KADİR HAS UNIVERSITY
SCHOOL OF GRADUATE STUDIES
DEPARTMENT OF BIOINFORMATICS AND GENETICS

IDENTIFICATION OF DISTINCT COMMUNICATION NETWORKS IN HUMAN β_2 ADRENERGIC RECEPTOR VIA MOLECULAR DYNAMICS SIMULATION

NURAY SÖĞÜNMEZ ERDOĞAN

PH.D. THESIS

İSTANBUL, AUGUST, 2020

IDENTIFICATION OF DISTINCT COMMUNICATION NETWORKS IN HUMAN β_2 ADRENERGIC RECEPTOR VIA MOLECULAR DYNAMICS SIMULATION

NURAY SÖĞÜNMEZ ERDOĞAN

PH.D. THESIS

Submitted to the School of Graduate Studies of Kadir Has University in partial fulfillment of the requirements for the degree of Ph.D. in the Program of Bioinformatics and Genetics

İSTANBUL, AUGUST, 2020

DECLARATION OF RESEARCH ETHICS /
METHODS OF DISSEMINATION

I, NURAY SÖĞÜNMEZ ERDOĞAN, hereby declare that;

this Ph.D. thesis is my own original work and that due references have been appropriately provided on all supporting literature and resources;

this PhD Thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;

I have followed "Kadir Has University Academic Ethics Principles" prepared in accordance with the "The Council of Higher Education's Ethical Conduct Principles".

In addition, I understand that any false claim in respect of this work will result in disciplinary action in accordance with University regulations.

Furthermore, both printed and electronic copies of my work will be kept in Kadir Has Information Center under the following condition as indicated below

The full content of my thesis will not be accessible for 2 years. If no extension is required by the end of this period, the full content of my thesis will be automatically accessible from everywhere by all means.

NURAY SÖĞÜNMEZ ERDOĞAN

_____

KADİR HAS UNIVERSITY

SCHOOL OF GRADUATE STUDIES

ACCEPTANCE AND APPROVAL

This work entitled **IDENTIFICATION OF DISTINCT COMMUNICATION NETWORKS IN HUMAN β_2 ADRENERGIC RECEPTOR VIA MOLECULAR DYNAMICS SIMULATION** prepared by **NURAY SÖĞÜNMEZ ERDOĞAN** has been judged to be successful at the defense exam held on **16 / 08 / 2020** and accepted by our jury as **Ph.D. THESIS.**

APPROVED BY:

(Prof. E. Demet Akten) (Advisor) (Kadir Has University) _____

(Prof. Kemal Yelekçi) (Kadir Has University) _____

(Prof. Özlem Keskin Özkaya) (Koç University) _____

(Prof. Türkan Haliloğlu) (Boğaziçi University) _____

(Assist. Prof. Bahar Delibaş) (Kadir Has University) _____

I certify that the above signatures belong to the faculty members named above.

_____

Prof. Füsun ALİOĞLU

Dean of School of Graduate Studies

DATE OF APPROVAL: / /

TABLE OF CONTENTS

IDENTIFICATION OF DISTINCT COMMUNICATION NETWORKS IN HUMAN β_2 ADRENERGIC RECEPTOR VIA MOLECULAR DYNAMICS SIMULATION

ABSTRACT

G-protein-coupled receptors (GPCRs) are seven-transmembrane (TM) proteins mediating a wide range of human physiological functions by transducing extracellular events into intracellular responses. The signal relay requires allosteric communication between two distant sites. We focused on the human β_2-adrenergic receptor (β_2AR), for which crystal structures were available in inactive and active states. Here, we investigated communication networks using five distinct trajectories of β_2AR, including both inactive and active conformations. Correlation and entropy-based approaches on Cα fluctuations and backbone/side-chain dihedral angle rotations were used to identify potential allosteric coupling and investigate information transfer due to causality. For conformations with freely moving ICL3, the information flow was directed from intracellular to extracellular parts, whereas those with constrained ICL3, this direction was reversed. Residues located on flexible domains generally had polar characteristics and contributed heavily to communication. Two independent runs conducted for the active state, which were complexed with partner G protein (*Gp*) displayed different dynamics.. Also, both the degree of communication and dominancy of polar residues in information transfer were decreased in active states, which was attributed to the presence of *Gp*. In addition, the mutual information (MI) based on Cα fluctuations was different from that based on dihedrals. Cα-based MI was dominant in regions with high RMSF values, whereas dihedral-based MI was dominant in loop regions where dihedral flexibility was high. Moreover, Cα-based transfer entropy (TE) was distinct in each state, whereas dihedral-based TE was high between loop and TM regions in all five states. These results led the way to a new approach in which the protein allostery was described with a combination of Cα- and dihedral-based MI and TE. This approach was suggested for applications including rational drug design and protein engineering.
Keywords: G-protein Coupled Receptors, β_2-adrenergic Receptor, Protein Allostery, Information Transfer, Causality

MOLEKÜLER DİNAMİK SİMÜLASYONU İLE İNSAN β2 ADRENERJİK RESEPTÖRÜNDEKİ BELİRGİN İLETİŞİM AĞLARININ TANIMLANMASI

ÖZET

G-protein-bağlı reseptörler (GPCR), hücre dışı ligand bağlanma işlemini hücre içi tepkilere dönüştürerek çok çeşitli insan fizyolojik fonksiyonlarına aracılık eden ve yedi transmembran (TM) yapısından oluşan proteinlerdir. Karşılıklı sinyal aktarımı, ancak iki uzak bölge arasında allosterik iletişimle oluşur. Hem inaktif hem de aktif kristal yapıların mevcut olduğu bir arketipik GPCR olan insan β_2-adrenerjik reseptörüne (β_2AR) odaklandık. β_2AR 'ın farklı konformasyonlarının yörüngelerini oluşturmak için moleküler dinamik (MD) simülasyonları gerçekleştirildi. Burada, β_2AR'a ait orijinal inaktif durum (Faz I), çok inaktif durum (Faz II), ara durum ve G-proteine bağlı aktif durumun yörüngelerini kullanarak potansiyel iletişim ağlarını araştırdık. Bu nedenle, bu tezde nedenselliğe bağlı potansiyel allosterik etkileşim ve bilgi aktarımını ortaya çıkarmak için proteindeki Cα dalgalanmaları ve omurga/yan zincir dihedral açı rotasyonları üzerinde hem korelasyon hem de entropi bazlı olasılıksal yaklaşımlar kullanılmıştır. Serbest ICL3 içeren yapılarda bilgi akışı yönü hücre içinden hücre dışına doğru iken, ICL3'ün hareketinin kısıtlandığı yapılarda bu akış yönünün tam tersine olduğu görüldü. Ayrıca, esnek alanlarda lokalize olmuş amino asitler genellikle polar özelliklere sahip olup iletişime büyük katkıda bulunmuşlardır. *aktif-Gp* için bağımsız iki çalışma, ICL3'ün z-yönünde hareketi ile birbirinden ayrılmış ve burada G proteini etkisinden dolayı iletişim ve polarite gücünün azaldığı saptanmıştır. Son olarak, mutlak Cα dalgalanma değerleri ve sterik engeller farklı dihedral açılarının oluşmasına neden olabildiğinden, Cα dalgalanmaları ve dihedral açıların ortak hareketi gözlemlenmemiştir. Bu nedenle de dihedral verilerde çoğunlukla ilmik alanlarının ortaya çıktığı görüntülenmiştir. Bu sonuçlar β_2AR 'daki allosterik iletişimi açıklamak için yapı temelli bir mekanizma sağlamakta ve dahası rasyonel ilaç tasarımı ve protein mühendisliği gibi uygulamalar için bir temel oluşturmaktadır.

Anahtar Sözcükler: G-protein Bağlı Reseptörler, β2-adrenerjik Reseptör, Protein Allosterisi, Bilgi Transferi, Nedensellik

ACKNOWLEDGEMENTS

I have been fortunate to receive exceptional mentorship during my time at the Department of Bioinformatics and Genetics at Kadir Has University. First and foremost, I must thank Prof. Demet Akten for her guidance, patience as well as her boundless enthusiasm and endless encouragement. Thanks also to the rest of my committee:

- Prof. Kemal Yelekçi, for his careful and thoughtful assessments over my whole graduate school tenure.
- Prof. Özlem Keskin Özkaya, for her support.
-Prof. Türkan Haliloğlu, for helping me to keep the "big picture" in mind.
-Assist. Prof. Bahar Delibaş, for her encouragement and perspective.

I am grateful to have worked with many great people at the Kadir Has University and to both current and former lab members for providing a truly enjoyable work environment, for many productive discussions, and for plenty of troubleshooting help. Thanks to Melis Gencel, Sercan Beytur, and Burak Servili. One could not ask for better friends. You truly made this journey fun and exciting. You will be dearly missed. Thanks to Funda for always being with me in this journey and listen to me whenever I was frusturated. Also, many thaks to my dearest co-workers; Ayhan, Büşra, and Merve, who were always there for me whenever I wanted to talk with someone.

I also want to thank to my brother, Turgay Söğünmez, to my father Ahmet Ali Söğünmez, and especially to my mother Nurten Söğünmez. I am very lucky to have you all in my life. I would like to express my gratetude to my parents for imparting the values of dedication and hard work in me, and always encouraging me during my doctoral training and research. I want to thank to my husband Alperen Erdoğan for his endless encouragement, love, motivation and support.

Finally, it was a great opportunity for me that my Ph.D. research has been supported by TÜBİTAK BİDEB 2211/C – Yurtiçi Öncelikli Alanlar Doktora Burs Programı.

DEDICATION

To my baby boy Uygar Erdoğan...

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AMBER	Assisted Model Building with Energy Refinement
APL	Area Per Lipid
β_2AR	β_2 Adrenergic Receptor
CHARMM	Chemistry at HARward Macromolecular Mechanics
Cmpd-15PA	Polyethylene glycol-carboxylic acid derivative
Cmpd-6FA	(R)-N-(4-amino-1-(4-(tert-butyl)phenyl)-4-oxobutan-2-yl)-5-(N-isopropyl-N-methylsulfamoyl)-2-((4-methoxyphenyl)-thio)benzamide
COM	Center Of Mass
CPU	Central Processing Unit
CTC	Conditional Time-Delayed Cross-correlation
ECL	Extracellular Loop
GPCR	G Protein-Coupled Receptor
GROMOS	GROningen MOlecular Simulation
H	Helix
ICL	Intracellular Loop
LJ	Lennard-Jones
MD	Molecular Dynamics
MI	Mutual Information
MM	Molecular Mechanics
NAMD	Nanoscale Molecular Dynamics
NMR	Nuclear Magnetic Resonance
NPT	Constant temperature and constant temperature
NVT	Constant volume and constant temperature
OPLS	Optimized Potentials for Liquid Simulations
PDB	Protein Data Bank
PME	Particle Mesh Ewald
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
Rc	Cutoff distance
RMSD	Root-Mean-Squared Deviation
RMSF	Root-Mean-Squared Fluctuation
TE	Transfer Entropy
TM	TransMembrane
vdW	van der Waals
VMD	Visual Molecular Dynamics

1. INTRODUCTION

In structure-based computational studies, a comprehensive analysis of the protein's intrinsic dynamics is fundamental. Although experimental work is the gold standard for structural analysis, no experimental techniques have yet been developed to demonstrate the full dynamic repertoire of a single protein. Currently, we do not have any visualization devices to display the time-dependent behavior of a protein. Bioinformatics and computational methods are evolving because experimental studies are costly and time-consuming. In order to understand the structure and function of a protein, we are now able to produce results rapidly with computational methods. By validating with experimental results, the procedures that took several years before are now completed within months or even days.

In this thesis, linear and probabilistic methods, such as mutual information and transfer entropy, were used to determine the causality between residue displacements. Eventually, the allosteric communication networks in inactive and active states of the human beta 2 adrenergic receptor (β_2AR) were identified. Based on this data, potential bindings sites were proposed. These proposed sites could be used in designing allosteric drugs. Several studies targeting β_2AR dynamics (Dror, Arlow, Borhani, Jensen, and Shaw, 2009; Dror *et al.*, 2011; Nygaard *et al.*, 2013) explored the receptor's activation mechanisms and its novel conformational states. However, β_2AR used in these studies generally lacked the longest intracellular loop, ICL3, which was shown to play a significant role in the receptor motion (Ozcan, Uyar, Doruker, and Akten, 2013). Here in this thesis, we used four distinct states of the receptor, including an original inactive state, a novel inactive state, an intermediate state and a G-protein bound active state, where ICL3 was incorporated. Hopefully, the results of this comprehensive data presented here will have an important contribution to the literature in the context of

GPCR dynamics, details of allosteric motions, and determination of potential allosteric target sites for drug molecules.

As discussed in Chapter 4, the linear and time-dependent correlation analysis based on backbone Cα fluctuations revealed driver-follower relationships between residue-pairs in two inactive states of β_2AR, which have distinct 3rd intracellular loop (ICL3) conformations and stiffness characteristics. Based on these calculations, critical protein parts became drivers and followers in specific delay times. A pathway emanating from 8th transmembrane helix (H8) → ICL3 → H5 → H6 in one inactive state was deduced from conditional time-delayed correlation (CTC) analysis whereas, the novel inactive state demonstrated an entirely different profile than the original inactive, with the second extracellular loop (ECL2) becoming a follower for H5 and H6.

The direction of information transfer in inactive and active conformers was analyzed based on backbone Cα fluctuations and the amino acids playing roles in this transfer were identified in Chapter 5. Locations of allosteric communications and information transfer in inactive states displayed a shift from intracellular to extracellular parts upon ICL3 closure in inactive conformations. Interestingly, the dominant role of residue polarity in information sharing was discovered, even though the calculations were only performed on Cα fluctuations.

In molecular dynamics (MD) simulation of the active receptor without G protein, the intracellular part of H6 ($H6_{Int}$) rapidly made an inward tilt and induced ICL3 to fluctuate through the core of the receptor. $H6_{Int}$, however, did not form complete bending as in inactive conformation. Moreover, the distance between orthosteric binding sites and the deviation of $H7_{Int}$, including the *NPxxY* domain, confirmed this state's resemblance to inactive state; however, it was not completely inactivated instead spanned intermediate conformations. Therefore, we called this state as *intermediate* throughout the thesis.

The *intermediate* state exhibited some interesting results. Surprisingly, both mutual information and transfer entropy values in the *intermediate* state were the highest

among all, indicating the highest allosteric communications present in this conformational state. Moreover, the allosteric communication and entropy transfer results of the *intermediate* state were similar to Phase I considering the loop regions, while similar to Phase II, considering the transmembranes. Both intracellular and extracellular parts of the receptor displayed entropy donor characteristics, supporting the intermediate characteristics of the trajectory. Similar to Phase I, ICL3 was one of the key domains for allosteric communications in the *intermediate* state. Surprisingly, ICL2 and $H5_{Ext}$ were discovered as other key domains, where ICL2 holds the positive allosteric modulator, *Cmpd-15FA*, and $H5_{Ext}$ is critical for orthosteric binding. Dominant roles of polar residues on information sharing were also observed in the analysis of the *intermediate* state. Finally, information transfer from intracellular parts and orthosteric ligand binding site towards the receptor's extracellular parts was disclosed.

Additionally, two independent simulations of active receptor were performed together with the trimeric G protein. These two runs with different initial states and velocities were used in both Cα- and dihedral based analysis to capture different dynamics of the active state. These two systems were denoted as *active Gp* and *active* Gp_{2nd}. In active phases, α-helical structure of G protein restrains the fluctuation of ICL3 in the direction perpendicular to the membrane surface. Interestingly, in one of the simulations, ICL3 displayed fluctuations in the direction of the membrane away from G-protein. In addition, the distance between the extracellular helices, H5 and H3, was smaller compared to Phase I and Phase II, and closer to the active state, supporting the hypothesis that GPCR/G protein complex preassembles at the very beginning of the activation process. The difference of this distance between *active-Gp* and Phase II was around 3 Å - 4 Å. Moreover, the extracellular H3 - H5 distance was also compared with respect to the distance in Phase I, where the extracellular part of the receptor was relatively narrow compared to Phase II. Interestingly, the extracellular H3 - H5 distance in *active-Gp* and *active-*Gp_{2nd} was also small, around 1 Å - 2 Å, compared to Phase I, indicating the effect of $H6_{Int}$ orientation along with ICL3 conformation, which opened up the intracellular part of the receptor but narrowed down the extracellular part. Still, the extracellular part was gradually widened, explaining the lack of agonist's stabilizing

effect on the active states. Thus, the presence of G protein in these two states preserved the active-like intracellular conformation more than the *intermediate* state; however, the agonist-free nature of the systems could not prevent the extracellular part from expansion. Thus, these two states were similar to active state intracellularly; however, they spanned intermediate conformations extracellularly.

Unexpectedly, high mutual information values were observed in H1, H2, $H3_{Ext}$, ECL2, and $H7_{Ext}$. ICL3 was also revealed in information sharing with smaller intensities. Interestingly, distal polar residues were not dominant in information sharing, which might be due to nearby G protein. The fluctuation of ICL3 in membrane direction in *active-*Gp_{2nd} notably reversed the characteristics of donors and acceptors compared to *active-Gp,* especially in residues locating on $H3_{Ext}$, ICL2, $H4_{Ext}$, $H7_{Ext}$, and H8. Except for ICL3, the entropy donors were observed to locate more at the extracellular parts, especially on $H2_{Ext}$, $H3_{Ext,}$ and $H4_{Ext}$ in *active-Gp* and on $H6_{Ext,}$ and $H7_{Ext}$ in *active-*Gp_{2nd} displaying unique information transfer characteristics compared to remaining phases.

The intracellular parts of the receptor became quite important after resolving the crystal structures of allosteric modulators. Positive allosteric modulator (PAM) (Xiangyu Liu *et al*., 2019) and negative allosteric modulator (NAM) (Xiangyu Liu *et al*., 2017) binding sites on β_2AR were studied in detail to understand the corresponding residues' entropic contributions to allosteric interactions. Since two modulators were observed to bind intracellularly, one can claim that receptor's intracellular part plays a significant role in receptor's overall allosteric coupling mechanism. In Phase I, the intracellular residues, which the interaction of NAM was displayed, showed entropy source features compared to Phase II where the binding site was occupied by the closure of ICL3. Similarly, NAM binding sites represented entropy source characteristics in the *intermediate* state; however, their net TE values were substantially decreased with the association of the trimeric G protein and the receptor in *active-Gp* and *active-*Gp_{2nd}. This represented the effect of free ICL3 on the PAM binding, where the association of G protein limited the loop mobility and occupy modulator binding site, thus silenced the residues' information source features.

Later on, the contribution of the backbone and side-chain dihedrals to the information sharing and transfer were examined using dihedral angle variations derived from the MD trajectories, as explained in Chapter 6. An important, but often neglected, contribution to the protein allostery is the loss/gain of entropy that results from accessible side-chain conformers. In addition to the displacement of Cα atoms, the angular changes in the backbone and side-chain rotatable bonds throughout the trajectory were considered. Several studies analyzed mutual information and conformational entropy of backbone and side-chain dihedrals (DuBay, Bothma, and Geissler, 2011; Lenaerts *et al*., 2008; Long and Tian, 2016; Long, Wang, and Tian, 2019; McClendon, Friedland, Mobley, Amirkhani, and Jacobson, 2009). These studies explained the long-range communications' transmission either through side-chains, by fixing the backbone motions, or explained the inter-protein communications only through backbone dihedrals. When backbone and side-chain dihedrals were both included in the mutual information analysis, the fixed binning approach was applied to describe the angular space in order to simplify the calculations (Long *et al*., 2019). However, both backbone and side-chain rotameric data were included in our calculations. Moreover, the number of bins was selected according to the angular frequencies of each rotameric state throughout the trajectory and to the data given in the rotamer library (van der Kamp *et al*., 2010).

Up to our knowledge, transfer entropy investigation on dihedral angles has not been studied in any of the protein systems, yet. Therefore, this will be the first study of transfer entropy based on the backbone and side-chain dihedral angles. In order to increase computational efficiency, we computed transfer entropy of backbone and side-chain dihedrals separately. Besides, the mutual information of backbone and side-chain dihedrals has not been studied on β_2AR, yet. Therefore, our results will have the utmost importance of considering residue-pair communications in β_2AR through dihedral data.

Accordingly, significant mutual information (MI) and transfer entropy (TE) were mostly observed between loop regions. On the other hand, transmembrane helices had either low or zero MI and TE data. The information profile of loops mostly overlapped with that of the information based on Cα atom displacements; however, the profiles on

transmembranes were mostly different. Thus not much direct relationship was found in Cα fluctuations and dihedral angles in terms of MI and TE. It was also supported the analysis, where the correlation between $\Delta\phi$, $\Delta\psi$, $\langle\Delta r^2\rangle$, relative solvent accessibility (RSA), and β-factor both in NMR and X-ray structures were calculated. The correlation of $\Delta\phi$, $\Delta\psi$ with $\langle\Delta r^2\rangle$ was 0.5 in NMR structures, while the correlation of $\Delta\phi$ with β-factor in X-ray structures was 0.29 and $\Delta\psi$ with β-factor was 0.27, where β -factor is known to be related with the Cα fluctuations. Additionally, the correlation between $\Delta\phi$ and $\Delta\psi$ was observed as 0.53 (T. Zhang, Faraggi, and Zhou, 2010)

Side-chain dihedral calculations revealed that residues with a higher number of rotatable bonds transferred more information compared to those with a low number or zero rotatable bonds. Inclusion of side-chain dihedrals to the calculations dramatically increased the MI values compared to backbone-only or Cα–only calculations. Additionally, a prominent increase in the effect of polar residues on mutual information was discovered. The MI values of polar residues were almost three times as high as the values of hydrophobic residues. Moreover, in terms of transfer entropy based on side-chain dihedrals, residues with at least three torsional angles notably became entropy donors. Otherwise, they were observed as acceptors, independent of being polar or hydrophobic. Besides the number of rotatable bonds, loop localization and polarity also had an effect on net TE values of side-chain dihedrals. The net TE value of polar residues with more than three rotatable bonds in a loop region was significantly higher than polar residues with more than three rotatable bonds in a transmembrane.

The high similarity between independent runs of active states was observed in terms of both mutual information and transfer entropy based on dihedral angle analysis; however, these values differed with the inactive phases and *intermediate* state. However, the same analysis based on Cα fluctuations yielded some differences in all five states. This might indicate that dihedral angle data could discriminate against each conformational state; however, it was not able to distinguish two independent runs of the same active state. However, since local changes can also be captured by Cα fluctuations along with global changes, two independent runs of the same active state yielded different results; thus, both minor and major conformational changes were

detected with Cα calculation data. Overall, the combinatorial effect of both Cα fluctuations and dihedral angles was considered in order to understand the full intrinsic protein dynamics and allosteric communications of five distinct conformational states of human β_2AR.

2. LITERATURE REVIEW OF PROTEIN ALLOSTERY AND G-PROTEIN COUPLED RECEPTORS (GPCRs)

2.1 Allostery In Proteins

The term allostery in proteins is defined as the relation of conformational changes among various distant protein domains (Gunasekaran, Ma, and Nussinov, 2004). The first allosteric models, the "Concerted (MWC)" and "Sequential (KNF)" models were developed in1960s (Koshland, Nemethy, and Filmer, 1966; Monod, Wyman, and Changeux, 1965). The term "dynamic allosters" was introduced in 1984. Cooper and his colleagues changed the concept of classical entropy calculations from *conformational change-related allostery* to *the effect of conformational-equilibrium states on the function* (Cooper and Dryden, 1984).

Members of protein families were compared based on their amino acid sequences in orthosteric and allosteric binding regions, and it was reported that orthosteric regions were evolutionarily more conserved than allosteric binding regions (Lockless and Ranganathan, 1999). Molecules binding to these allosteric binding regions include positive allosteric modulators (PAMs), negative allosteric modulators (NAMs), and silent/neutral modulators for protein subfamilies (See Figure 2.1). Since drugs developed against these conserved orthosteric regions can bind to different members of the same protein family, their selectivities could be low, and their side effects could be high. Nowadays, newly developed drugs targeting these allosteric regions are selectively binding to the target protein of interest, therefore significantly decreasing side effects (Flor and Acher, 2012; Nussinov and Tsai, 2012; Wenthur, Gentry, Mathews, and Lindsley, 2014).

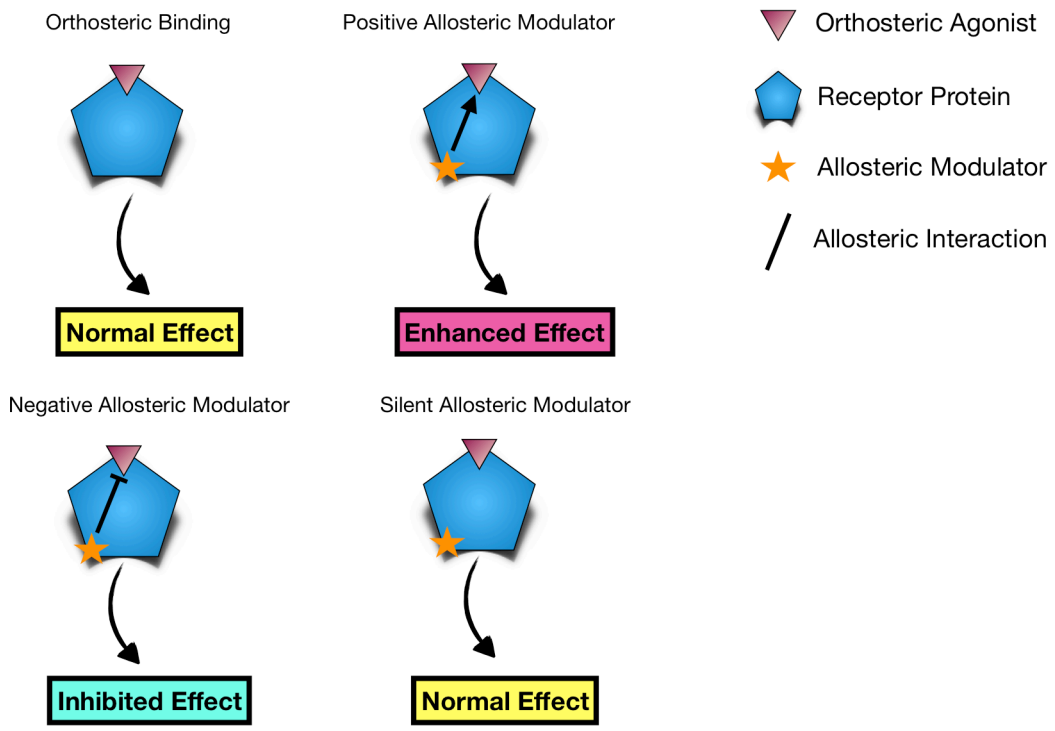


Figure 2.1 Representation of allostery. Types and effects of allosteric modulators on orthosteric ligand's function.

The allosteric motion in proteins is the direct result of the correlated motions between residue-pairs (Goodey and Benkovic, 2008; Kern and Zuiderweg, 2003; Koshland *et al.,* 1966; A. J. Wand, 2001). The information transfer due to amino acid fluctuation is explained by directionality and causality (Guarnera and Berezovsky, 2016; Kamberaj and van der Vaart, 2009). Residue-based correlations and allosteric motions of G protein-coupled receptors (GPCRs) and β_2AR have been studied in several studies (Balabin, Yang, and Beratan, 2009; Bruzzese, Gil, Dalton, and Giraldo, 2018; Manglik *et al.,* 2015) but still not resolved entirely. On the other hand, residue-based correlations are symmetric in nature; thus, they cannot tell us about the directionality of the information transfer.

All intra- and extracellular processes associated with protein-protein, protein-DNA, protein-drug interaction, and signal transduction require allosteric activity to preserve normal physiology. It is known that most cancer-causing mutations disrupt normal

allosteric interactions. In nature, the allosteric interactions arise from entropic and correlated atomic fluctuations (Motlagh, Wrabl, Li, and Hilser, 2014; Nussinov, Jang, and Tsai, 2014) Since signal transmission from correlated fluctuations is the universal characteristic of all proteins, the allostery can be regarded as an intrinsic property of each protein (Gunasekaran, Ma, and Nussinov, 2004). The terminology of entropy transfer used by Schreiber (Schreiber, 2000) is one of the widely used definitions of information flow in the literature. Using Schreiber's work, many studies analyzed the entropy flow in Ets-1 and its associated DNA (Kamberaj and van der Vaart, 2009), via the entropy flow in the Molecular Dynamics (MD) simulation of mutation analysis affecting autophosphorylation of ERK2 (Barr *et al.,* 2011), the entropy flow in the antibody-antigen relationship complexes (Corrada, Morra, and Colombo, 2013) and the entropy flow in Rho GTPase binding sites in the dimerization phase (L. Zhang, Centa, and Buck, 2014) have been conducted.

2.2 G Protein-Coupled Receptors (GPCRs)

Proteins can be categorized into four groups as structural proteins, enzymes, receptors, and other proteins. Among these four groups, receptors are the most widely studied protein group for many years. Receptors are found in the cell or nucleus membrane and respond to a variety of stimuli or molecules such as neurotransmitters, hormones, antigens, and ions. Amongst receptors, one of the most widely known protein groups and the greatest superfamily of all proteins are GPCRs. The GPCR superfamily is categorized in 5 subfamilies as Adhesion, Secretin, Frizzled/Tas2, Glutamate, and Rhodopsin as illustrated in Figure 2.2, which play a broad role in smell and vision, behavioral regulation, immune system and inflammation activity, autonomic nervous system stimulation, and apoptosis (Deupi and Kobilka, 2007). These wide-ranging functionalities attracted the pharmaceutical industry, as well. Today, almost half of the pharmaceutical market consists of drugs developed for the treatment of GPCR-based diseases (Rosenbaum *et al.,* 2009).

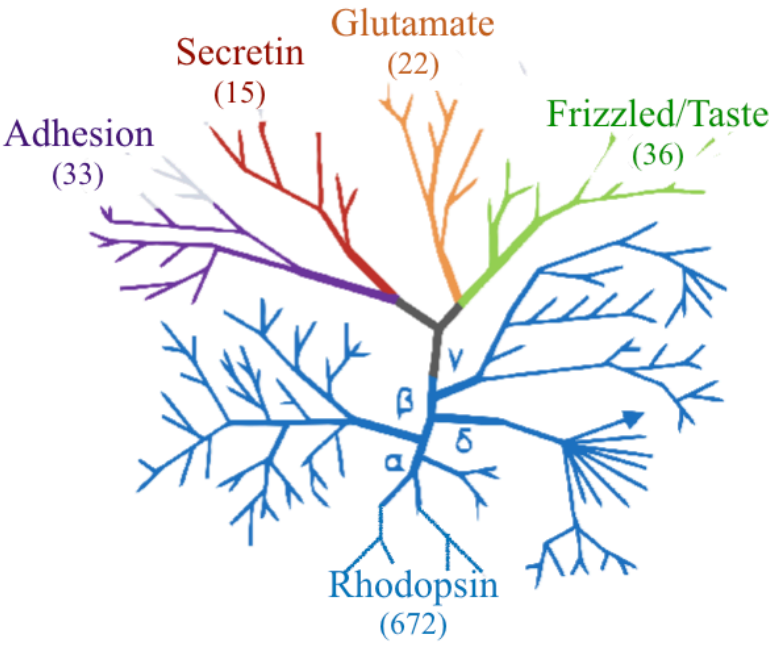
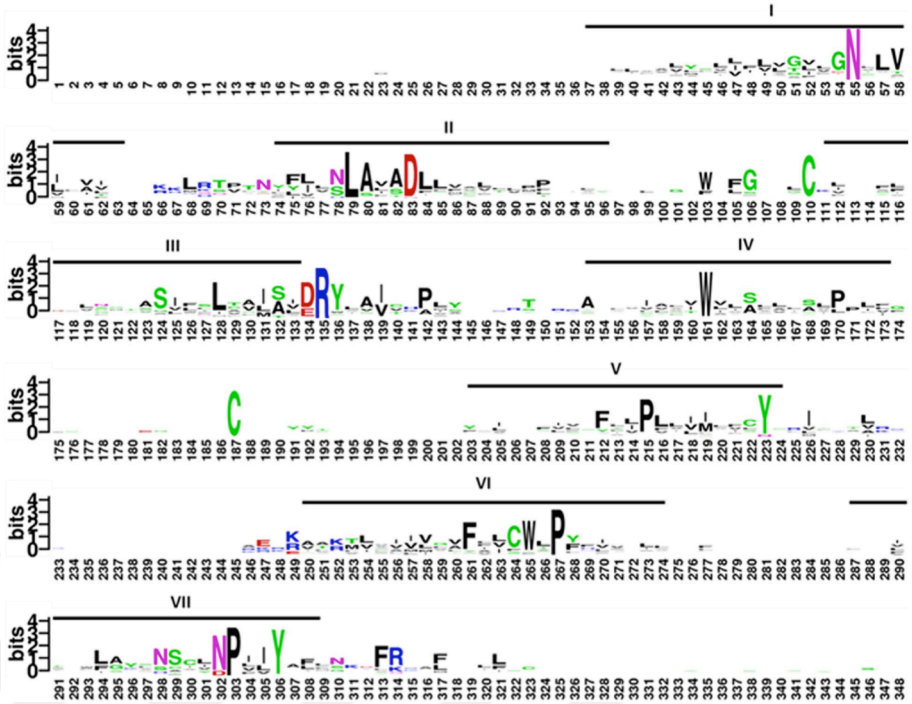
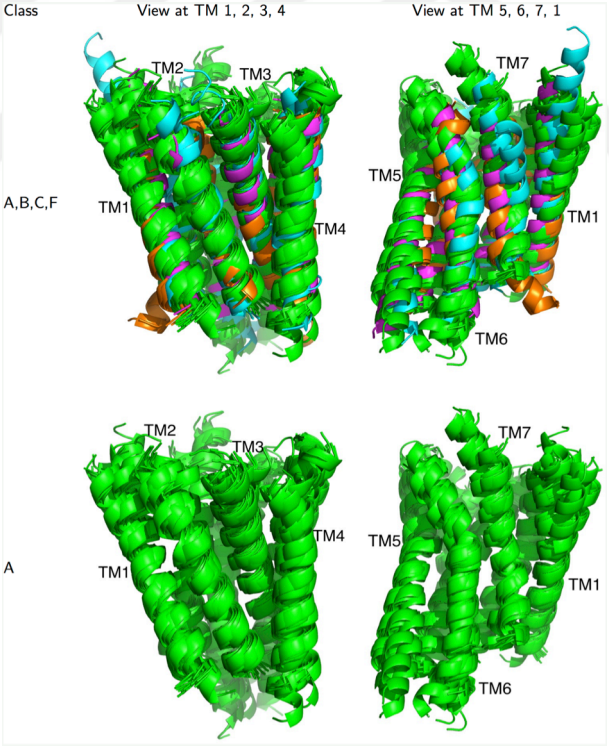


Figure 2.2 GPCR Superfamily. Five subgroups of GPCRs are colored separately. Known protein numbers in 5 subgroups are denoted in parentheses. Subbranches of the Rhodopsin subgroup were shown separately.

Drugs targeting GPCRs have to overcome selectivity and specificity issues. The multiple sequence analysis (MSA) based on both Statistical Coupling Analysis (SCA) (Süel, Lockless, Wall, and Ranganathan, 2003) and Generative REgularized ModeLs of proteINs (GREMLIN) (Moitra, Tirupula, Klein-Seetharaman, and Langmead, 2012) represent significant variations in amino acid sequences in GPCR members. These differences majorly observed in non-transmembrane substructures. Even in transmembranes of class A (Rhodopsin) family of GPCRs, MSA analysis revealed few conserved amino acids, while the rest of the residues are either relatively conserved or not conserved at all (See Figure 2.3a). However, structure-based sequence comparisons of GPCRs display a shared seven transmembrane helix (7TM) fold motif, even for receptors with a very low sequence similarity. For class A (Rhodopsin) GPCRs, the structural similarity is highly remarkable due to conserved inter-helical hydrogen bonds maintaining transmembrane packing. The similarity between transmembranes of experimentally determined GPCR structures was reported as 35%-100%; on the other hand, it was around 20%-30% when their sequence similarities were considered (See Figure 2.3b).



(a)



(b)

Figure 2.3 Multiple sequence alignment data of class A GPCRs (a). The conserved residues were given in capital letters (Süel *et al.,* 2003). Structure-based alignment of classes of A, B, C, and F (up) and only the members of class A (bottom) transmembranes (Moitra *et al.,* 2012) (b).

2.3 Beta-2 Adrenergic Receptor (β_2AR) Structure and Function

The GPCR superfamily is grouped into five subfamilies, and our target receptor, β_2AR from the rhodopsin subfamily, is a well-known member of GPCRs. β_2AR is mainly found in lung and cardiac tissues and is associated with diseases such as asthma, heart failure, and obesity (Koch, Lefkowitz, and Rockman, 2000; Kohm and Sanders, 2001; Xuan Liu *et al.,* 2008; Shang, Liu, and Liang, 2009). With the discovery of the high-resolution X-ray crystal structure of the inactive form of human β_2AR in 2007, which led to a Nobel Prize in 2012 (Cherezov *et al.,* 2007), a drastic increase in the number of studies focused on this receptor was observed. Later on, several inactive and active crystal structures of the same receptor were resolved (Hanson *et al.,* 2008; Rasmussen *et al.,* 2011; Rosenbaum *et al.,* 2011) and studies of the structure-function relationship of these receptors have been accelerated (Kobilka, 2013; Manglik and Kobilka, 2014).

High-resolution crystal structures of β_2AR provide invaluable information on the ligand and G-protein binding, GPCR structure, and protein dynamics. These structures demonstrated the hydrogen bonding of Carazolol, with *Asp*113, *Asn*312, and *Tyr*316 and hydrophobic interactions with *Val*114, *Phe*290, and *Phe*193 in the receptor (Cherezov *et al.,* 2007; Rasmussen *et al.,* 2007; Rosenbaum *et al.,* 2007) (Figure 2.4). Furthermore, BI-167107 binding results in a 2 Å inward movement at *Ser*207 in H5. This inward movement, in turn, maintains a hydrophobic interaction among the residues *Pro*211, *Ile*121, and *Phe*282. Also, it was reported in a study that, an endogenously bound agonist induces a shift in the position of *Asn*293 on $H6_{Ext}$ in a high-affinity nanobody bound active β_2AR crystal structure, as colored with green in the active state in Figure 2.4 (Ring *et al.,* 2013). Carazolol was colored with yellow, BI167107 was colored with red, inactive reference colored with blue, and the active state was colored with green.

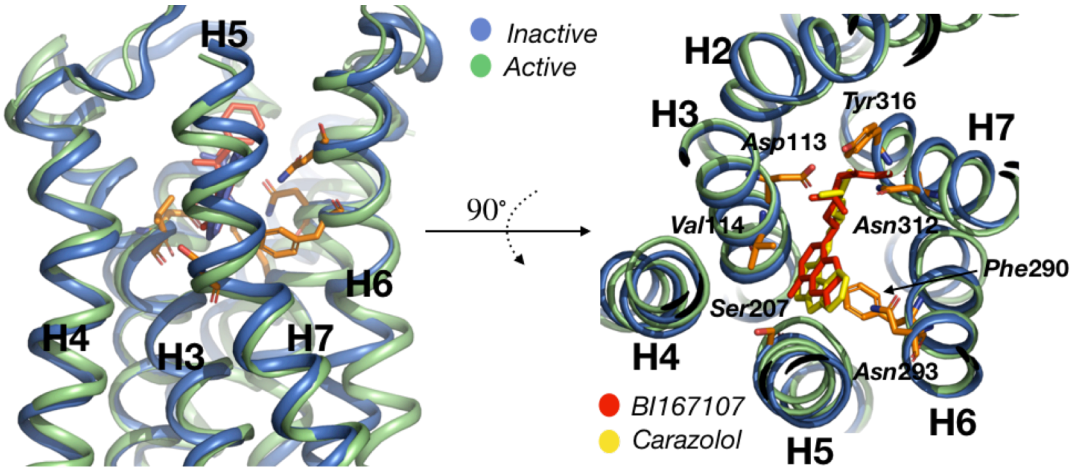


Figure 2. 4 Lateral (left) and extracellular wiew (right) of othosteric binding site of β_2AR together with agonist BI167107 (red) and inverse agonist Carazolol (yellow). Orthosteric binding sites were given as orange. Carazolol bound inactive state was colored in blue, and BI167107 bound active state was colored in green.

Additionally, one of the most important characteristic features of the receptors in the Rhodopsin family is the ionic lock between D[E]RY motif in the third helix (H3) region and the *Asp/Glu*268 in the H6 region (Figure 2.5b). It is also known that β_2AR has a ligand-independent basal activity (Samama, Bond, Rockman, Milano, and Lefkowitz, 1997). In the crystal structure of inactive β_2AR (PDB ID: 2RH1), this ionic lock is broken, and it is argued that this broken ionic lock may be the reason for the receptor's basal activity. In addition, bioinformatic analysis of βARs showed that the amino acid sequences of extracellular loops were different, but their structures were similar. Particularly, it was found that the 2nd extracellular loop (ECL2) was found to have common α-helix structures in β_1AR and β_2AR, which, in turn, did not exist in the rhodopsin. In β_2AR, the stabilization of ECL2 was maintained both by the presence of this helical structure and by two disulphate bonds between residues *Cys*184–*Cys*190 and *Cys*191–*Cys*106. It is suggested that the presence of helical structure and structural stability may play a role in reversible ligand binding and conformational changes, and differences in amino acid sequences of ECL2 in ARs may affect ligand specificity and selective binding.

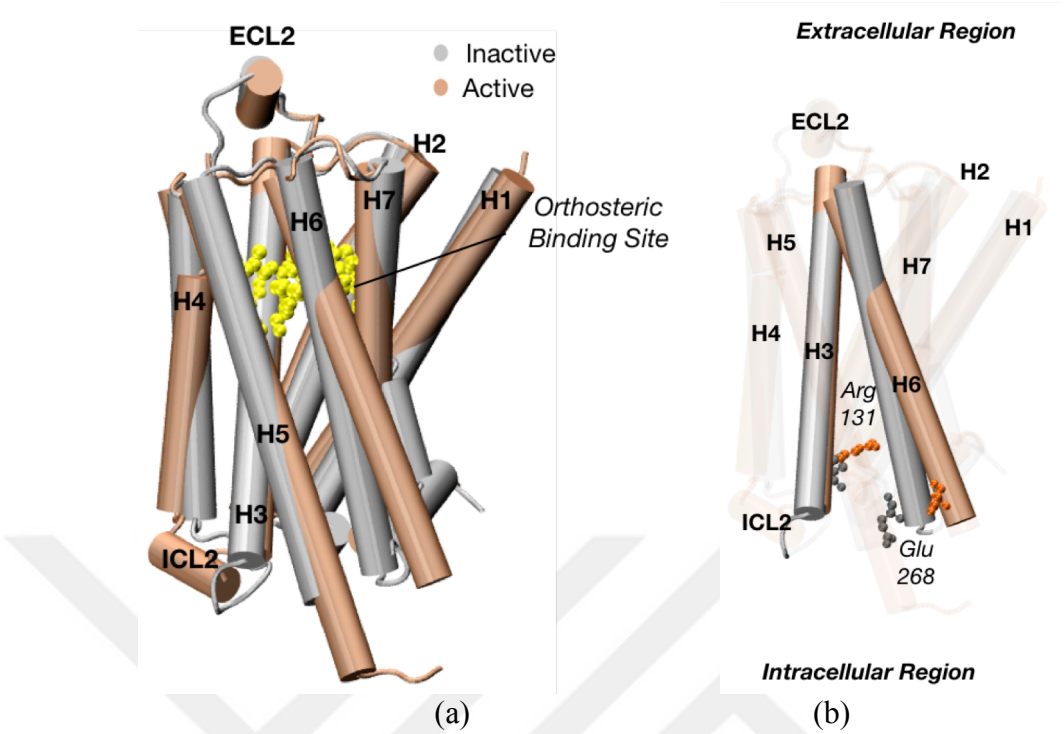
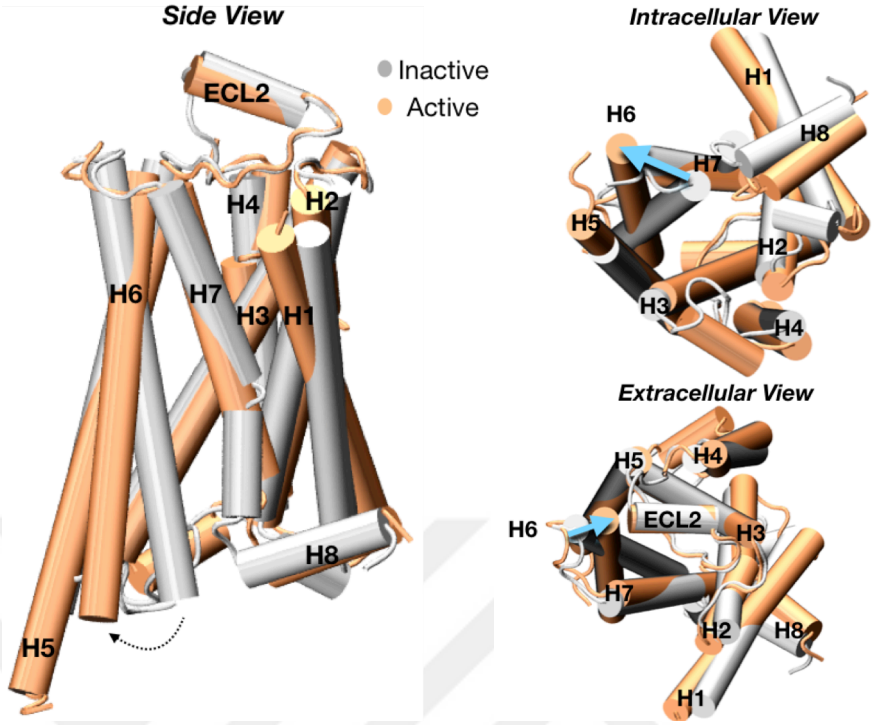


Figure 2.5 Orthosteric binding site for β_2AR (yellow) (a). Ionic lock and its conformations in inactive (PDB ID: 2RH1) and active (PDB ID: 3SN6) receptor states (b).

Besides, the structure of active and inactive β_2AR differs in terms of specific conformational rearrangements. Around 14 Å outward movement of intracellular H6 and the conformational change in ICL3, an inward movement of NPxxY domain of H7 and elongation and rotation of intracellular H5 induces the active conformation of β_2AR (Rasmussen *et al.,* 2011; Rasmussen *et al.,* 2011; Rosenbaum *et al.,* 2009) (Figure 2.6). To increase both the proteolytic stability and increase the yield of protein crystallizability, elimination and replacement of the ICL3 with the protein T4 lysozyme in β_2AR structure needed to be performed during the crystallization step. However, ICL3, which was atomistically modeled by the MODELLER program, has inevitable roles in protein dynamics (Šali, 1993; Eswar *et al.,* 2006).



(a) (b)

Figure 2.6 The active and inactive conformations from the resolved crystal structures of active (orange) and inactive (silver) β_2AR. Three major differences through activation were shown as the outward motion of $H6_{Int}$, *NPxxY* conformation in H7 and elongation of $H5_{Int}$.

The complete blockage of the intracellular G protein-binding site and the deviation of $H6_{Int}$ were found to be the major results of large ICL3 fluctuations on the receptor activation (Ozcan *et al*., 2013). As a result of the 1.5 µs long molecular dynamics (MD) simulation, three different conformational states of the β_2AR were discovered, and one of them was indicated as a "very inactive state" of β_2AR, which was not previously described in the literature. In this very inactive state, ICL3 was completely closed and stable for the rest of the simulation.

During activation of the receptor through ligand binding, experimental studies revealed the optimum distances of extracellular parts of H5 (*Ser*203, *Ser*204, and *Ser*207) and H3 (*Asp*113) for ligand binding as 8 Å - 10 Å, which could only be possible with an outward movement of ICL3/intracellular part of the H6 in the receptor due to its zipper-like motion. Ozgur and her co-workers further analyzed the allosteric coupling between extra- and intracellular domains of β_2AR and set up seven individual restrained MD

simulations on the distances between *Asp*113 in H3 and the residues *Ser*203, *Ser*204, and *Ser*207 in H5, *Phe*289, *Phe*290 in H6, and *Asn*312 in H7. Additionally, one more restrained distance was set up between *Phe*289 in H6 and *Asn*312 in H7. The restrained distances of 16 Å-17 Å between ligand binding sites *Asp*113 and *Ser*203 induced a rapid closure of ICL3, while 8 Å – 10 Å – 8 Å restraints between three Serines on H3 preserved the open conformation of ICL3 (Ozgur, Doruker, and Akten, 2016). Additionally, a study conducted by Dror *et al.* showed the fluctuation of intracellular H6 in open and intermediate forms in the basal structure and fully opening of intracellular H6 just after intracellular G-protein binding, which forced the extracellular binding site to be available for its ligand. If the ligand was an agonist, then the receptor stabilized its active state and maintained the intracellular signal relay (Dror *et al.,* 2011). Narrowing of the intracellular region and expansion of the extracellular part of β_2AR and vice versa indicates that there may be an allosteric communication and information flow between these two regions.

The increasingly growing interests of identification of allosteric regulators of β_2AR yielded two compounds as *Cmpd-15PA* and *Cmpd-6FA*, which play roles as negative and positive allosteric modulators, respectively. Both *Cmpd-15PA* and *Cmpd-6FA* bind to the receptor intracellularly, where the former one lies partially buried in a pocket formed by transmembrane helices H1, H2, H6, H7, H8 and ICL1, and the latter one in the transmembrane helices H2, H3, H4, and ICL2.

When the structural and functional analysis of these two compounds was conducted, it was observed that *Cmpd-15PA* sterically prevented G_s coupling and eventually blocked GPCR kinase and arrestin interactions. The negative effect of *Cmpd-15PA* on agonist binding affinity is probably coming from the stabilization of the inactive conformation of TM6. *Cmpd-15PA*, like Nb60, was shown to stabilize the inactive conformation of TM6, which is associated with reduced affinity for agonists (Ahn *et al.,* 2017). With the recent structures of CCR2 (Zheng *et al.,* 2016) and CCR9 (Oswald *et al.,* 2016), bound to intracellular modulators, these results conveyed the idea that the location of this intracellular allosteric pockets could be widely conserved amongst GPCRs and these pockets might have potential therapeutic value for drug discovery. On the other hand,

another remarkable finding related to *Cmpd-6FA* was that it enhances both orthosteric agonist binding and transducer coupling by stabilizing the active-state of the β_2AR. *Cmpd-6FA* stabilized the ICL2 α-helix with the help of inward displacement of *Pro*138 and enhanced the agonist affinity. This affinity increase was maintained by ~ 3 Å inward movement of TM3, which in turn led to an outward movement of TM5 and TM6 to prevent steric clashes between *Tyr*132 and *Ile*135 on TM3 and residues on TM5 and TM6. Since this outward movement is a key factor to check the GPCR activation, *Cmpd-6FA* binding increased the population of receptors adopting active conformations, which, in turn, had higher affinities for agonist (Rasmussen, Choi, *et al.,* 2011; Rasmussen, Devree, *et al.,* 2011; Ring *et al.,* 2013). It was also clear that the α-helix formation of ICL2 led to a structural rearrangement in the receptor, which would be disfavored when the orthosteric inverse agonist, carazolol, was bound. Further analysis on *Cmpd-6FA* revealed that the compound exerted its agonist trapping effect (Devree *et al.,* 2016) through stabilizing the active-state of the receptor with a closed hormone-binding site. Thus, these experimental findings on the β_2AR led us to investigate further the residues playing a role in intrinsic dynamics and allosteric coupling in the receptor.

3. MATERIALS AND METHODS

3.1 Molecular Dynamics

Molecular dynamics (MD) is a computational modeling method to study the physical movements of atoms and molecules. Since the first application by Karplus and coworkers (McCammon, Gelin, and Karplus, 1977), MD simulations have become into a sophisticated and practical tool for studying dynamics and energetics of biomacromolecules, especially proteins (Dror, Dirks, Grossman, Xu, and Shaw, 2012; Karplus and McCammon, 2002).

All possible conformations on the potential energy surface should be sampled in order to simulate a system effectively. Instead of calculating potential energy by solving the Schrodinger equation, which is computationally demanding for large systems, MD simulations seek interatomic potentials by a set of simple empirical energy functions known as force fields (Mackerell, 2004; Ponder and Case, 2003). Several popular force fields are currently available: CHARMM (Chemistry at HARward Macromolecular Mechanics) (MacKerell *et al.,* 1998), AMBER (Assisted Model Building with Energy Refinement) (Bayly *et al.,* 1995), GROMOS (GROningen MOlecular Simulation) (Oostenbrink, Villa, Mark, and Van Gunsteren, 2004), and OPLS (Optimized Potentials for Liquid Simulations) (Jorgensen, Maxwell, and Tirado-Rives, 1996). These force fields slightly differ in the energy functions and the way they were parameterized, but they all calculate total energy by summing bonded and non-bonded potentials as a function of the atomic coordinates. For instance, the total CHARMM force field follows the form (Brooks, B. R. and Brooks, III, C. L. and MacKerell, Jr. *et al.*, 2009):

$$\mathcal{V}(r^N) = \sum_{bond} \frac{k_l}{2}(l - l_0)^2 + \sum_{angle} \frac{k_\theta}{2}(\theta - \theta_0)^2 + \sum_{torsion} k_\varphi\,(1 + \cos(n\varphi - \gamma)) +$$

$$\sum_{improp} k_\omega\,(\omega - \omega_0)^2 + \sum_{i=1}^{N} [\![\sum_{j=1+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right]]\!] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \quad (3.1)$$

Bonded potentials include a bond (*b*), valence angle (θ), dihedral angle (φ), and improper dihedral angle (ω) contributions. Bond stretching and angle bending are represented by harmonic potentials in the first and second term of Eqn 3.1, respectively. The dihedral angle term is a sinusoidal expression; ω is the dihedral angle term, γ is the phase shift, *n* is the multiplicity term, and *k*'s are the respective force constants. The improper dihedral potentials are used to control out-of-plane motions (Buck, Bouguet-Bonnet, Pastor, and MacKerell, 2006; Mackerell, Feig, and Brooks, 2004; Alexander D. MacKerell, Feig, and Brooks, 2004). Non-bonded terms include the van der Waals (vdW) energy treated by Lennard-Jones (LJ) potential (Jones, 1924) and the electrostatic energy treated by Coulomb's law. In Coulombic term, q_i and q_j are atomic partial charges, ε_0 is the permittivity of vacuum, and ε is the relative dielectric term. The LJ term includes the r^{-12} term that describes the short-range repulsion and the r^{-6} that describes the long-range attraction; ε_{min} represents the well depth, r_{ij} is the *ij* interatomic distance, and R_{min} is the separation distance where the LJ-potential is minimum. Non-bonded interactions are evaluated for all non-covalent atom pairs within a user-defined cutoff distance.

MD generates a trajectory of molecules as a function of time by numerically solving Newton's equations of motion (Equation 3.2).

$$F_a = -\nabla_a \mathrm{U} = m_a a_a = m_a \frac{dv_a}{dt} = m_a \frac{d^2 r_a}{dt^2} \quad (3.2)$$

$$-\frac{dU}{dr_a} = m_a \frac{d^2 r_a}{dt^2} \quad (3.3)$$

A simplified algorithm of standard MD simulation is displayed in figure 3.1. Based on the equation above, with initial coordinate sets $\vec{r}_a$ are given, forces $\vec{F}_a$ acting on atoms are calculated using force fields. Accelerations $\vec{a}_a$ are then derived (Equation 3.2). From $\vec{r}_a$, $\vec{F}_a$, $\vec{a}_a$ and velocity $\vec{v}_a$ at time t, coordinates at time $t+\Delta t$ can be determined,

where U ($\vec{r}_a$) at new positions can be evaluated (Equation 3.3). These steps are repeated until the system propagates over time and generate a trajectory of molecular motion.

$$v_a = a_a t + v_0 = \frac{dr_a}{dt} \tag{3.4}$$

$$r_a = v_a t + r_0 = \frac{1}{2} a_a t^2 + v_0 t + r_0 \tag{3.5}$$

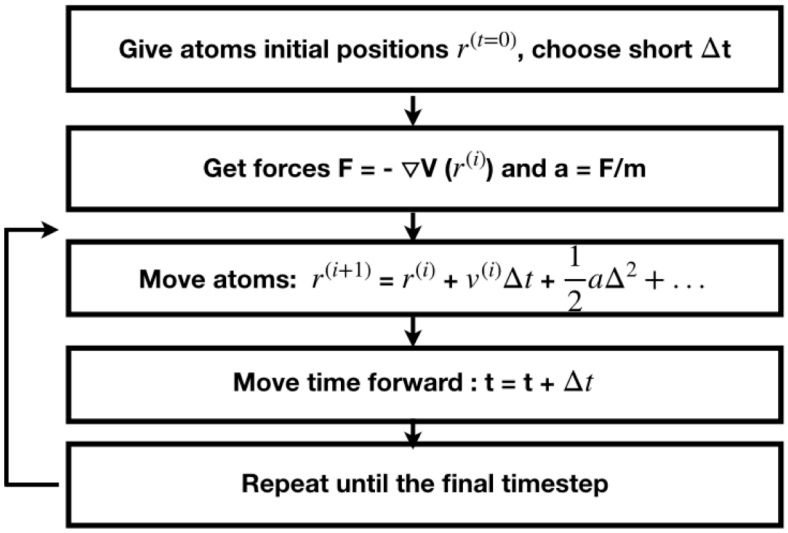


Figure 3.1 An algorithm of standard MD simulation

3.1.1 Verlet algorithm

Verlet method (Verlet, 1967) is a numerical approach uses Newton's equations of motion (Equation 3.2). Atomic trajectories in MD simulations are calculated using Verlet's algorithm since it offers great stability and important properties in physical systems such as time-reversibility and area-preserving properties. The basic idea is to write two third-order Taylor expansions for the positions r one forward and one backward in time. Based on Equations 3.4 and 3.5, we can derive:

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \cdots \tag{3.6}$$

The Verlet algorithm has been derived from Taylor expansion as:

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \tag{3.7}$$

$$r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \quad (3.8)$$

When Equations 3.7 and 3.8 are summed, the new position of the particle is obtained from the current position, previous position, and the particle's acceleration:

$$r(t+\delta t) = 2r(t) + r(t-\delta t) + a(t)\delta t^2 \quad (3.9)$$

3.2 System Preparation for Standard MD Simulation

3.2.1 Inactive human β_2AR system preparation (PDB ID:2RH1)

The system used was the crystal structure of inactive human β_2AR bound to a partial inverse agonist carazolol and T4-Lysozyme (PDB ID: 2RH1). A 1.5 μs long MD trajectory of β_2AR, which was generated in a prior MD study (Ozcan *et al.,* 2013; Ozgur, Doruker, and Akten, 2016), was used for analysis. T4 lysozyme was removed from the crystal structure and replaced by an atomistic model of ICL3 predicted using homology modeling via the MODELLER tool (Eswar *et al.,* 2006). The system was embedded in a POPC bilayer, solvated in a box of TIP3P water molecules, and then ionized with Na^{+2} and Cl^{-} to neutralize the total charge of the system. In order to describe protein and lipid interactions, CHARMM22 and CHARMM27 forcefields were used, respectively. Energy minimization, equilibrium, and production steps of the system have been performed using NAMD v2.7 (Phillips *et al.,* 2005). A sample configuration file used to initiate an MD run is given in Appendix A.

The system was equilibrated for 50 ns, and the root means square deviation (RMSD) was monitored during the equilibration phase. After the stabilization of the system, the production phase was initiated. At 450 ns, a conformational change in ICL3 made its first appearance. ICL3 began to pack itself beneath the receptor in a compact form while adopting short helical formations. Its complete closure was over at around 750 ns. In between, it gave rise to a transition phase of roughly 300 ns long. Upon its closure, a new phase started where ICL3 was persistently preserving its conformation with minor local fluctuations for the rest of the simulation, which lasted about 750 ns. Thus, the trajectory between 50 ns and 450 ns was designated as “Phase I”, which is the known inactive phase, while the trajectory between 750 ns and 1500 ns, was designated as

"Phase II", which is the novel inactive phase throughout this thesis. All calculations in this chapter were performed using backbone Cα atoms for simplicity, which was previously reported to give significantly similar results with fully atomistic calculations (Erman, 2015).

3.2.2 Active human β_2AR system preparation (PDB ID:3SN6)

The active receptor system was prepared from the crystal structure of the active β_2AR (PDB ID: 3SN6) (Rasmussen *et al.,* 2011). This active β_2AR crystal structure has 3.20 Å resolution with bound trimeric Gs protein, T4 lysozyme, and VHH Antibody (nanobody), which was resolved in 2011. The T4 lysozyme, nanobody and trimeric Gs protein were initially removed. The engineered mutations on the residues *Thr*96 and *Thr*98, which were used to increase T4- β_2AR expression levels for crystallization purposes, were mutated back to their natural state, which is *Met*. Similarly, *Glu*187 was mutated back to *Asn*. The missing ICL3 sequence between the residues *Arg*239 to *Cys*265 and the missing ECL2 sequence between the residues *Arg*175 to *Gln*179 were atomistically modeled by MODELLER (Eswar *et al.,* 2006). The new active receptor model was aligned with inactive receptor structure and embedded into the previous system, including water, ions, and POPC lipid membrane bilayer. The new system consists of 68,013 atoms. CHARMM22 and CHARMM27 forcefields were used to describe the interaction potential of protein and lipids. MD simulation for 2 μs was performed using the version of NAMD v2.12. The simulation was run at constant temperature (310 K) and pressure (1 atm) in NPT ensemble. During the production run, the time step and collection interval were set to 2 fs and 1 ps, respectively. The time interval between frames was selected as 0.2 ns, and RMSD of the protein was monitored. The initial 100 ns equilibration phase was eliminated, and the remaining 1900 ns were analyzed. This receptor-only structure was denoted as an *intermediate* state for the rest of the thesis.

3.2.3 System preparation for active human β_2AR with trimeric G protein (PDB ID:3SN6)

The active receptor state that was used here was also the crystal structure of human β_2-AR-trimeric Gs complex bound to the agonist BI-167107, extracellular T4 Lysozyme, and intracellular nanobody (Rasmussen *et al.,* 2011). T4 Lysozyme and nanobody, as well as the agonist, were removed from the initial complex. The missing extracellular (*Ala*176-*His*178) and intracellular (*Phe*239-*Phe*265) residues were atomistically modeled using MODELLER (Eswar *et al.,* 2006). Engineered mutations as *Thr*96, *Thr*98, and *Glu*187, which were used as a linker in the crystal structure formation process, were reversed with VMD version 1.9.3 Mutate Residue plugin to *Met*96, *Met*98, and *Asn*187. The system was embedded into POPC lipid bilayer and solvated with TIP3P water molecules and later ionized with 160 Na^{+2} and 154 Cl^{-1} counter ions for neutralization, which is necessary for Particle-Mesh Ewald summation method, which is used in the calculation of electrostatic energies. The system with the dimensions of 125 Å x 125 Å x 165 Å was prepared with a total of 228,299 atoms. CHARMM36 forcefield was used to describe the interaction potential of protein and lipids. Periodic boundary conditions were applied in an isothermal, isobaric NPT ensemble with a constant temperature of 310 K and a constant pressure of 1 bar. Temperature and pressure are controlled by Langevin thermostat and Langevin piston barostat, respectively. The equations of motion were integrated with a 2 fs time step, and SHAKE algorithm was used to constrain covalent bonds involving nonwater hydrogen bonds with a non-bonded cutoff 12 Å.

Two independent runs were initiated with three steps of initial energy minimizations under flexible cell conditions including 1) the melting of lipid tails when the rest of the atoms were fixed, 2) minimization and equilibration when protein was constrained but lipid, water, and ion atoms were released, and 3) minimization and equilibration in the release of all atoms, which was then followed by equilibrium and production runs under constant area according to membrane proteins' simulation protocol of NAMD (Phillips *et al.,* 2005) which was optimized in terms of atomic information according to CHARMM36 forcefield parameters. Lipid bilayer in the system was continuously

monitored in the minimization and equilibration steps until reaching 63.69 $Å^2$ area per lipid ratio, which was in the range of experimentally reported (Kučerka, Nieh, and Katsaras, 2011) and updated in CHARMM36FF POPC bilayer area per lipid ratio, which was 64.3 ± 1.3 $Å^2$. The production steps of the simulations were run at constant temperature (310 K) and pressure (1 atm) in NPT ensemble under constant area. During the production runs, the time step and collection interval were set to 2 fs and 1 ps, respectively.

First of all, RMSD between two equilibrated states of the active receptor, namely *active-Gp* and *active-Gp*$_{2nd}$, were calculated as 3.58 Å. The 1.5 μs long trajectories were aligned to their initial frames, and the RMSDs were monitored. For the first run, receptor RMSD was stabilized at 50 ns, and the stability was not changed for the remaining 1450 ns. The time interval between frames was selected as 0.2 ns, and the total number of frames to be analyzed was 7250 for this state. 7250 frames were also gathered from the second simulation for the comparison with the *active-Gp*. In order to compare active states with the inactive and *intermediate* receptor states, the G-protein subdomains were not included in the calculations.

3.3 Linear Analysis Methods

3.3.1 Root mean square deviation (RMSD)

Root mean square deviation was calculated by taking the relatively stable parts of the receptor, such as transmembrane regions' deviation into account. It is one of the best methods used for structural comparison. RMSD data gives how much target structures have deviated from the reference structure. It is calculated by:

$$RMSD\ (x, y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} |x_i - y_i|^2} \qquad (3.10)$$

where n is the total number of conformations.

3.3.2 Root mean square fluctuation (RMSF)

The amount of fluctuation from the well-defined average position of the residues is defined as the root mean square fluctuation (RMSF). It is calculated as:

$$RMSF(x) = \sqrt{\frac{1}{T}\sum_{t_j=1}^{T}\left(x_i(t_j) - \langle x_i \rangle\right)^2} \tag{3.11}$$

where t is the instantaneous positions of the residues and $< \cdots >$ denotes average position.

3.3.3 Distance difference calculation

Distances between residue-pairs were calculated for every snapshot and averaged for each phase of the trajectory as $< R_{ij} > = \left(\frac{1}{N}\right)\sum_{t=0}^{N-1} |\mathbf{R_i}(\mathrm{t}) - \mathbf{R_j}(\mathrm{t})|$. Then, the change in the residue-pair distance values from Phase I to Phase II were computed as $\Delta R_{ij} = \langle R_{ij} \rangle_{PhaseII} - \langle R_{ij} \rangle_{PhaseI}$. The information provided with this formula helps us determine the largest distance variations; therefore, the major residue-based displacements between two states.

3.3.4 Distance fluctuations

Distance fluctuations were represented by the *NxN* matrix with entries D_{ij}, where *N* is the number of residues. D_{ij} was calculated over the trajectory with the formula $D_{ij} = \langle \left(d_{ij} - \langle d_{ij} \rangle\right)^2 \rangle^{1/2}$. d_{ij} is the distance between each C_α atoms of amino acids *i* and *j*. Brackets indicate the time-averaged over the entire trajectory. Distance fluctuations help us to identify how much a residue-pair fluctuates from its average positions throughout the simulation, especially in the presence of large conformational changes in the protein subdomains.

3.3.5 Mechanical stiffness calculation

The effect of ICL3 closure on the mechanical stiffness of β_2AR was calculated with the changes in the 'stiffness' parameter k_{ij} between Phase I and Phase II with the formula;

$$k_{ij}^{-1} = \left\{\sum_m \left[\sum_n \left(\langle \Delta R_m^j \Delta R_n^j \rangle - \langle \Delta R_m^j \Delta R_n^i \rangle - \langle \Delta R_m^i \Delta R_n^j \rangle + \langle \Delta R_m^i \Delta R_n^i \rangle\right) u_m^{ij} u_n^{ij}\right]^2\right\}^{\frac{1}{2}} \Big/ \left|R^j - R^i\right| \quad (3.12)$$

Here, m and n are the Cartesian components of the residues *i* and *j*, ΔR_m^j is the m^{th} component of the instantaneous fluctuation of one residue, and ΔR_n^i is the n^{th} component of the other residue. u_m^{ij} is the m^{th} component of the unit vector from *i* to *j* $u_m^{ij} = \left(R_m^j - R_m^i\right)/\left|R^j - R^i\right|$.

The mechanical stiffness is inversely proportional to the RMSF. The differences of the spring constants in Phase I and Phase II are comparable in terms of the change of the stiffness with respect to the ICL3 movement.

3.3.6 Residue pair cross-correlation calculations from MD trajectory

Correlations between atomic fluctuations from average positions of two residues *i* and *j* were calculated using the following equation:

$$C_{i,j} = C\left(\Delta R_i, \Delta R_j\right) = \frac{\langle \Delta R_i(t) \cdot \Delta R_j(t) \rangle}{\sqrt{\langle (\Delta R_i)^2 \rangle \langle (\Delta R_j)^2 \rangle}} \quad (3.13)$$

The time average of the dot product of $\Delta R_i(t)$ and $\Delta R_j(t)$ is taken and then the obtained value is normalized. If $C_{ij} = 1$, then the fluctuations of atoms *i* and *j* are perfectly correlated (fluctuates in the same direction), if $C_{ij} = -1$, then the fluctuations of atoms *i* and *j* are perfectly anticorrelated (fluctuates in opposite directions), and if $C_{ij} = 0$, then the atoms *i* and *j* fluctuate independently.

3.3.7 Contact map generation

The cutoff distance (R_c) for heavy atoms (C, N, O, S) was taken as 6 Å, below which the atoms are considered to be in contact. Therefore, while determining the residue-pair contacts, the same R_c was used. The preference of heavy atom distance calculations instead of alpha carbons will prevent us from underestimating side-chain contacts. The formula used for contact map calculation was:

$$M_{i,j} = \begin{cases} 1, if\ \delta_{i,j} \leq R_c \\ 0, Otherwise \end{cases} \tag{3.14}$$

Contact percentages over the MD trajectory were calculated with $\sum_{n=1}^{nconf} M_{i,j}/nconf$ and the threshold was set to 75% of the whole trajectory in order to detect stable contacts.

3.3.8 Autocorrelations

The time delay required to decrease autocorrelation values of residues below 1/e was calculated with the formula 3.15. Different residues may need different time delays to lose their autocorrelations based on their flexibility during the trajectory.

$$\frac{C_{i,i}(\tau)}{C_{i,i}(0)} = \frac{\langle \Delta R_i(0) \cdot \Delta R_i(\tau) \rangle}{\langle \Delta R_i(0) \cdot \Delta R_i(0) \rangle} \tag{3.15}$$

3.4 Probabilistic Analysis of Backbone Fluctuations

3.4.1 Conditional time delayed cross-correlations

Correlations between atomic fluctuations residue-pairs at different time steps were calculated by:

$$C_{i,j\,(\tau)} = \frac{\langle \Delta R_i(0) \cdot \Delta R_j(\tau) \rangle_\chi}{\sqrt{\langle (\Delta R_i)^2 \rangle \langle (\Delta R_j)^2 \rangle}} \tag{3.16}$$

where χ indicates the conditional average and calculated as:

$$\langle \Delta R_i(0) \cdot \Delta R_j(\tau) \rangle \chi = \sum_k p\left(\Delta R_j(t_k + \tau) | \Delta R_i(t_k)\right) p(\Delta R_i(t_k)) \Delta R_i(t_k) \Delta R_j(t_k + \tau) \quad (3.17)$$

This formula was used to measure the probability of observing $\Delta R_j(t_k + \tau)$ at time $(t_k + \tau)$, with the condition of $\Delta R_i(t_k)$ is at time t_k. Therefore, a driver-follower relationship between the residues *i* and *j* can be considered only if the $C_{ij}(\tau) \neq C_{ji}(\tau)$. Otherwise, $C_{ij}(\tau) = C_{ji}(\tau)$ does not represent any net information transfer between the residue-pairs.

$p(\Delta R_i(t_k))$ was calculated by first computing the minimum and maximum values of ΔR_i, then dividing this interval with the residue-optimized bin numbers and lastly, determining the probability of occurrence of residue *i* in each bin. The frequency of occurrence of ΔR_i in a particular bin gives us the probability of finding the corresponding value of ΔR_i at time t_k. Additionally, $p\left(\Delta R_j(t_k + \tau) | \Delta R_i(t_k)\right)$ is the conditional probability and depends on the frequency of occurrence of $\Delta R_j(t_k + \tau)$ at time $t_k + \tau$, with a condition of observation of the frequency of occurrence of ΔR_i at time t_k, $p\ (\Delta R_i(t_k))$.

3.4.2 Shannon's entropy

Shannon's entropy is an important metric for information theory. Storage and transfer of information between variables are possible. Considering this idea, the idea of entropy can be derived as the amount of information stored in that variable. Shannon, therefore, proposed an algorithm to quantify the amount of information present in a variable. In our case, this information comes from MD trajectories by either quantifying the spontaneous deviation from the average Cα values or angular data from the backbone or side-chain dihedrals of any residue *i*.

Strictly speaking, Shannon's entropy formula is a measure for uncertainty, which increases with the number of bits needed to reveal the realizations of residues (Shannon, 1948). In order to measure the information flow between two processes, Shannon entropy is used with the approach of the Kullback-Leibler divergence (Leibler

and Kullback, 1951) and by considering that signals evolve according to a Markov process (Schreiber, 2000). In other words, if *i* and *j* denote two discrete random variables with possessing marginal probability distributions *p(i)* and *p(j)* and a joint probability distribution *p(i,j)*, whose dynamical actions correspond to stationary Markov processes as *i* and *j*. The Markov property implies that the probability of observing *i* at time t + 1 in state *i* with conditioning the *k* previous observations is;

$$p(i_{t+1} \mid i_t, \ldots, i_{t-k+1}) = p(i_{t+1} \mid i_t, \ldots, i_{t-k}) \tag{3.18}$$

The average number of bits needed to disclose the observation in *t + 1* if the previous *k* values are known is,

$$H_{i(k)} = \sum p(i_{t+1}, i_{(k)t}) \cdot log_2(p(i_{t+1} \mid i_{(k)t})) \tag{3.19}$$

As a summary, Shannon entropy for residue *i* is calculated with the formula;

$$H(I) = -\sum_{i}^{k} p\left(\Delta R_i(t_k)\right) log_2 \left(p\left(R_i(t_k)\right)\right) \tag{3.20}$$

Where *k* is the number of states. If $\log_2$ is used in the formula, the information is in *bits,* which is one of the mostly used parameters in information theory. Natural logarithm (*ln*), on the other hand, gives the information in *nats*.

3.4.3 The optimization of the residue-specific number of bins

Probability calculations require discretizing the sample space in order to calculate the *probability of occurrence* of an event at certain intervals, called *bins*. This number of bin determination is highly dependent on the sample distribution, and various approaches are used to define the optimum bin number. We used the Shannon entropy method, as explained below, to calculate the probability of residue *i* at time *t* in terms of the frequency of occurrence between the time *t-0.5τ* and *t+0.5 τ*, where *τ* is the size of the residue-specific number of bins. The number of bins is calculated by determining the bin number, which maximizes the Shannon entropy for each residue *i*.

The criterion for the number of bins was defined as the convergence of two sequential bin numbers as $|(H(N_{bins}+1)_i - H(N_{bins})_i)|/H(N_{bins})_i < 0.02$

3.4.4 Mutual information (MI)

MI is a measure for quatifying the information, which can be gathered about a variable *a* by observing another variable *b*. Probability distribution is essential in MI quantification and is sensitive to second and all higher-order correlations. Even though MI is a powerful model for identification of long-range communicating residue-pairs in proteins due to its sensitivity to high-order correlations, it says little about causal relationships due to both by being symmetrical and resulting in shared information by two signals. However, it still gives significant data in protein studies about how much information is shared between residue-pairs.

Mutual information between residue-pairs *i* and *j* was calculated using the following expression:

$$MI(i,j) = \sum_{k} p\left(\Delta R_i(t_k), \Delta R_j(t_k)\right) log_2 \frac{p\left(\Delta R_i(t_k), \Delta R_j(t_k)\right)}{p\left(\Delta R_i(t_k)\right).p\left(\Delta R_j(t_k)\right)} \quad (3.21)$$

Where $p(\Delta R_i(k), \Delta R_j(l))$ represents the joint probability of observing the fluctuation of residue *i* in state *k* and that of residue *j* in state *l*. Mutual information is a non-negative and symmetric quantity, and zero if the fluctuations of residue *i* are independent of the fluctuations of residue *j*.

3.4.5 Optimization of the time delay

The time lag parameter τ was optimized for each residue *i* independently as the time step at which the mutual information $MI(i_k, i_{k+\tau})$ has its first minimum (Cellucci, Albano, and Rapp, 2003; Kamberaj and van der Vaart, 2009; Noakes, 1991; Sauer, Yorke, and Casdagli, 1991), i.e., the future fluctuations of residue *i* became independent of its past fluctuations, in a sense the residue's memory of its own past weakens. Mutual information between two instances of a residue *i* separated by τ is defined as,

$$MI(i_k, i_{k+\tau}) = \sum_{k} p\big(\Delta R_i(t_k), \Delta R_i(t_{k+\tau})\big) log_2 \frac{p\big(\Delta R_i(t_k), \Delta R_i(t_{k+\tau})\big)}{p\big(\Delta R_i(t_k)\big).p\big(\Delta R_i(t_{k+\tau})\big)} \quad (3.22)$$

where $p(\Delta R_i(k), \Delta R_i(k+\tau))$ stands for the joint probability of observing the fluctuations of residue *i* in two different states separated by τ.

3.4.6 Transfer entropy (TE)

Information flow represents the reduction of uncertainty, which exists when one measures the value of a variable. Schreiber demonstrated that if one relates the prediction gain to uncertainty reduction, a measure for Wiener's causality could be derived. By definition, a causal dependence is related to the information being exchanged rather than shared. Therefore, Schreiber's information theory is a better approach for the dynamic structure of signals and identifying causality due to obtained directionality. Unlike standard measures of the residue-pairs' fluctuations, which are entirely symmetric, transfer entropy is inherently asymmetric. Moreover, it is able to detect asymmetries in the interaction of systems from observational data. Information flow from process *j* to process *i* is measured similarly by calculating the deviation from the generalized Markov property relying on the Kullback-Leibler distance (Schreiber, 2000). Thus, transfer entropy is given by;

$$TE_{i\to j}\,(k,l) = \sum_{i,j} p\big(i_{t+1}, i_{(k)t}, j_{(l)t}\big) \log \frac{p\big(\big(i_{t+1}, i_{(k)t}, j_{(l)t}\big)}{p\big(i_{t+1}|i_{(k)t}\big)} \tag{3.23}$$

where $TE_{j\to i}$ defines the information flow from *j* to *i* and $TE_{i\to j}$ defines the information flow from *i* to *j*. If we rewrite this equation based on the difference of two entropies we obtain:

$$TE_{i\to j}(\tau) = H\big(\Delta R_j(t)\big|\Delta R_j(t-\tau)\big) - H\left(\Delta R_j(t)|\Delta R_j(t-\tau), \Delta R_i(t-\tau)\right) \tag{3.24}$$

where $H\big(\Delta R_j(t)\big|\Delta R_j(t-\tau)\big)$ is the conditional entropy of residue *j* at time *t* given the values of $\Delta R_j(t-\tau)$ at time t$-\tau$ and the second term $H\left(\Delta R_j(t)|\Delta R_j(t-\tau), \Delta R_i(t-\tau)\right)$ is the conditional entropy of residue *j* at time *t* given the values of $\Delta R_j(t-\tau)$ and $\Delta R_i(t-\tau)$ at time t$-\tau$. Transfer entropy, $TE_{i\to j}(\tau)$, signifies the decrease in the amount of entropy in the trajectory of *j* due to a knowledge of the past fluctuations of residue *i*. Also, it is an asymmetric and non-negative quantity and varies between its

minimum value of 0, when fluctuations of residues *i* and *j* are completely independent, and its maximum value $H\left(\Delta R_j(t)\middle|\Delta R_j(t-\tau)\right)$ when fluctuations of residues *i* and *j* are completely coupled. The entropy values $H\left(\Delta R_j(t)\middle|\Delta R_j(t-\tau)\right)$ and the second term $H\left(\Delta R_j(t)\middle|\Delta R_j(t-\tau),\Delta R_i(t-\tau)\right)$ in Equation 3.24, were determined using the following two expressions,

$$H\left(\Delta R_j(t)\middle|\Delta R_j(t-\tau)\right) = -\langle log_2\, p\left(\Delta R_j(t),\Delta R_j(t-\tau)\right)\rangle + \langle log_2\, p\left(\Delta R_j(t-\tau)\right)\rangle \quad (3.25)$$

$$H\left(\Delta R_j(t)\middle|\Delta R_j(t-\tau),\Delta R_i(t-\tau)\right) = -\langle log_2\, p\left(\Delta R_j(t),\Delta R_j(t-\tau),\Delta R_i(t-\tau)\right)\rangle + \langle log_2\, p\left(\Delta R_j(t-\tau),\Delta R_i(t-\tau)\right)\rangle \quad (3.26)$$

where $\langle\cdots\rangle$ denotes conditional averages. As a result, the transfer entropy equation becomes

$$\begin{aligned} TE_{i\to j} = & -\langle log_2\, p\left(\Delta R_j(t),\Delta R_j(t-\tau)\right)\rangle + \langle log_2\, p\left(\Delta R_j(t-\tau)\right)\rangle \\ & + \langle log_2\, p\left(\Delta R_j(t),\Delta R_j(t-\tau),\Delta R_i(t-\tau)\right)\rangle \\ & - \langle log_2\, p\left(\Delta R_j(t-\tau),\Delta R_i(t-\tau)\right)\rangle \end{aligned} \quad (3.27)$$

Furthermore, in order to highlight the entropic activity of a residue, the transfer entropy, which came out of residue I to the rest of the receptor, was calculated using:

$$TE_{i(t-\tau)} = \sum_j TE_{i(t-\tau)\to j(t)} \quad (3.28)$$

Finally, the net transfer entropy from residue *i* to residue *j* is determined by taking the difference between the transfer entropy from *i* to *j* and that from *j* to *i* as:

$$NetTE_{(i\to j)} = TE_{(i\to j)} - TE_{(j\to i)} \quad (3.29)$$

3.4.7 Finite sampling effect

Systemic errors may affect the entropy estimation based on a finite sample size. Previous studies showed that correction was required to obtain true entropy values of the systems based on;

$$H^{True} \approx H^{Observed} + \frac{M-1}{2N} \quad (3.30)$$

where $H^{Obsertved}$ is the calculated entropy, M is the number of bins with non-zero probability, and N is the total number of conformations (Grassberger, 1988; Herzel, Schmitt, and Ebeling, 1994). Since mutual information $MI(i,j)$ can be written in terms of joint and singlet entropies as in $H_i + H_j - H_{ij}$, the correction for mutual information can be derived as;

$$MI_{ij}^{True} \approx MI_{ij}^{Observed} - \frac{M_{ij} - M_i - M_j + 1}{2N} \tag{3.31}$$

where $MI_{ij}^{Observed}$ is the observed mutual information and M_{ij}, M_i, and M_j are the corresponding histogram bins with non-zero probability in order to calculate H_{ij}, H_i, and H_j. A similar approach was used to derive the correction term for transfer entropy, which becomes;

$$TE_{ij}^{True} \approx TE_{ij}^{Observed} - \frac{M_{ijj} - M_{ij} - M_{jj} + M_j}{2N} \tag{3.32}$$

where M_{ijj}, M_{ij}, M_{jj}, and M_j are representing the bins with non-zero probabilities.

3.5 Probabilistic Analysis of Backbone and Side-chain Dihedrals

3.5.1 Backbone dihedrals

The backbone dihedral angles, ϕ and ψ, can provide important information about the three-dimensional structure of the protein. They vary from -180° to +180°, but they cannot adopt all possible values, because of steric restrictions. The famous Ramachandran plot (Ramachandran and Sasisekharan, 1968) illustrates the sterically allowed regions of the dihedral angles.

The backbone dihedrals were derived from the trajectories by using the Tcl code in the Appendix B. Basically, the angle ϕ is calculated between C-N-Cα-C atoms, and the angle ψ is calculated between N-Cα-C-N atoms in every frame (Figure 3.2).

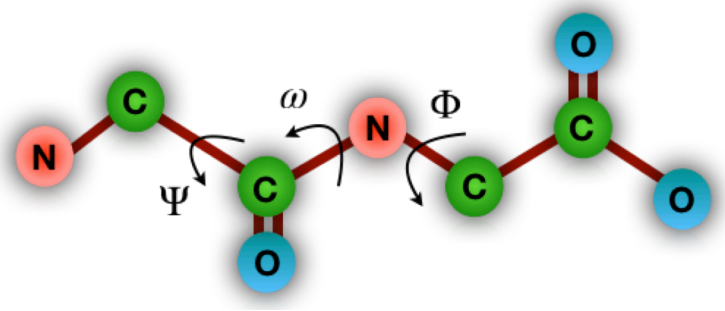


Figure 3.2 Backbone dihedral angles ϕ, ψ and w, and the corresponding atoms

3.5.2 Rotamer library for side-chain dihedrals

Side-chain conformations in proteins have been analyzed dated back that when the earliest crystal structures were resolved (Chandrasekaran and Ramachandran, 1970). With increasing amount of resolution of crystal structures, it was found that most side-chain dihedral angles cluster in discrete bins and that residues prefer certain combinations between each bin. Full and discrete side-chain conformations are labelled as a rotamer (Figure 3.3). Rotamer libraries are used to predict, build, design, and solve new protein structures and were created to detail the angles, frequency, and conformational variance (Ponder and Richards, 1987).

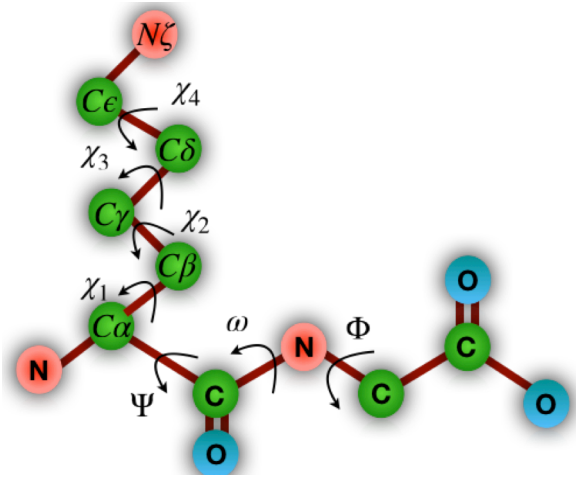


Figure 3.3 Backbone dihedral angles ϕ, ψ and w, together with side-chain torsional angles χ_1, χ_2, χ_3, χ_4, and the corresponding atoms.

We have undertaken a study of the side-chain behavior of proteins using the Dynameomics protein simulation dataset. Dynameomics was reported as an effort to simulate the native state and unfolding behavior of representatives of all autonomous

protein folds (Beck *et al.,* 2008; van der Kamp *et al.,* 2010). They have analyzed the native state molecular dynamics simulations of 807 proteins spanning essentially all known protein folds, each of which was simulated at 298 K for at least 31 ns (Beck *et al.,* 2008). The resulting Dynameomics rotamer library has 74% of the population shared with crystallographically derived rotamer libraries and a correlation coefficient of $R = 0.90$ with a collection of side-chain S^2 order parameters.

Basically, the backbone and side-chain angles were calculated with the Tcl code presented in Appendix B by using the residue-specific angle definitions given in Tables D and E. The number of bins for each residue was obtained from the recent data presented in the Dynameomics website. The details of the number of bins were given in Table D1.

3.5.3 Side-chain mutual information

MI is a measure of the high-order correlation of dihedral rotations between two residues *i* and *j*. Here, it reports on the degree of correlation between residues' rotameric states. The mutual information I_{ij} between residues *i* and *j* is calculated as;

$$MI_{i,j} = \sum_{\Theta_i} \sum_{\Theta_j} p(\Theta_i, \Theta_j)\, log_2 \left(\frac{p(\Theta_i, \Theta_j)}{p(\Theta_i), p(\Theta_j)} \right) \tag{3.33}$$

where $p(\Theta_i,\Theta_j)$ denotes the probability of each of the N_i:N_j joint states of residues *i* and *j*, and N_i is the number of rotatable *sp3-sp3* hybridized bonds in residue *i*. The number of bins for backbone dihedrals is set as 3 based on the distribution of the rotameric states and for side-chain dihedrals from 0 to 6, based on the rotameric information given in the reference rotamer library. Therefore, if the residue *Glu* is considered, the total number of rotatable bonds will be 5, including ϕ and ψ, χ_1, χ_2, χ_3, and the number of bins for each angle is 3, 3, 3, 3, 6, respectively (see Appendix E). Thus, for each amino acid *Glu*, $p(\theta_{Glu}) = 3x3x3x3x6$ possible states.

After rearranging Eq. 3.33, this becomes;

$$MI_{i,j} = H_{ij}^{(res)} - \left(H_i^{(res)} + H_j^{(res)} \right) \tag{3.34}$$

where $H_{ij}^{(res)}$ is the joint entropy related to the discrete rotameric states for residues *i* and *j*. Therefore, $H_{ij}^{(res)} = H_i^{(res)} + H_j^{(res)}$ when the fluctuations of the two residues are completely independent of one another, and $MI = 0$.

However, when the residues are correlated, their entropies are conjoint, and $MI_{ij} > 0$. One feature of mutual information is that calculated estimates of two completely uncorrelated variables only approaches zero at the limit of infinite sampling. For any finite sampling, a small amount of erroneous mutual information will be observed, regardless of the actual coupling between the two variables (Roulston, 1999). When calculating mutual information numerically, this inherently present bias must be justified in order to determine the mutual information's statistical significance. Thus, we followed the same finite sampling error estimation approach, as explained in 3.28. This approach counts the non-zero probability of bins in each entropy calculations and divides them with the double of the sample size (Grassberger, 1988).

3.5.4 Optimization of residue-specific time delay for dihedrals

For computing transfer entropy, a time delay parameter needs to be included in the calculation. Unlike the previous Cα calculations, where each residue-specific number of bins was optimized based on the atom displacements' probabilities, the number of bins here is not residue-specific. Since bin number is one of the major criteria for probability calculation, we compared the effect of residue-specific time delay and the time delay of 5ns on the transfer entropy (Figure 3.4). 5ns for delay time is selected based on the characteristic decay times of correlations defined by Erman and his coworkers (Hacisuleyman and Erman, 2017). Since the comparison of the 5 ns delay time with optimized time delays of the residues gave almost the same data, we continued our analysis by taking the time delay of 5ns for simplicity.

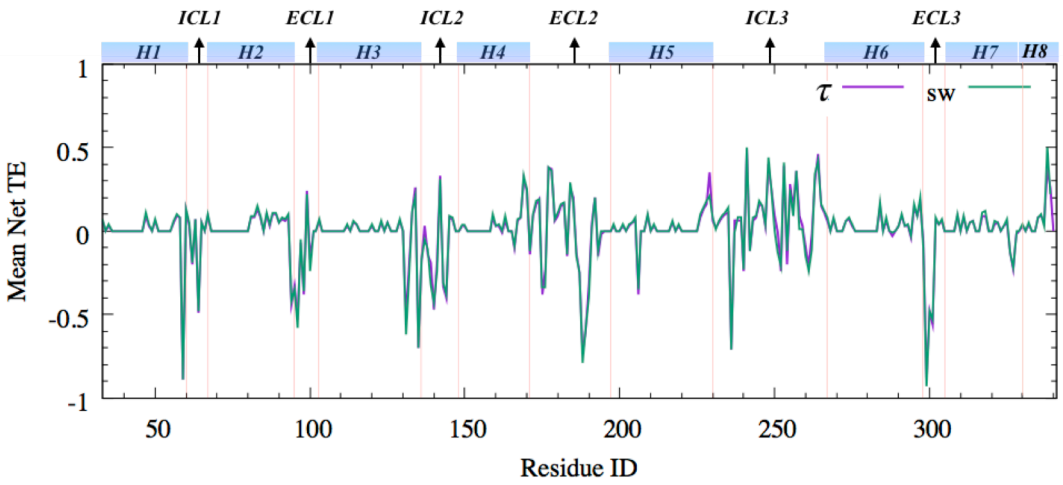


Figure 3.4 Residue specific time delay (τ) values versus time delay of sw = 5 ns on backbone transfer entropy.

3.5.5 Transfer entropy of dihedral angles

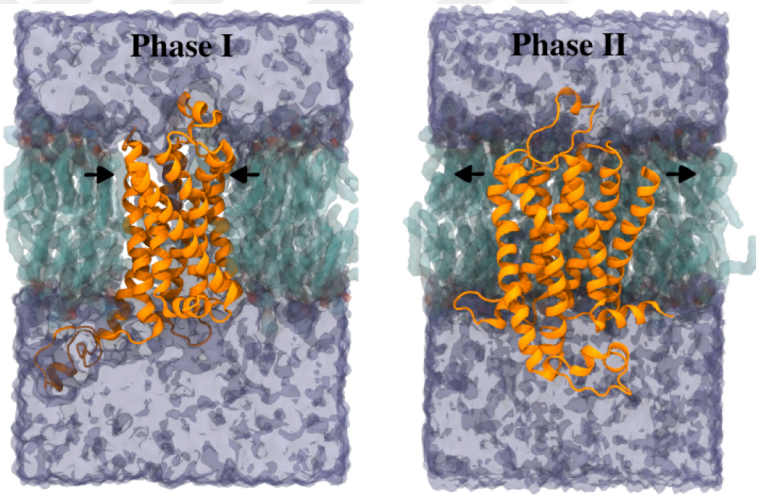
With side-chain transfer entropy, the question of how much information is transferred from the past state of one residue to the current state of another, when the past and the current states of the latter residue are known as important. As it was previously mentioned at Equation 3.22, transfer entropy is the difference of two conditional entropies.

TE is an asymmetric and non-negative quantity and takes values between 0, when fluctuations of residues i and j are completely independent, and $H(i_t|i_{t-\tau})$ when fluctuations of residues i and j are completely coupled. $H(i_t|i_{t-\tau}) = -\sum_{k=0}^{N_m} p(i_t, i_{t-\tau}) log_2 (p(i_t|i_{t-\tau})$, where $N_m = N - \tau$ for N number of snapshots. However, this time *p(i)* is $p(\theta_i)$, where θ includes the number of rotatable *sp3-sp3* bonds for each residue *i*. For example, the residue *Ala* has ϕ and ψ angles but no side-chain rotatable bonds; however, the residue *Lys* has both ϕ and ψ angles and χ_1, χ_2, χ_3, and χ_4 side-chain rotatable bonds. Therefore, $p(\theta_{Ala}) = 3x3$, but $p(\theta_{Lys}) = 3x3x3x3x3x3$ alternative conformations that are visited for each frame in the trajectory (see Appendix E).

However, transfer entropy calculation has a triple joint probability parameter denoted as $p(\Theta_{i(t-\tau)},\Theta_{j(t-\tau)}, \Theta_{j(t)})$. Based on the amino acid pairs such as *Lys-Arg*, the possible number of probabilities for backbone+side-chain TE becomes 3^{18}. Since our system has 311 amino acids for inactive and 312 amino acids for the active state with different numbers of *sp3-sp3* bonds, calculation of TE for backbone+side-chain becomes impractical for more than a few dihedrals. Not only does the dimensionality explode, but the amount of data required for statistics becomes too large. Therefore we split backbone and side-chain data and separately did transfer entropy calculations.

4. THE EFFECT OF RAPID ICL3 CLOSURE ON THE INTRINSIC DYNAMICS OF β_2AR

Protein dynamics and molecular mechanisms are unique to systems of interests, albeit general allosteric models have been developed. Furthermore, MD simulations are well-known techniques to be used for extracting dynamical information. In this chapter, we present various approaches to analyze the dynamic data and determine the hotspots of allosteric communication for inactive β_2AR.



(a) (b)

Figure 4.1 Inactive and "very inactive" conformations of β_2AR. Inactive conformation is denoted as Phase I (left) and "very inactive" conformation is denoted as Phase II (right). *(Adapted from Sogunmez N, Akten ED, 2019, J of Phys Chem B, 23(17), 3630. Copyright © 2019 ACS*)

Here we used the system, which was generated in a previous study (Ozcan *et al*., 2013; Ozgur, Doruker, and Akten, 2016) from the crystal structure of inactive human β_2AR bound to a partial inverse agonist carazolol and T4-Lysozyme (PDB ID: 2RH1). Previously, a 1.5 μs long MD simulation was run, and from this trajectory, a novel inactive state, where the longest loop of the receptor was tightly packed through the G protein-binding site of the receptor, was identified (Figure 4.1). For this chapter, we

used a known inactive state and a novel, very inactive state to understand the intrinsic dynamics of the inactive states of the receptor.

4.1 Results and Discussion

4.1.1 Major conformational shift in ICL3 reveals a new inactive state

ICL3 showed a large conformational shift in a simulation conducted by Ozcan and his coworkers (Ozcan *et al*., 2013). In Figure 4.2a, two distinct states in equilibrium, together with a transition phase were observed in the RMSD profile, where in the transition phase ICL3 is packed beneath the receptor. ICL3 is the longest loop in the receptor, and thus, it is the most mobile part of the receptor. In Phase I, which was the initial 450 ns long simulation trajectory, ICL3 formed transient helices and fluctuated within a window of 6-7 Å. The rest of the receptor was comparably stable with minor fluctuations. However, around 450 ns, ICL3 positioned itself towards the center of the receptor (Figure 4.2b). After ~7 Å of its deviation, ICL3 adopted a stable conformation, which was established by formation of multiple hydrogen bonds with the core of the receptor. This new state was denoted as Phase II. In Phase II, ICL3 was packed with stable helical formations and hydrogen bonds. This new ICL3 location was around the G protein-binding site, blocking its α-helical domain's accessibility from the intracellular part. This complete blockage of the intracellular part by ICL3 was discovered for the first time, where it results in a novel state and was characterized as "*highly inactive*". The receptor continued to remain in this state for about 750 ns. Since ICL3 formed multiple stable hydrogen bonds in Phase II, this novel state would persist beyond 1.5 μs if the system will be simulated further.

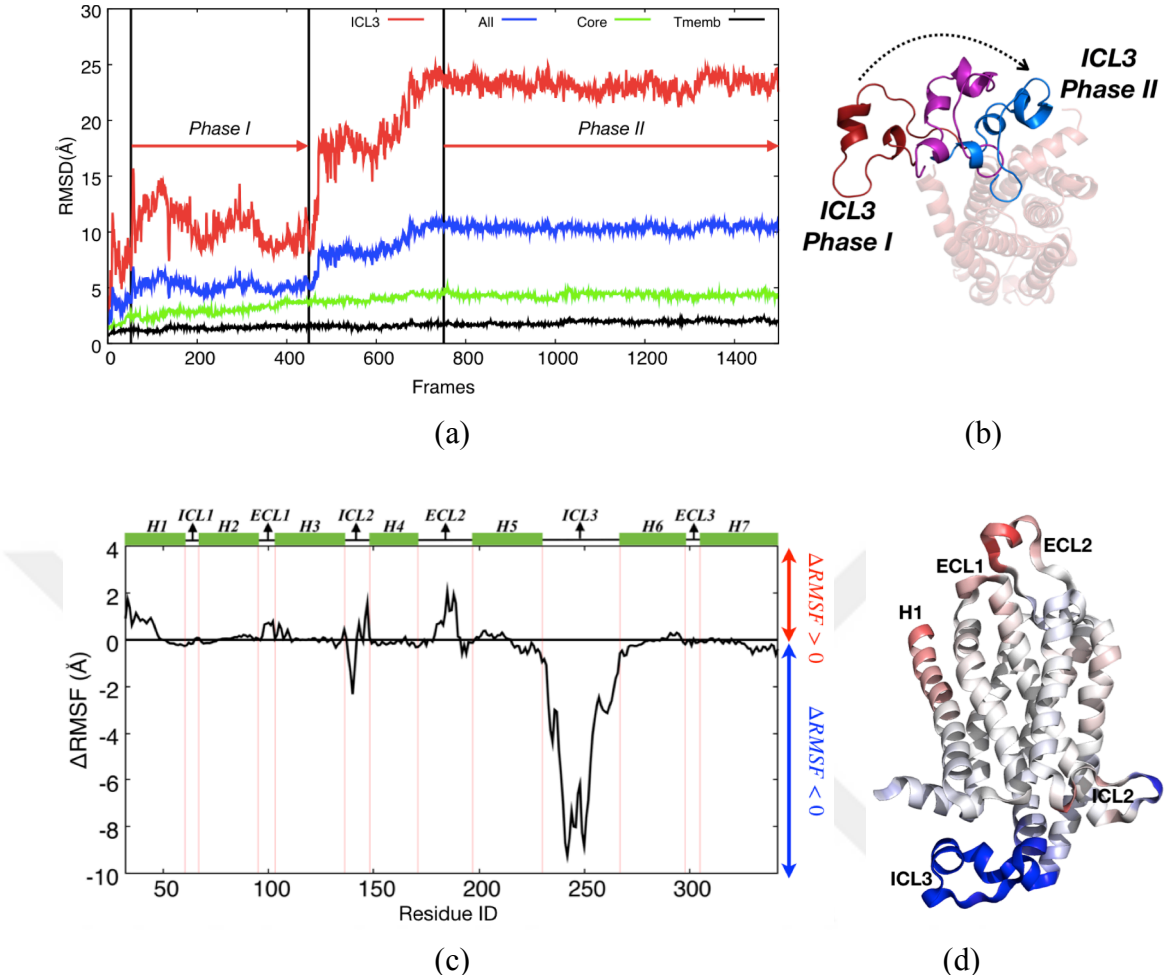
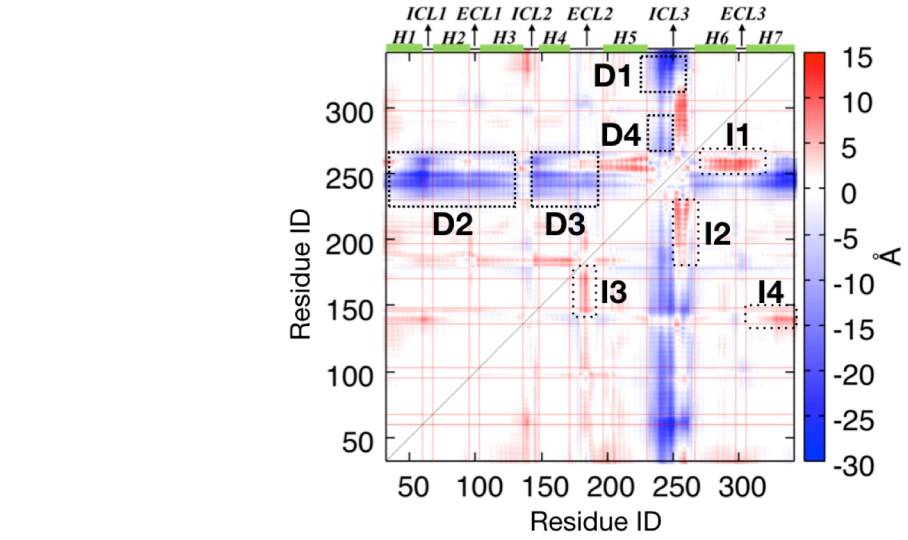


Figure 4.2 The effect of ICL3 on the receptor deviation and residue fluctuation. RMSD of the 1.5 µs simulation. Different parts of the receptor shown in different colors (a). Intracellular view of ICL3 closure (b). Net RMSF (Phase II – Phase I) representing the sites with increase/decrease mobility (c). Protein representation of Net RMSF (d). *(Adapted from Sogunmez N, Akten ED, 2019, J of Phys Chem B, 23(17), 3630. Copyright © 2019 ACS*)

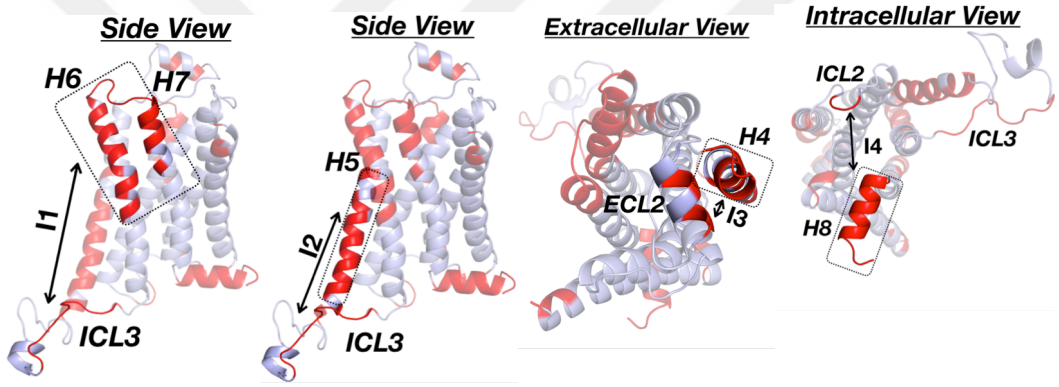
Additionally, RMSFs were investigated in terms of its change from Phase I to Phase II. It is visible in Figures 4.2c and 4.2d that the RMSF indicated the mobility decrease significantly for ICL3, by 8-10 Å, and residues of ICL2 adjacent to H3 were also decreased their mobilities moderately by 2.5 Å, while, remaining ECL regions and the extracellular part of the $H1_{Ext}$ displayed a mobility increase. A minor increase in ICL2 mobility adjacent to H4 was as a result of its secondary structure alteration, where *Thr*146 and *Lys*147 were loosened into an unstructured loop form in Phase II. Finally, the outweight of the stability increase in ICL2 and ICL3 was observed by a mobility increase in all ECLs and $H1_{Ext}$ shifting from Phase I to Phase II.

The distance map demonstrated all pairwise distances between the average positions of each Cα. Therefore, we used the distance map to investigate the overall effect of ICL3's closure to the structural reorganization. Then, significant changes upon shifting to Phase II were determined through two distance map differences between phases, as in Figure 4.3a. Regions colored in red, labeled as I1, I2, I3, and I4, indicated distance increases greater than 7 Å, whereas blue regions, labeled as D1, D2, D3, and D4, indicated distance decreases less than 10 Å. ICL3 closure led to the highest distance change, which was between 15 Å - 30 Å (Figures 4.3b and 4.3c).

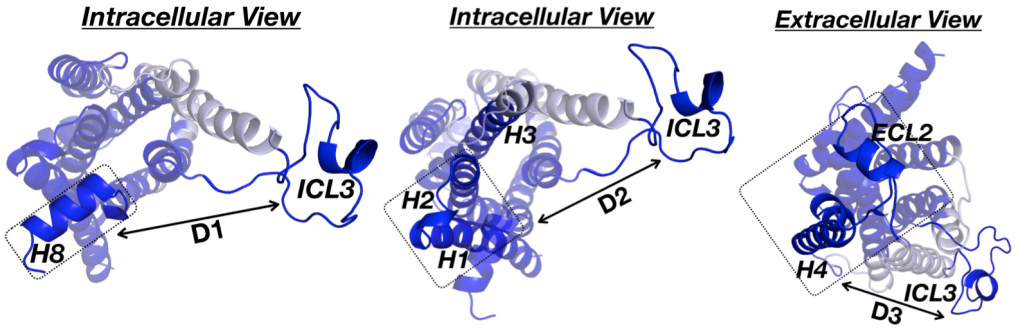
The highest decrease here corresponded to the distances between ICL3 and the TMs, while increased regions were observed as H5, $H6_{Ext}$ and $H7_{Ext}$, and ECL3. The discontinuity between D2 and D3 overlapped with ICL2, which was observed to move outward to leave enough space for ICL3 packing. Fragments with increased distance in Phase II, were ICL3-($H6_{Ext}$, ECL3, and $H7_{ext}$), ICL2-H8, ICL3-H5, and ECL2-$H4_{Ext}$, respectively (See Fig 4.3c). The domains that changed positions significantly relative to one another in the second Phase contain either one or two of these loops.



(a)



(b)



(c)

Figure 4.3 Distance difference between domains in the receptor. Heatmap representation of distance difference between Phase II and Phase I (a). Red and blue colors indicate an increase and decrease in distances between residue-pairs, respectively. Protein domains with increased distances (b) and protein domains with decreased distances (c). *(Adapted from Sogunmez N, Akten ED, 2019, J of Phys Chem B, 23(17), 3630. Copyright © 2019 ACS*)

4.1.2 Change in the mechanical stiffness upon ICL3 closure

The impact of ICL3 closure on the stiffening of overall structure is also important to highlight. This can be measured with the parameter called mechanical stiffness and denoted as k_{ij}. Mechanical stiffness is defined as a measure of rigidity between residues i and j. For Phase I and Phase II, k_{ij} was plotted as in Figures 4.4a and 4.4b, respectively. Their difference, Δk_{ij}, was also calculated to reveal the change in stiffness upon ICL3 closure and illustrated as in Figure 4.4c.

Finally, the average stiffening difference was computed by $\Delta\bar{k}_i = (\bar{k}_{i,Phase\ II} - \bar{k}_{i,Phase\ I})$ for each residue i and plotted as in Figure 4.4e. $\Delta\bar{k}_i$ was presented on the receptor as a color gradient in Figure 4.4d, which was to clearly define the receptor domains that stiffened (colored in red) or relaxed (colored in blue) from Phase I to Phase II. Clearly, ICL3, the adjacent residues on H5 and H6 to ICL3, and spatially nearby H8 stiffened, but the residues on ECL1, $H1_{Ext}$, and few residues of ICL2 adjacent to H4 relaxed significantly.

Similarly, stiffness in $H5_{Ext}$ and $H6_{Ext}$ decreased. This is most probably the reason for the enlargement observed at the orthosteric ligand-binding site in Phase II. Moreover, different parts of ECL2 showed dual charactersitics as both stiffening and flexibility, that the residues adjacent to transmembranes stiffened, but the middle part of the loop became flexible. The change in the average stiffness highly correlated with the change in average fluctuations (Figure 4.4c).

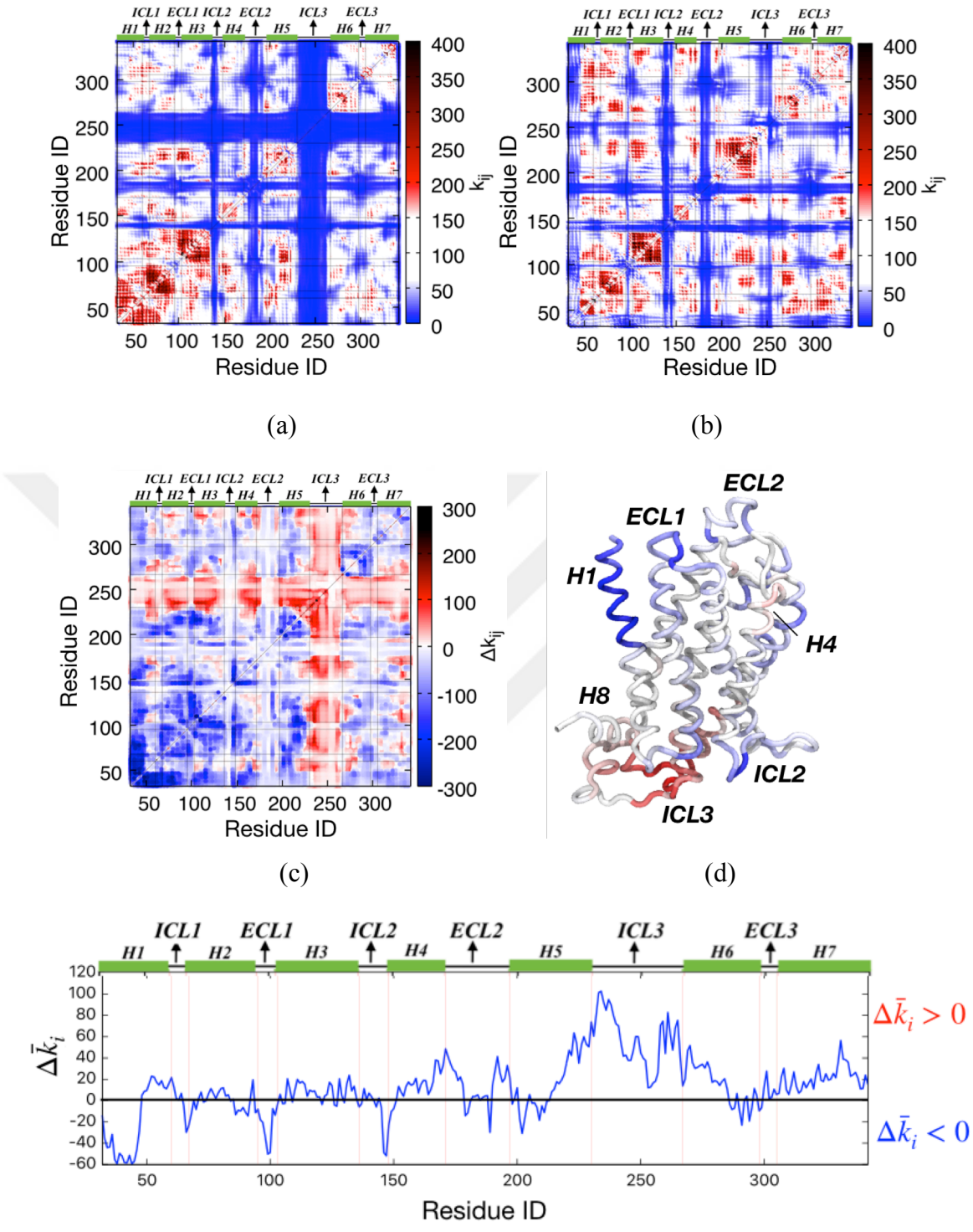
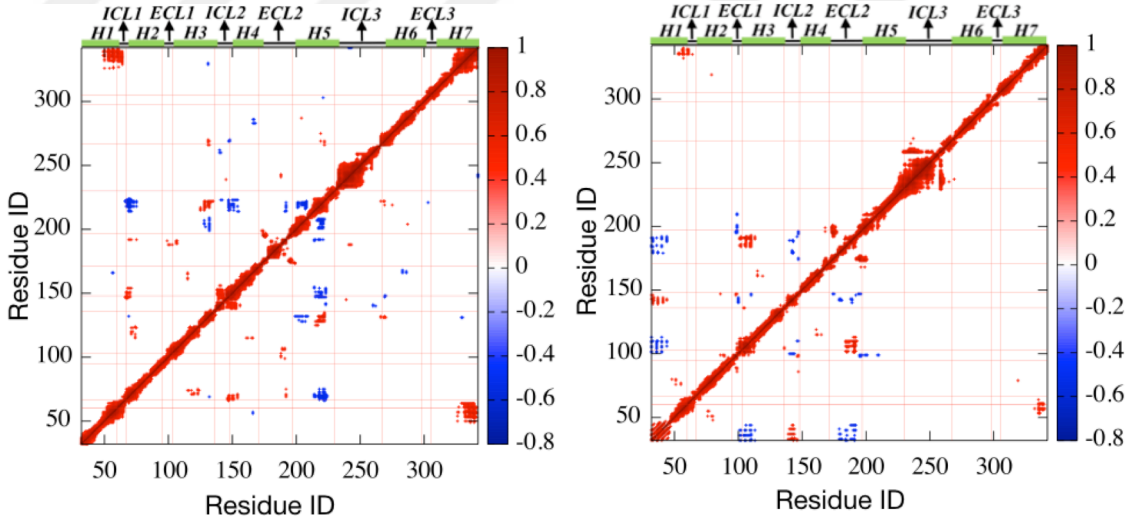


Figure 4.4 Stiffness heatmap for Phase I (a), Phase II (b), and Phase II – Phase I (c). Protein representation of net stiffness was shown in (d), red color indicated an increase and blue color indicated a decrease in stiffness, and mean net stiffness was plotted in (e). *(Adapted from Sogunmez N, Akten ED, 2019, J of Phys Chem B, 23(17), 3630. Copyright © 2019 ACS*)

4.1.3 The impact of ICL3 closure on residue pair cross-correlations

To understand the correlated motions of the subfragments in a receptor, we monitored the direction vector of atomic fluctuations throughout the molecular dynamics simulation. This fluctuation vector of each Cα atom from its average position was determined for each frame. Then, the dot products of atom *i* and atom *j* were calculated to obtain their average correlations using Equation 3.11. To emphasize the foremost important results, the strongest correlations (greater than +0.45 or lower than -0.45) were plotted only in Figures 4.5. In Phase I, the correlation between ICL3 with distant parts was observed as insignificant. However, it showed moderate correlations with a few distinct residues on loop regions, ICL2, and ECL2, as in Figure 4.5a. Upon its closure, the correlation in ICL3 was relatively increased with residues on H5 adjacent to ICL3, as illustrated in Figure 4.5b. Overall, correlations in Phase I, which were found in a few distinct areas on H2, H3, H4 and particularly on H5, shifted to H1, ECL2, ICL2, and ECL1 in Phase II. These correlated residues corresponded to regions displaying a significant amount of mobility increase.



(a) (b)

Figure 4.5 Residue-pair cross-correlations for Phase I (a) and Phase II (b). Only the correlations above/below 0.45 were plotted. Red color represents positive correlations, and blue color represents negative correlations. *(Adapted from Sogunmez N, Akten ED, 2019, J of Phys Chem B, 23(17), 3630. Copyright © 2019 ACS*)

Some other apparent changes were detected for residue-pairs on $H1_{Int}$ and $H8_{Int}$. These residues were positively correlated in both phases. However, the quantity of correlation

decreased in Phase II, which might be the end result of a stiffness increase in H8. Another important distinction between phases was the loss of the negatively correlated regions (colored in blue) in H5 and the remaining residues of the receptor (Figure 4.5a). This correlation loss might be the effect of the stiffness increase in $H5_{Int}$ when ICL3 was fully packed. Consequently, residue pair cross-correlations significantly associated with stiffness in those areas. Upon stiffening of a fragment in the receptor, the correlations in that part became constrained to adjacent residues as in ICL3 or totally disappeared as in H5. Alternatively, upon flexibility increase in any fragment, together with loop regions, did bring a correlated dynamic character.

4.1.4 Identification of highly correlated but spatially distant regions

The data of highly correlated but also spatially distant regions from each other was quite important in the understanding of the allosteric transitions since they represented the possible candidate sites communicating with each other through allosteric pathways. Contact maps were plotted in Figure 4.6, which were superimposed with the corresponding correlation maps shown in Figure 4.5, to show highly correlated $(C_{ij} > 0.45$ and $-0.45 < C_{ij})$ but spatially distant regions. Contact maps were calculated between heavy atoms with a distance threshold of 6 Å. Heavy atoms were preferred over Cα atoms in order not to miss the side-chain fluctuations.

In Figure 4.6a, it was shown that most of the positively correlated regions were also in close contact for Phase I. However, a few spatially distant residue-pairs exist with a moderate amount of correlations, which were between 0.46 and 0.54. *Glu*188, which was located on ECL2, correlated with three residues as *His*241, *Val*242, and *Gln*243 located on ICL3. However, *Asp*192, which was also located on ECL2, correlated with four residues located at the $H2_{Ext}$ as *Asn*69, *Tyr*70, *Phe*71, and *Leu*75. A residue pair, *Gln*197- *Ser*203, correlated positively with a value of 0.47. These residues, *Gln*197 and *Ser*203, on H5 were critical residues for ligand binding. Nevertheless, most of the negatively correlated and spatially distant regions, including the H5 region, played a critical role in agonist-driven conformational changes (Reynolds, Katritch, and Abagyan, 2009). Interestingly, $H5_{Ext}$ negatively correlated with its intracellular part.

Moreover, *Asp*192 on ECL2 was found to correlate with six residues at $H5_{Int}$ as *Ile*214, *Met*215, *Phe*217, *Val*218, *Tyr*219, and *Arg*221 in the range of - 0.54 to -0.45. However, none of these correlations were strong enough to mention a communication pathway.

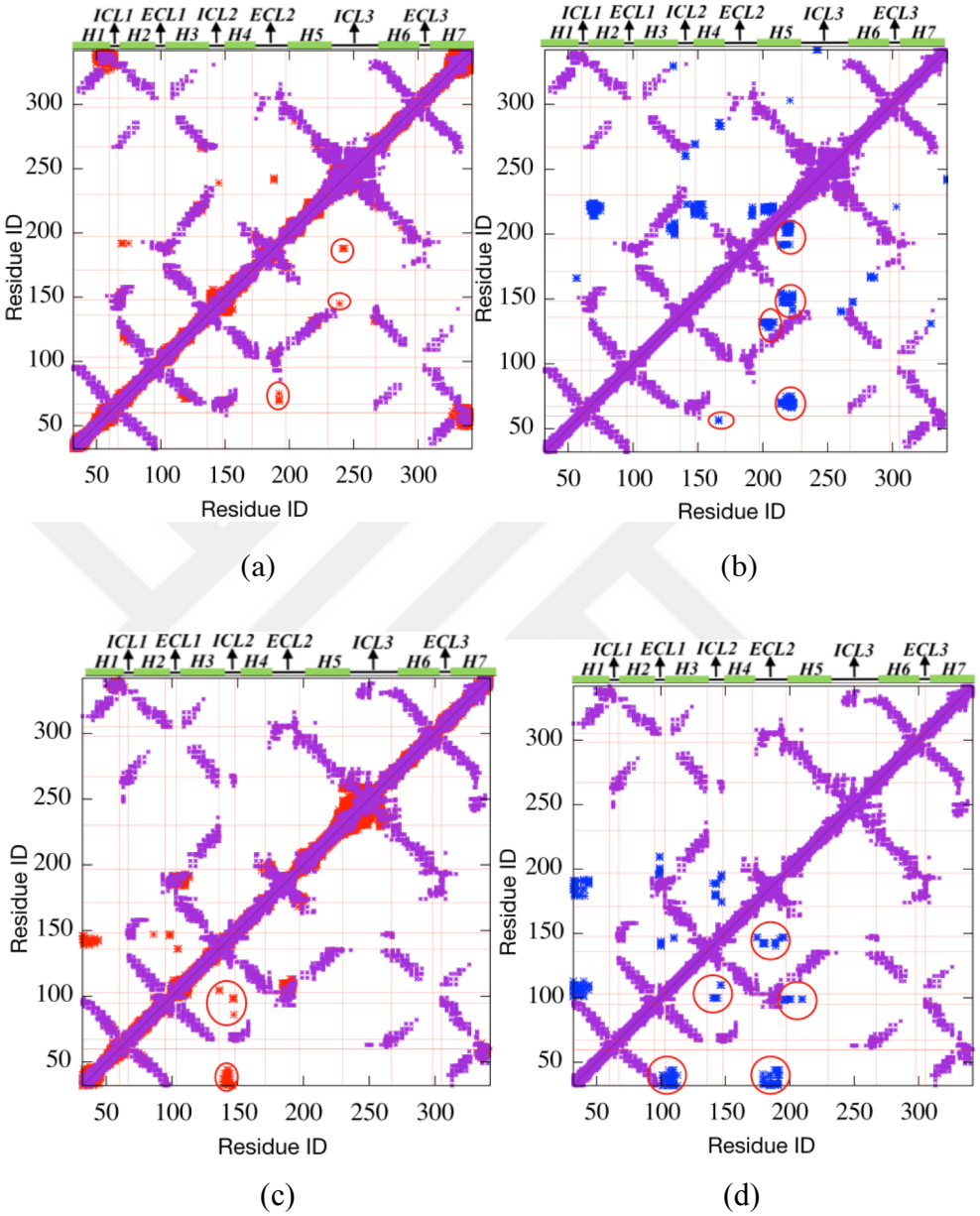


Figure 4.6 Contact map vs. residue-pair cross-correlations in detail. Purple dots are showing residue-pairs in contact for at least 75% of the simulation. (a-b) represent positive (red) and negative (blue) cross-correlations versus its contact map in Phase I and (c-d) represent positive (red) and negative (blue) cross-correlations versus its contact map in Phase II. *(Adapted from Sogunmez N, Akten ED, 2019, J of Phys Chem B, 23(17), 3630. Copyright © 2019 ACS)*

After the closure of ICL3, the majority of correlations, including H5, were shifted towards ECL1, ICL2, and ECL2. The regions that are spatially distant but still

correlated were generally locating on H1 and these three loop regions. $H1_{Ext}$ was positively correlated with ICL2. But, it was negatively correlated with the $H3_{Ext}$ and most of the residues on ECL2. Additionally, a few isolated residue-pairs display either positive or negative correlations, that were mostly located in ECL1, ICL2, and ECL2 (Figure 4.6c and 4.6d).

From Phase I to Phase II, all correlations existed between some transmembranes and H5 (see Figure 4.5a) shifted towards loops and the $H1_{Ext}$. It is important to observe that the stiffness of ICL3 was compensated by the correlation increase of smaller loop regions, instead of performing a random dynamic behavior. Besides, packed ICL3 started to fluctuate in harmony within itself and $H5_{Int}$. Since the strongest correlations between spatially distant regions were found between 0.5 and 0.6 in both phases, it might not be sufficient to suggest that these are potential allosteric sites. However, the significant decrease in the number of correlated regions in Phase II was noteworthy, which was apparently a result of the general increase in the receptor's stiffness.

Correlation is the dot product of two fluctuation vectors between two events that are observed exactly at the same time. However, two events can also be observed in perfect correlation at different time steps, i.e., with a distinct time delay. Thus, correlation is not adequate for understanding the time-delayed relationships of multiple events. For that purpose, when a delay time was introduced to the calculation to capture the correlated motions separated in both time and space, a new approach needs to be followed.

4.1.5 A decay of the autocorrelation in residue fluctuations

Before calculating time-delayed cross-correlations, the "time-delay" parameter needs to be characterized for each residue. This is possible by calculating the decay time of the autocorrelation for every residue by using Equation 3.13. The decay time, τ, is the time required for the autocorrelation to decay down to 1/e of its maximum initial value. This maximum initial value is adjusted to 1 when decay time τ=0. The decay time is an important parameter to be optimized for each residue and is going to be used to

compute time-delayed cross-correlations between residue *i* at time *t* and residue *j* at time *(t+τ)*. As shown in Figure 4.7, the decay times of residues in ECL1, ICL2 and ECL2, as well as $H1_{Ext}$, were significantly higher in Phase II than those in Phase I. Also, the decay times of residues in ICL3 and residues on H5 and H6 adjacent to ICL3 decreased significantly in Phase II. After ICL3 closure, regions displaying increased decay times were also observed with an increase both in their mobility and their correlations. Clearly, these fluctuations were not random and the memory of its initial fluctuations was hold for over 100 ns as in some loop regions.

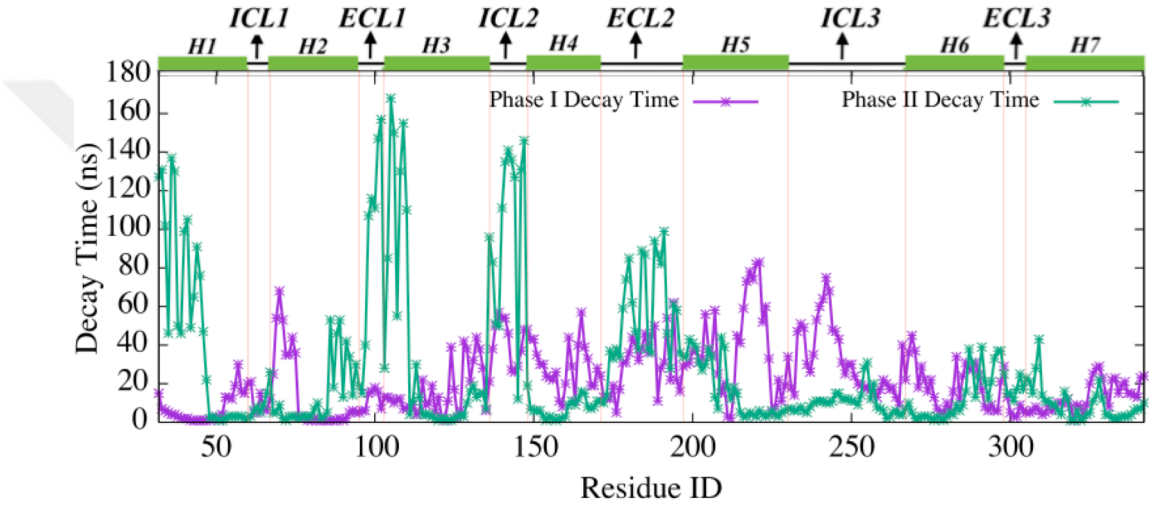


Figure 4.7 Decay of autocorrelations of residues in Phase I (purple) and Phase II (green). *(Adapted from Sogunmez N, Akten ED, 2019, J of Phys Chem B, 23(17), 3630. Copyright © 2019 ACS*)

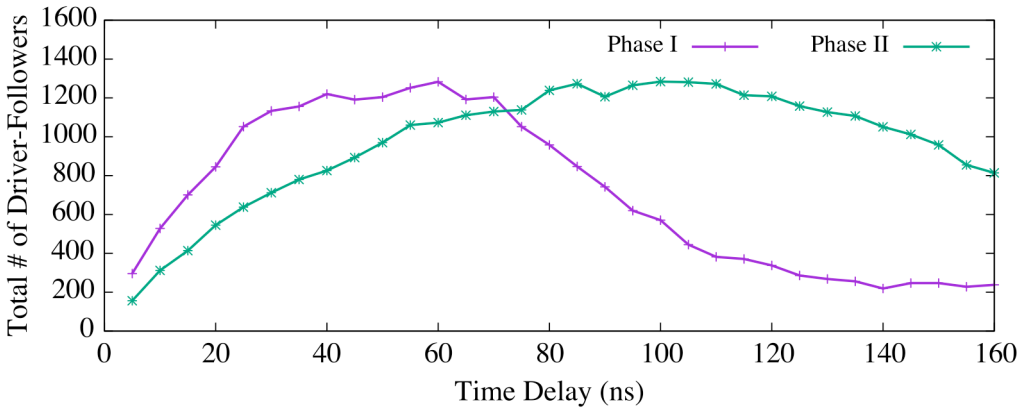
4.1.6 Conditional time-delayed cross-correlations (CTC) analysis and the driver-follower relationship between residue-pairs

Capturing the existence of a time-delayed correlation between residues gave us the information related to the direction of the information flow. Thus, strong time-delayed correlations between residues *i* and *j* separated by a delay time τ were denoted as $C_{ij}(\tau)$ and were computed using Equation 3.14. Accordingly, if $C_{ij}(\tau) > C_{ji}(\tau)$, then the past fluctuations in residue *i* drive the future fluctuations in residue *j*, and vice versa.

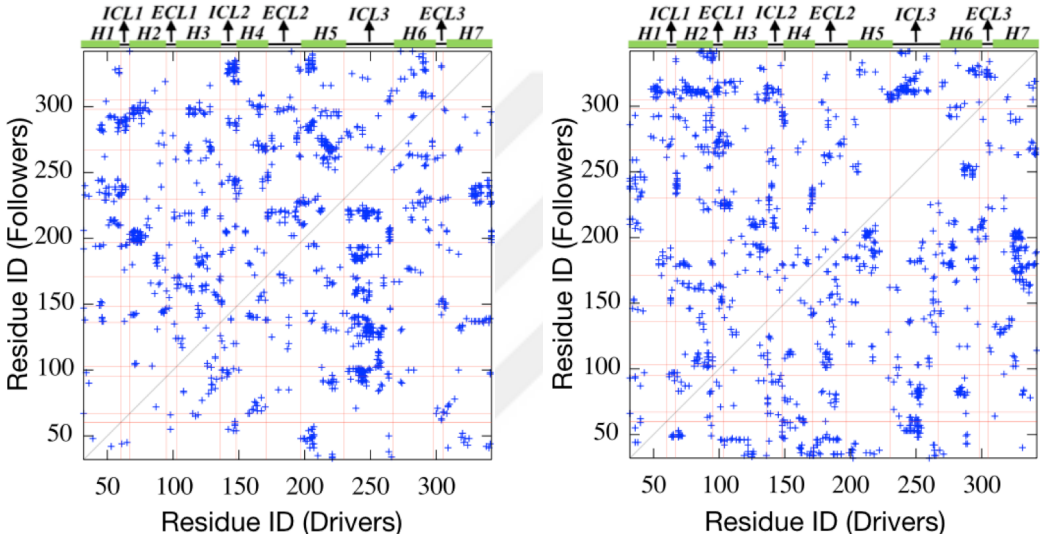
Different delay times were adopted between 5 ns and τ_{max} with an increment of 5 ns to discover τ_{max} for each phase. Interestingly, τ_{max} was detected as 80 ns for Phase I and

160 ns for Phase II, based on maximum values of the autocorrelation decay times given in Figure 4.7. In order to focus on the distinct directionalities, residue-pairs with $0 < C_{ij}(\tau_{max}) \leq 0.25$ and $C_{ji}(\tau_{max}) \geq 0.75$, were extracted. As given in Figure 4.8a, the total number of residue-pairs meeting abovementioned criteria was determined for all possible time delays. Our inactive states have 311 Cα atoms. Out of $311 * 310 = 96410$ possible residue-pairs, the τ values were 60 ns and 85 ns in Phase I and Phase II, respectively, for which a maximum number of driver-follower pairs was reached. Interestingly, the maximum number of pairs for both phases was 1283 for Phase I and 1284 in Phase II. The locations of these residue-pairs were given in Figures 4.8b and 4.8c.

Table 4.1 lists the highest amounts of driver-follower pairs with corresponding time delays given in parentheses. In Phase I, H8 is a moderate driver for ICL3 at $\tau = 5\ ns$ with 11 residue-pairs. However, it approached its maximum driving capacity through 82 different residue-pairs at $\tau = 25\ ns$ and then became persistent. Even at $\tau = 75\ ns$, which is a high delay time, it continued as a driver for ICL3 (Figure 4.9a). As its dual charactistics, ICL3 drove H3 and H5, as well, which are known as highly critical transmembrane helices in terms of orthosteric binding. Moreover, H6 was driven by both H2 and H5 in Phase I, and additionally, H2 was the driver of H5 via 44 residue-pairs at $\tau = 60\ ns$.



(a)



(b) (c)

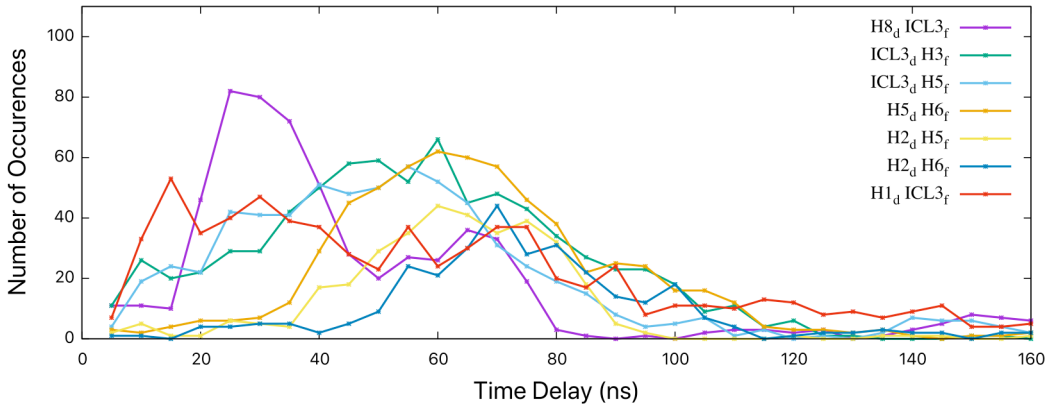
Figure 4.8 Conditional time-delayed cross-correlations. The change in the total number of driver-followers in time delay for Phase I (purple) and for Phase II (green) (a). Phase I driver-follower plot at a delay time of 60ns (b) and Phase II driver-follower plot at a delay time of 85ns (c). *(Adapted from Sogunmez N, Akten ED, 2019, J of Phys Chem B, 23(17), 3630. Copyright © 2019 ACS*)

Phase II showed a divergent profile compared to Phase I for driver-follower relationships. ICL3 strongly drove H7 and H1. Also, it drove H2 and H8 to a greater extent. Furthermore, ECL2 became a follower for both H5 and H6. Also, H2 was the driver previously and remained driver in Phase II, as well. But in Phase II, it drove H7. Also, H6 became an important driver region in Phase II by driving ICL3 and ECL2. Selected driver-follower residue-pairs, together with their correlation plots, were shown in Figure 4.10.

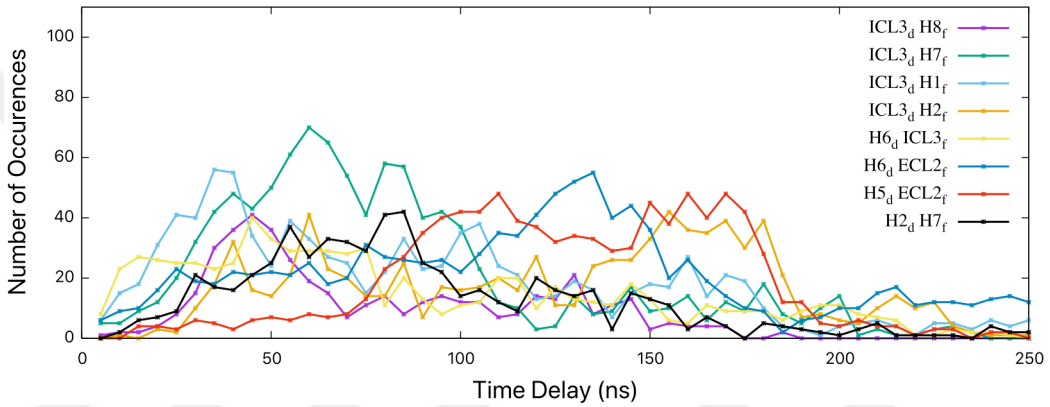
Table 4.1 Regions with the highest number of driver-follower residues according to their τ values.

PHASE I			PHASE II		
Driver	Follower	Max # of Occurrences (Time delay,τ)	Driver	Follower	Max # of Occurrences (Time delay,τ):
H8	**ICL3**	82 (25 ns)	**ICL3**	H7	70 (60 ns)
ICL3*	H3	66 (60 ns)	**ICL3**	H1	56 (35 ns)
H5	H6	62 (60 ns)	H6	**ECL2**	55 (135 ns)
ICL3	H5	57 (55 ns)	H5	**ECL2**	48 (110 ns)
H1	**ICL3**	53 (15 ns)	H2	H7	42 (85 ns)
H2	H6	44 (70 ns)	**ICL3**	H2	42 (155 ns)
H2	H5	44 (60 ns)	**ICL3**	H8	41 (45 ns)
			H6	**ICL3**	40 (45 ns)

* Intra- and extracellular loop regions were designated in bold.



(a)



(b)

Figure 4.9 Total number of driver-followers for different domains in Phase I (a) and Phase II (b). *(Adapted from Sogunmez N, Akten ED, 2019, J of Phys Chem B, 23(17), 3630. Copyright © 2019 ACS*)

An important study by Liu and his coworkers (Liu *et al*., 2017) revealed an allosteric site enclosed by H1, H2, H6, H7, H8, and ICL1 where a bound modulator called *Cmpd-15PA* was found to decrease the activity of the receptor. This NAM-binding site overlaped with our results of strong *driver-follower* interactions, mostly observed in intracellular parts of the helices H2, H5, H6, H7, and H8. ICL3 was a critical part occupying G protein-binding site, and playing a crucial role in receptor activation. Our CTC analysis showed that this region highly controlled two essential transmembranes; a structural and functional hub, H3, and a bridge for allosteric communication between extra- and intracellular parts of the receptor, H5. MD simulation by Liu and his coworkers also suggested that *Cmpd-15PA* binding stabilizes a more inward conformation of H6, which have shown that this motion was tightly coupled with and driven by H5 and H2 in Phase I.

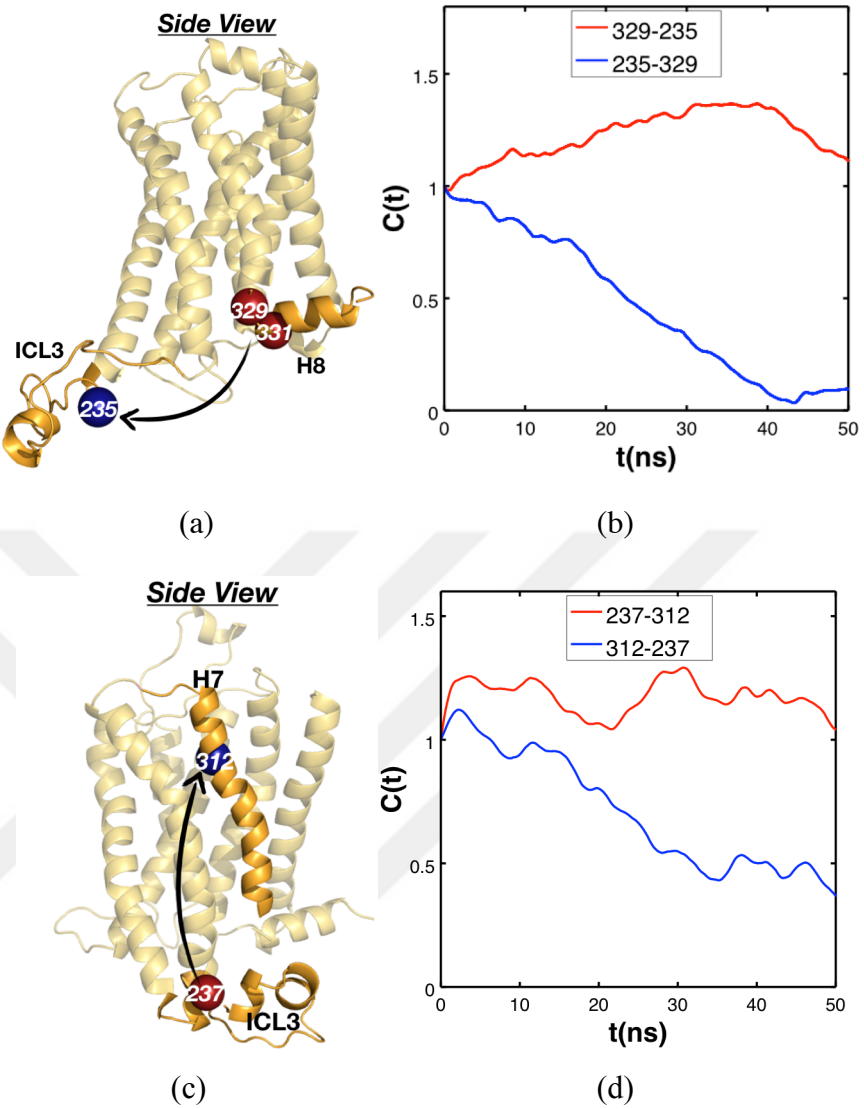


Figure 4.10 Representation of driver followers in Phase I (a) and Phase II (b). Time-delayed correlations of the residue-pairs in a (c). Time-delayed correlations of the residue-pairs in c (d). *(Adapted from Sogunmez N, Akten ED, 2019, J of Phys Chem B, 23(17), 3630. Copyright © 2019 ACS*)

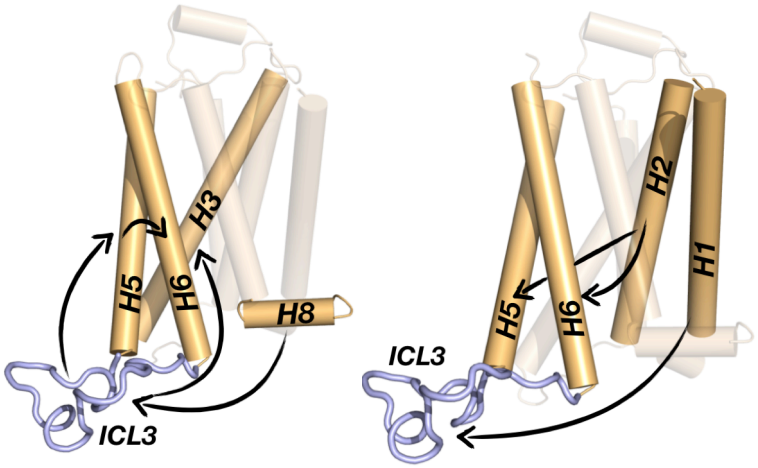
Upon ICL3 closure, the *driver-follower* interaction between H8 and ICL3 was reversed compared to Phase I. This time, future fluctuations of H8 was driven by previous fluctuations of ICL3, especially for $\tau = 45\ ns$ with 41 residue-pairs. Nevertheless, the intensity of this relation was stronger in Phase I, and was temporary in Phase II, where it faded away rapidly when τ values increased. Additionally, an interesting substructure in Phase II was ECL2, which was influenced by H5 and H6, which were two critical transmembranes in receptor activation and inactivation. Phase II was observed when the conformational shift of ICL3 through the core of the receptor occurred. This generated a

completely different profile of *driver-follower* couplings, especially for H5 and H6. Additionally, ECL2 is considered as a gate to the extracellular site, which controls the diffusions of small ligands inside the receptor (Dror *et al*., 2011). ECL2 became under the influence of both H5 and H6. Moreover, H6 also started to control ICL3. Therefore, when the receptor became *highly inactive*, the control over ICL3 by the nearby H8 weakened and shifted to H6 with moderate intensity, and also ECL2 allosterically coupled with H5 and H6.

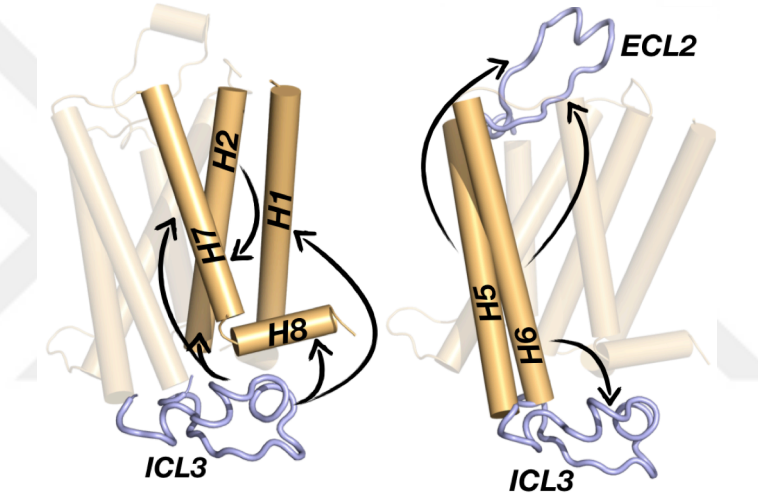
In Phase I, from the highest number of *driver-follower* pairs given as in Table 4.1, a possible allosteric communication pathway can be suggested as "H8 → ICL3 → H5 → H6" where the tip of the arrow indicates *driver→ follower* pairs (See Fig. 4.11a). The interesting fact in this suggested pathway that H8 had the highest driving capability in the receptor, which was followed by ICL3. Moreover, the large fluctuations in ICL3 directly controlled future fluctuations of $H5_{Int}$ between 25 ns and 65 ns delay time.

To suggest a communication pathway for Phase II was not easy. Upon tightly packing and its increased stiffness, the communication between ICL3 and H5 was lost and it started to control the fluctuations of H7 and H1. Furthermore, H5 and H6 drove the fluctuations of ECL2 at the extracellular side (See Fig. 4.11b). The mobility of ICL3 might indirectly affect the fluctuations of ECL2, where the allosteric communication between intracellular and extracellular regions of the receptor became detectable.

Therefore, in Phase II, together with intracellular binding sites, the extracellular part of the receptor was also regulated by other structural elements in the receptor. In β_2AR, the allosteric coupling between the intra- and the extracellular parts weakens upon activation of the receptor (Vaidehi and Bhattacharya, 2016). Since Phase II represented a complete inactive and a novel state of the receptor, the allosteric communication increase between distal parts was in great harmony with these general expectations.



(a)



(b)

Figure 4.11 Representation of major driver follower regions in Phase I (a) and Phase II (b). Arrows indicate driver and follower directions. *(Adapted from Sogunmez N, Akten ED, 2019, J of Phys Chem B, 23(17), 3630. Copyright © 2019 ACS*)

Finally, ICL3 highly drives H3 and H5 in Phase I (Table 4.1). H3 is known to play a central role in receptor activity and has evolutionarily conserved residues to form the H2-H3-H4 pocket. This pocket formation promoted the receptor stability, while the full activation of the receptor involves a more global motion of helices H5, H6, and H7. In a study conducted by Chelikani and his coworkers (Chelikani *et al*., 2007), the mutant forms of the receptor were obtained by replacing residues *Ala*128 on H3, *Ser*161, and *Ser*165 on H4 by either *Leu* or *Val*. As a result, these mutant receptors showed 3- to 4-fold less or no agonist-stimulated cAMP production, indicating a specific role of these residues in receptor activation.

Surprisingly, our calculations captured these three amino acids, which were intensely driven by ICL3. Our results also showed that the fluctuation of ICL3 drives the future fluctuations of *Met*215 to *Glu*225 on $H5_{Int}$. Finally, the same 11 amino acid long fragment was also adjacent to the conserved residue, *Ala*128.

5. MUTUAL INFORMATION AND TRANSFER ENTROPY UNDERLINE THE DIFFERENCES BETWEEN INACTIVE AND ACTIVE STATES IN COARSE GRAINED CALCULATIONS

In this chapter, initially, two inactive states and later on, two active states of β_2AR were analyzed using mutual information and transfer entropy approaches in order to identify potential allosteric sites and communication pathways. Two inactive states designated as Phase I, and Phase II distinguish from each other by the major conformational shift observed in ICL3. Two active states are designated as *intermediate* and *active-Gp* to make them distinguishable by the reader. The effect of residue type on mutual information was investigated in order to unravel shared characteristics of residues independent of the conformational state. Accordingly, polar residues contribute more to shared information than hydrophobic ones. Transfer entropy calculations were then used to identify the directionality of the information flow. Based on the in-depth analysis of driver-follower residue-pairs, a strong coupling between extra- and intracellular parts of the receptor was identified in the highly inactive state, and furthermore, the extracellular domains drove the future fluctuations of the intracellular domains.

5.1 Results and Discussion

5.1.1 MD simulations revealed distinct inactive states of β_2AR

In a previous MD run, Ozcan and her coworkers determined a unique inactive state of human β_2AR, conducted in fully atomistic detail prior to this study (Ozcan *et al.,*2013). The apo form inactive crystal structure (PDB ID: 2RH1) was used as the initial state, where missing ICL3 was atomistically modeled. The initial 50 ns was accepted as equilibration phase, and until around 450 ns, the ICL3 remained open and G protein-binding cavity was accessible. However, after 450 ns the loop started to pack itself

through the core of the receptor while it started to lose its mobility. And finally, after 750 the whole receptor shifted towards a distinct conformational state presented as "*highly inactive*". In this newly discovered inactive state, the intracellular G-protein binding cavity was completely occupied by the ICL3 region as it was fully packed under the core of the receptor and increased its stiffness. Together with this intracellular motion, the extracellular orthosteric ligand-binding site was increased in size considerably (Figure 5.1).

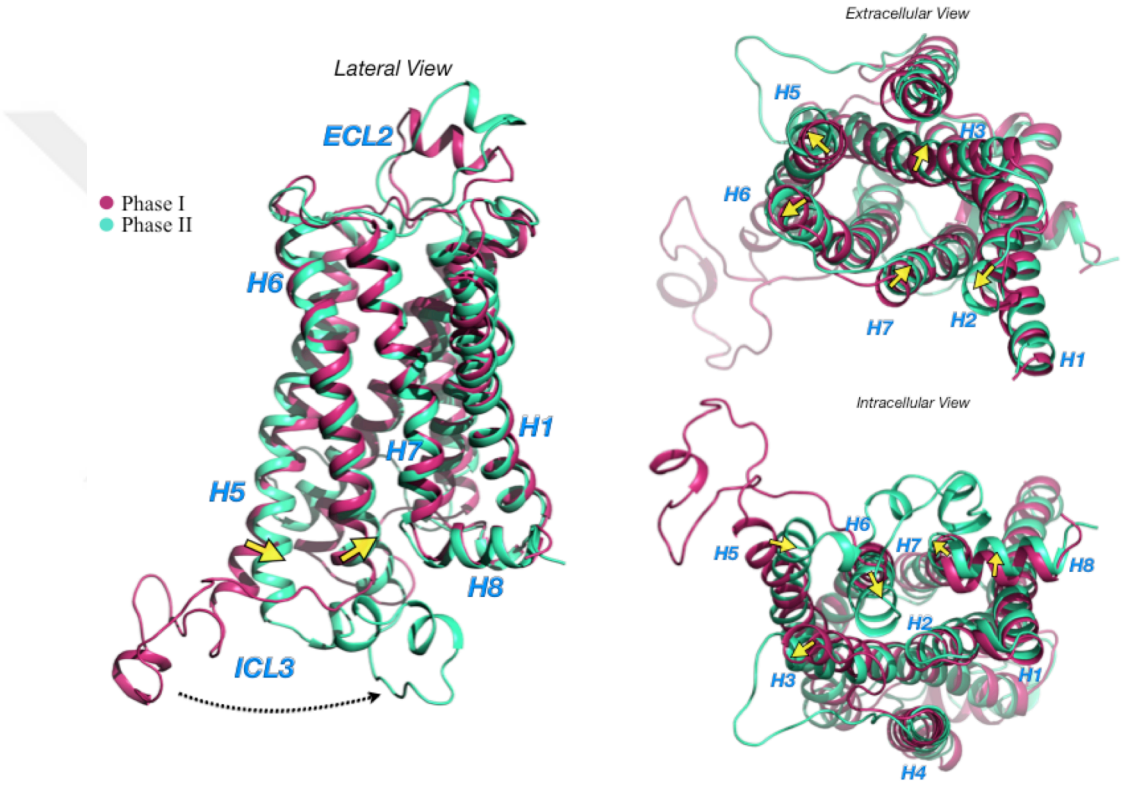


Figure 5.1 Overlay of Phase I (purple) and Phase II (green) to demonstrate the conformational changes between them. Lateral (a) Extracellular (b) and Intracellular (c) view of the phases (*Sogunmez and Akten, 2020*).

5.1.1.1 Mutual information (MI) increased in the inactive state, and polar residues share more information compared to hydrophobic residues

In this study, MI is defined as the mutual dependence between positional fluctuations of Cα atoms of two residues. Spatially distant but potentially allosteric residue-pairs were also captured with MI as unraveled in Figure 5.2 when mutual information was

superimposed with the contact map. In Phase I, where ICL3 remained open for 400 ns, ICL2, ICL3, and ECL2 comprise of potential allosteric sites. In Figure 5.2a, MI values between 1 and 1.5 correspond to distal pairs located especially on the loop regions, such as ICL2-ICL3, ICL2-ECL2, and ICL3-ECL2. The residues locating on H4 and H5, especially those adjacent to loops, also have high MI values, indicating the strong effect of loops on transmembrane helices.

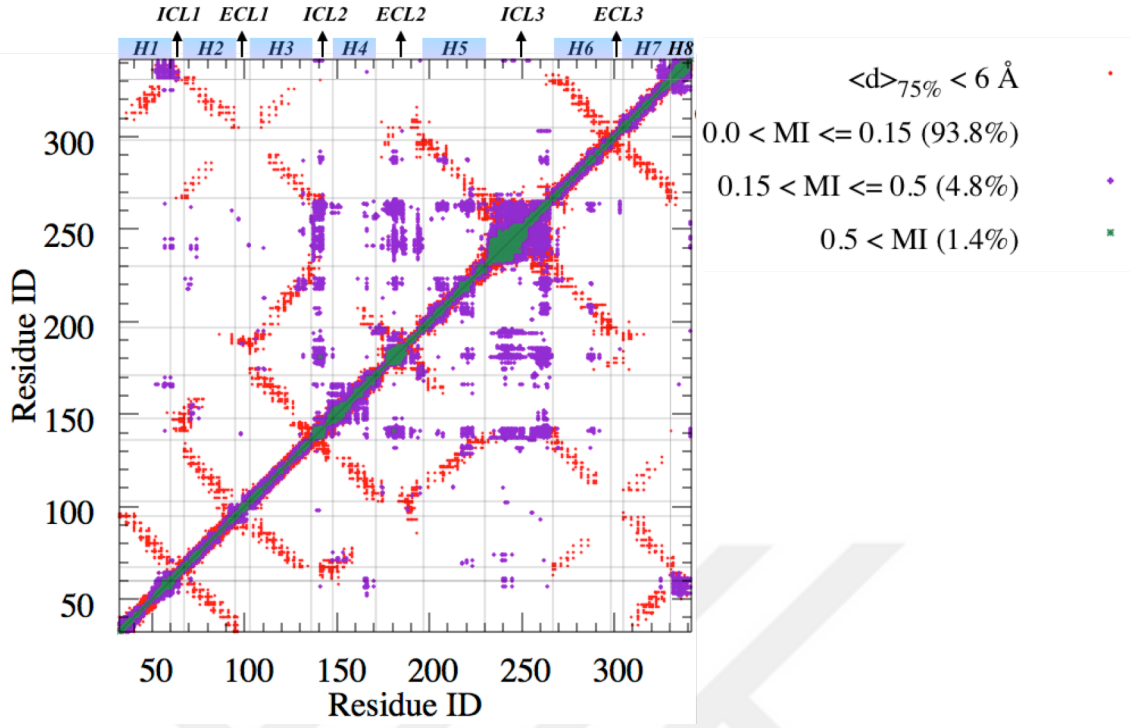
Through closure of ICL3, MI values of ICL2-ECL2 and ICL3-ECL2 still observed with increased intensity. However, the remaining intracellular loop, ICL1, also started to communicate with ECL2. In this phase $H1_{Ext}$ appeared to communicate especially with loops, such as ICL2 and ECL2, with MI values between 0.15 and 0.5. Therefore, a major conformational change in ICL3, thus the receptor, resulted in a significant regional shift in mutual information, where H1, ICL1, ICL2, and ECL2 came forward.

The highest MI intensity with a value above 0.5 was observed for residues spatially in contact by the thick diagonal line in Figure 5.2. In Figures 5.2a and 5.2b, the MI values between 0.0 and 0.15 were accepted as the lowest dependence and colored in white. This lowest dependence was observed in nearly 93.8% of the residue-pairs in Phase I and 95.8% in Phase II. The mutual information between 0.15 and 0.5 was illustrated with magenta dots and appeared as a moderate degree of communication. The number of communicating residues almost became half in quantity, from 4.8% to 2.9% upon shifting to Phase II. Mutual information with a high amount was represented with green and blue dots for values >0.5. These high degrees of correspondence have similar quantities in phases with 1.4% and 1.3% in Phase I and Phase II, respectively. Overall, the value of MI was higher in Phase I compared to Phase II. The average MI value of each residue with the rest of the receptor was calculated for each phase as $\langle M_i \rangle = \sum_{j \neq i}^{N} (MI)_{ij} / (N-1)$, where N is the total number of residues (Figure 5.2c). Then, $\langle MI_i \rangle_{Phase\ II} - \langle MI_i \rangle_{Phase\ I}$ was used to reveal the difference in the average MI upon ICL3 closure and plotted as in Figure 5.2d, where data was represented using a color gradient as in Figure 5.2e. Obviously, shared information in loops dominates transmembrane helices in both phases, and also except $H1_{Ext}$ and ECL2, the mutual correspondence significantly decreases going from Phase I and Phase II. Interestingly,

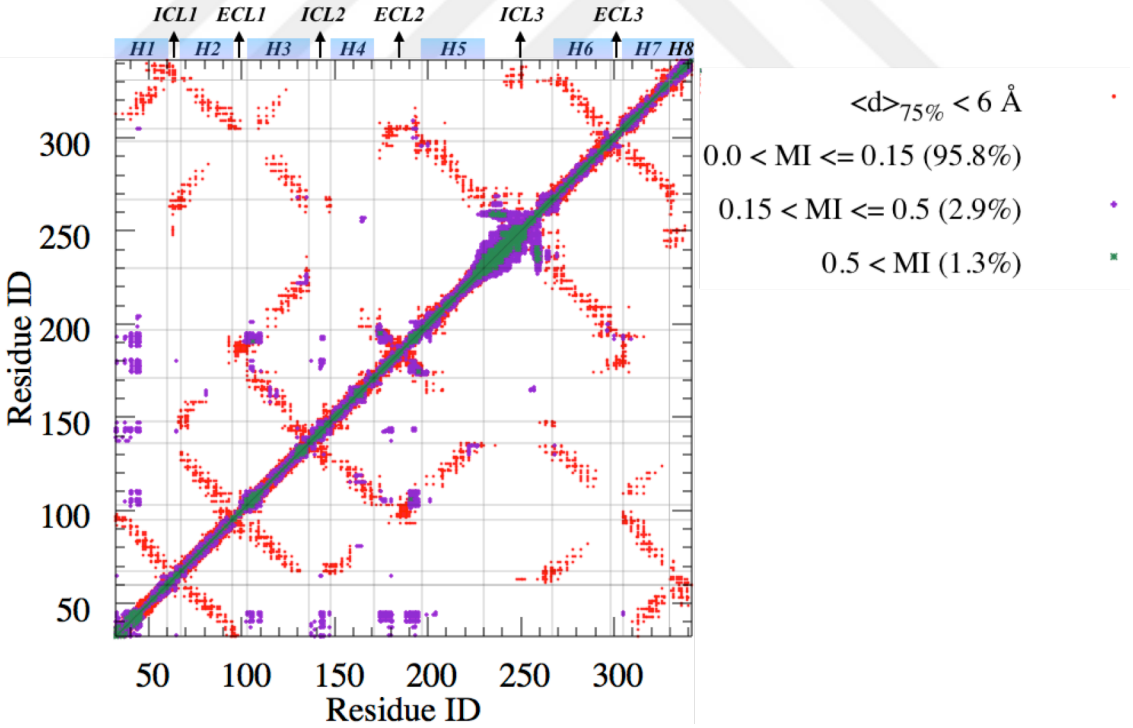
the intracellular extension of H7 denoted as H8, reported to have a role in receptor activation (Zhang *et al.,*2015).

Also, this small helical structure was driving the future fluctuations of the residues in the receptor (Sogunmez and Akten, 2019), and its communication with ICL1 vanished through Phase II. Notably, MI values were increased in ICL3, but information sharing was restricted between ICL3-$H5_{Int}$ and ICL3-$H6_{Int}$.

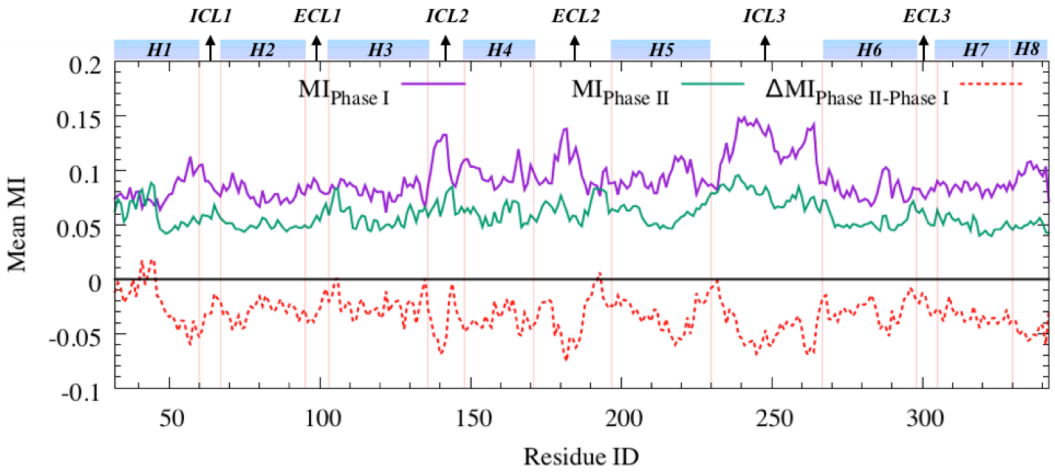
In this chapter, MI values were computed based on considering only Cα fluctuations. Side-chain angular motions were not included in the calculations. However, the effect of residue types based on their side-chains was also investigated on global Cα atoms. To understand this effect, MI values per residues were calculated, and then these residues were split into two groups as *hydrophobic* and *polar*. The hydrophobic residues include *Ala, Gly, Pro, Cys, Val, Ile, Leu, Phe, Trp, Tyr,* and *Met*, and polar residues include *Ser, Thr, Asn, Asp, His, Gln, Glu, Arg,* and *Lys.* Since MI captures native contacts as well as residues with allosteric interaction (Figures 5.2a and 5.2b), the values also further categorized based on the degree of separation of the two residues in the primary sequence, as *proximal* if 1 to 4 positions apart and otherwise as *distal*.


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Residue ID
50
100
150
200
250
300
<d>75% < 6 Å
0.0 < MI <= 0.15 (93.8%)
0.15 < MI <= 0.5 (4.8%)
0.5 < MI (1.4%)

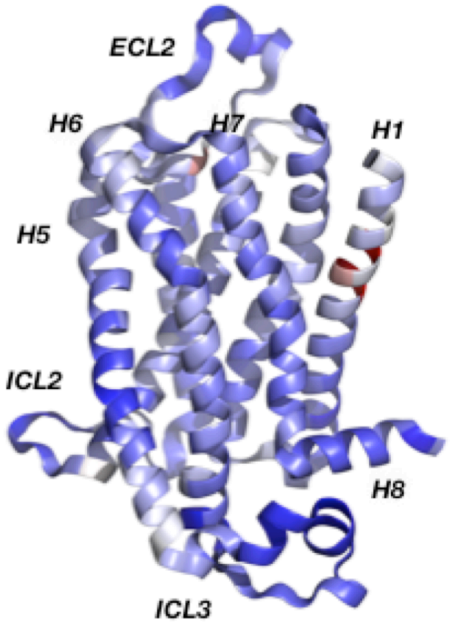
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Residue ID
50
100
150
200
250
300
<d>75% < 6 Å
0.0 < MI <= 0.15 (95.8%)
0.15 < MI <= 0.5 (2.9%)
0.5 < MI (1.3%)

(b)



(c)



(d)

Figure 5.2 Mutual information (MI) and contact map for Phase I, and Phase II (a-b). The contact map was colored in red, and mutual information was colored in purple. Mean MI values in both phases and their difference (Phase II – Phase I) (c) and protein representation of the mean MI differences (d). Positive values were shown in red and negative values were shown in blue *(Sogunmez and Akten,2020)*.

Surprisingly, polar residues had more contribution to MI compared to hydrophobic residues in both distal and proximal pairs. This difference between polar and hydrophobic residues was more divergent in Phase II proximal and distal pairs, as shown in Figure 5.3. As listed in Table 5.1, the average MI values for proximal polar

residues were 0.46 ± 0.07 and 0.41 ± 0.07 in Phase I and Phase II, respectively. On the other hand, the average MI values of proximal hydrophobic residues were determined as 0.37 ± 0.02 and 0.32 ± 0.03. Similarly, average MI values in distal polar residues with 0.087 ± 0.005 in Phase I, and 0.053 ± 0.002 in Phase II were observed to be higher compared to distal hydrophobic residues with 0.081 ± 0.004 in Phase I and 0.050 ± 0.002 in Phase II.

Further analysis based on the mutual information of both residue types in pairs was performed to understand the pairwise characteristics of the residues. Interestingly, polar-polar pairs have significantly higher average MI values compared to polar-hydrophobic and hydrophobic-hydrophobic pairs in both proximal and distal pairs. On the other hand, Phase I and Phase II represented the lowest average MI values in hydrophobic-hydrophobic pairs, as given in Table 5.2. In general, even in Cα-only fluctuations were analyzed, the contribution of polar residues was observed to be inevitable in β_2AR dynamics.

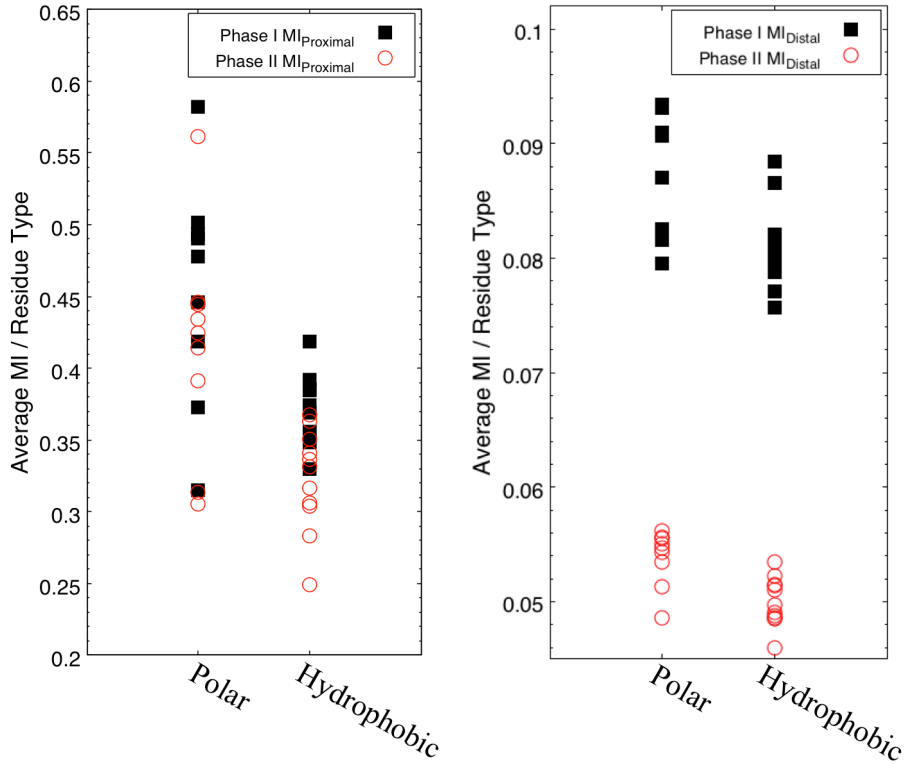
Table 5. 1 Polar and hydrophobic residues average MI value comparison based on their proximity.

	Type	Proximity	Mean Mutual Information	# of Residues
Phase I	Polar	Proximal	0.46 ± 0.07	912
		Distal	0.087 ± 0.005	34428
	Hydrophobic	Proximal	0.37 ± 0.02	1556
		Distal	0.081 ± 0.004	59514
Phase II	Polar	Proximal	0.41 ± 0.07	912
		Distal	0.053 ± 0.002	34428
	Hydrophobic	Proximal	0.32 ± 0.03	1556
		Distal	0.050 ± 0.002	59514

Table 5. 2 Residue-pair type average MI comparison based on their proximities.

	Type	Proximity	Mean Mutual Information	# of Residue-pairs
Phase I	Polar-Polar	Proximal	0.47 ± 0.23	370
		Distal	0.094 ± 0.01	12512
	Polar-Hydrophobic	Proximal	0.42 ± 0.21	1084
		Distal	0.084 ± 0.007	43832
	Hydrophobic-Hydrophobic	Proximal	0.33 ± 0.13	1014
		Distal	0.079 ± 0.005	37598
Phase II	Polar-Polar	Proximal	0.44 ± 0.24	370
		Distal	0.059 ± 0.007	12512
	Polar-Hydrophobic	Proximal	0.37 ± 0.17	1084
		Distal	0.051 ± 0.003	43832
	Hydrophobic-Hydrophobic	Proximal	0.29 ± 0.13	1014
		Distal	0.049 ± 0.003	37598

As it was observed in Figure 5.3, polar values in Phase I and Phase II were higher compared to hydrophobic values. The comparison in proximal and distal residues display the dominant effect of polar residues in average MI, and these results were verified statistically by using Kruskal-Wallis (Kruskal and Wallis, 1952) and Dunn's (Dunn, 1964) nonparametric statistical tests as given in Appendices H and I. The significance value for Kruskal-Wallis was selected as $p<0.05$, and for Dunn's test $p<0.01$.

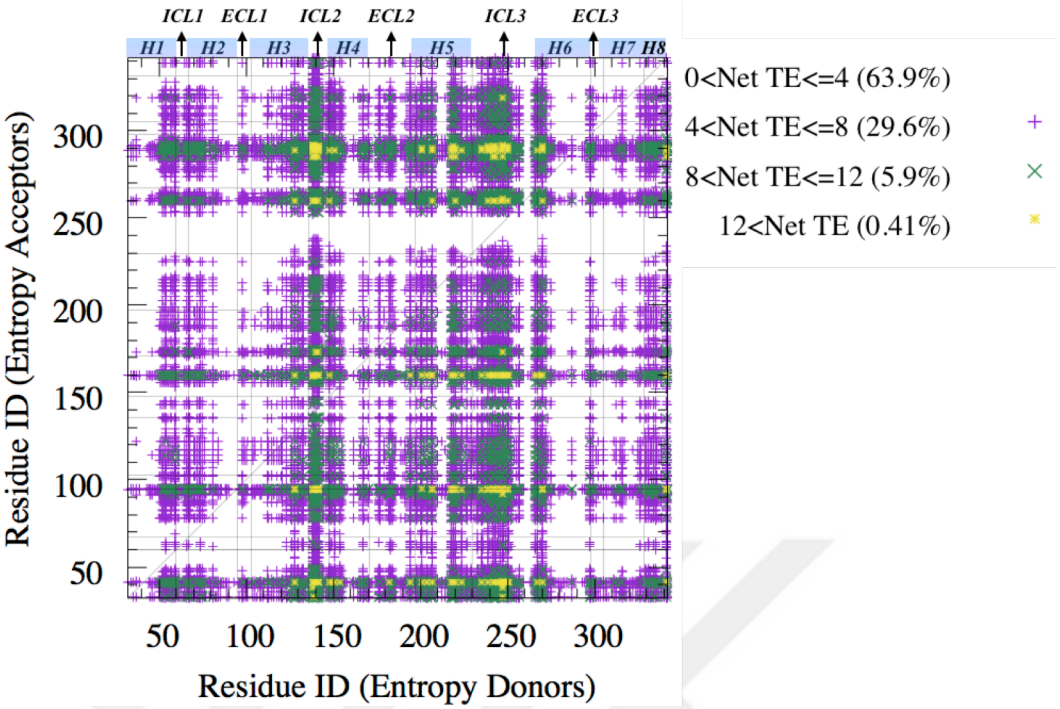


(a) (b)

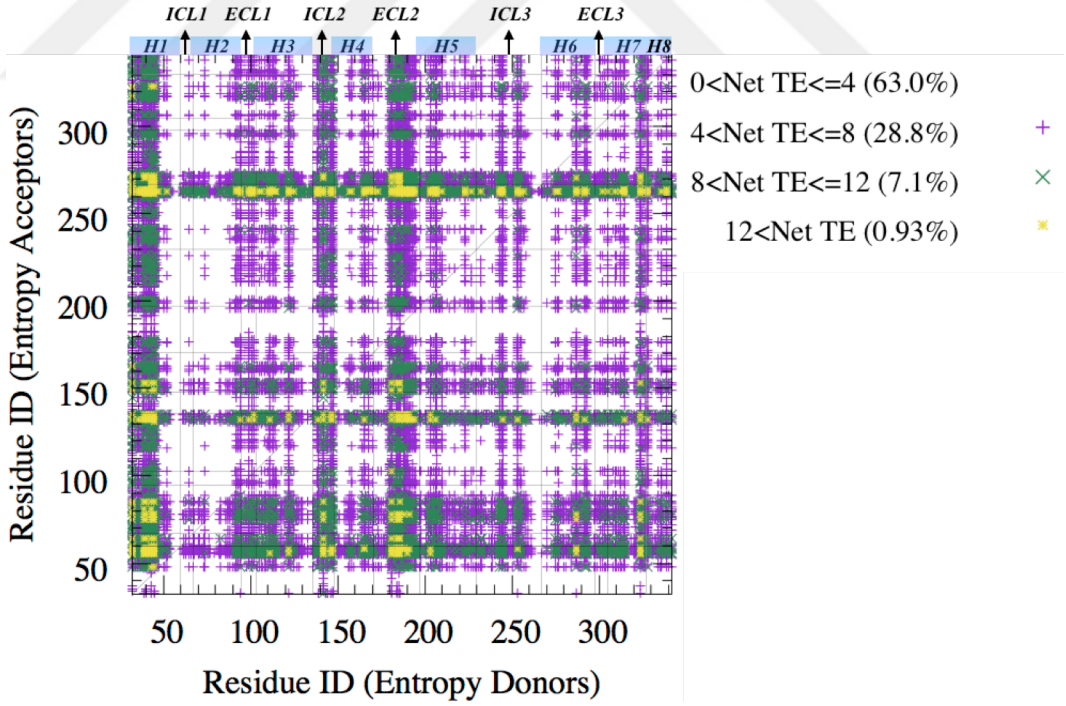
Figure 5.3 Proximal (within 5 amino acids) (a) and distal (above 5 amino acids) (b) residues average MI comparison based on polarity. Phase I was colored with black rectangles and Phase II with red circles *(Sogunmez and Akten,2020).*

5.1.1.2 Entropic activity adopted in the overall receptor through the transition from Phase I to Phase II

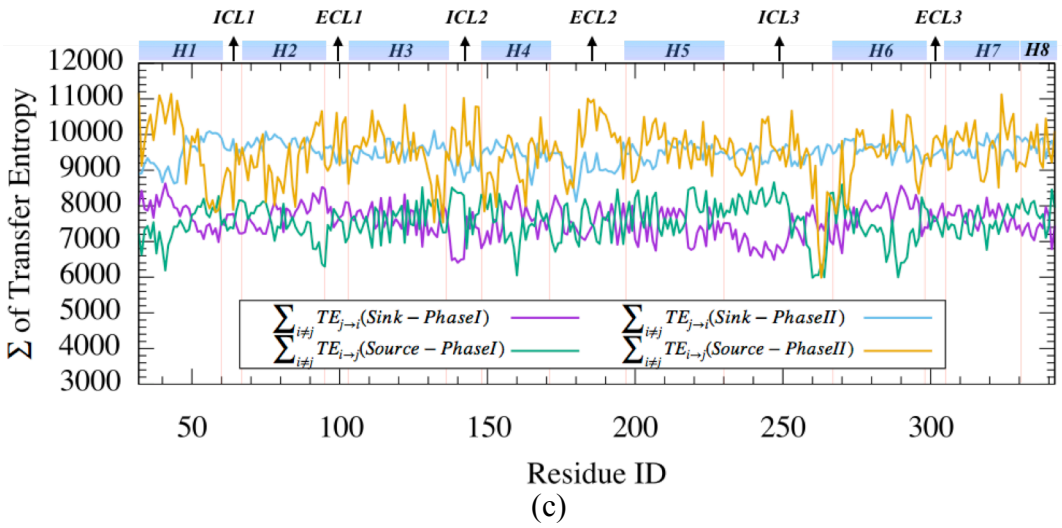
In both phases, the total amount of transfer entropy from source residues *i,* and sink residues *j* to the rest of the receptor was determined using Equation 3.26. As illustrated in Figure 5.4a, both sources' and sinks' total transfer entropy values in Phase II were higher in Phase I. This finding clearly indicates the entropy increases upon ICL3 closure, thus positively affecting the total information transfer.


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)
50
100
150
200
250
300
0<Net TE<=4 (63.9%)
4<Net TE<=8 (29.6%)
8<Net TE<=12 (5.9%)
12<Net TE (0.41%)

(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)
50
100
150
200
250
300
0<Net TE<=4 (63.0%)
4<Net TE<=8 (28.8%)
8<Net TE<=12 (7.1%)
12<Net TE (0.93%)

(b)



(c)

Figure 5.4 Pairwise Net TE values for Phase I (a) and for Phase II (b). From white to yellow, the net TE values increases. Sum of the sink and source TE values for both phases was shown (c) *(Sogunmez and Akten, 2020)*.

Net entropy transfer given in Figures 5.4b and 5.4c indicate a shift from intracellular parts of the receptor towards extracellular parts through the closure of ICL3. Here, abscissae denote entropy donors, and ordinates denote entropy acceptors, where the values were given in white ($0<NetTE\leq4$), magenta ($4<NetTE\leq8$), violet ($8<NetTE\leq12$) and yellow ($NetTE>12$). The percentages of residue-pairs with net TE values were given in pharentheses. The net TE values above 12 that were colored in yellow in Figures 5.4b and 5.4c only comprise 0.41% of overall net TE values in Phase I and 0.93% in Phase II. Additionally, it was observed that ~60% of net TE values were between 0 and 4.

5.1.1.3 Intracellular parts of H1, H3, H4, H6 and the adjacent segment of ICL3 becoming under the strict control of the rest of the protein in Phase II

Residues between *Arg*260 to *Phe*264 in ICL3 were observed as the strongest entropy acceptors in both phases. This short segment of ICL3 includes polar residues as *Arg, Ser* or *Lys*, and accepted information from all remaining residues, including even ICL3 itself. Through Phase II, residues adjacent to this short segment, such as *Glu*268, *His*269, *Ala*271, and *Leu*272, also became entropy acceptors (Figure 5.5a). Notably,

these residues on $H6_{Int}$ became entropy acceptors due to a significant amount of decrease in its mobility.

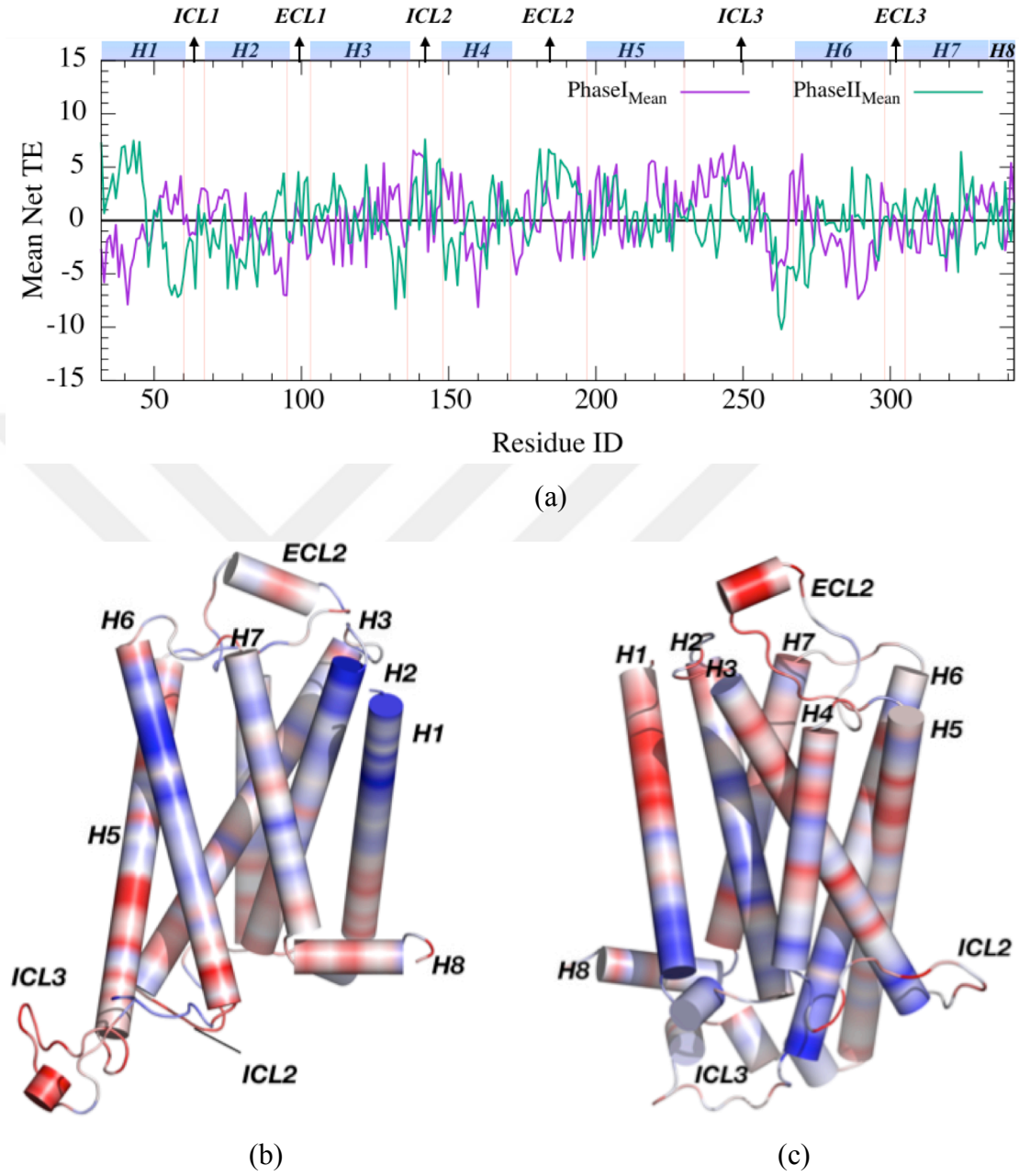


Figure 5.5 Mean Net TE for Phase I (purple) and Phase II (Green) was plotted in (a). Protein representation of entropy sources (red) and sinks (blue) for Phase I (b), and Phase II (c) *(Sogunmez and Akten, 2020)*.

In addition to $H6_{Int}$, $H3_{Int}$ and $H4_{Int}$ regions also appeared as entropy acceptors in Phase II, which was surprising since these regions are close to the ionic lock incorporating *Arg*131 on H3 and *Glu*268 on H6. The average net TE value of *Arg*131 was decreased from +1.9 to -5.9, shifting from Phase I to Phase II. Moreover, the other residue,

*Glu*268, playing a role in this ionic lock formation has a decrease in its average net TE value from -0.4 to -6.4, representing a strong acceptor characteristic in Phase II compared to either donor or mild acceptor characteristics in Phase I. Since the conformation of this ionic lock was reported as critical in receptor activation (Ballesteros *et al.,*2001), capturing its strong entropy acceptor feature in newly discovered *"highly"* inactive state is extremely significant.

Both "sink" and "source" residues discovered so far in both phases have almost the same intensity given as in Figures 5.5b and 5.5c as red and blue colors, respectively. However, the locations vary significantly. In Phase I given in Figure 5.5b, the red color was concentrated intracellularly, while it shifted to the extracellular part in Phase II, designated in Figure 5.5c. In Phase I, intracellular parts of the transmembrane helices, ICL2 and ICL3, majorly act as entropy donors, except its residues *Arg*260 to *Phe*264, which were recognized as entropy acceptors, previously. Conditional time-delayed cross-correlation (CTC) analysis given in chapter 4 (Sogunmez and Akten, 2019), also revealed a similar dual character for ICL3 as it intensely drove the future fluctuations of H3 and H5, while driven by H1 and H8. Similarly, H8 was also detected as an entropy source, which was also determined as a strong driver in Phase I. Interestingly, the association of H8 with arrestin binding was also reported in experimental studies (Zhou *et al.,*2017), which reveals the importance of this short intracellular helix.

5.1.1.4 The transition of extracellular ligand binding site to a donor in Phase II

Orthosteric ligand binding site on $H6_{Ext}$ was detected as entropy acceptor in Phase I including residues *Trp*286, *Pro*288, *Phe*289, *Phe*290, *Ile*291, and *Val*292. Especially residues *Trp*286, *Phe*289, and *Phe*290 are known as being orthosteric ligand binding sites in β_2AR by creating an extended aromatic network (Chelikani *et al.,*2007; Strader, Sigal, and Dixon, 1989; Suryanarayana, Daunt, Von Zastrow, and Kobilka, 1991). Strikingly, through ICL3 closure, the same residues became a major entropy source and started to transfer information through the rest of the receptor.

Additionally, $H5_{Ext}$ was also reported as the orthosteric ligand-binding site with the residues *Ser*203, *Ser*204, and *Ser*207. In Phase II, *Ser*203 and *Ser*207 both displayed entropy donor characteristics. Additionally, *Ile*121 in the connector site, which locates in the middle part of the receptor, displayed a slight increase in its donor characteristics from 11.1 to 14.6. The importance of this residue comes from the receptor activation process, where when *Ser*207 and *Pro*211 move, *Ile*121 moves towards H6, causes a deviation in the intracellular part of H6 away from H3. The motion from Phase I to Phase II was considered as a reverse activation; thus, a strong driving force from the extracellular orthosteric binding site, as *Ser*207 and from the middle part as *Ile*121 was represented through this inactivation process. This was further supported by the residues, which form ionic lock in $H3_{Int}$, ICL1, and ICL2, and became entropy acceptors in Phase II, while they were entropy donors in Phase I.

During activation mechanism of the receptor, intracellular to extracellular allosteric communication was proposed by several MD simulations and experimental studies which demonstrated that activation was in fact initiated when G protein binds to the receptor at the cytoplasmic side leading to the expansion of the intracellular region and the compression of the extracellular orthosteric ligand-binding site (Dror *et al.,*2011; Nygaard *et al.,*2013). Here, the opposite direction of regulation was observed, where extracellular parts of the receptor control the intracellular parts, thus further supports our finding that shifting to Phase II represented an extreme inactivation of the receptor.

Moreover, $H1_{Ext}$, $H2_{Ext}$, and ECL2 were also observed as entropy acceptors in Phase I but became entropy donors through ICL3 closure in Phase II. ECL2 also displayed a dual character with incorporating donors and acceptors in it. As it was observed in Figure 5.5c, entropy donors were colored in red and were populated in extracellular regions, while blue colors, representing entropy acceptors were clustered mostly around intracellular parts. Therefore, Phase II was the representative of the conformational states where entropy (or information) was transferred from extracellular parts of the receptor to the intracellular parts.

5.1.1.5 Association of mutual information and transfer entropy

MI is a measurement tool identifying the shared information of two residues at the same time, while transfer entropy reveals the amount of information gathered from future fluctuations of a residue when the previous fluctuation of another residue is presented. Therefore, the combination of these two provides allosteric sites of the receptor, where they share information at the same time but exchange information at different time points. In Phase I, the highest amount of MI between distal sites was detected between intra- (ICL2, ICL3) and extracellular (ECL2) loop regions. As it was presented, ICL3 and ECL2 displayed dual characteristics in terms of entropy transfer, some of which were observed as entropy donors, but some were as acceptors. Specifically, the residues of ICL3 adjacent to H6, which was detected as a major entropy sink region (or acceptor), also had the highest mutual information shared with either ICL2 or ECL2.

When shifting to Phase II, the most detectable change was the presence of the extracellular tip of H1 in the mutual information plot, indicating a mutual correspondence between H1 and ICL2 and ECL2. $H1_{Ext}$ also appeared in the transfer entropy map as a major entropy source in Phase II. Thus $H1_{Ext}$ emerged as a critical domain in the newly discovered inactive state's allosteric communicating network. While ECL2 represented both sink and source attributes in alternate orders, ICL2 became a sink region. Furthermore, the value of information shared between ECL2 and ICL2 was increased in this novel state. The entropy "sink" attribute of the segment from 263 to 269 still persisted with increased intensity in Phase II, and moreover, it also displayed an increase in the average MI. Overall, these results indicate that the regions, which displayed the high mutual information values, were also displayed major entropy "sink" or "source" features.

5.1.1.6 Orthosteric and allosteric binding sites were differentiated by the TE contributions of the residues in both phases

Several residues at the extracellular orthosteric binding site were discovered by agonist and/or antagonist binding. (Liapakis *et al.,* 2000; Strader *et al.,* 1987, 1988;

Suryanarayana and Kobilka, 1993) These residues were captured in this chapter as being important information donor/acceptor sites with different characteristics. Orthosteric binding sites were determined as entropy donors in Phase II. The net transfer entropy profiles of nine selected orthosteric binding residues were illustrated together as in Figure 5.7a and 5.7b for Phase I and Phase II, respectively. In Phase I, their net transfer entropy values were fluctuated in a wide window with different altitudes, whereas in Phase II, they almost merged into one unique silhouette. Most of the net TE values in Phase I appeared around or under zero, except *Ser*207 and *Asp*113, which acted as entropy donors in both phases. On the other hand, in Phase II, these nine residues shared positive values above zero, signifying them as major entropy donors. Furthermore, when shifting from Phase I to Phase II, the most significant change in the net transfer entropy appeared at the intracellular sites of H3, H6, and the short segment of ICL3, adjacent to H6, as illustrated in Figure 5.5 for *Ser*207. Evidently, in this novel inactive state, the *expanded* orthosteric ligand-binding site started to control the future fluctuations of the *contracted* intracellular site.

Importantly, another critical site in β_2AR was revealed recently in X-ray studies of Liu *et al.,* in 2017, which presented the first negative allosteric modulator *Cmpd-15PA* (polyethylene glycol-carboxylic acid derivative) bound to a pocket surrounded by the intracellular parts of H1, H2, H6, and H7, H8 and ICL1 (Liu *et al.,*2017). At this newly proposed allosteric site, it was observed that *Cmpd-15PA* forms polar interactions with *Thr*274, *Ser*329, *Asp*331, *Arg*63, and *Asn*69. Furthermore, *Val*54 and *Ile*58 at the end of H1, *Leu*64 in ICL1, *Ile*72 in H2, *Leu*275 in H6, and *Tyr*326 from the conserved *NPxxY* sequence in H7, and *Phe*332 in TM8 forms a hydrophobic pocket and stabilize the cyclohexyl and phenyl rings of *Cmpd-15PA*.

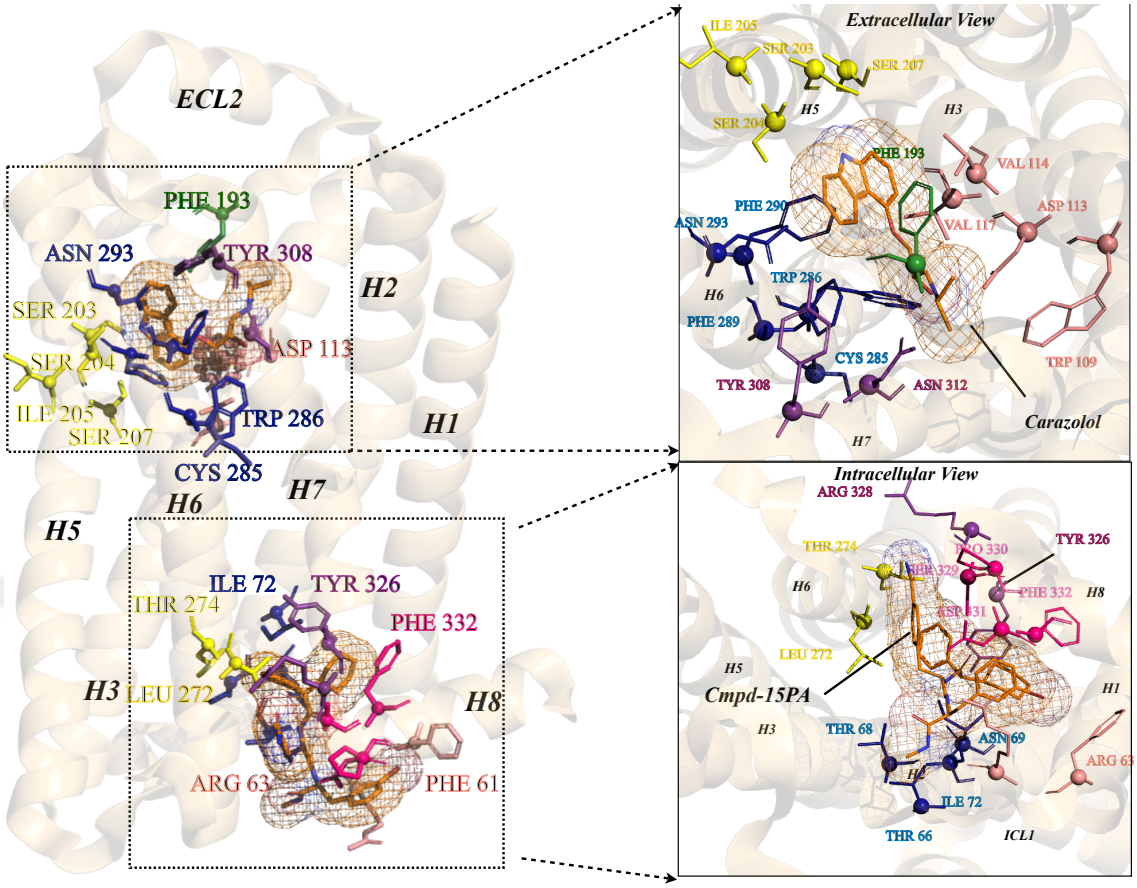
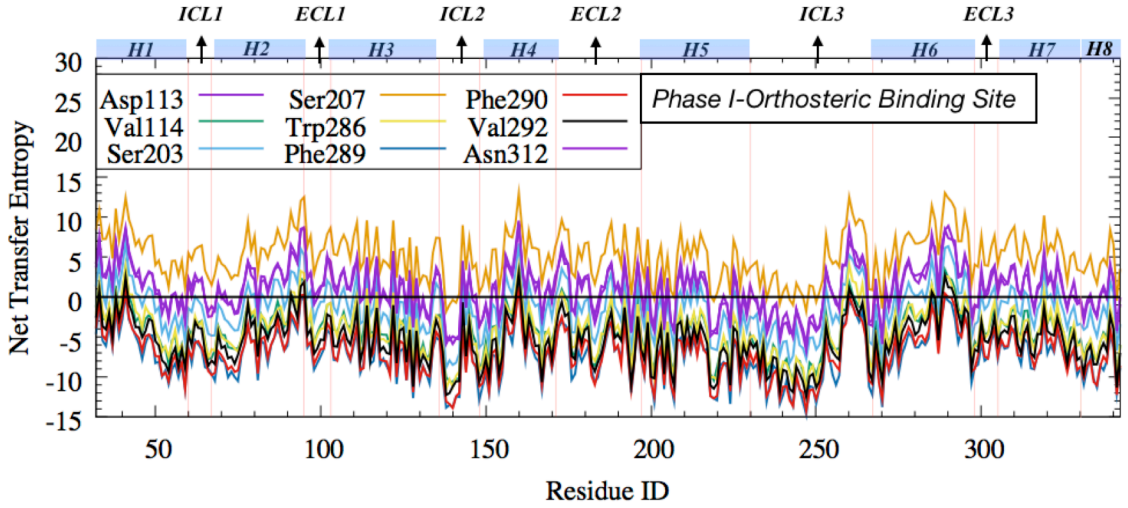
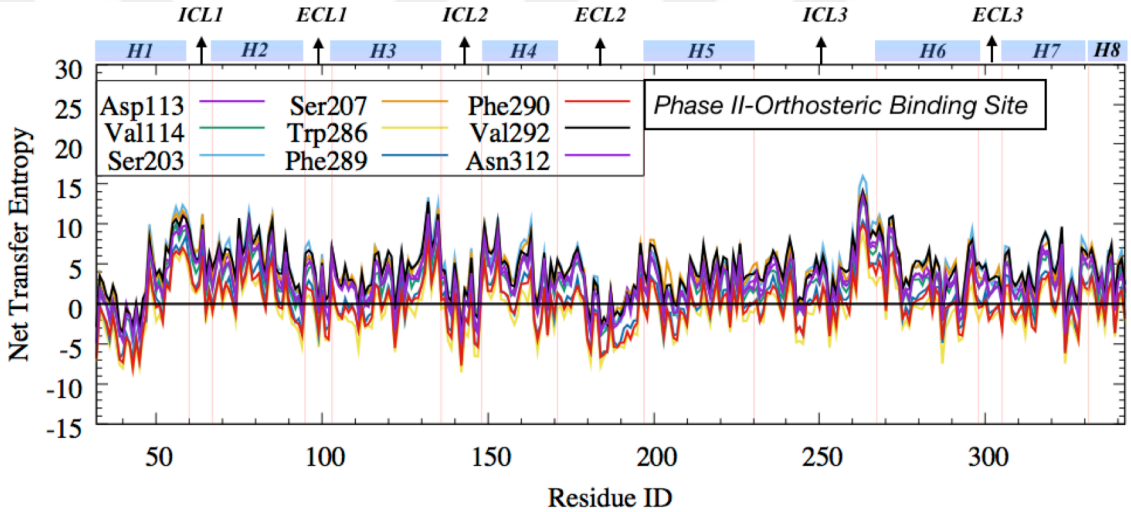


Figure 5.6 Inverse agonist Carazolol and Negative Allosteric Modulator *Cmpd-15PA* binding sites in β_2AR (PDB ID: 5X7D) *(Sogunmez and Akten, 2020)*.

Among all these NAM interacting residues, only *Leu*275, and *Tyr*326 displayed entropy donor features in Phase II, while the remaining ten residues appeared as entropy acceptor or *sink* sites with mean net transfer entropy values varying from -2.1 to -6.4. Since this NAM binding site was blocked by ICL3 in Phase II, these residues fluctuations became fully under the control of those at the extracellular region of the receptor.

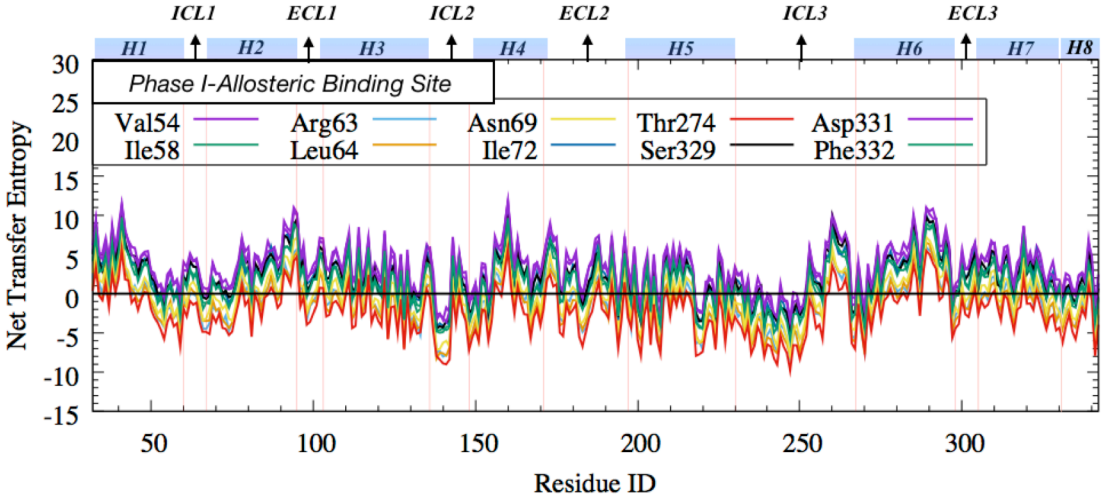


(a)

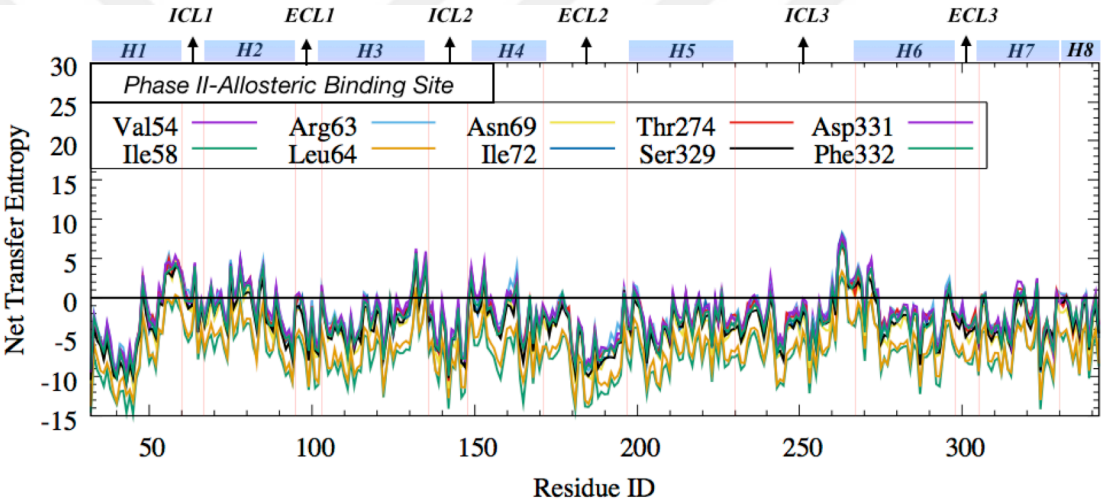


(b)

Figure 5.7 Phase I (a) and Phase II (b) orthosteric binding residues' net transfer entropy values *(Sogunmez and Akten, 2020)*.



(a)



(b)

Figure 5.8 Phase I (a) and Phase II (b) negative allosteric binding sites' net transfer entropy values *(Sogunmez and Akten, 2020)*.

5.1.2 Active receptor without intracellular binding partners (PDB ID:3SN6)

The crystal structure of active β_2AR was resolved in 2011, shortly after the inactive state was resolved (Rasmussen *et al.,*2007). The stabilization of the active states of GPCRs is difficult since the receptors have a tendency to stay in its inactive state. Even though the basal activity of β_2AR was identified without the presence of a ligand or intracellular binding partner, fully stabilization of the receptor requires both the ligand and intracellular binding partners in the environment. Moreover, the crystal structure of a protein represents a single snapshot in time and does not hold the information about

the dynamic repertoire incorporating several distinct conformational states or how much mobile or stable each conformational state upon ligand binding is still not clearly known. This uncertainty poses an obstacle to structure-based drug design approaches. Therefore, an MD simulation was performed by using an active β_2AR (PDB ID: 3SN6) system (Figure 5.9) by removing intracellular trimeric G protein to compare the dynamics with our findings obtained from the inactive states.

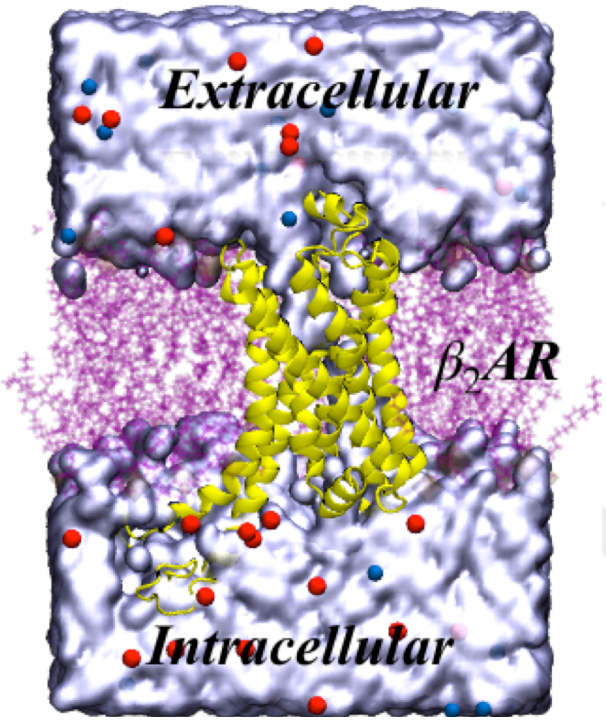
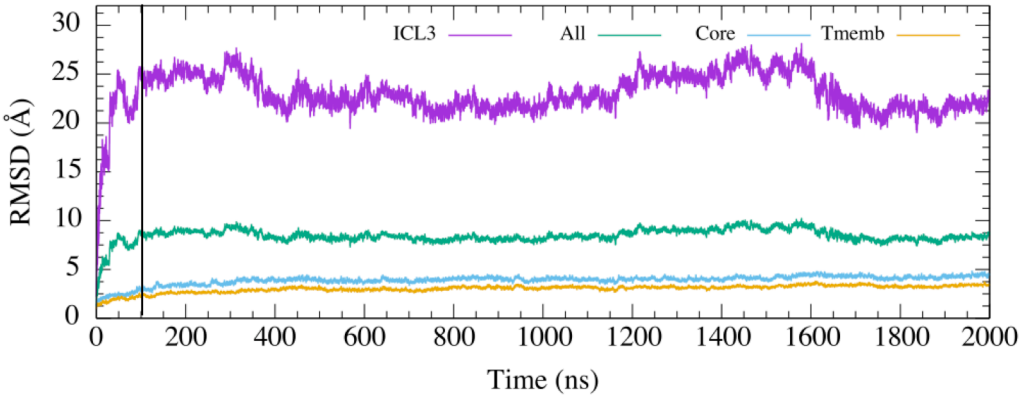


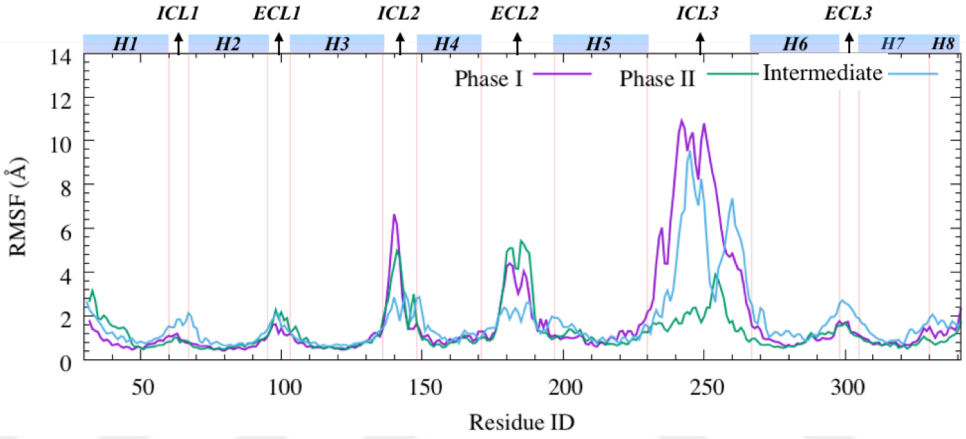
Figure 5.9 Active β_2AR (PDB ID: 3SN6) in POPC lipid bilayer surrounded by TIP3 water and Na^{+2} and Cl^{-1} ions.

5.1.2.1 Dynamics of the *intermediate* state originated from the large deviation of ICL3

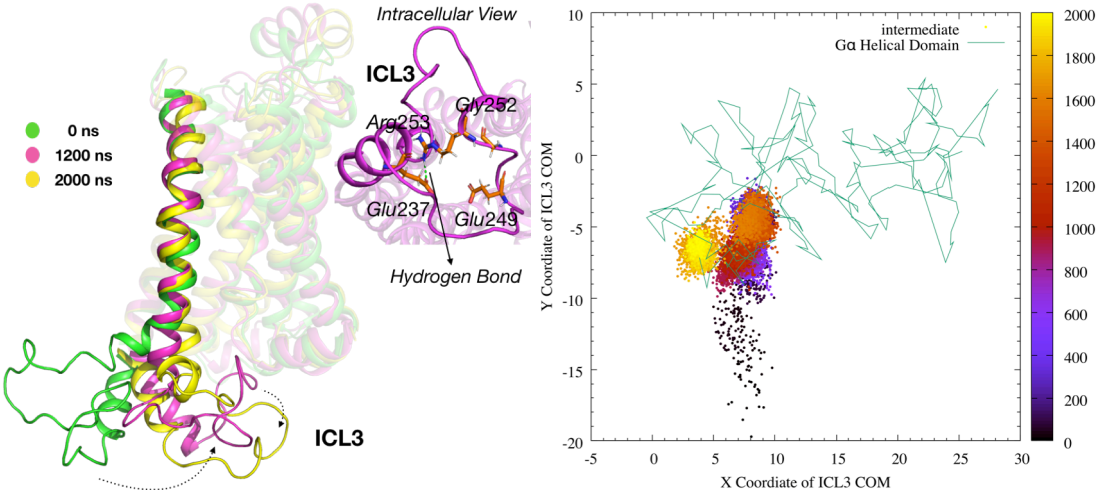
For visualization and analysis purposes, the MD simulation trajectory of the *intermediate* state was further collected with 0.2 ns intervals. A total of 10,000 frames was collected to represent 2 μs long trajectory. Similar to the inactive state of βAR (PDB ID: 2RH1), the dominancy of ICL3 mobility existed as well (Figure 5.10a). RMSD analysis for whole protein (All), protein without ICL3 (Core), transmembrane-only (Tmemb), and ICL3-only (ICL3), revealed that the highest deviation belonged to ICL3 where it also affected the whole protein deviation.



(a)



(b)



(c) (d)

Figure 5.10 RMSD of the *intermediate* state (a) and RMSF plots of Phase I (magenta), Phase II (green), and *intermediate* states (blue) (b). Representation of the trajectory of ICL3 with three snapshots as 0 ns (green), 1200 ns (pink), and 2000 ns (yellow) (c). Center of Mass of ICL3 with respect to Gα helical domain (d).

On the other hand, remaining Core and Transmembrane domains did not deviate as much as the ICL3 and whole protein. Based on the RMSD plot, the initial 100 ns of the trajectory were eliminated from the remaining calculations and accepted as the time for equilibration of the conformation. Similar to RMSD, RMSF calculation revealed ICL3 as the domain with the highest fluctuation, except for residues *Gly*252 and *Arg*253, which displayed a relatively low RMSF due to the prolonged interaction of *Glu*249 with the core of the receptor (Figure 5.10b). The fluctuation of ICL3 in the *intermediate* state was almost as high as the fluctuation of ICL3 in Phase I. The intracellular part of H6 and ICL3 displayed large fluctuations, first by deviating through the core of the receptor. Residue *Glu*249 formed hydrogen bonds with $H7_{Int}$ and ICL1 and then pulled the loop towards the core, thus maintained *Arg*253 to form hydrogen bonds with *Glu*237. After around 1600 ns, the bonds between *Glu*249 and the core were broken. Similarly, most of the hydrogen bonds between ICL3 and the core of the receptor were broken (Appendix C), and ICL3 was released slightly. However, the hydrogen bonds between *Glu*237 and *Arg*253 were still persistent, thus prevented *Gly*252 and *Arg*253 from fluctuating (Appendix C). Even though the RMSF of ICL3 reached up to 9.5 Å, the RMSF of *Gly*252 and *Arg*253 were 3.1 Å and 2.6 Å, respectively. The ICL3 release also induced short helical formation of residues in ICL3 adjacent to H6 (Figure 5.10c). The center of mass of ICL3 in Figure 5.10d also represented the rapid closure of ICL3 and its overlap with the Gα helical domain. Therefore, the occupied region of ICL3 was clearly observed as the region where G protein associates with the receptor. Also, the fluctuations of ICL2, ECL2, and extracellular H5 were relatively high, ~2 Å, compared to the transmembrane fluctuations in the *intermediate* state, but were well below the fluctuations in Phase I and Phase II, which were ≥ 4 Å. The amount of fluctuations in remaining short loops, ICL1, ECL1, and ECL3, were comparable with the fluctuations of ICL2 and ECL2; thus were low. Here, the effect of active conformation, used as the initial conformation, was clearly observed. In Phase I, ICL2, ECL2 and ICL3 were highly mobile, but upon ICL3 closure in Phase II, ICL2, and ECL2 remained flexible and ICL3 fluctuation significantly decreased. However, in the *intermediate* state, since its starting point was the active receptor, ICL3 fluctuation was as high as the Phase I, but the other loops remained more rigid compared to ICL3. This was also observed in *active-Gp* and *active-*Gp_{2nd} trajectories, which will be explained in the upcoming

sections, where the active receptor states were bound intracellularly to the trimeric G protein.

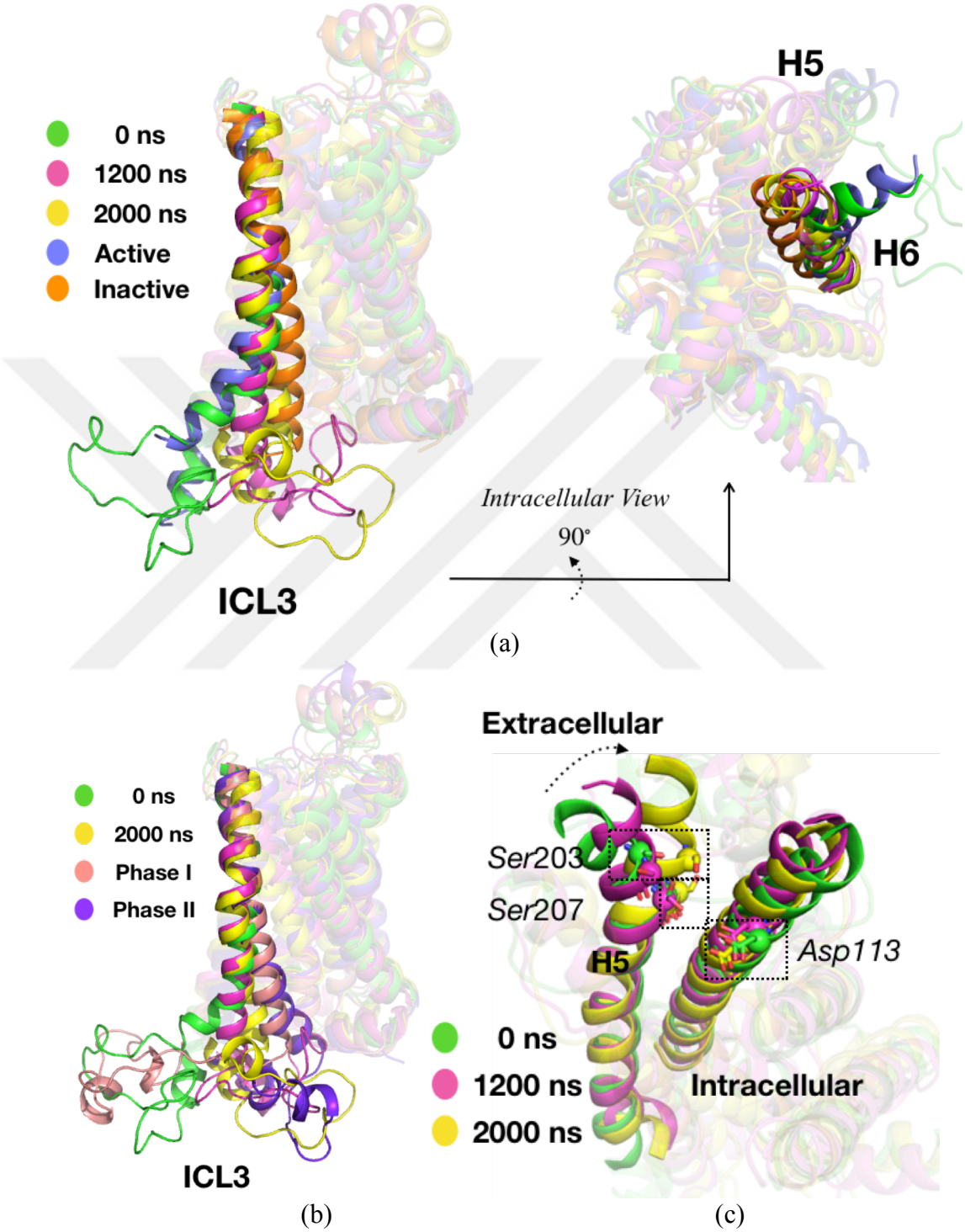


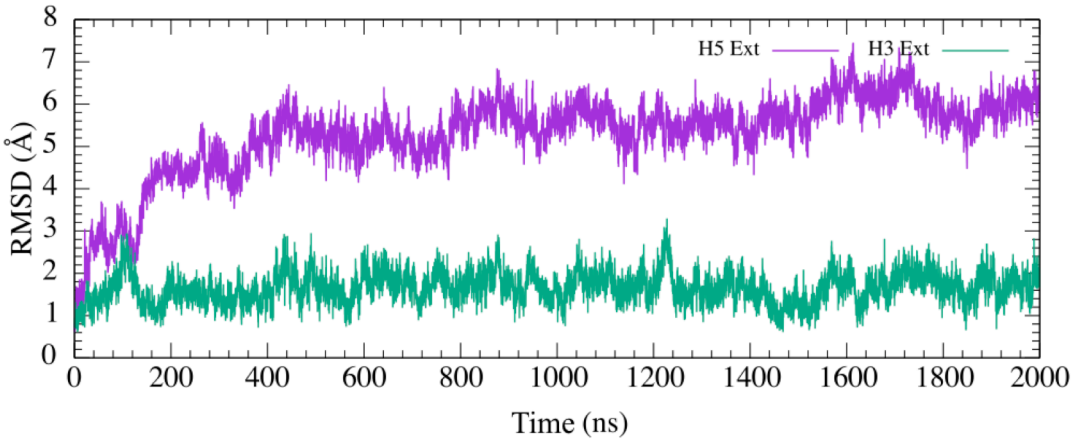
Figure 5. 11 Comparison of the conformational change of ICL3 and $H6_{Int}$ among 3 different snapshots in the intermediate frame, inactive, and active X-ray structures (a). The conformations given in (a) together with Phase I, and Phase II (b). The view of H3 and H5 together with *Asp*113, *Ser*203, and *Ser*207 in three time points as 0 ns, 1200 ns, and 2000 ns (c).

Moreover, the comparison of *intermediate* trajectory with active (PDB ID: 3SN6) and inactive (PDB ID: 2RH1) represented that the closure of $H6_{Int}$ due to ICL3 mobility was not as high as the $H6_{Int}$ in the inactive crystal structure (Figure 5.11a). The structural comparison of *intermediate* trajectory with Phase I and Phase II was also given in Figure 5.11b. Here, the initial conformation of the *intermediate* state, colored in green, represented the outward tilted ICL3 and $H6_{Int}$.

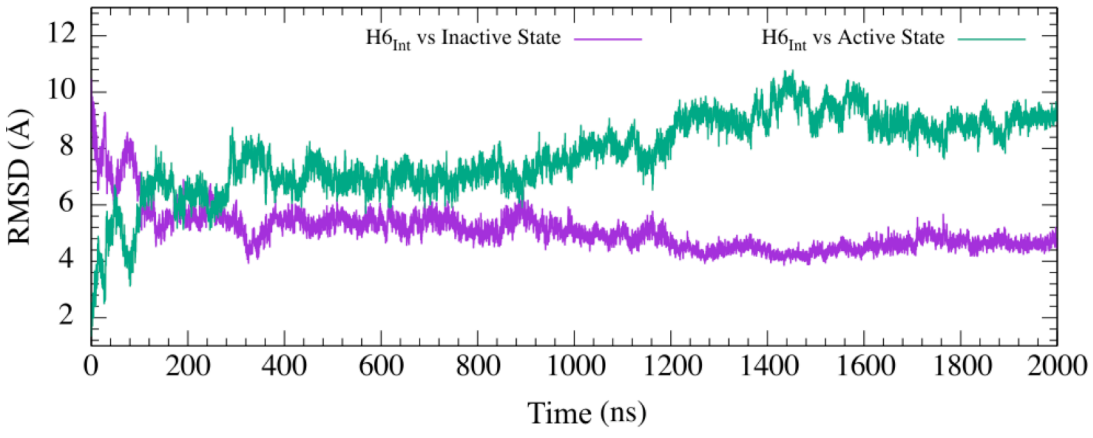
In time, the inward tilt in $H6_{Int}$ was observed as parallel with the motion of ICL3 (intermediate frame was in pink and final frame was in yellow). $H6_{Int}$ in Phase I (light-pink), and in Phase II (violet), was positionally oriented more vertical and through the core of the receptor compared to the final frame of the *intermediate* state's trajectory (Figure 5.11b). Thus, this state did not inactivate completely in time but remained as *intermediate* conformation, and that the fluctuations of ICL3 and $H6_{Int}$ in this *intermediate* state might be representing the basal activity of the receptor due to this persistent conformational fluctuation.

Moreover, the comparison of the deviation of extracellular parts of two key helices H3 and H5 revealed around 6 Å of an outward shift of H5, while H3 stayed relatively stable (Figures 5.11c and 5.12a). As it was shown in Figure 5.11c, the position of *Ser*203 and *Ser*207 changed in time, however *Asp*113 remained highly stable. The outward deviation of $H5_{Ext}$ was associated with the inward deviation of $H5_{Int}$ (Figure 5.11c) due to the closure of the ICL3. Even though a direct association between ICL3 and $H5_{Ext}$ could not be established, a delayed effect of ICL3 motion might be the reason for the alteration of the $H5_{Ext}$ position. Finally, the *intermediate* state's trajectory was compared with the resolved active and inactive crystal structures to understand its behavior in time. The $H6_{Int}$, the activity indicator of β_2AR, was assessed together with the active and inactive snapshots of the receptor and observed that the protein system resembles inactive conformation more than the active one. But still, as mentioned before, the complete closure of $H6_{Int}$ was not observed throughout the trajectory; therefore, this structure remained as one of the *intermediate* states of the receptors (Figure 5.12b).

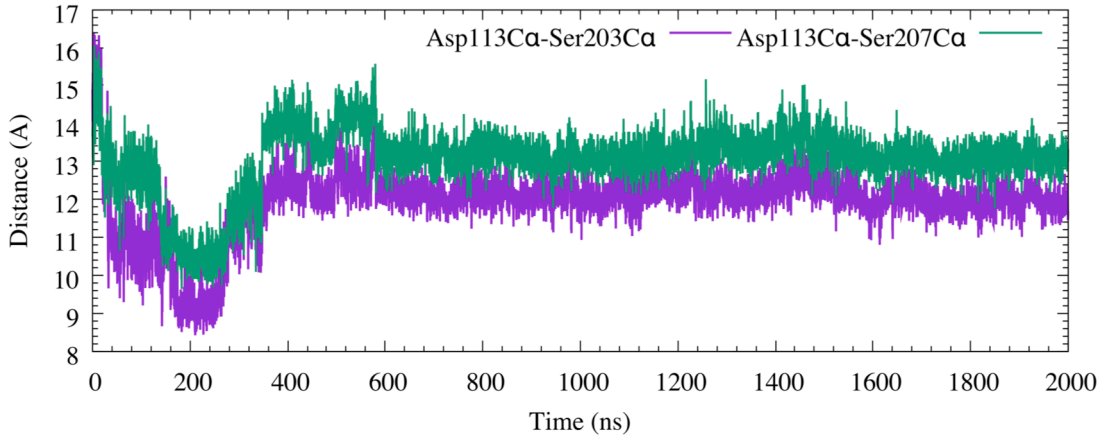
Moreover, to understand whether $H6_{Int}$ and extracellular ligand binding sites motions were correlated or not, the deviation of $H6_{Int}$ was plotted against the distance variations between Cα atoms of *Asp*113 and *Ser*207. Figure 5.12c indicated that up to 2 Å inward intracellular deviation of $H6_{Int}$, there was no response from $H5_{Ext}$, in contrast it became spatially close with $H3_{Ext}$. When $H6_{Int}$ deviation further increased, a configurational arrangement in the external part of H6 was observed, which apparently pushed the extracellular part of H5, thus increased the distance between *Asp*113 and *Ser*207 from 13.6 Å up to 15 Å and then the distance was stabilized at around 13 Å (Figure 5.12c), which was close to the distance observed in the crystal structure of the inactive state (PDB ID: 2RH1), which was 12.17 Å.



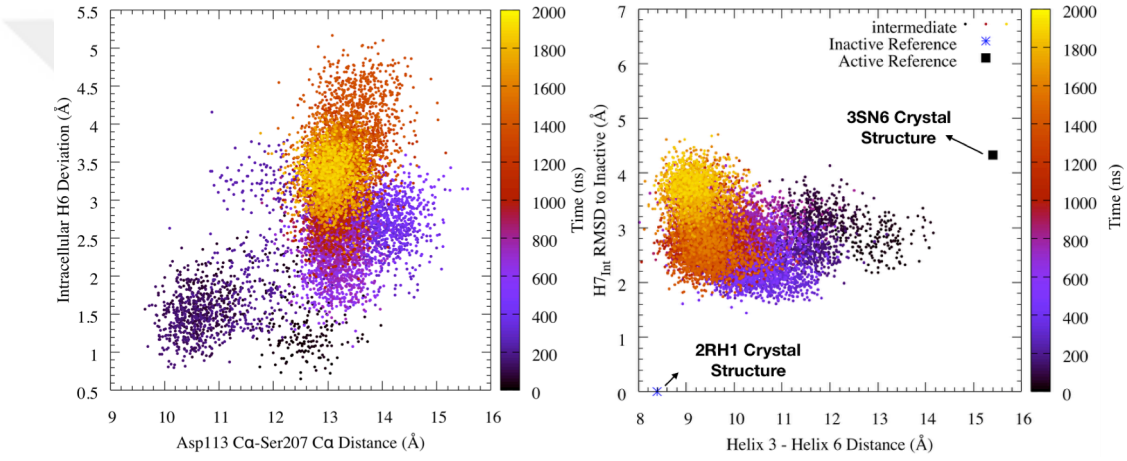
(a)



(b)



(c)



(d) (e)

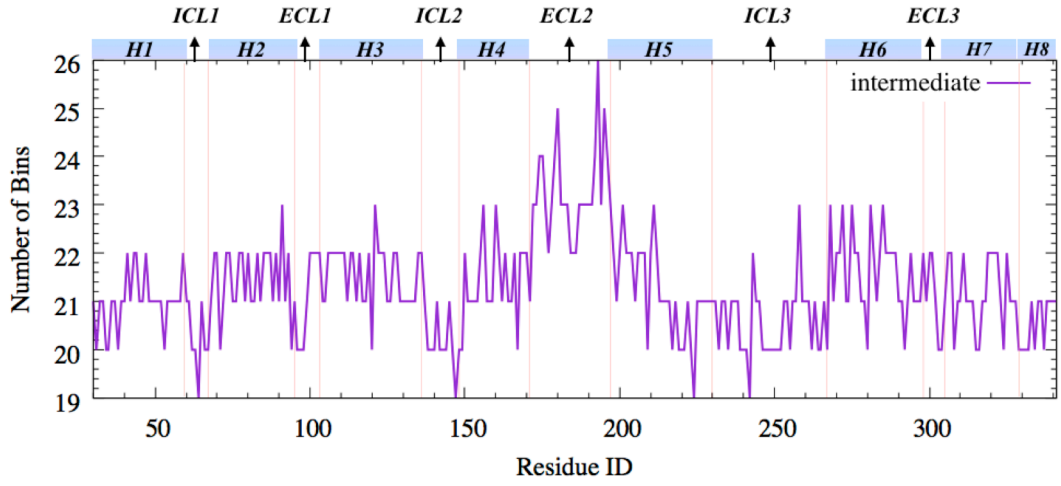
Figure 5. 12 Extracellular part of H5 and H3 deviation (a). The deviation of H5 was higher than H3. The deviation of intracellular H6 with respect to inactive (purple) and *intermediate* (green) conformations (b). The distance variation between *Asp*113 Cα and *Ser*203 Cα (magenta) together with the distance variation between *Asp*113 Cα and *Ser*207 Cα (green) (c). The distance between *Asp*113 Cα and *Ser*207 Cα versus $H6_{Int}$ deviation plot (d), and RMSD of $H7_{Int}$ versus *Arg*131- *Leu*272 Distance together with inactive (PDB ID: 2RH1) (blue stars) and active (PDB ID: 3SN6) (black square) references (e).

The distance between *Asp*113 and *Ser*203 also followed a similar trend as the distance between *Asp*113 and *Ser*207. Finally, one of the activity indicators, $H7_{Int}$ including the *NPxxY* domain, was plotted with respect to the distance change in *Arg*131-*Leu*272 in Figure 5.12d. The intermediate trajectory was observed to be in between the values of active (black square) and inactive (blue star shape) crystal structures, verifying that this trajectory was indeed an intermediate state of the receptor. Due to the deviation of $H6_{Int}$, the distance between *Arg*131 and *Leu*272 was altered from 13.7 Å to 9.3 Å. In parallel, the $H7_{Int}$ deviated up to 4 Å with respect to the inactive crystal structure.

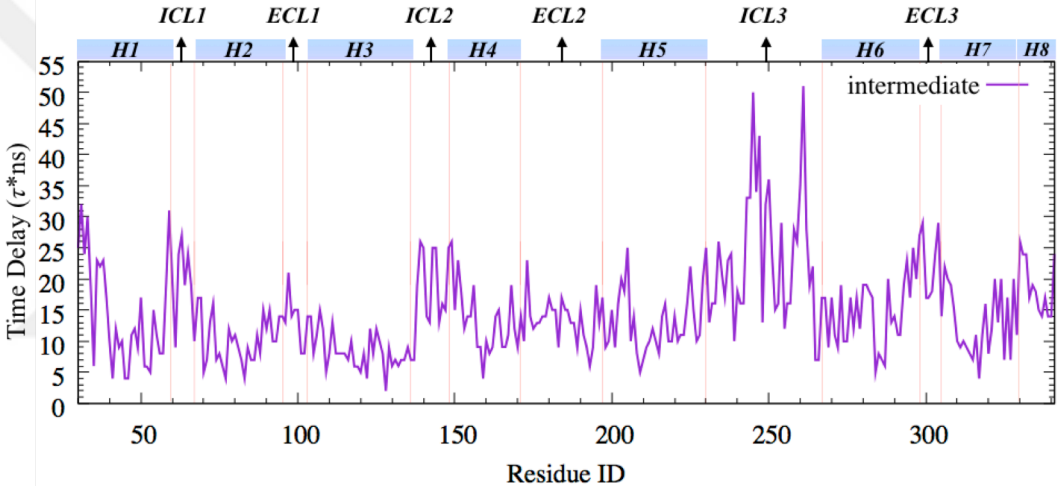
5.1.2.2 Optimized number of bins and time delay for residues in the *intermediate* state

Transfer entropy and mutual information calculation in protein dynamics have become frequent due to its definite outcomes, even though these calculations are tricky to some extent and difficult to optimize their parameters. The main problem in these calculations is the determination of the optimum number of bins and the optimum time delay between signals. The number of bins is directly related to the probability of occurrences of the events. Also, for the determination of the optimum time delays, different approaches have been proposed. One of the approaches is the autocorrelation method (Theiler, 1986), and another approach is finding the first minimum of the mutual information (Fraser and Swinney, 1986). The reason why time embedding is critical is that, an inappropriate choice can either result in false indications of structures where there is none, or fail in resolving structures even though they are present in the data. Since our protein system contains $311 \times 310 = 96410$ (for inactive states) and $312 \times 311 = 97032$ (for active states) different signaling pairs, the level of complexity increases further. Even though a definitive way of system-specific time delay determination has not been proposed, the approaches mentioned above produce approximate results, which are reliable for explaining information transfer. In this thesis, an entropy-based and residue-specific bin and time delay optimization techniques were performed to calculate our entropy transfers.

The number of bins and the time delay optimization for the residues were performed similarly as they were in the inactive trajectories. The residue-specific number of bins for the *intermediate* receptor was between 19 and 26 (Figure 5.13a), and the residue-specific time delay was between 2 and 51 (Figure 5.13b).



(a)



(b)

Figure 5. 13 Residue optimized number of bins (a) and time delay (b) plot for *intermediate* trajectory.

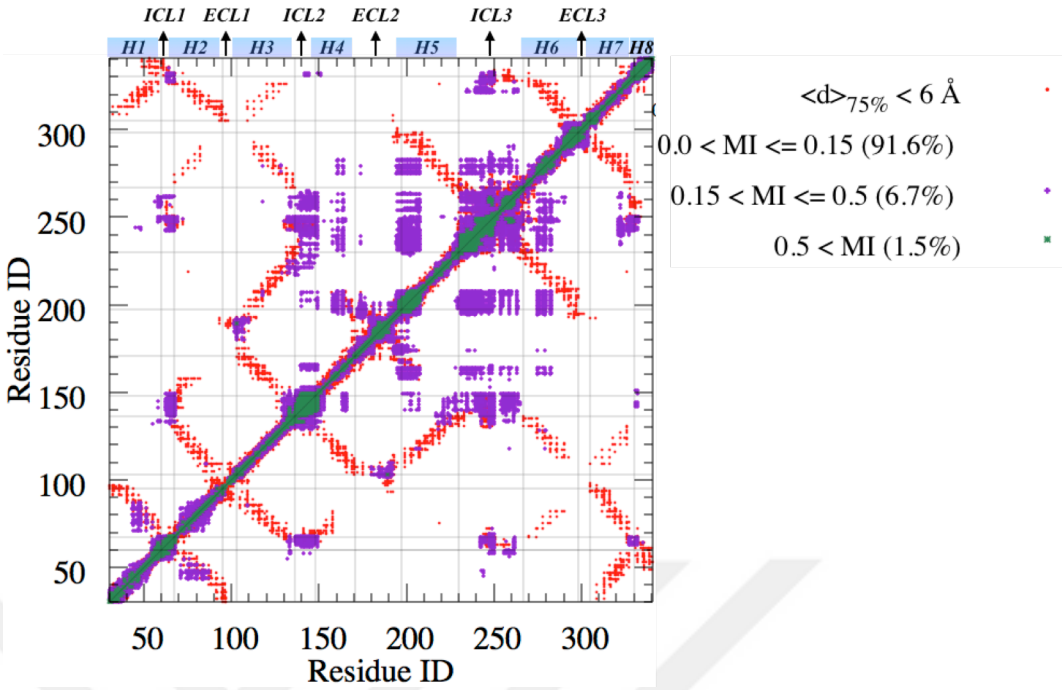
Among all domains, ECL2 had the highest number of bins, as plotted in Figure 5.13a. Also, it was observed in the time delay plot that the delay time in loops was higher compared to the transmembrane helices (Figure 5.13b). This can be explained by the high mobility of loops, where the first minimum of the MI value in time did not occur as rapidly as the transmembranes.

5.1.2.3 Shared information between the allosteric and orthosteric binding sites

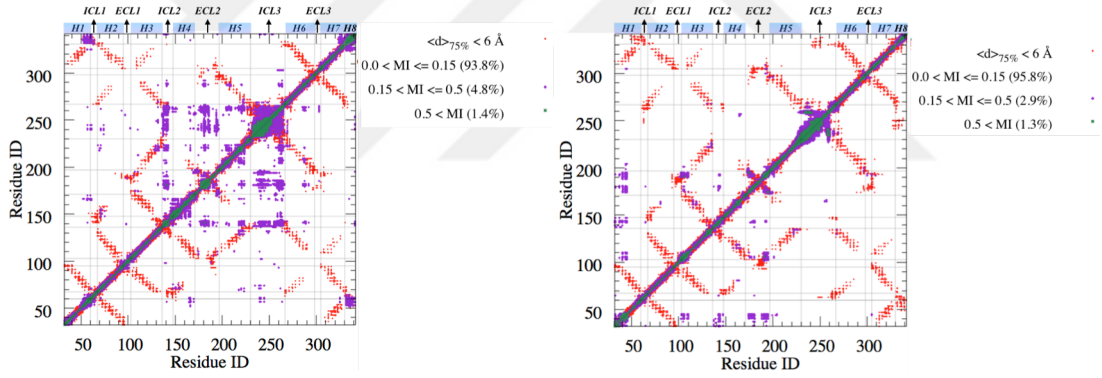
If the qualitative definition of entropy is the uncertainty, then mutual information is a method used for momentarily reduction of uncertainty. From our data, it was apparent that motions of loops could be understood by knowing the fluctuations of other loops. This finding proves that loops do not move randomly, but makes coordinated fluctuations with other loops. In the *intermediate* phase, mutual information existed between ICL3 and two other intracellular helices, ICL1 and ICL2, as well as the middle parts of H4 and H6 and the extracellular part of H5 (Figure 5.14a).

$H5_{Ext}$ also shared information with ECL2 and H4. By knowing the orthosteric binding molecules interact, especially with Serines on $H5_{Ext}$, capturing the role of $H5_{Ext}$ in *intermediate* state's pairwise communications was interesting. Similarly, $H3_{Ext}$ – ECL2 communication was also indicating the activity of orthosteric binding site, as ECL2 is known to function as a gate in the receptor signaling, and $H3_{Ext}$ comprises of the residues, which are critical for ligand binding.

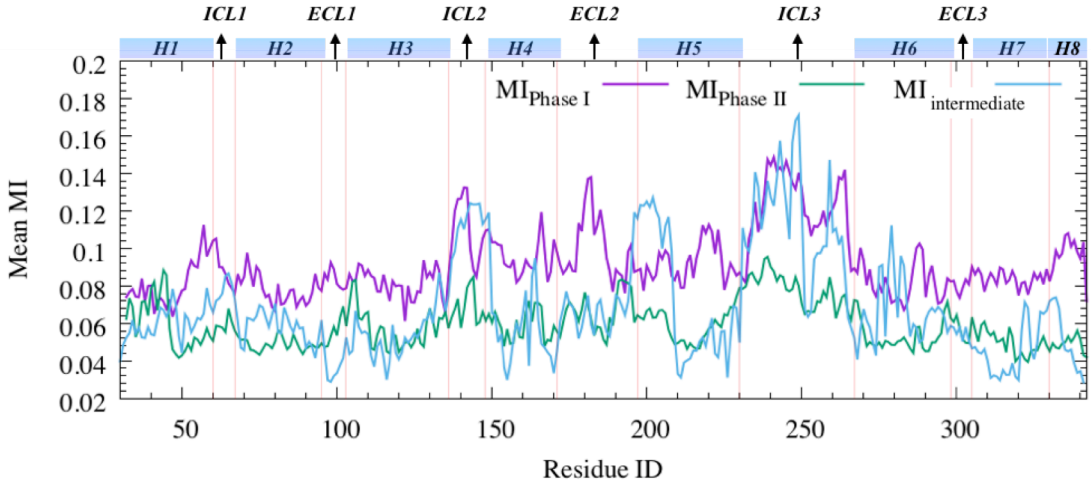
Mean MI in Figure 5.14c represented the residues with their average MI values. Based on this, ICL2, ICL3 and extracellular part of H5 shared a significant amount of information among each other, indicating that both intracellular and extracellular parts of the receptor play a role in communication in the *intermediate* state. Considering this state as the intermediate conformation based on the partial inward tilt of H6 and ICL3, the persistent communication between intracellular and extracellular parts was expected. Moreover, the comparison of *intermediate* state with inactive states revealed that mean MI values in ICL2 adjacent to H4 and ICL3, as well as extracellular part of H5, were higher in the *intermediate* state than inactive states.


ICL1 ECL1 ICL2 ECL2 ICL3 ECL3
H1 H2 H3 H4 H5 H6 H7 H8
Residue ID
Residue ID
<d>75% < 6 Å
0.0 < MI <= 0.15 (91.6%)
0.15 < MI <= 0.5 (6.7%)
0.5 < MI (1.5%)

(a)


Residue ID
Residue ID
<d>75% < 6 Å
0.0 < MI <= 0.15 (93.8%)
0.15 < MI <= 0.5 (4.8%)
0.5 < MI (1.4%)
<d>75% < 6 Å
0.0 < MI <= 0.15 (95.8%)
0.15 < MI <= 0.5 (2.9%)
0.5 < MI (1.3%)

(b)


ICL1 ECL1 ICL2 ECL2 ICL3 ECL3
H1 H2 H3 H4 H5 H6 H7 H8
MIPhase I
MIPhase II
MI intermediate
Mean MI
Residue ID

(c)

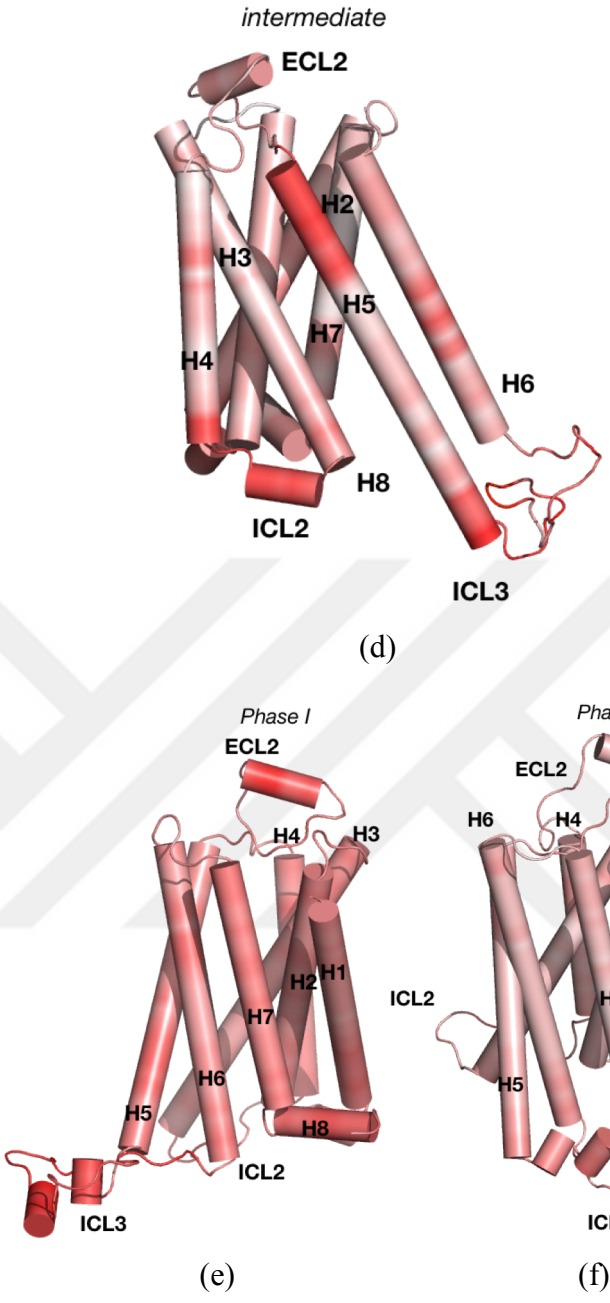


Figure 5. 14 MI heatmap for *intermediate* state superposed with pairwise contacts of heavy atoms (a). MI heatmap of Phase I (left), Phase II (right) (b). Mean MI value for Phase I (magenta), Phase II (green) and *intermediate* (blue) state (c). Protein representation for the distribution of mean MI values for *intermediate* state (d), Phase I (e), and Phase II (f). From white to red, the color intensity indicates the increase in MI values. The common color range for protein representations was taken from the MI values of the *intermediate* state, where the lowest MI value was 0.029, colored with white, and the highest MI value was 0.171, colored with red.

However, the overall mean MI value for Phase I was the highest among these three states followed by the *intermediate* state and Phase II, with values of 0.092 ± 0.018, 0.068 ± 0.028, and 0.059 ± 0.012, respectively. For the average MI, the value in the

intermediate state was closer to the value of Phase I regarding the residues on ICL3, ICL2 and $H5_{Ext}$, and even $H5_{Ext}$ shared more information in the *intermediate* state compared to Phase I (Figure 5.14a). Mutual information values of its other regions, including ECL2, appeared to resemble Phase II more than Phase I. The flexibility of $H5_{Ext}$ was observed to be higher than the flexibility of the remaining transmembranes in the i*ntermediate* state, as well as the flexibility of $H5_{Ext}$ in Phase I and Phase II. Thus, high MI values in $H5_{Ext}$ and its flexibility was related. Interestingly, ECL2 had significantly low MI value compared to the MI values of ICL2, $H5_{Ext}$, and ICL3. This might be the result of the lower fluctuation in ECL2 compared to other domains in the *intermediate* state. Moreover, Figure 5.14b was plotted here as a reminder and for the comparison of MI values in three states. As it was clearly observed both in Figure 5.14a and left plot in 5.14b, the MI values in the *intermediate* state highly resembled the MI values in Phase I with a pairwise distribution of high MI values.

Based on the results above, a relation between RMSF data and MI values is noticeable. A significant decrease in the fluctuation of ECL2, thus its MI value, was separating *intermediate* state from Phase I and Phase II. Additionally, a slight increase in RMSF of $H5_{Ext}$ and H6 had a major contribution to their mutual information values. Therefore, infinitesimal changes in transmembrane fluctuations were observed to strongly affect information sharing. Moreover, the RMSF of ICL2 in Phase I, and Phase II had narrow peaks, while its value was distributed through all residues in the *intermediate* state. The mean RMSF value of ICL2 in Phase I was 3.41 Å, in Phase II it was 2.94 Å, and in the *intermediate* state it was 2.38 Å. As was clearly observed, the average RMSF values were close in three phases, thus explained its high MI value in the *intermediate* state. The MI values in loop regions resembled Phase I; however, in transmembrane regions it resembled Phase II. The protein representation of average MI values per residues in the *intermediate* state, Phase I and Phase II were demonstrated in Figures 5.14d, 5.14e, and 5.14f using a common color gradient, the intensity of which represents the magnitudes of the mean MI value range of the *intermediate* state. Redder the color is the higher the average MI value. Based on the MI range of the *intermediate* state, which was between 0.029 to 0.171, the mutual information of Phase I was observed all over the receptor,

while in Phase II, the information sharing was significantly low compared to the remaining two states.

5.1.2.4 Information mostly shared by polar amino acids in the *intermediate* state

The types of amino acids in the *intermediate* state were also examined to understand their effect on information sharing and allosteric communications. As listed in Table 5.3, both proximal and distal residues showed higher information sharing attributes in polar residues compared to hydrophobic residues. Pairwise comparisons also showed that polar-polar interactions shared a higher amount of information than either polar-hydrophobic or hydrophobic-hydrophobic ones. These values were statistically verified by the nonparametric statistical test of Kruskal-Wallis and Dunn's post hoc tests. Figure 5.15 represents the polar and hydrophobic residue comparisons. In Figure 5.15a, both in proximal (red circles) and distal residues (black rectangles), polar residues were significantly higher than hydrophobic ones. The average MI values for proximal polar residues were in between [0.45-0.69], while distal polar residues were [0.055-0.075]. On the other hand, proximal hydrophobic residues average MI values were in between [0.36-0.50], and distal hydrophobic residues were in between [0.042-0.058]. Figure 5.15b represents pairwise comparisons, where polar–polar interactions were prevalent and statistically significant. Here, this figure shows that polar-polar interactions also resulted in a significantly higher MI value compared to hydrophobic ones both in proximal and distal amino acid pairs.

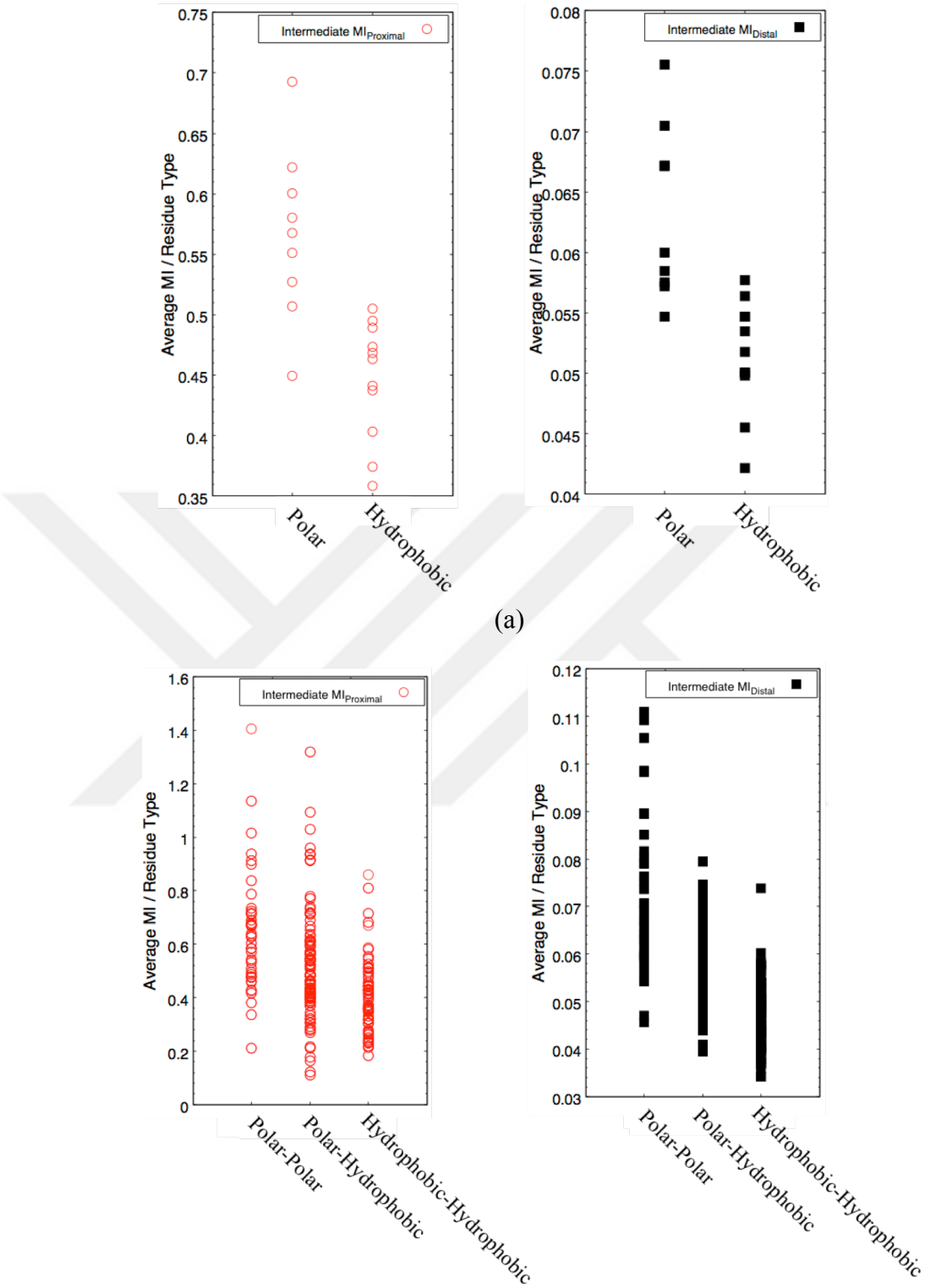


Figure 5. 15 Polar and hydrophobic residues' contribution to MI values in the *intermediate* state. (a). Residue-pair average MI values of polar and hydrophobic pairs (b). Proximal residue-pairs were colored in red and distal ones were black in all four plots.

Table 5. 3 Average mutual information of polar and hydrophobic residues in the *intermediate* state

	Type	Proximity	Mean Mutual Information	# of Residues
intermediate	Polar	Proximal	0.57 ± 0.36	915
		Distal	0.064 ± 0.070	34850
	Hydrophobic	Proximal	0.46 ± 0.29	1561
		Distal	0.052 ± 0.053	59706

Table 5. 4 Average mutual information of pairwise interactions based on polarity in the *intermediate* state.

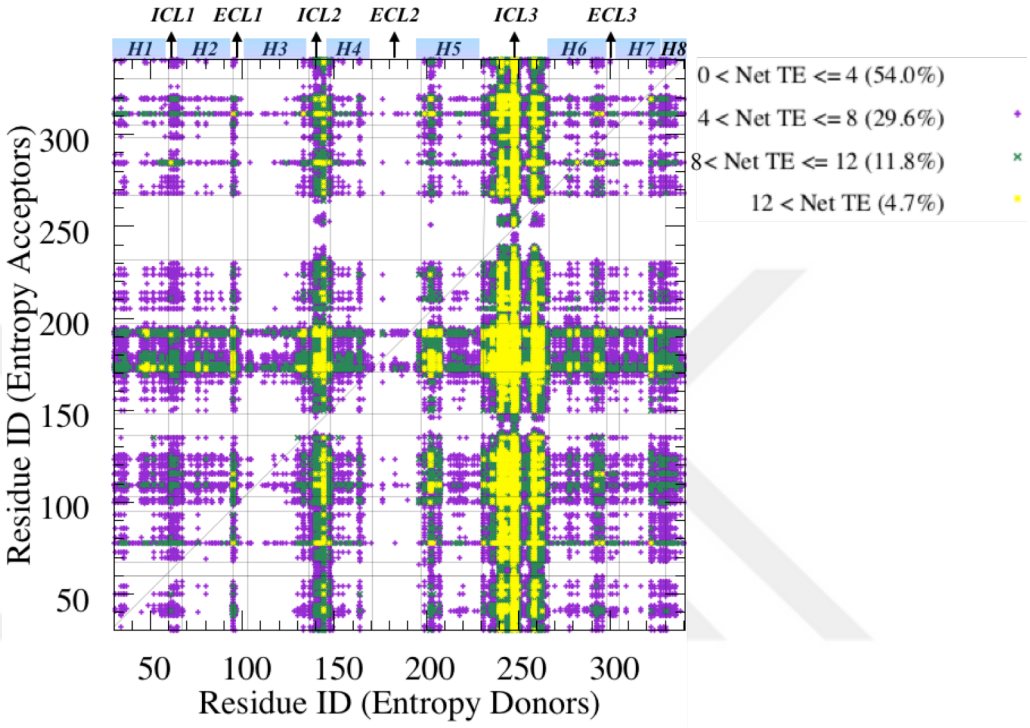
	Type	Proximity	Mean Mutual Information	# of Residues
Intermediate	Polar-Polar	Proximal	0.64 ± 0.38	370
		Distal	0.075 ± 0.084	12740
	Polar-Hydrophobic	Proximal	0.52 ± 0.34	1090
		Distal	0.058 ± 0.060	44220
	Hydrophobic-Hydrophobic	Proximal	0.43 ± 0.25	1016
		Distal	0.049 ± 0.048	37596

5.1.2.5 Entropy transfer in *intermediate* state

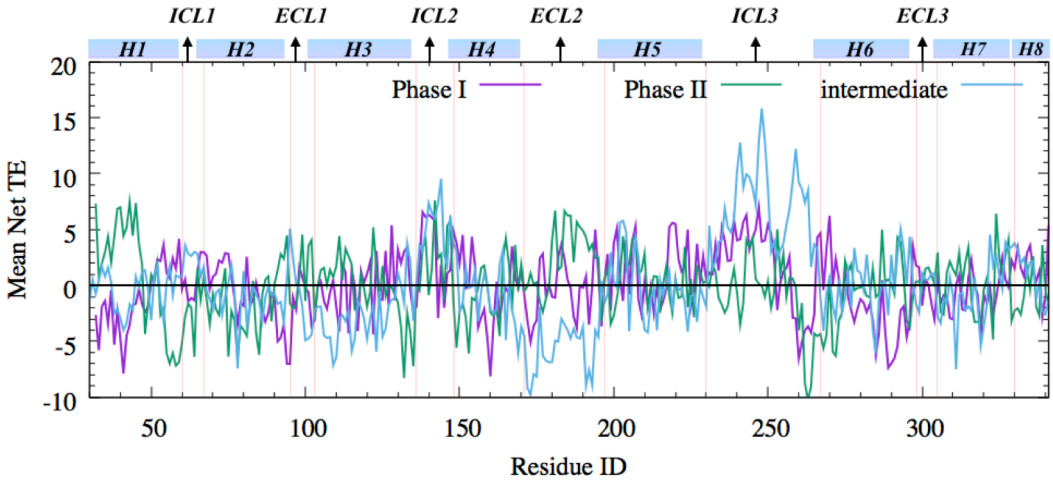
To identify the direction of information transfer as well as the source and sink residues, Erman's group TE formula based on Schreiber's approach was used (Hacisuleyman and Erman, 2017; Kamberaj and van der Vaart, 2009). Figure 5.16 represents the transfer entropy in the *intermediate* state. The residues in abscissa were entropy donors, and the residues in ordinate were entropy acceptors. More than half of the residue-pairs transferred a low amount of information with values between 0 and 4. Strong information transferring residues only constituted 4.7% of the overall residues, which were mainly localized intracellularly.

A strong information transfer was observed from residues in ICL2, $H5_{Ext}$, and ICL3, whereas the two ends of ECL2 were observed as the highest entropy acceptors among all domains. All three regions, ICL2, $H5_{Ext}$, and ICL3, were observed to have the highest mutual information sharing with the rest of the receptor, as it was shown in Figure 5.14. Here, they also appeared as transferring information from their previous states to the rest of the residues' future states. Additionally, the intracellular entropy donor features were observed on the critical residues, which associate with the positive

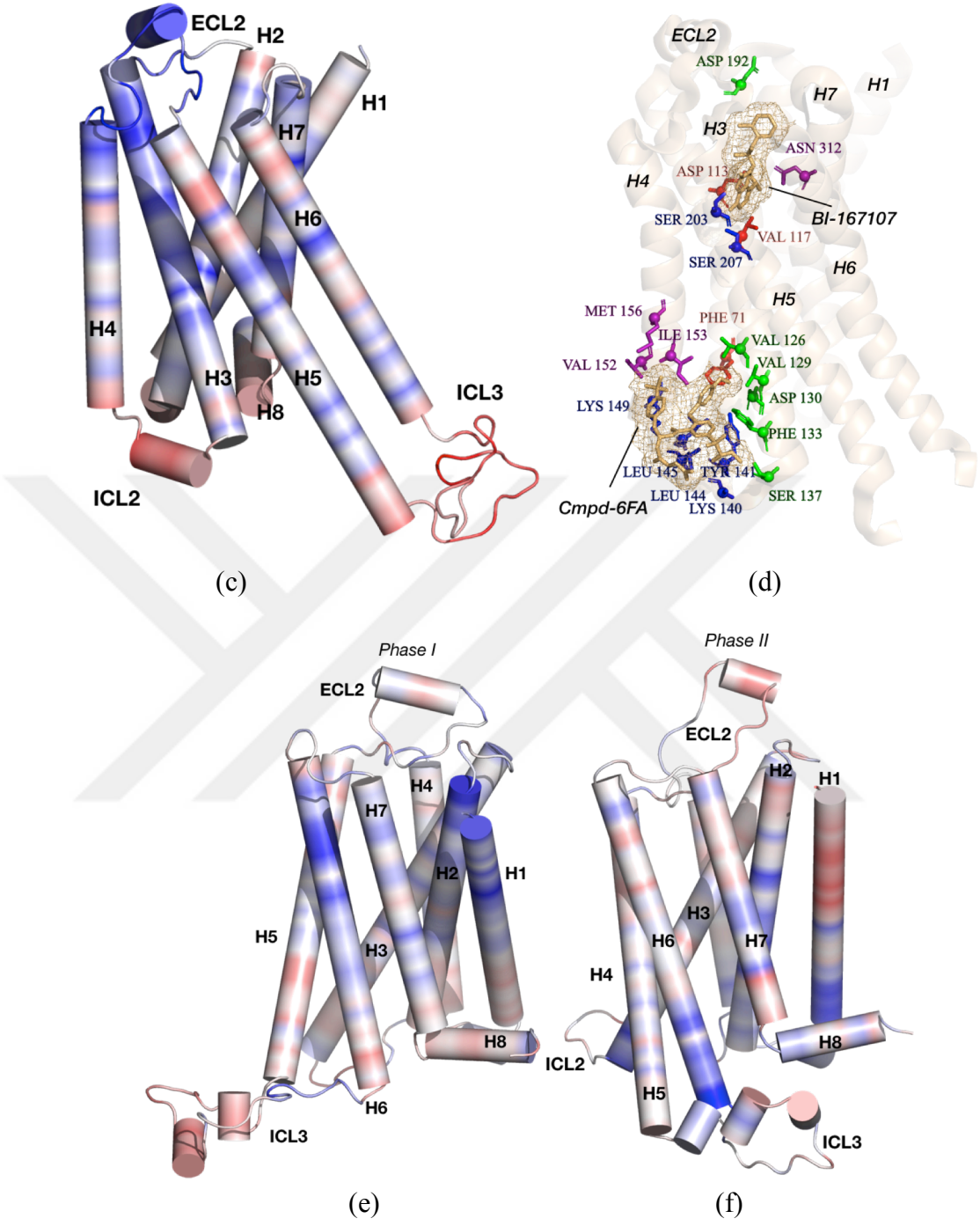
allosteric modulator, *Cmpd-6FA* (Figure 5.16d). It was reported that this molecule interacts with the receptor through the pocket enclosed by $H3_{Int}$, ICL2, $H4_{Int}$, which were observed as entropy donors in the *intermediate* state. The direction of the entropy transfer was determined from intracellular loops ICL2 and ICL3 to extracellular domains of H3 and ECL2 (Figure 5.16c).



(a)



(b)



.**Figure 5. 16** Entropy donors and acceptors in the *intermediate* state (a). Mean Net TE values for Phase I (magenta), Phase II (green), and *intermediate* state (blue) (b). Protein representation of the *intermediate* state indicating entropy donor residues (red) and entropy acceptor residues (blue) (c). Protein representation of agonist and Cmpd-6FA binding in β_2AR (PDB ID: 6N48) (d). Mean net TE protein representation of Phase I (e) and Phase II (f).

Additionally, the mean net transfer entropy calculations of individual residues in terms of their transfer entropy characteristics were plotted in Figure 5.16b. The intensity of

entropy donors with net TE values above 12 was almost ten times as high in the *intermediate* state compared to Phase I, which was 4.7% in the *intermediate* state, while 0.41% in Phase I. Especially the regions ICL2 and ICL3 displayed higher entropy donor features shown with a vertical bright yellow area in Figure 5.16a, whereas the highest entropy acceptor regions appeared as ECL2 (horizontal yellow area) in the *intermediate* state. Finally, *Leu*95 in ECL1 and *Asn*293 in $H6_{Ext}$ displayed entropy donor characteristics, showing that entropy transfer in the *intermediate* state was bidirectional and the mixture of the two inactive states.

Thus, based on the location of the highest entropy donors, the *intermediate* state resembled Phase I (Figures 5.16c and 5.16e). However, the entropy was transferred from these loops to a limited extracellular area enclosed by $H3_{Ext}$ and ECL2, where in Phase I, the entropy sources were distributed to wider extracellular parts of the receptor. Therefore, the *intermediate* state generally showed inactive characteristics in terms of the fluctuation and mutual information terms, however it also displayed unique features in terms of the information transfer.

Overall, the conformational change in the *intermediate* state represented an inactive state due to the distance between $H5_{Ext}$ and $H3_{Ext}$, as well as inward tilt of ICL3 and $H6_{Int}$. However, this inward deviation was not as high as the crystal structure of the inactive state, indicating this trajectory as one of the intermediate trajectories of the receptor. The RMSF data represented high fluctuations in ICL2, ICL3, and interestingly in $H5_{Ext}$. However, the fluctuation of ECL2 diminished in the *intermediate* state, which was one of the differences of *intermediate* state from Phase I and Phase II. Moreover, the relatively higher fluctuation of $H5_{Ext}$ compared to other transmembrane helices contribute to the information sharing. Even though the fluctuation of $H5_{Ext}$ was not as high as the loops ICL2 or ICL3, it was observed that a slight increase in the fluctuation of a transmembrane domain strengthened the mutual information of this domain with other residues.

Moreover, the average MI was highest in Phase I, followed by the *intermediate* state and then Phase II. However, in terms of net transfer entropy, the strongest entropy

transfer percentage belonged to the *intermediate* state with 4.7%, while this value was 0.41% in Phase I. Entropy donors and acceptors in the *intermediate* state were populated in specific regions, such as ICL2, $H5_{Ext}$ and ICL3 as donors and $H3_{Ext}$ and ECL2 as acceptors. On the other hand, donors and acceptors were distributed to different domains in Phase I and Phase II, where donors were mostly found in intracellular parts and acceptors were found in extracellular parts in Phase I and the other way around in Phase II. This clustering of donors and acceptors indicated a distinct but characteristic communication pattern between intracellular loops and ECL2, representing that the receptor in the *intermediate* state was awaiting intracellular signals to alter its function.

5.1.3 The outward position of ICL3 was protected through the interaction of α-helix of the trimeric G-protein in active receptor (PDB ID:3SN6)

Two different activation modes were proposed for GPCRs. The first one is through agonist binding on the extracellular part of a receptor, thus inducing G protein to associate intracellularly. In this model, the interaction was believed to occur via *collisional coupling*, where G protein and only activated GPCRs can couple by collusion (Gilman, 1987). However, the full activation was shown to occur in around milliseconds (Hille, 1992), which is a rapid event for a collisional incidence. Thus a second model, called *physical scaffold*, was introduced to explain the rapid activation of many G-protein-mediated signaling pathways. Based on this second hypothesis, specific protein components interact during activation and GPCR-promoted activation of G-proteins are available (Rebois and Hébert, 2003). Thus, G-protein is believed to preassemble with GPCRs before activation and then forms a stable protein complex at the very beginning of activation (Galés *et al.,* 2006). This was also supported by the results that agonist-bound active receptors span intermediate conformations and inactivated when G protein was not available or did not interact with the receptor-agonist complex (Manglik *et al.,* 2015; Nygaard *et al.,* 2013). Moreover, many experimental data support the existence of early GPCR / G protein complex (Hu *et al.*, 2010; Lachance *et al.,* 1999; Nanoff *et al.,* 1991; Pachov *et al.,* 2016; Qin *et al.,* 2011).

In this section, the active state of the receptor used in 5.1.2 was also used here together with the intracellularly associated G protein. The orthosteric-binding site was remained unoccupied to observe the tendency of the receptor when only trimeric G protein was associated intracellularly. This was the state where G protein was preassembled with the protein to stabilize it just before the agonist binding. As illustrated in Figure 5.17, the modeled ICL3 in the active receptor was extended distantly from the core of the protein due to the outward position of H6. Due to the presence of the α-helix subunit of the Ras-Like domain of Gα, this extended conformation of ICL3 was maintained throughout the simulation. Two independent simulations were run, and they represented slightly different motions. The first run was relatively stable, but ICL3 in the second one was more mobile throughout the simulation, as explained in upcoming sections.

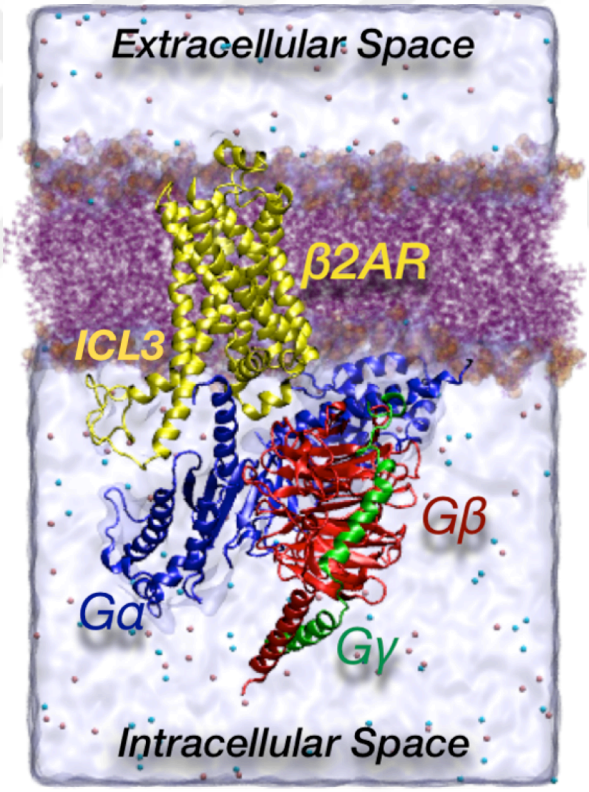


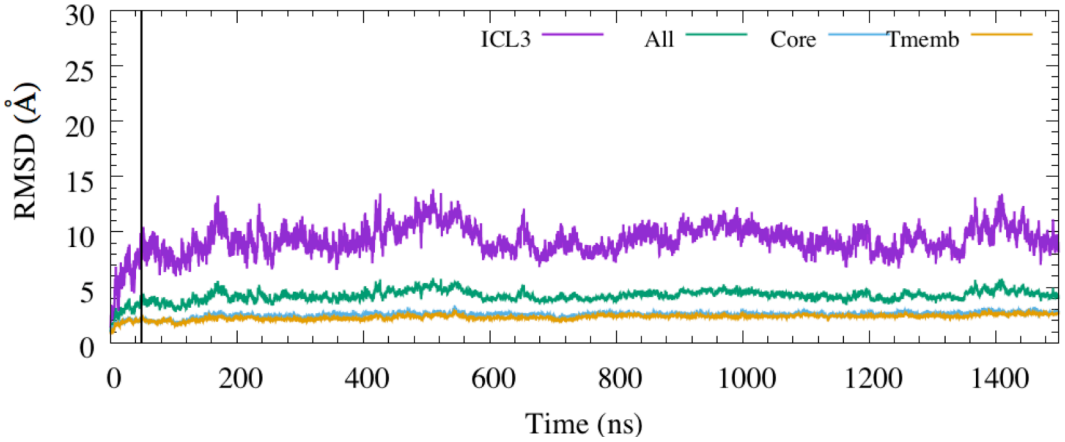
Figure 5. 17 Schematic illustration of active of β_2AR in complex with trimeric G protein (PDB ID:3SN6) (Rasmussen *et al.,* 2011). The Figure has been drawn with the graphics program VMD v1.9.3 (Humphrey *et al.,*1996).

5.1.3.1 RMSD and RMSF profiles in active receptor state

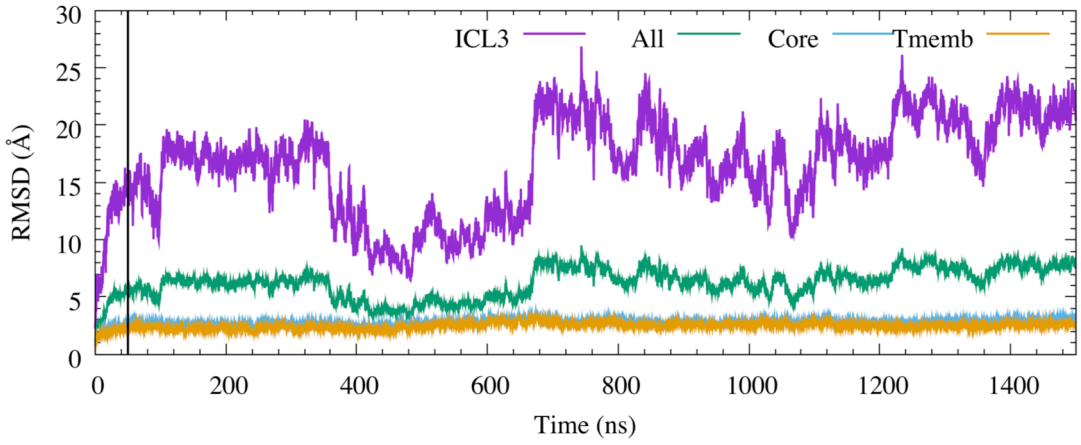
The active state of the receptor, together with the G protein, presented totally different characteristics compared to the free *intermediate* and inactive receptors. The interaction of the receptor with the α5-helical structure in the C-terminal of the Ras-like subunit in Gαs prevented the intracellular packing of ICL3 and intracellular transmembrane 6 in the active state but did not restrain the ICL3 movement, albeit maintain its outward tilted fluctuations. In the absence of trimeric G-protein, *intermediate* state, however, the receptor tended to form inactive conformation by a rapid closure of ICL3 through the core of the protein. In the *intermediate* state, ICL3 formed transient hydrogen bonds with ICL1, H7, and H8; however, these bonds were broken after ~1.5 μs. Still, the receptor stayed in an *intermediate* state with slightly outward-tilted intracellular H6 and semi-packed ICL3, which was fluctuating nearby H7 and H8 (Figure 5.11).

To evaluate the intrinsic causality of interacting residues in a state where ICL3 was stayed open, mutual information and information transfer analysis were performed on two independent MD simulations with different states and velocities. Data were collected every 0.2 ns, and a total of 7500 frames for each run was used to represent each 1.5 μs long trajectory. The system of the first run was equilibrated in 50 ns, and the remaining 1450 ns trajectory per simulation was used as a production phase. The same number of frames was also collected from the second run for comparison of two active states.

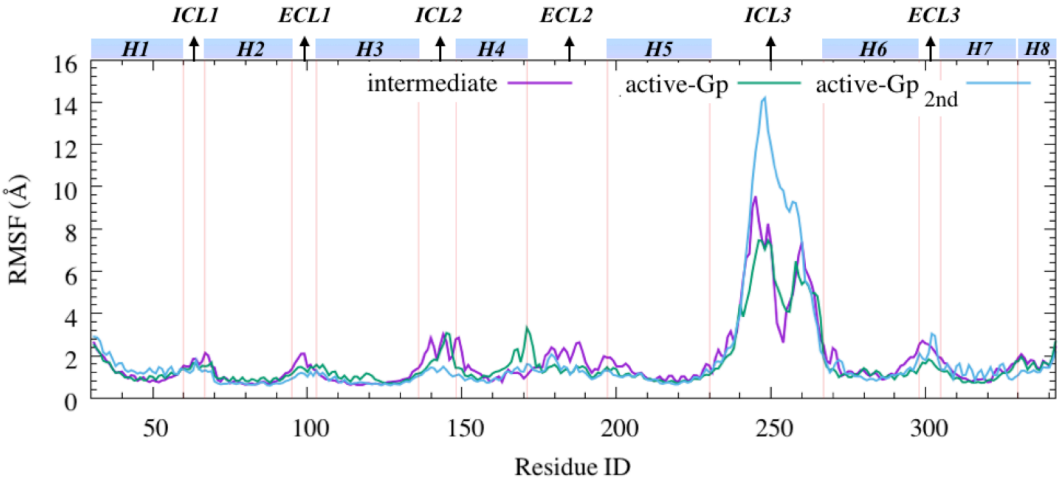
In Figure 5.18a and 5.18b, two distinct dynamics of the active state with bound G protein were detected. The system shown in Figure 5.18a was relatively stable, including ICL3, whereas the one represented in Figure 5.18b, ICL3 was fluctuated throughout the trajectory and not fully stabilized.



(a)



(b)



(c)

Figure 5. 18 RMSD of the first (a) and second (b) *active-Gp* runs. RMSF plot for *intermediate* state together with *active-Gp* and *active-Gp*$_{2nd}$ MD runs (c).

The RMSF profiles in Figure 5.18c indicated that the lower the fluctuation of free ICL3 was, the higher the fluctuation of ICL2 and the residues of ECL2 adjacent to H4. However, in the presence of G protein, the lower ICL3 fluctuation also meant higher fluctuations in ECL2 and also in $H5_{Ext}$. In the simulation of *active-Gp*$_{2nd}$, high fluctuation of ICL3 also induced small fluctuations in ECL3 and H7. Based on ECL2, where inactive phases and *intermediate* phase had higher fluctuations, the fluctuation of it in *active-Gp* was almost zero. When G protein was present and interacted with the receptor intracellularly, the extracellular part of the receptor stayed more rigid than a free *intermediate* state.

Additionally, the deviation of $H7_{Int}$ and *Arg*131-*Leu*272 distance with respect to the inactive state was also examined to compare the conformational changes of *active-Gp* with respect to the experimentally resolved active and inactive conformations. It was reported in the literature that $H7_{Int}$ makes an inward motion, and the distance between $H3_{Int}$ and $H6_{Int}$ will be higher in active receptor compared to the inactive one. Therefore, the result of this analysis yielded what type of structures we had in our hands and how to treat them in terms of information theory-based techniques.

In Figure 5.19a, the structural changes of the $H6_{int}$ region in *active-Gp* and *active-Gp*$_{2nd}$ simulations to resolved active and inactive crystal structures were plotted. The RMSD values were indicating that the *active-Gp* simulation resembles the active state based on $H6_{int}$ deviation. Furthermore, 5.19b and 5.19c presented the distributions of distances between $H3_{Int}$ and $H6_{Int,}$ and the deviation of $H7_{Int,}$ including the *NPxxY* region in two trajectories. The deviation of $H7_{Int}$ and the distance between $H3_{Int}$ and $H6_{Int}$ in *active-Gp* fluctuated around the values of the active receptor's crystal structure, whereas deviation of $H7_{Int}$ through the core of the receptor in *active-Gp*$_{2nd}$ was further increased upon the outward fluctuation of ICL3.

Moreover, Figures 5.19d and 5.19e represented the intracellular views of the *active-Gp* and *active-Gp*$_{2nd}$ states. The orientation of $H6_{Int}$ was similar, but the outward tilt of ICL3 was clearly different than that of *active-Gp*$_{2nd}$. ICL3 fluctuated through a narrow window in *active-Gp*. However, the deviation of ICL3 was in the z-direction and also

away from the G protein associated site in *active-Gp*$_{2nd}$. $H7_{Int}$ was also affected by the fluctuation of ICL3 as given in 5.19b and 5.19c. Based on the comparison of these parameters, one can suggest that the intracellular parts of *active-Gp* and *active-Gp*$_{2nd}$ resembled more like the active crystal structure, however, the extracellular parts, as shown in Figure 5.20a, still spanned the intermediate conformations, since the stabilizing effect of an agonist was missing.

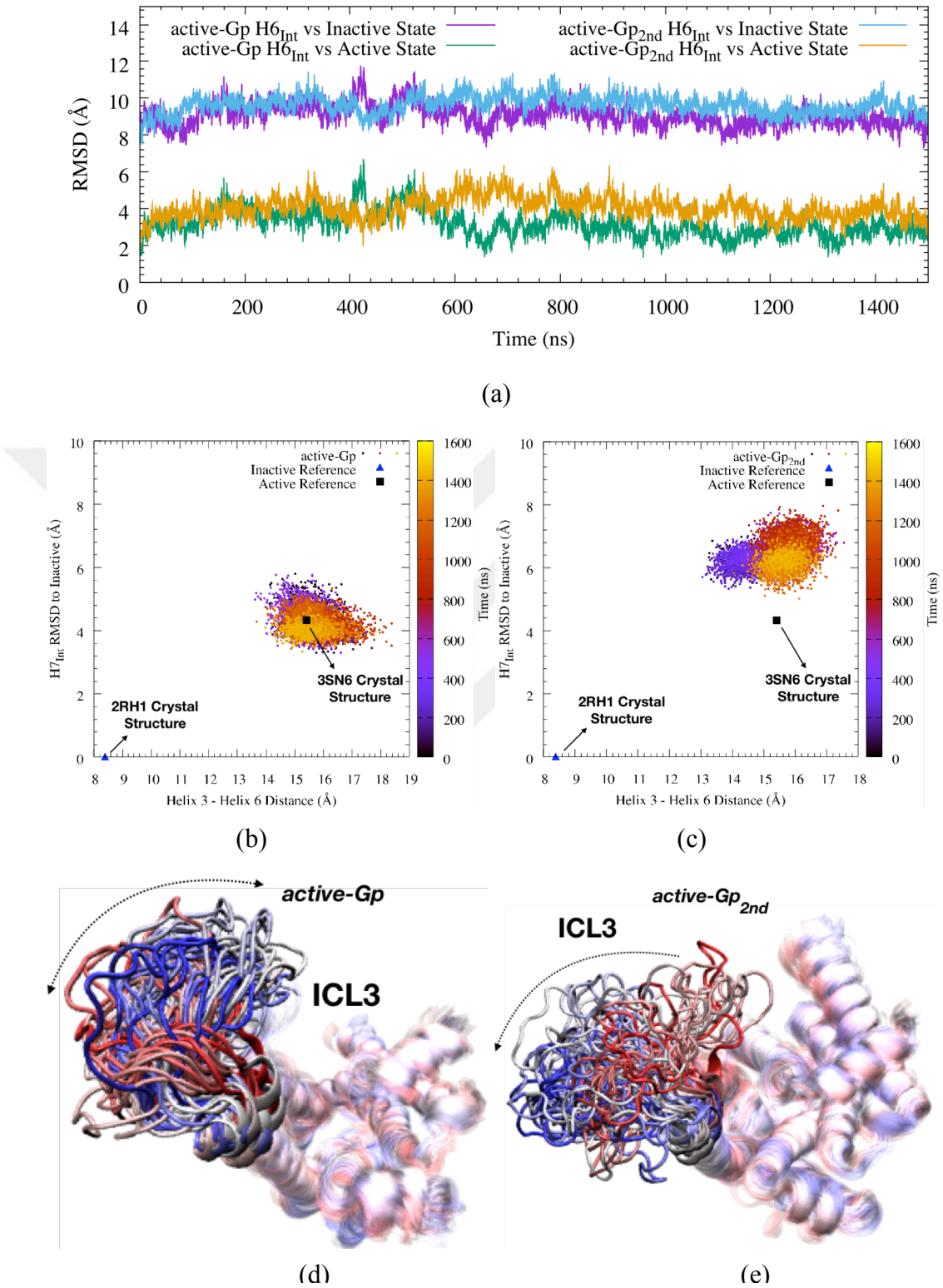
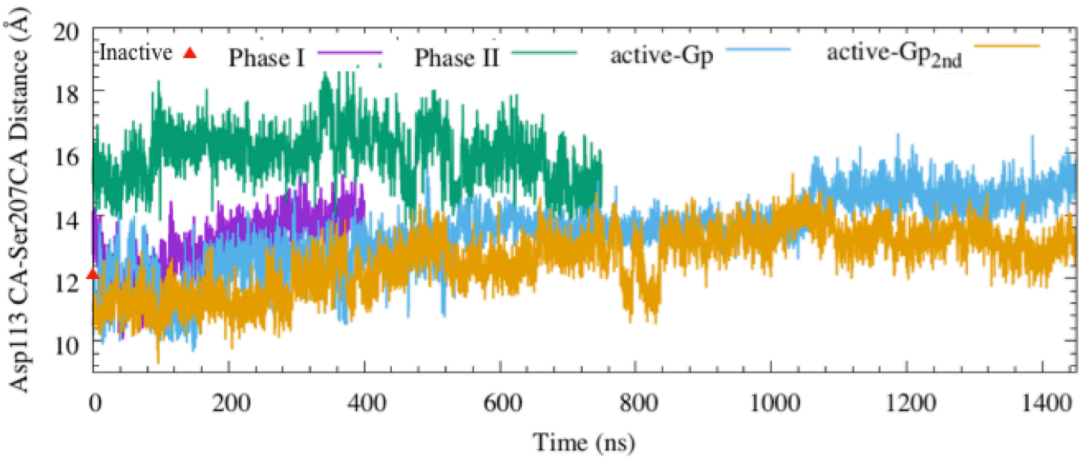


Figure 5. 19 The deviation of intracellular H6 with respect to active and inactive conformations (a). *Arg*131-*Leu*272 distance versus RMSD of $H7_{Int}$ region with respect to inactive crystal structure PDB ID: 2RH1 in *active-Gp* (b) and *active-Gp*$_{2nd}$(c). Intracellular views of trajectories of *active-Gp* (d) and *active-Gp*$_{2nd}$ (e). From red-white-blue, the frame number increases. The dashed arrows represented the directions of fluctuation of ICL3.

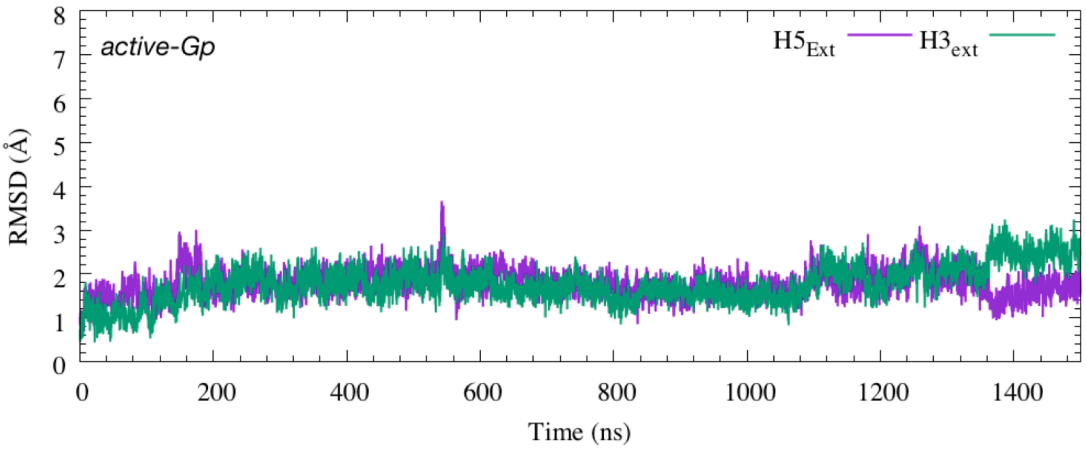
5.1.3.2 The dynamics of extracellular and intracellular parts of the *active-Gp* and *active-Gp$_{2nd}$*

The intracellular and extracellular parts of the receptor were moving in opposite directions, that is, when the intracellular part narrows down, the extracellular orthosteric ligand binding site expands, and vice versa. It was reported in the study of Ozcan and her coworkers that, in the very inactive state, where ICL3 was packed through the core of the receptor, the extracellular part of the receptor expanded. The distances between the orthosteric ligand bindings sites were higher than the native active state, indicating a zipper-like motion in the receptor dynamics. Here, the distance profiles of two critical residues, *Asp*113 and *Ser*207 were compared. These residues were known to involve in ligand binding. The corresponding distance in the inactive state crystal structure was ~12 Å, which was shown with the red triangle in Figure 5.20a.

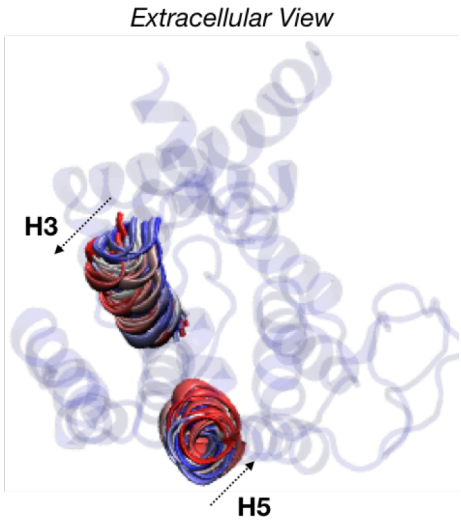
As in Figure 5.20a, the distance between Cα atoms of *Asp*113 and *Ser*207 was higher in Phase I, and Phase II compared to active phases. This distance displayed an increasing trend in the first 400 ns and was higher in Phase I compared to active trajectories. Moreover, it was as high as 18 Å in Phase II, while this distance in *active-Gp* runs increased slowly up to 16 Å at the end of the simulations. Even though this distance was increasing throughout the simulation, it did not reach the maximum value of Phase II. This distance increase in *active-Gp* was due to the increased deviation of both $H3_{Ext}$ and $H5_{Ext}$ in opposite directions. Even though the RMSD of $H5_{Ext}$ in *active-Gp* was not as high as the RMSD of $H5_{Ext}$ in the *intermediate* state, its coordinated motion with $H3_{Ext}$ managed to increase the distance between two critical residues in orthosteric ligand binding site (Figures 5.19b and 5.19c). This increase clearly indicates that the binding of the G protein solely is not sufficient to keep the orthosteric ligand binding site narrow as it is in the active state. Obviously, the receptor must be associate with both a small ligand molecule extracellularly and the G protein intracellularly in order to maintain it in the active state. This striking fact was observed in both of the simulations.


Inactive
Phase I
Phase II
active-Gp
active-Gp2nd
Asp113 CA-Ser207CA Distance (Å)
Time (ns)

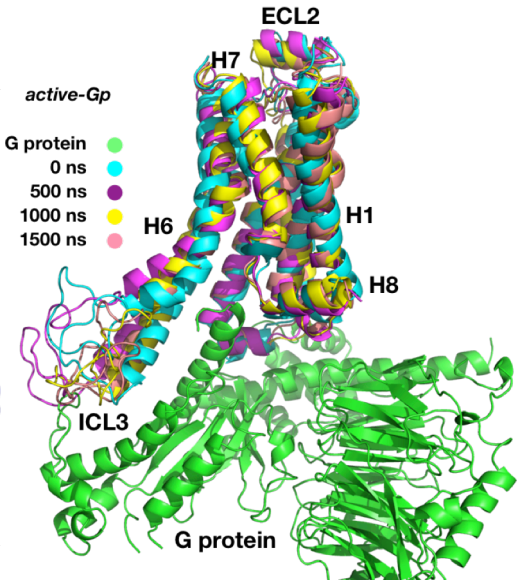
(a)


active-Gp
H5Ext
H3ext
RMSD (Å)
Time (ns)

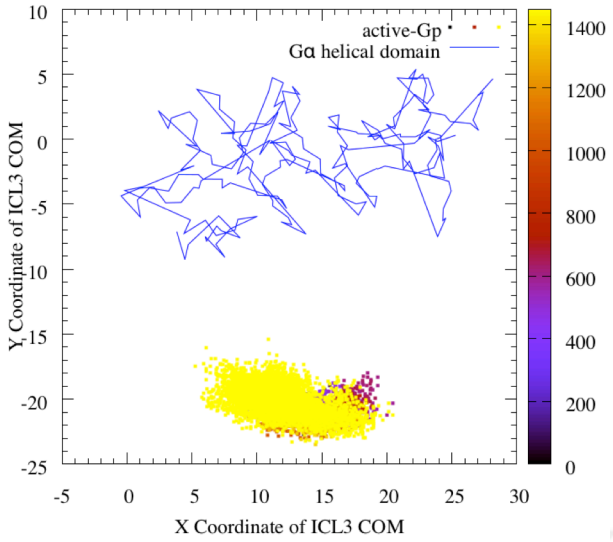
(b)


Extracellular View
H3
H5

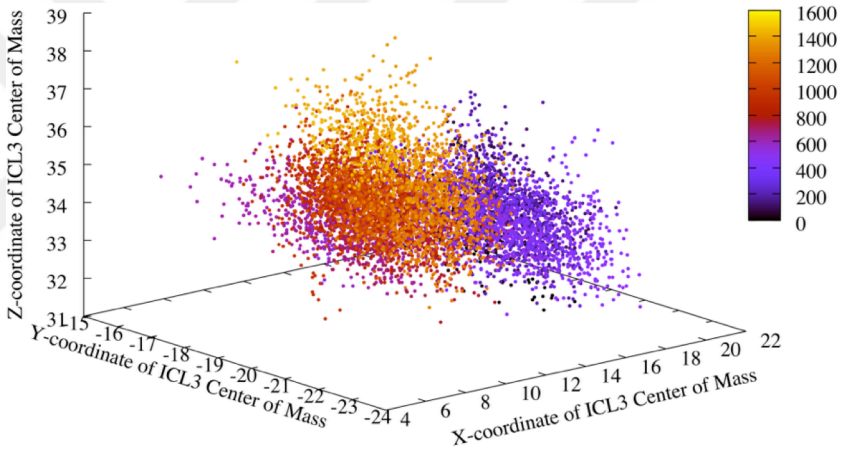
(c)


ECL2
H7
active-Gp
G protein
0 ns
500 ns
1000 ns
1500 ns
H6
H1
H8
ICL3
G protein

(d)



(e)



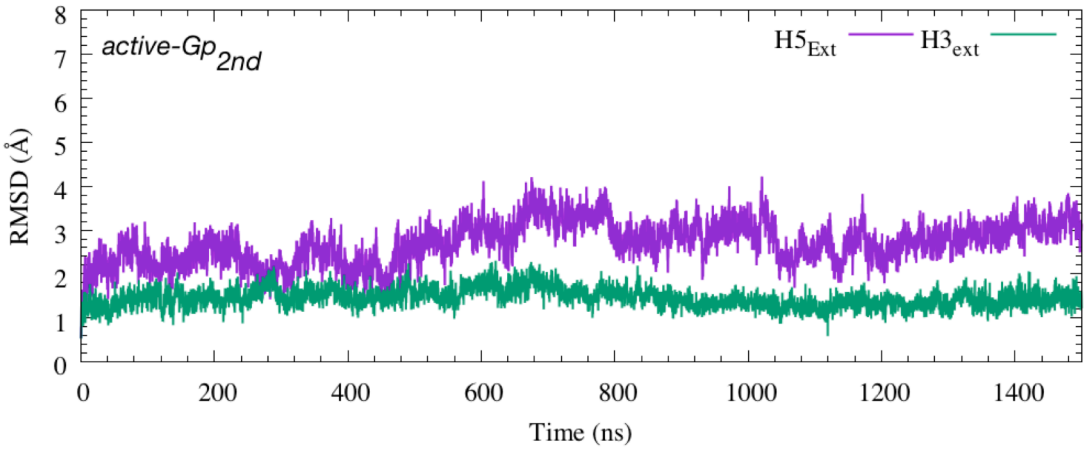
(f)

Figure 5. 20 The deviation of the distance between *Asp*113 Cα atom and *Ser*207 Cα atom compared in Phase I, Phase II, *active-Gp* and *active-Gp*$_{2nd}$ trajectories (a). The RMSD change of H5$_{Ext}$ and H3$_{Ext}$ (b). The extracellular view of the receptor depicting the mobility of H5$_{Ext}$ and H3$_{Ext}$ (c). The center of mass of ICL3 with respect to the Gα helical domain in *active-Gp* (d). Superposition of frames representing 0 ns, 500 ns, 1000 ns and 1500 ns of *active-Gp* trajectory with G protein (e). Cartesian coordinates of ICL3's center of mass in the *active-Gp* trajectory (f).

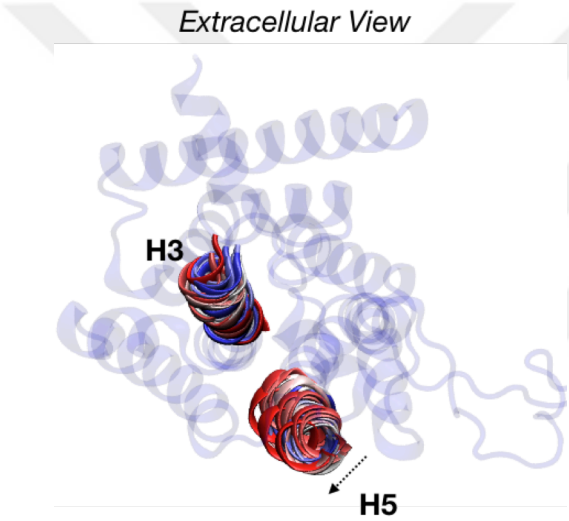
Figure 5.20d demonstrated the x- y- coordinates of ICL3's center of mass deviation with respect to the coordinates of Gα helical structure and 5.20c also represents the deviation in z-direction in *active-Gp* run. The mobility of ICL3 center was mostly in x-direction in the *active-Gp* with around 15 Å movement, while the change in z-direction was not as high as in the change in z-direction in *active-Gp*$_{2nd}$ as shown in Figure 5.21e.

Figure 5.21a represented RMSD of $H3_{Ext}$ and $H5_{Ext}$ in *active-Gp$_{2nd}$* trajectory. Here, there was around 2 Å RMSD change in $H5_{Ext}$; however, $H3_{Ext}$ was more stable than $H5_{Ext}$ in here and interestingly compared with RMSD of $H3_{Ext}$ in *active-Gp*. The increase in the distance between Asp113 Cα atom and Ser207 Cα, shown in Figure 5.20a, was based on the deviation of the $H5_{Ext}$ (Figure 5.21b). The change of ICL3's center of mass in x- and y-direction with respect to Gα helical domain of G protein in *active-Gp$_{2nd}$* trajectory was shown in Figure 5.21c and 5.21d. As shown in these figures below, the positioning of ICL3 did not overlap with the G protein; in contrast, moved away from the helical extention of its Gα domain in three directions. This mobility was not purely perpendicular to the membrane, but in all three x-, y- and z- directions. There were around 15 Å shift in x- direction, 5 Å shift in y-direction, and 9 Å shift in z-direction in the center coordinates of the ICL3 (Figure 5.21e).

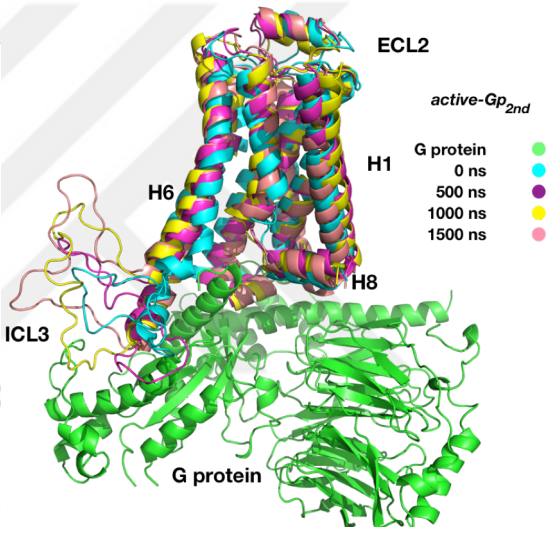
Obviously, the fluctuation of ICL3 in *active-Gp$_{2nd}$* in the z-coordinate was higher than the fluctuation in *active-Gp*, while the outward deviations of x- and y- coordinates of the loop were similar. The presence of Gα helical structure prevented ICL3 to pack itself through the core of the receptor, as it was in the *intermediate* state and Phase II, thus it stayed in the extended conformation.


active-Gp2nd
H5Ext
H3ext
RMSD (Å)
Time (ns)

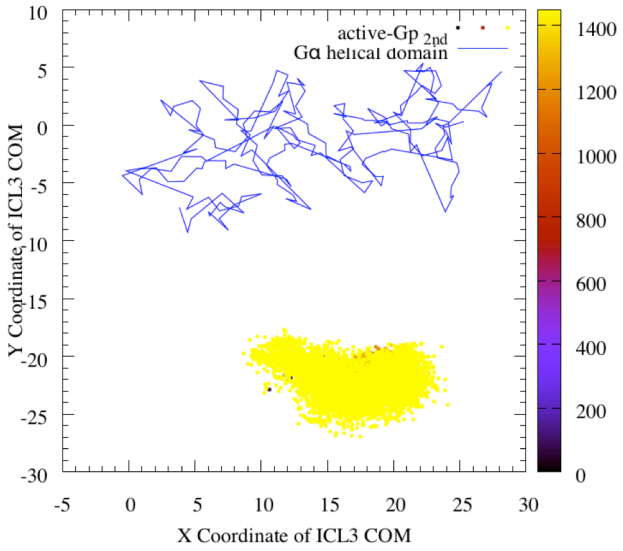
(a)


Extracellular View
H3
H5

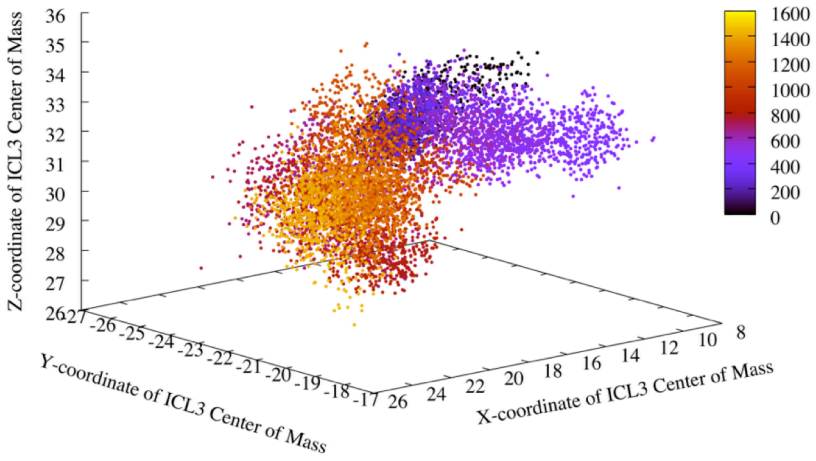
(b)


ECL2
active-Gp2nd
G protein
0 ns
500 ns
1000 ns
1500 ns
H1
H6
H8
ICL3
G protein

(c)


active-Gp 2pd
Gα helical domain
Y Coordinate of ICL3 COM
X Coordinate of ICL3 COM

(d)

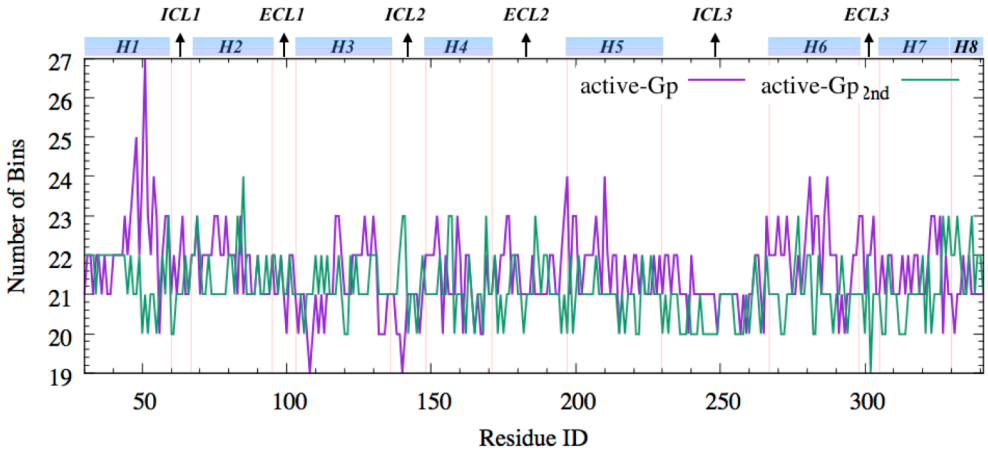


(e)

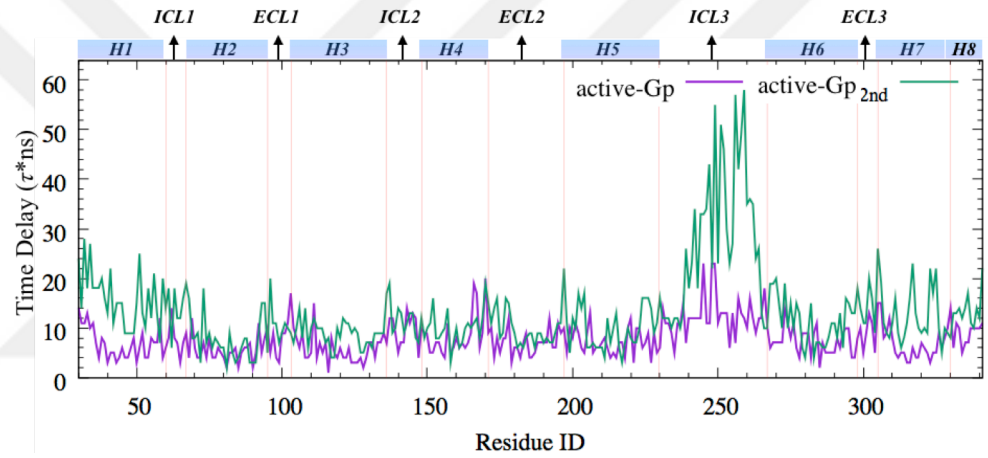
Figure 5. 21 The RMSD of the $H5_{Ext}$ and $H3_{Ext}$ in *active-Gp*$_{2nd}$ (a). The extracellular view of protein representation the RMSD deviation of $H5_{Ext}$ and $H3_{Ext}$ (b). Active receptor representation with 5 individual snapshots together with G protein. 0 ns was colored with green, 500 ns with pink, 100 ns with yellow, and finally, 1500 ns was colored with violet. G protein was colored in green (c). X- and y- coordinates of the center of mass of ICL3 in *active-Gp*$_{2nd}$ (d) Three dimensional plot of ICL3's center of mass change (e).

5.1.3.3 Optimal number of bins and time delay for *active-Gp* and *active-Gp*$_{2nd}$

The optimized number of bins for the active states was in between 19 and 27 in *active-Gp* and between 19 and 24 in *active-Gp*$_{2nd}$ (Figure 5.22a). The delay time τ was optimized for each residue by calculation of the first minimum of mutual information for each residue (Cellucci *et al.,*2003; Noakes, 1991; Sauer *et al.,*1991) Similar to the number of bins optimization, the delay time is also needed to be residue-specific since each residue has its own characteristic mobility in time. If τ is taken very small, then $TE_{i\to j} - TE_{j\to i}$ would be very small due to the persistence of autocorrelations of residue-pairs. Similarly, if τ is taken very large than the characteristic delay time, then $TE_{i\to j}$ values will be minimal since the correlations will be vanishingly small. Our residue-specific τ values were between 1 ns to 23 ns for the *active-Gp* and between 2 to 58 in *active-Gp*$_{2nd}$. As was seen in Figure 5.22b, the higher delay time was observed for residues with higher mobilities.



(a)



(b)

Figure 5. 22 Residue optimized number of bins (a) and time delay (b) for *active-Gp* (magenta) and *active-Gp*$_{2nd}$ (green)

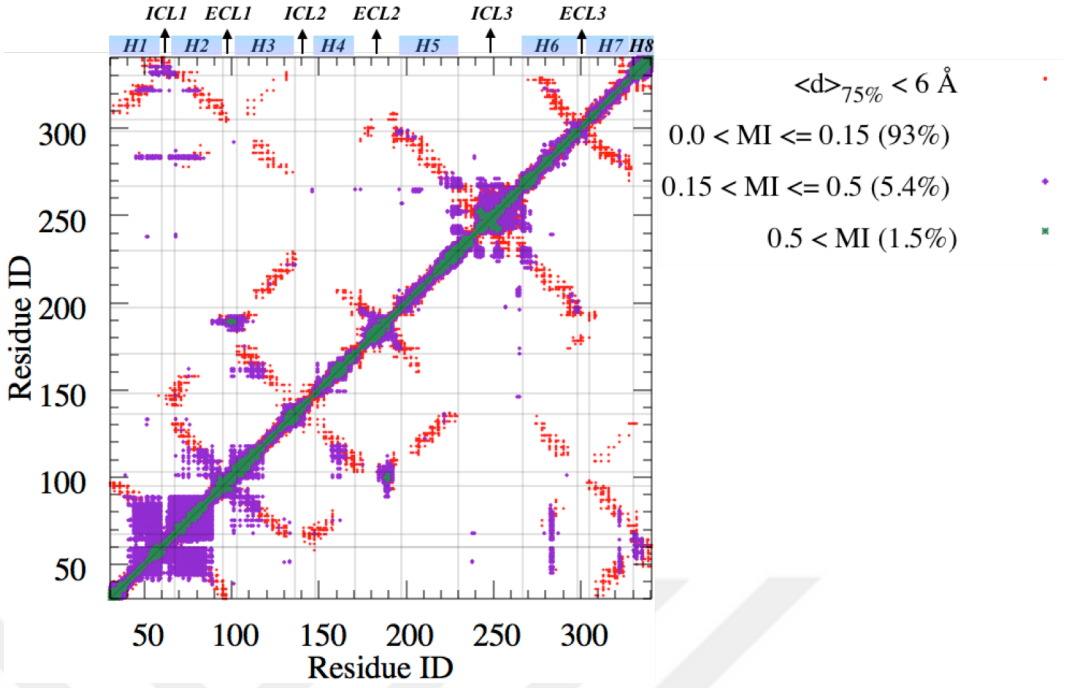
5.1.3.4 Proximal and distal communications in the presence of G protein

MI values above 0.15 were plotted in order to eliminate small allosteric interactions and for comparison of the data with the previous findings. Figure 5.23a and 5.23b show residues with strong allosteric communication, which were colored in colors magenta and green based on the corresponding MI values. Even though diagonal residues and spatially proximal residues had the highest MI values due to being directly in contact, it was observed that off-diagonal, distant, but allosterically communicating residue-pairs also experienced information sharing.

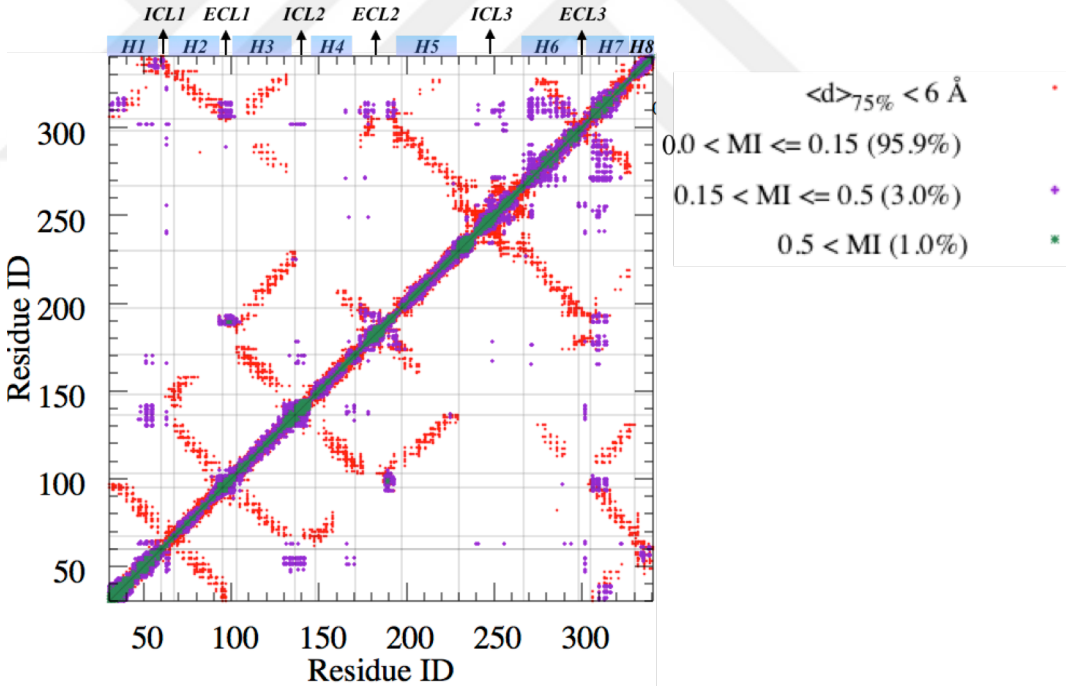
Interestingly, medium and high MI values were less in percentage in *active-Gp$_{2nd}$* run compared to Phase I, Phase II, *intermediate* state, and *active-Gp*. The percentage of MI above 0.15 was 6.9% in the *active-Gp* run and 4.0% in the *active-Gp$_{2nd}$*; however, they were 8.2% in the *intermediate* state, 6.2% in Phase I, and 4.2% in Phase II. Even though the fluctuation of ICL3 perpendicular to the membrane is higher in *active-Gp*$_{2nd}$ compared to *active-Gp,* its horizontal motion parallel to the membrane was similar in both runs.

The motion in z-direction was apparently reduced the communication of ICL3 with the remaining parts of the receptor. This lack of communication between ICL3 and the remaining parts of the receptor was reflected to the amount of information sharing, thus decreasing the percentage of the residue-pairs with MI values above 0.15 in *active-Gp*$_{2nd}$. Mostly, high information sharing pairs clustered around H1, H2, and few residues in H3 and H4, as well as between ECL1 and ECL2 in the *active-Gp* run. Additionally, H6 and H7 were observed to contribute to communication in *active-Gp*$_{2nd}$; however, the value was not as high as the MI value observed around H1 and H2 in *active-Gp*. Finally, residues *Ala*198 in ECL2 have strongly shared information with *Met*98 and *Trp*99 in ECL1, in *active-Gp$_{2nd}$* with a value of above 0.5, which indicates strong communication in the extracellular parts.

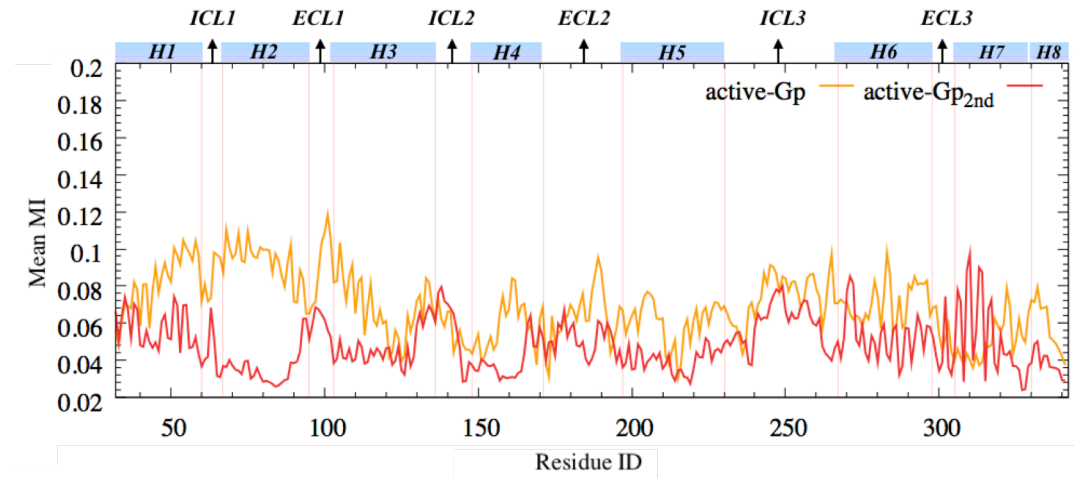
The high fluctuation resulted from the conformational change of ICL3 in the perpendicular direction with respect to the membrane (Figure 5.21), while the motion in the horizontal direction was restricted by the α-helical structure of Gα domain and stabilized. This relative stability in the horizontal direction of the ICL3 fluctuation, as it was observed in *active-Gp$_{2nd}$* and Phase II, reduced the information-sharing feature of ICL3.


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Residue ID
Residue ID
50
100
150
200
250
300
<d>75% < 6 Å
0.0 < MI <= 0.15 (93%)
0.15 < MI <= 0.5 (5.4%)
0.5 < MI (1.5%)

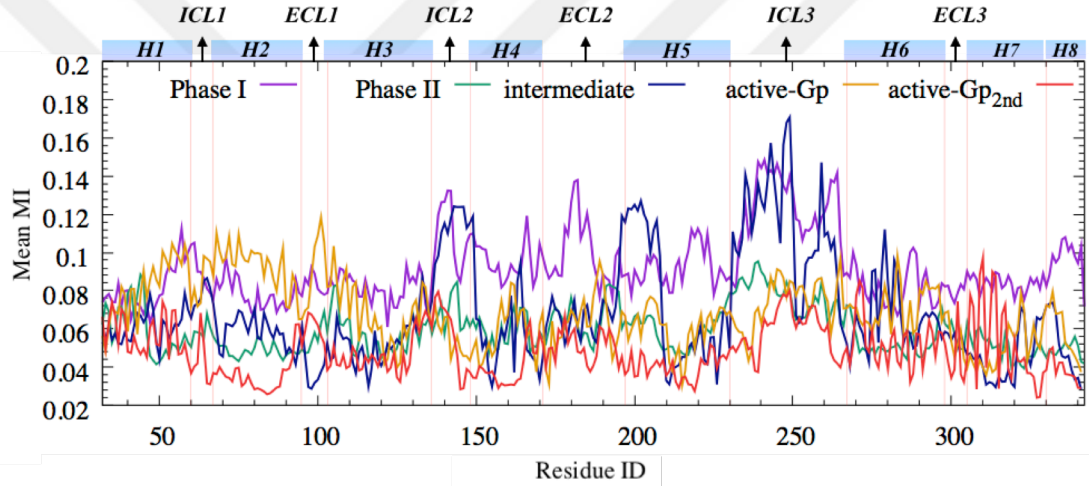
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Residue ID
Residue ID
50
100
150
200
250
300
<d>75% < 6 Å
0.0 < MI <= 0.15 (95.9%)
0.15 < MI <= 0.5 (3.0%)
0.5 < MI (1.0%)

(b)



(c)



(d)

Figure 5. 23 Pairwise MI in active states. MI > 0.15 was plotted from purple to green for *active-Gp* (a) and *active-Gp*$_{2nd}$ (b). Mean MI values for each residue for *active-Gp* (orange) and *active-Gp*$_{2nd}$ (red) (c). Mean MI values for all five states. Phase I (magenta), Phase II (green), intermediate (blue), *active-Gp* (orange), and *active-Gp*$_{2nd}$ (red) were given in (d).

Overall, as shown in figure 5.23d, the average mutual information decreased towards active phases. The highest mean MI belonged to Phase I (0.092 ± 0.018), followed by *active-Gp* (0.069 ± 0.018) and *intermediate* state (0.068 ± 0.028), and finally Phase II (0.059 ± 0.012) and *active-Gp*$_{2nd}$ (0.049 ± 0.014). Except for H1, H2, and few residues on H3, the average MI value in *active-Gp* was significantly low. Moreover, in Phase II, which was a unique highly inactive conformation, the stiffening of ICL3, and thus the overall protein, strongly affected the pairwise dependency. Therefore, one can suggest that when the activity of the protein was increased, then the communications were

restricted based on the finding that the increased activation of the receptor decreases the distal residues' information sharing (Vaidehi and Bhattacharya, 2016).

5.1.3.5 The effect of associated G protein on the polarity

Polar residues' contribution to information sharing compared to the hydrophobic residues was also analyzed for active states. As it was plotted in Figure 5.24 and Figure 5.25, polar and hydrophobic residues mean MI values were different in both proximal and distal pairs. As given in Table 5.5, average MI for proximal residues in *active-Gp* was 0.53 ± 0.28 for polar residues, and 0.48 ± 0.24 for hydrophobic residues. Among distal residues, on the other hand, polar residues had 0.057 ± 0.049 average MI, while hydrophobic residues had 0.058 ± 0.046. In *active-Gp_{2nd}*, these values were 0.40 ± 0.28 and 0.38 ± 0.25 for proximal residues and 0.041 ± 0.032 and 0.039 ± 0.033 for distal residues, respectively. The statistical validation for these comparisons revealed that the proximal residues' polar and hydrophobic MI distributions for *active-Gp_{2nd}* were not statistically different (H = 0.76, p = 0.38) as given in Appendix H. On the other hand, the distal residues' polar and hydrophobic MI distributions were statistically different as well as those in *active-Gp*; thus, the differences in average MI values were relevant. Interestingly, unlike previous states, the average MI for distal residues had a higher value for the hydrophobic ones compared to polar ones. Thus, intracellularly associated G protein might had an effect on the pairwise communications of active states.

Additionally, in *active-Gp*, proximal amino acid pairs of polar–polar characteristics had an average MI value of 0.57 ± 0.29, polar-hydrophobic characteristics had an average MI value of 0.51 ± 0.28, and hydrophobic-hydrophobic characteristics had an average MI value of 0.47 ± 0.22. Distal amino acid pairs, on the other hand, had 0.058 ± 0.052, 0.056 ± 0.046, and 0.059 ± 0.047 average MI values, respectively (Table 5.6). In proximal amino-acid pairs, *active-Gp_{2nd}* showed that polarity had a strong effect on information sharing. However, in the distal pairs, hydrophobic amino acids were dominated. Average MI for polar-polar pairs was lower than hydrophobic-hydrophobic pairs, but still, they were higher in polar-hydrophobic pairs between distal amino acids. Moreover, in *active-Gp_{2nd}*, both proximal and distal pairs had a higher mean MI value

for polar pairs. However, the difference in the distribution of proximal pairs was not statistically significant (H = 4.94, p = 0.084).

Observing polarity effect in Phase I, Phase II, and *intermediate* state, but losing it in *active-Gp* and *active-Gp$_{2nd}$* clearly pointed out the influence of G protein on the pairwise communications. The systems of *active-Gp* and *active-Gp*$_{2nd}$ were associated with trimeric G protein, and the receptor dynamics was affected by this large G protein. Thus, especially for polar residues locating intracellularly, their polarity might be affected by the electrostatic interactions with the amino acids of the G protein, and may not be able to present their polar characteristics. This effect was clearly observed in MI values of ICL2 and ICL3 in Figure 5.22c, that none of these two loops demonstrated strong information sharing characteristics, as they were in *intermediate* state and Phase I. Moreover, high information sharing in these two states was observed between transmembrane helices. Since transmembranes were rich in hydrophobic amino acids, these higher MIs in helices increased the possibility of observing more hydrophobic residues in the communication pathway.

Table 5. 5 The comparison of average mutual information of polar and hydrophobic residues in Active receptor with and without G protein

	Type	Proximity	Mean Mutual Information	# of Residues
intermediate	Polar	Proximal	0.57 ± 0.36	915
		Distal	0.064 ± 0.070	34850
	Hydrophobic	Proximal	0.46 ± 0.29	1561
		Distal	0.052 ± 0.053	59706
active-Gp	Polar	Proximal	0.53 ± 0.28	915
		Distal	0.057 ± 0.049	34850
	Hydrophobic	Proximal	0.48 ± 0.24	1561
		Distal	0.058 ± 0.046	59706
active-Gp$_{2nd}$	Polar	Proximal	0.40 ± 0.28	915
		Distal	0.041 ± 0.032	34850
	Hydrophobic	Proximal	0.38 ± 0.25	1561
		Distal	0.039 ± 0.033	59706

Table 5. 6 The comparison of average mutual information of polar and hydrophobic residue-pairs in Active receptor with and without G protein

	Type	Proximity	Mean Mutual Information	# of Residues
intermediate	Polar-Polar	Proximal	0.64 ± 0.38	370
		Distal	0.075 ± 0.084	12740
	Polar-Hydrophobic	Proximal	0.52 ± 0.34	1090
		Distal	0.058 ± 0.060	44220
	Hydrophobic-Hydrophobic	Proximal	0.43 ± 0.25	1016
		Distal	0.049 ± 0.048	37596
active-Gp	Polar-Polar	Proximal	0.57 ± 0.29	370
		Distal	0.058 ± 0.052	12740
	Polar-Hydrophobic	Proximal	0.51 ± 0.28	1090
		Distal	0.056 ± 0.046	44220
	Hydrophobic-Hydrophobic	Proximal	0.47 ± 0.22	1016
		Distal	0.059 ± 0.047	37596
active-Gp$_{2nd}$	Polar-Polar	Proximal	0.42 ± 0.29	370
		Distal	0.041 ± 0.033	12740
	Polar-Hydrophobic	Proximal	0.38 ± 0.27	1090
		Distal	0.039 ± 0.032	44220
	Hydrophobic-Hydrophobic	Proximal	0.38 ± 0.24	1016
		Distal	0.039 ± 0.034	37596

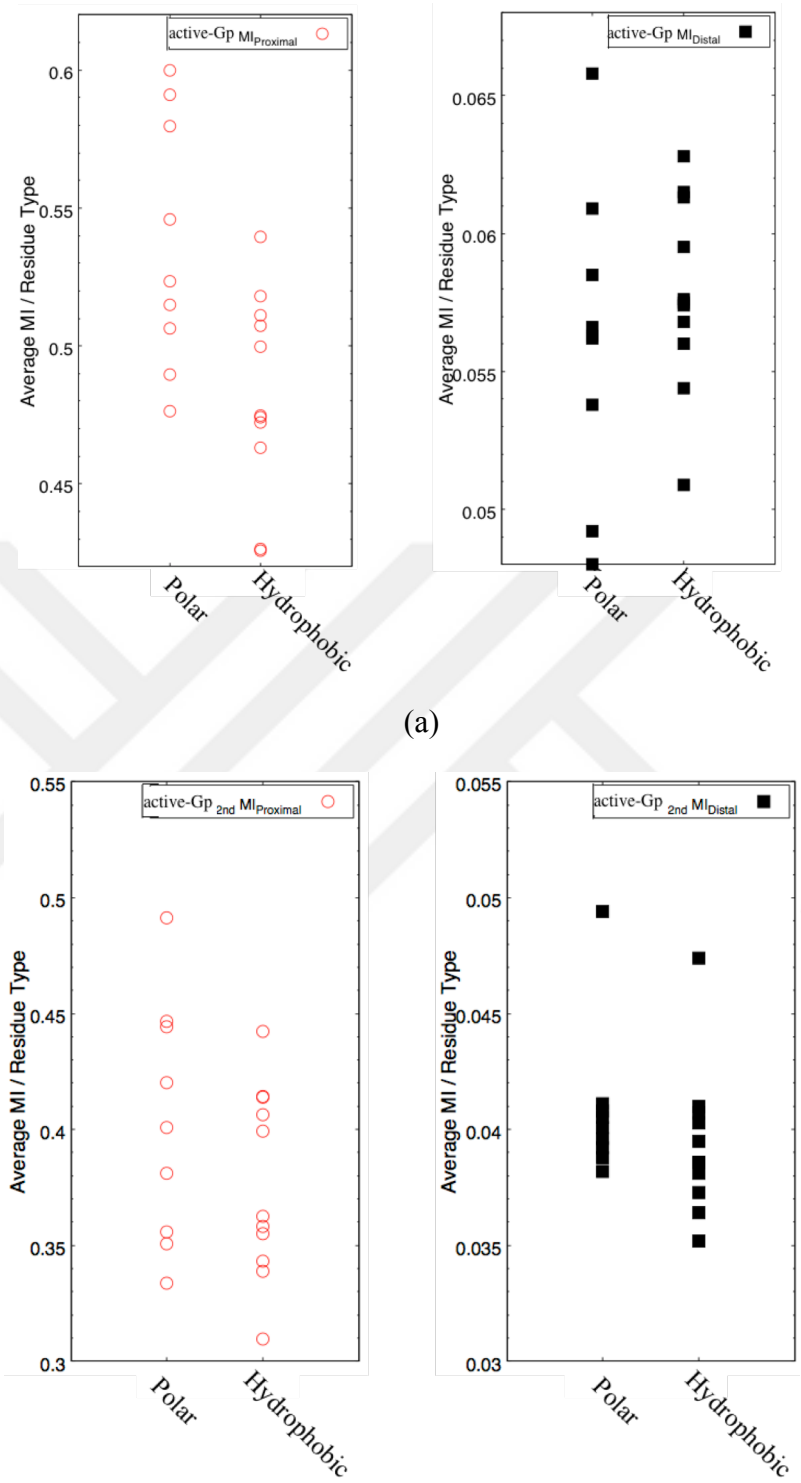
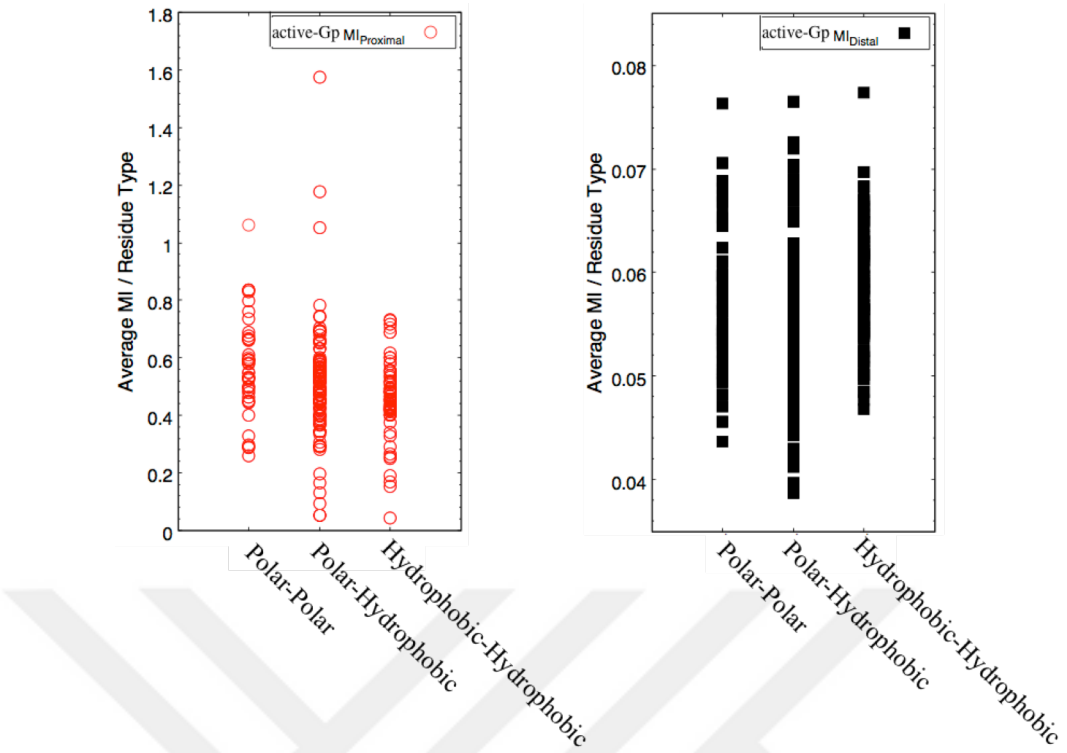
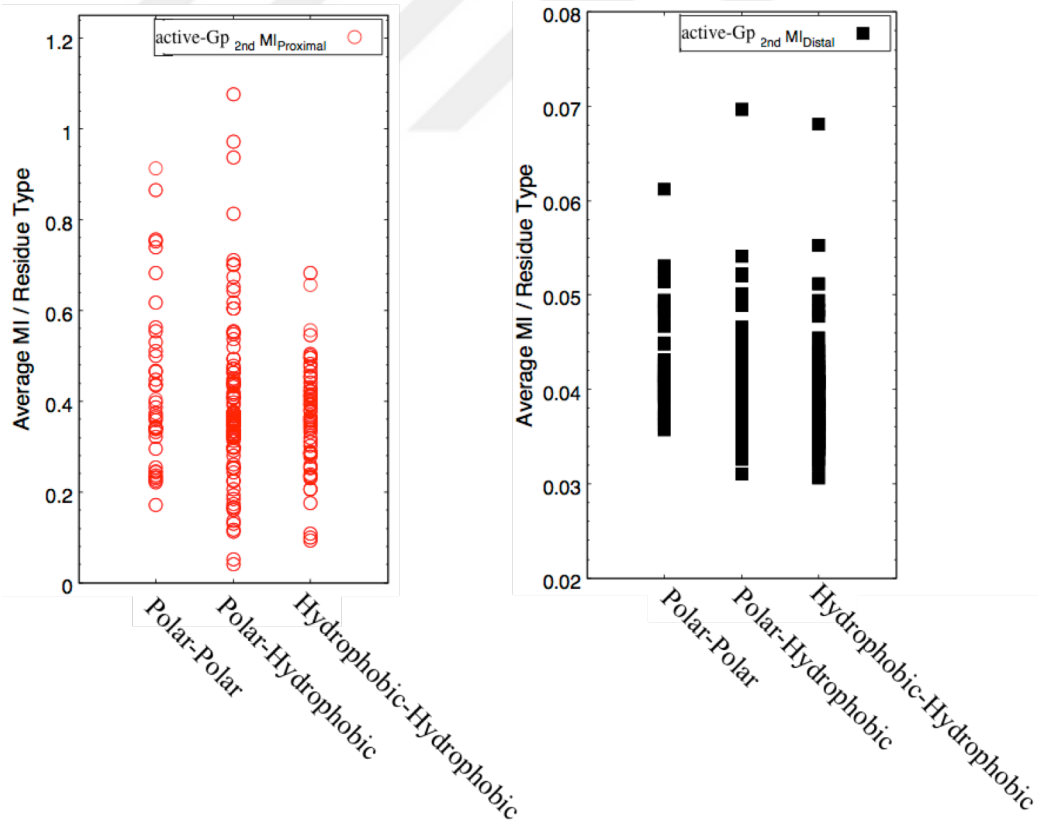


Figure 5. 24 The contribution of polar and hydrophobic residues to MI in *active-Gp* (a) *active-Gp*$_{2nd}$ (b). Proximal residues were colored in red and distal residues were colored in black.



(a)



(b)

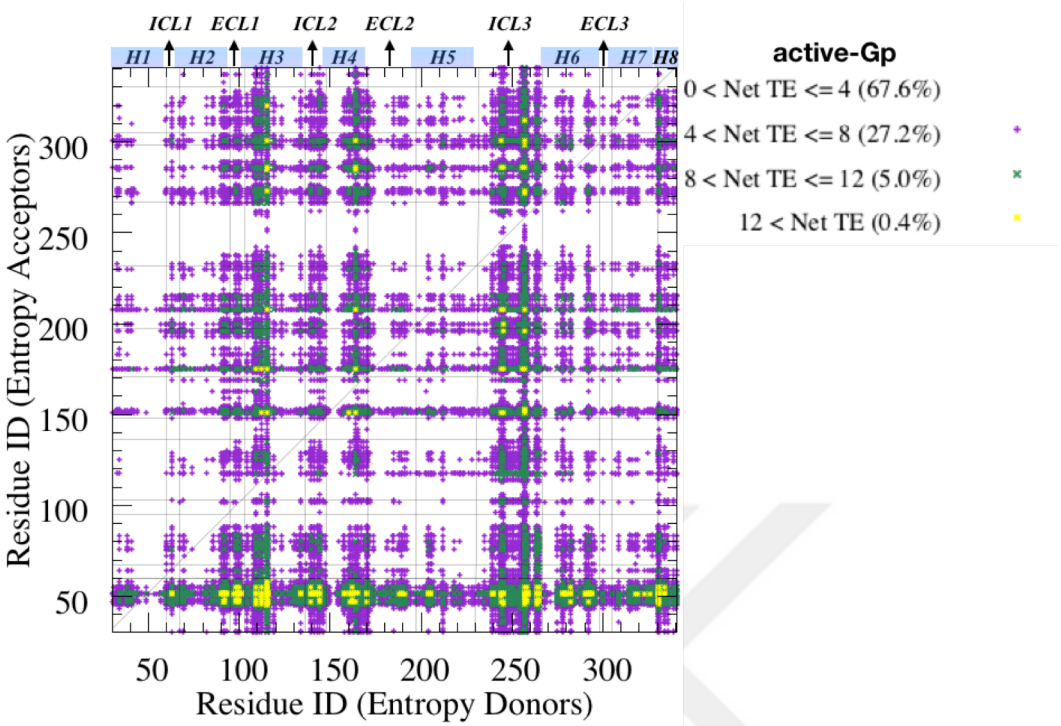
Figure 5. 25 Polar and hydrophobic residue-pairs contribution to MI in *active-Gp* (a) and *active-Gp*$_{2nd}$ (b). Proximal residue-pairs were colored in red and distal ones were in black.

5.1.3.6 Entropy transfer in active states

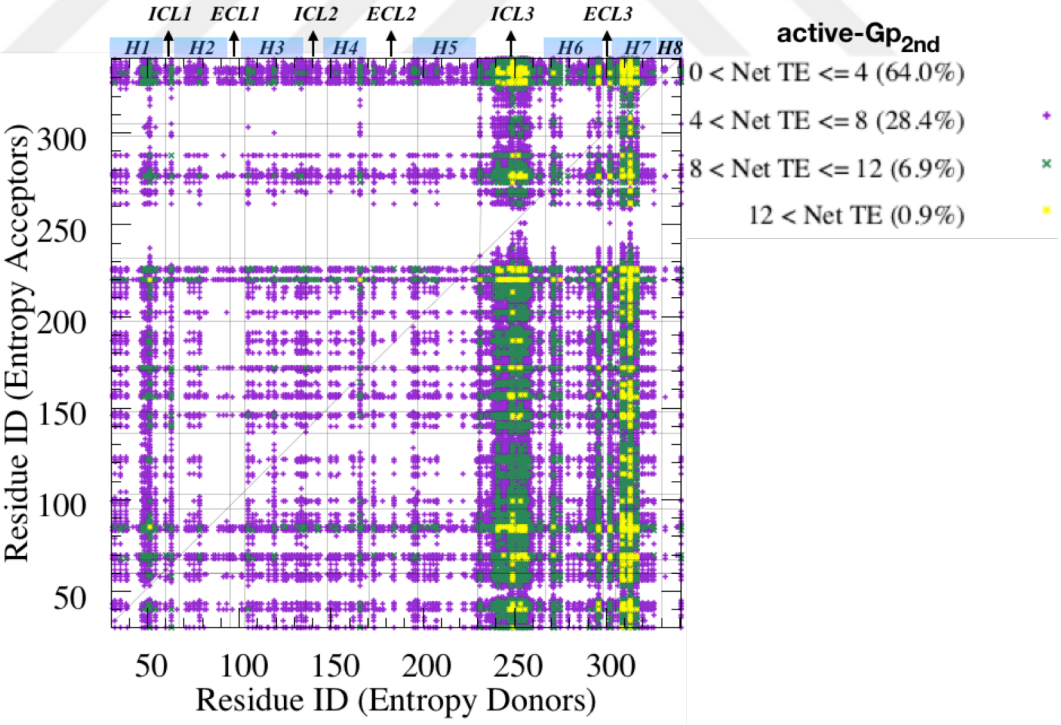
It was shown in our previous paper that ICL3 closure and its strong interaction with the core of the receptor induced a significant flexibility decrease in the overall system (Sogunmez and Akten, 2019), hence the overall number of driver-follower residue-pairs decreased accordingly. This decrease of flexibility due to ICL3 closure in Phase II and due to the restriction of inward motion of ICL3 by the α5-helical subunit of $G_{\alpha}s$ in the *active-Gp* phases were also thermodynamically reflected in transfer entropy.

Interestingly, as previously noted, the percentage of residue-pairs with high mutual information (MI > 0.15) was higher in the *intermediate* state compared to the remaining phases. Also, transfer entropy data displayed similar results, that was the highest percentages (Net TE >8.0) of mean net transfer entropy was belonging to the *intermediate* state with 16.5%, followed by Phase II with 8.0% and *active-Gp*$_{2nd}$ with 7.8%, and then, Phase I with 6.3% and finally *active-Gp* with 5.4%. Interestingly, the high MI percentages in *active-Gp* were higher than *active-Gp*$_{2nd}$ and Phase II. Surprisingly, a completely opposite profile was observed in here. In *active-Gp*$_{2nd}$, the fluctuation of ICL3 in z-direction increased the uncertainty of the loop; thus, it became entropy donor. Therefore, even though the momentarily dependence of ICL3 with other residues was low, it still managed to control the future fluctuations of the receptor. Also, $H6_{Ext}$ and H7 were observed as entropy donors, where their MI values were high, as well. On the other hand, ICL3, as well as ECLs in *active-Gp*, were relatively stable; thus, entropy transfer of this state was the lowest among five states.

Specifically, the net transfer of information, i.e., the gained information after the exchange, was computed by reciprocal subtraction of transfer entropies. From abscissa to ordinate in Figure 5.26a and 5.26b, the entropy donors (sources) in the *active-Gp* phases were shown with colors white, purple, green and yellow based on the increasing values, respectively. The strongest entropy donors were $H3_{Ext}$, ICL2, $H4_{Ext}$, and ICL3, while, $H1_{Int}$ was observed as the major entropy sink in the first run designated as *active-Gp*. In the *active-Gp*$_{2nd}$ trajectory, the highest donors became ICL3, and $H7_{Ext}$, while acceptors were the residues around H8.



(a)



(b)

Figure 5. 26 Entropy donors and acceptor residues in *active-Gp* (a) and *active-Gp*$_{2nd}$ (b).

Observing H1 as a strong acceptor and H2 as an acceptor in *active-Gp* also explains the strength of mutual information between these helices. Since mutual information measures the extent of uncertainty reduction and explains how much a variable tells us about another than the residue with significantly low uncertainty is expected to have high mutual information. Therefore, the low uncertainties of H1 and H2 naturally resulted in high mutual information data in Figures 5.23a and 5.23c.

In the *active-Gp*, where the closure of ICL3 was significantly constrained by the α5-helical bundle of G_{α}s, the percentages of the strong donors were not as high as it was in previous states. As previously noted, the donor characteristics of ICL2, ICL3 were stronger in the *intermediate* phase, where ICL3 was freely moving. However, in *active-Gp_{2nd}*, the perpendicular motion of ICL3 according to the membrane was observed (Figure 5.21), while horizontal motion through the core of the protein was also constrained by the presence of the α5-helical bundle of G_{α}s. The strength of donor characteristics of ICL3 was increased together with $H6_{Int}$ and $H7_{Ext}$ in *active-Gp_{2nd}*. The motion of ICL3, which was the only major difference between two active runs, changed the overall transfer entropy profiles. Even though the motion of ICL3 was independent of the remaining residues, it was still able to transfer strong entropy to the rest of the protein in *active-Gp_{2nd}*. However, the relative stability of ICL3 fluctuation in *active-Gp* lost the entropy donor characteristics of ICL3.

An interesting data was also obtained from the residues $H1_{Int}$, H8, $H7_{Ext}$, $H3_{Ext}$, and $H4_{Ext}$, which have opposite entropy transfer characteristics based on the change in ICL3 mobility between the active states. When ICL3 was more stable in *active-Gp*, $H1_{Int}$ was entropy acceptor, whereas H8 was entropy donor. On the other hand, when it was mobile in *active-Gp_{2nd},* $H1_{Int}$ was converted from acceptor do donor, and H8 was converted from donor to acceptor (Figure 5.27a and Figure 5.27b). A similar effect of ICL3 mobility was also observed in $H7_{Ext}$, where it converted to strong entropy donor upon an increase of the ICL3 mobility.

The comparison of average TE values with the *intermediate* state in Figure 5.27c indicated that the active states had lower TE values compared to the *intermediate* state.

Both donors in ICL2 and ICL3 and acceptors in ECL2 had significantly higher donor and acceptor characteristics in the *intermediate* state. Interestingly $H1_{Int}$, which was neutral in the *intermediate* state, became acceptor in the *active-Gp* and donor in the *active-*Gp_{2nd}. *active-*Gp_{2nd} was separated from these two states in terms of entropic features of ICL2, ECL2 and $H7_{Ent}$, where these regions acted similarly in the *active-Gp* and the *intermediate* state.

Finally, we also compared active states net TE values with inactive states in Figure 5.27d. The residues of ICL3 adjacent to H6 and residues of $H6_{Int}$ were entropy acceptors in inactive states. Interestingly, all amino acids on ICL3 were donors in active states. The entropy acceptor characteristics of the residues on $H6_{Int}$ were lower in active states compared to inactive states.

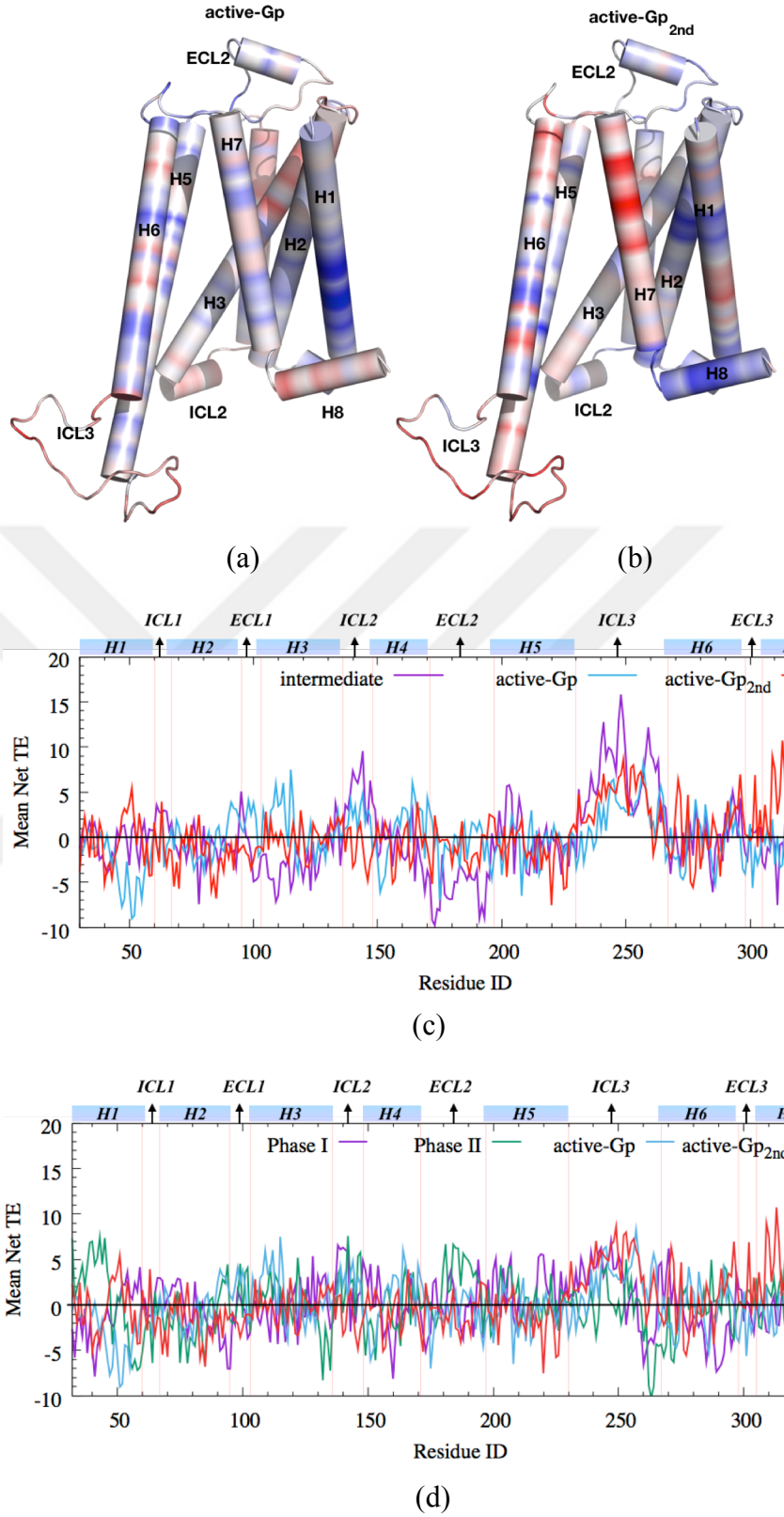
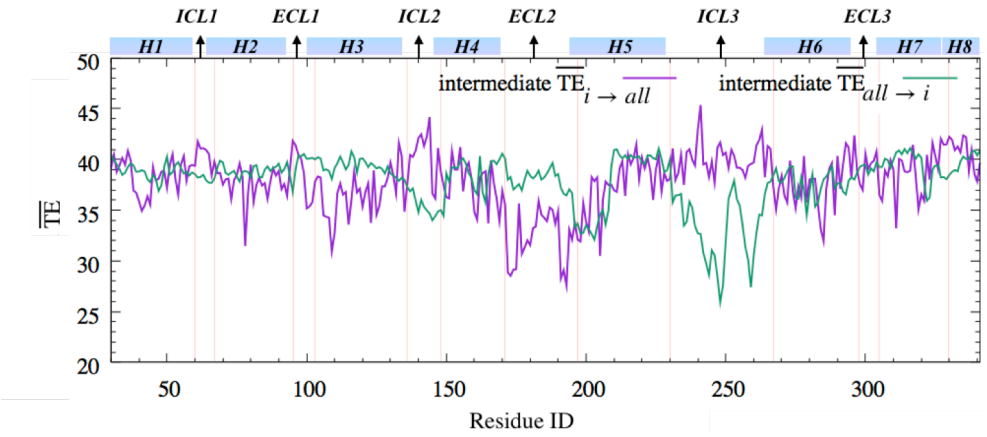


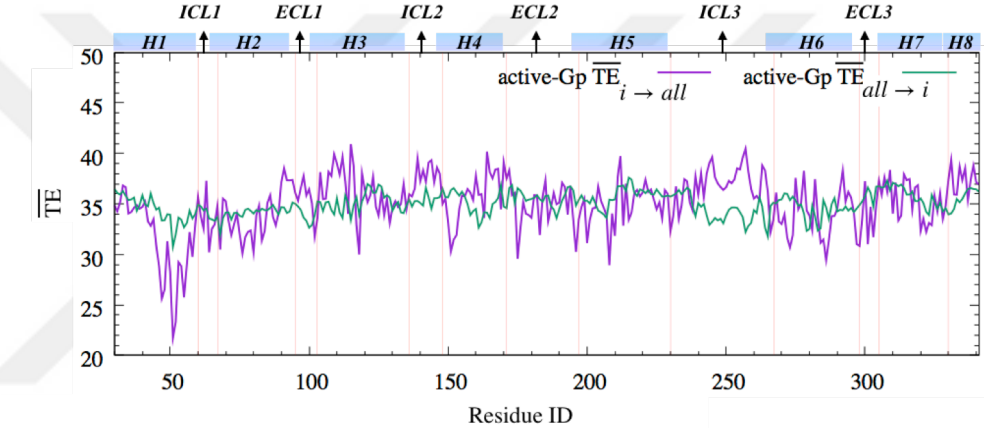
Figure 5. 27 Protein representation of the mean Net transfer entropy values for (a) *active-Gp*, and (b) *active-Gp*$_{2nd.}$ Red color indicates the entropy donors and blue color indicates the entropy acceptors. Mean net TE values for *intermediate* state, *active-Gp,* and *active-Gp*$_{2nd}$ (c). Mean net TE values for inactive and active phases(d).

Strangely, we observed from our previous analysis that the MI value of ICL3 in active states was lower compared to inactive and *intermediate* phases. Additionally, ICL3 was not sharing as high information as H1, H2, and H3 in *active-Gp* and H6 and H7 in *active-Gp*$_{2nd}$. However, here the net TE values were observed to be positively high, and this loop became one of the major entropy donors. To understand the underlying reason of this difference, the average transfer entropy values from residue *i* to all residues and vice versa were plotted in Figure 5.28.

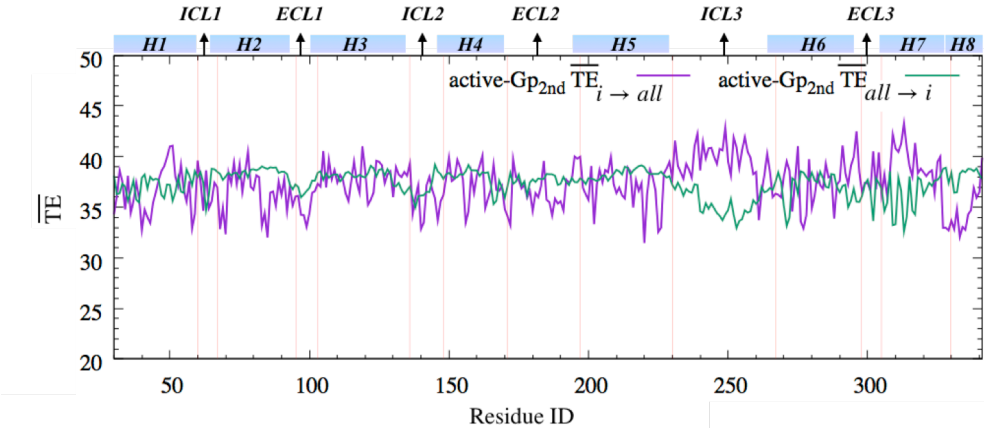
In the intermediate state, the transfer of information from ICL3 to the rest of the protein was higher compared to the transfer of information from the rest of the protein to ICL3. The opposite result was observed for ECL2. These larger gaps between one direction of entropy transfer to another direction made the *intermediate* state of being the state with the highest net TE values (Figure 5.28a). These gaps in active states were not as high as the gaps in the *intermediate* state. In *active-Gp*, H1$_{Int}$ was not donoring high information to the rest of the receptor, but it was accepting higher information. Therefore, the difference made this site as a major entropy acceptor. ICL3, however, donored slightly high information to the receptor compared to other domains. However, it was a donor because the past fluctuations of the rest of the residues did not give enough information about the future fluctuations of ICL3. Meaning that the past and future fluctuations of ICL3 could be used to predict the future fluctuations of the protein, but the past and future fluctuations of the receptor could not be used to predict the future fluctuations of ICL3 in active-Gp (Figure 5.28b). The same was true for *active-Gp*$_{2nd}$, as well. Interestingly, the gap was wider in this phase, indicating the perpendicular motion of ICL3 in *active-Gp*$_{2nd}$ was more important to understand the future conformational changes of the receptor compared to ICL3 motion in *active-Gp* (Figure 5.28c). Overall, ICL3 was not sharing high information instantaneously but transferring information from its previous conformations to the receptor. Thus ICL3 motions could predict the receptor mobility majorly.



(a)



(b)



(c)

Figure 5. 28. Transfer of entropy from each residue to the rest of the protein (magenta) and from the protein to each residue (green) in *intermediate* state (a) *active-Gp* (b) and *active-Gp*$_{2nd}$ (c).

Furthermore, experimentally resolved positive allosteric modulator was reported to interact with the intracellular part of the receptor enclosed by the helices H2, H3, H4, and loop ICL2. This molecule, known as *Cmpd-6FA*, binds to the active receptor and enhances both orthosteric agonist binding and transducer coupling by stabilizing the active-state of the β_2AR (Liu *et al.,* 2019) (Figure 5.29).

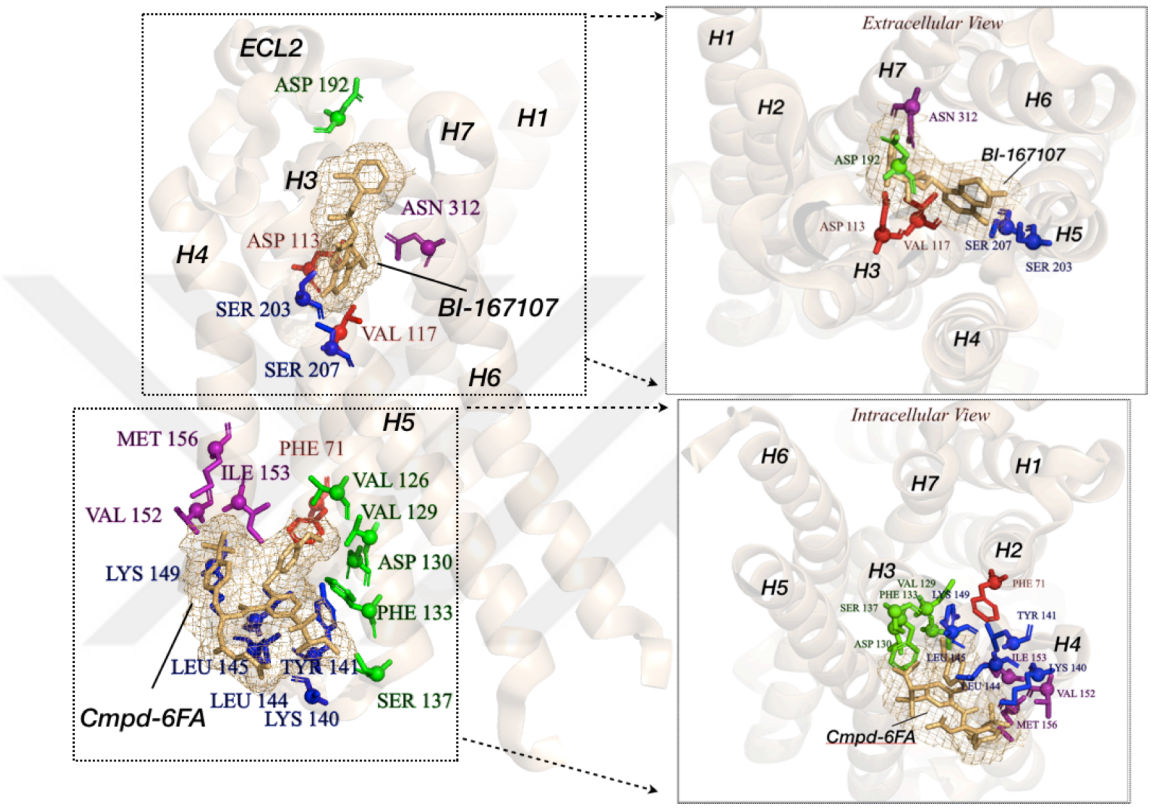
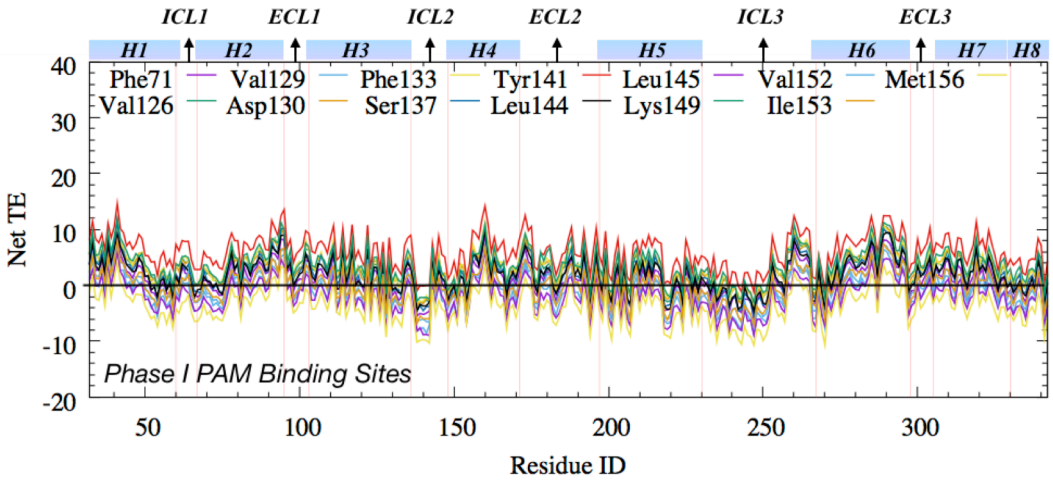
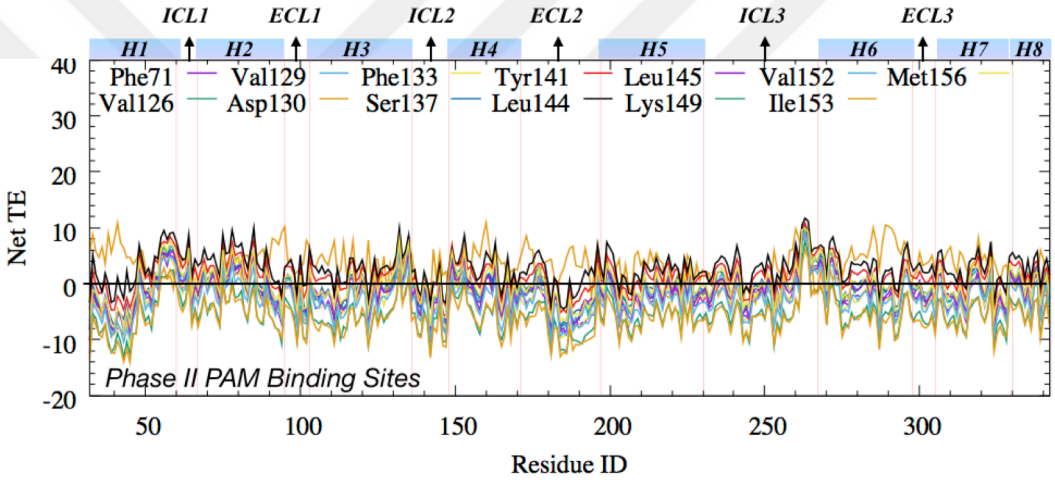


Figure 5. 29 Positive Allosteric Modulator *Cmpd-6FA* bound to the β_2AR (PDB ID: 6N48) together with agonist *BI-167107*. The interacted residues for *BI-167107* were zoomed extracellularly in upright and for *Cmpd-6FA*; they are zoomed intracellularly in the lower right side of the image.

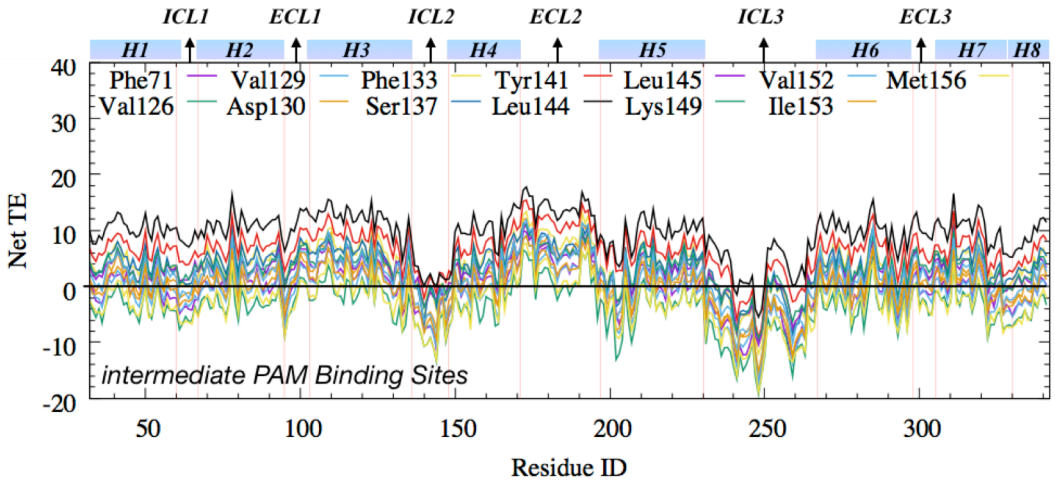
Inactive phases, *intermediate* trajectory, and *active-Gp* runs were compared (Figure 5.30), and a slightly higher entropy donor ability of PAM binding sites was observed in the *intermediate* state. However, in *active-Gp*, PAM binding residues in $H3_{Int}$ and ICL2 were entropy donors, while upon shifting to PAM binding sites on $H4_{Int}$, they became entropy acceptors. Thus residues interacting with PAM displayed donor characteristics when they were on H3 and ICL2 but acceptor characteristics when they were on H4. In *active-*Gp_{2nd}, PAM interacting residues showed a more confined and relatively neutral net transfer entropy values.


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Phe71
Val129
Phe133
Tyr141
Leu145
Val152
Met156
Val126
Asp130
Ser137
Leu144
Lys149
Ile153
Net TE
Phase I PAM Binding Sites
Residue ID

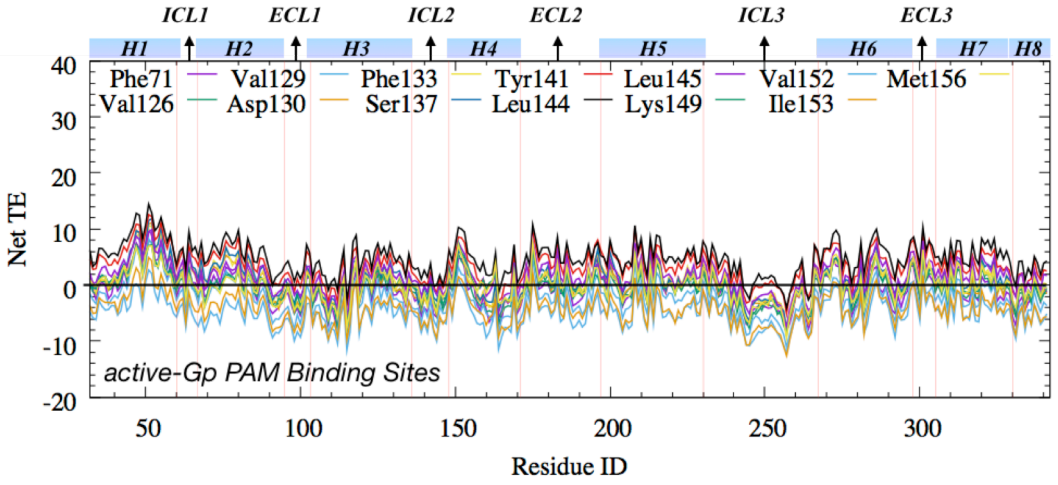
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Phe71
Val129
Phe133
Tyr141
Leu145
Val152
Met156
Val126
Asp130
Ser137
Leu144
Lys149
Ile153
Net TE
Phase II PAM Binding Sites
Residue ID

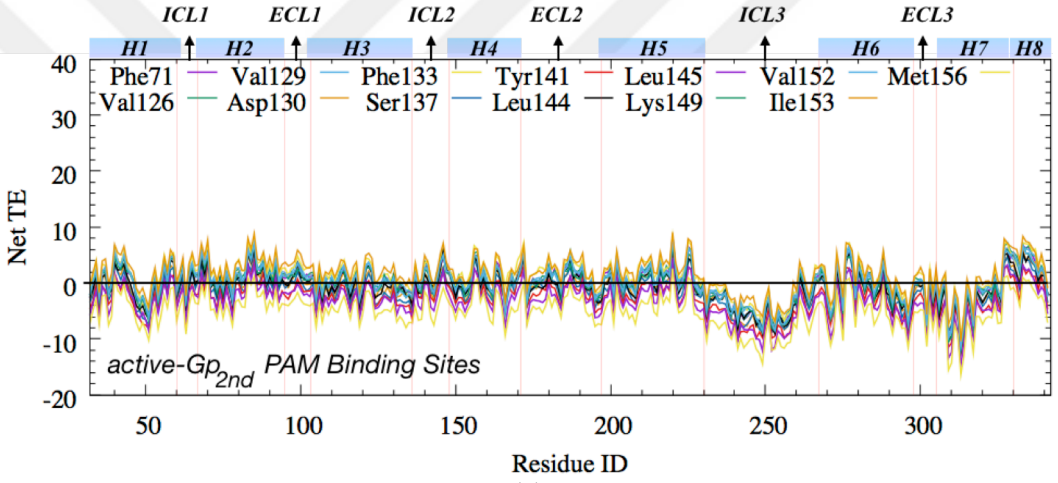
(b)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Phe71
Val129
Phe133
Tyr141
Leu145
Val152
Met156
Val126
Asp130
Ser137
Leu144
Lys149
Ile153
Net TE
intermediate PAM Binding Sites
Residue ID

(c)



(d)



(e)

Figure 5. 30 Net transfer entropies of the residues located in the PAM binding site. Net TE in Phase I (a), Phase II (b), *intermediate* State (c), *active-Gp* (d), and *active-Gp*$_{2nd}$ run (e).

The intracellular association of α5-helix of Gα might be affecting these structures by blocking the PAM binding site and decreasing their entropy transfer capabilities. Thus, the *intermediate* state was able to present this site as entropy donor due to the lack of intracellular binding partners; however, in G protein bound states, this site was occupied; thus, entropy characteristics might be interfered by G protein. In Phase II, on the other hand, PAM binding sites were acting as entropy acceptors, while these residues in Phase I generally displayed donor characteristics except *Met*156. This finding also supports the idea of the effect of G protein on the PAM binding site. Interestingly, *active-Gp*$_{2nd}$ net TE values were neutral in PAM binding sites. This was mainly because the fluctuation of H2$_{Int}$, H3$_{Int}$, H4$_{Int}$, and ICL2 in this state was almost

zero due to the fluctuation of ICL3; thus, no entropy transfer occurred. However, especially the fluctuation of ICL2 was higher in *active-Gp* compared to *active-Gp*$_{2nd}$, which displayed donor characteristics in Figure 5.30d.

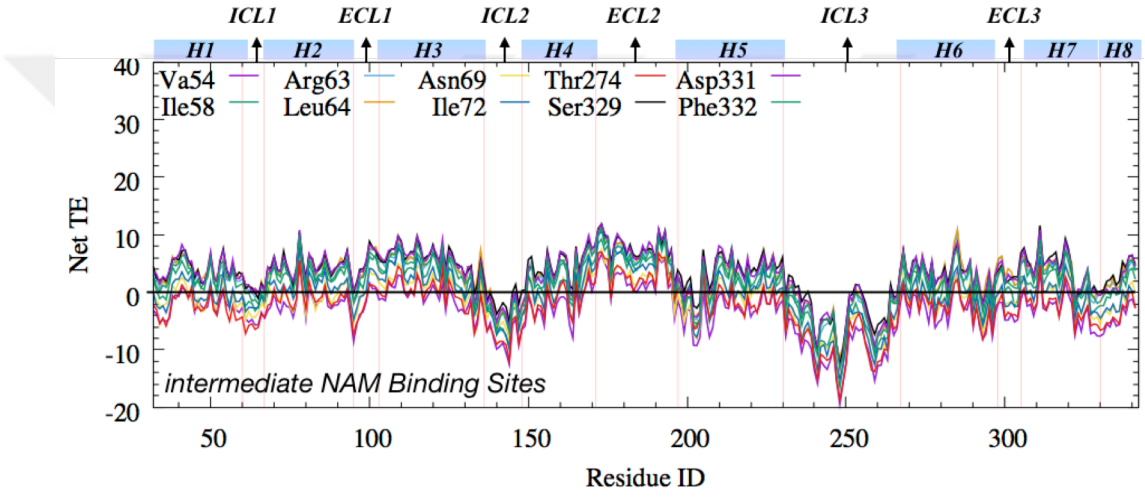
When NAM binding sites were plotted for *intermediate*, *active-Gp*, and *active-Gp*$_{2nd}$ in Figure 5.31, we observed that residues were mostly entropy donors except for *Val*54, *Asn*69, *Ile*72, and *Thr*274. In *active-Gp*, NAM bindings sites were observed to be acceptors except those in H8. Finally, *active-Gp*$_{2nd}$ also displayed these residues as acceptors. Thus, NAM bindings sites in active phases were observed as acceptors, in the *intermediate* state they were mostly donors, in Phase I, they were donors, and in Phase II, they were acceptors (Figure 5.8). Therefore, when the G protein binding site was occupied, intracellular modulator binding sites were acting as entropy acceptors; however, when this site was available, the same residues act as donors.

The orthosteric binding site plays a significant role in receptor activation, thus it highly important as allosteric modulator binding sites. It was observed in Figure 5.7 that donor features of the residues locating in orthosteric ligand binding site were present in Phase II, while these residues appeared as acceptors in Phase I. The comparison of *intermediate* state with active states based on the extracellular binding site revealed an interesting fact that *intermediate* state also displayed donor characteristics in its extracellular parts, especially with the residues *Ser*203 and *Ser*207 (Figure 5.32). The finding that both intracellular and extracellular residues in critical binding sites were donoring entropy further explains this state's intermediate characteristics representing both inactive and active receptor natures.

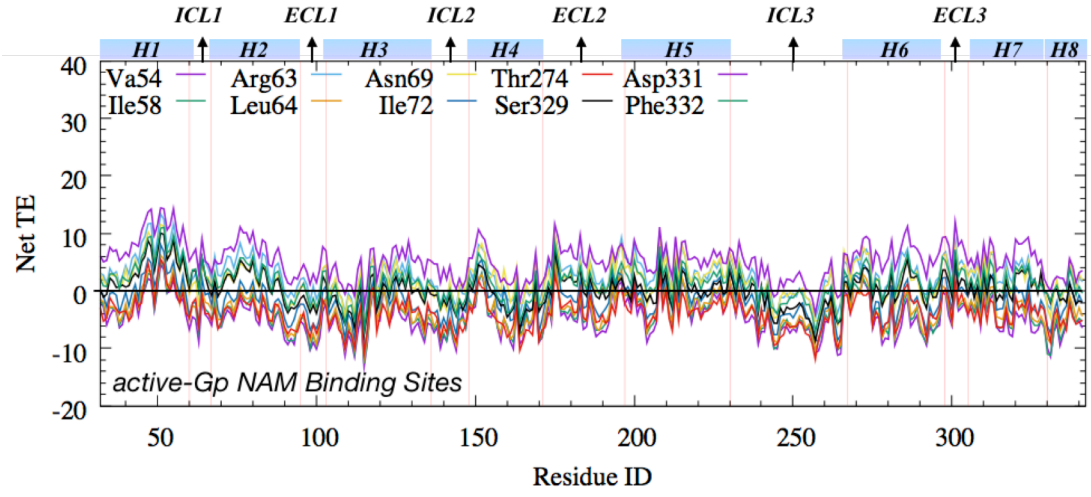
However, except *Asp*113, *Val*114, and *Val*292 in *active-Gp,* and *Asn*312 in *active-Gp*$_{2nd}$, residues in the orthosteric-binding site did not appear as strong entropy donors as they were in the *intermediate* state and Phase II. Moreover, similar to the allosteric binding sites, residues in *active-Gp*$_{2nd}$ were more neutral compared to *active-Gp*. Also, residues displayed more acceptor characteristics in *active-Gp* compared to *active-Gp*$_{2nd}$. These results indicated that the lower $H3_{Ext} - H5_{Ext}$ distance of *active-Gp*$_{2nd}$, displayed in Figure 5.20a, compared to *active-Gp,* might affect the extracellular binding sites

entropy traits. The closest distance between $H3_{Ext} - H5_{Ext}$ due to the ICL3 motion and G protein association in *active-Gp*$_{2nd}$ may result in a more active-like trajectory in *active-Gp*$_{2nd}$; thus, the orthosteric bing site was more stabilized.

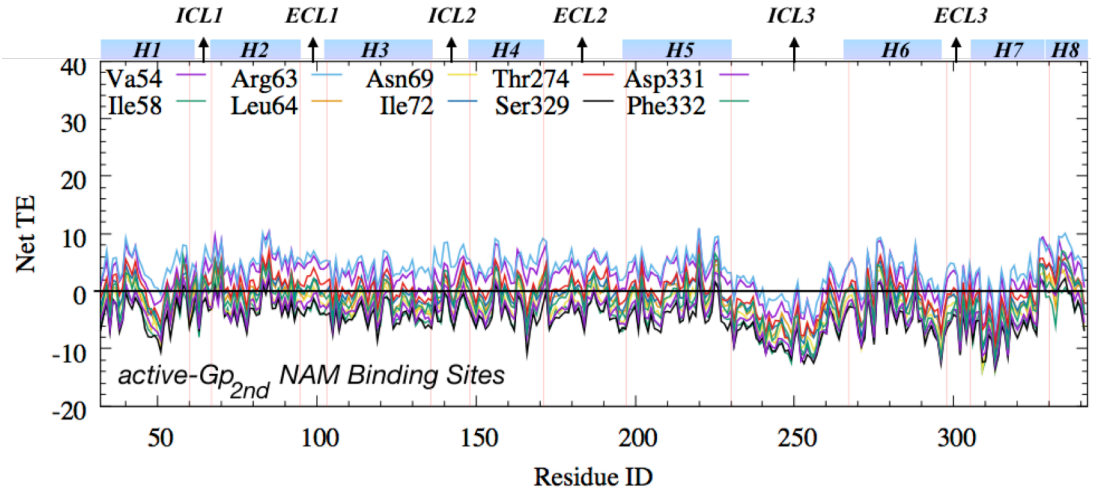
These results supported the allosteric pipeline between inactive, intermediate, and active states found by Bhattacharya group in 2014 (Bhattacharya and Vaidehi, 2014) by capturing the domains where they stated as in the allosteric pathway. In the same study, the active state demonstrated less correlated motions compared to the inactive phase, similar to our observations of fewer percentages of MI values in the *active-Gp* data.



(a)



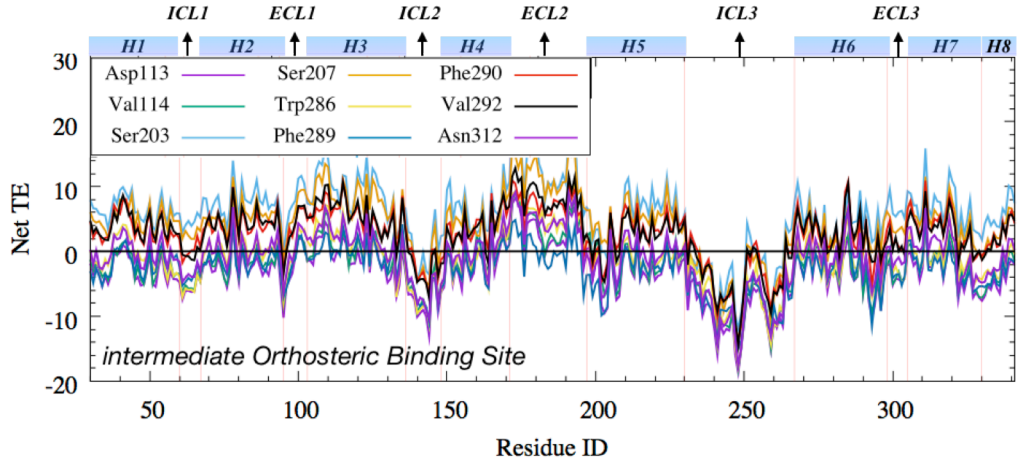
(b)



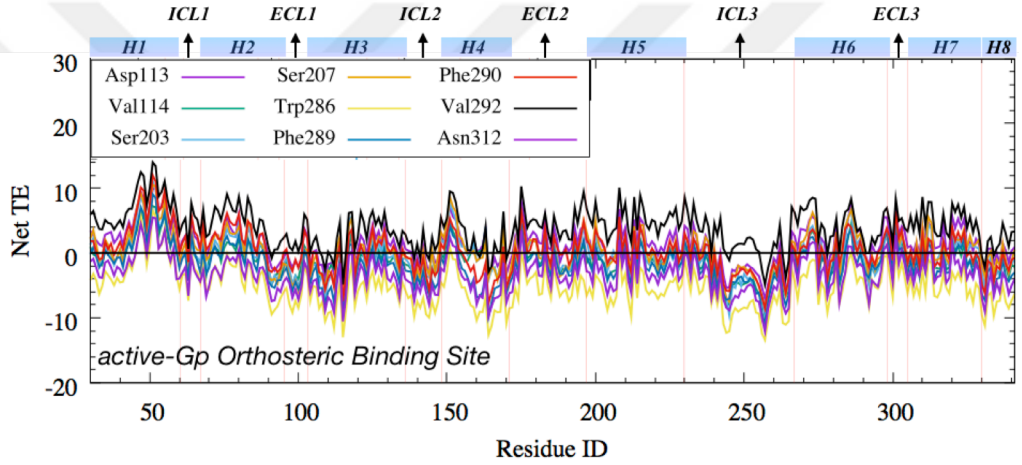
(c)

Figure 5. 31 Net transfer entropy of the residues located in the NAM binding site. Net TE in the *intermediate* state (a), *active-Gp* (b), and *active-Gp*$_{2nd}$ run (c).

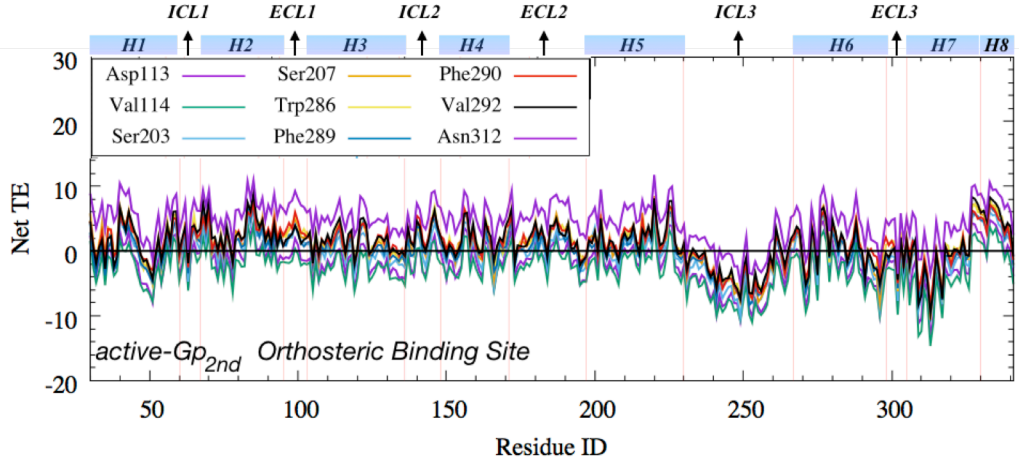
Additionally, it was reported in the same study that H5, H6, and intracellular parts of the receptor involved more in the allosteric pipeline in an inactive state, where it shifted to H7, H8, and the orthosteric binding sites in the active phase. In our case, ICL3 was included to the dynamical calculations, where the shared information between intra- and extracellular parts was discovered in all states. ICL3 maintained its donor characteristics from Phase I to the active states. In Phase II, however, it was packed, therefore, lost its entropy donor abilities. Additionally, the active state displayed fewer percentages of communications and uncorrelated ICL3 motions with the rest of the receptor. In information transfer analysis of the active phases, entropy donor feature of H7, H8, as well as extracellular parts of H3 and H4 was discovered, where these domains were given in the allosteric pipelines in the active phase.



(a)



(b)



(c)

Figure 5. 32 Net TE values of extracellular orthosteric binding sites in intermediate state (a), *active-Gp* (b) and *active-Gp*$_{2nd}$ (c).

6. COMMUNICATION NETWORK UNRAVELED THROUGH BACKBONE AND SIDE-CHAIN DIHEDRAL ANGLE FLUCTUATIONS

Side-chains are significantly less ordered than the backbone in folded proteins (Igumenova, Frederick, and Wand, 2006), and alternative side-chain configurations in protein crystals are prevalent (Lang *et al*., 2010). In addition, the involvement of side-chain fluctuations in long-range networks has been demonstrated through the mutational analysis of NMR studies (Clarkson, Gilmore, Edgell, and Lee, 2006; Millet, Mittermaier, Baker, and Kay, 2003). Nevertheless, MD studies have mostly neglected the contribution of side-chains in most of the structural analysis. Therefore, mutual information and transfer entropy based calculations were performed on the backbone and side-chain dihedrals for the comparison with the same analysis performed on the Cα fluctuations, as explained in Chapter 5. Here in Chapter 6, the dihedral data of the five trajectories of five distinct states of the receptor, namely Phase I, Phase II, *intermediate*, *active-Gp,* and *active-*Gp_{2nd}, was gathered. The analysis of dihedral angles was used to determine its contribution to information sharing and transfer. Differences in the results will help us to identify the effect of conformational changes on the allosteric communication network.

6.1. Results and Discussion

6.1.1 Backbone dihedral frequencies follow Ramachandran distributions

Backbone dihedral angle frequencies ϕ and ψ were calculated for each trajectory by using bin intervals of 0-120, 120-240, and 240-360 degrees. Then the frequencies of dihedral angles were derived and plotted, as shown in Appendix F. It was observed that each structure and trajectory have its own characteristic backbone motion and specific

ϕ and ψ angles. Most of the dihedral angles reside in between 240°-360°, which agrees with Ramachandran distribution for right-handed alpha helices (Ramachandran and Sasisekharan, 1968). It was reported that changes in backbone dihedrals were affected by the side-chain dihedrals, where χ_1 at values around –60 and 60° was observed to modulate ϕ, although its effect was expected to be smaller than its effect on ψ, since an -NH- group has a shorter range influence (Chakrabarti and Pal, 1998). Also, when χ_1 was 60° or 180° the γ position was shown near to the -CO- group in ψ angle, and thus, steric and/or electrostatic interactions were expected to occur between the side-chain atoms at the γ position and the -CO- groups with atoms bonded to it. Moreover, another study analyzed the correlation between $\Delta\phi$, $\Delta\psi$, $\langle\Delta r^2\rangle$, relative solvent accessibility (RSA), and B-factor, both in NMR and X-ray structures. The correlation of $\Delta\phi$ and $\Delta\psi$ with $\langle\Delta r^2\rangle$ was 0.5 in NMR structures, while the correlation of $\Delta\phi$ with B-factor in X-ray structures was 0.29 and $\Delta\psi$ with B-factor was 0.27, where B-factor is known to be related with the Cα fluctuations. The correlation between $\Delta\phi$ and $\Delta\psi$ was observed as 0.53 (Zhang *et al*., 2010). Thus, showing uncoordinated mobilities between Cα fluctuations and backbone dihedrals, which was reflected to the mutual information and transfer entropy data, will be explained in upcoming sections.

6.1.2 Side-chain dihedral frequencies depend on amino acid characteristics

Side-chain dihedral angles differ from amino acid to amino acid based on the number of rotatable bonds and the angular tendency of the particular rotatable bond due to the steric hindrance. The number of rotatable bonds per amino acid was given in Appendix E. Amino acids with the same number of rotatable bonds might have different angular positions. Fundamentally, side-chain dihedrals are named of being gauche (+), trans, and gauche (-) based on Newman projection. However, dihedral angles of some amino acids, such as *Glu*, are further classified based on Oxygen or Nitrogen atoms being cis or trans orientation according to their positioning to the corresponding Carbon atom. Thus, the occupied states of the dihedral angle differ based on the spatial atomic positions, which was given in Appendix E. Number of dihedral angles differ in amino acids. Residues with no side-chain dihedral are *Ala* and *Gly*, with one side-chain dihedral are *Pro, Cys, Val, Ser,* and *Thr*, with two side-chain dihedrals are *Ile, Leu, Phe,*

Trp, Tyr, Asn, Asp, and *His,* with three side-chain dihedrals are *Met, Gln,* and *Glu,* and finally with four side-chain dihedrals are *Arg* and *Lys*.

Some studies addressing side-chain dihedrals use three bins to simplify the calculations. However, as noticed in Figure given in Appendix G, frequency distributions of dihedral angles for some residues such as *Asn*, *Glu,* and *Gln* do not obey the three-binning approach and they require more number of bins for their χ angle-based probability calculations. Moreover, even though the selected number of bins for these types of residues is four or six, the initiation of the angle distribution should start at 45° or 30°, respectively. We followed a study of the side-chain behavior of proteins using the Dynameomics protein simulation dataset. (van der Kamp *et al*., 2010) They have analyzed the native state molecular dynamics simulations on 807 proteins spanning essentially all known protein folds, each of which was simulated at 298 K for at least 31 ns (Beck *et al*., 2008). The given angular intervals were further verified by calculating the angular information of backbone and side-chain dihedrals in our trajectories as four states' frequencies were given in Appendix G.

6.1.3 Mutual information analysis

6.1.3.1 Mutual information data based on backbone dihedrals

Backbone dihedrals based on ϕ and ψ angular motions were used in calculating mutual information. Here, the calculation depends on the angular variation between 0° and 360° for each dihedral angle with the bin size of 120° (i.e., bin number=3). Thus, unlike the mutual information of Cα fluctuations, where the absolute deviation from the average atomic positions was calculated, the angular information was used, as it was to calculate the probability of the corresponding angle. Moreover, since there were three bins for each ϕ and ψ angle, 9 different states were available for each residue *i* and 81 different states were available for each residue-pair *ij*. Furthermore, for *N* number of residues in a protein, the total number of pairs to be analyzed was *2N-2* due to the fact that the first amino acid in a protein lacks the ϕ angle, and the last residue lacks ψ angle.

As given in Figure 6.1, backbone mutual information for five distinct receptor states represented that pairwise communications were mostly between loops. As plotted in Figure 6.1a, strong communications in ECL1, ICL2, ECL2, and ICL3 was observed in Phase I. Moreover, the N- and C-terminal of the receptor also contributed to the information sharing. In the previous chapter, information sharing in similar non-contacting domains from the analysis of Cα fluctuations was observed, as well. Thus, these domains demonstrated both positional and angular communications.

In Phase II, however, ICL3 mostly lost its information-sharing characteristic except for the residues in between *Gly*252 to *Ser*261. Instead, ICL1 and ECL2 started to share information with themselves, with ECL2, and with residues *Gly*252 to *Ser*261 on ICL3. This shift of MI values from ICL3 to extracellular residues was also observed in Cα fluctuations, however high MI value of $H1_{Ext}$, observed in Cα analysis, was not detected in mutual information analysis of the backbone dihedrals. This is mostly because the helical structure of H1 prevented the dihedral angle fluctuations, thus decrease its entropic contribution to the calculations.

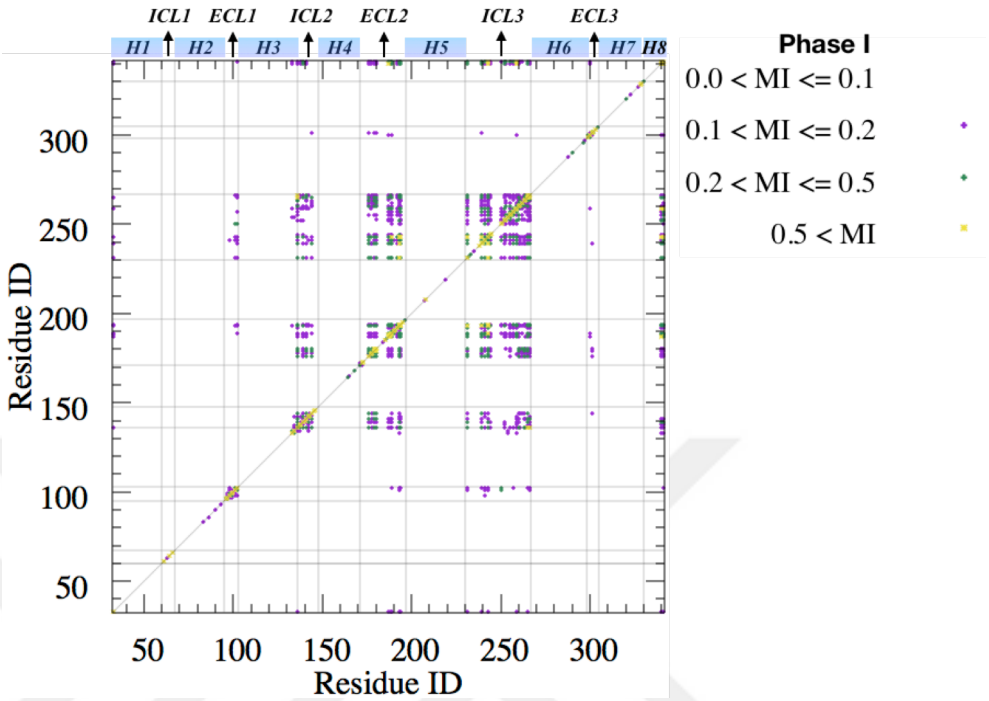
Furthermore, the *intermediate* state represented high MI sharing in ICL3 with itself and between ICL3 and other loops. However, the amount of information sharing was lower in the *intermediate* state compared to Phase I and Phase II. Thus, due to the selected cutoff in Figure 6.1c, all pairwise interactions could not be represented, but they were seen in mean MI values in Figure 6.1d. The mean backbone MI values for these three states showed the differences in communicating domains. Phase I had the highest backbone MI data, where it was mostly populated around ECL2, ICL3, and C-terminal domain. In Phase II, ECL1, ICL2, and ECL2 were the domains involved in communication, while in the *intermediate* state, ICL3 started to play a role in information sharing. Interestingly, *Gly*35 in H1, *Asn*196 in ECL2 adjacent to H5, and *Phe*321 in H7 had high mean MI values. The MI calculation in Cα fluctuations represented H5 with high MI value in the *intermediate* state. Interestingly, *Asn*196 in ECL2 adjacent to H5 represented high MI, which might be affected by Cα fluctuations.

The active states' MI values from backbone dihedrals were also plotted in Figure 6.1d and 6.1e. ICL3 had the biggest role in communication in active states. Interestingly, a few distinct residues in transmembrane helices were strongly sharing information in the *active-Gp* state, but their effect was missing in the *active-Gp*$_{2nd}$. Among them, *Ala*119 in H3 and ICL2 involved in information sharing in *active-Gp* but they were missing in *active-Gp*$_{2nd}$. Also, *Trp*99 in ECL1 *Ser*165 in H4, *Leu*272 in H6, and *Pro*323 – *Ser*329 in H7 were involved in *active-Gp*$_{2nd}$, but they were not present in *active-Gp (*Figure 6.1f). These residues were located critical residues on orthosteric and allosteric sites and playing a role in ligand binding. Since backbone dihedrals' higher MI values were observed at most in a few residues on the corresponding domain but not cover all domains, giving the protein domain instead of residue with high MI value could be wrong. For example, Only *Trp*99 had high MI value in ECL1; thus, specifically, the residue IDs were given.

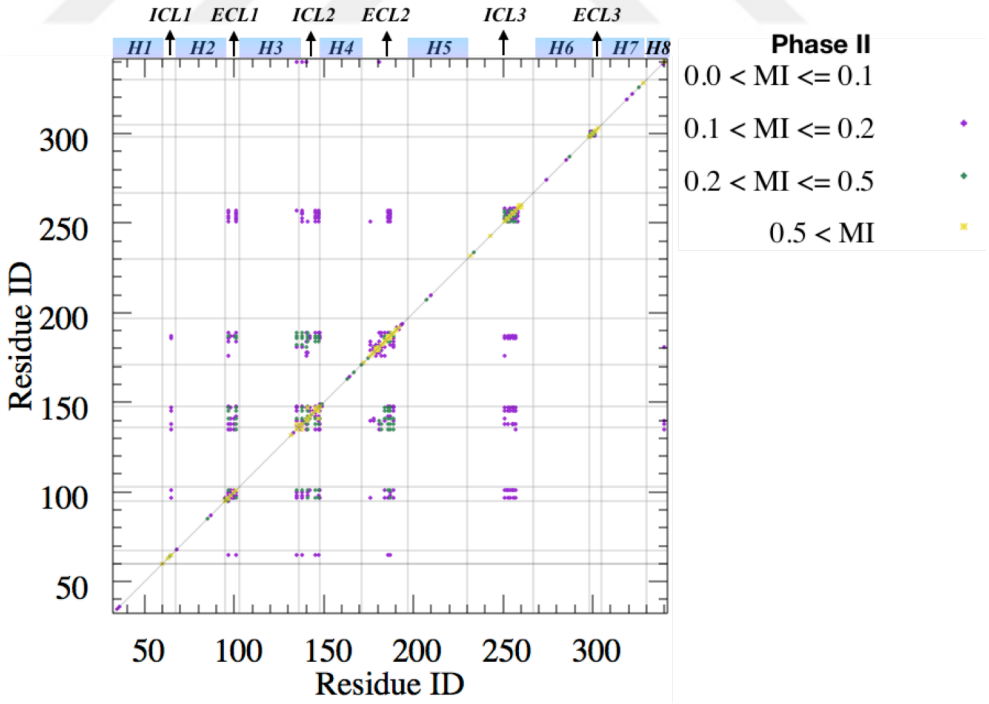
Comparison of the backbone MI data with Cα fluctuations revealed that the Cα fluctuations in H1, H2, and H3 did not affect backbone dihedral angles' transitions in *active-Gp.* Moreover, we observed the highest MI value was around $H7_{Ext}$ in Cα fluctuations of *active-Gp*$_{2nd}$, and it was also high in MI data of backbone dihedrals. Overall, backbone dihedrals' contribution to information sharing explains the communications were mostly between loops. Due to the rigidity of the transmembrane helices as a result of the non-bonded interactions, the high MI values observed in Cα fluctuations were not obtained from backbone dihedral angles' MI analysis. This stability in the helical residues decreases the angular entropy; thus, mutual information became significantly small or zero in transmembrane regions.

Furthermore, every residue had Cα fluctuations from average Cα positions. Some of these fluctuations were small, and some other fluctuations were large. Also, the number of bins was optimized based on these distributions. Thus entropy was depending on the distributions of Cα fluctuations in optimized bins. However, entropy in backbone dihedrals depends on angle distributions in three bins. Also, the absolute fluctuation values in Cα atoms in a residue might have different angular alterations. Therefore, MI values in Cα atoms and backbone dihedrals showed varieties in explaining information

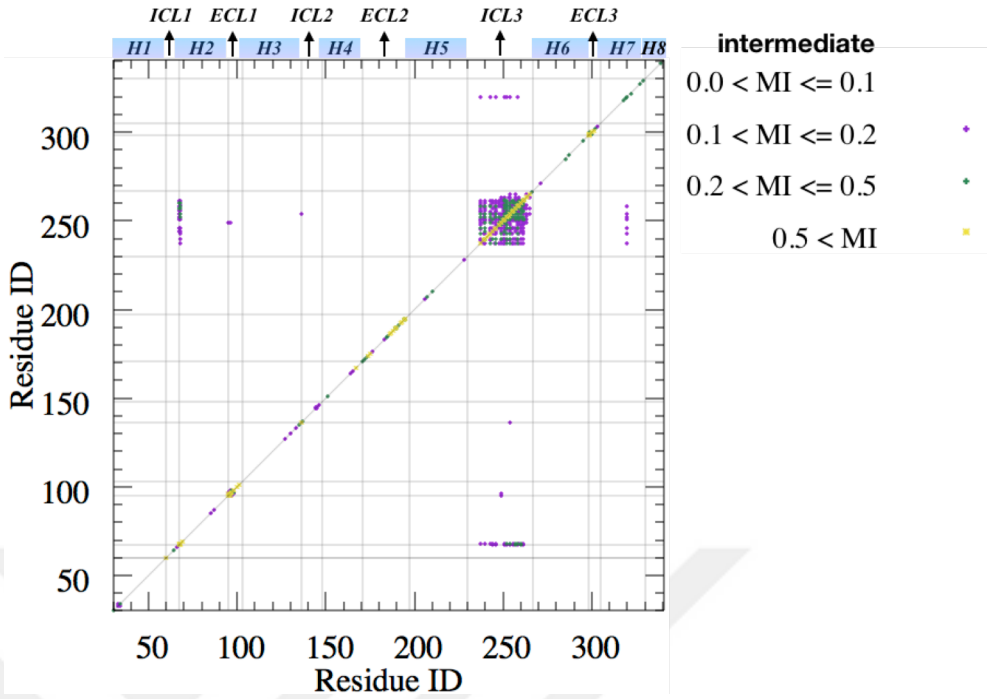
sharing. The discovery of the dihedral angle contribution to Cα fluctuation will give us the details of overall receptor dynamics.



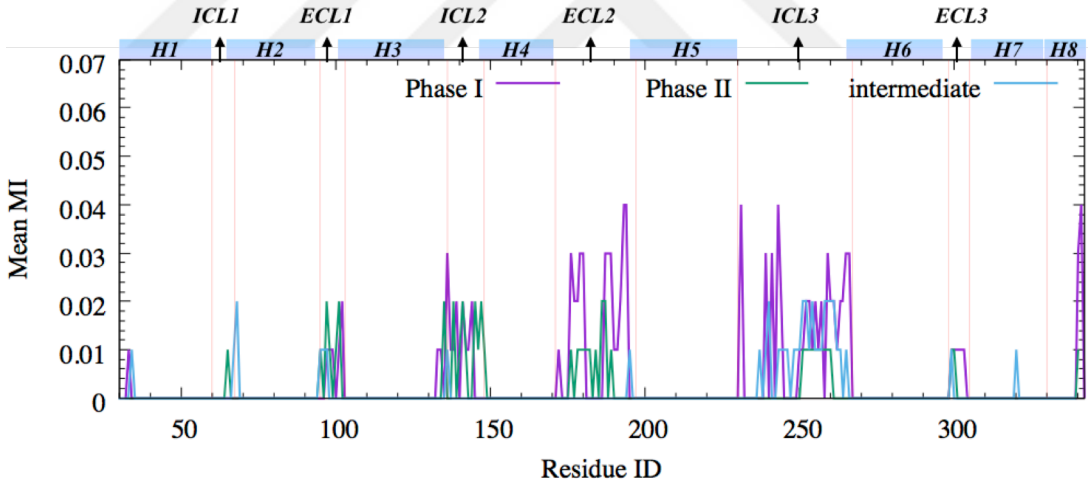
(a)



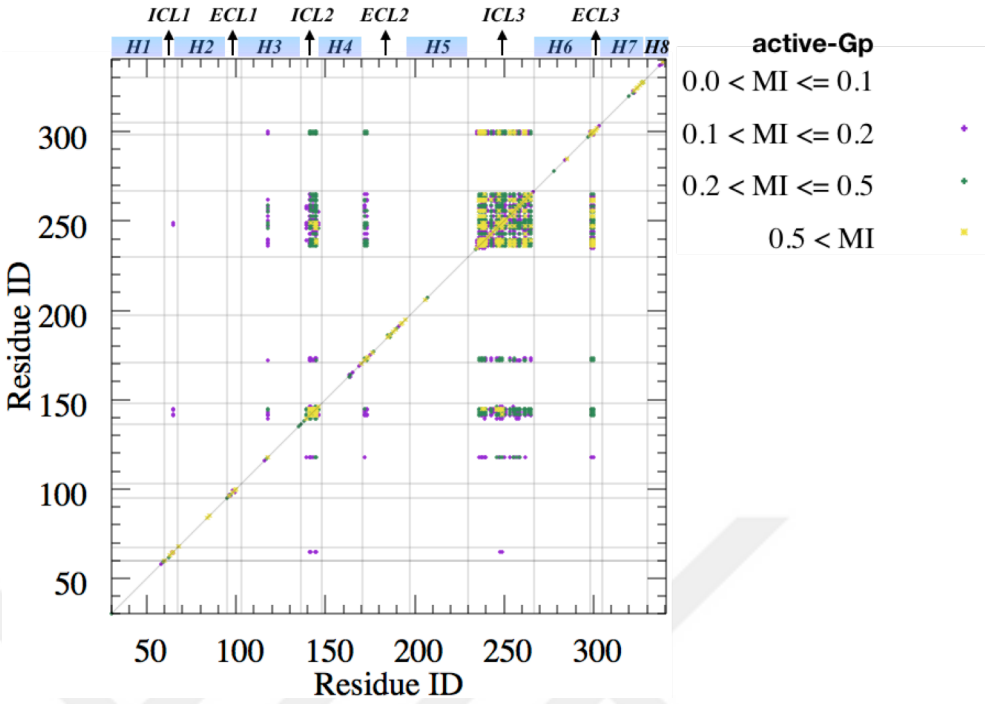
(b)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
intermediate
0.0 < MI <= 0.1
0.1 < MI <= 0.2
0.2 < MI <= 0.5
0.5 < MI
Residue ID
Residue ID
50
100
150
200
250
300

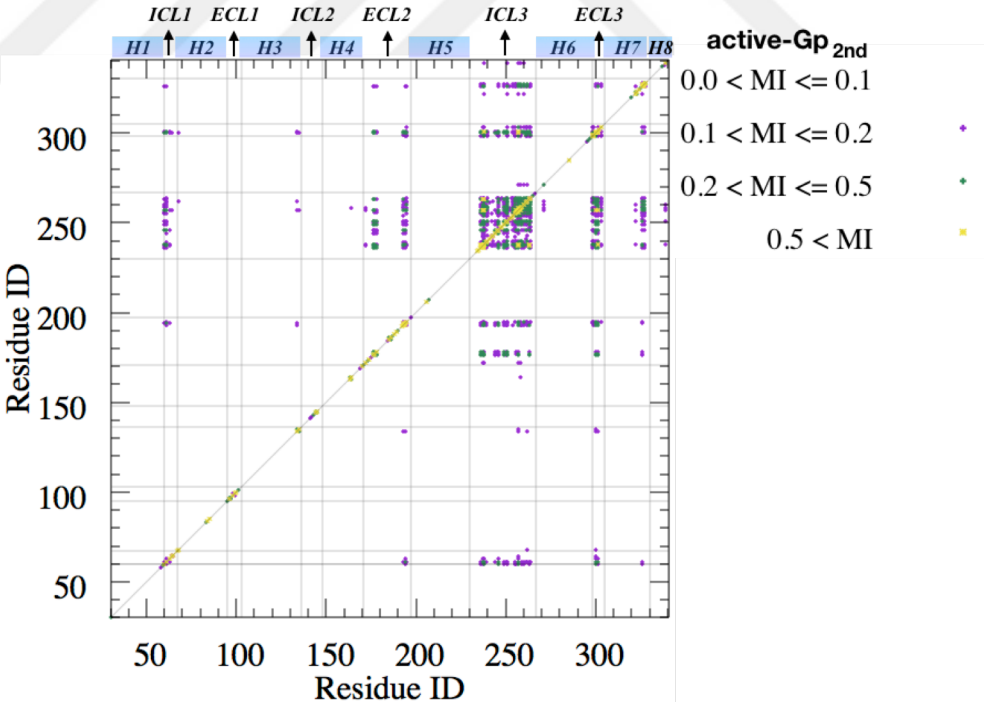
(c)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Phase I
Phase II
intermediate
Mean MI
Residue ID
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
50
100
150
200
250
300

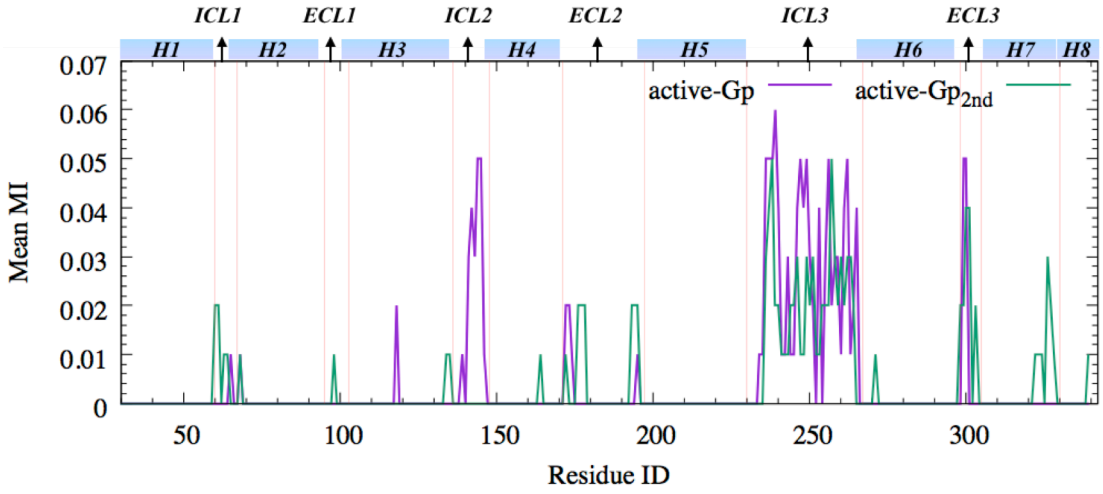
(d)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
active-Gp
0.0 < MI <= 0.1
0.1 < MI <= 0.2
0.2 < MI <= 0.5
0.5 < MI
Residue ID
50
100
150
200
250
300

(e)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
active-Gp 2nd
0.0 < MI <= 0.1
0.1 < MI <= 0.2
0.2 < MI <= 0.5
0.5 < MI
Residue ID
50
100
150
200
250
300

(f)



(g)

Figure 6. 1 Mutual information from backbone dihedral angle analysis for Phase I (a), Phase II (b), *intermediate* state (c), *active-Gp* (e), and *active-Gp*$_{2nd}$ (f). Mean MI values of the states Phase I, Phase II, and *intermediate* state (d). Phase I was colored in magenta, Phase II in green, and intermediate state in blue. Mean MI values for *active-Gp* and *active-Gp*$_{2nd}$. *active-Gp* was colored in magenta and *active-Gp*$_{2nd}$ was colored in green.

6.1.3.2 The effect of side-chain dihedrals to mutual information

Residue-pair MI values of the side-chain dihedral angles also gave substantial information about the conformational change of the receptor and the residues involving in these transitions. The increase in both the MI values and the number of residues involved in information sharing were observed in Figure 6.2a and 6.2b for Phase I. Higher information sharing occurred between mobile residue-pairs compared to stable ones, specifically those located on the loop regions (Figure 6.2b). The highest MI values were observed in ICL3 and ECL2 in Phase I. MI values in transmembrane helices were relatively low, but still, H3, H5, and H6 contain residues, which were sharing medium amount of information. Additionally, residue *Arg*328 connecting the intracellular part of H7 with H8 also shared a high amount of information with the rest of the protein. Figure 6.2c indicated that in the middle and buried parts of the transmembranes, the information sharing due to dihedral angle alterations was significantly low with MI<0.1.

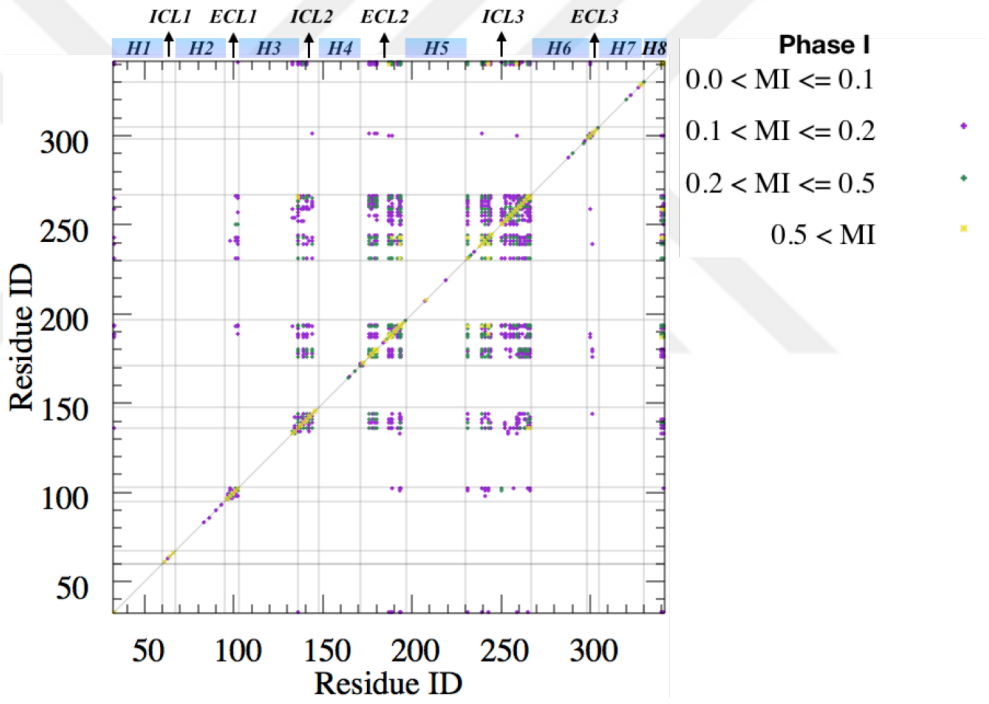
However, higher MI values were observed in the more exposed protein domains as loops or residues adjacent to loops. This data mostly coincides with the previous Cα fluctuation results in terms of loop data, where loops were mostly mobile, but the transmembrane domains were stiffer. One important aspect of these findings came from the alignment of the trajectories. The alignment of the trajectory was generally performed through selecting the most rigid protein parts and the backbone of these parts in order to eliminate the translational and rotational motions of the overall protein. Moreover, none of the alignments carried out by selecting the side-chains of the protein. Therefore, compared to the backbone, side-chains were relatively free throughout the trajectory. Thus, even though side-chains motions are believed to be irregular, the results given in here explained that the overall protein dynamics are actually coherent.

Mean MI values from the analysis of the fluctuation of Cα atoms in Phase I was superimposed with the MI values from the backbone and backbone+side-chain dihedral changes in Figure 6.2c. In the analysis of Cα fluctuations, atomic information per residue was counted. However, in backbone + side-chain dihedral analysis, one residue consists of the φ, ψ, and a number of χ angular information. The quantity of mean MI from backbone+side-chain dihedrals of flexible domains was generally higher compared to the mean MI from Cα fluctuations. This was due to the number of atoms and the number of possible visited states involved in the backbone + side-chain dihedral calculation.

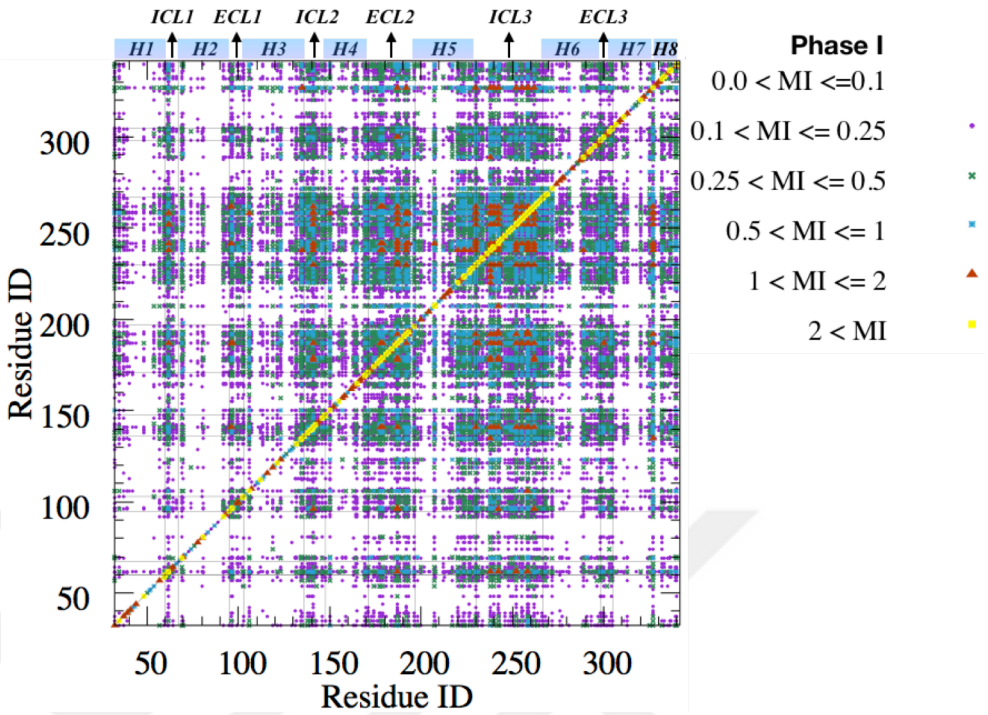
Moreover, the mean MI in Cα fluctuations was higher compared to the mean MI in backbone dihedrals. The optimized number of bins for each Cα atom was higher than the number of bins in backbone dihedrals. Thus, the entropy of each Cα was higher compared to the entropy of the corresponding backbone dihedral. In the end, both the number of bins and the number of atoms strengthen the communication between the residues. Moreover, it was observed that the MI in Cα and dihedral angles mostly demonstrated the same receptor domains, such as ICL3, ECL2, and ICL2, as the domains, which shared information. This is a significant observation, explaining these residues had both positional and angular communications in Phase I. This was not

always the case in different conformations, where positional information was lost in dihedral data, which will be explained in the upcoming sections.

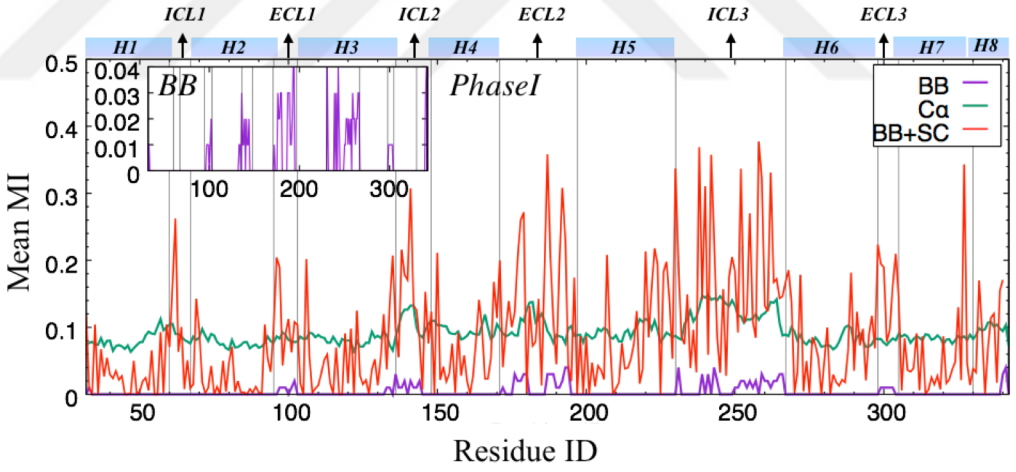
Protein representation of the residues with their backbone+side-chain MI intensities was given in Figure 6.2d. From white to red, MI was colored with increasing values. Residues with high MI value were observed to locate at the loops or the residues in the helices that are adjacent to the loops. These long-range communications between intra- and extracellular parts of the receptor was also shown in Cα fluctuations. Therefore, these sites were not only fluctuating from their mean positions more but also changed their dihedral angles more compared to the buried parts of the receptor.



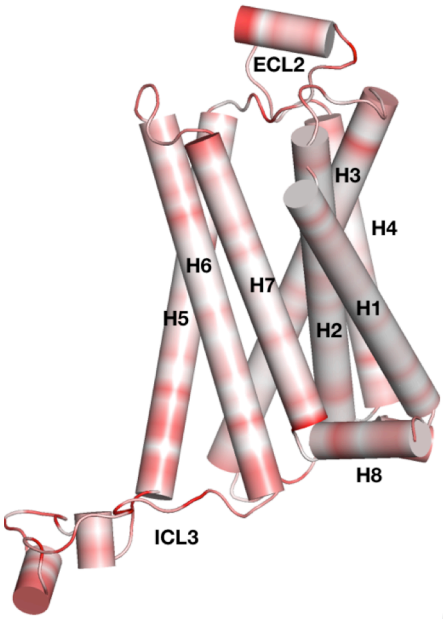
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Phase I
0.0 < MI <=0.1
0.1 < MI <= 0.25
0.25 < MI <= 0.5
0.5 < MI <= 1
1 < MI <= 2
2 < MI
Residue ID
Residue ID

(b)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
BB
PhaseI
BB
Cα
BB+SC
Mean MI
Residue ID

(c)

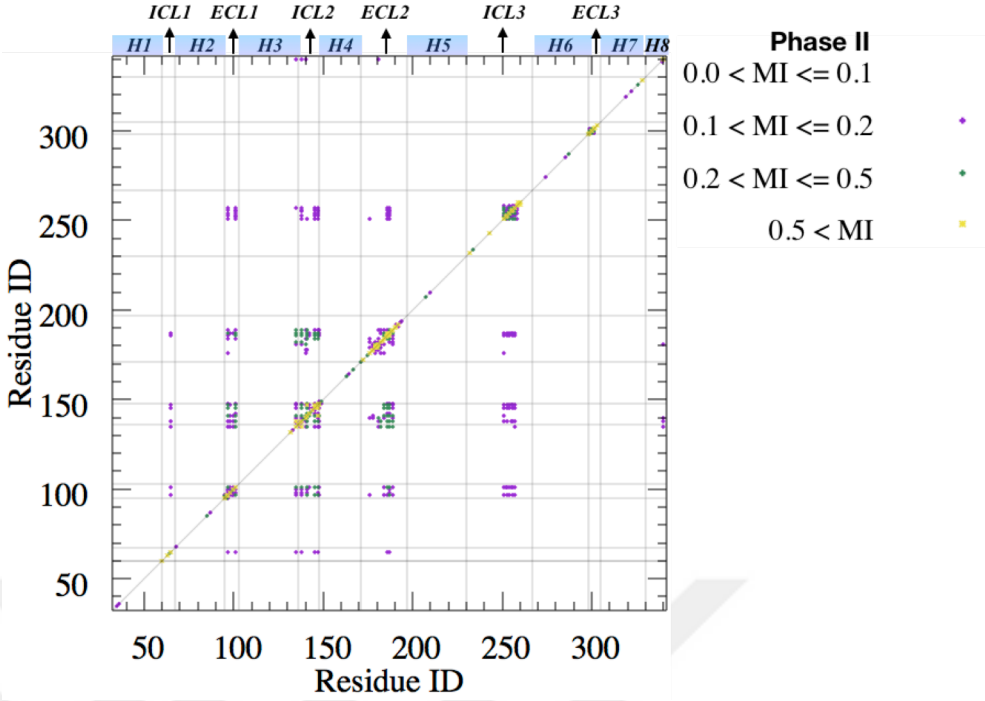


(d)

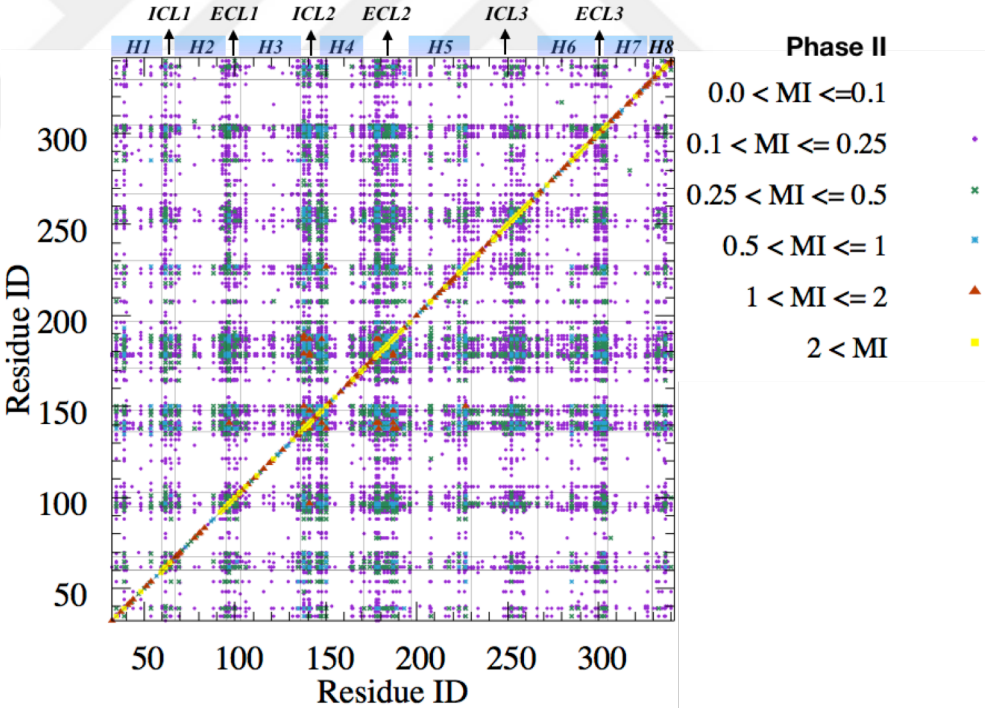
Figure 6. 2 MI plot of backbone (a) and backbone + side-chain dihedrals for Phase I (b). Colors white to yellow indicate an increase in MI values. Mean MI values for backbone dihedrals (magenta), C α fluctuation (green) and backbone + side-chain dihedrals (red)(c) Protein representation based on the mean MI values of backbone + side-chain dihedrals (d). From low to high, MI values were shown from a shade of red color.

In Phase II, except for the residues between *Gly*252 to *Ser*261, ICL3 lost its flexibility, thus information-sharing characteristics. Upon ICL3 closure, ICL1 and ECL2 started to communicate with themselves, with ECL2, and with the 10 residues on ICL3 adjacent to H6. In Cα fluctuations, however, high MI value of $H1_{Ext}$ was observed, but it was lost in backbone dihedrals, due to the constrain of the helical structure of H1 on dihedral angle fluctuations.

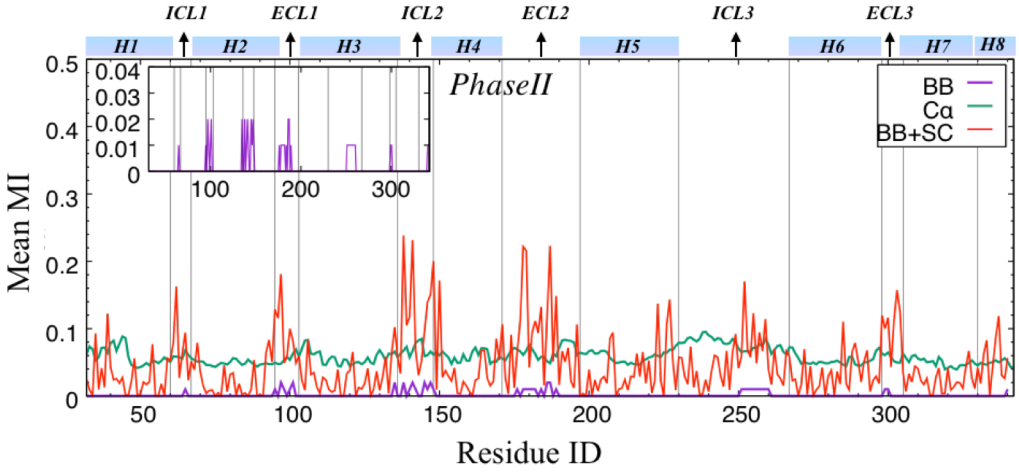
The MI value obtained from backbone and backbone+side-chain revealed a significant decrease both location-wise and value-based compared to Phase I. This was due to the decrease in general flexibility in the receptor upon ICL3 closure. Even though ECL1, ICL2, and ECL2 represented higher information sharing compared to other regions in Phase II, the number of pairwise communications was low in MI above 0.1 (Figures 6.3a and Figure 6.3b).


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Phase II
0.0 < MI <= 0.1
0.1 < MI <= 0.2
0.2 < MI <= 0.5
0.5 < MI
Residue ID
50
100
150
200
250
300
Residue ID

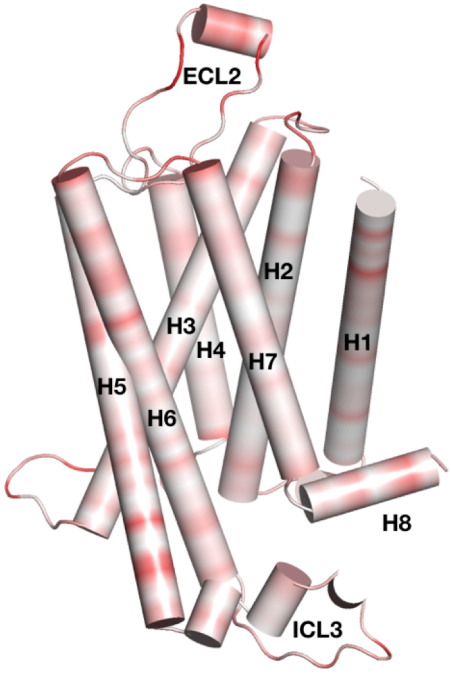
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Phase II
0.0 < MI <=0.1
0.1 < MI <= 0.25
0.25 < MI <= 0.5
0.5 < MI <= 1
1 < MI <= 2
2 < MI
Residue ID
50
100
150
200
250
300
Residue ID

(b)



(c)



(d)

Figure 6. 3 Heatmap of MI based on backbone (a) and backbone + side-chain dihedral angle fluctuations (b) for Phase II. (b) Superposition of average MI values of Cα fluctuations (green) with MI of the backbone (magenta) and backbone + side-chain dihedral angles (red) (c). Protein representation based on the average MI of backbone + side-chain dihedral fluctuations (d).

Dihedral data displayed relatively low mean MI values in residues of ICL3 adjacent to $H5_{Int}$ compared to the MI values of remaining loops. However, this region was sharing information with the remaining residues in Cα fluctuation data. Upon ICL3 closure, the previously unstructured loop started to form short helices in the regions adjacent to H5 and H6, thus along with stabilization of the receptor, ICL3 also stabilized its own

fluctuations. Therefore, this short-segmented helical formation, apparently, did affect the fluctuation of Cα atoms by decreasing its mean MI value from 0.13 to 0.09 (Figure 6.2c and 6.3c), as well as restricted the motion of its dihedral angles. High MI value was observed in ECL2, ICL2, and ECL1. However, these values were only half of the MI values of the same receptor parts in Phase I. Here, we observed differences in MI values of Cα fluctuations and backbone+side-chain dihedral data. Even though domain-wise communication patterns were similar in both positional and angular data, as $H1_{Ext}$, ICL2, ECL2, or ICL3, the residue-based communications have differed. The high MI values of the ICL2 residue in Cα fluctuations were not among the residues with the highest MI values in backbone+side-chain MI calculation. This was explained by the differences between atomic fluctuations and angular motions, where the correlation of these two was not significantly high.

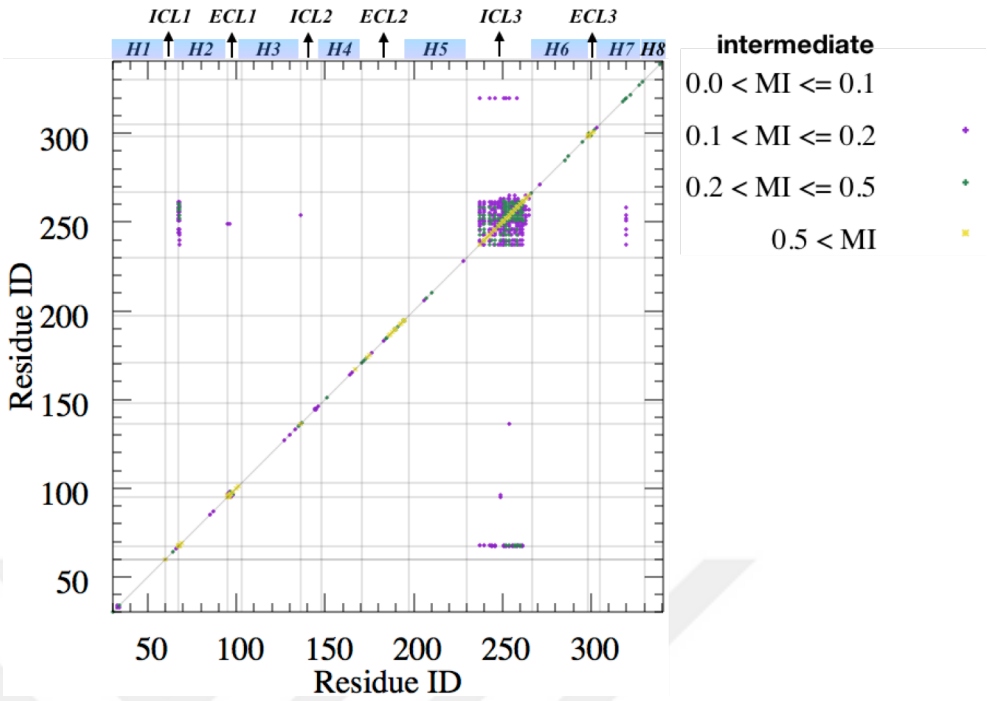
The RMSD and RMSF analysis of the *intermediate* state illustrated the fluctuation and inward tilt of ICL3 and $H6_{Int}$ throughout the trajectory. This state represented similar features with the original inactive state based on Cα fluctuations, but the continuous fluctuations of ICL3 and the partial inward deviation of $H6_{Int}$ diverged this phase from Phase I, and the conformations were counted as receptor's *intermediate* conformations. MI calculation of the dihedral angles in the *intermediate* state revealed high MI values in ICL3, similar to the data observed in Cα fluctuations (Figure 5.14a). MI values of backbone dihedrals represented communication of ICL3 with *Thr*68 and *Asn*69 in ICL1, *Met*96 and *Lys*97 in ECL1, *Ser*137 in ICL2, ICL3-itself and *Phe*321 in $H7_{Int}$ with backbone dihedral MI value above 0.1 (Figure 6.4c). This limited communication between residue-pair backbone dihedrals indicated that, except ICL3, the lower RMSF of loops was also displayed as stable dihedrals with low information sharing.

Interestingly, information-sharing abilities of ICL2 and $H5_{Ext}$, which were observed in Cα fluctuations, were diminished in dihedral angles (Figures 6.4a and 6.4b). Even though RMSF data showed slightly high fluctuations both in ICL2 and $H5_{Ext}$ compared to the fluctuations in transmembranes (Figure 5.10b), these positional fluctuations apparently were not sufficient enough to increase communications in dihedral angles, since the correlation between Cα fluctuations and backbone dihedrals were at most

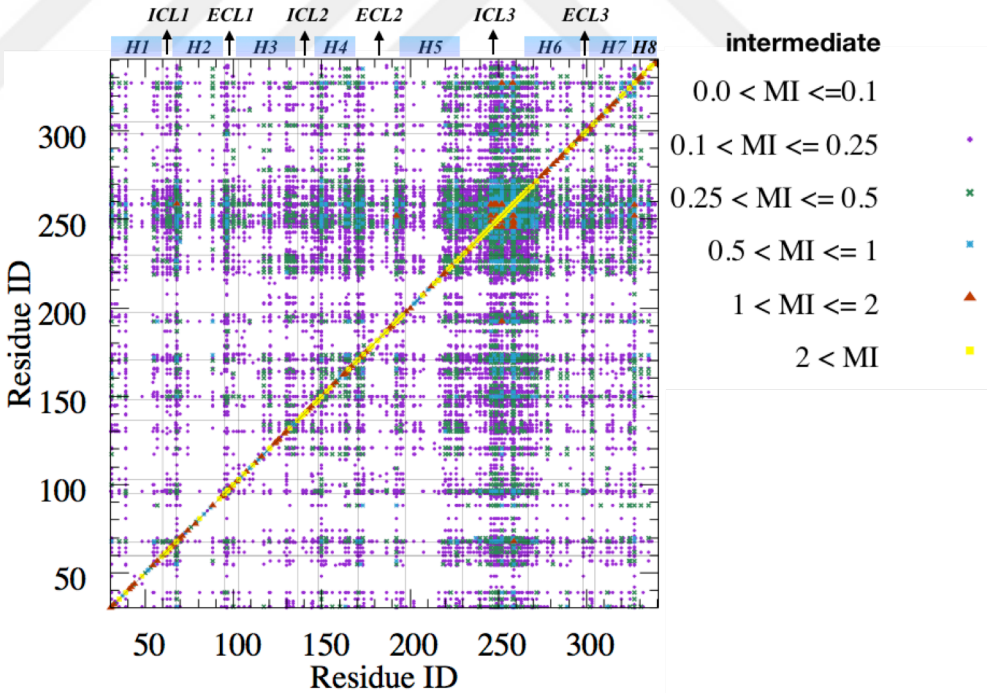
0.29. MI analysis in Chapter 5 was depending on Cα fluctuations; thus, it was observed to be associated with RMSF data. However, here dihedral angles were highly associated with side-chain dihedrals, which was also associated with the secondary structure of a protein part.

Additionally, the inclusion of side-chains to backbone dihedrals increased both the MI values and the number of communicating residues. In Figure 6.4a, ICL3 was observed to communicate with *Thr*68 and *Asn*69 in ICL1, *Met*96 and *Lys*97 in ECL1, *Ser*137 in ICL2, ICL3-itself and *Phe*321 in $H7_{Int}$ with backbone dihedral MI value above 0.1, however it was communicating with the rest of the protein based on both backbone + side-chain dihedrals in Figure 6.4b. As was represented in Figure 6.4b, the highest mean MI values were observed between ICL3-ICL1 and ICL3-$H7_{Int}$ in the *intermediate* state where a number of hydrogen bonds were formed and broken between these sites throughout the simulation.

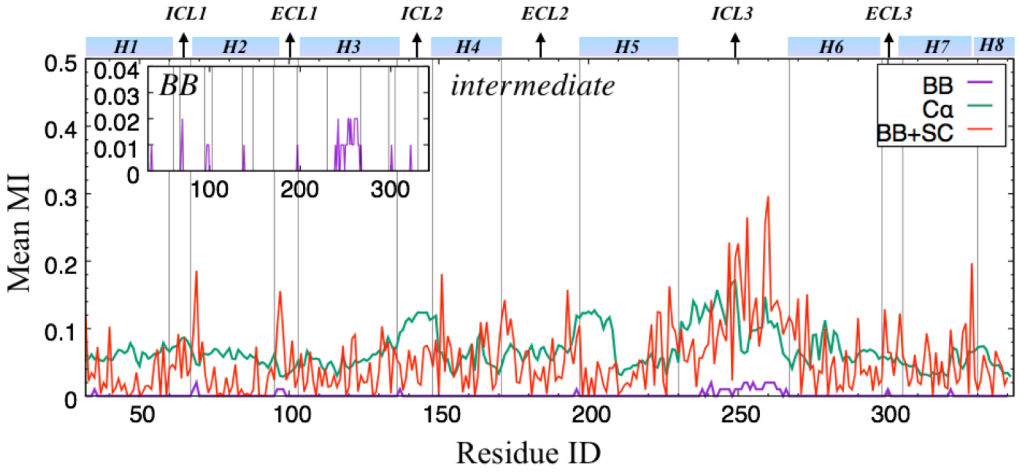
Furthermore, similar to the data obtained in Cα fluctuations, the highest MI values were observed between pairs in spatially close proximity. Therefore, most of the high information sharing residues with MI>1.0 were observed in the diagonal line. However, ICL3 and the residues in ICL1, as well as intracellular H7-H8 junction, shared information with a value of MI> 1.0. The two ends of ECL2 had high MI values. However, the internal helical structure of ECL2 was lost this feature.


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
intermediate
0.0 < MI <= 0.1
0.1 < MI <= 0.2
0.2 < MI <= 0.5
0.5 < MI
Residue ID
50
100
150
200
250
300

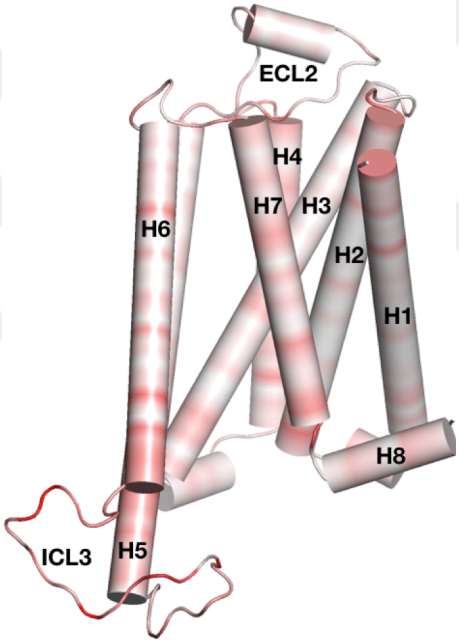
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
intermediate
0.0 < MI <=0.1
0.1 < MI <= 0.25
0.25 < MI <= 0.5
0.5 < MI <= 1
1 < MI <= 2
2 < MI
Residue ID
50
100
150
200
250
300

(b)



(c)



(d)

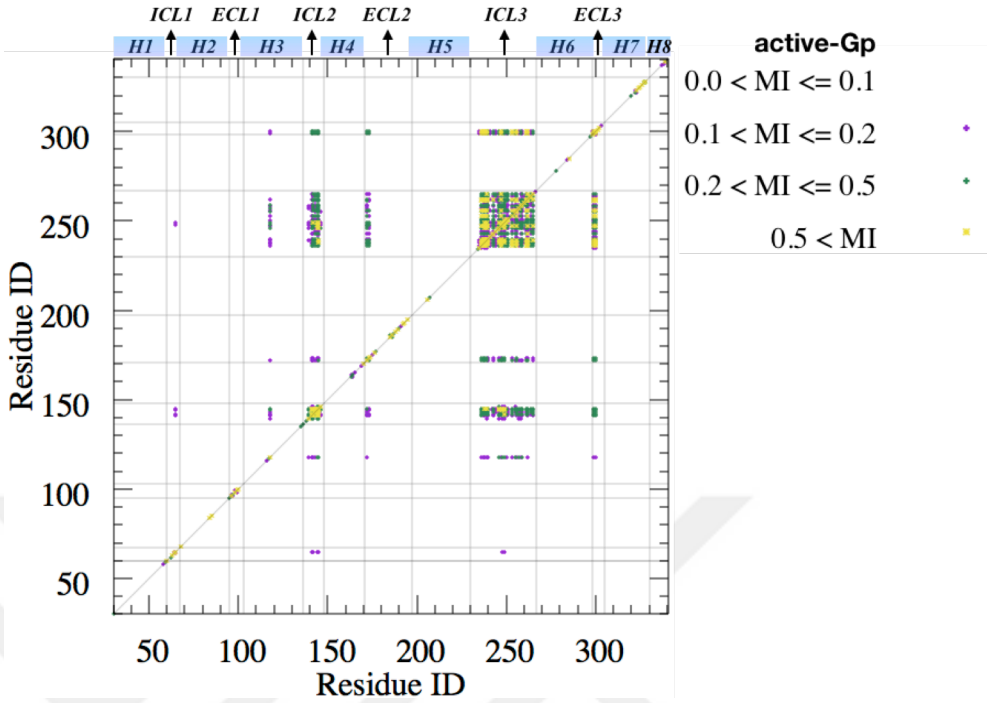
Figure 6. 4 The pairwise MI calculated from backbone (a) and backbone + side-chain dihedral angle fluctuations (b) for the *intermediate* state. Mean MI values of Cα fluctuations (green) were superposed with backbone (magenta) and backbone + side-chain dihedral angles (red) (c). Protein representation based on the mean MI values of backbone + side-chain dihedrals (d).

Among the examined three states, the lowest average MI value for backbone + side-chain dihedrals belonged to Phase II with a value of 0.045, followed by the *intermediate* state with a value of 0.053. The highest average MI was observed in Phase I with a value of 0.085. Similar results were also observed in MI values of Cα fluctuations, where Phase I shared the highest amount of information compared to Phase II and

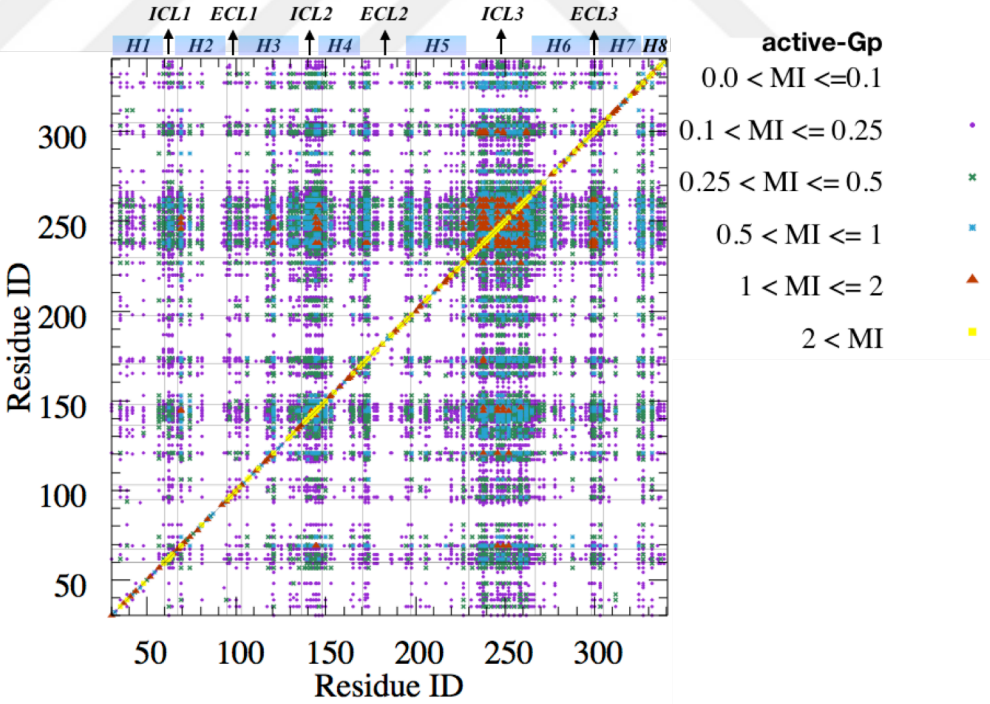
intermediate phase, indicating a correspondence of the mutual information between Cα atoms and dihedral angles. Moreover, almost all loops were involved in information sharing in Phase I, where ICL2 and ECL2 had the highest MI values in Phase II. Interestingly, ICL1, ICL3, and ECL3 were highly involved in information sharing in the *intermediate* state, supported the fluctuation data obtained, and presented in Chapter 5.

Furthermore, two independent MD runs were conducted for the active state bound to the G protein in order to fully capture the common dynamical features of the active receptor. Mean MI profile in *active-Gp* and *active-Gp$_{2nd}$* were highly similar (Figure 6.6e). Except ICL2, which had a higher mean MI value in *active-Gp*, compared to *active-Gp$_{2nd}$*, and ECL2, which had residues with a higher mean MI value in *active-Gp$_{2nd}$* compared to *active-Gp*, most of the remaining residues represented in Figures 6.5 and 6.6 yielded similar MI data. MI differences of ICL2 and ECL2 were due to the differences of their mobilities between two runs, where ICL2 had a higher fluctuation in *active-Gp*. Still, few residues in ECL2 adjacent to H4 was mobile in the *active-Gp$_{2nd}$*, as it was given in Chapter 5. Additionally, a few residues' average MI value of ICL3 was higher in *active-Gp$_{2nd}$* compared to *active-Gp* (Figure 6.6e). This was due to the increased motion of ICL3 in z-direction in the trajectory of *active-Gp$_{2nd.}$*

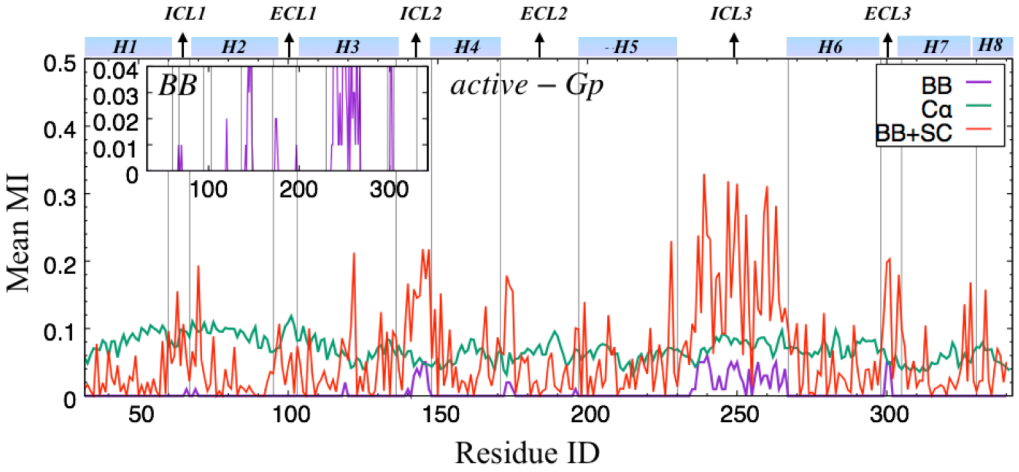
Residues in H1, H2, and H3 strongly communicating through Cα fluctuations in *active-Gp.* However, backbone and side-chain dihedrals in these helices were observed that they did not contribute to this information sharing, considering the dihedral angles were affected by secondary structures. Instead, loops and the residues adjacent to loops had higher MI values compared to transmembrane helices (Figures 6.5 and 6.6). The highest average MI value was observed in ICL1, ICL2, ICL3, and ECL3 in *active-Gp* (Figures 6.5a and 6.5b). ICL2, which interestingly had weak communication with the rest of the receptor in Cα level, started to strongly communicate with ICL3 in both backbone and backbone+side-chain dihedral levels in *active-Gp*. This communication in dihedral angles might be because the helical structure of ICL2 was loosened in the trajectory and reformed again, thus increased its dihedral angle alterations (Figure 6.5c).


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
active-Gp
0.0 < MI <= 0.1
0.1 < MI <= 0.2
0.2 < MI <= 0.5
0.5 < MI
Residue ID
50
100
150
200
250
300
Residue ID

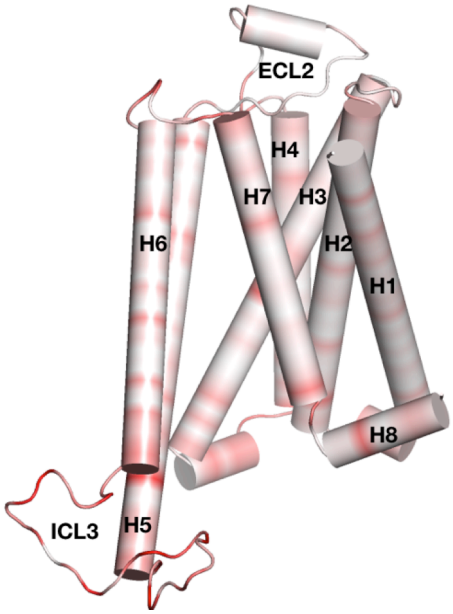
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
active-Gp
0.0 < MI <=0.1
0.1 < MI <= 0.25
0.25 < MI <= 0.5
0.5 < MI <= 1
1 < MI <= 2
2 < MI
Residue ID
50
100
150
200
250
300
Residue ID

(b)



(c)



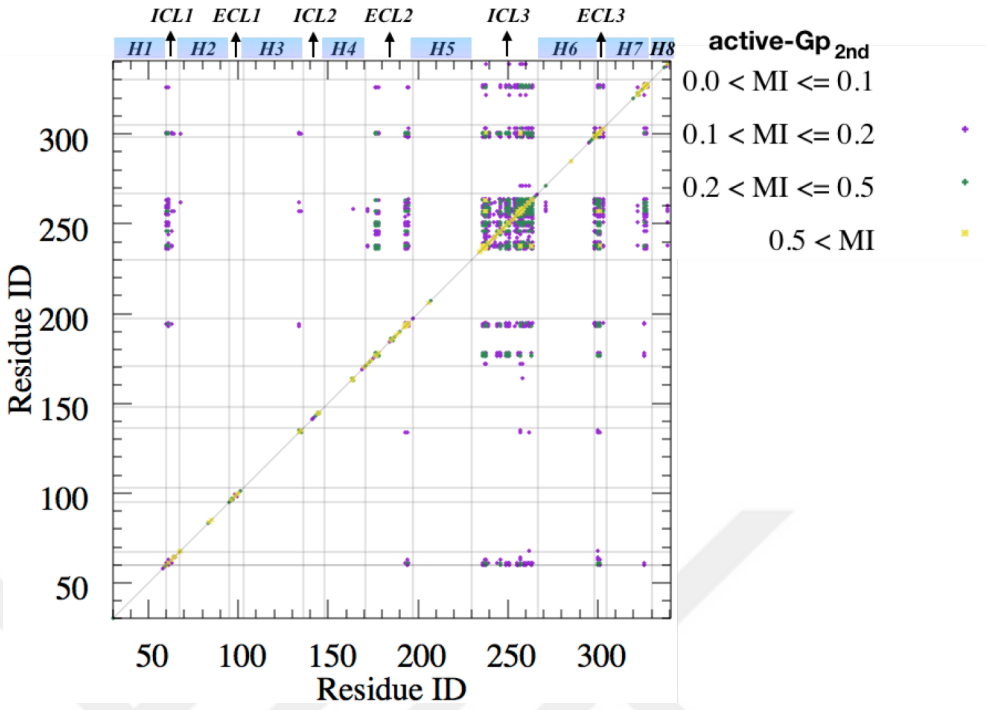
(d)

Figure 6. 5 The heatmap of the backbone (a) and backbone + side-chain dihedrals' (b) mutual information values. Superposition of backbone (magenta), Cα fluctuations (green), and backbone + side-chain (red) dihedrals mean MI values (c). Protein representation of mean MI for *active-Gp* (d).

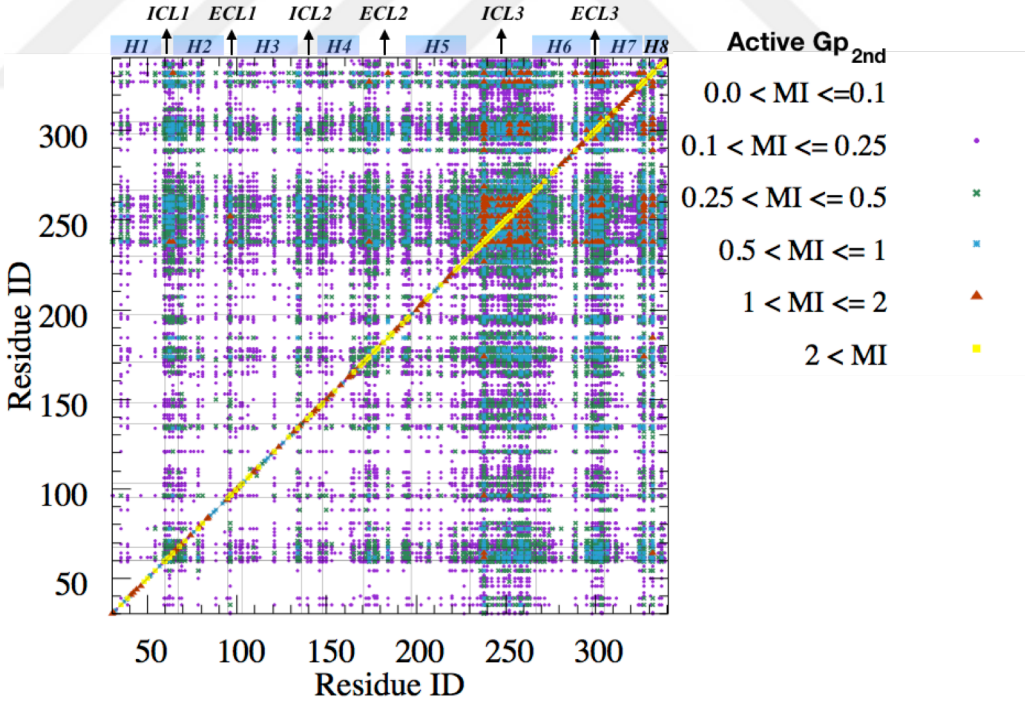
The information sharing in ECL3 and $H7_{Ext,}$ observed in Cα fluctuations, was also observed in the backbone and side-chain dihedral's analysis (Figure 6.6c). Even though the α-helical structure of the G-protein constrained the horizontal backbone mobility of ICL3 in active states, it, apparently, did not restrict the backbone and side-chain angular motions. Thus, it continued to communicate with itself and with other loops.

When active phases were examined according to dihedral angles, a dominant role of the ICL3 on MI was clearly observed. The average MI value in *active-Gp* was 0.059, and it was 0.075 in *active-Gp*$_{2nd}$. These values were higher than mean MI values of the *intermediate* state and Phase II, but lower than those of Phase I. This finding indicated that the highest allosteric communications were observed in the original inactive state in terms of dihedral angle fluctuations. In addition, the α-helical part of G protein in the *active-Gp* system was missing in the *intermediate* state. This intracellular association might induce communication between ICL3 and ICL2, where ICL3 was mostly free in the *intermediate* state except the presence of few transient hydrogen bonds between ICL3 and ICL1 and H7$_{Int}$. Thus, the communication between ICL2 and ICL3 was lost in the *intermediate* state due to their uncoupled motions.

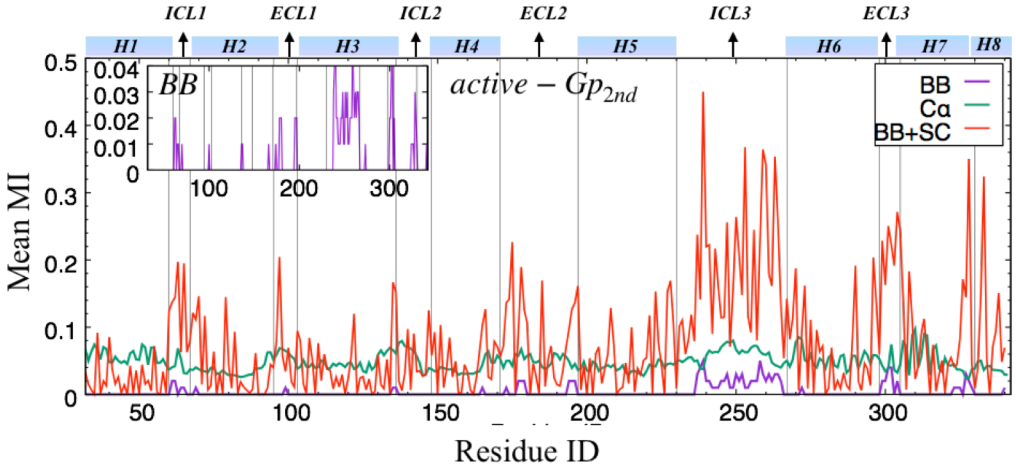
The contribution of H1, H2, and H3 communication in *active-Gp* increased the mean MI value of Cα fluctuations, which was not observed in dihedral data. Interestingly, in *active-Gp* runs, the allosteric interactions of dihedral angles were slightly different than those of Cα fluctuations. In Figures 6.5c and 6.6c, mean MI values of atomic fluctuations of both *active-Gp* and *active-Gp*$_{2nd}$ were plotted together as insets with mean MI values of the backbone (magenta) and backbone + dihedral angles (red). High MI values in Cα fluctuations of transmembranes were generally not observed in mean MI values in dihedral angles, where MI in dihedral angles demonstrated communications in loop structures. The uncorrelation between Cα and dihedral angles was obviously observed in our results as well, thus resulted in contributions of communications in different levels as Cα, backbone dihedrals, and side-chain dihedrals. The combinatorial effect of communications in these levels, consequently, results in the overall protein dynamics in each state.


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
active-Gp 2nd
0.0 < MI <= 0.1
0.1 < MI <= 0.2
0.2 < MI <= 0.5
0.5 < MI
Residue ID
Residue ID
50
100
150
200
250
300

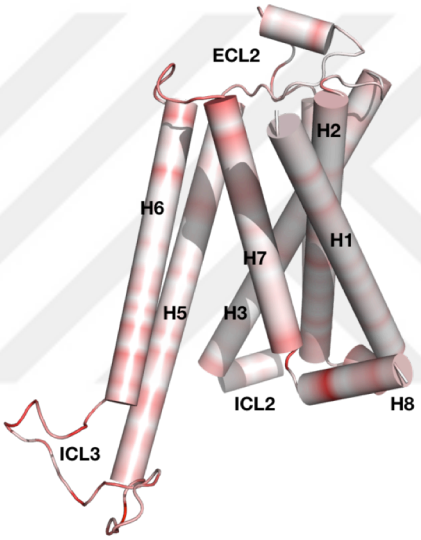
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Active Gp 2nd
0.0 < MI <=0.1
0.1 < MI <= 0.25
0.25 < MI <= 0.5
0.5 < MI <= 1
1 < MI <= 2
2 < MI
Residue ID
Residue ID
50
100
150
200
250
300

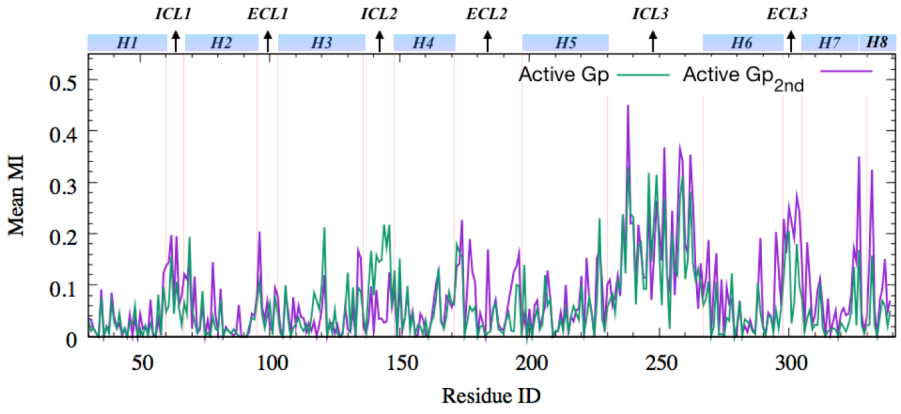
(b)



(c)



(d)



(e)

Figure 6. 6 The MI heatmap of backbone (a) and backbone + side-chain dihedrals (b). Superposition of backbone (magenta), Cα fluctuations (green), and backbone + side-chain (red) dihedrals' mean MI values (c). Protein representation of mean MI for *active-Gp*$_{2nd}$ (d). Mean MI values of backbone + side-chain MI values of *active-Gp* (green) and *active-Gp*$_{2nd}$ (magenta) (e).

Moreover, high information sharing among intracellular parts of the receptor might also be the reason for the intracellular binding of G protein. The analysis of the effect of G protein on the receptor dynamics can be helpful in understanding the environmental aspect of the allostery. However, this analysis is out of the scope of this thesis, thus will be separately examined. ICL2 fluctuation in *active-Gp$_{2nd}$* was the smallest among five states, due to the outward motion of ICL3, thus negatively affecting the communications in both Cα level. Additionally, its helical structure was preserved throughout the trajectory; therefore, limited the communication in the dihedral level, as well. Even the inclusion of side-chain data slightly increased the MI value of ICL2; however, it was still lower than the MI value of ICL2 in *active-Gp*.

6.1.3.3 Mutual information from dihedral angles of proximal residue types

Mutual information analysis for five distinct trajectories revealed the effect of polar residues and the number of rotatable bonds on the pairwise communications. Initially, the proximal pairwise communications among five trajectories were analyzed. In Figure 6.7, the polar-polar pairwise interactions, which were shown with red rectangles in abscissa and ordinate, displayed the highest MI values in all trajectories, even though the number of the proximal amino acid pairs is not significantly high in quantity (Table 6.1 and 6.2). Both hydrophobic (blue) and polar (red) residues were ordered according to the number of their rotatable bonds, as was also given in Chapter 5. In hydrophobic residues, the number of rotatable bonds was between 0 and 3, and in polar residues, this number was between 1 and 4. The mutual information in proximal amino acid pairs indicated that the *Glu-Glu* pairs share the highest information.

The higher mutual information between Cα atoms of the residues was also revealed the importance of polar residues in pairwise communications. It was observed in Chapter 5 that, average MI values between proximal pairs within a contact distance were statistically higher in polar residues compared to hydrophobic residues except for *active-Gp$_{2nd}$* (Tables 5.5 and 5.6), which was mainly due to the involvement of helices to the communications in active phases. Since the polar residues are mostly locating in loops, the helical communications in active states increased the effect of hydrophobic

residues on the mutual information. However, the side-chain information was not present in Cα analysis; but still, identification of the role of polarity in almost all states was valuable.

This time, the dihedral data clearly demonstrated the impact of polarity on pairwise communication. Between the proximal polar residue-pair interactions, the highest dihedral MI values were observed in *active-Gp*$_{2nd}$, followed by Phase I, then *active-Gp*, Phase II, and *intermediate* state (Table 6.1). The proximal hydrophobic pairs indicated that the highest MI was observed in Phase I, followed by *active-Gp*$_{2nd}$, *active-Gp*, *intermediate*, and finally, in Phase II. Thus, the pairwise communications within 5 residues were the highest in Phase I and interestingly in *active-Gp*$_{2nd}$, where *active-Gp*$_{2nd}$ had one of the lowest mean MI values based on Cα fluctuations (Figure 5.23d). This finding demonstrated the fact that, even though positional fluctuations of Cα atoms were limited in phases with extended ICL3 conformation, angular motions in the backbone and side-chain dihedrals were higher compared to the phases where ICL3 was packed.

The number of hydrophobic amino acids is 197 out of 311 residues in the inactive state, compared with 114 polar residues. However, the mutual information of polar residues dominated in all states. Still, a few distinct hydrophobic amino acids were clearly representing higher information sharing ability compared to others. For example, *Gly* and *Cys* were observed as sharing information with the proximally located polar amino acids in Phase I, *intermediate* state, and *active-Gp*.

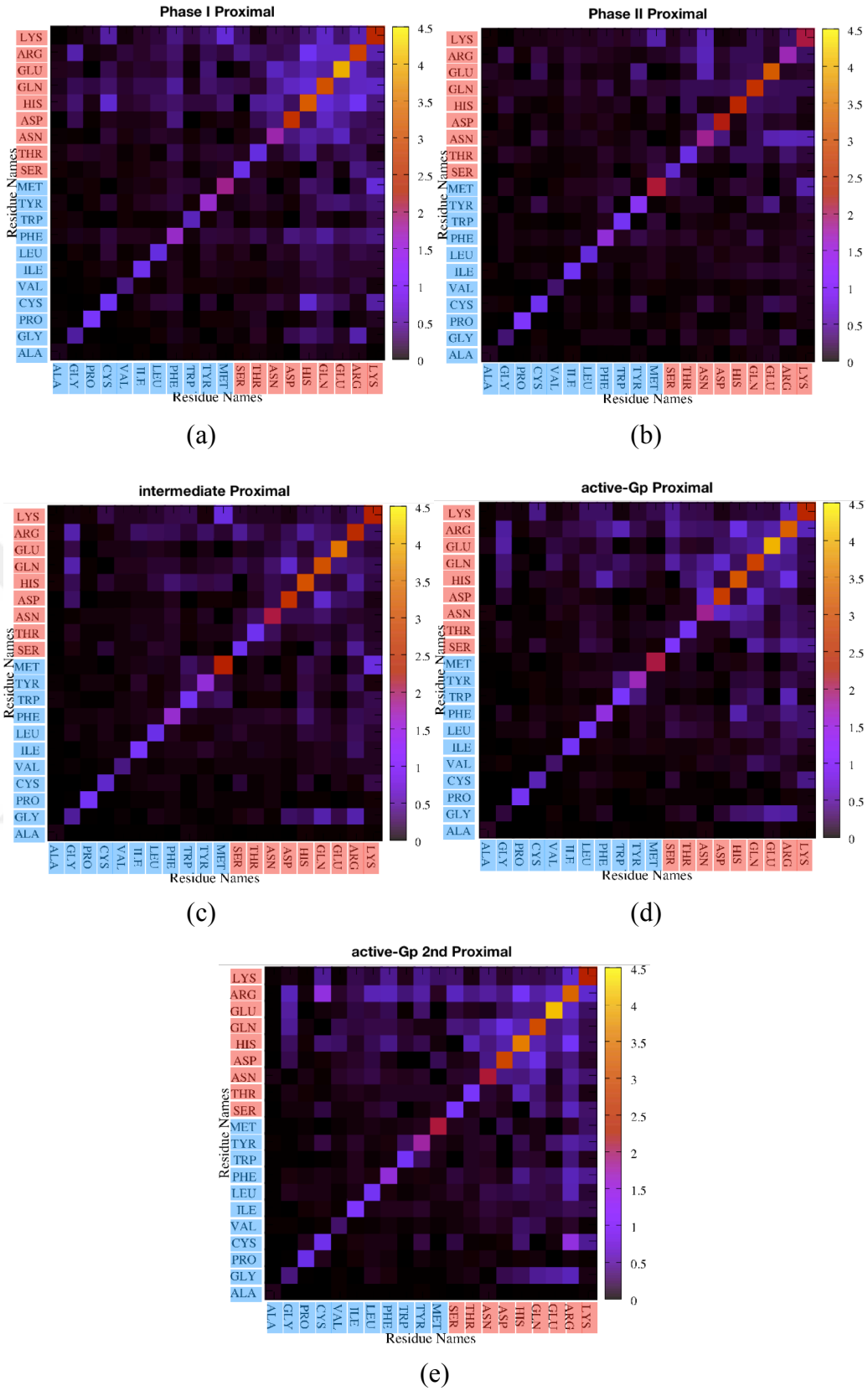


Figure 6. 7 Mutual information between proximal amino acid pairs in Phase I (a), Phase II (b), *intermediate* state (c) and *active-Gp* (d), and *active-Gp*$_{2nd}$ (e). Hydrophobic residues were shown in blue and polar residues were shown in red. The residues were ordered according to increasing rotatable bonds.

Figure 6.8 represented the average mutual information values per proximal residue types. The amino acids in the abscissa were ordered based on the increasing number of rotatable bonds similar to Figure 6.7. From Phase I to Phase II, the contribution of MI values of *His* and *Arg* decreased. *Gln* was observed to have the highest MI contribution among all amino acids (Figure 6.8a and 6.8b). In Phase I, the contribution of *Phe* to MI was also observed. Additionally, in the *intermediate* state, *Gln* had the highest MI value. However, the effect of *Phe* was decreased, and *Met* was increased (Figure 6.8c). Active phases, on the other hand, represented a slightly higher contribution of MI value for *Arg* residue compared to other states (Figures 6.8d and 6.8e). Thus, inactive and *intermediate* states displayed *Gln* as the highest communicating residue; however, active phases represented *Arg* with the highest MI. *Arg* was found in ICL3 as well as ICL1, $H3_{Int}$, $H7_{Int}$, and H8, adjacent to the G protein binding site.

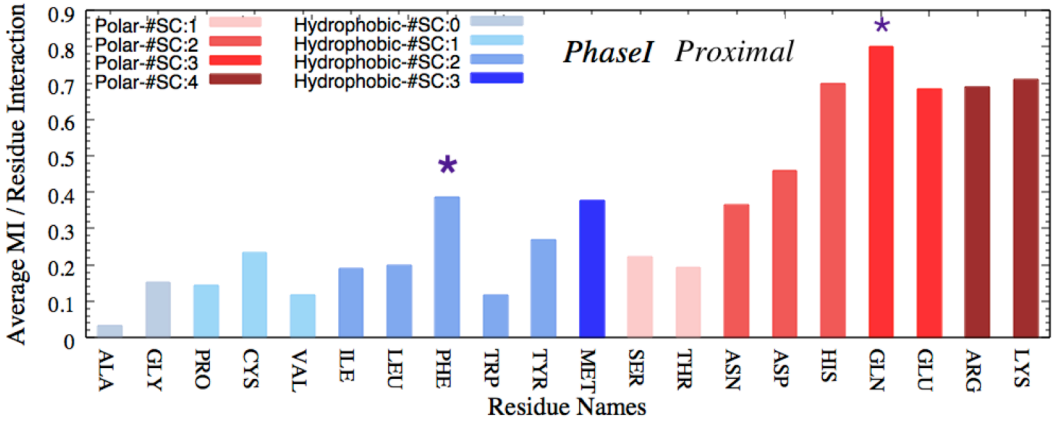
Upon association with the α-helical domain of G protein, these *Arg* residues became prominent in pairwise communications. Again, the average MI values of polar residues were more than double in quantity compared to hydrophobic residues as given in Tables 6.1 and 6.2 for proximal pairs. Moreover, the number of rotatable bonds seemed to have an effect on information sharing. Average MI values significantly increased along with the increase in the number of rotatable bonds both in hydrophobic and polar residues. Interestingly, *Phe* and *Gln* violated this finding in proximal residues. *Phe* had more mean MI value compared to *Met* in Phase I (Figure 6.8a); and *Gln*, had more MI value than *Arg* and *Lys* in three states (Figures 6.8a, 6.8b, and 6.8c). Still, the low number of residue types within five amino acids might be affecting these calculations in proximal pairs. Finally, hydrophobic residues were observed to have almost 30% contribution to the information sharing, while the remaining 70% belonged to polar residues.

Table 6. 1 Average MI values of proximal polar and proximal hydrophobic residues.

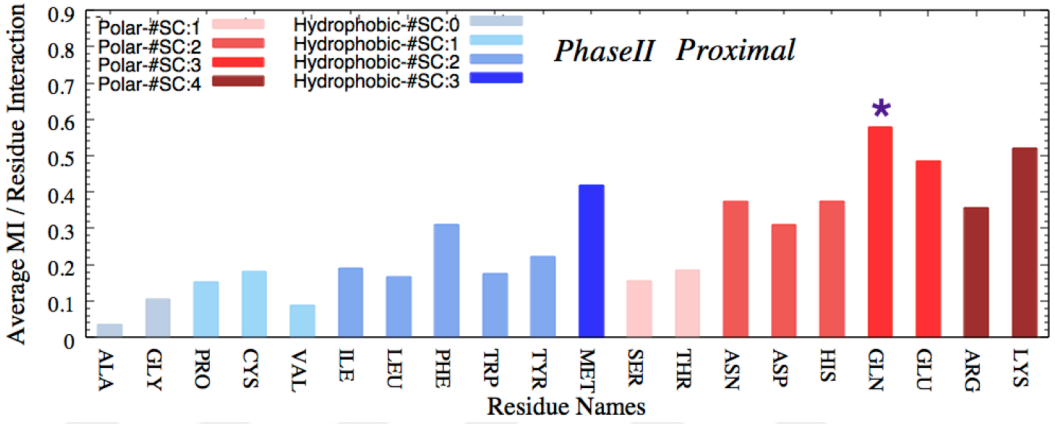
	Type	Proximity	Mean Mutual Information	# of Residues
Phase I	Polar	Proximal	0.228 ± 0.288	912
	Hydrophobic	Proximal	0.076 ± 0.141	1556
Phase II	Polar	Proximal	0.117 ± 0.185	912
	Hydrophobic	Proximal	0.050 ± 0.103	1556
intermediate	Polar	Proximal	0.151 ± 0.216	915
	Hydrophobic	Proximal	0.056 ± 0.108	1561
active-Gp	Polar	Proximal	0.206 ± 0.309	915
	Hydrophobic	Proximal	0.058 ± 0.146	1561
active-Gp$_{2nd}$	Polar	Proximal	0.232 ± 0.319	915
	Hydrophobic	Proximal	0.070 ± 0.160	1561

Table 6. 2 Mean MI values for pairwise and proximal residue types based on their polarity.

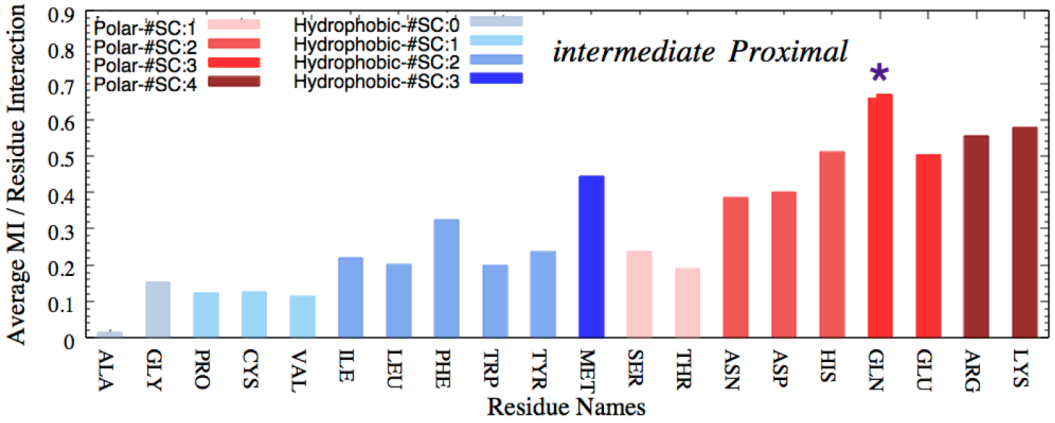
	Type	Proximity	Mean Mutual Information	# of Residue-pairs
Phase I	Polar-Polar	Proximal	0.370 ± 0.351	370
	Polar-Hydrophobic	Proximal	0.132 ± 0.181	1084
	Hydrophobic-Hydrophobic	Proximal	0.045 ± 0.102	1014
Phase II	Polar-Polar	Proximal	0.178 ± 0.228	370
	Polar-Hydrophobic	Proximal	0.076 ± 0.133	1084
	Hydrophobic-Hydrophobic	Proximal	0.036 ± 0.080	1014
intermediate	Polar-Polar	Proximal	0.238 ± 0.271	370
	Polar-Hydrophobic	Proximal	0.091 ± 0.140	1090
	Hydrophobic-Hydrophobic	Proximal	0.037 ± 0.080	1016
active-Gp	Polar-Polar	Proximal	0.337 ± 0.376	370
	Polar-Hydrophobic	Proximal	0.116 ± 0.209	1090
	Hydrophobic-Hydrophobic	Proximal	0.028 ± 0.083	1016
active-Gp$_{2nd}$	Polar-Polar	Proximal	0.366 ± 0.375	370
	Polar-Hydrophobic	Proximal	0.140 ± 0.232	1090
	Hydrophobic-Hydrophobic	Proximal	0.034 ± 0.080	1016


Polar-#SC:1
Polar-#SC:2
Polar-#SC:3
Polar-#SC:4
Hydrophobic-#SC:0
Hydrophobic-#SC:1
Hydrophobic-#SC:2
Hydrophobic-#SC:3
PhaseI Proximal
Average MI / Residue Interaction
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
ALA GLY PRO CYS VAL ILE LEU PHE TRP TYR MET SER THR ASN ASP HIS GLN GLU ARG LYS
Residue Names

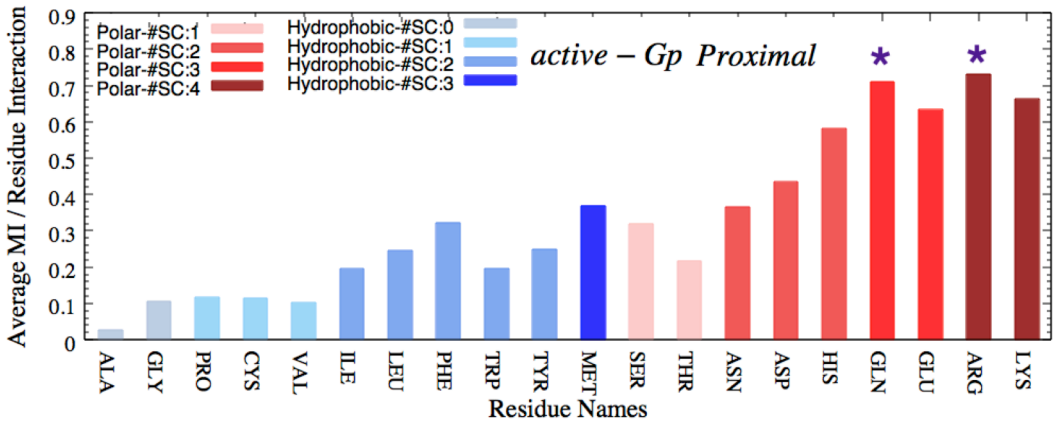
(a)


Polar-#SC:1
Polar-#SC:2
Polar-#SC:3
Polar-#SC:4
Hydrophobic-#SC:0
Hydrophobic-#SC:1
Hydrophobic-#SC:2
Hydrophobic-#SC:3
PhaseII Proximal
Average MI / Residue Interaction
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
ALA GLY PRO CYS VAL ILE LEU PHE TRP TYR MET SER THR ASN ASP HIS GLN GLU ARG LYS
Residue Names

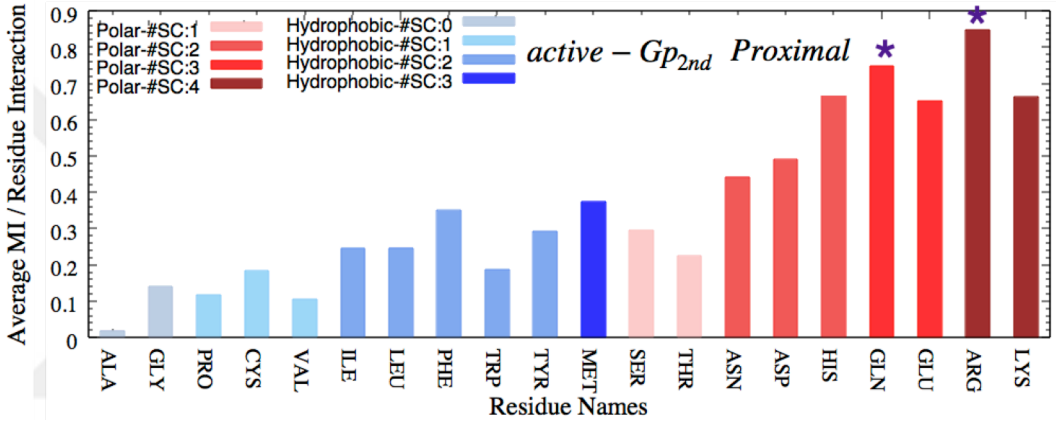
(b)


Polar-#SC:1
Polar-#SC:2
Polar-#SC:3
Polar-#SC:4
Hydrophobic-#SC:0
Hydrophobic-#SC:1
Hydrophobic-#SC:2
Hydrophobic-#SC:3
intermediate Proximal
Average MI / Residue Interaction
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
ALA GLY PRO CYS VAL ILE LEU PHE TRP TYR MET SER THR ASN ASP HIS GLN GLU ARG LYS
Residue Names

(c)



(d)



(e)

Figure 6. 8 Proximal polar and hydrophobic residues' average MI values in Phase I (a), Phase II (b), *intermediate* state (c) *active-Gp* (d), and.*active-Gp*$_{2nd}$ (e). Hydrophobic residues were shown in blue and polar residues were shown in red. #SC = number of rotatable bonds. Asterisks represented the residue types with high MI values.

6.1.3.4 Mean mutual information from dihedral angles of distal residues

In distal amino acid pairs, the MI values of polar-polar residue interactions were more detectable compared to the proximal data. Among five phases, the highest MI value was also belonging to Phase I, as it was in MI values of Cα fluctuations expressed in Chapter 5. Interestingly, the average MI in the *intermediate* state was as low as Phase II, as opposed to the mean MI value from the Cα fluctuations. Additionally, the strong information-sharing ability of *Arg* with other amino acids was observed in five phases. Even though *Gly* does not contain a torsional angle, it also represented a high

information sharing ability compared to *Ala* and *Pro,* shown in Figure 6.9. *Trp* in Phase I contributed less to the mutual information compared to other phases (Figure 6.9a). Also, *Ser* in Phase II, and *Thr* in *active-Gp* contributed less to the information sharing among all states (Figure 6.9b and 6.9d). These residue contribution differences might result in the conformational changes in these states. For example, less involvement of *Ser* to MI might be needed for the receptor to transform into a very inactive state by changing its internal communication pathways since it is known to play a role in agonist interaction during activation of the receptor.

Residues, given in table 6.3 and 6.4, represented significantly high mean MI values of polar residues compared to hydrophobic ones in all states. For instance, mean MI value for polar amino acids in Phase I was calculated as 0.125 ± 0.187; however, it was 0.051 ± 0.096 for hydrophobic residues. Mean MI values for polar – polar interactions were also observed to be higher than polar–hydrophobic and hydrophobic–hydrophobic pairs. Mean MI for polar–polar pairs in Phase I was 0.201 ± 0.245, while it was 0.081 ± 0.124 for polar–hydrophobic pairs and 0.034 ± 0.069 for hydrophobic–hydrophobic pairs. In Chapter 5, the calculation of mutual information in Cα fluctuations represented preliminary data to disclose the higher contributions of the polar residues, except active states, where the internal G protein association might be changing the electrostatics of the residues. However, dihedral data clearly demonstrated the effect of polarity on pairwise communications.

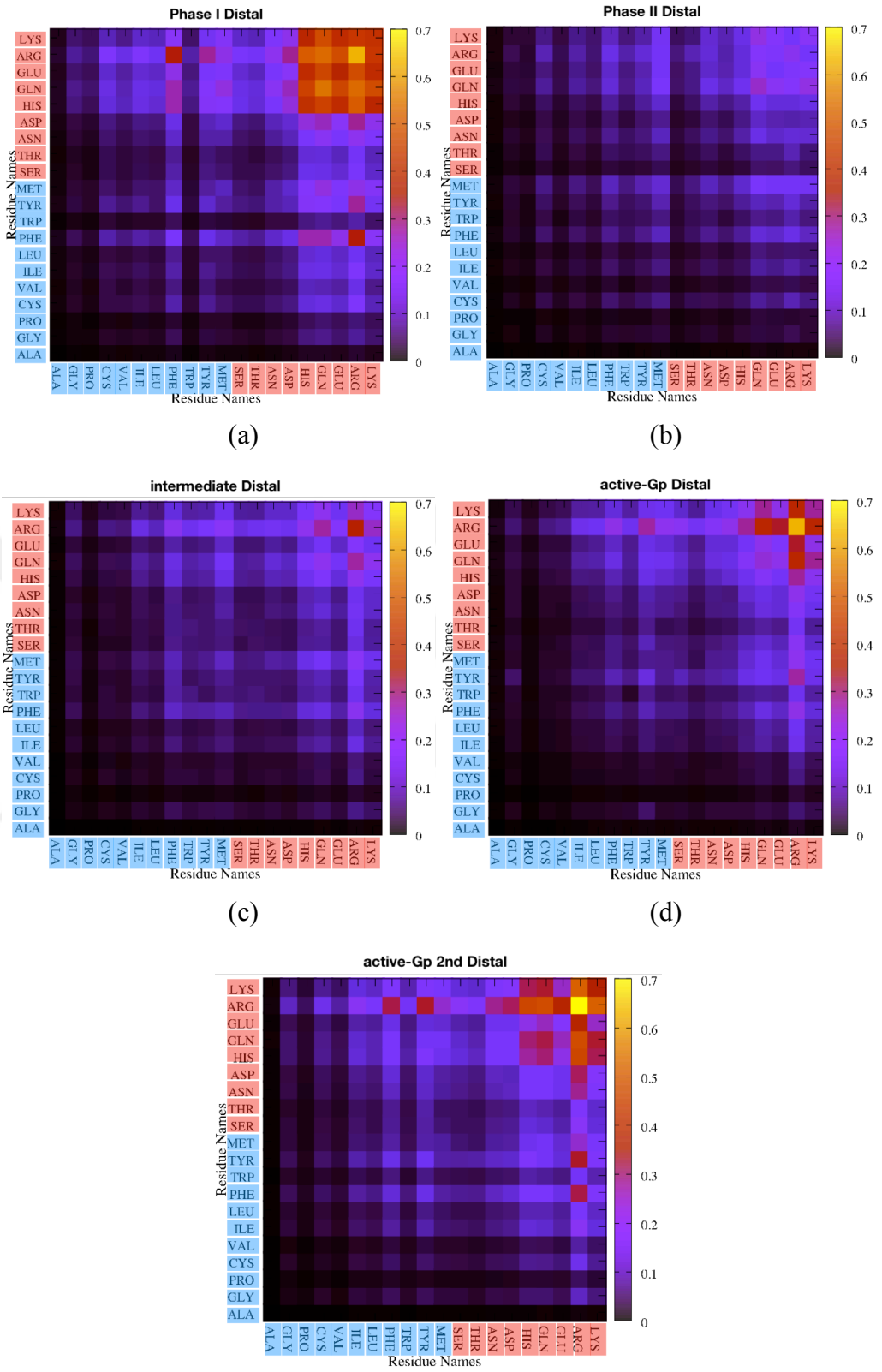


Figure 6. 9 MI heatmap between distal amino acid pairs in Phase I (a), Phase II (b), *intermediate* state (c), *active-Gp* (d), and *active-Gp*$_{2nd}$ (e). Hydrophobic residues were shown in blue and polar residues were shown in red with the order of increasing rotatable bonds.

Table 6. 3 Average MI values of distal polar and distal hydrophobic residues.

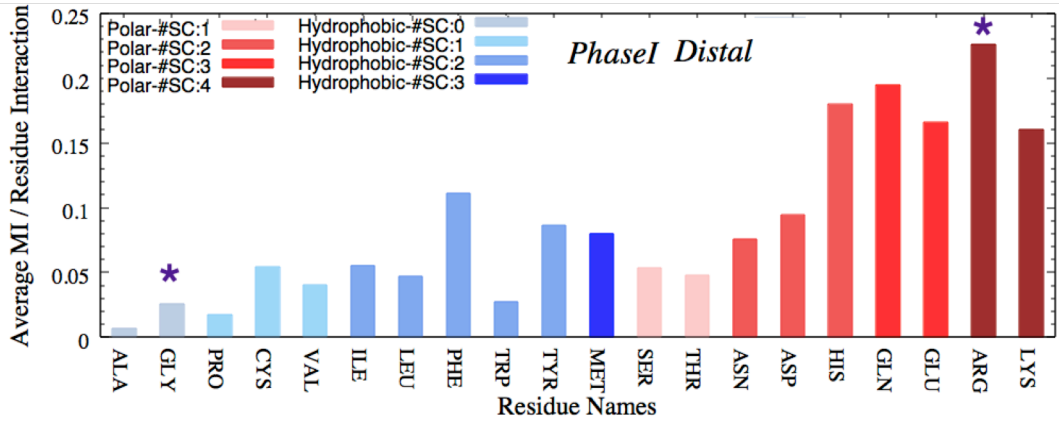
	Type	Proximity	Mean Mutual Information	# of Residues
Phase I	Polar	Distal	0.125 ± 0.187	34428
	Hydrophobic	Distal	0.051 ± 0.096	59514
Phase II	Polar	Distal	0.058 ± 0.099	34428
	Hydrophobic	Distal	0.029 ± 0.060	59514
intermediate	Polar	Distal	0.069 ± 0.107	34850
	Hydrophobic	Distal	0.033 ± 0.062	59706
active-Gp	Polar	Distal	0.086 ± 0.152	34850
	Hydrophobic	Distal	0.033 ± 0.083	59706
active-Gp$_{2nd}$	Polar	Distal	0.109 ± 0.174	34850
	Hydrophobic	Distal	0.044 ± 0.093	59706

Table 6. 4 Average MI values for pairwise and distal residue types based on their polarity.

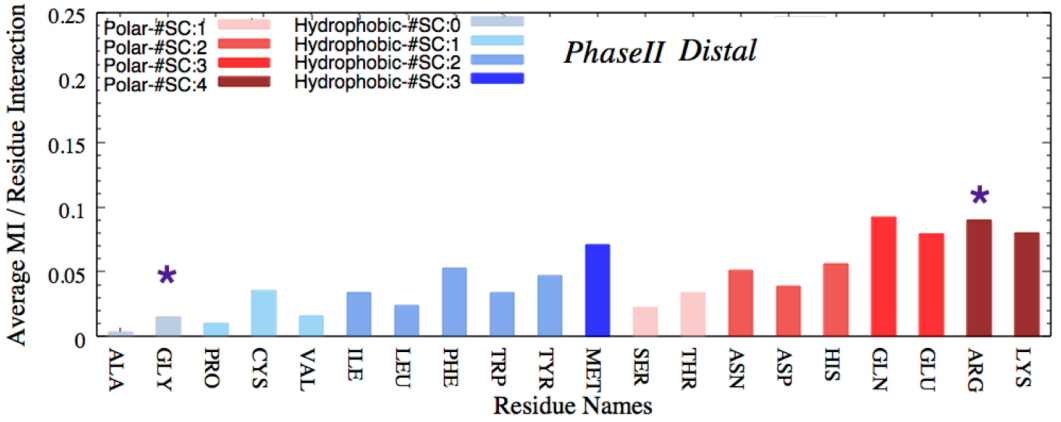
	Type	Proximity	Mean Mutual Information	# of Residue-pairs
Phase I	Polar-Polar	Distal	0.201 ± 0.245	12512
	Polar-Hydrophobic	Distal	0.081 ± 0.124	43832
	Hydrophobic-Hydrophobic	Distal	0.034 ± 0.069	37598
Phase II	Polar-Polar	Distal	0.087 ± 0.126	12512
	Polar-Hydrophobic	Distal	0.042 ± 0.074	43832
	Hydrophobic-Hydrophobic	Distal	0.021 ± 0.049	37598
intermediate	Polar-Polar	Distal	0.106 ± 0.135	12740
	Polar-Hydrophobic	Distal	0.049 ± 0.079	44220
	Hydrophobic-Hydrophobic	Distal	0.023 ± 0.047	37596
active-Gp	Polar-Polar	Distal	0.143 ± 0.196	12740
	Polar-Hydrophobic	Distal	0.054 ± 0.107	44220
	Hydrophobic-Hydrophobic	Distal	0.021 ± 0.062	37596
active-Gp$_{2nd}$	Polar-Polar	Distal	0.178 ± 0.224	12740
	Polar-Hydrophobic	Distal	0.070 ± 0.120	44220
	Hydrophobic-Hydrophobic	Distal	0.029 ± 0.069	37596

Under physiological conditions, *Arg* and *Lys* are two polar, positively charged basic amino acids and mostly localized on protein surfaces (Kumar *et al*., 2000; Yokota *et al*., 2006). These two amino acids are extremely useful in maintaining protein stability by forming ionic interactions and hydrogen bonds with other residues and with water molecules (Barlow and Thornton, 1983; Strickler *et al*., 2006). Even though *Arg* and

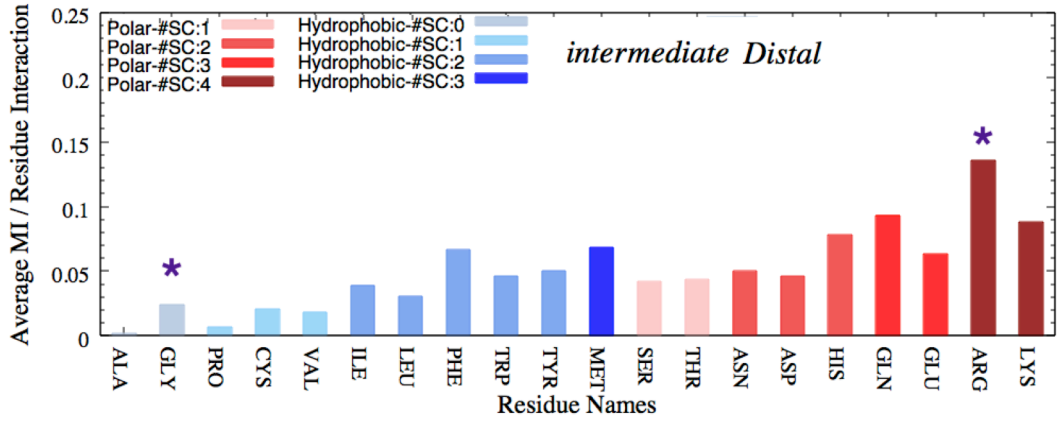
Lys are both basic residues, the *Arg* residue was reported to provide more stability to the protein structure than *Lys* due to its geometric structure. *Arg* residue contains guanidinium group that interacts in three possible directions by its three asymmetrical nitrogen atoms (N^{ε}, $N^{\eta 1}$, $N^{\eta 2}$), whereas *Lys* has only one functional group allowing only one direction of interaction (Borders *et al*., 1994; Donald, Kulp, and DeGrado, 2011). Therefore, *Arg* is able to form a number of salt-bridges and hydrogen bonds compared to *Lys*; thus, these interactions are going to be more persistent and strong compared to the interactions of *Lys*. Moreover, the equilibrium thermodynamic data was reported that the side-chain of *Arg* was still positively charged even when the residue was removed from bulk water and buried, at least partially, in the relatively hydrophobic interior of a globular protein (Harms *et al*., 2011). Thus, observing *Arg*, instead of *Lys*, with high MI values in all five conformationally distinct states can be explained by its persistent interactions with a number of residues and its significant role in protein dynamics and stability (Figures 6.9 and 6.10).


Polar-#SC:1
Polar-#SC:2
Polar-#SC:3
Polar-#SC:4
Hydrophobic-#SC:0
Hydrophobic-#SC:1
Hydrophobic-#SC:2
Hydrophobic-#SC:3
PhaseI Distal
Average MI / Residue Interaction
0.25
0.2
0.15
0.1
0.05
0
ALA GLY PRO CYS VAL ILE LEU PHE TRP TYR MET SER THR ASN ASP HIS GLN GLU ARG LYS
Residue Names

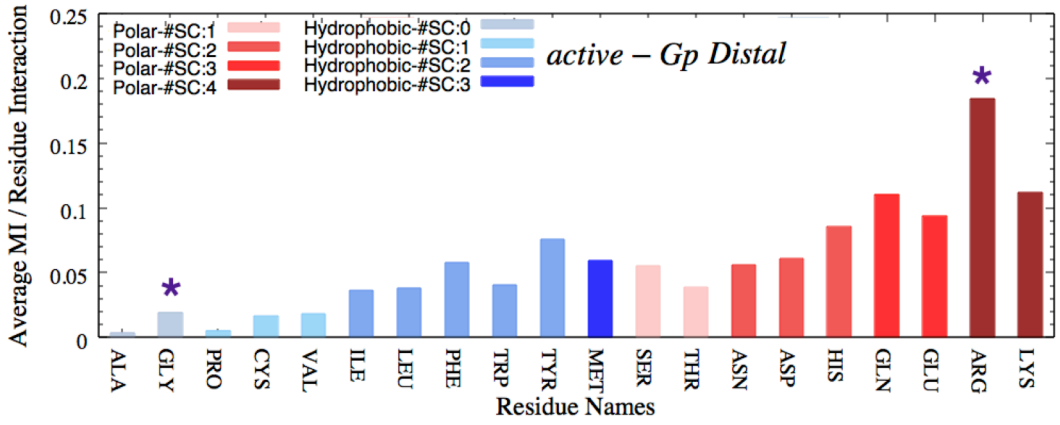
(a)


Polar-#SC:1
Polar-#SC:2
Polar-#SC:3
Polar-#SC:4
Hydrophobic-#SC:0
Hydrophobic-#SC:1
Hydrophobic-#SC:2
Hydrophobic-#SC:3
PhaseII Distal
Average MI / Residue Interaction
0.25
0.2
0.15
0.1
0.05
0
ALA GLY PRO CYS VAL ILE LEU PHE TRP TYR MET SER THR ASN ASP HIS GLN GLU ARG LYS
Residue Names

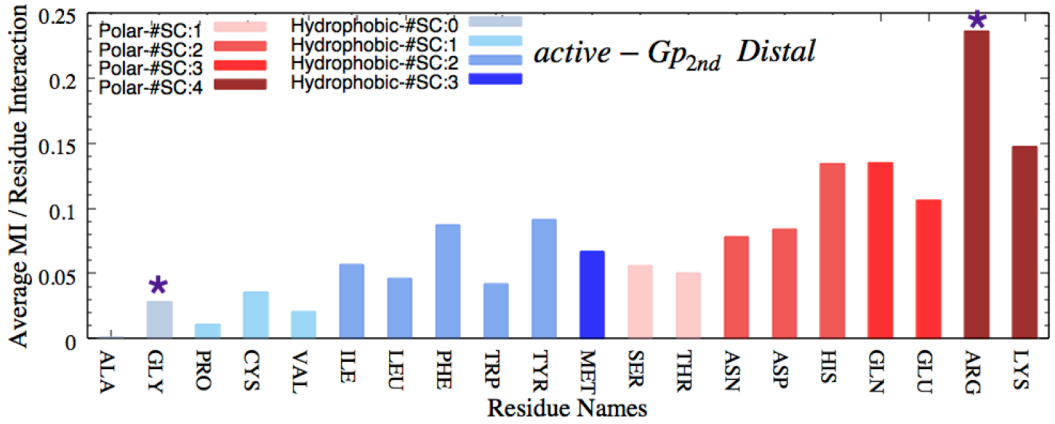
(b)


Polar-#SC:1
Polar-#SC:2
Polar-#SC:3
Polar-#SC:4
Hydrophobic-#SC:0
Hydrophobic-#SC:1
Hydrophobic-#SC:2
Hydrophobic-#SC:3
intermediate Distal
Average MI / Residue Interaction
0.25
0.2
0.15
0.1
0.05
0
ALA GLY PRO CYS VAL ILE LEU PHE TRP TYR MET SER THR ASN ASP HIS GLN GLU ARG LYS
Residue Names

(c)



(d)



(e)

Figure 6. 10 Average MI per residue type for distal residues in Phase I (a), Phase II (b), *intermediate* state (c), *active-Gp* (d), and *active-Gp*$_{2nd}$ (e). #SC = number of rotatable bonds. Hydrophobic residues were colored with blue and polar residues were colored with red.

It was reported that allosteric and orthosteric sites also have specific amino acid compositions that were related to the mechanisms and helpful in allosteric site predictions (Li *et al*., 2013; Song *et al*., 2017; Yang, Seo, Jang, Jung, and Kim, 2012). Apparently, allosteric sites contained more hydrophobic residues, as *Leu*, *Val, Ile*, *Phe,* and *Pro*, compared to orthosteric sites that were rich in *Ser, Tyr, Thr, Asp*, *His*, *Asn*, and *Glu.* These characteristic site compositions could be attributed to functionalities undertaken by the sites. Thus, besides the role of loops, the role of orthosteric and allosteric ligand binding sites can also be considered when analyzing the differences of MI contributions of the residue types between states. Moreover, it was shown that hydrophobic residues playing a role in maintaining the core structure of proteins; thus, they have less potential energy. However, polar interactions contribute mostly to

entropies. Thus, protein stability was dictated by a balance between hydrophobic interactions' energetic contribution and polar and charged interactions' entropic contributions.

With all this regard, our data was suggesting the higher communicating residues were polar ones, indicating those either locating in the loops or interacting with the ligands covalently (E.g., *Arg, Lys, Gln*). However, the hydrophobic residue types such as *Met, Tyr*, and *Phe* were reported to be present in binding pockets and play a role in the orientation of the allosteric or orthosteric ligands. Thus, the high value of information sharing between distal pairs as *Phe-Arg*, in Phase I, *Met-Arg* in the *intermediate* state, or *Tyr-Arg* and *Phe-Arg* in *active-Gp* could be preparing the systems for proper ligand orientation and interaction (Figure 6.10).

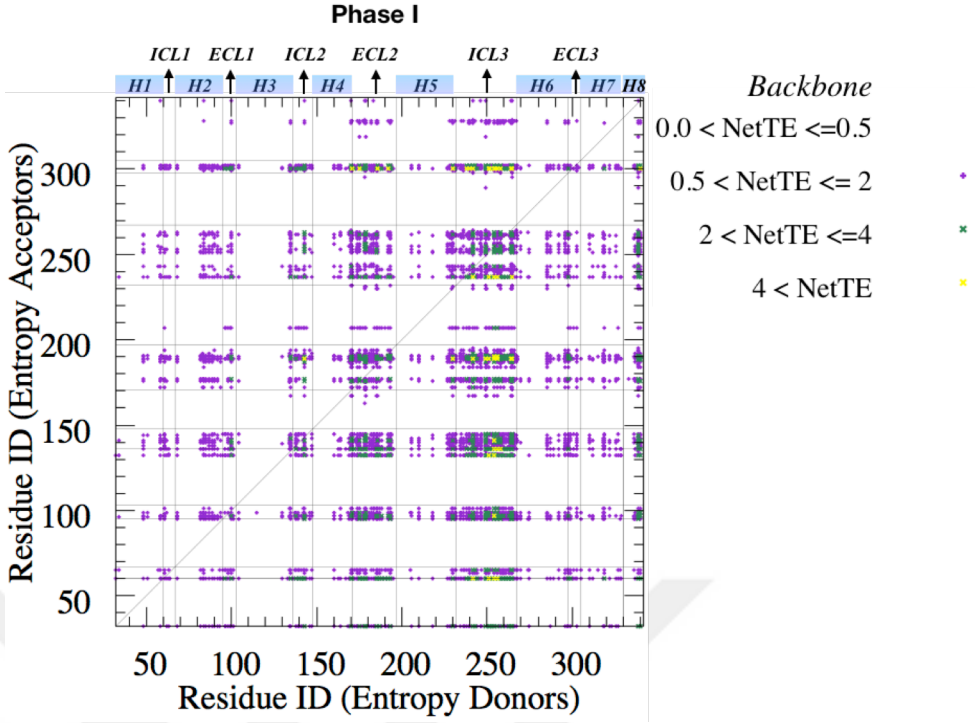
6.1.4 Net transfer entropy based on the backbone and side-chain dihedral fluctuations

6.1.4.1 Comparison of the entropy transfer profiles in the original inactive conformational states of β_2AR

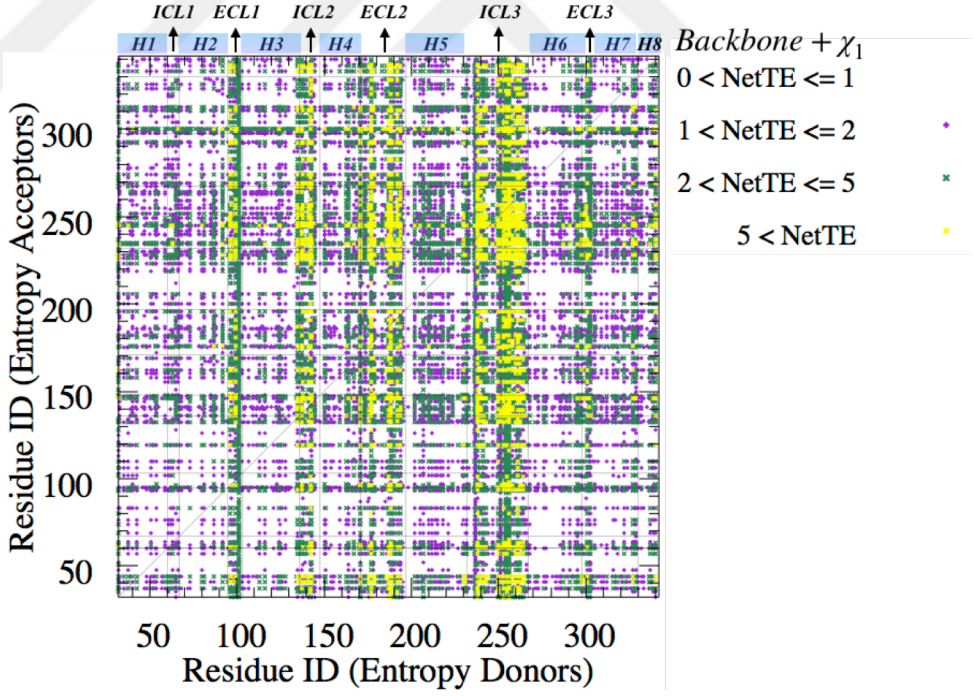
Changes in the conformational entropies of protein side-chains were believed to make important contributions to the overall change in the free energy of molecular events (A. Joshua Wand and Sharp, 2018). Therefore, analysis of the intrinsic dynamics of the receptor needed the data of side-chain dihedral angle calculations, as well. These calculations were based on the number of rotatable bonds and their occurrences in the pre-defined number of bins, i.e., degree of angles, of the amino acids. Transfer entropy, as mentioned before, is the calculation of the information contribution of the past state of one residue to the system, when past and future states of another residue were known. Entropy transfer from dihedrals of backbone and side-chain transfer entropy calculations were calculated and compared with the transfer entropy data from Cα fluctuations.

Transfer entropy for backbone and side-chain dihedrals were calculated based on the dihedral angle changes for residues *i* and *j*. Backbone dihedrals' net TE analysis revealed that ICL3 and remaining loops, as well as *Phe*208 in H5, shared strong information in Phase I. Residues on ECL2 adjacent to H4 were also involved in information sharing. On the other hand, ICL1, ICL2 ends adjacent to H3 and H4, and ECL3 were strong entropy acceptors in Phase I. Not much change was observed in terms of transfer entropy values of transmembrane regions since their mobilities were limited compared to the mobilities of loop regions (Figure 6.11a). This indicated the backbone dihedral angle changes in loops, including ECL2 and ICL3, affected the future changes in ICL1, ECL1, ICL2, and ECL3. Transmembrane helices, however, were neutral. Moreover, upon the inclusion of one rotatable bond, i.e., χ_1, to the backbone transfer entropy data significantly increased the net transfer entropy values. This time, non-negative net transfer entropy values of the transmembrane helices were also observed, as seen in Figure 6.11b and the inset with a blue line in Figure 6.11c. The inclusion of χ_1 data broadened the donor characteristics of loops. Especially, ECL1, ICL2, ECL2, and ICL3 transferred strong information to the rest of the protein. The entropy accepting the strength of the residues was not as high and populated as the entropy donor strength, which was also observed in backbone dihedral analysis.

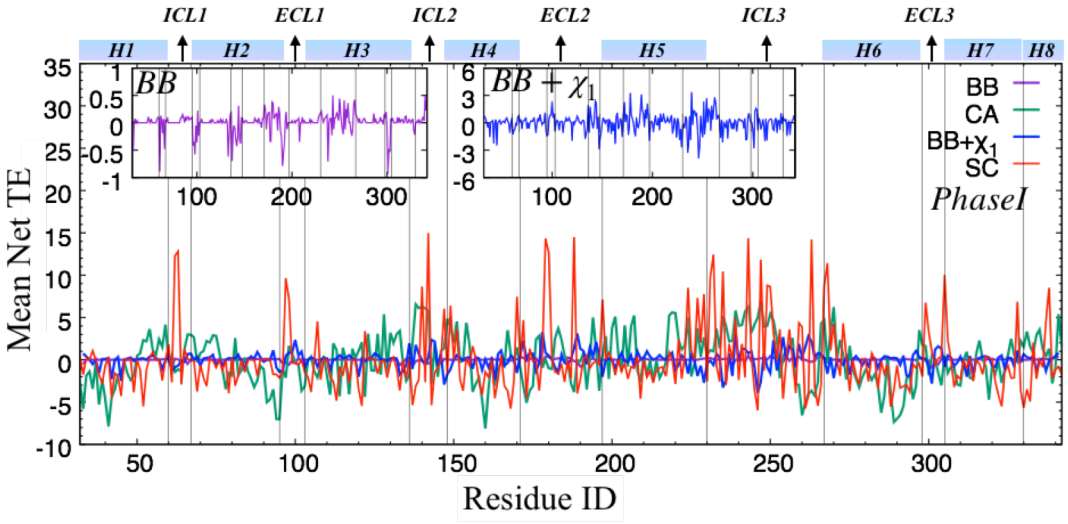
Comparison of the backbone + χ_1 net TE values with the Cα fluctuations in Figure 6.11c resulted that extracellular loops are donating information together with intracellular loops. It was observed in Cα fluctuations that the information transfer was from intracellular to extracellular in Phase I; however, here, the transfer from all loops to all transmembranes was noticed. Thus, net TE values of Cα fluctuations and dihedral data differed similar to the mutual information data, supporting the data that there are low correlations between these two parameters. Transfer entropy data for backbone dihedrals yielded more confined areas compared to backbone + χ_1 and side-chain dihedrals. The inclusion of only one side-chain torsional angle significantly changed the transfer entropy profile of the phase I, demonstrating the strong effect of the side-chain dihedrals on the overall protein dynamics.


Phase I
ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Backbone
0.0 < NetTE <=0.5
0.5 < NetTE <= 2
2 < NetTE <=4
4 < NetTE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)
50
100
150
200
250
300

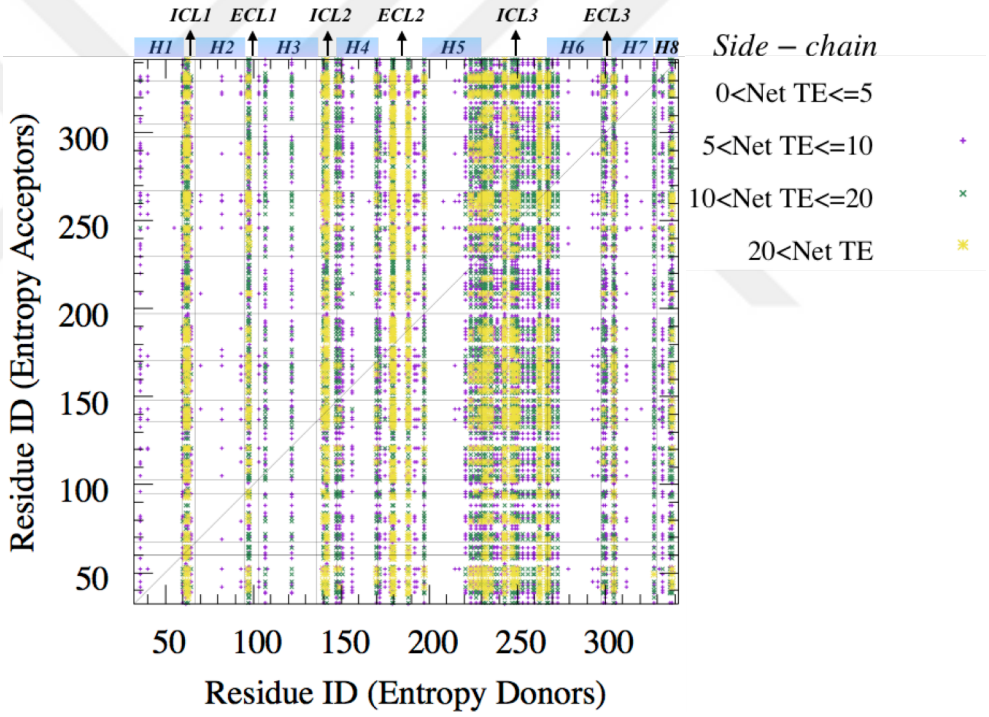
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Backbone + χ_1
0 < NetTE <= 1
1 < NetTE <= 2
2 < NetTE <= 5
5 < NetTE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)
50
100
150
200
250
300

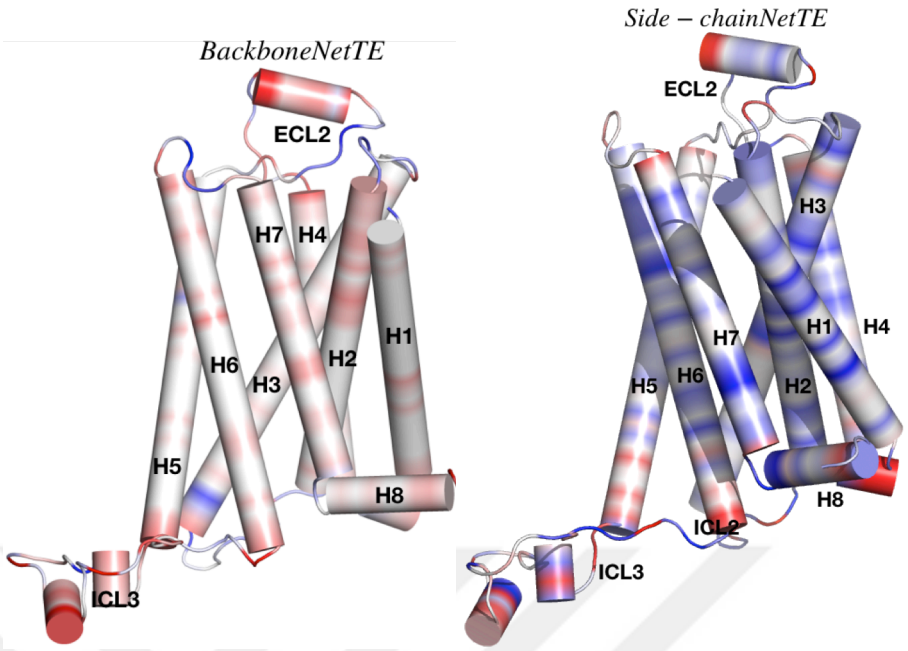
(b)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
BB
$BB + \chi_1$
BB
CA
BB+X1
SC
PhaseI
Mean Net TE
Residue ID

(c)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Side − chain
0<Net TE<=5
5<Net TE<=10
10<Net TE<=20
20<Net TE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)

(d)



(e)

Figure 6. 11 Heatmap representations of the pairwise net TE from backbone dihedrals (a), backbone+ χ_1 dihedrals (b). Mean net TE values for Backbone (magenta), Backbone+1SC(blue), Cα (green) and SC (red) (c). Heatmap representation of pairwise net TE values for side-chain dihedrals (d). Protein representations of backbone-only (left) and SC-only (right) net TE values (e).

The net TE analysis of only side-chain dihedral changes specifically showed the loops as strong entropy donors (Figure 6.11d). Also, the value of net TE was significantly higher than the previous two analysis data, giving that the number of atoms and the number of visited states involved in TE calculation increases the entropy output. Additionally, since backbone dihedrals were omitted in side-chain only dihedral analysis, the impact of transmembranes was apparently missing; instead, the highest mobile segments of the receptor were observed to contribute to entropy transfer.

As explained in the Materials and Methods section in Chapter 3, since dimensionality will explode and the amount of data required for statistics becomes too large in transfer entropy formula of backbone+side-chain dihedrals, backbone and side-chain dihedral data were computed separately. The impact of side-chains on the backbone data was examined through the inclusion of χ_1 on the backbone dihedral analysis. The comparison of backbone only and side-chain only protein representation indicated that the donors, even though their net TE values were low, could also be found in

transmembranes in backbone-only net TE analysis; however, positive net TE values were specifically observed in loops and in residues adjacent to the loops in side-chain dihedral data. Thus, both the backbone and side-chain angular motions were believed to have a correspondence with all atoms, and it was clear that their motions were not random.

Additionally, even though there was no direct relationship between dihedral data and Cα fluctuations, some similarities were observed. Net TE of backbone dihedrals captured the area that was also observed in transfer entropy data of Cα fluctuations. The intracellular part of the receptor became strong entropy donor together with ECL2 based on the data originated from the backbone dihedrals. However, the quantity of net transfer entropy obtained from backbone dihedrals was significantly low compared to the data obtained from Cα fluctuations and side-chain dihedrals. As it was plotted in Figure 6.11c, the highest net TE value was around 0.5 when backbone dihedrals were analyzed. The side-chain transfer entropy results were also interesting. Especially intracellular loops, together with few residues on ECL2, had high entropy donor abilities with a net TE > 4.

6.1.4.2 Comparison of the entropy transfer profiles in the novel inactive conformational state of β_2AR

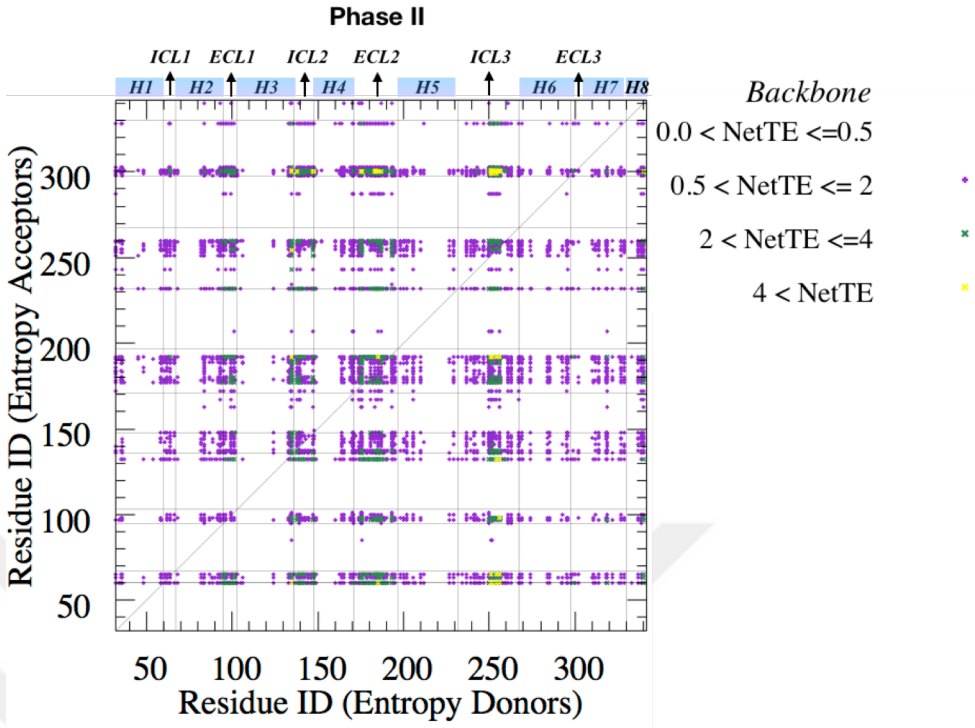
In Phase II, the entropy donors were shifted from ICL3 towards ICL2, ECL2, and intracellular parts of H3. Since the mobility of ICL3; thus, the overall flexibility of the receptor decreased upon its closure, the entropy transfer values of its dihedrals were decreased in parallel, except the middle part of the loop. Here in the highly inactive state, entropy acceptors were mostly located in ICL1, $H3_{Int}$, ECL2, and ECL3 (Figures 6.12a and 6.12b).

Transfer entropy from backbone and backbone+ χ_1 dihedrals gave interesting data regarding entropy donors and acceptors, where each loop acted as a donor and also as acceptor (Figure 6.12c). For example, ECL2 was observed as a strong donor; however, it was also an acceptor. This dual character of the domains implied that the type of

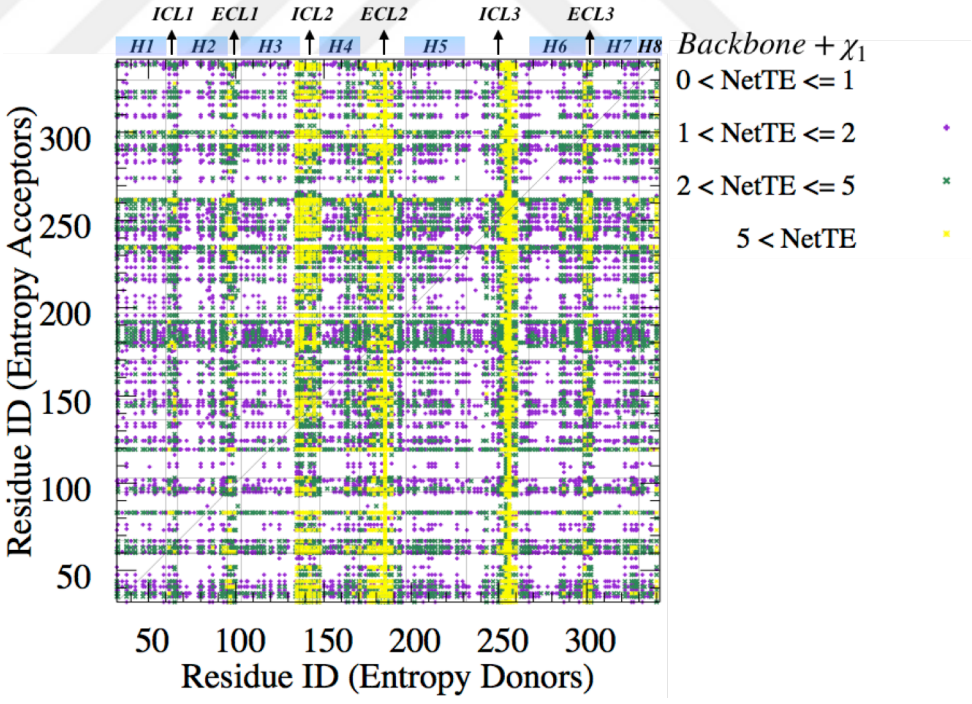
amino acids that loops contain influenced the entropy transfer based on side-chain dihedrals. The analysis of amino acid types will be performed in the upcoming sections.

Additionally, the protein representations of the transfer entropies displayed the increase in net TE of ICL2 upon closure of ICL3 (Figures 6.12e). In net transfer entropy data of side-chain dihedrals, however, a significant decrease in the net transfer entropy value of ICL3 was observed. Comparison of side-chain transfer entropy values with the transfer entropy values of Cα fluctuations resulted in a divergence in donor and acceptor domains. As shown in Chapter 5, residues in $H1_{Ext}$, ECL2, $H5_{Ext}$, $H6_{Ext}$, and $H7_{Ext}$ acted as entropy donors based on the fluctuations of Cα atoms. However, there was no distinct separation for the protein domains based on transfer entropy data of backbone + χ_1 dihedrals. This observation pinpointed the possible effect of the steric hindrance of dihedral angles on the information transfer. Cα atom of a residue can fluctuate, but its dihedral angles are not always capable of rotating freely. Especially in transmembranes, a correlated movement, including backbone and side-chain atoms, might not be possible due to the lack of space for side-chain atoms and the presence of bonded and non-bonded interactions between residues.

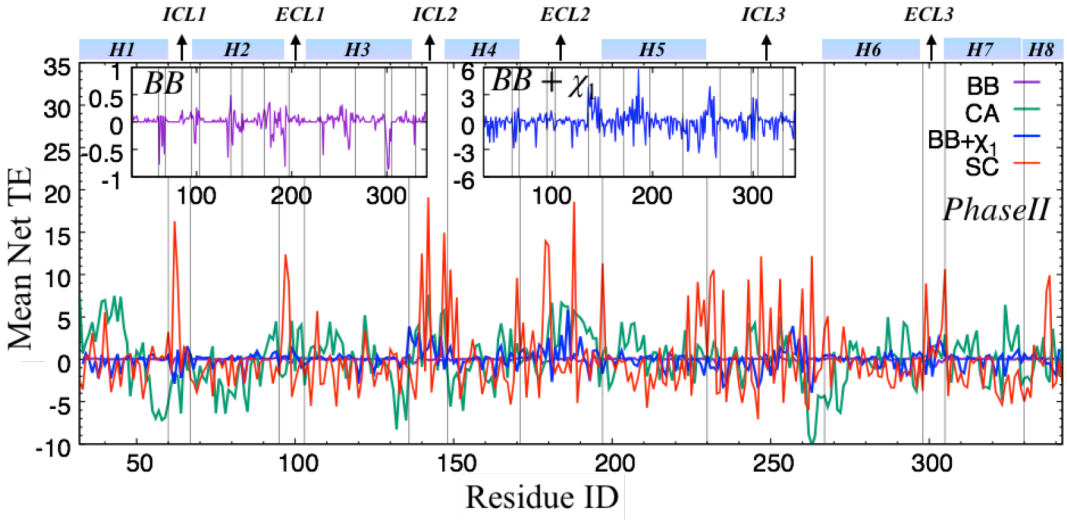
Net TE values of backbone dihedrals were lower than backbone + χ_1 dihedrals, and they were both lower compared to the net TE values from side-chain dihedrals. Interestingly, the net TE values of Cα fluctuations were also lower than the side-chain data. Both the number of atoms and the number of bins involved in the calculations can explain this entropy-based finding. The number of bins for Cα fluctuations was optimized for each residue based on the decrease of its Shannon entropy below 0.02. The bin numbers were obtained as above 19 for Phase II. However, when backbone dihedrals were considered, the maximum state that can be reached was $3x3 = 9$; for backbone+χ_1, it was $3x3x3 = 27$, and finally, for side-chains, it was $3x3x3x3 = 81$. Thus, the higher the number of states visited, the higher the entropy value that a residue may have.


Phase II
ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Backbone
0.0 < NetTE <=0.5
0.5 < NetTE <= 2
2 < NetTE <=4
4 < NetTE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)
50
100
150
200
250
300

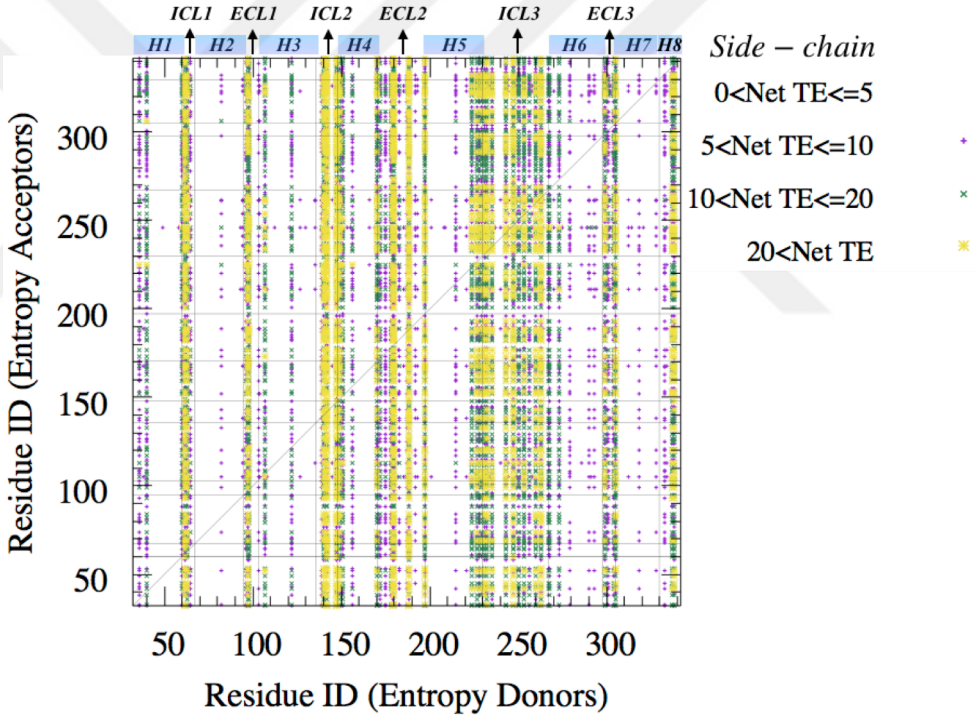
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Backbone + χ_1
0 < NetTE <= 1
1 < NetTE <= 2
2 < NetTE <= 5
5 < NetTE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)
50
100
150
200
250
300

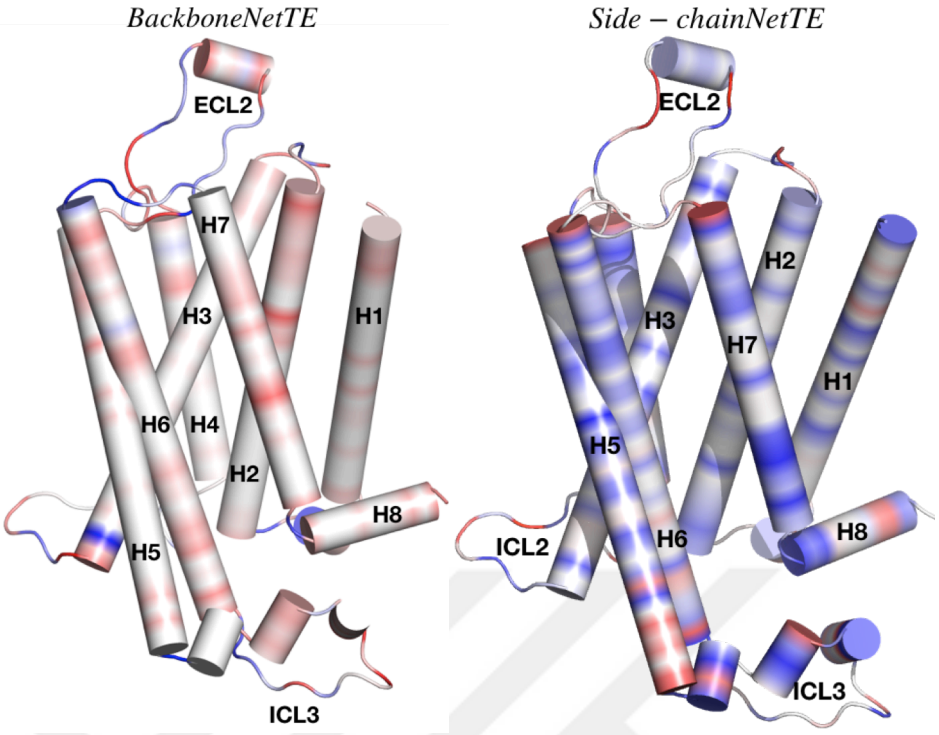
(b)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
BB
$BB + \chi_1$
BB
CA
BB+X1
SC
PhaseII
Mean Net TE
Residue ID

(c)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Side − chain
0<Net TE<=5
5<Net TE<=10
10<Net TE<=20
20<Net TE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)

(d)



(e)

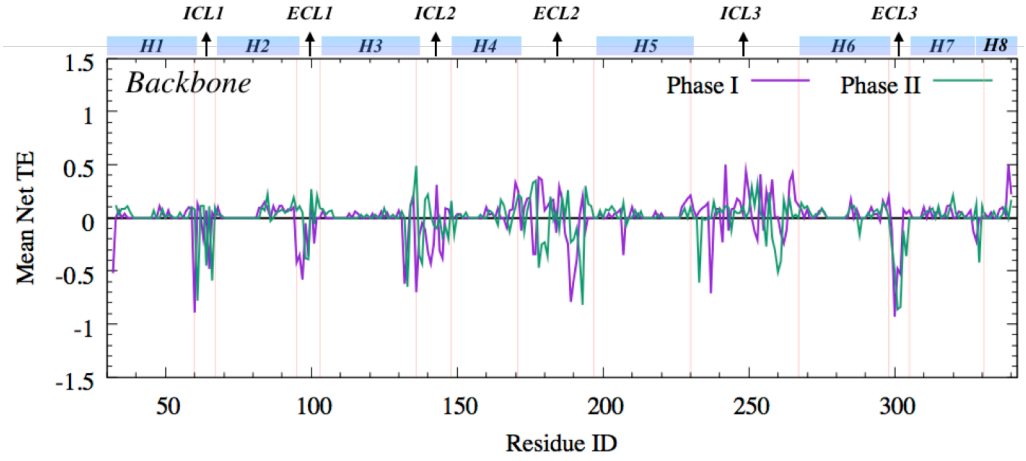
Figure 6. 12 Heatmap of the backbone (a) and backbone + χ_1 dihedrals' (b) net TE values. The abscissa represents donors and ordinate represent acceptors. From colors white to yellow, the intensity of the net TE value increase. The mean net TE values for backbone dihedrals (magenta), backbone + χ_1 dihedrals (blue), Cα fluctuations (green), and side-chain dihedrals (red) (c). The insets displayed the backbone and backbone + χ_1 net TE plots. Heatmap of side-chain dihedrals' net TE values (d). Protein representations of mean net TE of backbone dihedral (left) and side-chain dihedral (right) (e). Entropy donors were colored with red, and acceptors were colored in blue.

The comparison of entropy donors and acceptors due to backbone, backbone+ χ_1, and side-chain dihedral fluctuations between Phase I and Phase II was plotted in Figure 6.13. The net TE from backbone dihedrals of ICL3 was higher in Phase I compared to Phase II; however, including χ_1 increased the net TE values of ICL3 in Phase II in its middle segments.

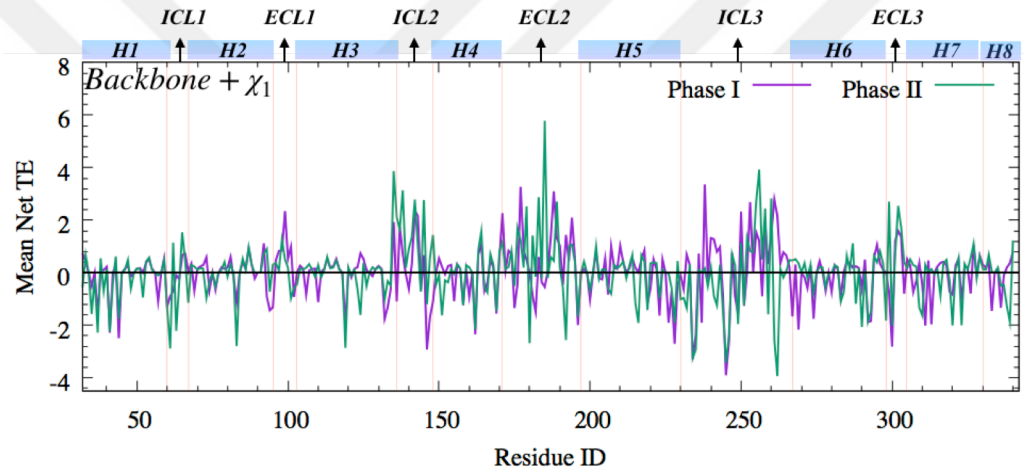
Moreover, the net TE of ICL2 and the middle part of ECL2 and $H6_{Int}$ became significantly higher in Phase II than Phase I. Transmembranes were observed as entropy acceptors in both phases, while loops were mostly entropy donors. Side-chain net TE values of ICL3 and $H7_{Int}$ differ in Phase I and Phase II, where the net TE values were higher in Phase I compared to Phase II. Thus, without the inclusion of side-chain to the calculations, transmembranes were either neutral or mild-entropy donors. When the

side-chains were included, they mostly became entropy acceptors, where some donor residues in the transmembranes existed with low net TE values.

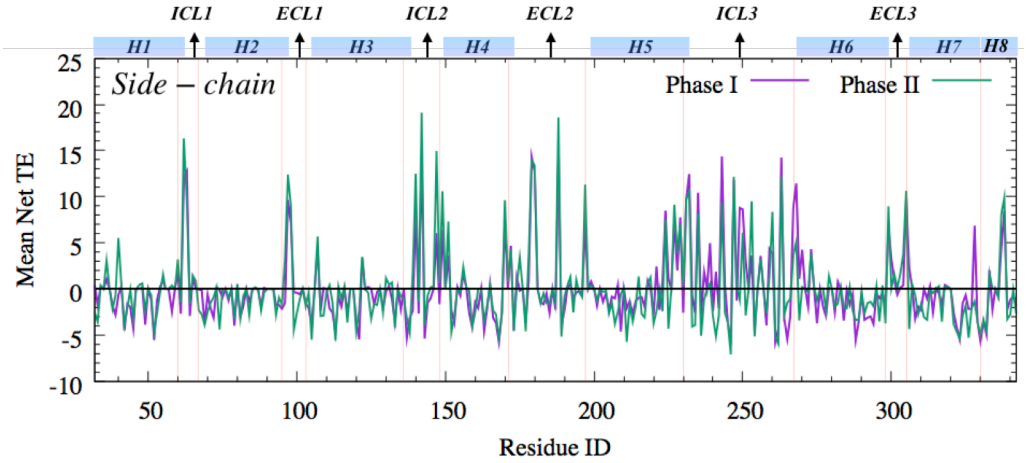
Figure 6.13 represented that side-chain net TE calculation revealed loops as entropy donors and transmembranes as entropy acceptors since side-chains of loops had high entropy to transfer. However, backbone dihedrals of loops indicate that this entropy mostly came from side-chains but not from backbones. Except for few residues on each loop, they behaved entropy acceptors based on backbone net TE computation. Inclusion of χ_1 to the calculation increased the values and changed the features of the residues, where transmembranes became acceptors and loops became more donors. Due to memory problems, the backbone+side-chain net TE could not be calculated. Still, Figure 6.13b and 6.13c guided as that side-chains had more entropy compared to backbone dihedrals, and this entropy was mostly stored in loops.



(a)



(b)



(c)

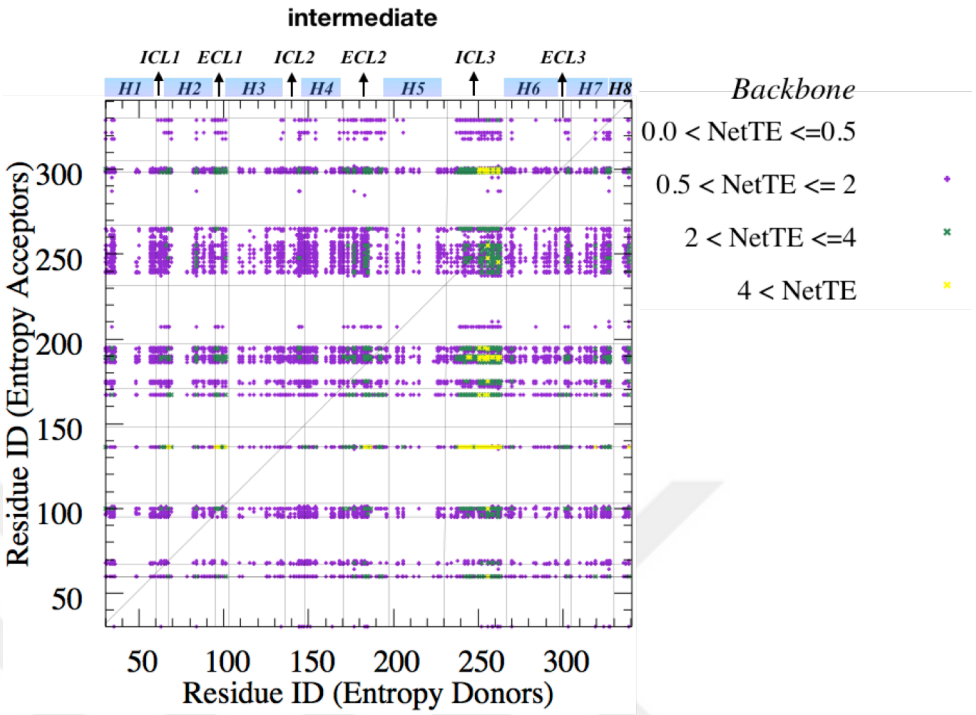
Figure 6. 13 Mean net TE for backbone (a), backbone + χ_1 (b) and side-chain (c) dihedrals for Phase I, and Phase II.

6.1.4.3 Comparison of the entropy transfer profiles in the *intermediate* conformational state of β_2AR

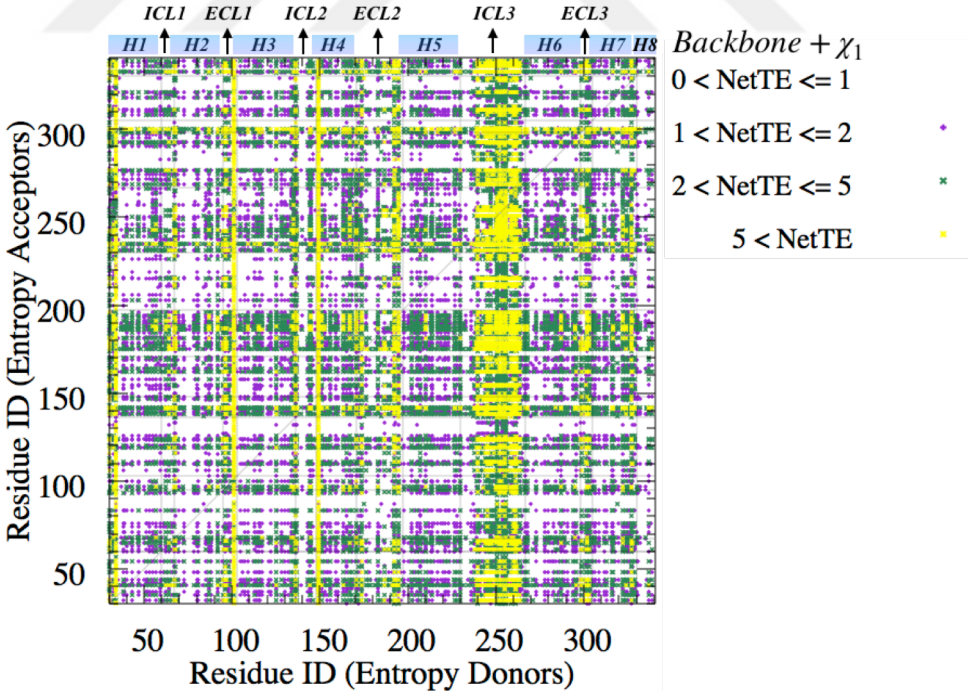
Going from inactive to active phases, an increase in net TE of dihedrals was observed. Precise TE from ICL3 to intracellular parts of H3, ECL2, and ECL3 was noticeable in backbones of the *intermediate* state (Figures 6.14a) similar to the data obtained from Cα fluctuations (Figure 5.16a). The dual characteristics of protein domains were also observed in the *intermediate* state, where the residues on ECL2 adjacent to H4 acted as an entropy source; however, those adjacent to H5 acted as entropy sinks. A similar pattern was also observed in ICL3. Residues adjacent to H5 acted as entropy acceptors, while the residues adjacent to H6 acted as entropy donors. However, ECL2 was observed purely as an entropy acceptor in Cα data and ICL3 as an entropy donor. Moreover, ICL2 was also observed as a donor in Chapter 5, but it was an acceptor in both backbone dihedrals and backbone+ χ_1 dihedral data, mostly due to its helical structure did not loosen throughout the trajectory, thus had a limited torsional fluctuation ability.

Backbone + χ_1 dihedral for *intermediate* state revealed that especially extracellular parts of the receptor donored information. Specifically, $H1_{Ext}$, ECL1, ECL2 adjacent to H5, and ECL3 were major entropy donors. ICL3 was also strongly donating information to the rest of the receptor. ICL2, ECL2 adjacent to H4, ICL3 adjacent to H5, $H6_{Ext,}$ and H8 were strong entropy acceptors. Thus, the inclusion of χ_1 to backbone dihedral angles changed the entropic features of the residues.

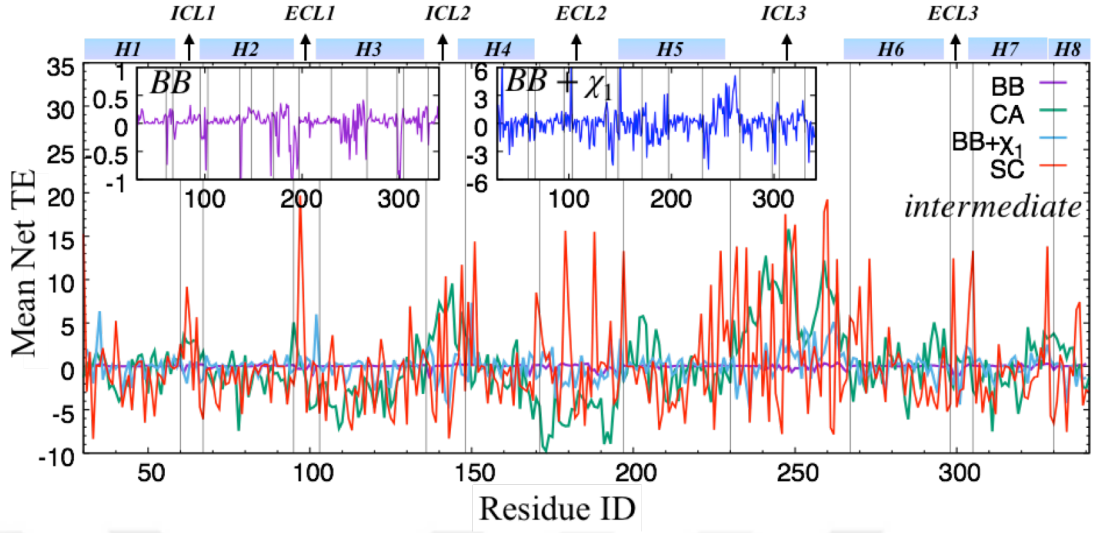
Only χ_1 angle may not be sufficient to explain all dynamics of the amino acids, since only *Val, Pro, Ser, Thr,* and *Cys* has χ_1 angle. Except for *Ala* and *Gly*, the remaining amino acids have $\chi_{2,}$ $\chi_{3,}$ and χ_4 angles. Therefore, the data from the χ_1 angle presented partial information about these amino acids. Side-chain dihedrals displayed how all-rotatable bonds of the residues transfer information (Figure 6.14d). Especially the intensity of positive net TE values was observed around ICL3. Unlike backbone dihedrals, where transmembranes showed little donor features, transmembranes exhibited acceptor characteristics except $H4_{Int}$, $H5_{Int}$, and $H6_{Int}$ (Figure 6.14e).


intermediate
ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Backbone
0.0 < NetTE <=0.5
0.5 < NetTE <= 2
2 < NetTE <=4
4 < NetTE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)
50
100
150
200
250
300

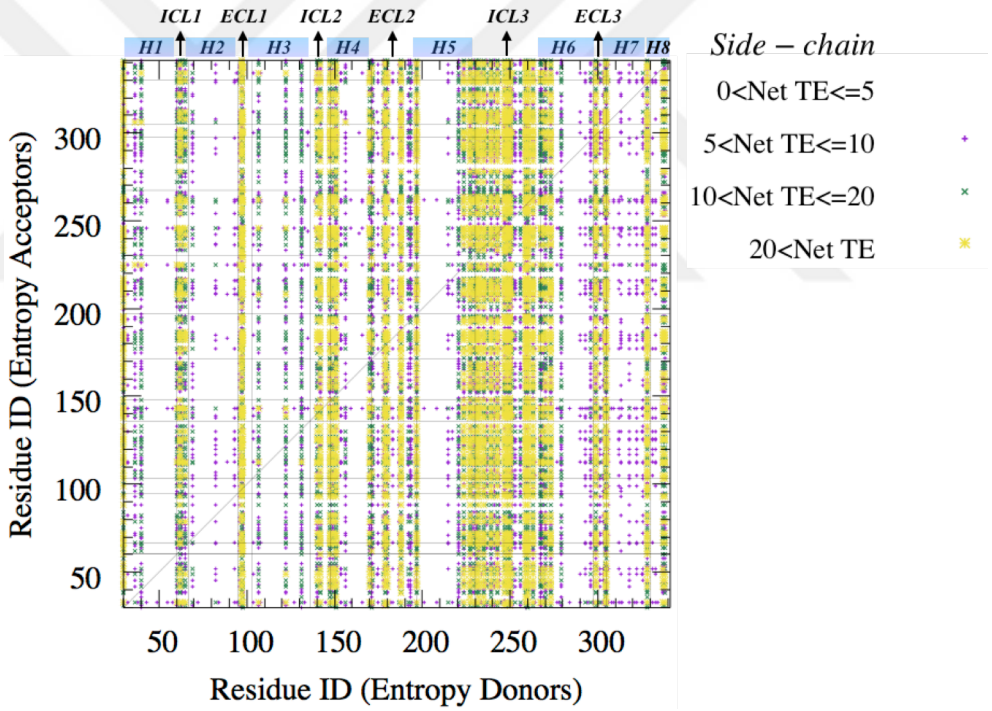
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Backbone + χ_1
0 < NetTE <= 1
1 < NetTE <= 2
2 < NetTE <= 5
5 < NetTE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)
50
100
150
200
250
300

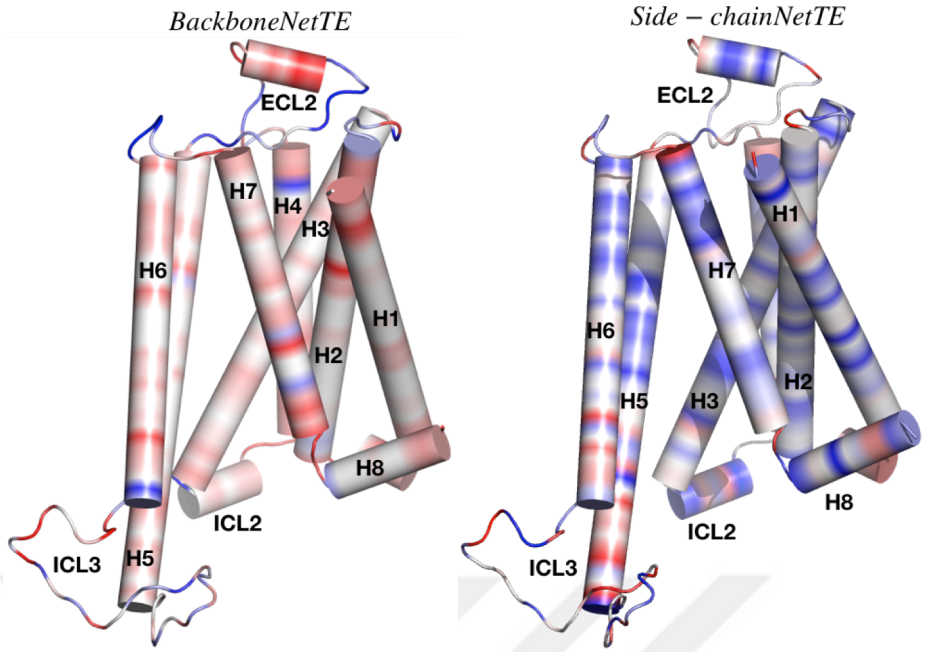
(b)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
BB
BB + χ1
BB
CA
BB+X1
SC
intermediate
Mean Net TE
Residue ID

(c)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Side − chain
0<Net TE<=5
5<Net TE<=10
10<Net TE<=20
20<Net TE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)

(d)

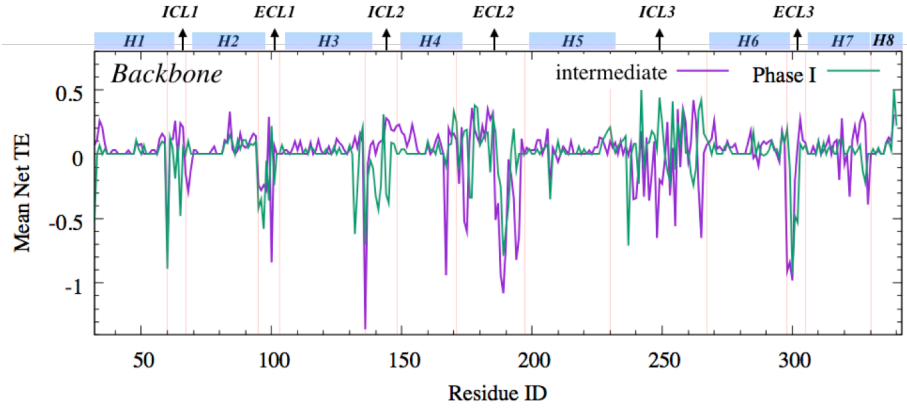


(e)

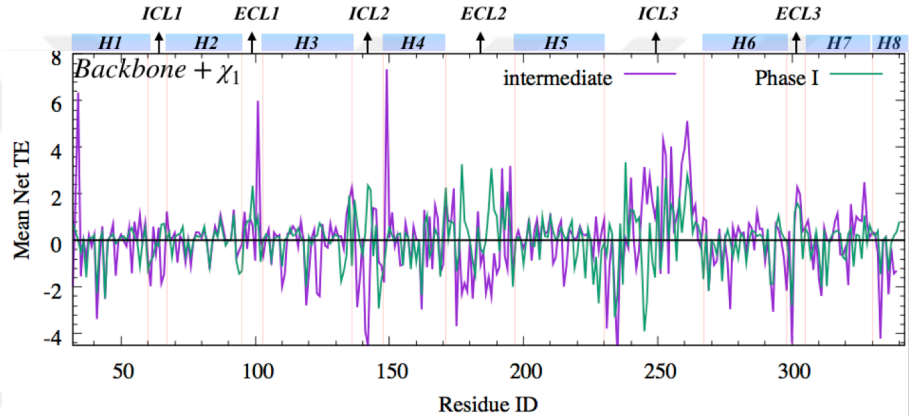
Figure 6. 14 Heatmap representation of *intermediate* backbone net TE values (a) and backbone + χ_1 net TE values (b). Mean net TE inset plot for backbone (magenta), backbone + χ_1 (blue), Cα fluctuations (green) and side-chain (red) (c). Heatmap for net TE values of side-chain dihedrals (d). Protein representation for mean backbone (left) and side-chain (right) mean net TE values (e).

As was shown in Chapter 5, the *intermediate* state resembled Phase I based on the information gathered from Cα fluctuations. Figure 6.15 presented that, in comparison to the *intermediate* state with Phase I, we observed differences in net TE profiles based on backbone dihedrals. These differences between phases might be originated from the variation of the dihedral flexibility data of the corresponding domains. The *intermediate* state displayed small fluctuations in loops except for ICL3, while ICL3 tended to pack itself through the core of the receptor; however, it stayed in *intermediate* conformation since $H6_{Int}$ did not orient itself vertically.

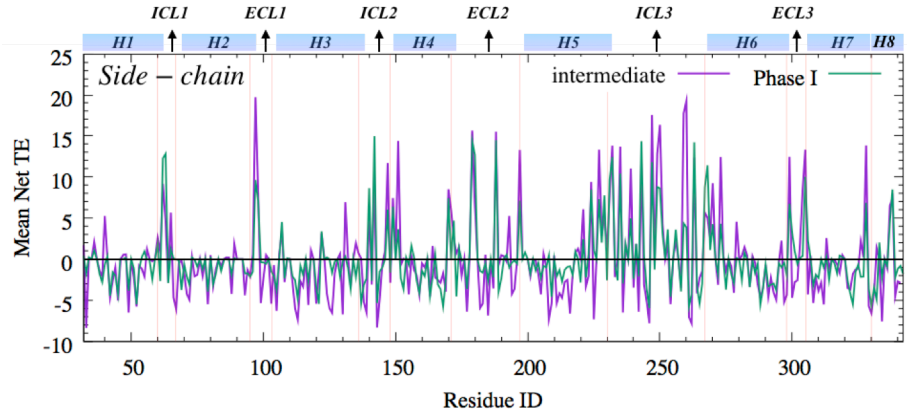
Additionally, backbone + χ_1 net TE comparisons clearly represented an increase in the net TE values of the *intermediate* state (6.15b). *Gly*35 in H1, *Gly*102 in ECL1, and *Ala*150 in H4 were the highest donors, followed by ECL2 and ICL3. It was interesting to observe the highest donors were the residues, which do not have χ_1 angle. However, it was not the case for Phase I. *Arg*239, *His*178, and *Thr*189 were the highest entropy donors in Phase I.



(a)



(b)



(c)

Figure 6. 15 Average entropy transfer in *intermediate* state (magenta) versus Phase I (green) in backbone (a), backbone + χ_1 (b), and side-chain (c) dihedrals.

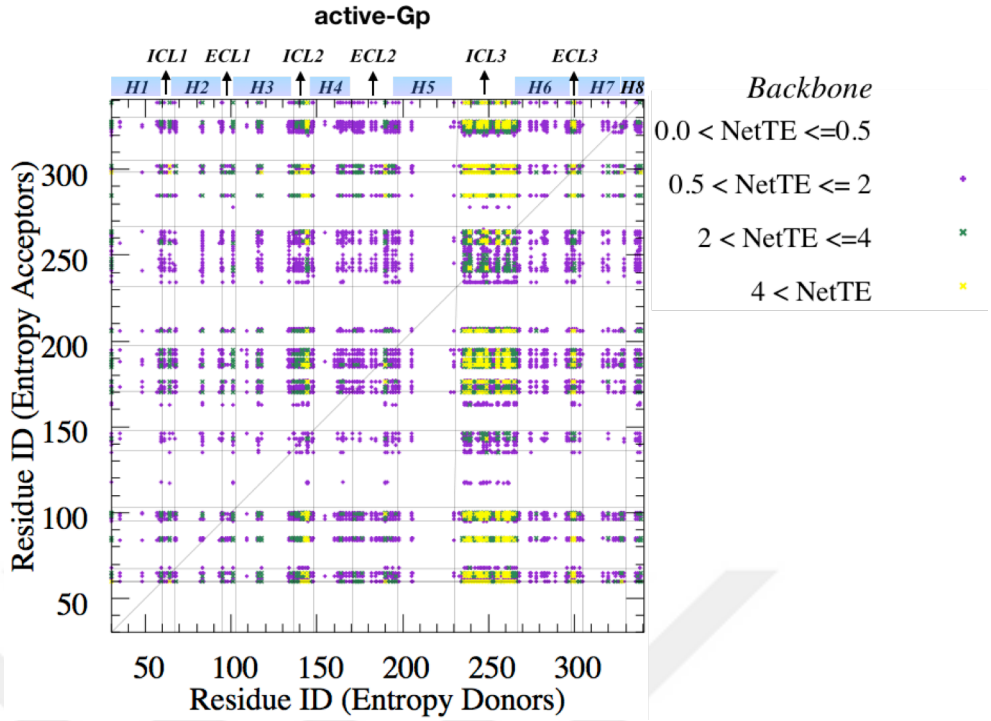
Net TE comparison for side-chain dihedrals also represented higher donor and acceptor characteristics in the *intermediate* state compared to Phase I. Figure 6.15c represented the comparison of side-chain dihedrals between the *intermediate* state and Phase I. The

value of net TE was generally higher in *intermediate* state, but the overall transfer entropy pattern of side-chain dihedrals was significantly similar to Phase I.

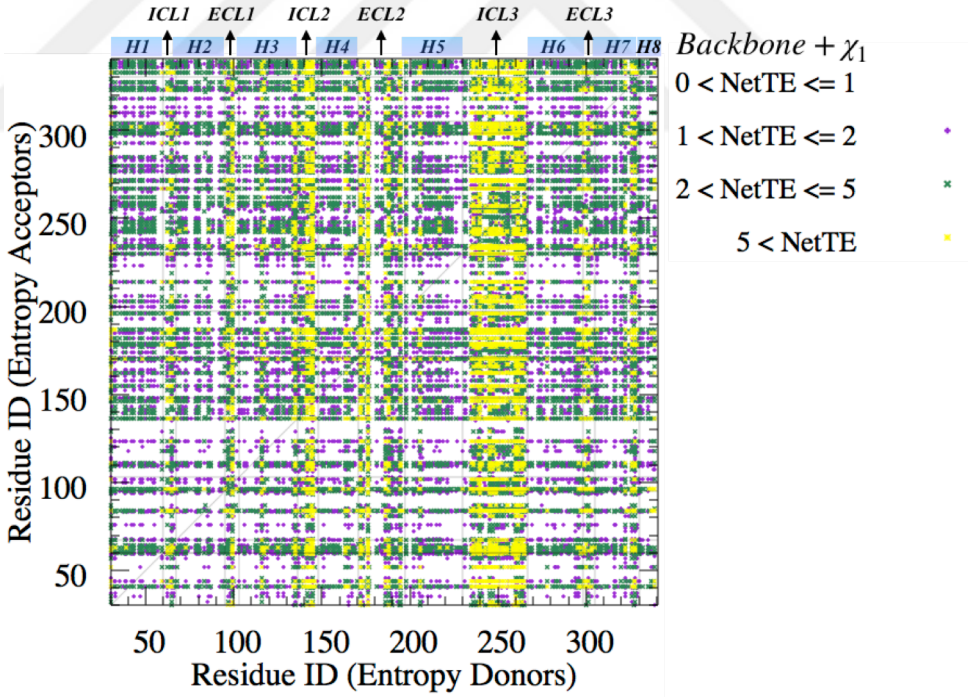
6.1.4.4 Comparison of the entropy transfer profiles in the active conformational state of β_2AR

One of the biggest entropy changes was observed in the *active-Gp* in terms of backbone dihedrals. ICL2 and ICL3 were observed as the highest entropy donors, while remaining loops were entropy acceptors (Figure 6.16a). Interestingly, the contribution of the transmembranes was observed as H2, H6, and H7. The dual characteristics of the entropy source and sinks were also observed in here in backbone analysis, where ICL3 was observed both as entropy donor and acceptor. Among analyzed five states, the highest donor capacity of ICL3 was observed in *active-Gp* state, when backbone dihedrals were considered. However, the inclusion of χ_1 did not change the profile, where ICL2 and ICL3 were donors together with ECL2 adjacent to H4. Side-chain dihedrals in Figure 6.16d demonstrated the donors locating on ICL1, ICL2, and ICL3. There were three residues in ECL2 and 1 residue in ECL3 showing donor features, however, the remaining parts of these loops were acceptors. The comparison of dihedral data with Cα data gave complementary data to explain the overall dynamics of the receptor. Net TE values from dihedrals mostly not overlapped with the Cα data, since Cα data was depending on absolute fluctuation from the average, while dihedral data was depending on direct angular data from MD trajectory. Based on Cα fluctuations, ICL3 and $H3_{Ext}$ were the entropy donors, and H1 was the strong entropy acceptor. However, we observed ICL3 as an entropy donor, but $H3_{Ext}$ did not donor, and H1 did not accept information based on dihedral data. Thus, it can be concluded that the combination of Cα fluctuation with dihedrals might answer the question of how global dynamics of the receptor occur. Therefore, the parameters need to be analyzed separately, and then their combinatorial effect needs to be understood.

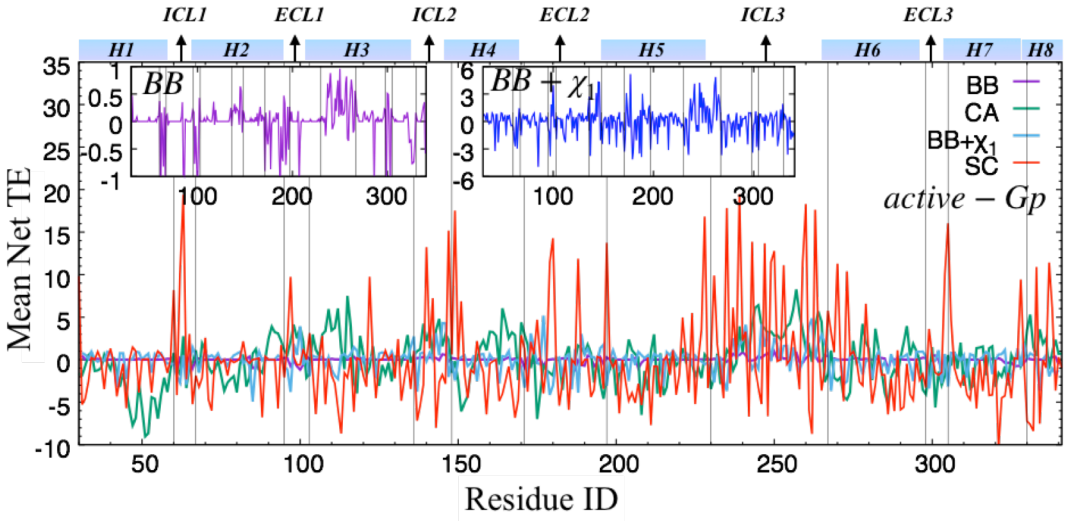
.


active-Gp
ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Backbone
0.0 < NetTE <=0.5
0.5 < NetTE <= 2
2 < NetTE <=4
4 < NetTE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)
50
100
150
200
250
300

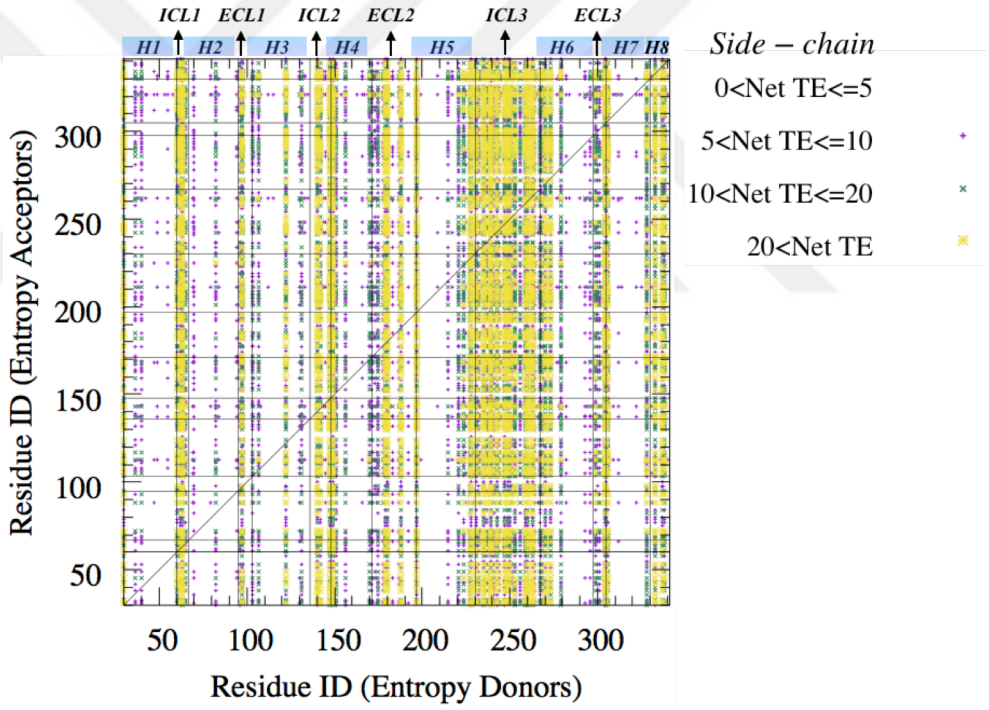
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Backbone + χ_1
0 < NetTE <= 1
1 < NetTE <= 2
2 < NetTE <= 5
5 < NetTE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)
50
100
150
200
250
300

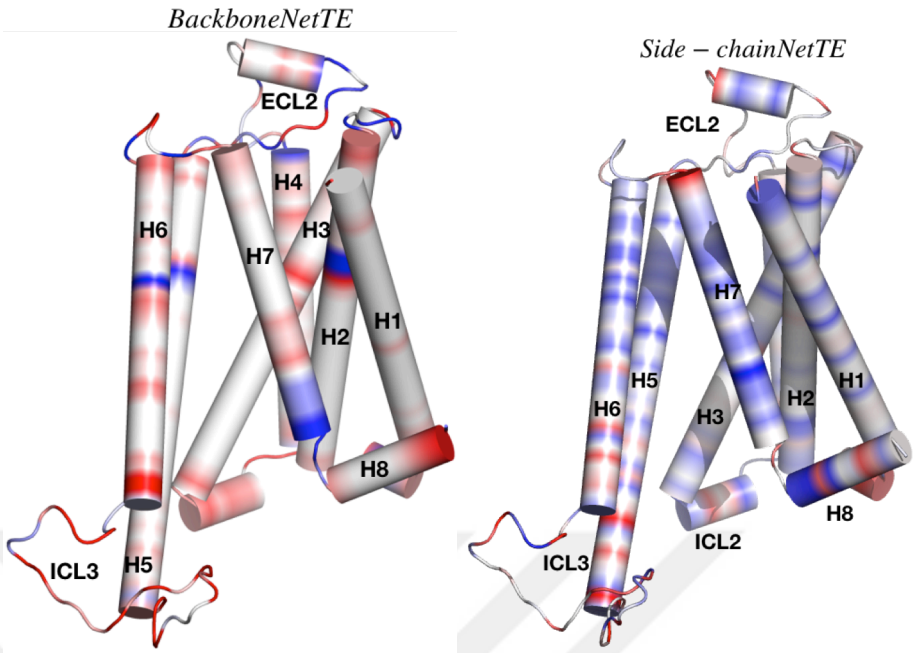
(b)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
BB
BB + χ1
BB
CA
BB+X1
SC
active − Gp
Mean Net TE
Residue ID

(c)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Side − chain
0<Net TE<=5
5<Net TE<=10
10<Net TE<=20
20<Net TE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)

(d)

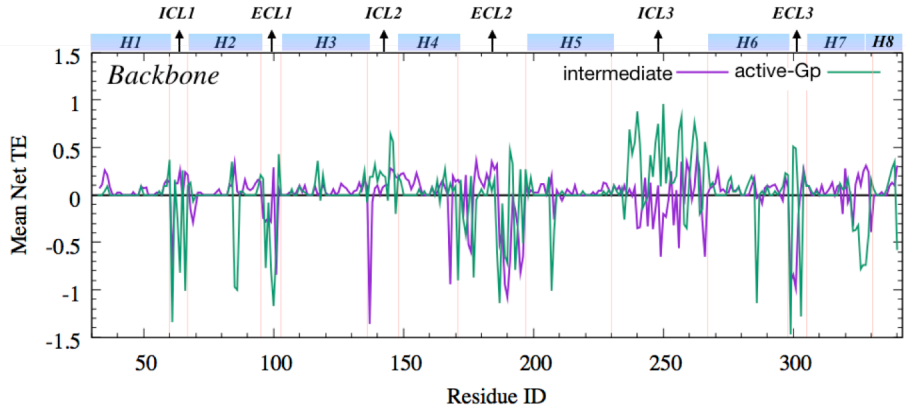


(e)

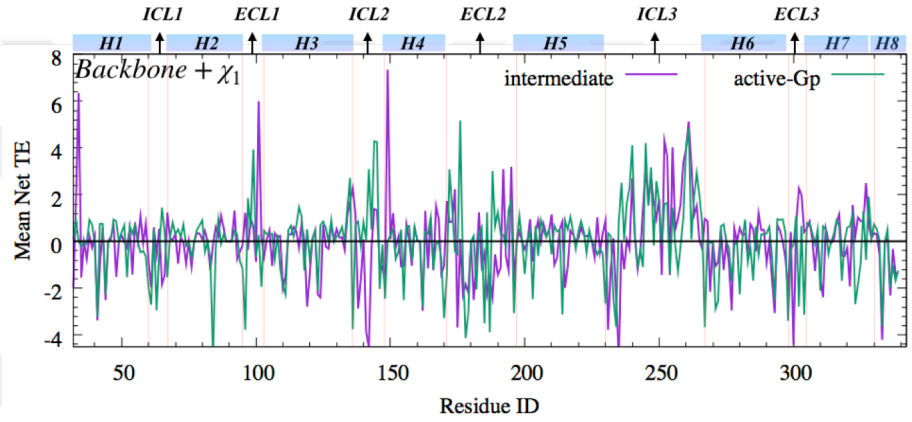
Figure 6. 16 Heatmap for net TE related to backbone dihedrals (a) and backbone + χ_1 dihedral (b) for *active-Gp*. Mean net TE values for backbone (magenta), backbone + χ_1 (blue), side-chain (red) dihedrals and Cα (green) fluctuation (c). Heatmap for side-chain dihedrals (d). Protein representation for mean backbone (left) and side-chain (right) net TE values (e).

The comparison of backbone dihedrals resulted that 1.1% of residue-pairs had net TE>4 in *active-Gp*. However, this value was 0.16% in the *intermediate* state, 0.09% in Phase II, and 0.04% in Phase I. Thus, ICL3 can be claimed as extremely significant in activity. The comparison of *intermediate* state with *active-Gp* displayed the role of dihedrals in ICL3. The backbone dihedral data showed that ICL2 and ICL3 were donors in *active-Gp*, while they were acceptors in the *intermediate* state. $H7_{Int}$ was observed as an acceptor in *active-Gp*, but it was a donor in the *intermediate* state (Figure 6.17a).

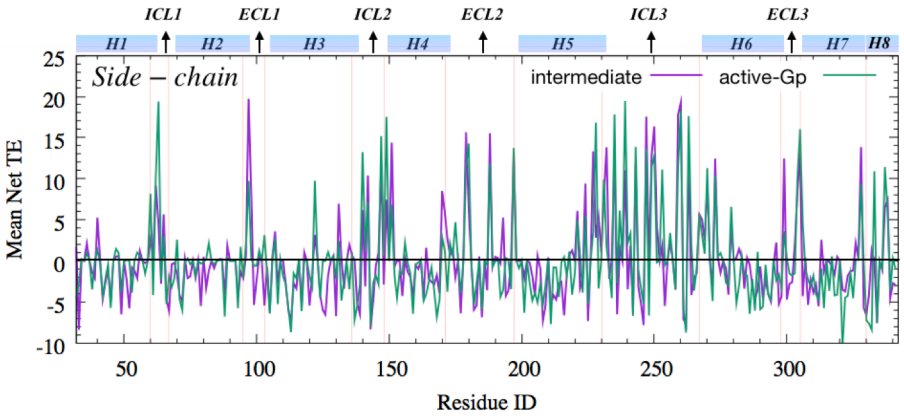
Inclusion of χ_1 changed the scenario, where loops, such as ECL1 and ICL2, as well as the extracellular part of H1, became strong donors in the *intermediate* state (Figure 6.17b). In *active-Gp*, the acceptor characteristics of transmembranes were higher compared to *intermediate*. Side-chain dihedrals, however, displayed similar profiles between these two states (Figure 6.17c).



(a)



(b)



(c)

Figure 6. 17 Mean net transfer entropy for backbone (a), backbone + χ_1 (b) and side-chain (c) dihedrals for *intermediate* state and *active-Gp*.

In the analysis of backbone dihedrals, the comparison of the net transfer entropy between *intermediate* state and *active-Gp* demonstrated the differences between ICL1,

ICL2, $H5_{Ext}$, ICL3, and NPxxY domain of the receptor (Figure 6.17). As it was stated above, these domains were entropy donors in *active-Gp*, while they were entropy acceptors in the *intermediate* state. On the other hand, ICL1 was entropy acceptor in *active-Gp*, whereas it appeared as entropy donor in the *intermediate* state.

In side-chain dihedrals; however, residues *Phe*193, *Ser*204, *Trp*286, *Phe*290, and *Tyr*316, which were key residues in orthosteric ligand binding, were entropy donors in the *intermediate* state but became entropy acceptors in *active-Gp*. For instance, the mean net TE value for *Phe*193 was 5.26 in the *intermediate* state; however, it was -3.55 in *active-Gp*.

In the analysis of active phase simulations, on the other hand, we observed more entropy donor role of the G protein binding site compared to the other loops. Even though all loops still had entropy source characteristics, $H5_{Int}$, ICL3, and $H6_{Int}$ dominated as entropy sources (Figure 6.17b). In *active-Gp*, also the donor characteristics of ICL1 was significantly increased. Transmembrane regions mostly acted as entropy sinks, expecting information from donor residues, generally from intracellular parts of the receptor.

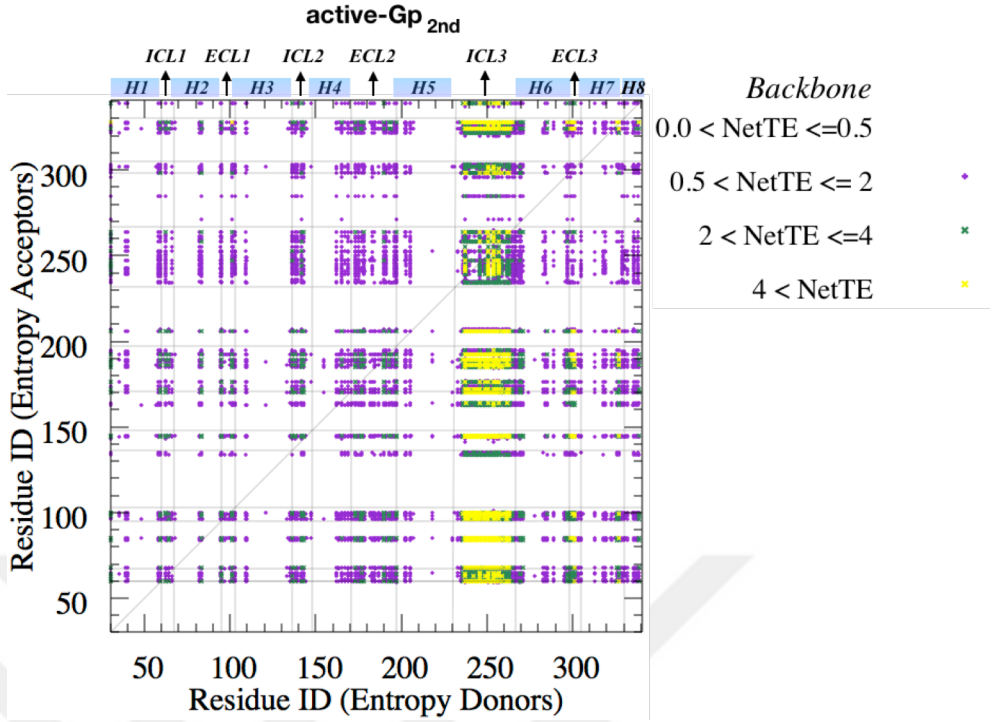
6.1.4.5 Comparison of the entropy transfer profiles in the second run of the active conformational state of β_2AR

Finally, *active-*Gp_{2nd} displayed the highest net TE data in all dihedral analysis. ICL3 was apparently the major entropy donor in backbone dihedrals (Figure 6.18a) together with ICL2. ECL1, the two ends of ECL2, ICL3, and $H7_{Int}$ became entropy donors upon inclusion of the χ_1 dihedral (Figure 6.18b). Interestingly, ICL2, which was entropy donor in *active-Gp*, was lost in *active-*$Gp_{2nd.}$ All side-chains, however, strongly displayed ICL3 and $H6_{Int}$ as donors.

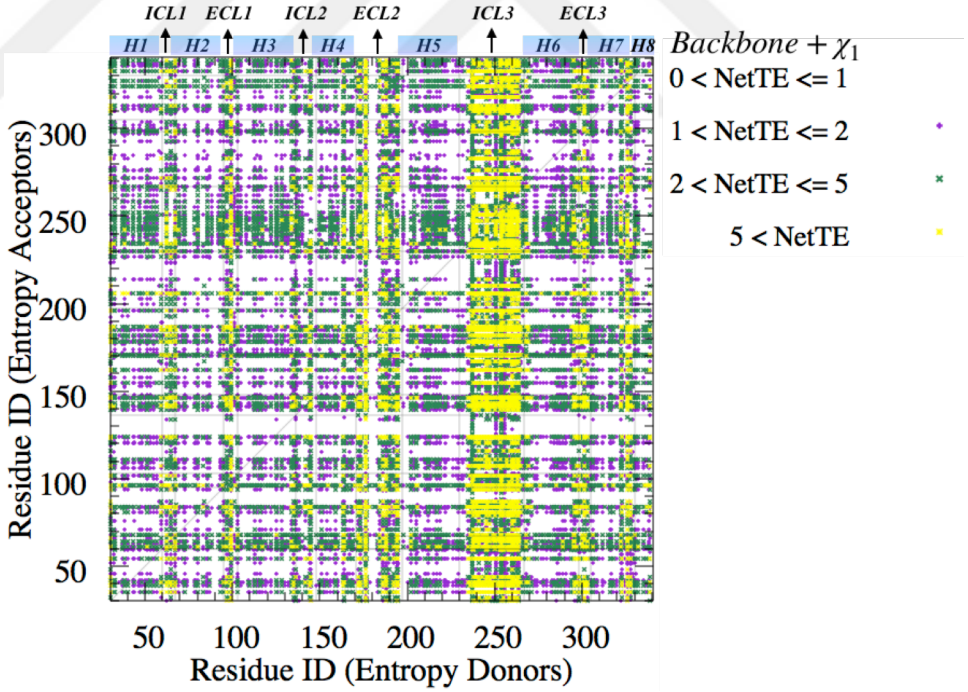
MD trajectories of *active-Gp* runs were also compared based on their dihedral entropy transfer characteristics. Side-chains of amino acids were more flexible compared to backbones due to the lack of rigid covalent bonds connecting these residues. Side-

chains were free to rotate until steric hindrance holds, but backbone atoms had a limitation due to ionic bonds. What we have compared in this part included the backbone and side-chain net transfer entropy values between two MD trajectories of G protein-bound active states. Both net TE results of backbone dihedrals and the results of side-chain dihedrals were highly similar except two ends of ICL2 became entropy acceptor in backbone dihedrals of *active-Gp_{2nd}*. The comparison of Cα data with dihedral data also revealed differences in net TE values. ICL3 and $H7_{Ext}$ were strong entropy donors in Cα fluctuations, but $H7_{Ext}$ was not visible as strong donors or acceptors in net TE values of dihedral data. Also, H8 appeared as acceptors in Cα fluctuations; however, it was observed as donors only in the side-chain dihedral analysis.

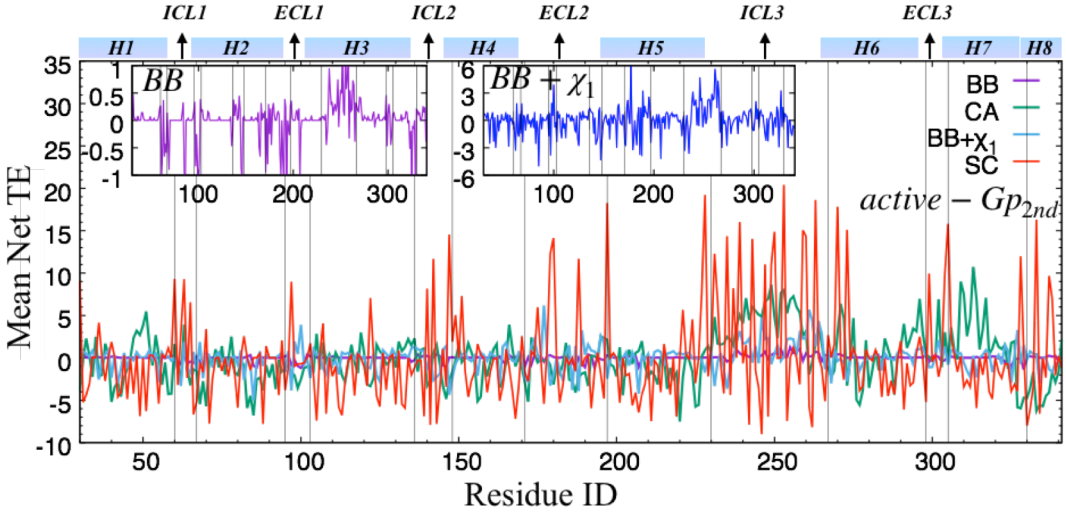
The mean net TE values of the backbone and side-chain dihedrals for these states were calculated and plotted in Figure 6.19. It was clearly visible that both backbone and side-chain entropy transfers' shapes were similar in both states, where positive and negative values were mostly overlapping.


active-Gp 2nd
ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Backbone
0.0 < NetTE <=0.5
0.5 < NetTE <= 2
2 < NetTE <=4
4 < NetTE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)
50
100
150
200
250
300

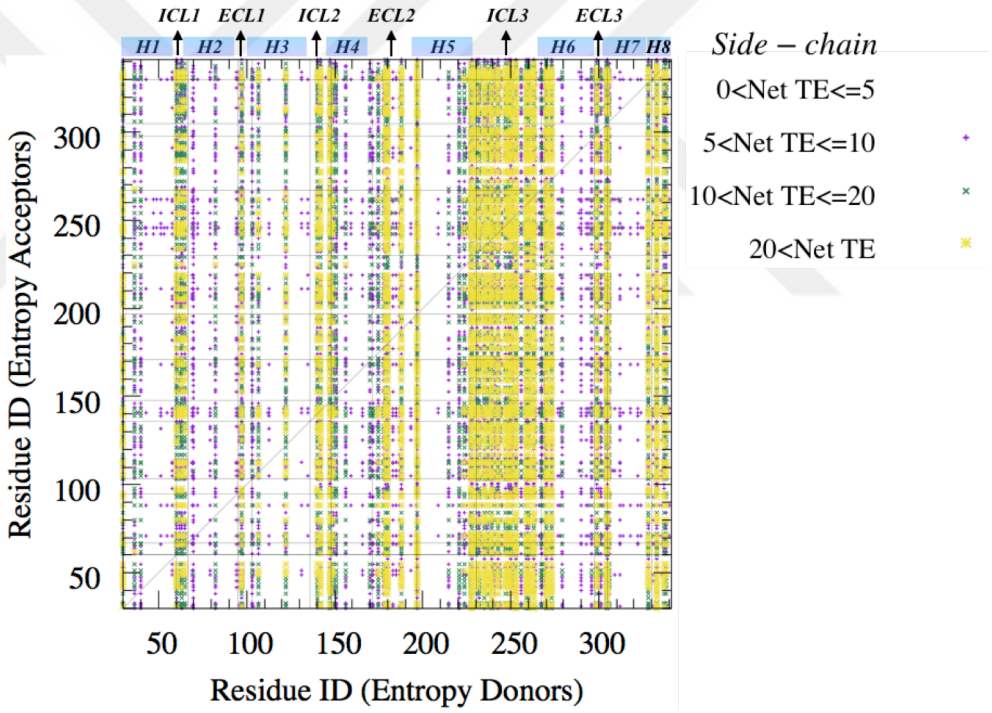
(a)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Backbone + χ_1
0 < NetTE <= 1
1 < NetTE <= 2
2 < NetTE <= 5
5 < NetTE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)
50
100
150
200
250
300

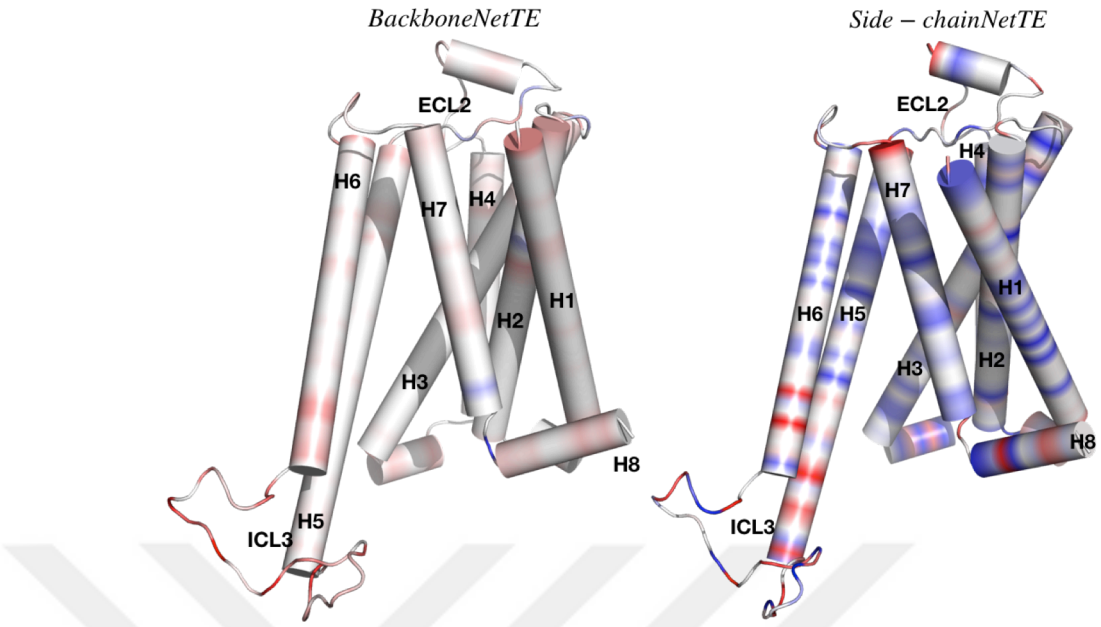
(b)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
BB
BB + χ1
BB
CA
BB+X1
SC
active − Gp2nd
Mean Net TE
Residue ID

(c)


ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
H1
H2
H3
H4
H5
H6
H7
H8
Side − chain
0<Net TE<=5
5<Net TE<=10
10<Net TE<=20
20<Net TE
Residue ID (Entropy Acceptors)
Residue ID (Entropy Donors)

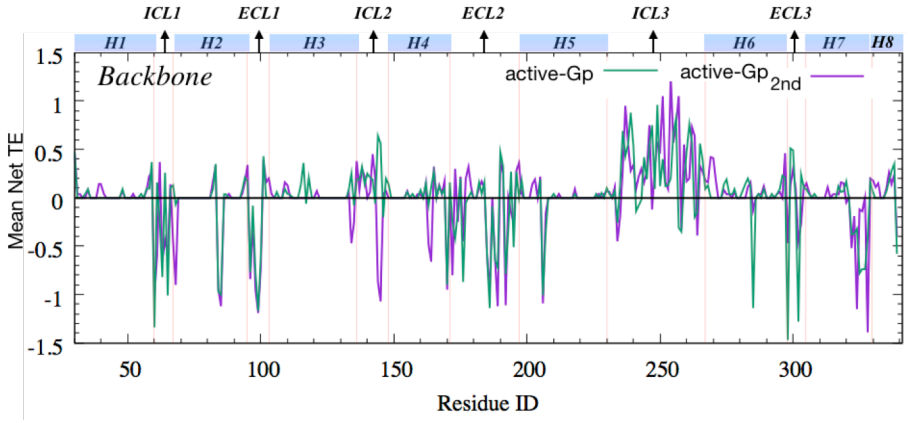
(d)



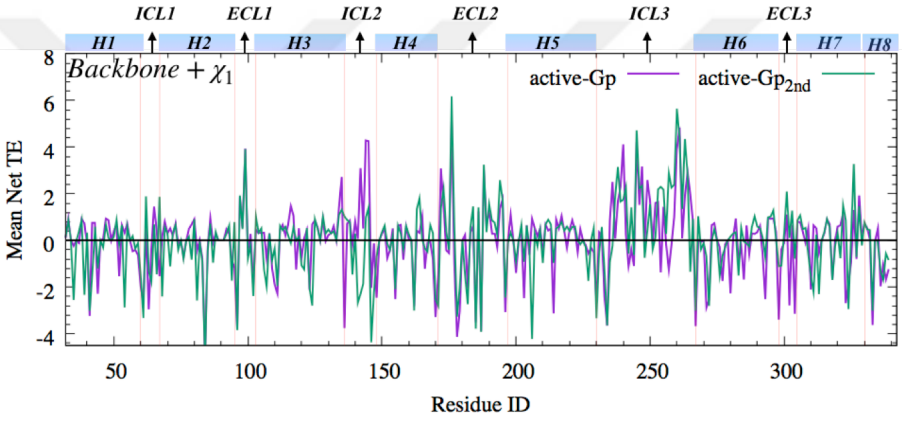
(e)

Figure 6. 18 Heatmap for net TE related to backbone dihedrals (a) and backbone + χ_1 dihedral (b) for *active-Gp*$_{2nd}$. Mean net TE values for backbone (magenta), backbone + χ_1 (blue), side-chain (red) dihedrals and Cα (green) fluctuation (c). Heatmaps for side-chain dihedrals (d). Protein representations for mean backbone (left) and side-chain (right) net TE values (e).

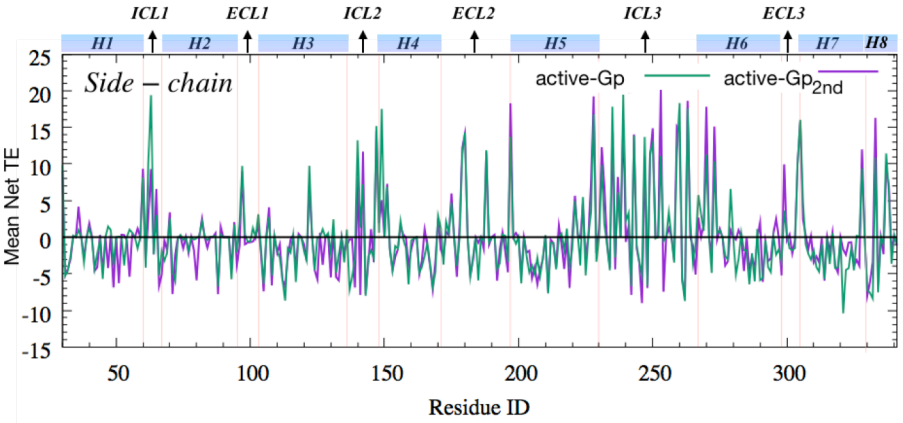
The green color in Figure 6.19 was from the *active-Gp,* and the magenta color was from *active-Gp*$_{2nd}$. Mean net TE values of side-chain dihedrals were significantly higher than the net TE values of backbone dihedrals. The comparison of these two states indicated that, except for ICL2 adjacent to H4, no significant differences were observed in mean Net TE values of backbone dihedrals. Loop and flexible regions were entropy donors. Especially intracellular regions, such as ICL1, ICL2, ICL3, and H8 had a higher number of donors. Besides them, higher donor characteristics of ECL2 and ECL3 were observed compared to ECL1 and transmembrane domains. As in the deviation and fluctuation data given in Chapter 5, *active-Gp* was relatively stable, but ICL3 in the *active-Gp*$_{2nd}$ trajectory was highly flexible. Thus, even though trajectories of the active states were different, both dihedral mutual information and transfer entropy data gave similar results. This similarity was the indication of dihedral analysis was able to capture and differentiate the active conformation from other conformations.



(a)



(b)



(c)

Figure 6. 19 Backbone (a) backbone + χ_1 (b), and side-chain (c) mean Net TE plots for two MD simulations of the active state. *active-Gp* was plotted with a green line, and *active-Gp*$_{2nd}$ was colored with magenta.

Allosteric communications, where information is conveyed from one site of the protein to another distal site, are affected by intra-protein communications. In Chapter 5, pairwise communications and information transfer analysis was performed on the absolute fluctuations of backbone Cα atoms from their average positions. Up to date, entropy-based studies to explain allostery were mostly performed on the Cα atoms, where these backbone motions have large structural changes compared to dihedral motions, which have subtle contributions to the protein dynamics.

Here in Chapter 6, we considered these small contributions to the intrinsic dynamics of the protein using five different MD trajectories, in order to understand the underlying mechanisms affecting the conformational differences. Both backbone and side-chain dihedrals behaved differently than Cα data since dihedral data was obtained from the angular changes of each torsional angle, while Cα data was obtained from absolute fluctuations of the backbone Cα atoms from their average positions. Still, dihedral data revealed the long-range allosteric interactions, especially between loop regions. The amount of dihedral flexibility was directly reflected in the MI results, where Phase II and *intermediate* state had the lowest information sharing among five states.

The net transfer entropy of dihedral angles, on the other hand, resulted in the strongest donor characteristics of ICL3 in backbone dihedrals of active states, the significant effect of the χ_1 inclusion both on the values and on the information transfer features of the residues, and finally, as it will be explained in the upcoming sections, the roles of residue types on the information transfer in side-chain dihedrals. It was reported that the correlations between the B-factor of the residues and backbone dihedrals were at most 0.29 (Zhang *et al*., 2010). Moreover, Dubay and his coworkers fixed the backbone of Calmodulin in order to study the allosteric contributions of side-chain dihedrals and obtained their correlated motions (DuBay *et al*., 2011). Even in the fixed backbone, the presence of correlated side-chain motions indicated the presence of interactions between residue-pairs. Thus, observing dissimilarity between dihedral data and Cα fluctuations is expected and their combinatorial effect generates the specific protein dynamics and conformations.

6.1.4.6 The analysis of residue type versus conformational state in transfer entropy data of backbone dihedrals

The contribution of amino acid type to the entropy transfer of backbone dihedrals was analyzed for five states as Phase I, Phase II, *intermediate* state, *active-Gp*, and *active-*Gp_{2nd}. Average Net TE per residue type calculations represented that the highest net TE value belongs to *active-*Gp_{2nd}, followed by *active-Gp* and *intermediate* state. In Figure 6.20, the net TE values of amino acid pairs were plotted. Hydrophobic amino acids were colored with blue colors and polar amino acids with red colors. From left to right, the number of rotatable bonds increases in both hydrophobic and polar groups.

The differences in these pairwise donor-acceptor relations might affect the conformational changes in the receptor. Therefore, focusing on the differences in donors and acceptors will help us to understand the impact of amino acid type on the particular conformation of the receptor.

In Phase I, net TE values of *Glu→Thr* pairs was the highest where *Glu* donors entropy to *Thr*. It shifted to *Gly→Asp* pairs in Phase II. In the *intermediate* state, the highest net TE was observed in *Gly→Pro* and *Arg→Pro* pairs; however, in *active-Gp*, it was observed in *His→Trp* pairs. Finally, *active-*Gp_{2nd} represented *Gly→Thr* and *His→Thr* pairs. Strong entropy donors in *active-Gp* and Phase I was observed as polar amino acids. Interestingly, both polar and hydrophobic amino acids were observed as donor in the *intermediate* state, and finally, polar and hydrophobic residıes, such as *His, Gln,* and *Gly* were observed as entropy donors in active states.

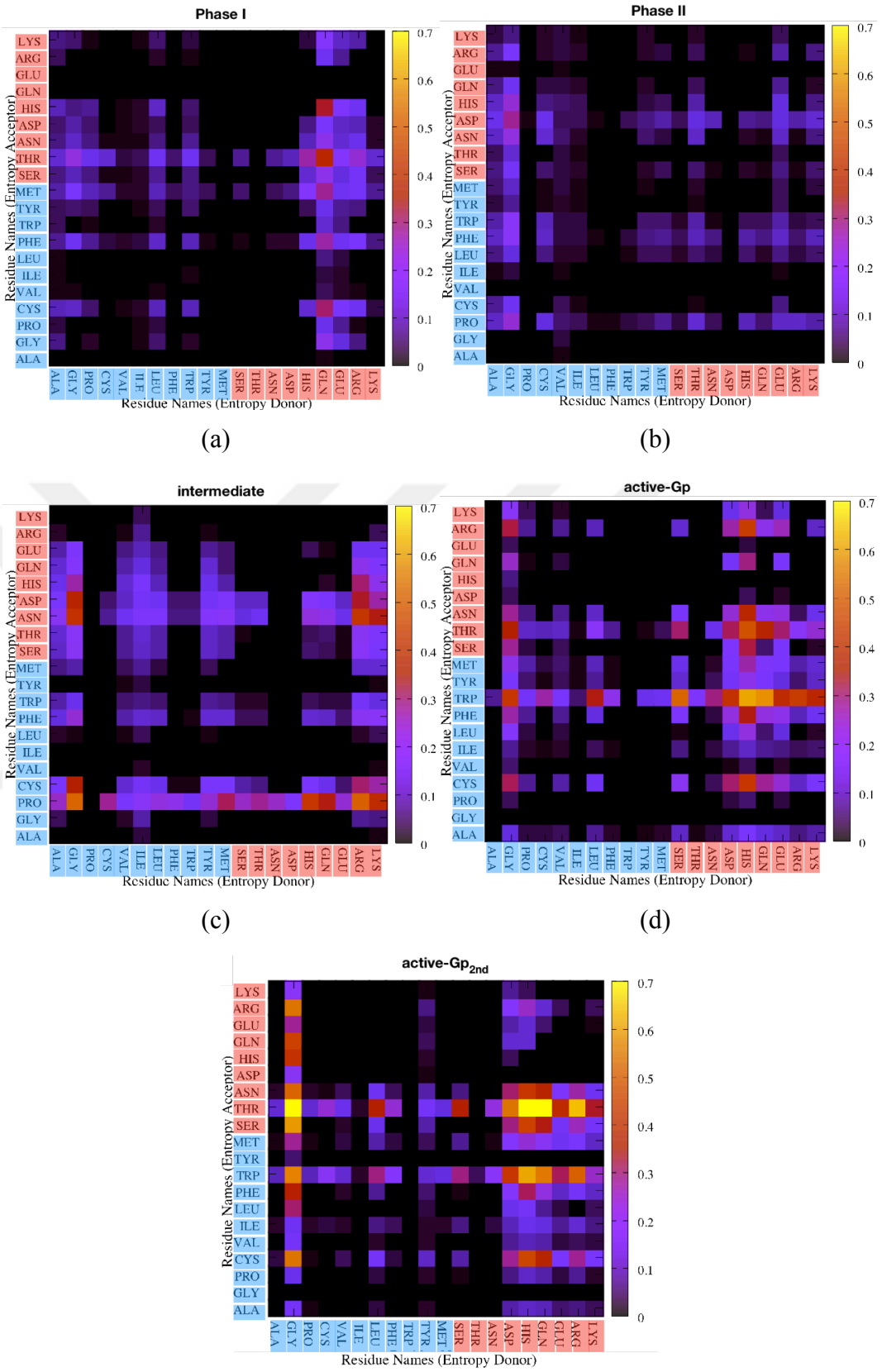


Figure 6. 20 Pairwise net TE values of backbone dihedrals based on residue types in Phase I (a), Phase II (b), *intermediate* state (c), *active-Gp* (d) and *active-Gp*$_{2nd}$ (e)

Increased steric hindrance might prevent large amino acids from changing their dihedral angles. Thus, hydrophobic amino acids were strong entropy donors in Phase II. The lowest net TE value was observed in Phase II, and it was observed that *Gly* had the highest donor characteristics among all residue types in this phase. The decrease in the flexibility of the structure; thus, the effect dihedral angles were determined. *Gly* locates on ICL3, $H7_{Int}$, $H1_{Int}$, $H1_{Ext}$, $H2_{Mid}$, $H2_{Ext}$, $H3_{Ext,}$ and $H4_{Mid}$, where high positive net TE values were observed.

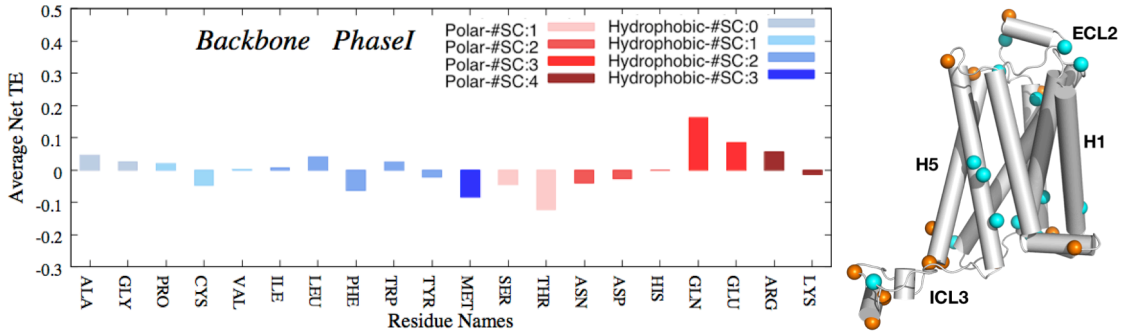
Interestingly, *Gly* and *Pro* residues were commonly found in TMs (Eilers *et al*., 2002; von Heijne, 1981), even though their role in water-soluble proteins is helix destabilization (Lyu *et al*., 1990; Pace and Scholtz, 1998; O'Neil and DeGrado, 1990). Strong helical hydrogen bonds are present in TMs, resulting in a limitation in helix packing (Kim and Cross, 2002; Page, Kim, and Cross, 2008). Therefore, *Gly* and *Pro* are present in the TMs to induce kinks. *Gly* has access to a large backbone dihedrals torsional space, and *Pro* lacks amide proton to form hydrogen bonds. Even though these kinks may seem to destabilize the helical structure, they actually increase the inter-helical surface area to enhance tertiary structural stability (MacKenzie and Fleming, 2008). Additionally, *Gly* was also significant in helix formation through maintaining backbones to come closer (Javadpour, Eilers, Groesbeek, and Smith, 1999). Thus, *Gly* is extremely significant in facilitating tertiary and quaternary structural stability in TMs. Therefore, observing *Gly* as a strong entropy donor in all phases might explain its roles in the receptor.

active-Gp_{2nd} was the state where the highest net TE value of the backbone dihedrals was observed, as plotted in Figure 6.21, which was followed by the *intermediate* state. Phase II, had the lowest net TE value, due to the stability of its structure. Average net TE per amino acid type given in Figure 6.21 demonstrated the residue-specific information on each conformational system. The residue type with different donor or acceptor characteristics in one state compared to other states may be associated with the formation of corresponding receptor conformation. Therefore, net TE values of five conformational states were analyzed in terms of the entropy characteristics of twenty

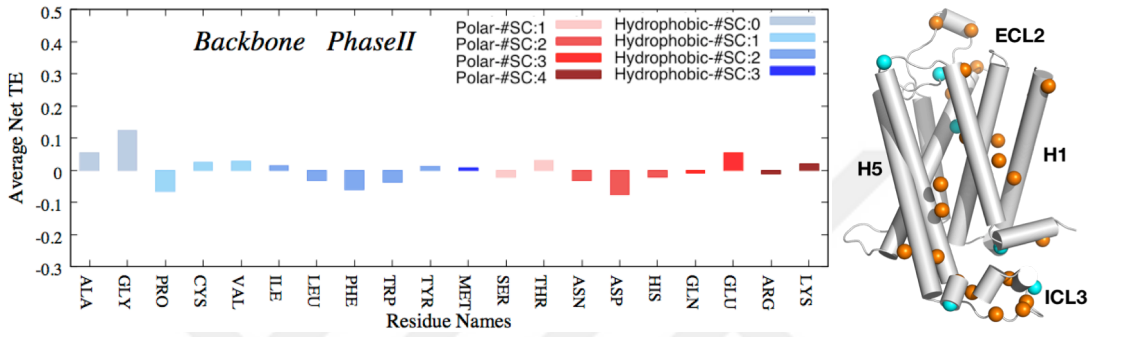
amino acids. The number of rotatable bonds was given as "#SC" in the key label. The polar residues were colored in red and hydrophobic residues were colored in blue.

In Phase I, which was known as the original inactive phase, residues *Gln* and *Glu* were the highest entropy donors, and *Met* was the highest entropy acceptor (Figure 6.21a). These residues mostly located in intracellular and extracellular loops and $H5_{Int}$, explaining the presence of high allosteric communications and the intracellular → extracellular direction of information transfer displayed in our previous studies. Moreover, Phase II, which is the novel inactive state, *Gly* and *Glu* were observed as the highest entropy donors, and *Pro* and *Asp* were the strongest acceptors. *Gly* locates in $H6_{Int}$, H7, H1, H2, as well as ICL3, where they were also determined as strong drivers (Sogunmez and Akten, 2019) and found to play a role in allosteric communication and information transfer in Cα fluctuation analysis.

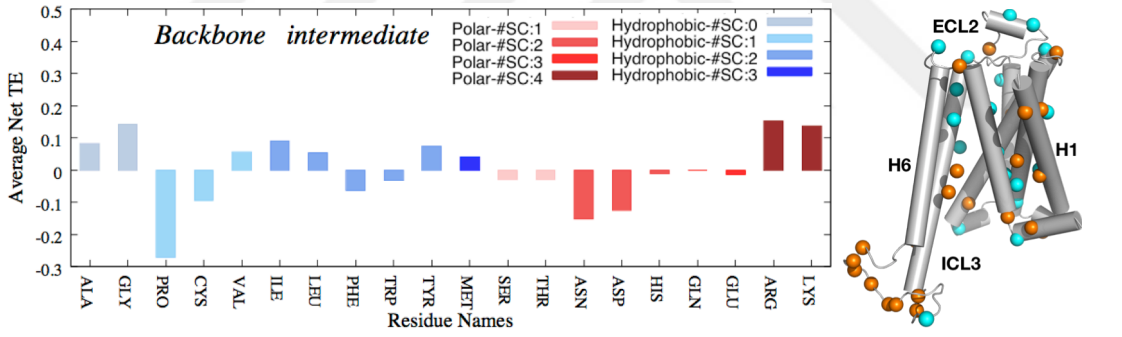
In the *intermediate* phase, *Gly*, *Arg*, and *Lys* were strong entropy donors, where they found on the intracellular and extracellular loops (Figure 6.21c). Even though the *intermediate* state and Phase I showed similar information sharing characteristics both in Cα fluctuation and dihedral angle analysis, roles of different polar amino acids were discovered. *Asn, Asp*, and *Pro* were observed as strong acceptors, where they located on H2, ECL2, $H6_{Ext}$, and H7, where these domains were also found as entropy sinks in the analysis of Cα fluctuations.


Backbone PhaseI
Polar-#SC:1
Polar-#SC:2
Polar-#SC:3
Polar-#SC:4
Hydrophobic-#SC:0
Hydrophobic-#SC:1
Hydrophobic-#SC:2
Hydrophobic-#SC:3
Average Net TE
Residue Names
ALA GLY PRO CYS VAL ILE LEU PHE TRP TYR MET SER THR ASN ASP HIS GLN GLU ARG LYS
ECL2
H1
H5
ICL3

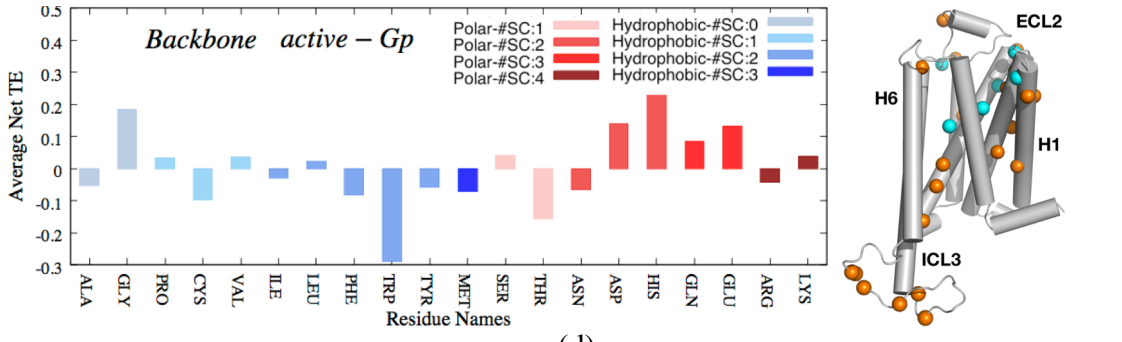
(a)


Backbone PhaseII
Polar-#SC:1
Polar-#SC:2
Polar-#SC:3
Polar-#SC:4
Hydrophobic-#SC:0
Hydrophobic-#SC:1
Hydrophobic-#SC:2
Hydrophobic-#SC:3
Average Net TE
Residue Names
ALA GLY PRO CYS VAL ILE LEU PHE TRP TYR MET SER THR ASN ASP HIS GLN GLU ARG LYS
ECL2
H1
H5
ICL3

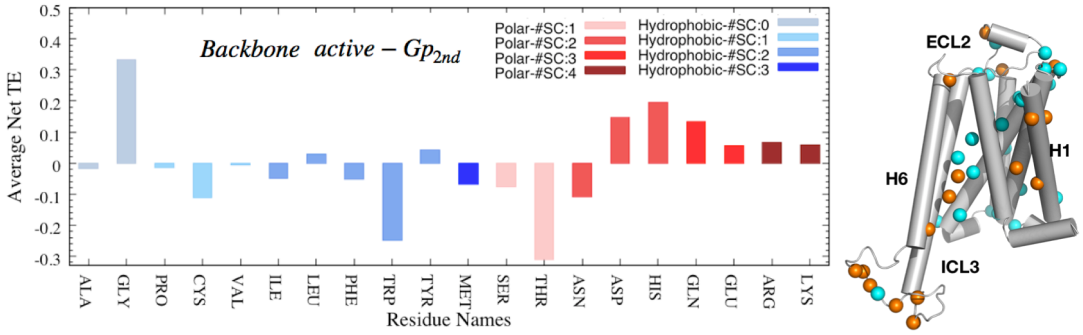
(b)


Backbone intermediate
Polar-#SC:1
Polar-#SC:2
Polar-#SC:3
Polar-#SC:4
Hydrophobic-#SC:0
Hydrophobic-#SC:1
Hydrophobic-#SC:2
Hydrophobic-#SC:3
Average Net TE
Residue Names
ALA GLY PRO CYS VAL ILE LEU PHE TRP TYR MET SER THR ASN ASP HIS GLN GLU ARG LYS
ECL2
H1
H6
ICL3

(c)


Backbone active – Gp
Polar-#SC:1
Polar-#SC:2
Polar-#SC:3
Polar-#SC:4
Hydrophobic-#SC:0
Hydrophobic-#SC:1
Hydrophobic-#SC:2
Hydrophobic-#SC:3
Average Net TE
Residue Names
ALA GLY PRO CYS VAL ILE LEU PHE TRP TYR MET SER THR ASN ASP HIS GLN GLU ARG LYS
ECL2
H1
H6
ICL3

(d)



(e)

Figure 6. 21 Boxplot of backbone dihedrals' net TE data for Phase I (a), Phase II (b), *intermediate* state (c), *active-Gp* (d), and *active-Gp*$_{2nd}$ (e). Polar residues were colored in red and hydrophobic residues were in blue. The amino acids in each group were ordered based on their increasing number of rotatable bonds.

active-Gp was the state, where a large amount of information exchange between receptor domains was observed (Figure 6.21d). *His, Gly, Asp*, and *Glu* were the strongest entropy donors, whereas *Trp* and *Thr* were the acceptors, indicating a direction of information transfer from loops to extracellular parts of the helices H1, H2, H3, H6, and H7. *active-Gp*$_{2nd}$, finally, showed that *Gly* was the strongest entropy donor, whereas *Thr* was the strongest acceptor.

6.1.4.7 The effect of residue type in the net transfer entropy values from backbone + χ_1 dihedrals

Inclusion of χ_1 angle to backbone dihedrals resulted in significantly different net transfer entropy profiles related to the residue types. Pairwise analysis of residue types displayed the importance of *His*, *Thr*, and *Arg* as entropy donors and *Gln*, *Asn,* and *Lys* as entropy acceptors (Figure 6.22). Even though a single rotameric bond is not sufficient to explain the whole dynamics of a side-chain, it can at least direct us to understand the trend of side-chain entropies together with backbone dihedrals. Among these amino acids, *His* is the most versatile amino acid in the protein architectures due to imidazole in its side-chain. The side-chain of *His* can form multiple interactions either with protonated amino acids as *Lys* or *Arg* (Crowley and Golovin, 2005; Priyakumar *et al*., 2004), or aromatic amino acids as *Phe*, *Tyr* and *Trp* (Janiak, 2000; Sygula *et al*., 2007).

Moreover, *Asn* and *Gln* were observed as forming transient side-chain-backbone bonds. Their side chain functionalities appeared similar to backbone amide groups; thus, these bonds were reported to mimic those formed by the hydrogen bonds between backbone atoms. Additionally, it was shown that *His* was mimicking *Asn* in formation of side-chain-backbone hydrogen bond motifs (Vasudev *et al*., 2012).

In Figure 6.22 the highest net transfer entropy value was observed in Phase I between *His→Gln* and *His→Lys*, with values of 4.61 and 4.58, respectively. Figures 6.22a, 6.22b, and 6.22c, entropy donor ability of *His* and *Arg* was observed, while *Gln* appeared as acceptor. Interestingly, these residues were not as strong donors in backbone dihedrals as here, and also, the acceptors were different in backbone dihedrals considering Phase I, Phase II, and *intermediate* state. However, in active states, entropy donor was shifted to *Thr*, while this time *Gln*, *Glu,* and *Lys* were acceptors. *His,* which was acting as entropy donor in backbone dihedrals, was also present here as a donor, but the acceptors were different. *Trp* was an entropy acceptor for backbone dihedrals; however, *Gln*'s acceptor characteristic appeared in here (Figures 6.22d and 6.22e). Since the effect of side-chain dihedrals was obvious, their transfer entropy contribution to the protein was planned to be analyzed. Unfortunately, backbone + side-chain net TE calculation is computationally hard due to memory issues; therefore, the results of net TE values from the side-chain only analysis was given in section 6.1.4.9

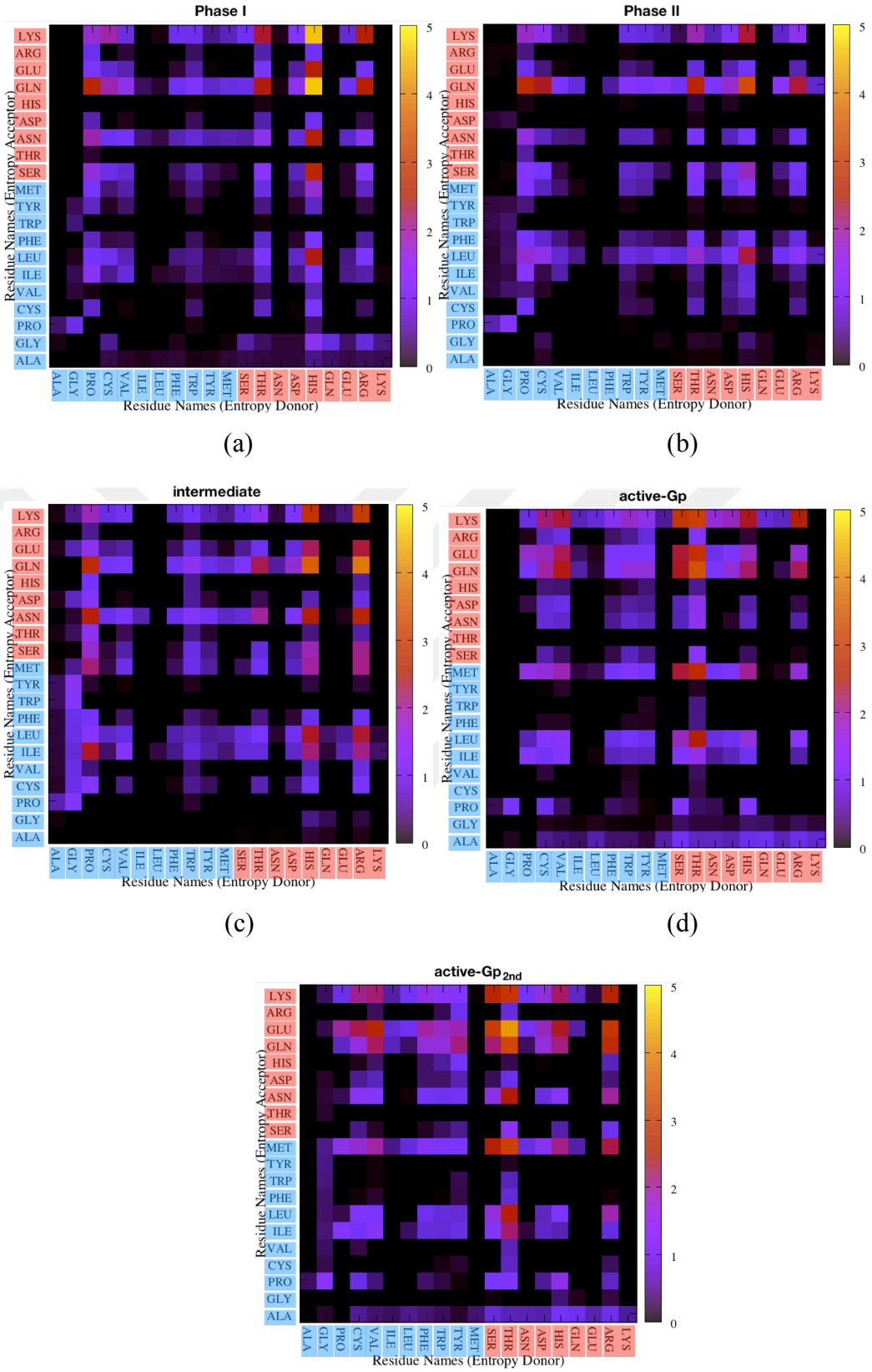


Figure 6. 22 Residue-pair net TE values of backbone + χ_1 dihedrals in Phase I (a), Phase II (b), *intermediate* state (c), *active-Gp* (d), and *active-Gp*$_{2nd}$ (e).

6.1.4.8 The effect of residue type on side-chain information transfer

Residue-pair heatmaps in side-chain transfer entropy analysis revealed that loops significantly donor information, while transmembranes were the major acceptors. This is mainly because the side-chain fluctuations in loops were higher compared to transmembranes, thus increasing the uncertainty. As shown in Figures 6.22 and 6.23, the entropy contribution of polar amino acids was higher than that of hydrophobic ones. Similar to the Cα fluctuation data, the net TE value in the *intermediate* state was observed to be the highest among the five states. *Ala* and *Gly* had no rotatable bonds; thus, their net TE values based on side-chain dihedral angles were zero. Additionally, the rotamer library used for this thesis (van der Kamp *et al*., 2010) reported one rotamer for the amino acid *Pro* with an angle fluctuation range of [-50:50] degrees. Thus we used the χ_1 data of the imidazole ring for the amino acid *Pro* in the transfer entropy calculations based on side-chain dihedrals. The number of rotamers for each amino acid was given in Appendix E.

Interestingly, the polarity and the number of rotatable bonds affected the net TE values. Generally, polar residues with above three rotatable bonds, such as *Gln, Glu, Arg,* and *Lys* were entropy donors in all five states. *Met,* as a hydrophobic amino acid, was also observed as an entropy donor mainly because its number of rotatable bonds is three. Surprisingly, *Pro*, among hydrophobic residues, and *Ser,* among polar residues, were observed as main two entropy acceptors. Since the analyzed five states belonged to the same receptor, the differences in the entropy transfer values and profiles would give us the causes of conformational changes. Therefore, we focused on the residue types, which differ in terms of net TE values in five states.

Pairwise analysis of the residue types demonstrated that the highest three net TE values belong to *Gln-Pro*, *Lys-Pro*, and *Gln-Asp* pairs in Phase I with values of 18.87, 18.34 and 17.76, respectively. Here, *Gln* and *Lys* were the entropy donors, and *Pro* and *Asp* were the entropy acceptors. Similarly, in Phase II, *Gln-Pro*, *Lys-Pro*, and *Gln-Asp* pairs were observed with the highest net TE values as 20.62, 19.79, and 19.03, respectively.

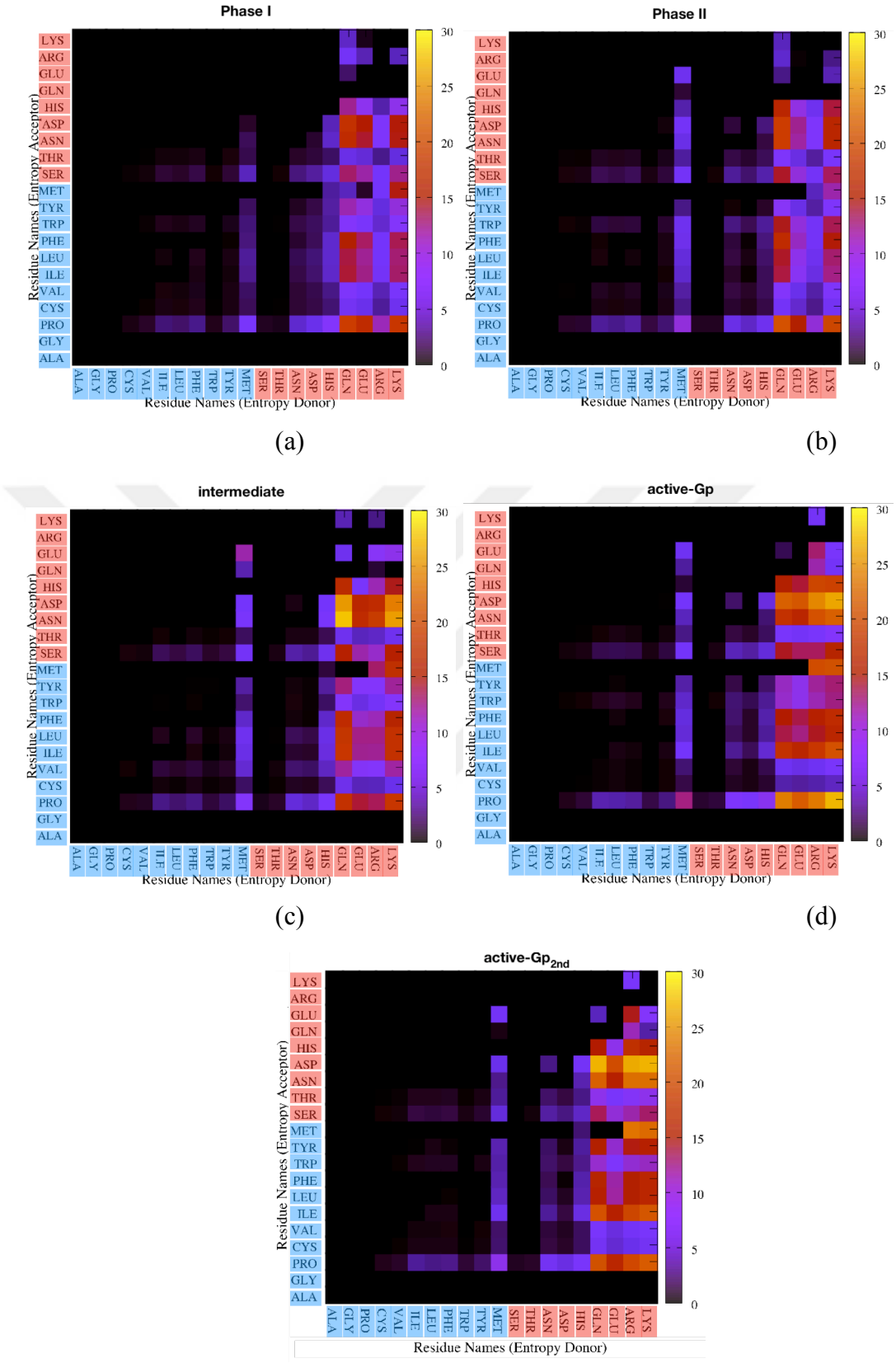


Figure 6. 23 Net TE values for side-chain dihedrals. Phase I (a), Phase II (b), *intermediate* state (c), *active-Gp* (d), and *active-Gp*$_{2nd}$ (e) were given.

In the *intermediate* phase, on the other hand, *Gln-Asn*, *Lys-Asn*, and *Gln-Asp* were observed with values 27.51, 25.63, and 25.49, respectively. In the *active-Gp* phase, *Lys-Pro*, *Lys-Asp*, and *Arg-Asp* pairs were observed with the values 27.04, 26.11, and 24.45, respectively. In *active-Gp$_{2nd,}$* finally, *Lys-Asp, Gln-Asp,* and *Arg-Asp* were the highest three pairs with a mean net TE values of 26.80, 26.60, and 26.22. Therefore, the highest transfer entropy was observed in the *intermediate* state followed by *active-Gp,* and *active-Gp$_{2nd}$*. Phase I has the lowest net TE values based on side-chain dihedrals.

When average net TE values per residue types were analyzed from the Figure 6.23, hydrophobic amino acids were mostly detected as entropy acceptors in Phase I, except *Met*, which slightly showed donor characteristics in Phase I. The amino acid pairs in side-chain net transfer entropy calculation reveals the fact that, *His*, even though the values were low, was an entropy donor in Phase I but became entropy acceptor in Phase II. Similarly, Phase II displayed hydrophobic amino acids as entropy acceptors, except *Met*. Among polar amino acids, when the number of rotatable bonds was above two, they became entropy donors; otherwise, they acted as entropy acceptors. This indicated the direction of information flow was from polar and large amino acids to polar and small, or hydrophobic amino acids, when these two inactive phases were concerned.

Similar to the inactive phases, *His* had dual characteristics in active phases, where it was entropy source in *intermediate* and Phase I state, but entropy sink in Phase II and active phases. Among polar residues, *Glu, Gln, Arg,* and *Lys* were entropy donors, and among hydrophobic residues, *Met* was the entropy donor. Compared to the other phases, where *Glu* and *Lys* had a higher entropy donor feature compared to *Arg*, this time, the values of *Arg* and *Lys* were almost similar. *Pro* was a major entropy sink among all amino acids. Interestingly, its mean TE value was the lowest in *active-Gp* trajectory.

Another concern was the regional difference in the net TE values of the side-chain dihedrals. The comparison of transmembrane and loops based on the residue types with the number of different rotatable bonds was given with violin plots in Figure K.1, generated with the Seaborn visualization library of Python programming language. In

all residues given in five states, it was observed that polar residues had higher net TE values compared to hydrophobic residues. Moreover, polar residues in loops also had higher net TE values compared to hydrophobic residues. When the number of rotatable bonds was compared, the rotatable bonds above three had higher net TE values compared to the number of rotatable bonds below three in both transmembranes and loops.

The number of rotatable bonds in hydrophobic residues is between 0 and 3, and in polar residues, it is between 1 and 4. Residues with one rotatable bond (red) and observed that they were mostly acceptors. Residues with two rotatable bonds (blue), and they were also acceptors. However, residues with three rotatable bonds (yellow) mainly were entropy donors together with residues with four rotatable bonds (grey). The average net TE values for each rotatable bond data was given with black circles. The wider the violin shape around zero is, the larger the distribution of the data around zero. Positive values belonged to donors, and negative values belonged to acceptors.

Tables L1 to L7 represented the summary of this chapter. In Tables L1 and L2, backbone and backbone+side-chain dihedrals' mutual information data in five states were given, respectively. The highest mean MI values of backbone dihedrals were observed in active states. When ICL3 was in an extended conformation, their mean MI values of backbone dihedrals were higher compared to its packed conformations. Inclusions of side-chain dihedrals' data, the MI values were increased significantly. The highest number of residue-pairs was contributed to the strong communications in Phase I, followed by active states. The MI intensities in Phase II were the lowest, as given in L2.

Moreover, net transfer entropy data of backbone, backbone + χ_1, and side-chain dihedrals presented significant data. In backbone dihedrals, Phase I, and Phase II differed in terms of ICL2 and ICL3, where ICL2 was acceptor in Phase I but donor in Phase II, on the other hand, ICL3 was a donor in Phase I, but it was generally acceptor in Phase II. Interestingly, ICL3 also possessed a distinct role in the *intermediate* state and active states, where it was majorly an acceptor in the *intermediate* state, while a

donor in active states. Also, $H7_{Int}$ was entropy donor in *intermediate* state, but it was acceptor in active states. The comparison of two active trajectories resulted in a difference in ICL2. The edges of ICL2 were entropy acceptors in *active-*$Gp_{2nd,}$ but they were entropy acceptors in *active-Gp.* The remaining donors and acceptors were also given in Table L3.

Inclusion of one dihedral angle, namely χ_1, increased the net TE values drastically. Loops contribution to the net TE was increased, and transmembranes were also observed as donors and acceptors, which were missing in backbone data. ICL3 became acceptor in Phase II, while it was generally a donor in the remaining four states. Interestingly, loops performed dual characteristics, where they both perform donor and acceptor features (Table L4). Even though χ_1, by itself, is not sufficient to explain all dynamics of the residues, it at least gave us an idea about the direction of information transfer when the side-chain dihedral was present.

Side-chain only dihedrals' net TE data was mostly captured loops as entropy donors in all five states. Residue-based differences were observed between phases, as given in Table L5. One interesting result obtained from the comparison of phases was that, same residues in *active-Gp* and *active-*Gp_{2nd} differed from the remaining three states. Such as *His*172 and *Glu*229 was observed as entropy acceptor in two active states, whereas it was a donor in Phase I. Since this trend was observed in all phase comparisons, we can conclude that dihedral angle alterations in active phases were the same, even though conformational differences based on ICL3 fluctuations were observed.

Additionally, residue-type differences in proximal MI values represented that active states differ from inactive and *intermediate* states in terms of the high MI value of *Arg*. Moreover, Phase II was isolated from the remaining states in terms of Gln's high MI value, where *Arg* had the highest MI value in other states in distal pairs (Table L6). The dominant role of polar residues in information sharing was observed both in proximal and in distal pairs, verifying the data that polar residues contributing entropically to the protein structure.

Finally, the residue type effect was more noticeable in net TE values. The effect of Gly and Pro was visible for backbone dihedrals since these two residues were important in facilitating structural stability in transmembrane helices. The inclusion of χ_1, and side-chain only net TE values; on the other hand, change the donor characteristics to more polar residues. Active states differ in backbone + χ_1 net TE in terms of *Thr* residues as entropy donors, compared to *His* in the remaining phases. In side-chain net TE values, we observed residues with at least three rotatable bonds were entropy donors, but also *Pro* was the entropy acceptor in all phases. As observed in Table L7, *Gln* and *Lys* were the major donors in Phase I, Phase II, and *intermediate* state, whereas *Arg* and *Lys* were the donors in active states when side-chain dihedrals were compared.

7. CONCLUSION

7.1 Intrinsic Dynamics in Inactive Human β_2AR Conformational States

The first analysis was based on determination of linear and time-delayed correlations based on alpha Carbon fluctuations in order to identify the driver and follower residue-pairs. The distribution of positive linear correlations overlapped mostly with the spatially proximal residue-pairs, where anti-correlations were shifted from H5 to $H1_{Ext}$, ECL1, ICL2, ECL2 and $H5_{Ext}$ upon ICL3 closure. Interestingly, as the overall flexibility of the receptor decreased in Phase II, the number of highly correlated residue-pairs were also decreased.

Driver-follower relationships between residue-pairs based on time-dependency was determined to understand the directionality of correlated motions. It explains how the previous fluctuations of a residue controls the future fluctuations of another residue. Time-delayed correlation analysis, denoted as CTC, yielded causalities between H8-ICL3, ICL3-H3, ICL3-H5, H5-H6, where the initial domain drives the future fluctuations of the second domain in Phase I. A plausible communication pathway from driver to follower was established as "H8 → ICL3 → H5 → H6". In phase II, on the other hand, driver-follower relations between ICL3-H7, ICL3-H1, ICL3-H2 and ICL3-H8 were discovered. It indicated that even in its closely packed form, ICL3 was still an important allosteric site, but now controlling a completely different part of the receptor. Interestingly, H5 and H6 were observed as drivers for ECL2, which is known to act as a gate in the process of orthosteric ligand binding. Finally, a relationship between the activity of the rreceptor and causality was observed, which led us to pursue the allosteric interactions and causality data in the receptor by including additional trajectories in our analysis.

7.2 Coarse-Grained Entropy Calculations for Inactive and Active Human β_2AR Trajectories

The same inactive phases used in Chapter 4 were also used in Chapter 5 for identification of the information sharing and then the information transfer between residues. These two phases were important in a way that the original inactive state acted as a general reference, where a number of studies were conducted and considerable amount of information was gathered. Additionally, the novel inactive conformational state provided a more complete picture of the allosteric communication in β_2AR. Additionally, we included two new conformational states to our analysis, namely *intermediate* state, where ICL3 in the active receptor-only system was rapidly closed and the system started to show intermediate system characteristics, and two simulations of active states, denoted as *active-Gp* and *active-Gp*$_{2nd}$, where active receptor was simulated together with an intracellularly bound trimeric G-protein in order to maintain the extended ICL3 conformation. The trajectory of these two active states differed in terms of the deviation of ICL3. ICL3 had a major perpendicular fluctuation coupled with horizontal motions in *active-Gp*$_{2nd}$, which was missing in *active-Gp*. As explained below, these five distinct conformations were compared based on the effect of their Cα fluctuations on pairwise information sharing and the direction of transfer entropy.

In two independent simulations of active receptor with intracellularly associated G protein, two distinct active state profiles were observed. One of the profiles was relatively stable in time, and the other one displayed constrained ICL3 fluctuations. However, these fluctuations of ICL3 were perpendicular to the membrane and away from the α-helical bundle of Gα domain, where the changes in x- and y- directions of ICL3 fluctuations were relatively stable compared to the changes in z- direction. Mutual information data displayed interesting results, where the amount of communication decreased significantly compared to other three states, Phase I, Phase II, and *intermediate* state. Moreover, the effect of polarity was also significantly reduced, indicating the effect of the presence of G protein, which might alter the

microenvironment of the system. Finally, the entropy donors were mostly observed in extracellular parts and in ICL3 in both states. $H1_{Int}$, H8, $H7_{Ext}$ changed their entropy transfer characteristics based on the mobility of ICL3. When ICL3 was stable, $H1_{Int}$ and $H7_{Ext}$ were acceptors, while H8 was a donor, however, when it was mobile, $H1_{Int}$ and $H7_{Ext}$ became donors but H8 became acceptor. ICL3 was also observed as entropy source in these two states, but the strength of entropy transfer in *active-*Gp_{2nd} was higher than *active-Gp*. This verified that unless it is strongly packed, ICL3 involves in the regulation of intrinsic dynamics of the receptor.

Moreover, all five conformationally distinct states represented activity specific information sharing and transfer patterns. The mobility of ICL3 and the deviation of $H6_{Int}$, as well as general flexibility of the receptor changed both the location of pairwise communications and the direction of information transfer. In terms of mutual information, pairwise dependency in Phase I was the highest, followed by *active-Gp* and *intermediate* state, then followed by Phase II and *active-*Gp_{2nd}. Except initial three helices in *active-Gp* and ICL2, ECL2, and ICL3 in *intermediate* state, the MI values of these states were significantly low. These residues with low MI values indicated the lack of communication due to independency between residue-pairs. However, in Phase II, the increased stiffness decreased entropy, thus the joint entropy of instantaneous events between two residues, hence the MI value was decreased. Thus, except Phase II, which was a novel and highly inactive state, the pairwise communications were decreased from original inactive to active states.

Interestingly, the mutual information was exchanged mostly through polar residues, except *active-Gp* and *active-*Gp_{2nd}, and the average amount of shared information between polar residue-pairs was significantly higher than that of hydrophobic-hydrophobic pairs. On the other hand, relatively high number of polar residues located on the loop regions, which positively influenced the mutual information values. Moreover, the average mutual information shared by proximal residues was higher compared to distal residues, indicating that mutual information represented distance dependency. For distal residues, although the mutual information value was low compared to the mutual information values of proximal pairs, the highest degree of

mutual information was shared between loop regions at distant sites. Interestingly, in *active-Gp* and *active-Gp*$_{2nd}$, the statistical validation of the positive effect of polar residues in communication was failed and this can be explained by the affect of G protein on the electrostatic distribution of especially intracellular parts of the receptor in these two runs. Without the presence of G protein, on the other hand, we observed strong communication around ICL2, ICL3 with ECL2 and H5$_{Ext}$.

Additionally, Phase II formed a conformational transition due to the packing of ICL3, and released extracellular parts of the receptor. Here we showed increased entropic control of the pocket surrounded by H1, H2, ECL2, and H7 by the residing residues, by transferring the information to the future states of the remaining amino acids in the receptor in Phase II. Also, *active-Gp* displayed entropy donors in some of the extracellular regions as mentioned for Phase II. The intracellular constrain on the ICL3 by association of trimeric G protein prevented horizontal ICL3 fluctuation, thus induced extracellular parts to involve in information transfer and allosteric coupling. Our entropy sources in the *active-Gp* and Phase II displayed common characteristics with the allosteric binding sites with Class A GPCRs, where the vestibule between ECL1/2 near H1, H2, H3, and H7 was also found to be the modulator-binding site for PAR2 receptor (Cheng *et al*., 2017), as well.

Besides, information acceptor sites located in ICL3-H6 intersection was observed in inactive phases. In Phase I, the acceptor sites were located in ICL3 followed by donors in H6$_{Int}$. In Phase II, on the other hand, they are in between ICL3-H6$_{Int}$ intersection. In the *intermediate* phase the acceptors lie around H6$_{Int}$ and in active phase it relatively shifted towards H6$_{Mid}$. This site might be acting as a switch control to both outward motion of H6 and release of ICL3 after receiving signals from the binding sites of the receptor. Additionally, it is known that ionic lock was observed between *Arg*131 and *Glu*268 in some inactive crystal structures of the β_2AR. The shifting of the acceptor to the *highly inactive state* through ICL3 closure might be involved in the ionic-lock formation process. The formation of more persistent ionic lock was observed in Phase II trajectory compared to Phase I, due to ICL3 packing induces a further inward tilt in intracellular part of H6.

The most exciting discoveries in β_2AR allostery were the resolution of the crystal structure of negative and positive allosteric inhibitors, *Cmpd-15PA*, and *Cmpd-6FA*, which were partially buried in the intracellular pocket of carazolol-bound β_2AR and *BI-167107* bound β_2AR, respectively. *Cmpd-15PA* was intracellularly interacting with the residues on H1, H2, H6, H7, H8, and ICL1 (Liu *et al*., 2017) and *Cmpd-6FA* was intracellularly interacting with the residues H2, H3, H4, and ICL2 (Liu *et al*., 2019). Our information transfer data, especially in Phase I and *intermediate* state, spotted on the specific intracellular pocket by introducing the entropy sources, which especially located in the intracellular domains of H1, H2, ICL2, H5, and H6. Moreover, ICL3, which was still missing in almost all X-ray structures, also found to act as an entropy source in all phases except Phase II. However, it acted as an entropy sink in Phase II especially with its residues adjacent to H6. Moreover, with the introduction of *Cmpd-6FA* as a positive allosteric modulator for β_2AR, the residues interacting with this modulator have become popular. Our analysis revealed these PAM interacted sites were entropy sources in *intermediate* state and Phase I, where they were neutral in the remaining states.

7.3 Dihedral Angle Based Information Entropy for Inactive and Active Human β_2AR Trajectories

The role of dihedral angles was mostly neglected in explaining protein dynamics. Therefore, together with Cα fluctuations, the entropy analysis of the backbone and side-chain dihedrals of five trajectories was performed. It was reported that the B-factors and backbone dihedrals had at most 0.29 correlation, and backbone dihedrals also had a correlation of 0.53, explaining the differences observed from the calculations of Cα fluctuations and backbone+side-chain dihedral motions. The mutual information data originated from backbone dihedrals mostly captured loops, where the highest backbone MI was observed in active states. Interestingly*, intermediate* state and Phase II had the lowest backbone MI values, indicated that the closure of ICL3 negatively affected the dependency of the residue-pairs' backbone dihedrals. The inclusion of all side-chain dihedral data to the backbone dihedral data resulted an increase in both the MI values

and the numbers of information sharing residue-pairs. Except Phase II, high information sharing between ICL3 and other residues were observed in all states. The highest number of residue-pairs with strong information sharing was observed in Phase I, whereas residue-pairs in Phase II was appeared to be least communicative among five states. Even though the large horizontal deviation of ICL3 was restricted by G protein in active states, ICL3 was still predicting the future fluctuations of the remaining residues of the receptor.

These observations pinpointed the possible effect of steric hindrance of dihedral angles on the information sharing and transfer. Minor fluctuations of Cα atom of a residue can be observed but dihedral angles in its backbone or side-chains do not always rotate freely. Especially in transmembranes, a coupled movement including Cα fluctuations and dihedral angles might not be possible due to the lack of space for side-chain atoms. Moreover, the fluctuation of Cα atoms were absolute numbers, where a number of combinations of x-, y- and z-directional motions can result in a single absolute fluctuation value. However each of these fluctuations, even though they have same values, did not suppose to result in the same angular motion. Thus a divergence between the MI and TE values of Cα fluctuations and dihedral data can be expected and was observed. Overall, the receptor dynamics was originated from the combinatorial effect of both Cα fluctuations and dihedral angles. Thus, each of the motions originated from inactive and active states had their own contributions to the receptor activation.

Both the mutual information and transfer entropy results of backbone and side-chain dihedrals represented similarities between two simulations of the active state. This finding indicated that dihedral angles could be successfully used in determination of the similarities between conformations. Moreover, the transfer entropy calculated from backbone dihedrals revealed the entropy donors and acceptors were located mostly in loops and transmembrane helices had small changes in their net TE values. Communications of intracellular and extracellular loops existed both in Phase I and Phase II, with a significant decrease of MI values in Phase II. The closure of ICL3 also decreased the information-sharing role of ICL3, as well as increased the MI values in ICL1 and ECL1. Additionally, the *intermediate* state displayed one of the lowest MI

features among five states, displaying self-communication in ICL3, ICL3-ICL1, and ICL3-$H7_{Int}$, which were the residues where ICL3 interacted and formed transient hydrogen bonds throughout the trajectory. Active states, where intracellular G-protein association existed, represented strong pairwise dependencies based on backbone dihedrals. Especially ICL3 and ECL3 communicated with H3, ICL2, and ECL2 adjacent to H4 in first run, but they communicated with ICL1, two edges of ECL2 and $H7_{Int}$ in the second run.

Inclusion of χ_1 angle to the backbone dihedral data enhanced the donor and acceptor characteristics of loops. Moreover, some residues changed their net TE features, such as ICL2 in Phase I. ICL2 was accepting information when backbone dihedrals were considered, however it became entropy donor upon considering the information from χ_1 angle. Transmembrane helices were also observed as donors and acceptors upon inclusion of one dihedral angle, while they were mostly displayed insignificant data when backbone dihedrals were considered. Based on this, ICL3 in *intermediate* and active states was the highest entropy donor, where ECL2 was mostly entropy acceptor. ICL2 and ECL2 was entropy donor in Phase II, and interestingly the average net TE value of Phase II was higher than Phase I. Considering the Cα fluctuations net TE data, almost no similarities exist in backbone+χ_1 angle net TE values, indicating the angular variations differ in the absolute fluctuations, and both of their contributions to the protein maintain the balance of the intrinsic protein dynamics. Even though only χ_1 angle was not sufficient to explain all contributions of side-chain dihedrals on receptor dynamics, it guided us how different receptor conformations were affected when χ_1 angle was presented together with backbone dihedrals.

Side-chain transfer entropy calculations located loops as entropy sources, independent of the conformation. But still, some residue-level differences between the phases were observed. When two inactive phases were considered, residues located in ICL3, ECL3, and H7 were entropy donors in Phase I but became acceptors in Phase II. *Met*96 located on ECL1 was entropy acceptor in Phase I but became entropy donor in Phase II. These types of changes were observed in residues, which have at least two but mostly three rotatable bonds. Additionally, the comparison of Phase I, Phase II, or *intermediate* state

with active states displayed that the residues with acceptor/donor characteristics were highly similar between active states compared to remaining three states. This indicated that two independent runs of different active states presented highly similar dihedral angle motions in their trajectories, however they differ when the receptor converted to different conformational states.

Moreover, side-chain dihedrals of polar residues, which were mostly locating around the loops, shared more information than the hydrophobic residues. Even though the number of residues in the loop regions is lower than that of the transmembrane regions, high information sharing between the loops or the residues adjacent to loops was seen. This was also observed in preceding analysis based on Cα fluctuations. However, data derived from dihedrals resulted significantly higher differences between polar and hydrophobic pairs, which had almost three times more information sharing capacity in polar residues than hydrophobic amino acids. Among the amino acids, *Arg, Lys*, and *Gln* had the highest average MI value compared with others, where *Arg* and *Lys* maintain protein stability by forming ionic interactions and hydrogen bonds with other residues and with water molecules, and *Gln* were known as forming transient side-chain-backbone bonds.

Additionally, the side-chain analysis of amino acid types revealed that polar residues were entropy donors, especially those with three and four rotatable bonds. However, residues with polar characteristics but below three rotatable bonds together with hydrophobic residues were mostly entropy acceptors. Except for *Met* in all phases, all hydrophobic residues were appeared as entropy sinks. Additionally, *His* in the polar group, slightly donored entropy in Phase I and *intermediate* state but accepted entropy in Phase II and Active Gp. *His* located in extracellular and intracellular loops. The limited motion in Phase II and *active-Gp* resulted in a decrease of uncertainty (or entropy) therefore might decreased in entropy transfer in these two phases. Also, the comparison of the polar and hydrophobic residues' location was also displayed the strong entropic contribution of loops over transmembranes. Overall, locating in loops, having at least three rotatable bonds and having polar characteristics are the three critical factors for a residue being entropy donor in side-chain dihedrals.

Consequently, five distinct conformational states of human β_2AR yielded extremely significant information based on the intrinsic dynamics of the receptor. With the inclusion of antagonists to the known inactive conformation and agonists to the active structure with G protein in upcoming analysis might present the differences of the fully inactive and fully active receptor behaviors of the receptor, respectively. Additionally, the transfer entropy analysis of backbone+side-chain dihedrals will also clarify how information is transferred in each conformational state. Therefore, this full TE analysis needs to be in the future plans to fully understand the combinatorial effect of Cα, backbone and side-chain dihedrals.

REFERENCES

Ahn, S., Kahsai, A. W., Pani, B., Wang, Q.-T., Zhao, S., Wall, A. L., et al. (2017). Allosteric "beta-blocker" isolated from a DNA-encoded small molecule library. *Proceedings of the National Academy of Sciences*, *114*(7), 1708–1713.

Andrej Šali. (1993). MODELLER A Program for Protein Structure Modeling. *Comparative Protein Modelling by Satisfaction of Spatial Restraints.*

Balabin, I. A., Yang, W., and Beratan, D. N. (2009). Coarse-grained modeling of allosteric regulation in protein receptors. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0901811106

Ballesteros, J. A., Jensen, A. D., Liapakis, G., Rasmussen, S. G. F., Shi, L., Gether, U., and Javitch, J. A. (2001). Activation of the β2-Adrenergic Receptor Involves Disruption of an Ionic Lock between the Cytoplasmic Ends of Transmembrane Segments 3 and 6. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M103747200

Barlow, D. J., and Thornton, J. M. (1983). Ion-pairs in proteins. *Journal of Molecular Biology*. https://doi.org/10.1016/S0022-2836(83)80079-5

Barr, D., Oashi, T., Burkhard, K., Lucius, S., Samadani, R., Zhang, J., et al. (2011). Importance of domain closure for the autoactivation of ERK2. *Biochemistry*. https://doi.org/10.1021/bi200503a

Bayly, C. I., Merz, K. M., Ferguson, D. M., Cornell, W. D., Fox, T., Caldwell, J. W., et al. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*. https://doi.org/10.1021/ja00124a002

Beck, D. A. C., Jonsson, A. L., Schaeffer, R. D., Scott, K. A., Day, R., Toofanny, R. D., et al. (2008). Dynameomics: Mass annotation of protein dynamics and unfolding in water by high-throughput atomistic molecular dynamics simulations. *Protein Engineering, Design and Selection*. https://doi.org/10.1093/protein/gzn011

Bhattacharya, S., and Vaidehi, N. (2014). Differences in allosteric communication pipelines in the inactive and active states of a GPCR. *Biophysical Journal*.

https://doi.org/10.1016/j.bpj.2014.06.015

Borders, C. L., Broadwater, J. A., Bekeny, P. A., Salmon, J. E., Lee, A. S., Eldridge, A. M., and Pett, V. B. (1994). A structural role for arginine in proteins: Multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Science*. https://doi.org/10.1002/pro.5560030402

Brooks, B. R., Brooks, III, C. L., Mackerell, Jr., A. D., Nilsson, L., Petrella, R. J., et al. (2009). CHARMM: The Biomolecular Simulation Program. *JOURNAL OF COMPUTATIONAL CHEMISTRY*.

Bruzzese, A., Gil, C., Dalton, J. A. R., and Giraldo, J. (2018). Structural insights into positive and negative allosteric regulation of a G protein-coupled receptor through protein-lipid interactions. *Scientific Reports*. https://doi.org/10.1038/s41598-018-22735-6

Buck, M., Bouguet-Bonnet, S., Pastor, R. W., and MacKerell, A. D. (2006). Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme. *Biophysical Journal*. https://doi.org/10.1529/biophysj.105.078154

Cellucci, C. J., Albano, A. M., and Rapp, P. E. (2003). Comparative study of embedding methods. *Physical Review E*, *67*(6), 66210.

Chakrabarti, P., and Pal, D. (1998). Main-chain conformational features at different conformations of the side-chains in proteins. *Protein Engineering*. https://doi.org/10.1093/protein/11.8.631

Chandrasekaran, R., and Ramachandran, G. N. (1970). Studies on the conformation of amino acids. XI. Analysis of the observed side group conformation in proteins. *International Journal of Protein Research*.

Chelikani, P., Hornak, V., Eilers, M., Reeves, P. J., Smith, S. O., RajBhandary, U. L., and Khorana, H. G. (2007). Role of group-conserved residues in the helical core of β2-adrenergic receptor. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0702024104

Cheng, R. K. Y., Fiez-Vandal, C., Schlenker, O., Edman, K., Aggeler, B., Brown, D. G., et al. (2017). Structural insight into allosteric modulation of protease-activated receptor 2. *Nature*. https://doi.org/10.1038/nature22309

Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Foon, S. T., Kobilka, T. S., et al. (2007). High-resolution crystal structure of an engineered human β2-adrenergic G protein-coupled receptor. *Science*. https://doi.org/10.1126/science.1150577

Clarkson, M. W., Gilmore, S. A., Edgell, M. H., and Lee, A. L. (2006). Dynamic coupling and allosteric behavior in a nonallosteric protein. *Biochemistry*. https://doi.org/10.1021/bi060652l

Cooper, A., and Dryden, D. T. F. (1984). Allostery without conformational change - A plausible model. *European Biophysics Journal*. https://doi.org/10.1007/BF00276625

Corrada, D., Morra, G., and Colombo, G. (2013). Investigating allostery in molecular recognition: Insights from a computational study of multiple antibody-antigen complexes. *Journal of Physical Chemistry B*. https://doi.org/10.1021/jp310753z

Crowley, P. B., and Golovin, A. (2005). Cation-π interactions in protein-protein interfaces. *Proteins: Structure, Function and Genetics*. https://doi.org/10.1002/prot.20417

Deupi, X., and Kobilka, B. (2007). Activation of G Protein-Coupled Receptors. *Advances in Protein Chemistry*. https://doi.org/10.1016/S0065-3233(07)74004-4

Devree, B. T., Mahoney, J. P., Vélez-Ruiz, G. A., Rasmussen, S. G. F., Kuszak, A. J., Edwald, E., et al. (2016). Allosteric coupling from G protein to the agonist-binding pocket in GPCRs. *Nature*. https://doi.org/10.1038/nature18324

Donald, J. E., Kulp, D. W., and DeGrado, W. F. (2011). Salt bridges: Geometrically specific, designable interactions. *Proteins: Structure, Function and Bioinformatics*. https://doi.org/10.1002/prot.22927

Dror, R. O., Arlow, D. H., Borhani, D. W., Jensen, M., and Shaw, D. E. (2009). Identification Of Two Distinct Inactive Conformations Of The Beta-2 Adrenergic Receptor Reconciles Structural And Biochemical Observations. *Biophysical Journal*. https://doi.org/10.1016/j.bpj.2008.12.1964

Dror, R. O., Arlow, D. H., Maragakis, P., Mildorf, T. J., Pan, A. C., Xu, H., et al. (2011). Activation mechanism of the β 2-adrenergic receptor. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1110499108

Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., and Shaw, D. E. (2012). Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics*. https://doi.org/10.1146/annurev-biophys-042910-155245

DuBay, K. H., Bothma, J. P., and Geissler, P. L. (2011). Long-range intra-protein communication can be transmitted by correlated side-chain fluctuations alone. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1002168

Dunn, O. J. (1964). Multiple Comparisons Using Rank Sums. *Technometrics*. https://doi.org/10.1080/00401706.1964.10490181

Eilers, M., Patel, A. B., Liu, W., and Smith, S. O. (2002). Comparison of helix interactions in membrane and soluble α-bundle proteins. *Biophysical Journal*. https://doi.org/10.1016/S0006-3495(02)75613-0

Erman, B. (2015). Effects of ligand binding upon flexibility of proteins. *Proteins: Structure, Function and Bioinformatics*. https://doi.org/10.1002/prot.24785

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M., et al. (2006). Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics*, *15*(1), 5–6.

Flor, P. J., and Acher, F. C. (2012). Orthosteric versus allosteric GPCR activation: the great challenge of group-III mGluRs. *Biochemical Pharmacology*, *84*(4), 414–424.

Fraser, A. M., and Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review A*. https://doi.org/10.1103/PhysRevA.33.1134

Galés, C., Van Durm, J. J. J., Schaak, S., Pontier, S., Percherancier, Y., Audet, M., et al. (2006). Probing the activation-promoted structural rearrangements in preassembled receptor-G protein complexes. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb1134

Gilman, A. (1987). G Proteins: Transducers Of Receptor-Generated Signals. *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev.biochem.56.1.615

Goodey, N. M., and Benkovic, S. J. (2008). Allosteric regulation and catalysis emerge via a common route. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.98

Grassberger, P. (1988). Finite sample corrections to entropy and dimension estimates. *Physics Letters A*. https://doi.org/10.1016/0375-9601(88)90193-4

Guarnera, E., and Berezovsky, I. N. (2016). Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1004678

Gunasekaran, K., Ma, B., and Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins? *Proteins: Structure, Function, and Bioinformatics*, *57*(3), 433–443.

Hacisuleyman, A., and Erman, B. (2017). Causality, transfer entropy, and allosteric

communication landscapes in proteins with harmonic interactions. *Proteins: Structure, Function, and Bioinformatics*, *85*(6), 1056–1064.

Hacisuleyman, A., and Erman, B. (2017). Entropy transfer between residue-pairs and allostery in Proteins: Quantifying allosteric communication in Ubiquitin. *PLoS Computational Biology*, *13*(1), e1005319.

Hanson, M. A., Cherezov, V., Griffith, M. T., Roth, C. B., Jaakola, V. P., Chien, E. Y. T., et al. (2008). A Specific Cholesterol Binding Site Is Established by the 2.8 Å Structure of the Human β2-Adrenergic Receptor. *Structure*. https://doi.org/10.1016/j.str.2008.05.001

Harms, M. J., Schlessman, J. L., Sue, G. R., and Bertrand García-Moreno, E. (2011). Arginine residues at internal positions in a protein are always charged. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1104808108

Herzel, H., Schmitt, A. O., and Ebeling, W. (1994). Finite sample effects in sequence analysis. *Chaos, Solitons and Fractals*. https://doi.org/10.1016/0960-0779(94)90020-5

Hille, B. (1992). G protein-coupled mechanisms and nervous signaling. *Neuron*. https://doi.org/10.1016/0896-6273(92)90158-A

Hu, J., Wang, Y., Zhang, X., Lloyd, J. R., Li, J. H., Karpiak, J., et al. (2010). Structural basis of G proteing-coupled receptorg-G protein interactions. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.385

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. https://doi.org/10.1016/0263-7855(96)00018-5

Igumenova, T. I., Frederick, K. K., and Wand, A. J. (2006). Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chemical Reviews*. https://doi.org/10.1021/cr040422h

Janiak, C. (2000). A critical account on n-n stacking in metal complexes with aromatic nitrogen-containing ligands. *Journal of the Chemical Society, Dalton Transactions*. https://doi.org/10.1039/b003010o

Javadpour, M. M., Eilers, M., Groesbeek, M., and Smith, S. O. (1999). Helix packing in polytopic membrane proteins: Role of glycine in transmembrane helix association. *Biophysical Journal*. https://doi.org/10.1016/S0006-3495(99)77009-8

Jones, J. E. (1924). On the Determination of Molecular Fields. I. From the Variation of the Viscosity of a Gas with Temperature. *Proceedings of the Royal Society A:*

Mathematical, Physical and Engineering Sciences. https://doi.org/10.1098/rspa.1924.0081

Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*. https://doi.org/10.1021/ja9621760

Kamberaj, H., and van der Vaart, A. (2009). Extracting the causality of correlated motions from molecular dynamics simulations. *Biophysical Journal*, *97*(6), 1747–1755.

Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural Biology*. https://doi.org/10.1038/nsb0902-646

Kern, D., and Zuiderweg, E. R. P. (2003). The role of dynamics in allosteric regulation. *Current Opinion in Structural Biology*. https://doi.org/10.1016/j.sbi.2003.10.008

Kim, S., and Cross, T. A. (2002). Uniformity, ideality, and hydrogen bonds in transmembrane α-helices. *Biophysical Journal*. https://doi.org/10.1016/S0006-3495(02)73969-6

Kobilka, B. (2013). The structural basis of G-protein-coupled receptor signaling (nobel lecture). *Angewandte Chemie - International Edition*. https://doi.org/10.1002/anie.201302116

Koch, W. J., Lefkowitz, R. J., and Rockman, H. A. (2000). Functional Consequences of Altering Myocardial Adrenergic Receptor Signaling. *Annual Review of Physiology*. https://doi.org/10.1146/annurev.physiol.62.1.237

Kohm, A. P., and Sanders, V. M. (2001). Norepinephrine and β2-adrenergic receptor stimulation regulate CD4+ T and B lymphocyte function in vitro and in vivo. *Pharmacological Reviews*.

Koshland, D. E., Nemethy, J. G., and Filmer, D. (1966). Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry*. https://doi.org/10.1021/bi00865a047

Kruskal, W. H., and Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.1952.10483441

Kučerka, N., Nieh, M. P., and Katsaras, J. (2011). Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochimica et Biophysica Acta - Biomembranes*.

https://doi.org/10.1016/j.bbamem.2011.07.022

Kumar, S., Tsai, C. J., and Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein Engineering*. https://doi.org/10.1093/protein/13.3.179

Lachance, M., Ethier, N., Wolbring, G., Schnetkamp, P. P. M., and Hébert, T. E. (1999). Stable association of G proteins with β2AR is independent of the state of receptor activation. *Cellular Signalling*. https://doi.org/10.1016/S0898-6568(99)00024-8

Lang, P. T., Ng, H. L., Fraser, J. S., Corn, J. E., Echols, N., Sales, M., et al. (2010). Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Science*. https://doi.org/10.1002/pro.423

Leibler, R. A., and Kullback, S. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*(1), 79–86.

Lenaerts, T., Ferkinghoff-Borg, J., Stricher, F., Serrano, L., Schymkowitz, J. W. H., and Rousseau, F. (2008). Quantifying information transfer by protein domains: Analysis of the Fyn SH2 domain structure. *BMC Structural Biology*. https://doi.org/10.1186/1472-6807-8-43

Li, X., Chen, Y., Lu, S., Huang, Z., Liu, X., Wang, Q., et al. (2013). Toward an understanding of the sequence and structural basis of allosteric proteins. *Journal of Molecular Graphics and Modelling*. https://doi.org/10.1016/j.jmgm.2012.12.011

Liapakis, G., Ballesteros, J. A., Papachristou, S., Chan, W. C., Chen, X., and Javitch, J. A. (2000). The forgotten serine. A critical role for Ser-2035.42 in ligand binding to and activation of the β2-adrenergic receptor. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M002092200

Liu, Xiangyu, Ahn, S., Kahsai, A. W., Meng, K. C., Latorraca, N. R., Pani, B., et al. (2017). Mechanism of intracellular allosteric β 2 AR antagonist revealed by X-ray crystal structure. *Nature*. https://doi.org/10.1038/nature23652

Liu, Xiangyu, Masoudi, A., Kahsai, A. W., Huang, L. Y., Pani, B., Staus, D. P., et al. (2019). Mechanism of β2AR regulation by an intracellular positive allosteric modulator. *Science*. https://doi.org/10.1126/science.aaw8981

Liu, Xuan, Wu, W. K. K., Yu, L., Sung, J. J. Y., Srivastava, G., Zhang, S. T., and Cho, C. H. (2008). Epinephrine stimulates esophageal squamous-cell carcinoma cell proliferation via β-adrenoceptor-dependent transactivation of extracellular signal-regulated kinase/cyclooxygenase-2 pathway. *Journal of Cellular Biochemistry*. https://doi.org/10.1002/jcb.21802

Lockless, S. W., and Ranganathan, R. (1999). Evolutionarily conserved pathways of

energetic connectivity in protein families. *Science*, *286*(5438), 295–299.

Long, S., and Tian, P. (2016). Nonlinear backbone torsional pair correlations in proteins. *Scientific Reports*. https://doi.org/10.1038/srep34481

Long, S., Wang, J., *and* Tian, P. (2019). Significance of triple torsional correlations in proteins. *RSC Advances*. https://doi.org/10.1039/c9ra02191d

Lyu, P. C., Liff, M. I., Marky, L. A., and Kallenbach, N. R. (1990). Side chain contributions to the stability of alpha-helical structure in peptides. *Science*. https://doi.org/10.1126/science.2237416

MacKenzie, K. R., and Fleming, K. G. (2008). Association energetics of membrane spanning α-helices. *Current Opinion in Structural Biology*. https://doi.org/10.1016/j.sbi.2008.04.007

Mackerell, A. D. (2004). Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry*. https://doi.org/10.1002/jcc.20082

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*. https://doi.org/10.1021/jp973084f

Mackerell, A. D., Feig, M., and Brooks, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. *Journal of Computational Chemistry*. https://doi.org/10.1002/jcc.20065

MacKerell, Alexander D., Feig, M., and Brooks, C. L. (2004). Improved Treatment of the Protein Backbone in Empirical Force Fields. *Journal of the American Chemical Society*. https://doi.org/10.1021/ja036959e

Manglik, A., Kim, T. H., Masureel, M., Altenbach, C., Yang, Z., Hilger, D., et al. (2015). Structural insights into the dynamic process of β2-adrenergic receptor signaling. *Cell*. https://doi.org/10.1016/j.cell.2015.04.043

Manglik, A., and Kobilka, B. (2014). The role of protein dynamics in GPCR function: insights from the β2AR and rhodopsin. *Current Opinion in Cell Biology*, *27*, 136–143.

McCammon, J. A., Gelin, B. R., and Karplus, M. (1977). Dynamics of folded proteins. *Nature*. https://doi.org/10.1038/267585a0

McClendon, C. L., Friedland, G., Mobley, D. L., Amirkhani, H., and Jacobson, M. P. (2009). Quantifying correlations between allosteric sites in thermodynamic ensembles. *Journal of Chemical Theory and Computation*. https://doi.org/10.1021/ct9001812

Millet, O., Mittermaier, A., Baker, D., and Kay, L. E. (2003). The effects of mutations on motions of side-chains in protein L studied by 2H NMR dynamics and scalar couplings. *Journal of Molecular Biology*. https://doi.org/10.1016/S0022-2836(03)00471-6

Moitra, S., Tirupula, K. C., Klein-Seetharaman, J., and Langmead, C. (2012). A minimal ligand binding pocket within a network of correlated mutations identified by multiple sequence and structural analysis of G protein coupled receptors. *BMC Biophysics*. https://doi.org/10.1186/2046-1682-5-13

Monod, J., Wyman, J., and Changeux, J.-P. (1965). On the nature of allosteric transitions: a plausible model. *J Mol Biol*, *12*(1), 88–118.

Motlagh, H. N., Wrabl, J. O., Li, J., and Hilser, V. J. (2014). The ensemble nature of allostery. *Nature*. https://doi.org/10.1038/nature13001

Nanoff, C., Jacobson, K. A., and Stiles, G. L. (1991). The A2 adenosine receptor: Guanine nucleotide modulation of agonist binding is enhanced by proteolysis. *Molecular Pharmacology*.

Nick Pace, C., and Martin Scholtz, J. (1998). A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophysical Journal*. https://doi.org/10.1016/s0006-3495(98)77529-0

Noakes, L. (1991). The Takens embedding theorem. *International Journal of Bifurcation and Chaos*, *1*(04), 867–872.

Nussinov, R., Jang, H., and Tsai, C. J. (2014). The structural basis for cancer treatment decisions. *Oncotarget*. https://doi.org/10.18632/oncotarget.2439

Nussinov, R., and Tsai, C.-J. (2012). The Different Ways through Which Specificity Works in Orthosteric and Allosteric Drugs. *Current Drug Metabolism*. https://doi.org/10.2174/138920012799362855

Nygaard, R., Zou, Y., Dror, R. O., Mildorf, T. J., Arlow, D. H., Manglik, A., et al. (2013). The dynamic process of β2-adrenergic receptor activation. *Cell*. https://doi.org/10.1016/j.cell.2013.01.008

O'Neil, K. T., and DeGrado, W. F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*.

https://doi.org/10.1126/science.2237415

Oostenbrink, C., Villa, A., Mark, A. E., and Van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry*. https://doi.org/10.1002/jcc.20090

Oswald, C., Rappas, M., Kean, J., Doré, A. S., Errey, J. C., Bennett, K., et al. (2016). Intracellular allosteric antagonism of the CCR9 receptor. *Nature*. https://doi.org/10.1038/nature20606

Ozcan, O., Uyar, A., Doruker, P., and Akten, E. D. (2013). Effect of intracellular loop 3 on intrinsic dynamics of human β2-adrenergic receptor. *BMC Structural Biology*. https://doi.org/10.1186/1472-6807-13-29

Ozgur, C., Doruker, P., and Akten, E. D. (2016). Investigation of allosteric coupling in human β2-adrenergic receptor in the presence of intracellular loop 3. *BMC Structural Biology*. https://doi.org/10.1186/s12900-016-0061-9

Pachov, D. V., Fonseca, R., Arnol, D., Bernauer, J., and Van Den Bedem, H. (2016). Coupled Motions in β2AR:Gαs Conformational Ensembles. *Journal of Chemical Theory and Computation*. https://doi.org/10.1021/acs.jctc.5b00995

Page, R. C., Kim, S., and Cross, T. A. (2008). Transmembrane Helix Uniformity Examined by Spectral Mapping of Torsion Angles. *Structure*. https://doi.org/10.1016/j.str.2008.02.018

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–1802.

Ponder, J. W., and Case, D. A. (2003). Force fields for protein simulations. *Advances in Protein Chemistry*. https://doi.org/10.1016/S0065-3233(03)66002-X

Ponder, J. W., and Richards, F. M. (1987). Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *Journal of Molecular Biology*. https://doi.org/10.1016/0022-2836(87)90358-5

Priyakumar, U. D., Punnagai, M., Mohan, G. P. K., and Sastry, G. N. (2004). A computational study of cation-π interactions in polycyclic systems: Exploring the dependence on the curvature and electronic factors. *Tetrahedron*. https://doi.org/10.1016/j.tet.2004.01.086

Qin, K., Dong, C., Wu, G., and Lambert, N. A. (2011). Inactive-state preassembly of

Gq-coupled receptors and G q heterotrimers. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.642

Ramachandran, G. N., and Sasisekharan, V. (1968). Conformation of Polypeptides and Proteins. *Advances in Protein Chemistry*. https://doi.org/10.1016/S0065-3233(08)60402-7

Rasmussen, Søren G.F., Choi, H. J., Fung, J. J., Pardon, E., Casarosa, P., Chae, P. S., … Kobilka, B. K. (2011). Structure of a nanobody-stabilized active state of the β2 adrenoceptor. *Nature*. https://doi.org/10.1038/nature09648

Rasmussen, Søren G.F., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., … Kobilka, B. K. (2007). Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature*. https://doi.org/10.1038/nature06325

Rasmussen, Sŏren G.F., Devree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., et al. (2011). Crystal structure of the β 2 adrenergic receptor-Gs protein complex. *Nature*. https://doi.org/10.1038/nature10361

Rebois, R. V., *and* Hébert, T. E. (2003). Protein complexes involved in heptahelical receptor-mediated signal transduction. *Receptors and Channels*. https://doi.org/10.1080/10606820308243

Reynolds, K. A., Katritch, V., and Abagyan, R. (2009). Identifying conformational changes of the β2 adrenoceptor that enable accurate prediction of ligand/receptor interactions and screening for GPCR modulators. *Journal of Computer-Aided Molecular Design*. https://doi.org/10.1007/s10822-008-9257-9

Ring, A. M., Manglik, A., Kruse, A. C., Enos, M. D., Weis, W. I., Garcia, K. C., and Kobilka, B. K. (2013). Adrenaline-activated structure of β 2-adrenoceptor stabilized by an engineered nanobody. *Nature*. https://doi.org/10.1038/nature12572

Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G. F., Foon, S. T., Kobilka, T. S., et al. (2007). GPCR engineering yields high-resolution structural insights into β2-adrenergic receptor function. *Science*. https://doi.org/10.1126/science.1150609

Rosenbaum, D. M., Rasmussen, S. G. F., and Kobilka, B. K. (2009). The structure and function of G-protein-coupled receptors. *Nature*, *459*(7245), 356.

Rosenbaum, D. M., Zhang, C., Lyons, J. A., Holl, R., Aragao, D., Arlow, D. H., et al. (2011). Structure and function of an irreversible agonist-β2 adrenoceptor complex. *Nature*. https://doi.org/10.1038/nature09665

Roulston, M. S. (1999). Estimating the errors on measured entropy and mutual information. *Physica D: Nonlinear Phenomena*. https://doi.org/10.1016/S0167-2789(98)00269-3

Samama, P., Bond, R. A., Rockman, H. A., Milano, C. A., and Lefkowitz, R. J. (1997). Ligand-induced overexpression of a constitutively active β2-adrenergic receptor: Pharmacological creation of a phenotype in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.94.1.137

Sauer, T., Yorke, J. A., and Casdagli, M. (1991). Embedology. *Journal of Statistical Physics*. https://doi.org/10.1007/BF01053745

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*. https://doi.org/10.1103/PhysRevLett.85.461

Shang, Z. J., Liu, K., and Liang, D. F. (2009). Expression of β2-adrenergic receptor in oral squamous cell carcinoma. *Journal of Oral Pathology and Medicine*. https://doi.org/10.1111/j.1600-0714.2008.00691.x

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Sogunmez, N., and Akten, E. D. (2019). Intrinsic Dynamics and Causality in Correlated Motions Unraveled in Two Distinct Inactive States of Human β 2 -Adrenergic Receptor. *Journal of Physical Chemistry B*. https://doi.org/10.1021/acs.jpcb.9b01202

Sogunmez, N., and Akten, E. D. (2020). Distinctive Communication Networks in Inactive States of β 2 -Adrenergic Receptor: Mutual Information and Entropy Transfer Analysis. *Proteins: Structure, Function, and Bioinformatics*. https://doi.org/10.1002/prot.25965

Song, K., Liu, X., Huang, W., Lu, S., Shen, Q., Zhang, L., and Zhang, J. (2017). Improved Method for the Identification and Validation of Allosteric Sites. *Journal of Chemical Information and Modeling*. https://doi.org/10.1021/acs.jcim.7b00014

Strader, C. D., Sigal, I. S., and Dixon, R. A. F. (1989). Structural basis of β-adrenergic receptor function. *FASEB Journal*. https://doi.org/10.1096/fasebj.3.7.2541037

Strader, C. D., Sigal, I. S., Register, R. B., Candelore, M. R., Rands, E., and Dixon, R. A. (1987). Identification of residues required for ligand binding to the β-adrenergic receptor. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.84.13.4384

Strader, C. D., Sigal, I. S., Rios Candelore, M., Rands, E., Hill, W. S., and Dixon, R. A. F. (1988). Conserved aspartic acid residues 79 and 113 of the β-adrenergic receptor have different roles in receptor function. *Journal of Biological Chemistry*.

Strickler, S. S., Gribenko, A. V., Gribenko, A. V., Keiffer, T. R., Tomlinson, J., Reihle, T., et al. (2006). Protein stability and surface electrostatics: A charged relationship. *Biochemistry*. https://doi.org/10.1021/bi0600143

Süel, G. M., Lockless, S. W., Wall, M. A., and Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Structural Biology*. https://doi.org/10.1038/nsb881

Suryanarayana, S., Daunt, D. A., Von Zastrow, M., and Kobilka, B. K. (1991). A point mutation in the seventh hydrophobic domain of the α2 adrenergic receptor increases its affinity for a family of β receptor antagonists. *Journal of Biological Chemistry*.

Suryanarayana, S., and Kobilka, B. K. (1993). Amino acid substitutions at position 312 in the seventh hydrophobic segment of the β2-adrenergic receptor modify ligand-binding specificity. *Molecular Pharmacology*.

Sygula, A., Fronczek, F. R., Sygula, R., Rabideau, P. W., and Olmstead, M. M. (2007). A double concave hydrocarbon buckycatcher. *Journal of the American Chemical Society*. https://doi.org/10.1021/ja070616p

Theiler, J. (1986). Spurious dimension from correlation algorithms applied to limited time-series data. *Physical Review A*. https://doi.org/10.1103/PhysRevA.34.2427

Vaidehi, N., and Bhattacharya, S. (2016). Allosteric communication pipelines in G-protein-coupled receptors. *Current Opinion in Pharmacology*. https://doi.org/10.1016/j.coph.2016.07.010

van der Kamp, M. W., Schaeffer, R. D., Jonsson, A. L., Scouras, A. D., Simms, A. M., Toofanny, R. D., et al. (2010). Dynameomics: A Comprehensive Database of Protein Dynamics. *Structure*. https://doi.org/10.1016/j.str.2010.01.012

Vasudev, P. G., Banerjee, M., Ramakrishnan, C., and Balaram, P. (2012). Asparagine and glutamine differ in their propensities to form specific side chain-backbone hydrogen bonded motifs in proteins. *Proteins: Structure, Function and Bioinformatics*. https://doi.org/10.1002/prot.24001

Verlet, L. (1967). Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review*. https://doi.org/10.1103/PhysRev.159.98

VON HEIJNE, G. (1981). Membrane Proteins: The Amino Acid Composition of Membrane-Penetrating Segments. *European Journal of Biochemistry*. https://doi.org/10.1111/j.1432-1033.1981.tb05700.x

Wand, A. J. (2001). Dynamic activation of protein function: A view emerging from NMR spectroscopy. *Nature Structural Biology*. https://doi.org/10.1038/nsb1101-926

Wand, A. Joshua, and Sharp, K. A. (2018). Measuring Entropy in Molecular Recognition by Proteins. *Annual Review of Biophysics*. https://doi.org/10.1146/annurev-biophys-060414-034042

Wenthur, C. J., Gentry, P. R., Mathews, T. P., and Lindsley, C. W. (2014). Drugs for allosteric sites on receptors. *Annual Review of Pharmacology and Toxicology*, *54*, 165–184.

Yang, J. S., Seo, S. W., Jang, S., Jung, G. Y., and Kim, S. (2012). Rational engineering of enzyme allosteric regulation through sequence evolution analysis. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1002612

Yokota, K., Satou, K., and Ohki, S. ya. (2006). Comparative analysis of protein thermostability: Differences in amino acid content and substitution at the surfaces and in the core regions of thermophilic and mesophilic proteins. *Science and Technology of Advanced Materials*. https://doi.org/10.1016/j.stam.2006.03.003

Zhang, L., Centa, T., and Buck, M. (2014). Structure and dynamics analysis on plexin-B1 Rho GTPase binding domain as a monomer and dimer. *Journal of Physical Chemistry B*. https://doi.org/10.1021/jp503668k

Zhang, P., Leger, A. J., Baleja, J. D., Rana, R., Corlin, T., Nguyen, N., et al. (2015). Allosteric activation of a G protein-coupled receptor with cell-penetrating receptor mimetics. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M115.636316

Zhang, T., Faraggi, E., and Zhou, Y. (2010). Fluctuations of backbone torsion angles obtained from NMR-determined structures and their prediction. *Proteins: Structure, Function and Bioinformatics*. https://doi.org/10.1002/prot.22842

Zheng, Y., Qin, L., Zacarías, N. V. O., De Vries, H., Han, G. W., Gustavsson, M., et al. (2016). Structure of CC chemokine receptor 2 with orthosteric and allosteric antagonists. *Nature*. https://doi.org/10.1038/nature20605

Zhou, X. E., He, Y., de Waal, P. W., Gao, X., Kang, Y., Van Eps, N., et al. (2017). Identification of Phosphorylation Codes for Arrestin Recruitment by G Protein-Coupled Receptors. *Cell*. https://doi.org/10.1016/j.cell.2017.07.002

CURRICULUM VITAE

Personal Information
Name Surname : NURAY SÖĞÜNMEZ ERDOĞAN
Place and Date of Birth: Karacabey, Bursa Turkey, 10 September 1985

Education
Undergraduate Education : Bogazici University (September 2004-June 2010)
Graduate Education : Bogazici University (September 2010-February 2013)
Foreign Language Skills : English (Advanced), Turkish (Native),
Italian (Beginner), German (Beginner)

Work Experiences
Name of Employer and Dates of Employment: **Health Institutes in Turkey (TÜSEB),**
December 2019-To Date

Acibadem Maslak Hospital-
Comprehensive Spine Center
August 2015- April 2017

Sentromer DNA LLC.
November 2013- August 2015

SACEM Life Technologies LLC.
August 2012-July 2013

Pfizer Inc.
January 2010-September 2010

Contact:
Telephone : +905322051507
E-mail Address : nuray.sogunmez@gmail.com

Publications

1. **Nuray Sogunmez**, E. Demet Akten (2020). Distinctive Allosteric Communication Networks in Inactive States of β2-Adrenergic Receptor: Mutual Information and Entropy Transfer Analysis. *PROTEINS: Structure, Function, and Bioinformatics Online ahead of print*, DOIMer

2. Caglar Yilgor, Irem N. Eroglu, **Nuray Sogunmez,** Kadir Abul, Ahmet Alanay (2019). Diagnosis and Treatment of Congenital Scoliosis. *Acibadem Universitesi Saglık Bilimleri Dergisi, 10(2)*

3. **Nuray Sogunmez,** E. Demet Akten (2019). Intrinsic Dynamics and Causality in Correlated Motions Unraveled in Two Distinct Inactive States of Human β2 – Adrenergic Receptor (β2AR). *J of Physical Chemistry B, 123(17), 3630-3642.*

4. Caglar Yilgor, **Nuray Sogunmez,** Yasemin Yavuz, Berk Baris Ozmen, Ibrahim Obeid, Frank S. Kleinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, Essg European Spine Study Group (ESSG) (2018). Global Alignment and Proportion (GAP) Score Better Correlates to HRQoL Scores and Better Predicts Mechanical Complications Compared to Schwab Sagittal Modifiers. *The Spine Journal, 17(10), S156*

5. Caglar Yilgor, Yasemin Yavuz, **Nuray Sogunmez**, Sleiman Haddad, Anne F. Mannion, Kadir Abul, Louis Boissiere, Ibrahim Obeid, Frank Kleinstück, Francisco Javier Sanches Perez-Grueso, Emre Acaroglu, Ferran Pellise, Ahmet Alanay, The European Spine Study Group (ESSG) (2018). Relative Pelvic Version (RPV): An Individualized Pelvic Incidence-Based Proportional Parameter That Quantifies Pelvic Version More Precisely Than Pelvic Tilt. *The Spine Journal, 18(10), 1787-1797*

6. Caglar Yilgor, **Nuray Sogunmez**, Lois Boissiere, Yasemin Yavuz, Ibrahim Obeid, Frank S. Kleinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Sleiman Haddad, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, European Spine Study Group (ESSG) (2017). Relative Lumbar Lordosis And Lordosis Distribution Index: Individualized Pelvic Incidence– Based Proportional Parameters That Quantify Lumbar Lordosis More Precisely Than The Concept Of Pelvic Incidence Minus Lumbar Lordosis. *Journal of Neurosurgery, 43(6), E5*

7. Caglar Yilgor, **Nuray Sogunmez**, Lois Boissiere, Yasemin Yavuz, Ibrahim Obeid, Frank S. Kleinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Sleiman Haddad, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, Essg European Spine Study Group (ESSG) (2017). Global Alignment and Proportion (GAP) Score: Development and Validation of a New Method of Analyzing Spinopelvic Alignment to Predict Mechanical Complications After Adult Spinal Deformity Surgery. *The Journal of Bone and Joint Surgery, 99(19):1661-1672*

8. Ahmet Alanay, Caglar Yilgor, Gokhan Ergene, **Nuray Sogunmez**, Barbaros Omer Cebeci (2017). Anterior Vertebral Body Tethering: A Promising Non-Fusion Scoliosis Treatment. *Spinal Surgery News*

Oral Presentations/ Posters

1. **Nuray Sogunmez**, E. Demet Akten. *Allosteric Communications in Inactive States of Human* β_2 Adrenergic Receptor (β2-AR).

- *The International Symposium on Health Informatics and Bioinformatics 2019 (HIBIT-2019)*. (Oral Presentation). (Travel Grant)

2. **Nuray Sogunmez**, E. Demet Akten. Protein Allostery in Human β_2 Adrenergic Receptor and Information Transfer Between Residues.

- *1st International Symposium on Graduate Research in Science 2018, Istanbul, Turkey.* (Poster Presentation)

3. Caglar Yilgör, **Nuray Sogunmez**, Yasemin Yavuz, Sleiman Haddad, Ibrahim Obeid, Frank Kelinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, European Spine Study Group (ESSG). Relative Pelvic Version (RPV): A new PI-based Proportional Parameter that Quantifies Pelvic Version More Precisely Compared to PT.

- *31th Turkish Neurosurgery Scientific Congress 2017, Antalya, Turkey* (Oral Presentation)
- *12th International Turkish Spine Association Scientific Congress 2017*, Antalya, Turkey (Oral Presentation)
- *Eurospine Congress 2017, Dublin, Ireland* (Quick Fire Presentation)

4. Caglar Yilgör, **Nuray Sogunmez**, Yasemin Yavuz, Louis Boissiere, Ibrahim Obeid, Frank Kleinstuek, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, Essg European Spine Study Group (ESSG). Relative Lumbar Lordosis (RLL): A new PI-based Proportional Parameter that Quantifies Lumbar Lordosis More Precisely Compared to PI-LL Concept.

- *31th Turkish Neurosurgery Scientific Congress 2017, Antalya, Turkey* (Oral Presentation)
- *12th International Turkish Spine Association Scientific Congress 2017, Antalya,Turkey* (Oral Presentation)
- *North American Spine Society Congress, 2017, Orlando, Florida, USA (E-poster Presentation)*
- *Eurospine Congress 2017, Dublin, Ireland* (E-poster Presentation)

5. Caglar Yilgor, **Nuray Sogunmez,** Yasemin Yavuz, Louis Boissiere, Ibrahim Obeid, Frank Kleinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, European Spine Study Group (ESSG). Relative Spinopelvic Alignment (RSA): A new PI-based Proportional Parameter that Quantifies Sagittal Alignment More Precisely Compared to SVA.

- *31th Turkish Neurosurgery Scientific Congress 2017, Antalya, Turkey* (Oral Presentation)
- *12th International Turkish Spine Association Scientific Congress 2017, Antalya,Turkey* (Oral Presentation)
- *Eurospine Congress 2017, Dublin, Ireland* (E-poster Presentation)

6. Caglar Yilgor, **Nuray Sogunmez**, Yasemin Yavuz, Can Berk Aşaroğlu, Ibrahim Obeid, Frank S. Kleinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, European Spine Study Group (ESSG). Global Alignment and Proportion (GAP) Score: Development and Validation of a New Method of Analyzing Spinopelvic Alignment to Predict Mechanical Complications after Adult Spinal Deformity Surgery.

- *31th Turkish Neurosurgery Scientific Congress 2017, Antalya, Turkey* (Oral Presentation) (Best Presentation Award)
- *12th International Turkish Spine Association Scientific Congress 2017, Antalya,Turkey* (Oral Presentation) (Best Presentation Award)
- *North American Spine Society Congress 2017, Orlando, Florida, USA* (Oral Presentation)
- *Eurospine Congress 2017*, *Dublin, Ireland* (Quick Fire Presentation)
- *Scoliosis Research Society Congress 2017, Philedelphia, USA* (E-poster Presentation)
- *The International Society for the Study of the Lumbar Spine Congress 2017, Athens, Greece* (Oral Presentation)

7. Caglar Yilgor, **Nuray Sogunmez**, Yasemin Yavuz, Berk Baris Ozmen, İbrahim Obeid, Frank S. Kleinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, European Spine Study Group (ESSG). Global Alignment and Proportion (GAP) Score Better Correlates to HRQoL Scores and Better Predicts Mechanical Complications Compared to SRS-Schwab Sagittal Modifiers.

- *31th Turkish Neurosurgery Scientific Congress 2017, Antalya, Turkey* (Oral Presentation) (Best Presentation Award)
- *12th International Turkish Spine Association Scientific Congress 2017, Antalya,Turkey* (Oral Presentation) (Best Presentation Award)
- *North American Spine Society Congress 2017, Orlando, Florida, USA* (Oral Presentation)
- *Eurospine Congress 2017, Dublin, Ireland* (Oral Presentation)
- *International Meeting on Advanced Spine Techniques Congress 2017, Cape Town, South Africa* (Oral Presentation)

8. Caglar Yilgor, **Nuray Sogunmez**, Yasemin Yavuz, İbrahim Obeid, Frank S. Kleinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, European Spine Study Group (ESSG). The Appropriate Sagittal Plane Correction to Individualized Proportionate Shape and Alignment Does Not Necessarily Mean More Complex Surgery in Adult Spinal Deformity.

- *12th International Turkish Spine Association Scientific Congress 2017, Antalya, Turkey* (Oral Presentation)
- *Eurospine Congress 2017, Dublin, Ireland* (Oral Presentation)

9. Caglar Yilgor, **Nuray Sogunmez**, Yasemin Yavuz, İbrahim Obeid, Frank S. Kleinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, European Spine Study Group (ESSG).

Failure to Validate the Age-Adjusted Alignment Thresholds Concept in an Adult Spinal Deformity Database.

- *12th International Turkish Spine Association Scientific Congress 2017, Antalya, Turkey* (Oral Presentation)
- *Eurospine Congress 2017, Dublin, Ireland* (Quick-Fire Presentation)
- *International Meeting on Advanced Spine Techniques Congress 2017, Cape Town, South Africa* (Whitecloud Award Nominee) (Oral Presentation)

10. Caglar Yilgor, **Nuray Sogunmez**, Yasemin Yavuz, İbrahim Obeid, Frank S. Kleinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, European Spine Study Group (ESSG). Sagittal Realignment Goals Should Be Set to Ideal Proportionate Shape and Alignment Independent of Age.

- *12th International Turkish Spine Association Scientific Congress 2017, Antalya, Turkey* (Oral Presentation)
- *Eurospine Congress 2017, Dublin, Ireland* (Oral Presentation)
- *Scoliosis Research Society Congress 2017, Philedelphia, USA* (Oral Presentation)

11. Caglar Yilgor, **Nuray Sogunmez,** Mohamed Dalla, Gülden Demirci Otluoglu, Murat Basarır, Ahmet Alanay, Memet Özek. Safety and Efficacy of Concurrent Spinal Deformity and Intra-Dural Surgery: Case Series.

- *31th Turkish Neurosurgery Scientific Congress 2017, Antalya, Turkey* (Oral Presentation)
- *12th International Turkish Spine Association Scientific Congress 2017, Antalya,Turkey* (Oral Presentation)
- *Eurospine Congress 2017, Dublin, Ireland* (E-poster Presentation)

12. Caglar Yilgor, **Nuray Sogunmez**, Peri Kindan, Binnaz Ay, Ahmet Alanay. Use of a Tailored Multimodal Recovery Pathway Accelerates Inpatient Recovery.

- *12th International Turkish Spine Association Scientific Congress 2017, Antalya,Turkey* (Oral Presentation)
- *Eurospine Congress 2017, Dublin, Ireland* (E-poster Presentation)

13. Barbaros Omer Cebeci, Gokhan Ergene, **Nuray Sogunmez,** Binnaz Ay, Caglar Yilgor, Ahmet Alanay. Non-fusion Growth Modulation with Anterior Vertebral Body Tethering (VBT) for Adolescent Idiopathic Scoliosis.

- *12th International Turkish Spine Association Scientific Congress 2017, Antalya,Turkey* (Oral Presentation)
- *31th Turkish Neurosurgery Scientific Congress 2017, Antalya, Turkey* (Oral Presentation)
- *Eurospine Congress 2017, Dublin, Ireland* (Oral Presentation)
- *North American Spine Society Congress 2017, Orlando, Florida, USA* (Oral Presentation)
- *The International Scientific Society on Scoliosis Orthopaedic and Rehabilitation Treatment 2017, Lyon, France* (Oral Presentation)

14. Caglar Yilgor, **Nuray Sogunmez**, Erdal Cosgun, İbrahim Obeid, Frank S. Kleinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, European Spine Study Group (ESSG). Critical Age Treshold(s) for Surgical Complexity and Complication Rates in Adult Idiopatic Scoliosis.
- *Eurospine Congress 2016, Berlin, Germany* (E-poster Presentation)

15. Caglar Yilgor, **Nuray Sogunmez**, Erdal Cosgun, İbrahim Obeid, Frank S. Kleinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, European Spine Study Group (ESSG). Global Alignment and Proportion (GAP) Score Predics Mechanical Failure in Adult Spinal Deformity Surgery.

- *Eurospine Congress 2016, Berlin, Germany* (E-poster Presentation)

16. Caglar Yilgor, **Nuray Sogunmez**, Erdal Cosgun, İbrahim Obeid, Frank S. Kleinstueck, Francisco Javier Sanchez Perez Grueso, Emre R. Acaroğlu, Anne F. Mannion, Ferran Pellisé, Ahmet Alanay, European Spine Study Group (ESSG). Which Factors Influence Surgery vs. Non-Surgery Decision for AIS Patients with Borderline (40^o - 55^o) Main Thoracic Curves?

- *Eurospine Congress 2016, Berlin, Germany* (E-poster Presentation)

17. Xalid Bayramli, **Nuray Sogunmez**, Stefan H. Fuss. Characterization of Zebrafish Class II OR Genes.

- *ZFIN 2011, Edinburgh, Scotland* (E-poster Presentation)

APPENDIX A

```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################

# Minimization and Equilibration of 3sn6_Model and G protein
#embedded in POPC membrane, ions and water.
# COMMENT ON YOUR SYSTEM HERE


#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure          protein_ionized.psf
coordinates        protein_ionized.pdb
set outputname     protein_product_01

set temperature    310

# Continuing a job from the restart files
if {1} {
set inputname      protein_eq_01
binCoordinates     $inputname.restart.coor
binVelocities      $inputname.restart.vel  ;# remove the "temperature" entry if you use
this!
extendedSystem     $inputname.restart.xsc
}

firsttimestep      1000000


#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input
paraTypeCharmm       on
parameters          par_all36m_prot.prm   ;# you may need to specify a path
parameters          par_all36_lipid.prm   ;# you may need to specify a path
parameters          toppar_water_ions_namd.str ;# you may need to specify a path
parameters          toppar_all36_carb_glycopeptide.str;# you may need to specify a path
parameters          par_all36_cgenff.prm;# you may need to specify a path
parameters          par_all36_na.prm;# you may need to specify a path
```

```
parameters          par_all36_carb.prm;# you may need to specify a path
#temperature         $temperature

# Periodic Boundary Conditions
if {0} {
cellBasisVector1    123.     0.     0.
cellBasisVector2      0.   124.     0.
cellBasisVector3      0.     0.   165.
cellOrigin          14.52 14.74 29.25
}
wrapWater           on
wrapAll             on

# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff              12.0
switching           on
switchdist          10.
pairlistdist        13.5

# Integrator Parameters
timestep            2.0  ;# 2fs/step
rigidBonds          all  ;# needed for 2fs steps
nonbondedFreq       1
fullElectFrequency  2
stepspercycle       20

#PME (for full-system periodic electrostatics)
if {1} {
PME                 yes
PMEGridSpacing     1.0
PMEGridSizeX       125
PMEGridSizeY       125
PMEGridSizeZ       165
}

# Constant Temperature Control
if {1} {
langevin            on    ;# do langevin dynamics
langevinDamping     1     ;# damping coefficient (gamma) of 5/ps
langevinTemp        $temperature
}

# Constant Pressure Control
if {1} {
useGroupPressure      yes ;# needed for 2fs steps
useFlexibleCell       yes  ;# no for water box, yes for membrane
```

```
useConstantArea       yes  ;# no for water box, yes for membrane

langevinPiston        on
langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod  200.
langevinPistonDecay   50.
langevinPistonTemp    $temperature
}


# Output
outputName          $outputname

restartfreq         10000     ;# 10000steps = every 2ps
dcdfreq             10000
outputEnergies      10000
outputPressure      10000

# Fixed Atoms Constraint (set PDB beta-column to 1)
if {0} {
fixedAtoms          on
fixedAtomsFile      protein_ionized.fix
fixedAtomsCol       B
fixedAtomsForces    on
}

#############################################################
## EXTRA PARAMETERS                                        ##
#############################################################

# Put here any custom parameters that are specific to
# this job (e.g., SMD, TclForces, etc...)

#eFieldOn yes
#eField 0 0 -0.155

#############################################################
## EXECUTION SCRIPT                                        ##
#############################################################
# For heating, use temperature reassignment.

# Minimization
if {0} {
minimize            1000
reinitvels          $temperature
}
run 750000000;
```

APPENDIX B

B.1 Tcl Code to Calculate the Dihedral Angles

```
source dihedral_angles_atom_names.tcl
proc get_backbone_dihedral_indices {resid resname chain mol} {
  set ressel [atomselect $mol "chain $chain and resid $resid and name CA"]
  set residue_index [$ressel get residue]
  $ressel delete
  set prev_residue_index [expr $residue_index-1]
  set next_residue_index [expr $residue_index+1]
  #Nitrogen
  set atomsel [atomselect $mol "chain $chain and residue $residue_index and name
\"N\""]
  set Ni [$atomsel get index]
  $atomsel delete
  #CA
  set atomsel [atomselect $mol "chain $chain and residue $residue_index and name
\"CA\""]
  set CAi [$atomsel get index]
  $atomsel delete
  #C(O)
  set atomsel [atomselect $mol "chain $chain and residue $residue_index and name
\"C\""]
  set Ci [$atomsel get index]
  $atomsel delete
  #if prevous residue exists
  if {$prev_residue_index>=0} then {
    #previous C(O)
    set atomsel [atomselect $mol "chain $chain and residue $prev_residue_index and
name \"C\""]
    set pCi [$atomsel get index]
    $atomsel delete
    lappend phi $pCi $Ni $CAi $Ci
  } else  {
    #set phi {}
    set phi 0.0
  }
#next N
  set atomsel [atomselect $mol "chain $chain and residue $next_residue_index and
name \"N\""]
  set nNi [$atomsel get index]
  $atomsel delete
  #next CA
  set atomsel [atomselect $mol "chain $chain and residue $next_residue_index and
name \"CA\""]
  set nCAi [$atomsel get index]
```

```
  $atomsel delete


  #if next residue exists
  if {$nNi>0} then {
    lappend psi $Ni $CAi $Ci $nNi
    lappend omega $CAi $Ci $nNi $nCAi
  } else  {
   # set psi {}
    #set omega {}
    set psi 0.0
    set omega 0.0
  }
  lappend result $phi $psi $omega
  return $result
}
proc get_sidechain_inidces {resid resname chain mol} {
  lappend result $resid $resname $chain
  lappend result [get_backbone_dihedral_indices $resid $resname $chain $mol]
  #puts $result
  switch $resname {
    GLY {lappend result [get_GLY_sidechain_indicies $resid $resname $chain $mol]}
    ALA {lappend result [get_ALA_sidechain_indicies $resid $resname $chain $mol]}
    SER {lappend result [get_SER_sidechain_indicies $resid $resname $chain $mol]}
    CYS {lappend result [get_CYS_sidechain_indicies $resid $resname $chain $mol]}
    VAL {lappend result [get_VAL_sidechain_indicies $resid $resname $chain $mol]}
    THR {lappend result [get_THR_sidechain_indicies $resid $resname $chain $mol]}
    ILE {lappend result [get_ILE_sidechain_indicies $resid $resname $chain $mol]}
    PRO {lappend result [get_PRO_sidechain_indicies $resid $resname $chain $mol]}
    MET {lappend result [get_MET_sidechain_indicies $resid $resname $chain $mol]}
    ASP {lappend result [get_ASP_sidechain_indicies $resid $resname $chain $mol]}
    ASN {lappend result [get_ASN_sidechain_indicies $resid $resname $chain $mol]}
    LEU {lappend result [get_LEU_sidechain_indicies $resid $resname $chain $mol]}
    LYS {lappend result [get_LYS_sidechain_indicies $resid $resname $chain $mol]}
    GLU {lappend result [get_GLU_sidechain_indicies $resid $resname $chain $mol]}
    GLN {lappend result [get_GLN_sidechain_indicies $resid $resname $chain $mol]}
    ARG {lappend result [get_ARG_sidechain_indicies $resid $resname $chain $mol]}
    HIS {lappend result [get_HIS_sidechain_indicies $resid $resname $chain $mol]}
    HSE {lappend result [get_HIS_sidechain_indicies $resid $resname $chain $mol]}
    HSP {lappend result [get_HIS_sidechain_indicies $resid $resname $chain $mol]}
    HSD {lappend result [get_HIS_sidechain_indicies $resid $resname $chain $mol]}
    PHE {lappend result [get_PHE_sidechain_indicies $resid $resname $chain $mol]}
    TYR {lappend result [get_TYR_sidechain_indicies $resid $resname $chain $mol]}
    TRP {lappend result [get_TRP_sidechain_indicies $resid $resname $chain $mol]}
    UNK {lappend result [get_UNK_sidechain_indicies $resid $resname $chain $mol]}
    CMT {lappend result [get_CMT_sidechain_indicies $resid $resname $chain $mol]}
    CMTS {lappend result [get_CMT_sidechain_indicies $resid $resname $chain $mol]}
    }
```

```
   return $result
}
proc get_all_sidechain_indices {resinfos chain mol} {
  set result {}
  foreach {resid resname} $resinfos {
    # puts  "$resid $resname $chain $mol"
     lappend result [get_sidechain_inidces $resid $resname $chain $mol]
  }
  return $result
}
proc print_all_sidechain_dihedrals {indices frame_num output_file} {
#  puts $indices
  foreach {resid resname chain backbone_dihedrals dihedrals} $indices {
  #  puts "| $resid $resname $chain $backbone_dihedrals $dihedrals|"

    set phi_indices [lindex $backbone_dihedrals 0]
    set psi_indices [lindex $backbone_dihedrals 1]
    #set omega_inidces [lindex backbone_dihedrals 2/]
    #first phi doesnt exist
     if {[llength $phi_indices]} then {
       set phi [measure dihed $phi_indices]
     } else {
      set phi 0.0
    }
    #last psi doesnt exist
    if {[llength $psi_indices]} then {
      set psi [measure dihed $psi_indices]
    } else {
      set psi 0.0
    }
    set outstring [format "%4.d%4.4s%9.4d phi %6.2f psi %6.1f" $resid $resname
$frame_num $phi $psi]
    #if not empty list
    if {[llength $dihedrals]} {
      foreach dihedral $dihedrals {
        set val [measure dihed $dihedral]
        set val_str [format " dihedrals %6.1f" $val]
        append outstring $val_str
      } ;#foreach dihedral
    } ;# if not empty list
    puts $output_file $outstring
  } ;#foreach residue
}
puts "calculate dihedrals 1.0"
puts "proc print_sidechain_dihedrals {chain residues {mol top} {output_file_name
\"stdout\"} {first_frame 0} {last_frame -1} {stride 1} {print_progress 0}}"
proc print_sidechain_dihedrals {chain residues {mol top} {output_file_name "stdout"}
{first_frame 0} {last_frame -1} {stride 1} {print_progress 0}} {
```

```
  #get residue list
  set residues [atomselect $mol "chain $chain and $residues and name CA"]
  set resinfos [join [$residues get {resid resname}]]
  #puts $resinfos
  if {($print_progress>0) && ($output_file_name != "stdout")} then {
      set counter [expr [llength $resinfos]/2]
      puts "Getting indices for $counter residues."
     }
  #get the atom indices of the dehidral (torsion) angles
  set indices [join [get_all_sidechain_indices $resinfos $chain $mol]]
  #puts $indices
  if {($print_progress>0) && ($output_file_name != "stdout")} then {
      set counter [expr [llength $resinfos]/2]
      puts "Done."
     }
#set output
  if {$output_file_name == "stdout"} then {
    set output_file stdout
  } else {
    set output_file [open $output_file_name w]}
  #get current mol frame
  set old_frame [molinfo $mol get frame]
  #don't update display
  display update off
  #default last_frame is the number of all frames
  if {$last_frame == -1} then {
    set last_frame [expr [molinfo $mol get numframes] -1]}
  set counter 0
  #iterate along through all frames
  for {set frame $first_frame} {$frame <= $last_frame} {incr frame $stride} {
     incr counter
     if {($print_progress>0) && ($counter>=$print_progress) && ($output_file_name !=
"stdout")} then {
       puts "Working on frame $frame."
       set counter 0
     }
     molinfo $mol set frame $frame
     print_all_sidechain_dihedrals $indices $frame $output_file
  }
  #restore frame
  molinfo $mol set frame $old_frame
  #restore update
  display update on
  if {$output_file_name != "stdout"} then {
  close $output_file}
  $residues delete
}
```

B.2 Dihedral Angle Atom Names

```
proc get_GLY_sidechain_indicies {resid resname chain mol} {
    set result {}

    return $result
  }

proc get_ALA_sidechain_indicies {resid resname chain mol} {
    set result {}

    return $result
  }

proc get_SER_sidechain_indicies {resid resname chain mol} {
    set result {}

     #do angle Chi1
     #

     set angle_list {}

       set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
       lappend angle_list [$atomsel get index]
       $atomsel delete

       set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
       lappend angle_list [$atomsel get index]
       $atomsel delete

       set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
       lappend angle_list [$atomsel get index]
       $atomsel delete

       set atomsel [atomselect $mol "chain $chain and resid $resid and name \"OG\""]
       lappend angle_list [$atomsel get index]
       $atomsel delete

     lappend result $angle_list

    return $result
  }

proc get_CYS_sidechain_indicies {resid resname chain mol} {
    set result {}

     #do angle Chi1
     #
```

```
    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"SG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

  return $result
}

proc get_VAL_sidechain_indicies {resid resname chain mol} {
  set result {}

    #do angle Chi1
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG1\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete
```

```
      lappend result $angle_list

    return $result
  }

proc get_THR_sidechain_indicies {resid resname chain mol} {
    set result {}

      #do angle Chi1
      #

      set angle_list {}

        set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
        lappend angle_list [$atomsel get index]
        $atomsel delete

        set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
        lappend angle_list [$atomsel get index]
        $atomsel delete

        set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
        lappend angle_list [$atomsel get index]
        $atomsel delete

        set atomsel [atomselect $mol "chain $chain and resid $resid and name \"OG1\""]
        lappend angle_list [$atomsel get index]
        $atomsel delete

      lappend result $angle_list

    return $result
  }

proc get_ILE_sidechain_indicies {resid resname chain mol} {
    set result {}

      #do angle Chi1
      #

      set angle_list {}

        set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
        lappend angle_list [$atomsel get index]
        $atomsel delete

        set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
        lappend angle_list [$atomsel get index]
```

```
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG1\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

    #do angle Chi2
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG1\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD.?\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

  return $result
}

proc get_PRO_sidechain_indicies {resid resname chain mol} {
  set result {}

    #do angle Chi1
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
```

```
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

lappend result $angle_list

#do angle Chi2
#

set angle_list {}

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

lappend result $angle_list

#do angle Chi3
#

set angle_list {}

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete
```

```
    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

   lappend result $angle_list

  return $result
 }

proc get_MET_sidechain_indicies {resid resname chain mol} {
  set result {}

   #do angle Chi1
   #

   set angle_list {}

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

   lappend result $angle_list

   #do angle Chi2
   #

   set angle_list {}
```

```
      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"SD\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

    #do angle Chi3
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"SD\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CE\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

  return $result
}

proc get_ASP_sidechain_indicies {resid resname chain mol} {
  set result {}

    #do angle Chi1
```

```
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

    #do angle Chi2
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"OD1\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

  return $result
}
```

```
proc get_ASN_sidechain_indicies {resid resname chain mol} {
    set result {}

     #do angle Chi1
     #

     set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

     lappend result $angle_list

     #do angle Chi2
     #

     set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"OD1\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

     lappend result $angle_list
```

```
    return $result
  }

proc get_LEU_sidechain_indicies {resid resname chain mol} {
    set result {}

     #do angle Chi1
     #

     set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

     lappend result $angle_list

     #do angle Chi2
     #

     set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD1\""]
```

```
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

  return $result
}

proc get_LYS_sidechain_indicies {resid resname chain mol} {
  set result {}

    #do angle Chi1
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

    #do angle Chi2
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
```

```
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

lappend result $angle_list

#do angle Chi3
#

set angle_list {}

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CE\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

lappend result $angle_list

#do angle Chi4
#

set angle_list {}

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CE\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete
```

```
      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"NZ\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

  return $result
}

proc get_GLU_sidechain_indicies {resid resname chain mol} {
  set result {}

    #do angle Chi1
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

    #do angle Chi2
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete
```

```
      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

    #do angle Chi3
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"OE1\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

  return $result
}

proc get_GLN_sidechain_indicies {resid resname chain mol} {
  set result {}

    #do angle Chi1
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete
```

```
  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

lappend result $angle_list

#do angle Chi2
#

set angle_list {}

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

lappend result $angle_list

#do angle Chi3
#

set angle_list {}

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
```

```
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"OE1\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

   lappend result $angle_list

  return $result
 }

proc get_ARG_sidechain_indicies {resid resname chain mol} {
  set result {}

   #do angle Chi1
   #

   set angle_list {}

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

   lappend result $angle_list

   #do angle Chi2
   #

   set angle_list {}

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
```

```
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

lappend result $angle_list

#do angle Chi3
#

set angle_list {}

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"NE\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

lappend result $angle_list

#do angle Chi4
#

set angle_list {}

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete
```

```
      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"NE\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CZ\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

  #do angle Chi5
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"NE\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CZ\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"NH1\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

   return $result
 }

proc get_HIS_sidechain_indicies {resid resname chain mol} {
   set result {}

    #do angle Chi1
    #

    set angle_list {}
```

```
      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

    #do angle Chi2
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"ND1\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

  return $result
}

proc get_PHE_sidechain_indicies {resid resname chain mol} {
  set result {}
```

```
    #do angle Chi1
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

    #do angle Chi2
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD1\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

  return $result
}
```

```
proc get_TYR_sidechain_indicies {resid resname chain mol} {
    set result {}

     #do angle Chi1
     #

     set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

     lappend result $angle_list

     #do angle Chi2
     #

     set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD1\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete
```

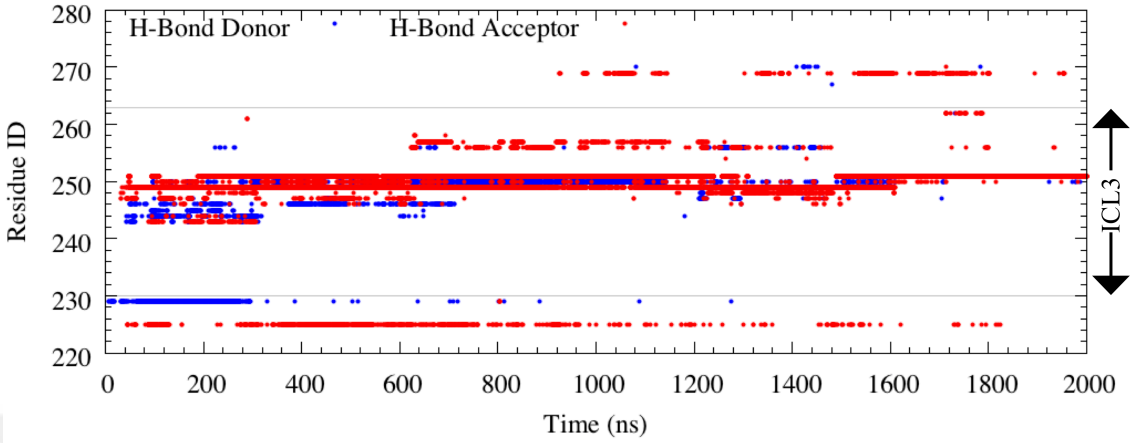
```
    lappend result $angle_list

   return $result
  }

proc get_TRP_sidechain_indicies {resid resname chain mol} {
   set result {}

    #do angle Chi1
    #

    set angle_list {}

     set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
     lappend angle_list [$atomsel get index]
     $atomsel delete

     set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
     lappend angle_list [$atomsel get index]
     $atomsel delete

     set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
     lappend angle_list [$atomsel get index]
     $atomsel delete

     set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
     lappend angle_list [$atomsel get index]
     $atomsel delete

    lappend result $angle_list

    #do angle Chi2
    #

    set angle_list {}

     set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
     lappend angle_list [$atomsel get index]
     $atomsel delete

     set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
     lappend angle_list [$atomsel get index]
     $atomsel delete

     set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CG\""]
     lappend angle_list [$atomsel get index]
     $atomsel delete
```

```
      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CD1\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

  return $result
}

proc get_CMT_sidechain_indicies {resid resname chain mol} {
  set result {}

    #do angle Chi1
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"N\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"SG\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

    lappend result $angle_list

    #do angle Chi2
    #

    set angle_list {}

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CA\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete

      set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
      lappend angle_list [$atomsel get index]
      $atomsel delete
```

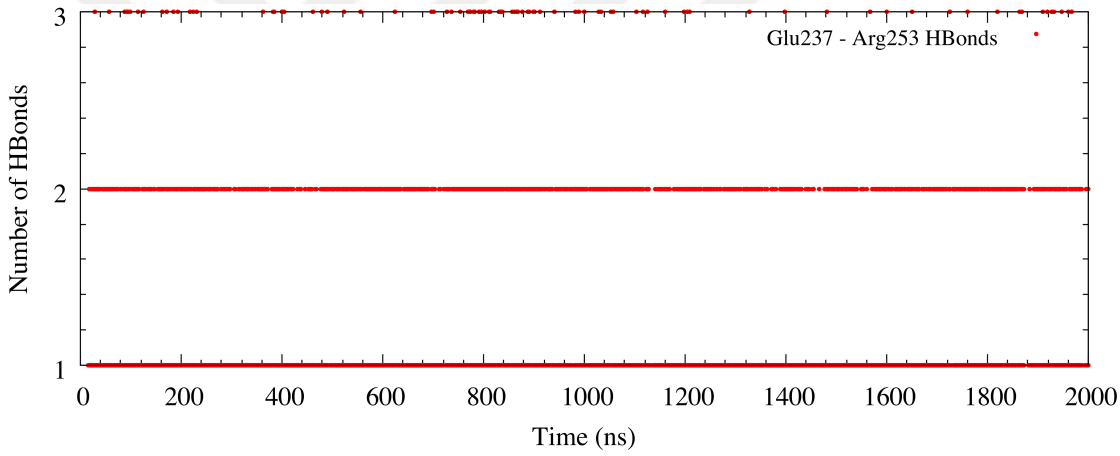
```
  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"SG\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"S1\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

lappend result $angle_list

#do angle Chi3
#

set angle_list {}

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"CB\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"SG\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"S1\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"C10\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

lappend result $angle_list

#do angle Chi4
#

set angle_list {}

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"SG\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete
  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"S1\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete

  set atomsel [atomselect $mol "chain $chain and resid $resid and name \"C10\""]
  lappend angle_list [$atomsel get index]
  $atomsel delete
```

```
    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"C3\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

  lappend result $angle_list

  #do angle Chi5
  #

  set angle_list {}

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"S1\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"C10\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"C3\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

    set atomsel [atomselect $mol "chain $chain and resid $resid and name \"C4\""]
    lappend angle_list [$atomsel get index]
    $atomsel delete

  lappend result $angle_list

 return $result
}
```

APPENDIX C



(a)



(b)

Figure C. 1 Hydrogen bonds between ICL3 and the core of the receptor (a). Blue color represented where ICL3 was hydrogen donor and red color represented where ICL3 was hydrogen acceptor. The number of hydrogen bonds between Glu237 and Arg253 (b). At least two hydrogen bonds were preserved between these residues throughout the trajectory.

APPENDIX D

Table D. 1 Sidechain χ angles and their ranges together with the atomic information defining the corresponding angle type

CHI1				
A.A.	**AXIS**	**ATOMS DEFINE ANGLE**	**ANGLE RANGE**	**MODIFIED ANGLE RANGE**
ARG	CA-CB	N-CA-CB-CG	from -180 to +180	from 0 to +360
ASN	CA-CB	N-CA-CB-CG	from -180 to +180	from 0 to +360
ASP	CA-CB	N-CA-CB-CG	from -180 to +180	from 0 to +360
CYS	CA-CB	N-CA-CB-SG	from -180 to +180	from 0 to +360
GLN	CA-CB	N-CA-CB-CG	from -180 to +180	from 0 to +360
GLU	CA-CB	N-CA-CB-CG	from -180 to +180	from 0 to +360
HIS	CA-CB	N-CA-CB-CG	from -180 to +180	from 0 to +360
ILE	CA-CB	N-CA-CB-CG1	from -180 to +180	from 0 to +360
LEU	CA-CB	N-CA-CB-CG	from -180 to +180	from 0 to +360
LYS	CA-CB	N-CA-CB-CG	from -180 to +180	from 0 to +360
MET	CA-CB	N-CA-CB-CG	from -180 to +180	from 0 to +360
PHE	CA-CB	N-CA-CB-CG	from -180 to +180	from 0 to +360
PRO	CA-CB	N-CA-CB-CG	CA-CB part of ring	CA-CB part of ring
SER	CA-CB	N-CA-CB-OG	from -180 to +180	from 0 to +360
THR	CA-CB	N-CA-CB-OG1	from -180 to +180	from 0 to +360
TRP	CA-CB	N-CA-CB-CG	from -180 to +180	from 0 to +360
TYR	CA-CB	N-CA-CB-CG	from -180 to +180	from 0 to +360
VAL	CA-CB	N-CA-CB-CG1	from -180 to +180	from 0 to +360
CHI2				
ARG	CB-CG	CA-CB-CG-CD	from -180 to +180	from 0 to +360
ASN	CB-CG	CA-CB-CG-OD1	from -180 to +180	from 0 to +360
ASP	CB-CG	CA-CB-CG-OD1	from -180 to +180	from 0 to +360
GLN	CB-CG	CA-CB-CG-CD	from -180 to +180	from 0 to +360
GLU	CB-CG	CA-CB-CG-CD	from -180 to +180	from 0 to +360
HIS	CB-CG	CA-CB-CG-ND1	from -180 to +180	from 0 to +360
ILE	CB-CG1	CA-CB-CG1-CD	from -180 to +180	from 0 to +360
LEU	CB-CG	CA-CB-CG-CD1	from -180 to +180	from 0 to +360
LYS	CB-CG	CA-CB-CG-CD	from -180 to +180	from 0 to +360
MET	CB-CG	CA-CB-CG-SD	from -180 to +180	from 0 to +360
PHE	CB-CG	CA-CB-CG-CD1	from -180 to +180	from 0 to +360
TRP	CB-CG	CA-CB-CG-CD1	from -180 to +180	from 0 to +360
TYR	CB-CG	CA-CB-CG-CD1	from -180 to +180	from 0 to +360
CHI3				
ARG	CG-CD	CB-CG-CD-NE	from -180 to +180	from 0 to +360
GLN	CG-CD	CB-CG-CD-OE1	from -180 to +180	from 0 to +360
GLU	CG-CD	CB-CG-CD-OE1	from -180 to +180	from 0 to +360
LYS	CG-CD	CB-CG-CD-CE	from -180 to +180	from 0 to +360
MET	CG-SD	CB-CG-SD-CE	from -180 to +180	from 0 to +360
CHI4				
ARG	CD-NE	CG-CD-NE-CZ	from -180 to +180	from 0 to +360
LYS	CD-CE	CG-CD-CE-NZ	from -180 to +180	from 0 to +360

APPENDIX E

Table E. 1 Rotameric states and the start and stop angles of corresponding states for 20 amino acid

A.A.	Angle	Conf.	Start Angle	Stop Angle	A.A.	Angle	Conf.	Start Angle	Stop Angle
ARG	Chi1	g+	0	120	LYS	Chi1	g+	0	120
	Chi1	t	120	240		Chi1	t	120	240
	Chi1	g-	240	360		Chi1	g-	240	360
	Chi2	g+	0	120		Chi2	g+	0	120
	Chi2	t	120	240		Chi2	t	120	240
	Chi2	g-	240	360		Chi2	g-	240	360
	Chi3	g+	0	120		Chi3	g+	0	120
	Chi3	t	120	240		Chi3	t	120	240
	Chi3	g-	240	360		Chi3	g-	240	360
	Chi4	g+	0	120		Chi4	g+	0	120
	Chi4	t	120	240		Chi4	t	120	240
	Chi4	g-	240	360		Chi4	g-	240	360
GLN	Chi1	g+	0	120	GLU	Chi1	g+	0	120
	Chi1	t	120	240		Chi1	t	120	240
	Chi1	g-	240	360		Chi1	g-	240	360
	Chi2	g+	0	120		Chi2	g+	0	120
	Chi2	t	120	240		Chi2	t	120	240
	Chi2	g-	240	360		Chi2	g-	240	360
	Chi3	Ng+	210	270		Chi3	g+	30	90
	Chi3	Og-	270	330		Chi3	g+	210	270
	Chi3	Nt	330	30		Chi3	t	150	210
	Chi3	Og+	30	90		Chi3	t	330	30
	Chi3	Ng-	90	150		Chi3	g-	90	150
	Chi3	Ot	150	210		Chi3	g-	270	330
MET	Chi1	g+	0	120	ASN	Chi1	g+	0	120
	Chi1	t	120	240		Chi1	t	120	240
	Chi1	g-	240	360		Chi1	g-	240	360
	Chi2	g+	0	120		Chi2	Ng+	210	270
	Chi2	t	120	240		Chi2	Og-	270	330
	Chi2	g-	240	360		Chi2	Nt	330	30
	Chi3	g+	0	120		Chi2	Og+	30	90
	Chi3	t	120	240		Chi2	Ng-	90	150
	Chi3	g-	240	360		Chi2	Ot	150	210

Table E. 1 (continued) Rotameric states and the start and stop angles of corresponding states for 20 amino acid

A.A.	Angle	Conf.	Start Angle	Stop Angle	A.A.	Angle	Conf.	Start Angle	Stop Angle
ASP	Chi1	g+	0	120	HIS	Chi1	g+	0	120
	Chi1	t	120	240		Chi1	t	120	240
	Chi1	g-	240	360		Chi1	g-	240	360
	Chi2	g+	30	90		Chi2	Ng+	30	90
	Chi2	g+	210	270		Chi2	Cg-	90	150
	Chi2	t	150	210		Chi2	Nt	150	210
	Chi2	t	330	30		Chi2	Cg+	210	270
	Chi2	g-	90	150		Chi2	Ng-	270	330
	Chi2	g-	270	330		Chi2	Ct	330	30
PHE	Chi1	g+	0	120	TYR	Chi1	g+	0	120
	Chi1	t	120	240		Chi1	t	120	240
	Chi1	g-	240	360		Chi1	g-	240	360
	Chi2	g	45	135		Chi2	g	45	135
	Chi2	g	225	315		Chi2	g	225	315
	Chi2	t	135	225		Chi2	t	135	225
	Chi2	t	315	45		Chi2	t	315	45
TRP	Chi1	g+	0	120	LEU	Chi1	g+	0	120
	Chi1	t	120	240		Chi1	t	120	240
	Chi1	g-	240	360		Chi1	g-	240	360
	Chi2	g+	0	120		Chi2	g+	0	120
	Chi2	t	120	240		Chi2	t	120	240
	Chi2	g-	240	360		Chi2	g-	240	360
ILE	Chi1	g+	0	120	CYS	Chi1	g+	0	120
	Chi1	t	120	240		Chi1	t	120	240
	Chi1	g-	240	360		Chi1	g-	240	360
	Chi2	g+	0	120	SER	Chi1	g+	0	120
	Chi2	t	120	240		Chi1	t	120	240
	Chi2	g-	240	360		Chi1	g-	240	360
THR	Chi1	g+	0	120	VAL	Chi1	g+	0	120
	Chi1	t	120	240		Chi1	t	120	240
	Chi1	g-	240	360		Chi1	g-	240	360
PRO	Chi1	g+	0	180					
	Chi1	g-	180	360					

APPENDIX F

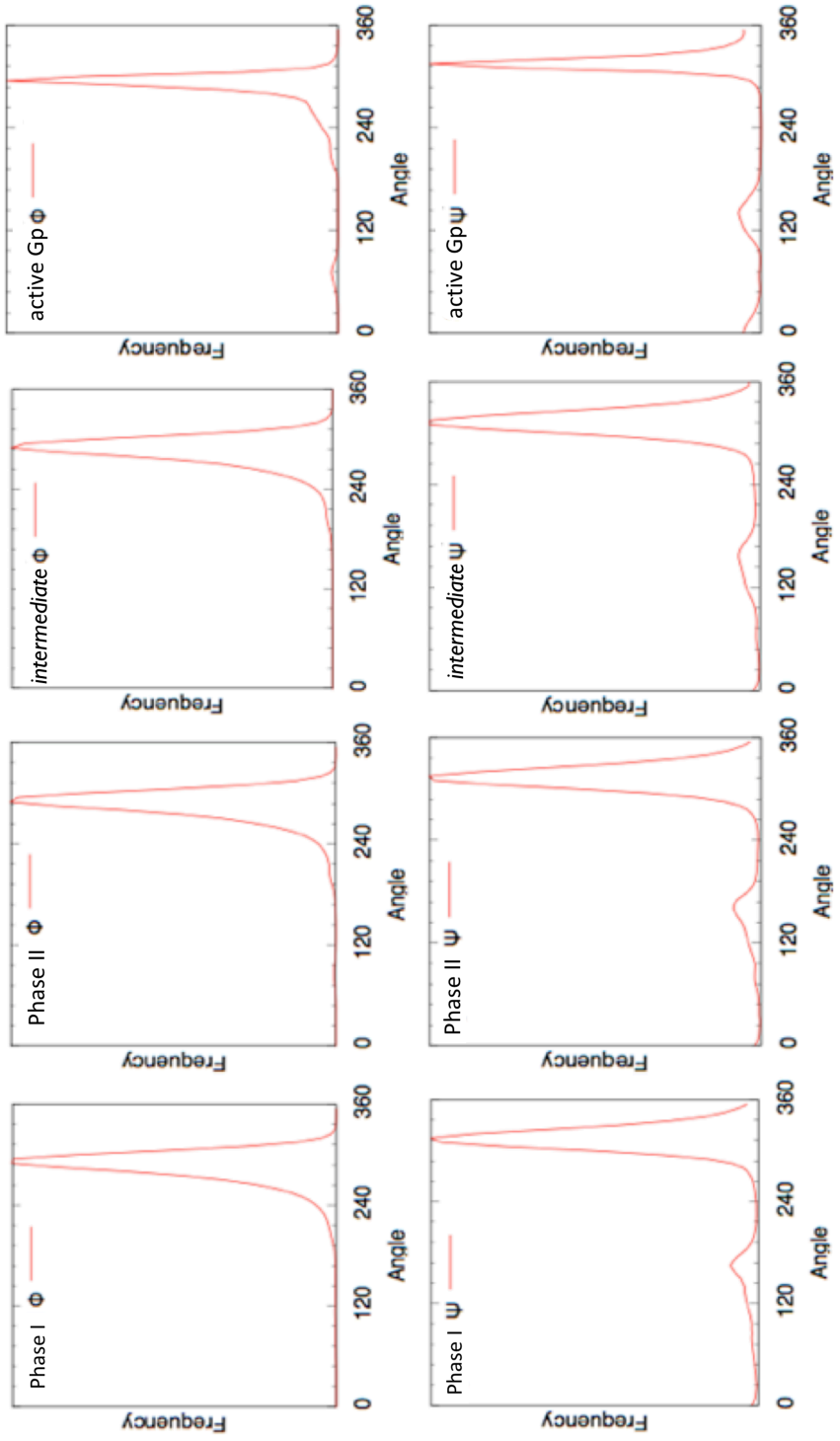


Figure F. 1 Backbone φ and ψ dihedral frequencies for 4 phases

APPENDIX G

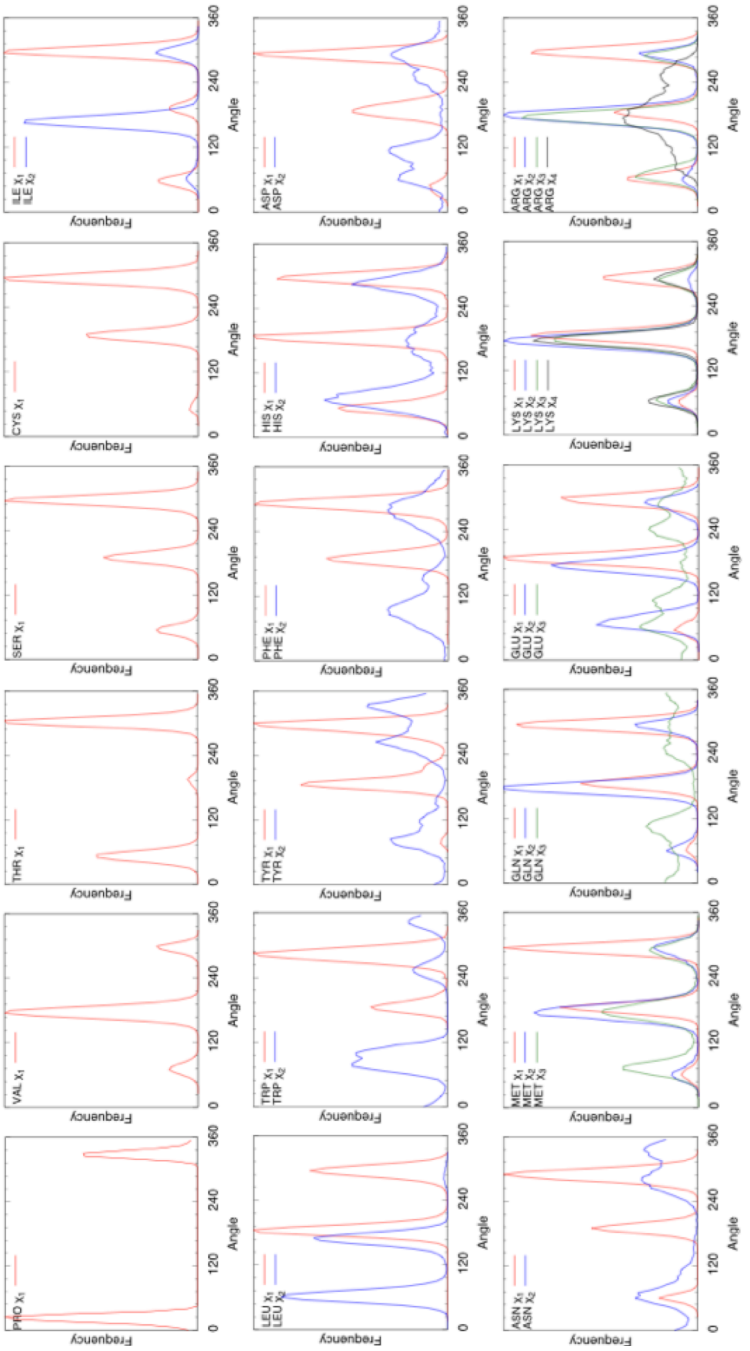


Figure G. 1 Sidechain χ dihedral angle frequencies for the amino acids in Phase I

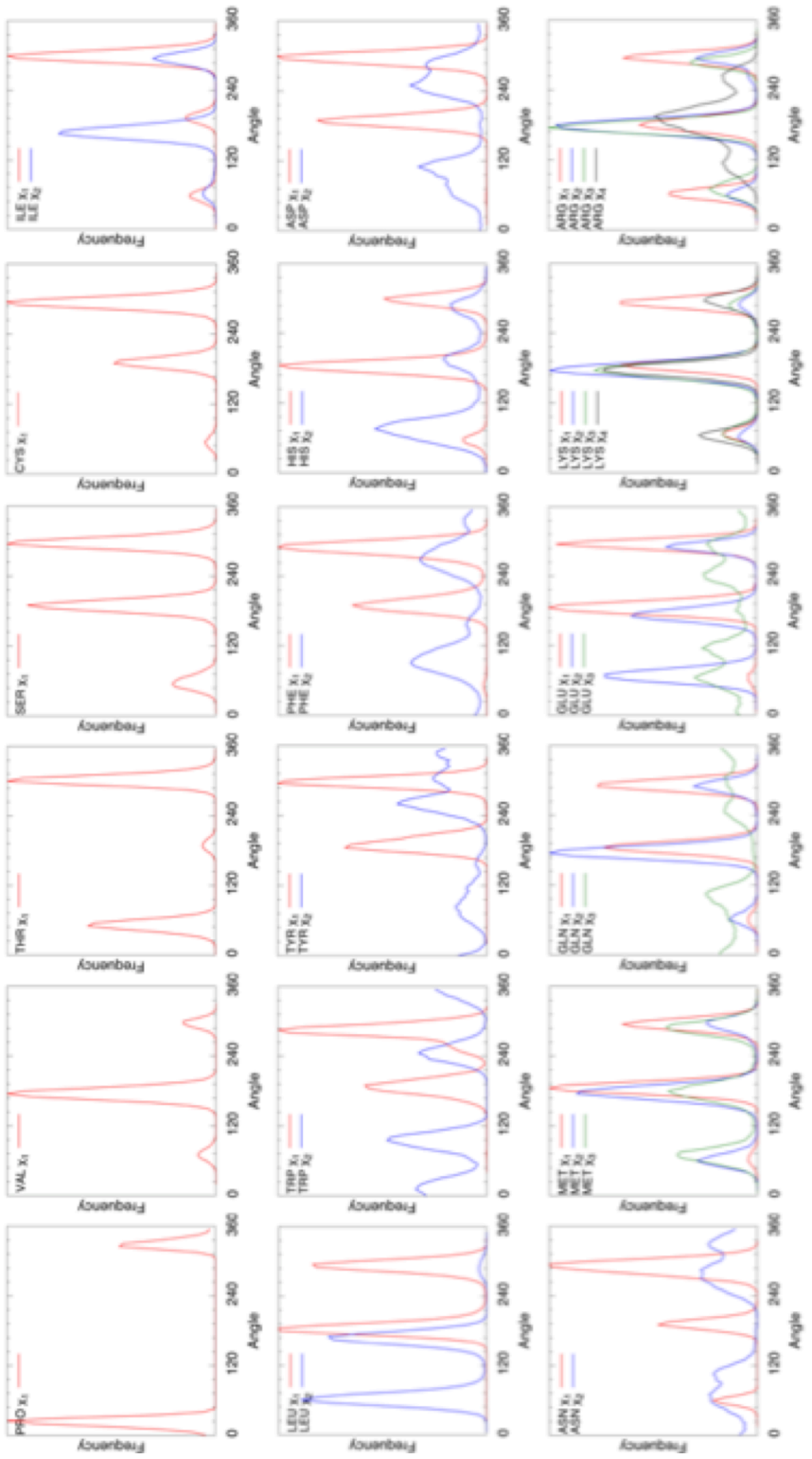


Figure G. 2 Sidechain χ dihedral angle frequencies for the amino acids in Phase II

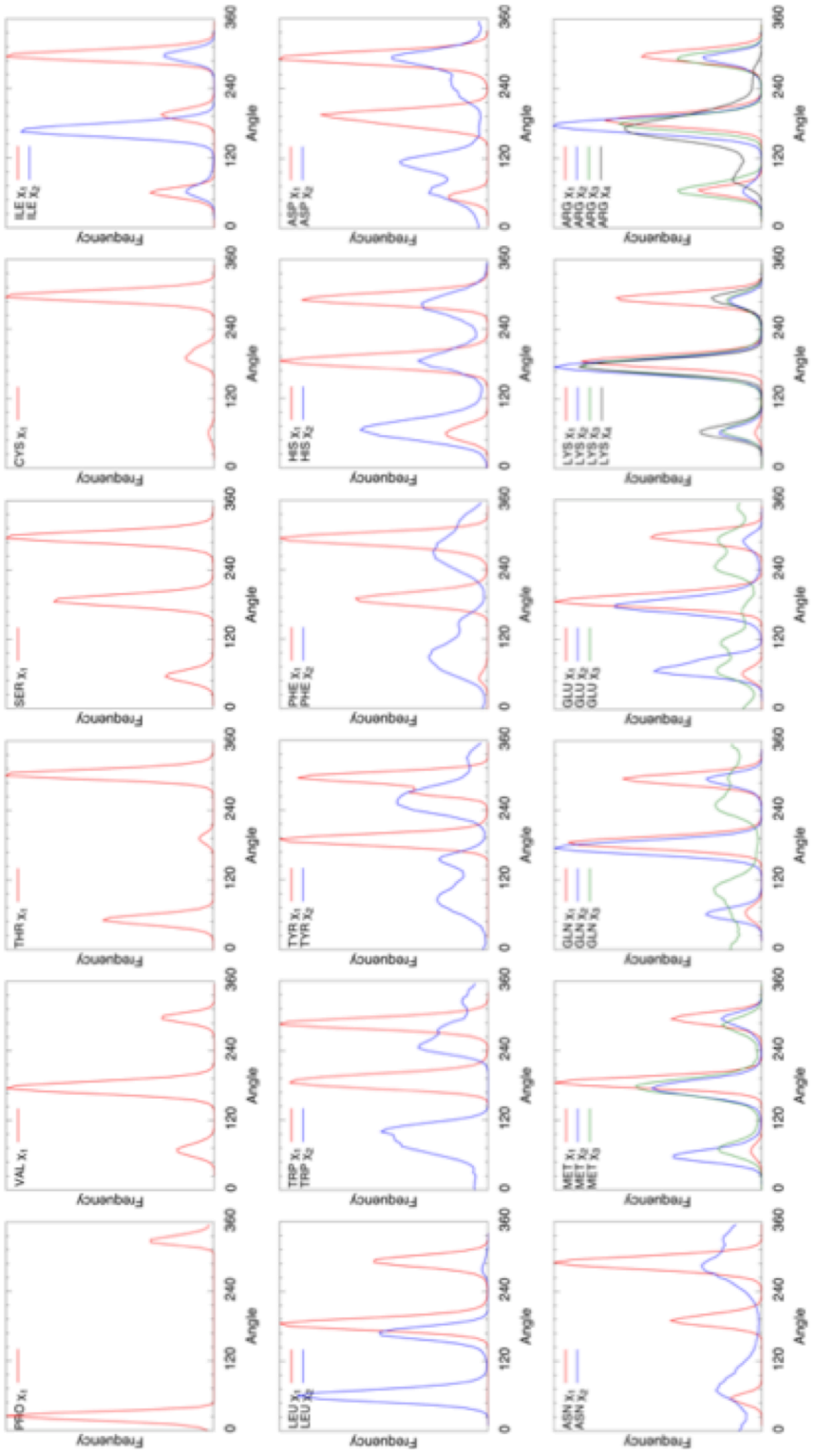


Figure G. 3 Sidechain χ dihedral angle frequencies for the amino acids in *intermediate* state

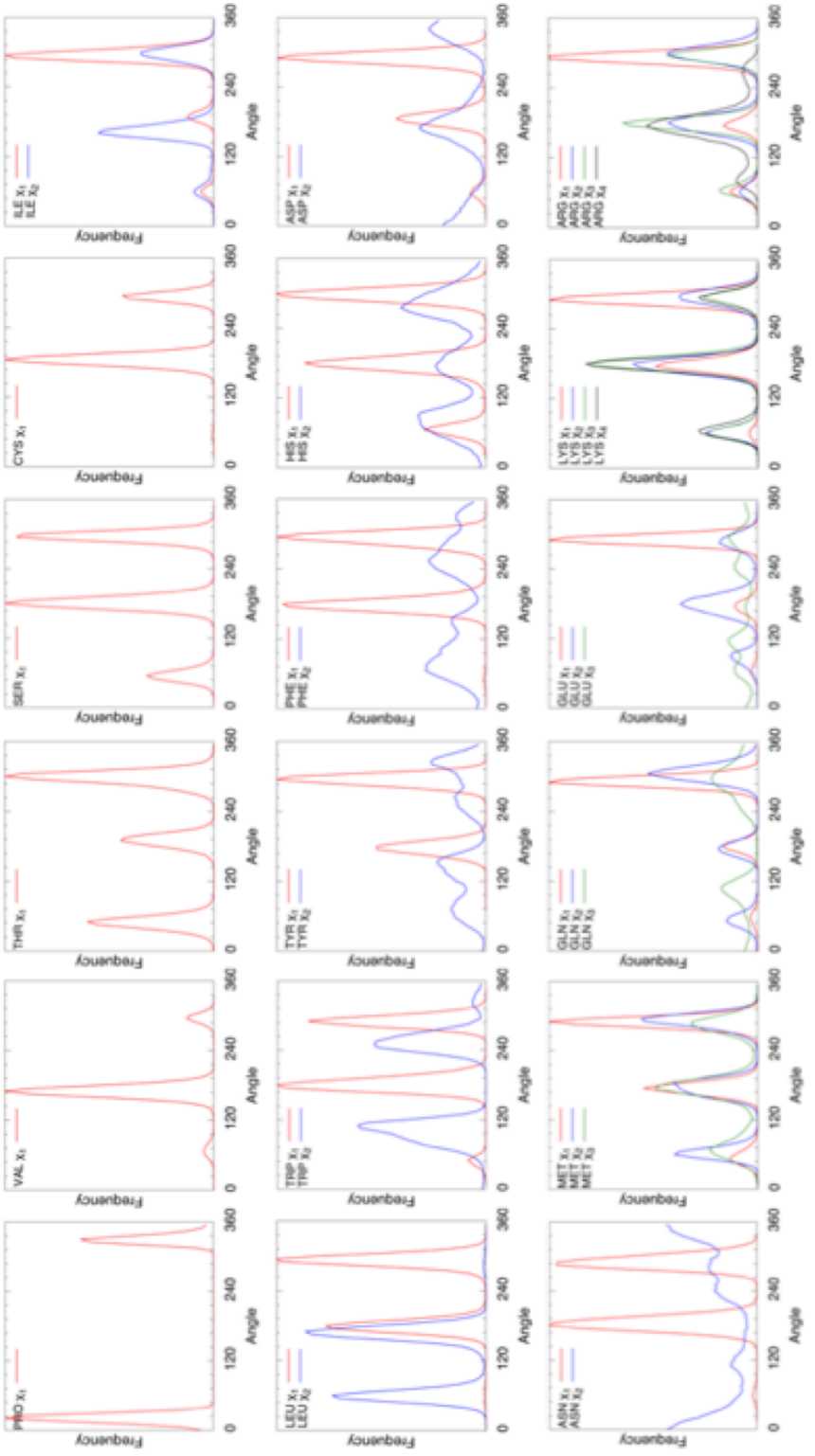


Figure G. 4 Sidechain χ dihedral angle frequencies for the amino acids in *active-Gp*

APPENDIX H

Table H. 1 Statistical validation of the significance of distribution differences between polar and hydrophobic residues using Kruskal Wallis nonparametric test

	Distribution Set	Proximity	H-Value	# of Samples	d.o.f*	p-Value** (Kruskal Wallis)
Phase I	Polar / Hydrophobic	Proximal	20.27	2468	1	6.7e-06
	Polar / Hydrophobic	Distal	846.31	93942	1	4.6e-186
Phase II	Polar / Hydrophobic	Proximal	31.31	2468	1	2.2e-08
	Polar / Hydrophobic	Distal	145.01	93942	1	2.14e-33
intermediate	Polar / Hydrophobic	Proximal	45.15	2476	1	1.82e-11
	Polar / Hydrophobic	Distal	680.867	94556	1	4.33e-150
Active Gp	Polar / Hydrophobic	Proximal	15.73	2476	1	7.30e-05
	Polar / Hydrophobic	Distal	78.26	94556	1	9.0e-19
Active Gp2nd	Polar / Hydrophobic	Proximal	0.76	2476	1	0.38
	Polar / Hydrophobic	Distal	176.67	94556	1	2.59e-40

*d.o.f: degree of freedom = (N-1), N = Number of Groups

**Significant if P <= 0.05 (Kruskal Wallis)

Table H. 2 Statistical validation of the significance of distribution differences between polar-polar, polar-hydrophobic and hydrophobic-hydrophobic residue-pairs using Kruskal Wallis nonparametric test

	Distribution Set	Proximity	H-Value	# of Samples	d.o.f	p-Value** (Kruskal Wallis)
Phase I	Polar-Polar / Polar-Hydrophobic / Hydrophobic-Hydrophobic	Proximal	38.99	2468	2	3.42e-09
	Polar-Polar / Polar-Hydrophobic / Hydrophobic-Hydrophobic	Distal	1720.5	93942	2	0.00
Phase II	Polar-Polar / Polar-Hydrophobic / Hydrophobic-Hydrophobic	Proximal	62.03	2468	2	3.39e-14
	Polar-Polar / Polar-Hydrophobic / Hydrophobic-Hydrophobic	Distal	328.53	93942	2	4.58e-72
intermediate	Polar-Polar / Polar-Hydrophobic / Hydrophobic-Hydrophobic	Proximal	86.87	2476	2	1.37e-19

	Polar-Polar / Polar-Hydrophobic / Hydrophobic-Hydrophobic	Distal	1375.73	94556	2	1.84e-299
Active Gp	Polar-Polar / Polar-Hydrophobic / Hydrophobic-Hydrophobic	Proximal	31.87	2476	2	1.20e-07
	Polar-Polar / Polar-Hydrophobic / Hydrophobic-Hydrophobic	Distal	197.41	94556	2	1.36e-43
Active Gp_{2nd}	Polar-Polar / Polar-Hydrophobic / Hydrophobic-Hydrophobic	Proximal	4.94	2476	2	0.084
	Polar-Polar / Polar-Hydrophobic / Hydrophobic-Hydrophobic	Distal	366.77	94556	2	2.26e-80

APPENDIX I

Table I. 1 Post Hoc validation of Kruskal Wallis H scores between polar and hydrophobic residue-pairs using Dunn's test[†]

Dunn's Test (fdr_tsbky[††])[***]	**Proximal**		**Distal**	
	Phase I	**Phase II**	**Phase I**	**Phase II**
Polar-Polar / Polar-Hydro	4.06e-04	4.36e-07	2.14e-121	4.95e-41
Polar-Polar / Hydro-Hydro	3.11e-09	1.69e-14	0.00	7.61e-73
Polar-Hydro / Hydro-Hydro	5.68e-04	1.60e-04	6.92e-137	6.06e-13
	active-Gp	**active-Gp_{2nd}**	**active-Gp**	**active-Gp_{2nd}**
Polar-Polar / Polar-Hydro	1.76e-04	0.084	2.45e-01	8.05e-33
Polar-Polar / Hydro-Hydro	5.74e-08	0.126	2.94e-21	2.15e-78
Polar-Hydro / Hydro-Hydro	0.013	0.56	1.20e-39	5.79e-25
	intermediate			***intermediate***
Polar-Polar / Polar-Hydro	1.012e-07			7.59e-89
Polar-Polar / Hydro-Hydro	2.64e-19			4.15e-275
Polar-Hydro / Hydro-Hydro	1.85e-07			9.33e-119

[***] Significant if P <= 0.01 (Dunn's Test)

[†] Self-comparison results are always -1 in Dunn's statistical analysis.

[††] False Discovery Rate 2-stage Benjamini-Krieger-Yekutieli

APPENDIX J

J.1 C Code for Number of Bins Optimization for Cα Fluctuations

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <string.h>


int main(int argn, char *arg[]) {

   FILE *fd, *fp1, *fp2, *fp3, *fp4;
   int i, j, k, m, n, ii, jj, kk, nres, nconf, sw, sw1, tau, nbins, c, fin[400], nc;
   char line[1024], key[128], name[128], fname[128], resi[128], resj[128];
   double sumx[400], sumy[400], sumz[400], binsize, binsize_res1, binsize_res2;
   double **ax, **ay, **az, *xb, *yb, *zb, SE[400], SEold[400], de[400];
   double avgx[400], avgy[400], avgz[400], **dr, **pDelR, ****pcDelRiRj,
***pcDelRjRj, *****pcDelRiRjRj;
   double **drx, **dry, **drz, a, b, aa, bb, aaa, bbb, pj, pj1, pj2, pj3, sum;
   double dridri[400], dridrj[400][400], **tr, **trtest, **trnew, *sumpj, **sumpj1,
*sumpj2, **sumpj3, Smin[400], Smax[400];

   if (argn!=2) {
      fprintf(stderr, "Usage: %s <filename>\n", arg[1]);
      return 1;
   }

// open for reading only
   fd = fopen(arg[1], "r");

   if (fd == NULL) {
      fprintf(stderr, "Cannot open %s.\n", arg[1]);
      return 1;
   }

   // open for writing only
   fp1 = fopen("Shannon_entropy_50ns-450ns.dat", "w");
   //fp2 = fopen("Shannon_entropy_resi.dat", "w");

// get the total number of lines
// so we allocate (more than) enough space for x,y,z arrays
   m = 0;
   while (fgets(line, 1024, fd)) {
      m++;
   }
   printf("%d\n",m);

   nconf = 2000;  // for 50 ns - 450 ns
   //nconf = 3750;  // for 750 ns - 1500 ns
   nres = 311;
   k = nconf*nres;
   printf("%d %d %d\n",nconf, nres, k);
   tau = 500;

   nbins = round(1 + log(nconf));  // Sturge's rule number of bins is equal to
(1+log2(number of data))
   printf("Number of bins is %5d\n", nbins);
   //nbins = 8;
   sw = 200;       // sliding window is 200 which corresponds to 200 * 0.2 = 40 ns

// allocate memory
   xb = malloc(sizeof(double)*k);
   yb = malloc(sizeof(double)*k);
   zb = malloc(sizeof(double)*k);

   ax = malloc(sizeof(double*)*nconf);
   ay = malloc(sizeof(double*)*nconf);
   az = malloc(sizeof(double*)*nconf);

   for (i=0;i<nconf;i++) ax[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) ay[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) az[i] = malloc(sizeof(double)*nres);

   drx = malloc(sizeof(double*)*nconf);
   dry = malloc(sizeof(double*)*nconf);
   drz = malloc(sizeof(double*)*nconf);

dr = malloc(sizeof(double*)*nconf);
   for (i=0;i<nconf;i++) drx[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) dry[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) drz[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) dr[i] = malloc(sizeof(double)*nres);

   pDelR = malloc(sizeof(double*)*(1501));
   for (i=0;i<(1501);i++) pDelR[i] = malloc(sizeof(double)*nres);


// rewind to the beginning of the file
   rewind(fd);

// now reading the coordinates for every conformation in the trajectory
   m = 0;
   while (fgets(line, 1024, fd)) {
      sscanf(line, "%s %*s %s %*s %*s %*s %lf %lf %lf %*s %*s", key, name,
&(xb[m]), &(yb[m]), &(zb[m]));
      if (strcmp(key, "ATOM")==0) {
         if (strcmp(name, "CA")==0){
            m++;
         }
      }
   }

   printf("%d\n",m);

   m = 0;
   for (i=0;i<nconf;i++) {
      for (j=0;j<nres;j++) {
         ax[i][j] = xb[m];
         ay[i][j] = yb[m];
         az[i][j] = zb[m];
         //printf("%lf %lf %lf %lf %lf %lf %d
%d\n",xb[m],yb[m],zb[m],ax[i][j],ay[i][j],az[i][j],i,j);
         m++;
      }
   }

   printf("%d\n",m);

// calculating the average position (over N snapshots) of each alpha Carbon atom in
residue i

   for (j=0;j<nres;j++){
      sumx[j] = 0.0;
      sumy[j] = 0.0;
      sumz[j] = 0.0;
      for (i=0;i<nconf;i++){
         sumx[j] = sumx[j] + ax[i][j];
         sumy[j] = sumy[j] + ay[i][j];
         sumz[j] = sumz[j] + az[i][j];
      }
      avgx[j] = sumx[j] / nconf;
      avgy[j] = sumy[j] / nconf;
      avgz[j] = sumz[j] / nconf;
      //printf("%d %lf %lf %lf\n",j, avgx[j], avgy[j], avgz[j]);
   }

// calculating the fluctuation vector from the average for each alpha Carbon of residue i
for each snapshot (Delta Ri)

   for (i=0;i<nconf;i++) {
      for (j=0;j<nres;j++) {
         drx[i][j] = ax[i][j] - avgx[j];
         dry[i][j] = ay[i][j] - avgy[j];
         drz[i][j] = az[i][j] - avgz[j];
         dr[i][j] = sqrt(drx[i][j]*drx[i][j] + dry[i][j]*dry[i][j] + drz[i][j]*drz[i][j]);  //
magnitude of the vector |Delta R(i)|
      }
   }

   // finding the min and the max values of |Delta R(i)|
   for (i=0;i<nres;i++) {
      Smin[i] = 1000.0;
      Smax[i] = -1000.0;
      for (k=0;k<nconf;k++) {
         if (dr[k][i] < Smin[i]){
            Smin[i] = dr[k][i];
         }
         if (dr[k][i] > Smax[i]){
            Smax[i] = dr[k][i];
         }
      }
      //printf("Smin Smax for res %5d are %5.3lf and %5.3lf \n", i+32, Smin[i],
Smax[i]);
      SEold[i] = 0.0;
      fin[i] = 0;
   }
```

```
   nc = 0;

 for (nbins=10;nbins<=1501;nbins++) {
     printf("%d\n", nbins);
// initialization of the counter
   for (i=0;i<nres;i++) {
      for (k=0;k<nbins+1;k++){
         pDelR[k][i] = 0.0;
      }
   }

// now calculating the probability of the fluctuations of residue i to be in state k,
p(Delta_Ri(tk)).
   for (i=0;i<nres;i++) {
      binsize_res1 = (Smax[i]-Smin[i])/nbins;
      for (k=0;k<nconf;k++) {
         for (ii=0;ii<nbins+1;ii++){
            a = Smin[i] + (ii-0.5)*binsize_res1;
            b = Smin[i] + (ii+0.5)*binsize_res1;
            if ((dr[k][i] >=a) & (dr[k][i] < b)){
               ++pDelR[ii][i];
            }
         }
      }
   }
   for (i=0;i<nres;i++) {
      for (ii=0;ii<nbins+1;ii++){
         pDelR[ii][i] /= nconf;      // p(Delta_Ri(tk))
         //printf("nbins pdel %d %12.8lf\n", nbins, pDelR[ii][i]);
      }
   }

   // testing .....
   /*double tot = 0.0;
   for (i=0;i<nres;i++) {
      tot = 0.0;
      for (ii=0;ii<nbins+1;ii++)
         tot += pDelR[ii][i];
      //printf("testing total %7.4lf\n", tot);
   }*/

   //printf("p(Delta_Ri(tk)) calculated!\n");

// Shannon Entropy
     for (i=0;i<nres;i++) {
        //printf("%d\n", i);
        if (fin[i] == 0) {
           SE[i] = 0.0;
           for (ii=0;ii<nbins+1;ii++) {
              if (pDelR[ii][i] != 0.0) {
                 SE[i] += -pDelR[ii][i] * log(pDelR[ii][i]) / log(2);
              }
           }
           de[i] = (SE[i]-SEold[i])/SEold[i]; // our interpretation
           //de[i] = SE[i]-SEold[i]; // Kamberaj paper
           //if (i == 147) {
             //  fprintf(fp2, "%d %d %12.8lf %12.8lf %12.8lf\n", i, nbins, SE[i], SEold[i],
de[i]);
           //}
           //printf("%d %d %12.8lf %12.8lf %12.8lf\n", i, nbins, SE[i], SEold[i], de[i]);
           if (fabs(de[i]) < 0.02){
              fprintf(fp1, "%d %d %12.8lf %12.8lf %12.8lf\n", i, nbins, SE[i], SEold[i],
de[i]);
              printf("%d %d %12.8lf %12.8lf %12.8lf\n", i, nbins, SE[i], SEold[i], de[i]);
              fin[i] = 1;
              ++nc;
              printf("%d\n", nc);
           }
           SEold[i] = SE[i];
        }
     }
    if (nc == nres) break;
}

  return 0;
}
```

J.2 C Code for Time Delay Optimization for Cα Fluctuations

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <string.h>


int main(int argn, char *arg[]) {

   FILE *fd, *fp1, *fp2, *fdT;
   int i, j, k, m, n, ii, jj, nres, nconf, sw, tau, nbins, c, nbn, nb[400], resid, ncount,
fmin[400];
   char line[1024], key[128], name[128], fname[128], resi[128], resj[128];
   double sumx[400], sumy[400], sumz[400], binsize, binsize_res1, binsize_res2;
   double **ax, **ay, **az, *xb, *yb, *zb, tot;
   double avgx[400], avgy[400], avgz[400], **dr, **pDelR, **pDelR_prev,
***pDelRiRi;
   double **drx, **dry, **drz,  a, b, aa, bb, pi, pj, pj2, pk, sum;
   double dridri[400], dridrj[400][400], Smin[400], Smax[400], **mi, **mi2, *fin,
*ncount_i, *ncount_j, *ncount_ij;

   if (argn!=3) {
      fprintf(stderr, "Usage: %s %s <filename>\n", arg[1], arg[2]);
      return 1;
   }

// open for reading only
   fd = fopen(arg[1], "r");
   fdT = fopen(arg[2], "r");

   if (fd == NULL) {
      fprintf(stderr, "Cannot open %s.\n", arg[1]);
      return 1;
   }

   if (fdT == NULL) {
      fprintf(stderr, "Cannot open %s.\n", arg[2]);
      return 1;
   }

   // open for writing only
   fp1 = fopen("mutual_info_50ns-450ns_Tau_Nbins0.02_testing2.dat", "w");

// get the total number of lines, to allocate (more than) enough space for x,y,z arrays
   m = 0;
   while (fgets(line, 1024, fd)) {
      m++;
   }
   printf("%d\n",m);

   nconf = 2000;  // for 750 ns - 1500 ns
   nres = 311;
   k = nconf*nres;
   printf("%d %d %d\n",nconf, nres, k);
   tau = 300;
   printf("Number of bins is %5d\n", nbins);
   nbins = 1500;

// allocate memory
   xb = malloc(sizeof(double)*k);
   yb = malloc(sizeof(double)*k);
   zb = malloc(sizeof(double)*k);

   ax = malloc(sizeof(double*)*nconf);
   ay = malloc(sizeof(double*)*nconf);
   az = malloc(sizeof(double*)*nconf);

   for (i=0;i<nconf;i++) ax[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) ay[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) az[i] = malloc(sizeof(double)*nres);

   drx = malloc(sizeof(double*)*nconf);
   dry = malloc(sizeof(double*)*nconf);
   drz = malloc(sizeof(double*)*nconf);
   dr = malloc(sizeof(double*)*nconf);

   for (i=0;i<nconf;i++) drx[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) dry[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) drz[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) dr[i] = malloc(sizeof(double)*nres);

   pDelR = malloc(sizeof(double*)*(nbins+1));
   for (i=0;i<(nbins+1);i++) pDelR[i] = malloc(sizeof(double)*nres);

   pDelR_prev = malloc(sizeof(double*)*(nbins+1));
   for (i=0;i<(nbins+1);i++) pDelR_prev[i] = malloc(sizeof(double)*nres);

   pDelRiRi = malloc(sizeof(double**)*(nbins+1));
   for (ii=0;ii<(nbins+1);ii++) pDelRiRi[ii] = malloc(sizeof(double*)*nres);
   for (ii=0;ii<(nbins+1);ii++){
      for (i=0;i<nres;i++) pDelRiRi[ii][i] = malloc(sizeof(double)*(nbins+1));
   }

   mi = malloc(sizeof(double*)*tau);
   for (i=0;i<tau;i++) mi[i] = malloc(sizeof(double)*nres);

   mi2 = malloc(sizeof(double*)*tau);
   for (i=0;i<tau;i++) mi2[i] = malloc(sizeof(double)*nres);

   fin = malloc(sizeof(double)*nres);

   ncount_i = malloc(sizeof(double)*nres);
   ncount_j = malloc(sizeof(double)*nres);
   ncount_ij = malloc(sizeof(double)*nres);

// now reading the number of bin values for every residue
   while (fgets(line, 1024, fdT)) {
      sscanf(line, "%d %d %*s %*s %*s", &resid, &nbn);
      nb[resid] = nbn;
   }

   // testing ...
   for (j=0;j<nres;j++)
      printf("%d %d\n", j, nb[j]);

// rewind to the beginning of the file
   rewind(fd);

// now reading the coordinates for every conformation in the trajectory
   m = 0;
   while (fgets(line, 1024, fd)) {
      sscanf(line, "%s %*s %s %*s %*s %*s %*s %lf %lf %lf %*s %*s", key, name,
&(xb[m]), &(yb[m]), &(zb[m]));
      if (strcmp(key, "ATOM")==0) {
         if (strcmp(name, "CA")==0){
            m++;
         }
      }
   }

   printf("%d\n",m);

   m = 0;
   for (i=0;i<nconf;i++) {
      for (j=0;j<nres;j++) {
         ax[i][j] = xb[m];
         ay[i][j] = yb[m];
         az[i][j] = zb[m];
         //printf("%lf %lf %lf %lf %lf %lf %d
%d\n",xb[m],yb[m],zb[m],ax[i][j],ay[i][j],az[i][j],i,j);
         m++;
      }
   }

   printf("%d\n",m);

// calculating the average position (over N snapshots) of each alpha Carbon atom in
residue i
   for (j=0;j<nres;j++){
      sumx[j] = 0.0;
      sumy[j] = 0.0;
      sumz[j] = 0.0;
      for (i=0;i<nconf;i++){
         sumx[j] = sumx[j] + ax[i][j];
         sumy[j] = sumy[j] + ay[i][j];
         sumz[j] = sumz[j] + az[i][j];
      }
      avgx[j] = sumx[j] / nconf;
      avgy[j] = sumy[j] / nconf;
      avgz[j] = sumz[j] / nconf;
      //printf("%d %lf %lf %lf\n",j, avgx[j], avgy[j], avgz[j]);
   }

// calculating the fluctuation vector from the average for each alpha Carbon of residue i
for each snapshot (Delta Ri)
   for (i=0;i<nconf;i++) {
      for (j=0;j<nres;j++) {
         drx[i][j] = ax[i][j] - avgx[j];
         dry[i][j] = ay[i][j] - avgy[j];
```

```
        drz[i][j] = az[i][j] - avgz[j];
        dr[i][j] = sqrt(drx[i][j]*drx[i][j] + dry[i][j]*dry[i][j] + drz[i][j]*drz[i][j]);  //
magnitude of the vector |Delta R(i)|
      }
   }

// now calculating the probability of observing Delta Ri in state k
// generating the histogram first

   // finding the min and the max values of |Delta R(i)|
   for (i=0;i<nres;i++) {
      Smin[i] = 1000.0;
      Smax[i] = -1000.0;
      for (k=0;k<nconf;k++){
         if (dr[k][i] < Smin[i])
            Smin[i] = dr[k][i];
         if (dr[k][i] > Smax[i])
            Smax[i] = dr[k][i];
      }
      printf("Smin1 Smax1 for res %5d are %5.3lf and %5.3lf \n", i+32, Smin[i],
Smax[i]);
   }

// initialization
   for (i=0;i<nres;i++){
      for (k=0;k<(nb[i]+1);k++){
         pDelR[k][i] = 0.0;
      }
      fmin[i] = 0;
   }

   for (i=0;i<nres;i++){
      for (k=0;k<(nb[i]+1);k++){
         pDelR_prev[k][i] = 0.0;
      }
   }

   for (i=0;i<nres;i++){
      for (ii=0;ii<(nb[i]+1);ii++){
         for (jj=0;jj<(nb[i]+1);jj++)
            pDelRiRi[ii][i][jj] = 0.0;
      }
   }
   printf("Initialization done!\n");

// calculating the mutual information for different tau value for each residue

  for (sw=0;sw<=250;sw++) {
     printf("sw is %5d\n", sw);
// now calculating the probability of the fluctuations of residue i to be in state k

     for (i=0;i<nres;i++) {
       if (fmin[i] == 0) {
       //printf("first i is %5d\n", i);
            binsize_res1 = (Smax[i]-Smin[i])/nb[i];
            ncount = 0;
            for (k=0;k<nconf;k++) {
               for (ii=0;ii<(nb[i]+1);ii++){
                  a = Smin[i] + (ii-0.5)*binsize_res1;
                  b = Smin[i] + (ii+0.5)*binsize_res1;
                  if ((dr[k][i] > a) & (dr[k][i] <= b)) {
                     ++pDelR[ii][i];
                     ++ncount;
                     ii = nb[i]+1;
                  }
               }
            }
       }
      //printf("%d %d\n", i, ncount);
     }

   // testing... total pDelR for each resiue should be equal to 1
     for (i=0;i<nres;i++) {
        tot = 0.0;
        for (ii=0;ii<nb[i]+1;ii++){
           pDelR[ii][i] /= nconf;
           tot += pDelR[ii][i];
        }
     //  printf("i and tot is %5d %5lf\n", i+32, tot);
     }

     for (i=0;i<nres;i++) {
        if (fmin[i] == 0) {
           //printf("first i is %5d\n", i);
           binsize_res1 = (Smax[i]-Smin[i])/nb[i];
           ncount = 0;
           for (k=0;k<(nconf-sw);k++) {
              for (ii=0;ii<(nb[i]+1);ii++){
                 a = Smin[i] + (ii-0.5)*binsize_res1;
                 b = Smin[i] + (ii+0.5)*binsize_res1;
                 if ((dr[k][i] > a) & (dr[k][i] <= b)) {
                    ++pDelR_prev[ii][i];
                    ++ncount;
                    ii = nb[i]+1;
                 }
              }
           }
        }
        //printf("%d %d\n", i, ncount);
     }

     // testing... total pDelR for each resiue should be equal to 1
     for (i=0;i<nres;i++) {
        tot = 0.0;
        for (ii=0;ii<nb[i]+1;ii++){
           pDelR_prev[ii][i] /= (nconf-sw);
           tot += pDelR_prev[ii][i];
        }
        //  printf("i and tot is %5d %5lf\n", i+32, tot);
     }

// now calculating the joint probability of the fluctuations of residue i to be in state l while
the same residue i is in state k, p(Delta_Ri(tk),DeltaRi(tk))
     for (i=0;i<nres;i++) {
        if (fmin[i] == 0) {
      //printf("second i is %5d\n", i);
           binsize_res1 = (Smax[i]-Smin[i])/nb[i];
           for (n=0;n<(nconf-sw);n++) {
              for (ii=0;ii<nb[i]+1;ii++){
                 a = Smin[i] + (ii-0.5)*binsize_res1;
                 b = Smin[i] + (ii+0.5)*binsize_res1;
                 for (jj=0;jj<nb[i]+1;jj++){
                    aa = Smin[i] + (jj-0.5)*binsize_res1;
                    bb = Smin[i] + (jj+0.5)*binsize_res1;
                    if ((dr[n][i] > a) & (dr[n][i] <= b)){
                       if ((dr[n+sw][i] > aa) & (dr[n+sw][i] <= bb)){
                          ++pDelRiRi[ii][i][jj];
                          jj = nb[i]+1;
                          ii = nb[i]+1;
                       }
                    }
                 }
              }
           }
        }
     }
   }

// testing... total pDelR for each resiue should be equal to 1
   for (i=0;i<nres;i++) {
      //tot = 0.0;
      for (ii=0;ii<(nb[i]+1);ii++){
         for (jj=0;jj<(nb[i]+1);jj++){
            pDelRiRi[ii][i][jj] /= (nconf-sw);
            //tot += pDelRiRi[ii][i][jj];
         }
      }
      //printf("i and tot is %5d %5lf\n", i+32, tot);
   }

// calculation of mutual information
   for (i=0;i<nres;i++){
      //printf("third i is %5d\n",i);
      if (fmin[i] == 0) {
      //printf("third i is %5d\n",i);
         mi[sw][i] = 0.0;
         mi2[sw][i] = 0.0;
         ncount_i[i] = 0.0;
         ncount_j[i] = 0.0;
         ncount_ij[i] = 0.0;
         fin[i] = 0.0;
      // finding the bin of the current DeltaRi
         // finding the joint probability of res i to be in state l at time t+sw while in state k
at time t
         for (ii=0;ii<nb[i]+1;ii++){
            if (pDelR[ii][i] !=0) {
               ++ncount_i[i];
            }
         }

         for (jj=0;jj<nb[i]+1;jj++){
            if (pDelR_prev[jj][i]!=0) {
               ++ncount_j[i];
            }
         }
         for (ii=0;ii<nb[i]+1;ii++){
            for (jj=0;jj<nb[i]+1;jj++){
               if (pDelRiRi[ii][i][jj] !=0) {
                  ++ncount_ij[i];
```

```
                }
            }
        }
        for (ii=0;ii<nb[i]+1;ii++){
            for (jj=0;jj<nb[i]+1;jj++){
                if (pDelR[ii][i] !=0 & pDelR_prev[jj][i] !=0 & pDelRiRi[ii][i][jj] !=0) {
                    pi = pDelR[ii][i];
                    pk = pDelR_prev[jj][i];
                    pj2 = pDelRiRi[ii][i][jj];
                    mi[sw][i] = pj2 * (log(pj2/(pk * pi))/log(2));
                    mi2[sw][i] += mi[sw][i];
                }
            }
        }
        fin[i] = ncount_i[i] + ncount_j[i] - ncount_ij[i] - 1; //Finite sampling error
estimation based on # of non-zero bins
        fin[i] /= 2*nconf;
        printf("%d %d %d %lf %lf %lf %lf\n", i, ii, jj, ncount_i[i], ncount_j[i],
ncount_ij[i], fin[i]);
        mi2[sw][i] += fin[i];
       // determine the tau value at the which the mutual information is at its first
minimum
        if (sw > 1){
            if (fmin[i] == 0){
                if ( ((mi2[sw][i] - mi2[sw-1][i]) > 0.0) && ((mi2[sw-2][i] - mi2[sw-1][i])
> 0.0) ) {
                    fmin[i] = sw-1;
                    printf("%d %d %9.4lf %9.4lf %9.4lf\n", i, fmin[i], mi2[sw-2][i], mi2[sw-
1][i], mi2[sw][i]);
                    fprintf(fp1, "%d %d %9.4lf %9.4lf %9.4lf\n", i, fmin[i], mi2[sw-2][i],
mi2[sw-1][i], mi2[sw][i]);
                }
            }
        }
    }
  }
 }

  return 0;
}
```

J.3 C Code for Calculation of Mutual Information from Cα Fluctuations

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <string.h>


int main(int argn, char *arg[]) {

   FILE *fd, *fp, *fdN, *fp1, *fp2, *fp3, *fp4, *fp5, *fp6, *fp7, *fp8;

   char line[1024], key[128], name[128], fname[128], resname[128],
**residue_type;

   int i, j, k, kk, m, n, ii, jj, nres, nconf, sw, tau, nbins, c;
   int resid, ncount, ntau, nbn, nb[400], nt[400], nt_type[20];
   double lev, lev1, lev2, lev3, lev4, lev5, lev6;
   double sum[311], avg[311], std[311];

   int nc_close[20][20], nc_far[20][20], sumnc_close[20], sumnc_far[20], total, *loc;
   int nc_close_tr[20][20], nc_far_tr[20][20], sumnc_close_tr[20],
sumnc_far_tr[20];
   int nc_close_trloop[20][20], nc_far_trloop[20][20], sumnc_close_trloop[20],
sumnc_far_trloop[20];
   int nc_close_loop[20][20], nc_far_loop[20][20], sumnc_close_loop[20],
sumnc_far_loop[20];

   double sumx[400], sumy[400], sumz[400], binsize, binsize_res1, binsize_res2;
   double **ax, **ay, **az, *xb, *yb, *zb, *ncount_j, *ncount_i, **ncount_ij, tot;
   double avgx[400], avgy[400], avgz[400], **dr, **pDelR, ****pDelRiRj;
   double **drx, **dry, **drz,  a, b, aa, bb, pi, pj, pj2, pk, **fin;
   double dridri[400], dridrj[400][400], Smin[400], Smax[400], **mi, **mi2;

   double sum_mi_close[20][20], sum_mi_far[20][20];
   double sum_mi_new_close[20][20], sum_mi_new_far[20][20], sumr_close[20],
sumr_far[20];
   double sum_mi_close_tr[20][20], sum_mi_far_tr[20][20],
sum_mi_close_trloop[20][20], sum_mi_far_trloop[20][20];
   double sum_mi_close_loop[20][20], sum_mi_far_loop[20][20];;
   double sum_mi_new_close_tr[20][20], sum_mi_new_far_tr[20][20],
sumr_close_tr[20], sumr_far_tr[20];
   double sum_mi_new_close_trloop[20][20], sum_mi_new_far_trloop[20][20],
sumr_close_trloop[20], sumr_far_trloop[20];
   double sum_mi_new_close_loop[20][20], sum_mi_new_far_loop[20][20],
sumr_close_loop[20], sumr_far_loop[20];

   const char *reslist[] = {"ALA", "GLY", "PRO", "CYS", "SER", "THR", "VAL", "ASN", "ASP",
"HIS", "ILE", "LEU", "PHE", "TRP", "TYR", "GLN", "GLU", "MET", "ARG", "LYS"};
  // const int rotlist[] = {0, 0, 0, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4};

   //New rotlist grouping based on charges
   //Nonpolar, aliphatic 0, Ala, Gly, Pro, Val, Ile, Leu, (Met), (Cys)
   //Nonpolar, aromatic 1, Phe, Trp, Tyr
   //Polar, uncharged 2, Ser, Thr, Asn, Gln
   //Negatively, charged 3, Asp, Glu
   //Positively, charged 4, His, Arg, Lys


   const int rotlist[] = {0, 0, 0, 0, 2, 2, 0, 2, 3, 4, 0, 0, 1, 1, 1, 2, 3, 0, 4, 4};

   // HYDROPHOBIC: 1
   // POLAR: 0
   const int reslist_HP[] = {1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0};

   if (argn!=3) {
      fprintf(stderr, "Usage: %s %s <filename>\n", arg[1], arg[2]);
      return 1;
   }

// open for reading only
   fd = fopen(arg[1], "r");
   fdN = fopen(arg[2], "r");

   if (fd == NULL) {
      fprintf(stderr, "Cannot open %s.\n", arg[1]);
      return 1;
   }

   // open for writing only
   fp = fopen("mutual_info_50ns-450ns_fin_testing.dat", "w");
   fp1 = fopen("mutual_info_50ns-450ns_residuepair_fin_testing.dat", "w");


   fp2 = fopen("mutual_info_50ns-450ns_residuepair_regions_far_fin_testing.dat",
"w");
   fp3 = fopen("mutual_info_50ns-450ns_residuepair_regions_close_fin_testing.dat",
"w");

   fp4 = fopen("mutual_info_50ns-450ns_residue_fin_testing.dat", "w");
   fp5 = fopen("mutual_info_50ns-450ns_residue_regions_far_fin_testing.dat", "w");
   fp6 = fopen("mutual_info_50ns-450ns_residue_regions_close_fin_testing.dat",
"w");

   fp7 = fopen("mean_mutual_per_resid_50ns-450ns_fin_testing.dat", "w");
   fp8 = fopen("MI_phasel_testing.dat", "w");

// get the total number of lines
// so we allocate (more than) enough space for x,y,z arrays
   m = 0;
   while (fgets(line, 1024, fd)) {
      m++;
   }
   printf("%d\n",m);

   nconf = 2000;  // for 50 ns - 450 ns
   //nconf = 3750;  // for 750 ns - 1500 ns
   nres = 311;
   k = nconf*nres;
   printf("%d %d %d\n",nconf, nres, k);
   tau = 300;

   //nbins = round(1 + log(nconf));  // Sturge's rule number of bins is equal to
(1+log2(number of data)
   printf("Number of bins is %5d\n", nbins);
   nbins = 100;

// allocate memory
   residue_type = malloc(sizeof(char*)*nres);
   for (i=0;i<nres;i++) residue_type[i] = malloc(sizeof(char)*128);

   fin = malloc(sizeof(double*)*nres);
   for (i=0;i<nres;i++) fin[i] = malloc(sizeof(double)*nres);

   loc = malloc(sizeof(int)*nres);

   xb = malloc(sizeof(double)*k);
   yb = malloc(sizeof(double)*k);
   zb = malloc(sizeof(double)*k);

   ax = malloc(sizeof(double*)*nconf);
   ay = malloc(sizeof(double*)*nconf);
   az = malloc(sizeof(double*)*nconf);

   for (i=0;i<nconf;i++) ax[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) ay[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) az[i] = malloc(sizeof(double)*nres);

   drx = malloc(sizeof(double*)*nconf);
   dry = malloc(sizeof(double*)*nconf);
   drz = malloc(sizeof(double*)*nconf);
   dr = malloc(sizeof(double*)*nconf);

   for (i=0;i<nconf;i++) drx[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) dry[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) drz[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) dr[i] = malloc(sizeof(double)*nres);

   mi = malloc(sizeof(double*)*nres);
   for (i=0;i<nres;i++) mi[i] = malloc(sizeof(double)*nres);

   mi2 = malloc(sizeof(double*)*nres);
   for (i=0;i<nres;i++) mi2[i] = malloc(sizeof(double)*nres);

   ncount_i = malloc(sizeof(double)*nres);
   ncount_j = malloc(sizeof(double)*nres);

   ncount_ij = malloc(sizeof(double*)*nres);
   for (i=0;i<nres;i++){
      ncount_ij[i] = malloc(sizeof(double)*nres);
   }

   pDelR = malloc(sizeof(double*)*(nbins+1));
   for (i=0;i<(nbins+1);i++) pDelR[i] = malloc(sizeof(double)*nres);

   pDelRiRj = malloc(sizeof(double***)*(nbins+1));
   for (ii=0;ii<(nbins+1);ii++) pDelRiRj[ii] = malloc(sizeof(double**)*nres);
   for (ii=0;ii<(nbins+1);ii++){
      for (i=0;i<nres;i++) pDelRiRj[ii][i] = malloc(sizeof(double*)*(nbins+1));
   }
```

```
   for (ii=0;ii<(nbins+1);ii++){
      for (i=0;i<nres;i++){
         for (jj=0;jj<(nbins+1);jj++)
            pDelRiRj[ii][i][jj] = malloc(sizeof(double)*nres);
      }
   }

// now reading the number of bin values for every residue (from Shannon Entropy
calculations)
   while (fgets(line, 1024, fdN)) {
      sscanf(line, "%d %d %*s %*s %*s", &resid, &nbn);
      nb[resid] = nbn;
   }

   // testing ...
   for (j=0;j<nres;j++)
      printf("%d %d\n", j, nb[j]);

// rewind to the beginning of the file
   rewind(fd);

// now reading the coordinates for every conformation in the trajectory
   m = 0;
   while (fgets(line, 1024, fd)) {
      sscanf(line, "%s %*s %s %s %*s %3d %lf %lf %lf %*s %*s", key, name,
resname, &resid, &(xb[m]), &(yb[m]), &(zb[m]));
      if (strcmp(key, "ATOM")==0) {
         if (strcmp(name, "CA")==0){
            if (m <= 310){
               if ((strcmp(resname, "HSE")!=0) && (strcmp(resname, "HSD")!=0))
                  strcpy(residue_type[m], resname);
               else
                  strcpy(residue_type[m], reslist[9]);
               if ( (((m+32)>=32) && ((m+32)<=60)) || (((m+32)>=67) &&
((m+32)<=95)) || (((m+32)>=103) && ((m+32)<=136)) || (((m+32)>=148)
&& ((m+32)<=171)) || (((m+32)>=197) && ((m+32)<=230)) ||
(((m+32)>=267) && ((m+32)<=298)) || (((m+32)>=305) &&
((m+32)<=328)) || (((m+32)>=330) && ((m+32)<=342)))
                  loc[m] = 1;
               else
                  loc[m] = 0;
            }
            m++;
         }
      }
   }

   for (i=0;i<nres;i++)
      printf("%d %d\n", i+32, loc[i]);

   printf("%d\n",m);
   for (k=0;k<20;k++) nt_type[k] = 0;;
   for (i=0;i<nres;i++){
      printf("res type %d %s\n", i, residue_type[i]);
      for (k=0;k<20;k++){
         if (strcmp(residue_type[i], reslist[k])==0)
            nt_type[k]++;
      }
   }
   printf("frequency\n");
   for (k=0;k<20;k++)
      printf("%s %lf\n", reslist[k], nt_type[k]/311.0);

   m = 0;
   for (i=0;i<nconf;i++) {
      for (j=0;j<nres;j++) {
         ax[i][j] = xb[m];
         ay[i][j] = yb[m];
         az[i][j] = zb[m];
         //printf("%lf %lf %lf %lf %lf %d
%d\n",xb[m],yb[m],zb[m],ax[i][j],ay[i][j],az[i][j],i,j);
         m++;
      }
   }

   printf("%d\n",m);

// calculating the average position (over N snapshots) of each alpha Carbon atom in
residue i
   for (j=0;j<nres;j++){
      sumx[j] = 0.0;
      sumy[j] = 0.0;
      sumz[j] = 0.0;
      for (i=0;i<nconf;i++){
         sumx[j] = sumx[j] + ax[i][j];
         sumy[j] = sumy[j] + ay[i][j];
         sumz[j] = sumz[j] + az[i][j];
      }
      avgx[j] = sumx[j] / nconf;
      avgy[j] = sumy[j] / nconf;
      avgz[j] = sumz[j] / nconf;
      //printf("%d %lf %lf %lf\n",j, avgx[j], avgy[j], avgz[j]);
   }

// calculating the fluctuation vector from the average for each alpha Carbon of residue i
for each snapshot (Delta Ri)
   for (i=0;i<nconf;i++) {
      for (j=0;j<nres;j++) {
         drx[i][j] = ax[i][j] - avgx[j];
         dry[i][j] = ay[i][j] - avgy[j];
         drz[i][j] = az[i][j] - avgz[j];
         dr[i][j] = sqrt(drx[i][j]*drx[i][j] + dry[i][j]*dry[i][j] + drz[i][j]*drz[i][j]);  //
magnitude of the vector |Delta R(i)|
      }
   }

//finite sample error calculation
   for(i=0;i<nres;i++){
      for(j=0;j<nres;j++){
         fin[i][j] = nb[i] + nb[j] -(nb[i] * nb[j]) - 1;
         fin[i][j] /= (2*nconf);
         //printf("%d %d %lf\n", i, j, fin[i][j]);
      }
   }


// now calculating the probability of observing Delta Ri in state k
// generating the histogram first

   // finding the min and the max values of |Delta R(i)|
   for (i=0;i<nres;i++) {
      Smin[i] = 1000.0;
      Smax[i] = -1000.0;
      for (k=0;k<nconf;k++){
         if (dr[k][i] < Smin[i])
            Smin[i] = dr[k][i];
         if (dr[k][i] > Smax[i])
            Smax[i] = dr[k][i];
      }
      //printf("Smin1 Smax1 for res %5d are %5.3lf and %5.3lf \n", i+32, Smin[i],
Smax[i]);
   }

// initialization
   for (i=0;i<nres;i++){
      for (k=0;k<(nb[i]+1);k++){
         pDelR[k][i] = 0.0;
      }
   }

   for (i=0;i<nres;i++){
      for (j=0;j<nres;j++){
         for (ii=0;ii<(nb[i]+1);ii++){
            for (jj=0;jj<(nb[j]+1);jj++)
               pDelRiRj[ii][i][jj][j] = 0.0;
         }
      }
   }
   printf("Initialization done!!\n");

// calculating the mutual information for different nbins value for each residue

// first calculating the probability of the fluctuations of residue i to be in state k

   for (i=0;i<nres;i++) {
      printf("first i is %5d\n", i);
      binsize_res1 = (Smax[i]-Smin[i])/nb[i];
      ncount = 0;
      for (k=0;k<nconf;k++) {
         for (ii=0;ii<(nb[i]+1);ii++){
            a = Smin[i] + (ii-0.5)*binsize_res1;
            b = Smin[i] + (ii+0.5)*binsize_res1;
            if ((dr[k][i] > a) & (dr[k][i] <= b)) {
               ++pDelR[ii][i];
               ++ncount;
               ii = nb[i]+1;
            }
         }
      }
      //printf("%d %d\n", i, ncount);
   }

   // testing... total pDelR for each resiue should be equal to 1
  for (i=0;i<nres;i++) {
     tot = 0.0;
     for (ii=0;ii<nb[i]+1;ii++){
        pDelR[ii][i] /= nconf;
        tot += pDelR[ii][i];
```

```
        }
        //printf("i and tot is %5d %5lf\n", i+32, tot);
    }

// now calculating the joint probability of the fluctuations of residue i to be in state l while
the same residue i is in state k, p(Delta_Ri(tk),DeltaRi(tk))
    for (i=0;i<nres;i++) {
        printf("second i is %5d\n", i);
        binsize_res1 = (Smax[i]-Smin[i])/nb[i];
        for (j=0;j<nres;j++){
            binsize_res2 = (Smax[j]-Smin[j])/nb[j];
            for (n=0;n<nconf;n++) {
                for (ii=0;ii<nb[i]+1;ii++){
                    a = Smin[i] + (ii-0.5)*binsize_res1;
                    b = Smin[i] + (ii+0.5)*binsize_res1;
                    for (jj=0;jj<nb[j]+1;jj++){
                        aa = Smin[j] + (jj-0.5)*binsize_res2;
                        bb = Smin[j] + (jj+0.5)*binsize_res2;
                        if ((dr[n][i] > a) & (dr[n][i] <= b)){
                            if ((dr[n][j] > aa) & (dr[n][j] <= bb)){
                                ++pDelRiRj[ii][i][jj][j];
                                jj = nb[j]+1;
                                ii = nb[i]+1;
                            }
                        }
                    }
                }
            }
        }
    }

// testing... total pDelR for each resiue should be equal to 1
    for (i=0;i<nres;i++) {
        for (j=0;j<nres;j++){
            tot = 0.0;
            for (ii=0;ii<(nb[i]+1);ii++){
                for (jj=0;jj<(nb[j]+1);jj++){
                    pDelRiRj[ii][i][jj][j] /= nconf;
                    tot += pDelRiRj[ii][i][jj][j];
                }
            }
            //printf("i, j and tot is %5d %5d %5lf\n", i+32, j+32, tot);
        }
    }

// calculation of mutual information between residue i and j
    for (i=0;i<nres;i++){
        printf("%5d\n",i);
        for (j=0;j<nres;j++){
            mi[i][j] = 0.0;
            mi2[i][j] = 0.0;
            ncount_i[i] = 0.0;
            ncount_j[j] = 0.0;
            ncount_ij[i][j] = 0.0;
            fin[i][j] = 0.0;
        // finding the bin of the current DeltaRi
            // finding the joint probability of res i to be in state l while residue j is in state k
            for (ii=0;ii<nb[i]+1;ii++){
                if (pDelR[ii][i]!=0) {
                    ++ncount_i[i];
                }
            }
            for (jj=0;jj<nb[j]+1;jj++){
                if (pDelR[jj][j]!=0) {
                    ++ncount_j[j];
                }
            }
            for (ii=0;ii<nb[i]+1;ii++){
                for (jj=0;jj<nb[j]+1;jj++){
                    if (pDelRiRj[ii][i][jj][j] !=0) {
                        ++ncount_ij[i][j];
                    }
                }
            }
            for (ii=0;ii<nb[i]+1;ii++){
                for (jj=0;jj<nb[j]+1;jj++){
                    if (pDelRiRj[ii][i][jj][j] !=0 & pDelR[ii][i]!=0 & pDelR[jj][j]!=0) {
                        pi = pDelR[ii][i];
                        pj = pDelR[jj][j];
                        pj2 = pDelRiRj[ii][i][jj][j];
                        mi[i][j] = pj2 * (log(pj2/(pi * pj))/log(2));
                        mi2[i][j] += mi[i][j];
                    }
                }
            }
            fin[i][j] = ncount_i[i] + ncount_j[j] - ncount_ij[i][j] - 1; //Finite sampling error
estimation based on # of non-zero bins
            fin[i][j] /= 2*nconf;
            printf("%d %d %d %d %lf %lf %lf %lf\n", i, j, ii, jj, ncount_i[i], ncount_j[j],
ncount_ij[i][j], fin[i][j]);
            fprintf(fp,"%5d %5d Before; %lf  ", i, j, mi2[i][j]);
            mi2[i][j] += fin[i][j];          // finite sampling correction

            fprintf(fp," After; %lf\n", mi2[i][j]);
        }
    }

    for (i=0;i<nres;i++){
        for (j=0;j<nres;j++){
            fprintf (fp8, "%d %d %lf %lf\n", i+32, j+32, mi2[i][j], mi2[j][i]);
        }
    }


    // determining the residue types
    // first initialization
    for (k=0;k<20;k++){
        for (kk=0;kk<20;kk++){
            nc_close[k][kk] = 0;
            sum_mi_close[k][kk] = 0.0;
            nc_close_loop[k][kk] = 0;
            sum_mi_close_loop[k][kk] = 0.0;
            nc_close_tr[k][kk] = 0;
            sum_mi_close_tr[k][kk] = 0.0;
            nc_close_trloop[k][kk] = 0;
            sum_mi_close_trloop[k][kk] = 0.0;
            nc_far[k][kk] = 0;
            sum_mi_far[k][kk] = 0.0;
            nc_far_loop[k][kk] = 0;
            sum_mi_far_loop[k][kk] = 0.0;
            nc_far_tr[k][kk] = 0;
            sum_mi_far_tr[k][kk] = 0.0;
            nc_far_trloop[k][kk] = 0;
            sum_mi_far_trloop[k][kk] = 0.0;
        }
    }

    double totMI_close = 0.0;
    double totMI_far = 0.0;

    for (i=0;i<nres;i++){
        for (j=i+1;j<nres;j++){
            for (k=0;k<20;k++){
                if (strcmp(residue_type[i], reslist[k])==0){
                    for (kk=0;kk<20;kk++){
                        if (strcmp(residue_type[j], reslist[kk])==0){
                            if (abs(i-j) < 5){
                                nc_close[k][kk]++;
                                sum_mi_close[k][kk] += mi2[i][j];
                                totMI_close += mi2[i][j];
                                if ((loc[i]+loc[j])==2){
                                    nc_close_tr[k][kk]++;
                                    sum_mi_close_tr[k][kk] += mi2[i][j];
                                }
                                else if ((loc[i]+loc[j])==1){
                                    nc_close_trloop[k][kk]++;
                                    sum_mi_close_trloop[k][kk] += mi2[i][j];
                                }
                                else if ((loc[i]+loc[j])==0){
                                    nc_close_loop[k][kk]++;
                                    sum_mi_close_loop[k][kk] += mi2[i][j];
                                }
                                else printf("Error %3d %3d!\n", loc[i], loc[j]);
                            }
                            if (abs(i-j) >= 5){
                                nc_far[k][kk]++;
                                sum_mi_far[k][kk] += mi2[i][j];
                                totMI_far += mi2[i][j];
                                if ((loc[i]+loc[j])==2){
                                    nc_far_tr[k][kk]++;
                                    sum_mi_far_tr[k][kk] += mi2[i][j];
                                }
                                else if ((loc[i]+loc[j])==1){
                                    nc_far_trloop[k][kk]++;
                                    sum_mi_far_trloop[k][kk] += mi2[i][j];
                                }
                                else if ((loc[i]+loc[j])==0){
                                    nc_far_loop[k][kk]++;
                                    sum_mi_far_loop[k][kk] += mi2[i][j];
                                }
                                else printf("Error %3d %3d!\n", loc[i], loc[j]);
                            }
                        }
                    }
                }
            }
        }
    }
```

```
    }

    total = 0;
    for (k=0;k<20;k++){
        for (kk=0;kk<20;kk++){
            total += nc_close[k][kk] + nc_far[k][kk];
            printf("%3d %3d %s %s %7.4lf %3d %7.4lf %7.4lf %7.4lf %3d %7.4lf
%7.4lf\n", k, kk, reslist[k], reslist[kk], sum_mi_close[k][kk], nc_close[k][kk],
sum_mi_close[k][kk]/nc_close[k][kk], sum_mi_close[k][kk]/totMI_close*100.0,
sum_mi_far[k][kk], nc_far[k][kk], sum_mi_far[k][kk]/nc_far[k][kk],
sum_mi_far[k][kk]/totMI_far*100.0);
        }
    }
    printf("tot %3d\n", total);
    printf("total MI for close %7.4lf\n", totMI_close);
    printf("total MI for far %7.4lf\n", totMI_far);

    // combining i - j nd j - i pairs
    for (k=0;k<20;k++){
        sumr_close[k] = 0.0;
        sumr_far[k] = 0.0;
        sumnc_close[k] = 0;
        sumnc_far[k] = 0;

        sumr_close_tr[k] = 0.0;
        sumr_far_tr[k] = 0.0;
        sumnc_close_tr[k] = 0;
        sumnc_far_tr[k] = 0;
        sumr_close_trloop[k] = 0.0;
        sumr_far_trloop[k] = 0.0;
        sumnc_close_trloop[k] = 0;
        sumnc_far_trloop[k] = 0;
        sumr_close_loop[k] = 0.0;
        sumr_far_loop[k] = 0.0;
        sumnc_close_loop[k] = 0;
        sumnc_far_loop[k] = 0;

        for (kk=0;kk<20;kk++){
            sum_mi_new_close[k][kk] = sum_mi_close[k][kk] + sum_mi_close[kk][k];
            sumnc_close[k] += nc_close[k][kk] + nc_close[kk][k];
            sumr_close[k] += sum_mi_new_close[k][kk];

            sum_mi_new_far[k][kk] = sum_mi_far[k][kk] + sum_mi_far[kk][k];
            sumnc_far[k] += nc_far[k][kk] + nc_far[kk][k];
            sumr_far[k] += sum_mi_new_far[k][kk];

        // mutual information between transmembrane regions
            sum_mi_new_close_tr[k][kk] = sum_mi_close_tr[k][kk] +
sum_mi_close_tr[kk][k];
            sumnc_close_tr[k] += nc_close_tr[k][kk] + nc_close_tr[kk][k];
            sumr_close_tr[k] += sum_mi_new_close_tr[k][kk];

            sum_mi_new_far_tr[k][kk] = sum_mi_far_tr[k][kk] + sum_mi_far_tr[kk][k];
            sumnc_far_tr[k] += nc_far_tr[k][kk] + nc_far_tr[kk][k];
            sumr_far_tr[k] += sum_mi_new_far_tr[k][kk];

        // mutual information between transmembrane and loop regions
            sum_mi_new_close_trloop[k][kk] = sum_mi_close_trloop[k][kk] +
sum_mi_close_trloop[kk][k];
            sumnc_close_trloop[k] += nc_close_trloop[k][kk] + nc_close_trloop[kk][k];
            sumr_close_trloop[k] += sum_mi_new_close_trloop[k][kk];

            sum_mi_new_far_trloop[k][kk] = sum_mi_far_trloop[k][kk] +
sum_mi_far_trloop[kk][k];
            sumnc_far_trloop[k] += nc_far_trloop[k][kk] + nc_far_trloop[kk][k];
            sumr_far_trloop[k] += sum_mi_new_far_trloop[k][kk];

        // mutual information between loop regions
            sum_mi_new_close_loop[k][kk] = sum_mi_close_loop[k][kk] +
sum_mi_close_loop[kk][k];
            sumnc_close_loop[k] += nc_close_loop[k][kk] + nc_close_loop[kk][k];
            sumr_close_loop[k] += sum_mi_new_close_loop[k][kk];

            sum_mi_new_far_loop[k][kk] = sum_mi_far_loop[k][kk] +
sum_mi_far_loop[kk][k];
            sumnc_far_loop[k] += nc_far_loop[k][kk] + nc_far_loop[kk][k];
            sumr_far_loop[k] += sum_mi_new_far_loop[k][kk];
        }
    }

    for (k=0;k<20;k++){
        for (kk=0;kk<20;kk++){
            fprintf(fp1, "%3d %3d %3d %3d %s %3d %s %3d %8.4lf %5d %8.4lf %8.4lf
%8.4lf %5d %8.4lf %8.4lf\n", k, kk, rotlist[k], rotlist[kk], reslist[k], reslist_HP[k],
reslist[kk], reslist_HP[kk], sum_mi_new_close[k][kk],
nc_close[k][kk]+nc_close[kk][k],
sum_mi_new_close[k][kk]/(nc_close[k][kk]+nc_close[kk][k]),
sum_mi_new_close[k][kk]/totMI_close*100.0, sum_mi_new_far[k][kk],
nc_far[k][kk]+nc_far[kk][k], sum_mi_new_far[k][kk]/(nc_far[k][kk]+nc_far[kk][k]),
sum_mi_new_far[k][kk]/totMI_far*100.0);

            // mutual information between far residues in different regions (tr-tr / tr-loop /
loop-loop) regions
            fprintf(fp2, "%3d %3d %3d %3d %s %3d %s %3d %8.4lf %5d %8.4lf %8.4lf
%8.4lf %5d %8.4lf %8.4lf %8.4lf %5d %8.4lf %8.4lf\n", k, kk, rotlist[k], rotlist[kk],
reslist[k], reslist_HP[k], reslist[kk], reslist_HP[kk], sum_mi_new_far_tr[k][kk],
nc_far_tr[k][kk]+nc_far_tr[kk][k],
sum_mi_new_far_tr[k][kk]/(nc_far_tr[k][kk]+nc_far_tr[kk][k]),
sum_mi_new_far_tr[k][kk]/totMI_far*100.0, sum_mi_new_far_trloop[k][kk],
nc_far_trloop[k][kk]+nc_far_trloop[kk][k],
sum_mi_new_far_trloop[k][kk]/(nc_far_trloop[k][kk]+nc_far_trloop[kk][k]),
sum_mi_new_far_trloop[k][kk]/totMI_far*100.0, sum_mi_new_far_loop[k][kk],
nc_far_loop[k][kk]+nc_far_loop[kk][k],
sum_mi_new_far_loop[k][kk]/(nc_far_loop[k][kk]+nc_far_loop[kk][k]),
sum_mi_new_far_loop[k][kk]/totMI_far*100.0);

            // mutual information between close residues in different regions (tr-tr / tr-loop /
loop-loop)
            fprintf(fp3, "%3d %3d %3d %3d %s %3d %s %3d %8.4lf %5d %8.4lf %8.4lf
%8.4lf %5d %8.4lf %8.4lf %8.4lf %5d %8.4lf %8.4lf\n", k, kk, rotlist[k], rotlist[kk],
reslist[k], reslist_HP[k], reslist[kk], reslist_HP[kk], sum_mi_new_close_tr[k][kk],
nc_close_tr[k][kk]+nc_close_tr[kk][k],
sum_mi_new_close_tr[k][kk]/(nc_close_tr[k][kk]+nc_close_tr[kk][k]),
sum_mi_new_close_tr[k][kk]/totMI_close*100.0, sum_mi_new_close_trloop[k][kk],
nc_close_trloop[k][kk]+nc_close_trloop[kk][k],
sum_mi_new_close_trloop[k][kk]/(nc_close_trloop[k][kk]+nc_close_trloop[kk][k]),
sum_mi_new_close_trloop[k][kk]/totMI_close*100.0,
sum_mi_new_close_loop[k][kk], nc_close_loop[k][kk]+nc_close_loop[kk][k],
sum_mi_new_close_loop[k][kk]/(nc_close_loop[k][kk]+nc_close_loop[kk][k]),
sum_mi_new_close_loop[k][kk]/totMI_close*100.0);
        }
    }

    for (k=0;k<20;k++){
        fprintf(fp4, "%3d %3d %s %3d %8.4lf %5d %8.4lf %8.4lf %8.4lf %5d %8.4lf
%8.4lf\n", k, rotlist[k], reslist[k], reslist_HP[k], sumr_close[k], sumnc_close[k],
sumr_close[k]/sumnc_close[k], sumr_close[k]/totMI_close*100.0, sumr_far[k],
sumnc_far[k], sumr_far[k]/sumnc_far[k], sumr_far[k]/totMI_far*100.0);

        // residue based mutual information for far residues in different regions (tr-tr / tr-
loop / loop-loop) regions
        fprintf(fp5, "%3d %3d %s %3d %8.4lf %5d %8.4lf %8.4lf %8.4lf %5d %8.4lf
%8.4lf %8.4lf %5d %8.4lf %8.4lf\n", k, rotlist[k], reslist[k], reslist_HP[k],
sumr_far_tr[k], sumnc_far_tr[k], sumr_far_tr[k]/sumnc_far_tr[k],
sumr_far_tr[k]/totMI_far*100.0, sumr_far_trloop[k], sumnc_far_trloop[k],
sumr_far_trloop[k]/sumnc_far_trloop[k], sumr_far_trloop[k]/totMI_far*100.0,
sumr_far_loop[k], sumnc_far_loop[k], sumr_far_loop[k]/sumnc_far_loop[k],
sumr_far_loop[k]/totMI_far*100.0);

        // residue based mutual information for close residues in different regions (tr-tr / tr-
loop / loop-loop) regions
        fprintf(fp6, "%3d %3d %s %3d %8.4lf %5d %8.4lf %8.4lf %8.4lf %5d %8.4lf
%8.4lf %8.4lf %5d %8.4lf %8.4lf\n", k, rotlist[k], reslist[k], reslist_HP[k],
sumr_close_tr[k], sumnc_close_tr[k], sumr_close_tr[k]/sumnc_close_tr[k],
sumr_close_tr[k]/totMI_close*100.0, sumr_close_trloop[k], sumnc_close_trloop[k],
sumr_close_trloop[k]/sumnc_close_trloop[k],
sumr_close_trloop[k]/totMI_close*100.0, sumr_close_loop[k], sumnc_close_loop[k],
sumr_close_loop[k]/sumnc_close_loop[k], sumr_close_loop[k]/totMI_close*100.0);
    }


    // percentage of MI values in defined ranges
    lev = lev1 = lev2 = lev3 = lev4 = lev5 = lev6 = 0;
    for (i=0;i<nres;i++){
        for (j=i;j<nres;j++){
            lev++;
            if ((mi2[i][j] > 0.0) && (mi2[i][j] <= 0.15)) lev1++;
            else if ((mi2[i][j] > 0.15) && (mi2[i][j] <= 0.5)) lev2++;
              // else if ((mi2[i][j] > 0.2) && (mi2[i][j] <= 0.5)) lev3++;
                //   else if ((mi2[i][j] > 1.0) && (mi2[i][j] <= 1.5)) lev4++;
                  //     else if ((mi2[i][j] > 1.5) && (mi2[i][j] <= 2.5)) lev5++;
                            else if (mi2[i][j] > 0.5) lev6++;
                                else printf("Error! %3d %3d %lf\n", i, j, mi2[i][j]);

        }
    }
    printf("%lf\n", lev);
   // printf("%lf %lf %lf %lf %lf %lf\n", lev1/lev*100, lev2/lev*100, lev3/lev*100,
lev4/lev*100, lev5/lev*100, lev6/lev*100);
    printf("%lf %lf %lf\n", lev1/lev*100, lev2/lev*100, lev6/lev*100);
   // printf("%lf\n",
lev1/lev*100+lev2/lev*100+lev3/lev*100+lev4/lev*100+lev5/lev*100+lev6/lev*100
);
    printf("%lf\n", lev1/lev*100+lev2/lev*100+lev6/lev*100);

    // Mean MI per residue
    for (i=0;i<nres;i++){
```

```
      sum[i] = 0.0;
      for (j=0;j<nres;j++){
         if (i!=j) sum[i] += mi2[i][j];
      }
      avg[i] = sum[i]/310.0;
   }
   // standard deviation
   for (i=0;i<nres;i++){
      std[i] = 0.0;
      for (j=0;j<nres;j++){
         if (i!=j) std[i] += (mi2[i][j] - avg[i])*(mi2[i][j] - avg[i]);
      }
      std[i] = sqrt(std[i]/310);
   }
   for (i=0;i<nres;i++) fprintf(fp7,"%3d %7.4lf %7.4lf\n", i, avg[i], std[i]);

  return 0;
}
```

J.4 C Code for Calculation of net TE from Cα Fluctuations

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <string.h>

int main(int argn, char *arg[]) {

   FILE *fd, *fp, *fdN, *fp1, *fp2, *fp3, *fp4, *fp5, *fp6, *fp7, *fp8;

   char line[1024], key[128], name[128], fname[128], resname[128],
**residue_type;

   int i, j, k, kk, m, n, ii, jj, nres, nconf, sw, tau, nbins, c;
   int resid, ncount, ntau, nbn, nb[400], nt[400], nt_type[20];
   double lev, lev1, lev2, lev3, lev4, lev5, lev6;
   double sum[311], avg[311], std[311];

   int nc_close[20][20], nc_far[20][20], sumnc_close[20], sumnc_far[20], total, *loc;
   int nc_close_tr[20][20], nc_far_tr[20][20], sumnc_close_tr[20],
sumnc_far_tr[20];
   int nc_close_trloop[20][20], nc_far_trloop[20][20], sumnc_close_trloop[20],
sumnc_far_trloop[20];
   int nc_close_loop[20][20], nc_far_loop[20][20], sumnc_close_loop[20],
sumnc_far_loop[20];

   double sumx[400], sumy[400], sumz[400], binsize, binsize_res1, binsize_res2;
   double **ax, **ay, **az, *xb, *yb, *zb, *ncount_j, *ncount_i, **ncount_ij, tot;
   double avgx[400], avgy[400], avgz[400], **dr, **pDelR, ****pDelRiRj;
   double **drx, **dry, **drz,  a, b, aa, bb, pi, pj, pj2, pk, **fin;
   double dridri[400], dridrj[400][400], Smin[400], Smax[400], **mi, **mi2;

   double sum_mi_close[20][20], sum_mi_far[20][20];
   double sum_mi_new_close[20][20], sum_mi_new_far[20][20], sumr_close[20],
sumr_far[20];
   double sum_mi_close_tr[20][20], sum_mi_far_tr[20][20],
sum_mi_close_trloop[20][20], sum_mi_far_trloop[20][20];
   double sum_mi_close_loop[20][20], sum_mi_far_loop[20][20];;
   double sum_mi_new_close_tr[20][20], sum_mi_new_far_tr[20][20],
sumr_close_tr[20], sumr_far_tr[20];
   double sum_mi_new_close_trloop[20][20], sum_mi_new_far_trloop[20][20],
sumr_close_trloop[20], sumr_far_trloop[20];
   double sum_mi_new_close_loop[20][20], sum_mi_new_far_loop[20][20],
sumr_close_loop[20], sumr_far_loop[20];

   const char *reslist[] = {"ALA", "GLY", "PRO", "CYS", "SER", "THR", "VAL", "ASN", "ASP",
"HIS", "ILE", "LEU", "PHE", "TRP", "TYR", "GLN", "GLU", "MET", "ARG", "LYS"};
   const int rotlist[] = {0, 0, 0, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4};

   //New rotlist grouping based on charges
   //Nonpolar, aliphatic 0, Ala, Gly, Pro, Val, Ile, Leu, (Met)
   //Nonpolar, aromatic 1, Phe, Trp, Tyr
   //Polar, uncharged 2, (Cys), Ser, Thr, Asn, Gln
   //Negatively, charged 3, Asp, Glu
   //Positively, charged 4, His, Arg, Lys

  // const int rotlist[] = {0, 0, 0, 2, 2, 2, 0, 2, 3, 4, 0, 0, 1, 1, 1, 2, 3, 0, 4, 4};

   // HYDROPHOBIC: 1
   // POLAR: 0
   const int reslist_HP[] = {1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0};

   if (argn!=3) {
      fprintf(stderr, "Usage: %s %s <filename>\n", arg[1], arg[2]);
      return 1;
   }

// open for reading only
   fd = fopen(arg[1], "r");
   fdN = fopen(arg[2], "r");

   if (fd == NULL) {
      fprintf(stderr, "Cannot open %s.\n", arg[1]);
      return 1;
   }

   // open for writing only
   fp = fopen("mutual_info_50ns-450ns_fin_testing.dat", "w");
   fp1 = fopen("mutual_info_50ns-450ns_residuepair_fin_testing.dat", "w");
   fp2 = fopen("mutual_info_50ns-450ns_residuepair_regions_far_fin_testing.dat",
"w");
   fp3 = fopen("mutual_info_50ns-450ns_residuepair_regions_close_fin_testing.dat",
"w");

fp4 = fopen("mutual_info_50ns-450ns_residue_fin_testing.dat", "w");
   fp5 = fopen("mutual_info_50ns-450ns_residue_regions_far_fin_testing.dat", "w");
   fp6 = fopen("mutual_info_50ns-450ns_residue_regions_close_fin_testing.dat",
"w");
   fp7 = fopen("mean_mutual_per_resid_50ns-450ns_fin_testing.dat", "w");
   fp8 = fopen("MI_phasel_testing.dat", "w");

// get the total number of lines
// so we allocate (more than) enough space for x,y,z arrays
   m = 0;
   while (fgets(line, 1024, fd)) {
      m++;
   }
   printf("%d\n",m);

   nconf = 2000;  // for 50 ns - 450 ns
   nres = 311;
   k = nconf*nres;
   printf("%d %d %d\n",nconf, nres, k);
   tau = 300;
   printf("Number of bins is %5d\n", nbins);
   nbins = 100;

// allocate memory
   residue_type = malloc(sizeof(char*)*nres);
   for (i=0;i<nres;i++) residue_type[i] = malloc(sizeof(char)*128);
   fin = malloc(sizeof(double*)*nres);
   for (i=0;i<nres;i++) fin[i] = malloc(sizeof(double)*nres);
   loc = malloc(sizeof(int)*nres);
   xb = malloc(sizeof(double)*k);
   yb = malloc(sizeof(double)*k);
   zb = malloc(sizeof(double)*k);
   ax = malloc(sizeof(double*)*nconf);
   ay = malloc(sizeof(double*)*nconf);
   az = malloc(sizeof(double*)*nconf);
   for (i=0;i<nconf;i++) ax[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) ay[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) az[i] = malloc(sizeof(double)*nres);
   drx = malloc(sizeof(double*)*nconf);
   dry = malloc(sizeof(double*)*nconf);
   drz = malloc(sizeof(double*)*nconf);
   dr = malloc(sizeof(double*)*nconf);
   for (i=0;i<nconf;i++) drx[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) dry[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) drz[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) dr[i] = malloc(sizeof(double)*nres);
   mi = malloc(sizeof(double*)*nres);
   for (i=0;i<nres;i++) mi[i] = malloc(sizeof(double)*nres);
   mi2 = malloc(sizeof(double*)*nres);
   for (i=0;i<nres;i++) mi2[i] = malloc(sizeof(double)*nres);
   ncount_i = malloc(sizeof(double)*nres);
   ncount_j = malloc(sizeof(double)*nres);
   ncount_ij = malloc(sizeof(double*)*nres);
   for (i=0;i<nres;i++){
      ncount_ij[i] = malloc(sizeof(double)*nres);
   }
   pDelR = malloc(sizeof(double*)*(nbins+1));
   for (i=0;i<(nbins+1);i++) pDelR[i] = malloc(sizeof(double)*nres);
   pDelRiRj = malloc(sizeof(double***)*(nbins+1));
   for (ii=0;ii<(nbins+1);ii++) pDelRiRj[ii] = malloc(sizeof(double**)*nres);
   for (ii=0;ii<(nbins+1);ii++){
      for (i=0;i<nres;i++) pDelRiRj[ii][i] = malloc(sizeof(double*)*(nbins+1));
   }
   for (ii=0;ii<(nbins+1);ii++){
      for (i=0;i<nres;i++){
         for (jj=0;jj<(nbins+1);jj++)
            pDelRiRj[ii][i][jj] = malloc(sizeof(double)*nres);
      }
   }
// now reading the number of bin values for every residue (from Shannon Entropy
calculations)
   while (fgets(line, 1024, fdN)) {
      sscanf(line, "%d %d %*s %*s %*s", &resid, &nbn);
      nb[resid] = nbn;
   }

// rewind to the beginning of the file
   rewind(fd);

// now reading the coordinates for every conformation in the trajectory
   m = 0;
   while (fgets(line, 1024, fd)) {
      sscanf(line, "%s %*s %s %s %*s %3d %lf %lf %lf %*s %*s", key, name,
resname, &resid, &(xb[m]), &(yb[m]), &(zb[m]));
      if (strcmp(key, "ATOM")==0) {
         if (strcmp(name, "CA")==0){
```

```
            if (m <= 310){
               if ((strcmp(resname, "HSE")!=0) && (strcmp(resname, "HSD")!=0))
                  strcpy(residue_type[m], resname);
               else
                  strcpy(residue_type[m], reslist[9]);
               if ( (((m+32)>=32) && ((m+32)<=60)) || (((m+32)>=67) &&
((m+32)<=95)) || (((m+32)>=103) && ((m+32)<=136)) || (((m+32)>=148)
&& ((m+32)<=171)) || (((m+32)>=197) && ((m+32)<=230)) ||
(((m+32)>=267) && ((m+32)<=298)) || (((m+32)>=305) &&
((m+32)<=328)) || (((m+32)>=330) && ((m+32)<=342)))
                  loc[m] = 1;
               else
                  loc[m] = 0;
            }
            m++;
         }
      }
   }

   for (k=0;k<20;k++) nt_type[k] = 0;;
   for (i=0;i<nres;i++){
      for (k=0;k<20;k++){
         if (strcmp(residue_type[i], reslist[k])==0)
            nt_type[k]++;
      }
   }
   m = 0;
   for (i=0;i<nconf;i++) {
      for (j=0;j<nres;j++) {
         ax[i][j] = xb[m];
         ay[i][j] = yb[m];
         az[i][j] = zb[m];
         m++;
      }
   }

// calculating the average position (over N snapshots) of each alpha Carbon atom in
residue i
   for (j=0;j<nres;j++){
      sumx[j] = 0.0;
      sumy[j] = 0.0;
      sumz[j] = 0.0;
      for (i=0;i<nconf;i++){
         sumx[j] = sumx[j] + ax[i][j];
         sumy[j] = sumy[j] + ay[i][j];
         sumz[j] = sumz[j] + az[i][j];
      }
      avgx[j] = sumx[j] / nconf;
      avgy[j] = sumy[j] / nconf;
      avgz[j] = sumz[j] / nconf;
   }

// calculating the fluctuation vector from the average for each alpha Carbon of residue i
for each snapshot (Delta Ri)
   for (i=0;i<nconf;i++) {
      for (j=0;j<nres;j++) {
         drx[i][j] = ax[i][j] - avgx[j];
         dry[i][j] = ay[i][j] - avgy[j];
         drz[i][j] = az[i][j] - avgz[j];
         dr[i][j] = sqrt(drx[i][j]*drx[i][j] + dry[i][j]*dry[i][j] + drz[i][j]*drz[i][j]);  //
magnitude of the vector |Delta R(i)|
      }
   }

//finite sample error calculation
   for(i=0;i<nres;i++){
      for(j=0;j<nres;j++){
         fin[i][j] = nb[i] + nb[j] -(nb[i] * nb[j]) - 1;
         fin[i][j] /= (2*nconf);
      }
   }


// now calculating the probability of observing Delta Ri in state k
// generating the histogram first

   // finding the min and the max values of |Delta R(i)|
   for (i=0;i<nres;i++) {
      Smin[i] = 1000.0;
      Smax[i] = -1000.0;
      for (k=0;k<nconf;k++){
         if (dr[k][i] < Smin[i])
            Smin[i] = dr[k][i];
         if (dr[k][i] > Smax[i])
            Smax[i] = dr[k][i];
      }
   }

// initialization
   for (i=0;i<nres;i++){
      for (k=0;k<(nb[i]+1);k++){
         pDelR[k][i] = 0.0;
      }
   }

   for (i=0;i<nres;i++){
      for (j=0;j<nres;j++){
         for (ii=0;ii<(nb[i]+1);ii++){
            for (jj=0;jj<(nb[j]+1);jj++)
               pDelRiRj[ii][i][jj][j] = 0.0;
         }
      }
   }

// calculating the mutual information for different nbins value for each residue
// first calculating the probability of the fluctuations of residue i to be in state k
   for (i=0;i<nres;i++) {
      printf("first i is %5d\n", i);
      binsize_res1 = (Smax[i]-Smin[i])/nb[i];
      ncount = 0;
      for (k=0;k<nconf;k++) {
         for (ii=0;ii<(nb[i]+1);ii++){
            a = Smin[i] + (ii-0.5)*binsize_res1;
            b = Smin[i] + (ii+0.5)*binsize_res1;
            if ((dr[k][i] > a) & (dr[k][i] <= b)) {
               ++pDelR[ii][i];
               ++ncount;
               ii = nb[i]+1;
            }
         }
      }
      //printf("%d %d\n", i, ncount);
   }

    // testing... total pDelR for each resiue should be equal to 1
   for (i=0;i<nres;i++) {
      tot = 0.0;
      for (ii=0;ii<nb[i]+1;ii++){
         pDelR[ii][i] /= nconf;
         tot += pDelR[ii][i];
      }
      printf("i and tot is %5d %5lf\n", i+32, tot);
   }

// now calculating the joint probability of the fluctuations of residue i to be in state l while
the same residue i is in state k, p(Delta_Ri(tk),DeltaRi(tk))
   for (i=0;i<nres;i++) {
      printf("second i is %5d\n", i);
      binsize_res1 = (Smax[i]-Smin[i])/nb[i];
      for (j=0;j<nres;j++){
         binsize_res2 = (Smax[j]-Smin[j])/nb[j];
         for (n=0;n<nconf;n++) {
            for (ii=0;ii<nb[i]+1;ii++){
               a = Smin[i] + (ii-0.5)*binsize_res1;
               b = Smin[i] + (ii+0.5)*binsize_res1;
               for (jj=0;jj<nb[j]+1;jj++){
                  aa = Smin[j] + (jj-0.5)*binsize_res2;
                  bb = Smin[j] + (jj+0.5)*binsize_res2;
                  if ((dr[n][i] > a) & (dr[n][i] <= b)){
                     if ((dr[n][j] > aa) & (dr[n][j] <= bb)){
                        ++pDelRiRj[ii][i][jj][j];
                        jj = nb[j]+1;
                        ii = nb[i]+1;
                     }
                  }
               }
            }
         }
      }
   }

// testing... total pDelR for each resiue should be equal to 1
   for (i=0;i<nres;i++) {
      for (j=0;j<nres;j++){
         tot = 0.0;
         for (ii=0;ii<(nb[i]+1);ii++){
            for (jj=0;jj<(nb[j]+1);jj++){
               pDelRiRj[ii][i][jj][j] /= nconf;
               tot += pDelRiRj[ii][i][jj][j];
            }
         }
         printf("i, j and tot is %5d %5d %5lf\n", i+32, j+32, tot);
      }
   }

// calculation of mutual information between residue i and j
   for (i=0;i<nres;i++){
      printf("%5d\n",i);
```

```
        for (j=0;j<nres;j++){
            mi[i][j] = 0.0;
            mi2[i][j] = 0.0;
            ncount_i[i] = 0.0;
            ncount_j[j] = 0.0;
            ncount_ij[i][j] = 0.0;
        // finding the bin of the current DeltaRi
            // finding the joint probability of res i to be in state l while residue j is in state k
            for (ii=0;ii<nb[i]+1;ii++){
                if (pDelR[ii][i]!=0) {
                    ++ncount_i[i];
                }
            }
            for (jj=0;jj<nb[j]+1;jj++){
                if (pDelR[jj][j]!=0) {
                    ++ncount_j[j];
                }
            }
            for (ii=0;ii<nb[i]+1;ii++){
                for (jj=0;jj<nb[j]+1;jj++){
                    if (pDelRiRj[ii][i][jj][j] !=0) {
                        ++ncount_ij[i][j];
                    }
                }
            }
            for (ii=0;ii<nb[i]+1;ii++){
                for (jj=0;jj<nb[j]+1;jj++){
                    if (pDelRiRj[ii][i][jj][j] !=0 & pDelR[ii][i]!=0 & pDelR[jj][j]!=0) {
                        pi = pDelR[ii][i];
                        pj = pDelR[jj][j];
                        pj2 = pDelRiRj[ii][i][jj][j];
                        mi[i][j] = pj2 * (log(pj2/(pi * pj))/log(2));
                        mi2[i][j] += mi[i][j];
                    }
                }
            }
            fin[i][j] = ncount_i[i] + ncount_j[j] - ncount_ij[i][j] - 1; //Finite sampling error
estimation based on # of non-zero bins
            fin[i][j] /= 2*nconf;
            printf("%d %d %d %d %lf %lf %lf %lf\n", i, j, ii, jj, ncount_i[i], ncount_j[j],
ncount_ij[i][j], fin[i][j]);
            fprintf(fp,"%5d %5d Before; %lf  ", i, j, mi2[i][j]);
            mi2[i][j] += fin[i][j];          // finite sampling correction

            fprintf(fp," After; %lf\n", mi2[i][j]);
        }
    }

    for (i=0;i<nres;i++){
        for (j=0;j<nres;j++){
            fprintf (fp8, "%d %d %lf %lf\n", i+32, j+32, mi2[i][j], mi2[j][i]);
        }
    }

    // determining the residue types
    // first initialization
    for (k=0;k<20;k++){
        for (kk=0;kk<20;kk++){
            nc_close[k][kk] = 0;
            sum_mi_close[k][kk] = 0.0;
            nc_close_loop[k][kk] = 0;
            sum_mi_close_loop[k][kk] = 0.0;
            nc_close_tr[k][kk] = 0;
            sum_mi_close_tr[k][kk] = 0.0;
            nc_close_trloop[k][kk] = 0;
            sum_mi_close_trloop[k][kk] = 0.0;
            nc_far[k][kk] = 0;
            sum_mi_far[k][kk] = 0.0;
            nc_far_loop[k][kk] = 0;
            sum_mi_far_loop[k][kk] = 0.0;
            nc_far_tr[k][kk] = 0;
            sum_mi_far_tr[k][kk] = 0.0;
            nc_far_trloop[k][kk] = 0;
            sum_mi_far_trloop[k][kk] = 0.0;
        }
    }

    double totMI_close = 0.0;
    double totMI_far = 0.0;

    for (i=0;i<nres;i++){
        for (j=i+1;j<nres;j++){
            for (k=0;k<20;k++){
                if (strcmp(residue_type[i], reslist[k])==0){
                    for (kk=0;kk<20;kk++){
                        if (strcmp(residue_type[j], reslist[kk])==0){
                            if (abs(i-j) < 5){
                                nc_close[k][kk]++;
                                sum_mi_close[k][kk] += mi2[i][j];
                                totMI_close += mi2[i][j];
                                if ((loc[i]+loc[j])==2){
                                    nc_close_tr[k][kk]++;
                                    sum_mi_close_tr[k][kk] += mi2[i][j];
                                }
                                else if ((loc[i]+loc[j])==1){
                                    nc_close_trloop[k][kk]++;
                                    sum_mi_close_trloop[k][kk] += mi2[i][j];
                                }
                                else if ((loc[i]+loc[j])==0){
                                    nc_close_loop[k][kk]++;
                                    sum_mi_close_loop[k][kk] += mi2[i][j];
                                }
                                else printf("Error %3d %3d!\n", loc[i], loc[j]);
                            }
                            if (abs(i-j) >= 5){
                                nc_far[k][kk]++;
                                sum_mi_far[k][kk] += mi2[i][j];
                                totMI_far += mi2[i][j];
                                if ((loc[i]+loc[j])==2){
                                    nc_far_tr[k][kk]++;
                                    sum_mi_far_tr[k][kk] += mi2[i][j];
                                }
                                else if ((loc[i]+loc[j])==1){
                                    nc_far_trloop[k][kk]++;
                                    sum_mi_far_trloop[k][kk] += mi2[i][j];
                                }
                                else if ((loc[i]+loc[j])==0){
                                    nc_far_loop[k][kk]++;
                                    sum_mi_far_loop[k][kk] += mi2[i][j];
                                }
                                else printf("Error %3d %3d!\n", loc[i], loc[j]);
                            }
                        }
                    }
                }
            }
        }
    }

    total = 0;
    for (k=0;k<20;k++){
        for (kk=0;kk<20;kk++){
            total += nc_close[k][kk] + nc_far[k][kk];
            printf("%3d %3d %s %s %7.4lf %3d %7.4lf %7.4lf %7.4lf %3d %7.4lf
%7.4lf\n", k, kk, reslist[k], reslist[kk], sum_mi_close[k][kk], nc_close[k][kk],
sum_mi_close[k][kk]/nc_close[k][kk], sum_mi_close[k][kk]/totMI_close*100.0,
sum_mi_far[k][kk], nc_far[k][kk], sum_mi_far[k][kk]/nc_far[k][kk],
sum_mi_far[k][kk]/totMI_far*100.0);
        }
    }
    printf("tot %3d\n", total);
    printf("total MI for close %7.4lf\n", totMI_close);
    printf("total MI for far %7.4lf\n", totMI_far);

    // combining i - j nd j - i pairs
    for (k=0;k<20;k++){
        sumr_close[k] = 0.0; sumr_far[k] = 0.0; sumnc_close[k] = 0; sumnc_far[k] =
0; sumr_close_tr[k] = 0.0; sumr_far_tr[k] = 0.0;
        sumnc_close_tr[k] = 0; sumnc_far_tr[k] = 0; sumr_close_trloop[k] = 0.0;
sumr_far_trloop[k] = 0.0;sumnc_close_trloop[k] = 0;
        sumnc_far_trloop[k] = 0;sumr_close_loop[k] = 0.0; sumr_far_loop[k] = 0.0;
sumnc_close_loop[k] = 0;sumnc_far_loop[k] = 0;

        for (kk=0;kk<20;kk++){
            sum_mi_new_close[k][kk] = sum_mi_close[k][kk] + sum_mi_close[kk][k];
            sumnc_close[k] += nc_close[k][kk] + nc_close[kk][k];
            sumr_close[k] += sum_mi_new_close[k][kk];

            sum_mi_new_far[k][kk] = sum_mi_far[k][kk] + sum_mi_far[kk][k];
            sumnc_far[k] += nc_far[k][kk] + nc_far[kk][k];
            sumr_far[k] += sum_mi_new_far[k][kk];

        // mutual information between transmembrane regions
            sum_mi_new_close_tr[k][kk] = sum_mi_close_tr[k][kk] +
sum_mi_close_tr[kk][k];
            sumnc_close_tr[k] += nc_close_tr[k][kk] + nc_close_tr[kk][k];
            sumr_close_tr[k] += sum_mi_new_close_tr[k][kk];

            sum_mi_new_far_tr[k][kk] = sum_mi_far_tr[k][kk] + sum_mi_far_tr[kk][k];
            sumnc_far_tr[k] += nc_far_tr[k][kk] + nc_far_tr[kk][k];
            sumr_far_tr[k] += sum_mi_new_far_tr[k][kk];

        // mutual information between transmembrane and loop regions
            sum_mi_new_close_trloop[k][kk] = sum_mi_close_trloop[k][kk] +
sum_mi_close_trloop[kk][k];
            sumnc_close_trloop[k] += nc_close_trloop[k][kk] + nc_close_trloop[kk][k];
            sumr_close_trloop[k] += sum_mi_new_close_trloop[k][kk];
```

```
            sum_mi_new_far_trloop[k][kk] = sum_mi_far_trloop[k][kk] +
sum_mi_far_trloop[kk][k];
            sumnc_far_trloop[k] += nc_far_trloop[k][kk] + nc_far_trloop[kk][k];
            sumr_far_trloop[k] += sum_mi_new_far_trloop[k][kk];

        // mutual information between loop regions
            sum_mi_new_close_loop[k][kk] = sum_mi_close_loop[k][kk] +
sum_mi_close_loop[kk][k];
            sumnc_close_loop[k] += nc_close_loop[k][kk] + nc_close_loop[kk][k];
            sumr_close_loop[k] += sum_mi_new_close_loop[k][kk];

            sum_mi_new_far_loop[k][kk] = sum_mi_far_loop[k][kk] +
sum_mi_far_loop[kk][k];
            sumnc_far_loop[k] += nc_far_loop[k][kk] + nc_far_loop[kk][k];
            sumr_far_loop[k] += sum_mi_new_far_loop[k][kk];
        }
    }

    for (k=0;k<20;k++){
        for (kk=0;kk<20;kk++){
            fprintf(fp1, "%3d %3d %3d %3d %s %3d %s %3d %8.4lf %5d %8.4lf %8.4lf
%8.4lf %5d %8.4lf %8.4lf\n", k, kk, rotlist[k], rotlist[kk], reslist[k], reslist_HP[k],
reslist[kk], reslist_HP[kk], sum_mi_new_close[k][kk],
nc_close[k][kk]+nc_close[kk][k],
sum_mi_new_close[k][kk]/(nc_close[k][kk]+nc_close[kk][k]),
sum_mi_new_close[k][kk]/totMI_close*100.0, sum_mi_new_far[k][kk],
nc_far[k][kk]+nc_far[kk][k], sum_mi_new_far[k][kk]/(nc_far[k][kk]+nc_far[kk][k]),
sum_mi_new_far[k][kk]/totMI_far*100.0);

            // mutual information between far residues in different regions (tr-tr / tr-loop /
loop-loop) regions
            fprintf(fp2, "%3d %3d %3d %3d %s %3d %s %3d %8.4lf %5d %8.4lf %8.4lf
%8.4lf %5d %8.4lf %8.4lf %8.4lf %5d %8.4lf %8.4lf\n", k, kk, rotlist[k], rotlist[kk],
reslist[k], reslist_HP[k], reslist[kk], reslist_HP[kk], sum_mi_new_far_tr[k][kk],
nc_far_tr[k][kk]+nc_far_tr[kk][k],
sum_mi_new_far_tr[k][kk]/(nc_far_tr[k][kk]+nc_far_tr[kk][k]),
sum_mi_new_far_tr[k][kk]/totMI_far*100.0, sum_mi_new_far_trloop[k][kk],
nc_far_trloop[k][kk]+nc_far_trloop[kk][k],
sum_mi_new_far_trloop[k][kk]/(nc_far_trloop[k][kk]+nc_far_trloop[kk][k]),
sum_mi_new_far_trloop[k][kk]/totMI_far*100.0, sum_mi_new_far_loop[k][kk],
nc_far_loop[k][kk]+nc_far_loop[kk][k],
sum_mi_new_far_loop[k][kk]/(nc_far_loop[k][kk]+nc_far_loop[kk][k]),
sum_mi_new_far_loop[k][kk]/totMI_far*100.0);

            // mutual information between close residues in different regions (tr-tr / tr-loop /
loop-loop)
            fprintf(fp3, "%3d %3d %3d %3d %s %3d %s %3d %8.4lf %5d %8.4lf %8.4lf
%8.4lf %5d %8.4lf %8.4lf %8.4lf %5d %8.4lf %8.4lf\n", k, kk, rotlist[k], rotlist[kk],
reslist[k], reslist_HP[k], reslist[kk], reslist_HP[kk], sum_mi_new_close_tr[k][kk],
nc_close_tr[k][kk]+nc_close_tr[kk][k],
sum_mi_new_close_tr[k][kk]/(nc_close_tr[k][kk]+nc_close_tr[kk][k]),
sum_mi_new_close_tr[k][kk]/totMI_close*100.0, sum_mi_new_close_trloop[k][kk],
nc_close_trloop[k][kk]+nc_close_trloop[kk][k],
sum_mi_new_close_trloop[k][kk]/(nc_close_trloop[k][kk]+nc_close_trloop[kk][k]),
sum_mi_new_close_trloop[k][kk]/totMI_close*100.0,
sum_mi_new_close_loop[k][kk], nc_close_loop[k][kk]+nc_close_loop[kk][k],
sum_mi_new_close_loop[k][kk]/(nc_close_loop[k][kk]+nc_close_loop[kk][k]),
sum_mi_new_close_loop[k][kk]/totMI_close*100.0);
        }
    }

    for (k=0;k<20;k++){
        fprintf(fp4, "%3d %3d %s %3d %8.4lf %5d %8.4lf %8.4lf %8.4lf %5d %8.4lf
%8.4lf\n", k, rotlist[k], reslist[k], reslist_HP[k], sumr_close[k], sumnc_close[k],
sumr_close[k]/sumnc_close[k], sumr_close[k]/totMI_close*100.0, sumr_far[k],
sumnc_far[k], sumr_far[k]/sumnc_far[k], sumr_far[k]/totMI_far*100.0);

        // residue based mutual information for far residues in different regions (tr-tr / tr-
loop / loop-loop) regions
        fprintf(fp5, "%3d %3d %s %3d %8.4lf %5d %8.4lf %8.4lf %8.4lf %5d %8.4lf
%8.4lf %8.4lf %5d %8.4lf %8.4lf\n", k, rotlist[k], reslist[k], reslist_HP[k],
sumr_far_tr[k], sumnc_far_tr[k], sumr_far_tr[k]/sumnc_far_tr[k],
sumr_far_tr[k]/totMI_far*100.0, sumr_far_trloop[k], sumnc_far_trloop[k],
sumr_far_trloop[k]/sumnc_far_trloop[k], sumr_far_trloop[k]/totMI_far*100.0,
sumr_far_loop[k], sumnc_far_loop[k], sumr_far_loop[k]/sumnc_far_loop[k],
sumr_far_loop[k]/totMI_far*100.0);

        // residue based mutual information for close residues in different regions (tr-tr / tr-
loop / loop-loop) regions
        fprintf(fp6, "%3d %3d %s %3d %8.4lf %5d %8.4lf %8.4lf %8.4lf %5d %8.4lf
%8.4lf %8.4lf %5d %8.4lf %8.4lf\n", k, rotlist[k], reslist[k], reslist_HP[k],
sumr_close_tr[k], sumnc_close_tr[k], sumr_close_tr[k]/sumnc_close_tr[k],
sumr_close_tr[k]/totMI_close*100.0, sumr_close_trloop[k], sumnc_close_trloop[k],
sumr_close_trloop[k]/sumnc_close_trloop[k],
sumr_close_trloop[k]/totMI_close*100.0, sumr_close_loop[k], sumnc_close_loop[k],
sumr_close_loop[k]/sumnc_close_loop[k], sumr_close_loop[k]/totMI_close*100.0);
    }

    // percentage of MI values in defined ranges
    lev = lev1 = lev2 = lev3 = lev4 = lev5 = lev6 = 0;
    for (i=0;i<nres;i++){
        for (j=i;j<nres;j++){
            lev++;
            if ((mi2[i][j] > 0.0) && (mi2[i][j] <= 0.02)) lev1++;
            else if ((mi2[i][j] > 0.02) && (mi2[i][j] <= 0.05)) lev2++;
                else if ((mi2[i][j] > 0.05) && (mi2[i][j] <= 0.1)) lev3++;
                    else if ((mi2[i][j] > 0.1) && (mi2[i][j] <= 1.0)) lev4++;
                        else if ((mi2[i][j] > 1.0) && (mi2[i][j] <= 3.0)) lev5++;
                            else if (mi2[i][j] > 3.0) lev6++;
                                else printf("Error! %3d %3d %lf\n", i, j, mi2[i][j]);

        }
    }

    // Mean MI per residue
    for (i=0;i<nres;i++){
        sum[i] = 0.0;
        for (j=0;j<nres;j++){
            if (i!=j) sum[i] += mi2[i][j];
        }
        avg[i] = sum[i]/310.0;
    }
    // standard deviation
    for (i=0;i<nres;i++){
        std[i] = 0.0;
        for (j=0;j<nres;j++){
            if (i!=j) std[i] += (mi2[i][j] - avg[i])*(mi2[i][j] - avg[i]);
        }
        std[i] = sqrt(std[i]/310);
    }
    for (i=0;i<nres;i++) fprintf(fp7,"%3d %7.4lf %7.4lf\n", i, avg[i], std[i]);

  return 0;
}
```

J.5 C Code for Calculation of Backbone Dihedral net TE

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <string.h>


int main(int argn, char *arg[]) {

   FILE *fd, *fdT, *fp, *fdN, *fp1, *fp2, *fp3, *fp4, *fp5, *fp6, *fp7, *fp8, *fp9;

   char line[1024], key[128], name[128], fname[128], resname[128],
**residue_type;

   int i, j, k, kk, m, n, ii, jj, ll, mm, nn, nres, nconf, sw, tau, nbins, nphi, npsi, c;
   int resid, ncount, ntau, nbn, nb[400], nt[400], nt_type[20];
   double lev, lev1, lev2, lev3, lev4, lev5, lev6, kb;
   double sum[311], avg[311], std[311];

   int nc_close[20][20], nc_far[20][20], sumnc_close[20], sumnc_far[20], total, *loc;
   int nc_close_tr[20][20], nc_far_tr[20][20], sumnc_close_tr[20],
sumnc_far_tr[20];
   int nc_close_trloop[20][20], nc_far_trloop[20][20], sumnc_close_trloop[20],
sumnc_far_trloop[20];
   int nc_close_loop[20][20], nc_far_loop[20][20], sumnc_close_loop[20],
sumnc_far_loop[20];

   double binsize, Smin, Smax;
   double **drx, **dry, **drz,  a, b, aa, bb, aaa, bbb, aaaa, bbbb, aaaaa, bbbbb,
aaaaaa, bbbbbb,  **fin, *ncount_j, **ncount_ij, *ncount_jj, **ncount_ijj;
   double **phi_, **psi_, **chi1_, **chi2_, **chi3_, **chi4_, **chi5_, *phi, *psi,
*chi1, *chi2, *chi3, *chi4, *chi5,tot;
   long double ***pdeg_phi_psi, ******pdeg_phi_psi_phi_psi_prev_iijj,
*****pdeg_phi_psi_phi_psi_prev_jjjj,
********pdeg_phi_psi_phi_psi_phi_psi_prev_iijjjj, NetTR, **tr, **tr2, **te, **te2, pj,
pij, pjj, pijj;

   double sum_tr_close[20][20], sum_tr_far[20][20];
   double sum_tr_new_close[20][20], sum_tr_new_far[20][20], sumr_close[20],
sumr_far[20];
   double sum_tr_close_tr[20][20], sum_tr_far_tr[20][20],
sum_tr_close_trloop[20][20], sum_tr_far_trloop[20][20];
   double sum_tr_close_loop[20][20], sum_tr_far_loop[20][20];;
   double sum_tr_new_close_tr[20][20], sum_tr_new_far_tr[20][20],
sumr_close_tr[20], sumr_far_tr[20];
   double sum_tr_new_close_trloop[20][20], sum_tr_new_far_trloop[20][20],
sumr_close_trloop[20], sumr_far_trloop[20];
   double sum_tr_new_close_loop[20][20], sum_tr_new_far_loop[20][20],
sumr_close_loop[20], sumr_far_loop[20];

   const char *reslist[] = {"ALA", "GLY", "PRO", "CYS", "SER", "THR", "VAL", "ASN", "ASP",
"HIS", "ILE", "LEU", "PHE", "TRP", "TYR", "GLN", "GLU", "MET", "ARG", "LYS"};
   const int rotlist[] = {0, 0, 0, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4};

   // HYDROPHOBIC: 1
   // POLAR: 0
   const int reslist_HP[] = {1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0};

   if (argn!=2) {
      fprintf(stderr, "Usage: %s <filename>\n", arg[1]);
      return 1;
   }

// open for reading only
   fd = fopen(arg[1], "r");

   if (fd == NULL) {
      fprintf(stderr, "Cannot open %s.\n", arg[1]);
      return 1;
   }

   // open for writing only
   fp = fopen("TE_phasel_backbone_minmut_sw25.dat", "w");

   fp1 = fopen("TE_phasel_backbone_residuepair_minmut_sw25.dat", "w");
   fp2 = fopen("TE_phasel_backbone_residuepair_regions_far_minmut_sw25.dat",
"w");
   fp3 = fopen("TE_phasel_backbone_residuepair_regions_close_minmut_sw25.dat",
"w");

   fp4 = fopen("TE_phasel_backbone_residue_minmut_sw25.dat", "w");
   fp5 = fopen("TE_phasel_backbone_residue_regions_far_minmut_sw25.dat", "w");
   fp6 = fopen("TE_phasel_backbone_residue_regions_close_minmut_sw25.dat", "w");

   fp7 = fopen("mean_TE_per_resid_phasel_backbone_minmut_sw25.dat", "w");
   fp8 = fopen("test_TE_phasel_minmut_sw25.dat", "w");
   fp9 = fopen("TE_phasel_probabilities_backbone_minmut_sw25.dat", "w");

// get the total number of lines
// so we allocate (more than) enough space for x,y,z arrays
   m = 0;
   while (fgets(line, 1024, fd)) {
      m++;
   }
   printf("%d\n",m);

  // nconf = 2000;  // for 50ns -450ns
   nconf = 3750;  // for 750 ns - 1500 ns
   nres = 309;
   k = nconf*nres;
   printf("%d %d %d\n",nconf, nres, k);
   tau = 300;

   //nbins = round(1 + log(nconf));  // Sturge's rule number of bins is equal to
(1+log2(number of data)
   printf("Number of bins is %5d\n", nbins);
   nbins = 4;
   nphi = 3;
   npsi = 3;
   binsize = 120;
   Smin = 0;
   Smax = 360;
   sw = 25;

// allocate memory
   residue_type = malloc(sizeof(char*)*nres);
   for (i=0;i<nres;i++) residue_type[i] = malloc(sizeof(char)*128);

   fin = malloc(sizeof(double*)*nres);
   for (i=0;i<nres;i++) fin[i] = malloc(sizeof(double)*nres);

   loc = malloc(sizeof(int)*nres);

   phi = malloc(sizeof(double)*k);
   psi = malloc(sizeof(double)*k);
   chi1 = malloc(sizeof(double)*k);
   chi2 = malloc(sizeof(double)*k);
   chi3 = malloc(sizeof(double)*k);
   chi4 = malloc(sizeof(double)*k);
   chi5 = malloc(sizeof(double)*k);
   printf ("memory alloc begins\n");
   phi_ = malloc(sizeof(double*)*nconf);
   psi_ = malloc(sizeof(double*)*nconf);
   chi1_ = malloc(sizeof(double*)*nconf);
   chi2_ = malloc(sizeof(double*)*nconf);
   chi3_ = malloc(sizeof(double*)*nconf);
   chi4_ = malloc(sizeof(double*)*nconf);
   chi5_ = malloc(sizeof(double*)*nconf);


   for (i=0;i<nconf;i++) phi_[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) psi_[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) chi1_[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) chi2_[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) chi3_[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) chi4_[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nconf;i++) chi5_[i] = malloc(sizeof(double)*nres);


  printf ("memory alloc cont.s\n");
   pdeg_phi_psi = malloc(sizeof(long double**)*(nres));
   pdeg_phi_psi_phi_psi_prev_iijj = malloc(sizeof(long double*****)*(nres));
   pdeg_phi_psi_phi_psi_prev_jjjj = malloc(sizeof(long double****)*(nres));
   pdeg_phi_psi_phi_psi_phi_psi_prev_iijjjj = malloc(sizeof(long
double*******)*(nres));

   for (i=0;i<(nres);i++) {
      pdeg_phi_psi[i] = malloc(sizeof(long double*)*(nbins+1));
      pdeg_phi_psi_phi_psi_prev_iijj[i] = malloc(sizeof(long double****)*(nres));
      pdeg_phi_psi_phi_psi_prev_jjjj[i] = malloc(sizeof(long double***)*(nbins+1));
```

```
      pdeg_phi_psi_phi_psi_phi_psi_prev_iiiiii[i] = malloc(sizeof(long
double******)*(nres));
   }
   for (i=0;i<(nres);i++) {
      for (ii=0;ii<(nbins+1);ii++){
         pdeg_phi_psi[i][ii] = malloc(sizeof(long double)*(nbins+1));
         pdeg_phi_psi_phi_psi_prev_jjjj[i][ii] = malloc(sizeof(long
double**)*(nbins+1));
      }
   }
   for (i=0;i<(nres);i++) {
      for (ii=0;ii<(nbins+1);ii++){
         for (jj=0;jj<(nbins+1);jj++){
            pdeg_phi_psi_phi_psi_prev_jjjj[i][ii][jj] = malloc(sizeof(long
double*)*(nbins+1));
         }
      }
   }
   for (i=0;i<(nres);i++) {
      for (ii=0;ii<(nbins+1);ii++){
         for (jj=0;jj<(nbins+1);jj++){
            for (kk=0;kk<(nbins+1);kk++){
               pdeg_phi_psi_phi_psi_prev_jjjj[i][ii][jj][kk] = malloc(sizeof(long
double)*(nbins+1));
            }
         }
      }
   }

   for (i=0;i<(nres);i++){
      for (j=0;j<(nres);j++){
         pdeg_phi_psi_phi_psi_prev_iiij[i][j] = malloc(sizeof(long
double***)*(nbins+1));
         pdeg_phi_psi_phi_psi_phi_psi_prev_iiiiij[i][j] = malloc(sizeof(long
double*****)*(nbins+1));
      }
   }

   for (i=0;i<(nres);i++){
      for (j=0;j<(nres);j++){
         for (ii=0;ii<(nbins+1);ii++){
            pdeg_phi_psi_phi_psi_prev_iiij[i][j][ii] = malloc(sizeof(long
double**)*(nbins+1));
            pdeg_phi_psi_phi_psi_phi_psi_prev_iiiiij[i][j][ii] = malloc(sizeof(long
double****)*(nbins+1));
         }
      }
   }
   for (i=0;i<(nres);i++){
      for (j=0;j<(nres);j++){
         for (ii=0;ii<(nbins+1);ii++){
            for (jj=0;jj<(nbins+1);jj++){
               pdeg_phi_psi_phi_psi_prev_iiij[i][j][ii][jj] = malloc(sizeof(long
double*)*(nbins+1));
               pdeg_phi_psi_phi_psi_phi_psi_prev_iiiiij[i][j][ii][jj] = malloc(sizeof(long
double***)*(nbins+1));
            }
         }
      }
   }
    printf ("Malloc\n");

   for (i=0;i<(nres);i++){
     //  printf ("malloc i is %d\n", i);
      for (j=0;j<(nres);j++){
         for (ii=0;ii<(nbins+1);ii++){
            for (jj=0;jj<(nbins+1);jj++){
               for (kk=0;kk<(nbins+1);kk++){
                  pdeg_phi_psi_phi_psi_prev_iiij[i][j][ii][jj][kk] = malloc(sizeof(long
double)*(nbins+1));
                  pdeg_phi_psi_phi_psi_phi_psi_prev_iiiiij[i][j][ii][jj][kk] =
malloc(sizeof(long double**)*(nbins+1));
               }
            }
         }
      }
   }
   for (i=0;i<(nres);i++){
    //    printf ("malloc 2 i is %d\n", i);
      for (j=0;j<(nres);j++){
         for (ii=0;ii<(nbins+1);ii++){
            for (jj=0;jj<(nbins+1);jj++){
               for (kk=0;kk<(nbins+1);kk++){
                  for (ll=0;ll<(nbins+1);ll++){
                     pdeg_phi_psi_phi_psi_phi_psi_prev_iiiiij[i][j][ii][jj][kk][ll] =
malloc(sizeof(long double*)*(nbins+1));
                  }
               }
            }
         }
      }
   }
   for (i=0;i<(nres);i++){
     //   printf ("malloc 3 i is %d\n", i);
      for (j=0;j<(nres);j++){
         for (ii=0;ii<(nbins+1);ii++){
            for (jj=0;jj<(nbins+1);jj++){
               for (kk=0;kk<(nbins+1);kk++){
                  for (ll=0;ll<(nbins+1);ll++){
                     for (mm=0;mm<(nbins+1);mm++){
                        pdeg_phi_psi_phi_psi_phi_psi_prev_iiiiij[i][j][ii][jj][kk][ll][mm]
= malloc(sizeof(long double)*(nbins+1));
                     }
                  }
               }
            }
         }
      }
   }

   tr = malloc(sizeof(long double*)*nres);
   tr2 = malloc(sizeof(long double*)*nres);
   te = malloc(sizeof(long double*)*nres);
   te2 = malloc(sizeof(long double*)*nres);

   ncount_ijj = malloc(sizeof(double*)*nres);
   ncount_ij = malloc(sizeof(double*)*nres);
   ncount_jj = malloc(sizeof(double)*nres);
   ncount_j = malloc(sizeof(double*)*nres);

   for (i=0;i<nres;i++) tr[i] = malloc(sizeof(long double)*nres);
   for (i=0;i<nres;i++) tr2[i] = malloc(sizeof(long double)*nres);
   for (i=0;i<nres;i++) te[i] = malloc(sizeof(long double)*nres);
   for (i=0;i<nres;i++) te2[i] = malloc(sizeof(long double)*nres);
   for (i=0;i<nres;i++) ncount_ijj[i] = malloc(sizeof(double)*nres);
   for (i=0;i<nres;i++) ncount_ij[i] = malloc(sizeof(double)*nres);


// rewind to the beginning of the file
   rewind(fd);

// now reading the coordinates for every conformation in the trajectory
   m = 0;
   while (fgets(line, 1024, fd)) {
      sscanf(line, "%*s %s %*s %*s %12lf %*s %12lf %*s %*lf %*s %*lf %*s %*lf
%*s %*lf %*s %*lf", resname, &(phi[m]), &(psi[m]));
            if (m <= 308){
               if ((strcmp(resname, "HSE")!=0) && (strcmp(resname, "HSD")!=0))
                  strcpy(residue_type[m], resname);
               else
                  strcpy(residue_type[m], reslist[9]);
               if ( (((m+33)>=33) && ((m+33)<=60)) || (((m+33)>=67) &&
((m+33)<=95)) || (((m+33)>=103) && ((m+33)<=136)) || (((m+33)>=148)
&& ((m+33)<=171)) || (((m+33)>=197) && ((m+33)<=230)) ||
(((m+33)>=267) && ((m+33)<=298)) || (((m+33)>=305) &&
((m+33)<=328)) || (((m+33)>=330) && ((m+33)<=341)))
                  loc[m] = 1;
                  else
                  loc[m] = 0;
            }
            m++;
         }

   m = 0;
   for (i=0;i<nconf;i++) {
      for (j=0;j<nres;j++) {
         phi_[i][j] = phi[m];
         psi_[i][j] = psi[m];
         m++;
         //  printf(" i is %d j is %d phi is %lf psi is %lf\n", i, j, phi_[i][j], psi_[i][j]);
      }
   }

   printf("%d\n",m);

   for (i=0;i<nres;i++) {
      printf("%d %d\n", i+33, loc[i]);
   }

   for (k=0;k<20;k++){
      nt_type[k] = 0;
   }

   for (i=0;i<nres;i++){
      printf("res type %d %s\n", i, residue_type[i]);
      for (k=0;k<20;k++){
         if (strcmp(residue_type[i], reslist[k])==0)
```

```
            nt_type[k]++;
        }
    }

    printf("frequency\n");
    for (k=0;k<20;k++){
        printf("%s %lf\n", reslist[k], nt_type[k]/309.0);
    }


    // initializations
// Initialization of  H(i_phi, i_psi)
    for (j=0;j<nres;j++){
        for (jj=0;jj<nphi;jj++){
            for (kk=0;kk<npsi;kk++){
                pdeg_phi_psi[j][jj][kk] = 0.0;
            }
        }
    }


    printf("Initialization 1 done!\n");

  // Initialization of  H(i_phi, i_psi, j_phi, j_psi)
    for (i=0;i<nres;i++) {
        for (j=0;j<nres;j++) {
            for (ii=0;ii<nphi;ii++){
                for (jj=0;jj<npsi;jj++){
                    for (kk=0;kk<nphi;kk++){
                        for (ll=0;ll<npsi;ll++){
                            pdeg_phi_psi_phi_psi_prev_iiijj[i][j][ii][jj][kk][ll] = 0.0;
                        }
                    }
                }
            }
        }
    }

    printf("Initialization 2 done!!\n");

    for (j=0;j<nres;j++) {
        for (ii=0;ii<nphi;ii++){
            for (jj=0;jj<npsi;jj++){
                for (kk=0;kk<nphi;kk++){
                    for (ll=0;ll<npsi;ll++){
                        pdeg_phi_psi_phi_psi_prev_jjjj[j][ii][jj][kk][ll] = 0.0;
                    }
                }
            }
        }
    }
    printf("Initialization 3 done!!\n");
    for (i=0;i<nres;i++) {
      //  printf("i is %d\n", i);
        for (j=0;j<nres;j++) {
            for (ii=0;ii<nphi;ii++){
                for (jj=0;jj<npsi;jj++){
                    for (kk=0;kk<nphi;kk++){
                        for (ll=0;ll<npsi;ll++){
                            for (mm=0;mm<nphi;mm++){
                                for (nn=0;nn<npsi;nn++){

pdeg_phi_psi_phi_psi_phi_psi_prev_iijjjj[i][j][ii][jj][kk][ll][mm][nn] = 0.0;
                                }
                            }
                        }
                    }
                }
            }
        }
    }
    printf("Initialization 4 done!!\n");


// first calculating H(i_nphi, i_npsi, j_nphi,j_npsi)

    for (j=0;j<nres;j++) {
        //printf("first i is %5d\n", i);
        ncount = 0;
        for (k=0;k<(nconf-sw);k++) {
            for (ii=0;ii<nphi;ii++){
                a = (ii)*binsize;
                b = (ii+1)*binsize;
                for (jj=0;jj<npsi;jj++){
                    aa = (jj)*binsize;
                    bb = (jj+1)*binsize;
                    if ((phi_[k][j] >= a) & (phi_[k][j] < b)) {
                        if ((psi_[k][j] >= aa) & (psi_[k][j] < bb)) {
                            ++pdeg_phi_psi[j][ii][jj];
                            ++ncount;
                            jj = npsi;
                            ii = nphi;
                        }
                    }
                }
            }
        }
    }
    printf("%d %d\n", j, ncount);
}

for (j=0;j<nres;j++) {
    for (ii=0;ii<nphi;ii++) {
        for (jj=0;jj<npsi;jj++) {
            pdeg_phi_psi[j][ii][jj] /= (nconf-sw);
        }
    }
}


// testing...
tot = 0.0;
for (j=0;j<nres;j++) {
    tot = 0.0;
    for (ii=0;ii<nphi;ii++){
        for (jj=0;jj<npsi;jj++){
            tot += pdeg_phi_psi[j][ii][jj];
        }
    }
    printf("i and tot phi_psi is %d %lf\n", j+33, tot);
}

for (i=0;i<nres;i++) {
    printf("i for phi,psi,phi,psi is %5d\n", i);
    ncount = 0;
    for (j=0;j<nres;j++) {
        for (n=0;n<(nconf-sw);n++) {
            for (ii=0;ii<nphi;ii++){
                a = (ii)*binsize;
                b = (ii+1)*binsize;
                for (jj=0;jj<npsi;jj++){
                    aa = (jj)*binsize;
                    bb = (jj+1)*binsize;
                    for (kk=0;kk<nphi;kk++){
                        aaa = (kk)*binsize;
                        bbb = (kk+1)*binsize;
                        for (ll=0;ll<npsi;ll++){
                            aaaa = (ll)*binsize;
                            bbbb = (ll+1)*binsize;
                            if ((phi_[n][i] >= a) & (phi_[n][i] < b)) {
                                if ((psi_[n][i] >= aa) & (psi_[n][i] < bb)){
                                    if ((phi_[n][j] >= aaa) & (phi_[n][j] < bbb)){
                                        if ((psi_[n][j] >= aaaa) & (psi_[n][j] < bbbb)){
                                            ++pdeg_phi_psi_phi_psi_prev_iijj[i][j][ii][jj][kk][ll];
                                            ++ncount;
                                            ll = npsi;
                                            kk = nphi;
                                            jj = npsi;
                                            ii = nphi;

                                        }
                                    }
                                }
                            }
                        }
                    }
                }
            }
        }
    }
  //  printf("%d %d\n", i, ncount);
}
for (i=0;i<nres;i++) {
    for (j=0;j<nres;j++) {
        for (ii=0;ii<nphi;ii++){
            for (jj=0;jj<npsi;jj++){
                for (kk=0;kk<nphi;kk++){
                    for (ll=0;ll<npsi;ll++){
                        pdeg_phi_psi_phi_psi_prev_iijj[i][j][ii][jj][kk][ll] /= (nconf-sw);

                    }
                }
            }
        }
    }
}
    // testing .....
tot = 0.0;
for (i=0;i<nres;i++) {
```

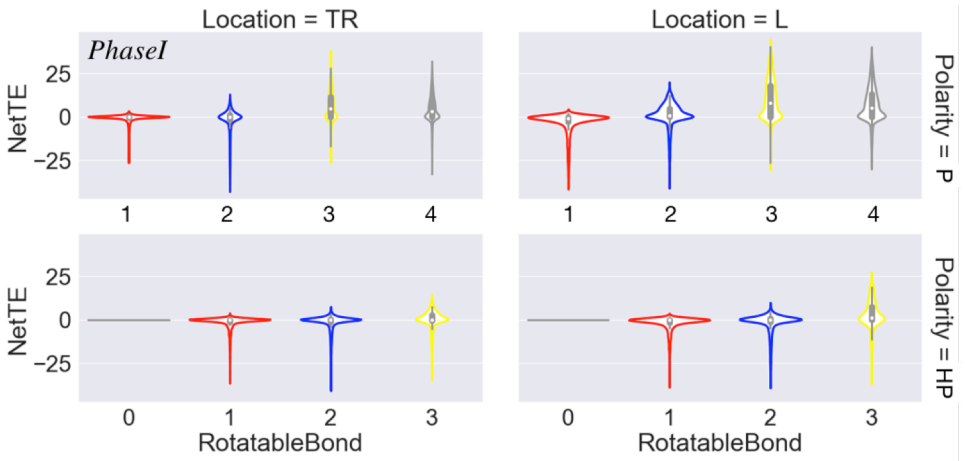
```
    for (j=0;j<nres;j++) {
       tot = 0.0;
       for (ii=0;ii<nphi;ii++){
          for (jj=0;jj<npsi;jj++){
             for (kk=0;kk<nphi;kk++){
                for (ll=0;ll<npsi;ll++){
                   tot += pdeg_phi_psi_phi_psi_prev_iijj[i] [j] [ii] [jj] [kk] [ll];
                }
             }
          }
       }
        printf("testing total i %d and j %d and phi_psi_phi_psi %7.4lf\n", i, j, tot);
    }
}


for (j=0;j<nres;j++) {
    printf("j for phi,psi,phi,psi is %5d\n", j);
    ncount = 0;
    for (n=0;n<(nconf-sw);n++) {
       for (ii=0;ii<nphi;ii++){
          a = (ii)*binsize;
          b = (ii+1)*binsize;
          for (jj=0;jj<npsi;jj++){
             aa = (jj)*binsize;
             bb = (jj+1)*binsize;
             for (kk=0;kk<nphi;kk++){
                aaa = (kk)*binsize;
                bbb = (kk+1)*binsize;
                for (ll=0;ll<npsi;ll++){
                   aaaa = (ll)*binsize;
                   bbbb = (ll+1)*binsize;
                   if ((phi_[n] [j] >= a) & (phi_[n] [j] < b)) {
                      if ((psi_[n] [j] >= aa) & (psi_[n] [j] < bb)){
                         if ((phi_[n+sw] [j] >= aaa) & (phi_[n+sw] [j] < bbb)){
                            if ((psi_[n+sw] [j] >= aaaa) & (psi_[n+sw] [j] < bbbb)){
                               ++pdeg_phi_psi_phi_psi_prev_jjjj[j] [ii] [jj] [kk] [ll];
                               ++ncount;
                               ll = npsi;
                               kk = nphi;
                               jj = npsi;
                               ii = nphi;
                            }
                         }
                      }
                   }
                }
             }
          }
       }
    }
    //  printf("%d %d\n", i, ncount);
}

for (j=0;j<nres;j++) {
    for (ii=0;ii<nphi;ii++){
       for (jj=0;jj<npsi;jj++){
          for (kk=0;kk<nphi;kk++){
             for (ll=0;ll<npsi;ll++){
                pdeg_phi_psi_phi_psi_prev_jjjj[j] [ii] [jj] [kk] [ll] /= (nconf-sw);

             }
          }
       }
    }
}

// testing .....
tot = 0.0;
for (j=0;j<nres;j++) {
    tot = 0.0;
    for (ii=0;ii<nphi;ii++){
       for (jj=0;jj<npsi;jj++){
          for (kk=0;kk<nphi;kk++){
             for (ll=0;ll<npsi;ll++){
                tot += pdeg_phi_psi_phi_psi_prev_jjjj[j] [ii] [jj] [kk] [ll];
             }
          }
       }
    }
    printf("testing total j %d and phi_psi_phi_psi %7.4lf\n", j, tot);
}


for (i=0;i<nres;i++) {
    printf("i for phi,psi,phi,psi,phi,psi is %5d\n", i);
    ncount = 0;
    for (j=0;j<nres;j++) {
       for (n=0;n<(nconf-sw);n++) {
          for (ii=0;ii<nphi;ii++){
             a = (ii)*binsize;
             b = (ii+1)*binsize;
             for (jj=0;jj<npsi;jj++){
                aa = (jj)*binsize;
                bb = (jj+1)*binsize;
                for (kk=0;kk<nphi;kk++){
                   aaa = (kk)*binsize;
                   bbb = (kk+1)*binsize;
                   for (ll=0;ll<npsi;ll++){
                      aaaa = (ll)*binsize;
                      bbbb = (ll+1)*binsize;
                      for (mm=0;mm<nphi;mm++){
                         aaaaa = (mm)*binsize;
                         bbbbb = (mm+1)*binsize;
                         for (nn=0;nn<npsi;nn++){
                            aaaaaa = (nn)*binsize;
                            bbbbbb = (nn+1)*binsize;
                            if ((phi_[n] [i] >= a) & (phi_[n] [i] < b)) {
                               if ((psi_[n] [i] >= aa) & (psi_[n] [i] < bb)){
                                  if ((phi_[n] [j] >= aaa) & (phi_[n] [j] < bbb)){
                                     if ((psi_[n] [j] >= aaaa) & (psi_[n] [j] < bbbb)){
                                        if ((phi_[n+sw] [j] >= aaaaa) & (phi_[n+sw] [j]
< bbbbb)){
                                           if ((psi_[n+sw] [j] >= aaaaaa) &
(psi_[n+sw] [j] < bbbbbb)){

++pdeg_phi_psi_phi_psi_phi_psi_prev_iijjjj[i] [j] [ii] [jj] [kk] [ll] [mm] [nn];
                                              ++ncount;
                                              nn = npsi;
                                              mm = nphi;
                                              ll = npsi;
                                              kk = nphi;
                                              jj = npsi;
                                              ii = nphi;
                                           }
                                        }
                                     }
                                  }
                               }
                            }
                         }
                      }
                   }
                }
             }
          }
       }
       //  printf("%d %d\n", i, ncount);
    }
    for (i=0;i<nres;i++) {
       for (j=0;j<nres;j++) {
          for (ii=0;ii<nphi;ii++){
             for (jj=0;jj<npsi;jj++){
                for (kk=0;kk<nphi;kk++){
                   for (ll=0;ll<npsi;ll++){
                      for (mm=0;mm<nphi;mm++){
                         for (nn=0;nn<npsi;nn++){

pdeg_phi_psi_phi_psi_phi_psi_prev_iijjjj[i] [j] [ii] [jj] [kk] [ll] [mm] [nn] /= (nconf-sw);
                         }
                      }
                   }
                }
             }
          }
       }
    }
    // testing .....
    tot = 0.0;
    for (i=0;i<nres;i++) {
       for (j=0;j<nres;j++) {
          tot = 0.0;
          for (ii=0;ii<nphi;ii++){
             for (jj=0;jj<npsi;jj++){
                for (kk=0;kk<nphi;kk++){
                   for (ll=0;ll<npsi;ll++){
                      for (mm=0;mm<nphi;mm++){
                         for (nn=0;nn<npsi;nn++){
                            tot +=
pdeg_phi_psi_phi_psi_phi_psi_prev_iijjjj[i] [j] [ii] [jj] [kk] [ll] [mm] [nn];
                         }
                      }
                   }
                }
             }
          }
          printf("testing total i %d and j %d and phi_psi_phi_psi %7.4lf\n", i, j, tot);
```

```
    }
  }

// calculation of mutual information between residue i and j
  for (i=0;i<nres;i++){
    printf("%5d\n",i);
    for (j=0;j<nres;j++){
      tr[i][j] = 0.0;
      tr2[i][j] = 0.0;
      te[i][j] = 0.0;
      te2[i][j] = 0.0;
      ncount_iij[i][j] = 0.0;
      ncount_ij[i][j] = 0.0;
      ncount_jj[j] = 0.0;
      ncount_j[j] = 0.0;
      fin[i][j] = 0.0;
    // finding the bin of the current Phi,Psi
      for (ii=0;ii<nphi;ii++){
        for (jj=0;jj<npsi;jj++){
          if (pdeg_phi_psi[j][ii][jj] !=0) {
            ++ncount_j[j];
          }
        }
      }
      for (ii=0;ii<nphi;ii++){
        for (jj=0;jj<npsi;jj++){
          for (kk=0;kk<nphi;kk++){
            for (ll=0;ll<npsi;ll++){
              if (pdeg_phi_psi_phi_psi_prev_iijj[i][j][ii][jj][kk][ll] !=0) {
                ++ncount_ij[i][j];
              }
            }
          }
        }
      }
      for (ii=0;ii<nphi;ii++){
        for (jj=0;jj<npsi;jj++){
          for (kk=0;kk<nphi;kk++){
            for (ll=0;ll<npsi;ll++){
              if (pdeg_phi_psi_phi_psi_prev_jjjj[j][ii][jj][kk][ll] !=0) {
                ++ncount_jj[j];
              }
            }
          }
        }
      }
      for (ii=0;ii<nphi;ii++){
        for (jj=0;jj<npsi;jj++){
          for (kk=0;kk<nphi;kk++){
            for (ll=0;ll<npsi;ll++){
              for (mm=0;mm<nphi;mm++){
                for (nn=0;nn<npsi;nn++){
                  if
(pdeg_phi_psi_phi_psi_phi_psi_prev_iijjjj[i][j][ii][jj][kk][ll][mm][nn] !=0) {
                    ++ncount_ijj[i][j];
                  }
                }
              }
            }
          }
        }
      }
      for (ii=0;ii<nphi;ii++){
        for (jj=0;jj<npsi;jj++){
          for (kk=0;kk<nphi;kk++){
            for (ll=0;ll<npsi;ll++){
              for (mm=0;mm<nphi;mm++){
                for (nn=0;nn<npsi;nn++){
                  if (pdeg_phi_psi[j][kk][ll] !=0 &
pdeg_phi_psi_phi_psi_prev_iijj[i][j][ii][jj][kk][ll] !=0 &
pdeg_phi_psi_phi_psi_prev_jjjj[j][kk][ll][mm][nn] !=0 &
pdeg_phi_psi_phi_psi_phi_psi_prev_iijjjj[i][j][ii][jj][kk][ll][mm][nn] !=0) {
                    pj = pdeg_phi_psi[j][kk][ll];
                    pjj = pdeg_phi_psi_phi_psi_prev_jjjj[j][kk][ll][mm][nn];
                    pij = pdeg_phi_psi_phi_psi_prev_iijj[i][j][ii][jj][kk][ll];
                    pijj =
pdeg_phi_psi_phi_psi_phi_psi_prev_iijjjj[i][j][ii][jj][kk][ll][mm][nn];
                    tr[i][j] = ((pjj * ((log(pj)/log(2)) - (log(pjj)/log(2)))) + (pijj
* ((log(pijj)/log(2)) - (log(pij)/log(2)))));
                    //  fprintf (fp9, "resi %d resj %d nphi_i(t-tau) %d npsi_i(t-tau)
%d nphi_j(t-tau) %d npsi_j(t-tau) %d nphi_i(t) %d npsi_i(t) %d pj %Lf pjj %Lf pij %Lf
pijj %Lf mi %Lf\n", i, j, ii, jj, kk, ll, mm, nn, pj, pjj, pij, pijj, tr[i][j]);
                    tr2[i][j] += tr[i][j];
                    te[i][j] = (pijj * (log ((pijj * pj) / (pij * pjj)) / log(2)));
                    te2[i][j] += te[i][j];

                  }
                }
              }
            }
          }
        }
      }
      fin[i][j] = ncount_ij[i][j] + ncount_jj[j] - ncount_iij[i][j] - ncount_j[j];
      fin[i][j] /= 2*nconf;
      printf("%d %d %d %d %lf %lf %lf %lf %lf\n", i, j, ii, jj, ncount_j[j], ncount_jj[j],
ncount_ij[i][j], ncount_iij[i][j], fin[i][j]);
      fprintf(fp,"%5d %5d before: tr %7.2Lf before: te %7.2Lf fin: %7.2lf  ", i, j,
tr2[i][j], te2[i][j], fin[i][j]);
      if (tr2[i][j] < -fin[i][j])
        tr2[i][j] = 0.0;
      else
        tr2[i][j] += fin[i][j];// finite sampling correction
      if (te2[i][j] < -fin[i][j])
        te2[i][j] = 0.0;
      else
        te2[i][j] += fin[i][j];
      fprintf(fp,"after: %7.2Lf %7.2Lf\n", tr2[i][j], te2[i][j]);
    }
  }

  for (i=0;i<nres;i++){
    for (j=0;j<nres;j++){
      NetTR = tr2[i][j]-tr2[j][i];
      fprintf(fp8," %5d %5d %7.2Lf %7.2Lf %7.2Lf %7.2Lf %7.2Lf %7.2Lf %7.2Lf\n", i, j,
tr2[i][j], tr2[j][i], NetTR, te2[i][j], te2[j][i], (te2[i][j] - te2[j][i]));
    }
  }
  // determining the residue types
  // first initialization
  for (k=0;k<20;k++){
    for (kk=0;kk<20;kk++){
      nc_close[k][kk] = 0;
      sum_tr_close[k][kk] = 0.0;
      nc_close_loop[k][kk] = 0;
      sum_tr_close_loop[k][kk] = 0.0;
      nc_close_tr[k][kk] = 0;
      sum_tr_close_tr[k][kk] = 0.0;
      nc_close_trloop[k][kk] = 0;
      sum_tr_close_trloop[k][kk] = 0.0;
      nc_far[k][kk] = 0;
      sum_tr_far[k][kk] = 0.0;
      nc_far_loop[k][kk] = 0;
      sum_tr_far_loop[k][kk] = 0.0;
      nc_far_tr[k][kk] = 0;
      sum_tr_far_tr[k][kk] = 0.0;
      nc_far_trloop[k][kk] = 0;
      sum_tr_far_trloop[k][kk] = 0.0;
    }
  }

  double tottr_close = 0.0;
  double tottr_far = 0.0;

  for (i=0;i<nres;i++){
    for (j=0;j<nres;j++){
      for (k=0;k<20;k++){
        if (strcmp(residue_type[i], reslist[k])==0){
          for (kk=0;kk<20;kk++){
            if (strcmp(residue_type[j], reslist[kk])==0){
              if (abs(i-j) < 5){
                nc_close[k][kk]++;
                sum_tr_close[k][kk] += tr2[i][j];
                tottr_close += tr2[i][j];
                if ((loc[i]+loc[j])==2){
                  nc_close_tr[k][kk]++;
                  sum_tr_close_tr[k][kk] += tr2[i][j];
                }
                else if ((loc[i]+loc[j])==1){
                  nc_close_trloop[k][kk]++;
                  sum_tr_close_trloop[k][kk] += tr2[i][j];
                }
                else if ((loc[i]+loc[j])==0){
                  nc_close_loop[k][kk]++;
                  sum_tr_close_loop[k][kk] += tr2[i][j];
                }
                else printf("Error %3d %3d!\n", loc[i], loc[j]);
              }
              if (abs(i-j) >= 5){
                nc_far[k][kk]++;
                sum_tr_far[k][kk] += tr2[i][j];
                tottr_far += tr2[i][j];
                if ((loc[i]+loc[j])==2){
                  nc_far_tr[k][kk]++;
                  sum_tr_far_tr[k][kk] += tr2[i][j];
                }
                else if ((loc[i]+loc[j])==1){
                  nc_far_trloop[k][kk]++;
```

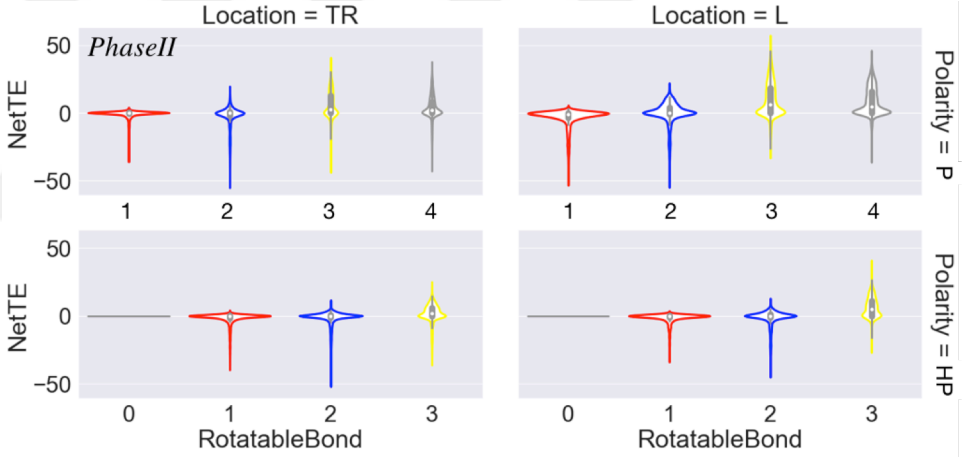
```
                    sum_tr_far_trloop[k][kk] += tr2[i][j];
                }
                else if ((loc[i]+loc[j])==0){
                    nc_far_loop[k][kk]++;
                    sum_tr_far_loop[k][kk] += tr2[i][j];
                }
                else printf("Error %3d %3d!\n", loc[i], loc[j]);
            }
        }
    }
  }
  }

  total = 0;
  for (k=0;k<20;k++){
    for (kk=0;kk<20;kk++){
      total += nc_close[k][kk] + nc_far[k][kk];
      printf("%3d %3d %s %s %7.2lf %3d %7.2lf %7.2lf %7.2lf %3d %7.2lf
%7.2lf\n", k, kk, reslist[k], reslist[kk], sum_tr_close[k][kk], nc_close[k][kk],
sum_tr_close[k][kk]/nc_close[k][kk], sum_tr_close[k][kk]/tottr_close*100.0,
sum_tr_far[k][kk], nc_far[k][kk], sum_tr_far[k][kk]/nc_far[k][kk],
sum_tr_far[k][kk]/tottr_far*100.0);
    }
  }
  printf("tot %3d\n", total);
  printf("total tr for close %7.2lf\n", tottr_close);
  printf("total tr for far %7.2lf\n", tottr_far);

  // combining i - j nd j - i pairs
  for (k=0;k<20;k++){
    sumr_close[k] = 0.0;
    sumr_far[k] = 0.0;
    sumnc_close[k] = 0;
    sumnc_far[k] = 0;

    sumr_close_tr[k] = 0.0;
    sumr_far_tr[k] = 0.0;
    sumnc_close_tr[k] = 0;
    sumnc_far_tr[k] = 0;
    sumr_close_trloop[k] = 0.0;
    sumr_far_trloop[k] = 0.0;
    sumnc_close_trloop[k] = 0;
    sumnc_far_trloop[k] = 0;
    sumr_close_loop[k] = 0.0;
    sumr_far_loop[k] = 0.0;
    sumnc_close_loop[k] = 0;
    sumnc_far_loop[k] = 0;

    for (kk=0;kk<20;kk++){
      sum_tr_new_close[k][kk] = sum_tr_close[k][kk] + sum_tr_close[kk][k];
      sumnc_close[k] += nc_close[k][kk] + nc_close[kk][k];
      sumr_close[k] += sum_tr_new_close[k][kk];

      sum_tr_new_far[k][kk] = sum_tr_far[k][kk] + sum_tr_far[kk][k];
      sumnc_far[k] += nc_far[k][kk] + nc_far[kk][k];
      sumr_far[k] += sum_tr_new_far[k][kk];

    // mutual information between transmembrane regions
      sum_tr_new_close_tr[k][kk] = sum_tr_close_tr[k][kk] +
sum_tr_close_tr[kk][k];
      sumnc_close_tr[k] += nc_close_tr[k][kk] + nc_close_tr[kk][k];
      sumr_close_tr[k] += sum_tr_new_close_tr[k][kk];

      sum_tr_new_far_tr[k][kk] = sum_tr_far_tr[k][kk] + sum_tr_far_tr[kk][k];
      sumnc_far_tr[k] += nc_far_tr[k][kk] + nc_far_tr[kk][k];
      sumr_far_tr[k] += sum_tr_new_far_tr[k][kk];

    // mutual information between transmembrane and loop regions
      sum_tr_new_close_trloop[k][kk] = sum_tr_close_trloop[k][kk] +
sum_tr_close_trloop[kk][k];
      sumnc_close_trloop[k] += nc_close_trloop[k][kk] + nc_close_trloop[kk][k];
      sumr_close_trloop[k] += sum_tr_new_close_trloop[k][kk];

      sum_tr_new_far_trloop[k][kk] = sum_tr_far_trloop[k][kk] +
sum_tr_far_trloop[kk][k];
      sumnc_far_trloop[k] += nc_far_trloop[k][kk] + nc_far_trloop[kk][k];
      sumr_far_trloop[k] += sum_tr_new_far_trloop[k][kk];

    // mutual information between loop regions
      sum_tr_new_close_loop[k][kk] = sum_tr_close_loop[k][kk] +
sum_tr_close_loop[kk][k];
      sumnc_close_loop[k] += nc_close_loop[k][kk] + nc_close_loop[kk][k];
      sumr_close_loop[k] += sum_tr_new_close_loop[k][kk];

      sum_tr_new_far_loop[k][kk] = sum_tr_far_loop[k][kk] +
sum_tr_far_loop[kk][k];
      sumnc_far_loop[k] += nc_far_loop[k][kk] + nc_far_loop[kk][k];
      sumr_far_loop[k] += sum_tr_new_far_loop[k][kk];
    }
  }

  for (k=0;k<20;k++){
    for (kk=0;kk<20;kk++){
      fprintf(fp1, "%3d %3d %3d %3d %s %3d %s %3d %8.2lf %5d %8.2lf %8.2lf
%8.2lf %5d %8.2lf %8.2lf\n", k, kk, rotlist[k], rotlist[kk], reslist[k], reslist_HP[k],
reslist[kk], reslist_HP[kk], sum_tr_new_close[k][kk],
nc_close[k][kk]+nc_close[kk][k],
sum_tr_new_close[k][kk]/(nc_close[k][kk]+nc_close[kk][k]),
sum_tr_new_close[k][kk]/tottr_close*100.0, sum_tr_new_far[k][kk],
nc_far[k][kk]+nc_far[kk][k], sum_tr_new_far[k][kk]/(nc_far[k][kk]+nc_far[kk][k]),
sum_tr_new_far[k][kk]/tottr_far*100.0);

      // mutual information between far residues in different regions (tr-tr / tr-loop /
loop-loop) regions
      fprintf(fp2, "%3d %3d %3d %3d %s %3d %s %3d %8.2lf %5d %8.2lf %8.2lf
%8.2lf %5d %8.2lf %8.2lf %8.2lf %5d %8.2lf %8.2lf\n", k, kk, rotlist[k], rotlist[kk],
reslist[k], reslist_HP[k], reslist[kk], reslist_HP[kk], sum_tr_new_far_tr[k][kk],
nc_far_tr[k][kk]+nc_far_tr[kk][k],
sum_tr_new_far_tr[k][kk]/(nc_far_tr[k][kk]+nc_far_tr[kk][k]),
sum_tr_new_far_tr[k][kk]/tottr_far*100.0, sum_tr_new_far_trloop[k][kk],
nc_far_trloop[k][kk]+nc_far_trloop[kk][k],
sum_tr_new_far_trloop[k][kk]/(nc_far_trloop[k][kk]+nc_far_trloop[kk][k]),
sum_tr_new_far_trloop[k][kk]/tottr_far*100.0, sum_tr_new_far_loop[k][kk],
nc_far_loop[k][kk]+nc_far_loop[kk][k],
sum_tr_new_far_loop[k][kk]/(nc_far_loop[k][kk]+nc_far_loop[kk][k]),
sum_tr_new_far_loop[k][kk]/tottr_far*100.0);

      // mutual information between close residues in different regions (tr-tr / tr-loop /
loop-loop)
      fprintf(fp3, "%3d %3d %3d %3d %s %3d %s %3d %8.2lf %5d %8.2lf %8.2lf
%8.2lf %5d %8.2lf %8.2lf %8.2lf %5d %8.2lf %8.2lf\n", k, kk, rotlist[k], rotlist[kk],
reslist[k], reslist_HP[k], reslist[kk], reslist_HP[kk], sum_tr_new_close_tr[k][kk],
nc_close_tr[k][kk]+nc_close_tr[kk][k],
sum_tr_new_close_tr[k][kk]/(nc_close_tr[k][kk]+nc_close_tr[kk][k]),
sum_tr_new_close_tr[k][kk]/tottr_close*100.0, sum_tr_new_close_trloop[k][kk],
nc_close_trloop[k][kk]+nc_close_trloop[kk][k],
sum_tr_new_close_trloop[k][kk]/(nc_close_trloop[k][kk]+nc_close_trloop[kk][k]),
sum_tr_new_close_trloop[k][kk]/tottr_close*100.0, sum_tr_new_close_loop[k][kk],
nc_close_loop[k][kk]+nc_close_loop[kk][k],
sum_tr_new_close_loop[k][kk]/(nc_close_loop[k][kk]+nc_close_loop[kk][k]),
sum_tr_new_close_loop[k][kk]/tottr_close*100.0);
    }
  }

  for (k=0;k<20;k++){
    fprintf(fp4, "%3d %3d %s %3d %8.2lf %5d %8.2lf %8.2lf %8.2lf %5d %8.2lf
%8.2lf\n", k, rotlist[k], reslist[k], reslist_HP[k], sumr_close[k], sumnc_close[k],
sumr_close[k]/sumnc_close[k], sumr_close[k]/tottr_close*100.0, sumr_far[k],
sumnc_far[k], sumr_far[k]/sumnc_far[k], sumr_far[k]/tottr_far*100.0);

    // residue based mutual information for far residues in different regions (tr-tr / tr-
loop / loop-loop) regions
    fprintf(fp5, "%3d %3d %s %3d %8.2lf %5d %8.2lf %8.2lf %8.2lf %5d %8.2lf
%8.2lf %8.2lf %5d %8.2lf %8.2lf\n", k, rotlist[k], reslist[k], reslist_HP[k],
sumr_far_tr[k], sumnc_far_tr[k], sumr_far_tr[k]/sumnc_far_tr[k],
sumr_far_tr[k]/tottr_far*100.0, sumr_far_trloop[k], sumnc_far_trloop[k],
sumr_far_trloop[k]/sumnc_far_trloop[k], sumr_far_trloop[k]/tottr_far*100.0,
sumr_far_loop[k], sumnc_far_loop[k], sumr_far_loop[k]/sumnc_far_loop[k],
sumr_far_loop[k]/tottr_far*100.0);

    // residue based mutual information for close residues in different regions (tr-tr / tr-
loop / loop-loop) regions
    fprintf(fp6, "%3d %3d %s %3d %8.2lf %5d %8.2lf %8.2lf %8.2lf %5d %8.2lf
%8.2lf %8.2lf %5d %8.2lf %8.2lf\n", k, rotlist[k], reslist[k], reslist_HP[k],
sumr_close_tr[k], sumnc_close_tr[k], sumr_close_tr[k]/sumnc_close_tr[k],
sumr_close_tr[k]/tottr_close*100.0, sumr_close_trloop[k], sumnc_close_trloop[k],
sumr_close_trloop[k]/sumnc_close_trloop[k],
sumr_close_trloop[k]/tottr_close*100.0, sumr_close_loop[k], sumnc_close_loop[k],
sumr_close_loop[k]/sumnc_close_loop[k], sumr_close_loop[k]/tottr_close*100.0);
  }


  // percentage of tr values in defined ranges
  lev = lev1 = lev2 = lev3 = lev4 = lev5 = lev6 = 0;
  for (i=0;i<nres;i++){
    for (j=i;j<nres;j++){
      lev++;
      if ((tr2[i][j] >= 0.0) && (tr2[i][j] <= 0.1)) lev1++;
      else if ((tr2[i][j] > 0.1) && (tr2[i][j] <= 0.2)) lev2++;
        else if ((tr2[i][j] > 0.2) && (tr2[i][j] <= 0.5)) lev3++;
          else if ((tr2[i][j] > 0.5) && (tr2[i][j] <= 1.0)) lev4++;
            else if ((tr2[i][j] > 1.0) && (tr2[i][j] <= 2.0)) lev5++;
              else if (tr2[i][j] > 2.0) lev6++;
                else printf("Error! %3d %3d %Lf\n", i, j, tr2[i][j]);

    }
```

```
    }
    printf("%lf\n", lev);
    printf("%lf %lf %lf %lf %lf %lf\n", lev1/lev*100, lev2/lev*100, lev3/lev*100,
lev4/lev*100, lev5/lev*100, lev6/lev*100);
    printf("%lf\n",
lev1/lev*100+lev2/lev*100+lev3/lev*100+lev4/lev*100+lev5/lev*100+lev6/lev*100
);

    // Mean tr per residue
    for (i=0;i<nres;i++){
        sum[i] = 0.0;
        for (j=0;j<nres;j++){
            if (i!=j) sum[i] += (tr2[i][j] - tr2[j][i]);
        }
        avg[i] = sum[i]/308.0;
    }
    // standard deviation
    for (i=0;i<nres;i++){
        std[i] = 0.0;
        for (j=0;j<nres;j++){
            if (i!=j) std[i] += ((tr2[i][j] - tr2[j][i]) - avg[i])*((tr2[i][j] - tr2[j][i]) - avg[i]);
        }
        std[i] = sqrt(std[i]/308);
    }
    for (i=0;i<nres;i++) fprintf(fp7,"%3d %7.2lf %7.2lf\n", i, avg[i], std[i]);

  return 0;
}
```

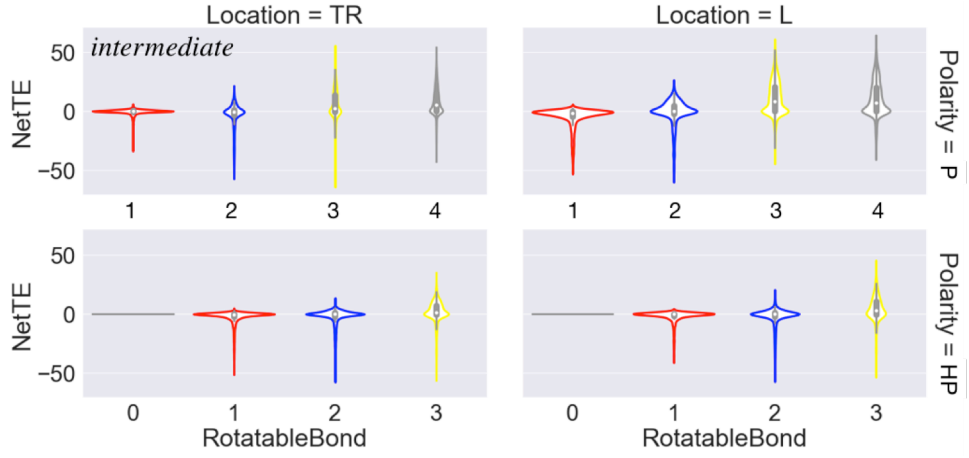
APPENDIX K



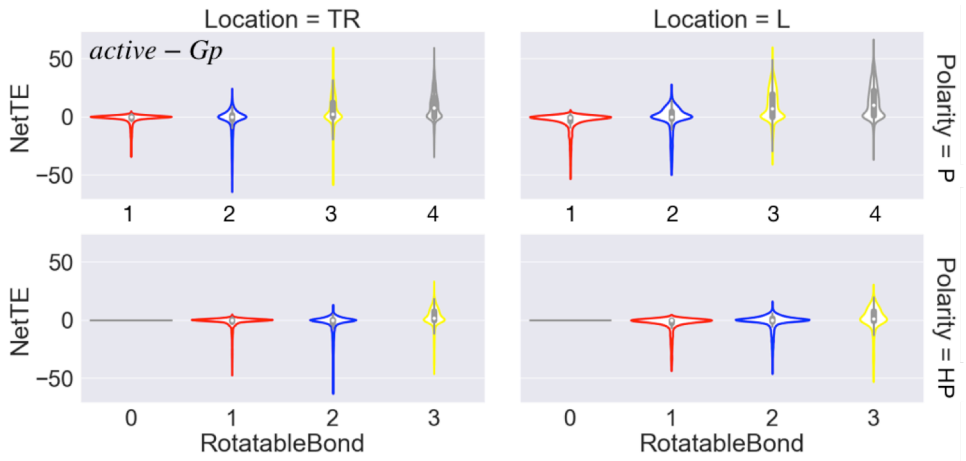
(a)



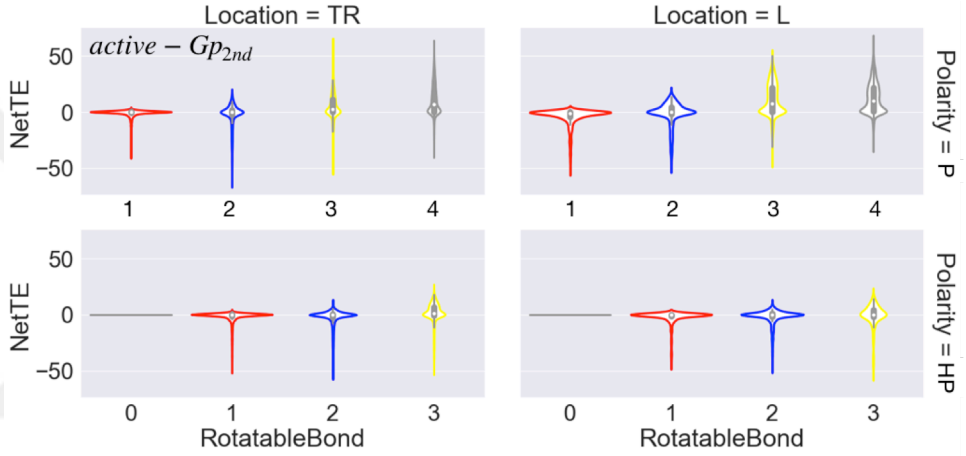
(b)



(c)



(d)



(e)

Figure K. 1 The distribution of net TE with respect to number of rotatable bonds in Phase I (a), Phase II (b), *intermediate* state (c), *active-Gp* (d), and *active-Gp*$_{2nd}$ (e). Polar residues were denoted with P and hydrophobic residues were denoted with HP. Residues with one rotatable bond were colored with red, two were with blue, three were with yellow, and four were with grey.

APPENDIX L

Table L. 1 Protein regions mutual information intensities based on backbone dihedrals

Backbone			
Mutual Information			
	(0<MI<-0.01)	**(0.01≤MI<-0.03)**	**(0.03≤MI)**
Phase I	$H1_{Ext}$, ECL3	ECL1, ICL2, ECL2, ICL3	ECL2, ICL3, H8
Phase II	ICL1, ECL2	ECL1, ICL2, ECL2	
intermediate	$H1_{Ext}$, ICL2, ECL2, ICL3, ECL3, $H7_{Int}$	ICL1, ICL3	
active-Gp	ICL1, ECL2	$H3_{Mid}$, ECL2	ICL2, ICL3, ECL3
active-Gp$_{2nd}$	ICL1, ECL1, $H3_{Int}$, $H4_{Int}$, $H6_{Int}$, H8	ICL1, ECL2, $H7_{Int}$	ICL3, ECL3

Table L. 2 Mutual information intensities in the protein regions in bacibone + side-chain dihedrals

Backbone + Side-chain		
Mutual Information		
	(0.1≤MI<-0.3)	**(0.3≤MI)**
Phase I	ICL1, ECL1, $H3_{Ext}$, ICL2, $H4_{Ext}$, $H4_{Int}$, ECL2, $H5_{Int}$, ECL3, H8	ICL2, ECL2, ICL3, $H7_{Int}$
Phase II	ICL1, ECL1, ICL2, ECL2, $H5_{Int}$, ICL3, ECL3, H8	
intermediate	$H2_{Ext}$, ECL1, $H4_{Int}$, ECL2, $H5_{Int}$, ICL3, $H6_{Int}$, ECL3, $H7_{Int}$	
active-Gp	ICL1, $H2_{Int}$, $H3_{Mid}$, ICL2, $H4_{Ext}$, $H4_{Int}$, ECL2, $H5_{Ext}$, $H5_{Int}$, ICL2, ECL3, $H7_{Int}$, H8	ICL3
active-Gp$_{2nd}$	ICL2, $H2_{Int}$, ECL1, $H3_{Mid}$, ICL2, ECL2, $H5_{Int}$, ICL3, $H6_{Int}$, $H6_{Ext}$, ECL3, $H7_{Ext}$, H8	ICL3, $H7_{Int}$, H8

Table L. 3 Donor and acceptor regions based on net transfer entropy of backbone dihedrals

Backbone		
Net Transfer Entropy		
	Donor	**Acceptor**
Phase I	ECL2 adjacent to H4, ICL3 adjacent to H6, H8	ICL1, ECL1, ICL2, ECL2 adjacent to H5, ICL3 adjacent to H5
Phase II	ICL2 adjacent to H3, ECL2, $ICL3_{Mid}$	ICL1, ECL1, ECL2, ECL3
intermediate	$H1_{Ext}$, ICL1, $H2_{Ext}$, $H4_{Int}$, ECL2, $H7_{Int}$	ICL1, ECL1, ICL2, ECL2 adjacent to H4, ECL2 adjacent to H5, ICL3, ECL3
active-Gp	$H3_{Mid}$, ICL2, ICL3,	ICL1, ICL2, ECL2, $H6_{Ext}$, ECL3, $H7_{Int}$
active-Gp$_{2nd}$	ICL2, ICL3, ECL3	ICL1, H2Ext, ECL1, ECL2, $H5_{Ext}$, $H7_{Int}$

Table L. 4 Donor and acceptor regions based on net TE of backbone + χ_1-chain dihedrals.

Backbone + χ_1		
Net Transfer Entropy		
	Donor	**Acceptor**
Phase I	ECL1, ICL2, ECL2, ICL3	$H1_{Ext}$, ICL2, $H5_{Int}$
Phase II	ICL2, ECL2, $ICL3_{Mid}$	$H1_{Int}$, ICL3
intermediate	$H1_{Ext}$, ECL1, ICL3	ICL2, ECL2, H8
active-Gp	ECL1, ICL2, ICL3	ICL1, $H2_{Mid}$, ECL2
active-Gp$_{2nd}$	ECL1, ECL2, ICL3, $H7_{Int}$	ICL2, ECL2

Table L. 5 The differences of donor and acceptor characteristics of residues between states based on side-chain net TE values

Side-Chain				
Net Transfer Entropy				
	Phase I - Phase II	**Phase I - *intermediate***	**Phase I - *active-Gp***	**Phase I - *active-Gp$_{2nd}$***
Donor → Acceptor (Donor ≥2 & Acceptor ≤-1)	*Arg*221 *Glu*249 *Asp*251 *Glu*306 *Arg*328	*Asp*300 *Glu*306	*His*172 *Gln*229	*His*172 *Gln*229
Acceptor → Donor (Donor ≥2 & Acceptor ≤-1)	*Met*96	*Phe*193 *His*296 *Tyr*326	*Met*82 *Met*96 *Asn*103 *Trp*173 *Phe*240	*Met*82 *Ile*94 *His*296
		Phase II - *intermediate*	**Phase II - *active-Gp***	**Phase II - *active-Gp$_{2nd}$***
Donor → Acceptor (Donor ≥2 & Acceptor ≤-1)		*Arg*175	*His*172 *Gln*229	*Met*96 *His*172 *Gln*229
Acceptor → Donor (Donor ≥2 & Acceptor ≤-1)		*Phe*193 *Arg*221 *Glu*249 Asp251 *His*269 *His*296 *Tyr*326 *Arg*328	*Tyr*70 *Asn*103 *Trp*173 *Arg*221 *Glu*237 *Phe*240 *His*241 *Glu*249 *Glu*306 *Arg*328	*Tyr*70 *Arg*221 *Glu*237 *Glu*249 *His*296 *Glu*306 *Arg*328
			Intermediate - active-Gp	***Intermediate - active-Gp$_{2nd}$***
Donor → Acceptor (Donor ≥2 & Acceptor ≤-1)			*His*172 *Phe*193 *Gln*229 *His*296	*Met*171 *His*172 *Phe*193 *Gln*229 *Tyr*326
Acceptor → Donor (Donor ≥2 & Acceptor ≤-1)			*Tyr*70 *Met*82 *Asn*103 *Arg*175 *Phe*240 *Glu*306	*Tyr*70 *Met*82 *Ile*94 *Arg*175 *Glu*306
				active-Gp - active-Gp$_{2nd}$
Donor → Acceptor (Donor ≥2 & Acceptor ≤-1)				*Met*96
Acceptor → Donor (Donor ≥2 & Acceptor ≤-1)				*Phe*290 *His*296

Table L. 6 Residue type comparisons based on backbone + side-chain dihedral angles

	Backbone + Side-chain Proximal MI (Residue Type with highest MI value)	**Backbone + Side-chain Distal MI (Residue Type with highest MI value)**
Phase I	Gln	Arg
Phase II	Gln	Gln
intermediate	Gln	Arg
active-Gp	Arg	Arg
active-Gp$_{2nd}$	Arg	Arg

Table L. 7 Donor and acceptor residue types on net transfer entropy data

	Backbone Net TE Donor	**Backbone Net TE Acceptor**	**Backbone + χ_1 Net TE MI Donor**	**Backbone + χ_1 Net TE MI Acceptor**	**Side-chain Net TE MI Donor**	**Side-chain Net TE MI Acceptor**
Phase I	Gln	Thr	His	Gln	Gln, Lys	Pro
Phase II	Gly	Asp, Pro	His	Gln	Gln, Lys	Pro
intermediate	Gly, Arg	Pro	His, Arg	Gln, Asn	Gln, Lys	Pro
active-Gp	His, Gly	Trp	Thr	Gln, Glu, Lys	Arg, Lys	Pro
active-Gp$_{2nd}$	Gly, His	Thr, Trp	Thr	Glu, Gln, Lys	Arg, Lys	Pro