

# **Structural Mechanism of Activation of G Protein-Coupled Receptors**

**Argha Mitra**

*Laboratory of Chemical Biology, Institute of Biochemistry,*

*Biological Research Centre, Szeged*

Supervisor

**Dr. Attila Borics**

*Doctoral School of Theoretical Medicine*

*Albert Szent-Györgyi Medical School, University of Szeged*

**Summary of the Ph.D. Thesis**

Szeged

2021

## Introduction

The pharmaceutical relevance of G protein-coupled receptors (GPCRs) is clearly demonstrated by the fact that, according to the latest reports, approximately 34% of FDA approved marketed drugs target members of this receptor family. Consequently, enormous efforts were invested for the three-dimensional structural characterization of GPCRs. As a result, a remarkable amount of detailed structural data is available for many types and classes of GPCRs, which provide a solid basis for the in-depth analysis of their structural mechanism of function.

From the structural perspective, GPCRs share similar architecture in their transmembrane (TM) domains and possess high sequential and structural diversity in their extra- and intracellular loops (ECLs and ICLs) and the extracellular (N-terminal) and cytosolic (C-terminal) domains. These variable domains are proposed to be responsible for ligand and signaling protein specificity, whereas the transmembrane domain controls the transmission of external signal to the intracellular surface of the protein.

The recently published x-ray crystallographic, NMR spectroscopic and cryo-electron microscopic data provided invaluable information about the specificities of the active and inactive states of the receptor. However, the dynamic mechanism of transition between these states is unattainable with the common experimental techniques. Molecular dynamics (MD) simulations of GPCRs share a common theme by confirming the structural ensemble of active, inactive and intermediate states. The populations of these states are shifted upon ligand binding, depending on the functional properties of the ligand. The 5<sup>th</sup>, 6<sup>th</sup> and 7<sup>th</sup> transmembrane helices (TM5, TM6 and TM7, respectively) have been emphasized for their foremost interplay in signal transduction. The flexible domains, missing from experimental structures of GPCRs, was not explicitly taken into account in GPCR simulations. The application of single-component membrane bilayers or very simplistic membrane models dominates the practice of MD simulation studies of GPCRs. However, the need for more accurate membrane representation is clearly suggested by several recent reports.

It is also still unclear how ligands of similar physico-chemical properties exert diverse cellular response and how exactly the signal is transmitted from the binding pocket to the G protein coupling interface.

### **Aims and Scope**

In order to respond to the most recent challenge of rational drug design and to develop high-affinity, high-efficacy, and functionally selective GPCR ligands, a quantitative model of the activation mechanism is needed. This necessitates the introduction of new perspectives. In this thesis a new model is proposed. According to this model the activation of GPCRs is accompanied by a shift of macroscopic polarization in a shielded central duct of the TM domain, initiated by ligand binding and propagated to the intracellular G protein-binding interface. This hypothesis was formulated on the basis of the abundance of highly conserved polar/ionizable species which were indicated to be key participants of the activation mechanism.

The main objective of the study presented in this thesis is to test the feasibility of the above hypothesis and to reveal further key determinants of the mechanism of activation of GPCRs by utilizing large scale MD simulations in which the physiological state and environment of the receptor is represented as accurately as possible. Evidently, MD simulations employing fixed point charge force fields cannot reveal exact or quantitative details of processes involving charge shift. Nevertheless, statistical analysis of the dynamic motions of key molecular parts could provide convincing support for the interplay of the above-mentioned polar species.

With regard to the above considerations our specific aims are as follows

1. To build stable and realistic MD simulation systems, which represent the native state and environment of the  $\mu$ -opioid (MOP),  $\beta_2$ -adrenergic ( $\beta_2$ AR) and type 1 cannabinoid (CB1) receptors, representative class A GPCRs addressed in this study.

- 1.1 To model and include the flexible N- and C-terminal domains and ICL3 to study the dynamics of the full sequence receptors.

- 1.2 To obtain plausible ligand-receptor complex structures via molecular docking if experimental data of such complexes are unavailable.

- 1.3 To apply native conditions for protein function through the explicit inclusion of all post translational modifications (PTMs), such as phosphorylation, glycosylation, and lipidation.
- 1.4 To approximate native membrane environment through the application of explicit caveolar (raft-like) membrane composition during simulations.
2. To utilize large scale MD simulations in order to investigate specific structural transitions of GPCRs in the presence/absence of ligands and intracellular signaling proteins that correlate with the activation state of the receptor.
3. To design and employ specific analysis methods or strategies in order to identify further potential parameters of GPCR signaling which could rationally extend the current state-of-the-art.

## **Materials and Methods**

The active and inactive state crystallographic structures of the MOP bound to endomorphin 2 (EM2) and the  $G_i$  protein complex or  $\beta$ -arrestin-2 were subjected to MD simulations. Furthermore, the active and inactive  $\beta_2$ AR bound to epinephrine and the  $G_s$  protein or  $\beta$ -arrestin-2, and the type1 cannabinoid receptor (CB1) complexed with 2-arachidonoylglycerol (2-AG) and the  $G_i$  protein or  $\beta$ -arrestin-2 were investigated.

The missing loops and mutated residues were retrieved utilizing loop modeling and the Swiss-PdbViewer program, respectively. The missing N- and C-terminal domains were modeled by performing folding simulations using GROMACS 5.1.4 and 2018.3 program packages.

Due to the lack of experimentally derived relative orientation of EM2 and 2-AG bound to MOP and CB1, respectively, docking was used to predict their binding mode. The final poses which were used as starting structures of MD simulations were selected by comparison with experimental data of ligands of similar functional properties.

All the PTMs, such as phosphorylation on serine and threonine residues, N-linked complex-type glycosylation on asparagines and lipidations on cysteine residues were attached to the corresponding sites in CHARMM-GUI.

In order to represent the native, raft-like membrane environment a membrane slab comprising neutral lipids (POPC, POPE), negatively charged lipids (POPI, POPS), sphingomyelin (PSM), ganglioside (GM3) and cholesterol was prepared using the CHARMM-GUI membrane builder. Receptor complexes were embedded into the membrane and solvated with 0.15M NaCl / TIP3P water solution.

The resultant complex systems were energy minimized thoroughly by performing consecutive steepest descent and conjugate gradient minimizations. Then the systems were subjected to a six-step equilibration protocol with gradually decreasing positional restraints on the heavy atoms of the proteins and membrane constituents.

Ten, eleven and four independent simulation systems were prepared for MOP,  $\beta_2$ AR and CB1 receptors, respectively. 1 $\mu$ s long all-atom MD simulations were performed to investigate the conformational dynamics of the receptor complexes using the GROMACS 5.1.4 and 2018.3 program packages, and the CHARMM36 force field.

Domain movements of the agonist-bound and ligand-free receptors were assessed on the basis of the root mean square deviations (RMSD) of backbone atoms. Specific interactions such as hydrogen bonds (threshold distance and angle: 0.35 nm and 60.0 degrees) and salt bridges (threshold distance and angle: 0.4 nm and 90.0 degrees) were calculated. The evolution of secondary structure was determined using the DSSP method. The penetration of Na<sup>+</sup> ions into the allosteric Na<sup>+</sup> binding site, D<sup>2.50</sup>, was checked with a distance threshold of 0.4 nm. The structural parameters of helical segments were also calculated and compared. Dynamic cross correlation matrix (DCCM) analysis was performed for the amino acids of the TM domain. The threshold of assignment of correlation was red color intensity corresponding to > 0.65 value of the correlation coefficient for all the receptors. All of the above analysis of was performed in the analysis suite of GROMACS.

Membrane thickness and area per lipid head group values were calculated using the FATSLiM program. Multiple sequence alignment was performed using the Clustal Omega program and results were compared using the Jalview software.

## Results and Discussion

### *Membrane Properties*

Rapid vertical contraction of the phospholipid bilayer was observed at the beginning of the equilibration, while lateral contraction was more gradual. In the first few nanoseconds of the production simulation, a slight vertical expansion was observed which stabilized at  $\sim 4.30$  nm thickness of the membrane. Lateral contraction of the membrane continued in the initial production phase until stabilization at approximately  $0.49$  nm<sup>2</sup> after 30 ns. The initial imbalance of membrane during the production simulation was attributed to the relaxation of the after positional restraints of the proteins were removed. No artificial trends reflecting this initial membrane imbalance were observed in the results.

### *N- and C-terminal Domain Dynamics*

The highly flexible intrinsically disordered N- and C-terminal domains were generally excluded from the experimental structures as well as MD simulation studies. We have modeled these domains using repeated folding simulations to create approximate structures. Convergent results were obtained in the folding simulations of MOP,  $\beta_2$ AR and CB1 in terms of the evolution of backbone RMSD, radius of gyration, number of H-bonds and secondary structure. Although it cannot be guaranteed that the folding of these domains was correct and complete in the available time frame, their effect on TM dynamics, primarily exerted by their mass was satisfactorily taken into account by using these approximate structures. Partial unfolding was often observed during production simulations, but periodic images of the terminal domains did not contact each other.

### *Simulation System Integrity*

To address the integrity of simulation systems, dissociation of molecular components was checked in all MD trajectories. No dissociation of ligand or nucleotide (GTP or GDP) or notable displacement of protein components was observed for the MOP and CB1 receptors. The systems including the  $\beta_2$ AR were found to be intact for most simulations as no dissociation of epinephrine or the nucleotide or notable relative

displacement of macromolecular components were observed. Epinephrine dissociated from the orthosteric binding pocket of  $\beta_2$ AR on two occasions. Epinephrine was first observed to depart from the binding pocket during the simulation of the inactive,  $G_s$  protein-bound  $\beta_2$ AR after 600 ns, and then after 100 ns during the 3<sup>rd</sup> replica simulation of the active  $\beta_2$ AR bound to the  $G_s$  protein. Interestingly, in the 2<sup>nd</sup> replica simulation of the active  $G_s$  protein-bound  $\beta_2$ AR, the ligand took an opposite orientation in the binding pocket compared to the X-ray crystallographic structure of the  $\beta_2$ AR-epinephrine complex. These simulations were not excluded from analysis but the results were interpreted accordingly.

### ***Allosteric Na<sup>+</sup> Binding***

Na<sup>+</sup> penetration into the allosteric binding site (D<sup>2.50</sup>) was not observed while agonists were bound to the orthosteric binding site of the MOP,  $\beta_2$ AR and CB1 receptors. Conversely, Na<sup>+</sup> rapidly localized at the allosteric site in the absence of bound agonists, following the transition of the receptor to an intermediate structural state. This suggested that the entry of Na<sup>+</sup> was only possible through the orthosteric binding site. In the case of the MOP bound simultaneously by EM2 and Na<sup>+</sup> the frequency of close contacts between Na<sup>+</sup> and conserved polar/amphipathic residues further down towards the intracellular surface (N328<sup>7.45</sup>, N332<sup>7.49</sup>, and Y336<sup>7.53</sup>) have increased significantly.

### ***TM6 Dynamics***

The disposition of TM6 of the transmembrane domain is a general indicator of the activation state of the class A GPCRs as it was shown by previous experimental results. Atomic displacement analysis of this TM helix of the EM2-bound MOP receptor revealed, that TM6 assumed intermediate conformations during simulations with minor changes from the corresponding starting structures. TM6 of the ligand-free receptor, however, underwent much larger changes, suggesting remarkable stabilizing effect of the agonist. Similar results were observed for  $\beta_2$ AR, but the extent of TM6 displacement was more comparable in the agonist-bound and ligand-free states of the active receptor possibly due to the lower binding affinity, smaller molecular size and reduced stabilizing effect of epinephrine compared to that of EM2 bound to the MOP. The above trends for

the disposition of TM6 were followed by the CB1 receptor too, although no reference simulations have yet been done for the ligand-free CB1.

### ***Loop Dynamics and Intermolecular Interactions***

The first, second and third intracellular loops (ICL1, ICL2 and ICL3, respectively as well as the cytosolic helix (H8) form an interaction interface for G proteins on the intracellular surface of GPCRs. The evolution of secondary structure of these loops was assessed using the DSSP method to reveal their structural preferences with regard to the bound signaling protein and the activation state of the receptor. The secondary structures of ICL1, ICL3 and H8 were maintained throughout the simulations of MOP,  $\beta_2$ AR and CB1 receptors with respect to their starting structures. ICL2 of the active MOP bound to EM2 was found to adopt stable  $\alpha$ -helical structure when bound to  $\beta$ -arrestin-2 and partially unfolded upon interaction with the  $G_i$  protein, independent of the state of the receptor. During the 1  $\mu$ s simulation of the epinephrine-bound  $\beta_2$ AR- $G_s$  protein complex, partial unfolding of ICL2 was observed for inactive state and  $\alpha$ -helical structure was maintained in the active state. The structural adaptation of ICL2 could be related to the specificity for intracellular proteins and corresponding signaling pathways.

The analysis of specific intermolecular interactions between the studied GPCRs and the  $G\alpha$  subunit revealed, that the  $\alpha$ -helical conformation of ICL2 in the active, ligand-free states facilitates strong interactions between the receptors and  $G\alpha$ . This supports the hypothesis of pre-coupled GPCR-G protein complexes in the absence of a bound ligand.

### ***Intramolecular Interactions***

The role of intramolecular interactions between conserved TM motifs was described in detail in previous reports. The interaction between R165<sup>3.50</sup> of the E/DRY motif and D340<sup>8.47</sup> of H8 was not reported previously, but it was found to be present in the  $G_i$  protein-bound active state of MOP. In the  $\beta_2$ AR and CB1 simulation systems no such interaction was indicated suggesting that this interaction between R<sup>3.50</sup> and D<sup>8.47</sup> is a specific property of the MOP receptor.



### ***Transmembrane Helix Properties***

According to the hypothesis of this study, the activation of GPCRs is accompanied by a shift of macroscopic polarization in a shielded central duct of the TM domain. The inherent dipole moment of  $\alpha$ -helices can promote various conduction processes along the helix axis. Generally, the more ordered an  $\alpha$ -helical segment is, the higher its dipole moment. The TM7 was found to be the most ordered among TM helices of the active MOP bound to the  $G_i$  protein and EM2. Furthermore, the helicity of TM7 was closest to ideal when the receptor was  $G_i$  protein-bound, and least ideal when it was complexed by beta-arrestin-2, presumably providing TM7 with the highest dipole moment in the  $G_i$  protein-bound state, compared to all other receptor states. This indirectly supports the involvement of the inherent dipole moment of TM7 in the activation mechanism.

Such a proposed role of TM7 was not corroborated by the results obtained for  $\beta_2$ AR and CB1 receptors, suggesting that the above-described potential role of TM7 is a specific property of the MOP receptor.

### ***NPxxY Motif Dynamics***

The conserved NPxxY motif in the junction of TM7 and H8 is associated with stabilizing intermediate conformations in ligand-free class A GPCRs that facilitate G protein insertion. The surprisingly high disorder of TM7 observed in  $\beta_2$ AR prompted the analysis of the disposition of this helical segment to see if the dynamics of this motif shows any correlation with the activation state of the receptor or with the presence of agonist and/or intracellular proteins. Thus, the displacement of this motif have been calculated in the  $\beta_2$ AR simulation systems. Interestingly, relatively large dispositions ( $\sim 0.4$  nm RMSD of backbone atoms) of the NPxxY motif were found to coincide with the intense concerted dynamics of the second segment of a series of conserved polar amino acid side chains located close to the intracellular surface of the receptor. This suggests, that the elevated mobility of the NPxxY motif could be associated either with receptor activation or constitutional activity.

### ***Correlated Side-chain Motions in the Transmembrane Domain***

Dynamic cross-correlation matrix (DCCM) analysis revealed that the orthosteric binding pocket is connected to the intracellular surface of the MOP receptor through a channel of polar amino acid residues, of which motions are highly correlated. Such concerted motions were observed only for the active receptor-G<sub>i</sub> protein complex and none for the other reference systems. Residues of the above identified polar signaling channel are located mostly on the TM7, in the inner core of the TM helical bundle, shielded from the surrounding membrane environment. Analysis of the individual dynamics of these specific side chains revealed that the observed movements are small and mostly occur without the transition between rotameric states. However, small conformational changes of polar or charged species could result in significant alterations in the local electronic structures, which could then propagate along the molecule. The involvement of protons and water molecules in the transmission of signal could also be presumed, since the orthosteric binding pocket and the G protein-binding interface is connected through a hydrated pathway and water molecules were observed to exchange rapidly between the internal cavities of the TM domain during simulations. In light of the above results, the initial hypothesis of this study proposing a parallel change of macroscopic polarization in the TM domain during GPCR activation could be deemed plausible. Impaired G protein signaling, or elevation of constitutational activity, was observed for mutant receptors previously, where residues of the above of mentioned polar signaling channel were replaced, while receptor activity was preserved in double mutants, where the net charge of channel residues was kept intact.

The high degree of conservation of polar signaling channel residues suggests that the above theory could be extended to other class A GPCRs. DCCM analysis of the MD simulation trajectories of  $\beta_2$ AR and CB1 receptors provided corroborating results. Although, several differences between the channel residues of the three receptors were also identified.

A specific feature of the polar signaling channels of  $\beta_2$ AR and CB1 receptors is that they could be subdivided into two segments. The first segment spans the orthosteric and allosteric binding pockets and the NPxxY motif, while the second segment shares the

last residue of the NPxxY motif and includes the tip of TM7 and three H8 residues. The rationale behind this subdivision is given by the NPxxY disposition data obtained for  $\beta_2$ AR, presented earlier. The relatively large, approximately 0.4 nm (RMSD) disposition of the NPxxY motif coincided with intense concerted motions in the second segment of the polar signaling channel. Such movements were observed in inactive states and in the absence of ligand, whereas the full sequence of correlated motions was incomplete in those systems. This suggests that the concerted dynamics of this second segment indicate the constitutive activity of the  $\beta_2$ AR and CB1 receptors. This was supported by that correlated motions of this second segment were decoupled upon  $\beta$ -arrestin-2 binding to  $\beta_2$ AR and CB1 and/or if epinephrine was bound to  $\beta_2$ AR in the wrong relative orientation, presumably stabilizing a conformational state of the receptor that is inappropriate for signaling.

## Summary of Findings

1. The structure of ICL2 has an important role in signaling protein specificity of the MOP and  $\beta_2$ AR receptors.
2. Localization  $\text{Na}^+$  ions in the allosteric binding site of class A GPCRs takes place from the extracellular side. The allosteric site is only accessible through the orthosteric binding pocket and its entrance is blocked by a bound ligand, whereas cytosolic access to the TM domain is closed by the bound intracellular signaling proteins.
3. The precise orientation of the agonist is crucial to initiate signal transduction of GPCRs. Minor displacements from the functional binding mode could arrest receptor function through stabilizing a conformational state that is inappropriate for signaling.
4. The orthosteric binding pocket of the three GPCRs studied here is connected to the intracellular G protein-binding interface through a channel of conserved polar amino acid residues, of which motions are highly correlated. This polar signaling channel may assist the shift of macroscopic polarization in a shielded central duct of the TM domain, leading to G protein activation.
5. The polar signaling channel could be subdivided into two segments. The first segment starting from the orthosteric binding pocket could be responsible for the transmission of the agonist induced signal, whereas the second segment, closer to the intracellular surface, could be related to the constitutive activity of the receptor.

### Publications related to the thesis

**1. Mitra A, Sarkar A, Szabó MR and Borics A.**

Correlated motions of conserved polar motifs lay out a plausible mechanism of G Protein-Coupled Receptor activation. *Biomolecules*, 11 (5), 670 (2021).

IF: 4.879 [Q2]

**2. Mitra A, Sarkar A and Borics A.**

Universal properties and specificities of the  $\beta_2$ -adrenergic receptor-G<sub>s</sub> protein complex activation mechanism revealed by all-atom molecular dynamics simulations. *Int J Mol Sci*, 22(19), 10423 (2021).

IF: 5.923 [Q1]

### Publications not related to the thesis

**1. Yeggoni DP, Rachamalla A, Dubey S, Mitra A and Subramanyam R.**

Probing the interaction mechanism of menthol with blood plasma proteins and its cytotoxicity activities. *J Biomol Struct Dyn*, 36 (2), 465-474. (2018) IF: 3.39

### Conference / Poster Abstracts

**1. Mitra A, Sarkar A, Szabó MR and Borics A.**

Mu-opioid receptor activation mechanism: a new perspective and possible role of polar amino acids in the receptor core.

Advances and Challenges in Biomolecular Simulations; EMBO Workshop, Brno, Czech Republic, 2021

**2. Mitra A, Sarkar A and Borics A.**

Universal properties and specificities of the  $\beta_2$ -adrenergic receptor-G<sub>s</sub> protein complex activation mechanism.

Scientific Meeting of Peptide Chemistry Working Committee; Balatonszemes, Hungary, 2021.

**3. Tóth BI, Janssens A, Lisztes E, Kelemen B, Herczeg M, Nagy MT, Kövér K, Mitra A, Borics A, Tamás B, Voets T.**

Rapamycin activates TRPM8 via novel mechanisms.

The 4<sup>th</sup> International TRP Meeting (Hybrid Edition); Leuven, Belgium, 2021.

**4. Borics A, Mitra A, Sarkar A, Szabó, MR.**

A plausible structural mechanism of G protein-coupled receptor activation.

49<sup>th</sup> Membrane Transport Conference; Sümeg, Hungary, 2019.

**5. Mitra A, Szabó, MR and Borics, A.**

Building simulation-systems to study mu-opioid receptor activation.

Scientific Meeting of Peptide Chemistry Working Committee; Balatonszemes, Hungary, 2017.