

Chapter VI

*Molecular Dynamics Simulation Receptor-Binding C-Terminal
Domain from Clostridium difficile Toxin A*

6.1: Introduction

Gastrointestinal diseases like antibiotic-associated diarrhea are caused by the gram positive bacteria like *Clostridium difficile* (Bartlett *et al.*, 1978; George *et al.*, 1978; Bartlett *et al.*, 1980a; Kelly & LaMont 1998). A strong correlation between pseudomembranous colitis, antibiotic therapy, *C. difficile* colonization and cytotoxin production are reported by various researchers (Aronsson *et al.*, 1981; Bartlett *et al.*, 1980b; Burdon *et al.*, 1981; Don & Devis 1981; George *et al.*, 1982; Meyers *et al.*, 1981; Tedesco, 1981; Willey & Bartlett 1979). Protective antibodies produced by the host immune system against the toxins of *C. difficile* play an important role in reducing the severity of disease. Despite the lack of understanding of the epidemiology of *C. difficile* disease, there has been a considerable progress towards the understanding of organism's mechanism of virulence. The toxins, TcdA and TcdB were studied intensively as major *C. difficile* virulence factors. *C. difficile* toxins A and B belong to the family of large clostridial cytotoxins. They are often called clostridial glucosylating toxins, as their toxic potency depends on their glucosyltransferase activity (Von Eichel-Streiber *et al.*, 1996; Schirmer & Aktories 2004; Aktories & Barbieri 2005; Aktories & Just 2005; Voth Ballard 2005). *C. difficile* disease is caused by TcdA and TcdB and is detected in the stools of patients by antibody-based and cytotoxicity assays.

These two toxins, along with the other members of the large clostridial toxin family of toxins, target the Ras super family of small GTPases for modification via glycosylation. This irreversible modification inactivates these small regulatory proteins and disrupts the signaling pathways in the cell. Along with enzymatic modification of targets, there are also several important steps in receptor-binding and cell entry necessary for intoxication.

In *in vivo* studies of toxins, it was found that toxin B was unable to cause disease in the absence of toxin A. In contrast, purified toxin A was capable of causing disease even in the absence of toxin B (Lyerly *et al.*, 1985).

The experimental evidence for receptor-binding by the C-terminus of TcdA comes from a series of studies showing the neutralizing capacity of monoclonal antibodies to this region of the protein (Frey & Wilkins 1992) and the ability of recombinant fragments from this region to protect against the toxin. TcdA-induced cytotoxicity is blocked by a fragment of C-terminus by the competitive inhibition of receptor-binding (Sauerborn *et al.*, 1997). C-terminal receptor-binding domain was shown to be necessary for both receptor-binding and effective endocytosis of the holotoxin (Frisch *et al.*, 2003).

The C-terminal Repetitive Domain (CRD) binds carbohydrates on colonic epithelial cells as an initial step in pathogenesis. Various oligosaccharides, including the linear B-type 2-trisaccharide α -Gal-(1, 3)- β -Gal-(1, 4)- β -GlcNAc, bind specifically to TcdA (Tucker & Wilkins 1991; Teneberg *et al.*, 1996). The functional importance of carbohydrates binding to TcdA is supported by previous studies. TcdA-f2 consists of nine Short Repeats (SR) and two Long Repeats (LR). First, diethylpyrocarbonate modification of histidine residues in TcdA specifically abolishes cytotoxicity and receptor-binding activities (Roberts & Shone 2001).

Ho *et al.* (2005) solved the crystal structure of a 127-aa fragment of C-terminal repeats from *C. difficile* toxin A with the 1.85 Å resolution. The structure of a 127-residue C-terminal fragment of TcdA (TcdA-f1) is highly repetitive which contains SR and LR. The crystal structure of TcdA-f1 reveals that each SR or LR contains a single β -hairpin consisting of a pair of five- to six-residue antiparallel β -strands connected by a tight turn. The first short repeat (SR1) starts from Gly14 residue and ends with Gly33 residue. The second short repeat (SR2) consists of Arg64 residue to Thr83 residue. The third short repeat (SR3) starts

from Gly84 and ends at Gly105. The fourth short repeat (SR4) starts from Gly106 to Gly126. Each SR consists of two-beta strands connected by a turn and one repeat unit is connected with other repeat by a loop. There is an LR, which is in between the two short repeats SR1 and SR2. The LR starts from Gly33 residue and ends with Arg64 residue. It has also two strands connected by a turn.

The 3D structure of TcdA-f1 suggests that the boundaries of each SR or LR should be defined to coincide with the beginning of the β -hairpin and the end of the connecting loop preceding the following β -hairpin. Each β -hairpin interacts with both the preceding and following β -hairpins, except for the N- and C-terminal hairpins. The N-terminal end of TcdA-f1 adopts a non-natural structure due to the truncation of the protein. There is a hairpin in N-terminal hairpin, comprising hydrophobic residues which form a non-natural intermolecular interface with a few exposed hydrophobic residues of C-terminal.

In the present work, we have done molecular dynamics simulation and Principal Component Analysis (PCA) of C-terminal domain from *Clostridium difficile* toxin A to understand its global and local motional properties which will help to understand its function and mode of action.

6.2: Materials and Methods

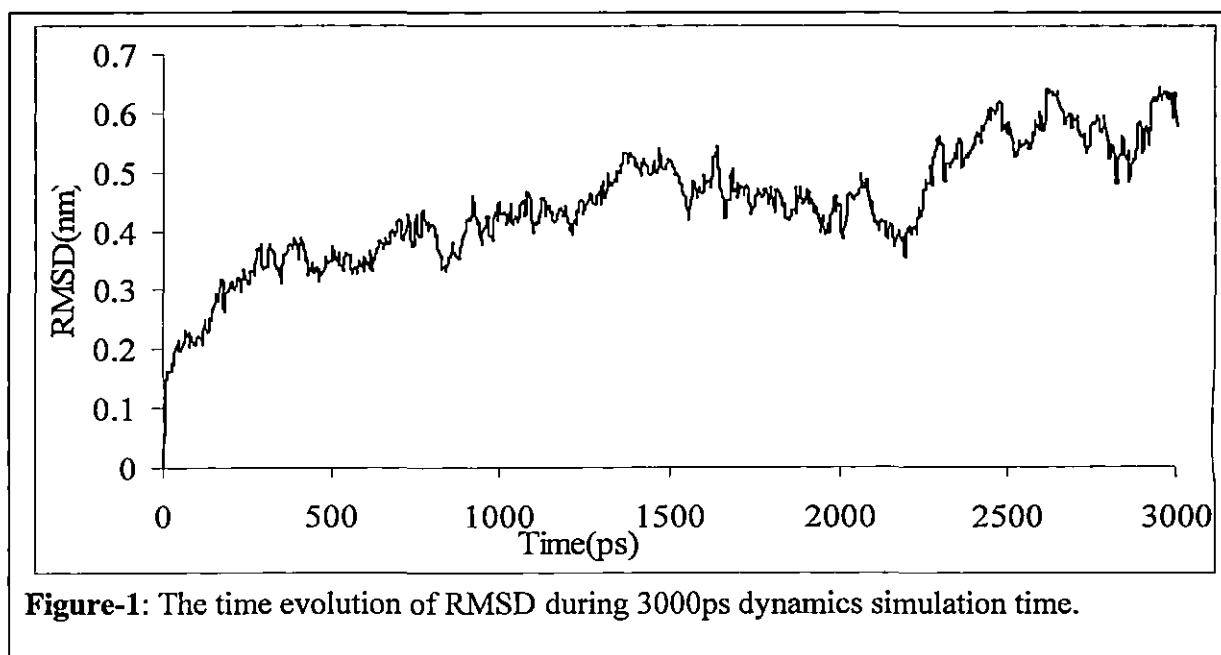
The 1.85 Å resolution x-ray structure of *C. difficile* toxin A (Protein Data Bank code 2f6e) (Ho *et al.*, 2005) was used as a starting structure. A single monomer was solvated with SPC water molecules in a cubic box, having an edge length of 35 Å. The simulation was performed using Groningen Machine for Chemical Simulation (GROMACS) (Lindahl *et al.*, 2001). The LINCS algorithm was used to constrain all bond lengths (Hess *et al.*, 1997). The simulation was conducted at a constant temperature (300 K), coupling each component separately to a temperature bath using the Berendsen coupling method (Berendsen *et al.*, 1984). A cutoff of 0.9 nm was used for Lennard Jones interaction and 1.0 nm for Coulomb

interaction (Darden *et al.*, 1993). The time step was 2 fs, with coordinates stored after every 4 ps. MD simulation was performed for 3 ns. Before running simulation, an energy minimization was performed in steepest descent method (converged at 523 steps), followed by conjugate gradient method (converged at 8 steps) (Essmann *et al.*, 1995; Bothra *et al.*, 1998); and this was followed by 1.0 ns of simulation imposing positional restraints on the non-H atoms. The positional restraints were then released and 3 ns production run were obtained and analyzed. Analysis programs from GROMACS were used and PCA was performed with the MD trajectory.

6.3: Results

The overall structural stability of the protein during the simulation was monitored using several parameters like Root Mean Square Deviation (RMSD), Radius of gyration (Rg), Root Mean Square Fluctuation (RMSF), etc.

The time evolution of RMSD is computed taking the constrained structure of the whole protein as initial structure and presented in Fig.1.



The time evolution of Rg is presented in Fig.2,

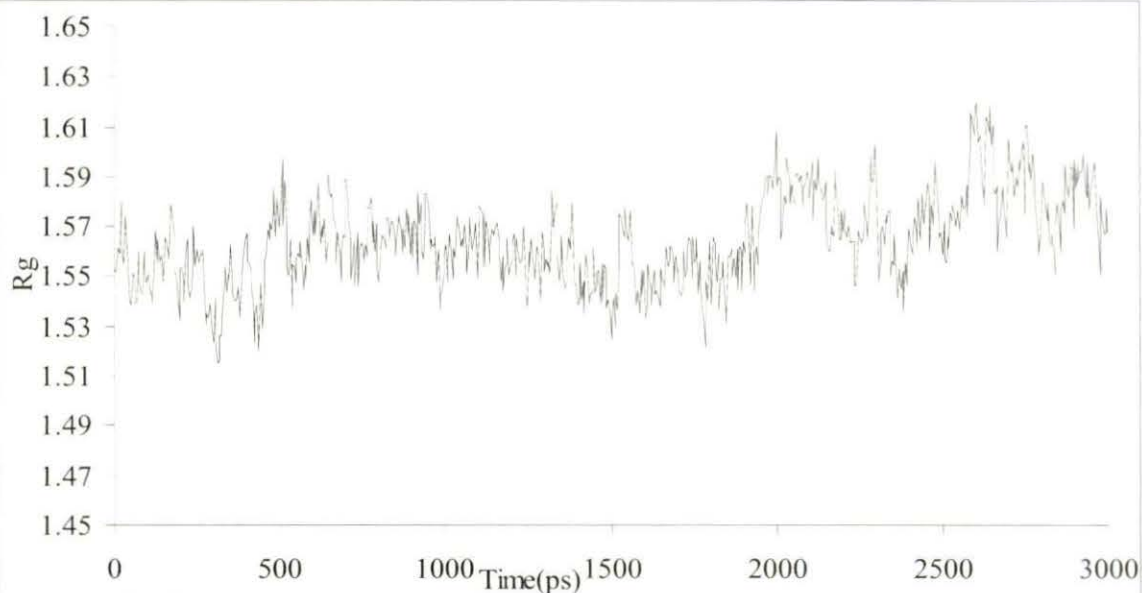


Figure-2: Time evolution radius of gyration changes in aqueous medium during 3000ps dynamics simulation.

RMSF indicates the flexibility of the protein. RMSF of $C\alpha$ and B factors are presented as a function of residue numbers in Fig.3.

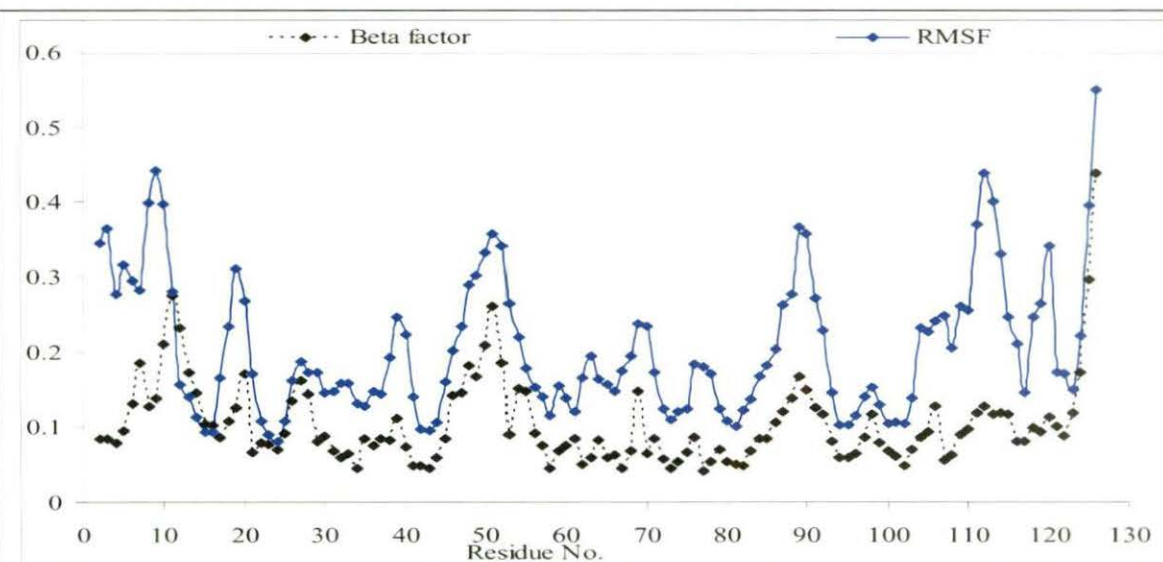
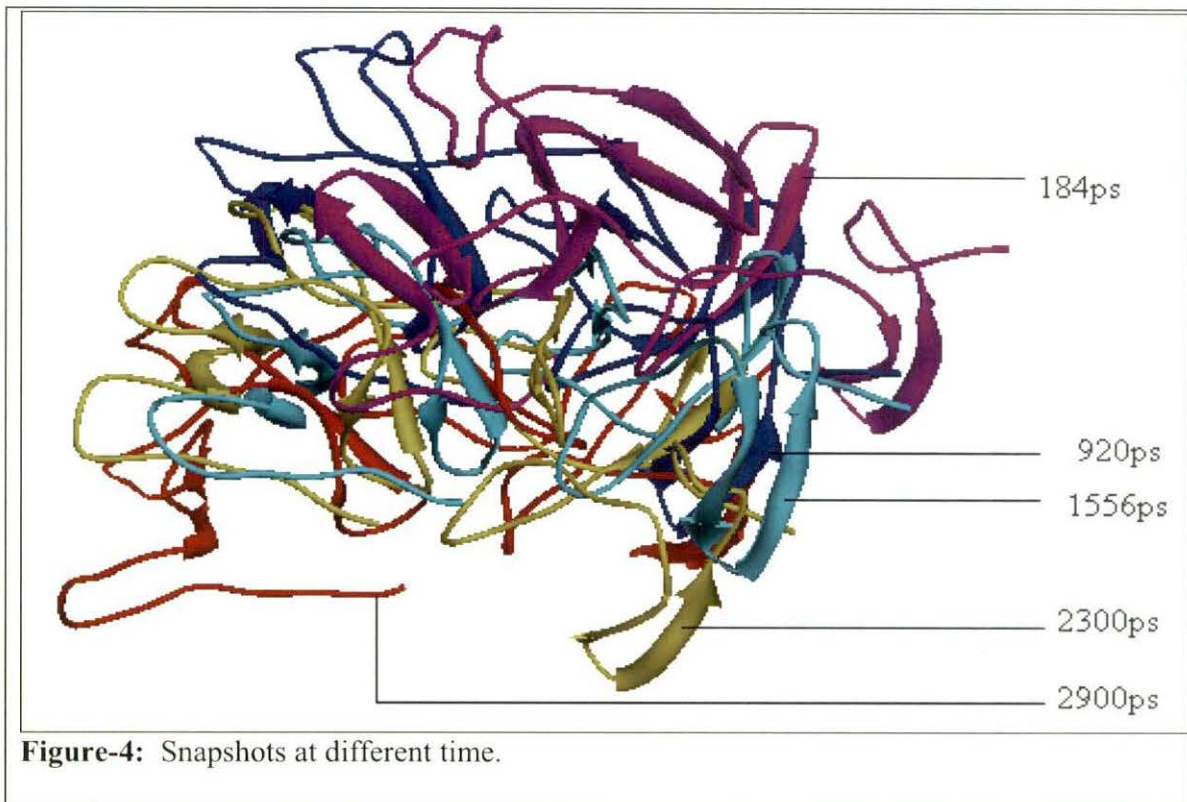


Figure-3: Plot of the RMSF of $C\alpha$ atoms value in aqueous medium and B factors is presented as a function of residue number of *Clostridium difficile* toxin A in starting X-ray structure.

We have taken some selected snapshots from the dynamics trajectory considering time evolution of RMSD as a guideline. The snapshots were taken when the RMSD from initial structure was high, and they are represented in Fig.4.



We have also calculated the RMSD of the SR1 and LR. We have determined the RMSD values of the SR and LR and their structural elements, which is given in Table 1.

Table 1: Mean RMSD values of different repeating unit including strand, turn and Loop

Repeat Unit	Overall	Strand I	Strand II	Turn	Loop
SR1	0.196788	0.107284	0.173618	0.145548	0.170353
LR	0.236153	0.103893	0.110976	0.115819	0.254105
SR2	0.239528	0.212295	0.139324	0.169818	0.216462
SR3	0.170861	0.124877	0.143989	0.123328	0.165811
SR4	0.230572	0.109931	0.141252	0.104894	0.245512

Time evolution of RMSD for different repeating units are presented in Fig.5.

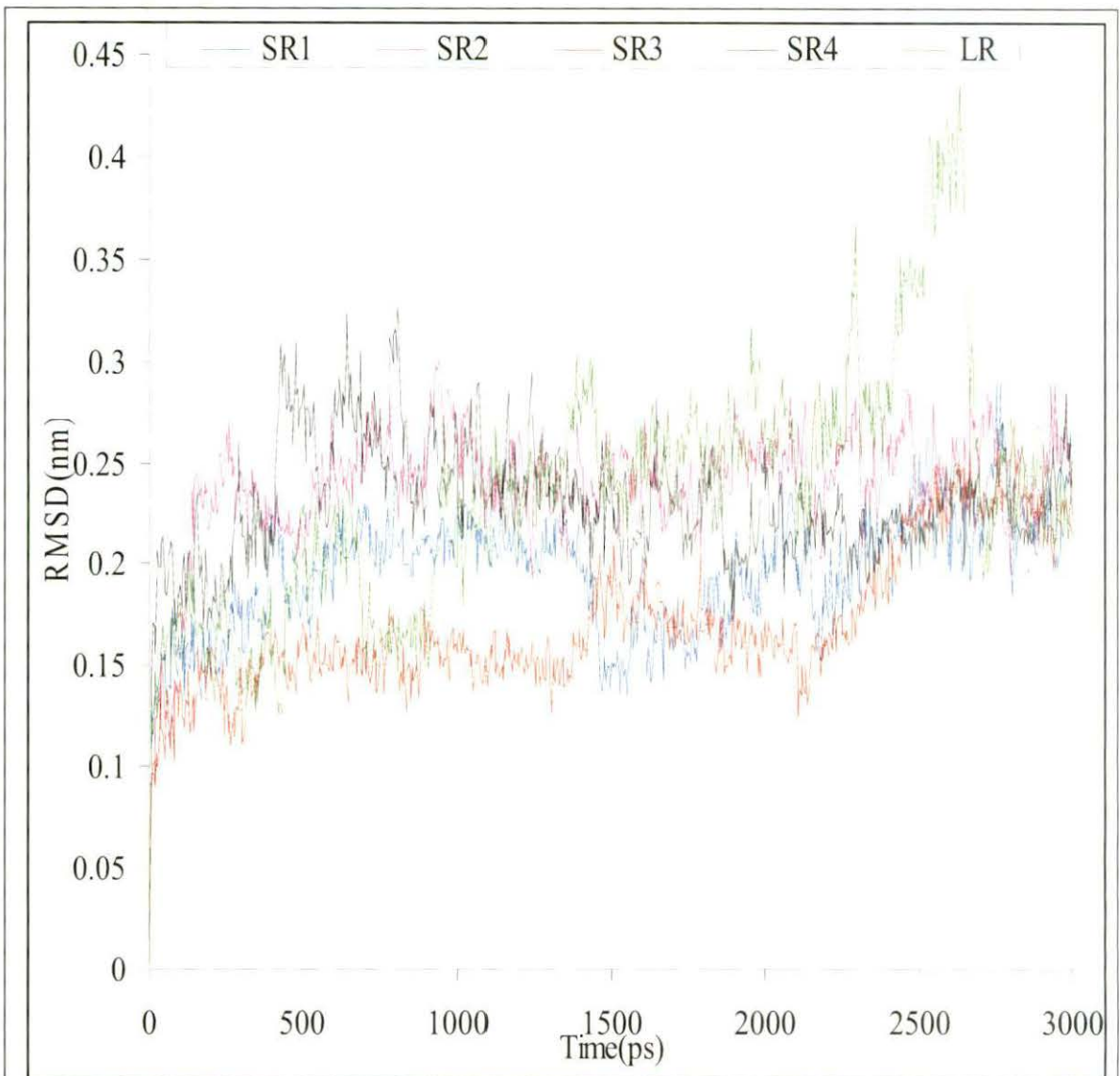


Figure-5: Time evolution of RMSD of different repeating unit of *Clostridium difficile* toxin A.

The average RMSD for the overall strands of different repeating units are represented in a histogram (Fig.6) and it is seen that strand I of SR2 and strand I of LR showed the highest and lowest fluctuations respectively among all strands. Time evolution of RMSD values of strand I of SR2 and strand I of LR are given in Fig.7.

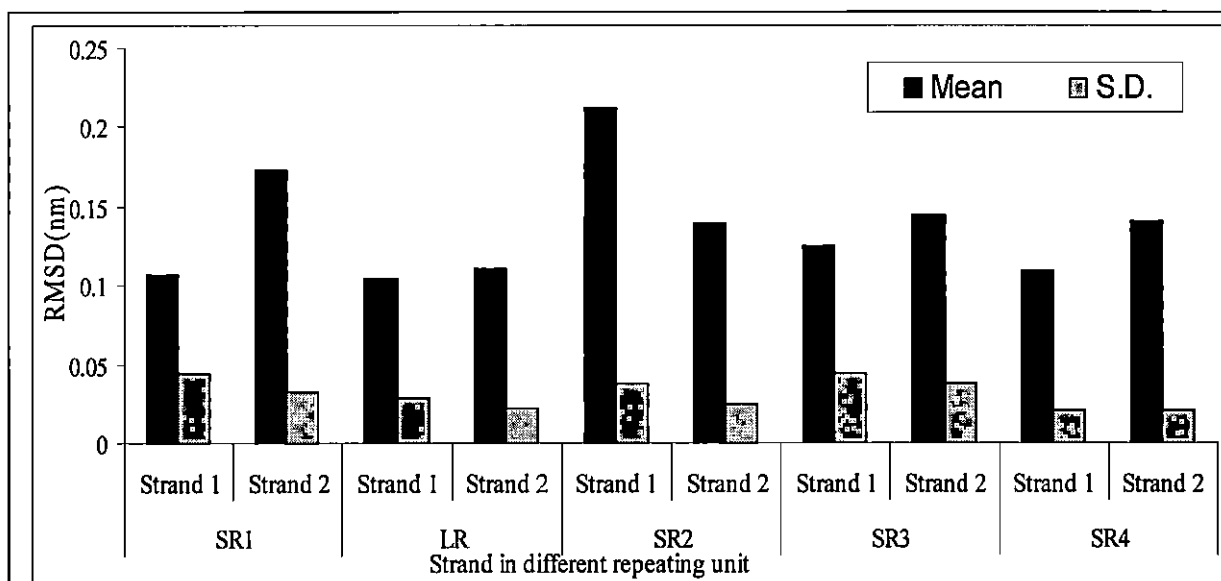


Figure-6: Histogram of RMSD for the all the strands of different repeating unit

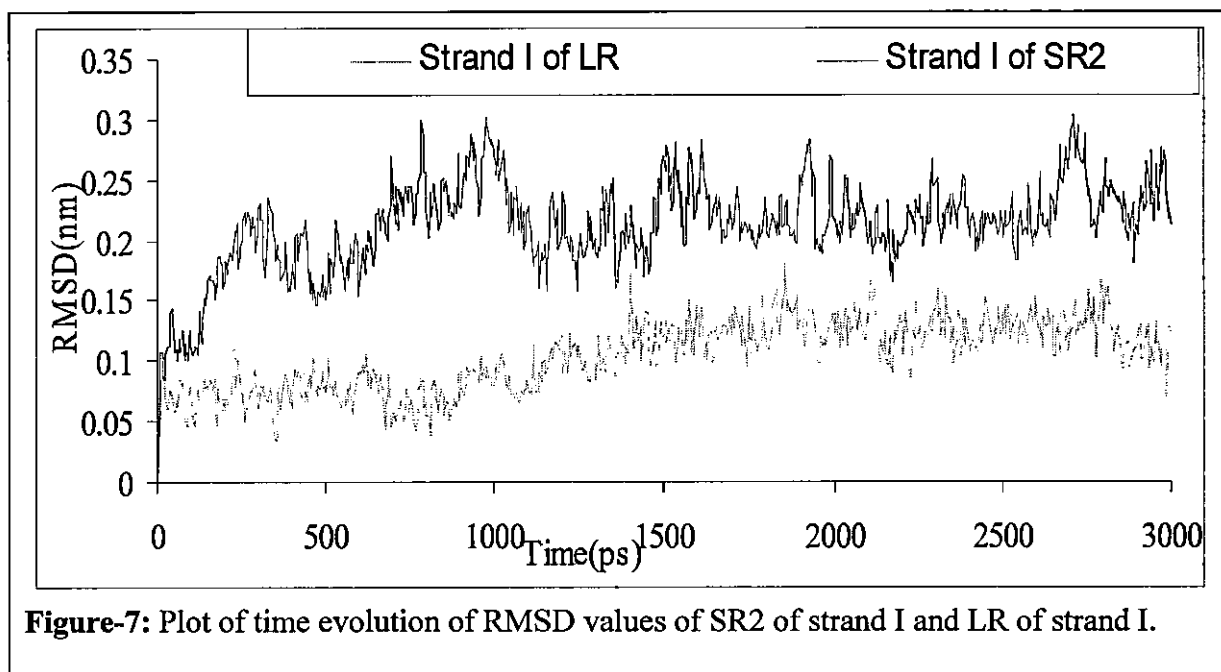


Figure-7: Plot of time evolution of RMSD values of SR2 of strand I and LR of strand I.

A common approach in the identification of the major motions of a protein is the use of PCA (Garcia, 1992; and Amadei et al., 1993). PCA reduces the dimensionality of a complex data set and applied to decompose a complex motion of proteins, which are characterized by an eigenvector and an eigenvalue. The eigenvalue for a given motion represents the contribution of the corresponding eigenvector to the global motion of the protein. PCA of the *C. difficile* toxin A simulation reveals that the first 10 eigenvectors account for 87.16%

of the global motion and that the first eigenvector corresponds to 49.68% of the total motion, the second to 15.32%, and the third to a further 5.90% (Fig.8).

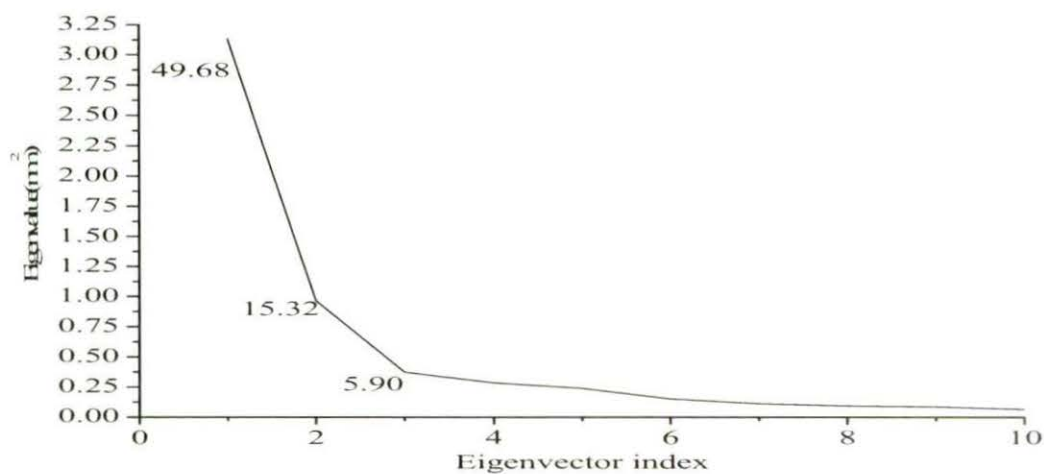


Figure-8: Plot of eigenvalues with eigenvector indexes.

The protein comprises one LR and four SR. Each repeat consists of two beta strands followed by a loop.

The projection of RMSF on four vectors is presented in Fig.9.

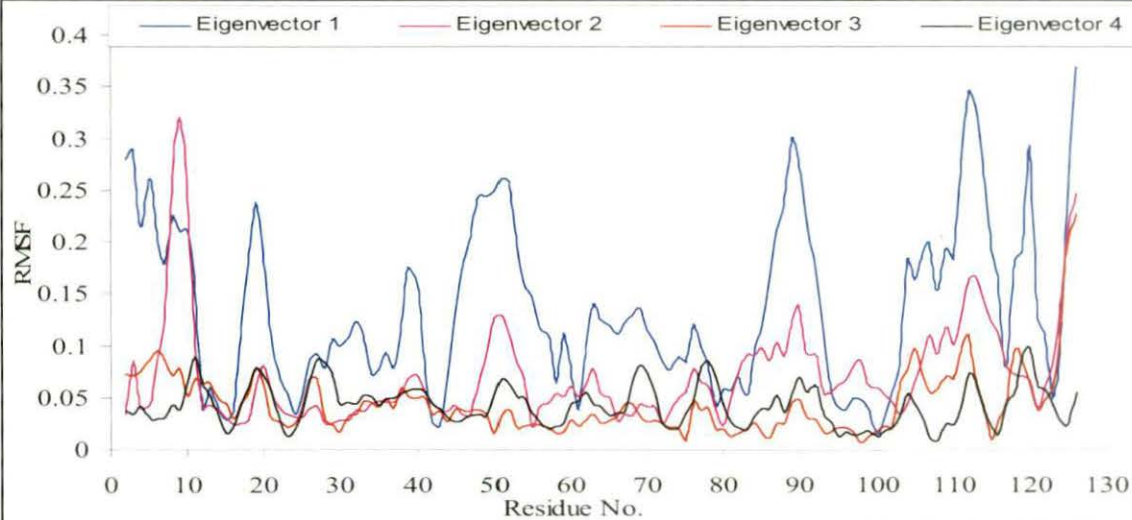


Figure-9: Plot of the projection of RMSF on four vectors.

Time evolution of principal component 1 (PC1), principal component 2 (PC2), principal component 3 (PC3) and principal component 4 (PC4) in water is represented in Fig.10.

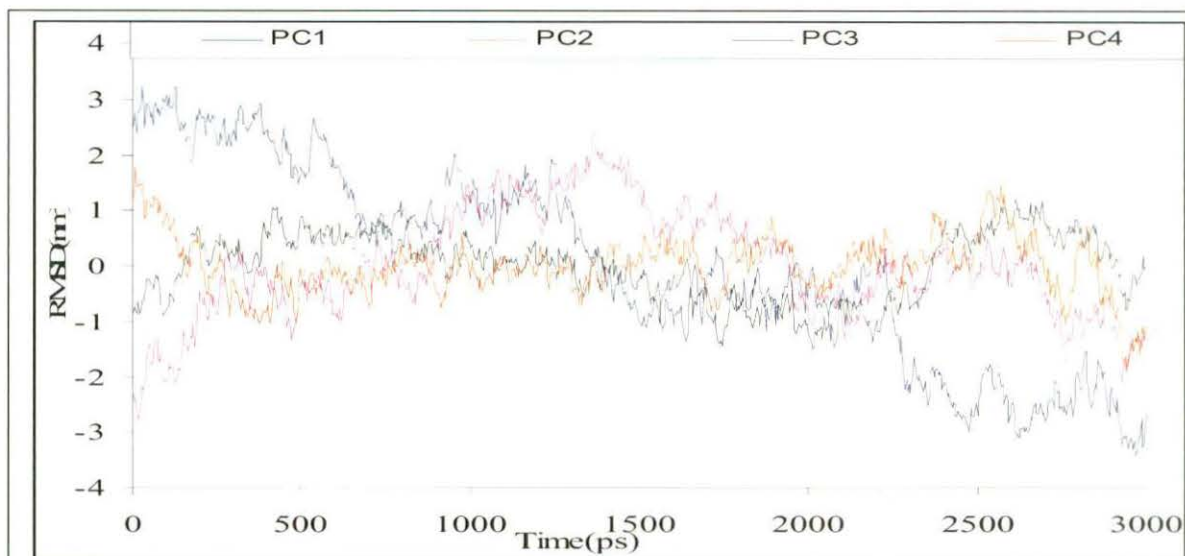


Figure-10: Plot of four principle components (nm) with simulation time

The projection of the dynamics trajectory onto the first two eigenvectors was analyzed and presented in Fig.11.

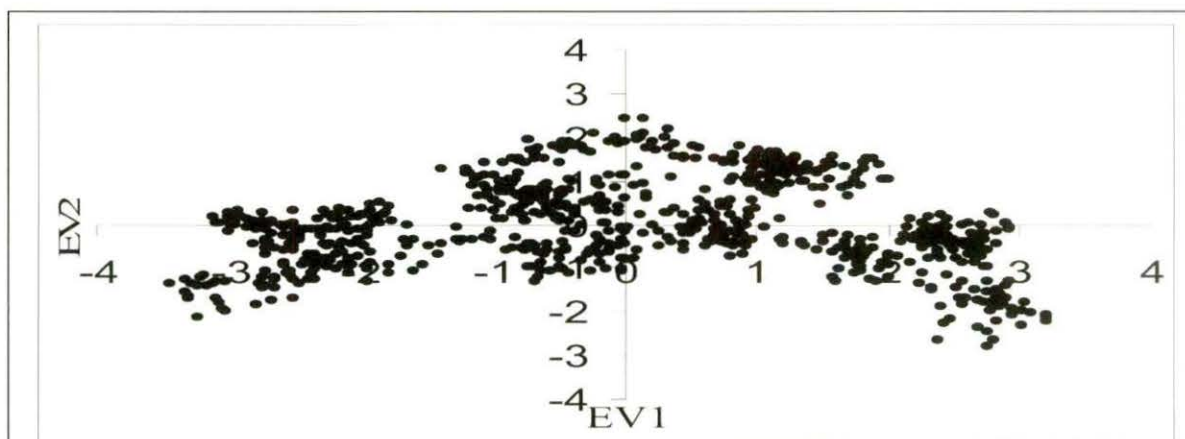
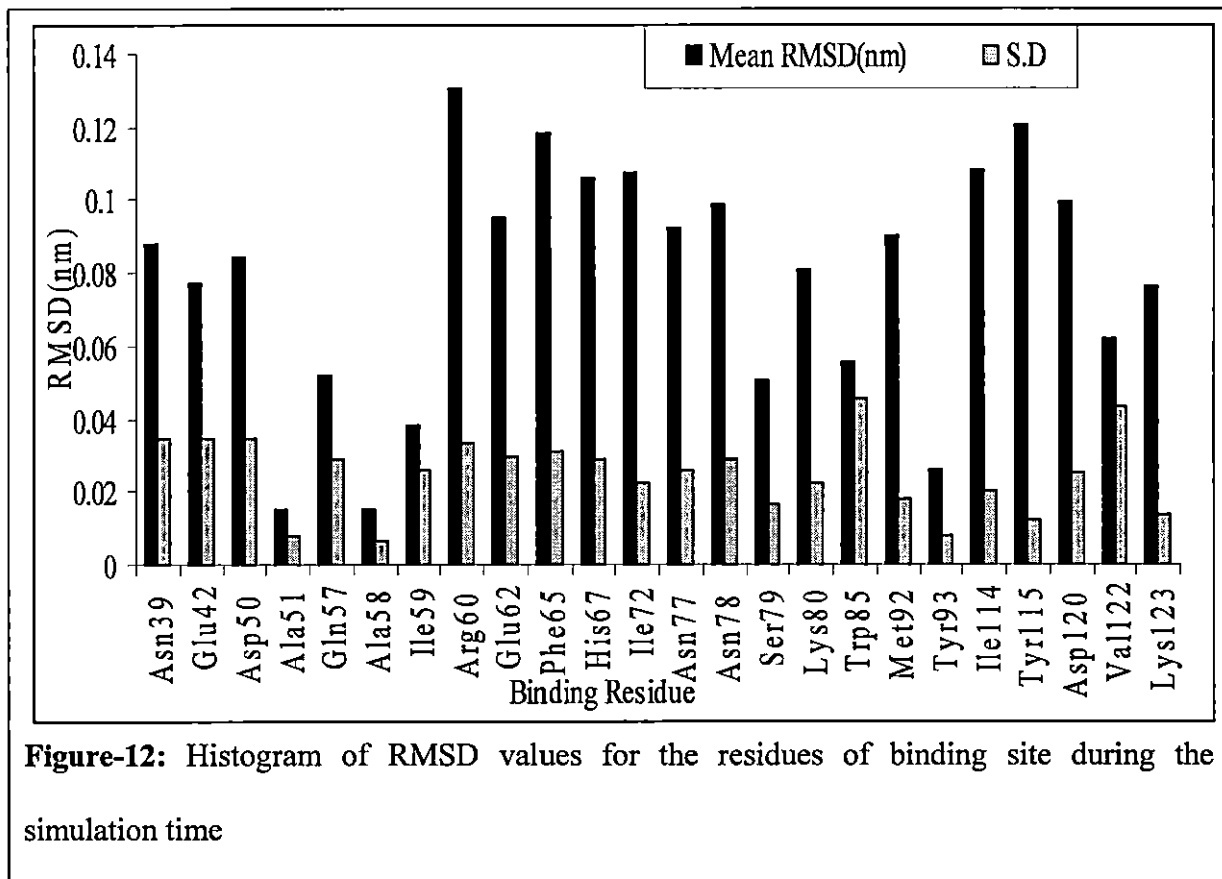


Figure-11: Plot of eigenvector1 with eigenvector 2.

According to the report of Greco et al. (2006), the crystal structure of the C-terminal 255 residues of TcdA from *C. difficile* strain 48489, toxinotype VI (TcdA-f2) found to be a synthetic derivative of the natural carbohydrate receptor, α -Gal-(1, 3)- β -Gal-(1, 4)- β -GlcNAcO(CH₂)₈CO₂CH₃ (CD-grease). The structural feature and sequence similarity of LR as well as the similar binding mode indicates the conserved nature of the seven

carbohydrate binding sites in TcdA-f2 which consist of an LR and the hairpin turn of the following SR.



To get insight about the binding mechanism, the residues of binding site were critically analyzed. The residues Lys80, Arg60, Ser79, Asp50, and Gln57 are very important for binding processes, as pointed out by Greco et al. (2006). From the RMSD values during the simulation time (Fig.12), it is evident that the residues, Ala51, Ala 58, Ile59 and Tyr93 have very low fluctuation. The residues, Arg60, Phe65, His67, Ile72, Ile114 and Tyr115 have high fluctuation, and the fluctuation was moderate for the residues Asn39, Glu42, Asp50, Gln57, Ser79, Trp85, Val122 and Lys123 during the whole simulation time (Mean RMSD and SD value is given in Table 2).

Table-2: Mean RMSD and S.D. of different residue binding site.

Binding residue	Mean RMSD(nm)	S.D
Asn39	0.087781	0.034847
Glu42	0.077564	0.034463
Asp50	0.084088	0.034517
Ala51	0.01535	0.007662
Gln57	0.052091	0.028836
Ala58	0.015217	0.006695
Ile59	0.038305	0.026329
Arg60	0.130462	0.033052
Glu62	0.095121	0.029304
Phe65	0.118562	0.031224
His67	0.106327	0.028664
Ile72	0.107694	0.022652
Asn77	0.09213	0.025975
Asn78	0.098776	0.02911
Ser79	0.050197	0.016265
Lys80	0.080878	0.022682
Trp85	0.055926	0.045661
Met92	0.090154	0.018117
Tyr93	0.026039	0.007814
Ile114	0.108123	0.020269
Tyr115	0.120166	0.012418
Asp120	0.099466	0.02529
Val122	0.062328	0.043586
Lys123	0.076151	0.01389

The 6-OH of β -galactose accepts hydrogen from Lys80 which has RMSD value 0.080878. The B-face of β -galactose packs against the apolar proximal portions of Arg60 which has RMSD value 0.130462, and Ser79 has RMSD value of 0.050197. The α -galactose at the nonreducing end lies in a pocket formed by highly conserved residues in the loop after the β -hairpin of the LR. Its hydrophobic B-face packs against Ile59 with an average RMSD value 0.038305 and the 6-OH forms hydrogen bonds with Asp50 (RMSD 0.084088) and Arg60 (RMSD 0.130462), whereas the 4-OH accepts hydrogen from Gln57 (RMSD 0.052091). During the simulation, several hydrogen bonds broke and formed. It is found that the number of hydrogen bonds ranged from 54 to 85 (Fig.13).

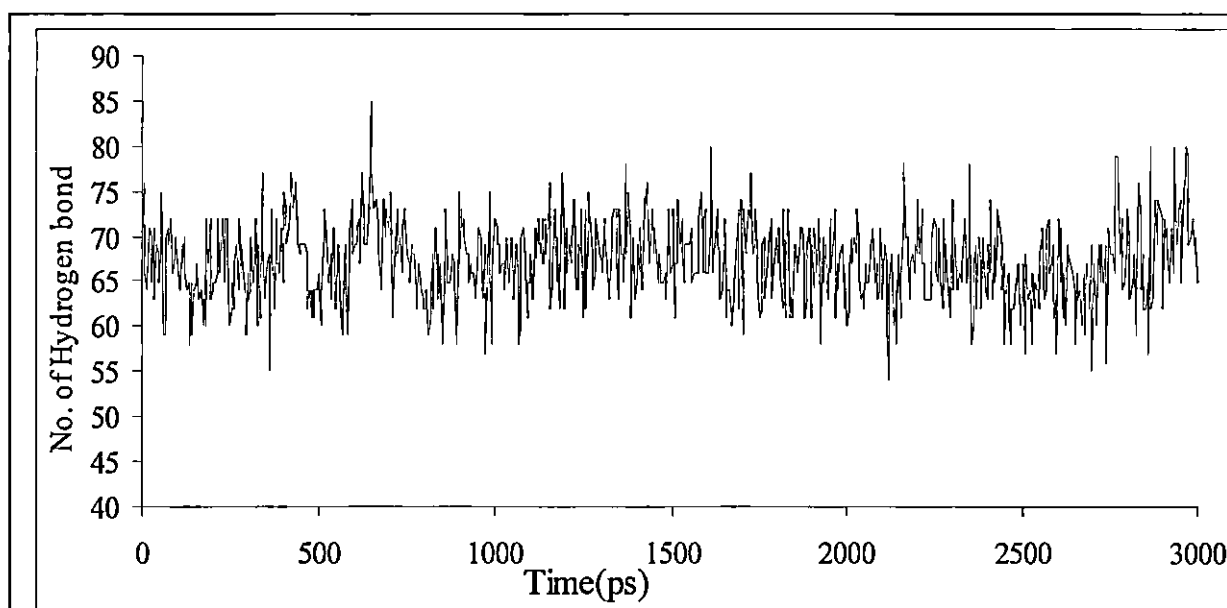
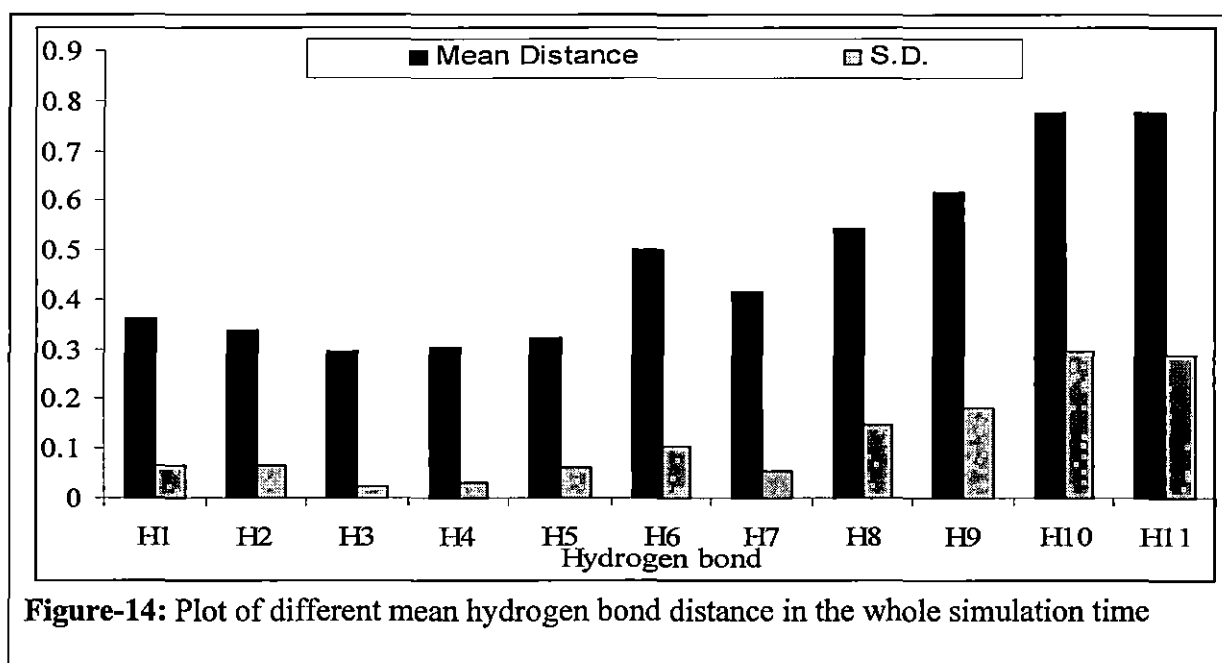


Figure-13: Number of hydrogen bond during the whole simulation time.

A series of hydrogen bonds were observed in the side chains of highly conserved amino acids of 18 residue loop of LR in *C. difficile* toxin A (Ho *et al.*, 2005). We have monitored the distance of several hydrogen bonds during the simulation time. Some of the hydrogen bonds were very strong, but some were broken during the simulation time. It was found that the hydrogen bonds between amide nitrogen of Asn53 and the carbonyl oxygen atoms of Pro46 (H1), amide nitrogen of Asn53 and the carbonyl oxygen atoms of Ile 54 (H2), carbonyl oxygen of Asn53 and the amide nitrogen atoms of Thr49 (H3), carbonyl oxygen of Asn53 and the amide nitrogen atoms of Asp50 (H4) and the side chain of Thr49 and the main chain peptide groups of Pro46 (H5) remain intact during the whole simulation time, whereas the hydrogen bonds between side chain of Glu55 and the main chain peptide groups of Ala47 (H6), the side chains of Gln57 and the main chain peptide groups of Ile 54 (H7 and H8) and hydrogen bonds formed between the side chains of Asn48 and Asn52 (H9), as well as between the side chains of Asp50 and Arg60 (H10 and H11), were broken during the whole simulation time (Fig.14).



4. Discussion

It is evident from Fig.1 that RMSD increased slowly up to 2,000 ps, then there was a small jump at about 2,500 ps and remained at that state with low fluctuation. The initial drift in RMSD may be due to the difference of crystal structure with solution structure. From Fig.2, it is clear that Rg does not show much variation during the simulation time, which indicates that the protein is not much flexible.

The plot of B-factor and RMSF (Fig.3) presents a similar trend across the sequence, although there are differences in the environmental condition of the protein molecule in the x-ray and MD studies. From RMSF, it is evident that the first and last residue fluctuates considerably. Interestingly, pronounced fluctuations are observed along some amino acid stretches (16-23, 45-57, 84-94, and 110-117), which indicate the flexibility of the toxin in that region.

From the snapshots (Fig.4), it is also clear that there was no major change in the protein conformation.

Among the SR units, SR1 and SR3 have low RMSD values. Strand I and turn of SR2 have approximately double RMSD in comparison to the strand I of other SR. Strand II of SR1 has higher RMSD value amongst the strand II of the entire SR. The loop that connects LR and SR2 has the highest RMSD value among different loops (see Table-1).

It is clear from the Fig.5 that SR1 and SR3 have less fluctuation among the SR.

From Fig.9, it is evident that fluctuation is high in the projection on 1st vector. It indicates the analysis of 1st vector will provide more information regarding the collective motion of the protein.

Fig.10 indicates PC1 fluctuates remarkably in comparison to PC2, PC3 and PC4.

From Fig.11, it is clear that the toxin is sampling different conformational space during the simulation. It is also clear from the Fig.11 that the protein traverses one conformational space around the origin, second one at the right side of the origin, and the third one at the left side of the origin which are not much scattered, indicating less conformational freedom of the toxin; this is also revealed by different snapshots of the different structures extracted along the simulation trajectory in different times (Fig.4).

The hydrogen bond network is almost conserved during the simulation time. Extensive study of the residues of binding site was performed in order to understand the binding mechanism. From Table 2, it was found that Ser79 and Gln57 have low RMSD values. The Lys80 residue, which accepts hydrogen from the 6-OH of β -Galactose and donates hydrogen to a water molecule, has a moderate RMSD value indicating the conformation change of this residue is possible during binding process. Asp50 has a moderate RMSD value. Arg60 has much higher flexibility than the other binding residues (Lys80, Ser79, Asp50, and Gln57). The side chain of Gln57 is involved in hydrogen bond with main chain peptide group of Ile54. Arg60 has hydrogen bond with Asp50 which often breaks during

simulation and has high conformational freedom, and this is also supported by its RMSD value.

The binding site is much flexible so the binding may be initiated by those residues which have a very crucial role in inducing binding ligand. The other residues in binding site may be responsible for the stability of ligand toxin complex. These analyses will further help to understand the binding mechanism and overall motional properties of the protein. From the overall study of the dynamics trajectory such as Rg, RMSD of the toxin and PCA, it is clear that *C. difficile* toxin A is not much flexible.

6.5: References

- Aktories K & Barbieri JT (2005) Bacterial Cytotoxins: Targeting Eukaryotic Switches. *Nat. Rev. Microbiol.*, **3**: 397-410.
- Aktories K & Just I (2005) Clostridial Rho-Inhibiting Protein Toxins. *Curr. Top. Microbiol. Immunol.*, **291**:113-145.
- Amadei A, Linssen ABM & Berendsen HJC (1993) Essential Dynamics of Proteins. *Proteins*, **17**:412-425.
- Aronsson B, Mollby R & Nord CE (1981) Occurrence of Toxin Producing *Clostridium difficile* in Antibiotic-Associated Diarrhea in Sweden. *Med. Microbiol. Immunol.(Berlin)*, **170**:27-35.
- Bartlett JG, Moon N, Chang TW, Taylor N & Onderdonk AB (1978) Role of *Clostridium difficile* in Antibiotic-Associated Pseudomembranous Colitis. *Gastroenterology*. **75**:778-782.
- Bartlett JG, Taylor NS, Chang T & Dzink J (1980a) Clinical and Laboratory Observations in *Clostridium difficile* Colitis, *Am. J. Clin. Nutr.*, **33**:2521-2526.
- Bartlett JG, Tedesco FJ, Shull S, Lowe B & Chang T (1980b) Symptomatic Relapse After Oral Vancomycin Therapy of Antibiotic-Associated Pseudomembranous Colitis. *Gastroenterology*, **78**:431-434.
- Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A & Haak JR (1984) Molecular Dynamics With Coupling to an External Bath. *J. Chem. Phys.*, **81**:3684-3690.
- Bothra AK, Roy S, Bhattacharyya B & Mukhopadhyay C (1998) Molecular Dynamics Simulation of Colchicinoids. *Journal of Biomolecular Structure and Dynamics*, **15**:999-1008.

- Burdon DW, George RH, Mogg GA, Arabi Y, Thompson H, Johnson M, Alexander-william J & Keighley MR (1981) Faecal Toxin and Severity of Antibiotic-Associated Pseudomembranous Colitis. *J. Clin. Pathol.*, **34**: 548-551.
- Darden T, York D & Pedersen L (1993) Particle Mesh Ewald—An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.*, **98**:10089-10092.
- Don GJ & Devis AE (1981) The Association Between Antibiotic-Associated Diarrhoea and *C. difficile* Toxin in Children. *Aust. N. Z. J. Med.*, **11**:433-434.
- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H & Pedersen LG (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**:8577–8593
- Frey S M & Wilkins T D (1992) Localization of Two Epitopes Recognized by Monoclonal Antibody PCG-4 on *Clostridium difficile* Toxin A. *Infect. Immun.*, **60**:2488-2492.
- Frisch C, Gerhard R, Aktories K, Hofmann F & Just I (2003) The Complete Receptor-Binding Domain of *Clostridium difficile* Toxin A is Required for Endocytosis. *Biochem. Biophys. Res. Commun.*, **300**:706-711.
- Garcia AE (1992) Large-Amplitude Nonlinear Motions in Proteins. *Phys. Rev. Lett.*, **68**:2696-2699.
- George WL, Sutter VL, Goldstein EJ, Ludwig SL & Finegold SM (1978) Aetiology of Antimicrobial- Agent-Associated Colitis. *Lancet*, **311**:802-803.
- George WL, Rolfe RD, Harding GK, Klein R, Putnam CW & Finegold SM (1982) *Clostridium difficile* and Cytotoxin in Feces of Patients with Antimicrobial Agent-Associated Pseudomembranous Colitis. *Infection*, **10**:205-208.
- Greco A, Ho JGS, Lin S, Palcic MM, Rupnik M & Ng KK (2006) Carbohydrate Recognition by *Clostridium difficile* Toxin A. *Nature Structural & Molecular Biology*, **13**:460-461.

- Hess B, Bekker H, Berendsen HJC & Fraaije JGE (1997) LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**:1463-1472.
- Ho JGS, Greco A, Rupnik M and Ng KK-S (2005) Crystal Structure of Receptor- Binding C-Terminal Repeats from *Clostridium difficile* Toxin A. *PNAS*, **102**:18373-18378.
- Kelly CP & LaMont JT (1998) *Clostridium difficile* Infection. *Annu. Rev. Med.*, **49**:375-390.
- Lindahl E, Hess B & van der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Modeling* **7**:306-317.
- Lyerly DK, Saum KE, MacDonald DK & Wilkins TD (1985) Effects of *Clostridium difficile* Toxins Given Intra-gastrically to Animals. *Infect. Immun.*, **47**:349-352.
- Meyers S, Meyers L, Bottone E, Desmond E & Janowitz HD (1981) Occurrence of *Clostridium difficile* Toxin During the Course of Inflammatory Bowel Disease. *Gastroenterology*, **80**:697-670.
- Roberts AK & Shone CC (2001) Modification of Surface Histidine Residues Abolishes the Cytotoxic Activity of *Clostridium difficile* Toxin A. *Toxicon*, **39**:325-333.
- Sauerborn M, Leukel P & von Eichel-Streiber C (1997) The C-Terminal Ligand- Binding Domain of *Clostridium difficile* Toxin A (TcdA) Abrogates TcdA-Specific Binding to Cells and Prevents Mouse Lethality. *FEMS Microbiol. Lett.*, **155**:45-54.
- Schirmer J & Aktories K (2004) Large Clostridial Cytotoxins: Cellular Biology of Rho/Ras-Glucosylating Toxins. *Biochim. Biophys. Acta.*, **1673**:66-74.
- Tedesco FJ (1981) Antibiotic-Associated Colitis: An Abating Enigma. *J. Clin. Gastroenterol.*, **3**:221-224.
- Teneberg S, Lönnroth I, Torres López JF, Galili U, Halvarsson MO, Angstrom J & Karlsson K (1996) Molecular Mimicry in the Recognition of Glycosphingolipids by Gal3Gal β 4GlcNAc β -Binding *Clostridium difficile* Toxin A, Human Natural Anti-

Galactosyl IgG and the Monoclonal Antibody Gal-13: Characterization of a Binding-Active Human Glycosphingolipid, Non-Identical with the Animal Receptor. *Glycobiology*, **6**: 599-609.

Tucker KD & Wilkins TD (1991) Toxin A of *Clostridium difficile* Binds to the Human Carbohydrate Antigens I, X, and Y. *Infect. Immun.*, **59**:73-78.

Von Eichel-Streiber C, Boquet P, Sauerborn M and Thelestam M (1996) Large Clostridial Cytotoxins—A Family of Glycosyl Transferases Modifying Small GTPBinding Proteins. *Trends Microbiol.*, **4**:375-382.

Voth D E & Ballard JD (2005) *Clostridium difficile* Toxins: Mechanism of Action and Role in Disease. *Clin. Microbiol. Rev.*, **18**:247-263.

Willey SH & Bartlett JG (1979) Cultures for *Clostridium difficile* in Stools Containing a Cytotoxin Neutralized by *Clostridium sordellii* Antitoxin. *J. Clin. Microbiol.*, **10**: 880-884.