

GPCRmd - A web platform for collection,
visualization and analysis of molecular
dynamics data for G protein-coupled
receptors

Bridging the gap between receptor dynamics and
receptor functionality

Ismael Rodríguez Espigares

TESI DOCTORAL UPF / ANY 2016

DIRECTORS DE LA TESI

Dr. Ferran Sanz Carreras i Dra. Jana Selent

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA
SALUT



This thesis is dedicated to my family, friends and workmates who put up with me and supported me for more than four years.

Acknowledgements

I acknowledge Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya (2015 FI_B00145) and for its financial support.

We acknowledge financial support from the Instituto de Salud Carlos III FEDER (CP12/03139 and PI15/00460).

I would like to thank my supervisors Prof Jana Selent and Prof Ferran Sanz, for her guidance and aid during the development of this thesis. I would also like to thank my workmates at GPCR Drug Discovery group and at Research Programme on Biomedical Informatics. I would like to also thank GPUGRID volunteers for providing us the resources for part of our computations.

Abstract

In this thesis, we present GPCRmd, an online repository with a submission system and visualization platform specifically designed for storing and providing easy access to molecular dynamics (MD) data of G protein-coupled receptor (GPCR). This database stores MD trajectories and necessary metadata (e.g. force-field, simulation software, integration time-step) for posterior analysis, ensuring data reproducibility and integrity. Importantly, we demonstrate the usefulness of implemented analysis tools in two case studies related to GPCR signaling bias and membrane-induced GPCR modulation. These tools enabled us to detect important structural rearrangement in the initial phase of β -arrestin signaling in the δ -opioid receptor. In addition, we captured relevant molecular mechanisms which are responsible for cholesterol-induced modulation of the 5-HT_{2A} receptor.

Resum

En aquesta tesi, presentem el GPCRmd, un repositori en línia amb un sistema de dipòsit i una plataforma de visualització dissenyats específicament per oferir l'emmagatzematge i un fàcil accés a dades dinàmica molecular (en anglès MD) de receptors acoblats a proteïnes G (en anglès GPCRs). Aquesta base de dades emmagatzema trajectòries de MD i les metadades (per exemple, el programari de simulació, el camp de força o el temps d'integració) necessàries per a una anàlisi posterior, garantint la reproductibilitat i la integritat de les dades. És important destacar la utilitat de les eines d'anàlisi implementades en dos estudis de cas relacionats amb el biaix de senyalització de les GPCRs i la modulació de les GPCRs induïda per membrana. Aquestes eines ens han permès detectar una important reordenació estructural en la fase inicial de senyalització per β -arrestina en el receptor δ -opioide. A més, hem capturat el mecanismes moleculars rellevants responsables de la modulació induïda per colesterol del receptor 5-HT_{2A}.

Prologue

This thesis focuses on the development of a new online platform specialized on G protein-coupled receptors (GPCRs) that aims to make molecular dynamics (MD) data easily accessible to the public while ensuring data integrity and reproducibility.

GPCRs are the largest superfamily of human cell receptors and they play an important role in modulating physiological processes spanning from sensory and neurological to endocrine ones.

In consequence, more than 30% of all FDA-approved drugs act through GPCRs. Furthermore, as most GPCRs within human biology are still unexploited in therapies, they have a huge potential for future drug development. In this respect, a better understanding of GPCR functionality is of high value to the scientific community.

In this regard, MD techniques have become extremely useful to provide insights on the mechanisms that modulate GPCR functionality by complementing the findings obtained by experimental means. In addition, the increase in computational power experimented in the past years suggests that in a close future the large amount of available MD data would require integration, classification, indexing and new tools capable of the analysis of this data in its integrated form.

For this reason, we implemented GPCRmd with features the following feature: 1) submission tools which aid GPCRmd contributors in their task of providing new simulation data, 2) a simulation browsing by system components such as proteins and small molecules, and 3) several tools that allow to perform basic analyses on the data submitted to our database.

Then, in order to demonstrate the usefulness of the implemented analysis tools for unraveling GPCR functionality, we focused in two case studies related to GPCR signaling bias and membrane-induced GPCR modulation.

The study of δ -opioid receptor (δ -OR) in complex with naltrindole (a δ -OR antagonist) and a mutant variant where the latter ligand becomes a strong antagonist lead to the detection of an important

structural rearrangement in the initial phase of β -arrestin signaling in this receptor.

Regarding the case study of membrane-induced effects on GPCR, we captured relevant molecular mechanisms which are responsible for cholesterol-induced modulation of the 5-HT_{2A} receptor.

All in all, we proved that MD techniques and our analysis tools were useful to provide new information about GPCR functionally that might generate, for example through drug rational design, new safer and drugs with a higher efficacy.

Contents

Acknowledgements	v
Abstract	vii
Prologue	ix
List of Figures	xiii
1. Introduction	1
1.1. G protein-coupled receptors (GPCRs)	1
1.2. Conformational states in GPCRs	3
1.3. GPCR activation allosteric network	4
1.4. Molecular Dynamics	7
2. Objectives	9
2.1. GPCRmd server - a platform for MD GPCRs submission, visualization and analysis	9
2.2. Elucidating GPCR functionality in terms of signaling bias and membrane effects	10
3. GPCRmd: a GPCR specialized Molecular Dynamics database and analysis tool	11
4. Rapid network rearrangements drive the initial phase of β- arrestin signaling in the δ-opioid receptor	65
5. Membrane cholesterol effect on the 5-HT_{2A} receptor: Insights into the lipid-induced modulation of an antipsychotic drug target	99
6. Discussion	135
6.1. The GPCRmd database	135

6.2. β -arrestin signaling	137
6.3. Membrane effects.....	137
7. Conclusions	139
8. List of communications	141
9. Appendix: other publications	143
9.1. Membrane cholesterol access into a G-protein-coupled receptor	143
Bibliography.....	145

List of Figures

Figure 1.1 Structure of 5-hydroxytryptamine 1B (5-HT _{1B}) receptor in complex with ergotamine	2
Figure 1.2 GPCR micro-switches.....	6

Chapter 1

Introduction

1.1. G protein-coupled receptors (GPCRs)

G protein-coupled receptors (GPCRs), also known as seven-transmembrane domain receptors are the largest superfamily of human cell receptors. They play an important role in modulating physiological processes spanning from sensory and neurological to endocrine ones.^{1,2} Thereby, their function is to transduce extracellular signals into multiple cell signaling pathways by responding either to physical (e.g. light) or chemical (e.g. small molecules, hormones, peptides) extracellular stimuli.

Despite the functional diversity of this superfamily, they have a conserved fold consisting of seven transmembrane helices (Figure 1.1). The helical domains are responsible for the signal transduction across the membrane which is commonly initiated by ligand binding deep inside the receptor.

Another characteristic property of GPCRs transmembrane helices is that they contain several highly conserved groups residues that are highly important for receptor functionality. For this reason, these conserved microdomains are groups of residues commonly numbered according to the Ballesteros & Weinstein nomenclature. This numbering consist on a first digit that identifies the TM helix number (from 1 to 7) and a second one that identifies the position of the residue in the TM with respect to the most conserved residue in that TM (arbitrary defined as 50) assigned from sequence alignment sequentially from N to C-terminus.³ Another available numbering

scheme is GPCRdb⁴ general numbering. The different GPCR proteins usually present insertions and deletions that shift Ballesteros–Weinstein numbering. This issue is solved by GPCRdb numbering as it is based on structural alignments and adds an extra number at the end of the conserved Ballesteros–Weinstein numbering in case of an insertion. In addition, the dot is replaced by “x” in order to avoid confusion (e.g. 2x50).⁵

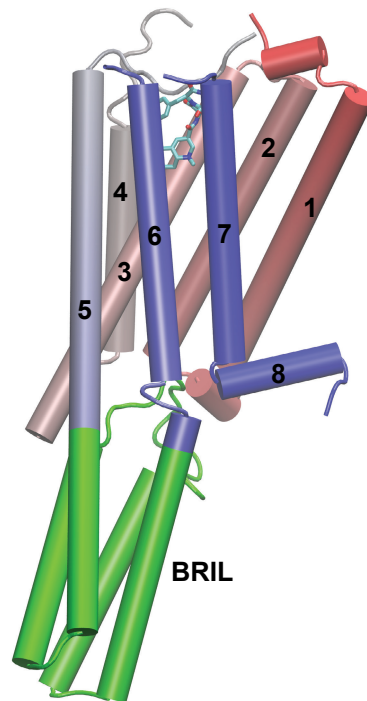


Figure 1.1 Structure of 5-hydroxytryptamine 1B (5-HT_{1B}) receptor in complex with ergotamine. PDB ID: 4IAR.⁶ Cartoon color represents residue number in red grey scale going from red (N-terminal) to blue (C-ter). BRIL protein function is shown in green cartoon and ergotamine in licorice. Numbers represent transmembrane helices (TM) 1-7 and helix 8.

GPCRs can be classified based on sequence homology into classes A to F⁷: Glutamate (A), Rhodopsin (B), Adhesion (C) and

Frizzled(F)/Taste2. An alternative classification system is GRAFS⁸: Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2 and Secretin for human receptors.

The class with the largest number of different receptors is Class A accounting for nearly 85% of the GPCR genes.²

Because of their abundance and their participation in multiple physiological processes, more than 30% of all FDA-approved drugs act through GPCRs.^{9,10} Furthermore, as most GPCRs within human biology are still unexploited in therapies¹¹, they have a huge potential for future drug development. In this respect, a better understanding of GPCR functionality is of high value to the scientific community.¹²

1.2. Conformational states in GPCRs

Important aspects of GPCR conformational space have been discussed in a previous review by us.¹³

In that review, we stated that GPCRs conformational flexibility is crucial for its functionality, and consequently, they present a large number of receptor conformational states. The sum of this conformational states describes a conformational universe that expands from inactive to active states. The fact that different active states exists explains why receptors are able to couple to different signaling proteins. One relevant case of this property of GPCRs is biased-signaling by β -arrestin instead of activating or inactivating a G protein. Biased agonism (or functional selectivity) of GPCRs is related to their ability to preferentially elicit a subset of responses of all possible receptor responses.¹⁴ This observation has opened new avenues for producing more efficacious and safer drugs.¹⁵

Furthermore, there is increasing evidence that inactive and active states are separated by intermediate transient states,^{16,17} but their function in receptor activation and functionally is yet not well understood.

Exploring this conformational universe and deciphering the mechanisms which controls the transitions between these states is a major challenge. However, this would give the opportunity to modulate receptor signaling with grate specificity and to discover drugs safer therapeutic profiles.^{18,19}

One way to search for evidence of the existence of these different conformational states is by looking at the different crystal structures that have been obtained by protein engineering and crystallographic techniques. However, most of these structures capture low-energetic states, which are predominately inactive states. Hence, only few GPCRs are crystallized in states different to the inactive one. β 2-adrenergic receptor (β 2AR) in complex with the stimulatory G protein (Gs) for adenylyl cyclase (PDB ID: 3SN6)²⁰ and rhodopsin in complex with visual arrestin (PDB ID: 4ZWJ).²¹ Based on those structures, we know that, when G proteins and arrestin insert themselves into the intracellular side of the receptor, some conformational changes are observed: 1) TM6 of the β 2AR opens 14 Å in the intracellular region. 2) In arrestin-bound rhodopsin, this movement is less pronounced (around 10 Å). 3) A comparison between arrestin-bound rhodopsin to the active receptor C-terminal-coupled to G-transducin reveals additional conformational differences in TM1, TM4, TM5 and TM7 which might be related to signaling bias.²¹

The combination of these active crystal structures with other inactive and intermediate states that have been crystallized in complex with antagonists, partial agonists or biased agonists, give a sparse glance into different states of the conformational universe.

All in all, exploring the conformational space of GPCRs is crucial in order to understand GPCR activation process and its functionality. High resolution experimental techniques together with computational approaches such as molecular dynamics (MD) can be used to explore the conformational universe and provide obtain relevant information on how these proteins work. In fact, experiments have suggested already that transitions between different states involve conformational changes in micro-switches and the formation of an internal water channel as explained below.

1.3. GPCR activation allosteric network

In our last review,¹³ we also addressed the known structural insights regarding the GPCR activation allosteric network.

We summarized the principles of signal propagation through the receptor and across the membrane in class A GPCRs. The work by Rasmussen *et al*²² suggested that an allosteric network connects the intracellular and extracellular regions of the receptor. This allosteric network is partially conserved among several GPCR receptor types.²³ Further analysis of this network revealed, that it contains conserved structural elements, also known as micro-switches.^{24,25} Structural studies first using computational techniques and later crystallography and mutational assays, have revealed that changes in those elements correlate with the receptors activation state^{26,27}

In our last review,¹³ we summarized available mechanistic insights linked about the mentioned micro-switches:

(i) the “ionic lock” between Arg3.50 D/E3.49 and D/E6.30²⁴ (D/E)R(Y/M) motif (Figure 1, red VdW), which has been described as *closed* in the inactive conformations of some receptors. (ii) the hydrophobic arginine cage (Figure 1, yellow VdW) — conserved hydrophobic amino acids at positions 3.46²⁴ and 6.37²⁸ which restrains the conformation of Arg3.50; (iii) the NPxxYxF motif in TM7, responsible for the direct interaction of Tyr7.53 in TM7 with Phe7.60 in H8 (Figure 1, lime) and with the side chain and backbone (via water molecule) of Arg2.40 in TM2.^{29–31} Furthermore, the conformational state of Tyr7.53 has been related to the presence of a water channel and receptor activation.^{32,33}

We describe also (iv) the Rotamer Toggle Switch (Figure 1, purple) formed by clustered aromatic residues in TM6 which sense the binding of distinct ligand, and in response present an unified rearrangement³⁴. Changes in this switch, might translate downwards, towards the extracellular part of the GPCR, and regulate ionic lock formation.³⁵ The described group of aromatic residues in TM6 near Trp6.48 of the CWxP motif triggers a conformational transition of Trp6.48 from pointing towards TM7 in the inactive state to pointing towards TM5 in the active state.^{26,36,37} Finally, we end with (v) the “P-I-F” motif (Figure 1, pink), a hydrophobic connector switch comprised of Ile3.40, Pro5.50 and Phe6.44, suggested to link the agonist binding pocket with the intracellular domain.^{22,17}

As a final remark, we point out that the above-mentioned micro-switches conformations change leading into different GPCR activation states and that formation/disruption of a water channel that has been related to receptor activation.^{32,33,38}

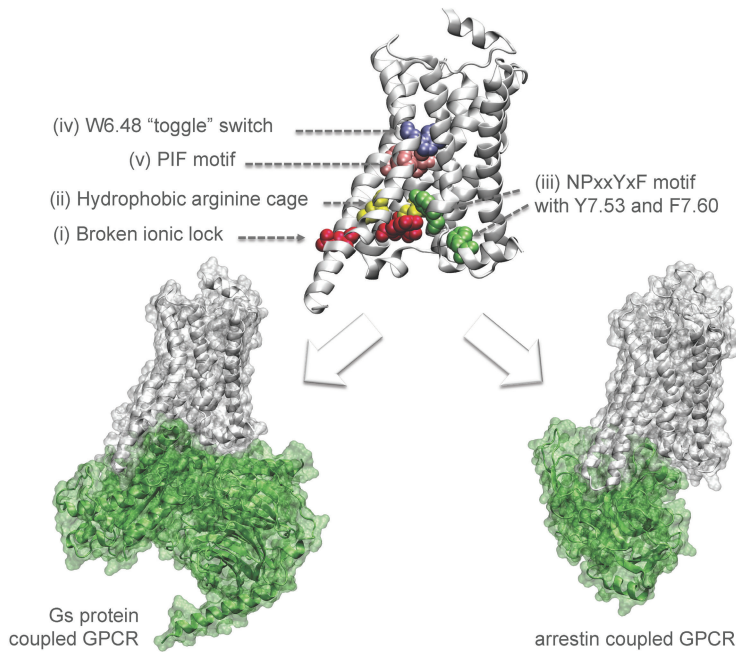


Figure 1.2 GPCR micro-switches. Highly conserved receptor regions have been reported to be structural indicators for distinct conformational receptor states that are related to the coupling of selected extracellular signaling transducers (e.g. Gs or arrestin). Reproduced from Rodríguez-Espigares et al.¹³

Highly conserved receptor regions that can also be considered micro-switches have been reported to be structural indicators for distinct conformational receptor states that are related to the coupling of selected extracellular signaling transducers (e.g. Gs or arrestin).

Regarding bias signalling, several computational and *in-vitro* studies and have shown that residues of TM 7, in particular Tyr7.53^{39,40} is key in transducing β -arrestin signalling.⁴¹⁻⁴⁴ Furthermore, class A GPCRs present an allosteric site for sodium ion next to Asp2.50. Experimental work suggested that the present of this allosteric site is connected to G protein/arrestin signaling bias.⁴⁵ The relevance in the study of bias signalling relies in the fact that can be used as new mechanism of action for safer and more efficacious drugs.⁴⁶

Finally, membrane composition is a factor that can influence GPCR signaling properties. Several experiments show that the function and/or ligand binding properties of serotonin receptors are modulated

by cholesterol concentration in the membrane⁴⁷, i.e. 5-HT_{1A}^{48,49}, 5-HT_{2A}⁵⁰, or 5-HT_{7A}⁵¹. However, it is still unknown if cholesterol modulates GPCR functionality by directly interaction with the receptor or indirectly by changing the membrane environment.^{52–55} These changes may lead to conformational changes that alter the dynamics of micro-switches and in the whole receptor. Thus, it is important also to study the membrane effects GPCRs to know they alter the GPCR conformational space.

1.4. Molecular Dynamics

Molecular Dynamics (MD) is a simulation method that allows the study of molecular and biomolecular processes, that is difficult to be observed using classical experimental techniques like X-ray crystallography or NMR.^{56,57}

In classical MD, trajectories of molecules and atoms are obtained by numerically solving classical equations of motion (Newton's equations). Interaction between particles is simplified by describing it through simple harmonic forces.

The potential energy of the ensemble only depends on particle positions. The bonded and non-bonded interactions between particles, are described in sets called “force-fields”. The parameters used in a force-field for computing forces and energies are usually obtained from quantum mechanics computations or fitting experimental data.

Some for the force-fields used for MD of membrane proteins include CHARMM^{58–60}, AMBER⁶¹, GROMOS⁶² and OPLS⁶³.

Regarding simulation software the most popular packages are CHARMM⁶⁴, AMBER⁶¹, NAMD⁶⁵, GROMACS⁶⁶, ACEMD⁶⁷ or LAMMPS.⁶⁸

MD like every experimental technique has its limitations: the simplification of molecular process to molecular classical mechanics might produce biased or artefacts during simulation. Also sufficient computational resources are needed to sample a simulate process enough, for the results to be relevant. However, many times the results obtained provide structural and kinetic insights that help understand molecular processes better.

Chapter 2

Objectives

2.1. GPCRmd server - a platform for MD GPCRs submission, visualization and analysis

Molecular dynamics (MD) simulations are becoming a popular tool to study the physiology and structure of GPCRs. However, there is no repository for GPCR MD trajectory which makes dynamics data available to researchers in a way that a simple query could retrieve a set of simulations of a specific ligand-GPCR complex. Such GPCR MD repository could help overcome problems of data reproducibility, integrity and availability which are often found for molecular dynamic studies.

Therefore, one of goals of this thesis is the design and implementation of the GPCRmd database, an online deposition repository for GPCR MD data which not only stores trajectories but also other files and metadata needed for ensuring integrity and reproducibility and, at the same time, allows and search of MD simulation by the molecular components present.

In addition, this thesis aims to provide a trajectory visualization platform together with several analysis tools for the GPCR MD analysis of the simulation stored in the database. This way we try to ensure the availability of the data to the maximum possible number of users, which is the ultimate objective of this platform.

2.2. Elucidating GPCR functionality in terms of signaling bias and membrane effects

GPCR molecular dynamics are often carried out to obtain insights into GPCR functionality which cannot be obtained through experimental techniques. These structural insights can serve to explain phenomena such as functional selectivity, or the influence of membrane composition on receptor properties. In this respect, the thesis focuses on the impact of cholesterol-induced membrane alterations on the dynamics of serotonin 5-hydroxytryptamine 2A receptor (5-HT_{2A}). In addition, the thesis studies the molecular determinants of β -arrestin signaling in the δ -opioid receptor. Ultimately, the obtained structural information can help design safer and more efficacious drugs through means of rational design.

Chapter 3

GPCRmd: a GPCR specialized Molecular Dynamics database and analysis tool

In this chapter, we present our work in form of a journal article:

I. Rodríguez-Espigares, M. Torrens-Fontanals, A. Varela-Rial, J. M. Ramírez-Anguita & J. Selent. GPCRmd: a GPCR specialized Molecular Dynamics database and analysis tool. (Manuscript in preparation)

Summary: Here, we present GPCRmd an online repository with a submission system and visualization platform specifically designed for GPCR molecular dynamics (MD). This database not only stores MD trajectories, it also includes necessary metadata (e. g. force-field simulation software, integration time-step) for posterior analysis after publication, ensuring data reproducibility and integrity.

In order to make this data easily browsable, information about the components that are present in MD simulation are stored and indexed. These components are classified as either protein or small-molecule (non-protein). Protein information includes name, sequence, references to UniprotKB entries and mutations. Small-molecules are stored at two levels of detail: first as chemical substance and second molecular entity. We describe the molecule with cheminformatics notations (SMILES and InChI) and link it to PubChem and ChEMBL identifiers. The combination of these two elements allows the user to browse intuitively for the desired simulations.

Furthermore, this platform includes an online visualization tool based on NGL which allows to play MD trajectories using a modern

internet browser software without need of external add-ons through WebGL technology. These visualization tool is complemented by a set of analysis tools that allow Root Mean Square Deviation and interatomic distances measurements and the study of receptor-ligand interactions and the hydrogen bond network.

Finally, as an example of application two sets of MD simulation data previously published is analyzed using the GPCRmd online tools (see also Chapter 4). In this work, I supervised the development of the platform together with my thesis supervisors, designed the submission system, simulation browser and server backend and implemented most of the submission application code.

GPCRmd: a GPCR specialized Molecular Dynamics database and analysis tool

Ismael Rodríguez-Espigares*, Mariona Torrens-Fontanals*, Alejandro Varela-Rial*, Juan Manuel Ramírez-Anguita* and Jana Selent*

Abstract

GPCRs are major targets for the pharmaceutical industry, being involved in many physiological processes and diseases. However, the underlying mechanisms responsible for their functionality and differential interactions with ligands and intracellular signaling proteins remain elusive. Molecular dynamics (MD) simulations are a promising approach for the understanding of such molecular processes, providing a temporal and structural resolution greater than what, currently, is achievable by experimental methods. In this context, a web platform dedicated to the diffusion of GPCR dynamics data can be of high interest to the scientific community.

GPCRmd is a database of GPCR molecular dynamics capable to foster data from all-over the world with the purpose to support and stimulate GPCR research and the discovery of new drugs. In this paper, we present the features of such database and its web-based platform for simulation submission and online visualization by the GPCRmd viewer, specially design for GPCR MD data.

* GPCR Drug Discovery group, Research Programme on Biomedical Informatics (GRIB), Universitat Pompeu Fabra (UPF)-Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain.

1. Introduction

G protein-coupled receptors (GPCRs), also known as are the largest superfamily of human cell receptors. They participate in multiple physiological processes spanning from sensory and neurological to endocrine function by transducing binding information of a broad spectrum of extracellular ligands into multiple cell signaling pathways.^{1,2}

Although there are many receptors in GPCR superfamily, they all share a common conserved tertiary structure: seven transmembrane helices, which are responsible for the signal transduction across the membrane, separated by intracellular and extracellular loops. These last ones contain part of the binding pockets for signaling molecules that target GPCRs.³ GPCRs are usually classified by sequence homology into the classes A to F,⁴ or with the alternative system GRAFS (Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2, Secretin) for human receptors⁵ (Table 1). The class with the largest number of different receptors is class A, which accounts for nearly 85% of the GPCR genes.

Because of their abundance, participation in multiple physiological processes, accessibility at the cell surface and druggability, more than 30% of all FDA-approved drugs act on a GPCR.^{6,7} However, most of them are still unexploited in therapies.⁸ Thus, this superfamily has a huge potential for future drug development, and a better understanding of GPCR functionality is of high value to the scientific community.

Table 1 GPCR classification by A-F and GRAFS systems.

A-F	GRAFS*
A	Rhodopsin Secretin
B	Adhesion
C	Glutamate
D	-
E	-
F	Frizzled
Other	Taste 2

*Note that GRAFS only includes human GPCRs, and classes D and E do not exist in humans.

Around 200 GPCR crystal structures in the Protein Data Bank (PDB) provide insights on the activation process of GPCRs and the mechanism of the signal transduction. However, crystal structures on its own can only provide snapshots of the behavior of the receptors while these proteins can undergo a whole universe conformational changes. Hence, additional information is needed to know more about the physiological relevance of the conformations obtained by crystallographic methods and for deciphering the underlying molecular and structural mechanisms responsible for many processes including signal transduction, allosteric modulation, functional selectivity, and constitutive activity. One way to obtain this necessary conformational data is through Molecular Dynamics (MD) simulations. These simulations provide a time dimension to the 3D conformational space that allows us to observe processes which otherwise would be difficult or impossible, such as protein folding or the transition between the different conformations of a protein.⁹ The progress that has been made in recent years in computer hardware and MD force field parameters definition, which characterizes the interactions among the particles involved, has allowed an increase on simulation time. This means that in the following years will be able to achieve enough sampling to study slow and complex processes such as GPCR activation.

Usually, GPCR MD data is scattered among supplementary data of various articles, online file repositories or even not available online. Such as in other fields, data integrity, availability and reproducibility is important also in molecular dynamics, but these principles are usually difficult to accomplish due to the different formats and standards, software and force fields available to generate MD data, and the large quantities of data that are obtained.

Here we present the GPCRmd database, an online deposition repository for GPCR all-atom force-field MD data which not only stores trajectories but also other files and metadata needed for ensuring integrity and reproducibility and, at the same time, allows trajectory visualization and provides several tools for GPCR MD analysis.

Other similar projects to this new database have been also developed in the past. Some examples of databases that have offered computed analysis data or analysis tools and MD trajectories are the MoDEL (Molecular Dynamics Extended Library),¹⁰ where the addition of membrane proteins is in progress; Dynameomics,¹¹ that contains trajectories of different unique protein folds and BiosymGrid,¹² a general MD deposition platform. More recent examples of MD deposition and analysis databases focused on non-protein biomolecules are Cyclo-lib¹³ and BIGNASim.¹⁴

The GPCRmd (freely-available at www.gpcrmd.org) offers new features respect some of these databases: 1) allows data deposition through a customized submission system for GPCR MD, 2) it is a specialized database, thus it can provide tools and analysis which are useful specifically for the study of GPCRs; 3) matches trajectories to their molecular components providing links to external databases such UniprotKB¹⁵, PubChem Compound¹⁶ or ChEMBL¹⁷, and 4) supplies visualization and analysis tools that only require a modern web browser software compatible with WebGL technology without need of plugins or external software. The latter is possible thanks to the NGL viewer, a MD visualization software created by A.S. Rose and P.W. Hildebrand.¹⁸

In this article, we are going to present the features of the GPCRmd database and the methods used for their development along with two study cases for the application of analysis tools of GPCRmd involving three different GPCR receptors: a comparison of interaction events in 5-hydroxytryptamine receptor 1B (5-HT_{1B}), and 5-hydroxytryptamine receptor 2B (5-HT_{2B}) and the analysis of interaction events associated with δ -type opioid receptor (δ -OR).

2. Methods and features

a) Web platform and database software

GPCRmd database web interface have been developed using Django Web Framework (v1.9) based on Python (v3.4) and JavaScript libraries jQuery 1.9 and jQuery UI 1.11.2. PostgreSQL 9.3 is used as database engine.¹⁹

The structure of the database is based on main five objects: 1) *protein* objects identified by their sequence and their relationship with UniprotKB entries, 2) molecular entities (*molecule* object in GPCRmd) identified by an InChI²⁰ generated with forced explicit hydrogen connectivity, 3) chemical species (*compound*) identified by standard InChI, 4) crystalized assembly (*model*) and 5) molecular dynamics simulations (*dynamics*). Entity Relationship (ER) diagram can be found on Supplementary Information (Figure S1 to Figure S8). Furthermore, we have added experimental data to the simulated systems such as Ki, Kd, IC50 or EC50 values which were obtained from IUPHAR²¹ and BindingDB.²² Also, we linked each main database object to a set of bibliographic references. Finally, some tables from GPCRdb^{23,24} have been included for the treatment of GPCR sequence residue numbers.

b) Submission process

In order to submit a MD simulation, a series of steps have to be taken (Figure 1):

(1) First, the simulated protein -or proteins- are declared, using its UniprotKB identifier and specifying its mutations with respect to the UniprotKB sequence, if any. To accomplish that, the user can provide a sequence to align with the UniprotKB sequence or an alignment in FASTA format. An alignment between the canonical sequence of that UniprotKB identifier and the sequence that has been introduced. If required, the alignment is performed with the pairwise alignment function of BioPython module,²⁵ and the result is visually displayed to the user with the MSA viewer tool²⁶ for user inspection.

(2) In the next step, the small molecules of the simulation (every molecular entity in the simulation except for the declared proteins) are submitted by uploading their MDL Molfile (mol) or SDF files. Each molecular entity needs to be defined as co-crystallized or bulk only molecule. Molecular descriptors and identifiers (canonical OpenSMILES,²⁷ InChI and InChIkey)²⁰ are generated from the uploaded file with RDKit²⁸ Python library and with Open Babel²⁹ 2.3.2 for OpenSMILES. Using these codes as input, we can retrieve the PubChem Compound identifier, the ChEMBL identifier and compound names through their corresponding web services.^{16,17}

Once this step is completed, a PDB file³⁰ of the protein model and co-crystallized molecules is uploaded. The user has to declare the segments of the protein by defining an unambiguous interval of PDB

residues, using the residue identifier, chain and segment identifier (a compatible legacy field in ATOM and HETATM records from PDB 1996 standard³¹) or any subset of these three elements which ensures a unique residue reference across the whole PDB file.

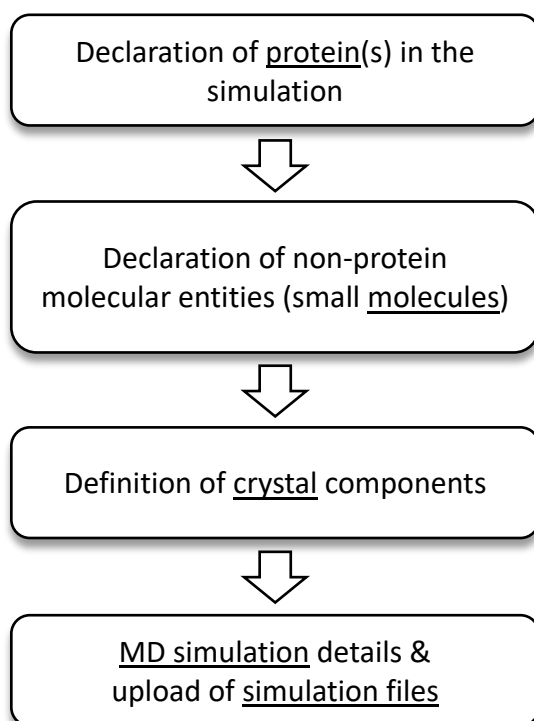


Figure 1 Main steps in the simulation submission process. Key stored items are underlined.

We understand by segment any continuous region of residues present on the PDB file split by one or more gaps when the sequence from PDB file MODEL record is aligned against the complete sequence of the protein. These gaps should be present only if some part of the protein has no structural data available and cannot be modelled or if the missing region of the protein is not crucial for the study. Thus, the user defined segments may be associated with the corresponding interval in the submitted sequence from the first step. The user also should indicate if the segments are covalently bound by peptide bond

between them. These last two steps can be done manually, although we provide an automatic suggestion. Once every segment in the PDB has been matched to an interval of the submitted sequence, our script checks if there are any insertions, deletions or mismatches in each pair. This script also suggests the number of segments that should be defined and in which line they can be found in the uploaded PDB. The presence of residue names not following PDB or IUPAC standard -like 'HSP', protonated histidine use in CHARMM force field-,³² which we may not have included as protein residues, prevents us from doing this segment definition automatically, as there is no way to include all possible modified residue names. We have included the most common ones, including the cited 'HSP' for CHARMM³² and AMBER³³ force fields. If a mismatch is found between the PDB segment and the submitted sequence, and it is due to an unknown amino acid residue name, we allow the user to proceed with the simulation submission. Any other mismatch, deletion or insertion raises an error. The script can handle non-standard NAMD³⁴ Psfgen tool hexadecimal numbering and insertion code numbering, which are frequently used when the residue identifier of the PDB is greater than 9999. Later in this same step, the user associates the declared co-crystallized molecules with their corresponding residue name in this same PDB file.

(4) In the last step, the user uploads all the trajectories files of the simulation, the PDB file of the built system (which includes all the molecules in MD simulation such as solvation water, bulk ions and bulk lipids in addition to the crystal and co-crystallized molecules from the previous step) and the force-field native topology and parameter files. Then, one must define the association between the bulk molecules defined in step two and the residue names in the built PDB file. Finally, the user adds some information about the MD methodology such as the integration time-step, force-field version, simulation software or solvent and membrane type (implicit or explicit) used.

On steps 3 and 4, a tool is provided to verify if, whenever is possible, that submitted MDL Molfile (mol) or SDF files matches with the molecules present on PDB file. For this reason, is mandatory to include all hydrogen atoms in PDB files.

c) Simulation browser and integrated pages

The simulation browser is a two-step search engine that allows users to find simulations by defining a desired composition of proteins and molecules as well as some features related to the nature of the simulation.

In the first step, the user introduces the name or identifier of the molecule or protein. For molecules, InChI, InChIkey, PubChem Compound, ChEMBL and SMILES codes are allowed, as well as most of synonyms included in PubChem and the IUPAC name. For proteins, UniprotKB accession numbers and common names can be used. Molecule results are split into two categories: standard form and specific state. The former is a molecular entity that represents the standardized molecular representation of a chemical substance (compound). The latter is each one of the molecular entities which represent the same chemical substance but with differences due to (de)protonation or tautomerization state or isotopic composition. The search performed in this first step is done using Haystack, a Django module which independently indexes the desired tables in the database for a better performance.

From the results of this first step, one can add the desired protein, standard form or specific state to the second step by clicking a button. This way, one can build the desired composition of the simulation in order to perform the second step search yielding simulations as search results.

Moreover, Boolean operators (AND, OR, NOT) for the first step search and advanced options for filtering simulation according to other metadata are also available.

In order to display the information stored, a set of pages were built. The simulation page includes information about the software and the force field used to generate it, the time step (the trajectory integration lapse), the delta time (lapse between frames) and the trajectory, topology and parameter files of the simulation, as well as the experimental data retrieved from BindingDB or IUPHAR databases, if available. More importantly, it has a link to the pages of the molecules and proteins that are present in the simulation, and to the complex structure page, which is basically the internal object that represents the crystalized assembly (*model*) of the simulation. These pages have also links to other related objects in the database making navigation more intuitive.

Molecule pages include the 3Dmol.js viewer to visualize the molecules in 3D³⁵ when 3D coordinates are available and their corresponding identifiers in ChEMBL and PubChem. Finally, the protein page includes its sequence, its organism name and whether it is a mutant or not.

d) Structure and MD data visualization

Visualization of MD trajectories deposited can be done through GPCRmd viewer, which renders graphical representations of 3D structural data including the coordinates contained in the MD trajectories. Furthermore, it provides a series of analysis tools some of them integrated into the visualizer functionality (Figure 2a).

Our viewer is based on NGL viewer¹⁸ and MDsrv³⁶ 0.3 software. When a trajectory time-point (frame) requires to be rendered, NGL viewer sends a request to MDsrv and the latter delivers the binary coordinate data corresponding to that frame. In this manner, a frame can be rendered as soon as it is downloaded, so previous downloading of the whole trajectory file is no longer needed. Once data has been downloaded, it is kept in browser cache for later visualization.

NGL viewer uses WebGL, a JavaScript API which have been integrated into the web standards of web browsers and offers hardware-accelerated graphics on supported web browsers without external plugins (see <http://caniuse.com/#feat=webgl> for browser availability details). Furthermore, NGL provides a web embedded version with a developer JavaScript API for controlling the viewer, which allowed the addition of new selection features that exploit the metadata stored in the databases. These new features are:

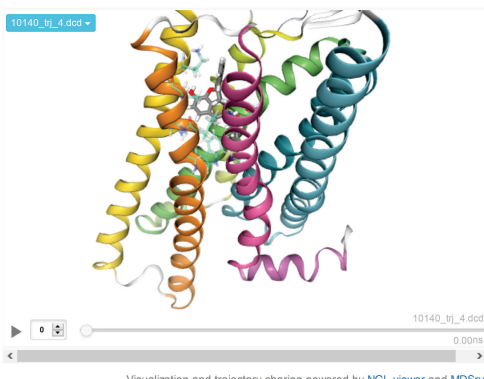
- **Ballesteros–Weinstein and GPCRdb numbering residue selection:** Ballesteros–Weinstein numbering is based on the presence of highly conserved residues in each of the seven transmembrane helices in GPCRs. It consists of two number separated by a dot. The first one is the helix numbering starting from N-terminal while the second one is 50 for the most conserved residue on the helix (e. g. 3.50). For the remaining residues, the second number decreases towards N-terminal or increases towards C-terminal along receptor protein sequence. However, when comparing GPCR sequences, some receptor present insertions or deletions that produce shifts in Ballesteros–Weinstein numbering. To tackle this issue, GPCRdb implemented a new residue numbering, based on the

Ballesteros–Weinstein scheme (e.g. 2x50) and structural alignments.²⁴

These two numberings are available for custom user selections when creating graphical representations for GPCR protein residues, together with the combination of the two of them (e.g. 2.50x50). Furthermore, selecting by the residue number assigned on the corresponding PDB file is also available. Numbering assignment for mutated residues in GPCRs is done during submission (see section “b)” on 2 Methods).

- **Quick-selection buttons:** We have implemented buttons with representations tailored to the molecular entities of the simulated system, along with a “binding site” button for showing protein residues close to ligand molecules if present (Figure 2b).
- **Distance between components selection:** This highlights molecules from a user selected molecular entity if one of their atoms are within a user defined distance of a user defined atom selection (Figure 2c). The distance filter is applied each time that a trajectory frame is displayed and it updates which molecules are currently highlighted.
- **Custom representation:** User can set any atom selection using the NGL selection language, which has been complemented with the possibility to use different generic GPCR residue numbering schemes to refer to GPCR residues. (Figure 2d). In addition, several NGL viewer representation styles (e. g. cartoon, licorice) are available.
- **Residue selector tool:** This tool allows to select residues directly from their protein sequence in different protein segments (Figure 2e). The protein sequences of this segments present a curated numbering associated to the actual residue numbers in the submitted PDB file.
- **Motif selector:** Our visualization tool provides buttons with predefined selections for highlighting sequence motifs or domains conserved in the different GPCR family classes (e.g. Class A PIF and DRY motifs, Figure 2f).

a) Delta-type opioid receptor



10140_trj_4.dcd

10140_trj_4.dcd 0.00ns

Visualization and trajectory sharing powered by NGL viewer and MDSrv

b) Quick selection

Receptor Natrindole POPC Chloride Sodium Ion

Water Binding Site All Clear

c) Show residues within A of Natrindole

d) Atom selection

Specify your selection

Licorice Element

e) Sequence selection

G	S	P	G	A	R	S	A	S	S	L	A	L	A	I	A	I	T	A	L
38	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Y	S	A	V	C	A	V	G	L	L	G	N	V	L	V	M	F	G	I	V
56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
R	Y	T	K	M	K	T	A	T	N	I	Y	I	F	N	L	A	L	A	D
76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
A	L	A	T	S	T	L	P	F	Q	S	A	K	Y	L	M	E	T	W	P
96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115
F	G	E	L	L	C	K	A	V	L	S	I	D	Y	Y	N	M	F	T	S
116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135
I	F	T	L	T	M	S	V	D	R	V	I	A	V	C	H	P	V	K	
136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155
A	L	D	F	R	T	P	A	K	A	K	L	I	N	I	C	I	W	V	L
156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175

f) GPCR conserved positions

Class A Class B Class C Class F

Highly conserved residues and motifs

D3.32 I3.40 D3.49 E6.30L F6.44 W6.48 Y7.53

PIF DRY NPxxY

Most conserved residues

N1.50 D2.50 R3.50 W4.50 P5.50 P6.50 P7.50

Figure 2 Selection and visualization tools of the GPCRmd viewer . Example of the selection and visualization tools available in the case of the δ -type opioid receptor.

d.1) Distance between atom pairs

This tool calculates distance between atom pairs through the different frames (time-points) of a trajectory. Results are presented as a plot of distance by time or by frame (Figure S9a), which can be exported as an image or csv file, and the user can set a stride value in order to skip frames and speed up the computation. Moreover, the computed distances are displayed by the viewer as lines connecting the atom pairs and indicating, for each frame, the distance value.

d.2) Root-Mean-Square Deviation

Root-Mean-Square Deviation (RMSD) tool computes using MDTraj the RMSD of protein (alpha carbon, backbone, non-hydrogen or all atoms) or all ligand atom coordinates along a MD trajectory using a user defined trajectory frame as reference. As in the case of distance tool, results are shown in a plot of RMSD as a function of time or frame (Figure S9b). The data can be exported as a plot image or a csv file, and the user can set a stride or a trajectory frame range to speed up the analysis.

d.3) Ligand-residue interaction frequency

This tool allows computing the frequency of the contacts or interactions between protein residues and ligand along the simulation. The maximum distance at which ligand atoms are considered to be interacting with the protein is given by the user with a default value of 4Å.

The user may choose between taking into account only heavy (non-hydrogen) or all atoms, and can define a frame stride.

Results are presented as table and a plot, showing the residues found to interact with the ligand and the contact frequency (Figure S9c). Moreover, the interacting residues can be displayed, giving a more visual out-come and the obtained data can also be downloaded.

d.4) Salt bridges

Salt bridges tool finds salt bridges between protein residues and computes the contact frequencies of the pair of atoms involved along a trajectory frame range given by the user. We consider as salt bridge

any contact formed between one atom of the set {Arg-NH1, Lys-NZ, His-NE2, His-ND1} and another of the set {Glu-OE2, Asp-OD2} which lays closer than 4Å. Only protonated histidines are considered.^{38,39} The results are displayed in a table, indicating the salt bridges and the frequency in which they are found. Also, the salt bridges can be displayed by the viewer as lines connecting the interacting atoms.

d.5) Hydrogen bonds

The GPCRmd viewer also makes it possible to identify hydrogen bonds and to compute the frequency of their formation through a MD simulation. This relies on MDTraj Wernet-Nilsson method that assigns hydrogen bonds based on distance and angle cutoffs considering any combination of donor atoms (NH or OH) and acceptor atoms (N or O) which holds the condition expressed in (Equation 1):

$$r_{DA} < 3.3 - 0.00044 \cdot (\delta_{HDA})^2 \quad (1)$$

Where r_{DA} is the distance in Angstroms between donor and acceptor heavy atoms, and δ_{HDA} is the angle formed by the hydrogen atom, donor, and acceptor atoms in degrees.⁴⁰ Additionally, users can add more levels of restriction, such as:

- Not considering hydrogen bonds formed between neighbor residues (i.e. less than 4 residues apart), which are usually the bonds that stabilize alpha helices.
- Considering only those bonds formed between acceptor and donor atoms belonging to side chains or lipids.
- Returning only hydrogen bonds that are found in a frequency higher than a threshold given by the user.

The results are exposed in two different tables: intramolecular bonds, which includes those bonds formed between protein residues, and intermolecular bonds, including all the rest. The hydrogen bonds obtained can also be represented by the viewer.

d.6) Hydrogen bond interaction flare plot

The GPCRmd viewer integrates a tool for the study and representation of intra-protein interactions based on a Javascript library developed at the Stanford University by Dr. Fonseca and Dr. Venkatakishnan. This approach makes it possible to obtain a highly

visual depiction of complex data. This library uses a circular interactive graph, named flare plot, to display of interactions between protein residues throughout MD simulations.

Applied to the study of GPCRs, flare plots are capable to display networks that group protein residues according to the helix to which they belong. Furthermore, these plots differentiate interactions between residues of the same helix from inter-helix interactions. These two features are highly useful for exploring how residue-residue interactions evolve over the course of a simulation, and yield qualitative data on the frequency of each of such interactions.

Particularly, GPCR viewer represents GPCR hydrogen bonds interaction data pre-computed after trajectory submission by means of the MDTraj Python library, followed by the parsing of the obtained results in order to generate data in JSON (JavaScript Object Notation) format. Such JSON data is finally used to render the flare plot displaying the interaction network.

e) Interaction general criterion for case studies

In order to capture the different interaction types, we use a general criterion based on distance. This approach makes it possible to retrieve such interaction results on the fly within a reasonable time, but have the draw-back of not specifying the type of interactions that are found, which would require a longer computational time.

Different interaction types occur at different distances. Thus, in order to detect all of them, we need to set an inclusive distance threshold. However, an excessively wide threshold would imply accepting as interaction events that are not, generating background noise or false interactions. With the purpose of establishing what distance threshold should be set, we revised the characteristic distances between atoms in the main interaction types:

- Hydrogen bonds: The accepted and most frequently observed geometry for a hydrogen bond implies a distance ranging from 2.4 to 3.5Å between the two non-hydrogen atoms.⁴¹
- Ionic interaction (salt bridges): In this type of interaction distance must be within the range of 4Å between non-hydrogen atoms.³⁹
- Hydrophobic interaction: There is no consensus concerning the distance dependence and effective range for hydrophobic

interactions. However, it has been described that, for this interaction type, heavy atoms within 4.5Å can be considered to be within bonding distance.⁴²

- Aromatic interactions: In interactions between aromatic compounds, the centers of the aromatic rings have been found to be separated by a distance from 4.5Å to 7Å.⁴³

Overall, most interaction types should be detected if we accept as interaction any ligand-residue distance around 4Å.

Code availability:

Code for the web platform will be available at <https://github.com/GPCRmd/GPCRmd>.

3. Results

As a demonstration of the tools available at the GPCRmd viewer, here we present two interaction case studies based on the analysis of MD data of a selection of members of the GPCR family:

- Comparison of interaction events in 5-hydroxytryptamine receptor 1B (5-HT_{1B}) and 5-hydroxytryptamine receptor 2B (5-HT_{2B}).
- Interaction events associated with δ -type opioid receptor (δ -OR) sodium modulation (GPCRmd dynamics id:4).

When analyzing the interaction events between residues of the considered receptors and their ligand, we applied the ligand-residue interaction frequency analysis tool. This tool is intended to provide an overall idea of all the interactions that take place between a ligand and the protein residues throughout a simulation. For this purpose, a general criterion, as explained above in Methods, is applied. With this criterion, based on accepting as interaction any ligand-residue distance around 4Å, most interaction types should be detected. Aromatic interactions may be more difficult to detect using this distance threshold, but increasing the threshold much more would imply adding a lot of false positives. To illustrate the effect of different thresholds when trying to identify interactions, we will perform our analysis with thresholds ranging from 3 to 4.5Å. Only heavy atoms will be considered in the analysis.

In order to evaluate the results obtained, we present our results in comparison with the ligand-residue interaction data offered by the GPCRdb,²³ which is an important and reliable source of information for the study of GPCRs. Since, in contraposition to our MD analysis, GPCRdb information is based on static structural data, we will use our case studies as an opportunity to discuss the idea that MD data is a useful approach that can add a new dimension of description to structural data – that is, information about time.

Additionally, the case studies are complemented with an intramolecular protein interaction analysis based on the generation of interaction networks, represented as flare plots.

a) 5-hydroxytryptamine receptor 1B and 5-hydroxytryptamine receptor 2B

Ergotamine is an anti-migraine drug which activates G protein coupling and β -arrestin signaling when binding to 5-HT_{1B} receptor. Differently, at the 5-HT_{2B} receptor ergotamine favors β -arrestin over G protein coupling. Understanding the interaction process of these receptors with ergotamine is of high interest since agonism at the 5-HT_{1B} receptor has been associated to the anti-migraine effect of ergotamine while agonism at the 5-HT_{2B} receptor seems to be related to undesired valvulopathic effects.^{44,45}

Previous MD studies on these receptors show that ergotamine adopts a similar position in the orthosteric binding pocket of both receptors, but establishes increased contacts with an extended binding pocket in the extracellular part of transmembrane helix 5 (TM5) of the 5-HT_{2B} receptor, which are not present in the 5-HT_{1B} receptor. In turn, this impacts the on orientation and conformational properties of the PIF motif, which plays a critical role in the activation state of GPCRs.⁴⁶

a.1) Ligand-residue interaction analysis

In this analysis, we applied the GPCRmd viewer ligand-residue interaction analysis to find which residues of 5-HT_{1B} and 5-HT_{2B} receptors interact with ergotamine and at which frequency. Five MD trajectories were analyzed for each receptor using different distance thresholds to capture the interactions. All trajectories contained 5000 frames, representing a simulation time of 500 nanoseconds (ns) per

trajectory. The results generated were compared with those obtained at the GPCRdb for determining if there is correspondence between the outcomes of both approaches.

Interestingly, results (Table S1) show that the ligand-receptor interactions formed at both receptors are mostly similar, with some essential exceptions. Residues 5.38x39 and 5.39x40 (positions indicated in combined Ballesteros–Weinstein/structural-based nomenclature), which belong to TM5, show a highly frequent hydrophobic interaction with ergotamine on 5-HT_{2B} receptor, while almost no noticeable interaction is found in the case of 5-HT_{1B}.

Our results largely match with the GPCRdb interaction data. Moreover, our tool adds on top valuable information about the frequency of interaction. For instance, our analysis reveals that some residues indicated as accessible to the ligand at GPCRdb are hardly engaged in residue interaction over time.

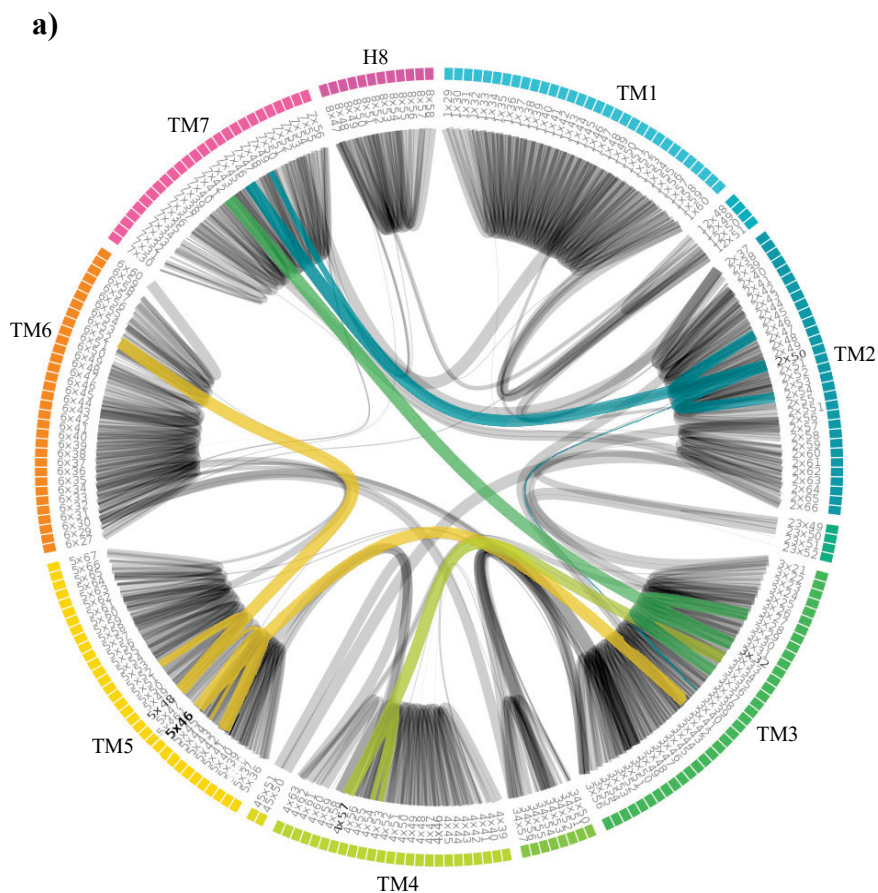


Figure 3 Intramolecular hydrogen bonds of 5-HT_{1B} and 5-HT_{2B}. Flare plots representing the network of hydrogen bonds among the residues of 5-HT_{1B} (a) and 5-HT_{2B} (b). Helices are indicated in colors ranging from light blue (transmembrane helix 1, TM1) to purple (helix 8, H8). Residue positions are expressed in the GPCRdb structure-based numbering scheme. Lines connecting residues represent hydrogen bonds formed between them, and their width the frequency of such contacts. Residue contacts showing most significant differences between the two receptors are highlighted according to the helix where they belong: 2x50 (dark blue), 3x32 (dark green), 4x57 (light green), 5x45 and 4x48 (yellow).

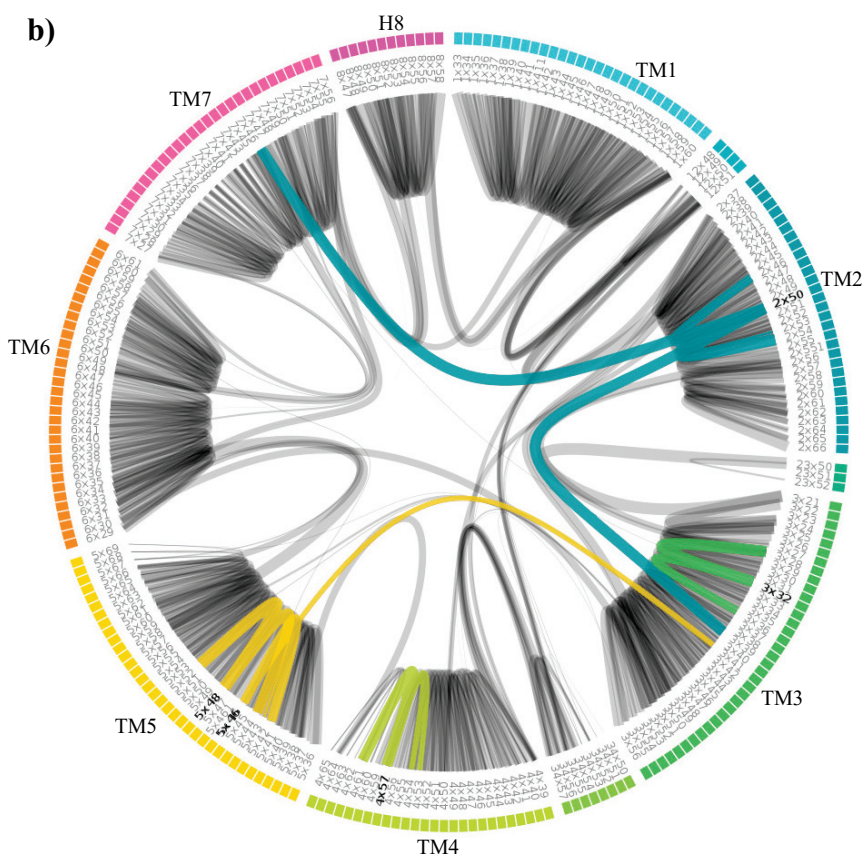


Figure 3 Intramolecular hydrogen bonds of 5-HT_{1B} and 5-HT_{2B} (continued) .

a.2) Intramolecular interaction analysis

In order to provide an insight on the differences in intra-molecular interactions that occur at 5-HT_{1B} and 5-HT_{2B} receptors, flare plots representing hydrogen bonds between protein residues were generated. Each flare plot was computed using one MD trajectory containing 5000 frames (simulation time of 500 ns).

Such interaction networks show important differences between these two receptors (Figure 3a and b, Table 2a). For instance, residues 5x46 and 5x48 at 5-HT_{1B} TM5 present strong inter-helix interactions that at 5-HT_{2B} either are not found or only in a low frequency. Another noticeable difference is found at residue 3x32, which at 5-HT_{1B} form a very stable interaction with 7x42, but at 5-HT_{2B} the same

interaction is extremely unstable. Similarly, 4x57 interaction with 3x34 is only found at 5-HT_{1B}. Finally, another disparity worth to mention is the hydrogen bond linking residues 2x50 and 3x49, being this interaction far more frequent at the 5-HT_{2B} receptor than in 5-HT_{1B}.

Importantly, the observed differences in the interhelical network may be responsible for the different signaling outcome. Thus, stronger 5x48-6x52, 5x46-3x41, 4x57-3x34 and 3x32-7x42 interactions in the 5-HT_{1B} receptor could promote G protein coupling and β -arrestin signaling. In contrast, stronger interaction of residues 2x50 and 3x49 could favor arrestin in the 5-HT_{2B} receptor. These insights can be valuable for understand signaling bias and the rational design of novel ligands.

b) δ -type opioid receptor

The δ -type opioid receptor (δ -OR) is a GPCR essential for the regulation of nociception, mood and awareness. Interestingly, it was found that the sodium ion has a fundamental role in mediating allosteric control of receptor functional selectivity and constitutive activity. Particularly, sodium acts stabilizing a receptor state with reduced agonist binding and low β -arrestin constitutive activity. Site-directed mutagenesis of key sodium allosteric site residues such as Aspartate 2.50x50 to Alanine (Asp2x50Ala) were found to dramatically change GPCR functional activity, transforming classical δ -opioid antagonists such as naltrindole into potent β -arrestin-biased agonists. Such studies revealed an essential role for allosteric sodium anchoring residues at specifying GPCR signal

transduction and pharmacology. A deeper knowledge on sodium allosteric modulation process at atomic level is of interest for the design of new δ -OR ligands and allosteric modulators with improved selectivity and functional profiles.⁴⁷

b.1) Ligand-residue interaction analysis

In this study, we considered the δ -OR in complex with naltrindole. Given the important role of Asp2x50 in sodium modulation, we analyzed the frequency of interaction of naltrindole with δ -OR

residues through MD simulations in wild type and Asp2x50Ala mutated receptor in order to examine if sodium modulation affects on the ligand binding process from a dynamic, time-dependent perspective. 96 replicates of MD stridden trajectories containing the last 16 frames (64 ns) – consisting in a simulation time of 6.144 μ s ns – was analyzed for each condition at different distance thresholds. Again, our results were compared with those obtained at GPCRdb from static wild type δ -OR as a reference.

Little noteworthy differences were found between wild type δ -OR and Asp2x50Ala mutant concerning ligand-residue interaction (Table S2). However, there appeared to be some perceptible variations on the frequency of some of the interactions. First of all, residues belonging to helix 7 of the mutant δ -OR (mainly 7.39x40 and 7.42x41) seemed to interact with naltrindole at a lower frequency than the wild type receptor. Moreover, it could also be argued that residue 5.39x40 shows a lower interaction frequency at the mutant δ -OR, although more analysis should be done to clarify it. Overall it does appear that the allosteric effect of sodium on δ -OR do not imply notable qualitative changes in the residue interactions process with naltrindole, but might have some type of quantitative impact in terms of interaction frequency.

b.2) Intramolecular interaction analysis

With the purpose to obtain an insight on the conformational changes that take place as consequence of sodium allostherism, we studied δ -OR intramolecular contacts by creating flare plots of residue-residue hydrogen bonds. Each flare plot was computed using 96 replicates of MD stridden trajectories containing the last 16 frames (64 ns), which represents a simulation time of 6.144 μ s.

As could be expected, results (Figure S10a and b, Table 2b) show, in the mutant receptor, the loss of contacts of the mutated residue, Asp2x50Ala, in comparison with the wild type. Further-more, the hydrogen bond between 6x38 and 5x58 is formed at a lower frequency at the mutant, as happens with the contact 4x64-3x36. Overall, it seems that intramolecular interactions are reduced in the mutant receptor, which may have an impact on eliciting arrestin bias

Table 2 Intramolecular hydrogen bonds of 5-HT_{1B}, 5-HT_{2B}, wild type δ -OR and Asp2x50Ala mutant δ -OR.

a)

Residues	freq 5-HT1B	freq 5-HT2B
2x50-3x39	9.64	88.38
2x50-7x46	96.7	88.2
2x50-7x49	48.18	0
3x32-7x42	99.12	4.24
4x57-3x34	95.18	0
5x46-3x41	98.62	39.84
5x48-6x52	98.58	0

b)

Residues	freq wt	freq mt
2x50-7x46	30.38	0.62
2x50-7x49	29.22	0
4x64-3x26	27.72	20.52
6x38-5x58	12.24	3.84

4. Discussion and conclusions

We have developed a web platform intended to assist the needs of researchers interested in the study of GPCRs, especially those focused on molecular dynamics. Our platform provides users with multiple tools that make it possible to visualize and analyze the MD data stored at the GPCRmd database without needing to first download the dataset. On the one hand, it incorporates numerous tools for the selection and customization of the representation of GPCRs and the molecules interacting with them. Moreover, it offers tools for the analysis of the basic features of the simulations, and presents the descriptive data generated in such analysis in a visual and interactive manner. The services offered to the users by our platform are designed to be interactive and intuitive with the intention to approach GPCR MD data to all researchers studying this protein family, including those who are not familiarized with computational tools or complex MD analysis software.

GPCRdb, the database of reference for the study of GPCRs, contains a vast amount of information about this protein family, focusing on

structural and sequence data. It includes descriptive data, diagrams and multiple web tools. GPCRmd aims to extend this information by adding a new level of description: the time-dependent behavior of GPCRs. The differences between the data stored at GPCRdb and GPCRmd (static vs. dynamic data) determine the divergence of the tools offered by them. For instance, while both databases benefit from the NGL viewer to allow the visualization of molecular structures, the GPCRmd viewer has been more centered on the implementation of selection and customization tools that allow the creation of elaborate molecule representations, which is necessary given the complexity of MD data. Considering the analysis tools, GPCRmd is focused on the study of the evolution of the variable being calculated over time.

Other databases contain information about GPCR MD simulations and provide extensive visualization tools and analysis results. However, they are not specifically directed to the study of GPCRs, and are designed to fit a broader group of proteins. An example which is worth mentioning is MemProtMD,⁴⁸ a database of membrane protein structures inserted into simulated lipid bilayers. Being a database of membrane proteins, MemProtMD contains information about myriad GPCRs, but is focused on the study of the behavior of proteins within a lipid bilayer membrane environment. Thus, its visualization platform is centered on the depiction of proteins and lipidic structures, but not ligands and other molecules. Similarly, the analyses presented are focused on identification of lipid binding sites, local bilayer deformation by membrane proteins, and other related studies. Finally, other molecular dynamics databases, such as BioSimGrid¹² and Dymeomics,¹¹ apart from not being specific for GPCRs, differ from GPCRmd as they are created to be accessed mainly in a programmatic manner. Therefore, their web interface is modest and does not contain online predefined visualization and analysis tools.

The small case studies presented in this article were conducted with the aim to present some examples of the applicability of our platform. Moreover, they can yield a good illustration of how MD data can complement structural data by providing information on the evolution of different molecular processes throughout time. In the case of ligand-residue interaction events, MD data makes it possible to obtain an approximation of the frequency at which contacts take place, which gives us an idea about whether interactions found at static structures are as common as thought or, on the contrary, are

sporadic. On the other side, MD analysis can be computationally expensive and imply a high computation time, especially for long simulations. In web-based analysis tools, it is important to retrieve results within a relatively short time, which restricts the possibility to provide complex analysis. However, comparing our simplistic ligand-residue interaction analysis based only on distances with data provided by the GPCRdb, it appears it is a good approach, given the fact that it is able to detect approximately the same interactions. We do not obtain information on the type of the interactions found (hydrogen bond, hydrophobic etc.) but, in exchange, the analysis is relatively fast and reasonably sensible.

Overall, given the complexity of GPCRs functionality and signal transduction, MD studies are a useful approach that can help us solve the enigmas that still exist about these proteins, and thus the development of tools for their study can be of value to the scientific community.

Acknowledgements

Authors would like to thank all the members of the GPCR drug design group at the Research Programme on Biomedical Informatics. Also, the developers of the NGL viewer, and specially Dr. Johanna Tiemann, for their help and assistance during the implementation of their application. Finally, we would like to thank Instituto de Salud Carlos III for funding this project.

Funding

IR-E acknowledges Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya (2015 FI_B00145) for its financial support.

This work has been supported by Instituto de Salud Carlos III, El Fondo Europeo de Desarrollo Regional (FEDER) (CP12/03139 and PI15/00460).

Conflict of Interest: none declared.

References

1. Katritch, V., Cherezov, V. & Stevens, R. C. Structure-

- Function of the G Protein–Coupled Receptor Superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53**, 531–556 (2013).
2. Venter, J. C. *et al.* The sequence of the human genome. *Science* (80-.). **291**, 1304–51 (2001).
 3. Gether, U. Uncovering Molecular Mechanisms Involved in Activation of G Protein-Coupled Receptors. *Endocr. Rev.* **21**, 90–113 (2000).
 4. Kolakowski, L. F. GCRDb: a G-protein-coupled receptor database. *Receptors Channels* **2**, 1–7 (1994).
 5. Fredriksson, R., Lagerström, M. C., Lundin, L.-G. & Schiöth, H. B. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
 6. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discovery* **16**, 19–34 (2017).
 7. Bentham Science Publisher, B. S. P. Latest Development in Drug Discovery on G Protein-coupled Receptors. *Curr. Protein Pept. Sci.* **7**, 465–470 (2006).
 8. Rask-Andersen, M., Masuram, S. & Schiöth, H. B. The Druggable Genome: Evaluation of Drug Targets in Clinical Trials Suggests Major Shifts in Molecular Class and Indication. *Annu. Rev. Pharmacol. Toxicol.* **54**, 9–26 (2014).
 9. Piana, S., Klepeis, J. L. & Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **24**, 98–105 (2014).
 10. Meyer, T. *et al.* MoDEL (Molecular Dynamics Extended Library): A Database of Atomistic Molecular Dynamics Trajectories. *Structure* **18**, 1399–1409 (2010).
 11. van der Kamp, M. W. *et al.* Dymeomics: A Comprehensive Database of Protein Dynamics. *Structure* **18**, 423–435 (2010).
 12. Ng, M. H. *et al.* BioSimGrid: Grid-enabled biomolecular simulation data storage and analysis. *Futur. Gener. Comput. Syst.* **22**, 657–664 (2006).
 13. Mixcoha, E., Rosende, R., Garcia-Fandino, R. & Piñeiro, Á. Cyclo-lib: a database of computational molecular dynamics simulations of cyclodextrins. *Bioinformatics* **32**, 3371–3373 (2016).
 14. Hospital, A. *et al.* BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic*

- Acids Res.* **44**, D272–D278 (2016).
15. Consortium, U. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).
 16. Kim, S. *et al.* PubChem Substance and Compound databases. *Nucleic Acids Res.* **44**, D1202–D1213 (2016).
 17. Bento, A. P. *et al.* The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* **42**, D1083–D1090 (2014).
 18. Rose, A. S. & Hildebrand, P. W. NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.* **43**, W576–W579 (2015).
 19. PostgreSQL Global Development Group. PostgreSQL: Documentation: 9.3: What is PostgreSQL? (2013).
 20. Heller, S. R., McNaught, A., Pletnev, I., Stein, S. & Tchekhovskoi, D. InChI, the IUPAC International Chemical Identifier. *J. Cheminform.* **7**, 23 (2015).
 21. Southan, C. *et al.* The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. *Nucleic Acids Res.* **44**, D1054–D1068 (2016).
 22. Gilson, M. K. *et al.* BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **44**, D1045–D1053 (2016).
 23. Munk, C. *et al.* GPCRdb: the G protein-coupled receptor database - an introduction. *Br. J. Pharmacol.* **173**, 2195–2207 (2016).
 24. Isberg, V. *et al.* Generic GPCR residue numbers – aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.* **36**, 22–31 (2015).
 25. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
 26. Yachdav, G. *et al.* MSAViewer: interactive JavaScript visualization of multiple sequence alignments. *Bioinformatics* **32**, btw474 (2016).
 27. James, C. A. *et al.* OpenSMILES specification. (2016). at <<http://opensmiles.org/opensmiles.html>>
 28. RDKit: Open-source cheminformatics. (2016). at <<http://www.rdkit.org>>
 29. O’Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminformatics 2011 31* **37**, W504–W509 (2011).
 30. Worldwide Protein Data Bank. Protein Data Bank Contents

- Guide: Atomic Coordinate Entry Format Description Version 3.3. (2012). at <<https://www wwpdb.org/documentation/file-format-content/format33/v3.3.html>>
31. Callaway, J. *et al.* *Protein Data Bank Contents Guide*. (1996). at <https://cdn.rcsb.org/wwpdb/docs/documentation/file-format/PDB_format_1996.pdf>
 32. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–90 (2010).
 33. Case, D. A. *et al.* AMBER 2017. (2017). at <<http://ambermd.org/>>
 34. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
 35. Rego, N. & Koes, D. 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* **31**, 1322–1324 (2015).
 36. Rose, A. & MDsrv Contributors. mdsrv v0.2. (2016). doi:v
 37. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528–32 (2015).
 38. Costantini, S., Colonna, G. & Facchiano, A. M. ESBRI: a web server for evaluating salt bridges in proteins. *Bioinformation* **3**, 137–8 (2008).
 39. Kumar, S. & Nussinov, R. Close-Range Electrostatic Interactions in Proteins. *ChemBioChem* **3**, 604 (2002).
 40. Wernet, P. *et al.* The Structure of the First Coordination Shell in Liquid Water. *Science (80-.)*. **304**, 995–999 (2004).
 41. Berg, J. M., Tymoczko, J. L. & Stryer, L. *Chemical Bonds in Biochemistry. Biochemistry* (W H Freeman, 2002).
 42. Israelachvili, J. & Pashley, R. The hydrophobic interaction is long range, decaying exponentially with distance. *Nature* **300**, 341–342 (1982).
 43. Anjana, R. *et al.* Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: a study based on orientation and distance. *Bioinformation* **8**, 1220–1224 (2012).
 44. Ramírez Rosas, M. B., Labrujere, S., Villalón, C. M. & Maassen Vandenbrink, A. Activation of 5-hydroxytryptamine1B/1D/1F receptors as a mechanism of action of antimigraine drugs. *Expert Opin. Pharmacother.* **14**, 1599–610 (2013).
 45. Roth, B. L. Drugs and Valvular Heart Disease. *N. Engl. J.*

- Med.* **356**, 6–9 (2007).
46. Martí-Solano, M. *et al.* Detection of New Biased Agonists for the Serotonin 5-HT_{2A} Receptor: Modeling and Experimental Validation. *Mol. Pharmacol.* **87**, 740–6 (2015).
 47. Fenalti, G. *et al.* Molecular control of δ -opioid receptor signalling. *Nature* **506**, 191–196 (2014).
 48. Stansfeld, P. J. *et al.* MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. *Structure* **23**, 1350–1361 (2015).

Supplementary information. GPCRmd: a GPCR specialized Molecular Dynamics database and analysis tool

Ismael Rodríguez-Espigares*, Mariona Torrens-Fontanals*, Alejandro Varela-Rial*, Juan Manuel Ramírez-Anguita* and Jana Selent*

1. GPCRmd platform structure and tools

GPCRmd database web interface have been developed using Django Web Framework (v1.9) based on Python (v3.4) and JavaScript libraries jQuery 1.9 and jQuery UI 1.11.2. PostgreSQL 9.3 is used as database engine.

The structure of the database is based on main five objects:1) *protein* objects identified by their sequence and their relationship with UniprotKB entries, 2) molecular entities (*molecule* object in GPCRmd) identified by an InChI¹ generated with forced hydrogen connectivity, 3) chemical species (*compound*) identified by standard InChI, 4) crystalized assembly (*model*) and 5) molecular dynamics simulations (*dynamics*). Entity Relationship (ER) diagram can be found on Supplementary Information (Figure S1 to Figure S8). Furthermore, we have added experimental data to the simulated systems, obtained from IUPHAR² and BindingDB³, and we also linked each main object to a set of bibliographic references. Finally, some tables from GPCRdb have been included for the treatment of GPCR sequence residue numbers.

* GPCR Drug Discovery group, Research Programme on Biomedical Informatics (GRIB), Universitat Pompeu Fabra (UPF)-Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain.

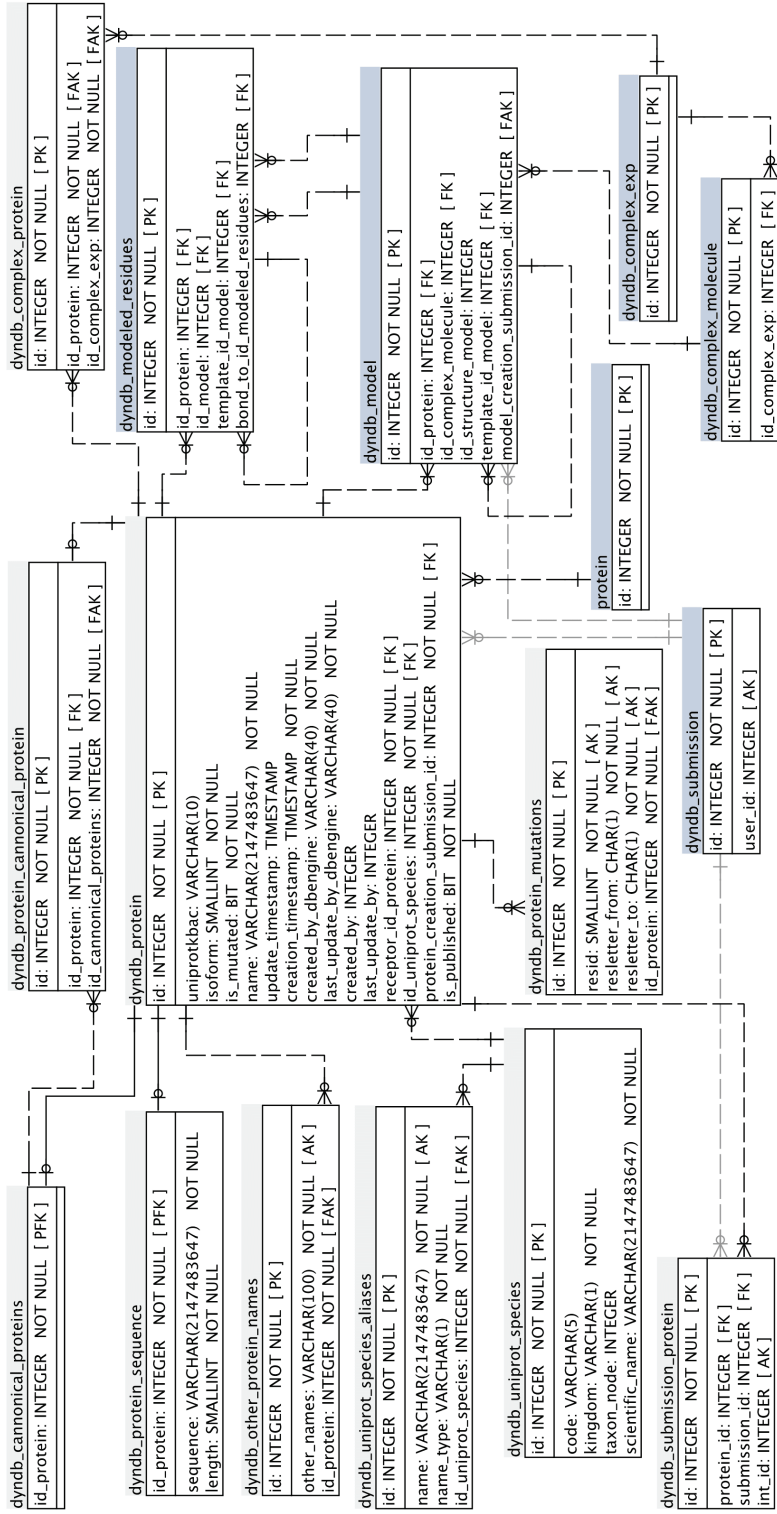


Figure S1 GPCRMd ER diagram of entities related with protein objects. Tables in blue only display fields forming part of relationships.

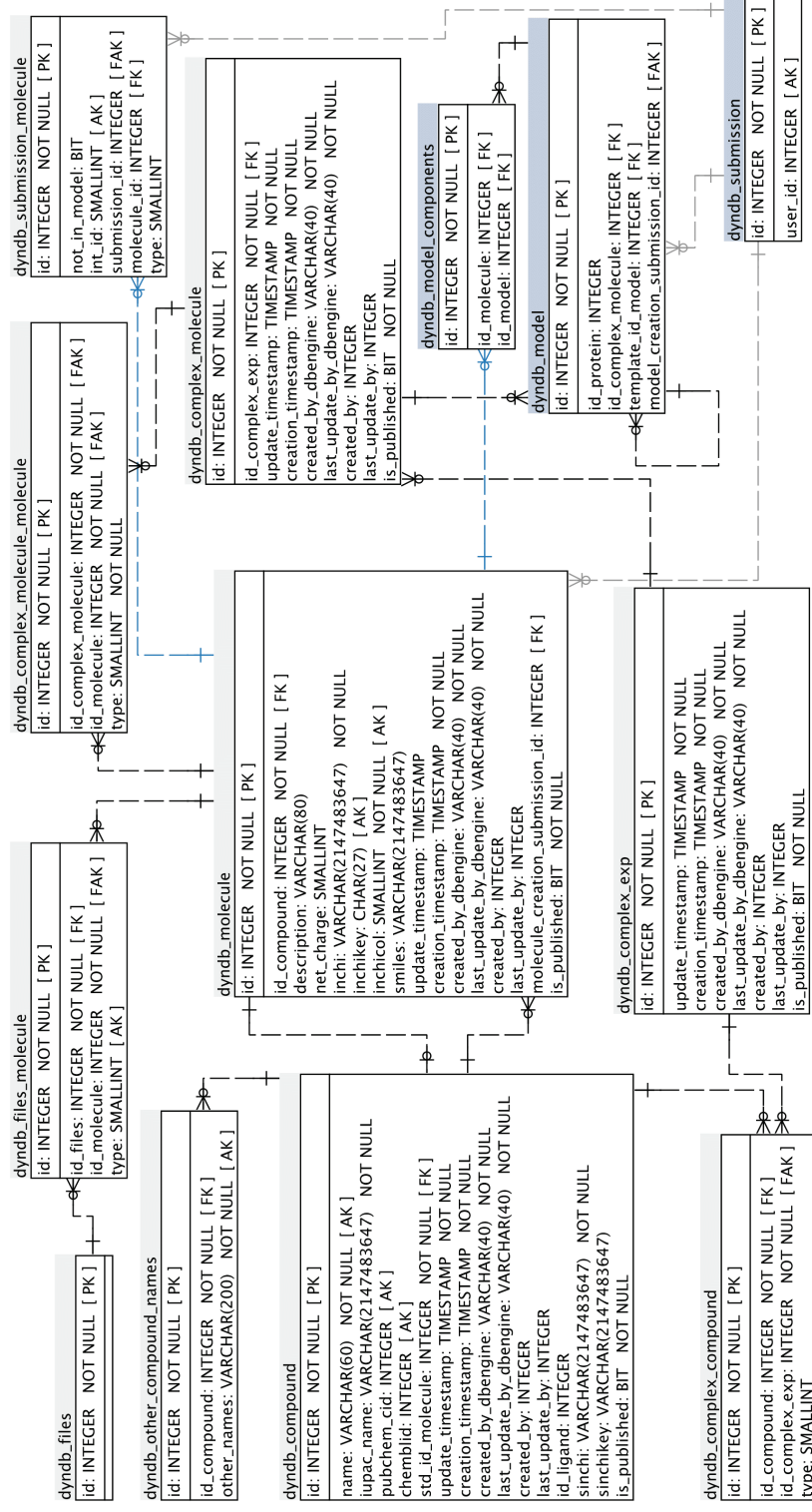


Figure S2 GPCrmd ER diagram of entities related with molecule objects. Tables in blue only display fields forming part of relationships.

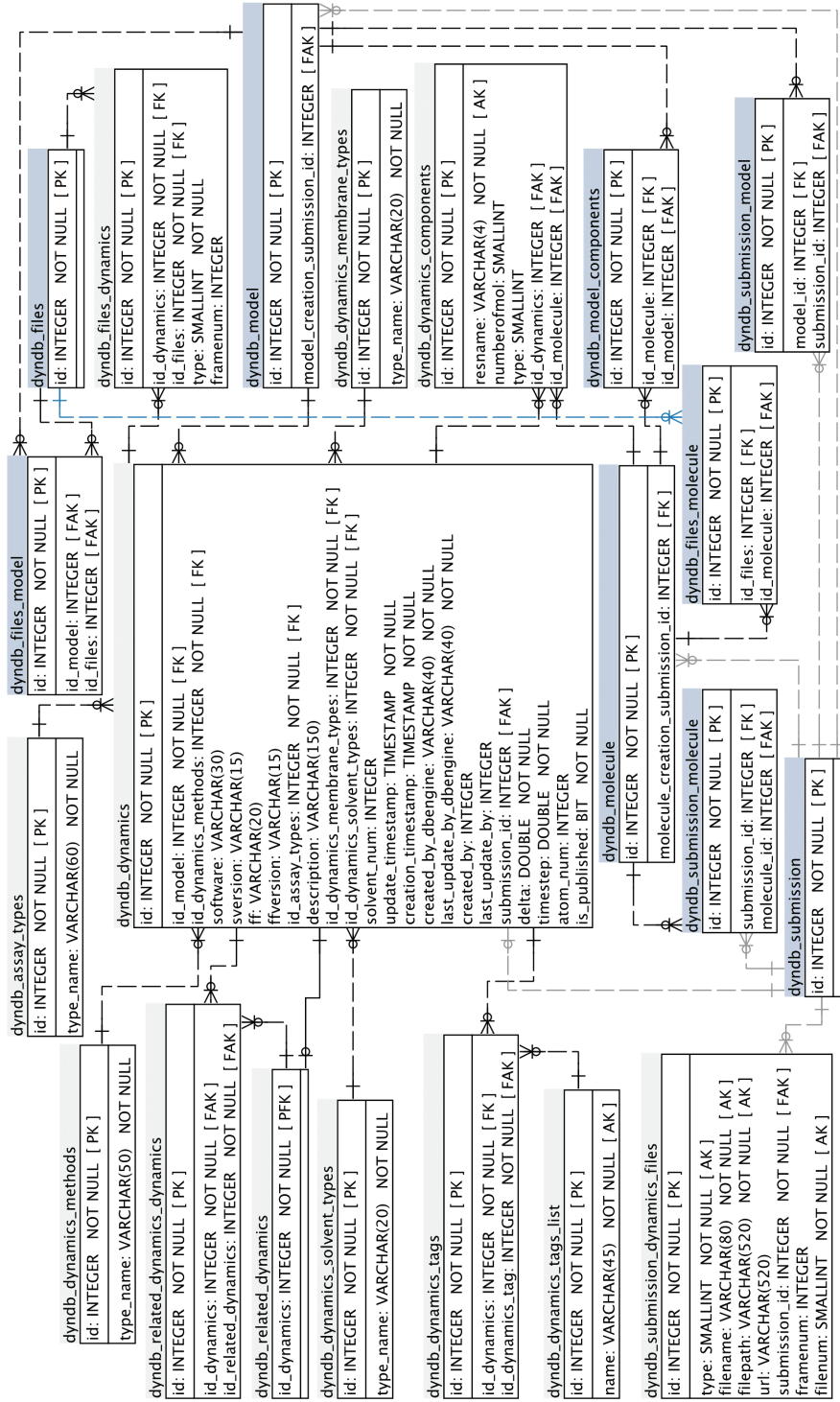


Figure S4 GPCrmd ER diagram of entities related with dynamics objects. Tables in blue only display fields forming part of relationships.

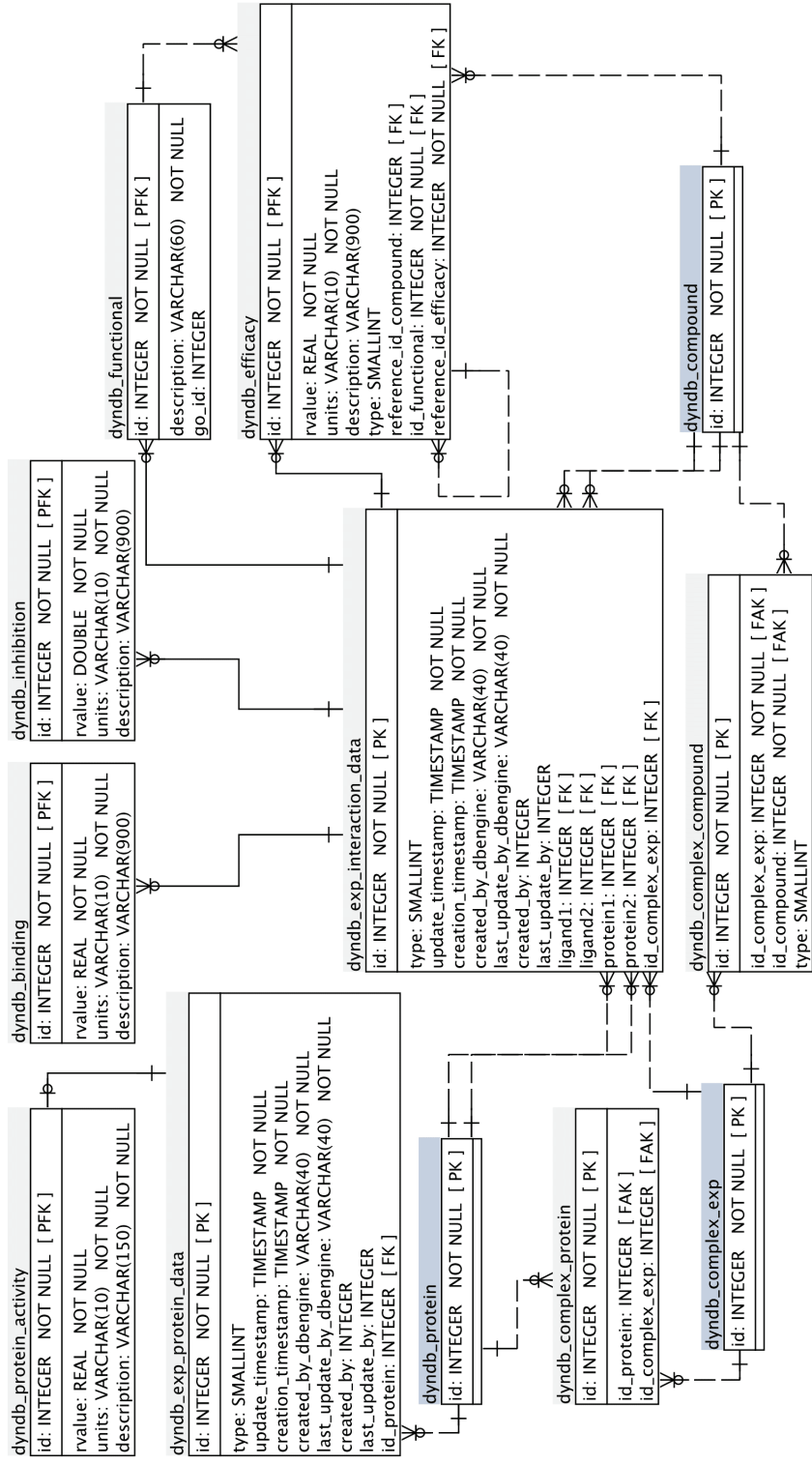


Figure S5 GPCRmd ER diagram of entities related with experimental data. Tables in blue only display fields forming part of relationships.

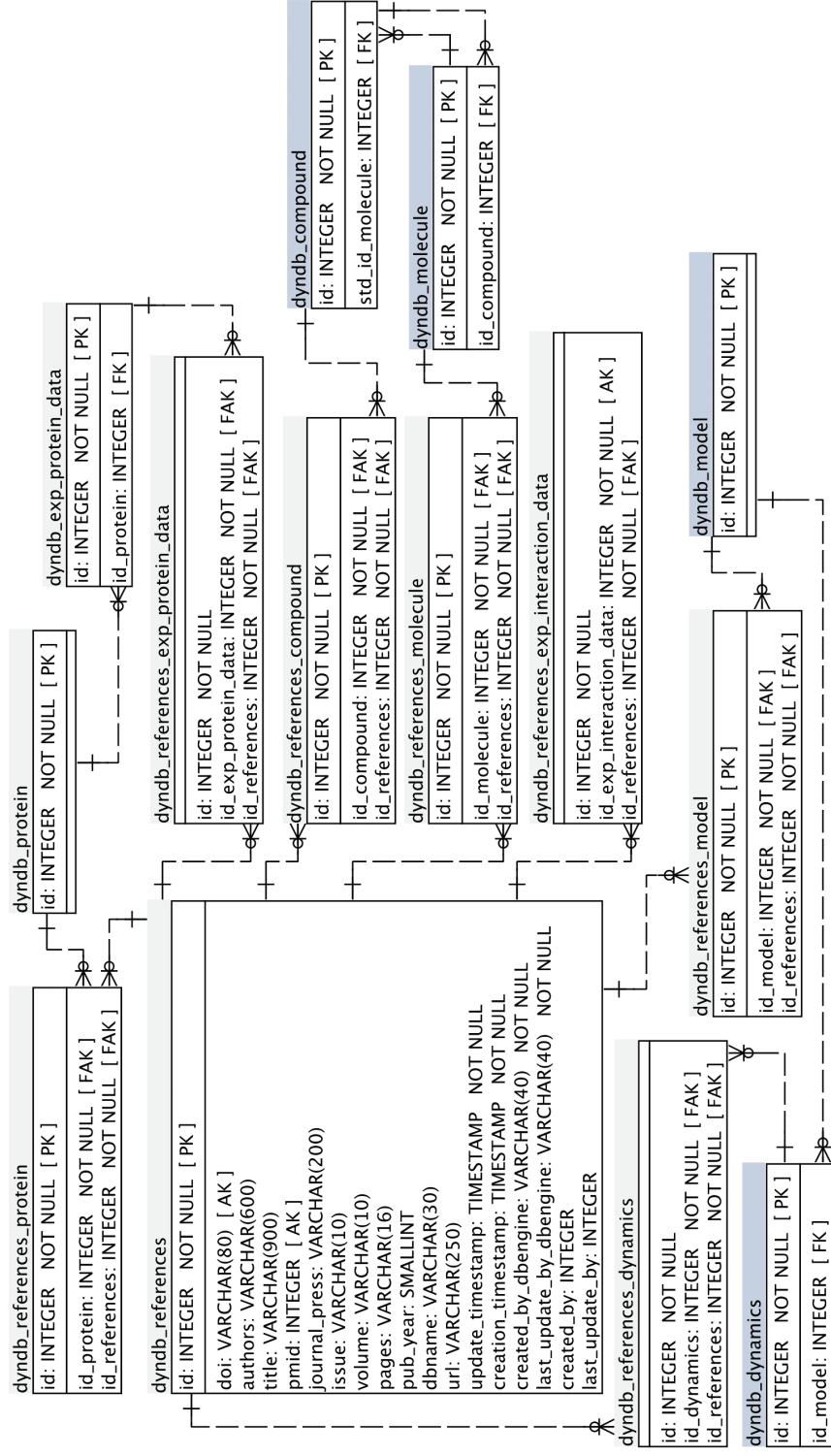


Figure S6 GPCRmd ER diagram of entities related with references objects. Tables in blue only display fields forming part of relationships.

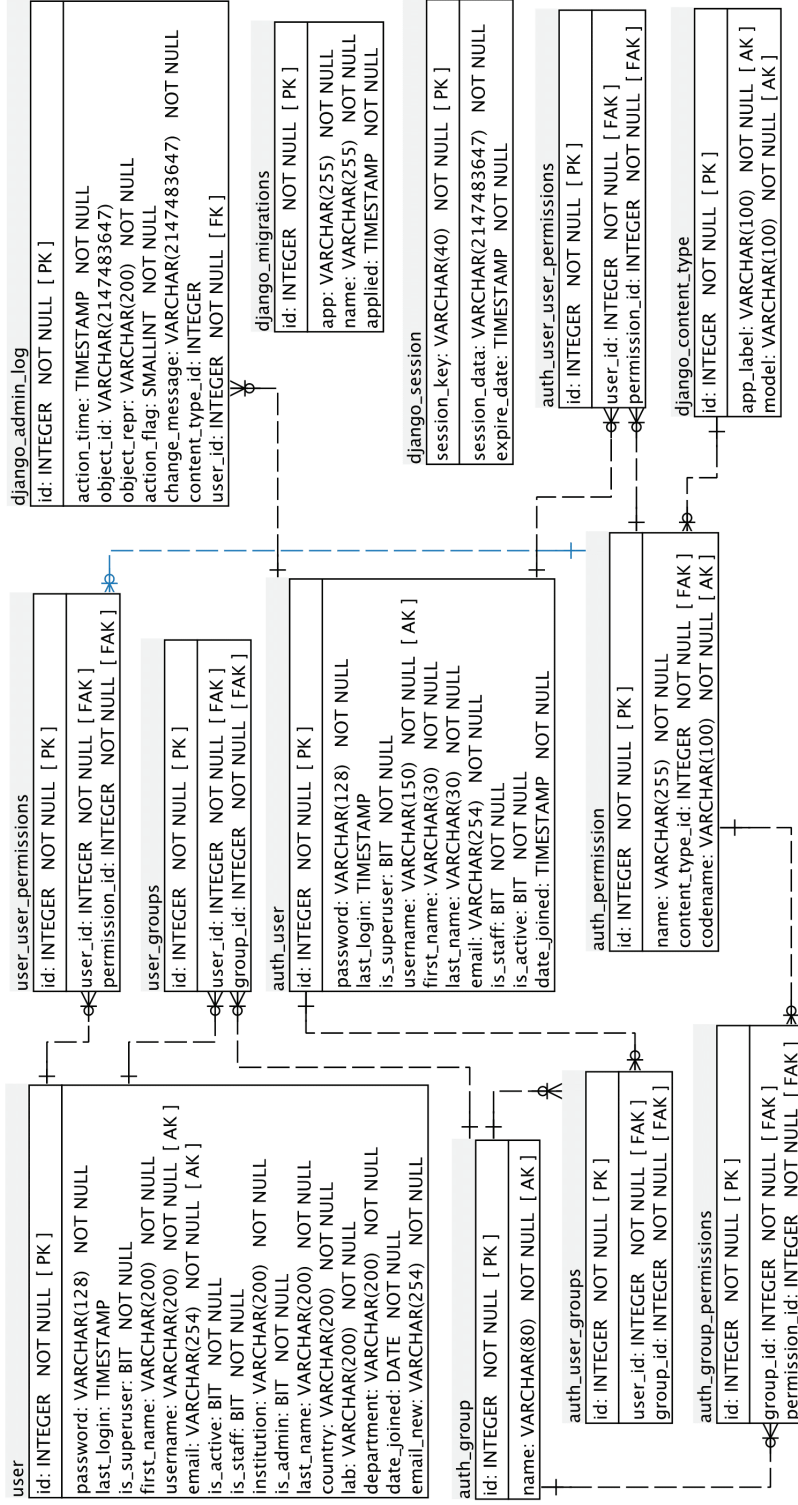


Figure S7 GPCRmd ER diagram user and Django tables.

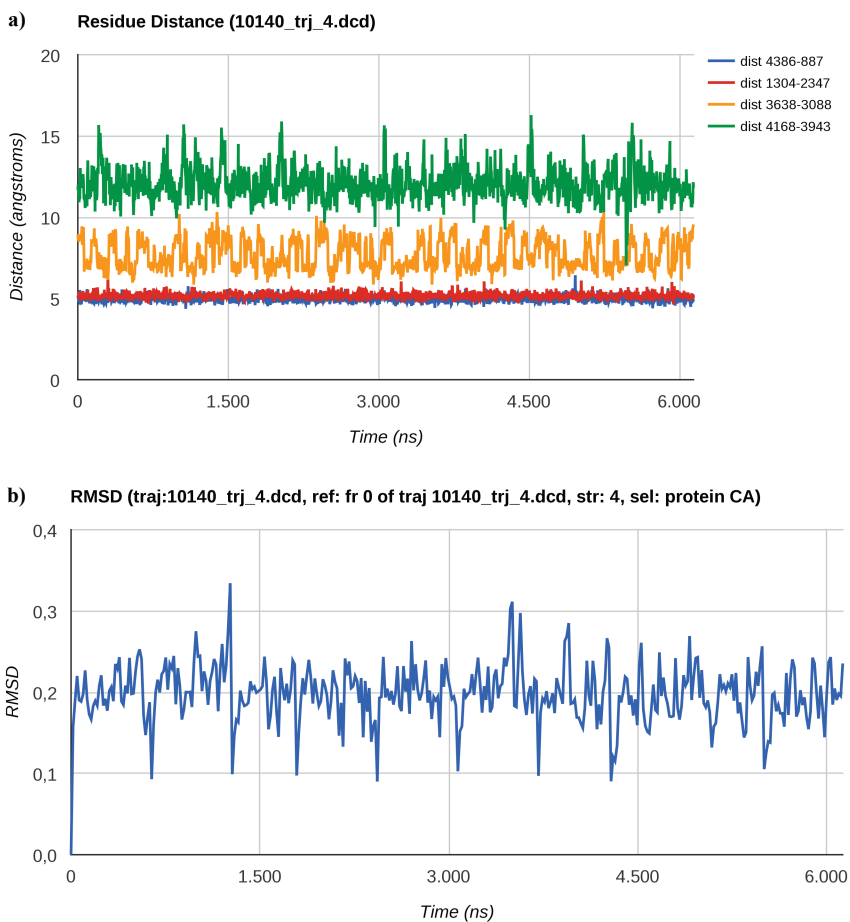


Figure S9 Plots generated by the GPCRmd viewer analysis tools . Example of plots generated when applying different analysis tools to the δ -type opioid receptor. The trajectories were previously concatenated and stridden to 4 ns/frame to reduce the number of trajectory frames. (a) Distances between alpha carbons of residues 2x50-7x46 (blue), 3x26-4x64 (red), 5x58-6x38 (orange) and 6x57-7x30 (green). Numbers at the legend indicate the atom indexes. (b) RMSD using the first trajectory frame as reference and considering only protein alpha carbons.

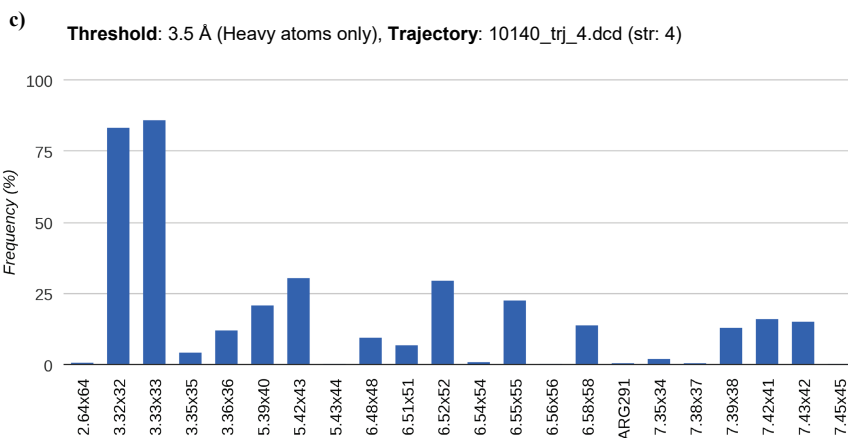


Figure S9 Plots generated by the GPCRmd viewer analysis tools (continued). (c) Interaction of the receptor the ligand (Naltrindole). Residues were considered to interact with the ligand when the distance between any of the residue atoms and the ligand is smaller than 3.5Å. Only heavy atoms are considered.

2. Molecular dynamics simulation methods

a) Serotonin 5-HT_{1B} and 5-HT_{2B} receptors

The analyzed simulations were previously published by M. Martí-Solano, F. Sanz, M. Pastor and J. Selent⁴ and their methods are reproduced here:

System preparation. 5-HT_{1B} and 5-HT_{2B} receptors (PDB IDs 4IAR⁵ and 4IB4⁶) were retrieved from the Protein Data Bank. Residues were assigned numbers according to the numbering scheme proposed by Ballesteros and Weinstein⁷ and the region of the crystal corresponding to the fusion protein BRIL was removed for the subsequent simulations. The protonation state of titratable groups was predicted for a pH value at 7.4 based on PROPKA⁸ using the implemented prediction tool of the MOE package.⁹ Subsequently, in order to place both receptors into the bilayer membrane, a hole was generated in a pre-equilibrated palmitoylcholine (POPC) bilayer – generated using the CHARMM-GUI Membrane Builder¹⁰ – by removing POPC molecules. Lipids which were in close contact with the protein atoms (<1 Å distance from any protein atoms) were deleted.

Finally, the coordinates for water and ions were generated using the solvate and autoionize modules of VMD 1.9.1.¹¹ The ionic strength was kept at 0.15 M by NaCl and we used the TIP3 water model. The all-atom models of each system were generated by using the Amber99SB force-field parameters and ergotamine, LSD and POPC were parameterized using Antechamber from AmberTools 11.¹² To obtain the systems of LSD in complex with both 5-HT receptors, we simply modified ergotamine to create LSD. This involved deletion of N-substitution of the lysergamide of ergotamine and subsequent addition of two ethyl groups yielding N,N-diethyl-lysergamide (LSD). The system was parameterized as mentioned above.

Molecular dynamics simulations. Simulations were performed using ACEMD¹³ using the following protocol: In a first stage, each system was submitted to a minimization procedure for 3000 steps. In a second stage, the system was equilibrated using the NPT ensemble with a target pressure equal to 1.01325 bar, a time-step of 2 fs and using the RATTLE algorithm for the hydrogen atoms. In this stage, the harmonic constraints applied to the heavy atoms of the protein and ligand were progressively reduced from an initial value of 10 kcal/mol/Å until an elastic constant force equal to 0 kcal/mol and the temperature was increased to 300 K. The purpose of this relaxation phase is to allow for a complete adjustment of membrane lipids to the receptor, thus filling non-physiological gaps between receptor and membrane lipids. All the simulations were conducted using the same non-bonded interaction parameters, with a cutoff of 9 Å, a smooth switching function of 7.5 Å and the non-bonded pair list set to 9 Å. The periodic boundary conditions were set to a size of 78×78×88, and for the long range electrostatics we used the PME methodology with a grid spacing of 1 Å. In a third stage, production phases were performed using the NVT ensemble with aforementioned parameters but a time-step of 4 fs, and a hydrogen scaling factor of 4. This timestep is possible due to the implementation of the hydrogen mass repartitioning scheme in the ACEMD code.¹³ Simulations were performed for 500 ns for individually generated starting structures (by performing stages 1 to 3). Importantly, individually-generated starting structures allow a more robust statistical analysis and thus the detection of relevant dynamic events that are independent from the starting structures.

b) δ -Opioid receptor (δ -OR)

System preparation: The systems were generated using CHARMM-GUI.^{10,14} We have used the crystal structure of δ -opioid receptor in complex with naltrindole [PDB code: 4N6H].¹⁵ To generate the mutated D2.50A δ -opioid receptor, we have inserted the mutation using the CHARMM-GUI panel. A sodium ion was placed in the allosteric binding site of the wild-type receptor. It was not placed in the D2.50A receptor, as mutational data show, that this mutation makes the receptor sodium insensitive.¹⁶ The receptor was embedded in a $\sim 80 \times 80 \text{ \AA}$ POPC bilayer and solvated with TIP3 water molecules. The ionic strength of the solution was kept at 0.15 M with NaCl ions. Parameters for the simulation were obtained from the CHARMM36 forcefield.¹⁷ Parameters for the ligand were assigned from the CGenFF forcefield automatically by the ParamChem tool implemented in CHARMM-GUI.^{18,19}

Molecular dynamic simulations: The systems were first equilibrated in conditions of constant pressure (NPT, 1.01325 bar) for 20 ns, preceded by an initial 1000 step minimization. After the NPT step we have carried out simulations in conditions of constant volume (NVT) in 100 replicates of 128 ns and in 3 replicates of 1500 ns for each of the systems. The simulations were run in ACEMD.²⁰ In both steps we used a time-step of 4 fs. Such a time-step was possible due to the hydrogen mass repartitioning scheme being employed in ACEMD.²¹ A non-bonded interaction cutoff was set at 9 \AA . A smooth switching function for the cut-off was applied, starting at 7.5 \AA . The size of the cell was set to prevent non-bonding interactions between the protein and its periodic boundary image. Long-distance electrostatic forces were calculated using the Particle Mesh Ewald algorithm. The algorithm had grid spacing of 1 \AA . The bond lengths of hydrogen atoms were kept constrained using the RATTLE algorithm. Simulations were carried out at a temperature of 300K.

3. Ligand-residue interaction analysis

5-HT_{1B} and 5-HT_{2B} residues interacting with the ligand ergotamine. Interaction data of crystalized structures, extracted from GPCRdb, indicate the presence or absence of interaction and its type: polar, aromatic, hydrophobic (hydroph.) or accessible (acc., contact of the ligand with a residue backbone) (Table S1). Data obtained from the MD ligand-residue interaction frequency analysis, implemented at the GPCRmd viewer, indicate the average frequency in which the interaction is found along the simulations. The MD interaction frequency analysis was performed at different distance thresholds: 3, 3.5, 4 and 4.5 Å (total of 5 trajectories of 5000 frames analyzed for each receptor at each threshold, each trajectory represented a simulation time of 500 ns). Cell colors indicate the stability regarding frequency of the interaction, ranging from darkest red (no interaction) to darkest green (interaction found at all frames). Residues are indicated using a combination of Ballesteros–Weinstein and structural-based numbering schemes, when available. Positions where numeration of the two residues does not coincide are indicated as 5-HT_{1B} position/5-HT_{2B} position, and the same method is applied when the interaction types differ.

Table S1 Ligand-Residue interaction analysis showing the residues of 5-HT_{1B} and 5-HT_{2B} receptors interacting with ergotamine.

Seq. position (5-HT _{1B} /5-HT _{2B})	GPCRdb	3Å 5-HT _{1B}	3Å 5-HT _{2B}	3.5Å 5-HT _{1B}	3.5Å 5-HT _{2B}
TYR2/ILE0	-	0.28	0.00	8.51	0.00
TYR4/GLU2	-	0.11	0.00	12.90	0.03
2.64x63	Accessible	0.01	0.00	1.11	0.46
2.65x64	-	0.00	0.00	0.00	0.44
3.28x28	Accessible	0.00	0.00	2.76	4.24
3.29x29	Accessible	0.00	0.00	0.62	1.17
3.32x32	Polar	98.24	96.78	99.85	99.95
3.33x33	Hydrophobic	0.05	0.03	16.60	8.75
3.36x36	Hydroph/Acc	0.00	1.94	22.16	55.14
3.37x37	Polar	5.40	9.61	20.98	39.27
3.40x40	Accessible	0.00	0.00	0.02	0.00
4.56x56	Accessible	0.00	0.00	0.10	0.01
45.50x50	-	0.16	0.61	17.38	30.78
45.51x51	Hydrophobic	1.12	0.49	82.92	63.75
45.52x52	Polar	53.30	65.36	94.70	97.50
ASN166/THR164	Accessible	0.00	0.70	0.03	10.91
THR167/LYS165	Accessible	0.41	0.05	18.40	6.26
5.35x36	Accessible	0.00	0.00	0.00	0.04
5.36x37	Accessible	0.00	0.00	0.00	0.01
5.38x39	Hydrophobic	0.00	0.47	0.19	48.14
5.39x40	Hydrophobic	0.00	0.03	0.03	23.79
5.42x43	Accessible	29.87	9.65	71.25	31.73
5.43x44	Accessible	0.10	0.66	21.50	18.61
5.46x461	Hydrophobic	1.00	1.58	54.99	61.39
6.48x48	Hydroph/Acc	0.29	0.07	27.34	10.18
6.51x51	Hydrophobic	0.58	0.11	74.99	48.67
6.52x52	Aromatic	0.00	0.00	12.44	12.51
6.54x54	-	0.00	0.00	0.09	0.00
6.55x55	Accessible	0.01	2.10	9.96	42.18
6.58x58	Hydroph/Acc	0.10	0.01	21.68	9.48
6.59x59	Accessible	0.00	0.00	0.24	1.11
TRP245/7.28x27	-	0.29	0.38	24.31	1.86
7.32x31	Accessible	0.07	37.63	20.09	62.72
7.35x34	Hydrophobic	1.04	0.34	68.22	38.18
7.36x35	Accessible	0.04	0.12	14.83	14.59
7.39x38	Accessible	4.21	0.01	50.70	18.59
7.43x42	Accessible	0.57	0.24	18.72	23.36

Table S1 Ligand-Residue interaction analysis showing the residues of 5-HT_{1B} and 5-HT_{2B} receptors interacting with ergotamine (continued).

Seq. position (5-HT _{1B} /5-HT _{2B})	4Å 5-HT _{1B}	4Å 5-HT _{2B}	4.5Å 5-HT _{1B}	4.5Å 5-HT _{2B}
TYR2/ILE0	14.69	0.00	18.38	0.00
TYR4/GLU2	29.05	0.15	32.76	0.47
2.64x63	5.94	3.94	24.34	6.82
2.65x64	0.00	5.38	0.03	11.14
3.28x28	41.69	39.66	81.48	68.39
3.29x29	13.33	25.38	51.96	61.77
3.32x32	99.88	99.99	99.96	100.00
3.33x33	91.42	77.04	99.93	99.08
3.36x36	91.91	95.83	99.72	99.86
3.37x37	39.63	67.90	68.34	91.38
3.40x40	0.61	0.07	2.10	1.09
4.56x56	2.46	0.34	12.54	4.07
45.50x50	55.36	65.67	86.59	83.14
45.51x51	98.06	98.48	99.17	99.58
45.52x52	96.16	99.42	96.26	99.86
ASN166/THR164	0.35	21.82	1.19	44.29
THR167/LYS165	51.76	36.56	61.45	56.34
5.35x36	0.00	0.40	0.00	1.51
5.36x37	0.00	0.24	0.00	0.86
5.38x39	1.62	90.84	5.89	98.72
5.39x40	1.18	80.15	10.95	98.04
5.42x43	91.21	67.38	97.26	88.42
5.43x44	67.98	63.09	84.59	92.55
5.46x461	94.98	95.20	99.62	99.38
6.48x48	82.79	48.20	97.42	79.19
6.51x51	99.90	97.89	100.00	99.92
6.52x52	83.20	89.27	98.68	99.61
6.54x54	7.89	0.04	29.82	0.25
6.55x55	49.35	73.72	71.45	92.37
6.58x58	76.76	52.26	91.91	76.36
6.59x59	2.54	7.89	3.80	11.58
TRP245/7.28x27	73.59	7.67	88.69	12.11
7.32x31	76.80	81.47	90.76	88.50
7.35x34	99.02	93.58	99.99	99.88
7.36x35	70.39	86.28	89.97	99.15
7.39x38	93.33	71.42	98.62	93.65
7.43x42	45.66	52.80	63.16	70.64

δ -OR residues interacting with the ligand naltrindole, considering both the wild type receptor (wt) and the Asp2x50Ala mutant (mt). Interaction data of crystalized structures, extracted from GPCRdb, indicate the presence or absence of interaction and its type: H bond (* indicates charge-assisted H bond), aromatic, hydrophobic or accessible (contact of the ligand with a residue backbone) (Table S2). Data obtained from the MD ligand-residue interaction frequency analysis, implemented at the GPCRmd viewer, indicate the frequency in which the interaction is found along the simulations.

Table S2 Ligand-Residue interaction analysis showing the residues of wild type and Asp2x50Ala mutant δ -OR interacting with naltrindole.

	GPCRdb	3Å wt	3Å mt	3.5Å wt	3.5Å mt
3.32x32	H bond *	60.030	63.890	81.320	85.350
3.33x33	H bond	25.200	22.470	85.940	80.240
3.35x35	-	0.000	0.060	2.860	6.380
3.36x36	Hydrophobic	0.000	0.000	13.020	12.630
5.39x40	Accessible	0.070	0.130	20.050	14.020
5.42x43	Accessible	0.070	0.060	31.770	30.240
6.48x48	Hydrophobic	0.000	0.000	8.590	7.640
6.51x51	Hydrophobic	0.000	0.000	4.300	4.480
6.52x52	Aromatic	0.330	0.250	25.260	20.450
6.54x54	-	0.000	0.000	0.460	0.250
6.55x55	Hydrophobic	0.000	0.000	20.310	19.570
6.58x58	Aromatic	0.070	0.060	13.480	13.640
6.59x59	Accessible	0.000	0.000	0.070	0.250
7.35x34	Accessible	0.000	0.000	2.670	1.640
7.38x37	-	0.000	0.000	0.460	0.380
7.39x38	Accessible	0.000	0.000	15.040	10.350
7.42x41	Accessible	0.000	0.130	13.930	8.520
7.43x42	Accessible	0.000	0.060	15.690	12.370

Table S2 Ligand-Residue interaction analysis showing the residues of wild type and Asp2x50Ala mutant δ -OR interacting with naltrindole (continued).

	4Å wt	4Å mt	4.5Å wt	4.5Å mt
3.32x32	95.960	96.780	99.480	99.430
3.33x33	97.070	96.090	99.220	98.740
3.35x35	11.780	19.950	23.760	32.640
3.36x36	78.450	82.320	92.710	96.210
5.39x40	73.700	62.880	94.210	89.650
5.42x43	84.110	81.500	96.090	94.890
6.48x48	72.850	66.410	88.670	86.300
6.51x51	76.890	74.240	97.790	97.540
6.52x52	79.170	75.380	96.030	95.140
6.54x54	6.180	8.020	18.290	21.150
6.55x55	91.600	91.790	99.410	99.870
6.58x58	65.620	70.390	86.460	89.650
6.59x59	0.460	0.950	1.040	1.770
7.35x34	31.250	28.030	67.510	63.950
7.38x37	4.430	1.700	12.630	4.730
7.39x38	70.770	56.440	90.950	83.650
7.42x41	72.660	53.980	94.140	81.060
7.43x42	77.280	68.180	89.520	84.600

The MD interaction frequency analysis was performed at different distance thresholds: 3, 3.5, 4 and 4.5 Å (one trajectory of 5000 frames analyzed for each receptor at each threshold, representing a simulation time of 500 ns). Cell colors indicate the strength of the interaction, ranging from darkest red (no interaction) to darkest green (interaction found at all frames).

Residues are indicated using a combination of Ballesteros–Weinstein and structural-based numbering schemes, when available.

4. Intramolecular interaction analysis

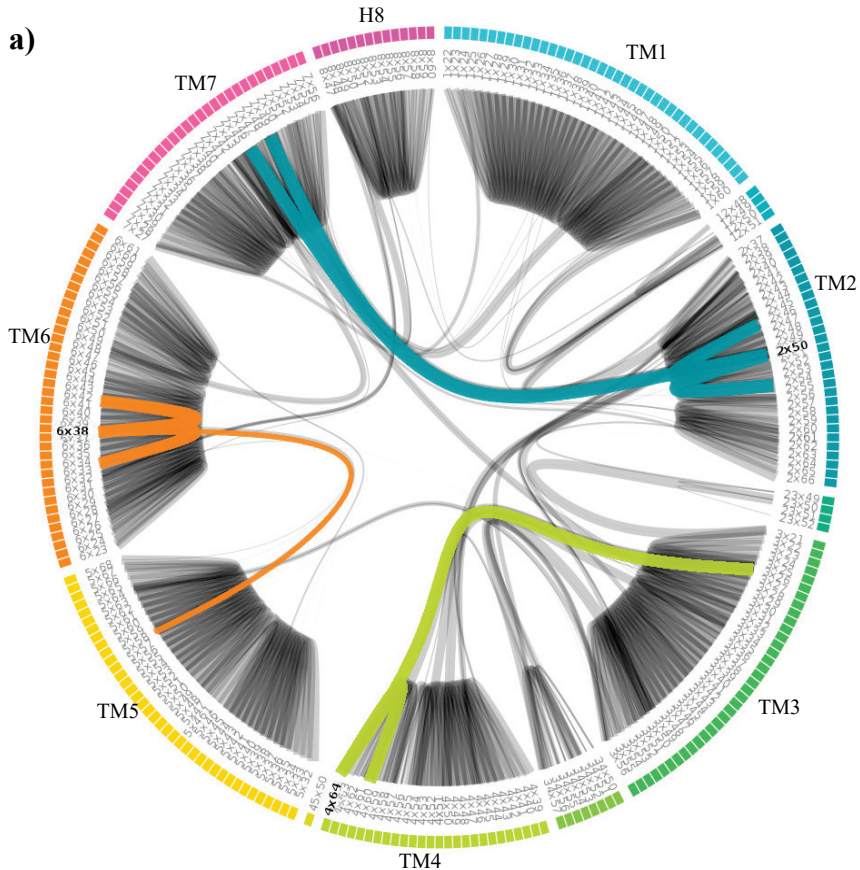


Figure S10 Intramolecular hydrogen bonds of δ -OR. Flare plots representing the network of hydrogen bonds among the residues of wild type δ -OR (a) and Asp2x50Ala mutant (b). Helices are indicated in colors ranging from light blue transmembrane helix 1, TM1) to purple (helix 8, H8). Residue positions are expressed in the GPCRdb structure-based numbering scheme. Lines connecting residues represent hydrogen bonds formed between them, and their width the frequency of such contacts. Residue contacts showing most significant differences between the two receptors are highlighted according to the helix where they belong: 2x50, which is the mutated residue (dark blue), 4x64 (light green) and 6x38 (orange).

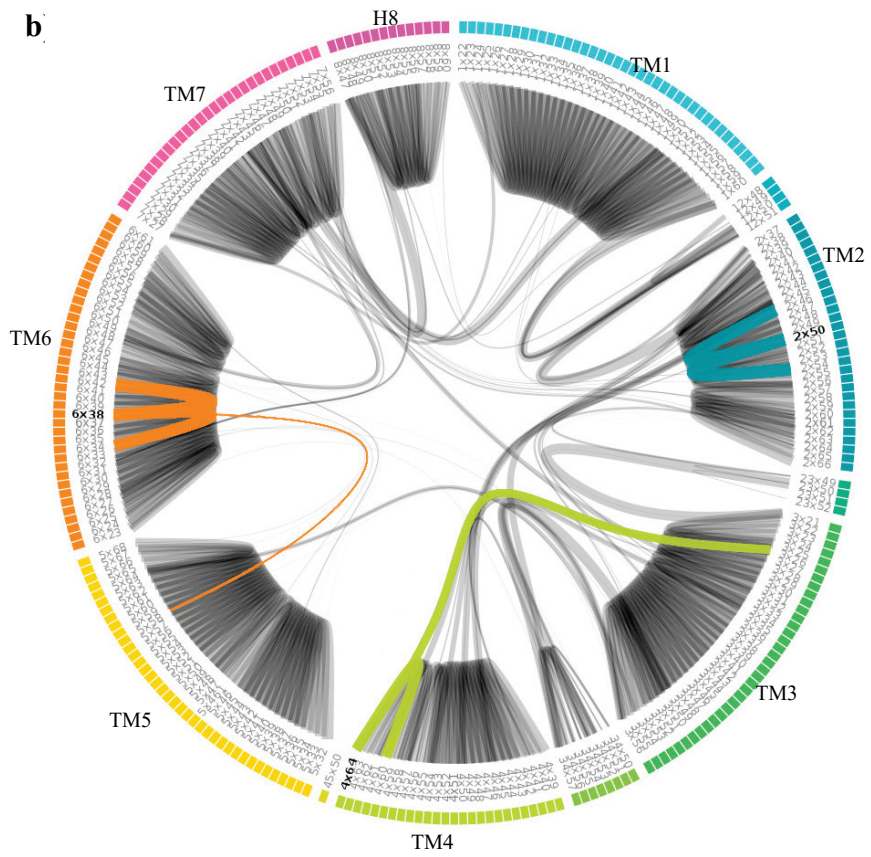


Figure S10 Intramolecular hydrogen bonds of δ -OR (continued).

5. References

1. Heller, S. R., McNaught, A., Pletnev, I., Stein, S. & Tchekhovskoi, D. InChI, the IUPAC International Chemical Identifier. *J. Cheminform.* **7**, 23 (2015).
2. Southan, C. *et al.* The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. *Nucleic Acids Res.* **44**, D1054–D1068 (2016).
3. Gilson, M. K. *et al.* BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **44**, D1045–D1053 (2016).
4. Martí-Solano, M., Sanz, F., Pastor, M. & Selent, J. A dynamic view of molecular switch behavior at serotonin receptors: implications for functional selectivity. *PLoS One* **9**, e109312 (2014).
5. Wang, C. *et al.* Structural Basis for Molecular Recognition at Serotonin Receptors. *Science (80-.)*. **340**, 610–614 (2013).
6. Wacker, D. *et al.* Structural features for functional selectivity at serotonin receptors. *Science* **340**, 615–9 (2013).
7. Ballesteros, J. A. & Weinstein, H. in *Methods in Neurosciences* **25**, 366–428 (1995).
8. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins Struct. Funct. Bioinforma.* **61**, 704–721 (2005).
9. Chemical Computing Group ULC. Molecular Operating Environment (MOE). (2014). at <https://www.chemcomp.com/MOE-Molecular_Operating_Environment.htm>
10. Wu, E. L. *et al.* CHARMM-GUI membrane builder toward realistic biological membrane simulations. *Journal of Computational Chemistry* **35**, 1997–2004 (2014).
11. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
12. Case, D. A. *et al.* AMBER 12. (2012). at <<http://ambermd.org/>>
13. Harvey, M. J., Giupponi, G. & Fabritiis, G. De. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **5**, 1632–1639 (2009).

14. Lee, J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **12**, 405–413 (2016).
15. Fenalti, G. *et al.* Molecular control of δ -opioid receptor signalling. *Nature* **506**, 191–196 (2014).
16. Fenalti, G. *et al.* Molecular control of δ -opioid receptor signalling. *Nature* **506**, 191–196 (2014).
17. Klauda, J. B. *et al.* Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
18. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–90 (2010).
19. Yu, W., He, X., Vanommeslaeghe, K. & MacKerell, A. D. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.* **33**, 2451–2468 (2012).
20. Harvey, M. J., Giupponi, G. & De Fabritiis, G. ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* **5**, 1632–1639 (2009).
21. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **20**, 786–798 (1999).

Chapter 4

Rapid network rearrangements drive the initial phase of β -arrestin signaling in the δ -opioid receptor

In this chapter, we present the work in form of a journal article:

Stepniewski, T.M, Rodríguez-Espigares I., Martí-Solano, M., Troya-Bruguer, A., Torrens-Fontanals, M., De Fabritiis, G., Filipek, S. & Selent, J. Rapid network rearrangements drive the initial phase of β -arrestin signaling in the δ -opioid receptor. 2017 (submitted).

Summary: Here, we propose a model for the mechanism behind the initiation of β -arrestin signaling in the δ -opioid receptor (δ -OR) which might also apply for other class A GPCRs. This model is based on data yielded by all-atom molecular dynamics (MD) simulations of naltrindole in complex with wild-type (WT) δ -OR and D2.50A δ -OR. This mutation converts the naltrindole (a δ -OR antagonist) into a potent β -arrestin agonist. The study of two hundred replicas obtained shows that rapid network rearrangements at a nanosecond timescale drive the initial phase of β -arrestin signaling in δ -OR. The initial part of β -arrestin signaling seems to begin with the destabilization of contacts between transmembrane helix (TM) 2 and 7 which are likely caused by changes in intramolecular water networks and interhelical interactions. This produces higher TM7 fluctuations near the binding site of intracellular signaling proteins, a behavior that has been linked to β -arrestin biased signaling by biophysical experiments. To obtain a mechanistic explanation of this phenomenon, I focused on the Markov State Model analysis of the TM7. My analysis identified two new main conformational states in D2.50A δ -OR characterized by a

rotation of Y7.53 similar to the one found in inactive rhodopsin⁶⁹ and arrestin coupled rhodopsin crystal structures.⁷⁰ Altogether, these results show that molecular dynamics multiple simulation replicates of one hundred nanoseconds can be useful to predict the functional outcome of a given drug.

Rapid network rearrangements drive the initial phase of β -arrestin signaling in the δ -opioid receptor

Tomasz Maciej Stepniewski^{[a]‡}, Ismael Rodríguez-Espigares^{[a]‡},
Maria Martí-Solano^{[a][b]}, Anna Troya-Bruguer^[a], Mariona Torrens-
Fontanals^[a], Gianni De Fabritiis^[c], Slawomir Filipek^[d] and Jana
Selent^{[a]*}

Abstract: Signaling bias is an emerging concept for obtaining more efficacious drugs targeting GPCRs. However, its structural fundamentals remain poorly understood. Here, we use all-atom molecular dynamics simulations with hundreds of replicas to show that rapid network rearrangements at nanosecond timescale drive the initial phase of β -arrestin signaling in the δ -opioid receptor. First steps involve a destabilization of contacts between transmembrane helix (TM) 2 and 7 caused by alterations in intramolecular water networks and interhelical interactions. Together this translates into higher TM7 fluctuations near the binding site of intracellular signaling proteins, a behavior that has been associated to β -arrestin biased signaling by biophysical experiments. Ultimately, the existence of specific network rearrangements during the initial signaling phase creates new opportunities to explore ligand binding effects on receptor signaling at shorter timescales than currently thought.

G-protein coupled receptors (GPCRs) are flexible transmembrane proteins that exist in an equilibrium of conformations with different signaling properties. Certain

GPCR ligands, known as biased agonists, can stabilize receptor conformations that preferentially couple to specific intracellular signaling proteins. Given their pathway selectivity, biased agonists could act as more efficacious and safer drugs.^[1] However, for their rational design, we first need to understand the molecular determinants of GPCR signaling bias.

In the present study, we address this question by studying β -arrestin (β -arr) bias at the δ -opioid receptor (δ OR), a member of class A GPCRs, using molecular dynamics simulations. We take advantage of a D2.50A receptor mutant in which the antagonist naltrindole converts into a potent β -arr biased agonist.^[2] The mutated site is a well-known allosteric sodium binding site in class A GPCRs.^[3,4] Thus, inducing sodium decoupling seems to structurally mimic β -arr biased agonism in the δ OR.^[2]

Previous computational studies identified the formation of a water channel as an key event in GPCR activation.^[5] To assess its importance in β -arr signaling events, we studied the β -arr biased mutant D2.50A and the inactive wild-type (WT) δ OR in complex with naltrindole

-
- [a] TM Stepniewski, I Rodríguez-Espigares, Dr. M Martí-Solano, A Troya Bruguer, M Torrens-Fontanals, Dr. J Selent
Research Programme on Biomedical Informatics, Hospital del Mar Medical Research Institute & Department of Experimental and Health Sciences, Pompeu Fabra University
Dr. Aiguader 88, Barcelona, E 08003 (Spain)
E-mail:jana.selent@upf.edu
- [b] Dr. M Martí-Solano
Department of Pharmaceutical Chemistry, Philipps-University, Marbacher Weg 6, Marburg 35032 (Germany)
- [c] Prof. Dr. G De Fabritiis
Department of Experimental and Health Sciences, Pompeu Fabra University
Dr. Aiguader 88, Barcelona, E 08003 (Spain)
E-mail:jana.selent@upf.edu
- [d] Prof. Dr. S Filippek
Laboratory of Biomodeling, Faculty of Chemistry & Biological and Chemical Research Centre, University of Warsaw Warsaw (Poland)

± Authors contributed equally to this work.

Supporting information for this article is given via a link at the end of the document.

performing three molecular dynamics (MD) simulation replicates of 1.5 μ s per system. Interestingly, these first simulations show differences in local water occupancies comparing both receptors (Figure S1B and C, highlighted by arrows).

To further explore these observations, we next carried out 100 replicates of 128 ns for each system (Table S1). This short time interval is suitable to capture rapid water rearrangements (Figure S2). Results from these simulations suggest that the D2.50A mutation alters the water network of the connector region located between the orthosteric ligand binding site and the intracellular coupling site of signaling proteins. In the WT δ OR, we find pronounced water occupancies in positions 1 to 3 and 5, 6 (Figure 1A). These positions correlate well with locations of waters in the δ OR crystal structure (Figure S3A), supporting the reliability of our model. In contrast, in the β -arr biased δ OR, we find an overall destabilization of the water network, in particular in positions 2, 3 and 5 (Figure 1B). Destabilization of these waters indicates higher solvent fluctuation in these positions (Figure S4B). In the inactive WT structure, water molecule 2 forms a hydrogen bond with N3.55, S7.46 and water molecule 3, whereas water molecule 3 is further coordinated via W6.48, N7.45 and the sodium ion (Figure 1A). Despite the overall decrease of water network stability in the β -arr biased δ OR, we find one additional water molecule of high occupancy in position 4, which replaces the sodium ion found in the WT δ OR. Our data clearly suggest that the D2.50A mutation results in a significant alteration of the water network. However, it is not clear how this contributes to β -arr bias. It is tempting to speculate that distinct changes in the

conserved water network transmit a structural message through the receptor. To address this, we studied the interhelical hydrogen bonding network of the WT and the β -arr biased δ OR using Flare plots,^[6] which emphasize differences between receptors related to conserved activation motifs (D2.50, D3.32, Y7.43, N7.46 N7.49, R6.32 and D8.47, Figure 2AB). Figure 2C shows how these differences contribute to altered dynamic properties of the δ OR in terms of C α root mean square fluctuation (RMSF). This analysis reveals overall higher fluctuation (in red, Figure 2C) in the β -arr biased δ OR compared to the inactive δ OR. Surprisingly, the most pronounced difference is not in the mutated D2.50A site in TM2 but in the middle-lower part of TM7. This area contains three highly conserved residues - N7.45, S7.46 and N7.49 (Figure S3), with N7.49 being part of the conserved NPXXY motif. All three residues are in proximity to the mutated residue 2.50 and involved in direct or indirect interaction with the sodium ion.^[2] Distances between C α atoms of N7.45, N7.49 and the C α atom of residue 2.50 (Figure 2B, distance 3 and 4) fluctuate significantly more in the β -arr biased δ OR compared to the WT δ OR (Figure 2C). Higher fluctuations of N7.45 and N7.49 (Figure 1) are likely the result of a disruption of sodium and water-mediated interactions (molecules 2, 3 and 5). In addition, there is a loss of direct interaction between N7.45, S7.46 and D2.50.

In light of these observations, we next analyzed how far this disturbance extends within the δ OR during the initial phase of β -arr signaling. Analysis of 100 short replicates indicates that there is substantial propagation at a nanosecond scale to extra- and intracellular sides.

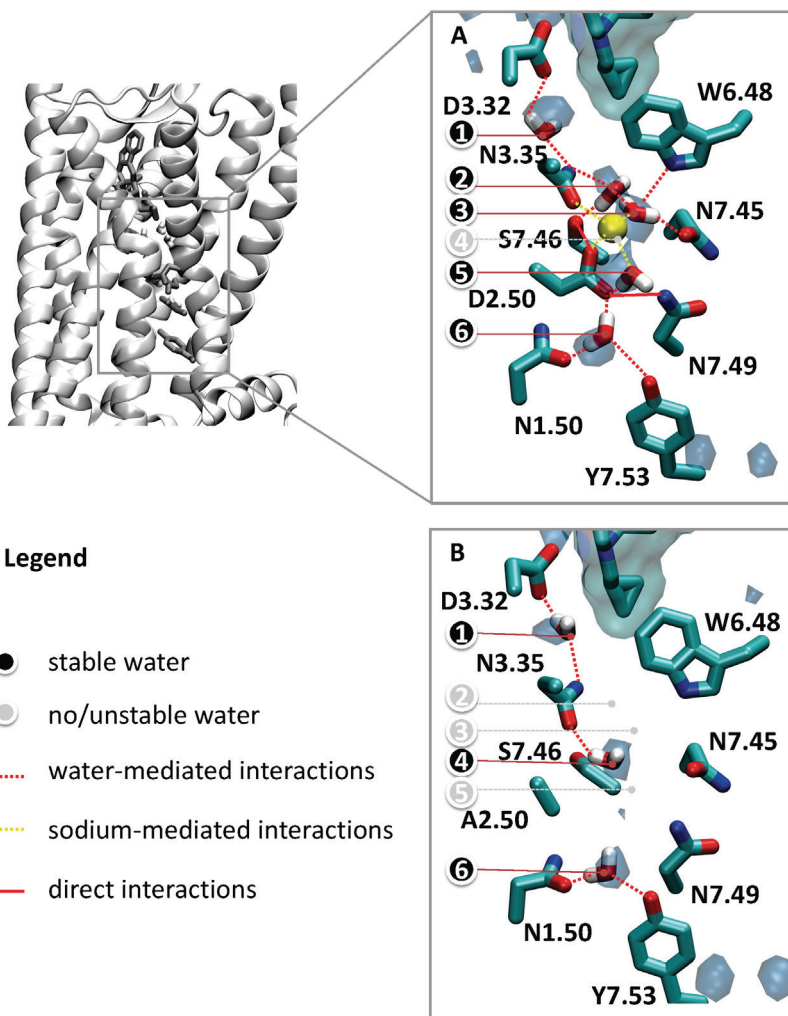


Figure 1. Water occupancy map of the connector region for the WT (A) and β -arr biased δ OR (B) plotted at 40% water occupancy. The connector region links the orthosteric binding site (D3.32) to receptor regions close to the intracellular coupling site of signaling proteins (Y7.53). Direct polar interactions between residues are displayed as solid lines whereas indirect (water or sodium-mediated) polar interactions are displayed as dashed lines. A summary of absolute values of water occupancies can be found in Figure S4A.

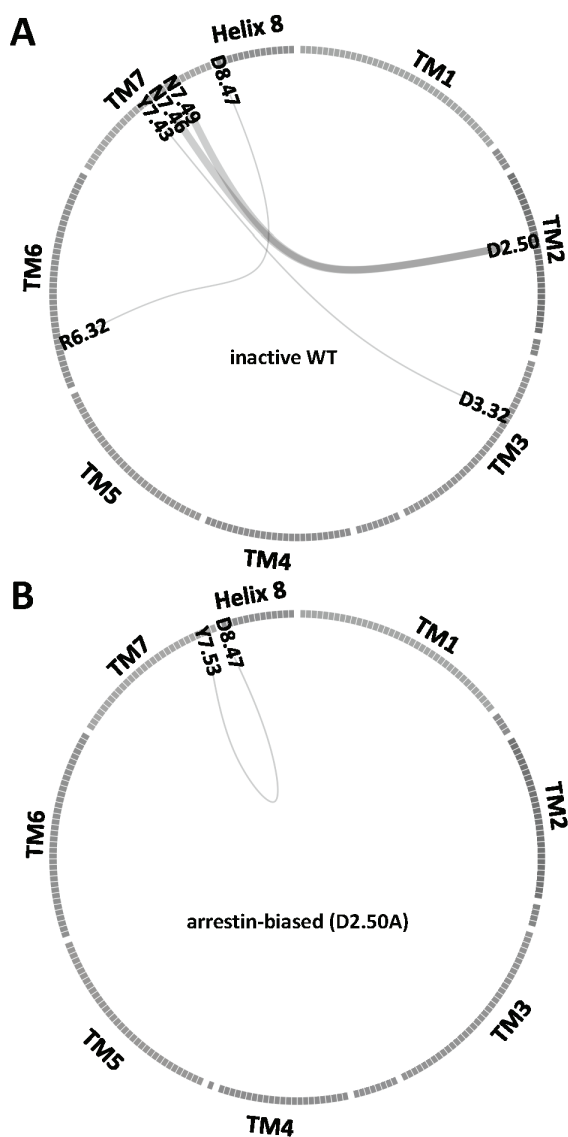


Figure 2. Dynamics for β -arr biased and inactive δ OR. Intra- and interhelical hydrogen bonding network for the inactive WT (A) and the β -arr biased (B) δ OR. Line thickness relates to how strongly an interaction is preserved over the simulation time in comparison to its preservation in the other receptor. Complete network plots are found in Figure S5.

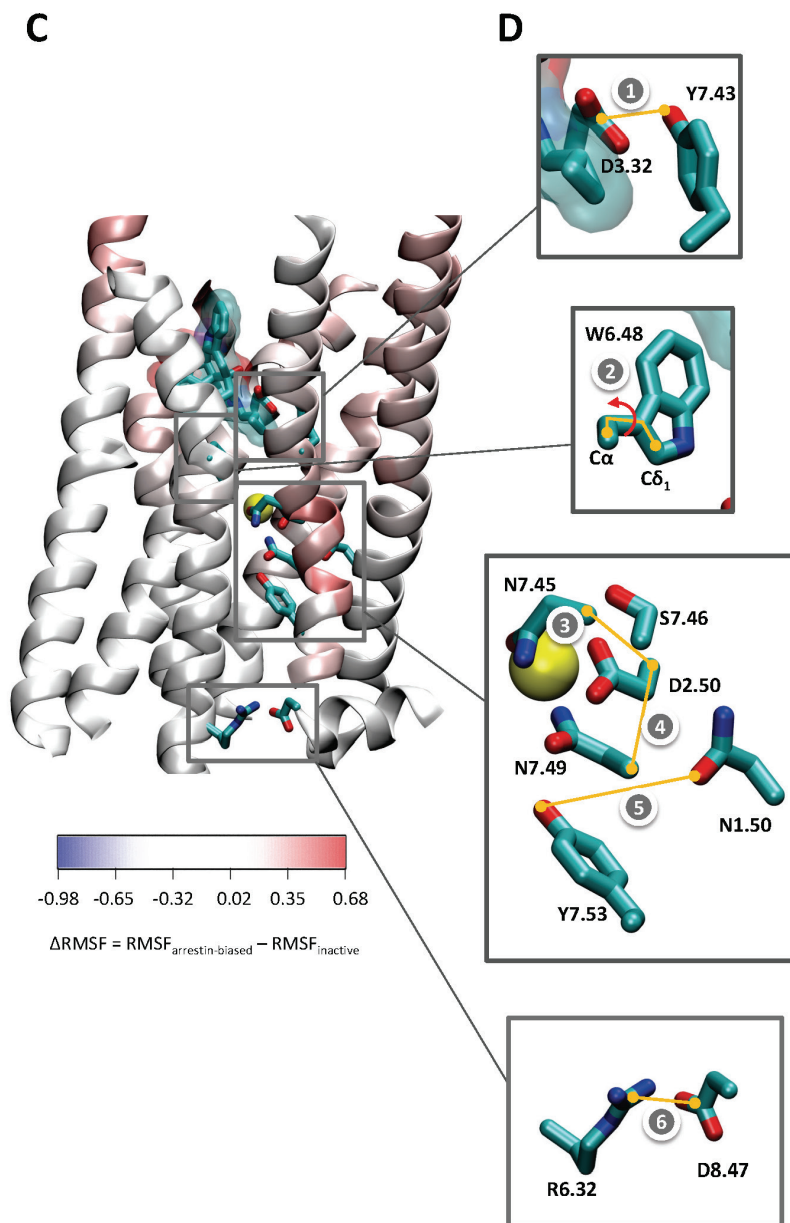


Figure 2. (continued) (C) Difference of C α root mean square fluctuation ($\Delta\text{RMSF} = \text{RMSF}_{\beta\text{-arr biased}} - \text{RMSF}_{\text{inactive}}$) between receptors. Positive values (red color) indicate regions of higher fluctuation in the β -arr biased receptor. (D) Structural depiction of selected relevant residues that link the orthosteric site towards the intracellular δ OR end. (E) Distances and torsion angles of selected residues over the studied simulation time (6.4 μs) plotted using a 4 ns window (1600 data points).

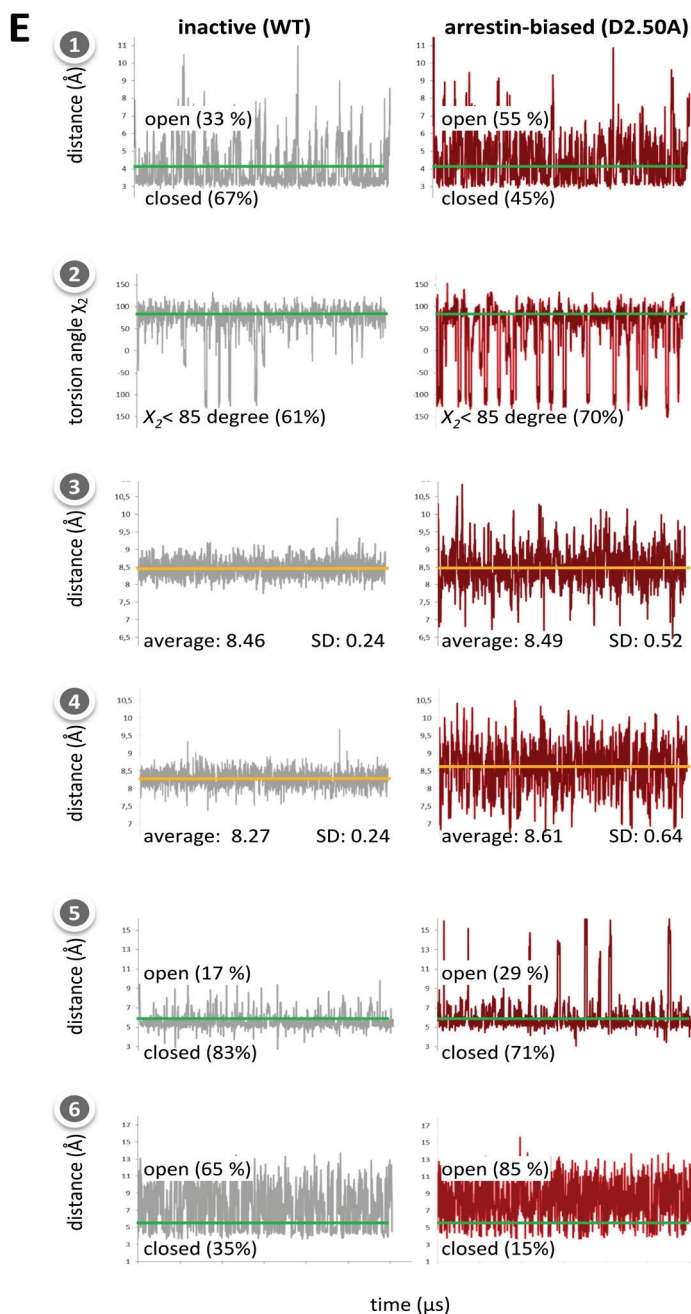


Figure 2. (continued) (E) Distances and torsion angles of selected residues over the studied simulation time (6.4 μ s) plotted using a 4 ns window (1600 data points).

Towards the extracellular side, we observe a higher fluctuation of the so called “rotamer toggle switch” W6.48

(Figure 2E, torsion angle 2) in the β -arr biased δ OR as well as a higher opening frequency of the D3.32-Y7.43 bridge

(55%) compared to the inactive state (33%) (Figure 2C, distance 5). Higher mobility of W6.48 seems to be a result of the destabilization of water molecule 3, which in the WT receptor is stabilized by the sodium ion. This goes along with recent computational studies.^[7] Moreover, movement of W6.48 has been previously related to GPCR activation.^[8] In this respect, our study suggests that higher W6.48 fluctuation is also an initial step of β -arr signaling. In fact, mutagenesis of W6.48 into L, F, A or D in the δ OR results in a loss of ligand-specific activity for the β -arr pathway, while G-protein signaling is maintained.^[7] Tyrosine 7.43 is another highly conserved residue, which forms an ionic lock with D3.32 in the δ OR.^[2] Previous work on the μ and κ OR highlights the importance of this lock in opioid receptor functionality.^[9] In our simulations, this lock remains more frequently open in the β -arr biased δ OR (Figure 2C, top). This is the result of a destabilized network between the conserved residues D3.32, N3.35, W6.48, S7.46 and water molecules 1, 2 and 3 (Figure 1A and B).

Towards the intracellular side, we observe a similar tendency of signal propagation. N1.50 and Y7.53 form a conserved lock mediated via a water molecule at the intracellular end of the δ OR (Figure S6). The studied lock is part of the NPXXY^{7.53} motif and forms an important gateway for water molecules entering the receptor from the intracellular side.^[5] In our simulations, we find that this lock, which is maintained in a closed state via water molecule 6 (Figure 1A and B), fluctuates more in the β -arr biased δ OR, visiting the open state during 29% of the simulation time compared to 17% in the inactive δ OR simulations (Figure 2C, plot 5). A similar trend is observed for the polar lock between TM6 (R6.32) and helix 8 (D8.47) (Figure 2C, plot 6).

All in all, our data indicate that initial destabilization of the connector region (D2.50A mutation) causes higher fluctuations in the middle-lower part of TM7 towards the intracellular side. Importantly, higher TM7 fluctuation has been related to β -arr bias using ¹⁹F-NMR experiments.^[10] In their study, authors report that β -arr biased ligands predominantly impact the conformational state of TM7 whereas changes in TM6 state are related to G protein signaling. In line with this finding, we propose that an initial step of β -arr bias involves higher fluctuation of TM7 at a nanosecond timescale. In turn, this might facilitate global conformational changes leading to β -arr coupling in a timescale of seconds.^[11]

To study in more detail TM7 fluctuations observed in the β -arr biased mutant, we used a Markov State Model (MSM) analysis. This analysis revealed different populations of δ OR macrostates (Figure S7). In particular, we find two macrostates, which are exclusively found for the β -arr biased δ OR and involve a partial anti-clockwise rotation of Y7.53 (Figure S8, state 4 and 5). This rotation is sampled for 1.2 μ s within a total of 12.8 μ s and is similar to the observed rotation in the arrestin-coupled rhodopsin structure (Figure S8C). Such Y7.53 rotation impacts also the formation of new intrahelical bonds between TM7 and helix 8 as captured in the Flare plots (Figure 2B). In addition, the adjacent residue towards the intracellular side (7.54) significantly changes its environment. This finding goes along with previously mentioned ¹⁹F-NMR experiments, which report pronounced shifts of the NMR signal in position 7.54 upon binding of β -arr biased agonists.^[10] Note that the observed Y7.53 rotation (state 4 and 5) occurs sporadically during receptor fluctuation and indicates altered δ OR dynamics, which might initiate

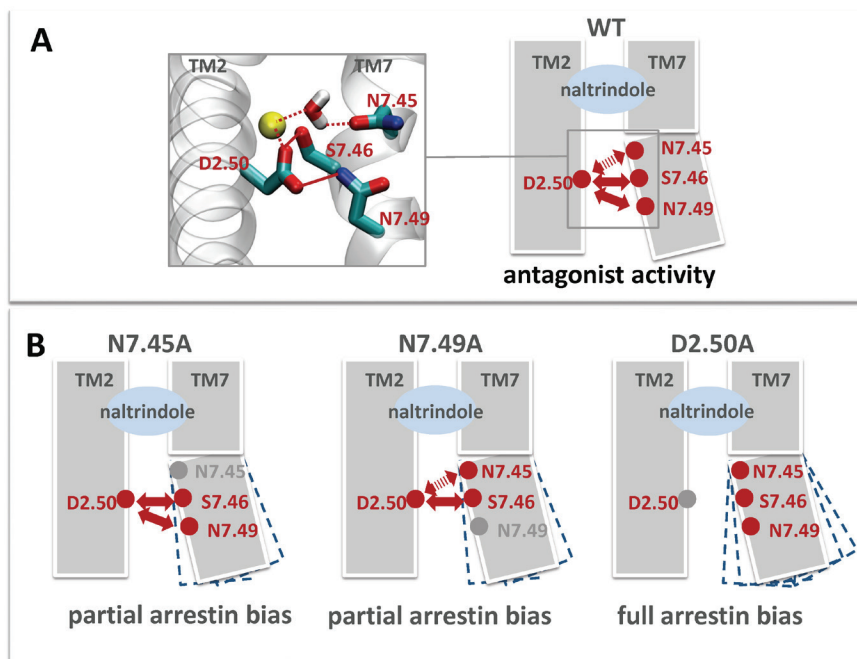


Figure 3. Potency of β -arrestin bias correlates with the destabilization of TM2-TM7 contacts. Direct interactions are shown in solid lines and water-mediated interactions in dashed lines.

global conformational changes leading to β -arr coupling.

Mechanistically (Figure 3A), higher TM7 fluctuation in the δ OR can be related to the loss of three important interactions between TM2 and TM7 comprising sodium and water-mediated indirect (dashed line) and direct (solid line) interactions, which stabilize the antagonist-bound δ OR. This finding is supported by the observation that mutations of residues linking TM2 and TM7 (e.g. N7.45 and N7.49) also promote β -arr bias in the δ OR.^[2] Surprisingly, TM7 mutations result in lower β -arr bias than TM2 mutations (Table S2). In this respect, our data provides a plausible explanation for the different impact of mutations on β -arr bias (Figure 3). In the fully inactive δ OR, TM7 is stabilized by three interactions between D2.50 and N7.45, S7.46 and

N7.49 (Figure 3A). These three interactions are lost upon D2.50A mutation (Figure 3B), which strongly weakens TM2-TM7 interaction and consequently increases TM7 fluctuation leading to full β -arr bias. In contrast, mutation of N7.45 or N7.49 abolishes only a single interaction. This would cause only a partial reduction in TM7 fluctuation and thus partial β -arr coupling which is in line with experimental data (Table S2).

The relevance of the connector region around the sodium-binding site D2.50 for signaling bias appears to be partially conserved among class A GPCRs. An interesting case is the neurokinin 1 receptor.^[12] Despite being sodium-insensitive (i.e. no sodium binding at position 2.50), E2.50A mutation transforms the natural agonist SP, into a β -arr biased agonist. Residue E2.50

stabilizes TM2-TM7 interactions by direct contacts with S7.45 and N7.49. Moreover, mutating either residue N7.49 or S7.45 into alanine leads to a loss of Gs signaling while maintaining β -arr signaling. These results point to modulation of TM2-TM7 interaction as a common mechanism to elicit β -arr signaling bias in GPCRs.

In summary, a key finding of our study is the existence of a series of rapid network rearrangements at the initial phase of β -arr signaling which can be captured at a timescale of nanoseconds. In addition, we provide a molecular mechanism on how arrestin bias propagates through the δ OR leading to higher TM7 fluctuation at the intracellular end – an event which has been previously related to arrestin-bias.^[10] In hundreds of replicates, we show that an important trigger for β -arr bias is the disruption of direct and indirect interactions between TM2 and TM7 in the connector region, which otherwise stabilize TM7 in the inactive δ OR. This initial event of β -arr bias can be probed on a relatively short timescale of nanoseconds. In this regard, our study indicates that the impact of a signaling stimulus such as ligand binding can be approximated by simulating network rearrangements and structural fluctuations in the initial phases of receptor activation. Ultimately, this finding opens new avenues to study receptor signaling and drug effects by multiple short simulations instead of monitoring the entire activation process, an approach that is still out of reach for conventional simulation methods.

Experimental Section

Experimental Details are found in the Supplemental Material.

[1] M. Martí-Solano, D. Schmidt, P.

Kolb, J. Selent, *Drug Discov. Today* **2016**, *21*, 625–631.

- [2] G. Fenalti, P. M. Giguere, V. Katritch, X.-P. Huang, A. Thompson, V. Cherezov, B. L. Roth, R. C. Stevens, *Nature* **2014**, *506*, 191–196.
- [3] V. Katritch, G. Fenalti, E. E. Abola, B. L. Roth, V. Cherezov, R. C. Stevens, *Trends Biochem. Sci.* **2014**, *39*, 233–244.
- [4] J. Selent, F. Sanz, M. Pastor, G. De Fabritiis, *PLoS Comput. Biol.* **2010**, *6*, 1–6.
- [5] S. Yuan, S. Filipek, K. Palczewski, H. Vogel, *Nat. Commun.* **2014**, *5*, 4733.
- [6] F. Rasmuss, A. J. Venkatakrishnan, <https://gpcrviz.github.io/flareplot/> **2017**.
- [7] X. Sun, G. Laroche, X. Wang, H. Ågren, G. Bowman, P. M. Giguère, Y. Tu, *Chem. - A Eur. J.* **2017**, *23*, 4615–4624.
- [8] X. Deupi, J. Standfuss, *Curr. Opin. Struct. Biol.* **2011**, *21*, 541–551.
- [9] S. Yuan, K. Palczewski, Q. Peng, M. Kolinski, H. Vogel, S. Filipek, *Angew. Chemie - Int. Ed.* **2015**, *54*, 7560–7563.
- [10] J. J. Liu, R. Horst, V. Katritch, R. C. Stevens, K. Wuethrich, *Science* **2012**, *335*, 1106–1110.
- [11] S. Nuber, U. Zabel, K. Lorenz, A. Nuber, G. Milligan, A. B. Tobin, M. J. Lohse, C. Hoffmann, P. Group, T. Unit, et al., *Nature* **2016**, *531*,

661–664.

- [12] L. Valentin-Hansen, T. M. Frimurer, J. Mokrosinski, N. D. Holliday, T. W. Schwartz, *J. Biol. Chem.* **2015**, *290*, 24495–24508.

Acknowledgements

We are grateful to the people who volunteered their computer time to GPUGRID.net for this research.

Keywords: GPCR • opioid receptor • interaction network • signaling bias • molecular dynamics

Supporting Information

Rapid network rearrangements drive the initial phase of β -arrestin signaling in the δ -opioid receptor

Tomasz Maciej Stepniewski, Ismael Rodríguez-Espigares, Maria Martí-Solano, Anna Troya-Bruguer, Mariona Torrens-Fontanals, G. De Fabritiis, Sławomir Filipek and Jana Selent*

Abstract: Signaling bias is an emerging concept for obtaining more efficacious drugs targeting GPCRs. However its structural fundamentals remain poorly understood. Here, we use all-atom molecular dynamics simulations with hundreds of replicas to show that initial structural rearrangements leading to β -arrestin signaling in the δ -opioid receptor occur at a nanosecond timescale. These first steps involve a destabilization of contacts between transmembrane helix (TM) 2 and 7 caused by alterations in intramolecular water networks and in interhelical interactions. This translates into higher TM7 fluctuations near the binding site of intracellular signaling proteins, a behavior that has been associated to β -arrestin biased signaling by biophysical experiments. Ultimately, the existence of rapid network rearrangements during the initial phase of signaling bias creates new opportunities to explore receptor signaling and the impact of ligand binding effects at shorter timescales than currently thought.

Table of Contents

1. Overview of all-atom molecular dynamics simulations
2. Long-time simulations
3. Water diffusion and equilibration
4. Highly conserved residues in the connector region of the δ OR
5. Water occupancies and fluctuation
6. Hydrogen bonding network
7. Y7.53 forms a water mediated bond with N1.50
8. Metastable states during receptor fluctuation detected via MSM analysis
9. Impact of TM2 and TM7 mutation on the signaling profile of naltrindole
10. Experimental Procedures

SUPPORTING INFORMATION

1. Overview of all-atom molecular dynamics simulations

Table S1. Amassed simulation time for studied systems

system	Time [μ s]	replicates
Wt δ OR	1,5	3
D2.50A δ OR	1,5	3
Wt δ OR	0,128	100
D2.50A δ OR	0,128	100
Total time	34,6	

2. Long-time simulations

Despite water influx, we do not observe the formation of a stable continuous water channel connecting the extra- with the intracellular side for none of the receptors (Figure S1B and C). This is not surprising as our simulation were carried out in the presence of an antagonist (Figure S1B) or in the presence of a biased agonist but in the absence of an intracellular signaling protein (Figure S1C). An intracellular agent is needed to facilitate the opening of the intracellular receptor regions and thus the formation of a continuous water channel.^[1] Instead, we can appreciate differences in the structure of the partially formed water channel when comparing the inactive wild-type delta opioid receptor (WT δ OR) with the β -arrestin (β -arr) biased mutant (Figure S1B and C, highlighted by arrows).

SUPPORTING INFORMATION

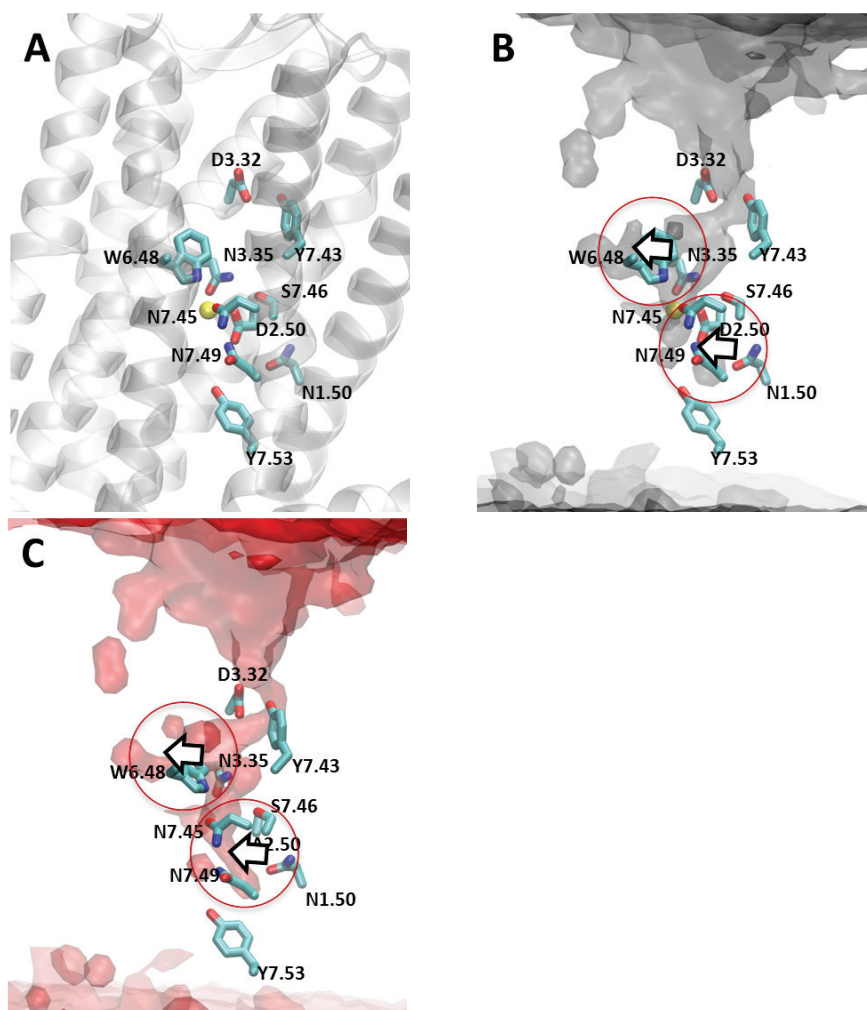


Figure S1. Water occupancy in the inactive WT and β -arr biased D2.50A mutant receptors. The δ OR WT with relevant residues (displayed as licorice) and a sodium ion (yellow sphere) (A). (B) and (C) show water occupancy maps for the WT and D2.50A δ OR, respectively. The maps are computed over 4.5 μ s (3 replicates \times 1.5 μ s) per system. Differences in water occupancy between inactive and biased δ OR are highlighted by arrows. Water occupancy is plotted at a value of 0.2 (i.e. water is present in at least 20% of the simulation).

SUPPORTING INFORMATION

3. Water diffusion and equilibration

In our studies we have utilized the TIP3P water model.^[2] This model has been shown to reproduce well solvent properties.^[3] One of the parameters that characterizes water is the self-diffusion coefficient. This parameter takes into account the mean square displacement (MSD) of molecules that occurs at a certain time-step. Experimental studies have shown that this parameter equals 2.299 at 298.15K and 2.597 at 303.15K.^[4] We have studied the progression of the water self-diffusion coefficient in one of our replicates. The parameter has been computed with Newton's equation:

$$D(\tau) = \text{MSD}(\tau) / 6\tau$$

in which D is the self-diffusion coefficient and τ is the timestep between two distinct states of the system.

In our system, we computed MSD of water oxygens once every 2.5 ns. By studying the progression of this parameter (see below), we can observe that waters converge to a stable rate of diffusion after around 25 ns, with a self-diffusion parameter of around $2.7 \text{ m}^2 \cdot 10^{-9} / \text{s}$ which is slightly higher than the value observed in nature. This is consistent with previous studies that show slight increase in mobility of the TIP3P waters.^[5] The below graph (Figure S2) clearly shows that at the beginning of the replicate there is an equilibration phase, during which the waters in the system converge, towards a dynamic equilibrium. The slightly increased mobility of the waters in the system on one hand shows us that our computational models still has limitations. On the other hand, more mobile waters ensure a faster and more efficient sampling of different positions of this molecules. Thus the short time interval that we use in our simulation scheme is suitable to capture rapid water rearrangements.

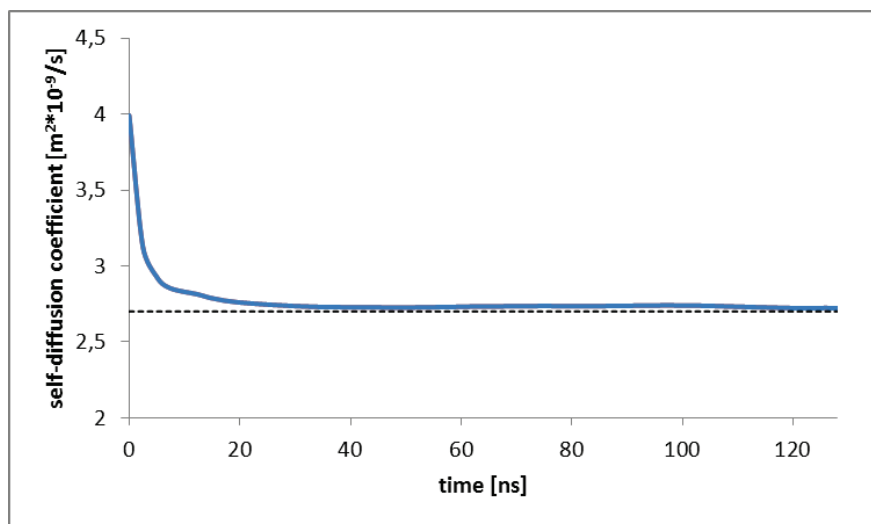


Figure S2. Progression of the water self-diffusion parameter studied over 2.5 ns intervals in a single replicate, of the simulated WTÖOR

SUPPORTING INFORMATION

4. Highly conserved residues in the connector region of the δ OR

To compare the water network observed in our simulations with the one observed in the δ OR crystal structures (PDB code: 4N6H) we have superimposed the WT δ OR simulations coordinates used for the water network analysis with the crystal structure using the protein backbone as reference. We superimposed the water occupancy map generated for the protein (Figure S3A, shown in blue) with water molecules present in the crystal structure (Figure S3A, shown as red spheres). The comparison highlights that our simulations reproduce the water network observed in the crystal structure. In Figure S3B, we show the conservation of residues involved in maintaining this network.

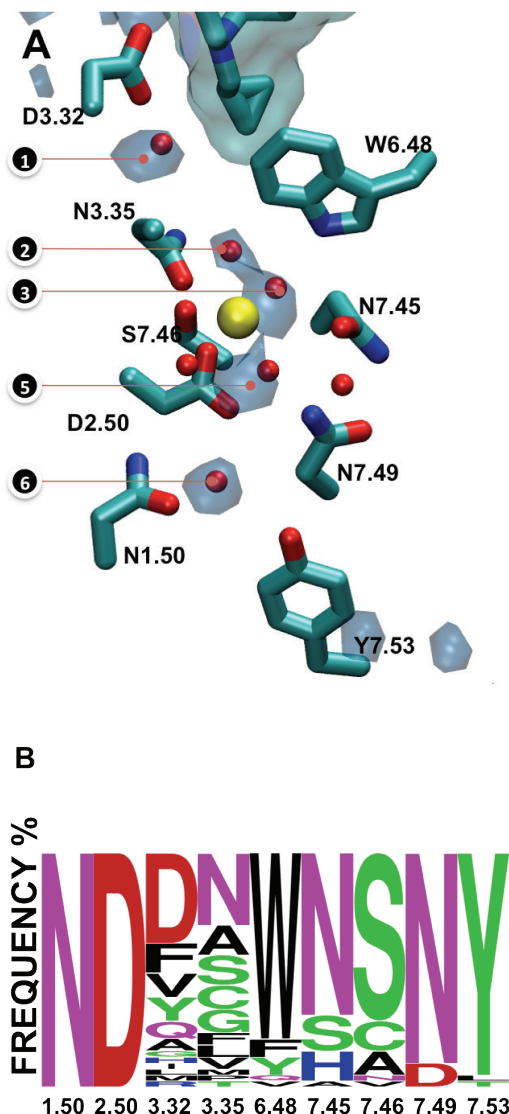


Figure S3. Highly conserved residues in the connector region of the δ OR. Water occupancy map in the connector region of the WT δ OR was generated for coordinates obtained from the short-replicate simulations at an occupancy threshold of 40%. The studied system was aligned to the δ OR crystal structure (PDB code: 4N6H) (A). Waters present in the crystal structure are displayed as red spheres. In (B) we show evolutionary conservation of residues in the connector region computed among class A GPCRs with available crystal structures.

SUPPORTING INFORMATION

5. Water occupancies and fluctuation

Occupancies values for WT (grey) and arrestin-biased receptor are summarized in Figure S4A. It can be seen that occupancies of water numbers 1,2,3,5 and 6 are lower in the arrestin-biased receptor compared to the WT δ OR. Lower occupancy suggests higher fluctuation in the connector region. In fact, plotting the transition of waters through defined slices in 100 ns (i.e. water fluctuation) supports this statement (Figures S4B, right top). In particular in slices 5 to 8, we observe higher water fluctuations whereas slices 1 to 4 and 9 to 10 are not significantly different between arrestin-biased and WT receptor.

Surprisingly, we observe that occupancy of water molecule 6 which mediates the interaction between N1.50-Y7.53 lock is similar comparing the WT and β -arr biased δ OR (Figure 1B and Figure S4A) despite the lock being more open in the β -arr biased δ OR (Figure 2E, point 5). However, measuring the fluctuation rate of waters in the different compartments shows that highest rates are found at the surroundings of the N1.50-Y7.53 lock (Figure S4B, slices 7 and 8). This high fluctuation might contribute to a partial disruption of the N1.50-Y7.53 lock and on a larger scale to the initiation of global TM7 conformational changes.

In addition, we find that slices with pronounced differences in fluctuation (Figure S4B, right top: slices 5 to 9) have a lower average number of waters (Figure S4B, right bottom: slices 5 to 9) whereas slices with no significant differences in fluctuation (Figure S4B, right top: slices 1 to 4 and 9 to 11) have high numbers of waters (Figure S4B, right bottom: slices 1 to 4 and 9 to 11).

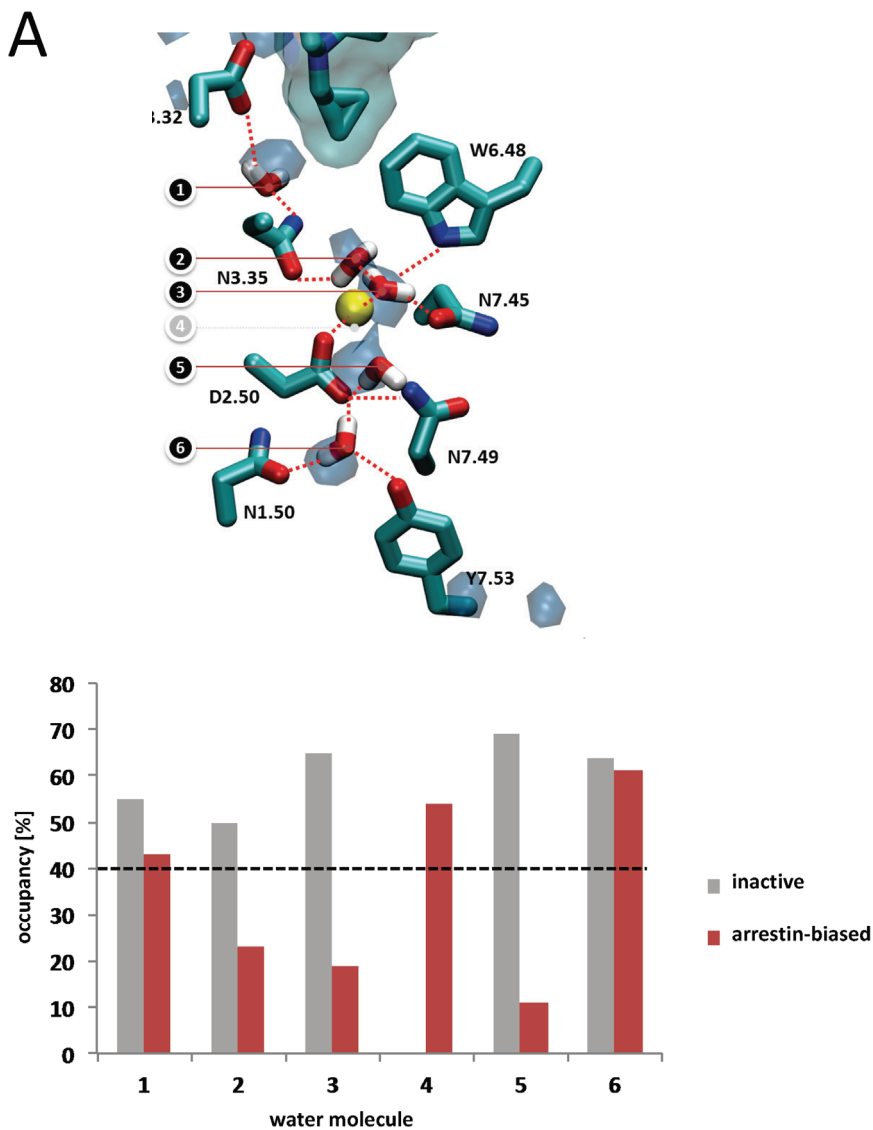


Figure S4. Occupancy of water molecules in the connector region of the δ OR (A). The number of water molecule corresponds to its position in Figure 2 in the main text. In (B) we show the average number of waters and water fluctuation per slice.

SUPPORTING INFORMATION

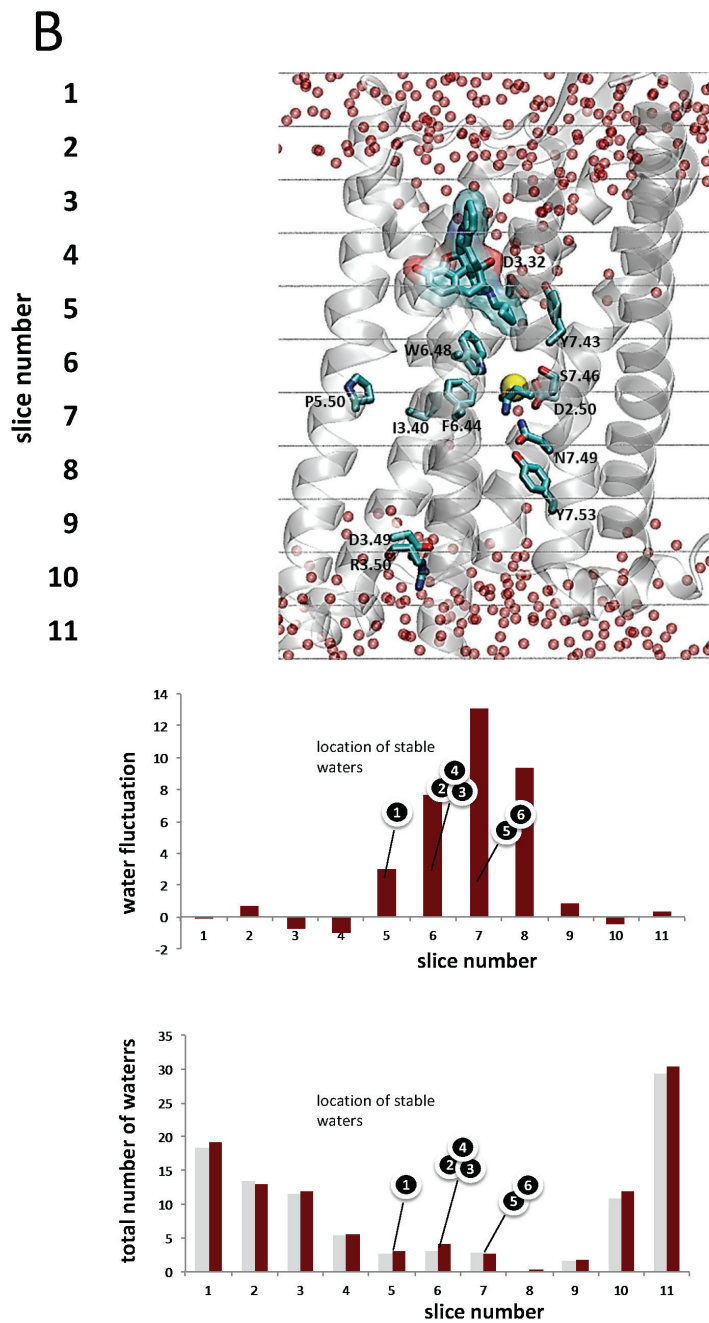


Figure S4. (continued) In (B) we show the average number of waters and water fluctuation per slice.

6. Hydrogen bonding network

Flare plots reveal in general less interhelical interactions in the mutated receptor (Figure S5B) compared to the WT receptor (Figure S5A). We can see that direct polar interactions between D2.50 and residue N7.49 and S7.46 are not maintained in the β -arr biased receptor (Figure 2C). Lack of those interactions likely contributes to increased fluctuations observed in the lower part of TM7. In the upper part, we see that the polar bond between Y7.53 and D3.32 is less stable in the WT receptor. This is in line with our observation that this conserved ionic lock is more often broken in the mutated receptor (Figure 2E, plot 1). We also observe a general decrease of interhelical interactions in the β -arr biased receptor (i.e. interactions between TM2 and TM3, or TM3 and TM4). Such a result highlights the importance of the sodium ion in stabilizing the interaction network of the δ OR.

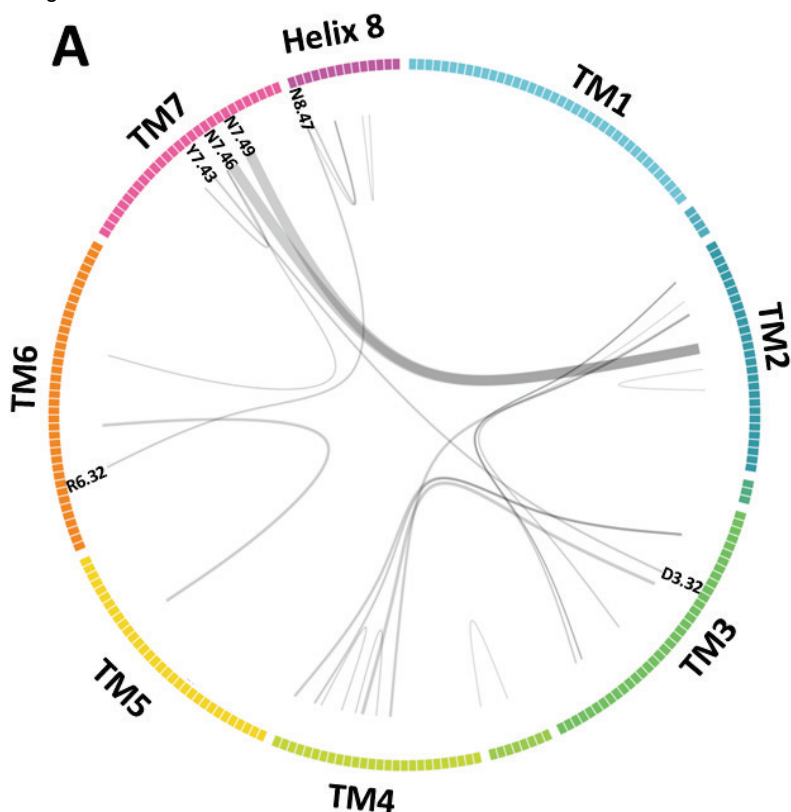


Figure S5. Differences in the intra- and interhelical hydrogen bonding network for inactive WT (A) and β -arr biased (B) δ OR. To obtain the graphs, we have generated interaction frequencies for both the WT and D2.50A δ OR, and then subtracted the resulting frequencies between the systems as follows: WT receptor (Δ frequency = frequency_{inactive} - frequency _{β -arr biased}) (A) and β -arr biased δ OR: (Δ frequency = frequency _{β -arr biased} - frequency_{inactive}). The graphs are plotted with a threshold of Δ frequency > 10%. Line thickness relates to the degree of interaction difference between the WT (A) and β -arr biased δ OR (B)

SUPPORTING INFORMATION

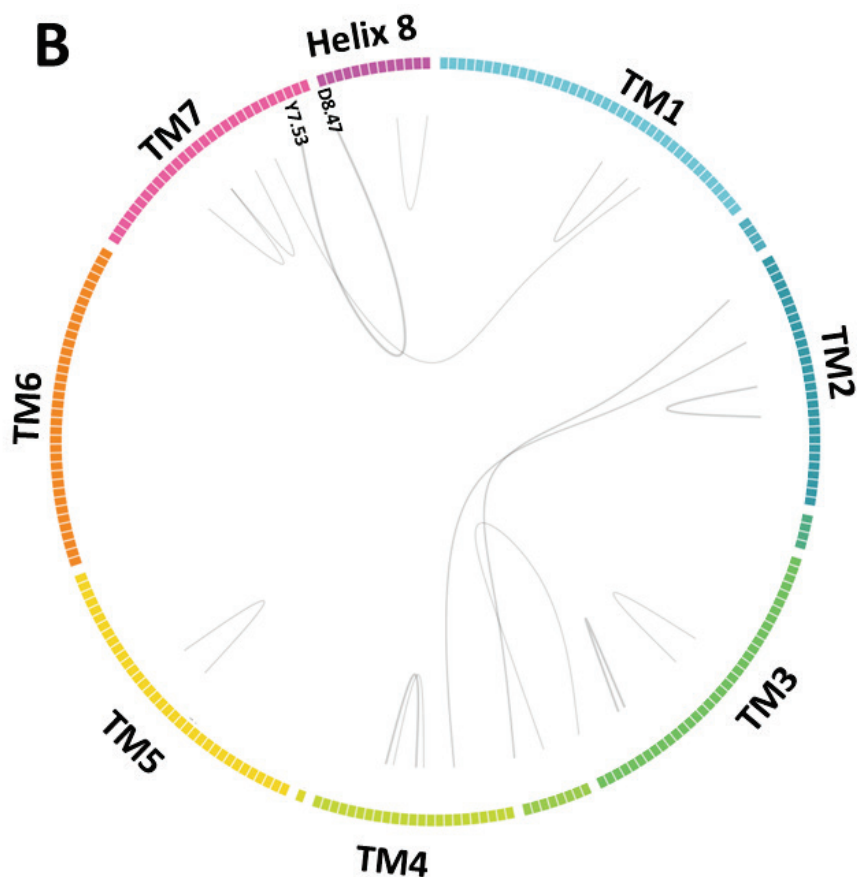


Figure S5. Differences in the intra- and interhelical hydrogen bonding network for inactive WT (A) and β -arr biased (B) δ OR (continued). To obtain the graphs, we have generated interaction frequencies for both the WT and D2.50A δ OR, and then subtracted the resulting frequencies between the systems as follows: WT receptor (Δ frequency = frequency_{inactive} - frequency _{β -arr biased}) (A) and β -arr biased δ OR: (Δ frequency = frequency _{β -arr biased} - frequency_{inactive}). The graphs are plotted with a threshold of Δ frequency > 10%. Line thickness relates to the degree of interaction difference between the WT (A) and β -arr biased δ OR (B)

Curiously, we also observe differences in stability of intrahelical interactions between the studied systems. Intrahelical interactions are primarily interactions between backbone oxygens and nitrogens that stabilize the helical conformation. Thus, the observed changes may correspond to conformation differences between the studied receptors. We can observe pronounced differences in TM7 in respect of those interactions. This changes can be related to difference in fluctuations (Figure 1C) and different population of macrostates for this region (Figure S7) observed in the D2.50A receptor.

SUPPORTING INFORMATION

7. Y7.53 forms a water mediated bond with N1.50

In the inactive δ OR^[6,7] as well as in the closely related inactive μ OR structure,^[8] Y7.53 forms a water mediated bond with the highly conserved N1.50. An open state of this lock has been related to receptor activation in the μ OR.^[9] (Figure S5)

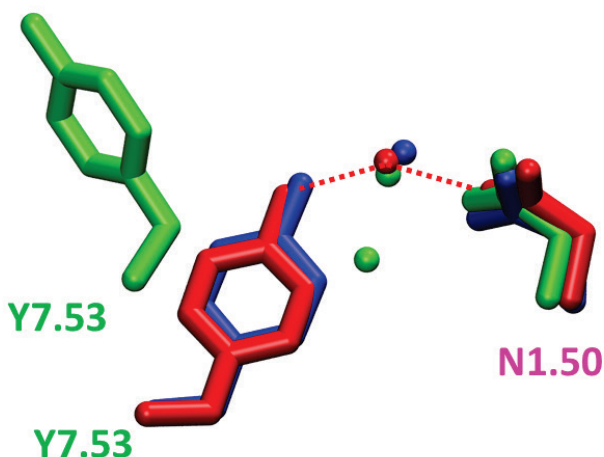


Figure S6. Conservation of the lock between Y7.53 and N1.50 in opioid receptors mediated by a water molecule. Superimposition of the crystal structures of the inactive δ -opioid receptor (PDB code: 4N6H, colored blue), inactive μ -opioid receptor (PDB code: 4DKL, colored red), and active μ -opioid receptor (PDB code: 5C1M – colored green). Crystallized waters within 5Å distance of N1.50 are displayed as spheres.

8. Metastable states during receptor fluctuation detected via MSM analysis

Previous analysis showed that arrestin-bias is related to higher fluctuation in particular TM7 (Figure 2, main text). In order to explore potential metastable states and their transitions during receptor fluctuation, we performed Markov State Models (MSMs) analysis on ϕ and Ψ dihedral angles of δ OR residues Y7.43 - L7.56 (Figure S6A, see also experimental section). MSM analysis and PCCA+ yield 7 macrostates for WT δ OR and 11 macrostates for the arrestin-biased δ OR (D2.50A mutant). Most relevant macrostates in terms of frequency (state 1 to 5) and net fluxes suggested by Transition Pathway Theory (TPT) analysis (see tables S3, S4, S5) from state 1 to 3 and 5 are shown in Figure S7B. We can observe that the most stable state 1 is more populated in the WT receptor than in the arrestin-biased receptor (WT/ β -arr-Bias: 59%/34%). This tendency is also observed in the second most populated state 4 (WT/ β -arr-Bias:

SUPPORTING INFORMATION

20%/16%). In contrast, state 2 (WT/ β -arr-Bias: 3%/11%) becomes more stable in the arrestin-biased receptor. Importantly, states 3 and 5 appear exclusively in arrestin-biased δ OR. These are characterized by helix movement and rotation of residues F7.44 – F7.55 with a pronounced conformational change in Y7.53. This residue is part of the NPXXY motif. Interestingly, the observed anti-clockwise rotation in the arrestin-biased δ OR (D2.50A mutant) is also observed in rhodopsin when coupling to arrestin (Figure S7C). Note that state 4 and 5 together count for 9 % of simulation data in the arrestin-biased δ OR while these states have not been found in the WT δ OR. Considering the total analyzed simulation time of 12.8 μ s obtained from 100 replicates, state 4 and 5 occur for a significant time of 1.2 μ s for the arrestin-biased δ OR (D2.50A mutant).

SUPPORTING INFORMATION

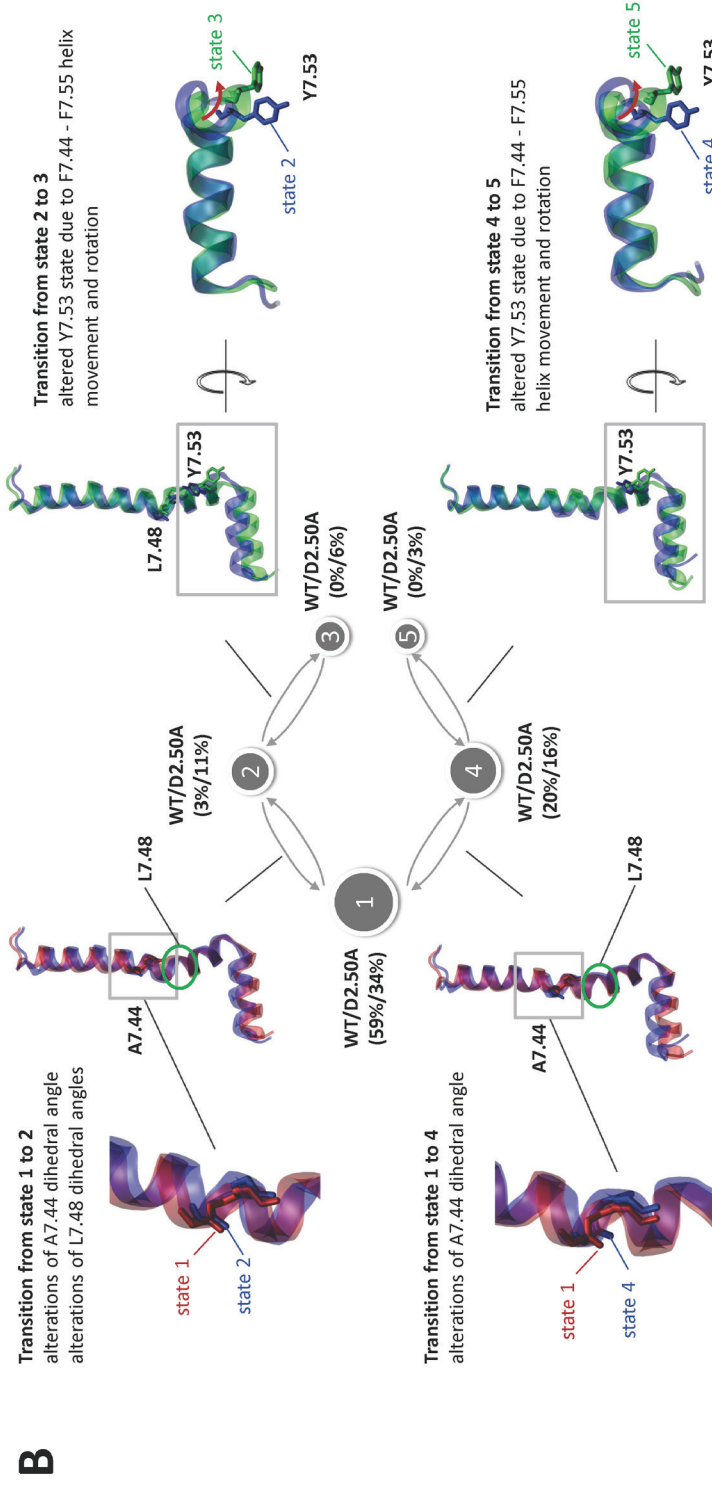


Figure S7. Metastable states of TM7 detected during receptor fluctuation via MSM analysis. δ OR residues Y7.43 - L7.56 used for MSM analysis are labelled (A). Most relevant macrostates based on frequency (state 1 to 5) and pathways net fluxes suggested by TPT analysis (B). Superimposition of the crystal structures of inactive rhodopsin (PDB 1F88) and arrestin coupled rhodopsin (PDB 4ZWW) (C).

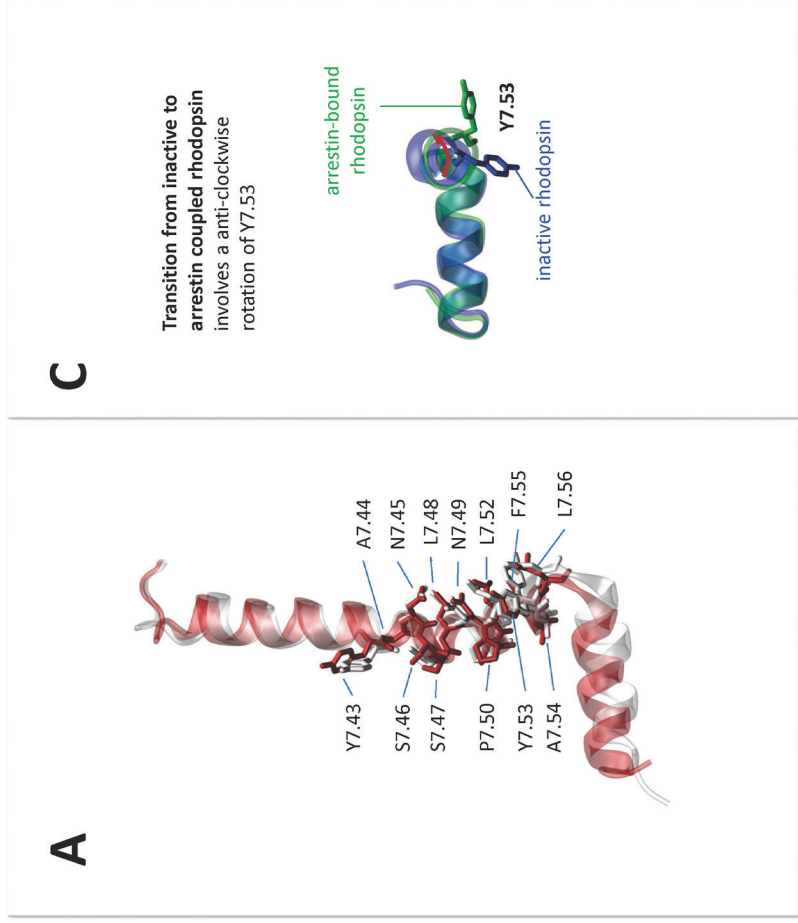


Figure S7. Metastable states of TM7 detected during receptor fluctuation via MSM analysis (continued).

9. Impact of TM2 and TM7 mutation on the signaling profile of naltrindole

Table S2. Impact of TM2 and TM7 mutation on the signaling profile of naltrindole. Reproduced from [6]

	β -arrestin activation Emax (%) of BW373U86) \pm SEM	β -arrestin activation EC50 (nM)	G α_i activation Emax (% of BW373U86) \pm SEM	G α_i activation EC50 (nM)
WT	N/A ^a	N/A ^a	53.4 \pm 2,8	21.9
D2.50A	64.3 \pm 1	4.2	N/A ^a	N/A ^a
N7.45A	45 \pm 1	50	N/A ^a	N/A ^a
N7.49A	25.5 \pm 1,4	36.7	N/A ^a	N/A ^a

[a] "N/A" indicated that no activation was observed or the ligand concentration dependence curve lacked a sigmoidal characteristic

10. Experimental Procedures

System preparation: The systems were generated using CHARMM-GUI.^[10,11] We have used the crystal structure of δ -opioid receptor in complex with naltrindole [PDB code: 4N6H]. To generate the mutated D2.50A δ -opioid receptor, we have inserted the mutation using the CHARMM-GUI panel. A sodium ion was placed in the allosteric binding site of the wild-type receptor. It was not placed in the D2.50A receptor, as mutational data show, that this mutation makes the receptor sodium insensitive.^[6] The receptor was embedded in a $\sim 80 \times 80 \text{ \AA}$ POPC bilayer and solvated with TIP3 water molecules. The ionic strength of the solution was kept at 0.15 M with NaCl ions. Parameters for the simulation were obtained from the CHARMM36 forcefield.^[12] Parameters for the ligand were assigned from the CGenFF forcefield automatically by the ParamChem tool implemented in CHARMM-GUI.^[13,14]

Molecular dynamic simulations: The systems were first equilibrated in conditions of constant pressure (NPT, 1.01325 bar) for 20 ns, preceded by an initial 1000 step minimization. After the NPT step, we have carried out simulations in conditions of constant volume (NVT) in 100 replicates of 128 ns and in 3 replicates of 1500 ns for each of the systems. The simulations were run in ACEMD.^[15] In both steps we used a time-step of 4 fs. Such a time-step was possible due to the hydrogen mass repartitioning scheme being employed in Acemd.^[16] A non-bonded interaction cutoff was set at 9 \AA . A smooth switching function for the cut-off was applied, starting at 7.5 \AA . The size of the cell was set to prevent non-bonding interactions between the protein and its periodic boundary image. Long-distance electrostatic forces were calculated

SUPPORTING INFORMATION

using the Particle Mesh Ewald algorithm. The algorithm had grid spacing of 1 Å. The bond lengths of hydrogen atoms were kept constrained using the RATTLE algorithm. Simulations were carried out at a temperature of 300K.

Water fluctuation analysis: To analyze fluctuations, the coordinates obtained in multiple short simulations runs were aligned to the structure available in the OPM database.^[17] For analysis only, the area corresponding to the internal water channel of the protein ($x > -4$ Å and $x < 10$ Å and $y > -12$ Å and $y < 12$ Å) was taken into account. The area was divided into 5 Å slices based on the value of the Z coordinate, with the eleventh slice starting at $Z = -27,5$ Å and ending at $Z = -22,5$ Å. To quantify water fluctuations, the number of waters that either enter or exit the studied area between two subsequent frames was measured (for this analysis, we have strided the simulations, so that there is a 0.4 ns time skip between frames). To simplify quantification, we only considered the oxygen atom for describing the position of a water molecule. We divided the obtained fluctuations by the average amount of water present in the slice in the two frames used to compute fluctuations. By averaging the values obtained from all of the coordinates, we get the number of fluctuations in a slice, per water molecule occurring over 0.4 ns of the simulation. The values were then multiplied, to obtain the average number of fluctuations per slice that occur in 100 ns.

Flare plot analysis: To generate flare plots (Figure 2AB, and Figure S5) we have used the tool developed by Dr. Fonseca and Dr Venkatakishnan (<https://gpcrviz.github.io/flareplot/>). Flare plots are based on the computation of hydrogen bonds by means of the MDTraj Python library.^[18] Hydrogen bonds are identified as any combination of donor atoms (NH or OH) and acceptor atoms (N or O) which have a donor-acceptor distance < 2.5 Å and the angle formed between the acceptor atom, the hydrogen atom and the atom covalently bound to the polar hydrogen has a value higher than 120° .

Markov State Models: Sines and cosines of ϕ and Ψ dihedral angles of arrestin-biased δ OR and WT δ OR residues Y7.43 - L7.56 for the whole 100 replicates of 128 ns per system were used as raw input data for dimensionality reduction. Raw data from both arrestin-biased δ OR and WT δ OR was processed together by Time-lagged Independent Component Analysis (TICA)^[19-21] with a lag-time 0.2 ns yielding the first three time-lagged Independent Coordinates (tIC) for both systems that summarized the slowest processes of the system accounting for a total kinetic variability of 52%. The three tICs of each system were clustered independently to 1000 approx. discretized states by mini-batch K-means clustering.^[22] Transition matrixes were estimated by maximum likelihood method considering a lag-time of 5 ns. Then, several macrostates were obtained by Robust Perron Cluster Analysis (PCCA+) for both systems.^[23] Furthermore, Transition Pathway Theory (TPT)^[24-26] analysis was performed to obtain a coarse-grained model for the transition between macrostates. Software used was HTMD 1.9.4^[27] for MD data processing, MSM estimation and PCCA+ and PyEMMA 2.4^[28] for TPT analysis.

SUPPORTING INFORMATION

Markov state model estimation and TPT analysis:

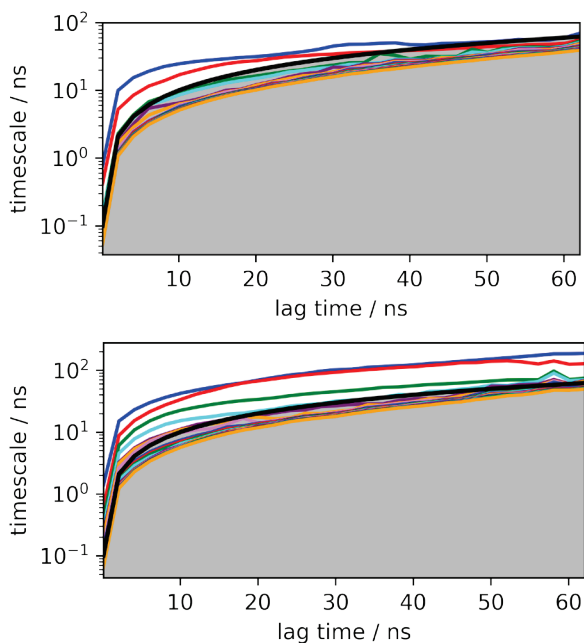


Figure S8. 20 first implied timescales computed for every 2 ns as a function of the lag time. Grey area and the limiting black line is defined by lag time > timescale and represent the area where the processes under investigation have already decayed and the estimation is not reliable. Left: Implied timescales of the markov model estimated from WT δ OR simulations. Right: Implied timescales of the markov model estimated from WT δ OR simulations arrestin-biased δ OR. Nearly constant timescales show process markovianity.

Table S3. Most relevant net fluxes of direct transitions between states in the pathways from state 1 to 3 and 5 obtained by TPT analysis for arrestin-biased δ OR.

State transitions	Net flux (10^{-5} transitions/ lag time*)	Transition time (1/Net flux, μ s)
1→3 pathways		
1→2	5.35	93.30
2→3	7.63	65.50
1→4	2.51	198.88
4→2	1.51	331.91
4→3	1.03	484.93
1→5 pathways		
1→4	4.08	103.46

SUPPORTING INFORMATION

4→5	4.83	103.46
1→5	1.12	448.20
1→2	0.93	535.00
2→4	0.69	726.28

*Lag time = 5 ns

Table S4. Most relevant net fluxes of direct transitions between states in the pathways from state 1 to 2 and 4 obtained by TPT analysis for WT δ OR.

State transitions	Net flux (10^{-5} transitions/ lag time*)	Transition time (1/Net flux, μ s)
1→2 pathways		
1→2	12.18	41.06
1→4	6.83	73.26
4→2	6.84	73.10
1→4 pathways		
1→4	77.37	6.46
1→2	4.96	100.76
2→4	4.95	101.05

*Lag time = 5 ns

Table S5. Total transition rates of the different pathways analyzed by TPT analysis starting from state 1.

Pathways	Total transition rate (k_{AB}) (10^{-5} transitions/ lag time*)	MFPT† (μ s)
1→3 arrestin-biased δ OR	9.52	52.54
1→5 arrestin-biased δ OR	6.93	72.18
1→3+5 arrestin-biased δ OR	16.00	31.26
1→2 WT δ OR	19.59	25.52
1→4 WT δ OR	89.75	5.57

*Lag time = 5 ns

†MFPT = Mean First Passage Time. MFPT = $1/k_{AB}$

References

- [1] S. Yuan, S. Filipek, K. Palczewski, H. Vogel, *Nature communications* **2014**, *5*, 4733.
- [2] W. L. Jorgensen, *Journal of the American Chemical Society* **1981**, *103*, 335–340.
- [3] D. J. Huggins, *Journal of Chemical Physics* **2012**, *136*, 1–30.
- [4] M. Holz, S. R. Heil, A. Sacco, *Physical Chemistry Chemical Physics* **2000**, *2*, 4740–4742.
- [5] P. Florová, P. Sklenovský, P. Banáš, M. Otyepka, *Journal of Chemical Theory and Computation* **2010**, *6*, 3569–3579.
- [6] G. Fenalti, P. M. Giguere, V. Katritch, X.-P. Huang, A. a Thompson, V. Cherezov, B. L. Roth, R. C. Stevens, *Nature* **2014**, *506*, 191–196.
- [7] S. Granier, A. Manglik, A. C. Kruse, T. S. Kobilka, F. S. Thian, W. I. Weis, B. K. Kobilka, *Nature* **2012**, *485*, 400–4.
- [8] A. Manglik, A. C. Kruse, T. S. Kobilka, F. S. Thian, J. M. Mathiesen, R. K. Sunahara, L. Pardo, W. I. Weis, B. K. Kobilka, S. Granier, *Nature* **2012**, *485*, 321–6.
- [9] W. Huang, A. Manglik, a J. Venkatakrishnan, T. Laeremans, E. N. Feinberg, A. L. Sanborn, P. Gmeiner, H. E. Kato, K. E. Livingston, T. S. Thorsen, et al., *Nature* **2015**, *524*, 315–21.
- [10] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, et al., *Journal of Chemical Theory and Computation* **2016**, *12*, 405–413.
- [11] E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. M. Venable, et al., *Journal of Computational Chemistry* **2014**, *35*, 1997–2004.
- [12] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, R. W. Pastor, *Journal of Physical Chemistry B* **2010**, *114*, 7830–7843.
- [13] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, et al., *Journal of Computational Chemistry* **2010**, *31*, 671–690.
- [14] W. Yu, X. He, K. Vanommeslaeghe, A. D. MacKerell, *Journal of Computational Chemistry* **2012**, *33*, 2451–2468.
- [15] M. J. Harvey, G. Giupponi, G. De Fabritiis, *Journal of Chemical Theory and Computation* **2009**, *5*, 1632–1639.
- [16] K. A. Feenstra, B. Hess, H. J. C. Berendsen, *Journal of Computational Chemistry* **1999**, *20*, 786–798.
- [17] M. A. Lomize, A. L. Lomize, I. D. Pogozeva, H. I. Mosberg, *Bioinformatics* **2006**, *22*, 623–625.
- [18] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L. P. Wang, T. J. Lane, V. S. Pande, *Biophysical Journal* **2015**, *109*, 1528–1532.
- [19] L. Molgedey, H. G. Schuster, *Physical Review Letters* **1994**, *72*, 3634–3637.
- [20] C. R. Schwantes, V. S. Pande, *Journal of Chemical Theory and Computation* **2013**, *9*, 2000–2009.
- [21] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, *The Journal of chemical physics* **2013**, *139*, 15102.
- [22] D. Sculley, D., in *Proceedings of the 19th International Conference on World Wide Web - WWW '10*, ACM Press, New York, New York, USA, **2010**, p. 1177.
- [23] S. Röblitz, M. Weber, *Advances in Data Analysis and Classification* **2013**, *7*, 147–179.
- [24] W. E., E. Vanden-Eijnden, *Journal of Statistical Physics* **2006**, *123*, 503–523.
- [25] P. Metzner, C. Schütte, E. Vanden-Eijnden, *Multiscale Modeling & Simulation* **2009**, *7*, 1192–1219.
- [26] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, T. R. Weigl, *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 19011–6.

SUPPORTING INFORMATION

- [27] S. Doerr, M. J. Harvey, F. Noé, G. De Fabritiis, *Journal of chemical theory and computation* **2016**, *12*, 1845–52.
- [28] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, F. Noé, *Journal of chemical theory and computation* **2015**, *11*, 5525–42.

Author Contributions

TMS and IRE equally contributed to microsecond scale simulations, data curation, data analysis, data interpretation and manuscript writing. MMS, MTF and ATB supported data analysis. GDF launched production runs on GPUgrid (100 replica of 128 ns for WT and biased δ OR). SF helped in designing the study and data interpretation. JS designed the study, analysed/interpreted data and wrote the manuscript.

Chapter 5

Membrane cholesterol effect on the 5-HT_{2A} receptor: Insights into the lipid-induced modulation of an antipsychotic drug target

Ramírez-Anguita, J.M., Rodríguez-Espigares, Ismael, Guixà-González, R., Bruno, A., Torrens-Fontanals, M., Varela-Rial, Alejandro & Selent, J. Membrane cholesterol effect on the 5-HT_{2A} receptor: Insights into the lipid-induced modulation of an antipsychotic drug target. *Biotechnology and Applied Biochemistry* (2017). doi:10.1002/bab.1608 (accepted for publication)

Summary: In this article, we aim to decipher how cell membrane composition affects GPCR functionality focusing on the serotonin 5-hydroxytryptamine 2A (5-HT_{2A}) receptor. 5-HT_{2A} is a GPCR relevant for the treatment of Central Nervous System (CNS) disorders⁷¹ and, as happens in other class A GPCRs, its functioning is heavily modulated by membrane composition in the brain.⁵⁰

Since cholesterol is an essential component of neuronal membranes,⁷² we have studied its effect on 5-HT_{2A} receptors structural fluctuation through all-atom Molecular Dynamics (MD) simulations. Here, I focused in the relationship between hydrogen bond formation among different receptor protein segments and the increased receptor conformational variability.

We find that the presence of cholesterol in the membrane increases receptor conformational variability in most receptor segments. Together with other structural analysis explained in this article, our results indicate that conformational variability seems to go along with the destabilization of hydrogen bonding networks not only within the receptor but also between receptor and lipids. Our findings contribute

to a better understanding of membrane-induced alterations of receptor dynamics, which might affect receptor functionality, and points to cholesterol-induced stabilizing and destabilizing effects on the conformational variability of GPCRs.

Ramírez-Anguita JM, Rodríguez-Espigares I, Guixà-González R, Bruno A, Torrens-Fontanals M, Varela-Rial A, et al. [Membrane cholesterol effect on the 5-HT 2A receptor: Insights into the lipid-induced modulation of an antipsychotic drug target](#). *Biotechnol Appl Biochem*. 2018 Jan;65(1):29–37. DOI: 10.1002/bab.1608

Chapter 6

Discussion

The present thesis sets the framework to provide the GPCR community with a web platform for molecular dynamics data related to GPCR functionality with the aim to accelerate GPCR research and drug discovery. Our website offers visualization and analysis tools for analyzing GPCRmd database trajectories. In two case studies (Chapters 3 and 4), these tools have been proved to be useful for analyzing the molecular dynamics (MD) data related to cholesterol-induced membrane effects and functional selectivity.

6.1. The GPCRmd database

GPCRmd is not only a repository, that helps preserve data integrity and availability, it also provides the tools for further visualization and analysis. This interface is user-friendly, and thus gives access to MD data to the whole GPCR community, regardless of specialization.

However, different simulation force-fields and software make it difficult to adapt structural data to a single file format. The alternative is to complement standard formats with complementary files which usually are force-filed and/or software specific. This is one of the main challenges that our platform should face for its maintenance as new force-fields and simulation software are developed. Also, different naming of non-protein residue modifications and atom types complicate the verification of the protein topology. This needs to be addressed in future implementation efforts.

Furthermore, increasing production and deposition of MD data expected in the future will require increasing data storage space, internet bandwidth and computational power. Finally, ensuring

quality of the submitted MD data through curation would require an immense effort. Thus, quality analysis pipelines should be developed although these could not be sufficient. For this reason, the submitter will be required to publish his data in peer-reviewed journal prior to the submission to GPCRmd. Although we must face these technical problems, further development, continuous code maintenance and hardware upgrades would be able to solve these problems.

As an example of application, we presented two case studies: 1) a comparison of interaction events in 5-hydroxytryptamine receptor 1B (5-HT_{1B}) and 5-hydroxytryptamine receptor 2B (5-HT_{2B}) both in complex with ergotamine, and 2) interaction events associated with δ -type opioid receptor (δ -OR) in complex with naltrindole in comparison with the same complex with D2.50A δ -OR.

Case study 1: Ergotamine is an anti-migraine drug which activates G protein coupling and β -arrestin signaling when binding to 5-HT_{1B} receptor. Differently, at the 5-HT_{2B} receptor ergotamine favors β -arrestin over G protein coupling.^{73,74} Furthermore, previous MD studies on these receptors show that ergotamine adopts a similar position in the orthosteric binding pocket of both receptors, but establishes increased contacts with an extended binding pocket in the extracellular part of transmembrane helix 5 (TM5) of the 5-HT_{2B} receptor, which are not present in the 5-HT_{1B} receptor.

Ligand-residue interaction analysis of molecular dynamics trajectories retrieves mostly the same interactions as the ones observed in the originally crystallized structure (interaction tool in GPCRdb).⁴ However, GPCRmd adds an important level of information about the strength of those interactions in terms of frequency. This information is very valuable when designing novel strategies to target these GPCRs. In addition to this structural information, GPCRmd revealed interesting information about differences in the interhelical interaction network which may be responsible for the different signaling outcome.

Case study 2: δ -OR findings on intramolecular hydrogen bond analysis are thoroughly discussed in Chapter 4 and Section 6.2.

6.2. β -arrestin signaling

In Chapter 4, we provide a model for a molecular mechanism on how arrestin bias propagates through the δ -opioid receptor (δ -OR) producing higher TM7 fluctuation at the intracellular end—an event which has been previously related to arrestin-bias.⁷⁵ We propose that this model is also valid for other class A GPCRs which conserve residues N7.45, S7.46 and N7.49 (Ballesteros-Weinstein notation).⁷⁶

For this study, we compared the dynamics of the wild-type δ -OR in complex with naltrindole (inactive receptor) with a δ -OR mutant which carries a single point mutation in position 2.50 (arrestin biased receptor). According to our data, β -arrestin signaling begins with the destabilization of contacts between TM2 and 7. This seems to be a consequence of changes in the intramolecular water networks around the D2.50 sodium allosteric-site^{77,78} and interhelical interactions. Importantly, Markov State Model (MSM) analysis of TM7 residues identified two new main conformational states which are specific to the arrestin biased δ -OR. They are characterized by a rotation of Y7.53 similar to the one found in inactive rhodopsin⁶⁹ and arrestin coupled rhodopsin crystal structures.⁷⁰ Thus, MSM combined with Transition Path Theory analysis proved useful for describing mechanisms of residue allosteric networks.

All in all, our study suggests that the initial phase of signaling bias can be captured at a scale of nanoseconds. We find that this phase is specific for a distinct signal outcome (i.e. inactive or arrestin signaling) which has important implication for future in-silico drug screening efforts. Instead of simulating the entire and slow process of receptor activation, drug action can be predicted by focusing on the initial and rapid signaling phase.

6.3. Membrane effects

Previous experiments done by Zocher *et al* have found, using single-molecule force spectroscopy, that the presence of cholesterol increases the kinetic, energetic, and mechanical stability of almost every structural segment in the β_2 AR during the process of unfolding.⁷⁹ (see Chapter 3 Experimental section for definition of the receptor segments). Interestingly, despite increased mechanical

receptor stability, the authors also found higher conformational variability for all receptor segments in the presence of cholesterol, with significantly pronounced values for transmembrane helix 1 (TM1), TM2 and the extracellular loop 2 (ECL2) at a time-scale of seconds.⁸⁰

In Chapter 3, we ran MD simulations to study how changes in membrane composition —specifically the presence (25% of membrane lipids) to absence of cholesterol— affects GPCR dynamics and functionally, in this case serotonin 5-hydroxytryptamine 2A (5-HT_{2A}) receptor dynamics. Interestingly, Root Mean Square Fluctuation (RMSF) analysis of receptor α -carbons shows similar changes in fluctuation to the ones reported by Zocher *et al* although results might not be comparable due to the very different methodology and timescales.

Nevertheless, our results suggest that destabilization of receptor intramolecular hydrogen bonds and the reduction of hydrogen bond interactions with membrane lipids seem to decrease receptor stability and increase receptor conformational variability. This increase in receptor fluctuation is present in many residues of the extracellular site. Note that, alterations in particular at extracellular receptor regions might impact the pathway of ligand entrance as well as its final binding pose in the orthosteric binding site.

Chapter 7

Conclusions

1. GPCRmd, an online platform for molecular dynamics GPCR data submission, visualization and analysis, is presented. This repository makes submitted data available for the whole GPCR community and provide tools for online visualization and analysis. Furthermore, its tools proved to be useful for studying GPCR functionally.
2. Molecular dynamics can be used to provide possible explanations to phenomena found in previous experimental findings which helps understanding better GPCR functionally.
3. Addition of cholesterol to GPCR membranes might produce a decrease on intramolecular receptor hydrogen bonds and lipid-receptor hydrogen bond interactions leading to receptor destabilization and an increase in conformation variability. These leads to changes in receptor dynamics affecting GPCR functionality.
4. A molecular mechanism on how β -arrestin signaling initiates in some of class A GPCRs is proposed. First, destabilization of contacts between transmembrane helix (TM) 2 and TM7 is produced due to alterations in intramolecular water networks and interhelical interactions. Then, arrestin-bias propagates through TM7 producing higher fluctuations at the helix intracellular end, which includes a helix rotation that alters Y7.53 orientation.

Chapter 8

List of communications

Articles

1. Ramírez-Anguita, J. M., Rodríguez-Espigares, I., *et al.* Membrane cholesterol effect on the 5-HT 2A receptor: Insights into the lipid-induced modulation of an antipsychotic drug target. *Biotechnology and Applied Biochemistry* (2017). doi:10.1002/bab.1608
2. Guixa-Gonzalez, R., Rodríguez-Espigares, I., MEMBPLUGIN: studying membrane complexity in VMD. *Bioinformatics* **30**, 1478–1480 (2014).

Reviews

1. Rodríguez-Espigares, I., Kaczor, A. A. & Selent, J. In silico Exploration of the Conformational Universe of GPCRs. *Molecular Informatics* **35**, 227–237 (2016).

Posters

1. A first prototype of GPCRmd. Spring GLISTEN conference. March 2017. Universidade do Porto. Porto. Portugal.

2. GPCRmd database: dynamic insights into static GPCR complexes. Autumn GLISTEN conference. Vysoká škola chemicko-technologická v Praze. September 2016. Prague. Czech Republic.
3. The GPCRmd Database. 3rd Workshop on High-Throughput Molecular Dynamics (HTMD). November 2016. Acellera. Barcelona. Spain.
4. Capturing the action of current antipsychotic drugs at G protein-coupled receptors (GPCRs) by means of Markov State Model analysis. 3rd International Work-Conference on Bioinformatics and Biomedical Engineering. April 2015. Universidad de Granada. Granada. Spain.

Chapter 9

Appendix: other publications

This section lists a publication in which I have contributed in a minor part.

9.1. Membrane cholesterol access into a G-protein-coupled receptor

Guixà-González, R., Albasanz, J.L., Rodriguez-Espigares, I., Manuel, P., Sanz, F., Martí-Solano, M., Manna, M., Martinez-Seara, H., Hildebrand, P.W., Martín, M. & Selent, J. Membrane cholesterol access into a G-protein-coupled receptor. *Nature Communications* **8**, 14505 (2017).

Abstract: Cholesterol is a key component of cell membranes with a proven modulatory role on the function and ligand-binding properties of G-protein-coupled receptors (GPCRs). Crystal structures of prototypical GPCRs such as the adenosine A2A receptor (A2AR) have confirmed that cholesterol finds stable binding sites at the receptor surface suggesting an allosteric role of this lipid. Here we combine experimental and computational approaches to show that cholesterol can spontaneously enter the A2AR-binding pocket from the membrane milieu using the same portal gate previously suggested for opsin ligands. We confirm the presence of cholesterol inside the receptor by chemical modification of the A2AR interior in a biotinylation assay. Overall, we show that cholesterol's impact on A2AR-binding affinity goes beyond pure allosteric modulation and unveils a new interaction mode between cholesterol and the A2AR that could potentially apply to other GPCRs.

Bibliography

1. Katritch, V., Cherezov, V. & Stevens, R. C. Structure-Function of the G Protein-Coupled Receptor Superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53**, 531–556 (2013).
2. Venter, J. C. *et al.* The sequence of the human genome. *Science (80-)*. **291**, 1304–51 (2001).
3. Ballesteros, J. A. & Weinstein, H. in *Methods in Neurosciences* (ed. Sealfon, S. C.) **25**, 366–428 (Academic Press, 1995).
4. Munk, C. *et al.* GPCRdb: the G protein-coupled receptor database - an introduction. *Br. J. Pharmacol.* **173**, 2195–2207 (2016).
5. Isberg, V. *et al.* Generic GPCR residue numbers – aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.* **36**, 22–31 (2015).
6. Wang, C. *et al.* Structural basis for molecular recognition at serotonin receptors. *Science* **340**, 610–4 (2013).
7. Kolakowski, L. F. GCRDb: a G-protein-coupled receptor database. *Receptors Channels* **2**, 1–7 (1994).
8. Fredriksson, R., Lagerström, M. C., Lundin, L.-G. & Schiöth, H. B. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
9. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discovery* **16**, 19–34 (2017).
10. Bentham Science Publisher, B. S. P. Latest Development in Drug Discovery on G Protein-coupled Receptors. *Curr. Protein Pept. Sci.* **7**, 465–470 (2006).
11. Rask-Andersen, M., Masuram, S. & Schiöth, H. B. The Druggable Genome: Evaluation of Drug Targets in Clinical Trials Suggests Major Shifts in Molecular Class and Indication. *Annu. Rev. Pharmacol. Toxicol.* **54**, 9–26 (2014).
12. The druggable genome: an update. *Drug Discovery Today* **10**, 1607–1610 (2005).
13. Rodríguez-Espigares, I., Kaczor, A. A. & Selent, J. In silico Exploration of the Conformational Universe of GPCRs. *Mol. Inform.* **35**, 227–237 (2016).
14. Rodríguez-Espigares, I., Kaczor, A. A., Stepniowski, T. M. &

- Selent, J. in *Computational Methods for GPCR Drug Discovery* (ed. Heifetz, A.) **1705**, (Humana Press, 2018).
15. Martí-Solano, M., Guixà-González, R., Sanz, F., Pastor, M. & Selent, J. Novel insights into biased agonism at G protein-coupled receptors and their potential for drug design. *Curr. Pharm. Des.* **19**, 5156–66 (2013).
 16. Nygaard, R. *et al.* The Dynamic Process of β 2-Adrenergic Receptor Activation. *Cell* **152**, 532–542 (2013).
 17. Dror, R. O. *et al.* Activation mechanism of the β 2-adrenergic receptor. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18684–9 (2011).
 18. Martí-Solano, M., Schmidt, D., Kolb, P. & Selent, J. Drugging specific conformational states of GPCRs: challenges and opportunities for computational chemistry. *Drug Discovery Today* (2016). doi:10.1016/j.drudis.2016.01.009
 19. Martí-Solano, M., Guixà-González, R., Sanz, F., Pastor, M. & Selent, J. Novel insights into biased agonism at G protein-coupled receptors and their potential for drug design. *Curr. Pharm. Des.* **19**, 5156–66 (2013).
 20. Rasmussen, S. G. F. *et al.* Crystal structure of the β 2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–55 (2011).
 21. Kang, Y. *et al.* Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523**, 561–7 (2015).
 22. Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the β (2) adrenoceptor. *Nature* **469**, 175–80 (2011).
 23. Latek, D., Pasznik, P., Carlomagno, T. & Filipek, S. Towards improved quality of GPCR models by usage of multiple templates and profile-profile comparison. *PLoS One* **8**, e56742 (2013).
 24. Ballesteros, J. *et al.* Functional microdomains in G-protein-coupled receptors. The conserved arginine-cage motif in the gonadotropin-releasing hormone receptor. *J. Biol. Chem.* **273**, 10445–53 (1998).
 25. Filizola, M., Visiers, I., Skrabanek, L., Campagne, F. & Weinstein, H. in *Molecular Neuropharmacology: Strategies and Methods* (eds. Schousboe, A. & Bräuner-Osborne, H.) 235–266 (Humana Press, 2004). doi:10.1007/978-1-59259-672-0
 26. Weinstein, H. Hallucinogen actions on 5-HT receptors reveal distinct mechanisms of activation and signaling by G protein-

- coupled receptors. *AAPS J.* **7**, E871-84 (2005).
27. Bruno, A. & Costantino, G. Molecular Dynamics Simulations of G Protein-Coupled Receptors. *Mol. Inform.* **31**, 222–230 (2012).
 28. Krumm, B. E., White, J. F., Shah, P. & Grisshammer, R. Structural prerequisites for G-protein activation by the neurotensin receptor. *Nat. Commun.* **6**, 7895 (2015).
 29. Angel, T. E., Gupta, S., Jastrzebska, B., Palczewski, K. & Chance, M. R. Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14367–72 (2009).
 30. Angel, T. E., Chance, M. R. & Palczewski, K. Conserved waters mediate structural and functional activation of family A (rhodopsin-like) G protein-coupled receptors. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8555–60 (2009).
 31. Prioleau, C., Visiers, I., Ebersole, B. J., Weinstein, H. & Sealfon, S. C. Conserved helix 7 tyrosine acts as a multistate conformational switch in the 5HT_{2C} receptor. Identification of a novel ‘locked-on’ phenotype and double revertant mutations. *J. Biol. Chem.* **277**, 36577–84 (2002).
 32. Martí-Solano, M., Sanz, F., Pastor, M. & Selent, J. A dynamic view of molecular switch behavior at serotonin receptors: implications for functional selectivity. *PLoS One* **9**, e109312 (2014).
 33. Yuan, S., Filipek, S., Palczewski, K. & Vogel, H. Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway. *Nat. Commun.* **5**, 4733 (2014).
 34. Kim, S.-K., Riley, L., Abrol, R., Jacobson, K. A. & Goddard, W. A. Predicted structures of agonist and antagonist bound complexes of adenosine A₃ receptor. *Proteins* **79**, 1878–97 (2011).
 35. S.P. Sansom, M. & Weinstein, H. Hinges, swivels and switches: the role of prolines in signalling via transmembrane α -helices. *Trends Pharmacol. Sci.* **21**, 445–451 (2000).
 36. Pellissier, L. P. *et al.* Conformational toggle switches implicated in basal constitutive and agonist-induced activated states of 5-hydroxytryptamine-4 receptors. *Mol. Pharmacol.* **75**, 982–90 (2009).
 37. Ruprecht, J. J., Mielke, T., Vogel, R., Villa, C. & Schertler, G. F. X. Electron crystallography reveals the structure of

- metarhodopsin I. *EMBO J.* **23**, 3609–20 (2004).
38. Bai, Q. *et al.* Ligand induced change of β_2 adrenergic receptor from active to inactive conformation and its implication for the closed/open state of the water channel: insight from molecular dynamics simulation, free energy calculation and Markov state model analysis. *Phys. Chem. Chem. Phys.* **16**, 15874–85 (2014).
 39. Hothersall, J. D. *et al.* Residues W320 and Y328 within the binding site of the μ -opioid receptor influence opiate ligand bias. *Neuropharmacology* **118**, 46–58 (2017).
 40. Venkatakrishnan, A. J. *et al.* Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature* **536**, 484–487 (2016).
 41. Residues W320 and Y328 within the binding site of the μ -opioid receptor influence opiate ligand bias. *Neuropharmacology* **118**, 46–58 (2017).
 42. Bartuzi, D., Kaczor, A. A. & Matosiuk, D. Interplay between Two Allosteric Sites and Their Influence on Agonist Binding in Human μ Opioid Receptor. *J. Chem. Inf. Model.* **56**, 563–570 (2016).
 43. Woo, A. Y.-H. *et al.* Tyrosine 308 is necessary for ligand-directed Gs protein-biased signaling of β_2 -adrenoceptor. *J. Biol. Chem.* **289**, 19351–63 (2014).
 44. Kumari, P. *et al.* Core engagement with β -arrestin is dispensable for agonist-induced vasopressin receptor endocytosis and ERK activation. *Mol. Biol. Cell* **28**, 1003–1010 (2017).
 45. Katritch, V. *et al.* Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci.* **39**, 233–44 (2014).
 46. Bologna, Z., Teoh, J.-P., Bayoumi, A. S., Tang, Y. & Kim, I.-M. Biased G Protein-Coupled Receptor Signaling: New Player in Modulating Physiology and Pathology. *Biomol. Ther. (Seoul)*. **25**, 12–25 (2017).
 47. Björk, K., Sjögren, B. & Svenningsson, P. Regulation of serotonin receptor function in the nervous system by lipid rafts and adaptor proteins. *Exp. Cell Res.* **316**, 1351–6 (2010).
 48. Sjögren, B., Csöreg, L. & Svenningsson, P. Cholesterol reduction attenuates 5-HT_{1A} receptor-mediated signaling in human primary neuronal cultures. *Naunyn. Schmiedeberg's Arch. Pharmacol.* **378**, 441–6 (2008).
 49. Prasad, R., Paila, Y. D., Chattopadhyay, A. & Jafurulla, M.

- Membrane cholesterol depletion enhances ligand binding function of human serotonin_{1A} receptors in neuronal cells. *Biochem. Biophys. Res. Commun.* **390**, 93–6 (2009).
50. Sommer, B. *et al.* Extraction of membrane cholesterol disrupts caveolae and impairs serotonergic (5-HT_{2A}) and histaminergic (H-1) responses in bovine airway smooth muscle: role of Rho-kinase. *Can. J. Physiol. Pharmacol.* **87**, 180–195 (2009).
 51. Sjögren, B., Hamblin, M. W. & Svenningsson, P. Cholesterol depletion reduces serotonin binding and signaling via human 5-HT_{7(a)} receptors. *Eur. J. Pharmacol.* **552**, 1–10 (2006).
 52. Bruno, A., Costantino, G., de Fabritiis, G., Pastor, M. & Selent, J. Membrane-Sensitive Conformational States of Helix 8 in the Metabotropic Glu₂ Receptor, a Class C GPCR. *PLoS One* **7**, e42023 (2012).
 53. Khelashvili, G., Grossfield, A., Feller, S. E., Pitman, M. C. & Weinstein, H. Structural and dynamic effects of cholesterol at preferred sites of interaction with rhodopsin identified from microsecond length molecular dynamics simulations. *Proteins* **76**, 403–17 (2009).
 54. Shan, J., Khelashvili, G., Mondal, S., Mehler, E. L. & Weinstein, H. Ligand-Dependent Conformations and Dynamics of the Serotonin 5-HT_{2A} Receptor Determine Its Activation and Membrane-Driven Oligomerization Properties. *PLoS Comput. Biol.* **8**, e1002473 (2012).
 55. Mondal, S., Khelashvili, G., Johner, N. & Weinstein, H. *G Protein-Coupled Receptors - Modeling and Simulation. Advances in Experimental Medicine and Biology* **796**, (Springer Netherlands, 2014).
 56. Hospital, A., Goñi, J. R., Orozco, M. & Gelpí, J. L. Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.* **8**, 37–47 (2015).
 57. Johnston, J. M. & Filizola, M. Showcasing modern molecular dynamics simulations of membrane proteins through G protein-coupled receptors. *Curr. Opin. Struct. Biol.* **21**, 552–8 (2011).
 58. Lim, J. B., Rogaski, B. & Klauda, J. B. Update of the cholesterol force field parameters in CHARMM. *J. Phys. Chem. B* **116**, 203–210 (2012).
 59. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys.*

- Chem. B* **114**, 7830–7843 (2010).
60. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2016).
 61. Case, D. A. *et al.* AMBER 2017. (2017). at <<http://ambermd.org/>>
 62. Soares, T. A., Daura, X., Oostenbrink, C., Smith, L. J. & van Gunsteren, W. F. Validation of the GROMOS force-field parameter set 45A3 against nuclear magnetic resonance data of hen egg lysozyme. *J. Biomol. NMR* **30**, 407–422 (2004).
 63. Banks, J. L. *et al.* Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **26**, 1752–1780 (2005).
 64. Brooks, B. R. *et al.* CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–614 (2009).
 65. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
 66. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
 67. Harvey, M. J., Giupponi, G. & De Fabritiis, G. ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* **5**, 1632–1639 (2009).
 68. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
 69. Palczewski, K. *et al.* Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739–45 (2000).
 70. Kang, Y. *et al.* Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523**, 561–567 (2015).
 71. Mestre, T. A., Zurowski, M. & Fox, S. H. 5-Hydroxytryptamine 2A receptor antagonists as potential treatment for psychiatric disorders. *Expert Opin. Investig. Drugs* **22**, 411–421 (2013).
 72. Zhang, J. & Liu, Q. Cholesterol metabolism and homeostasis in the brain. *Protein Cell* **6**, 254–64 (2015).
 73. Ramírez Rosas, M. B., Labrujere, S., Villalón, C. M. & Maassen Vandenbrink, A. Activation of 5-hydroxytryptamine1B/1D/1F receptors as a mechanism of action of antimigraine drugs. *Expert Opin. Pharmacother.* **14**, 1599–610 (2013).
 74. Roth, B. L. Drugs and Valvular Heart Disease. *N. Engl. J.*

- Med.* **356**, 6–9 (2007).
75. Valentin-Hansen, L., Frimurer, T. M., Mokrosinski, J., Holliday, N. D. & Schwartz, T. W. Biased Gs versus Gq proteins and β -arrestin signaling in the NK1 receptor determined by interactions in the water hydrogen bond network. *J. Biol. Chem.* **290**, 24495–24508 (2015).
 76. Ballesteros, J. A. & Weinstein, H. in *Methods in Neurosciences* **25**, 366–428 (1995).
 77. Katritch, V. *et al.* Allosteric sodium in class A GPCR signaling. *Trends in Biochemical Sciences* **39**, 233–244 (2014).
 78. Selent, J., Sanz, F., Pastor, M. & De Fabritiis, G. Induced effects of sodium ions on dopaminergic G-protein coupled receptors. *PLoS Comput. Biol.* **6**, 1–6 (2010).
 79. Zocher, M., Zhang, C., Rasmussen, S. G. F., Kobilka, B. K. & Müller, D. J. Cholesterol increases kinetic, energetic, and mechanical stability of the human β 2-adrenergic receptor. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E3463–72 (2012).
 80. Manna, M. *et al.* Mechanism of allosteric regulation of β 2 - adrenergic receptor by cholesterol. *Elife* **5**, e18432 (2016).