Master en Pharmacie

Travail Personnel de Recherche

Structural Insight into the Allosteric Activation of LHCGR by ORG43553

présenté à la

Faculté des sciences de
L'Université de Genève

par

Raphaël Paul Feser

Unité de recherche:
Computational Biophysics
School of Pharmacy
University College London

Directeur de l'unité:
Prof. Shozeb Haider

Autres responsables

Prof. Francesco Luigi Gervasio

Genève

Année académique 2022-2023

Abstract

The luteinizing hormone-choriogonadotropin receptor (LHCGR) plays an important role in sexual development and pregnancy. Although LHCGR naturally recognizes endogenous LH and CG peptides, in vitro and mouse assays have shown that the small molecule ORG43553 is capable of allosterically activating the receptor and displays biased agonism. This ligand holds promise for the development of an orally active drug to replace the use of CG and LH which has limitations in the treatment of reproductive disorders and assisted reproductive technologies. Here we characterize elements of the binding and the effect of ORG43553 in the mouse LHCGR using homology modelling and molecular dynamics simulation. The results suggest differences in the human and murine binding sites, implying that the use of mice animal models is not optimal for predicting ORG43553 structure-related effects. Furthermore, an active-like conformation and mechanisms to stabilize this conformation have been demonstrated, suggesting that the ligand mediates its activity by stabilizing a specific conformation of the LHCGR transmembrane domain. Interactions with P520, F592 and I589 appear to be important in this regard. These findings, together with further research, may guide the design of safer and more effective drugs.

Résumé

Le récepteur de l'hormone lutéinisante et de la choriogonadotrophine (LHCGR) joue un rôle important dans le développement sexuel et la grossesse. Alors que le LHCGR reconnaît naturellement les peptides endogènes LH et CG, des essais in vitro ainsi que de essais chez la souris ont montré que la petite molécule ORG43553 est capable d'activer allostériquement le récepteur et se comporte comme un agoniste biaisé. Ce ligand est prometteur dans le développement d'un médicament actif par voie orale car il pourrait remplacer l'utilisation limitée de la CG et de la LH dans le traitement des troubles de la reproduction et des technologies de reproduction assistée. Dans le cadre de ce travail, nous caractérisons des effets liés à la liaison de ORG43553 dans le LHCGR de la souris en utilisant la modélisation par homologie ainsi que la simulation de dynamique moléculaire. Les résultats suggèrent des différences dans les sites de liaison humains et murins, ce qui indique que l'utilisation de modèles animaux murins n'est pas optimale pour prédire les effets liés à la structure d'ORG43553. En outre, une conformation de type active et des mécanismes de stabilisation de cette même conformation ont été mis en évidence. Cela suggère que le ligand exerce son activité en stabilisant une conformation spécifique du domaine transmembranaire des LHCGR. Les interactions avec P520, F592 et I589 semblent être importantes à cet égard. Ces résultats, ainsi que d'autres futures recherches, pourraient guider la conception de médicaments plus sûrs et plus efficaces.

Acknowledgement

I would also like to express my gratitude to my co-supervisors Prof. Shozeb Haider and Prof. Francesco Luigi Gervasio who actively accompanied and supported me in the process of coming to the this research unit in London. I appreciated this opportunity which, without them, would not have been possible.

I am also thankful to Prof. Shozeb Haider for including me in two exciting projects, one of which is discussed in this paper, and for remaining available to help me throughout.

I very much appreciate the valuable assistance of my colleague Saleh Alyemni in several of the challenges I had to face. I also enjoyed working with Jing Gu, Fedaa Attana, Shuang Chen and all the other colleagues in the research laboratory with whom I had enlightening discussions.

I acknowledge the Swiss European Mobility Program (SEMP) for providing me with scholarship funding.

Lastly, a special thanks to my parents for their encouragement and support.

Table of Contents

1 Introduction 9
1.1 G protein-coupled receptors 9
1.2 The Luteinizing hormone/choriogonadotropin receptor 10
1.3 Protein flexibility 11
1.3.1 Protein dynamics 11
1.3.2 Molecular switches 11
1.3.3 Biased agonism 11
1.4 Molecular Dynamic Simulation 12
1.5 Homology Modelling 13
1.5.1 Context 13
1.5.2 Templates Identification and selection 13
1.5.3 Backbone Generation 13
1.5.4 Refinement 13
1.6 Hypothesis and Objectives 14
2 Material and methods 15
2.1 Homology model 15
2.2 MD simulation 15
2.3 Structural mobility 17
2.4 Principal component analysis 17
2.5 Protein interactions 18
2.6 Data visualization 18
3 Results 19
3.1 Homology model and ligand stability 19
3.2 Differences in protein flexibility are observed comparing bound and unbound proteins 22
3.3 Effect of the bound and unbound state of the proteins on the sampled conformations 24
4 Discussion 29
5 Conclusions 31
6 References 32
7 Appendices 37

1 Introduction

1.1 G protein-coupled receptors

G protein-coupled receptors (GPCRs) are the largest family of proteins in humans and mediate the majority of external cellular stimuli (1). They are classified into six classes (labelled A-F) (2). Of the ~800 distinct GPCR genes in the human genome, over 700 encode GPCRs of class A, making them an important drug target. To mediate physiological responses, GPCRs are activated by a variety of external stimuli which includes peptides, small molecules, lipids and light (3). Class A GPCRs structures share an extracellular domain that varies significantly between the proteins, and a highly conserved transmembrane domain (TMD) consisting of 7 helices connected by intracellular and extracellular loops (figure 1 A). (3) Once activated by an external stimulus, the TMD initiates a signalling pathway by coupling to the α subunit of a G protein ($G\alpha$) (figure 1 B). Consequently, a conformational change will dissociate $G\alpha$ from its complex. This will allow the $G\alpha$ to modulate the concentration of intracellular signalling molecules by binding to an effector protein. Among the $G\alpha$ subunits, $G\alpha s$ interacts with the effector protein CA while $G\alpha q$ subunits interact with PLC (1). Other signalling pathways have been identified, some of which do not include the G protein (4).

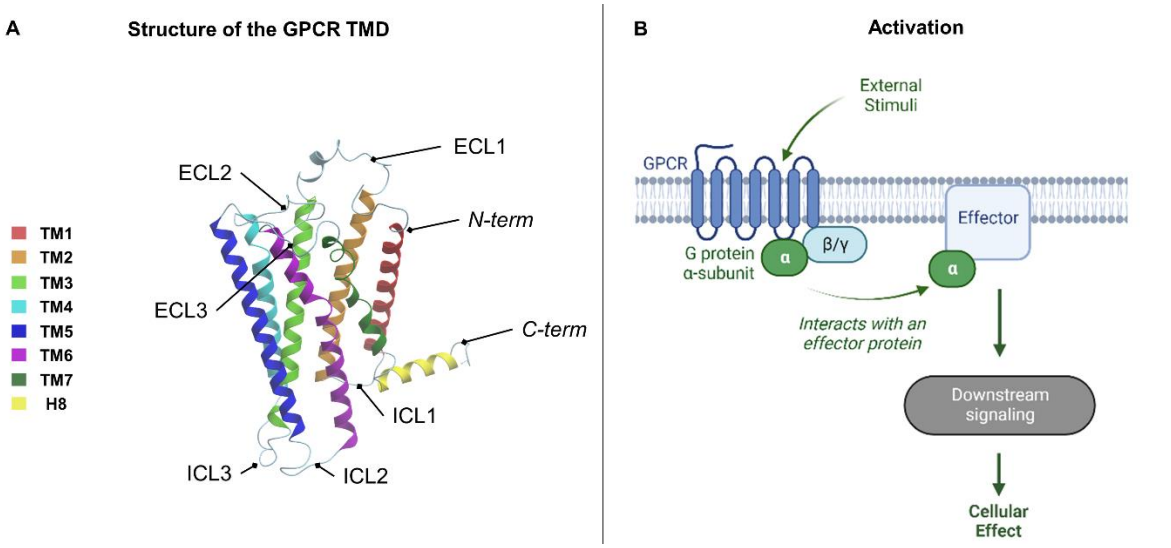


Figure 1: *Class A GPCR structure and activation*

A) Cartoon representation of a GPCR TMD (PBD ID 7FIJ). Seven transmembrane helices are interconnected by Intracellular loops (ICL) and Extracellular loops (ECL). B) Illustration of the GPCR activation mechanism through $G\alpha$ created with BioRender.com. External stimuli induce a conformational change in the TMD. This allows the TMD to form a complex with a neighbouring G protein by binding its α subunit. Consequently, another conformational change occurs which release and activate $G\alpha$. The latter modulates the concentration of intracellular signalling molecules by binding to an effector protein. The nature of the effector protein and the $G\alpha$ subunit mediating the signal determines the produced cellular effect.

1.2 The Luteinizing hormone/choriogonadotropin receptor

The Luteinizing hormone/choriogonadotropin receptor (LHCGR) is a G protein-coupled receptor (GPCR) of the class A subfamily that plays an important role in sexual development and pregnancy (5). The receptor is activated by Luteinizing Hormone (LH) and Chorionic Gonadotropin (CG), both of which bind to its large extracellular domain (ECD). Remarkably, LHCGR can trigger several activation pathways, notably the adenylyl cyclase and the phospholipase C pathway by coupling with $G\alpha s$ and $G\alpha q$ proteins, respectively (6,7). The structure of active LHCGR in complex with CG and $G\alpha s$ was recently solved (8). It suggests that upon binding CG, the ECD triggers conformational changes in the TMD i.e., an outward movement of TM6 and TM5, and an inward movement of TM7. Additionally, TM6 kinks at M582 and D578 as a result of those conformational changes. LH and CG are used as therapeutic agents in female and male reproductive disorders and assisted reproductive technologies (9). Due to their peptide nature, they are subject to post-translational modifications that can alter their activity, also, they require parenteral administration, something that is known to decrease patient compliance. These disadvantages should be avoided with the use of small orally active agonists (10). ORG43553 is a thienopyrimidine derivative that was found to induce ovulation in rodents by selectively stimulating the adenyl cyclase pathway. It is hypothesized that ORG43553 achieves its signalling selectivity by stabilizing a conformation specifically recognized by the $G\alpha s$ protein (11). In the same study that solved the structure of the CG-LHCGR-$G\alpha s$ complex (8), the structure of LHCGR bound to ORG43553 was solved, showing that the molecule binds to LHCGR at an allosteric site on the extracellular side of the TMD. (Figure 2). Interestingly, the binding site is similar to the orthosteric site of many GPCRs (8,11).

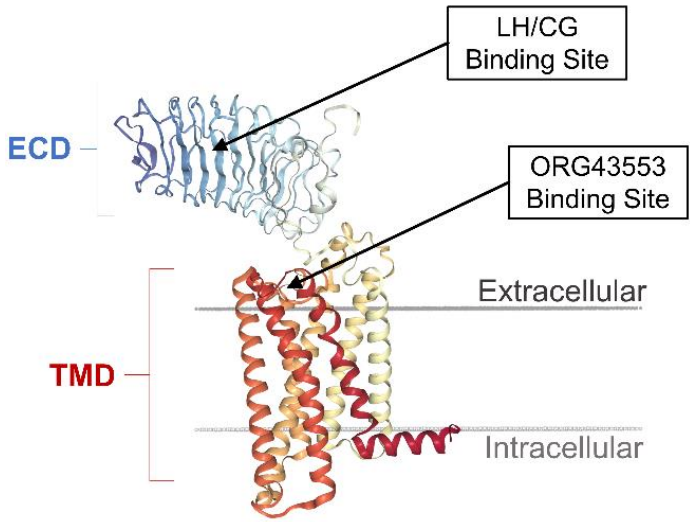


Figure 2: *Cartoon representation of the LHCGR receptor with its large extracellular domain (ECD) and the transmembrane domain (TMD). The binding site of LH, CG and ORG43553 are indicated with an arrow. It is similar to the orthosteric site of other class A GPCRs*

1.3 Protein flexibility

1.3.1 Protein dynamics

Proteins are in constant motion in living organisms. This is particularly true for GPCRs, which sample an immense number of inactive, intermediate and active conformations (12,12). The well-established flexibility of proteins renders obsoletes the "Lock and Key" model, in which proteins are rigid structures that recognize specific molecular patterns. Models that take flexibility into account are preferable since flexibility has been demonstrated to determine their various functions (13,14). In "Induced Fit models" the binding of a ligand induces a conformational change in the protein structure while in the "Conformational Selection models", the binding is allowed by naturally occurring conformations in the protein. Both models describe many events, and it should be noted that they are not mutually exclusive. The true protein-ligand binding mechanism can be a compromise between the two models, with maybe one of the two being dominant (15,16). Studying the dynamics of proteins provides interesting information about their behaviour, for example by defining desirable protein conformations for pharmacological targeting or by predicting ligand-induced effects. In this way, the characterization of dynamic protein-ligand interactions is fundamental to drug discovery and can guide the design of safe and effective drugs.

1.3.2 Molecular switches

Experiments indicate that even unbound, GPCRs can sample active conformations, this explains why, even in the absence of ligands, LHCGR and many GPCRs exhibit basal activity (17) Intramolecular interactions are known to stabilize inactive conformations, thereby limiting the basal activity of GPCR, these interactions are known as molecular switches and are characterized in many GPCRs (18).

1.3.3 Biased agonism

As mentioned above, ORG43553 selectively induces particular signalling in LHCGR, this is a typical example of biased agonism and might have important clinical implications (19,20). Evidence suggests that biased agonism may in part be due to the ability to favour the disruption of distinct molecular switches, resulting in the recruitment of specific G proteins (18).

1.4 Molecular Dynamic Simulation

Molecular dynamics (MD) simulation is a computational technique that allows the temporal evolution of biological systems to be captured on the basis of Newton's equations of motion. MD simulations are conceptually deterministic, they rely on molecular models referred to as force fields. Force fields are complex energy functions that govern the motion of a system by modelling the forces applied to each particle at any given time. They consider bonded and non-bonded interactions based on experimental and empirical parameters (21,22). All-atom force fields provide detailed atomistic information by computing the energy at the scale of each atom, considering them as single particles, however, they are computationally expensive. Coarse-grained force fields simplify the models and thus require fewer computational resources by considering particular groups of atoms as single particles. They are used to study the process of large biological systems, MD Simulation programs work as an iterative process where, for each timestep, forces are derived from the force field and the velocities and positions of the atoms are integrated by numerically solving Newton's equations of motion. The programs typically use Verlet-derived numerical integrators such as the Velocity Verlet algorithm, used in NAMD and ACEMD.

$$\begin{aligned} \mathbf{x}(t+\Delta t) &= \mathbf{x}(t) + \mathbf{v}(t)\,\Delta t + 1/2\ \mathbf{a}(t)\Delta t^2, \\ \mathbf{v}(t+\Delta t) &= \mathbf{v}(t) + \frac{\mathbf{a}(t)+\mathbf{a}(t+\Delta t)}{2}\Delta t. \end{aligned} \qquad (1)$$

The wise choice of the timestep represented in (eq. 1) by Δt is crucial for a simulation, the timestep should sample the fastest motions which are the bonds involving hydrogen. For this, a timestep of 1-4 fs is advised. Too large values would destabilize the algorithm and leads to an energy convergence problem. while too small timesteps would require too much computational power and limit the extent of the simulation. There are however several tricks to increase the timestep and still maintain accuracy, for instance, holonomic constraints can be applied to the hydrogens bond with algorithms such as SHAKE, or one can decrease the frequency of hydrogen motions by using hydrogen mass partitioning (23,24). The deterministic nature of MD simulation makes the choice of the initial structure crucial as it will mostly determine the simulation outcome. For this reason, the actual launching of the simulation is preceded with a setup where we carefully prepare the systems to best represent the studied process and an equilibration where a restraint simulation is run to remove bad contacts and clashes.

1.5 Homology Modelling

1.5.1 Context

Protein 3D configurations are the principal determinant of their function, knowing them allows for rational therapeutic design (25). For this reason, a lot of attention is given to constructing protein structure models; Experimental structural prediction using X-ray crystallography, NMR spectroscopy, and electron microscopy techniques can provide high-resolution structures (26) while having the disadvantage of being tedious and expensive. As a result of these limitations, there is a gap between the number of annotated protein-coding genes and the number of experimentally known structures (25). To date, many computational methods have been developed to build fast and efficient protein structure models. Homology modelling methods are good for predicting the structure of proteins based on a template that shares sequence and phylogenetic homology (27). The strength of homology modelling methods comes from the observation that in related proteins, structures are much more conserved than their sequence i.e. if two related sequences have a high sequence similarity, one can assume an even higher high structural similarity (28).

1.5.2 Templates Identification and selection

In this step, homologous proteins are identified in online databases based on sequence similarity. Due to the enormous size of these databases, heuristic algorithms have been developed such as the Basic Local Alignment Search Tool (BLAST) which is the most frequently used (29). After shortlisting candidate experimental structures, several factors were considered other than just the sequence similarity. Features such as the resolution of the structure, and the environment of the experiment are important. One should not forget to check for phylogenetic information when selecting templates. Finding a protein that shares the same function as our target in a closely related species can only strengthen the before-mentioned assumption on structural conservation.

1.5.3 Backbone Generation

The backbone is first created and then the side chains are inserted into the model (predict side chain conformers). The different “kinds” of methods used to generate the models based on the target-template(s) alignment. Rigid-body assembly methods generate the backbone of the model by assembling the core of small rigid bodies gathered in aligned regions. SWISS-MODEL is an example of a program using the rigid-body assembly method (30). Another approach used in the Modeller program is to construct the model by optimizing the target conformation so that it satisfies, at best, the spatial constraints obtained from the target-model alignment and other sources (31).

1.5.4 Refinement

A refinement of the model is usually performed to eliminate big errors, especially steric clashes. For this purpose, energy minimization methods are performed, using restricted Molecular Dynamics (section 1.3) or Monte Carlo simulation. In the entire process of homology, modelling

errors can be introduced. Model error detection tools such as PROCHECK allow us to evaluate the final model by analysing the stereochemistry (32).

1.6 Hypothesis and Objectives

ORG43553 is a small molecule that was found to selectively bind and activate LHCGR. Such molecules are interesting for the development of drugs, providing a more convenient alternative to the current parenteral administration of CG and LH, in the treatment of female and male reproductive disorders as well as assisted reproductive technology.

Whereas endogenous CG and LH bind to the ECD of LHCGR, ORG43553 binds to an allosteric site in the TMD. The molecule functions as a biased agonist. It has been hypothesized that this ability is exerted by stabilizing a particular conformation of the TMD which is directly responsible for G protein recruitment. A great effort has been made to understand and characterize the binding of ORG43553. The activity of the ligand has been evaluated in several assays in mice, with promising results.

In the present work, a homology model of the mouse ORG43553-bound LHCGR TMD (LHCGRCpx) was constructed and a model of the Mouse LHCGR TMD Apo (LHCGRApo) was retrieved from the alphafold database (33). Both models were prepared to enable the study of their dynamic behaviour using MD simulation. The aim was to better understand the binding and effect of ORG43553 on the receptor.

For further analysis, we hope to compare these results with the MD simulations of the human LHCGR that have been set up and are still running. Such information could also be useful in the design of safe and effective drugs and in evaluating the use of mice as in-vivo human models.

2 Material and methods

2.1 Homology model

Ligands are absent from Alphafold structure predictions, as the the latter is not designed to predict non-protein components. For this reason, a homology model of the Mouse LHCGR TMD bound to ORG43553 ($LHCGR_{Cpx}$) was constructed using the crystal structure of the Human LHCGR TMD bound to ORG43553 (PDB ID 7FIH) as a template. (8). The model was aligned and built using Modeller, five models were generated of which the best was selected, based on the DOPE score (31). The model was then evaluated with PROCHECK and considered a reasonable model (figure 21). To predict the protonation state of the protein, the Protein Prepare tool from PlayMolecule was used (34). Based on that prediction I protonated D409 and E455. A disulphide bond was added between C443 and C518 and the last two cysteines i.e., C647 and C648 were palmitoylated. The ligand was parameterized using CHARMM General Force Field.

2.2 MD simulation

The molecular model of the Mouse LHCGR TMD Apo ($LHCGR_{Apo}$) was retrieved from the alphafold database and prepared using the same following procedures (33). CHARMM-GUI membrane builder was used to prepare the lipid bilayer and assemble it with the protein in a periodic cell with water and ions (35). During the process, the protein was oriented in the membrane using the PPM2.0 predictions (35). We used a homogeneous 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer membrane for its common use and simplicity. It's a good model and it is abundant in animal membrane environments (36). The constructed membrane consisted of 200 and 202 lipids respectively in the upper and lower leaflet. The protein was embedded in the constructed membrane using the replacement method. The complex was solvated with the TIP3P CHARMM water model on top and below to obtain a rectangular periodic cell of 122.3x122.3x148.5. (Figure 3) The system was finally neutralized with the corresponding of 0.15M of KCl ions.

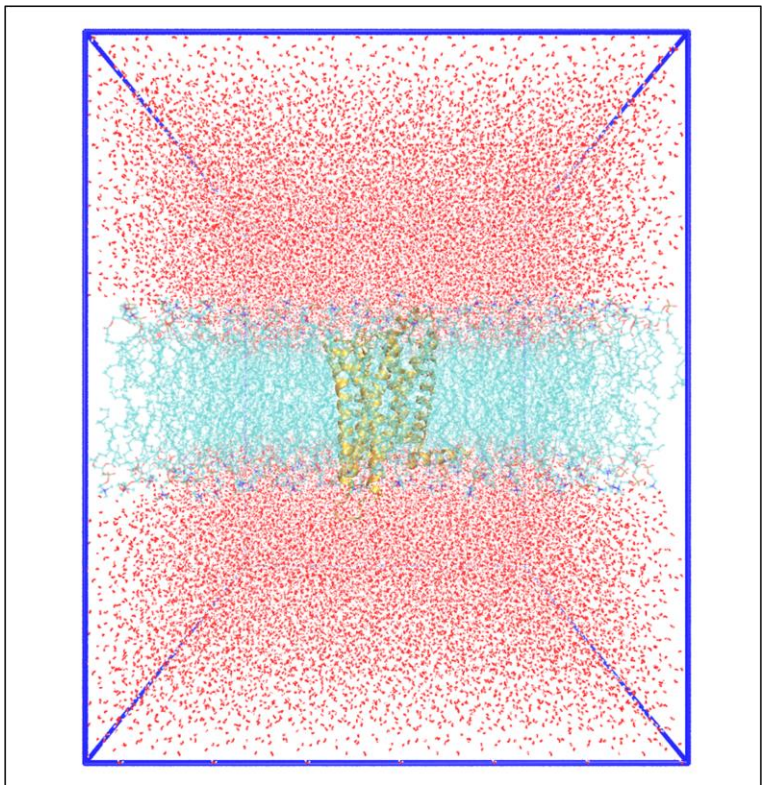


Figure 3: *The simulation cell prepared with CHARMM-GUI. The blue frame delimits the simulation cell, LHCGR is shown in yellow cartoon representation, and membrane lipids are shown as lines with the hydrophobic heads in dark blue and the hydrophobic tail in cyan. Water is represented in red dots.*

The system was equilibrated in NAMD using the standard CHARMM-GUI input files. It consists of six simulations with gradually reduced harmonic restraints. The timestep was 1 and 2fs respectively for the three first and last steps. The first two are under constant volume and temperature (NVT) and the last 4 under constant pressure, area and temperature (NVET). All bonds involving hydrogen were constrained using the SHAKE algorithm (37).

Table 1: *Equilibration steps*

Step	Ensemble	Number of steps	dt [fs]	time [ns]
1	NVT	12'500	1	0.125
2	NVT	12'500	1	0.125
3	NPT	12'500	1	0.125
4	NPT	25'000	2	0.50
5	NPT	25'000	2	0.50
6	NPT	25'000	2	0.50

Both the equilibration and simulation were run using the CHARMM36 force field under periodic boundary conditions (38). The NVET equilibration steps were kept at 1.01325 atm by a Langevin piston with a period of 50 fs and a damping time scale of 15 fs. The NVT equilibration and the simulation were kept at 310 K using a Langevin thermostat with damping of respectively 1 and 0.1 ps-1, electrostatics were approximated using the Particle-Mesh Ewald method with a cut-off of 9 Å for short range and Van Der Waals interactions. The simulation was run for 1000 ns in duplicates without restraints using ACEMD engine, the timestep was 2fs and the frames were gathered every 0.1 ns. The trajectories resulting from the MD simulations were post-processed using the PyTraj package (39).

2.3 Structural mobility

The root-mean-square deviation (RMSD) is a metric used to compare the structural deviation of two aligned protein structures. It can also be useful for following the time evolution of a trajectory by aligning it and comparing each frame to the starting frame. The root-mean-square fluctuation (RMSF) gives a measure of the flexibility of a group of atoms and averages it by its number. RMSF of the proteins and RMSD of the ligand were measured using Bio3D (40).

RMSD and RMSF are both useful indicators of stability and flexibility however their limitation is that they entirely rely on the rigid body superposition to the reference. To give complementary and additional insights, the MDlovofit package was used (41). It allows us to detect the subset of atoms presenting the smallest trajectory averaged RMSD for any given fraction size, the RMSD over time and the subset of atoms can be retrieved. A plot representing the trajectory averaged RMSD as a function of the fraction size was made and indicated that it was possible to align 70% of $LHCGR_{Apo}$ and $LHCGR_{Cpx}$ to less than 1.5 Å yFigure B1). Based on this observation, the fraction size of 70% was used to separate the protein and measure the RMSD over time of both fractions using MDlovofit.

2.4 Principal component analysis

Principal component analysis (PCA) is a non-parametric statistical method for reducing the dimensionality of large multivariate datasets such as protein trajectories (42). The algorithm standardizes the data, calculates the covariance for each set of dimensions and constructs a covariance matrix. The matrix is then diagonalized to obtain the corresponding eigenvalues and eigenvectors. The descending ranking of the eigenvectors according to their eigenvalues reveals a first eigenvector, representing the direction accounting for the largest variability, a second eigenvector, representing the second largest variability, and so on. By projecting the data onto each eigenvector, we obtain the principal components (PC) (43). The interest of a PCA is that for example in trajectory analysis, we can represent a large part of the protein dynamics with a few sets of PC (42).

PCA was performed on the coordinates of the self-superimposed c-alpha backbone of $LHCGR_{Apo}$ and the self-superimposed c-alpha backbone of $LHCGR_{Cpx}$ using R and Bio3d (40).

2.5 Protein interactions

Unless otherwise specified, all distance data were calculated between the two c-alpha of interest using Python and the MDAnalysis package (44,45). 2D ligand binding site diagrams were generated using the ICM-Pro software (Molsoft LLC, La Jolla, CA). To guide the analysis, protein-protein and Lipid-protein interactions were calculated using respectively MDCiao and PyLipID packages (46,47).

2.6 Data visualization

For visual consistency, all the plots were generated in R. Figures representing the proteins were generated using VMD and ICM-Pro (Molsoft LLC, La Jolla, CA) (48).

3 Results

3.1 Homology model and ligand stability

The homology model shared a 96% similarity with the template. The backbone superposition of the homology model and the template are completely aligned with the template with an RMSD of 0.2 Å (figure 4).

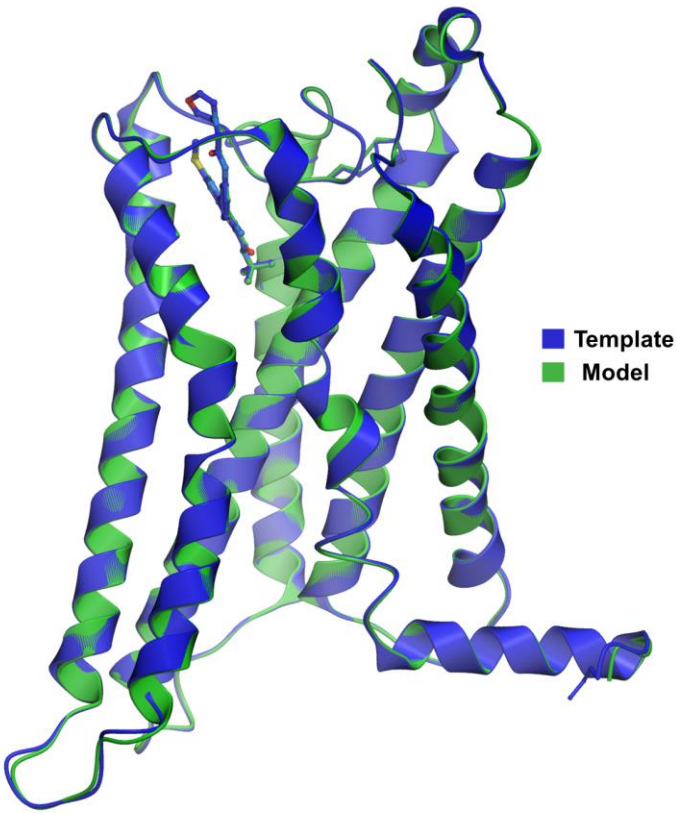


Figure 4: *Modelling of the LHCGR TMD. The selected homology model (green) is aligned and superimposed onto the template 7FIH (blue)*

In the homology model, the ligand occupies the same pocket as in the crystal template. To evaluate the ligand stability in this pocket over the simulation, the Root Mean-Square Deviation (RMSD) of ORG43553 fitted on the protein backbone was computed. The resulting plot suggests that the ligand undergoes an important rearrangement in the first 4ns and then remains relatively stable during the remaining simulation time (figure 5).

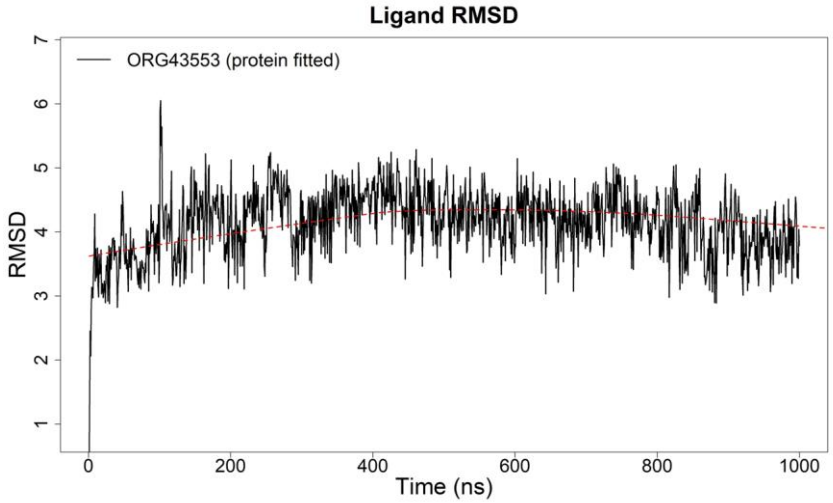


Figure 5: *RMSD values along the simulation time for ORG43553 fitted on the protein backbone (black lines). RMSD values after a LOWESS smoothing (red dots)*

The residues forming the binding site are well conserved between the two species (figure 6). There are, however, some key differences. Firstly, human F515 is permuted to L519 in the mouse gene. Secondly, human Y612 does not have its Y616 homolog in the mouse binding site, just as mouse I589 does not have its I585 homolog in the human binding site. Finally, in the human LHCGR template, the ligand shares a hydrogen bond with K595 and S604 (K599 and S608 in mice). In the mouse model, both the residues have different orientations, and the hydrogen bond interaction is lost.

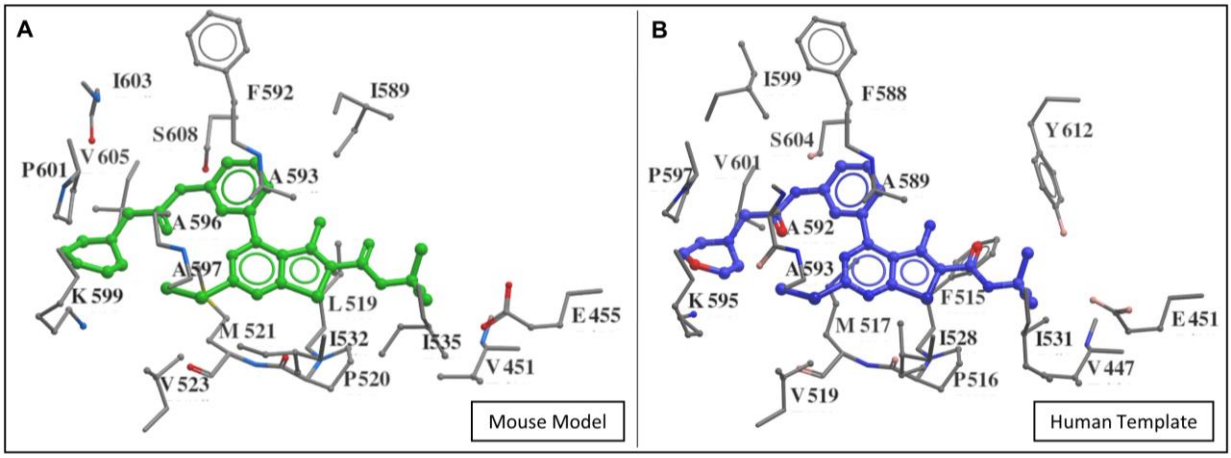


Figure 6: *Comparison of the model and the template binding sites.*
A) Binding site of the homology model ligand (green). B) Binding site of the crystal structure template ligand (blue).

Along the simulation, the distance between the ligand and K595 and S604 increases, until about 350 ns, where it remains stable (figure 7).

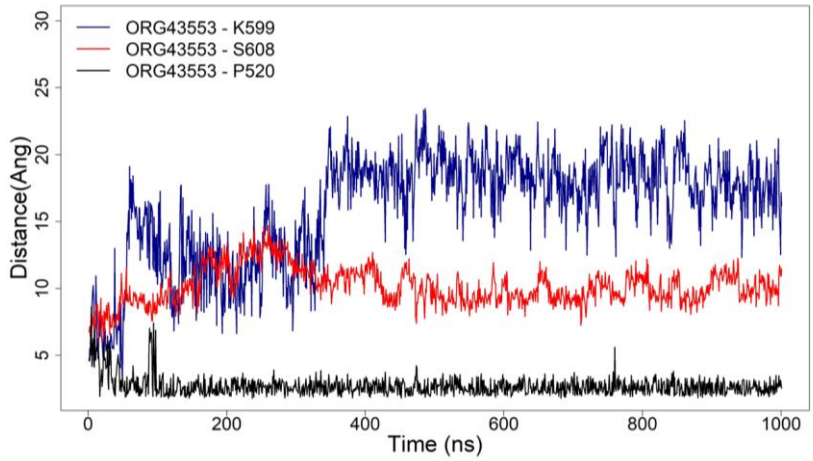


***Figure 7**: Temporal evolution of the distance separating ORG43553 and key residues in LHCGR$_{Cpx}$. The distance between the ORG43553 and the residues K599, S608, P520 are respectively represented in blue, red and dark*

The stabilised binding site at 350 ns is shown (figure 8 A), as well as the starting binding site (figure 8 B). It appears that the starting binding site is not stable for the ligand, leading it to change its pose and move closer towards P520 of TM3 and engage in a stable hydrogen bonding. We notice that in both binding sites the residues surrounding the ligand are predominantly hydrophobic, however, it should be noted that there are several water molecules in the ligand's environment that were not represented in the figure. They are known to actively participate in ligand binding and activation mechanisms of GPCRs (49).

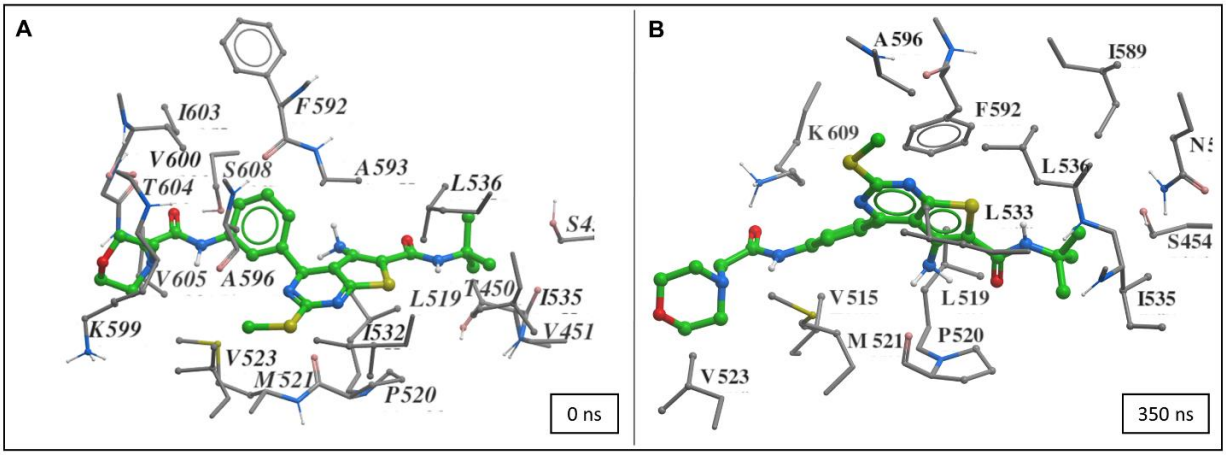


***Figure 8:** Evolution of the ligand binding domain.*
A) ORG43553 binding domain in LHCGRCpx before the simulation (0nss) B) ORG43553 binding domain in LHCGRCpx during the simulation after 350 ns

A different view on the starting and the 350 ns binding pose is shown in figure 9. On both panels, the starting structure of LHCGR$_{Cpx}$ is shown in grey and the 350 ns structure is in blue, only the ligand is changing. We observe that there is an outward movement of TM6 and interestingly, the pose of the ligand at 350 ns occupies the space of the starting structure TM6.

This movement is not observed in the $LHCGR_{Apo}$ simulation and seems to be a ligand-induced effect. The figure also allows us to see an inward movement of TM5. (50)

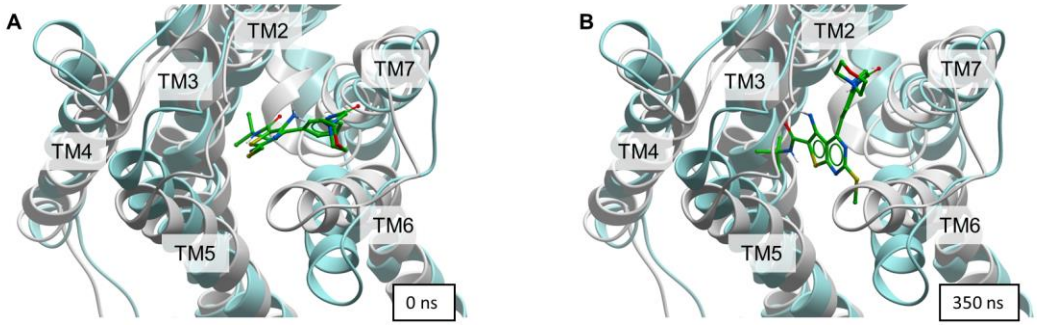


Figure 9: *Another view on the binding domain evolution, we see the two different poses at the beginning (0 ns) and during the simulation (350 ns). $LHCGR_{Cpx}$ at 0 ns (grey) is superimposed on itself at 350 ns (blue). A) In green, the ligand pose was observed at 0 ns. B) In green, the ligand pose observed at 350 ns is represented.*

3.2 Differences in protein flexibility are observed comparing bound and unbound proteins

As described in (section 2.3) the Package MDlovofit was used to separate the 70% best-aligned fraction from the most flexible one. Both the fractions were mapped to the 3D structure of $LHCGR_{Apo}$ and $LHCGR_{Cpx}$ (figure 10 A, C). We can see the extent of the outward motion on the extracellular end of TM6 which is not observable in the $LHCGR_{Apo}$ structure. Looking at the intracellular end of TM6 and TM7, we observe that a large part of the most flexible fraction is located here, especially in $LHCGR_{Apo}$. It is worth noting that the intracellular end of TM7 moving away from TM7 is described as particularly important in the activation of GPCR (18).

To evaluate the stability of the protein Root Mean Square Deviation RMSD was calculated, $LHCGR_{Apo}$ and $LHCGR_{Cpx}$ have a similar RMSD average (respectively 2.6 Å and 2.8 Å), when watching the time evolution of the RMSD, it appears that $LHCGR_{Cpx}$ structure is relatively stable throughout the simulation compared to $LHCGR_{Apo}$, furthermore, it seems that the latter undergoes a significant change at the beginning and approximately at 600 ns (figure 10 B, D).

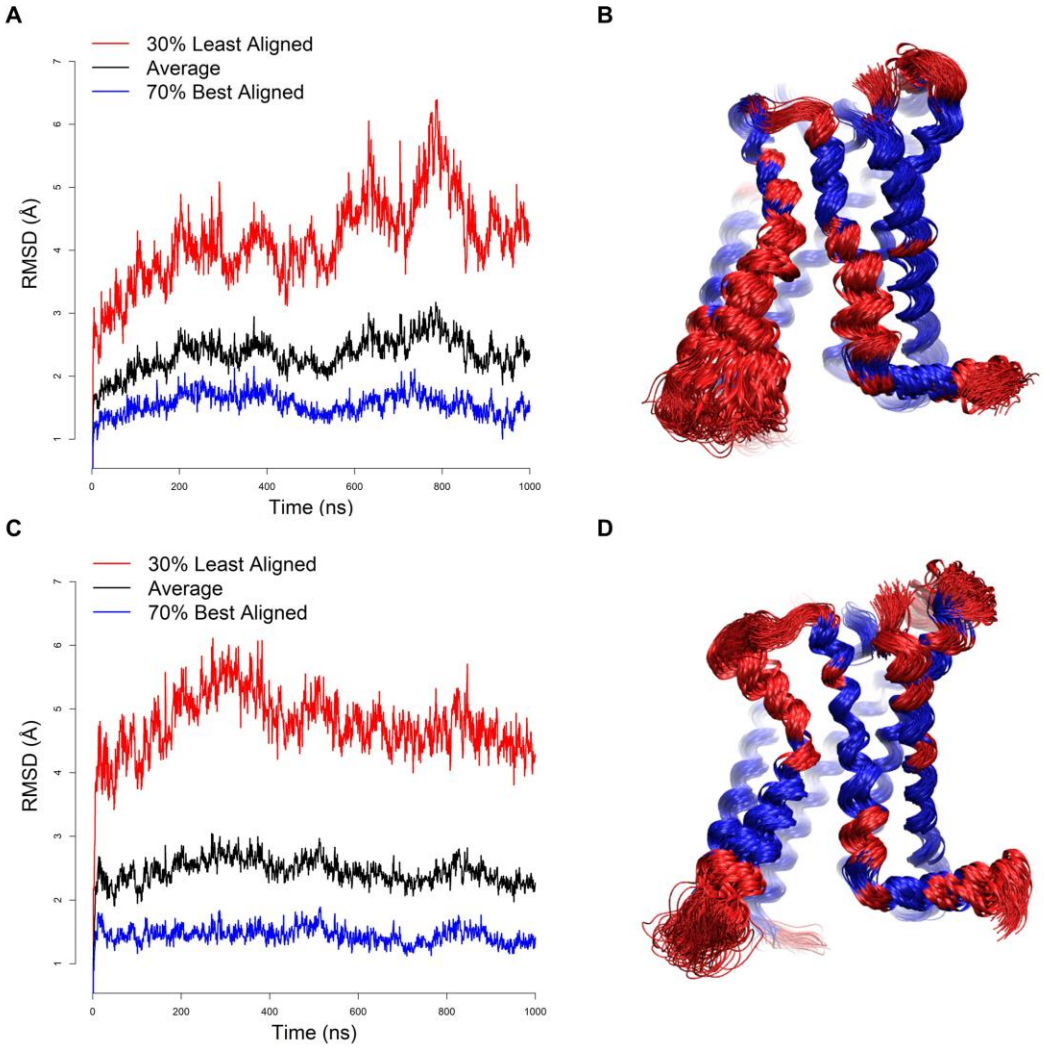


Figure 10: *Analysis of the protein's mobility A) RMSD values along the simulation time using MDlovofit alignment. The average RMSD is shown in black, the RMSD of the 70% best-aligned fraction is shown in blue and the RMSD of the 30% most flexible fraction is shown in red B) Superposition of LHCGR$_{Apo}$ simulation frames aligned using MDlovofit alignment. The 70% best-aligned fraction is shown in blue, and the 30% most flexible fraction is shown in red C) RMSD values along the simulation time using MDlovofit alignment. The average RMSD is shown in black, the RMSD of the 70% best-aligned fraction is shown in blue and the RMSD of the 30% most flexible fraction is shown in red D) Superposition of LHCGR Ligand bound simulation frames aligned using MDlovofit alignment. The 70% best-aligned fraction is shown in blue, and the 30% most flexible fraction is shown in red*

The RMSF plot allows us to see the respective deviation of each residue, without surprise the loops are the more flexible segments. Focusing on the TMs, with few exceptions we observe that the TM move less in LHCGR$_{Cpx}$. The extracellular end of TM6 (close to ICL3) exhibits a significantly higher deviation in LHCGR$_{Apo}$ compared to LHCGR$_{Cpx}$ and inversely in the intracellular end (close to ECL4).

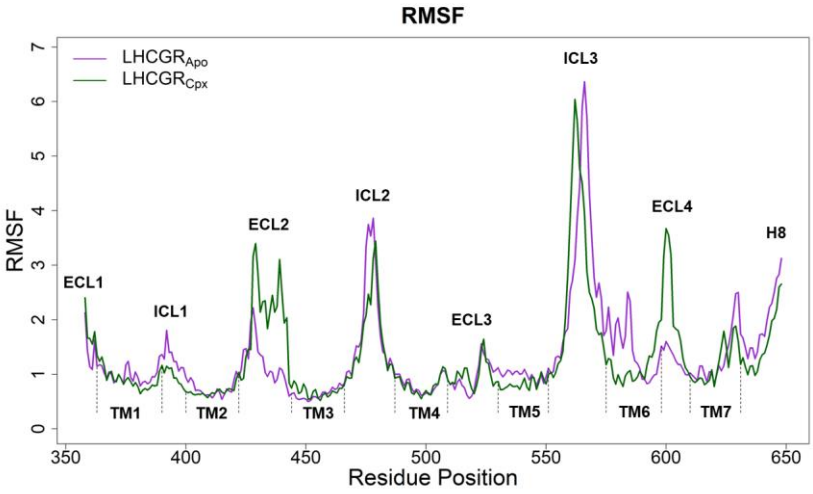


***Figure 11**: The respective RMSF values of each residue of LHCGR$_{Apo}$ (purple) and LHCGR$_{Cpx}$ (green)*

3.3 Effect of the bound and unbound state of the proteins on the sampled conformations

To probe the ligand effect on the protein conformational landscape, a PCA of the whole structure c-alpha backbone was performed. The space formed by the first two PC is represented in figure 12. In the plots, the colours spectrum adds a notion of time with blue being the first frames and red the lasts. The red frames are more clustered in LHCGR$_{Cpx}$ than in LHCGR$_{Apo}$. It looks like both undergo a change of the overall conformational state, but unlike LHCGR$_{Apo}$ the new state of LHCGR$_{Cpx}$ is stabilized. This supports what was seen earlier with RMSD and suggests that the ligand rigidifies the structure by stabilizing it in a more restricted set of conformations. TM6 outward motion and TM5 inward motion are considered activation paths in many GPCRs.

In (figure 13) we observe the displacement in the extracellular end and the intracellular end of the proteins. All the segments are displaced when we compare the LHCGR$_{Apo}$ structure (purple) to the LHCGR$_{Cpx}$ structure (green). In the superposition of the extracellular end (figure 13 A), the bigger gap is observed in TM6, TM7 and ECl1. Unlike TM7, the displacement of TM6 and ECL1 leaves more space in the binding site. In the intracellular end (figure 13 A), similar gaps are observed namely with TM6 and TM7 which undergo an inward and outward displacement respectively. The previously described displacement resembles the ones described as activation motions in Duan and al. research where they compared the human CG-bound LHCGR to the inactivated LHCGR (8).

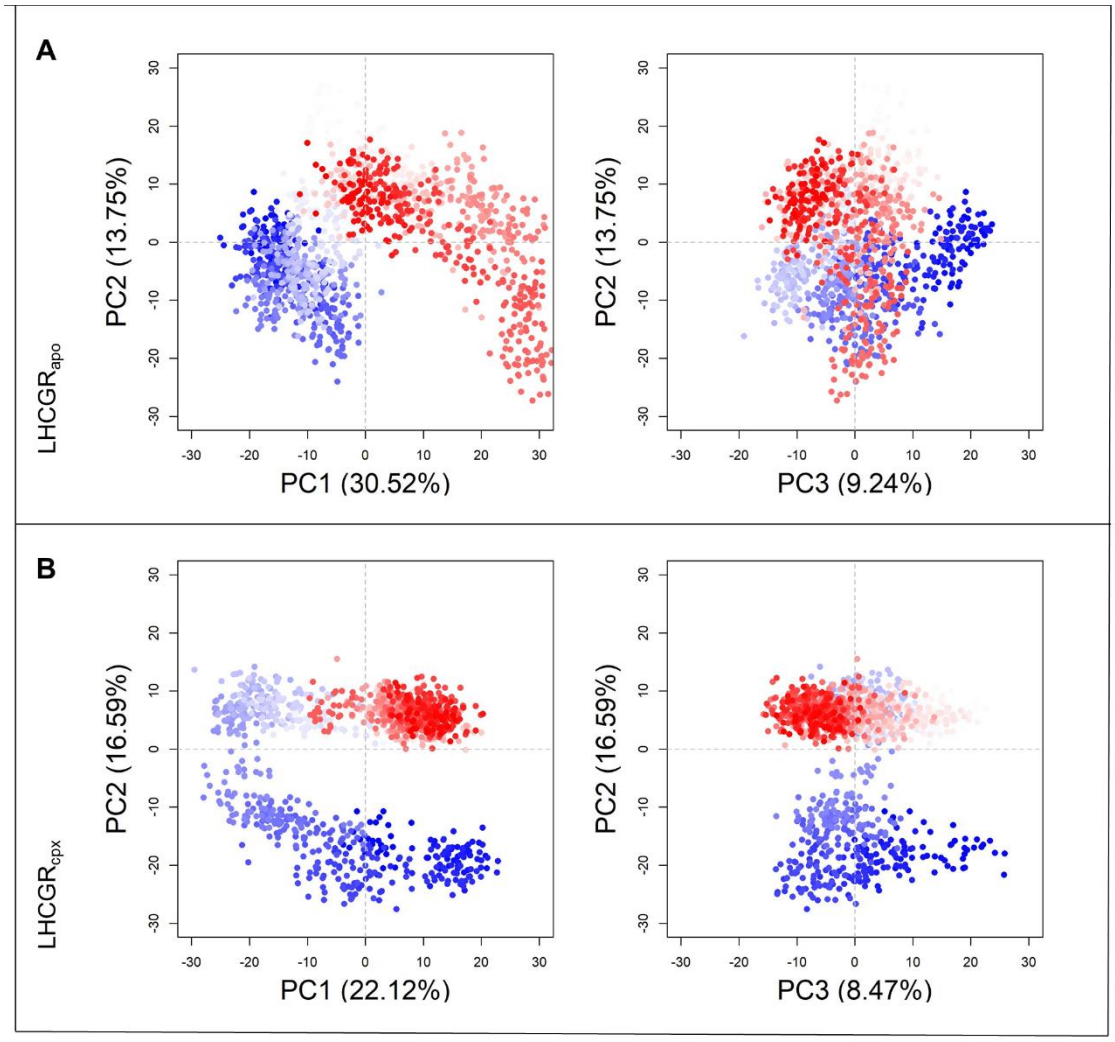


Figure 12: *PCA plot. The colour spectrum gives a notion of time (start with blue, then white and end with red) A) LHCGR$_{Apo}$ in the space of PC1-PC2 and PC2-PC3 B) LHCGR$_{Cpx}$ in the space of PC1-PC2 and PC2-PC3*

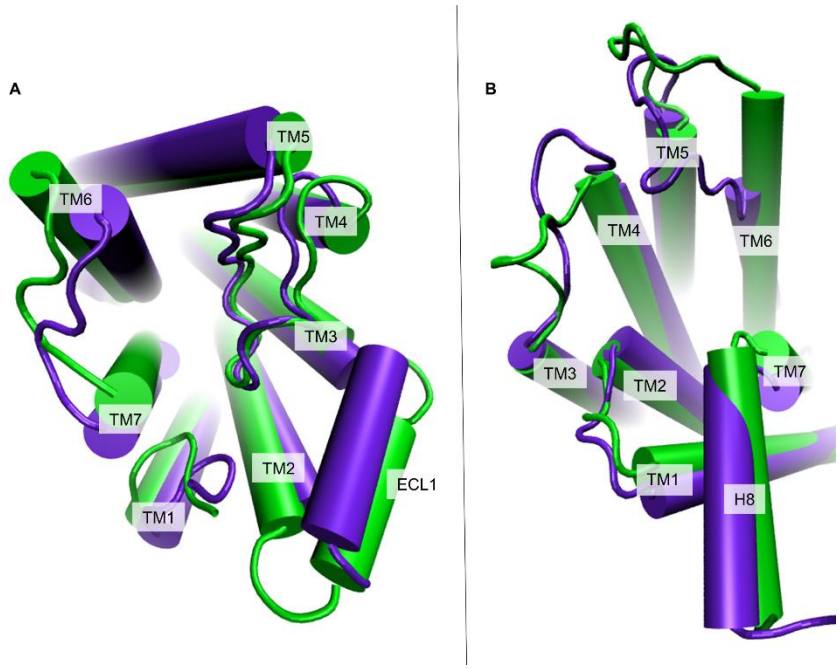


Figure 13: *$LHCGR_{Apo}$ and ligand-bound conformation.*
A) Extracellular view on the superimposition of $LHCGR_{Apo}$ in purple cylinder and $LHCGR_{Cpx}$ in green cylinder bs. Extracellular view on the superimposition of $LHCGR_{Apo}$ in purple cylinder and $LHCGR_{Cpx}$ in green cylinder

In figure 14 we see another view on the $LHCGR_{Cpx}$ TM6 outward displacement. The structure representation allows us to see $LHCGR_{Cpx}$ TM6 interacting with ORG43553 and $LHCGR_{Apo}$ TM6 interacting with TM3. TM6 conformational difference in $LHCGR_{Apo}$ and $LHCGR_{Cpx}$ seems to be reflected in a phospholipid binding site situated in between TM6 and TM7. In both structures, a phospholipid interacts with K573 and inserts in between the helix. While in $LHCGR_{Apo}$ the phospholipid is fully inserted and stabilized in the binding site, in $LHCGR_{Cpx}$ the phospholipid alternates between semi-inserted and not inserted state. This can be seen by looking at the phospholipid distance to their respective binding site (figure 15). The small distance between TM6 and TM3 in $LHCGR_{Cpx}$ compared to $LHCGR_{Apo}$ seems to be the cause (figure 16).

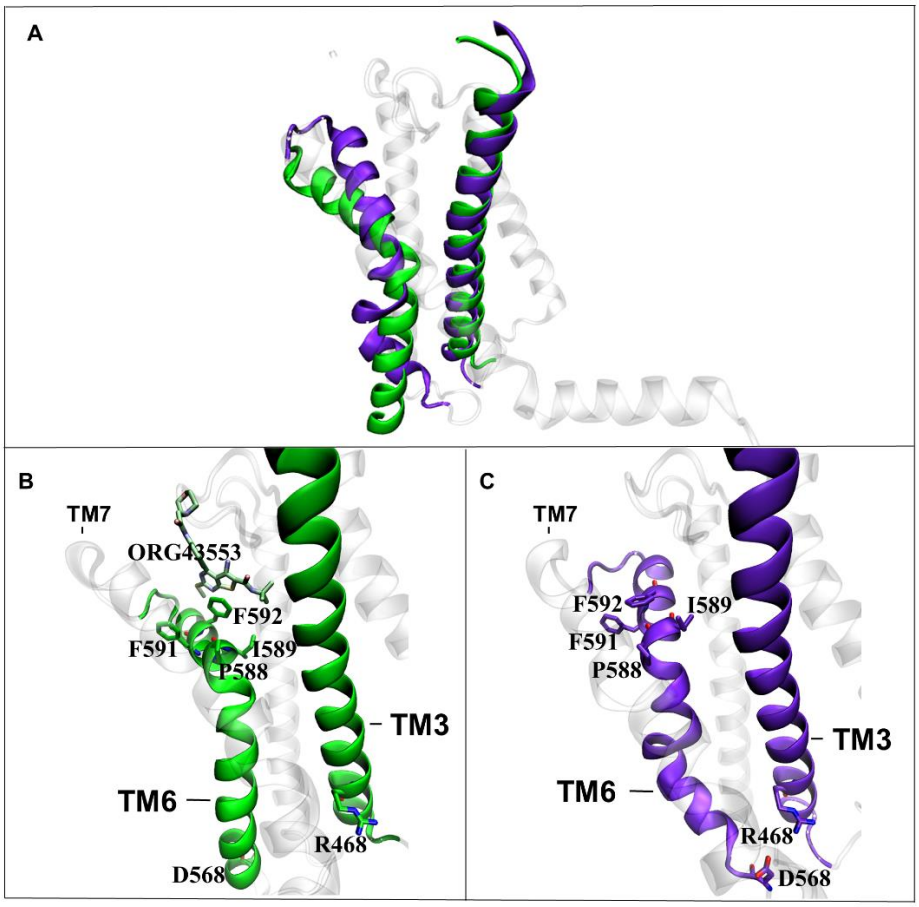


Figure 14: *Side views on TM6 bending.*
A) Superposition of $LHCGR_{Cpx}$ (TM3 and TM6 in green) and $LHCGR_{Apo}$ (TM3 and TM6 in purple) B) $LHCGR_{Cpx}$ TM3 and TM6. TM6 is shown interacting with ORG43553 and exhibiting a characteristic kink after the P588. C) $LHCGR_{Apo}$ TM3 and TM6. TM6 is shown bending and interacting with TM3 via the ionic lock (D568 and R468).

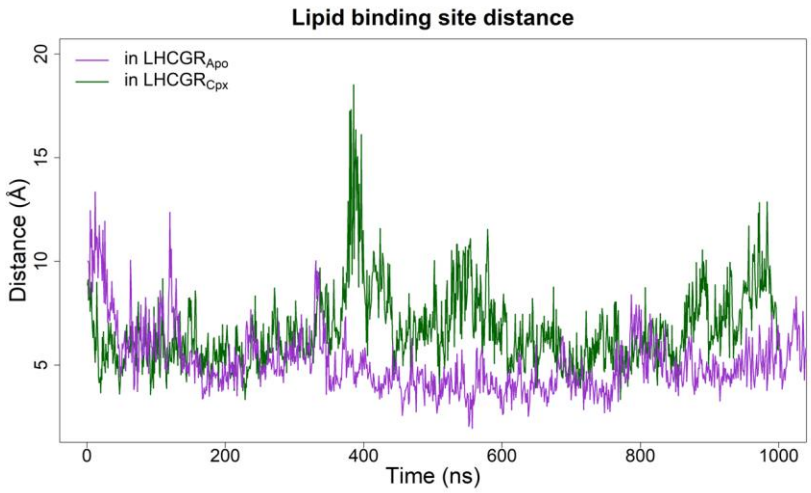


***Figure 15**: Temporal evolution of the distance separating interhelical phospholipids from their respective binding sites. The binding site is defined as the residues within 4 Å of the phospholipid at a given time. The purple line represents the distance observed in LHCGR$_{Apo}$ that separates the centre of mass of the phospholipid of interest and the centre of mass of its binding site. The green line represents the distance observed in LHCGR$_{Cpx}$ that separates the centre of mass of the phospholipid of interest and the centre of mass of its binding site.*

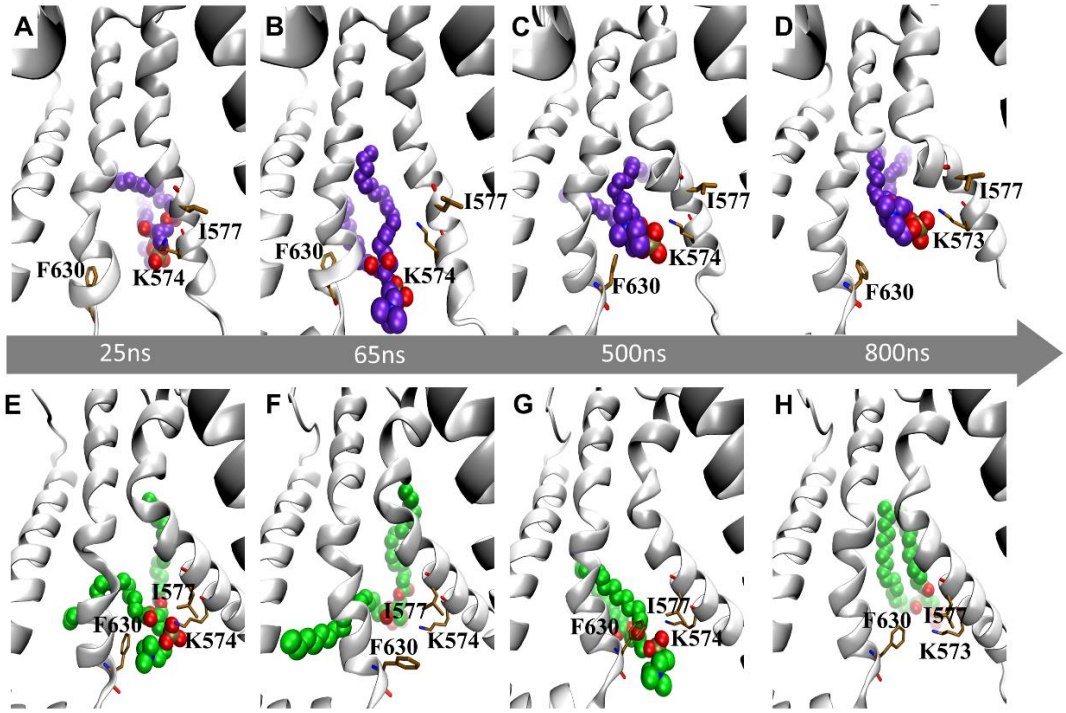


***Figure 16**: Phospholipid interhelical binding site. A-D) K574 attracts the POPC in the binding site by interacting with its hydrophilic head. F630 and I577 are sufficiently distant to allow the POPC to enter. E-H) K574 attracts the POPC in the binding site by interacting with its hydrophilic head but the POPC can't fully enter the binding site notably because of F630 and I577 which are close to each other.*

4 Discussion

In the past few years, efforts have been made to develop an orally active LHCGR agonist that could in the treatment of reproductive disorders and in assisted reproductive technologies. ORG43553 has shown promising results, it induces ovulation in mice, allosterically binds LHCGR and selectively triggers the adenyl cyclase pathway. In this study, we aimed to characterize the binding and effect of this drug on LHCGR. For this purpose, we built a homology model of the mouse receptor and used MD simulation to study its motion.

Looking first at the ligand binding in the mouse model, it starts with the same pose as the human crystal structure (figure 6) and is then reoriented towards other residues including P520. Several hypotheses could explain this, including the possibility that the binding does not occur at the same place in mice and humans, and the possibility that the ligand changes its pose over time both in humans and mice (figure 8). If the homology model is correct, the second hypothesis seems unlikely as the ligand is involved in two hydrogen bonds in the human predicted structure, the hypothesis could however be tested by examining the human LHCGR simulation. Alternatively, if the first hypothesis is verified, a difference in the ligand binding poses would entail a different structure-activity relationship between mouse and human, and therefore have major implications for drug development. For instance, the LHCGR in-vivo mouse model may not accurately predict the structure-related effects of ORG43553. A similar conclusion has been suggested by the research of Jia, X C et al. where they found that equine, rat and ovine CG and LH were able to competitively interact with the rat LHCGR but not the human one (51). Under either hypothesis, the results show the value of studying ligand binding in an MD simulation, such rearrangement over time would have been overlooked in static molecular docking.

The differences in stability between the $LHCGR_{Apo}$ and $LHCGR_{Cpx}$ structures observed with RMSD (figure 5), RMSF (figure 11) are consistent with the assumption that the ligand rigidifies the protein. Looking at the PCA (figure 12), the conformations representing the ending simulation frames (in red) are more clustered than the beginning frames (in blue). This supports the hypothesis that attempts to explain biased agonism, according to which the ligand is able to stabilise a specific conformational state only recognized by $G\alpha s$. Moreover, and in the same direction, the stabilised conformational state shown in figure 13 resembles the active conformation described in other GPCRs, especially with the outward movement of TM6 and the inward movement of TM7 (8,19).

Phospholipids have a role in the activating and deactivating mechanisms of GPCRs (52). In the simulations, we observe that a membrane phospholipid is inserted between TM6 and TM7 (figure 16), and the phospholipid even remains stable in $LHCGR_{Apo}$, which has the effect of distorting the helix. In $LHCGR_{Cpx}$ the phospholipid in question fails to remain in the binding site and repeatedly moves in and out (figure 15). It is thought, together with the following elements, that the conformation stabilized by the ligand, restricts access to its binding site. The interactions between ORG43553 thienopyrimidine nucleus and the side chain of F592, as well as the interaction between ORG43553 tert-butylamine and the side chain of I589, could be behind

the different orientation of the top of TM6 seen in $LHCGR_{Apo}$ and $LHCGR_{Cpx}$ (figure 14). This different orientation seems to propagate down the helix and to be the cause of the diminished distance between residues F630 in TM7, I577 and K573 in TM3 by a rotating movement. As a consequence, a steric hindrance occurs which prevents the phospholipid from stabilizing between the two helices. This result implies that the ligand can prevent the membrane from distorting TM6 which would highlight the important role of the allosteric ligand's interactions with F592 and I589. Going further, one could think that it is through this kind of mechanism that ORG43553 manages to function as a biased agonism and that a lot more of them exist. It is possible that by binding the receptor to the TDM, which is directly responsible for the recruitment of the G protein, it could induce a specific conformation, in a more precise manner than the endogenous ligand that binds the large ECD.

The "ionic lock" is a molecular switch involving an arginine of a highly conserved D-R-Y motif in TM3 (permutated to E-R-W in LHCGR) and an aspartate in TM6. The ion lock is closed when the two residues interpenetrate. GPCR inactivation induced by ion lock closure has been demonstrated in many GPCRs using experiments such as mutagenesis, crystallography, and MD simulations. For example, in LHCGR, arginine mutation has been linked to higher constitutive activity and has led to pathologies such as male-limited precocious puberty or Leydig cell adenoma. In the inactive crystal structure of human LHCGR, the ion lock is surprisingly open, the same is true for $LHCGR_{Apo}$ and, as expected, $LHCGR_{Cpx}$. However, in $LHCGR_{Apo}$, considering the time evolution of the simulated trajectory, we observe that, although they are far apart in the starting structure, the distance between R568 and D468 approaches less than 4 Å and then transits between more than 10 Å and less than 3.5Å. (figure B2) This is not observed in $LHCGR_{Cpx}$, so it would support the existence of this ionic lock and could explain what has been seen in the mutagenesis studies. Inexplicably, this transient interaction appears to be completely lost after 800 ns.

More generally, the results of the present work suggest that all these mechanisms are closely interconnected, perhaps without the phospholipid binding site, the discrimination between the conformations sampled by $LHCGR_{Apo}$ and $LHCGR_{Cpx}$ would be reduced. There are many other aspects to examine however time did not allow for further analysis, for example, as mentioned above, it would have been interesting to look at the involvement of water in the binding site which as mentioned before is known to be important in GPCRs activation. It would also have been interesting to look more closely at the movement of all the seven TMs and also the loops, which have not been much discussed. In $LHCGR_{Cpx}$ ECL2 and ECL4 have a much higher RMSF than in $LHCGR_{Apo}$ and there are probably elements in the trajectory that can explain this. Structural clustering could also provide useful insights into the protein's motions. There are also many validated methods for clustering trajectories into a set of related conformations based on geometric data (53). Some methods even include kinetic data, namely the Markov State Models (54).

5 Conclusions

Although the LHCGR receptor naturally recognises endogenous LH and CG peptides, in vitro and mice essays showed that the small molecule ORG43553 is able to activate it allosterically and display biased agonism. This project was undertaken to characterize the binding and the effects of ORG43553 on the mouse LHCGR. In this regard, a homology model was built for LHCGRCpx and LHCGRApo was retrieved from the alphafold2 database. The motions of both structures were sampled in an MD simulation experiment. The results indicate that ORG43553 stabilizes the protein in what appears to be an active conformational state. The interaction of ORG43553 with F582 and I589 respectively with its thienopyrimidine nucleus and its butyl end is important and involved in a stabilization mechanism. The simulation revealed a phospholipid binding site by which the membrane model deforms TM6 in the absence of a ligand. The results suggest that these mechanisms are closely interconnected and are not the only ones that exist. Other mechanisms could be highlighted by studies focusing on each system component and by using clustering methods. Additionally, the study of human LHCGR simulations would provide interesting information on the possible difference with its mouse homolog, just as the simulations of Apo FSHR and Apo TSHR could provide interesting information on the selectivity of ORG43553.

6 References

1. Rosenbaum DM, Rasmussen SGF, Kobilka BK. The structure and function of G-protein-coupled receptors. Nature. 2009;459(7245):356–63.

2. Kolakowski LF. GCRDb: A G-protein-coupled receptor database. Receptors & channels. 1994;2(1):1–7.

3. Fredriksson R, Lagerström MC, Lundin L-G, Schiöth HB. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. Molecular Pharmacology. 2003;63(6):1256–72.

4. Avet C, Mancini A, Breton B, Le Gouill C, Hauser AS, Normand C, et al. Effector membrane translocation biosensors reveal G protein and *B*arrestin coupling profiles of 100 therapeutically relevant GPCRs. eLife. 2022;11:e74101.

5. Troppmann B, Kleinau G, Krause G, Gromoll J. Structural and functional plasticity of the luteinizing hormone/choriogonadotrophin receptor. Human Reproduction Update. 2013;19(5):583–602.

6. Structural and functional plasticity of the luteinizing hormone/choriogonadotrophin receptor | Human Reproduction Update | Oxford Academic. https://academic.oup.com/humupd/article/19/5/583/613397;

7. Breen SM, Andric N, Ping T, Xie F, Offermans S, Gossen JA, et al. Ovulation Involves the Luteinizing Hormone-Dependent Activation of Gq/11 in Granulosa Cells. Molecular Endocrinology. 2013;27(9):1483–91.

8. Duan J, Xu P, Cheng X, Mao C, Croll T, He X, et al. Structures of full-length glycoprotein hormone receptor signalling complexes. Nature. 2021;598(7882):688–92.

9. Anderson RC, Newton CL, Anderson RA, Millar RP. Gonadotropins and Their Analogs: Current and Potential Clinical Applications. Endocrine Reviews. 2018;39(6):911–37.

10. Nataraja SG, Yu HN, Palmer SS. Discovery and Development of Small Molecule Allosteric Modulators of Glycoprotein Hormone Receptors. Frontiers in Endocrinology. 2015;6.

11. van Koppen CJ, Zaman GJR, Timmers CM, Kelder J, Mosselman S, van de Lagemaat R, et al. A signaling-selective, nanomolar potent allosteric low molecular weight agonist for the human luteinizing hormone receptor. Naunyn-Schmiedeberg's Archives of Pharmacology. 2008;378(5):503–14.

12. Bu Z, Callaway DJE. Proteins MOVE! Protein dynamics and long-range allostery in cell signaling. In: Advances in Protein Chemistry and Structural Biology. Elsevier; 2011. p. 163–221.

13. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Protein Function. Molecular Biology of the Cell 4th edition. 2002;

14. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE. Long-timescale molecular dynamics simulations of protein structure and function. Current Opinion in Structural Biology. 2009;19(2):120–7.

15. Morando MA, Saladino G, D'Amelio N, Pucheta-Martinez E, Lovera S, Lelli M, et al. Conformational Selection and Induced Fit Mechanisms in the Binding of an Anticancer Drug to the c-Src Kinase. Scientific Reports. 2016;6:24439.

16. Nussinov R, Ma B, Tsai C-J. Multiple conformational selection and induced fit events take place in allosteric propagation. Biophysical Chemistry. 2014;186:22–30.

17. Yao XJ, Vélez Ruiz G, Whorton MR, Rasmussen SGF, DeVree BT, Deupi X, et al. The effect of ligand efficacy on the formation and stability of a GPCR-G protein complex. Proceedings of the National Academy of Sciences. 2009;106(23):9501–6.

18. Schwartz TW, Frimurer TM, Holst B, Rosenkilde MM, Elling CE. Molecular Mechanism of 7tm Receptor Activationa Global Toggle Switch Model. Annual Review of Pharmacology and Toxicology. 2006;46(1):481–519.

19. Lu S, He X, Yang Z, Chai Z, Zhou S, Wang J, et al. Activation pathway of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design. Nature Communications. 2021;12(1):4721.

20. Smith JS, Lefkowitz RJ, Rajagopal S. Biased signalling: From simple switches to allosteric microprocessors. Nature Reviews Drug Discovery. 2018;17(4):243–60.

21. Schlick T. Force Fields. In: Schlick T, editor. Molecular Modeling and Simulation: An Interdisciplinary Guide. New York, NY: Springer; 2002. p. 225–58. (Interdisciplinary Applied Mathematics).

22. Braun E, Gilmer J, Mayes HB, Mobley DL, Monroe JI, Prasad S, et al. Best Practices for Foundations in Molecular Simulations [Article v1.0]. Living journal of computational molecular science. 2019;1(1):5957.

23. Elber R, Ruymgaart AP, Hess B. SHAKE parallelization. The European physical journal Special topics. 2011;200(1):211–23.

24. Balusek C, Hwang H, Lau CH, Lundquist K, Hazel A, Pavlova A, et al. Accelerating Membrane Simulations with Hydrogen Mass Repartitioning. Journal of Chemical Theory and Computation. 2019;15(8):4673–86.

25. Agnihotry S, Pathak RK, Singh DB, Tiwari A, Hussain I. Chapter 11 - Protein structure prediction. In: Singh DB, Pathak RK, editors. Bioinformatics. Academic Press; 2022. p. 177–88.

26. Baker D, Sali A. Protein structure prediction and structural genomics. Science. 2001;294(5540):93–6.

27. Vyas VK, Ukawala RD, Ghate M, Chintha C. Homology modeling a fast tool for drug discovery: Current perspectives. Indian Journal of Pharmaceutical Sciences. 2012;74(1):1–17.

28. Illergård K, Ardell DH, Elofsson A. Structure is three to ten times more conserved than sequence–a study of structural response in protein cores. Proteins. 2009;77(3):499–508.

29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990;215(3):403–10.

30. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Research. 2018;46(W1):W296–303.

31. Fiser A, Šali A. Modeller: Generation and Refinement of Homology-Based Protein Structure Models. In: Methods in Enzymology. Academic Press; 2003. p. 461–91. (Macromolecular Crystallography, Part D; vol. 374).

32. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. Journal of Applied Crystallography. 1993;26(2):283–91.

33. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research. 2022;50(D1):D439–44.

34. PlayMolecule ProteinPrepare: A Web Application for Protein Preparation for Molecular Dynamics Simulations | Journal of Chemical Information and Modeling. https://pubs.acs.org/doi/abs/10.1021/acs.jcim.7b00190;

35. Wu EL, Cheng X, Jo S, Rui H, Song KC, Dávila-Contreras EM, et al. CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations. Journal of computational chemistry. 2014;35(27):1997–2004.

36. Lipid Bilayers: Clusters, Domains and Phases - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4377075/;

37. Scalable Molecular Dynamics with NAMD - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2486339/;

38. CHARMM: The biomolecular simulation program - Brooks - 2009 - Journal of Computational Chemistry - Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1002/jcc.21287;

39. Roe DR, Cheatham TEI. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of Chemical Theory and Computation. 2013;9(7):3084–95.

40. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: An R package for the comparative analysis of protein structures. Bioinformatics (Oxford, England). 2006;22(21):2695–6.

41. Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis | PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0119264;

42. Palma J, Pierdominici-Sottile G. On the Uses of PCA to Characterise Molecular Dynamics Simulations of Biological Macromolecules: Basics and Tips for an Effective Use. ChemPhysChem. n/a(n/a):e202200491.

43. Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016;374(2065):20150202.

44. Gowers RJ, Linke M, Barnoud J, Reddy TJE, Melo MN, Seyler SL, et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference. 2016;98–105.

45. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. Journal of computational chemistry. 2011;32(10):2319–27.

46. Mdciao: Accessible Analysis and Visualization of Molecular Dynamics Simulation Data | bioRxiv. https://www.biorxiv.org/content/10.1101/2022.07.15.500163v1;

47. PyLipID: A Python Package for Analysis of Protein from Molecular Dynamics Simulations | Journal of Chemical Theory and Computation. https://pubs.acs.org/doi/10.1021/acs.jctc.1c00708;

48. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics. 1996;14(1):33–8, 27–8.

49. Pardo L, Deupi X, Dölker N, López-Rodríguez ML, Campillo M. The Role of Internal Water Molecules in the Structure and Function of the Rhodopsin Family of G Protein-Coupled Receptors. ChemBioChem. 2007;8(1):19–24.

50. Katritch V, Cherezov V, Stevens RC. Structure-Function of the G Protein. Annual Review of Pharmacology and Toxicology. 2013;53(1):531–56.

51. Jia X-C, Oikawa M, Bo M, Tanaka T, Ny T, Boime I, et al. Expression of Human Luteinizing Hormone (LH) Receptor: Interaction with LH and Chorionic Gonadotropin from Human but not Equine, Rat, and Ovine Species. Molecular Endocrinology. 1991;5(6):759–68.

52. Dawaliby R, Trubbia C, Delporte C, Masureel M, Van Antwerpen P, Kobilka BK, et al. ALLOSTERIC REGULATION OF GPCR ACTIVITY BY PHOSPHOLIPIDS. Nature chemical biology. 2016;12(1):35–9.

53. Shao J, Tanner SW, Thompson N, Cheatham TE. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. Journal of Chemical Theory and Computation. 2007;3(6):2312–34.

54. Progress and challenges in the automated construction of Markov state models for full protein systems - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2766407/#c8;

55. UniProt: The universal protein knowledgebase in 2021 | Nucleic Acids Research | Oxford Academic. https://academic.oup.com/nar/article/49/D1/D480/6006196;

56. Tao A, Huang Y, Shinohara Y, Caylor ML, Pashikanti S, Xu D. ezCADD: A Rapid 2D/3D Visualization-Enabled Web Modeling Environment for Democratizing Computer-Aided Drug Design. J Chem Inf Model. 2019 Jan 28;59(1):18–24.

7 Appendices

Appendix A

Binding site definitions

Table A1: *ORG43553 initial binding site interactions (0 ns).*

Residue	type	Distance (Å)
T450	hydrophobic	3,83
V451	hydrophobic	4,16
S454	hydrophobic	3,89
L519	hydrophobic	4,13
M521	hydrophobic	3,86
V523	hydrophobic	4,10
I532	hydrophobic	4,43
I535	hydrophobic	3,93
F592	hydrophobic	3,87
A593	hydrophobic	3,66
A596	hydrophobic	3,29
A597	hydrophobic	4,14
K599	hydrophobic	3,85
V600	hydrophobic	4,35
P601	hydrophobic	4,32
I603	hydrophobic	3,90
T604	hydrophobic	3,66
V605	hydrophobic	3,88
S608	hydrophobic	3,87
Y616	hydrophobic	4,46

Table A2: *ORG43553 Binding site interactions after 350 ns of simulation.*

Residue	type	Distance (Å)
P520	h-bond	3,05

Residue	type	Distance (Å)
V451	hydrophobic	4,37
S454	hydrophobic	4,03
V515	hydrophobic	3,73
L519	hydrophobic	3,77
M521	hydrophobic	3,55
V523	hydrophobic	3,96
I532	hydrophobic	4,40
L533	hydrophobic	3,95
I535	hydrophobic	3,81
L536	hydrophobic	4,12
I589	hydrophobic	4,09
F592	hydrophobic	3,37
S595	hydrophobic	4,27
A596	hydrophobic	4,10
I603	hydrophobic	3,90

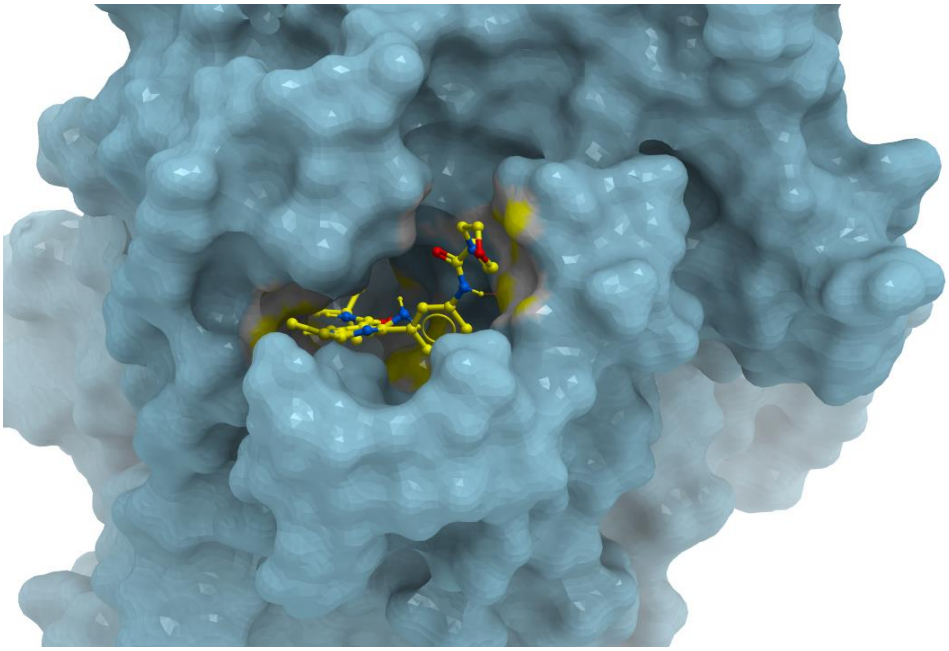
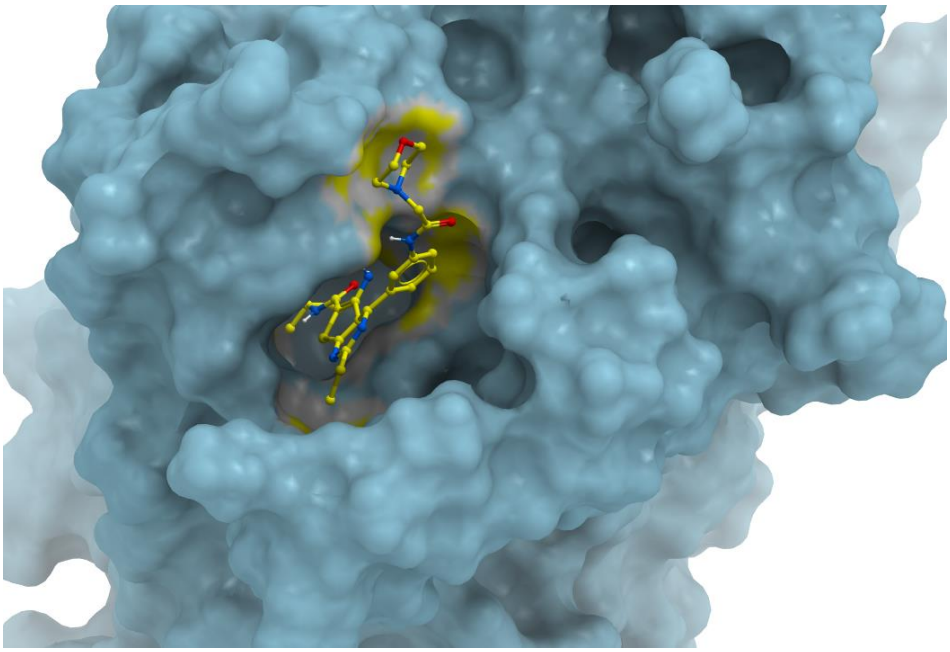


Figure A1: *Trajectory-averaged RMSD as a function of the fraction size using the structural alignment of the MDlovofit package*



Appendix B

Elements used in the analysis

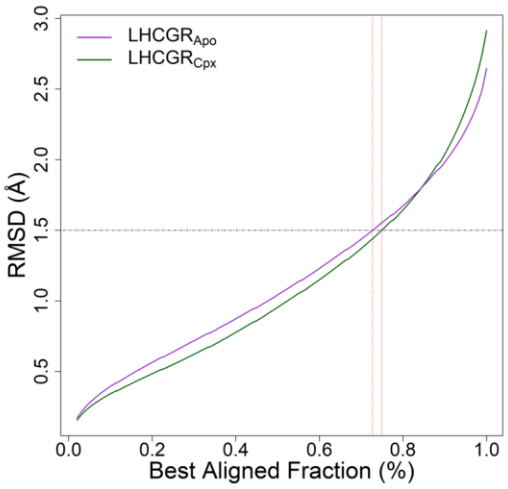


Figure B1: *Trajectory-averaged RMSD as a function of the fraction size using the structural alignment of the MDlovofit package*

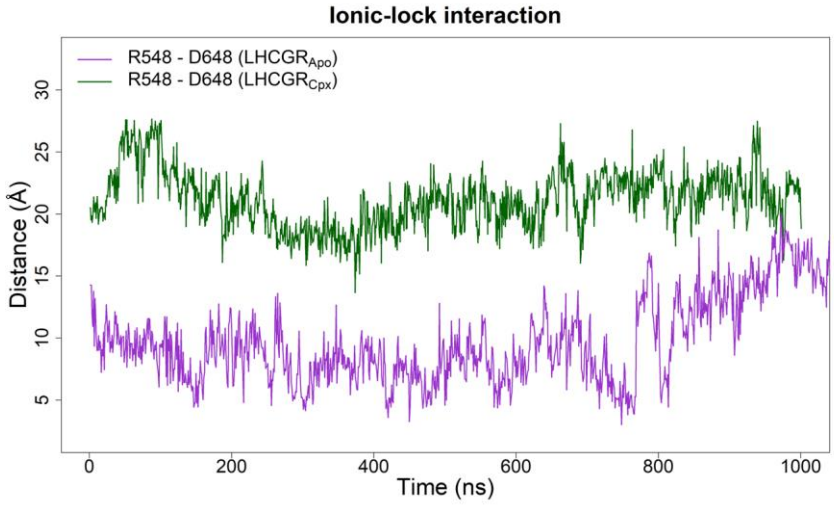


Figure B2: *Plot of the distance along the simulation time separating R548 and D648 of the ionic lock motif*

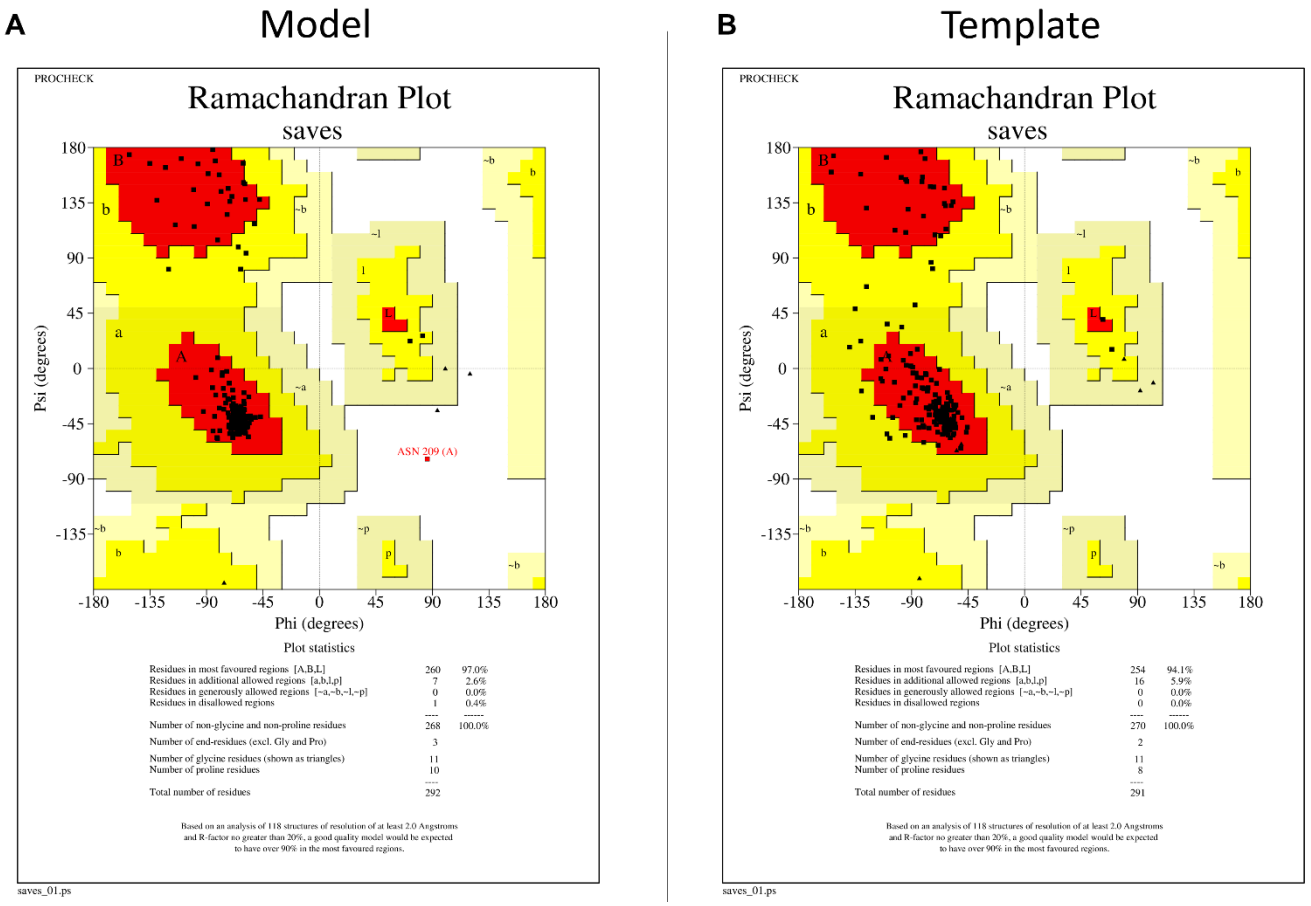


Figure B3: *Ramachandran plots generated by the PROCHECK program. The output shows that the Model have 97 % of its residue's dihedral angles in the most favoured regions. It is higher than 90% and higher than the template. The Model was considered to be reasonable*

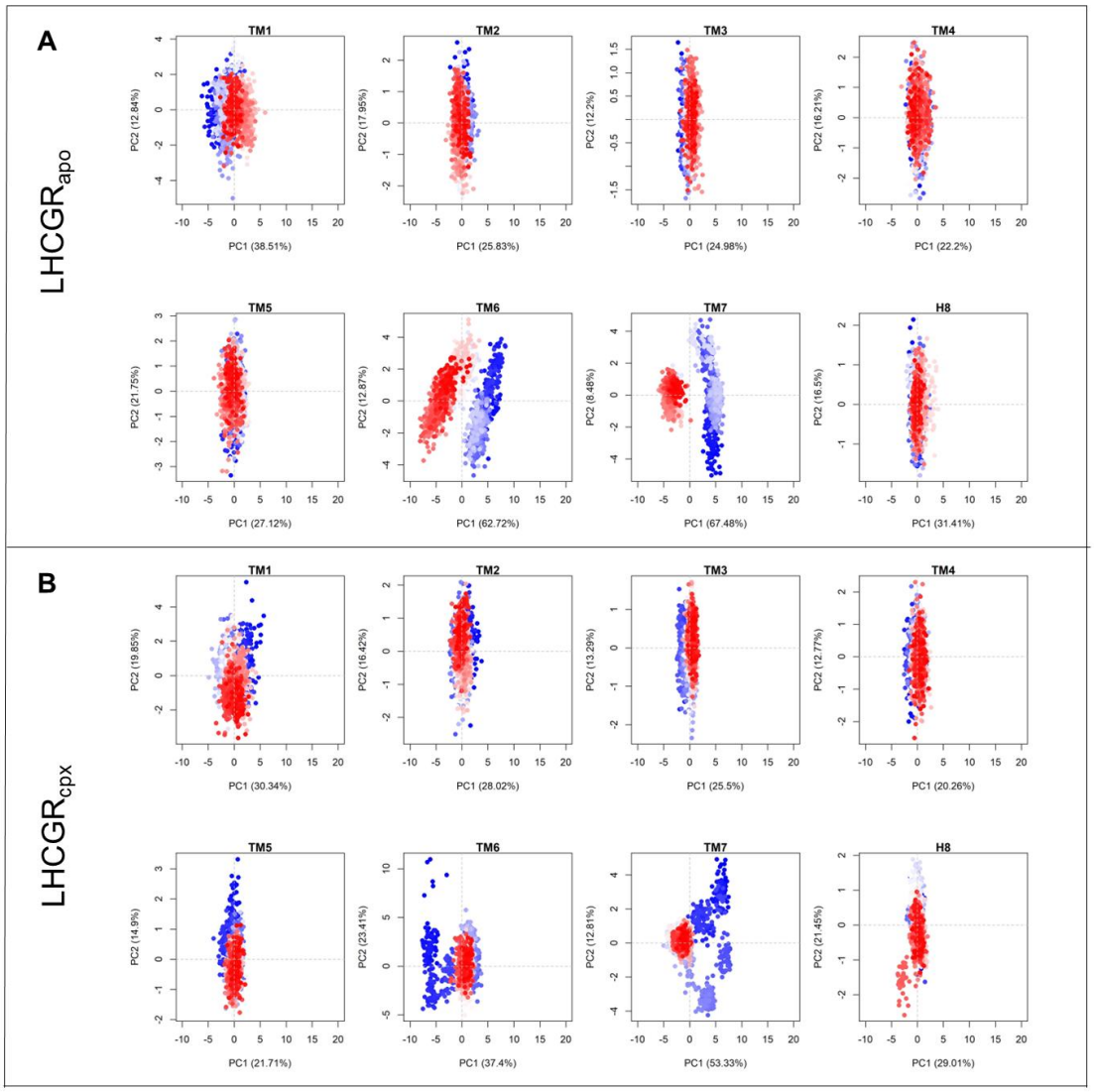


Figure B4: *PCA on each helix. A) LHCGR$_{Apo}$ TM. B) LHCGR$_{Cpx}$.*

To resolve the motion of the helix, a Principal Component Analysis was performed on each helix. It gives an interesting insight into helix mobility as we can see the time evolution of the structure (from blue to white, to red) along the first two PC axes. In both structures in the quadrant of TM6 and TM7, we see differences in the space occupied at the beginning and end of the simulation. This shows that TM6, TM7 particularly have interesting motions that seems to be clear change of conformational state. However, it should be noted that PC1 and PC2 represent only a small part of the variability of the data, so interpretation is limited.