CHAPTER 4

RESULTS AND DISCUSSION

4.1 Molecular modeling

4.1.1 Homology modeling of the missing region

4.1.1.1 The one-domain structures

Note that the CORE structure (PDB code 1BL3 chain C) considerably selected in this study was a complete chain, free from missing residues. The N-terminal domain in E-form (PDB code 1WJC, residues 1 – 55), NTER, and the C-terminal domain (PDB code 1IHV, residues 219 – 280), CTER, were also selected to compare with the two-domain structures as well as the full-length one.

4.1.1.2 The two-domain structures

4.1.1.2.1 CORE-N

The CORE-N contains two missing regions, residues 47 – 55 and 140 – 148. The structure including the homology model for these two missing regions was shown in Figure 4.1 which the two regions were zoomed out.

4.1.1.2.2 CORE-C

The CORE-C has only one missing part, residues 142 – 144. With the homology modeling in the Insight II package, the missing region was modeled and shown in Figure 4.2.

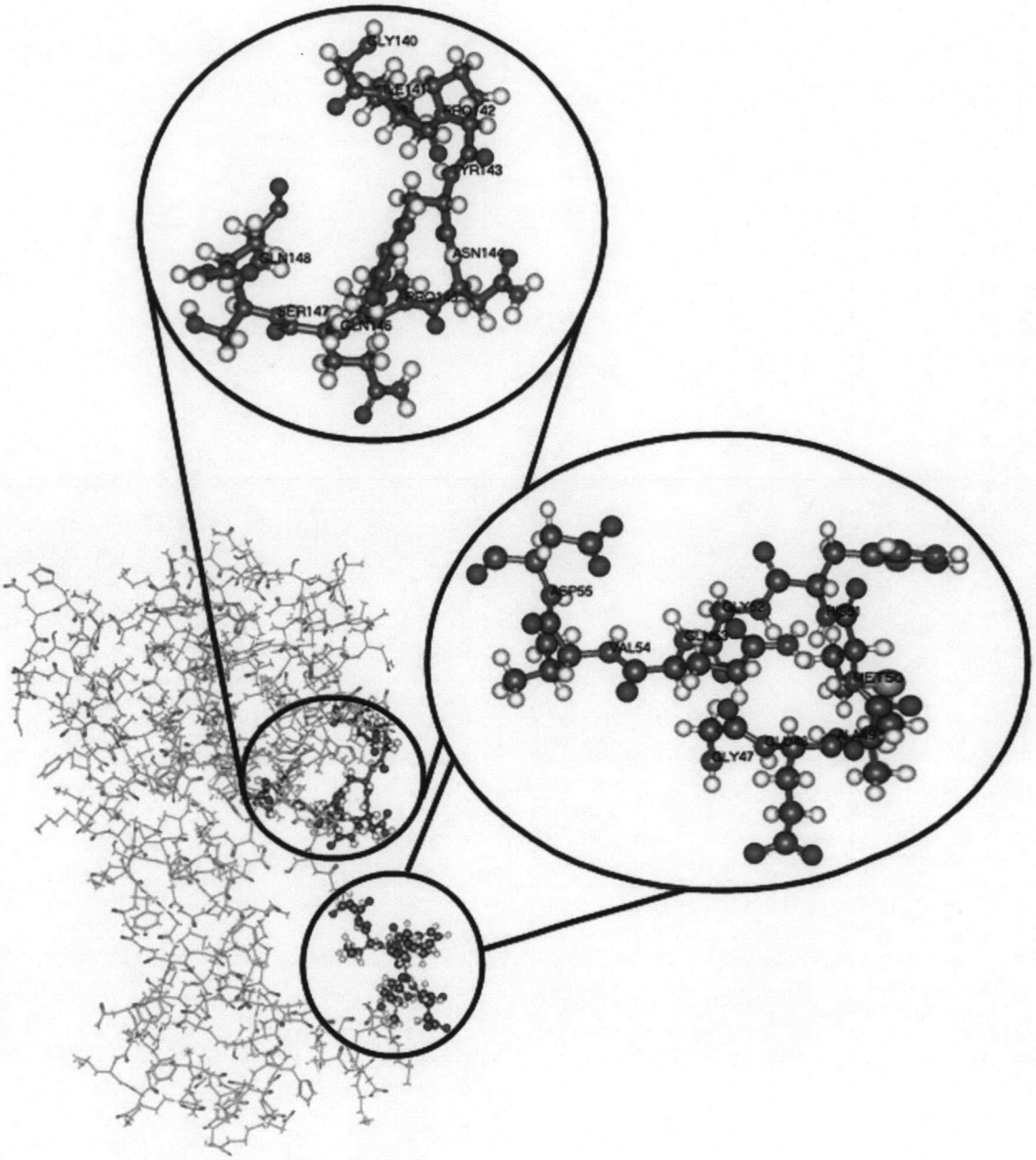


Figure 4.1 Structure of the CORE-N, two missing regions (residues 47 – 55 and 140 – 148) are zoom out and shown in ball and stick model.

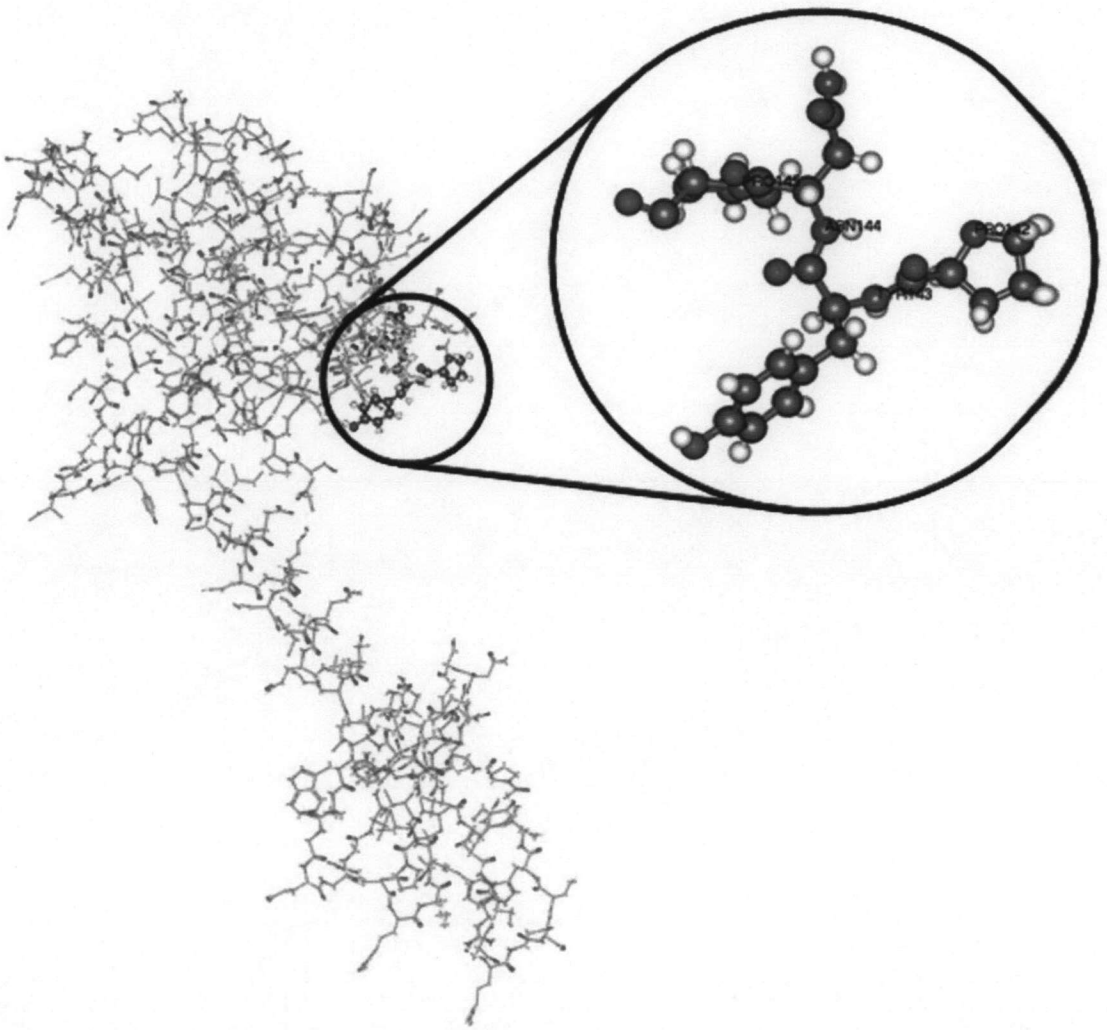


Figure 4.2 Structure of the CORE-C, missing region (residues 142 – 145) is shown in ball and stick model and zoom out.

4.1.2 Modeling of the full-length HIV-1 IN

4.1.2.1 The Full-length monomer

To the best of our knowledge, the full-length structure of the HIV-1 composing of three domains is still not available to date due to the experimentally difficulty resolved. The full-length structure composes of 288 amino acid residues. The amino acid sequence of the full-length structure was shown in Figure 4.3. There are several groups proposed the full-length models, as well as its complex with DNA **(63, 64)**. Schematic representation of the domain full-length structure and the superimposition between two models proposed by Podtelezhnikov *et al.* and Luca *et al.* were shown in Figure 4.4. Although, the synthesized full-length HIV-1 IN from both groups were based on the same experimental data, superimposition of these two proposed structures shows significant differences, especially in the orientations of the C- and N- terminals (Figure 4.4 b). These differences indicate that further studies are needed. The full-length model proposed in this study was displayed in Figure 4.5 (see Chapter 3 for more details)

```
        10         20         30         40         50         60         70
         |          |          |          |          |          |          |
FLDGIDKAQD EHEKYHSNWR AMASDFNLPP VVAKEIVASC DKCQLKGEAM HGQVDCSPGI WQLDCTHLEG

        80         90        100        110        120        130        140
         |          |          |          |          |          |          |
KVILVAVHVA SGYIEAEVIP AETGQETAYF LLKLAGRWPV KTIHTDNGSN FTGATVRAAC WWAGIKQEFG

       150        160        170        180        190        200        210
         |          |          |          |          |          |          |
IPYNPQSQGV VESMNKELKK IIGQVRDQAE HLKTAVQMAV FIHNFKRKGG IGGYSAGERI VDIIATDIQT

       220        230        240        250        260        270        280
         |          |          |          |          |          |          |
KELQKQITKI QNFRVYYRDS RNPLWKGPAK LLWKGEGAVV IQDNSDIKVV PRRKAKIIRD YGKQMAGDDC

    288
     |
VASRQDED
```

Figure 4.3 Amino acid sequence of the full-length HIV-1 IN **(62)**.

(a)

N-terminal domain	Core domain	C-terminal domain
HTH fold (HHCC)	Rnase fold (DD35E)	SH3 fold
1 50	51 212	213 288

(b)

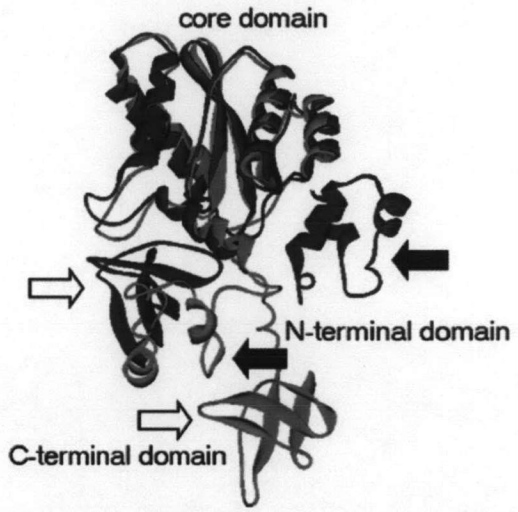


Figure 4.4 (a) Schematic representation of the full-length structure and (b) superimposition of the two available proposed full-length models **(63, 64)**.

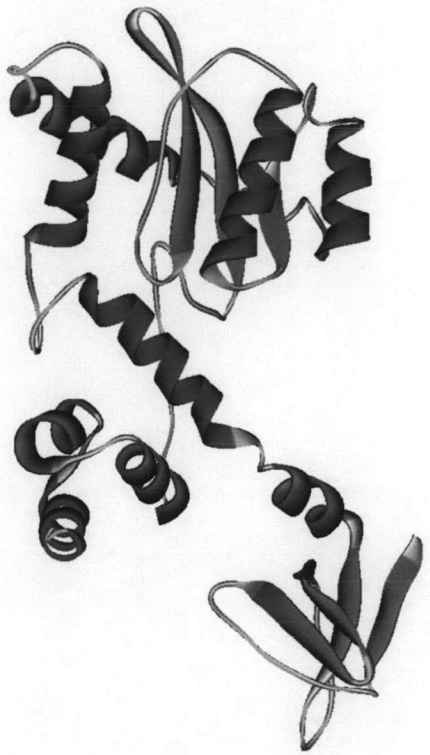


Figure 4.5 Full-length structure proposed in this study.

4.1.2.2 The full-length dimer

Full-length dimer HIV-1 IN was built with the same procedure in which the based on the crystal structure of the CORE-C (Figure 4.6). The two monomer subunits of the core and the N-terminal domains have some interfaces. The N-terminal domain of two monomer subunits interacts with each other. This supports the evidence that the zinc binding domain promotes tetramerization **(32, 64)**. Unlike those two domains, each monomer chain of the C-terminal domain point away from each other. The active site of both monomer units were shown in Figure 4.7.

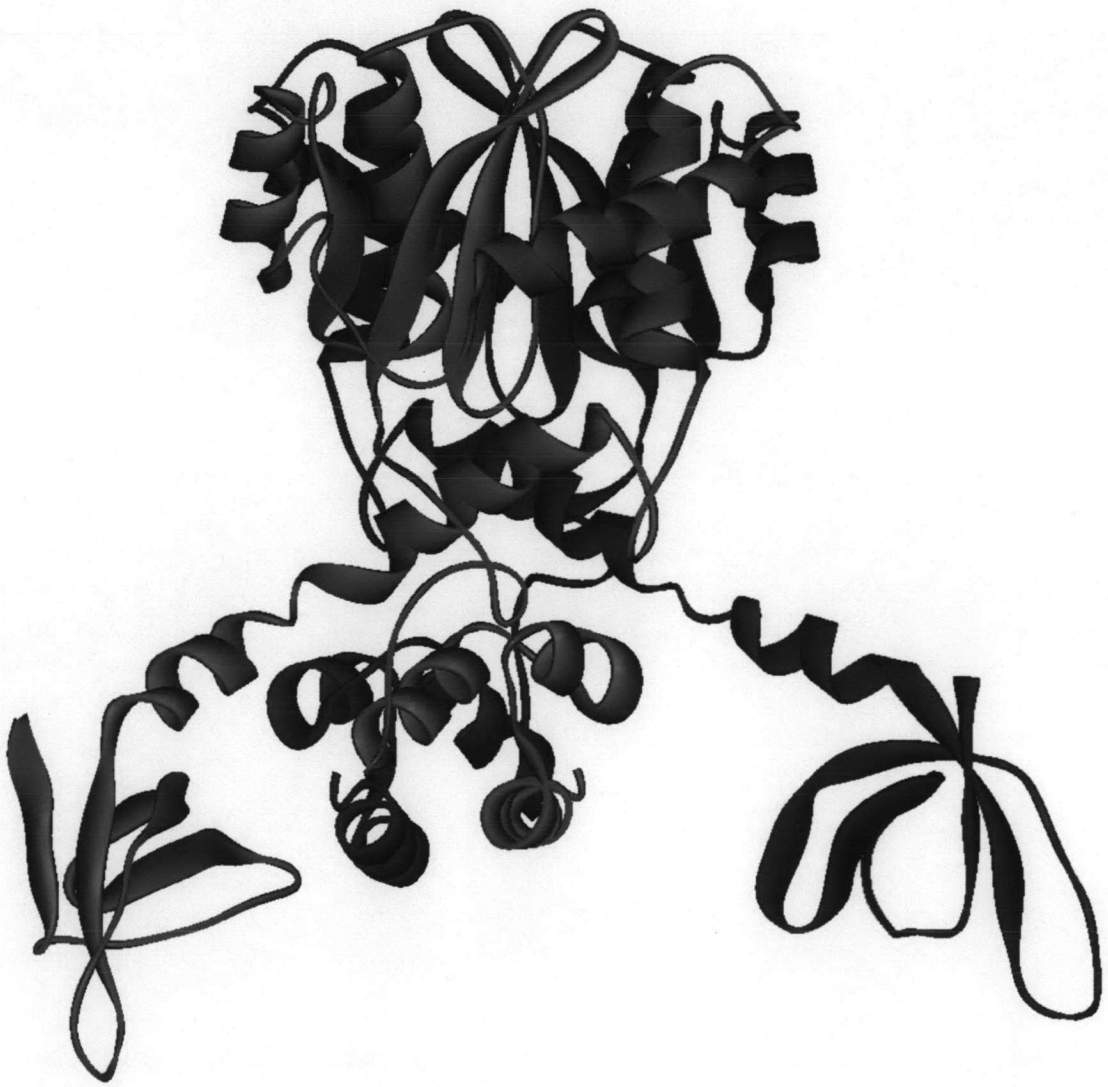


Figure 4.6 Dimer model of the full-length HIV-1 IN, chain A (red) and chain B (blue).

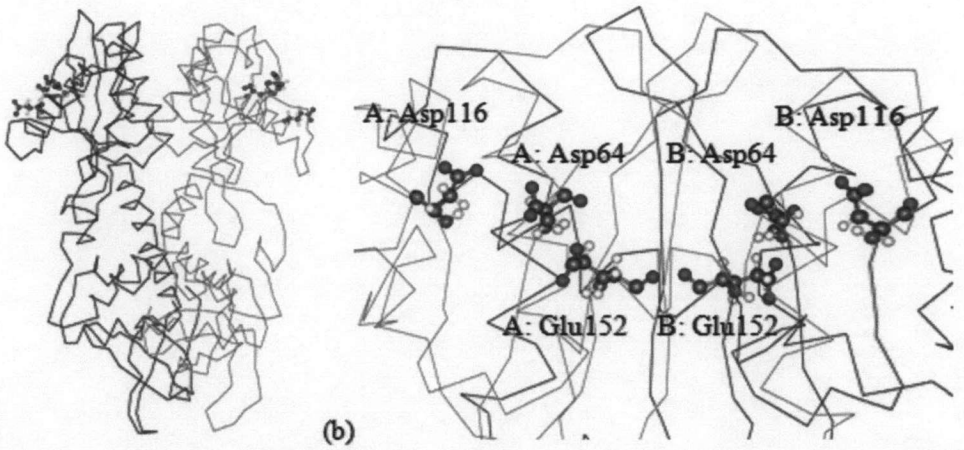


Figure 4.7 The orientation of the catalytic residues (Asp64, Asp116 and Glu152) in the core domain region of both chain A (red) and chain B (blue). The three acidic residues were shown in ball and stick model and labeled.

4.1.2.3 The full-length tetramer

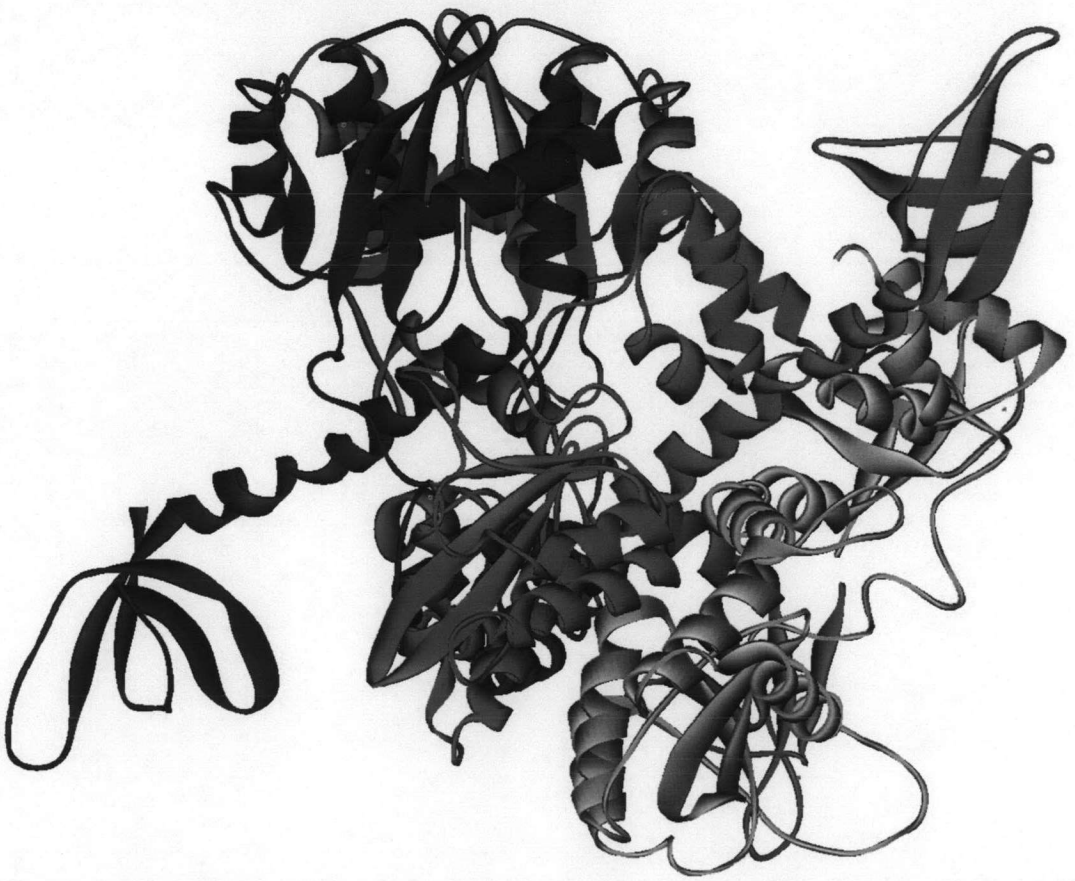


Figure 4.8 Tetrameric form of the full-length HIV-1 IN, chain A (red), chain B (blue), chain C (green) and chain D (yellow).

The structure of the tetrameric form of full-length HIV-1 IN (Figure 4.8) was modeled in this study based on the crystal structures of the two-domain structures, CORE-N and CORE-C with the same procedure.

4.1.3 Comparison of modeling structure

Seven starting structures which are CORE, NTER, CTER, CORE-N, CORE-C, FULL and FULL+ION were physically compared. By considering individually one domain, three sets of comparisons were made based upon, *(i)* core region, *(ii)* N-terminal region and *(iii)* C-terminal region. The analyses were described as the following. Note that the structures were optimized after adding the missing atoms.

4.1.3.1 The core domain

Considering the structure of the core domain from various systems, residues 56 – 209, superimposition studies reveal that the missing region (residues 140 – 148) in the core connected to the terminal end domains for the CORE-N, CORE-C, FULL and FULL+ION systems are different from the core only domain, CORE (Figure 4.9).

The different between the CORE and the core domain in the two-domain structure, CORE-N (black line) and CORE-C (red line) are in the same range (2.0 Å < RMSD < 7.0 Å) which is lower than that of the core from the three-domain ones (2.0 Å < RMSD < 10.0 Å), FULL (green line) and FULL+ION (blue line). The big different of the core region between the two-domain and the three-domains structures is also found in the flexible region, residues 188 – 192 (Figure 4.9). The dissimilar among the CORE-C, FULL and FULL+ION was found in the range of 1.5 – 4.0 Å whereas this value is higher for the CORE-C structure, *i. e.*, RMSD values reach up to 9.0 Å. Note that the first few residues, residues 56 – 58, which is connection region to the N-terminal domain of the CORE-N and CORE-C are significantly different from the CORE. These observations are clearly seen in Figure 4.10.

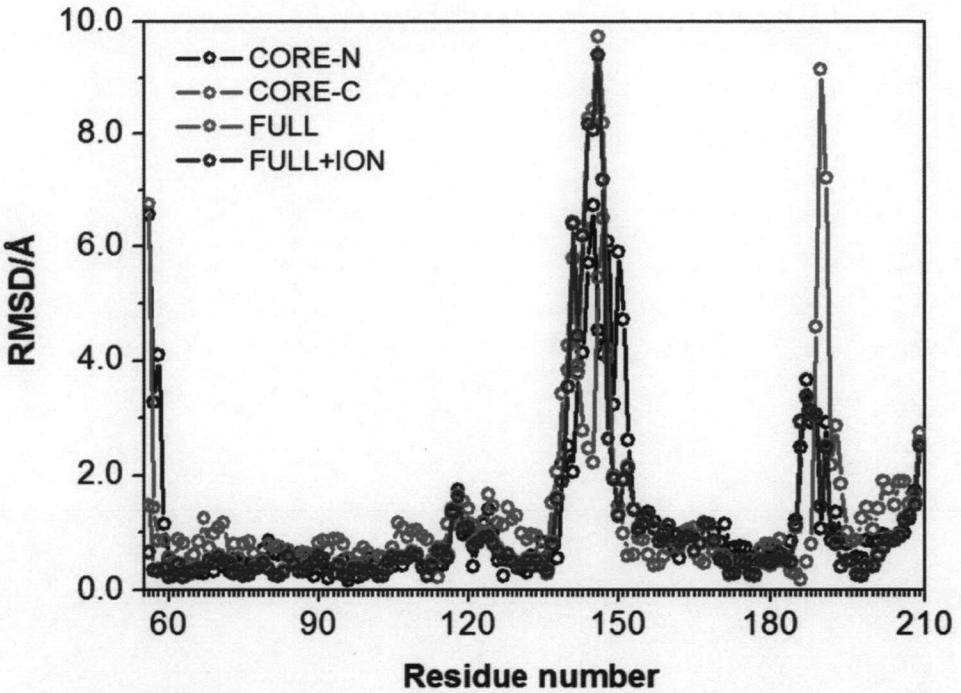


Figure 4.9 Superimposition between the core only domain, CORE and various core structures from the following structures; CORE-N (black line), CORE-C (red line), FULL (green line) and FULL+ION (blue line).

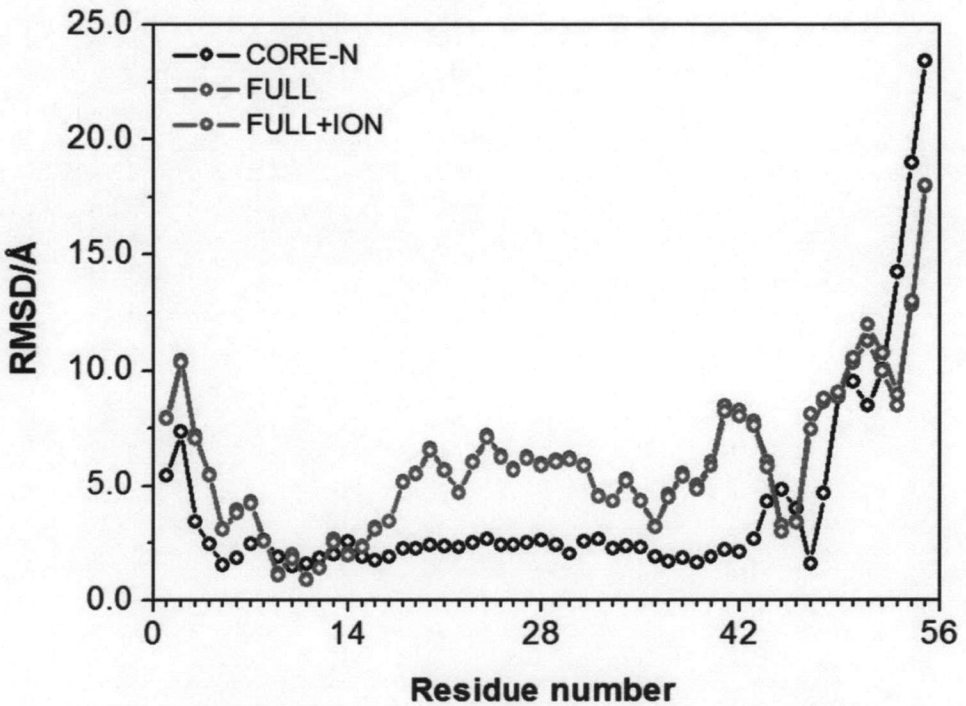


Figure 4.10 Superimposition of the various N-terminal regions in CORE-N (black line), FULL (red line) and FULL+ION (green line) onto the NTER only domain.

4.1.3.2 The N-terminal domain

Figure 4.10 shows the superimposition plot of the N-terminal regions from various structures; CORE-N (black line), FULL (red line) and FULL+ION (green line) with the one-domain NTER. The N-terminal region in the two-domain is less differing from the NTER in residues 14 – 42 than that of the three-domain ones. While the differences increase in the core connection region, residues 50 – 55 in which the RMSD is in the range of 10 – 25 Å (Figure 4.10).

4.1.3.3 The C-terminal domain

The comparisons were made in the same manner for the C-terminal region. The one-domain, CTER, was used as a referent point for the superimposition of the C-terminal region in CORE-C (black line), FULL (red line) and FULL+ION (green line), see Figure 4.11. The differences between these domains in various structures are in the same range, *i. e.*, RMSD values fluctuate with the same pattern. No difference was observed between the FULL and the FULL+ION in the C-terminal region. Addition of the third parts, N-terminal domain, seems to affect the structure of the C-terminal domain.

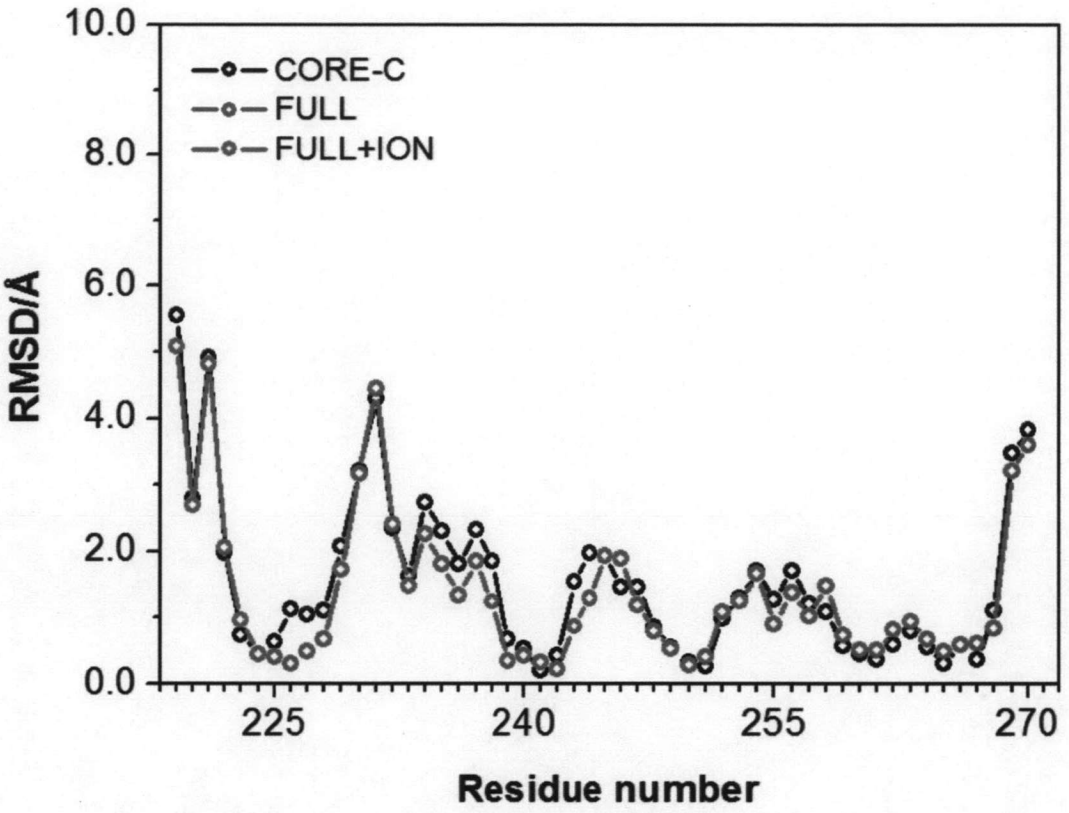


Figure 4.11 Superimposition of the various C-terminal regions in CORE-C (black line), FULL (red line) and FULL+ION (green line) onto the CTER only domain.

4.1.3.4 The two-domain structure

Since the core connected to the N-terminal part of the CORE-N was used to build up the full-length model, the differences between the full-length models were found in two regions, residues 47 – 55 and 140 – 148, which are the two missing regions. With the RMSD values of 5 – 22.5 Å, the connection area between the N-terminal domain and the core domain are greatly differing, while the second missing region has lower RMSD. The two regions which have high RMSD values were shown in Figure 4.12. Except for those two regions, the RMSD values for the rest residues are lower than 2 Å.

Similar manner were made to the two-domain structure, CORE-C, three big differences were found. The result shows that the connection region to the N-

terminal domain between the CORE-C and the CORE-N (in term of the FULL and the FULL+ION) are diverse. As expect, the 140 – 148 and the 188 – 192 regions show high different. Note that the dissimilar in the first region are less than that of using the CORE-N one as a reference (Figure 4.13, see also Figure 4.12).

4.1.3.5 The full-length

The differences were found in the 47 – 55 linkage region between the N-terminal domain and the core domain and the 140 – 148 missing region (Figure 4.14). There is not any significantly different among the rest part of the structure.

Comparisons were made to our model with the full-length structures proposed by Luca *et al.* **(63)** and Podtelezhnikov *et. al.* **(64)**. Superimposed structures between our model and those two structures were shown in Figure 4.23; note that only monomer subunit A from each structure was taken into consideration. It was clearly seen that overall architecture of our model look similar to that of Luca *et al.*'s one (Figure 4.15 a, b). On the other hand, the structure from Podtelezhnikov et al. has different orientation between the two end domain, the N- and the C-terminal domains (see Figure 4.15 c).

However, there are some detail differences between our full-length and Luca *et al.*'s model, particularly in the linkage connection between the N-terminal and the core domain and the elbow linkage between the core and the C-terminal domain. The RMSD per residue plot (Figure 4.15 b) shows high values in these two regions, as 5 Å < RMSD < 20 Å at the N-core linkage and RMSD above 5 Å in the C-terminal region. Interestingly, the CORE-N and the full-length model has high different in this region. The divergence in this region was thus checked for the CORE-N and Luca *et al.*'s model, the result shows that the orientation of the N-core linkage in the two structures was different (Figure 4.15 d). The difference in the C-terminal region comes from the difference in residues 210 – 219 as the flexibility of this region. Such flexibility was suggested to play role in DNA binding during the integration process **(29)**.

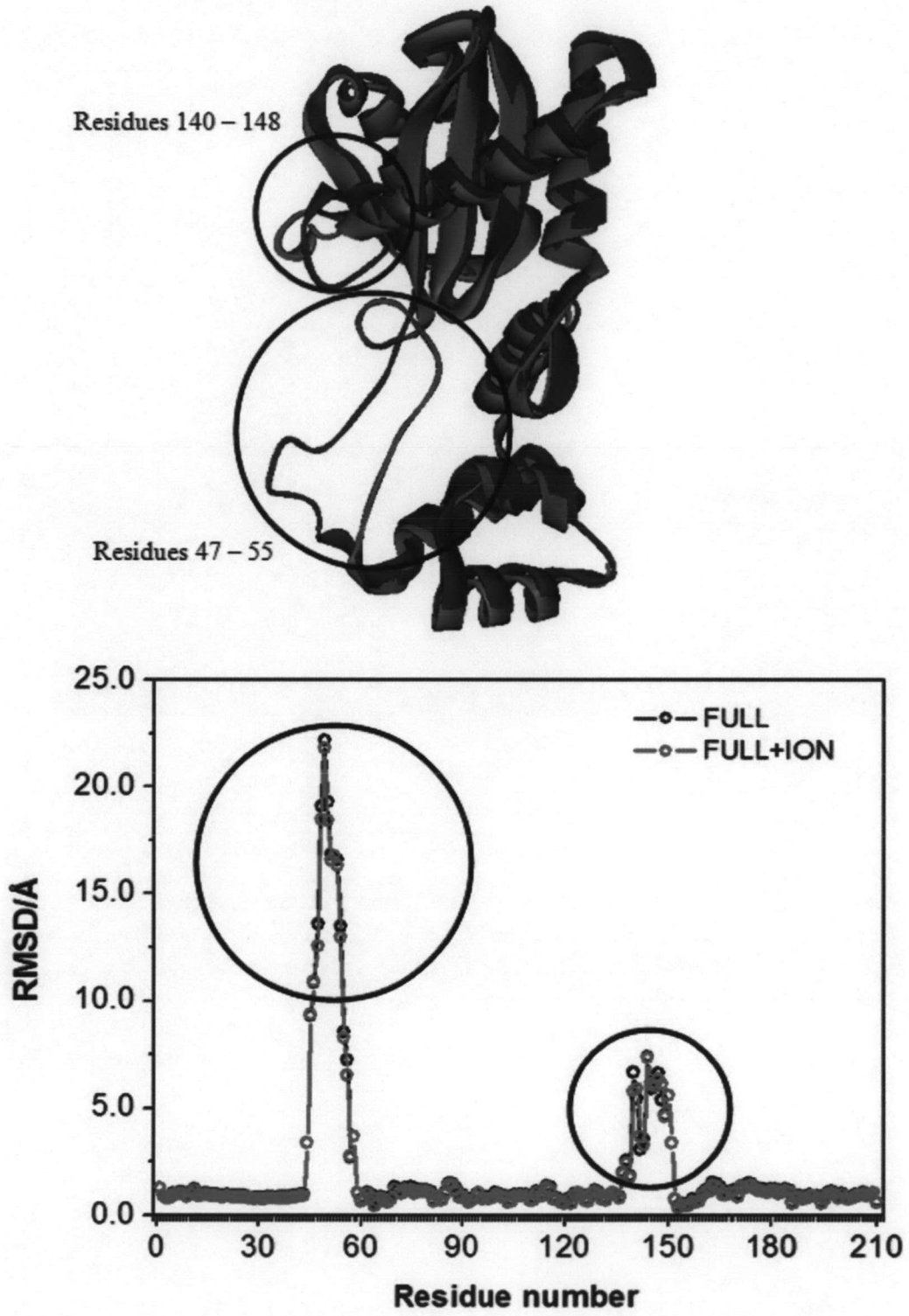


Figure 4.12 Superimposition of the two-domain region, the core connected to the N-terminal domain, CORE-N with those regions in the FULL (black line) and FULL+ION (red line). The two high-RMSD regions were shown in circle.

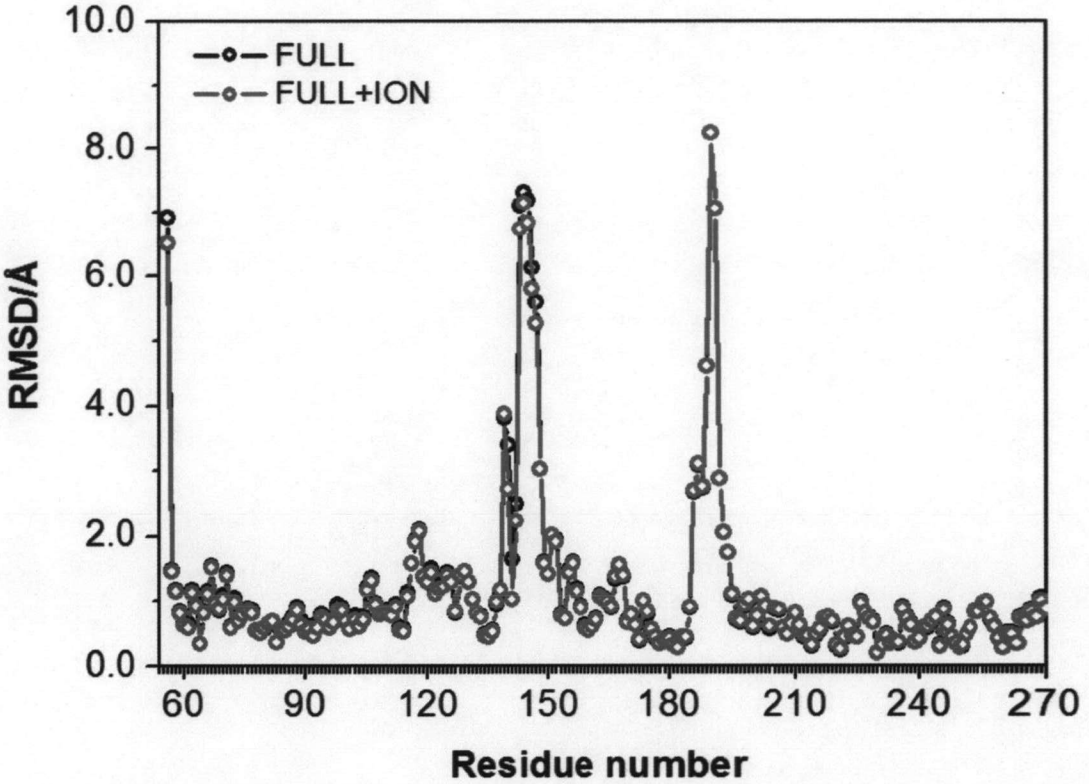


Figure 4.13 Superimposition of the two-domain region, the core connected to the C-terminal domain, CORE-C with those regions in the FULL (black line) and FULL+ION (red line).

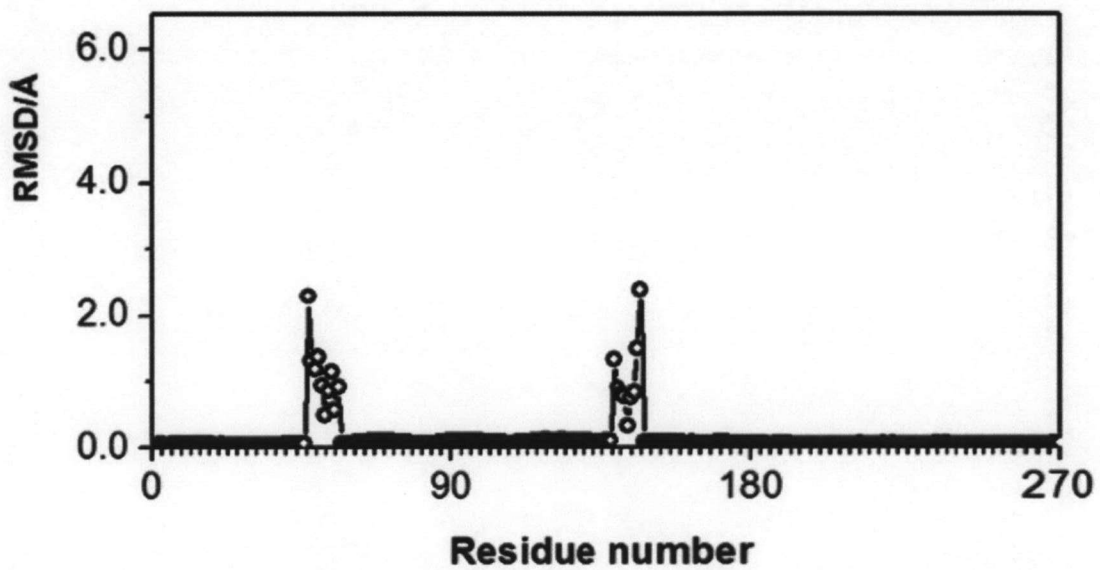


Figure 4.14 Superimposition of the full-length structure without (FULL) and with ions (FULL+ION).

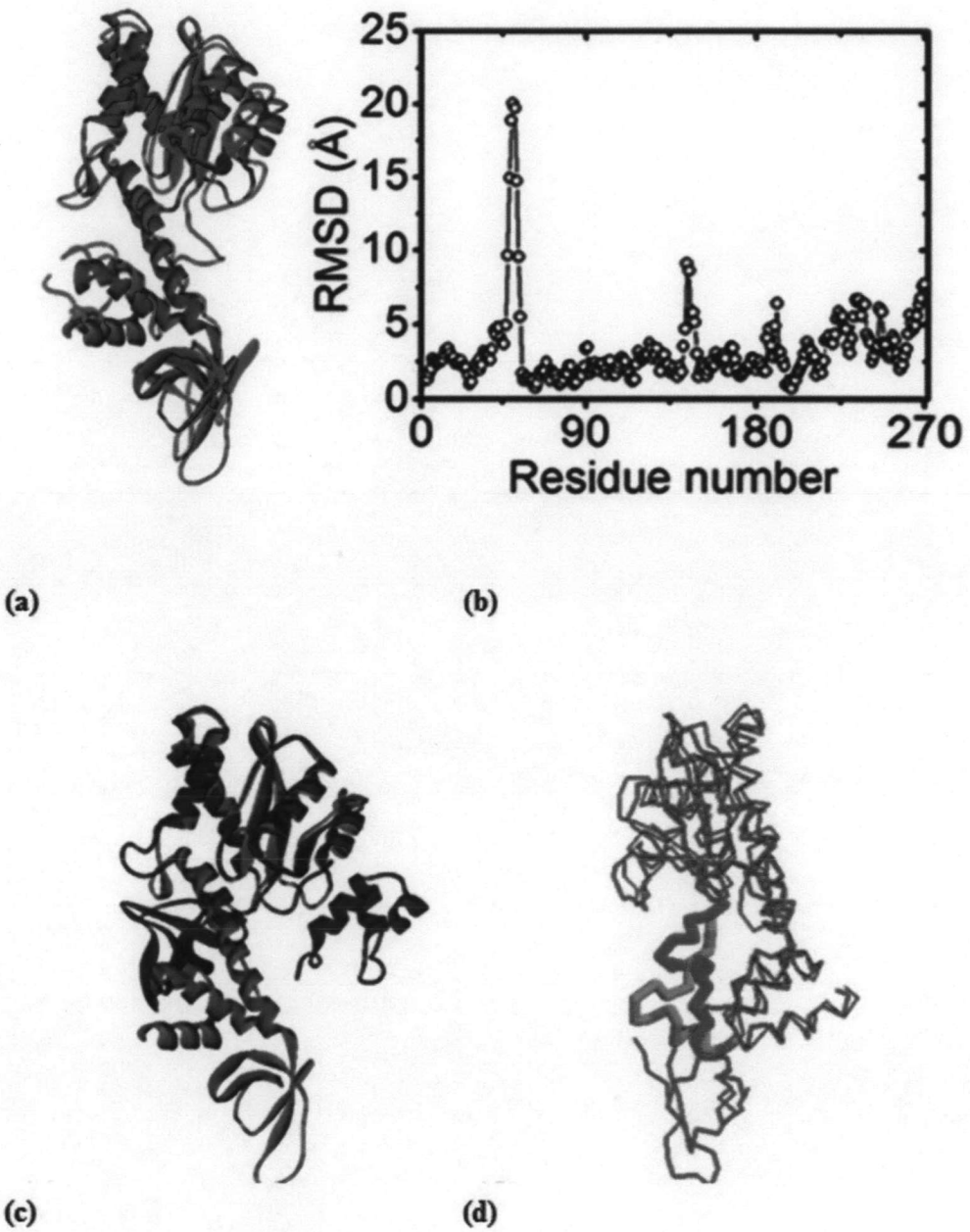


Figure 4.15 Superimposed structures between our FULL+ION model (red) with Luca *et al.*'s model (green) and their correspond RMSD per residue (b), superimposed structure between our model with Podtelezhnikov *et al.*'s structure (blue) (c). The different between the core-N linkage in the CORE-N (orange) and Luca *et al.*'s (green) structure was also displayed (d).

4.2 MD simulations

4.2.1 The one-domain system

4.2.1.1 CORE

A plot of total, potential and kinetic energies as well as temperature calculated over 2 ns simulation time are shown in Figure 4.16. The results indicate that our system propagates properly. The mean values for the total energy, kinetic energy and potential energy are -5.81 x 10^4, 1.45 x 10^4 and -7.26 x 10^4 kcal/mol, respectively. The mean temperature of the system is 299.98 K.

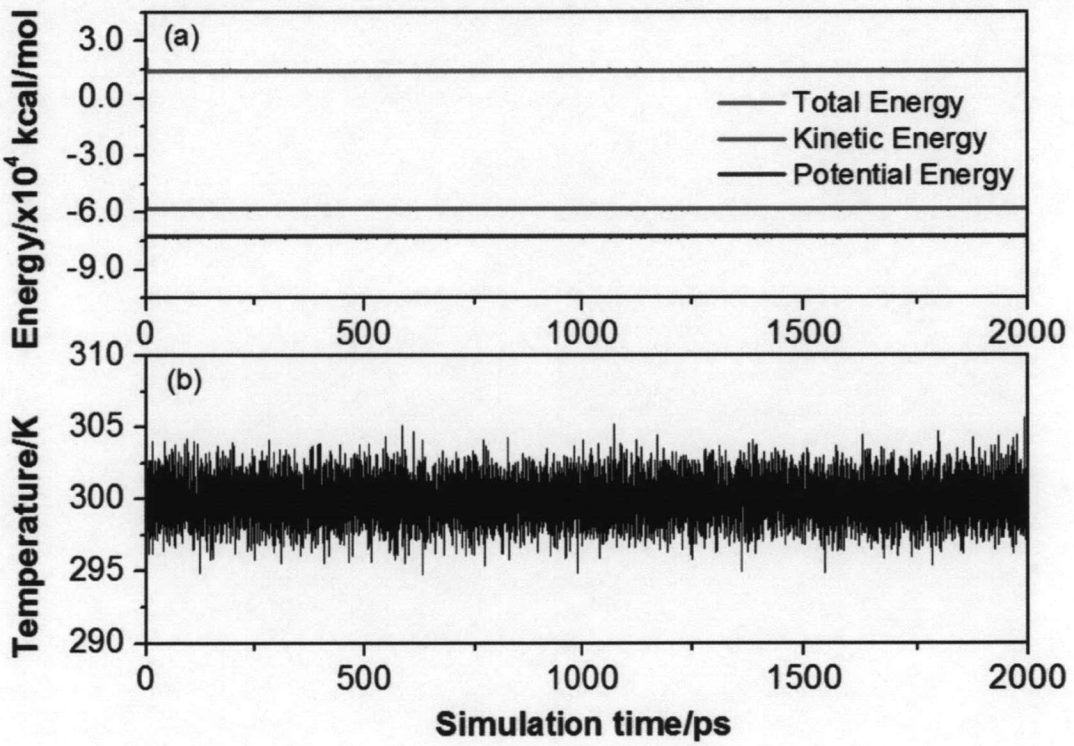


Figure 4.16 The total (red), kinetic (green) and potential (blue) energies (a), and the temperature (b) over the 2-ns MD simulation for the CORE system.

The average structure was generated from the 2-ns trajectory. The structure was then energy optimized and used as the reference structure for calculating the atomic deviation. Figure 4.17 shows a global root mean square deviation (RMSD)

value with respect to the time-averaged structure plotted against the 2-ns simulation time. The RMSD of the CORE system fluctuates in a range of 1.26 – 2.19 Å with the mean value of 1.62 Å.

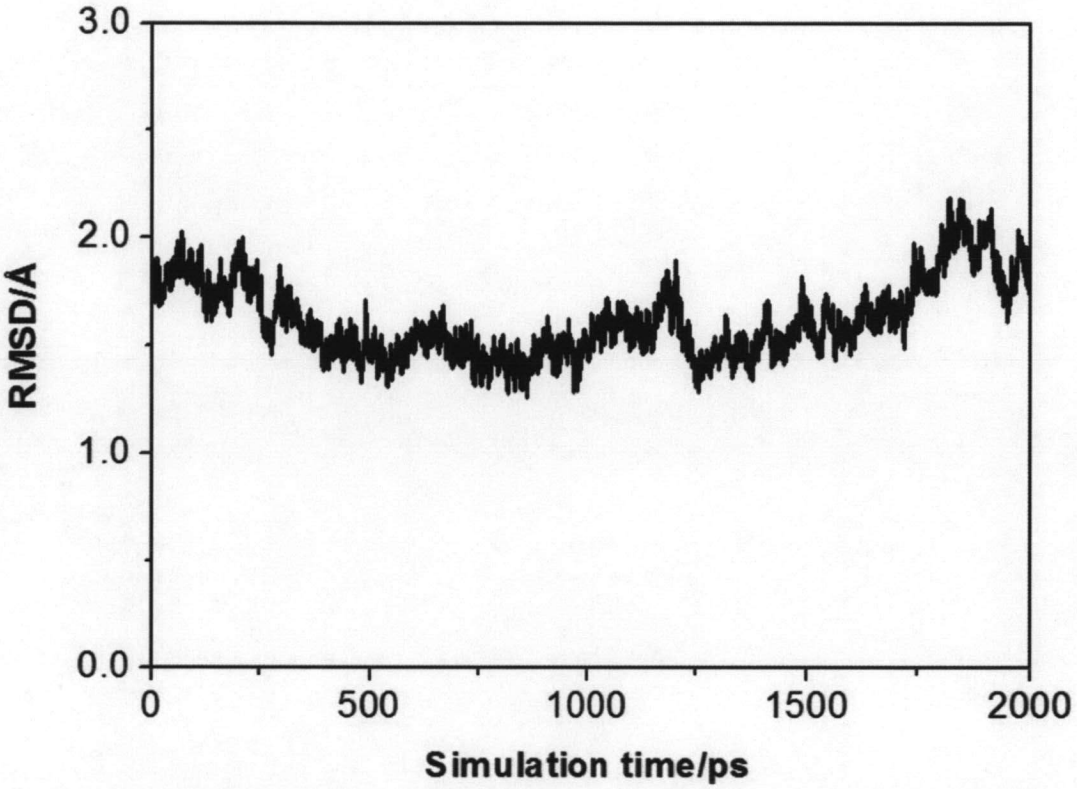


Figure 4.17 Global RMSD value of the CORE system with respect to the average structure over the time range of 2-ns.

Apart from the RMSD value of the whole system with respect to the average structure, the RMSD per residue with respect to the average structure, which enable details analysis on dynamics of the system, was additionally calculated (Figure 4.18). The RMSD values for most residues are in a range of around 0.5 – 2.0 Å. However, there are three regions with RMSD of over 3 Å. The first region with the highest RMSD value is located at residues 188-192. From the experimental data, these residues are flexible loop **(60, 91)**. This is therefore in well agreement with out data indicating that these residues are very flexible. The second region is the residue 210. This is clearly because it is an open end, i.e. it is not connected to C-terminal so it has more degree of freedom. The last region is at around residue 146. This residue is missed in the

experimental structure indicating that it is flexible. 20 structures were snapshooting every 100 ps from 2 ns MD trajectory and superimposed (Figure 4.19). It can be clearly seen that the residues at the connection end as well as the two flexible regions (residues 140 – 148 and 188 – 192) show high fluctuation.

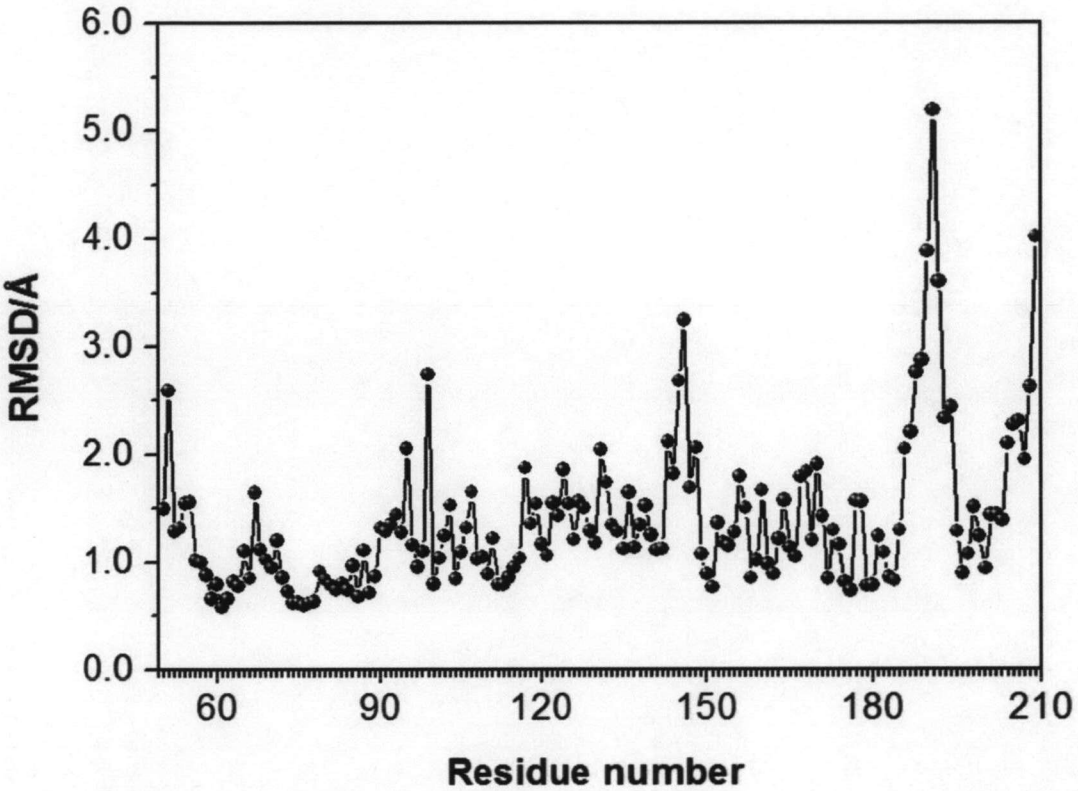


Figure 4.18 RMSD per residue of the CORE system with respect to the average structure over the time range of 2-ns.

Due to the fact that integrase enzyme requires Mg^{2+} for its function **(35)**, dynamics of this metal ion is of interest. Therefore, the distances from Mg^{2+} to OD1 and OD2 atoms of Asp64 and of Asp116 along the 2-ns simulation time were monitored and are displayed in Figure 4.20 a. The OD1 and OD2 atoms of Asp64 and the OD2 atom of Asp116 are closed to the Mg^{2+} with distances of around 1.7 – 2.1 Å. Thus these atoms have strong interactions with the metal ion leading to less flexibility as indicated by small fluctuations of *d1*, *d2*, and *d4* distances. On the other hand, the OD1 atom of Asp116 is quite far away from the Mg^{2+}, i.e. distance of longer than 3.5 Å, so the interaction is quite weak. As a result, this atom has more flexibility and large fluctuation of *d3* distance was

observed. Figure 4.21 shows distances, *d1*, *d2*, *d3*, *d4* and Figure 4.22 shows torsional angle, χ_a measurement for the CORE systems.

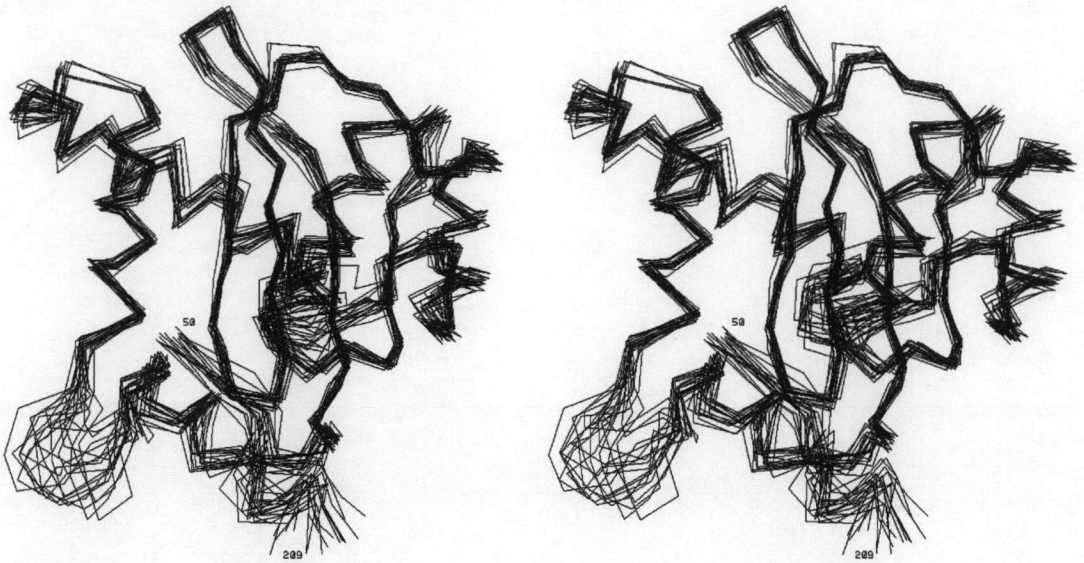


Figure 4.19 Stereo views showing the superimposition of the CORE structure taken every 100 ps through 2 ns MD trajectory.

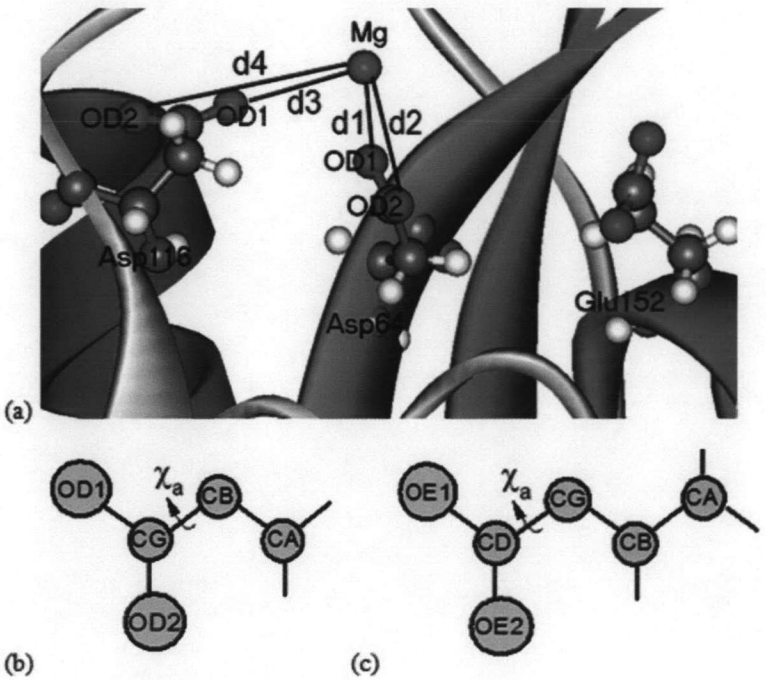


Figure 4.20 Distances between the Mg^{2+} and carbonyl oxygen atoms of Asp64 and Asp116 (a), torsional angle measurement for Asp64, Asp116 (b) and Glu152 (c).

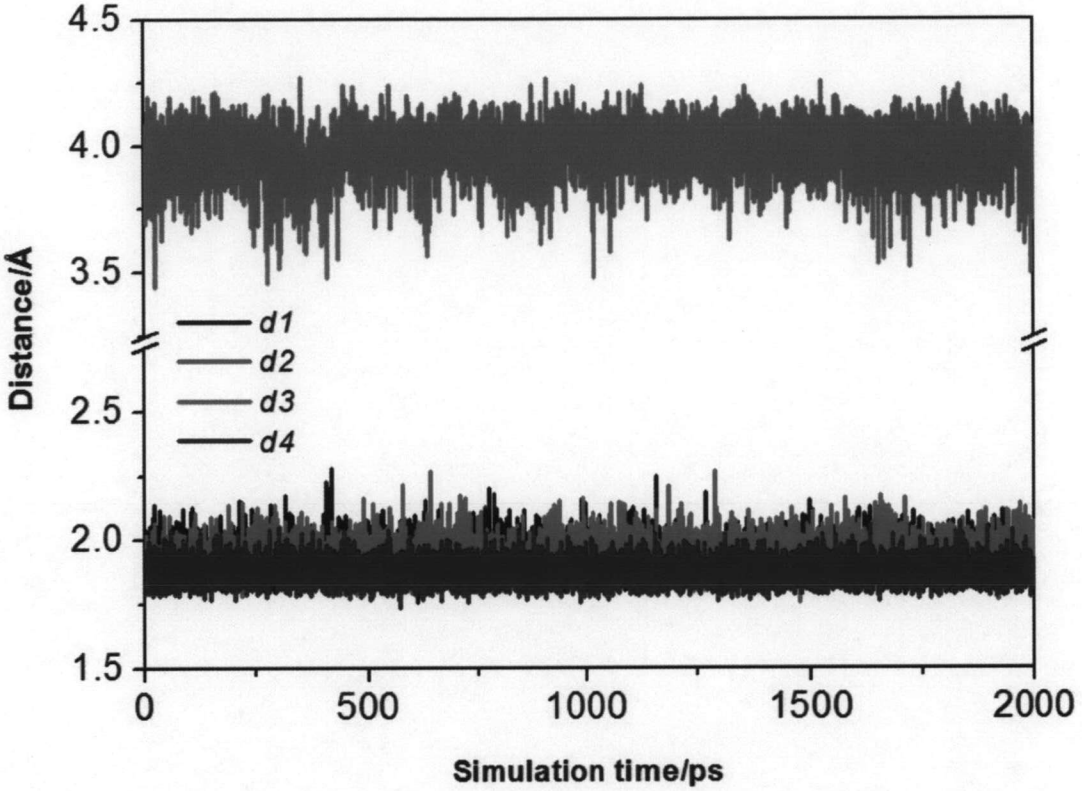


Figure 4.21 Distances between the Mg^{2+} and the carbonyl oxygen atoms of the two aspartate residues Asp64 (*d1* and *d2* as for OD1 and OD2, respectively) and Asp116 (*d3* and *d4* for OD1 and OD2, respectively) of the CORE system.

Dynamics of the catalytic residues: Asp64, Asp116 and Glu152, were investigated by measuring involved torsion angles as shown in Figure 4.20 b and c. The distribution plot of all three angles is displayed in Figure 4.22. The most probable value of both χ_{64} and χ_{116} is 0°. These two torsions have small fluctuations indicating their low flexibility. On the contrary, the χ_{152} has very high flexibility. For example, the torsion flipped from 90° to -90° at around 200 ps. Then, it changed to 0° and to 180°. After that it suddenly rotated back to -90° again. This is possibly because of weak interaction between Glu152 and Mg^{2+} as described above.

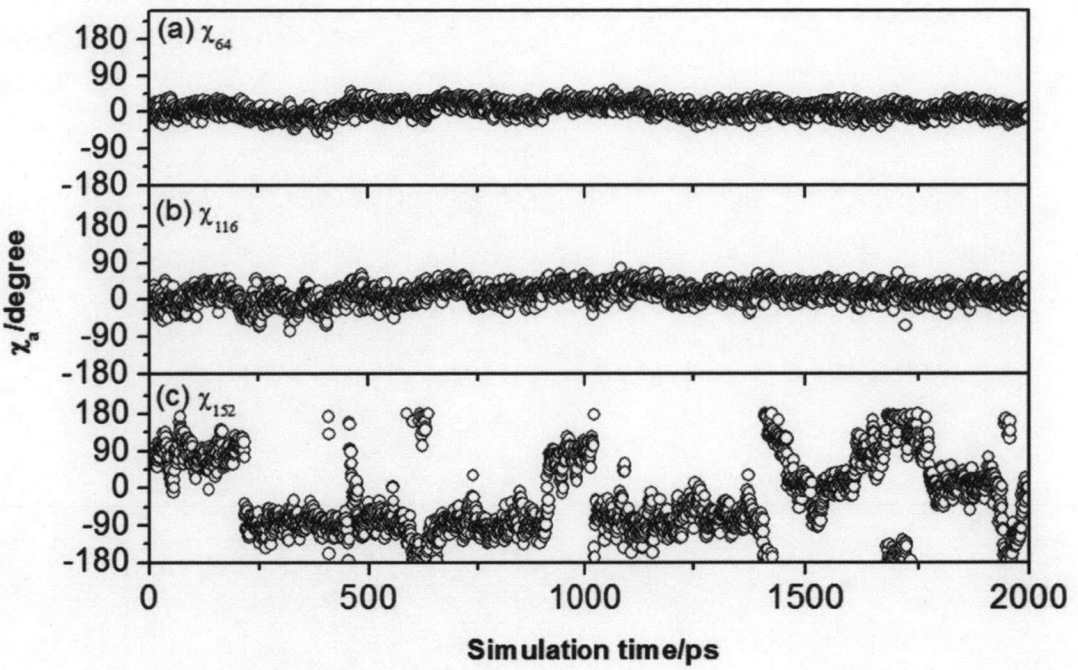


Figure 4.22 Torsional angles (χ_a) plot of three catalytic residues, Asp64 (a), Asp116 (b) and Glu152 (c) of the CORE system.

4.2.2 The two-domain system

Similar to the one-domain system, dynamical properties of the two-domain systems, CORE-N and CORE-C, were investigated in the same manner.

4.2.2.1 CORE-N

The mean values of the total, kinetic and potential energies are -7.98 x 10^4, 1.98 x 10^4 and -9.96 x 10^4 kcal/mol, respectively, and the mean temperature is 299.99 K (Figure 4.23). The RMSD over the simulation time plot start the fluctuation at about 2 Å and appear to increase (2 Å < RMSD < 3 Å) during 700 – 900 ps and converged to fluctuate between 1 – 2 Å (Figure 4.24). Other dynamical properties were investigated in the same manner as the CORE system.

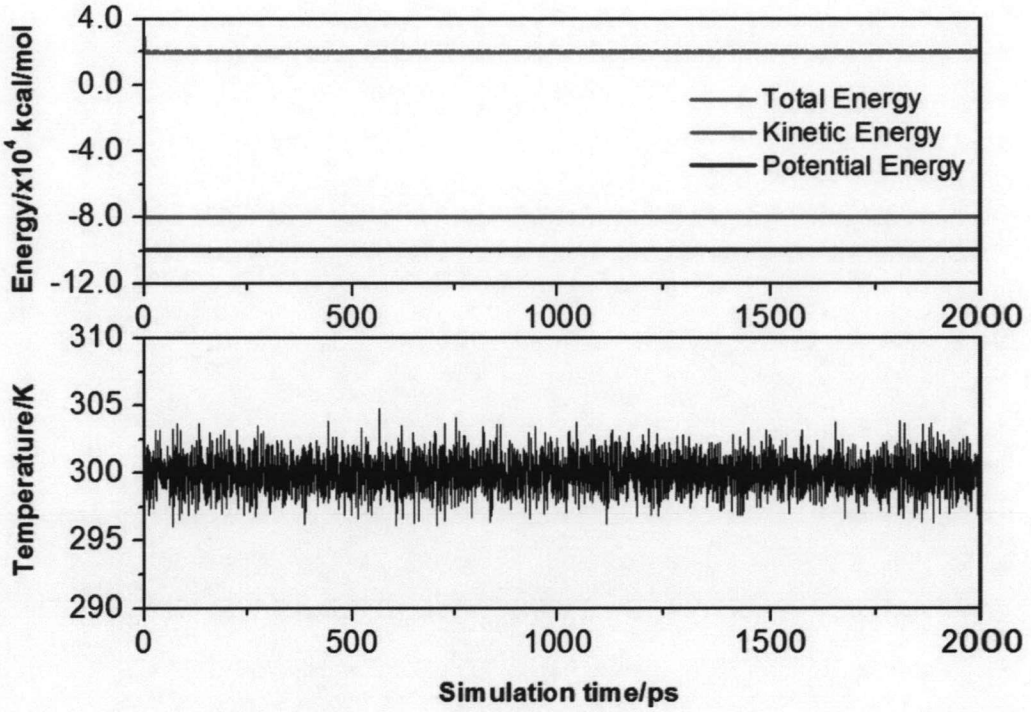


Figure 4.23 The total (red), kinetic (green) and potential (blue) energies (a), and the temperature (b) over the 2-ns MD simulation for the CORE-N system.

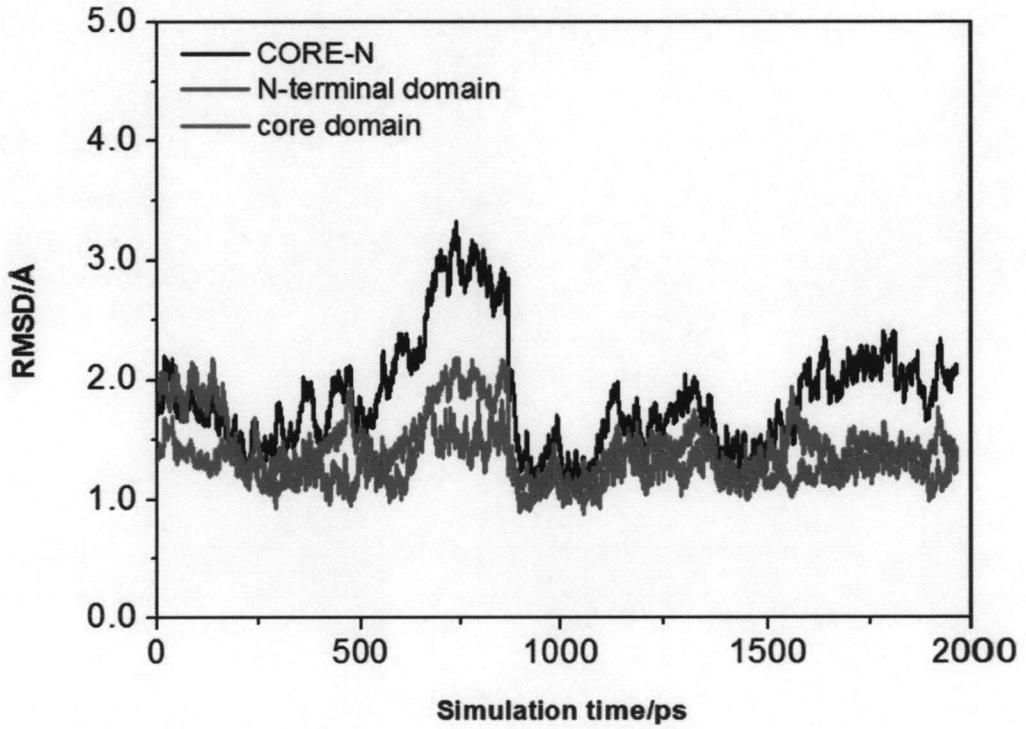


Figure 4.24 Global RMSD value of the CORE-N system with respect to the average structure over the time range of 2-ns.

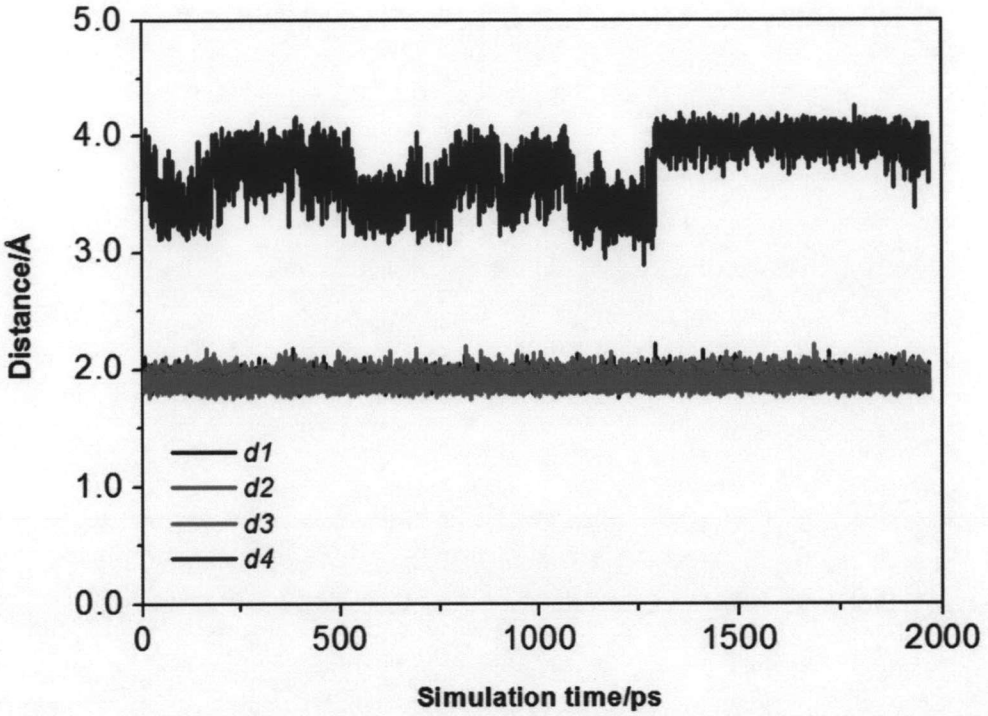


Figure 4.25 Distances between the Mg^{2+} and the carbonyl oxygen atoms of the two aspartate residues Asp64 (*d1* and *d2* as for OD1 and OD2, respectively) and Asp116 (*d3* and *d4* for OD1 and OD2, respectively) of the CORE-N system.

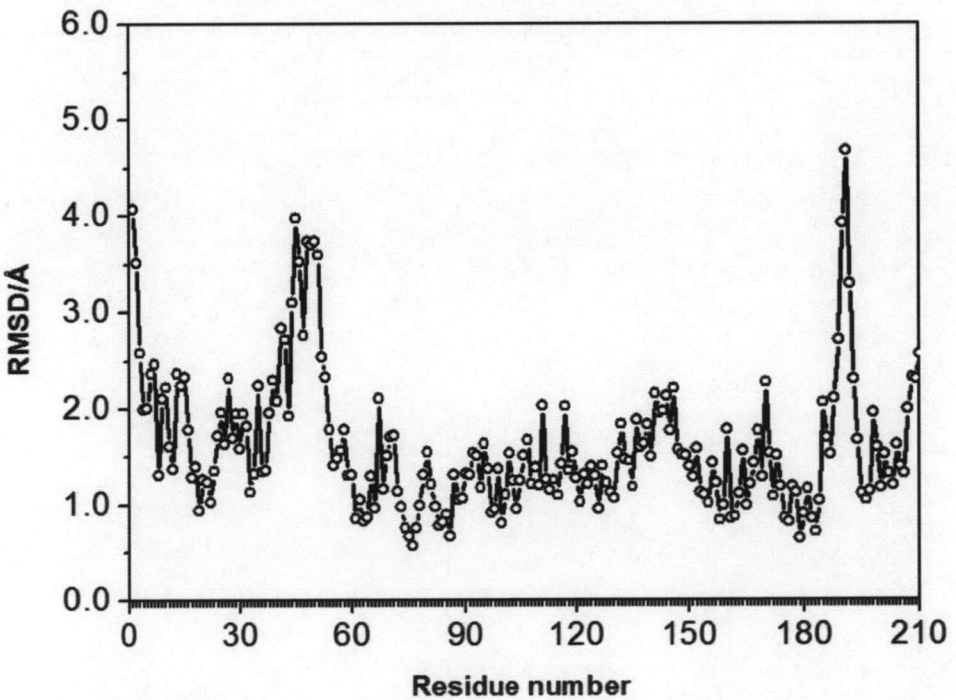


Figure 4.26 RMSD per residue of the CORE-N system with respect to the average structure over the time range of 2-ns.

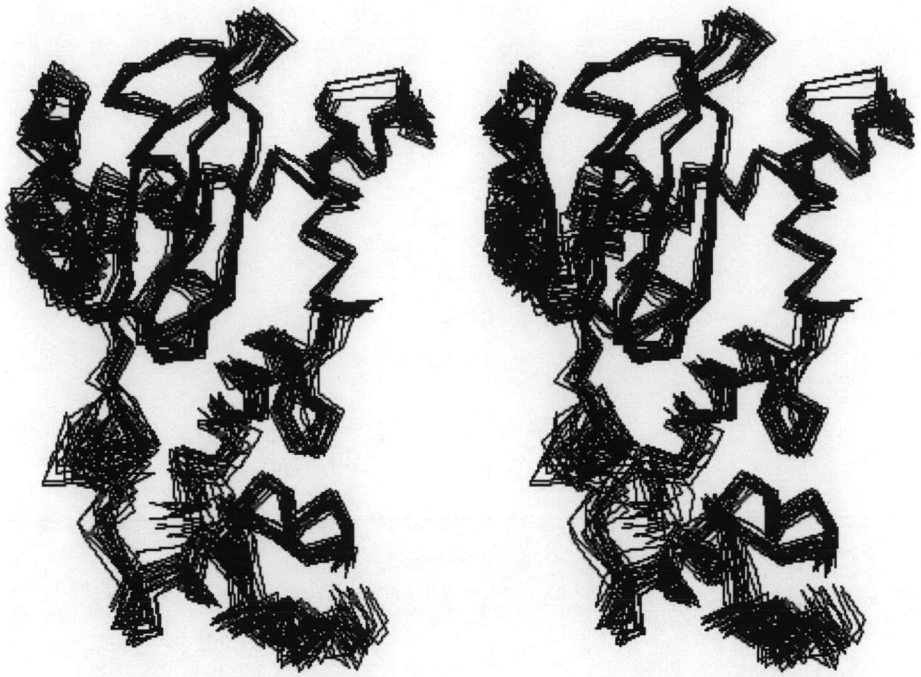


Figure 4.27 Stereo views showing the superimposition of the CORE-N structure taken every 100 ps through 2 ns MD trajectory.

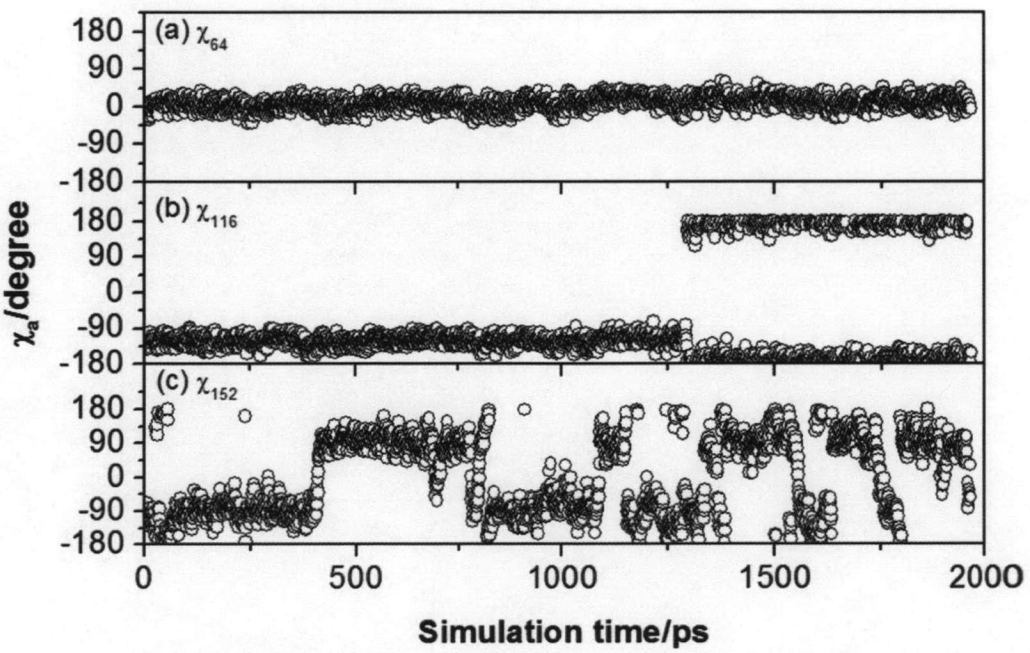


Figure 4.28 Torsional angles (χ_a) plot of three catalytic residues, Asp64 (a), Asp116 (b) and Glu152 (c) of the CORE-N system.

Distance between the Mg^{2+} and the carboxyl oxygen atoms of the catalytic residues, Asp64 and Asp116 as designated by *d1*, *d2*, *d3* and *d4*, were plotted in Figure 4.25. Mg^{2+} bound in the active site region in the octahedral fashion, *i.e.* two aspartate residues bind Mg^{2+} with three oxygen atoms (*d1*, *d2* and *d3*) and allowing three water molecules to coordinate. The RMSD per residue plot shows two regions having high RMSD, residues 47 – 55 and 188 – 192 (Figure 4.26). Interestingly, in this structure, region composing of residues 140 – 148 does not show high fluctuation. 20 Sub-structures taken from MD trajectories superimposed onto together displays that the structures do not change throughout the simulation (see Figure 4.27).

Similar to the CORE system, the two aspartate residues, Asp64 and Asp116, have only one preference. The torsional angle Asp64 is almost the same throughout the entire simulation where $\chi_{64} = 0$ (Figure 4.28 a), while the conformation of the Asp116 change from -135 at 1250 ps to be -180 and stable for the left 750 ps (Figure 4.28 b). The glutamate residue, Glu152, is flexible throughout the whole simulation (Figure 4.28 c). This observation supports the previous data for the flexibility of the Glu152.

4.2.2.2 CORE-C

From a plot of total, potential and kinetic energies and temperature calculated over 2 ns simulation time (Figure 4.29), it is clearly seen that the simulation of CORE-C system was appropriately run. The mean values for the total energy, kinetic energy and potential energy are -8.96×10^4, 2.20×10^4 and -11.15×10^4 kcal/mol, respectively. The mean temperature of the system is 299.91 K.

The RMSD values of each individual domain: core and C-terminal domains, with respect to its corresponding average structure along the 2-ns simulation time were calculated and are shown in Figure 4.30. Both domains have relatively low RMDS values, i.e. 1.0 – 1.5 Å, indicating low flexibility of the structures. In average, the C-terminal domain has slightly higher RMSD than the core domain. Compared to the CORE system with RMSD of around 1.5 – 2.0 Å, core domain in the CORE-C system has lower RMDS value.

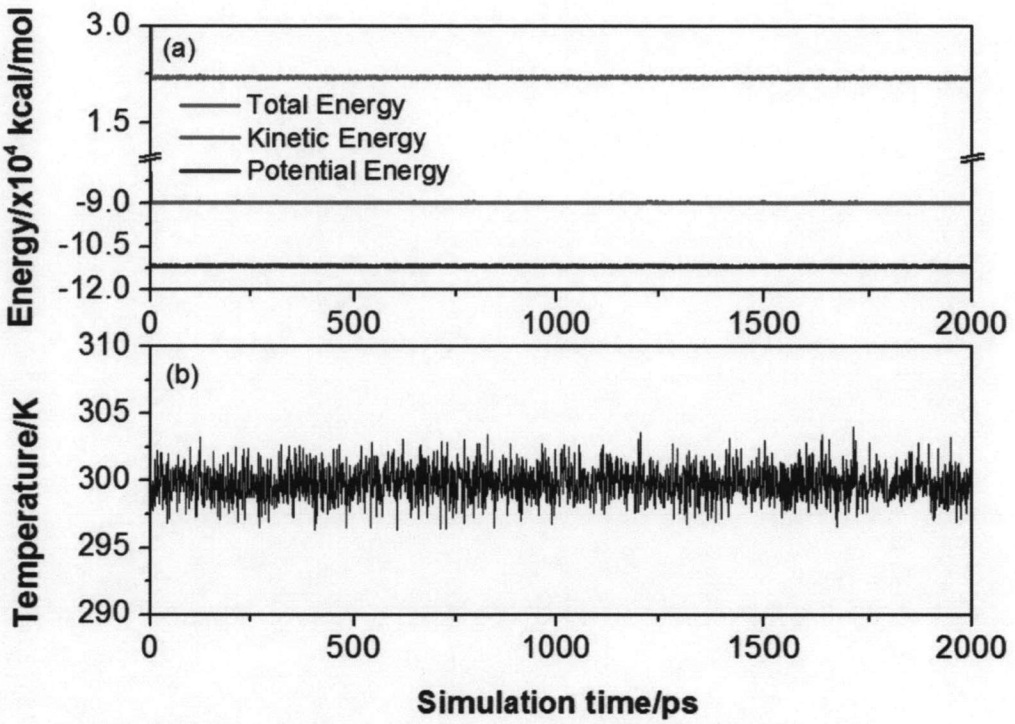


Figure 4.29 The total (red), kinetic (green) and potential (blue) energies (a), and the temperature (b) over the 2-ns MD simulation for the CORE-C system.

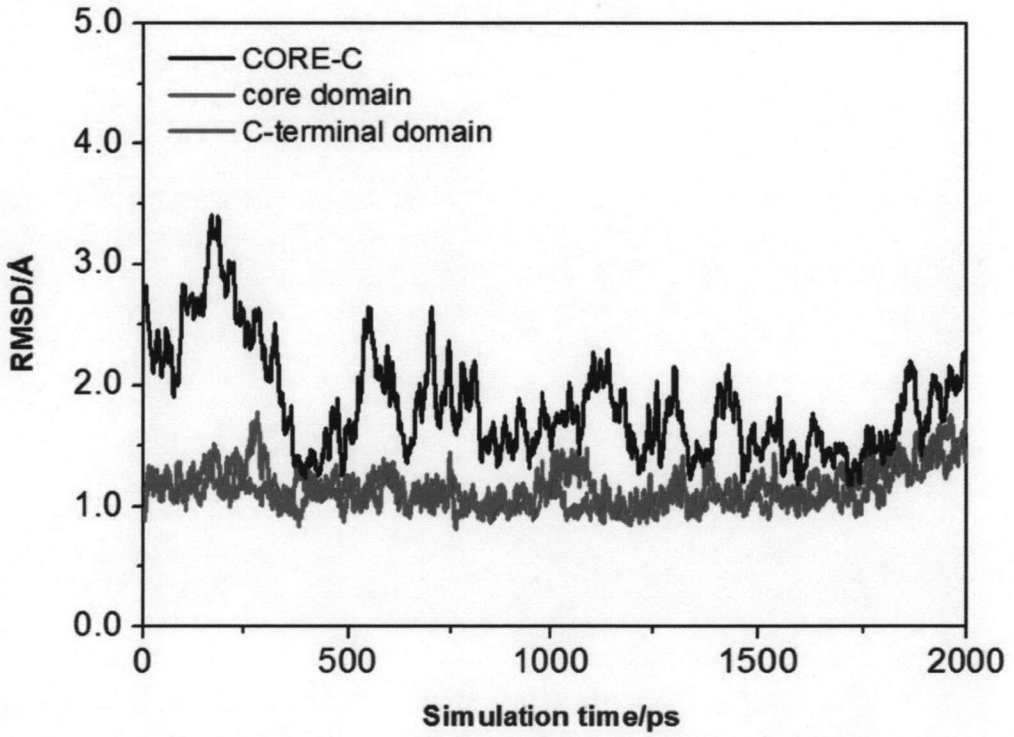


Figure 4.30 Global RMSD value of the CORE-C system in core region (a) and C-terminal region (b) with respect to the average structure over the time range of 2-ns.

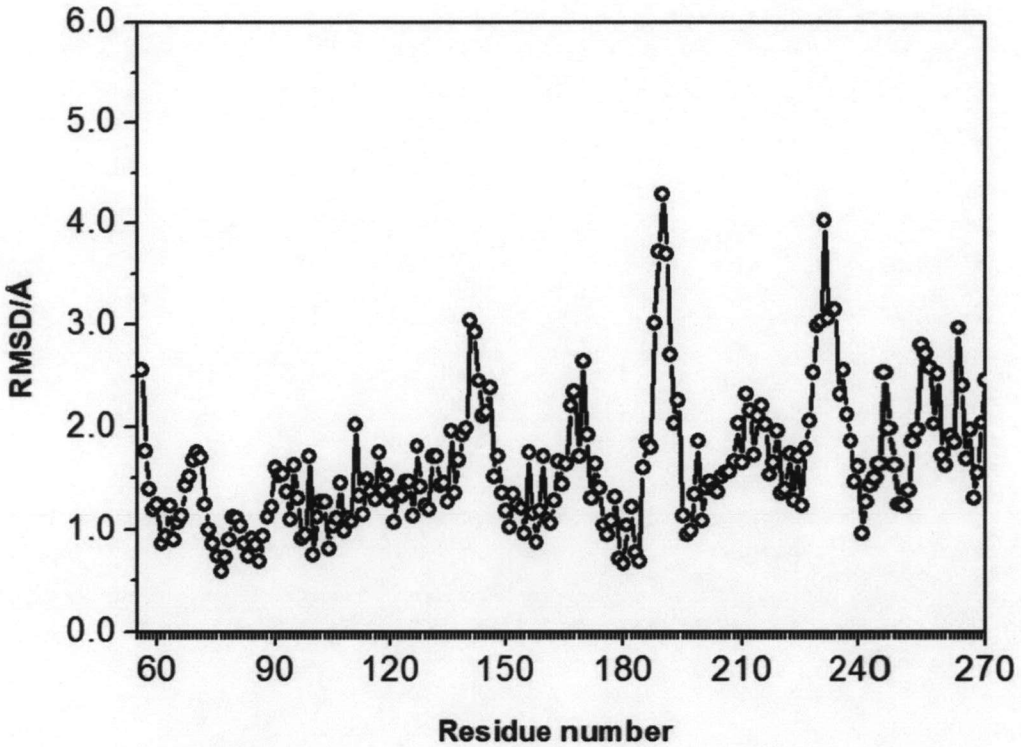


Figure 4.31 Global RMSD per residue of the CORE-C system with respect to the average structure over the time range of 2-ns.

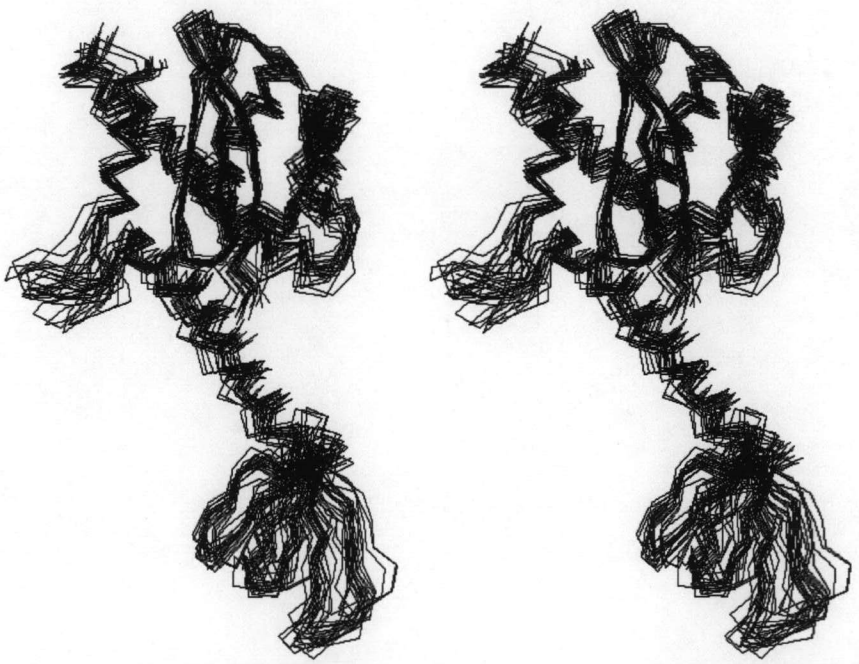


Figure 4.32 Superimposition of the structure taken every 100 ps from 2 ns MD trajectory for the CORE-C.

The RMSD per residue with respect to the average structure of the CORE-C system is illustrated in Figure 4.31. Most residues in the core domain have RMSD values in a range of 0.5 – 2.0 Å as in the CORE system. However, their RMSD values are in average lower than those of the CORE. Moreover, this core domain has only 2 regions with RMSD of over 3 Å. The fist region with the highest RMSD value is at around residues 188-192, the experimental flexible loop. This is similar to the CORE system. The second one is at around residue 150, which is missing in the experimental structure. Unlike the CORE system, the RMSD values of residues in core domain connected to C-terminal domain are lower than 2.0 Å. The reason is certainly due to an effect of the C-terminal domain that limits its movement. Considering the C-terminal domain itself, higher RMSD in a wide region, residues around 225 to 240, was observed. These can be clearly seen in Figure 4.32, similar to those two systems described above, the snapshot structures show high flexibility of the missing regions (residues 140 – 148) and region 188 – 192. The elbow linking region (residues 210 – 219) was reported to be flexible and may play role in DNA binding during integration process. Superimposed structures show the high fluctuation of the C-terminal domain (Figure 4.32). Such fluctuation is as a result of the flexible of the elbow linkage. This is confirmed by the stable RMSD of the C-terminal domain region (Figure 4.30).

The distances from Mg^{2+} to OD1 and OD2 atoms of Asp64 and of Asp116 along the 2-ns simulation are displayed in Figure 4.33. Comparing to the CORE system, the *d1*, *d2* and *d4* distances have smaller fluctuations while the *d3* distance varied in a broader range. Moreover, all distances are shorter especially the *d4* distance with reduction of around 0.5 Å.

The torsion angles involving residues 64, 116, and 152 were measured and are displayed in Figure 4.34. The most probable value of χ_{64} is around 120°, which is significantly different from that of the CORE system, 0°. This may be due to the effect of C-terminal domain. The dynamics of χ_{116} is similar to that of the CORE. The χ_{152} has the most probable value of around 90° while it is -90° in the CORE. Moreover, the χ_{152} has less flexibility than the CORE one, i.e. it flipped only 3 times.

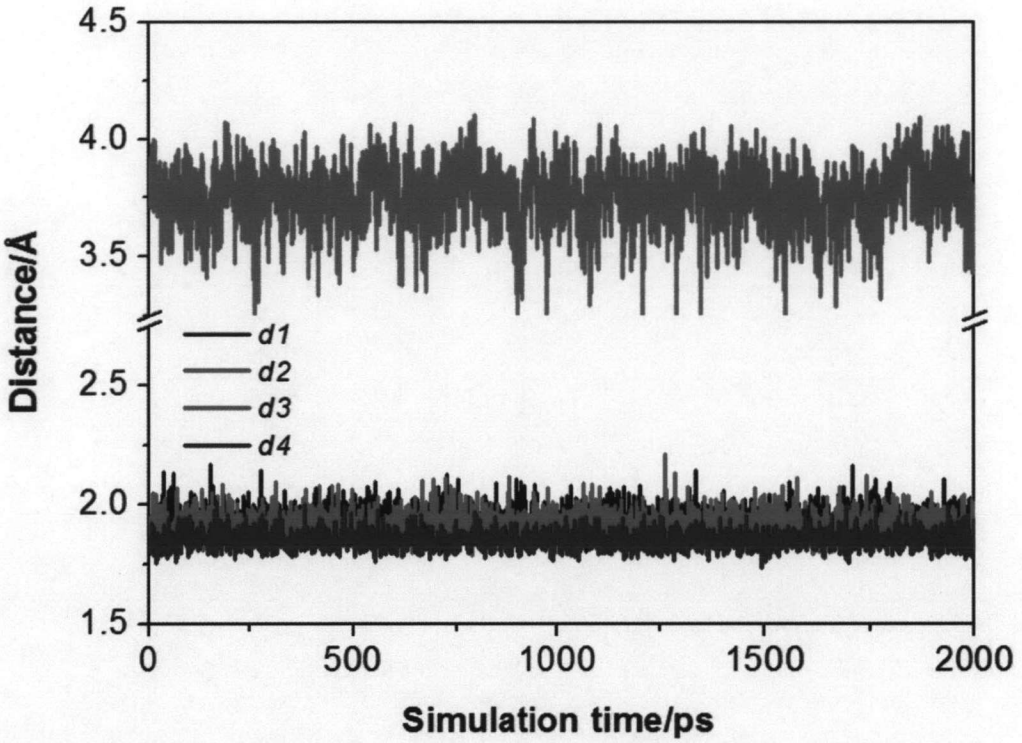


Figure 4.33 Distances between the Mg^{2+} and the carbonyl oxygen atoms of the two aspartate residues Asp64 (*d1* and *d2* as for OD1 and OD2, respectively) and Asp116 (*d3* and *d4* for OD1 and OD2, respectively) of the CORE-C system.

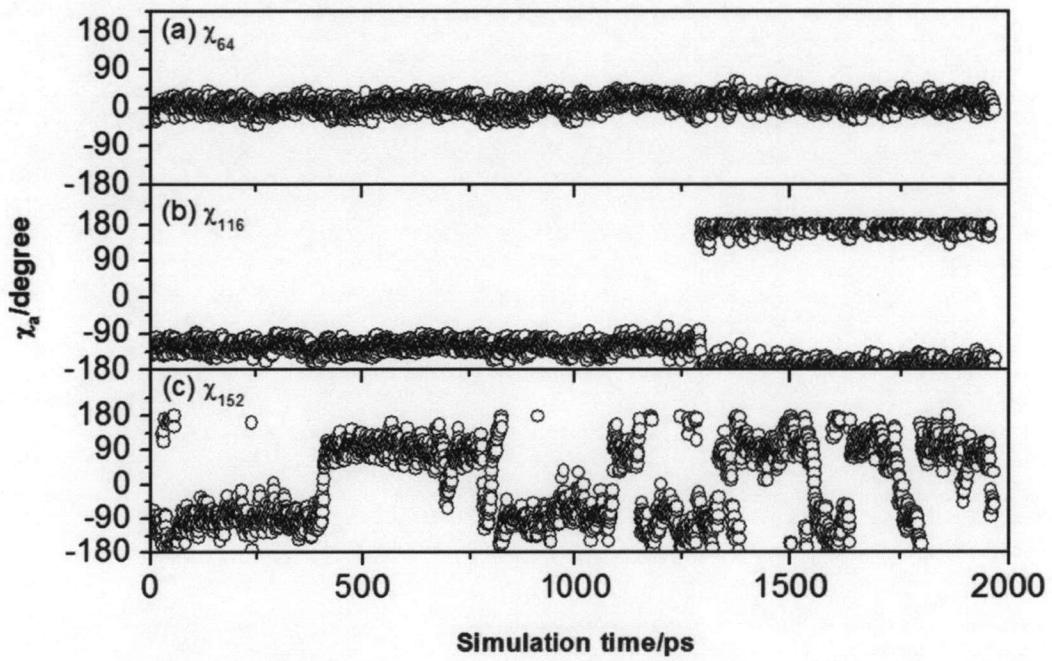


Figure 4.34 Torsional angles (χ_a) plot of three catalytic residues, Asp64 (a), Asp116 (b) and Glu152 (c) of the CORE-C.

4.2.3 The three-domain structure: Full-length system

4.2.3.1 FULL

Similar to other systems, a plot of total, potential and kinetic energies and temperature over 2 ns simulation time (Figure 4.35) assures the proper simulation. The mean values for the total energy, kinetic energy and potential energy are -10.19 x 10^4, 2.54 x 10^4 and -12.74 x 10^4 kcal/mol, respectively. The mean temperature of the system is 300.02 K.

If each individual domain was considered, the RMSD values of N-terminal and core domains are quite small while that of C-terminal domain is higher (Figure 4.36). However, if we included all domains, the RMSD was moderately increased. Compared to the CORE and CORE-C systems, the RMSD for both core and C-terminal domains are quite similar.

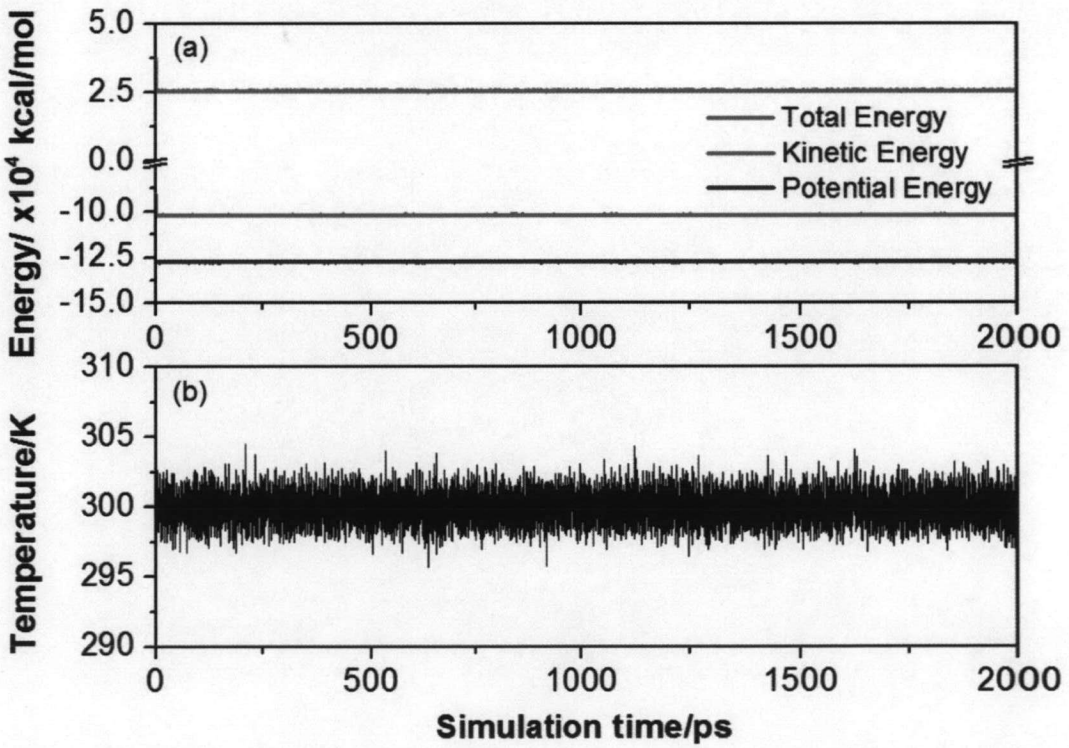


Figure 4.35 The total (red line), kinetic (green line) and potential (blue line) energies (a), and the temperature (b) over the 2-ns MD simulation for the FULL system.

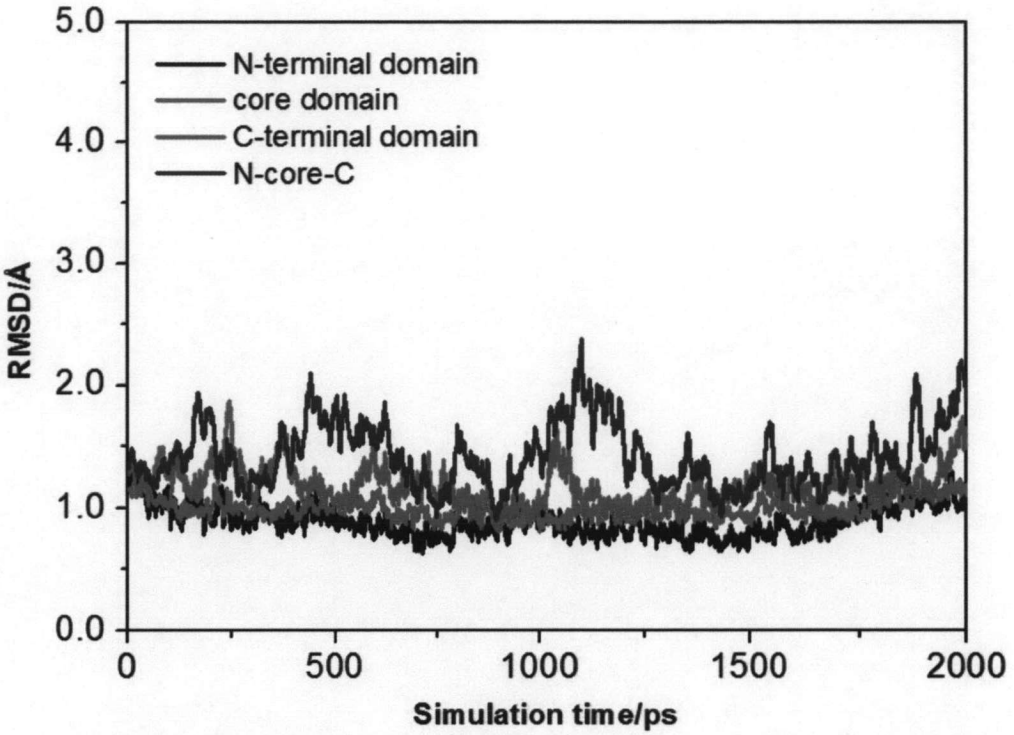


Figure 4.36 RMSD value of the FULL system in the N-terminal domain (black line), core domain (red line), C-terminal domain (green line) and the whole structure N-core-C (blue line) with respect to the average structure over the time range of 2-ns.

The RMSD of most residues in the FULL system is higher than that of the CORE and the CORE-C systems (Figure 4.37). In general, the RMSD pattern of all three systems is similar. However, there is one significant different point at around the residues 180. The broad and high RMSD peak at this position in the CORE-C is disappeared in the FULL system. The high RMSD in the C-terminal region comes from the flexible elbow linkage similar to the CORE-C system (Figure 4.37; see also Figure 4.31 for CORE-C). The superimposed structure over the trajectory shows that the C-terminal domain has high fluctuation (Figure 4.38).

Both χ_{64} and χ_{116} are significantly different from the CORE and CORE-C systems. As the FULL system does not contain Mg^{2+}, which normally binds to both Asp64 and Asp 116, so both residues have more flexibility (Figure 4.39 a and b). The χ_{152} was fluctuated from 180° to -180° indicating its high flexibility, similar to other systems (Figure 4.39 c).

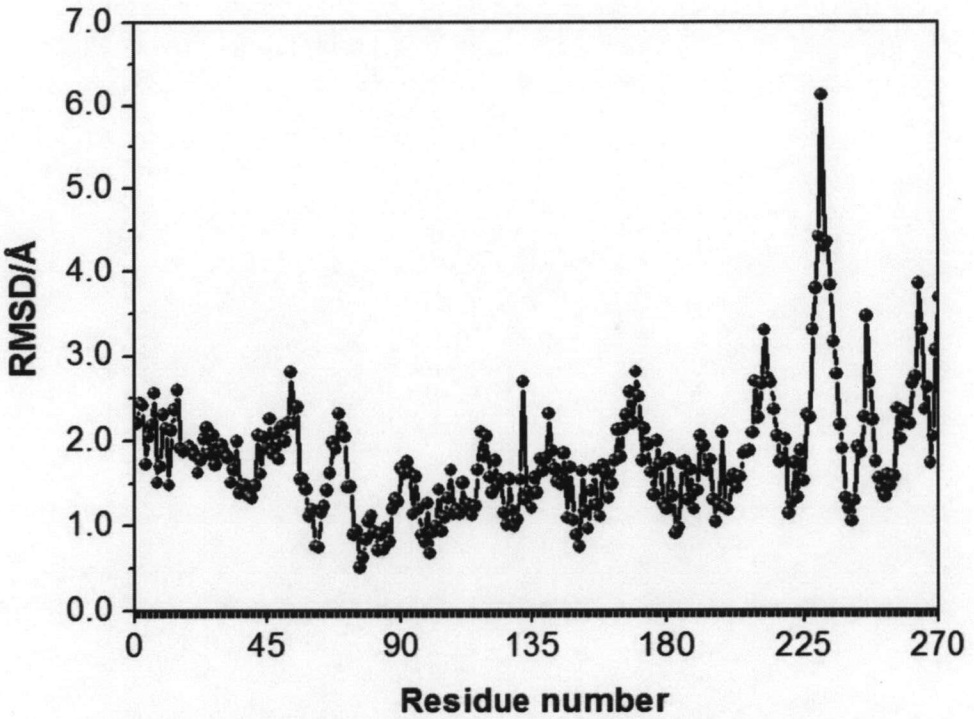


Figure 4.37 RMSD per residue of the FULL system with respect to the average structure over the time range of 2-ns.

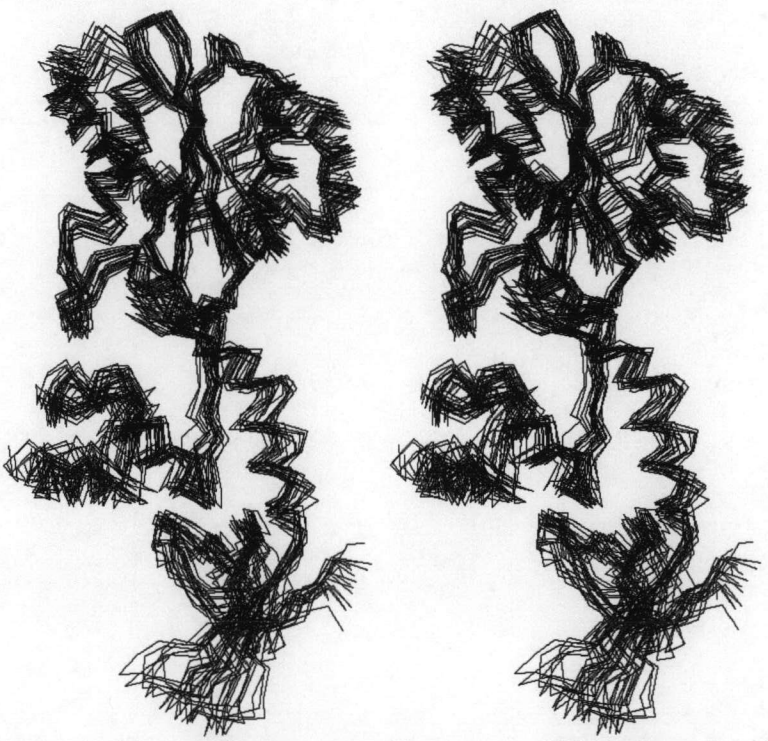


Figure 4.38 Superimposition of the structure taken every 100 ps from 2 ns MD trajectory of the FULL.

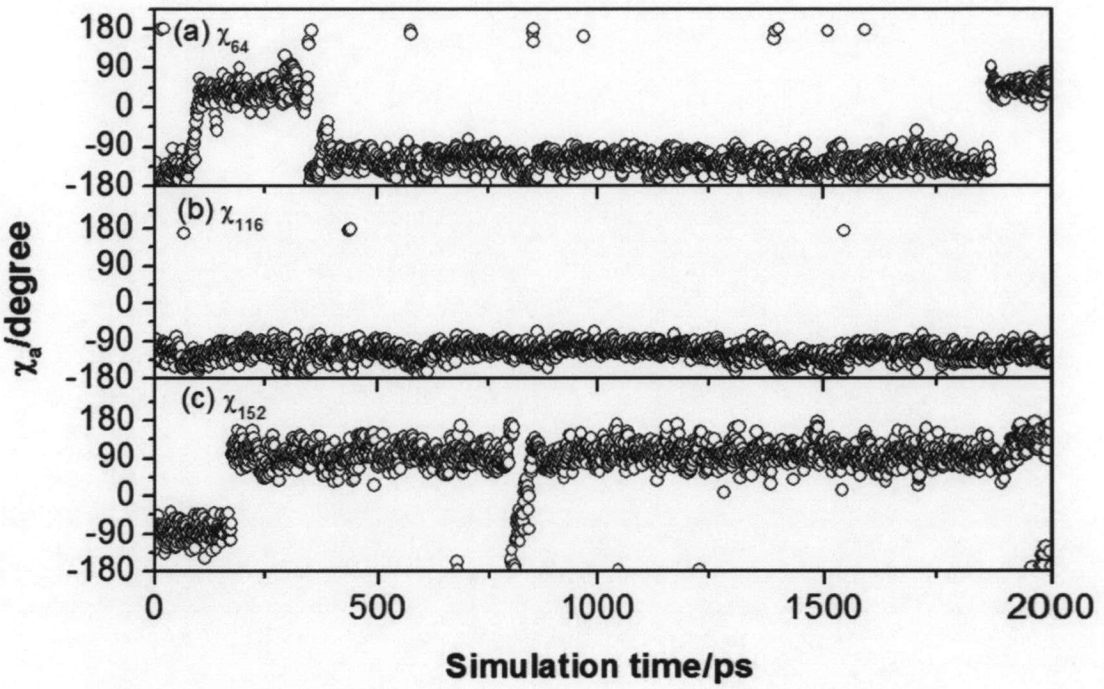


Figure 4.39 Torsional angles (χ_a) plot of three catalytic residues, Asp64 (a), Asp116 (b) and Glu152 (c) of the FULL.

4.2.3.2 FULL+ION

The quality of our simulation was validated by consistency of total, potential and kinetic energies and temperature over 2 ns simulation time (Figure 4.40). The mean values for the total energy, kinetic energy and potential energy are -10.41 x 10^4, 2.56 x 10^4 and -12.97 x 10^4 kcal/mol, respectively. The mean temperature of the system is 300.02 K. Comparing the RMSD plot over time of the FULL+ION (Figure 4.41) with that of the FULL (Figure 4.36), the N-terminal (black line) and the core (red line) domains showed similar RMSD values while the C-terminal (green line) and the whole structure (blue line) of the FULL+ION have higher RMSD than that of the FULL. Similar to the previous two systems containing the C-terminal domain, the C-terminal has large fluctuation due to the flexible elbow linkage. Interestingly, unlike the FULL which no metal ion is present in the active site, structures of the FULL+ION changes throughout the simulation (Figure 4.42, Figure 4.43). The RMSD of most residues in the FULL+ION system (Figure 4.42) is higher than that of the FULL, especially at the N-terminal and the C-terminal domains.

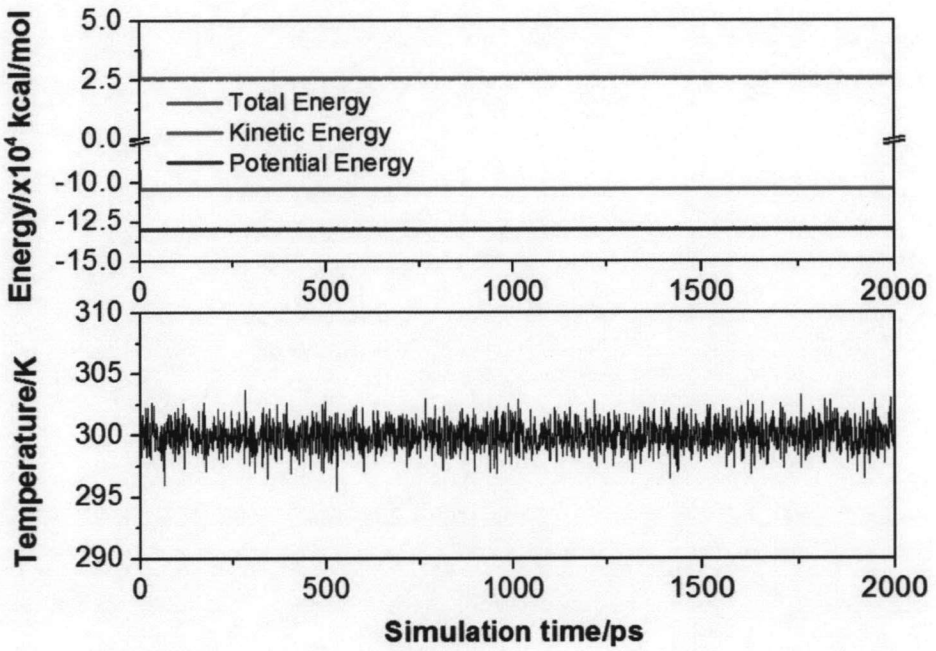


Figure 4.40 The total (red), kinetic (green) and potential (blue) energies (a), and the temperature (b) over the 2-ns MD simulation for the FULL+ION system.

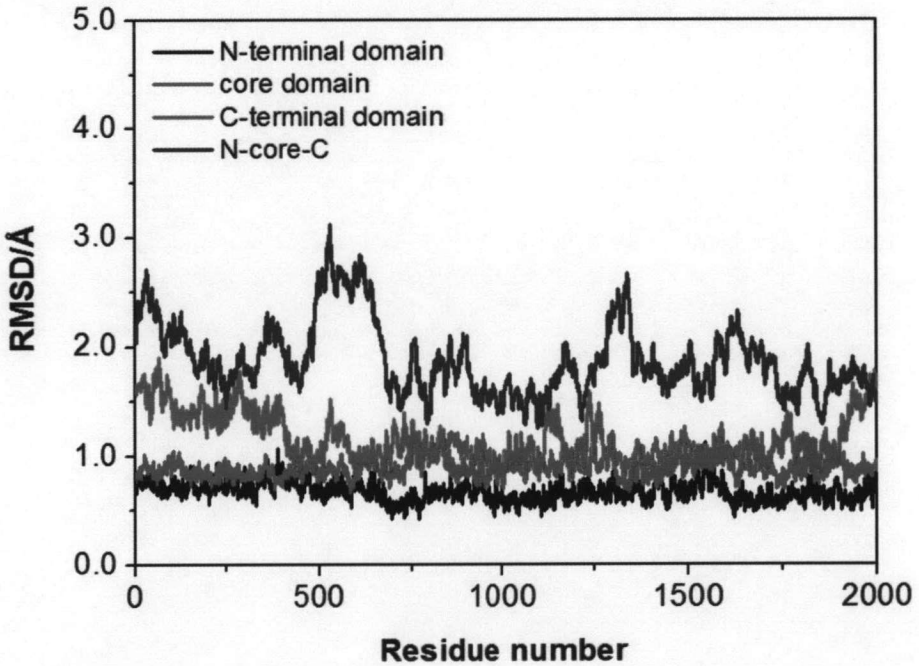


Figure 4.41 Global RMSD value of the FULL+ION system in the N-terminal domain (black line), core domain (red line), C-terminal domain (green line) and the whole structure N-core-C (blue line) with respect to the average structure over the time range of 2-ns.

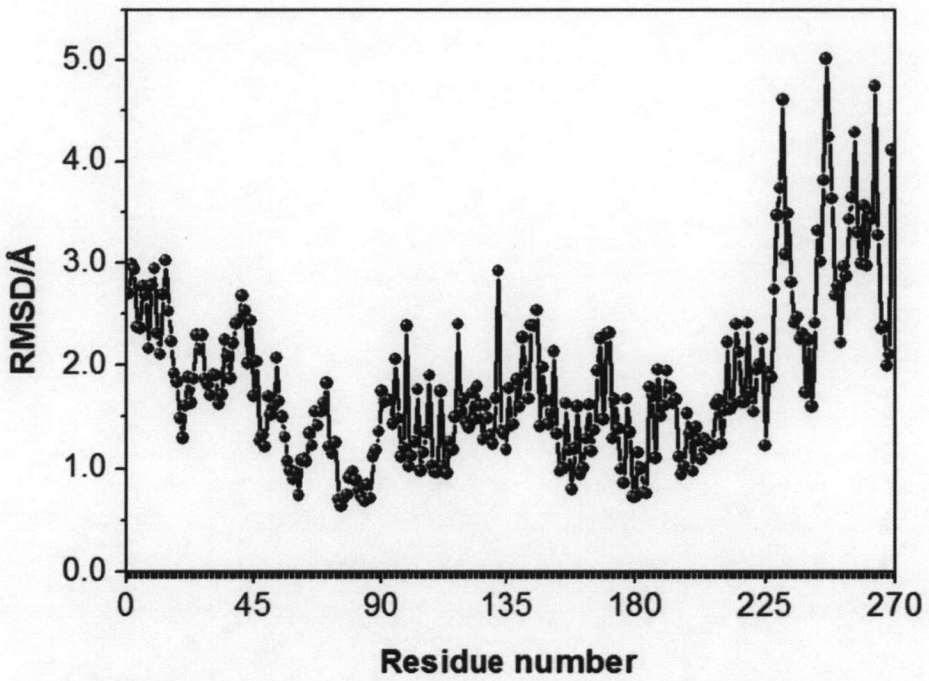


Figure 4.42 RMSD per residue of the FULL+ION system with respect to the average structure over the time range of 2-ns.

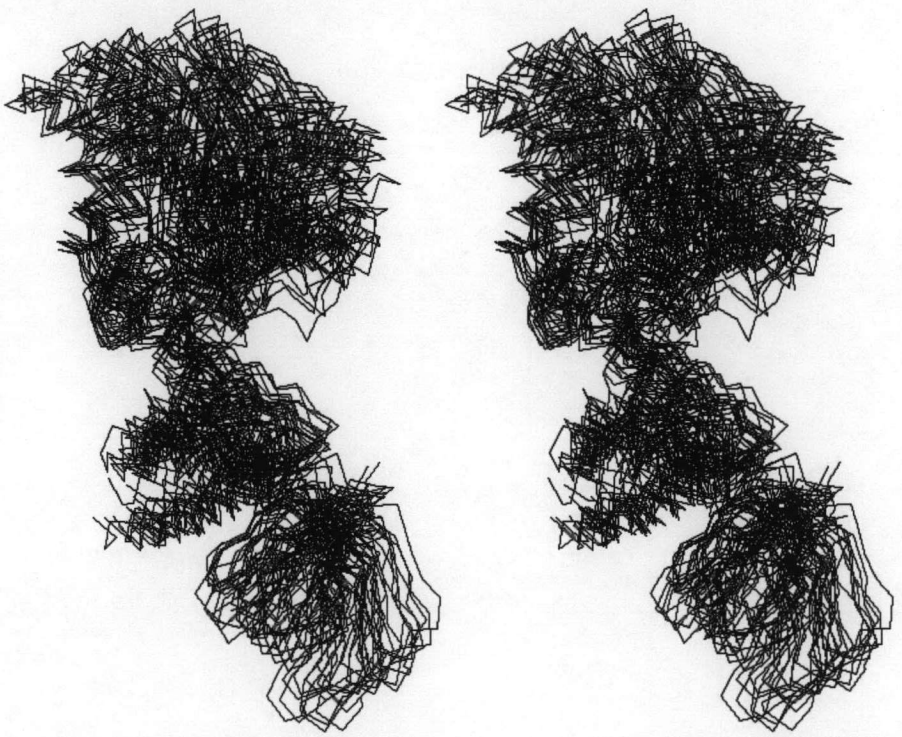


Figure 4.43 Superimposition of the structure taken every 100 ps from 2 ns MD trajectory of the FULL+ION.

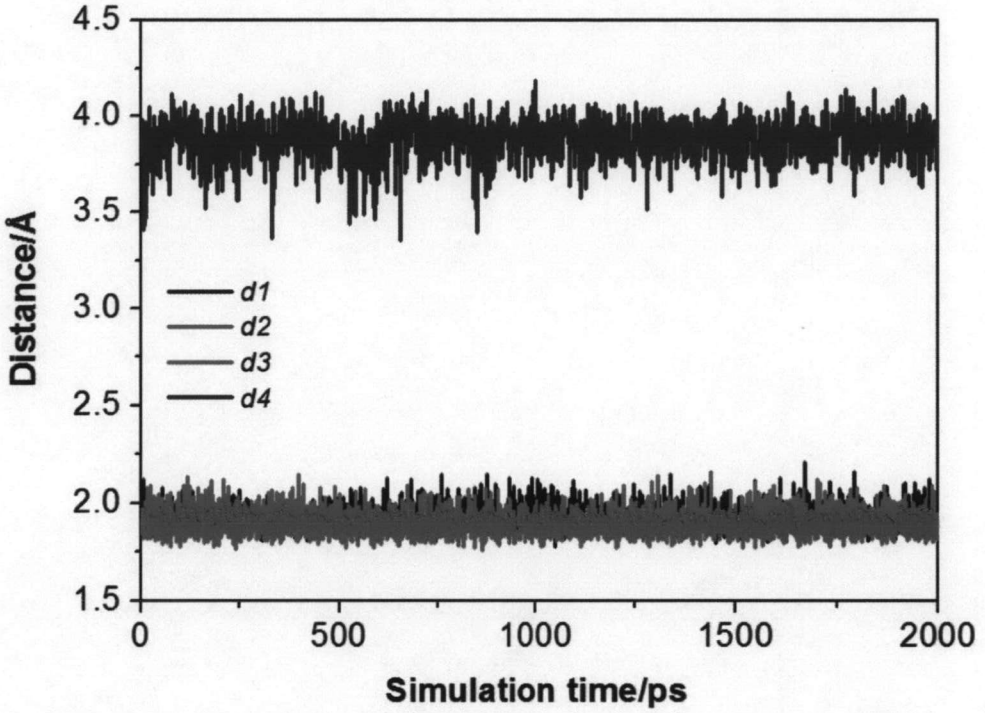


Figure 4.44 Distances between the Mg^{2+} and the carbonyl oxygen atoms of the two aspartate residues Asp64 (*d1*, *d2*) and Asp116 (*d3*, *d4*) of the FULL+ION system.

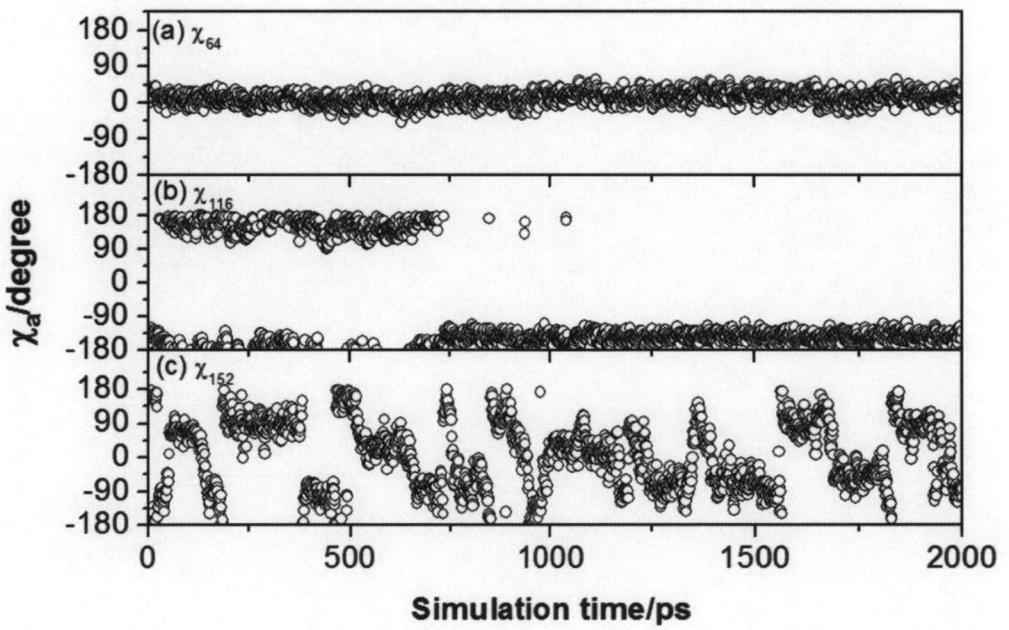


Figure 4.45 Torsional angles (χ_a) plot of three catalytic residues, Asp64 (a), Asp116 (b) and Glu152 (c).

Interestingly, compared to the CORE and the CORE-C systems, the Mg^{2+} switched to interact with the OD2 of Asp116 (*d4*). Moreover, this *d4* distance has lower fluctuation while the other distances are similar to that of the two systems (see Figure 4.44 for more details).

The χ_{64} showed similar pattern to that of the CORE and CORE-C systems, in which all systems have Mg^{2+} binding to the Asp64. The χ_{116} exhibited totally different behavior from other systems. This torsion is fluctuated in an interval around 180° and -180°, which is exactly the same angle, while the most probable value of other systems is at around 0°, the opposite side. For the χ_{152}, it has the most flexibility than the other systems as indicated by many flip throughout the simulation (Figure 4.45).

4.2.4 Detailed analysis

Structural analysis was considerably applied to the set of structures including CORE, the two-domain (CORE-N and CORE-C) and the FULL+ION to elucidate how each end terminal domain affects the central core domain. The same analysis was made to the two-domain (CORE-N and CORE-C) and the full-length (FULL+ION) in order to investigate the effect of addition of the third part to the structure.

The two linkage regions (residues 47 – 55 and 211 – 219) were considered as parts in the N-terminal and the C-terminal, respectively, while the comparisons were made to the domain fragments in all system as their high fluctuation. These two regions might cause the high RMSD values due to their flexibility. Table 4.1 shows the RMSD values from superimposition each domain of the FULL+ION and the other systems. The average structure obtained from 2 ns MD trajectories from each system was selected. An individual domain, core, N- and C-terminal domains of the CORE-N, CORE-C were superimposed to such domain of the FULL+ION. Table 4.1 lists the RMSD value for the different between the individual domains in all systems classified as one-domain, two-domain and three-domain of the FULL+ION.

Table 4.1 indicates that each individual one-domain in the FULL+ION differs from each component in CORE, CORE-N and CORE-C. Large difference was

found in the N-terminal region of the CORE-N and the FULL+ION (RMSD = 5.07 Å) while the RMSD for the FULL is lower (RMSD = 1.67 Å). Considering the core domain region, the CORE differs from the CORE-N and the FULL+ION in the same range while the difference is less in the CORE-C and FULL. For the C-terminal region, CORE-C differs from the FULL+ION (RMSD = 2.03) while the structure of this region in the FULL and FULL+ION is almost similar (RMSD = 1.61).

Table 4.1 Mean RMSD values for each comparison made between the domain in CORE, CORE-N, CORE-C and FULL with the individual domain in the FULL+ION structure.

	One-domain in FULL+ION		
	N-terminal (Residues 1 – 55)	Core (Residues 56 – 209)	C-terminal (Residues 210 – 270)
CORE	-	2.38	-
CORE-N	5.07	2.35	-
CORE-C	-	1.96	2.03
FULL	1.67	0.88	1.61
	Two-domain in FULL+ION		
	N-terminal + Core (Residues 1 – 209)		C-terminal
CORE-N	5.71		-
FULL	3.19		-
	N-terminal		Core + C-terminal (Residues 56 – 270)
CORE-C	-		2.32
FULL	-		4.64
	Three-domain in FULL+ION		
	N-terminal + core + C-terminal (Residues 1 – 270)		
FULL	4.76		

Table 4.2 Distances (*d1* – *d4*) between the Mg^{2+} and the carboxyl oxygen of Asp64 and Asp116.

Structure / chain		*d1*	*d2*	*d3*	*d4*	Coordination number of Mg^{2+}		Domain	Method
						IN	Water*		
1BIU / A		3.57	3.87	5.07	4.45	-	*n*		
1BIU / B **(35)**		2.31	3.66	3.63	2.10	2	*n*	core	X-ray
1BIU / C		2.82	3.53	3.90	2.30	2	*n*		
1BL3 / A		2.42	3.78	4.01	2.39	2	4		
1BL3 / B **(77)**		2.34	3.30	4.34	2.91	2	*n*	core	X-ray
1BL3 / C		2.20	3.53	3.93	2.32	2	4		
1QS4 / A **		2.32	3.54	2.25	3.77	2	4		
1QS4 / B **(78)**		2.37	3.49	2.34	3.92	2	4	core	X-ray
1QS4 / C		2.59	3.69	2.43	3.86	2	*n*		
Lin *et al.*/ A **(24)**		-	-	-	-	2	4	core	MD
Luca *** / A **(102)**	ion1	2.04	2.38	2.17	2.09	4	*n*	Full-length	Minimization
	ion 2	4.21	2.11	2.13	2.12	3			
Luca / B	ion 1	3.47	2.10	2.11	2.06	3	*n*		
	ion 2	2.10	2.99	2.16	2.14	4			
FULL+ION		1.91	1.92	1.86	3.89	3	3	Full-length	MD
CORE		1.93	1.93	3.98	1.87	3	3	core	MD
CORE-N		1.91	1.93	1.87	3.72	3	3	core + N	MD
CORE-C		1.93	1.92	3.76	1.85	3	3	core + C	MD

* *n* : not available, ** complexed with inhibitor, *** complexed with DNA

4.2.4.1 Effect of the divalent metal ion on the catalytic triad

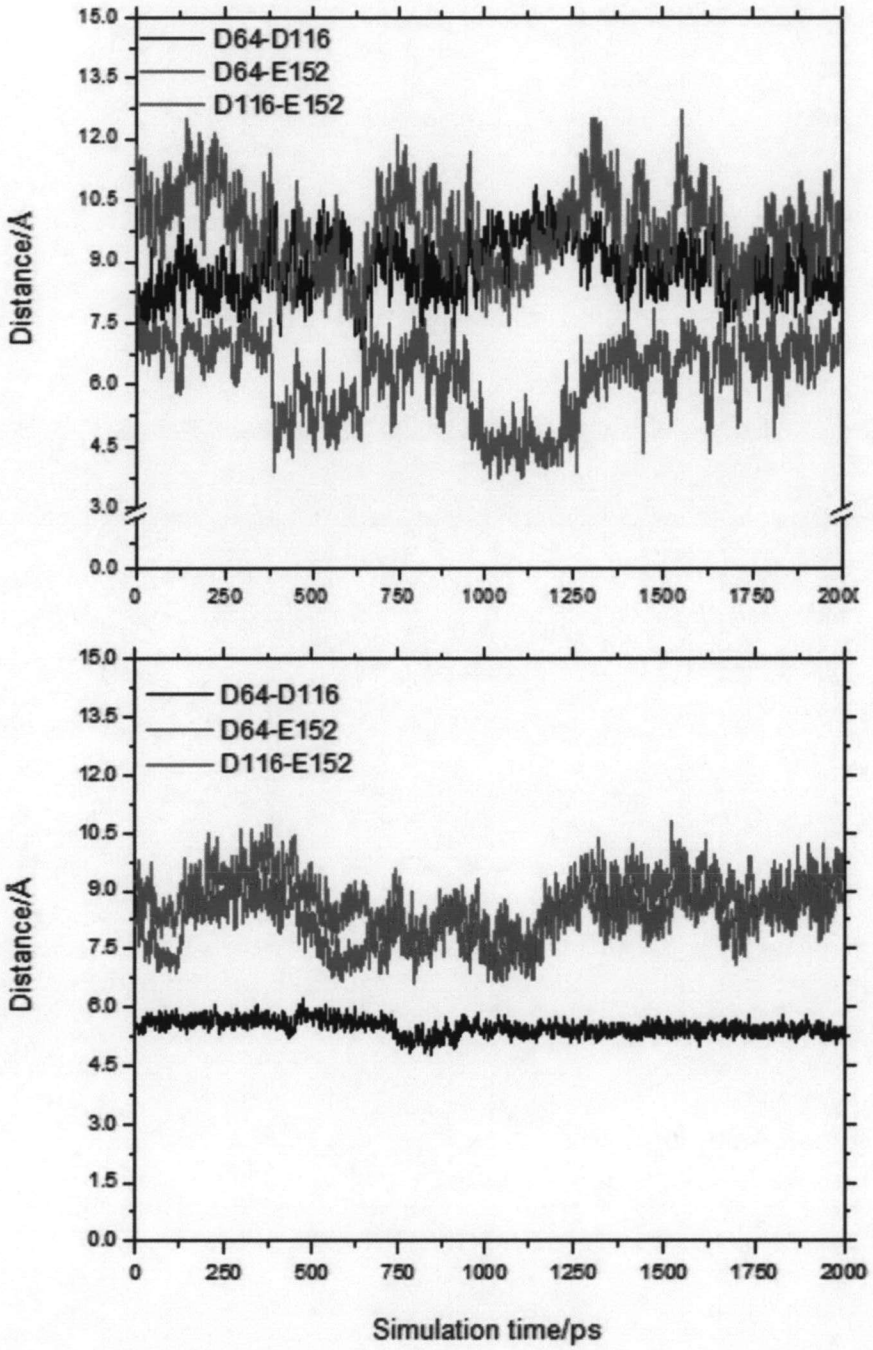


Figure 4.46 Distance between each catalytic residues, Asp64-Asp116 (black line), Asp64-Glu152 (red line) and Asp116-Glu152 (green line) over the 2 ns simulation time for the FULL (top) and the FULL+ION (bottom).

In order to investigate the effect of the divalent metal ion (Mg^{2+}) in the active site of the core domain region, the full-length structure without and with Mg^{2+} in the core domain were studied (see 4.2.3 for more details). The full-length structure in the absence of the Mg^{2+} in the active site, FULL, were compared to the FULL+ION. It was shown in Table 4.1 that the differences between the core domains in both system are not as much as in the two-domain systems (RMSD = 0.88 Å). The major observation is at the torsion angle of Asp64 and Asp116 (see Figures 4.39 and 4.45). Binding of Mg^{2+} in the active site causes Asp64 to have only one conformational preference. The size of the binding pocket is also altered, see 4.2.4.3 for more details.

The distances between each catalytic residues (Asp64, Asp116 and Glu152) were measured and plotted in Figure 4.46. The distance between the catalytic residues (Figure 4.46) reveals that in the absence of Mg^{2+} (Figure 4.46, top) the fluctuations are larger than that of the system with Mg^{2+} (FULL+ION, Figure 4.46, bottom). In the present of the Mg^{2+}, Glu52 fluctuates in the wide range as it does not bind to the Mg^{2+}. The distance involving this glutamate residue shows higher fluctuation.

4.2.4.2 Effect of the two end domains

4.2.4.2.1 Effect on the structure

The active site regions, Asp64, Asp116 and Glu152, in the catalytic core domain from the five systems in our study were superimposed and compared. The distances between the Mg^{2+} and the carboxyl oxygen atoms of Asp64 and Asp116 of the four simulated systems were monitored in order to elucidate the binding pocket and summarized in Table 4.2. The previous experimental data are also given in the table. Previous data suggested that Mg^{2+} bound to the active site in an octahedral fashion. It is obviously seen that, for the crystal structure of the core domain, Mg^{2+} coordinates with two oxygen atoms, one from Asp64 and another from Asp116, and four oxygen atoms from water molecules. Previous MD simulation of the core domain by Lins *et al.* **(24)** also showed the same result. However, our simulation results displayed that the octahedral fashion of metal binding is formed by the three coordination's from carboxyl oxygen atoms of the two aspartate residues. This is different from the other in which

another three oxygen atoms come from water molecules (see Table 4.2). All systems, CORE, CORE-N, CORE-C and FULL+ION show the same binding indicating that the active site binding was not altered.

Considering the C-terminal DNA binding domain, some selected key residues, *i.e.*, Ser230, Glu246 and Arg262 in which reported to bind DNA **(53, 54)**, were taken into account to investigate how the N-terminal domain affect the structure of the C-terminal domain. Distances between these key residues were measured and distributions of such values in the CORE-C and the FULL+ION system were plotted in Figure 4.47. The results show that the conformation of the C-terminal domains in the FULL+ION is wider than that of the CORE-C. This implies that addition of the N-terminal domain may affect such alteration.

For the N-terminal domain, it should be note that the major different is at the linkage region (residues 47 – 55). The coordination of Zn^{2+} in the HHCC binding pocket was kept constant throughout the simulation.

4.2.4.2.2 Effect on the mobility

Mobility of the core domain from the various systems, which are CORE, CORE-C, CORE-N and FULL+ION, was plotted together in Figure 4.48. The RMSD of the core regions clearly show that the core only domain (CORE, black line) has highest fluctuation for the entire simulation. The core domain of the two-domain fragment (CORE-N, green and CORE-C, red) has higher RMSD than that of the core of the FULL+ION one. Note that, there may have some large conformational changes in the CORE-N occuring during 700 – 900 ps. These results indicate that the presence of the one end domain, either N- or C-terminal domain, led to the higher mobility of the core region. It is obviously seen that the presence of both two end domains cause the core region to less flexible.

The same manners were plotted for the two end domains, N- and C-terminal domains, as illustrated in Figure 4.49 and also to the two-domain, CORE-N and CORE-C (see Figure 4.50)

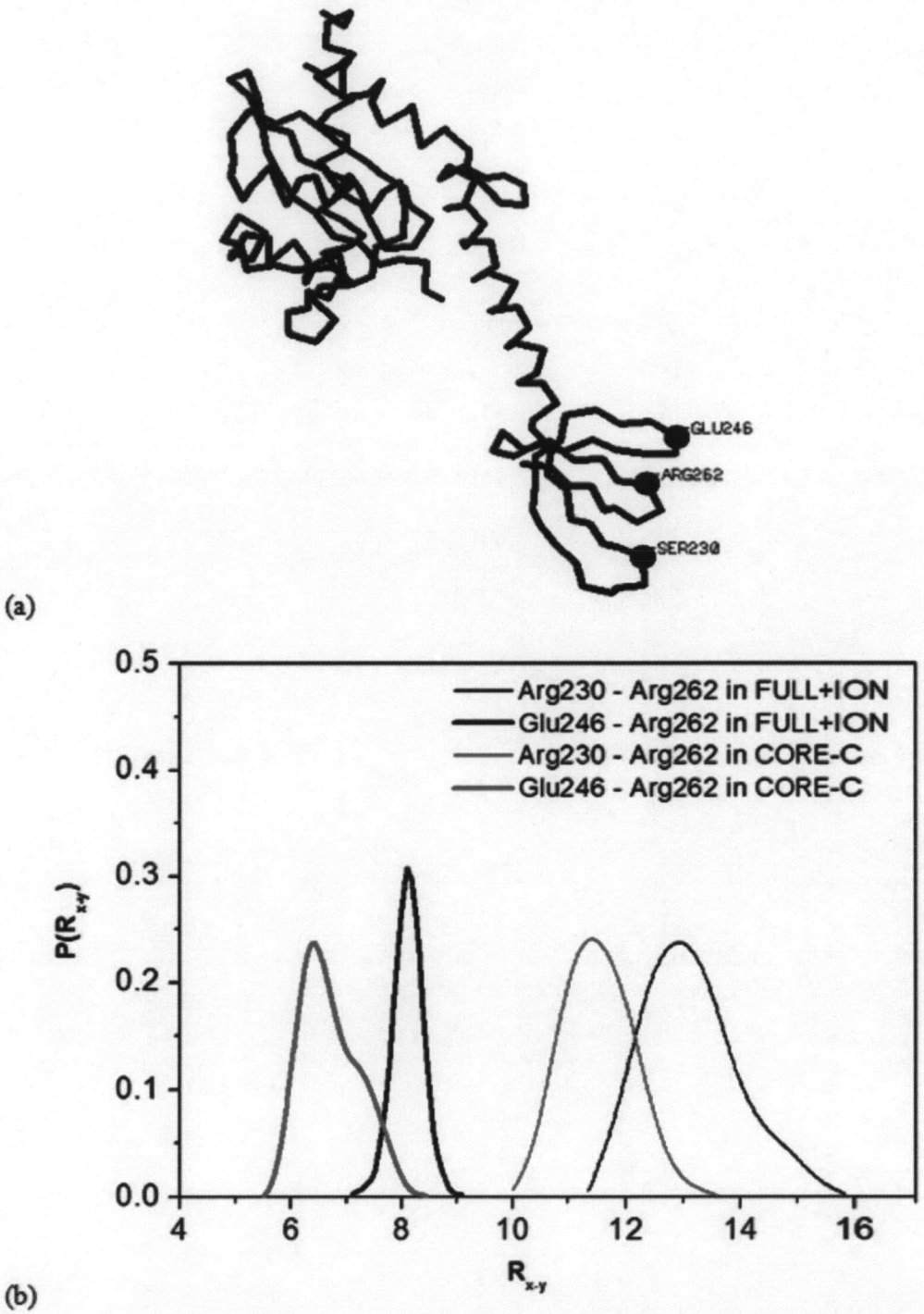


Figure 4.47 Structure of the CORE-C where the labelled residues in the C-terminal region (Ser230, Glu246 and Arg262) was reported to bind with DNA (a) and the distribution of the distance between these key residues in the two-domain CORE-C and the FULL+ION (b).

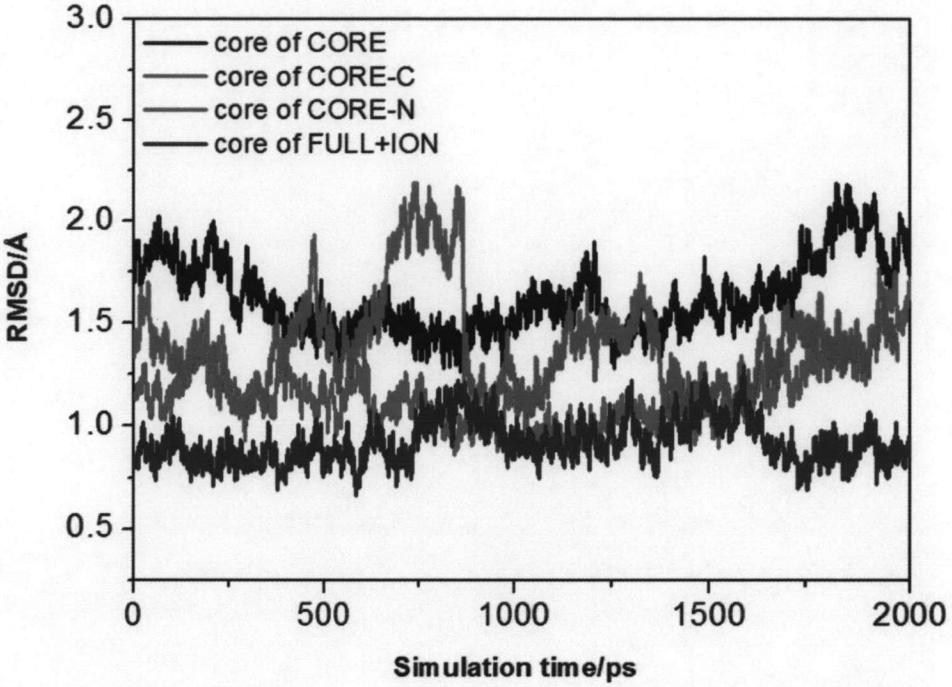


Figure 4.48 RMSD over the trajectories of the core domain region (residues 56 – 209) of CORE (black line), CORE-C (red line), CORE-N (green line) and FULL+ION (blue line).

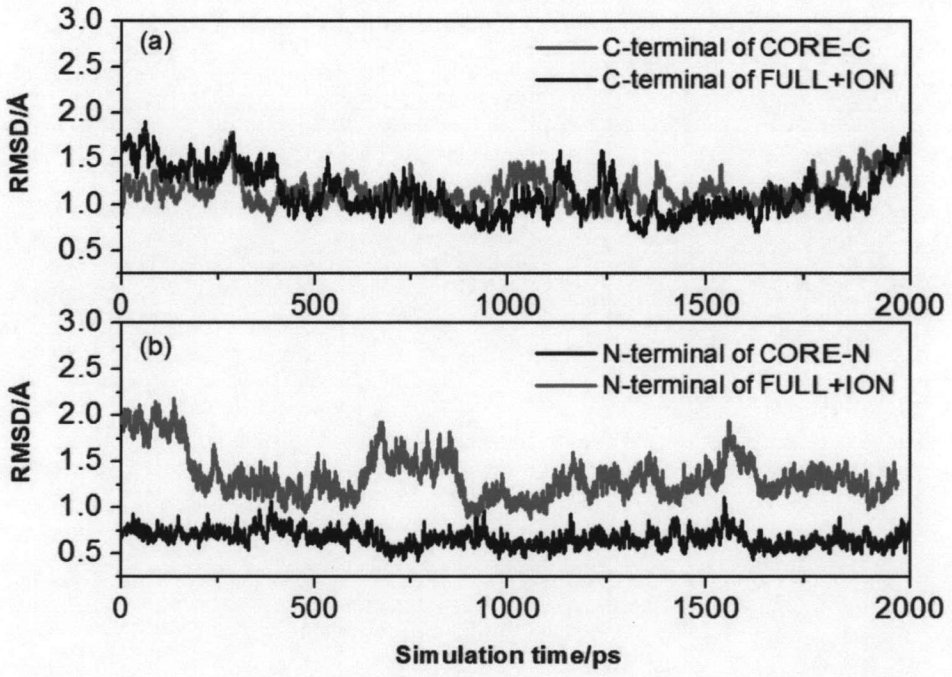


Figure 4.49 RMSD over the trajectories (a) of the C-terminal domain (residues 210 – 270) of CORE-C (green line) and FULL+ION (black line) and (b) of the N-terminal domain (residues 1 – 55) of CORE-N (red line) and FULL+ION (black line).

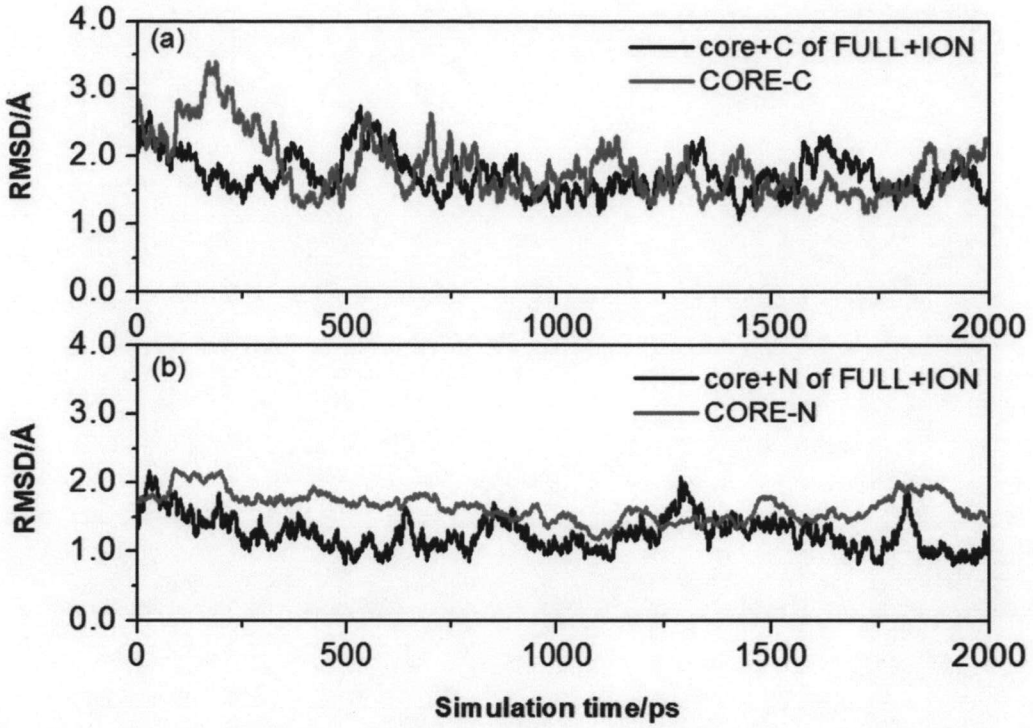


Figure 4.50 RMSD over the trajectories (a) of the two-end terminal domain (CORE-C, residues 56 – 270) of CORE-C (green line) and FULL+ION (black line) and (b) of the CORE-N (residues 1 – 210) of CORE-N (red line) and FULL+ION (black line).

4.2.4.3 Similar and dissimilar among the full-length models

Comparisons were made to the full-length structure proposed by different groups. Our results indicated that there is no significant difference among the four models (our FULL and FULL+ION, Podtelezhnikov's **(64)** and Luca's **(63)** models) in terms of geometry and orientation within each domain. This is not surprise since each single domain structure was taken from x-ray data. However, as indicated above the differences are the relative orientations of the three separated domains, core connected to N-terminal and core connected to C-terminal.

To elucidate this aspect more quantitatively, relative orientations between the core and the two end domains were measured in terms of the angle α, formed by the two vectors pointing from *a* to *b* and *a* to *c*, where *a*, *b* and *c* are the averaged locations from all atoms of the core, N-terminal and C-terminal domains,

respectively. The distribution of the angle α extracted from the MD simulations was displayed in Figure 4.51. The same angle was also calculated from Podtelezhnikov *et al.*'s and Luca *et al.*'s models and the values were also plotted in the same figure for comparison.

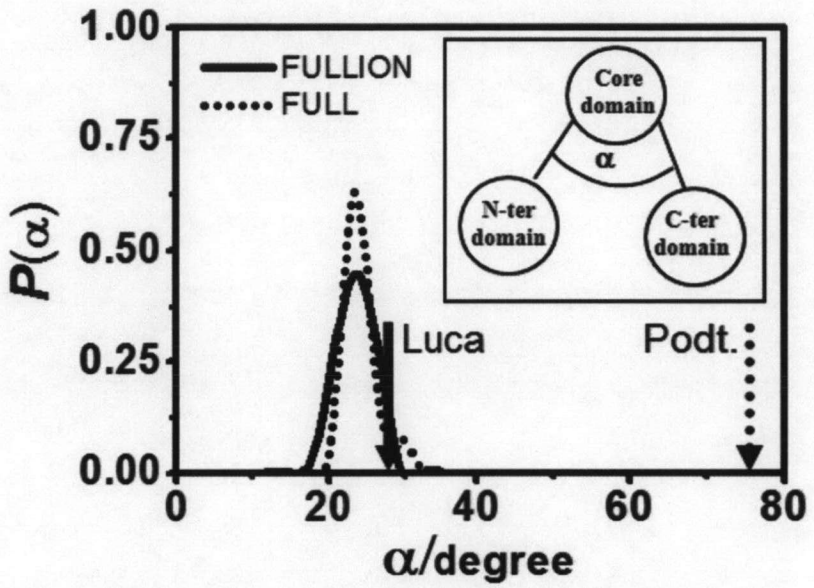


Figure 4.51 Probability distribution of the angle α representing relative orientation of the core, N-terminal and C-terminal domains of our full-length HIV-1 integrase, FULL+ION (thick line) and FULL (thin line). The arrows are the corresponding values calculated from the Podtelezhnikov's and Luca's models.

The distribution plot for the FULL+ION shows the α fluctuated in the range 18° - 27° with the average value of 23°, whereas, that of the free form, FULL, yields the fluctuation range of 19° – 31° and the average value at 24°. The corresponding values for the angle α are 76° and 28°, for the Podtelezhnikov *et al.*'s and Luca *et al.*'s models, respectively. The results indicate that our models are consistent with those from Luca *et al.* whereas those from Podtelezhnikov *et al.* show a large deviation. This implies that generating of the full-length structures from the two connected CORE-N and CORE-C domains may provide more consistent structures than that from the separated core, N-terminal and C-terminal domains.

To understand the structure of the catalytic region, especially how the Mg^{2+} cation affects the cavity for inhibitor binding of this region, distributions of the

selected distances and angles over the simulation time were evaluated and plotted in Figures 4.52 – 4.54. Corresponding values were also extracted from Podtelezhnikov *et al.* and Luca *et al.* structures and were also plotted for comparison. Note that the distances, $R_{x\text{-}y}$, were measured from a carboxyl carbon atom in the –COO– groups of residue x to the same type of atom of residue y where subscripts x and y can be Asp64, Asp116 or Glu152 (Figure 4.52). The coordination of the cation in the active site and the torsional angles were defined in an inset of Figure 4.53 and 4.54.

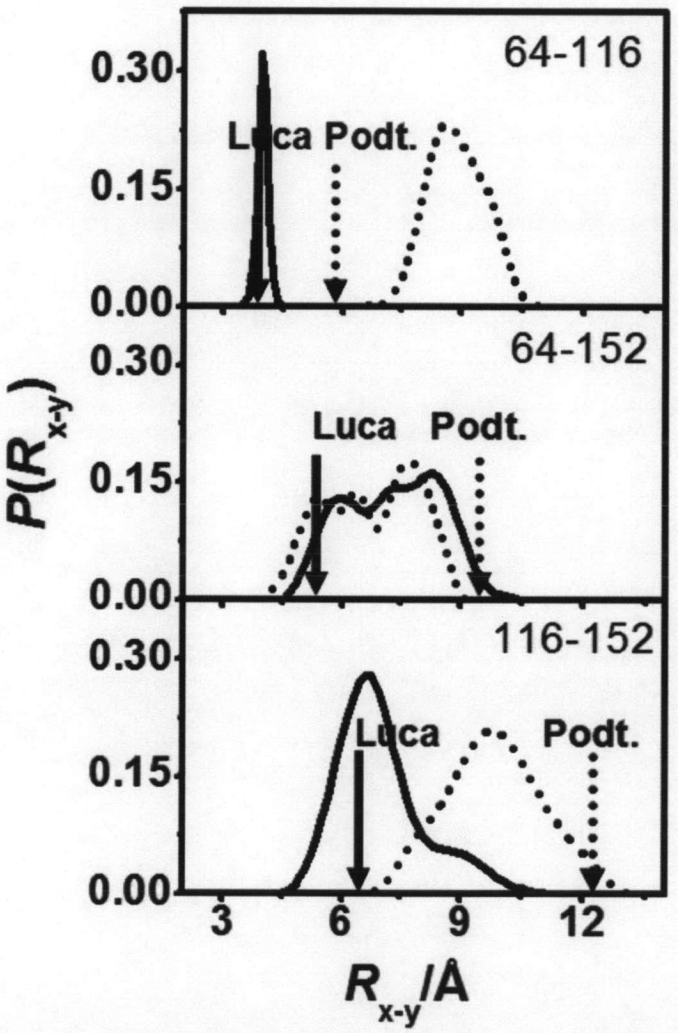


Figure 4.52 Distribution of the distances $P(R_{x\text{-}y})$ between each catalytic triad (see text for details); Asp64-Asp116 (a), Asp64-Glu152 (b) and Asp116-Glu152 (c) throughout the 2-ns simulations for the FULL+ION (solid line) and the FULL (dash line). Single value extracted from Podtelezhnikov's (dash arrow) and Luca's (solid arrow) models were also displayed.

For the $R_{64\text{-}116}$ (Figure 4.52 a), the FULL+ION system has a single sharp peak at 3.93 Å (solid line) due to the tight binding between the Mg^{2+} ion and two aspartate residues, *i.e.* both residues were moved consistently together resulting in the narrow distribution of its distance. In contrast, the FULL system showed a broad peak (dash line) indicating more flexibility. In the case of $R_{64\text{-}152}$ (Figure 4.52 b), both the FULL+ION (solid line) and the FULL (dash line) systems displayed broad distribution patterns with no prominent peak. The reason may be due to the rather flexible Glu152 residue indicated by its torsion angle as discussed in more detail below while the Asp64 is rather rigid since it is bound to the Mg^{2+} ion. Considering the $R_{116\text{-}152}$, the FULL+ION (Figure 4.52 c, solid line) showed a broad single peak at around 6.81 Å. This distance in the FULL system (Figure 4.52 c, dash line) is, as expected, shifted to a longer distance and showed a broader peak than that of the FULL+ION. In comparison to the available data, our three distances of the FULL+ION system are in agreement with those reported by Luca *et al.* On the other hand, the FULL system exhibits substantially different distances from those of the Podtelzhnikov *et al.* model, especially the $R_{64\text{-}116}$ distance.

In terms of the cation's coordination, distributions of the distances between electron donor atoms of the catalytic residues and the Mg^{2+} ion were plotted in Figure 4.53. Comparisons with the available data, the cation's coordination from the three crystal structures of the core domain only; 1BIU, 1BL3 and 1QS4 and from the simulation results by Lin *et al.* were also summarized in Table 4.2. Note that all three X-ray structures are in trimer forms, chains A, B and C. It can be seen from Table 4.2 that the X-ray data of the three chains, A, B and C, as well as the simulation results by Lin *et al.* **(24)** predict that the Mg^{2+} ion coordinates with 4 oxygen atoms from the water molecules and 2 carboxyl oxygen atoms of the Asp64 and Asp116. The experimental Mg^{2+}-O distances are in the range 2.10 – 2.91 Å while those of Lin *et al.* were not reported. For the Luca *et al.* dimer full-length structure where two Mg^{2+} ions per monomer were taken into account (Table 4.2), the coordination number was not reported. However, with an assumption that binding could be formed when the Mg^{2+} - O distance is shorter than 3 Å, we measured all the Mg^{2+} - O distances in their modeled structure and found that the coordination number can be either 3 or 4. The first Mg^{2+} of chain A coordinates with four oxygen atoms of Asp64 and Asp116 while the second Mg^{2+} of

chain A forms three bonds with oxygen atoms, one from Asp64 and the other two from Asp116. In contrast, for chain B, the first Mg^{2+} ion coordinates with one oxygen atom of Asp64 and two oxygen atoms of Asp116 and the second ion forms 4 bonds with four carboxyl oxygen atoms, two from Asp116 and the other two with Asp64 (See Table 4.2). From the distribution plot (Figure 4.53) and the optimal Mg^{2+} - O distances (maximum of the peak) (Table 4.2), our results also indicate that the coordination number for the Mg^{2+} ion is 6 but the specific binding structure is different from the previous studies. The octahedral sites of our data of the Mg^{2+} contain two oxygen atoms of Asp64, one oxygen atom of Asp116 and three oxygen atoms of the water molecules. However, our solvation properties are, quantitatively, in the range obtained from the MD results of the full-length structure by Luca *et al.* which observed 1-4 bonds between the Mg^{2+} ion and the catalytic residues and the corresponding Mg^{2+} - O distances in the range of 2.06 – 2.99 Å.

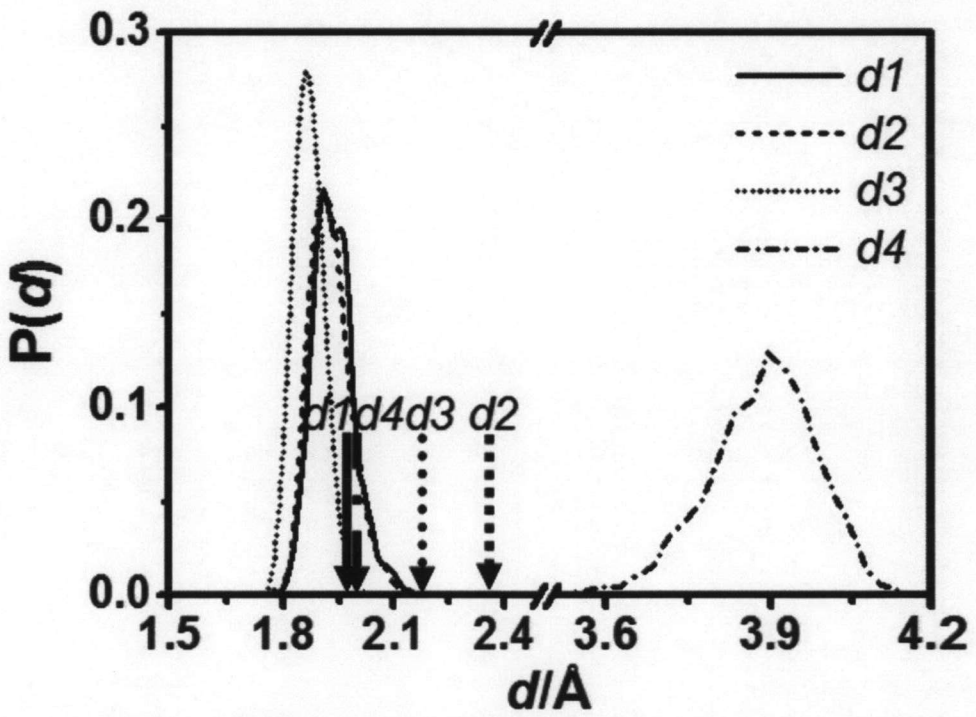


Figure 4.53 Distribution of the distance P(*d*) between Mg^{2+} ion and the carboxyl oxygen of the catalytic triad residues, *d1*, *d2*, *d3* and *d4*, as for Asp64:OD1, Asp64:OD2, Asp116:OD1, Asp116:OD2, respectively. The arrows illustrate the data extracted from Luca's structure.

In addition to the ion's coordination numbers and the related distances which represent the size of the catalytic site of the HIV-1 IN, binding affinity between the enzyme and the inhibitor was known to depend, obviously, on the orientation of the catalytic residues. Such information was evaluated and displayed in Figure 4.54, in terms of torsional angles as defined in the inset. The plot indicates clearly that the presence of cation leads to dramatic changes of the orientation of the catalytic residues, especially for the Asp64 and Asp116. Without the Mg^{2+}, two preferential conformations of the Asp64 at the torsional angles 65° and -155° were observed (Figure 4.54 a, dash line). For the FULL+ION system, the presence of the Mg^{2+} ion induces the two preferential conformations of the Asp64 in the FULL to give only one stable conformation at the torsional angle of 13°. (Figure 4.54 a, solid line). The angle of 30° and 41° were extracted from the structure published by Podtelezhnikov *et al.* and Luca *et al.*, respectively. The conformations of the Asp64 in both the FULL and FULL+ION systems are comparable to those of the two published full-length models. In the Asp116 catalytic residue, the two preferential conformations in the FULL system (χ_{116} = -125° and 65°, Figure 4.62 b, dash line) were shifted by the Mg^{2+} to form the most stable structure at -145° and another minor conformation at 155° (Figure 4.54 b, solid line). The situation is different for the Glu152 torsional angle in which its conformation is highly flexible in both FULL+ION and FULL systems (Figure 4.54 c). This is understandable because this residue is not located within the solvation shell of the Mg^{2+}. The detected data agree well with the previous report on the catalytic core domain by Lins *et al.* which stated that the Glu152 shows large fluctuations. **(24)**

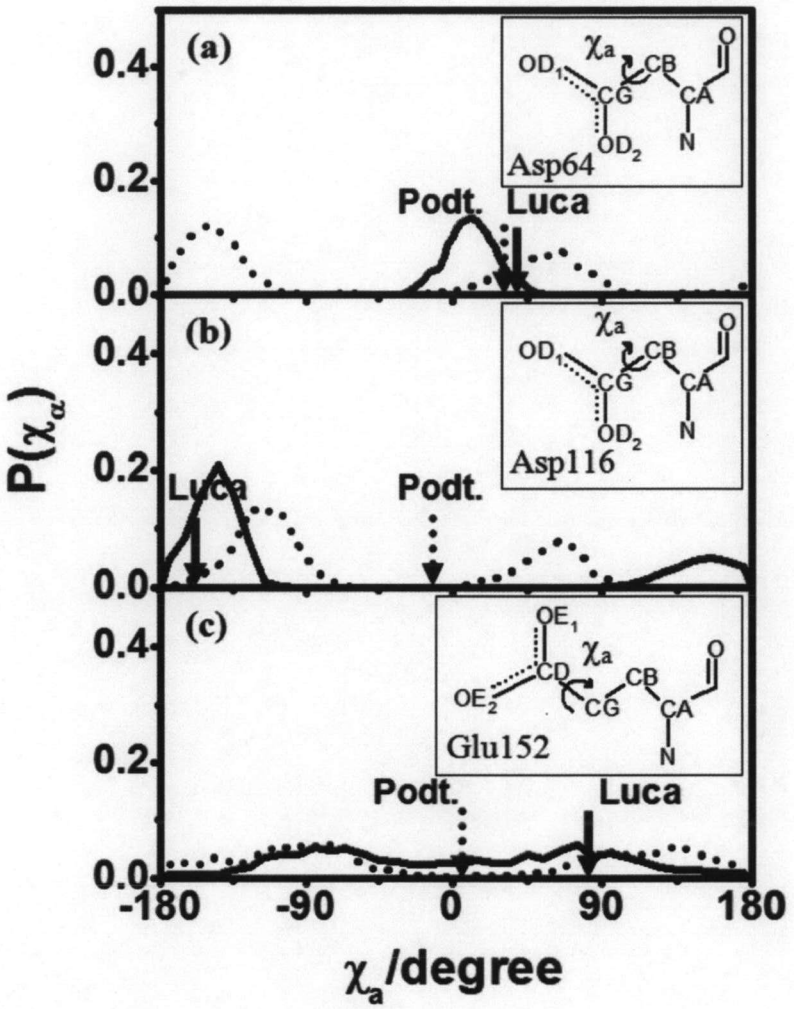


Figure 4.54 Distribution of the torsional angles (χ_a) of each catalytic triad; χ_{64} (a), χ_{116} (b) and χ_{152} (c), for the FULL+ION (solid line) and the FULL (dash line) systems. The values measured from Podtelezhnikov's (dash arrow) and Luca's (solid arrow) models were also plotted.

4.3 Molecular docking

4.3.1 Dimer model bound to DNA

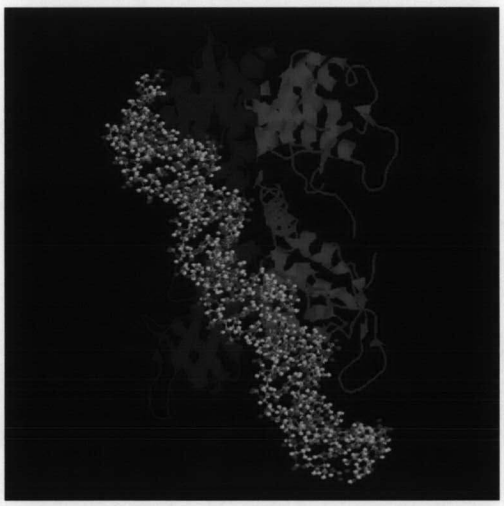
(a)  (b) 

Figure 4.55 Dimer full-length complexed with viral DNA in side view (a) and front view (b). Subunits A and B of the dimer were colored as red and blue, respectively.

The dimer structure complexed with viral DNA with the sequence TAGTCAGTGTGGAAAATCTCTATAGCAGT **(102)** was also reported herein (Figure 4.55). The DNA was built up in B-form using nucgen module in AMBER7. The viral model was manually docked into the dimer structure followed by fully optimization. The docked structure shows that the viral DNA binds to the HIV-1 IN dimer in *trans* action similar to what proposed by previous studies **(54, 63)**. The HIV-1 IN chain B contributes the catalytic core domain while chain A contributes the C-terminal domain. This binding character allows the viral DNA to have some contact with the N-terminal DNA which is suggested by the cross-linking studies that this region may also involve in DNA binding. The amino acids which have some contact with the viral DNA were reported in Table 4.3.

4.3.2 Tetramer model bound to DNA

The wild type tetramer model bound to DNA proposed in this study was shown in Figure 4.56. The DNA molecules interact with chain B (blue) and chain D

(green) of the tetramer full-length. Table 4.4 shows the amino acid which interacts with both viral and host DNA.

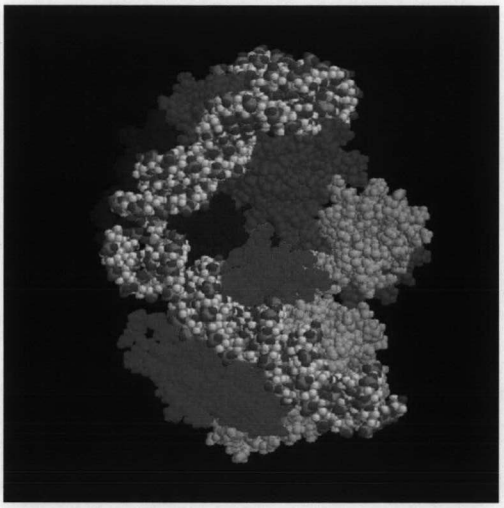
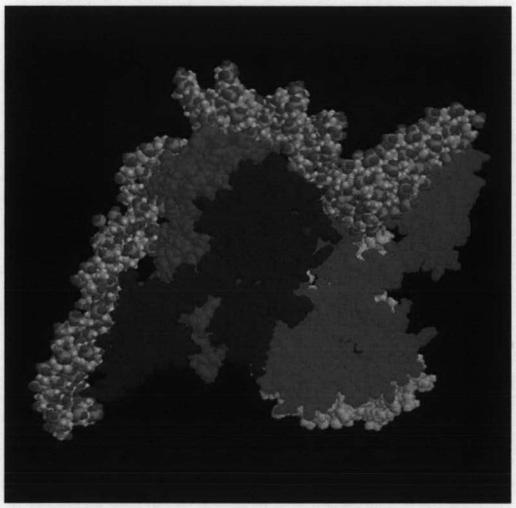
(a)  (b) 

Figure 4.56 A wild type tetramer HIV-1 IN complexed with both viral and host DNA proposed in this study, top view (a) and side view (b). Each monomer unit was colored as red, blue, yellow and green for chains A, B, C, and D, respectively.

4.3.3 HIV-1 IN-DNA interaction

In the dimer complexed with viral DNA, both monomer subunits have contribution in binding with DNA as described above. In the case of the tetramer model, the host DNA contact with the core domain of subunits A and B while the viral DNA interacts with core domain and C-terminal domain of another chain. It also contact with the N-terminal domain (Figure 4.57). List of the interaction was given in Table 4.3. The interactions reported from previous studies were also included in the Table.

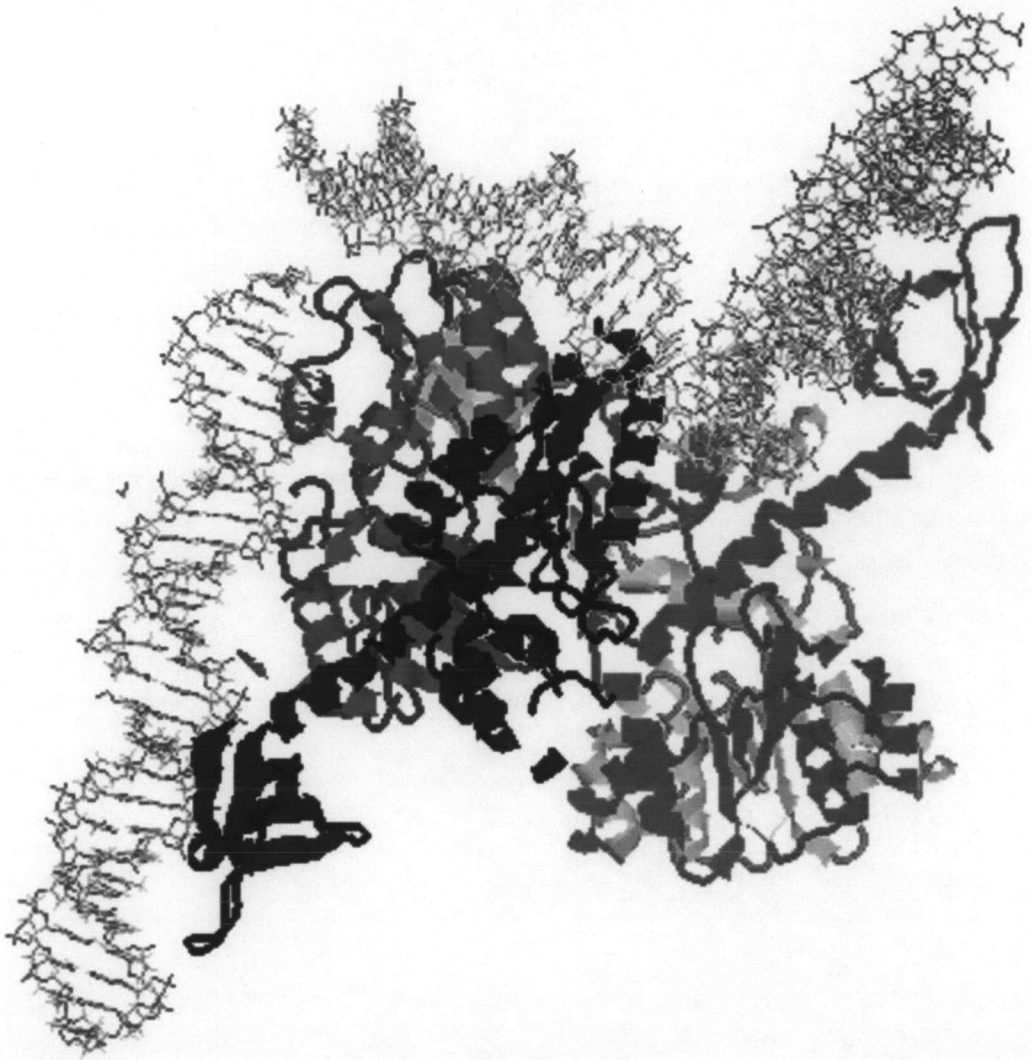


Figure 4.57 A model of wild type tetrameric HIV-1 IN complexed with DNA. The integrase monomers are colored as red, blue, yellow and blue as for chains A, B, C and D, respectively.

Table 4.3 List of the amino acid residues which interact with the viral and host DNA from various studies.

HIV-1 IN model	DNA	Method	Contact	Ref.
Core and C-terminal model	DNA substrate, Single viral DNA end	Photo-cross-linking	His67, Asn117, Glu152, Ser153 Arg263, Lys264, Ser230,	**54**
Core and C-terminal model	Viral cDNA, Target DNA	Cross-linking	Glu246, Ser230	**52**
Full-length (tetramer)	Viral DNA Target DNA	Modeling	Glu246, 247 – 270	**64**
Full-length (dimer)	Viral DNA	Docking	Chain B: Arg20 Chain B: Lys156, Lys159, Lys160, Lys186, Lys188 Chain A: Ser230, Arg231, Trp243, Lys244, Lys263 and Lys264	**63**
Core domain	Viral DNA, Genomic DNA	Auto docking	Lys111, Lys136, Glu138, Lys155, Lys 148, Leu160, Lys186, Lys182	**62**
Core domain	Viral DNA	Fast molecular docking	Asp64, Asp116, Glu152, Lys156 and Lys159	**104**

Table 4.3 (cont.)

HIV-1 IN model	DNA	Method	Contact	Ref.
C-terminal		Auto Docking	R228, K244, E246, R262, R263, K264, K266, I267 and R269	**61**
FULL+ION (dimer)	Viral DNA	Docking	Chain B: His12, Tyr15, His16, Arg20 Chain B: Lys156, Chain A: Pro261, Arg262, Lys264,	**This study**
FULL+ION (tetramer)	Viral DNA Target DNA	Docking	Chain B: His12, His16, Asn18, Arg20, Asp25, Lys156, Lys46, Ile191, Gly190, Gly192, Chain A: Ser230, Arg231,Trp243, Pro261, Arg262, Arg263, Lys264	**This study**

4.4 Summary

4.4.1 Molecular modeling

We built up the missing regions in the crystal structure of the CORE-N and CORE-C. By comparison with the one-domain structure, CORE, and the three-domain full-length system, these regions show high different in the structure. These observations support the previous studies which stated that these regions were experimentally difficult to be solved.

We have also proposed an alternative model structure of the full-length HIV-1 IN. This model was carefully built up based on the experimental structures of two-domain fragments, which provided reasonable linkages between each domain. By comparing to the two previously published structures, the molecular backbone of our full-length model is consistent with that proposed by Luca *et al.* while different from that of Podtelezhnikov *et al.* Despite the similarity in the protein backbone, significant differences were observed in the detailed structure and orientation of the catalytic residues between our model and that of Luca *et al.* The dimer and tetramer forms of the HIV-1 IN were also reported herein. All divalent metal ions in the active site were placed to the model based on the relative position in the available crystal data of the one- and two-domain structures.

4.4.2 MD simulations

The structural and dynamical behavior of the HIV-1 IN was elucidated by the employment of the MD technique. The 2-ns MD simulations of the one-, two- and three-domain systems were carried out. Thermodynamic properties such as energies and temperature were monitored for all system during the 2-ns MD trajectories. The average temperature of the system is at about 300 K for all systems.

The dynamical behaviors of the various domain of the HIV-1 IN were figured out from the MD simulation trajectories. The RMSD values indicate the flexibility of the structure. The results show that the RMSD over the simulation time of all systems fluctuate in the range of 1.0 – 3.0 Å, which is the usual phenomenon for the protein dynamics, implying the reliability of the MD results.

The RMSD per residues reveals that, among the three domains, the C-terminal domain has largest fluctuation. This fluctuation is due to the flexibility of the elbow linkage between the core and the C-terminal domains. Such fluctuation may be involved in the binding of DNA during the integration process. The larger fluctuations are due to the two regions corresponding to residues 140 – 148 and 188 – 192 as can be seen from the RMSD per residue plot extracted from the MD trajectories from the simulations.

The active site region containing the two aspartate residues, Asp64 and Asp116, binds the Mg^{2+} while a glutamate residue, Glu152, does not interact with this ion. The torsional angle plots of these catalytic residues display the high flexibility of the Glu152. This result agrees well with the previous MD studies and the experimental data.

For the effect of the metal ion in the active site, two 2-ns MD simulations of the full-length HIV-1 IN without and with metal ions in the active site were performed. The results provide insight into how a metal ion affects both structure and dynamics of the active site region. The size of the active site pocket is decreased by the presence of the metal ion. Unlike the previous studies **(24, 57)**, the Mg^{2+} ion in the active site of the core domain in all systems has an octahedral coordination configuration with two oxygen atoms of Asp64, one oxygen atom of Asp116 and three oxygen atoms of the water molecules. Analysis in details of the dynamical properties, the metal ion was observed to influence the mobility of the active site residues, in which the mobilites of the Asp64 and Asp116 increase while that of the Glu152 decreases.

4.4.3 Molecular docking

It should be noted that substrate recognition by the retroviral DNA is critical for retroviral integration. HIV-1 IN must recognize and acts on two types of substrate, viral DNA and host DNA. HIV-1 IN interacts with DNA in both core and C-terminal regions. The interaction between the viral DNA and the N-terminal domain was observed. The main function is to promote the multimerization of the structure. The interaction is consistent with the previous model structure and the experimental data. Here, the alternative model for the dimer complexed with viral

DNA was examined. Interactions between the following amino acid residues, His12, Tyr15, His16, Arg20 in the N-terminal region, Lys156 in the core domain of chain B and Pro261, Arg262, Lys264 in the C-terminal domain of chain A were observed. The tetramer full-length HIV-1 IN complexed with both viral and host DNA was also reported herein. The target DNA was docked to the core domain at the sites of integration, where viral and target DNA are joined near the active site region. The viral DNA interacts with all three domains. The structure was subjected to optimize prior to MD simulations.