5-2016

Understanding Allosteric Modulation of G-Protein Coupled Receptors

Vibhor Agrawal
Clemson University, vibhora@clemson.edu

Understanding Allosteric Modulation of G-Protein Coupled Receptors

---

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Chemistry

---

by
Vibhor Agrawal
May 2016

Accepted by:
Dr. Brian N. Dominy, Committee Chair
Dr. Steve J. Stuart
Dr. Emil Alexov
Dr. Jason McNeill

Abstract

G protein-coupled receptors (GPCRs) which are seven-transmembrane allosteric machine constitutes largest and diverse family of membrane proteins. GPCR participate in activating a diverse range of signaling pathways, in response to ligand perturbation which ranges from neurotransmitters, hormones to photons. The role of GPCRs in a wide range of key physiological processes and their ubiquity in mammalian genome makes them attractive pharmaceutical targets. Signal transduction in GPCR occur mainly, via G proteins and leads to a cascade of signaling. In addition to the orthosteric site, GPCRs also possesses a topographically distinct allosteric site which contributes to allosteric modulation, i.e long distant ligand binding for activating G proteins and trigger GDP release. The mechanism that governs allosteric activation triggering GDP release is yet uncertain. Differential ligands bind to GPCR's orthosteric sites and can modulate allosteric signaling. Ligands that increase or decrease the GPCR signaling are classified as agonists and antagonists respectively. Compared to orthosteric ligand allosteric modulator through electrostatic repulsion, steric hindrance or conformational stability can select subsets of signaling responses.

We in this study are trying to understand the basis of ligand-biased signaling or functional selectivity that leads to long-distance signaling in a receptor. Using the information from crystal structures of the receptor, combined with molecular dynamics simulations, we performed a systematic analysis to identify the basis of

conformational selectivity for allosteric bias in GPCRs. Our study explores the conformational landscape of GPCRs as a function of the activity of the receptor. Normal modes analysis (NMA) was used to identify low-frequency modes that describe conformational changes due to large-scale domain motions in the receptor. NMA characterized changes in correlated motions of residues in the first six global modes and revealed conformation shift starting from the inactive structure.

We used MD simulations coupled with network analysis to reveal correlated motion between G-protein Coupling site and ligand binding site. Changes in dynamic correlated residue motion in allosteric networks reveals the characteristic feature of receptor activity in GPCRs. Single point mutations studies were aimed to analyze the changes in the structural scaffold of GPCRs as a result of mutations. Mutational studies facilitated in determining the basis of functional selectivity and changes in the allosteric communication as a result of allosteric binding to the receptor. Single point mutations also revealed residues critical for functional activity of GPCRs. Inter-residue contact network responsible for biased signaling using microsecond atomic level simulations reveals differential allosteric modulation. Finally, comparative analysis using mutual information in the internal coordinates of mutants and wild types helped to quantify the allosteric modulation and long-range cooperativity between binding sites in GPCRs.

Dedication

I would like to dedicate this dissertation to my loving parents and brother, without whose support none of the work would have been possible.

Acknowledgments

I am deeply grateful to my advisors Professor Brian Dominy and Professor Steve Stuart for their infectious scientific and intellectual curiosity. They have been a constant source of inspiration throughout my graduate career. I would like to thank all of the members of the Dominy-Stuart group and other scientists at Clemson University for creating a scientifically exciting environment for graduate work in computational biophysics and bioinformatics. A few people who worked with me on various projects in no particular order are Zhe Jia, Tingting Han, Linna Liu. I wish to thank my dissertation committee members Professor Emil Alexov, Professor Jason McNeil for many helpful discussions related to the various research projects presented here. Many thanks to Professor Brian Kobilka and Professor Robert Lefkowitz for being an inspiration and reason to pursue graduate research in GPCRs. I would thank my undergraduate thesis advisors and Dr. Aridaman Pandit, for introducing me to the exciting world of computational biology and encouraging me to jump straight into research even though I had limited biological training at the time. I would like to thank my friends and family who were constant support during my journey as a graduate student.

The NSF and NIH were my sources of support at various times. Specific grants as well as the articles that provided much of the source material are acknowledged in the appropriate chapters.

Table of Contents

List of Figures

Chapter 1

Introduction

G-protein coupled receptor (GPCR) is one of largest and most diverse member of membrane protein family. They are and are encoded by more than 800 genes in the human genome [24]. GPCRs in response to signaling molecule like hormones and neurotransmitters host diversity of cellular response. They are responsible for signal transduction and play important role in the carrying various physiological activities ranging from sensory perception, chemo taxis, to neurological responses. Their involvements in this diverse and important set of responses also make them largest class of drug targets [24]

1.1 GPCR classification system

Guidelines of International Union of Basic and Clinical Pharmacology classify GPCRs broadly into four families: class A rhodopsin-like, class B or secretin-like, class C or metabotropic glutamate like, and frizzled receptors [24]. Rhodopsin-like class A forms the largest super family and is one the most studied class. Common Class A GPCR includes rhodopsins, adrenergic receptor and cannabinoid receptors and

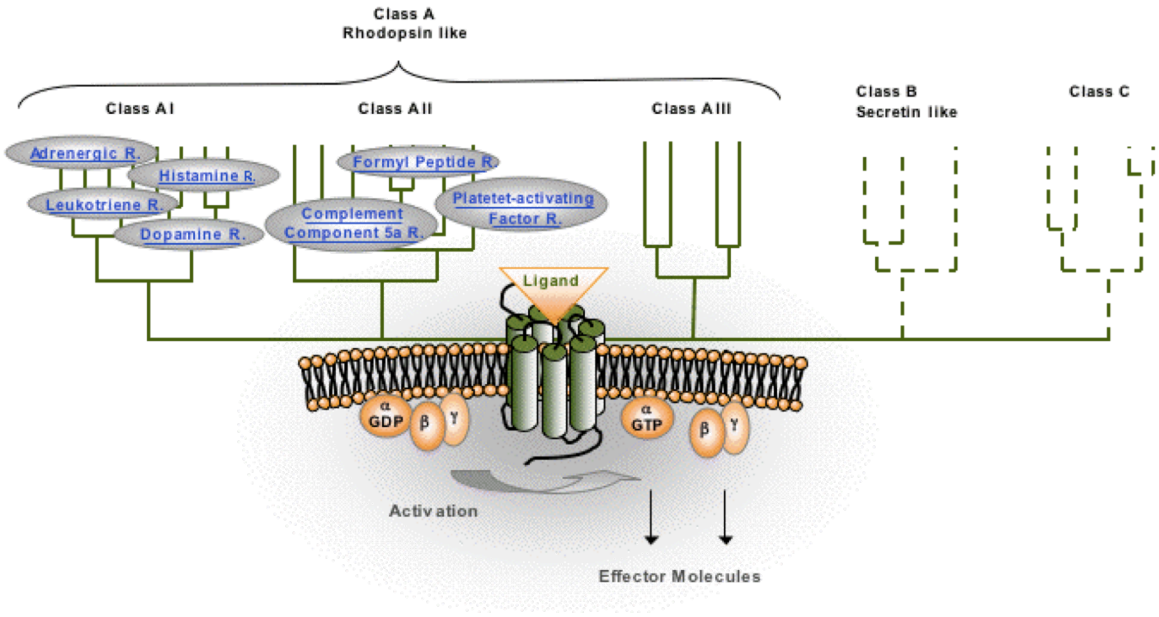


Figure 1.1: Classification of GPCR

Extremely diverse nature of the superfamily of G-protein-coupled receptors (GPCRs) have common structural features which include seven helical transmembrane domains and broadly classified into four main families based on pharmacological properties: class A rhodopsin-like, class B secretin-like, class C metabotropic glutamate/pheromone, and frizzled receptors. (modified from jenabioscience.com)

.

share regions of residues conserved across the subfamily which include Glu/Asp-Arg-Tyr (E/DRY) motif on the intracellular side of TM3, along with Asn-Pro-X-X-Tyr (NPXXY) motif on the intracellular side of TM7 [65]. Within the family cysteine residue the carboxy terminus is often palmitoylated for attaching the receptor to membrane. Helix 8 (H8) is α-helix originating at start of the carboxyl terminus is common in the subfamily. The class B superfamily comprises of secretin receptor, calcitonin receptor, parathyroid receptor, and the glucagon receptor. Class B superfamily comprises mostly of peptide hormones and large extracellular loops and often with many well-characterized protein domains like EGF domains [50]. The amino terminus plays important role in modulating ligand binding and interacts with other proteins like chaperons, receptor activity modifying proteins (RAMPs) which act as accessory proteins. Class C receptor is characterized by long amino acid terminal and majorly constituted by calcium sensing receptor (CASR), eight metabotropic glutamate receptors (GRM), two aminobutyric acid (GABA) receptors. The long amino termini which can be 600 residue plays major role in ligand recognition. Frizzled class common include conserved cysteine rich domain mainly include receptors that participate in Wnt-signalling pathway [44]. When activated receptor in turn lead to activation of Dishevelled in the cytosol.

These isolation of crystal structures provides enormous amount insights towards structural and functional diversity of receptors . With the advancement in structural biology coupled with the power of computation to determine the structural details of receptor is helping to address fundamental question about the topology of the GPCR fold and its role in towards receptor function and what are the molecular changes take place in receptor biogenesis? With availability of structures of diverse GPCRs we are in unique position to understand and comprehensively explore the structure function relationship of receptor at molecular level.

1.2 Common structural features of GPCR

Generally, GPCRs consist of seven transmembrane α-helices (7TM) spanning across the membrane. The transmembrane region composed mainly of hydrophobic core of amino acids. These TMs are connected by three extracellular loops (EC1-EC3) and three intracellular loops (IC1-IC3). The extracellular region (EC) plays crucial role in ligand binding. The intracellular (IC) region is crucial in interaction with G proteins, β-arrestins and other ligands. Generally receptor also includes cytoplasmic H8 and a carboxyl terminus. Carboxy terminus generally provides sites for palmitoylation. The receptor stabilization in the membrane is achieved through non-covalent interactions. Non covalent interaction define the GPCRs fold. The GPCRs fold contains a number of kinks, mostly induced by Pro residues. EC regions is involve in ligand binding and plays important role in receptor activation and undergo small conformational movement has highest sequence diversity IC region involve in binding of G-protein and β-arrestin is more conserved in GPCR family and undergo large conformational movement upon receptor activation.

1.3 Solved structures of GPCR and Reason for crystallography success

Solving and isolating crystal structure has been a challenge in past two decades. Since 2007, with advancement of crystallographic method and innovation in protein engineering method there is exponential increased in solved crystal structure [10]. Current state of art to isolate the GOCR structure include

- creating receptorT4 lysozyme complex [10, 61].

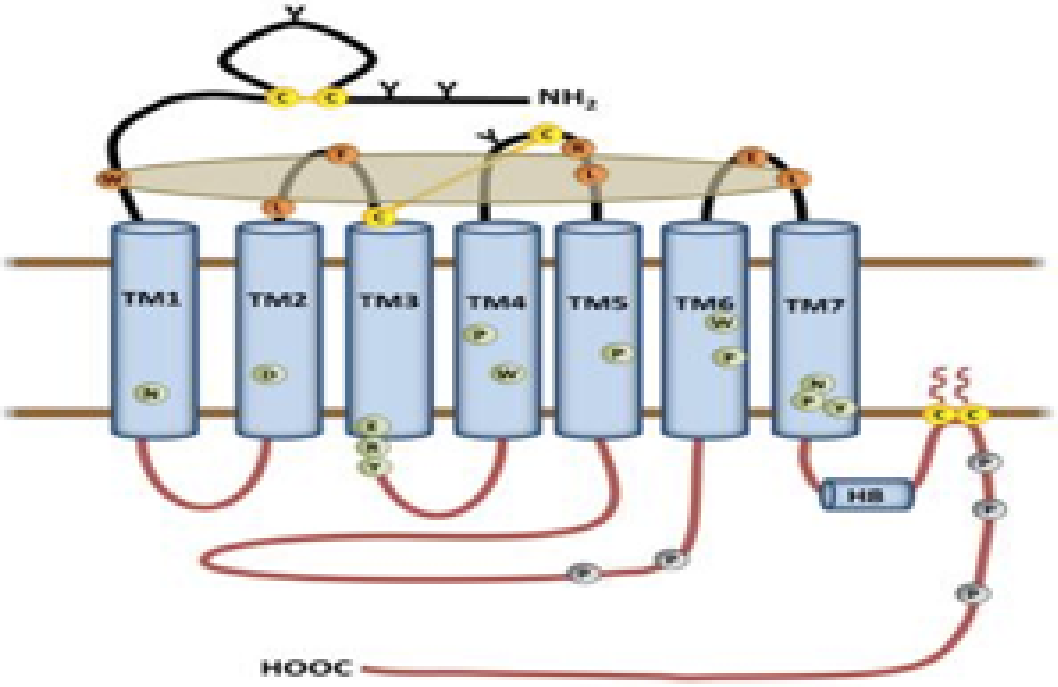


Figure 1.2: General outline of GPCR structure highlighting seven transmembrane helices with helix 8 and three extracellular loops and three intracellular loops

- receptorapocytochrome chimaeras [77]
- co-crystallization with monoclonal antibody fragments from mouse or cameloid [57, 58]
- thermo-stabilizing GPCRs by systematic scanning mutagenesis [84, 20] or using disulphide bridges [16].

The above method is usually achieved my truncating the flexing IC loop to add the high affinity ligand that enhance receptor stability. Other refinement approach uses the use of lipid cubic [8] phase and new detergents [9] has improved the isolation of crystal more likely. Current use of micro-crystallography [47] helps to obtain higher resolution diffraction from smaller crystals.

With above described advancements there has been upsurge of crystal structure since 2007. The recent crystal structure of class A GPCRs solved are mostly inactive conformation bound to either antagonist that leads to decrease in basal activity or receptor bound to neutral inverse agonist [12]. NTSR-1 was isolated with agonist with increase of basal activity [65]. Few of the class A GPCRs structures solved in recent times are

1. rhodopsin (bovine rhodopsin [52] and squid rhodopsin [49]
2. β-adrenoceptors and human β-adrenoceptors [10]
3. muscarinic acetylcholine receptors (human and rat muscarinic acetylcholine receptors [29, 38]
4. Histamine receptor [70] and dopamine receptor [11]
5. a nucleoside-binding GPCR: rat neurotensin receptor (NTSR1), opioid receptors, human adenosine A2A, human CXCR4 chemokine receptor and human

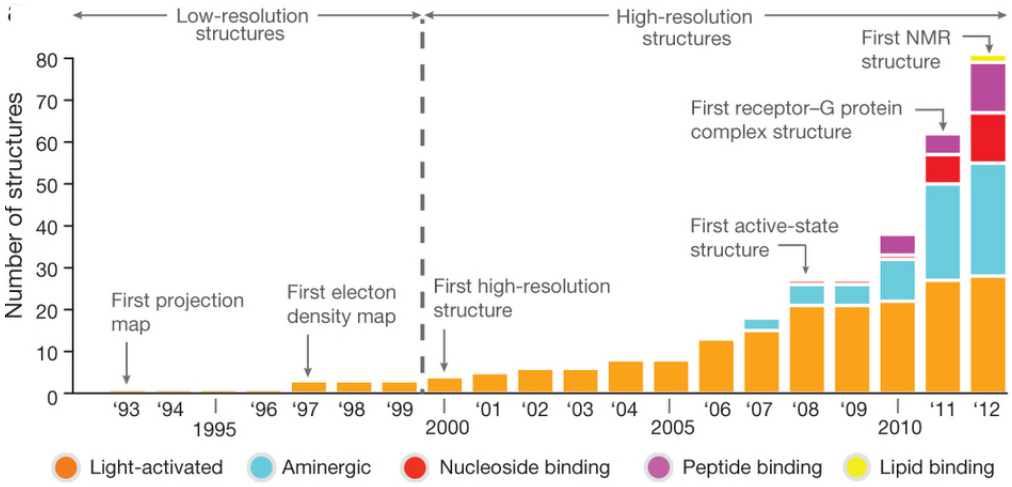


Figure 1.3: following chart indicates increasing number of crystal structure in recent time. (modified from AJ Venkatakrishnan et al. Nature 494, 185-194 (2013)

protease activated receptor

6. lipid-binding GPCR: human sphingosine-1 phosphate(S1P1) receptor

Although With these successful state of art method in isolating crystal structure of GPCR there are certain limitation to above recombinant methods.These limitations include removal of post translation modification during isolation, loop truncations limits understanding of structure-functional mapping of these regions, insertion of T4-lyzozyme and adding the themostablizing mutant might bias the conformation ensemble during isolation.

1.4 Ballesteros-Weinstein Numbering

The Ballesteros-Weinstein [3] number is a general numbering system that facilitate consistent residue numbering across GPCRs independent of individual sequential numbers. Ballesteros-Weinstein number scheme works on principle that on multiple

sequence alignment residues with the same general residue number have equivalent locations. In our present work we have used Ballesteros-Weinstein scheme where residue numbers represented in format 'TM.n', where the 'TM' denotes the transmembrane helix number, and the number after the decimal point indicates residue position with respect to the most conserved residue position in transmembrane helix.

1.5 Molecular Changes during Receptor biogenesis

Ligands are the molecules that when bind to receptor regulate the receptor activity. Ligands bind to the orthosteric site within transmembrane domain which induces structural change that correspond to functional activity. The diverse ligand include neurotransmitters, light-sensitive compounds, odors, pheromones, hormones. Size of ligands range from small molecule to large proteins. Ligand binding mainly to the extracellular region of the TM region results in small conformational change. These conformational change further lead to the larger rearrangements in the cytoplasmic side of the receptor that further facilitate the binding of G-protein or β-arrestins. Active state of GPCR is when the receptor couples and stabilizes with the effector molecule like G-protein. Several studies have been successful in isolating crystal structure in active and intermediate state. These crystal structures details to a certain extent provide insight to activation of receptor. Combined with current biophysical and computational studies details

Few GPCRs have constitutive activity, i.e. activity in the absence of agonist. The constitutive activity can be equated as an equilibrium of a GPCR population between receptors in the ground state with no activity (R) and receptors with fully

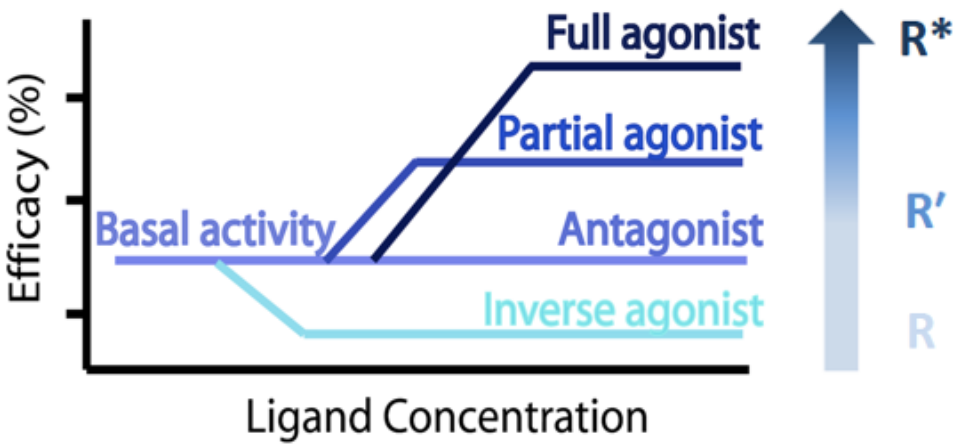


Figure 1.4: Activation State in GPCR.
Agonist binding to receptor lead to the active state $R^{\star}$ that enhance the efficacy and promote signal transduction. Upon binding with inverse agonist there is reduce in basal activity R'. Inverse agonist though bind to orthosteric site they neither alter $R \leftrightarrow R^{\star}$ equilibrium nor directly modulate signaling.

activated state ($R\star$). Experiment on β-adrenergic has shown that specific sub states that are sparsely populated in the ligand free receptor subspace are stabilized by agonist. The agonist or inverse agonist ligands have ability to shift conformation equilibrium from one active state to other extent of activity. The ligand upon binding to receptor not only stabilizes the conformation they also induce larger conformational rearrangement in the intracellular side of transmembrane.

Investigation of conformational changes is often quantified by quantitative mass spectrometry. It quantifies the chemical response of side chains of individual amino acid. Multiple ligand specific conformations have been identified in β adrenergic receptor by identifying distinct pattern of reactivity towards probe with different

amino acid [33] which further provides supports that ligand binding induce structure variability and stabilizes a specific conformation from conformation ensemble. Hydrogen-deuterium exchange mass spectrometry studies with β_2 adrenergic receptor not only confirms the existence of multiple conformation of receptor but also went ahead to explain the agonist induces more conformational mobility while inverse agonist are more stabilizing than agonist.

Above finding were critical in establishing the multiple conformational states in GPCR though further studies with crystal structure of agonist bound β-adrenergic receptor [62, 83, 82] were found to be inactive which suggest stabilizing the fully active state may not be sufficient by agonist but also depend and the conformation landscape for each specific GPCR [81]. GPCR binding with G-protein ,or β-arrestin is in active state. Computational simulation of receptor ligand complex elucidates, ligands with differential efficacies modulate free energy landscape differentially. The free energy landscape of receptor usually shifts the conformation equilibrium depending on ligands response.

To study the structural and molecular changes in the receptor study can be further divided into three parts.

- **Extracellular Region of the Receptor**
- **Transmembrane Region of the Receptor**
- **TM-intracellular Region of the Receptor**

1.5.1 Extracellular Region of the Receptor

Extracellular(EC) region is one of active segments during receptor biogenesis[85]. In rhodopsin where activation is induced to 11-cis-retinal to all-trans-retinal by light-induced isomerization, ECL-2 under go conformational changes [57]. In β_2-adrenergic

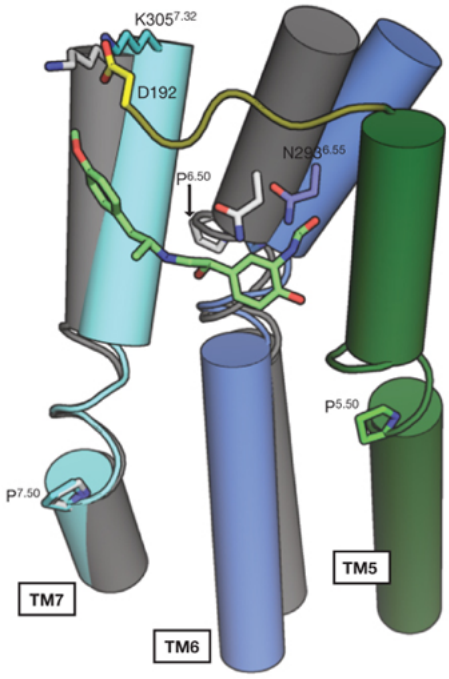


Figure 1.5: Changes in the EC region during receptor activation.

Activation model of β_2-adrenergic by agonist(formoterol). Grey helices and white side chains represent active-state model. Green sticks indicate agonist ligand and yellow indicates ECL2. modified from *Nature463, 108 − 112(7January2010)*

.

with agonist, antagonist and neutral ligands ECL-2 and ECl-3 has different conformation of salt bridge lining the ligand-binding pocket in different conformation[6]. Extracellular region act as drug targeting zone by multiple drugs to select specific substate from conformation ensemble. Its plausibe that drug stabilize distinct conformational states of EC hence modulating receptor activity which further lead to allosteric signaling [79].

1.5.2 TM Region of the Receptor

Previous studies on rhodopsin, β_2-adrenergic[73, 16] and α_{2A}-adrenergic highlights common structural rearrangement near the agonist binding site (towards the extracellular part of TM).These common changes include

1. Structural changes in $Pro^{5.50}$ induces distortion in TM5.The changes at TM5 are then transmitted to receptor cytoplasmic side

2. TM3 and TM7 are relocated during the receptor biogenesis

3. Binding of agonist lead to translation or rotation of TM5 and TM6. These rearrangement of TM-TM6 is mainly associated with rotation of near $Phe^{6.44}$. This change in translated to relocation of cytoplasmic end of TM6 and this further assist the binding of G-protein by opening the cleft. [38]

Water molecules also help in facilitating conformation specific contact. Water mediated hydrogen-bonding networks involving transmission switch plays crucial role in the receptor activation of class A receptors. Although these are common structural changes, agonist interact differently with receptor to induce these common structural changes. The extracellular region of TM3,TM5 and TM7 are "pulled" together when agonist interact with β_2-adrenergic and α_{2A}-adrenergic, in rhodopsin light induced retinal isomerization increases the volume of binding pocket [16] which helps in rearrangement of hydrophobic and aromatic residue in the receptor core these lead to further changes to the TM3-TM5 and rearrangement of TM5-TM6. The extent of TM6 movement after structural rearrangement can be upto $14\mathring{A}$.

1.5.3 TM-intracellular Region of the Receptor

ICL2 and TM3 plays critical role in receptor activation. ICL2 Residues and cytoplasmic end residues of TM3 interact with downstream signaling effectors like G-protein after receptor activation [31]. Residues of ICL2 in β_2-adrenergic interact with the N terminus of the Gα-subunit highlights role of ICL during receptor activation [59]. DRY motif $Asp^{3.50}$ of TM3 forms salt bridge with conserved Arg in ICL2. Mutation of analogous residue $Tyr149^{ICL-2}$ in avian b1-AR decreases receptor stability highlights the role of ICL2 in receptor activation'[84]. Furthermore phosphorylation of $Tyr141^{ICL-2}$ in human b2-AR activates the receptor complex by promoting binding of G-protein. Studies comparing the α_{2A}-adrenergic-T4lysozyme chimaera complex with α_{2A} bound to antibody shows distortion of cytoplasmic end of TM5 and TM6 is primarily achieved in presence of ICL3 and absence of ICL3 fails to form salt bridge with TM5 and TM6' [48].

Computational studies comparing active-state structures of the G-protein bound to β_2-adrenergic and of rhodopsin (metarhodopsin II) bound to a peptide identified TM consensus formed by movement of TM5 and TM6 residues that interact with with G-protein [81]. Along with these consensus residues during receptor activation, N terminal end of the G protein interacts with intracellular loop highlighting importance of ICL2 crucial for receptor biogenesis [48].

1.6 Allosteric Modulation in GPCRS

Conventionally signaling event in GPCRs is when endogenous ligand like G-protein, β-arrestins binds to orthosteric binding site. Certain ligands instead of binding to orthosteric binding site binds at distant site and still successfully initiate sig-

naling pathway, phenomenon termed as allostery and ligand associated with it are termed as allosteric modulators. Allosteric ligands that can selectively trigger different signaling pathways are often can be drug candidate of high potential. There are three types of allosteric modulators. (a) Positive allosteric Modulators that binds to allosteric distinct site and enhance affinity and efficacy of orthosteric agonist. (B) Negative allosteric Modulators that binds to allosteric distinct site and decrease affinity and efficacy of orthosteric ligands. (C) Silent allosteric Modulators that binds to allosteric distinct site and have no affinity and efficacy effect on binding of orthosteric ligands.

Allosteric site are better accessible and can be targeted by drugs better than orthosteric site. Allosteric ligand with limited cooperativity allows high degree of titrabilty and modulator can be administered in large doses with lower magnitude of target toxicity than orthosteric ligands.

Allosteric modulators mediate receptor biogenesis and modulate activity of allosteric ligands [36, 88]. The phenomenon is known as allosetric agonism. Studies shows ligand McN-A-343 and AC-42 inhibits binding of N-methylscopolamine to the M2 and human M1 mAChRs and act as partial agonist in absence of orthosteric ligand. Allosteric modulation and role of allosteric ligand in modulating activity of orthosteric ligand has been shown in past,complete mechanistic understanding of biased signaling is yet to be deciphered completely. In first structures of β_2-adrenergic complexed with arrestin-biased ligands, allosteric ligand have weaker interactions with TM5 compared with full agonists. Allosteric ligands interact with minor binding pocket near TM7and involving ECL2 and have role in arrestin-biased signaling [63]. Additionally allosteric ligandreceptor interact with TM7 and activation pathway is triggered through TM2 or TM7 conserved pathway [17]. Recent experiments with 19F-NMR spectroscopy show that biased ligands when bind to receptor alters the

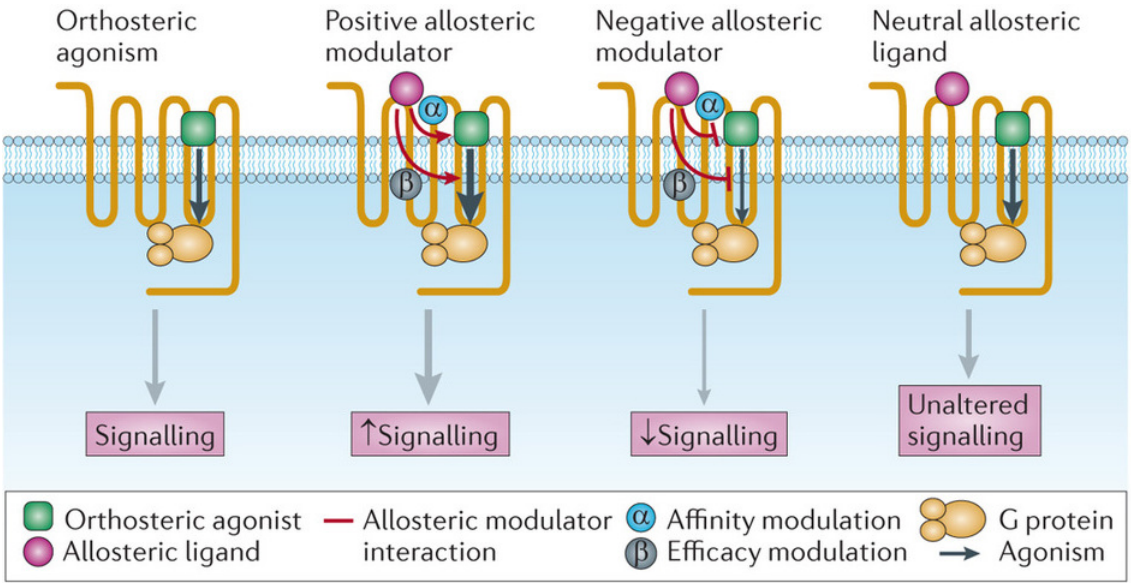


Figure 1.6: Biased Signaling in GPCRs.

Endogenous ligand bind to the GPCR, this induced the conformational change which result in downstream signaling. Positive allosteric modulators (PAMs) are allosteric ligands that bind to site distant from orthosteric agonist and enhance the affinity(α) and efficacy(β) of orthosteric ligand. Negative allosteric modulators (PAMs) are allosteric ligands that bind to site distant from orthosteric agonist and decrease the affinity(α) and efficacy(β) of orthosteric ligand. When allosteric ligand has no effect on the signaling its neutral allosteric ligand.$modifiedfromNatureReviewsDrugDiscovery12, 630644(2013)$

.

local environment and chemical shift was observed of modified Cys at the TM7H8 interface [42]. Resonance energy transfer spectroscopy compared arginine-vasopressin type 2 receptor complexed with agonist witharginine-vasopressin type 2 receptor complexed with biased agonist reveal that the receptor have different TM6 and TM7-H8 conformation which highlight different receptor activation pathway with agonist and biased agonist [55].

1.7 Functional Selectivity in GPCR

Existence of several conformational states of GPCR have been established recently. Pharmacological evidences suggest that GPCRs can adopt multiple conformation that can then lead to differential downstream signaling. The down stream signaling pathway is dependent upon the type of ligand bound to receptor causing functional selectivity of ligand and biased signaling. The idea of functional selectivity is based on ability of receptor to bind with different ligand, activate selective pathways and then lead to multiple downstream signaling. Studies on β-arrestin suggest apart from its role in cellular internalization, it can be also signaling protein and select its own subset of signaling molecule. Thus there exist a differential pathways which are independent from G protein binding.

Other examples in biased signaling is of serotonin type 2A and 2C receptors. Studies found out phospholipase C-mediated inositol phosphate (IP) or phospholipase A2-mediated arachidonic acid (AA) release can be regulated by different ligand binding. Maximal response of biased ligand for each signalling pathways was different to indicated biased signaling.

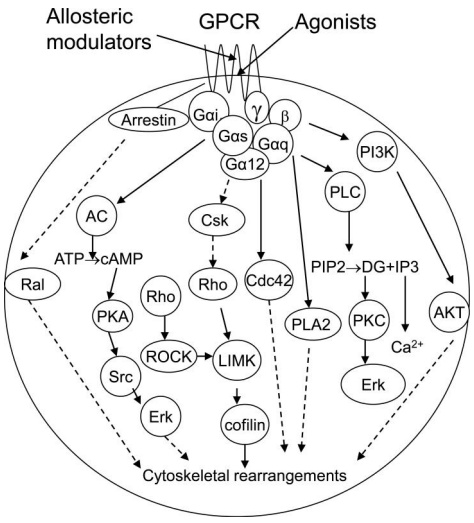


Figure 1.7: Functional selectivity in GPCRs.

GPCRs interact with different G protein families (i.e. Gi/o, Gs, and Gq). Activated GPCR-G proteins complex result in downstream signaling like adenylyl cyclase (AC), mitogen-activated protein kinase (MAPK) and phosphatidylinositol 3-kinase (PI3K) pathways. The downstream signaling pathway leads to signaling cascade like activation of MAPK and PI3K pathways. The MAPK and PI3K signaling leads to phosphorylation of extracellular signal-regulated kinases (ERKs) and protein kinase B (PKB, also . GPCR alternatively can also interact with arrestins to lead independent signaling pathway other than G protein. figure modified from [25]

.

1.8 GPCR life cycle

GPCR at cell surface when activate by ligands like hormones, neurotransmitter, nucleotides, light-sensitive compounds, chemokines and growth factors undergo structural conformational changes which lead to activation of hetrotrimeric guanine-nucleotide which is enabled by GTP -GDP exchange on α subunit. Thus further downstream signaling is initiated until G-protein in inactivated by intrinsic GTPase of the -subunit hydrolyzes the bound GTP to GDP. and inactivates the G protein.

Rapid desensitization of GPCRs is acheived after ligand is in the binding pocket for long. Phosphorylation of the intracellular residue(mainly serine and threonine residues) of ICL3 and carboxy terminus lead to deactivation [41]. In case of rhodopsin desensitization is achieved by phosphorylation and is a crucial for regulation of receptor.

The phosphorylation lead to modification of intracellular side of receptor and which further leads to β-arrestin recruitment and then desensitization and internalization results in blocking the receptors from binding to G-protein [41]. GPCR-kinases and β-arrestins are recruited as desensitization of GPCRs and removal of active ligand-GPCRs from the plasma membrane.

The GPCR-kinases(GRK) family constitute seven members that are structural and sequential homolog of each other. homology (reviewed in (Ferguson, 2001; Penela et al., 2006)). GRK2 and GRK3 reside are cytosolic but translocate to the membrane upon activation of GPCR. Activation of GPCRs in mediated by heterotrimeric G protein β-subunits. GRK5 and GRK6,are located in plasma membrane are constitutively localized to the plasma membrane. Electrostatic interaction between the carboxyl terminus of GRK5 and phospholipids at the plasma membrane is step towards phosphorylation while in GRK6 palmitoylation is required for membrane association. For

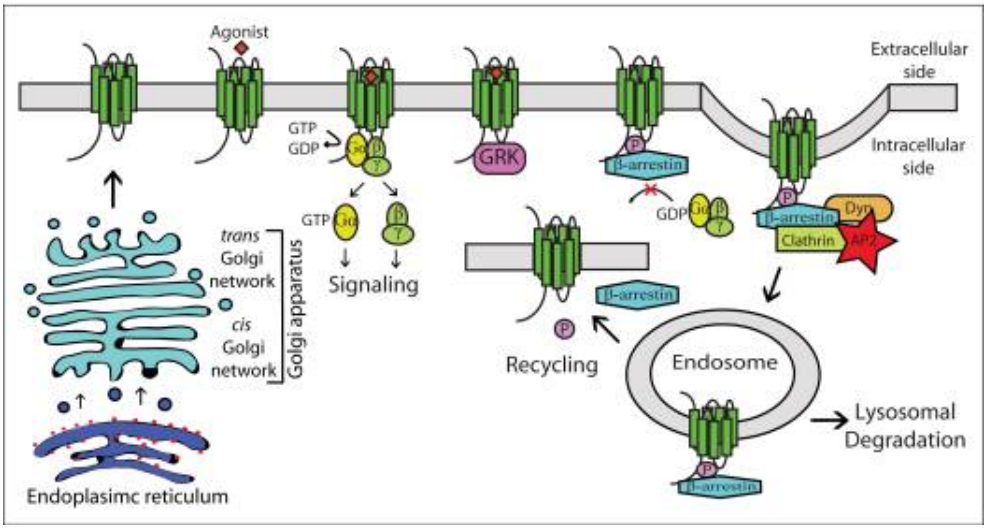


Figure 1.8: Life Cycle of GPCRs.

Synthesis of GPCRs starts from endoplasmic reticulum (ER). Receptors is transported from the ER to plasma membrane through via Golgi complex. Agonist stimulation is when GPCRs bind and activate a G protein, results in dissociation leading downstream signaling pathways. Long exposure to agonist ligands leads to desensitization and internalization of the receptor via phosphorylation and recruitment of β-arrestin recruit. modified from ([72])

.

desensitization β arrestin serves as recruiter and in establishing link between the receptor and endocytic machinery.β arrestin exist in two sub forms.β-1 and β-2. β-arrestin participate in recpetor internalization through direct interaction of clathrin coated complex, namely AP-2 complex [40].

After internalization, receptors traffic to lysosomes for degradation or follow a recycling route back up to the cell surface [13]. Recent study indicate stimulation of β-adrenergic receptor by promoting direct association of the carboxy terminus of the receptor with first PDZ domain with Na+/H+-exchanger regulatory factor-1 (NHERF1) [18].

Chapter 2

Activation and Dynamic Networks in Cannabinoid Receptors

2.1 Abstract

G-protein Coupled Receptors (GPCRs) are the largest and one of most important class of membrane protein that serve as drug target for multiple diseases. GPCRs in response to neurotransmitter and hormones undergo conformational changes resulting in activation of complex and initiate signaling which further results in a cellular response. The ligand binding often regulated biased signaling, which leads to same receptor preferentially activating different signaling pathways depending on the ligand bound to the receptor. The basis of theses biases is yet to be completed deciphered. Brain cannabinoid receptor (CB1) mediates neurological processes in response to psychoactive compounds like marijuana. CB1 is structurally similar to rhodopsin class GPCR. Ligand specific CB1 interaction has been used to develop an improved, yet still incomplete, understanding of the signaling and activation mechanisms of GPCRs. CB1 can serve as crucial model to study the characteristic dynamics and

catalytic efficiency of GPCRs. Over the past several years many studies combining mutation experiments and computational modeling have been employed to explain the activation mechanism and signaling of GPCRs, yet still much has to be learned about both the CB1 in particular and GPCRs in general. Our study explored the conformational landscape of CB1 as a function of activity and catalytic efficiency of CB1 receptor. Change in correlated motion CB1 residues in lowest frequency vibrational global modes starting from inactive structure highlighted presence of intermediate to active structure .Biochemical studies and crystallography method coupled molecular dynamic simulations have characterized role of TM3 and so called "molecular switches" in receptor activation. Highly conserved residues $D101^{3.49}$, $R102^{3.50}$, $Y103^{3.51}$ of TM3 forms the basis of receptor activation . $D101^{3.49}$ towards the cytoplasmic end of GPCRs interacts with ICL2 while $R102^{3.50}$ interacts with the carbonyl backbone of G protein. Biochemical evidences indicate interaction of TM3 residues with the residues of TM4-TM6-TM7 is critical for receptor function. Mutations in many of the residues of TM3 helix has lead to modulation in receptor activity.

These evidence suggests that the TM3 is critical for the functions of CB1. Experimental evidences studying the allostery mechanism with differential ligands highlight biased in allosteric signaling. Molecular dynamic simulation coupled with network analysis reveals different ligands can interact with G-protein coupling site using different allosteric pathway. Our community network analysis derived from running molecular dynamic simulation on CB1 reveals allosteric network in CB1.studies on these network indicate correlated motion G-protein Coupling site and ligand binding site in CB1. Coupled with previous experimental studies our study which incorporates normal mode analysis and Molecular dynamic simulation highlights dynamic allosteric communication pathways and sub networks of residues critical for the presence of receptor activity in CB1.These sub networks and communication pathway can

be further crucial to understand allosteric communication pathway in GPCRs.

2.2 Introduction

Cannabinoids(CB1) Receptors[34] responsible for the numerous neurological responses result of psychoactive ligand like marijuana belongs to class A seven transmembrane(TM) G-protein Coupled receptor(GPCR). CB1 when couples with G-proteins inhibit the activity of adenylyl cyclase and in turn regulate the different ion channel activity[60] including potassium ion channel[15] and cAMP dependent calcium ion channel[51] .Through different signaling pathways CB1 is involved in multiple diseases including drug addiction,anxiety disorders, chronic pains and obesity.[75] Recent studies have expanded our understanding of GPCR signaling.Experimental evidences suggest receptor can adopt different conformations depending upon ligands bound to it. Differential signaling corresponds to activity of receptor.These ligand induce different conformation on same receptor can explain functional selectivity or biased signaling in GPCRs.Study of dynamic information from correlated motion of residues in GPCRs can further provide detailed picture of allosteric communication that can further be related to experimental observation

Cannabinoid receptors shows partial constitutive activity -i.e there exist basal activity independent of agonist binding.Presence of basal activity in a receptor without ligand suggests existence of conformation ensemble of structures in different functional states [66]. Presence of basal activity even in the absence of ligand indicates the conformational shift in structures due to the active functional domain of receptor. Binding of agonist of agonist and antagonist triggers the structural changes that corresponds to the either increase in signaling (agonist)or decrease in signaling (inverse agonist).

The Cannabinoid receptors are distribute mainly in central peripheral nervous system[34] and displays partial constitutive activity. CB1 is found mainly in brain and on activation by the marijuana and related psychoactive compounds induces hypertension,appetite stimulation and anxiety [75]. Cannabinoid Receptors shares a structural similarity with class- A GPCR[26]. which indicate the activation role similar to the beta-2 adrenergic receptors and rhodopsin. Lack of current crystal structures of cannabinoid receptor makes understanding of receptor's activation and it coupling to its G-protein challenging .Structural similarities and conservation of the motif between CB1 and beta-2 adrenergic receptors formed the basis of structural modeling of CB1. This will provide insight to the activation and role of residues in functional activity of Cannabinoid receptors .

GPCRs starts with N-terminal of extracellular side of membrane with seven transmembrane region (TM1 through TM7)spanning inside the bi-lipid layer with helix 8 located at C terminal of intracellular region.Transmembrane region are connected with 3 extracellular loops(ECL1 to ECL3) and three intracellular loops(ICL1 to ICL3)[26].

With the advances in protein engineering and novel methods in crystallography, high resolution crystal structures of many class A receptors have been isolated.Most of the class-A crystal structures have been in inactive conformation bounded to ligand that maintains basal activity or reduces it .The class A GPCR structures of rhodopsin[74],beta adrenergic[57] and muscarinic[38]Haga2012 receptors provide significant information about the inactive state of GPCR. Bovine rhodopsin,human beta-2 adrenergic, alpha -2 adrenergic ,rat NTSR1 and M-2 muscarinic receptors are among fewer crystal structures isolated in active or partial active state. With fewer known crystal structure in the active state our knowledge about the activation mechanism is limited.The previous studies simulated active beta-2 adrenergic

receptor to successfully obtain conformational ensemble representing inactive beta-2 adrenergic receptor in micro timescale simulation from active structure [21].The study on beta adrenergic receptor obtain three different conformation representing active ,inactive and transition intermediate.Study highlighted significant difference in terms of structural rearrangement of G-protein binding sites and helix orientation in three states.Study on muscarinic receptors also captured binding of the antagonist to extracellular vestibule .Though activation in muscarinic receptor was not detected and receptor remain inactive and ligand free through out the simulation[38].Above two mentioned studies highlight the complexity of identifying the GPCR activation even in the longest simulations and required much more computational resource with no definitive certainty.

Role of CB1 in diseased state like drug addiction,anxiety ,depression and in chronic pain is been highlighted in previous studies[75]. Due to lack of crystal structures and details of the its activation mechanism to couple with its G-protein is still unclear.The role of correlated motion for allosteric regulating CB1 activity from the extracellular binding site to G-protein coupling site and the conformational changes in the receptor upon activity is an open ended question. Our present study aims to understands the activation mechanism and how allosteric regulation in CB1 is achieved. Our study also characteristically determine and identify the role of intermediate thermodynamic ensemble during activation . Normal mode analysis (NMA) is method to analyze collective motion in protein expressed by superimposing some collective variable. These collective variable is often termed as normal modes and are obtained as linear combination of atom based coordinates. [28]. On assumption of system is stabilized in harmonic potential ,normal mode explores the functional motions of protein[2]. Full atomic molecular dynamic (MD) simulation of CB1 receptor is performed to identify the role of the functional residues previous reported from

experimental mutation studies through community network analysis.

In the present study,we used Normal mode analysis (NMA) coupled with correlated motion analysis to explore the origin of conformations transition in cannabinoid receptor. NMA analysis revealed the correlated motion between allosteric binding site and G- Protein coupling site revealing allostery in CB1. Role of TM3 as function hub of CB1 was established by studying correlated motion in thermodynamic ensemble of CB1.

We identified highly dynamic allosteric network in CB1 and our present study also explored the functional communities in the network. The functional communities relates to the community that are formed through highly correlated residue. The community networks helps in identifying the information transfer in the Cannabinoid receptor. Community network analysis based on cross correlation motion analysis also asses the strength of communication between multiple local substructure in protein. Allosteric signaling in CB1 involved different community network based on ligand binding to the allosteric binding site.

2.3 Materials and Method

2.3.1 Input Structure and Model Preparation

Due to absence of crystal structure of Cannabinoid receptor was homology modelled. The sequence alignment was performed using Clustal-Omega[71] to align the CB1 receptor sequence. MODELLAR[64] was used to homology model CB1.The structural information from x ray structure of beta-2 adrenergric [37] receptor was incorporated.For helical bundle in CB1 ,helical boundaries were assigned by STRIDE.Beta-2 adrenergic receptor was used to define helical boundaries.Extracellular loops(ECL)

and intracellular loops (ICL) are integral part of receptor as it stabilizes the helical region of receptor.Loops were assigned based on the previously published work by SHIMJY[68, 69]. Modeller was used for building loops for the receptor.Structural information for loops were also derived from the structural information from beta-2 adrenergic receptor. N-terminus and C-terminus of the receptor were capped with acetyl group and an N-methyl group respectively.All the other residues were kept in their ionizable state.in the C-terminal end, helix H8 after Helix H7 upto C300 was retained.The cannabinoid receptor was inserted into a palmitoyl-oleoyl-phosphatidyl-choline (POPC) bilayer.Overlapping lipid molecules were removed using the Membrane plugin in VMD[32].Using 0.5 M NaCl the system was ionized and charges were then neutralized. Solvate plugin was used in VMD [**?**] the process. Initially for simulation Cannabinoid Receptor measured nearly 93 x 93 x 90 $Å^3$ and contained 126 Na+, 140 Cl- and 13,167 water molecules, for a total of 67,126atoms.Periodic boundary condition were applied in the simulations.

2.3.2 Molecular Dynamic Simulation

For Molecular Dynamic simulations NAMD2.9[53] was used.The parameters set for simulation used was CHARMM32 for protein and POPC lipid,TIP3 model for water molecule. 12 Å distance cutoff was set for the van der Waals, particle mesh Ewald summation method was used to compute long-range and and short-range electrostatic interactions [78] with grid point density of 1Å.For calculation of the van der Waals and electrostatic interactions pair list was set to 13.5 Å. A 2fs integration time-step was used for all MD simulations,multiple-time-stepping algorithm[53]was employed with long-range electrostatic interactions and short-range non bonded interactions computed very two time-steps and one- time steps respectively.SHAKE

[1]algorithm was used to be applied on all hydrogen bonds. For first 0.5 ns with only lipid allowed to move and all other atom fixed ,lipid tail was minimized for 2000 step with NVT run and at 300K lipid tails were allowed to be melt using conjugated gradient algorithm .The system was further protein restrained with force constant of 2kcal/molÅ for 1.5 ns protein , with membrane allowed to move with first 0.5 ns external force applied using NPT ensemble at 300 K .Remaining 1 ns protein backbone was harmonic position restrain at 310 K in NPT run allowing further equilibration .After the minimization and equilibration production run was performed on system and simulated at 310 K for 94 ns in NPT ensemble.

2.3.3 Normal Mode Analysis

Normal mode analysis of the Canabinoid receptor was performed using two step process. Initial protein was minimized in vacuum with initial constrain of 110 kcal/molÅ and relaxed gradually with decrease of 1kcal/molÅ. minimized structure was used to compute the global modes of protein. Vibran module in CHARMM [7] was used to compute hessian matrix and the eigen value and cross correlation analysis.The global normal modes were computed and the modes were shown to converge using correlation matrix with 50 global modes.The cross correlation matrix was computed with individual modes and with the superimposition of first six nontrivial modes and first 44 nontrivial modes .

2.3.4 Generalized Cross-Correlation Analysis

To calculation of the cross-correlations of residues in the CB1 receptor are calculated based on information between all C-alpha atoms in the receptor using the

generalized correlation analysis approach.Corman module from CHARMM was used

$$C_{(i,j)} = \frac{< \Delta r_i \bullet \Delta r_j >}{< \Delta r_i^2 >^{1/2} < \Delta r_j^2 >^{1/2}} \tag{2.1}$$

2.3.5 Community Network Analysis

Allosteric network in the CB1 receptor is analyzed by community network analysis using NetworkView module in VMD [?]. A network graph with each protein residue represented as node is constructed. Edges are defined in the network by connecting pairs of nodes.There exist a edge between two nodes when any two heavy atom are in contact within 4.5 Å for greater than 75% of the simulation time. The weight of edge is defined by the correlation values of the two nodes wij = -($|logCij|$) . The Girvan-Newman algorithm is used for determine the communities in the the network.The algorithm divides the protein network into communities of highly intra-connected but loosely inter-connected nodes. The betweenness or number of shortest path that cross the edge is calculates and then the modularity score is calculated by iteratively removing the node of highest betweenness. Critical nodes are present at the interface of neighboring communities and connect one community to other through an edge.These edges describe the probability of information transfer and allosteric communications

2.3.6 Characteristic Path Length Analysis

Allosteric signal transmission in CB1 involves the interaction and communication of dynamical information within protein.The allosteric binding site of the receptor initiate the signal which gradually gets transmitted into orthosteric binding

site. To identify residues that have the largest effect on communication across the interface, change in characteristic path length is used as metric. Changes in CPL is calculated by removing all the interlinks from a defined interface to any given residue while other connection in networks is not perturbed.

$$CPL = \frac{\sum (i,j) D(i,j)}{Npairs} \tag{2.2}$$

CPL is the characteristic path length and is mean distance between all residue pair in protein network. where D_{ij} between i^{th} and j^{th} node equals to the summation of edge weights between the consecutive nodes (a,b) along the path: Dij $=\sum_{a,b} w_{a,b}$

2.3.7 Modifications Alter the Network

Mutation of the functional residues contributes to the reduce in the signaling of the receptor.Role of functional residues on the activity of the receptor and the community network is assessed. Edges between functional residues and any residue on the receptor in the network was weakened to modify the network. network analysis computationally without carrying any MD. The community organization was assessed and the change in the characteristic path length was analyzed.The The contribution of a residue k in communication within the network is analyzed by recalculating the CPL after removing residue k or its connecting contacts. Role of the robustness of the community was assessed by removing the residue and quantifying the change in the community structure through community repartition difference.

2.4 Result and Discussion

Starting from modeled inactive structure of Cannabinoid Receptor from beta-2 adrenergic receptor (PDB code : 2RH1) we examined vibrational frequency global mode to understand the conformational shift among its various conformation state. Our results indicate the lowest frequency vibrational mode contributes the most to the change in conformational changes and correlated motion in a protein is linear combination of the first six vibrational modes. Past studies also support our study as it has shown that the low-frequency modes describing the large-scale motions of a protein and can be related to fundamental biological characteristics and low frequency modes overlap with protein conformational change in real world.

2.4.1 Changes in TM3 correlation reveals its functional role in receptor biogenesis.

The role of TM3 as functional hub and structural hub in GPCRs have been previously established[81]. The interaction of TM3 with TM5 ,TM6 and TM7 have been critical in receptor activation[81, 61]. The receptor activation leads to common structural changes in transmembrane regions.Small structural distortion in TM5 induced by local structural changes in Pro5.50; TM3 helix and TM7 helix relocation, translation of TM5 and rotation of TM1/TM5/TM6 are few events characteristic of receptor activation and receptor ligand interaction. The changes in interaction of residues of TM3-TM6-TM7 is required for receptor biogenesis.

We analyzed changes in correlation motion of TM3 with TM5 ,TM6 in low frequency modes to identify the biogenesis of receptor activation. On mapping interaction of residues of TM3 with TM6 residues we identified the residues with changes

(greater than 0.8) in correlation of residues(Fig. 2.1). Significant changes in correlation seems to provide insight towards biogenesis and allosteric modulation of CB1. Our current findings Fig. 2.1C suggest change in the interaction of TM3-TM6 mainly involves F77$^{3.25}$ G82$^{3.30}$ G83$^{3.31}$ and I242$^{6.46}$(with all three above mentioned residues), G245$^{6.49}$(with F77$^{3.25}$ and G83$^{3.31}$), P246$^{6.50}$(with G83$^{3.31}$) as shown in Fig. 2.1C. The changes in interaction of TM3 and TM5 with respect to correlated motion in low frequency modes are mapped to identify conformational transition as these contacts are critical for receptor biogenesis Fig. 2.1B.

Our studies with normal mode analysis on CB1 highlights the interaction involving L175$^{5.51}$,L174$^{5.50}$,and $^{5.58}$ with G83$^{3.31}$ and interaction of Y182$^{5.58}$ and M183$^{5.59}$ with multiple TM3 contacts Fig. 2.1B. On mapping interaction of TM3-TM4, residues L140$^{4.61}$ P139$^{4.6}$ C126$^{4.47}$ were involved in change in multiple interaction mostly with the middle portion TM3 residues highlighting TM3 as functional hub of receptor(Fig. 2.1A). These local interaction are translated into larger helical movement and are likely responsible for allosteric signaling and receptor activation pathway of the receptor.

Past studies on family of GPCRs indicate the relocation of TM7 and TM3 after activation. The residues of TM2/TM3/TM7 forms minor binding pocket. We mapped how the changes in correlation of TM3 and TM7 have induced the the changes in the residues of minor binding pocket as mentioned above is critical for the activation. Previous studies on CB1 employing Alanine mutagenesis highlighted L81$^{3.29}$, T85$^{3.33}$, F267$^{7.35}$ might be ligand contact resides. Significant changes(greater than 0.80) in interaction of G83$^{3.31}$ with V280$^{7.48}$, N281$^{7.49}$, I283$^{7.51}$,I284$^{7.52}$, Y285$^{7.53}$ similarly residue F267$^{7.35}$ was not much affected but nearby residues F269$^{7.37}$, C270$^{7.38}$, S271$^{7.39}$ were experienced significant change in correlated motion with TM3 residues indicating there might be the intermediate conformation in first six low frequency

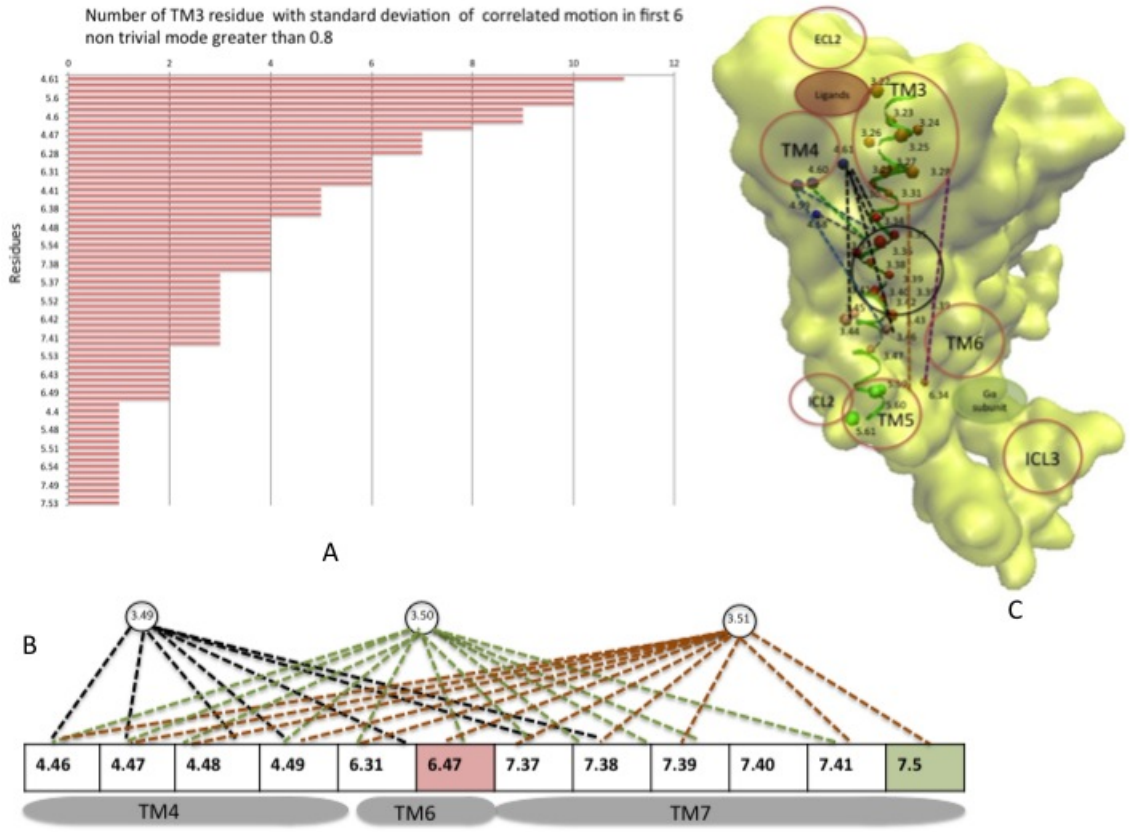


Figure 2.1: Change in the interaction of TM3 highlights the modes representing structural variation towards activity

Figure indicates the number of residues in TM3 with standard deviation of correlated motion in first 6 non trivial mode greater than 0.8 with TM4 TM5 TM6 TM 7.TM3 is functional hub and the its interaction with TM4,TM5,TM6,TM7 is suggested to as requirement for biogenesis, stability or functionality. of GPCRs. Mapping the change in correlated motion interaction in these region for first 6 non trivial mode will give us better understanding to biogenesis of receptor activation. B) The figure highlight the residues standard deviation of correlated motion of residues in TM4 TM5 TM6 with highest number of changes in interaction with number of residues in TM3.The ligand binding site and G-protein are shown in red and green respectively. B) Figure details Role of Molecular Switches in terms of standard deviation in correlated motion with interacting residues of TM4 TM5 TM6 TM7.Molecular switches in TM6 TM7 are shown in red and green respectively

.

mode. (Table 2.1, Figure 2.2, Figure 2.1D).

We mapped interaction of the TM3 residues whose correlated motions were significantly affected by TM4,TM5 TM6 TM7. We identified correlation motion based interaction which significantly changed with residues L140$^{4.61}$, L138$^{4.59}$,P139$^{4.6}$ in TM4 , residues M183$^{5.59}$,Y184$^{5.6}$,I185$^{5.61}$ in TM5 and A230$^{6.34}$ in TM6 with multiple contacts (greater than 0.8) with TM3 residues. The correlated motion induced structural change in these six lowest frequency modes with TM5, TM6 residues more prominent change in interaction with the extracellular side of TM3 and TM4 with central side of TM3 helical bundle. These correlated motion changes in six low frequency modes with change near ligand site and G alpha binding site enabled to observe presence of multiple conformation starting from initial inactive structure. These local structural changes close to binding site of receptor are shown to translate resulting in large helical movement in G-protein coupes receptors. The observed changes near the ligand binding site and the change in the G-protein binding site in six low frequency modes is also informs about the existence of multiple conformation state and the distinct signaling in receptors is induced after the binding of ligand. With our results indicates the first six modes are associated with conformational change in CB1 receptors these ligand induced structural rearrangement in the TM helices of receptor can be represented by the low frequency global modes.

Ligand	Functional Residue	Position
anandamide	F77	3.25
	K80	3.28
	L95	3.43
	T98	3.46
	D101	3.49
	Y163	5.39
	Y182	5.58
	A230	6.34
CP55940	H69	E1
	R70	E1
	D72	E1
	V76	3.24
	F77	3.25
	K80	3.28
	L95	3.43
	T98	3.46
	D101	3.49
	C145	E2
	C152	E2
	F156	E2
	P157	E2
	H158	E2
	I159	E2
	Y163	5.39
	Y182	5.58
	A230	6.34
	C243	6.47
	S271	7.39
WIN55212-2	D51	2.5
	K80	3.28
	G83	3.31
	F88	3.36
	L94	3.42
	T98	3.46
	D101	3.49
	Y163	5.39
	W167	5.43
	V170	5.46
	Y182	5.58
	A230	6.34
SR141716A	K80	3.28
	F88	3.36
	L95	3.43
	T98	3.46
	C145	E2
	C152	E2
	W167	5.43
	W244	6.48
	C274	7.42

Table 2.1: Experimental studies revealed key functional residues of Cannabinoid receptor that are crucial for receptor function

2.4.2 Role of Molecular switches in low frequency modes towards structural and functional changes in TM

Molecular switches are the highly conserved residues in GPCRs [46]. Previous research on activation mechanism of GPCRs highlight the role of molecular switches. Conformational changes in the receptor requires destabilization of normal arrangement. The molecular switches in TM3 through ionic lock with TM6 stabilizes the inactive state of receptor. Ionic lock are formed as result of interaction between an arginine in transmembrane TM3 (R3.50) and a negatively charged residue in TM6 (D/E6.30)[22]. Breaking of ionic lock has been suggest to activate the receptor. In six low frequency modes we examine the change of interaction with respect to cross correlation to find out the the the molecular switch in the TM3 have significant change in interactions with TM4, TM6, TM7.

Residues F125$^{4.46}$, C126$^{4.47}$, L127$^{4.48}$, M128$^{4.49}$ from TM4, residue I227$^{6.31}$ and C243$^{6.47}$ (molecular switch of TM6) from TM6 and residues from TM7 F269$^{7.37}$,C270$^{7.38}$, S271$^{7.39}$, M272$^{7.40}$, L273$^{7.41}$, P281$^{7.5}$(molecular switch in TM7). Change in interaction of TM3 molecular switches with TM4 TM6 TM7 highlights the structural rearrangement in these low frequency modes. These changes that suggest as local structural rearrangements that can lead to partial active or fully active receptor.(See Fig. 2.2) Past studies indicate the rearrangement of the molecular switch [81] during the receptor activation. Above mentioned structures change must therefore relate either structural changes to an initial receptor-ligand complex or to intermediate active states that has not yet undergone conformational changes which will allow eventually binding of the hetrotrimeric G protein. Thus, it seems molecular switch facilitate the receptors not only to reach fully active conformation upon G-protein bind but also constitute as a common feature of GPCR activation.[22]

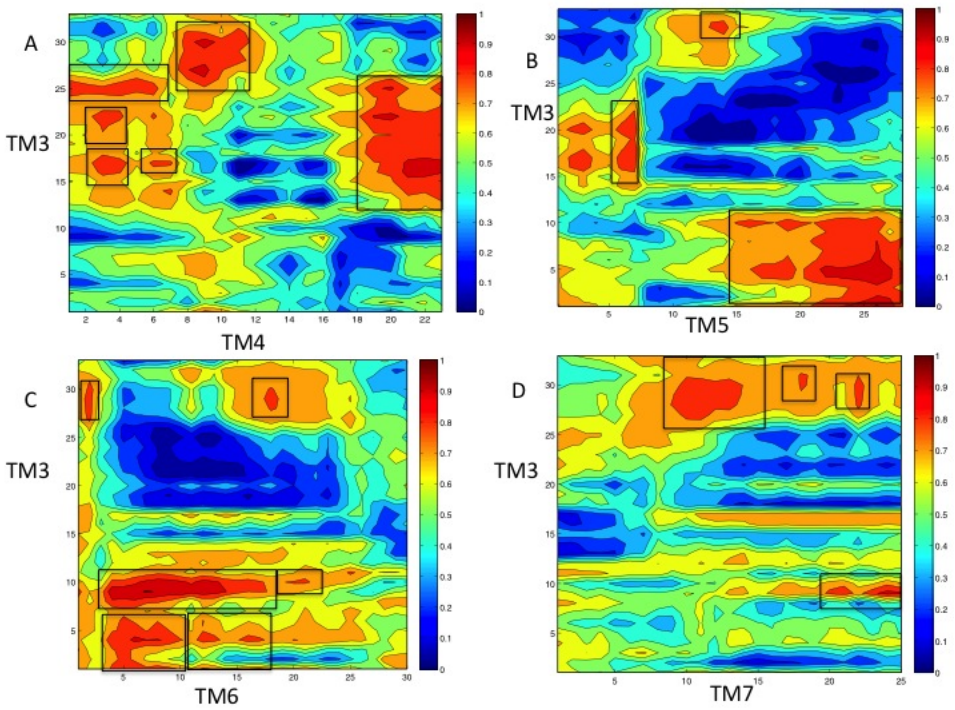


Figure 2.2: Mapping Role of TM3 interaction with TM4-TM5-TM6-TM7s

Role of position in TM3 is indicated to be important for maintaining structure or function of GPCRs. Large tilt-angle of TM3 contributes largely to this behavior. The consensus contact network formed by the residues defines GPCR fold. TM3 facilitate contacts with widely diverse pharmacology ligands and have shown to the form of the "ligand-binding cradle" [81]. We here map the changes in TM3-TM6-TM7, TM3-TM4 and TM3-TM5.Above mentioned regions are the residues with the stronger standard deviation in turn represent the conformational change that occur in the low frequency motion through mode 7 to 12.

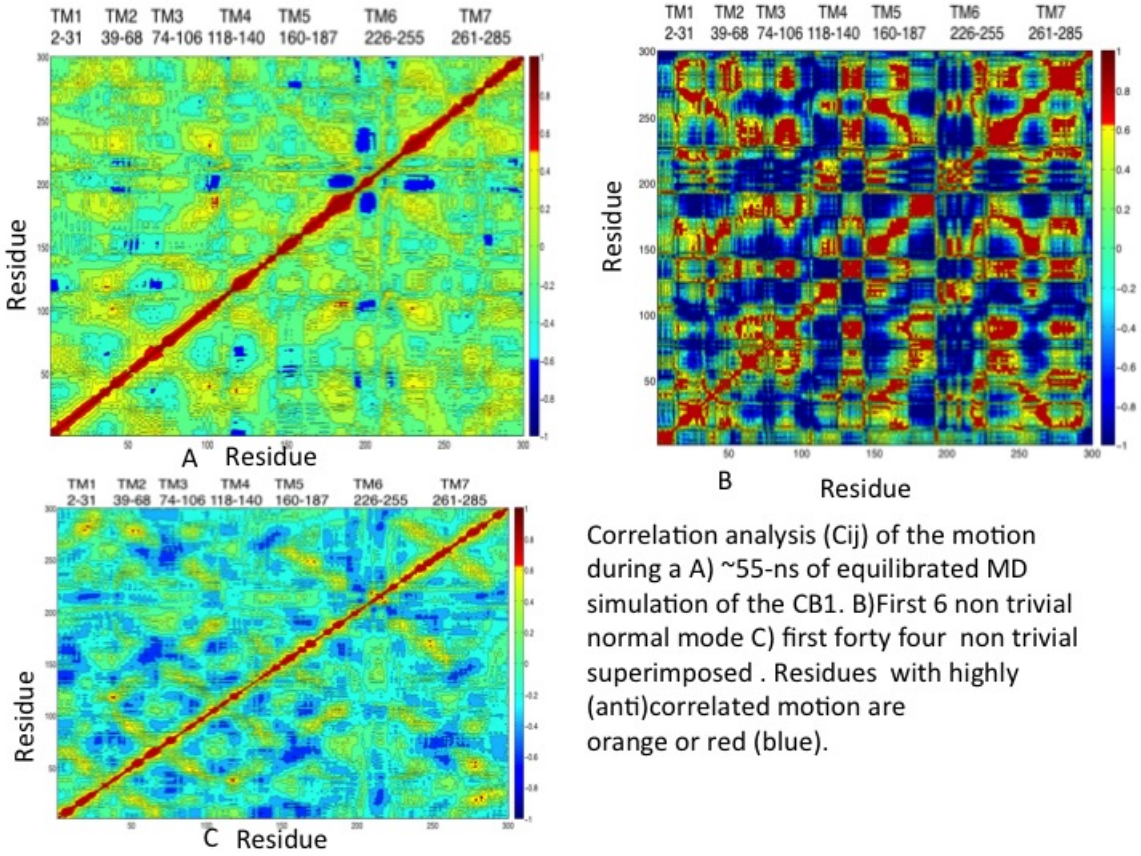


Figure 2.3: Cross correlation analysis of CB1

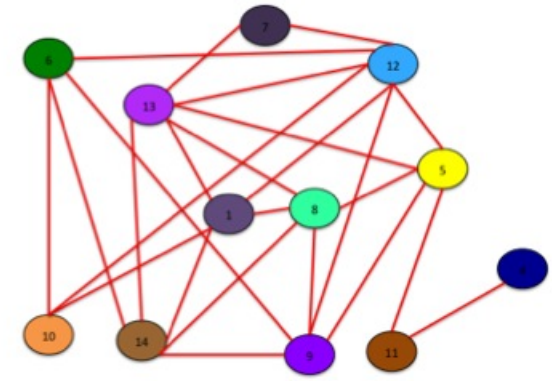
Positive Correlated motions (with distance greater than 10 A)are identified between residues in the TM3 with TM5 and TM6 regions identified interaction zone where allosteric ligand binds.Regions of intracellular loops are strongly correlated with residues TM4 in molecular dynamic trajectory. The region here is color-coded dynamic map of residues' cross-correlations. Residue pairs with cross correlation greater than 0.6 is colored red

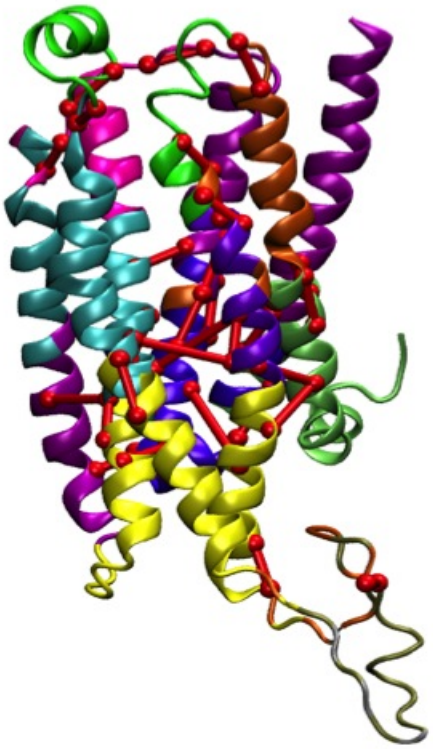
.

2.4.3 Correlation motion between allosteric binding site and G-Protein coupling site indicates allostery in CB1

Allosteric signaling is one of the important aspect related to activation of Cannabinoid receptors. Single receptor through structural conformation changes can be involved in different signaling pathway. Different signaling pathways are initiated with different allosteric ligand. We used inactive structure to explore the residues in receptor involved in allosteric signaling.

Transmission of allosteric signal in a receptor should involve coupled motion between allosteric binding site and G-protein coupling site. The degree of correlation was measured by cross correlation matrix which recorded the atomic fluctuation over the length of molecular dynamic simulation. The distant coupling parts are shown in boxed region in Figure 2.3. Motion of residues in TM3 helices have positive correlation with the TM5 residues region. The residues from molecular switch of TM3 3.51 and nearby residue 3.52, 3.53, 3.54 is in strong correlated motion with distal end of TM5 and initial TM6 residues comprising of residues 5.58, 5.59, 5.60, 5.61, 6.28, 6.29. Residue 3.48 from TM3 domain is strongly coupled with the intracellular part of TM5. The correlated motion between residues helps to indicate the correlated motion in G-protein binding site. Ligand binding in the allosteric binding site lead to the movements and results in the shift of the TM3 region. These shift is driven mainly by rearrangement of conserved hydrophobic cluster and aromatic residues and by the strong correlation between the molecular switch of TM3 3.51 and nearby residue 3.52, 3.53, 3.54. The strong correlated motion of TM3 with distal end of TM5 and TM6 residues comprising of residues 5.58, 5.59, 5.60, 5.61, 6.28, 6.29 leads to further rearrangement at the TM3-TM5 interface, which eventually lead to formation of new non-covalent contacts at the TM5-TM6 interface.





Community 1: part of (H1+ H2 + ECL1)
Community 2: part of ICL3
Community 3: part of ICL3
Community 4: part of ICL3
Community 5:part of (H3+ICL2+H5+ I3+H6
Community 6:part of (E2+ H6 + E3)
Community 7: part of (H2 + I2+ H4)
Community 8: part of (H1+H7+H8)
Community 9: part of H6
Community 10:part of (E1+ H3+H4+E2)
Community 11: part of ICL3
Community 12: part of (H3+H4 +E2 + H5)
Community 13: part of (H1+ I1+ H2+H3)
Community 14: part of (H2+ H6+ E3+H7)

Figure 2.4: Dynamic Network of CB1

Dynamic Network in CB1 is highlighted according to community membership calculated. Community are color coded and the helixes and loops corresponding to the community are highlighted in the structure. The critical nodes and edges are colored by red and signifies the communication across communities.

.

Past studies reveals the the many of the residues in this transmission switch are highly conserved in class A GPCRs, and are believe to be associated as a common feature of GPCR activation [16, 17, 57] Based on previous studies on beta-adrenergic receptors and rhodopsin identified G-protein binding site in class A receptors are mostly formed with residues near TM3 (3.50, 3.53, 3.54), TM5(5.61,5.64,5.65), TM6(6.26, 6.29, 6.32, 6.33, 6.36, 6.37, 6.40) [57]. These transmembrane residues form contact with ligands and the extent of these have shown to correlate with ligand efficacy. Thus, we conclude these key positions in TM3, TM6 and TM7 not only contribute ligand specificity but also determine allosteric differential signaling in the receptors.

Other highly positive coupling between distant parts exist between ICL1 and TM3 residues. Coupling between Residues of TM1 and TM7 and extracellular domain of TM6 and TM2 indicates the presence of long distance communication between the distal domains of receptor. Increased correlation motion of extracellular domain of TM6 and extracellular domain TM5 is observed .Weak correlated coupling (cross correlation >0.6) with extracellular domain of TM3 (3.28,3.29,3.30) and ECL2 highlights receptor is in inactive state and allosteric communication is weak(Fig. 2.3).

2.4.3.1 Highly Dynamic Allosteric Network in Cannabinoid Receptor

Correlation analysis provided indicated signals of allostery in the inactive receptor. To completely understand the signal transmission in the receptor further analysis is required and correlation map can't complete elucidate allosteric signaling in receptor. We identified highly dynamic network in CB1 via community network analysis.

Allosteric signal in a receptor involves transfer of dynamic information with in a receptor. The dynamic network in CB1 was identified using community network analysis. The residues in the network represent a node and strength of correlation between the two nodes defines the edge between them. The edges reduces as the correlation between the two nodes increases. The community network analysis revealed the receptor to be the highly connected cluster of 14 major communities.Fig. 2.4. The residues in same communities can constitute from different helical domain of the protein. The residues in same cluster are shown to be highly correlated. Among 14 communities community number 5 is identified as one of the community which carry signaling pipeline from extracellular region to the G-protein binding site. The Community 5 consist of residues from TM3, TM5, TM6. Residue in community 5 are 3.50, 3.53, 3.54 from TM3, 5.61 from TM5 and 6.29, 6.32, 6.33, 6.37,6.40 from TM6 which constitute G-protein binding region for class A receptor.

In Cannabinoid the ligand binding pocket varies with the ligands interacting with the receptor. The ligand binding pocket in cannabinoid receptors has not been well defined. Binding pockets varies and binding region depends upon the class of the ligand to the receptor with some partial overlap of residues participating in the ligand binding. Our community network result highlight the different pathways are involved with different participating ligands. We assessed the role of functional

residues involved in the interaction of different allosteric interactions.

These residues that are involved in the communication pipelines from the extracellular binding region to the G-protein coupling site near TM5 are distinctly different based on the allosteric ligand interacting with the ligand binding pocket. In the receptor, the communication of the clusters shows a distinct allosteric communication pathway by which the the signal propagates from the extracellular domain of TM3 (3.28,3.29,3.30) and ECL2 to the G-protein coupling site of cluster 5.(Fig 2.5 . We also observed that in the cannabinoid receptor the contact point 6.40 on TM6, allosterically communicates with 5.61 and 5.51 on TM5. These interaction have also been found in the allosteric networks for M2-AcR [46]. The published accelerated MD simulations reveals two major clusters though which the communication in the muscarinic residues were found. These cluster include TM3 and TM5 cluster, and TM6 and TM7 residues cluster. These two allosteric clusters also communicate via 5.61-6.40,5.51-3.30, similarly in the $\beta - 2$adrenergic receptor has been found to allosteric communicate via 5.61-6.40 and 5.51-3.30. Our analysis also identifies the contact region of 5.61-6.41,3.40 as connecting bridge that receives the signal from the ligand binding site and transmit signals to the G-protein coupling site. Thus, our results is in strong agreement and show similar behavior as with other published simulation studies on Muscarinic and Beta-2 adrenergic receptor. Studies with site-directed mutagenesis

Above mentioned residues reveal that interaction of 5.61 with TM6 residues near TM5 is critical for G-protein selectivity and helps to preference for Gi/o over Gq pathways [42]. Therefore, the dynamic network network involving 5.61 and 6.40 important for the allosteric network of GPCR but also essential for specificity for G-protein-coupling conformation in Cannabinoid receptor.

2.4.4 Different Allosteric Interaction involves different Community network

Cannabinoid receptors interacts with natural as as synthetic ligands. We examined four allosteric ligand Anandamide (endogenous cannabinoid neurotransmitter), CP55940(synthetic agonist neurotransmitter), SR141716 (synthetic inverse agonist), WIN55212-2 (synthetic agonist) through the functional residues previously identified crucial for allosteric activity.

In past functional residues in CB1 were determined using mutagenesis data. These functional residues play role in either ligand binding,stabilizing the receptor or in receptor activation. Using the correlation based community network analysis we identified the communities that participated in the allosteric interaction pathways through functional residues of the ligand binding activity. With Community 5 corresponding G protein coupling site we mapped the critical residues require for allosteric communication between the ligand binding region and the G-protein binding community .

Mapping allosteric pipeline in inactive receptor on interaction with Anandamide indicates strong community interaction in community 5,10,12. Fig.2.5A indicated the functional residues which are mainly located in TM3 ,TM5 and TM6 interaction mainly through the critical residues that connects and help in exchange of information in between the communities (Fig.2.5) .

With agonist ligand like CP55940 using available experimental data on functional residues the information is located in community 1,5,6,10,12,14.The interaction which involve s ECL2 ,ECL2 coupled with the extracellular TM3 interact with binding site in TM5 ,TM6 and TM7 region Figure 2.5B. With agonist like WIN55212-2 as shown in Figure 2.5C the pathways maps through communities 5,10,12,14.Interac-

tion of diffract communities indicate the multiple pathways can lead to the allosteric changes in the receptor. Experimental evidences highlight the role of functional residues located in TM2 TM3 TM5 and TM6 have played role in binding affinity of WIN55212-2 in the receptor. The receptor through the functional residues involved with WIN55212-2 forms the network of communities 5, 10, 12, 13, 14.

Inverse agonist like SR141716 have ECL2 ,TM3,TM5,TM6,TM7 involved with communities 5 ,6, 9,10,12,14 as seen in Fig. 2.5D. These finding with the help of network based approach suggest even in the inactive state of receptor there is path from agonist binding site to the the G-protein binding site. Different ligands have multiple pathways to communicate between two sites.The use different community pathway can be identified by using the functional residue that affect agonist and antagonist binding in the cannabinoid receptors.Our results suggest that different allosteric interaction can process by different community pathway interactions.(See Figures 2.5)

Allosteric Interaction with in ECL2 is a major initiator of allosteric communication in all four the ligand based community network. Residues of ECL2 communicates with TM3 via 3.29 and 3.32 near the ligand binding pocket site in all the four ligand based interactions. These interaction of ECL-2 is crucial to the receptor biogensis and allosteric communication and has been conserved in different previously studied GPCR system including muscarinic and beta-2 adrenergic receptor. In all the four ligand based allosteric communication the community 5 which is G-protein coupling site and the community 10 which is ligand binding site is strongly correlated. The difference in the communication of the two cluster is often lead to ligand binding specificity. The allosteric communication from ligand binding to the G-protein coupling is often characterized by the breaking of the salt bridge between 7.32 and ECL2 residues in GPCR systems. Also, recent findings by Dror et al.shows that the commu-

nication between the ligand binding site and G-protein coupling site involves common set of features ranging from rearrangement of TM3/TM5 to outward movement of TM6 however the extent of these changes are depended on the ligand specificity and the path of allosteric communication.

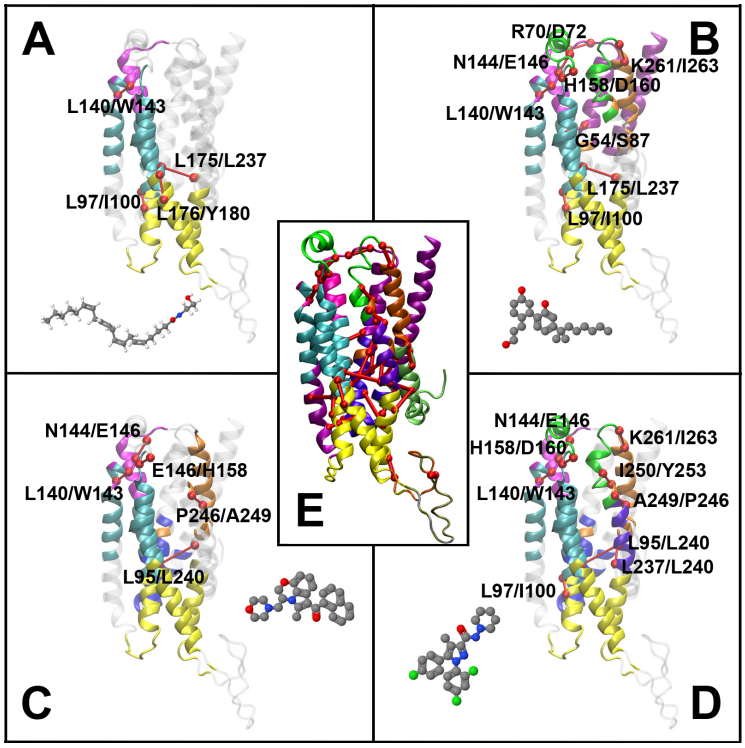


Figure 2.5: Different Allosteric interaction involves different Dynamic Community Networks

The communities involved in with functional residues in inactive state of recently homology modeled of CB1 receptor. The labeled residue represents the critical residues involved in the communication of different community that are regulated differently based on different ligands in A) Anandamide shows community 10 (magenta), community 12 (cyan) and community 5 (yellow) are functionally critical. Critical residues that participate and the helices and loop that involve in the information transfer are labeled B) CP55940 indicates role of communities: 1(purple) 10(magenta) 12 (cyan) 5(yellow) 6 (green)14(orange).CP55940 completes the information transfer while region of communities from extracellular region to the communities involving the intracellular regions.(C)WIN55212-2 initiates communications in communities 5 (yellow), 10(magenta) ,12(cyan),13 (blue),14 (orange).D)SR141716A involves communities 9(violet), 10(magenta), 12(cyan), 5 (yellow),6(green),14(orange) for information transfer through critical labelled critical residues.

Chapter 3

Identifying Allosteric Signatures in Muscarinic receptor

3.1 Abstract

G-protein Coupled Receptors(GPCRs) facilitate trigger release of GDP by allosteric activation of G-protein.GPCRs correspond to the largest class of membrane protein with nearly 800 human GPCRs and form largest super family of drug targets.Given its role in multiple physiological pathways and mediating cellular responses we explored residues critical for allosteric mechanism of GPCRs.Here we investigated the evolution of allosteric modulation and biased signaling within the muscarinic family of GPCRs.We provide demonstration that allosteric bias in muscarinic receptor is due to key residues that act as allosteric signatures in receptor conformations. This might explain and provide a basis for quantifying the allosteric modulation in GPCRs for stabilizing its particular inactive or active conformation

3.2 Author Summary

GPCRs is one of the largest family of receptor and play important role in the cell signaling. GPCR signaling is activated i) directly when G-protein binds to GPCRs (orthosteric site) 2) indirectly when the ligand binds distant to the orthosteric binding site (allosteric site) and facilitate the signaling .The differential signaling in GPCRs has been shown to exist with different ligand interacting with GPCRs. These communication is mediated via conformation changes and reorganization of contacts among the residues in receptor. Although orthosteric binding have been potential drug target G-protein Coupled Receptor (GPCR), ease of targeting allosteric sites in GPCRS makes allosteric modulation by ligands in these sites interesting candidates for better receptor selectivity and improved binding affinity. Allosteric Modulation is known to exist in GPCRs through agonist and antagonist. While there is still wide uncertainty in path of allosteric communication and how allosteric signals propagate inside the receptor recent experimental and biophysical studies identified role of "few residues" that have contributed significantly towards activity of GPCRs. Our current work focuses on identifying the allosteric signature in GPCRs. These efforts are to identify evolution of these communications in Muscarinic receptors and how different ligands triggers biased signaling allosteric pathways in muscarinic receptors. Our findings identify correlated motion in receptor conformations, identify allosteric sites in statistically robust manner which can further enable targeting novel allosteric sites which are of biological and therapeutic importance in receptor.

3.3 Introduction

G-Protein Coupled Receptors (GPCRs) comprise one of the largest families of the membrane proteins. They are responsible for signal transduction and play important role in carrying out various physiological activities ranging from sensory perception, chemo taxis, to neurological responses [57]. Their involvement in this diverse and important set of responses also makes them the largest class of drug targets [74]. GPCRs can be characterized structurally by the presence of seven transmembrane helices connected by three extracellular and three intracellular loops1. The well-characterized existence of active, inactive, and partially active conformational states highlights the functional importance of dynamics associated with these proteins [26]. Changes in conformation have been observed upon binding of naturally occurring ligands resulting in a signaling cascade through G-protein or Beta arrestin. Research into the function of various GPCRs has resulted in astonishing insight regarding the complex functional properties of this class of receptor. Only recently, however, has it become practical to explore the allosteric mechanism in GPCRs.

Allostery in GPCR is achieved through cooperative binding of ligands on allosteric site at receptors.These binding induces the ligand-mediated shift in conformational [12]. These distant binding event further impacts the function and dynamic on to the distant site of the receptor. Reorganization of non-covalent contacts between residues of the protein is one of the ways for allosteric interaction in proteins [59].

Experiment characterization of allosteric site in proteins are often involves mutational studies to identify the functional change and thermodynamic characterization of protein before and after the mutation. Approaches like Statistical Couple Analysis uses mutated pairs of protein residues from multiple sequence alignment to identify the coupling between protein sites [76]. Allosteric sites prediction using this

method though has been widely use but require large multiple sequence alignment and predicted site in a protein might not be relevant in the alignment. These methods have provide detailed insights towards understanding of dynamic processes and existence of multiple conformation of the receptor as well as there is complex equilibrium between the receptor conformational ensembles.

Molecular Dynamics(MD) approach to study the thermodynamic ensemble of structures and using cross correlation analysis has been successfully adopted in the past. These studies aim to understand the allosteric signal and approximate extent of functional changes driven by ligand induced structure changes provide insight to allosteric interactions. Molecular dynamic simulations studies on β_2-AR identified the loosely coupled allosteric network. Studies on these allosteric network reveal small perturbations near the ligand-binding site leads to significant change in conformational at the intracellular G-protein-binding site [81]. Molecular dynamic studies also provide insights into intermediate states of receptor which are and could be difficult to isolate and experimentally [22] exploration. MD approach coupled with computational studies have explored ligand interactions with receptor. These studies reveal ligand interact with different efficacies with receptor and differential modulation of free-energy landscape of the receptor by ligand [80]. Other Computational approach to identify allosteric interactions and participating residues is analyzing the generated ensemble using principal component analysis to understand the collective motion of allosteric sites to reveal the correlation between role of allosteric sites and how modified confirmation induce functional response in GPCR. [86]. Principal component analysis gives set of eigen vectors that represents collective movement of selected atoms along a principal axes. Principal component analysis may be suited to explain the small-backbone fluctuation but large conformation fluctuation that involves motion of side chain, motion in flexible loop regions couldn't be appropriately explained

by studying principal components.

To study contributions arising from non mutable groups of atoms during the ligand binding events or to analyze protein residues that participate in ligand binding process Molecular MechanicsGeneralized Born Surface Area (MM-GBSA) can be quite helpful. MM/GBSA allows the decomposition of the electrostatic solvation thus facilitate in detail study of protein ligand interaction at the residue level. This can further facilitate studying of total interaction energies at residue levels in protein and will help in understanding the contribution of residues towards allostery in GPCRs [19].

Our Study also uses mutual information as metric to quantify correlated motions. Mutual information is quantified by analyzing both backbone and sidechain torsions angle of residues. In an equilibrium ensemble mutual information approach aims to identify residues pair having correlated conformations. Configurational entropy metric is further used to identify mechanisms behind cooperative allosteric binding of ligand and modulator in muscarinic receptor.

Mutinf package uses mutual information approach to identify allosteric networks [45]. This approach helps in quantifying correlation between the conformations of residues located at different sites of protein. Mutinf which is entropy based approach pioneered by Jacobson lab helps to analyze thermodynamic ensembles of protein conformers generated using MD simulations. Mutinf employs analyzing internal coordinated focusing on dihedral angle to identify anharmonic correlated motion in sidechain rotamers. Dihedral angle is mostly responsible for low frequency motion and thus help to identify the unique pair residue pairs with correlated structural conformation in an ensemble at equilibria. Recent advances in crystallographic methods with use of T4L receptor complexes, monoclonal fused receptors, and GPCR targeting nano bodies [57], have resulted in atomically detailed GPCR structures from various

families. Our understanding regarding the ability of ligands to mediate variable responses from GPCRs has been greatly augmented through experimental structural studies on beta-2 adrenergic receptor [59], bovine rhodopsin [73] , and alpha-2 adrenergic receptors [20]. Studies on bovine rhodopsin have spectroscopically characterized multiple distinct conformational and functional intermediates [20]. Beta-2 adrenergic and alpha-2 adrenergic receptors have experimentally identified mutants that still retain activity independent of ligand binding. The conformational selectivity of different mutants highlights the importance of dynamics and multiple conformational sub states of GPCRs. Their importance extends to medicinal applications as the degree of conformational changes in GPCRs has been linked to the efficacy of ligands [24].

Recent crystal structures of muscarinic family receptors, namely M2, in bound and unbound states, provides the basis for studying the role of the modulator in allosteric signaling [39]. Muscarinic receptors belong to the acetylcholine class of receptor as they regulate the release of acetylcholine from prostaglandins receptors [39, 17].For M2 receptors one of the challenge is to identify the residues participating in the allosteric interaction and identify its contribution in achieving allosteric signal propagation from extra cellular ligand binding site to the intra cellular G-protein coupling site. The M2 and M3 receptor structures will help identify allosteric signatures through a systematic analysis of dynamic properties such as thermodynamic interaction energies, inter residues contacts, and correlated motion of residues in two closely related mammalian receptors. A comparative analysis of muscarinic receptors with known activity to adrenergic receptors will also give a broader understanding of class A GPCRs. These approaches will promise better therapeutics by providing the basis for identifying the drug candidates that target the specific GPCR conformational state associated with the disease state.

Despite the evidence of multiple conformational states and their role in ligand

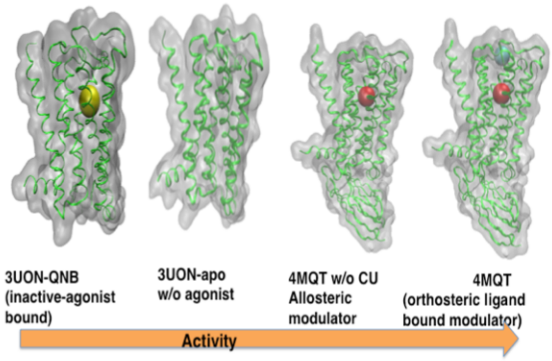


Figure 3.1: Structural information about muscarinic receptors with various ligands in bounded states is available ,which is further used in our study.We have used M-2 muscarinic receptor in apo form ,agonist bounded and orthosteric ligand bounded with allosteric modulator (PDB id: 4MQT)

efficacy, the evolution of allosteric bias and how that bias propagates in the system is still unclear. Past studies on identifying the characteristic features of ligand binding that are common in a family of GPCRs has been very informative, but limited to static crystal structures. In this present study we have applied MD to simulate the M2 receptor and studied changes to allosteric in presence of allosteric ligand and modulator and in absence of allosteric ligand and modulator. Cross correlation analysis was used to observe the cross correlation residue motion and information flow in the receptor. Our study also quantified correlation between allosteric site in thermodynamic ensemble in GPCRs and identified clusters of correlated residues that are crucial for allostery in GPCRs by studying changes in configuration entropy between the side chain and torsion angles of residue pair. Our study reveals evolution of allostery in GPCRs and presence of modulator or ligand differential set of residues trigger information flow in GPCRs.

3.4 Materials and Methods

3.4.1 System Setup.

X-ray crystal structure of M2 bounded to iperoxo and allosteric modulator LY2119620 that was solved at 3.7 resolution (PDB: 4MQT) was used for carrying out simulations. The "G-protein mimetic" nano-body Nb9-8 that was fused into the receptor to replace intracellular loop 3 (ICL3) to assist crystallizing was omitted from all simulations. Previous findings suggest that removal of the bulk of ICL3 does not appear to affect GPCR function. Chain termini of receptor were capped with neutral groups. Disulphide bonds (Cys963.25-Cys176ECL2 and Cys4136.61-Cys4167.29) initially resolved in the crystal structure of were maintained in the carrying out the simulations.

The muscarinic receptor was inserted into a palmitoyl-oleoyl-phosphatidylcholine (POPC) bilayer. Lipid molecules which were overlapping were removed using the Membrane plugin in VMD. The charges in the system were neutralized at 0.15 M KCl using the Solvate plugin in VMD. Initial Box size $95\mathring{A}$ X 95 $\mathring{A}$ x 105 $\mathring{A}$ .Simulation was performed using CHARMM36 force field parameters.73 K+ ,83 Cl-, 230 lipid molecule.Periodic boundary conditions were applied to the system during the simulations. Modulator LY2119620 was removed from the starting structure for the second system without modulator. QNB ligand was removed from X-ray crystal structure of QNB-M2 complex (PDB: 3UON) that was solved at $3\mathring{A}$ resolution to obtain apo state of muscarinic receptor for simulation.

3.4.2 Simulation Details.

Molecular dynamics (MD) simulations were performed using NAMD2.9(12). The CHARMM36 parameter set with CMAP terms was used for the protein,POPC. TIP3P model was used for the water molecules. Force field parameters of iperoxo,LYS2119620 that was obtained using ParamChem web server. 12 distance cutoff was set for the van der Waals. particle-mesh Ewald summation method using a grid point density of $1\mathring{A}$ was used for calculating long-range electrostatic interactions. A 2fs integration time-step was used for all MD simulations. Multiple-time-stepping method was used and bonded and short-range non-bonded interactions was computed at every time-step while long-range electrostatic interactions was calculated at every two time-steps [54]. The SHAKE algorithm was applied to all hydrogen-containing bonds. Molecular dynamic simulations was started with equilibration of the system in NAMD. Equilibration was performed first. First, all atom fixed the lipid tail was energy minimized for 2000 steps for 1.0 ns. Melting was NVT run at 310K. Further equilibration was carried using NPT run at 310K with 10 kcal/mol.2 on Backbone and 5 kcal/mol.2 on rest of the system for 10ns. Restraints on backbone and rest of the system was further reduced for equilibration in following ways (i) 5 kcal/mol.2 and 2.5 kcal/mol.2 respectively for 10 ns (ii) 2.5 kcal/mol.2 and 1 kcal/mol.2 respectively for 10ns (iii) 1.0 kcal/mol.2 and 0.5 kcal/mol.2 respectively for 10ns (iv) 0.5 kcal/mol.2 and 0.1 kcal/mol.2 respectively for 10ns (v) 0.01 kcal/mol.2 on for 10ns. After above minimization and equilibration procedures, equilibrated system was prepared for Anton run. The charmm36 forcefield was used for protein and lipid and added equilibrated system using VIPARR. The production run were performed on all three systems using Multigrator on ANTON super computer for 4 μs at 1 atm and 310 K.

3.4.3 Correlation Analysis.

Correlation between residues in all three system was analyzed for 4 μs MD trajectory by calculating the normalized covariance:

$$C_{ij} = \frac{Cov_{ij}}{((< \Delta r_i\,(t)^2 >) \cdot < \Delta r_j\,(t)^2 >))^{1/2}} \tag{3.1}$$

where

$$Cov_{ij} = (< \Delta r_i\,(t)^2 > \cdot < \Delta r_j\,(t)^2 >) \tag{3.2}$$

and

$$\Delta r_i(t) =< \overrightarrow{ri(t)} - \overrightarrow{ri(t)} > . \tag{3.3}$$

where ri(t) is the position vector of C_α of ith residue and the time average of the quantity is in the <.> brackets.

The correlation matrix is generated that characterize the motion of protein residues.The correlation ranges between -1 to 1. If the residue move in same direction the motion positive correlated and if they move in opposite direction the motion is anti correlated.

3.4.4 MM/GBSA

MM/GBSA method was used to calculate the binding affinity between (i) receptor and ligand (ii) receptor and ligand-modular complex. MM/GBSA was estimated using CHARMM. Following equation weere used to estimate binding free energy.

$$\Delta\Delta G_{binding} = \Delta G_{complex} - \Delta G_{receptor} - \Delta G_{ligand+modulator} \tag{3.4}$$

$$\Delta G_{MM/GBSA} =< \Delta E_{VDW} > + < \Delta E_{ELEC} > + < \Delta E_{INT} > + < G_{solv}^{polar} > + < G_{SA}^{non-polar} > -T\Delta S \quad (3.5)$$

where ΔE_{VDW} is energy term of sum of Van-der Waals energy, ΔE_{ELEC} id sum of Coulombic energy , ΔE_{INT} reperesnts other energy terms such as bond energy and angle energy, where no nonbond cutoff was applied to these energy calculations. G_{solv}^{polar} is solvation energy term calculated through GBSW module in CHARMM. $G_{solv}^{nonpolar}$ is calculated by evaluating the solvent accessible surface area (SASA) with a radius probe of 1.4. ensemble average of the quantity is in the <.> brackets

Energy term was calculated by the following equation:

$$\Delta G_{SA}^{non-polar} = \gamma SASA + b \quad (3.6)$$

where γ is the surface tension term and value $0.00542 kcal mol^{-1} \mathring{A}^{-2}$ and constant b the value of 0.92 kcal/mol was used The snapshots taken from trajectory at 1 ns and final free energy was estimated by averaging the values of energy terms of ensemble.Based on previous studies we assumed that change in entropy upon protein-ligand binding is significantly small and henceforth was neglected.

3.4.5 Mutual Information Analysis.

Mutual information analysis metric is used as metric to evaluate of long-range coupling between residues.The evaluation is done by calculating the correlation between the motions of GPCR protein backbone and side-chain torsion angles.These calculations are further used to identify allosteric couplings in muscarinic recep-

tors.Mutual Information(MI) between two residue is given by:

$$MI(x,y) = \sum_{x \epsilon x} \sum_{y \epsilon y} p(x,y) log(\frac{p(x,y)}{p(x)p(y)}) \tag{3.7}$$

p(x,y) represent the joint probability distribution of residues X and Y. p(x) is the marginal probability distribution function of residue X and p(y) is the marginal probability distribution function of residue Y. In here X and Y represent one rotameric state of residue X and Y respectively.

We quantified correlation between residues torsion using McClendon et.al. mutinf package. Mutual information approach connects between information theory and thermodynamics and calculates configurational entropies from conformational ensembles. To calculate configurational entropy, internal coordinates of the system is used. Second-order terms is derived from entropy expansion (mutual information) and is further used to determine residue pairs with correlated conformations. The method uses molecular configurational entropy over torsion angle changes in residues. The mutual information represents correlations between degrees of freedom. Mutual Information quantifies amount of information gained by one degree of freedom by knowing another. Summation of torsion entropy can be calculated by

$$\begin{aligned} S_{conf} = & \sum_{i}^{n} \int_{0}^{2\pi} p(\phi) ln p(\phi) d\phi \\ & - \sum_{i}^{n} \sum_{j}^{n} \int_{0}^{2\pi} \int_{0}^{2\pi} p(\phi_1, \phi_2) ln \frac{p(\phi_1, \phi_2)}{p(\phi_1)p(\phi_2)} d\phi_1 d\phi_2 + ... - ... \end{aligned} \tag{3.8}$$

where indices i and j represents torsions of residue, n is the number of torsions for residue. We used $(\phi, \psi, \chi_1, \chi_2)$. The second-order term is sum of the mutual information of each pair of torsions. The mutual information explains correlated relationship between degrees of freedom. By help of mutual information we aim

to quantify the gain of information about one degree of by knowing the another. Mutual information values in here are in terms of entropy, which is eventually related to free energy hence the the mutual information in here Equation 4.9 is in units of kT. [45] Mutual information is a distribution-free analysis method and robust statistical measure that doesn't rely on assumption about the sampling data from a given probability distribution, enables us to understand the changes in conformation.

3.4.6 Allosteric Network based on Mutual Information

Mutual information between the torsion angle of residue was used to study the allosteric contribution of each residue. We constructed the allosteric network based on the pair wise correlation between the residues. Cross correlation of torsion angle in this approach is defined in kT unit. To study major contribution and reduce noise, our network construction limit the correlation greater than kT. Weighted edge between two residue is defined in terms of kT and degree of a node is number of edges passing through each node. In an effort to understand the protein ligand interaction and to identify which functional site can alter the conformation or dynamics of receptor by perturbation hierarchical clustering was used to identify the group of residues which are critical for correlation between the sites important for receptor function. Hierarchical clustering of constructed allosteric network was performed using cytoscape. Euclidean distance metric was used to cluster residues with similar patterns of correlations with other residues

3.5 Results and Discussion

3.5.1 Correlation Motion from MD simulation Identifies the communication between Ligand Binding site and G-Protein Coupling site

The allosteric signal transmission within the muscarinic receptor complex is proposed to be coupled motion between residues in active site residues and parts of receptor that interacts with ligand. Beside local coupling long distant correlated coupling motion is been identified in the distant part of receptor complex with the ligand and allosteric modulator.

Correlation of residue motion were identified between G-protein coupling site and ligand binding site in micro second trajectories of receptor. We used generalized cross-correlation analysis and compared the apo-receptor with agonist bound receptor with modulator and with agonist bound receptor without modulator.Figure 3.2 shows the dynamic cross correlation map of MD simulation for three receptor. In apo receptor the receptor is in unbounded form and exhibit significantly higher correlation compared to modulator bound and agonist bound receptor. Higher correlated residue motion is observed in the TM5,TM6 and TM7 region.

Previous studies highlight the functional importance of TM3 domain and we too identified the TM3 as functional hub.Correlation of TM3 and intracellular TM5 confirms the communication and its role in receptor activation. Similarly the extra-cellular region of TM3 domain and TM4 domain is correlated to ECL2 region. By examining thermodynamic ensemble in the bounded and ligand free form in muscarinic receptors, we were able to identify the key allosteric sites and the allosteric

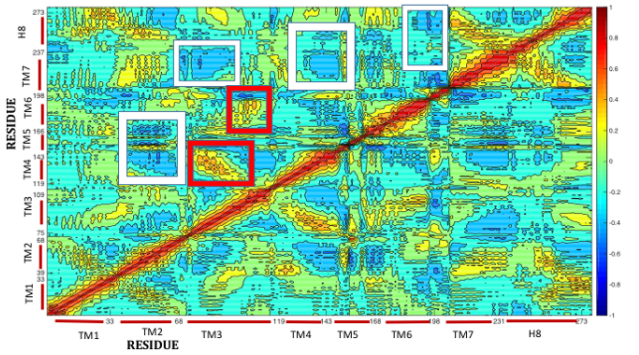
bias in the muscarinic receptors. In this current study of M2 receptor we identified allosteric sites to quantify correlation and determine long range allosteric signals in thermodynamic ensemble. These result will help us to detail the understanding of long range communication. The ligand bound to receptor increase the allosteric response and in turn promotes the G-alpha binding to the orthosteric binding site. Allosteric modulator (LY2119620) not only enhances the affinity of muscarinic receptor for iperoxo and can also directly activate the muscarinic receptor.

Residue motions in G-protein-coupling sites and ligand-binding is analyzed using cross-correlation of C_α residues. This analysis facilitate in understanding long distance signaling between the G-protein coupling site and ligand binding site. Role of TM3 has functional hub towards the activity of receptor has been previous established. Our results indicate the TM3 also plays critical role in allostery. Our studies explores the dynamics and long range signaling in three different states (i) apo form (ii) bounded with a ligand (iii) bounded with ligand and allosteric modulator. The significant difference in the cross correlation of C_α atom of residues indicates the role of ligand and modulator towards the long range signaling in receptor.

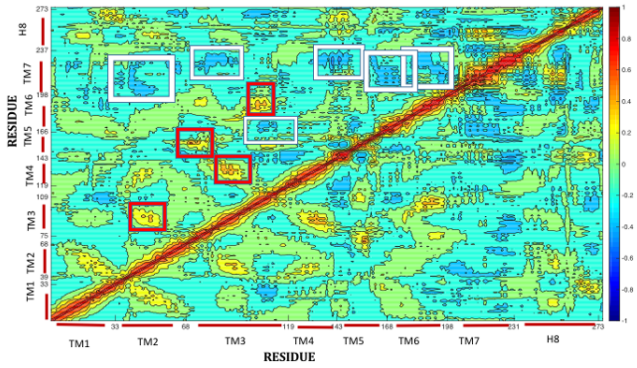
Cross correlation analysis reveal the role of intracellular residues of TM6 towards the allostery. In agonist bounded complex the residues of TM6 are correlated to residues of TM5, TM6 and TM7 near the binding pocket(Fig. 3.2b). The change in correlation can be attributed to the rotation of Trp6.48 leading to cleavage of the Asp3.32 - Tyr7.43, which leads to allosteric signal. Change in the strength of correlation between the TM6 and binding pocket marks the difference in allosteric signaling in receptor in ligand-modulator bounded receptor complex and ligand bounded receptor (Fig. 3.2b,Fig. 3.2c). Our correlation analysis suggest the role of TM1 ,TM5 ,TM7 critical for the allostery. Studying Correlation of N terminal end of TM5 with the TM1 can highlight the signs of allosteric interaction in the receptor. Break-

ing of $\pi - \pi$ clustering between TM1 and TM5 which is maintained in the inactive state is highlight of the allosteric interaction pipeline.This clustering is facilitated by $Phe^{5.8}, His^{1.12}, Phe^{2.6}$ and $Phe^{3.3}$.Our studies indicate the change in cross correlation of these residues across all three system as a significant allosteric signature interaction to identify receptor activity.

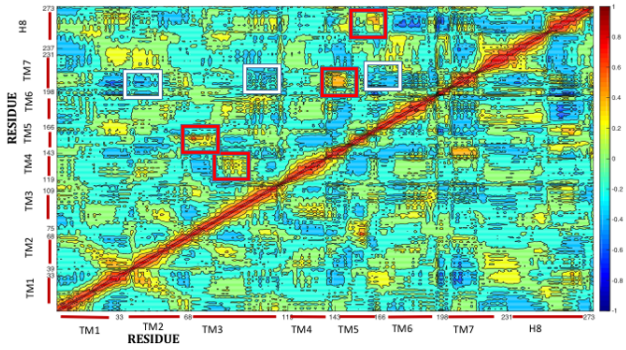
Correlation analysis suggests muscarinic receptor has long range signaling which is triggered by the change in correlation at allosteric binding site leads to change in interaction at G-protein binding site. The change in the correlation between TM1 and TM5 residues serves as key step in identifying allostery. Past studies indicated the disruption of the contact between TM1 and TM5 can be associated with activation of receptor leading to release of GDP. Studying interaction of these residues which associate G-protein binding site with allosteric binding site can help in understanding allosteric pathway. These residue that consistently re-organized upon receptor binding relate the structural changes in receptor with the allostery. However these residues from TM3, TM5, TM7 is critical for allostery the role of non conserved residues can't be ignored in allosteric interaction. Our correlation study identifies the residue critical for allosteric interaction but this basic frame work can be further improved by understanding the ligand binding interface and allosteric pathway can also vary by different ligand.



(a) muscarinic apo receptor



(b) muscarinic receptor bound to allosteric ligand



(c) muscarinic receptor bound to allosteric ligand and modulator

Figure 3.2: Correlation motion between receptor residues are identified in extracellular ligand binding domain and intracellular G-protein coupling site.

Color coded dynamic map of GPCR residues for (A) unbounded apo receptor (B) ligand bound receptor without allosteric modulator and (C) ligand and modulator bound muscarinic receptor. Anti correlation is displayed by white boxes and strong correlation is displayed by red boxes. Receptor residues of helices TM1 to TM7 are highlighted by red bars on the left and below

3.5.2 Mapping structural functional relationship reveal role of molecular switch towards Allostery.

One of the key changes towards GPCR activation is an outward movement of the intracellular portion of transmembrane (TM) helices 5 and 6. These outward movements creates cavity that are large enough to accommodate the carboxyl end of the G$-\alpha$subunit . Previous studies on agonist-bound receptor crystal structure have revealed agonist bind can also induce these outward movement and facilitate binding of G-protein alpha subunit. These allosteric binding induces structural changes in the receptor are often seen in the several active crystal structures GPCRs which have been crystallized in complex with agonists. The outward displacement at the intracellular side of TM6 and a rearrangement of TM7 around the NPxxY motif is often used as characteristic to study receptor activation. Inward movement of the NPxxY motif in the inner region of TM7 is examined to identify conformational changes in the receptor. This will further able to demonstrate the allosteric bias in the Muscarinic receptor.

To understand how allosteric behaviors has changed after the ligand (QNB) was removed from the 3uon:Inactive receptor and how the allosteric activity differ in the absence of modulator we plotted RMSD of NPxxY motif in muscarinic receptor v/s TM3-TM6 distance. Change in the structural motif and TM3-TM6 distance can be related to the activity change. Along with our results we combined the pre published Anton trajectory on QNB bound system. The TM3-TM6 distance is calculated between the C_α atoms of Arg121 and Thr386 in the cytoplasmic end. Comparing conformational sampling of 3uon and currently simulated 3uon-apo indicates the larger change in the TM3-TM6 distance v/s RMSD (Fig. 3.3A). This preliminary data indicates the presence of intermediate state in 3uon-apo and more conforma-

tional sampling might result in conformation corresponding to active structure. No significant structural changes in TM3-TM6 distance corresponding to NPxY motif was identified between MQT and MQTw/o allosteric modulators Fig. 3.3B Fig. 3.3C. This suggest there exist activity in 4MQT receptor even after removal of allosteric modulator indicating presence of allosteric bias in the receptor .

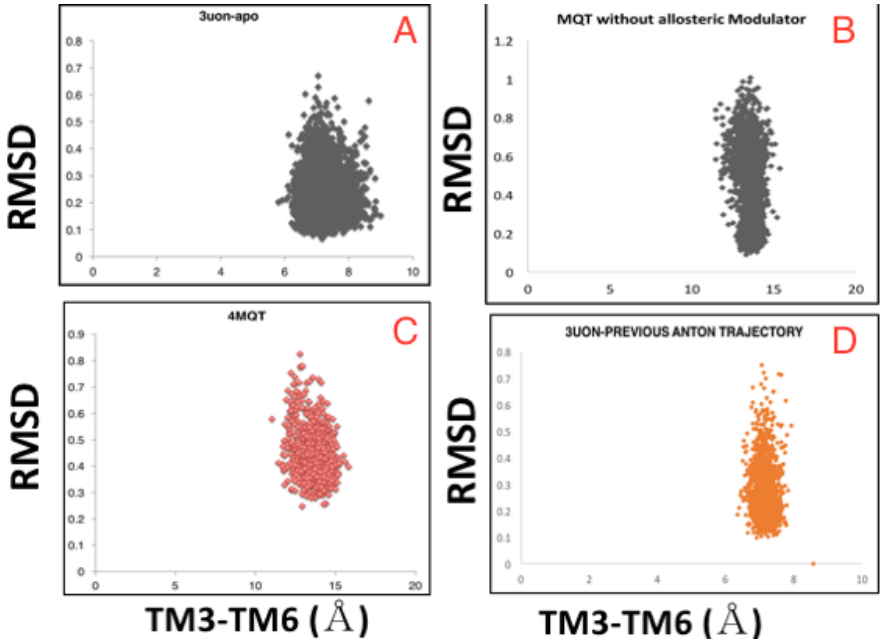


Figure 3.3: Mapping Structure function relation in muscarinic receptors

Plots for the TM3-TM6 distance v/s RMSD of the NPxxY motif (A) 4μs of unbounded muscarinic receptor (B) 4μs of bounded muscarinic receptor without allosteric modulator (C)4μs of bounded muscarinic receptor with modulator and ligand (D) previously published 16.4μs of QNB-bound M2 receptor

3.5.3 MM/GBSA interaction energy analysis reveals key residues responsible for differential Allosteric modulation in receptor.

To identify the key residues responsible for the binding process of allosteric ligands, the per residue contribution of the receptor to the binding free energy between ligand and receptor was computed using MM-GBSA decomposition process. The per residue binding free energy interaction information in muscarinic complex with and without modulator is shown in Fig. 3.5 and the contributions of key residues for two complexes is highlighted. The residues at positions 3.30, 3.37, 6.51 and 7.42 forms the contact with ligand and are near the binding pocket. Residues 5.47, 6.48, 7.42 has shown to significantly impact as key interaction towards allostery Fig. 3.5. Binding of allosteric ligands is accompanied by the outward movement of the TM3 and TM6 helices. These binding lead to rearrangement of hydrophobic cluster and aromatic residue primarily made of the 3.30, 3.37, 6.51 and 7.42, 3.40, 5.51, 6.44 and 6.48. Similarly, binding of modulator to the receptor-ligand complex lead to the minor conformation changes due to breaking of the ionic lock which further is responsible for biased signaling.

Past studies with β-1 adrenergic receptor bounded to arrestin-biased ligands and G-protein based ligand also shows that modulator have contact with minor biding pocket (formed by TM7-ECL2 residues) [49]. These residue in minor binding pocket is also possibly involved in arrestin-biased signaling. These modulator bound receptor complex structures do not show large conformational changes in the cytoplasmic side of the receptor but additional modulator driven interaction with TM7 is shown to drive the biased allostery via conserved activation pathway mainly through TM2 or

TM7. Modulator binding alters the local environment by inducing chemically shift in Cys residue at the TM7H8 interface. These binding of biased ligand is further shown to stabilize conformation of TM6 and TM7/H8 that is distinct from agonist bound receptor. [83]

Along these lines past studies reveals these residues role in receptor activation.These residue interactions and conformational changes is critical towards GPCR biogenesis [81]. These changes of the paired interactions and its role in local conformation changes as in allostery is further studied using MM/GBSA per residue interaction method as it helps in detailing the interaction of the ligand with the receptor. The binding pocket of the receptor interacts with the ligand directly . These interaction further promotes the structural changes as seen in the Fig. 3.4 to the structural G-protein binding site. The binding of ligand is most likely,not only stabilize the specific conformation of the receptor but also modulate the activity of receptor. In the case of muscarinic receptor we have seen the presence of the modulator in the binding cavity is stabilized and have distinct energy distribution. The residue near the binding pocket have distinctly contributed in ligand binding. In particularly residues at position 3.30, 3.36, 3.37, 6.48, 6.51 and 7.38 make contacts with ligand and stabilizes the corresponding state of muscarinic receptor.(Fig. 3.5)

Apart from residues in TM, the resides in ECL-2 plays a critical role in allosteric activity of the protein.In absence of the modulator there exist shift in allosteric activity in terms of energy distribution and showed that the some residues in some sites displayed different energy between modulator bounded and unbounded state. Difference between the two state reveal the allosteric sites and residues contributing to the functional activity of muscarinic receptor. Our study revolves around the assumption that the interaction of ligand with residues in the binding pocket lead to changes in the interaction energy of residues not near binding pocket but also lead to

the changes in interaction energy distant to ligand binding site. Mapping these interactions in the modulator-ligand bounded receptor and in ligand bounded receptor is key feature to identify allosteric bias and allosteric signature in the receptor.

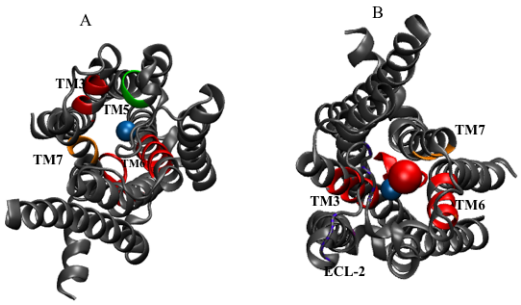


Figure 3.4: MM/GBSA interaction energy contributed by individual residues on ligand bounded MQT (A) and (B) ligand and modulator bounded ligand, in kcal/mol. The residues that significantly participate in interaction are highlighted previously and participating helix are colored here

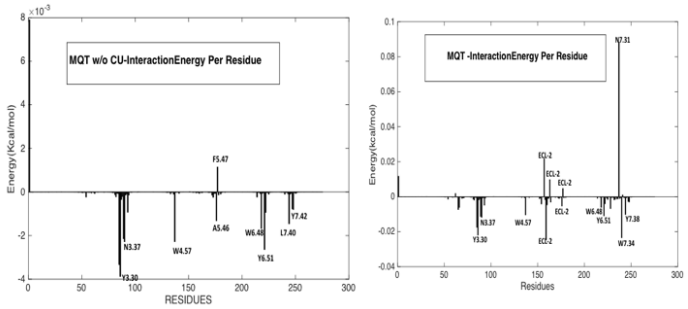


Figure 3.5: Mapping per residue contribution in MM/GBSA Structure function relation in muscarinic receptors

Comparison of per-residue energy decomposition for key residues for two systems: (A) 4μs of ligand bounded muscarinic receptor without allosteric modulator (B)4μs of ligand bounded muscarinic receptor with modulator

3.5.4 Mutual Information from MD mapped significant Long-Range communication

Thermodynamic ensemble of receptor in ligand-modulator bounded, ligand bounded and in apo state was used to analyze correlation between the residues of receptor. From MD simulations for each pair of residues, mutual information was calculated between all pairs of ϕ, ψ, χ torsion angles of receptors. The mutual information is depicted in units of kT, due to relationship of mutual information with entropy which is subsequently connected to binding free energy. The receptor exhibit significant difference in correlation between it apo receptor form, ligand bounded-recpetor complex and ligand-modulator complexed receptor.In contrast to bounded ligand apo receptor indicates increased dynamics in TM3,TM5 and TM7 region.

The Fig. 3.6, Fig. 3.7, Fig. 3.8 indicates the plot of mutual information between torsions of residue pairs in muscarinic receptor. In contrast to apo-receptor, in ligand bounded receptor complex we found a small subset of highly correlated residue pairs and many residues with marginal correlation. The plots indicate the residues which are in the binding pocket are significantly correlated. Comparing Fig. 3.6, Fig. 3.7, Fig. 3.8 shows that that apo-receptor have lower MI than ligand bounded state and in the modulator-ligand bound receptor, and which shows that residue movements are more uncorrelated in apo form, leading to a more possible conformation state of the muscarinic receptor. This finding is in agreement with recent studies of M2 receptor which that show that the apo receptor is more dynamic than active state of Muscarinic receptor[46].

In apo receptor(Fig. 3.6) the residues in the TM3, TM4 region are significantly correlated with the TM7 region. In the modulator bound receptor complex(Fig. 3.8) residues in ECL2 specifically ARG169 forms significantly correlation

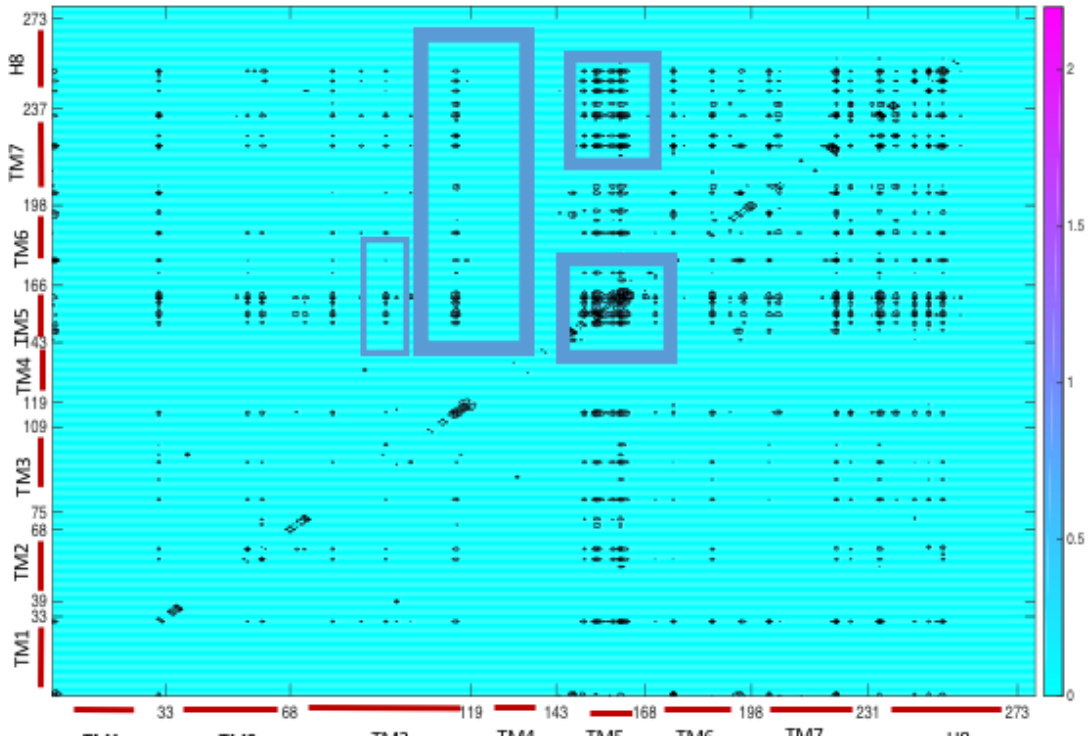


Figure 3.6: Mutual Information Matrix for apo muscarinic receptor

Figure shows Mutual information between residues' torsion computed using the mutinf as detailed in Methods in muscarinic apo receptor. Correlated residues in simulation are surrounded by blue box depicted significant correlation in all three systems.

with TM1(ILE38P), TM5(ARG211$^{5.63}$), TM7(ASN436$^{7.49}$), H8(LEU455$^{8.58}$). The ECL2 loop due to it flexible nature and ability to adopt at multiple conformations is revealed to be critical for the allosteric activity of muscarinic receptor.

Previous biochemical studies have highlighted mutation of above positions in the ligand binding pocket not only effect the ligand-binding affinity, but also plays important role towards conformation selectivity of the receptor. In modulator bound complex, the intracellular region of complex is strongly correlated to extracellular sides (Fig. 3.8). One of the reason of these strong correlation can be due the modulator

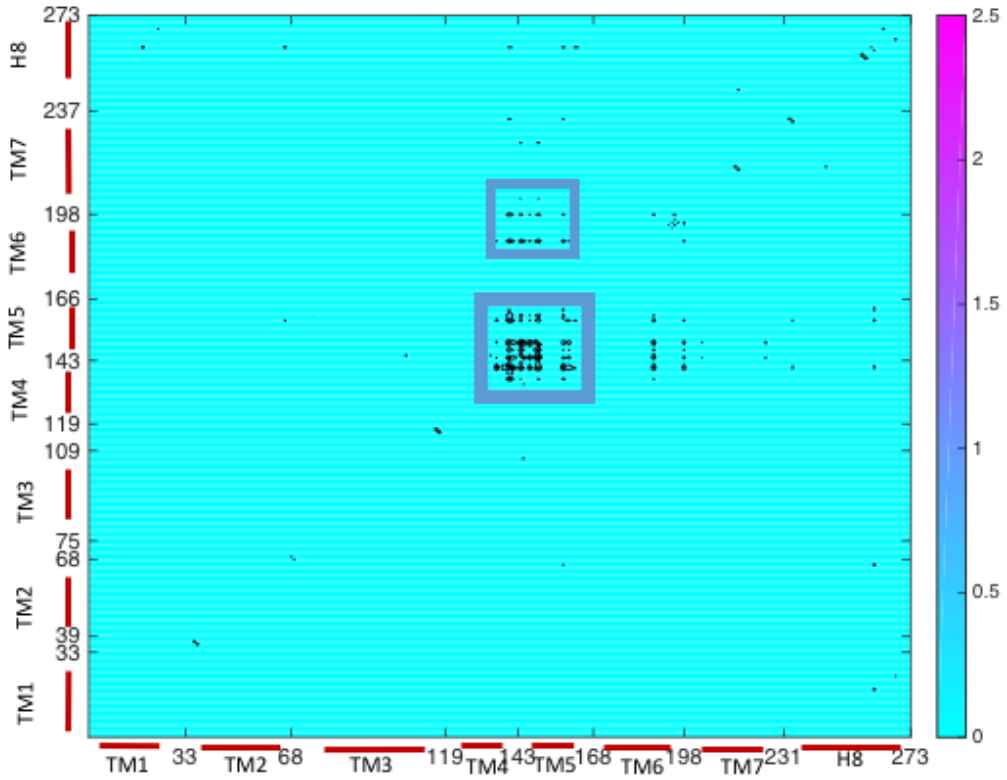


Figure 3.7: Mutual Information Matrix for muscarinic receptor bound to ligand

Figure shows Mutual information between residues' torsion computed using the mutinf as detailed in Methods in muscarinic receptor bound to ligand. Correlated residues in simulation are surrounded by blue box depicted significant correlation in all three systems.

binding in the extracellular side of ECL2 further facilitates association of the G protein in the intracellular regions and that lead to stabilization of the receptor in active conformation. In the muscarinic receptor role of ECL2 has been establishes as the allosteric modulator binding site and in biogenic amine receptor the ECL loops also form ligand entry and exit domain.

Past studies on microseconds time scale of MD simulations with muscarinic receptor bound to agonist reveals role of ECL2 (PHE181,

TYR177, ARG169) towards allostery [37]. TRY177, ARG169 is key contact for mod-

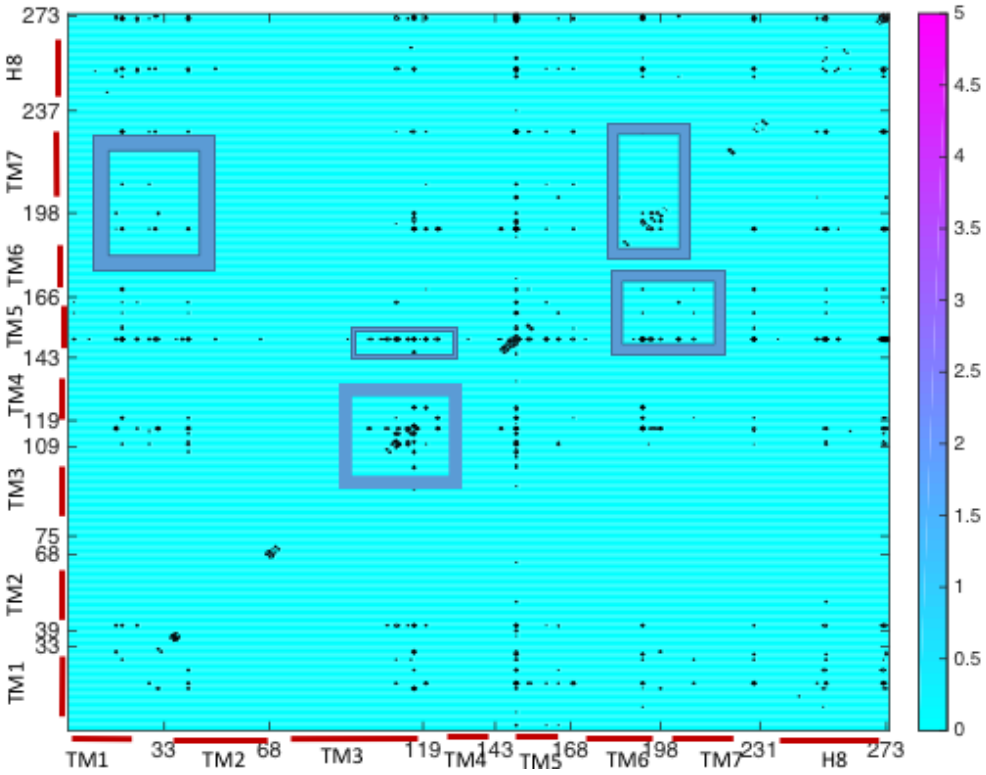


Figure 3.8: Mutual Information Matrix for muscarinic receptor bound to ligand and modulator

Figure shows Mutual information between residues' torsion computed using the mutinf as detailed in Methods in muscarinic receptor bound to ligand and modulator. Correlated residues in simulation are surrounded by blue box depicted significant correlation in all three systems.

ulating agonist binding in muscarinic receptor [25] and in beta-2 adrenergic receptor and by breaking of the salt bridge between ECL2 residues and TM7 residues, LYS305$^{7.32}$ is too found critical towards allostery of GPCRs.

Also, our findings in muscarinic receptor bound to allosteric modulator (Fig. 3.8) suggests role of ASN436$^{7.49}$, H8(LEU455$^{8.58}$ and TM5(ARG211$^{5.63}$) towards long range allosteric communication as one of the critical interactions for the receptor activation. This is consistent with previous structural studies which suggest the inter-

action of ASN436$^{7.49}$ of the NPxxY motif in TM7, the aromatic-aromatic interaction between Tyr440$^{7.53}$ of the NPxxY motif in TM 7 with H8(LEU455$^{8.58}$) critical for allosteric regulations.

Of the mutual Information between pair of residues in muscarinic receptor our analysis identified a very small subset of residues is significantly correlated. Using torsion angles facilitate in examining structural changes and comparing the distribution of the side chain angle in apo-receptor, ligand bounded and modulator-ligand bounded receptor. Mutual information between the residue pairs help to quantify the allosteric changes and contribution of residues toward allostery. Allosteric network based on the mutual information between the pair of residues indicates long-range signaling.

Figure 3.9 depict allosteric network of three systems. The residues in three systems with degree of connection greater than 4 are highlighted in our study. Based on mutual information correlation between the residues in extracellular loop regions is significantly high. Increase information flow among the residues of extracellular loop region indicates great variability of apo receptor crystal structure. Since loops are partially disordered it indicates it can adopt multiple conformation state in apo for than in ligand bounded form. The modulator-ligand bound receptor network highlights the set of residue that forms conserved network from residues of TM4, TM6, TM7 helix and highlights their role in allostery.

In modulator bounded complex the allosteric network(Fig. 3.9C) primarily consist of ILE435$^{7.48}$, TYR440$^{7.53}$, ASN436$^{7.49}$, ARG211$^{5.63}$, LEU455$^{8.58}$, Val133$^{\mathrm{ICL2}}$, MET139$^{4.41}$, MET456$^{8.59}$, MET112$^{3.41}$, MET143$^{4.45}$, ARG169$^{\mathrm{ECL2}}$, ASN183$^{\mathrm{ECL2}}$, GLN179$^{\mathrm{ECL2}}$, LEU43$^{1.52}$, LYS49$^{1.58}$. In particular in modulator receptor complex the residues towards extracellular end of TM3 forms conserved disulphide linkage with extracellular loop(ECL2). These interaction of ECL2 residues lead of

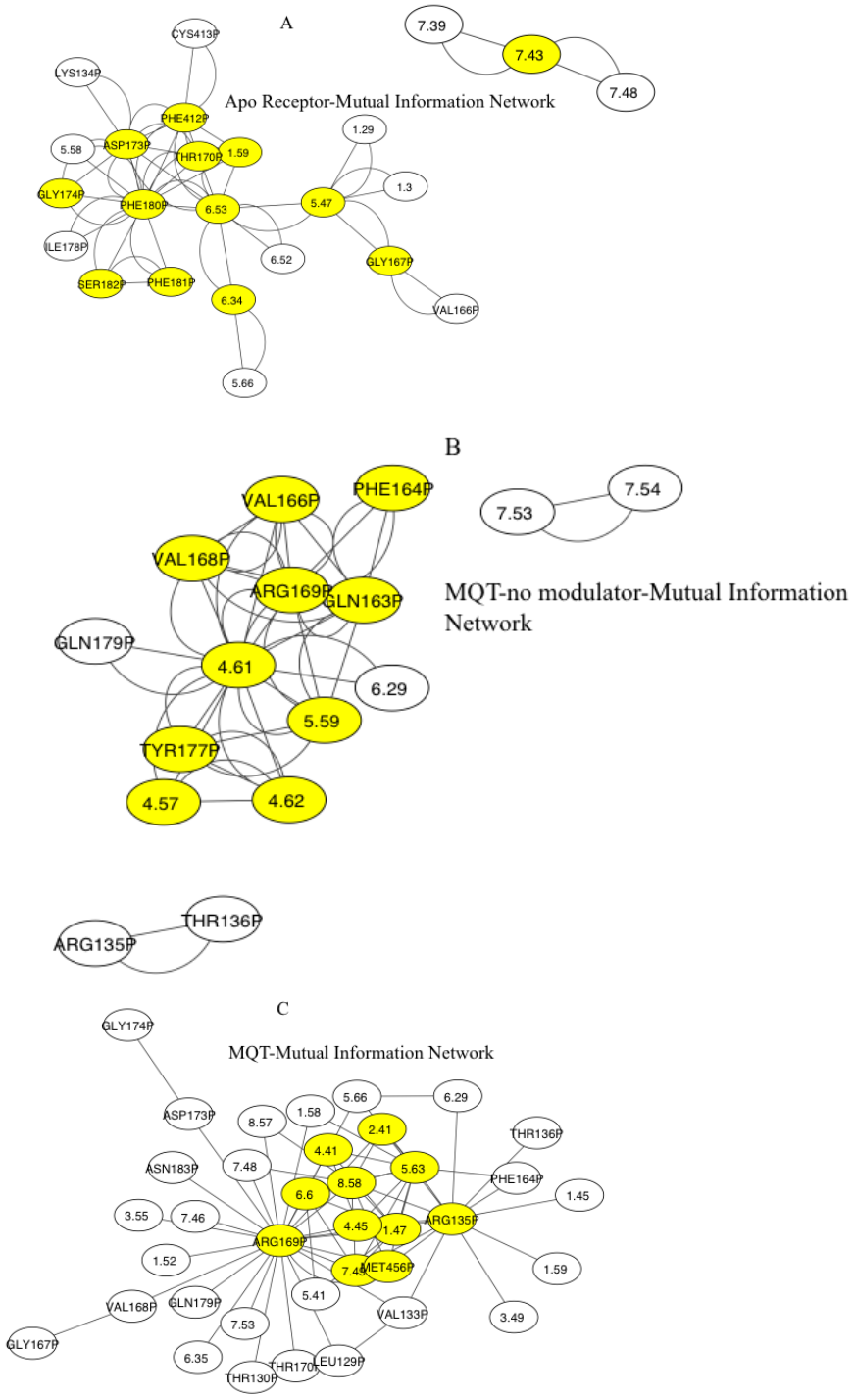


Figure 3.9: Mutual Information based network in Muscarinic Receptor

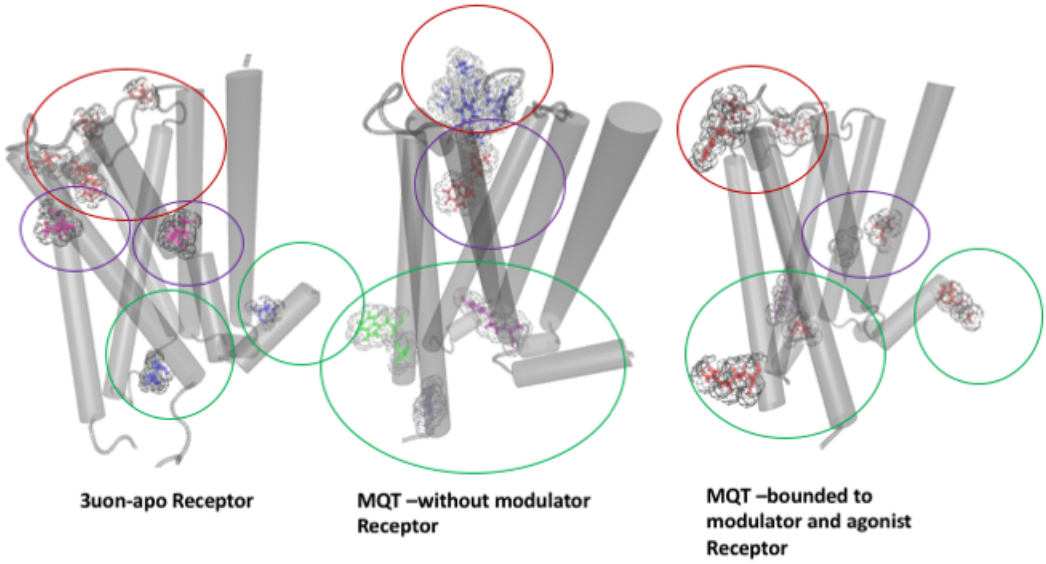


Figure 3.10: Functional region based on Mutual information network in muscarinic receptor reveal differential region responsible for the allosteric modulation in muscarinic receptor.

Red circle represent Extracellular region magneta circle represents middle region of the receptor and green circle represent intracellular region.

relocation of conserved residues near the G-protein coupling site primarily in TM7 ILE435$^{7.48}$, TYR440$^{7.53}$, ASN436$^{7.49}$) and TM8.

Compared to the modulator-agonist bound complex weaker communication is found in the allosteric bound complex(Fig. 3.9B). Weak interaction in receptor agonist complex network primary consist of ARG135$^{\mathrm{ICL2}}$, THR136$^{\mathrm{ICL2}}$, PHE164$^{\mathrm{ECL2}}$, TYR177$^{\mathrm{ECL2}}$, ILE149$^{4.61}$, ARG381$^{6.29}$, GLN179$^{\mathrm{ECL2}}$, VAl166$^{\mathrm{ECL2}}$, TYR440$^{7.53}$, ALA441$^{7.54}$. Mutual Information Network reveals that allosteric signaling has a differential pathway and is based on the type of agonist bound to the allosteric binding site. Analysis of apo receptor(Fig. 3.9A) reveals prominent role of extracellular loop region for allosteric interaction and information flow. Significant correlations between residues in the extracellular loop region indicates the great variability in the crystal structures of

the muscarinic receptor. Loops are partially disordered and flexible in most structures which further indicates it can adopt at least several conformation and hence can play role in allosteric interaction in the receptor.

Previous biophysical experiments on GPCRS suggest upon agonist binding lead to receptor stabilization. Agonist binding further facilitate receptor activation and binding of G-protein coupled receptor [38, 87, 23] and allostery in GPCRs. Our mutual network based analysis validate these finding and suggest allosteric binding not only lead to perturbation of residues in binding pocket but also lead to changes in interaction of residue distant from ligand binding site. These set of residue we term as "functional sites" participate in significant long-range signaling, resulting into change in configurational entropy of residues near the G-protein coupling site.

3.5.5 Hierarchical Clustering identifies Dynamic allosteric Hot-Spots In Muscarinic Receptor

In an effort to further study the allosteric signatures in muscarinic receptor and predicting sites, which on perturbation (for e.g. by ligand binding) has potential to bring structural changes and functional changes in dynamics of receptor we performed hierarchical clustering. These Sites can be identified by studying the ensemble of residues responsible for correlations between functional regions in the receptor. Functional hubs represents group of residues whose interaction among themselves is greater than average number of interaction. Interaction/connection between the two residues in our allosteric network analysis is measured by the mutual information between residue pairs and degree is defined as number of connection residue has with other residues.

We hypothesized in an allosteric network, the clusters of residues correlated to number of other residues can serve as "allosteric signatures and could act either as potential allosteric site or can act as mediators responsible for allosteric modulation by affecting other sites. Hierarchical clustering provided insights towards finding such allosteric signatures. Hierarchical clustering analysis were performed a allosteric network based on the matrix of mutual information values between residues. Euclidean distance metric was used to cluster residues that residues shows similar patterns of correlations with other residues.

Hierarchical cluster in the receptor with modulator and ligand bound system show coupling of ligand binding site with the G-protein coupling site indicating the dynamic clusters. Residues clusters, colored yellow are residues in the flexible loop between helices ECL2 and N-terminal tail of receptor and a residue in C-terminal loop of the receptor. Mutual information between two in these torsions is less than

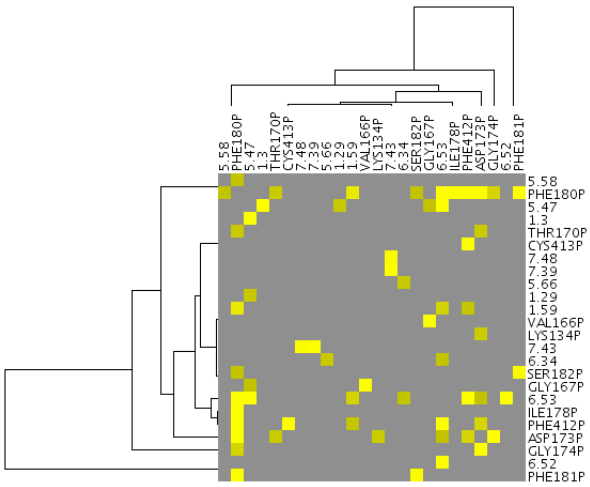
their self-entropies, ECL2 and flexible regions often have high mutual information with other residues. This yellow cluster constitutes a functional hubs or hotspot as it is highly correlated to other residue clusters.(Fig. 3.11c)

Difference in hierarchical cluster between agonist bound and agonist with allosteric ligand modulator bound receptor reveals difference in allostery induced by modulator Fig. 3.11c. In the modulator bound to receptor-agonist complex Fig. 3.11c reveal set of residue comprising of ILE435$^{7.48}$, TYR440$^{7.53}$, ASN436$^{7.49}$, ARG211$^{5.63}$, LEU455$^{8.58}$, Val133$^{\mathrm{ICL2}}$, MET139$^{4.41}$, MET456$^{8.59}$, MET112$^{3.41}$, MET143$^{4.45}$, ARG169$^{\mathrm{ECL2}}$, ASN183$^{\mathrm{ECL2}}$, GLN179$^{\mathrm{ECL2}}$, LEU43$^{1.52}$, LYS49$^{1.58}$ compared to weak cluster that are elucidated in in receptor agonist complex (Fig. 3.11b) ARG135$^{\mathrm{ICL2}}$, THR136$^{\mathrm{ICL2}}$, PHE164$^{\mathrm{ECL2}}$, TYR177$^{\mathrm{ECL2}}$, ILE149$^{4.61}$, ARG381$^{6.29}$, GLN179$^{\mathrm{ECL2}}$, VAl166$^{\mathrm{ECL2}}$. Interestingly the similar set of residue cluster were found in apo receptor (Fig. 3.11a) when we compared the clusters in apo muscarinic receptor to modulator bound to receptor-agonist complex cluster with weaker configuration entropy.

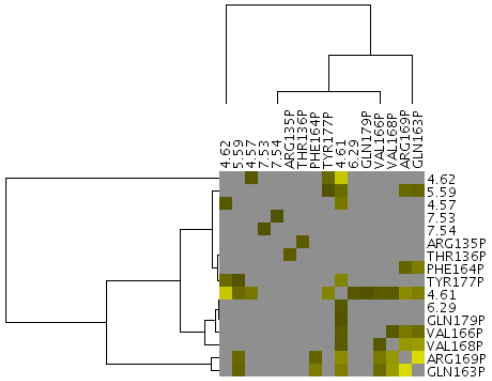
The hierarchical cluster constitutes residues from TM1-ECL2-TM3-TM4-TM5-ICL2-TM7-H8. The hierarchical clusters we identified constitutes region of the ligand binding site formed from ECL2-TM3 residues connected by the connector region which is located between the ligand-binding and G-protein-coupling sites and regions of G-protein binding site TM5-TM6-TM7 residues which also helps in identifying the path of communication from allosteric binding site to G-protein coupling site and in turns imparts allosteric bias in modulator bound complex. Furthermore, these residue cluster can thus be said to be allosteric signatures that can either act as potential allosteric site in the receptor or can be mediators indirect allosteric modulation in muscarinic receptors.

The general significance of our results and our previous findings that have explored allosteric interaction in GPCRs are supported in recent work from Flock et. al

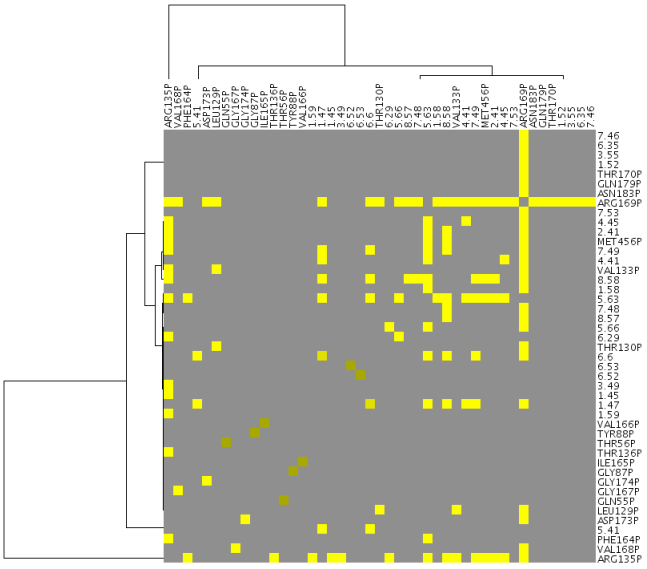
and by Vadehi and coworkers[23, 4, 5]. They have demonstrated that maintaining the intracellular and extracellular TM helical contact is critical for information transfer between ligand-receptor interface contact to the G-protein binding region and perturbation of these contacts may lead to the change in the allosteric modulation in GPCRs. This work supports their finding that the number of key/critical interactions in maintaining allostery GPCRs may be few than than previously hypothesized.



(a) Muscarinic Receptor in apo-form



(b) MQT no modulator hierarchical cluster



(c) muscarinic receptor bounded to ligand and modulator

Figure 3.11: Comparison of hierarchical clusters in Muscarinic Receptors.

3.6 Conclusion

The allosteric pathway of the Muscarinic Receptor has been identified using MD simulations. Residue motions using cross correlation analysis was used to study the Ligand(iperoxo) and modulator(LY2119620) binding. We identified allosteric lig-

ands coupled with modulator, modulates essential structural interactions that are critical to allostery in receptor. These modulation involve rearrangements of flexible extracellular loop and aromatic face to face pi-stacking interaction of hydrophobic region in the modulator with the hydrophobic interaction of transmembrane 2 residues near allosteric site. Other interaction include ionic interactions of allosteric site with the modulator that lead to allosteric modulation in residues near the G-protein coupling site. The residues in the G-protein coupling site that participate in these allosteric modulation includes residues from the TM5 and the TM6 helices in muscarinic receptor. The interaction of TM3-TM5-TM6 residues near the agonist binding site interacts with ligand residues to modulate allostery in the GPCRs. The entropically driven modulator binding enhances the inter domain motion in TM3-TM6-TM7. In receptor-ligand complex , residue in the intracellular of the receptor remains significantly more correlated compared to the extracellular region while in In the apo-form receptor these cross correlation motions of residues are found to be significantly much less. These extracellular binding of ligand induced motion in GPCR which results in protein motions partially promoting the change in molecular switch is mainly responsible for allosteric transition from inactive state to the allosterically active receptor . Our finding identified single amino acid residues, such as Arg169 of ECL2 loop, $R211^{5.63}$, $F412^{6.6}$, $I435^{7.48}$ and K491.58, to play a crucial part in the allosteric modulation. Biochemical studies revealed the role of ionic interaction involving residues Arg169ECL2 in stabilizing the active state of receptor complex , polar R2115.63 interact with G-protein coupling site in active state while $F412^{6.6}$, $I435^{7.48}$ assist in G-protein coupling site via non covalent interaction with the residues in G-protein coupling site. Thus these residues presents itself to be good candidates for experimental and computational mutation studies. Therefore, our proposed allosteric signatures could provides critical information that can lead to design

of small molecules which can interact with allosteric ligands and can modulates the allosteric signal propagation of GPCR.

Chapter 4

Role of Mutations Towards Identification of Allosteric Bias in Muscarinic receptor

4.1 Abstract

Although orthosteric binding have been potential drug target G-protein Coupled Receptor(GPCR), ease of targeting allosteric sites in GPCRS makes allosteric modulation by ligands in these sites interesting candidates for better receptor selectivity and improved binding affinity. Allosteric Modulation is known to exist in GPCRs through agonist and antagonist. While there is still wide uncertainty in path of allosteric communication and how allosteric signals propagate inside the receptor recent studies identified role of " few residues" that have contributed significantly towards activity of GPCRs. Our current work focuses on identifying the allosteric signature in GPCRs. These efforts are to identify evolution of these communications in Muscarinic receptors and how different ligands triggers biased signaling allosteric pathways

in muscarinic receptors. In recent years new insights in muscarinic receptor's physiology, pharmacology and structure have promising potential towards understanding of allostery. The advances in muscarinic receptors include the first crystal structural information of muscarinic receptors in both inactive and active conformations.The crystal structures however do not outline differences in the biogenesis mechanism of the mutant receptors compared to the wild type muscarinic receptor, which had been observed experimentally. These differences in allosteric activation and bias mainly originate from the dynamic behavior of the mutant receptors. Allosteric bias and conformational stability of mutants is yet to understood. Allosteric communication and how the mutations confer thermo-stability in GPCR can open new avenue in drug design. We have used atomic level molecular dynamic simulations to elucidate the dynamic behavior of the agonist bound muscarinic receptor as its difficult to identify these answers by experimental studies. Our results provide an understanding of the basis of allosteric mechanism in muscarninc receptor. Our analysis indicates mutations that affect the ligand efficacy, but not the binding affinity, forms the part of allosteric communication pipelines. This further explains the role of these mutation which are yet to be deciphered.

4.2 Introduction

Muscarinic acetylcholine receptor belongs to G-Protein Coupled Receptors (GPCRs) family which regulate physiological activities like sensory perception, to chemo taxis and neurological responses [57, 84]. Structurally GPCR is seven transmembrane receptor connected by three intracellular and three extracellular loops. Allosteric machinery can be define as communication of ligand binding event in the extracellular (EC) domain to IC domain that further initiate the coupling of effector

proteins [14]. Existence of multiple conformational states of these receptor have been established in past studies. These conformation includes active ,inactive and intermediate which are stabilized by binding of ligand. Existence of multiple conformation highlights the functional importance of ligands and how it regulate the dynamics of associated receptors [39]. Ligands not only plays role in conformation stability of a receptor structure that is otherwise sparsely populated in conformation subspace but also select the allosteric communication pathway. Communication pathway when ligand binding in the EC domain and how the information is communicated to residues in the IC domain has not been completely answered and is difficult to get deciphered from experiments. Allosteric modulation to increase or decreases the activity of orthosteric ligand is critical for efficient drug design [37]. Allosteric modulator don't interact with orthosteric ligand directly but the affect of activation is communicated by interacting with residues which communicate the change all the way to intracellular domain. Mapping these ligand-residue interaction and communication pathway between allosteric binding site to orthosteric binding site will provide insight towards allosteric bias as well aid in identifying therapeutic agent.

Distance of propagation of activation signaling from the orthosteric site to the inner region of G-protein binding site of the receptor is nearly 30 $\mathring{A}$ in class A GPCRs [58]. Experimental evidence on the activation mechanism of β-2 adrenergic and rhodopsin has led to suggest that the minor perturbation at agonist binding site is transmitted to the intracellular interfaces via residue contacts and mutations of the residues which are not in direct contact of ligand binding site plays role in receptor activation. Role of the mutant and the residues involved in receptor activation is not clear. Our analysis used mutual information (MI) on torsion angles and cross correlation analysis of C-α calculated from all-atom molecular-dynamics (MD) simulations of muscarinic receptor to elucidate the dynamic communication pathway.

Past computational approaches to map the allosteric modulation and identify the allosteric networks in proteins include bioinformatics method, normal mode analysis elastic network models, stress analysis, analysis of contact maps, force distribution analysis, and correlated residue motions. Identifying functional residues using computational method that involve purely bioinformatics methods, like sequence similarity and evolution analysis has limitation as that its unable to distinguish residues from structural role.

In an effort to understand the fluctuation among the protein residues and to identify allosteric interactions and participant residue one of the most commonly used method involves using principal component analysis(PCA). PCA understand the collective motion of allosteric sites to reveal the correlation between role of allosteric sites and how modified confirmation induce functional response in GPCR. [86]. Principal component analysis gives set of eigen vectors that represents collective movement of selected atoms along a principal axes. Principal component analysis may be suited to explain the small-backbone fluctuation but large conformation fluctuation that involves motion of side chain, motion in flexible loop regions couldn't be appropriately explained by studying principal components.

To study contributions arising from non mutable groups of atoms during the ligand binding events or to analyze protein residues that participate in ligand binding process Molecular MechanicsGeneralized Born Surface Area (MM-GBSA) can be quite helpful. MM/GBSA allows the decomposition of the electrostatic solvation thus facilitate in detail study of protein ligand interaction at the residue level. This can further facilitate studying of total interaction energies at residue levels in protein and will help in understanding the contribution of residues towards allostery in GPCRs [19]

In our present study the mutual information is derived using Mutinf pack-

age approach to identify allosteric networks [45]. This approach helps in quantifying correlation between the conformations of residues located at different sites of protein. Mutinf which is entropy based approach pioneered by Jacobson lab,it helps to analyze thermodynamic ensembles of protein conformers generated using MD simulations. Internal coordinated focused on dihedral angle is used to identify correlated motion in side chain rotamers is calculated by mutinf. Dihedral angle plays role for low frequency motion and in turn allows to help to identify the residue pairs with correlated structural conformation in an ensemble at equilibria.

Conformational changes is induced when ligand bind to the receptor. These changes is not limited to the ligand binding site but also throughout the receptor [45]. Based on experimental studies, allostery is triggered by interactions between binding pocket and ligand, which are mostly then propagated to the interior region near the G-protein binding site located far away from the ligand binding site. Long distance coupling forms the basis of allosteric regulation. Biophysical computational studies of many receptor ligand complex identify the role of complex to undergo conformational transition that are induced by inter-residue interactions. These inter residue contacts plays a role in identifying the pathways of signal transduction however mechanism through which these inter residue contact communicate and conservation of these communication pathways is yet to be established. Our present study demonstrates that importance of the residues within the muscarinic receptor that are essential for information transfer in the receptor.ligand bounded state. Inclusion of dynamic correlation information from receptor ligand complex gives idea to physical communication network, overall topology of protein network and helps to approximately quantify the allosteric signal strength that can be further corrobrated by experimental findings.

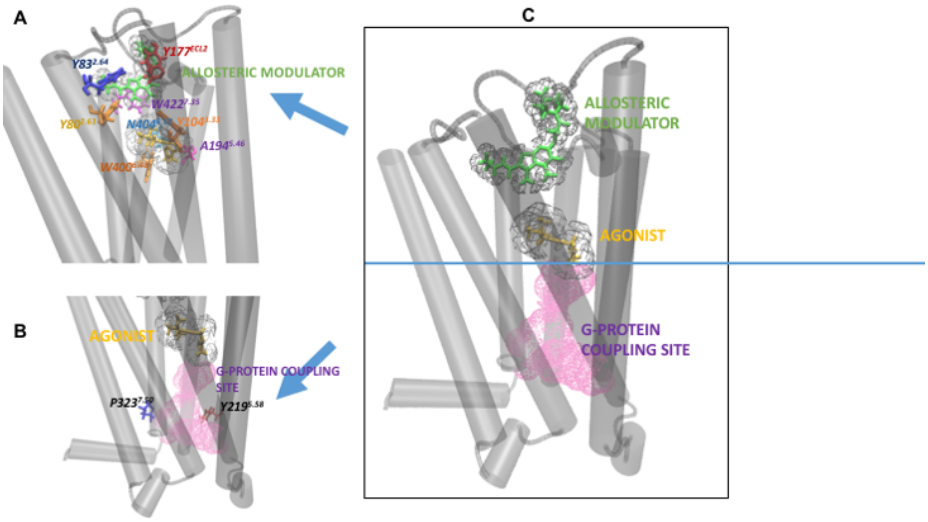


Figure 4.1: Allosteric signatures in Muscarinic Receptor.

(A) Expanded view of the allosteric region located in the EC loops of M2, showing the key allosteric hub residues. (B) Allosteric pockets near the G-protein interface in M2 receptor. The residues that define the G-protein binding site pockets are highlighted. (C) Allosteric pockets identified in the active conformation of Muscarinic Receptor. Interacting residues of allosteric binding sites in M2, along with LY2119620 in M2-AcR is highlighted.

4.3 Methods

4.3.1 Molecular dynamic simulation protocol

Starting from muscarinic receptor crystal structure inactive struncturs bound to inverse aginist(PDB id: 3uon), active human M2 muscarinic acetylcholine receptor bound to the agonist iperoxo (PDB id: 4MQS), active human M2 muscarinic acetylcholine receptor bound to the agonist iperoxo and allosteric modulator LY2119620(PDB id: 4MQT) was used in this analysis. The muscarinic receptor was inserted into a palmitoyl-oleoyl-phosphatidyl-choline (POPC) bilayer. Lipid molecules which were overlapping were removed using the Membrane plugin in VMD. The charges in the system were neutralized at 0.15 M KCl using the Solvate plugin in

VMD. Initial Box size 95 X 95 x 105 .Simulation was performed using CHARMM36 force field parameters.46 K+ , 58Cl-, 230 lipid molecule.Periodic boundary conditions were applied to the system during the simulations. Modulator QNB was removed from the starting structure for the apo system without inverse agonist.

Molecular dynamics (MD) simulations were performed using NAMD2.9(12). The CHARMM36 parameter set with CMAP terms was used for the protein,POPC. For the water molecules we used TIP3P. Parameters of iperoxo,LYS2119620 that was modeled using ParamChem web server. 12 distance cutoff was set for the van der Waals. particle-mesh Ewald summation method using a grid point density of $1\mathring{A}$ was used for calculating long-range electrostatic interactions. A 2fs integration time-step was used for all MD simulations. Multiple-time-stepping method was used and bonded and short-range non-bonded interactions was computed at every time-step while long-range electrostatic interactions was calculated at every two time-steps [54]. The SHAKE algorithm was applied to all hydrogen-containing bonds. The SHAKE algorithm was applied to all hydrogen-containing bonds. Molecular dynamic simulations equilibration was first performed on the system in NAMD. First, all atom fixed the lipid tail was energy minimized for 2000 steps for 1.0 ns. Melting was NVT run at 310K. Further equilibration was carried using NPT run at 310K with 10 kcal/mol.2 on Backbone and 5 kcal/mol.2 on rest of the system for 10ns. Restraints on backbone and rest of the system was further reduced for equilibration in following ways (i) 5 kcal/mol.2 and 2.5 kcal/mol.2 respectively for 10 ns (ii) 2.5 kcal/mol.2 and 1 kcal/mol.2 respectively for 10ns (iii) 1.0 kcal/mol.2 and 0.5 kcal/mol.2 respectively for 10ns (iv) 0.5 kcal/mol.2 and 0.1 kcal/mol.2 respectively for 10ns (v) 0.01 kcal/mol.2 on for 10ns. After above minimization and equilibration procedures, equilibrated system was prepared for production run. Systems were then run for nearly 200 ns for production run.

4.3.2 Cross-Correlation Analysis

Correlation between residues in receptor ligand system was analyzed for 200 ns MD trajectory by calculating the normalized covariance:

$$C_{ij} = \frac{Cov_{ij}}{((< \Delta r_i (t)^2 >) \cdot < \Delta r_j (t)^2 >))^{1/2}} \tag{4.1}$$

where

$$Cov_{ij} = (< \Delta r_i (t)^2 > \cdot < \Delta r_j (t)^2 >) \tag{4.2}$$

and

$$\Delta r_i(t) =< \overrightarrow{ri(t)} - \overrightarrow{ri(t)} > . \tag{4.3}$$

where ri(t) is the position vector of C_α of ith residue and the time average of the quantity is in the <.> brackets. The correlation matrix is generated that characterize the motion of protein residues.The correlation ranges between -1 to 1. If the residue move in same direction the motion positive correlated and if they move in opposite direction the motion is anti correlated.

4.3.3 MM/GBSA

MM/GBSA method was used to calculate the binding affinity between (i) receptor and ligand (ii) receptor and ligand-modular complex. MM/GBSA was estimated using CHARMM. Following equation weere used to estimate binding free energy.

$$\Delta\Delta G_{binding} = \Delta G_{complex} - \Delta G_{receptor} - \Delta G_{ligand+modulator} \tag{4.4}$$

where ΔE_{VDW} is energy term of sum of Van-der Waals energy

$$\Delta G_{SA}^{non-polar} = \gamma SASA + b \tag{4.5}$$

The value of 0.00542kcal/mol/ $\mathring{A}^2$ was used for the surface tension γ and the value of 0.92kcal/mol was used for the constant b. We assume that the change of entropy upon protein-ligand binding is small and can be neglected, the equation for calculating binding affinity is equation

$$\begin{aligned} \Delta G_{MM/GBSA} = &<\Delta E_{VDW}> + <\Delta E_{ELEC}> + <\Delta E_{INT}> \\ &+ <\Delta G_{solv}^{polar}> + <\Delta G_{SA}^{non-polar} - T\Delta S \end{aligned} \tag{4.6}$$

The snapshots taken from trajectory at 0.25ns and final free energy was estimated by averaging the values of energy terms of ensemble.

4.3.4 Ligand.Receptor Network

A network is defined in terms of nodes that are connected with edges. Nodes here means amino acid residues and edges is defined if connecting pairs of nodes are in contact. Two nonconsecutive nodes are connected in our network if any non hydrogen heavy atoms is within $4.5\mathring{A}$ of each other for at least in 75% analyzed simulation. Minor changes in network and community distribution was observed when change in parameters that define network contacts. Nearest neighbors in sequence is ignored to reduce noise resulting from number of trivial sub optimal paths in the weighted protein network. The networks was build using information from 200 ns trajectories of the receptor-ligand complexes. The dynamical proteinligand network is a weighted network in which edge between nodes i and j is information transfer probability across the connection and is measured in terms of correlation strength between the two nodes

:

$$wij = log(|Cij|) \tag{4.7}$$

4.3.5 Mutual Information Analysis.

Mutual information analysis metric is used as metric to evaluate of long-range coupling between residues.The evaluation is done by calculating the correlation between the motions of GPCR protein backbone and side-chain torsion angles.These calculations are further used to identify allosteric couplings in muscarinic receptors.Mutual Information(MI) between two residue is given by:

$$MI(x,y) = \sum_{x \epsilon x} \sum_{y \epsilon y} p(x,y) log(\frac{p(x,y)}{p(x)p(y)}) \tag{4.8}$$

p(x,y) represent the joint probability distribution of residues X and Y. p(x) is the marginal probability distribution function of residue X and p(y) is the marginal probability distribution function of residue Y. In here X and Y represent one rotameric state of residue X and Y respectively.

We quantified correlation between residues torsion using McClendon et.al. mutinf package. Mutual information approach connects between information theory and thermodynamics and calculates configurational entropies from conformational ensembles. To calculate configurational entropy, internal coordinates of the system is used. Second-order terms is derived from entropy expansion (mutual information) and is further used to determine residue pairs with correlated conformations. The method uses molecular configurational entropy over torsion angle changes in residues. The mutual information represents correlations between degrees of freedom. Mutual Information quantifies amount of information gained by one degree of freedom by

knowing another. Summation of torsion entropy can be calculated by

$$\begin{aligned} S_{conf} = & \sum_{i}^{n} \int_{0}^{2\pi} p(\phi) ln p(\phi) d\phi \\ & - \sum_{i}^{n} \sum_{j}^{n} \int_{0}^{2\pi} \int_{0}^{2\pi} p(\phi_1, \phi_2) ln \frac{p(\phi_1, \phi_2)}{p(\phi_1) p(\phi_2)} d\phi_1 d\phi_2 + ... - ... \end{aligned} \tag{4.9}$$

where indices i and j represents torsions of residue, n is the number of torsions for residue. We used $(\phi, \psi, \chi_1, \chi_2)$. The second-order term is sum of the mutual information of each pair of torsions. The mutual information explains correlated relationship between degrees of freedom. By help of mutual information we aim to quantify the gain of information about one degree of by knowing the another. Mutual information values in here are in terms of entropy, which is eventually related to free energy hence the the mutual information in here Equation 4.9 is in units of kT. [45] Mutual information is a distribution-free analysis method and robust statistical measure that doesn't rely on assumption about the sampling data from a given probability distribution, enables us to understand the changes in conformation.

4.3.6 Allosteric Network based on Mutual Information

Mutual information between the torsion angle of residue was used to study the allosteric contribution of each residue. We constructed the allosteric network based on the pair wise correlation between the residues. Cross correlation of torsion angle in this approach is defined in kT unit. To study major contribution and reduce noise, our network construction limit the correlation greater than kT. Weighted edge between two residue is defined in terms of kT and degree of a node is number of edges passing through each node. In an effort to understand the protein ligand interaction and to identify which functional site can alter the conformation or dynamics

of receptor by perturbation hierarchical clustering was used to identify the group of residues which are critical for correlation between the sites important for receptor function. Hierarchical clustering of constructed allosteric network was performed using cytoscape. Euclidean distance metric was used to cluster residues with similar patterns of correlations with other residues

4.3.7 Community Repartition Difference

Mutants of muscarinic receptor and ligand receptor interface of the mutant and receptor as well for examination of the robustness of the community algorithm, we used a community repartition difference(CRD). CRD helps to quantify and evaluate the difference between a reference community partition (c1) of muscarinic receptor (4MQT) and a new partition (c2) of the same set of nodes of wild type. Partition measure helps to evaluate difference in community structure among mutant and wild type. Method is based on algorithm developed by Knox [35] and Rand [56]. The community repartition difference (CRD) is defined as

$$CRD(c_1, c_2) = 1 - \frac{\sum_{ni,nj} z(ni, nj, c_1) z(ni, nj, c_2)}{\sum_{ni,nj} z(ni, nj, c_1)} \tag{4.10}$$

4.4 Result and Discussion

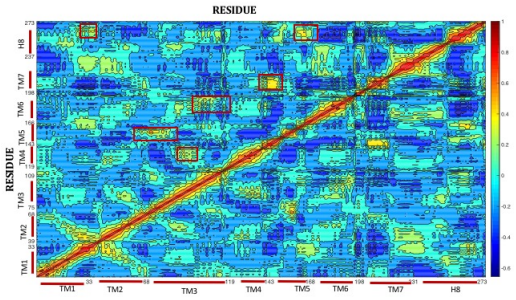
4.4.0.1 Correlated Motion between the Ligand Binding site and G-Protein Coupling site reveals differential long distant signaling among Mutants

Correlation motion among residues were identified between allosteric binding site and G-protein coupling site. In the receptor bound to the orthosteric agonist iperoxo and the positive allosteric modulator LY2119620 (2CU), where receptor stays in active conformation with increased basal activity residue motions. In the wild type (Figure 4.2a, the intracellular domain of transmembrane regions TM5, TM6 and TM7 is significantly correlated to the extracellular domains of muscarinic receptor. Intracellular domain of TM5 and TM6 is correlated to the TM3 and TM5 respectively. Significant correlation is identified between extracellular domain of the TM5 and ex-

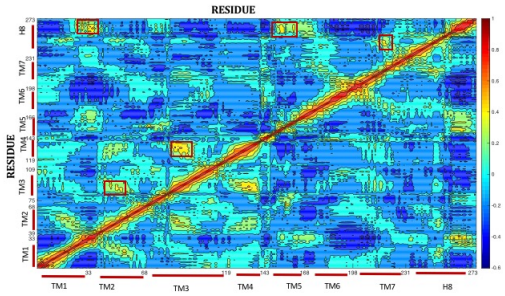
tracellular domain of TM6 and TM7. These above mentioned region for the ligand binding site in the receptor. Past experimental studies determined the ligand binding pocket for the allosteric ligand is formed mainly residues of TM3,TM5, TM6 and plays crucial role towards allosteric regulation [39, 81, 5]. Allosteric modulator has been shown to interact with the extracellular cavity, formed mainly by the TM2, ECL2, TM7 residues and upon binding increases the basal activity of allosteric ligand.Table 4.1 enlist residues participating in then receptor ligand interaction. These allosteric induce correlated motion changes near the ligand binding site is further translated into large TM domain motion [22]. The correlated motion changes between the cytoplasmic residues of TM5 and TM3 residues of allosteric binding site indicates the long range information transfer. The changes in residues of TM5 is transmitted via helix to the cytoplasmic end rearrangement of TM5. Also, the correlated motion of TM5-TM6 has linked rotation of TM6 near Phe6.44, results in the relocation of the cytoplasmic end of TM6 which leads to receptor biogenesis [22].

In the M2 receptor Ser210 plays crucial role in the activation of $G_{q/11}$. SER210 lies towards bottom of the TM5 and make contact with carboxy terminal helix of the $G\alpha_s$ [38]. Past studies reveal SER210 is critical. In receptor activation, G-protein bind event leads to displacement $Ser210^{5.62}$ cytoplasmic end of TM5 approximately $4\mathring{A}$. The SER210Y mutant exhibit significantly higher correlation between the cytoplasmic end of the TM5 and Tm7 with the ligand binding end of the TM3. The higher correlation explains the outward movement of the Tm5 helix which is triggered by the allosteric binding and movement of TM5. SER210Y mutation result in the change in correlation of the $T195^{ECL2}$ with Helix H8. $T195^{ECL2}$ located above the orthosteric binding site increased the correlation in mutant indicates long distant signaling in the receptor.(Figure 4.2d)

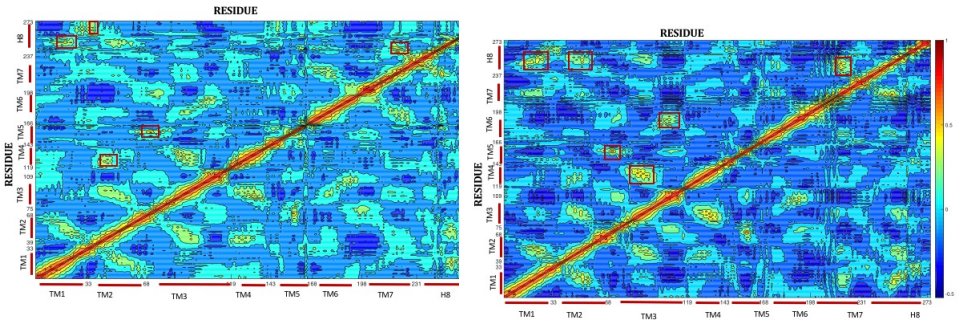
The residue $THR110^{3.29}$, near the agonist-binding site has increase positive



(a) 4MQT



(b) THR190A



(c) ASN419D (d) SER210Y

Figure 4.2: Cross correlation of residue motion is identified between residues in the extracellular region near the G-protein-coupling sites and ligand binding site located in the intracellular part of the M2 receptor and mutants bounded to allosteric ligand and modulator.

Figure here is dynamic map of residue pairs cross-correlations that are color coded based on correlation strength in ligand and modulator bounded muscarinic receptor and in mutants. Residues in the TM1 to TM7 and helix H8 is shown by red bars on the top and right. The red box indicates the long range coupling between the ligand binding site and G-protein binding site.

correlation with the EC loops and the G-protein-coupling region. These correlation among residues in wild type indicates the two residues are critical for the that show allosteric communication. The other residues that involve in long distant communication are $Y316^{7.43}$ and $S207^{5.46}$. $Y316^{7.43}$. The above mentioned residues is important for allosteric communication in muscarinic receptor and has shown to interact with the agonist directly via hydrogen bond with the protonated amine of the ligand while the residue $S204^{5.43}$ indirectly modulates the allosteric activity in the active state(Figure 4.2b).

The residue ASN419 is essential for allosteric modulation. Previous studies on muscarinic acetylcholine Receptors to understand the selectivity of the Allosteric interactions has revealed ASN419 play important role in the gallamine, an allosteric modulator binding [27]. $ASN419^{7.32}$ when mutated to Aspartate has been reported to have higher affinity towards allosteric ligand than asparagine [27]. Correlation studies to understand the role of these residues towards allosteric modulation reveals the distant long range signaling. Compared to wild type, mutant receptor has increased positive correlation between ECL2 loop residue above the binding site with the intracellular region of TM7. Change in the positive correlation of $TYR430^{7.43}$ with the TM2 residues indicates the increase allosetric modulation as flipping of side chain of $Tyr430^{7.43}$ from ligand binding cavity to toward TM7TM2 interface. $TYR430^{7.43}$ plays crucial role in receptor activity as its side chain reorients the lead to displacement of the NPxxY with subsequent modulation in receptor activity.(Figure 4.2c)

Increased correlation of motifs in the intracellular domain of TM5-TM7 and TM6 also indicates the long range coupling in in AS419D mutant compared to wild type. However in the mutant, $Tyr206^{5.58}$ and $Tyr440^{7.53}$ which exhibit significant correlations in wild type is weakly correlated . This indicates that the weaker correlation $Tyr206^{5.58}$ and $Tyr440^{7.53}$ lead to decrease receptor activity. Relocation of side chains

of Tyr206$^{5.58}$ and Tyr440$^{7.53}$ residue via hydrogen-bonding interaction leads to TM6 tilts [39, 37, 16] lead to the receptor biogensis and receptor activity in wild type receptor ligand complex. These changes further assist in G-protein binding are distinct in wild type while mutant exhibit different path towards allosteric modulation.

THR190 plays a crucial role in the acetylcholine binding binding to the receptor.Acetylcholine binding is mainly driven by ionion interaction between the negatively charged aspartate (D103) of TM3 and positively charged amino head group of muscarinic ligands. Hydrogen binding between the ligand and The THR190 stabilizes these interaction [30]. THR190A alter binding of agonists reveals 3 fold to 21 fold decrease in the the binding affinity [30].

Correlation motion study of THR190A reveals the change in correlation near the binding pocket and in the TM1-TM7 reveals change in the THR190$^{5.52}$ induce change in long range signaling in TM1-TM7 and towards the agonist binding site at TM3-ECL2 regions. Structural changes THR190$^{5.52}$ seems to change the correlation of TM5 residue Pro5.50 which induce distortion of TM5 and translation and rotation of TM5 leading to receptor activity. THR190A mutation leads to decrease in the correlation of Pro5.50 with TM3 and TM7 residues which further suggesting role of these residue in decrease in receptor's ligand binding affinity.(Figure 4.2b)

Ligand	Amino Acid	Sequence Number	Generic Number	Segment	Interaction	Interaction Slug
LY2119620	Y	80	2.61x60	TM2	polar (hydrogen bond)	polar_donor_protein
LY2119620	Y	83	2.64x63	TM2	hydrophobic	hyd
iperoxo	Y	104	3.33x33	TM3	hydrophobic	hyd
LY2119620	Y	177	45.51x51	ECL2	aromatic (face-to-face)	aro_ff
iperoxo	A	194	5.46x461	TM5	hydrophobic	hyd
iperoxo	W	400	6.48x48	TM6	hydrophobic	hyd
iperoxo	N	404	6.52x52	TM6	polar (hydrogen bond)	polar_donor_protein
LY2119620	W	422	7.35x34	TM7	aromatic (face-to-face)	aro_ff

Table 4.1: Ligand Receptor Interaction table

4.4.1 Dynamic Allosteric Networks in Muscarinic Receptor

A highly dynamic network is revealed in muscarinic GPCR using community network analysis. There is significant differences in the network in allosteric modulator -ligand-bounded muscarinic receptor and in the mutants. The residues of muscarinic receptor and in mutants are distributed into highly connected clusters(communities) and the cluster do evolves in mutants that suggest allosteric bias in receptor. The strength between the cluster constituting G-protein-coupling sites and ligand binding sites changes in mutants. We hypothesized that the mutation of residues that result in decrease in agonist efficacy upon binding, should help in stabilizing the active conformation of the receptor, and mutation of residues which result in increase agonist efficacy should stabilize the inactive conformation state of muscarinic receptor.The strength between the cluster in the allosteric network will also change based on the changes in efficacy of receptor.

Although there exist dynamic modulation of the community there exist allosteric signature which can be defined as the residue that form common framework for allosteric information transfer among receptor and mutants. The intracellular domains of TM3 ,TM5,TM6,TM7 is strongly connected and strong network is identified which is connected via SER118$^{3.47}$-Tyr206$^{5.58}$, Val106$^{3.35}$-Thr399$^{6.47}$, Phe195$^{5.47}$-Asn404$^{6.52}$, Thr399$^{6.47}$-Leu205$^{5.57}$, Ile392$^{6.40}$- Tyr440$^{7.53}$.(Fig. 4.3D) In the ligand-binding site of the agonist allosteric ligand iperoxo the clusters are strongly connected via critical residue connecting extracellular domain of TM5,TM6,TM7 to TM3. Extracellular region of TM6 and TM7 is strongly connects to TM3 strongly through Val106$^{3.35}$ which is very near to ligand binding site. Strong information network is identified between TM6-TM5 via the Asn402$^{6.52}$ interactions((Fig. 4.3B), mutation at this point has reported to reduce agonist binding affinity by $>$ 10-fold [30].

With the mutants in the Muscarinic receptor, there is significant alteration of the levels of network communication. In the SER210Y the the extracellular domains of TM3 and ECL2 and the allosteric modulator binding site is strongly connected via the ILE178^{E2}Tyr178^{E2} interaction (Fig. 4.3F). Connector region identified in TM3-TM5, is between the G-protein-coupling sites and ligand binding region [39] exhibit medium network ILE392$^{6.40}$-Y4407.53 interaction. However, in the intracellular region and towards the G-protein coupling site and the inner regions of the TM3, TM5, TM6, and TM7 is connected weakly to each other (Fig. 4.3H).

In the mutant THR190A which is near to the allosteric binding site extracellular residue of TM3 merges with the ECL2 loops residue critical for ligand binding.This cluster of extracellular TM3 and ECL2 is very weakly connected to the extracellular TM6 residues interacting with the allosteric modulator via Y403$^{6.43}$-PHE181^{ECL2}(Figure 4.3E). This could possibly explain the decrease in the binding affinity of mutant towards allosteric modulator. Intracellular region of TM3-TM5-TM7 is loosely clustered into different community and cross community transfer is facilitated by residues ILE389$^{6.37}$, ASN432$^{7.45}$, ILE432$^{7.43}$, LEU62$^{2.43}$ as it connects as it facilitate in binding of G-protein coupled receptor (Figure4.3G).

ASN419 to Aspartate mutations has been reported to have higher affinity towards allosteric ligand than asparagine [27]. The network analysis of ASN41D mutant shows it exhibit stronger community interaction in extracellular side of the receptor, weak allosteric network is identified towards the intracellular near the G-protein coupling site (Fig.4.3A, Fig.4.3C). Extracellular region of TM3-TM7 that participate in the ligand interaction and includes residue Y104$^{3.33}$, W422$^{7.34}$ is strongly connected via ASN432$^{7.45}$ - SER433$^{7.46}$ PHE195$^{5.47}$-TRP400$^{6.48}$ connects the TM5-TM6 cluster interacting with allosteric modulator responsible for increase activity allosteric ligand(Fig.4.3A). Towards the intracellular region of ASN419D the moderate

network strength is found between the TM3-TM5-TM6 regions. The residues near the G-protein coupling site breaks into separate clusters and contact is essentially facilitated via $Y122^{3.51}$-Ile2095.61-$SER380^{6.28}$(Fig.4.3C).

Mutations on muscarinic receptor leads to the significant repartition in the community network of mutants. As measured by the community repartition difference, In ASN419D 40% of the node pairs remain clustered in their previous same community, while 65% of the node pairs is reorganized into different separate communities. In TH190A 50% of the node pairs remain clustered in their previous same community, while 50% of the node pairs is reorganized into different separate communities and in SER210 60% of the node pairs remain clustered in their previous same community, while 40% of the node pairs are further reorganized into different separate communities.

In THR190A allosteric bind site reconnects with the extracellular loop in same community while the orthosteric residues in TM5-TM7 binding site in the mainly $Pro198^{5.50}$, $SER433^{7.46}$, $THR433^{7.47}$, $PRO437^{7.50}$ which were in the strong same community C-8 in the wild type now redistribute in C-1 and into newly formed C-10 communities(Fig.4.3E, Fig.4.3G). While in SER210Y multiple TM3 and TM5 and TM6 residue of the binding site reconfigure in multiple community resulting in the weaker interaction and differential biased signaling(Fig.4.3F, Fig.4.3H).

In ASN419D 65% of the node pairs reorganizes while 35% of the node pairs ares in previous same community. Compared to the community distribution in the wild-type muscarinic network, there is slight variation in community boundaries in the mutant. Significant changes occur near to community C-5 and community C-8 and community C-10. In new community C-10 and C-11 contains many strong contacts that C-10 previously constituted with C-5($TYR403^{6.51}$), C-9($ILE209^{5.61}$) and C-8 ($TRP422^{7.34}$).

While different Allosteric Interaction involves different Community network among Mutants and wild-type, allosteric signatures are identified that are critical for information transfer and allosteric communication in muscarinic receptor and mutants. Allosteric signature constitute
ICL1(LEU54P), TM2(ASN59$^{2.40}$), TM3(TYR104$^{3.33}$), ECL2(PHE181),
TM6(TRP400$^{6.48}$, TYR403$^{6.51}$), TM7(ASN432$^{7.45}$, TYR440$^{7.53}$). (Fig. 4.4). Allosteric signature which can be defined as the residue that form common framework for allosteric information transfer among receptor and mutants.

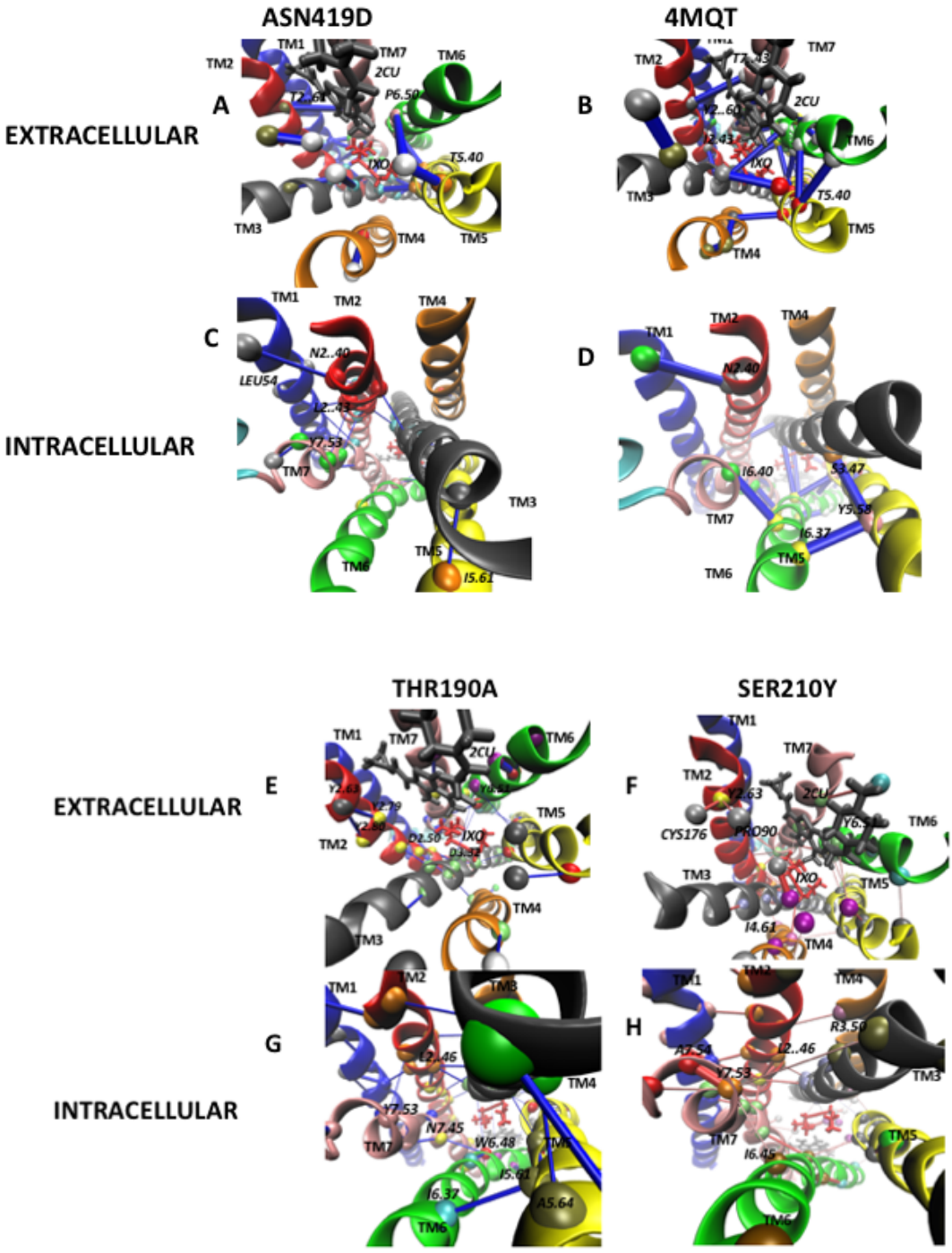


Figure 4.3: Dynamic network in Muscarinic Receptor

Figure 4.3 identifies Highly dynamic network is identified in the ligand(iperoxo), modulator(2CU) bounded muscarinic receptor(M2) through community network analysis. Extra-cellular views of the allosteric binding site for the mutants (A,E,F) and the active muscarinic receptor (B) states are depicted, and corresponding view from intracellular perspective of the ligand-binding site is depicted in C,G,H. Allosteric network exhibit significant difference between the allosteric bounded form and the mutants. Particularly, the allosteric network between clusters in the receptor and mutants are significantly different and weaker in intracellular domains. In network communities, the receptor states are colored coded using unique index. Critical nodes are mainly located at the interface of near by neighboring communities is represented spheres and is marked using residue number. Connecting edges between he nodes are shown by blue lines and its width is weighted by betweeness. The weighted edge here represents the probability of information transfer across residue communities.

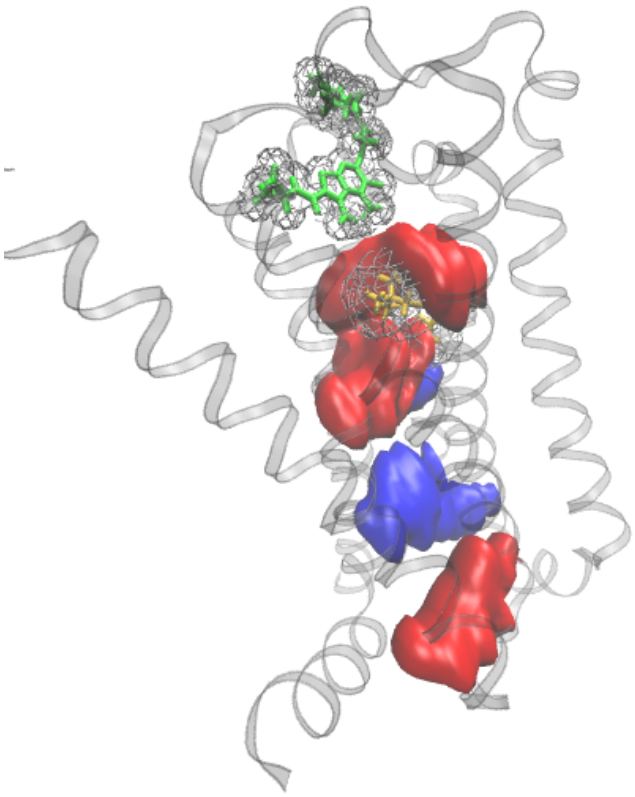


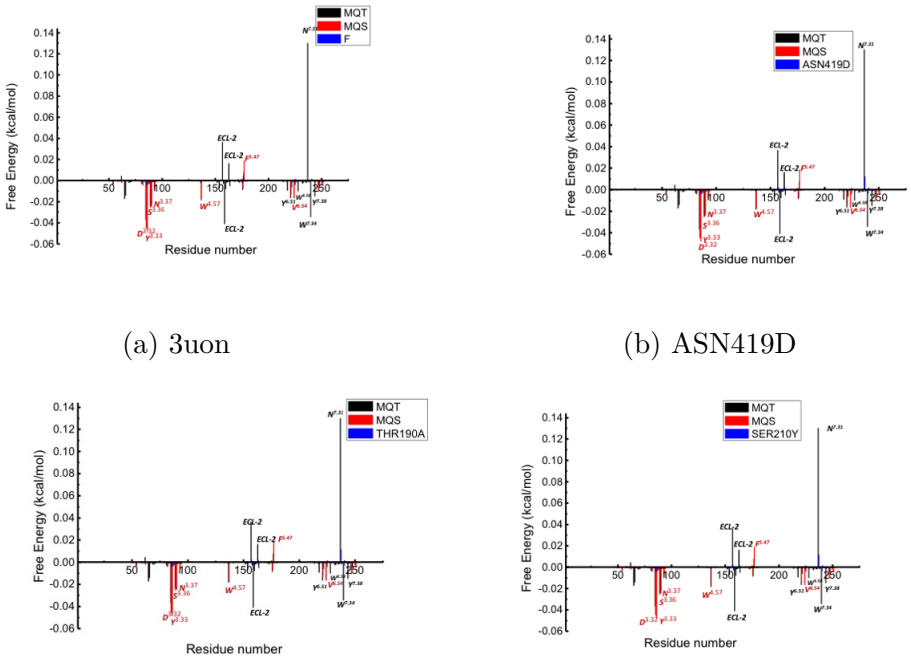
Figure 4.4: Allosteric signatures in Muscarinic Receptor.

Allosteric signatures are identified in muscarinic receptor based upon the information flow between allosteric binding site and G-protein coupling site.The red surface in GPCR is identified as signature composed of ICL1(LEU54P), TM2(ASN59$^{2.40}$), TM3(TYR104$^{3.33}$), ECL2(PHE181), TM6(TRP400$^{6.48}$, TYR403$^{6.51}$), TM7(ASN432$^{7.45}$, TYR440$^{7.53}$) These residues connect allosteric modulator(green) interaction site,allosteric ligand (yellow) binding site to G-protein binding site(blue)

4.4.1.1 Per Residue Interaction Energy reveals the Allosteric sites in Muscarinic Receptor

For the purpose of obtaining the detailed presentation of the receptor/iperoxo interactions, free energy decomposition analysis was employed to decompose the total binding free energies into per residue basis. The quantitative information of each residues contribution is critical in understanding the allostery and information flow in the receptor.

In order to identify important muscarinic receptor residues that interacted with allosteric ligand and modulator, we analyzed the interactions between muscarinic receptor and modulator for M2R-iperoxo-LYS2110. Among the experimentally reported residues, which are important to agonist binding [39], allosteric ligand and allosteric modulator interacted strongly with $L82^{2.62}$, $T84^{2.64}$, $W250^{4.57}$, $A^{5.47}$, $Y403^{6.51}$, $M406^{6.54}$, $N410^{6.58}$, $W422^{7.34}$. Besides, the interactions between receptor and ligand.modulator ECL-2 loops residues, $S107^{3.36}$, $N108^{3.37}$, $Y426^{7.38}$. These interactions are consistent with the recently reported agonist bound muscarinic receptor and agonist and modulator bound muscarinic receptor structure [39], which also supports our model. These interaction important for the allostery and to investigate factors besides inter-molecular interactions that are involved in conferring specificity we binding free energy to explore which residues make a significant inter molecular contribution towards differential binding. Our observation as expected in general, reveals mutations of residues near the active site residues are highly unfavorable with all four mutants there is reduce in the binding free energy with the ligand. Though there is the reduce in the binding free energy but per residues energy decomposition reveal that residues identified as critical in wild type muscarinic, also in mutant have significant more contribution towards allostery than other residues.



(a) 3uon

(b) ASN419D

(c) THR190AA

(d) SER210Y

Figure 4.5: Comparison of decomposition of the binding free energies on a per-residue basis for the Muscarinic receptor and Mutants

Figure compare per residue contribution towards free energy in (a) QNB-bound muscarinic receptor/allosteric ligand and modulator bound receptor (MQT)/allosteric ligand bound receptor (MQS) (b)MQS/MQT/ASN419D (c) MQS/MQT/THR190A and (d) MQS/MQT/SER210Y.

4.4.2 Significant Long-Range Correlations in Muscarinic Receptor is Identified using Mutual Information from MD simulation

Mutual information study was used to analyze unbiased MD simulation of muscarinic receptor. We analyzed correlations between residues in muscarinic receptor to reveal cooperative sites. Mutual information also helped in exploring correlations between in ligand modulator bound receptor and mutants.

Mutual information calculation for each pair of residue involved calculation of $\phi, \psi, \chi 1, \chi 2$ torsion angles in our simulations of muscarinic receptor and mutants. Metric for measuring mutual information is kT, due to the relationship between mutual information and entropy which is further related to binding free energy. significant mutual information was identified between small subset pair of residue in receptor and mutants.

In the THR190A mutant strong correlation exist to the intracellular regions of TM3-TM5 is strongly correlated. DRY motif, which is responsible for receptor activation and G-protein binding is correlated with the intracellular region of TM5 S210$^{5.62}$, V204$^{5.56}$. Intracellular region of TM6-TM7 strongly correlation with allosteric modulator binding site TM4-TM7-ECL2 regions consolidate the long distance signaling between G-protein coupling site and allosteric binding site.(Fig 4.6b).

Compared to wild type in THR190A mutant there is significant loss in the correlation between the connector region, which is region between ligand binding and G-protein coupling site region in muscarinic receptor [22]. In the wild type extracellular side where agonist modulator bind is very strongly coupled.Residue from ECL2 (Arg169),TM5(ALA184$^{5.37}$)-TM7(VAL421$^{7.33}$, ILE431$^{7.49}$) is strongly coupled with

intracellular region of TM5 (PHE188$^{5.41}$, ARG211$^{5.63}$), TM6 PHE412$^{6.60}$ and to the connector region signifies the strong allosteric signaling in receptor.(Fig 4.6b,4.7a). This finding explain the change in allosteric modulation between the mutant and wild type and emerging of new allosteric correlation in mutant.

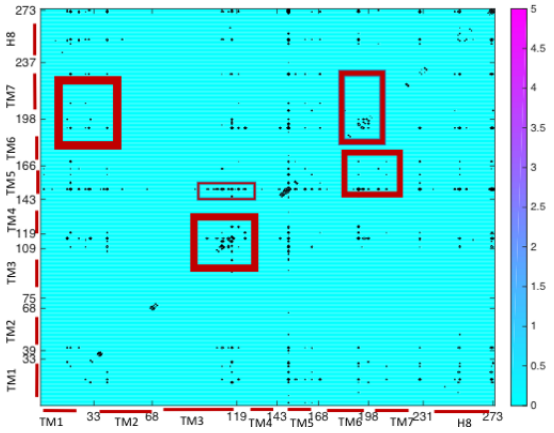
In SER210Y mutant the intracellular region of TM5 -TM6-TM7 is strongly correlated and extracellular regions of allosteric binding site and G-protein coupling site is connected via ECL2-TM5-TM6-TM7 correlation of residue ASN444$^{7.57}$, LEU393$^{6.41}$, ARG211$^{5.63}$. In the connector region the G-protein-coupling sites and ligand-binding has very weak correlation between TM3 and TM5 via the Val111$^{3.40}$ Pro198$^{5.50}$ interaction. (Fig. 4.6d). The difference in the correlation between residue pairs in SER210Y mutant and wild type reveals that though there exhibit the allosteric communication and long distance correlation between the ligand binding site and G-protein coupling site, the path of these allosteric modulation varies significantly compared to wild type.

In ASN419D there is significant loss in correlation among the residue pairs and number of weak correlation increased in the and the receptor. Despite significant loss in significant correlation the allosteric communication of extracellular binding site to intracellular region is mainly carried out through strong coupling of ECL2 loops (PHE181P, SER182, GLN179, PHE181, SER182),TM7 (TYR426$^{7.38}$ and TM5 (ILE397$^{5.45}$) residue is correlated and the intracellular regions TM5 (SER210$^{5.62}$), TM7 (ALA441$^{7.54}$). These loss in correlation between the residue pair corresponding to ligand binding site and G-protein coupling site highlights the importance of role of ECL-2 in allosteric communication which is responsible for propagating the allosteric modulation in mutant receptor.

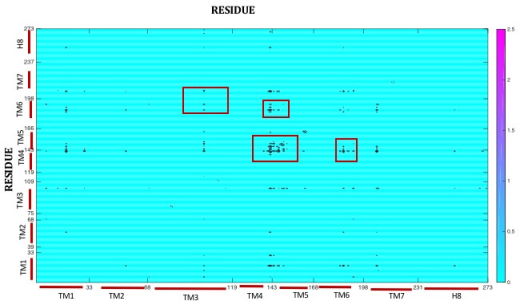
These allosetric behavior is consistent with thermodynamic considerations in G-protein coupled receptors, where conformational changes in ligand binding site or in

G-protein coupling site serves in optimizing allosteric coupling between distant sites in receptor. These allosteric coupling mostly are long distant signaling and mostly constitute by network of residue that doesn't require linking of the sites [45]. Since for residues' self-fluctuations mutual information are not normalized, is the reason why regions of ECL2,ICL1 and flexible residues connecting transmembrane regions, in some cases showed couplings $\rangle$ kT while intervening and regions of transmembrane helical residues did not.

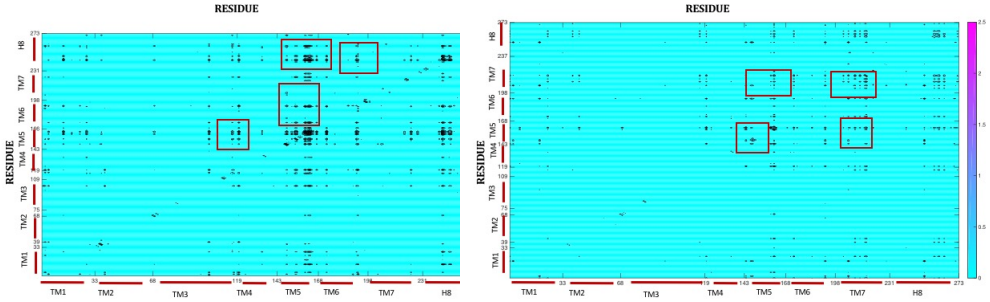
Our current approach using mutual information in thermodynamic ensembles identifies correlations between the functional sites in the receptor. Mechanism on how these changes in the mutant, indirectly mediate these correlation is beyond the scope of this study but area for future work.



(a) 4MQT



(b) THR190A



(c) ASN419D

(d) SER210Y

Figure 4.6: Correlated motions is revealed between residues in the intracellular G-protein-coupling sites and in extracellular regions of ligand-binding of the muscarinic receptor and mutants bounded to allosteric ligand and modulator.

Figure above is a dynamic map of muscarinic receptor and its mutant.The correlation strength between residues' pair is color coded and represent cross-correlations for the ligand and modulator bounded muscarinic receptor and mutants. Residues in the transmembrane helix (TM1 to TM7) and in helices H8 are depicted by red bars on the top and right. The red box indicates the long range coupling between the ligand binding site and G-protein binding site.

4.4.3 Mutual Information Based Network Reveals Differential Sites for Allosteric Network

Allosteric network and identifying residues that are crucial for the information flow in the receptor is important as it can help in identifying the therapeutic targets and potential drugabble sites. Targeting allosteric site can cause shape or flexibility changes at distant site which can help in characterizing desired clinical response. Identifying allosteric sites can also help in determining potential druggable site in terms of identifying the interaction of ligand with therapeutic modulator and determining better fit and hydrophobicity for small therapeutic drug like small molecules.

Our mutual information based network is aimed to predict residues and functional site that on perturbation have better chance in altering conformation and dynamics of the receptor. Our analysis aim to identify set of residues crucial for correlations between functional sites in receptor. Network based on residue pair with greater than kt correlation in muscarinic receptor bound to ligand and muscarinic receptor identifies helps in identifying such sites.

In this Mutual information based network we found the functional hubs in receptor whose interaction among themselves is greater than average number of interaction. Based on previous protein interaction networks, we hypothesized these functional hubs could act either as potential allosteric site or can act as mediators responsible for allosteric modulation by affecting other sites. In muscarinic receptor the set of residues that are potential allosteric hotspots(Fig. 4.7) lie majorly in intracellular region of TM2-TM4-ICL2 and H8-TM5-TM7 near the G-protein coupling site and allosteric modulator perturbing these site can modulate the receptor biogenesis.

Allosteric communication in THR190A mutant is different compared to wild type. The hotspot identified in THR190A mutant is based on mutual information

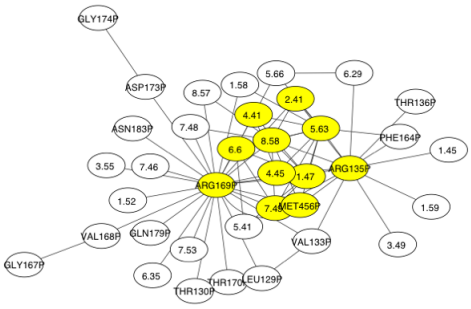
network is primary located near the G-protein binding pocket. Binding of allosteric ligand stabilizes the G-protein coupled site and hence the receptor. Because the mutual information between two torsion angles in allosteric binding pocket and G-protein coupling site is greater than self-entropies of the residues it forms the core of the dynamic allosteric exchange.Fig. 4.7d

Other mutants ASN419D, SER201Y identifies the possible hotspots in the muscarinic receptor for the allosteric interaction. While the ASN419 identifies the ECL2 and extracellular TM7 residues sites as allosteric signatures, in SER201Y mutant the potential hotspot of allosteric site primarily composed of the residue in the intracellular regions of the receptor around the G-protein Coupling site Fig. 4.7c, 4.7b.

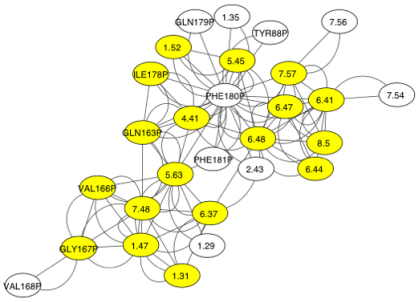
These pairwise couplings was used to identify putative allosteric sites in receptor also highlights the differential cluster arises in the mutant and how the correlations between residue changes in mutant clusters. While the residues involved in the allosteric pipelines in wild type agonist and modulator bound receptor is distinctly different compared with the other three mentioned mutatnts, they all involve pipeline connecting set of residues from allosteric bindsite to the connector region to the G-protein coupling site . In the wild type active state, the extracellular domain forms formed by ECL2-TM3 recognizes the ligand-modulator. We also observed that in the wild type the allosteric communication near the connector region involve TM5 region Arg211$^{5.63}$ that allosterically communicates with TM6-TM7 via PHE412$^{6.6}$ and NPxxY motif ASN436$^{7.49}$ which leads to distortion and significant displacement of TM3-TM6 helix in the receptor. Similar set of allosteric communication was identified in previous trajectory of muscarinic, which observed communication network involving TM3-TM5 and TM6-TM7 residues. Similar to our identified communication network pathway, beta-2 adrenergic receptor also have allosteric network found involving TYR219$^{5.58}$ and MET279$^{6.41}$, and PRO211$^{5.50}$ILE121$^{3.40}$. . Besides extra-

cellular regions, ARG211$^{5.63}$ and ARG387$^{6.35}$, VAL111$^{3.4}$ forms a connector residue in our allosteric communication pipeline that receives the signal from the orthosteric site. Thus, our results correlate with previous simulation results. Also, site-directed mutagenesis study reveals that mutation that the interaction of Y206$^{5.58}$ with TM6 motif is critical for G-protein selectivity [43].

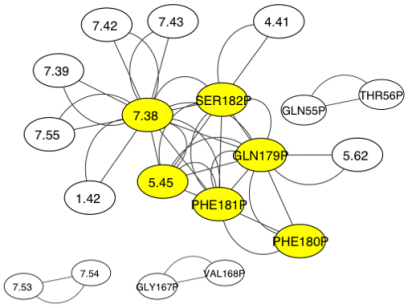
Comparing the changes brought in the inner G-protein coupling site by the above three mutants suggest the allosteric pertubation by drug molecule on allosteric binding site could help in potential regulation of receptor biogenesis.



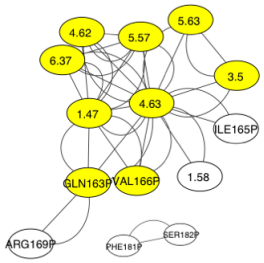
(a) 4MQT



(b) SER210Y



(c) ASN419D



(d) THR190A

Figure 4.7: Mutual Information network for ligand bounded muscarinic receptor and in mutants

Figure shows mutual information network (A) Receptor bounded to ligand- modulator. (B) SER210Y mutant (C) ASN419D mutant. (D) THR190A

4.4.4 Conclusion

The allosteric communication pathways lead to coordinated motion between ligand binding site and distant regions of the G-protein complexes. These pathways are highly degenerate and few nodes termed as critical node that occur at inter-community junctions is responsible for information flow within the receptor complex. The protein residue nodes that are part of major number of suboptimal paths between G-protein coupling site and ligand binding site of receptor are predicted to play critical role in allostery.

The community network analysis based on correlation residue motion and network based on mutual information between the residue pair facilitated coarse-graining of protein network and help towards understanding allosteric bias in muscarinic receptor. These approaches can be used to compare muscarinic receptor with mutants which are topologically similar to further understand function reorganization and information flow in complex. Our studies of the dynamical networks in GPCRs provides several metrics for comparing the allosteric signaling in GPCRs and in mutant systems.

On comparing the the allosteric network of mutants and wild type we found that there is significant difference between the allosteric bounded form and the mutants. While in the wild type, the allosteric modulation is via intracellular region of TM2-TM4-ICL2 and H8-TM5-TM7 near the G-protein coupling site, in the mutants the information flow varies ranging from network primarily near the G-protein binding pocket. In THR190A, regions of ECL2 and extracellular TM7 residues in ASN419D to intracellular part around the G-protein Coupling site in SER201Y. Our studies also reveal that ligand induce changes in the receptor are primarily carried our via residue 2.62 ,2.64, $W250^{4.57}$, $ALA^{5.47}$, 6.51, 6.54, 6.58, 7.34. Besides these residues in-

teractions between receptor and ligand.modulator also involve ECL-2 loops residues, 3.36,3.37, 7.38. These above mentioned residues, we termed as allosteric signatures towards allosteric bias as they are responsible for allosteric bias in the signaling.

Our community network analysis explained the changes in observed binding affinity of mutants compared to wild-type. As reported ASN419 when mutated to Aspartate have higher affinity towards allosteric ligand than asparagine is also via our network analysis where network 65% reorganizes itself while in THR190A mutant little decrease in the binding affinity of mutants towards allosteric modulator network 50% reorganizes and SER210Y where mutant results in 15 fold increase in binding affinity towards allosteric ligand and modulator has 60% reorganization. Our finding also led us conclude that mutation in the receptor lead to wide range of changes in the allosteric communication compared to the wild type receptor. There is decrease in correlated residue motion which are involved in wild type allosteric pathway and new communication pathways emerges in the mutants.

Chapter 5

Conclusions

Allosteric modulation of GPCR is key towards developing novel therapeutic agent that will help in controlling biological process associated with GPCRs [38]. Due to the ease of access of allosteric site compared to the orthosteric site it offers itself as great prospect for potential dug target with increase specificity and by binding to less conserved part of receptor interface [12]. Due to largely existing problems with GPCR crystallization, in-silico structure approach towards understanding allostery makes drug development an existing challenge [81]. Challenges in crystallization of GPCRs complexed with agonist have created opportunity for computational approaches using currently available recent structures, to understand allostery and identify druggable site that can modulate receptor binding and modulate signaling . In this work, we are focused to understand allostery in GPCRs and to map the residues in the receptor that are critical for information flow in the receptor. Many of the allosteric modulators plays role in activation without the direct contact with G-protein site or orthosteric binding site. These allosteric modulators could affect activation is by interacting with the residues that mediate allosteric communication between the orthosteric binding site and the G-protein coupling site [12]. Mapping of these allosteric communication

pathways in GPCRs would provide insights into allostery and drug discovery.

Molecular dynamic simulations,in which system follows Newton's equation of motion has been used to study interactions between atoms and it can help in constructing physical models using molecular mechanics force fields [53]. MD simulation provides unique strength of describing of bio molecules on timescales that can range from the femtosecond (fs) to the millisecond (ms) and provide atom-scale view for both the biomolecules and the solvent. MD simulation over the last two decades has helped in development of multiscale models for complex chemical systems and been critical in structural biology and protein science by providing computational microscope for proteins and their complexes. Other approaches to study allostery involves focusing on vibrations of protein around minimized static conformation of receptor, usually often through coarse-grained approach of normal mode analysis. Normal mode analysis is an approach to calculate effects of perturbation on allostery and analyze residue clusters critical for allosteric process [2]. These effects can be evaluated by studying structures trapped in different states of multiple possible conformation [2].

Protein structure, its topology and dynamics of proteins has been described using network paradigm, extensively [2, 67, 45]. The intramolecular non-covalent interactions in a protein collectively represented in the form of a network and these network can be crucial in determining physical communication and studying strength of interaction. Network is represented where the residues are the nodes of the is connected by edges that depend on their interaction strength. Several methods have been used to identify networks and paths of communication in receptor ensembles, collected through molecular dynamic simulations. Metrics like linear correlation and mutual information between the residue pairs to estimate interaction strength and to trace long distant communication in protein has been coupled to analyze these thermodynamic ensemble. Approach like Principal component analysis (PCA) is also

helpful in yielding a set of eigen vectors to represent non-redundant set of motions from the MD trajectory for a set of selected atoms. PCA helps in determine the observed equilibrium conformational fluctuations resemble the fluctuation between during allosteric fluctuation of receptor.

Our study with Cannabinoid focused on understanding the ligand specific interaction in CB1 . CB1 is used as allosteric models to understand signaling and activation mechanisms of GPCRs. We employed molecular dynamic simulation coupled with network analysis to reveal different ligands can interact with G-protein coupling site using different allosteric pathway. Our community network analysis was derived from running molecular dynamic simulation on CB1 revealed dynamic allosteric network. Studying these network elucidated correlated motion between G-protein Coupling site and ligand binding site in CB1.

Coupled with previous experimental studies our study also incorporated normal mode analysis and Molecular dynamic simulation to highlights dynamic allosteric communication pathways and sub networks of residues that are critical for the presence of receptor activity in CB1.

Our studies with normal mode in CB1 highlights the interaction involving L175$^{5.51}$, L174$^{5.50}$ and Y182$^{5.58}$ with G83$^{3.31}$ and interaction of Y182$^{5.58}$ and M183$^{5.59}$ with multiple TM3 contacts and interaction of TM4 residues L140$^{4.61}$, P139$^{4.6}$ C126$^{4.47}$ with TM3 connector regions as crucial local interaction. These interactions are translated into larger helical movement and most likely responsible for receptor biogenesis.

Based on our findings, residues from TM3 (3.50, 3.53, 3.54), TM5(5.61, 5.64, 5.65), TM6(6.26, 6.29, 6.32, 6.33, 6.36, 6.37, 6.40) form the communication pathway to transmit allosteric signal between allosteric binding site and G-protein coupling site in the CB1 receptor. These sub networks and communication pathway can be further crucial to understand allosteric biases in CB1. Having identified a series of potential

allosteric communication, this work will serve as a springboard for structure and fragment-based lead identification methods. Virtual screening of existing drug-like compound libraries for potential high-affinity ligands for allosteric binding pocket and G-protein coupling site can be assayed for identifying binding and allosteric activity.

Other approaches can include designing of novel compounds using the poses of docked probe molecules and performing fragment-based techniques to look for potential therapeutic agent. Probe molecule can be grown into high-affinity small molecules that have potential to interact with further GPCR residues for maximum specificity.

Availability of muscarinic receptor crystal structures bounded to agonist, inverse agonist and agonist-modulator has provided support to understand allostery and allosteric interaction in GPCR as well explore communication pathway that trigger variable signaling response in the same receptor. Our study using the structural information from muscarinic receptor in bounded and unbounded sate revealed allosteric pathway. we analyzed molecular dynamic simulations, and the cross correlation of residue motions to explore the interaction of receptors with allosteric ligand and modulator LY2119620 and how binding modulate essential structural interactions that are critical for the allosteric mechanism. Our finding suggest theses allosteric communication involve rearrangements of flexible ECL2 and the aromatic face to face interaction hydrophobic region at the modulator and hydrophobic interaction of transmembrane 2 residues near allosteric site, ionic interactions near the G-protein coupling site , including residues in the TM5 and the TM6 helices in muscarinic receptor , and in the TM3 in re as well as hydrogen bonding between the Tm6 residues with agonist and the TM3-TM6-TM6 residues adjacent to the agonist binding site.

The entropically driven modulator binding enhances the inter domain motion in TM3-TM6-TM7 of the receptor. During the receptor activation intracellular region

of the receptor-ligand complex remaining significantly more closed compared to extracellular region while in apo-form these motions are found to be significantly much less. These extracellular binding of ligand induced motion in GPCR results in protein motions partially promoting the change in molecular switch is mainly responsible for allosteric transition from inactive state to the allosterically active receptor. Using configuration entropy coupled with hierarchical clusters we the proposed allosteric signatures provides information that is crucial for the design of small molecules. These molecules can further can further interact with allosteric ligands and can modulates the allosteric signal propagation of GRPCR. These allosteric communication is mediated via conformation changes and reorganization of contacts among the residues in receptor. Although orthosteric binding have been potential drug target G-protein Coupled Receptor (GPCR), ease of targeting allosteric sites in GPCRS makes allosteric modulation by ligands in these sites interesting candidates for better receptor selectivity and improved binding affinity.

To identify allosteric ligand-selective signaling bias and site in receptor that contribute towards allosteric modulation receptor , we studied role of mutants in orthosteric and allosteric site in M2 muscarinic acetylcholine receptors. Our studies combined atomistic molecular dynamic simulation coupled with analysis of change in torsion angles and correlation between residue pair. These metric not only helped in estimation configuration entropy of thermodynamic ensemble in different receptor state but also helped in allosteric hotspot that are responsible for allosteric modulation. We elucidated the allosteric communication pathways in receptor is primary a coordinated correlated motion between binding site of ligand/modulator and regions in G-protein coupling site that are distant from these allosteric bindings. These communication pipelines are highly degenerate and few nodes termed as critical node that occur at inter-community junctions is responsible for information flow within the

receptor complex. Our finding identified single amino acid residues, such as Arg169 of ECL2 loop, R211$^{5.63}$, F412$^{6.6}$, I435$^{7.48}$ and K491.58 plays crucial part in the allosteric modulation and forms allosteric communication pipeline in the receptor. Our finding also suggest the critical nodes participate in the number of the sub-optimal paths between the ligand binding site and G-protein coupling site of receptor and predicted to play critical role in allostery.

Our studies of the dynamical networks in GPCRs provides several metrics for comparing the allosteric signaling in GPCRs and in mutant systems. Community network like approach compares muscarinic receptor with mutants which are topologically similar to further understand function reorganization and information flow in complex. The community network analysis based on correlation residue motion and allosteric network based on mutual information between the residue pair provides a coarse-grained view of the network. We used community network analysis as one of the metric to develop understanding allosteric bias in muscarinic receptor. Mutational studies on muscarinic receptor led us to conclude that mutation of functional residues lead to wide range of changes in the allosteric communication compared to the wild type receptor. Mutation of ASN419, which when mutated to Aspartate have higher affinity towards allosteric ligand than asparagine lead to reorganization of network and 65% of network reorganizes itself. While in THR190A mutant which has been reported to cause little decrease in the binding affinity of mutants also reorganizes 50%. SER210Y mutant which has been reported to cause 15 fold increase in binding affinity has 60% reorganization in community network analysis.

Our study thus not only helps in understanding activation mechanisms of GPCRs but also explores role of allosteric clusters that are responsible for modulation in GPCRs. Our studies can further be used to design of allosteric pharmacophores with novel activity that can control receptor biogensis and play crucial role

in GPCR pharmacology. Identifying allosteric pipelines, will help in allosteric drug design and will help in knitting structural and dynamical information on GPCRs with understanding of their function.

Bibliography

[1] Hans C Andersen. Rattle: A velocity version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52(1):24–34, 1983.

[2] Ivet Bahar, Timothy R. Lezon, Ahmet Bakan, and Indira H. Shrivastava. Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. *Chemical Reviews*, 110:1463–1497, 2010.

[3] Juan A Ballesteros and Harel Weinstein. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in g protein-coupled receptors. 1995.

[4] Supriyo Bhattacharya and Nagarajan Vaidehi. Differences in allosteric communication pipelines in the inactive and active states of a gpcr. *Biophysical journal*, 107(2):422–434, 2014.

[5] Supriyo Bhattacharya and Nagarajan Vaidehi. Mapping allosteric communication pipelines in gpcrs from microsecond timescale molecular dynamics simulations. *Biophysical Journal*, 106(2):635a, 2014.

[6] Michael P Bokoch, Yaozhong Zou, Søren GF Rasmussen, Corey W Liu, Rie Nygaard, Daniel M Rosenbaum, Juan José Fung, Hee-Jung Choi, Foon Sun Thian, Tong Sun Kobilka, et al. Ligand-specific regulation of the extracellular surface of a g-protein-coupled receptor. *Nature*, 463(7277):108–112, 2010.

[7] Bernard R Brooks, Charles L Brooks, Alexander D MacKerell, Lennart Nilsson, Robert J Petrella, Benoît Roux, Youngdo Won, Georgios Archontis, Christian Bartels, Stefan Boresch, et al. Charmm: the biomolecular simulation program. *Journal of computational chemistry*, 30(10):1545–1614, 2009.

[8] Martin Caffrey. Crystallizing membrane proteins for structure determination: use of lipidic mesophases. *Annual review of biophysics*, 38:29–51, 2009.

[9] Pil Seok Chae, Søren GF Rasmussen, Rohini R Rana, Kamil Gotfryd, Richa Chandra, Michael A Goren, Andrew C Kruse, Shailika Nurva, Claus J Loland,

Yves Pierre, et al. Maltose-neopentyl glycol (mng) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *nAture methods*, 7(12):1003–1008, 2010.

[10] Vadim Cherezov, Daniel M Rosenbaum, Michael a Hanson, Søren G F Rasmussen, Foon Sun Thian, Tong Sun Kobilka, Hee-Jung Choi, Peter Kuhn, William I Weis, Brian K Kobilka, and Raymond C Stevens. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science (New York, N.Y.)*, 318(5854):1258–65, 2007.

[11] Ellen YT Chien, Wei Liu, Qiang Zhao, Vsevolod Katritch, Gye Won Han, Michael A Hanson, Lei Shi, Amy Hauck Newman, Jonathan A Javitch, Vadim Cherezov, et al. Structure of the human dopamine d3 receptor in complex with a d2/d3 selective antagonist. *Science*, 330(6007):1091–1095, 2010.

[12] Arthur Christopoulos and Terry Kenakin. G protein-coupled receptor allosterism and complexing. *Pharmacological reviews*, 54(2):323–374, 2002.

[13] Audrey Claing, Stéphane A Laporte, Marc G Caron, and Robert J Lefkowitz. Endocytosis of g protein-coupled receptors: roles of g protein-coupled receptor kinases and ß-arrestin proteins. *Progress in neurobiology*, 66(2):61–79, 2002.

[14] P Jeffrey Conn, Arthur Christopoulos, and Craig W Lindsley. Allosteric modulators of gpcrs: a novel approach for the treatment of cns disorders. *Nature reviews Drug discovery*, 8(1):41–54, 2009.

[15] Dirk G Demuth and Areles Molleman. Cannabinoid signalling. *Life sciences*, 78(6):549–563, 2006.

[16] Xavier Deupi, Patricia Edwards, Ankita Singhal, Benjamin Nickle, Daniel Oprian, Gebhard Schertler, and Jörg Standfuss. Stabilized g protein binding site in the structure of constitutively active metarhodopsin-ii. *Proceedings of the National Academy of Sciences*, 109(1):119–124, 2012.

[17] Xavier Deupi, Jörg Standfuss, and Gebhard Schertler. Conserved activation pathways in g-protein-coupled receptors. *Biochemical Society Transactions*, 40(2):383–388, 2012.

[18] Gary J Doherty and Harvey T McMahon. Mechanisms of endocytosis. *Annual review of biochemistry*, 78:857–902, 2009.

[19] Brian N Dominy and Charles L Brooks. Development of a generalized born model parametrization for proteins and nucleic acids. *The Journal of Physical Chemistry B*, 103(18):3765–3773, 1999.

[20] Andrew S Doré, Nathan Robertson, James C Errey, Irene Ng, Kaspar Hollenstein, Ben Tehan, Edward Hurrell, Kirstie Bennett, Miles Congreve, Francesca Magnani, et al. Structure of the adenosine a 2a receptor in complex with zm241385 and the xanthines xac and caffeine. *Structure*, 19(9):1283–1293, 2011.

[21] R O Dror, D H Arlow, P Maragakis, T J Mildorf, A C Pan, H F Xu, D W Borhani, and D E Shaw. Activation mechanism of the beta(2)-adrenergic receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 108:18684–18689, 2011.

[22] Ron O Dror, Daniel H Arlow, Paul Maragakis, Thomas J Mildorf, Albert C Pan, Huafeng Xu, David W Borhani, and David E Shaw. Activation mechanism of the β2-adrenergic receptor. *Proceedings of the National Academy of Sciences*, 108(46):18684–18689, 2011.

[23] Tilman Flock, Charles NJ Ravarani, Dawei Sun, AJ Venkatakrishnan, Melis Kayikci, Christopher G Tate, Dmitry B Veprintsev, and M Madan Babu. Universal allosteric mechanism for g [alpha] activation by gpcrs. *Nature*, 524(7564):173–179, 2015.

[24] Robert Fredriksson, Malin C Lagerström, Lars-Gustav Lundin, and Helgi B Schiöth. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Molecular pharmacology*, 63(6):1256–72, 2003.

[25] Zhan-Guo Gao and Kenneth A Jacobson. Allosteric modulation and functional selectivity of g protein-coupled receptors. *Drug Discovery Today: Technologies*, 10(2):e237–e243, 2013.

[26] U Gether. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. *Endocrine reviews*, 21:90–113, 2000.

[27] Ann L Gnagey, Margaret Seidenberg, and John Ellis. Site-directed mutagenesis reveals two epitopes involved in the subtype selectivity of the allosteric interactions of gallamine at muscarinic acetylcholine receptors. *Molecular pharmacology*, 56(6):1245–1253, 1999.

[28] N Go, T Noguti, and T Nishikawa. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proceedings of the National Academy of Sciences of the United States of America*, 80:3696–3700, 1983.

[29] Kazuko Haga, Andrew C. Kruse, Hidetsugu Asada, Takami Yurugi-Kobayashi, Mitsunori Shiroishi, Cheng Zhang, William I. Weis, Tetsuji Okada, Brian K. Kobilka, Tatsuya Haga, and Takuya Kobayashi. Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist, 2012.

[30] Freddy Heitz, James A Holzwarth, Jean-Pierre Gies, Rebecca M Pruss, Suzanne Trumpp-Kallmeyer, Marcel F Hibert, and Chantal Guenet. Site-directed mutagenesis of the putative human muscarinic m 2 receptor binding site. *European journal of pharmacology*, 380(2):183–195, 1999.

[31] Klaus Peter Hofmann, Patrick Scheerer, Peter W Hildebrand, Hui-Woog Choe, Jung Hee Park, Martin Heck, and Oliver P Ernst. Ag protein-coupled receptor at work: the rhodopsin model. *Trends in biochemical sciences*, 34(11):540–552, 2009.

[32] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[33] Alem W Kahsai, Kunhong Xiao, Sudarshan Rajagopal, Seungkirl Ahn, Arun K Shukla, Jinpeng Sun, Terrence G Oas, and Robert J Lefkowitz. Multiple ligand-specific conformations of the β2-adrenergic receptor. *Nature chemical biology*, 7(10):692–700, 2011.

[34] Thomas W. Klein, Cathy Newton, and Herman Friedman. Cannabinoid receptors and immunity, 1998.

[35] EG Knox and MS Bartlett. The detection of space-time interactions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 13(1):25–30, 1964.

[36] Lotte Bjerre Knudsen, Dan Kiel, Min Teng, Carsten Behrens, Dilip Bhumralkar, János T Kodra, Jens J Holst, Claus B Jeppesen, Michael D Johnson, Johannes Cornelis de Jong, et al. Small-molecule agonists for the glucagon-like peptide 1 receptor. *Proceedings of the National Academy of Sciences*, 104(3):937–942, 2007.

[37] Andrew C Kruse, Jianxin Hu, Albert C Pan, Daniel H Arlow, Daniel M Rosenbaum, Erica Rosemond, Hillary F Green, Tong Liu, Pil Seok Chae, Ron O Dror, et al. Structure and dynamics of the m3 muscarinic acetylcholine receptor. *Nature*, 482(7386):552–556, 2012.

[38] Andrew C. Kruse, Jianxin Hu, Albert C. Pan, Daniel H. Arlow, Daniel M. Rosenbaum, Erica Rosemond, Hillary F. Green, Tong Liu, Pil Seok Chae, Ron O. Dror, David E. Shaw, William I. Weis, Jürgen Wess, and Brian K. Kobilka. Structure and dynamics of the M3 muscarinic acetylcholine receptor, 2012.

[39] Andrew C Kruse, Brian K Kobilka, Dinesh Gautam, Patrick M Sexton, Arthur Christopoulos, and Jürgen Wess. Muscarinic acetylcholine receptors: novel opportunities for drug development. *Nature Reviews Drug Discovery*, 13(7):549–560, 2014.

[40] Stéphane A Laporte, Robert H Oakley, Jie Zhang, Jason A Holt, Stephen SG Ferguson, Marc G Caron, and Larry S Barak. The β2-adrenergic receptor/βarrestin complex recruits the clathrin adaptor ap-2 during endocytosis. *Proceedings of the National Academy of Sciences*, 96(7):3712–3717, 1999.

[41] Robert J Lefkowitz, Kristen L Pierce, and Louis M Luttrell. Dancing with different partners: protein kinase a phosphorylation of seven membrane-spanning receptors regulates their g protein-coupling specificity. *Molecular pharmacology*, 62(5):971–974, 2002.

[42] Jeffrey J Liu, Reto Horst, Vsevolod Katritch, Raymond C Stevens, and Kurt Wüthrich. Biased signaling pathways in β2-adrenergic receptor characterized by 19f-nmr. *Science*, 335(6072):1106–1110, 2012.

[43] Jie Liu, Bruce R Conklin, Nathalie Blin, June Yun, and Jurgen Wess. Identification of a receptor/g-protein contact site critical for signaling specificity and g-protein activation. *Proceedings of the National Academy of Sciences*, 92(25):11642–11646, 1995.

[44] Craig C Malbon. Frizzleds: new members of the superfamily of g-protein-coupled receptors. *Frontiers in bioscience: a journal and virtual library*, 9:1048–1058, 2004.

[45] Christopher L McClendon, Gregory Friedland, David L Mobley, Homeira Amirkhani, and Matthew P Jacobson. Quantifying correlations between allosteric sites in thermodynamic ensembles. *Journal of chemical theory and computation*, 5(9):2486–2502, 2009.

[46] Yinglong Miao, Sara E Nichols, Paul M Gasper, Vincent T Metzger, and J Andrew McCammon. Activation and dynamic network of the m2 muscarinic receptor. *Proceedings of the National Academy of Sciences*, 110(27):10982–10987, 2013.

[47] Rouslan Moukhametzianov, Manfred Burghammer, Patricia C Edwards, Sebastien Petitdemange, Dimitri Popov, Maikel Fransen, Gregory McMullan, Gebhard FX Schertler, and Christian Riekel. Protein crystallography with a micrometre-sized synchrotron-radiation beam. *Acta Crystallographica Section D: Biological Crystallography*, 64(2):158–166, 2008.

[48] Rouslan Moukhametzianov, Tony Warne, Patricia C Edwards, Maria J Serrano-Vega, Andrew GW Leslie, Christopher G Tate, and Gebhard FX Schertler. Two distinct conformations of helix 6 observed in antagonist-bound structures of a β1-adrenergic receptor. *Proceedings of the National Academy of Sciences*, 108(20):8228–8232, 2011.

[49] Midori Murakami and Tsutomu Kouyama. Crystal structure of squid rhodopsin. *Nature*, 453(7193):363–367, 2008.

[50] JOJI M. OTAKI and STUART FIRESTEIN. Length analyses of mammalian g-protein-coupled receptors. *Journal of Theoretical Biology*, 211(2):77 – 100, 2001.

[51] Uberto Pagotto, Giovanni Marsicano, Daniela Cota, Beat Lutz, and Renato Pasquali. The emerging role of the endocannabinoid system in endocrine regulation and energy balance. *Endocrine reviews*, 27(1):73–100, 2006.

[52] Krzysztof Palczewski, Takashi Kumasaka, Tetsuya Hori, Craig A Behnke, Hiroyuki Motoshima, Brian A Fox, Isolde Le Trong, David C Teller, Tetsuji Okada, Ronald E Stenkamp, et al. Crystal structure of rhodopsin: Ag protein-coupled receptor. *science*, 289(5480):739–745, 2000.

[53] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16):1781–1802, 2005.

[54] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16):1781–1802, 2005.

[55] Rita Rahmeh, Marjorie Damian, Martin Cottet, Hélene Orcel, Christiane Mendre, Thierry Durroux, K Shivaji Sharma, Grégory Durand, Bernard Pucci, Eric Trinquet, et al. Structural insights into biased g protein-coupled receptor signaling revealed by fluorescence spectroscopy. *Proceedings of the National Academy of Sciences*, 109(17):6733–6738, 2012.

[56] William M Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.

[57] Søren G F Rasmussen, Hee-Jung Choi, Daniel M Rosenbaum, Tong Sun Kobilka, Foon Sun Thian, Patricia C Edwards, Manfred Burghammer, Venkata R P Ratnala, Ruslan Sanishvili, Robert F Fischetti, Gebhard F X Schertler, William I Weis, and Brian K Kobilka. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature*, 450:383–387, 2007.

[58] Søren GF Rasmussen, Brian T DeVree, Yaozhong Zou, Andrew C Kruse, Ka Young Chung, Tong Sun Kobilka, Foon Sun Thian, Pil Seok Chae, Els Pardon, Diane Calinski, et al. Crystal structure of the [bgr] 2 adrenergic receptor-gs protein complex. *Nature*, 477(7366):549–555, 2011.

[59] Søren GF Rasmussen, Brian T DeVree, Yaozhong Zou, Andrew C Kruse, Ka Young Chung, Tong Sun Kobilka, Foon Sun Thian, Pil Seok Chae, Els Pardon, Diane Calinski, et al. Crystal structure of the [bgr] 2 adrenergic receptor-gs protein complex. *Nature*, 477(7366):549–555, 2011.

[60] M H Rhee, M Bayewitch, T Avidor-Reiss, R Levy, and Z Vogel. Cannabinoid receptor activation differentially regulates the various adenylyl cyclase isozymes. *Journal of neurochemistry*, 71:1525–1534, 1998.

[61] Daniel M Rosenbaum, Søren G F Rasmussen, and Brian K Kobilka. The structure and function of G-protein-coupled receptors. *Nature*, 459:356–363, 2009.

[62] Daniel M Rosenbaum, Cheng Zhang, Joseph A Lyons, Ralph Holl, David Aragao, Daniel H Arlow, Søren GF Rasmussen, Hee-Jung Choi, Brian T DeVree, Roger K Sunahara, et al. Structure and function of an irreversible agonist-[bgr] 2 adrenoceptor complex. *Nature*, 469(7329):236–240, 2011.

[63] Mette M Rosenkilde, Tau Benned-Jensen, Thomas M Frimurer, and Thue W Schwartz. The minor binding pocket: a major player in 7tm receptor activation. *Trends in pharmacological sciences*, 31(12):567–574, 2010.

[64] Roberto Sanchez and Andrej Sali. Evaluation of comparative protein structure modeling by modeller-3. *Proteins Structure Function and Genetics*, 29(s 1):50–58, 1997.

[65] Torsten Schöneberg, Günter Schultz, and Thomas Gudermann. Structural basis of g protein-coupled receptor function. *Molecular and cellular endocrinology*, 151(1):181–193, 1999.

[66] Caitlin E. Scott, Ravinder Abrol, Kwang H. Ahn, Debra A. Kendall, and William A. Goddard. Molecular basis for dramatic changes in cannabinoid CB1 G protein-coupled receptor activation upon single and double point mutations. *Protein Science*, 22:101–113, 2013.

[67] Anurag Sethi, John Eargle, Alexis A Black, and Zaida Luthey-Schulten. Dynamical networks in trna: protein complexes. *Proceedings of the National Academy of Sciences*, 106(16):6620–6625, 2009.

[68] Joong Youn Shim. Transmembrane helical domain of the cannabinoid CB1 receptor. *Biophysical Journal*, 96:3251–3262, 2009.

[69] Joong Youn Shim, James Rudd, and Tomas T. Ding. Distinct second extracellular loop structures of the brain cannabinoid CB 1 receptor: Implication in ligand binding and receptor function. *Proteins: Structure, Function and Bioinformatics*, 79:581–597, 2011.

[70] Tatsuro Shimamura, Mitsunori Shiroishi, Simone Weyand, Hirokazu Tsujimoto, Graeme Winter, Vsevolod Katritch, Ruben Abagyan, Vadim Cherezov, Wei Liu, Gye Won Han, et al. Structure of the human histamine h1 receptor complex with doxepin. *Nature*, 475(7354):65–70, 2011.

[71] Fabian Sievers, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular systems biology*, 7(1), 2011.

[72] Rebecca Stadel, Kwang H Ahn, and Debra A Kendall. The cannabinoid type-1 receptor carboxyl-terminus, more than just a tail. *Journal of neurochemistry*, 117(1):1–18, 2011.

[73] Jörg Standfuss, Patricia C Edwards, Aaron DAntona, Maikel Fransen, Guifu Xie, Daniel D Oprian, and Gebhard FX Schertler. The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature*, 471(7340):656–660, 2011.

[74] R. E. Stenkamp, D. C. Teller, and K. Palczewski. Crystal structure of rhodopsin: A G-protein-coupled receptor, 2002.

[75] Lesley A Stevenson, Morag Easson, Martin R Price, Gemma L Baillie, Richard Goodwin, Ad Mclean, Lorraine Mcintosh, Gillian Goodwin, Glenn Walker, Paul Westwood, Julia Marrs, Fiona Thomson, Phillip Cowley, Arthur Christopoulos, Roger G Pertwee, and Ruth A Ross. Allosteric Modulation of the Cannabinoid CB 1 Receptor. 68(5):1484–1495, 2005.

[76] Gürol M Süel, Steve W Lockless, Mark A Wall, and Rama Ranganathan. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Structural & Molecular Biology*, 10(1):59–69, 2003.

[77] Aaron A Thompson, Wei Liu, Eugene Chun, Vsevolod Katritch, Huixian Wu, Eyal Vardy, Xi-Ping Huang, Claudio Trapella, Remo Guerrini, Girolamo Calo, et al. Structure of the nociceptin/orphanin fq receptor in complex with a peptide mimetic. *Nature*, 485(7398):395–399, 2012.

[78] Abdulnour Y Toukmaji and John A Board. Ewald summation techniques in perspective: a survey. *Computer physics communications*, 95(2):73–92, 1996.

[79] Hamiyet Unal and Sadashiva S Karnik. Domain coupling in gpcrs: the engine for induced conformational changes. *Trends in pharmacological sciences*, 33(2):79–88, 2012.

[80] Nagarajan Vaidehi and Terry Kenakin. The role of conformational ensembles of seven transmembrane receptors in functional selectivity. *Current opinion in pharmacology*, 10(6):775–781, 2010.

[81] a J Venkatakrishnan, Xavier Deupi, Guillaume Lebon, Christopher G Tate, Gebhard F Schertler, and M Madan Babu. Molecular signatures of G-protein-coupled receptors. *Nature*, 494:185–94, 2013.

[82] Tony Warne, Patricia C Edwards, Andrew GW Leslie, and Christopher G Tate. Crystal structures of a stabilized β 1-adrenoceptor bound to the biased agonists bucindolol and carvedilol. *Structure*, 20(5):841–849, 2012.

[83] Tony Warne, Rouslan Moukhametzianov, Jillian G Baker, Rony Nehmé, Patricia C Edwards, Andrew GW Leslie, Gebhard FX Schertler, and Christopher G Tate. The structural basis for agonist and partial agonist action on a [bgr] 1-adrenergic receptor. *Nature*, 469(7329):241–244, 2011.

[84] Tony Warne, Maria J Serrano-Vega, Jillian G Baker, Rouslan Moukhametzianov, Patricia C Edwards, Richard Henderson, Andrew GW Leslie, Christopher G Tate, and Gebhard FX Schertler. Structure of a &bgr; 1-adrenergic g-protein-coupled receptor. *Nature*, 454(7203):486–491, 2008.

[85] Mark Wheatley, Denise Wootten, Matthew T Conner, John Simms, R Kendrick, RT Logan, David R Poyner, and James Barwell. Lifting the lid on gpcrs: the role of extracellular loops. *British journal of pharmacology*, 165(6):1688–1703, 2012.

[86] Xin-Qiu Yao, Rabia Malik, Nicholas W Griggs, Lars Skjaerven, John R Traynor, Sivaraj Sivaramakrishnan, and Barry J Grant. Dynamic coupling and allosteric networks in the alpha subunit of heterotrimeric g proteins. *Journal of Biological Chemistry*, pages jbc–M115, 2015.

[87] Shuguang Yuan, Slawomir Filipek, Krzysztof Palczewski, and Horst Vogel. Activation of g-protein-coupled receptors correlates with the formation of a continuous internal water pathway. *Nature communications*, 5, 2014.

[88] Yongqin Zhang, Alice L Rodriguez, and P Jeffrey Conn. Allosteric potentiators of metabotropic glutamate receptor subtype 5 have differential effects on different signaling pathways in cortical astrocytes. *Journal of Pharmacology and Experimental Therapeutics*, 315(3):1212–1219, 2005.