INVESTIGATION OF ALLOSTERIC COUPLING IN BETA-2 ADRENERGIC RECEPTOR THROUGH MOLECULAR DYNAMICS SIMULATIONS

by

Canan Özgür

B.S. , Chemical Engineering, Boğaziçi University, 2012

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computational Science and Engineering

Boğaziçi University

2014

ACKNOWLEDGEMENTS

I would like to present my intimate thanks to my thesis advisor Prof. Pemra Doruker Turgut and my co-advisor Assoc. Prof. Demet Akten Akdoğan, who helped me throughout this work in all possible ways. I am very grateful to be able to work with them during this thesis under their guidance.

I also want to express my gratitude to Prof. Türkan Haliloğlu, Prof. Levent Kurnaz and Assoc. Prof. Elif Özkırımlı Ölmez for reading my thesis and commenting on it.

I also want to mention my gratitude for graduate scholarship ensured by TÜBİTAK for my M. S. studies via the project number 213M544.

I would like to thank to Mehmet Tarık Can, Hüseyin Doğa Fındık, Arzu Uyar, Zeynep Kürkçüoğlu and Fidan Sümbül for their support and help whenever I needed. I wish to express my thanks to all of the members of Polymer Research Center; Burcu A. Fas, Burçin Acar, Gökçe Ezeroğlu, Melis Yıldırım, Müberra Yılmaz, Nurver Tezcan, Serdar Özsezen, Seren Soner, Şefik Ovalı and Zeynep Erge Akbaş for their sincere friendship and kindness. I want to thank to Canan and Bilgi Dedeoğlu and Yüksel Yıldırım for their smiling faces and positive attitudes to me all the time.

I want to thank to Gaye Genç Kara and Yunus Emre Kara for supporting me throughout my thesis and believing in my work.

Lastly, I would like to thank to my family Sebaye Özgür, Selçuk Özgür and Nurcan Özgür for their trust in me and support all the time. Many thanks to Cesur Özdemir for everything he is. This thesis is dedicated to them.

ABSTRACT

INVESTIGATION OF ALLOSTERIC COUPLING IN BETA-2 ADRENERGIC RECEPTOR THROUGH MOLECULAR DYNAMICS SIMULATIONS

The purpose of this thesis is to investigate the allosteric coupling between intra- and extracellular parts of G-protein coupled human β_2 adrenergic receptor (β_2AR) in the presence of intracellular loop ICL3, which is missing in all crystallographic structures. In a recent study, a 1 μs MD run showed that closure of the G-protein binding site due to tight packing and closure of ICL3 under the receptor was coupled with the expansion of the ligand binding site. In this study, two independent 500 ns runs, which started from the final snapshot of the original 1 μs MD run with closed ICL3, showed that ICL3 stayed mostly in closed position, pointing out that it is a stable state blocking the binding cavity of G-protein. This allosteric coupling between ICL3 opening/closure and ligand binding site compression/expansion was further investigated by imposing specific distance constraints on the ligand binding site of the receptor. A total of seven constraints were applied between the residues on helices 3, 5, 6 and 7 that form the binding pocket. Specifically, these constraints were between residue couples Ser203O^{γ}-Asp113C^{γ}, Ser204O^{γ}-Asp113C^{γ}, Ser207O^{γ}-Asp113C^{γ}, Asn293C^{β}-Asp113C^{β}, Phe289C^{β}-Asp113C^{β}, Asn312C^{β}-Asp113C^{β}, and Phe289C^{β}-Asp312C^{β}. Constraining the binding site to an open position forced ICL3 to close from its fully open position. Constraining the binding site to a relatively closed position kept ICL3 around its open form. Using one extra constraint on backbone atoms (between residues Ser207C^{α}-Asp113C^{α} of helices 5 and 3) facilitated the opening of the binding site, and thus accelerated the closure of ICL3 from a relatively open position. However, attempting to compress the binding pocket did not lead to a fully open ICL3. Through this work, the allosteric coupling between the ligand binding regions in the extracellular parts of H5 and H6 and the intracellular part of the receptor including ICL3 was demonstrated one more time.

ÖZET

BETA2-ADRENERJİK RESEPTÖRÜNDE ALLOSTERİK EŞLEŞMENİN MOLEKÜLER DİNAMİK SİMÜLASYONLARIYLA ARAŞTIRILMASI

Bu tezin amacı G-proteine bağlanan beta-2 adrenerjik reseptör proteininin (β_2AR) hücre içine ve dışına bakan kısımları arasındaki allosteric eşleşmenin bütün kristalografik yapılarda eksik olan hücre içine bakan düğümü ICL3'ün varolduğu durumda incelenmesidir. Önceden, 1 μs'lik bir MD gidişizinin (*Ozcan et al.,* 2013) gösterdiğine göre, ICL3'ün reseptörün altına sıkıca paketlenip G-protein bağlanma bölgesini kapatması ligand bağlanma bölgesinin genişlemesiyle ilişkilendirilmiştir. Bu çalışmada, kapalı ICL3'e sahip orjinal 1 μs MD gidişizinin son konformasyonundan başlayan iki bağımsız 500 ns gidişizine göre, ICL3 daha çok G-proteinin bağlanma boşluğunu bloke edip, kararlı bir durum olan kapalı pozisyonu tercih ettiği gösterilmiştir. Ayrıca bu çalışmada, ICL3'ün açılması/kapanması ve ligand bağlanma bölgesinin sıkıştırılması/genişletilmesi arasındaki bu allosterik eşleşme, reseptörün ligand bağlanma bölgesine belirli uzaklık kısıtlamaları uygulayarak daha detaylı incelenmiştir. Bağlanma bölgesini oluşturan 3, 5, 6 ve 7. sarmallar üzerinde bulunan rezidüler arasına toplamda yedi uzaklık kısıtlaması getirilmiştir. Bu kısıtlandırmalar Ser203O^{γ}-Asp113C^{γ}, Ser204O^{γ}-Asp113C^{γ}, Ser207O^{γ}-Asp113C^{γ}, Asn293C^{β}-Asp113C^{β}, Phe289C^{β}-Asp113C^{β}, Asn312C^{β}-Asp113C^{β}, and Phe289C^{β}-Asp312C^{β} rezidü çiftleri arasında uygulanmıştır. Bağlanma bölgesini açık bir pozisyonda kısıtlandırmak ICL3'ü tamamen açık pozisyonundan kapalı bir pozisyona geçmeye zorlamıştır. Bağlanma bölgesini nispeten kapalı bir pozisyonda kısıtlandırmak ise ICL3'ü açık pozisyonu etrafında tutmuştur. Omurga atomlarına (sarmal 5 ve 3'te bulunan Ser207C^{α}-Asp113C^{α} rezidüleri arası) uygulanan ekstra bir uzaklık kısıtlandırması bağlanma bölgesinin açılmasını kolaylaştırıp, nispeten açık bir pozisyonda duran ICL3'ün kapanmasını hızlandırmıştır. Fakat, bağlanma bölgesini daraltmaya çalışmak ICL3'ün tamamen açılmasını sağlayamamıştır. Bu çalışma sayesinde, 5. ve 6. sarmalların hücre dışına bakan ligand bağlanma bölgeleri ile ICL3'ün de içinde olduğu hücre içine bakan bölgesi arasındaki allosterik eşleşme bir kez daha gösterilmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

a_i	Acceleration of particle i
A	Eigenvector group
C^{α}	Alpha carbon atom
d	Derivative operation
E	Energy
E_{bonded}	Internal energy including angles bond rotations and bond stretching
$E_{electrostatic}$	Energy of electrostatic interactions
$E_{non-bonded}$	External energy including electrostatic and Van der Waals interaction energies
$E_{van\ der\ walls}$	Energy of Van der Waals interactions
F_i	Force acting on atom *i*
j	PCA mode indice
k	Force constant
k_b	Bond force constant
k_θ	Force constant of the instantaneous angle θ
k_ϕ	Torsion force constant
l	Bond length
l_0	Reference bond length
m_i	Molecular mass of atom *i*
n	Function multiplicity
N	Number of C^{α} atom pairs
N_p	Number of point charges
p	Variance
q_i	Charges of atom *i*
q_j	Charges of atom *j*
r_{ij}	Distance between atoms *i* and *j*
r_{ref}	Reference atom position
$r_{t,i}$	Position of the atom *i* at time *t*
$\overline{r_i}$	Average position

s_i^2	Variance of the *i*th variable
s_{ij}	Covariance between the *i*th and *j*th variables
S	Covariance matrix
t	Time
T	Temperature
T_s	Total number of snapshots
u_i	Vector
u_j^A	Eigenvectors of *A*
u_{jk}^{AB}	Overlap matrix
u_k	Eigenvector
u_k^B	Eigenvectors of *B*
U	Potential energy
U_m	Matrix
v_i	Speed of the particle i
x	Distance
x_m	Data matrix of the C^α coordinates
$\overline{x_m}$	Time average of the C^α coordinates in a single run
x_p	*p* correlated variable
x_{ref}	Specified restraint distance
z_p	*p* new uncorrelated variable
α	Alpha
β	Beta
γ	Gamma
ε_0	Effective dielectric constant for the medium
ε_{ij}	Well depth of the potential between atoms *i* and *j*
θ	Instantaneous angle of the bond
θ_0	Equilibrium angle of the bond
δ	Integration operator
δ_p	Phase shift
Δ	Difference operator
σ_{ij}	Finite distance at which the potential between atoms *i* and *j* is zero

ϕ	Torsional angle
$\nabla_{r_i} U$	Gradient of the potential energy U with respect to r_i

LIST OF ACRONYMS / ABBREVIATIONS

3D	Three-dimensional
Å	Angstrom
Asn	Asparagine
Asp	Aspartic Acid
ANM	Anisotropic Network Model
β_2AR	Human Beta-2 Adrenergic Receptor
DNA	Deoxyribonucleic Acid
ECL2	Extracellular Loop Two
ECL3	Extracellular Loop Three
GPCR	G-protein Coupled Receptor
G_s	Alpha Subunit of G-protein
H	Helix
ICL2	Intracellular Loop Two
ICL3	Intracellualr Loop Three
MD	Molecular Dynamics
MOE	Molecular Operating Environment
MSM	Markov State Model
NMR	Nuclear Magnetic Resonance
OPM	Orientations of Proteins in Membranes
PCA	Principal Component Analysis
PDB	Protein Data Bank
Phe	Phenylalanine
RMSD	Root Mean Square Devation
RMSF	Root Mean Sqaure Fluctuation
Ser	Serine
GPCR	G-protein-coupled Receptor
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine

1. INTRODUCTION

GPCRs are the largest family of eukaryotic signal transduction proteins. They communicate signals across cell membranes responding to a variety of extracellular stimuli such as peptides, light, proteins, ions, hormones and small molecules. Upon activation, they trigger responses inside the cell, interacting with their G-protein couples. The structure of a GPCR recognizing a diffusible ligand can help for the design of new drugs but cannot explain how it triggers the internal signaling pathways inside the cell. The changes in the receptor conformation upon activation are crucial to investigate. Many GPCRs are thought to exist in dimer forms in the plasma membrane, but the exact location of the subunit interface region is still in debate. Besides, β_2 adrenergic receptor (β_2AR) couples to G_s efficiently as a monomer so it can be concluded that dimerization reduces the efficiency of β_2AR coupling to G_s (Cherezov *et al.,* 2007). All GPCRs include seven transmembrane-spanning alpha helices, an intracellular C terminus, and extracellular N terminus and three inter-helical loops. Human GPCRs can be classified into five main families of receptors which are glutamate, secretin, adhesion and frizzled-taste-2, and rhodopsin, which is the most studied and largest family (Oldham *et al.,* 2008).

Drugs acting on GPCRs, especially class A adrenergic receptors responding to hormones adrenaline and noradrenaline, are in the mainstream of research because these receptors are the targets of current cardiac and asthma drugs with several side effects. The first explained crystallographic structures of human β_2AR were in the inactive state, two with Fab5 complex in monoclonal antibody assisted crystallization (Rasmussen *et al.,* 2007, PDB ids: 2R4R, 2R4S) and one with the inverse agonist carazolol and T4-lysozyme (mutant) in 2007 (Cherezov *et al.,* 2007; PDB id: 2RH1, respectively). A year later, a cholesterol bound form of the inactive state with the inverse agonist timolol and T4-lysozyme (mutant) was discovered (PDB id: 3D4S, Hanson *et al*., 2008). Later, four more inactive states were found in 2010; 3NY8 with the inverse agonist ICI118551, 3NY9 with an inverse agonist, 3NYA with the antagonist alprenolol (Wacker *et al*., 2010) and 3KJ6 with an unresolved inverse agonist in monoclonal antibody assisted crystallization (Bokoch *et al.,* 2010). Finally, 3 active states, which are 3P0G with the agonist BI-167107 and G-protein mimic nanobody bound on the intracellular side of the activated form of the

receptor, 3PDS with the agonist FAUC50 which is a covalently bound ligand and 3SN6 engineered by N-term fusion with the agonist BI-167107, $G^{\alpha\beta\gamma}$ and nanobody bound to the activated form of the receptor, were discovered with T4-lysozyme (Katritch *et al.*, 2012). Recently, three more active states with agonists were explained; 4LDE, 4LDL and 4LDO coupled with camelid antibody fragment (Ring *et al.*, 2013). Although an impressive amount of X-ray crystallographic structures was revealed, the conformational transition between the inactive and active states is still not fully resolved (Dror *et al.*, 2011).

Ozcan *et al.* used the inactive crystal structure of the human β_2AR in complex with carazolol with PDB id of 2RH1 in order to understand the effect of the intracellular loop 3 (ICL3) on the intrinsic dynamics of the receptor. Carazolol and T4-lysozyme were removed prior to the molecular dynamics (MD) simulations, which were performed on two different models called *loop* and *clipped* models. In *loop* model, the missing ICL3 region was modeled, while in the *clipped* model the two open ends of the transmembrane helices 5 (H5) and 6 (H6) were covalently bonded to each other. Both models embedded into POPC bilayer membrane with explicit water were subjected to a 1 μs MD simulation at 310 K. As a result, *loop* model found a very inactive conformation after around 600 ns, when H6 moved towards the receptor's core region and ICL3 packed underneath the membrane blocking the G-protein's binding site completely. Meanwhile, the Ser207-Asp113 distance at the binding site of the receptor's extracellular region increased from 11 Å to 18 Å. This result showed that there is a strong coupling between the extracellular and intracellular regions of the receptor. In *clipped* model, no such transition to the inactive state was observed (Ozcan *et al.*, 2013).

In this thesis, atomistic molecular dynamics (MD) simulations were conducted on the same *loop* model (Ozcan *et al.*, 2013) to explore the conformational dynamics of the apo form of the receptor. The aim was to observe the dynamics of the intracellular part of the receptor including the ICL3 region by MD simulations under various distance constraints, which were applied to the ligand binding site of the receptor and to confirm the coupling between the intracellular and extracellular parts of the receptor. The intracellular loop ICL3, which has an important role in G-protein recognition but is missing in crystal structures, is present as an unstructured loop in all runs.

The other chapters in this thesis are Chapter 2, where theoretical background on GPCR and β_2AR is given, Chapter 3 in which methods are explained for analysis of molecular dynamics simulations, Chapter 4 giving results and discussion of MD simulations' analyses, and lastly Chapter 5 including conclusion and future studies.

2. THEORETICAL BACKGROUND

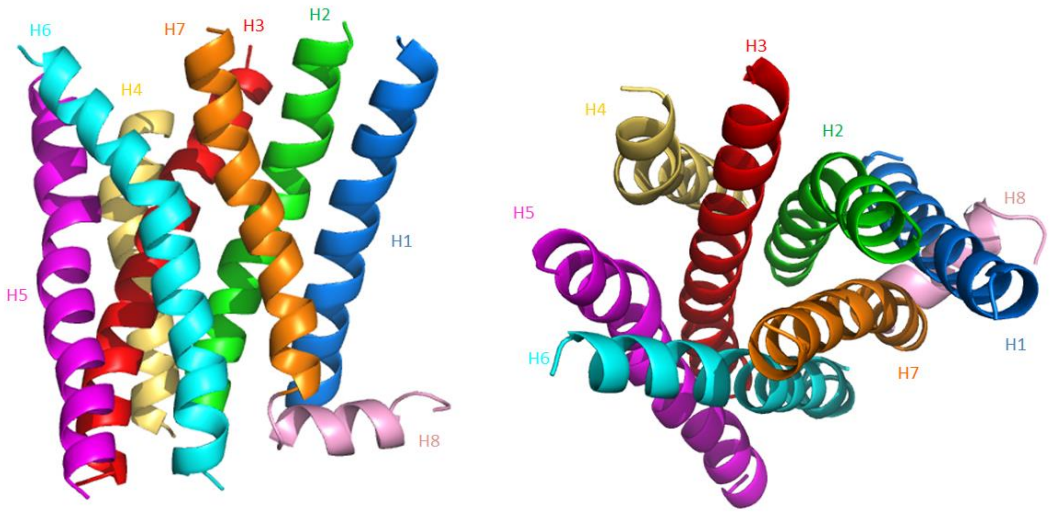
2.1. GPCR Superfamily

The largest protein superfamily in mammalian genomes belongs to G-protein coupled receptors (GPCRs). Different extracellular ligands have led to the diversity of more than 800 human GPCRs. Most physiological processes such as smell, taste, vision, cardiovascular, endocrine, reproductive and neurological functions are dependent on GPCR signaling. As of 2014, crystallographic studies have resolved 16 different receptors. Biochemical and biophysical techniques such as hydrogen-deuterium exchange coupled with mass spectroscopy and nuclear magnetic resonance have helped to investigate the ligand-dependent dynamic equilibrium among different functional states. High-resolution structures reveal that the receptors are controlled also by lipids, ions and cholesterol as well as ligands (Katritch *et al.,* 2013). In order to understand the conformations of the active states, many biochemical and physicochemical approaches have been used. Some of them are disulfide cross-linking, site-directed spin labeling and engineered zinc binding (Simpson *et al.*, 2011).

2.2. GPCR Structure

GPCR superfamily is a collection of proteins, which have common functional and structural characteristics but lack obvious sequence similarity. The sequences of more than 1000 different GPCRs are already known. They are composed of a single chain that is usually 400-500 amino acids long. The major similarity of GPCRs is the presence of seven hydrophobic segments, which are separated from each other by hydrophilic segments of about 20-25 amino acids long. Thus, GPCRs have seven transmembrane-spanning α helices, also called TM domains, connected by intracellular and extracellular loops. The carboxy terminus is located on the intracellular side, while the amino terminus is on the extracellular side. Helices are organized sequentially in a counter-clockwise mode, which forms a flattened circle around a central pocket seen from the extracellular region. H3 is almost in the center of the molecule and is tilted. The helical order opens up and makes a cavity for the ligand to bind on the extracellular region (Vauquelin, 2007). The three

dimensional structure of a GPCR called 2RH1 is demonstrated in Figure 2.1 from side and top views with color coded cartoon representations of helices using PyMOL (DeLano, 2004; DeLano and Lam, 2005).



(a) (b)

Figure 2.1. Three dimensional structure of GPCRs (2RH1) (a) from side view, (b) from top view.

GPCRs were divided into five main families, whose protein sequences share more than 20% sequence identity in their transmembrane domains (Vauquelin, 2007). The largest one is family A (class I), which is divided into groups such as opsin, peptide, amine, cannabinoid and olfactory receptors according to ligand specificity (Cherezov *et al.*, 2007). Being the rhodopsin-like receptor class, it is also the most studied one. A disulfide bridge connects the second and third extracellular loops. Among all class A receptors, the overall homology is low and restricted to a number of highly conserved key residues. Their conservation points out their essential role for the functional and structural coherence of the receptors (Vauquelin, 2007). Usually, allosteric sites are located in less-conserved regions of the protein compared to orthosteric sites. Therefore allosteric sites are less amenable for homology modeling (Katritch *et al.*, 2012).

Until 2007, rhodopsin (Palczewski *et al.*, 2000) was the only GPCR, whose structure was found out by X-ray crystallography from diffraction data. Progresses in protein engineering and stabilization led to discoveries of other GPCR models. In 2007, inactive

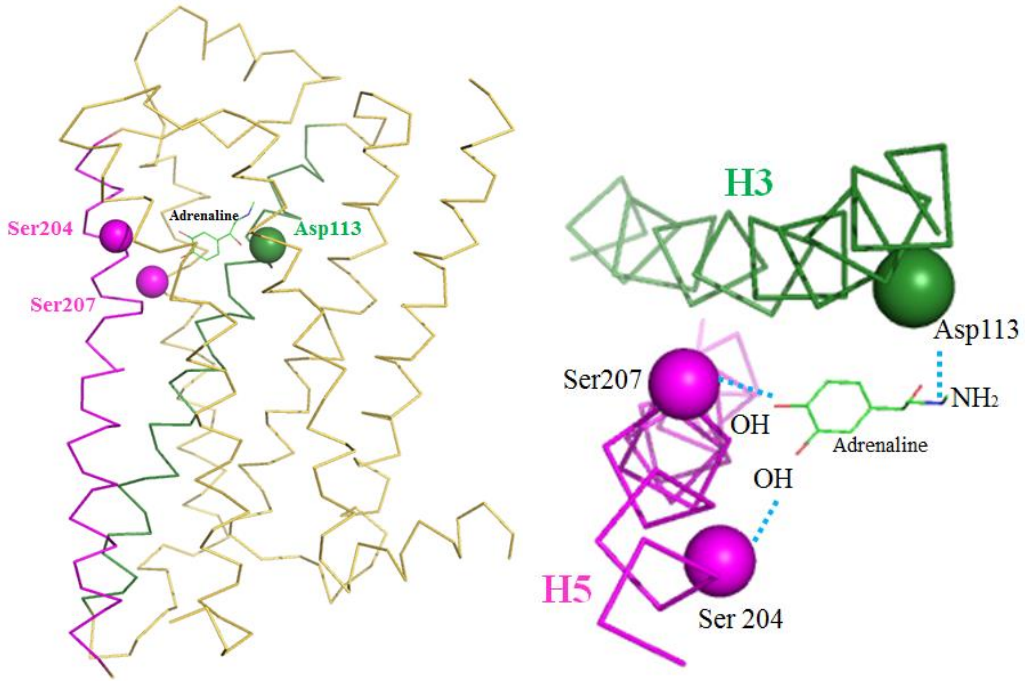
structures of β_2AR from family A were discovered. Also, a number of structures with bound agonists corresponding to active states were observed for rhodopsin, β_2AR and A_{2A} adenosine receptor (A_{2A}AR), too. The structures of active β_2AR and A_{2A}AR show agonist-induced conformational changes similar to rhodopsin in the binding pocket. Small local structural changes in the binding pocket are revealed by the crystal structure of the nanobody-stabilized β_2AR. The differences between the active and inactive structures are very trivial for the binding pocket, but the main difference is seen in the lower regions of the protein. The differences in the binding pocket are related to changes in an interhelical hydrogen bond network involving residues in the transmembrane helices H3, H5, H6 and H7. The agonist interacting with Ser203 and Ser207 stabilizes a receptor conformation showing an inward movement of H5 at Ser207 and Pro211 relative to the inactive structure (Deupi *et al*., 2011).

2.3. Ligand Interactions with Family A Receptors

The chemical diversity among the endogenous ligands is very large, including peptides, lipids, ions, biogenic amines, nucleotides, proteases and glycoproteins. An exogenous stimulus such as odor, taste and light is also recognized by GPCRs. Ligand size is very effective on location and nature of binding. Small molecules such as pharmacological agents bind within the transmembrane region of the receptor, while large ligands such as peptides and proteins bind to the extracellular loop scaffold. Peptides can show a mixed binding in which they bind mainly to the extracellular loops while some of the structure gets into the transmembrane region (Vauquelin, 2007).

During its ligand attachment, rhodopsin is different than the other GPCRs in terms of the binding crevice formed by the transmembrane helices. This may be the reason for rhodopsin's very rapid response to light. Small molecule transmitters, such as epinephrine, norepinephrine, serotonin, acetylcoline, histamine and dopamine, have binding sites in the central pocket of their GPCRs. Therefore, these transmitters contact with amino acid residues of especially the transmembrane helices H3, H5, H6 and H7. Mutation studies with β_2AR have shown that at least H3 and H5 are related to the binding of agonists, such as isopropanol and adrenaline. The secondary amine interacts with Asp113 in H3 and the hydroxyls on the aromatic ring interact with Ser204 and Ser207 in H5 as shown in Figure

2.2 from side and top views. Phe289 and Phe290 on H6 may make π-stacking interactions with the catechol ring. Asp113 in H3 of β_2AR is conserved among the biogenic amine receptors. On the other hand, the large-sized peptide messengers bind to extracellular domains including the three extracellular loops and the N-terminal part of GPCRs (Vauquelin, 2007). Large glycoprotein hormones bind to a domain, which is Leu-rich and has a large N-terminal (Oldham *et al.*, 2008).



(a) (b)

Figure 2.2. Adrenaline bound β_2AR from (a) side and (b) top views.

According to their effects on receptor signaling, the orthosteric ligands are classified as agonists, partial agonists, inverse agonists and neutral antagonists. Agonists maximize the activity of the receptor and stabilize it by shifting the equilibrium to a more active conformation. They promote G-protein binding to the receptor. Many GPCRs show some basal activity independent of the agonist. Partial agonists cause submaximal activity. Inverse agonists change the conformation to a more inactive state, thereby decreasing the basal activity. Neutral antagonists do not affect the baseline activity of the receptor, while blocking the activity of other ligands sterically (Rasmussen *et al.*, 2011).

Kohlhoff *et al*. found out that agonist-bound simulations stabilize active-like conformations, while inverse agonist and apo simulations do not show active state conformations at all. The agonist BI-167107 strengthened correlations between intracellular and extracellular residue groups, while the inverse agonist carazolol disconnected them. The apo receptor showed only irrelevant correlations. In summary, it was shown that ligands modulate the receptor dynamics to choose different signaling pathways (Kohlhoff *et al.,* 2013).

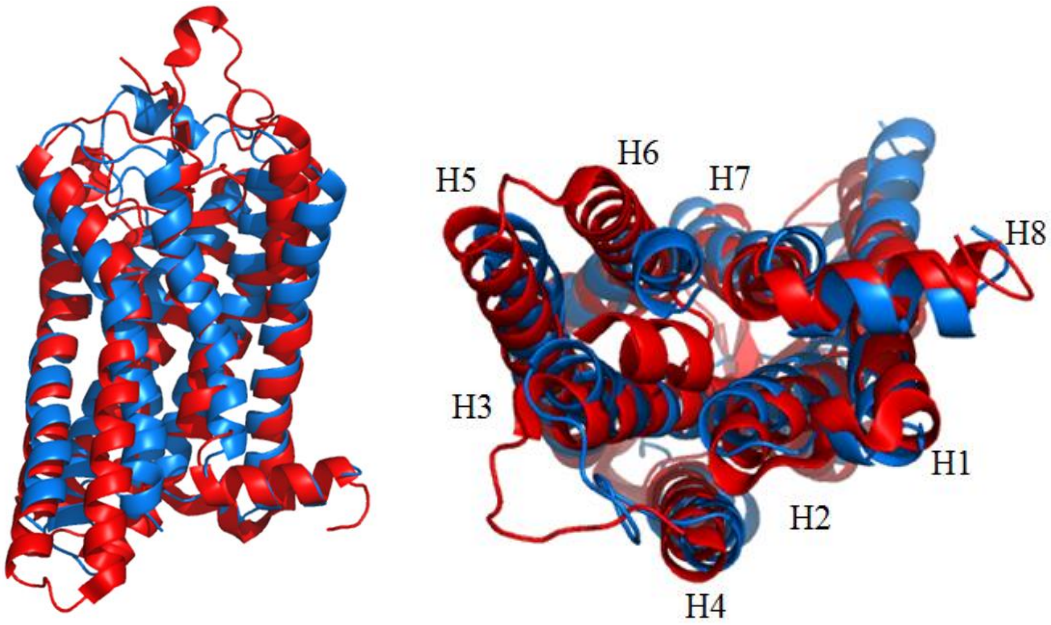
Vanni and his colleagues applied MD simulations to β1 and β2 adrenergic receptors bound to the full agonist isoprenaline and in their unliganded form. They determined a novel agonist binding mode differing from poses of in silico docking studies on rigid receptors and from the one for antagonists in the crystal structures. The stabilization of novel interactions between ligand and receptor both at the interface of helices 5 and 6 is contributed by internal water molecules. The interaction network between isoprenaline and the anchor site is different between agonists and inverse agonists/antagonists because of two water molecules entering the cavity and contributing to the stabilization of a network of interactions (Vanni *et al.,* 2011).

2.4. Receptor Activation

Receptors are transmission channels for the chemical signals into the cell (Oldham *et al.,* 2008). Among GPCRs, the molecular mechanism of activation differs to a great extent. It is believed that in the ground state, a network of intramolecular interactions constrain the receptors in an inactive conformation. An agonist releases these inactivating constraints and reinforces the activating interactions to form and/or stabilize the active state. The activation process probably includes a multistep sequence. It has been proposed that peptides create activator contacts while small molecules destabilize constraining interactions. The protonated state of the aspartic acid/glutamic acid at the cytoplasmic end of H3 has been suggested to be one of the key events in the activation of β_2AR. Biochemical and spectroscopic approaches to investigate activation of the receptor include the generation of artificial bridges between two helices to prevent their separation, introducing pairs of histidines to form zinc-binding sites, observing receptors with

cysteines to find out whether the environment of a side-chain is aqueous or hydrophobic (Vauquelin, 2007).

Conformational sampling and the energetics of activation can be different among GPCRs. The conformational equilibria of β_2AR and rhodopsin are different due to the fact that β_2AR cannot take a fully active conformation in the absence of G-protein, while rhodopsin can. Their alignment is shown in Figure 2.3 from side and bottom views in which rhodopsin (pdb id: 3PQR) is red and β_2AR (pdb id: 2RH1) without T4-lysozyme chimera is blue. It is enough for an agonist to interfere with only one key intramolecular interaction necessary for the stabilization of the inactive state. Thus, these stabilizing interactions by agonists lowers the energy barrier between the inactive and active states and the probability of a receptor interacting with a G-protein is increased (Rasmussen *et al.*, 2011).



(a) (b)

Figure 2.3. Rhodopsin (PDB id: 3PQR, red) and β_2AR (PDB id: 2RH1, blue) aligned from (a) side and (b) bottom views.

G-protein activation occurs through the conformational changes of the GPCR on whose extracellular surface an agonist binds. The common result regardless of the ligand binding mode is that the conformational change occurs in the cytoplasmic regions of the receptor. An outward movement of H6 in the intracellular region is the result of receptor

activation according to the site-directed spin labelling experiments of bovine rhodopsin. As a result, a crevice opens up within the intracellular surface of the receptor (Oldham *et al.*, 2008). Similar results have been found for β_2AR (Jensen *et al.*, 2001). Another observation is on H5 moving outward. The largest movement is 11-14 Å in H6 of β_2AR complexes with nanobody and G-protein. Figure 2.4 demonstrates alignment of 2RH1 (red) to 3SN6 (cyan), which is used as the active state crystallographic structure in this thesis, from a lower view. The movement of H6 can be observed clearly. Conformational changes in H3 and H7 depend on the receptor and the bound ligand, which may play a role in G-protein binding (Rasmussen *et al.*, 2011).

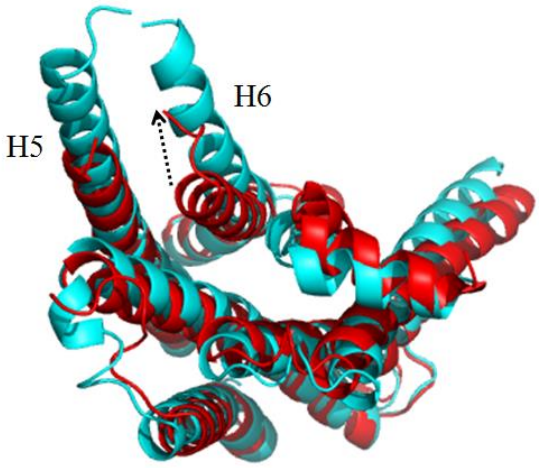


Figure 2.4. Alignment of 2RH1 (red) to 3SN6 (cyan) from a lower view.

Dror and his coworkers (Dror *et al*., 2011) have proposed that small perturbations at the extracellular drug-binding site are linked to large conformational changes at the intracellular G-protein binding site by a coupled allosteric network of three regions, each with multiple distinct conformations. Their molecular dynamics (MD) simulation studies have shown that the activation process starts with the binding of a G-protein to an inactive receptor. They have also observed that the first structural changes occur in the intracellular part of the receptor with the outward motion of H6 during receptor activation. The agonist-bound β_2AR changes from an active to an inactive state upon removal of the G-protein. This finding confirms that in the absence of an intracellular binding partner the energetically favored conformation of agonist-bound β_2AR agrees with the inactive state determined by crystallography. Moreover, three functionally relevant regions, which are the ligand-binding site, the G-protein binding site and a connector region between them, have been found to be loosely coupled. This situation explains why it is very difficult to

stabilize β_2AR in a homogeneous active state with an agonist. The active state of the ligand-binding site caused by an agonist lock is not enough to lock the whole receptor in its active state. Besides, ICL3 normally interacts with the G-protein and may influence the activation mechanism (Dror *et al.*, 2011).

An apo β_2AR missing ICL3 was investigated in a 1.02 μs all-atom simulation by Romo *et al.* in order to understand the inactive structure. Although the structure was rigid, a rapid influx of water into the core of the protein and a slight expansion of the protein compared to the crystal structure were observed. The structure returns to its inactive conformation. Also, the authors observed a single activation-like event in which the intracellular end of H6 tightened into a 3-10 helix and moved the side chain Glu268 far from its salt bridge partner. The helix returned to a more ordinary α-helical state roughly around 200 ns later, and the salt bridge continued its previous noisy association (Romo *et al*., 2010).

Simpson and his coworkers (Simpson *et al*., 2011) generated an active conformation of β_2AR combining the information from GPCR and opsin crystal structures. In MD simulations, they used experimental data obtained from site-directed spin labeling and zinc binding studies. In the simulations, initial velocities were given at low temperature to a minimized model and new velocities were given at slightly higher temperatures until 310 K was reached. As a result, intracellular ends of H5 and H6 closed the cavity between H3, H5 and H6, and moved inward. This result may be due to the lack of G-protein that binds to this cavity during receptor activation. Therefore, harmonic distance restraints were applied in order to keep the distances at the intracellular end of the TM bundle constant between H3, H5 and H6. The experiments of manual docking agreed on the produced model. Site-directed mutagenesis experiments confirmed the interacting residues for both agonist binding to the orthosteric site and the binding of the C terminal peptide of Gsα to the intracellular region of the receptor. Virtual screening and CHARMM-based manual docking interactions validate the active model (Simpson *et al.*, 2011).

An agonist-bound, active-state crystal structure of the human β_2AR-nanobody (camelid antibody fragment) complex exhibiting G-protein-like behavior was generated by Rasmussen and his coworkers (Rasmussen *et al*., 2011). Subtle changes in the binding

pocket were revealed when it was compared with the inactive β_2AR structure. These small changes are related to an outward movement of the cytoplasmic end of H6 and rearrangements of H5 and H7 similar to those observed in opsin, an active form of rhodopsin (Rasmussen *et al.*, 2011).

In order to understand the mechanism of β_2AR activation, Kohlhoff and his coworkers (Kohlhoff *et al.*, 2013) used an alternative approach to MD using Google's Exacycle cloud-computing to simulate two milliseconds of β_2AR dynamics. MD simulations describe intrinsic receptor dynamics in atomistic detail, and Markov State Models (MSM) give a summarized view and enable to find out the key conformational states of the receptor. Most of the simulations were done without ICL3 loop since β_2AR remains functional even in its absence. Both active and inactive crystal structures of β_2AR were used. After generating 150 μs trajectories, it was found out that activation and deactivation go through multiple pathways and pass through metastable intermediate states. Also, MSMs simplify the details of 3000-states model by discarding fast conformational dynamics in order to give lumped and kinetically relevant details of receptor dynamics with 10-state model. During activation, H6 moved away from H3 (Kohlhoff *et al.*, 2013).

In summary, H3 and H6 play crucial roles for transition of β_2AR to the fully active state. H3 is long and tilted, thus it makes multiple interhelical contacts. These helices rotate counter-clockwise and their cytoplasmic ends move outward, enabling the G-proteins to interact with previously inaccessible ICL2, ICL3 and C-terminal residues (Vauquelin, 2007).

2.5. Interaction with G-Proteins

The key step towards G-protein activation and initiation of downstream signaling pathways is nucleotide exchange, which is GDP dissociation from the G-protein α-subunit. Inactive and active conformations of several heterotrimeric G-proteins were biochemically and biophysically studied intensely. However, the molecular mechanism of G-protein activation is still not fully comprehended (Chung *et al*., 2011).

Many studies showed that different ligands bound to the receptor cause GPCRs to change into different conformations, which in turn induces specific intracellular signaling pathways. Dror *et al.* suggested that in an activation process the G-protein may bind to the receptor when this binding site is in an intermediate state. Their simulation's intermediate conformation has a larger cleft in the G-protein-binding site (between the intracellular ends of helices 2, 3, 5, 6 and 7) compared to that of an active conformation. The transition from the intermediate to the active state may be prompted by this G-protein-binding (Dror *et al.*, 2011).

Gαsi, part of the Gs heterotrimers, and Gαi, part of the Gi heterotrimers, are the two G-proteins that β_2AR activates. They regulate adenylate cyclase which produces cyclic AMP (cAMP) activating G-protein kinase A (PKA). This kinase regulates β_2AR and Ca^{2+} channel. Specific phosphodiesterase proteins (PDEs) downregulate cAMP levels. β_2AR activation lets a GPCR kinase (GRK) phosphorylate β_2AR. It also enables enough arrestin coupling. Arrestin has many functions such as preventing the activation of G-proteins and activating extracellular signal-regulated kinases (ERK) (Rosenbaum *et al*., 2009).

In the absence of a G-protein, an agonist-bound active-state GPCR is difficult to obtain due to its instability. In order to obtain the active state conformation, Rasmussen and his coworkers generated a nanobody (Nb80) with G-protein-like behaviors to get an agonist-bound, active-state crystal structure of the protein (Rasmussen *et al.*, 2011). This β_2AR-T4L-Nb80 complex, where T4L stands for T4-lysozyme, was compared with the inactive β_2AR. The nanobody Nb80 did not recognize the inactive conformation of β_2AR but bound to β_2AR occupied by agonist. Also, its effect on β_2AR binding to the inverse agonist was little. The energetic coupling of agonist and G-protein binding was imitated by Nb80, which also stabilized a conformation of β_2AR very similar to that of G-protein. Due to the steric hindrance by T4L, G-protein coupling in β_2AR-T4L cannot be studied. Nevertheless, the results showed that T4L did not prevent Nb80 binding. In short, it was concluded that small changes around the agonist-binding area are related to the binding of Nb80 and G-protein. Since Nb80 and G-protein stabilize or induce similar structural changes in the protein, Nb80 and G-protein might recognize similar agonist stabilized conformations. Also, the fact that the transmembrane regions of β_2AR and rhodopsin show

similar structural changes upon activation supports that the agonist-bound β_2AR-T4L-Nb80 represents an active conformation (Rasmussen *et al.,* 2011).

In order to understand G-protein activation, Chung *et al*. applied peptide amide hydrogen-deuterium exchange mass spectrometry to observe changes in the heterotrimeric bovine G-protein Gs, while forming a complex with agonist-bound β_2AR. Gs receives higher levels of hydrogen-deuterium exchange than would be predicted from the crystal structure of β_2AR-Gs complex. Thus, the determined mechanism of nucleotide exchange includes the receptor disturbing the structure of the amino-terminal region of Gs' α-subunit and altering the "P-loop", which binds the β-phosphate in GDP. P-loop stabilization and β-phosphate coordination are key features in determination of GDP and GTP binding affinity (Chung *et al*., 2011).

Feng *et al*. performed 20 ns MD simulations of the β_2AR-Gs protein complex with agonist in a water and lipid environment to find out the activation mechanism. The interaction between β_2AR and Gs protein is stable with a nanobody. However, when there is no nanobody stabilizing the complex, the agonist triggers β_2AR's conformational changes from the extracellular region to the intracellular region. Especially, because of their direct interaction with Gs protein's α5-helix, the intracellular parts of H3, H5, H6 and H7 undergo changes. The residues related to the hydrophobic interactions between Gs protein and receptor also change their conformations (Feng *et al.,* 2012).

2.6. GPCR Dimerization

Dimerization can occur between identical proteins (homodimerization) or different proteins (heterodimerization). Many evidences were found out in recent years on GPCRs forming dimers, although it is not clear when and where they do so under physiological conditions. Dimerization is not necessary for ligand recognition or signaling, but it can facilitate many functions. Some of the reasons why a receptor wants to dimerize for ligand recognition are specificity through interactions of ligands with two subunits in a dimer, increased affinity through many contact points and increased ligand types via different binding partners. Dimerization also has roles in receptor activation such as the changes in distance between subunits and crossphosphorylation as an activation mechanism. Multiple

contact points and specificities for intracellular signaling-proteins and crossregulation of activity in protomers are the signal transduction mechanisms, which can favor dimerization. Dimerization also contributes to changes in cell surface delivery, crossregulation in cell surface mobility and receptor internalization (Lohse, 2010).

Interactions that hold a dimer together differs among the receptors. It can be either by covalent (disulfide) and/or non-covalent (hydrophobic) interactions. For example, dimerization of the metabotropic glutamate receptors depends on the formation of disulfide bonds between cysteines in their large amino-terminal domains. During the heterodimerization process, some receptors seem to swap some of their transmembrane domains. Co-expression of truncated β-adrenergic receptors with H1 to H5 and H6 to H7 restores its binding and functional activities such as adenylate cyclase stimulation (Vauquelin, 2007).

Homo- and hetero-dimerization are types of GPCR interactions, which can modulate the signaling properties of the receptors and intervene the relation between GPCR pathways (Katritch *et al.,* 2012). Modeling of dimerization has many technical challenges as the simulation system needs to be much larger. This increases the computational cost and reduces the length of the trajectory significantly. Since the protein system is larger, it will have slower fluctuation modes, requiring a longer trajectory to get equal statistical sampling. Even for monomeric systems, it is a significant problem to generate good statistics (Grossfield, 2011). Identification of the functionally relevant interfaces has been very difficult due to the transient mode of the interactions and technical problems of separating binding in a crowded membrane environment. Nonspecific and partial dimerization of GPCRs in crystallization studies can be risky for obtaining crystals because it brings out heterogeneity into the samples (Katritch *et al.,* 2012). Since the nature of the lipid environment can change the thermodynamic balance between monomer and dimer easily, the nature of the physiologically relevant state is not clear (Botelho *et al.,* 2006).

2.7. Class A Family GPCR: β2-Adrenergic Receptors

GPCRs bind their endogenous ligands in different binding modes, even within class A to which β_2AR belongs as one of the most researched GPCR (Simpson *et al.,* 2011).

Adrenergic receptors are in the amine group, consisting of two main subfamilies α and β. They differ in ligand specificity, tissue localization, G-protein coupling and the following downstream mechanisms. Diseases such as hypertension, asthma and heart failure are seen due to genetic modifications of adrenergic receptors (Cherezov *et al.,* 2007). β_2AR is crucial in medicinal chemistry such that β_2ARs expressed on the airway smooth muscle are targets for β agonists used in the treatment of asthma. It is also widely used in signal transduction studies (Simpson *et al.,* 2011). Large ligands such as peptides bind closer to the membrane surface near the ECLs, while smaller ligands bind deeper in the space which the transmembrane domain helices create (Cherezov *et al.,* 2007).

β_2AR has seven transmembrane helices (H) with intracellular (ICL) and extracellular (ECL) loops in between. They are H1 (32-60), ICL1 (61-66), H2 (67-96), ECL1 (97-102), H3 (103-136), ICL2 (137-146), H4 (147-171), ECL2 (172-196), H5 (197-229), ICL3 (230-266), H6 (267-298), ECL3 (299-304), H7 (305-328) and a horizontal small helix H8 (330-341), which are common to all rhodopsin-like GPCRs. Also there is a short helical piece in the middle of ECL2; not present in rhodopsin. Helices H2, H5, H6 and H7 have proline-induced kinks at conserved positions. The kinks are thought to play a role in G-protein activation through structural rearrangements.

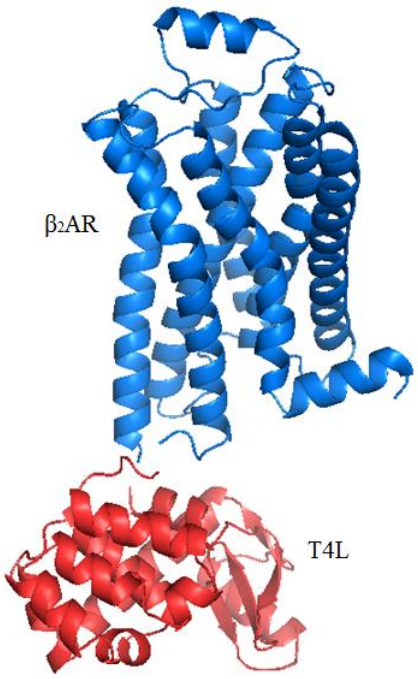


Figure 2.5. High resolution crystal structure of β_2AR-T4L.

In the crystal structure of β_2AR-T4L (R2H1), electron density was excellent for residues 29 to 342 of β_2AR, as well as the ligand carazolol and the two disulfide bonds Cys106-Cys191 and Cys184-Cys190. The electron density quality was lower for the other regions, which are the N terminus between 1-28 and the majority of the C terminus between 343-365. These regions were not detected due to invisibility and disorder (Cherezov *et al.*, 2007). In H3, residues Asp130 and Arg131 form a salt bridge and it stays intact in the active β_2AR nanobody complex. The other salt bridges are formed between Arg131-Glu268, and Asp192-Lys305, connecting ECL2 and ECL3 (Bokoch *et al.*, 2010). Also, hydrogen bond interactions found between Arg131 side chain and polar residues in H6 may have an important role in receptor signaling regulation (Katritch *et al*., 2013).

Since it is a challenge to obtain high-resolution structural data for membrane proteins, Cherezov and his coworkers engineered β_2AR and inserted T4 lysozyme in place of ICL3 in order to reduce conformational heterogeneity and facilitate crystal nucleation (Cherezov *et al.*, 2007).They used a diffusible ligand, the partial inverse agonist carazolol. ECL2 was kept out of the binding cavity by a pair of closely spaced disulfide bridges. Also, a short helical segment belonging to ECL2 helped the ligand binding (Cherezov *et al.*, 2007).

3. METHODS AND MATERIALS

In this section, theoretical background of molecular dynamics (MD), constrained MD, CHARMM forcefield and analysis tools like Principal Component Analysis (PCA) will be given.

3.1. MD Simulations

MD simulations are used to predict the time-dependent events in a molecular system. Atoms are permitted to interact with each other in MD simulations using forcefields or potential energy functions. The integration of Newton's equation of motion supplies the constitution of the molecular system's trajectory (Gunsteren and Berendsen, 1990):

$$F_i(t) = m_i \times a_i(t) = m_i \times \frac{dv_i(t)}{dt} = m_i \times \frac{d^2 r_i(t)}{dt^2} \tag{3.1}$$

$$F_i = -\nabla_{r_i} U \tag{3.2}$$

In Equation 3.1, F_i is the force acting on atom *i* exerted by all the molecules, m_i is the molecular mass of the particle, a_i is the acceleration of the particle, v_i is the speed of the particle *i*, r_i is the position of the atom *i* and *t* is time (Gunsteren and Berendsen, 1990). *F* is calculated from the gradient of the potential energy *U* with respect to r_i as in Equation 3.2.

The initial distribution of velocities, the initial positions of the atoms and the acceleration designated by the potential energy function's gradient are needed in order to calculate the trajectory. The velocities and positions at time *t* are determined by the velocities and positions at the previous time step, meaning that the equations of motion are deterministic. Experimental structures such as the solution structure, which are resolved by NMR spectroscopy or the x-ray crystal structure of the protein, give the initial positions (Stote *et al.*, 1999).

In order to calculate the initial distribution of velocities, Boltzmann distribution with the magnitudes conforming to the required temperature is used and corrected so that there is no overall momentum (Stote *et al.,* 1999).

Verlet algorithm, based on the addition and subtraction of the Taylor series expansions, calculates new positions at time *t+dt* by using accelerations at time *t* and positions from time *t-dt* (Stote *et al.,* 1999):

$$r_i(t + \delta t) \approx 2r_i(t) - r_i(t - \delta t) + \frac{\Delta t^2}{m_i} F_i \tag{3.3}$$

One handicap of Verlet is that the updating of velocities is one step behind. The velocities at time t can be calculated only after the positions at time $t + \delta t$ become available (Binder *et al.,* 2004). The "velocity Verlet Algorithm" synchronizes the updating velocities and positions (Stote *et al.,* 1999).

3.2. ForceField: CHARMM

CHARMM is a force field treating molecules in solutions, molecules in crystalline solids and isolated molecules (Brooks *et al*., 1983). The original nucleic acid force field related to the program CHARMM was designed to be used in vacuo. The next nucleic acid force field in CHARMM22 was designed to contain an explicit representation of the solvent. One of the popular releases of CHARMM force field for nucleic acids is called CHARMM27 which represents a complete reoptimization of the CHARMM22 force field and emphasizes dihedral and nonbonded parameters. It has been used in MD studies on DNA, DNA-lipid systems and DNA-protein complexes (MacKerell *et al.,* 2001). The latest release is in 2014 with the name CHARMM40.

While calculating CHARMM potential energy, the energy denoted by *E* is a function of the atomic positions *r* of all the atoms in the system expressed by Cartesian coordinates. Potential energy U is calculated by adding E_{bonded}, internal energy including the angles, bond rotations (torsions) and bond stretching in a molecule, to $E_{non-bonded}$, external

energy accounting for interactions between atoms separated by three or more covalent bonds or nonbonded atoms (Stote *et al.*, 1999):

$$U = E_{bonded} + E_{non-bonded} \tag{3.4}$$

E_{bonded} can be explained as:

$$E_{bonded} = \sum_{bonds} k_b(l - l_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} k_\phi \left(1 + \cos(n\phi - \delta_p)\right) \tag{3.5}$$

In Equation 3.5, the first term, where l is the bond length, l_0 is the reference bond length and k_b is bond force constant, is modelled by a harmonic potential giving the increase in energy. The second term sums up the overall valence angles in the molecule modelled again using a harmonic potential based on Hooke's Law. A valence angle is the one formed between the atoms A, B and C when A and C are bonded to B. θ is the instantaneous angle and θ_0 is the equilibrium angle of the bond. k_θ is the force constant of the instantaneous angle θ. The third term adds the torsions modelling the energy changes during bond rotation. k_ϕ is the torsion force constant, *n* is the function multiplicity, ϕ is the torsional angle and δ_p is the phase shift. (Leach, 2001).

In the CHARMM potential function, $E_{non-bonded}$ has two components which are the electrostatic and Van der Waals interaction energies (Stote *et al.*, 1999):

$$E_{non-bonded} = \sum_{i=1}^{N_p} \sum_{j=i+1}^{N_p} \left(\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right) \tag{3.6}$$

Equation 3.6 shows two different potentials one of which is the Coulomb potential for electrostatic interactions between two molecules or between different parts of the same molecule. It is calculated as a sum of interactions between pairs of point charges. N_p's are the numbers of point charges in the two molecules. r_{ij} is the distance between atoms *i* and *j* having charges q_i and q_j. ε_0 is the effective dielectric constant for the medium. The second

potential is Lennard-Jones 12-6 potential function of Van der Waals interactions. It includes two adjustable parameters: the well depth of the potential ε_{ij} and σ_{ij}, the finite distance at which the potential between atoms *i* and *j* is zero. The twelfth power term is quite reasonable for rare gases but is too steep for other systems such as hydrocarbons, while the sixth power is the same power-law relationship found for the leading term in theoretical treatments of the dispersion energy. The 6-12 potential is widely used, especially for large system calculations (Leach, 2001). Van der Waals interactions occur due to a balance between the attractive and repulsive forces of two atoms. At short distances, the repulsive force is seen because of the strength of electron-electron interaction. The attractive force or dispersion force occurs since the electron clouds have fluctuating charge distributions. These two effects become zero at infinite atomic separation and become important when the distance increases. The range of the attractive force is longer but the repulsive interaction becomes dominant as the distance gets shorter (Stote *et al.*, 1999).

3.3. MD with Constraints and Restraints

Under normal conditions, all the bonds fluctuate more or less, while the constraints are never-fluctuating and frozen bonds. Thus, the bond between each hydrogen and the heavy atom to which it is bonded can be constrained, i.e. kept rigid in order to speed up the calculations. Although these bonds exist and fluctuate under physiological conditions, their contributions are minor and thus can be neglected. Similarly, in the case of waters, the hydrogen-oxygen and the hydrogen-hydrogen distances are generally constrained to the angle or length in the parameter file and the molecules become rigid completely (Phillips *et al*., 2005). In RATTLE algorithm, the H atom is kept near to the heavy atom it is bonded to. If the H atom moves too fast, then it cannot keep the assigned bond constraint and the algorithm fails (Andersen, 1983).

Meanwhile, restraints are not totally frozen bonds. The bonds are artificial and the bond length fluctuates around the set value during the simulation. They are defined in a specific file for every extra bond, angle, dihedral and improper terms by giving the atom indices, the applied distance or angle, and the force constant *k*. Force constants are applied according to the distance applied between two atoms. If the force constant is kept low, the

expected bond length cannot be reached. If it is kept high, the system is forced too much. It means that in the case of one of the atoms moving a little bit, the distance will increase and the system cannot tolerate it because of being too stiff. Such restraints are designated as extra terms in the form of harmonic potentials:

$$U(x) = k(x - x_{ref})^2 \tag{3.7}$$

where x_{ref} is the specified restraint distance.

However, throughout this work, the "restraints" applied to the ligand binding site will be named also as constraints.

3.4. MD Trajectory Analysis

3.4.1. Principal Component Analysis

Principal component analysis (PCA), or essential dynamics method, is one of the most popular methods for the reduction of the dimensionality of a complex system (Altis *et al.,* 2007). PCA's goal is to extract the crucial information from a data set to represent it as a set of new orthogonal variables named principal components (Abdi *et al.,* 2010). The covariance matrix, which PCA is based on, provides information on the two-point correlations of the system. The PCA shows a linear transformation, which diagonalizes the covariance matrix with dimensions $3N \times 3N$ and takes out the instantaneous linear correlations from the variables. The eigenvalues of the transformation are decreasingly ordered, and a large part of the system's fluctuations can be described in terms of only some principal components (Altis *et al.,* 2007). PCA shows the pattern of similarity of the observations as points in maps. Principal components obtained as linear combinations of the original variables are new variables computed by PCA. The first principal component has the largest possible variance. The second component has the constraint of having the largest possible inertia and being orthogonal to the first component. The other components are constituted via the same ways. Factor scores are the values of these new variables for the observations and they can be geometrically evaluated as the projections of the observations onto the principal components (Abdi *et al.,* 2010).

In order to carry out PCA on a set of data, the sample means, sample variances and the covariance between the two methods are necessary (Alakent *et al*., 2004). C^{α} coordinates are used separately for each run to form the covariance matrix. The eigenvalue decomposition is applied and by projecting the C^{α} coordinates displacements onto a principal component, the collective motion along the specific principal component is found (Alakent *et al*., 2004). In the equations below, vectors will be shown in lower case letters while matrices in upper case. The starting point of the statistical applications is the sample covariance matrix *S*:

$$S = < (X - < X >)(X - < X >)^T > \tag{3.8}$$

X is the coordinate of C^{α} atoms, and $< X >$ is the mean value of the coordinates.

For a problem with *p* variances, *S* is also written as (Jackson, 1980):

$$S = \begin{bmatrix} s_1^2 & s_{12} \cdots & s_{1p} \\ s_{12} & s_2^2 \ldots & s_{2p} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ s_{1p} & s_{2p} \cdots & s_p^2 \end{bmatrix} \tag{3.9}$$

s_i^2 is the variance of the *i*th variable and s_{ij} is the covariance between the *i*th and *j*th variables. If the covariances are not zero, it means that there is a relationship between those two variables (Jackson, 1980).

p correlated variables $x_1 \ldots, x_p$ will be transformed by a principal axis transformation into p new uncorrelated variables $z_1, \ldots, z_p$. These new variables' coordinate axes are described by the vectors u_i making up the matrix U_m of direction cosines used in the transformation (Jackson, 1980):

$$z = U_m{}'(x_m - \overline{x_m}) \tag{3.10}$$

x_m is the data matrix of the C^{α} coordinates, and $\overline{x_m}$ is the time average of the C^{α} coordinates in a single run (Alakent *et al*., 2004). The transformed variables are named as the principal components of *x* (Jackson, 1980).

3.4.2. Root Mean Square Deviation

In computational biology, in order to compare two conformations of the same protein, the structure of one protein is needed to be overlaid onto another. While computing the root mean square deviation (RMSD), the two structures should be brought together as closely as possible by rigid space motions which are a proper rotation and translation (Coutsias *et al.,* 2004). The RMSD of the models along the trajectories are calculated after the alignment of all the C^{α} atoms of each snapshot to the initial frame of the specific model and the optimal rigid body superposition in globular protein conformation studies. RMSD plot signifies the conformational changes and the amount of deviation from the initial conformation throughout the trajectory.

The initial equilibration period before the dynamic equilibrium is also defined by this RMSD plot. A very large RMSD value of two protein structures indicates that they are not similar. Zero value shows that they are conformationally identical (Maiorov and Crippen, 1994). The formulation of the RMSD is the square root of the sum of the squares of the distances between atoms x and y, giving the atomic displacement between two conformations of the protein. In the equation, *N* corresponding atom positions are given from structure x and y separately (Blessing, 1995):

$$RMSD_t = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} \left(r_{t,i} - r_{ref}\right)^2} \tag{3.11}$$

In Equation 3.11, $r_{t,i}$ is the atom position of atom *i* at any time *t*, r_{ref} is the reference atom position, *N* is the number of C^{α} atom pairs.

3.4.3. Root Mean Square Fluctuation

Protein mobility along the simulation can be expressed the most effectively by root mean square of the average fluctuation, RMSF, of each individual residue, which is calculated from the production phase of the trajectory (Stote *et al*., 1999). Prior to the calculation, all snapshots are aligned to the initial structure (frame) of the simulations of

only C^{α} atoms. Then an average of all the snapshots is calculated. The square root of the variance of the fluctuation around the average position is taken. In other words, the square of the fluctuations' average is calculated and then its square root is found:

$$RMSF_i = \sqrt{\frac{1}{T_s} \times \sum_{t=1}^{T_s} (r_{i,t} - \overline{r_i}\ )^2} \tag{3.12}$$

In Equation 3.13, T_s is the total number of snapshots during the trajectory over which the RMSF value is calculated, $\overline{r_i}$ is the average position, $r_{i,t}$ is the position of C^{α} atoms in atom *i* after superposition on the reference. The average of the square distance is taken over the selection of atoms (which is C^{α} for the case of this work) in the residue (Viswanathan *et al*., 1996).

4. RESULTS AND DISCUSSION

In this thesis, the allosteric coupling of human β_2-adrenergic receptor is investigated via molecular dynamics (MD) simulations. The effect of different constraints applied to the receptor's extracellular domain on the dynamics of intracellular domain, which includes the intracellular loop (ICL3), will be presented through different simulation trajectories. Followed by the simulation details, the first section includes the results of two unconstrained MD simulations, which are the extensions of the previously performed MD simulation by Ozcan *et al.* (Ozcan *et al*., 2013). Next, the results of the constrained MD simulation results will be presented.

4.1. Simulation Details

In Ozcan's work, the initial conformation was the X-ray crystallographic structure of human β_2AR in complex with T4 lysozyme (PDB id: 2RH1) at 2.4 Å resolution (Cherezov *et al.*, 2007). T4L between helices 5 and 6 was removed and the ICL3 region, which was taken out during crystallographic experiments, was added between residues 230 and 266 via homology modeling tool, MODELLER (Narayanan *et al*., 2009). The areas, which are disordered and not visible in the X-ray crystallography, are the N-terminus (residues 1 to 28) and the C-terminus (residues 343 to 365) tails. Since cholesterol bound form of human β_2AR (PDB id: 3D4S) is chosen to be the best scoring template by MODELLER homology modeling server, the structure in this work was completed by homology modeling according to 3D4S's template structure.

Knowledge of protonation states of protein, water molecules and ligands are necessary for the rational three-dimensional design of drugs and to understand structure-function relationships (Ahmed *et al.*, 2007). In Ozcan's work, the protonation states were designated using the programs MOE (Molecular Operating Environment) (Labute, 2007) and H++ web server (Bashford and Karplus, 1990).

The Orientations of Proteins in Membranes (OPM) database gives a collection of transmembrane, peripheral and monotopic proteins from Protein Data Bank (PDB), a

computer-based archival file for macromolecular structures (Bernstein *et al*., 1977). The surface of the cell membrane was perpendicular to the z-axis of the coordinate file from OPM. Thus, the oblique angle between the z-axis and the receptor's main principal component along the cell membrane was determined to be 8^{o}. The receptor in Ozcan's work was aligned to the structure given by OPM in order to have a tilted protein (Ozcan, 2011).

15 water molecules were located at the inner regions of the structure since it was also experimentally found. These water molecules were not neglected since they make critical hydrogen bonds with the most conserved residues. In order to neutralize the protein and equalize the net charge of the system to zero, Na^{+} and Cl^{-} ions were added with a concentration of 0.154 mol/L using VMD's Autoionize module v1.2. When the ions were added, the system had 17 Na^{+} and 24 Cl^{-}. The total electric charge of the system was tried to be neutralized because of the fact that Particle-Mesh Ewald (PME) summation method, which is used to compute long range electrostatic interactions, can only work under these conditions. This method was influenced by particle-particle particle-mesh method based on dividing the total electrostatic energy into local interactions and long-range interactions (Essmann *et al*., 1995).

Nanoscale Molecular Dynamics (NAMD) v2.6 software tool, which is useful for high-performance simulations of large biomolecular systems was used for all our MD runs (Phillips *et al.,* 2005). The force field used was CHARMM27 (Foloppe *et al*., 2000) for lipids and CHARMM22 (Mackerell *et al*., 2001) for proteins. TIP3P model was used for water in the system (Neria *et al.,* 1996). According to the Berendsen weak-coupling approach, the pressure was kept at 1 bar (1 atm) due to the fact that it is the same pressure inside the cell (Berendsen *et al.,* 1984). Using a heat thermostat, the temperature of the simulation was kept constant at 310 K, which is the physiological temperature. Also, the type of the ensemble chosen is crucial in an MD simulation. In this work, the isothermal-isobaric (NPT) ensemble was employed (Stote *et al*., 1999). Langevin dynamics were used in order to keep the temperature constant with a Langevin damping coefficient (gamma) of 5/ps for all non-hydrogen atoms. The pressure was kept constant at 1 atm using a Nose-Hoover Langevin piston with 100 fs period and 50 ps damping timescale (Feller *et al.*, 1995).

The system in Ozcan's work was equilibrated through a simulation protocol that consists of melting of lipid tails, minimization and equilibration with protein constrained, equilibration with protein released and lastly the production runs using an integration time step of 2 fs (Ozcan, 2011). In reality, the vibrational frequency of the hydrogen atoms is higher than 2 fs. The usage of the RATTLE algorithm facilitating the Velocity Verlet integration scheme which constraints the hydrogen molecules' mobility enabled the usage of 2 fs time steps. Every 200 ps the Cartesian coordinates of atoms and energies were recorded (Andersen, 1983).

As the main component of the biological cell membrane, a double-layered 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC) phospholipid cell membrane was generated with a specific thickness at the direction of z-axis using the membrane Plug-in v1.9.1 of VMD (Visual Molecular Dynamics) visualization program in Ozcan's work (Humphrey *et al*., 1996). It was measured to be 0.65 nm^2 per a lipid molecule by Lantzsch and his coworkers (Lantzsch *et al.*, 1994). The dimensions of the box in x and y directions were adjusted according to the protein's dimensions in the same directions. The box dimension in the z direction was chosen as 100 Å according to the amount of the water molecules. This value was arranged so that water thickness would be 15 Å at the upper and lower regions of the protein and the lipid system in the z direction. After all these stages, experimental value of 0.65 nm^2 was approached. Table 4.1 shows the dimensions of the protein, cell membrane and the system.

Table 4.1. Dimensions of the system, cell membrane and the protein.

Box dimension (Å) (xyz)	**Cell membrane (Å) (xy)**	**Protein Dimension (Å) (xyz)**	**Protein (Å) (x_{min}/x_{max})**	**Protein (Å) (y_{min}/y_{max})**	**Protein (Å) (z_{min}/z_{max})**
86×86×100	86×86	56×57×71	25.4/30.0	36.0/21.2	33.3/38.0

Two independent 500 ns runs named as *continued_1* and *continued_2* were performed using the final snapshot of the original run with a closed ICL3 loop, which also represents a very inactive state of the receptor (Ozcan *et al.*, 2013). No constraints were

applied to the binding site. The goal here was to observe how long ICL3 would remain in this closed state, which would in turn indicate how stable this inactive state is.

In order to study the effect of constraints on the dynamics of β_2AR, especially of ICL3, harmonic distance constraints are applied to seven pairs of residues located at the ligand-binding cavity that are known to be critical in binding small signaling molecules. The distance values measured by experimental approaches in the receptor's active state (Liapakis *et al.*, 2000; Gouldson *et al*., 1995) are adopted in this study. Two different sets of constraints are applied, first to the initial state of the receptor, in which ICL3 is in an open position (*c1_initial_open, c2_initial_closed*), and later to an intermediate state, in which ICL3 is partially open (*c3_interm_closed, c4_interm_open, c5_200ns_interm_open,* and *c6_200ns_interm_closed*). *c3_interm_closed*, *c4_interm_open, c5_200ns_interm_open* and *c6_200ns_interm_closed* started from the 2351st frame (470 ns) of the original 1000 ns MD run. The force constants used for the constraints are chosen as 50, 60 or 65 according to the system's response to constraints.

The total number of atoms in the models was 5.055 for the proteins, 20.77 for the lipids, 42.135 for the water, and 41 for the ions, which makes a total 68.001 atoms in the system. General information about the MD runs is given in Table 4.2. Here, the MD run called *original* was performed by Ozcan and his coworkers (Ozcan *et al*., 2013).

Table 4.3 demonstrates the distances observed in crystallographic structures and the constraints applied to the MD runs of 500 ns long each. The inactive and active state distances were noted for the inactive 2RH1 and the active 3SN6 crystal structures. γ oxygen, and β and γ Carbon atoms are taken into consideration (Cherezov *et al.*, 2007; Rasmussen *et al*., 2011).

In the constraint model *c1_initial_open*, Ser203-Asp113 distance was set to 16 Å for the first 300 ns. For the next 200 ns, it was increased to 17 Å. The remaining distances were set to the observed values in 2RH1.pdb.

Table 4.2. General information about the models and the simulation.

	Name	**Total duration (ns)**	**Constrained / Ligand Binding Site**	**Initial structure**
1	*original*	1000	No	crystal structure (2RH1)
2	*continued_1*	500	No	last frame of original
3	*continued_2*	500	No	last frame of original
4	*c1_initial_open*	500	Yes / open	0th frame
5	*c2_initial_closed*	500	Yes / closed	0th frame
6	*c3_interm_open*	500	Yes / open	2351st frame (470th ns)
7	*c4_interm_closed*	500	Yes / closed	2351st frame (470th ns)
8	*c5_200ns_interm_open*	200	Yes / open	2351st frame (470th ns)
9	*c6_200ns_interm_closed*	200	Yes / closed	2351st frame (470th ns)

For *c2_initial_closed*, initially, the values of 8 Å, 10 Å and 8 Å were assigned for Ser203-Asp113, Ser204-Asp113 and Ser207-Asp113, respectively. However, the simulation became unstable with an error message of "*constraint failure in RATTLE algorithm*". In order to overcome this unstable state, the simulation was restarted with slightly larger distances, which were 10 Å, 12 Å and 10 Å. The simulation was carried out for 70.6 ns. Then, the constraints were changed back to 8 Å, 10 Å and 8 Å and the stability was established. The simulation was carried on for a total of 500 ns.

For *c3_interm_open*, the distances were set as shown in Table 4.3 and the system was stable throughout the simulation.

For *c4_interm_closed*, the constraints were applied incrementally, similar to *c2_initial_closed.* First, the three distances (Ser203-Asp113, Ser204-Asp113 and Ser207-Asp113) were set to 10 Å, 10.5 Å, and 10 Å, respectively. The simulation was carried on for 26.2 ns. Then, for the next 473.8 ns, these three distances were decreased to 8 Å, 9 Å, 8 Å, respectively.

The constraints of *c5_200ns_interm_open* were the same as *c3_interm_open,* except that an additional 17 Å distance constraint on the backbone of Ser203-Asp113 pair was incorporated.

For *c6_200ns_interm_closed*, the final state of the 26.2 ns long MD run in *c4_interm_closed* was taken. Additionally, Ser203-Asp113 backbone distance constraint of 7.5 Å was incorporated, as well. Thus the first four distances were constrained to 10 Å,10.5 Å, 10 Å and 7.5 Å for 26.2 ns, respectively. Later, the simulation was carried on for 12.4 ns with a new set of constraints 7 Å, 8 Å, 7 Å and 7.5 Å, additional to the first 26.2 ns. Then, for the last 161.4 ns, the first four distances were further decreased to 6 Å, 7 Å, 6 Å and 6 Å, respectively, as reported in Table 4.3.

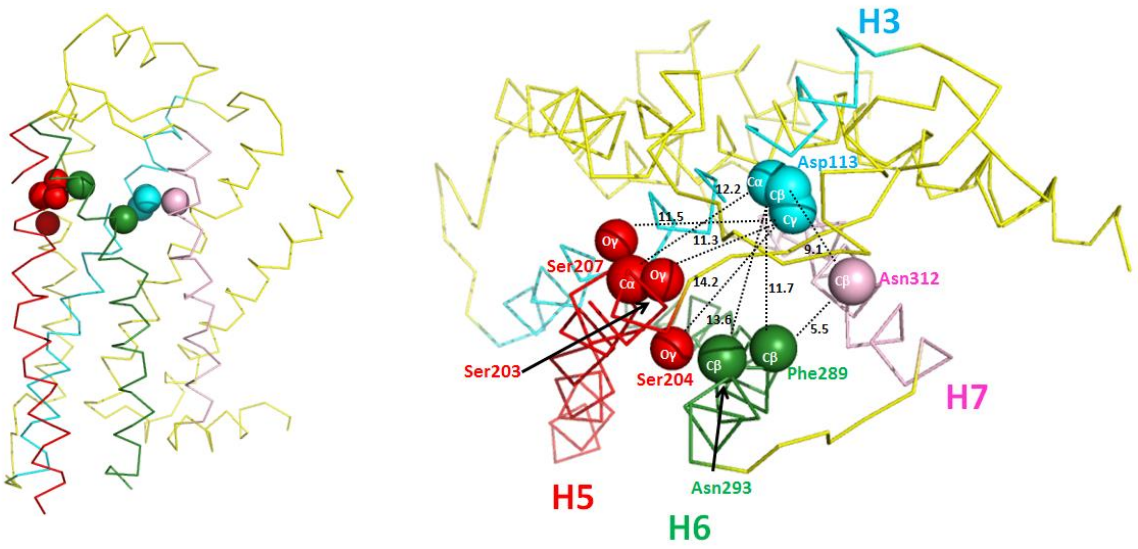
Table 4.3. Distances in crystallographic structures and bond constraints.

	Distances in crystallographic structures (Å)		**Constraint distances (Å)**					
Residue pair	**Inactive state (PDB id: 2RH1)**	**Active state (PDB id: 3SN6)**	**c1 (O)**	**c2**	**c3 (O)**	**c4**	**c5 (O)**	**c6**
Ser203O^{γ}-Asp113C^{γ}	11.25	10.34	17	8	17	8	17	6
Ser204O^{γ}-Asp113C^{γ}	14.18	12.37	14	10	14	9	14	7
Ser207O^{γ}-Asp113C^{γ}	11.51	10.37	11.7	8	11.7	8	11.7	6
Ser207C^{α}-Asp113C^{α}	12.17	11.97	-	-	-	-	17	6
Asn293C^{β}-Asp113C^{β}	13.59	13.59	13.5 *(14)*	13.5 *(15)*	13.5 *(14)*	8 *(9.5)*	13.5 *(15)*	10 *(11)*
Phe289C^{β}-Asp113C^{β}	11.73	12.31	11.7 *(13)*	12 *(14)*	11.7 *(13)*	8.5 *(10)*	11.7 *(13)*	8.5 *(9)*
Asn312 C^{β}-Asp113C^{β}	9.09	8.62	9 *(10)*	8 *(9)*	9 *(10)*	8 *(8.5)*	9 *(10)*	8 *(9)*
Phe289C^{β}-Asn312C^{β}	5.5	5.46	5.5	5.5	5.5	8	5.5	8

These six different constraint runs are referred to as c1, c2, c3, c4, c5 and c6, in short in Table 4.3, among which runs with constraints forcing the binding site to open position is indicated by (O). In some of the runs, slightly larger distances than the imposed constraints

were achieved, which are indicated in italic under the constraint distances. The constraint distances reported in the table represent the final set value in cases where the constraints have been changed during the run.

The distance measurements for the inactive crystallographic state of 2RH1 are shown in Figure 4.1 both from the side and top views, looking down from the extracellular side.



(a) (b)

Figure 4.1. Distance measurements for the inactive crystallographic state of 2RH1 from (a) side and (b) top views.

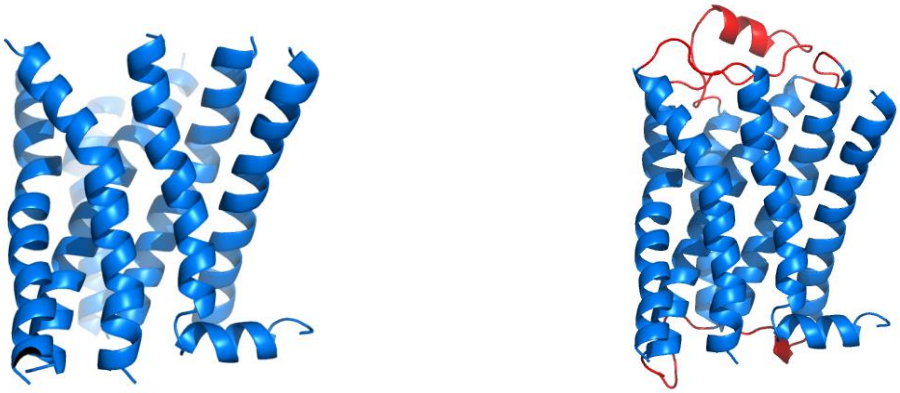
4.2. Unconstrained MD Runs

This section introduces the results of two 500 ns long unconstrained MD simulations, which are carried out as an extension of the 1 μs long simulation (Ozcan *et al.*, 2013), as described in Simulation Details, Section 4.1. Previously, the ICL3 started to pack under β_2AR at around 600 ns and kept a stationary state until the end of 1 μs. The aim here is to determine how long this stationary, restricted state would last. The extended simulations called *continued_1* and *continued_2* start from the final snapshot of the so-called *original* 1 μs long simulation with different velocities, in order to improve the sampling. The *continued_1* starts using the final velocities of the *original* run and carries out for another 500 ns, during which ICL3 does not change its packed form. The *continued_2* is performed with a different velocity distribution, causing ICL3 to slightly move away from the protein

molecule at around 220 ns (1100th frame) before returning back to its packed position shortly. In comparison, ICL3 stays almost fixed at a closed position in *continued_1*.

4.2.1. RMSD Profiles

In order to calculate the RMSD values along a trajectory, a specific region of the protein is first selected and aligned to the same region in the initial frame. Next, the RMSD value of the whole protein or a portion of it is calculated. This procedure is repeated for all snapshots in the trajectory. Accordingly, in the "*transmembrane_fit_transmembrane RMSD*" profile in Figure 4.2a, the alignment was based on the transmembrane region of the protein, which is composed of the parts of helices embedded inside the lipid membrane. The RMSD value was then calculated for the transmembrane region only. Likewise, in the "*all_fit_core RMSD*" in Figure 4.2b, the core region of the protein, which is the whole protein except the ICL3 region, is aligned to its initial frame first, and later the RMSD value is calculated for the whole protein.



(a) (b)

Figure 4.2. Specific regions used in RMSD calculations; (a) transmembrane region including no loops, (b) core region including all the loops (colored in red) except ICL3.

Figure 4.3 shows *all_fit_core, core_fit_core* and *transmembrane_fit_transmembrane RMSD* profiles along with a few snapshots generated by VMD program (Humphrey *et al.*, 1996) from intracellular side for *original, continued_1* and *continued_2*. In RMSD profiles, red indicates *all_fit_core*, green is *core_fit_core*, purple is *transmembrane_fit_transmembrane RMSD*. The snapshots colored red, white and blue indicate the initial, middle and final frames aligned to each other, respectively. For

original, the *core_fit_core* and *transmembrane_fit_transmembrane RMSD* profiles are nearly stable around 4 Å and 2 Å, respectively, while in *all_fit_core RMSD* profile, there is a major increase from 2 Å to 10 Å, which is caused by the RMSD of ICL3 region. The snapshots support this increase by the change in the position of ICL3. Figure 4.3b shows the RMSD profiles of *continued_1*, which stay around 2-3 Å throughout the simulation. This is an expected result since ICL3 is rolled up under the receptor and stays there for the whole trajectory keeping its closed position as the initial (red) and the last (blue) snapshots prove being in the same region in Figure 4.3b. In Figure 4.3c of *continued_2*, ICL3 causes the a*ll_fit_core RMSD* to have a sharp increase around 230 ns from 2 Å to nearly 6 Å and to stabilize around 4 Å. The reason for this increase may be related to the sudden opening and closing of ICL3 as seen in the corresponding snapshot colored in white. When ICL3 closes back, it adopts a slightly different conformation, which is indicated by an RMSD of 4 Å instead of 2 Å according to the starting conformation in the RMSD profile.

The lower parts of H5 and H6 have a crucial role in the simulations due to their location and contact with ICL3 throughout the runs. Therefore, RMSD profiles of the lower parts of H5 and H6 are determined in order to observe their dynamics in relation with ICL3 mobility. Two different alignments are done according to the transmembrane region using active (PDB id: 3SN6) and inactive (PDB id: 2RH1) structures of the receptor.

Figure 4.4 shows the RMSD results for the lower side of H6 and the corresponding snapshots for *original, continued_1* and *continued_2*. Purple indicates the RMSD with respect to active crystal structure and green indicates the RMSD with respect to inactive crystal structure for the lower part of H6. For *original* in Figure 4.4a, the RMSD with respect to inactive structure is around 1 Å for nearly 500 ns and then starts to increase to 4 Å towards the end. The RMSD with respect to active structure starts around 7 Å and moves to 10 Å starting from 500 ns. The upper and lower limits of this value are shown on the figure for *original* as a reference for the other runs. Thus, H6 moves to a final position farther away from both active and inactive crystal structures, which is named as a ‘very inactive state’ in previous work (Ozcan *et al.*, 2013). The *continued_1* and the *continued_2* in Figure 4.4b and Figure 4.4c are also quite stable at this very inactive state, both at around RMSD of 4 Å with respect to inactive structure and RMSD of 10 Å with respect to

active structure values. As a result, the most visible change is seen in the RMSD of lower part of H6 starting from around 610 ns of *original* due to ICL3 closure towards the end of the trajectory. The slight opening of ICL3 in *continued_2* does not have a visible effect on RMSD of lower part of H6 which seems to be at the same position throughout the trajectory.

Figure 4.5 shows the RMSD values of the lower part of H5 with respect to inactive and active structures along with their corresponding snapshots. These two RMSD values are almost the same for the three runs. For *original* in Figure 4.5a, generally the RMSD is around 2.5 Å. Initially, it is around 3 Å and decreases to 1 Å around 38 ns. It stays there until around 233 ns and starts to increase to around 2 Å. Then a decrease to 1 Å is seen again around 500 ns, and it increases to 2.5 Å within 100 ns and stabilizes there for the rest of the trajectory. Interestingly, the beginning of the plateau coincides with the packing of ICL3. In *continued_1* and *continued_2*, both RMSD results are stable around 2.5 and 2 Å, respectively.

4.2.2. RMSF Profiles

The peaks in Figure 4.6 correspond to the regions fluctuating the most in the protein during the simulation, which are the residues with the highest root mean square fluctuations. As expected, ICL3 between residues 230-266 fluctuates the most in all of the trajectories due to its conformational flexibility although ICL3 and ECL2 have nearly the same RMSF values due to the closely packed ICL3 in the case of *continued_1*. In general, secondary structure elements such as alpha helices and beta strands are more rigid than loops, so their fluctuations are less compared to loop regions. The highly mobile intracellular and extracellular loops such as ICL1, ICL2, ECL1, ECL2 and ECL3 also show high RMSF values compared to the helical regions. ICL2 and ECL2 are the other mobile loops with the highest RMSF values after ICL3. The most mobile ICL3 belongs to *orig_1st half* (first 500 ns of the *original*), whose ICL3 moves freely in a fully open conformation before adopting its packed state during *orig_2nd half* (last 500 ns of the *original*). Furthermore, the residues, which are more mobile than the other helices' residues are located on H4 and H5. The mobility of ICL3 may affect the mobility of the other loops and helices within the protein, bringing conformational flexibility to the whole

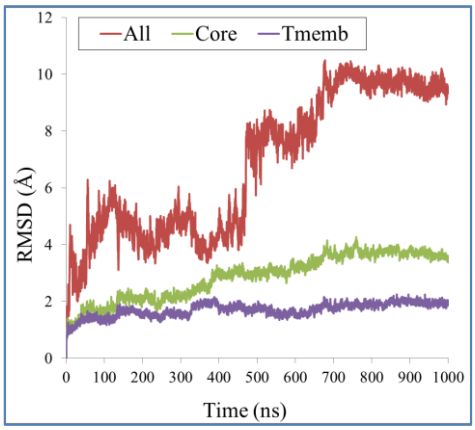
structure. The mobility of H5 may be due to the fact that its upper part was displaced away from the binding site towards the end of the simulation, which is around the same time that the lower part of the receptor became much more restricted by ICL3 packing.

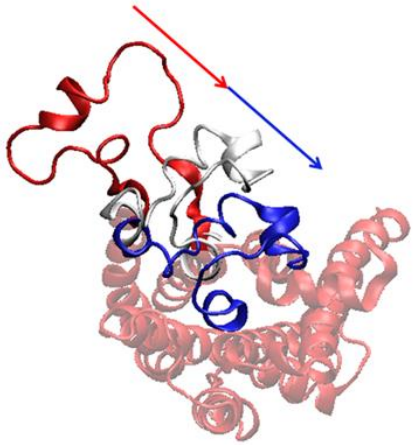
4.2.3. Distance Profiles For Ser207-Asp113 Residue Pair

The distance between Ser207 (O^{γ}) in H5 and Asp113 (C^{γ}) in H3, shown in Figure 4.7, is also crucial in β_2AR due to the fact that these critical residues are in the binding region and interact with both agonists and antagonists. H5 including Ser207 is shown in red and H3 with Asp113 is in green. In *original*, during the time when ICL3 and the lower part of H6 move inwards, blocking the G-protein-binding cavity, serine residues (Ser203, Ser204 and Ser207) as well as the upper part of H5 move away from the ligand binding site and cause the widening of the binding cavity. The increase of the distance between Ser207 and Asp113 can be observed in Figure 4.8, where *original* has an increasing profile from 11 Å to 15 Å as the simulation goes on, while *continued_1* and *continued_2* have more stable profiles around 15 Å.

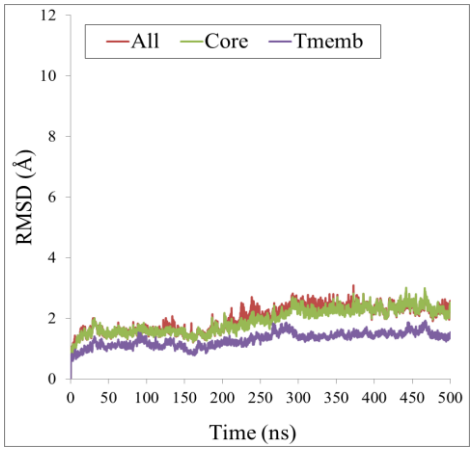
In order to emphasize the allosteric coupling of the lower and upper parts of the receptor as mentioned in the previous paragraph, the change in the distance value between Ser207-Asp113 is plotted against RMSD of the lower part of H6 with respect to active structure in Figure 4.9. RMSD results of the lower part of H6 are shown separately for the first and second 500 ns of the long run as *orig_1st half* (green dots) and *orig_2nd half* (orange dots). The *original* simulation has a wider range starting from around 6 Å and increasing to around 10 Å with Ser207-Asp113 distance increasing from 8 Å to 19 Å, while *continued_1* (blue) and *continued_2* (red) have narrower ranges. In Figure 4.10, data for continuation runs are shown separately for better comparison. For *continued_1*, the distance between Ser207 and Asp113 varies from 11.5 Å to 19 Å, while RMSD of lower part of H6 differs from 8 Å to 10 Å. For *continued_2*, which spans the smallest area, the distance between Ser207 and Asp113 varies from 12 Å to 18 Å, while RMSD of lower part of H6 differs from 9 Å to 10 Å. Since *continued_1* and *continued_2* are the continuations of *original*, they coincide with the last cluster of *orig_2nd half* (between 9-10 Å for

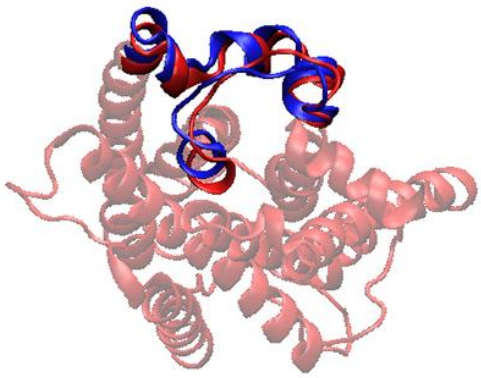
RMSD of lower part of H6 and 13-18 Å for distance between Ser207-Asp113). This means that the upper and lower parts of the protein stay stable around the last closed position.



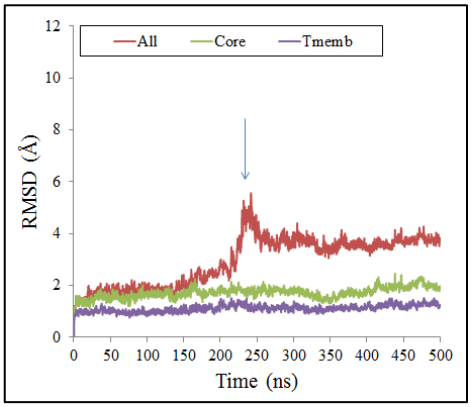


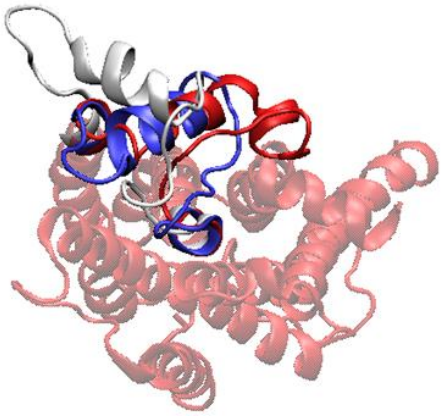
(a)





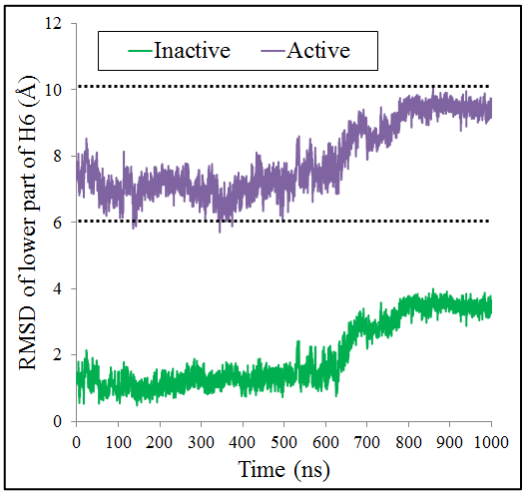
(b)

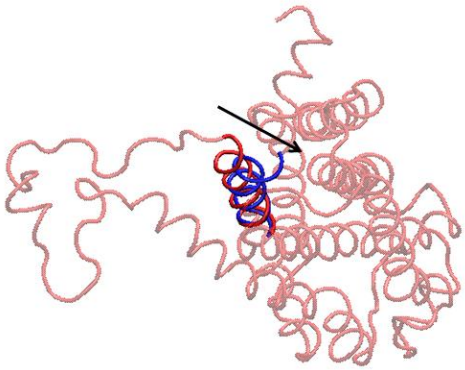




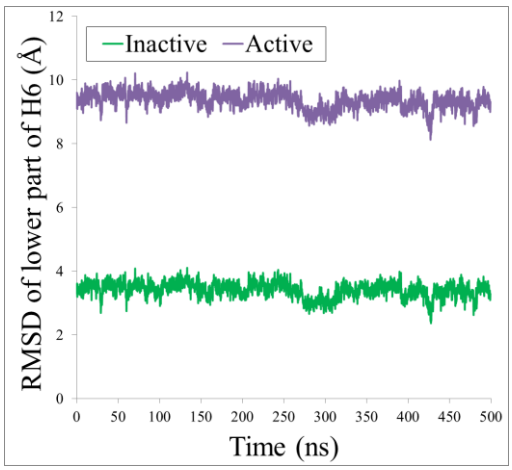
(c)

Figure 4.3. RMSD and snapshots of (a) *original*, (b) *continued_1* and (c) *continued_2* from the initial energy minimized structure.

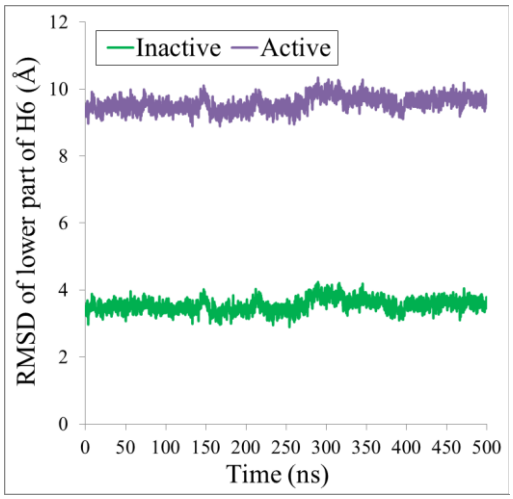


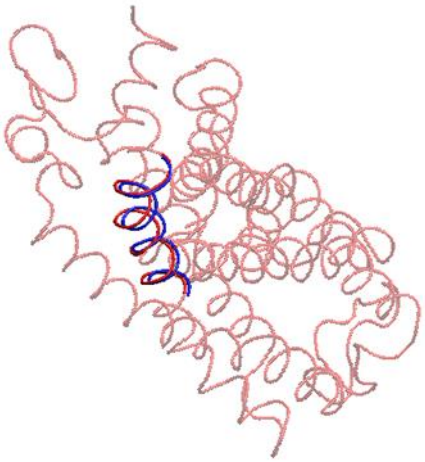


(a)



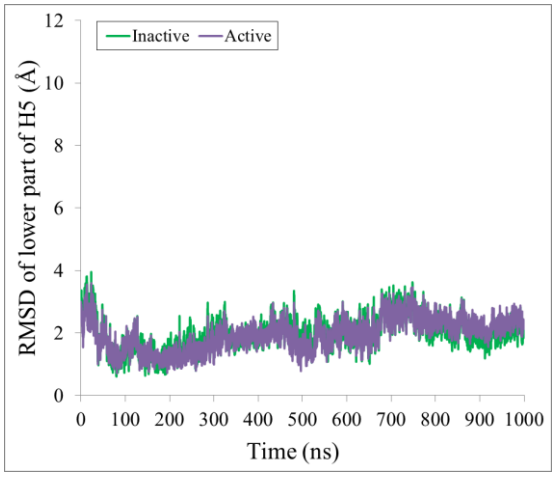
(b)

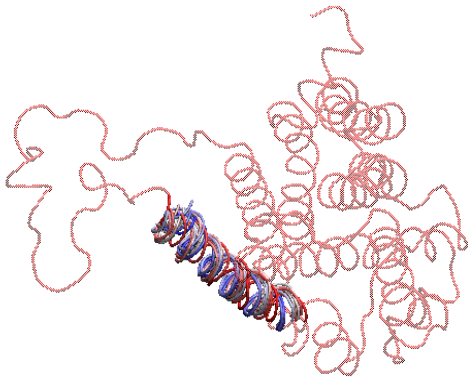




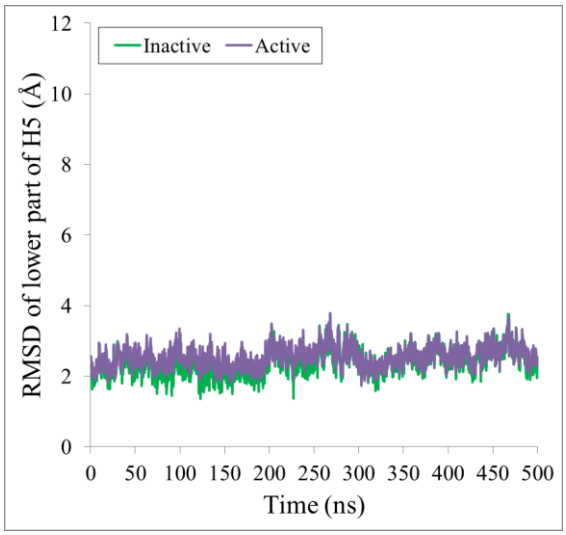
(c)

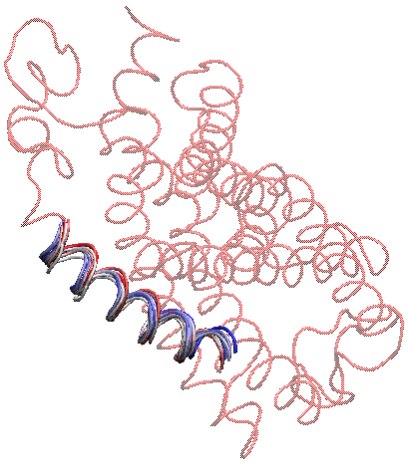
Figure 4.4. RMSD of lower part of H6 and snapshots for (a) *original*, (b) *continued_1* and (c) *continued_2* from active and inactive crystal structures.



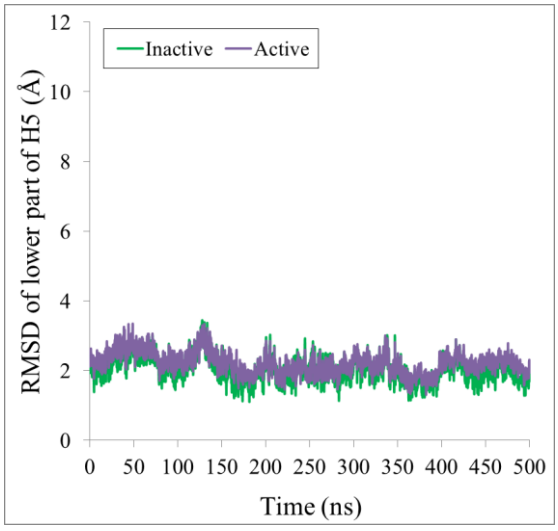


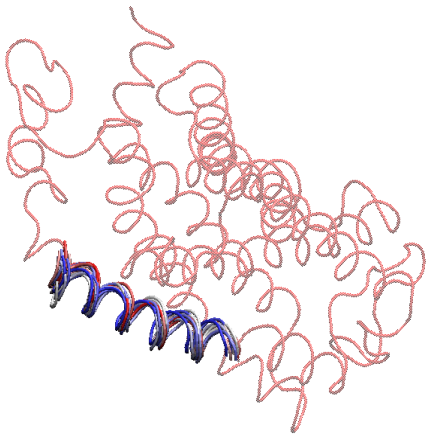
(a)





(b)





(c)

Figure 4.5. RMSD of lower part of H5 and snapshots for (a) *original*, (b) *continued_1* and (c) *continued_2* from active and inactive crystal structures.

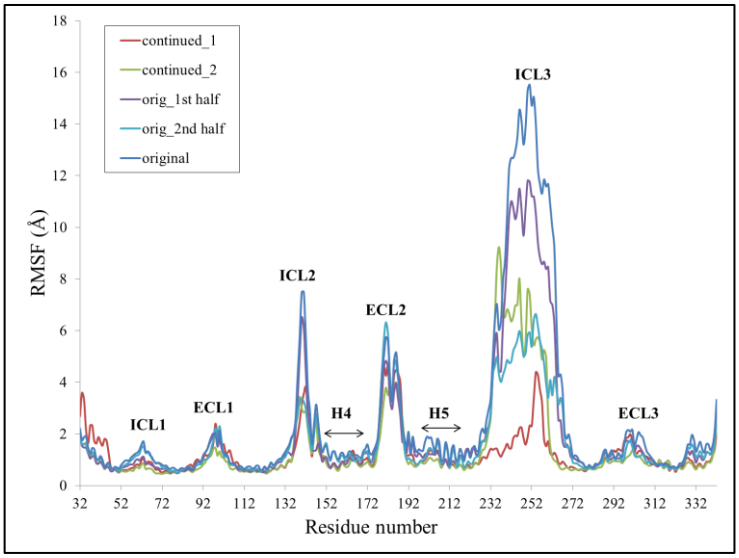
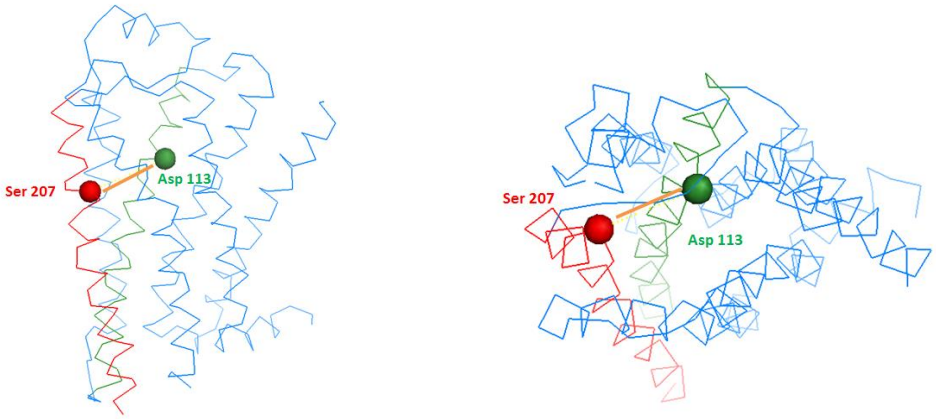


Figure 4.6. RMSF values of the 500 ns long unconstrained models.



(a) (b)

Figure 4.7. Ser207-Asp113 distance in β_2AR from (a) side view and (b) upper view.

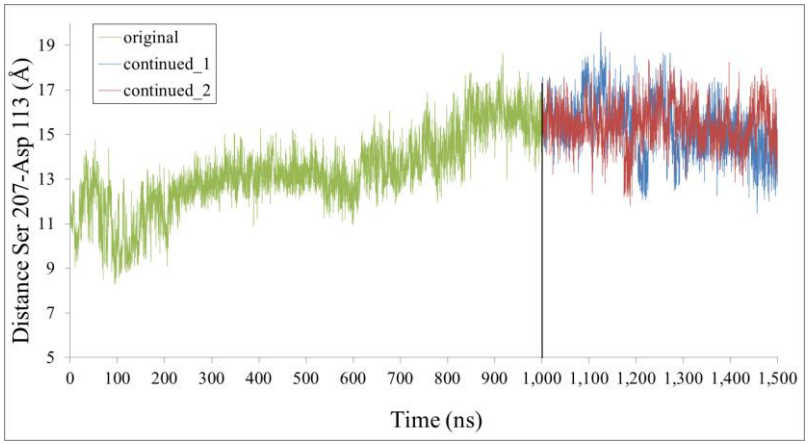


Figure 4.8. Ser207-Asp113 distance profiles for unconstrained models.

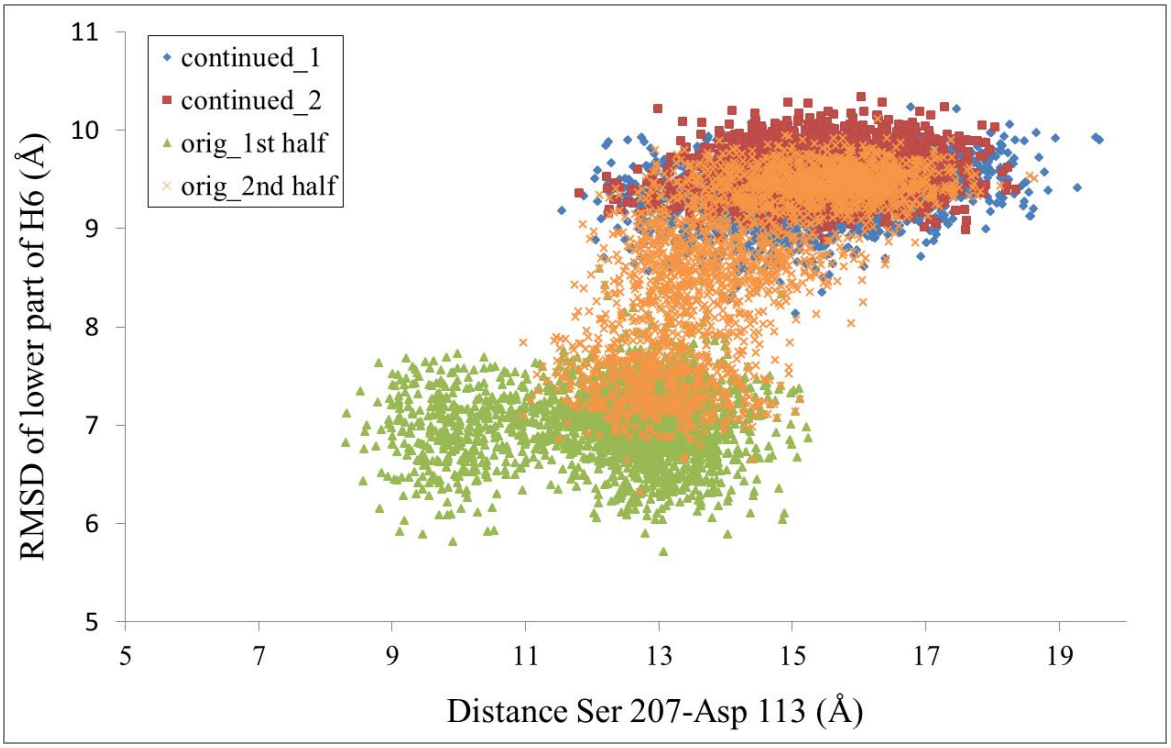
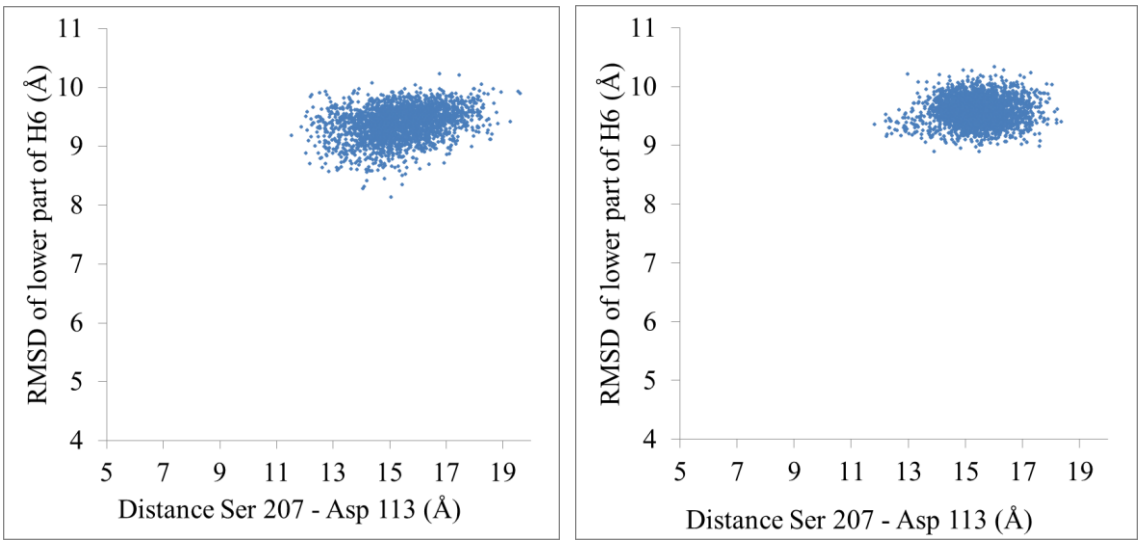


Figure 4.9. Ser207-Asp113 distance vs RMSD of lower part of H6 results for unconstrained models.



(a) (b)

Figure 4.10. Ser207-Asp113 distance vs RMSD of lower part of H6 (A) for (a) *continued_1* and (b) *continued_2.*

4.2.4. Monitoring the Position of ICL3 Through Its Center of Mass

With the inward movement of H6, ICL3 also makes a major conformational shift towards the core of the receptor as observed in *original* MD. In order to monitor this displacement, the center of mass of ICL3 is calculated and plotted for each snapshot of the extended trajectories *continued_1* and *continued_2* and also *original* in Figure 4.11. The x- and y- coordinates determine the lateral position (movement) of the ICL3 loop, since the z-axis is almost parallel to the principal axis of GPCR's transmembrane region. In the same figures, the position of a portion of G-protein interacting with β_2AR is also depicted, in order to show the closure of the G-protein binding site as a result of this displacement. The gray lines give a tentative position for G-protein by connecting its residues' x and y coordinates. G-protein is obtained from the crystal structure with PDB id 3SN6 and aligned to the first frame of *original*. The lower right side of the graph indicates the open position of ICL3, while the upper left side shows the closed position. Thus, a very clear closure is seen for ICL3 of *original* as explained by the frame numbers in Figure 4.11a. The movement of ICL3 is divided into three main parts for a better view. Between 673 ns and 1000 ns, the ICL3 region stabilizes at a closed position under the receptor. In Figure 4.11b, *continued_1*'s ICL3 is in a very closed pack in the left upper region, while that of *continued_2* in Figure 4.11c starts from the same position of *continued_1* and travels towards a more open region and stabilizes between 254-500 ns. Finally, Figure 4.12 shows the center of mass profiles of all trajectories in one place. The last bulk of *original* coincides with *continued_1* and the last bulk of *continued_2* as mentioned before.

4.2.5. Conformational Dynamics and PCA

PCA is facilitated to reduce the parameters of the data if the system includes too many parameters. The essential modes give information about the proteins' cooperative conformational motions, which are also related to biological functionalities of proteins and enzymes (Berendsen *et al.,* 1984). These anharmonic or essential modes are effective in the overall dynamics of the structure among the 3N-6 normal modes, too. These modes correspond to the collective motions that have high variance and generally are functionally relevant, disregarding the random fluctuations.

In Figure 4.13, the cumulative explanation values of principal modes are compared for the unconstrained models. According to the results, the model *original* has the highest cumulative percentage result around 70% for the first PCA mode. It means that the first slow PCA mode explains 70% of the movement giving the protein's overall dynamics with a good result. Also, the first two PCA modes of *original* explain around 80% of the movement cumulatively. *continued_2* and *orig_2nd half* have nearly the same results for 20 PCA modes although a slight opening and later closing of ICL3 is observed in *continued_2*. Since no movement is seen in *continued_1*, the least cumulative PCA percentage belongs to it with a result around 32%.

Figure 4.14 shows the eigenvectors of the first principal component on the receptor for *original*, where H3 is shown in orange, H5 in green, H6 in cyan, and H7 in pink. It does not include the loops for a better view of the movements of the helices. Upper and lower views of the models are shown separately in order to distinguish the displacements from both intracellular and extracellular sides, respectively. Additionally, helices 5 and 6 are shown together with ICL3 in order to see its movement relative to H5 and H6, as well. In all vector presentations, the magnitude of the residue deformations are adjusted according to the variance (eigenvalue) of the first mode for the specific run discussed. Later a common scaling of vectors is also applied to each run, so that the helical deformations can be seen more clearly as they are much smaller compared to ICL3 movement. Therefore, the magnitudes of the vectors from different runs can be compared with each other. The mode vectors show that the upper region of H5 moves outside while the lower region which is pulled by ICL3 moves inside. Similarly, the upper part of H6 moves outside, while the lower side of H6 moves inside also pulled by ICL3. The long mode vectors on ICL3 show why H5 and H6 move in the same direction with it.

Ten snaphsots of *original* shown from top and bottom views in Figure 4.15 support the mode vectors. The helices except H3, H5 and H6 are transparent as well as ICL3 for a better view. The arrows on H5 and H6 shows the directions of the movements.

Figure 4.16 shows the first principal mode vectors of *continued_1*, which has smaller and less mode vectors compared to *original*. The movement of ICL3 is not correlated with

H5 or H6 in general for *continued_1*. Moreover, there is not a coupling in *continued_1* like in *original* in which the lower region closed while the upper region of the receptor opened.

Figure 4.17 shows the first principal mode vectors of *continued_2*, which has a relatively mobile ICL3 region compared to *continued_1*. This can be due to the new velocity given to *continued*_2 which samples different conformations, while *continued_1* had the same velocity of *original*, thus having the same initial conformation, a closely packed up ICL3.

In general, these PCA mode vector figures support the snapshot profiles of the simulations shown in Figure 4.3. Some of the mode vectors are not shown on mode vector graphics due to their low values compared to the other regions' mode vectors. Nevertheless, these figures enlighten the perspective of investigating the movement of each helix separately and also of ICL3.

Snapshots of *continued_1* and *continued_2* in Figure 4.18 and Figure 4.19 show the stabilities of the helices, supporting the first principal mode vectors explained before.

4.3. Constrained Runs with an Open ICL3

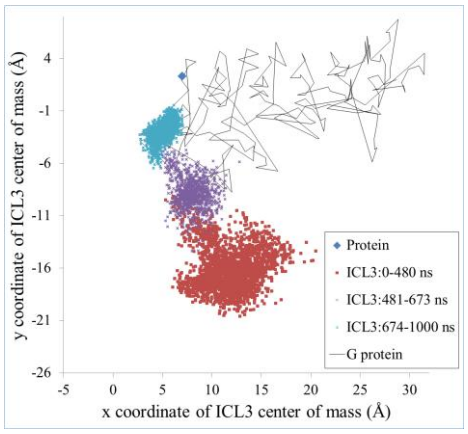
Two independent runs are performed by applying two different set of constraints to *original'*s first snapshot with open ICL3, which will be discussed in this part. The first run is *c1_initial_open* with its set of constraints given in Table 4.3. In this set, six distances are constrained to their values observed in the inactive crystal structure 2RH1, and a comparatively larger distance than the inactive state is used for Ser203O^{γ}-Asp113C^{γ}.

Ser203 on H5 is one of the key residues of the receptor in terms of ligand binding. It is located on the next turn above Ser207 towards the extracellular region. In previous work (Ozcan *et al*., 2013) performed on *original* 1 μs run, the distance of Ser203-Asp113 has increased to 16 Å from its initial value of around 11 Å in inactive crystal structure due to restraining conditions during system preparation including insertion in the membrane and several stages of equilibration. In *original'*s starting conformation both the binding site based on Ser203-Asp113 distance and the ICL3 were in an open position. Later, it was

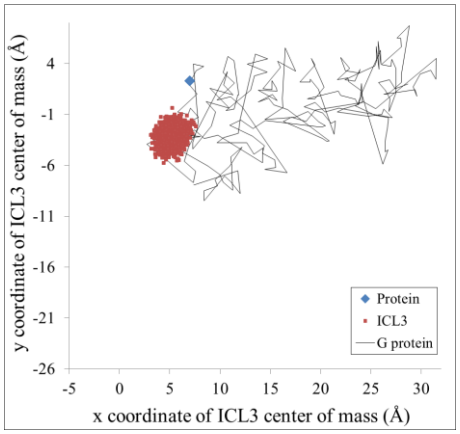
observed in four independent runs (1 μs and three other 100 ns runs) that this distance of 16 Å decreased back to around 11 Å in a short time period of about 10-20 ns, during which ICL3 stayed in its initial open position. This observation pointed to an allosteric coupling between the ligand binding site and the ICL3 loop. In other words, the Ser203-Asp113 distance in binding site cannot stay open as long as ICL3 is open, which is consistent with the observation that the Ser207-Asp113 distance increases as ICL3 closes or vice versa. These findings further lead to the question of whether ICL3 could be guided to closure by constraining the Ser203-Asp113 distance at its initially open position around 16 Å. So in *c1_initial_open*, the distance Ser203-Asp113 is kept at 16 Å for the first 300 ns (and slightly increased to 17 Å after 300 ns), which is much larger than 11.25 Å observed in the inactive state 2RH1. Ser204-Asp113 and Ser207-Asp113 distances are kept near to the inactive state values as well as the other distance constraints.

In Figure 4.20a, all the constraint distances are shown together for *c1_initial_open*. The distances are set to the inactive state values except Ser203-Asp113 distance which is set to a much higher value of 17 Å. Since the observed distances are closer to the distances observed in the inactive state of the receptor, further adjustment was not necessary. As a result, the figure shows that all the Ser couples almost show the set values. The other distance constraints show slightly higher profiles. In addition, Figure 4.20b shows the unconstrained distances between backbone C^{α} atoms, which fluctuate more than the constrained distances between side chains.

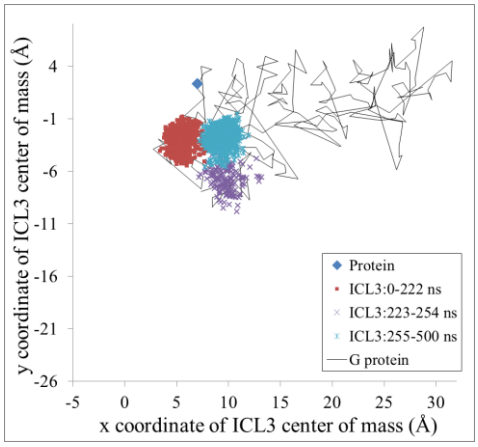
As expected, the observed change in this system of 500 ns simulation (*c1_initial_open*) is the closure of ICL3, shown best in Figure 4.21a of ICL3 center of mass. According to this figure, ICL3 starts to close at around 134 ns, pointing out to a much faster closure compared to *original*, which showed the same type of closure around 600 ns. Figures 4.21b and 4.21c show a total of 20 snapshots for *c1_initial_open* shown from side and lower views. The closure of ICL3 is recognized very clearly with the changing colors of the snapshots from red (initial snapshot) to blue (final snapshot).



(a)



(b)



(c)

Figure 4.11. Variations in ICL3's center of mass for (a) *original*, (b) *continued_1*, and (c) *continued_2* compared to coordinates of G-protein residues and receptor's center of mass.

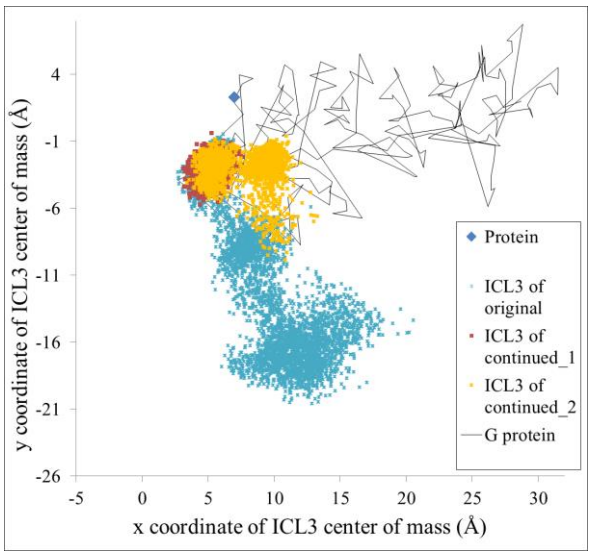


Figure 4.12. Variations in ICL3's center of mass for unconstrained runs.

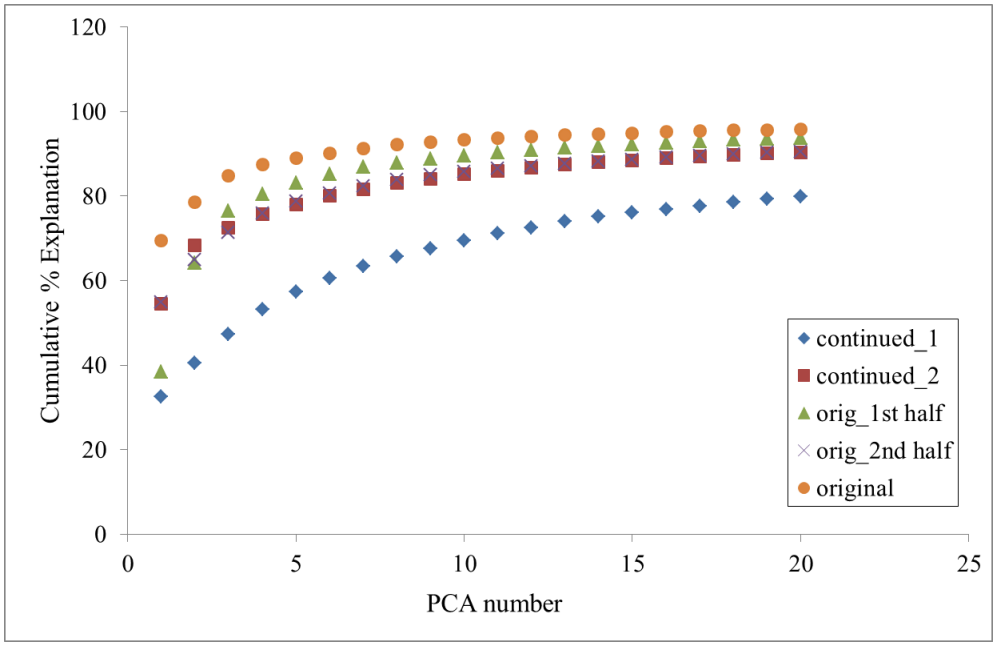


Figure 4.13. Cumulative % explanation of principal components for unconstrained models.

In the second run called *c2_initial_closed*, another set of constraints to narrow the ligand binding site is applied to the same initial structure. This time the expected result is ICL3's staying in an open position as opposed to closure. As shown in Table 4.3, the first three Ser distances (Ser203, 204 and 207 in H5 with Asp113 in H3) are set to lower distances than those of the inactive state. The other four distances are kept at the inactive state values. In Figure 4.22a, all the constraint distances are shown together for *c2_initial_closed*. Sudden drops are seen at around 17 ns and 167 ns in three constrained distances (Asn293-Asp113, Phe289-Asp113 and Asn312-Asp113). The reason for these changes is that the structure fails to adjust these distances to constrained values, and stabilizes them at slightly higher values. The constraint distances may affect each other, too. Similar behavior is observed for three distances of Asn293-Asp113, Phe289-Asp113 and Asn312-Asp113 in *c1_initial_open*. Figure 4.22b shows that the unconstrained C^{α} distances fluctuate more than the constraint distances that are applied in *c2_initial_closed* just like in *c1_initial_open*.

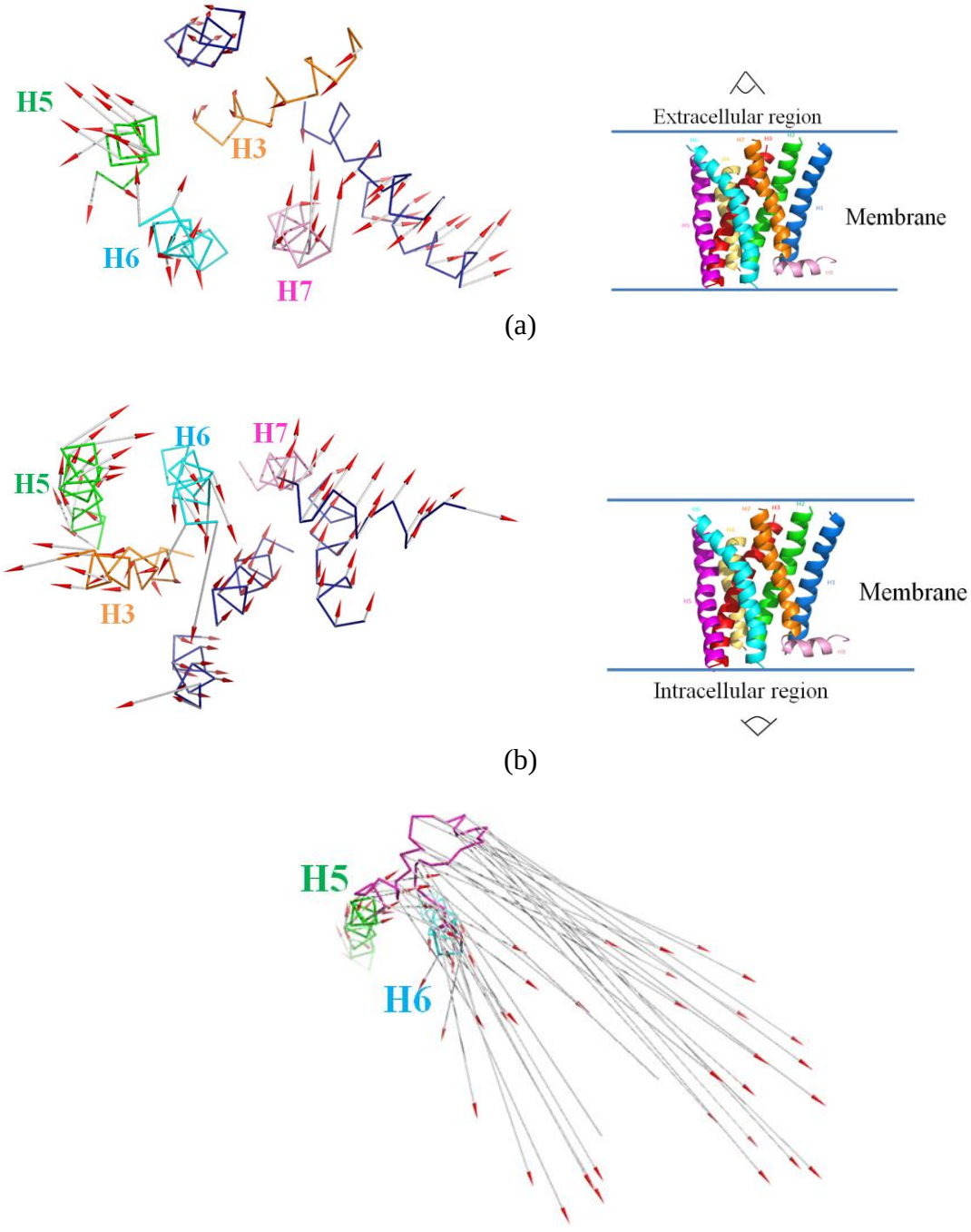
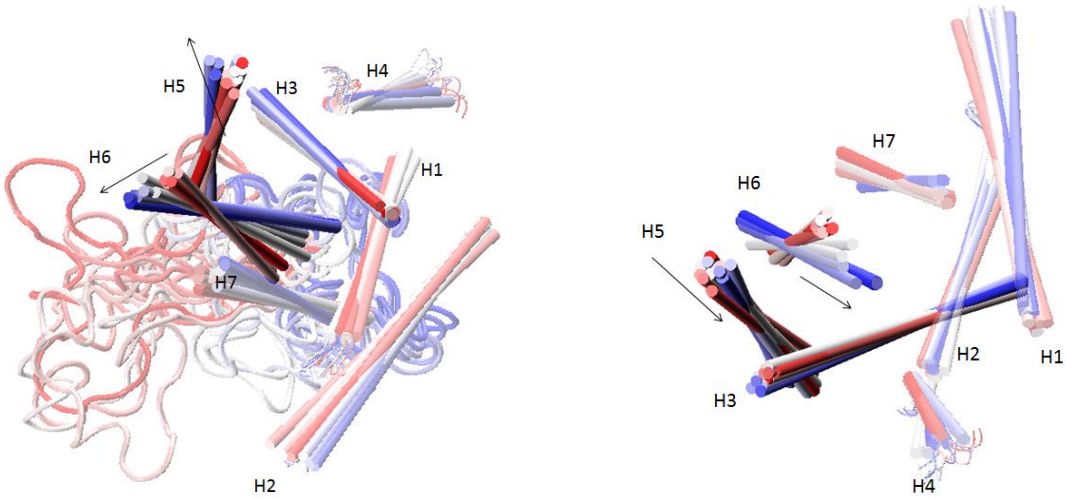
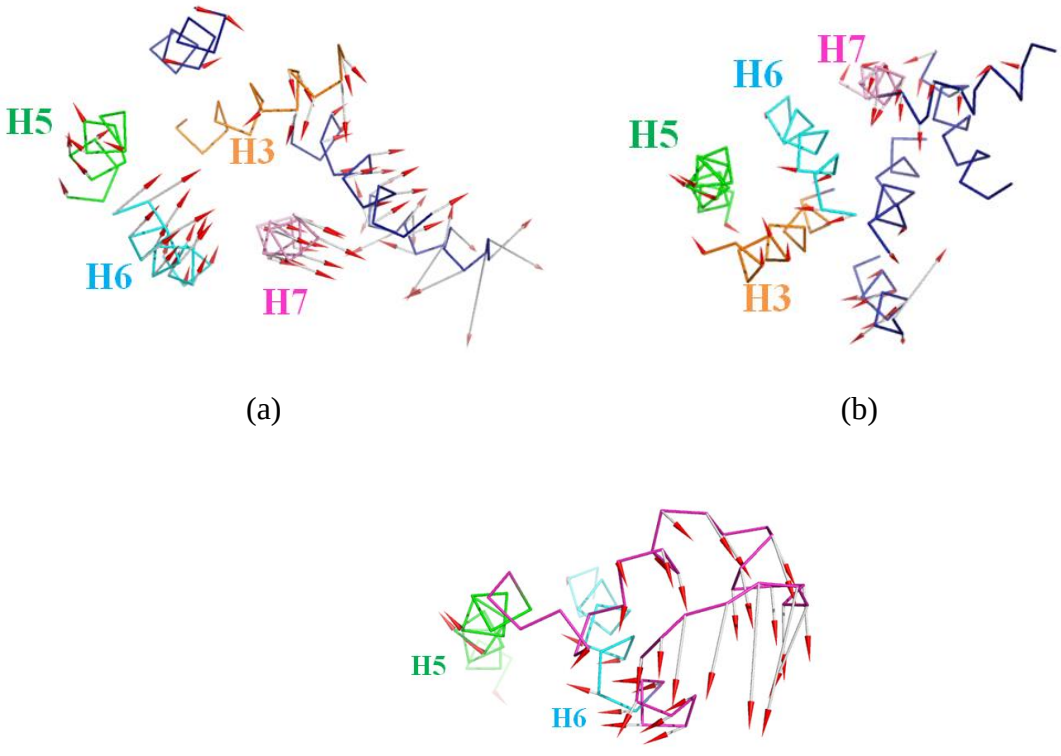


Figure 4.14. First principal mode vectors of *original* (a) upper view, (b) lower view, (c) lower view with H5, H6 and ICL3.



(a) (b)

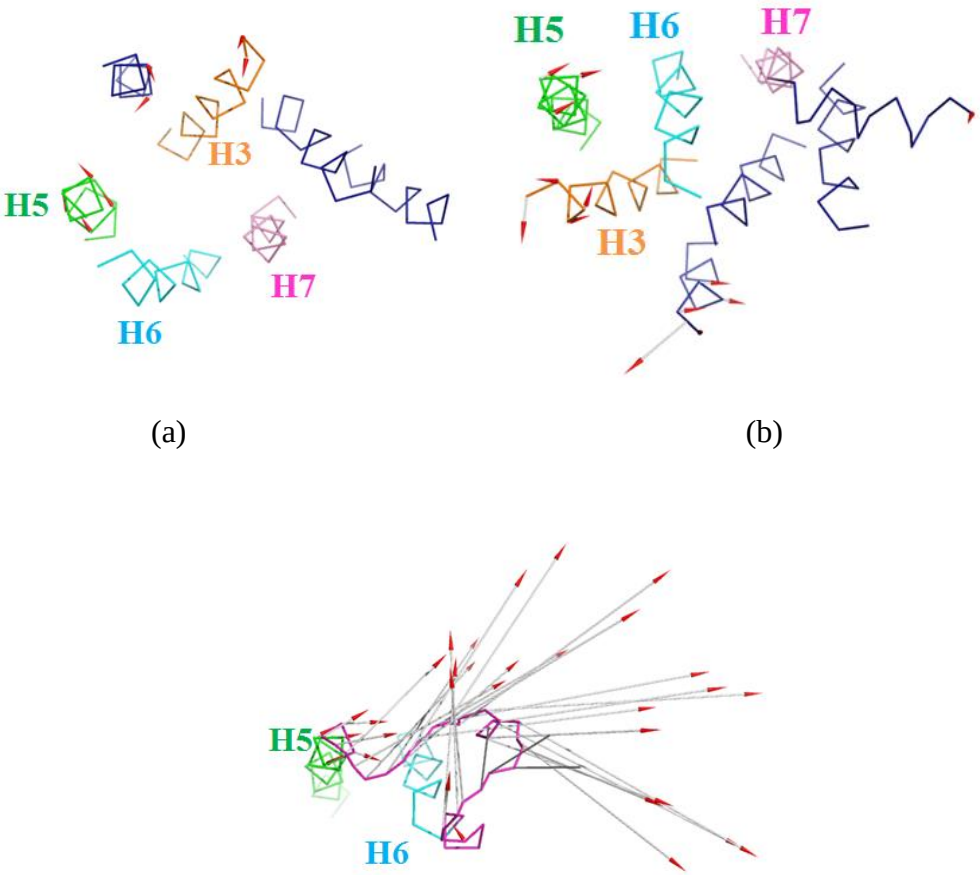
Figure 4.15. Snapshots of *original* with detailed H3, H5 and H6 from (a) top and (b) bottom views.



(a) (b)

(c)

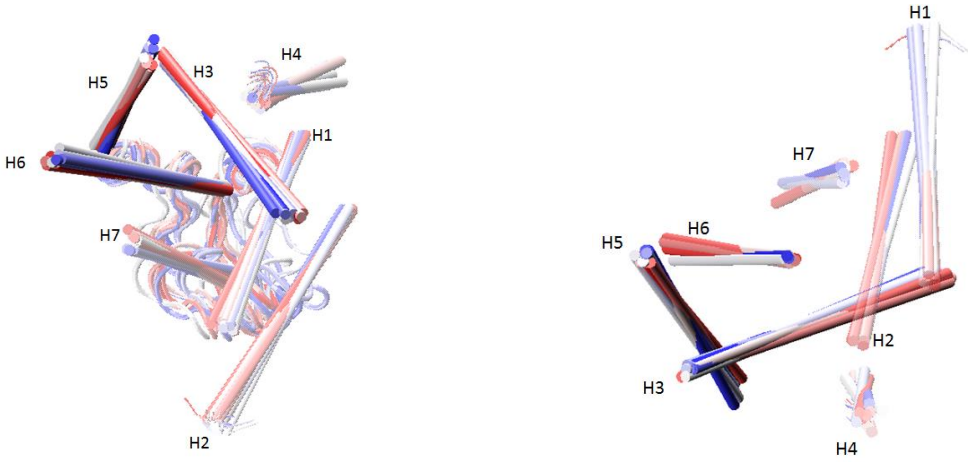
Figure 4.16. First principal mode vectors of *continued_1* (a) upper view, (b) lower view, (c) lower view with H5, H6 and ICL3.



(a) (b)

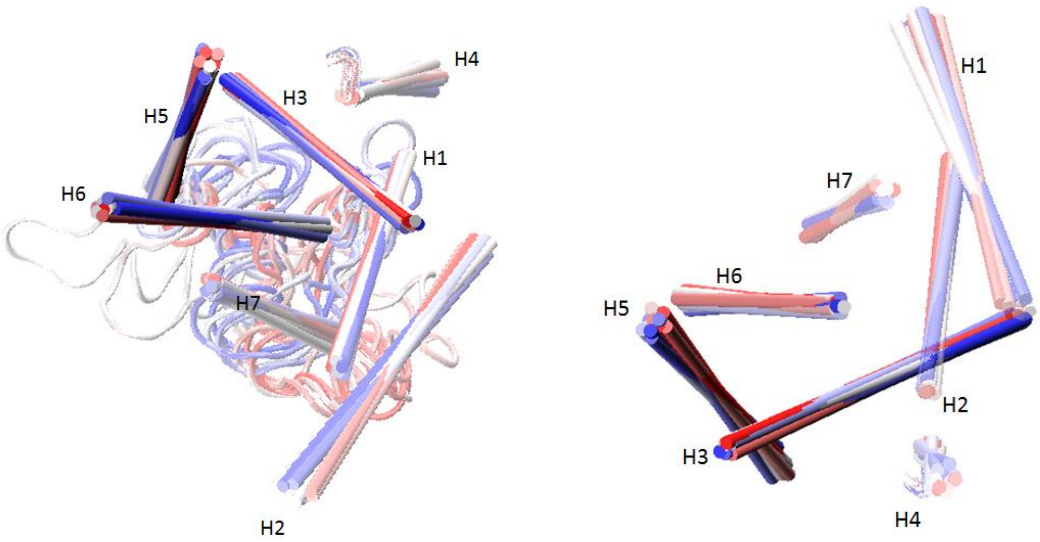
(c)

Figure 4.17. First principal mode vectors of *continued_2* (a) upper view, (b) lower view, (c) lower view with H5, H6 and ICL3.



(a) (b)

Figure 4.18. Snapshots of *continued_1* with detailed H3, H5 and H6 from (a) top and (b) bottom views.



(a) (b)

Figure 4.19. Snapshots of *continued_2* with detailed H3, H5 and H6 from (a) top and (b) bottom views.

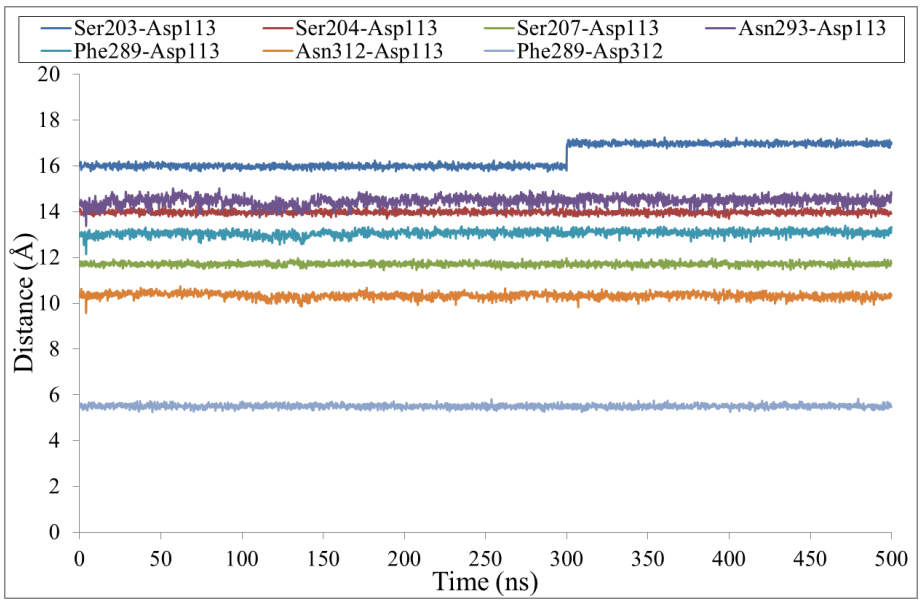
When center of mass profile is analyzed in Figure 4.23a, the open and stable position of ICL3 can be observed easily. Figures 4.23b and 4.23c show 20 different snapshots of *c2_initial_closed* from side and lower views. ICL3 region stays open and away from the receptor's G-protein binding site.

4.3.1. RMSD Profiles

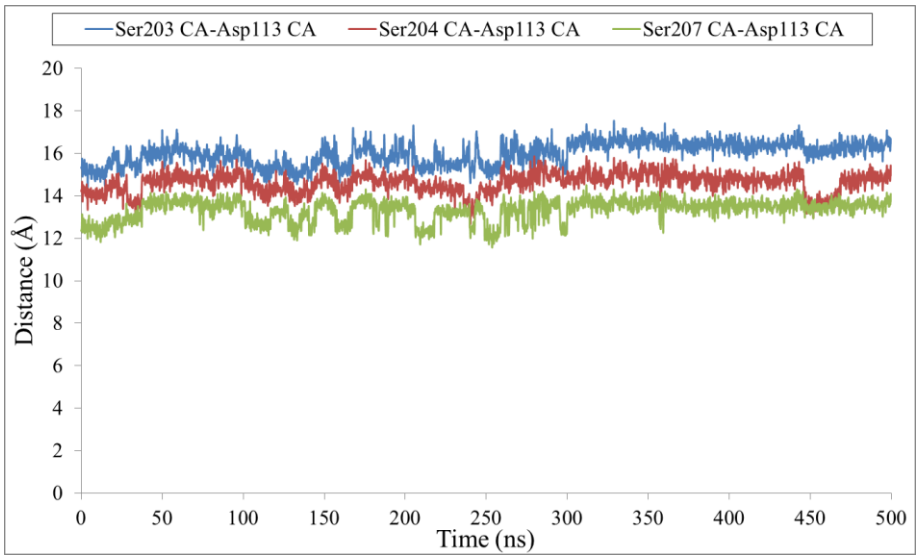
The RMSD profile of *c1_initial_open* in Figure 4.24a shows that there is a constant increase in the *all_fit_core RMSD* values because of the presence of ICL3. The other two curves, which do not include ICL3 region do not show an increase and are stable. The increase to 6, 8 and 10 Å in the *all_fit_core RMSD* profile of *c1_initial_open* is similar to that of *original* (See Figure 4.3a).

RMSD of lower part of H6 profile of *c1_initial_open* in Figure 4.24b shows that this region is stable compared to *original*'s results in Figure 4.4a, where the RMSD with respect to inactive structure is around 1 Å for nearly 500 ns and then increases to 4 Å towards the end. The RMSD with respect to active structure starts around 7 Å and increases to 10 Å starting around 500 ns. In *c1_initial_open*'s profile, the RMSD with

respect to inactive structure starts at around 1 Å and goes to 2 Å and stays there until 292 ns. Then it increases to 3 Å and stays around there for the rest of the trajectory. The RMSD with respect to active structure starts around 8 Å and increases to 9 Å around 292 ns at the same time with the RMSD with respect to inactive structure. These results show that the RMSD with respect to active structure of *original* has 3 Å increase while that of *c1_initial_open* has only 1 Å increase. The lower part of H6 of *c1_initial_open* is not affected by the movement of ICL3 as much as *original*.

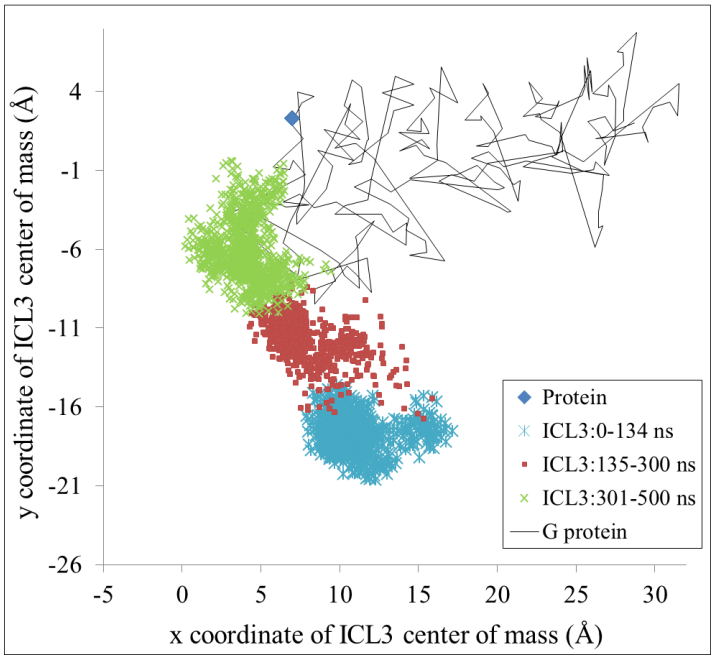


(a)

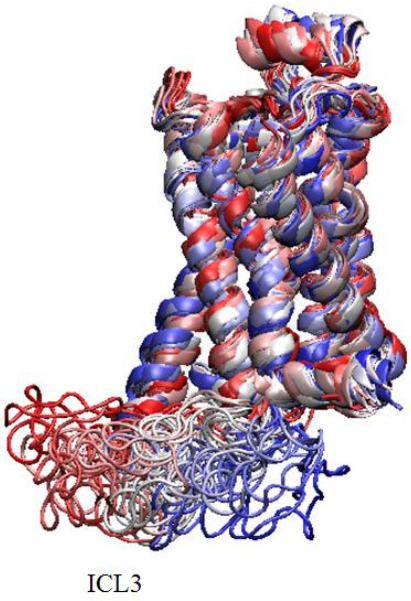


(b)

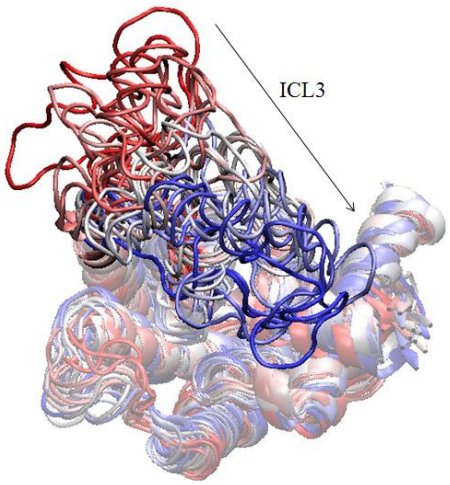
Figure 4.20. Distance profiles of *c1_initial_open* for (a) seven different distance constraints and for (b) C^{α}'s (CA) of three different unconstrained distances.



(a)

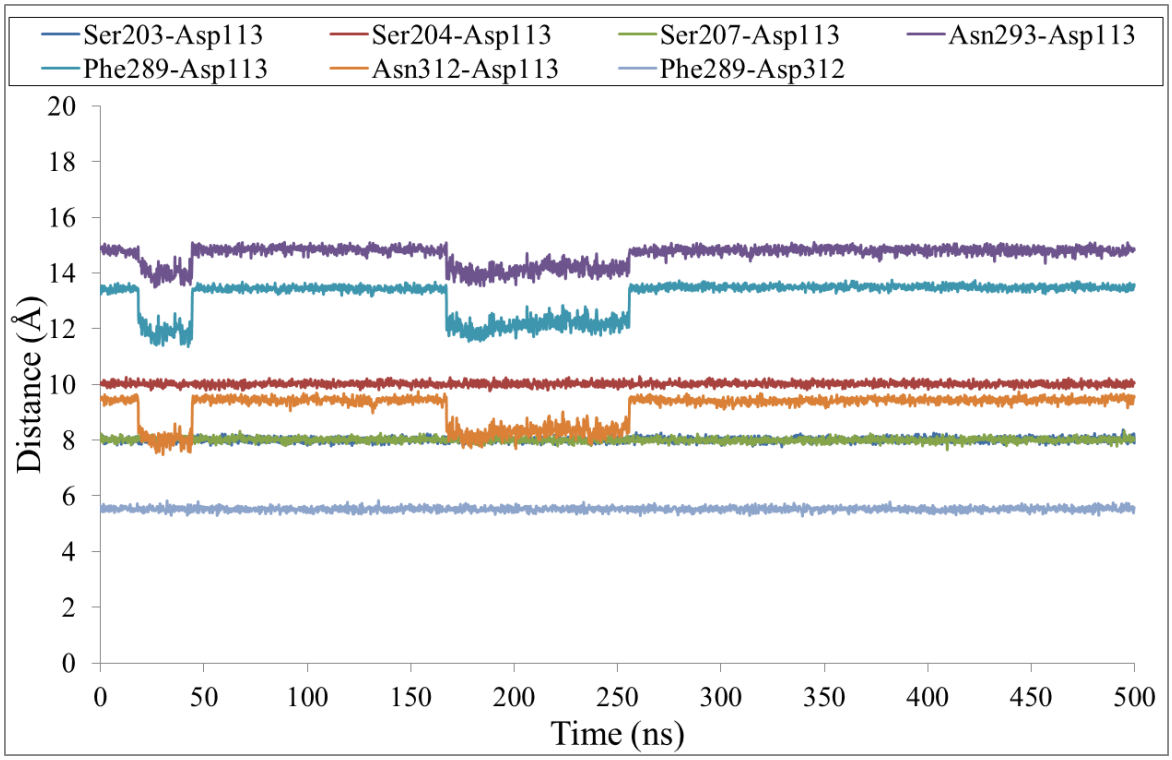


(b)

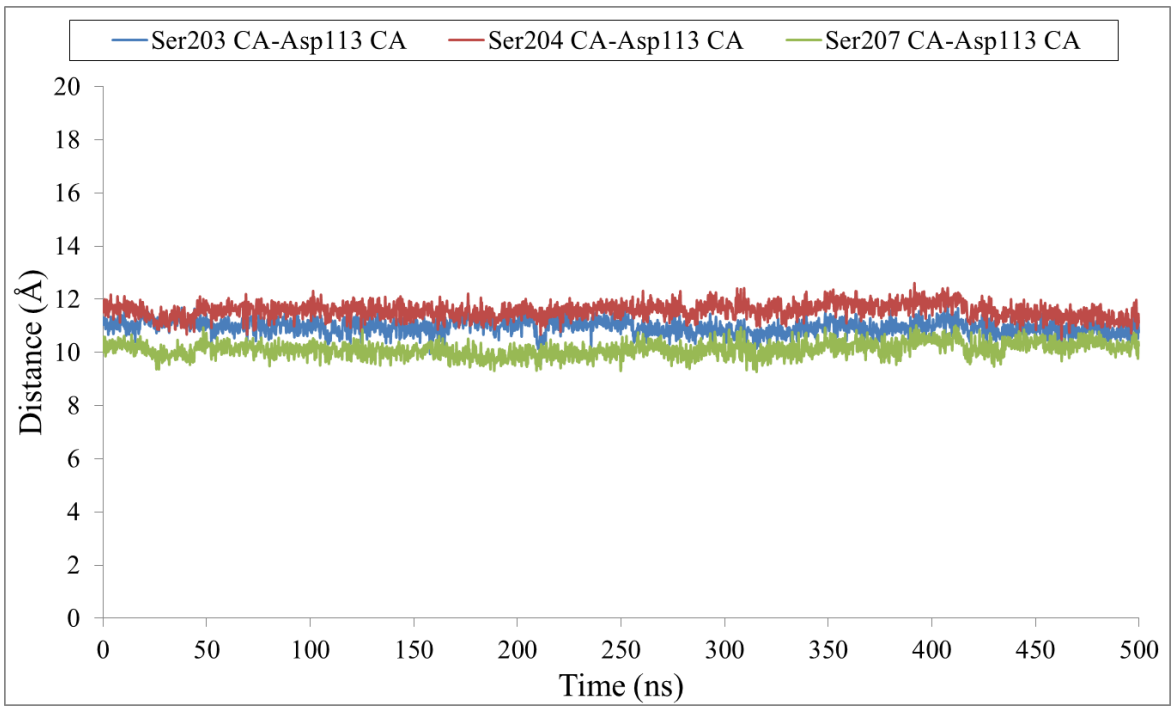


(c)

Figure 4.21. *c1_initial_open*'s (a) variations in ICL3's center of mass profile and snapshots from (b) side and (c) lower views.

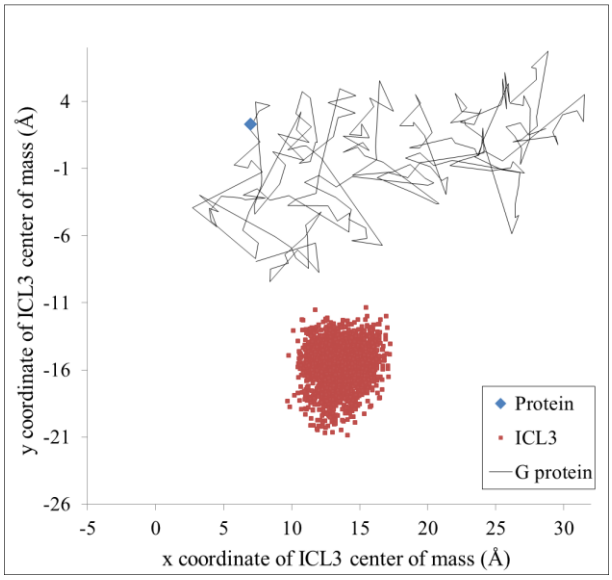


(a)

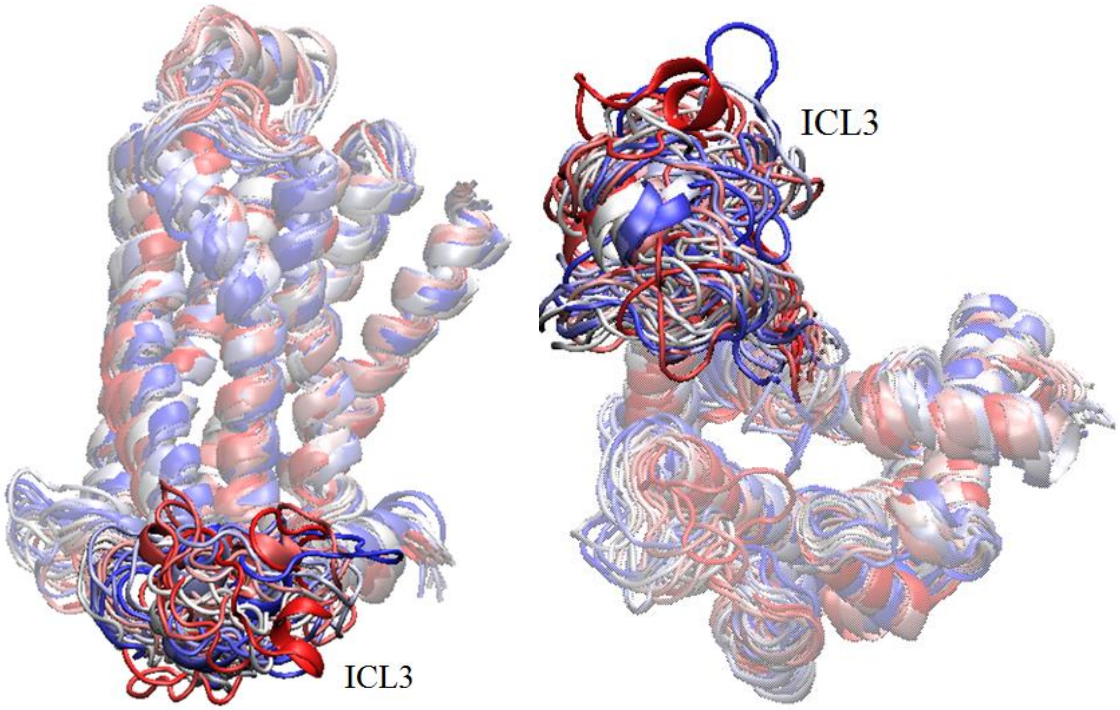


(b)

Figure 4.22. Distance profiles of *c2_initial_closed* for (a) seven different distance constraints and for (b) C^{α}'s (CA) of three different unconstrained distances.



(a)



(b) (c)

Figure 4.23. *c2_initial_closed*'s (a) variations in ICL3's center of mass profile and snapshots from (b) side and (c) lower views.

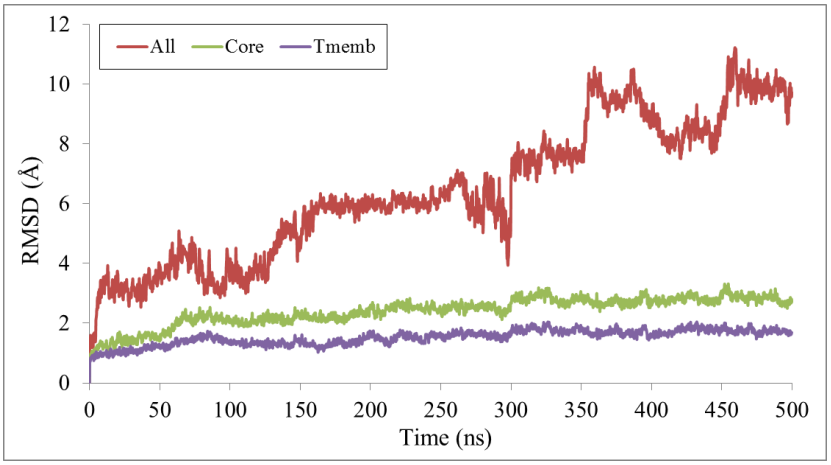
Figure 4.24c supports Figure 4.24b because no major change in the lower part of H6 is observed from also the snapshot of *c1_initial_open*. Ten snapshots are aligned and only the lower part of H6 is shown clearly for a better understanding from side and lower views. The initial conformation is shown in red, the intermediate in white and the final in blue.

Besides the lower region of H6, the lower region of H5 in Figure 4.25 is also investigated since it has a contact with ICL3, too. Generally, an increase is observed although there is a slight decreasing tendency between 105-201 ns. Before 105 ns, the RMSD value is around 2.5 Å while after 201 ns this value becomes around 3.5 Å. When these results of *c1_initial_open* are compared with the results of *original*, it can be concluded that there is not a major difference between the two profiles although the RMSD of lower part of H5 of *original* fluctuates more. The snapshots do not show a distinct profile, either.

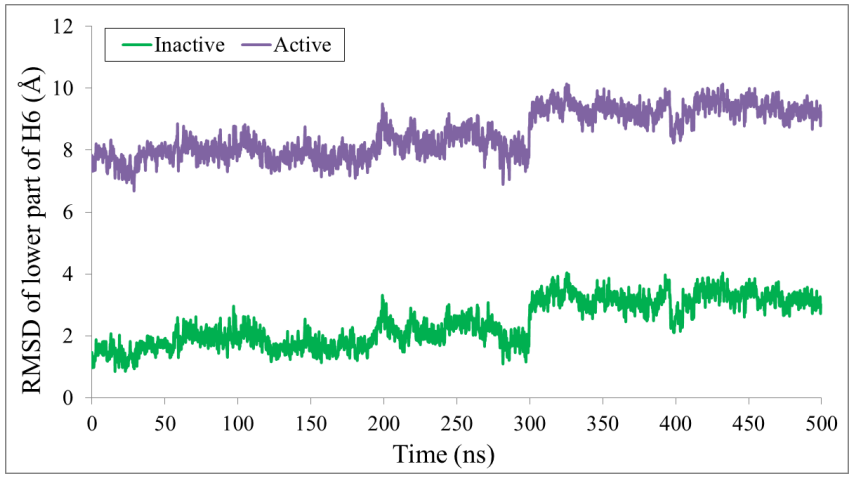
In order to see if the last snapshots of *original* and *c1_initial_open* coincide with each other after alignment, Figure 4.26 is prepared with *original* in red and *c1_initial_open* in green. It can be concluded that their ICL3 coincide well blocking the G-protein's binding site.

The RMSD profile of *c2_initial_closed* in Figure 4.27a shows that the *all_fit_core RMSD* values are almost stable around 5 Å because ICL3 does not move as much as in *c1_initial_open*. The *core_fit_core RMSD* and *transmembrane_fit_transmembrane RMSD* also seem to stay stable around 3 Å and 2 Å, respectively.

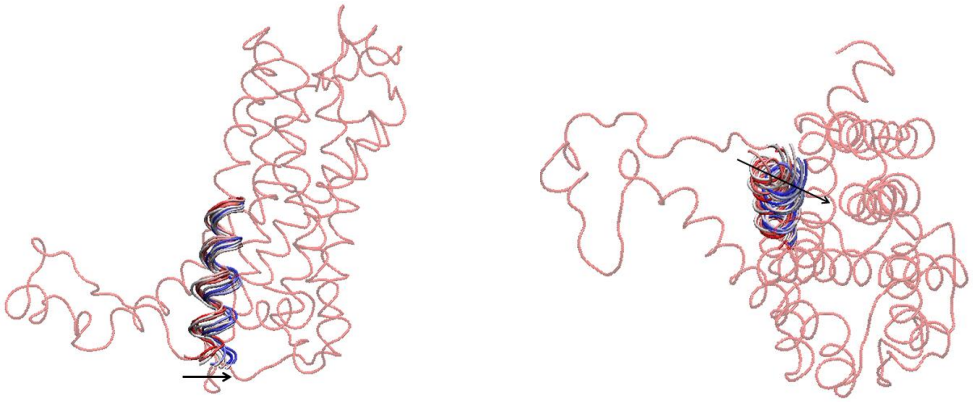
RMSD of lower part of H6 profile of *c2_initial_closed* in Figure 4.27b shows that the increases resemble slightly to those of *original'*s results in Figure 4.3a where the RMSD with respect to inactive structure increases from 1 Å to 4 Å. The RMSD with respect to active structure increases from 7 Å to 10 Å. In *c2_initial_closed*'s profile, the RMSD with respect to inactive structure starts at around 2 Å and increases to 4 Å around 365 ns and stays there for the rest of the trajectory. The RMSD with respect to active structure starts around 8 Å and increases to around 10 Å around 365 ns as the RMSD result with respect to inactive structure. These results show that both of the models' last RMSD values are the same.



(a)

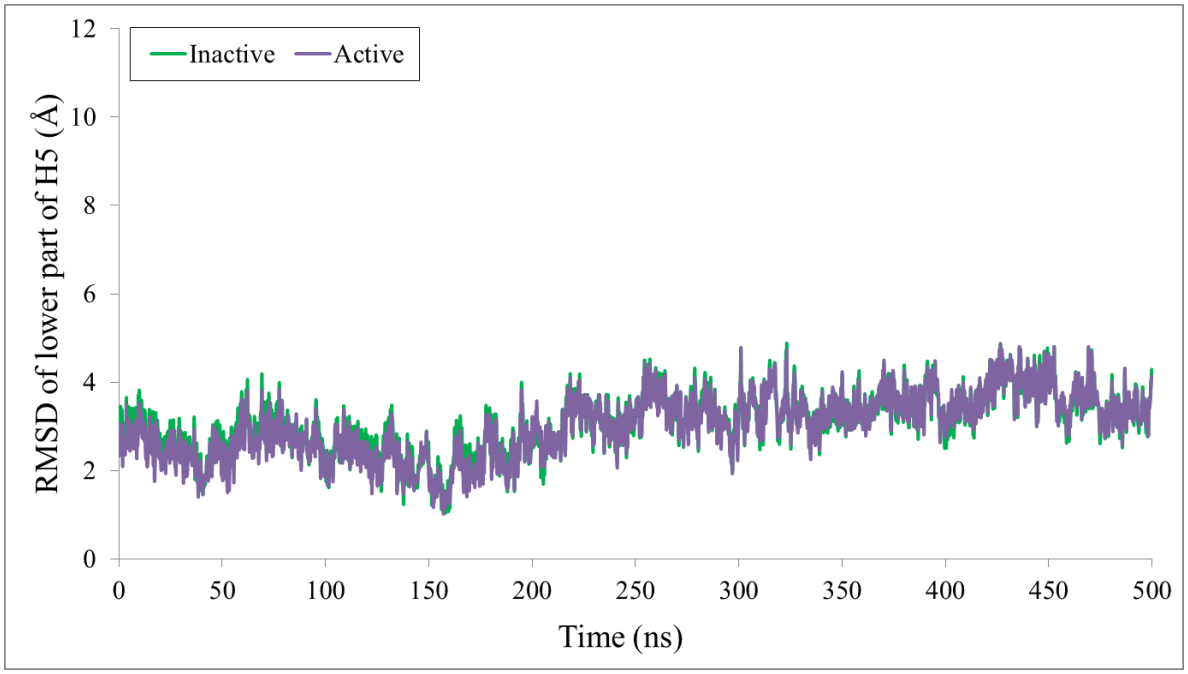


(b)

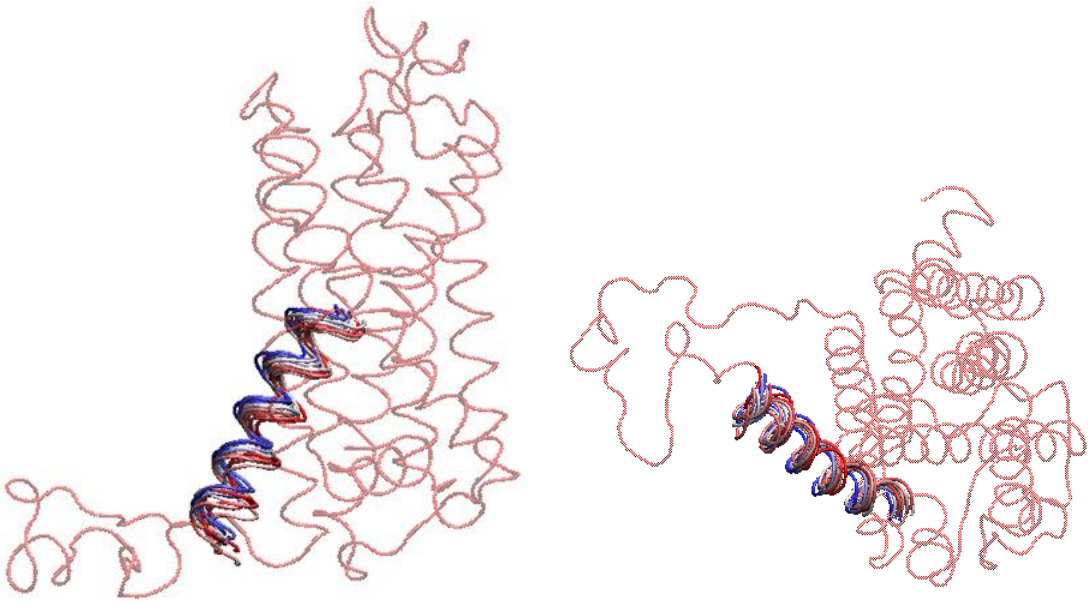


(c) (d)

Figure 4.24. *c1_initial_open*'s (a) RMSD profile, (b) RMSD of lower part of H6 profile and ten aligned snapshots from (a) side and (d) lower views.

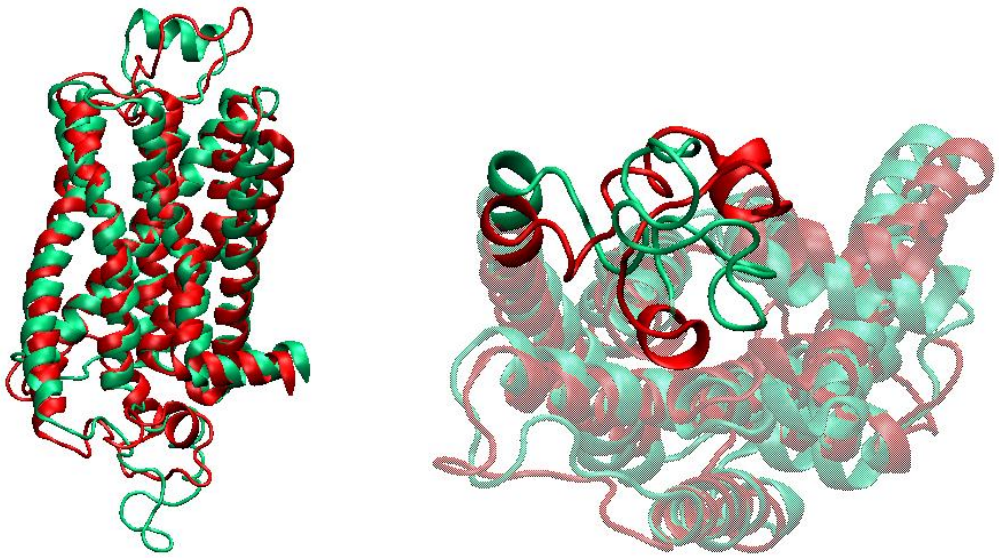


(a)



(b) (c)

Figure 4.25. *c1_initial_open*'s (a) RMSD of lower part of H5 and ten aligned snapshots from (b) side and (c) lower views.



(a) (b)

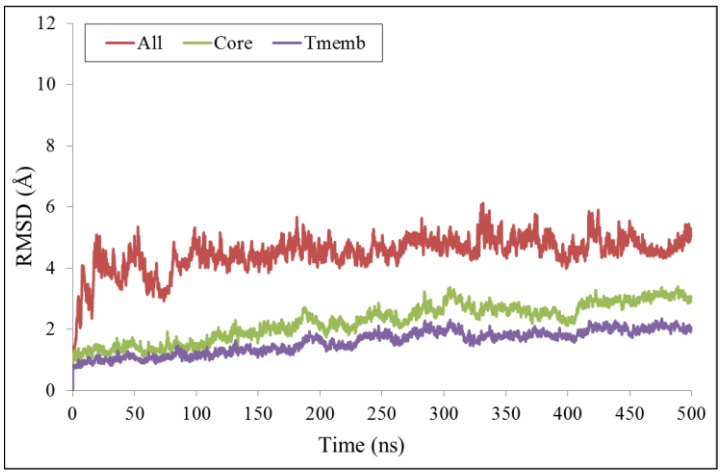
Figure 4.26. Snapshot profiles of the last snapshots of *original* and *c1_initial_open* from (a) side and (b) lower views.

1.5 Å increase in the RMSD of lower part of H6 in Figure 4.27b can be caused by the deformation of the helix near ICL3 as seen in Figure 4.27c. The last snapshot seems to deform towards the inside of the receptor. The initial and middle snapshots are nearly in the same position.

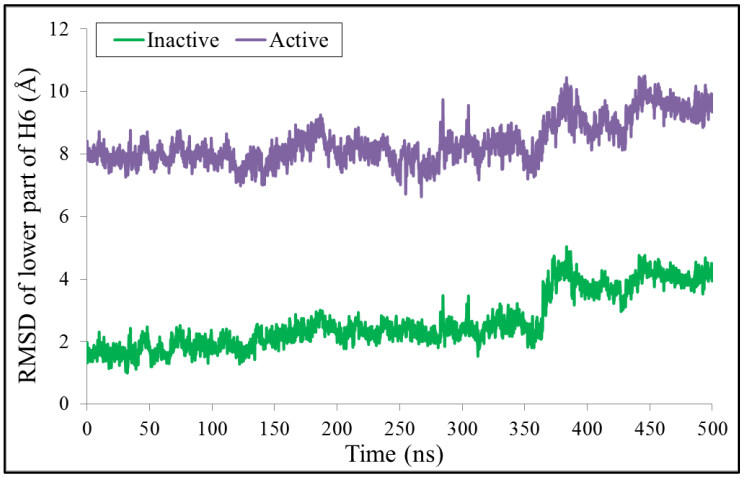
The lower region of H5 profile of *c2_initial_closed* in Figure 4.28a shows that the RMSD of lower part of H5 results start around 2.5 Å and it moves to 3 Å around 40 ns. It stays there until 265 ns when it increases to around 3.5 Å. Then it has a slight increasing profile to 5 Å till the end. Therefore, it can be generalized that there is an increase in the RMSD results. The RMSD value of *c1_initial_open* increases from 2.5 Å to 3.5 Å, while that of *c2_initial_closed* increases more from 2.5 Å to 5 Å. This result can be seen from the snapshots where the last snapshots in Figure 4.28b seem to move outside the receptor, indicating that ICL3 pulled them outwards while it stayed open.

Previously, the constrained distance between Ser207-Asp113 for *c1_initial_open* and *c2_initial_closed* were reported to be 11.7 Å and 8 Å, respectively. Figure 4.29 shows these distance values vs the RMSD of lower parts of H6 for the two models. For *c1_initial_open*, the RMSD changes between 6.5 Å and 10 Å, while that of

c2_initial_closed changes between 6.5 Å and 10.5 Å showing that *c2_initial_closed* has a wider movement of lower part of H6.



(a)



(b)

(c) (d)

Figure 4.27. *c2_initial_closed*'s (a) RMSD profile, (b) RMSD of lower part of H6 profile and ten aligned snapshots from (c) side and (d) lower views.

4.3.2. Conformational Dynamics and PCA

According to principal component analysis results shown in Figure 4.30, the percentage explanation of principal components for *c1_initial_open* coincides with that of *original*. The reason may be that similar type of ICL3 closing is observed. In contrast, *c2_initial_closed* has lower cumulative % explanation results for PCA modes.

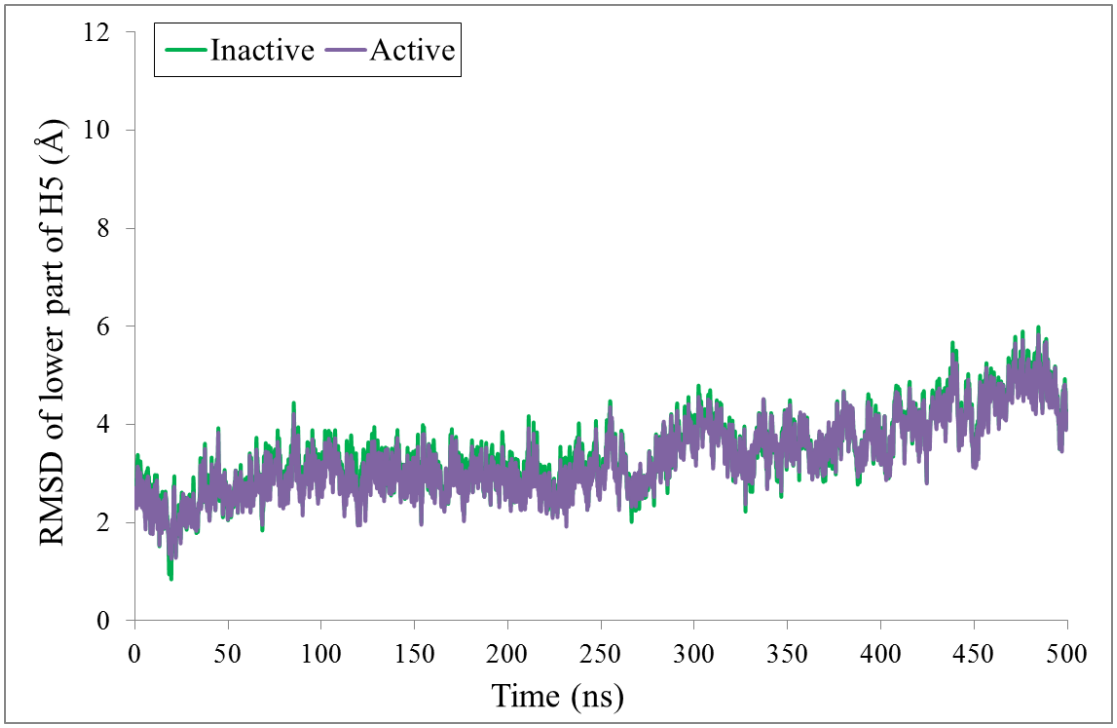
The overlap matrix u_{jk}^{AB} can be calculated by taking the dot product or the inner product of eigenvectors u_j^A and u_k^B for every PCA mode.

$$u_{jk}^{AB} = u_j^A \cdot u_k^B \tag{4.1}$$

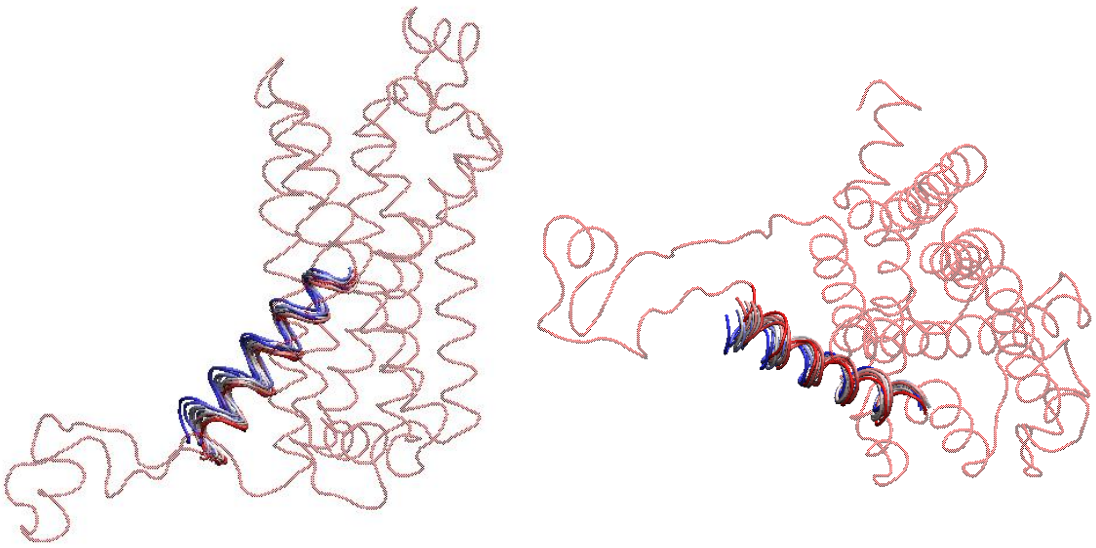
A and *B* are distinct eigenvector groups from independent trajectories. *j* and *k* are the PCA mode indices. The values near to 1 indicate a good match between the two modes of two different trajectories. The values near to 0 show bad overlaps. The overlap matrix of *original*'s first 20 PCA modes and *c1_initial_open*'s first 20 PCA modes are calculated and found as in Figure 4.31. Here, the first PCA modes of both models are very alike with a score of around 0.9, supporting the PCA % explanation.

Figure 4.32 shows the first principal mode vectors of *c1_initial_open,* which exhibits the longest ICL3 vectors so far. The ICL3 moves inside and it is directed by the lower side of H6, which also moves inside. The allosteric relation between the ligand binding site and the ICL3 of *original* can be reported for this model, too. When the ligand binding site is kept at an open specific constraint distance, ICL3 closes under the receptor and stays there.

The first principal mode vectors of *c2_initial_closed* are shown in Appendix A Figure A.1 because of the fact that *c2_initial_closed* has very low cumulative % explanation for the first PCA mode (around 20%).



(a)



(b) (c)

Figure 4.28. *c2_initial_closed*'s (a) RMSD of lower part of H5 and ten aligned snapshots from (b) side and (c) lower views.

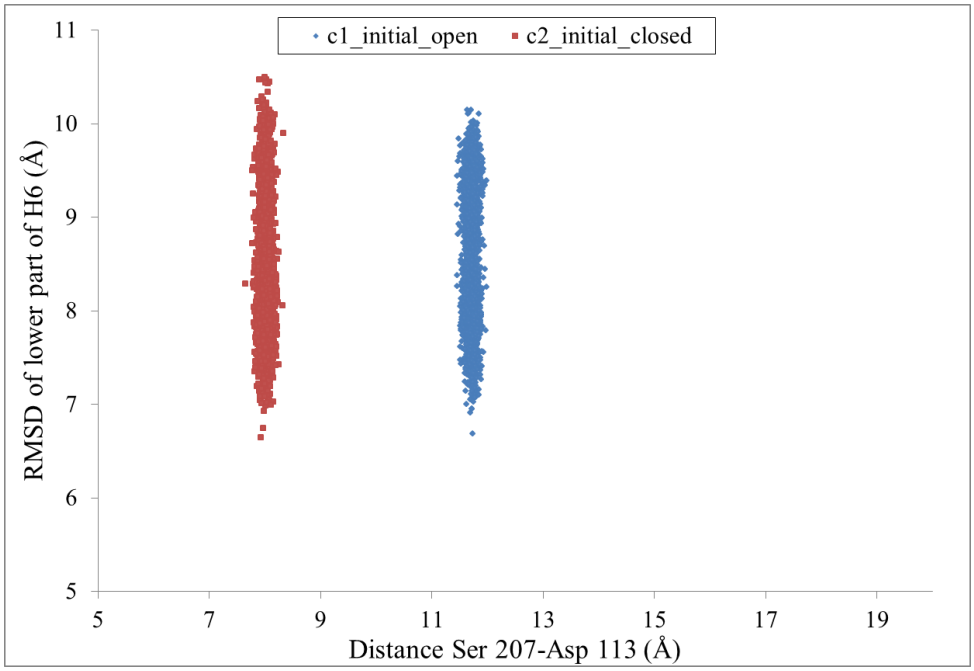


Figure 4.29. Distance Ser207-Asp113 vs RMSD of lower part of H6 profiles of *c1_initial_open* and *c2_initial_closed*.

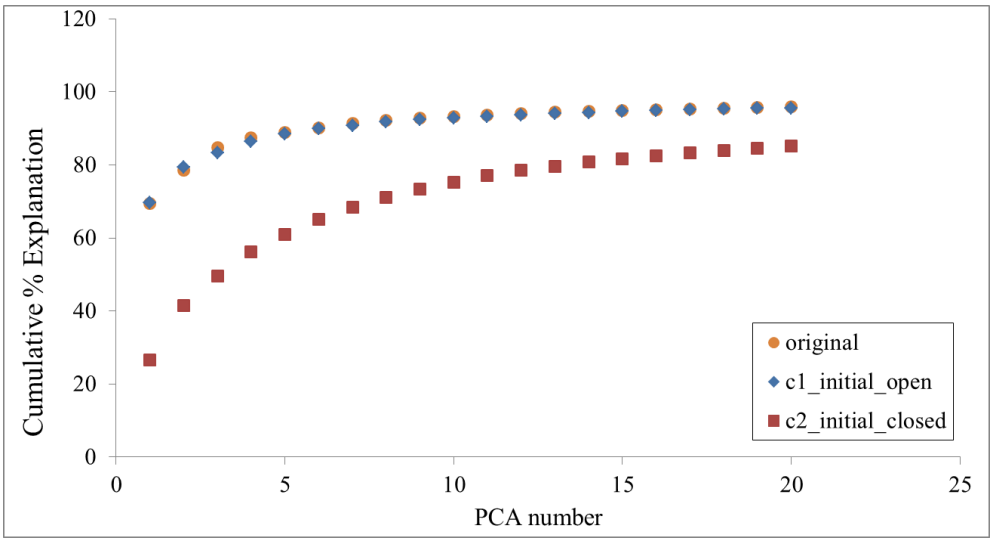


Figure 4.30. Cumulative % explanation of principal components for *c1_initial_open* and *c2_initial_closed* compared with *original*.

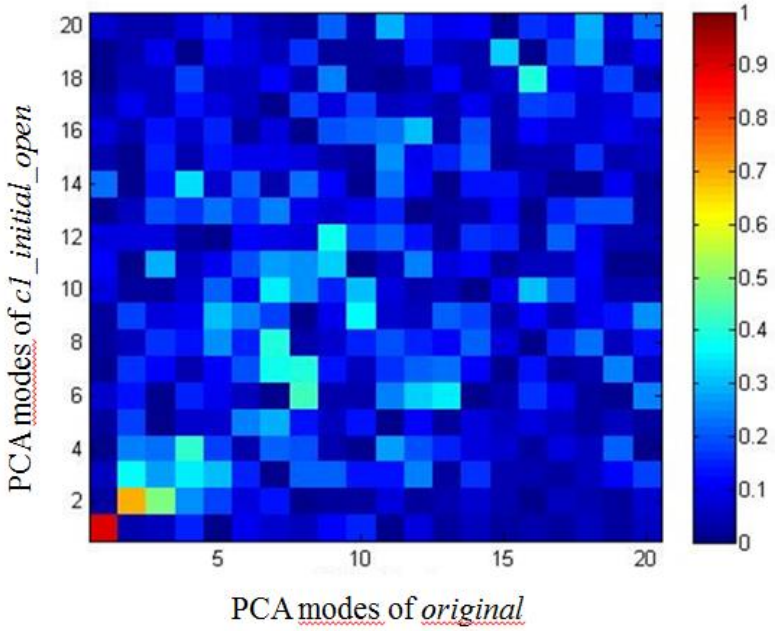


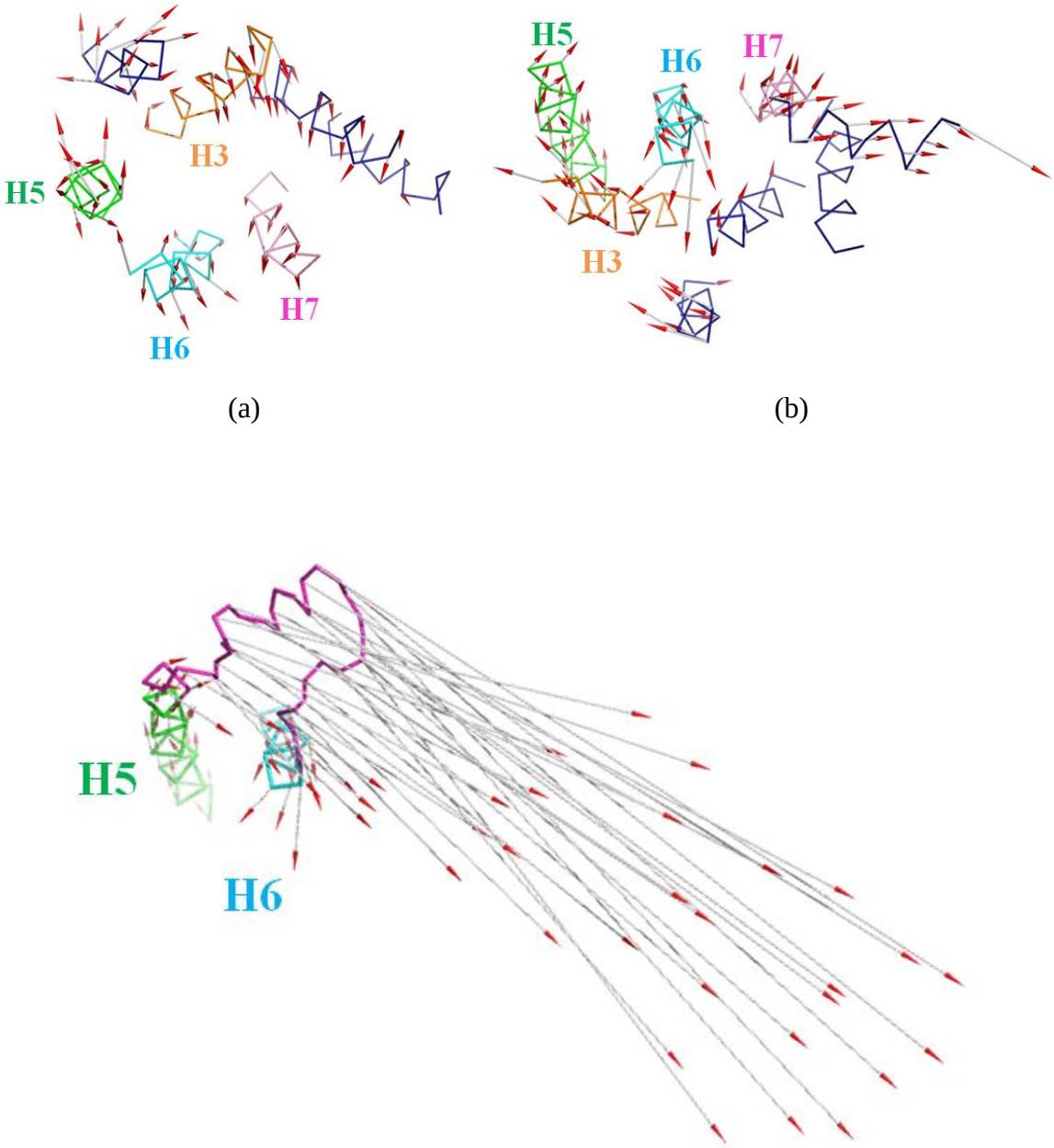
Figure 4.31. Overlap matrix between PCA modes of *original* and *c1_initial_open*.

Figure 4.33 show the displacements of H3, H5 and H6 relative to each other in ten snapshots for *c1_initial_open* where the upper parts of H5 and H6 move outside and the lower parts of H5 and H6 move inside, showing the direction where ICL3 pulls them very clearly.

Figure 4.34 shows snapshots of *c2_initial_closed* where H5 seems to be broken on the upper side because of the constraints applied on the Ser residues. The lower parts of H5 and H6 move outside since ICL3 pulls them away from the receptor. The upper part of H6 seems to move correlated with the lower part.

4.3.3. RMSF Profiles

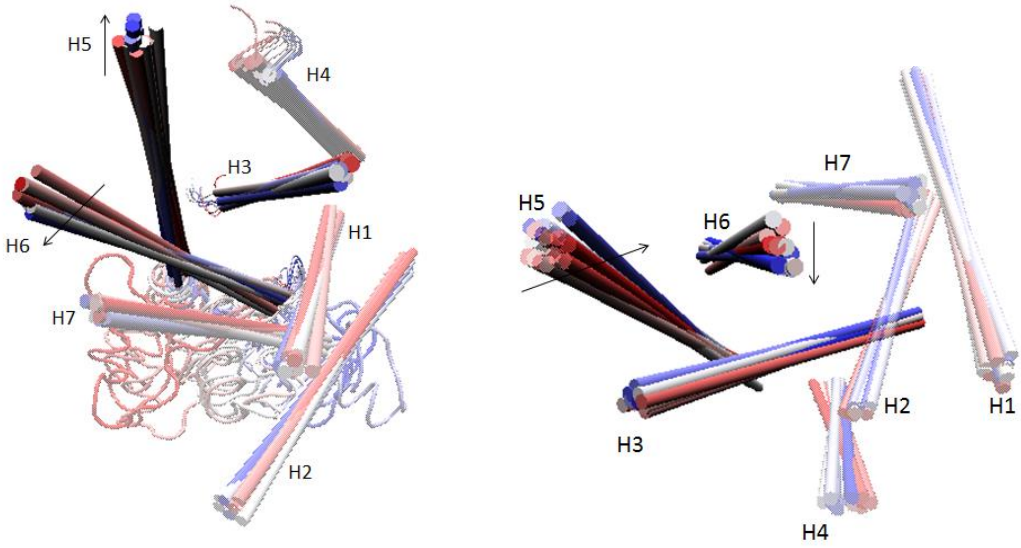
The RMSF results of 500 ns long constrained models in Figure 4.35 show that the highest RMSF value for ICL3 region belongs to *c1_initial_open* due to its rapid closure under the receptor. It is normally expected that when the ligand binding area is closely constrained, ICL3 should move to an open position, or if it is in an open position, it should stay where it is. This idea is supported by *c2_initial_closed* which has the lowest amount of RMSF for ICL3 region.



(a) (b)

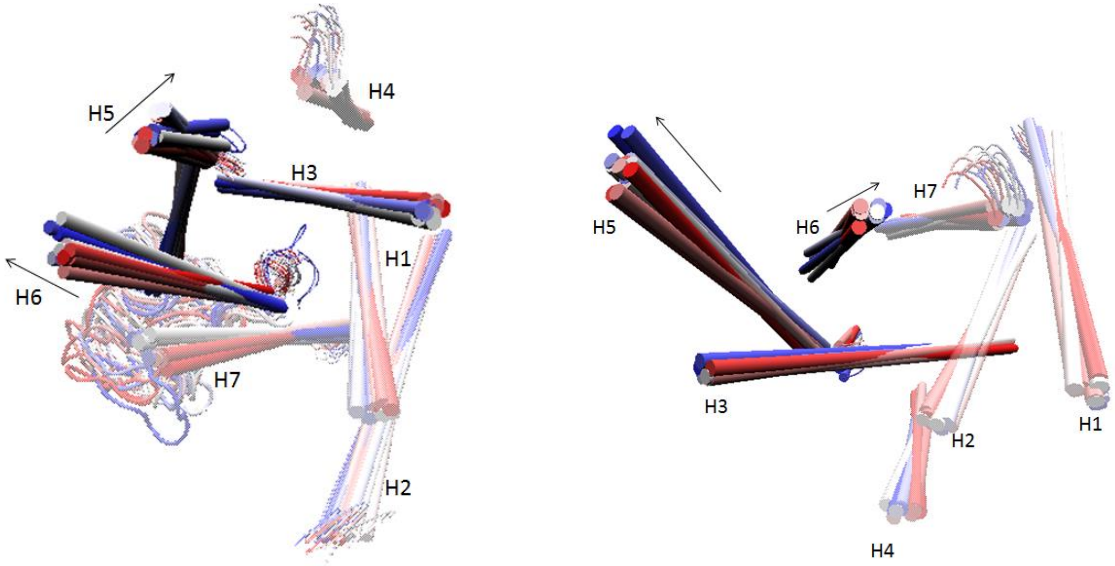
(c)

Figure 4.32. First principal mode vectors of *c1_initial_open* with (a) upper view, (b) lower view, (c) lower view with H5, H6 and ICL3.



(a) (b)

Figure 4.33. Snapshots of *c1_initial_open* with detailed H3, H5 and H6 from (a) top and (b) bottom views.



(a) (b)

Figure 4.34. Snapshots of *c2_initial_closed* with detailed H3, H5 and H6 from (a) top and (b) bottom views.

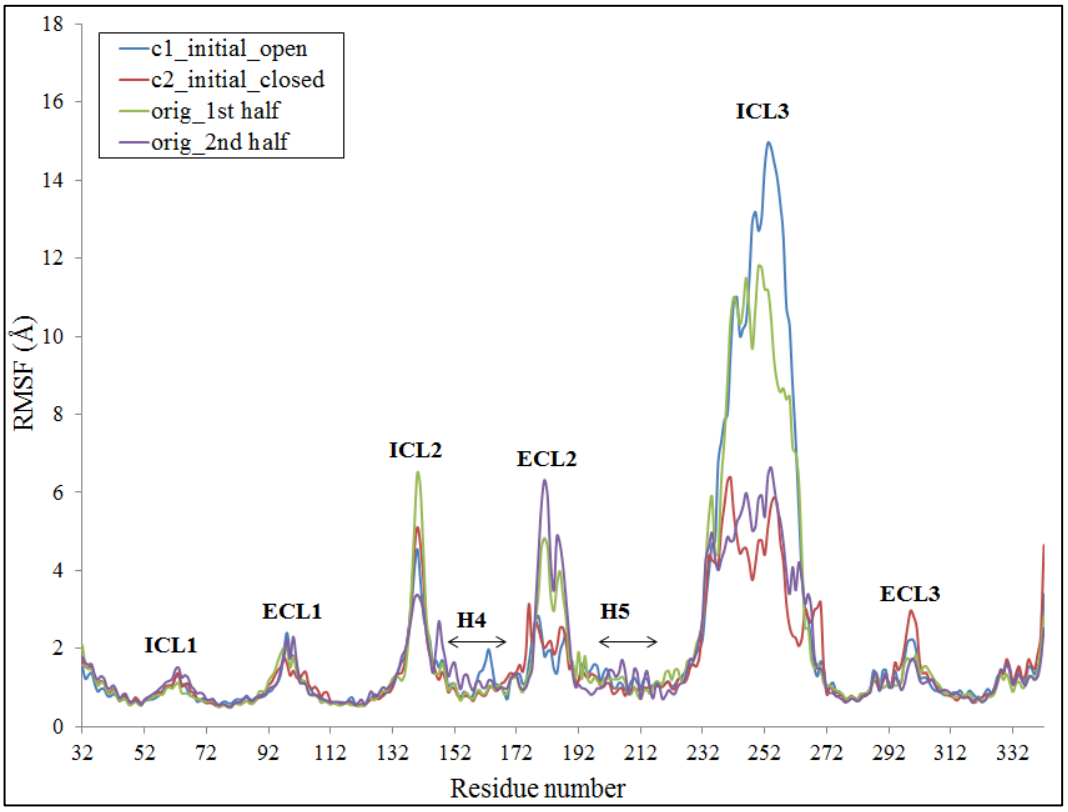


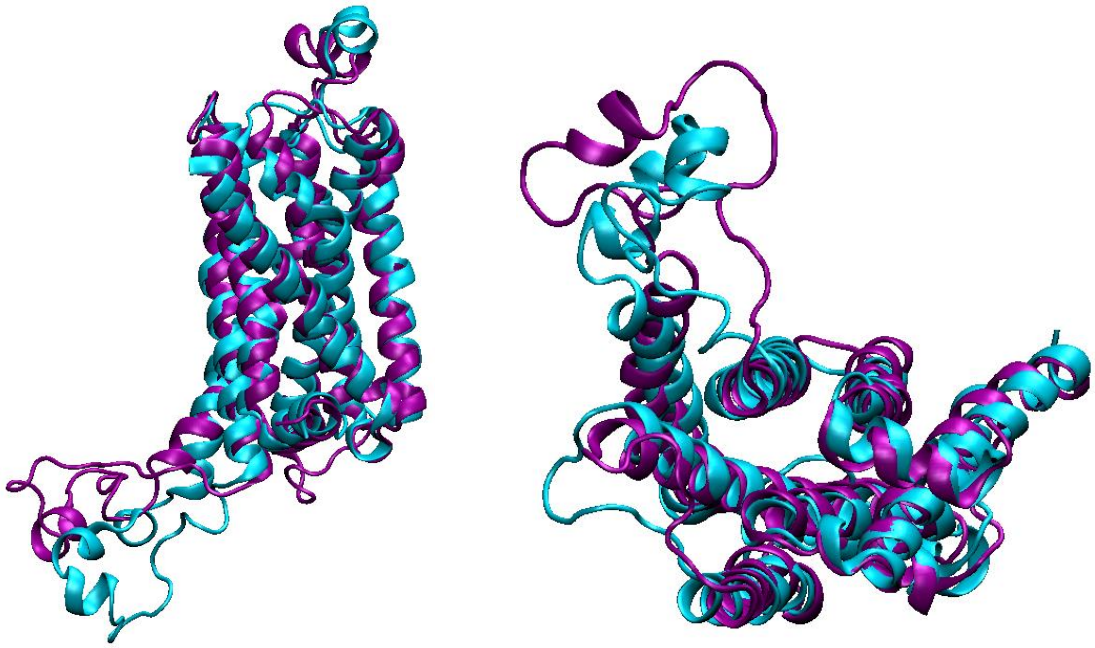
Figure 4.35. RMSF values of the 500 ns long initial constrained models.

4.4. Constraint Runs with a Half-closed (Intermediate) ICL3

The two models in this section, namely *c3_interm_open* and *c4_interm_closed*, are run in order to see the effect of starting from another snapshot of *original*, which is at 470 ns with an ICL3 position between open and closed. It is expected that when the ligand binding site is kept open, ICL3 with an intermediate position on the way to close would close faster. When the ligand binding site is kept in a narrow position, ICL3 would further open. Specifically, in *c3_interm_open*, Ser203-Asp113 distance is forced to be in a more open position (17 Å) compared to the inactive state of the receptor (11 Å). In *c4_interm_closed*, all the distances are decreased around 8-9 Å, which are 3-5 Å lower than the inactive state distances. The constraints applied are analogous to the ones discussed in the previous section. However, the conformational change in ICL3 is not as expected as the two previous constrained MD results, which will be explained next.

The difference between the initial and intermediate structures can be understood from Figure 4.36, where their first snapshots are aligned to each other. The initial structure

in purple is in a more open position compared to the intermediate structure in cyan both in the side and lower views.

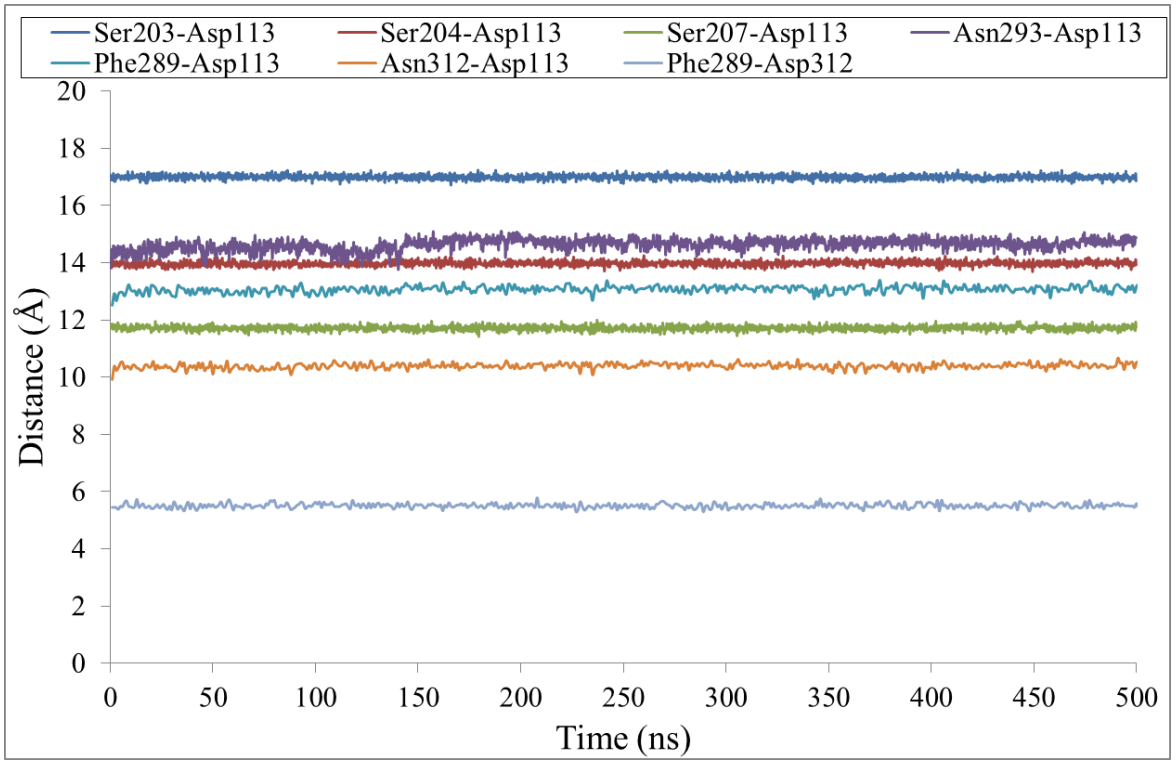


(a) (b)

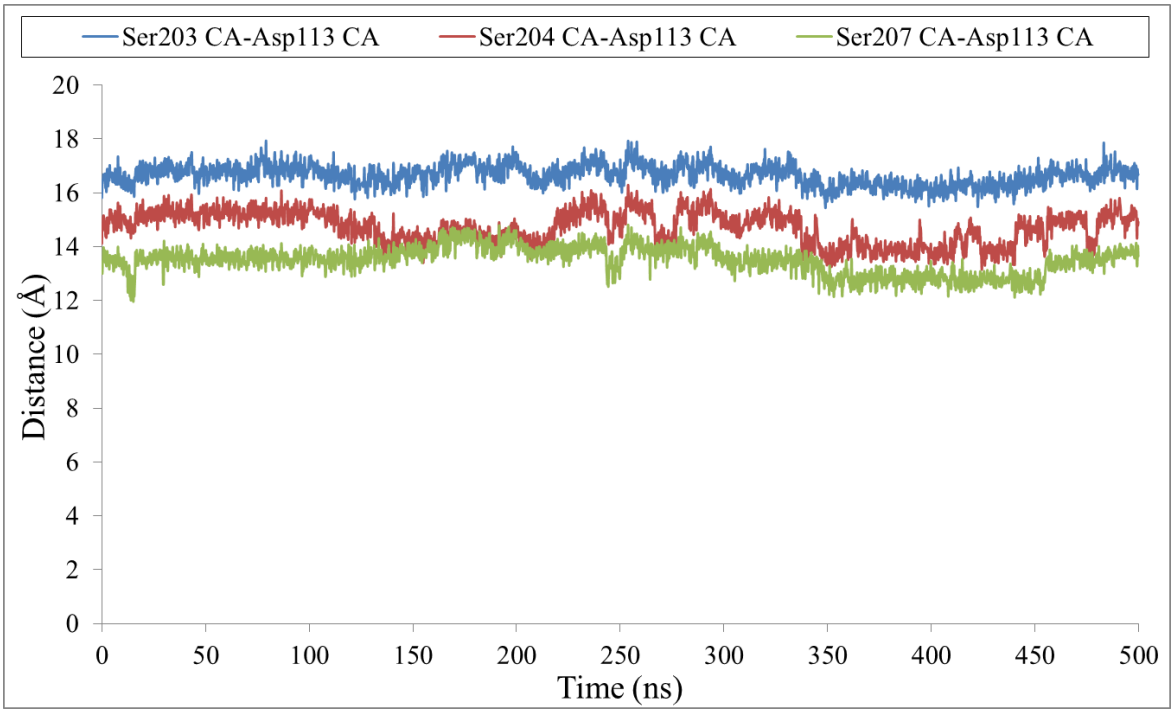
Figure 4.36. The initial snapshots of initial (in magenta) and intermediate (in cyan) structures from (a) side and (b) lower views.

Figure 4.37a shows the distance profiles during *c3_interm_open*. The plotted distances for the Ser residue couples comply with the constraints applied during the run. All the distances except Ser203-Asp113 which is set to 17 Å are kept at the inactive structures' distances. As a result, all the constraint distances except the Ser couples and Phe289-Asp312 have a tendency to increase. In Figure 4.37b, the unconstrained C^{α} distances of Ser couples fluctuate around their constraint couples' distance values.

Figure 4.38a shows the distance profiles of *c4_interm_closed* during the simulation. All the distances are set to lower values than the inactive state distances, also lower than *c2_initial_closed* at some points. All three of the Ser couples' distances are seen as they are set. The other distances show slightly higher profiles. Asn293-Asp113 and Phe289-Asp113 distances can come to the set values only between 300-350 ns. The distance Phe289-Asp312 stays in the set value all the time. Figure 4.38b shows the C^{α} distance profiles of *c4_interm_closed*. All three of these distances are higher than the constrained distance profiles of the same model. They also show the fluctuations between 300-350 ns.



(a)



(b)

Figure 4.37. Distance profiles of *c3_interm_open* for (a) seven different distance constraints and for (b) C^{α}'s (CA) of three different unconstrained distances.

c3_interm_open is analyzed first for the locations of ICL3. Its center of mass profile is separated into different time frames in order to be able to distinguish the movements of ICL3 along the trajectory. In Figure 4.39a, it starts from an intermediate position shown in red, closes a little bit shown in yellow and goes to another half open or intermediate position shown in blue. The expectation based on imposed constraints was that it would close down instead of going back to a relatively open conformation. The snapshots in Figure 4.39b and 4.39c support the center of mass of ICL3 profile and fully closed conformers are not observed. The snapshots form a compact network in which it is hard to distinguish the positions of intermediate and final (blue) snapshots. The initial snapshot (red) has the most open conformation.

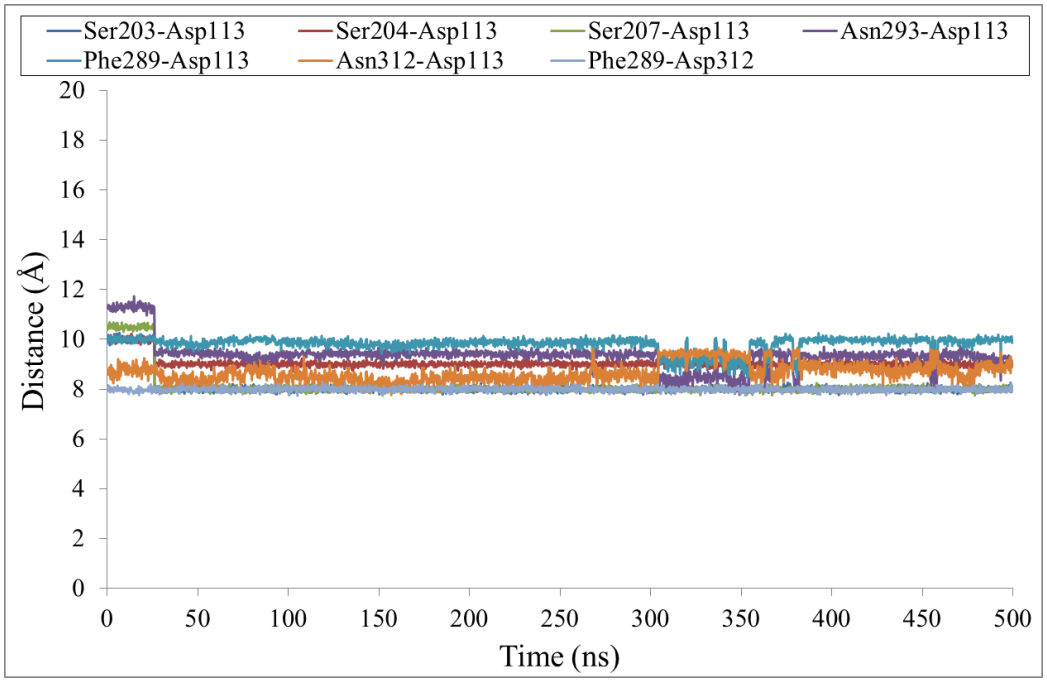
In *c4_interm_closed*, it was expected that ICL3 would open up. However, with these results, ICL3 travels around the same locations like the one in *c3_interm_open*, and it does not open. In Figure 4.40a, the movement is separated into six different clusters for a better understanding of the time ICL3 changed its positions. ICL3 starts from the same intermediate position as in *c3_interm_open*, and after going through a slight closure, it finds another half-closed conformational state (shown with dark blue cluster) instead of going to a full open position as expected. Figures 4.40b and 4.40c support this observation as well.

4.4.1. RMSD Profiles

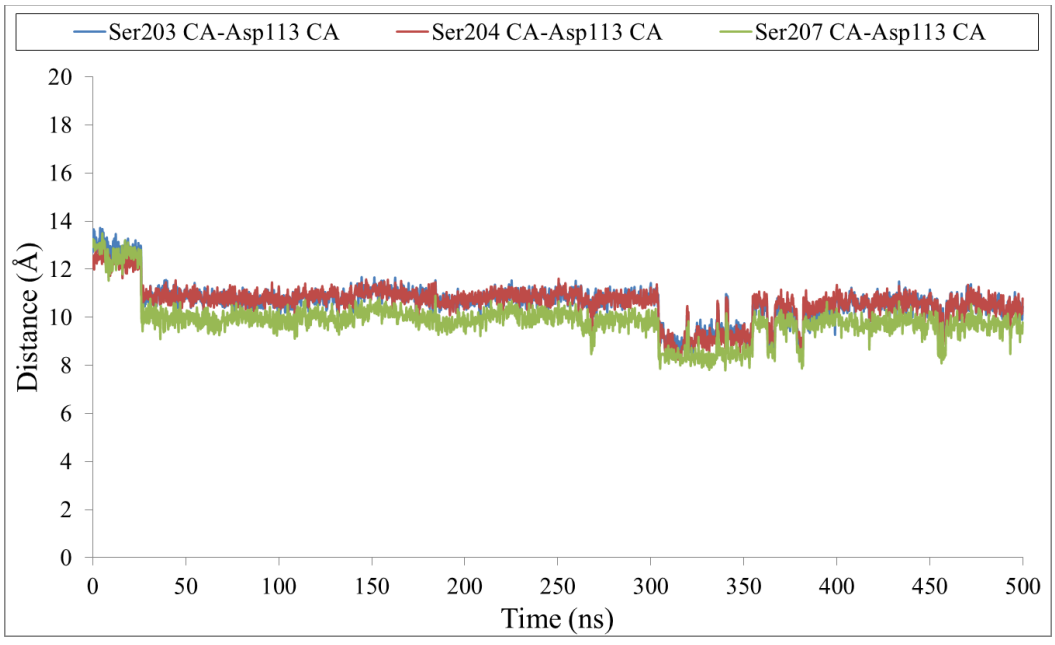
c3_interm_open's *all_fit_core RMSD* in Figure 4.41a mainly changes between 2-7 Å. There is a certain increase until 96 ns during, which ICL3 shows a closing movement as seen in the center of mass of ICL3 profile. Later, the RMSD stays almost constant until 336 ns, after which high fluctuations are observed in the RMSD value. The *core_fit_core* and *transmembrane_fit_transmembrane RMSD* values are almost constant and around 2 Å and 2.5 Å, respectively.

The RMSD of lower part of H6 for *c3_interm_open* in Figure 4.41b moves from 7 to 10 Å, while the RMSD with respect to inactive structure moves from 2 Å to 4 Å. The increases are seen around 39 ns, which is an early time in the trajectory. The initial

snapshot in Figure 4.41c shows a more distinct profile from the intermediate and final snapshots, supporting the RMSD increases.

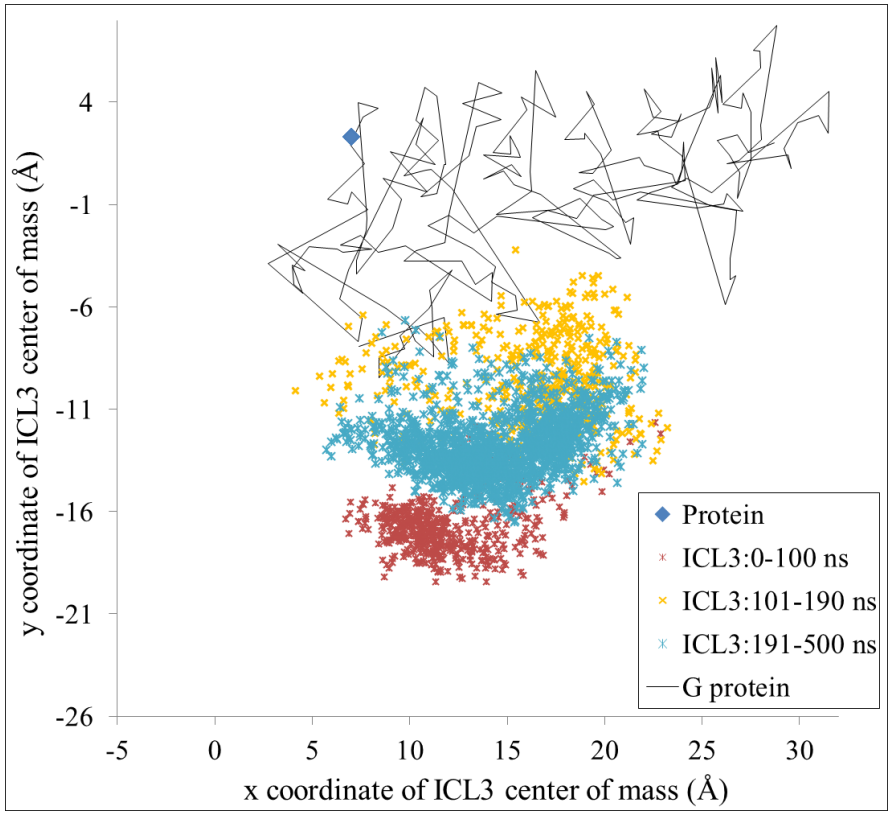


(a)

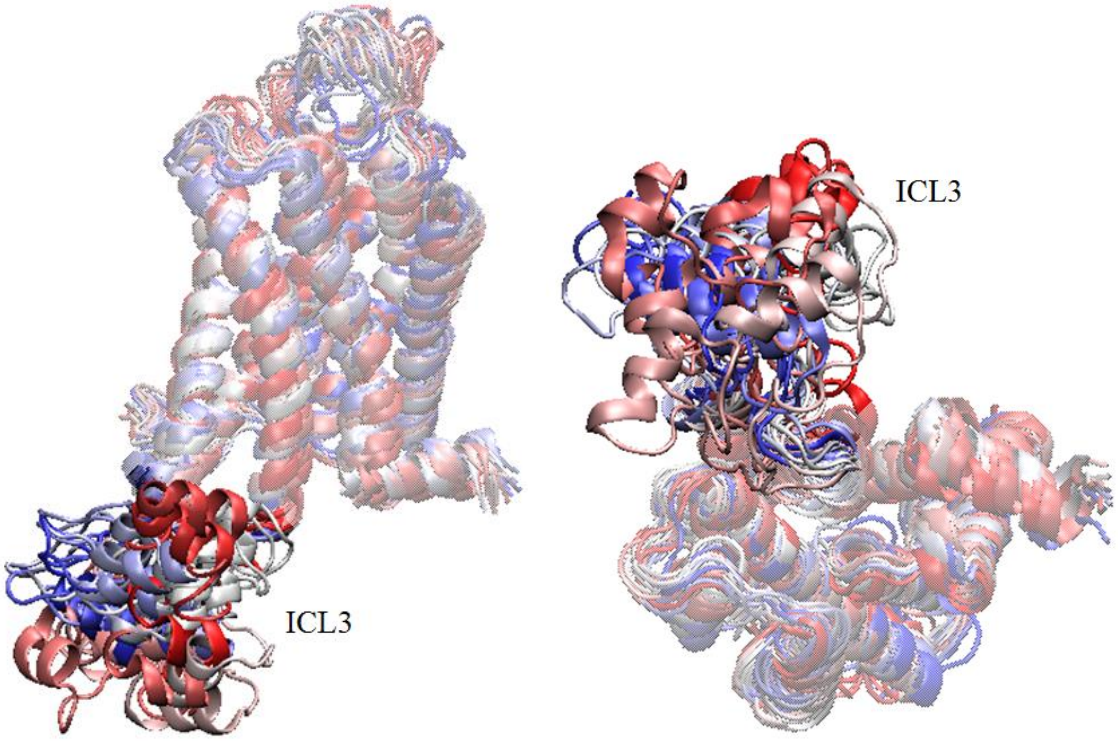


(b)

Figure 4.38. Distance profiles of *c4_interm_closed* for (a) seven different distance constraints and for (b) C^{α}'s (CA) of three different unconstrained distances.

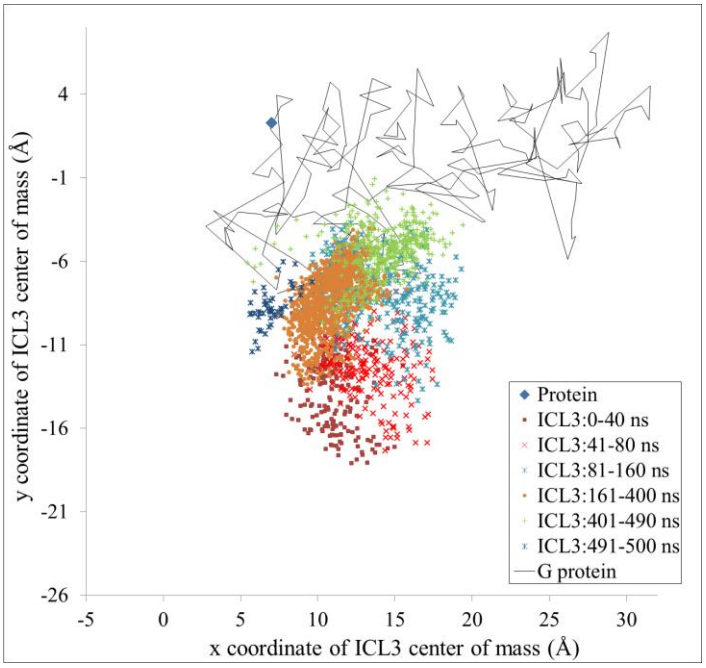


(a)

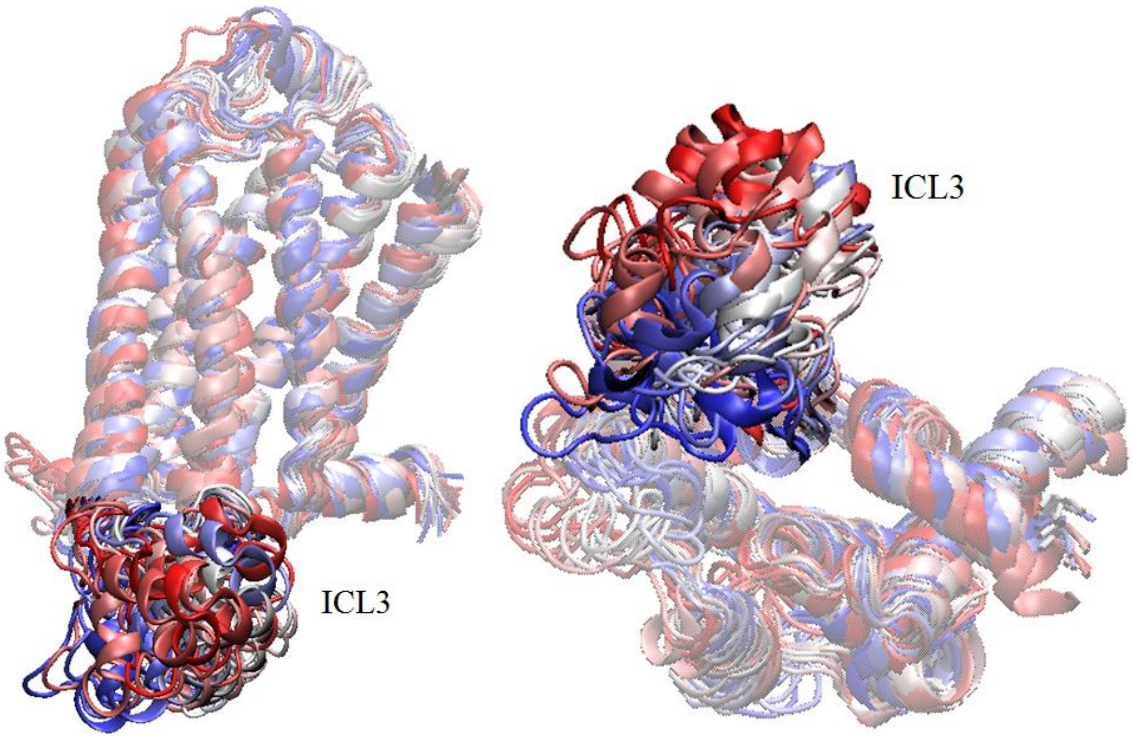


(b) (c)

Figure 4.39. *c3_interm_open*'s (a) variations in ICL3's center of mass profile and snapshots from (b) side and (c) lower views.



(a)



(b) (c)

Figure 4.40. *c4_interm_closed*'s (a) variations in ICL3's center of mass profile and snapshots from (b) side and (c) lower views.

Figure 4.42a shows the RMSD of lower part of H5 with respect to active structure for *c3_interm_open*. The value changes between 2-4.5 Å throughout the trajectory and it finishes around 4 Å. This profile can be interpreted as quite stable as the snapshots prove, too, in Figure 4.42b.

c4_interm_closed's *all_fit_core RMSD* value in Figure 4.43a has more fluctuations compared to *c3_interm_open* in the second half of the trajectory. The first halves of the models resemble very much with values between 2-6 Å. Differently, *c4_interm_closed* has a slight increase between 231-264 ns, a decrease at 374 ns. After 396 ns, it has a big increase from nearly 3.5 Å to 7 Å.

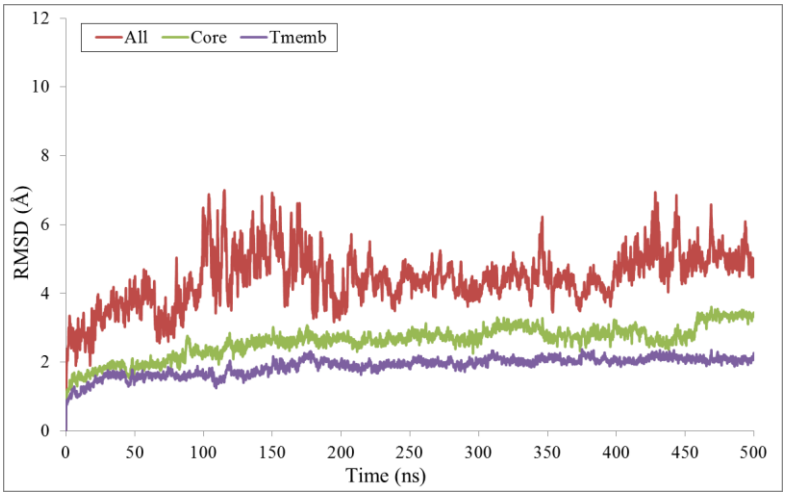
In 4.43b, the RMSD of lower part of H6 of *c4_interm_closed* with respect to active structure fluctuates between 6-9 Å while the RMSD with respect to inactive structure of *c4_interm_closed* changes between 1-3 Å. After 130 ns, the RMSD stays almost constant. The snapshots of the lower part of H6 in Figure 4.43c support Figure 4.43b with the different initial snapshot from the other ones, since the RMSD of lower part of H6 was fluctuating more during the initial 130 ns.

The RMSD of lower part of H5 with respect to active structure for *c4_interm_closed* in Figure 4.44a fluctuates between 1-4 Å and it is more constant than the other regions between 269-349 ns. The snapshots in Figure 4.44b seem relatively similar with each other.

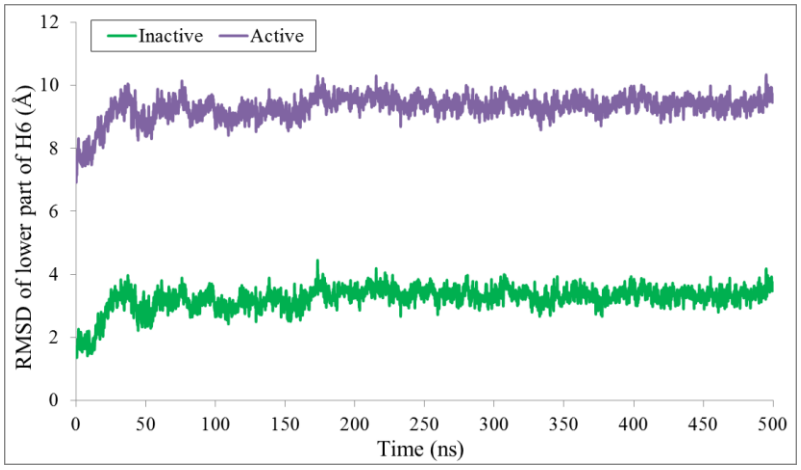
4.4.2. Conformational Dynamics and PCA

The profile for cumulative % explanation of the first principal components for *3_interm_open* and *c4_interm_closed* compared with *original* is given in Appendix A Figure A.2. The snapshots in Figure 4.45 are more clear compared to the first principal mode vectors of *c3_interm_open* in Appendix A Figure A.3. H5 on the upper region moves inside although an open constraint of 17 Å applied between Ser203-Asp113 was expected to cause H5 to move outside. On the other hand, the upper region of H3 moves outside, keeping the constraint distance almost constant at 17 Å throughout the simulation. This model does no work because the constraints applied to the intermediate position do not

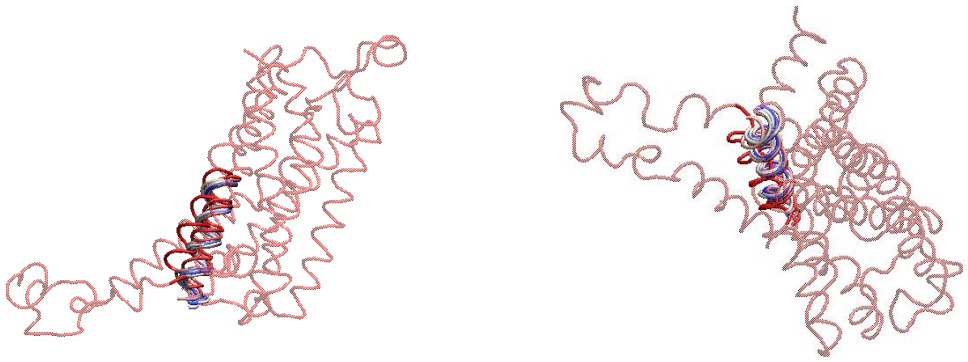
cause the expected displacement of the extracellular part of H5. As a result, ICL3 does not adopt to the closed position.



(a)

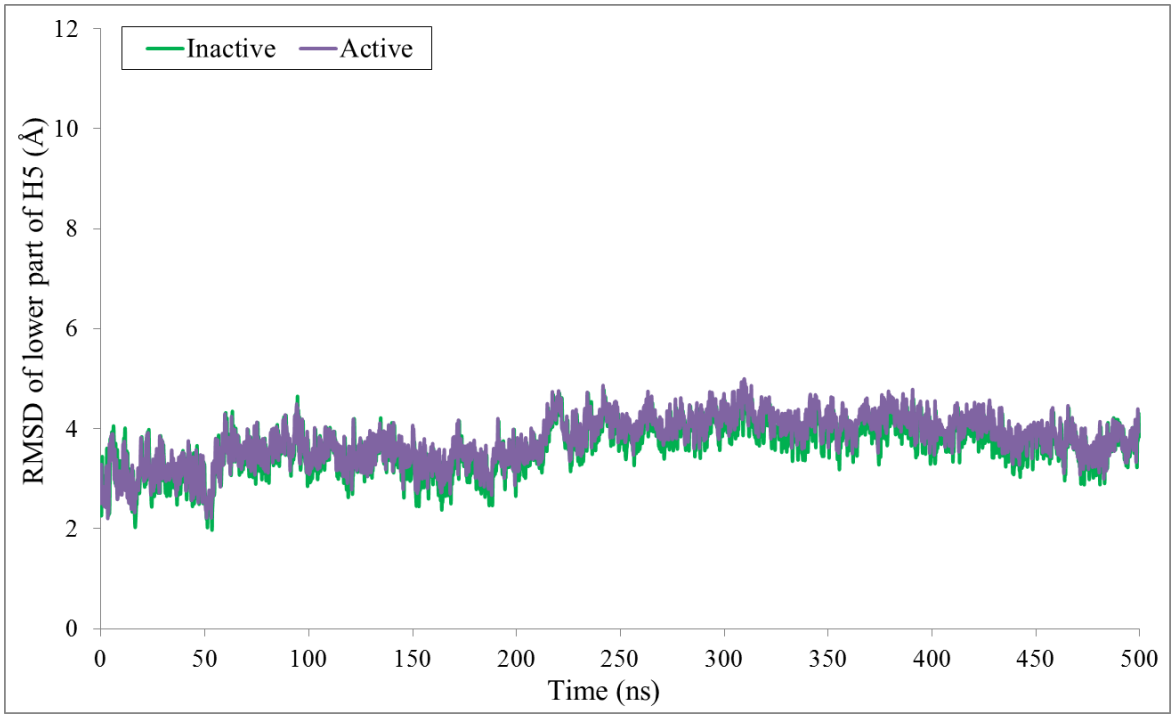


(b)

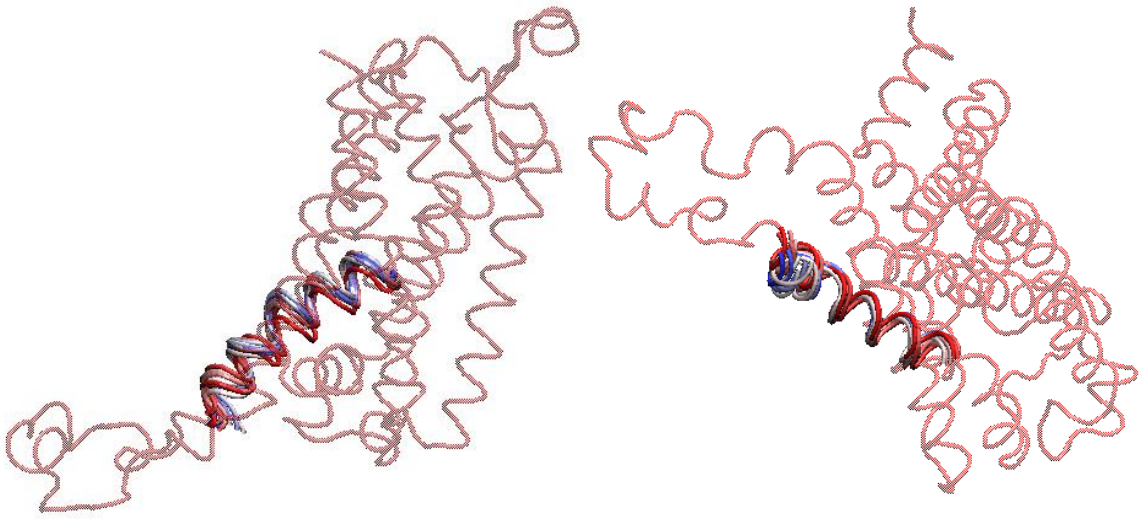


(c) (d)

Figure 4.41. *c3_interm_open*'s (a) RMSD profile, (b) RMSD of lower part of H6 profile and ten aligned snapshots from (c) side and (d) lower views.

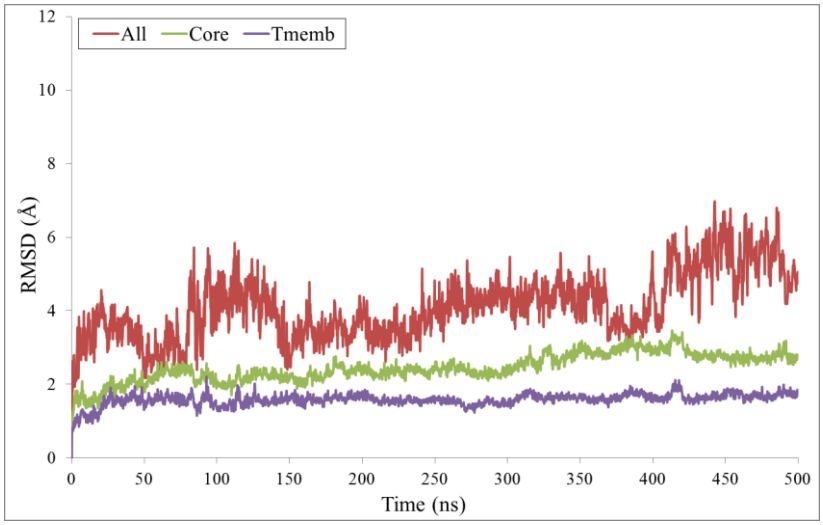


(a)

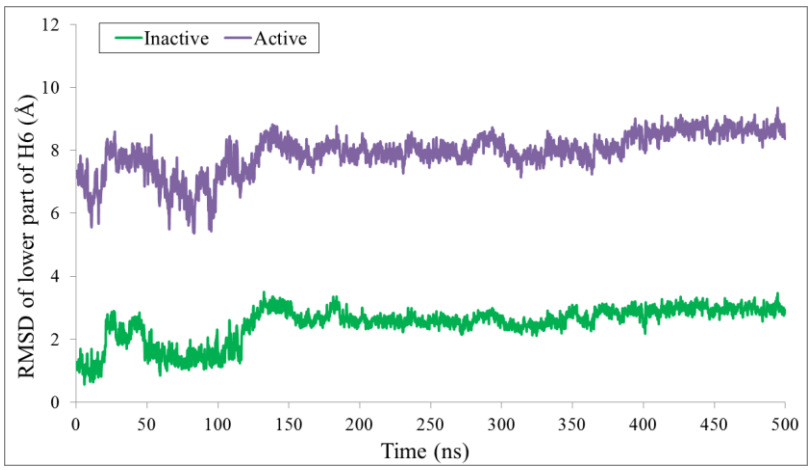


(b) (c)

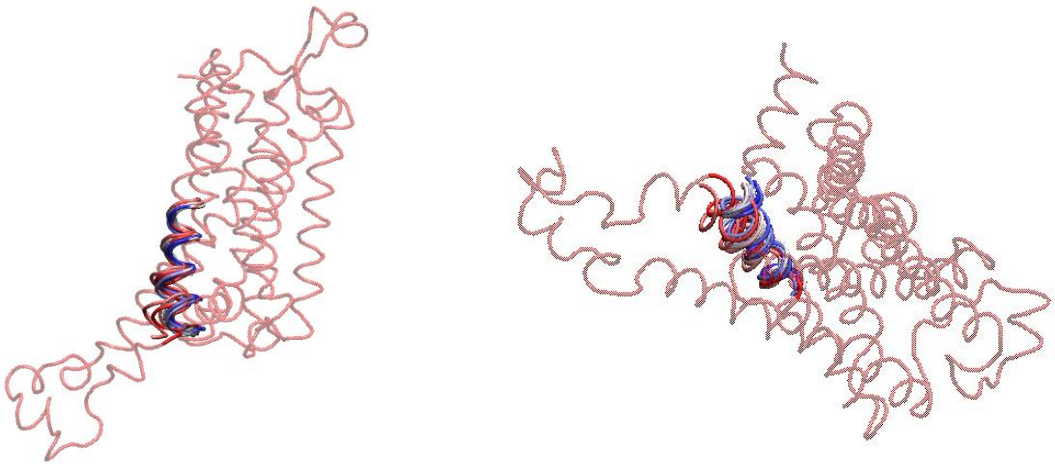
Figure 4.42. *c3_interm_open*'s (a) RMSD of lower part of H5 and ten aligned snapshots from (b) side and (c) lower views.



(a)

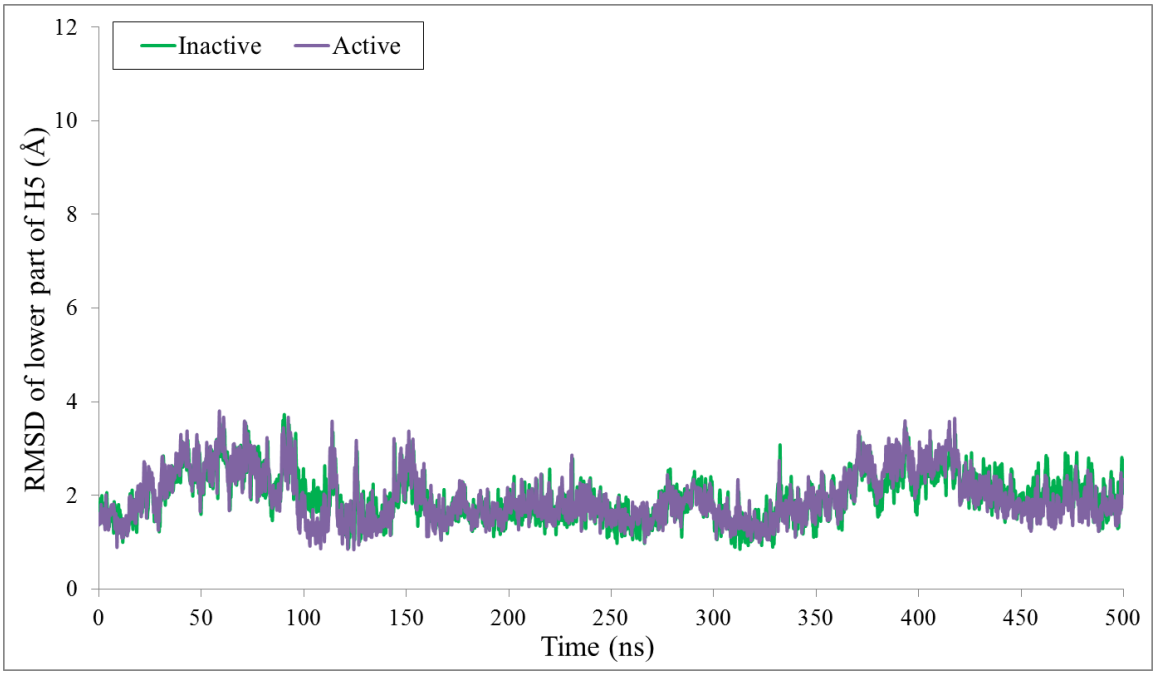


(b)

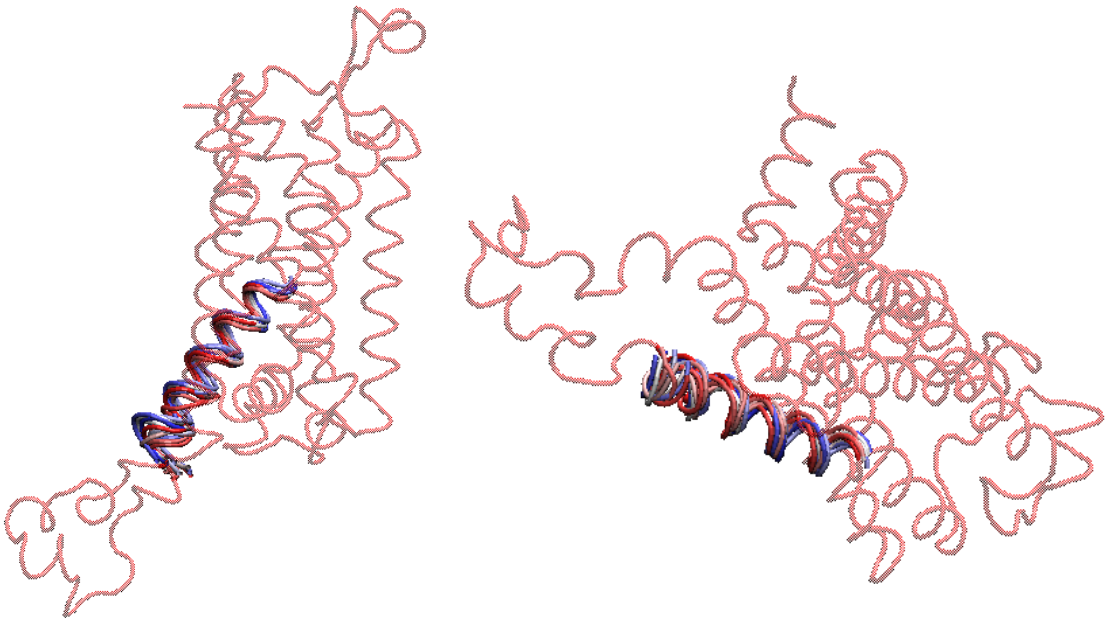


(c) (d)

Figure 4.43. *c4_interm_closed*'s (a) RMSD profile, (b) RMSD of lower part of H6 profile and ten aligned snapshots from (c) side and (d) lower views.

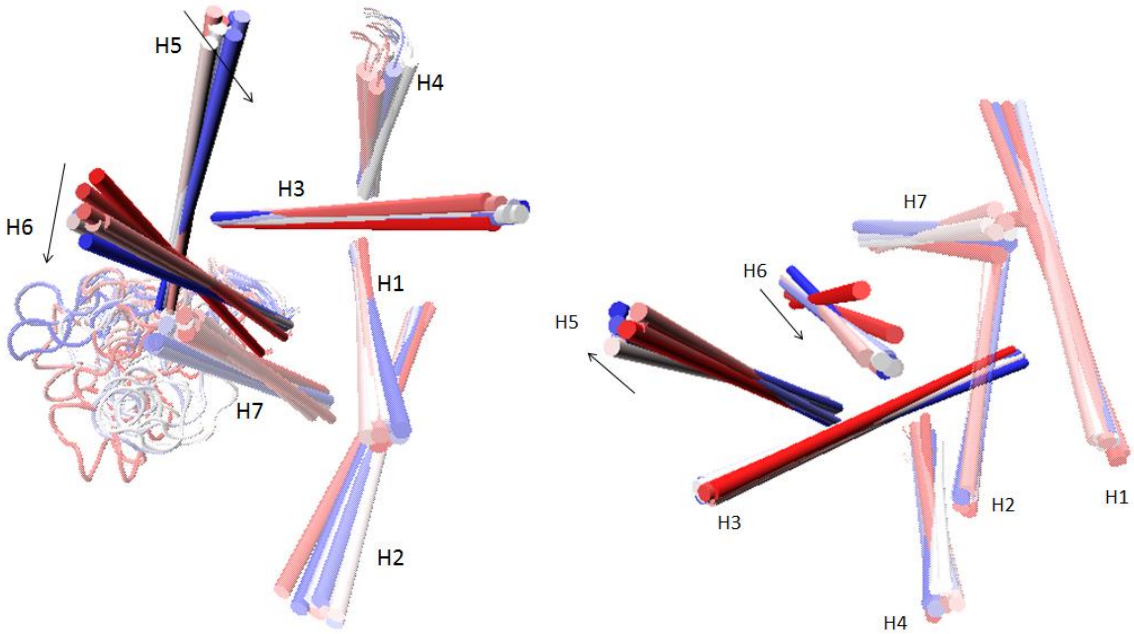


(a)



(b) (c)

Figure 4.44. *c4_interm_closed*'s (a) RMSD of lower part of H5 and ten aligned snapshots from (b) side and (c) lower views.



(a) (b)

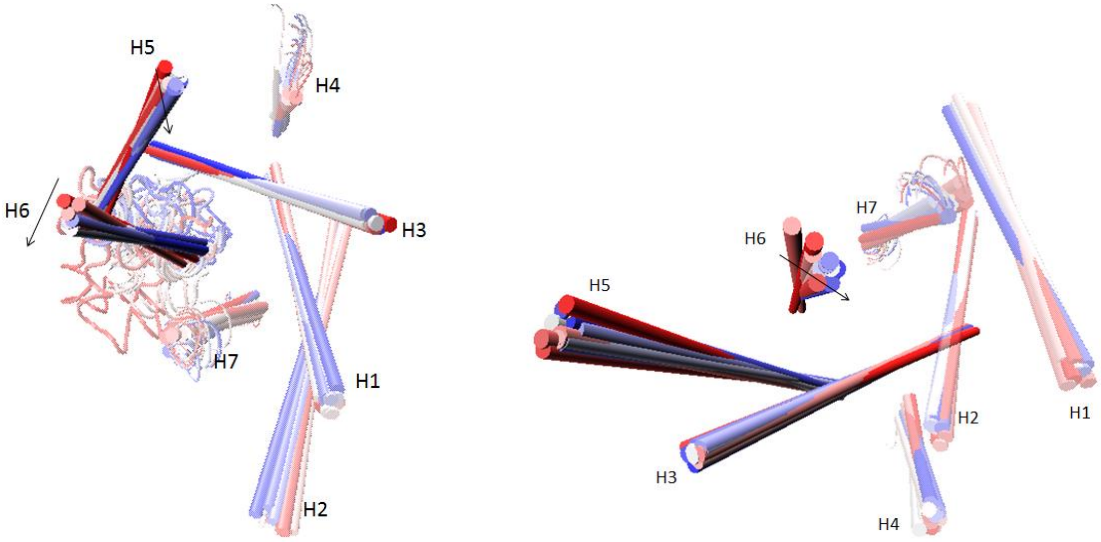
Figure 4.45. Snapshots of *c3_interm_open* with detailed H3, H5 and H6 from (a) top and (b) bottom views.

Figure 4.46 shows the snapshots of *c4_interm_closed* with the upper region of H5 moving inside and the upper region of H6 moving outside. The lower region of H5 does not have a specific movement while that of H6 moves inwards. Here, the expected closure of ICL3 is not observed, although the constraints applied are successful to keep the extracellular part of H5 closer to the binding site. Apparently, the open state of ICL3 is energetically difficult to be reached from a closed position and the allosteric coupling that exists between the intra- and the extracellular parts of the receptor is not enough to cause the desired opening in ICL3. The first principal mode vectors of *c4_interm_closed* are given in Appendix A Figure A.4.

4.4.3. RMSF Profiles

When the RMSF values of the intermediate and initial constrained models are compared in Figure 4.47, it can be concluded that the ICL3 regions' RMSF values for *c3_interm_open* and *c4_interm_closed* are nearer to *c1_initial_open* instead of

c2_initial_closed whose ICL3 moves very slightly. *c3_interm_open* has a higher RMSF value than *c4_interm_closed* for the ICL3 region.



(a) (b)

Figure 4.46. Snapshots of *c4_interm_closed* with detailed H3, H5 and H6 from (a) top and (b) bottom views.

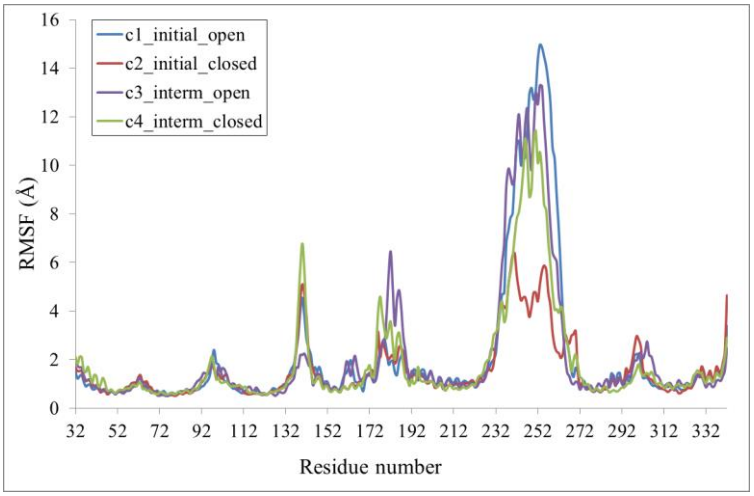


Figure 4.47. RMSF profiles of the constrained trajectories.

4.5. Additional Constraint Applied to β_2AR with a Half-closed ICL3

The two models studied in this section starts from the same intermediate position of ICL3 as in the previous section. The difference from the previous constraint models is an

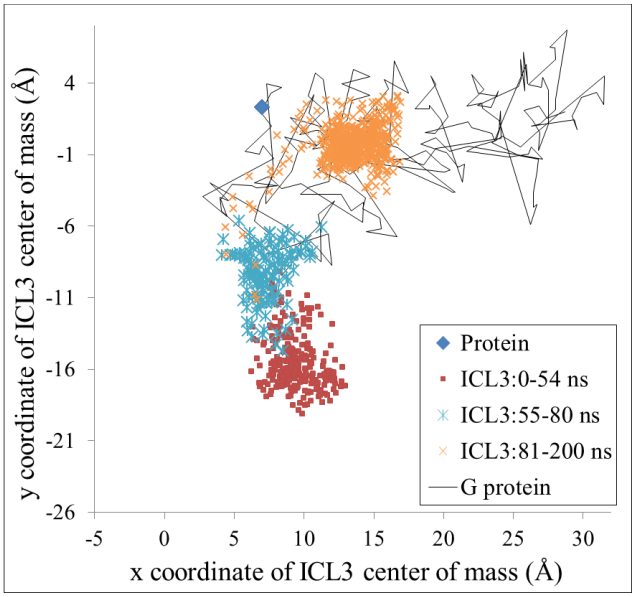
additional constraint between the C^{α} atoms of Ser207 and Asp113 at 17 Å which is observed in Figure 4.9 upon ICL3 closure. The aim is to strengthen the constraints in the ligand binding site and therefore force ICL3 to act as wanted. Also, *c6_20ns_interm_closed* which is the closed model of this section has more narrowly constrained distances between the Ser residues and Asp113 because of the failure in the less constrained intermediate model *c4_interm_closed*. The outcome of this new constrained MD is satisfactory leading to expected open/closed states in ICL3 within 200 ns, which is a relatively shorter time frame. These new MD simulations starting with the intermediate structures are simply called as *c5_200ns_interm_open* and *c6_200ns_interm_closed*.

The center of mass for ICL3 profile of *c5_200ns_interm_open* in Figure 4.48a shows a similar profile to that of *original*. ICL3 center of mass for *c6_200ns_interm_closed* is generally on the open side of ICL3 as expected from a receptor with a closed ligand binding site. As a result, the first half of *c5_200ns_interm_open* is at nearly the same location with *c6_200ns_interm_closed*. Thus, it can be concluded that the constraints applied on the backbone C^{α} atoms work for the intermediate models.

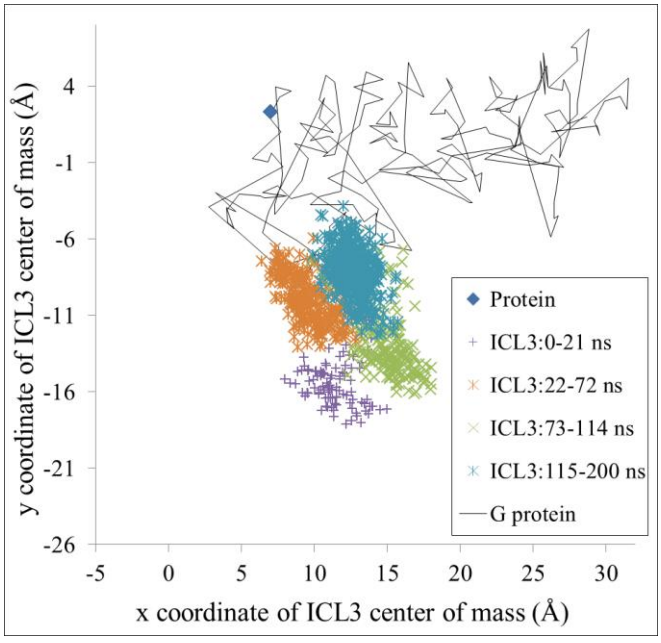
4.5.1. Conformational Dynamics and PCA

Figure 4.49 shows the cumulative % explanation of *c5_200ns_interm_open* and *c6_200ns_interm_closed* compared to that of *original*. The highest result belongs to *c5_200ns_interm_open*. In comparison, *c6_200ns_interm_closed*'s first PCA mode does not explain the movement of the trajectory well.

PCA mode vector profiles are also investigated for the last two 200 ns constrained models. *c5_200ns_interm_open* in Figure 4.50 is a specific model differing from *c3_interm_open* with only a C^{α} backbone distance constraint of 17 Å. It lasts for 200 ns because of the clear behavior of ICL3 which closes immediately and stays there for the whole trajectory because of this specific backbone C^{α} distance constraint.



(a)



(b)

Figure 4.48. Variations in ICL3's center of mass for (a) *c5_200ns_interm_open* and (b) *c6_200ns_interm_closed*.

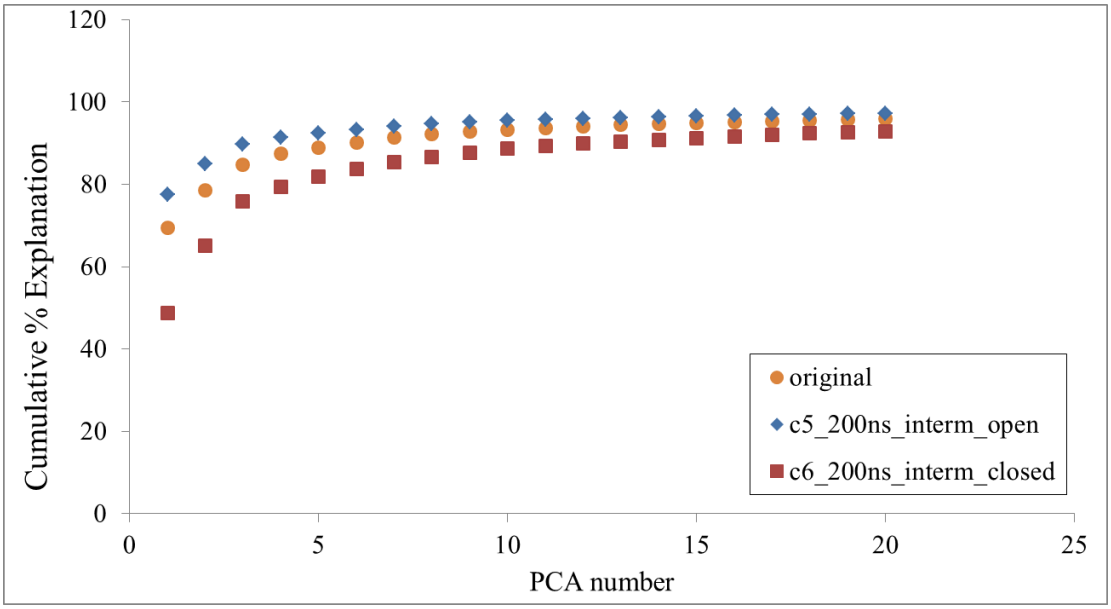


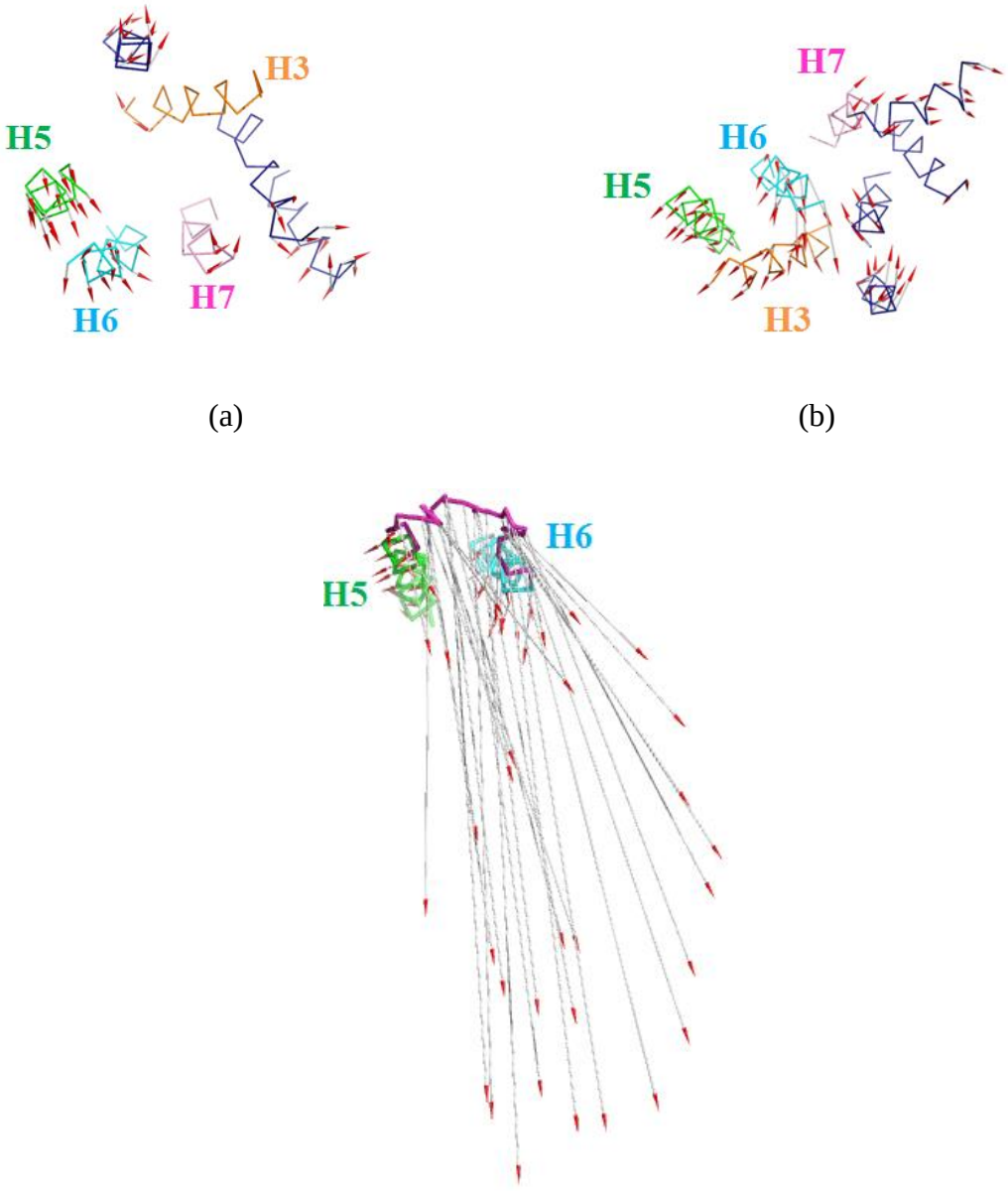
Figure 4.49. Cumulative % explanation of principal components for *c5_200ns_interm_open* and *c6_200ns_interm_closed* compared with *original*.

The movement of ICL3 is supported by mainly the lower part of H6 and partially the lower part of H5 according to the first principal mode vectors. The upper regions of H5 and H6 are expected to move outside because of giving an anticorrelated movement with the lower parts due to the pulling of ICL3.

In order to compare *original* with *c5_200ns_interm_open*, their overlap matrix is drawn as in Figure 4.51. It is observed that the first PCA mode of *original* has a high overlap score of 0.7 with the second PCA mode of *c5_200ns_interm_open*. Thus, the second principal mode vectors are drawn and shown in Appendix A Figure A.5. According to these figures, the directions of ICL3 mode vectors are more similar to those of *original* because the closures of ICL3's are more towards the side of the receptor. However, the second principal mode vectors of *c5_200ns_interm_open*'s ICL3 are more towards the inside of the receptor.

c6_200ns_interm_closed has one additional distance constraint of backbone C^{α} atom to the distance constraints in *c4_interm_closed*. The first three Ser distance couples are more narrowly set than those of *c4_interm_closed*. In Figure 4.52, the outwards movement of ICL3 is seen clearly. It stays there for the whole trajectory. Both of the lower parts of

H5 and H6 move outside pulled by ICL3. As expected, the upper region of H6 moves inside anticorrelated to the lower part of it.



(a) (b)

(c)

Figure 4.50. First principal mode vectors of *c5_200ns_interm_open* with (a) upper view, (b) lower view, (c) lower view with H5, H6 and ICL3.

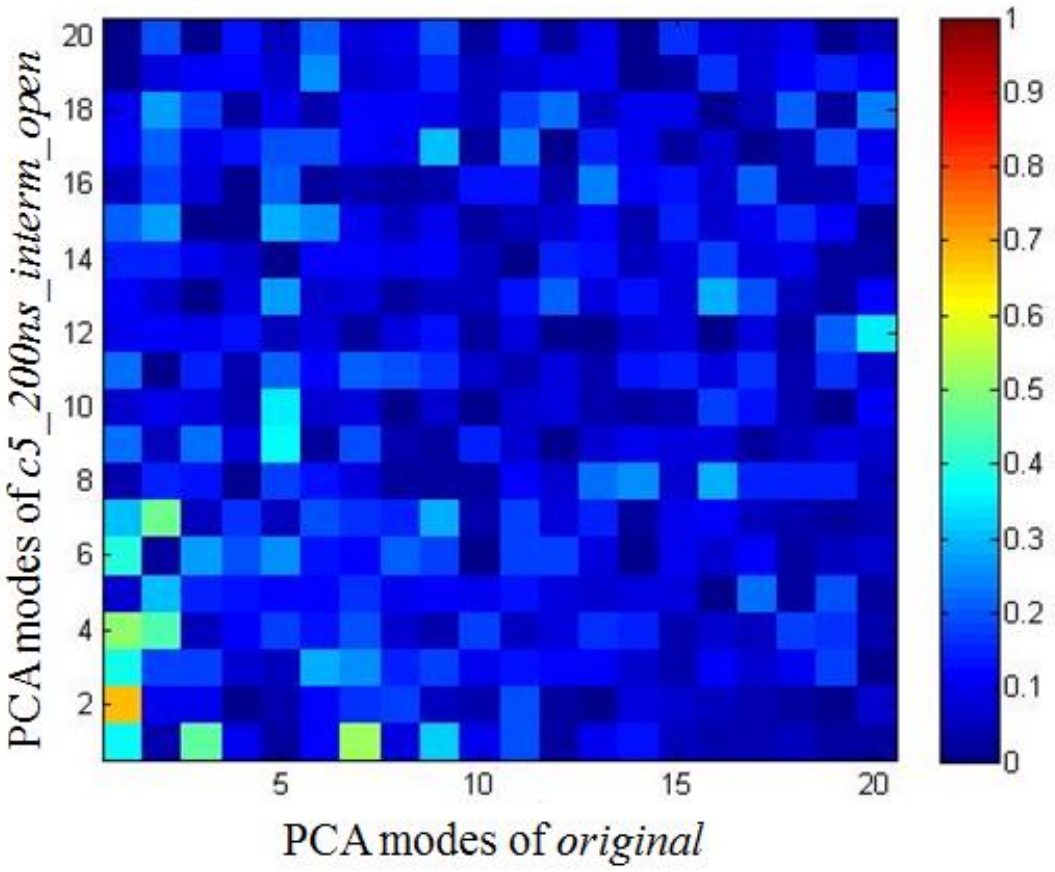


Figure 4.51. Overlap between *original* and *c5_200ns_interm_open*.

The snapshots of *c5_200ns_interm_open* in Figure 4.53 shows the movement of ICL3 very clearly just like *c1_initial_open*. The effect of the backbone C^{α} atom constraint shows its results very distinctly. The middle and the last snapshots are very similar compared to the beginning snapshot, thus a huge portion of the simulation has a closing ICL3. The overlap results in Figure 4.51 for *original* and *c5_200ns_interm_open* also shows the similarity between the first PCA mode of *original* and the second PCA mode of *c5_200ns_interm_open* with a high score. Before, the overlap and the similarity between the first PCA modes of *original* and *c1_initial_open* were shown, too. Therefore, *original*, *c1_initial_open* and *c5_200ns_interm_open* can be evaluated in the same category of closing ICL3.

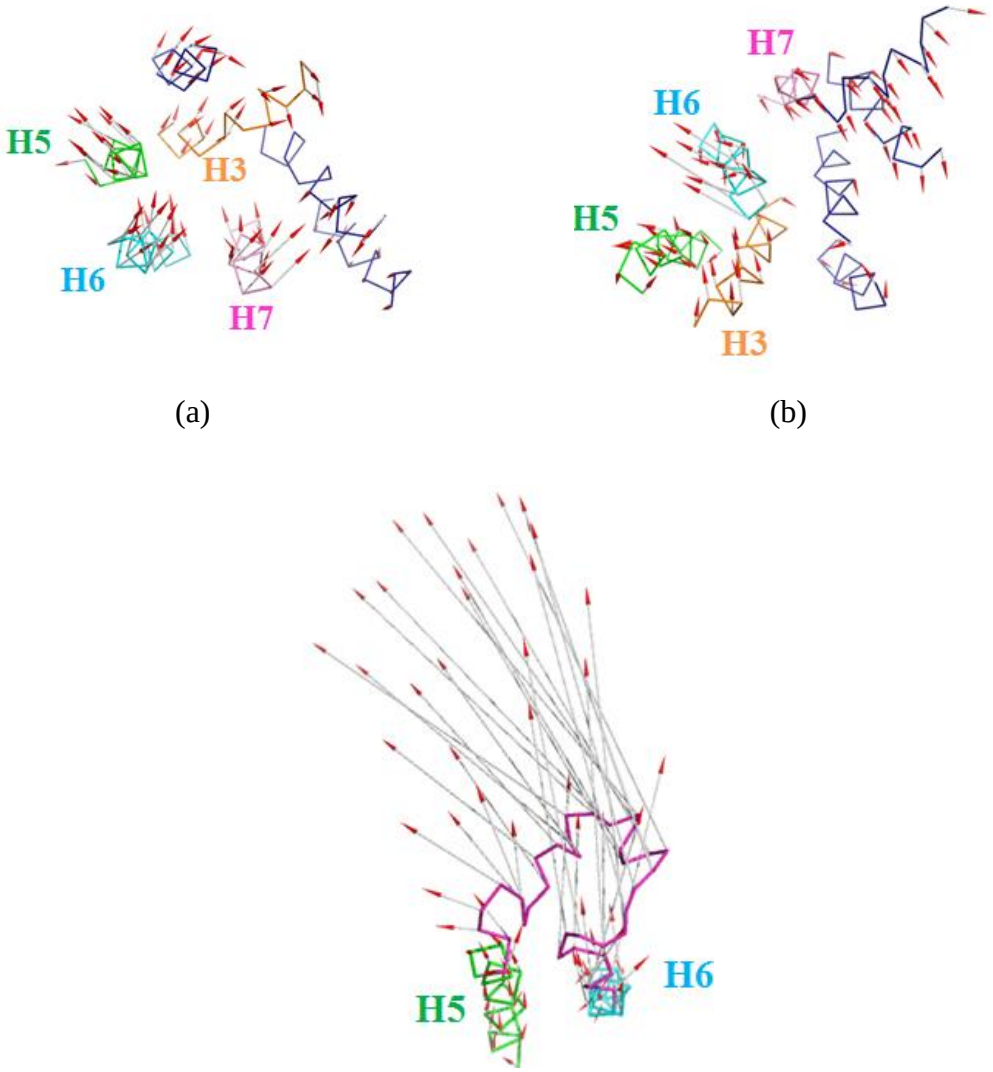
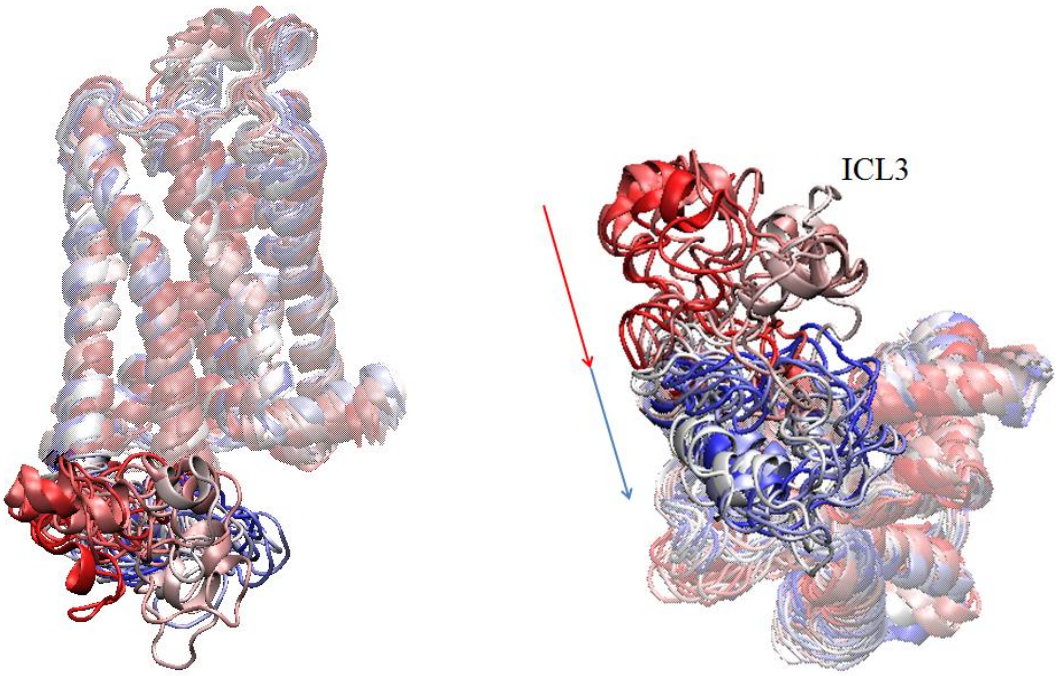


Figure 4.52. First principal mode vectors of *c6_200ns_interm_closed* with (a) upper view, (b) lower view, (c) lower view with H5, H6 and ICL3.

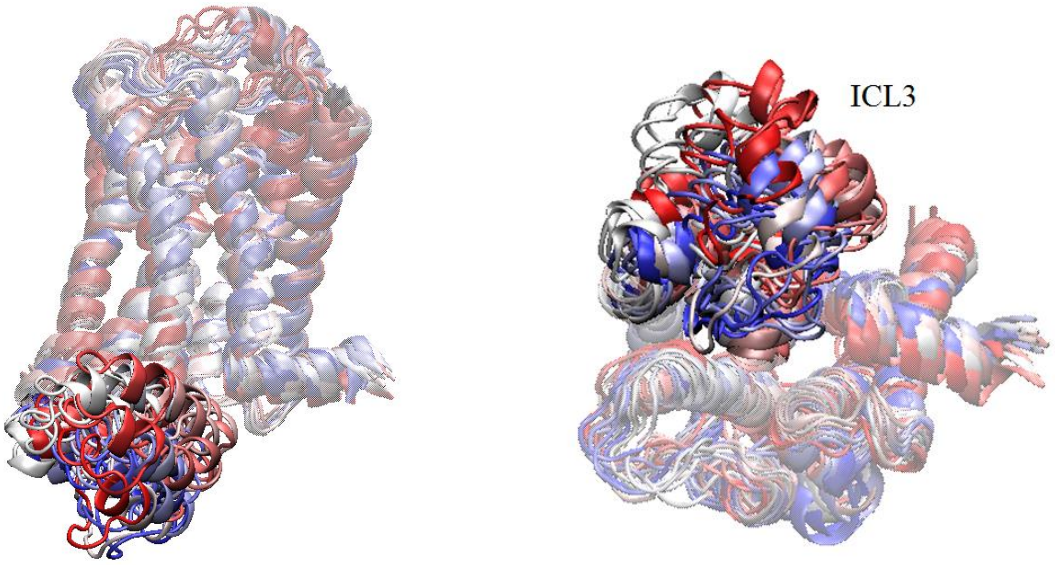
c6_200ns_interm_closed in Figure 4.54 also shows a very distinct open ICL3 with the last snapshot slightly trying to close it. ICL3 is expected to stay open throughout the trajectory due to the C^{α} backbone atom constraint. But in general, it can be considered as open.

For *c5_200ns_interm_open*, the outer movements of the upper regions of H5 and H6 as well as the inner movements of them on the lower region are very well seen in Figure 4.55.



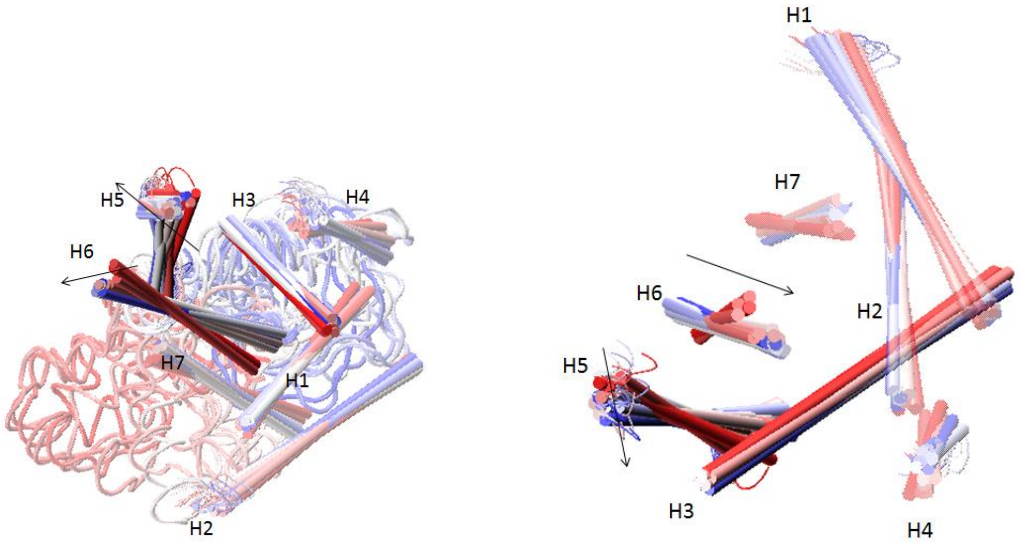
(a) (b)

Figure 4.53. Snapshots from (a) side and (b) lower views for *c5_200ns_interm_open*.



(a) (b)

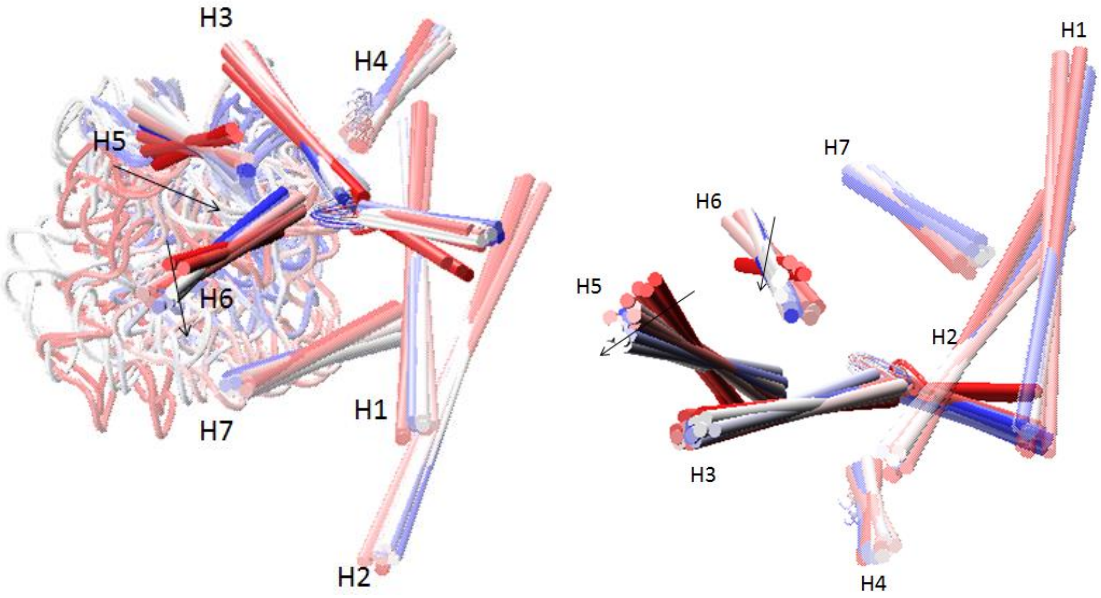
Figure 4.54. Snapshots from (a) side and (b) lower views for *c6_200ns_interm_closed*.



(a) (b)

Figure 4.55. Snapshots of *c5_200ns_interm_open* with detailed H3, H5 and H6 from (a) top and (b) bottom views.

The inner movement of H5 and H6 on the upper part and the outer movements of them on the lower part are also very clear in Figure 4.56 for *c6_200ns_interm_closed*.



(a) (b)

Figure 4.56. Snapshots of *c6_200ns_interm_closed* with detailed H3, H5 and H6 from (a) top and (b) bottom views.

4.5.2. Distance Profiles

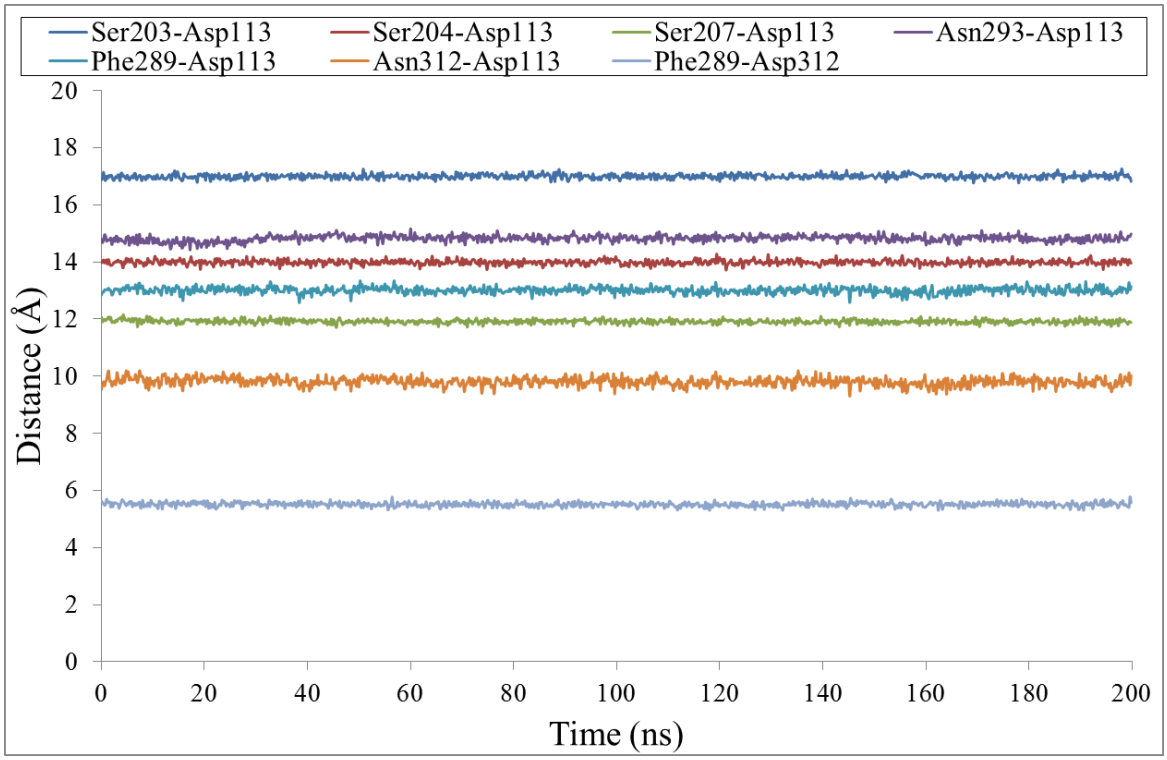
The distance profiles of *c5_200ns_interm_open* in Figure 4.57a show that all the distance values are the same with those of the inactive structure except Ser203-Asp113 and the C^{α} distance of Ser207-Asp113 which are both 17 Å. The distant constraints of *c1_initial_open* are also the same except that it does not include the C^{α} distance. The distances are constant throughout the trajectory. The Ser distance couples show the same set constraints. The other distances show slightly higher values than the set values. Phe289-Asp312 distance shows the set distance as in many other constrained runs. In Figure 4.57b, the Ser207-Asp113 distance is very slightly lower than the set value.

The distant profile of *c6_200ns_interm_closed* is more variable than *c5_200ns_interm_open*. All the distances are set to values much lower than those of the inactive structure. In Figure 4.58a and Figure 4.58b, the Ser couples including the C^{α} atom provide the set value of constraint distances. The other constraint distances are slightly higher than the set values.

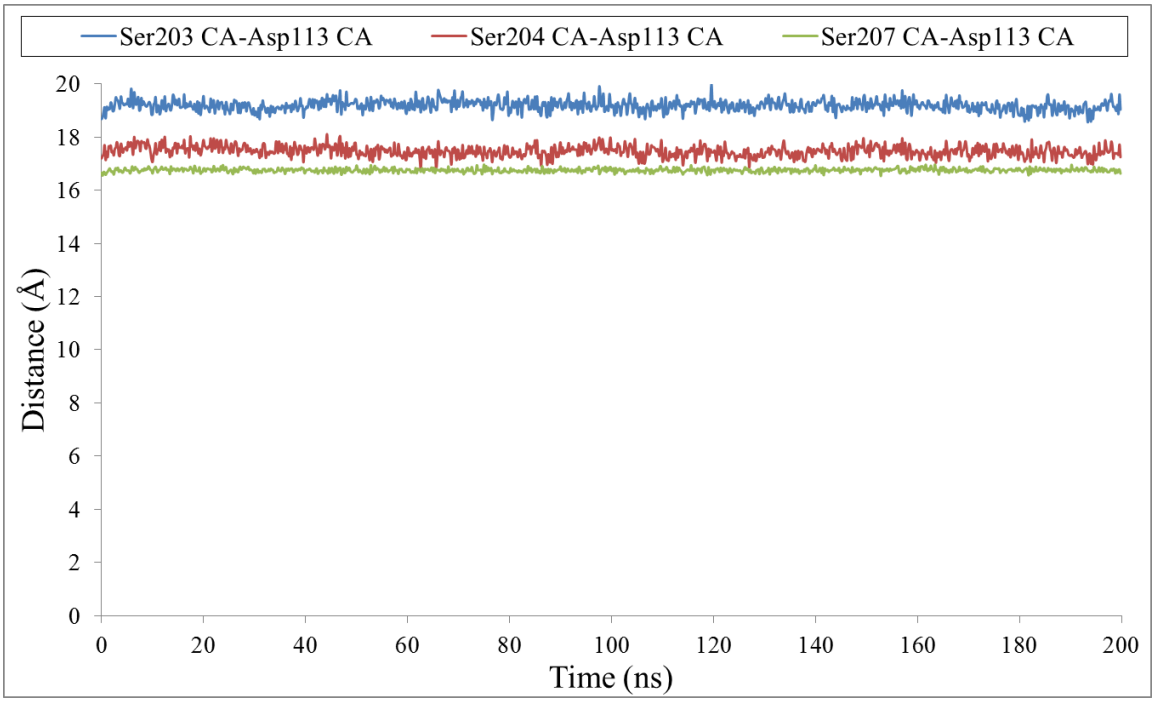
4.5.3. RMSD Profiles

The RMSD results of 200 ns long constrained models give expected results. *c5_200ns_interm_open* in Figure 4.59a has a sharp increase in *all_fit_core RMSD* from 4 to 10 Å around 80 ns among all the models since its constraint is applied on the backbone C^{α}, forcing a faster and sharper result. The applied constraint forces ICL3 to close and stay in a stable position under the receptor. *c1_initial_open's* and *original*'s *all_fit_core RMSD* trends are very similar to *c5_200ns_interm_open* since all show the closure of ICL3 very clearly.

c5_200ns_interm_open's RMSD of lower part of H6 with respect to active structure in Figure 4.59b moves from 8 to 10 Å starting around 23 ns while the RMSD with respect to inactive structure moves from 2 to 4 Å. Its trends are very similar to those of *original.* Thus it can be concluded that *c5_200ns_interm_open* shows an expected profile for the closure of ICL3. The discrete snapshots in Figure 4.59c reinforce the results.

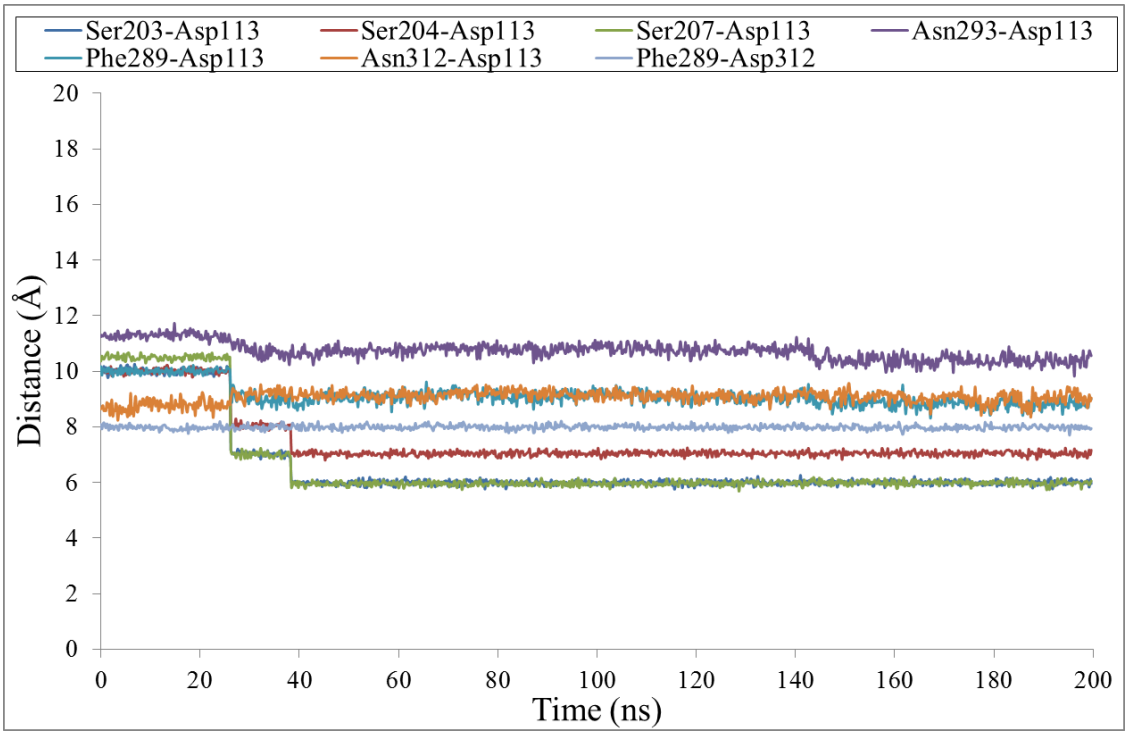


(a)

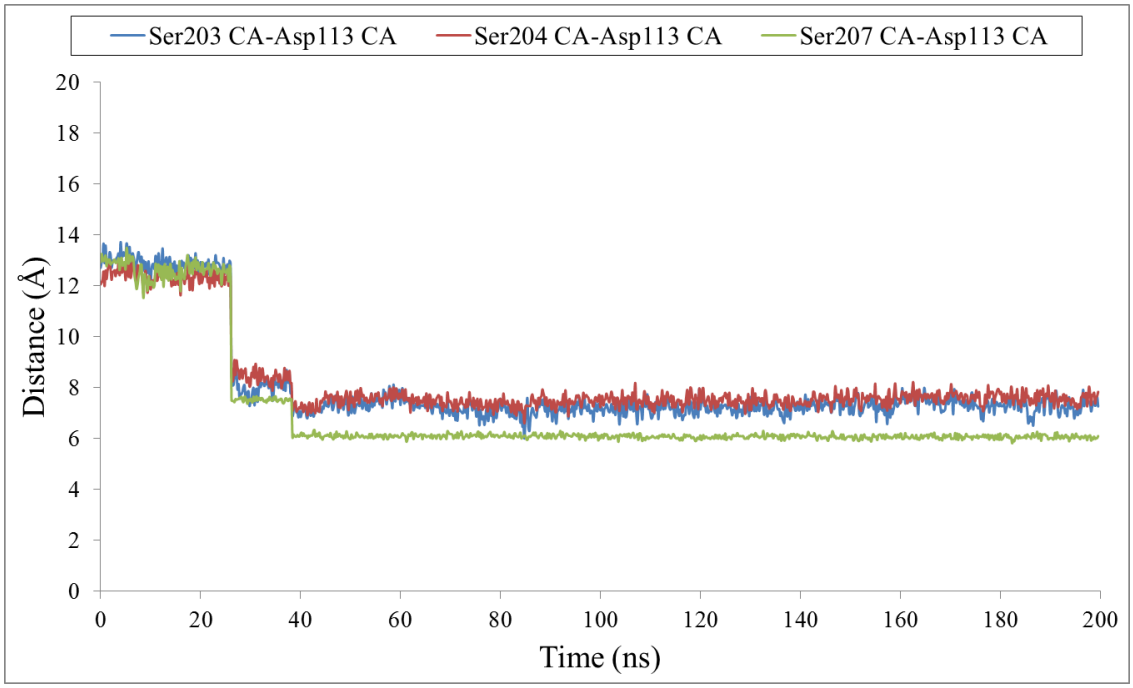


(b)

Figure 4.57. Distance profiles of *c5_200ns_interm_open* for (a) seven different distance constraints and for (b) C^{α}'s (CA) of three different unconstrained distances.



(a)



(b)

Figure 4.58. Distance profiles of *c6_200ns_interm_closed* for (a) seven different distance constraints and for (b) C^{α}'s (CA) of three different unconstrained distances.

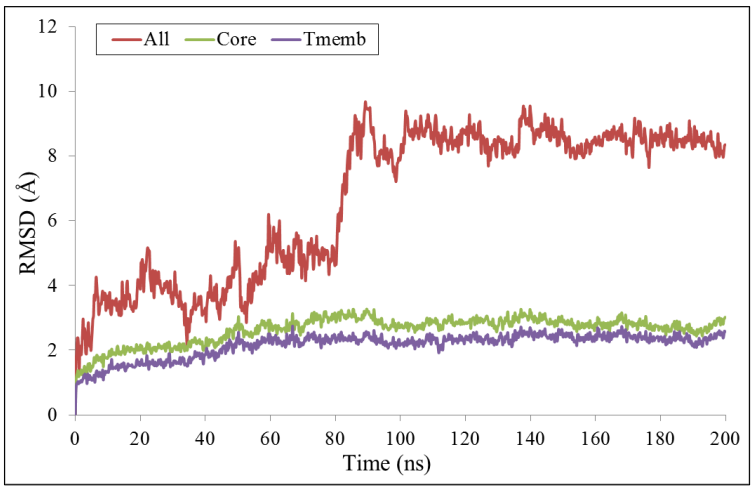
The RMSD of lower part of H5 with respect to active and inactive structures in Figure 4.60a fluctuates between 2-4 Å and mainly is around 3 Å with a fairly stable profile. The snapshots in Figure 4.60b sustain this stability in the lower part of H5.

c6_200ns_interm_closed in Figure 4.61a has very stable and similar RMSD values like *orig_2nd half* and *c2_initial_closed*. It is an expected result that since ICL3 does not move in *c6_200ns_interm_closed*, its RMSD values are also stable. The *all_fit_core RMSD* starts around 2 Å and moves to 4 Å around 22 ns after which it starts to fluctuate very slightly around 4 Å. The *core_fit_core* and *transmembrane_fit_transmembrane RMSD* values seem to stabilize around 2.5 Å. In Figure 4.61b, *c6_200ns_interm_closed'*s RMSD with respect to active structure fluctuates between 6-9 Å while the RMSD with respect to inactive structure fluctuates between 1-3 Å. This model has the most fluctuating RMSD profiles among all the models until now. The outcome of this fluctuation can be seen in the snapshots in Figure 4.61c where the initial structure is separate from the other snapshots.

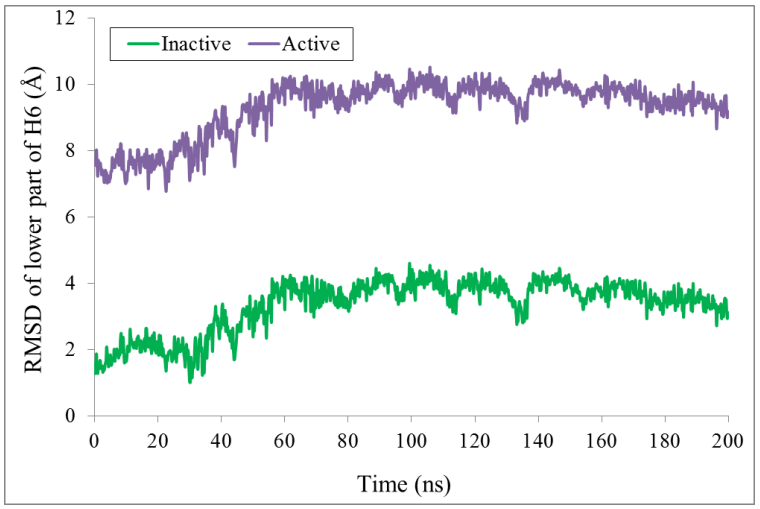
Figure 4.62a shows the RMSD of lower part of H5 with respect to active and inactive structures for *c6_200ns_interm_closed.* The value changes between 1.5-4 Å and does not show a specific difference in the snapshot profile of H5in Figure 4.62b.

4.5.4. RMSF Profiles

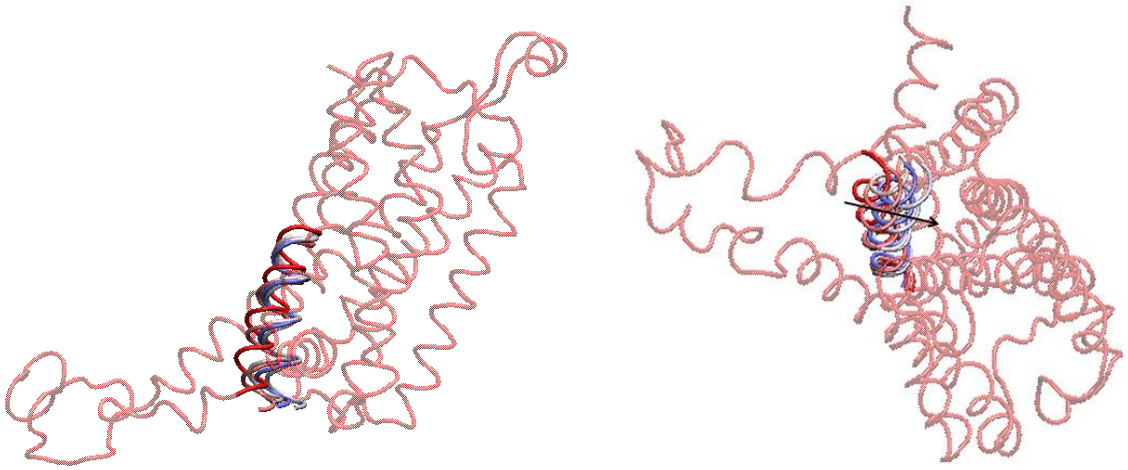
RMSF of ICL3 region in Figure 4.63 shows that the highest RMSF belongs to *c5_200ns_interm_open* whose ICL3 closes immediately after the run starts, meaning that a constraint applied to the backbone C^{α} atoms affects the movement of the ICL3 very strongly. The lowest RMSF in the ICL3 region belongs to *c2_initial_closed*, which is followed by *c6_200ns_interm_closed. c1_initial_open* has the highest values after *c5_200ns_interm_open.* This RMSF profile gives very satisfactory outcome since it explains the effect of the constraints applied on the C^{α} backbone chains of the receptors.



(a)

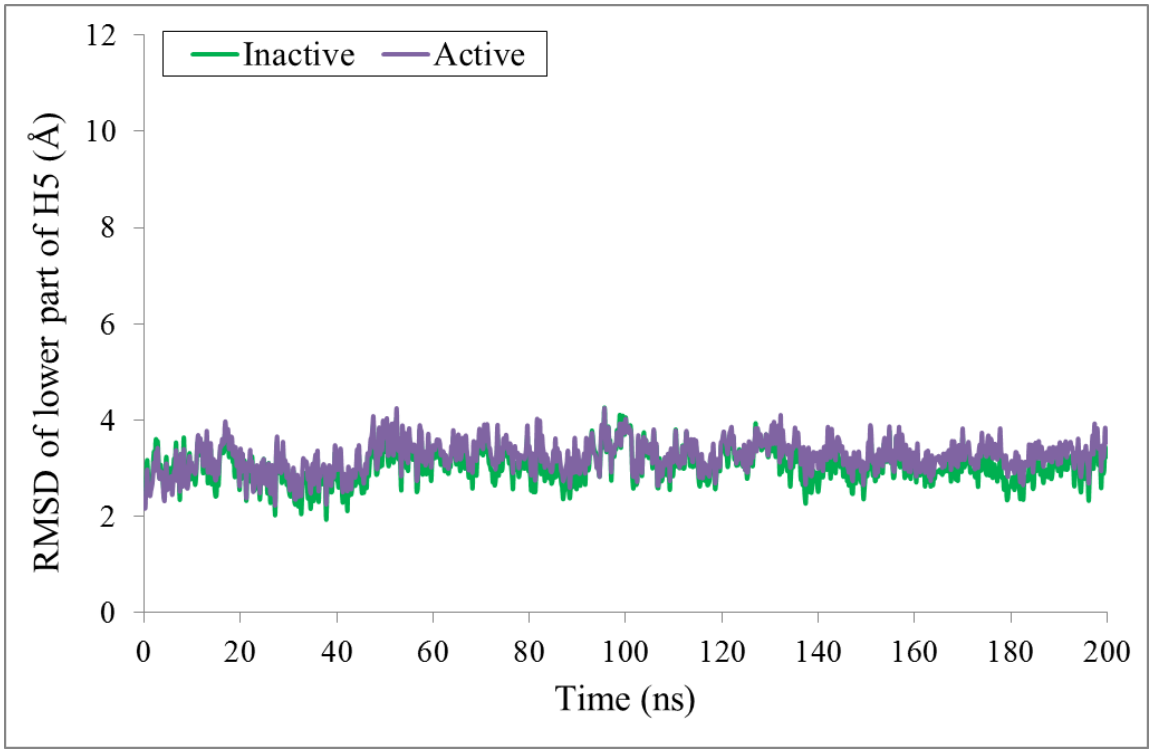


(b)

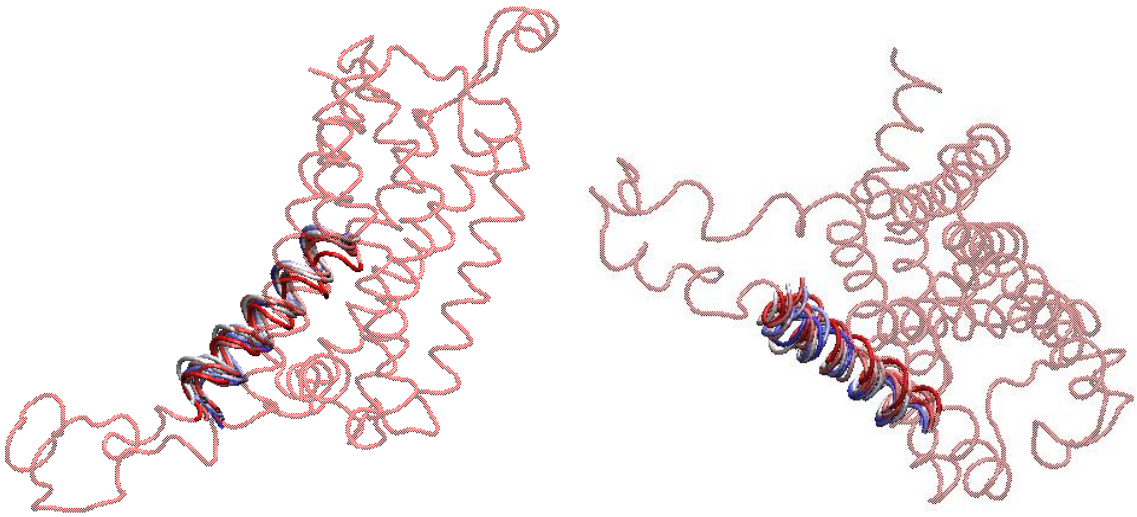


(c) (d)

Figure 4.59. *c5_200ns_interm_open*'s (a) RMSD profile, (b) RMSD of lower part of H6 profile and ten aligned snapshots from (c) side and (d) lower views.

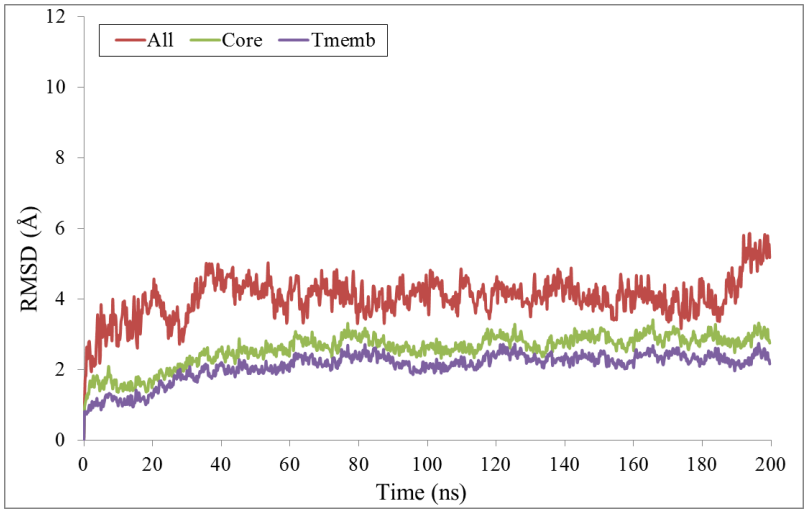


(a)

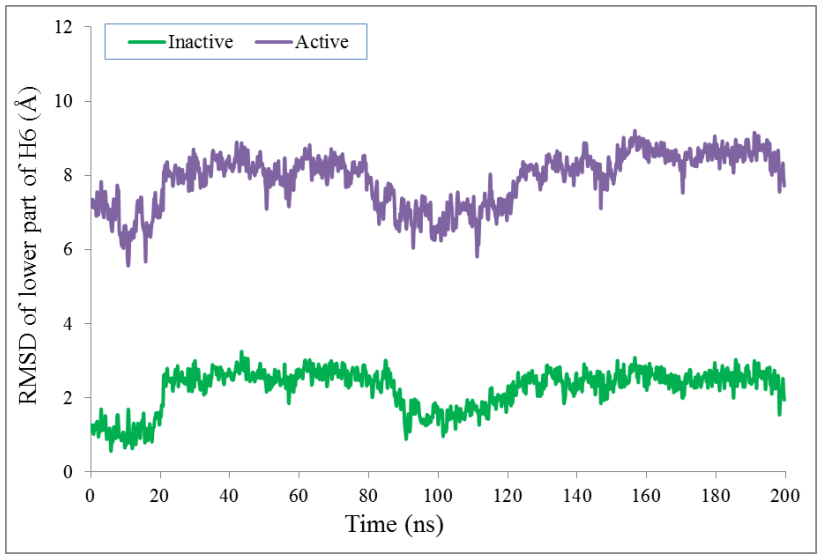


(b) (c)

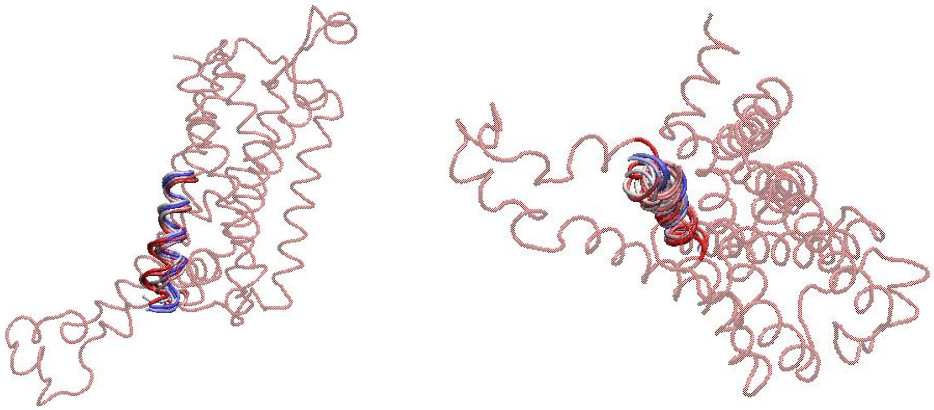
Figure 4.60. *c5_200ns_interm_open*'s (a) RMSD of lower part of H5 profile and ten aligned snapshots from (b) side and (c) lower views.



(a)

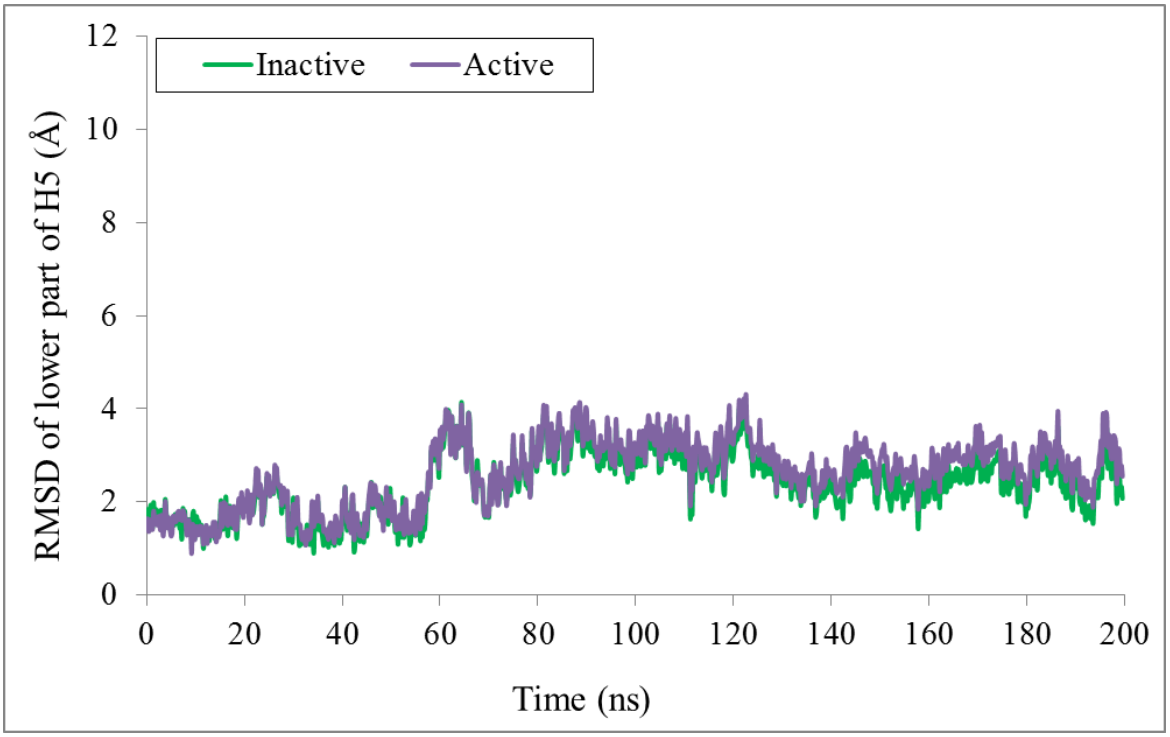


(b)

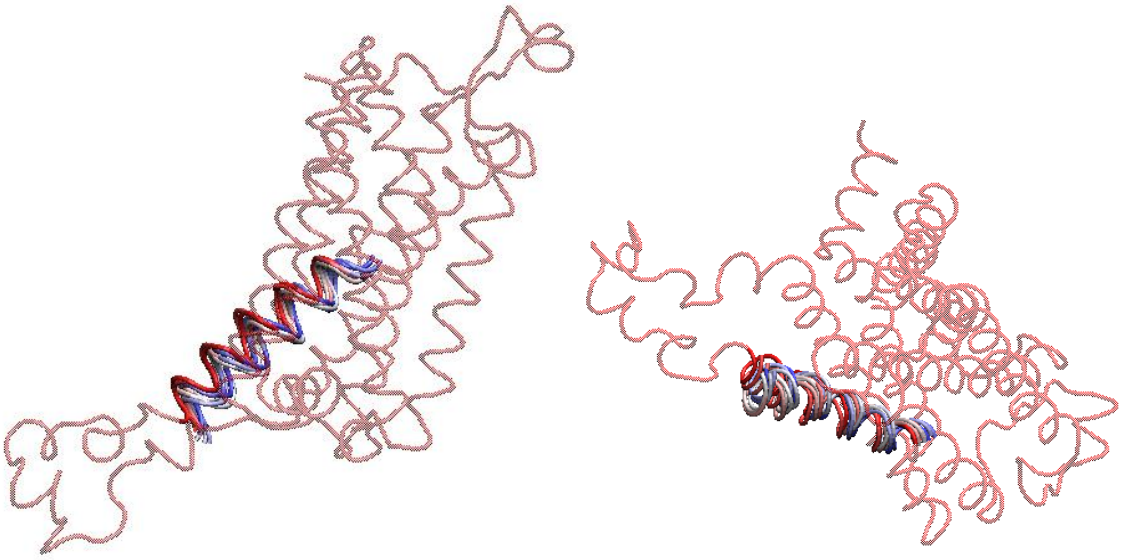


(c) (d)

Figure 4.61. *c6_200ns_interm_closed*'s (a) RMSD profile, (b) RMSD of lower part of H6 profile and ten aligned snapshots from (c) side and (d) lower views.



(a)



(b) (c)

Figure 4.62. *c6_200ns_interm_closed'*s (a) RMSD of lower part of H5 profile and ten aligned snapshots from (b) side and (c) lower views.

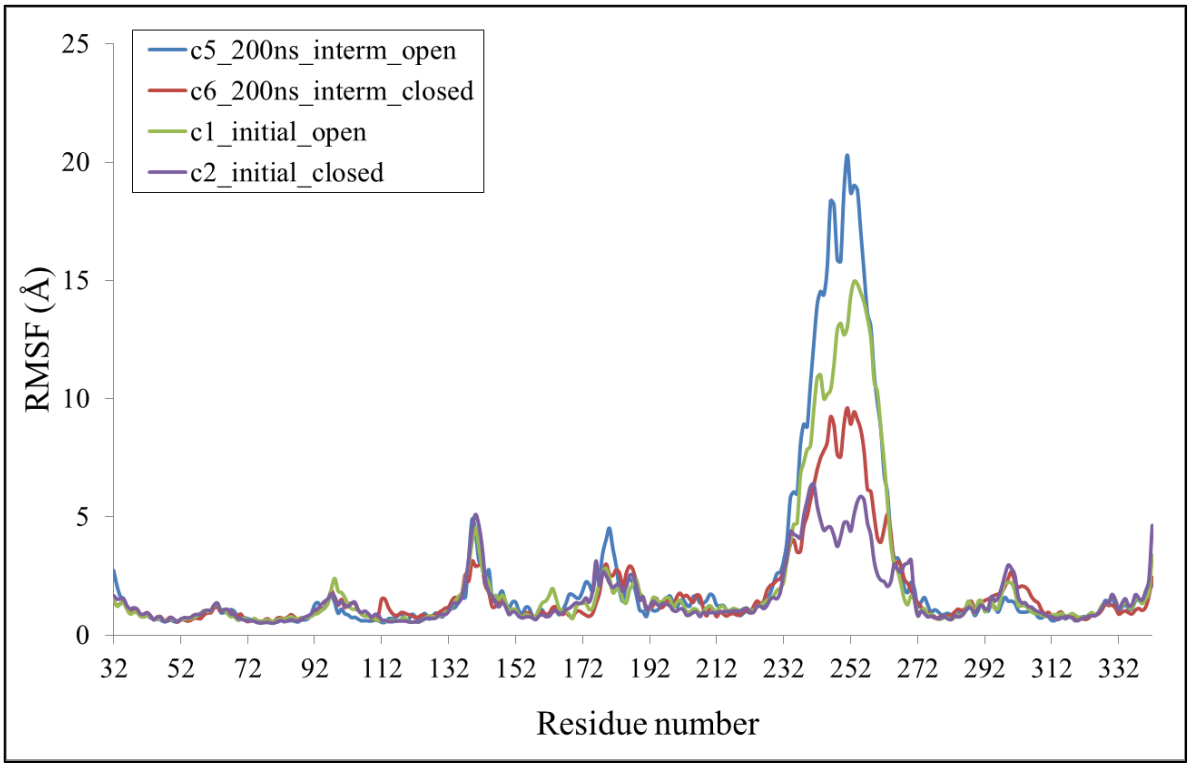


Figure 4.63. RMSF profile of 200 ns long constrained trajectories with intermediate structure compared to 500 ns long trajectories with initial structure.

4.6. Hydrogen Bond Profiles

Additionally, because of their roles in allosteric mechanism, H5 and H6 are observed in detail according to their hydrogen bonds stabilizing the helical structure. Intact H-bonds on the backbone between residues *i* and *i+3*, *i* and *i+4*, *i* and *i+5* represent 3-10 helices, α helices and π helices, respectively. The radius of 3-10 helices is less than that of α helices because two residues *i* and *i+3* are nearer to each other than *i* and *i+4*, decreasing the diameter of one spiral of the helical structure. If none of these hydrogen bonds exist for residue *i*, it can be said that the helical structure is locally broken, which may be indicative of local unfolding or a hinge region on the helix. In the following figures, the white regions indicate the residues with no hydrogen bond formation on the backbone. Green, yellow and red areas on the figures indicate the presence of *i* and *i+3*, *i* and *i+4*, *i* and *i+5* hydrogen bonds, respectively. The residues where two different colors coincide mean that there are two bond types for that specific residue *i*. The expression "upper" on the figures symbolizes the receptor near to the extracellular region, while "lower" indicates the receptor near to the intracellular region.

Figure 4.64 shows the hydrogen bond profiles of H5 for *orig_1st half, orig_2nd half, continued_1,* and *continued_2*. The x coordinate gives the frame number while the y coordinate is the residue number which is *i* (acceptor) in this case. Generally, the results show that the upper part of H5 is more distorted then the lower part. The disruptions in the helices are clearly seen in all the models between the residues 205-207 acting like a hinge region allowing the upper and lower parts of the helix to move around the pivot. In *orig_2nd half,* there is another distortion of hydrogen bonds between residues 200-203 and frames 250-1700 (50-340 ns). As known from the other analyses, the upper region of H5 moves away from the receptor slightly with the closure of ICL3 towards the end of the simulation *original.* The occurrence of this deteriorated region of H5 almost coincides with the same time. Around 350 ns (frame 1750), this region turns into pi helices (red) with the occurrence of *i* and *i+5* hydrogen bonds between residues 198-205. *i+3* bonds in green populate at the lower part of H5 near ICL3, especially at residue 225. *i+4* bonds indicating α helices dominate the general view of H5 in yellow regions. *i+5* bonds in red (π helices) are between residues 198-205, starting from *orig_1st half* and increasing towards the ends of *orig_2nd half.* The two extended simulations *continued_1* and *continued_2* are very similar, keeping the same amount of *i+5* bonds till the end of the trajectories.

In Figure 4.65, the hydrogen bond profiles of H5 of all the constrained simulations are shown separately. The same disruption seen in the unconstrained models is seen here between the residues 205-207, the pivot region. *i+3* bonds gather on the lower part of H5 especially on residue 225, *i+4* bonds dominate all the figures except the pivot region. *i+5* bonds dominate at the upper region, however they are less than the unconstrained runs' *i+5* bonds. *i+5* bonds of *c3_interm_open* and *c4_interm_closed* increase between residues 225-228 on the lower part of H5 near ICL3 compared to the initial constrained models. The pivot regions where there is a lack of hydrogen bonds are generally same; however, there is an increase in *c3_interm_open* between residues 199-201 and 219-221. The upper region of H5 of *c5_200ns_interm_open* has a popularity of *i+5* bonds between residues 199-200 compared to the whole H5 structure of this model. The disruptions are between residues 200-201 and 207-209. The pivot regions of *c6_200ns_interm_closed* are nearly at the same sites of residues 206-208 and at some parts of residue 227. There is a popularity of *i+3* bonds in green at residue 225 coinciding with *i+5* bonds in red at some points for both models. Generally, the models resemble the initial constrained models.

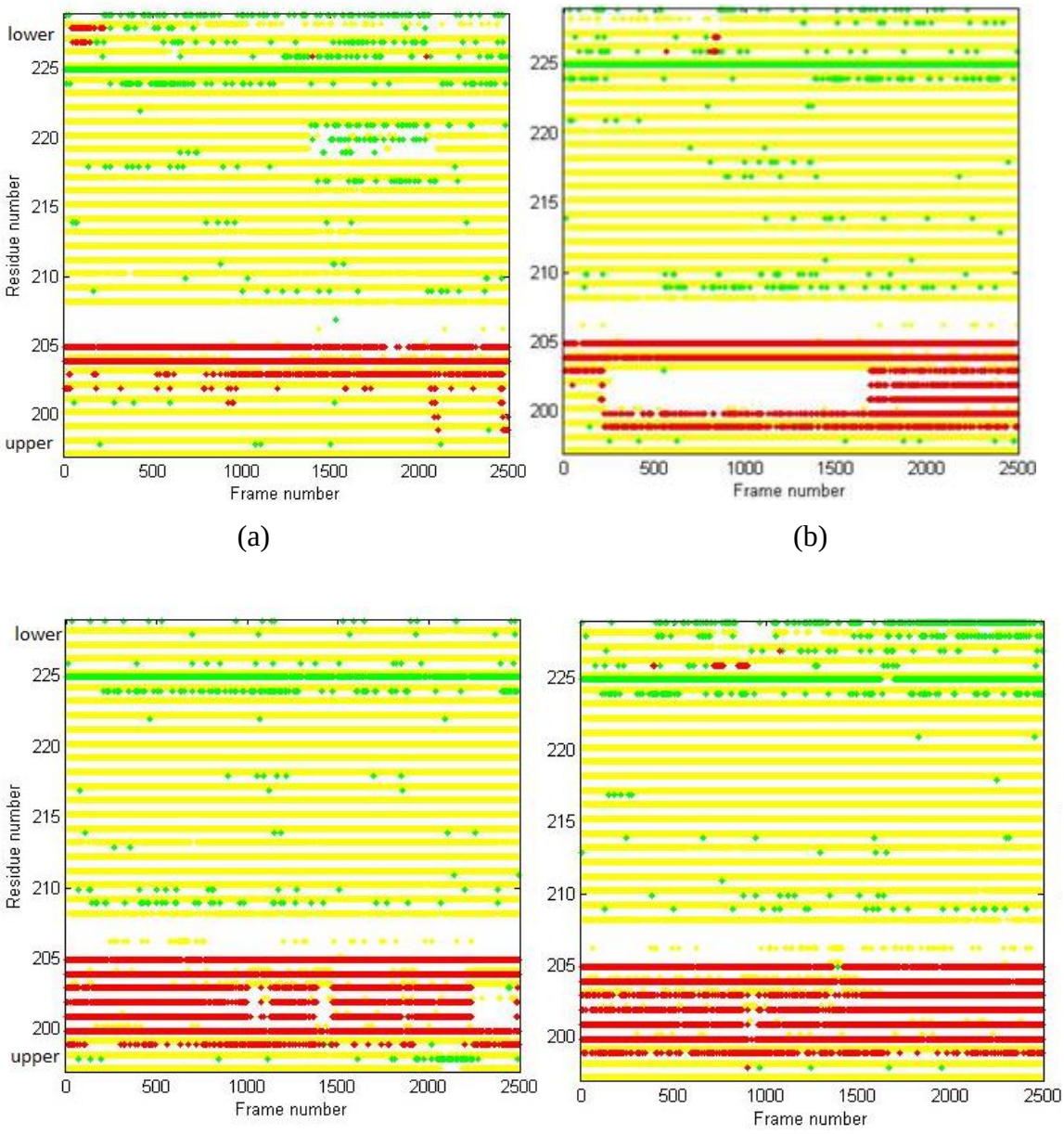


Figure 4.64. Hydrogen bond profiles of H5 for (a) *orig_1st half*, (b) *orig_2nd half*, (c) *continued_1*, and (d) *continued_2*.

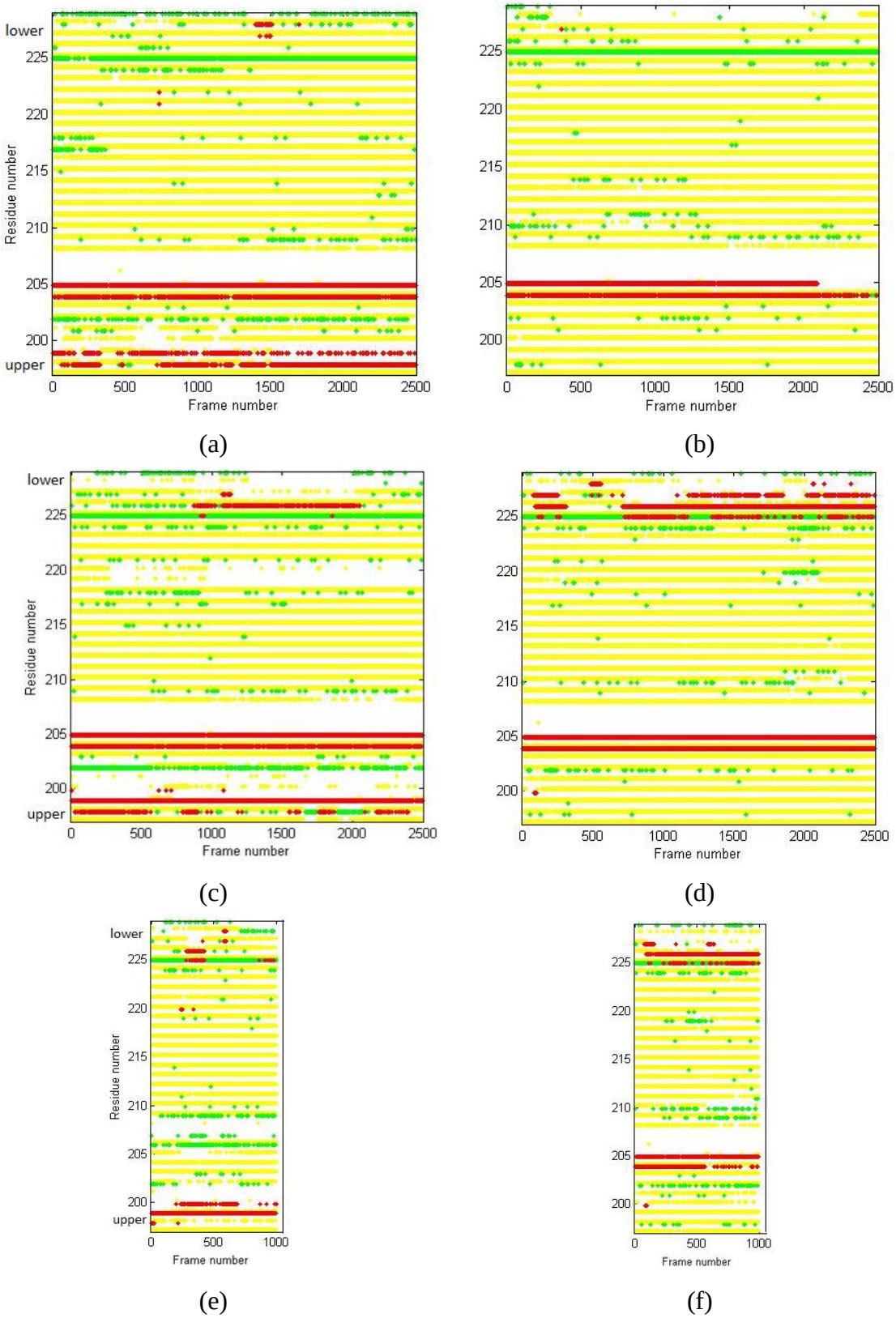


Figure 4.65. Hydrogen bond profiles of H5 for (a) *c1_initial_open*, (b) *c2_initial_closed*, (c) *c3_interm_open*, (d) *c4_interm_closed*, (e) *c5_200ns_interm_open*, (f) *c6_200ns_interm_closed*.

Figure 4.66 shows the hydrogen bond disruptions of H6 for the unconstrained models. The regions without hydrogen bonds are on the upper region of H6, between residues 283-286 and 297-298. *i+3* bonds are more widespread compared to H5 and populate at the pivot region of residues 283-286. *i+4* bonds again dominates the helical structure of H6 like in H5 while *i+5* bonds are at the upper regions of the second half of *orig_2nd half*, *continued_1* and *continued_2* between residues 291-296. It could be expected for the lower part of H6 that there would be more white areas indicating the loss of hydrogen bonds, since this region is more near to ICL3, which pulls it and maybe breaks some hydrogen bonds.

In Figure 4.67, the hydrogen bonds on H6 are shown for the constrained runs. *i+5* bond formation is more popular on the upper region of H6 just like in the unconstrained models. The pivot regions are also coinciding with the unconstrained models, except the formations in *c2_initial_closed* between residues 266-270 after frame 750 (150 ns) and residues 290-292. This can be explained by the unfolding of the lower part of H6 at the last frames seen in the previously shown snapshots of *c2_initial_closed*. Generally, *i+5* bonds are decreased on the upper part of H6 for the intermediate constrained runs, however there is an increase in *c4_interm_closed* between residues 268-270. There is a disruption of hydrogen bonds at the lower side of H6 between residues 270-273 for *c4_interm_closed*. Both of the 200 ns long intermediate runs seem to have the least amount of *i+5* bonds in H6 just like *orig_1st half* compared to the previously studied simulations. *c5_200ns_interm_open* has two pivot regions at the middle and upper parts of the helix. *c6_200ns_interm_closed* has the same pivot regions at the upper and middle parts of the helix as well as an additional part at some frames of the lower site.

Normally, residues 203-204 are considered to be pivot regions, but none of the hydrogen bond profiles showed this area free of hydrogen bonds. Figure 4.68 shows the pivot regions common to all the models in H5 and H6. The hinge region in H5 shown in blue is between residues 205-207 and the one in H6 shown in green is between residues 284-286.

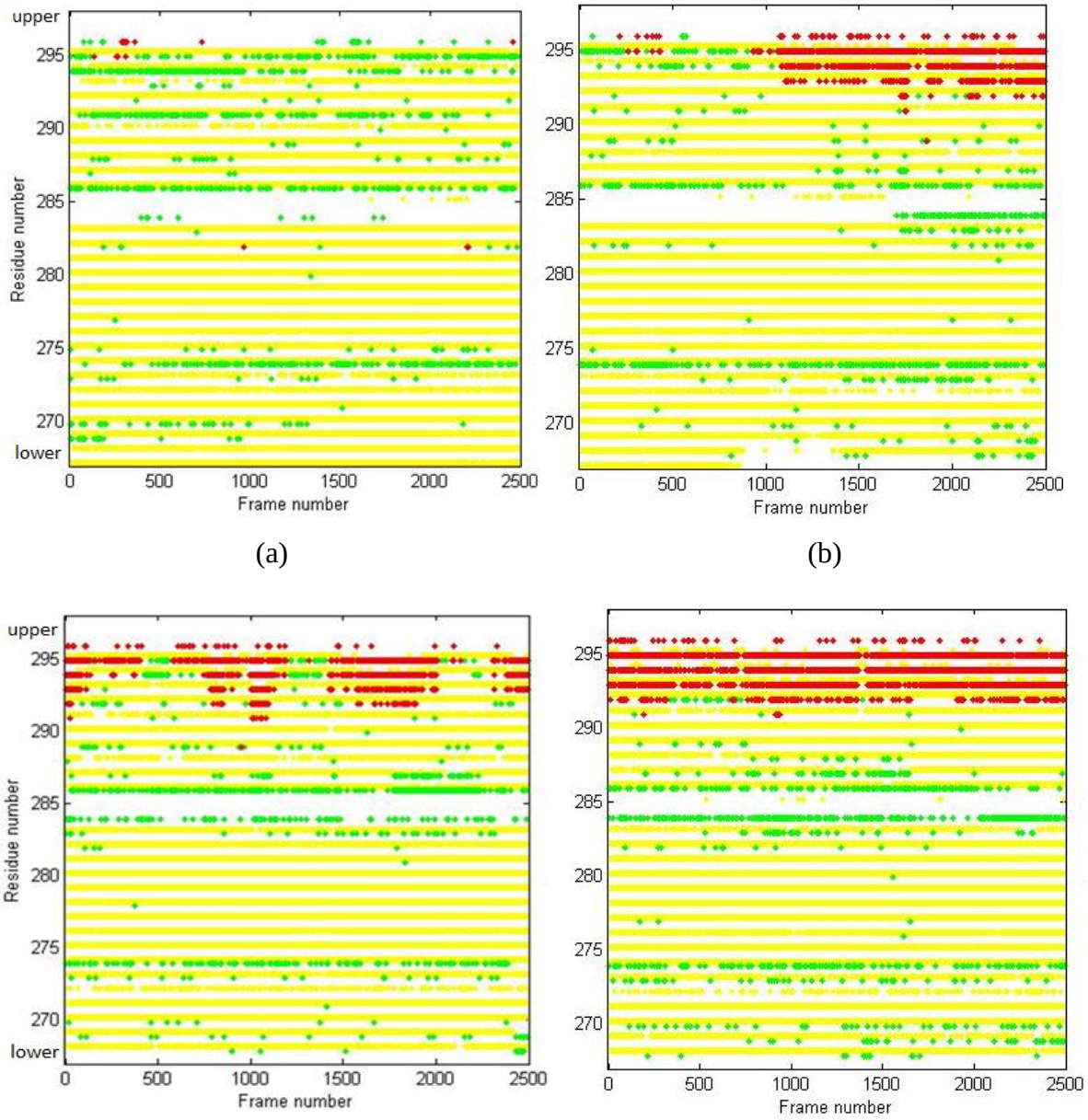


Figure 4.66. Hydrogen bond profiles of H6 for (a) *orig_1st half*, (b) *orig_2nd half*, (c) *continued_1*, and (d) *continued_2*.

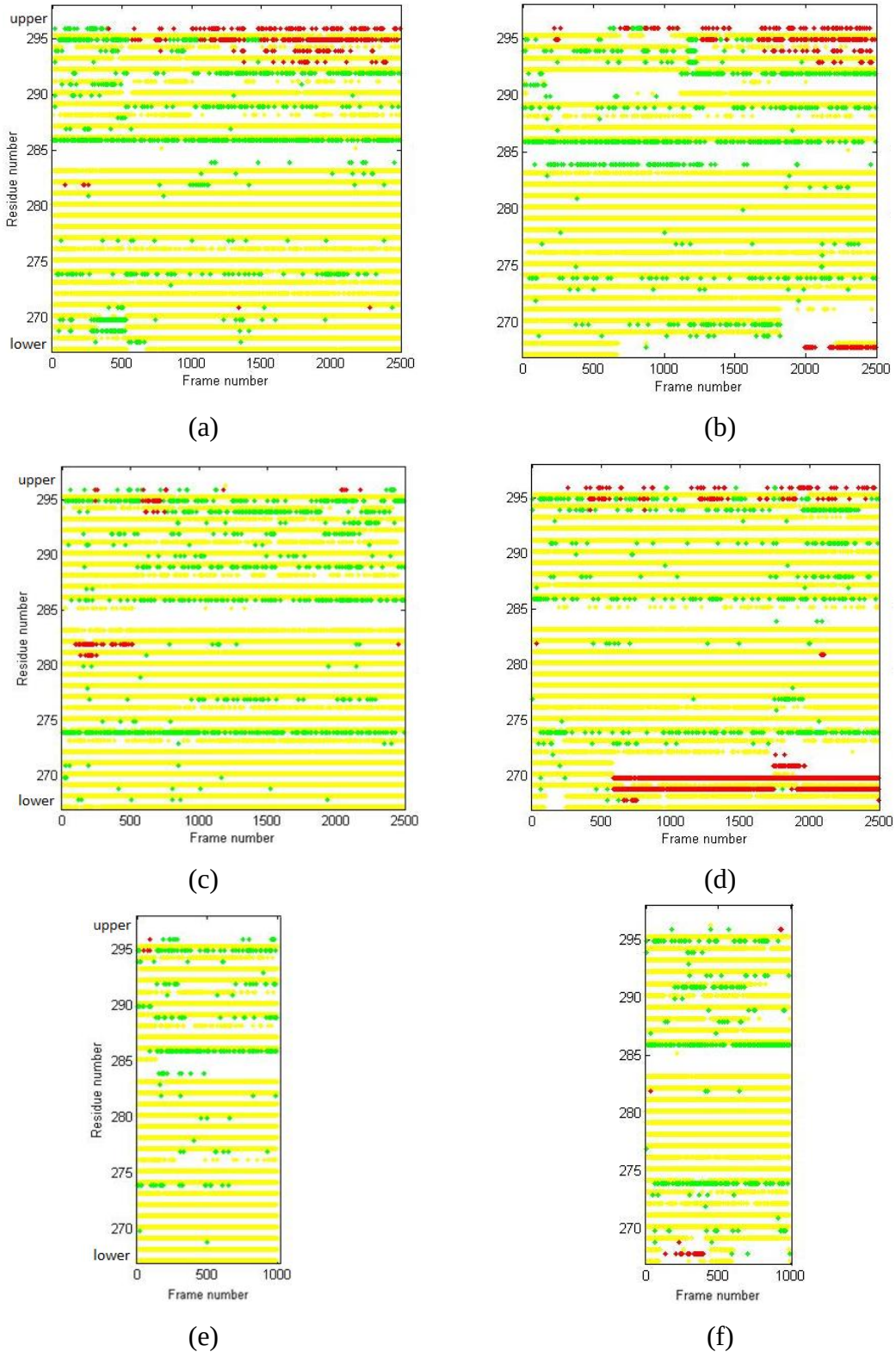


Figure 4.67. Hydrogen bond profiles of H6 for (a) *c1_initial_open*, (b) *c2_initial_closed.* (c) *c3_interm_open*, (d) *c4_interm_closed*, (e) *c5_200ns_interm_open*, (f) *c6_200ns_interm_closed.*

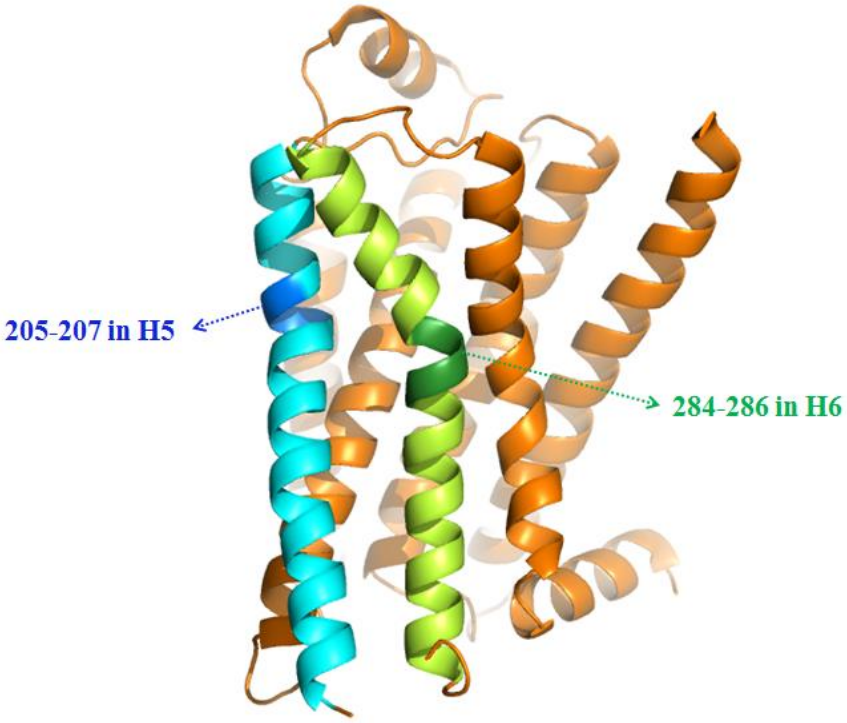


Figure 4.68. Hinge regions in H5 and H6.

5. CONCLUSION AND FUTURE STUDIES

5.1. Conclusion

β_2AR is a receptor molecule of the GPCR family with crucial roles in human metabolism. Several unconstrained and constrained MD simulations were performed on β_2AR in order to investigate the allosteric coupling between extracellular and intracellular regions of the receptor.

The unconstrained models (*continued_1* and *continued_2*) generated from the last snapshot of *original* showed that the closed ICL3 did not change its position much under the receptor, thus representing a quite stable state. The distance between Ser207-Asp113 increased throughout the trajectory for *original*, meaning that the upper regions of H5 and H3 moved away from each other during the closure of ICL3, while this distance stayed in a relatively open position during the 500 ns long continued runs. The first principal mode of *original* showed that the intracellular parts of H5 and H6 were pulled towards inside by the movement of ICL3. There were other coupled movements of the transmembrane helices, which were not observed in the continuation runs with ICL3 closed.

In order to see the effect of constraints on the movement of ICL3, harmonic constraints were applied on seven pairs of residues on the ligand binding site of the receptor, located on the upper regions of H5 (residues Ser203, Ser204 and Ser207), H3 (Asp113), H6 (Phe289 and Asn293) and H7 (Asn312). Two constrained runs started from the initial conformation of *original* with ICL3 in fully open position. *c1_initial_open*, whose ligand binding site was constrained in an open position showed the closure of ICL3 consistent with that observed in *original* but on a shorter time scale. *c2_initial_closed*, whose ligand binding site was narrowed down compared to the inactive state, showed an ICL3 profile changing position in the vicinity of the open conformation.

Two other constrained runs were performed starting with the 470 ns snapshot of *original* with ICL3 in an intermediate (relatively open) position. Even though the expectation was observing a faster opening or closure of ICL3 due to constraints applied,

c3_interm_open and *c4_interm_closed* showed that ICL3 adopted intermediate positions, i.e. not full closure or opening, in both runs. Because of these unexpected results, an extra constraint was applied between C^{α} backbones of Ser207-Asp113 for generating two new 200 ns trajectories starting from intermediate ICL3. Specifically, *c5_200ns_interm_open* lead to ICL3 closure within 80 ns and *c6_200ns_interm_closed* kept its ICL3 in relatively open position. The success of the runs with an extra constraint on the backbone indicates that constraints only on side chains may not transfer the allosteric effect to the intracellular side of the receptor.

5.2. Future Studies

Imposing extra bond constraints to the selected pair of residues did not always yield the desired conformation. In our study, the hypothetical bond between Ser203 on H5 and Asp113 on H3 constrained to a larger value did not always result in an outward motion of the extracellular part of H5, which is believed to have an allosteric effect on the closure of ICL3. Thus, to overcome the difficulty of obtaining the desired outward motion, an external force can be applied on one or more atoms in H5, in a steered MD simulation.

Furthermore, all seven/eight constraints can be removed altogether in order to investigate the stability of the final state of the receptor at the end of a constrained simulation. As an example, in the first constrained run where the Ser203-Asp113 distance was constrained to 17 Å, a satisfactory closure of ICL3 was observed. However, to investigate whether this final closed state could still be preserved in the absence of any constraints, the final snapshot could be subjected to another MD simulation with no constraints applied. The stability of the final state would simply indicate that the constraints increase the probability of the receptor to sample such a closed state, which would require longer time frames under normal conditions.

Moreover, the constrained residues and helices may be changed in order to see the constraint effects of other regions on the behavior of the protein and the ICL3 region.

Docking and screening experiments can be applied to closed state including the snapshots of *original* and its continuation runs, in which ICL3 is packed under the receptor. By stabilizing this closed state, G-protein binding may be inhibited. Thus, novel antagonists can be discovered.

APPENDIX A: PRINCIPAL COMPONENT ANALYSIS RESULTS

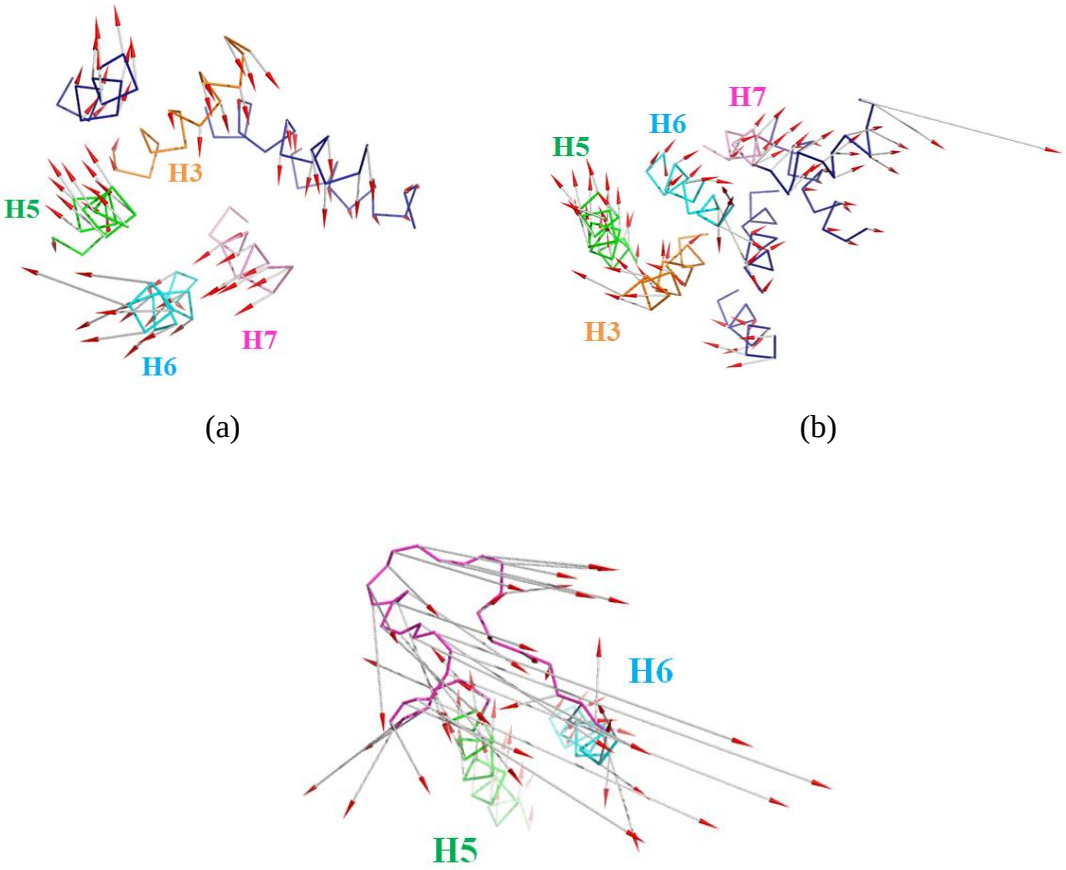


Figure A.1. First principal mode vectors of *c2_initial_closed* with a) upper view, b) lower view, c) lower view with H5, H6 and ICL3.

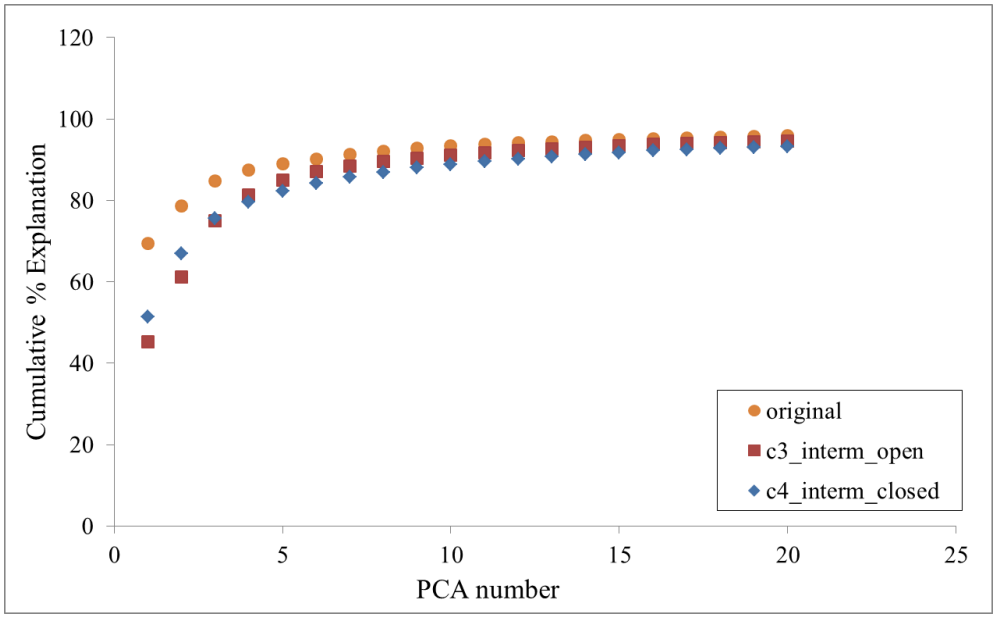
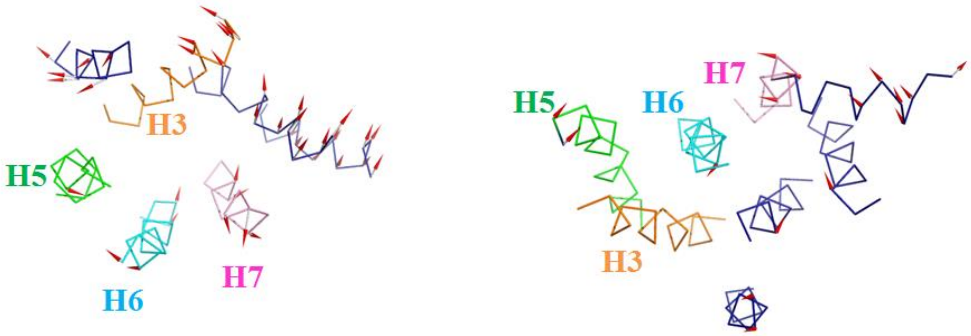
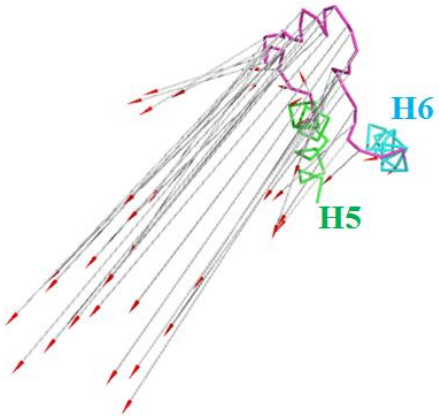


Figure A.2. Cumulative % explanation of principal components for *3_interm_open* and *c4_interm_closed* compared with *original.*

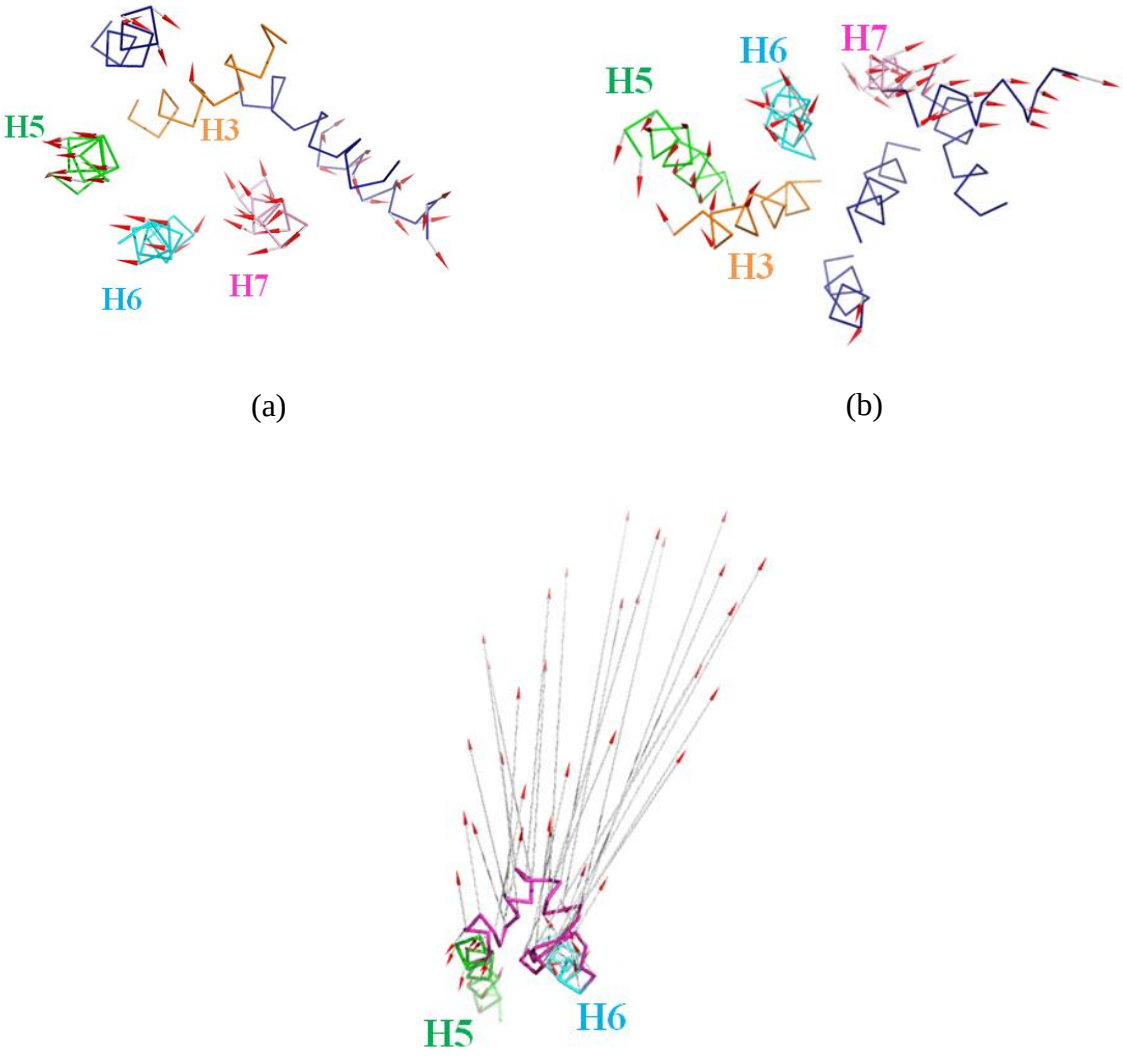


(a) (b)



(c)

Figure A.3. First principal mode vectors of *c3_interm_open* with a) upper view, b) lower view, c) lower view with H5, H6 and ICL3.



(a) (b)

(c)

Figure A.4. First principal mode vectors of *c4_interm_closed* with a) upper view, b) lower view, c) lower view with H5, H6 and ICL3.

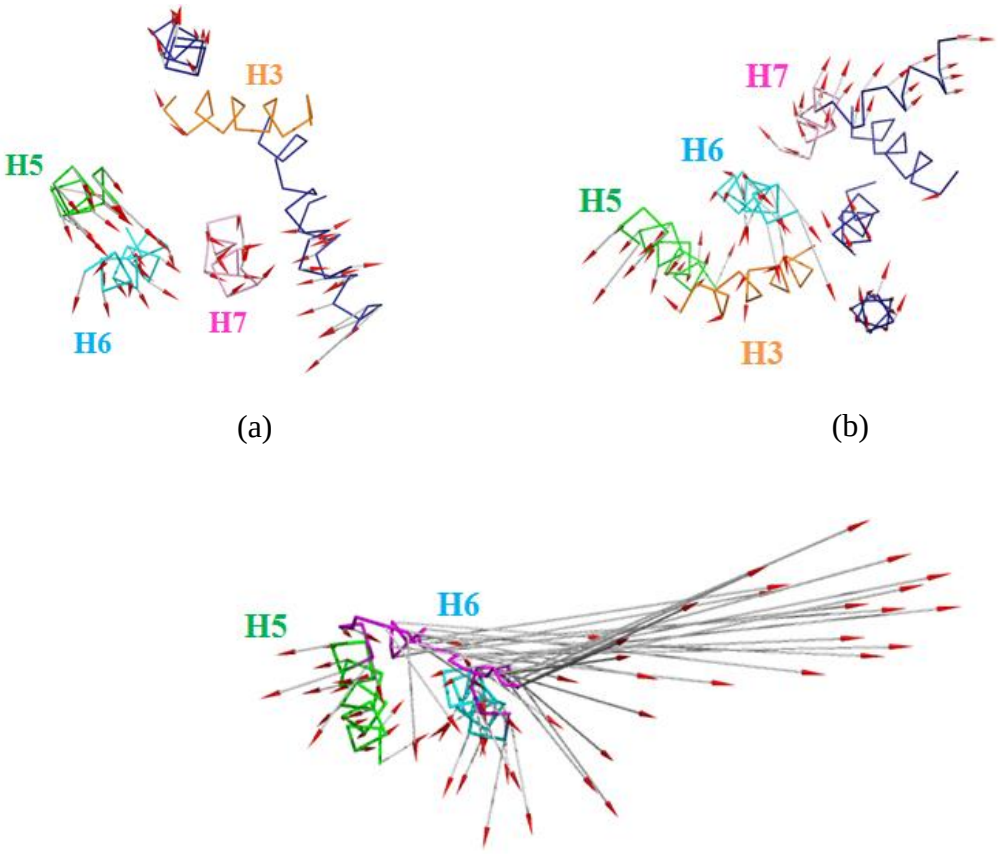


Figure A.5. Second principal mode vectors of *c5_200ns_interm_open* with a) upper view, b) lower view, c) lower view with H5, H6 and ICL3.

REFERENCES

Abdi, H., L. J. Williams, 2010, "Principal Component Analysis", *John Wiley & Sons, Inc., WIREs Computational Statistics*, Vol. 2, No. 4, pp. 433-459.

Ahmed, H. U., M. P. Blakeley, M. Cianci, D. W. J. Cruickshank, J. A. Hubbard and J. R. Helliwell, 2007, "The Determination of Protonation States in Proteins", *Acta Crystallographica*, Vol. 63, No.8, pp. 906-922.

Alakent, B., P. Doruker, and M. C. Camurdan, 2004, "Application of Time Series Analysis on Molecular Dynamics Simulations of Proteins: A Study of Different Conformational Spaces by Principal Component Analysis", *The Journal of Chemical Physics,* Vol. 121, No. 10, pp. 4759-4769.

Altis, A., P. H. Nguyen, R. Hegger, and G. Stock, 2007, "Dihedral Angle Principal Component Analysis of Molecular Dynamics Simulations", *The Journal of Chemical Physics,* Vol. 126, No. 244111, pp.1-10.

Andersen, H. C., 1983, "Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations", *Journal of Computational Physics*, Vol. 52, pp. 24-24.

Bashford, D. and M. Karplus, 1990, "pK_a's of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model", *Biochemistry*, Vol. 29, pp. 10219-10225.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. and Haak, J. R., 1984, "Molecular Dynamics with Coupling to an External Bath", *The Journal of Chemical Physics,* Vol. 81, pp. 3684-3690.

Bernstein, F. C., T. F. Koetzle, G. J. Williams, E. E. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, 1977, "The Protein Data Bank: A

Computer-based Archival File For Macromolecular Structures," *European Journal of Biochemistry,* Vol. 80, pp. 319-324.

Binder, K., J. Horbach, W. Kob, W. Paul and F. Varnik, 2004, "Molecular Dynamics Simulations", *Journal of Physics: Condensed Matter,* Vol. 16, pp. 429-453.

Blessing, R. H., 1995, "An Empirical Correction for Absorption Anisotropy", *Acta Crystallographica,* Section A51, pp. 33-38.

Bokoch, M. P., Y. Zou, S. G. F. Rasmussen, C. W. Liu, R. Nygaard, D. M. Rosenbaum, J. J. Fung, H. J. Choi, F. S. Thian, T. S. Kobilka, J. D. Puglisi, W. I. Weis, L. Pardo, R. S. Prosser, L. Mueller, B. K. Kobilka, 2010, "Ligand-Specific Regulation of the Extracellular Surface of a G-Protein-Coupled Receptor", *Nature,* Vol. 463, pp. 108-114.

Botelho, A. V., T. Huber, T. P. Sakmar, M. F. Brown, 2006, "Curvature and Hydrophobic Forces Drive Oligomerization and Modulate Activity of Rhodopsin in Membranes, *Biophysical Journal.,* Vol. 91, pp. 4464-4477.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, 1983, "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations", *Journal of Computational Chemistry,* Vol. 4, No. 2, pp. 187-217.

Cherezov, V., D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka and R. C. Stevens, 2007, "High-Resolution Crystal Structure of An Engineered Human Beta2-Adrenergic G-Protein-Coupled Receptor", *Science,* Vol. 318, pp. 1258-1265.

Coutsias, Evangelos A., C. Seok, K. A. Dill, 2004, "Using Quaternions to Calculate RMSD", *Journal of Computational Chemistry,* Vol. 25, No. 15, pp. 1849-1857.

DeLano, W. L., 2004, "Use of Pymol as a Communications Tool for Molecular Science", *Abstracts of Papers of the American Chemical Society*, Vol. 228, No. U313-U314.

DeLano, W. L. and J. W. Lam, 2005, "Pymol: A Communications Tool for Computational Models", *Abstracts of Papers of the American Chemical Society*, Vol. 230, No. U1371-U1372.

Deupi, X.,and J. Standfuss, 2011, "Structural Insights into Agonist-Induced Activation of G-Protein-Coupled Receptors", *Current Opinion in Structural Biology*, Vol. 21, pp. 541-551.

Dror, R. O., D. H. Arlow, P. Maragakis, T. J. Mildorf, A. C. Pan, H. Xu, D. W. Borhani, and D. E. Shaw, 2011, "Activation Mechanism of the β_2-Adrenergic Receptor", *Pnas*, Vol. 108, No. 46, pp. 18684-18689.

Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, 1995, "A Smooth Particle Mesh Ewald Method", *The Journal of Chemical Physics*, Vol. 103, No. 19, pp. 8576-8593.

Feller, S. E., Y. Zhang, and R. W. Pastor, 1995, "Computer-Simulation of Liquid/Liquid Interfaces II. Surface-Tension Area Dependence of a Bilayer and Monolayer", *The Journal of Chemical Physics*, Vol. 103, pp. 10267-10276.

Feng, Z., T. Hou, and Y. Li, 2012, "Studies on the Interactions between β_2 Adrenergic Receptor and Gs Protein by Molecular Dynamics Simulations", *Journal of Chemical Information and Modeling*, Vol. 52, pp. 1005-1014.

Foloppe, N., and A. D. Mackerell, 2000, "All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data", *Journal of Computational Chemistry*, Vol. 21, No. 2, pp. 86-104.

Gouldson, P. R., P. J. Winn, and C. A. Reynolds, 1995, “A Molecular Dynamics Approach to Receptor Mapping: Application to the $5HT_3$ and β2-Adrenergic Receptors”, *Journal of Medicinal Chemistry*, Vol. 38, pp. 4080-4086.

Grossfield, A., 2011, “Recent Progress in the Study of G Protein-Coupled Receptors with Molecular Dynamics Computer Simulations”, *Biochimica at Biophysica Acta*, Vol. 1808, No. 7, pp. 1868-1878.

Gunsteren, W. F., and H. J. C. Berendsen, 1990, “Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry”, *Angewandte Chemie International Edition*, Vol. 29, No. 9, pp. 992-1023.

Hanson, M. A., V. Cherezov, C. B. Roth, M. T. Griffith, V. P. Jaakola, E. Y. T. Chien, J. Velasquez, P. Kuhn and R. C. Stevens, 2008, “A Specific Cholesterol Binding Site is Established by the 2.8 Å Structure of the Human β_2-Adrenergic Receptor in an Alternate Crystal Form”, *Structure*, Vol. 16, No. 6, pp. 897-905.

Humphrey, W., Dalke, A. and Schulten, K., 1996, "VMD - Visual Molecular Dynamics", *Journal of Molecular Graphics*, Vol. 14, pp. 33-38.

Jackson, J. E., 1980, “Principal Component and Factor Analysis: Part I - Principal Components”, *Journal of Quality Technology*, Vol. 12, No. 4, pp. 201-213.

Jensen, A.D., F. Guarnieri, S. G. F. Rasmussen, F. Asmar, J. A. Ballesteros and U. Gether, 2001, “Agonist-Induced Conformational Changes at the Cytoplasmic Side of Transmembrane Segment 6 in the β_2 Adrenergic Receptor Mapped by Site-Selective Fluorescent Labeling”, *Journal of Biological Chemistry*, Vol. 276, pp. 9279-9290.

Katritch, V., V. Cherezov, and R. C. Stevens, 2012, “Diversity and Modularity of G Protein-Coupled Receptor Structures”, *Trends in Pharmacological Science*s, Vol. 33, No. 1, pp. 17-27.

Katritch, V., V. Cherezov, R. C. Stevens, 2013, "Structure-Function of the G Protein-Coupled Receptor Superfamily", *Annual Review of Pharmacology and Toxicology.*, Vol. 53, pp. 531-556.

Kohlhoff, Kai J., D. Shukla, M. Lawrenz, G. R. Bowman, D. E. Konerding, D. Belov, R. B. Altman, V. S. Pande, 2013, "Cloud-Based Simulations on Google Exacycle Reveal Ligand Modulation of GPCR Activation Pathways", *Nature Chemistry,* Vol. 6, pp. 15-21.

Labute, P., 2007, *Protonate 3D: Assignment of Macromolecular Protonation State and Geometry,* Chemical Computing Group Incorporation, http://www.chemcomp.com/journal/proton.htm, [Accessed July 2014].

Lantzsch, G., H. Binder, H. Heerklotz, 1994, "Surface Area per Molecule in Lipid/C12En Membranes as Seen by Fluorescence Resonance Energy Transfer", *Journal of Fluorescence,* Vol. 4, No. 4, pp. 339.

Leach, A. R., 2001, *Molecular Modelling: Principles and Applications,* 2nd edition, Prentice Hall, Harlow.

Lohse, M. J, 2010, "Dimerization in GPCR Mobility and Signaling", *Current Opinion in Pharmacology,* Vol. 10, No. 1, pp. 53-58.

MacKerell, A. D., N. Banavali, and N. Foloppe, 2001, "Development and Current Status of the CHARMM Force Field for Nucleic Acids", *Biopolymers,* Vol. 56, No. 4, pp. 257-265.

Maiorov, V. N., and G. M. Crippen, 1994, "Significance of Root-Mean-Square Deviation in Comparing Three-dimensional Structures of Globular Proteins", *Journal of Molecular Biology,* Vol. 235, No. 2, pp. 625-634.

Narayanan, E., B. John, N. Mirkovic, A. Fiser, V. A. Ilyin, U. Pieper, A. C. Stuart, M. A. Marti-Renom, M. S. Madhusudhan, B. Yerkovic and A. Sali, 2003, "Tools for

Comparative Protein Structure Modeling and Analysis", *Nucleic Acids Research,* Vol. 31, No. 13, pp. 3375-3380.

Neria, E., Fischer, S., and Karplus, M., 1996, "Simulation of Activation Free Energies in Molecular Systems", *The Journal of Chemical Physics*, Vol. 105, pp. 1902-1921.

Oldham, W. M., H. E. Hamm, 2008, "Heterotrimeric G Protein Activation by G-protein-Coupled Receptors", *Nature Reviews Molecular Cell Biology,* Vol. 9, No. 1, pp. 60-71.

Ozcan, O., *Exploring the Intrinsic Dynamics of Human Beta-2 Adrenergic G-Protein Coupled Receptor*, M.S. Thesis, Bogazici University, 2011.

Ozcan O., A. Uyar, P. Doruker, E. D. Akten, 2013, "Effect of Intracellular Loop 3 on Intrinsic Dynamics of Human β_2-Adrenergic Receptor", *BMC Structural Biology,* Vol. 13, No. 29, pp. 1-16.

Palczewski, Krzysztof, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. L. Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, and M. Miyano, 2000, "Crystal Structure of Rhodopsin: A G-Protein-Coupled Receptor", *Science*, Vol. 289, pp. 739-745.

Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. J. Schulten, 2005, "Scalable Molecular Dynamics with NAMD", *Journal of Computational Chemistry*, Vol. 26, No. 16, pp. 1781-1802.

Rasmussen S. G. F, H. J. Choi, M. D. Rosenbaum, T. S. Kobilka, S. F. Thian, C. P. Edwards, M. Burghammer, R. P. V. Ratnala, R. Sanishvili, F. R. Fischetti, F. X. G. Schertler, I. W. Weis, B. K. Kobilka, 2007, "Crystal Structure of the Human β_2 Adrenergic G-Protein-Coupled Receptor", *Nature,* Vol. 450, No. 7168, pp. 383 – 388.

Rasmussen S. G. F, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah ST, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK, 2011, "Crystal Structure of the β_2 Adrenergic Receptor-Gs Protein Complex", *Nature*, Vol. 477, No. 7366, pp. 549-555.

Ring, A. M., A. Manglik, A. C. Kruse, M. D. Enos, W. I. Weis, K. C. Garcia and B. K. Kobilka, 2013, "Adrenaline-Activated Structure of the β_2-Adrenoceptor Stabilized by an Engineered Nanobody", *Nature*, Vol. 502, No. 7472, pp. 575-579.

Romo, T. D, A. Grossfield, and M. C. Pitman, 2010, "Concerted Interconversion Between Ionic Lock Substates of the Beta-2 Adrenergic Receptor Revealed by Microsecond Timescale Molecular Dynamics", *Biophysical Journal*, Vol. 98, pp. 76-84.

Rosenbaum, D. M., S. G. F. Rasmussen, B. K. Kobilka, 2009, "The Structure and Function of G-Protein-Coupled Receptors", *Nature*, Vol. 459, pp. 356-363.

Ryckaert, J. P., G. Ciccotti, H. J. C. Berendsen, 1977, "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes", *Journal of Computational Physics*, Vol. 23, pp. 327-341.

Simpson, L. M., I. D. Wall, F. E. Blaney, and C. A. Reynolds, 2011, "Modeling GPCR Active State Conformations: The β_2-Adrenergic Receptor", *Proteins*, pp. 1441-1457.

Stote, R. H., A. Dejaegere, and D. Kuznetsov, 1999, *Charmm, Molecular Dynamics Simulations*, http://www.ch.embnet.org/MD_tutorial, Version 1.0, Part 1, [Accessed July 2014].

Vanni, S., M. Neri, I. Tavernelli, and U. Rothlisberger, 2011, "Predicting Novel Binding Modes of Agonists to β Adrenergic Receptors Using All-Atom Molecular Dynamics Simulations", *PLoS Computational Biology*, Vol. 7, No. 1, e1001053.

Vauquelin, G., B. Mentzer, 2007, *G Protein-Coupled Receptors, Molecular Pharmacology from Academic Concept to Pharmaceutical Research*, John Wiley & Sons Inc, West Sussex.

Viswanathan, G. M., V. Afanasyev, S. V. Buldyrev, E. J. Murphy, P. A. Prince and H. E. Stanley, 1996, "Levy Flight Search Patterns of Wandering Albatrosses", *Nature*, Vol. 381, PP. 413-415.

Wacker, D., G. Fenalti, M. A. Brown, V. Katritch, R. Abagyan, V. Cherezov, R. C. Stevens, 2010, "Conserved Binding Mode of Human β_2-Adrenergic Receptor Inverse Agonists and Antagonist Revealed by X-ray Crystallography", *Journal of American Chemical Society*., Vol. 132, No. 33, pp. 11443-11445.