

KADIR HAS UNIVERSITY  
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING



EXPLORING DISTINCT CONFORMERS OF  $\beta_2$ -ADRENERGIC RECEPTOR VIA  
COARSE-GRAINED MOLECULAR DYNAMICS SIMULATIONS

SİBEL ÇAKAN

July, 2012

Sibel Çakan

M.S. Thesis

2012

EXPLORING DISTINCT CONFORMERS OF  $\beta_2$ -ADRENERGIC RECEPTOR VIA  
COARSE-GRAINED MOLECULAR DYNAMICS SIMULATIONS

SİBEL ÇAKAN

Submitted to the Graduate School of Science and Engineering  
in partial fulfillment of the requirements for the degree of  
Master of Science in Computational Biology and Bioinformatics

KADIR HAS UNIVERSITY

July, 2012

KADIR HAS UNIVERSITY GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

EXPLORING DISTINCT CONFORMERS OF  $\beta_2$ -ADRENERGIC RECEPTOR VIA  
COARSE-GRAINED MOLECULAR DYNAMICS SIMULATIONS

SİBEL ÇAKAN

APPROVED BY:

Assist. Prof. Dr. Demet Akten Akdoğan  
(Thesis Supervisor)

---

Prof. Dr. Kemal Yelekçi

---

Prof. Dr. Pemra Doruker

---

APPROVAL DATE:

# EXPLORING DISTINCT CONFORMERS OF $\beta_2$ -ADRENERGIC RECEPTOR VIA COARSE-GRAINED MOLECULAR DYNAMICS SIMULATIONS

## Abstract

$\beta_2$  adrenergic receptor ( $\beta_2$ AR) is a G protein-coupled receptor, which belongs to the largest family of membrane proteins and is the target of many drugs.  $\beta_2$ AR is highly flexible and, able to recognize a wide range of ligands through its conformational variations. Although recent crystallographic experiments have revealed active and inactive conformations, they are not sufficient for deciphering the whole receptor's dynamics. Molecular dynamics (MD) simulation is an alternative and efficient method to understand the protein dynamics. However, traditional all-atom simulations do not reach the millisecond time scales at which many biological processes occur. Thus, coarse-grained (CG) modeling is used to reduce the number of degrees of freedom. The system was composed of  $\beta_2$ AR embedded into a palmitoyl-oleoyl-phosphatidylcholine (POPC) membrane bilayer with surrounding water. Main purpose of using a CG model is to explore a wider conformational space that would not be reachable via all-atom models. The local fluctuations were in good agreement with all-atom simulations. Four snapshots were selected and reverse-mapped to all-atom representations. Each was later subjected to 100 ns MD simulation for equilibration. RMSD clustering yielded distinct receptor conformers that are both energetically and structurally acceptable. PCA analysis of CG-MD simulations showed that the first five principle modes explained only 50% of the overall dynamics compared to 85% in all-atom simulations. Maximum overlap value between eigenvectors of CG and all-atom was determined as 0.46. Normalized orientational cross-correlations between residue fluctuations revealed weaker correlations in CG simulations compared to all-atom

# $\beta_2$ -ADRENERJİK RESEPTÖRÜN KABA TANELİ MOLEKÜLER DİNAMİK SİMÜLASYONU İLE FARKLI KONFORMASYONLARININ ARAŞTIRILMASI

## Özet

$\beta_2$ AR, G protein bağlantılı reseptör ve birçok ilaç için hedef moleküldür. Reseptörün son derece esnek olan yapısı bir çok ligant molekülünü tanıma özelliği sağlar. Son yıllarda yapılan kristalografik çalışmalar reseptörün aktif ve inaktif yapısını ortaya çıkarmasına rağmen bu çalışmalar reseptörün tüm dinamiğini çözmek için yeterli değildir. Moleküler dinamik (MD) metodu reseptörün tüm dinamiğini anlamak için alternatif ve verimli bir yöntemdir. Ancak geleneksel atomistik simülasyonlar birçok biyolojik olayın gerçekleştiği zaman aralığı olan milisaniye seviyelerine ulaşamaz. Bu nedenle, bu çalışmada serbestlik derecesini azaltan kaba taneli modelleme kullanıldı. Sistem POPC membran tabakası içine gömülü  $\beta_2$ AR ve sulardan oluşturuldu. CG model kullanılmasının asıl amacı atomistik modellerde mümkün olmayan daha geniş yapısal alanı ortaya çıkarmaktır. Reseptörün bölgesel hareketleri atomistik simülasyonlarla uyum içindedir. CG simülasyondan dört görüntü seçilmiş ve geri eşleme yöntemi ile atomistik modele çevrilmiştir. Daha sonra herbiri 100 ns uzunluğunda bir MD simülasyonuna tabi tutulmuştur. Enerjik ve yapısal olarak farklı reseptör yapıları ortaya çıkmıştır. CG MD simülasyonunun PCA analizi, ilk beş birincil bileşenin tüm dinamiğin %50 sini açıklarken, atomistik simülasyonların %85 ini açıkladığını göstermiştir. CG ve atomistik öz-vektörlerin maksimum örtüşme değeri 0.46 dir. CG modelde atomistik modele göre korelasyonlar daha zayıftır.

*To my family*

## **Acknowledgements**

I wish to express my deepest gratitude, first and foremost, to my dissertation supervisor, Assist. Prof. Demet Akten Akdođan, for her valuable advice, sincere comments and guidance of this study, and her encouragement that helped me to complete and writing of this thesis.

I would like to thank my thesis committee members: Prof. Pemra Doruker and Prof. Kemal Yelekçi for their participation and valuable comments.

I am grateful to my friends, İlker Ümit Yılmaz, Ayça Gençaydın, Duygu Demirci, Burcu Ohri, for their affection, support and motivation. My special thanks go to Seda Demirci, Serkan Altuntas, Ayça Korođlu, Bora Buyukturk, Cagla Midik, whose friendship I deeply value. I would also like to thank Mark Wyers for his support and friendship.

The last but not least, I am deeply indebted to my parents, Semiha and Haldun Çakan, my sister and brothers, Handan-Selim Pilten and Haluk Çakan for their love, patience, and encouragement.

This study was supported by TÜBİTAK through project 109M281.



## Table of Contents

<b>Abstract</b> .....	<b>i</b>
<b>Özet</b> .....	<b>ii</b>
<b>Dedication</b> .....	Error! Bookmark not defined.
<b>Acknowledgements</b> .....	<b>iv</b>
<b>Table of Contents</b> .....	<b>v</b>
<b>List of Tables</b> .....	<b>viii</b>
<b>List of Figures</b> .....	<b>ix</b>
<b>List of Abbreviations</b> .....	<b>xiv</b>
<b>Introduction</b> .....	<b>1</b>
<b>Materials and Methods</b> .....	<b>4</b>
2.1.Molecular Dynamics Simulations .....	4
2.2. Coarse-Grained Molecular Dynamics Force Field.....	4
2.3. Reverse Mapping.....	8
2.4. Trajectory Analysis .....	8
2.4.1. Root Mean Square Deviation (RMSD) .....	8
2.4.2. Mean Square Fluctuations (RMSF).....	9
2.4.3. Clustering .....	9
2.4.4. Principal Component Analysis (PCA).....	10

2.4.5. Cross Correlations .....	10
<b>Results and Discussions.....</b>	<b>12</b>
3.1. Construction of the Residue Based Coarse-Grained Model.....	12
3.2. Coarse-Grained Molecular Dynamics Simulation Details .....	16
3.3. Clustering of CG Trajectory .....	17
3.4. Reverse Mapping of Coarse-Grained Models to Their Atomistic Representation .....	20
3.4.1. System Preparation for Reverse-Mapped Molecular Dynamics (MD) Simulation.....	20
3.4.2. Molecular Dynamics Simulation Details of Reverse-Mapped Structures.....	21
3.4.2.1. Melting of Lipid Tails .....	22
3.4.2.2. Minimization and Equilibration with Protein Constrained .....	22
3.4.2.3. Equilibration with Protein Released .....	22
3.4.2.4. Production Runs .....	23
3.5. Root Mean Square Deviation (RMSD) of CG Simulations .....	24
3.6. Root Mean Square Fluctuations of CG Simulations (RMSF).....	27
3.7. RMSD of Reverse-Mapped (RM) Simulations .....	29
3.8. RMSF of Reverse-Mapped Simulations.....	33
3.9. Energy Profiles of RM and FA Simulations.....	33
3.10. Clustering of All FA and RM Trajectories.....	37
3.11. Structural Agreement of 2RH1 and RM Simulations.....	41
3.12. RMSD of Helices of RM Structures.....	45
3.13. Binding Pocket of $\beta_2$ AR.....	47
3.14. Structural Agreement of CG Model and 2RH1 .....	48

3.15. RMSD of Helices of CG Simulations .....	55
3.16. RMSD of Binding-site Regions.....	57
3.17. Distance Between Residues D113 and S203-S207 .....	57
3.18. Principal Component Analysis (PCA).....	62
3.19. Overlap Calculations .....	67
3.20. Cross Correlations .....	71
3.21. Applicability of Martini Force Field on Proteins .....	76
3.22. Comparison of Simulation Costs Between Fully-Atomistic and Coarse-Grained Models .....	77
<b>Conclusions .....</b>	<b>78</b>
<b>References .....</b>	<b>87</b>

## List of Tables

Table 2.1. Interaction matrix (Adapted from Marrink et.al.[17]) [23] .....	7
Table 3.1. System size details of fully atomistic (FA) and coarse-grained (CG) models.....	16
Table 3.2. Simulation system size and run length. ....	16
Table 3.3. Clustering information for four different regions of the receptor. ....	18
Table 3.4. The dimensions of the reverse-mapped protein and surrounding system. All values are given in Angstrom.....	21
Table 3.5 Total numbers of atoms in reverse-mapped structures.....	21
Table 3.6. Terminology used for various RMSD calculations.....	24
Table 3.7. RMSD values of alignments based on transmembrane and helices. RMSD-1 shows the alignment according to the core region, RMSD-2 shows the helices alignments. RMSD-2 values higher than 3.0 Å are typed in bold characters. ...	55
Table 3.8. Percentage of the total motion explained by first five principal components of each simulations.....	64
Table 3.9. First three overlap values of all simulations.....	70

## List of Figures

Figure 2.1. All atom model (a) coarse-grained model (b) and (c) both models superposed for Aspartic acid residue.....	8
Figure 3.1. Fully atomistic (a) and Coarse-grained (b) representation of the $\beta_2$ AR and POPC lipids. ....	13
Figure 3.2. Fully atomistic (a) and Coarse-grained (b) representation of whole system of $\beta_2$ AR (embedded in a POPC membrane, solvated and ionized).....	15
Figure 3.3. Binding-site (a) and transmembrane (b) regions of the protein. ....	18
Figure 3.4. Cluster profile of CG MD trajectory according to different regions of $\beta_2$ AR (a) core region (RMSD=4Å), (b) ICL3 region (RMSD=3Å), (c) transmembrane region(RMSD=3Å), (d) binding-site region (RMSD=2.5).....	19
Figure 3.5. Cell membrane representation before (a) and after (b) the melting of the lipid tails. ....	22
Figure 3.6. Cell membrane representation before (a) and after (b) the equilibration of the whole system. ....	23
Figure 3.7. RMSD <i>All Fit All</i> profile of CG simulations. ....	25
Figure 3.8. RMSD <i>Core Fit Core</i> profile of CG simulations.....	25
Figure 3.9. RMSD <i>ICL3 Fit Core</i> profile of CG simulation. ....	26
Figure 3.10. The snapshot from 2 $\mu$ s is colored in violet while the yellow colored structure is from 4 $\mu$ s.....	26

Figure 3.11. RMSF profiles of CG MD and FA MD simulations. (First 30 ns, 97 ns, 200 ns and 48 ns were excluded from FA's trajectory, 290 K, 310 K, and 323 K CG's trajectories, respectively) .....	28
Figure 3.12. Cartoon representation of $\beta_2$ AR with loop regions.....	28
Figure 3.13. RMSF profiles of CG MD, long and short FA MD simulations. (First 30 ns were excluded from FA long trajectory, 2 ns from FA short trajectories.) ...	29
Figure 3.14. RMSD <i>All Fit All</i> profile of FA (a) and RM (b) simulations.....	30
Figure 3.15. RMSD <i>Core Fit Core</i> profiles of FA (a) and RM (b) simulations. ....	31
Figure 3.16. RMSD <i>ICL3 Fit Core</i> profiles of FA (a) and RM (b) simulations. ....	32
Figure 3.17. RMSF profiles of RM simulations and FAS1 .....	33
Figure 3.18. Total energy profiles of FA short (a) and RM simulations (b). ....	34
Figure 3.19. Kinetic energy profiles of FA short (a) and RM simulations (b).....	34
Figure 3.20. Potential energy profiles of FA short (a) and RM simulations (b) .....	35
Figure 3.21. Electrostatic energy profiles of FA short (a) and RM simulations (b)...	35
Figure 3.22. The van der Waals (vdW) energy profiles of FA short (a) and RM simulations (b).....	36
Figure 3.23. Bond stretching energy profiles of FA short (a) and RM simulations (b). .....	36
Figure 3.24. Angle bending energy profiles of FA short (a) and RM simulations (b) .....	37
Figure 3.25. Dihedral energy profiles of FA short (a) and RM simulations (b).....	37
Figure 3.26. Improper energy profiles of FA short (a) and RM simulations (b).....	37
Figure 3.27. Cluster profile of all simulations (FA and RM). (a) based on core RMSD = 5 Å, (b) transmembrane RMSD = 3 Å, (c) bindingsite RMSD = 2.8 Å, (d) ICL3 RMSD = 5.9 Å. ....	40

Figure 3.28. Alignments of RM4's and 2RH1 according to the core region (a) and the helices (b). .....	41
Figure 3.29. The structural variations of RM1 (a), RM2 (b), RM3 (c), RM4 (d).Color bar on the right is given as: 0 = T (turn), 1 = C (coil), 2 = B (isolated bridge), 3 = E (beta sheet), 4 = H (alpha helix), 5 = G (3-10 helix), 6 = I (pi helix). .....	44
Figure 3.30. The structural variations of FAL simulation. ....	44
Figure 3.31. (cont'd) RMSD profiles of helices during the RM1 (a), RM2 (b), RM3 (c), RM4 (d) simulations. ....	47
Figure 3.32. Five representative structures taken from five distinct clusters aligned to 2RH1 (shown in gray). RMSD of the binding site region is FAL (a), RM1 (b), RM1 (c), RM2 (d), RM3 (e). ....	48
Figure 3.33. Alignments of 290 K CG trajectory's first snapshot and 2RH1 according to the transmembrane region (a) and the helices (b). ....	49
Figure 3.34. Alignments of 290 K CG trajectory's last snapshot and 2RH1 according to the transmembrane region (a) and the helices (b). ....	50
Figure 3.35. Alignments of 310 K CG trajectory's first snapshot and 2RH1 according to the transmembrane region (a) and the helices (b). ....	51
Figure 3.36. Alignments of 310 K CG trajectory's last snapshot and 2RH1 according to the transmembrane region (a) and the helices (b). ....	52
Figure 3.37. Alignments of 323 K CG trajectory's first snapshot and 2RH1 according to the transmembrane region (a) and the helices (b). ....	53
Figure 3.38. Alignments of 323 K CG trajectory's last snapshot and 2RH1 according to the transmembrane region (a) and the helices (b). ....	54
Figure 3.39. RMSD profiles of helices for 290 K (a), 310 K (b), 323 K (c) CG simulations. ....	56

Figure 3.40. Binding-site RMSD of all FA and RM simulations.....	57
Figure 3.41.(cont'd) FAL (a) and CG (b) simulation's variation of distances with time between residues Asp113-Ser203 and Asp113-Ser207.....	59
Figure 3.42. FA simulation's variation of distances with time for Asp113-Ser203 (a) and Asp113-Ser207 (b).....	60
Figure 3.43. (cont'd) RM simulation's variation of distances with time for Asp113- Ser203 (a) and Asp113-Ser207 (b).....	62
Figure 3.44. The explanation percentages of the protein motion of the first 20 principal modes for all simulations. (a) individual, (b) cumulative. ....	63
Figure 3.45.(cont'd) Collective motions on the first and second principal components of FAL (a), FAS1 (b), CG (c) and RM4 (d) simulations. Initial conformation is represented in blue color, last conformation is in red. ....	66
Figure 3.46. (cont'd) Overlap matrix of first 20 modes between CG and FAL (a), CG and FAS2 (b), CG and RM2 (c) simulations.....	69
Figure 3.47. Cross-correlations between residue fluctuations (a) first ten modes of CG simulation, (b) first mode of FAL simulation.....	72
Figure 3.48. Cross-correlations between residue fluctuations (a) first two modes of FAS1 simulation, (b) three modes of FAS2 simulation, (c) two modes of FAS3 simulation. ....	73
Figure 3.49.(cont'd) Cross-correlations between residue fluctuations (a) first two modes of RM1 simulation, (b) two modes of RM2 simulation, (c) first three modes of RM3 simulation, (d) two modes of RM4 simulation.....	75
Figure 3.50. RMSD profiles of T4-lysozyme.....	76



Figure 3.51. Alignments of T4-lysozyme's first and last snapshots according to the all structure (a) and helices (b). First snapshot is shown in grey, last snapshot is in red color.....77

## List of Abbreviations

CG	Coarse-Grain
ECL2	Extracellular Loop Two
ECL3	Extracellular Loop Three
ICL2	Intracellular Loop Two
ICL3	Intracellular Loop Three
MD	Molecular Dynamics
PDB	Protein Data Bank
RM	Reverse Map
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
$\beta_2$ AR	Human Beta-2 Adrenergic Receptor
GPCR	G-Protein Coupled Receptor
Asp	Aspartic Acid
POPC	palmitoyl-oleoyl-phosphatidylcholine

## **Chapter 1**

### **Introduction**

Twenty five percent of the eukaryotic genomes encode the membrane proteins that have significant roles such as transporting, signaling and cell-cell interactions in the biological cells [1,2]. They also constitute the largest class of drug targets, approximately 50% of all drugs in the market [3]. In spite of their physiological and pharmaceutical importance, very few crystallographic structures are reported [4].

The interactions with lipids and also the bilayer properties such as hydrophobicity or lipid composition affect the protein function in membrane proteins [5]. Crystallization has often been performed as membrane protein-detergent complexes, however in most cases only a few tightly bound lipid molecules remain. Thus, explicit information about where the protein is located in the lipid bilayer is obtained from the crystal structures [6]. Still it is difficult to understand the details of protein-membrane interactions using experimental methods. Thus, embedded proteins in a lipid bilayer (membrane proteins) are ideal systems for computer simulations [7].

With current computational power, it is not possible to sample all intermediates of adrenergic receptors along the activation pathway (that are in the millisecond time scale) via traditional MD simulations. Just early rearrangements such as ligand positioning and initializing of protein activation can be observed in an atomistic level simulations [8-11]. Several recent studies indicate that atomistic detail approaches is not enough to obtain a full dynamics of protein-membrane interactions, which require at least microsecond time scale. On the other hand, coarse-grained (CG) models provide a favorable approach for increasing the time scale from nanoseconds

to microseconds by grouping several atoms into one or two particles, while interaction potentials are modeled similar to those describing the original atoms. CG models significantly reduce the number of degrees of freedom and thus increase the time scale of biomolecular simulations [12,13]. In a CG model, collective motions of the protein in larger time scales can be observed due to removal of high-frequency motions. CG models have been applied to different biomolecules such as lipids, membranes, proteins and DNAs.

For proteins, there exist several approaches of CG modeling which have different agreement between accuracy and transferability, with degrees of independence from the reference structure. The earliest approach of the coarse-graining for proteins is the Elastic Network Models and Go-like models whose bias towards a reference structure makes them only weakly transferable to general dynamics studies [14]. In 1970s Levitt built a transferable coarse-grained model with a knowledge-based parameterization that inspired many successive researches [15]. The coarsest approach is one-bead model that was evolved from Go-like models, and includes more sophisticated potentials, but still was not enough. To improve the specificity of the local interactions, one more bead was added on the centroid of the side chain (two beads model). Statistical analyses of the experimental structures were used for developing the force field of the two-bead models [16]. For lipids, Marrink has improved a CG model where four heavy atoms are represented on average by one CG particle [17, 18]. It is an off-lattice model that was improved first by Smit et al [19].

Marrink *et al.* developed and employed a CG model to examine the effects of physicochemical features of the lipid bilayer on self-assembly of visual rhodopsin molecules that is a member of GPCR. They used GROMACS software package and obtain that interaction of membrane bilayer with the receptor is near trans-membrane helices of 2,4 and 7. They concluded that future application of the CG MD method may contribute to a better understanding of the role of lipid diversity and protein structure in lipid-mediated protein-protein interactions [20]. In another work of Marrink and coworkers, a CG MD simulation was performed for Kv1.2 (a voltage

gated ion channel). The study revealed a possible gating mechanism with a coarse-grained model [21]. Another CG MD method was developed by Scott *et al.* for self assembly of lipid bilayers around proteins. They predicted precisely position of the protein in the bilayer with a range of different size and architectures of membrane proteins [22].

In this thesis, a CG model for protein-lipid-water model developed by Shih *et al.* was used for studying the dynamics of beta-2 adrenergic receptor embedded in a lipid bilayer [23]. Shih's model originates from Marrink's CG lipid-water model, which was extended to include proteins as well. It was further implemented into NAMD Molecular Dynamics software tool [24].

In this study, the crystal structure of  $\beta_2$ AR (PDB id: 2rh1) was first converted into a coarse-grained model using CG-Builder plug-in tool of VMD visualization software [25]. Then, a 6  $\mu$ s MD simulation was performed using NAMD v2.7. In Shih's model, the Lennard-Jones (LJ) potential was used for non-bonded interactions and the Coulombic potential for charged groups. Shih's approach was successfully applied to the system of self-assembly nanodiscs. Although  $\beta_2$ AR's interaction with lipids ( $\beta_2$ AR is an embedded membrane protein) is completely different from nanodiscs, MD results were comparable with experimental and also fully-atomistic MD simulations' results. There are two types of coarse-grained modeling tool in VMD CG-Builder, one is residue-based (RBCG) and the other, shape-based coarse-graining (SBCG) [25]. In this thesis residue-based modeling where each residue has two interaction sites, one on backbone and the other on side chain was used. The main goal of this thesis is to obtain distinct conformers of  $\beta_2$ AR by exploring a wider conformational space through coarse-grained molecular dynamics (MD) simulation.

## Chapter 2

### Materials and Methods

#### 2.1. Molecular Dynamics Simulations

Molecular dynamics (MD) simulation is a powerful computational tool to compute macroscopic behavior of molecular systems from microscopic interactions. MD simply solves Newton's equations of motion (second law) for molecular systems which results in trajectories that specifies how the positions and velocities of the atoms in the system vary with time [26]. The equation of motion applied to each atom in the system is:

$$\frac{d^2x_i}{dt^2} = \frac{Fx_i}{m_i} \quad (1)$$

This equation describes the motion of a particle of mass  $m_i$ , along one coordinate ( $x_i$ ) with  $Fx_i$  being the force on the particle in that direction [27].

#### 2.2. Coarse-Grained Molecular Dynamics Force Field

Force field refers to combination of a mathematical formula and associated parameters that are used to describe the energy of the system as function of its atomic coordinate [28]. The force field used in this thesis was developed first by Shih and its coworkers. It originates from the lipid-water model of Marrink *et.al.* It was extended to treat proteins and implemented into NAMD software tool by Shih and coworkers. Marrink's model uses a four-to-one mapping. Four heavy atoms

(non-hydrogen) are represented by a single interaction center, called “beads”. The Marrink CG model defines only four main types of beads based on properties such as hydrophobicity, hydrogen bonding, or charge and also bead classes are used to determine the strength of non-bonded interactions between any two beads in the system. Existing parameters for lipid and water coarse-graining are extended to describe the proteins as well. The CG beads are considered as point masses and Newton’s second law describes their dynamics. The interaction potentials between CG beads is given by:

$$V = \sum_j [V_{bond}^j + V_{angle}^j + V_{dihedral}^j] + V_{non-bonded} \quad (2)$$

where the index  $j$  represents one of the four components of the system; lipid, protein, water, and ions. While the term  $V_{non-bonded}$  in Eq.(2) describes the non-bonded interactions between CG beads, other terms in brackets describe the covalently bonded CG beads.  $V_{bond}^j$  in Eq. (2) accounts the bond lengths between CG beads,  $V_{angle}^j$  the forces maintaining certain angles between sets of three bonded CG beads,  $V_{dihedral}^j$  describes the potential of dihedral angles for quadruples of bonded CG beads. Only non-bonded interactions are accounted for water and ions that do not have any bonds in this model. The term  $V_{bond}^j$  in Eq. (2) describing the bonds for lipids and proteins, is given by:

$$V_{bond}^j = \sum_i \frac{1}{2} K_i^j (R_i^j - L_i^j)^2 \quad (3)$$

Where  $j$  represents protein or lipid and the summation is over all bonds  $i$ ;  $R_i$  is the distances between the bonded beads for bond  $i$ ,  $K_i$  is the force constant, and  $L_i$  is the equilibrium bond length. The term  $V_{angle}^j$  in Eq. (2), describing harmonic angle term for lipids and proteins, is given by:

$$V_{angle} = \sum_k M_k (\theta_k - \Theta_k)^2 \quad (4)$$

Where  $\theta_k$  is the angle,  $M_k$  is the force constants, and  $\Theta_k$  is the equilibrium angle. The term  $V_{dihedral}^j$  in Eq. (2), describing the dihedral potentials only for any four bonded beads in the quadruple,  $l$ , on protein backbone, is given by:

$$V_{dihedral} = \sum_l \Phi_l (1 + \cos(n\chi_l - \delta_l)) \quad (5)$$

For dihedral potential  $\Phi_l$  is the force constant,  $n$  is the multiplicity, and  $\delta_l$  is the phase shift;  $\chi_l$  is the angle between the plane formed by the first three beads in the quadruple  $l$  and the plane formed by the last three. The non-bonded interaction potentials between CG beads, which consists of Lennard-Jones (LJ) and the Coulomb term, is given by:

$$V_{non-bonded} = \sum_{m,n} 4\epsilon_{mn} \left[ \left( \frac{\sigma_{mn}}{r_{mn}} \right)^{12} - \left( \frac{\sigma_{mn}}{r_{mn}} \right)^6 \right] + \sum_{m,n} \frac{q_m q_n}{4\pi\epsilon\epsilon_0 r_{mn}} \quad (6)$$

Here,  $r_{mn}$  is the distance between beads  $m$  and  $n$  (all components of the system including: lipids, proteins, water and ions),  $\epsilon_{mn}$  and  $\sigma_{mn}$  are the van der Waals (vdW) parameters for the interaction between beads  $m$  and  $n$ ,  $q_m$  is the charge of the  $m$ th bead, and the sum over  $m$  and  $n$  runs over all pairs of CG particles in the system.  $\epsilon_0$  is the vacuum dielectric permittivity and a relative dielectric constant  $\epsilon$  is set to 20 everywhere. The charges  $q_m$  is set to 0.7 times the total charge of the atomic group represented by  $m^{\text{th}}$  bead. The charge scaling of 0.7 and  $\epsilon = 20$  are used to mimic the screening and polarization effects due to the electrostatic interactions at the atomic level [17]. Each CG bead is grouped depending on the properties of the group of atoms represented (hydrophobic-hydrophilic, charged-uncharged, and hydrogen bonding) and they are: polar (P), non-polar (N), apolar (C), and charged (Q). Non-polar and charged groups are further broken down into normal (0), hydrogen-bond donor (d), hydrogen-bond acceptor (a), and donor-acceptor (da) groups. The force constant,  $\epsilon_{mn}$ , of vdW interactions between beads  $m$  and  $n$  is assigned to one of five levels (*I to V*) depending on the class the bead belongs to, as tabulated in Table 2.1. The values of the force constants for each of the levels are  $\epsilon_{mn} = 5$  kJ/mol for level *I*,  $\epsilon_{mn} = 4.2$  kJ/mol for *II*,  $\epsilon_{mn} = 3.4$  kJ/mol for *III*,  $\epsilon_{mn} = 2.6$  kJ/mol for *IV*, and  $\epsilon_{mn} = 1.8$  kJ/mol for *V*. The vdW radius is  $\sigma_{mn} = 4.7$  Å for any pair of CG beads.



Table 2.1. Interaction matrix (Adapted from Marrink et.al.[17]) [23]

		N					Q				
P		0	d	a	da	C	0	d	a	da	
P		I	IV	III	III	II	V	I	I	I	I
	0	IV	III	III	III	III	III	III	III	III	III
N	d	III	III	II	II	II	IV	III	III	II	II
	a	III	III	II	II	II	IV	III	II	III	II
	da	II	III	II	II	I	V	III	II	II	I
C		V	III	IV	IV	V	III	V	V	V	V
	0	I	III	III	III	III	V	III	III	III	II
Q	d	I	III	III	II	II	V	III	III	II	I
	a	I	III	II	III	II	V	III	II	III	I
	da	I	III	II	II	I	V	II	I	I	I

The model of lipid, water and ions are the same as the Marrink's model. Lipid molecules are represented by twelve CG beads: one for the choline group, one for the phosphate group, two for each of the glycerol groups and eight to represent the two hydrocarbon tails [27]. A single polar (class P) CG bead represents four water molecules. Each ion, with its first hydration shell (six water molecules), is represented by one charged (class Q) CG bead. As in the Marrink's lipid-water model, the mass of any CG bead is equal to 72 amu for lipids, water and ions.

For coarse-graining of proteins, an amino acid residue is mapped onto two CG beads. Backbone beads belong to Nda class and side-chain beads belong to a variable class. There is an exception for Glycine residue that is represented by a single backbone CG bead. Bead's placements can be seen clearly in Figure 2.1 where Aspartic acid (ASP) residue of the protein is represented by and all-atom and coarse-grained models.

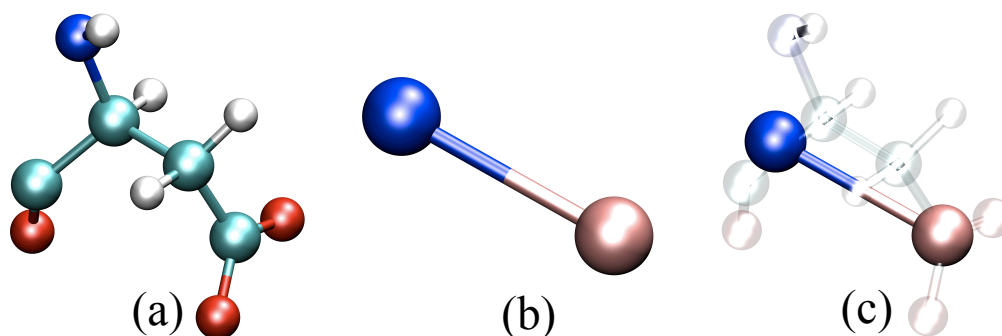


Figure 2.1. All atom model (a) coarse-grained model (b) and (c) both models superposed for Aspartic acid residue

### 2.3. Reverse Mapping

Reverse mapping is a method that reintroduces atomistic details into a CG structure, which lack some important structural details [29]. Reverse mapping process simply recovers the lost properties of the structure in atomistic detail. In reverse mapping process the beads are replaced with the atom groups they represent. The bead position is used as the center of mass, with random orientation. For preserving the topology, the atom groups are reconnected. Then an optimization is carried out geometrically. Because of the initially high stresses in the system, restraints are used to keep the centers-of-mass near the original bead positions. In order to get the energy to a reasonable minimum, minimizing and annealing is used to further equilibrate the system. In this thesis, the coarse-grained model is reverse-mapped back to a fully-atomistic model using CG Tools Plugin of VMD.

### 2.4. Trajectory Analysis

#### 2.4.1. Root Mean Square Deviation (RMSD)

The degree of similarity of two proteins' three-dimensional structures is usually measured with the root-mean-square distance between equivalent atom pairs. RMSD is calculated by:

$$rmsd = \sqrt{\frac{\sum_i d_i^2}{n}} \quad (7)$$

Here  $d$  is the distance between corresponding atoms in two optimally superposed structures and  $n$  is the total number of such pairs. The rmsd value increases as the two structures become structurally different and 0 for identical structures [30].

#### 2.4.2. Mean Square Fluctuations (RMSF)

The mean square fluctuation (MSF) is a measure of the deviation between the position of particle  $i$  from the average structure. MSF is defined as:

$$MSF = \left\langle \left( R_i(t) - \langle R_i \rangle \right)^T \left( R_i(t) - \langle R_i \rangle \right) \right\rangle \quad (8)$$

Where  $R_i$  is the vector of time average of Cartesian coordinates of the C $\alpha$  atom of the  $i^{th}$  residue.  $R_i(t)$  is the vector of Cartesian coordinates of the C $\alpha$  atom of the same residue at time  $t$ . An alternative measure is given by its square root of MSF as RMSF

#### 2.4.3. Clustering

The main purpose of using clustering method is to classify the different conformations obtained from molecular dynamics simulations. Clustering reduces the conformational space, thus similar states of a system can be collected in different clusters. In this study, kclust (k-means clustering) module of Multiscale Modeling Tools of Structural Biology (MMTSB) Tool Set is used for this purpose [31].

In k-means clustering, the data is separated into  $k$  clusters. Randomly selected frames are initially assigned as centroids for each cluster. Each frame is then assigned to its closest cluster center. Each cluster center is updated to be the mean of its constituent frames. Then the algorithm converges when there is no further change in assignment of frames to clusters [32].

#### 2.4.4. Principal Component Analysis (PCA)

Principal component analysis (PCA) is a statistical method that reduces the size of data for revealing the components, which mostly explains the variability of the data. The variability in an MD simulation is the expression of the deviation  $x$ ,  $y$  and  $z$  coordinates of the each atom of the protein from the average values. The most important conclusion from this analysis is the most dominant motion of the protein throughout the simulation.

PC calculation includes two basic steps. First the covariance matrix ( $\mathbf{C}$ ) of the positional deviations is calculated, and then this matrix is diagonalized [33,34]. The 3N dimensional covariance matrix is calculated based on ensemble of protein structures, and the elements of  $\mathbf{C}$  are defined as:

$$C^{(ij)} = \begin{bmatrix} \langle \Delta x_i \Delta x_j \rangle & \langle \Delta x_i \Delta y_j \rangle & \langle \Delta x_i \Delta z_j \rangle \\ \langle \Delta y_i \Delta x_j \rangle & \langle \Delta y_i \Delta y_j \rangle & \langle \Delta y_i \Delta z_j \rangle \\ \langle \Delta z_i \Delta x_j \rangle & \langle \Delta z_i \Delta y_j \rangle & \langle \Delta z_i \Delta z_j \rangle \end{bmatrix} \quad (9)$$

where  $x_i$  and  $x_j$  are atomic coordinates.  $\langle x_i \rangle$  and  $\langle x_j \rangle$  are the ensemble averages. The diagonalization of the symmetric matrix  $\mathbf{C}$  is equivalent to solving the eigenvalue problem:

$$A^T C A = \lambda \quad (10)$$

Here  $\mathbf{A}$  represents the eigenvectors and  $\lambda$  the associated eigenvalues [35]. PCA calculations were performed via ProDy Python package [36].

#### 2.4.5. Cross Correlations

The normalized orientational cross-correlation  $C(i, j)$  between residue fluctuations defined as:

$$C(i,j) = \frac{[\langle \Delta R_i \Delta R_j \rangle]}{[\langle \Delta R_i \Delta R_i \rangle \langle \Delta R_j \Delta R_j \rangle]^{1/2}} \quad (11)$$

The cross-correlations vary in the range [-1,1] with the lower and upper limits indicating that fully anti-correlated and correlated fluctuations in terms of orientation, respectively.  $C(i,j) = 0$  gives uncorrelated fluctuations in terms of orientation.

## Chapter 3

### Results and Discussions

#### 3.1. Construction of the Residue Based Coarse-Grained Model

The X-ray crystallographic structure of human  $\beta_2$ AR with 2.40 Å resolution (PDB entry: 2RH1) is used as the starting conformation [37]. 2RH1 includes a protein (T4 lysozyme) in place of intracellular loop III (ICL3) region. The insertion of T4L after removal of ICL3 has helped the crystallization process via decreasing the protein's high mobility and increasing the protein's total polar surface area. To restore the receptor in its native state, T4L was removed and the ICL3 was estimated via MODVEB homology modeling server in a previous study [39]. Prior to Molecular Dynamics simulations, the receptor was embedded in a phosphatidylglycerol membrane (POPC) and solvated using TIP3 water model. A total of 800 ns (long) and three independent 100 ns (short) MD simulations were performed using NAMD v2.7 software tool [31]. In this thesis, the system equilibrated in a previous study was used as an initial structure. The protein and lipids were extracted and converted into a residue based coarse-grained model using the CG Tools Plugin of VMD [25]. All-atom and coarse-grained representation of the protein and lipid system are shown in Figure 3.1. The decreased number of atoms can be seen clearly. The system size was reduced significantly: 68,001 atoms in fully atomistic (FA) model are represented by 6,868 CG beads in the coarse grained (CG) model only. CG Tools Plugin produces two output files: the first one is the PDB file, which includes coarse-grained beads instead of all atom molecules. The second file is the "Reverse Coarse Graining File" that is necessary for reverse mapping procedure that converts the coarse-grained system back to all-atom representation. First output file was used for primary studies.

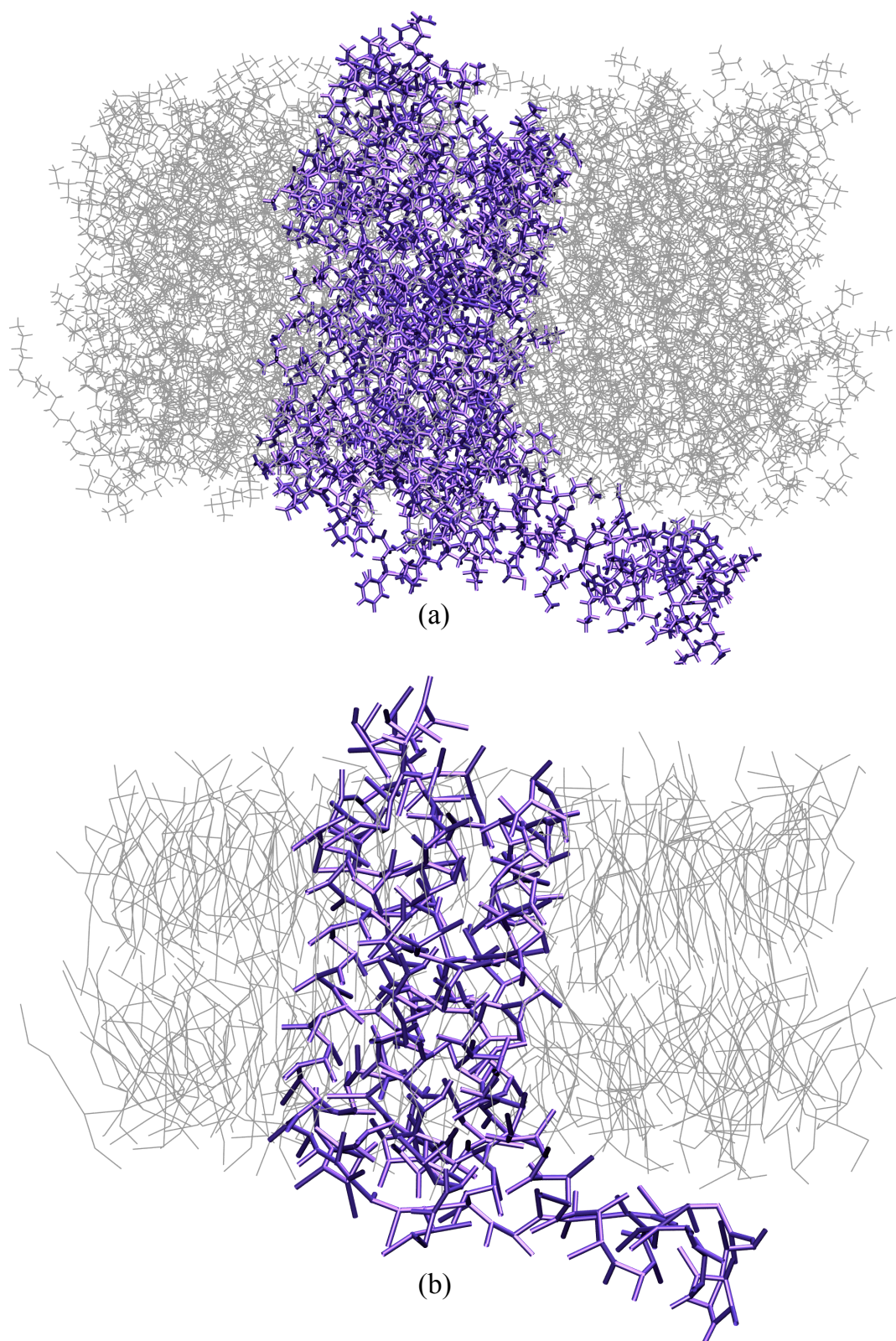
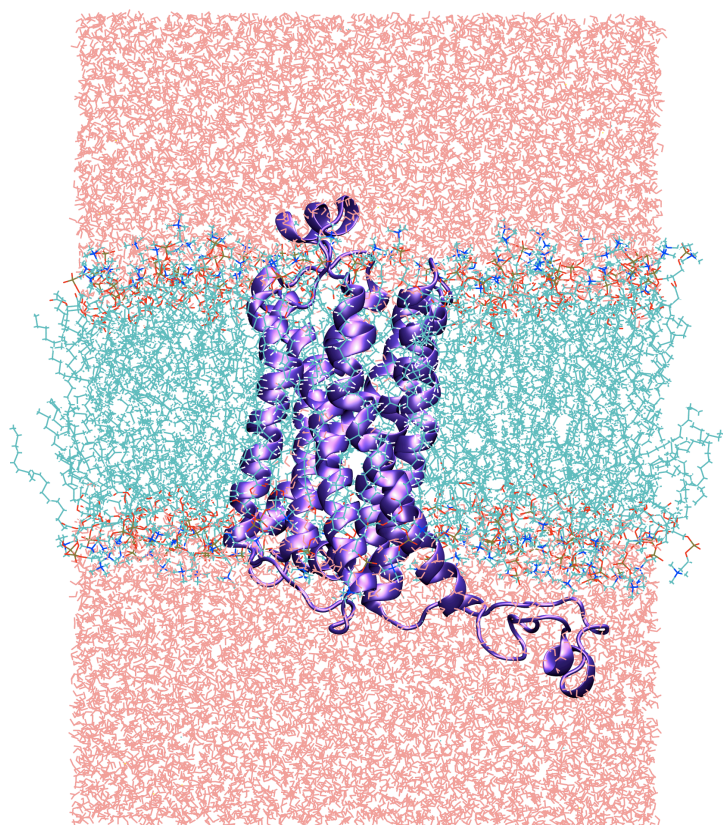


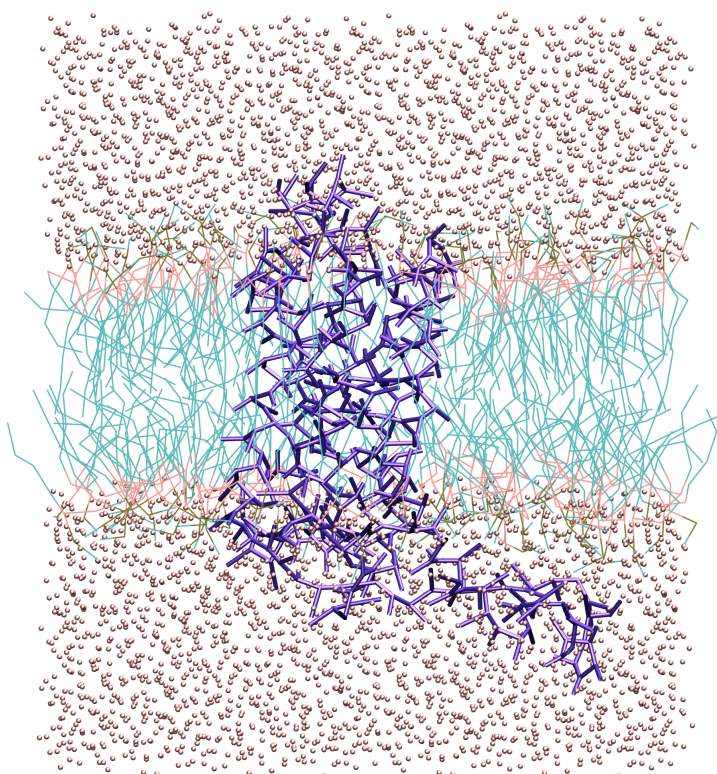
Figure 3.1. Fully atomistic (a) and Coarse-grained (b) representation of the  $\beta_2$ AR and POPC lipids.

Coarse-grained protein and lipid systems were then solvated with CG water molecules with VMD's Solvate module v1.2. Then the system was neutralized with a concentration of 0,154 mol/L Na<sup>+</sup> and Cl<sup>-</sup> ions. After ionization, the coarse-grained model has 13 Na<sup>+</sup> and 20 Cl<sup>-</sup> ions. The whole system included protein, cell membrane, water molecules and ions as shown in Figure 3.2 in both all-atom and coarse-grained representation. After coarse-graining the system, the number of atoms in protein was decreased from 5055 to 607, and for lipids from 20770 to 1860. For solvation a total of 4368 CG water molecules were added. Overall, the system size was reduced significantly: 68,001 atoms in fully atomistic (FA) model represented by 6,868 CG beads in coarse-grained (CG) model. The periodic box was taken as 89 Å x 98 Å x 101 Å. Details about the number of atoms and the box dimensions for both fully-atomistic and coarse-grained models are listed in Table 3.1 for comparison.





(a)



(b)

Figure 3.2. Fully atomistic (a) and Coarse-grained (b) representation of whole system of  $\beta_2$ AR (embedded in a POPC membrane, solvated and ionized)

Table 3.1. System size details of fully atomistic (FA) and coarse-grained (CG) models.

Run	Periodic Box Dimension (Å)	Number of Protein	Number of Lipids	Number of Waters	Number of Ions	Total Number of Atoms
FA	86x86x100	5.055	20.770	42.135	41	68.001
CG	89x98x101	607	1860	4368	33	6868

### 3.2. Coarse-Grained Molecular Dynamics Simulation Details

Before performing the main CG MD simulations, we had carried out some trial runs. First, the fully atomistic system's size, 86 Å x 86 Å x 100 Å, was used for maintaining the experimental surface area of the lipid molecules. However, the system crashed at an early stage. A summary of the performed trial runs with different system size is given in Table 3.2. Increasing the periodic box dimensions to 89 Å x 98 Å x 101 Å helped to maintain the system under equilibrium.

Table 3.2. Simulation system size and run length.

System Size	Time Step (fs)	Temperature (K)	Simulation Time
71x80x114	5	323	100 ns*
71x80x114	10	323	86 ns*
71x80x114	15	323	2 μs*
89x98x101	5	323	5 μs (completed)
89x98x101	10	323	158 ns*
89x98x101	15	323	85 ns
89x98x101	5	290	5 μs (completed)
89x98x101	5	310	6 μs (completed)

\*: system crashed at an early stage.

The time step for a fully atomistic simulation is generally 1 or 2 fs. In contrast, this value can be increased up to 25 fs in a CG simulation. The 800 ns FA simulation used 2 fs integration time step. In CG MD simulations, we first started to use 20 fs time step, however the system crashed at early stages. Decreasing the time step to 5 fs helped the system to run for 5-6  $\mu$ s. Three independent CG MD simulations were performed at different temperatures, that is 290, 310 and 323 K, to define the most proper environment for the system. Although it was suggested to use a temperature value which is  $\sim$ 10-20 K higher than 310 K in order to reproduce the results of an equivalent all-atom simulation [17, 23], we obtained similar results for three temperatures. All simulations were performed using the Nanoscale Molecular Dynamics (NAMD) v2.7 simulation tool. The system was gradually heated from 30 K to target temperature (290, 310, 323 K) to raise the kinetic energy with minimal conformational changes. The system achieved the final temperature after 14 ns. Then the system was minimized for 30,000 steps via Conjugate Gradient algorithm to eliminate the steric clashes between system components. CG-MD simulations were performed in NPT ensemble at 290, 310, 323 K temperature and 1 atm pressure. Langevin Dynamics with a damping coefficient of  $5\text{ ps}^{-1}$  was used for maintaining the temperature constancy; constant pressure was maintained using a Langevin Nosè-Hooverpiston with a period of 1000 fs and a decay time of 500 fs. 1-2 exclusion was used for non-bonded interactions with a cut off value of 12 Å with shifting starting at 9 Å. Simulations at 290 and 323 K were run for 5  $\mu$ s, while 310 K was run for 6  $\mu$ s with 5 fs integration time step. The overall system size was set to 89 Å x 98 Å x 101 Å with  $\beta_2$ AR, a phosphatidylglycerol membrane (POPC), TIP3 water and ions. According to RMSF profiles, 310 K simulation was considered to be the most compatible one for forward analysis.

### 3.3. Clustering of CG Trajectory

*K*-means algorithm implemented in *kclust* module of Multiscale Modeling Tools of Structural Biology Tool Set was used for clustering the 6  $\mu$ s CG MD trajectory at 310 K. Four different clustering based on different significant regions of the protein (core, ICL3, transmembrane, binding site) were performed. Binding-site and

transmembrane regions are illustrated in Figure 3.3. Total length of the CG simulation is 6  $\mu$ s that produces 5929 frames. To get exactly four clusters, RMSD values were set to 4 Å, 3 Å, 3 Å, and 2.5 Å for core, ICL3, transmembrane and binding site regions, respectively. Backbone atoms of the CG model were taken into account in clustering. The information about clustering process such as the aligned region and RMSD thresholds are given in Table 3.3.

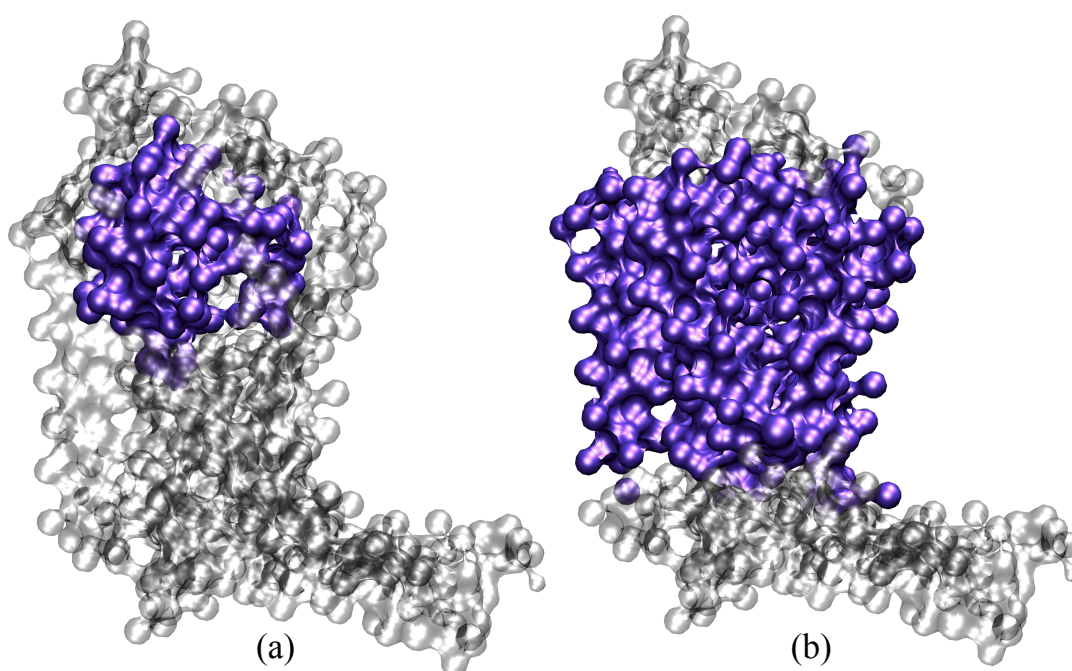


Figure 3.3. Binding-site (a) and transmembrane (b) regions of the protein.

Table 3.3. Clustering information for four different regions of the receptor.

Clustering procedure	Alignment	RMSD Calculation	RMSD Threshold (Å)	Number of Clusters
1	Core <sup>1</sup>	Core	4	4
2	Core	ICL3 <sup>2</sup>	3	4
3	Transmembrane <sup>3</sup>	Transmembrane	3	4
4	Bindingsite <sup>4</sup>	Bindingsite	2.5	4

<sup>1</sup>:Core region index: 35-58 69-93 107-134 150-169 200-228 270-296 308-325

<sup>2</sup>:ICL3 region index: 231-262

<sup>3</sup>:Transmembrane region index: 35-58 72-95 107-129 151-174 197-220 275-298 306-329

<sup>4</sup>:Binding-site region index: 82 86 90 93 106 109-119 121 164 165 169 174 191-197 199-209 286 289 290 292 293 294 297 308 309 311 312 313 315 316

Figure 3.4 displays the distribution of clusters of core (a), ICL3 (b), transmembrane (c) and binding-site (d) regions according to frame number. Clusters of the core, trans-membrane and binding-site, which can be seen in Figure 3.4 (a), (c) and (d) respectively, display a similar profile. Early stages of the simulation, which include the equilibration period, fall into different three small clusters, while production run creates one large cluster, which is dominant for the last 4500 ns. On the other hand, clustering of ICL3 (Figure 3.4b) displays a completely different profile. The simulation is mainly divided into two distinct clusters, the first half being in one cluster, and the second half in another cluster.

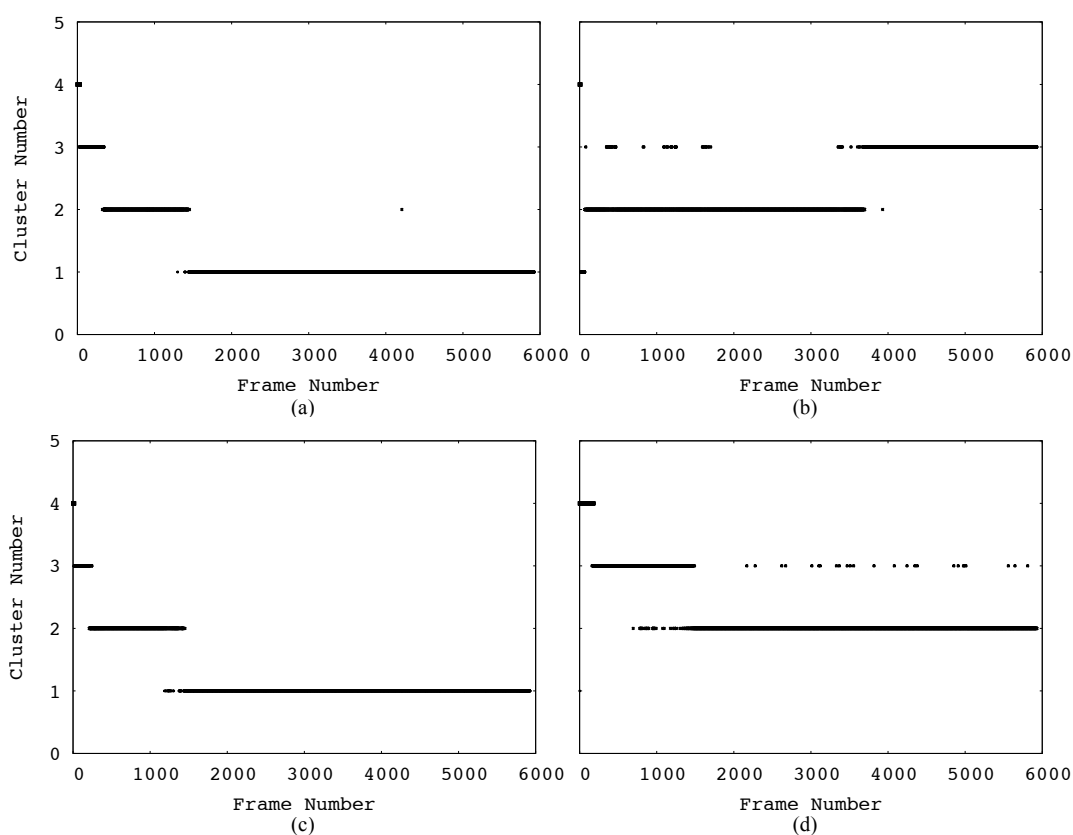


Figure 3.4. Cluster profile of CG MD trajectory according to different regions of  $\beta_2AR$ . (a) core region (RMSD=4Å), (b) ICL3 region (RMSD=3Å), (c) transmembrane region (RMSD=3Å), (d) binding-site region (RMSD=2.5).

### **3.4. Reverse Mapping of Coarse-Grained Models to Their Atomistic Representation**

The atomistic details of the CG model were recovered via CG Tools Plugin of VMD. First, the CG model for protein was extracted from the system and reverse-mapped alone. Then, the protein was embedded into a new bilayer membrane (POPC), solvated and ionized. Following the three preparation phases, which include melting lipids, constraining the protein and releasing the protein, the system was ready for a 100 ns MD simulation for further equilibration. Four CG models were chosen for reverse mapping and they were extracted from four different clustering, which had been performed according to core, ICL3, trans-membrane, and binding-site region. They are simply named as RM1, RM2, RM3, and RM4, respectively. The conformations were chosen as the element, which was nearest to the centroid from the most crowded cluster.

#### **3.4.1. System Preparation for Reverse-Mapped Molecular Dynamics (MD) Simulation**

The VMD Membrane Builder Plugin was used to generate the cell membrane. 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC) doubled-layered cell membrane was built with a 100 Å thickness in the direction of  $z$ -axis. The membrane's  $x$  and  $y$  dimensions were defined according to the protein's dimensions in the same directions. To give an example, minimum and maximum  $x$  coordinates of the protein were determined as -53.6 Å and -0.9 Å, which makes the dimension of the protein as ~55 Å. As the minimum distance between the protein and the boundary of the periodic cell was set to 15 Å, the box dimension in  $x$  direction became 85 Å (=55+15+15). The box dimension in  $y$  direction was calculated similarly.

After protein and membrane were aligned, the system was solvated with a thickness of 15 Å water molecules at both directions of the  $z$ -axis. Table 3.4 displays the protein's dimensions and the system size of all RM systems in detail. Finally, the systems were ionized with a certain number of Na<sup>+</sup> and Cl<sup>-</sup> ions with a concentration

of 0,154 mol/L to make the total charge of the system equal to zero. VMD's Solvate module v1.2 was used for that purpose. RM systems' total numbers of atoms can be seen in Table 3.5.

Table 3.4. The dimensions of the reverse-mapped protein and surrounding system. All values are given in Angstrom.

	<b>Protein</b> $x_{min} / x_{max}$	<b>Protein</b> $y_{min} / y_{max}$	<b>Protein</b> $z_{min} / z_{max}$	<b>Protein Dimension</b> (xyz)	<b>Box Dimension</b> (xyz)
<b>RM1</b>	-53.66/-0.96	-57.89/-1.33	-16.24/52.70	55x59x68	85x89x100
<b>RM2</b>	-65.32/-10.27	-62.22/-5.91	-14.62/53.18	55x56x69	85x86x100
<b>RM3</b>	-48.57/4.98	-73.44/-14.47	-29.70/38.02	54x59x68	84x89x100
<b>RM4</b>	-50.40/3.39	-71.25/-10.73	-31.57/33.83	54x61x65	84x91x100

Table 3.5 Total numbers of atoms in reverse-mapped structures.

	<b>Protein</b>	<b>Lipid</b>	<b>Water</b>	<b>Na<sup>+</sup></b>	<b>Cl<sup>-</sup></b>	<b>Total</b>
<b>RM1</b>	5055	18760	40311	39	46	64211
<b>RM2</b>	5055	19028	41247	40	47	65417
<b>RM3</b>	5055	19430	42174	41	48	66748
<b>RM4</b>	5055	20100	43584	42	49	68830

### 3.4.2. Molecular Dynamics Simulation Details of Reverse-Mapped Structures.

VMD Membrane Builder does not provide an equilibrated membrane patch. Thus, before running the MD simulations of the RM structures, the systems had been prepared for equilibration. Preparation stages were carried out in the following three steps.

### 3.4.2.1. Melting of Lipid Tails

Normally the interior of a lipid bilayer is highly fluid and hydrocarbon chains of phospholipids are disordered. In contrast, lipids generated by VMD are in a highly ordered structure, which is unrealistic. To create a natural state of lipids, we performed a simulation in which everything (protein, water, ions, lipid head groups) except lipid tails, was fixed in the first stage. After 1000 steps of minimization, 0.5 ns MD simulation was performed. As a result, a fluid-like, disordered cell membrane was obtained as shown in Figure 3.5.

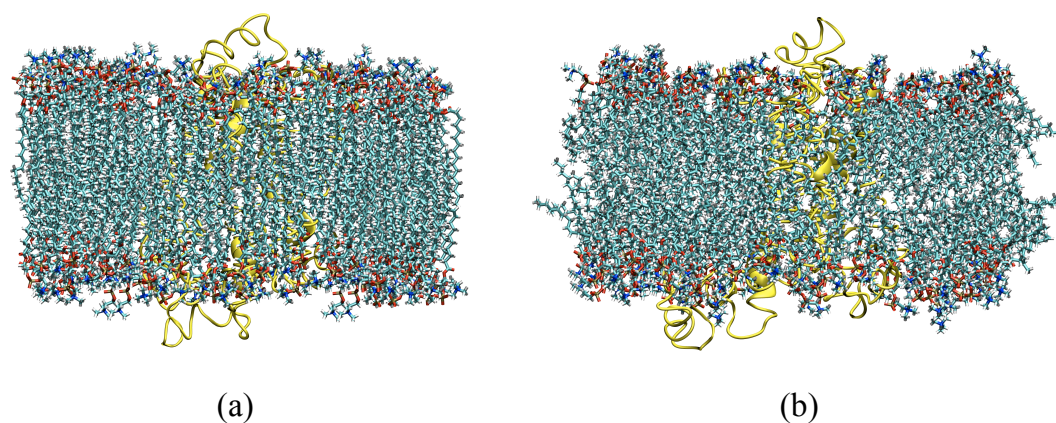


Figure 3.5. Cell membrane representation before (a) and after (b) the melting of the lipid tails.

### 3.4.2.2. Minimization and Equilibration with Protein Constrained

Still the system is not ready for an MD simulation due to many unnatural atomistic positions. For this reason, a harmonic constraint was applied on the protein, while lipids, waters and ions were released. This process allows cell membrane and water molecules to adapt to the protein in a shorter time interval. A 1000 steps of minimization was followed by 0.5 ns MD simulation.

### 3.4.2.3. Equilibration with Protein Released

Lipid molecules were well packed after protein-constrained simulation. In this final step, the protein's harmonic constraints were released and equilibration was



performed for the whole system. As in the previous steps, after 1000 steps of minimization, an MD simulation was carried out for 0.5 ns. After equilibration, the lipid molecules became well packed around the protein as seen in Figure 3.6. During the equilibration, the volume decreases in  $xy$  direction, while in  $z$  direction; the volume increases to keep the initial volume and the pressure at 1 atm.

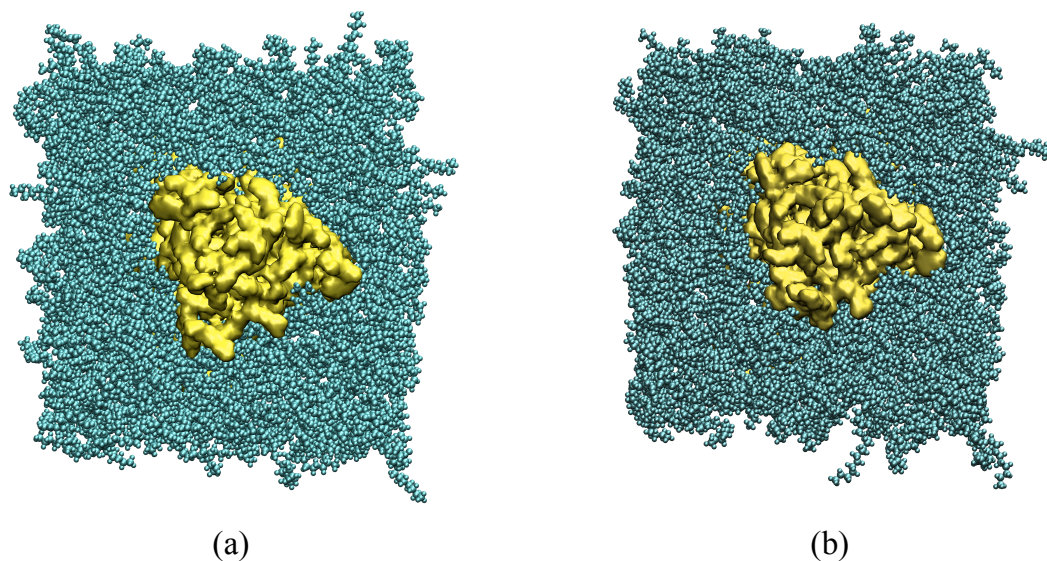


Figure 3.6. Cell membrane representation before (a) and after (b) the equilibration of the whole system.

#### 3.4.2.4. Production Runs

All three-preparation steps were performed for four different reverse-mapped (RM) structures (RM1, RM2, RM3, RM4). 100 ns MD simulations were performed using NAMD v2.7 software. For interaction potentials, CHARMM27 (for lipids) and CHARMM22 (for proteins) force fields were used. All simulations were carried out at 310 K.

### 3.5. Root Mean Square Deviation (RMSD) of CG Simulations

To define the simulations' equilibrium states and conformational changes throughout the trajectory, Root Mean Square Distance (RMSD) was calculated. In Table 3.6, different RMSD calculations and their terminology are listed. To give an example, RMSD “*All Fit Core*” stands for RMSD value of the whole receptor with respect to its initial structure, after aligning the receptor based on the core region. All alignments are performed using backbone atoms only.

Table 3.6. Terminology used for various RMSD calculations.

<b>RMSD Value</b>	<b>Fitting Value</b>	<b>Terminology</b>
All	All	All Fit All
Core	Core	Core Fit Core
ICL3	Core	ICL3 Fit Core

Figure 3.7 shows RMSD *All Fit All* values that stabilize around 4.25 Å, 5.5 Å, and 6 Å for 290 K, 310 K, and 323 K respectively. All CG simulations at different temperatures reach equilibrium states after a short simulation time. The lowest RMSD value was obtained for the lowest temperature value of 290 K.

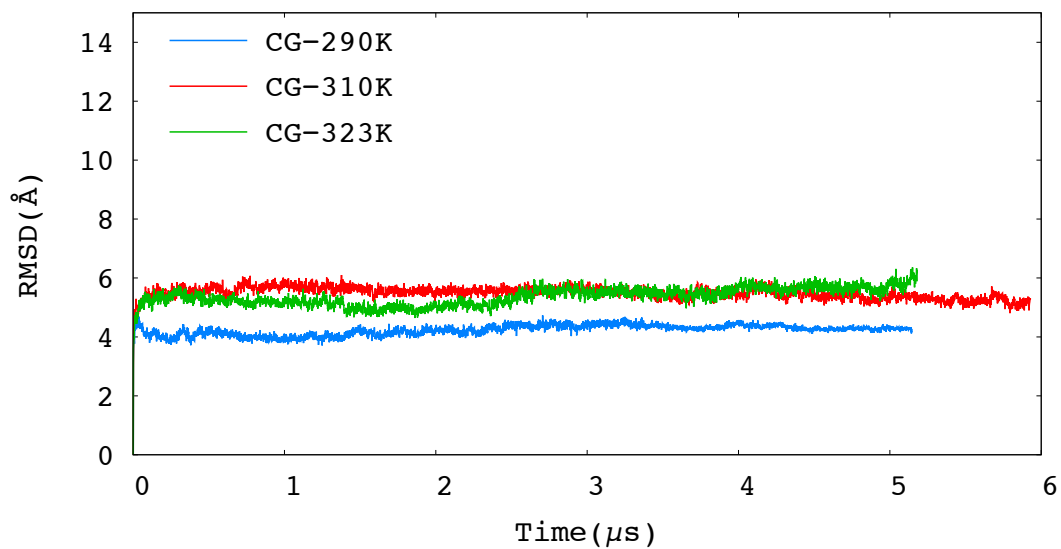


Figure 3.7. RMSD *All Fit All* profile of CG simulations.

Whole protein except ICL3 loop region is named as the “core” region. Figure 3.8 shows RMSD *Core Fit Core* profile in which ICL3 region is excluded on RMSD calculation. The simulations have RMSD values stabilizing at around 4, 5 and 5 Å for 290, 310, 323 K simulations respectively. The effect of the ICL3 on RMSD values of the protein is nearly negligible with a 0.5 Å difference.

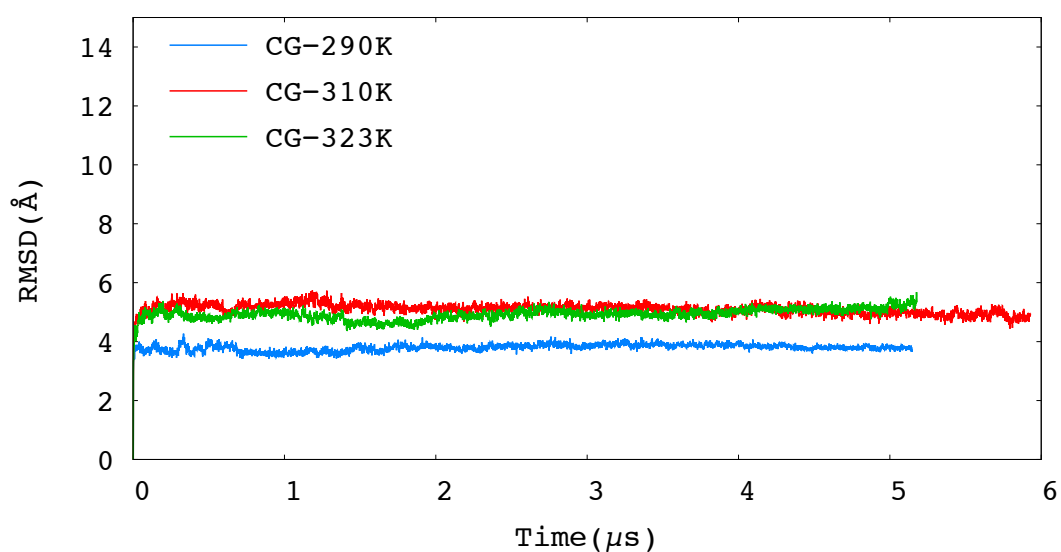


Figure 3.8. RMSD *Core Fit Core* profile of CG simulations.

Finally, the RMSD values were calculated using only the ICL3 region after aligning the backbone atoms of the core region to the initial frame and shown in Figure 3.9. As expected, the most flexible part of the protein is the ICL3 loop region. RMSD values fluctuate the most at 323 K, within the range of [6-13] Å. At around 2.5  $\mu$ s, the RMSD value for 323 K simulation gradually increases by 3-4 Å. This conformational change in ICL3 can also be seen in Figure 3.10.

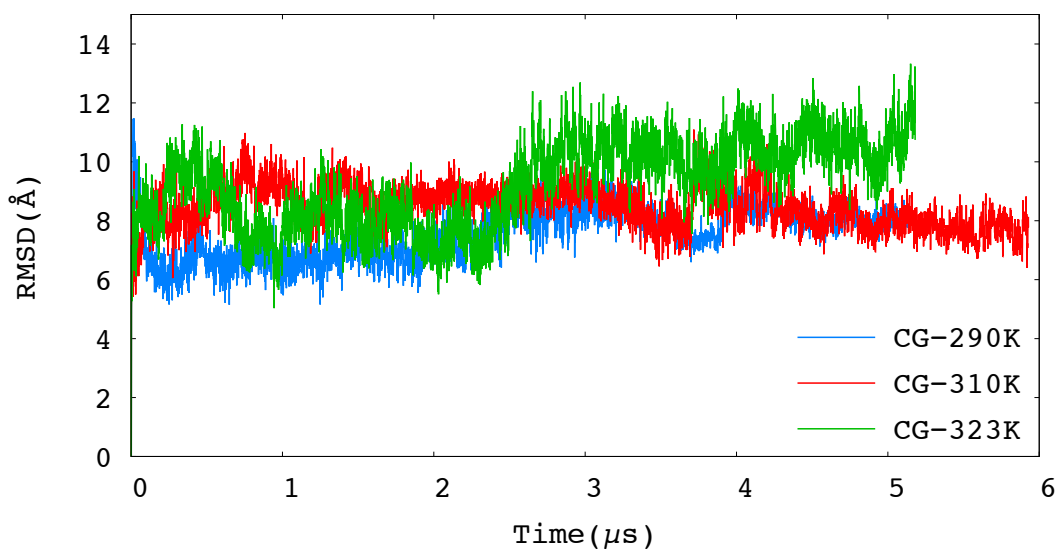


Figure 3.9. RMSD *ICL3 Fit Core* profile of CG simulation.

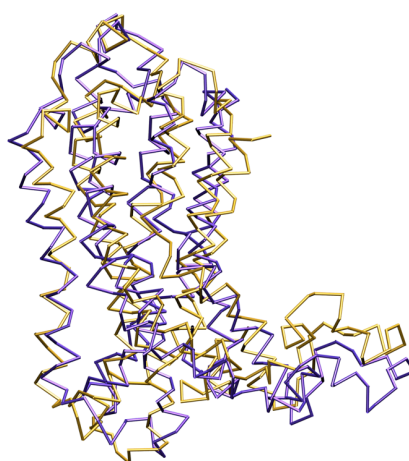


Figure 3.10. The snapshot from 2  $\mu$ s is colored in violet while the yellow colored structure is from 4  $\mu$ s.

### 3.6. Root Mean Square Fluctuations of CG Simulations (RMSF)

The root mean square fluctuation profile gives the best elucidation for quantitatively expressing the protein's mobility along the simulation. Position in space during the simulation of each residue in the protein is calculated by taking the average squared deviation from the reference point. Here, the reference point is the average position of the protein in space along the simulation. All conformations over the trajectory had been aligned to the average structure of the simulations before calculating the RMSF values. Alignment was performed using only  $C_{\alpha}$  atoms for fully atomistic simulations and backbone atoms for CG simulations.

RMSF values of CG simulations were compared with fully atomistic long (FAL) simulation as illustrated in Figure 3.11. Accordingly, the intracellular loop III (ICL3) region has the highest mobility for all CG and FA simulations. As expected, the fluctuation values of CG MD simulations are less than the fluctuation value of FA MD simulation result. The greater masses of the CG particles can lead to fewer fluctuations than in a fully atomistic model. While ICL3 fluctuation of fully-atomistic model is up to 15.2 Å, in coarse-grained model these values reach only 3.8 Å, 3 Å, 2.3 Å for 310, 323, 290 K simulations respectively. The RMSF profile of 310 K simulation was found to be the most compatible with FA simulation. Thus, 310 K simulation was chosen in future comparison analysis.

Other mobile regions for fully atomistic model are the intracellular loop II (ICL2) and extracellular loop II (ECL2). These loop regions are displayed in Figure 3.12. CG model (310 K) shows some fluctuations in ICL2, but not in ECL2 region.

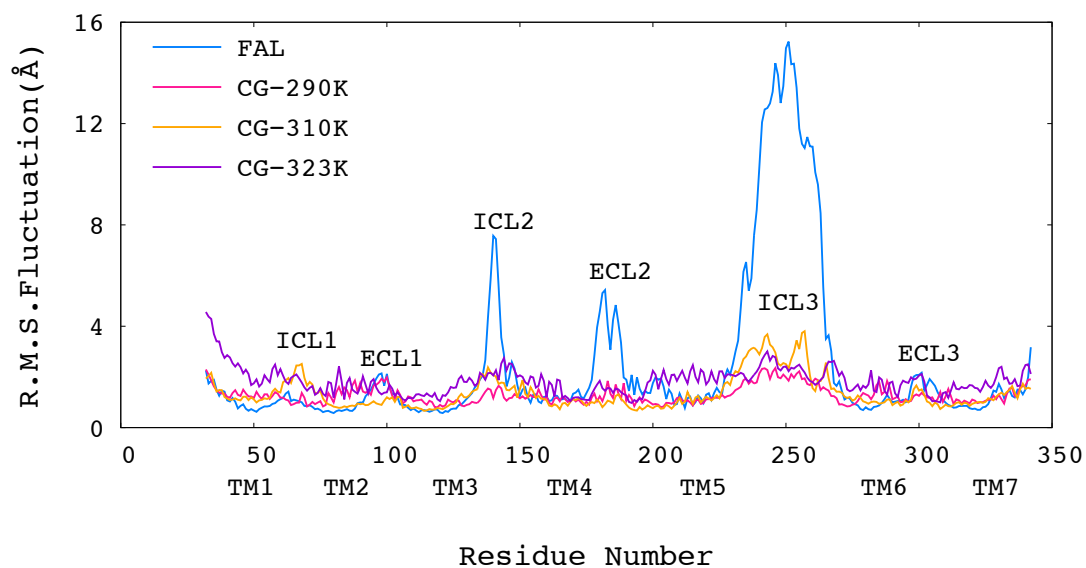


Figure 3.11. RMSF profiles of CG MD and FA MD simulations. (First 30 ns, 97 ns, 200 ns and 48 ns were excluded from FA's trajectory, 290 K, 310 K, and 323 K CG's trajectories, respectively)

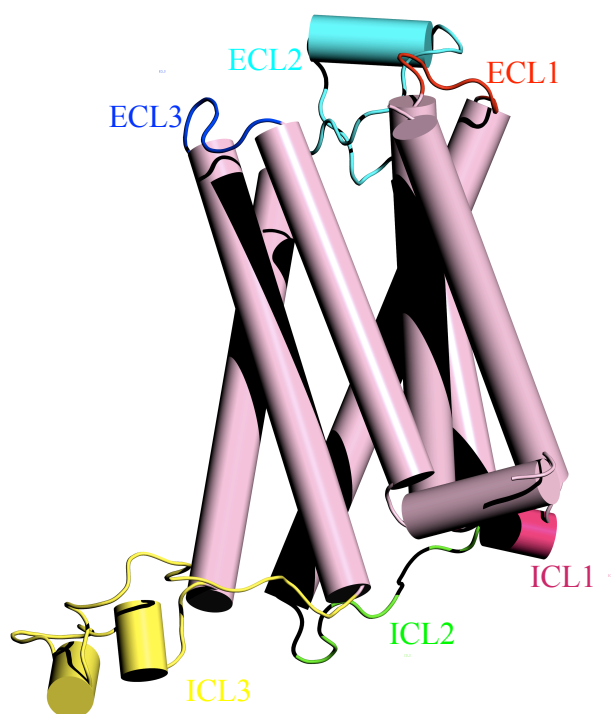


Figure 3.12. Cartoon representation of  $\beta_2$ AR with loop regions

In previous studies, three independent 100 ns long fully-atomistic simulations were performed to determine whether the conformational sampling would be sufficient

and be able to find different conformers within a 100 ns period. It is a well-known fact that simulations started at different initial conditions (such as different initial velocities) will enrich the conformational sampling [39]. Figure 3.13 displays RMSF values for one long (FAL) and three short (FAS1, FAS2, FAS3) fully atomistic MD and CG MD simulations. Overall, the mobility seen in short FA simulations is comparable to that in long FA simulation. The highest mobility of the protein is observed at the ICL3 for all short and long FA MD simulations. However, the ICL3's mobility in the second short FA (FAS2) simulation is lower with respect to other FA models and comparable to that in CG model.

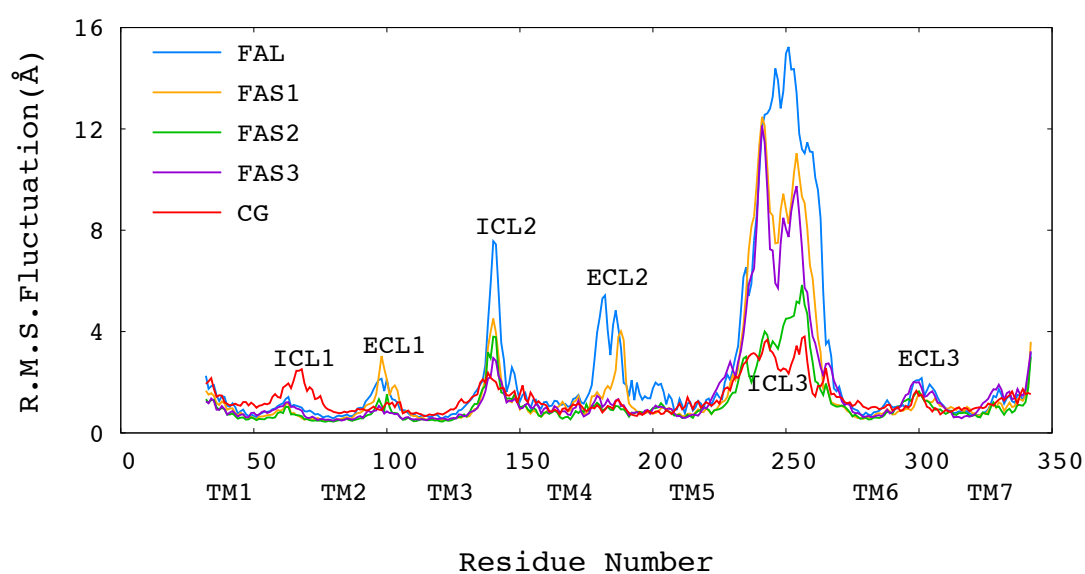
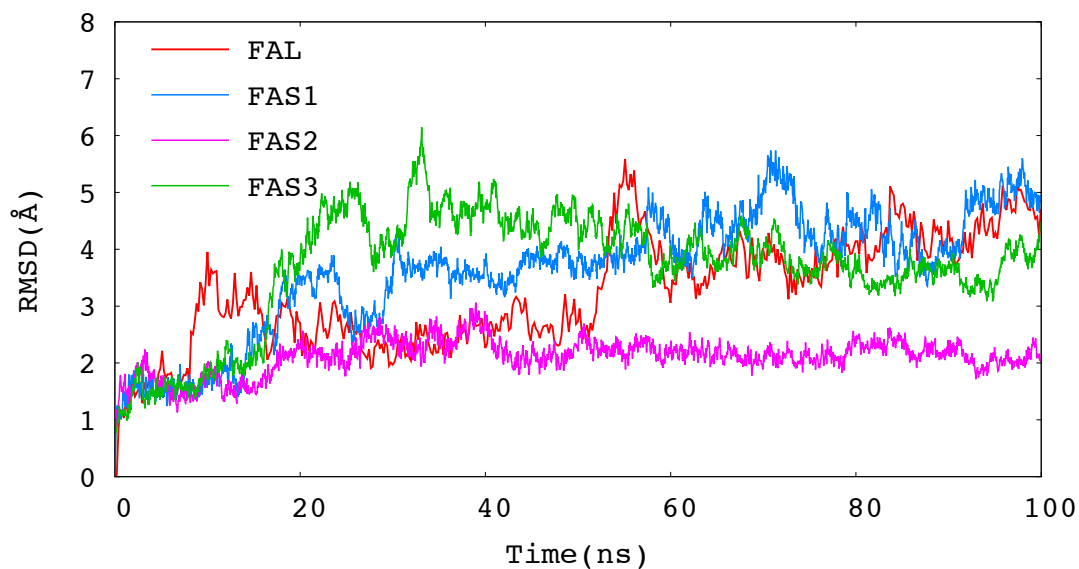


Figure 3.13. RMSF profiles of CG MD, long and short FA MD simulations. (First 30 ns were excluded from FA long trajectory, 2 ns from FA short trajectories.)

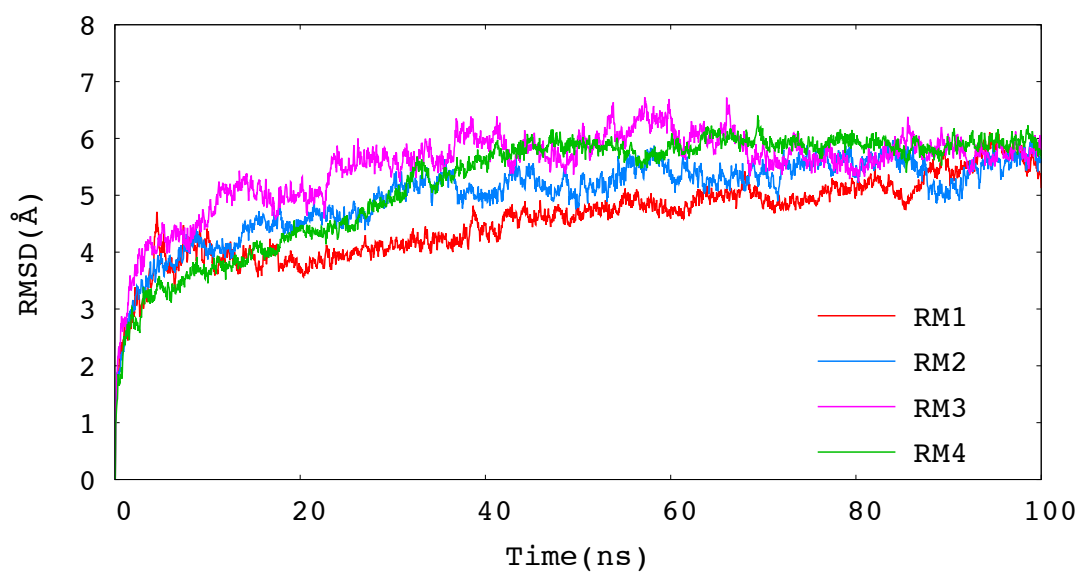
### 3.7. RMSD of Reverse-Mapped (RM) Simulations

RMSD values have been calculated for four different RM MD simulations and compared with those of fully atomistic simulations. Figure 3.14 (a) shows RMSD *All Fit All* profiles of fully atomistic simulations. FAL simulation's first 100 ns were considered while comparing with other short simulations. FAS2 reaches equilibrium state earlier than the others. FAL simulation reaches equilibrium state around 60 ns.

FAS1 and FAS3 reach equilibrium state around 40 ns. RMSD values fluctuate within a range of [2 Å-6 Å].



(a)



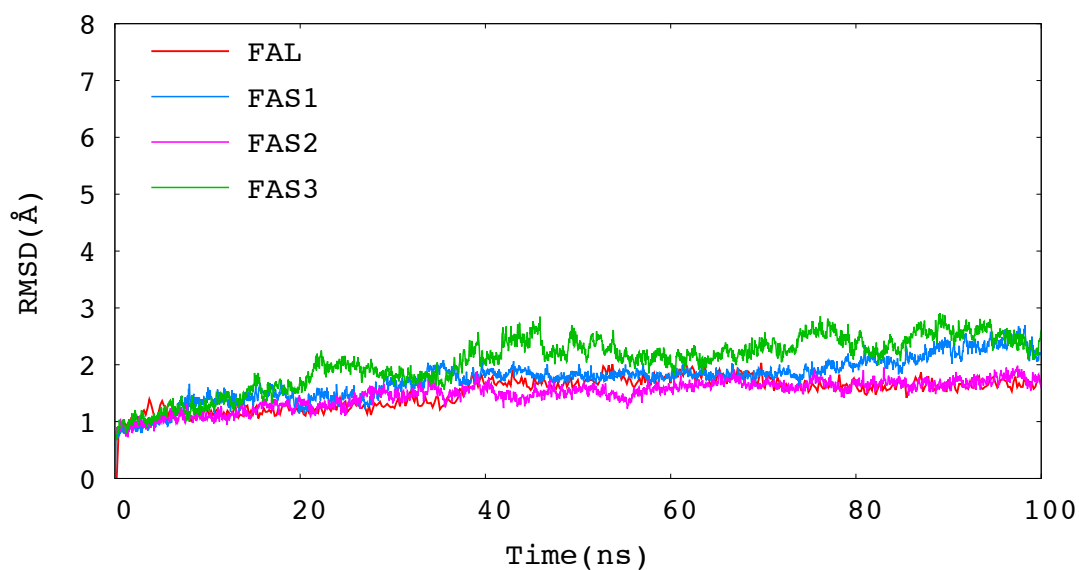
(b)

Figure 3.14. RMSD *All Fit All* profile of FA (a) and RM (b) simulations.

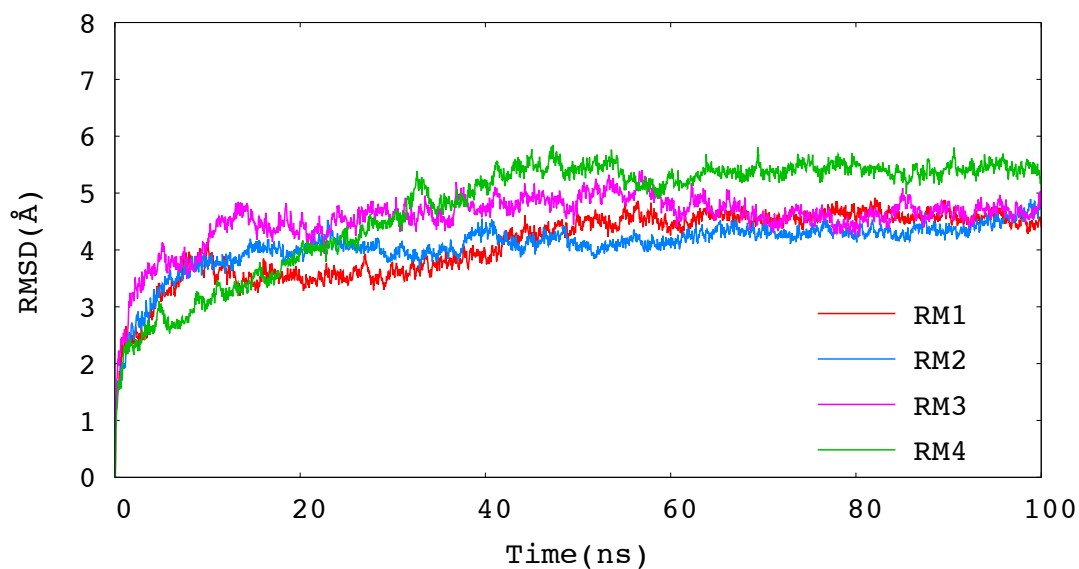
The RMSD (*all-fit-all-RMSD*) profiles of all four RM MD trajectories are shown in Figure 4.14 (b). All RM structures reach equilibrium state in a shorter time at around



10 ns. Also, the RMSD fluctuations are less in the FA simulations.



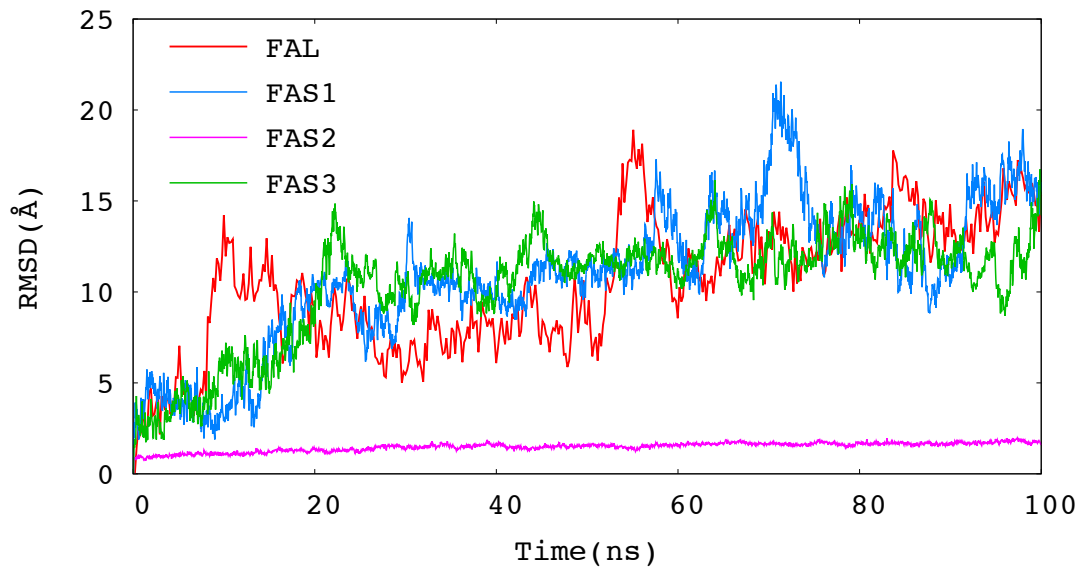
(a)



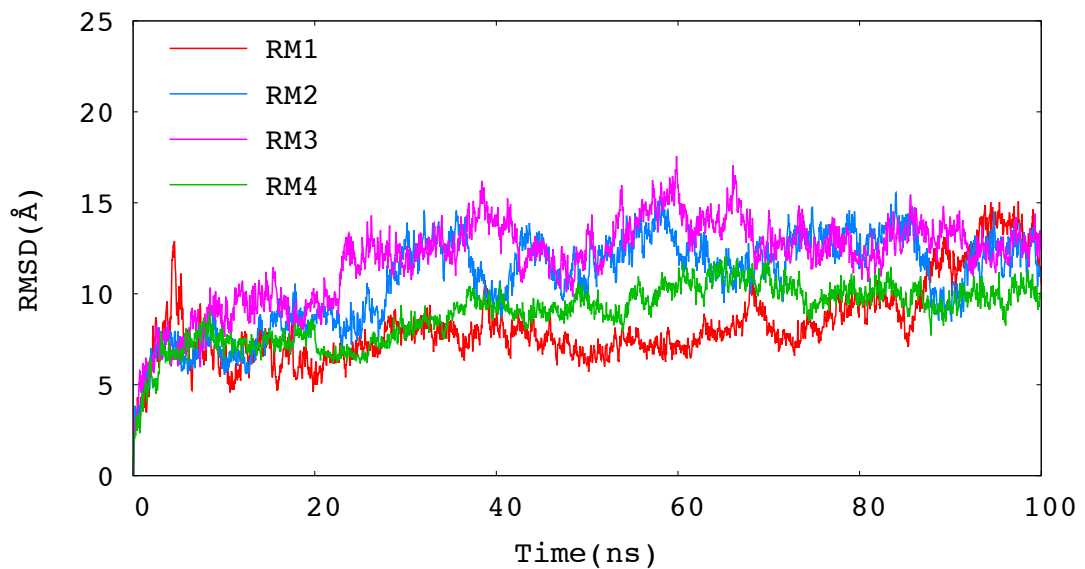
(b)

Figure 3.15. RMSD *Core Fit Core* profiles of FA (a) and RM (b) simulations.

Core region's RMSD profiles (*core-fit-core-RMSD*) of FA simulations and RM simulations are illustrated in Figure 3.15. All four FA simulations' RMSD stabilize at around 2 Å, while the RMSD value of RM simulation stabilize at 5 Å.



(a)



(b)

Figure 3.16. RMSD *ICL3 Fit Core* profiles of FA (a) and RM (b) simulations.

The most flexible part of the protein is ICL3 as can be seen in Figure 3.16, which displays the loop region's RMSD profiles (*ICL3-fit-core-RMSD*). FA simulations' RMSD values fluctuate within a 10 Å window, while RM simulations' RMSD values fluctuate within a 5 Å range. Surprisingly, in simulation of FAS2, ICL3 was stabilized at a very early stage at around 2 Å.

### 3.8. RMSF of Reverse-Mapped Simulations

RMSF values of RM simulations are compared with FAS1 as displayed in Figure 3.17. RM simulations indicate a good compliance with FAS1. As expected, the highest mobile region is the ICL3 for all RM simulations. The best compliant with FAS1 is the RM2. While ICL3 fluctuation of FAS1 is up to 13 Å, for RM simulations the fluctuation is between 5 and 10 Å.

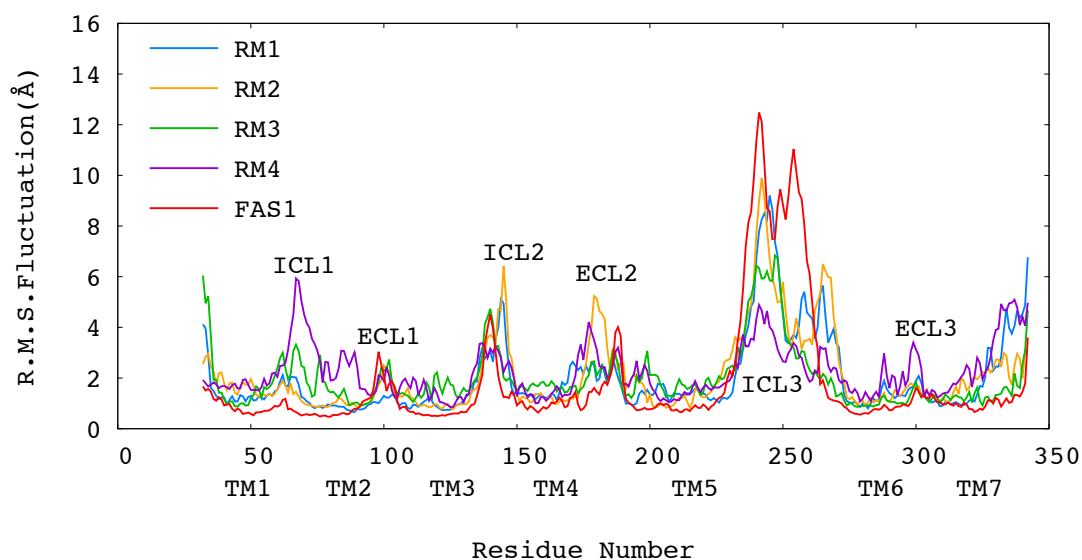


Figure 3.17. RMSF profiles of RM simulations and FAS1

### 3.9. Energy Profiles of RM and FA Simulations

The intra-molecular and intermolecular interaction energy profiles of RM simulations were compared with FA short simulations to check the energetic state of the models. Figure 3.18 reveals that the total energy value of for RM1, RM2, RM3 and RM4 simulations are stable at around -100800, -103000, -105500, -109000 kcal/mol respectively, while the energy value of all three FA simulations is stable at around -98000 kcal/mol.

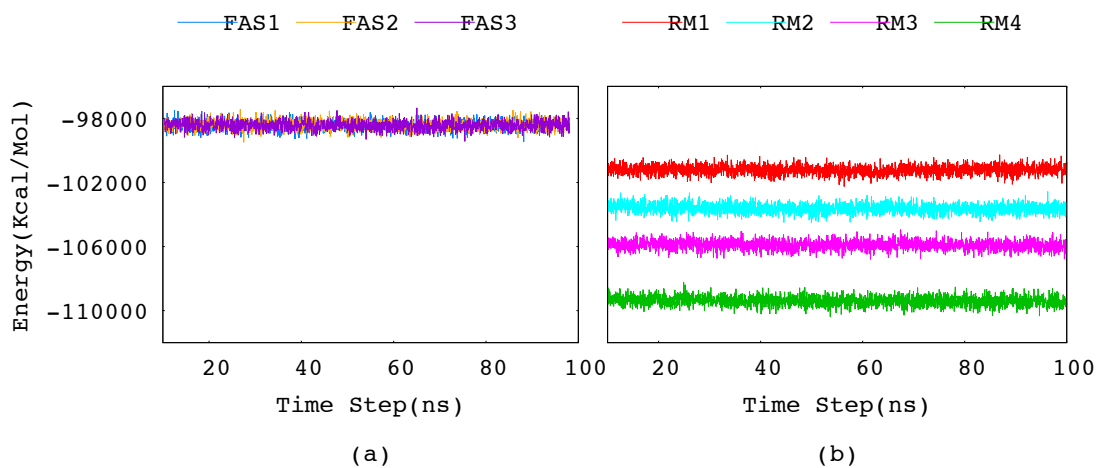


Figure 3.18. Total energy profiles of FA short (a) and RM simulations (b).

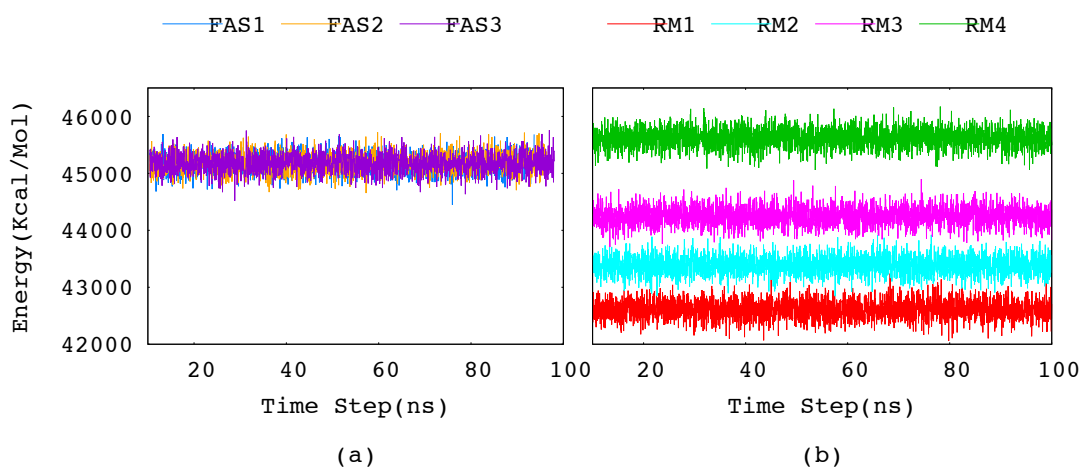


Figure 3.19. Kinetic energy profiles of FA short (a) and RM simulations (b).

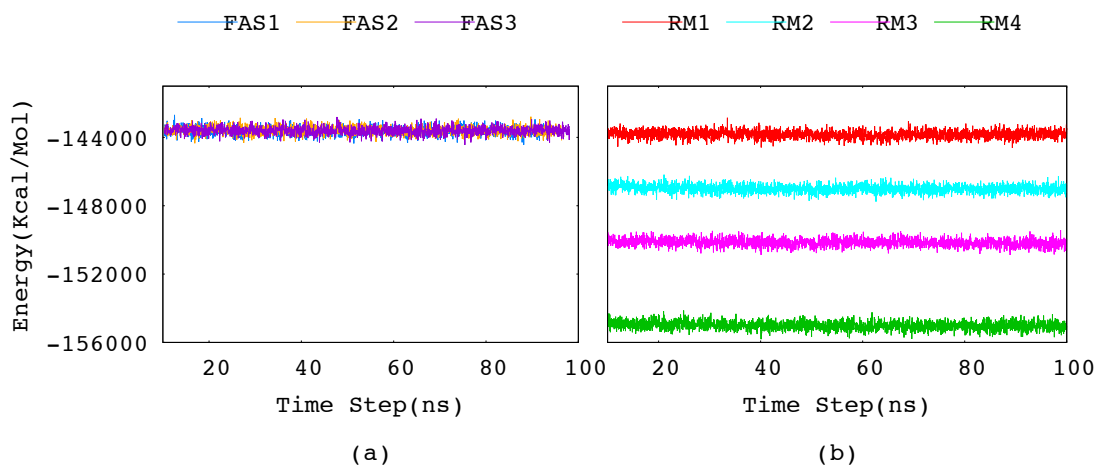


Figure 3.20. Potential energy profiles of FA short (a) and RM simulations (b). In order to understand the nature of the difference in total energy in RM simulations, the total energy is first decomposed into its two components, which are kinetic, and potential energy. Figure 3.19 shows the kinetic energy values stabilized at 42,500, 43,500, 44,300 and 45,500 kcal/mol for RM1, RM2, RM3 and RM4 simulations, respectively. Similarly, Figure 3.20 shows the potential energy values stabilized at -143200, -146500, -149800 and -154800 kcal/mol. The amount of difference between the lowest and the highest value for kinetic and potential energy are calculated as 3000 and 11,000 kcal/mol, respectively. Thus, the highest contribution to the difference is coming from the potential energy component.

The potential energy is further decomposed into its electrostatics, vdW and bonded energy terms (bond, angle, dihedral and improper). Figure 3.21 shows the electrostatics energy profile of all four RM simulations in comparison to FA simulations. Similarly, all three FA simulations are stable at around a single value (-176,000 kcal/mol) while the four RM simulations are stable at four different energy values, -172900, -176500, -181000 and -186000 kcal/mol, respectively. The amount of difference between the lowest and the highest value for electrostatic energy was determined as 13,000 kcal/mol. Clearly, the energy component that contributed most to the difference of 8,000 kcal/mol in total energy is the electrostatic energy.

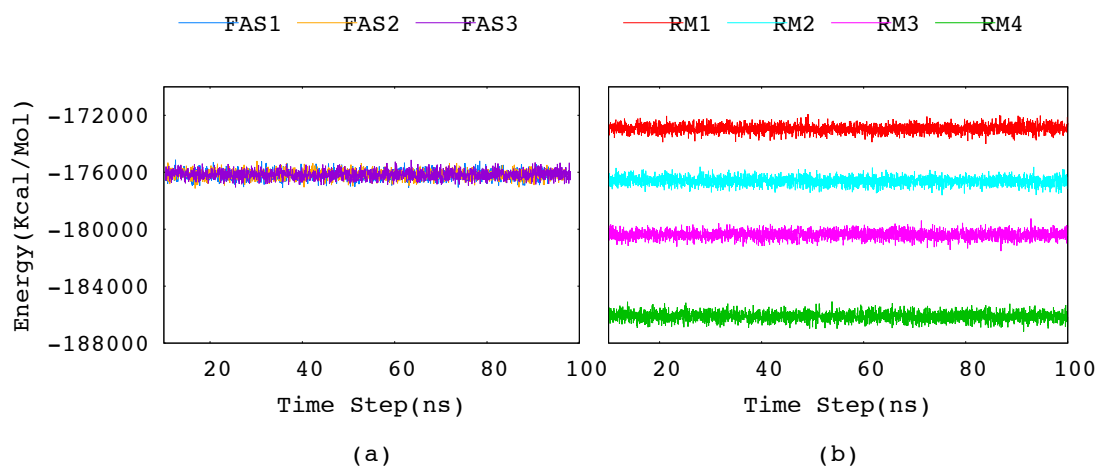


Figure 3.21. Electrostatic energy profiles of FA short (a) and RM simulations (b).

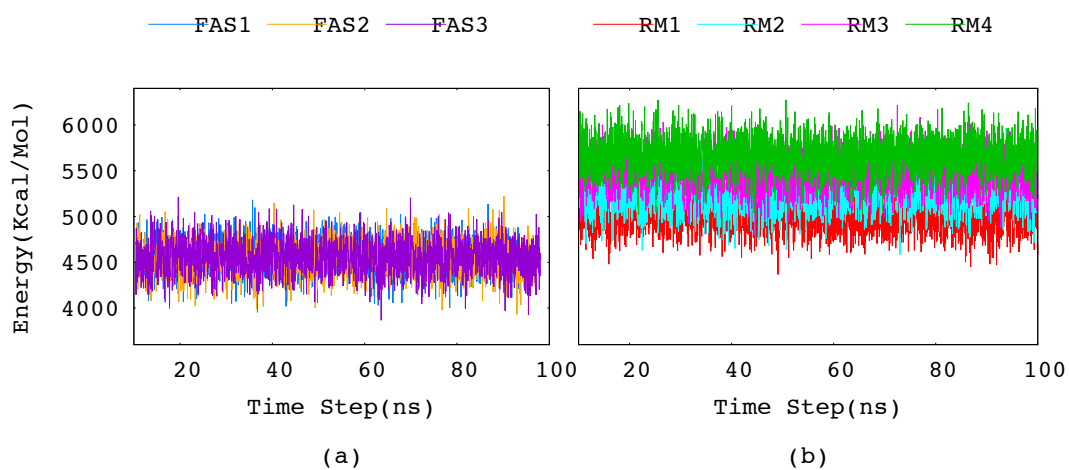


Figure 3.22. The van der Waals (vdW) energy profiles of FA short (a) and RM simulations (b).

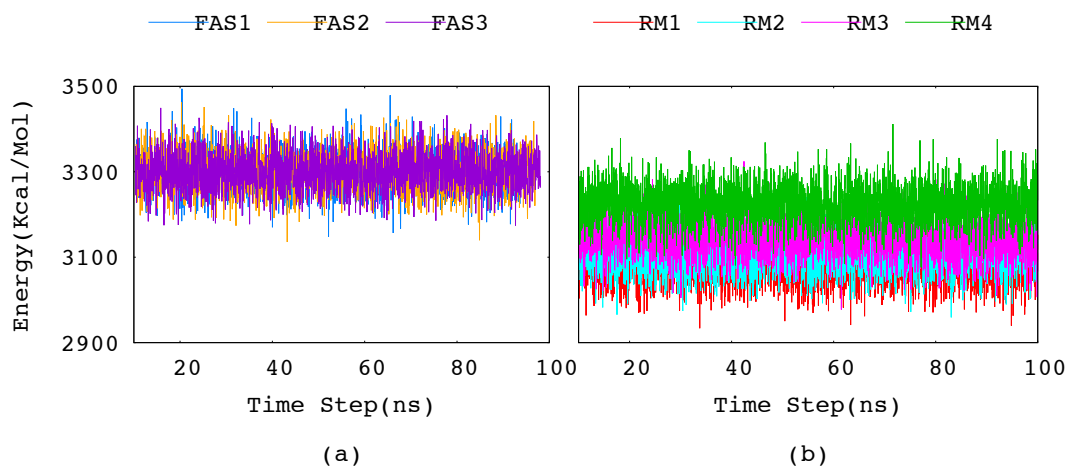


Figure 3.23. Bond stretching energy profiles of FA short (a) and RM simulations (b).

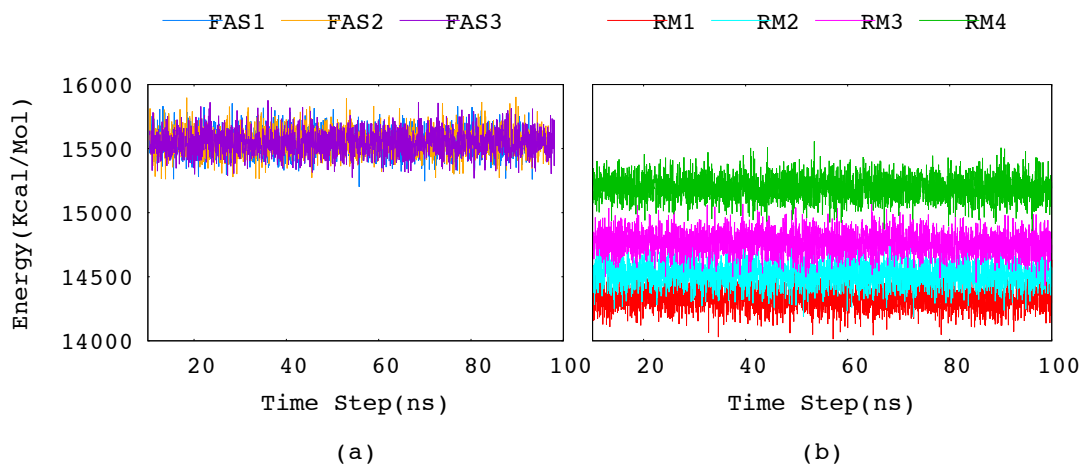


Figure 3.24. Angle bending energy profiles of FA short (a) and RM simulations (b)

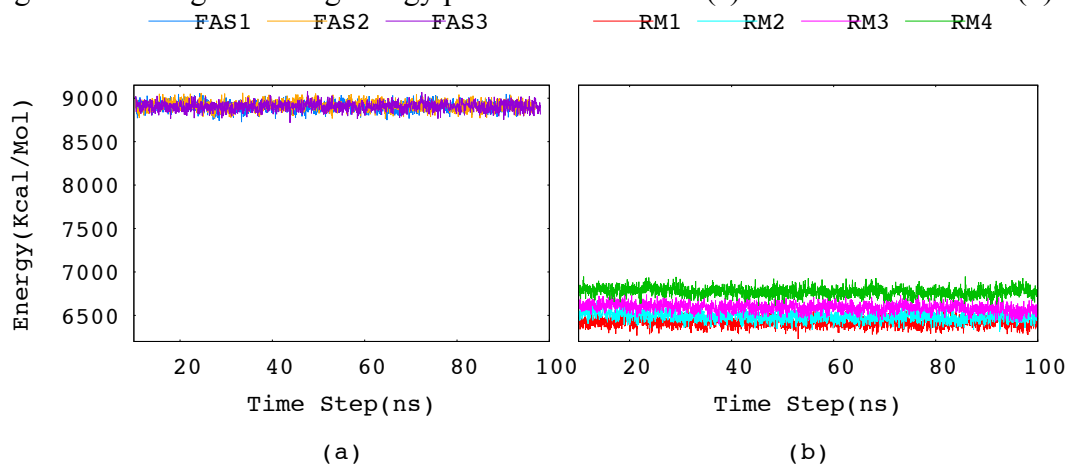


Figure 3.25. Dihedral energy profiles of FA short (a) and RM simulations (b).

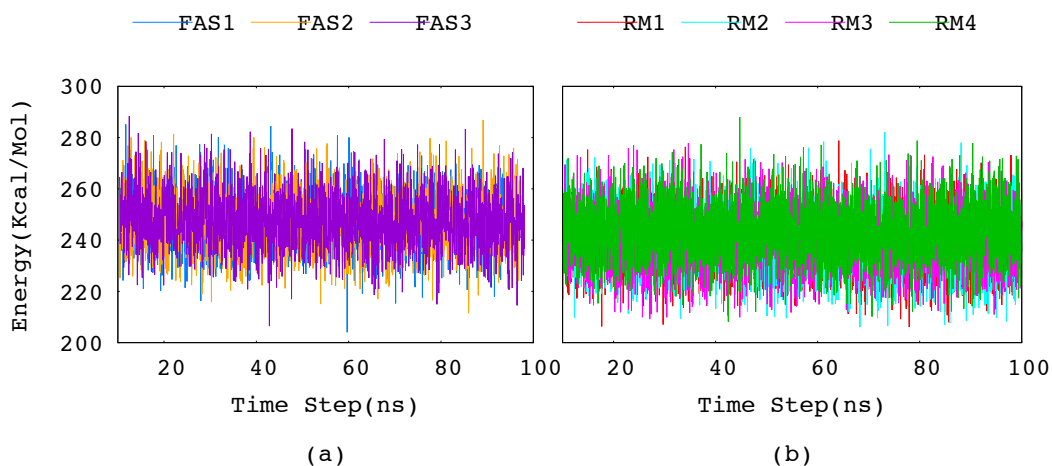


Figure 3.26. Improper energy profiles of FA short (a) and RM simulations (b).

### 3.10. Clustering of All FA and RM Trajectories

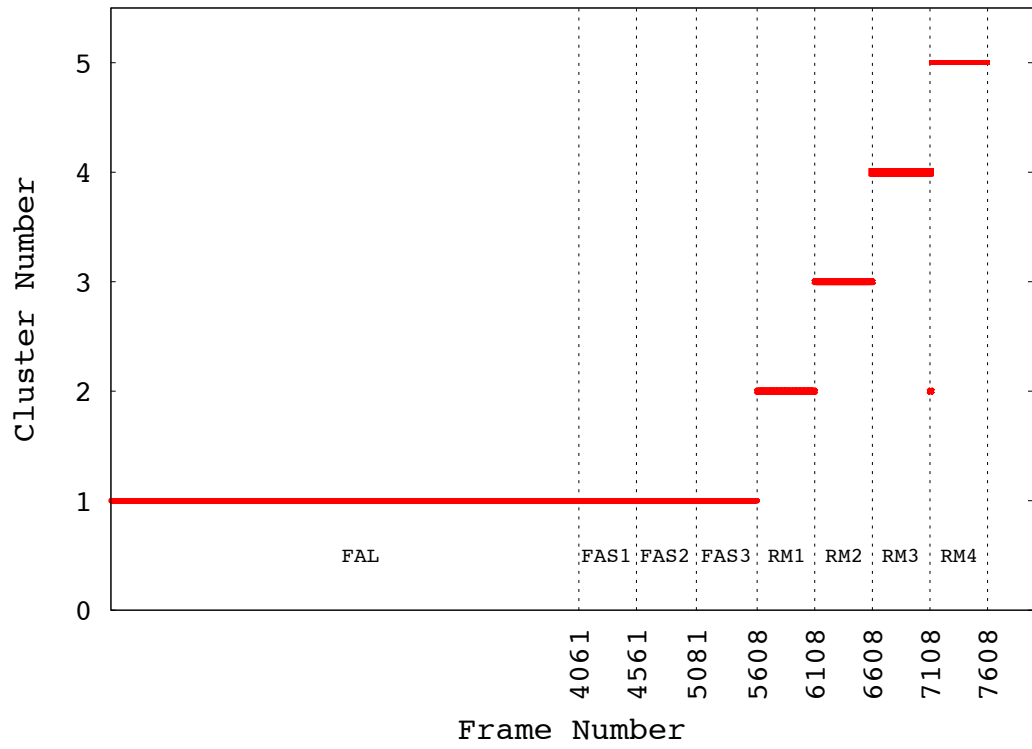
In order to reveal distinct conformers of  $\beta_2\text{AR}$ , all RM and FA simulation trajectories were clustered via *k-clust* algorithm using Multiscale Modeling Tools of Structural Biology Tool Set (MMTSB). All MD snapshots were clustered based on the core, trans-membrane, binding-site and ICL3 regions at different RMSD thresholds and their cluster profiles are shown in Figure 3.27 a-d respectively. The RMSD threshold values for the core, ICL3, trans-membrane and binding-site regions were set to 5 Å,

5.9 Å, 3 Å and 2.8 Å, respectively. These threshold values were chosen such that they would produce 5 or 6 clusters in total. In all four cluster profiles, 800 ns long FA simulation with 4061 snapshots are shown in the first range [1-4061], followed by three 100 ns long FA simulations each having around 500 conformers, thus occupying the intervals of [4062-4561], [4562-5081] and [5082-5608]. The last four intervals belong to 100 ns long RM simulations and they are located at [5609-6108], [6109-6608], [6609-7108] and [7109-7608].

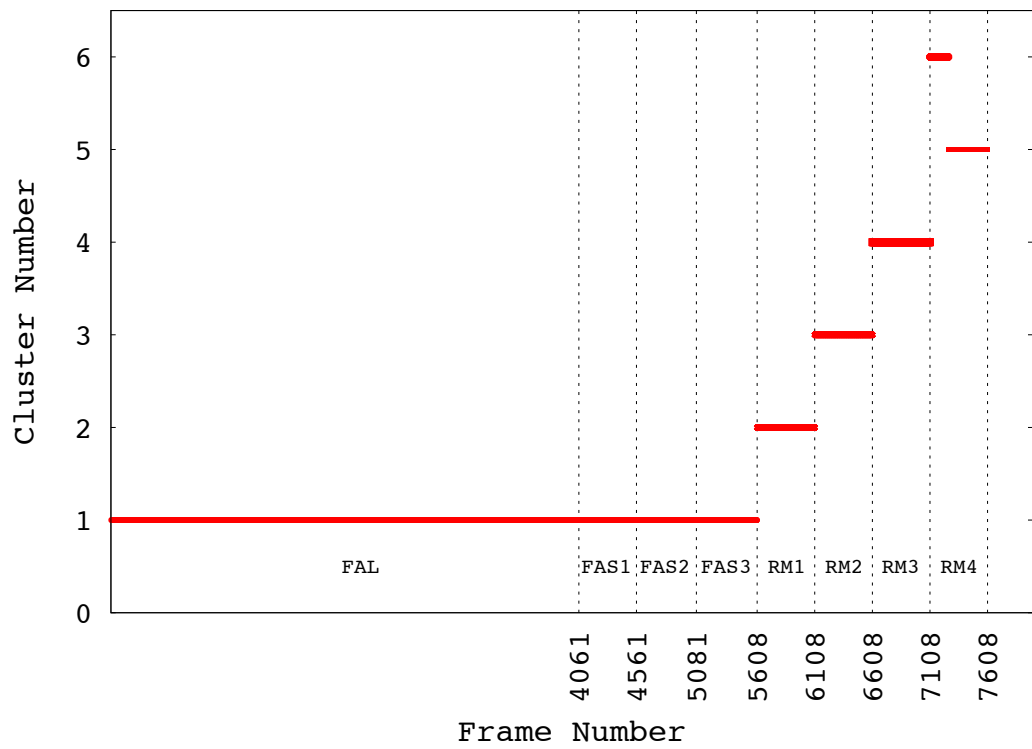
Figure 3.27a displays the clustering profile performed based on the core region with an RMSD value of 5 Å. Clearly, all FA simulations fall into the same cluster, while each RM simulation is clustered into distinct clusters. Similar profiles were obtained for clustering based on trans-membrane and binding site regions (Figs. 3.27b, 3.27c). Finally, the cluster profile based on ICL3 region shows that long FA simulation is divided into two distinct clusters (Fig 3.27d). Also, the three short FA simulations are found in two distinct clusters where one of the clusters also contains a few snapshots from RM FA simulation.

The conformations obtained from RM FA snapshots are 3 Å and 2.8 Å distinct from FA simulations based on trans-membrane and binding site regions, respectively. However, the distinctiveness may either indicate a different conformational rearrangement of the receptor or a structural deformation in some part of the receptor. From the energy profiles, RM simulations were as stable as FA simulations; however, there was a significant difference in electrostatic energy value, which is 13,000 kcal/mol, between two models. RMSF profiles indicated a structural mobility in agreement with FA model. Still, it would be difficult to conclude the nature of the distinctiveness by looking at the energy and RMSF profiles, alone. Thus, a structural comparison was performed and the results are presented in the next section.

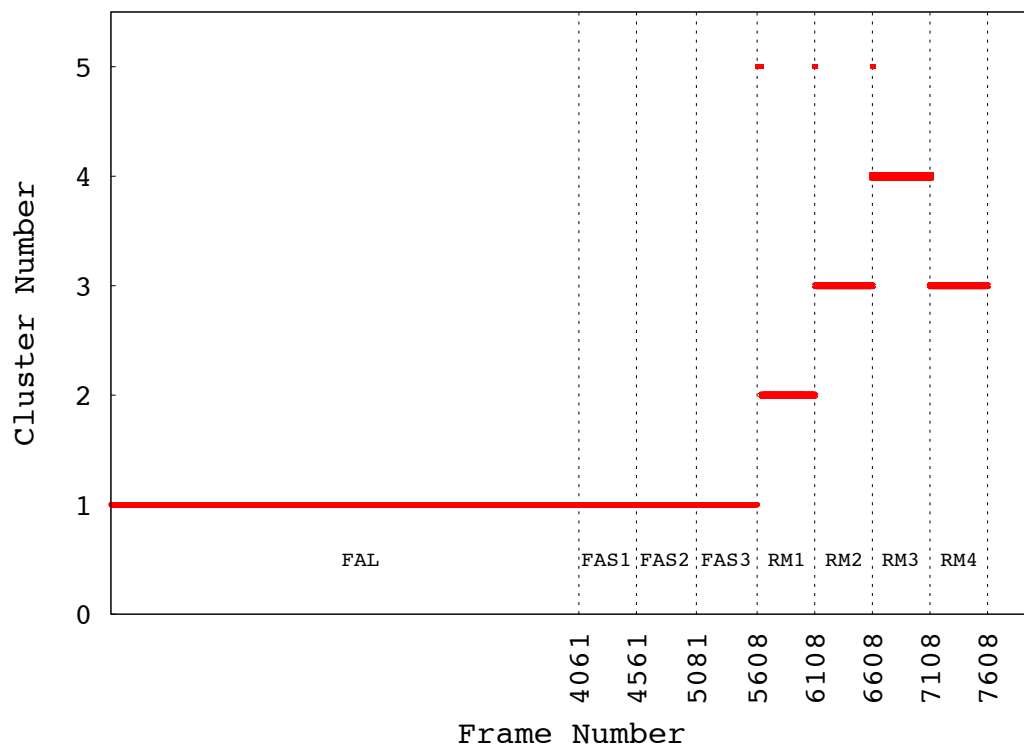




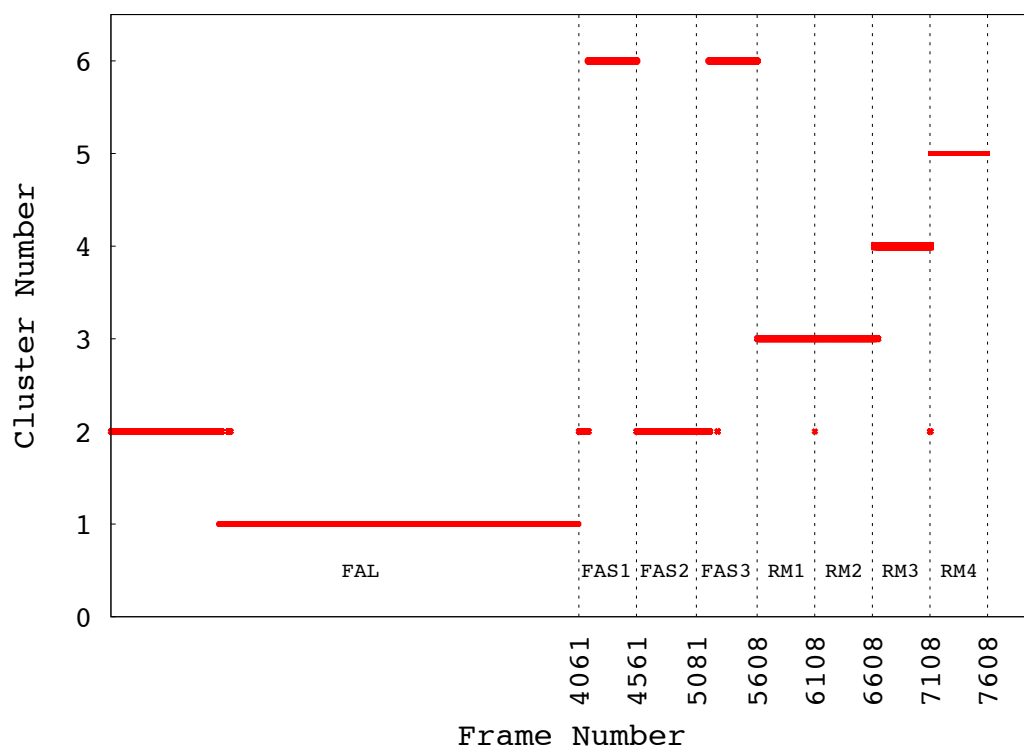
(a)



(b)



(c)



(d)

Figure 3.27. Cluster profile of all simulations (FA and RM). (a) based on core RMSD = 5 Å, (b) transmembrane RMSD = 3 Å, (c) bindingsite RMSD = 2.8 Å, (d) ICL3 RMSD = 5.9 Å.

### 3.11. Structural Agreement of 2RH1 and RM Simulations

For structural comparison, 2RH1 was aligned to RM4. First, two structures were aligned according to the core region and RMSD value was calculated for the whole structure and each helix, as shown in Figure 3.28a. Then just helices were aligned to each other to see the secondary structural change in the helix itself (Figure 3.28b). Except helix 1, 2 and 5 (H1, H2, H5), all four helices of RM4 are in a good agreement with 2RH1 with acceptable RMSD values (1.54 Å - 2.46 Å). H1 and H5 could not keep their original structures, which have broken around the middle. On the other hand, helices 3, 4, 6 and 7 were well preserved with RMSD values smaller than 2.5 Å.

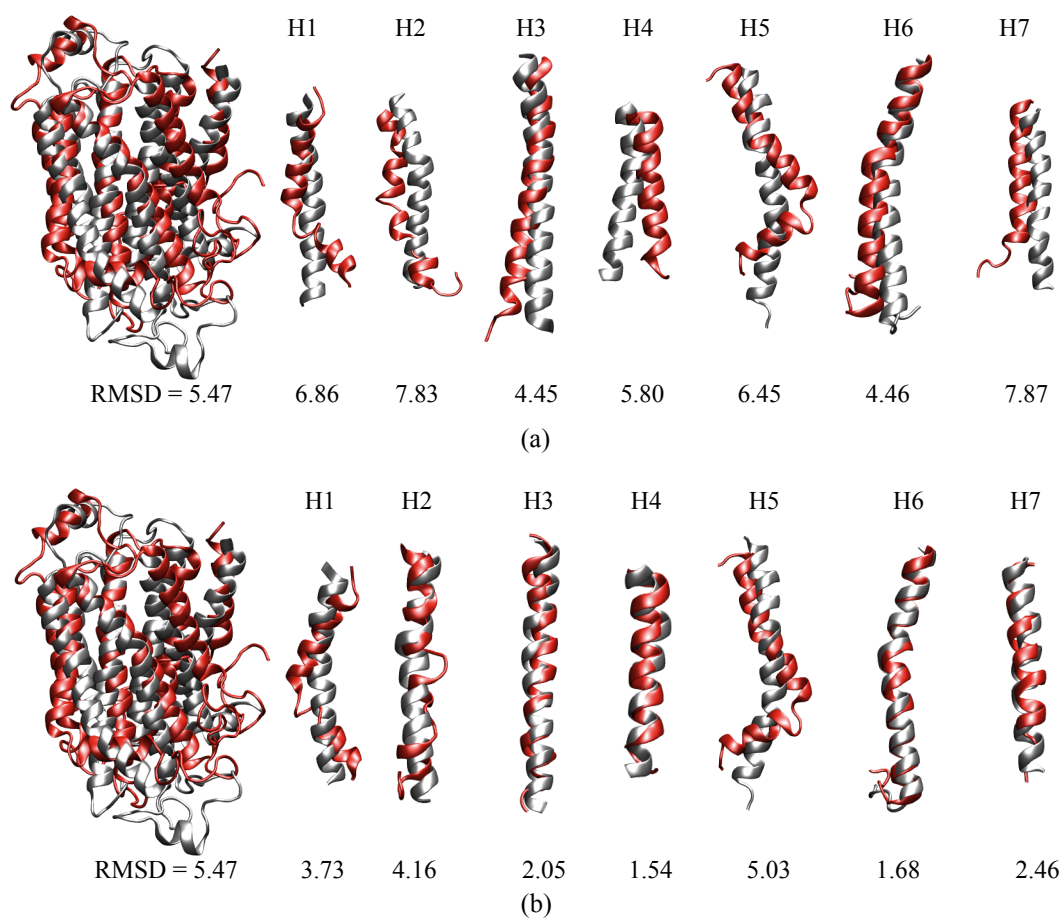
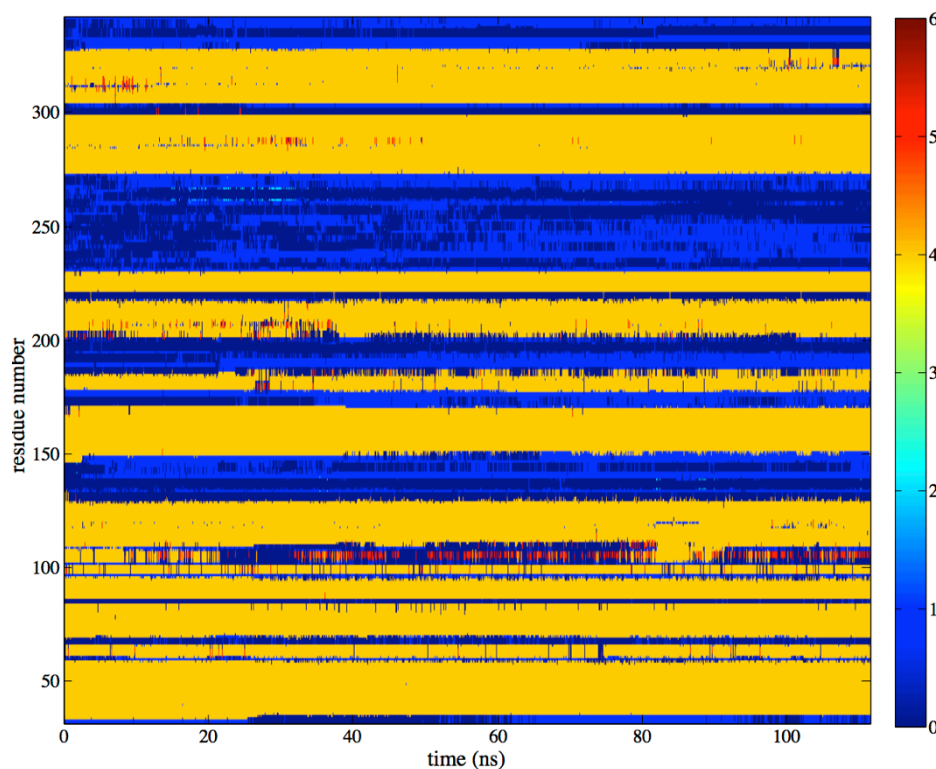


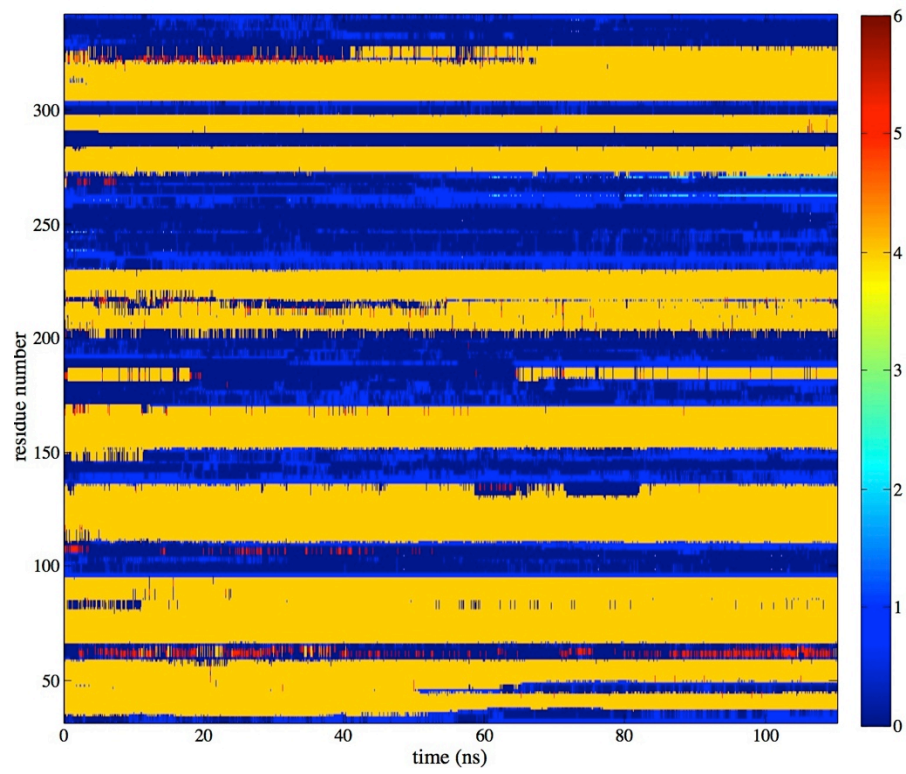
Figure 3.28. Alignments of RM4's and 2RH1 according to the core region (a) and the helices (b). All values are in Angstrom unit.

In addition to the comparison of the last snapshots, the secondary structural variations in each RM model are plotted as a function of time (Figure 3.29). The color code explains the structure type as: 0 = T (turn), 1 = C (coil), 2 = B (isolated bridge), 3 = E (beta sheet), 4 = H (alpha helix), 5 = G (3-10 helix), 6 = I (pi helix). Also, the secondary structure profile of 800 ns long FA simulation is given in Figure 3.30 for comparison. The structural deformations of helices could be clearly seen in RM simulations. Especially, helix 6 is shortened in RM1 model since its lower part has lost its helical motif. In RM2 model, helix 6 is broken in the middle into two short helices. In all RM models, helix 5 is broken in the middle. Also, helix 1 has some serious deformations in RM2, RM3 and RM4 models. However, in all secondary structure profiles, the deformations already existed at the beginning of the simulation. Thus, these deformations most likely occurred during the recovery of the atomistic detail. The Molecular Dynamics simulation simply could not fix the deformations in the secondary structure due to necessity of large energy jumps, however was able to preserve the main tertiary contacts between helices and those between the receptor and its surroundings.

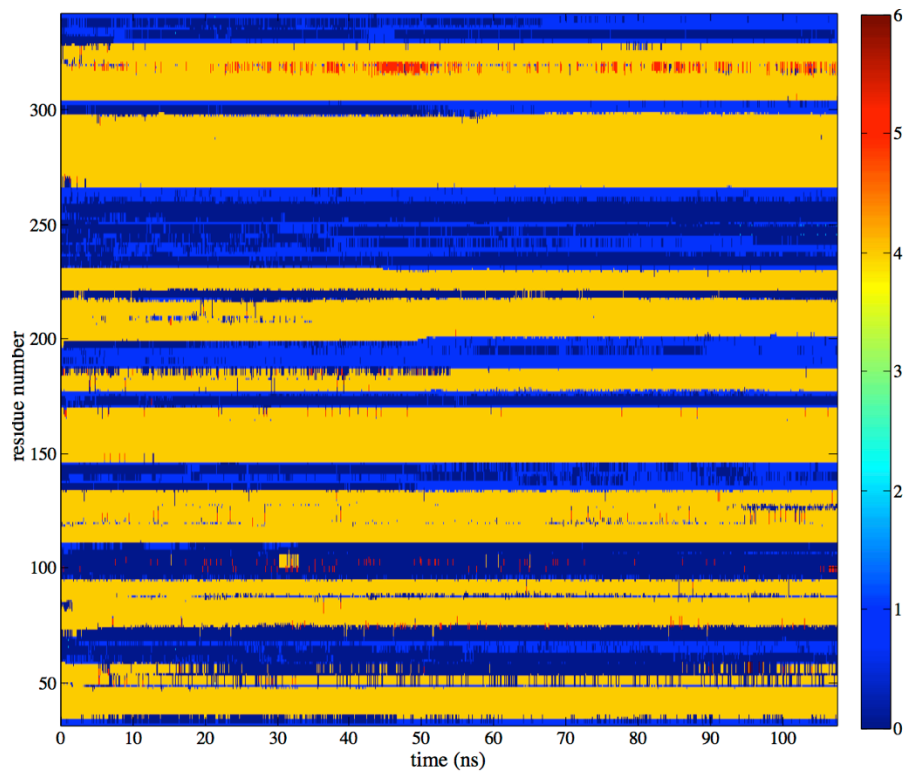


(a)

Figure 3.29

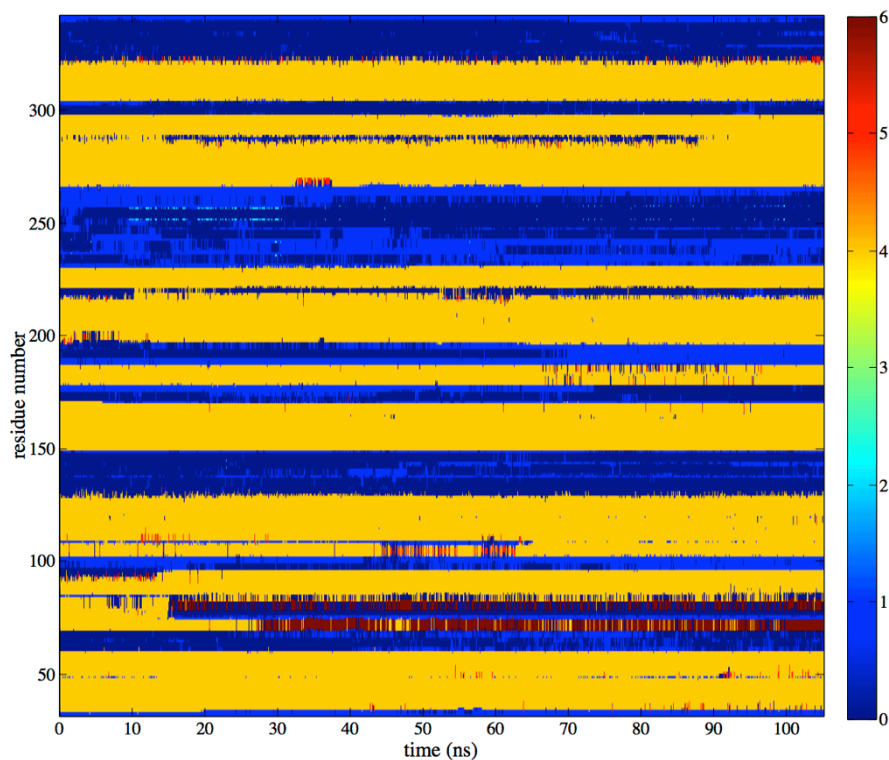


(b)



(c)

Figure 3.29



(d)

Figure 3.29. (cont'd) The structural variations of RM1 (a), RM2 (b), RM3 (c), RM4 (d). Color bar on the right is given as: 0 = T (turn), 1 = C (coil), 2 = B (isolated bridge), 3 = E (beta sheet), 4 = H (alpha helix), 5 = G (3-10 helix), 6 = I (pi helix).

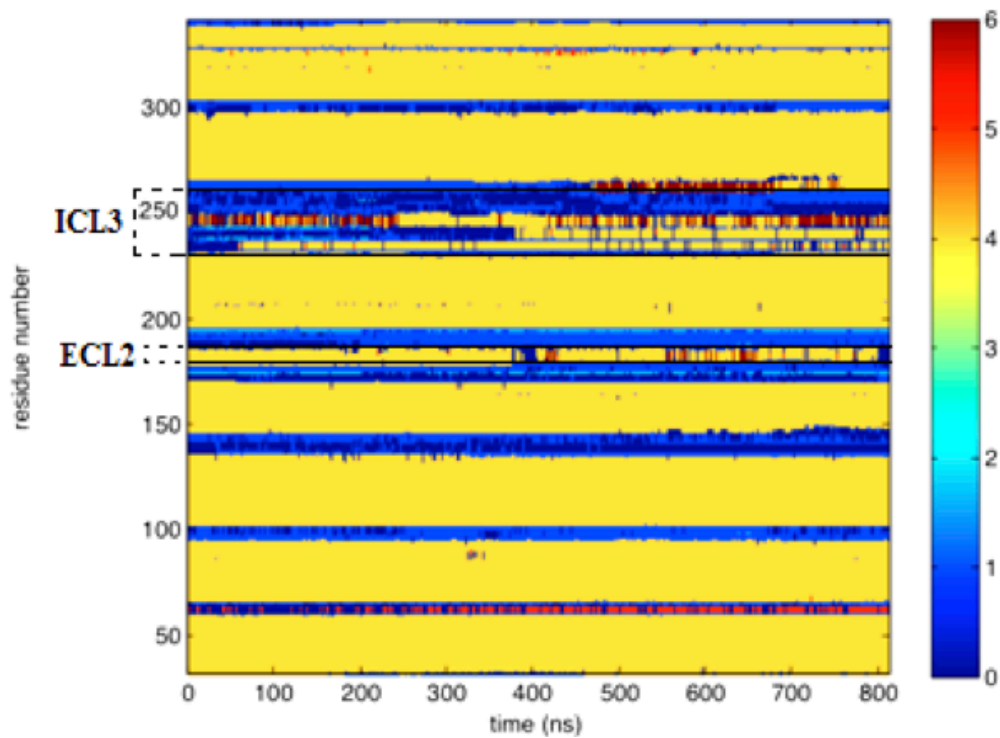
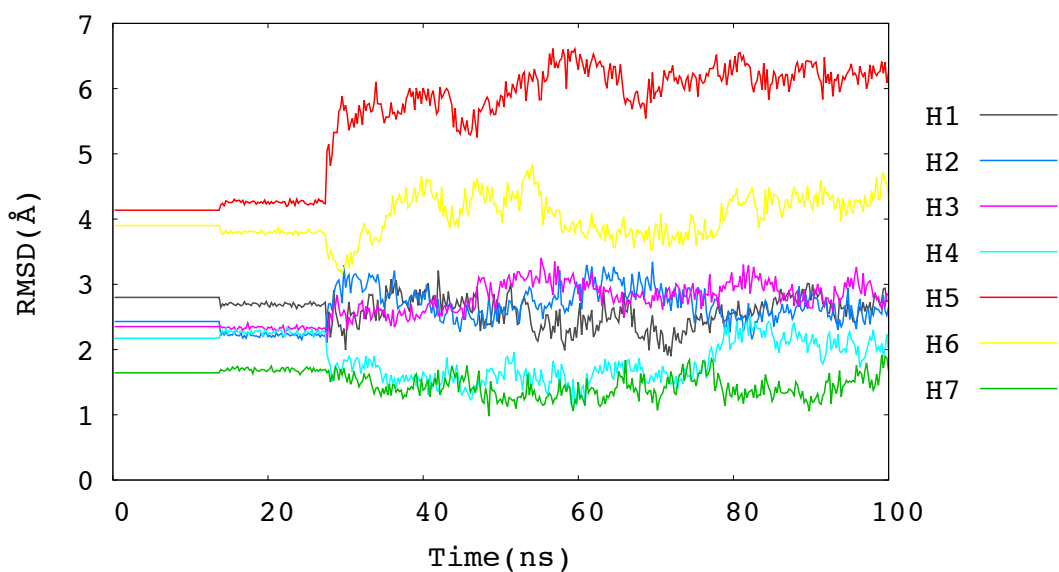


Figure 3.30. The structural variations of FAL simulation.

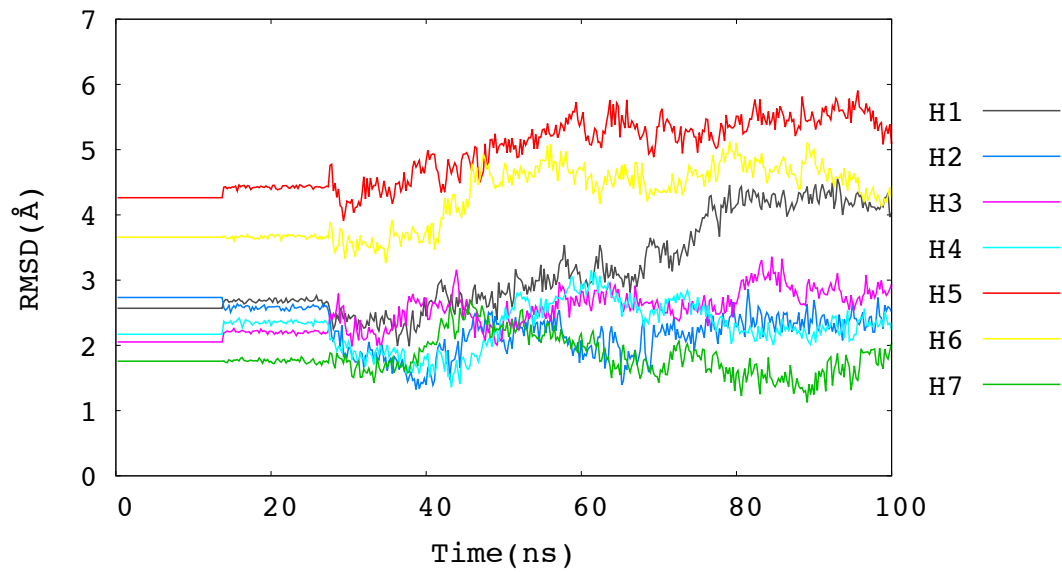
### 3.12. RMSD of Helices of RM Structures

To reveal the time at which these structural deformations occurred, the RMSD profiles of each helices in reverse-mapped structures are calculated with respect to the crystal structure 2rh1. Accordingly, most helices deformed just after the preparation phase, not during the reverse mapping procedure. Deformation of H5 is similar in all RM structures. Just after reverse mapping, H5's RMSD value is around 4 Å and towards the end of the simulation, it increased up to the 6 Å. Generally, the initial RMSD value determines the final state of the helical structure. If the initial RMSD is greater than 2.5 Å, it becomes more difficult to recover during the simulation. Exceptionally, RMSD value of H6 decreased from around 4 Å to around 2 Å in RM3 and RM4 simulations.

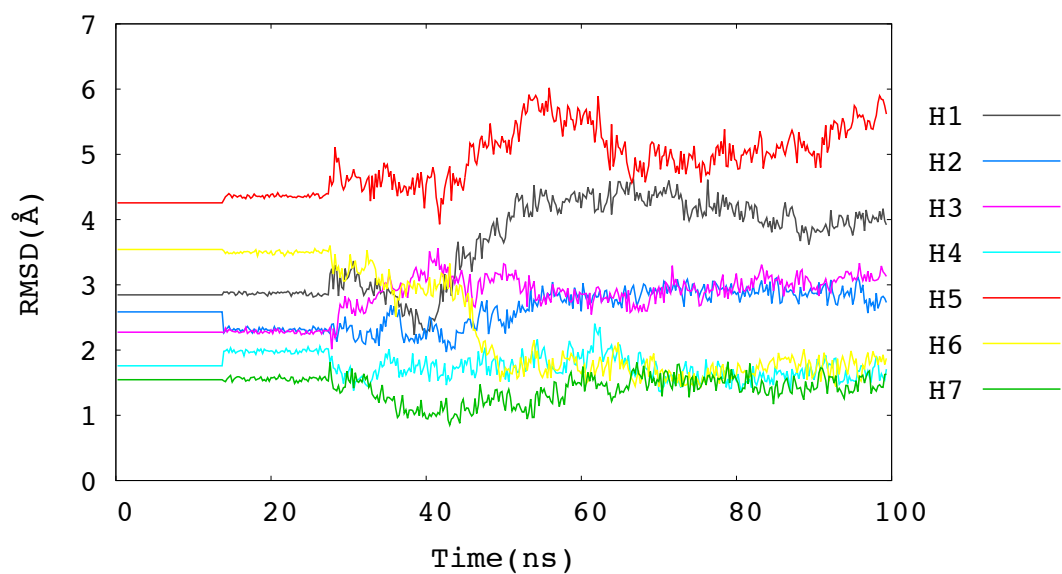


(a)

Figure 3.31



(b)



(c)

Figure 3.31



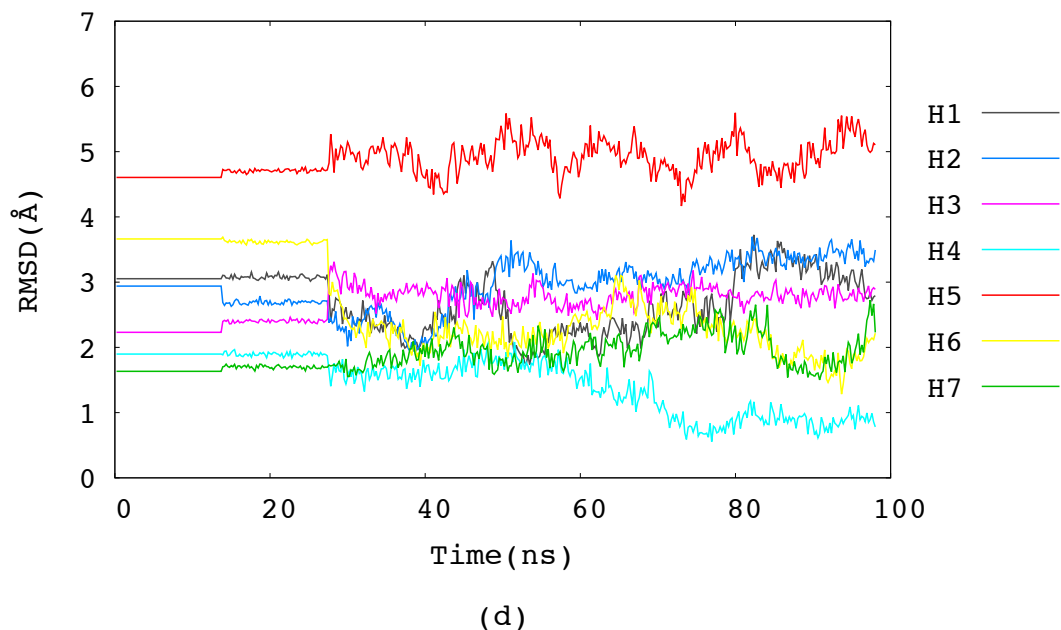


Figure 3.31. (cont'd) RMSD profiles of helices during the RM1 (a), RM2 (b), RM3 (c), RM4 (d) simulations.

### 3.13. Binding Pocket of $\beta_2$ AR

To reveal the distinct conformers of the binding site region obtained from reverse-mapped structures, five representative structures taken from five different clusters are aligned to the crystal structure 2RH1 (colored in gray) as shown in Figure 3.32. Clustering was performed based on binding site regions using all snapshots of the simulations including FAL, three FAS and 4 RM. As a result, with a RMSD cutoff of 2.8 Å, five clusters were obtained as previously illustrated in 3.27c. RMSD values of the binding site regions with respect to 2RH1 are calculated as 1.3, 3.0, 3.5, 3.7, 3.5 for FAL, RM1, RM1, RM2, RM3 simulation snapshots which were taken from clusters closest to the centroid.

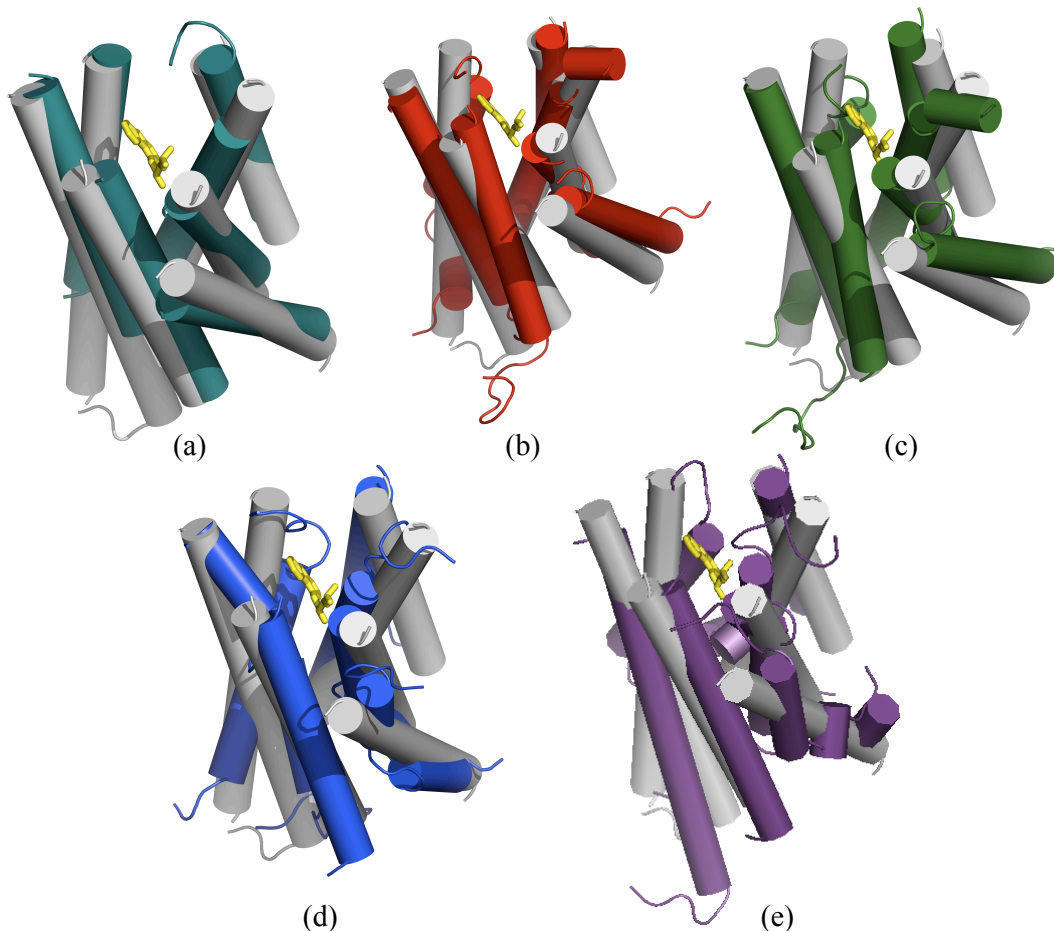
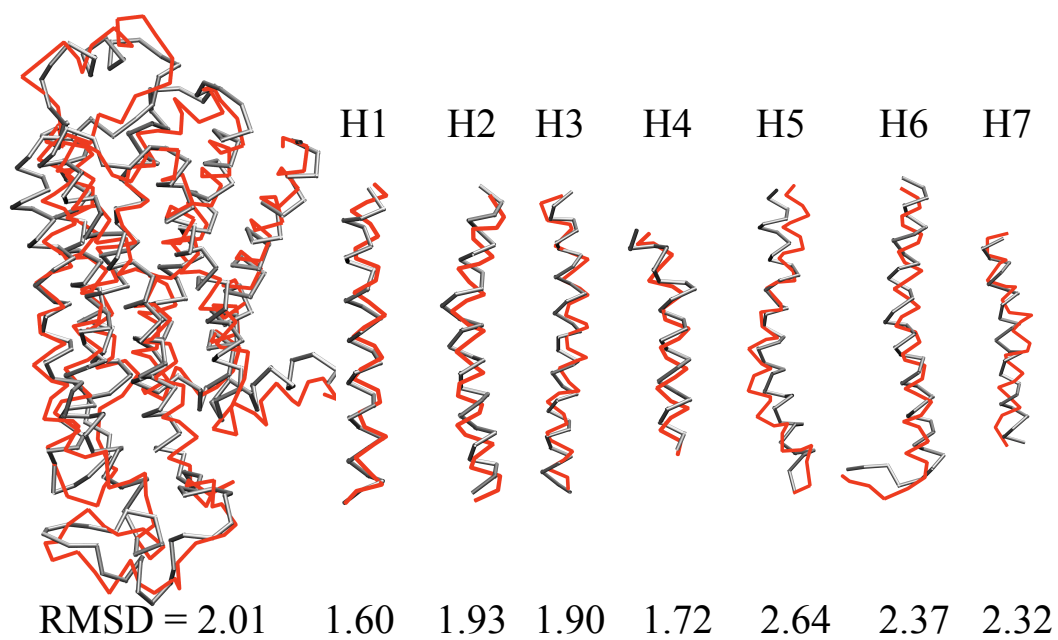


Figure 3.32. Five representative structures taken from five distinct clusters aligned to 2RH1 (shown in gray). RMSD of the binding site region is FAL (a), RM1 (b), RM1 (c), RM2 (d), RM3 (e)

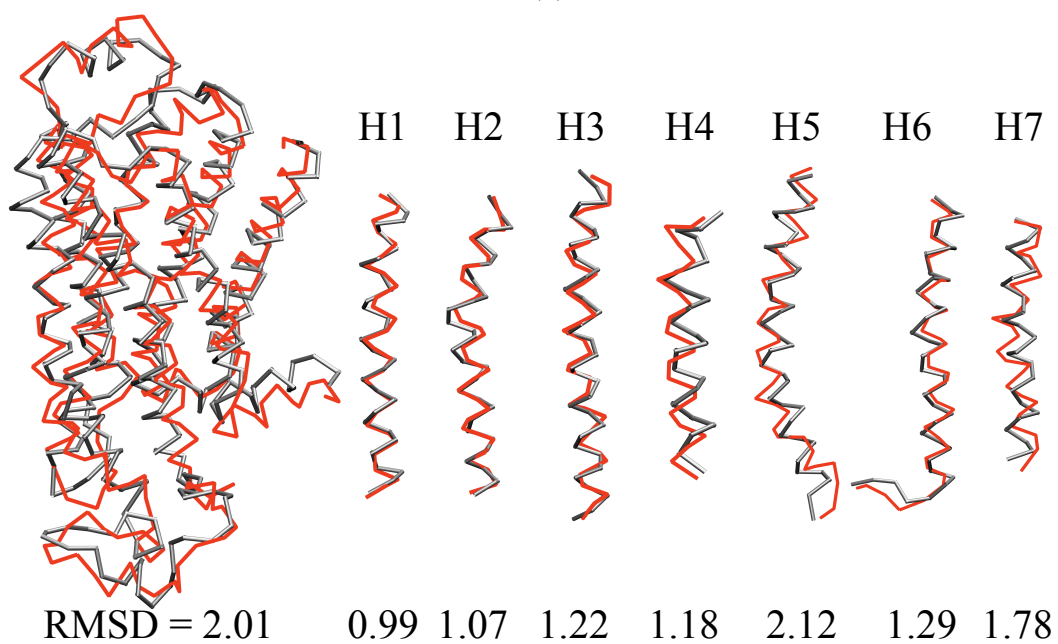
### 3.14. Structural Agreement of CG Model and 2RH1

All three CG trajectory's (at 290, 310 and 323 K) first and last snapshots were aligned with 2RH1 (colored in gray) to understand the effect of CG modeling procedure on structural deformations. The structures were aligned based on the trans-membrane region and RMSD value was calculated for whole structure and each helix as can be seen in Figure 3.36a, then just helices were aligned with each other as shown in Figure 3.36b for 290 K CG simulation's first snapshot. In Figure 3.37, alignment of CG simulation's last snapshot and 2RH1 is shown. For 310 K and 323 K simulations, the alignments with 2RH1 can be seen in Figure 3.38, 3.39 and Figure 3.40, 3.41 respectively. All RMSD values of the alignments are tabulated in Table

3.8. The least deformation on the helices is observed in 323 K simulation.

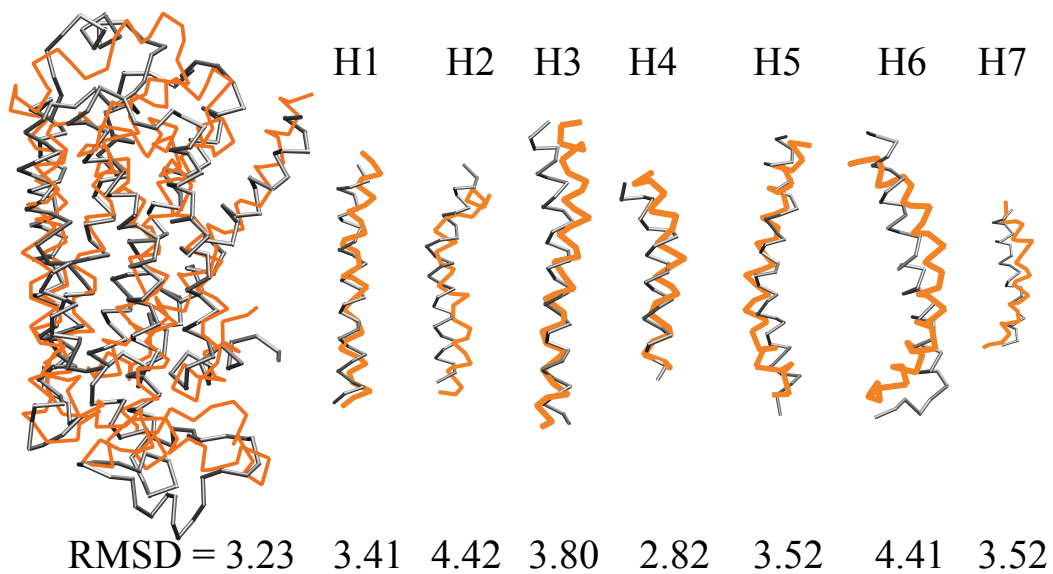


(a)

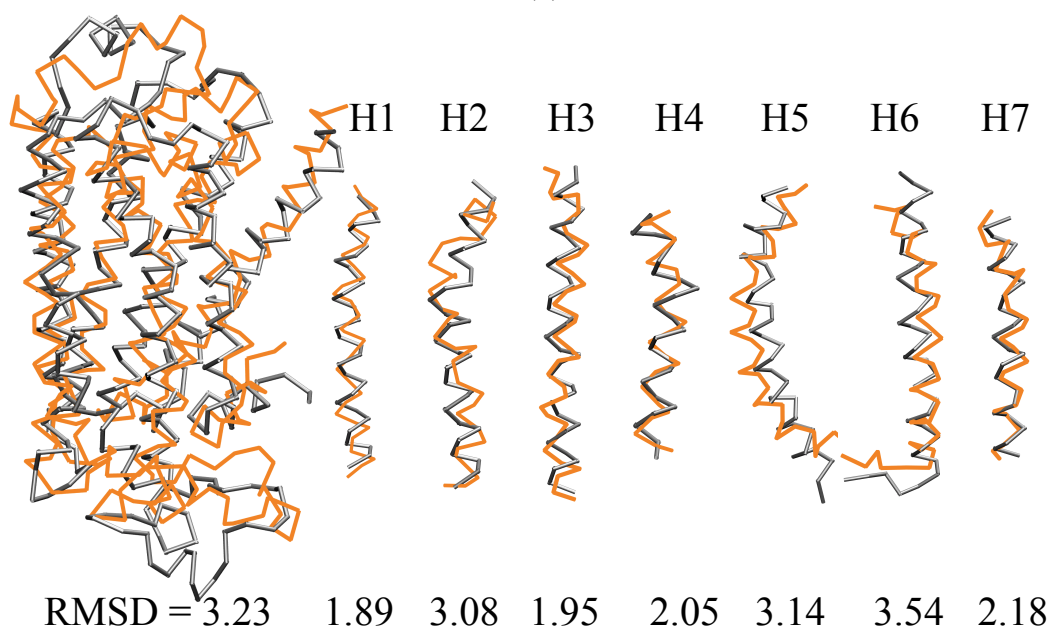


(b)

Figure 3.33. Alignments of 290 K CG trajectory's first snapshot and 2RH1 according to the transmembrane region (a) and the helices (b).

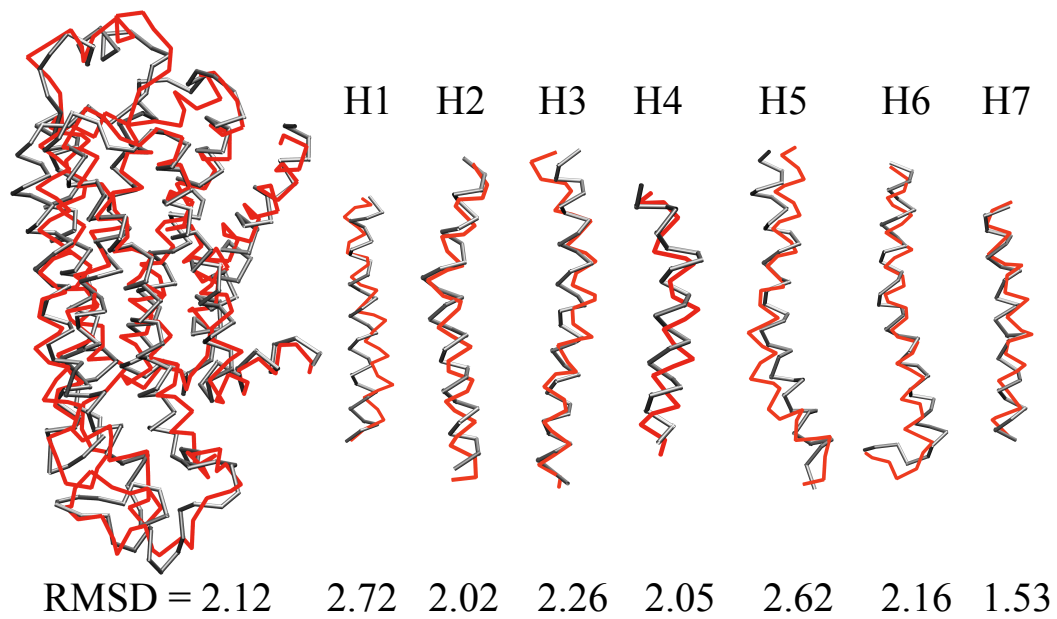


(a)

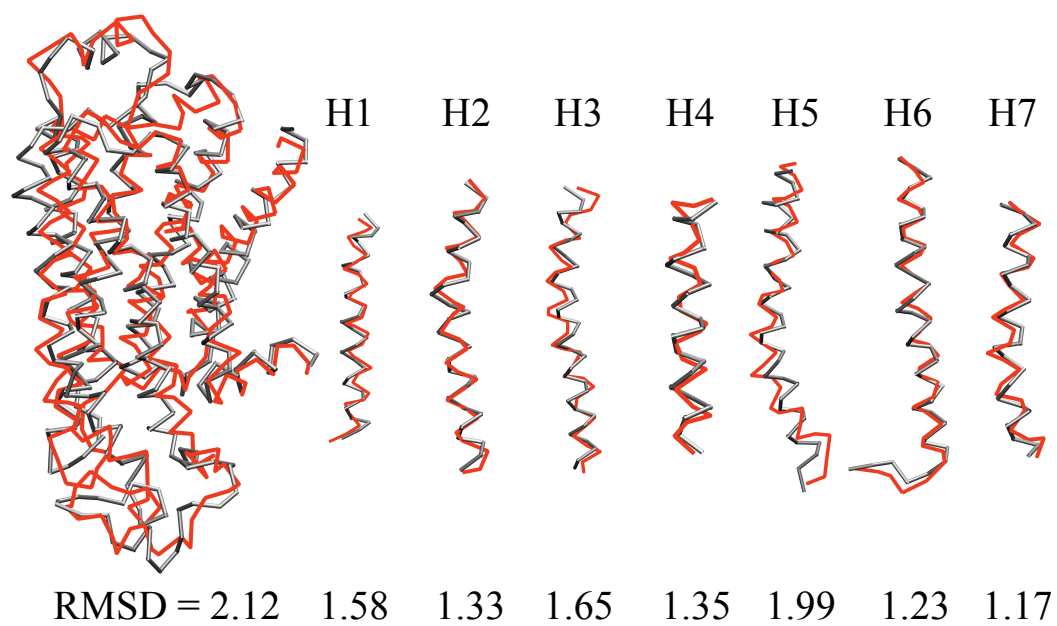


(b)

Figure 3.34. Alignments of 290 K CG trajectory's last snapshot and 2RH1 according to the transmembrane region (a) and the helices (b).

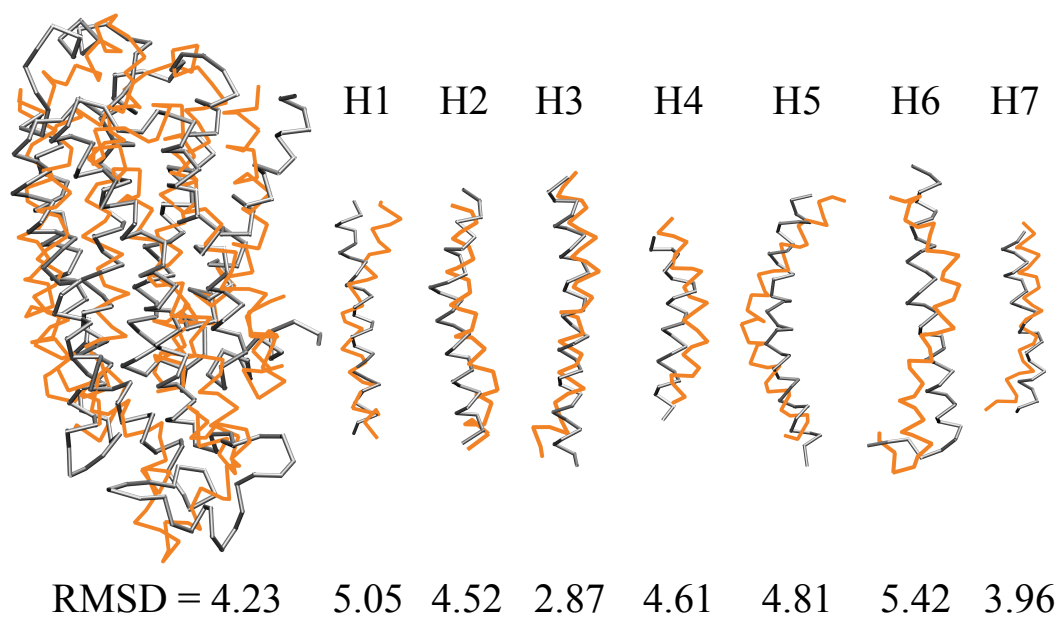


(a)

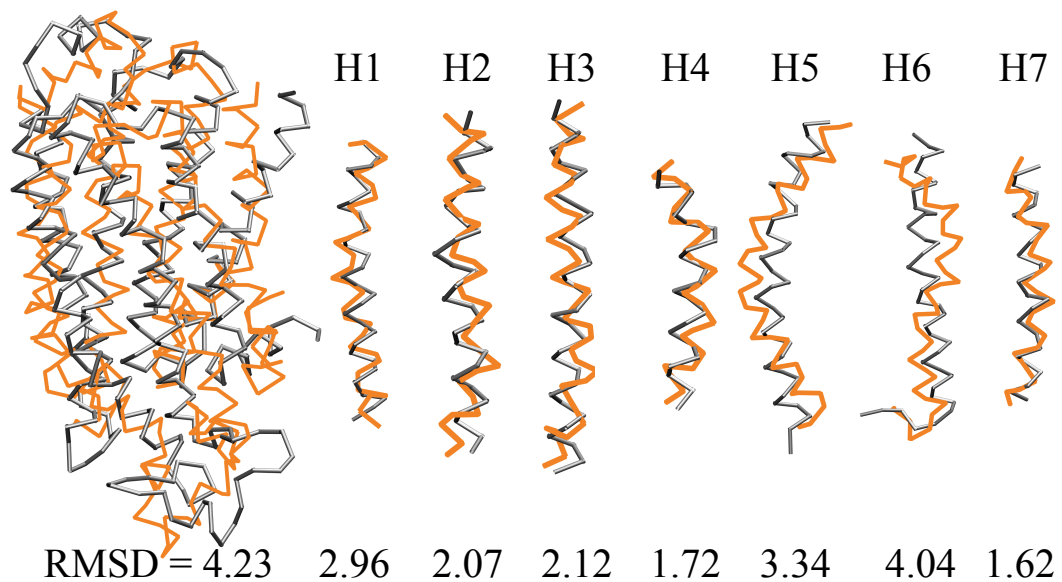


(b)

Figure 3.35. Alignments of 310 K CG trajectory's first snapshot and 2RH1 according to the transmembrane region (a) and the helices (b).

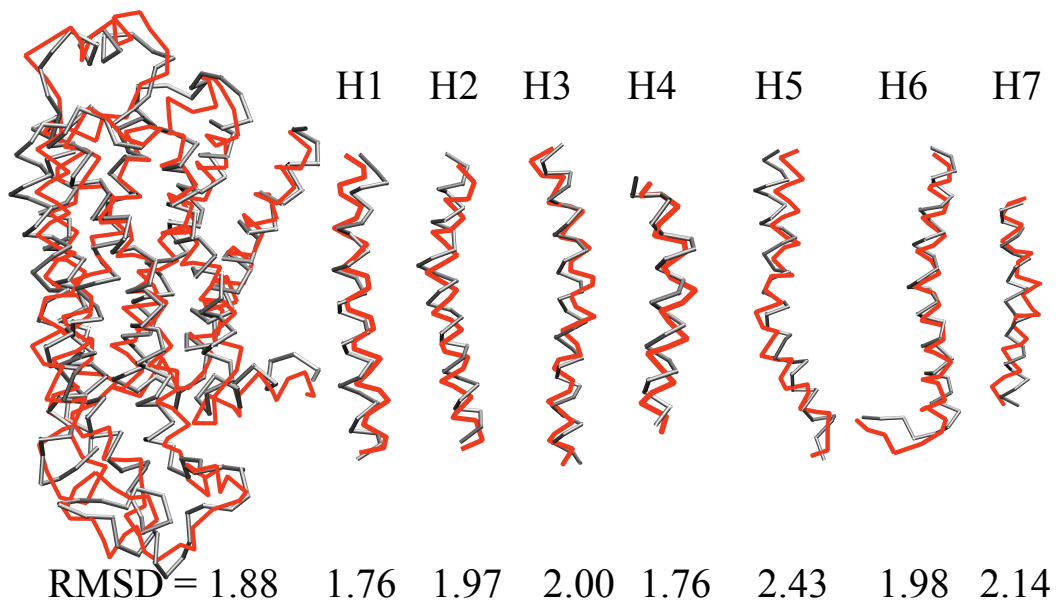


(a)

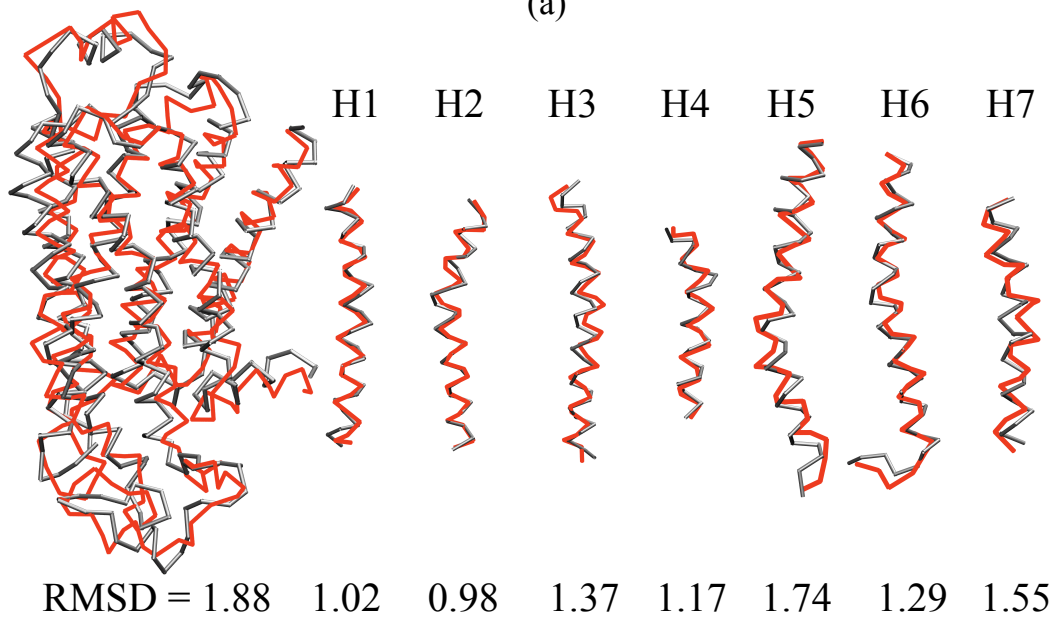


(b)

Figure 3.36. Alignments of 310 K CG trajectory's last snapshot and 2RH1 according to the transmembrane region (a) and the helices (b).

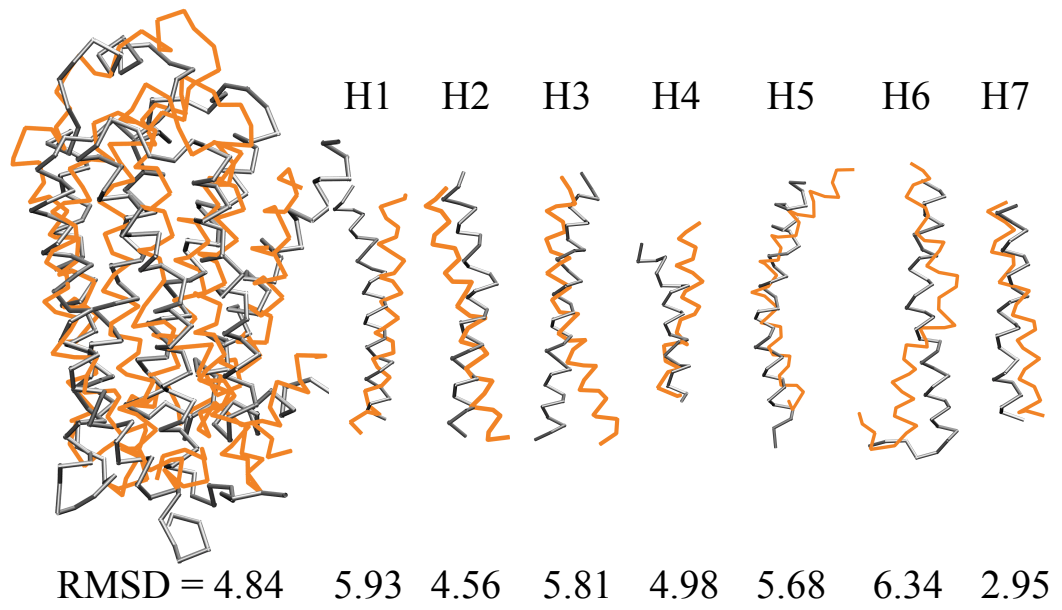


(a)

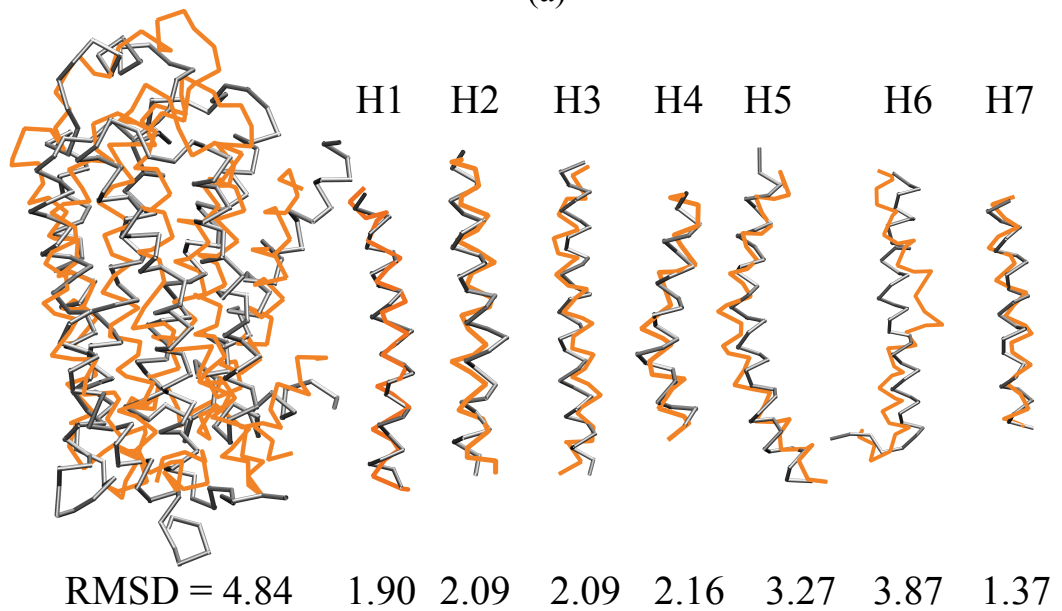


(b)

Figure 3.37. Alignments of 323 K CG trajectory's first snapshot and 2RH1 according to the transmembrane region (a) and the helices (b).



(a)



(b)

Figure 3.38. Alignments of 323 K CG trajectory's last snapshot and 2RH1 according to the transmembrane region (a) and the helices (b).

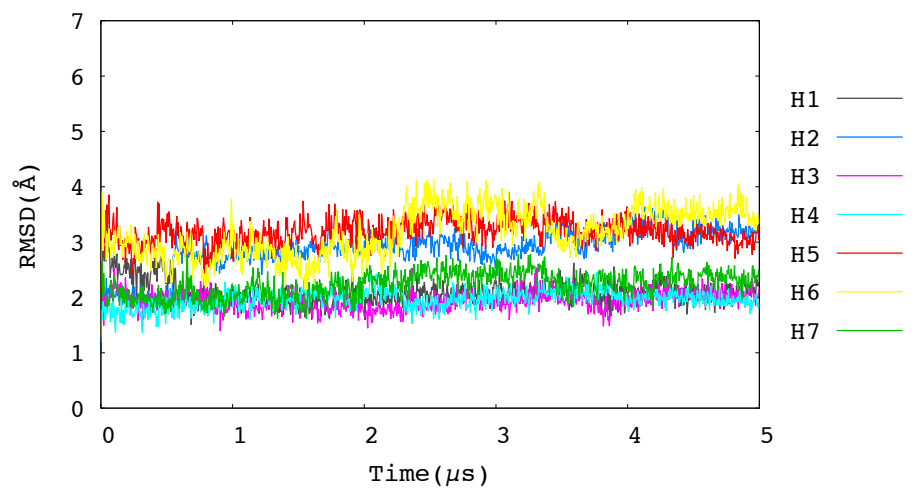


Table 3.7. RMSD values of alignments based on transmembrane and helices. RMSD-1 shows the alignment according to the core region, RMSD-2 shows the helices alignments. RMSD-2 values higher than 3.0 Å are typed in bold characters.

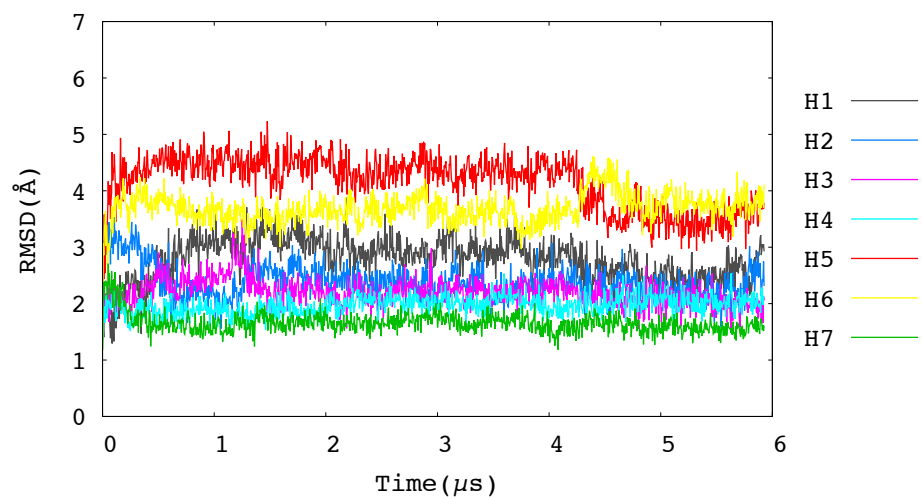
	290 K				310 K				323 K			
	FIRST		LAST		FIRST		LAST		FIRST		LAST	
	RMSD1	RMSD2	RMSD1	RMSD2	RMSD1	RMSD2	RMSD1	RMSD2	RMSD1	RMSD2	RMSD1	RMSD2
<b>H1</b>	1.6	0.9	3.4	1.9	2.7	1.6	5.1	2.9	1.8	1.0	5.9	1.9
<b>H2</b>	1.9	1.1	4.4	<b>3.1</b>	2.0	1.3	4.5	2.1	1.9	0.9	4.6	2.1
<b>H3</b>	1.9	1.2	3.8	1.9	2.3	1.7	2.9	2.1	2.0	1.4	5.8	2.1
<b>H4</b>	1.7	1.2	2.8	2.1	2.1	1.4	4.6	1.7	1.7	1.2	4.9	2.2
<b>H5</b>	2.6	2.1	3.5	<b>3.1</b>	2.6	1.9	4.8	<b>3.3</b>	2.4	1.7	5.7	<b>3.3</b>
<b>H6</b>	2.4	1.3	4.4	<b>3.5</b>	2.2	1.2	5.4	<b>4.0</b>	1.9	1.3	6.3	<b>3.9</b>
<b>H7</b>	2.3	1.8	3.5	2.2	1.5	1.2	3.9	1.6	2.1	1.6	2.9	1.4

### 3.15. RMSD of Helices of CG Simulations

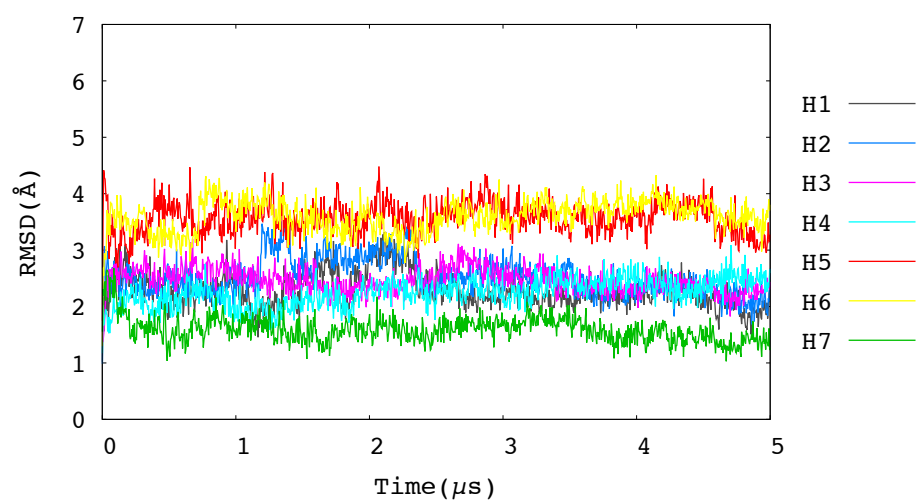
RMSD values of each helices were calculated and shown in Figure 3.39 for 290 K (a), 310 K (b), 323 K (c) CG simulations. RMSD values were calculated with respect to 2RH1. H5 and H6 have the highest RMSD values in CG simulations. RMSD values of helices fluctuate at around 2-3.5 Å, 1.5-4.5 Å, and 1.5-4 Å for 290 K, 310 K and 323 K simulations, respectively.



(a)



(b)



(c)

Figure 3.39. RMSD profiles of helices for 290 K (a), 310 K (b), 323 K (c) CG simulations.

### 3.16. RMSD of Binding-site Regions

The RMSD profile of binding-site region for all FA and RM simulations is displayed in Figure 3.40. RMSD values are calculated with respect to the initial structure, that is 2RH1. Binding site region is stable in all simulations. RMSD values fluctuate at around 1-1.5 Å for FA simulations, while RM simulations have higher RMSD values at around 3.5-4.0 Å.

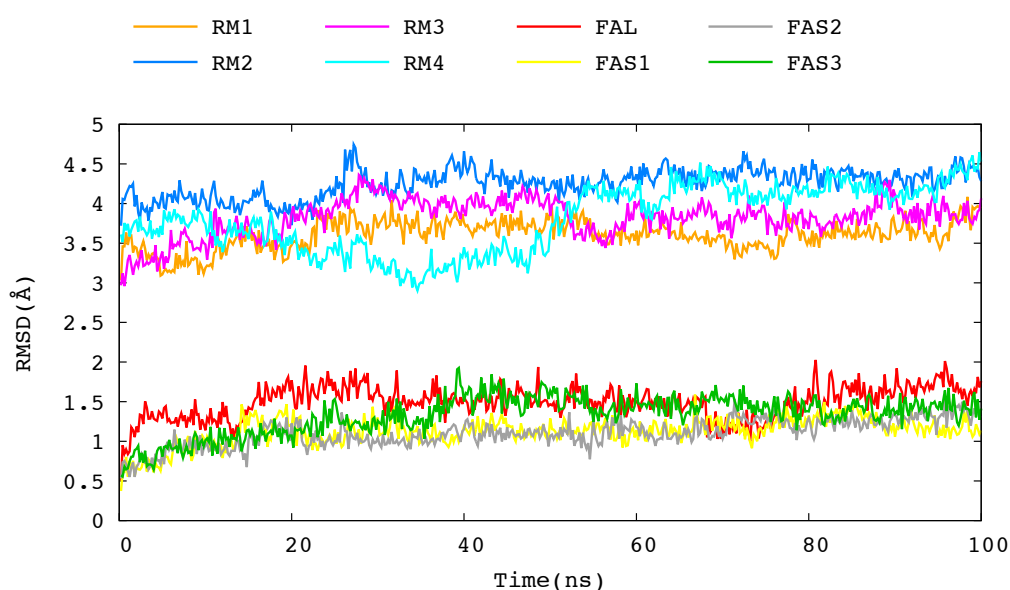


Figure 3.40. Binding-site RMSD of all FA and RM simulations.

### 3.17. Distance Between Residues D113 and S203-S207

For the CG, FA and RM simulations, the variation in the distance is calculated between critical residues, Asp113 on helix 3 and Ser203, Ser207 on helix 5. Asp113, Ser203 and Ser207 are the key residues in the binding pocket that interact with both agonists and antagonists. Distance calculation was performed based on the gamma carbon (CG) atom of Aspartic acid and gamma oxygen (OG) of Serine residue for fully atomistic model. For CG model, the side chain bead of Aspartic acid and Serine

residue was taken into account. The non-bonded distances between these residues are informative about the active/inactive state of the receptor. When the receptor is in an inactive state, the distance varies around 11 Å, while in active form the distance varies between 8 Å and 10 Å. The smaller distance value in active form is more suitable to an agonist molecule, which is smaller in size in comparison to an antagonist.

Figure 3.41a shows the time variation of distances between the residues Asp113 and Ser203-Ser207 during the FAL (800 ns) simulation. Towards the end of the simulation, both distances have increased from 10 Å to around 15 Å. At the initial stages of the simulation the receptor is in an inactive form. Towards the end, the distances becomes out of range for an inactive form and represent a “very” inactive form of the receptor.

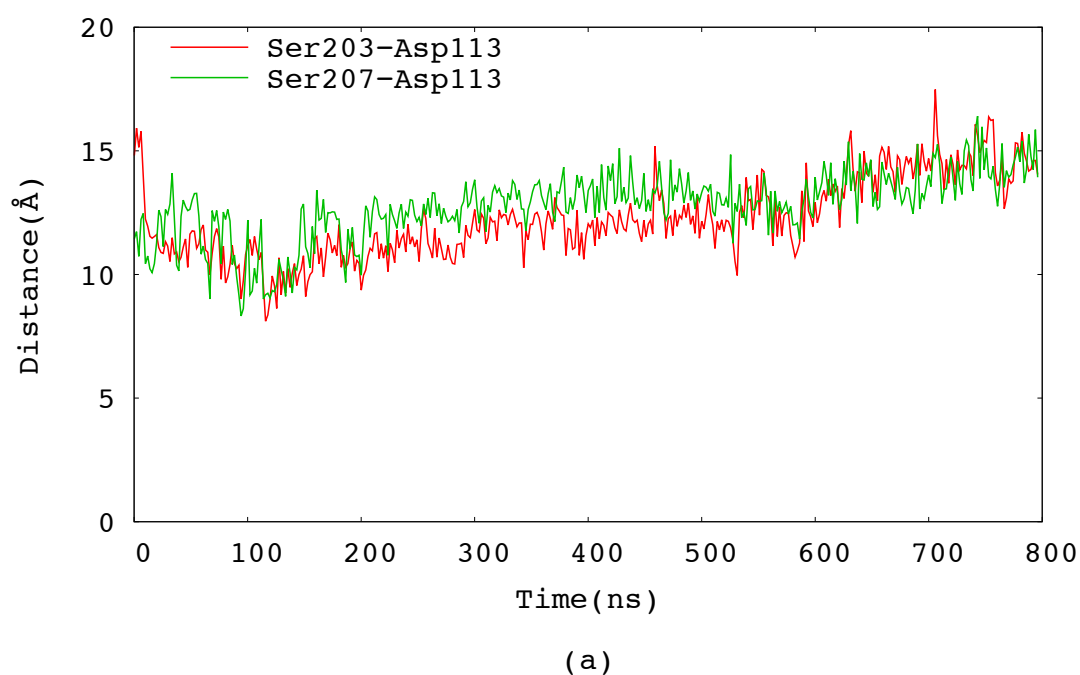
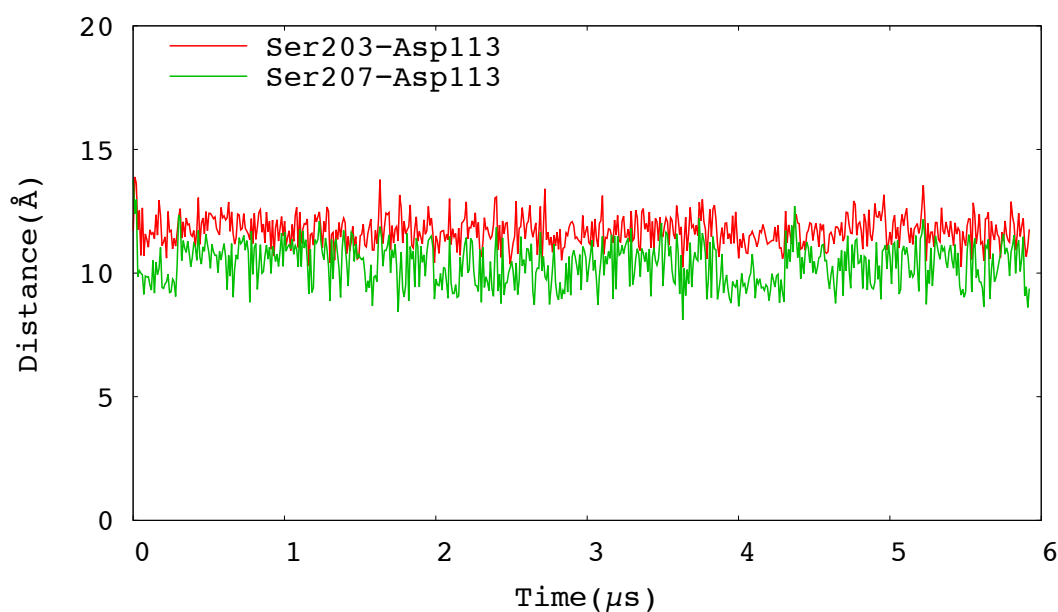


Figure 3.41



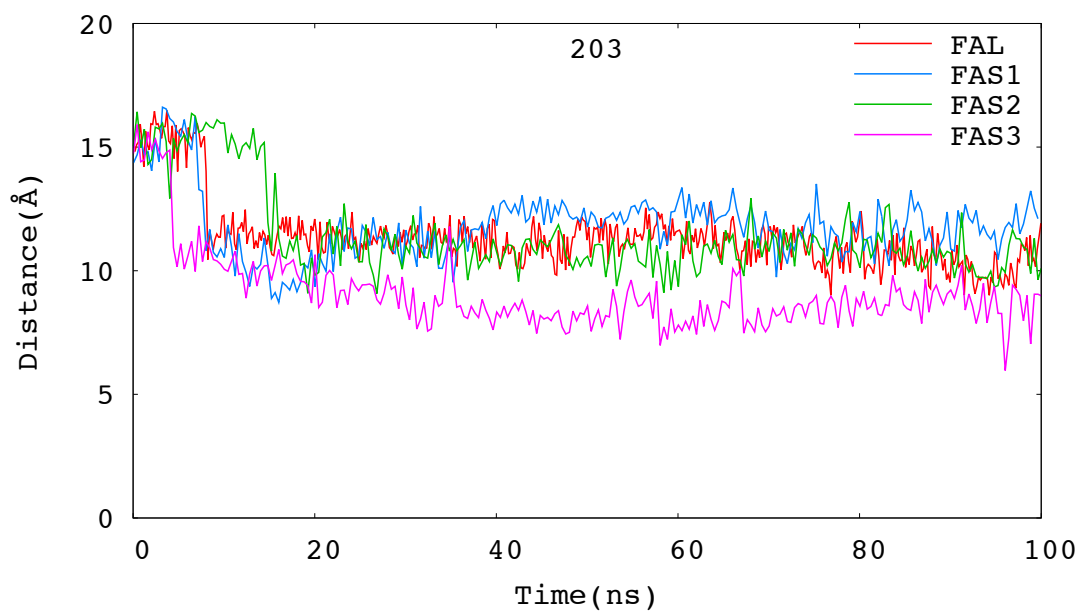
(b)

Figure 3.41.(cont'd) FAL (a) and CG (b) simulation's variation of distances with time between residues Asp113-Ser203 and Asp113-Ser207.

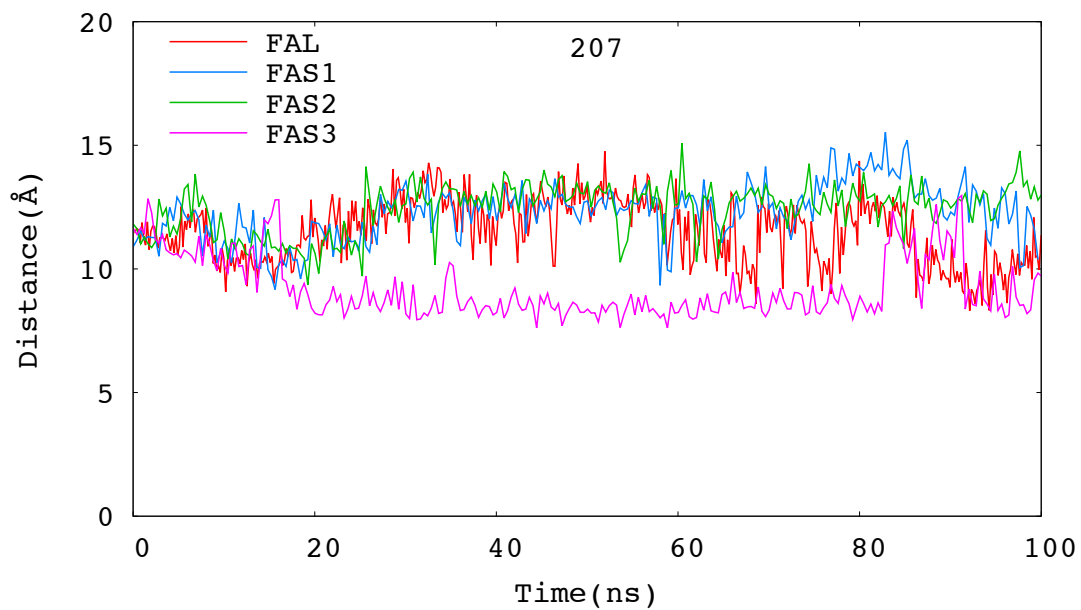
Figure 3.41b displays a similar profile for 6  $\mu\text{s}$  long CG simulation. The distances between the Asp113 and Ser203 is approximately stable around 12  $\text{\AA}$ , while the distance between Asp113 and Ser207 is stable around 11  $\text{\AA}$ . Thus, in CG simulations, a similar “very” inactive form of the receptor has not been observed, in other words, the receptor has stayed in its inactive form throughout the simulation.

Figure 3.42 shows the variation of distances between the key residues during all FA (long and shorts) simulations. FAL simulation's first 100 ns were used for comparison alongside 100 ns long FA short simulations. The distance between the residues Asp113 and Ser203 fluctuates around 10  $\text{\AA}$  towards the end of the trajectory. FAS3's distance variation is slightly below the other FA simulations below 10  $\text{\AA}$ . Furthermore, the distance profiles between the residues Asp113 and Ser207 vary between 10  $\text{\AA}$  and 15  $\text{\AA}$ , as shown in Figure 3.42b and similarly, FAS3's distance variation is slightly below the other FA simulations at around 8  $\text{\AA}$ . According to the distance variations, it can be said that FAS3 system has more closed binding cavity than other systems and may represent an active state of the

receptor.



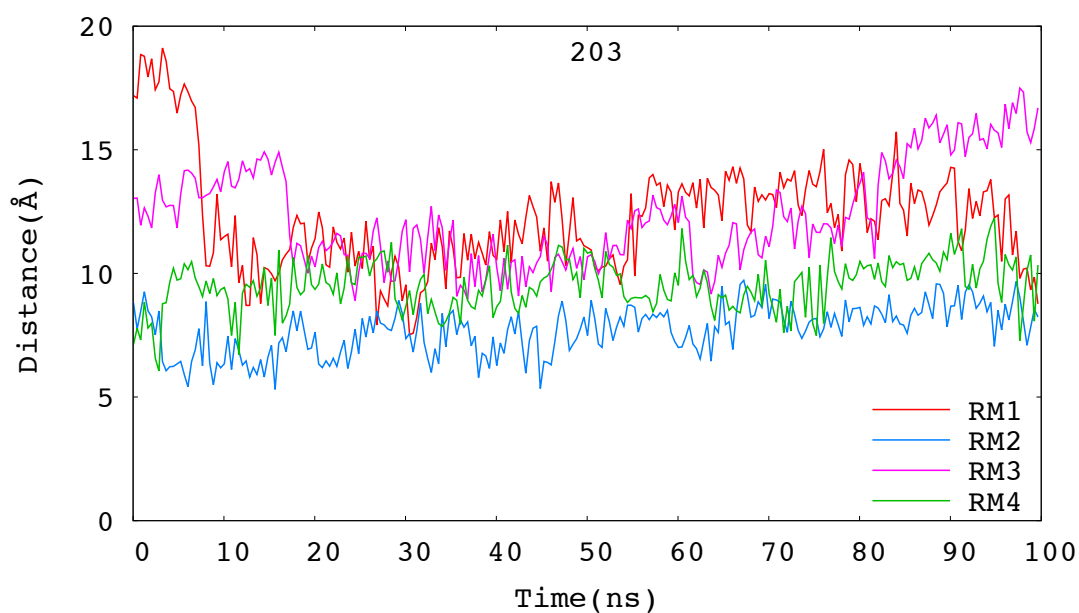
(a)



(b)

Figure 3.42. FA simulation's variation of distances with time for Asp113-Ser203 (a) and Asp113-Ser207 (b).

Figure 3.43 displays the variation of distances between the residues Asp113 and Ser203-Ser207 for RM simulations. The Asp113-Ser203 and Asp113-Ser207 distances fluctuate within a wider range in comparison to FA simulations. In RM3, both Asp113-Ser203 and Asp113-Ser207 distances increase up to 18 Å towards the end of the simulation. However, as in FA models, the open form of the binding site at the upper region does not necessitate a closed form at the lower part of the receptor. In other words, the receptor's two important regions for signaling, the binding site (upper) and the G protein binding site with ICL3 (lower) are dynamically uncorrelated.



(a)

Figure 3.43

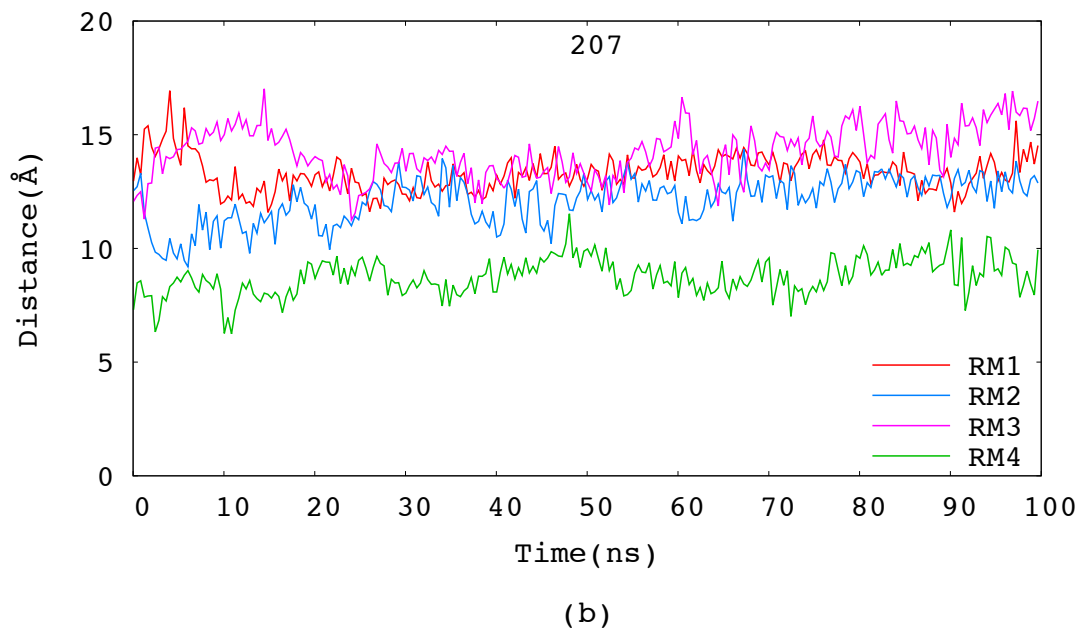
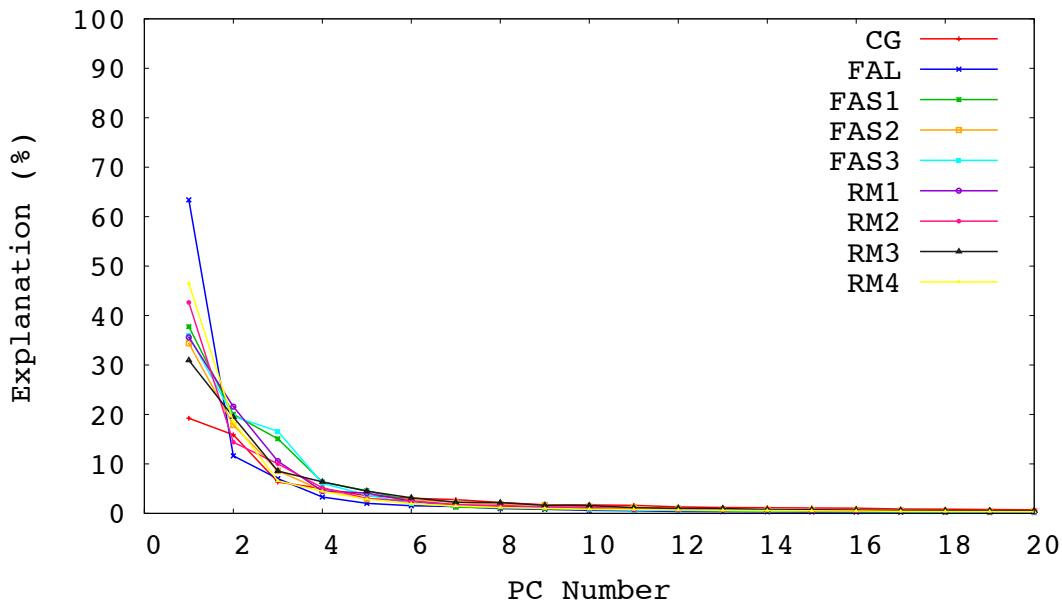


Figure 3.43. (cont'd) RM simulation's variation of distances with time for Asp113-Ser203 (a) and Asp113-Ser207 (b).

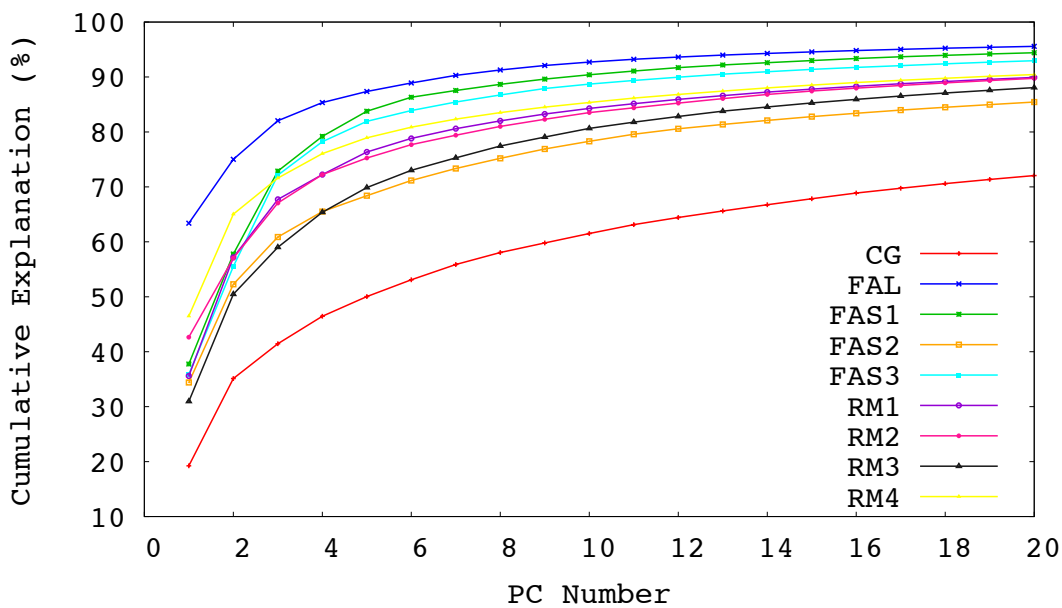
### 3.18. Principal Component Analysis (PCA)

For CG, four FA and four RM FA simulations, individual explanation values of the protein motion by the first twenty principal modes is shown mode by mode in Figure 3.44a and cumulatively in Figure 3.44b. The variance percentages of the first five principal components are also tabulated in Table 3.9. According to Figure 3.44a FAL, FAS1, FAS3 and RM4's percentages of the cumulative values of the first five principal modes explain approximately 80-85% of the protein's entire dynamics. In contrast, CG simulation's descriptive power of the first twenty modes only reaches 70%.





(a)



(b)

Figure 3.44. The explanation percentages of the protein motion of the first 20 principal modes for all simulations. (a) individual, (b) cumulative.

The first mode of the FAL explains 63% of the protein's entire motion, while the first mode of CG explains only 19%. As for the FAS simulations, their explanation percentage values are 38, 34 and 36%, which are all lower than the FAL simulation. The low values are the result of the insufficient length of the simulation. RM simulations have similar explanation values for the first mode, which are as 36, 43,

31, and 47%. Although 50% of the protein motion can be explained by the first five principal modes in CG model, in RM models this value reaches 76% for RM1, 75% for RM2, 70% for RM3, and 79% for RM4 which are comparable with FA long and short simulations. According to these results, it can be considered that RM simulations give reasonable information about the protein dynamics.

Table 3.8. Percentage of the total motion explained by first five principal components of each simulations.

Simulations	Principal Component Number	%Explanation Individual	%Explanation Cumulative	Simulations	Principal Component Number	%Explanation Individual	%Explanation Cumulative
CG	PC1	<b>19.23</b>	19.23	RM1	PC1	<b>35.59</b>	35.59
	PC2	15.89	35.12		PC2	21.58	57.17
	PC3	6.33	41.45		PC3	10.55	67.72
	PC4	5.00	46.45		PC4	4.52	72.24
	PC5	3.58	<b>50.03</b>		PC5	4.10	<b>76.34</b>
FAL	PC1	<b>63.39</b>	63.39	RM2	PC1	<b>42.66</b>	42.66
	PC2	11.63	75.02		PC2	14.39	57.05
	PC3	7.02	82.04		PC3	10.02	67.07
	PC4	3.31	85.35		PC4	5.17	72.24
	PC5	2.02	<b>87.37</b>		PC5	3.01	<b>75.25</b>
FAS1	PC1	<b>37.76</b>	37.76	RM3	PC1	<b>30.96</b>	30.96
	PC2	20.02	57.78		PC2	19.50	50.46
	PC3	15.09	72.87		PC3	8.55	59.01
	PC4	6.33	79.20		PC4	6.37	65.38
	PC5	4.57	<b>83.77</b>		PC5	4.49	<b>69.87</b>
FAS2	PC1	<b>34.40</b>	34.40	RM4	PC1	<b>46.48</b>	46.48
	PC2	17.87	52.27		PC2	18.56	65.04
	PC3	8.60	60.87		PC3	6.56	71.60
	PC4	4.64	65.51		PC4	4.45	76.05
	PC5	2.88	<b>68.39</b>		PC5	2.88	<b>78.93</b>
FAS3	PC1	<b>35.90</b>	35.90				
	PC2	19.62	55.52				
	PC3	16.62	72.14				
	PC4	6.09	78.23				
	PC5	3.70	<b>81.93</b>				

Movement towards the first two principal components of the FAL, FAS1, CG and RM4 simulations is shown in Figure 3.45. The 75% of the movements of the protein in FAL simulation can be explained by just the first two principal components, while this value is only 58% for FAS1, 35% for CG, and 65% for RM4 simulation. All intracellular and extracellular loops, including ICL3, are the most mobile regions on the protein for the fully atomistic (FAL and FAS1) simulations. On the other hand, not just loop regions, but also helices are mobile in CG and RM4 simulations. Especially RM4's first PC displays larger fluctuations in the helices compared to CG and other simulations. Each principal mode takes place at the RM simulation has different dynamics than those take place in FA and CG simulations.

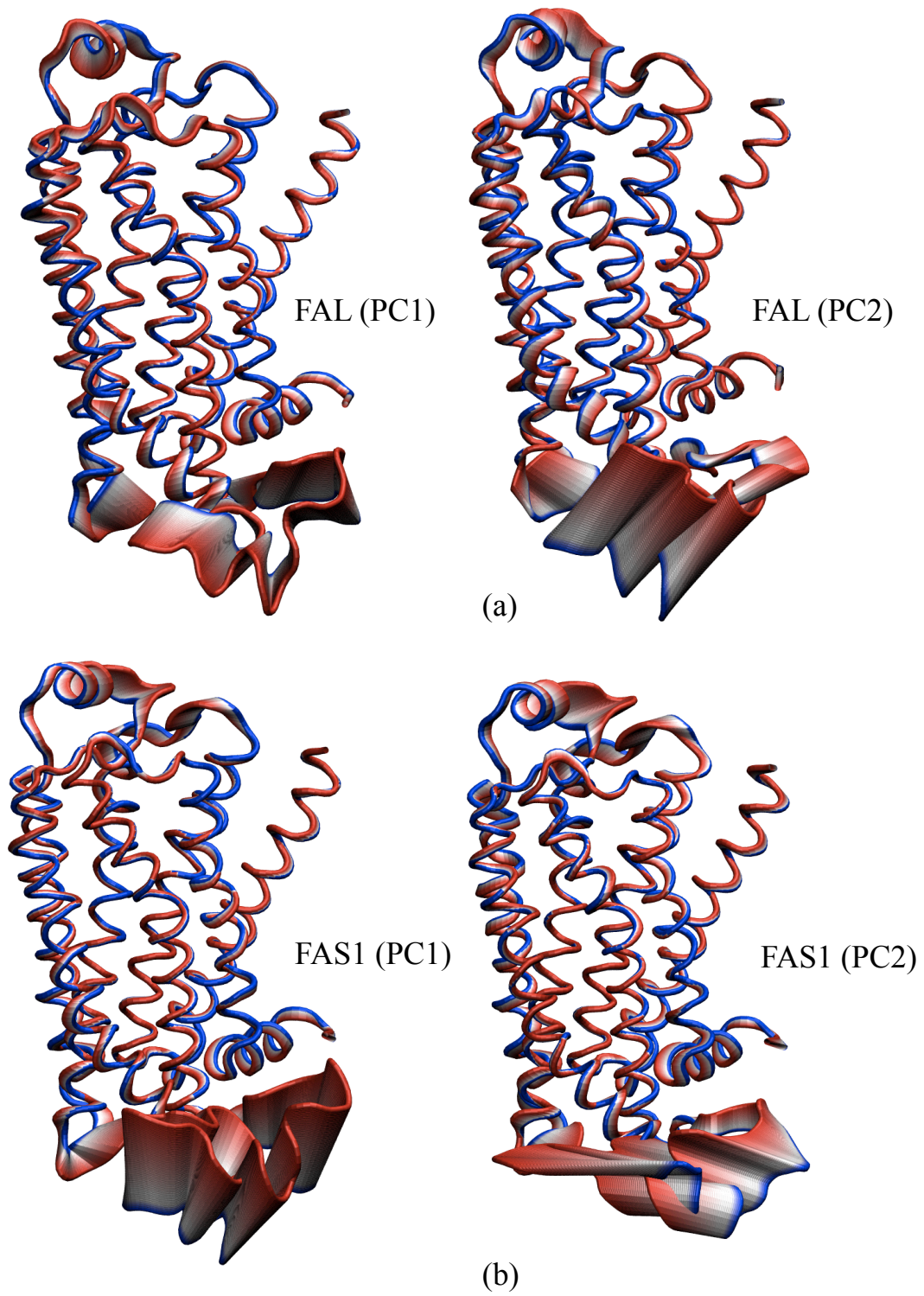


Figure 3.45

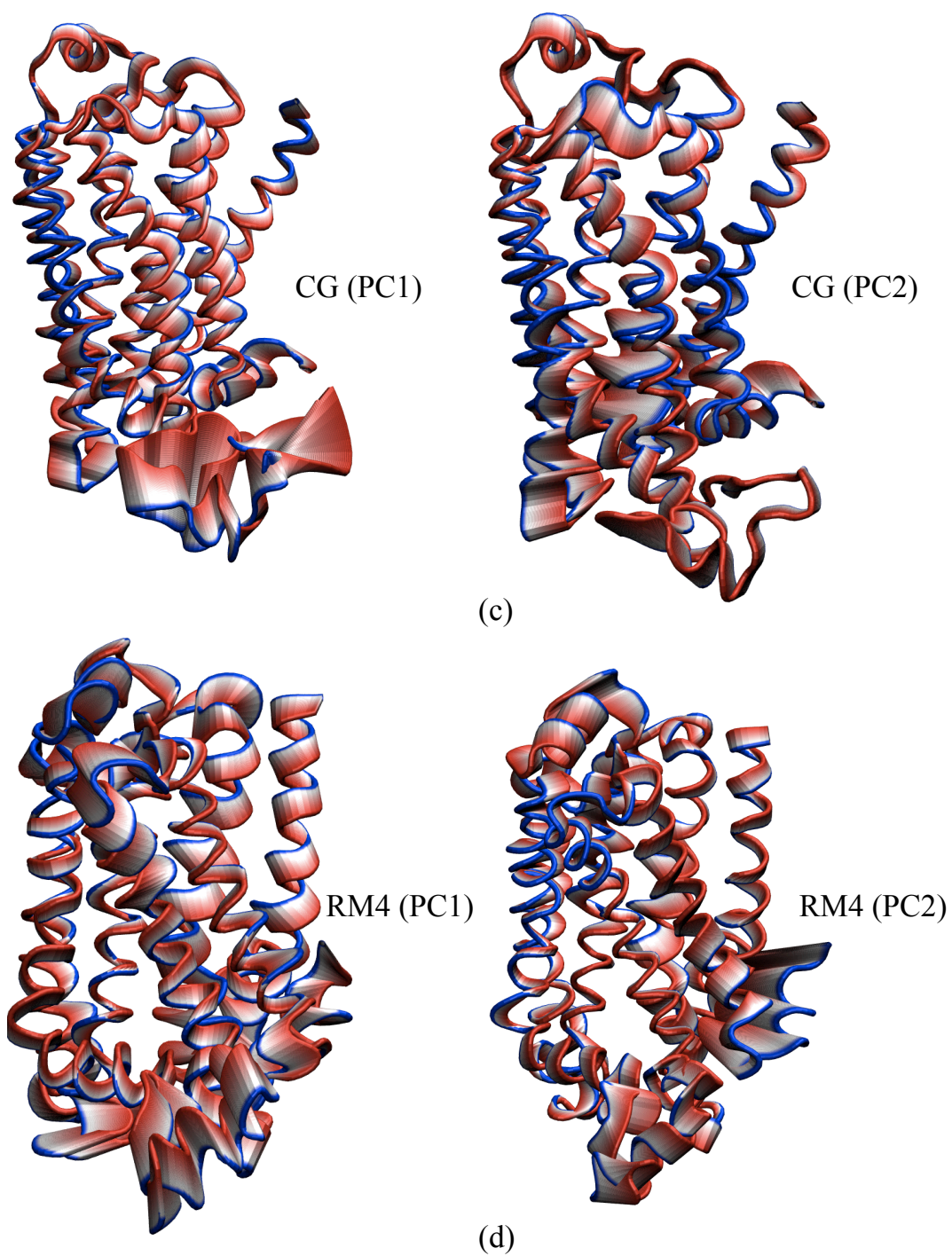


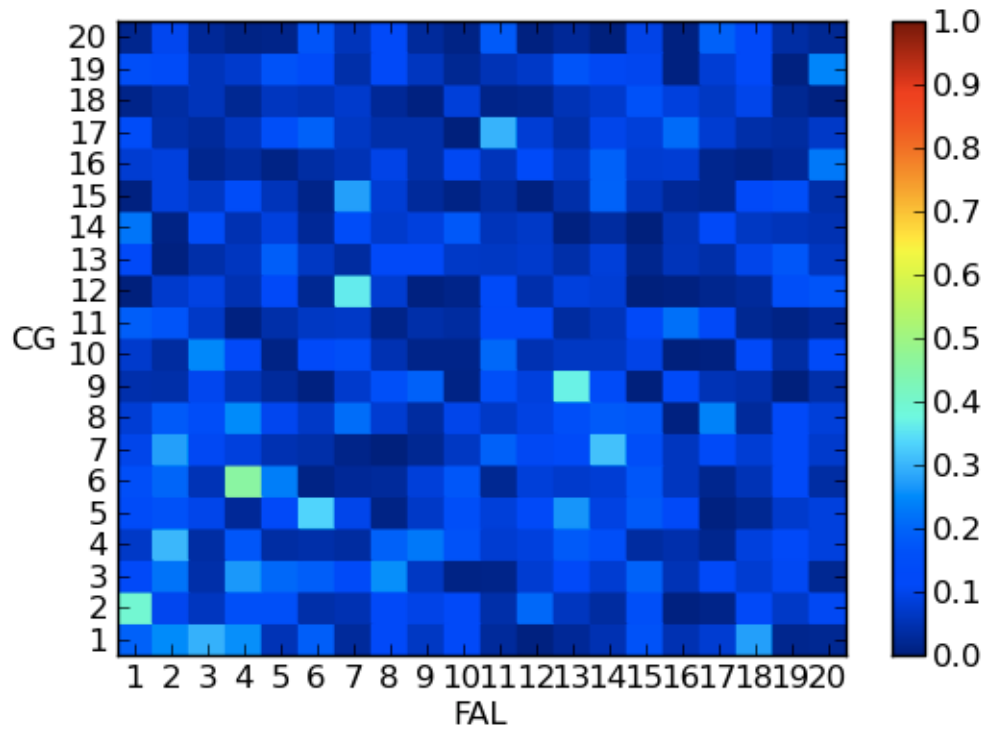
Figure 3.45.(cont'd) Collective motions on the first and second principal components of FAL (a), FAS1 (b), CG (c) and RM4 (d) simulations. Initial conformation is represented in blue color, last conformation is in red.

### 3.19. Overlap Calculations

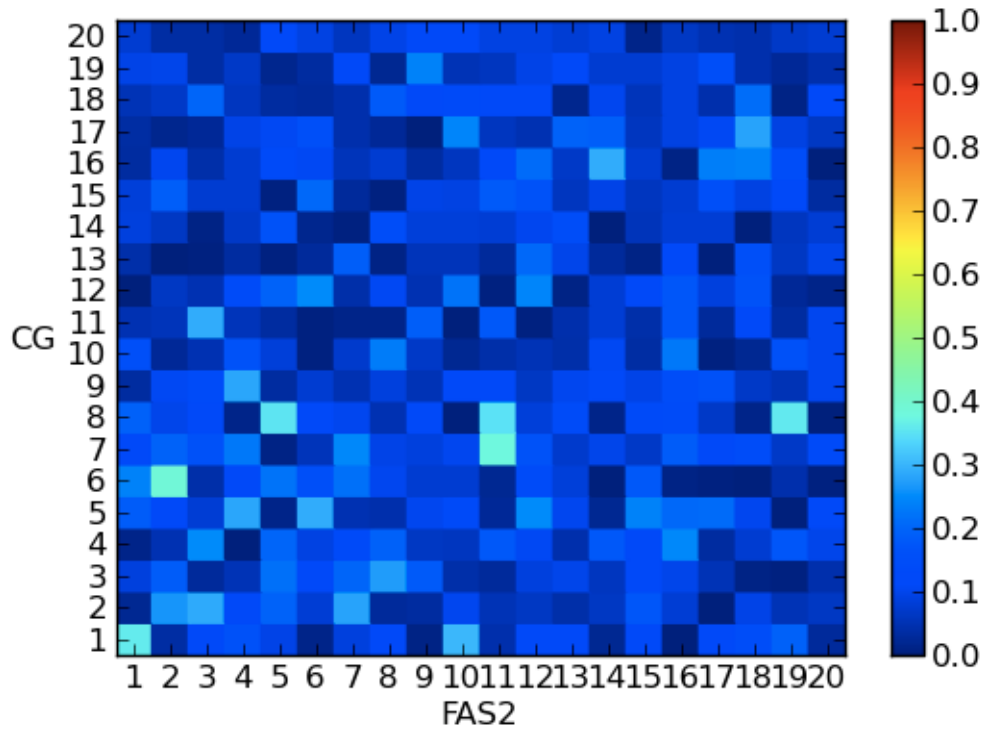
Eigenvectors obtained from PCA, in other words movement directions of the atoms in the principal components (modes) are used for revealing the similarities and differences between two different simulations. The dot product of two eigenvectors called as “overlap value”, varies between -1 and 1 and indicates the cosine of the angle between the vectors. For two vectors pointing in the same direction, the overlap value is 1, pointing in the opposite direction it is equal to -1, and 0 if they are perpendicular to each other. The overlap values are reported in Table 3.9 just for first three principal components of all trajectories.

The compatibility of the principal components of the CG simulation with FAL is also shown in Figure 3.46a for the first twenty modes as well. Overall, the principal modes of CG simulation reflect a low overlap with the principal modes of FAL. The maximum overlap value is 0.45, which is between CG’s sixth mode and FAL’s fourth mode. In Figure 3.46b, the overlap between the first twenty modes of CG and FAS2 simulations is illustrated. The overlap values vary between 0 and 0.4, the maximum overlap being 0.39 between CG’s sixth and FAS2’s second modes. Similarly, the overlap between CG and RM2 simulation is found to be weak (Fig. 3.49c), the highest overlap value being 0.36, which is between CG’s first mode and RM2’s second mode.

Similarly overlap values are given in Appendix A between CG and FAL, CG and RM, CG and FAS simulations. The highest overlap values for the first twenty modes between two FA simulations are found to be 0.74 (FAS2-FAS3)



(a)



(b)

Figure 3.46.

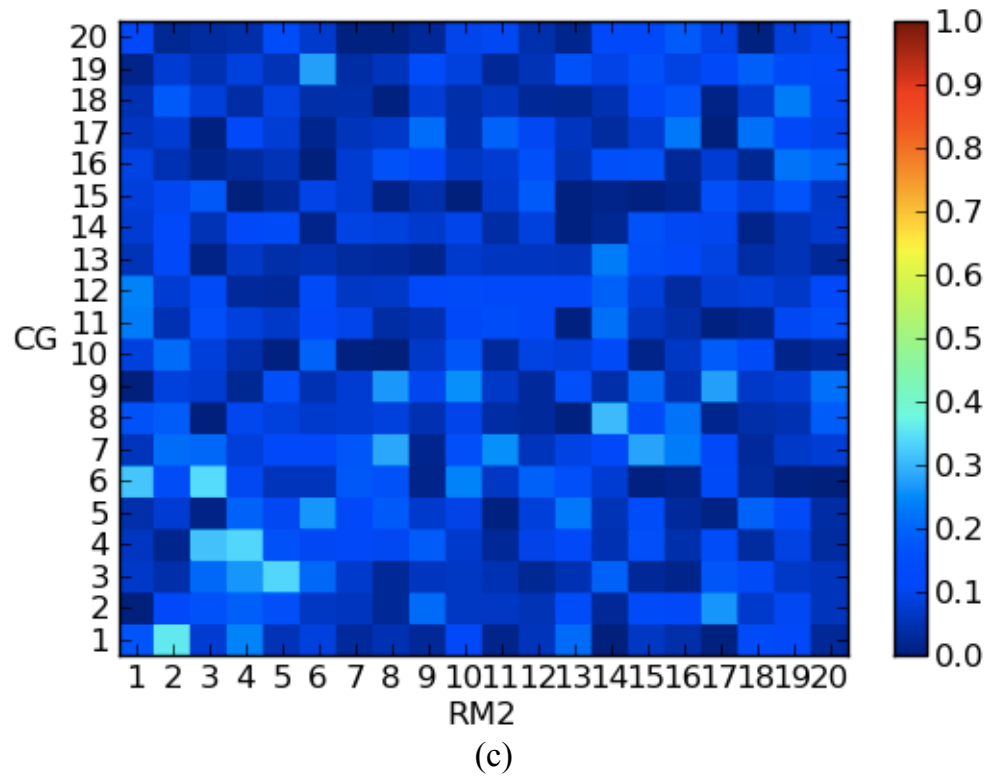


Figure 3.46. (cont'd) Overlap matrix of first 20 modes between CG and FAL (a), CG and FAS2 (b), CG and RM2 (c) simulations

Table 3.9. First three overlap values of all simulations

	FAL			FAS1			FAS2			FAS3			RMI			RM2			RM3			RM4		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
<b>CG</b>	1	-0.19	0.25	0.29	0.06	-0.26	0.36	0.04	0.13	0.21	-0.21	-0.10	-0.29	0.02	0.23	0.17	0.36	0.08	-0.08	0.11	-0.26	0.28	-0.04	0.31
	2	0.40	0.11	-0.06	0.27	0.03	0.29	-0.02	-0.26	0.25	0.08	0.01	0.18	0.11	-0.30	0.00	-0.11	0.16	-0.10	-0.10	0.27	0.03	0.20	0.02
	3	0.13	-0.22	0.04	-0.04	0.09	-0.09	0.09	-0.18	0.03	-0.13	-0.03	0.16	-0.01	-0.04	0.27	0.07	-0.04	-0.20	0.05	-0.07	0.06	-0.10	0.08
<b>FAL</b>	1	0.50	0.03	0.62	0.03	0.62	-0.09	-0.51	0.32	0.46	0.22	-0.07	-0.08	0.09	0.07	-0.03	-0.36	0.28	-0.45	0.24	0.18	-0.06	0.01	0.03
	2	0.56	-0.49	-0.23	0.49	0.31	0.41	0.31	0.41	0.23	-0.52	-0.45	-0.37	-0.47	0.35	-0.26	0.40	0.11	-0.04	0.38	-0.39	0.18	0.21	0.02
	3	0.41	0.18	-0.36	0.54	-0.09	0.10	0.17	-0.45	0.31	-0.45	0.31	-0.26	-0.05	0.03	0.07	0.32	-0.10	-0.02	0.20	-0.14	0.05	0.07	0.19
<b>FAS1</b>	1	0.57	-0.29	0.42	0.57	-0.29	0.42	0.65	-0.45	-0.19	-0.41	-0.19	-0.41	-0.24	0.29	-0.21	0.29	0.21	-0.30	0.38	-0.22	0.15	0.12	0.10
	2	-0.09	-0.65	-0.14	0.40	0.38	0.56	-0.03	0.58	0.05	0.28	0.13	-0.02	0.28	0.05	0.28	-0.12	0.27	-0.22	-0.33	0.15	0.02	-0.26	0.18
	3	-0.55	-0.35	-0.18	0.16	0.69	-0.12	0.09	0.28	0.13	-0.02	-0.55	0.42	-0.41	-0.02	-0.41	-0.02	0.27	-0.41	-0.02	0.27	-0.06	-0.09	-0.06
<b>FAS2</b>	1	0.18	-0.74	0.14	0.18	-0.74	0.14	-0.28	-0.34	0.07	-0.12	0.50	-0.06	-0.34	0.07	-0.12	0.50	-0.06	0.00	0.33	-0.36	0.04	0.23	0.09
	2	-0.49	-0.42	-0.36	0.04	-0.49	-0.36	0.04	-0.49	-0.19	-0.13	0.29	-0.32	0.39	-0.19	-0.13	0.29	-0.32	0.39	0.03	-0.17	0.04	0.18	-0.11
	3	0.26	-0.08	-0.16	0.26	-0.08	-0.16	-0.36	0.03	0.31	-0.12	0.17	0.34	-0.15	0.19	-0.12	0.17	0.34	-0.15	0.19	0.03	0.35	0.05	0.10
<b>FAS3</b>	1	-0.34	0.10	0.25	-0.34	0.10	0.25	0.10	0.25	0.10	0.25	0.11	0.00	0.39	-0.47	0.16	0.00	0.39	-0.47	0.16	-0.07	0.15	-0.07	0.21
	2	0.24	0.56	0.09	0.24	0.56	0.09	0.20	-0.54	0.33	-0.26	-0.54	0.33	0.20	-0.54	0.33	-0.54	0.33	-0.26	-0.31	0.32	-0.05	-0.22	-0.02
	3	0.03	0.36	-0.21	0.03	0.36	-0.21	0.22	-0.14	0.12	0.06	-0.14	0.12	0.22	-0.14	0.12	0.06	-0.14	0.12	0.06	-0.17	-0.02	-0.08	0.11
<b>RMI</b>	1	-0.07	-0.39	0.11	-0.07	-0.39	0.11	-0.07	-0.39	0.11	-0.14	0.15	-0.17	0.11	-0.14	0.15	-0.17	0.11	-0.14	0.15	-0.17	0.06	-0.15	-0.15
	2	0.37	-0.27	0.26	0.37	-0.27	0.26	0.37	-0.27	0.26	-0.13	-0.31	0.19	-0.13	-0.31	0.19	-0.13	-0.31	-0.13	-0.31	0.19	-0.02	-0.36	0.25
	3	-0.18	0.01	0.28	-0.18	0.01	0.28	-0.18	0.01	0.28	-0.34	0.11	-0.01	-0.18	0.01	0.28	-0.34	0.11	-0.34	0.11	-0.01	0.13	-0.05	-0.11
<b>RM2</b>	1	-0.01	-0.18	-0.15	-0.01	-0.18	-0.15	-0.01	-0.18	-0.15	-0.01	-0.18	-0.15	-0.01	-0.18	-0.15	-0.01	-0.18	-0.01	-0.18	-0.15	-0.16	-0.12	0.32
	2	0.24	0.03	-0.29	0.24	0.03	-0.29	0.24	0.03	-0.29	0.24	0.03	-0.29	0.24	0.03	-0.29	0.24	0.03	0.24	0.03	-0.29	0.21	0.13	0.09
	3	-0.44	-0.05	0.18	-0.44	-0.05	0.18	-0.44	-0.05	0.18	-0.44	-0.05	0.18	-0.44	-0.05	0.18	-0.44	-0.05	-0.44	-0.05	0.18	0.24	-0.06	0.05
<b>RM3</b>	1	-0.01	0.01	-0.06	-0.01	0.01	-0.06	-0.01	0.01	-0.06	-0.01	0.01	-0.06	-0.01	0.01	-0.06	-0.01	0.01	-0.01	0.01	-0.06	0.01	0.01	-0.06
	2	-0.04	0.13	0.03	-0.04	0.13	0.03	-0.04	0.13	0.03	-0.04	0.13	0.03	-0.04	0.13	0.03	-0.04	0.13	-0.04	0.13	0.03	-0.04	0.13	0.03
	3	0.09	-0.08	-0.16	0.09	-0.08	-0.16	0.09	-0.08	-0.16	0.09	-0.08	-0.16	0.09	-0.08	-0.16	0.09	-0.08	0.09	-0.08	-0.16	0.09	-0.08	-0.16



### 3.20. Cross Correlations

The cross-correlation between the residue fluctuations including the first ten modes for the CG model is shown in Figure 3.47a. First ten modes of the CG model constitute 62% of the overall motion in protein. On the other hand, the first mode of FAL model that constitutes 63% of the protein's motion is used in the cross-correlation profile shown in Figure 3.47b. Strong positive correlations exist between helices H1-H2, H1-H7, H2-H7, H6-H7 and weak positive correlations between helices H3-H5, H3-H6 in CG model. There are also many uncorrelated regions as well. In contrast, FAL model includes more strong positively and negatively correlated regions than CG model. Helices H1-H5, H1-H7, H2-H5, H2-H7, H5-H7 are positively correlated, while H1-H2, H1-H3, H2-H6, H3-H6 are negatively correlated in FAL model.

In Figure 3.48 (a), (b) and (c) the cross correlations between residue fluctuations of FAS1, FAS2 and FAS3 simulations are shown, respectively. The maps have been generated using the first two, three and two principal components, that explain the 58, 61 and 56% of the overall motion of the receptor for FAS1, FAS2 and FAS3, respectively. In FAS2 simulation, the protein has weakly correlated regions compared to FAS1 and FAS3. The most correlated motion belongs to the FAS3 in all short FA simulations.

In Figure 3.49, RM simulations' cross correlations between residue-based fluctuations are displayed. For RM1, RM2 and RM4, first two PCs were mapped, which explains the 57%, 57%, 65% of the overall motion of the protein, respectively. While in simulation RM3, first three PCs that explain the 59% of the overall dynamics, were used for calculations. RM2 has the most correlated motions, while RM3 has the most uncorrelated motions among the four RM simulations.

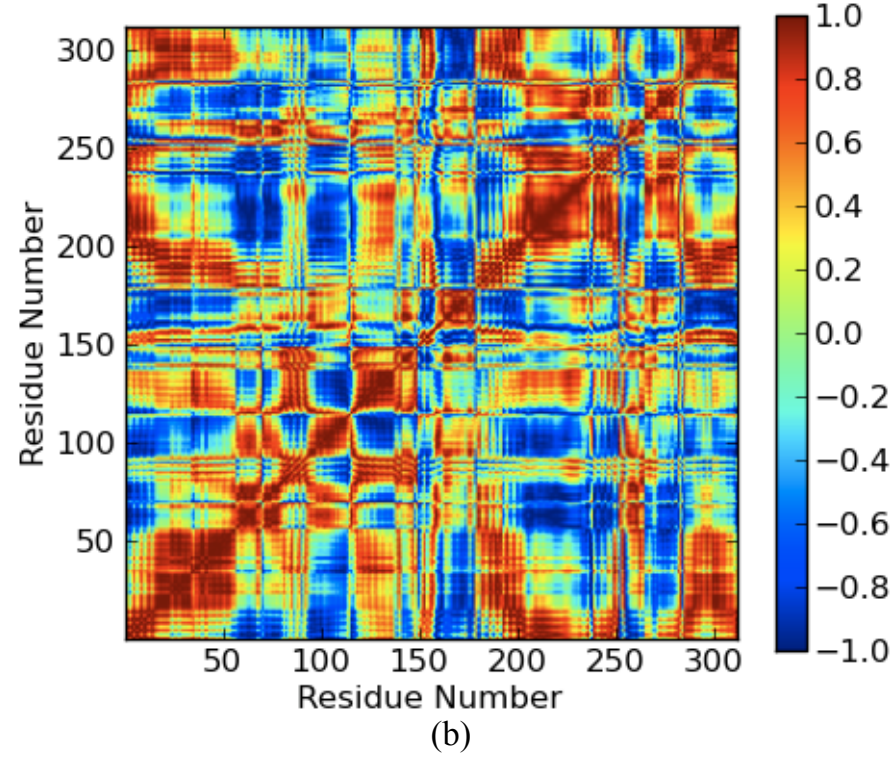
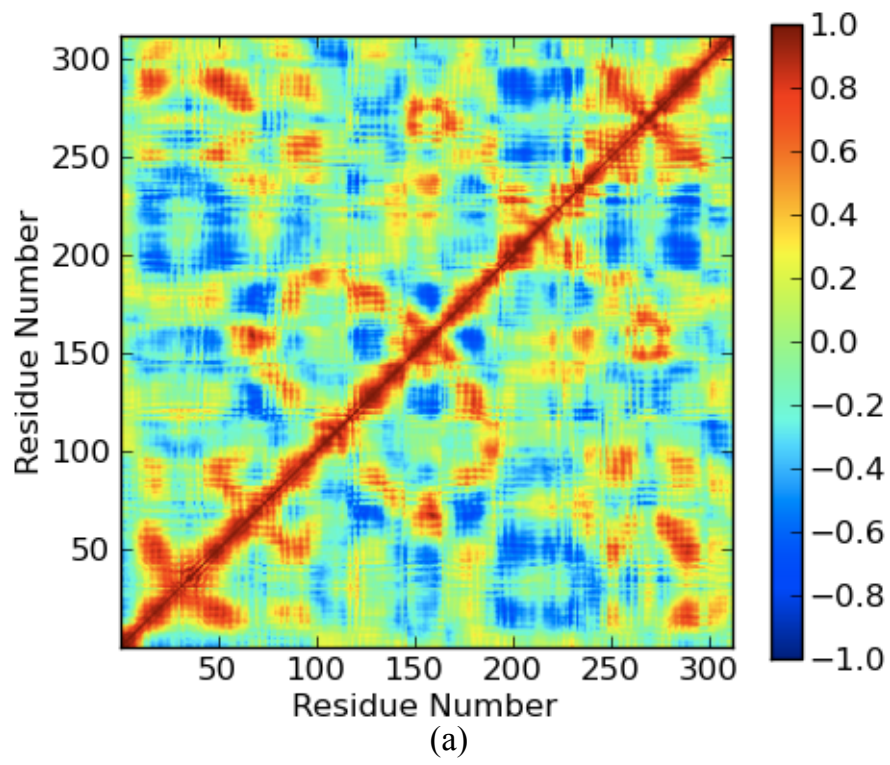


Figure 3.47. Cross-correlations between residue fluctuations (a) first ten modes of CG simulation, (b) first mode of FAL simulation.

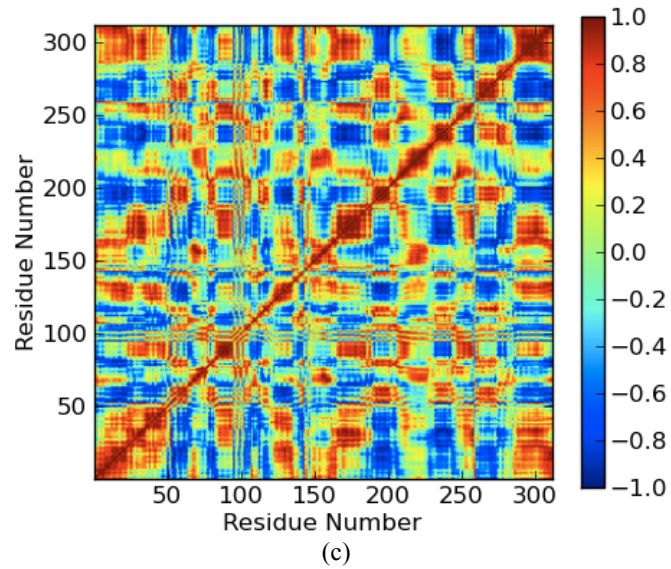
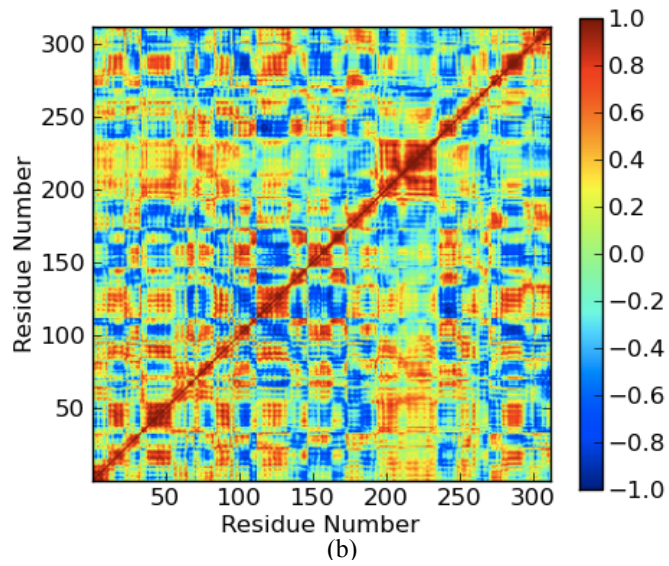
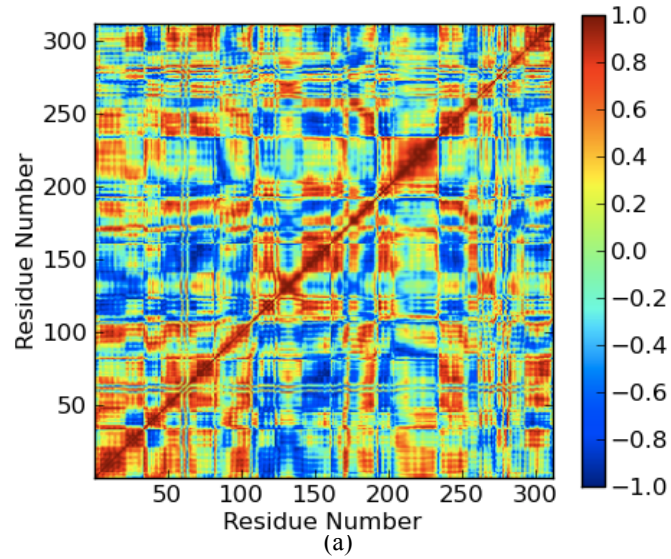


Figure 3.48. Cross-correlations between residue fluctuations (a) first two modes of FAS1 simulation, (b) three modes of FAS2 simulation, (c) two modes of FAS3 simulation.

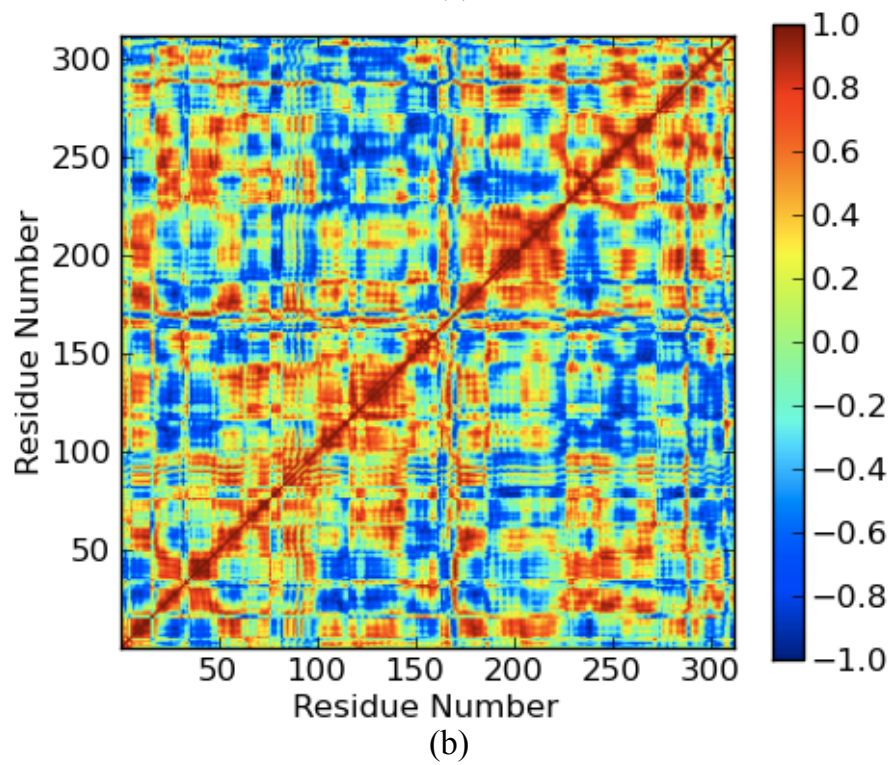
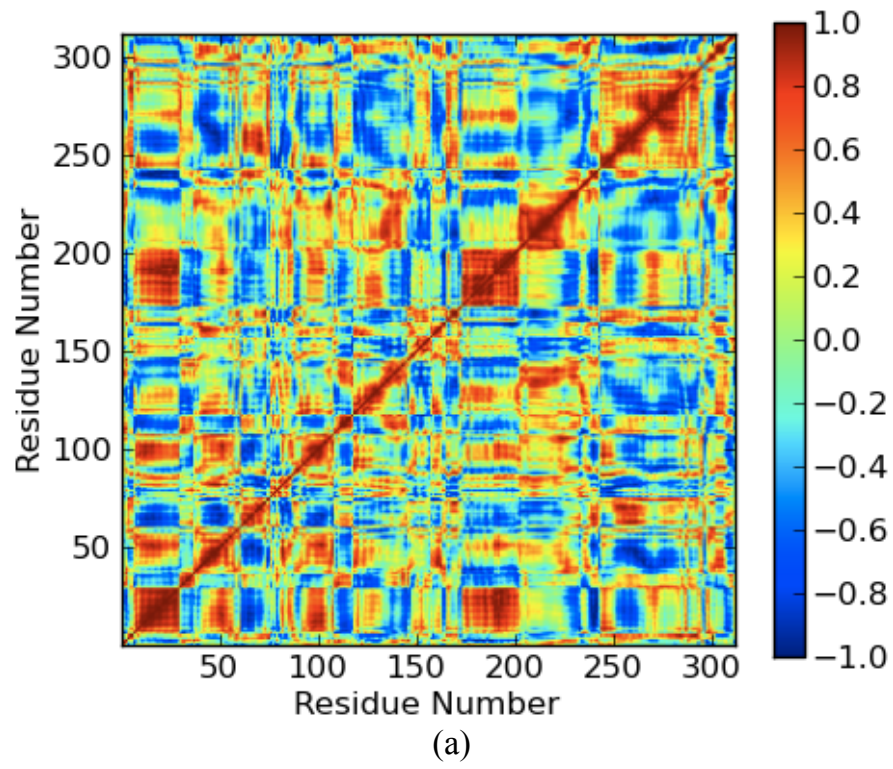


Figure 3.49

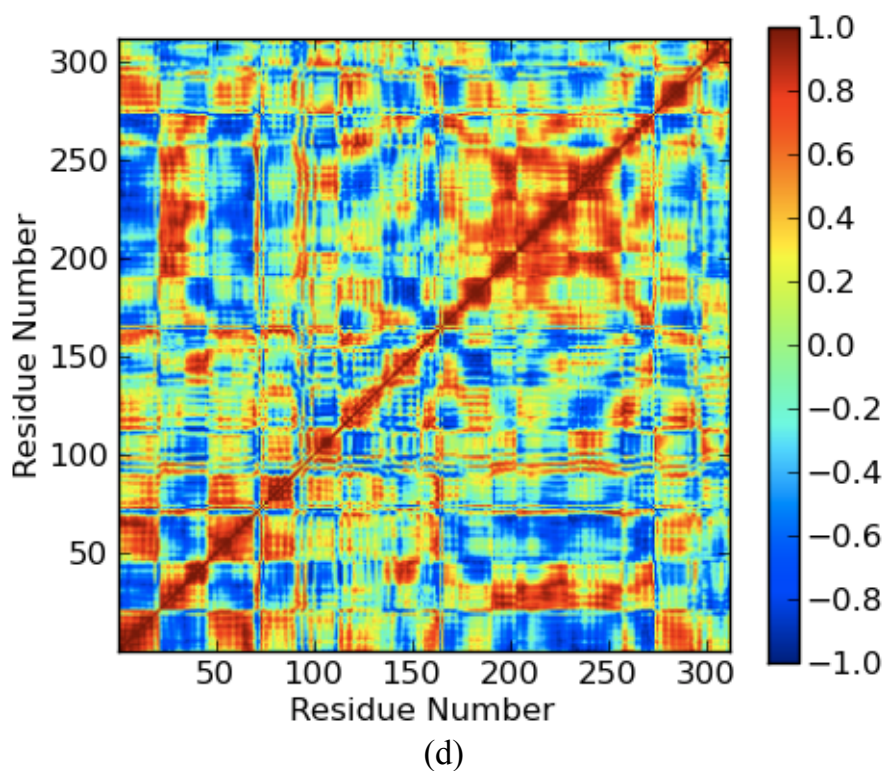
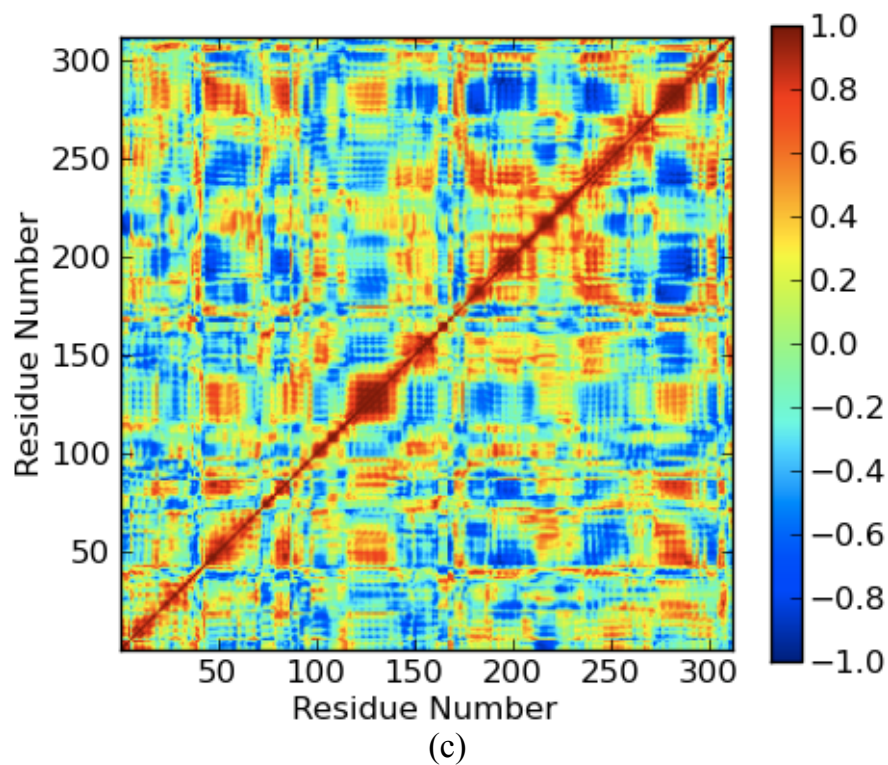


Figure 3.49.(cont'd) Cross-correlations between residue fluctuations (a) first two modes of RM1 simulation, (b) two modes of RM2 simulation, (c) first three modes of RM3 simulation, (d) two modes of RM4 simulation.

### 3.21. Applicability of Martini Force Field on Proteins

The distortion of the structural motif of trans-membrane helices in  $\beta_2$ AR may occur either from inadequate representation of protein-lipid interactions or protein's internal energy in the Martini force field. To identify the cause, a residue-based coarse-grained model of T4-lysozyme, which is a G protein for  $\beta_2$ AR, was constructed and simulated alone in a CG water environment for 6  $\mu$ s at 310 K. Figure 3.50 shows the RMSD profile of each snapshot with respect to the initial state. A sudden structural deviation up to 5-6 Å has occurred at the initial stages. The system has reached its equilibrium at around 500 ns at 7 Å. A sudden increase to 9 Å was observed at around 2000 ns, but then the system was stabilized after 2500 ns at around 8 Å.

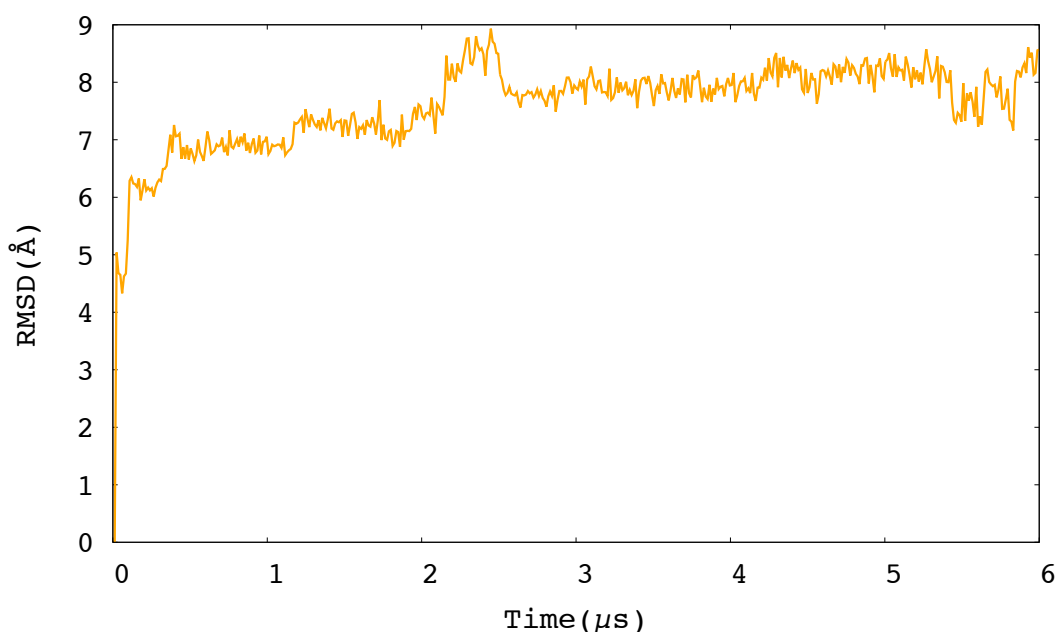


Figure 3.50. RMSD profiles of T4-lysozyme.

Structural deformations of helices similar to  $\beta_2$ AR occurred at initial stages of the simulation. Figure 3.51 shows the first and the last snapshots aligned based on all structure and three selected helices. In addition, the tertiary structure of the protein has not been maintained as well as in  $\beta_2$ AR. The same helical distortions that

occurred in the absence of lipids suggest an inadequate representation of the protein's internal energy in the force field rather the result of an inadequate representation of protein-lipid interactions.

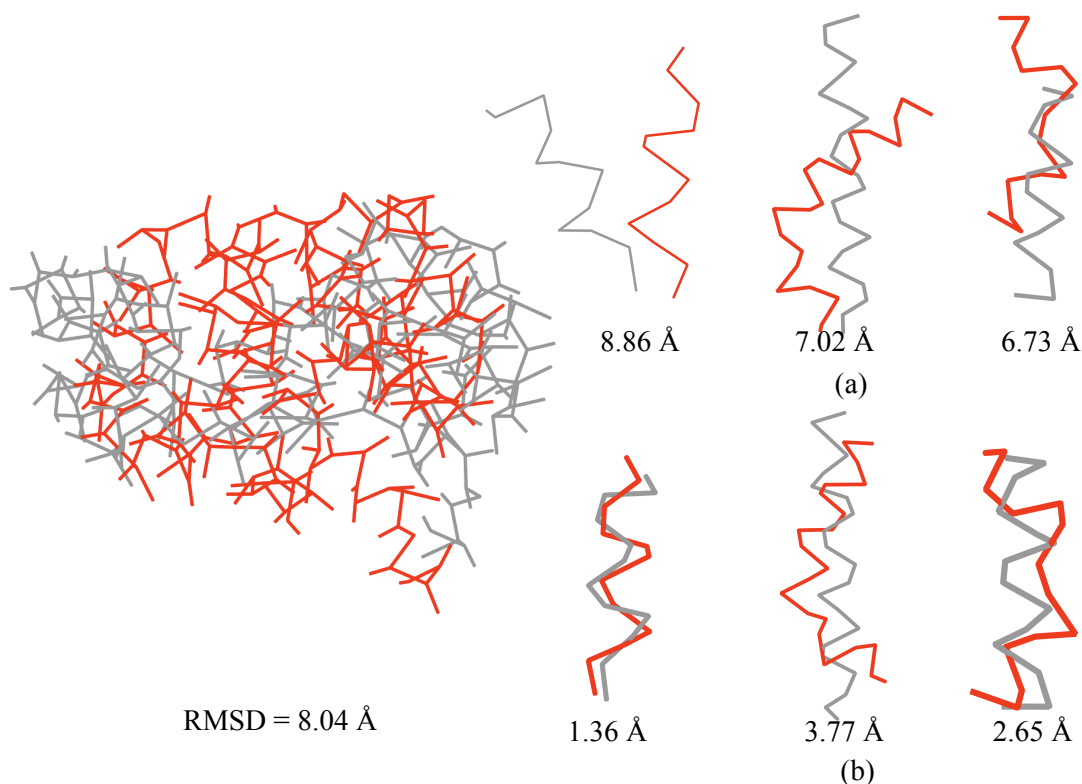


Figure 3.51. Alignments of T4-lysozyme's first and last snapshots according to the all structure (a) and helices (b). First snapshot is shown in grey, last snapshot is in red color.

### 3.22. Comparison of Simulation Costs Between Fully-Atomistic and Coarse-Grained Models

Although the main goal of preferring the CG model to the FA model was to explore a wider conformational space, another benefit of CG modeling is the significant gain in the computational speed. While a 1  $\mu$ s MD simulation with a fully atomistic model can be completed nearly in 1 year 9 months using 12 cores on a supercomputer, a 1  $\mu$ s CG simulation can be completed in just two days.

## Chapter 4

### Conclusions

CG simulations conducted at 310 K yielded a less mobile receptor in all its regions, including the ICL3 compared to an all-atom simulation. But still, RMSF profile clearly indicates more flexibility at the loop regions relative to trans-membrane region. RMSD clustering of CG MD trajectory based on core, trans-membrane, and binding site regions do not reveal structural diversity with only one large cluster that constitutes 75% of all snapshots. The analysis of four reverse-mapped structures that have been subjected each to 100 ns MD simulation show that RMSF values agree well with those obtained from 100 ns long fully atomistic MD simulations. Their energy values were stable throughout the simulation. The overall three dimensional structure of the receptor was maintained in all four reverse-mapped models, yet some deformations have been observed especially in helix 5 and helix 6. These deformations already existed at the initial structure just after reverse mapping. Thus, it is likely that they originate from coarse-grained simulations. The reverse-mapped simulations simply could not recover the structural motif due to high-energy barriers.

RMSD clustering of all RM snapshots with FA snapshots reveal distinct clusters for RM conformations. Also, the RMSD profile of binding site region calculated with respect to the crystal structure (PDB id: 2rh1) for each RM FA simulation showed that the RMSD value becomes stable at around 4 Å - 4.5 Å. Despite a few helical deformations, the three-dimensional structure of the binding site remains stable throughout the simulation, thus the structural diversity presented by RM models is found to be satisfactory and can be further tested in virtual screening experiments to reveal their potential in extracting the known agonists and antagonists.



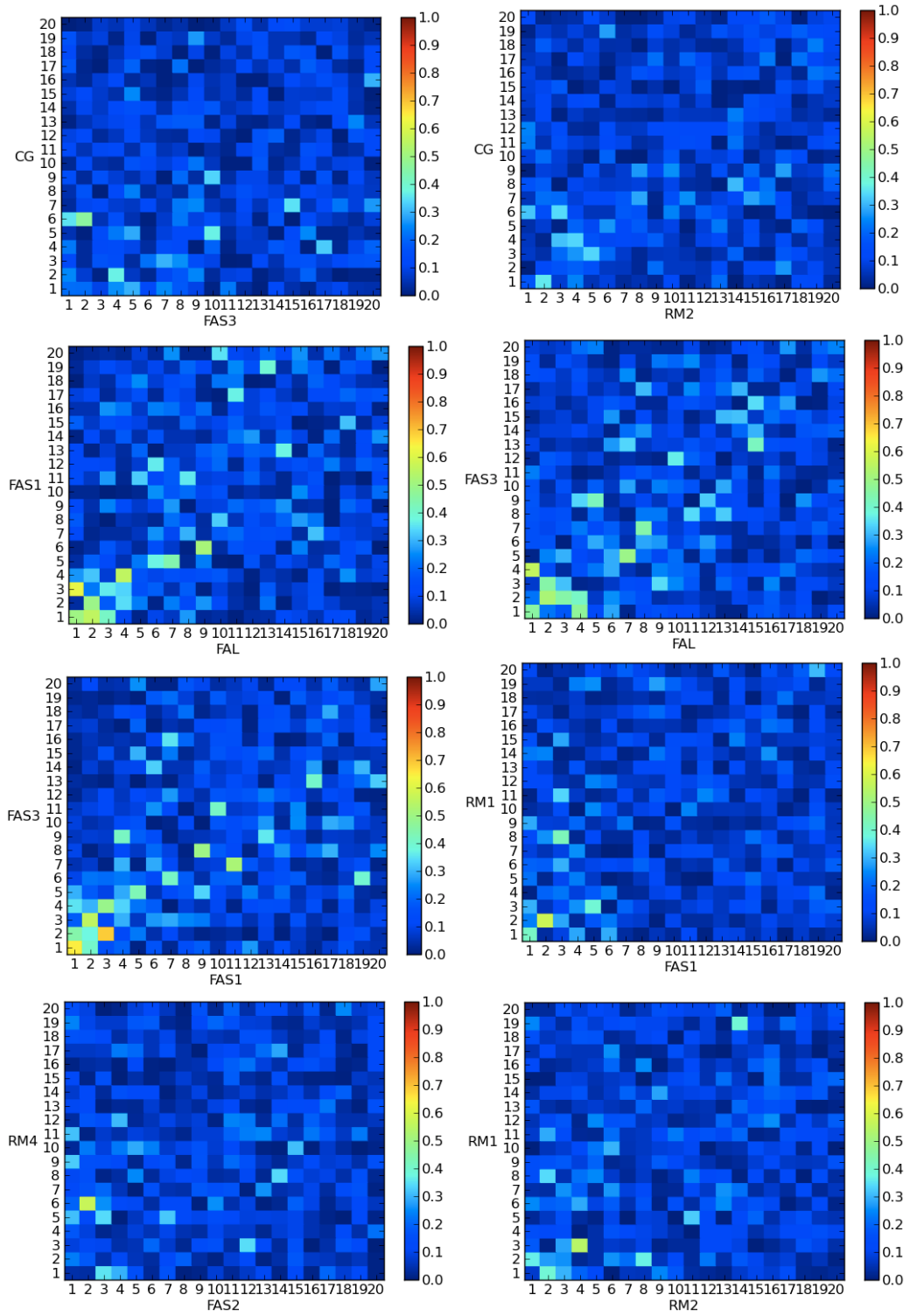
The distance between two anchor sites Asp113 and Ser203-Ser207 was calculated for each RM simulations. Compared to all-atom simulations, the distances of RM models fluctuate within a wider range between 5 Å and 10 Å. In one reverse-mapped model, both Asp113-Ser203 and Asp113-Ser207 distances increase up to 18 Å towards the end of the simulation. However, as opposed to FA models, when the binding site is expanded at the upper part, the lower part of the receptor is not found narrowed or vice versa. In other words, the receptor's binding site (upper) and the G-protein binding site with ICL3 (lower) do not oscillate in correlation around equilibrium state.

In the second part, PCA method was performed to get the most dominant motion of the receptor throughout CG, FA and RM simulations. In FAL simulation, the first mode explains 63% of the receptor's entire motion, while in CG simulation it only explains 19%. Three FAS simulations' first mode explain only 38, 34 and 36% of the whole receptor's dynamics which indicates that 100 ns is not sufficient to capture the protein's global dynamics in the slow modes. Similarly, RM simulations have explanation values for the first mode as 36, 43, 31, and 47 %, which are not as descriptive as FAL.

In order to get the level of agreement between the collective motions of CG, FA and RM simulations, overlap values are calculated. CG simulation reflects a low overlap with both FA and RM simulations. The maximum overlap value is between CG's sixth mode and FAL's fourth mode and is equal to 0.45, while this value is 0.39 between the CG's sixth and FAS2's second modes and 0.36 between the CG's first and RM2's second modes. Additionally, orientational cross-correlation between the residue-based fluctuations are calculated for CG, FA, and RM simulations. In CG model, correlation between the residues is relatively weak compared to FA and RM models. FAL model have the strongest negatively and positively correlated regions. Also FAS1, FAS3, RM1, RM2, and RM4 have strong negatively and positively correlated regions.

Although CG modeling is an alternative and efficient way to study protein dynamics and explore a wider conformational space in short times, further improvements are necessary for improving the Martini force field described for proteins in order to avoid structural deformations. Still, the deformation in the binding site is tolerable and may represent an alternative model to be used in virtual screening.

## APPENDIX A: OVERLAP MATRIX OF CG, FA AND RM SIMULATIONS



## **APPENDIX B: TUTORIAL OF COARSE-GRAIN MODELLING**

**Step 1:** Protein and lipid molecules were extracted from .pdb file of the system.

**Step 2:** .psf file was created for protein and lipids.

**Step 3:** With new .psf and .pdb files, protein and lipid molecules were converted into coarse-grained model via VMD CG Builder using protein.cgc and lipid.cgc files. Then .pdb and rcg files were obtained.

**Step 4** .psf file was created.

**Step 5:** With cgsolvate.tcl and solv\_remove.tcl files, protein and lipid molecules were solvated with CG water molecules.

**Step 6:** With cgionize.tcl, the system was neutralized with CG ions.

## APPENDIX C: Scripts used in Principal Component Analysis

```
from prody import *
import MDAnalysis

startLogfile('CG_310K-6ms_EDA')

universe=MDAnalysis.Universe('CG_ionizedeq-03-
CONT_lastframe_IONIZED.psf', 'cg_310-6MS.dcd')

# Select atoms of interest
# This selection must be "name CA" for fully-atomistic structures.
universe_ca = universe.selectAtoms('name BB')

# Get coordinates of CA atoms
ca_coords = universe.trajectory.timeseries(universe_ca, format='fac')

reference_str = parsePDB('ionizedeq-03-CONT_lastframe_protein_CA.pdb')

ensemble = Ensemble('MD-CG-310K-6ms-Snapshots')

# Add all coordinate sets to ensemble
ensemble.addCoordset(ca_coords)

# Set reference coordinates
ensemble.setCoords(reference_str)
ensemble.setAtomGroup(reference_str)
ensemble.select('name BB and resnum 32 to 230 263 to 342')

# Perform iterative superimposition
ensemble.superpose()

ensemble.select('name BB')

# Calculation RMSF
rmsf = ensemble.getRMSFs().round(20)
f = open('RMSF_CG-310K-6ms.txt','w')

for i in rmsf:
    f.write(str(i) + "\n")
f.close()
```

```
eda = EDA('EDA-CG-310K-6ms')
eda.buildCovariance(ensemble)
eda.calcModes()

# Eigenvectors
ev = eda.getEigenvectors().round(10)
f = open('Eigenvectors_CG-310K-6ms.txt','w')

for i in ev:
    f.write(str(i) + "\n")
f.close()

saveModel(eda)
saveEnsemble(ensemble)
writeNMD('md_eda-CG-310K-6ms.nmd', eda[:20], reference_str)
writeArray('CG-310K-6ms_pca_modes.txt', eda.getArray(), format = '%8.3f')
writeArray('CG-310K-6ms_pca_eigenvectors.txt', eda.getEigenvectors())

closeLogfile('CG_310K-6ms_EDA')
```

## APPENDIX D: CONFIGURATION FILES OF CG AND FA SIMULATIONS

Structure	CG.psf	Structure	FA.psf
Coordinates	CG.pdb	Coordinates	FA.pdb
set temperature	310	set temperature	310
set outputname	CG_MD	outputName	FA_MD
firsttimestep	0	firsttimestep	0
# Continuing a job from the restart files if {0} { set inputname binCoordinates \$inputname.restart.coor binVelocities \$inputname.restart. extendedSystem \$inputname.restart.xsc } # Input paraTypeCharmm on parameters rbcg-2007.par temperature \$temperature		# Continuing a job from the restart files if {0} { set inputname binCoordinates \$inputname.restart.coor binVelocities \$inputname.restart.vel extendedSystem \$inputname.restart.xsc } # Input paraTypeCharmm on parameters par_all27_prot_lipid.prm temperature \$temperature	
# Force-Field Parameters exclude 1-2 cutoff 12.0 switching on switchdist 9.0 pairlistdist 16.0 margin 5.0		# Force-Field Parameters exclude scaled1-4 1-4scaling 1 cutoff 12. switching on switchdist 10. pairlistdist 13.5	
# Integrator Parameters timestep 5.0 #rigidBonds all nonbondedFreq 1 stepspercycle 20		# Integrator Parameters timestep 2.0 rigidBonds all nonbondedFreq 1 fullElectFrequency 2 stepspercycle 20	
# Constant Temperature Control langevin on langevinDamping 5 langevinTemp \$temperature langevinHydrogen off		# Constant Temperature Control langevin on langevinDamping 1 langevinTemp \$temperature	
# Periodic Boundary Conditions if {1} { cellBasisVector1 89.2 0. 0. cellBasisVector2 0. 97.6 0. cellBasisVector3 0. 0. 101.2           }		# Periodic Boundary Conditions if {1} { cellBasisVector1 89.2. 0. 0. cellBasisVector2 0. 97.6. 0. cellBasisVector3 0. 0. 101.2           }	

cellOrigin	0.3 -1.0 2.8	cellOrigin	0.3 -1.0 2.8
}		}	
wrapAll	on	wrapWater	on
		wrapAll	on
# Constant Pressure Control (variable volume)		# Constant Pressure Control (variable volume)	
if{1}{		if {1} {	
useGroupPressure	no	useGroupPressure	yes
useFlexibleCell	yes	useFlexibleCell	yes
useConstantArea	yes	useConstantArea	yes
langevinPiston	on	langevinPiston	on
langevinPistonTarget	1.01325	langevinPistonTarget	1.01325
langevinPistonPeriod	1000.	langevinPistonPeriod	200.
langevinPistonDecay	500.	langevinPistonDecay	50.
langevinPistonTemp	\$temperature	langevinPistonTemp	\$temperature
# Output		# Output	
outputName	\$outputname	outputName	\$outputname
restartfreq	10000	restartfreq	10000
dcdfreq	10000	dcdfreq	10000
xstFreq	10000	xstFreq	10000
outputEnergies	10000	outputEnergies	500
outputPressure	10000	outputPressure	500
#heating protocol		#PME (for full-system periodic electrostatics)	
reassignFreq	10000	if {1} {	
reassignTemp	30	PME	yes
reassignIncr	1	PMEGridSizeX	90
reassignHold	310	PMEGridSizeY	100
		PMEGridSizeZ	120
		}	
#script		# Minimization	
minimize	30000	if {1} {	
run	200000000	minimize	1000
		reinitvels	\$temperature
		}	
		run	50000000



## References

- [1] Wallin, E.; von Heijne, G., “Genome-wide analysis of integral membrane proteins from eubacterial, archean, and eukaryotic organisms” *Protein Sci.* 7:1029-1038, 1998.
- [2] Nilsson, J.; Persson, B.; von Heijne, G., “Comparative analysis of amino acid distributions in integral membrane proteins from 107 genomes”, *Proteins*, 60:606-616, 2005.
- [3] Cherezov, V.; Rosenbaum, D.M.; Hanson, M.A.; Rasmussen, S.G.F.; Thian, F.S.; Kobilka, T.S.; Choi, H.J.; Kuhn, P.; Weis, W.I.; Kobilka, B.K.; Stevens, R.C., “High-resolution crystal structure of an engineered human beta(2)-adrenergic G protein-coupled receptor”, *Science*, 318:1258-1265, 2007.
- [4] White, S.H.; “The progress of membrane protein structure determination” *Protein Sci.* 13:1948–1949, 2004.
- [5] Hunte C., *Biochem.Soc.Trans.*, 33:938, 2005.
- [6] Palsdottir, H.; Hunte, C., *Biochim.Biophys.Acta*, 1666, 2, 2004.
- [7] Lindahl, E.; Sansom, M. SP.; “Membrane proteins: molecular dynamics simulations”, *Current Opinion in Structural Biology*, 18:425-431, 2008.
- [8] Spijker, P.; van Hoof, B.; Debertrand, M.; Markvoort, A.J.; Vaidehi, N.; Hilbers, P.A.J.; “Coarse Grained Molecular Dynamics Simulations of Transmembrane Protein-Lipid Systems”, *Int. J. Mol. Sci.*, 11:2393-2420, 2010.
- [9] Vaidehi, N.; Floriano, W.B.; Trabanino, R.; Hall, S.E.; Freddolino, P.; Choi, E.J.; Zamanakos G.; Goddard, W.A., III, “Prediction of structure and function of G protein-coupled receptors” *Proc. Natl. Acad. Sci.*, 99:12622–12627, 2002.
- [10] Freddolino, P.L., Kalani, M.Y.S., Vaidehi, N., Floriano, W.B., Hall, S.E.,

- Trabanino, R.J., Kam, V.W.T., Goddard, W.A., III, "Predicted 3D structure for the human  $\beta$ 2 adrenergic receptor and its binding site for agonists and antagonists" *Proc. Natl. Acad. Sci.*, 101:2736–2741, 2004.
- [11] Kalani, M.Y.S.; Vaidehi, N.; Freddolino, P.L.; Hall, S.E.; Trabanino, R.J.; Floriano, W.B.; Kam, V.W.T.; Kalani, M.A.; Goddard, W.A., III, "The predicted 3D structure of the human D2 dopamine receptor and the binding site and binding affinities for agonists and antagonists" *Proc. Natl. Acad. Sci.*, 101:3815–3820, 2004.
- [12] Nielsen, S.O.; Lopez, C.F.; Srinivas, G.; Klein, M.L.; "Coarse grain models and the computer simulation of L-alpha-to-inverted phase transition using coarse-grain molecular dynamics, *Biophys. J.*, 87:2107-2115, 2004b.
- [13] Bond, P.J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M.S.; "Coarse-grained molecular dynamics simulations of membrane proteins and peptides", *J. Struct. Biol.*, 157(3):593-605, 2006.
- [14] Tozzini, V.; "Coarse-grained models for proteins", *Curr. Opin. Struct. Biol.*, 15:144-150, 2005.
- [15] Levitt, M.; "A simplified representation of protein conformations for rapid simulation of protein folding", *J. Mol. Biol.*, 104:59-107, 1976
- [16] Bahar, I.; Jernigan; "Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation" *J. Mol. Biol.*, 266:195-214, 1997.
- [17] Marrink, S.J.; de Vries, A.H. ; Mark, A.E.; "Coarse grained model for semi-quantitative lipid simulations", *J. Phys.Chem.B*, 108:750-760, 2004.
- [18] Marrink, S.J.; Risselada, H.J.; Yefimov, S.; Tieleman, D.P.; de Vries, A.H.; "The MARTINI forcefield: coarse grained model for biomolecular simulations", *J. Phys.Chem.B*, 111:7812-7824, 2007.
- [19] Smit, B.; Hilbers, A.J.; Esselink, K.; Rupert, L.A.M.; Van Os, N.M.; Schlijper, G.; "Computer simulations of a water/oil interface in the presence of micelles", *Nature*, 348:624–625, 1990.
- [20] Periolo, X.; Huber, T.; Marrink, S.J.; Sakmar, T.P. "G Protein-Coupled Receptors Self-Assemble in Dynamics Simulations of Model Bilayers", *J. Am. Chem. Soc.*, 129:10126-10132, 2007.
- [21] Treptow, W.; Marrink, S.J.; Mounir, T., "Gating Motions in Voltage-Gated Potassium Channels Revealed by Coarse-Grained Molecular Dynamics

- Simulations”, *J. Phys. Chem. B*, 112:3277-3282, 2008.
- [22] Scott, K.A.; Bond, P.J.; Ivetac, A.; Chetwynd, A.P.; Khalid, S.; Sansom, M.S.P., “Coarse-Grained MD Simulations of Membrane Protein-Bilayer Self-Assembly”, *Structure* 16(4):621-30, 2008.
- [23] Shih, A.Y.; Arkhipov, A.; Freddolino, P.L.; Schulten, K., “Coarse Grained Protein-Lipid Model with Application to Lipoprotein Particles”, *J. Phys. Chem. B*, 110:3674-3684, 2006
- [24] Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kale, L.; Schulten, K. *Journal of Computational Chemistry*, 26:1781-1802, 2005.
- [25] Humprey, A.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics*, 14:33-38, 1996.
- [26] Gunsteren, W.E.; Berendsen, H.J.C., “Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry”, *Angew. Chem. Int. Ed. Engl.*, 29:992-1023, 1990
- [27] Leach, A.R., “Molecular Modelling-Principles and Applications”, 2nd edition, 2001.
- [28] Guvench, O.; MacKerell, A.D.Jr., “Comparison of Protein Force Fields for Molecular Dynamics Simulations”, *Methods in Molecular Biology*, Volume 443, 1:63-88, 2008.
- [29] Bezkorovaynaya, O.; Lukyanov, A.; Kremer, K.; Peter, C., “Multiscale simulation of small peptides: consistent conformational sampling in atomistic and coarse-grained models”, *J. Comput. Chem.*, 33(9):937-49, 2012
- [30] Oliviero Carugo, O.; Pongor, S. A., “A normalized root-mean-square distance for comparing protein three-dimensional structures”, *Protein Sci.*, 10(7): 1470-1473, 2001
- [31] Feig, M. J.; Karanicolas, J.; Brooks, C. L., “MMTSB Tools Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology”, *Journal of Molecular Graphics and Modelling*, 22:377-395, 2004.
- [32] Wagstaff, K.; Cardie, C.; Rogers, S.; Schroedl, S., “Constrained K-means Clustering with Background Knowledge”, *Proceedings of the Eighteenth International Conference on Machine Learning*, 577–584, 2001

- [33] Ichiye, T.; Karplus, M.; “Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations”, *Proteins*, 11:205–217, 1991.
- [34] Amadei, A.; Linssen, A.B.; Berendsen, H.J.; “Essential dynamics of proteins”, *Proteins*, 17:412–425, 1993
- [35] Skjaerven, L.; Martinez, A.; Reuter, N., “Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit”, *Proteins*, 79:232–243, 2011
- [36] Bakan A, Meireles LM, Bahar I., ProDy: Protein Dynamics Inferred from Theory and Experiments *Bioinformatics*, 27(11):1575-1577, 2011
- [37] Cherezov, V.; Rosenbaum, D.M.; Hanson, M.A.; Rasmussen, S.G.; Thian, F.S.; Kobilka, T.S.; Choi, H.J.; Kuhn, P.; Weis, W.I.; Kobilka, B.K.; Stevens, R.C., “High Resolution Crystal Structure of An Engineered Human Beta2Adrenergic G-Protein-Coupled Receptor”, *Science*, 318:1258-1265, 2007.
- [38] Ozcan, O., “Exploring the intrinsic dynamics of human beta-2 adrenergic receptor and its potential use in computational drug design studies”, M.S. Thesis, *Bogazici University*, 2011.
- [39] Caves, L.S.D.; Evanseck, J.D.; Karplus, M., “Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin”, *Protein Sci* 7:649-666, 1998.

## **Curriculum Vitae**

Sibel akan was born on 16 August 1980, in Malatya. She received her BS degree in Biochemistry in 2004 from Ege University. She worked at marketing department of different pharmaceutical companies from 2004 to 2010. Then she has worked as a research assistant at the Department of Computational Biology and Bioinformatics of Kadir Has University since 2010.