

EXPLORING THE INTRINSIC DYNAMICS OF HUMAN BETA-2 ADRENERGIC  
G-PROTEIN COUPLED RECEPTOR

by

Özer Özcan

B.S, Chemical Engineering, Yıldız Technical University, 2007

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  
2011

## ACKNOWLEDGEMENTS

First of all, I would like to express my truthful gratitude to my thesis supervisors Prof. Pemra Doruker Turgut and co-supervisor Assist. Prof. Demet Akten Akdoğan, who devoted their valuable time to guiding me, helping me and motivating me all the time. It was a great opportunity for me to work with them during my thesis, where I learned from them great expertise and experiences in this study.

I also thank to Prof. Viktorya Aviyente, Prof. Levent Kurnaz and Assist. Prof. Nevra Özer, who have read and commented on my thesis.

Special thanks to Arzu Uyar, who devoted her valuable time to guide and help me all the time. I also wish to express my gratitude to other members of Polymer Research Center (PRC), Fidan Sümbül, Seren Soner, Zeynep Kürkçüoğlu, Şefik Kerem Ovalı, Duygu Kocaman, Gülgün Ural, Cihan Kaya, Mehmet Ünal Güneş, Çağlar Meriçer, Mehmet Tarık Can, Harun Ferit Özbakır and also Ahmet Özcan, Belkız Merve Eropak, Bahar Nalbantoğlu, Nilay Aktürk, Hasan Özgen Sicim, Arman Savran, Osman Yüksel, Erdem Eren for giving me their everlasting friendship and joyful help all the time.

Cordial thanks for Canan Dedeoğlu, Avni Ercan, Bilgi Dedeoğlu and for their friendly attitudes and contributions.

Finally, I would like to thank my family for their unrequited support and trust in me throughout my thesis like my whole life. This work is dedicated to them, without whom it would have never been possible.

The graduate scholarship provided by TÜBİTAK for my M.S. studies deserves thankful recognition. Financial support provided by TÜBİTAK through project 109M281, by Boğaziçi University through project BAP 06A509D are gratefully acknowledged.

## ABSTRACT

### EXPLORING THE INTRINSIC DYNAMICS OF HUMAN BETA-2 ADRENERGIC G-PROTEIN COUPLED RECEPTOR

G protein-coupled receptors (GPCRs) represent the single largest family of cell surface receptors involved in signal transduction. It is estimated that several hundred distinct members of this receptor family in humans direct responses to a wide variety of chemical transmitters, including biogenic amines, amino acids, peptides, lipids, nucleosides, and large polypeptides. These transmembrane receptors are key controllers of such diverse physiological processes as neurotransmission, cellular metabolism, secretion, cellular differentiation, and growth as well as inflammatory and immune responses. GPCRs therefore represent major targets for the development of new drug candidates with potential application in all clinical fields. In this thesis, the crystal structure of a human  $\beta_2$ -adrenergic receptor ( $\beta_2$ AR) complexed with a partial inverse agonist carazolol was used as a starting conformation (PDB ID: 2rh1). The missing intracellular loop III (ICL3), which plays an essential role in G protein recognition, was estimated via homology modelling. An alternative model of the receptor with missing loop, which called as “clipped” model, was also used in order to understand the effect of the loop on the dynamics. The purpose of this study is to explore the dynamics of the receptor and the effect of the generated loop on the whole structure the shed light on the results of MD simulations. Here we characterize 0.8  $\mu$ s and 0.5  $\mu$ s all-atom MD simulation of an apo- $\beta_2$ AR for looped and clipped model respectively. Also Asn187 residue has been replaced with Glu187 to facilitate crystallization for the clipped model. Though it is estimated that a single residue will not affect the dynamics of the system significantly, to obtain a definite outcome, the system, undergoing no mutation, has been formed with its natural contents, and being placed within cell membrane, it has been exposed to a 0.5  $\mu$ s all-atom MD simulation of an apo- $\beta_2$ AR. From MD studies, it was shown that the global orientation of the loop region (ICL3) changed considerably with respect to the core structure. The maximum mobility was observed for ICL3 and short loops ICL2, ECL2 and ECL3 connecting the transmembrane helices.

## ÖZET

### **BETA-2 ADRENERJİK RESEPTÖR PROTEİNİN İÇSEL DİNAMİK YAPISININ ARAŞTIRILMASI**

G-proteinine bağlanan reseptör proteinleri hücrenin sinyal aktarımına dahil olan en geniş reseptör ailesini temsil eder. Birkaç yüz farklı üyeye sahip olan bu reseptör ailesi insanda çeşitli fizyolojik işlemlerin örneğin, nörotransmisyon, hücrenel metabolizma, hücrenel salgılama, hücrenel farklılaşma ve gelişim işlemlerinin yanı sıra iltihaplanma ve bağışıklık sisteminde de kilit elemanlarıdır. Ayrıca bu reseptörler tüm klinik alanlarda uygulanabilirlik potansiyeline sahip yeni geliştirilen ilaçlarında hedefinde olan yapılardır. Bu tezde  $\beta_2$  adrenerjik reseptörün kristal yapısı, bu reseptöre bağlanan ve reseptörü inhibe eden carazolol ile birlikte başlangıç yapısı olarak kullanılmıştır. Reseptördeki eksik olan ve hücrenin iç kısmına bakan düğüm bölgesi ki bu bölge G-protein ile reseptör arasındaki temasta önemli rol oynar, homoloji modellemesi yolu ile tamamlanmıştır. Bu “düğüm” modele alternative olması ve ayrıca reseptörün eksik bölgesini tamamlamak amacıyla oluşturulan düğüm bölgesinin, reseptörün dinamiği üzerindeki etkisini anlamak amacıyla bir model daha oluşturulmuştur. Bu model, eksik olan düğüm bölgesi yerine eksik bölgenin iki ucu birleştirilerek oluşturulmuş ve dolayısıyla adına da “kırpık” model denilmiştir. Bu çalışmadaki ana amaç, reseptörün dinamiğini ve oluşturulan düğüm bölgesinin reseptörün dinamiği üzerindeki etkisini moleküler dinamik (MD) simülasyonu ile incelemektir. Bu nedenle reseptörün ligandsız haline düğüm model için 0.8 ve kırpık model için 0.5 mikro saniyelik MD simülasyonlar uygulanmıştır. Ayrıca, tek bir rezidünün mutasyonu sonucu reseptörün dinamiğinin değişip değişmeyeceğini anlayabilmek için ise kırpık modelin 187. rezidüsü mutasyona uğratarak yeni bir kırpık model yaratılmış ve bu model de 0.5 mikro saniyelik bir MD simülasyonuna tabi tutulmuştur. Bu çalışmalar sonucu, oluşturulan düğüm bölgesinin reseptörün dinamiğini önemli ölçüde etkilediği ve tüm bölgeler içinde en hareketli bölge olduğu sonucuna varılmıştır. Ayrıca hücre içine ve dışına bakan diğer düğüm bölgelerde, örneğin, ICL2, ECL2 ve ECL3, yüksek hareketlilik görülmüştür.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
ABSTRACT .....	iv
ÖZET .....	v
LIST OF FIGURES .....	ix
LIST OF TABLES .....	xix
LIST OF SYMBOLS .....	xxi
LIST OF ACRONYMS/ABBREVIATIONS .....	xxiii
1. INTRODUCTION .....	1
2. THEORETICAL BACKGROUND .....	3
2.1. G-protein Mediated Signal Transduction Pathway .....	5
2.2. Receptor - Ligand Interactions .....	6
2.3. Dimerization of GPCR .....	7
2.4. Class A Family GPCRs .....	7
2.4.1. The $\beta_2$ Adrenergic Receptor .....	9
2.4.2. Overall Topology of The Human $\beta_2$ AR .....	11
2.4.3. Signaling Mechanism of The Human $\beta_2$ AR .....	12
2.4.4. Activation of The Human $\beta_2$ AR (Ligand Interaction) .....	13
3. THE MODELS AND THE METHODS .....	17
3.1. System Setup for Molecular Dynamics .....	17
3.1.1. Homology Modeling (MODWEB) .....	18
3.1.2. Generating Model Structures .....	20
3.1.3. Determination of Protonation States .....	23
3.1.4. Determination of the Orientation of Protein in Membrane .....	24
3.1.5. System Preparation for Molecular Dynamics (MD) Simulations .....	26

3.2. Theoretical Background of MD Simulations .....	29
3.2.1. Force Fields .....	29
3.2.2. Initialization of the System .....	31
3.3. Running of MD Simulation .....	32
3.3.1. Melting of Lipid Tails .....	33
3.3.2. Minimization and Equilibration with Protein Constrained .....	33
3.3.3. Equilibration with Protein Released .....	34
3.3.4. Production Runs .....	35
3.4. Trajectory Analysis .....	36
3.4.1. Root Mean Square Deviation (RMSD) .....	36
3.4.2. Mean Square Fluctuation (MSF) .....	36
3.4.3. Principal Component Analysis (PCA) .....	37
3.4.4. Clustering .....	37
3.5. Anisotropic Network Model (ANM) .....	38
4. RESULTS AND DISCUSSION .....	41
4.1. Construction of the Models .....	41
4.2. Anisotropic Network Model (ANM) for Monomer Models .....	42
4.2.1. B-Factor Profiles .....	42
4.2.2. Collective Modes .....	44
4.2.3. Overlap Matrix .....	46
4.2.4. Orientational Cross-Correlations .....	51
4.2.5. Fast Modes (Hot Spots) .....	55
4.3. Anisotropic Network Model (ANM) for Dimer Models .....	57
4.3.1. Generating The Dimer Structure of $\beta_2$ AR .....	57
4.3.2. B-Factor Profiles .....	58

4.3.3. Slow Modes ( Collective Deformation) .....	59
4.3.4. Overlap Matrix .....	60
4.4. Molecular Dynamics (MD) Simulations .....	64
4.4.1. Simulation Details .....	64
4.4.2. Root Mean Square Distance (RMSD) .....	65
4.4.3. Root Mean Square Fluctuations (RMSF) .....	71
4.4.4. Ionic Lock .....	74
4.5. Principal Component Analysis (PCA) .....	81
4.5.1. Slow (Collective) Modes .....	83
4.5.2. Overlap Calculations .....	88
4.5.2.1. Overlap between Looped and Clipped Models .....	89
4.5.2.2. Overlap between Clipped (Natural, Mutated) Models .....	91
4.5.3. Cross Correlations .....	93
4.5.3.1. Difference of Orientational Cross-Correlation between Looped and Clipped Models .....	93
4.5.3.2. Difference of Orientational Cross-Correlation between two Clipped Models .....	97
4.6. Clustering .....	101
4.6.1. Clustering of Overall MD Trajectory .....	101
4.6.2. Clustering of Looped Model .....	108
4.6.3. Clustering of Clipped (Natural) and Clipped (Mutated) Models .....	110
5. CONCLUSIONS AND FUTURE STUDIES .....	119
5.1. Conclusion .....	119
5.2. Future Studies .....	123
APPENDIX A: CROSS-CORRELATION TABLES FOR BOTH MODELS .....	124
REFERENCES .....	127

## LIST OF FIGURES

Figure 2.1.	Schematic representation of a GPCR with a trademark of 7-TM helices that are embedded inside of the cell membrane. ....	3
Figure 2.2.	Schematic diagram of GPCRs for (a) two-dimensional model of bovine rhodopsin, (b) cartoon view of rhodopsin, (c) two-dimensional description of $\beta_2$ AR (Palczewski <i>et al.</i> , 2000; Teller <i>et al.</i> , 2001; Rasmussen <i>et al.</i> , 2007). ....	4
Figure 2.3.	Some of the important signalling pathways controlled by GPCRs. The $\alpha$ -unit of the trimeric G-proteins initiates different pathways. ....	5
Figure 2.4.	Efficiency of different ligands (Rosenbaum <i>et al.</i> , 2009). ....	6
Figure 2.5.	Superimposed crystal structures of bovine rhodopsin and $\beta_2$ AR. Bovine rhodopsin is shown in blue and $\beta_2$ AR is shown in purple. ....	8
Figure 2.6.	Classification of adrenergic receptors. ARs can be broadly classified into two major categories for alpha ( $\alpha$ ) and beta ( $\beta$ ) adrenergic receptor. ....	10
Figure 2.7.	The $\beta_2$ AR-T4L. The receptor and T4-L are colored dark gray and green. ....	11
Figure 2.8.	Intracellular signaling pathway of the $\beta_2$ AR (Johnson, 2006). ....	13
Figure 2.9.	Ionic lock formation between Asp130-Arg131 and Arg131-Glu286 in $\beta_2$ AR. ....	15
Figure 2.10	Ionic lock formation between Asp192-Lys305 in $\beta_2$ AR. ....	16

Figure 3.1.	Steps in comparative protein structure modeling (March <i>et al.</i> , 2000). .	18
Figure 3.2.	The flowchart of the ModPipe (Eswar <i>et al.</i> , 2006). .....	19
Figure 3.3.	Initial sequence that includes overall missing residues of $\beta_2$ AR except the sequence of T4L. The first 28 missing residues of $\beta_2$ AR (red), the missing sequence of ICL3 between 231 <sup>st</sup> to 262 <sup>nd</sup> residues (green) and the missing sequence between 343 <sup>th</sup> to 365 <sup>th</sup> residues (blue). .....	21
Figure 3.4.	The conformation of $\beta_2$ AR with estimated ICL3 region. The missing ICL3 region between helices V and VI is shown in purple. ....	21
Figure 3.5.	Estimated ICL3 model obtained from MD study (Dror <i>et al.</i> , 2008). ....	22
Figure 3.6.	Clipped model of $\beta_2$ AR. Peptide bond is in between the C and N atoms belonging to Leu230 and Lys263 residues and is highlighted with green and orange spheres in turn. ....	22
Figure 3.7.	The schematic overview of the interaction between His241 and Asn244. Dotted line represents the bond between N $\delta$ atom included in His241 and backbone nitrogen included in Asn244. ....	24
Figure 3.8.	The inclined structure of $\beta_2$ AR that has been derived by its alignment with the model obtained from OPM. ....	25
Figure 3.9.	Water molecules, which are located in inner regions, are shown here as part of X-ray crystallographic experiments conducted for $\beta_2$ AR. ....	26
Figure 3.10.	The position of the crystallographic structure of the $\beta_2$ AR in the cell membrane depicted as an orange cartoon. ....	27

Figure 3.11.	The systems consisting of $\beta$ 2AR, cell membrane are shown along with its periodic image in the direction of x. The polar heads of the lipid and water, protein and lipid hydrocarbon ends are illustrated are in blue, yellow, magenta and pink respectively. ....	28
Figure 3.12.	Interactions included in representative potential energy function for MD simulations (Steinbach, 2010). ....	30
Figure 3.13.	Representations of the cell membrane, (a) before the melting of lipid tails, (b) after the melting of lipid tails. ....	33
Figure 3.14.	Representation of the xy plane of the cell membrane, (a) before the third preparation stage, (b) after equilibration. ....	34
Figure 3.15.	Change of surface area calculated per lipid through MD simulations. ...	35
Figure 4.1.	Structures of (a) clipped (green), and (b) looped (blue) models. ....	41
Figure 4.2.	Theoretical and experimental B-factor values for clipped and looped models. ....	43
Figure 4.3.	Most mobile loop regions of the structure, (a) surface representation, (b) cartoon representation. ....	43
Figure 4.4.	All transmembrane helices and loop regions of the looped model. ....	44
Figure 4.5.	Mean square fluctuations based on the cumulative action of 10 low-frequency modes for looped and clipped models. ....	45
Figure 4.6.	MSF profile is calculated by taking between the fifth and fourteenth slowest modes for looped model and the first 10 slowest modes for clipped model. ....	45

Figure 4.7.	Overlap between low-frequency modes of the clipped and looped models. ....	46
Figure 4.8.	The directions and magnitudes of eigenvectors of the models for (a) the first slow mode of clipped model, (b) the fifth slow mode of looped model. ....	48
Figure 4.9.	The directions and magnitudes of eigenvectors of the models for (a) the third slow mode of clipped model, (b) the eighth slow mode of looped model. ....	49
Figure 4.10.	The directions and magnitudes of eigenvectors of the low-frequency modes of looped model for (a) the first mode, (b) the second mode. ....	50
Figure 4.11.	The directions and magnitudes of eigenvectors of the low-frequency modes of looped model for (a) the third mode, (b) the fourth mode. ....	51
Figure 4.12.	Orientational cross-correlation map for (a) the first ten modes of clipped model, (b) the first ten modes of looped model. ....	53
Figure 4.13.	Orientational cross-correlation map based on modes five through fourteen for looped model. ....	54
Figure 4.14.	Difference cross-correlation plot of clipped and looped models. ....	54
Figure 4.15.	Normalized MSF profile is calculated by taking the 20 high-frequency modes of the models for (a) Asp130, Arg131, Glu268 and Phe282 (b) Leu272, Gly276, Gly280 and Leu284 residues are shown on the profile. ....	56
Figure 4.16.	The position of the hot spot residues on the looped model. ....	56

Figure 4.17.	The best symmetric mate of the monomer with cholesterol (yellow) and palmitic acid (pink) for looped model: (a) Front view, (b) ICL view and (c) ECL view. ....	57
Figure 4.18.	Ionic interaction between the charged amine group of Lys60 in helix I and the carboxylate of Glu338 in the helix VIII of the symmetry-related receptor. ....	58
Figure 4.19.	Comparison of B-factor values of two dimer models. ....	59
Figure 4.20.	Cumulative 10 low-frequency modes for the dimer models. ....	60
Figure 4.21.	Most flexible regions of both dimers for (a) clipped dimer model, (b) looped dimer model. ....	60
Figure 4.22.	Overlap matrix of the two dimer models. ....	61
Figure 4.23.	Eigenvector representation of the first modes for (a) clipped and (b) looped models. ....	61
Figure 4.24.	Eigenvector representation of the second modes for (a) clipped and (b) looped models. ....	62
Figure 4.25.	Overlap matrix between the monomer and dimer structures for the looped model. ....	63
Figure 4.26.	Overlap matrix between the monomer and dimer structures of the clipped model. ....	63
Figure 4.27.	RMSD for looped model from the initial minimized structure. ....	66
Figure 4.28.	RMSD of the snapshots from the initial energy minimized structures for all runs of (a) the core regions, (b) the transmembrane regions. ....	66

Figure 4.29.	Regions considered in RMSD calculations for (a) core region includes all the loop regions except ICL3, (b) transmembrane region excludes all loop regions for both models. ....	67
Figure 4.30.	The alteration of the secondary structure along the simulation of the looped model. ....	68
Figure 4.31.	The structural alterations of certain loop regions based on initial (green) and final (purple) conformations of the looped simulation: (a) Changes in ICL3, (b) changes in ECL2. ....	68
Figure 4.32.	The alteration of the secondary structure along the simulation for (a) clipped model (natural), (b) clipped model (mutated). ....	69
Figure 4.33.	The pivot angle between the axis of two half helix. ....	70
Figure 4.34.	RMSF about the average position plotted for $C\alpha$ atoms of residues for (a) MD result, (b) ANM result for both models. ....	71
Figure 4.35.	The position of the most mobile regions on the structure for (a) loop, (b) TM4 and TM5. ....	72
Figure 4.36.	RMSD profiles of (a) TM4, (b) TM5. ....	73
Figure 4.37.	The N-O (black) and $C\alpha$ - $C\alpha$ (red) distances (N-O reflecting the open/closed states of the salt bridge between Asp130 and Arg131 along the simulations for (a) clipped (Natural), (b) clipped (Mutated), and (c) looped models. ....	76
Figure 4.38.	The N-O (black) and $C\alpha$ - $C\alpha$ (red) distances (N-O reflecting the open/closed states of the salt bridge between Arg131 and Glu268 along the simulations for (a) clipped (Natural), (b) clipped (Mutated), and (c) looped models. ....	77

Figure 4.39.	The N-O (black) and C $\alpha$ -C $\alpha$ (red) distances (N-O reflecting the open/closed states of the salt bridge between Asp192 and Lys305 along the simulations for (a) clipped (Natural), (b) clipped (Mutated), and (c) looped models. ....	78
Figure 4.40.	The change, along the simulation of looped, in main chain $\phi$ (phi) and $\psi$ (psi) angles and in the side chain $\chi_1$ (chi) angle for Arg131, which form the salt bridge for (a) $\phi$ and $\psi$ , (b) $\chi_1$ . ....	79
Figure 4.41.	The change, along the simulation of looped, in main chain $\phi$ (phi) and $\psi$ (psi) angles and in the side chain $\chi_1$ (chi) angle for Glu268, which form the salt bridge for (a) $\phi$ and $\psi$ , (b) $\chi_1$ . ....	80
Figure 4.42.	The change, along the simulation of clipped (natural), in main chain $\phi$ (phi) and $\psi$ (psi) angles and in the side chain $\chi_1$ (chi) angle for Glu268, which form the salt bridge for (a) $\phi$ and $\psi$ , (b) $\chi_1$ . ....	81
Figure 4.43.	The graphical representation of variance percentages of the four models for (a) individual variance, (b) cumulative variance. ....	82
Figure 4.44.	Collective motions on the first and second PC of looped and clipped models. ....	85
Figure 4.45.	The percentage of residue-based motion explained by PC1 and cumulative two PC for each models for (a) looped, (b) clipped (natural), (c) clipped (mutated) models. ....	86
Figure 4.46.	The residue-based explanation values for (a) the first PC, (b) cumulative 2 PC of looped and cumulative 11 PC of each clipped model. ....	87
Figure 4.47.	Overlap matrix of the first 20 modes between (a) looped and clipped (natural), (b) looped and clipped (Mutated). ....	90

Figure 4.48.	Collective motions on PC for (a) the sixth PC of looped, (b) the third PC of clipped (natural), (c) the second PC of clipped (Mutated). .....	91
Figure 4.49.	Overlap matrix between clipped (natural) and clipped (mutated) models. ....	92
Figure 4.50.	Collective motion on PC for (a) the fifth principal mode of clipped (natural), (b) the fourth principal mode of clipped (mutated), (c) the third principal mode of clipped (natural), (d) the seventh principal mode of clipped (mutated). ....	93
Figure 4.51.	Difference cross-correlation for (a) looped and clipped (natural) models and between (b) looped and clipped (mutated) model. ....	94
Figure 4.52.	Difference of the cross correlation between Figure 4.51 (a) and (b). ....	95
Figure 4.53.	Orientalional cross-correlation map of cumulative first 11 principal modes for (a) clipped (natural) model, (b) clipped (mutated) model. ....	98
Figure 4.54.	Difference of the cross correlation between Figures 4.53 (a) and (b). ...	99
Figure 4.55.	Clustering of the all simulations according to binding-site region using a RMSD threshold of 2.15 Å. ....	102
Figure 4.56.	The best representative members of the clusters are aligned on crystal structure. (a) Crystal structure (blue) with carazolol (yellow) encircled; Representative from (b) Cluster1 (pink); (c) Cluster2 (green); (d) Cluster3 (orange); (e) Cluster4 (gray). ....	103
Figure 4.57.	Clustering of the all simulations according to transmembrane region using a RMSD threshold of 1.5 Å. ....	104

Figure 4.58.	Clustering of the all simulations according to overall structure excluding ICL3 using a RMSD threshold of 3.6 Å. ....	105
Figure 4.59.	The representative members of the clusters of overall structure: (a) Blue: Crystal structure; yellow: Carazolol; (b) red: Cluster1; (c) green: Cluster2; (c) pink: Cluster3; (e) orange: Cluster4. ....	107
Figure 4.60.	The distribution of cluster for the looped model according to frame number obtained from clustering based on (a) binding-site region and (b) transmembrane region. ....	108
Figure 4.61.	The best representative members of the clusters of binding-site region are aligned on the crystal structure. (a) Blue: Crystal structure; yellow: Carazolol; (b) green: Cluster1; (c) red: Cluster2; (d) pink: Cluster3; (e) orange: Cluster4. ....	109
Figure 4.62.	The distribution of cluster of clipped (natural) model according to frame number for (a) binding-site region (RMSD threshold 1.6 Å) and (b) transmembrane region (RMSD threshold 1.51 Å). ....	111
Figure 4.63.	The best representative members of the clusters of binding-site region of clipped (natural) are aligned on the crystal structure. (a) Blue: Crystal structure; yellow: Carazolol; (b) green: Cluster1; (c) red: Cluster2; (d) pink: Cluster3; (e) orange: Cluster4. ....	112
Figure 4.64.	The best members of the clusters of transmembrane region of clipped (natural) are aligned on the crystal structure. (a) Blue: Crystal structure; yellow: Carazolol; (b) red: Cluster1; (c) green: Cluster2; (c) pink: Cluster3; (e) orange: Cluster4. ....	113
Figure 4.65.	The distribution of cluster of clipped (mutated) model according to frame number for (a) binding-site region (RMSD threshold 1.75 Å), (b) transmembrane region (RMSD threshold 1.58 Å). ....	115

- Figure 4.66. The best representative members of the clusters of binding-site region of clipped (mutated) are aligned on the crystal structure. (a) Blue: Crystal structure; yellow: Carazolol; (b) green: Cluster1; (c) red: Cluster2; (d) pink: Cluster3; (e) orange: Cluster4. .... 116
- Figure 4.67. The best members of the clusters of transmembrane region of clipped (mutated) are aligned on the crystal structure: (a) Blue: Crystal structure; yellow: Carazolol; (b) red: Cluster1; (c) green: Cluster2; (c) pink: Cluster3; (e) orange: Cluster4. .... 117

## LIST OF TABLES

Table 2.1.	Comparison of characteristic between $\beta_2$ AR and bovine rhodopsin. ....	9
Table 3.1.	The dimensions of the protein, cell membrane and system for each model. ....	28
Table 3.2.	Total number of atoms, lipid and water molecules at both models. ....	36
Table 4.1.	Simulation system details. ....	64
Table 4.2.	The average values of pivot angles of transmembrane helices for all models. ....	70
Table 4.3.	The percentage of time during which the salt bridge is in the closed state (N-O distance is less than 3.5 Å). ....	75
Table 4.4.	Percentage of the total motion explained by the first five principal components of each model. ....	83
Table 4.5.	Percentage of the total motion explained by the first two PC of each model. ....	84
Table 4.6.	The cross-correlation values of cumulative two PCs of looped and eleven PCs of clipped (natural) models. ....	96
Table 4.7.	The cross-correlation values of cumulative two PCs of looped and eleven PCs of clipped (mutated) models. ....	97
Table 4.8.	Cross-correlation values of cumulative eleven PCs of both clipped models. ....	100

Table 4.9.	The number of matching frames between binding-site and transmembrane-based clustering. ....	105
Table 4.10.	The number of matching frames between loop region, binding-site and transmembrane-based clustering. ....	106
Table 4.11.	The number of matching frames between binding-site and transmembrane-based clustering for looped model. ....	110
Table 4.12.	The number of matching frames between binding-site and transmembrane-based clustering for clipped (natural) model. ....	114
Table 4.13.	The number of matching frames between binding-site and transmembrane-based clustering for clipped (mutated) model. ....	118
Table A.1.	The cross correlation values of cumulative two PCs of looped model. .	124
Table A.2.	The cross correlation values of cumulative eleven PCs of clipped (natural). ....	125
Table A.3.	The cross-correlation values of cumulative eleven PCs of clipped (mutated). ....	126

## LIST OF SYMBOLS

$a$	Acceleration of an atom
$a_i$	Effective Born radii
$C\alpha$	Alpha-carbon
$C(i,j)$	Orientalional cross-correlations
$C\beta$	Beta-carbon
$F_i$	Force on particle $i$
$h(x)$	Heavy side step function
$H$	Hessian matrix
$k$	Mode
$k_B$	Boltzman constant
$k_i$	Force constant that describes the stiffness of the specific bond
$l_i$	Instantaneous bond length
$l_{i,0}$	Equilibrium length of a bond
$m_i$	Mass of particle $i$
$N$	Number of Particles
$q_i, q_j$	Charges of atoms
$Q$	$3N \times 3N$ matrix of left singular vectors
$r_c$	Cutoff distance
$r_{ij}$	The distance between partial charges
$R_i$	Position vector of residue $i$
$S$	The diagonal matrix of the $3N$ singular values
$T$	Temperature
$T$	Time
$u_k$	Eigenvector
$U(R_1, \dots, R_N)$	Potential energy function
$v_i$	Velocity of an atom $i$
$V$	Potential energy as a function form
$V$	The matrix of right singular vectors
$W$	Instantaneous torsion angle
$x_i$	The coordinate at which the particle is moving

$\Delta G^{GB}_{elec}$	Generalized Born Solvation energy
$\Delta R$	Matrix of instantaneous fluctuations of $\Delta R_i$
$\Delta R_i$	Displacement vector for $R_i$
$\Delta R_{ij}$	Fluctuation in the distance vector
$\alpha$	Alpha helix
$\beta$	Beta strand
$\gamma$	Phase factor
$\Gamma$	Uniform Hookean force constant
$\varepsilon$	The solvent dielectric constant
$\theta_t$	Instantaneous angle
$\theta_{i,0}$	Equilibrium angle
$\Theta$	Angle
$\sigma_{ij}$	Lennard-Jones collision parameter
$\Phi$	Phi
$\chi$	Chi
$\Psi$	Psi

**LIST OF ACRONYMS/ABBREVIATIONS**

3D	Three-dimensional
Å	Angstrom
ANM	Anisotropic Network Model
Arg	Arginine
Asn	Asparagine
Asp	Aspartic Acid
β <sub>2</sub> AR	Human Beta-2 Adrenergic Receptor
cAMP	Cyclic Adenosine Monophosphate
cSrc	Cellular Sarcoma
ECL2	Extracellular Loop Two
ECL3	Extracellular Loop Three
EN	Elastic Network
ICL3	Intracellular Loop Three
Glu	Glutamic Acid
GNM	Gaussian Network Model
GPCR	G-protein Coupled Receptor
Lys	Lysine
MAPK	Mitogen-activated Protein Kinase
MD	Molecular Dynamics
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
Raf	Proto-oncogene Serine / Threonine-protein Kinase
RAS	RAt Sarcoma
RMSD	Root Mean Square Deviation
SOS	Son of Sevenless

## 1. INTRODUCTION

G-protein-coupled receptors (GPCRs) represent the largest family of membrane proteins in the human genome, with members grouped into five classes based on sequence and functional similarity (Fredriksson *et al.*, 2003). Structurally, all GPCRs are characterized by the presence of seven membrane-spanning  $\alpha$ -helical segments with an extracellular N terminus and an intracellular C terminus. These receptors couple the binding of agonists to heterotrimeric G-protein activation thereby playing tremendously significant roles in cellular responses to hormones, neurotransmitters, and other molecules involved in a wide variety of physiological processes (Oldham *et al.*, 2008). Therefore, GPCRs, being main contributors to the information flow into cells, are associated with a plenty of diseases that make members of this family significant pharmacological targets (Cherezov *et al.*, 2007).

One of the major challenges in drug development for GPCRs is to design subtype specific drugs. Since GPCRs of one particular function have many subtypes, design of subtype specific drugs calls for structural information on the target GPCRs. Unfortunately there is very little structural information on GPCRs. In fact, bovine rhodopsin is one of the first experimental 3-D structures of GPCRs (Palczewski *et al.*, 2000; Teller *et al.*, 2001). Unlike rhodopsin, however, determination of most GPCRs' functions is very difficult and many studies about GPCRs have been dedicated to investigate functionally important parts of the structures (Romo *et al.*, 2010; Kobilka *et al.*, 2007; Rosenbaum *et al.*, 2009).

Although rhodopsin is one of the first experimental three dimensional structures of GPCRs; many unknown aspects remain on the conformational changes between different activation states for each receptor. There are many structural differences, which depend on binding the very large diversity of ligands, between receptors. To overcome these unknown aspects and questions about the differentiation of activation states for receptors, the human  $\beta_2$ -adrenergic receptor ( $\beta_2$ AR) was modified with inserting T4-lysozyme (T4L) in place of the third intracellular loop ( $\beta_2$ AR-T4L) and solved the three-dimensional structure in the presence of a partial inverse agonist carazolol (2-propanol, 1-(9H-carbazol-4-yloxy)-3-[(1-methylethyl) amino]) at 2.4 Å resolution (Cherezov *et al.*, 2007).

The  $\beta_2$ AR is a GPCR activated by adrenaline that plays substantial parts in cardiovascular and pulmonary physiology and it is one of the most widespread characterized members of the family of membrane protein. All family of  $\beta$ -ARs among GPCRs is one of the most significant therapeutic targets such that they comprise 50% of drugs in the market. Genetic modifications of adrenergic receptors are generally associated with asthma, hypertension, and heart failure. Also,  $\beta_2$ AR, which is the member of the subtype of  $\beta$ -AR family, locate largely in smooth muscle throughout the body, and  $\beta_2$ AR agonists are used in the treatment of asthma and preterm labor (Rasmussen *et al.*, 2007).

In this thesis, the crystal structure of the human  $\beta_2$ AR complexed with a partial inverse agonist carazolol was used as a starting conformation (PDB ID: 2rh1). The missing intracellular loop (ICL3), which plays an essential role in G protein recognition, was modeled via MODBASE web-server in order to obtain appropriate homologues of the structure of  $\beta_2$ AR (Pieper *et al.*, 2009). An alternative model of the receptor with missing loop, which called as “clipped” model, was also used in order to understand the effect of the ICL3 loop on the dynamics by a comparison of the looped and clipped models. The purpose of this thesis is to explore the conformational dynamics of the apo- $\beta_2$ AR by the coarse-grained anisotropic network model (ANM) (Atilgan *et al.*, 2001) and by atomistic molecular dynamics (MD) simulations.

## 2. THEORETICAL BACKGROUND

GPCRs are referred to seven transmembrane (7TM) receptors and comprise the largest family of membrane proteins. Seven membrane-spanning  $\alpha$ -helical segments with an extracellular N terminus and an intracellular C terminus are shown in Figure 2.1. These receptors help to transmit an extracellular signal into intracellular part of the cell. They are classified into five main subfamilies on the basis of sequence homology and/or pharmacological characteristics (Fredriksson *et al.*, 2003; Kristiansen, 2004). The main subfamilies include in the rhodopsin-like receptors (family A), the secretin-like receptors (family B) and the glutamate receptor-like GPCRs (family C) (Pierce *et al.*, 2002). Other two subfamilies are the minor subfamilies of 7TM (Gether, 2000; Horn *et al.*, 1998).

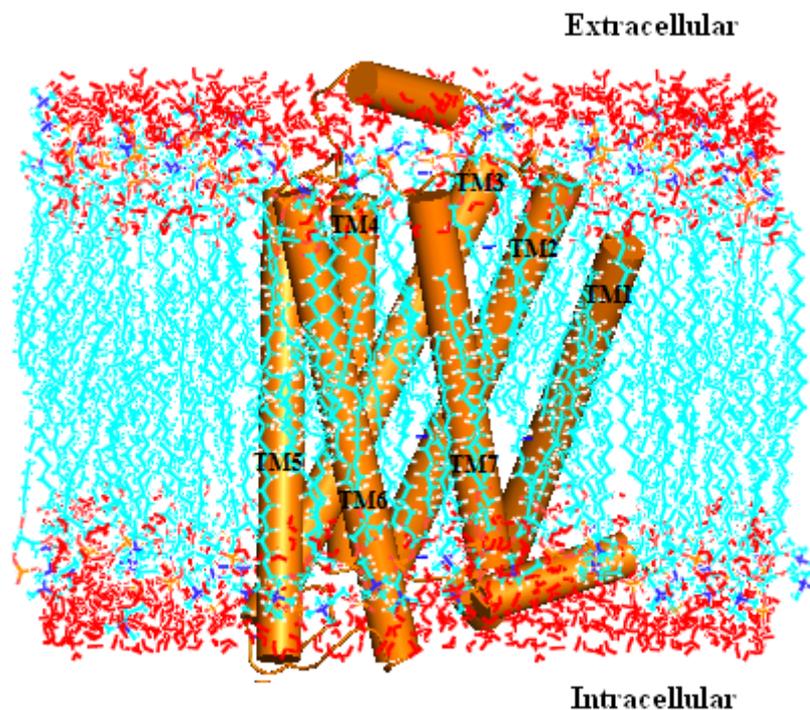


Figure 2.1. Schematic representation of a GPCR with a trademark of 7-TM helices that are embedded inside of the cell membrane.

The first GPCR that was structurally elucidated is bovine rhodopsin (Palczewski *et al.*, 2000), the light-activated photoreceptor found in the mammalian eye. Until recently, bovine rhodopsin has been the only member of GPCR family that has crystal structural

information available and served as a template for other GPCRs (Rasmussen *et al.*, 2007). Despite the remarkable structural diversity of GPCRs, all GPCRs share a common molecular architecture consisting of seven transmembrane helices connected by three intracellular and three extracellular loops with an extracellular N terminus and an intracellular C terminus (Figure 2.2).

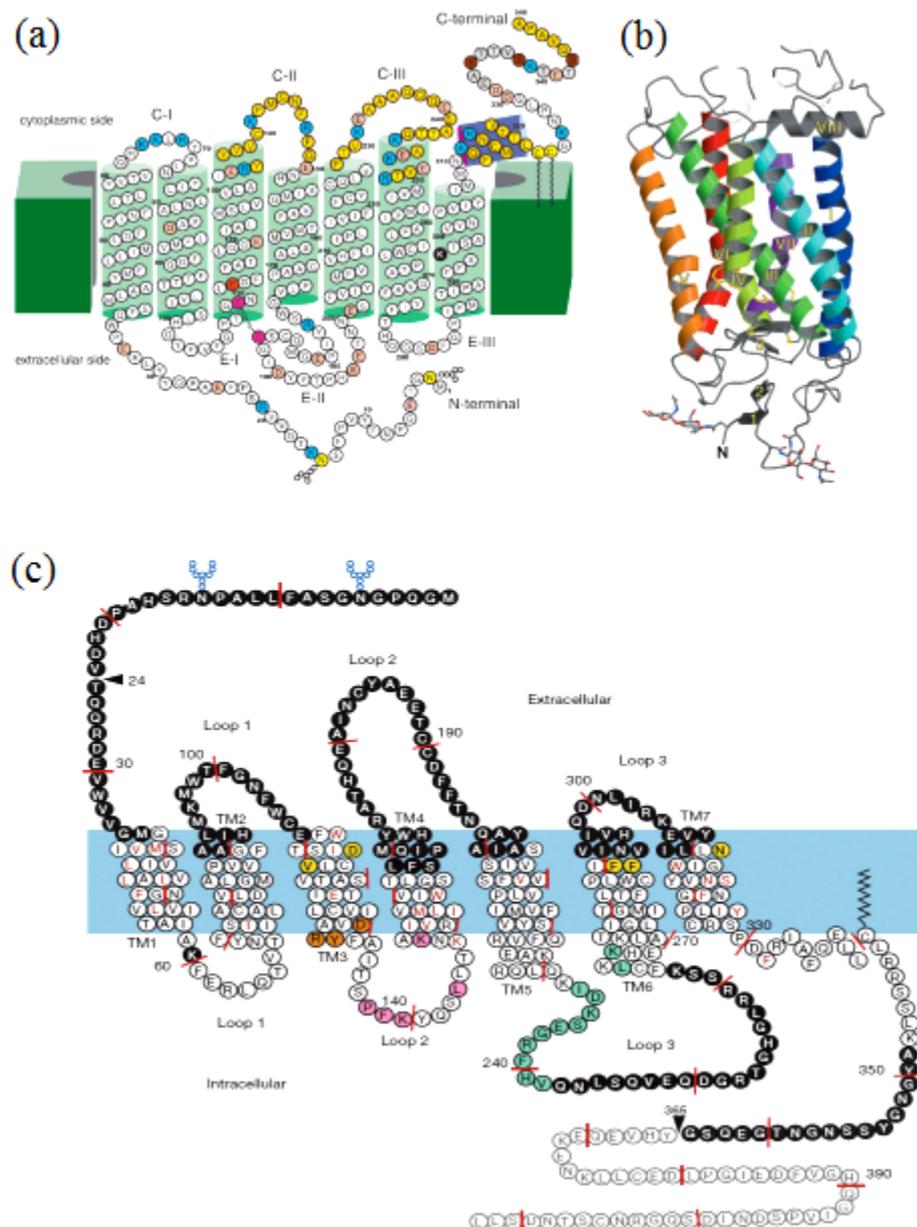


Figure 2.2. Schematic diagram of GPCRs for (a) two-dimensional model of bovine rhodopsin, (b) cartoon view of rhodopsin, (c) two-dimensional description of  $\beta_2$ AR (Palczewski *et al.*, 2000; Teller *et al.*, 2001; Rasmussen *et al.*, 2007).

## 2.1. G-protein Mediated Signal Transduction Pathway

GPCRs play critical roles in transducing extracellular signals to inside the cell. It is important to identify the GPCRs involved in human health and target them for therapeutic intervention. When a GPCR is activated by an agonist, it triggers a cascade of responses inside the cell, particularly through interactions with heterotrimeric G-protein regulators that are switched on and off. Heterotrimeric G-proteins are involved in a second messenger cascade and are made up of alpha ( $\alpha$ ), beta ( $\beta$ ), and gamma ( $\gamma$ ) subunits. These G-proteins switch on and switch off the cell signaling initiated by GPCRs by alternating between an inactive guanosine diphosphate (GDP) and an active guanosine triphosphate (GTP) bound state, ultimately regulating downstream cellular processes (Figure 2.3).

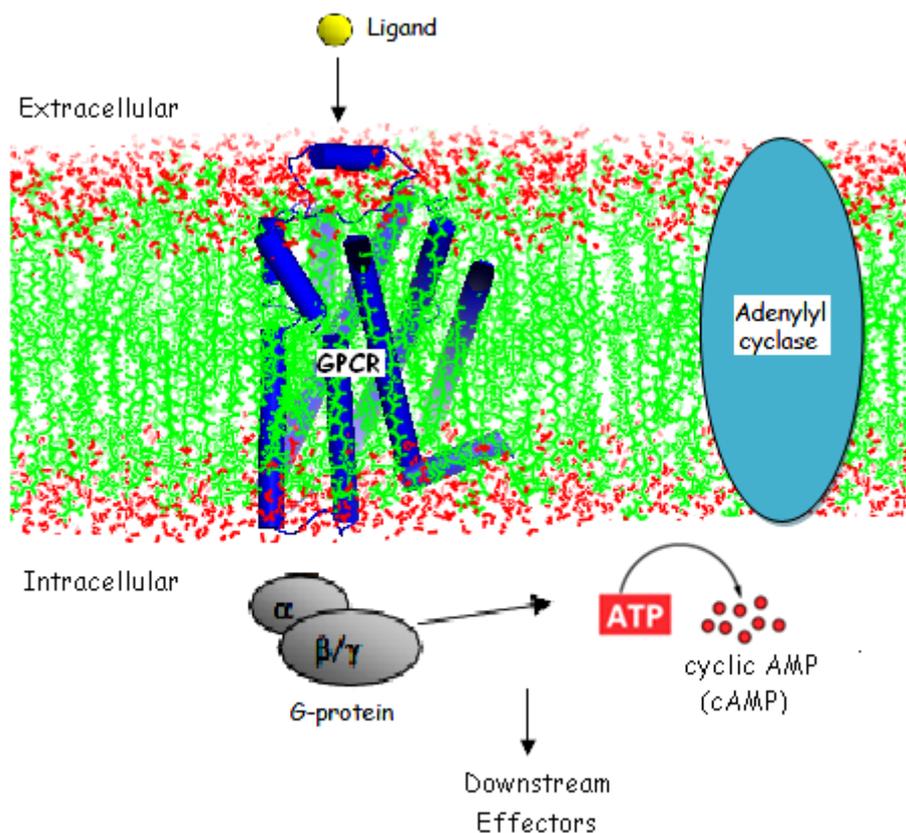


Figure 2.3. Some of the important signalling pathways controlled by GPCRs. The  $\alpha$ -unit of the trimeric G-proteins initiates different pathways.

## 2.2. Receptor - Ligand Interactions

Many therapeutic agents such as activating (agonist) or blocking (antagonist) GPCRs are used to get information about the biochemical events underlying cellular signalling. However, the understanding of the molecular interactions between ligands and receptor protein and the effect of agonists and inverse agonists on the receptor activation is still immature. Hence, most of work in this area has focused on understanding the receptor responsible for drug binding affinity. Although the first action of drug design is to examine the binding of ligand molecules to specific receptors, the efficacy of binding remains the only determinant of therapeutic utility (Hebert *et al.*, 1998).

Receptor activation by an agonist and partial agonist is believed to cause a conformational change in the receptor's three-dimensional structure, shifting the equilibrium towards a more active conformation and thus lead to an increased signaling activity. This structural change is responsible for activating the G protein. G protein activation is characterized by an exchange of GDP for GTP on the alpha subunit and subsequent dissociation of the G protein from the receptor (Strader *et al.*, 1994). In contrast to this, inverse agonists change the conformation to a more inactive state, hence inhibiting the basal activity (Figure 2.4). Antagonists do not affect the baseline level of activity of the receptor, but sterically block the binding/activation of other ligands (Kobilka & Deupi, 2007; Rosenbaum *et al.*, 2009)).

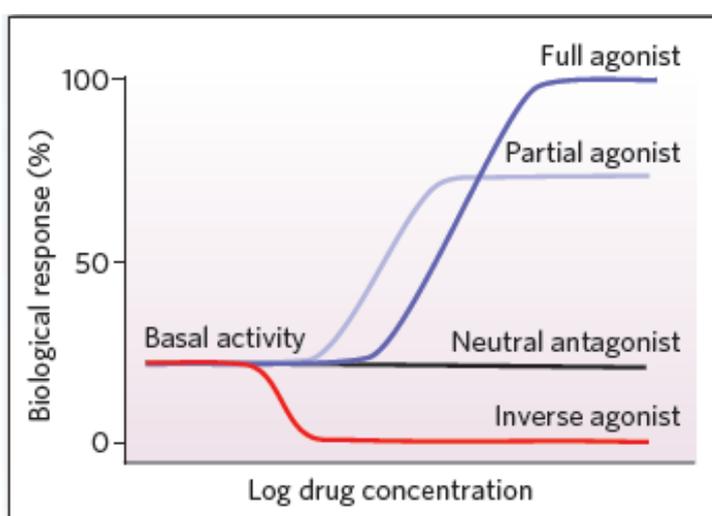


Figure 2.4. Efficiency of different ligands (Rosenbaum *et al.*, 2009).

### 2.3. Dimerization of GPCR

Many proteins, which belong to the family of GPCR, have also been observed in dimeric states experimentally. According to some dimerization or oligomerization studies, dimerization can affect different stages of a receptor's life and regulate receptor activity and lead to increase the functional properties of proteins (Nikbin *et al.*, 2003). The interaction region between monomers is still unknown. Some dimeric structures, which are obtained from crystallographic studies, do not reflect the real native state, but simply the result of crystallographic conditions.

The crystal structure of  $\beta_2$ AR showed symmetry-related dimers (Cherezov *et al.*, 2007). The interaction between the charged amine group of Lys60 in TM1 (transmembrane helix 1) from one monomer and the carboxylate of Glu338 in TM8 (transmembrane helix 8) from the second monomer was critical in the formation of the dimeric structure. TM8 is a cytoplasmic helix formed in the membrane-proximal part of the tail and it is perpendicular to the TM bundle. This is thought to be a common feature in all rhodopsin-like GPCRs (Katragadda *et al.*, 2004). However, the dimerization interface of GPCRs becomes a highly controversial issue, since some studies suggested different localization, such as TM6 as the dimer interface for the  $\beta_2$ AR (Hebert *et al.*, 1996).

The principle of receptor dimerization or oligomerization has important implications for drug design studies. Furthermore, the discovery of differential or ligand biased signal transduction from receptors involved in dimeric complexes might also be able to explain some unexpected side effects of currently marketed drugs. The molecular mechanism and functional consequences of the dimerization of GPCRs might shed light on the new therapeutic applications and guide novel ways in drug design (Maggio *et al.*, 2005; Prinster *et al.*, 2005).

### 2.4. Class A Family GPCRs

GPCRs have been categorized into five classes according to their sequence conservation with class A being the largest and most diverse subfamily of G-protein coupled receptors. Furthermore, class A receptors are classified into groups according to their specific ligand specificities (Cherezov *et al.*, 2007).

Class A family is formed by the rhodopsin like receptors. The basic structural features of these receptors were identified with the use of two dimensional models of rhodopsin (Schertler *et al.*, 1993). However, significant improvements were made by studying the three dimensional crystal structure of rhodopsin (Palczewski *et al.*, 2000) and the  $\beta_2$ AR (Rasmussen *et al.*, 2007), which revealed a cytoplasmic eighth helix. Additionally, these studies showed that some motifs among family A receptors, such as the E/DRY motif on the transmembrane domain 3 as well as the NPxxYx motif on the transmembrane domain 7 are highly conserved (Fritze *et al.*, 2003; Gether, 2000). The highest variability was found in the carboxyl terminus, the amino terminus and the intracellular loop spanning transmembrane domain 5 and 6. Comparison of the structures of rhodopsin and  $\beta_2$ AR showed that there are substantial alterations in their TM helices, which might account for their different activation states. A comparison of the crystal structures of the two proteins (Figure 2.5) shows that the  $\beta_2$ AR has a longer intracellular loop and a longer C-terminus than that of rhodopsin.

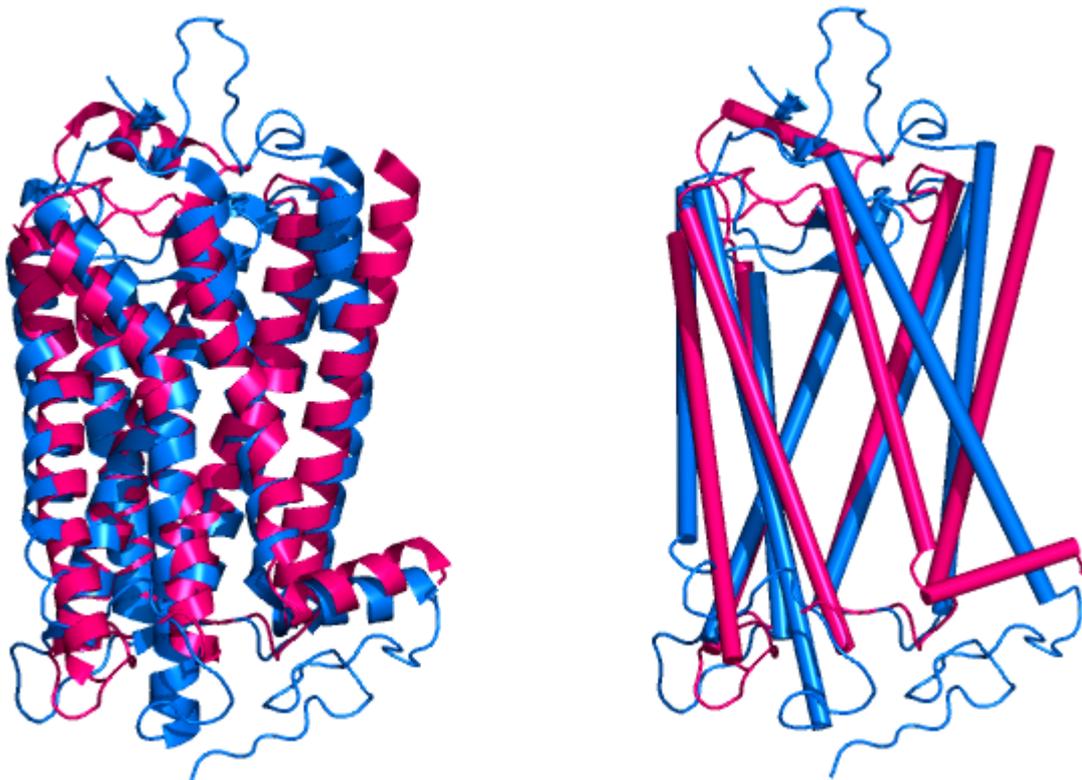


Figure 2.5. Superimposed crystal structures of bovine rhodopsin and  $\beta_2$ AR. Bovine rhodopsin is shown in blue and  $\beta_2$ AR is shown in purple.

Furthermore, it was observed that both receptors show major differences on their extracellular termini (Rasmussen *et al.*, 2007; Cherezov *et al.*, 2007; Rosenbaum *et al.*, 2007). The  $\beta_2$ AR exhibits a high degree of structure similarity to bovine rhodopsin. Table 2.1 summarized the comparison between rhodopsin and  $\beta_2$ AR.

Table 2.1. Comparison of characteristic between  $\beta_2$ AR and bovine rhodopsin.

Characteristic	Beta-2 Adrenergic Receptor	Bovine Rhodopsin
1 <sup>st</sup> messenger	Norepinephrine	Retinal
G-protein coupled receptor	Beta-2 adrenergic receptor	Opsin
G-protein	Gs protein	Transducin
Activated Enzyme	Adenylyl cyclase	Phosphodiesterase
2 <sup>nd</sup> messenger	Cyclic AMP	Cyclic GMP
Ion channel is permeable to	K <sup>+</sup>	Na <sup>+</sup> , K <sup>+</sup> , Ca <sup>2+</sup>

Nevertheless, rhodopsin and  $\beta_2$ AR were also reported to share some features, such as a network of hydrogen bonds ranging from their ligand binding pocket to their cytoplasmic side, which was proposed to be involved in conformational changes upon activation (Li *et al.*, 2004; Pardo *et al.*, 2007; Kobilka *et al.*, 2008). The crystal structure of rhodopsin was an important step in order to determine the activation states of all 7TM/GPCRs. Later, the crystal structure of  $\beta_2$ AR was solved and showed some features that are shared by both rhodopsin and  $\beta_2$ AR, but also some important differences between the two receptors.

#### 2.4.1. The $\beta_2$ Adrenergic Receptor

The  $\beta_2$ AR is the most studied member of the adrenergic family of receptors. The  $\beta_2$ AR is known to control a variety of biological functions with respect to physiological relevance. Mostly,  $\beta_2$ AR regulates the smooth muscle of the airway and vasculature and it is activated by catecholamines in addition to numerous exogenously administered adrenergic drugs. Adrenergic receptors are the most intensely studied classes of membrane

protein and are often used as a model in order to examine the signaling mechanisms of GPCR (Mcgraw *et al.*, 2005).

Over the past few decades, most studies have been based on the immunomodulatory properties of  $\beta_2$ ARs. Importantly, adrenoceptors, which are the subtype of  $\beta_2$  class receptors, have been identified on several immunocompetent cell types and the effects of them on the inflammatory immune response of these cells are well-known. Although primarily characterized by differences in function and distribution, research has demonstrated that members of the adrenergic family of receptors share several similarities with respect to structural properties. As shown in Figure 2.6, there are three major  $\beta$ -AR subtypes ( $\beta_1$ ,  $\beta_2$  and  $\beta_3$ ); each of which has been identified through a combination of biochemical and pharmacological techniques (Tan *et al.*, 2007). All of these receptor subtypes are composed of a single polypeptide chain that is approximately 400 to 500 amino acid residues in length. The primary structure of all adrenergic receptors contains seven stretches of hydrophobic amino acid residues. These highly conserved hydrophobic stretches correspond to seven alpha helical transmembrane regions that span the lipid bilayer of the cell upon which the receptor resides (Strosber, 1995; Kristiansen, 2004; Gether, 2000; Klabunde *et al.*, 2002).

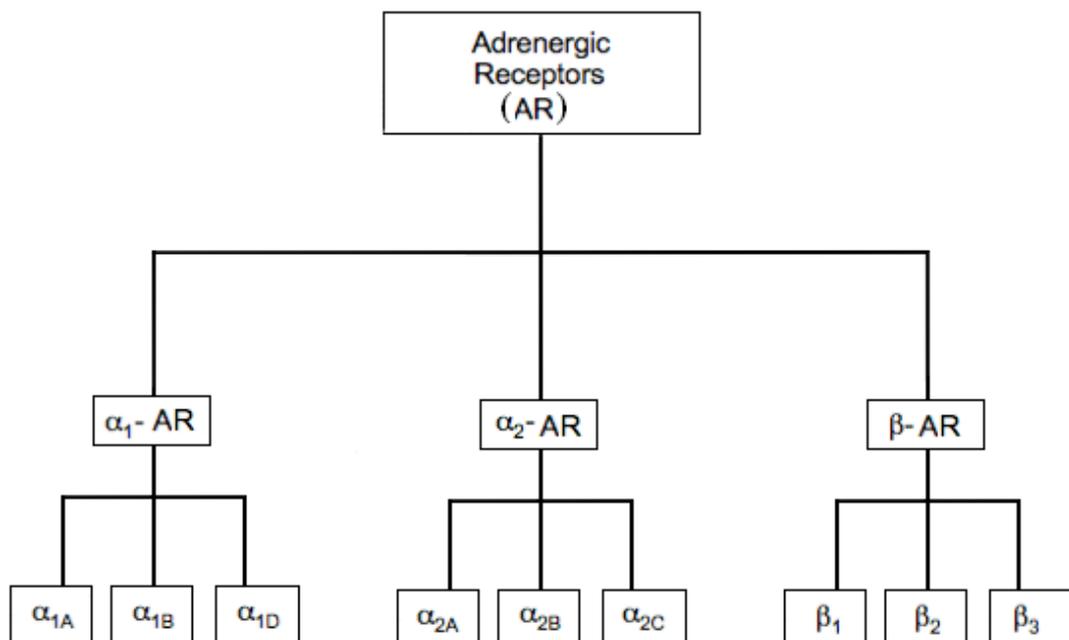


Figure 2.6. Classification of adrenergic receptors. ARs can be broadly classified into two major categories: Alpha ( $\alpha$ ) and beta ( $\beta$ ) adrenergic receptor.

All ARs have three intracellular and three extracellular loops. Each AR possesses an extracellular N-terminal domain and an intracellular C-terminal domain. Both of these terminal domains are of variable length and sequence depending on AR subtype (Strosberg, 1995).

#### 2.4.2. Overall Topology of The Human $\beta_2$ Adrenergic Receptor

The latest crystal structure of  $\beta_2$ AR-T4-lysozyme, which is shown in Figure 2.7, includes 442 amino acids and also includes a palmitic acid covalently bound to Cys341 (Cystine) and an acetamide bound to Cys265. Moreover, the crystal structure of  $\beta_2$ AR includes carazolol molecule, three cholesterol molecules, two sulfate ions and two butanediol molecules.

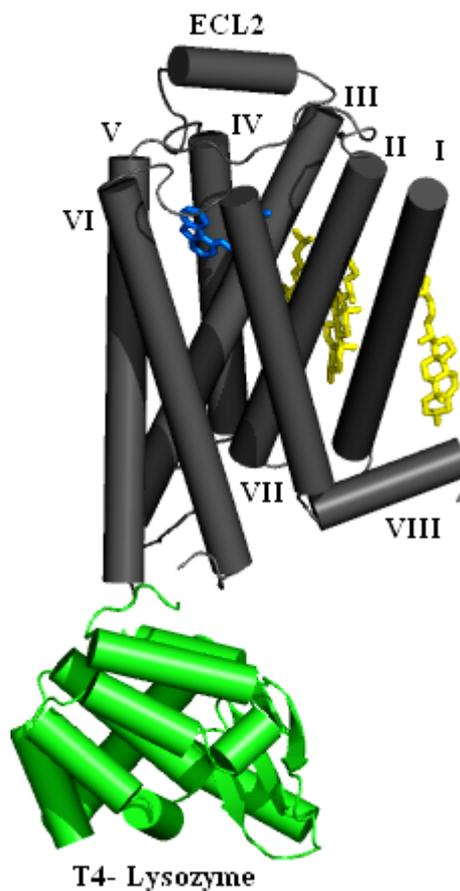


Figure 2.7. The  $\beta_2$ AR-T4L. The receptor and T4L are colored dark gray and green.

Carazolol (blue) and lipid molecules (yellow) bound to the receptor are shown. The crystal structure of the  $\beta_2$ AR without T4L includes 311 residues (residue 29 to 342) because of the N-terminus (residue 1 to 28) and the C-terminus (residue 343 to 365) are disordered, therefore cannot be detected experimentally. Also  $\beta_2$ AR includes ligand carazolol and the two disulfide bonds Cys106 – Cys191 and Cys184 – Cys190 (Cherezov *et al.*, 2007).

### 2.4.3. Signaling Mechanism of The Human $\beta_2$ Adrenergic Receptor

Classically, the signaling pathway of the  $\beta_2$ AR involves agonist-promoted binding of the receptor to the heterotrimeric guanosine triphosphate-binding protein Gs, activation of adenylyl cyclase, and production of cyclic adenosine monophosphate (cAMP). Current evidence suggests that the signaling of  $\beta_2$ AR is regulated by interaction with multiple proteins. These interactions are classified into three main groups. The first one includes Guanosine triphosphate-binding proteins such as Gs and Gi. The second one includes protein kinases such as the cAMP-dependent protein kinase, protein kinase C, GPCR kinases and tyrosine kinases and the last one includes adaptor proteins such as arrestins, A-kinase anchoring proteins, and the Na<sup>+</sup>/H<sup>+</sup> exchanger regulatory factor (Johnson, 2006).

The majority of  $\beta_2$ AR mediated signaling, which is shown in Figure 2.8, occurs via Gs-proteins and subsequent cAMP-dependent mechanisms (Hall *et al.*, 1998; Tan *et al.*, 2007; Liggett, 2002; Strosberg, 1993). The most notable alternative signaling pathway is the Gi-dependent pathway that results in the activation of the mitogen-activated protein kinase (MAPK) pathway (Azzi *et al.*, 2003; Tan *et al.*, 2007). This Gi dependent pathway requires the phosphorylation of the  $\beta_2$ AR by PKA and is mediated by the  $\beta\gamma$ -subunit of the associated G-protein. This subunit, along with  $\beta$ -arrestin, serves as a scaffold for other signaling molecules such as SOS, cSrc, RAS and Raf (Daaka *et al.*, 1997; Hein *et al.*, 1995).

Recent data suggests that the MAPK pathways can also be activated by Gs-dependent mechanisms. This signaling pathway is complex but leads to MAPK activation via the B-Raf signaling cascade. Some studies have shown that  $\beta_2$ AR signaling can occur via G-protein independent mechanisms. There is no doubt that, the complexity of the  $\beta_2$ AR

signaling mechanisms is mirrored by the diverse role of these receptors (Tan *et al.*, 2007).

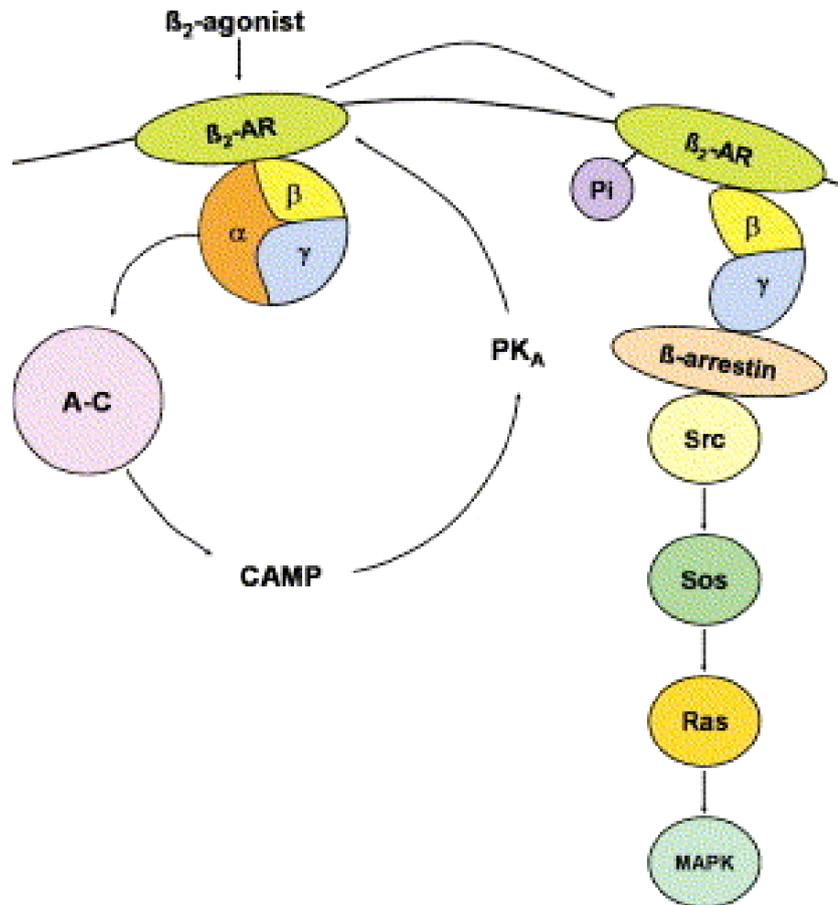


Figure 2.8. Intracellular signaling pathway of the  $\beta_2$ AR (Johnson, 2006).

#### 2.4.4. Activation of The Human $\beta_2$ -Adrenergic Receptor (Ligand Interaction)

Traditionally, the theory of the interaction between receptor and ligand was used to describe agonist activation of  $\beta_2$ ARs. This theory, which was called “lock and key”, proposed the  $\beta_2$ AR agonist would bind the receptor in such a way that the receptor would adopt a conformation that is better suited to associate with Gs. However, current studies suggested that the receptor actually toggles between distinct conformational states in the absence of ligand (Mcgraw *et al.*, 2005). Moreover,  $\beta_2$ ARs may be active even in the absence of receptor agonist (Soudijn *et al.*, 2005). Predominantly, the active and inactive states are in equilibrium with the inactive state of the receptor. Many studies supported that the agonists of  $\beta_2$ AR such as isoproterenol, epinephrine, norepinephrine, and carmoterol

exert their effects by binding to and stabilizing the active form of the receptor.

The ligands of the  $\beta_2$ AR determine the way to interact with the receptor and its binding domain. For example, hydrophilic agonists can access the  $\beta_2$ AR binding site directly from the aqueous extracellular region. They are often referred to as “short-acting” agonists due to direct access and rapid action of them. On the other hand, lipophilic agonists, which are often referred to as “long-acting”, are readily taken up into the cell membrane. Once within the cell membrane, the  $\beta_2$ AR agonist slowly leaches out into the active site of the receptor (Johnson, 2001).

On the other hand, inverse agonist or antagonists of the  $\beta_2$ AR such as carazolol, pindolol, propranolol, alprenolol, timolol so forth, bind to the inactivated form of the receptor thus moving the equilibrium further away from the active form of the receptor (Mcgraw *et al.*, 2005). Actually, agonists and antagonists should not be as a competitor to bind to the same receptor. Instead, these ligands bind to different forms of the  $\beta_2$ AR and shift the receptor conformation equilibrium in their favor. In recent years, significant regions of the  $\beta_2$ AR have been identified for ligand-binding domain with the hydrophobic core of the  $\beta_2$ AR protein (Houslay *et al.*, 2000; Leineweber *et al.*, 2004).

Adrenergic drugs could originally be classified as full agonists, partial agonists, antagonists, full inverse agonist, and partial inverse agonist (Parra *et al.*, 2007). Full agonists bind and activate a particular signaling pathway in the receptor. Partial agonists bind and activate the receptor to different degrees, typically less than that of the full agonist. Inverse agonism is defined as the ability of a ligand to reduce the basal level of signaling activity following receptor-ligand binding. On the other hand, a  $\beta_2$ AR antagonist is unable to modify constitutive receptor activity alone, but it is able to block the agonist-induced activity of the  $\beta_2$ AR (Bhattacharya *et al.*, 2008).

To better understand distinct conformational changes of  $\beta_2$ AR, which is triggered by structurally different ligands, some computational methods were developed. These computational studies show that, the receptor conformational state depends on the structure and efficacy of the ligand for a given signaling pathway. The conformational switches in  $\beta_2$ AR depend on the breaking of the ionic lock between Arginine 131 (R) at the

intracellular end of TM3 (part of the E/DRY motif) and Glutamic acid 268 (E) on TM6, and the rotamer toggle switch on Tryptophan 286 (W) on TM6. Moreover, the ligand-stabilized receptor conformations of different ligands show that conformational changes of the receptor are completely dependent on particular ligands and tremendously significant for developing functionally specific drugs (Bhattacharya *et al.*, 2008). In addition, recent studies mentioned that many of the conserved residues of the  $\beta_2$ AR have a dual role. They constrain the seven-TM bundle in its inactive conformation and are main determinants of the structural changes that occur on receptor activation. One of the most significant structural changes occurring during transition from inactive to active state is the break of the two salt bridges (Asp130-Arg131 and Arg131-Glu268) (Figure 2.9).

Although the ionic lock between Arg131-Glu268 is broken upon receptor activation, the crystal structure, which is inactive, lacked these contacts. Broken ionic lock has raised questions about the true conformation(s) of the inactive state and the role of the ionic lock in receptor activation and signaling. Recent results, which were obtained from long-range MD simulations, suggest that inactive state of  $\beta_2$ AR exists in equilibrium between conformations with the lock formed and the lock broken, whether or not the co-crystallized ligand is present (Dror *et al.*, 2008).

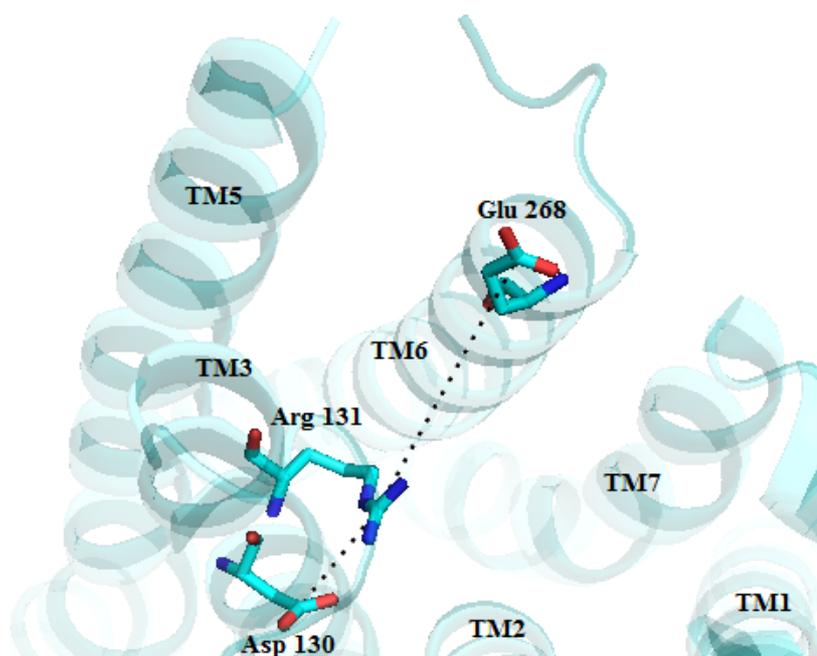


Figure 2.9. Ionic lock formation between Asp130-Arg131 and Arg131-Glu286 in  $\beta_2$ AR.

Another study suggests that the lock breaks and reforms along with the conformational changes within the TM helices. Some biochemical evidence and hypothesized models suggest that these conformational changes are associated with the activation of the receptor. Moreover, the ionic lock was classified into three main states: closed (or locked), semi-open with bridging water, and fully open. Hence, the toggling and breaking ionic lock is thought to be the results of the characterization of motions of  $\beta_2$ AR (Romo *et al.*, 2010).

A recent experimental study detected a new salt bridge, facing the extracellular side between Asp192 and Lys305, which connects ECL2 to ECL3, as seen in Figure 2.10. The functional role of the salt bridge linking extracellular loops (ECLs) two and three in receptor activation is remarkably diverse, and therefore represents an ideal target for the discovery of subtype-selective drugs (Bokoch *et al.*, 2010).

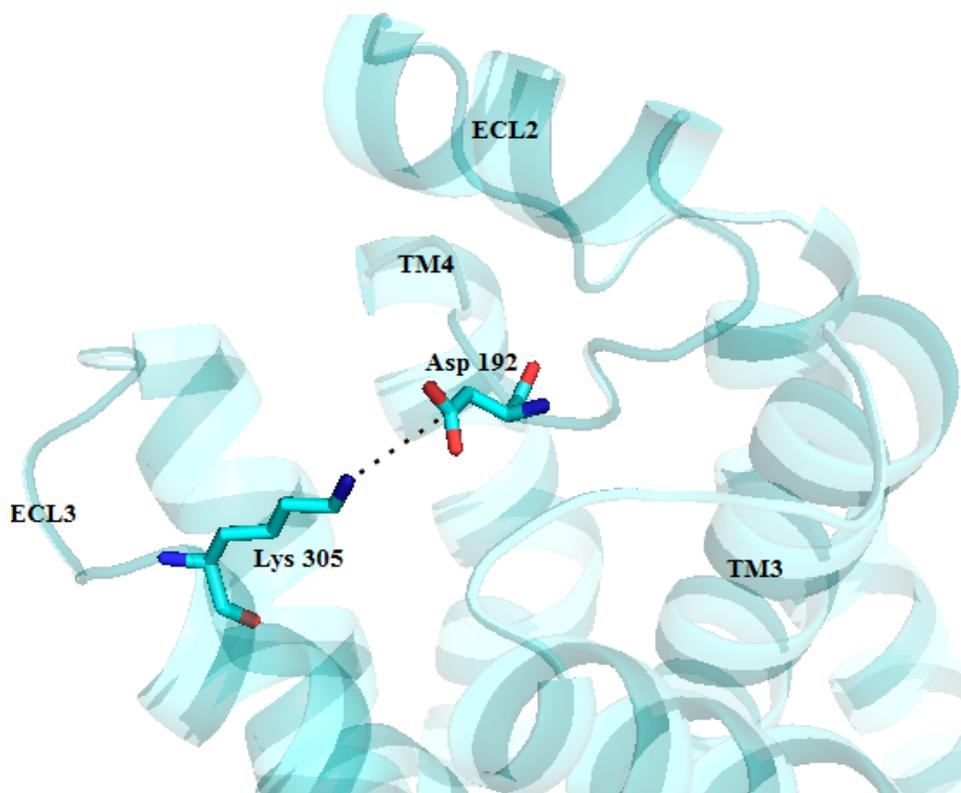


Figure 2.10. Ionic lock formation between Asp192-Lys305 in  $\beta_2$ AR.

### 3. THE MODELS AND THE METHODS

In this thesis, the structure-function relationship of the human  $\beta_2$ AR is addressed, via two different computational approaches, which are Anisotropic Network Model (ANM) and Molecular Dynamics (MD) simulations.

#### 3.1. System Setup for Molecular Dynamics

The human  $\beta_2$ AR is a cell membrane protein that belongs to the GPCR protein family. Its physiological surroundings have a major influence on the dynamic structure of the protein. Thus, the protein has been studied along with lipid cell membrane and water molecules.

The three-dimensional structure of the human  $\beta_2$ AR has been obtained from X-ray crystallographic experiments and has been deposited in Protein Data Bank with the PDB code of 2rh1 (Cherezov *et al.*, 2007; Berman *et al.*, 2000). When this structure is examined, it comes to light that the intracellular loop III region (ICL3), which is mainly formed by residues between 231 and 262, has been cut off and that the protein T4 lysozyme has been attached to the protein at this region. Because of the protein's utmost elastic structure and its low surface area due to the fact that half of its total volume is buried in the cell membrane, it is impossible to crystallize the protein on its own. Mobility of the protein has decreased as a result of the attachment of the T4L and the total polar surface area of the system has increased. Consequently, crystallization of the protein has been performed successfully. Following the methods related to genetic engineering, T4 lysozyme has been expressed and isolated in a structure that rendered it attached to the  $\beta_2$ AR.

To model a  $\beta_2$ AR close to its natural state, it is required that the T4L has been replaced with the ICL3 loop region. The loop region has been generated with one of the popular homology modeling tools called MODWEB (Pieper *et al.*, 2009). Homology modeling is the estimation of the three-dimensional (3-D) structure of a protein by amino acid sequence alignment to proteins whose 3-D structures are already resolved experimentally. Homology modeling algorithm also completes flexible loop regions in the

protein that are hard to detect experimentally. Hence, in this study, homology modeling is solely used to complete the missing loop region ICL3, since the 3-D structure of  $\beta_2$ AR is already known.

### 3.1.1. Homology Modeling (MODWEB)

Functional characterization of a protein sequence is one of the most frequent problems in biology. This task is usually facilitated by an accurate 3-D structure of the studied protein. In the absence of an experimentally determined structure, comparative or homology modeling (Figure 3.1) often provides a useful 3-D model for a protein that is related to at least one known protein structure (March *et al.*, 2000).

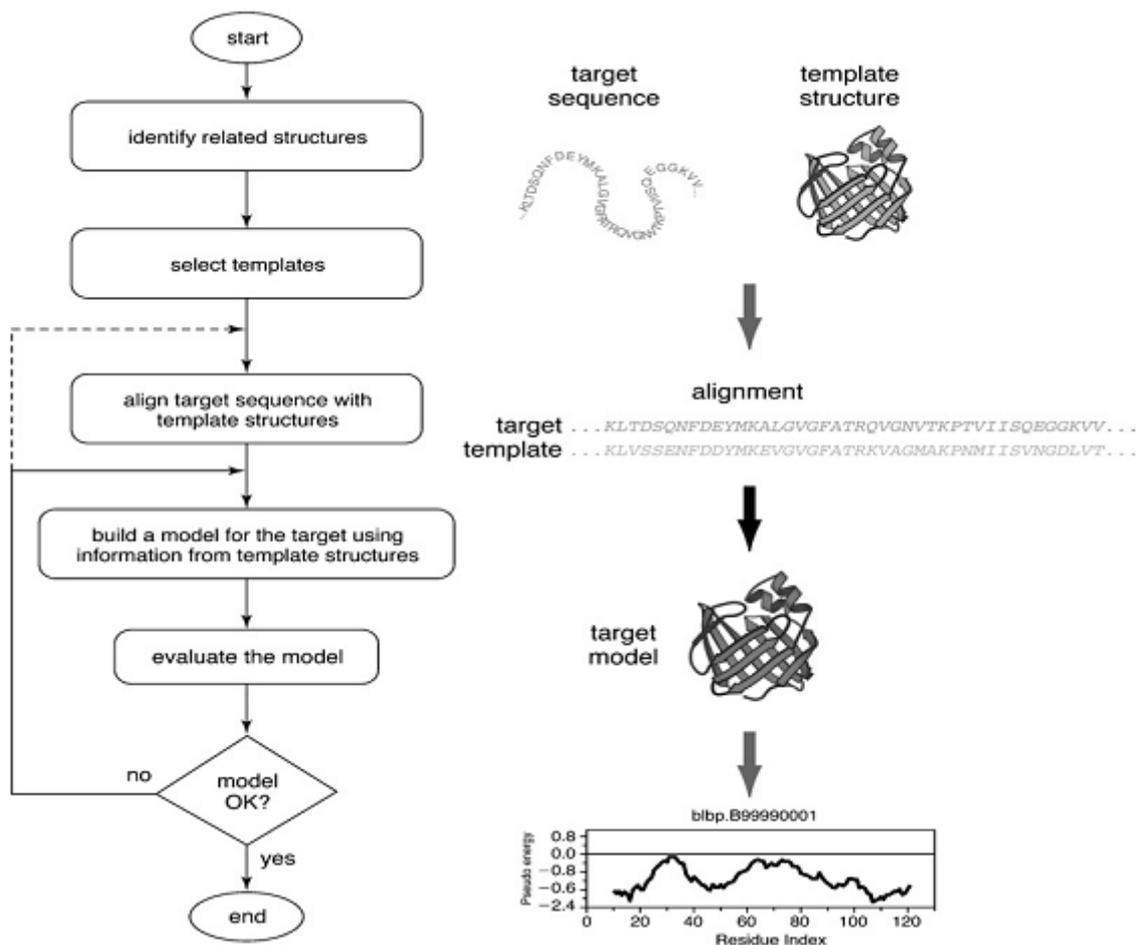


Figure 3.1. Steps in comparative protein structure modeling (March *et al.*, 2000).

Comparative modeling predicts the 3-D structure of a given protein sequence (target)

based primarily on its alignment to one or more proteins of known structure (templates) and consist of four main steps (March *et al.*, 2000). Several computer programs and web servers automate the comparative modeling process. In this thesis, MODWEB server is used in order to be able to build complete models of the human  $\beta_2$ AR including the ICL3 loop region.

MODWEB is a server for comparative protein structure modeling. It depends on the large-scale protein structure-modeling pipeline, ModPipe, for its functionality (Eswar *et al.*, 2006). The flowchart of ModPipe is shown in Figure 3.2. The structural templates used to build models in ModPipe consist of a set of non-redundant chains extracted from structures in the Protein databank. Sequence-structure matches are established using multiple variations of sequence-sequence, profile-sequence, sequence-profile and profile-profile alignment methods. For the sake of efficiency, the fold assignment methods and sequence-structure alignment methods have been combined. Models are built for each one of the sequence-structure matches using comparative modeling by satisfaction of spatial restraints as implemented in Modeller (Sali and Blundell, 1993).

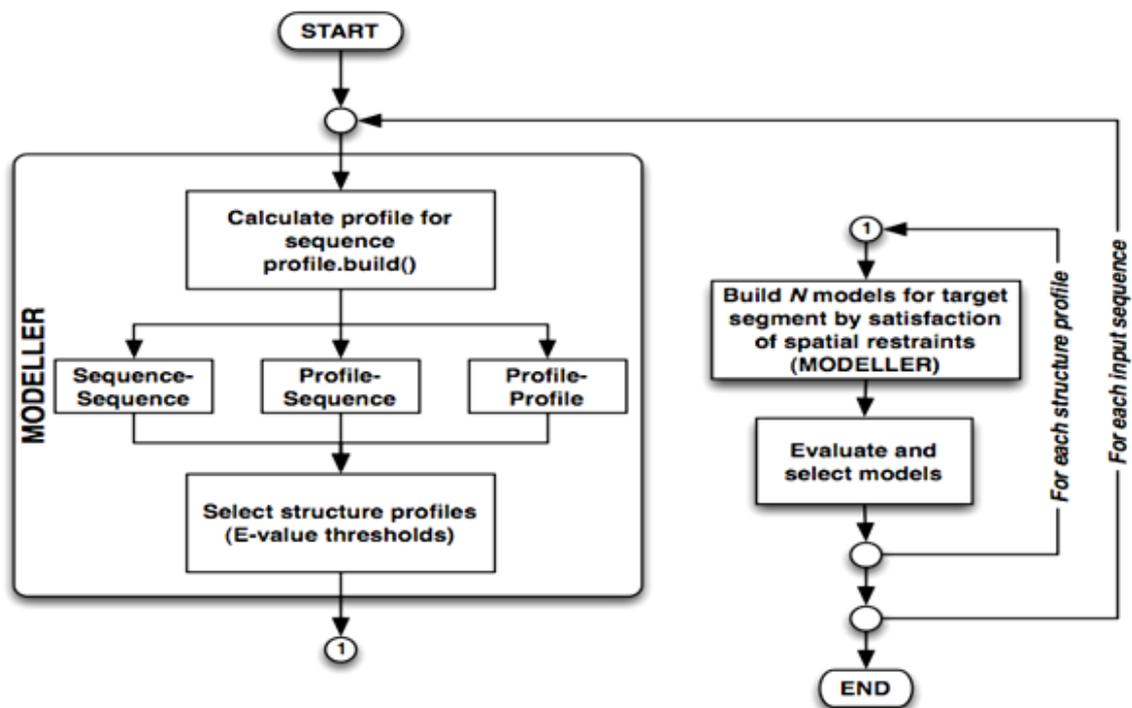


Figure 3.2. The flowchart of the ModPipe (Eswar *et al.*, 2006).

The flowchart depicting the ModPipe protocol is as follows:

- (i) Calculates a profile (multiple sequence alignment) using the *profile.build ()* method in Modeller and/or PSI-BLAST.
- (ii) Calculates fold-assignments and sequence-structure alignments using several optional variations of sequence-sequence, profile-sequence, sequence-profile, profile-profile, sequence-consensus methods. All these methods are implemented in Modeller.
- (iii) Hits (input sequence to template structure matches) with an E-value better than a specified cut-off value are selected.
- (iv) The selected hits/alignments are clustered to remove redundancy using specified thresholds.
- (v) A specified number of models are calculated for each selected hit/alignment. If multiple models are calculated, one of the many model assessment scores implemented in Modeller can select the best model (for a given hit/alignment).
- (vi) All the calculated models are then subjected to several fold assessment tests. Some of these scores are part of Modeller and some are specific to ModPipe.
- (vii) The resulting profiles, alignments, models and all scores are written out to files in specific locations within the ModPipe file system hierarchy.
- (viii) Finally, the resulting models are evaluated using several model assessment schemes and the best scoring models are returned to the user.

### **3.1.2. Generating Model Structures**

Given the whole amino acid sequence of  $\beta_2$ AR in FASTA format as input, the MODWEB server generated the best scoring models based on appropriate templates (Figure 3.3). One of the templates is the cholesterol bound form of human  $\beta_2$ AR with the

PDB ID code of 3d4s. The estimated conformation of ICL3 region is shown in purple in Figure 3.4. What is encouraging is that it displays a considerable similarity with the conformation obtained from MD simulations conducted by Dror *et al* (Dror *et al.*, 2008) (Figure 3.5).

**MGQPGNGSAFLLAPNGSHAPDHDVTQQR**DEVWVVGMGIVMSLIV  
 LAIVFGNVLVITAIKFERLQTVTNYFITSLACADLVMGLAVVPPFGAA  
 HILMKMWTFGNFWCEFWTSIDVLCVTASIE TLCVIAVDRYFAITSPFK  
 YQSLLTKNKARVILMVWTVSGLTSFLPIQMHWYRATHQEAINCYAN  
 ETCDFFTNQAYALASSIVSFYVPLVIMVFVYSRVFQEAKRQL**QKIDK**  
**SEGRFHVQNLSQVEQDGRTGHGLRRSSKFCLKEHKALKTLGIIMGT**  
 FTLCWLPFFIVNIVHVIQDNLIRKEVYILLNWIYVNS**GFNPLIYCRSP**  
**DFRIAFQELLCLRRSSLKAYGNGYSSNGNTGEQSGYHVEQEKENKL**  
 LCEDLPGTEDFVGHQGTVPSDNIDSPGRNCSTNDSL

Figure 3.3. Initial sequence that includes overall missing residues of  $\beta_2$ AR except the sequence of T4L. The first 28 missing residues of  $\beta_2$ AR (red), the missing sequence of ICL3 between 231<sup>st</sup> to 262<sup>nd</sup> residues (green) and the missing sequence between 343<sup>th</sup> to 365<sup>th</sup> residues (blue).

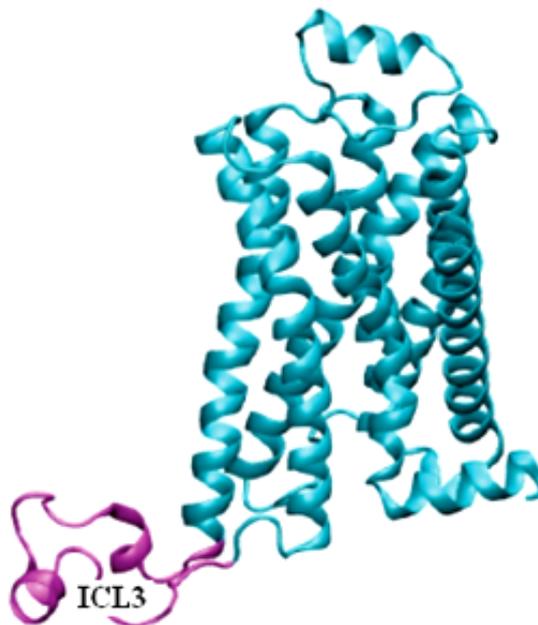


Figure 3.4. The conformation of  $\beta_2$ AR with estimated ICL3 region. The missing ICL3 region between helices V and VI is shown in purple.

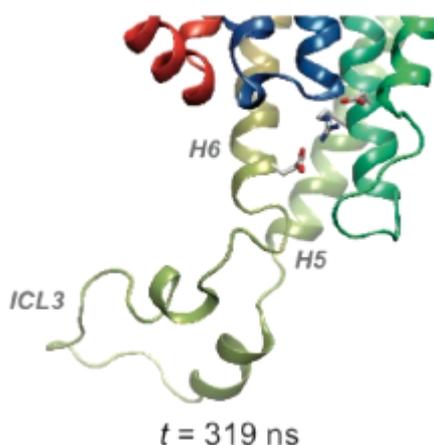


Figure 3.5. Estimated ICL3 model obtained from MD study (Dror *et al.*, 2008).

As a second input, the amino acid sequence of  $\beta_2$ AR without the sequence of ICL3 loop region is given to the server. This has created a second alternative model, where Leu230 and Lys263 residues, located at both ends of the ICL3 loop region, are connected to each other with a peptide bond (Figure 3.6). In other words, 32 residues which form the missing loop region have been extracted from the model.

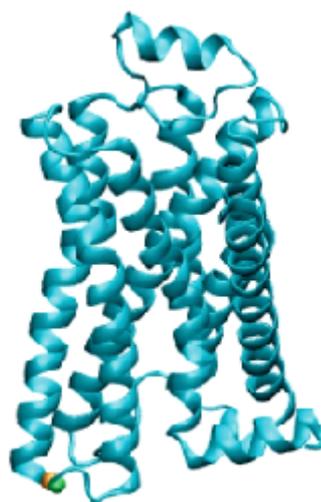


Figure 3.6. Clipped model of  $\beta_2$ AR. Peptide bond is in between the C and N atoms belonging to Leu230 and Lys263 residues and is highlighted with green and orange spheres in turn.

Throughout the studies, we designated the model with the ICL3 loop region as ‘looped’ model, while we designated the second model without the ICL3 region as

‘clipped’ model. Comparing the outcomes of the simulation for these two models, we wanted to reveal the influence of ICL3 region on protein’s intrinsic dynamics. We estimate that this region will have a substantial influence on protein dynamics, as this region is located on signal transmission pathway where the protein is connected to G proteins.

### 3.1.3. Determination of Protonation States

The simulation surroundings have been prepared similarly for both models. Before this procedure, the “protonation” state of the protein has been determined. Along with  $\alpha$ -amino and  $\alpha$ -carboxyl groups, which are located at both ends of the protein chain, seven amino acids in total (Glu, Asp, His, Cys, Tyr, Lys and Arg) have side groups that could release protons depending on pH, temperature and chemical environments they are in. Though pH (=7) and temperature (= 310 Kelvin) are constant, the chemical surroundings differ, thus have a major effect on the protonation state (i.e. pKa values) of these groups.

The programs used for this procedure are MOE (The Molecular Operating Environment) (Labute, 2007) and H++ web server (Bashford and Karplus, 1990), with the latter being conducted on the internet. According to results obtained from these programs, no amino acid has been found to indicate a change in its standard pKa value. Glu, Asp, His, Cys, Tyr, Lys and Arg residues have standard pKa values as 4.1, 4.1, 6.0, 8.3, 10.0, 10.8 and 12.5. In a neutral environment where pH is 7, Glu, Asp and His have their protons released, whereas Cys, Tyr, Lys and Arg residues are noted to have preserved their protons.

Histidine could bear two different neutral states according to its micro-surroundings. Both nitrogen atoms at histidine’s imidazole ring (called as N $\epsilon$  and N $\delta$ ) could release its hydrogen atom. Which one will be protonated or deprotonated depends on its chemical surroundings. Our systems consist of seven different His structures. Five of them (residues 93, 172, 178, 241 and 296) have H atom on N $\epsilon$ , thus the residue name becomes ‘HIE’, while residues 256 and 269 have H atom on N $\delta$ , thus the residue name is called as ‘HID’. As an example, His241 could be given as shown in Figure 3.7. The distance between N $\delta$  atom and the backbone nitrogen atom of Asn244 nearby is approximately 2.82 Å. The distance mentioned is not convenient for an H bond to be formed. If N $\delta$  of His241 was protonated, repulsive interactions would occur between N $\delta$  of His241 and protonated

backbone N of Asn244. Thus,  $N_{\epsilon}$  of His241 is more suitable for protonation.

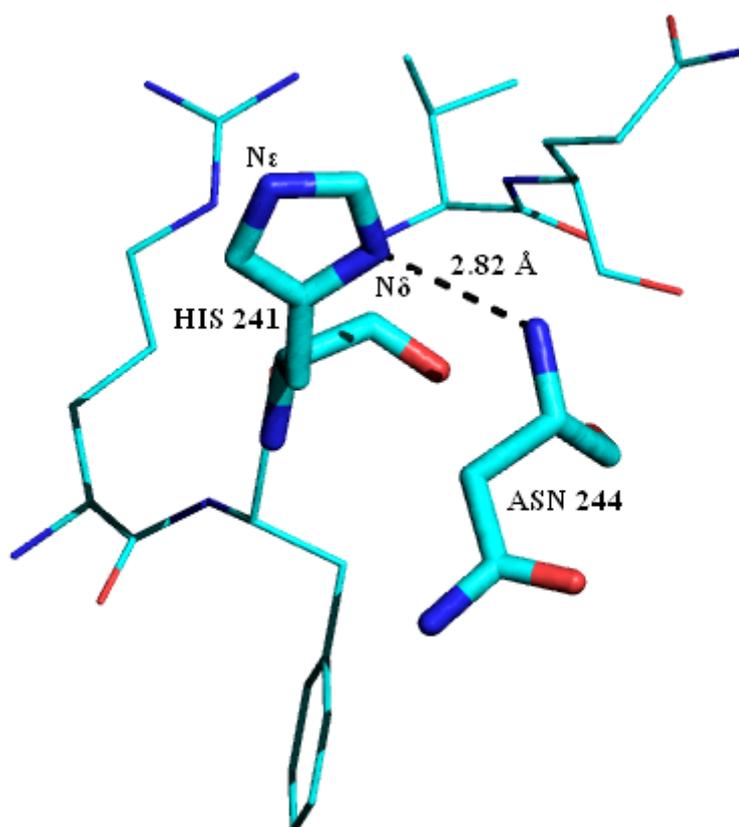


Figure 3.7. The schematic overview of the interaction between His241 and Asn244. Dotted line represents the bond between  $N_{\delta}$  atom included in His241 and backbone nitrogen included in Asn244.

#### 3.1.4. Determination of the Orientation of Protein in Membrane

Receptor protein stands in the cell membrane with an oblique angle of  $6^{\circ}$ . This value has been obtained from theoretical calculations and presented in OPM (Orientations of Proteins in Membranes) databank (Lomize *et al.*, 2006). The surface of the cell membrane has been situated to be perpendicular with z-axis of the coordinate file given in OPM. The new position of the receptor structure is determined such that the oblique angle between its main principal component along the cell membrane and the z-axis is  $6^{\circ}$ . In order to obtain the same tilt in our two models ‘looped’ and ‘clipped’ derived from homology modeling, they have been aligned to the structure given in OPM. The clipped model after the alignment is shown in Figure 3.8.

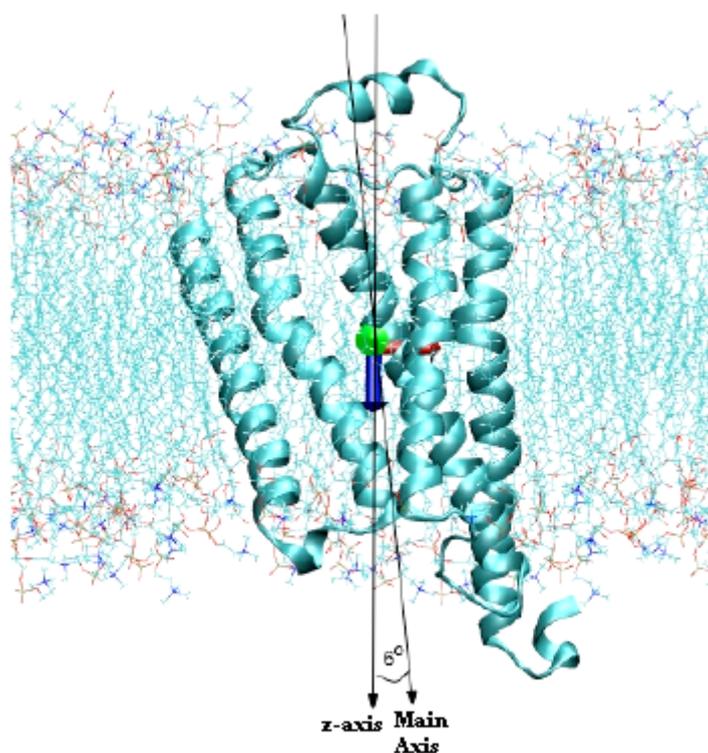


Figure 3.8. The inclined structure of  $\beta_2$ AR that has been derived by its alignment with the model obtained from OPM.

It is estimated that posing a role in the mutual interactions of seven transmembrane helices, water molecules in the interior part of the receptor affect the dynamics of the receptor and therefore the signal transduction considerably. Moreover, water molecules are found to be at the evolutionarily conserved regions of the protein. A similar case has been detected for rhodopsin, another protein from the same family. For this reason, we have decided to incorporate 15 water molecules, which have been detected experimentally at the inner regions of the receptor. As shown in Figure 3.9, water molecules, which are located at the inner parts of the structure, form critical hydrogen bonds with the most conserved residues.

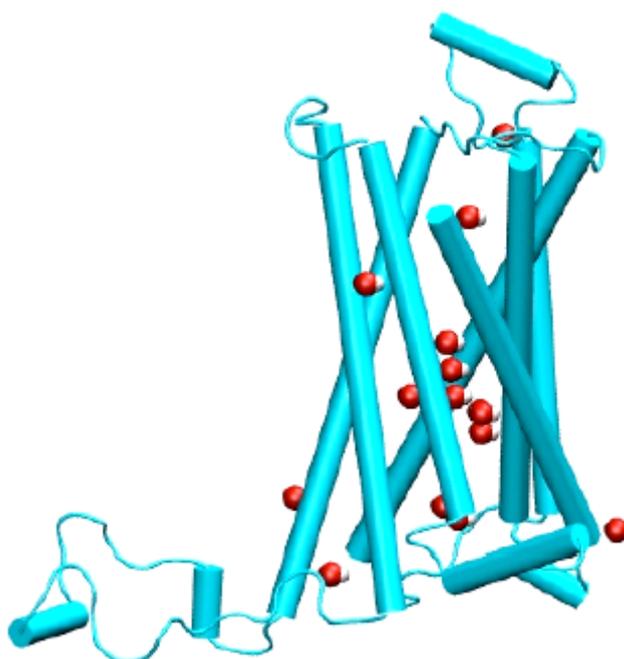


Figure 3.9. Water molecules, which are located in inner regions, are shown here as part of X-ray crystallographic experiments conducted for  $\beta_2$ AR.

### 3.1.5. System Preparation for Molecular Dynamics (MD) Simulations

The Membrane Plug-in v1.1 of VMD visualization program has been used to generate the cell membrane. A double-layered palmitoyl-oleoyl-phosphatidylcholine (POPC) phospholipid cell membrane has been generated at the direction of z-axis with a constant thickness. These molecules are the main components of the biological cell membranes. The position of the receptor in the cell membrane is shown in Figure 3.10.

The cell membrane's dimensions in the z and y directions have been determined according to the protein's dimensions in the same directions. As an example, for the looped model, min and max coordinates of the protein in the x direction are found to be  $-25.4 \text{ \AA}$  and  $30 \text{ \AA}$ , which makes a distance of  $55.4 \text{ \AA}$  between two farthest atoms in the protein. Consequently, the protein's dimension in the x direction is determined to be  $56 \text{ \AA}$ . The membrane's dimension in this direction is adjusted in such a way that the min distance between the protein and the boundary of the periodic cell is  $15 \text{ \AA}$ , thus a cell membrane is set to  $86 \text{ \AA}$  ( $= 56+15+15$ ). This way, any interaction between the protein and its own image in the periodic box at this direction is definitely avoided. A similar adjustment was made

for the y direction.

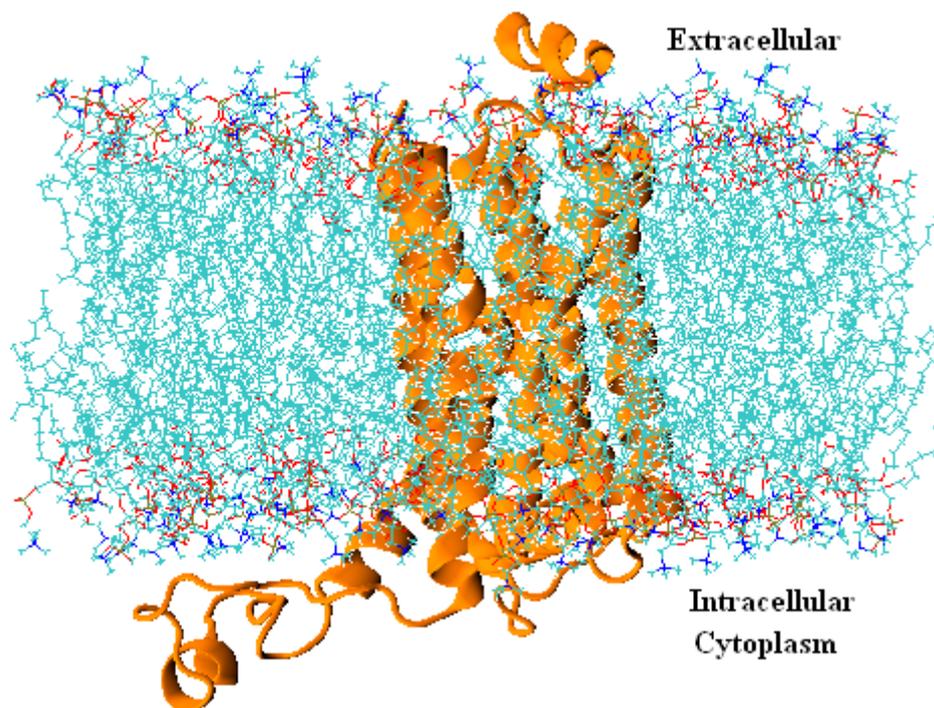


Figure 3.10. The position of the crystallographic structure of the  $\beta_2$ AR in the cell membrane depicted as an orange cartoon.

After the protein is placed inside the membrane, a solution with water molecules is formed in order to simulate the actual surroundings. Water molecules with a thickness of around 15 Å for looped model and 12 Å for clipped model at both directions of z-axis are added to the protein and the cell membrane. As an example, if minimum and maximum coordinates of the protein in the z direction are -33.3 Å and 38.0 Å, its dimension at this direction is determined as around 71 Å. As the total water thickness is 30 Å, the dimension of the system at this direction is set to 100 Å. Table 3.1 reflects the minimum and maximum values for both models of the protein at x, y and z directions, approximate dimensions, dimensions of the system and the cell membrane formed. For the clipped model, the final image of the system, which consists of cell membrane and water molecules, is shown in Figure 3.11 along with its periodic images at the x direction.

Lastly, by VMD's Autoionize module v1.2, the system has been added with a certain number of  $\text{Na}^+$  and  $\text{Cl}^-$  ions with a concentration of 0.154 mol/L to make the total net

charge of the system to be equal to zero. This necessity arises from the fact that PME (Particle-Mesh Ewald) summation method, which is used for electrostatic energy calculations during MD simulation, is only operable when the total electric charge of the system is zero. With the ions added, the looped model now has 17 Na<sup>+</sup> and 24 Cl<sup>-</sup>, and the clipped model has 10 Na<sup>+</sup> and 14 Cl<sup>-</sup>.

Table 3.1. The dimensions of the protein, cell membrane and system for each model.

Model	Protein $x_{\min} / x_{\max}$	Protein $y_{\min} / y_{\max}$	Protein $z_{\min} / z_{\max}$	Protein Dimension (xyz)	Cell Membrane (xy)	Box Dimension (xyz)
Looped	-25.4/30.0	-36.0/21.2	-33.3/38.0	56x57x71	86x86	86x86x100
Clipped	-25.9/20.7	-18.6/20.9	-33.1/32.9	47x39x66	77x69	77x69x90

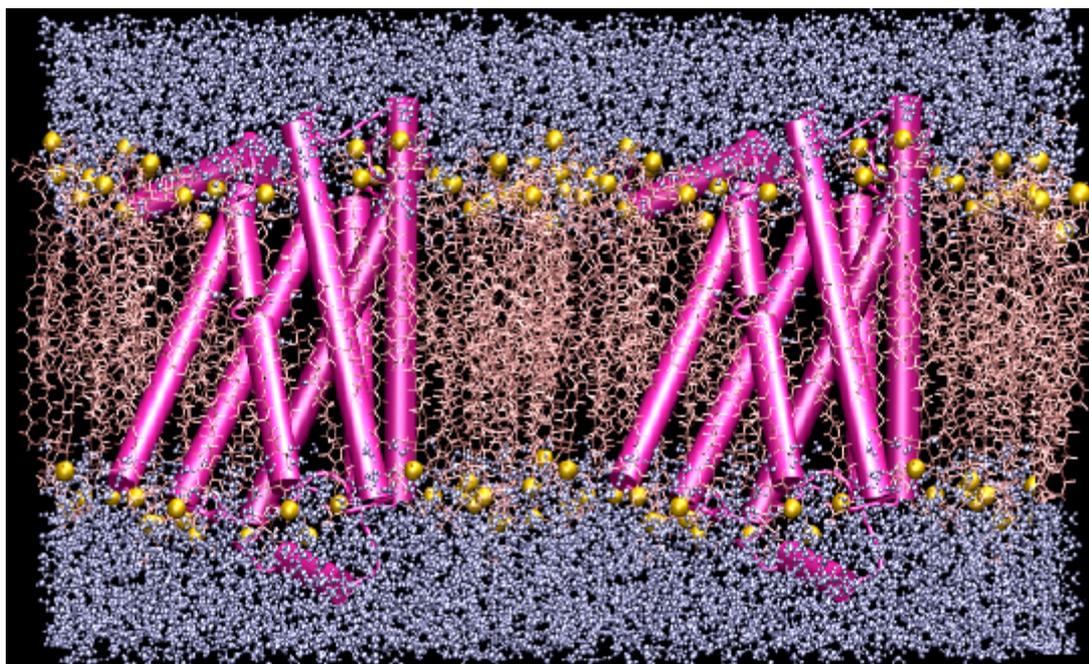


Figure 3.11. The systems consisting of  $\beta$ 2AR, cell membrane are shown along with its periodic image in the direction of x. The polar heads of the lipid and water, protein and lipid hydrocarbon ends are illustrated are in blue, yellow, magenta and pink respectively.

### 3.2. Theoretical Background of MD Simulations

One of the principal tools in the theoretical study of dynamic behavior of biological molecules is the MD simulation. This computational method calculates the time dependent behavior of a molecular system. Atoms are allowed to interact with each other using empirical potential energy functions or force fields, from which the forces acting on atoms are calculated for a given configuration. A successive configuration of the system along the time trajectory is obtained by the integration of Newton's equation of motion, which is

$$\frac{d^2 R_i}{dt^2} = \frac{F_i}{m_i} . \quad (3.1)$$

Here,  $R_i$  represents the position of particle  $i$ , and  $F_i$  is the total force acting on particle  $i$  exerted by all other molecules, and  $m_i$  its the molecular mass.

#### 3.2.1. Force Fields

Force fields describe the potential energy of a system as a function of the atomic positions/coordinates. MD simulations are based on an empirical model of interactions within a system involving stretching of bonds, bending, as well as non-bonded interactions, as shown in Figure 3.12.

$$\begin{aligned} V(R_1, \dots, R_N) = & \sum_{\text{bonds}} \frac{k_{li}}{2} (l_{li} - l_{i,0})^2 + \\ & \sum_{\text{angles}} \frac{k_{\theta i}}{2} (\theta_i - \theta_{i,0})^2 + \\ & \sum_{\text{torsions}} \frac{V_n}{2} (1 + \cos(n\phi_i - \delta_i)) + \\ & \sum_{i=1}^N \sum_{j=i+1}^N \left( 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) \end{aligned} \quad (3.2)$$

In the above equation,  $V(R_1, \dots, R_N)$  indicates the potential energy, which is a function of the positions ( $R_i$ ) of  $N$  atoms or particles. The first term in the equation describes the interaction of pairs of bonded atoms, where  $l_i$  is the bond length. The second term, which is based on Hooke's Law, is the summation of the overall angles in the molecule modeled using a harmonic potential, where  $\theta_i$  and  $\theta_{i,0}$  are the instantaneous and the equilibrium angle of the bond, respectively. The third term in the equation represents the torsional potential and the fourth equation defines the non-bonded interactions. The non-bonded interactions are defined by two different potentials. The former one is the Lennard-Jones 12-6 potential function that accounts for van der Waals interactions, whereas the latter is the Coulomb potential for electrostatic interactions (Leach, 2001).

## Empirical Potential Energy Function

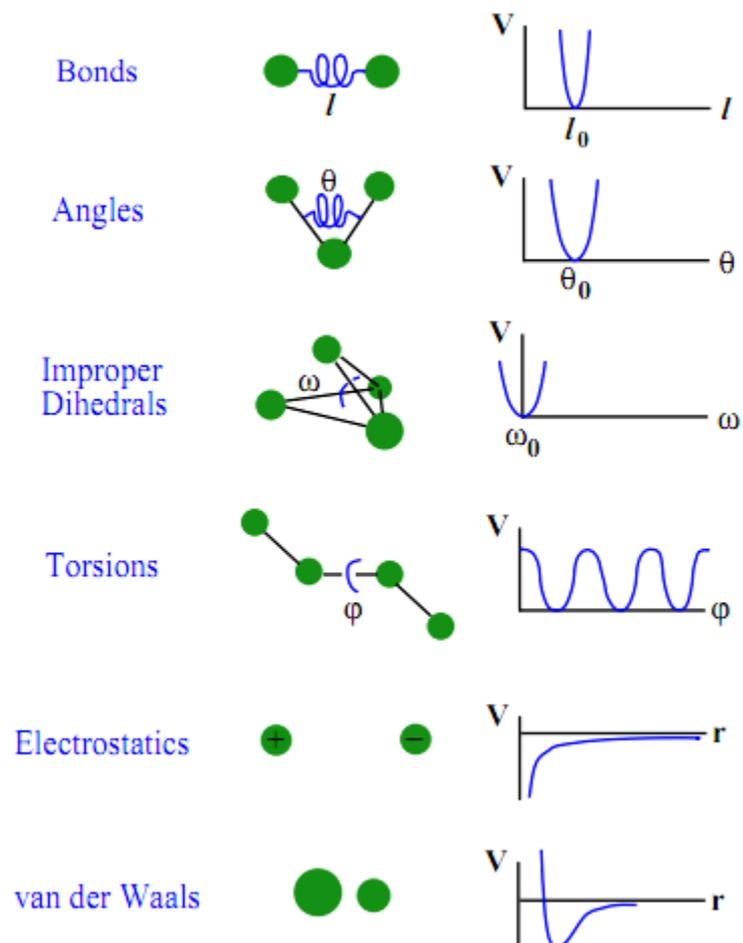


Figure 3.12. Interactions included in representative potential energy function for MD simulations (Steinbach, 2010).

### 3.2.2. Initialization of the System

In order to perform an MD simulation, it is essential to determine an initial configuration of the system by specifying  $3N$  atomic coordinates ( $R_i$ ) at time zero. The initial configuration can be generated by using experimental data such as a crystal structure. After the initial configuration of the system is minimized in a solvent box or cell membrane, the initial velocities are assigned according to the Maxwell-Boltzmann distribution at the initial temperature (Leach, 2001). After setting up the system, the potential energy of the system is calculated and the force on each atom from the derivative of potential energy is determined by

$$F_i = -\nabla U_i(R_1, \dots, R_N) = \frac{\partial U(R_1, \dots, R_N)}{\partial R_i} . \quad (3.3)$$

Once the force on each atom at time  $t$  is calculated, the next step is to generate the new conformation at time  $t+\Delta t$  according to the Equation 3.1. There are several numerical algorithms in order to perform the integration of Equation 3.1. Verlet algorithm is one of the most commonly used algorithms in MD simulations (Verlet, 1967) Verlet algorithm is based on the addition and subtraction of the Taylor series expansions for the time dependence of the coordinates  $R_i$  at times  $t-\Delta t$  and  $t+\Delta t$ .

$$R_i(t + \Delta t) = R_i(t) + v_i(t)\Delta t + \frac{1}{2}a_i(t)\Delta t^2 + \dots \quad (3.4)$$

$$R_i(t - \Delta t) = R_i(t) - v_i(t)\Delta t + \frac{1}{2}a_i(t)\Delta t^2 - \dots \quad (3.5)$$

Adding these equations together and combining with Equation 3.1 produces

$$R_i(t + \Delta t) \approx -R_i(t - \Delta t) + 2R_i(t) + \Delta t^2 a_i(t) . \quad (3.6)$$

The velocities can be calculated by the difference of the positions at  $t-\Delta t$  and  $t+\Delta t$

$$v_i(t) \approx \frac{1}{2\Delta t} [R_i(t + \Delta t) - R_i(t - \Delta t)] . \quad (3.7)$$

Slightly modified versions of the Verlet algorithm have been proposed to increase the accuracy in calculation of positions and velocities such as Leap-frog algorithm (Hockney, 1970) and the velocity Verlet method (Swope *et al.*, 1982). The velocity Verlet method, positions, velocities and accelerations at time  $t+\Delta t$  are obtained from the same quantities at time  $t$ . Furthermore, this method does not involve precision (Leach 2001).

$$R_i(t + \Delta t) = R_i(t) + v_i(t)\Delta t + \frac{1}{2}a_i(t)\Delta t^2 \quad (3.8)$$

$$v_i(t + \Delta t) = v_i(t) + \frac{1}{2}\Delta t[a_i(t) + a_i(t + \Delta t)] \quad (3.9)$$

The algorithm is applied as a three-stage procedure, since as seen from Equation 3.8, the accelerations at  $t$  and  $t+\Delta t$  is required to calculate new velocities. First, positions at  $t+\Delta t$  are calculated according to Equation 3.7 and the velocities at time  $t+\Delta t/2$  are evaluated by using the following equation:

$$v_i(t + \frac{\Delta t}{2}) = v_i(t) + \frac{1}{2}\Delta ta_i(t) \quad (3.10)$$

Then, forces are computed from the current positions to obtain the acceleration,  $a(t+\Delta t)$ . In the final step, the velocities at time  $t+\Delta t$  are calculated:

$$v_i(t + \Delta t) = v_i(t + \frac{1}{2}\Delta t) + \frac{1}{2}\Delta ta_i(t + \Delta t) \quad (3.11)$$

### 3.3. Running of MD Simulation

The running of an MD simulation consists of two phases; equilibrium and production. As the time to reach the equilibrium could be long for large and complex systems, several preliminary stages prior to equilibration are necessary in order to shorten the equilibration time. At this stage, the system is exposed to energy minimization and MD simulation for one session or more depending on the system properties. For example, for a system consisting of protein and water only, protein is kept immobile, while water molecules are allowed to be mobile. Energy minimization followed by a short MD

simulation is then applied. As a result, since water molecules respond to outside sources faster than protein does, the system reaches equilibrium a lot faster. Right after, protein is moved alongside water and the main ‘reaching equilibrium and production’ phases begin. As the system we are working on includes a cell membrane in addition to protein and water, various preparation stages are needed.

### 3.3.1. Melting of Lipid Tails

First stage is the melting of the lipid hydrocarbon ends. Here melting refers to a process that gives the cell membrane a more realistic liquid form. When the cell membrane has firstly been generated by VMD, it is made of lipid molecules aligned unrealistically on a flat line (Figure 3.13a). For this reason, at the first preparation stage, only lipid molecules are allowed to be mobile and the system under these conditions is exposed to a 1000 step minimization and later to an MD simulation of 0.5 ns. Within this period, as could be seen in Figure 3.13b, lipid molecules have shown a considerable amount of melting, with a more disordered, liquid-like structure.

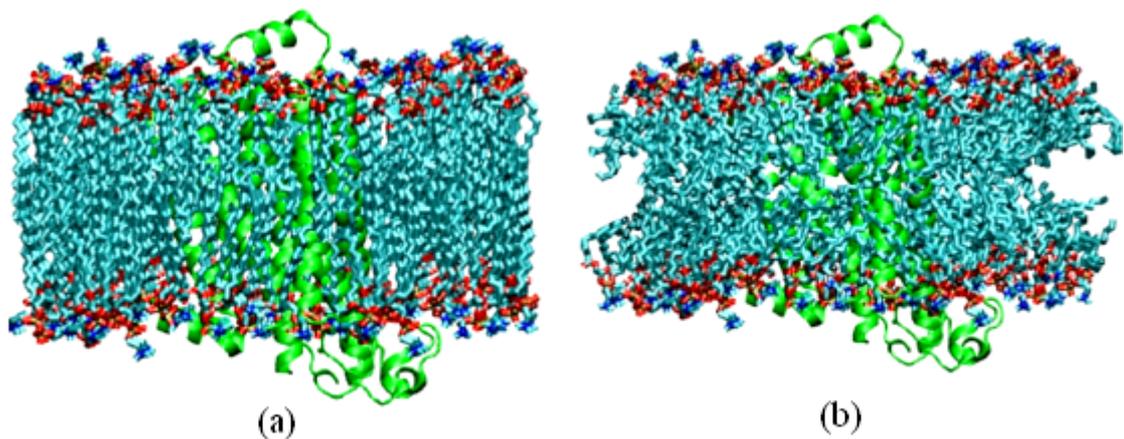


Figure 3.13. Representations of the cell membrane, (a) before the melting of lipid tails, (b) after the melting of lipid tails.

### 3.3.2. Minimization and Equilibration with Protein Constrained

The harmonic constraints have been imposed on the protein at the second preparation stage and the cell membrane and water molecules have been released. Similar to the first stage, the system under these conditions is exposed to a 1000 step minimization

followed by an MD simulation of 0.5 ns. Simultaneously, some forces are applied to water molecules in order to prevent hydration of the membrane-protein interface during equilibration. These water molecules are close enough to get into the cell membrane and these forces stop the water molecules from entering inside the cell membrane. At the end of this stage, putting the protein aside, the system reaches the local minimum faster.

### 3.3.3. Equilibration with Protein Released

At the third preparation stage, the protein released along with surrounding molecules is exposed to a short MD simulation of 5 ns. In this period, a slight shrink has occurred at the surface area of xy plane of the cell membrane (Figure 3.14). The reduction is caused by the lipid molecules, which tend to get closer to the protein. The decrease in volume in x-y directions is compensated by an increase in volume in the z direction, so that the system could keep its initial volume and pressure of 1 atm.

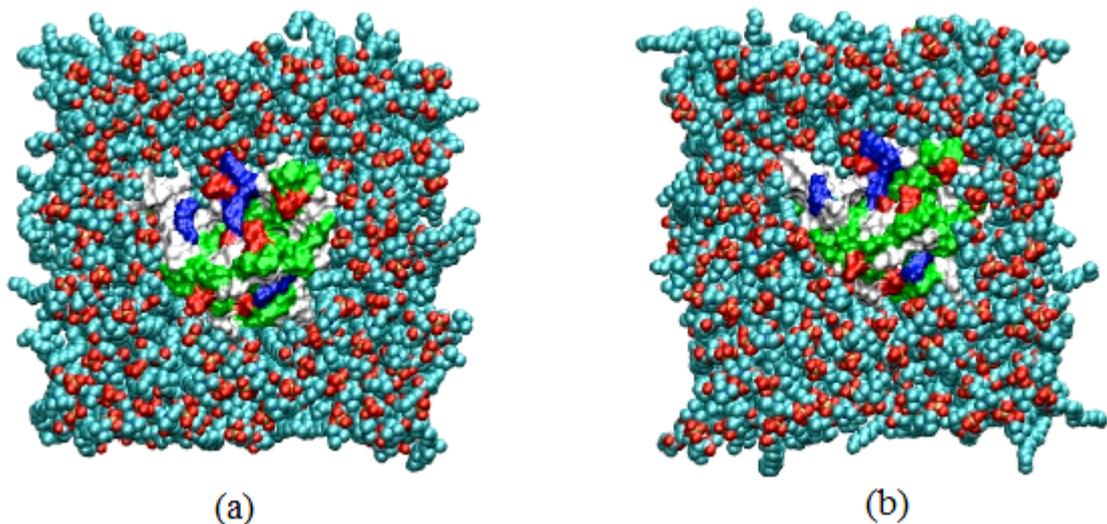


Figure 3.14. Representation of the xy plane of the cell membrane, (a) before the third preparation stage, (b) after equilibration.

Experimentally, the density of 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC) double-layered cell membrane was measured as  $0.65 \text{ nm}^2$  per one lipid molecule (Lantsch *et al.*, 1994). This value has been calculated in an approximate way through MD simulation and determined as shown in Figure 3.15. It has been noted that throughout the

simulation, there has been a decline in the dimensions through x and y directions of the periodic box but that no considerable change has been observed for the related dimensions of the protein. Thus, the most effective factor, which has caused the density to increase, has been the reduction of the surface area of the lipid molecules. As shown in Figure 3.15, surface area is stabilized at  $0.635 \text{ nm}^2/\text{lipid}$  at around 2 ns. The fact that this value is below the experimental value of  $0.65 \text{ nm}^2$ , can be explained by the slightly underestimated area of the lipids.

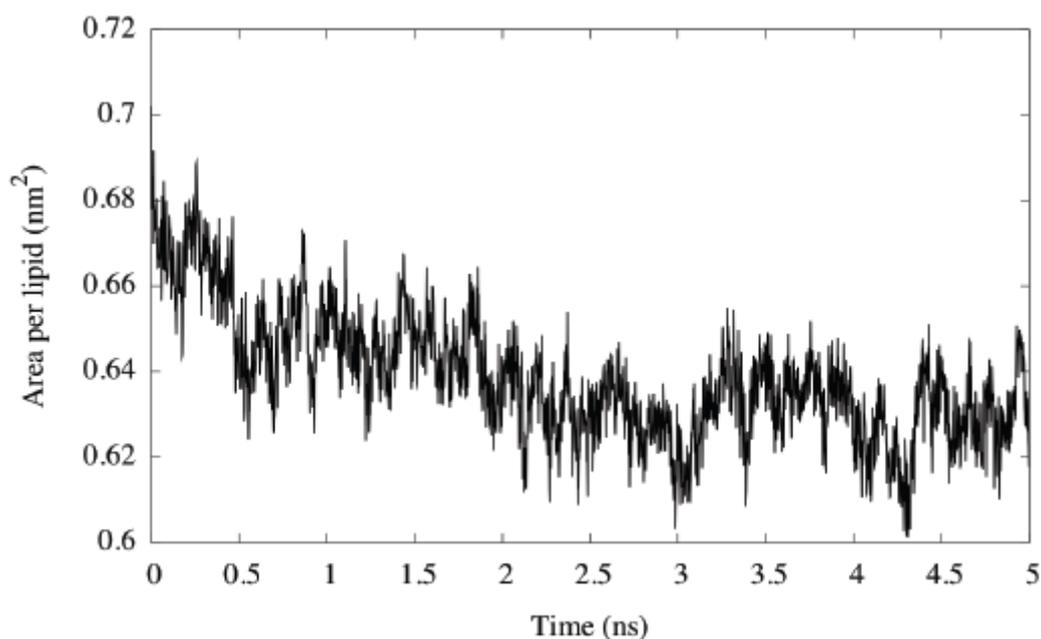


Figure 3.15. Change of surface area calculated per lipid through MD simulations.

### 3.3.4. Production Runs

After a detailed preparation stage consisting of melting of lipid tails, relaxation of water and lipids and adjustment of cell membrane density to the experimental value, an MD simulation of 500 ns for two separate clipped models (mutant and natural) and 800 ns for the looped model have been performed. The total number of atoms in each system is given in Table 3.2. In this thesis, NAMD software tool is used for MD simulations. NAMD is a parallelized MD program which is designed to run on multiple cores (Phillips *et al.*, 2005; Humphrey *et al.*, 1996). For the interaction potentials, the CHARMM27 force field was used. Three independent runs were performed for apo  $\beta_2\text{AR}$  at 310 K.

Table 3.2. Total number of atoms, lipid and water molecules at both models.

Model	Protein	Lipid	Water	Ion	Total
Looped	5.055	20.77	42.135	41	68.001
Clipped (Natural)	4.551	13.4	24.726	24	42.701
Clipped (Mutated)	4.550	13.4	24.726	24	42.700

### 3.4. Trajectory Analysis

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

Root mean square deviation is the square root of the sum of the squares of the distances between corresponding atoms of x and y (Equation 3.12). It is a measure of average atomic displacement between two conformations. Formally, given N atom positions from structure x and the corresponding N atoms from structure y; the RMSD is defined as

$$rmsd = \frac{\sqrt{\sum_{i=1}^N (d_i^2)}}{N} \quad (3.12)$$

#### 3.4.2. Mean Square Fluctuation (MSF)

The mean square fluctuation (MSF) is a measure of the variation of the position of the atoms from the average structure. MSF is proportional to the mobility. It is calculated by the following equation

$$MSF = \left\langle \left( (R_i(t) - \langle R_i \rangle)^T \left( (R_i(t)) - \langle R_i \rangle \right) \right) \right\rangle \quad (3.13)$$

where  $\langle R_i \rangle$  is the vector of time average of the cartesian coordinates of the C $\alpha$  atom of the  $i^{\text{th}}$  residue, and is the vector of the cartesian coordinates of the C $\alpha$  atom of the same residue at time  $t$ . Root mean square fluctuation (RMSF) is the square root of MSF.

### 3.4.3. Principal Component Analysis (PCA)

PCA, which is an eigenvalue decomposition of the covariance matrix (C), gives a good description of collective motions in the trajectory. This method is used to reduce large dimensional data sets to data sets with a few dimensions that still have most of the information in the original data matrix. The covariance matrix is defined as:

$$C^{(ij)} = \begin{bmatrix} \langle \Delta x_i \Delta x_j \rangle & \langle \Delta x_i \Delta y_j \rangle & \langle \Delta x_i \Delta z_j \rangle \\ \langle \Delta y_i \Delta x_j \rangle & \langle \Delta y_i \Delta y_j \rangle & \langle \Delta y_i \Delta z_j \rangle \\ \langle \Delta z_i \Delta x_j \rangle & \langle \Delta z_i \Delta y_j \rangle & \langle \Delta z_i \Delta z_j \rangle \end{bmatrix} \quad (3.14)$$

Covariance matrix composed of  $N \times N$  sub elements each having a size of  $3 \times 3$  for a  $N$  residue long protein system, and  $\langle \Delta x_i \Delta x_j \rangle$  is the cross correlation between x components of the fluctuation vectors for residue  $i$  and  $j$  ( $\Delta R_i$  and  $\Delta R_j$ ) (Bakan and Bahar, 2009).

First part consists of superimposing the structures taken as ensembles to a randomly selected reference structure and calculating an average structure. This procedure is continued in an iteration to obtain deserved average coordinates for the system.

After that, C is decomposed to its eigenvalues and eigenvectors for each dataset as:

$$C = \sum_{i=1}^{3N} \sigma_i p^{(i)} p^{(i)T} \quad (3.15)$$

where  $\sigma_i$  is the  $i^{\text{th}}$  eigenvalue of C,  $p^{(i)}$  is  $i^{\text{th}}$  eigenvector of C and  $p^{(i)T}$  is its transpose.

### 3.4.4. Clustering

Large number of conformations is created via MD simulations.  $k$ -means clustering method that is part of the kclust module of Multiscale Modeling Tools for Structural Biology (MMTSB) Tool Set (Feig *et al.*, 2004) is used to reduce the conformational space and identify a few distinct clusters or conformational states that are generated during the

simulation. For each cluster, there exists a centroid or an average structure of all the members in that cluster. Frames are assigned to clusters based on their RMSD value with respect to the centroids of the clusters.  $k$ -means clustering aims to minimize the within cluster sum of squares of distances of each element from the centroid. This procedure is repeated until every single frame is assigned to a cluster. The centroids and the clusters are updated after each iteration step. Finally, the conformation, which is closest to the centroid, is selected as the representative snapshot of that cluster.

### 3.5. Anisotropic Network Model (ANM)

Anisotropic Network Model (ANM) is improved to obtain collective motions that can be related with biological function and mechanisms of the proteins (Atilgan *et al.*, 2001). This model, which incorporates the anisotropy of residue fluctuations, is a 3-D extension of the Gaussian Network Model (GNM) with isotropic fluctuations in x-y-z directions. The potential energy of a structure with N interaction sites is expressed as a Gaussian form (Atilgan *et al.*, 2001).

$$V = (\gamma/2) \Delta R^T H \Delta R \quad (3.16)$$

Here  $\Delta R$  is a 3N-dimensional vector of the fluctuations  $\Delta R_i$  in the position vectors  $R_i$  of all sites ( $1 \leq i \leq N$ ),  $\Delta R^T$  being its transpose, and H is the Hessian matrix. Following formulation, which was proposed by Tirion (Tirion, 1996), gives information about the collective motions and large-scale conformational changes of proteins, which are functionally important. According to this formula, a universal force constant  $\gamma$  is adopted for all the interactions in the system among close-neighboring bonded and non-bonded pairs. Note that in this definition, the components of the Hessian matrix do not contain  $\gamma$ . More specifically, all pairs of sites that are closer than a cutoff distance,  $r_c$ , are connected by springs, which sum up to the potential energy:

$$V = (\gamma/2) \sum_i \sum_j h(r_c - R_{ij}) (\Delta R_j - \Delta R_i)^2 \quad (3.17)$$

$h(x)$  is the Heaviside step function [ $h(x) = 1$  if  $x \geq 0$ , and zero otherwise],  $R_{ij}$  is the distance between the  $i^{\text{th}}$  and  $j^{\text{th}}$  centers. H is composed of 3x3 super-elements of  $H_{ij}$ .

$$\mathbf{H}_{ij} = \begin{bmatrix} \mathbf{H}_{11} & \cdots & \mathbf{H}_{1N} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{N1} & \cdots & \mathbf{H}_{NN} \end{bmatrix} \quad (3.18)$$

The  $ij^{\text{th}}$  super-element for  $i \neq j$   $H_{ij}$  of H is

$$\mathbf{H}_{ij} = \begin{bmatrix} \partial^2 V_s / \partial X_i \partial X_j & \partial^2 V_s / \partial X_i \partial Y_j & \partial^2 V_s / \partial X_i \partial Z_j \\ \partial^2 V_s / \partial Y_i \partial X_j & \partial^2 V_s / \partial Y_i \partial Y_j & \partial^2 V_s / \partial Y_i \partial Z_j \\ \partial^2 V_s / \partial Z_i \partial X_j & \partial^2 V_s / \partial Z_i \partial Y_j & \partial^2 V_s / \partial Z_i \partial Z_j \end{bmatrix} \quad (3.19)$$

Here  $X_i$ ,  $Y_i$ ,  $Z_i$  are the components of the distance vector  $R_i$ . The elements of H are calculated by using the coordinates of the alpha carbon atoms. Singular value decomposition (SVD) (Kitao *et. al.*, 1991) of H yields  $3N-6$  non-zero eigenvalues and  $3N-6$  eigenvectors that represent the respective frequencies and shapes of individual modes respectively. Six eigenvalues are equal to zero due to degree of freedom for the set of the equations are six.

The inverse of the Hessian matrix (H) gives correlations between the fluctuations at sites  $i$  and  $j$  as

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \left( \frac{k_B T}{\gamma} \right) \text{tr}[\mathbf{H}^{-1}]_{ij} \quad (3.20)$$

where  $k_B$  is the Boltzmann constant,  $T$  is the absolute temperature and  $\text{tr}[\mathbf{H}^{-1}]_{ij}$  is the trace of the  $ij^{\text{th}}$  sub matrix  $[\mathbf{H}^{-1}]$  of  $\mathbf{H}^{-1}$ . By performing modal decomposition of the Hessian matrix, the overall motion can be expressed as a sum over the  $(3N-6)$  individual internal fluctuations modes, i.e.

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \sum_{k=1}^{3N-6} [\Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j]_k \quad (3.21)$$

The contribution of the  $k^{\text{th}}$  mode correlation is:

$$[\Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j]_k = \left( \frac{k_B T}{\gamma} \right) \text{tr}[\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T]_{ij} \quad (3.22)$$

where  $k$  is the  $k^{\text{th}}$  nonzero eigenvalue of  $H$  and  $u_k$  is the corresponding eigenvector. Each eigenvalue is related to the frequency of an individual mode and the corresponding eigenvector describes how the positions of the  $N$  sites of the structure are affected as a result.

The mean square fluctuation for the  $i^{\text{th}}$  residue can be given as following in terms of coordinates as a summation over all normal modes:

$$\langle \Delta R_i^2 \rangle = (k_B T / \gamma) \sum_{l=1}^{3N-6} \frac{S_{il}^2}{\lambda_l} \quad (3.23)$$

Here,  $\langle \Delta R_i^2 \rangle$  is the mean square fluctuation of the  $i^{\text{th}}$  residue,  $k_B$  is the Boltzmann constant,  $T$  is the absolute temperature and  $\lambda_l$  is the normal mode frequency (eigenvalue) for the none zero  $l^{\text{th}}$  mode. Gamma ( $\gamma$ ) is used to scale the theoretical fluctuations considering the experimentally observed B-factors. The cumulative fluctuations for  $k$  slowest modes can be calculated as:

$$\langle \Delta R_i^2 \rangle_k = \sum_{l=1}^k \frac{S_{il}^2}{\lambda_l} \quad (3.24)$$

A simple model like coarse grained NMA with ANM is proved in many applications to be a promising tool for describing the collective dynamics of a wide range of biomolecular systems. It can be useful to obtain fluctuations dynamics rather than fully atomistic model because; fully atomistic approaches become inefficient when the system is large or complex (Doruker and Jernigan, 2003).

## 4. RESULTS AND DISCUSSION

In this thesis, the structure-function relationship of the human  $\beta_2$ AR (2RH1) is addressed via two different computational approaches, namely the anisotropic network model (ANM) and molecular dynamics (MD) simulations. Corresponding results and discussions will be presented in two separate sections. For both methods, the effect of the intracellular loop (ICL3) will be presented through a comparison of dynamic properties of “clipped” and “looped” models.

### 4.1. Construction of the Models

As explained in detail in the previous section, the X-ray crystallographic structure of human  $\beta_2$ AR in complex with T4 lysozyme (PDB code: 2RH1) at 2.40 Å resolution (Cherezov *et al.*, 2007) is used as the initial conformation. After removal of T4L, the missing intracellular loop region (ICL3) between residues 321 to 342 is reconstructed via homology modeling. The N-terminus (residues 1 to 28) and the C-terminus (residues 343 to 365) tails are highly disordered, therefore are not visible in the X-ray.

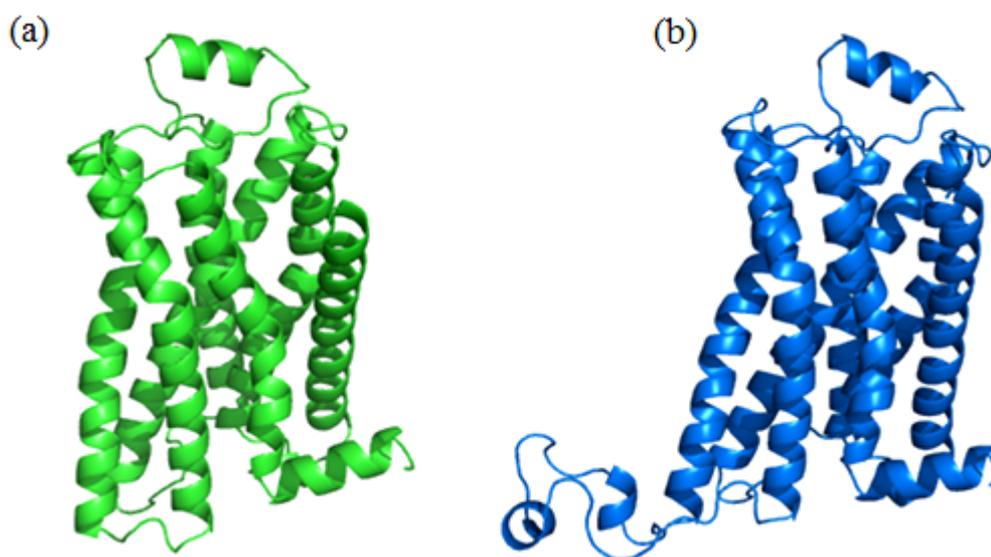


Figure 4.1. Structures of (a) clipped (green), and (b) looped (blue) models.

The structure is completed by homology modeling according to the template structure

(Cholesterol bound form of human  $\beta_2$ AR, PDB code: 3D4S), which is proposed as the best scoring template by MODWEB homology modeling server. Finally, two different structures, which are the clipped and looped models obtained from MODWEB (Figure 4.1), are used to analyze their dynamics.

The end points of the helix V and VI are bounded to each other when clipped model is built because of missing residues of the third loop. Therefore, loop prediction method is used to generate the looped model and missing residues of the third loop (residue 343-365) is completed.

## 4.2. Anisotropic Network Model (ANM) for Monomer Models

In residue-based ANM calculation, the connectivity matrices are constructed using the coarse-grained native structure represented by C-alpha atoms. The cutoff distance, which defines the range of coarse-grained interactions, is taken as  $r_{cut} = 13 \text{ \AA}$  for ANM calculations. The force constant is set to a universal fixed value of  $1 \text{ kcal}/(\text{mol \AA}^2)$ , which is later scaled so as to match the theoretical B-factors with the experimental values. After diagonalization of the Hessian matrix, the mean-square fluctuations (MSF) of residues can be predicted by summing over all modes.

### 4.2.1. B-Factor Profiles

The theoretical B-factors (or temperature factors) are compared with the experimental values from the X-ray structures in order to see the validity of each elastic network model for the clipped and looped monomer conformations. In Figure 4.2, X-ray temperature factors of the template structure and theoretical results from ANM are displayed for both models. The missing residues of the loop are indicated with gaps on the plots. There is generally a good agreement between the B-factor values of ANM and experiments. As expected, the most flexible regions of the clipped and looped model are located in the exposed loop regions of the structures. The regions with B-factor values represent the more restricted parts that correspond to the membrane-embedded helical parts of the structure. Greater mobility is in extracellular loop two (ECL2) regions for both models. This region, which is located between 172<sup>nd</sup> and 196<sup>th</sup> residues, is shown in Figure 4.3 together with other intra and extracellular loops.

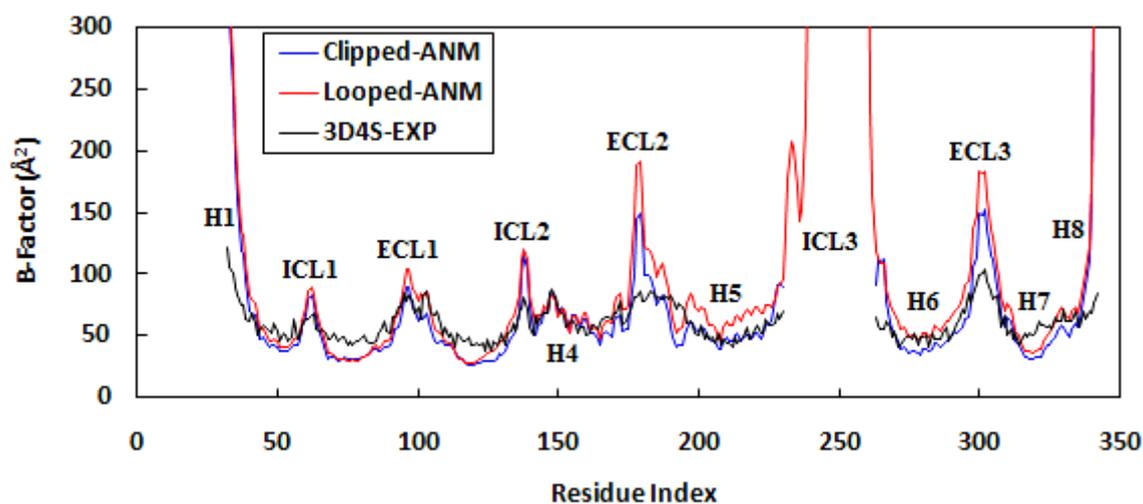


Figure 4.2. Theoretical and experimental B-factor values for clipped and looped models.

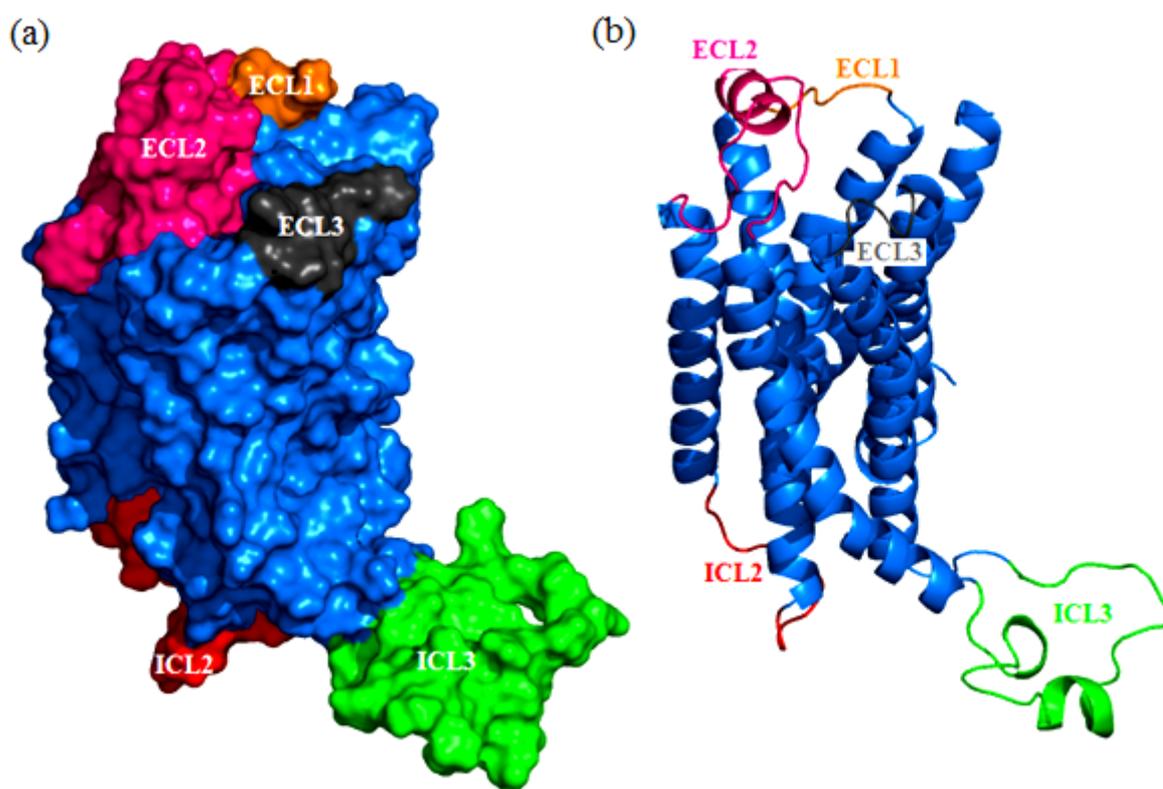


Figure 4.3. Most mobile loop regions of the structure, (a) surface representation, (b) cartoon representation.

The ECLs and amino termini of GPCRs, together with the extracellular halves of the transmembrane helices, define the ligands-binding site of each receptor (Angers *et al.*, 2000). Therefore, the ECLs play an important role in the overall pharmacology of any

particular receptor. In general, small molecule ligands are thought to bind deeper within the space created by the transmembrane domain helices, whereas larger ligands such as peptides bind closer to the membrane surface near the ECLs (Gether *et al.*, 2000).

Another extracellular loop (ECL3), located between the sixth and seventh helices, is shown in Figure 4.4. H6 and H7 have a broken in the middle of the structure because of having proline residue that is capable of breaking the helical structure, located in the middle parts of these helices, and conserved in evolutionary terms.

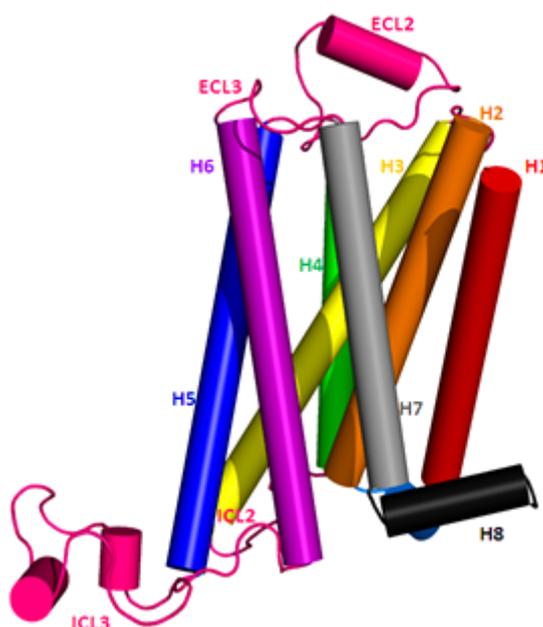


Figure 4.4. All transmembrane helices and loop regions of the looped model.

Conformational changes of thought to be made possible by this broken part of the receptor during the switching activated receptor forms. Therefore, it is not a coincidence that the region, which has a maximum mobility, is between the two helices.

#### 4.2.2. Collective Modes

The low-frequency (slow) modes give insights about the cooperative conformational motions of proteins around their native state. These motions are relevant to biological functions of many enzymes and proteins (Bahar *et al.*, 1998). These modes also contribute most to the overall dynamics of the structure among the  $(3N-6)$  normal modes.

In Figure 4.5, MSF values are calculated based on the cumulative action of the first ten slowest modes. Since B-factor and MSF are correlated, the highly mobile loops also exhibit high MSF in the slow modes. The functionally important parts of the structures, namely the intracellular (ICL1, ICL2, ICL3) and extracellular loop regions (ECL1, ECL2, ECL3) regions are highly mobile parts of the structures. ICL3 shows the highest MSFs among all loops in the looped model due to its conformational flexibility. Because of the dominant behavior of ICL3, the MSF values for other loop regions are not pronounced.

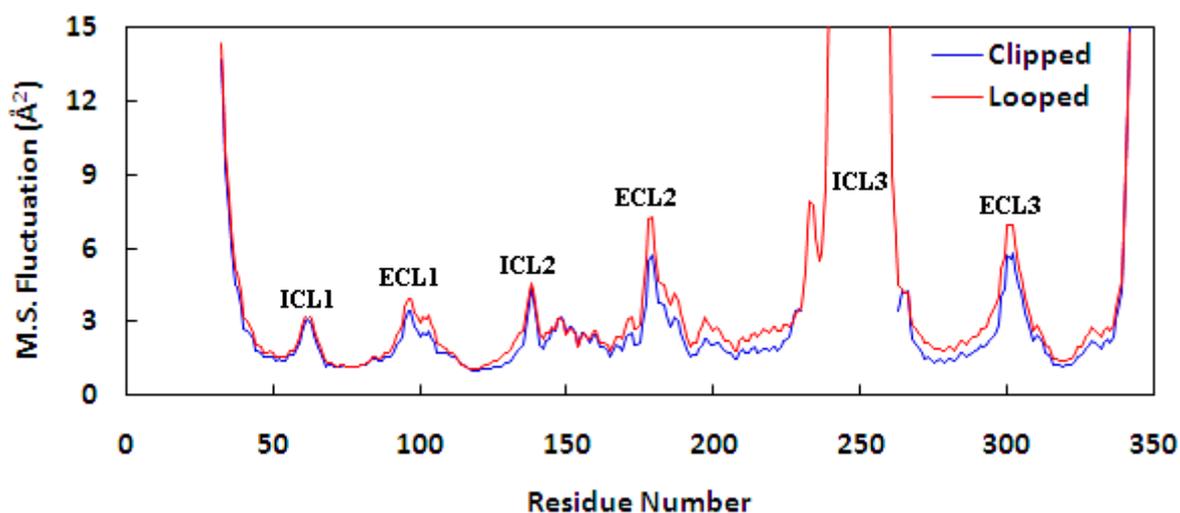


Figure 4.5. Mean square fluctuations based on the cumulative action of 10 low-frequency modes for looped and clipped models.

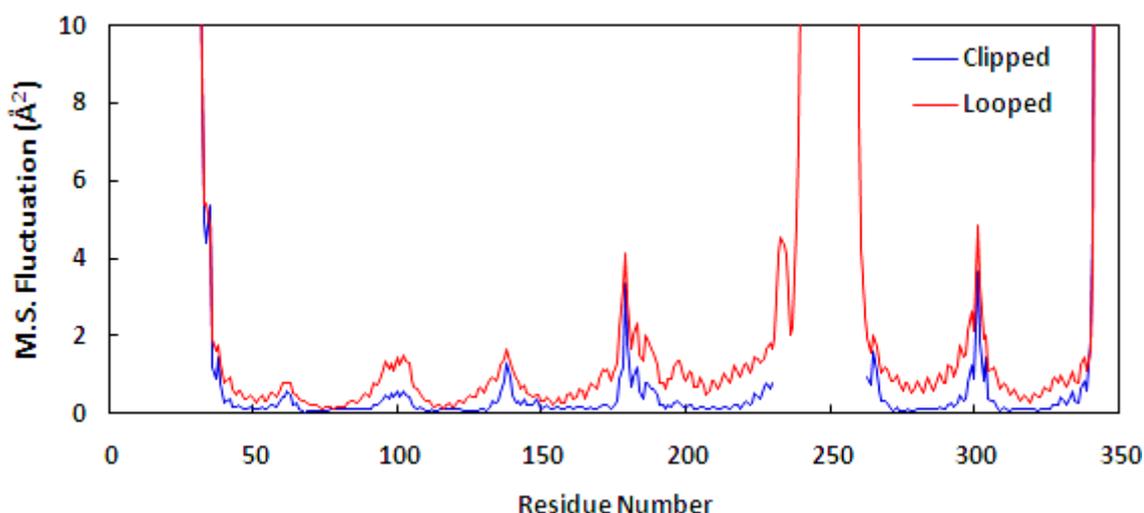


Figure 4.6. MSF profile is calculated by taking between the fifth and fourteenth slowest modes for looped model and the first 10 slowest modes for clipped model.

MSF profile is calculated by taking the first 10 slowest modes for both models in Figure 4.5, but the first four modes of looped model are not available in clipped model because they represent the movement of ICL3 only. Therefore, to compare it with the clipped model as shown in Figure 4.6, MSF values of the looped model are recalculated by considering the modes between the fifth and fourteenth.

### 4.2.3. Overlap Matrix

To observe the correspondence between the slow modes of the looped and clipped models, the overlap matrix is calculated. The level of agreement between the collective motions of clipped and looped model is calculated using the following overlap definition (Amadei *et al.*, 1993; Hess, 2000).

$$u_{jk}^{AB} = \mathbf{u}_j^A \cdot \mathbf{u}_k^B \quad (4.1)$$

Here A and B indicate different eigenvector sets from independent ANM calculations, and subscripts  $j$  and  $k$  indicate the slow mode indices of each eigenvector. The limits of  $u_{AB}$  are 0 and 1, which indicate random directions and perfect overlap, respectively. In Figure 4.7, the similarity between low-frequency modes for clipped and looped models is displayed using blue and red scale (red indicated perfect overlap).

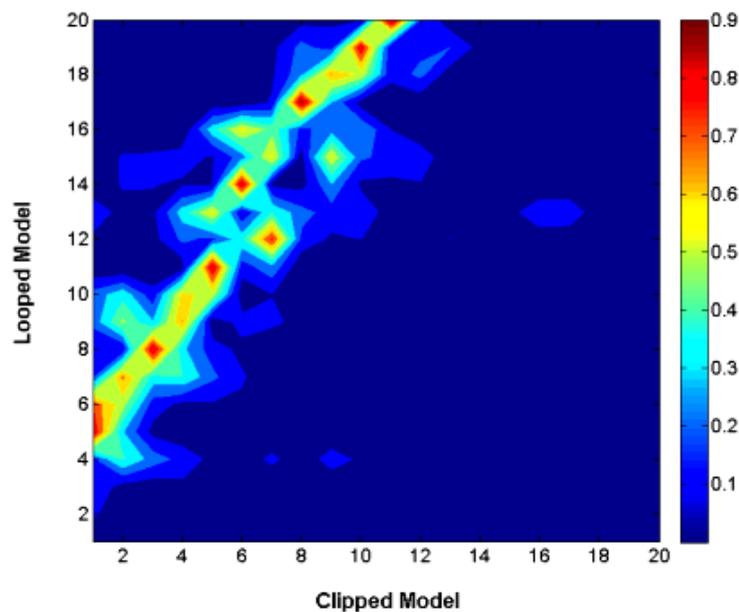


Figure 4.7. Overlap between low-frequency modes of the clipped and looped models.

The first four modes of looped model are unique and do not exhibit any correspondence with the slow modes of clipped model. Similarities between modes of the clipped and looped models appear after the fourth mode of looped model. For instance, fifth and eighth modes of looped model are well compatible with the first and third modes of clipped model, respectively.

The movements of some corresponding modes of the clipped and looped models are shown with the directions and magnitudes of eigenvectors in Figures 4.8 and 4.9. The first low-frequency mode of clipped model and the fifth low-frequency mode of looped model are shown in Figures 4.8a, b respectively. Similarly, the third mode of clipped model and the eighth mode of looped model are shown in Figure 4.9. Furthermore, the first four unique low-frequency modes of the looped model are shown in Figure 4.10a, b and 4.11a, b. As expected, the first four unique modes reflect the effect of the ICL3 loop on the overall dynamics of the protein.

The core region of the receptor consists of two domains because some transmembrane helices such as TM6 and TM7 have a broken in the middle of the structure because of having proline residue that is capable of breaking the helical structure, located in the middle parts of these helices. In the first mode of clipped and the fifth mode of looped model, the region, where is the middle of the receptor, is acting as a hinge region (shown as black dashed line) and at that point motions of the first and fifth modes of clipped and looped model describe a counter-rotation of the upper and lower parts of the receptor named as twisting motion can be observed (Figure 4.8a, b).

ICL3 region of the looped model highly mobile and it moves independently from the core region of the looped model. Also, TM1 is highly mobile and turning opposite to the rest of the receptor in these modes. Similarly, the same motion is observed in these modes of the models for TM8. In the third mode of clipped model and the eight mode of looped model, hinge-bending motion of the upper and the lower parts of the receptor towards each other is observed (Figure 4.9a, b). Hinge bending motion of thought to be made possible by this broken part of the receptor. This hinge region of the receptor is shown as black dashed line in Figure 4.9a, b. Also, domain motions, such as hinge bending or closure, are known to play an important role in the function of many proteins (Hayward, 1999).

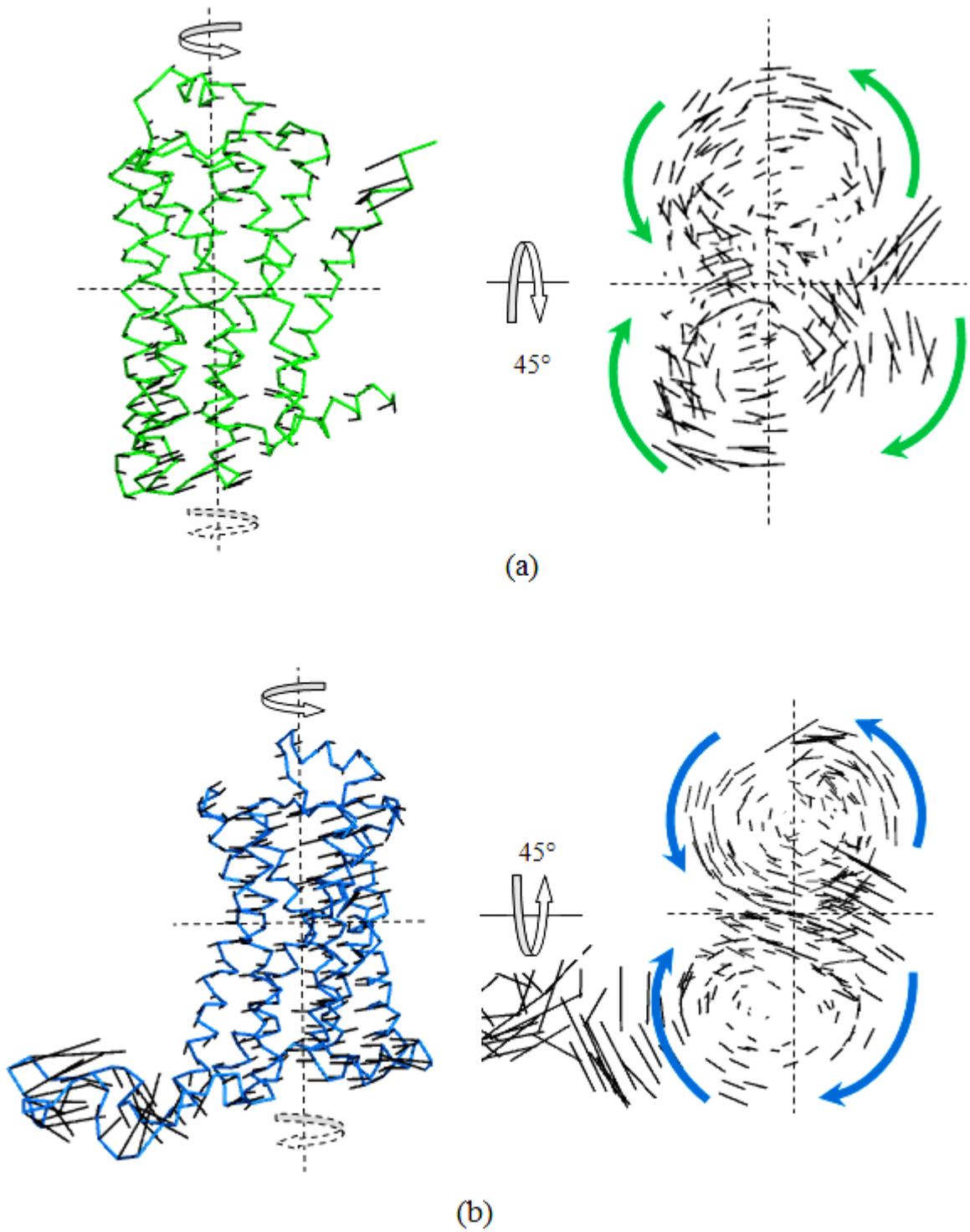


Figure 4.8. The directions and magnitudes of eigenvectors of the models for (a) the first slow mode of clipped model, (b) the fifth slow mode of looped model.

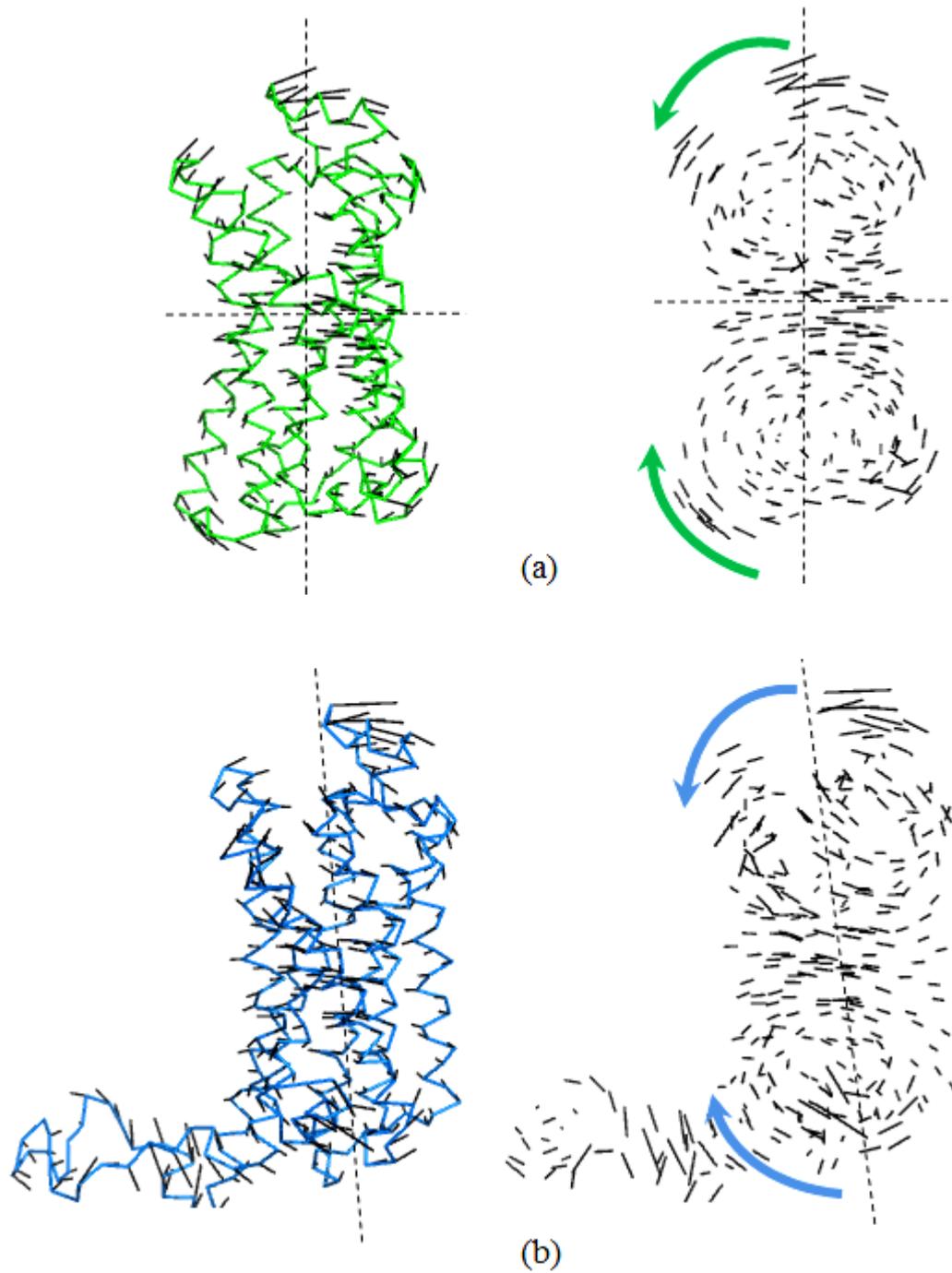


Figure 4.9. The directions and magnitudes of eigenvectors of the models for (a) the third slow mode of clipped model, (b) the eighth slow mode of looped model.

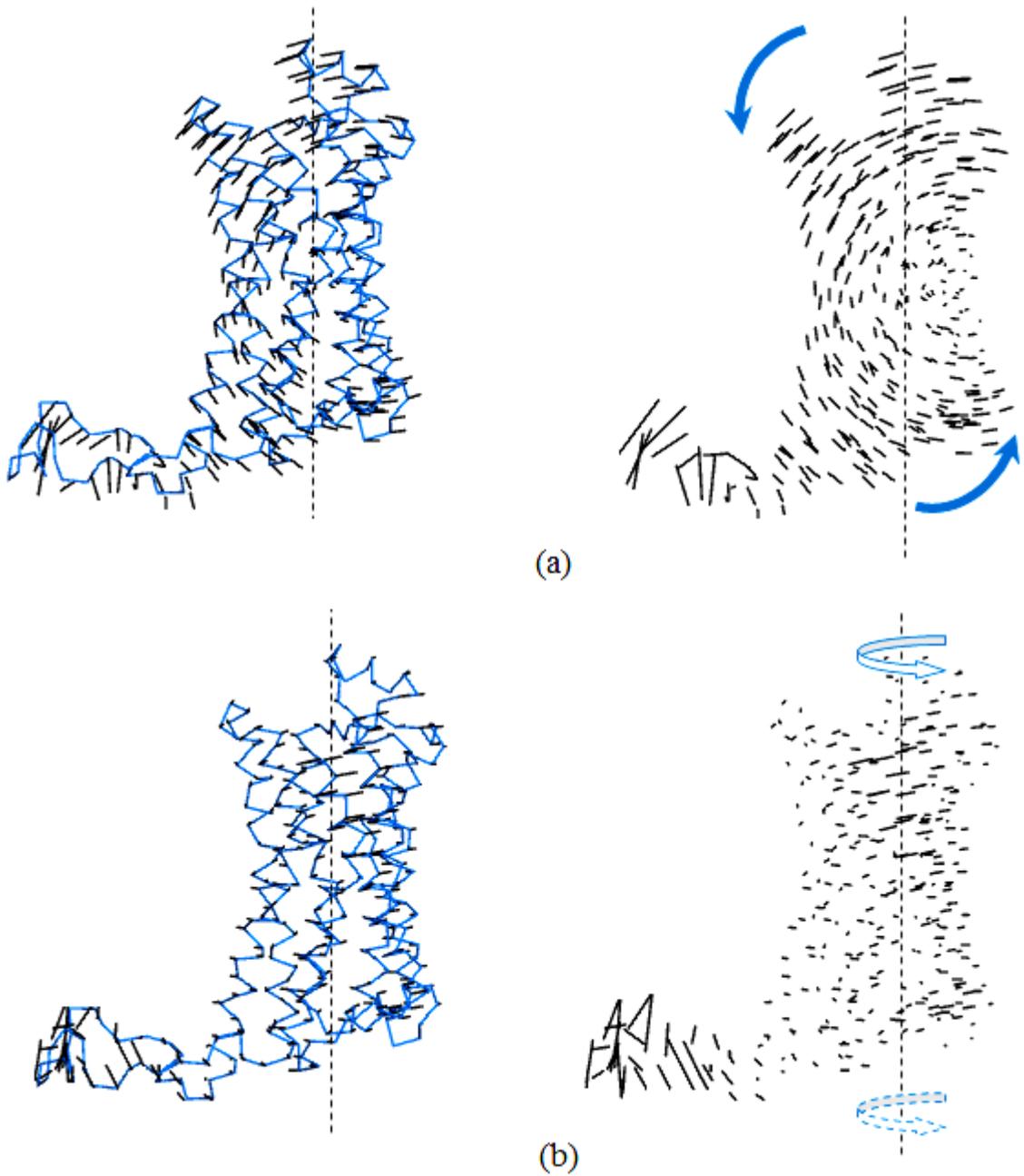


Figure 4.10. The directions and magnitudes of eigenvectors of the low-frequency modes of looped model for (a) the first mode, (b) the second mode.

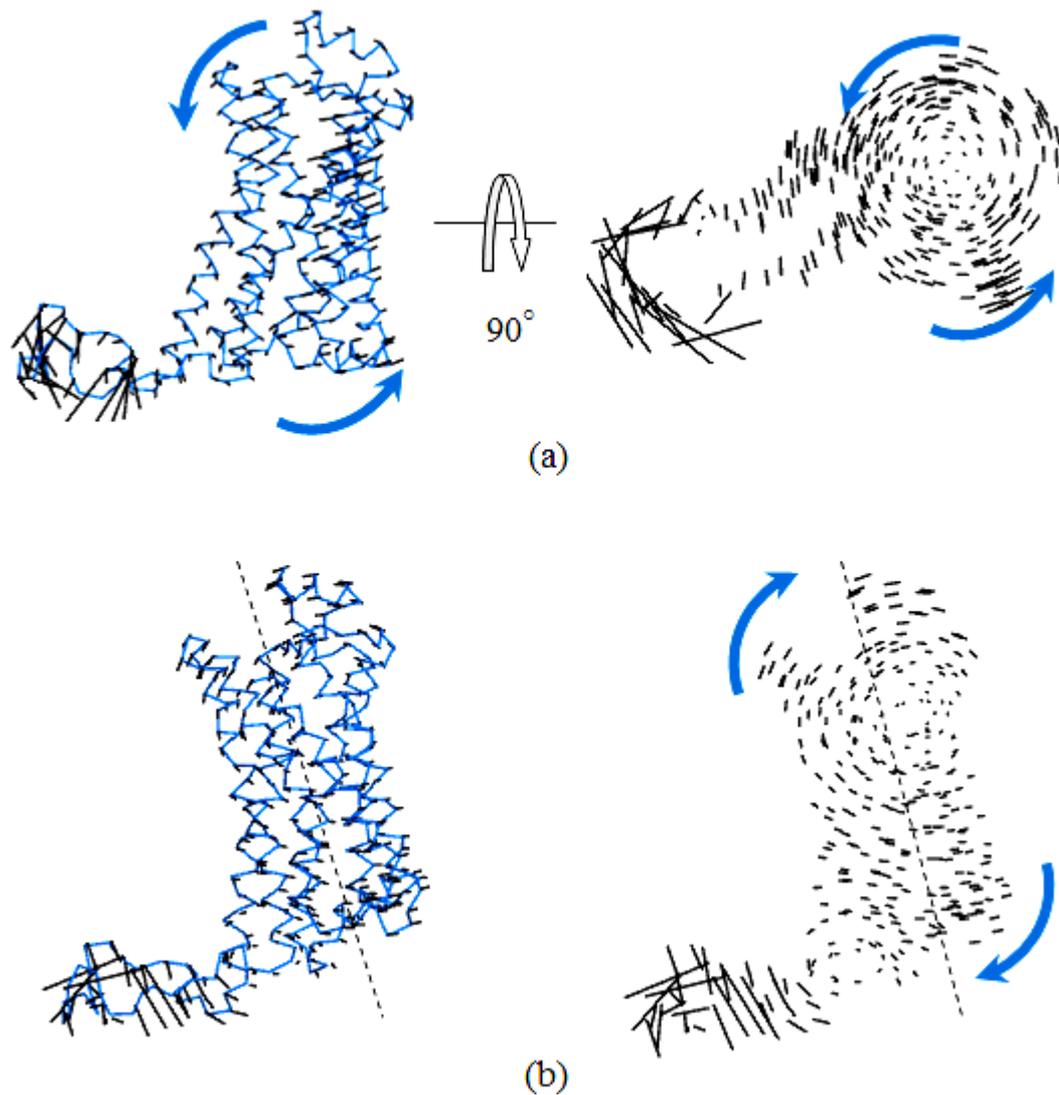


Figure 4.11. The directions and magnitudes of eigenvectors of the low-frequency modes of looped model for (a) the third mode, (b) the fourth mode.

#### 4.2.4. Orientational Cross-Correlations

The normalized orientational cross-correlations,  $C(i,j)$ , between residue fluctuations are calculated by the following equation

$$C_{(i,j)} = \frac{\langle \Delta R_i \cdot \Delta R_j \rangle}{[\langle \Delta R_i \cdot \Delta R_i \rangle \langle \Delta R_j \cdot \Delta R_j \rangle]^{1/2}} \quad (4.2)$$

The cross-correlations vary within the range  $[-1, 1]$  with the lower and upper limits

indicating respective fully anti-correlated and correlated fluctuations, whereas “0” indicates uncorrelated fluctuations in terms of orientation. The gaps are used to show the missing residues when the cross correlation maps are plotted. In all cross-correlation figures, the positively correlated fluctuations are shown in shades of red, the negative correlations are shown in shades of blue for clarity.

In Figure 4.12a, the cross-correlation plot using the first ten slow modes are shown for the clipped model, where all helices (H1 to H8) act positively correlated with each other. In contrast, the correlation pattern of helices seems to be concentrated in specific helical pairs of looped model in Figure 4.12b. These helical pairs are H1-H2, H1-H7, H3-H5, H3-H6, H4-H5, H5-H6, H5-H7, H6-H7 and H6-H8.

As explained previously, the first four modes of the looped model are unique in that they represent the movement of intracellular loop three (ICL3). Resultingly, the cross-correlations between helical regions are either weakened or strengthened compared to the clipped model. In order to compare the correlation maps of the two models based on the matching ten modes of both models, the cross-correlation map is re-created by taking the cumulative effect of modes five through fourteen for the looped model. Comparison of Figures 4.12a and 4.13 indicate that the correlation values of helical regions are similar and the correlated pairs concentrate in the same areas for both models. Accordingly, the intrinsic dynamics of the helical regions, which provide the signal transduction, are not affected significantly by ICL3 region when the extra modes of the looped model are excluded.

To better understand the consistency between the compatible modes of clipped and looped model, orientational cross-correlation maps are calculated by taking the difference between the cross-correlation values of clipped model and the cross-correlation values of looped model (See Figure 4.14). The difference cross-correlation map is shown in Figure 4.14 and the difference cross-correlation values vary within the range  $[-2, 2]$  with the lower and upper limits. As previously mentioned, the correlations between transmembrane helices are compatible with each other and the distributions of the correlated helical pairs on the difference cross-correlation map concentrate in almost the same areas for both models. Although these transmembrane helices help to transmit a signal to the intracellular

part of the receptor, this movement is more specific for this helical region. In addition, it can be said that the movement is independent from the intracellular loop three (ICL3) which is part of the protein facing the cytoplasm and allows the interaction with G-protein.

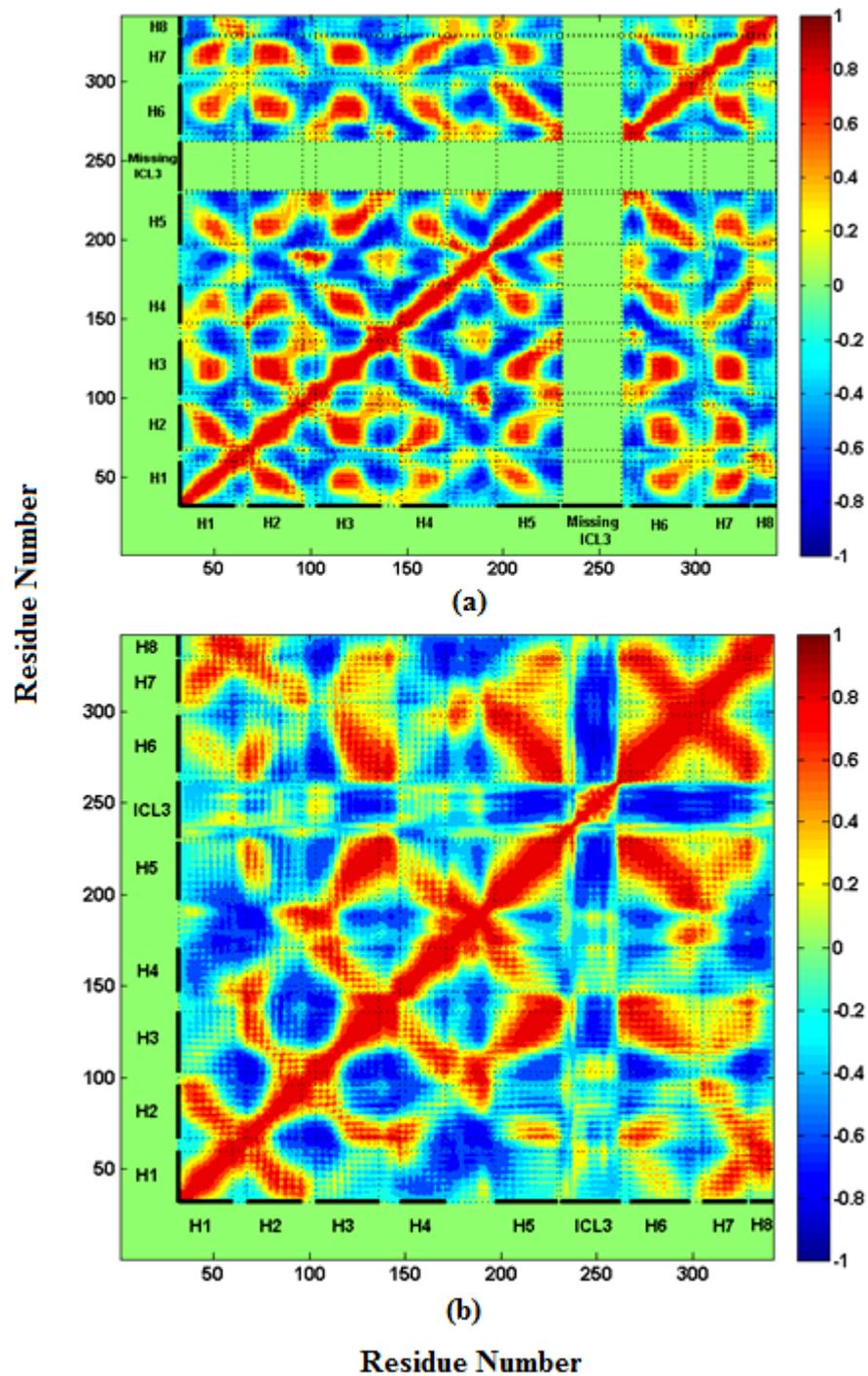


Figure 4.12. Orientational cross-correlation map for (a) the first ten modes of clipped model, (b) the first ten modes of looped model.

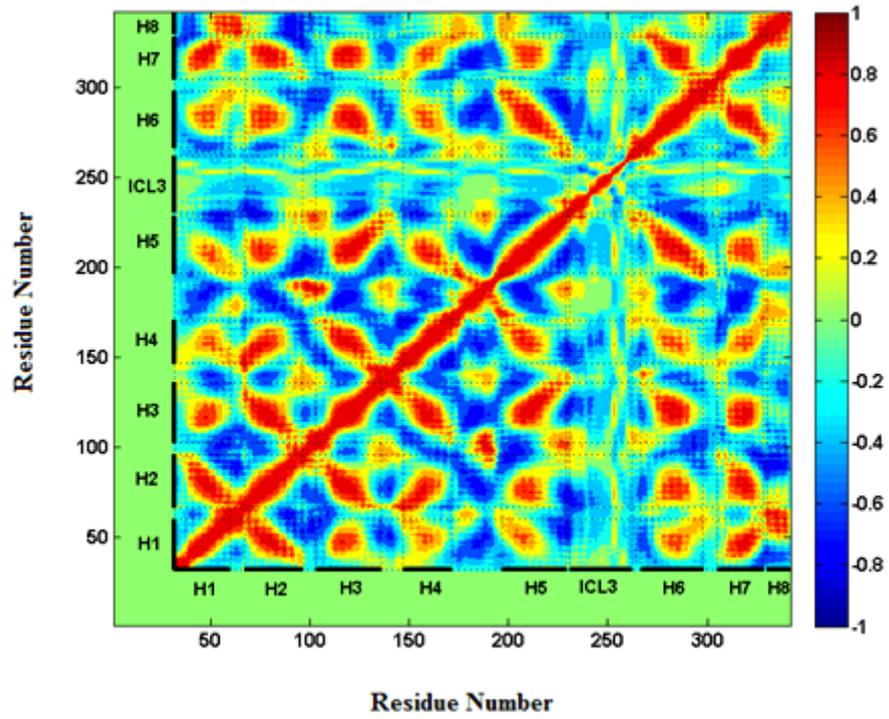


Figure 4.13. Orientational cross-correlation map based on modes five through fourteen for looped model.

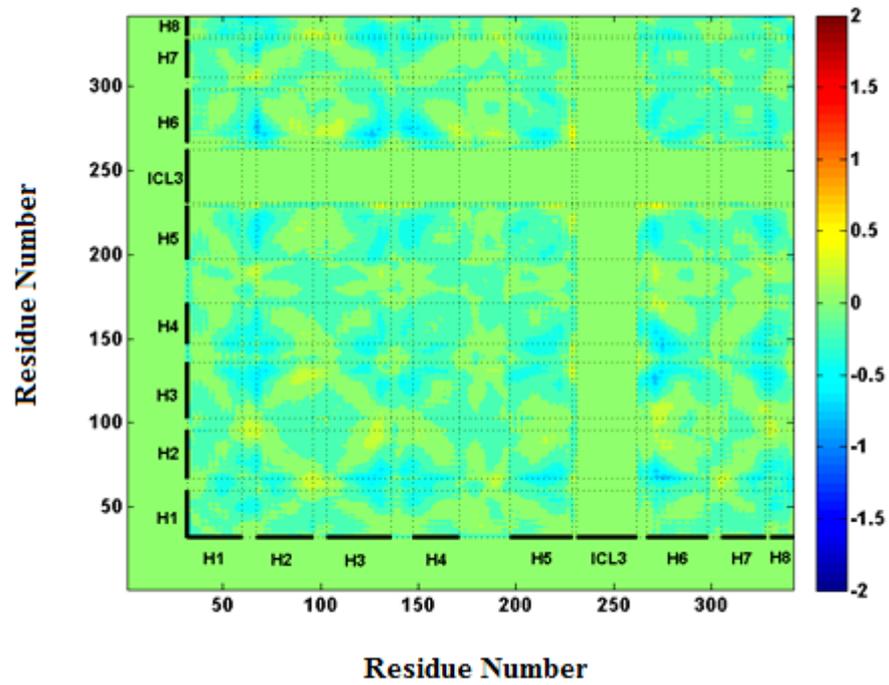


Figure 4.14. Difference cross-correlation plot of clipped and looped models.

#### 4.2.5. Fast Modes (Hot Spots)

In elastic network models, the residues with high MSF in the fast modes are termed as hot spot residues (Bahar *et al.*, 1998; Haliloglu *et al.*, 2005). These highly fluctuating residues in the fast modes, which are important for the stability of the protein (Bahar *et al.*, 1998), are found to be evolutionarily conserved residues, thus they are important for mutation studies (Hu *et al.*, 2000; Ma *et al.*, 2003; Keskin *et al.*, 2004; Haliloğlu *et al.*, 2005).

Figure 4.15 shows the MSF profiles, which are calculated by taking the 20 fastest modes for both models. As seen in the MSF plot, looped model fluctuates almost in the same range with the clipped model. However, high MSF values for high frequency modes do not indicate that the more flexible regions. The residues located on the peaks generally play a critical role to protect and maintain the stability of the structure.

Phe282 has a high MSF value, approximately 0.01, as seen in Figure 4.15a. This residue is located on the sixth helix and plays an important role in the interaction of the third and the sixth helices. As a result of the mutation of this residue to glycine, asparagine, alanine, or leucine, the interaction between these two helices becomes weak and then the protein is activated (Chen *et al.*, 2000). Similar situation also applies to Asp130, Arg131 and Glu268 residues. The salt bridge between Arg131 and Glu268 strengthens the interaction between the third and the sixth helices. Also the third helix is more stable by another salt bridge which is between two neighboring residues, Asp130 and Arg131 (Rasmussen *et al.*, 1999).

As shown in Figure 4.15a, the MSF values of the critical four residues (Asp130, Arg131, Glu268 and Phe282) are more than 0.004. Only the MSF of Glu268 is lower than expected for both model. Nevertheless it is near the region which has high MSF values. In Figure 4.15b, four residues (Leu272, Leu284, Gly276 and Gly280) are displayed on the same MSF profile. They belong to the following sequence **LKTLGIIMGTFTL**, which is suggested to be the evolutionarily conserved dimerization motif involved in monomer-monomer interaction at the interface (Hebert *et al.* 1996).

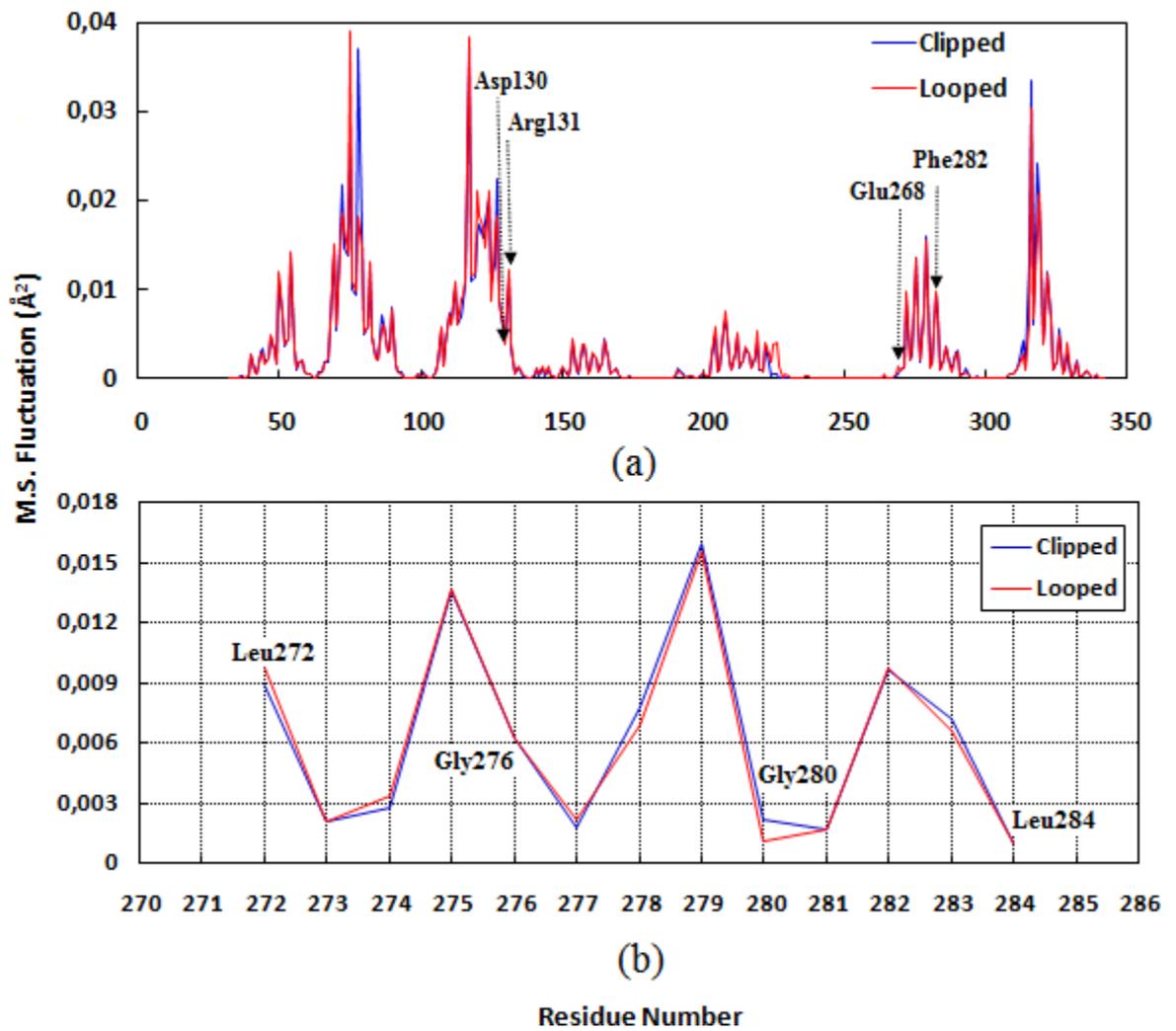


Figure 4.15. Normalized MSF profile is calculated by taking the 20 high-frequency modes of the models for (a) Asp130, Arg131, Glu268 and Phe282 (b) Leu272, Gly276, Gly280 and Leu284 residues are shown on the profile.



Figure 4.16. The position of the hot spot residues on the looped model.

### 4.3. Anisotropic Network Model (ANM) for Dimer Models

In this section, residue-based ANM calculation is applied to the proposed dimeric structures of the clipped and looped models. The connectivity matrices are constructed using the same cutoff distance of 13 Å as for the monomer case.

#### 4.3.1. Generating The Dimeric Structure of $\beta_2$ AR

The interaction between symmetry-related receptor was proposed as potential physiological dimer interfaces in the crystal structure of  $\beta_2$ AR (Cherezov *et al.*, 2007). There are four distinct crystal-packing interactions within each layer according to the studies of Cherezov. Three of these are mediated by T4-lysozyme. The fourth interaction is between two receptor molecules related by a crystallographic two-fold rotation axis. This is the interaction between symmetry-related receptor, and the direct receptor contact involves a 2.7 Å pair of ionic interaction between the charged amine group of Lys60 in helix I and the carboxylate of Glu338 in the helix VIII of the symmetry-related receptor (Cherezov *et al.*, 2007).

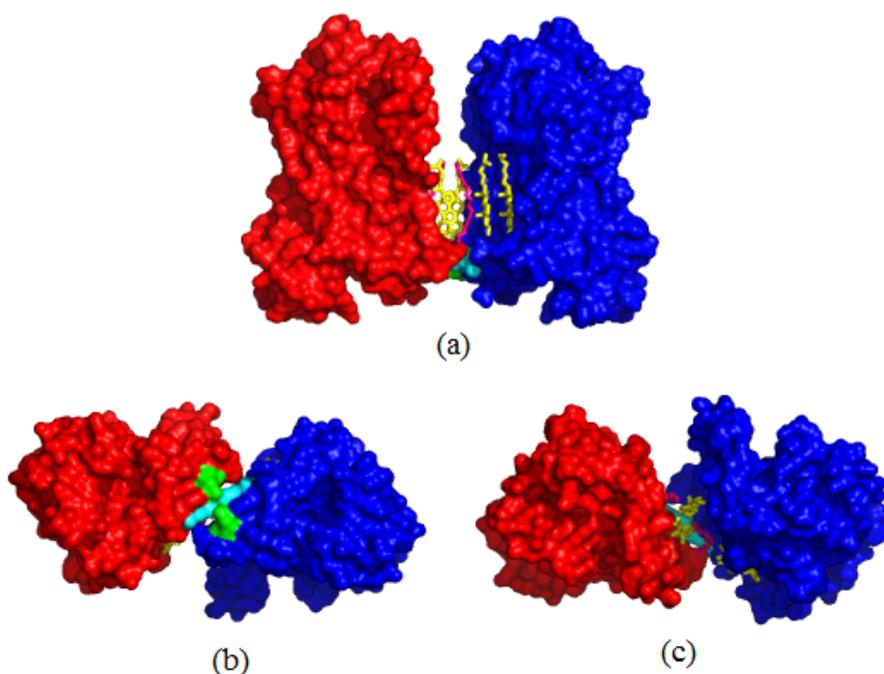


Figure 4.17. The best symmetric mate of the monomer with cholesterol (yellow) and palmitic acid (pink) for looped model: (a) Front view, (b) ICL view and (c) ECL view.

These studies revealed that the biological dimer possesses a symmetrical interaction where helix I and helix VIII of monomer A interact with helix I and helix VIII of monomer B. In this thesis, the PyMOL Molecular Graphics System (Delano, 2002) was used to generate the best symmetric mate of the monomer for both clipped and looped models according to the interaction distance of helix VIII. Hence, as shown in Figure 4.17 and 4.18, there are 4 Å spatial differences between the generated symmetric mates with cholesterol and palmitic acid for looped model and the main monomer structures. Cholesterol has very significant role on the structure. For example it influences on membrane fluidity and curvature on the macro level and also it increases the packing interactions for helix II-IV and thermal stability. However, cholesterol is not in the packing interface which means they are structurally relevant cholesterol-binding site between helices 1, 2, 3 and 4.

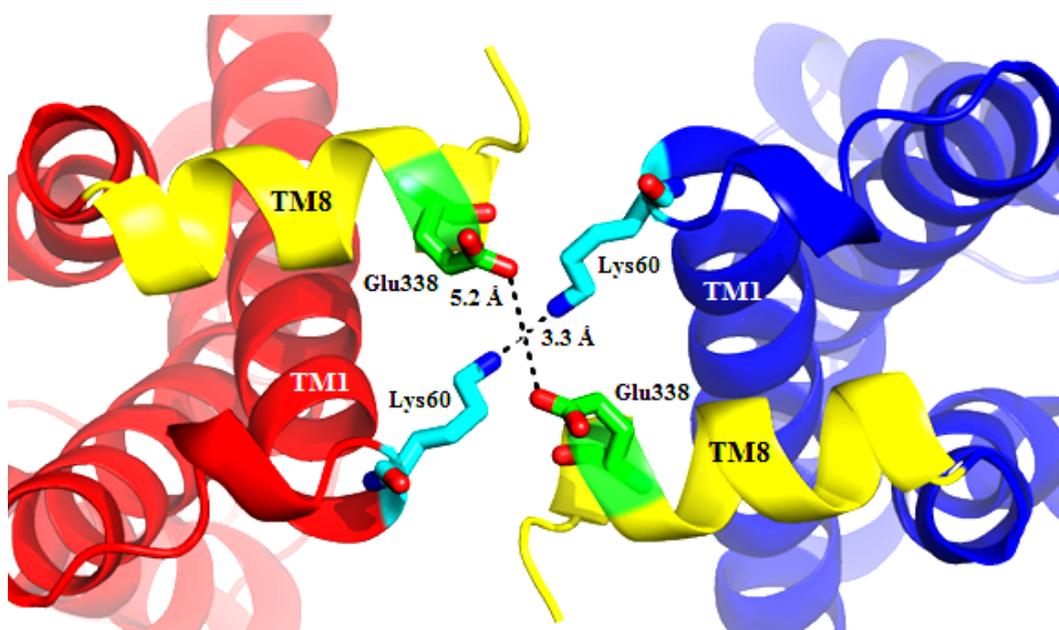


Figure 4.18. Ionic interaction between the charged amine group of Lys60 in helix I and the carboxylate of Glu338 in the helix VIII of the symmetry-related receptor.

#### 4.3.2. B-Factor Profiles

B-factor values of the dimer models are compared to each other in Figure 4.19 in order to determine the differences between the dynamics of monomer and dimer models. The most flexible regions of the dimeric structures of clipped and looped model such as

extracellular loop II (ECL2), the end points of helix VIII (TM8), helix I (TM1) for the both models and intracellular loop III (ICL3) for the looped model locate in highly mobile parts of the structures. However, the movement of TM8 is not observed clearly from the dimeric structure of clipped and looped models. While the B-factor value of TM8 is more than 200 Å for the monomer models, this value is between 50 to 100 Å for the dimeric models. Since TM8 locates at the interface of the symmetrical interaction of two monomers, its movement is restricted.

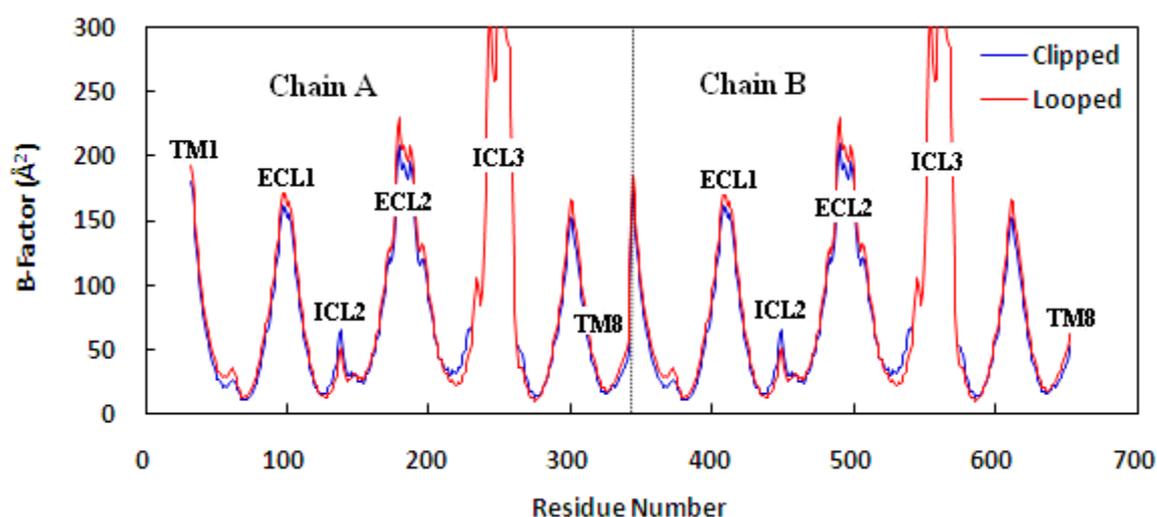


Figure 4.19. Comparison of B-factor values of two dimer models.

#### 4.3.3. Slow Modes (Collective Deformation)

These flexible residues play a very important role for the cooperative conformational motions of the structures. According to the slow modes of the dimer models in Figure 4.20, the most flexible residues are in the same regions for both models. These residues locate at the ends of transmembrane helices (TM1, TM7) and the extracellular loop II (ECL2) for both dimer models and at the intracellular loop III (ICL3) for looped dimer models (See Figure 4.21). The range of the mean square fluctuation values display similarities for slow modes of both dimer models excluding generated loop region. Moreover, the slow modes of the monomer models are in shown with dimers at the same plot in order to see the dimerization effects on mobile regions.

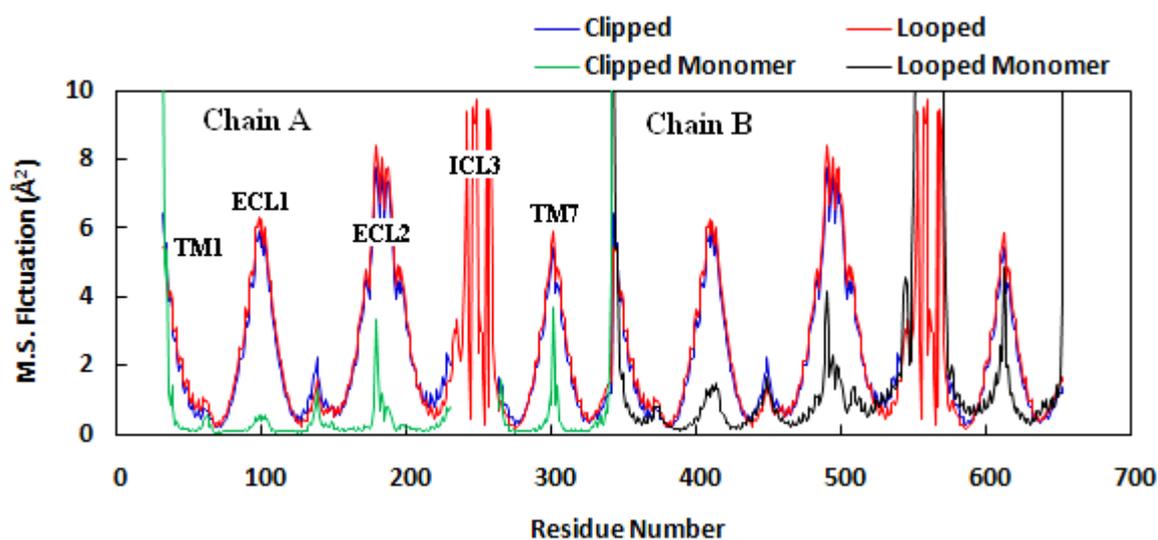


Figure 4.20. Cumulative 10 low-frequency modes for the dimer models.

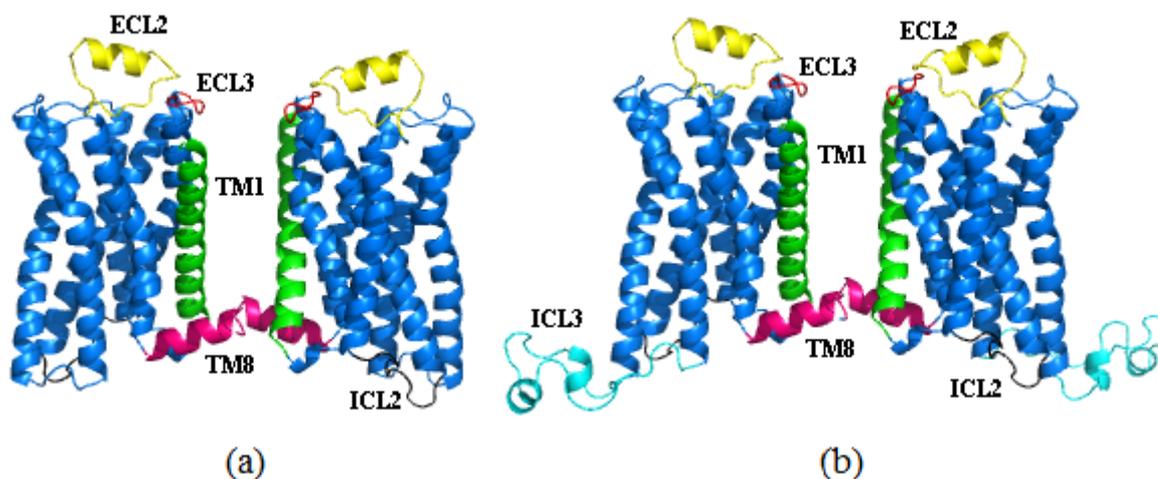


Figure 4.21. Most flexible regions of both dimers for (a) clipped dimer model, (b) looped dimer model.

#### 4.3.4. Overlap Matrix

In this part, the overlap matrix (Figure 4.22) of the dimer models is evaluated using Equation 4.1. Accordingly, the first and second modes of clipped and looped dimer models are compatible with each other. Thus, the eigenvector representations of these modes are shown in Figures 4.23 and 4.24.

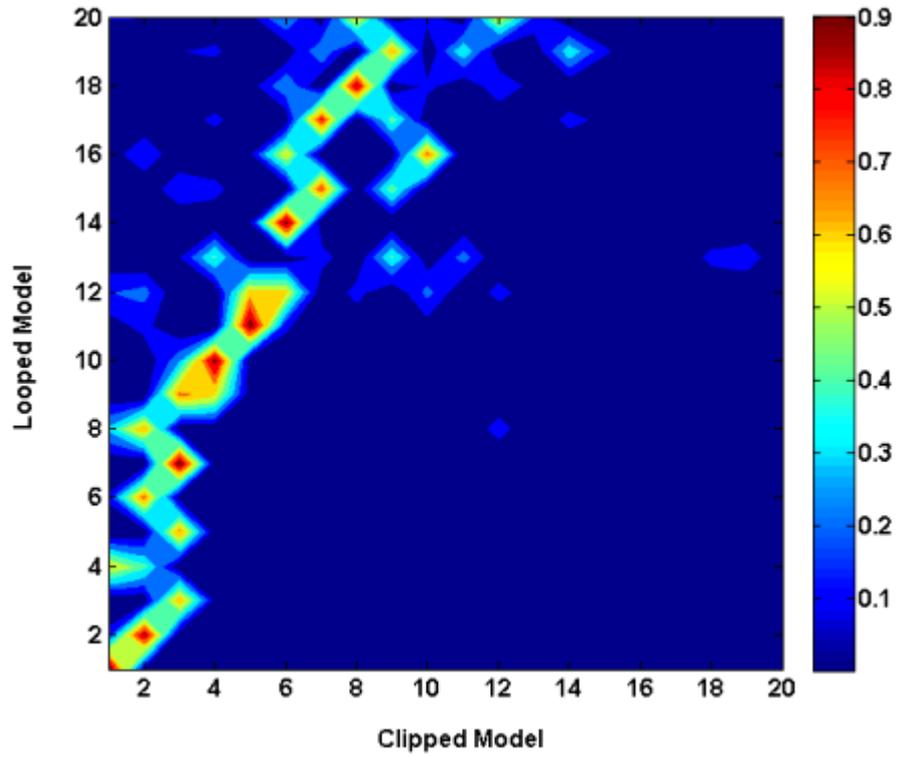


Figure 4.22. Overlap matrix of the two dimer models.

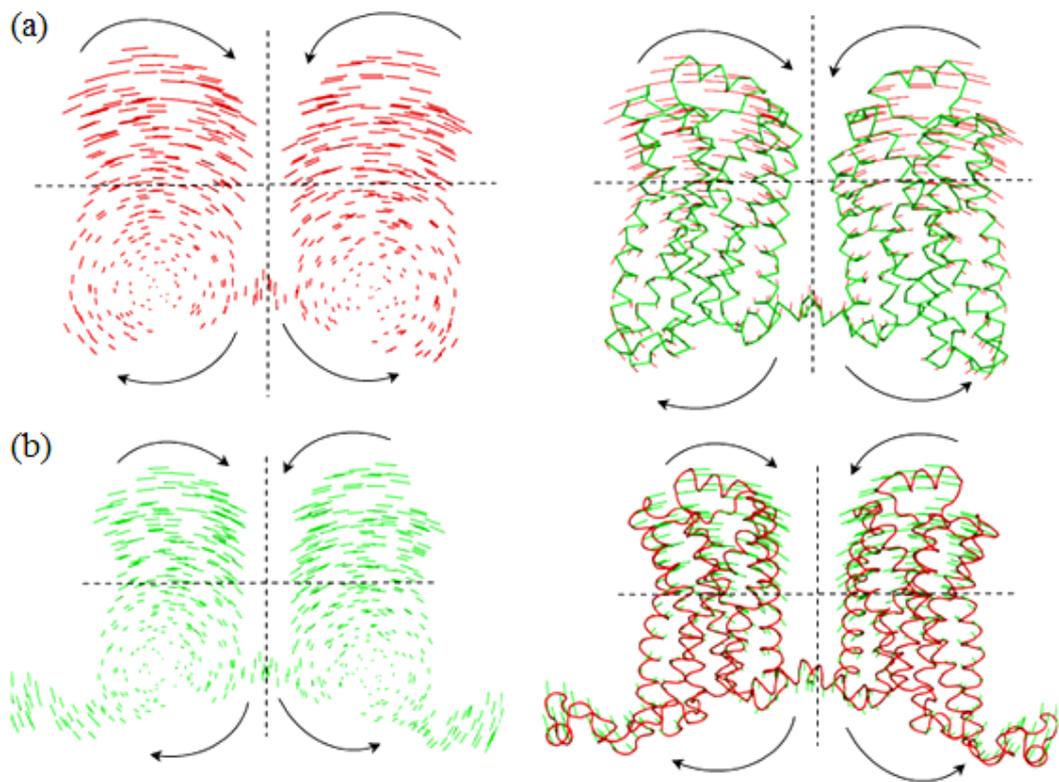


Figure 4.23. Eigenvector representation of the first modes for (a) clipped and (b) looped models.

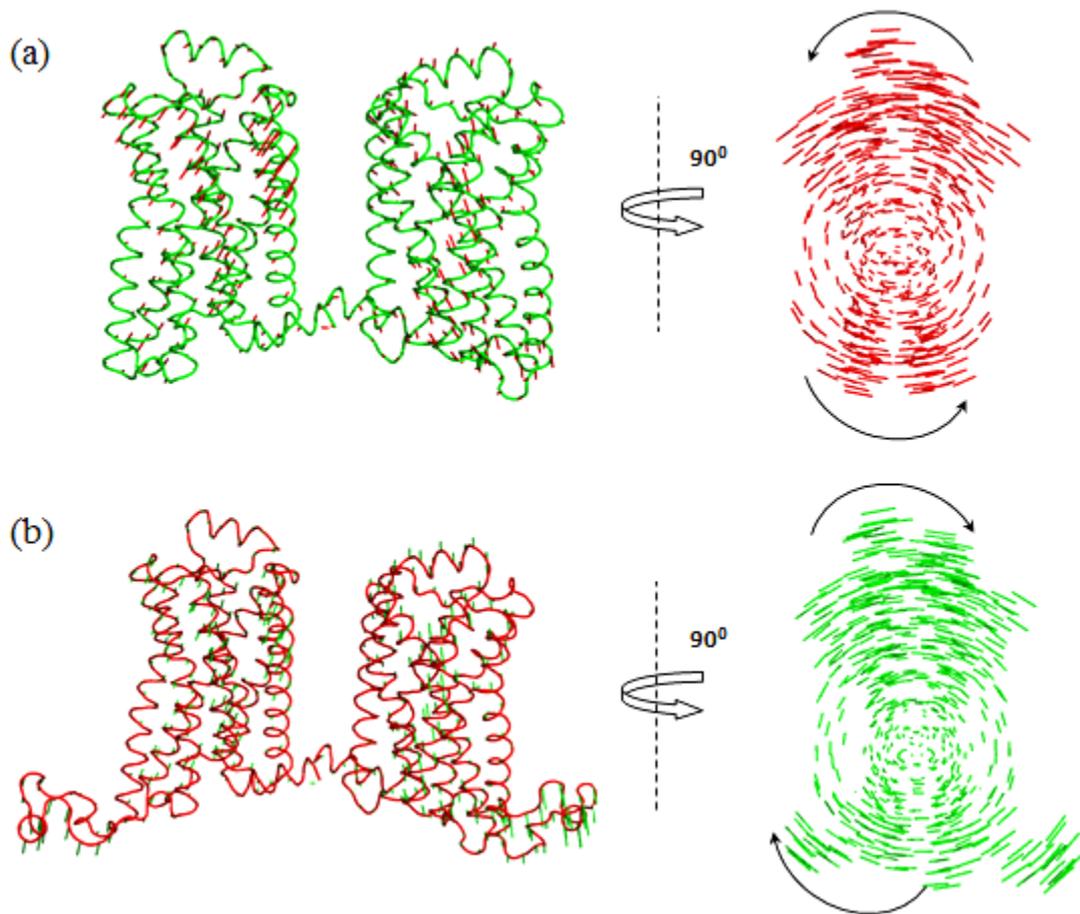


Figure 4.24. Eigenvector representation of the second modes for (a) clipped and (b) looped models.

To better understand this difference, one monomer of the dimeric structures was randomly selected and the eigenvectors of low-frequency modes of this monomer were extracted in order to compare the level of agreement with the modes of monomer models (See Figure 4.25). Accordingly, the first two modes of dimer looped model are not compatible with the first two modes of the monomer looped. These two modes of the dimer looped model represent unique conformations of dimeric structure. A similar analysis is made for the dimer clipped model (See Figure 4.26). Hence, the first four modes of the dimer clipped model do not overlap with any modes of the monomer clipped. Consequently, the effect of the dimeric structure is more pronounced in clipped model and more specific conformations are observed in dimeric structure of clipped model according to the unique conformation of monomer clipped model. This result can be explained as an effect of the absence of the intracellular loop III (ICL3) in clipped model.

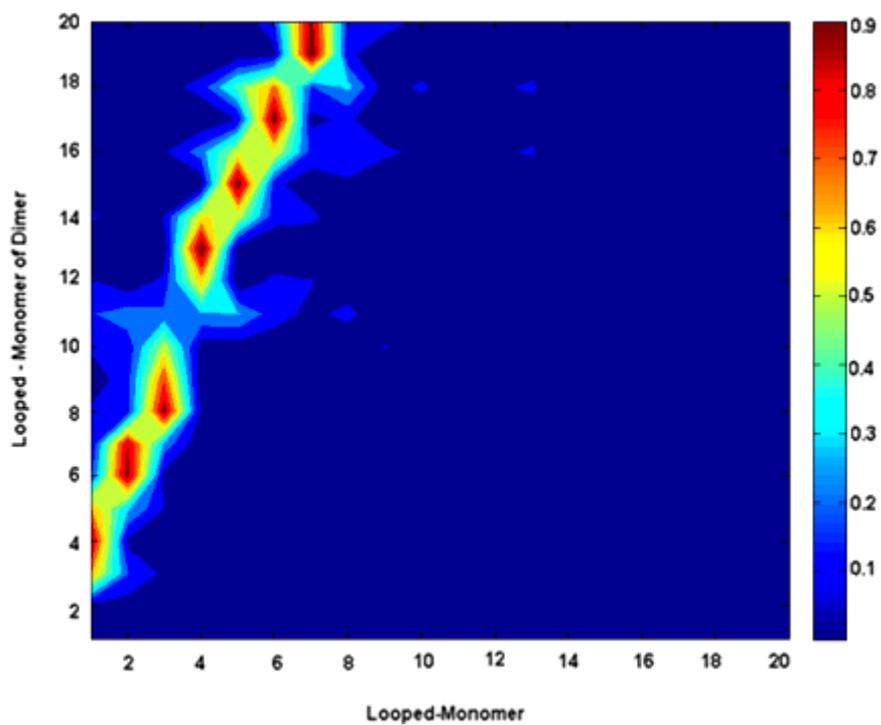


Figure 4.25. Overlap matrix between the monomer and dimer structures for the looped model.

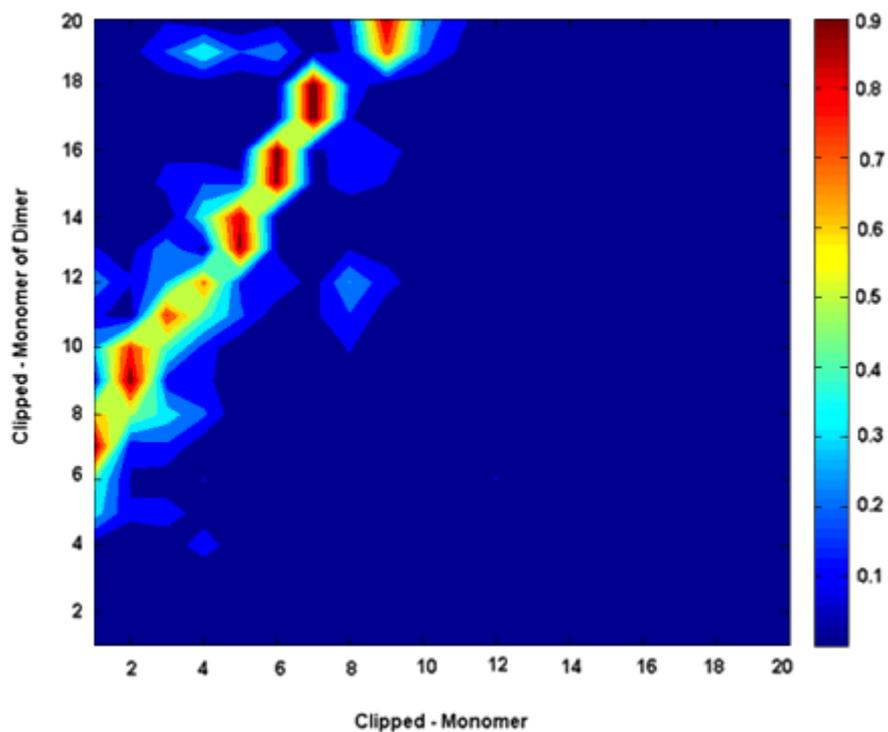


Figure 4.26. Overlap matrix between the monomer and dimer structures of the clipped model.

#### 4.4. Molecular Dynamics (MD) Simulations

To facilitate crystallization of the ‘clipped’ model, 187<sup>th</sup> residue asparagine (Asn) has been replaced with glutamic acid (Glu) (Cherezov *et al.*, 2007). Although it is estimated that a single residue will not affect the dynamics of the system significantly, the clipped system without mutation, named as ‘natural’ in this thesis, was regenerated and exposed to an MD simulation of 500 ns. The ‘mutated’ clipped model was also simulated for 500 ns. And another simulation of 800 ns was carried out for the loop model without any mutation.

##### 4.4.1. Simulation Details

MD simulations were carried out using the Nanoscale Molecular Dynamics (NAMD) software package, which can be parallelized for high-performance simulations of large biomolecular systems. For the interaction potentials, the CHARMM27 force field was used. As explained above, three independent runs were performed for apo  $\beta_2$ AR at 310 K: (i) 800 ns for the looped model, (ii) 500 ns for the clipped model (natural) and (iii) 500 ns for the clipped model (mutant). The apo structure of human  $\beta_2$ AR (PDB code: 2RH1) with 2.4 Å resolution was used as the starting conformation. The  $\beta_2$ AR-clipped models were generated simply by deleting the T4-lysozyme residues and by connecting the exposed termini of Leu230 and Lys263. The loop region of the looped model was generated via homology modeling after removal of the T4-lysozyme. The simulation details are summarized in Table 4.1 and the procedure is as follows:

Table 4.1. Simulation system details.

Run	Simulation length (ns)	Box Dimension (xyz)	Number of atoms in protein	Number of water molecules	Number of lipids	Number of ions	Total number of atoms
Looped	800	86x86x100	5.055	42.135	20.770	41	68.001
Clipped	500	77x69x90	4.551	24.726	13.400	24	42.701

The structures were surrounded by a periodic box of 86×86×100 Å for looped model and 77x69x90 Å for clipped models and used a phosphatidylglycerol membrane (POPC) and the TIP3P water model (Neria *et al.*, 1996). Long range electrostatic interactions were

computed by the Particle-Mesh-Ewald (PME) method. The pressure was kept at 1 bar by Berendsen weak-coupling approach (Berendsen *et al.*, 1984). A time step of 2 *fs* was used by the implementation of SHAKE algorithm (Ryckaert *et al.*, 1977) for the bonds involving hydrogen. The Cartesian coordinates of atoms and energies were recorded at every 1 *ps*.

#### 4.4.2. Root Mean Square Distance (RMSD)

The RMSD of the looped model along the trajectory, which is calculated after aligning the C $\alpha$  atoms of each snapshot to the initial frame, is plotted in Figure 4.27. The RMSD plot reflects the extent of deviation from the initial conformation and the conformational changes throughout the trajectory. This plot is also used to identify the initial equilibration period prior to the dynamic equilibrium. In Figure 4.27, the RMSD values are calculated based on the alignment of a specified region, e.g., “RMSD Core Fit All” indicates that the whole protein structure is aligned to its initial frame first, and then the RMSD value is calculated using only the core region (whole protein excluding the ICL3 loop region). It is clear that exception the highly mobile loop region (light blue curve); the loop model has reached equilibrium state around 50 ns. The overall RMSD plot (RMSD All Fit All), which is displayed by dark blue, displays a change about 8.7 Å. The RMSD plot of “RMSD Core Fit Core” (purple line) shows that the most stable part of the protein is the core region.

The RMSD values have been calculated according to the core and transmembrane region for both clipped and looped models (Figure 4.28a, b). In Figure 4.28a, the alignment is based on only core region for looped model, i.e., the protein does not include the ICL3; in Figure 4.28b, the alignment is based on only transmembrane regions for all models, i.e., transmembrane regions include only the parts of the helices embedded inside the membrane. Core and transmembrane regions are shown in Figure 4.29. In Figure 4.28a, the clipped models achieve equilibrium at approximately 2-2.5 Å, but the equilibrium level of the looped model achieves approximately around 4 Å due to the effect of the ICL3 on the dynamics of the protein.

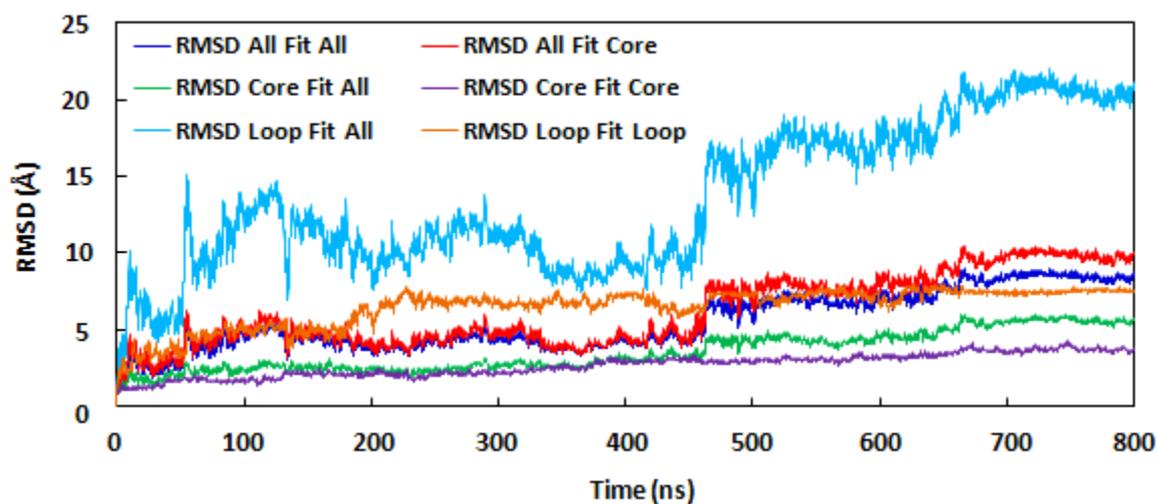


Figure 4.27. RMSD for looped model from the initial minimized structure.

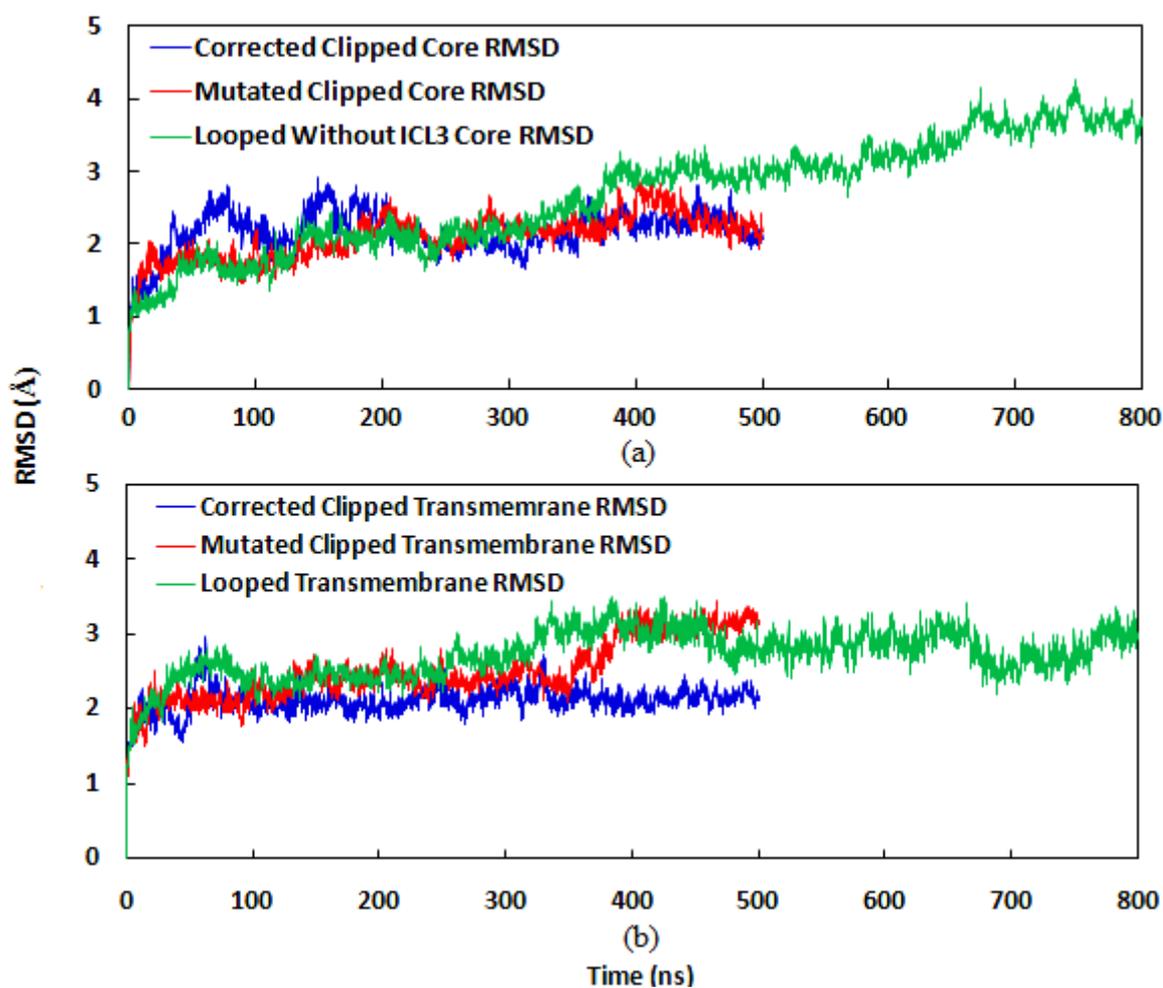


Figure 4.28. RMSD of the snapshots from the initial energy minimized structures for all runs of (a) the core regions, (b) the transmembrane regions.

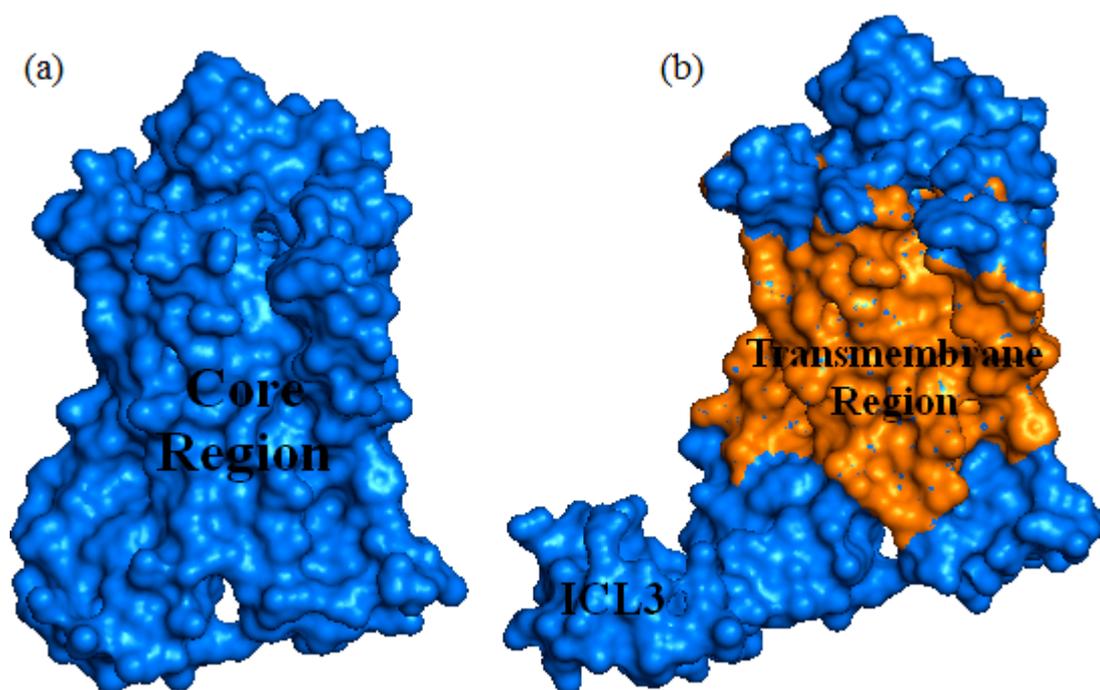


Figure 4.29. Regions considered in RMSD calculations for (a) core region includes all the loop regions except ICL3, (b) transmembrane region excludes all loop regions for both models.

In Figure 4.28b, the effect of ICL3 on the other loop regions can be observed. All models indicate similar levels of RMSD (maximum 3 Å) when only their transmembrane regions are taken into account. Going back to Figure 4.28a, this indicates that the conformations of other loop regions are less constrained in the looped model than clipped models, which is an effect of ICL3 on protein's conformational dynamics.

From Figure 4.27, there is an increase of about 2-2.5Å at around 450 ns in the overall RMSD plot (RMSD All Fit All). At the same time, the RMSD plot of ICL3 (RMSD Loop Fit Loop) shows an increase of about 5-7.5Å, which refers to the significant changes in the secondary structure of ICL3. This structural variation of looped model can be seen in Figure 4.30, which shows the conformational state (secondary structure) of each residue as a function of time. The color code displayed is as follows: 0 = T (turn), 1 = C (coil), 2 = B (isolated bridge), 3 = Extended (beta sheet), 4 = H (alpha helix), 5 = G (3-10 helices), 6 = I (Pi helix). Some conformational changes are observed in mainly two parts of the looped structure, namely ICL3 and ECL2. The first region including the residues from 231 to 262 is the ICL3 appears as a dark blue line at the beginning of the simulation that represent to

turn and coil helix. However, ICL3 is converted into alpha and 3-10 helix structure especially in the range from 400 ns to 450 ns. Moreover, the second region including the residues from 172 to 196 is ECL2, which in the beginning are of alpha helix structure at all three of the models. It is converted into a ‘Turn’ or ‘PI’ helix structure especially during the second half of the simulation in looped model. The conformational changes of these loops are also shown in Figure 4.31 using initial (green) and last (magenta) frames of MD trajectory. The structural alteration at ICL3 is shown in Figure 4.31a and the structural alteration at ECL2 is shown in Figure 4.31b.

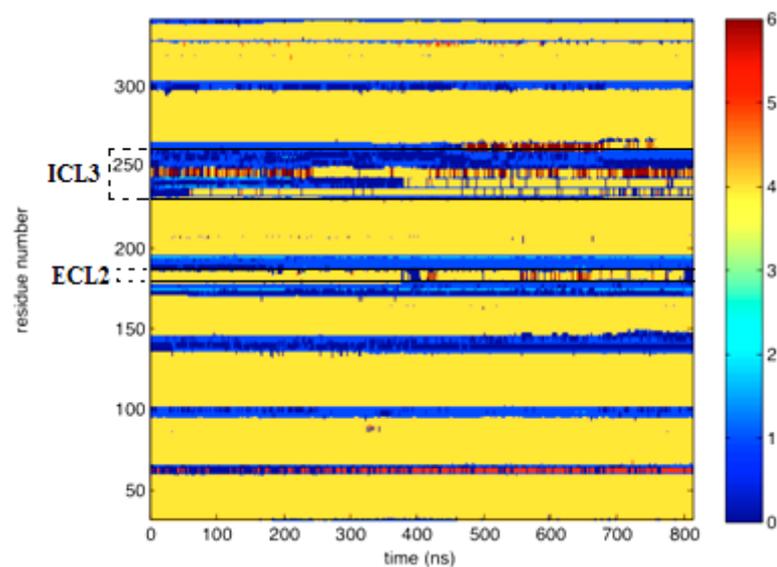


Figure 4.30. The alteration of the secondary structure along the simulation of the looped model.

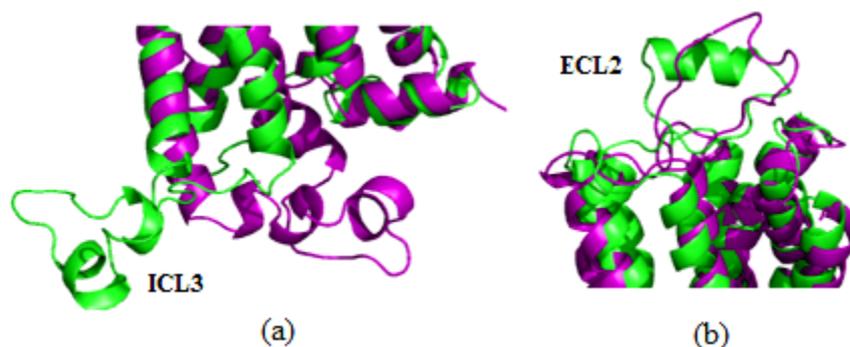


Figure 4.31. The structural alterations of certain loop regions based on initial (green) and final (purple) conformations of the looped simulation: (a) Changes in ICL3, (b) changes in ECL2.

The main difference between the clipped and looped simulations is the conformational changes of the extracellular ECL2 loop region. This region is well preserved in clipped models, whereas it is converted into a ‘Turn’ or ‘PI’ helix structure especially during the second half of the simulation in looped model. The structural variation of clipped models can also be seen in Figure 4.32a, b, respectively. This difference may be due to the effect of ICL3. No considerable variation has been noticed at other regions and the structural stability has been preserved throughout the simulation.

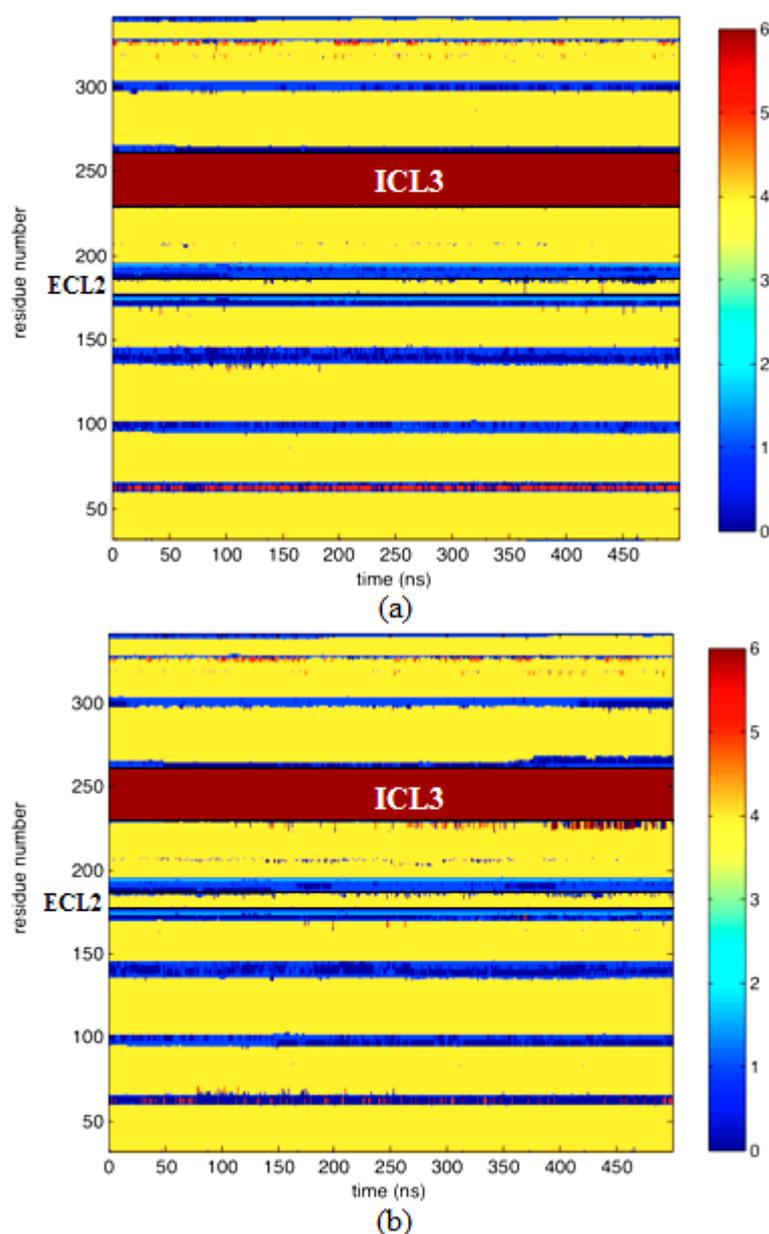


Figure 4.32. The alteration of the secondary structure along the simulation for (a) clipped model (natural), (b) clipped model (mutated).

The comparison of the structural variations between three models, are also discussed in terms of the pivot angles of each helix. The pivot region is where the helix breaks and this point often intersects with the proline residue. The change in pivot angle, that has an important role in signal transduction, has been obtained by calculating the angle between the axes of two half helices at both sides of the pivot (See Figure 4.33). Only the pivot angle at the eighth helix has been indicated as the angle between the axes of the seventh and eighth helices. The value of the pivot angle is averaged over the snapshots and presented with its standard deviation in Table 4.2. Accordingly, the helices where the break is detected to be the highest, i.e., the helices that have their angles closest to 90° are TM2, TM6 and TM7. The pivot values of TM7 and TM8 are close to each other in both clipped models, and different from that of the looped model. This may suggest a possible effect of the ICL3 region at the looped model's dynamics on TM7 and TM8.

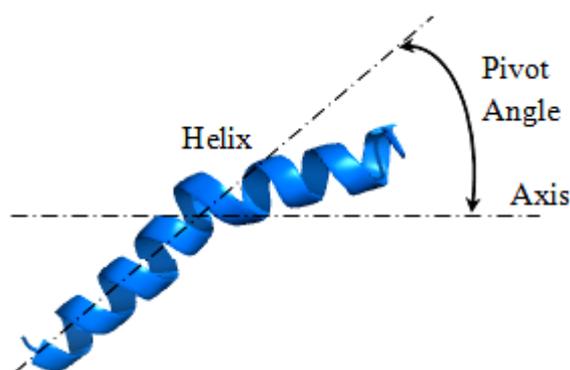


Figure 4.33. The pivot angle between the axis of two half helices.

Table 4.2. The average values of pivot angles of transmembrane helices for all models.

Helix Number	Looped Model	Clipped Model (Natural)	Clipped Model (Mutant)
TM1	168.495 ± 4.913	168.085 ± 5.946	170.168 ± 3.860
TM2	142.884 ± 5.697	142.767 ± 4.924	149.178 ± 4.231
TM3	168.980 ± 4.695	165.590 ± 4.165	170.677 ± 4.770
TM4	163.472 ± 6.409	163.808 ± 5.321	159.440 ± 5.527
TM5	165.545 ± 5.815	163.797 ± 4.034	163.776 ± 7.031
TM6	148.921 ± 7.051	143.954 ± 4.977	153.589 ± 4.675
TM7	151.115 ± 4.730	145.450 ± 4.403	141.759 ± 5.063
TM8	104.481 ± 6.476	113.355 ± 4.595	116.3135.568

#### 4.4.3. Root Mean Square Fluctuations (RMSF)

The root mean square fluctuations (RMSF) are calculated from the production phase of the trajectory. Calculations are carried out over the conformations aligned to the average structures of the simulations using only C $\alpha$  atoms. The best way to numerically express the protein mobility along the simulation is the root mean square of the average fluctuations or simply called the RMSF profile as shown in Figure 4.34. RMSF values, which are obtained from MD simulations, of looped and clipped models (Figure 4.34a) are compared to the experimental RMSF values, which are obtained from ANM, of both models (Figure 4.34b).

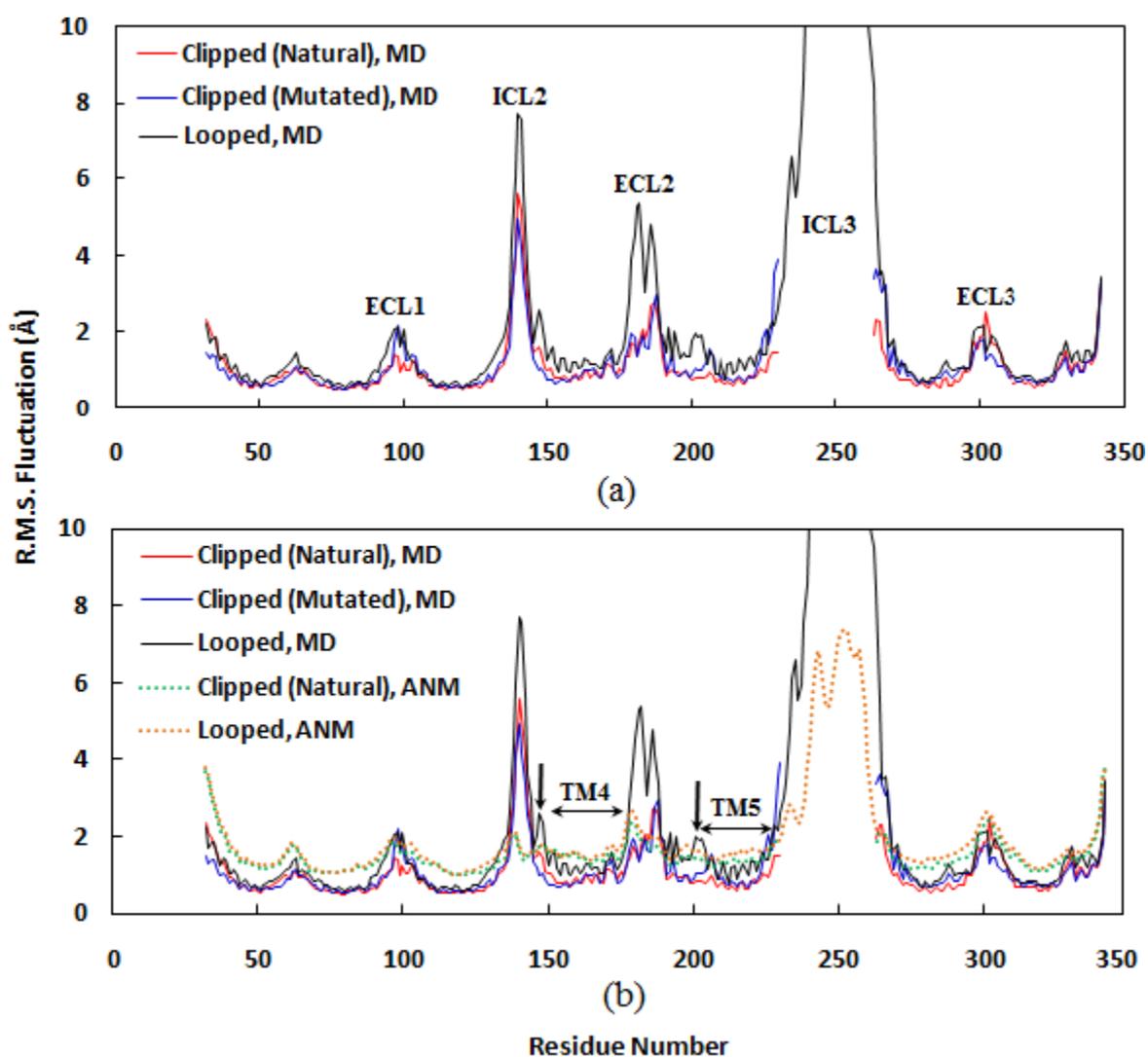


Figure 4.34. RMSF about the average position plotted for C $\alpha$  atoms of residues for (a) MD result, (b) ANM result for both models.

According to these profiles, the highest mobility of the protein is observed at the intracellular loop III (ICL3) of the looped model, as expected. After ICL3, the second mobile region for both models is detected at intracellular loop II (ICL2), facing inside the cell. The third highly mobile region is the ECL2 region, which faces outside the cell and is situated at the connection point of the signal molecules. Furthermore, RMSF values of ANM result are well compatible with the RMSF values of MD result for each models. Although, the fluctuation values of MD simulations are less than the fluctuation values of ANM result, in particularly the values of loops tend to fluctuate quite compatible with each other for both methods. Moreover, highly mobile residues, which belong to TM4 and TM5, respectively, are displayed with arrow in Figure 4.34b. These residues have the highest fluctuation values in these helices. The mobile regions are also shown on the structure of looped model in Figure 4.35. In general, the mobility of specific loops, namely ICL2 and ECL2, is higher in the looped model. Along the RMSD profile, the helices are generally located in between the peaks. Among all transmembrane helices, the most mobile ones are the fourth and the fifth helices (TM4 and TM5) in the looped model (Figure 4.36).

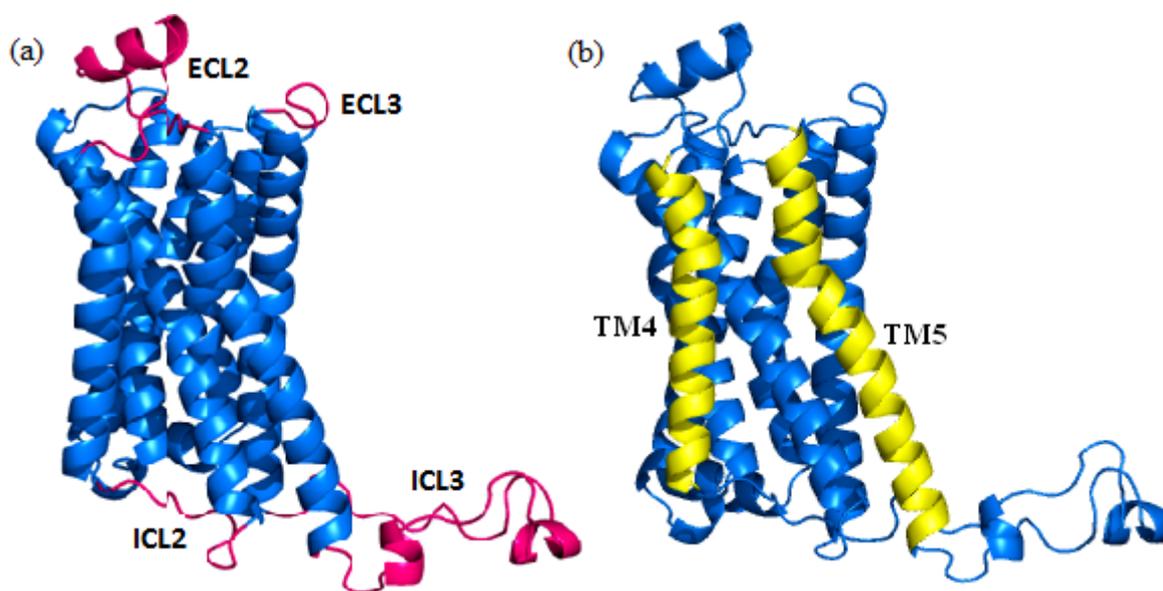


Figure 4.35. The positions of the most mobile regions on the structure for (a) loop, (b) TM4 and TM5.

The increase in the mobility of certain loops and helices in the looped model may be a result of ICL3, which possibly incorporates conformational flexibility to the whole protein. The mobility at TM4 and TM5 is even more obvious in the RMSD profiles

displayed in Figure 4.36a and 4.36b. The alignment required for RMSD calculations has been performed using the target protein, but the region including ICL3 has been excluded.

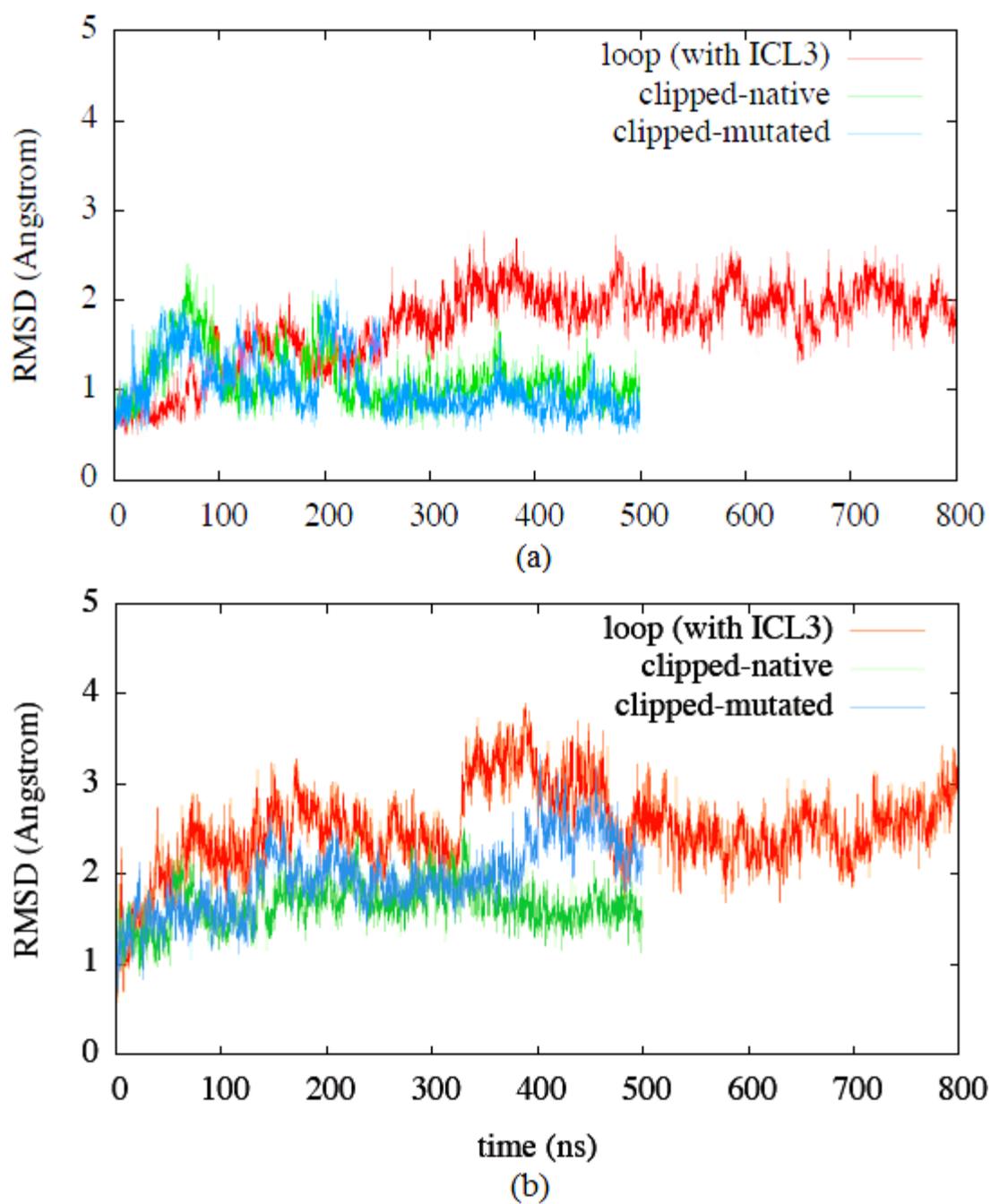


Figure 4.36. RMSD profiles of (a) TM4, (b) TM5.

The RMSD values obtained in this fashion display the position of TM4 and TM5 helices were altered in time compared with the position of other helices, taking the initial

conditions as the reference. This alteration may have taken place at the pivot break point in the middle of the helix and/or at both ends of the helix. Another consequence is that TM4 remained at a stable position for approximately 300 ns at all three models, whereas TM5 was unable to reach a stable position throughout the simulation. TM5 being unable to reach a stable position supports its effective role in structural changes occurring during the agonist binding of this helix (Katritch *et al.*, 2009).

#### 4.4.4. Ionic Lock

One of the most significant structural changes occurring during the deactivation process of protein is the degradation of the salt bridges. Located close to interior part of the cell, two different salt bridges have been mentioned in the preliminary studies, which are situated between Asp130 and Arg131 on TM3, and between Arg131 upon TM3 and Glu268 upon TM6. In a recent study (Bokoch *et al.*, 2010); the existence of a third bridge between Asp192 and Lys305 that connect ECL2 to ECL3 has been detected at the region facing outside the cell. The open/closed states (distance between positive and negative charges) of these three different salt bridges are shown in Figure 4.37- 4.39 for clipped and looped models. Moreover, the percentages of time (snapshots) during which the salt bridges are closed throughout the simulation are listed in Table 4.3.

The criterion used for the closed state, i.e. an intact salt bridge, is that the distance between N and O atoms is less than 3.5 Å. In Figure 4.37- 4.39, the minimum distance between all possible nitrogen and oxygen atom pairs in the side groups of arginine and glutamic acid is represented by the black line. Among the six different N-O distances between three nitrogen atoms (at guanidium group of Arg) and two oxygen atoms (at the carboxyl group of Glu), the smallest distance is plotted for each snapshot. Moreover, the distance between the alpha carbon atoms of the same residues forming the salt bridge is represented by the red line in the same figures.

The salt bridge in Figure 4.37 is between Asp 130 and Arg 131, which are located on the same helix (TM3). There is no considerable change in the distances between alpha carbon atoms, which is an indication that the helix structure remains intact in all simulations. When the distances between side chains are considered, this salt bridge remains open for the longest time in the clipped (mutated) model (Table 4.3) in 71% of all

the snapshots. Whether this is a consequence of the single mutation (D187E) or just the result of an independent simulation still remains an open question.

Table 4.3. The percentage of time during which the salt bridge is in the closed state (N-O distance is less than 3.5 Å).

Model	Simulation lengths (ns)	Asp130 - Arg131 (%)	Arg131 – Glu268 (%)	Asp192 – Lys305 (%)
Looped	800	95.2	60.3	28.6
Looped	500	96.3	55.8	54.7
Clipped (Natural)	500	98.2	83.6	64.1
Clipped (Mutated)	500	70.6	94.4	79.3

Opening of the Arg131-Glu268 is one of the critical conformational changes required for the receptor to get activated (Dror *et al.*, 2008). The state of this salt bridge is presented in Figure 4.38. This bridge is observed as being open in the looped model for the longest time (40% open, 60% closed during the 800 ns simulation, Table 4.3). In the contrast, the bridge is open in 16% and 6% of the snapshots in 500 ns runs of both clipped models. These percentages are based on the N-O distances (black) rather than the backbone distances (red), which seem to be correlated to the N-O distances to a certain extent. This difference between looped and clipped is clearly a consequence of the ICL3, which incorporates conformational flexibility to the protein. Thus, almost a periodic opening/closing action of the gate is observed in Figure 4.38c with respect to time.

The change in the length of the third salt bridge, which connects the regions of ECL2 and ECL3 to each other, is shown in Figure 4.39 for both models. This bond connects Asp192 and Lys305 residues to each other and has a considerably different profile to the one that takes place between Arg131 and Glu268. Asp192-Lys305 salt bridge suggests that both of the clipped models remain open for a longer period of time. As for the looped model, it mostly stays shut in the first 250 ns, but it remains constantly open after the first 250 ns till the end of the simulation. At the part of the simulation between 250 and 800 ns, the distance of N-O lies at a wide range of 6-18 Å. Another important difference is that the length between N-O is 18 Å at the looped model, while the length between C $\alpha$  atoms reaches 16-17 Å.

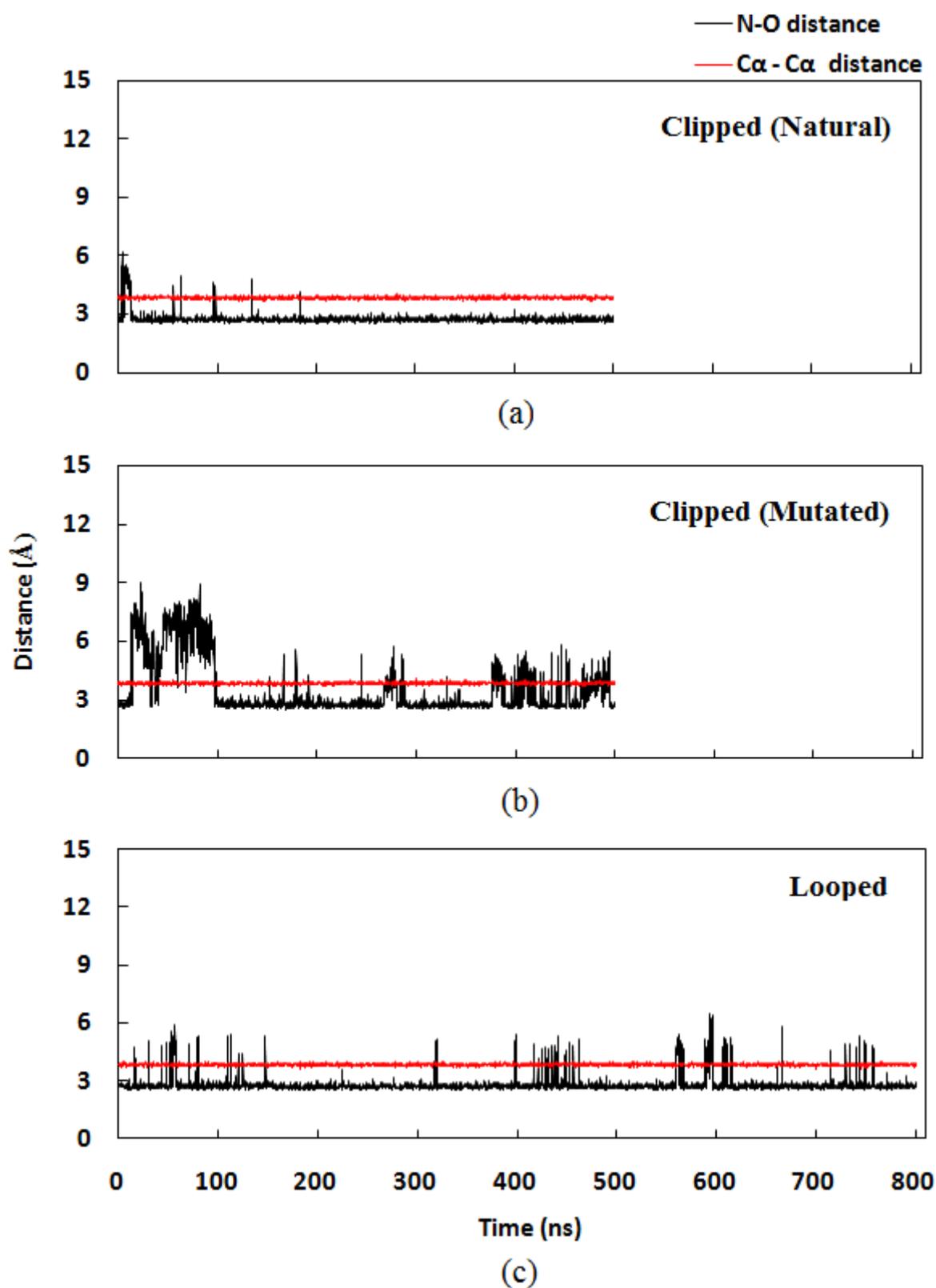


Figure 4.37. The N-O (black) and C $\alpha$  - C $\alpha$  (red) distances (N-O reflecting the open/closed states of the salt bridge between Asp130 and Arg131 along the simulations for (a) clipped (natural), (b) clipped (mutated), and (c) looped models.

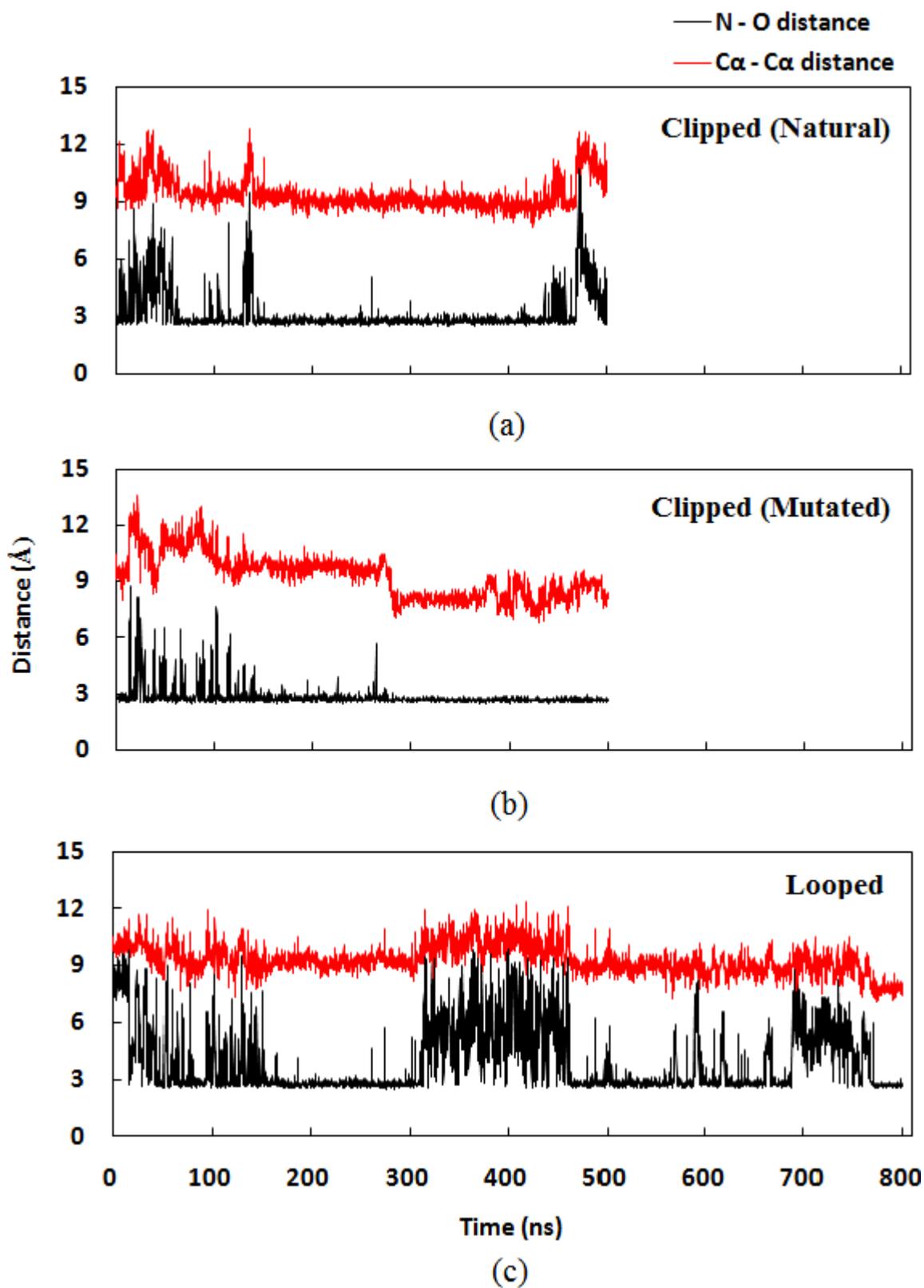


Figure 4.38. The N-O (black) and Ca - Ca (red) distances (N-O reflecting the open/closed states of the salt bridge between Arg131 and Glu268 along the simulations for (a) clipped (natural), (b) clipped (mutated), and (c) looped models.

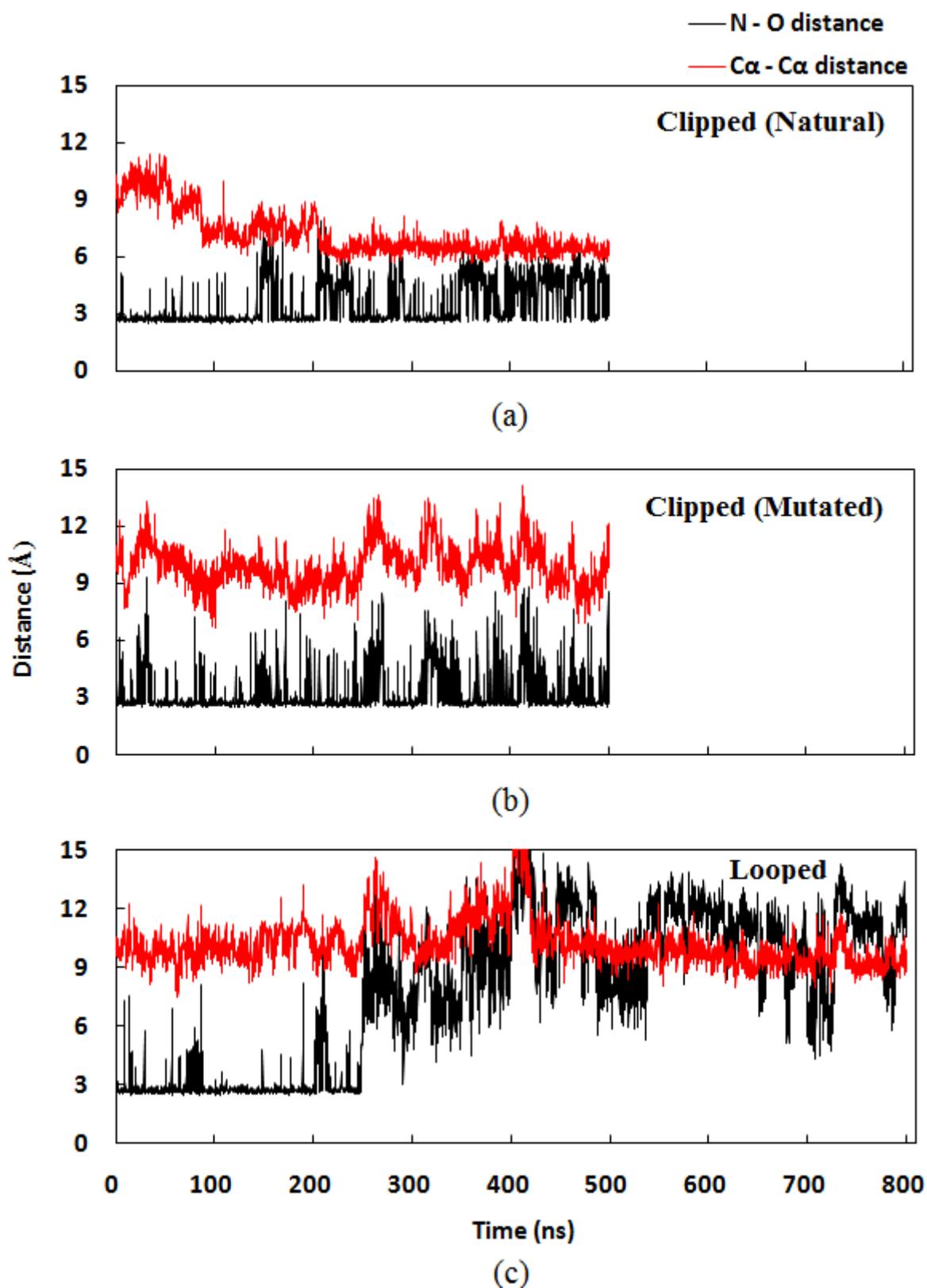


Figure 4.39. The N-O (black) and  $\text{Ca} - \text{Ca}$  (red) distances (N-O reflecting the open/closed states of the salt bridge between Asp192 and Lys305 along the simulations for (a) clipped (natural), (b) clipped (mutated), and (c) looped models.

The largest N–O distance for the Arg131–Glu268 salt bridge does not exceed 10 Å (the maximum length of C $\alpha$ –C $\alpha$  is around 12 Å). In contrast, it is an expected outcome considering the fact that Asp192–Lys305 salt bridge is situated on ECL2 and ECL3 loop regions, both of which are almost devoid of restrictions for mobility. Apart from that, no obvious correlation has been noticed on three salt bridges regarding the durations they remained closed/opened and it has been revealed that they act rather independently without being affected from each other. To reveal the conformational changes that are coupled to the opening/closing of the salt bridge in the looped model, the changes in main chain  $\phi$  (phi) and  $\psi$  (psi) angles of Arg131 and Glu268 residues and the angle  $\chi_1$  (chi) of the side chains are analyzed. As displayed in Figure 4.40a and 4.41a, there is no considerable change in the  $\phi$  or  $\psi$  angles of both residues.

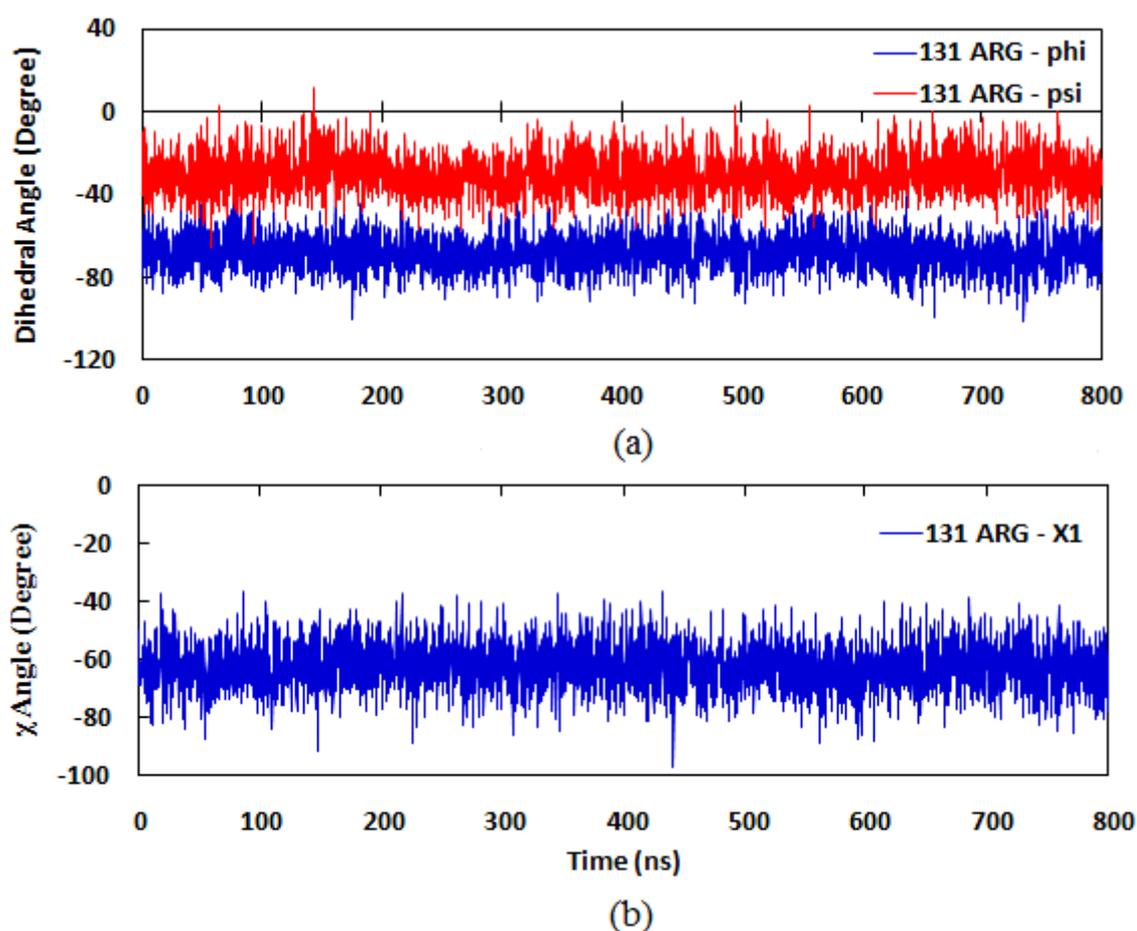


Figure 4.40. The change, along the simulation of looped, in main chain  $\phi$  (phi) and  $\psi$  (psi) angles and in the side chain  $\chi_1$  (chi) angle for Arg131, which form the salt bridge for (a)  $\phi$  and  $\psi$ , (b)  $\chi_1$ .

However, the change in angle  $\chi_1$  of Glu268 in Figure 4.38b seems to be in correlation with the change in the N-O distance shown for the looped model in Figure 4.38c. In other words, the switching of the salt bridge from closed to open states and the elevation of the angle  $\chi_1$  of Glu268 from  $-160^\circ$  to  $-80^\circ$  occurs simultaneously. Consequently, opening of the salt bridge stems mainly from the change in angle  $\chi_1$ . The same correlation is observed in angle  $\chi_1$  of Glu268 with the change in the  $C\alpha - C\alpha$  distance.

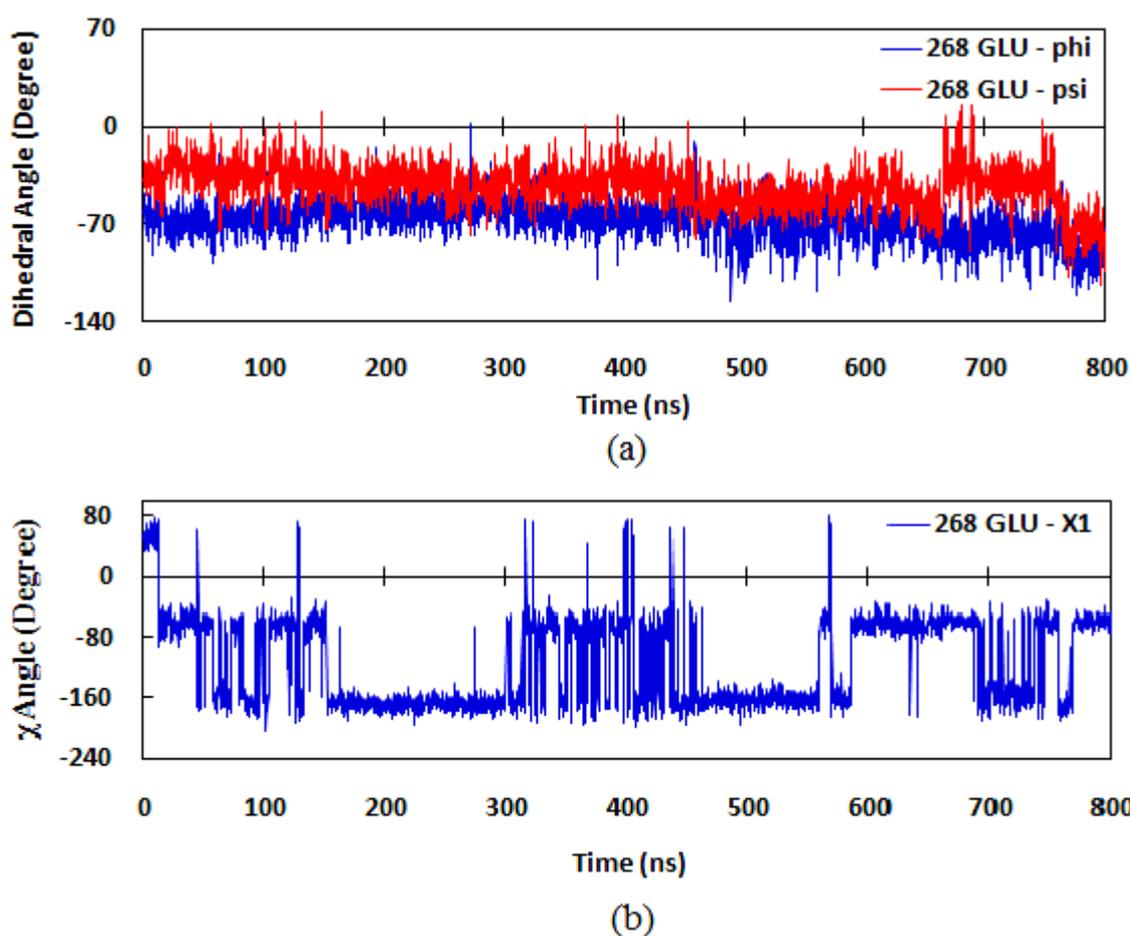


Figure 4.41. The change, along the simulation of looped, in main chain  $\phi$  (phi) and  $\psi$  (psi) angles and in the side chain  $\chi_1$  (chi) angle for Glu268, which form the salt bridge for (a)  $\phi$  and  $\psi$ , (b)  $\chi_1$ .

On the other hand, the change in angle  $\chi_1$  of Glu268 in Figure 4.42b seems to be in correlation with the change in the N-O distance shown for the clipped (natural) model in Figure 4.38a. The switching of the salt bridge from closed to open states and the elevation of the angle  $\chi_1$  of Glu268 from  $-160^\circ$  to  $-80^\circ$  occurs simultaneously. Consequently, opening of the salt bridge stems mainly from the change in angle  $\chi_1$ . The same situation is observed

in angle  $\chi_1$  of Glu268 with the change in the  $C\alpha - C\alpha$  distance. However, the same correlation between the change in any angle  $\chi_1$  of ionic lock residues and the change in any salt bridge distances is not obtained for clipped (mutated).

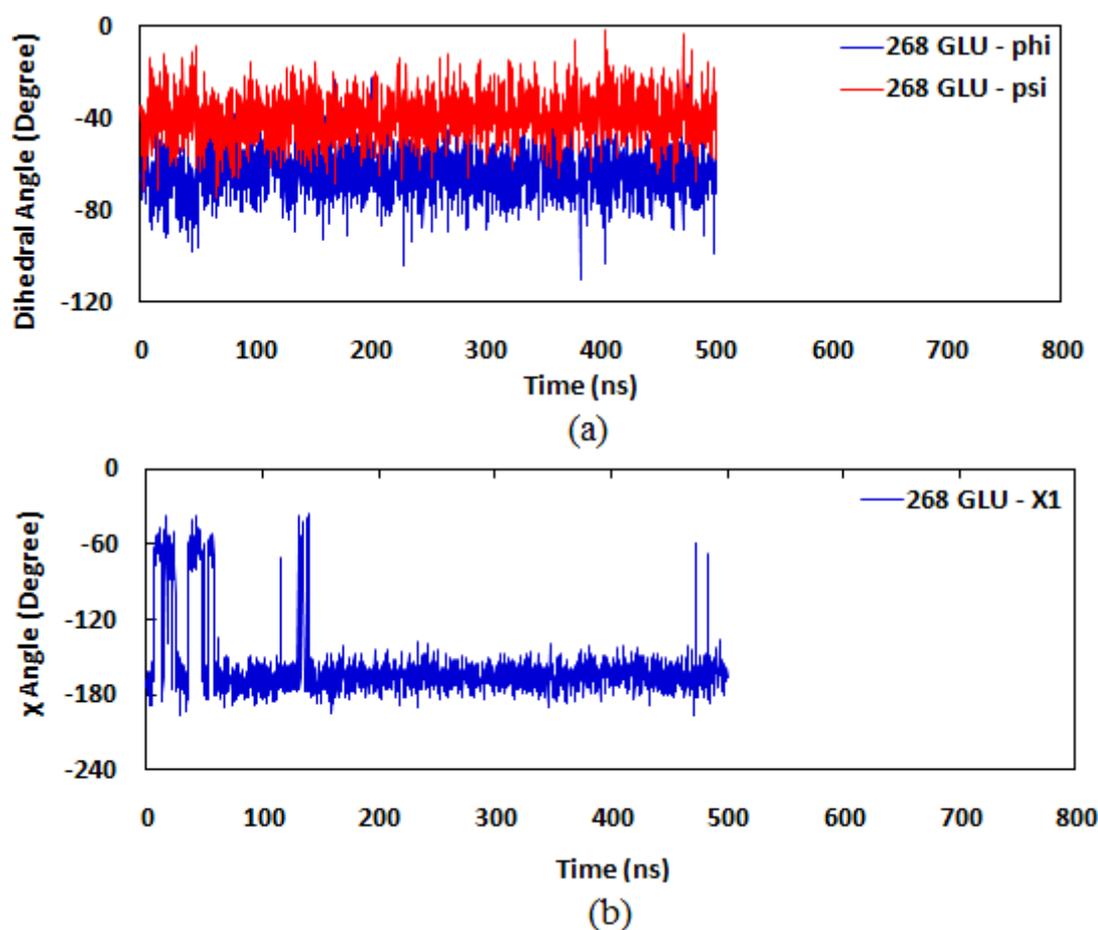


Figure 4.42. The change, along the simulation of clipped (natural), in main chain  $\phi$  (phi) and  $\psi$  (psi) angles and in the side chain  $\chi_1$  (chi) angle for Glu268, which form the salt bridge for (a)  $\phi$  and  $\psi$ , (b)  $\chi_1$ .

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

Individual and cumulative explanation values of low-indexed principal modes are shown in Figure 4.43. The variance percentages of the first five PC are also listed in Table 4.4 for each model. According to these results, looped model has the highest explanation value in the first mode. The percentage explanation value of the first PC included in the looped model is 59.3. In addition, having the highest explanation values, the first PC of this model is considered to explain the protein's overall dynamics to a considerable degree.

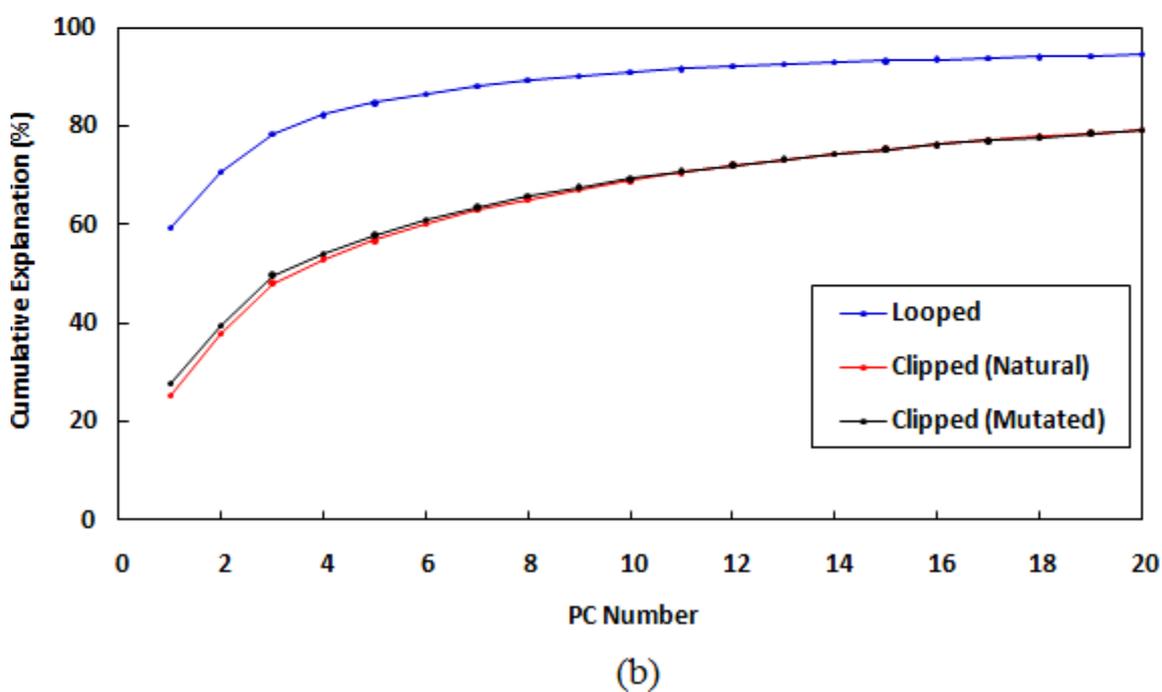
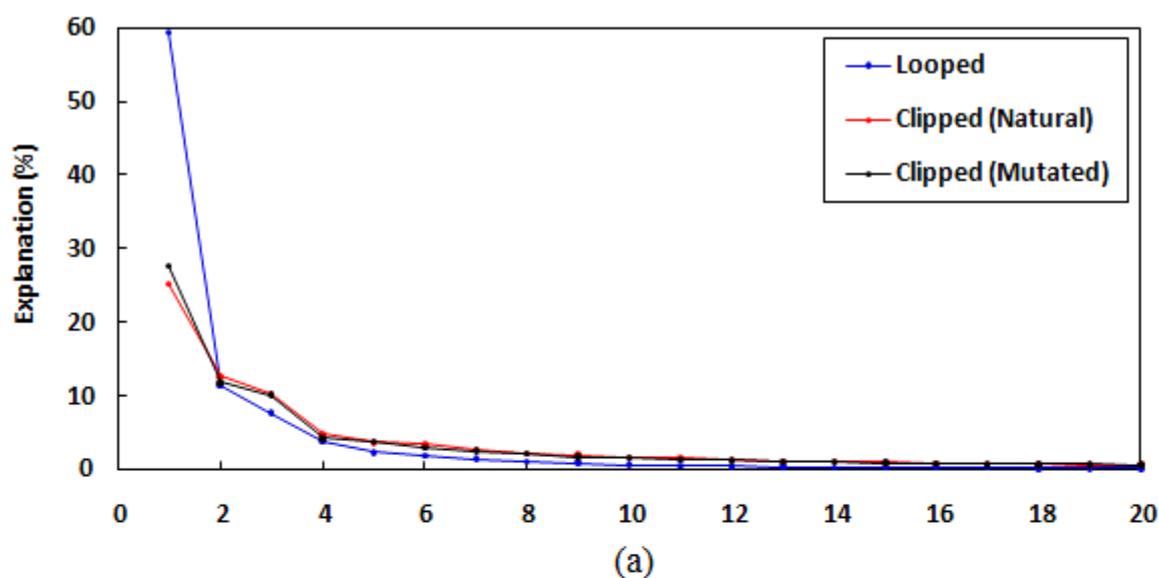


Figure 4.43. The graphical representation of variance percentages of the models for (a) individual variance, (b) cumulative variance.

Specifically the percentages of the cumulative values of the first five modes indicate that first five modes could satisfactorily explain the protein's entire dynamics. As for the clipped models, their explanation values are comparatively lower than the clipped models, which also indicate the effect of ICL3 on the conformational dynamics of the protein. There is also a slight difference between the explanation values of the mutated and natural clipped models (the first principal components are 27.7 and 25.3 for the mutant and the

natural models). Evaluating the explanation values cumulatively, the values for the first five modes are exceedingly close to each other.

Table 4.4. Percentage of the total motion explained by the first five principal components of each model.

<b>Simulations</b>	<b>Principal Component Number</b>	<b>% Variance Capture-Individual</b>	<b>% Variance Capture-Total</b>
<b>Looped</b>	PC1	59.3	59.3
	PC2	11.4	70.7
	PC3	7.7	78.4
	PC4	3.8	82.2
	PC5	2.3	84.6
<b>Clipped (Natural)</b>	PC1	25.1	25.1
	PC2	12.7	37.8
	PC3	10.3	48.1
	PC4	4.8	53
	PC5	3.7	56.7
<b>Clipped (Mutated)</b>	PC1	27.7	27.7
	PC2	11.8	39.5
	PC3	10.2	49.7
	PC4	4.2	54
	PC5	3.8	57.8

#### 4.5.1. Slow (Collective) Modes

Slow modes give information about the collective conformational motions of proteins around their native state. These collective motions represent biological functions of many enzymes and proteins (Bahar *et al.*, 1998). These modes also contribute to address the overall dynamics of the structure. In this section of the thesis, the first two slow modes of the models, which are obtained from PCA, are examined in order to gather knowledge about cooperative conformational motions of the models that these modes have the highest explanation values of overall dynamics of the structure. The explanation percentages of the first two principal components are also listed in Table 4.5 for each model. According to these results, looped model has the highest explanation value in the first mode. Also the cumulative explanation values of the first two principal modes reach approximately 71%

that is almost adequate to give insight about the overall dynamics of the looped model. On the other hand, there is not a big difference between the explanation values of the first principal modes of the clipped models. These modes are capable of describing one fourth of the overall dynamics of the structure. Moreover, the cumulative explanation values of the first two principal modes are able to represent about 40% of the overall dynamics of the clipped models.

Table 4.5. Percentage of the total motion explained by the first two PC of each model.

<b>Simulations</b>	<b>Principal Component Number</b>	<b>% Explanation Individual</b>	<b>% Explanation Cumulative</b>
<b>Looped</b>	PC1	59.3	59.3
	PC2	11.4	70.7
<b>Clipped (Natural)</b>	PC1	25.1	25.1
	PC2	12.7	37.8
<b>Clipped (Mutated)</b>	PC1	27.7	27.7
	PC2	11.8	39.5

In Figure 4.45, the collective motions of the 1<sup>st</sup> and 2<sup>nd</sup> principal modes are displayed using 11 snapshots of each principal component. The structures marked with red specify the initial snapshot, while the structure marked with green specifies the last snapshot. The actions of the modes vary going from red snapshot to green snapshot, and the structures in between are marked with light blue. Particularly, ICLs and ECLs are very mobile in looped model. Also TM helices move to be compatible with loop regions in terms of collective motions. On the other hand, loop regions of both clipped models have the highest mobility but the mobility of TM helices is lower than loops.

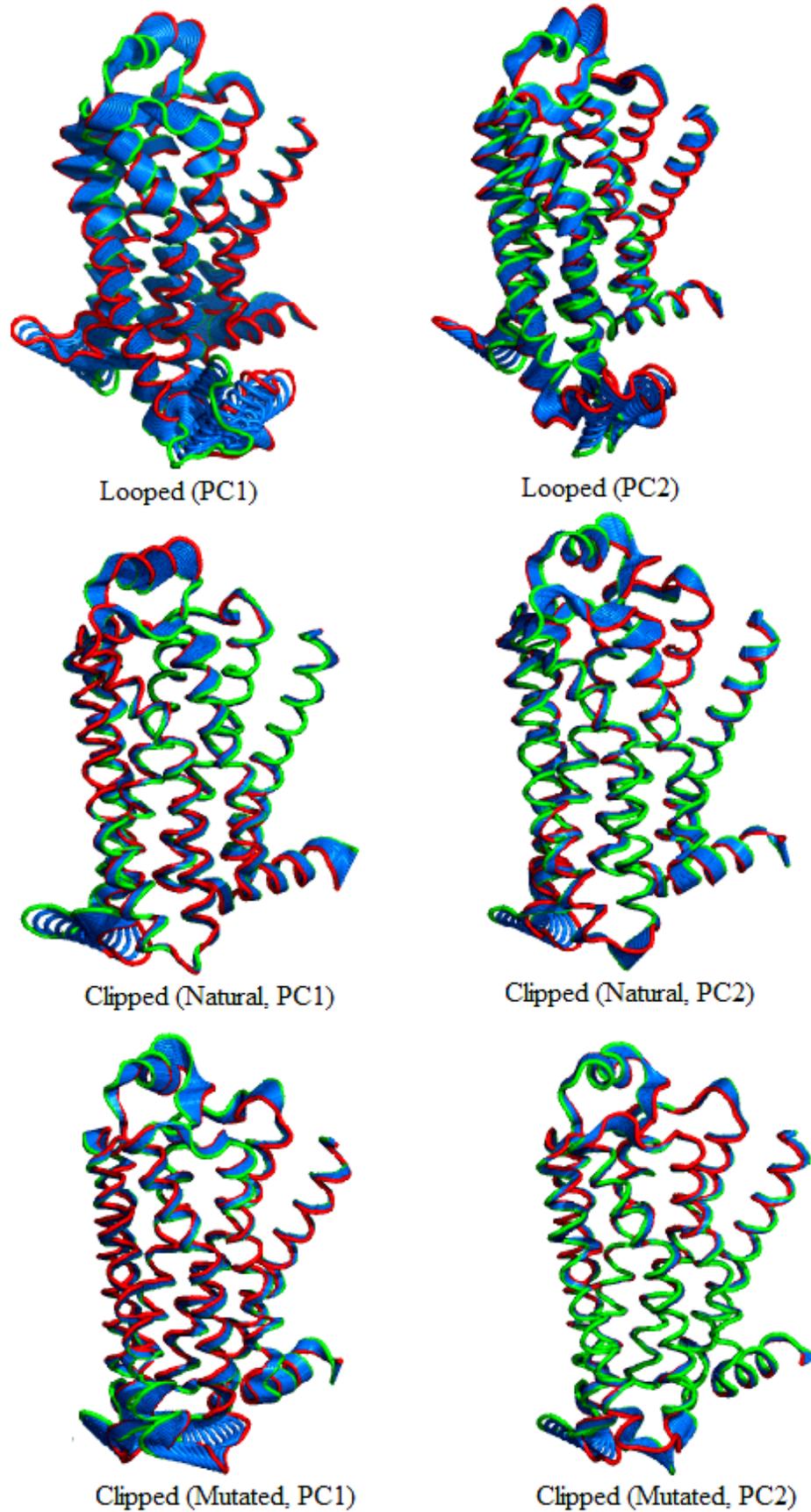


Figure 4.44. Collective motions on the first and second PC of looped and clipped models.

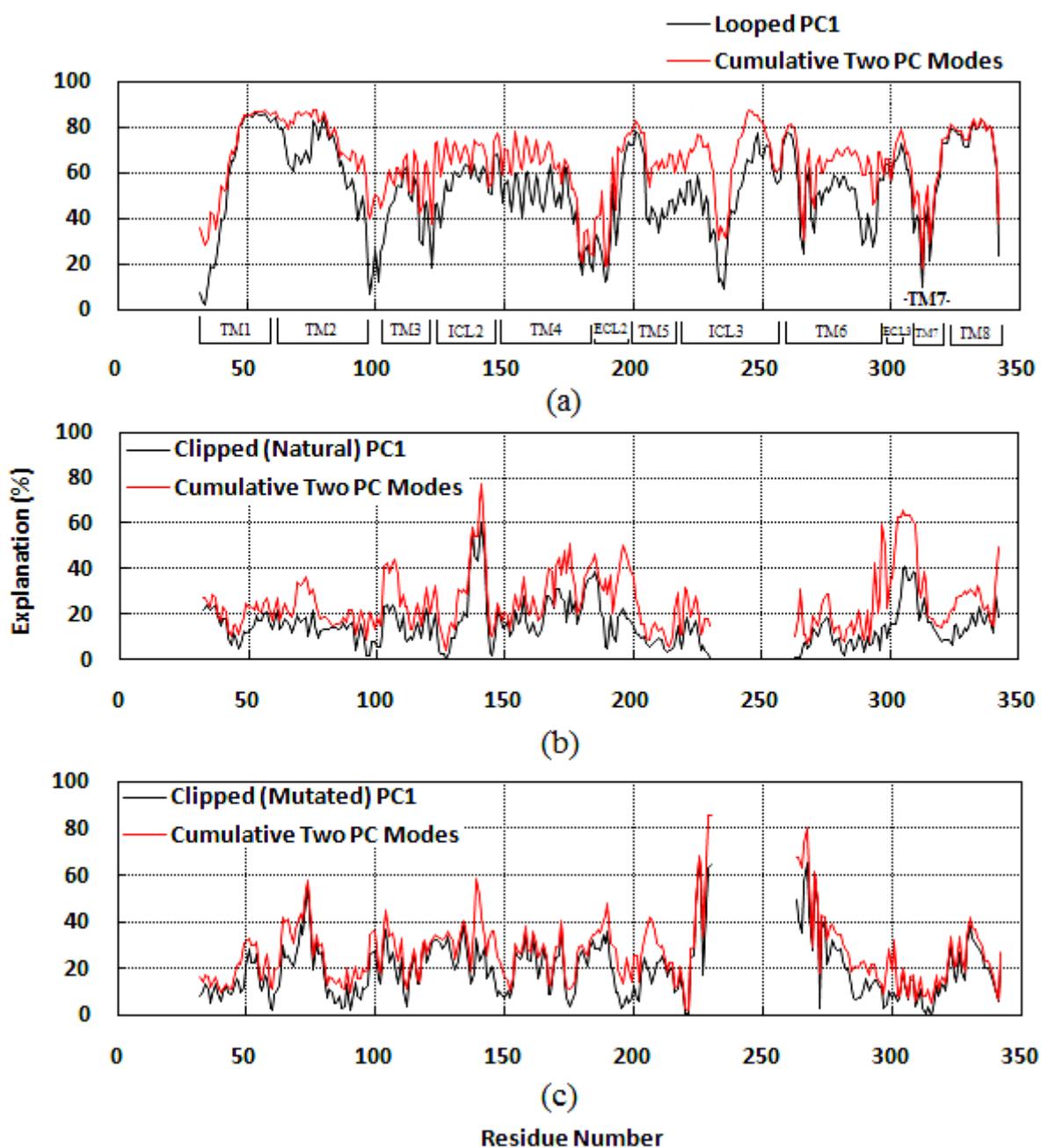


Figure 4.45. The percentage of residue-based motion explained by PC1 and cumulative two PC for (a) looped, (b) clipped (natural), (c) clipped (mutated) models.

The residue-based explanation values of each model are shown in Figure 4.45a, b, c. Also the most mobile regions are shown in Figure 4.45a. Examination of the explanatory power of PC1 on residue basis shows that about 80% of the motions of the regions that are the range of 137-146 (ICL2), 231-266 (ICL3) and 299-304 (ECL3) are explained, whereas about 40% of the motion of region that is the range of 172-196 is explained by PC1 for looped model. Furthermore, the motions of the transmembrane helices such as TM3, 4, 5, 6

and 7 are explained by almost 60% in PC1 for looped model. The explanation values of these ECLs and ICLs and TM helices are shown in detail in Figure 4.45. The overall motion of the looped model is described by the first PC. PC2 explains about 12% of the motion.

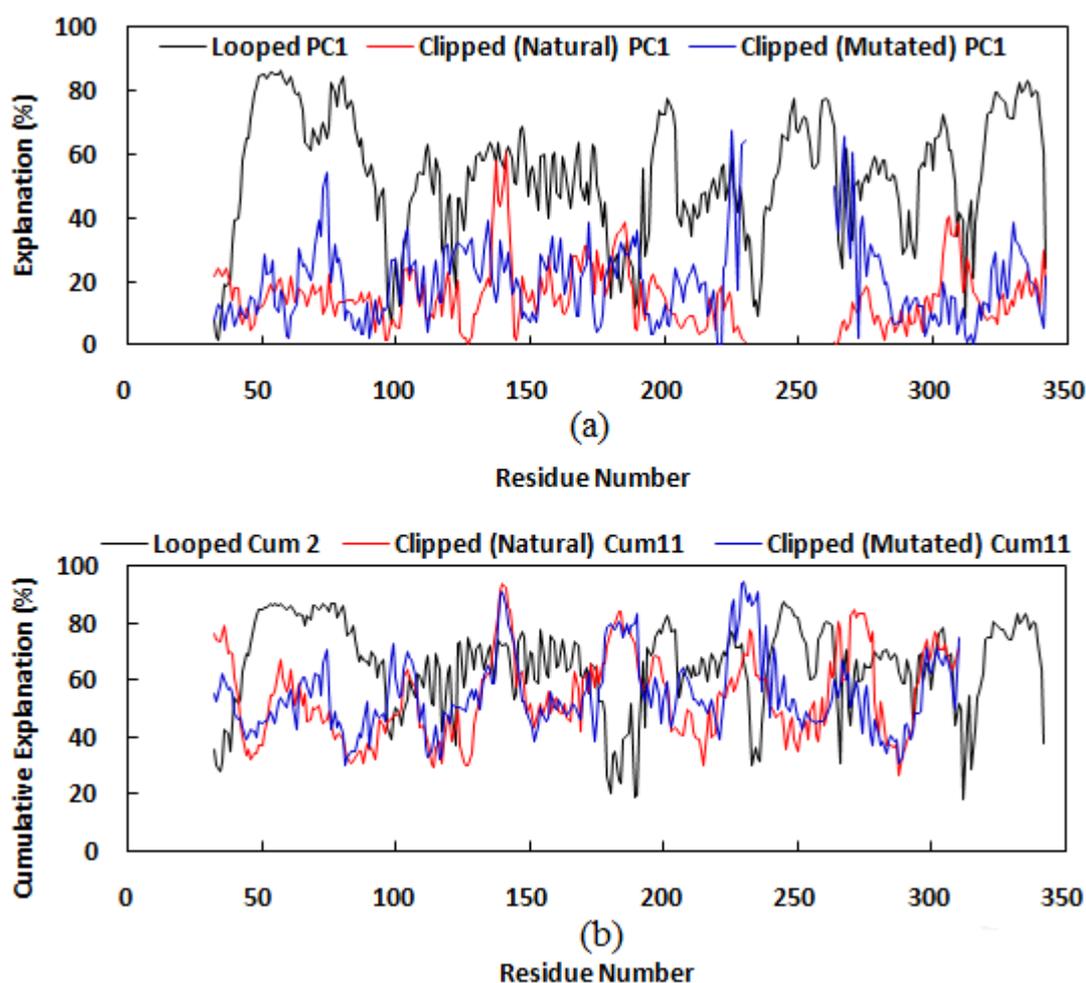


Figure 4.46. The residue-based explanation values for (a) the first PC, (b) cumulative 2 PC of looped and cumulative 11 PC of each clipped model.

In Figure 4.45b, c, examination of the explanatory power of PC1 on residue basis shows that about 40% of the motions of the regions are generally explained for both clipped model. To better understand the explanation values, the residue-based explanatory power of PC1 of each model and cumulative two PC of looped and 11 PC of each clipped model are shown in Figure 4.46. Accordingly, explanatory power of PC1 on residue basis shows that about 60% of the motions of the residues are explained for looped, whereas about 40% of the motions are explained for both clipped models by PC1. The overall

motions of the clipped models are described by the first PC, which has the percentage of 28. Also, PC2 explains about 12% of the motion.

In light of this evaluation, it could be concluded that each and every principal mode taking place at the looped model has different dynamics than those of the principal modes taking place at the clipped models and that this distinction poses a very important role in the overall dynamics of the structure. It is possible to understand this difference easily looking at the cumulative explanation values (Table 4.5). The explanatory power of the cumulative two principal modes shows that about 70% of the motions of the looped model are explained by these cumulative two principal modes, whereas about 40% of the motions of the clipped (natural) and clipped (mutated) models are explained. The explanation percentage of the loop regions is approximately 80 that they are highly mobile parts of the structure. The movements of these loop regions are observed in the first principal mode of the looped model. These regions play a very significant role to represent the overall dynamics of the structure. On the other hand, loop regions and transmembrane helices are more rigid than looped model due to lack of ICL3 in the clipped models. As for the clipped models, their explanation values are comparatively lower than the looped model, which also indicates the effect of ICL3 on the conformational dynamics of the protein. There is also a slight difference between the explanation values of the mutated and natural clipped models (the first principal components have the percentage of 27.7 and 25.3 for the mutant and the natural models). Evaluating the explanation values cumulatively, the values for the first five modes are exceedingly close to each other (Table 4.4).

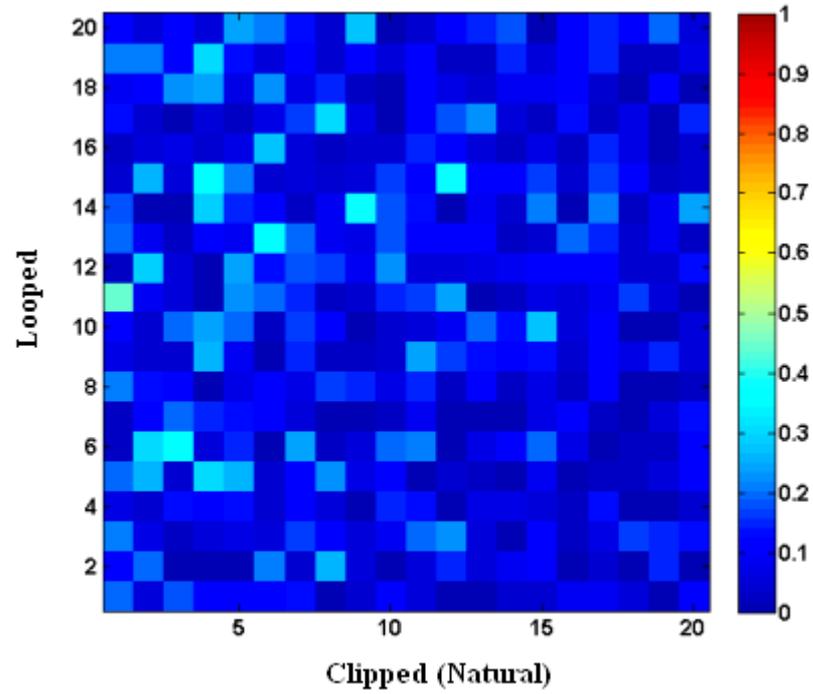
#### **4.5.2. Overlap Calculations**

In this part, we evaluate the overlap matrix that is obtained from the eigenvectors of the models after principal component analysis. Overlap matrix is used to determine whether the movements of the modes are compatible with each other. The level of agreement between the collective motions of clipped models and looped model are calculated using overlap definition in Equation 4.2. In figure of overlap matrix, the [0:1] range is indicated with color scale from blue to red color and shows the similarity between modes for models.

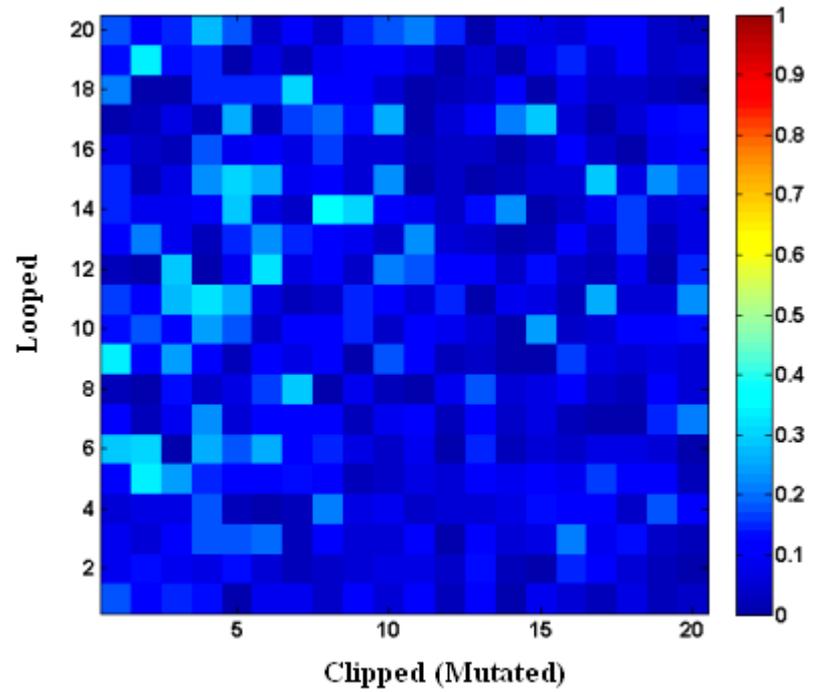
#### 4.5.2.1. Overlap between Looped and Clipped Models

The compatibility of the principal components of the looped model with those of the clipped models is shown in Figure 4.47. Compatible and incompatible principal modes between the looped model and the clipped (natural) model are shown in Figure 4.47a, while compatible and incompatible principal modes between the looped model and clipped (mutated) model are shown in Figure 4.47b. According to these figures, the principal modes of looped model reflect a low overlap with both of the principal modes of clipped models. This distinction stems from ICL3 taking place at the looped model. The first five principal modes of the looped model have a low overlap value with the principal modes of the clipped models and some incoherence seems to be in place. The modes after the sixth principal mode seem to be relatively more coherent. This coherence lays out overlap values ranging from 0.3 to 0.5. Taking this gap into consideration, it would be right to state that 20 principal modes belonging to looped models display quite a low coherence with the 20 principal modes of clipped models.

The principal modes taking place at the clipped models, which displays a partial coherence with the collective motion of the principal mode of the looped model, are shown in Figure 4.48. Collective motion on the sixth principal mode of the looped model is shown in Figure 4.48a. Also collective motion on the third and the second principal modes of clipped (Natural) and clipped (Mutated) models are shown in Figure 4.48b, c respectively. In light of this evaluation, it could be concluded that each and every principal mode, taking place at the looped model has different dynamics than those of the principal modes taking place at the clipped models and that this distinction poses a very important role in the overall dynamics of the structure.



(a)



(b)

Figure 4.47. Overlap matrix of the first 20 modes between (a) looped and clipped (natural), (b) looped and clipped (Mutated).

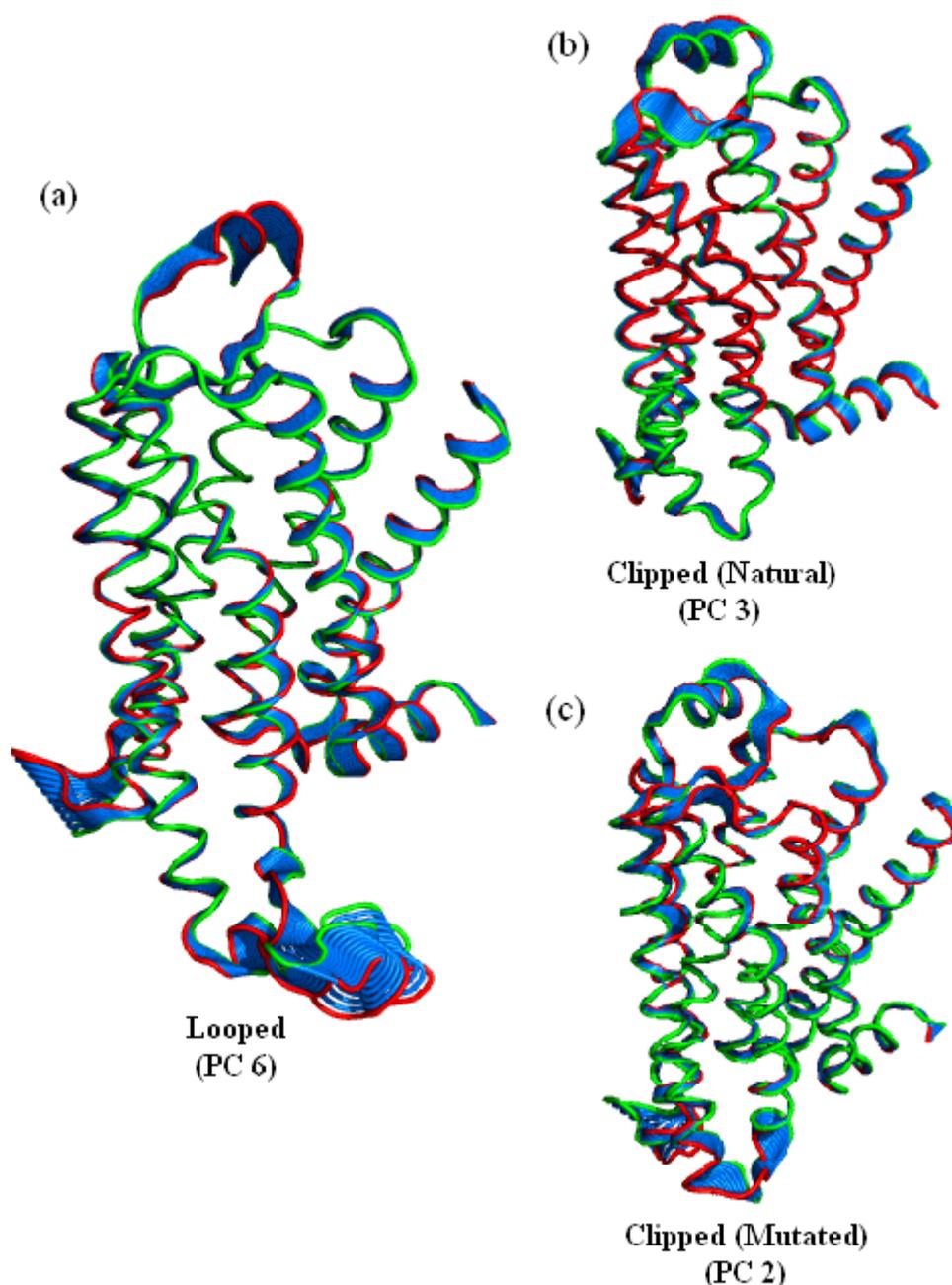


Figure 4.48. Collective motions on PC for (a) the sixth PC of looped, (b) the third PC of clipped (natural), (c) the second PC of clipped (Mutated).

#### 4.5.2.2. Overlap between Clipped (Natural, Mutated) Models

Overlap matrix, which displays the reciprocal compatibility values of the modes at clipped models, is shown in Figure 4.49. Adding the reminder that this was mentioned in our study before, the only distinction among clipped models is that Asn187 has been replaced with Glu187 to be mutated. The aim here was to see that one single residue would

not affect the dynamics of the system considerably. To reach a definite conclusion, clipped (mutated) model has been formed as a result of the mutation of this residue and just like clipped (natural) model, has been exposed to an MD simulation of 500 ns. Overlap matrix has been calculated with the results obtained from the principal component analysis and it shows that the modes of two clipped models are in an exact compatibility. The compatibility is specifically intense on the diagonal axis of the matrix, which is an expected outcome. Though the compatibility values of the modes are not high, two models move in a correlated manner. It could also be said that different collective motions of the principal modes between two models do not stem from mutation, but from the MD simulations conducted at different initial conditions of the two identical systems. Figure 4.50 shows four modes taken from the clipped models, which are in a high compatibility. The compatibility values between the fifth mode for the clipped (natural) model and the fourth mode for the clipped (mutated) model, which are shown in Figure 4.50a, range from 0.6 to 0.8. Similarly, the compatibility values between the third mode of the clipped (natural) model and seventh mode of the clipped (mutated) model, which are shown in Figure 4.50b, range from 0.6 to 0.8.

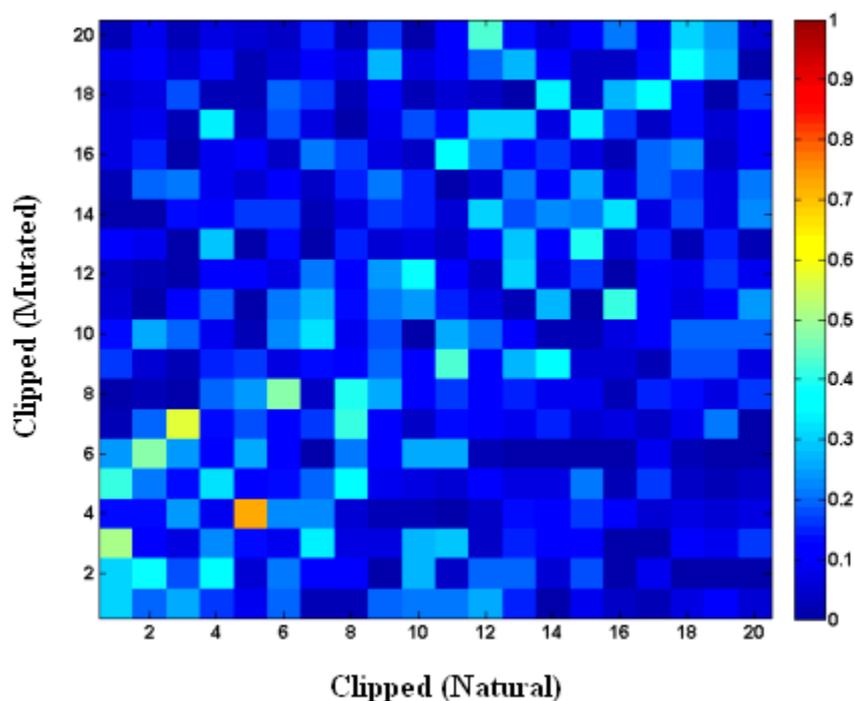


Figure 4.49. Overlap matrix between clipped (natural) and clipped (mutated) models.

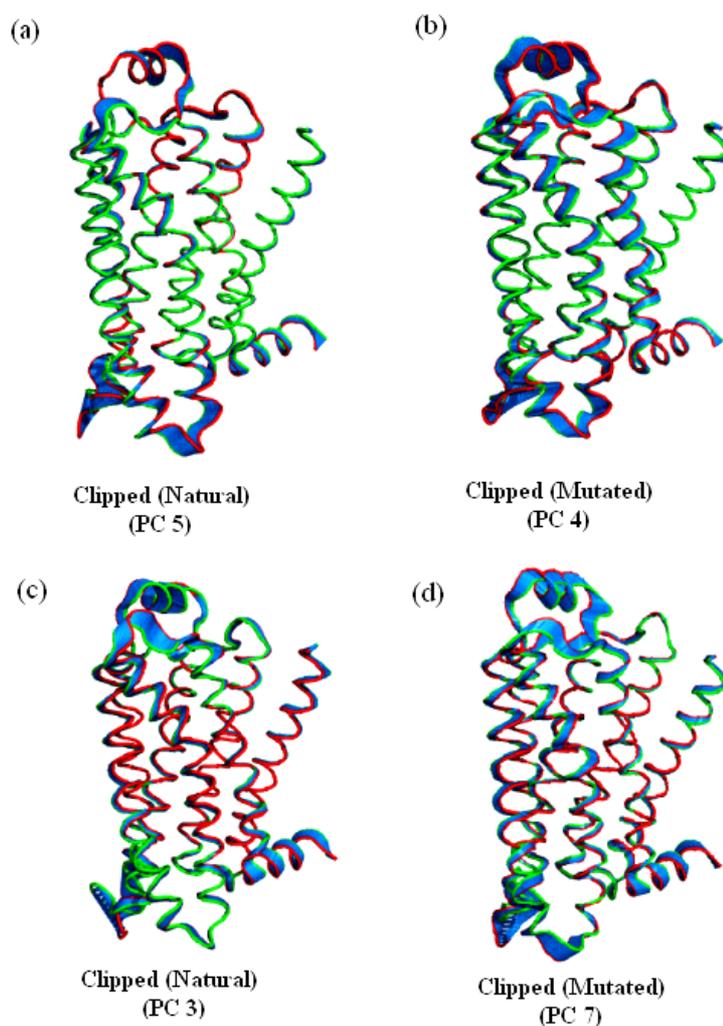


Figure 4.50. Collective motion on PC for (a) the fifth principal mode of clipped (natural), (b) the fourth principal mode of clipped (mutated), (c) the third principal mode of clipped (natural), (d) the seventh principal mode of clipped (mutated).

### 4.5.3. Cross Correlations

The normalized orientational cross-correlations,  $C(i,j)$ , between residue fluctuations are calculated by Equation 4.2.

#### 4.5.3.1. Difference of Orientational Cross Correlations between Looped and Clipped Models

The difference of the orientational cross-correlation between the cumulative first two principal modes of the looped model and the cumulative first eleven modes of the clipped

(natural) model are shown in Figure 4.51a, while the difference of the cross-correlation maps of the cumulative first two principal modes between the looped model and the cumulative first eleven modes of the clipped (mutated) model are shown in Figure 4.51b.

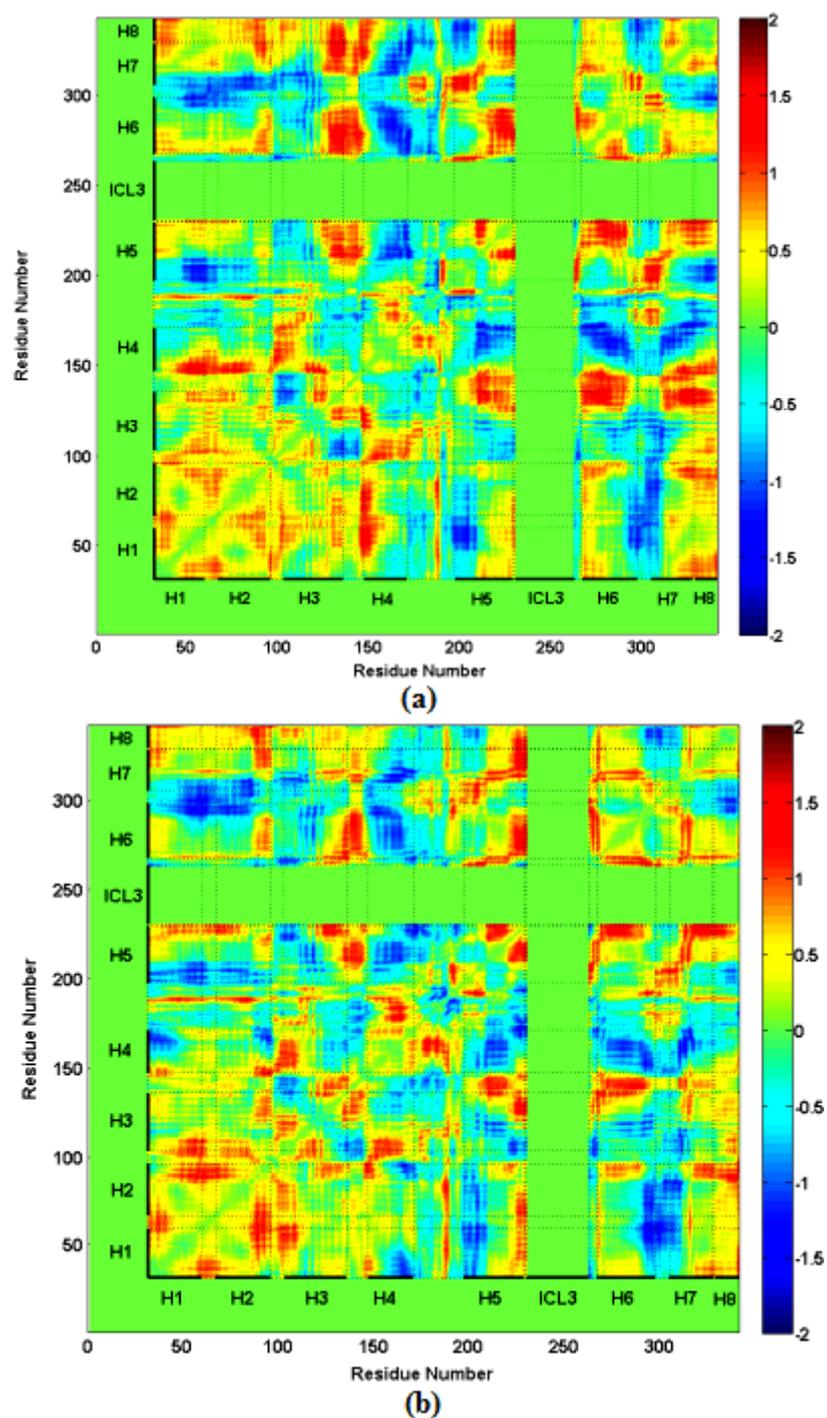


Figure 4.51. Difference of the cross-correlation between (a) looped and clipped (natural) model and between (b) looped and clipped (mutated) model.

Moreover, the map, which has been generated by extracting these two difference maps from each other, is shown in Figure 4.52, expressing the difference of the cross-correlation maps between clipped (natural) model and clipped (mutated) model.

According to this, the difference of the cross-correlation between looped and clipped (natural) model is higher than the difference of the cross-correlation between looped and clipped (mutated) model. As expected, the existence of the intracellular loop (ICL3) has a significant impact on the looped model's dynamics. The regions with the highest difference in cross correlations between looped and clipped models are shown as red and blue spots that dominate in Figure 4.52 over the green spots which signify no difference.

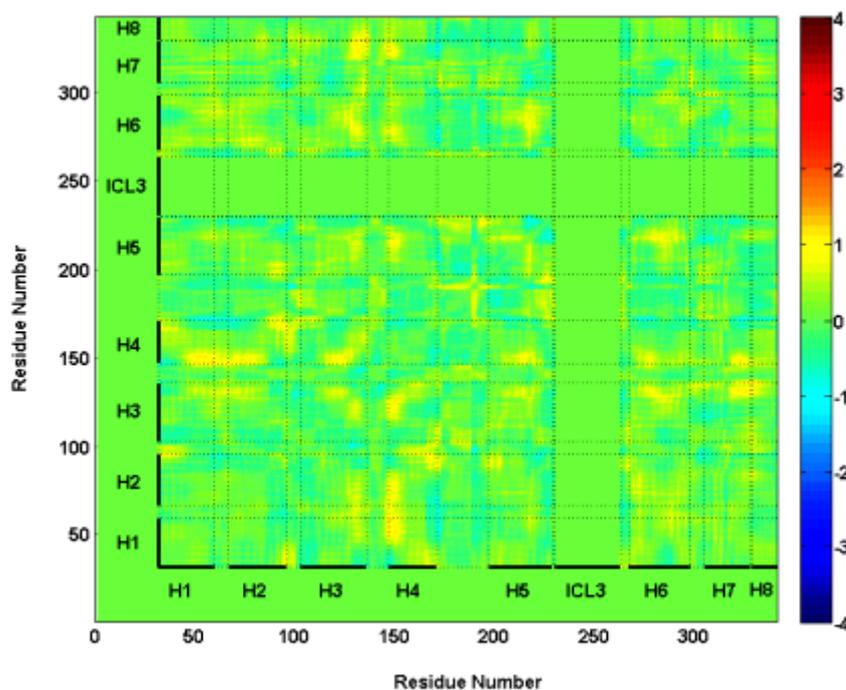


Figure 4.52. Difference of the cross correlations between Figure 4.51a, b.

To further clarify the difference between looped and clipped models, the cumulative cross-correlation values of the first two principal components of looped model and the first eleven components of both clipped models are listed in Tables 4.6 and 4.7. In these tables, the high positive (0.6, 1), high negative (-0.6, -1), mediate positive (0.4, 0.6), mediate negative (-0.4, -0.6), low positive (0.2, 0.4) and low negative (-0.2, -0.4) correlation values are shown with the signs (++ , - -), (+, -) and ((+), (-)) that represent the high, mediate and low positive and negative correlation values, respectively. More detailed cross-correlation



Table 4.7. The cross-correlation values of cumulative two PCs of looped model and eleven PCs of clipped (mutated) models.

		Cumulative 2 PC of Looped											
		H1	H2	H3	ICL2	H4	ECL2	H5	ICL3	H6	ECL3	H7	H8
Cumulative 11 PC of Clipped (Mutated)	H1		++	(+)			(-)		-		--	(+)	++
	H2	(+)		+		(+)	(-)		-		--		++
	H3					(+)					-		+
	ICL2	(-)	(-)			-	(-)	+		+		+	(+)
	H4	(+)					(+)	-		-	-	-	
	ECL2	(-)	(-)	(-)	(+)			(-)	(+)	(-)	(+)	(-)	-
	H5				(-)					+	(+)	+	
	ICL3										(+)	(-)	-
	H6				(-)						(+)	+	
	ECL3				(-)								--
	H7	(+)	(+)				(-)				(+)		(-)
	H8	(+)					(-)					(+)	

#### 4.5.3.2. Difference of Orientational Cross Correlations between two Clipped Models

The orientational cross correlation maps of two clipped models are shown in Figure 4.53. These maps have been generated using the cumulative first eleven principal components, which express 70% of the overall dynamics of the protein. As in looped model, high positive and negative values are encountered in both clipped models. However, the cumulative cross correlation of clipped (natural) model seem to be lower than that of clipped (mutated) model.

Still, some specific regions at both of these models are correlated to each other. For example, H1 is in a high positive correlation with H2 and H8 at both models. H2, on the

other hand is in a high negative correlation with ICL2 and ECL2. However, some correlation values seem to act in opposition to what is expected, i.e., the correlation value of two regions belonging to a clipped (natural) model is negative, while it is positive for the clipped (mutated) model.

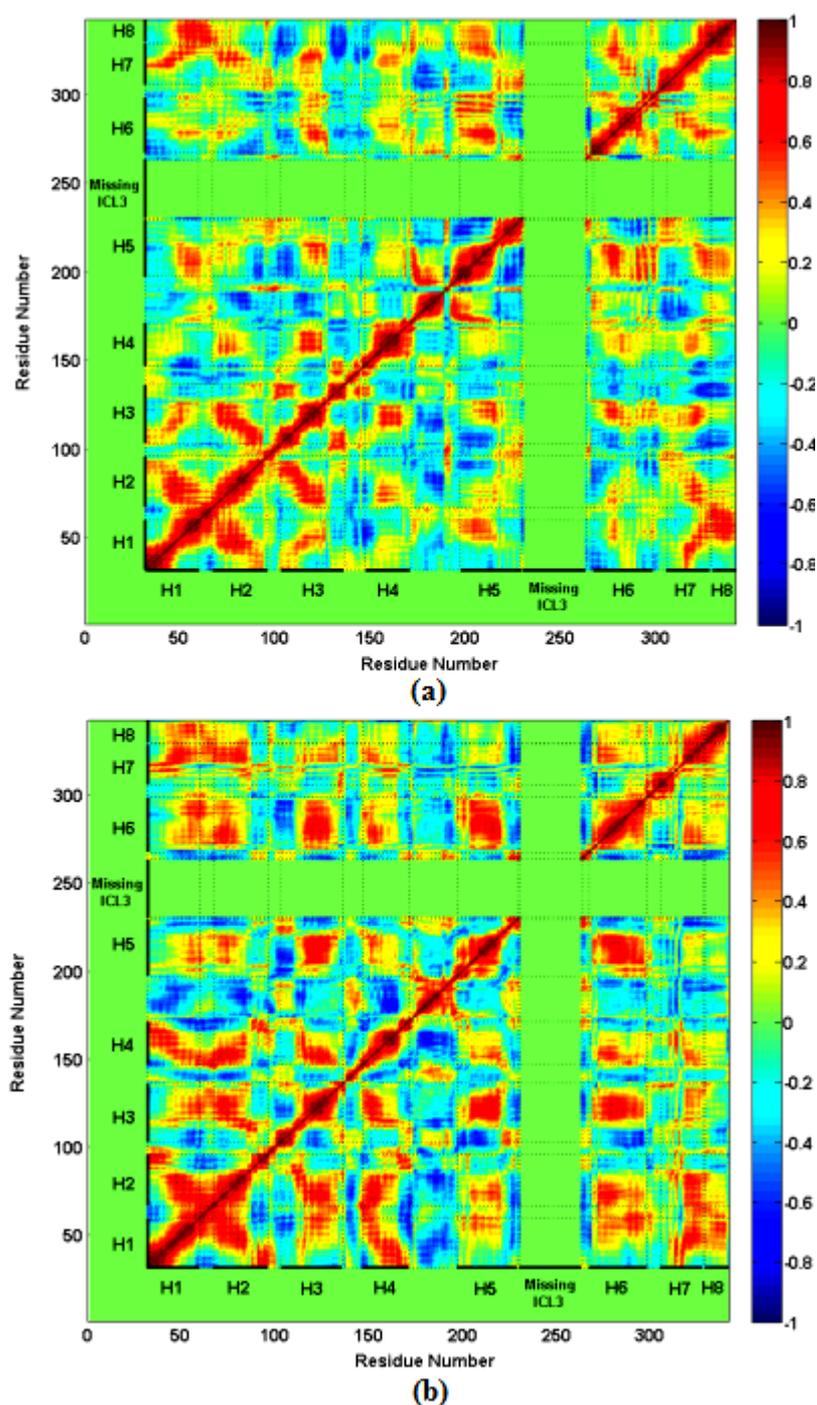


Figure 4.53. Orientational cross-correlation map of cumulative first 11 principal modes for (a) clipped (natural) model, (b) clipped (mutated) model.

For example, H1 region is negatively correlated with H4 and H6 regions in the clipped (natural) model, while it is positively correlated with the same helices in the clipped (mutated) model. Similarly, H6 of the clipped (natural) seems to be negatively correlated with H1, H2 and H3, while it is positively correlated with the same helices in the clipped (mutated) model. Figure 4.54 displays the difference cross-correlation map belonging to the 11 cumulative modes of these two models.

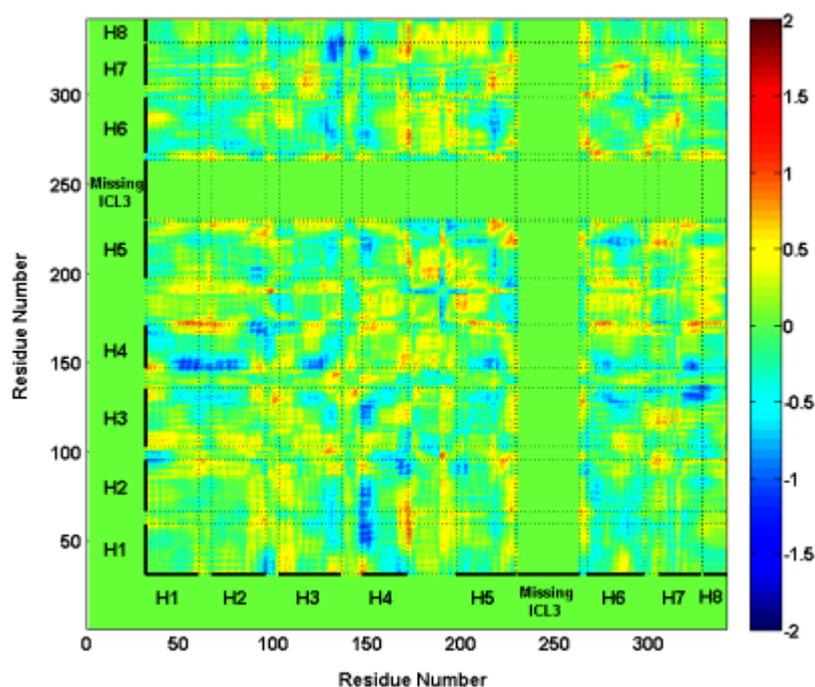


Figure 4.54. Difference of the cross correlation between Figures 4.53a, b.

As seen in Figure 4.54, several transition regions having positive and negative correlations are spotted more clearly. Regions that have a positive correlation value at one model and a negative correlation value at the other one have a higher correlation value at the difference map and are expressed in dark red and dark blue in the map. Other regions have lower correlations and have values close to zero. It could be stated that this differentiation does not stem from mutation. It is also supported by the results obtained from two MD runs with different initial velocities belonging to two identical initial structures. Their overall dynamics are virtually the same; however, regional differentiations could lead to a number of different outcomes. To better understand the difference between clipped (natural) and clipped (mutated), Table 4.8 is created by calculating the cumulative cross-correlation values of the first 11 PC modes of both

clipped models. In this table, only the highest positive and negative difference correlation values are shown with (+) and (-) signs. More detailed cross-correlation tables are shown in Table A.2 and A.3.

Table 4.8. Cross-correlation values of cumulative eleven PCs of both clipped models.

		Cumulative 11 PC of Clipped (Natural)										
		H1	H2	H3	ICL2	H4	ECL2	H5	H6	ECL3	H7	H8
Cumulative 11 PC of Clipped (Mutated)	H1		(+)									(+)
	H2	(+)		(+)	(-)		(-)				(+)	
	H3						(-)					(-)
	ICL2	(-)	(-)					(-)	(-)		(-)	(-)
	H4	(+)										
	ECL2	(-)	(-)	(-)	(+)						(-)	
	H5				(-)							
	H6				(-)							
	ECL3				(-)						(+)	
	H7	(+)	(+)				(-)			(+)		
	H8	(+)					(-)				(+)	

## 4.6. Clustering

The main purpose of clustering is to represent the long MD trajectory with a few numbers of distinct conformational states. It is performed via  $k$ -means algorithm implemented in `kclust` module of Multiscale Modeling Tools of Structural Biology (MMTSB) Tool Set (Feig *et al.*, 2004). In  $k$ -means clustering algorithm,  $n$  numbers of conformations or MD snapshots are clustered into a certain number of clusters (assume  $k$  clusters). For each cluster, a centroid is defined. The next step is to assign each conformation to the nearest centroid, based on RMSD value. After each conformation is assigned,  $k$  new centroids are calculated and each conformation is reassigned to the  $k$  new centroids. This creates a loop where centroids are updated and conformations are reassigned. When there is no more change in the centroids, the loop exits. As the RMSD threshold increases, naturally, the number of clusters will also decrease. In this thesis, selected RMSD values are set to get about four or five clusters.

### 4.6.1. Clustering of Overall MD Trajectory

To understand the extent of conformational change, all simulations of looped and clipped models are merged and clustered at different RMSD thresholds. MD snapshots are clustered based on the binding-site residues located within 8 Å of carazolol, and the transmembrane region in order to get distinct conformations. These conformations can be later used in docking studies in order to distinguish the active (agonist-bound) form from the inactive (antagonist-bound) form of the receptor. The total length of the simulations for all looped and clipped models is 2.1  $\mu$ s. The length of the first looped simulation is 800 ns that produce 4060 conformations. The lengths of the second, third and fourth MD simulations for looped model are each 100 ns that lead to nearly 500 conformations each. These form successively the frames numbered between 4060-4560, 4560-5080 and 5080-5607 that follow the 800 ns simulation for the looped model. It is aimed to sample a different region of the conformational space by these three independent 100 ns simulations, which have different initial velocity assignments. Two separate simulations of 500 ns each are conducted for the clipped (natural) and clipped (mutated) models, respectively. A total of 5000 snapshots are collected for clipped models. In order to align and cluster clipped and looped conformations together, the ICL3 region is extracted from all looped model frames. A total of 10608 frames for clipped and looped models are first aligned based on

the transmembrane region, which comprises all residues located within the boundaries of the cell membrane. The frames are then clustered based on the binding-site region, which consists of residues located within 8 Å of carazolol in the crystal structure of  $\beta_2$ AR. Four clusters are obtained when a threshold of 2.15 Å is selected for the RMSD with all atoms.

The distribution of clusters according to the frame numbers is shown in Figure 4.55. The beginning of all simulations, which include the equilibration periods of both looped and clipped models, fall to the cluster number 3. However, the conformations of the production part of the looped simulations are mostly found in two clusters (numbered 1 and 2), which are distinct from the conformations of the clipped models (Figure 4.55). This situation points to the effect of ICL3 on the conformational dynamics of the binding-site region of the structure. The representative member of each cluster is presented as a surface in Figure 4.56 in comparison to the crystal structure with carazolol, which is added later in order to highlight the changes around carazolol.

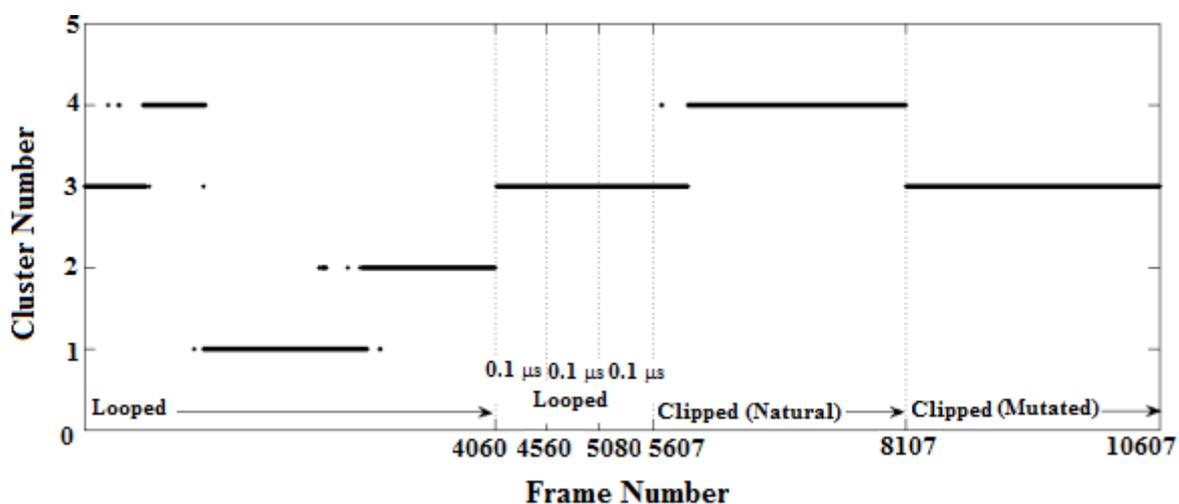


Figure 4.55. Clustering of the all simulations according to binding-site region using a RMSD threshold of 2.15 Å.

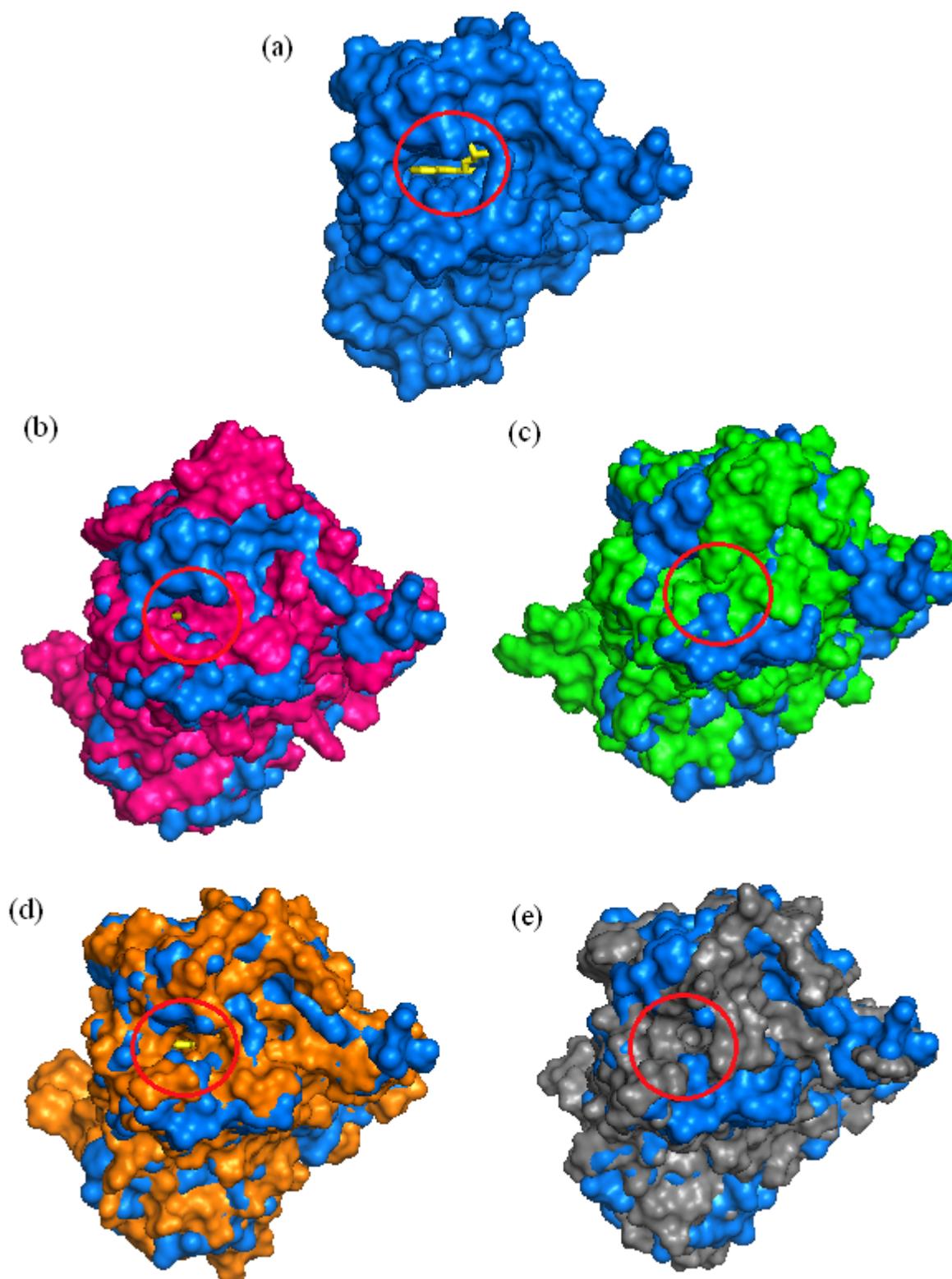


Figure 4.56. The best representative members of the clusters are aligned on crystal structure. (a) Crystal structure (blue) with carazolol (yellow) encircled; Representative from (b) Cluster1 (pink); (c) Cluster2 (green); (d) Cluster3 (orange); (e) Cluster4 (gray).

Similar clustering is performed based on the transmembrane region only using an RMSD value of 1.5 Å for all atoms. As shown in Figure 4.57, the clustering profile reveals four clusters. Similar to binding site based clustering; distinct conformations could not be observed from three short MD runs of the looped model and two clipped models. The distribution of the conformations into the clusters does not follow exact the same profiles as in the case of the binding site-based clustering. The correlation in the profiles between binding site and transmembrane regions is stronger in clipped models than in looped models.

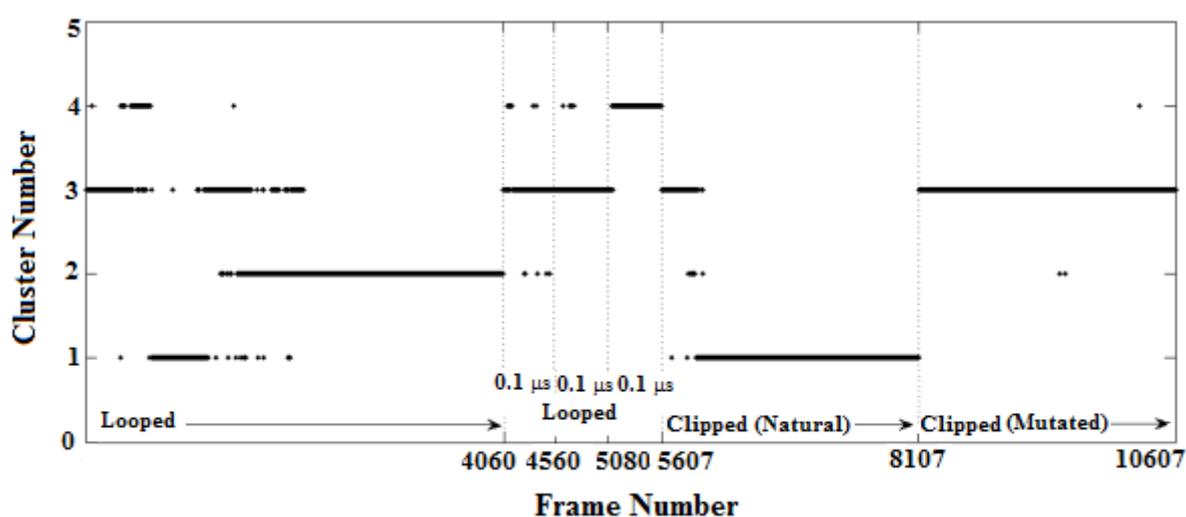


Figure 4.57. Clustering of the all simulations according to transmembrane region using a RMSD threshold of 1.5 Å.

The correlation in the profiles is quantified by calculating the number of matching frames between two clusters, one from the binding site-based clustering, and one from the transmembrane-based clustering. The percentage values, as well as the number of matching frames, are listed in Table 4.9, in two separate sheets. In this table, some cells that have the highest percentage of matching frame are highlighted to display the correlation between clusters. The upper sheet calculates the percentage of frames in Centroid 1 that belongs to transmembrane-based clustering found in four different centroids that belong to binding-site-based clustering. The lower sheet is vice versa. The correlation is observed to be higher in the lower sheet, where almost all frames from Centroids 2, 3, and 4 from binding-site-based clustering are observed in distinct centroids of transmembrane-based clustering, such as Centroids 2, 3 and 1, respectively.

Table 4.9. The number of matching frames between binding-site- and transmembrane-based clustering.

Binding-Site (Radius:2.15Å)	Transmembrane (Radius:1.5Å)							
	Centroid1 (2711)		Centroid2 (2475)		Centroid3 (4715)		Centroid4 (706)	
	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame
Centroid1 (1550)	0.81	22	45.55	1135	8.31	392	0.14	1
Centroid2 (1321)	0	0	53.37	1321	0	0	0	0
Centroid3 (4993)	0.44	12	0.73	18	91.05	4293	94.90	670
Centroid4 (2743)	98.75	2677	0.04	1	0.54	50	4.96	35

Transmembrane (Radius:1.5Å)	Binding-Site (Radius:2.15Å)							
	Centroid1 (1550)		Centroid2 (1321)		Centroid3 (4993)		Centroid4 (2743)	
	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame
Centroid1 (2711)	1.42	22	0	0	0.24	12	97.59	2677
Centroid2 (2475)	73.23	1135	100	1321	0.36	18	0.04	1
Centroid3 (4715)	25.29	392	0	0	85.98	4293	1.09	30
Centroid4 (706)	0.06	1	0	0	13.42	570	1.28	35

All 10608 MD snapshots are also clustered based on loop regions excluding ICL3. The distribution of clusters with respect to the frame number as shown in Figure 4.58 reveals four clusters when a threshold of 3.6 Å is selected for the RMSD of all atoms. No new cluster has been observed in the simulation of two clipped models and three short MD runs.

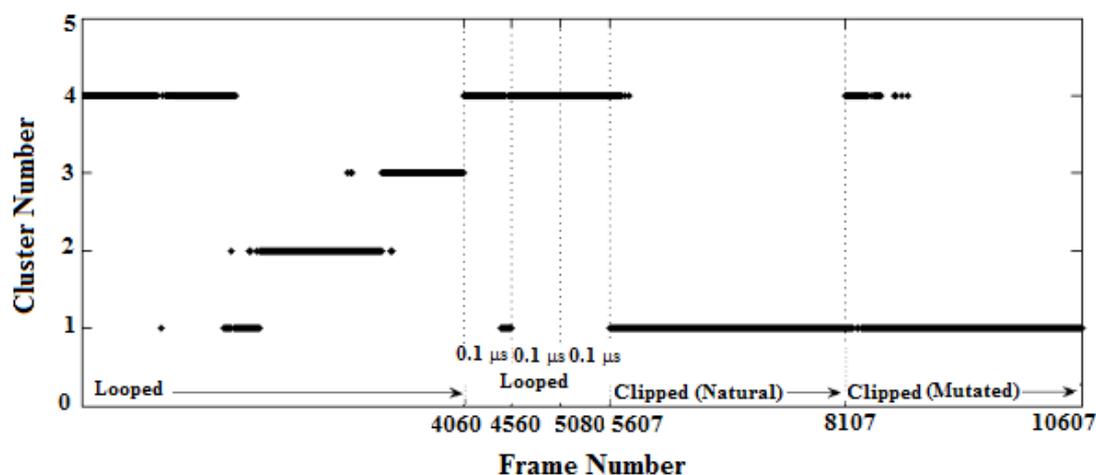


Figure 4.58. Clustering of the all simulations according to the overall structure excluding ICL3 using a RMSD threshold of 3.6 Å.

The same cluster also includes the conformers of the looped model from the long MD trajectory (800 ns). Three short independent MD runs for the looped model did not provide any new clusters. They are mostly found on the fourth cluster, which also include the conformers of the looped model from the first half of the long MD trajectory. The percentage of matching frames between clusters obtained from clustering based on loop, binding-site and transmembrane regions are shown in Table 4.10. The representative conformation of each cluster is presented as a cartoon in Figure 4.59 in comparison to the crystal structure with carazolol, which is added later in order to highlight the changes around carazolol.

Table 4.10. The number of matching frames between loop region, binding-site and transmembrane-based clustering.

Loop Regions (Radius: 3.6Å)	Binding-Site (Radius:2.15Å)							
	Centroid1 (1550)		Centroid2 (1321)		Centroid3 (4993)		Centroid4 (2743)	
	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame
Centroid1 (5013)	16.65	258	0	0	52.09	2601	78.53	2154
Centroid2 (1310)	55.68	863	33.84	447	0	0	0	0
Centroid3 (874)	0	0	66.16	874	0	0	0	0
Centroid4 (3410)	27.68	429	0	0	47.91	2392	21.47	589

Loop Regions (Radius: 3.6Å)	Transmembrane (Radius: 1.5Å)							
	Centroid1 (2711)		Centroid2 (2475)		Centroid3 (4715)		Centroid4 (706)	
	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame
Centroid1 (5013)	79.45	2154	9.82	243	55.46	2615	0.14	1
Centroid2 (1310)	0.15	4	50.26	1244	1.31	62	0	0
Centroid3 (874)	0	0	35.31	874	0	0	0	0
Centroid4 (3410)	20.40	553	4.61	114	43.22	2038	99.86	705

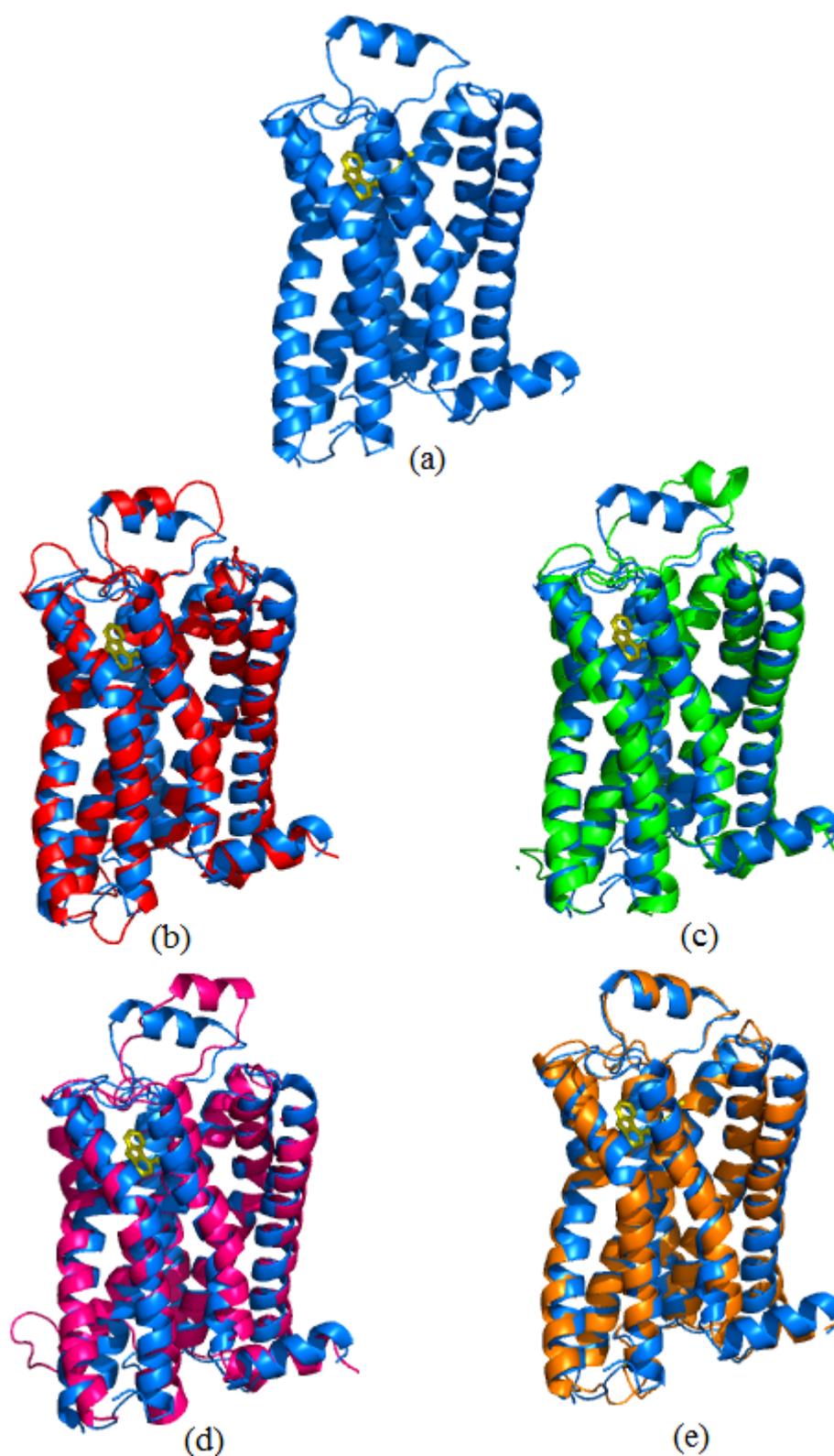


Figure 4.59. The representative members of the clusters of overall structure: (a) Blue: Crystal structure; yellow: Carazolol; (b) red: Cluster1; (c) green: Cluster2; (c) pink: Cluster3; (e) orange: Cluster4.

#### 4.6.2. Clustering of Looped Model

To better understand the transitions between conformations and to look for correlations between different regions of the structure, looped model is clustered at different RMSD thresholds. The distribution of clusters obtained from clustering based on binding-site and transmembrane regions are shown in Figure 4.60.

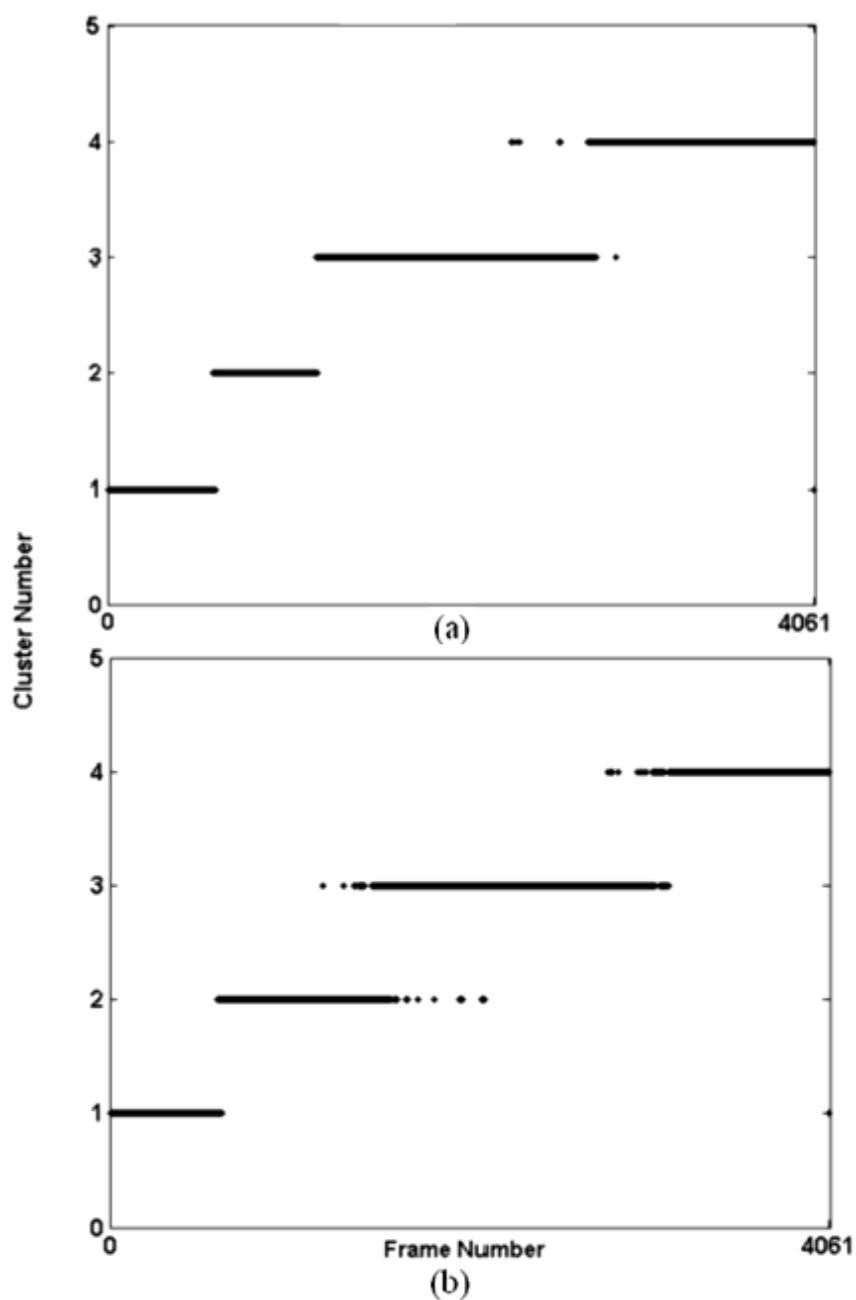


Figure 4.60. The distribution of clusters for the looped model according to frame number obtained from clustering based on (a) binding-site region, and (b) transmembrane region.

The length of the simulation is set to 800 ns for the looped model and produces 4061 conformations. MD snapshots are clustered based on the binding-site and transmembrane region in order to see the correlation between these two regions. The representative conformers from each cluster of the looped model can be seen in Figure 4.61.

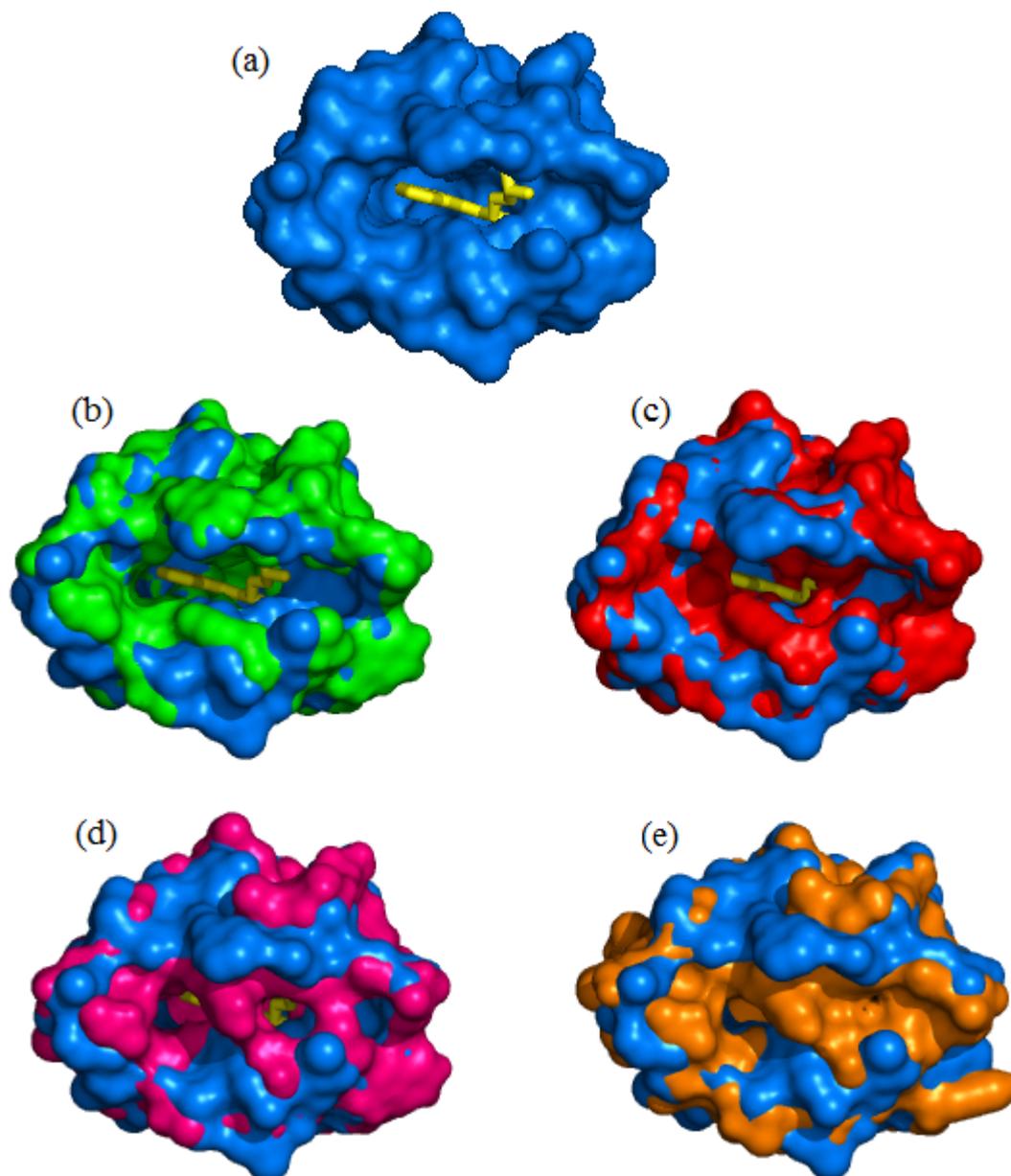


Figure 4.61. The best representative members of the clusters of binding-site region are aligned on the crystal structure. (a) Blue: Crystal structure; yellow: Carazolol; (b) green: Cluster1; (c) red: Cluster2; (d) pink: Cluster3; (e) orange: Cluster4.

The clustering analysis of both binding-site region and transmembrane region, which is shown in Figure 4.60a and 4.60b, revealed four clusters when 1.8 Å threshold is selected for the RMSD for all atoms. The members of the clusters appear at different times of the trajectory of looped model for both clustering. Although, the transitions between conformations are more distinct for the clustering of binding-site than the clustering of transmembrane region, the distribution of the conformations into the clusters follows the same profiles in Figure 4.60a, b. All clusters are perfectly compatible with each other. Hence, it is possible to observe a considerable amount of correlation between binding-site and transmembrane region. The percentage of matching frames between the clusters for binding-site and transmembrane regions are calculated in order to quantify their correlation as shown in Table 4.11.

Table 4.11. The number of matching frames between binding-site and transmembrane-based clustering for looped model.

Binding-Site (Radius:1.8Å)	Transmembrane (Radius:1.3Å)							
	Centroid1 (629)		Centroid2 (920)		Centroid3 (1554)		Centroid4 (957)	
	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame
Centroid1 (616)	97.93	616	0	0	0	0	0	0
Centroid2 (580)	2.07	13	61.85	567	0	0	0	0
Centroid3 (1569)	0	0	38.15	351	78.38	1218	0	0
Centroid4 (1295)	0	0	0	0	21.62	336	100	957

#### 4.6.3. Clustering of Clipped (Natural) and Clipped (Mutated) Models

Clipped models are clustered separately at different RMSD thresholds. The lengths of the simulations are set to 500 ns and produce 2500 conformations for each model. MD snapshots are clustered based on the binding-site and transmembrane region in order to get distinct conformational samplings and display the correlation between these two regions. The distributions of clusters of binding-site and of transmembrane regions according to frame number are shown for clipped (natural) in Figure 4.62a, b respectively. Four different clusters with clear transitions are obtained for the clustering of binding-site and transmembrane regions of clipped (natural) model. 1.6 Å and 1.51 Å RMSD thresholds are selected for binding-site and transmembrane regions respectively. From the clusters profile, it was shown that there is no significant correlation between the clusters of the transmembrane and binding-site region for clipped (natural) model.

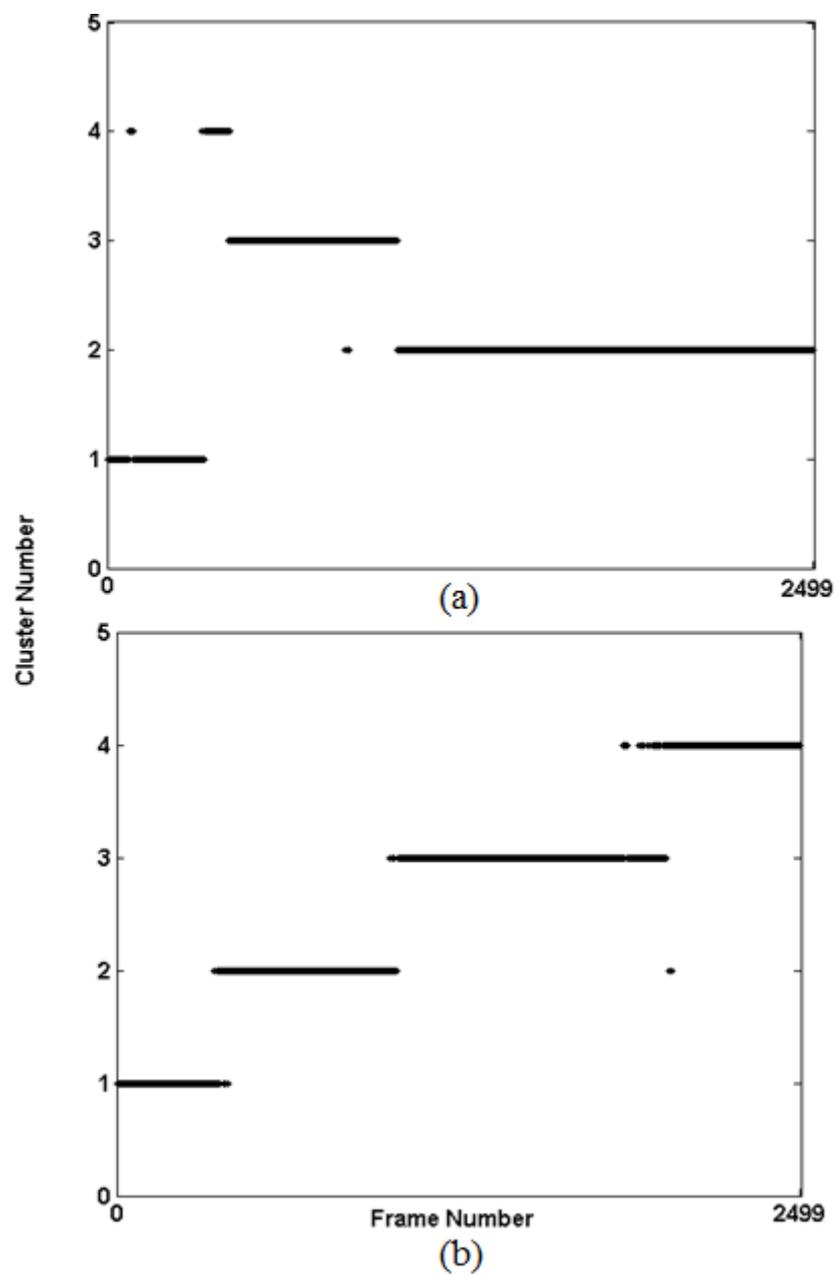


Figure 4.62. The distribution of cluster of clipped (natural) model according to frame number for (a) binding-site region (RMSD threshold 1.6 Å) and (b) transmembrane region (RMSD threshold 1.51 Å).

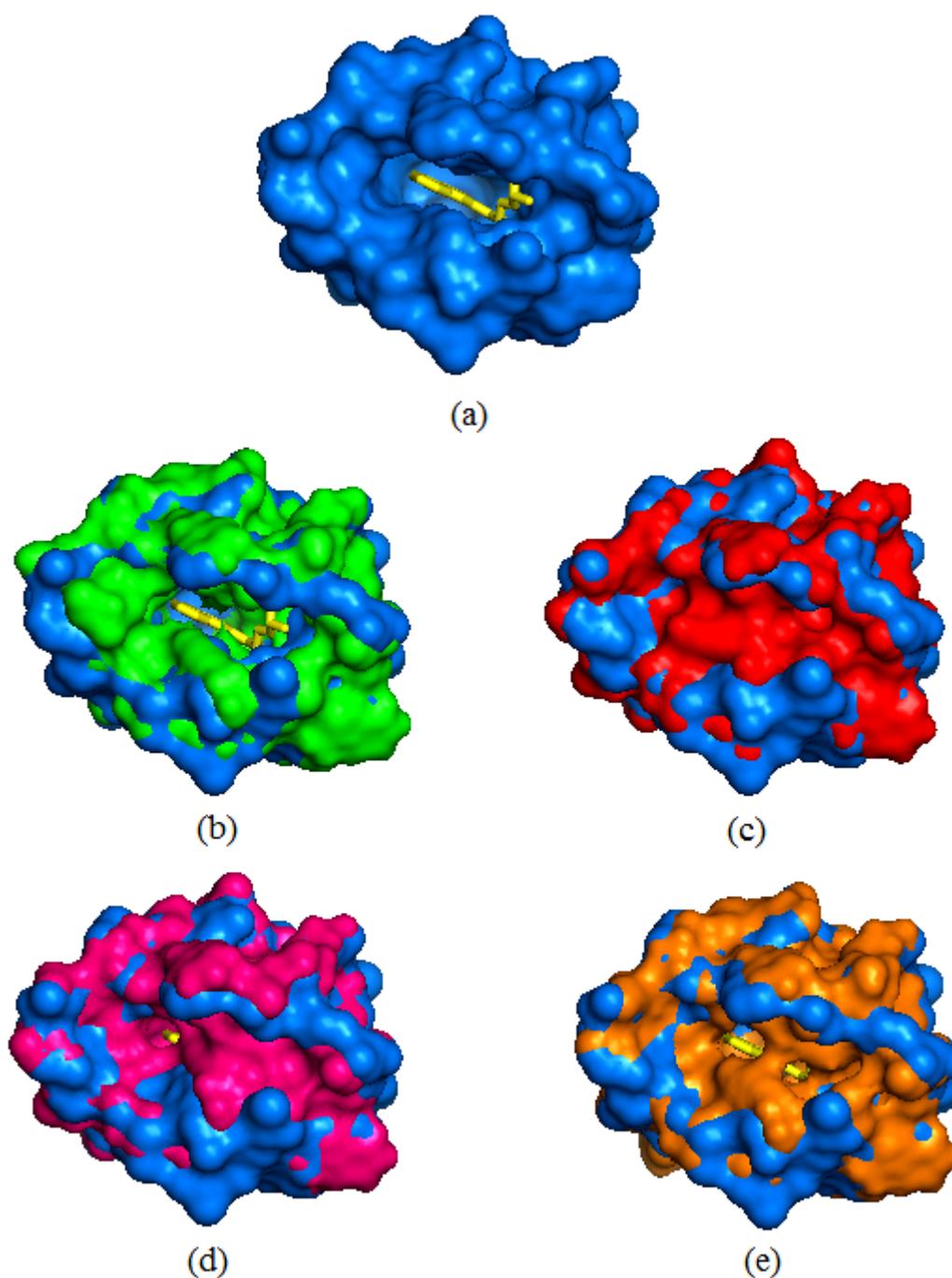


Figure 4.63. The best representative members of the clusters of binding-site region of clipped (natural) are aligned on the crystal structure. (a) Blue: Crystal structure; yellow: Carazolol; (b) green: Cluster1; (c) red: Cluster2; (d) pink: Cluster3; (e) orange: Cluster4.

The representative conformers of each cluster superimposed with the crystal structure are illustrated in Figure 4.63 and Figure 4.64 for binding-site and transmembrane regions in surface and cartoon representation respectively.

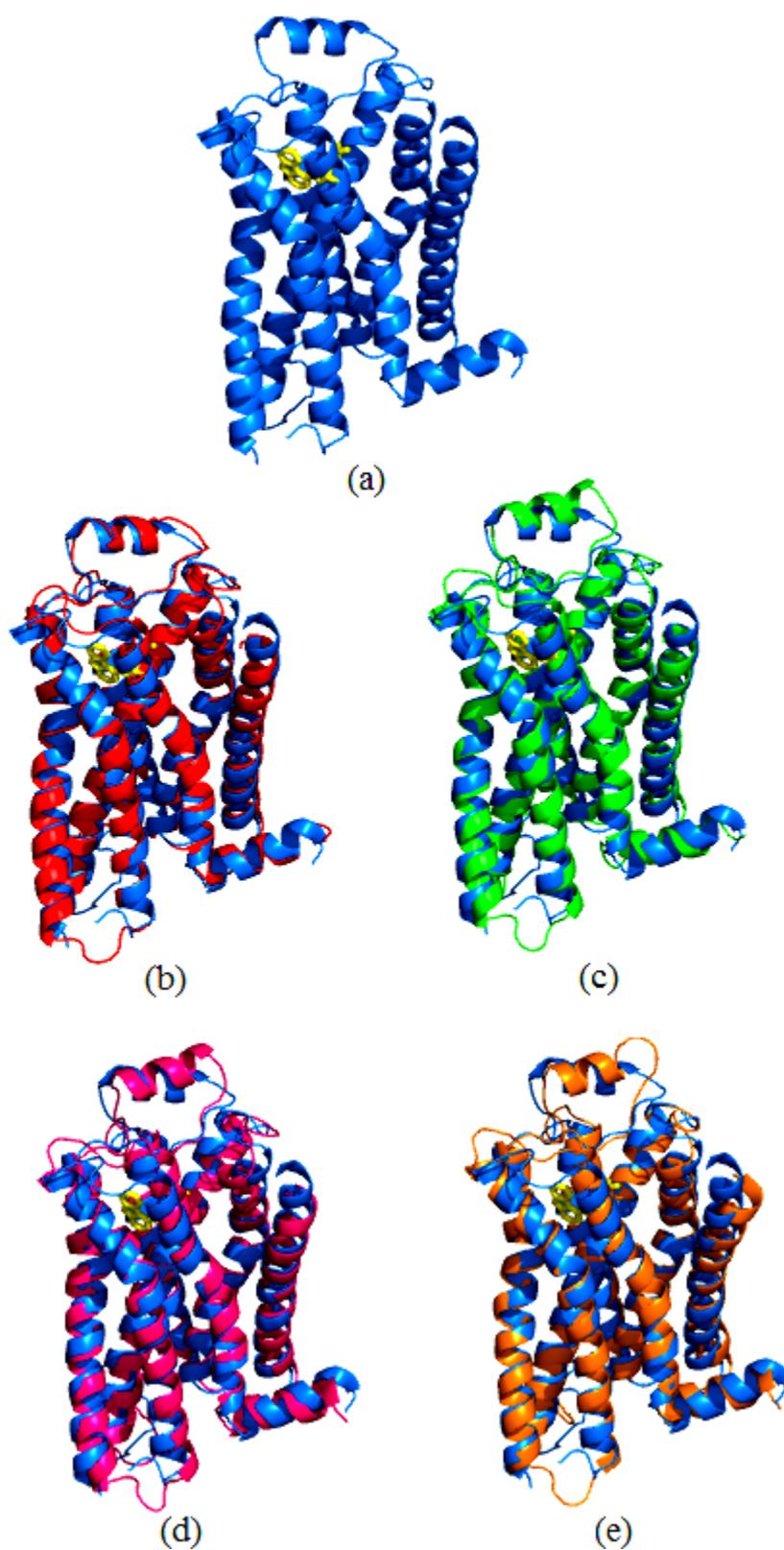


Figure 4.64. The best members of the clusters of transmembrane region of clipped (natural) are aligned on the crystal structure. (a) Blue: Crystal structure; yellow: Carazolol; (b) red: Cluster1; (c) green: Cluster2; (c) pink: Cluster3; (e) orange: Cluster4.

In Figure 4.62, it is clear that the trajectory is divided into four distinct regions. This indicates that once the receptor adopts a certain conformation, it remains in the same cluster, until there is a transition to another conformation. As mentioned before, there is not a strong correlation between the distribution of the clusters for the transmembrane and the binding-site regions. The extent of correlation is also given in Table 4.12 as the percentage of the matching frames between two clusters coming from clustering based on transmembrane and the binding-site regions.

Table 4.12. The number of matching frames between binding-site and transmembrane-based clustering for clipped (natural) model.

Transmembrane (Radius:1.51Å)	Binding-Site (Radius:1.6Å)							
	Centroid1 (327)		Centroid2 (1479)		Centroid3 (589)		Centroid4 (105)	
	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame
Centroid1 (375)	100	327	0	0	0	0	45.71	48
Centroid2 (651)	0	0	0.61	9	99.32	585	54.29	57
Centroid3 (943)	0	0	63.49	939	0.68	4	0	0
Centroid4 (531)	0	0	35.90	531	0	0	0	0

Similar clustering analysis is performed for the clipped (mutated) model for which the conformations are clustered based on binding-site and transmembrane regions for all atoms with 1.75 Å and 1.58 Å RMSD thresholds, respectively. The distributions of clusters of binding-site and of transmembrane regions according to frame number are shown in Figure 4.65a, b respectively. In Figure 4.65a, the transitions between conformations are more distinct for the clustering of binding-site than transmembrane region. The conformations of the clustering of binding-site region distributed to each cluster are approximately different. However, in Figure 4.65b, a clear transition is not observed in the clustering of transmembrane regions. The conformations passed into each cluster and the conformational ensemble shifted to different space of the same cluster. Hence, there appeared to be absolutely no relationship between the clusters of the binding-site and transmembrane regions for clipped (mutated) model. The extent of correlation is also given in Table 4.13, which shows the weakest correlation among all three models.

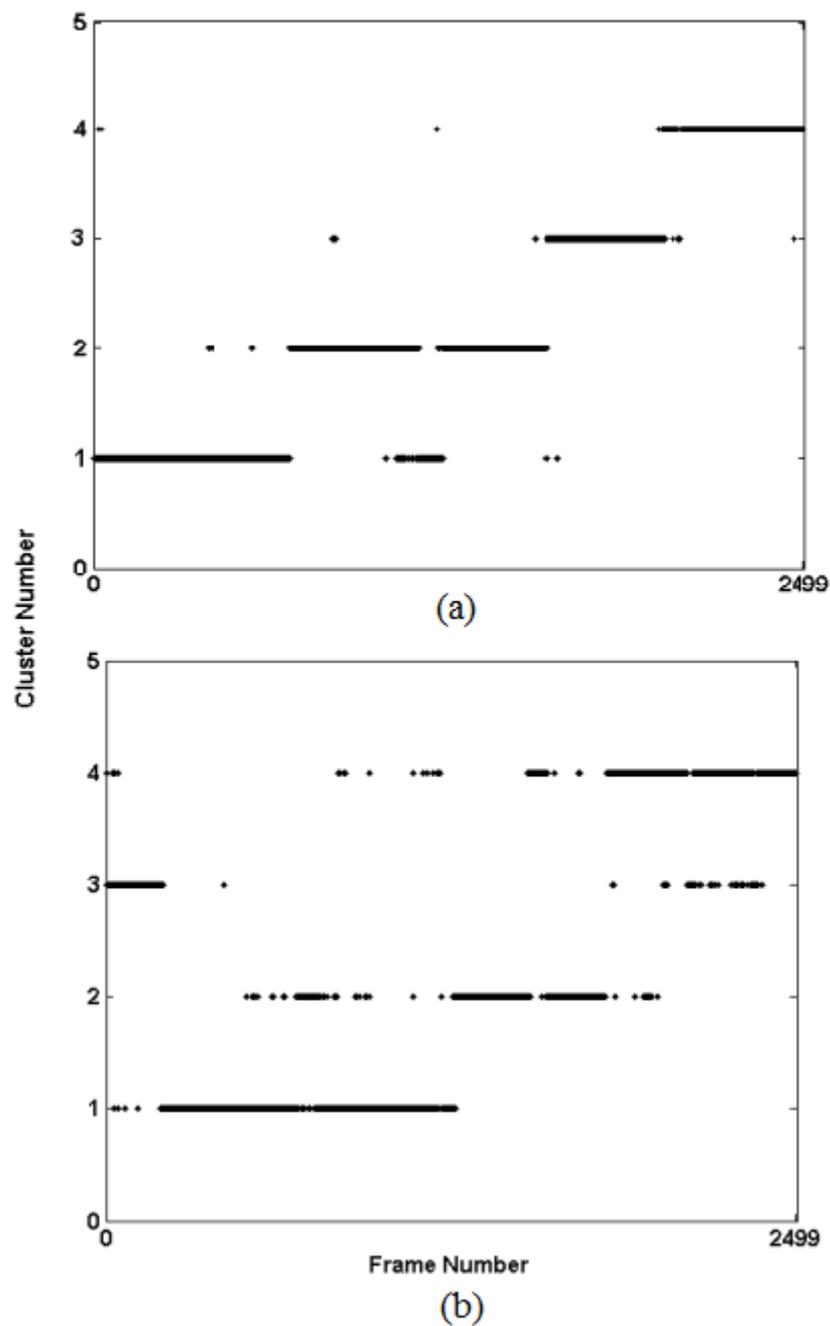


Figure 4.65. The distribution of cluster of clipped (mutated) model according to frame number for (a) binding-site region (RMSD threshold 1.75 Å), (b) transmembrane region (RMSD threshold 1.58 Å).

The representative conformers of each cluster superimposed with the crystal structure are illustrated in Figure 4.66 and 4.67 for binding-site and transmembrane regions in surface and cartoon representation, respectively.

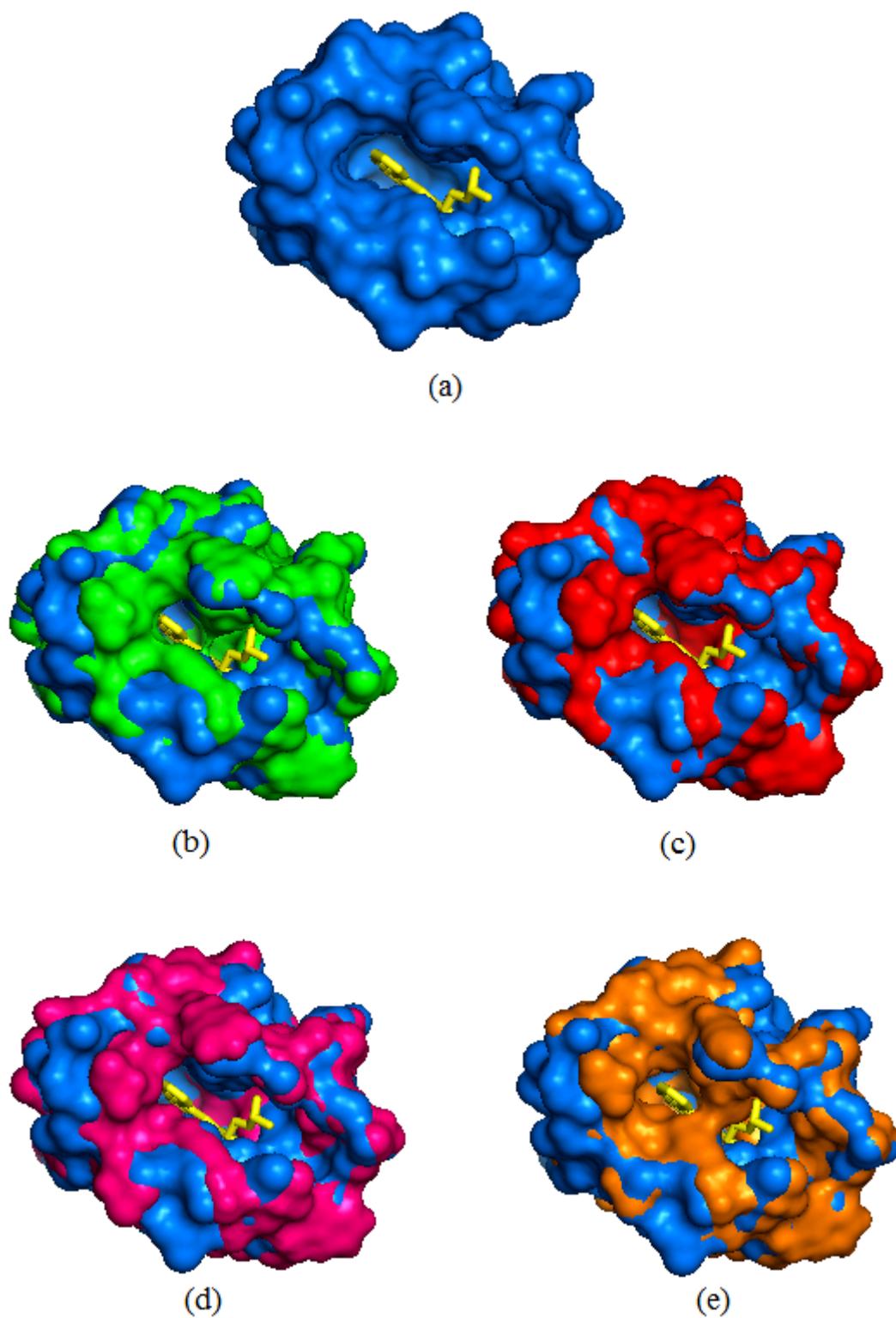


Figure 4.66. The best representative members of the clusters of binding-site region of clipped (mutated) are aligned on the crystal structure. (a) Blue: Crystal structure; yellow: Carazolol; (b) green: Cluster1; (c) red: Cluster2; (d) pink: Cluster3; (e) orange: Cluster4.

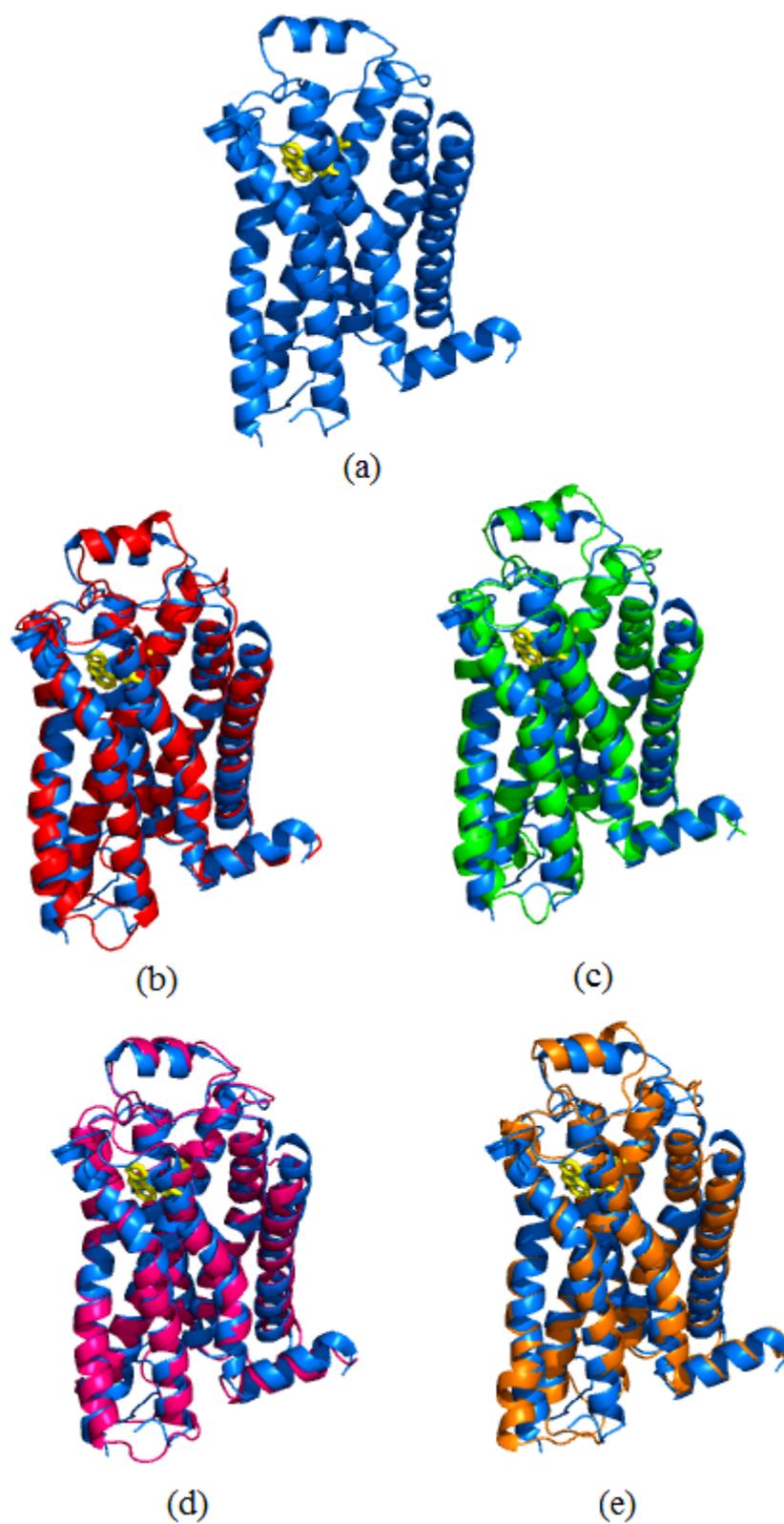


Figure 4.67. The best members of the clusters of transmembrane region of clipped (mutated) are aligned on the crystal structure: (a) Blue: Crystal structure; yellow: Carazolol; (b) red: Cluster1; (c) green: Cluster2; (d) pink: Cluster3; (e) orange: Cluster4.

Table 4.13. The number of matching frames between binding-site and transmembrane-based clustering for clipped (mutated) model.

Binding-Site (Radius:1.75Å)	Transmembrane (Radius:1.58Å)							
	Centroid1 (938)		Centroid2 (609)		Centroid3 (272)		Centroid4 (681)	
	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame	Matching (%)	Matching Frame
Centroid1 (778)	58.96	553	2.63	16	69.12	188	3.08	21
Centroid2 (810)	40.62	381	60.10	366	0	0	9.25	63
Centroid3 (413)	0.32	3	37.27	227	0.74	2	26.58	181
Centroid4 (499)	0.11	1	0	0	30.15	82	61.09	416

## 5. CONCLUSIONS AND FUTURE STUDIES

### 5.1. Conclusions

$\beta_2$ AR is the most studied member of the adrenergic family of GPCR. With regard to physiological relevance, it is known to influence a variety of biological functions. Hence, in order to explore the dynamics of the receptor and determine the transition pathway of the structure from active to inactive states, namely the Anisotropic Network Model (ANM) and Molecular Dynamics (MD) simulations are carried on for looped and clipped models. The major aim of this thesis is to explore the structure-function relationship of the human  $\beta_2$ AR and the effect of the generated loop on receptor's dynamics.

In the first part of the results, the structure-function relationship of the receptor is addressed via Anisotropic Network Model (ANM). Accordingly, the greatest mobility is in intracellular loop III (ICL3) for looped model. Also, intracellular (ICL1, ICL2, ICL3) and extracellular (ECL1, ECL2, ECL3) loop regions are highly mobile parts of both looped and clipped models. The loop regions represent functionally important parts of the receptor, having a critical role in the activation and signaling process.

In order to observe the correspondence between the slow modes of the looped and clipped models, the level of agreement between the collective motions of clipped and looped model is calculated. Accordingly, high compatibility between the modes of looped and clipped models appear after fifth mode of the looped model, because the first four modes are unique to the looped model since they represent the movement of ICL3 only.

Similar ANM analysis was carried out for the dimeric form of the receptor, which is generated using the C2 symmetry operation. To observe the agreement between the collective motions of the dimer and monomer, one monomer of the dimeric structure was randomly selected and the overlap values are calculated between monomer of the dimeric structure and monomer for both models. Accordingly, the first two modes of the monomer of the dimeric looped model are not compatible with any modes of monomer looped model. Similarly, the first four modes of the monomer of the dimeric clipped model are not compatible with any modes of the monomer clipped model. These two modes of looped

and four modes of clipped models represent unique conformations of dimer structure. However, the effect of dimerization is more pronounced in clipped models and more specific conformations are observed due to lack of ICL3.

In the second part of the results, all-atom MD simulations of an apo- $\beta_2$ AR were carried on for 0.8  $\mu$ s for looped models and 0.5  $\mu$ s for each clipped model. The clipped models achieved equilibrium at an RMSD value of approximately 2-2.5 Å, but the equilibrium level of the core region, which does not include ICL3, of the looped model achieves the equilibrium at an RMSD value of around 4 Å due to the effect of the ICL3 on the dynamics of the protein. Clearly, the ICL3 region increases the mobility of the core region. As expected, the highest mobility of the protein is observed at ICL3 of the looped model. Following ICL3, the next mobile regions for both models are detected at ICL2, ECL2 and ECL3. The calculation of the core and transmembrane regions for both models indicates that the conformations of the other loop regions are less constrained in the looped model than clipped models, because these regions of the looped model have greater mobility than clipped model, which is an effect of ICL3 on the conformational dynamics of the receptor. In addition, some conformational changes are observed in mainly two parts of the looped structure, namely ICL3 and ECL2. The first region is ICL3, whose RMSD increases to about 2-2.5 Å around 450 ns and it is converted into an alpha and 3-10 helix structure. At around the same time, the second region is ECL2, well preserved in both clipped models, is converted into a 'Turn' or 'PI' helix structure in looped model. Thus, the change in conformation observed in ECL2 may be related to the conformational change in ICL3. No considerable variation has been noticed at other regions and the structural stability has been preserved throughout the simulation.

One of the most significant structural changes occurring during the deactivation process of protein is the degradation of the salt bridges that are situated between Asp130 and Arg131, between Arg131 and Glu268 and between Asp192 and Lys305. Opening of the Arg131-Glu268 is one of the critical conformational changes required for the receptor to get activated. This bridge is observed as being open in the looped model for the longest time (40% open, 60% closed during the 800 ns simulation). In the contrast, the bridge is open in 16% and 6% of both clipped models. This difference between looped and clipped is clearly a consequence of the ICL3, which incorporates conformational flexibility to the

protein. However, no obvious correlation has been noticed on three salt bridges regarding the durations they remained closed/opened and it has been revealed that they act rather independently without being affected from each other. The switching of the salt bridge (Arg131-Glu268) from closed to open states stems mainly from the change in angle  $\chi_1$ . The change in angle  $\chi_1$  of Glu268 seems to be in correlation with the change in the N-O distance shown for the looped and clipped (natural) model. The same correlation is observed in angle  $\chi_1$  of Glu268 with the change in the C $\alpha$  - C $\alpha$  distance. However, correlation between the change in any angle  $\chi_1$  of ionic lock residues and the change in any salt bridge distances is not obtained for clipped (mutated).

Next, PCA method is used to reduce large dimensional data set of MD trajectory to a few dimensions in order to get the important and more pronounced motions of the receptor. Accordingly, looped model has the highest explanation value in the first mode. The percentage explanation value of the first PC, which is considered to explain the protein's overall dynamics to a considerable degree, included in the looped model is 59.3. Specifically, the first five modes could satisfactorily explain the protein's entire dynamics with a cumulative percentage explanation value of 84.6%. As for the clipped models, their explanation values are comparatively lower than the looped model, which also indicate the effect of ICL3 on the conformational dynamics of the protein. There is also a slight difference between the explanation values of the mutated and natural clipped models (the first principal components are 27.7 and 25.3 for the mutant and the natural models, respectively). Loop regions and transmembrane helices are more rigid than looped model due to lack of ICL3 in the clipped models.

The level of agreement between the collective motions of clipped models and looped model are examined using overlap calculation. PC of looped model reflects a low overlap with both PC of clipped models. This distinction stems from ICL3 taking place at the looped model. The first five PCs of the looped model have a low overlap value with the PCs of the clipped models. The modes after the sixth PC seem to be relatively more coherent. This coherence lays out overlap values ranging from 0.3 to 0.5. In light of this evaluation, it could be concluded that each and every PC that takes place at the looped model has different dynamics than those of the PCs taking place at the clipped models and that this distinction poses a very important role in the overall dynamics of the structure.

Asn187 has been replaced with Glu187 to be mutated. The aim here was to see that one single residue would not affect the dynamics of the system considerably. To reach a definite conclusion overlap matrix has been calculated with the results obtained from PCA and it shows that the modes of two clipped models are in an exact compatibility. Though the compatibility values of the modes are not high, two models move in a correlated manner. It could also be said that different collective motions of the principal modes between two models do not stem from mutation, but from the MD simulations conducted at different initial conditions of the two identical systems.

Also, orientational cross-correlations between residue fluctuations are calculated for both models. Being in motion as coupled structures with each other, these regions have high correlation values for the both of the models. However, these values of the looped model tend to be highly positive and negative, while they are seen to have diminished to almost a quarter of them in the clipped models. This could be explained by the fact that the dynamic structure of the looped model is completely different than those of the clipped models because of the utmost mobility of region ICL3.

In the second part of the results, *k*-means clustering method is used to determine the distinct conformational states from the MD trajectory. All simulations, which include the trajectories of looped and clipped models, are clustered together based on binding-site, transmembrane and core regions. Although the members of the clusters may appear at different times of the trajectory for the clustering of binding-site, transmembrane and core regions of the receptor, slight correlation between binding-site and transmembrane regions and binding-site and core regions are obtained. No new conformational state is obtained from the short MD simulations of the looped models and the MD simulations of the two clipped models.

Afterwards, each model is clustered within itself. For the looped model, the distribution profile of the clusters along the trajectory is similar for binding-site and transmembrane regions, i.e., there is a strong correlation in the cluster-to-cluster transitions between binding-site and transmembrane regions. For the clipped (natural) model, the same correlation is moderate, whereas for the clipped (mutated) model, no correlation is observed.

## 5.2. Future Studies

In the future, the protein will be inserted in the membrane and ANM will be performed to observe the constraints imposed on the conformational dynamics. Also, atomistic ANM will be applied to protein-ligand complex in order to determine the ligand-binding effects (such as adrenaline, noradrenaline, carazolol) on the dynamics of the receptor.

Oligomerization of GPCR is a well-established fact based on both experimental and theoretical findings. However, it is still unclear whether the complex form presents a new binding site, potential new ligands for activation and alternative signaling pathways. Here, the dimeric form of  $\beta_2$ AR, as a representative of the minimal oligomeric arrangement will be explored using the same methodology as for the monomeric form. However, MD simulation of the oligomeric form of the receptor in atomistic detail is a challenging task, which will be overcome by residue-based coarse-grained modeling.

## APPENDIX A: CROSS-CORRELATION TABLES FOR BOTH MODELS

Table A.1. The cross-correlation values of cumulative two PCs of looped model.

	H1	H2	H3	ICL2	H4	ECL2	H5	ICL3	H6	ECL3	H7	H8
H1		87,82 0,00 12,18 0,00 0,72	23,53 0,81 65,21 10,45 0,31	17,24 0,00 55,52 27,24 0,15	21,10 4,55 35,45 38,90 0,10	4,97 39,03 11,72 44,28 -0,33	5,02 32,60 53,29 9,09 -0,07	5,50 49,67 10,25 34,58 -0,39	1,94 19,72 58,41 19,94 0,00	0,00 94,25 0,00 5,75 -0,75	50,72 20,98 18,82 9,48 0,22	94,83 0,00 5,17 0,00 0,77
H2	87,82 0,81 12,18 0,00 0,72		45,69 0,00 50,88 3,43 0,47	19,33 0,00 49,33 31,33 0,18	37,07 3,60 32,00 27,33 0,26	7,87 29,60 9,33 53,20 -0,28	6,46 25,35 49,29 18,89 -0,04	0,99 53,51 9,46 36,04 -0,46	13,54 20,31 46,77 19,38 0,02	0,00 89,44 0,00 10,56 -0,75	46,53 23,06 20,69 9,72 0,18	96,67 0,00 3,33 0,00 0,75
H3	23,53 0,81 65,21 10,45 0,31	45,69 0,00 50,88 3,43 0,47		39,71 27,65 16,76 15,88 0,09	48,94 17,41 18,12 15,53 0,25	8,82 21,29 31,65 38,24 -0,11	25,85 23,80 19,88 30,48 0,02	5,25 37,92 22,89 33,94 -0,30	33,64 24,54 13,24 28,58 0,05	0,00 35,78 2,94 61,27 -0,39	32,35 18,01 15,32 34,31 0,05	36,76 0,00 54,20 9,03 0,39
ICL2	17,24 0,00 55,52 27,24 0,15	19,33 0,00 49,33 31,33 0,18	39,71 27,65 16,76 15,88 0,09		4,40 54,00 11,60 30,00 -0,41	0,80 44,00 14,40 40,80 -0,36	63,94 2,73 23,94 9,39 0,47	9,46 37,03 12,97 40,54 -0,29	75,00 1,88 15,94 7,19 0,55	0,00 16,67 63,33 20,00 -0,03	52,50 3,75 39,58 4,17 0,46	20,71 0,00 68,57 10,71 0,27
H4	21,10 4,55 35,45 38,90 0,10	37,07 3,60 32,00 27,33 0,26	48,94 17,41 18,12 15,53 0,25	4,40 54,00 11,60 30,00 -0,41		24,00 2,40 44,16 29,44 0,20	0,00 46,06 12,97 40,97 -0,44	16,43 21,41 30,27 31,89 -0,04	0,25 61,38 13,25 25,13 -0,52	0,00 44,00 11,33 44,67 -0,39	5,50 52,33 18,50 23,67 -0,38	20,00 8,00 32,86 39,14 0,06
ECL2	4,97 39,03 11,72 44,28 -0,33	7,87 29,60 9,33 53,20 -0,28	8,82 21,29 31,65 38,24 -0,11	0,80 44,00 14,40 40,80 -0,36	24,00 2,40 44,16 29,44 0,20		12,85 33,58 16,48 37,09 -0,21	37,62 10,49 31,03 20,86 0,22	6,25 36,50 20,50 36,75 -0,27	32,67 10,00 44,00 13,33 0,25	6,33 45,17 19,33 29,17 -0,30	3,71 50,00 10,57 35,71 -0,40
H5	5,02 32,60 53,29 9,09 -0,07	6,46 25,35 49,29 18,89 -0,04	25,85 23,80 19,88 30,48 0,02	63,94 2,73 23,94 9,39 0,47	0,00 46,06 12,97 40,97 -0,44	12,85 33,58 16,48 37,09 -0,21		19,00 34,81 17,12 29,07 -0,12	61,74 0,19 25,57 12,50 0,56	37,37 0,51 11,62 50,51 0,24	55,30 1,52 16,04 27,15 0,38	33,33 25,97 12,99 0,09
ICL3	5,50 49,67 10,25 34,58 -0,39	0,99 53,51 9,46 36,04 -0,46	5,25 37,92 22,89 33,94 -0,30	9,46 37,03 12,97 40,54 -0,29	16,43 21,41 30,27 31,89 -0,04	37,62 10,49 31,03 20,86 0,22	19,00 34,81 17,12 29,07 -0,12		16,89 39,10 18,07 25,93 -0,18	33,33 0,90 52,25 13,51 0,35	15,54 47,18 19,71 17,57 -0,23	6,56 61,78 5,02 26,64 -0,45
H6	1,94 19,72 58,41 19,94 0,00	13,54 20,31 46,77 19,38 0,02	33,64 24,54 13,24 28,58 0,05	75,00 1,88 15,94 7,19 0,55	0,25 61,38 13,25 25,13 -0,52	6,25 36,50 20,50 36,75 -0,27	61,74 16,89 39,10 18,07 25,93 -0,18	19,00 34,81 17,12 29,07 -0,12		27,08 0,00 31,77 41,15 0,23	61,20 3,78 18,62 16,41 0,47	30,13 17,19 40,85 11,83 0,16
ECL3	0,00 94,25 0,00 5,75 -0,75	0,00 89,44 0,00 10,56 -0,75	0,00 35,78 2,94 61,27 -0,39	0,00 16,67 63,33 20,00 -0,03	0,00 44,00 11,33 44,67 -0,39	32,67 10,00 44,00 13,33 0,25	37,37 0,51 11,62 50,51 0,24	33,33 0,90 52,25 13,51 0,35	27,08 0,00 31,77 41,15 0,23		29,17 12,50 8,33 50,00 0,03	0,00 98,81 0,00 1,19 -0,76
H7	50,72 20,98 18,82 9,48 0,22	46,53 23,06 20,69 9,72 0,18	32,35 18,01 15,32 34,31 0,05	52,50 3,75 39,58 4,17 0,46	5,50 52,33 18,50 23,67 -0,38	6,33 45,17 19,33 29,17 -0,30	55,30 1,52 16,04 27,15 0,38	15,54 47,18 19,71 17,57 -0,23	61,20 3,78 18,62 16,41 0,47	29,17 12,50 8,33 50,00 0,03		59,82 20,54 10,12 9,52 0,30
H8	94,83 0,00 5,17 0,00 0,77	96,67 0,00 3,33 0,00 0,75	36,76 0,00 54,20 9,03 0,39	20,71 0,00 68,57 10,71 0,27	20,00 8,00 32,86 39,14 0,06	3,71 50,00 10,57 35,71 -0,40	33,33 25,97 27,71 12,99 0,09	6,56 61,78 5,02 26,64 -0,45	30,13 17,19 40,85 11,83 0,16	0,00 98,81 0,00 1,19 -0,76	59,82 20,54 10,12 9,52 0,30	

Table A.2. The cross-correlation values of cumulative eleven PCs of clipped (natural) model.

	H1	H2	H3	ICL2	H4	ECL2	H5	H6	ECL3	H7	H8
H1		23,91	4,06	0,00	0,00	0,00	3,96	0,96	0,00	13,65	26,35
		0,00	0,81	12,07	5,38	0,83	1,22	2,59	0,57	0,00	0,00
		58,74	54,87	35,86	56,00	25,24	39,45	42,05	31,03	49,86	40,64
		17,36	40,26	52,07	38,62	73,93	55,38	54,41	68,39	36,49	33,00
		0,28	0,07	-0,14	-0,02	-0,13	0,00	-0,06	-0,10	0,14	0,21
H2	23,91		21,08	0,00	3,47	1,07	0,29	0,83	1,67	9,03	3,10
	0,00		0,20	7,00	1,73	10,93	3,63	4,26	5,00	0,28	0,00
	58,74		49,02	19,67	54,80	19,33	61,27	40,09	58,33	71,53	63,33
	17,36		29,71	73,33	40,00	68,67	34,80	54,81	35,00	19,17	33,57
	0,28		0,20	-0,20	0,04	-0,20	0,03	-0,07	-0,02	0,19	0,11
H3	4,06	21,08		5,00	11,41	0,12	4,07	5,72	1,47	5,76	0,00
	0,81	0,20		0,00	2,24	8,94	4,93	3,10	8,33	9,44	12,82
	54,87	49,02		45,29	44,24	26,24	45,85	38,40	41,18	45,22	27,52
	40,26	29,71		49,71	42,12	64,71	45,16	52,78	49,02	39,58	59,66
	0,07	0,20		0,00	0,05	-0,17	-0,02	-0,02	-0,08	0,00	-0,19
ICL2	0,00	0,00	5,00		4,40	0,00	0,29	0,00	0,00	0,00	0,00
	12,07	7,00	0,00		1,20	0,40	3,24	3,06	0,00	12,08	5,71
	35,86	19,67	45,29		23,60	47,20	14,12	10,28	11,67	14,17	10,71
	52,07	73,33	49,71		70,80	52,40	82,35	86,67	88,33	73,75	83,57
	-0,14	-0,20	0,00		-0,11	-0,02	-0,18	-0,22	-0,18	-0,22	-0,22
H4	0,00	3,47	11,41	4,40		1,12	0,59	0,89	0,00	1,00	0,00
	5,38	1,73	2,24	1,20		1,92	1,76	0,33	0,00	2,00	0,57
	56,00	54,80	44,24	23,60		31,84	56,00	65,44	54,67	65,17	19,71
	38,62	40,00	42,12	70,80		65,12	41,65	33,33	45,33	31,83	79,71
	-0,02	0,04	0,05	-0,11		-0,11	0,03	0,09	0,02	0,09	-0,13
ECL2	0,00	1,07	0,12	0,00	1,12		9,76	1,56	2,67	1,33	0,29
	0,83	10,93	8,94	0,40	1,92		5,18	0,44	8,67	11,67	1,71
	25,24	19,33	26,24	47,20	31,84		36,12	57,44	30,00	18,17	66,00
	73,93	68,67	64,71	52,40	65,12		48,94	40,56	58,67	68,83	32,00
	-0,13	-0,20	-0,17	-0,02	-0,11		-0,02	0,05	-0,13	-0,23	0,07
H5	3,96	0,29	4,07	0,29	0,59	9,76		14,87	9,31	5,39	0,42
	1,22	3,63	4,93	3,24	1,76	5,18		5,47	0,00	0,74	2,73
	39,45	61,27	45,85	14,12	56,00	36,12		51,88	56,86	55,15	59,66
	55,38	34,80	45,16	82,35	41,65	48,94		27,78	33,82	38,73	37,18
	0,00	0,03	-0,02	-0,18	0,03	-0,02		0,13	0,14	0,08	0,09
H6	0,96	0,83	5,72	0,00	0,89	1,56	14,87		9,26	7,87	1,19
	2,59	4,26	3,10	3,06	0,33	0,44	5,47		0,00	0,12	0,99
	42,05	40,09	38,40	10,28	65,44	57,44	51,88		50,93	56,94	55,75
	54,41	54,81	52,78	86,67	33,33	40,56	27,78		39,81	35,07	42,06
	-0,06	-0,07	-0,02	-0,22	0,09	0,05	0,13		0,15	0,11	0,04
ECL3	0,00	1,67	1,47	0,00	0,00	2,67	9,31	9,26		20,83	0,00
	0,57	5,00	8,33	0,00	0,00	8,67	0,00	0,00		0,00	0,00
	31,03	58,33	41,18	11,67	54,67	30,00	56,86	50,93		55,56	60,71
	68,39	35,00	49,02	88,33	45,33	58,67	33,82	39,81		23,61	39,29
	-0,10	-0,02	-0,08	-0,18	0,02	-0,13	0,14	0,15		0,21	0,07
H7	13,65	9,03	5,76	0,00	1,00	1,33	5,39	7,87	20,83		18,15
	0,00	0,28	9,44	12,08	2,00	11,67	0,74	0,12	0,00		0,00
	49,86	71,53	45,22	14,17	65,17	18,17	55,15	56,94	55,56		31,85
	36,49	19,17	39,58	73,75	31,83	68,83	38,73	35,07	23,61		50,00
	0,14	0,19	0,00	-0,22	0,09	-0,23	0,08	0,11	0,21		0,12
H8	26,35	3,10	0,00	0,00	0,00	0,29	0,42	1,19	0,00	18,15	
	0,00	0,00	12,82	5,71	0,57	1,71	2,73	0,99	0,00	0,00	
	40,64	63,33	27,52	10,71	19,71	66,00	59,66	55,75	60,71	31,85	
	33,00	33,57	59,66	83,57	79,71	32,00	37,18	42,06	39,29	50,00	
	0,21	0,11	-0,19	-0,22	-0,13	0,07	0,09	0,04	0,07	0,12	

Table A.3. The cross-correlation values of cumulative eleven PCs of clipped (mutated) model.

	H1	H2	H3	ICL2	H4	ECL2	H5	H6	ECL3	H7	H8
H1		34,94	5,68	0,00	25,93	0,00	2,94	7,09	2,87	22,84	10,84
		1,49	3,96	6,90	6,21	19,59	2,64	2,87	0,00	0,14	0,00
		38,16	44,02	27,24	51,86	11,45	48,48	51,15	41,38	45,83	73,65
		25,40	46,35	65,86	16,00	68,97	45,94	38,89	55,75	31,18	15,52
		0,27	0,02	-0,16	0,24	-0,29	0,03	0,07	-0,01	0,21	0,25
H2	34,94		21,08	2,00	26,00	0,53	4,61	11,11	0,00	24,31	1,19
	1,49		0,00	13,67	3,60	15,73	3,43	3,33	5,00	0,00	0,95
	38,16		43,53	29,33	48,67	11,07	50,10	48,70	38,33	41,25	62,86
	25,40		35,39	55,00	21,73	72,67	41,86	36,85	56,67	34,44	35,00
	0,27		0,18	-0,17	0,24	-0,28	0,04	0,09	-0,07	0,19	0,07
H3	5,68	21,08		1,76	17,41	0,35	20,07	19,53	0,98	6,13	0,63
	3,96	0,00		2,35	0,24	6,94	9,95	7,19	1,47	0,86	2,31
	44,02	43,53		37,35	48,35	21,29	37,28	41,67	44,12	55,76	39,71
	46,35	35,39		58,53	34,00	71,41	32,70	31,62	53,43	37,25	57,35
	0,02	0,18		-0,09	0,15	-0,17	0,08	0,12	-0,04	0,10	-0,07
ICL2	0,00	2,00	1,76		2,80	1,60	0,29	0,00	0,00	0,00	0,00
	6,90	13,67	2,35		12,40	0,80	0,88	6,67	1,67	0,00	2,14
	27,24	29,33	37,35		21,20	77,60	23,53	22,78	20,00	36,67	34,29
	65,86	55,00	58,53		63,60	20,00	75,29	70,56	78,33	63,33	63,57
	-0,16	-0,17	-0,09		-0,21	0,18	-0,17	-0,19	-0,17	-0,06	-0,08
H4	25,93	26,00	17,41	2,80		1,12	3,65	8,00	0,67	10,33	0,00
	6,21	3,60	0,24	12,40		27,52	1,06	2,22	2,00	0,17	2,00
	51,86	48,67	48,35	21,20		18,40	58,24	57,44	53,33	57,33	62,00
	16,00	21,73	34,00	63,60		52,96	37,06	32,33	44,00	32,17	36,00
	0,24	0,24	0,15	-0,21		-0,29	0,08	0,10	0,00	0,13	0,02
ECL2	0,00	0,53	0,35	1,60	1,12		2,24	1,22	0,00	0,00	0,00
	19,59	15,73	6,94	0,80	27,52		2,94	5,33	0,67	14,00	6,57
	11,45	11,07	21,29	77,60	18,40		38,00	24,89	24,00	11,67	19,43
	68,97	72,67	71,41	20,00	52,96		56,82	68,56	75,33	74,33	74,00
	-0,29	-0,28	-0,17	0,18	-0,29		-0,05	-0,14	-0,12	-0,27	-0,18
H5	2,94	4,61	20,07	0,29	3,65	2,24		29,82	0,98	0,00	1,05
	2,64	3,43	9,95	0,88	1,06	2,94		3,02	0,00	11,03	3,15
	48,48	50,10	37,28	23,53	58,24	38,00		36,76	52,94	45,34	46,85
	45,94	41,86	32,70	75,29	37,06	56,82		30,39	46,08	43,63	48,95
	0,03	0,04	0,08	-0,17	0,08	-0,05		0,20	0,02	-0,08	-0,03
H6	7,09	11,11	19,53	0,00	8,00	1,22	29,82		6,02	6,48	0,20
	2,87	3,33	7,19	6,67	2,22	5,33	3,02		3,70	8,68	5,36
	51,15	48,70	41,67	22,78	57,44	24,89	36,76		51,39	51,62	72,62
	38,89	36,85	31,62	70,56	32,33	68,56	30,39		38,89	33,22	21,83
	0,07	0,09	0,12	-0,19	0,10	-0,14	0,20		0,05	0,04	0,08
ECL3	2,87	0,00	0,98	0,00	0,67	0,00	0,98	6,02		22,22	0,00
	0,00	5,00	1,47	1,67	2,00	0,67	0,00	3,70		1,39	0,00
	41,38	38,33	44,12	20,00	53,33	24,00	52,94	51,39		41,67	63,10
	55,75	56,67	53,43	78,33	44,00	75,33	46,08	38,89		34,72	36,90
	-0,01	-0,07	-0,04	-0,17	0,00	-0,12	0,02	0,05		0,18	0,11
H7	22,84	24,31	6,13	0,00	10,33	0,00	0,00	6,48	22,22		19,05
	0,14	0,00	0,86	0,00	0,17	14,00	11,03	8,68	1,39		0,00
	45,83	41,25	55,76	36,67	57,33	11,67	45,34	51,62	41,67		56,55
	31,18	34,44	37,25	63,33	32,17	74,33	43,63	33,22	34,72		24,40
	0,21	0,19	0,10	-0,06	0,13	-0,27	-0,08	0,04	0,18		0,24
H8	10,84	1,19	0,63	0,00	0,00	0,00	1,05	0,20	0,00	19,05	
	0,00	0,95	2,31	2,14	2,00	6,57	3,15	5,36	0,00	0,00	
	73,65	62,86	39,71	34,29	62,00	19,43	46,85	72,62	63,10	56,55	
	15,52	35,00	57,35	63,57	36,00	74,00	48,95	21,83	36,90	24,40	
	0,25	0,07	-0,07	-0,08	0,02	-0,18	-0,03	0,08	0,11	0,24	

## REFERENCES

- Atilgan A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin and I. Bahar, 2001, “Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model”, *Biophysical Journal*, Vol. 80, No. 1, pp. 505–515.
- Azzi, M., P. G. Charest, S. Angers, G. Rousseau, T. Kohout, M. Bouvier and G. Pineyro, 2003, “Beta-Arrestin-Mediated Activation of MAPK by Inverse Agonists Reveals Distinct Active Conformations For G-Protein-Coupled Receptors”, *Proceedings of The National Academy of Sciences*, Vol. 100, No. 20, pp. 11406-11411.
- Bahar, I., A. R. Atilgan and B. Erman, 1997, “Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential”, *Folding & Design*, Vol. 2, No. 3, pp. 173–181.
- Bakan, A. and I. Bahar, 2009, “The Intrinsic Dynamics of Enzymes Plays a Dominant Role in Determining the Structural Changes Induced upon Inhibitor Binding”, *Proceedings of the National Academy of Sciences*, Vol. 106, No. 34, pp. 14349–14354.
- Bashford, D. and M. Karplus, 1990, “pKa of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model”, *Biochemistry*, Vol. 29, No. 4, pp: 10219-10225.
- Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, 2000, “The Protein Data Bank”, *Nucleic Acids Research*, Vol. 28, No. 1, pp. 235-242.
- Bhattacharya S., S. E. Hall, H. Li and N. Vaidehi, 2008, “Ligand-Stabilized Conformational States of Human Beta2 Adrenergic Receptor: Insight into G-Protein-Coupled Receptor Activation”, *Biophysical Journal*, Vol. 94, No. 6, pp. 2027–2042.

- Branden, C. and Tooze J., 1999, *Introduction to Protein Structure*, Garland Publishing Inc., New York.
- Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. J. Karplus, 1983, "Charmm: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations", *Journal of Computational Chemistry*, Vol. 4, No. 2, pp. 187-217.
- Cansu, S. and P. Doruker, 2008, "Dimerization Affects Collective Dynamics of Triosephosphate Isomerase", *Biochemistry*, Vol. 47, No. 5, pp. 1358-1368.
- 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 Beta-2 Adrenergic G-Protein Coupled Receptor", *Science*, Vol. 318, No. 5854, pp. 1258-1265.
- Chou K. C., 2005, "Prediction of G-Protein-Coupled Receptor Classes", *Journal of Proteome Research*, Vol. 4, No. 4, pp. 1413-1418.
- Daaka, Y., L. M. Luttrell and R. J. Lefkowitz, 1997, "Switching of the Coupling of the Beta2-Adrenergic Receptor to Different G-Proteins by Protein Kinase A", *Nature*, Vol. 390, No. 6655, pp. 88-91.
- Doruker, P. and R. L. Jernigan, 2003, "Functional Motions Can Be Extracted from On-Lattice Construction of Protein Structures", *Proteins*, Vol. 53, No. 2, pp.174-181.
- Eswar, N., M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper and A. Sali, 2006, "Comparative Protein Structure Modeling with MODELLER", *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., Vol. 15, pp. 1-30.

- Feig, M. J., J. Karanicolas and C. L. Brooks, 2004, "MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology", *Journal of Molecular Graphics and Modeling*, Vol.22, No. 5, pp. 377-395.
- Fredriksson R, M. C. Lagerstrom, L. G. Lundin and H. B. Schioth, 2003, "The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups and Fingerprints", *Molecular Pharmacology*, Vol. 63, No. 6, pp. 1256–1272.
- Fritze, O., S. Filipek, V. Kuksa, K. Palczewski, K. P. Hofmann and O. P. Ernst, 2003, "Role of The Conserved NPxxY(x) 5,6 F Motif In The Rhodopsin Ground State and During Activation", *Proceedings of The National Academy of Sciences*, Vol. 100, No.5, pp. 2290-2295.
- Gether, U., 2000, "Uncovering Molecular Mechanisms Involved In Activation of G-Protein-Coupled Receptors", *Endocrine Reviews*, Vol. 21, No. 1, pp. 90-113.
- Hall, R. A., R. T. Premont, C. W. Chow, J. T. Blitzer, J. A. Pitcher, A. Claing, R. H. Stoffel, R. H. Barak, S. Shenolikar, E. J. Weinman, S. Grinstein and R. J. Lefkowitz, 1998, "The Beta2-Adrenergic Receptor Interacts with The Na<sup>+</sup>/H<sup>+</sup> Exchanger Regulatory Factor to Control Na<sup>+</sup>/H<sup>+</sup> Exchange", *Nature*, Vol. 392, No. 6676, pp. 626-630.
- Hancock, J. T., 1997. "Cell Signaling", *Longman Scientific and Technical*, Harlow, England.
- Hayward, S., 1999, "Structural Principles Governing Domain Motions in Proteins", *Proteins*, Vol. 36, No. 4, pp. 425-435.
- Hebert T. E., S. Moffet, J. P. Morello, T. P. Loisel, D. G. Bichet, C. Barret and M. Bouvier, 1996, "A Peptide Derived From a  $\beta_2$ -Adrenergic Receptor Transmembrane Domain Inhibits Both Receptor Dimerization and Activation", *Jornal of Biological Chemistry*, Vol. 271, No. 27, pp. 16384-16392.

- Hein, L., and B. K. Kobilka, 1995, "Adrenergic Receptor Signal Transduction and Regulation", *Neuropharmacology*, Vol. 34, No. 4, pp. 357-366.
- Houslay, M. D. and W. Kolch, 2000, "Cell-Type Specific Integration of Cross-Talk Between Extracellular Signal-Regulated Kinase and cAMP Signaling", *Molecular Pharmacology*, Vol. 58, No. 4, pp. 659-668.
- Hockney, R. W., 1970, "The Potential Calculations on Some Applications", *Methods in Computational Physics*, Vol. 9, pp. 136-143.
- Humphrey, Q., A. Alke and K. Schulten, 1996, "VMD: Visual Molecular Dynamics", *Journal of Molecular Graphics and Modeling*, Vol. 14, No.1, pp. 33-38.
- Jeffrey, L. B., 2002, "Novel  $\beta_2$ -Adrenergic Receptor Signaling Pathways", *Journal of Allergy and Clinical Immunology*, Vol. 110, No. 6, pp. 229-235.
- Johnson, M., 2006, "Molecular Mechanisms of  $\beta_2$ AR Function, Response and Regulation", *Journal of Allergy and Clinical Immunology*, Vol. 17, No. 1, pp. 18-24.
- Johnson, M., 2001, "Beta2-Adrenoceptors: Mechanisms of Action of Beta2-Agonists", *Paediatric Respiratory Review*, Vol. 2, No. 1, pp. 57-62.
- Katragadda, M., M. W. Maciejewski and P. L. Yeagle, 2004, "Structural Studies of The Putative Helix VIII In The Human  $\beta_2$  Adrenergic Receptor an Nuclear Magnetic Resonance Study", *Biochimica et Biophysica Acta*, Vol. 1663, No. 1-2, pp. 74-81.
- Kenakin, T., 2004, "Principles: Receptor Theory in Pharmacology", *Trends in Pharmacological Sciences*, Vol. 25, No. 4, pp. 186-192.
- Kitao, A., F. Hirata and N. Go, 1991, "The Effects of Solvent on the Conformation and the Collective Motions of a Protein: Normal Mode Analysis and MD Simulations of Melting in Water and in Vacuum", *Chemical Physics*, Vol. 158, No. 2-3, pp. 447-472.

- Klabunde, T. and G. Hessler, 2002, "Drug Design Strategies for Targeting G-Protein-Coupled Receptors", *Chembiochem*, Vol. 3, No. 10, pp. 928-944.
- Kobilka, B. K. and X. Deupi, 2007, "Conformational Complexity of G-Protein-Coupled Receptors", *Trends in Pharmacological Sciences*, Vol.28, No. 8, pp. 397-406.
- Kristiansen, K., 2004, "Molecular Mechanisms of Ligand Binding, Signaling and Regulation Within the Superfamily of G-Protein-Coupled Receptors: Molecular Modeling and Mutagenesis Approaches to Receptor Structure and Function", *Pharmacology and Therapeutics*, Vol. 103, No. 1, pp. 21-80.
- Labute, P., 2009, "*Protonate 3D: Assingment of Macromolecular Protonation State and Geometry*", <http://www.chemcomp.com>, accessed at June 2010.
- Lantzsch, G., H. Binder and 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" 2<sup>nd</sup> Ed; *Prentice Hall*, Harlow, England.
- Leineweber, K. and O. E. Brodde, 2004, "Beta2-Adrenoceptor Polymorphisms: Relation Between in Vitro and in Vivo Phenotypes", *Life Sciences*, Vol. 74, No. 17, pp. 2803-2814.
- Li, J., P. C. Edwards, M. Burghammer, C. Villa and G. F. Schertler, 2004, "Structure of Bovine Rhodopsin in a Trigonal Crystal Form", *Journal of Molecular Biology*, Vol. 343, No. 5, pp. 1409-1438.
- Liggett, S. B., 2002, "Update on Current Concepts of the Molecular Basis of Beta2-Adrenergic Receptor Signaling", *Journal of Allergy and Clinical Immunology*, Vol. 110, No. 6, pp. 223-227.

- Lomize, M. A., A. Lomize, I. D. Pogozheva and H. I. Mosberg, 2006, "Orientations of Proteins in Membranes Database", *Bioinformatics*, Vol. 22, No. 5, pp. 623-625.
- Maggio, R., F. Novi, M. Scarselli and G. U. Corsini, 2005, "The Impact of G-Protein-Coupled Receptor Hetero-Oligomerization on Function and Pharmacology", *FEBS Journal*, Vol. 272, No. 12, pp. 2939-2946.
- Marti-Renom, M. A., A. Stuart, A. Fiser, R. Sánchez, F. Melo and A. Sali, 2000, "Comparative Protein Structure Modeling of Genes and Genomes", *Annual Review of Biophysics and Biomolecular Structure*, Vol. 29, No. 1, pp. 291-325.
- McGraw, D. W. and S. B. Liggett, 2005, "Molecular Mechanisms of Beta2-Adrenergic Receptor Function and Regulation", *Proceedings of the American Thoracic Society*, Vol. 2, No. 4, pp. 292-296.
- Oldham W. M. and H. E. Hamm, 2008, "Heterotrimeric G-Protein Activation by G-Protein-Coupled Receptors", *Nature Reviews Molecular Cell Biology*, Vol. 3, No. 9, pp. 60-71.
- Palczewski, K., T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, T. I. Le, 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, No. 5480, pp. 739-745.
- Pardo, L., X. Deupi, N. Dolker, M. L. Lopes-Rodriguez and M. Campillo, 2007, "The Role of Internal Water Molecules In The Structure and Function of The Rhodopsin Family of G-Protein-Coupled Receptors", *Chembiochem*, Vol. 8, pp. 19-24.
- Parra, S. and R. A. Bond, 2007, "Inverse Agonism: From Curiosity to Accepted Dogma, But Is It Clinically Relevant?", *Current Opinion in Pharmacology*, Vol. 7, No. 2, pp. 146-150.

- Pieper, U., N. Eswar, B. M. Webb, E. David, L. Kelly, D. T. Barkan, H. Carter, P. Mankoo, R. Karchin, M. A. Marti-Renom, F. P. Davis and A. Sali, 2008, "MODBASE, a Database of Annotated Comparative Protein Structure Models and Associated Resources", *Nucleic Acids Research*, Vol. 37, No. 10, pp. 347-354.
- 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.
- Prinster, S. C., C. Hague and R. A. Hall, 2005, "Heterodimerization of G-Protein-Coupled Receptors: Specificity and Functional Significance", *Pharmacological Reviews*, Vol. 57, No. 3, pp. 289-298.
- Rasmussen, S. G., H. J. Choi, D. M. Rosenbaum, T. S. Kobilka, F. S. Thian, P. C. Edwards, M. Burghammer, V. R. Ratnala, R. Sanishvili, R. F. Fischetti, G. F. Schertler, W. I. Weis and B. K. Kobilka, 2007, "Crystal Structure of The Human Beta2 Adrenergic G-Protein-Coupled Receptor", *Nature*, Vol. 450, No. 7168, pp. 383-387.
- Rosenbaum, D. M., S. G. Rasmussen and B. K. Kobilka., 2009, "The Structure and Function of G-Protein-Coupled Receptor", *Nature*, Vol. 459, No. 7245, pp. 356-363.
- Rosenbaum, D. M., V. Cherezov, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, X. J. Yao, W. I. Weis, R. C. Stevens and B. K. Kobilka, 2007, "G-Protein-Coupled Receptor Engineering Yields High-Resolution Structural Insights Into Beta2-Adrenergic Receptor Function", *Science*, Vol. 318, No. 5854, pp. 1266-1273.
- Sali, A. and T. L. Blundell, 1993, "Comparative Protein Modeling by Satisfaction of Spatial Restraints", *Journal of Molecular Biology*, Vol. 234, No. 1, pp. 779-815.
- Schertler, G. F. X., C. Villa and R. Henderson, 1993, "Projection Structure of Rhodopsin", *Nature*, Vol. 362, No. 7, pp. 770-772.

- Soudijn, W., I. V. Wijngaarden and A. P. Ijzerman, 2005, "Structure-Activity Relationships of Inverse Agonists for G-Protein-Coupled Receptors", *Medical Care Research and Review*, Vol. 25, No. 4, pp. 398-426.
- Steinbach, J. P., 2010, "Introduction to Macromolecular Simulation", <http://cmm.cit.nih.gov/intro-simulation>, accessed at July 2010.
- Strader C. D., T. M. Fong, M. R. Tota, D. Underwood and R. A. Dixon, 1994, "Structure and Function of G-Protein-Coupled Receptors", *Annual Review of Biochemistry*, Vol. 63, No. 2, pp. 101-132.
- Strosberg, A. D., 1995, "Structural and functional Diversity of Beta-Adrenergic Receptors", *Annals of the New York Academy of Sciences*, Vol. 757, No. 13, pp. 253-260.
- Strosberg, A. D., 1993, "Structure, Function and Regulation of Adrenergic Receptors", *Protein Science*, Vol. 2, No. 1, pp. 1198-1209.
- Swope, W. C., C. A. Hans, H. B. Peter and R. W. Kent, 1982, "A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters", *Journal of Chemical Physics*, Vol. 76, No. 4, pp. 637-650.
- Szidonya, L., M. Cserző and L. Hunyady, 2008, "Dimerization and Oligomerization of G-Protein Coupled Receptors: Debated Structures with Established and Emerging Functions", *Journal of Endocrinology*, Vol. 196, No. 3, pp. 435-453.
- Tan, K. S., A.G. Nackley, K. Satterfield, W. Maixner, L. Diatchenko and P. M. Flood, 2007, "Beta2 Adrenergic Receptor Activation Stimulates Pro-Inflammatory Cytokine Production In Macrophages Via PKA and NF-KappaB-Independent Mechanisms", *Cell Signal*, Vol. 19, pp. 251-260.

- Teller, D., T. Okada, C. Behnke, K. Palczewski and R. Stenkamp, 2001, "Advances in Determination of a High-Resolution Three-Dimensional Structure of Rhodopsin, A Model of G-Protein-Coupled Receptors (GPCRs)", *Biochemistry*, Vol. 40, No. 26, pp. 7761-7772.
- Tirion, M. M., 1996, "Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis", *Physical Review Letters*, Vol. 77, No. 9, pp. 1905–1908.
- Verlet, L., 1967, "Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules", *Physical Review*, Vol. 159, No. 1, pp. 98-103.
- Zhang, X., E. Y. T. Chien, M. J. Chalmers, B. D. Pascal, J. Gatchalian, R. C. Stevens and P. R. Griffin, 2010, "Dynamics of the  $\beta_2$ -Adrenergic G-Protein Coupled Receptor Revealed by Hydrogen-Deuterium Exchange", *Analytical Chemistry*, Vol: 82, No. 4, pp: 1100-1108.