the trajectories

The trajectories of all system were analyzed using the PTRAJ and CARNAL modules in AMBER7. Some thermodynamics properties of the system vary with the simulation time were extracted from the MD trajectories. Three dimensional structures of the enzyme were visualized using the programs Molmol **(98)**, Rasmol **(99)** and Swiss-PDB Viewer **(100)**. The quality of the protein structures were evaluated using the Procheck program **(101)**.

Table 3.3 Number of water molecules, cations, counter ions, water molecules and total atoms together with the dimension of simulation boxes of the seven simulations, CORE, CORE-N, CORE-C, FULL+ION and FULL.

System	Number of				Dimension of the simulation box ($Å^3$)
	Cations	Counter ions	Water molecules	Total atoms	
CORE	Mg^{2+}	1 Cl^-	7104	23780	68.71 x 69.84 x 63.40
CORE-N	Mg^{2+}, Zn^{2+}	1 Na^+	9717	32366	88.59 x 68.10 x 68.61
CORE-C	Mg^{2+}	2 Cl^-	10908	36141	61.98 x 76.28 x 97.70
FULL+ION	Mg^{2+}, Zn^{2+}	7 Cl^-	12854	41940	104.00 x 66.05 x 77.69
FULL	-	3 Cl^-	12435	41562	103.07 x 64.74 x 77.69

3.3 Molecular Docking

The tetramer model proposed in this study was used as a starting point for DNA binding. The five mutation points were change to wild type residues using the Insight II program **(94)**. The wild type tetramer with divalent metal ions in the active site regions and the viral and host DNA were then used, respectively, as a receptor and ligand. The Hex, an interactive and molecular superposition program, was used for docking calculation. Polar Fourier correlations were used in this program to accelerate docking calculations **(74)**. Both ligand and receptor are of full rotation. The shape and the grid potential were taken into consideration.

Step for docking calculations using Hex program is described as follow:

Step 1 Load up the receptor molecule

The molecules of dimer and tetramer HIV-1 IN were used as receptor for binding with DNA (see 3.1 for more details).

Step 2 Load up the ligand molecule

Both viral and host DNA were taken into account for the integration process in which catalyzed by the HIV-1 IN. The viral DNA for the dimer HIV-1 IN, sequence TAGTCAGTGTGGAAAATCTCTAGCAGT **(102)**, was built up in B-form using nucgen module in AMBER7 and were used as the viral LTR ends for the tetramer HIV-1 IN while the sequence of plasmid pH1IN1: 5' AGATCGGCTAAGATCGG and 5' CCGATCTTAGCCGATCT was used as host DNA **(103)**.

Step 3 Perform docking calculation

In the docking calculation, we perform full rotation for search mode. All correlation type which are shape, grid potential and electrostatics were selected for each docking run. Step size for each run was set to be 0.75 and the grid dimension was 0.6. The ranges for moving both receptor and ligand were set to be maximum values (180). The twist range

for both systems was 360 degree and distance range was 30 Å. The main window and the docking control panel were display in Figure 3.16.

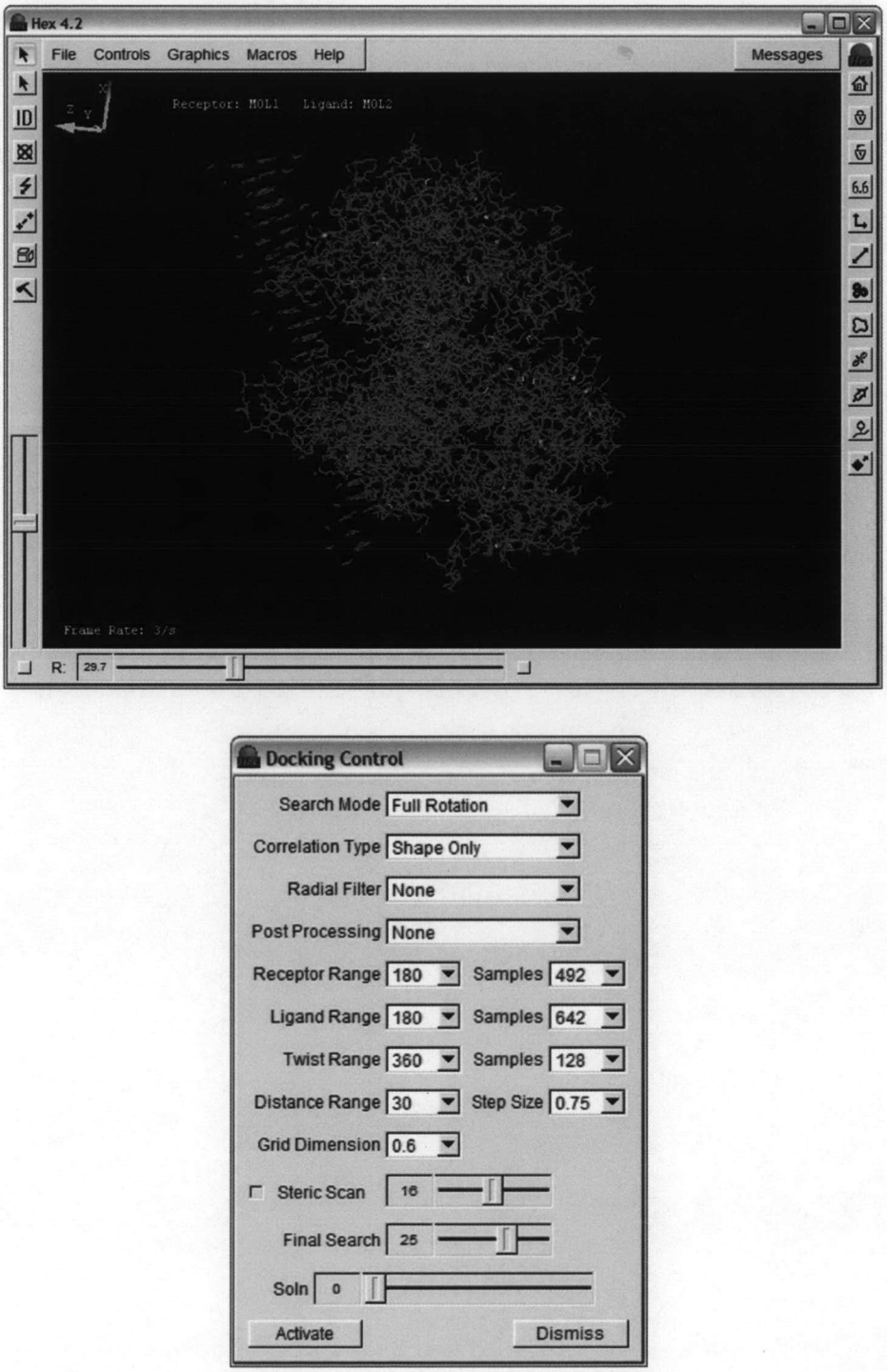


Figure 3.16 The Hex main window and the docking control panel.