This electronic thesis or dissertation has been downloaded from the King's Research Portal at https://kclpure.kcl.ac.uk/portal/

Modelling modulating phenomena for protein function in a serotonin-gated ion channel with molecular dynamics and enhanced sampling methods

Crnjar, Alessandro

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

Unless another licence is stated on the immediately following page this work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 13. Jan. 2025

King's College London

School of Natural and Mathematical Sciences - Department of Physics

Modelling Modulating Phenomena for Protein Function in a Serotonin-Gated Ion Channel with Molecular Dynamics and Enhanced Sampling Methods

Ph.D. in *Physics*

Alessandro Crnjar

Supervisor: *Professor Carla Molteni*

Second Supervisor: *Doctor Chris Lorenz*

Contents

List of Figures

List of Tables

Preliminary Statements

I declare that the work described in the following is my own, and it comprises the work published in the following papers:

- A. Crnjar, F. Comitani, W. Hester, and C. Molteni, "Trans-Cis Proline Switches in a Pentameric Ligand-Gated Ion Channel: How They Are Affected by and How They Affect the Biomolecular Environment.", *The Journal of Physical Chemistry Letters* 10 (3), 694-700, 2019.[7]
- A. Crnjar, and C. Molteni, "Cholesterol content in the membrane promotes key lipid-protein interactions in a pentameric serotonin-gated ion channel", *Biointerphases* 15 (6), 061018, 2020.[3]
- A. Crnjar, S. M. Mesoy, S. C. R. Lummis, and C. Molteni "A Single Mutation in The Outer Lipid-Facing Helix of a Pentameric Ligand-Gated Ion Channel Affects Channel Function Through a Radially-Propagating Mechanism", accepted in Frontiers in Molecular Biosciences in the early 2021.[8]

The work described in chapter 3 consists in a development of a project started by Dr. Federico Comitani, whose results are described in his PhD thesis.[9]

I am also first author of a review on the *in-silico* techniques at different levels of resolution used to complement experimental results of mutagenesis electrophysiology experiments in ligand-gated ion channels. These techniques include *ab-initio* methods, molecular dynamics and enhanced sampling techniques.[10]

During my PhD, I worked on two side projects not concerning ligand-gated ion channels. One focuses on the dynamics of ellagic acid, a polyphenol molecule, and its interaction with an exemplary double-stranded DNA, in different solvents, namely water and methanol. This has been part of a collaborative project with the overall goal of gaining access to the DNA preserved in the wood of forest trees, and evaluating its epigenetic information content with respect to the growth and therefore to climate conditions at the time of wood formation. This work will be submitted to PLOS ONE.[11] The other side projects regards the study of Foerster resonance energy transfer (FRET)-related phenomena in the green fluorescent protein (GFP). This work has been published in Biophysical Journal in 2020.[12]

Abstract

Pentameric ligand-gated ion channels (pLGICs) are important proteins, embedded in the lipid membranes of nerve cells, which are of great pharmacological relevance. They mediate fast synaptic transmission and are involved in many diseases. For this reason, their fundamental understanding at an atomistic level is paramount for the design of new, targeted pharmaceuticals. Phenomena of interest include ligand binding/unbinding, opening and closing of the inner ion channel, mechanical signal propagation, allosteric modulation by lipids of the cell membrane and/or by other ligands, and more. Depending on the time scales of the specific phenomenon of interest, a variety of computational (or *in-silico*) techniques can be employed for their study. One of the most common is molecular dynamics, which involves the solution of Newton's equations of motion for every particle that belongs to the system under investigation. This technique allows for the study of equilibrium and out-of-equilibrium phenomena, typically in the range of nanoseconds-to-microseconds, depending on the size of the system (between orders of 10^3 atoms to orders of 10^6 atoms in the case of bio-systems). However, rare events are hardly sampled by the usage of molecular dynamics, thus requiring the introduction of the so called enhanced sampling techniques. Among these are metadynamics, umbrella sampling, steered molecular dynamics, and many others.

The goal of this PhD project is the investigation of several phenomena revolving around one pLGIC, namely the serotonin-activated type 3 A ion channel, or 5-HT$_{3A}$ receptor. The importance of this receptor is proven by its involvement in diseases such as schizophrenia, drug abuse, and many others. In this system, the binding of serotonin triggers a cascade of conformational changes that ultimately culminates with the opening of a hydrophobic gate, leading to a flow of ions through the inner ionic channel. This cascade of events typically occurs over a time window of milliseconds: in this context, molecular dynamics is an ideal tool to study single phenomena that make up the

overall cascade. Its complete mechanism is, however, far from being fully understood. We focus here on studying phenomena that occur in different parts of this protein, and that, globally, contribute the overall mechanism of this molecular machine.

The first chapter of this thesis aims at providing all relevant information about pLGICs and relative ligands, with a particular focus on the 5-HT$_{3A}$ receptor, while the second chapter provides theoretical details on all-atom molecular dynamics and on the enhanced sampling methods used. The subsequent chapters describe the results obtained in each of the projects carried out during the PhD years, providing an introduction to the specific phenomena studied, their relevance, and a relative literature review.

Specifically, in the third chapter, we investigate the possibility that the propagation of mechanical signal across the extracellular and the transmembrane domains of the receptor is mediated by the isomerisation of a proline located at their interface. This consists in the rotation of a torsional angle linking this amino-acid and the subsequent residue. This study is carried out both with molecular dynamics and with metadynamics, with the aim of capturing structural differences that the different proline isomers induce in the surroundings, and of reconstructing the free energy associated with the isomerisation process. In this way, we are able to assess how the protein environment and the proline isomerisation influence each other.

In the fourth chapter, we focus our attention to what happens at the interface between the protein transmembrane domain and the lipid membrane. In recent years, the modulation by the lipids (phospholipids and cholesterol in particular) onto the working mechanism of several pLGICs has been proved both *in-silico* and experimentally. In this context, our goal is to characterise, via molecular dynamics simulations, lipid-protein interactions, i.e. the binding of specific lipids to the surface of the protein, to assess the effects that lipids have on structural and dynamical properties of the protein, and to show how these are influenced by the exact concentrations of the lipid species.

In the fifth chapter, we investigate the role of the outermost M4 α helices of the transmembrane domain on the protein function. According to experimental results, disrupting crucial interactions in this section of the protein may alter the overall protein function. Here, we perform molecular dynamics simulations to assess how a specific mutation on the M4 helix, i.e. Y441A, is able to alter structural and dynamical properties of the M4 itself and of its surroundings, with the aim of shedding light on how the M4 is relevant for the mechanism of the serotonin-activated receptor.

In the sixth chapter, we present a preliminary study of the unbinding of serotonin from the orthosteric pocket, located in the extracellular domain of the receptor. We perform this study by means of both molecular dynamics and metadynamics simulations, with the aim of understanding the fundamental interactions involved in the process.

Overall, the original work carried out and explained in this thesis contributes to the fundamental understanding of the 5-HT$_{3A}$ receptor, and by extension of pLGICs in general.

Chapter 1

Introduction

1.1 The Brain, Neurons, and Neuroreceptors

The complex functioning of the human brain is set by 10^{12} nerve cells, which are able to form 10^3 to 10^4 connections each, providing for a large variety of functions. These span across heart rate, sexual function, emotions, learning, memory, and many others. The brain is also thought to influence the response of the immune system to diseases and treatments.[1]

The building block of this complex network is the *neuron*, a highly specialised cell devised to transmit information to other nerve cells, muscle, or gland cells. It is made up of several parts: a *cell body*, *dendrites*, and an *axon*.

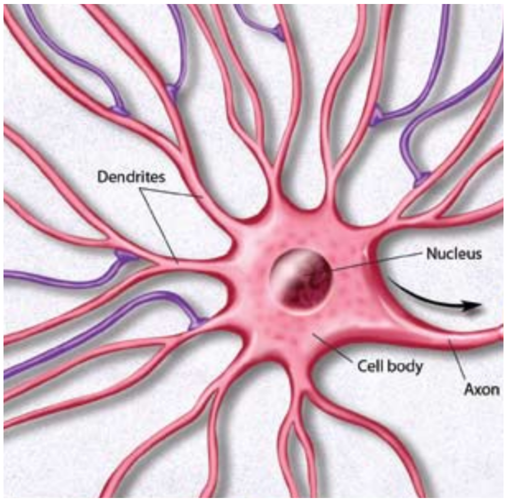


Figure 1.1: Representation of a neuron, including the cell body, the nucleus, the axon, and dendrites. Adapted from reference [1].

The *cell body* comprises the nucleus and the cytoplasm. The *axon*, which is electrically excitable, extends from the cell body, and in most cases into separate smaller branches, and finally ends at the nerve terminals. The *dendrites* also extend from the cell body, and receive messages from the other neurons. The *synapses* are the contacts that allow communication among neurons: about 10^{12} to 10^{14} of these are found in the human brain, making up both a complex and an adaptive network. Both dendrites and the cell body are covered by the synapses formed by the ends of the axons of other neurons.

A neuron is able to give rise to an electrical impulse depending on the difference in charge between the inside and the outside of the cell. At the beginning of a nerve impulse, a significant reversal in the electrical potential takes place at one point on the cells membrane, while the neuron switches from an internal negative charge to a positive one. Such change, referred to as an *action potential*, is then transmitted along the membrane of the axon. These impulses are fired by neurons several times per second. Once at the end of an axon, the voltage change induces the release of small molecules referred to as *neurotransmitters*.

The communication among neurons is provided by proteins embedded in their lipid membranes, called *neuroreceptors*, whose activation can occur either upon the binding, from the exterior of the cell, of neurotransmitters (also referred to as *ligands*), or due to a gradient in the electrical potential on the two sides of the membrane.

The receptors which are activated by ligand binding are divided into *G-protein coupled receptors* (GPCRs) and ligand-gated ion channels (LGICs), which differ in both structure and functioning. G-protein coupled receptors are *metabotropic*, i.e. the message they carry is transmitted by means of a chain of events which involve other compounds as well. Conversely, LGICs are *ionotropic*: their structure revolves around a central ion permeable channel able to alternate open and closed conformation through the so called *gating* process. LGICs are the focus of the work presented in this thesis.

1.2 Pentameric Ligand-gated Ion Channels and the Cys-Loop Superfamily

Pentameric ligand-gated ion channels (pLGICs) are proteins embedded in the lipid membrane of nerve cells, that are crucial for neuronal communication and fast synaptic transmission. Examples of such ion channels are nicotinic acetylcholine receptors (nAChRs), serotonin type 3 receptors (5-HT$_{3A}$Rs), γ-amminobutyric acid type A receptors (GABA$_A$Rs), glycine receptors (GlyRs), glutamate-gated chloride (GluCl) receptors, and zinc-activated ion channels.[13] Notable prokaryotic examples are the *Gloeobacter violaceus* (GLIC) and *Erwinia chrysanthemi* (ELIC) ligand-gated ion channels.

PLGICs are ubiquitous in the nervous system, and participate in a number of functions, such as sensory and motor processing, central autonomous control, heart rate, blood pressure, lung function, movement, regulatory process, memory, attention, sleep and wakefulness, reward, pain, anxiety, emotions, and cognition. Mutations of these proteins can thus alter their function and ultimately lead to certain pathologies, like congenital myasthenic syndrome, nocturnal front lobe epilepsy, hyperekplexia, autism, schizophrenia, Alzheimer's disease and many others. Because of this, pLGICs are the target of therapeutic drugs (but also of addictive ones): in particular, nAChRs are responsible for tobacco and nicotine addiction, GABA$_A$Rs mediate part of ethanol effects and are the target of the benzodiazepine class of drugs, and 5-HT$_{3A}$R antagonists are used for the treatment of nausea (e.g. in patients undergoing chemotherapy).

1.2.1 Structure of Pentameric Ligand-Gated Ion Channels

PLGICs are made up of five subunits (thus the name *pentameric*) arranged around the central ion channel, and span a cylinder-like volume of roughly 12 nm in height and 7 nm in diameter. As shown in figure 1.2, which depicts a X-ray structure of the 5-HT$_{3A}$ receptor, they are organised in three domains, from N-terminus to C-terminus: the *extracellular domain* (ECD), which is mainly made up of β sheets and that carries the orthosteric binding site for the neurotransmitter, the *transmembrane domain* (TMD), which is typically made up of four α helices (labelled M1, M2, M3, and M4) per subunit, and, often, an *intracellular domain* (ICD).

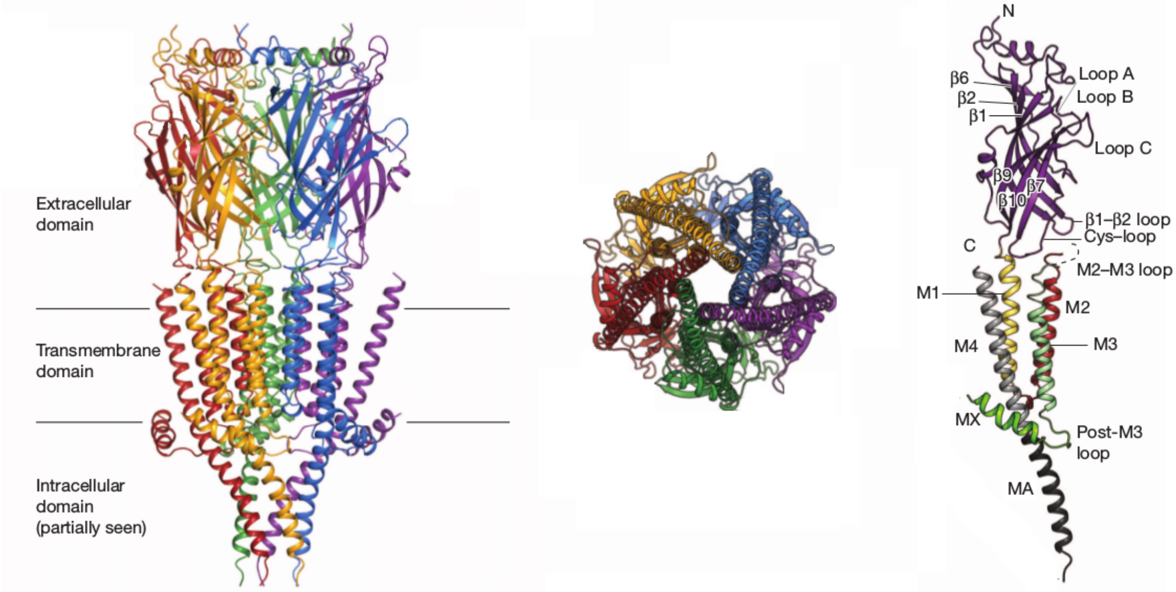


Figure 1.2: The 5-HT_{3A} receptor divided in its three domains (left), and seen from the top (centre); one single subunit out of five (right). Adapted from reference [2].

The ECD and the TMD form an interface through several well-conserved loops: the $\beta1$-$\beta2$ loop, the $\beta6$-$\beta7$ loop, the $\beta8$-$\beta9$ loop (also referred to as loop F), and, at the interface itself, the M2-M3 loop. The $\beta6$-$\beta7$ loop, also referred to as Cys-loop, bears an ultra-conserved disulphide-bridge between two cysteines 13 amino-acids apart and gives the name to the eukaryotic pLGIC *Cys-loop receptor superfamily.*

The TMD is embedded in the lipid membrane of the cell, with its innermost M2 α-helices lining the ion channel. Here, the hydrophobic side chains of certain highly conserved hydrophobic amino-acids (specifically leucines or isoleucines) are thought to constitute a barrier for ion translocation, and thus are the key feature of a hydrophobic gating mechanism.[14, 15]

1.2.2 Function of Pentameric Ligand-Gated Ion Channels

The function of pLGICs is to control the transmembrane potential via the flow of ions such as K^+, Na^+ and Cl^- through their domains. Ion channels, in fact, select specific ions to flow towards the interior of the cell, i.e. they contain channel-facing girdles, called *selectivity filters*, of charged residues, arranged one per subunit, which alter the electrical potential alongside the channel axis and promote in this way the passage of ions through it. According to the type of ions they select, pLGICs can be divided in two classes: cationic channels (which include nACh receptors and 5-HT_{3A} receptors) and anionic channels (which include glycine and GABA receptors). Thanks to the mechanism operated by pLGICs, complex behaviours emerge in terms of body response, including muscle contraction, or the production of enzymes.

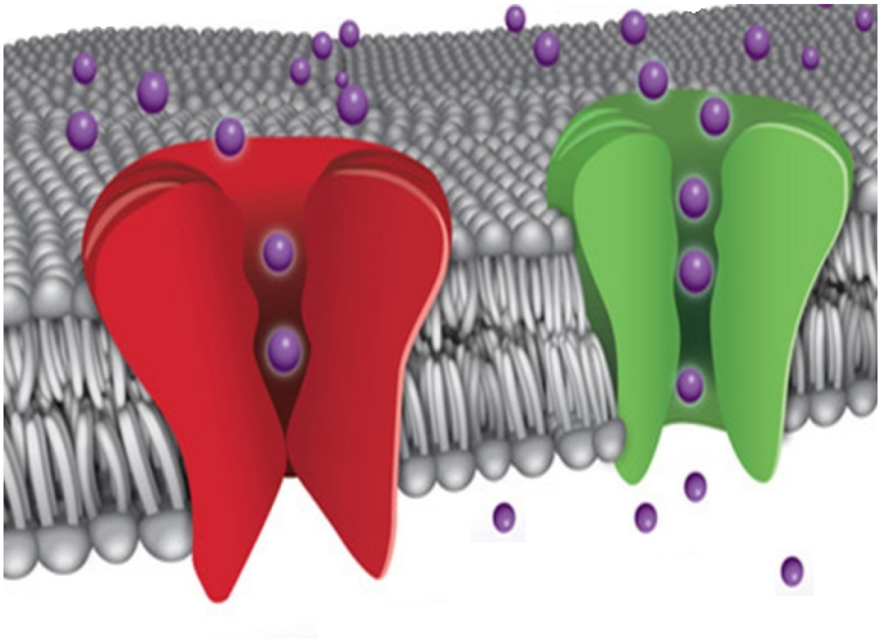


Figure 1.3: Representation of two ion channels of different conductive states, embedded in a lipid bilayer (grey): the open channel (green) lets the ions (purple) through, while the closed channel (red) does not. Figure adapted from the website www.cambridgenetwork.co.uk.

The flow of ions is controlled by the opening and closing of the transmembrane ion permeable pore, a conformational rearrangement referred to as the *gating mechanism* (represented in figure 1.3). This is triggered by the binding of a a *ligand* (the *neurotransmitter*, or *agonist*) in a specific cavity in the extracellular domain referred to as the *orthosteric* binding pocket, located at the interface between adjacent subunits. This binding, in turn, stabilises an open channel conformation by inducing in the receptor a series of subsequent conformational changes.

After the ligand binds into the orthosteric pocket, it interacts with the surrounding amino-acids via hydrogen bonds and π interactions (see Appendix A).[16] These interactions provoke a cascade of events that must propagate from the ECD to the TMD, where the gate is located. This propagation takes place over tens of Angstroms of distance and over time windows typically in the range of ms.

It is not yet clear how many agonists (one per pocket) are necessary to fully induce a gating mechanism in a pLGIC; however experimental evidence suggests that the maximal mean channel open time is observed when three binding sites at non-consecutive subunit interfaces are bound, while intermediate responses are exhibited with three consecutive or two non-consecutive interfaces bound.[17]

Artificial ligands can be designed to compete or substitute the natural ones in order to induce certain response into the receptors: this makes these receptors the target of drug design studies.

Apart from the agonists, other ligands, called *antagonists* (or *inhibitors*), may exhibit a high affinity for the orthosteric binding site of one of these receptors, although would not trigger any conformational change cascade in the protein, while preventing an agonist to bind in the same pocket. An *allosteric* modulation can also take place in the receptor, i.e. the events cascade may also be helped (or disrupted) by binding events on alternative sites. In case of a *positive modulation*, the ligand takes the name of *potentiator*. Thus, the binding-triggered gating induces the cell hyperpolarisation (i.e. a negative change in the transmembrane potential), that decreases the chance for an action potential to occur, or conversely depolarisation.

To account for the agonist-triggered activation, two models have been proposed. The *induced-fit theory* postulates that ligand binding is a prerequisite for the activation and drives conformational changes, while the *Monod-Wyman-Changeux (MWC) theory*[18] postulates that pLGICs spontaneously vary between different states (such as active, resting, and desensitised), and that ligands shift the equilibrium towards the state for which they have the highest binding affinity. In the active (or open) state, the receptor lets ions through, in the resting (or closed) state, it does not not; while the desensitised state constitute a similarly-closed conformation but is characterised by a higher affinity for the ligand.

The working mechanism of pLGIC at the atomistic level is still poorly understood. The role of specific amino-acids of their sequence can be successfully studied experimentally by means of mutagenesis and electrophysiology experiments. These experiments aim at understanding the role of specific amino-acids in the working mechanism of pLGICs, by mutating them in order to alter their properties: typical cases include the removal of aromatic or charged side chains with non-polar ones (for example alanine) with the aim of inhibiting the capability of forming interactions such as hydrogen bonds or π interactions. The response of the mutated receptors is measured via the application of a current in presence of different ligand concentrations. A useful quantity typically measured is the EC_{50}, the ligand concentration required to have half of the maximum current. While these recordings allow for the determination of what residues may be good candidates for important interactions for receptor function, *in-silico* studies are still required to validate their results. This can be done at different level of resolution, that may vary from the application of quantum mechanical methods, to that of classical ones.[10]

1.2.3 Crystallography, Cryo-electron Microscopy, and Homology Modelling

Atomistic structures of pLGICs can be experimentally obtained with different techniques. One of the main ones is *X-ray crystallography*, which is based on Braggs Law of X-ray diffraction by crystals, well-ordered packing of homogenous molecules in three dimensions. Illuminated by a beam of X-ray light, the crystal diffracts the light at various angles, generating a diffraction pattern, which reflects the structural arrangement of the atoms within the crystal. The quality of the produced structure heavily depends on the degree of order of the crystal: a large amount of highly purified macromolecules may thus be necessary, as well as the engineering of the molecules (including stability promotion, side-chain modification, and proteolysis). Until now, most of the atomic resolution macromolecular structures have been solved by X-ray crystallography. The other main technique is *cryo-electron microscopy* (cryo-EM), which can be used to examine non-crystalline structure: this technique uses high energy electrons as a source to illuminate very thin specimens in a transmission electron microscope. From hundreds of thousands of single particle cryo-EM images of one type of macromolecule, statistical analysis can be performed to classify, align, and average the images to reconstruct the common features into a three-dimensional virtual structure. While X-ray crystallography needs large amount of materials to optimise the crystallisation condition, cryo-EM needs much smaller macromolecule samples, and typically does not need rigorous molecular engineering. Moreover, the cryo-EM specimen is made by fast freezing biological samples in liquid nitrogen temperature directly from the solution, thus maintaining the macromolecules in their soluble states. Therefore, cryo-EM has the advantage to produce structures in more close-to-native state with respect to X-ray crystallography, but also the disadvantage of having to deal with intrinsic structural heterogeneity due to thermal fluctuations.[19]

Another way to build the structure of a biological macromolecule is to make use of *homology* *modelling.* This technique relies on the fact that proteins with similar evolutionary paths are characterised by similar folding patterns even when they have poor amino-acid sequence similarity. This allows for the reconstruction of the structure of a protein whose sequence is known even if its spatial disposition is not.

However, in the past years a number of experimentally solved structures of pLGICs have been made available, especially thanks to the diffusion of cryo-EM.

Structures are now available for the nicotinic acetylcholine receptor,[20] for the γ-aminobutiric receptor,[21, 22, 23] the glycine receptor,[24, 25, 26, 27] for the GluCl receptor,[28, 29] for the *Erwinia chrisanthemi* (ELIC) ion-channel,[30, 31, 32] and for the *Gloeobacter violaceus* ion-channel (GLIC).[33, 34, 35, 36, 37]

Many structures were made available for the 5-HT$_{3A}$ receptor as well. Since they are of particular relevance for the topics presented in this thesis, they are discussed in more details later on.

1.3 Serotonin and Serotonin Receptors

The goal of this Ph.D. project is to contribute to the currently limited understanding of the activation mechanisms in pLGICs by means of state-of-the-art and innovative computer simulation methods, focusing on a representative ion channel: the *5-HT$_{3A}$ receptor.*

5-HT (or serotonin) *receptors* are activated by the binding of 5-hydroxytryptamine (5-HT) or *serotonin.* These receptors, present in most animals, modulate the release of excitatory and inhibitory neurotrasmitters, like GABA, dopamine, acetylcholine, and glutamate. They are involved in the regulation of several processes, including learning, anxiety, emotional and cognitive control, memory, sleep, appetite, and sexual behaviour.

While most 5-HT receptors are G-protein coupled receptors, the 5-HT$_{3A}$ receptor is a cationic pentameric ligand-gated ion channel of the Cys-loop superfamily. This receptor may be composed of five identical 5-HT$_{3A}$ subunits (*homopentameric*) or a mixture of 5-HT$_{3A}$ and one of the other four 5-HT$_{3B}$, 5-HT$_{3C}$, 5-HT$_{3D}$, or 5-HT$_{3E}$ subunits (*heteropentameric*).[38] It is involved in neurological disorders like schizophrenia, drug abuse, and is the target of nausea-suppressant drugs for patients undergoing chemotherapy or subject of anesthetics. Its activation typically occurs with a frequency of 400 s^{-1},[39] i.e. one activation would occur on the ms time-scale.

Serotonin, represented in figure 1.4, is the 5-hydroxy derivative of tryptamine. It is a monoamine molecular messenger, a primary amino compound, a member of phenols, and a member of hydroxyindoles and of tryptamines.

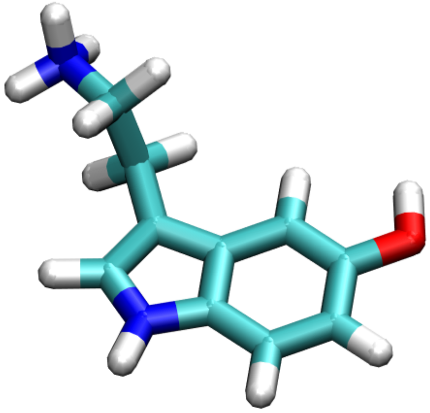


Figure 1.4: A serotonin molecule, in its positively-charged state.

This compound is ubiquitous in the nervous system, found in most animals, plants, and fungi. It is involved in a variety of diseases, like migraine, mood disorders, and others. Outside of the nervous system, it participates in a variety of biological functions including digestion, bone metabolism and organs growth.

Serotonin is a very polar molecule: it features two aromatic rings, one of which contains a hydroxyl group, and the other contains an aliphatic tail terminating with an NH_2 group (NH_3 in the protonated form). The hydroxyl group can easily react to form other compounds, or form intermolecular hydrogen bonds. The rings have a great electronegative potential due to the presence of a pair of lone electrons on the nitrogen of the ring; these electrons are located in a sp2 orbital, perpendicular to the π-bonds of the aromatic system, and for this reason are also available to form bonds. This compound is therefore able to give rise to a variety of interactions with its surroundings (e.g. the binding pocket of a serotonin receptor), including hydrogen bonds, cation-π and π-π interactions (see Appendix A). These interactions help serotonin find its way inside the binding pocket, but also stabilise it once inside, in turn giving rise to a chain of conformational changes within the receptor.

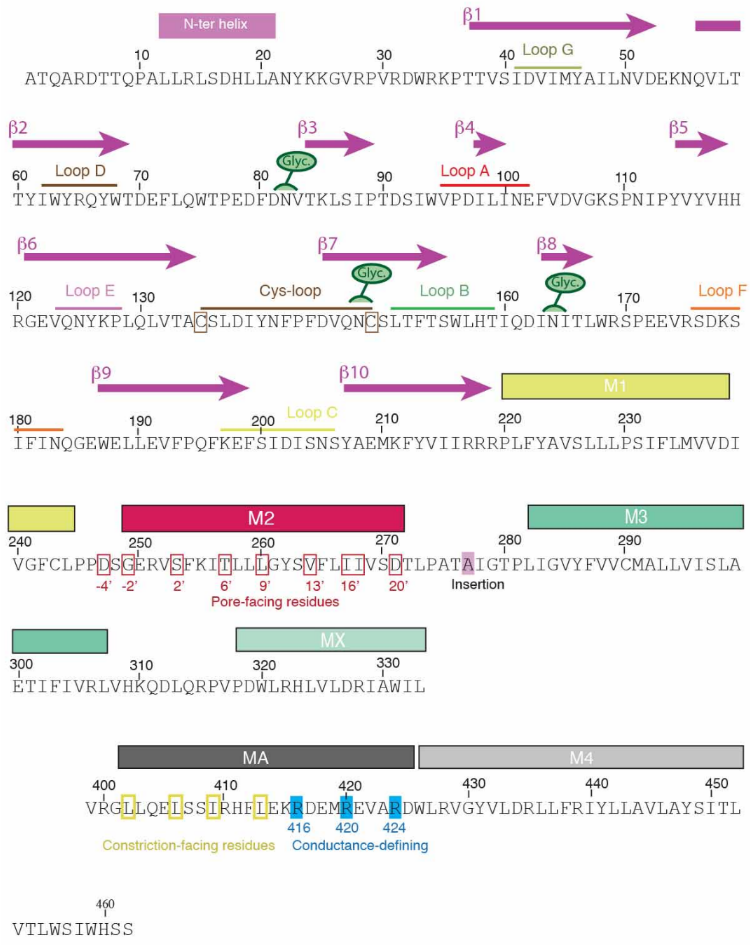


Figure 1.5: Sequence of the 5-HT$_{3A}$ receptor, as reported by Hassaine et al. for the X-ray crystallography structure in 2014. Adapted from reference [2].

While serotonin acts as the main agonist for the 5-HT$_{3A}$ receptor, several competing agonists and/or antagonists (e.g. varenicline, setrons, etc.) for this receptor can be either identified or engineered.

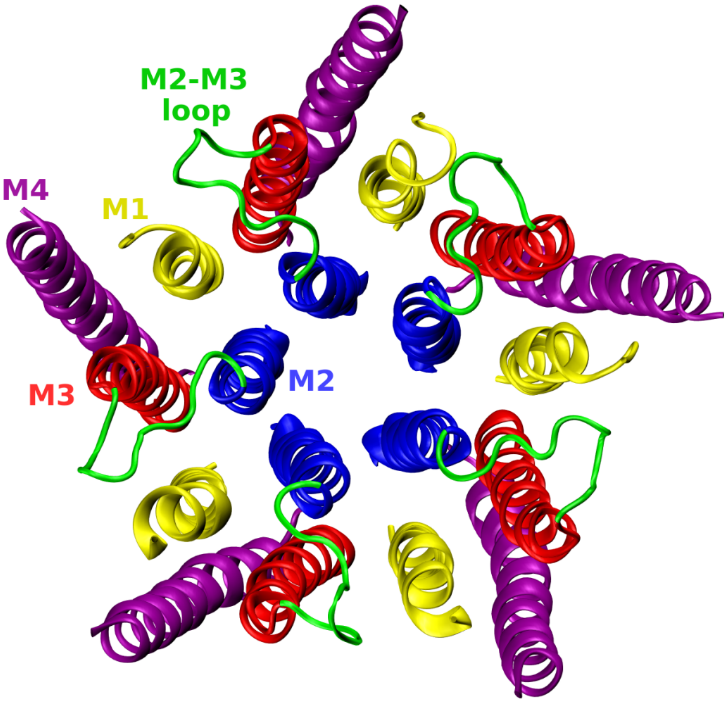


Figure 1.6: View of the TMD of the 5-HT$_{3A}$ receptor (PDB entry: 4PIR[2]) orthogonal to the channel axis, showing the four helices M1 (yellow), M2 (blue), M3 (red) and M4 (purple), and the M2-M3 loop (green) (figure from reference [3]).

The sequence of the mouse 5-HT$_{3A}$ receptor is reported in figure 1.5; its secondary structure is arranged in a similar way to other pLGICs of the Cys-loop superfamily, including the M1 to M4 α helices in the TMD (figure 1.6), helices MX and MA in the ICD, and a series of beta strands (labelled β1 to β10) within the ECD.

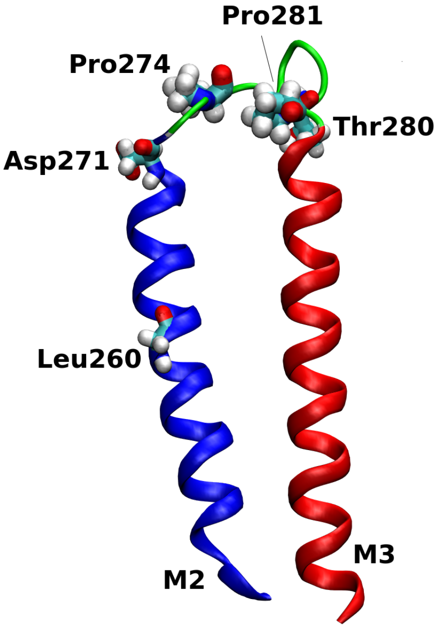


Figure 1.7: The M2 and M3 helices (in blue and red respectively), connected by the M2-M3 loop (in green), with relevant residues (PDB entry: 4PIR[2]).

Several notable amino-acids can be identified within the sequence, which are of relevance for the results presented in chapter 3. In the section comprising the M2, the M3 and the M2-M3 loop (figure 1.7), Asp271, located at the top of the M2 helices, may act as a selectivity filter for Na^+ ions thanks to its side chain, whose negative charge is due to the two terminal oxygen atoms. Its function was suggested by a study on the single-ion potential of mean force (for both Na^+ and Cl^-) as a function of the channel axis.[40] Furthermore, the M2-M3 loop contains two prolines, namely Pro281 and Pro274. Proline is a peculiar amino-acid due to its unique aromatic side chain, and, among other features, its relatively higher population at room temperature, with respect to other amino-acids, of the *cis* isomer of the *trans*-*cis* isomerisation process.[41] Morevoer, Pro274 is also a highly conserved amino-acid within the Cys-loop superfamily, while Pro281 is only found in 5-HT$_{3A}$ receptors and in nACh receptors. Finally, Leu260, found approximately at middle height of the M2 helix, makes up the hydrophobic gate of the 5-HT$_{3A}$ receptor.

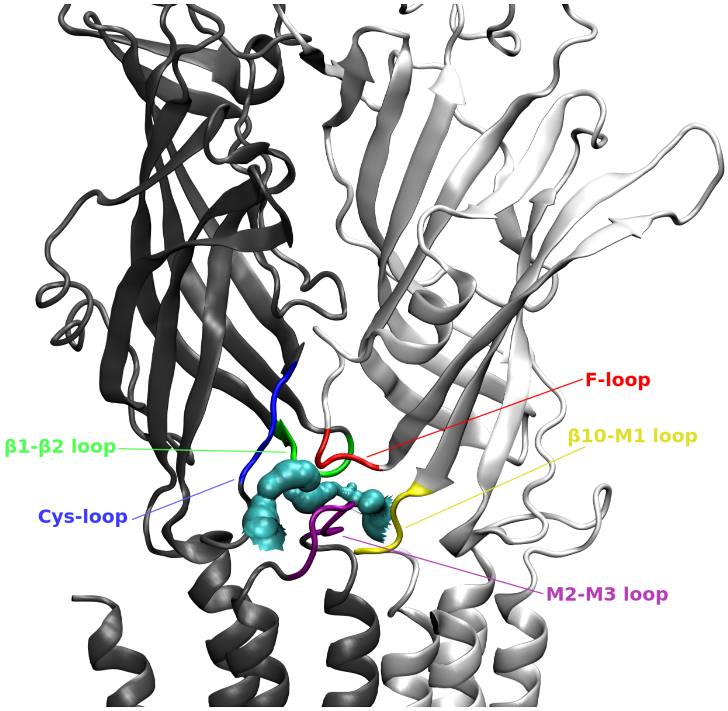


Figure 1.8: Loops surrounding the lateral ion portals between two adjacent subunits (subunit 1, in white, and subunit 2, in grey): the β1-β2 loop (green), the loop F (red), the β10-M1 loop (yellow), the Cys-loop (blue) and the M2-M3 loop (purple) (PDB entry: 4PIR[2]). The empty volume among these loops, that constitutes the portal, was calculated with the software HOLE2,[4] and is shown in cyan.

At the interface between TMD and ECD, ions can enter the receptor through lateral portals across adjacent subunits (represented in figure 1.8), as pointed out by an *in-silico* study.[40] Several residues will make up these portals: Asp52, Glu53, Lys54 and Asn55 in the β1-β2 loop, Gln184, Gly185 and Glu186 in the loop F, Arg219, Pro220, Leu221 and Phe222 in the β10-M1 loop, Cys135, Ser136, Leu137 and Asp138 in the Cys-Loop and from Pro274 to Thr280 in the M2-M3 loop.

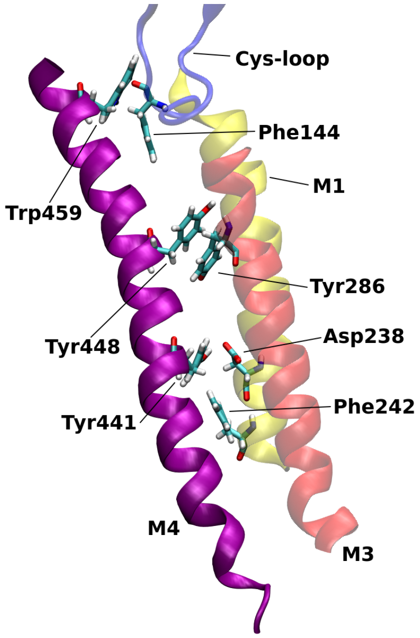


Figure 1.9: Important amino-acids for the pinning of M4 to M1, to M3, and to the Cys-loop (PDB entry: 6DG8[5]).

On the receptor outermost helices M4, important amino-acids allows for the pinning of such helices to the M1s and M3s (figure 1.9). Among these, Tyr441 forms π interactions and/or hydrogen bonds with Asp238 and with Phe242 (both belonging to M1); Tyr448 forms π-π interactions with Tyr286 (also belonging to M1). At the tip of the M4 helix, Trp459 can interact either with hydrogen bonds or via π-π interactions with amino-acids of the Cys-loop, such as Phe144.

Finally, in the ECD, at the interface between two adjacent subunits, the orthosteric binding pocket is found (figure 1.10). Here, the ligand is surrounded by several loops. In one subunit, we find Loop B (also called $\beta7$-$\beta8$), and Loop C ($\beta9$-$\beta10$). On the other subunit we find Loop G ($\beta1$), Loop D ($\beta2$), Loop E ($\beta6$), and Loop F ($\beta8$-$\beta9$). Of these, Loop C is of particular interest as it partially shields the ligand from the solvent.

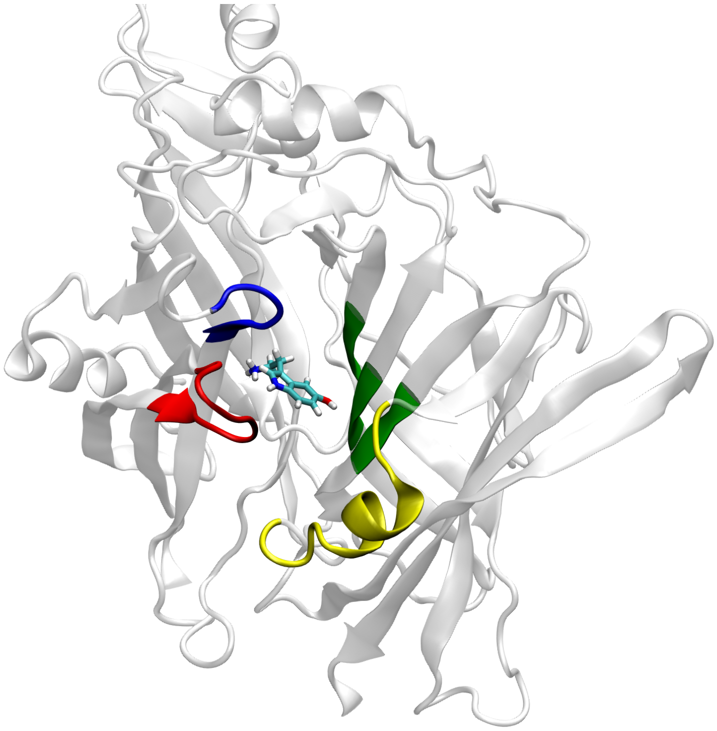


Figure 1.10: 5-$HT_{3A}R$ orthosteric binding pocket in between two adjacent subunits in the ECD, where a serotonin molecule is bound (PDB entry: 6DG8[5]). Important sections are represented in different colours: Loops G, D, E (green), Loop B (blue), Loop F, (yellow), and Loop C (red).

1.3.1 Available Structures of the 5-HT_{3A} Receptor

In recent years, the number of available structures of the homomeric 5-HT_{3A} receptor has significantly increased both in number and in variety, providing the unprecedented possibility of comparing different functional states of this receptor: in its apo-state or in complex with different agonists or antagonists, and in various conductive states including open, closed, desensitised, and other arguably classifiable intermediate ones. It is worth of mention that unambiguously assigning the correct functional state to a solved structure remains difficult. In fact, together with the observation of bound agonists/antagonists (or lack thereof), this would depend on the correct interpretation of distinctive conformational features.

The choice of a structure as a model for molecular dynamics simulations depends on a variety of factors, including the reliability of its experimental determination (given by the resolution at which it was originally solved), eventual refinement procedures, the eventual presence of unlikely side-chain rotamers or atomic clashes, the possible absence of certain parts of the amino-acid sequence, the presence/absence of ligands, and the deemed functional conformation.

In the following, the structures of the 5-HT$_{3A}$ receptor available as of 2018 (i.e. upfront to the last project started) are reported.

In 2014, Hassaine et al. reported the very first 5-HT$_{3A}$ receptor structure ever made available. This is an X-ray chrystallography apo-structure of the mouse serotonin receptor, and it was solved at 3.5 Å of resolution (PDB entry: 4PIR).[2] Due to its early release, this structure has been used as a model for various *in-silico* studies.[40, 42, 43, 7, 44, 3] This structure presents llama-derived antibodies (VHH) as crystallization chaperons; due to the presence of these antagonists, and because of the width at the hydrophobic gate in the TMD, this structure has been deemed to be in a non-conductive (inactive) state. At the same time, the overall backbone conformation appear to resemble those of open conformations of GLIC[33] or GluCl[28] receptors, and differs from those of closed conformations of GLIC (both closed[36] and locally-closed[35]) and ELIC[30]. Moreover, Hassaine et al. did not preclude the possibility that small fluctuations (in molecular simulations using this structure as starting model) may lead to the opening of the hydrophobic gate. We used this structure for our own works, namely those described in chapters 3 and 4.

Kudryashev et al. produced, in 2016, the structure of the mouse 5-HT$_{3A}$ receptor in lipid vesicles without antibodies, by means of cryo-electron tomography and sub-tomogram averaging.[45] In this case, the resolution is 12 Å, and the channel is in a open conformation.

Polovinkin et al. solved, in 2018, four cryo-EM structure of the mouse 5-HT$_{3A}$ receptor in different complexes, labelled as follows.[46] Complex “T” (PDB entry: 6HIS) is resolved at 4.5 Å , and comprises the receptor, in a closed conformation, in complex with the antagonist tropisetron, which is an anti-emetic drug. Complex “F” (PDB entry: 6HIN) is resolved at 4.1 Å , and includes the receptor, in an open conformation, and serotonin. Complex “I1” (PDB entry: 6HIO), resolved at 4.2 Å , comprises the receptor, in a closed (maybe desensitised) conformation, and serotonin. Finally, complex “I2” (PDB entry: 6HIQ), resolved at 3.2 Å , includes the receptor, in an intermediate conformation, in complex with serotonin and TMPPAA, a positive allosteric modulator.

Also in 2018, Basak et al. solved three cryo-EM structures of mouse 5-HT$_{3A}$ receptor. One is its apo-state, at 4.3 Å resolution (PDB entry: 6BE1).[47] With respect to the crystal structure, its ECD is twisted clockwise around the pore axis, in the direction towards a closed conformation.

Furthermore, Basak et al. resolved two distinct serotonin-bound conformations of the mouse receptor (PDB entries: 6DG7 and 6DG8, referred to as state 1 and state 2), respectively at 3.32 and 3.89 Å resolution. Compared to the 6BE1 apo-structure, these two states are characterised by a global twisting of the ECD and of the TMD, respectively in the counter-clockwise and in the clockwise direction. This leads to the repositioning of interfacial loops, while the TMD helices results expanded away from the central channel. Loop C, that closes the orthosteric binding pocket, is closed in both states 1 and 2, while it exhibits an outward or "open" orientation in the apo-structure. As for the conductive state of the two structures, state 2 was deemed conductive (i.e. open), while state 1 was described as non-conductive, although Basak et al. also suggested that side-chain fluctuations of the Leu260 side chains may give rise to a small conductance. State 1 was thus described as a possible pre-open state, or as a possible desensitised one. The binding poses of 5-HT in the two states are, however, similar. State 2 was also used in our works, specifically in those described in chapters 5 and 6.

PDB	**Authors**	**Year**	**Technique**	**Res. [Å]**	**Cond. state**	**Ligands and chaperones**
4PIR	Hassaine et al.	2014	X-ray chryst.	3.5	closed	Llama-derived antibodies
6HIS	Polovinkin et al.	2018	cryo-EM	4.5	closed	tropisetron
6HIN	Polovinkin et al.	2018	cryo-EM	4.1	open	5-HT
6HIO	Polovinkin et al.	2018	cryo-EM	4.2	closed/des.	5-HT
6HIQ	Polovinkin et al.	2018	cryo-EM	3.2	intermediate	5-HT and TMPPAA
6BE1	Basak et al.	2018	cryo-EM	4.3	closed	apo-structure
6DG7	Basak et al.	2018	cryo-EM	3.32	pre-open/des.	5-HT
6DG8	Basak et al.	2018	cryo-EM	3.89	open	5-HT

Table 1.1: The available 5-HT$_{3A}$ receptor structures, with relevant information. ("Res.": resolution, "Cond.": conductive, "chryst.": chrystallography, "des.": desensitised)

Figure 1.11 shows the width of the channel at the level of the TMD in all the 5-HT$_{3A}$ receptor structures, by representing the M2 helices of two non-consecutive subunits. These are a key section of the receptor for the classification of the conductive state, thus allowing for a visual comparison of the described PDB entries.

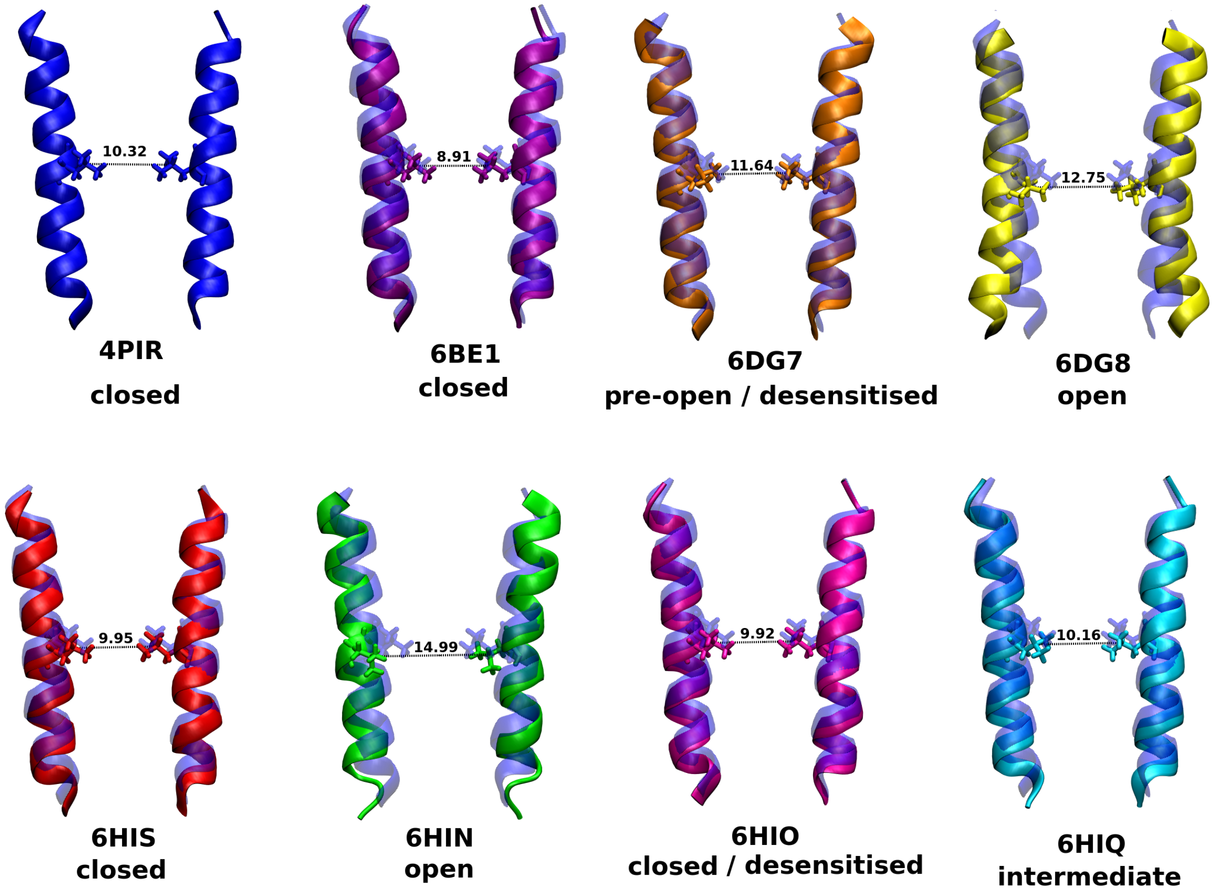


Figure 1.11: M2 helices of two non-consecutive subunits for representative X-ray and cryo-EM structures of the 5-HT$_{3A}$ receptor in different functional states, labelled according to their PDB entry. For comparison, the corresponding aligned helices of 4PIR are also shown in transparency for each structure, except for 4PIR itself. The hydrophobic gate (Leu260) is also displayed in the two subunits, together with the distance in Å of the two γ carbons.

Having here described the context of pLGICs, in the next chapter we will focus our attention to the *in-silico* methods of molecular dynamics and enhanced sampling techniques used in our project for the study of the 5-HT$_{3A}$R.

Chapter 2

Methods

2.1 Molecular Dynamics

The main computational technique used in the projects described in this thesis is *molecular dynamics* (MD), a powerful *in-silico* method widely used in modern simulations of bio-molecules as well as in soft matter, in condensed matter physics and in materials science in general. Molecular dynamics deals with a collection of N particles representing the system of interest, characterised by positions $\vec{r}_i$ and velocities $\vec{v}_i$:

$$\vec{r}_1, \vec{r}_2, ...\vec{r}_i, ...\vec{r}_N, \vec{v}_1, \vec{v}_2, ...\vec{v}_i, ...\vec{v}_N \;\;,\;\; i = 1, ..., N \tag{2.1.0.1}$$

and consists in the numerical integration of Newton's equations of motion for every particle:[6]

$$m_i \ddot{\vec{r}_i} = \vec{F}_i = -\nabla_i U(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N) \tag{2.1.0.2}$$

In equation 2.1.0.2, m_i is the mass of the i-th particle, $\vec{r}_i$ its position, $\ddot{\vec{r}_i}$ its acceleration, $\vec{F}_i$ the applied force on it, and U the inter-particle potential between the particles that make up the system.

Since computers are only able to implement processes on an iterative manner (i.e. by means of `for` loops over repeated code sections), the integration of Newton's equations of motion requires the discretisation of time, i.e. considering only equally-spaced instants t_n:

$$t_n = t_0 + n\Delta t \tag{2.1.0.3}$$

Here, t_n is the time after n iterations of the simulation, t_0 is the time at the beginning of the simulation, and Δt is the duration of one time step. This needs to be chosen as small as possible, as to minimise the error due to the discretisation. In biological systems, a typical choice is $\Delta t = 2 \cdot 10^{-15}$ s, which requires the constraint of all the molecular bonds containing hydrogen. This is because the vibrational frequency of such bonds cannot be sufficiently sampled with such a time step. Typical algorithms that allows for this constraint are *SHAKE* (used in these projects) and *RATTLE*.[48, 49]

Several algorithms for the integration have been developed, the most common of which, *Velocity Verlet* (used in our projects), is available in common simulation packages such as NAMD. The Velocity Verlet algorithm prescribes, for the update of positions and velocities, the following recipe:

$$\begin{cases} r_{n+1} = r_n + v_n\Delta t + \frac{1}{2}a_n\Delta t^2 \\ v_{n+1} = v_n + \frac{1}{2}(a_{n+1} + a_n)\Delta t \end{cases} \tag{2.1.0.4}$$

where r, v and a refer, respectively, to position, velocity and accelerations of a given particle for a certain time step n and its subsequent $n+1$. Velocity Verlet is applied through the following steps:

- an initial set of positions r_n and velocities v_n is given.
- by using this initial set, a_n (which is a function of particles positions) is computed.
- r_n is evolved into r_{n+1}.
- a_{n+1} is computed.
- v_n is evolved into v_{n+1}.
- the cycle is started again for the next value of n.

This algorithm well conserves the energy, and it is time-reversible (as required by the reversibility of classical equations of motion). Moreover, positions and velocities at the same time step are available with respect to Verlet. These characteristics makes it a reliable algorithm for molecular simulations.

2.1.1 Conserving the Temperature and the Pressure

The integration of Newtons equations of motion can be implemented with different algorithms which give rise to different *ensembles*, i.e. conserve different macroscopical quantities such as the total energy of the system, its temperature, its volume, or its pressure. Without the additional introduction of a *thermostat* (for temperature control) or of a *barostat* (for pressure control), Velocity Verlet (or another similar algorithm) gives rise to the *microcanonical ensemble* (or NVE, in which the number of particles N, the volume V of the system and the total energy E are constant). However, different integration schemes can be implemented for the discrete update of positions and velocities, according to the ensemble that we are interested in sampling. Typical cases use the *canonical ensemble* (or NVT, in which the number of particles N, the volume V of the system and the temperature T are constant), and the *isothermal-isobaric ensemble* (or NPT, in which the number of particles N, the pressure P and the temperature T are constant). The NPT ensemble is typically chosen for the simulation of biological systems in order to maintain the temperature at physiologically reasonable values and the pressure at the atmospheric value. The temperature is normally chosen to be either room temperature, around 300 K, at which many experiments are performed, or body temperature, around 310 K, at which physiological processes take place *in-vivo*. Systems containing lipid membranes require extra care, since lipid species are characterised by a *gel transition temperature*, that is the temperature required to induce a change in the lipid physical state from the ordered gel phase, where the hydrocarbon chains are fully extended and closely packed, to the disordered liquid crystalline phase, where the hydrocarbon chains are fluid and randomly oriented.

The temperature T can be easily computed at a given time t by accessing the particles velocities:

$$T(t) = \sum_{i}^{N} \frac{m_i v_i^2(t)}{k_B N_f} \tag{2.1.1.1}$$

where k_B is the Boltzmann constant, and N_f is the number of degrees of freedom ($N_f = 3N - 3$ for a system of N particles).

A simple method to maintain the temperature constant is the *Velocity Rescaling* algorithm, for which, after a certain number of steps, the modulus v of the velocity of every particle is rescaled according to the following equation:

$$v' = v\sqrt{\frac{\frac{3}{2}k_B T N}{E_k}} \tag{2.1.1.2}$$

where v' is the rescaled velocity, and E_k is the total kinetic energy. While working effectively, this algorithm does not provide a correct implementation of a canonical ensemble.

Another algorithm for the control of the temperature is the *Berendsen thermostat*,[50] which introduces a differential equation for the evolution of the kinetic energy:

$$\frac{dE_k}{dt} = -\frac{\left(-E_k - \frac{3}{2}k_B T N\right)}{\tau} \tag{2.1.1.3}$$

where τ is the relaxation time of the thermostat. This leads to the following rescaling of the velocity:

$$v = v \cdot \sqrt{\left(1 - \frac{\frac{3}{2}k_B T N}{E_k}\right)\exp(-\Delta t/\tau) + \frac{3}{2}\frac{k_B T N}{E_k}} \tag{2.1.1.4}$$

The Berendsen algorithm constitutes an improvement of the velocity rescaling, however it still does not sample the desired ensemble correctly. The *Andersen* thermostat tries to solve this issue by re-assigning the values of each velocity at each new time step according to:

$$v_i = \sqrt{\frac{k_B T}{m_i}}g \tag{2.1.1.5}$$

where g is a random number characterised by a Guassian distribution with mean equal to 0 and variance equal to 1. While this allows for a correct sampling, the trajectories result disrupted, since velocities are not rescaled, but rather re-assigned.

Combining the principles of rescaling the velocities and assigning their values from a given distribution leads to the introduction of *Langevin dynamics.* The *Langevin thermostat* adds a friction and a noise term to Newtons equations of motion. On a conceptual level, this thermostat can be thought as if the particles of the system were coupled with virtual particles of a bath, with which stochastic collisions may occur. We thus get *Langevin's equation*:

$$m_i \frac{d^2\vec{r_i}}{dt^2} = -m_i\gamma_i \frac{d\vec{r_i}}{dt} + \vec{F_i}(\vec{r}) + \tilde{\vec{r_i}} \tag{2.1.1.6}$$

where γ_i is the friction constant (with dimension of an inverse time), and $\tilde{\vec{r_i}}$ is a Gaussian noise process with zero mean and variance $\langle \tilde{\vec{r_i}}(t)\, \tilde{\vec{r_j}}(t+t') \rangle = 2m_i\gamma_i k_B T \delta(t')\delta_{ij}$. When $1/\gamma_i$ is large with respect to the time scales of the phenomena under investigation, molecular dynamics with stochastic temperature-coupling is being applied. The sampling performed with this dynamics will give rise to the desired ensemble. When $1/\gamma_i$ is small with respect to said time scales, the sampling is still correct, but the dynamics will differ from an NVE implementation.

We now turn to the conservation of pressure. Instead of conserving the total energy of the system, in an NPT ensemble the enthalpy $H = U + PV$ (where U is the potential energy) is the conserved quantity. This implies that, in order to implement such an ensemble, the volume of the system must be allowed to vary throughout time. In an atomistic simulation, the pressure associated with the system is computed by means of the *virial equation* that links such quantity to other accessible variables:

$$P = \frac{Nk_BT}{V} + \frac{1}{Vd}\langle \sum_{i<j} \vec{F_{ij}} \cdot \vec{r_{ij}} \rangle \tag{2.1.1.7}$$

In this equation, d is the dimensionality of the system (typically $d = 3$), $\vec{F_{ij}}$ is the internal force that the j-th particle exerts on the i-th particle, and r_{ij} is the vector joining them. The first is an ideal gas term, while the second is an interaction term. The brackets indicate an ensemble average.

Different methods can be employed for the control of the pressure. In the simplest case, the volume can be varied by scaling isotropically the particles positions. The *Berendsen barostat* couples the system with a reservoir at pressure P and with coupling constant τ_P:

$$\dot{V} = \frac{\kappa}{\tau_P}(P(t) - P_{bath})V \tag{2.1.1.8}$$

Here, κ is the isothermal compressibility, given by:

$$\kappa = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \tag{2.1.1.9}$$

The *Nosé-Hoover* barostat implements an additional term to the Lagrangian, referred to as a piston, and the *Parrinello-Rahman barostat* implements this in an independent way for each of the dimensions of the unit cell, thus allowing for shape variations. Both these approaches are more accurate, but slower, with respect to the Berendsen one.

The *Langevin piston*[51] is a modification of the Nosé-Hoover one, implemented by applying Langevin dynamics for the piston fluctuations. In this case, the following equations hold:

$$\dot{\vec{r_i}} = \frac{\vec{p_i}}{m_i} + \frac{1}{3}\frac{\dot{V}}{V}\vec{r_i} \tag{2.1.1.10}$$

$$\dot{\vec{p_i}} = \vec{F_i} - \frac{1}{3}\frac{\dot{V}}{V}\vec{p_i} \tag{2.1.1.11}$$

$$\ddot{V} = \frac{1}{W}(P(t) - P_{bath}) - \gamma\dot{V} + R(t) \tag{2.1.1.12}$$

In here, W is the fictitious mass of the piston, which has dimensions $[ML^{-4}]$, γ is a collision frequency and $R(t)$ is a random force whose value is extracted from a Gaussian distribution with mean equal to zero and variance equal to:

$$\langle R(0)R(t)\rangle = \frac{2\gamma k_B T\delta(t)}{W} \tag{2.1.1.13}$$

The fictitious mass of the piston is also linked to the piston oscillation period τ_P by equation:

$$W = 3N\tau_P^2 k_B T \tag{2.1.1.14}$$

If no random fluctuations of pressure take place and no friction occurs, the cell volume would oscillate with an angular period of value τ_P.

Langevin dynamics is the technique implemented in all projects described in this thesis for the control of both the temperature and the pressure of the systems. It is worth of mention that, while either the temperature or the pressure (or both) are guaranteed to be conserved up to given fluctuations, Langevin dynamics are unsuitable for the accurate study of diffusion processes and can only be employed for equilibrium processes.

Besides, because of the presence of a stochastic term in the Langevin dynamics, the generated trajectory will be dependent on the random numbers generator seed, thus the usage of different seeds allows for the sampling of different realisations of the same ensemble. Hence, the sampling of a given phenomenon can, in principle, be done in two ways: either with one long simulation, or with multiple shorter ones (provided that each short simulation is still long enough to generate meaningful statistics of an equilibrium phenomenon). While the first approach provides a robust statistics of equilibrium properties, or conversely allows for unbiased sampling of rare, non-equilibrium events, the latter approach may prove more effective for the exploration of stochastic phenomena, e.g. specific interactions or binding events of molecules characterised by high mobility within the system.

2.1.2 Energy Minimisation and Equilibration

Crystal structures of biological molecules are solved in conformations associated with specific active or inactive conformations (e.g. different conductive states, in the case of pLGICs) and/or are co-crystallized with ligands or chaperones. However, these structures normally require additional adjustments before any MD simulation can be carried out. These include the addition of a solvation box, ions, and in the case of pLGICs and other membrane proteins, a lipid bilayer; in other words all complementary molecules that contribute in making the system more realistic.

An *energy minimisation* procedure is therefore carried out as a preliminary stage. This is typically done by means of two algorithms, the *steepest descent algorithm* and the *conjugated gradient algorithm.* The former is more efficient in the first stages of minimisation, while the latter becomes more effective closer to the energy minimum, thus they are typically employed one after the other.

Different systems require different minimisation protocols. Typically, the larger and more complex the system, the more articulated the protocol should be. In the case of pLGICs, three minimisation steps are typically implemented: first, water and ions are minimised, while freezing the particles of both the protein and the lipid membrane (i.e. not implementing the integration of the equations of motion for such selected particles), although ions are typically placed in energetic minima around the complex of protein and membrane. Then, everything except the backbone atoms of the protein (i.e. the carbon of the carbonyl group, the α-carbons, and the nitrogen of the amino group, for every amino-acid), or everything except the α-carbons, are released. Finally, every particle of the structure is minimised. Such a protocol aims at adjusting solvent and lipid membrane first, while at the same time trying to preserve to the global conformation of the membrane protein, whose crystal structure is supposed to be already representative of a realistic state.

After minimisation, the resulting structure is found in an energetically convenient conformation, which is less likely to be unstable upon starting the simulation. However, this conformation does not likely resemble a representative state of the system at the wanted macroscopic parameters of the desired ensemble, such as temperature and pressure. For this reason, another stage, referred to as *equilibration*, is typically necessary after minimisation. This state normally includes an initial heating up to the target temperature, followed by an NVT equilibration at the desired temperature, and by an NPT equilibration at the desired temperature and pressure. Depending on the specific system requirements and complexity, the length, order, and extra conditions (like: the application of harmonic restraints on specific sections of the system) of each of these sub-stages may vary.

Once temperature and pressure have reached a stable plateau, a long simulation, referred to as *production*, can be carried out within the desired ensemble, and analysis can be performed over it. Typically this stage lasts over time windows of hundreds of nanoseconds to microseconds, in order to provide enough sampling of the phenomena of interest.

2.2 Force-fields

In classical MD, the particles of a given system interact with each other by means of an empirical potential, which can be modelled in different ways: we speak, therefore, of different *force-fields*. In the case of the projects here described, the *Amber* force-field potential [52] was used (figure 2.1), which is given by the following equation:

$$U_{AMBER} = U_B + U_{NB} \tag{2.2.0.1}$$

Here, U_B represents the so called *bonding terms*:

$$
\begin{aligned}
U_B = &\sum_{bonds} \frac{1}{2} K_b (l - l_0)^2 + \\
&\sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \\
&\sum_{dihedrals} \frac{U_n}{2} (1 + \cos(n\omega - \gamma))
\end{aligned}
\tag{2.2.0.2}
$$

and U_{NB} represents the so called *non-bonding terms*:

$$
\begin{aligned}
U_{NB} = &\sum_{j=1}^{N-1} \sum_{i=j+1}^{N} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \\
&\sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}
\end{aligned}
\tag{2.2.0.3}
$$

In the bonding terms, K_b and K_θ are the force constants for the bonds and for the bond angles, respectively, l and θ are the bond length and the bond angle, l_0 and θ_0 are the equilibrium bond length and the equilibrium bond angle, ω is the dihedral angle, U_n is the corresponding force constant, n is the periodicity, and γ is the phase angle (which is either equal to $0°$ or to $180°$).

The non-bonding terms describe two contributions. One term represents the Coulomb interaction between the partial charges q_i and q_j of the corresponding atoms, which are due to their electronegativity, and that are computed, for the Amber force-field, from electrostatic potential fitting of high-level *ab-initio* quantum mechanics calculations. The other term is the *Lennard-Jones* potential, an empirical potential which represents the Van der Waals interaction at long distances and the repulsion due to Pauli's exclusion principle at short distances. In this term, ϵ_{ij} is the depth of the potential well due to the Lennard-Jones potential, and σ_{ij} is the distance, between the i-th particle and the j-th particle, at which the potential is zero. These two quantities are computed for each ij pair by mixing σ_i, σ_j and ϵ_i, ϵ_j of the single particles via certain mixing rules. For the Amber force-field, the *Lorentz/Berthelot* mixing rules hold:

$$
\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \tag{2.2.0.4}
$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \tag{2.2.0.5}$$

The non-bonded terms are computed for all particle pairs that are either separated by more than three bonds or are not bonded. For the interactions between particles separated by three bonds, the *one to four interactions*, the electrostatic and Van der Waals terms are reduced by a certain factor, that depends on the specific implementation of the force-field (in this case, by a factor of 1/1.2).

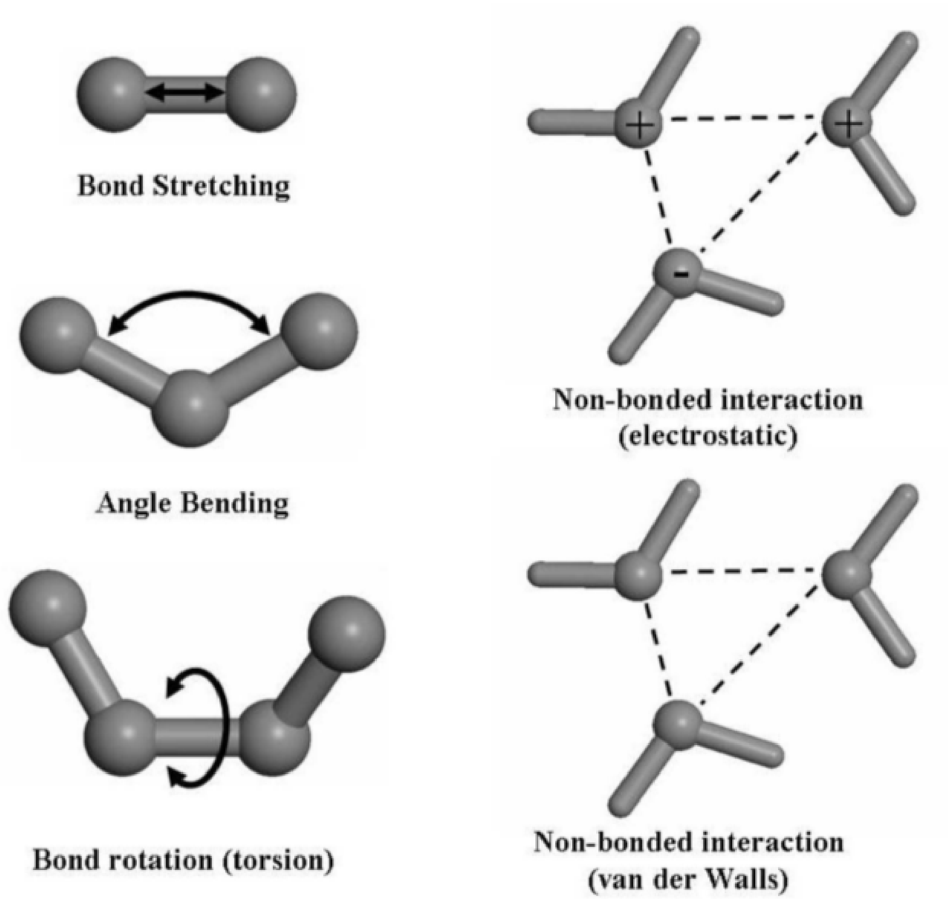


Figure 2.1: Schematic representation of the Amber force-field terms.

2.2.1 Cut-offs and Boundary Conditions

As we said, the acceleration that all particles undergo is ultimately due to the inter-particle potential that each particle exert on each other, which is, again, modelled by means of a force-field. Thus, in principle, the evolution of a single time step requires a computational time that is proportional to the square of the number of particles. This slows down significantly the computation for large systems. However, useful approximations can be made regarding non-bonded interactions to speed up the calculation.

The Lennard-Jones potential depends on the 6th and 12th powers of inter-particle distances. This makes this potential a relatively short-range interaction, meaning that the contribution of far away particles is tiny and can be neglected for the sake of the computation speed. Thus, to do that, interactions are disregarded if the two particles distance is greater than a given radius of "cut-off" r_C, which is typically chosen as 10 or 12 Å. In this way, rather than calculating $N(N-1)/2$ interactions (which is the number of pairs in a list of N particles), we only have to calculate $M \cdot N$ interactions (where M is the number of nearest neighbours of a given particle): the scaling of the computational time with the number of particles is now linear rather than quadratic.

The truncation of the interactions beyond the cut-off r_C results in an error in the estimation of the total potential energy. If the interactions decay rapidly with increasing distance, it is possible to correct for this systematic error by adding a tail contribution to the total potential energy.

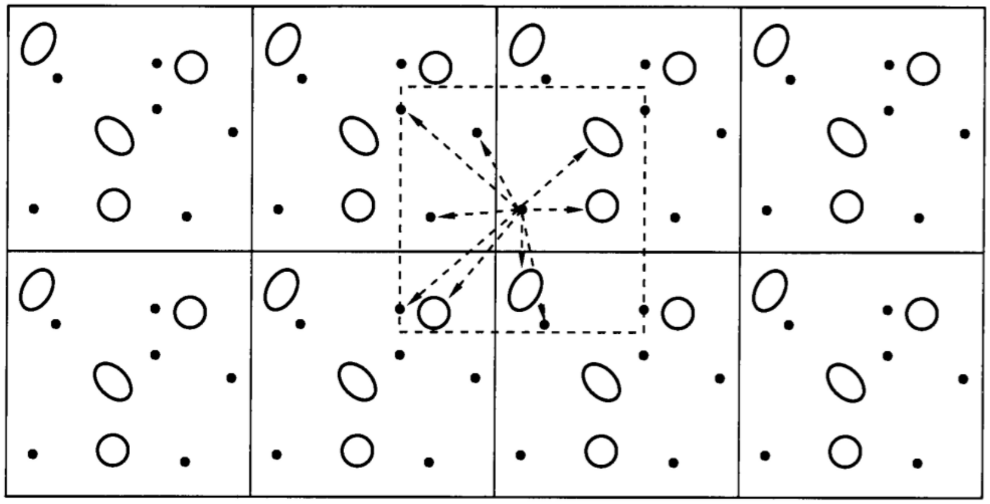


Figure 2.2: Schematic representation of periodic boundary conditions (adapted from reference [6]).

When it comes to cut-offs, talking about *periodic boundary conditions* (PBCs) becomes relevant. These are a computational artifact that is introduced in order to avoid spurious effects at the surface of the volume of the system, as well as allowing for the simulation of periodic systems. They consist in having the volume containing the N particles of the system treated as the *primitive supercell* of an infinite periodic lattice of identical cells, so that when a particle leaves the simulation box, its image enters from the opposite side. Figure 2.2 provides a representation of the implementation of PBCs. In principle, a given particle of the system interacts with all the other particles of this infinite lattice.

PBCs introduce the problem that if the potential that binds the particles is pairwise additive, then the total potential energy of the system would be given by an infinite sum. The introduction of a cut-off accounts for this problem. For a system with PBCs, it must be $r_C < L/2$ (i.e. lower than half of the side of the periodic box) so that every particle interacts only with the nearest periodic image of any other particle.

In the case of explicit-solvent molecular dynamics, the unit cell is filled with solvent molecules, typically water, and ions in a measure to counterbalance the net charge of the system, as well as to account for physiological conditions. Normally, the shape of the unit cell is orthorhombic. Another commonly chosen shape is the truncated octahedron, which results convenient for large systems, as it minimises the number of particles, given that it is the space-filling shape closest to a sphere. This shape is also useful for systems which are subject to tumbling and rotations. Whatever the chosen shape for the unit cell, its "buffer" (i.e. the minimum radius from the protein, at the centre of the cell, to the edges of the cell) must be large enough so that the protein does not interact in unphysical ways with its own images by means of long and short range interactions (for the latter, the choice of a cut-off is also inherent to this issue). Typical choices for the buffer are 12 or 15 Å.

PBCs are also important to be consider when performing quantitative analysis on the systems, being especially important for dynamical quantities such as the root mean square displacement (discussed in Appendix A): in this case, it becomes necessary to "unwrap" the trajectory of the system, which means removing the image and treat the diffusion of each particle as if PBCs were not imposed.

2.2.2 Long-range Interactions

While the Lennard-Jones is a short-range potential and can thus be truncated beyond a certain cut-off, this is not possible for long-range interactions, such as the Coulomb potential, which decays as $1/r$ and accounts for the other non-bonded term of the Amber potential. For this reason, the *Ewald sums* are introduced.

The concept at the basis of this method is that a set of point-like charges (which is the case of most systems of interest) is a sum of δ functions. The assumption made is that every i-th particle with charge q_i is surrounded by a diffuse charge distribution of the opposite sign, so that the total charge of this cloud counterbalance q_i. In this way, the i-th particle contributes to the electrostatic potential only for the fraction of the charge that is not screened by the cloud, so that it rapidly decays for large distances. Normally, the chosen charge distribution for the cloud is Gaussian. Therefore, the electrostatic potential in a certain position due to the screened point-like particles gets simply summed. However, the introduced fictitious charge distribution requires a correction, in order to compute the electrostatic potential of the real charge set. For this reason, a compensating charge density is also introduced. This is shown schematically in figure 2.3.

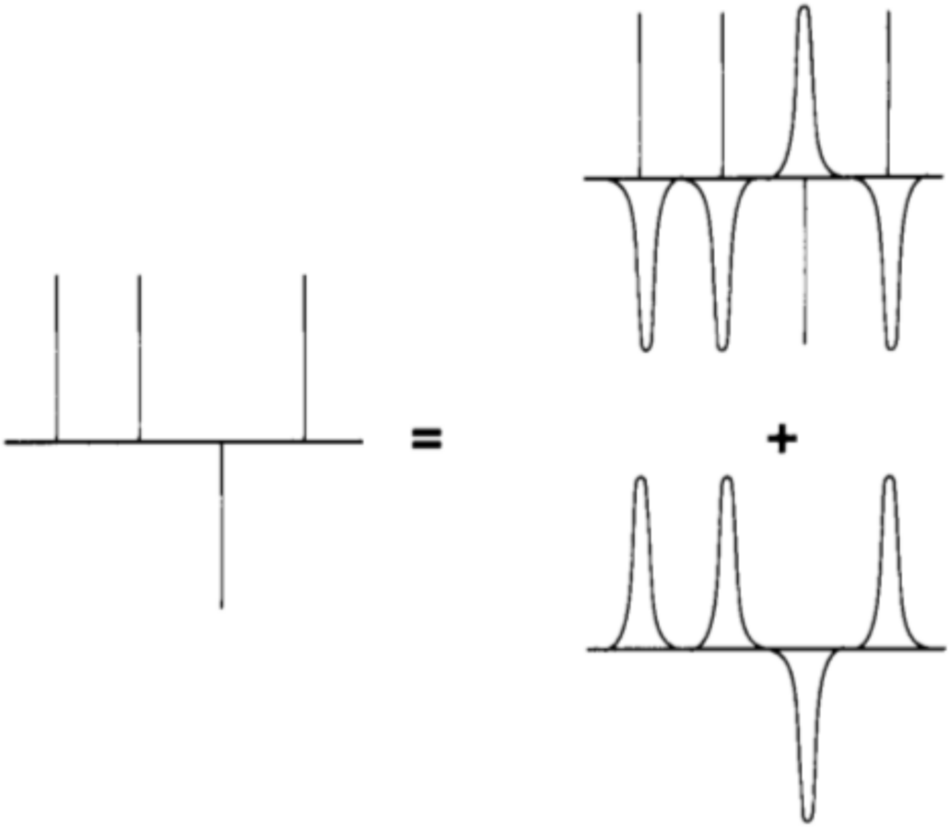


Figure 2.3: Schematic representation of the concept of Ewald sums: a set of point charges can be regarded as a set of screened point charges minus the smooth screening background (adapted from reference [6]).

The usage of this method allows for a calculation of the long-range electrostatic interactions without the need of a cut-off.

2.3 All-atom Force-fields Parametrisations

Molecular dynamics simulations can be either *all-atom* (AA), *coarse-grained* (CG), or *united atoms*, whose difference resides in the choice of the elementary particles whose evolution in time is implemented. *All-atom* MD (AAMD) deals with the exact atoms (grouped into "residues") that make up the systems, relying on a parametrisation performed from more precise *ab-initio* methods, and/or to fit experimentally-evaluated properties. *Coarse-grained* MD (CGMD) deal with "beads", i.e. bulkier elementary particles that typically corresponds to chemical groups (as in the MARTINI force-field[53]) or even larger clusters of atoms, depending on the specific CG method implemented. In this thesis, only AAMD has been used.

Each atom of an AAMD simulation has an associated *atom type*, which takes into account not only the atomic species (e.g. H, O, C, N, ...), but also the chemical identity of the atom, which is given by its topological location within the residue/molecule. Atom types defines which bonded and non-bonded parameters (fitted by *ab-initio* calculations) must be assigned to the corresponding atoms, allowing for the integration of the equations of motions.

Three Amber force-fields, each one having its own subset of atom types to be assigned to specific residues, have been employed in the projects described in this thesis: the general Amber force-field, `ff14SB`, and `LIPID14`.

The general Amber force-field (GAFF) was developed by Wang et al. in 2004 as a general purpose force-field for the description of a variety of organic molecules, to be as compatible as possible with Amber protein force-fields.[54] This force-field comprises of 35 basic atom types: five carbons, eight nitrogens, three oxygens, five sulfurs, four phosphori, six hydrogens, and one atom type for each of the four common halogens. These atom types were chosen according to element type, hybridization, aromaticity, and chemical environment. For example, fluorine, chlorine, bromine, and iodine have only one atom type, while carbon has five atom types: `c3` (sp3 hybridization), `c1` (sp1 hybridization), `c` (sp2 hybridization, and specific chemical environment, i.e., bonded to S or O), `ca` (sp2 hybridization, and belonging to an aromatic ring), and `c2` (sp2 hybridization). The GAFF also includes other 22 special atom types, that were introduced for different purposes.

The force-field `ff14SB`[55] was released by Maier et al. in 2015 for the parametrisation of protein residues, i.e. amino-acids. It is an improvement of the `ff99SB`[56] force-field, which did not provide a good energetic balance between helical and extended parts of the backbones of proteins and peptides. Moreover, `ff14SB` offers updated side chain dihedral corrections for lysine, arginine, glutamate, glutamine, methionine, serine, threonine, valine, tryptophan, cysteine, phenylalanine, tyrosine, and histidine, and increases reproduction of experimentally indicated geometries with respect to `ff99SB`. It is worth mentioning that `ff19SB` has been released in 2020 as the newest protein force-field within the Amber family:[57] this better reproduces backbone profiles for all amino-acids and improves the capability to differentiate amino-acid-dependent properties such as helical propensities.

Membrane proteins, which are embedded in lipids, require the introduction of a suitable force-field for the simulation of lipid molecules. Traditionally, for such molecules Amber has either employed the Charmm parameters,[58] or adaptations of the general Amber force-field, with limited success. From 2011 onwards with the introduction of `LIPID11`, ad-hoc Amber force-fields for the simulations of lipids have been developed, with `LIPID14` in 2014 as the latest release. `LIPID14`[59] is a modular force-field for cholesterol as well as phospholipids. Such modularity allows, in the case of phospholipids, for many combinations of lipid head groups and tail groups, so that each head and each tail makes up for a separate residue. With respect to its predecessors, several van der Waals and dihedral angle parameters have been refined to fit experimental data and quantum energies, while new partial charges have been obtained for the head and tail groups. `LIPID14` has been designed in order to be compatible with the other pairwise-additive Amber force-fields for proteins, nucleic acids, carbohydrates, and small molecules.

Water is very often present in molecular dynamics simulations, as the most common solvent used. As such, it requires a dedicated parametrisation to take into account many of its properties. Several force-fields have been implemented for water. One of the most used in biological simulations, including all the projects described in this thesis, is TIP3P,[60] which is a rigid 3-site model for water, with molecules made up of one oxygen and two hydrogens, according to its chemical formula. Despite its simplicity and convenient computational cost, this model fails to correctly reproduce certain features (among others, the oxygen-oxygen pair correlation function, and water viscosity[61, 62]). While 4-site models, such as TIP4P,[60] have been introduced to best reproduce water properties, TIP3P remain a useful model for the solvation of biological system and is thus extensively used.

2.4 Modelling Ligands

As previously mentioned, the aim of molecular dynamics simulations and enhanced sampling techniques is often to study the interaction of proteins with smaller molecules that can bind to their surfaces or within certain pockets. While proteins amino-acids are frequently-occurring residues whose parametrisation is well studied and provided by force-fields such as `ff14SB`, these smaller molecules constitute non-standard residues whose parametrisation can be done according to the general Amber force-field. Here we describe a procedure to treat these residues by using Amber software as well as the Gaussian package.[63]

As for the partial charges to be assigned to each of the atoms, they are typically calculated from *ab-initio* calculations. These consists normally of two different stages: first, a *geometry optimisation* is carried out in order to find a local energy minimum with respect to the original conformation; once this is done, a *single point energy calculation* is performed at a fixed geometry. The first stage can make use for example of any density functional theory level and any basis set, while the single point energy calculation is usually performed at the Hartree Fock level with the 6-31G* basis set for consistency with the Amber force-field parametrisation (6-31G* refers to the usage of 6 Gaussian type orbitals (GTO) for inner shell, 3 GTOs for inner valence, 1 GTO for outer valence, and the symbol * means that d-type functions are added on to atoms other than hydrogens). The partial charges so calculated for the parametrisation are the ESP (electrostatic potential) charges. These charges well reproduce the quantum mechanically determined multipole moments as well as the intermolecular interactions with surrounding molecules. However, ESP charges present certain flaws, including the scarce transferability between common functional groups in related molecules, the frequent conformation dependency, and the frequent occurrence of large charges that can prove problematic for the simulation of intra-molecular interactions. The origin of these flaws is to be searched in the fitting process: buried atoms charges can vary wildly in order to obtain tiny fit quality improvements. To overcome this issue, the restrained electrostatic potential (RESP) charges were proposed,[64] whose idea is to introduce a penalty function, whose aim is to restraint non-hydrogen atomic charges to a target value. For such target, Mulliken charges were initially proposed, but resulted ineffective. In their place, charges with zero value were proposed as targets (i.e., restraining the charges according to their magnitude).

For flexible molecules, it is important to consider that the charge distribution may be different alongside the molecule depending on its instantaneous conformation. Thus, the user may want to fit the RESP charges in such a way that this is taken into account. This can be done by a multiconformational RESP fitting,[65] which consists in calculating ESP charges for selected different meaningful conformations, and then performing a weighted fit. Each conformation i is assigned a weight, that can take the form of the Boltzmann factor w_i:

$$w_i = \frac{\exp(-(E_i - E_0)/k_B T)}{\sum_j \exp(-(E_j - E_0)/k_B T)} \tag{2.4.0.1}$$

where E_i is the i-th conformation energy, k_B is the Boltzmann constant, T is the desired temperature for the simulation, and E_0 is the minimum energy conformation.

2.5 Enhanced Sampling Techniques

Molecular dynamics covers, depending on size of the system, time spans of orders between tens/hundreds of nanoseconds to several microseconds. This prevents it from sampling processes that require the overcoming of significant energetic barriers: these so called *rare events*, in fact, typically occurs in times of the order of microseconds, milliseconds or larger. Thus, such processes can be conveniently reconstructed only by means of more advanced techniques, referred to as *enhanced sampling techniques*. Most of these are based on standard molecular dynamics, with the addition of certain algorithms that aim at modifying the energetic landscape, in order to provide easier sampling of the phenomena of interest. Among these methods, some examples are umbrella sampling,[66] metadynamics (and all of its variations and improvements),[67, 68] steered molecular dynamics,[69] and the string method.[70, 71] Some of these are used in the projects of this thesis, and are described in the following.

2.5.1 Metadynamics

A commonly used enhanced sampling technique for the acceleration of rare events in molecular dynamics simulations is *metadynamics* (MetaD),[67, 68] which also allows for the reconstruction of the free energy landscape of the sampled phenomenon.

In this method, suitable reaction coordinates, referred to as *collective variables* (CVs), must be identified. These quantities must be able to accurately distinguish the energetic basins separated by the high-energy barriers associated with the rare event of interest, and that need to be overcome. Thus, these collective variables can be used to build a history-dependent *bias potential*, that is an additive term for the Hamiltonian of the system. This potential consists of a sum of Gaussian repulsive terms centred on the positions in the CVs space that the system visits throughout the simulation:

$$U_{bias}(\vec{s}, t) = \sum_{n\tau < t} W \exp\left(- \sum_{i=1}^{d} \frac{(s_i - s_i(\vec{q}(n\tau)))^2}{2\sigma_i^2} \right) \tag{2.5.1.1}$$

Here, τ is the deposition time, $s_i(\vec{q}(n\tau))$ is the i-th collective variable, which is a function of the phase space coordinates $\vec{q}$ at time $n\tau$ (n being an integer positive number), d is the total number of collective variables, σ_i is the width of the Gaussian associated to the i-th CV, and is typically set as a fraction of the fluctuations in an unbiased simulation, and W is the Gaussian height, whose choice is a compromise between the required computational time and the resulting error for the free energy landscape reconstruction.

In a typical case, the simulation would start in an energy basin that would get progressively filled by the bias potential. Eventually, this basin will be completely full, allowing for the system to move into new basins, as represented in figure 2.4. On the long run, the entire CVs space will be filled with Gaussians, so that the system will be able to sample all conformations associated with the CVs in the so called *diffusive state.*

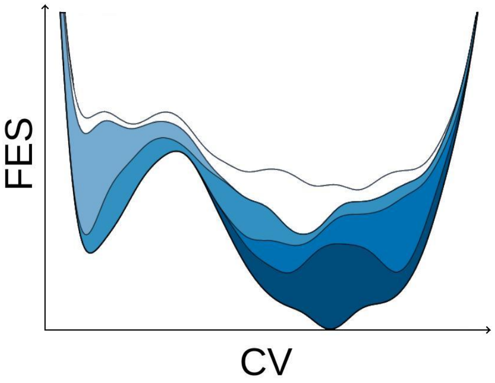


Figure 2.4: Free energy basins are filled with progressive Gaussian deposition during a metadynamics simulation. The bias potential is represented in shades of blue, from dark to light according to the simulation time lagging.

Thus, metadynamics allows for the reconstruction of the free energy surface (FES) as a function of the selected collective variables, since, in the limit of long simulation time, the bias potential will fluctuate around the value of $F(\vec{s})$ plus a constant C that increases with time:

$$U_{bias}(\vec{s}, t \rightarrow \infty) = -F(\vec{s}) + C \tag{2.5.1.2}$$

2.5.2 Well-tempered Metadynamics

An improved version of metadynamics is *well-tempered metadynamics*,[72] in which the Gaussian height W is a function of time:

$$W(t) = W_0 \exp(-U_{bias}(\vec{s}(\vec{q}(n\tau)))/k_B \Delta T)\,,\ n\tau < t \tag{2.5.2.1}$$

Here, ΔT is a parameter with the dimensions of a temperature. It can be demonstrated that well-tempered metadynamics samples an ensemble characterised by a fictitious temperature $T+\Delta T$. Two limit cases can be identified: for $\Delta T \rightarrow \infty$, standard metadynamics is being applied, while for $\Delta T \rightarrow 0$ no bias is applied, therefore a Boltzmann sampling is being performed.

In the limit of long simulation times, the bias potential can again be used to estimate the FES, which is in this case given by the following equation:

$$U_{bias}(\vec{s}, t \rightarrow \infty) = \frac{-\Delta T}{T + \Delta T} F(\vec{s}) + C \tag{2.5.2.2}$$

where C is a time-dependent constant.

While in a standard metadynamics simulation the free energy will fluctuate, over long times, around the correct value, well-tempered metadynamics will eventually converge to the exact value, resulting in an optimal use of computer time and making the post-processing of the results not necessary (with longer simulation times resulting in enhanced statistics).

The well-tempered metadynamics setup is controlled by choosing a value for the bias factor γ:

$$\gamma = \frac{T + \Delta T}{T} \tag{2.5.2.3}$$

Values of γ between 5 and 15 are typically reasonable for biological systems, with a value of 15 helping speed up the convergence in large systems. This is the value used in the well-tempered metadynamics simulations described in this thesis.

2.5.3 Reweighting the Free Energy Landscape

A metadynamics run yields a correct distribution (i.e. according to the implemented ensemble) only for the CVs used to first deposit the bias potential. The recovery of the unbiased distribution of any other observable in such a run can be achieved by performing a reweighting of the free energy landscape.[73] This can be done by considering the relation between the Boltzmann distribution $P_0(\vec{r}) = e^{-\beta U(\vec{r})}/Z$ and the probability distribution $P(\vec{r}, t)$ at time t:

$$P(\vec{r}, t) = P_0(\vec{r}) e^{-\beta(U_{bias}(\vec{s}(\vec{r}), t) + C(t))} \tag{2.5.3.1}$$

where $U(\vec{r})$ is the system internal potential, $U_{bias}(\vec{s}(\vec{r}), t)$ is the bias potential, and $C(t)$ is the time-dependent bias offset, defined as:

$$C(t) = -\frac{1}{\beta} \log \left(\frac{\int d\vec{s} e^{-\beta F(\vec{s})}}{\int d\vec{s} e^{-\beta(F(\vec{s}) + U_{bias}(\vec{s}, t))}} \right) \tag{2.5.3.2}$$

2.5.4 Time Reweighting with the Acceleration Factor

In a metadynamics run, the simulation time is not representative of the physical time since the sampling potential is being artificially altered. However, when one is not interested in the reconstruction of the free energy profile associated with a rare event, but rather in the calculation of the rates of rare event barrier crossing (e.g. in the unbinding times of a ligand from an orthosteric pocket), the reconstruction of the real time from the metadynamics one becomes paramount. This can be done, provided that the bias deposition is *infrequent*,[74] so that the assumption that no bias at all is deposited on the transition state region holds. When this assumption is valid, the distribution of the unbinding times follows a Poissonian distribution; this can be checked *a posteriori* by means of a *Kolmogorov-Smirnov* (KS) statistical test,[75, 76] which has been established as a standard checking method for the quality of the reweighted time distribution.[77]

The reweighting of time from a metadynamics simulation can be performed by means of a simple equation:[78]

$$t = t_{METAD} \cdot \alpha(t_{METAD}) \tag{2.5.4.1}$$

where t_{METAD} is the metadynamics time α is the so called *acceleration factor*, that can be computed from the bias as:

$$\alpha = \langle e^{U_{bias}(\vec{s}, t_{METAD})/k_B T} \rangle_V \tag{2.5.4.2}$$

where $\vec{s}$ is the CVs vector at time t_{METAD}. In practice, the reweighting of time is performed as:

$$t = \sum_{i}^{n_{METAD}} dt e^{U_{bias}(\vec{s}(t_{iMETAD}), t_{iMETAD})/k_B T} \tag{2.5.4.3}$$

where n_{METAD} is the number of metadynamics simulation steps and $t_{iMETAD} = idt$ is the metadynamics time at the i-th step.

In practice, a certain number of independent metadynamics runs with infrequent bias depositions are started from the same conformation and stopped as soon as the barrier crossing occurs. An average of the obtained rescaled crossing times can be used as the characteristic time τ of the Poisson distribution. The empirical cumulative distribution function (ECDF) is then built using the different simulations, and a theoretical cumulative distribution time (TCDF) is built from a large number of times randomly generated according to a Poisson distribution with characteristic time τ:

$$TCDF = 1 - e^{-t/\tau} \tag{2.5.4.4}$$

In this way, a two-sample KS test is carried out to verify the null hypothesis that the ECDF and the TCDF do not show any significant difference and thus correspond to the same distribution. The null hypothesis is rejected if the p-value resulting from the KS test is lower than a certain significance level α, which is usually chosen to be 0.05.

If significant bias is added in the transition state region, or if the choice of the CVs does not properly describe certain slow motions that remain hidden, the Kolmogorov-Smirnov test would not produce a low enough p-value. A correct frequency of bias deposition cannot be determined *a priori* and can only be verified *a posteriori* by means of a KS test.

2.5.5 Adaptive-frequency Metadynamics

The requisite of not depositing any bias on the transition state can be fulfilled with an infrequent metadynamics run. However, such a simulation would present the disadvantage of a slow deposition at early stages, when the initial basin is being filled. For this reasons, the *adaptive-frequency* scheme was introduced as to bridge normal and infrequent metadynamics.[79] In this scheme, the deposition time $\tau_{dep}(t)$ at the simulation time t is:

$$\tau_{dep}(t) = \min(\tau_0 \max(\alpha(t)/\theta, 1), \tau_c) \tag{2.5.5.1}$$

where τ_0 is the initial deposition time, τ_c is a cut-off value for the deposition time (to be considered equal to the typical duration of a barrier crossing), and θ is a threshold value for the acceleration factor to trigger the switch from frequent to infrequent metadynamics. θ should be as large as possible in order to delay the change, although too large θ may lead to bias being deposited on the transition state. An estimation for a reliable value of θ, given an expected transition time τ_{exp} (obtained either from experimental measure or from an estimation), the simulation time τ_{sim} (dependent on computational resources) and a “safety coefficient” C_s, is given by:

$$\theta \leq \frac{\tau_{exp}}{\tau_{sim} C_s} \tag{2.5.5.2}$$

In the case of ligand unbinding, one can estimate τ_{exp} to be on the order of 1 s,[79] C_s can be taken equal to 10^2 as to avoid the risk of a transition time falling in the long tail of a Poissonian distribution, and τ_{sim} may be set in the range of tens to hundreds of nanoseconds of metadynamics run. This yields a value of about 10^5 for the upper limit of θ.

In equation 2.5.5.1, the deposition is a function of the acceleration function. Since in practice significant fluctuations of $\tau_{dep}(t)$ may be observed (which may be problematic at the transition state), the final scheme is implemented as:

$$\tau'_{dep}(t) = \max(\tau_{dep}((n-1)\Delta t), \tau_{dep}(n\Delta t)) \tag{2.5.5.3}$$

where $\tau(n\Delta t)$ is the deposition time at step n of duration Δt.

Having so far provided the context of pLGICs and of the *in-silico* techniques used to study the 5-HT$_{3A}$R, in the following chapters we will discuss each project carried out, starting with pre-existing knowledge from the literature, describing the methods used, highlighting our findings and drawing our conclusions and suggestions for future perspectives. All the described simulations of the receptor-membrane complex in solution were carried out in the High Performance Computing resources ARCHER and Thomas, by using the NAMD package[80] with 12 nodes and 288 CPU processors (i.e. 24 processors per node), yielding 10 ns per day.

Chapter 3

Mutual Effects of a Proline Switch and the Protein Environment of the 5-HT$_{3A}$ Receptor

3.1 Introduction

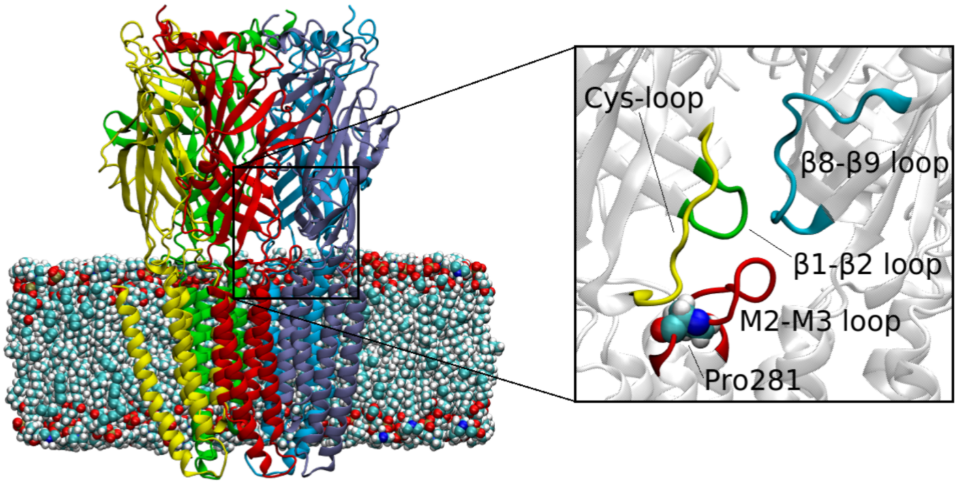


Figure 3.1: On the left, a section of the 5-HT$_{3A}$R model created for this work, embedded in a POPC lipid bilayer. Subunit 1 is in red, 2 in yellow, 3 in green, 4 in light blue, and 5 in blue. On the right, a zoomed view of the ECD-TMD interface is shown, with the Cys-loop represented in yellow, the β1-β2 loop in green, the β8-β9 loop in blue, and the M2-M3 loop in red. Pro281 is shown as van der Waals spheres. Figure from reference [7].

The M2-M3 loop, depicted in figure 3.1, is an important part of the 5-HT$_{3A}$R sequence, as it plays a role, thanks to its strategic position at the interface between ECD and TMD, in the transduction of the mechanical signal from the binding pocket in the ECD down to the hydrophobic gate in the TMD. This is also supported by the presence of conserved residues in its sequence.

Among the residues in its sequence (shown in figure 3.2), two prolines are found, at positions 274, close to the M2 top, and 281, just above the top of the M3. This amino-acid is renowned since its *cis* isomer is more accessible at room temperature, with respect to other amino-acids. *Trans-cis isomerisation* is the process by which the dihedral angle around the prolyl bond that connect two adjacent amino-acids changes from 180° (*trans*) to 0° (*cis*) and the other way around. This process has been, throughout the years, associated to a variety of biological processes,[81, 82, 83, 84, 85, 86, 87] including protein folding,[41] domain-domain rearrangements,[85] antibody binding,[86] amyloidogenesis,[87] uncoating and recruitment of the human HIV-1 capsid protein in the virions,[82] and modulation of association/dissociation rates in intrinsecally disordered proteins.[88]

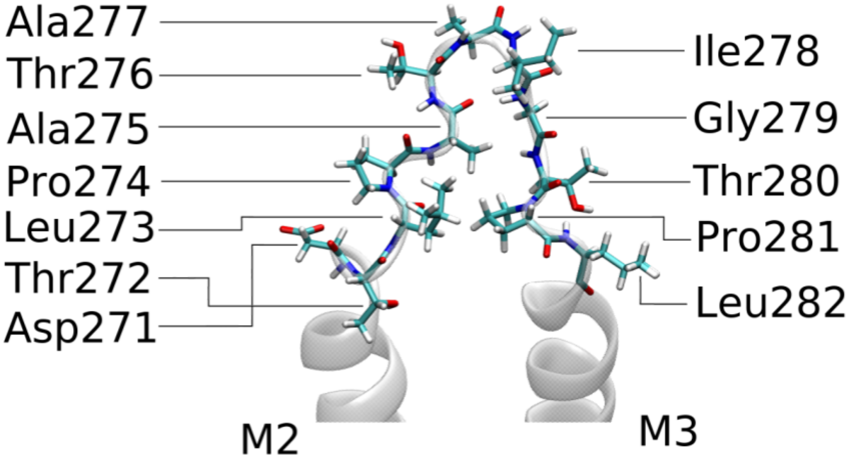


Figure 3.2: Representation of the 12 residues of the M2-M3 loop. Figure from reference [7].

The higher population of *cis* isomers of proline at room temperature with respect to other amino-acids is due to three reasons:[41] first, unfavourable interactions between atoms directly attached to the peptide unit are present both in *cis* and in *trans*; this is essentially due to the fact that the proline C$_\alpha$ (on the backbone) and the proline C$_\delta$ (on its aromatic side chain), both bonded to the proline amide nitrogen, provide a similar steric repulsion. Second, the O$_i$ to C'$_{i+1}$

electrostatic interactions result less favourable in *trans* (figure 3.3 shows proline atoms in a proline dipeptide). Third, the entropy loss from *trans* to *cis* is smaller for X-Pro peptide groups with respect to non-prolyl peptide groups.

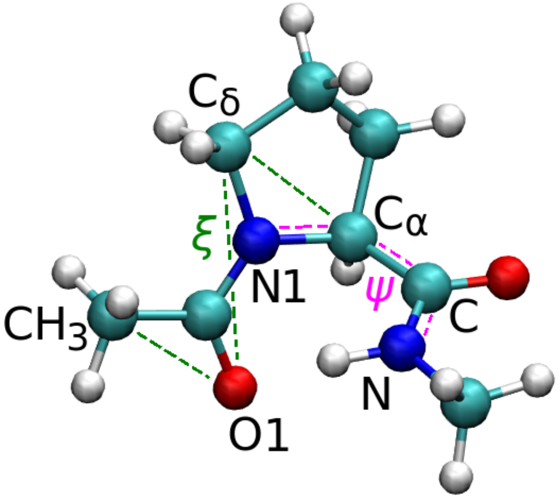


Figure 3.3: Proline dipeptide in water. In green, the dihedral angle ξ (defined by the atoms CH_3-O1-C_δ-C_α) and in magenta the dihedral angle ψ (defined by the atoms N1-Cα-C-N). Figure from reference [7].

Mutagenesis electrophysiology experiments[89] on the mouse 5-HT$_{3A}$ receptor highlighted how substituting Pro281 (Pro8* in reference [89], where 8* refers to the position within the M2-M3 loop, from N-terminus to C-terminus) affects the functionality of the channel, while leaving the ligand binding properties unaffected. When substituted with proline analogues that promote or diminish isomerisation propensity,[89] the effective concentration half (EC$_{50}$) showed a correlation with such propensity; when the substitution was made in favour of analogues strongly favouring the *trans* isomer (or of the amino-acids Gly, Ala, Cys, Val), non-functioning channels were observed, implying that isomerisation to *cis* is somehow due to occur for the channel to open. Free energy calculations of proline analogue dipeptides in acqueous solutions provided a support to these observations,[81] by finding a correlation between *trans*-*cis* energy difference in proline analogues and the experimental EC$_{50}$.

A possible hypothesis that can be drawn from these data is that proline isomerisation within the M2-M3 loop is an intermediate step that helps transmit the mechanical signal, activated by ligand binding in the ECD, down to a region which ultimately controls the activation state of the channel, i.e. the ring of five leucines at position 9 along the M2 helix (here referred to as Leu260). This ring has been suggested to control the ion flux depending on how large the pore is between their side chains.[15, 14] A further hypothesis[89] suggested that the β1-β2 loop and the Cys-loop may pin Pro281, while ligand binding may displace the Cys-loop, leading to the release of this amino-acid.

Despite the role of proline isomerisation in channel function has been proved, it remains controversial. Mutagenesis electrophysiology experiments on the homologous human 5-HT$_{3A}$ receptor revealed that substituting Pro8* with a histidine or a tryptohan did not alter channel function significantly, although other effects were observed, namely an altered dependence on the extracellular concentration of Ca^{2+} ions and an increase of the rate of desensitisation.[90] Further data from mutagenesis experiments prove that the substitution of a proline at 8* position can be rescued by charged amino-acids, in the specific by Glu and His (although not by Asp, Lys, nor Arg). Glu and His, with respect to the other charged amino-acids, allow for the formation of hydrogen bonds with the backbone amide NH of the Ala adjacent to Pro274 (i.e. Ala275), which is therefore argued to be responsible for the compensation of proline absence.[91] Furthermore, the proline switch may not be a part of the mechanism culminating in gating that applies to all pLGICs: in the nicotinic acetylcholine receptor, mutagenesis experiments revealed how the equivalent Pro8* can play a role in the coupling of ligand binding to channel gating due to its steric and hydrophobic compatibility with the surrounding residues, rather than to isomerisation.[92] Mutagenesis electrophysiology experiments performed on ELIC showed that Pro-to-Ala mutations on Pro253 and on Pro256 of the M2-M3 loop do not ablate function.[93]

Another controversy concerns the time scales over which proline isomerisation occurs, since intrinsic prolyl *trans-cis* isomerisation rates in peptides are only an order of magnitude smaller than gating time scale.[89] It is possible, in principle, that the receptor environment surrounding the proline may affect the energetic process associated with its isomerisation, thus making it a possible step in the gating mechanism.[89, 90, 39, 94] Finally, it is worth pointing out that, in all 5-HT$_{3A}$R structures available,[2, 46, 47, 5] no matter their conductive state, the five subunits were found in *trans.*

In this context, computer simulations can provide useful insights at the atomistic level while more experiments are needed to determine unambiguously what is the role of proline isomerisation.[95] Thus, in this chapter, we focus our attention to the modeling of this phenomenon within the receptor, by means of all-atom MD as well as metadynamics simulations. The work presented in this chapter was published in The Journal of Physical Chemistry Letters in 2019, appropriate references are made where relevant.[7]

3.2 Model and Simulation Details

For this work, we built a model, depicted in figure 3.1, of the 5-HT$_{3A}$ receptor from the X-ray structure at 3.5 Å resolution determined by Hassaine et al. (PDB entry: 4PIR).[2]

Its structure features the extracellular domain (ECD), a transmembrane domain (TMD), and an incomplete intracellular domain (ICD), that lacks the MX-MA loops.

As previously described in the introductory chapter, the pore width at the level of the hydrophobic gate, as well as the fact the the crystallisation was performed with antagonists Llama-derived single-chain antibodies (VHH), would suggest that the channel is found in a non-conductive state. No ligands are present in the orthosteric sites, nor were added in the model built.

As previously stated, more recent structures of the 5-HT$_{3A}$ receptor, which have been deemed in different conductive states, although this project was started upfront of their release. When compared to the cryo-EM apo-structure,[47] the X-ray structure the ECD appeared to be twisted counterclockwise around the pore axis with respect to the deemed closed conformation, thus may constitute a more intermediate step from the closed to the open conformations. Considering that the role of proline isomerisation is likely to be played as an intermediate step between ligand binding and channel opening, the X-ray structure thus constitutes, even *a-posteriori*, a valid choice for building our model.

Four residues of the M2-M3 loop, namely Thr276, Ala277, Ile278 and Gly279, are missing in the X-ray structure, thus required to be manually added: the optimisation of their structure and the refinement of their rotameric state were performed with the web-based software NQ-flipper.[96] The five Pro281 and the five Pro274 were all in the *trans* isomer in the PDB structure.

Several parts of the original PDB were removed, including the crystallisation chaperones, the non-native fragments, and the MA and MX helices which make up part of the ICD. The long loop connecting the helices MA and MX was missing even in the solved X-ray structure, due to its removal by proteolysis in the experiments. The removal of the ICD domain is not expected to affect our project, since its role mainly consists in a final selectivity filter for the ion flow,[97, 98] and its action on the ions is beyond the scope of our investigation.

The M3 and M4 helices were directly connected via an artificial linker, of sequence VHKQDLQRD, in a similar way to what is observed in pLGICs that lack the ICD, such as ELIC and GluCl. Thus, the residues 316-333 and 399-424 were omitted.

The built model was protonated at neutral pH, embedded in a lipid bilayer composed of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) by means of the web-based CHARMM-GUI software,[99] then solvated in a orthorhombic supercell of 12 Å of buffer of TIP3P water and 0.15 M of Na^+ and Cl^- ions to resemble physiological conditions.

The final model contained 183,207 atoms in total, and was simulated with the NAMD 2.9 package for molecular dynamics[80] and the Amber `ff14SB`[55] and `LIPID14`[59] force-fields.

The integration step chosen was 2 fs, the bonds containing hydrogen were constrained by means of the SHAKE algorithm,[48] while Particle Mesh Ewald was employed for the calculation of the electrostatic interactions and a cut off of 10 Å was implemented for the non-bonded interactions.

The structure initially underwent an energy minimisation, followed by a slow heating and a partially restrained equilibration. Simulations productions followed beyond this point, within the isothermal-isobaric NPT ensemble at 300 K and 1 atm in order to match the experiments (performed at room temperature) that this project aims at complementing with *in-silico* methods. The temperature was conserved with a Langevin thermostat with collision frequency of 1.0 ps^{-1}, and the pressure was maintained with a Langevin piston barostat,[51] with an oscillation period of 200 fs and damping time constant of 100 fs. The productions followed separate schemes, described in the following.

In order to study how the receptor environment influences the free energy of proline isomerisation, we studied such process by enhancing its sampling, by biasing two dihedral angles as collective variables (CVs): ξ and ψ (shown in figure 3.3). Thus, well-tempered metadynamics[72] simulations were performed on Pro281 in subunit 3 within the receptor. As a control system, proline dipeptide (N-acetylproline methylamide) in a orthorhombic box of water (12 Å of buffer), which consists of a proline residue capped by acetyl (ACE) and N-methyl (NME) groups, as depicted in figure 3.3, was also simulated. Well-tempered metadynamics was performed for both systems at 300 K with an initial Gaussian height of 0.8 kJ/mol, a width of 9° on both CVs, a bias factor of 15, and a deposition rate of 1 ps.

The improper diheadral angle ξ is similarly defined in the receptor, where it describes the torsion around the prolyl peptide bond which connects Thr280 to Pro281. Despite not being the dihedral that includes the two atoms of the prolyl bond, in both the receptor and the dipeptide the rotation of ξ accounts for the process of proline isomerisation, with values around $\xi \simeq \pm 180^\circ$ making up for the *trans* isomer, while values around $\xi \simeq \pm 0^\circ$ making up for *cis* one. This angle was chosen as it describes not only the isomerisation but also the pyramidalization (i.e. the out-of-plane deformation of the imide nitrogen) of the amide nitrogen N1, and it allows for the removal of the amplitude of pyramidalisation as an independent degree of freedom.

The dihedral angle ψ constitutes the proline backbone torsion angle, i.e. the torsion about the bond that, in the receptor, precedes the peptide bond between Pro281 and Leu282. It defines the amide orientation, that may affect the rate of transitions between *trans* and *cis* isomers. Such angle has been used in the past as a chaperone of ξ;[81, 82, 100] the reason for its suitability as a chaperone collective variable depends on the fact that it controls the interactions between the proline C-terminal group and the nitrogen lone pair in the transition state.[100]

The convergence of these well-tempered metadynamics runs resulted to be very different for proline dipeptide in water and proline within the receptor. In fact, the protein environment, made up of flexible loops and comprising a complex network of interactions, significantly slows down the full sampling in the case of the receptor simulation. Thus, while in the dipeptide, convergence was reached within 40 ns of well-tempered metadynamics, in the receptor it was reached over 800 ns. In order to reduce the error on the reconstructed energy profiles, the free energy surfaces reconstructed over the two CVs were averaged for proline dipeptide and for the proline in the receptor environment over the last 10 ns and 100 ns respectively. Well-tempered metadynamics simulations were performed using the Plumed 2.0 plugin.[101]

While the metadynamics of the proline isomerisation is able to reconstruct the free energy of the process, it does not leave the proline residue in a single isomer long enough, as to propagate mechanical modifications to adjacent residues that would be unambiguously observable. Moreover, since it focuses on one subunit only, it is not able to capture possible cumulative effects that the isomerisation occuring in multiple subunits may have on the channel. A combined study of the isomerisation in multiple subunits would enhance the sampling of equilibrated structural modifications that are, or may be, due to isomerisation. In order to explore this possibility, we also performed 500 ns long molecular dynamics simulations for four different models. The first model, labelled 5Trans, is characterised by the five Pro281 in *trans* and was straightforwardly obtained from the minimisation and equilibration procedure, as the *trans* isomers are the ones originally found in the 4PIR PDB structure. By means of metadynamics simulations, with 0.2 kcal/mol of Gaussian height, 9° of width, and with a deposition stride of 1 ps, the *cis* isomer was then forced on the various subunits in order to generate other models: in particular, a model with the Pro281 of subunits 1, 3 and 4 in *cis* (labelled 3Cis) and one with all five Pro281 in *cis* (5Cis). The reason for the investigation of the model 3Cis is that it is not known how many ligands are required for triggering the gating, nor, assuming that proline isomerisation would follow ligand binding in the mechanical cascade, how many of the five prolines are required to be in the *cis* conformation. It is plausible that there might be an additive effect, i.e. every global structural modification that the protein undergoes upon isomerisation becomes more pronounced with a larger number of subunits with the proline in *cis*.[13, 17] Finally, since Pro281 is not universally conserved within the Cys-loop superfamily,[25] with the same method we also generated a model where the five Pro281 are in *trans*, but the five, more conserved Pro274 are in *cis*. This model is labelled 5Trans5Cis (5T5C). In accordance to what observed in the free energy produced with the well-tempered metadynamics run (described in the following), once the *cis* isomer had been induced in a given subunit, it was observed as maintained for the duration of the simulation. No restriction was imposed, in any of the four models, on the values of ψ, since its basins are only separated by small barriers (as shown in the following).

The results regarding the four MD models were averaged, considering snapshots every 100 ps, over the last 400 out of 500 ns of simulation and over the five subunits.

3.3 Enhanced Sampling and Free Energy of Proline Isomerisation

Firstly, we aimed at sampling the process of proline isomerisation in the receptor, in order to assess the effect of its environment onto the process. Well-tempered metadynamics allowed for the reconstruction of the free energy landscape associated with the isomerisation in the receptor and in the control proline dipeptide in water, as a function of the two chosen CVs. The free energy surfaces (FESs) as functions of ξ and ψ in the case of proline dipeptide and for Pro281 of subunit 3 in the receptor are reported in figure 3.4, on the left and on the right respectively.

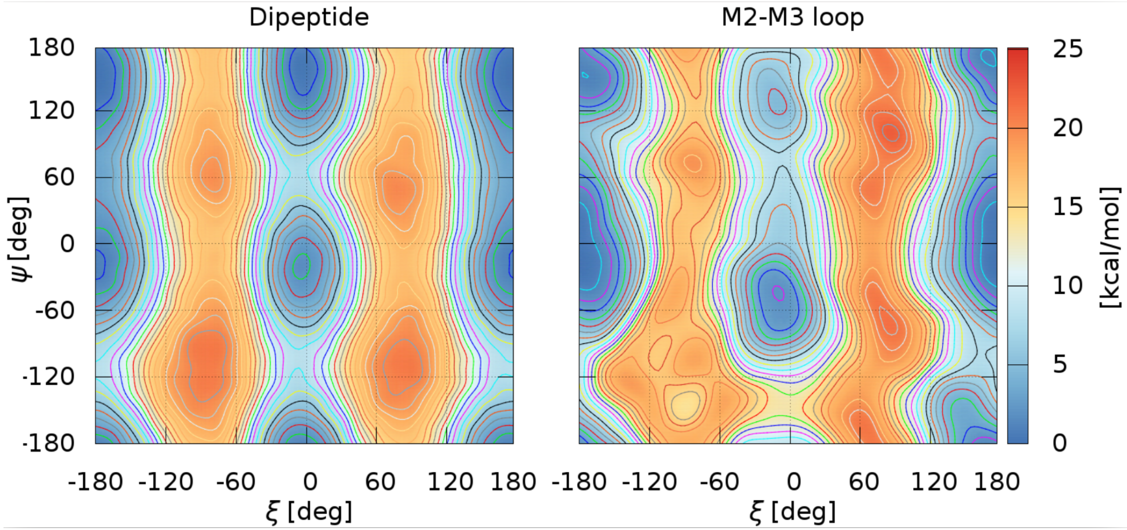


Figure 3.4: Free energy surfaces of the control proline dipeptide in water (left) and of Pro281 in the receptor environment (right) as a function of the torsional angles ξ and ψ. The absolute minimum has been set to zero, and the contour lines are drawn every kcal/mol. Figure from reference [7].

In both cases, the ξ *cis* basins (around 0°) are separated by high free energy barriers from the *trans* basins (± 180 °). In the dipeptide, these barriers are symmetric in both direction of isomerisation, clockwise (left-to-right) and anticlockwise (right-to-left), reaching around 20 kcal/mol in both cases. Compared to other biological processes, this value is significant and would imply that the overcoming of the barrier may be considered as a rare event, which may not occur over shorter time scales compared to those of the gating process in the 5-HT$_{3A}$R (on the millisecond scale). It is also worth noticing the relatively low separation (around 10 kcal/mol) within the same basin for the *cis* and the *trans* isomers of ψ.

In the case of Pro281, the FES gets significantly deformed with respect to that of the control dipeptide in water. Despite maintaining the overall shape, most relevantly in the form of high energetic barriers separating the ξ *cis* basins from the *trans* ones, the FES looses its symmetry. The clockwise barrier reaches higher values, around 25 kcal/mol, while the anticlockwise barrier keep the same height of 20 kcal/mol. Moreover, on this left-hand barrier new saddle points are formed. The centre-top *cis* basin also results higher, implying that ξ-*cis*-ψ-*trans* may be a more convenient metastable state, while a new higher saddle point is formed to divide the two *cis* minima in the counterclockwise direction.

The increase of the right-hand barrier, together with the formation of new saddle points on the other barrier, suggests that there may be a preferential direction of isomerisation in the protein environment (the one that would force the aromatic side chain of Pro281 to move under the M2-M3 loop). The increase in energy of the ξ-*cis*-ψ-*trans* basin may also imply a modification of the isomerisation rate. The differences observed may depend on interactions and/or steric hindrance with nearby sections of the protein. A word of caution should be said regarding the modification of the isomerisation rate, since it cannot be assessed by the FES alone: an evaluation of the real time of barrier crossing could only be reconstructed from the reweighted metadynamics time.[78] A statistics of first passage times should be sampled via several metadynamics runs, and the distribution of the first passage time should pass the Kolmogorov-Smirnov test to verify that it is a Poisson distribution. Future work may focus on this aspect.

3.4 Structural Features and Interactions as Functions of the Isomerisation Collective Variable

The Pro281 isomerisation influences the orientation of the moieties of the adjacent residue, Thr280 (shown in figure 3.5 b), which can form hydrogen bonds with the surroundings via its hydroxyl group, its amino group, and the oxygen of the carbonyl group. A block average, over the simulation time, of the number of hydrogen bonds formed by Thr280 with other parts of the protein was calculated as a function of the isomerisation angle ξ. This provides information on the interactions that this residue is able to form at different orientations, due to the isomerisation process. Intra-protein hydrogen bonds were defined as having a donor-acceptor distance less than 3.5 Å and a donor-hydrogen-accetor angle larger 100° (which is more permissive with respect to the standard cut-off of 120°, in order to better capture interactions with close-by residues, such as those of the M2-M3 loop).

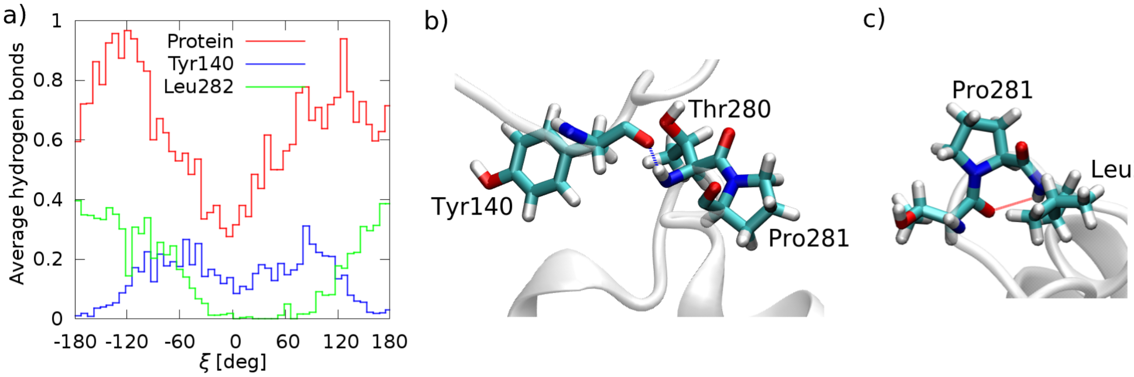


Figure 3.5: (a) Average number of hydrogen bonds formed by Thr280 in the metadynamics simulation with other protein residues (in red), with Tyr140 in the Cys-loop (in blue), and with Leu282 in the M2-M3 loop (in green). (b) and (c) Exemplary conformations and hydrogen bonds for the *cis* and *trans* Pro281 isomers with Tyr140 and Leu282, respectively. Figure from reference [7].

It is important to comment that these hydrogen bonds are evaluated from a metadynamics run, thus they are merely indicative as their lifetimes are not physical. Moreover, a reweighting of the free energy cannot be applied to this quantity in order to properly evaluate their distribution, since hydrogen bonds, for a given time step, can only take discrete values.

The average number of hydrogen bonds between the Thr280 and the rest of the protein is maximum in correspondence of the transition states between *cis* and *trans*. It is interesting to notice how the hydrogen bonds with the protein (red in figure 3.5 a) resemble the shape of the energetic barriers that separate the *cis* isomer from the *trans*. The major contributions to these hydrogen bonds is also different at different values of ξ. Namely, in the *cis* state and in transition states, the major contribution is from the carbonyl oxygen of Tyr140 (figure 3.5 b), which belongs to the Cys-loop in the ECD, while in the *trans* state it is the NH group of Leu282 (figure 3.5 c), which lies the end of the M2-M3 loop, at the top of the M3 helix. The latter is also observed in the dipeptide simulations.

Tyr140 is of particular interest because of its belonging to the Cys-loop, which is part of the ECD and has been previously hinted as relevant for the propagation of the mechanical signal that propagates from the ligand binding pocket.[89] The alleged implication is that the proline isomerisation effect is to alternatively connect Thr280 with different parts of the protein, thus acting as a mediator for the mechanical signal.

Next, we turned our attention to the possible structural modifications that Pro281 may induce in the M2-M3 loop of subunit 3. We calculated the averaged positions and relative standard deviations of the α carbons of its 12 residues, in the x-y plane orthogonal to the protein main axis (shown in figure 3.6). The averages were computed over the frames for which Pro281 is found in *trans* (defined as -180° $< \xi <$ 150° or 150° $< \xi <$ 180°) and in *cis* (defined as -30° $< \xi <$ 30°).

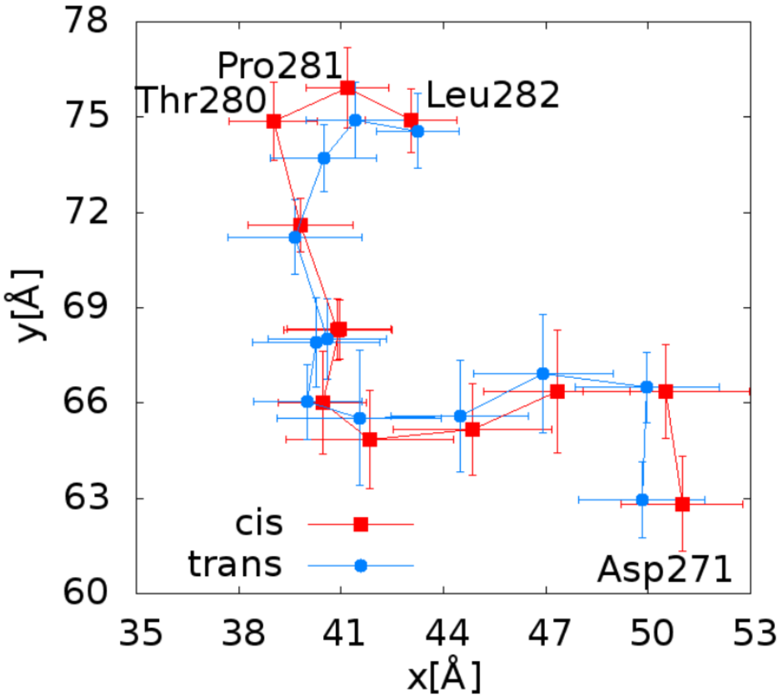


Figure 3.6: Average positions of the α carbons of the M2-M3 loop residues of subunit 3 in the x-y plane orthogonal to the ion channel, in the well-tempered metadynamics simulation, divided for *cis*-Pro281 frames (in red) and *trans*-Pro281 frames (in blue). Figure from reference [7].

As expected, the main sites of structural variations are found at the isomerisation site (Pro281 and Thr280). The next largest variation is observed at the far Asp271, which makes up the top of the M2 helix and may constitute a selectivity filter just before the TMD section of the ion channel. While this structural difference is small compared to the standard deviation, it is worth mentioning that the two Pro281 isomers keep alternating with one another in the metadynamics run, therefore no structural modification is ever to be considered as equilibrated, especially when located this far from the site. It may be possible that this subtle modification can be a consequence of a different steric hindrance (in presence of the two different proline isomers) between the tip of the M3 (approximately where Pro281 is found) and the tip of the M2 (where Asp271 is located). This is supported by the MD runs, which will be described in the following.

3.5 Helix Bend Angle: Reweighting the Free Energy Landscape

To verify if the influence of proline isomerisation extends beyond Asp271, along the M2 helix, we investigated the possibility of a direct connection between proline isomerisation and the change in the bend angle of the M2s, as the literature suggests an involvement of this quantity in the gating process [102]. Thus, we reweighted the FES obtained from the metadynamics run as a function of ξ and of the angle defined by the α carbons of Asp271, Leu260 (the pivotal point) and Arg251, which is found at the other extreme of the M2 helix. The reweighted FES is shown in figure 3.7.

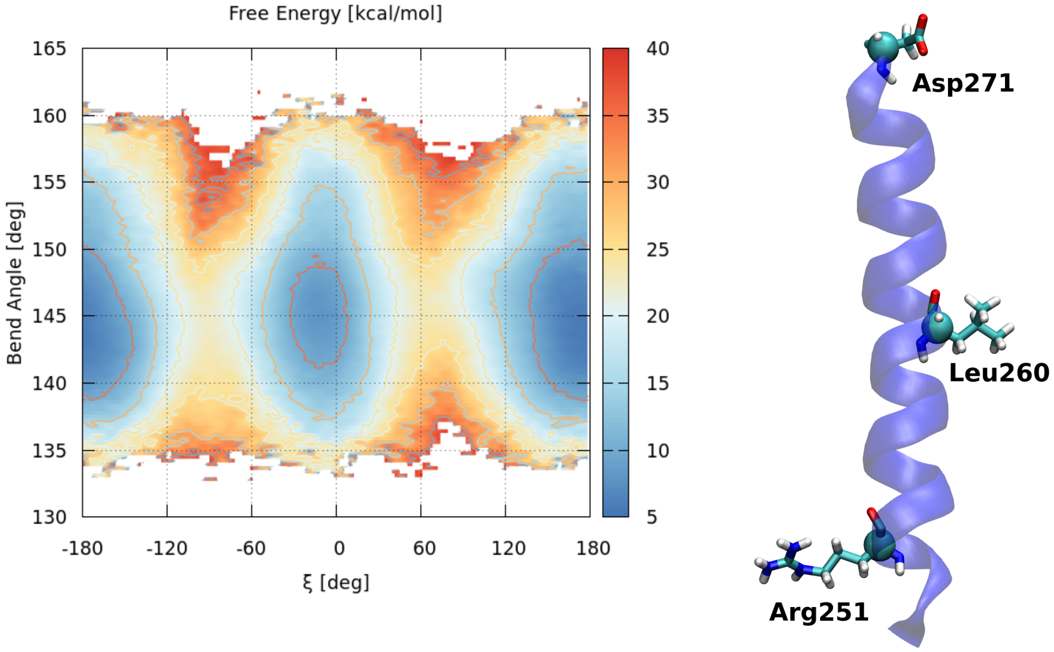


Figure 3.7: Left: Reweighted free energy surface as a function of ξ and of the bend angle of the M2 helix (Asp271 - Leu260 - Arg251) in subunit 3. Right: representation of M2 residues (with their α carbons as large spheres) selected for the definition of M2 bend angle.

In the graph, the bend angle can assume a wide range of values in both isomerisation basins (*cis* and *trans*), thus not providing a proof of correlation between isomerisation and the structural change in the upper part of the helix within a single subunit.

3.6 Stationary Features of the Receptor with Different Proline Isomers

We now turn our attention to the MD productions of the four models previously introduced: 5Trans, 3Cis, 5Cis, and 5Trans5Cis. In order to assess the stability of the models, the root mean square displacement (RMSD) of the protein backbone was calculated with respect to the first NPT step, and averaged over the last 400 ns, for the four models 5Trans, 3Cis, 5Cis and 5Trans5Cis. The values, together with the standard deviations, are reported in Table 3.6.

Model	**RMSD** [Å]
5Trans	1.91 ± 0.08
3Cis	1.94 ± 0.18
5Cis	1.52 ± 0.14
5Trans5Cis	1.73 ± 0.14

Table 3.1: Average root mean square displacement of the four models.

Given that the M2-M3 loop is of particular interest, the root mean square fluctuations (RMSF) of its 12 residues were also calculated, with respect to the equilibrated structure and averaged over the five subunits, as shown in figure 3.8.

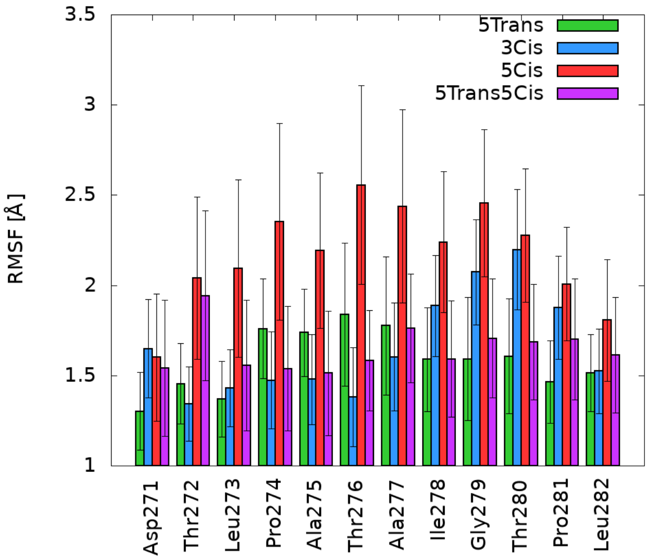


Figure 3.8: Root mean square fluctuations of the backbone of the M2-M3 loop residues. Errors were evaluated with error propagation from the five standard deviations of each of the subunits.

The M2-M3 loop was flexible during the simulations, with the largest fluctuations observed in the 5Cis model.

We calculated the time-average internal ion channel width as a function of the channel axis (shown in figure 3.9), with the zero set at the level of the hydrophobic gate (i.e. the five Leu260 girdle). For this calculation, we made use of the HOLE2 software,[4] with sampling performed every 1 ns over the last 400 ns of simulation.

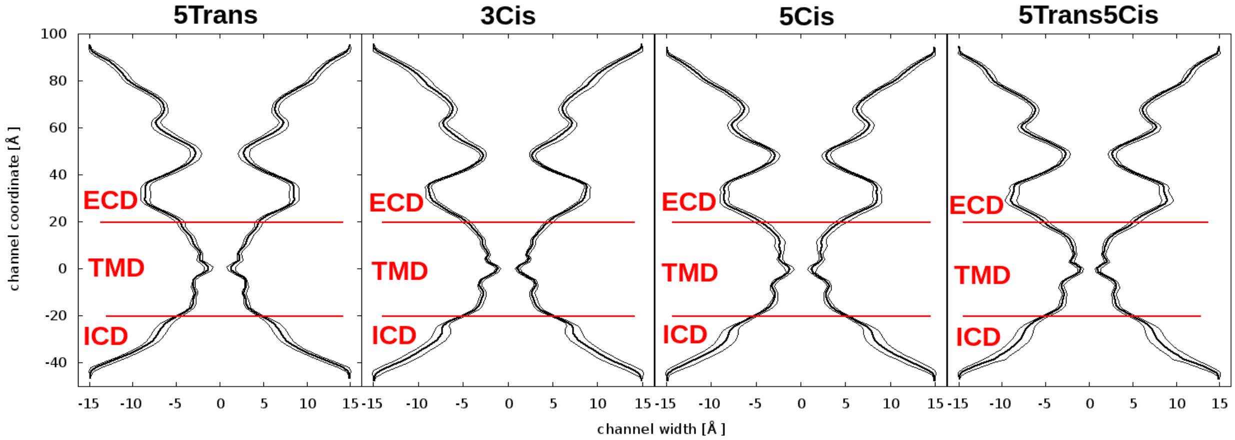


Figure 3.9: Time-averaged channel width profiles of (from left to right) the 5Trans, 3Cis, 5Cis, and 5Trans5Cis models in the MD simulations. The zero of the channel axis is set at the height of the hydrophobic gate (Leu260). The profile is mirrored for ease of visualisation. Figure from reference [7].

The four models do not show any modification at the level of the hydrophobic gate, whose width seems to confirm that the structure remains closed over simulation time and upon isomerisation (given that the initial X-ray structure was deemed as closed as it was crystallised with antagonists, and also because of the measured gate radius of 4.6 Å). This also confirms the results of past simulations on a model based on the same crystal structure.[40] It is interesting to notice, however, that a slight modification at the top of the M2 is observed, namely, a gradual restriction seems to happen at the level of the five Asp271 from 5Trans to 3Cis to 5Cis. This constriction is also featured by 5Trans5Cis.

In order to gain deeper insight into the Asp271 ring, we evaluated the position distributions of the α carbons of the five Asp271 in the x-y plane, orthogonal to the protein axis, shown in figure 3.10 a), together with their average and standard deviations along the x and y directions. Figure 3.10 b) reports the distributions for the five Leu260.

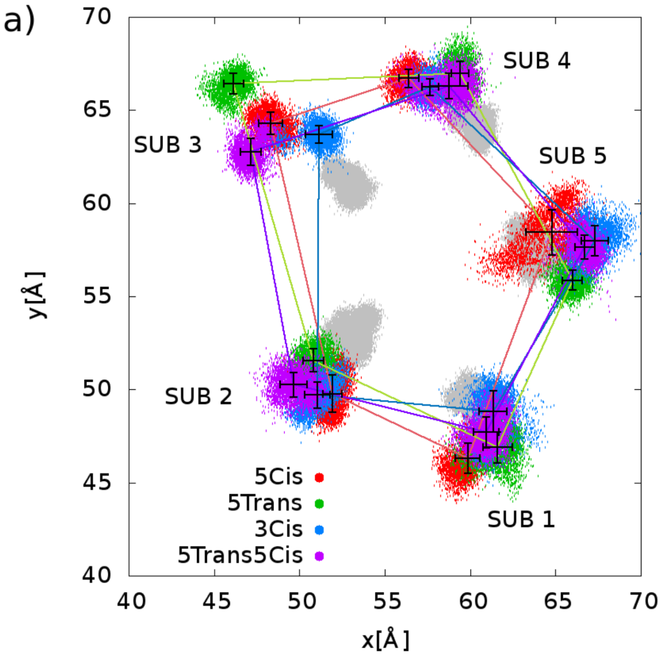


Figure 3.10: Positions in the *x*-*y* plane orthogonal to the ion channel of Asp271 (including average and standard deviation of their distributions) (in green for 5Trans, in blue for 3Cis, in red for 5Cis, and in purple for 5Trans5Cis) and of Leu260 (in grey for all models) in the MD simulations. Figure from reference [7].

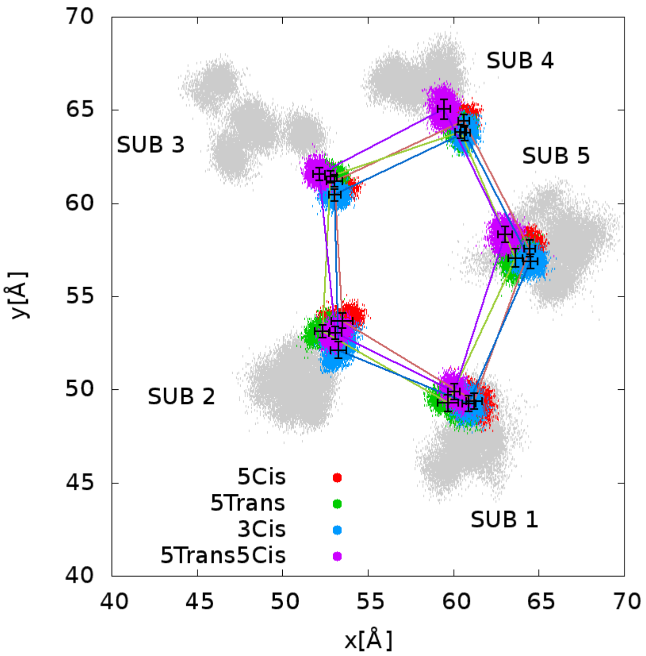


Figure 3.11: Positions in the *x*-*y* plane orthogonal to the ion channel of Leu260 (including average and standard deviation of their distributions) (in green for 5Trans, in blue for 3Cis, in red for 5Cis, and in purple for 5Trans5Cis) and of Asp271 (in grey for all models) in the MD simulations.

The minimisation and equilibration of the crystal structure without any induced isomerisation (i.e., the 5Trans model) results in an asymmetrical selectivity filter at the level of the five Asp271. The other models appear to stabilise in different conformations. Overall, the pentagon area defined by the five subunits is considerably reduced in the other three models. This is mostly due to the fact that in 5Cis, in 3Cis and in 5Trans5Cis the Asp271 of subunit 3 gets in a closer position with respect to 5Trans. The consequent area reduction of the corresponding pentagons make these become more symmetric. It is also worth noticing the more extended spread in subunit 5 for 5Cis. In the three remaining subunits (1, 2, 4), no particular features can distinguish the four models. The distributions of Leu260 exhibit no significant differences.

These results reflect those shown by the channel widths comparison, and the qualitative implications pointed out by the analysis of the M2-M3 loop within the well-tempered metadynamics run. The fact that a mechanism is propagated from Pro281 far away until Asp271, at 10 residues of distance, could be explained by means of interactions and/or contacts between the top of the M2 and M3 helices (while this mechanism is likely to be different for the much closer Pro274).

The reduced area at the level of the Asp271 selectivity filter, upon isomerisation, may lead to an increased electric field, thus to an enhanced attraction of Na^+ towards the entrance of the TMD portion of the channel.

To quantify the effect of the different arrangements of the ring of Asp271 on the attraction of the Na^+ ions, the map of electric potential generated by the Asp271 ring was studied in the four cases: 5Trans, 3Cis, 5Cis and 5Trans5Cis. Such analysis was carried out with the `pmepot` package of VMD,[103] which calculates the value of the electrical potential over every point of a grid defined in the simulation cell (see Appendix A). The value for every point of the map is computed for the last 400 ns of simulation, and then averaged.

Given that the electrostatic potential scales like $1/r$, and given that aspartic acid is a negatively-charged amino-acid due to its side chain oxygens, the electrostatic potential in between this five-amino-acid ring should decrease towards the centre, and the value of the minimum should be affected by the relative distance between the subunits, thus being lower for more constricted selectivity filters. In order to provide a meaningful comparison between the four models, only a section of the potential map was represented as a function of x and y, the coordinates perpendicular to the channel width, for a certain value of $z = z_0$, the parallel coordinate. The value z_0 was chosen, for each of the four cases, in order to correspond to the section were the electric potential register a minimum value. As expected, for all four models, z_0 was found at the level of Asp271.

The results are represented in figure 3.12.

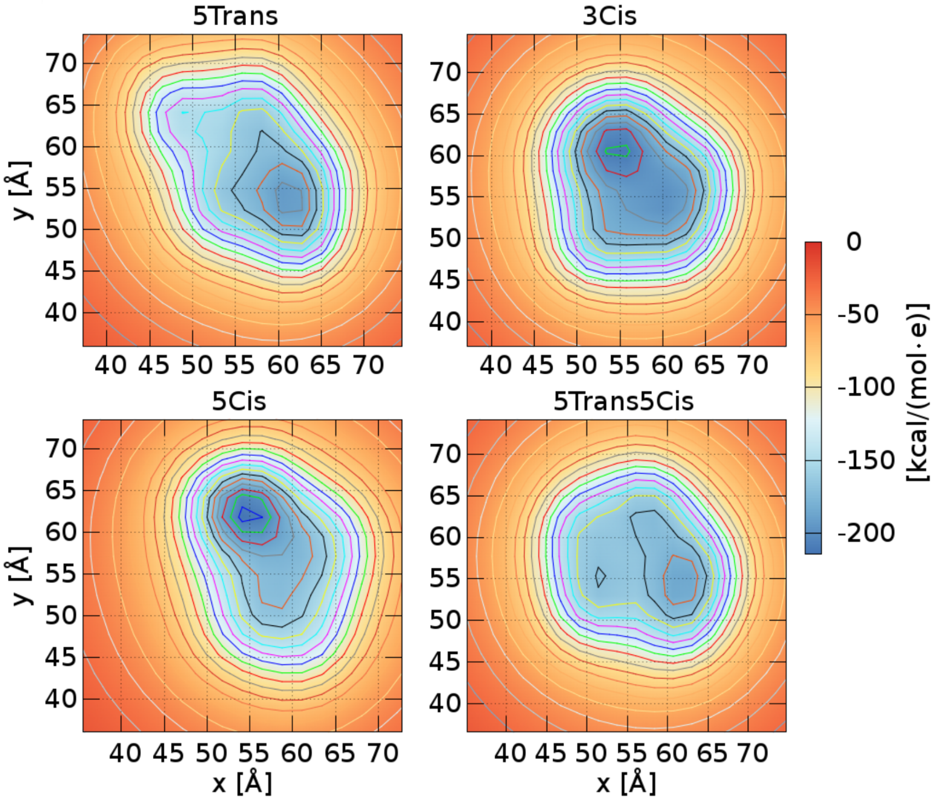


Figure 3.12: Maps of the electrostatic potential generated by the two side chain oxygens of the five Asp271 in 5Trans (top left), in 3Cis (top right), in 5Cis (bottom left), and in 5Trans5Cis (bottom right). The residues themselves lie roughly along the isolines characterised by a potential of -120 kcal/(mol $\cdot$ e). Figure from reference [7].

The minimum for the electric potential is found among the five Asp271 for all models, whose values are shown in Table 3.2.

Model	V_{min} **[kcal/(mol $\cdot$ e)]**	ΔV_{min} **[kcal/(mol $\cdot$ e)]**
5Trans	193.778	0.000
3Cis	212.510	18.732
5Cis	214.323	20.545
5Trans5Cis	186.368	-7.410

Table 3.2: V_{min} and ΔV_{min} (i.e. the difference with respect to 5Trans) for the four models. All values were rounded to the third decimal digit for a cleaner representation.

The resulting sections prove that upon isomerisation of Pro281, the minimum of this potential decreases and shifts towards subunits 3 and 4. The isomerisation of Pro274, as investigated by model 5Trans5Cis, does not affect significantly the map, while slightly increasing the minimum. This stresses the fact that not only the decrease in the area in between the selectivity filter (associated with the isomerisation), but also the asymmetry of the channel can be an important feature associated to the capture of Na^+ ions.

3.7 Conclusions and Future Perspectives

In summary, molecular dynamics and well-tempered metadynamics have been used to shed light on the mutual influence of a potential molecular switch, namely proline isomerisation within the M2-M3 loop, and the surrounding protein environment in the 5-HT_{3A} receptor. With respect to a control system (proline dipeptide) in water, the energy barriers that separate the *cis* and the *trans* isomers loose their symmetry because of steric hindrance and/or interactions with neighbouring residues. The modified energy landscape could hint at an enhanced rate of isomerisation. As past works previously suggested,[89] this proline switch may help pin the Cys-loop in a convenient position that would enable the propagation of the mechanical signal from the ECD ligand binding pocket down to the TMD hydrophobic gate. Our work seemed to confirm this, revealing how Pro281 is able to influence the possibility of Thr280 to form hydrogen bonds with Tyr140 of the Cys-loop. At the same time, Tyr140 side chain movements can hinder or clear the path for Pro281 isomerisation, explaining the altered free energy landscape of this process.

The fact that the five Pro281 are found in *trans* in all the currently known 5-HT_{3A} receptor structures suggests that the two isomers (*cis*, *trans*) may not be univocally associated with the conductive state of the channel. Instead, Pro281 isomerisation may be a kinetic process, with the interchange between *trans* and *cis* allowing for, or removing, a mean of connection between the M2-M3 loop and the Cys-loop.

We could speculate that the effects described, including interface loops properties, the isomerisation free energy barriers, the *cis* isomer stability, the rate of isomerisation, and the network of interactions/contacts between the M2-M3 loop and the Cys-loop, could be further influenced by the perturbation induced by ligand binding in the ECD, as well as by a relative twisting of the ECD with respect to the TMD, as suggested for other pLGICs.[102] In fact, by using an apo-structure (the only one available of the 5-HT$_{3A}$R when this work was started), the investigation we carried out could not, in principle, reveal the mechanistic connection between ligand binding and proline isomerisation. Such connection is, anyway, expected to occur over long time scales, and enhancing its sampling would require a separate investigation.

While proline isomerisation did not appear to be able to influence the hydrophobic gate in the investigated time-scale, a subtle effect was observed at the level of the selectivity filter ring made up by the five Asp271, which acts as an attractor for the Na$^+$ ions at the top of the M2 helix. This effect was hinted at in both MD and metadynamics simulations Interesting enough, a constriction at the level of the Asp271s is also observed in the cryo-EM structure 6DG8, despite having the five Pro281 in the *trans* isomer (figure 1.11).[5]

While our simulations clearly proved that the Asp271 ring assumes different shapes in the four models, more simulations (e.g. more replicas of each model), or more sophisticated techniques for dynamical analysis (for example by studying signal transmission via information-theoretic analysis of amino-acid dynamics[104]), would be needed to unambiguously relate this to proline isomerisation: in fact, pentagonal symmetry/asymmetry may be,in principle, the cumulative result of various factors (like the flexibility of the M2-M3 loop, or interacting nearby lipids). Moreover, while kinetics calculation have not yet proved if proline isomerisation in the M2-M3 loop is a kinetic process, rather than a thermodynamic one, we cannot yet assess the time scales over which a single isomer of Pro281 (or Pro274) can propagate mechanical modifications to the top of the M2 helix.

As previously said, other works hinted at the possibility that the effect of proline at position 8* can be rescued in a variety of ways. In fact, charged residues substituted at position 8* may promote hydrogen bonding with Ala275.[91] Histidine or tryptohan mutations at position 8* preserve channel function in the 5-HT$_{3A}$R,[90] while Pro8* equivalent in the nAChR may play a role due to features other than isomerisation.[92]

Alternatively, lipids in the site may connect the two loops via hydrogen bonds (more on this in the next chapter). Thus, proline isomerisation within the M2-M3 loop most likely constitutes one of many cumulative and/or alternative phenomena that can occur to instate a bridge between the M2-M3 loop and the Cys-loop.

Overall, the work presented in this chapter helped understand the role of key prolines within the

M2-M3 loop, whose isomerisation involvement in protein function was pointed at by mutagenesis electrophysiology experiments. Our results may prove of more general interest, given the presence of Pro8* in the nACh receptors, and that of Pro1* (our Pro274) in other pLGICs as well.

Future investigations may approach the enhanced sampling of Pro281 isomerisation with the goal of reconstructing the kinetics of this process. This would require, rather than a long metadynamics run, more replicas of first barrier crossing, for every possible pair initial basin-end basin.

Moreover, future studies may regard Pro281 isomerisation in a different conductive state of the 5-HT$_{3A}$R, contributing to shed light on how different conformations of the receptor may affect and be affected by this process.

Other links between the M2-M3 loop and the Cys-loop can be established in the 5-HT$_{3A}$R: n the next chapter, we will focus our attention on how lipid molecules can bind and interact with these structures, and how these binding events are affected by membrane composition.

Chapter 4

Lipid Membrane Composition Influence on Lipid-Protein Interactions in the 5-HT$_{3A}$ Receptor

4.1 Introduction

While ligand binding into the ECD orthosteric pocket remains the ultimate trigger for channel gating, growing evidence supports the fact that the lipid membrane surrounding pLGICs may contribute to the overall mechanisms, by means of interactions at the membrane-protein interface, as found for many other membrane proteins.[105, 106]

Systems investigated in this context include the nicotinic acetylcholine receptor, which has so far been the most studied among pLGICs,[107, 108, 109, 110, 111, 112, 113] the 5-HT$_{3A}$ receptor,[42] the GABA$_A$ receptor,[114] the Glycine receptor,[115] the prokaryotic *Erwinia chrysantemi* (ELIC) ion channel[116, 117, 118] and *Gloeobacter violaceous* (GLIC) receptor,[119, 120] G-protein coupled receptors (GPCRs),[121, 122, 123, 124] mechanosensitive channels,[125], amyloid precursor proteins,[126] heterotrimeric G proteins,[126, 127] and others.

However, the mechanisms behind lipid modulation in pLGICs are still not fully understood,[106], thus requiring additional investigations. Since state-of-the-art realistic models of the "average" human membrane contain about 60 lipid species,[128] modelling the cell membrane is a complex task. While progress has been made for modelling it in a realistic way at different resolution levels, from atomistic to coarse grained,[129] even relatively simple models, with more than one lipid species, pose challenges for state-of-art simulations and can provide useful preliminary information.

Phospholipids and cholesterol (CHL) are among the major components of animal cell membranes. In particular, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) are prevalent phosphatidylcholines and phosphatidylethanolamines, respectively. POPC, POPE and CHL are shown in figure 4.1.

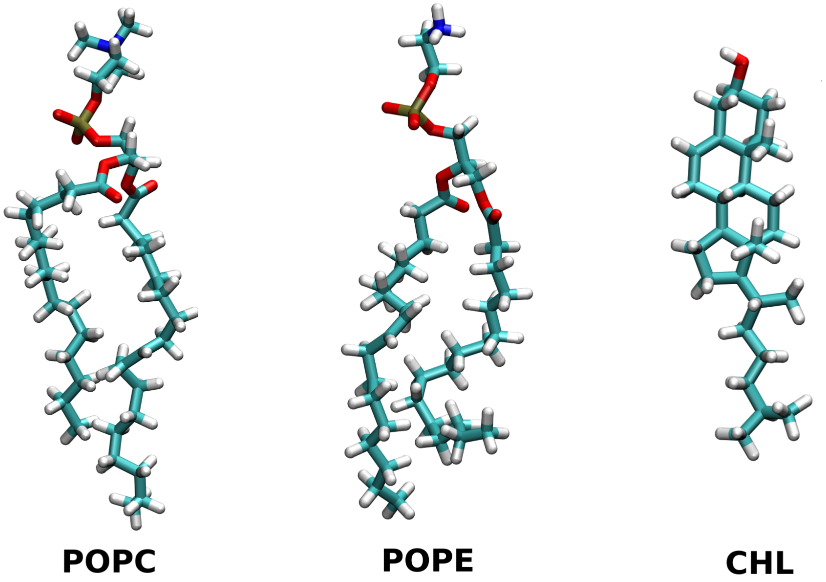


Figure 4.1: Representation of POPC (left), POPE (centre), and cholesterol (CHL, right) molecules. Figure from reference [3].

POPC is a neutral phospholipid consisting of two hydrophobic tails, joined to a hydrophilic head group, a choline. POPE only differs from it in the head group, which consists of an ethanolamine. POPC and POPE have therefore different abilities to form hydrogen bonds. Among others, this has been demonstrated to affect the structural conformation of the mycobacterium tuberculosis mechanosensitive channel, by means of molecular dynamics simulations of such protein embedded in POPE membranes, manually turned into POPC during the simulation.[125] A pure-POPC membrane has also been proved, with single-particle cryoelectron microscopy, to decouple the ECD and the TMD of ELIC, preventing the mechanical signal induced by agonist binding to propagate from the former to the latter.[116] In the same study POPC heads were observed, in molecular dynamics simulations, to form interactions with the protein which in turn would affect the closed-open-desensitised equilibrium constants.[116] The binding of the glycerophospholipid POPG to ELIC was proved responsible for the stabilisation of its open state with respect to the desensitised state, by means of native mass spectrometry, coarse-grained simulations and functional assays.[117] The head group of phospholipids was proved able to affect agonist and antagonist binding and receptor activation in a GPCR, by means of receptor reconstruction in high-density-lipoparticles.[124] A comparison of the X-ray structures of the *Caenorhabditis elegans* glutamate-gated chloride channel (GluCl) in apo-state and in POPC-bound state demonstrated how lipids can potentiate agonist binding and promote an expanded open-like conformation of the transmembrane domain.[29] Mutagenesis and electrophysiology experiments and kinetic analysis revealead how interactions between the arginines R64 and R89 and the anionic phosphatidic acid (PA) bound to the non-annular (i.e. in between transmembrane helices) lipid binding sites of the K^+-channel KcsA reduces inactivation in macroscopic current recordings, causes longer open channel lifetimes and decreases closing rate constants; moreover, molecular dynamics simulations revealed that R64 and R89 mediate persistent binding of PA to the non-annular sites.[130] Umbrella sampling simulations were used to assess the effect of anionic phospholipids 1-palmitoyl-2-oleoyl-phosphatidylserine (POPS) on ion translocation in the bacterial KcsA, determining that anionic lipids do not affect the energetic barriers of K^+ ions through the selectivity filter, and that the occupancies of each of the selectivity filters are influenced by the presence and identity of lipids in between transmembrane helices.[131]

A membrane composition comprising both POPC and POPE (with the addition of cholesterol) constitutes a valuable minimal choice, and has been used in a large variety of *in-silico* works.[121, 132, 125, 133, 122, 42, 120] However, other minimal compositions can be chosen in principle, for example by considering phospholipids with the same head group but different hydrophobic tails, e.g. POPC and SDPC.[42] Alternatively, more complex compositions, possibly requiring a coarse-grained simulation method, can be employed:[115] in fact, a larger number of lipid species would require larger membranes in order to capture a meaningful sampling of lipid-protein interactions for each of the present species, significantly slowing down the computation of an all-atom molecular dynamics simulation.

Cholesterol exhibits a much shorter structure with respect to POPC and POPE, comprising four aromatic rings, an aliphatic tail and a hydroxyl group head. The importance of this particular lipid species comes from its high concentrations in membranes of human brain cells.[134, 135]

Cholesterol modulates membrane viscosity, affecting the lateral diffusion of all the lipids making up the membrane.[136] It has typical preferred binding sites within membrane proteins, referred to as “Cholesterol Recognition/interaction Amino acid Consensus sequence”, or “CRAC” domains, and as “inverted CRAC” (“CARC”) domains.[137, 107, 138, 139] The sequence of CRAC domains is defined as (L/V)-X_{1-5}-(Y)-X_{1-5}-(K/R) in the N-terminus to C-terminus direction, where L/V is either a leucine or a valine, X_{1-5} is any number between 1 and 5 of any amino-acid, Y is a tyrosine, and K/R is either a lysine or an arginine. Instead, the sequence of CARC domains is defined as (K/R)-X_{1-5}-(Y/F)-X_{1-5}-(L/V), from the N-terminus to the C-terminus.

Cholesterol can affect membrane proteins in a variety of ways. For example, in the G-protein coupled 5-HT_{1B} receptor, it interacts with a transmembrane helix, thus modifying the serotonin binding mode in the extracellular pocket, as revealed by MD simulations.[133] Microseconds-long simulations of the 5-HT_{3A} receptor, embedded in a membrane with phospholipids POPC and SDPC, proved that cholesterol was able to intercalate within one subunit of the TMD during conformational changes upon serotonin binding, contributing to the stabilisation of a suggested preactive state (a conformation which resembles an initial stage of channel opening), suggesting its necessity with respect to pure-POPC membranes.[42] *In-silico* investigations on the adenosine type 2A receptor found that cholesterol binding affect the binding poses of caffeine, and its mobility in the pocket.[132]. Other *in-silico* studies showed that cholesterol binding may promote pore opening in the $GABA_A$ receptor,[114] while in nicotinic acetylcholine receptor it allows for contacts between the ligand-binding domain and the pore, thus proving necessary for structural preservation.[109]

Moreover, there has been much suggestion that the outermost M4 helix of pLGICs can act as a lipid sensor. For example, in the nicotinic acetylcoline receptor, experimental characterisation suggested that M4/Cys-loop interactions may be favoured by lipid-dependent properties.[112] Other experiments on this pLGIC showed that the thickness of phosphatidylcholine membranes affects transitions between uncoupled and coupled conformations.[113] Mutagenesis experiments proved how the M4 helix in ELIC contains a lipid-binding site, whose disruption accelerates desensitisation.[118]

In general, there is consensus that various lipids, in particular phospholipids and cholesterol, can affect the function of different membrane proteins, including pLGICs, thanks to their ability to bind to them, and in this way can mediate between different protein domains, or alter their properties.

In this chapter, we continue our investigation on the 5-HT$_{3A}$ receptor, focusing on the characterisation of the lipid-protein interface, and enquiring on the role of membrane composition on its features. This study was performed by means of all-atom MD simulations, with a pure-POPC membrane and two mixed membranes of increasing viscosity, that also include POPE and cholesterol. The work presented in this chapter was published in Biointerphases in 2020, appropriate references are made where relevant.[3]

4.2 Three Models with Different Lipid Membrane Composition

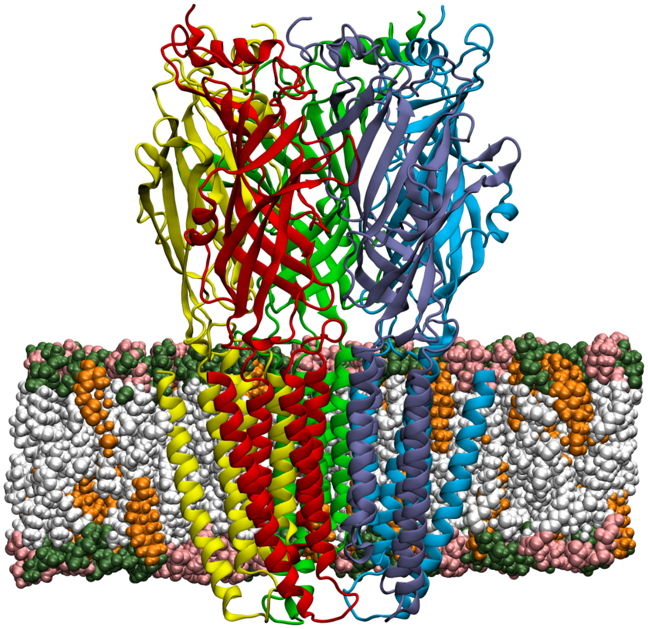


Figure 4.2: The model of 5-HT$_{3A}$ receptor embedded in the mixed membrane of model B (35% POPC, 35% POPE and 30% CHL). POPC heads are represented in pink, POPE heads in dark green, phospholipids tail (common to POPC and POPE) in white, and cholesterol in orange. The five subunits are coloured in anticlockwise order in red, yellow, green, light blue and blue. Figure from reference [3].

For this project, the model of the 5-HT$_{3A}$R described in the previous chapter has been re-introduced and embedded into three different membrane models, labelled A, B and C: model A features a membrane that contains only POPC, the one of model B contains 35% POPC, 35% POPE and 30% cholesterol, and the one of model C contains 25% POPC, 25% POPE, and 50% cholesterol. The resulting membrane areas were around 125 Å by 128 Å for model A, 122 Å by 123 Å for model B, and

121 Å by 122 Å for model C. Model A is the least complex, as a pure-POPC membrane constitute a standard choice previously simulated.[40, 43, 7] Model B is characterised by an intermediate composition where cholesterol is present, but the phospholipids are still predominant. Its 6:7:7 concentration has been used in past simulations of membranes with cholesterol (6), POPC (7) and a third lipid (7) for the study of serotonin receptors.[42, 140] Model C contains a higher percentage of cholesterol, that represents the upper CHL concentration in synaptic membranes (whose cholesterol content varies between 33% and 50%),[135, 132, 134] and would also be useful for enhancing the sampling of cholesterol-related phenomena. These three models are expected to exhibit different viscosity of the lipid membrane, increasing from A to C.

The systems were solvated in an orthorhombic supercell, with a 15 Å buffer of TIP3P water and 0.15 M of Na^+ and Cl^- ions that reproduce physiological conditions. The final 5-HT_{3A}R models contained, in total, 183,207 atoms (model A), 182,877 atoms (model B), and 175,904 atoms (model C).

The systems were simulated with the NAMD 2.9 MD package,[80] making use of the Amber `ff14SB`[55] and `LIPID14` force-fields.[59]

The simulation time step was 2 fs, and the bonds containing hydrogen were constrained using the SHAKE algorithm. Particle Mesh Ewald was employed for the electrostatic interactions, and a cut off of 10 Å was used for the non-bonded interactions.

The three models initially underwent an energy minimisation procedure, then a slow heating up to 310 K and a partially restrained equilibration, i.e. with the protein α carbons restrained and the lipids free to diffuse. Such equilibration stage lasted for 100 ns, as prescribed for the equilibration of lipid membranes, whose diffusion typically occurs over times of the order of tens to hundreds of nanoseconds.[141, 142] The restraints were subsequently released.

Next, an unrestrained simulation was performed within the isothermal-isobaric NPT ensemble for each model at 310 K, which is above the gel transition temperature for the three lipid species considered,[143, 144] and at a pressure of 1 atm. The temperature was maintained by making use of a Langevin thermostat with a collision frequency of 1.0 ps^{-1}, while the pressure was controlled by means of a Langevin piston barostat with an oscillation period of 200 fs and a damping time constant of 100 fs.

For each of the three models, two replicas (with different random number generator seeds for the Langevin dynamics) were carried out for the unrestrained NPT simulation, one of 500 ns (RX), and one of 250 ns (RY). This choice was made since lipids diffuse differently in separate replicas, thus allowing to explore additional statistics of the processes of both lipid diffusion and lipid-protein interaction. For both replicas, the first 50 ns were considered as unrestrained protein equilibration, thus the production would last 450 ns for RX and 200 ns for RY. Some of the analysis were performed over the conjunction of the 50-to-250 ns time-windows of RX and RY (RXY400), and over the time window 50-to-450 ns of replica RX (RX400). Comparing RXY400 and RX400 allowed us to test whether these two different approaches (namely, one long run or two shorter ones) would yield qualitatively similar results in terms of time- and subunit-averaged quantities.

Those analyses which explicitly express an observable as a function of time were carried out over the 50-to-500 ns window of RX (RX450) and over the 50-to-250 ns window of RY (RY200). This comparison instead provided information on how the observed phenomena and events depend on the specific NPT realisation, obtained with stochastic thermostat and barostat.

Trajectories were sampled every 50 ps and analyzed with the Cpptraj[145] and MDAnalysis[146, 147] software. Hydrogen bonds were defined by a donor-acceptor distance smaller than 3.5 Å and by a donor-hydrogen-acceptor angle larger than 120°, as described in Appendix A alongside other typical quantities.

4.3 Protein and Membrane Dynamical and Structural Features

First, we evaluated the stability of the protein, by calculating the RMSD of the backbone atoms over the unrestrained NPT trajectory for different sections of the protein, namely the TMD, the ECD, the protein as a whole, and for each of the TMD helices (M1, M2, M3, M4). Figure 4.3 reports the average RMSD over the production run for RXY400 and RX400, where the standard deviations represent the fluctuations around the averages. The sequence of the protein sections was defined as in reference [2].

The three models were stable in both replicas, as seen in the two averaging methods RXY400 and RX400, which produced similar results. The variability observed among the five subunits reflects different statistical sampling within the NPT ensemble, alongside different lipid environment. Being the outermost helices, only being covalently linked to the rest of the protein thanks to the solvent-exposed M3-M4 loop, and being exposed to the fluid lipid membrane, the M4s moved more than the other three helices, with the largest fluctuations mostly due to their terminals that extend outside the membrane.

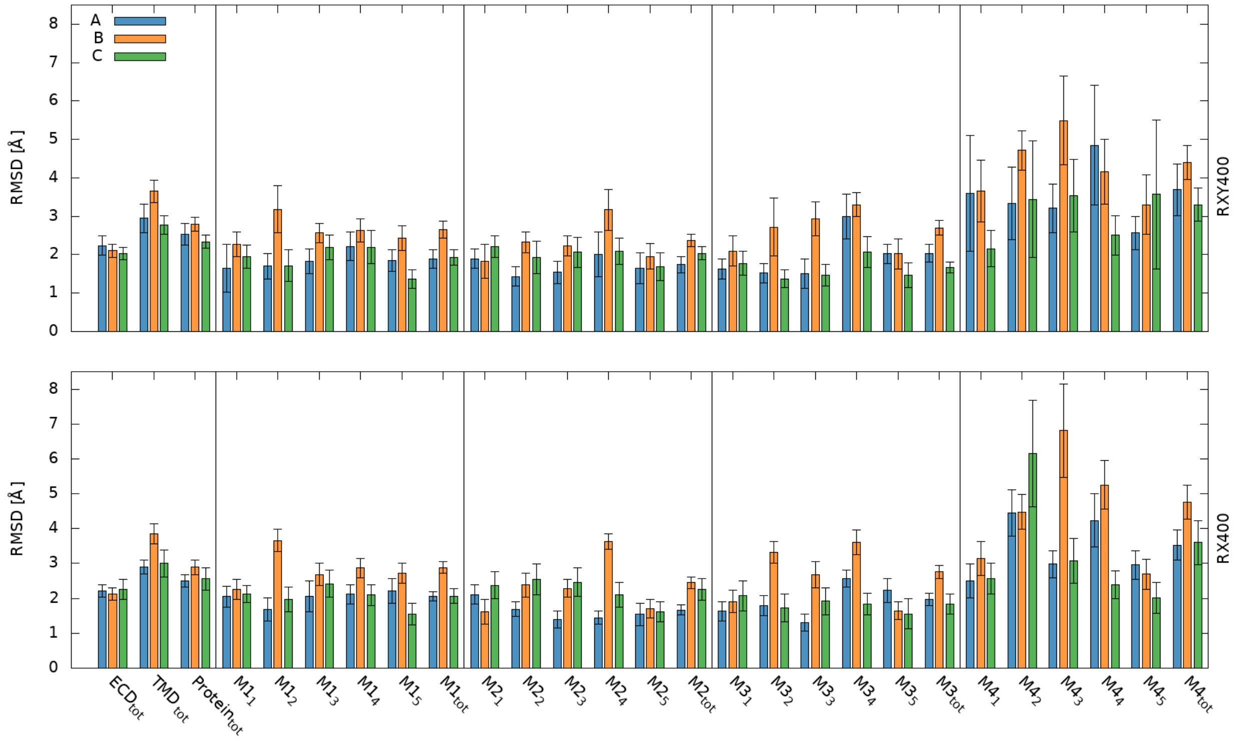


Figure 4.3: Average RMSD, calculated for the unrestrained NPT production runs, of the 5-HT$_{3A}$ receptor backbone, the ECD, the TMD, the whole protein, and the four helices M1, M2, M3 and M4. These quantities are shown for the three models A (blue), B (orange) and C (green), for RXY400 (left) and RX400 (right), for each subunit and averaged over the five of them. Figure from reference [3].

The lipid membranes of three models were characterised by the expected relative fluidity, as proved by the lipid lateral mean square displacements (MSD), shown in figure 4.4: the membrane of model C, having the highest content of cholesterol, is the most viscous, while the one of model A is the least viscous. The lateral diffusion coefficients were calculated via Einstein's equation:

$$2nD_L = \lim_{t\to\infty} \frac{MSD}{t} \tag{4.3.0.1}$$

where n is the number of dimensions (2, for a lateral diffusion), t is the time, and D_L is the lateral diffusion coefficient. In the simulations, the lateral diffusion coefficients were calculated for each model and each replica by linear regression over different time windows of 50 ns, 100 ns, and 200 ns, shifted of 10 ns from each other. These are defined after the first 10 ns (in order to discard the initial non-linear transients), and end at half simulation time (250 ns for RX, and 125 ns for RY). The values of the lateral diffusion coefficients are of the order of $\mu m^2 s^{-1}$, in agreement with previous calculations of lipid systems.[108, 148, 149] However, their absolute values are affected by the usage of Langevin dynamics, the small size of the membrane, the limited time-scale explored, and the presence of the embedded protein.

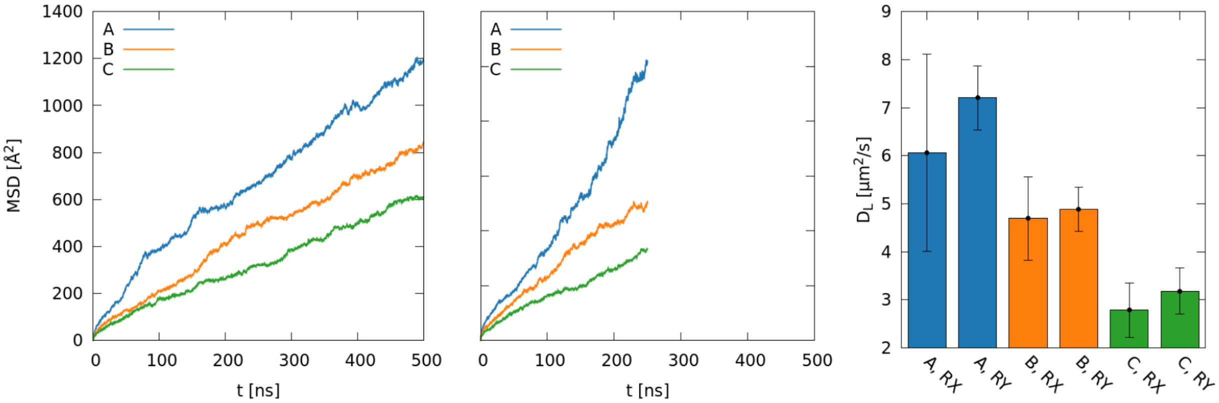


Figure 4.4: Lateral mean square displacement (MSD) of lipids as a function of time over the unrestrained NPT simulations (the last 500 ns for RX, on the left, and the last 250 ns for RY, in the centre), for model A (blue), B (orange) and C (green). The averages and standard deviations of the lateral diffusion coefficients D_L are reported in the graph on the right. Figure from reference [3].

The average channel width along the channel axis, calculated with the software HOLE2,[4, 150] is shown in figure 4.5: it proves that the constriction in the TMD at the level of the Leu260 residues within the M2 helices, that act as a hydrophobic gate for the ions,[14, 15] resembles a closed channel in all three models, thus the hydrophobic gate average diameter is not affected by the composition of the lipid membrane. This results from both RX400 and RXY400 averages, and is consistent with previous studies based on the same crystal structure of the 5-HT$_{3A}$R in a pure-POPC membrane, including the work described in the previous chapter (performed at different temperature), which also reported a close state of the channel.[40, 7]

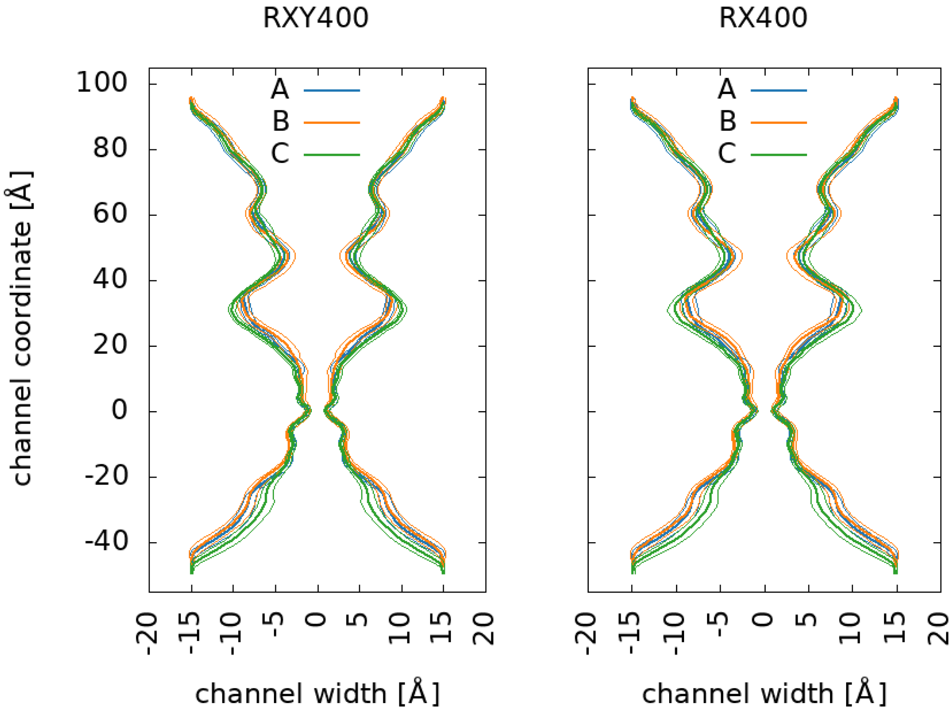


Figure 4.5: Average channel width (thick line) with its standard deviation (thin lines) as a function of the channel axis for RXY400 (left) and RX400 (right) and in the three models: A (blue), B (orange) and C (green). The zero of the channel axis is set at the height of the hydrophobic gate (Leu260).

The channel width along the channel axis is similar for the three models, over the explored time window: any difference in the ECD and TMD is small compared to the standard deviation. The only notable difference is found in the lower part of the channel, where model C presents a narrower channel. However, being the ICD substituted with an artificial linker, no significant conclusion can be drawn.

A previous *in-silico* investigation on the same crystal structure revealed how ions can enter the receptor through ion portals at the interface between the ECD and the TMD.[40] Since this part of the protein is close to the membrane, lipids could in principle play a role onto these portals as well. Therefore, we studied the matrix of distances of the α carbons of the residues that had been previously established as those making up the portals, at the interface between adjacent subunits. The matrices for the three models, averaged over the five subunits, are shown in figure 4.6 for RXY400 and for RX400.

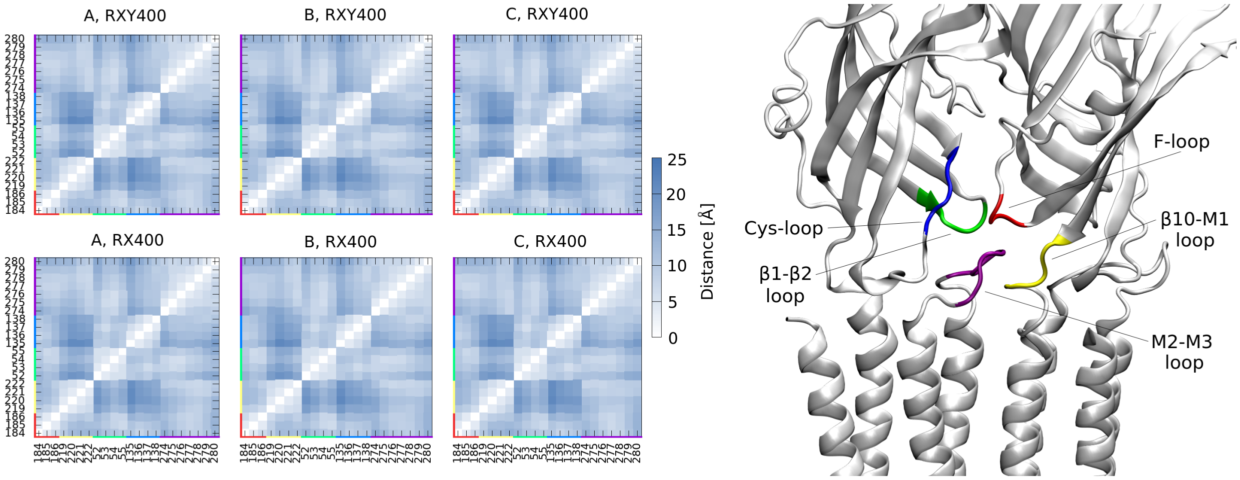


Figure 4.6: Average inter-residue distances for models A, B and C for the residues belonging to the β1-β2 loop (green, in the structure on the right showing two adjacent subunits), the F-loop (red), the β10-M1 loop (yellow), the Cys-loop (blue) and the M2-M3 loop (purple). The matrices are calculated for RXY400 (top) and for RX400 (bottom).

The distance matrices only exhibit minimal differences between the models, which are not significant with respect to thermal fluctuations. Therefore, we conclude the lateral portals are not influenced by the composition of the lipid membrane, over the investigated time windows. This holds for both RXY400 and RX400.

4.4 Lipid-Protein Interactions

Lipids proximity to the helices of the TMD is influenced by bulk properties of the membrane, such as its viscosity and the pressure it exerts on the receptor, and by lipid-protein interactions, including hydrogen bonds, hydrophobic effects, van der Waals and electrostatic interactions, and, in the case of cholesterol, π interactions (like π-π interactions and cation-π interactions). As the lipids diffuse in a stochastic way, thus resulting in a random distribution around the receptor TMD, a different local lipid environment is expected around each subunit of the protein. This can be considered an increasing of the sampling.

The number of hydrogen bonds, averaged over time and over the five subunits, is shown in figure 4.7, grouped for every lipid species and protein section: the four helices M1, M2, M3 and M4, the M2-M3 loop, and the Cys-loop. Errors were evaluated with error propagation from the five standard deviations of time-averaged data for each of the single subunits.

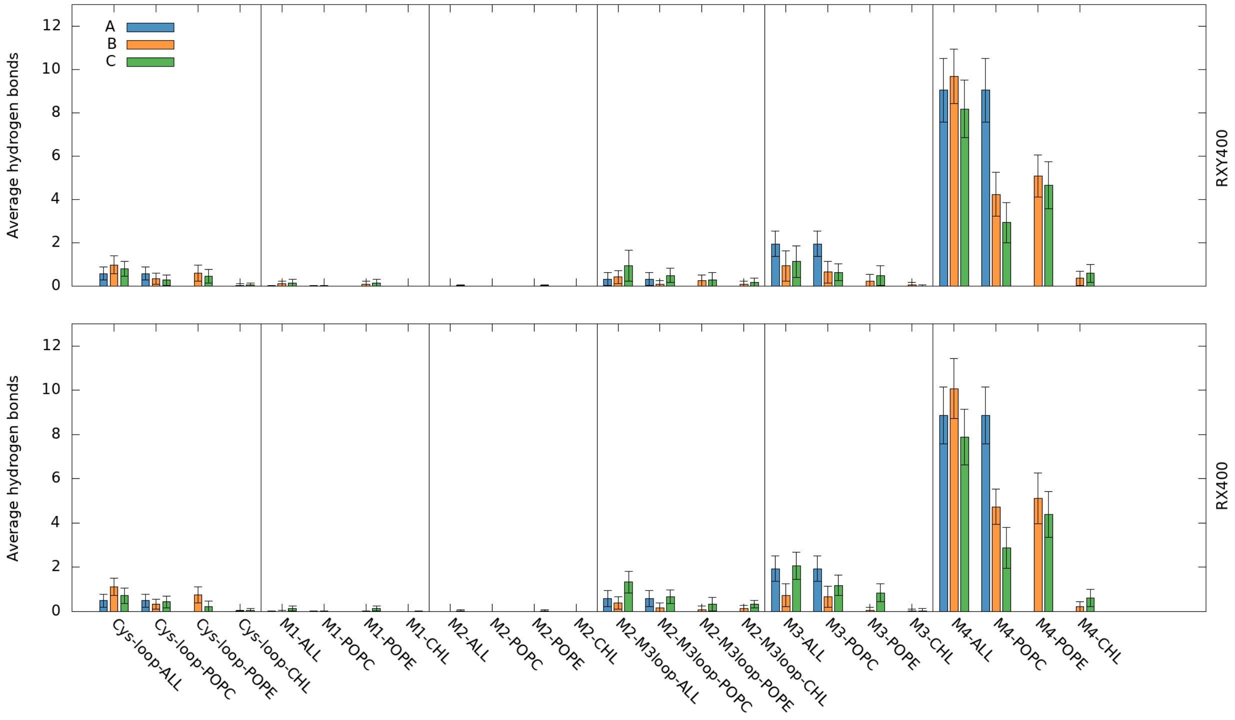


Figure 4.7: Average number of hydrogen bonds between receptor and lipids for RXY400 (top) and for RX400 (bottom), and for the three models: A (blue), B (orange), and C (green). Figure from reference [3].

On one hand, no major differences are observed neither among the three models nor between the averaging methods, on the other hand different interesting features can be noted regarding the way hydrogen bonds are formed between lipids and the protein. Most hydrogen bonds are formed with the M4s, which are the outermost helices. The M3s and M1s form a lower number of hydrogen bonds with lipids (this is truer for M1s), as both are found at the interface between adjacent subunits and thus are less exposed to the membrane. The hydrogen bonds of M4 with POPC are more frequent in model A as expected, because of its higher concentration in such model. Model B, on the other hand, exhibits the largest average number of hydrogen bonds with M4, due to the higher hydrogen bonding ability of POPE. Phospholipids are able to reach out for the ECD, interacting with the Cys-loop, and for the M2-M3 loop, both of which have been suggested as crucial for the receptor activation mechanism.[89, 81, 7] Within the M2-M3 loop, two residues are of particular interest: Thr280, that flanks Pro281, whose *trans-cis* isomerisation may be part of the gating mechanism as described in chapter 3,[89, 81, 7] and Leu273, that flanks Pro274, which is highly-conserved within the Cys-loop superfamily[25] and could also play a role in the gating mechanism thanks to of its interaction with the β1-β2 loop.[29, 25, 102, 151, 152]

While the large error bars reflect the variability among the five subunits, the average number of hydrogen bonds with the M2-M3 loop for any lipid species is larger for model C. The average number with the Cys-loop is smaller in model A compared to both B and C (in this case, A dominates in the sole case of POPC alone as expected). When comparing the two phospholipids species, POPC and POPE, the former appears to form more hydrogen bonds with the M2-M3 loop with respect to the later, while the opposite occurs for the Cys-loop. A phospholipid molecule in close proximity with both the Cys-loop and the M2-M3 loop may form alternate, or even simultaneous hydrogen bonds with them. This may provide a bridge between these two key protein segments. This is shown for a representative MD snapshot in figure 4.8. Cholesterol appears to be less able to form hydrogen bonds with both the ECD and the M2-M3 loop, with respect to phospolipids, possibly due to its shorter length.

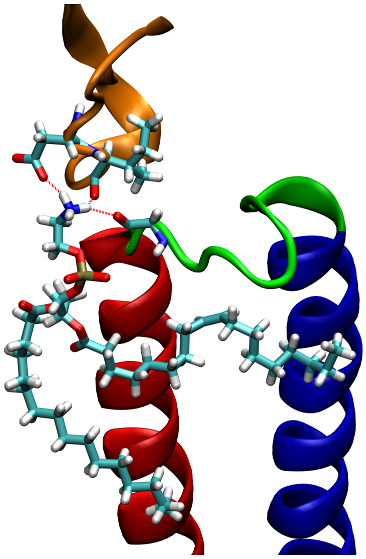


Figure 4.8: A POPE molecule forms hydrogen bonds with residues of the M2-M3 loop (Gly279) and of the Cys-loop (Asp138, Ile139) in an MD frame in model B, replica RY. Figure from reference [3].

As we said, interactions and membrane properties can promote proximity between molecules. Proximity between specific atoms and/or residues can be studied by means of atomic contacts (see Appendix A), for which we used here a cut-off of 3.5 Å in agreement to previous studies.[153] The average number of contacts between a given helix and a given lipid species (e.g. M1-POPC) or two given helices (e.g. M1 and M3) were calculated for the five subunits, and for the three models, for a total of 15 values for each lipid species-helix (or helix-helix) pair. Their averages over the five subunits, shown in figure 4.9, can be compared with those of hydrogen bonds. Errors were calculated via error propagation from the five standard deviations of time-averaged data for each of the single subunit.

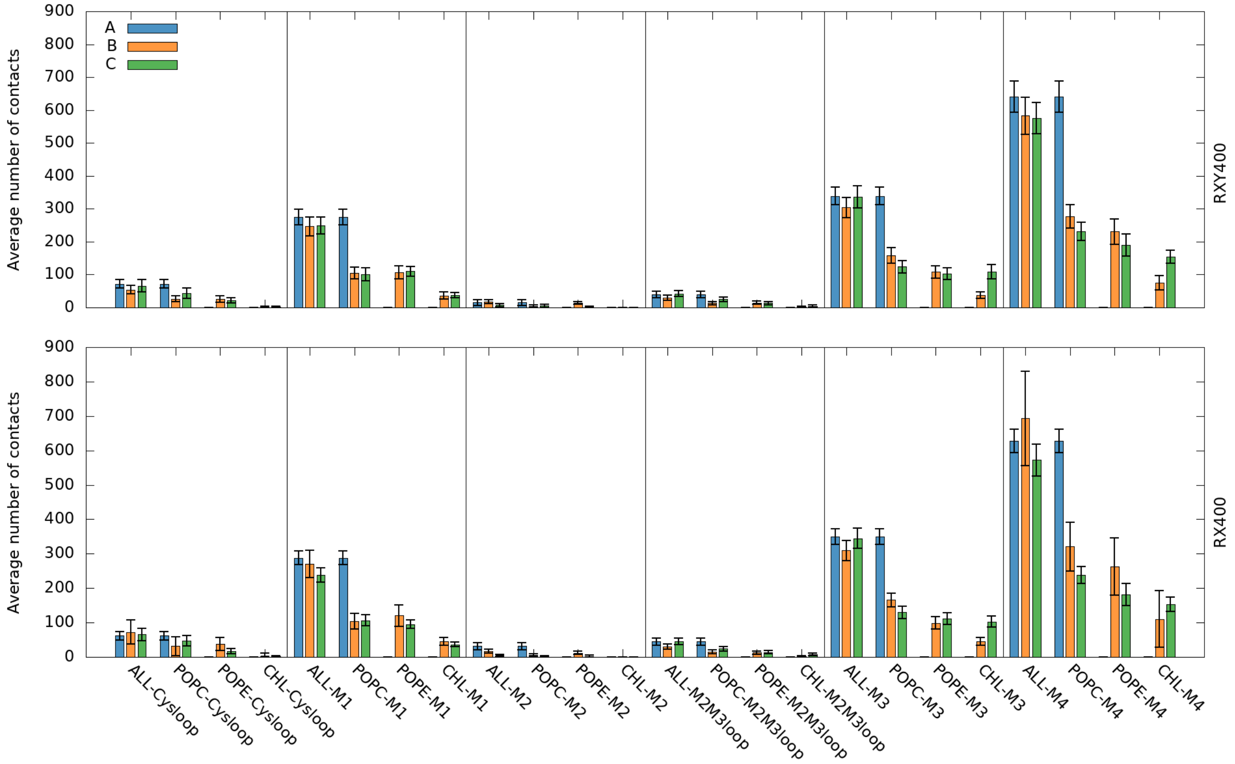


Figure 4.9: Number of contacts, averaged over the five subunits, between helices/loops and lipids, for RXY400 (top) and for RX400 (bottom). Figure from reference [3].

The M1 helices are able to make contacts with lipids, most of which are not hydrogen bonds. The M2 helices form no hydrogen bonds, and in general make sporadic contacts with lipids: this is due to their location deep within the interior of the protein. Both M3 and M4 form contacts and hydrogen bonds with the lipids, mostly with POPC. As for the M2-M3 loop and the Cys-loop, their location in the protein puts them in the position to interact with the heads of the upper leaflet, thus via hydrogen bonds.

4.5 Correlations among Contacts

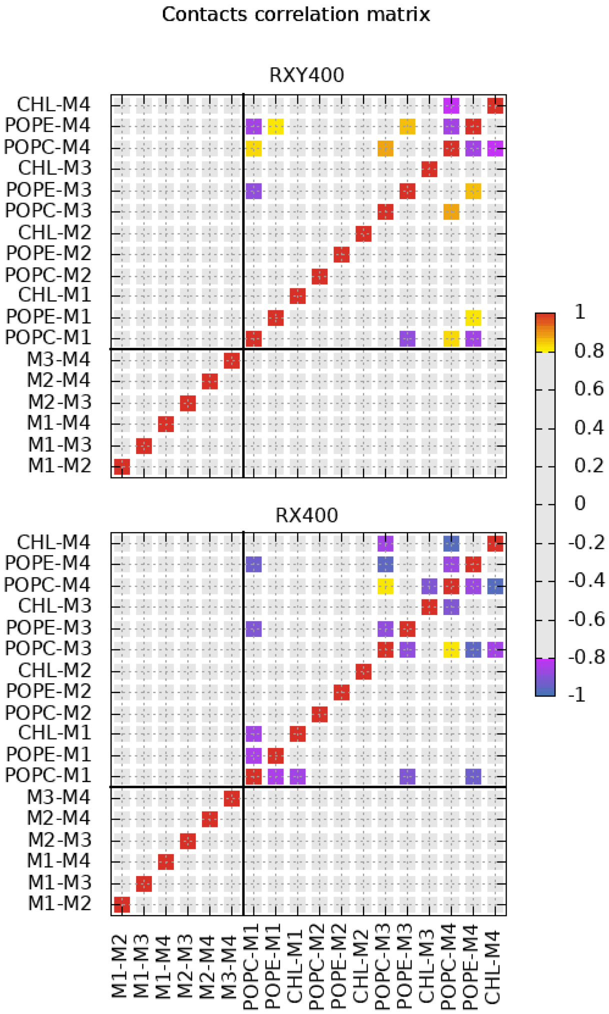


Figure 4.10: Correlation matrix for lipids-helix and helix-helix contacts for RXY400 (top) and RX400 (bottom).

The 15 lipid species-helix numbers of contacts can be used to search for trends of a given pair over the five subunits and over the three models. A comparison between these trends can be performed by calculating, for each two-pair comparison, the Pearson correlation coefficient ρ, which can take value comprised between -1 (full anti-correlation) and 1 (full correlation); here we considered correlations for $\rho > 0.8$, and anti-correlations for $\rho < -0.8$. The correlation between a given lipid species-helix (helix-helix) pair and that of another pair would indicate that an increased number of contacts for the former goes hand-in-hand with an increase of the latter, potentially involving a mechanistic relationship between the two. An anti-correlation, on the other hand, would mean that the increase of one is related to the decrease of the other, and may also imply a relationship between the two pairs. Figure 4.10 shows the Pearson coefficient for every two-pair comparison.

Correlations are found for lipid-helix pairs characterised by the same lipid species and whose helices are in close proximity to one another (e.g. POPC-M3 and POPC-M4): this is observed since once a lipid molecule is found in the proximity of a given helix, it likely forms contacts with the adjacent ones too. On the other hand, anti-correlations are seen for different lipid species competing for the proximity to a given helix (e.g. POPC-M4 and POPE-M4), or to adjacent helices (e.g. POPC-M1 and POPE-M3). Neither correlations nor anti-correlations are observed between any pair lipid-helix and any pair helix-helix, as reflected by the empty rectangles in figure 4.10, over the explored time windows. Thus, the interactions with the three lipid species do not appear to have an effect on the intra-TMD contacts. These results are qualitatively valid for both RXY400 and RX400: while differences are observed between the two methods, they only depend on the specific contacts formed in the different replicas.

4.6 Lipid-Binding Events

The lipids that give rise to the most stable interactions with the protein would have high proximity lifetimes. Thus, the interface between lipids and the TMD can be characterised in details by calculating the distributions of proximity lifetimes of those lipids with any atom within 3.5 Å of the TMD helices. These distributions were computed for RX450 and RY200, given that the lifetimes would depend on the specific replica, and are shown by figure 4.11.

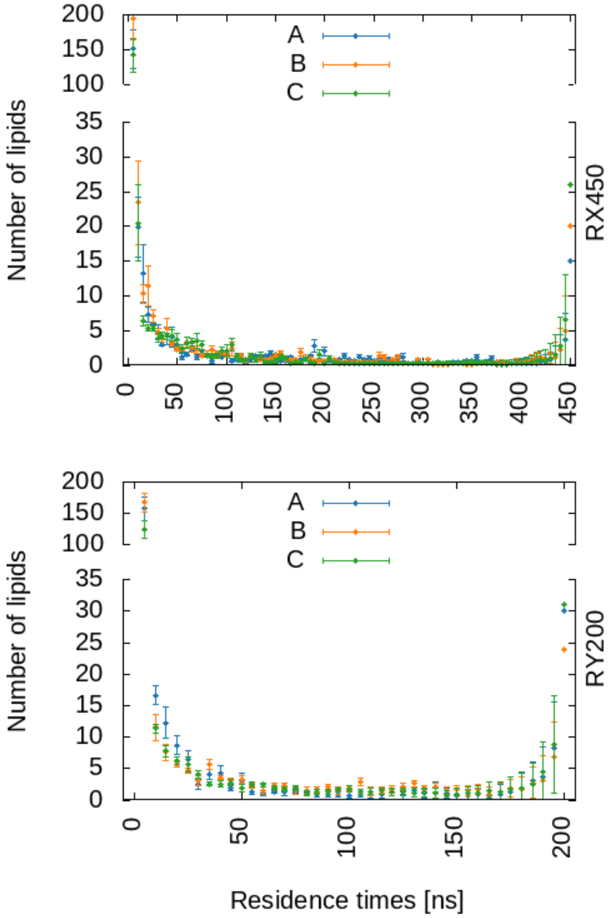


Figure 4.11: Proximity lifetimes distributions of any lipid in close proximity to the TMD in the three models: A (blue), B (orange) and C (green), for RX450 (top) and RY200 (bottom). Figure from reference [3].

For the calculation of the average number of lipids for a given residence time in close proximity to the protein, and its associate uncertainty, we considered multiple time windows of different size (shifted of $\Delta t = 5$ ns, the binning size): for example, for RY, one of 200 ns, two of 195 ns (0-195, 5-200), three of 190 ns (0-190, 5-195, 10-190), and so on. In particular, if $\mathrm{n}_{i,j}$ is the number of lipids whose residence time has $i\Delta t$ duration and $j\Delta t$ starting point (where both i and j are integer numbers), the average number of lipids with residence time i is calculated as:

$$\langle N_i \rangle = \frac{1}{(I-i+1)} \sum_{j=1}^{I-i+1} n_{i,j} \ , \quad i = 1, ..., I = \frac{t_{sim}}{\Delta t} \tag{4.6.0.1}$$

Here, t_{sim} is the production time (200 ns for RY, and 450 for RX). Similarly, the average squared number of lipids with i-th residence time is:

$$\langle N_i^2 \rangle = \frac{1}{(I-i+1)} \sum_{j=1}^{I-i+1} n_{i,j}^2 \tag{4.6.0.2}$$

so that the uncertainty associated to the i-th residence time is:

$$\sigma_i = \sqrt{\langle N_i^2 \rangle - \langle N_i \rangle^2} \tag{4.6.0.3}$$

Having only one point available, the histogram corresponding to the highest residence time cannot have an associated uncertainty. A similar error analysis is performed for RX450.

The distributions for the three models decay fast well before 50 ns, which means that very few lipids spend long times close to the TMD. Notably, the value of the rightmost histogram, corresponding to the longest time sampled, is higher, implying that a certain number of lipids spend the entire production time in direct contact to the protein. Their number varies both between the two replicas and among models. The lipids corresponding to the last histogram are therefore worth of further analysis. As lipids are typically characterised by high mobility, a measure of their dynamics would shed light on the stability of these bound lipids. Therefore, their RMSFs are calculated over RX450 and RY200 with respect to their average position, and shown in figures 4.12 to 4.17 (the lipids are labelled as in the MD data). The RMSFs of phospholipids are calculated separately for the head and the two tails, since each of these parts may exhibit different mobility due to different interactions with or hindrance by the protein and/or other lipids.

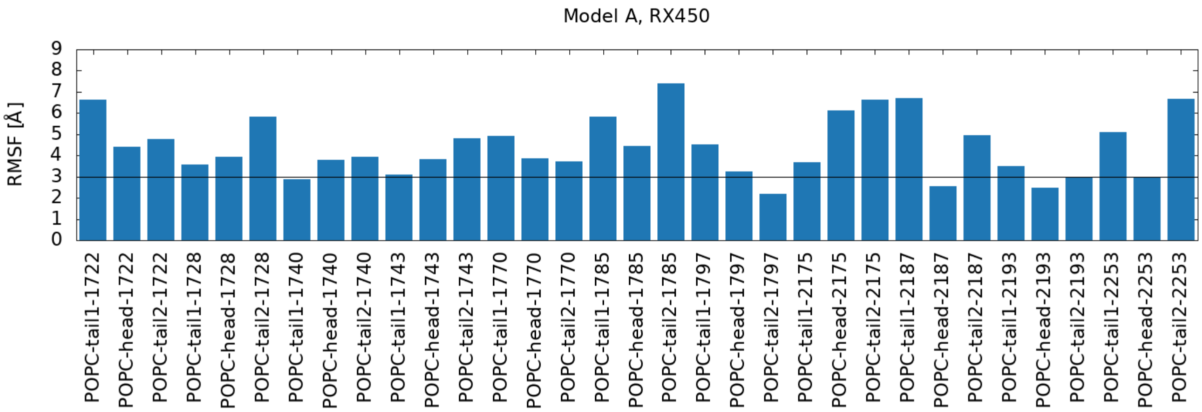


Figure 4.12: RMSF of lipid residues for model A, RX450. Figure from reference [3].

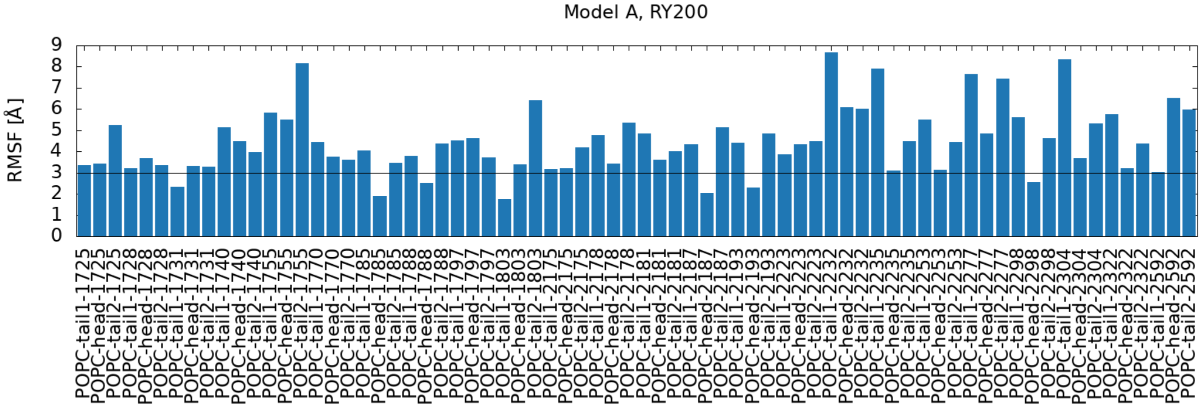


Figure 4.13: RMSF of lipid residues for model A, RY200. Figure from reference [3].

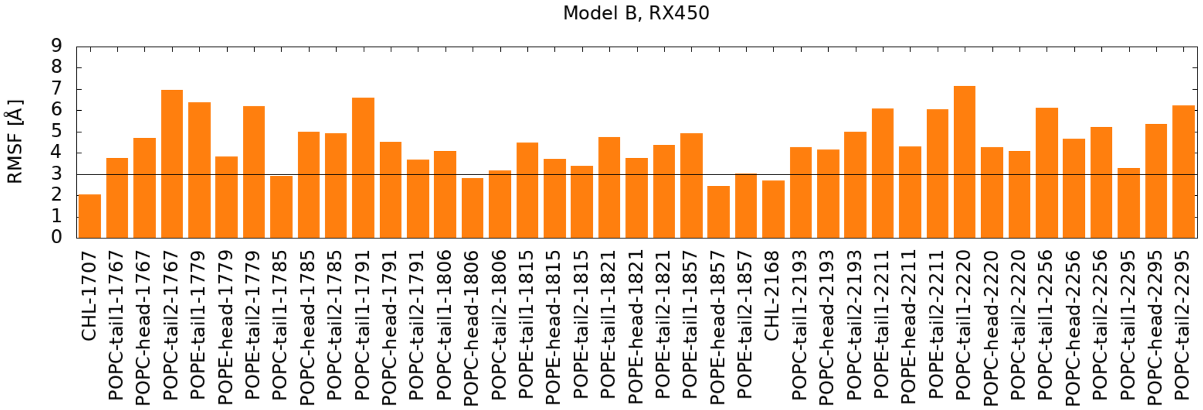


Figure 4.14: RMSF of lipid residues for model B, RX450. Figure from reference [3].

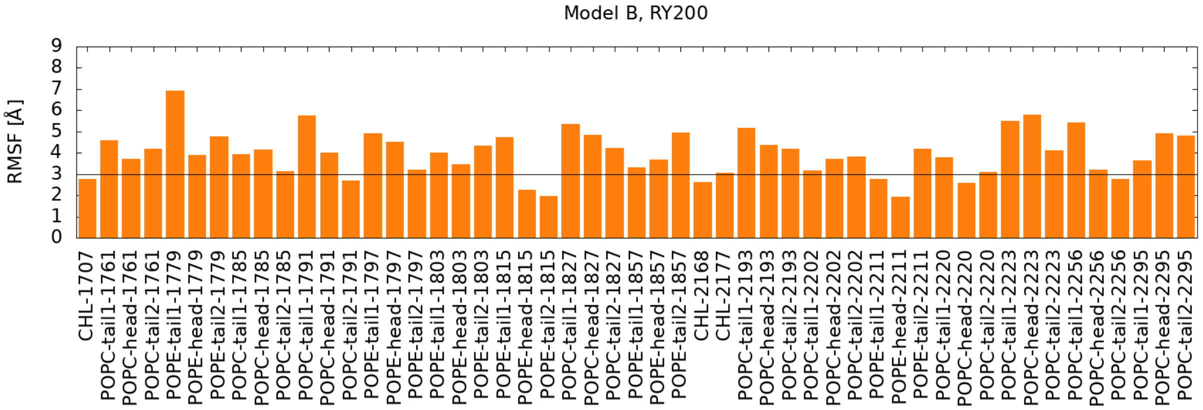


Figure 4.15: RMSF of lipid residues for model B, RY200. Figure from reference [3].

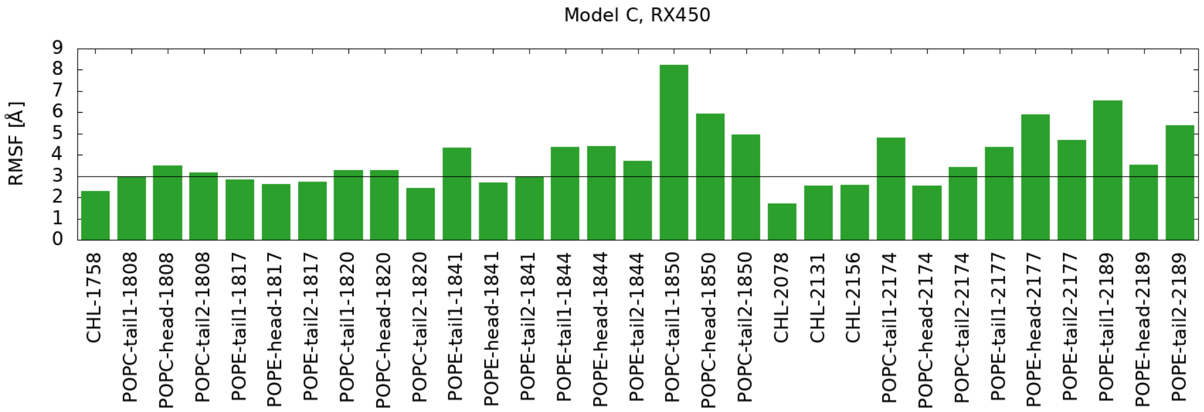


Figure 4.16: RMSF of lipid residues for model C, RX450. Figure from reference [3].

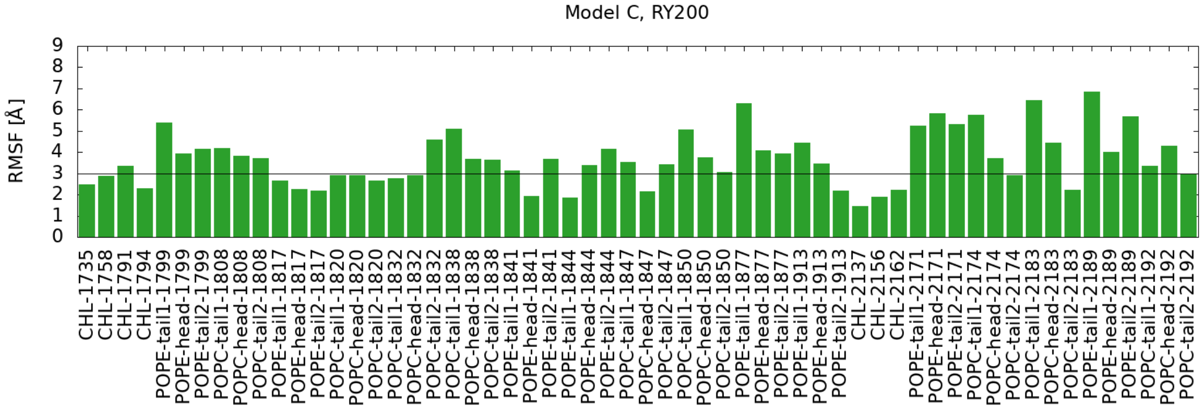


Figure 4.17: RMSF of lipid residues for model C, RY200. Figure from reference [3].

The RMSF of bound lipids range from about 1.5 to about 8.5 Å. The bound lipids with at least one residue (head or tail for phospholipids, whole molecule for cholesterol) whose RMSF was below a certain value (i.e. smaller than 3.0 Å) are those characterised by the lowest fluctuations and are therefore the most stable ones. Their number is shown in table 4.1.

Model	**Lipid**	**N. of sites (RX450)**	**N. of sites (RY200)**
A	POPC	5	8
	POPE	0	0
	CHL	0	0
	ALL	5	8
B	POPC	2	3
	POPE	1	2
	CHL	2	2
	ALL	5	7
C	POPC	3	6
	POPE	2	4
	CHL	4	6
	ALL	9	16

Table 4.1: Number of lipids (with RSMF smaller than 3.0 Å) bound to the protein for the whole duration of RX450 and RY200 replicas for models A, B and C. Table from reference [3].

Given the stochastic nature of lipid diffusion simulated with the Langevin thermostat and barostat in each replica, different binding events were observed in the various replicas. Their number is larger in replica RY200 with respect to RX450, as expected because of its shorter time scale. It is interesting to note that this number significantly increases, almost doubling, for model C with respect to A and B, and this effect results in both RX200 and RY450.

This fits well with the increased viscosity related to a larger content of cholesterol. Therefore, the content of this lipid is an important bulk property of the membrane which promotes the persistence of binding events of any lipid species.

The binding sites of every stably-bound lipid involving one or more helices/loops of one or two adjacent subunits, as reported by tables 4.2 and 4.3.

Lipid	**Binding site**
A, RX450	
POPC1740	$M1_2$ $M3_2$
POPC1797	$M1_1$ $M3_1$ $M4_1$
POPC2187	$M1_5$ $M3_1$ $M4_5$
POPC2193	$M1_3$ $M3_4$
POPC2253	$M4_2$
B, RX450	
POPC1785	$M1_1$
POPC1806	$M1_3$
POPE1857	$M1_4$ $M4_4$
CHL1707	$M1_5$ $M4_5$ (figure 4.18a)
CHL2168	$M1_4$ $M3_5$
C, RX450	
POPC1808	$M4_1$
POPC1820	$M1_5$ $M4_5$
POPC2174	$M3_4$
POPE1817	$M1_1$ $M4_1$ (figure 4.18b)
POPE1841	$M1_1$ (figure 4.18c)
CHL1758	$M1_3$ $M4_3$
CHL2078	$M3_5$ $M4_5$ (figure 4.18d)
CHL2131	$M1_4$ $M3_5$
CHL2156	$M3_2$

Table 4.2: Lipid binding sites to the receptor in replicas RX450, characterised by one or more helices/loops of one or two subunits (indicated as subscripts). The sites shown are only those whose corresponding lipid residue has a RMSF below 3.0 Å . Table from reference [3].

Lipid	**Binding site**
A, RY200	
POPC1731	$M1_1$ M2-M3loop$_2$
POPC1785	$M4_4$
POPC1788	$M4_5$
POPC1803	$M1_4$ $M3_4$ $M4_4$ Cys-loop$_4$
POPC2187	$M1_5$ $M4_5$
POPC2193	$M1_3$ $M3_4$
POPC2298	$M4_3$
POPC2592	$M4_5$
B, RY200	
POPC1791	$M1_1$ $M4_1$
POPC2220	$M3_2$ $M4_2$
POPC2256	$M4_5$
POPE1815	$M1_2$ $M2_3$ $M3_3$ M2-M3loop$_3$ (figure 4.18e)
POPE2211	$M1_3$ $M3_4$ $M4_3$
CHL1707	$M4_5$
CHL2168	$M3_5$
C, RY200	
POPC1820	$M1_5$ $M4_5$ Cys-loop$_5$ (figure 4.18f)
POPC1832	$M1_4$
POPC1847	$M3_4$
POPC2174	$M3_4$
POPC2183	$M4_2$
POPC2192	$M3_3$ $M4_3$
POPE1817	$M1_1$ $M3_1$ $M4_1$ (figure 4.18g)
POPE1841	$M1_1$ $M3_2$
POPE1844	$M1_5$
POPE1913	$M3_5$ $M4_5$
CHL1735	$M3_2$
CHL1758	$M1_3$
CHL1794	$M3_2$
CHL2137	$M3_2$
CHL2156	$M1_1$ $M3_2$
CHL2162	$M3_4$

Table 4.3: Lipid binding sites to the protein in replicas RY200, characterised by one or more helices/loops of one or two subunits (indicated as subscripts). The sites shown are only those whose corresponding lipid residue has a RMSF below 3.0 Å . Table from reference [3].

Most binding sites result from a single helix/loop, nevertheless several exceptions are found, especially for the longer and more flexible phospholipids. Lipid persistent binding occurs, therefore, in various poses. Representative examples of the persistent binding events are displayed in figure 4.18, and are described in the following.

In model B, replica RX, CHL1707 binds to $M1_5$ and to $M4_5$, and, for a certain part of the

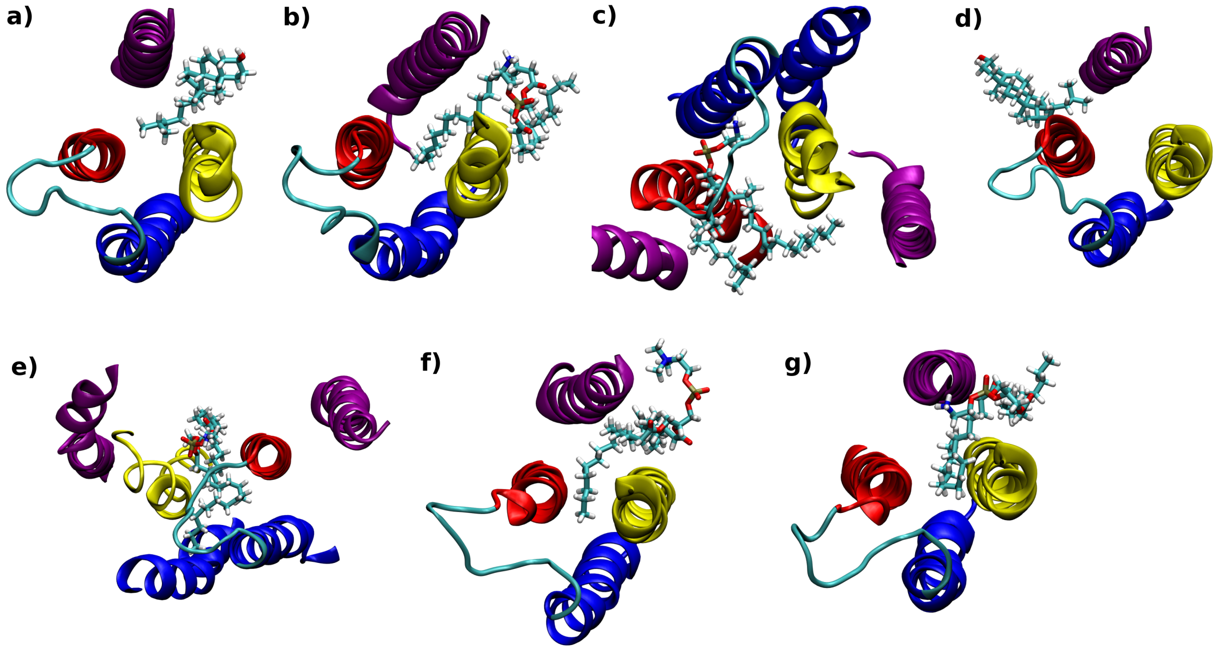


Figure 4.18: Representative examples of persistent binding lipid molecules to the receptor: a) CHL1707 in model B, RX. b) POPE1817 in model C, RX. c) POPE1841 in model C, RX. d) CHL2078 in model C, RX. e) POPE1815 in model B, RY. f) POPC1820 in model C, RY. g) POPE1817 in model C, RY. Figure from reference [3].

trajectory, partially intercalates within them, reaching out out for $M3_5$ (figure 4.18 a). In model C, replica RX, POPE1817 binds to $M1_1$ and to $M4_1$, and penetrates between them reaching out for $M3_5$ for a fraction of the trajectory with one of its tails (figure 4.18 b). POPE1841 binds to $M1_1$, and its head can intercalate, at times, in the direction of the interior $M2_1$ (figure 4.18 c). This may lead to the formation of a hydrogen bond with Leu273 of M2-M3 loop. Finally, CHL2078 binding site may result in the intercalation among the M1, M3 and M4 of subunit 5 (figure 4.18 d). In model B, replica RY, one tail of POPE1815 intercalates between $M1_2$ and $M3_3$ reaches out for the M2 helices of subunits 2 and 3 (figure 4.18 e). In model C, replica RY, one tail of POPC1820 intercalates between $M1_5$ and $M4_5$, and at times reaches out for $M2_5$ (figure 4.18 f). POPE1817 binds to $M1_3$ and $M4_3$, with one of its tails penetrating in between them and reaching out for $M3_3$, although without also reaching for the M2 helix (figure 4.18 g). It is worth of mention that intercalation events within the TMD were also briefly observed for other, more mobile, lipid molecules.

Overall, intercalation mostly occurs for phospholipids, due to the length and flexibility of their tails. On the other hand, over longer times, cholesterol may fully penetrate among the helices of one subunit, thanks to its smaller size.[42] Such an event was observed for cholesterol, POPC and SDPC, in the 5-HT$_{3A}$ receptor over a timescale of tens of microseconds.[42]

It is also worth noticing that the binding sites are similar among the three models, as they would be specific of the protein and of the three lipid species investigated. However, as previously stated, any of these binding events is more likely to be observed when the membrane contains a higher cholesterol content.

POPC has been previously observed to penetrate within the TMD of membrane proteins, for example in GLIC,[120], while cholesterol binding events were observed in the adenosine receptor type 2A,[132] in the G-protein coupled 5-HT$_{1B}$ receptor,[133] in the GABA$_A$ receptor,[114] and in ELIC.[118] A lipid-M2 interaction was also observed for the GABA$_A$ receptor.[114]

4.7 Conclusions and Future Perspectives

In this chapter, we continued our investigation of the 5-HT$_{3A}$ receptor, this time focusing on a well-established phenomenon, namely lipid-protein interaction and its implications on pLGICs function.[106] Conversely to proline isomerisation, the nature of this process regards a much more extended region of the protein, that is the whole interface between its TMD and the lipid membrane. In order to perform this investigation, we carried out molecular dynamics simulations with the aim to assess the effects of the model membrane composition at the interface with the receptor. We compared a pure-POPC membrane and two different mixed ones, which also contains the phospholipid POPE and cholesterol in different content. In particular, they spanned the lower and upper limits of this lipid concentration in synaptic membranes. The constructed three models allowed for the distinction of the effects due to the mere presence of a given species from those due to its exact concentration.

We introduced two different ways to evaluate statistical averages: one considers two shorter replicas, and one considers a single longer single replica, in this case covering double the time of a short one. When it came to the evaluation of global or averaged quantities, no qualitative nor quantitative difference appeared from the usage of these two methods.

Structural stability, as evaluated through RMSDs, and features such as channel width, ion portals, and correlations among contacts, did not show significant differences over the simulated time scales; the receptors remained in a non-conductive state provided that the mixed lipid bilayers were properly pre-equilibrated before releasing the receptor. Over much longer time scales, which can access slow structural rearrangements, differences between varied membrane composition may emerge, with potential roles played by intercalated lipids for example in the stabilisation of pre-active states,[42] or by patterns of interactions with different functional states of the receptor.[115]

The time scales explored in the simulations allowed however for a meaningful sampling of lipid diffusion,[141, 142] and consequently of the lipid-protein interactions and binding events. A multitude of binding events of POPC, POPE and cholesterol molecules to the receptor TMD were observed in both replicas of all three models. The stochastic nature of lipid diffusion results in a different local lipid environment around each of the five subunits, so that distinct replicas allow for a larger sampling of such events. Model C, i.e. the one characterised by the highest viscosity due to the highest cholesterol content, results to be the one with the highest number of lipid binding events to the receptor. The binding sites involved one or more of the helices/loop of one or two adjacent subunits. Certain binding events lead to intercalation within the TMD, either within the helices of one subunit, or within the grooves between two adjacent ones, potentially forming contacts and hydrogen bonds with inner structures like the M2-M3 loop and the Cys-loop of the ECD, which are thought to be relevant protein sections in the signalling transmission cascade of events triggered by ligand binding and culminating in channel gating.[7, 89]

Our findings also highlighted the different role of phospholipids with respect to cholesterol, since the three different species compete for binding and interactions with the protein. While cholesterol affects membrane viscosity, increasing the number of any lipid binding event, phospholipids form the largest number of hydrogen bonds, due to their higher number of suitable moieties, and are more able to intercalate within the protein, due to their length and/or the flexibility of their tails, thus being more likely to reach out for protein sections such as the Cys-loop in the ECD and the M2-M3 loop in the TMD, and possibly bridging these two loops. The viscosity of the membrane has been previously reported to influence membrane protein structural features: molecular dynamics simulations demonstrated how a change from an initial fluid phase to gel phase of the membrane bilayer affects the structure of the glucose transporter GLUT1, in particular it reduces the size of the cavities and tunnels traversing and connecting the external regions of the transporter and the central binding site.[154] Furthermore, the importance of cholesterol has been experimentally proven: its depletion affects the distribution on the cell surfaces of 5-HT$_{3A}$R and of other pLGICs. It also decreases serotonin-induced current response and desensitisation rates.[106]

Like the investigation described in the previous chapter, this study focused on a non-conductive apo-state state and, over the simulated times, the receptor structure was stable and did not changed significantly.

Despite the full picture regarding lipid-protein interactions is still far from being completely unravelled, our results contributed to the understanding of the phenomenon. Further investigations aiming at completing the picture may explore more accurate force-fields,[155] coarse-grained methods for the exploration of longer time scales, different functional states of the receptor,[115] a larger variety of membrane compositions, or additional replicas for further sampling. Enhanced sampling methods may also be employed for the study of lipid intercalation or binding (e.g. to quantify the associated free energy).[156]

In the next chapter, we will focus our attention on the role of the outermost, lipid-facing M4 helices on receptor function. Among other things, we will explore the possibility that this role is also influenced by the lipid environment.

Chapter 5

A Point-mutation on the M4 Helix Influences Channel Function Through a Radial Mechanism

5.1 Introduction

According to growing evidence in recent literature, the outermost α-helices M4 play a role in pLGICs function. Besides the role of amphipathic barrier to the hydrophobic lipid environment, these helices may also play a role in channel gating: in fact, mutations in this section of the receptor have been shown alter channel opening: in most cases they diminish or inhibit it,[157, 158, 159] while in one case they promote it.[160] Nevertheless, the mechanism of how this occurs is still to be understood.

Several experiments were carried out in *Xenopus* oocytes (i.e. female gametocytes) as well as in human embryonic kidney (HEK) 293 cells, which are a specific cell line originally derived from human embryonic kidney cells grown in tissue culture taken from an aborted female fetus. Mutations in cationic pLGICs M4 helices that do not produce wild-type-like receptors are slightly beneficial in receptors expressed in *Xenopus* oocytes, while being uncoupling in HEK293 cells: in the α7 nAChR, 11 mutations to alanine improve channel function in oocytes,[160] but 8 mutations to alanine in the M4 of the $\alpha 4\beta 2$ nAChR receptor abolish channel function in HEK293 cells.[161]

Mutations to alanine in the M4 of the α subunit of the muscle nAChR expressed in oocytes show both gains and losses of function [162]. These changes in function are all comparatively small, perhaps because the mutations are only present in two subunits per channel. This difference in channel function of M4 mutants between expression systems has not been observed in anionic or bacterial channels. In anionic channels, mutations to alanine, especially of aromatic residues, generally diminish channel function, regardless of expression system.[163, 158] In the bacterial pLGICs assayed in the same system (*Xenopus* oocytes), M4 mutations result in opposite effects: 15 mutations to alanine in the GLIC M4 reduce channel function, while 26 mutations to alanine in the M4 of ELIC enhance channel function.[164]

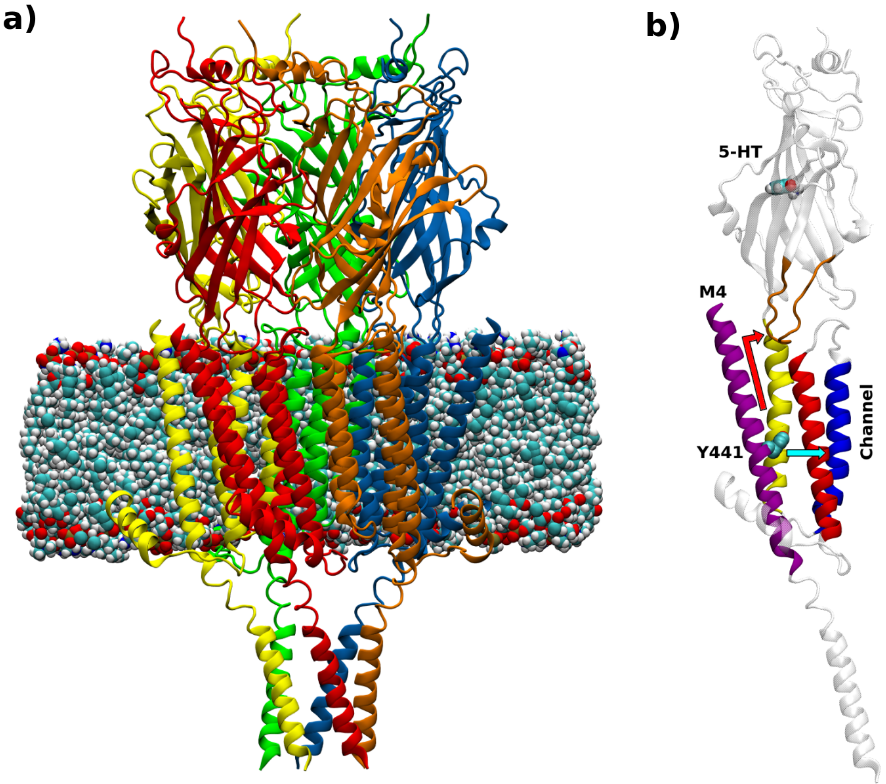


Figure 5.1: a) Model of the open structure of the 5-HT$_{3A}$ receptor used for the calculations described in this chapter (PDB entry: 6DG8). b) Schematic representation of the two proposed mechanisms of M4 coupling as inferred from the Y441A mutation: "radial" (in light blue) and "vertical" (in red). Figure from reference [8].

Two main mechanisms have been proposed for M4 coupling ligand binding to channel gating. For one mechanism, first described by DaCosta et al. in a work on the nAChR,[112] the C-terminal of the M4, located at the level of the ECD-TMD interface, is needed for interactions between the M4 and the Cys-loop: the removal of the M4 tip would prevent the channel opening signal from reaching the TMD. This model finds support in the fact that allosteric positive modulators has been observed to bind in a cavity in between M3 and M4 helices of the $\alpha4\beta2$ nAChR, and to communicate with the Cys loop through interactions between the last residue of the M4 tip. [165] While the C-terminal end of the M4 can be removed without abolishing function in both ELIC[164] and the *Torpedo* nAChR expressed in *Xenopus* oocytes,[166] removal or mutation to alanine of C-terminal residues abolishes function in the 5-HT$_{3A}$ and $\alpha4\beta2$ receptor receptor in HEK293 cells.[167, 168, 159, 161] However the first two studies were performed in *Xenopus* oocytes, and the latter four in HEK293 cells, indicating that the necessity for the C-terminal domain may depend more on the expression system than on the specific channel.

The alternative mechanism involves the interaction of M4 residues with residues on the other helices of the TMD: the signal that can result from such interactions may propagate towards the centre of the channel and affect channel opening. One M4 mutation, C418W, has been experimentally observed in the *Torpedo* nAChR, expressed in oocytes, as energetically coupled to S226 and T229, both on the M1 helix. This mutation, however, does not alter the interactions between the M4 tip and the Cys-loop.[169]

Overall, while there is consensus that the M4 helices play a significant role in pLGICs function, this role may depend on many factors, including the specific receptor and the expression system.

In this chapter, we continue our investigation on the 5-HT$_{3A}$R, focusing on the role of the M4 helices and aiming at probing the likelihood of both mechanisms to occur in this receptor. The work described here consists in the *in-silico* part of a joint computational and experimental project, which also involved mutagenesis and functional assays performed by Susanne M. Mesoy and Sarah C. R. Lummis of the Biochemistry Department at the University of Cambridge. The experiments proved that the 5-HT$_{3A}$ receptor may constitute a unique case when it comes to the mechanism of action of the M4 helix, in fact they showed that a single mutation to alanine on the M4 helices inhibit channel opening on ligand: the experiments proved that, when mutating into alanine each of the 26 residues of the M4 section 434-461, 24 did not result in significant channel function alteration, 2 affected channel expression, and a single mutation to alanine, specifically Y441A, abolished channel gating upon ligand binding.[159] This residue (located at the centre of the M4, as shown in figure 5.1 b) was therefore identified as a potential mediator for the entire M4 effects on channel function, in this particular receptor.

Past experimental work performed on this very residue showed that the mutation Y441F resulted in no different responses than a wild-type (WT) receptor, while Y441A resulted in functional receptors in oocytes, but not in HEK293T cells.[159] Y441 would form interactions with residues of M1 helix, specifically with F242, M235, and D238. Mutations of these M1 residues were tested as well: F242A receptors resulted in similar properties to WT ones, as well as M235A receptors with RIC-3 (a chaperone which has been shown to assist in the expression of 5-HT$_{3A}$R),[170] while D238A produced non-responsive receptors with or without RIC-3, thus proving to provide a crucial interaction between Y441 and D238.

On one hand, Y441 may propagate a mechanical signal to the M4 tip, inducing it to interact with the Cys-loop, while this may not occur in a Y441A mutated receptor. This can be seen as inducing (or not inducing) a "vertically-propagating" chain of events from residue 441 to the M4-tip (where the vertical direction is parallel to the main protein axis of the receptor). The other mechanism, that would involve a chain of interactions with residues of the other helices, can be instead described as "radially-propagating". Both mechanisms are represented in figure 5.1 b.

In this context, we performed molecular dynamics simulations on the wild-type and on a Y441A-mutated receptor in order to highlight structural and dynamical differences between the two, shedding light on the role of this mutation (and of the M4 in general) on receptor function. We found hints that the vertically-propagating mechanism may not occur in the 5-HT$_{3A}$R, and that this mutation may instead act on D238 on the M1 to affect K255, on the innermost M2. This "radial" mechanism found confirmation in the experimental observations. Considering the different lipid composition that characterise the membranes of HEK cells and oocytes,[171, 172] we also investigated the possibility that some lipid-related effect may take place around residue 441.

The work presented in this chapter was accepted as a paper in the early 2021, appropriate references are made where relevant.[8]

5.2 Receptor Models and Simulations

For the study of the role of M4 helices in the 5-HT$_{3A}$R, we built a model that is based on the cryo-EM serotonin-bound, open structure resolved by Basak et al. at 3.89 Å resolution (pdb entry:

6DG8 [5]), shown in figure 5.1 a). This system contains five serotonins (5-HTs), one per pocket, which are located at the interface between adjacent subunits.

The choice of an open model was based on the assumption that this structure would be representative of a late stage in the cascade of events that leads to gating, and on the assumption that the Y441A would play its role at a much later stage with respect to ligand binding. This structure comprises the TMD, the ECD, and part of the ICD: the highly flexible residues 333-396 (in the ICD), which include the long MX-MA loops, were not resolved and not reconstructed by us, not only because of their considerable length (which would result in a large solvation box), but also because they would unlikely influence the M4 helices.

The model was protonated at neutral pH, and embedded in a lipid membrane containing 35% POPC, 35% POPE, and 30% cholesterol, by using the CHARMM-GUI web-based membrane builder,[173] resulting in an area of about 124 Å by 127 Å . This 6:7:7 concentration was chosen as to resemble the composition of HEK cells membranes: its cholesterol/phospholipid ratio is 0.42 molarity, closer to the value of 0.48 in HEK cells[172] than to that of 0.6-0.7 in oocytes.[171] Furthemore, this ratio has been used in past simulations of membranes with cholesterol (6), POPC (7) and another lipid (7) for the study of serotonin receptors, including our own described in the previous chapter.[42, 140, 3] Finally, cholesterol is important as this lipid is present in high concentration in brain cells membranes,[134, 135] as previously mentioned, and a mixed membrane could prove important for a cooperative modulation of the effects of the Y441A mutation.

The system was then solvated in an orthorhombic supercell, with a 15 Å buffer of TIP3P water and with 0.15 M of Na^+ and Cl^- ions to reproduce physiological conditions, together with 5 Cl^- counterions to counterbalance the positive charge of the five bound serotonins. Once we prepared this wild-type receptor (WTR) model, we turned the five Y441 into alanine by using PyMOL,[174] in order to create a second, mutated receptor (MR) model. In total the WTR model had 226,082, while the MR model had 226,027 atoms.

The systems were simulated with the NAMD 2.13 molecular dynamics (MD) package[80], and the Amber `ff14SB`[55] and `LIPID14` force-field.[59] The five serotonins in the binding pockets were parametrised as described in the following section.

The simulation time step was 2 fs, and the bonds containing hydrogen were constrained with the SHAKE algorithm. Particle Mesh Ewald was used for the electrostatic interactions and a cut-off of

10 Å was employed for the non-bonded interactions. At the beginning, both models underwent an energy minimisation procedure, then a slow heating and a partially restrained equilibration (with the protein α carbons and the serotonin rings restrained, while the lipids were free to diffuse). Since the diffusion of lipid membranes occurs over time scale of the order of tens to hundreds of nanoseconds,[141, 142] the equilibration stage lasted 148 ns in total, while slowly releasing the chosen restraints. Table 5.1 describes the full equilibration procedure.

Restraints [kcal/mol$\AA^2$]	**Duration [ns]**
25	50
Rescaling	3
15	50
Rescaling	4
8	10
Rescaling	4
4	10
Rescaling	7
1	10
	Total: 148

Table 5.1: Full equilibration procedure for both the WTR and the MR. The restraints were applied on protein α carbons and on 5-HT rings. Table from reference [8].

After the equilibration, a production run was performed for both models. Open-channel models require particular care in order to prevent the collapse of the structure, including the possible closure of the channel.[115] For this reason, the production was simulated with 1.0 kcal/mol$\AA^2$ restraints on the α carbons of M2, MA, and MX helices.

This was carried out within the isothermal-isobaric (NPT) ensemble for every model at a temperature of 310 K, which is above gel transition temperature for all lipid species,[143, 144] and at a pressure of 1 atm. The temperature was controlled by means of a Langevin thermostat with a collision frequency of 1.0 ps^{-1}, and the pressure was maintained with a Langevin piston barostat with an oscillation period of 200 fs and a damping time constant of 100 fs.

For both the WTR and the MR, two replicas of 250 ns each were simulated, referred in the following as R0 and R1. This choice was made as a consequence of the stochastic nature of the Langevin dynamics, which would produce different trajectories in different replicas, affecting the diffusion of lipid molecules in particular, to whom the outermost M4 helices are exposed. Most of the analysis were performed over the conjunction of the time windows 50-to-250 ns of both R0 and

R1 (R01-400), similarly to what done in the previous chapter. However, the analysis of quantities that needed to be expressed as functions of time, or that would be dependent on the order of the union of R0 and R1, were carried out separately for R0 and for R1.

5.2.1 Serotonin Parametrisation

Since it is not a standard residue, the serotonin molecule (5-HT) had to be parametrised, i.e. its atoms needed to be assigned with atom types from the general Amber force field [54] and their partial charges needed to be evaluated with *ab-initio* calculations. This was done with the packages Antechamber[175] and Gaussian09E respectively.[63] The serotonin binds to the receptor in its protonated form, i.e with a terminal, positively charged amine group. Its structure can have, in principle, some conformational variability due to the rotations of its hydroxyl group, and to the flexibility of its tail. Thus, we calculated the partial charges by considering 12 conformations, each with a different orientation of those two groups.

Each of the serotonin structures (characterised by a unique orientation of the tail and of the hydroxyl group) was optimized at the density functional theory level with the b3lyp exchange and correlation functional[176] and a 6-31G* basis set. A single point energy calculation was then carried out on the optimized geometry at the Hartree Fock level with the same 6-31G* basis set (for consistency with the Amber force field parametrization). A multiconformational RESP fitting was then performed for the calculation of the partial charges.[64] Atom types and partial charges are reported in Appendix B.

5.3 Protein and M4 Stability

As a preliminary analysis, the stability of the protein, and of the M4 helices in particular, was studied by looking at the RMSD of the backbone atoms with respect to the last step of the equilibration protocol outlined in table 5.1, as a function of time, for both the WTR and the MR, and in replicas R0 and R1. They demonstrated that the simulated models are stable over the explored time windows.

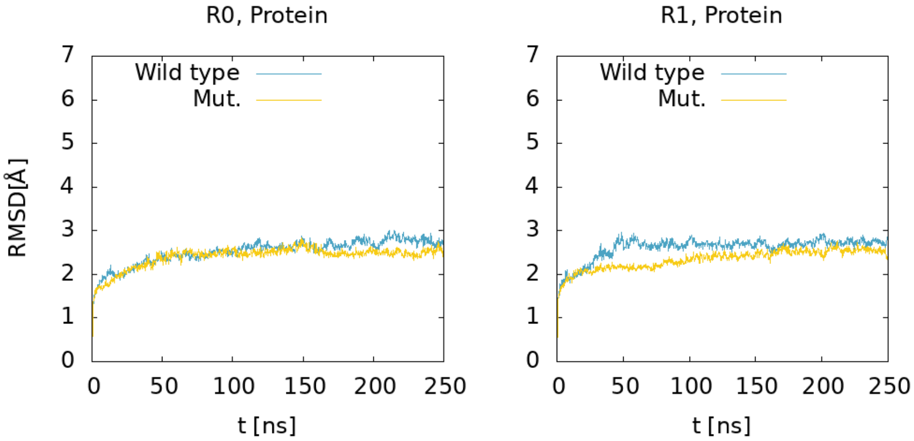


Figure 5.2: RMSD of the protein in the WTR and in the MR, in replicas R0 and R1. Figure from reference [8].

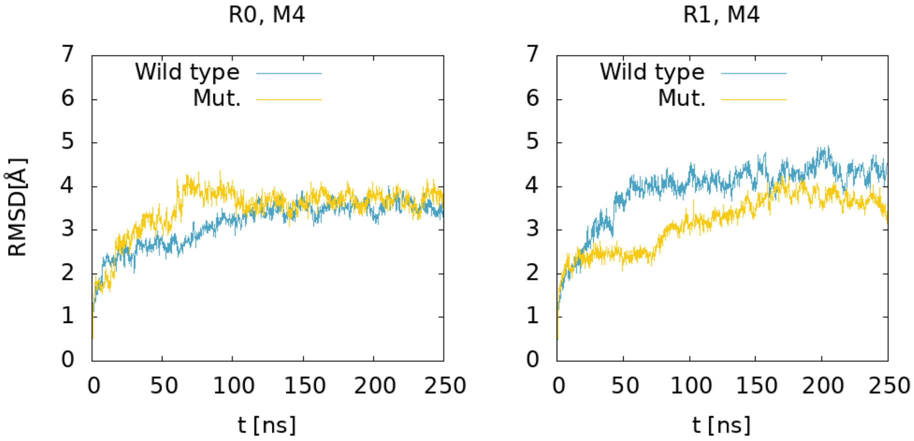


Figure 5.3: RMSD of the M4 helices in the WTR and in the MR, in replicas R0 and R1. Figure from reference [8].

5.4 The Vertical Mechanism

We first studied the likelihood of a vertically-propagating mechanism from the Y441(A) to the M4 tip (shown by figure 5.1), by investigating dynamics and interactions of the M4 residues. In the overall project, this part regarded *in-silico* calculations only; its hypothesis being based on the findings of experiments performed upfront of the project.[159] The dynamics of the M4 can be examined by means of the RMSF (Appendix A) of the backbone atoms of each of its residues. This was calculated over R01-400 with respect to the post-equilibration positions, for the backbone atoms of those residues, and averaged over each of the five subunits (the errors were calculated as maximum semidifferences). The results are shown in figure 5.4.

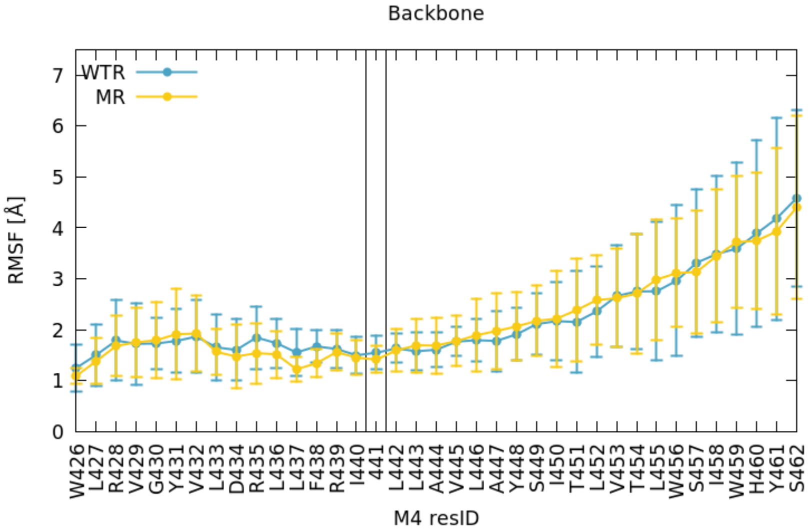


Figure 5.4: RMSF of backbone atoms of M4 amino-acids. Figure from reference [8].

Since the bottom of M4 is restrained, similarities or differences would be spurious and thus not to be considered in this region. Around 441, the residue fluctuations are around 1 to 2 Å, and increase (as well as their variations over subunits) towards the top of the helix, which is located at the interface between the lipid membrane and the solvent, and thus allowed to move more freely. This is shown by figure 5.5, that gives the average z component (with the z axis being the parallel direction to the main protein axis) of the distance between the centre of mass of the ring given by a certain M4 residue in the five subunits and the nearest layer of lipid heads (phosphorus for POPC and POPE, oxygen for cholesterol), calculated over R01-400. The results hold for both the WTR and the MR, which therefore show almost no difference in their dynamics. Most interestingly, on average, the 441 residue itself does not seem to be affected by the lack of the side chain.

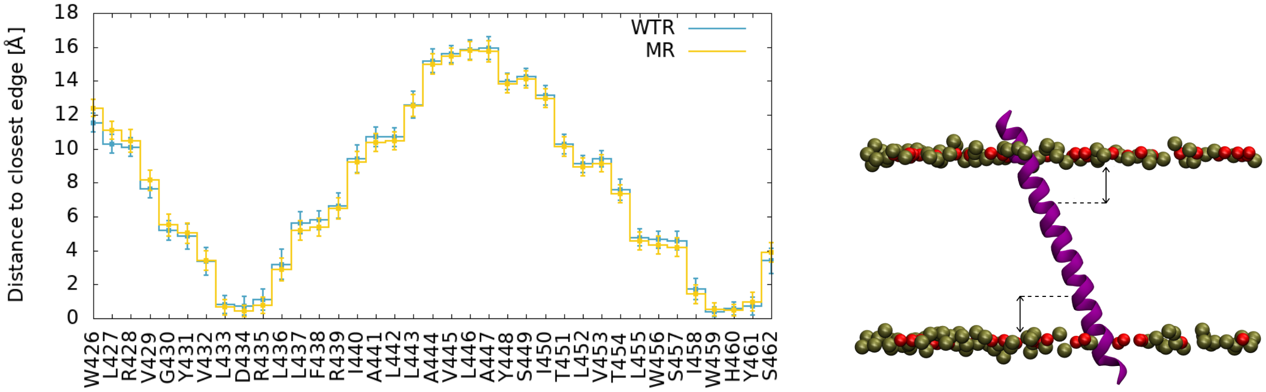


Figure 5.5: Average z component of the distance between the centre of mass of the ring given by a certain M4 residue in the five subunits and the nearest layer of lipid heads (identified by phosphorus for POPC and POPE, and oxygen for cholesterol). This is shown for the WTR (blue) and the MR (yellow), for each of the M4 residues.

The top of the M4 helix, then, appeared not to depend on the properties of residue 441. To confirm this, we aimed at evaluating the effect of the dynamics of residue 441 backbone atoms on the residues above, alongside M4. Thus, we calculated the time-averaged dynamical correlation (see Appendix A) of this residue (in both R0 and R1 and in the five subunits) with respect to the backbone atoms of two representative amino-acids; Y448 and W459 (figure 5.6).[177] Y448 is a residue facing inward that can prove as an important pinning level (e.g. by means of π interactions with Y286), while W459 is a relevant candidate for π interactions and hydrogen bonds with the Cys-loop. Mutation of either of these two residues diminishes or abolishes cell surface expression.[159]

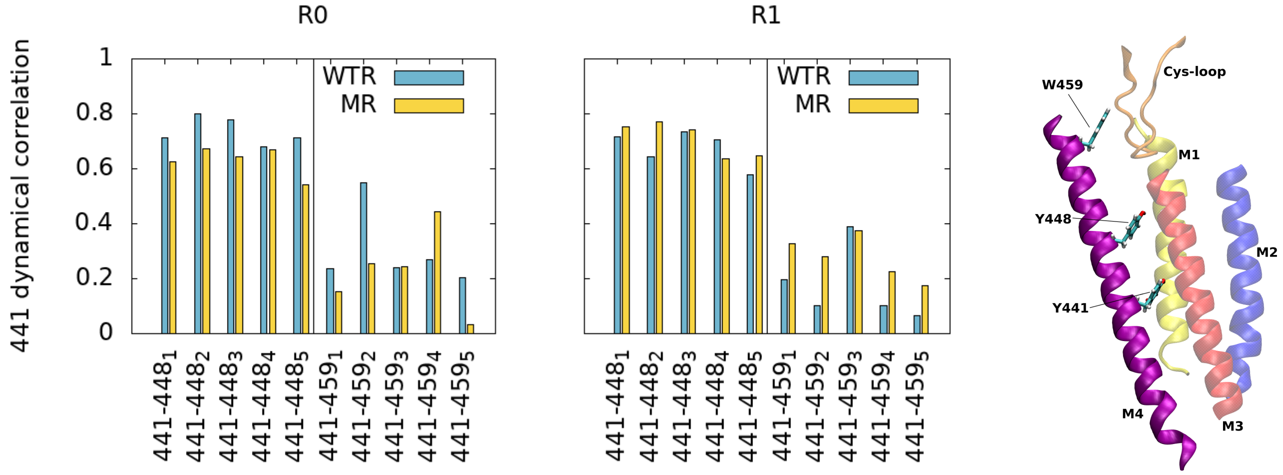


Figure 5.6: Left and centre: average dynamic correlation of residue 441 backbone atoms with those of residues 448 and 459 (in replicas R0 and R1 respectively). Right: location of residues 441, 448 and 459 on the M4 helix. Subscripts refer to subunits. Figure from reference [8].

The correlations between 441 and 448, 459 resulted subunit- and replica-specific, although the 441-448 correlations are consistently higher than the 441-459 ones, given that Y448 is closer to residue 441 with respect to W459. For the pair 441-448 they reach values around 0.8 (with an average of 0.69 ± 0.07), while for the pair 441-459 they never surpass the relatively low value of 0.6 (with an average of 0.24 ± 0.13). This implies that the motions of 441 are not correlated with those of the top of the helix.

Next, we calculated, over R01-400, the time- and subunit-averaged interactions of relevant M4 residues, including hydrogen bonds, π-π interactions and anion-π interactions, shown in figures 5.7 and 5.8).

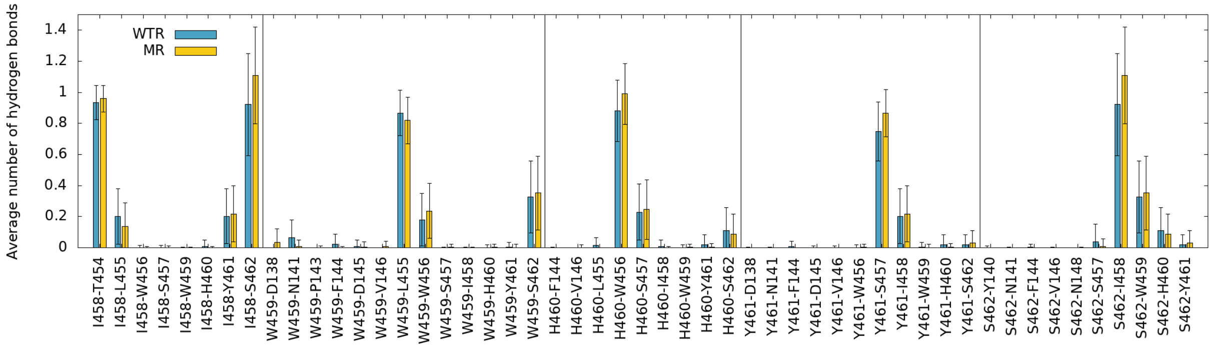


Figure 5.7: Hydrogen bonds of M4 tip residues, from 458 to 462. Figure from reference [8].

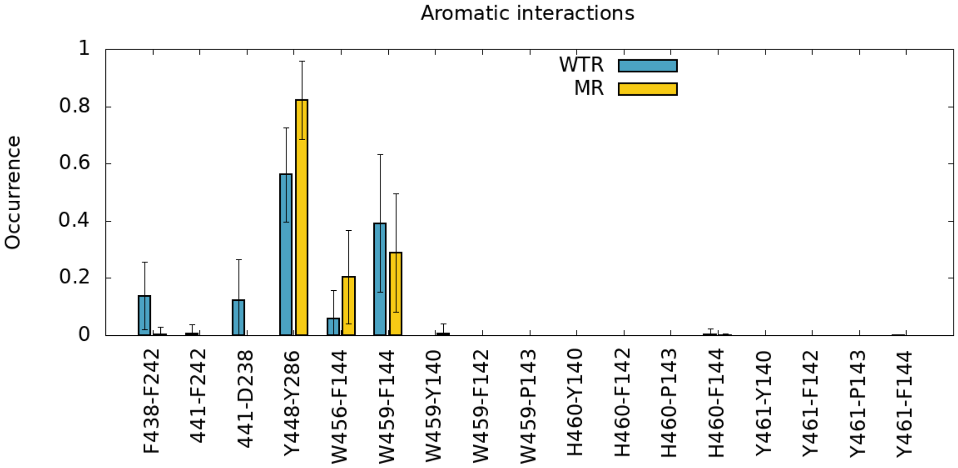


Figure 5.8: π-π interactions (F438-F424, 441-F242, Y448-Y286, W456-F144, W459-F144) and anion-π interactions (441-D238) of different residues of the M4 helix. Figure from reference [8].

Figure 5.7 shows that W459 is indeed the only M4 tip residue that can form hydrogen bonds with the Cys-loop, despite their frequency are not particularly high. However, no major differences are observed between the WTR and the MR when it comes to either hydrogen bonds and aromatic interactions, except for the fact that the lack of the side chain of residue 441 in the MR prevents this residue from reaching out for and forming interactions with its neighbouring amino-acids.

Overall, these findings pointed out that a vertically-propagating mechanisms would not take place in this receptor.

5.5 The Radial Mechanism

Then, we studied the likelihood of a radially-propagating mechanisms from the Y441(A) (shown by figure 5.1 b), by investigating differences in the interactions between this amino-acid with close residues on the other helices M1, M2, and M3. Mutagenesis experiments in HEK293 cells revealed that mutations to alanine of all residues potentially interacting with Y441 (M235, D238, F242, C290 and S297, shown in figure 5.9) proved that only the mutation D238A has a similar effect to Y441A, i.e. the two corresponding mutated receptors did not open in response to ligand addition, despite they both bind the ligand at the cell surface.[8]

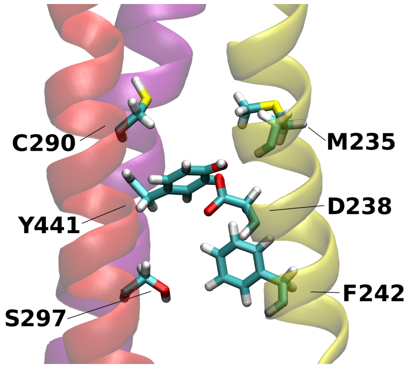


Figure 5.9: Y441 residue in the M4 helix and the amino-acids investigated with mutations: M235, D238, and F242 (on the M1 helix), and C290 and S297 (on the M3 helix). Figure from reference [8].

Given the experimental evidence, we decided to investigate the possible effects that take place between the residues 441 and 238, and that could be altered by Y441A mutation.

We investigated the time- and subunit-averaged hydrogen bonds of residue 441 and D238 with other residues of the four helices, over R01-400 (figure 5.10) for both the WTR and the MR. The associated errors are calculated via error propagation from the five standard deviations of time-data for each of the single subunits.

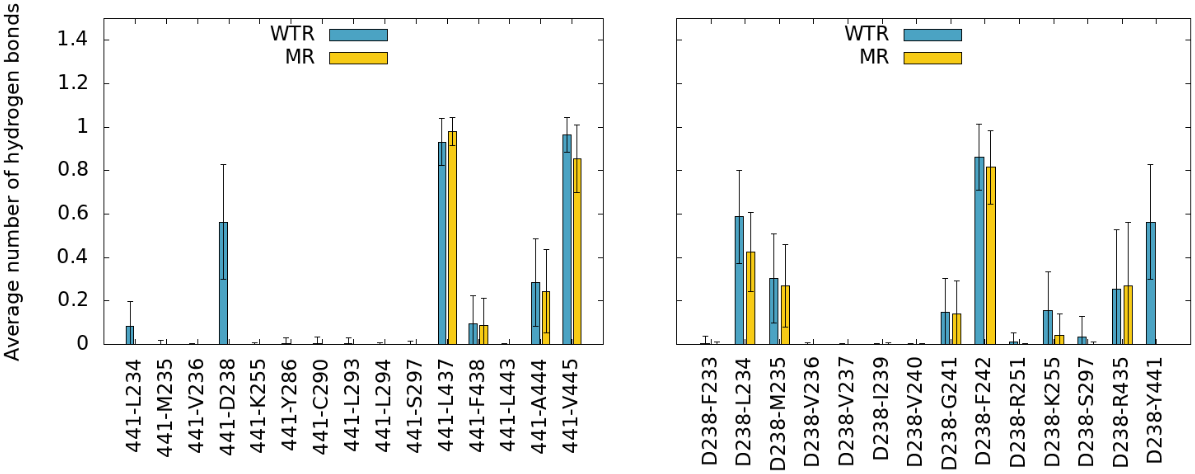


Figure 5.10: Hydrogen bonds of residue 441 (left) and D238 (right) with nearby protein residues. Figure from reference [8].

Between the WTR and the MR, no major differences were noticed, except for the lack of hydrogen bonds between 441 and 238 in the MR, which is simply due to the missing side chain of residue 441 in the MR. However, both residues 238 and 441 appeared able to make interactions with

K255, a positive lysine on the M2 helix, only five residues away from L260, which (as described in previous chapters), makes for a suggested hydrophobic gate of the channel.[14, 15] The side chain of K255 stretches in between M1 and M3, and is directed towards the middle of the four TMD helices. Its terminal nitrogen is, therefore, in a convenient position for the formation of hydrogen bonds with close residues, as shown in figure 5.11.

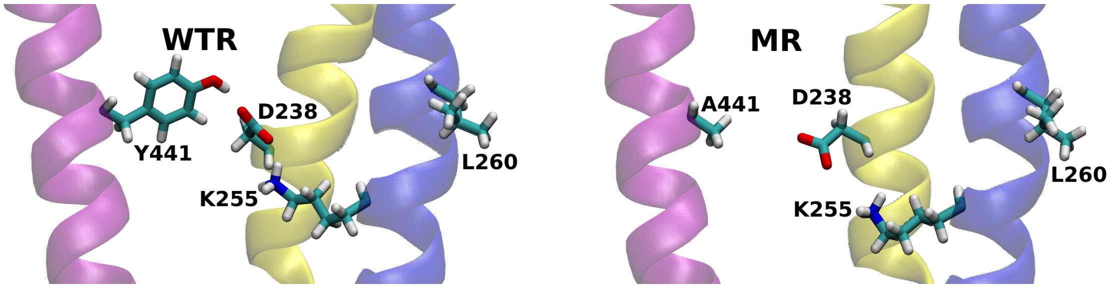


Figure 5.11: Representation of residues 441, K255, and D238, in the WTR and in the MR. Figure from reference [8].

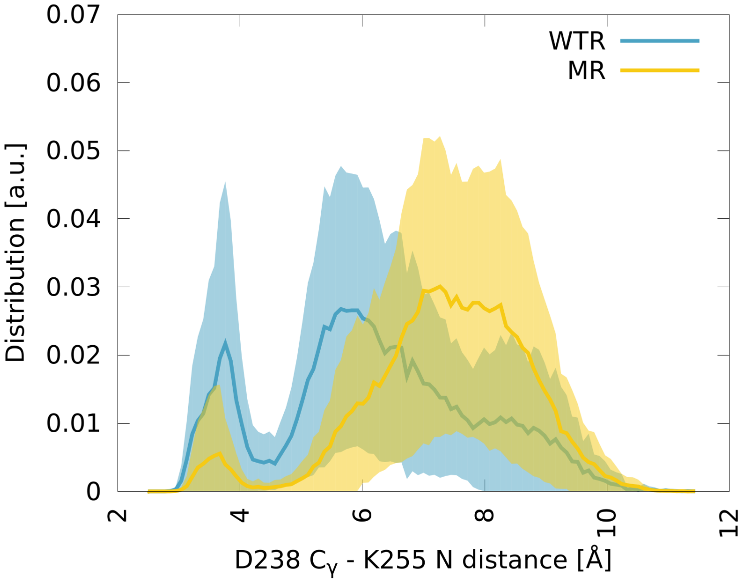


Figure 5.12: Distributions of the distance between the C_γ of D238 and the side chain nitrogen of K255, in the two models. Figure from reference [8].

Direct hydrogen bonds between Y441 and K255 are possible in the WTR, however they are barely within reach with each other and thus they only occurred for a tiny fraction of the simulation time explored. On the other hand, the D238 terminal oxygens formed hydrogen bonds for a considerable fraction of time. The distribution over time and subunits, evaluated over R01-400, of the distance between the C_γ of D238 and the side chain nitrogen of K255 showed two peaks in both the WTR and the MR (figure 5.12). One peak was observed around 7.5 Å for the MR and around 5.5 Å for the WTR, and proved how the absence of Y441 side chain allows for D238 to be farther away from K255 with respect to the WTR. The other peak was found around 3.5 Å for both the WTR and the MR, and was higher for the WTR than for the MR, reflecting the hydrogen bond between K255 and D238, which was observed for longer times in the WTR. Here, errors were calculated via error propagation from the ten standard deviations of data over time for each of the single subunits and for the two replicas R0 and R1.

Further experiments were made in order to assess the effect of K255 mutations in HEK293 cells.[8] While K255A behaves like the WTR, mutations to amino acids with shorter side chains that preserve the ability to form hydrogen bonds (namely C, E, and Q), resulted in a reduced EC_{50}. Moreover, the mutated K255L receptor, while expressed to cell surface, is non-responsive upon ligand binding. These observations suggested that K255 is in some measure implicated in the coupling effect of Y441 and D238.

5.6 Lipid Influence on the Y441 Mechanism

Experiments in *Xenopus* oocytes revealed how, unlike what observed in HEK293 cells, Y441A and K255L resulted in wild-type-like responses. This seemed to point out that the Y441-mediated coupling is not necessary in *Xenopus* oocytes, where a potential rescuing effect might take place, i.e. an effect that would compensate for the lack of side chain of Y441 by conveying a similarly-acting mechanism.

We deemed the membrane composition as a likely candidate for a rescue/inhibitory factor that differs between the two cases. Thus, we investigated possible lipid-related effects on Y441 and nearby residues in both the WTR and the MR. While the two models have a membrane of the same composition, and their starting conformation use the same lipid distribution around the protein, the local lipid environment around each of the five subunits would be different over time, given the random initial distribution and the stochastic nature of lipid diffusion during the simulations.

To evaluate the fitness of 441 as a binding site for lipid molecules, we calculated the proximity lifetime distributions of lipid molecules around it. This calculation was performed for the two replicas R0 and R1 separately, since the lifetime depends on the specific replica. For this calculation, we considered multiple time windows, shifted by 5 ns, and evaluated averages and standard deviations for each 5 ns time period, similarly to what performed in the previous chapter. The results are shown separately for phospholipids and cholesterol in figure 5.13: POPC and POPE were considered together, since residue 441 is at the level of their tails (as shown in the following), which are indistinguishable for POPC and POPE.

All lifetimes distributions were characterised by fast decays. Cholesterol was characterised by a single binding event that lasts for around 15 ns in replica R1 in the WTR, while phospholipids exhibited a certain number of binding events that lasted up to 25 ns. Nonetheless, no event was observed for any longer duration, this could be a hint that interactions with lipids at the level of 441 are feeble.

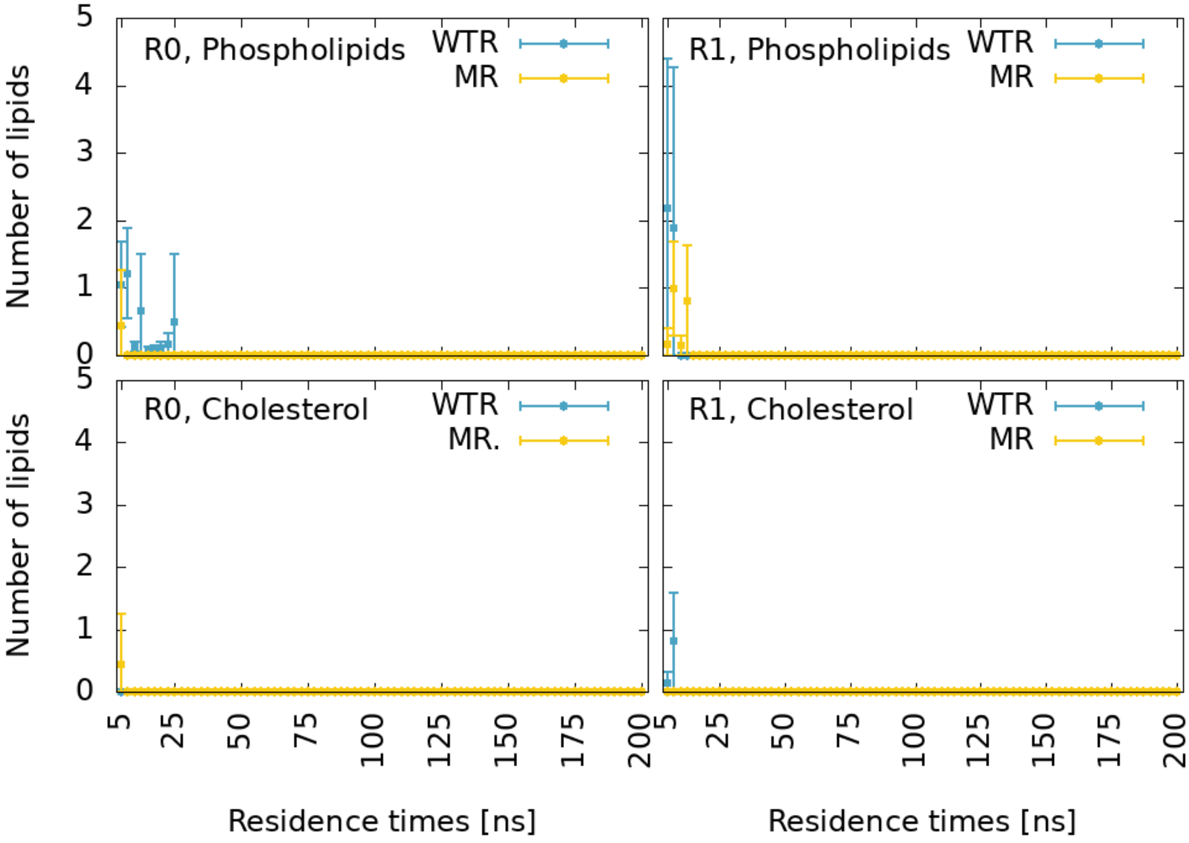


Figure 5.13: Proximity lifetimes distributions of any phospholipid (top) and of any cholesterol (bottom) close to residue 441 in the WTR model (blue) and in the MR model (yellow), and for replicas R0 (left) and R1 (right). The distributions are shown for residence times larger than 5 ns. Figure from reference [8].

We calculated the z component of the distance from the centre of mass of the five residues 441 to the centre of mass of lipids selections of the upper and lower leaflet: cholesterol centre of mass (COM), cholesterol oxygen (O1), phospholipid (PHOSP) centre of mass (COM), phospholipid phosphorus (P31), and phospholipid head nitrogen (N31). The results are shown for R01-400 in figure 5.14.

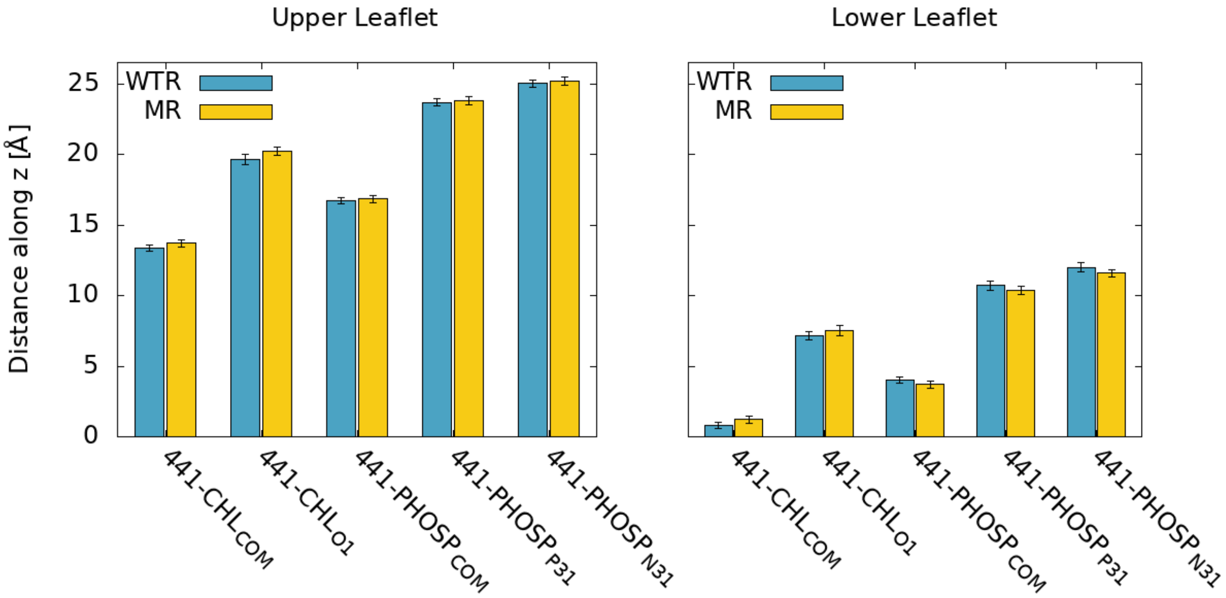


Figure 5.14: z component of the distance from residues 441 centre of mass of to that of lipids selections of the upper and lower leaflet, for the WTR (blue) and the MR (yellow). Figure from reference [8].

Information about the relative height of lipid molecules with respect to residue 441 can be used to make assumptions on the type of lipid binding events and interactions that can take place at the level of this amino-acid. The graph shows that this residue was found at the same height of cholesterol molecules centre of mass and phospholipid tails of the lower leaflet. Phospholipid tails could disrupt the interactions between residue 441 and D238, perhaps when intercalating within the subunit thanks to their flexibility (as observed in the previous chapter). Phospholipid heads are, on average, farther from residue 441, but could still form sporadic hydrogen bonds with this residue or with D238. Cholesterol, on the other hand, could form π-π interactions with the side chain of residue 441 in the WTR, hydrogen bonds with either residue 441 or D238, or could intercalate within the subunit. Cholesterol was previously observed interacting with 441 by means of π-π interactions as well as hydrogen bonds prior to intercalation within the subunit.[42]

The hydrogen bonds between residue 441 (or D238) and lipid molecules, calculated throughout replicas R0 and R1 for the two models, are shown in figure 5.15, and represented via MD snapshot in figure 5.16.

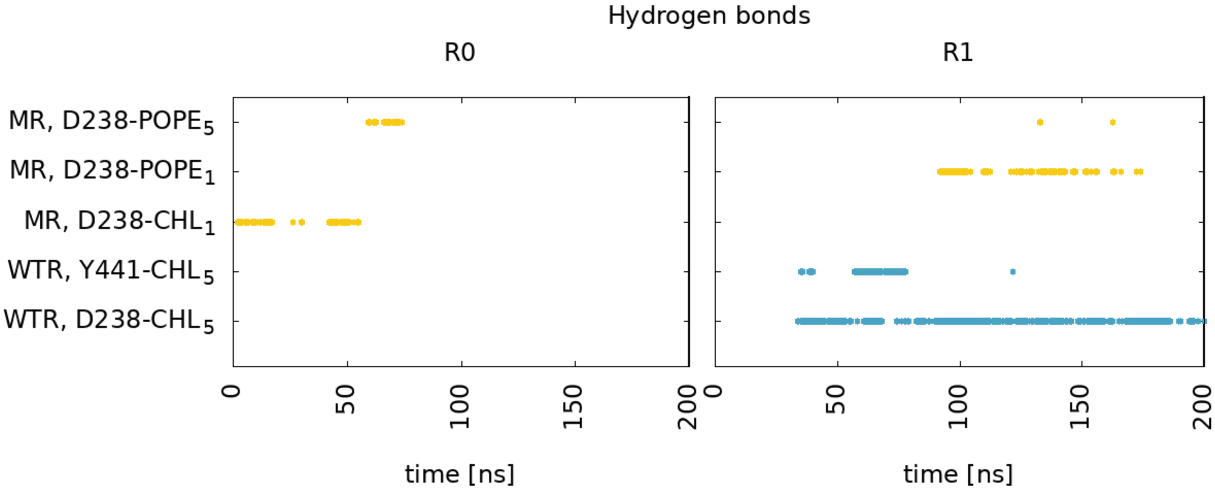


Figure 5.15: Hydrogen bonds between residue 441 (or D238) and lipids.

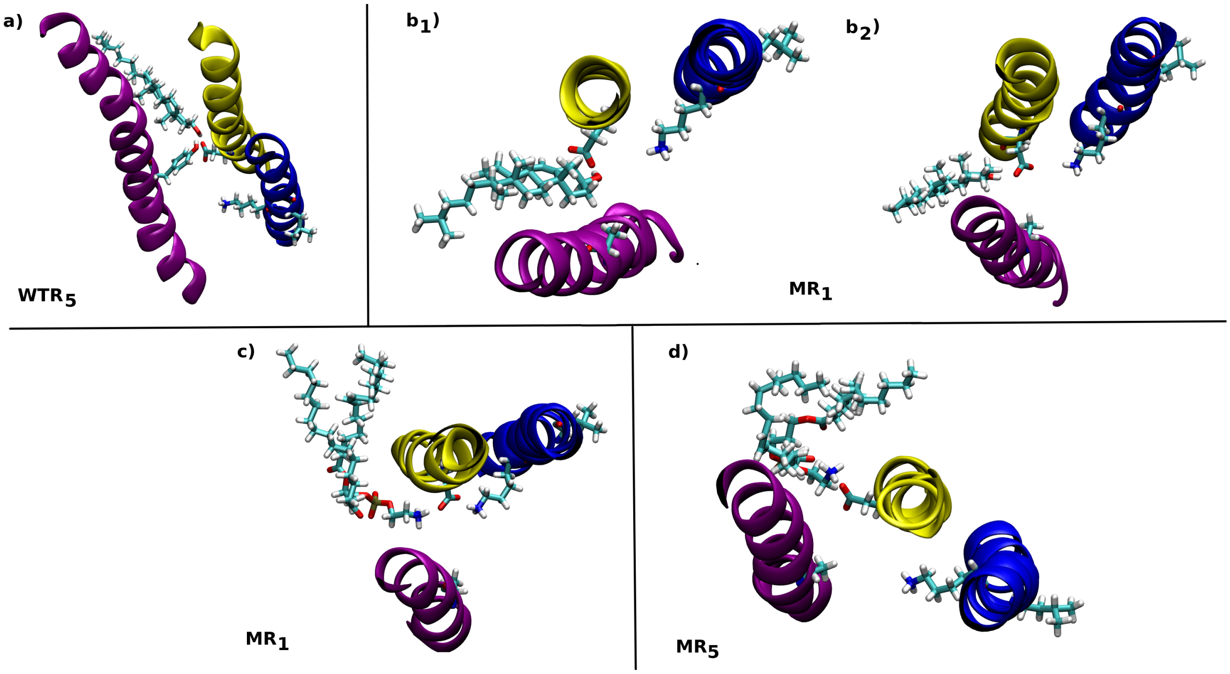


Figure 5.16: Representation of the hydrogen bonds between lipid molecules and residues 441 and 238 in the WTR and in the MR. Figure from reference [8].

In the simulations, only feeble hydrogen bonds were seen between lipid molecules and residues

441 and 238. In subunit 5 of the WTR, a POPE molecule formed hydrogen bonds with both Y441 and D238, pulling D238 away from K255 (figure 5.16 a). In the MR, other hydrogen bonds were observed. In subunit 1, a cholesterol molecule made hydrogen bonds, able to push D238 outward and away from K255 (figure 5.16 b_1), but this did not result in a pulling of D238 in other timeframes (figure 5.16 b_2). Besides, a POPE molecule in subunit 1 formed hydrogen bonds with D238 that did not prevent it from also interacting with K255 (figure 5.16 c). Another POPE molecule interacted with D238 of subunit 5, which instead resulted in pulling this residue outward (figure 5.16 d).

Thus, lipids interactions in this region appeared to be isolated phenomena, and may not necessarily promote nor inhibit the interactions between D238 and K255. Cholesterol, whose content in the membrane constitutes a major difference between HEK cells and oocytes, appeared to be able to interact with both Y441 and D238, but with unclear effect on the hydrogen bond between D238 and K255.

5.7 Conclusions and Future Perspectives

In this chapter, we aimed at elucidating the role and mechanism of the M4 helix on receptor function, by focusing on the uncoupling Y441A mutation, experimentally suggested as a possible proxy of the whole M4 coupling mechanism. We demonstrated that the M4 mediation, conveyed by Y441, does not occur by means of interactions of the M4 tip with the Cys-loop via a “vertically-propagating” mechanism, conversely our simulations support a “radially-propagating” mechanism of action through the M1 residue D238, which in turn interacts with the M2 residue K255.

The results obtained, supported by the experiments, lead to the suggestion that the Y441-mediated mechanism relies on the interactions between D238 and K255. These may help pull the hydrophobic gate (L260), located only 1.5 helical turns away from K255, in the process of opening. Mutations of residue 255 with shorter residues prone to form hydrogen bonds (C, E, and Q), might still allow for the M2 helices to bend away from the pore, reducing the work required to open the channel. However, in our simulations the M2 helices were restrained, thus preventing the check of any bending. As for the mutation K255A, there might still be ample room for the M2 helix to move on channel opening, thus resulting in wild-type-like behaviour. As for the mutation K255L, the bulky nature of the hydrophobic side chain of leucine may block outward M2 movements, thus preventing channel gating.

We argue that a functional mechanism involves the middle of the M4 helix, and, according to the experimental findings, is necessary in HEK cells but not in *Xenopus* oocytes. This would imply either that some factor present in oocytes can rescue these mutants, or that some other factor present in HEK cells inhibits the mutants, but not the wild-type receptors. Considering the closeness of the mutated residue 441 to the lipid membrane, together with the current knowledge that lipids interact with pLGICs (described in the previous chapter), it was reasonable to consider a lipid-related effect that would take place only in oocytes or HEK cells. Cholesterol in particular makes for a significant difference in the composition of oocytes membranes with respect to HEK cells, being more abundant in the former; besides, the addition of cholesterol to reconstituted membranes enhances pLGIC function.[178, 179]

Neither a rescuing factor that would be present in oocytes, nor an inhibitory factor present in HEK cells would necessarily be observable in the simulations presented here. This is because lipid interactions and other lipid-related effects depend on the local lipid environment around each of the five subunits, and only provide information on the role of the involved lipid species. From the simulations, our conclusion is that only sporadic lipid binding events and interactions took place at the level of Y441. Given the results, presented in the previous chapter, that cholesterol content enhances the likelihood of any lipid binding event to the surface of the receptor, we speculate that the membrane of oocytes, with higher cholesterol content, may promote additional binding events close to residue 441. However, we cannot conclude that these events would indeed be responsible for a difference in the mechanism induced by residue 441.

Future *in-silico* studies may focus on providing an explanation, at the atomistic level, of the working mechanism (or lack thereof) of other K255 mutations, or may try and shed light on the role of lipids in this area by means of longer simulations, in order to produce a larger statistics of possible lipid interactions with or near Y441. Enhanced sampling methods may also be employed for evaluating the free energy of lipid binding to or around residue 441.

In the next chapter, we will focus our attention on the ECD of the 5-HT$_{3A}$R, considering the same open, serotonin-bound model we built for this project. In particular, we will focus on the unbinding of 5-HT from its orthosteric pocket.

Chapter 6

The Unbinding of Serotonin from the Orthosteric Binding Site of the 5-HT$_{3A}$ Receptor

6.1 Introduction

In drug design, great importance is attributed to the phenomena of binding and unbinding of a ligand from the binding sites of receptors, not only because it is important to identify ligands with large affinity for the target proteins (measured by means of the *thermodynamic equilibrium binding constant* K_b), but also because the residence time in the pocket is significant.[180, 181] In fact, *in-vivo* drug selectivity is best achieved when the residence time is large for the target protein in comparison to other, non-desired ones.[182] However, engineering the chemical structure of a ligand in order to tune its unbinding kinetics remains challenging.

The *binding rate constant* k_{on} can be calculated with different methods, including molecular dynamics, brownian dynamics, and Markov State Models, while the *unbinding rate* k_{off} is more difficult, due to the presence of large free energy wells that lead to large unbinding times, making the process a rare event that unbiased MD simulation struggle to sample.[183] Moreover, it requires either bound structures available for the creation of computational models, or the usage of docking techniques to place the ligand inside the pocket in apo-structures. One way to evaluate it consists in estimating k_{on} first, and then obtaining it from the relation $k_{on}/k_{off} = K_b$. On top of the challenge of reconstructing the binding constant K_b from simulations, this method is also affected by large uncertainties. Alternatively, k_{off} can be calculated directly. The challenge here stems from the presence of large free energy wells, which result in large unbinding times. To tackle this problem, different methods have been proposed, including Markov State Models[184] and metadynamics.[67]

A standard protocol for studying the unbinding of a ligand from its pocket would start off with unbiased molecular dynamics simulations.[185, 183, 44] These would first assess the stability of said molecule and of the binding pocket, and provide a picture of the most important interactions and/or events that take place in the bound state. Afterwards, exploratory metadynamics can be used to identify slow-varying degrees of freedom that characterise the unbinding process. With this knowledge, several replicas of metadynamics can be produced, each stopping at the moment of unbinding. The metadynamics times can then be reweighted, so to reconstruct the unbinding time distribution. For such a process, under the condition of no bias being deposited at the transition state, the law of rare events holds: hence, if collective variables are correctly identified, the distribution is expected to be Poissonian.[77] Typically, several states (i.e. a collection of simulation frames, clustered according to the RMSD of the ligand, which are characterised by specific interactions) along the unbinding paths can be identified. A Markov State Model can then be constructed to calculate the transition rates among these states, by restarting new metadynamics replicas from each of the identified states. A distribution of transition times between each i-th and j-th state can be constructed, and thus an associated kinetic rate k_{ij}. Then, the different rates k_{ij} can be used to build a matrix of the full kinetic process: k_{off} is estimated as the fundamental relaxation rate of the kinetic model, i.e. the negative of the largest nonzero eigenvalue of the transpose matrix.[183, 186, 185, 187]

This protocol has been successfully used in the past.[183, 185] Casasnovas et al. studied the unbinding of an inhibitor of the pharmacologically relevant target p38 MAP kinase (a mitogen-activated enzyme),[183] reconstructing the kinetics of the whole process and providing an *in-silico* estimation of k_{off} in good agreement with its experimental value. In this work, unbinding metadynamics made use of two sets of CVs. In one case, the chosen CVs were the coordination number with water oxygens within 8 Å of the ligand centre of mass, together with the distance between the centres of mass of the pocket and of the ligand. In the other case, the path CVs s and z, as described for the first time by Branduardi et al.,[188] were employed, leading to similar results. Tiwary et al. performed a similar work on the trypsin-benzamidine complex, making use of the path CV s and a torsional angle that defines the orientation between the ligand and the pocket.[185] Tiwary et al. also studied the unbinding of the anticancer drug dasatinib from the c-Src kinase, by means of infrequent metadynamics, using the distance between the centre of mass of the ligand and the centre of mass of hydrogen bond-forming moieties of amino-acids of the pocket, together with the solvation state of the pocket, revealing the importance of water-mediated hydrogen bonding as a switch that controls unbinding.[187] The usage of infrequent metadynamics for ligand dissociation was benchmarked by Pramanik et al.,[74] proving how this method produces similar binding free energy profile, dissociation pathway, and ligand residence time as unbiased molecular dynamics, with an improvement in terms of computational cost.

For (un)binding problems, other methods can also be employed for the calculation of the binding/unbinding rate, or for estimating the associated free energy, which can provide useful insights into these processes. Free energy calculations can be performed with metadynamics (for example in the form of "funnel" metadynamics), steered molecular dynamics, umbrella sampling, and other methods. However, the reconstruction of the binding/unbinding free energy (for example in a metadynamics run) requires to thoroughly sample the process, leading to multiple crossings of the binding/unbinding free energy barrier in order to reach a diffusive state in a single run, unlike the multiple, first-passage runs required for the reconstruction of the kinetics of unbinding processes. Examples of works using metadynamics are listed in the following. Comitani et al. carried out funnel metadynamics simulations in order to study the binding free energy of γ-aminobutiric acid to the insect resistance to dieldrin (RDL) receptor, a pLGIC,[16] revealing the sequence of events leading to binding and highlighting the importance of charged residues (arginine and glutamic acid) for the capture of the zwitterionic ligand from the solvent and its positioning into the binding pocket. Brotzakis et al. computed the binding free energy and the unbinding rate in the benzamidine/trypsin system by combining a variational approach with funnel metadynamics.[189] Bernetti et al. combined Markov State Models and path metadynamics to calculate the binding free energy for the antagonist alprenolol to the β_2-adrenergic receptor, also identifying the minimum free energy path leading to the bound state.[190] Capelli et al. studied the unbinding of the agonist iperoxo to the human neuroreceptor M_2, reconstructing the associated free energy landscape by means of funnel metadynamics.[191] Capelli et al. also developed an extension to funnel metadynamics that makes use of spherical coordinates of the ligand centre of mass as CVs, thus probing different exiting paths. This methods was tested on a lysozyme T4 variant in complex with a benzene molecule, evaluating the binding free energy and identifying novel as well as previously reported pathways.[192] Haldar et al. proposed a method that combines enhanced sampling simulations and quantum mechanics/molecular mechanics (QM/MM) methods for the free energy and kinetics of binding, tested on the complex of the anticancer drug Imatinib and the Src kinase.[193] Zhou et al. introduced a variational-implicit solvent model for the study of (un)binding free energies, and described how it can be combined with the string method for obtaining the minimum energy paths and transition states between the various metastable hydration states.[194]

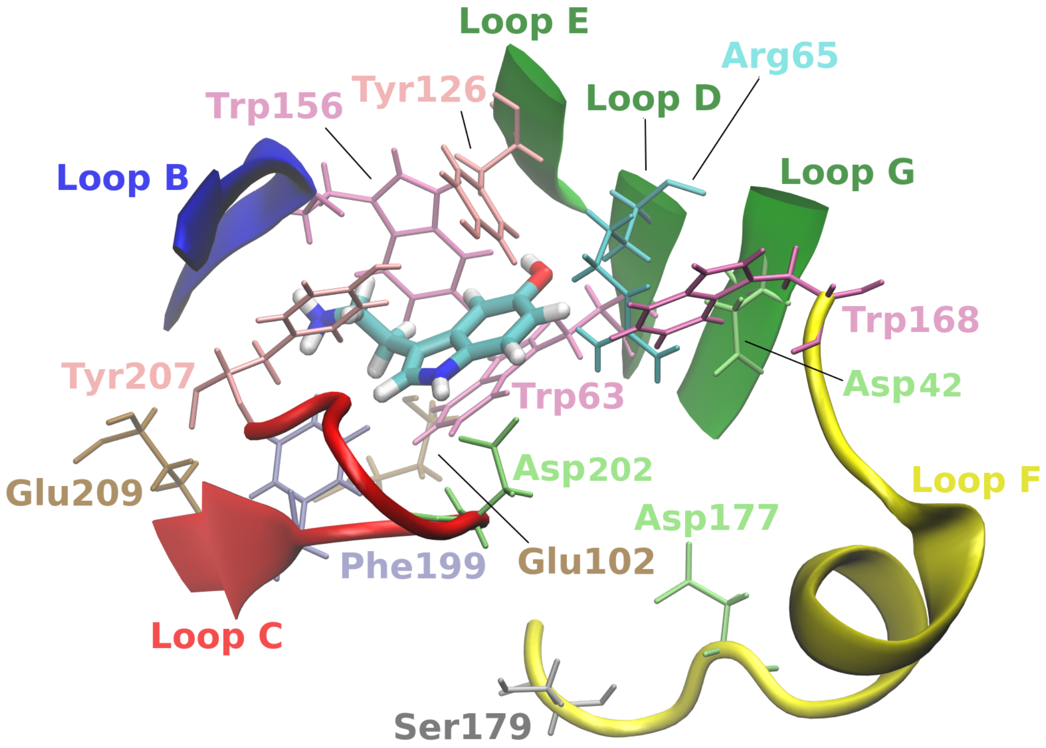


Figure 6.1: Representation of loops of the orthosteric binding pocket of the 5-TH$_{3A}$R (PDB entry 6DG8) and important amino-acids for 5-HT binding according to the literature: loop C (red), loop B (blue), loop F (yellow), loops G, D, E (green). Tyrosines are shown in pink, tryptophans in mauve, aspartic acids in lime, glutamic acids in brown, arginines in cyan, phenylalanines in ice blue, and serines in silver.

In this chapter, we focus our investigation on the behaviour of serotonin (5-HT) in the orthosteric pocket of the 5-HT$_{3A}$R, located in the ECD in between adjacent subunits (as depicted in figure

1.10 in the introductory chapter). Our goal is to characterise its interactions within this site, and to set up a preliminary set of simulations that would provide the basis for a future study of the kinetics of unbinding for serotonin from the 5-HT$_{3A}$R, and for the calculations of the associated k_{off}. To date, no experimental value is yet available for this quantity for 5-HT and the 5-HT$_{3A}$R.

Previous studies on the key interactions between 5-HT and the pocket are present in literature. According to mutagenesis experimental data, Glu129 (here Glu102) of Loop A interacts with 5-HT,[195] as well as Trp183 (here Trp156) of Loop B,[196, 197] Tyr234 (here Tyr207) of Loop C would provide part of the aromatic box of the pocket,[198] while several residues of loops D, E and F would contribute to the binding of the competitive antagonist granisetron,[199] among which Trp63 and Arg65 (loop D),[200] and Trp168, Asp177 and Ser179 (loop F).[201] The cocrystal structure of the 5-HT$_3$R-mimicking acetylcholine binding protein in complex with the anti-emetic drug palonosetron revealed the importance of residues Trp156 and Tyr126, the former interacting by means of a cation-π interaction and the latter interacting via hydrogen bonds.[202] Docking techniques proposed Tyr143 (here Tyr116), Tyr234 (here Tyr207), Trp183 (here Trp156), Tyr153 (here Tyr126), as relevant residues for 5-HT binding.[203] An *ab-initio* study identified three possible models for the interactions network: model A with Tyr153 (here 146), Lys154 (here 147), and Asn152 (here 145); model B with Tyr143 (here 136), Val142 (here 135), and Val144 (here 137); and model C with Ala235 (here 228), Ser233 (here 226), and Tyr234 (here 227).[204] A further *in-silico* work studied the binding pocket by docking 5-HT to the crystal structure of the 5-HT$_{3A}$ receptor (PDB entry: 4PIR), and performing subsequent molecular dynamics simulations.[44] This identified key interactions with Tyr126, Trp156, Arg65, Trp63, Glu209, Phe199, and Tyr207. Moreover, microseconds-long simulations of the crystal structure of the 5-HT$_{3A}$ receptor in complex with a multitude of 5-HT molecules revealed a number of binding events into the pocket,[42] highlighting interactions with Glu209, Arg65, Tyr207, Phe199, Glu102, Trp63, Asp202, Trp156, and also showing how a salt-bridge between Asp202 and Arg65 helps stabilising loop C. Recently, molecular dynamics simulations were performed to assess the binding, in the 5-HT$_{3A}$R, of the antagonists palonosetron, granisetron, dolasetron, ondansetron, and cilansetron, highlighting the importance of residues Glu102 (Loop A), Trp156 (Loop B), Tyr207 and Asp202 (Loop C), Trp63 and Arg65 (Loop D), Asp42 (Loop G), Tyr126 (Loop E), and Trp168, Pro171, Leu167, Trp168, Arg169, Asp177, Lys178 (Loop F).[205]

Overall, there is consensus on what are the key residues for serotonin binding in the 5-HT$_{3A}$ receptor. Figure 6.1 provides a representation of some of the most important residues of the orthosteric pocket according to the literature.

6.2 Models and Unbiased Simulations

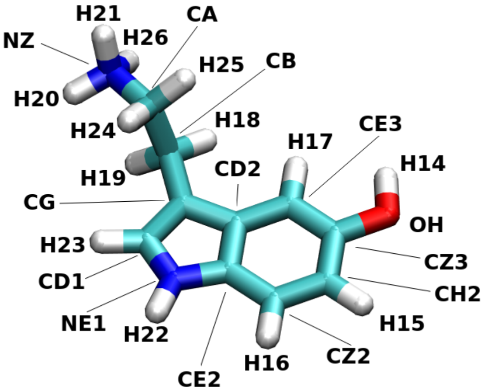


Figure 6.2: Representation of serotonin, with atom names.

In order to assess stability and interactions of serotonin in the orthosteric binding pocket, we first performed molecular dynamics simulations. The model and the MD simulations are the same described in the previous chapter. The model is based on the open, serotonin-bound cryo-EM structure (PDB entry: 6DG8).[5] This structure was ideal for the scope, as serotonin is bound and we did not need to apply approximate Docking methods to place it inside the orthosteric pocket. We made use of both the wild-type receptor (WTR) and the Y441A mutated receptor (MR), each with two NPT replicas of 250 ns each (with the first 50 ns considered as equilibration). Since the Y441A mutation is the only feature that distinguishes the MR from the MR, and is located in the TMD far away from the orthosteric binding pocket, we assumed that it would not have any significant influence on the serotonin interactions and therefore all four replicas could be used for statistics. Besides, having one serotonin in each of the five pockets (all starting from the same conformation, in the cryo-EM structure) also allows for a large statistics. Except for those quantities which are explicit functions of time, all analysis were performed on the conjunction of the four 200-ns-long time windows of the four replicas (R4).

6.2.1 Serotonin Dynamics

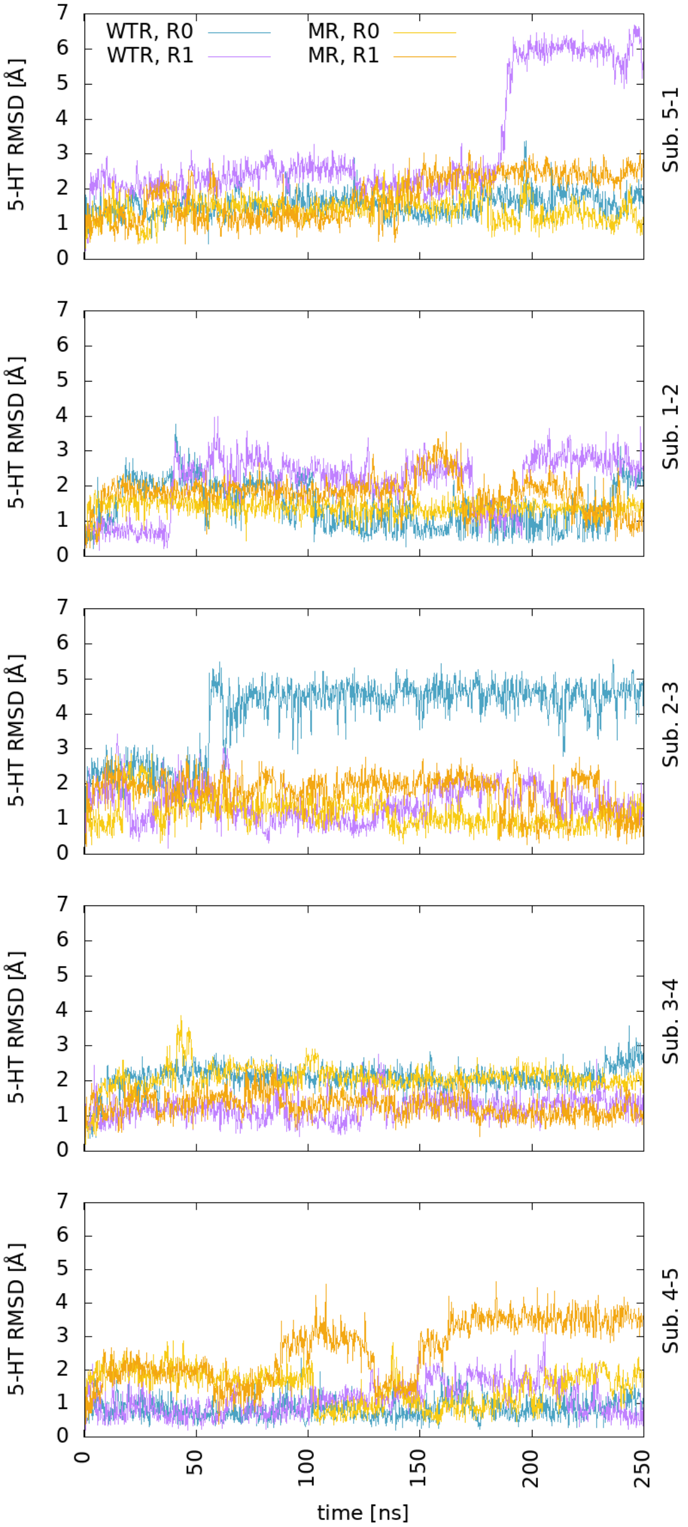


Figure 6.3: RMSD of 5-HT rings at the interface between subunits (5 and 1, 1 and 2, 2 and 3, 3 and 4, and 4 and 5) in the WTR and in the MR, replicas R0 and R1.

We first assessed the stability of serotonin in the pocket in the MD simulations, by studying the RMSD of its aromatic atoms over the four replicas, and considered separately for each subunit interface, with respect to the first time step of each 250-ns-long replica. The results are shown in figure 6.3.

No 5-HT molecule spontaneously exited the binding pocket in any of the simulations, and for the most part of the simulations were stable in the pocket, as the initial pose is kept for the vast part of simulation time in all replicas. However, different stable poses were identified, corresponding to the plateaus at higher RMSD in subunits 5-1, WTR replica R1, subunits 2-3, WTR replica R0, and subunits 4-5, MR replica R1. These poses will be discussed in the following.

Next, we assessed the stability of the pocket itself, by looking at the RMSD of the backbone atoms of the pocket main loops: B, C, D, E, F, G. The results are separately considered for the various replicas with respect to the first production step, and shown in figure 6.4.

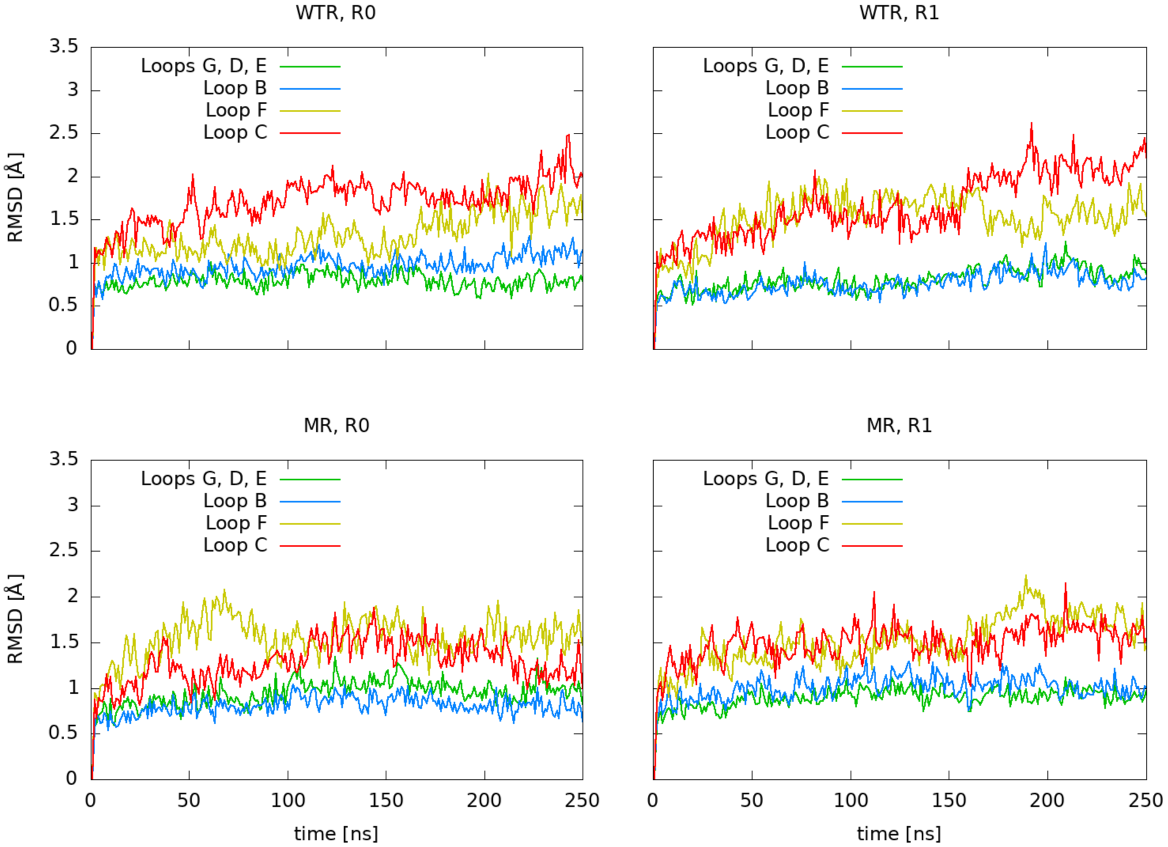
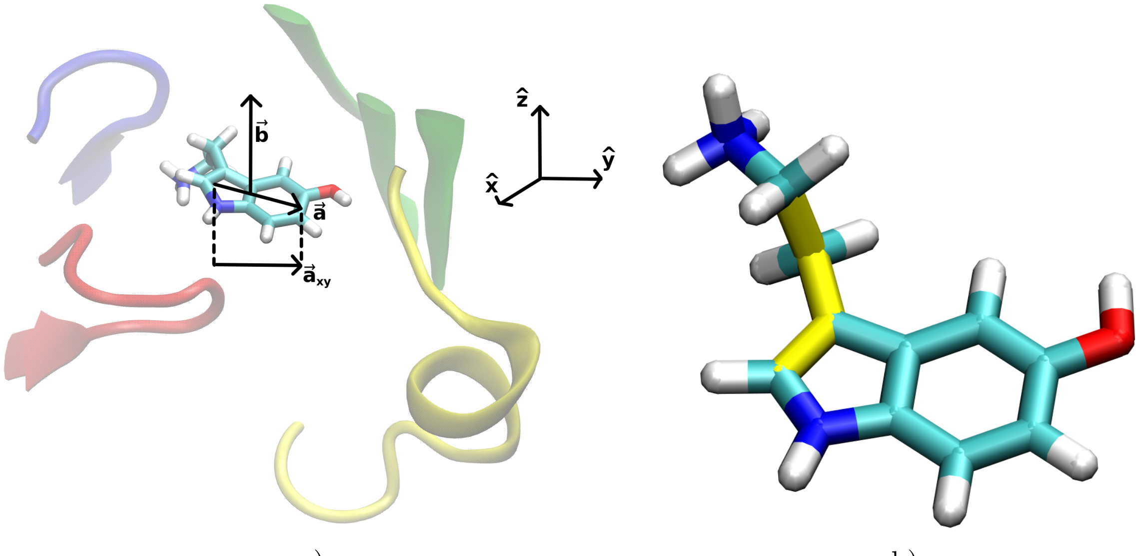


Figure 6.4: RMSD of loops B, C, D, E, F, G in the WTR and MR, replicas R0 and R1. Loops G, D, E are grouped together considering their closeness.

While all loops are overall stable in every replica, loop C and loop F result more dynamical then the others. Loop C, in particular, has been previously suggested as a flexible loop that acts a shield from the external solvent, and that opens and closes during the binding/unbinding process.[206, 9, 16]

To further assess the dynamics of 5-HT within the pocket, we looked at the fluctuations of the angles formed by 5-HT main axis (defined as the longer axis that crosses its two aromatic rings) with the x- and the z-axis. The vectors used for defining these angles are shown in figure 6.5 a: $\vec{a}$ is the vector parallel to 5-HT aromatic rings main axis, $\vec{a_{xy}}$ is its projection onto the xy-plane, and $\vec{b}$ is the vector normal to the aromatic rings. We considered two angles, one between $\vec{a_{xy}}$ and the x-axis, and one between $\vec{b}$ and the z-axis. The fluctuations (i.e. the standard deviations) of these angles, considered over R4, were around 8° and around 1° respectively, demonstrating that the orientation of 5-HT with respect to the pocket residues is well maintained during the MD simulations. Moreover, we investigated the fluctuations of the dihedral angle (represented in figure 6.5 b) of 5-HT tail to enquire about its torsion, which resulted around 40°, proving how 5-HT tail has some freedom to rotate even when 5-HT is in its stable pose.



a) b)

Figure 6.5: a) Schematic representation of vectors $\vec{a}$ (parallel to 5-HT principal axis), its projection $\vec{a_{xy}}$ onto the xy-plane, and $\vec{b}$ (normal to 5-HT aromatic rings). z is parallel to the protein main axis. b) Dihedral angle (in yellow) chosen to investigate torsions of the tail of 5-HT.

6.2.2 Serotonin Interactions in the Binding Pocket

Interactions are the way a ligand triggers structural modifications within the protein, and help stabilising it in the orthosteric pocket. Thus, for (un)binding problems, the understanding of the main interactions that occur between the ligand and the residues of the pocket is of paramount importance. This can be easily achieved by studying a bound structure with unbiased molecular dynamics simulations.

Serotonin can engage in hydrogen bonds through the NH_3 group on its tail (whose nitrogen is here labelled as NZ), its hydroxyl group (labelled as OH) and its amine group (whose nitrogen is labelled as NE1) (all 5-HT labels are shown in figure 6.2). It can also engage in π interactions thanks to its positively-charged tail (which provides the cation in cation-π interactions) and its two aromatic rings (which provide the aromatic part in π-π, cation-π, and anion-π interactions).

Figure 6.6 reports the time-, subunit-, and replica-averaged hydrogen bonds between serotonin and the residues of the pocket, while figure 6.7 reports the water-mediated hydrogen bonds with the pocket. Both were calculated over R4. Errors were evaluated as standard deviations over the time-averaged and replica-averaged data, and propagated over the five subunits.

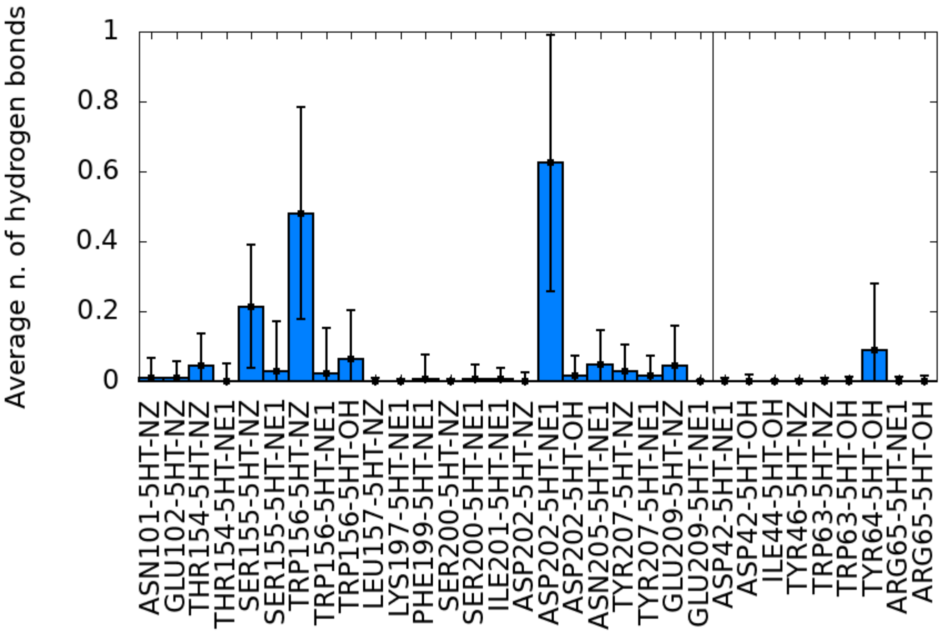


Figure 6.6: Average number of hydrogen bonds between 5-HT and the residues of the pocket.

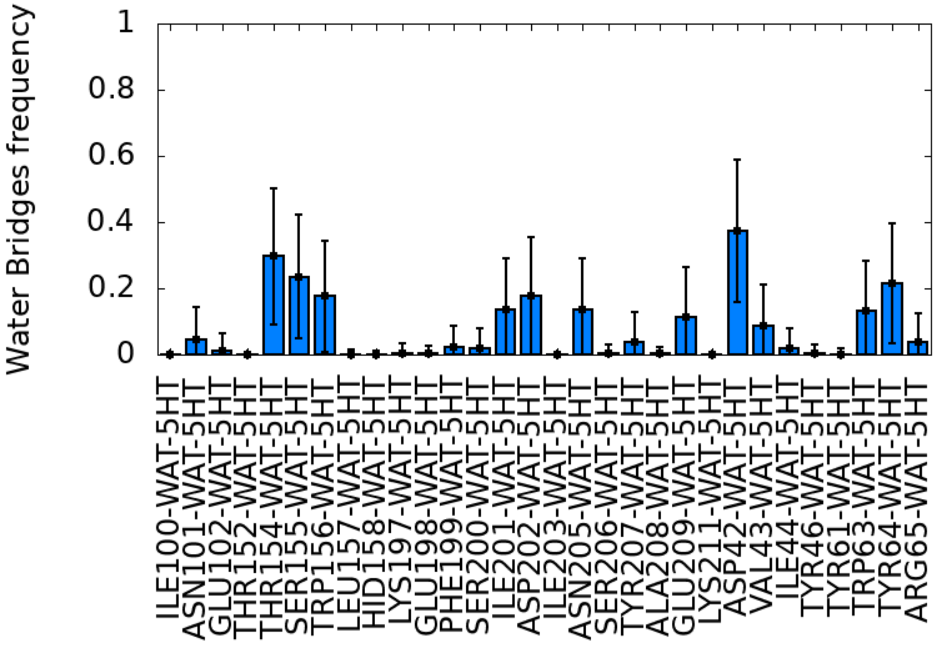


Figure 6.7: Average number of water-mediated hydrogen bonds between 5-HT and the residues of the pocket.

Serotonin was able to engage in a variety of direct hydrogen bonds with the surrounding pocket, implying that some variation was observed in the explored replicas. The most prominent of these bonds (around 0.6) is between 5-HT aromatic nitrogen and Asp202, which belongs to Loop C. Other relevant hydrogen bonds were formed by 5-HT tail with Ser155 and Trp156 (around 0.2 and 0.45 respectively), which belong to Loop B. Several water-mediated hydrogen bond bridges were also observed, in particular with residues Thr154, Ser155, Trp156, Ile201, Asp202, Asn205, Glu209, Trp63, Tyr64, and most prominently with Asp42. Thus, water molecules may also contribute to the stabilisation of 5-HT in the pocket by allowing for the formation of bridges with the highlighted residues.

Figure 6.8 shows the time-, subunit-, and replica- averages of aromatic interactions that 5-HT formed with the residues of the pocket, calculated over R4. A few π-π interactions were observed, in particular with Trp63 (loop D) and Tyr126 (loop E), which lie right above and right beneath the ligand aromatic rings, respectively. Less prominent π-π interactions were observed with Trp156 (loop B), whose aromatic side chains lies side-ways with respect to 5-HT, in the innermost part of the pocket; with Phe199 (loop C), and with Tyr207 (at the very end of loop C). A relevant cation-π

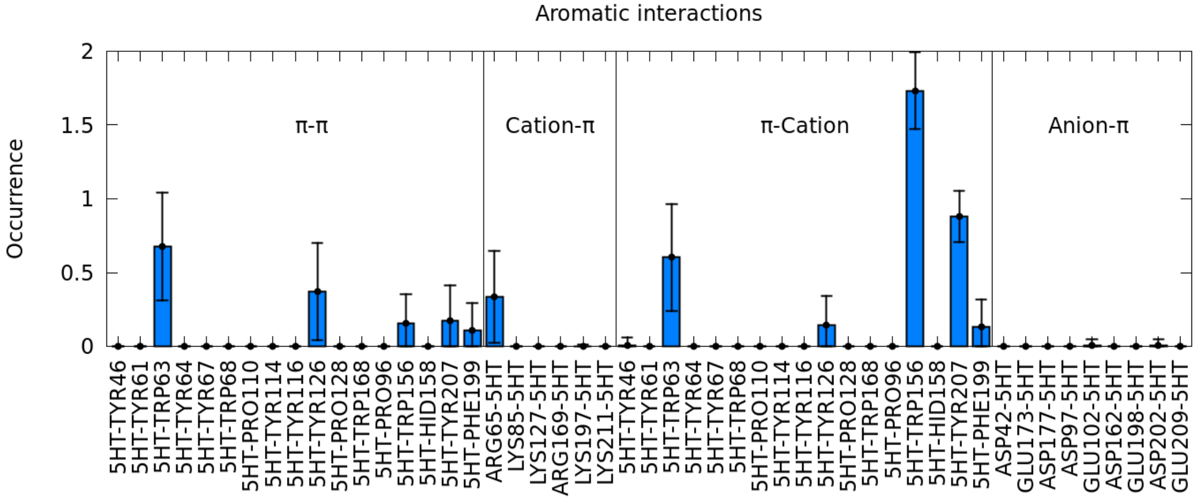


Figure 6.8: Average aromatic interactions formed by 5-HT with the residues in the pocket.

interaction was observed with Arg65 (loop D). Several other cation-π interactions, for which 5-HT provides the positive moiety (referred to as "π-cation" interactions in figure 6.8), resulted important: Trp63, when its side chain moves away from the space below 5-HT aromatic rings, Tyr207, and most importantly Trp156, whose two aromatic rings both form a persistent interaction with 5-HT tail. Tyr126 and Phe199 also engaged in π-cation interactions, although less prominently. Cation-π interactions were similarly found as relevant for the stabilisation of ligands in their pocket in other pLGICs.[207, 16]

The most relevant interactions that we found contribute to the stabilisation of a main binding pose (pose "I"), shown in figure 6.9. Our findings agree for the most part with past studies: we found matches in the interactions with Tyr126,[203, 44], Trp156,[203, 44, 42, 199, 205], Asp202,[42, 205], Arg65,[44, 42, 199, 205] Trp63[44, 42, 199, 205], and Tyr207.[203, 44, 42, 199, 205] In our model, Phe199 appears to form little to no direct and indirect hydrogen bonds, while forming π-π interactions for a short fraction of time. Glu102 was not found to interact with 5-HT. Residues of the more external Loop F were not found to play a role in stabilising 5-HT in the pocket, but will be shown to play a role in the unbinding process in the following.

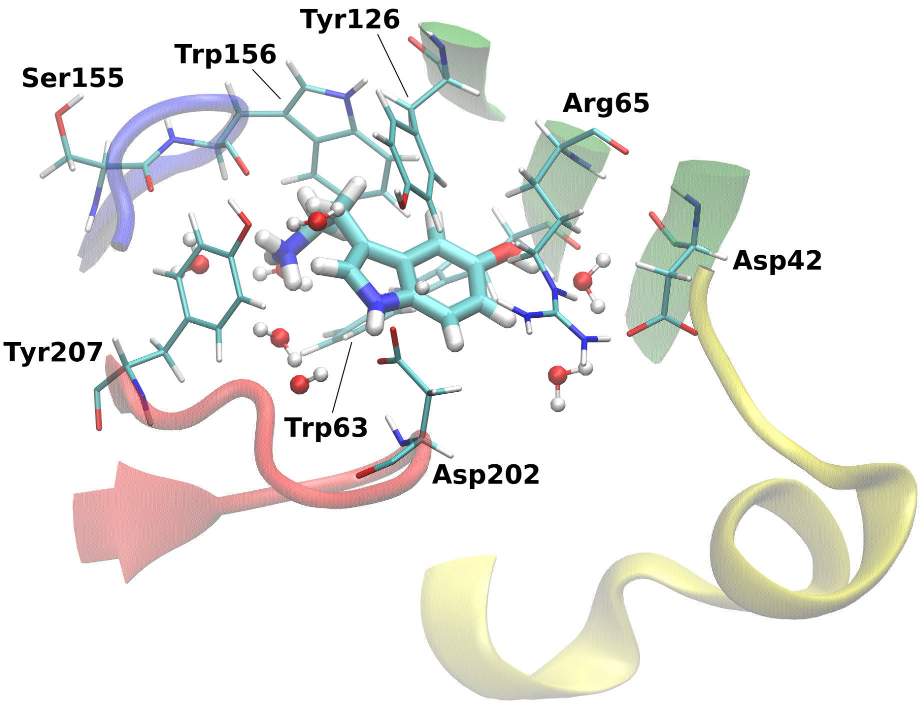


Figure 6.9: Representation of serotonin in the binding pocket, together with some of the most relevant residues for interactions and the closest water molecules (pose "I").

Figure 6.10 shows the minor (but still stable) poses of 5-HT in the pocket, corresponding to the cases of higher RMSD: one pose (pose"II") observed in the WTR replica R1, subunits 5-1, and the other pose (pose "III") observed in the WTR replica R0, subunits 2-3, and in the MR replica R1, subunits 4-5. Both were caused by random fluctuations that resulted in the formation of alternative interactions. In the first pose, while the tail of 5-HT kept its hydrogen bonds with the residues of Loop B, the aromatic rings turned in such a way that the hydroxyl interacted with Asp202. Conversely, in the second pose, 5-HT rotated around an axis parallel to its aromatic rings, making its tail break its interactions. This pose was stabilised by a hydrogen bond between the aromatic nitrogen and the residues of Loop B, while keeping the aromatic interactions.

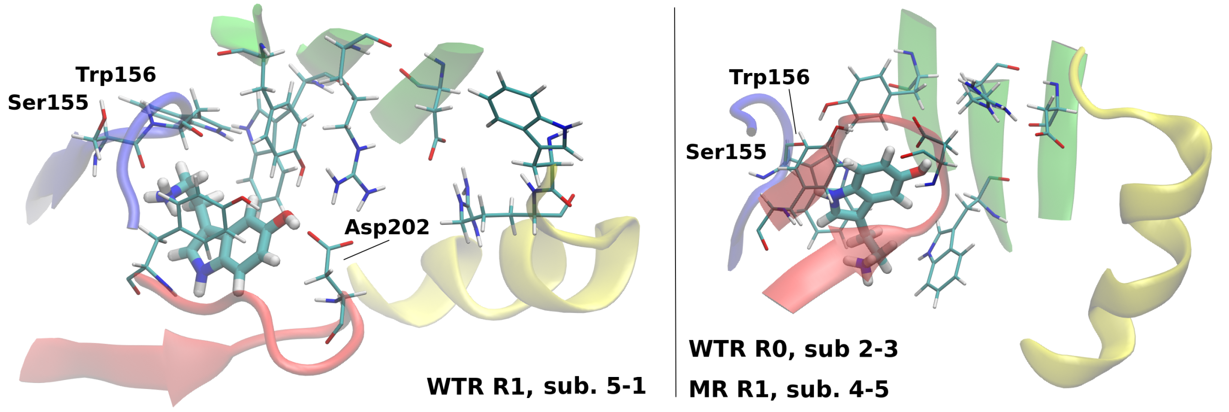


Figure 6.10: Representation of minor binding poses of 5-HT in the pocket: pose “II” (left) and “III” (right).

As previously said, the interactions between the ligand and the surrounding pocket constitute a major factor for the stability of the former into the latter: thus, any unbinding process would start with their breaking, as they give rise to a free energy well that separates the bound from the unbound state. It is reasonable, then, to consider certain distances between interacting moieties of 5-HT to those of the relevant pocket residues as candidate collective variables (CVs) for a biased run to enhance the sampling of the unbinding process. The literature suggests that for many systems one CV can be taken as the distance between the centre of mass of the ligand and the centre of mass of the atoms in the pocket that give rise to the most stable interactions.[183, 187] We defined a similar CV for our system, d_P, as the distance between the centre of mass of the two aromatic rings of 5-HT, and the centre of mass of the interacting moieties of the most important residues of the pocket: in particular, we considered the two side chain oxygens of Asp202, the aromatic rings atoms of Trp63, the carbonyl oxygens of Ser155 and of Trp156, and the aromatic rings atoms of Tyr126. The choice of these particular residues is not unique, as there are others which give rise to significant interactions as well; this particular selection was picked considering the closeness of its centre of mass of that of 5-HT. The evolution of d_P over the production time in the four replicas and for each subunit pair is shown by the left-hand panels of figure 6.11.

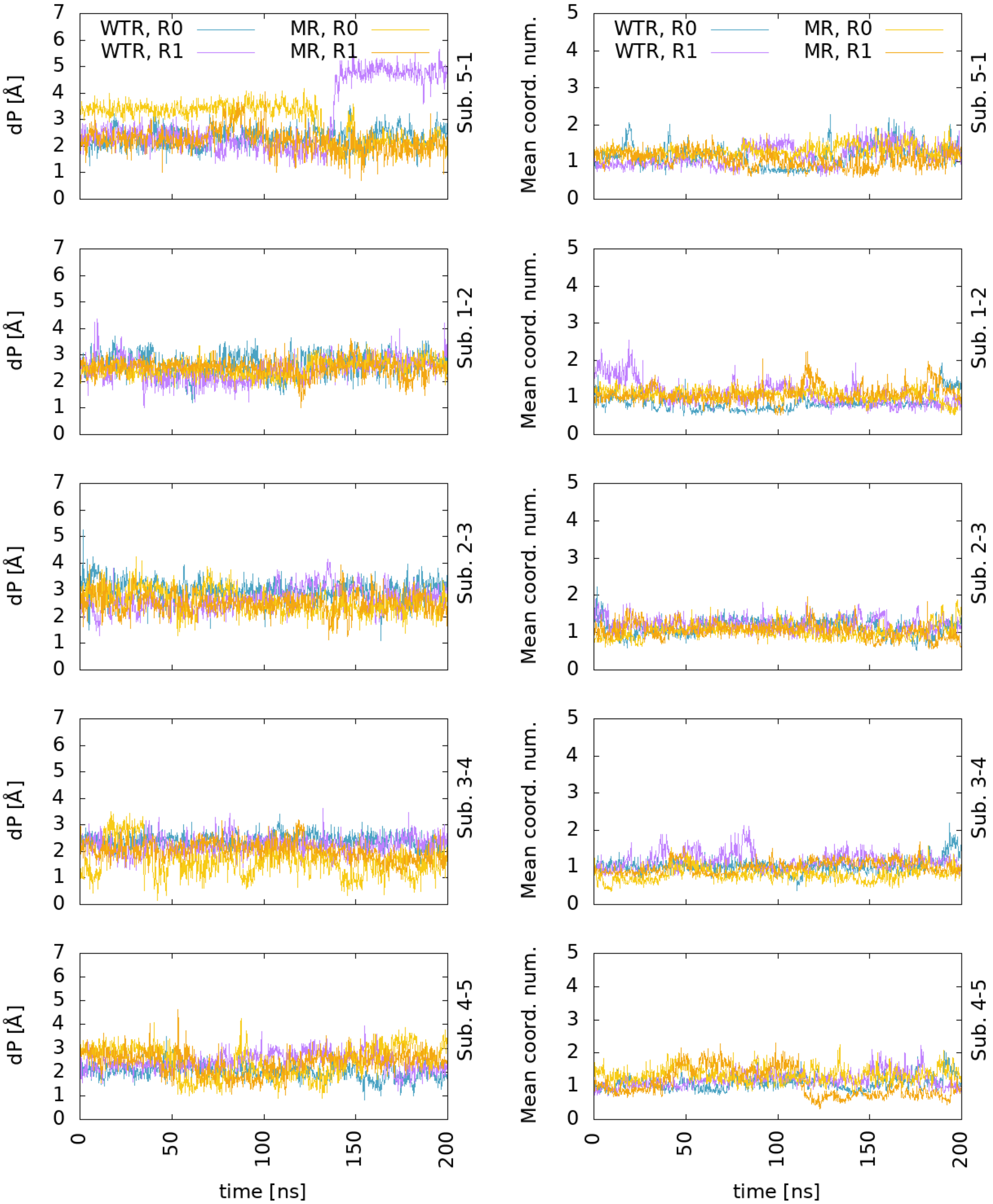


Figure 6.11: Left panels: time evolution of d_P in WTR and MR, in replicas R0 and R1. Right panels: mean coordination number of water oxygens within 3.5 Å from the five 5-HTs in the WTR and MR, in replicas R0 and R1.

This potential CV was stable in most of the observed cases, however some instability was exhibited by WTR replica R1 subunits 5-1. As seen in this subunit pair (for MR replica R0 and especially for WTR replica R1), d_P may be characterised by more than one equilibrium value when the system is in its bound state. This is consistent with what observed in the RMSD of 5-HT aromatic rings (i.e. the presence of one main binding pose, and two minor ones); however the fluctuations of the RMSD are not necessarily related to the shifts in the position of the ligand centre of mass with respect to that of the chosen pocket residues, as they are also linked to rotations. Finally, its time-, subunit- and model-average resulted 2.4 ± 0.6 Å.

We finally analysed solvation of 5-HT in the pocket (waters in the pocket are also shown in figure 6.9). We did this first by inspecting the number of water oxygens withing a certain distance (e.g. 3.5 Å) from the 5-HT heavy atoms in the pocket, finding an average of 3.9 ± 0.5. The number of water oxygens, however, as a discrete variable cannot constitute a potential CV for an enhanced sampling run, so we used a smooth cutoff function as typical in these cases, described in Appendix A. With this, we then calculated the mean coordination number μ (also explained in Appendix A) of 5-HT and water oxygens calculated over the two models, the two replicas, and the five subunit pairs (shown by the right-hand panels of figure 6.11). This quantity, which is scaled by the smooth cutoff function so it does not represent the actual number of water oxygens, appeared to be stable in the various replicas.

In summary, our unbiased simulations revealed the main interactions between 5-HT and the residues of the pocket, largely agreeing with both experiments and previous *in-silico* studies, and were used to suggest a CV to be used in exploratory biased sampling of the unbinding process, discussed in the following.

6.3 Exploratory Metadynamics in a Simplified Model

We then turned our attention to a biased sampling of the unbinding process. In order to do this, we first created a simplified model that only contains the extracellular domain, solvated in a orthorombic box of water with a buffer of 15 Å, together with Na^+ and Cl^- ions. The cut of the structure was performed in a conformation of the WTR after 50 ns of replica R0, after residue 220, i.e. the first residue of the M1 helix. In order to preserve the stability of the model, harmonic restraints were applied to the backbone atoms of the last five residues (residues 216 to 220) for each subunit, with force constant 1.0 kcal/mol$\AA^2$, as previously done for models of pLGICs ECDs.[207]

We first calculated the RMSD of the ECD and of the rings of five 5-HTs, in a trial 100-ns-long MD simulation of the simplified model, shown in figure 6.12 and in figure 6.13 respectively. The RMSD of 5-HT resulted slightly larger in the reduced model with respect to the full-receptor model, but still proved the stability of the ligand in the pocket (which was also stable).

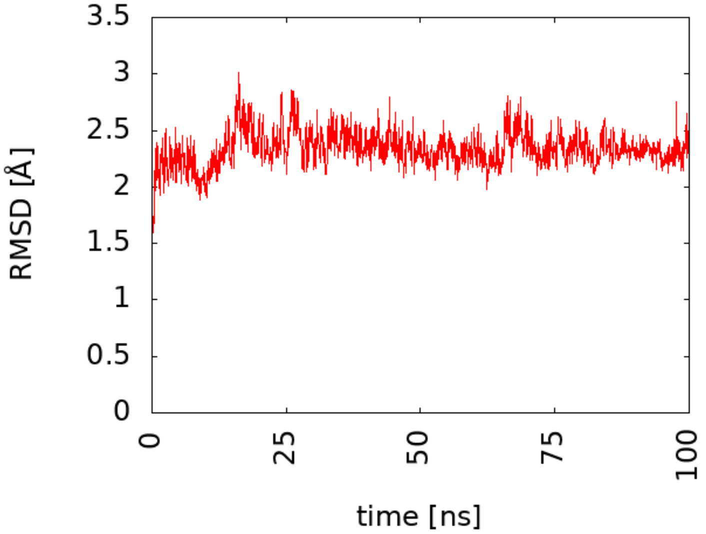


Figure 6.12: RMSD of the ECD as a function of time in the simplified model.

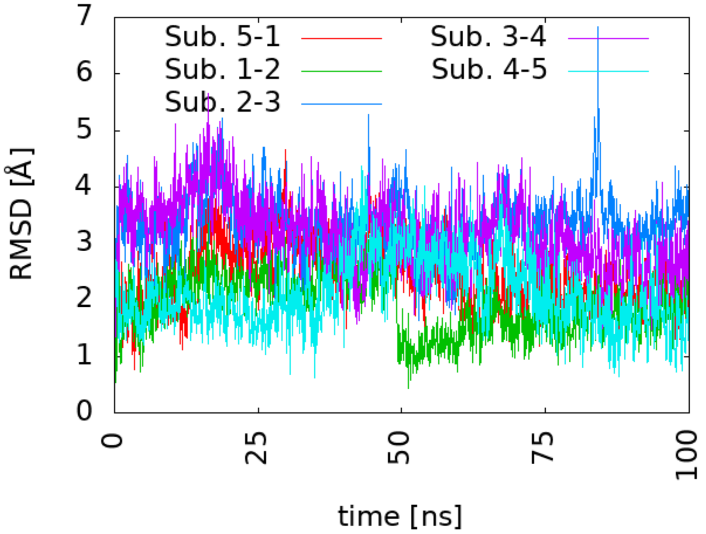


Figure 6.13: RMSD of the rings of the five 5-HTs as a function of time in the simplified model.

We then performed 12 exploratory well-tempered metadynamics runs, labelled A to L. These were perfomed with NAMD 2.13[80] patched with plumed 2.5[101] in an NVT ensemble (using a volume of about 114 Å by 113 Å by 104 Å). The temperature was controlled by means of a Langevin thermostat with a collision frequency of 1.0 ps^{-1}. In order to force the expulsion of 5-HT from its pocket, we made use of a single collective variable, the previously defined d_P. We enhanced the sampling of the process of unbinding for one of the five 5-HTs, specifically the one originally found in the binding pocket between subunits 3 and 4, starting from a conformation obtained after 50 ns of unrestrained equilibration. We used an initial Gaussian height of 0.2 kcal/mol (in accordance to previous works using a similar CV[183]), a bias factor of 15, and a Gaussian width of 0.1 Å chosen as a fraction of the observed fluctuations of d_P in MD. The adaptive-frequency scheme was adopted, with a starting deposition rate of 1 ps, a θ of 200,000 (as prescribed by Wang et al. for a simulation estimated to result in ligand unbinding at around 50 ns[208], described in the second chapter), and a τ_c of 10 ps (a value used by Casasnovas et al. in a past study of ligand unbinding[183]). The 12 runs reached the unbound state at different metadynamics time, but within 70 ns for all of them.

Figure 6.14 shows d_P as well as the mean coordination number μ as functions of time in the exploratory runs. d_P oscillated around separate values when 5-HT is still in the bound state; the low d_P values depending on 5-HT being, at times, fully "sandwiched" by Tyr126 and Trp63 (whose rings are intimately related to the definition of d_P). At long metadynamics times, d_P finally increased dramatically when 5-HT was finally lead to unbinding. μ was plotted as a chaperone quantity: in runs A, D and H, it oscillated around the full-solvation value previously described, indicating that 5-HT was fully immersed in the solvent. In the other runs, while the 5-HT molecule was indeed unbound, it remained in proximity of the protein surface. Indeed, to define the exact instant of unbinding remains an issue, as there is no consensus on this matter in the literature.[183, 187] One may consider the instant when the ligand reaches full solvation, although this may have problems if it remains bound to the protein surface; one may consider a cut-off of the ligand-pocket distance; or may consider when the main interactions between ligand and pocket are broken.

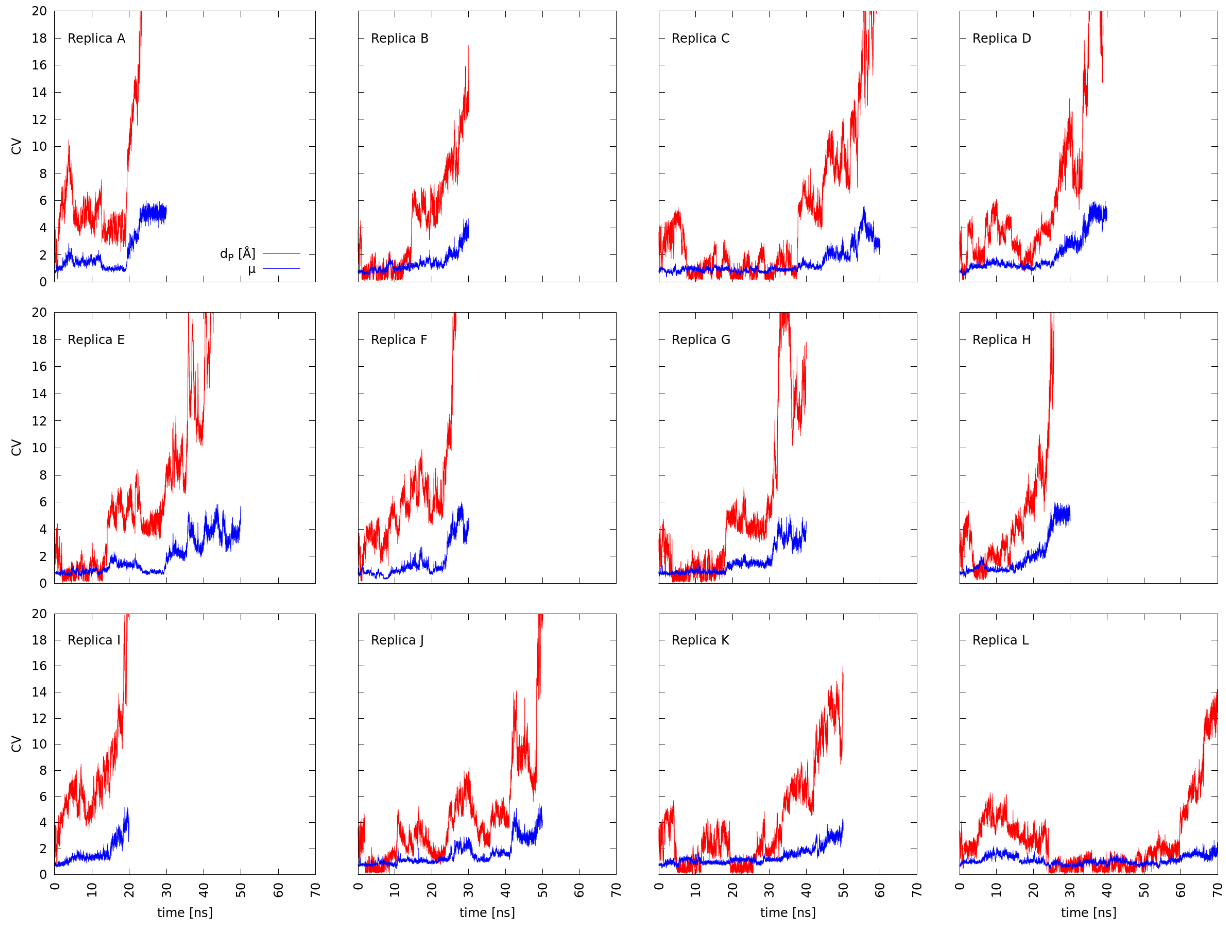


Figure 6.14: Evolution of the CV d_P and of the mean coordination number μ in the exploratory metadynamics runs.

6.3.1 The Free Energy Wells and the Breaking of the Interactions

We then evaluated the potential of mean force (PMF), i.e. a sum of the deposited Gaussians as a function of the CV. While this is not expected to be converged (as this would require several bindings and unbindings in the same metadynamics run), it is a reasonable way to preliminarily evaluate the free energy well associated with the breaking of all stabilising interactions, that separates the bound states (at low values of d_P) from the unbound one (characterised by a plateau at large values of d_P). The average PMF of the various runs is shown in figure 6.15.

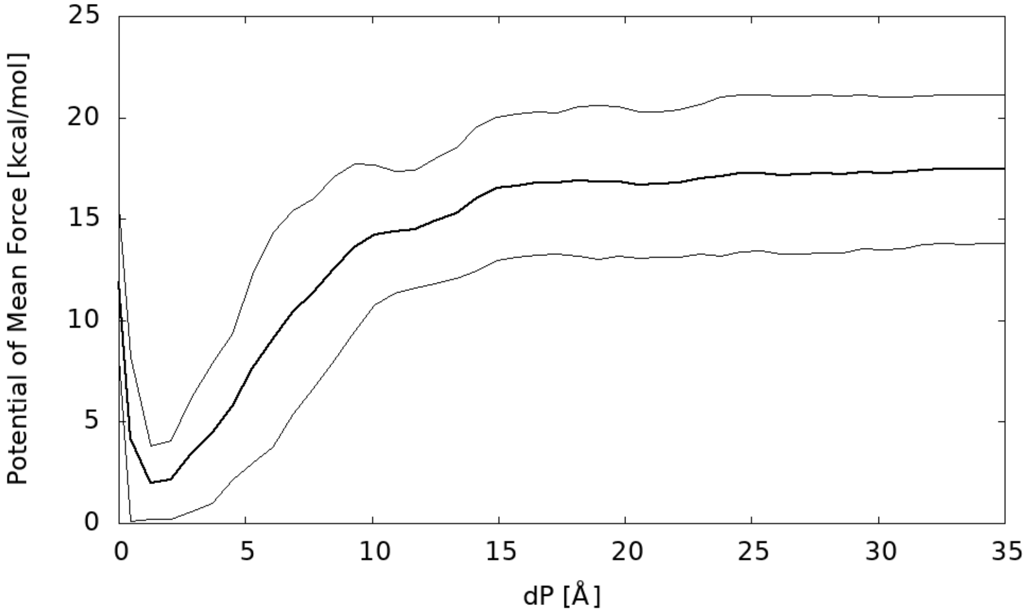


Figure 6.15: Average potential of mean force (central thick line) as a function of the CV d_P for the exploratory metadynamics runs, together with its standard deviation (thin lines).

The shape and depth of these wells resembled those of other unbinding cases described in the literature.[183, 209] The standard deviation of the average PMF highlighted how different trajectories resulted in wells of different depths, comprised between values of around 10 kcal/mol to around 20 kcal/mol.

An estimation of each of the wells can be given by considering at what value of d_P the most important interactions are broken (namely hydrogen bonds with Asp202, Ser155, Trp156, and π interactions with Trp63, Tyr126, Trp156, and Tyr207). Then, the corresponding value of the PMF can be taken as an indicative value: these are reported for each run in table 6.1.

Run label	Free Energy Well Depth [kcal/mol]	Corresponding dP [Å]
A	16.529	13.192
B	17.681	13.333
C	22.776	14.260
D	17.507	15.038
E	16.481	12.898
F	13.753	13.713
G	17.745	13.218
H	16.722	13.124
I	11.351	16.819
J	19.345	7.296
K	19.428	14.565
L	24.145	10.963

Table 6.1: Estimation of the free energy well depth (and corresponding d_P associated to the breaking of the main interactions that stabilise 5-HT in the binding pocket, for each of the 12 exploratory runs. All values were rounded to the third decimal digit for a cleaner representation.

Different free energy well depths imply the presence of different unbinding paths, in terms of timing of rupture of the interactions that help stabilise the ligand in the pocket. The unbinding paths may also be characterised by the presence of useful pinning interactions, as previously observed in past works.[16] The observed interactions during each of the 12 runs can thus be monitored, by making a distinction in terms of paths characterised by shallow free energy wells or deep free energy wells. The observed hydrogen bonds and π interactions, along simulation time, are shown in figure 6.16 and in figure 6.17 respectively.

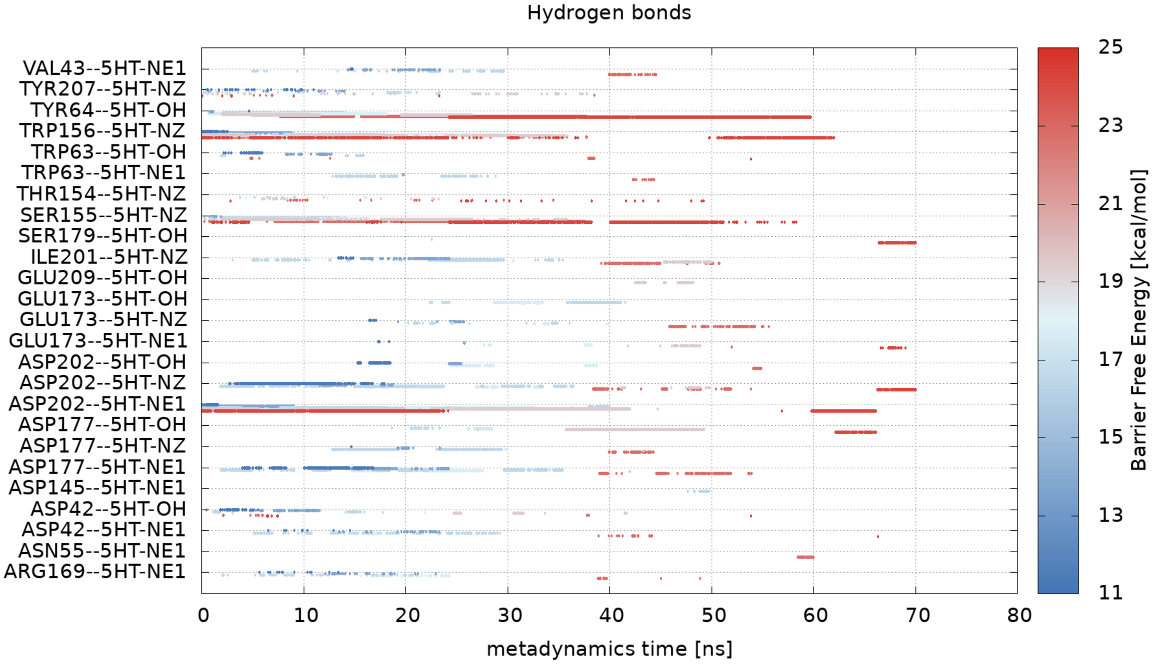


Figure 6.16: Hydrogen bonds over time formed by 5-HT and the residues in the pocket in the different exploratory runs, coloured according to the depth of the free energy well as reported by table 6.1. A coloured dot represents at least one hydrogen bond formed by that residue and 5-HT. Runs of different free energy well depths were partially shifted along the y-axis with respect to one another in order not to overlap with each other.

One feature that could be immediately taken from these results was that several interactions were present for both shallow-well runs (SWRs) and for deep-well runs (DWRs), with the only difference of lasting longer for the latter ones.

Several interactions were present at the beginning of the runs but were broken earlier in SWRs than in DWRs: this was the case of the hydrogen bonds between the hydroxyl group of 5-HT with Tyr64, the 5-HT tail nitrogen with Trp156 and with Ser155, and the 5-HT aromatic nitrogen with Asp202. It was also the case for the cation-π interactions between 5-HT and Tyr207, Trp156, Tyr126, Trp63, Arg65, and for the π-π with Trp63.

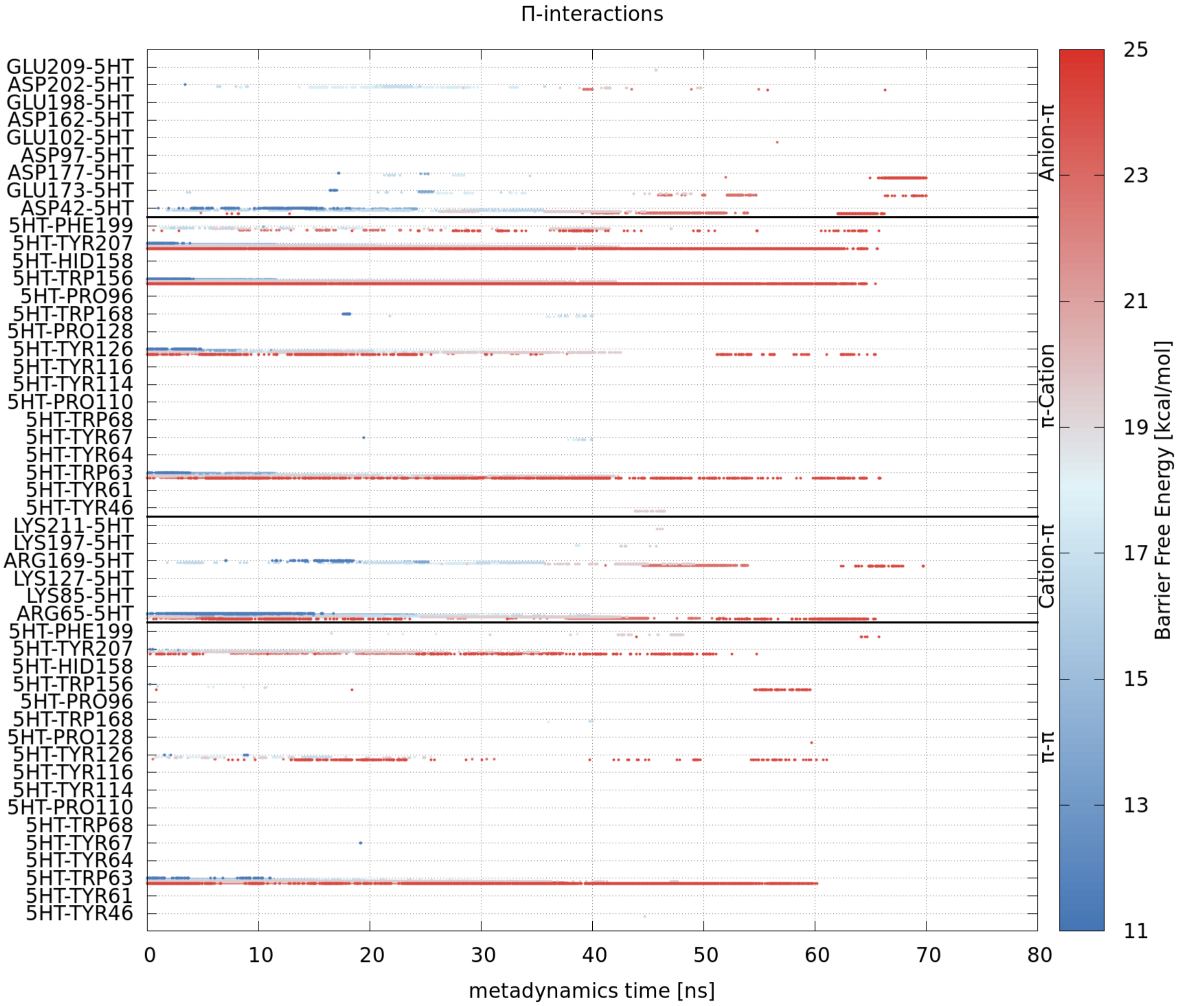


Figure 6.17: π interactions (π-π, cation-π, π-cation, and anion-π) formed by 5-HT and the residues in the pocket in the different exploratory runs, coloured according to the depth of the free energy well as reported by table 6.1. The four types of π interactions are separated by black horizontal lines. Runs of different free energy well depths were partially shifted along the y-axis with respect to one another in order not to overlap with each other.

Other interactions were formed at early stages of SWRs, but at later stages of DWRs: this included the hydrogen bonds between the 5-HT tail nitrogen with Ile201 and Asp202, and Asp177 with any polar moiety of 5-HT. It also regarded anion-π interaction with Asp42 and cation-π interaction with Arg169.

An exemplary SWR trajectory is represented in figure 6.18 by means of four exemplary frames (numbered 1 to 4): initially, the 5-HT tail engages in interactions with Trp156, Tyr126, Tyr207 and Ser155 (frame 1). Then, the flexible tail allows the NH_3 group to be significantly displaced, and to engage with hydrogen bonds with Asp202 (frame 2): this occurs at very early stages in SWRs and could possibly be the triggering event of the whole unbinding cascade of events. Once this pinning has occured, 5-HT rotates around the axis perpendicular to its rings, moving out of the interior of the pocket but still engaging in interactions with multiple loops (frame 3). Asp42, Asp202 and Asp177 may provide useful pinning interactions (e.g. hydrogen bonds or anion-π) that allow for the final unbinding (frame 4).

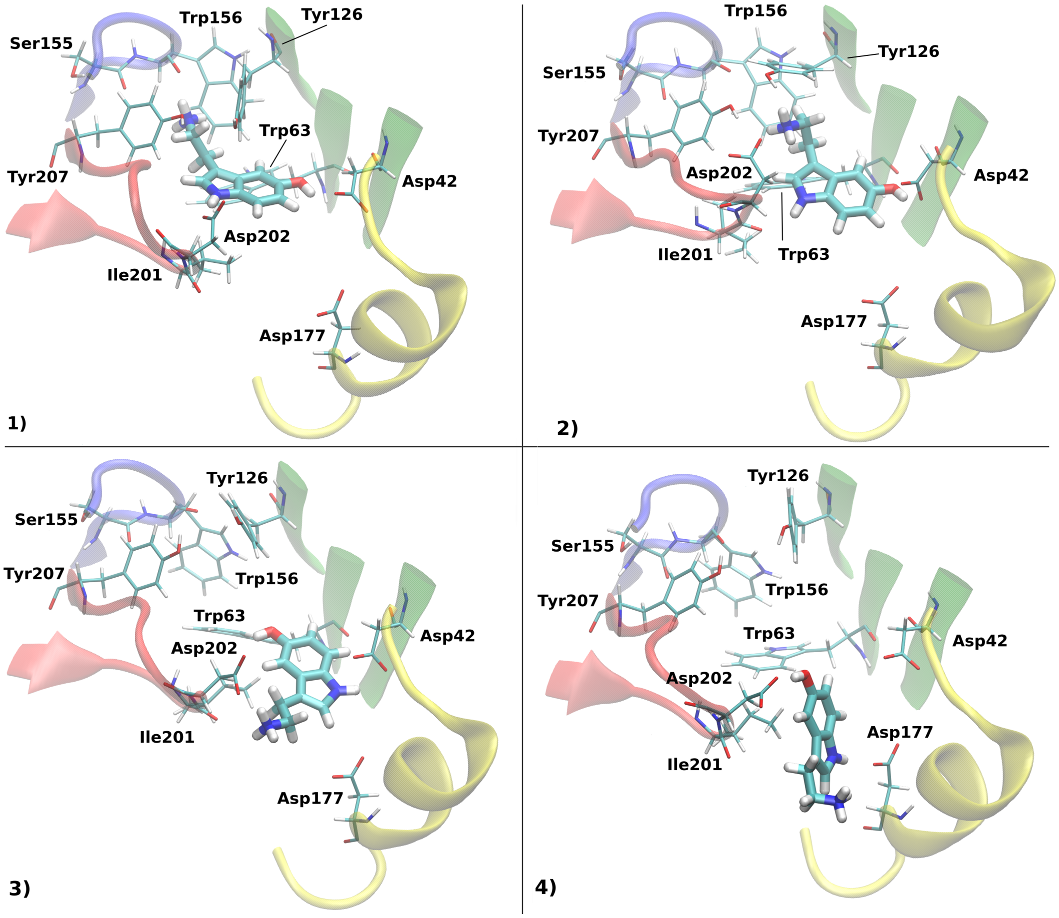


Figure 6.18: Representation of a shallow-well trajectory (replica I), with 4 exemplary frames numbered in progressive order (1 to 4).

6.4 Conclusions and Future Perspectives

In this final chapter, we have concluded our study on the 5-HT$_{3A}$R by focusing on the ligand orthosteric pocket, located in the extracellular domain. We made use of a bound cryo-EM structure to build our models and we performed unbiased as well as enhanced sampling simulations.

We were able to assess the properties of 5-HT in the binding pocket as well as its main stabilising interactions. The identified residues are in agreement with previous findings present in the literature.

Moreover, we set up exploratory well-tempered metadynamics simulations to provide a picture of the paths of interactions that would lead to the unbinding of 5-HT. We revealed the existence of a variable free energy well that separates the bound state from the unbound state, whose depth depends on the particular pattern of interactions formed by 5-HT alongside its path.

We characterised an exemplary low-energy path that sees an initial breaking of the interactions formed by the 5-HT tails with Trp156, Tyr126, Tyr207 and Ser155, a rotation of 5-HT leading to hydrogen bonding with Asp202, and a final breaking of π interactions within the pocket resulting in the expulsion of 5-HT from it. Alongside the path out, other interactions on both sides (namely with residues Asp42, Asp202 and Asp177) may result in useful intermediate pinnings.

The reconstruction of the unbinding time distribution (and thus of the unbinding rate k_{off} from a subsequent reconstrucion of a Markov State Model of the full kinetic process) requires the sampling of the same, low-energy path over several enhanced sampling runs. The presence of different paths in the exploratory runs is mainly due to 5-HT being a relatively large molecule with respect to other pLGICs ligands, such as the γ-aminobutiric acid or the zwitterionic glycine. Thus, the single CV d_P, or any other set of CVs that would identify the position of 5-HT centre of mass within the pocket, may not be enough to identify the precise state of the ligand-pocket complex. Considering its stability in the bound state, and its consistent increase upon unbinding, serotonin solvation allows to distinguish it from the unbound state, but, like d_P, would not be able to convey information regarding the instantaneous interactions between 5-HT and the pocket.

However, our findings allowed for the identification of low-energy paths that can be repeatedly sampled with enhanced sampling techniques such as path metadynamics.[188] An unbinding path may be described in terms of a map of contacts: it may feature a first state where the 5-HT tail interacts with Trp156 in Loop B, a second state where it interacts with Asp202 in Loop C (thus, inducing a rotation of 5-HT around an axis normal to its aromatic rings), a third state where it forms pinning interactions with residues of Loop C and/or F, and a final state where no contacts with the protein are featured. Resulting in a similar pattern of breaking/formation of interactions, the free energy well depth would be similar among the different runs and would thus allow for the reconstruction of the unbinding time distribution.

Chapter 7

Overall Conclusions

In this thesis, we have focused on a pentameric ligand-gated ion channel, the 5-HT$_{3A}$ receptor, by using molecular dynamics simulations as well as enhanced sampling techniques, specifically metadynamics. We investigated different phenomena in different sections of this protein, which may contribute to or modulate the overall working mechanism of the channel.

We have studied the mutual influence of the receptor environment and the *trans*-*cis* isomerisation of prolines within the M2-M3 loop at the interface between TMD and ECD. We demonstrated that, with respect to the simpler case of a proline dipeptide in water, proline *trans*-*cis* isomerisation free energy surface becomes asymmetrical and is characterised by new features that may hint at an enhanced rate of isomerisation. Its effect may propagate to the tip of the M2 helices, where an aspartic acid ring acts as a selectivity filter. Kinetics calculations may continue from our work in order to calculate the rates of isomerisation within the receptor environment.

We have characterised the lipid-protein interface in presence of membranes of different composition, whose viscosity increases with cholesterol content. We proved how this enhances the occurrence of binding events to the surface of the TMD. We also shown how bound phospholipids may reach out for the M2-M3 loop and the Cys-loop, potentially providing a bridge within them. Further calculations may aim at studying the free energy associated with these binding events, or the differences in lipid-protein interactions in different functional states.

We have studied the properties of the outermost M4 helix of the TMD, focusing on the role of a point mutation, Y441A, which has been characterised experimentally. We made two hypothesis: that the effect of this mutation could "vertically" propagate alongside the M4 itself, as to influence its interactions with the Cys-loop, or that a "radial" chain of interactions may reach out for the interior of the TMD, possibly influencing the hydrophobic gate. We found hints that the second hypothesis alone could take place in this particular receptor. Longer simulations, or different techniques (such as enhanced sampling methods for evaluating the free energy of lipid binding to or around residue 441), may be further used to characterise the role of lipids in this phenomenon.

Finally, we have studied the unbinding mechanism of serotonin from the binding pocket in the ECD by means of unbiased as well as biased molecular simulations, by making use of a novel, serotonin-bound structure. The stabilising ligand-pocket interactions that we found were in agreement with past experiments and *in-silico* calculations. We also performed exploratory metadynamics unbinding runs which provided information on a potential path to be sampled with more sophisticated techniques such as path metadynamics.[188] This may be used in the future for the kinetics characterisation of the unbinding process.

Altogether, these projects contributed to the fundamental understanding of the 5-HT$_{3A}$ receptor at the atomistic level, by focusing on separate phenomena that take place in the overall mechanism of this molecular machine. Some of this processes, like proline isomerisation in the M2-M3 loop or the modulation by the M4 residue Y441 (or lack of modulation of A441), occur in narrow regions of space, and were demonstrated as mediated by single residues. However, our work here also highlighted the relevance of phenomena taking place at domain interfaces, where either averaged interactions or bulk phenomena (as in the modulation by the lipid membrane) also become relevant for the protein function.

In particular, the interface between the transmembrane domain and the extracellular domain, essentially made up by the M2-M3 loop and Cys-loop, has been demonstrated to be a crucial area for protein function. Here, proline isomerisation may be responsible for alternatively connect and disconnect the two loops, thus controlling the propagation of mechanical signal, originally triggered by serotonin binding in the extracellular domain. Bound phospholipids in this region can also provide a connection between them. The interactions between the M4 helix and the Cys-loop do not appear to be controlled by the Y441A mutation, the only mutation to alanine that was experimentally found able to inhibit receptor function. Instead, this mutation may modulate the gate opening at the level of Leu260 by a chain of hydrogen bonds with D238 and K255.

Future studies on the 5-HT$_{3A}$ receptor may focus on the mutual influence of these different processes, together with the kinetics associated with each of them. This could be achieved by the usage of longer time scales (e.g. orders of microseconds) and multiple replicas, made accessible thanks to the increasing availability of GPU technology. Alternatively, it could be achieved by the implementation of novel enhanced sampling techniques that would use slowly-varying collective variables determined by means of Markov State Models,[184] time-structure based independent component analysis (tICA),[210] or machine learning.[211] Moreover, the still recent release of new structures of the 5-HT$_{3A}$ receptor also allows for a comparison of these phenomena in different functional states.

Thus, the projects here described contribute to a modular understanding of the 5-HT$_{3A}$ receptor and by extension of pLGICs.

Appendix A - Quantities and Methods for Analysis

Root Mean Square Displacement (RMSD)

For a system of N atoms in different conformations, the root mean square displacement (RMSD) is a quantity that measures how much every atom deviates from a reference position:

$$RMSD = \sqrt{\frac{\sum_i^N (\vec{r_i} - \vec{r_{i0}})^2}{N}} \tag{7.0.0.1}$$

Here, $\vec{r_i}$ is the position of the i-th atom, and $\vec{r_{i0}}$ its reference position. Normally, the different conformations are due to the evolution in time of the system, so that both $\vec{r_i}$ and the RMSD itself are functions of time:

$$RMSD(t) = \sqrt{\frac{\sum_i^N (\vec{r_i}(t) - \vec{r_{i0}})^2}{N}} \tag{7.0.0.2}$$

A common choice for the reference conformation is the one at the first simulation step. Thus, the RMSD is able to provide a measure of the stability of the system, provided that a structural alignment is preventively performed as to remove the effect of translations and rotations.

Root Mean Square Fluctuations (RMSF)

For a system of N atoms in different conformations, the root mean square fluctuation (RMSF) is a quantity that measures how much a given selection of $M \leq N$ atoms deviates from a certain reference conformation, typically the average one that these atoms assume over the simulation time.

In this case:

$$RMSF(i) = \sqrt{\frac{\sum_{t=1}^{N_T} (\vec{r_i}(t) - \overline{\vec{r_i}})^2}{N_T}} \tag{7.0.0.3}$$

where the sum runs over all the time steps t, N_T is the total number of time steps, and $\overline{\vec{r_i}}$ is the time average of the position $\vec{r_i}$ of the i-th atom. The RMSF is typically calculated for selections of residues. This quantity provides a measure of the mobility of the selected residues.

Hydrogen Bonds

Hydrogen bonds (represented in figure 7.1) are formed by an acceptor heavy atom (A), a donor hydrogen (H), and a donor heavy atom (D), with the hydrogen forming a covalent bond with the donor atom (although the opposite definition may appear in other works). A hydrogen bond is considered formed if the A to D distance is less than a given distance cut-off and the A-H-D angle is greater than a certain angle cut-off. Standard choices for these values are 120° and 3.5 Å. These values have been used in a variety of previous works in pLGICs[7, 10, 207, 83, 16] and are the defaults of analysis software such as MDAnalysis,[146, 147] but they can be tuned in order to suit the specific physical system of interest. For example, if the hydrogen bonds are searched between atoms of same residue (or molecule), hydrogen bonds can be searched with less restrictive values, i.e. 3.5 Å for the distance and 100° for the angle.

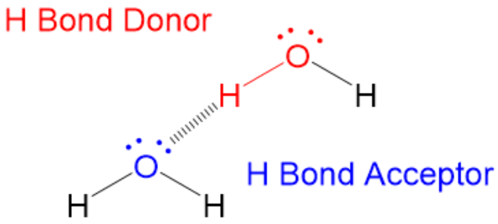


Figure 7.1: Representation of hydrogen bonding.

π Interactions

It is not infrequent to deal with organic compounds that feature aromatic rings. Among these, we find the aromatic amino-acids tryptophan, tyrosine, phenylalanine, histidine and proline. These rings are able to form a variety of interactions known as π *interactions*, which are due to the electron clouds locate above and below said rings. Serotonin, the agonist of the 5-HT$_{3A}$ receptor, is also capable of making π interactions.

Among the most important interactions we find π-π, *cation*-π, and *anion*-π. These often appear in biological systems, e.g. in the orthosteric binding pocket of receptors, where they contribute to stabilise the ligand. In the following, we describe these interactions one by one, and we outline their calculation methods applied in the project described in this thesis.

π-π Interaction

π-π interactions occur between aromatic residues of different molecules or residues, as depicted in figure 7.2. Among these, the ligand serotonin, and the amino-acids histidine, proline, phenylalanine, tryptophan, and tyrosine.

This interaction occurs because of the electron-rich sp2 orbitals of such rings, resulting in a mutual attraction. One likely conformation adopted by the rings by means of this interaction is the one with the two rings in parallel, although other conformations are possible.

Ab-initio calculations on the benzene dimer provided that the ring-to-ring distance is 4.96 Å.[212] Thus, in the calculations performed for the projects described in this thesis, the distance between the centres of mass (COMs) of the two rings was considered as lower than 6.0 Å, in order to account for small variations from the expected about 5.0 Å.

As for the orientation of the rings with respect to each other, here we considered the angle between normals angle as smaller than 45° or greater than 135°. The normal to an aromatic ring was defined as the average vector product among every pair of bond within the aromatic rings. For those residues including more than one ring (such as tryptophan, or serotonin), each of the two rings were considered separately, and their contributions to the π interaction was afterwards summed.

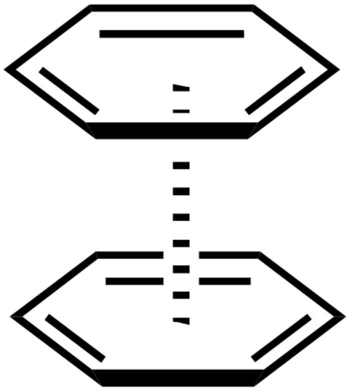


Figure 7.2: Representation of the π-π interaction.

Cation-π Interaction

Cation-π interactions (represented in figure 7.3) occur between an aromatic residue and the positively-charged moiety of another. Serotonin in its protonated form carries a positively-charged NH_3 group, which can engage in cation-π interactions with a cage of aromatic residues. Arginine and lysine, when standard protonation is applied, are the amino-acids with positively-charged side chains, so they can form these interactions too. Here, we considered cation-π interactions on the basis of the distance between the aromatic ring and the charged atom, and the angle between this distance and the ring average normal. We considered the distance as smaller than 6.0 Å, and the angle smaller than 45° or greater than 135°, on the basis of previous studies applying these same cut-offs.[213, 214, 215]

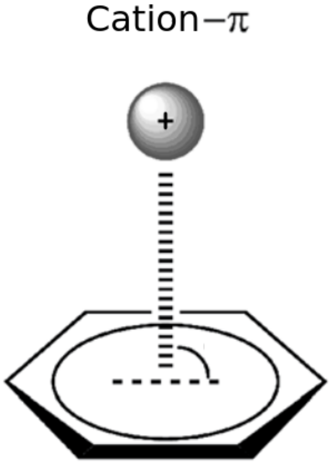


Figure 7.3: Representation of the cation-π interaction.

Anion-π Interaction

Similarly to cation-π interactions, anion-π interactions (represented in figure 7.4) take place between an aromatic ring and a negatively-charged moiety, although their relevance in bio-systems has been established more recently.[216] The amino-acids aspartic acid and glutamic acid can give rise to anion-π interactions thanks to their negatively-charged side chains. Serotonin can also form such interactions by means of its aromatic rings. We considered anion-π interactions on the basis of the distance between the aromatic ring and the charged atom, together with the angle between this distance and the ring average normal. The former would be smaller than 5.0 Å, while the latter would be smaller than 40° or greater than 140°, according to past studies using similar cut-offs.[216]

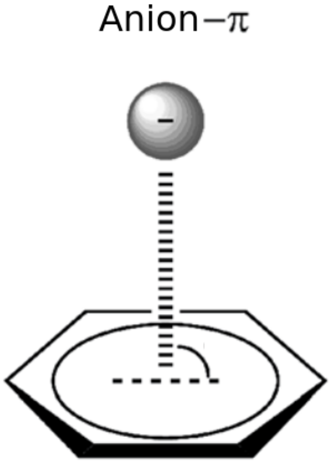


Figure 7.4: Representation of the anion-π interaction.

Atomic Contacts

In biological systems, different interactions can bring different molecules or different structures closer to each other. Many of these interactions, such as hydrogen bonds and π interactions, are directional, and as such can be studied with the application of distance cut-offs (and angle cut-offs, when appropriate). Others, like the hydrophobic effect between non-polar moieties, are instead non directional. In order to account for the presence of any type of interaction, without making and explicit discrimination on its exact nature, it is useful to calculate the number of contacts between two selections of atoms. A "contact", in this context, is defined as any pair of atom (one belonging to one selection, the other belonging to the other) whose distance is below a certain value. Previous studies on lipid-protein complexes made use of 3.5 Å as a cut-off,[153] and it was thus applied in all calculations in these projects as well.

Channel Width with HOLE2

It is often desirable to investigate the shape and time evolution of long cavities that may be featured within certain proteins. This is the case for the ion channel and ECD vestibule of pLGICs, whose conformation is intimately related to the protein function. The channel width as a function of the position along the channel axis is thus worth investigating. A dedicated package, HOLE2,[4] is available for this purpose. The working principle of this software is described in the following: given an initial point inside the cavity (the channel, in the case of pLGICs) and a vector that is approximately its direction, the software generates, for a series of points along the channel direction, the radius of the largest sphere that does not overlap with any atom. The procedure can be repeated for every time step of the simulation, which can be used to calculate the average channel profile and the associated error.

Electric Potential with PMEPOT

The `pmepot` package of VMD,[103] calculates the value of the electrical potential over every point on a grid defined in the simulation cell, thanks to an algorithm that considers the partial charge q_i of the selected atoms as a Gaussian-distributed charge:

$$\rho_i(\vec{r}) = q_i \left(\frac{\beta}{\sqrt{\pi}} \right)^3 \exp(-\beta^2 |\vec{r} - \vec{r_i}|) \tag{7.0.0.4}$$

where β, the Ewald factor, is a positive parameter that determines the width of the Gaussians and $\vec{r_i}$ is the position of the i-th atom. The potential as a function of the position $\vec{r}$ is calculated by solving Poisson's equation:

$$\nabla^2 V(\vec{r}) = 4\pi \sum_i \rho_i(\vec{r}) \tag{7.0.0.5}$$

where the sum runs over all selected atoms. This equation is solved numerically on the grid using Fast Fourier transforms: the grid density determines the accuracy of the resulting potential.

This software has been used for the results described in Chapter 3, for which the Ewald factor was set to 0.25 Å^{-1}.

Dynamical Correlation

The dynamical correlation is a quantity calculated between two selections of atoms, and can take values between -1 and 1: values close to 1 mean that the two atoms selections move consistently in the same direction over time, acting as a rigid body, values close to 0 indicate that the dynamics of the two atom selections never feature any correlation, and values close to -1 indicate that the two selections consistently move in opposite directions.

The dynamical corrrelation between two atoms i and j is defined as:[177]

$$C_{ij} = \frac{\langle \vec{r_i} \vec{r_j} \rangle - \langle \vec{r_i} \rangle \langle \vec{r_j} \rangle}{\sqrt{(\langle \vec{r_i}^2 \rangle - \langle \vec{r_i} \rangle^2)(\langle \vec{r_j}^2 \rangle - \langle \vec{r_j} \rangle^2)}} \tag{7.0.0.6}$$

Continuous Coordination Number

A way to define a continuous solvation (e.g. of a ligand and its surrounding water molecules), previously used as CV for studying unbinding problems,[183, 187] is to consider a *switching function* $f(r_{ij})$ of the distance r_{ij} between the ligand i-th atom and the oxygen of the j-th water molecule instead of a sharp cutoff:

$$f(r_{ij}) = \frac{1 - (r_{ij}/R_0)^n}{1 - (r_{ij}/R_0)^m} \tag{7.0.0.7}$$

Typical choices for the parameters of the switching function, used for our works here and suggested as defaults of the package Plumed,[101] are n=6, m=12, R_0=3.5 Å (the corresponding switching function being represented in figure 7.5).

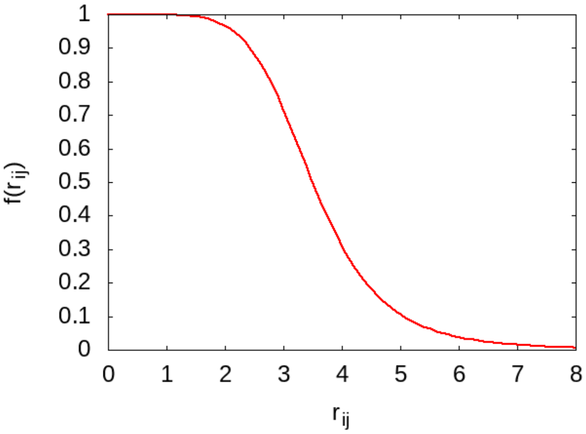


Figure 7.5: Switching function with n=6, m=12, R_0=3.5 Å.

In this way, one can consider a switching function between each atom pair of two given atom selections, and then consider functions of their distribution, such as the mean coordination number.

Appendix B - 5-HT Atom Types and Charges

Atom name	Atom type	Partial charge [electron units]
CZ2	ca	-0.065657
H16	ha	0.178673
CH2	ca	-0.467747
H15	ha	0.216645
CZ3	ca	0.561771
OH	oh	-0.655987
H14	ho	0.468385
CE3	ca	-0.623363
H17	ha	0.275391
CD2	ca	0.192953
CE2	ca	0.055571
NE1	na	-0.468545
H22	hn	0.417089
CD1	cc	-0.105692
H23	h4	0.215775
CG	cd	-0.000861
CB	c3	-0.439953
H18	hc	0.180300
H19	hc	0.180300
CA	c3	0.141805
H24	hx	0.094446
H25	hx	0.094446
NZ	n4	-0.451878
H20	hn	0.335378
H21	hn	0.335378
H26	hn	0.335378

Table 7.1: General Amber force-field atom types and charges of the serotonin molecule.

Acknowledgements

For the work presented in this thesis, I would like to thank my supervisor Carla Molteni, for the excellent mentoring provided during the four years of Ph.D.; Luciano Colombo, for being my first mentor in the academic world and for letting me know about this Ph.D. opportunity; my second supervisor Chris Lorenz, for useful discussions and support; Sarah C. R. Lummis, Susanne M. Mesoy, Klaus Suhling, Yurema Teijeiro González, Reiner Schulz, and Federico Rossi, for the valuable collaborations developed in the past years; the current and past members of the group, for the great work done together, and all the people of the Physics and Chemistry Departments of King's College London, as well as my former colleagues, for the brilliant brainstorming sessions.

I want to acknowledge how a great part of what I have become is the result of the affectionate, and enlightening, upbringing that my family gave me, and of their endless support through past and present struggles. A special thanks is also for all the people that supported me through the difficulties of the last year of Ph.D.

I would like to express how lucky I feel for having been part of the thrilling, diverse and exciting social and professional environment of the city of London, and I would like to thank all the friends and acquaintances that contributed to a priceless and unmissable personal growth as well as great fun.

Lastly, I would like to express my happiness and pride for having found a great sense of meaning thanks to the work I have carried out in the context of academia in the past four years, with the hope of being able to bring it with me in the days ahead and to communicate its value to others.

Bibliography

[1] Brain facts. *Society for Neuroscience*, 2008.

[2] Ghérici Hassaine, Cédric Deluz, Luigino Grasso, Romain Wyss, Menno B. Tol, Ruud Hovius, Alexandra Graff, Henning Stahlberg, Takashi Tomizaki, Aline Desmyter, Christophe Moreau, Xiao-Dan Li, Frédéric Poitevin, Horst Vogel, and Hugues Nury. X-ray structure of the mouse serotonin 5-HT$_3$ receptor. *Nature*, 512(7514):276–81, 2014.

[3] Alessandro Crnjar and Carla Molteni. Cholesterol content in the membrane promotes key lipid-protein interactions in a pentameric serotonin-gated ion channel. *Biointerphases*, 15(6):061018, 2020.

[4] O.S. Smart, J.M. Goodfellow, and B.A. Wallace. The pore dimensions of gramicidin A. *Biophysical Journal*, 65(6):2455 – 2460, 1993.

[5] Sandip Basak, Yvonne Gicheru, Shanlin Rao, Mark S. P. Sansom, and Sudha Chakrapani. Cryo-EM reveals two distinct serotonin-bound conformations of full-length 5-HT3A receptor. *Nature*, 563(7730):270–274, 2018.

[6] Daan Frenkel and Berend Smit. *Understanding Molecular Simulations - From Algorithms to Applications.* Academic Press, 2002.

[7] Alessandro Crnjar, Federico Comitani, William Hester, and Carla Molteni. Trans–cis proline switches in a pentameric ligand-gated ion channel: How they are affected by and how they affect the biomolecular environment. *The Journal of Physical Chemistry Letters*, 10(3):694–700, 2019.

[8] Alessandro Crnjar, Susanne M. Mesoy, Sarah C. R. Lummis, and Carla Molteni. A single mutation in the outer lipid-facing helix of a pentameric ligand-gated ion channel affects channel function through a radially-propagating mechanism. *Frontiers in Molecular Biosciences (accepted for publication)*, 2021.

[9] Federico Comitani. *Computational Modelling of Activation Mechanisms in Ligand Gated Ion Channels*. PhD Thesis, King's College London, 2016.

[10] Alessandro Crnjar, Federico Comitani, Claudio Melis, and Carla Molteni. Mutagenesis computer experiments in pentameric ligand-gated ion channels: the role of simulation tools with different resolution. *Interface Focus*, 9(3):20180067, 2019.

[11] F. Rossi, A. Crnjar, F. Comitani, R. Feliciano, L. Jahn, G. Malim, L. Southgate, E. Kay, R. Oakey, R. Buggs, A. Moir, L. Kistler, A. Rodriguez Mateos, C. Molteni, and R. Schulz. Extraction and high-throughput sequencing of oak heartwood DNA: assessing the feasibility of genome-wide DNA methylation profiling. *PLOS ONE (manuscript)*, 2021.

[12] Yurema Teijeiro-Gonzalez, Alessandro Crnjar, Andrew J. Beavil, Rebecca L. Beavil, Jakub Nedbal, Alix Le Marois, Carla Molteni, and Klaus Suhling. Time-Resolved Fluorescence Anisotropy and Molecular Dynamics Analysis of a Novel GFP Homo-FRET Dimer. *Biophysical Journal*, 2020.

[13] Ákos Nemecz, Marie S Prevost, Anaïs Menny, and Pierre-Jean Corringer. Emerging Molecular Mechanisms of Signal Transduction in Pentameric Ligand-Gated Ion Channels. *Neuron*, 90(3):452–70, 2016.

[14] Prafulla Aryal, Mark S.P. Sansom, and Stephen J. Tucker. Hydrophobic gating in ion channels. *Journal of Molecular Biology*, 427(1):121 – 130, 2015.

[15] Jemma L. Trick, Prafulla Aryal, Stephen J. Tucker, and Mark S. P. Sansom. Molecular simulation studies of hydrophobic gating in nanopores and ion channels. *Biochemical Society Transactions*, 43(2):146–150, 2015.

[16] Federico Comitani, Vittorio Limongelli, and Carla Molteni. The free energy landscape of gaba binding to a pentameric ligand-gated ion channel and its disruption by mutations. *Journal of Chemical Theory and Computation*, 12(7):3398–3406, 2016.

[17] Diego Rayes, María José De Rosa, Steven M. Sine, and Cecilia Bouzat. Number and Locations of Agonist Binding Sites Required to Activate Homomeric Cys-Loop Receptors. *Journal of Neuroscience*, 29(18):6022–6032, 2009.

[18] Jacque Monod, Jeffries Wyman, and Jean-Pierre Changeux. On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*, 12(1):88 – 118, 1965.

[19] Hong Wei Wang and Jia Wei Wang. How cryo-electron microscopy and X-ray crystallography complement each other. *Protein Science*, 26(1):32–39, 2017.

[20] Nigel Unwin. Refined Structure of the Nicotinic Acetylcholine Receptor at 4 Å Resolution. *Journal of Molecular Biology*, 346(4):967 – 989, 2005.

[21] Shaotong Zhu, Colleen M. Noviello, Jinfeng Teng, Richard M. Walsh, Jeong Joo Kim, and Ryan E. Hibbs. Structure of a human synaptic GABA A receptor. *Nature*, 559(7712):67–72, 2018.

[22] Duncan Laverty, Rooma Desai, Tomasz Uchański, Simonas Masiulis, Wojciech J. Stec, Tomas Malinauskas, Jasenko Zivanov, Els Pardon, Jan Steyaert, Keith W. Miller, and A. Radu Aricescu. Cryo-em structure of the human $\alpha1\beta3\gamma2$ GABA A receptor in a lipid bilayer. *Nature*, 565(7740):516–520, 2019.

[23] Simonas Masiulis, Rooma Desai, Tomasz Uchański, Itziar Serna Martin, Duncan Laverty, Dimple Karia, Tomas Malinauskas, Jasenko Zivanov, Els Pardon, Abhay Kotecha, Jan Steyaert, Keith W. Miller, and A. Radu Aricescu. GABA A receptor signalling mechanisms revealed by structural pharmacology. *Nature*, 565(7740):454–459, 2019.

[24] Juan Du, Wei Lü, Shenping Wu, Yifan Cheng, and Eric Gouaux. Glycine receptor mechanism elucidated by electron cryo-microscopy. *Nature*, 526(7572):224–229, 2015.

[25] Xin Huang, Hao Chen, Klaus Michelsen, Stephen Schneider, and Paul L. Shaffer. Crystal structure of human glycine receptor-α3 bound to antagonist strychnine. *Nature*, 526(7572):277–280, 2015.

[26] Xin Huang, Paul L. Shaffer, Shawn Ayube, Howard Bregman, Hao Chen, Sonya G. Lehto, Jason A. Luther, David J. Matson, Stefan I. McDonough, Klaus Michelsen, Matthew H. Plant, Stephen Schneider, Jeffrey R. Simard, Yohannes Teffera, Shuyan Yi, Maosheng Zhang, Erin F. DiMauro, and Jacinthe Gingras. Crystal structures of human glycine receptor α3 bound to a novel class of analgesic potentiators. *Nature Structural & Molecular Biology*, 24(2):108–113, 2017.

[27] Xin Huang, Hao Chen, and Paul L. Shaffer. Crystal structures of human glyrα3 bound to ivermectin. *Structure*, 25(6):945–950.e2, 2017.

[28] Ryan E. Hibbs and Eric Gouaux. Principles of activation and permeation in an anion-selective Cys-loop receptor. *Nature*, 474(7349):54–60, 2011.

[29] Thorsten Althoff, Ryan E. Hibbs, Surajit Banerjee, and Eric Gouaux. X-ray structures of GluCl in apo states reveal a gating mechanism of Cys-loop receptors. *Nature*, 512(7514):333–337, 2014.

[30] Ricarda J. C. Hilf and Raimund Dutzler. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature*, 452(7185):375–9, 2008.

[31] Haidai Hu, Ákos Nemecz, Catherine Van Renterghem, Zaineb Fourati, Ludovic Sauguet, Pierre-Jean Corringer, and Marc Delarue. Crystal structures of a pentameric ion channel gated by alkaline pH show a widely open pore and identify a cavity for modulation. *Proceedings of the National Academy of Sciences*, 115(17):E3959–E3968, 2018.

[32] Mieke Nys, Eveline Wijckmans, Ana Farinha, Özge Yoluk, Magnus Andersson, Marijke Brams, Radovan Spurny, Steve Peigneur, Jan Tytgat, Erik Lindahl, and Chris Ulens. Allosteric binding site in a Cys-loop receptor ligand-binding domain unveiled in the crystal structure of ELIC in complex with chlorpromazine. *Proceedings of the National Academy of Sciences of the United States of America*, 113(43):E6696–E6703, 2016.

[33] Nicolas Bocquet, Hugues Nury, Marc Baaden, Chantal Le Poupon, Jean-Pierre Changeux, Marc Delarue, and Pierre-Jean Corringer. X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation. *Nature*, 457(7225):111–114, 2009.

[34] Jianjun Pan, Qiang Chen, Dan Willenbring, David Mowrey, Xiang-Peng Kong, Aina Cohen, Christopher B. Divito, Yan Xu, and Pei Tang. Structure of the pentameric ligand-gated ion channel GLIC bound with anesthetic ketamine. *Structure*, 20(9):1463–1469, 2012.

[35] Marie S. Prevost, Ludovic Sauguet, Hugues Nury, Catherine Van Renterghem, Christèle Huon, Frederic Poitevin, Marc Baaden, Marc Delarue, and Pierre-Jean Corringer. A locally closed conformation of a bacterial pentameric proton-gated ion channel. *Nature Structural & Molecular Biology*, 19(6):642–649, 2012.

[36] Ludovic Sauguet, Azadeh Shahsavar, Frédéric Poitevin, Christèle Huon, Anaïs Menny, Àkos Nemecz, Ahmed Haouz, Jean-Pierre Changeux, Pierre-Jean Corringer, and Marc Delarue. Crystal structures of a pentameric ligand-gated ion channel provide a mechanism for activation. *Proceedings of the National Academy of Sciences of the United States of America*, 111(3):966–971, 2014.

[37] Ludovic Sauguet, Zeineb Fourati, Thierry Prang, Marc Delarue, and Nathalie Colloc'h. Structural Basis for Xenon Inhibition in a Cationic Pentameric Ligand-Gated Ion Channel. *PLOS ONE*, 11(2):1–17, 2016.

[38] Beate Niesler, Jutta Walstab, Sandra Combrink, Dorothee Möller, Johannes Kapeller, Jens Rietdorf, Heinz Bönisch, Manfred Göthert, Gudrun Rappold, and Michael Brüss. Characterization of the Novel Human Serotonin Receptor Subunits 5-HT3C, 5-HT3D, and 5-HT3E. *Molecular Pharmacology*, 72(1):8–17, 2007.

[39] Ken Solt, Dirk Ruesch, Stuart A. Forman, Paul A. Davies, and Douglas E. Raines. Differential Effects of Serotonin and Dopamine on Human 5-HT3A Receptor Kinetics: Interpretation within an Allosteric Kinetic Model. *Journal of Neuroscience*, 27(48):13151–13160, 2007.

[40] Danilo Di Maio, Balasubramanian Chandramouli, and Giuseppe Brancato. Pathways and Barriers for Ion Translocation through the 5-HT3A Receptor Channel. *PLOS ONE*, 10(10):1–23, 2015.

[41] W J Wedemeyer, E Welker, and H A Scheraga. Proline cis-trans isomerization and protein folding. *Biochemistry*, 41(50):14637–44, 2002.

[42] Nicholas B. Guros, Arvind Balijepalli, and Jeffery B. Klauda. Microsecond-timescale simulations suggest 5-HT-mediated preactivation of the 5-HT3A serotonin receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 117(1):405–414, 2020.

[43] M. Yu. Antonov, A. V. Popinako, and G. A. Prokopiev. Molecular dynamics simulation of the structure and dynamics of 5-HT3 serotonin receptor. *AIP Conference Proceedings*, 1773(1):060001, 2016.

[44] Shuguang Yuan, Slawomir Filipek, and Horst Vogel. A Gating Mechanism of the Serotonin 5-HT3 Receptor. *Structure*, 24(5):816–825, 2016.

[45] Mikhail Kudryashev, Daniel Castaño-Díez, Cédric Deluz, Gherici Hassaine, Luigino Grasso, Alexandra Graf-Meyer, Horst Vogel, and Henning Stahlberg. The Structure of the Mouse Serotonin 5-HT3 Receptor in Lipid Vesicles. *Structure*, 24(1):165–170, 2016.

[46] Lucie Polovinkin, Ghérici Hassaine, Jonathan Perot, Emmanuelle Neumann, Anders A. Jensen, Solène N. Lefebvre, Pierre-Jean Corringer, Jacques Neyton, Christophe Chipot, Francois Dehez, Guy Schoehn, and Hugues Nury. Conformational transitions of the serotonin 5-HT3 receptor. *Nature*, 563(7730):275–279, 2018.

[47] Sandip Basak, Yvonne Gicheru, Amrita Samanta, Sudheer Kumar Molugu, Wei Huang, Maria la de Fuente, Taylor Hughes, Derek J. Taylor, Marvin T. Nieman, Vera Moiseenkova-Bell, and Sudha Chakrapani. Cryo-EM structure of 5-HT3A receptor in its resting conformation. *Nature Communications*, 9(1):514, 2018.

[48] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23:327–341, 1977.

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

[50] Berendsen H.J.C. *Transport Properties Computed by Linear Response through Weak Coupling to a Bath. In: Meyer M., Pontikis V. (eds), Computer Simulation in Materials Science. NATO ASI Series (Series E: Applied Sciences), vol 205.* Springer, Dordrecht, 1991.

[51] Scott E. Feller, Yuhong Zhang, Richard W. Pastor, and Bernard R. Brooks. Constant pressure molecular dynamics simulation: The langevin piston method. *The Journal of Chemical Physics*, 103(11):4613–4621, 1995.

[52] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

[53] Siewert J. Marrink and D. Peter Tieleman. Perspective on the martini model. *Chem. Soc. Rev.*, 42:6801–6822, 2013.

[54] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.

[55] James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015.

[56] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3):712–725, 2006.

[57] Chuan Tian, Koushik Kasavajhala, Kellon A. A. Belfon, Lauren Raguette, He Huang, Angela N. Migues, John Bickel, Yuzhang Wang, Jorge Pincay, Qin Wu, and Carlos Simmerling. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum

Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation*, 16(1):528–552, 2020.

[58] Michael F. Crowley, Mark J. Williamson, and Ross C. Walker. CHAMBER: Comprehensive support for CHARMM force fields within the AMBER software. *International Journal of Quantum Chemistry*, 109(15):3767–3772, 2009.

[59] Callum J. Dickson, Benjamin D. Madej, Åge A. Skjevik, Robin M. Betz, Knut Teigen, Ian R. Gould, and Ross C. Walker. Lipid14: The Amber Lipid Force Field. *Journal of Chemical Theory and Computation*, 10(2):865–879, 2014.

[60] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[61] Miguel Angel Gonzlez and Jos L. F. Abascal. The shear viscosity of rigid water models. *The Journal of Chemical Physics*, 132(9):096101, 2010.

[62] Yanmei Song and Lenore L. Dai. The shear viscosities of common water models by non-equilibrium molecular dynamics simulations. *Molecular Simulation*, 36(7-8):560–567, 2010.

[63] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Jr. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09, revision e.01, 2013.

[64] Christopher I. Bayly, Piotr Cieplak, Wendy Cornell, and Peter A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *The Journal of Physical Chemistry*, 97(40):10269–10280, 1993.

[65] Christopher A. Reynolds, Jonathan W. Essex, and W. Graham Richards. Atomic charges for variable molecular conformations. *Journal of the American Chemical Society*, 114(23):9075–9079, 1992.

[66] Marc Souaille and Benot Roux. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer Physics Communications*, 135(1):40 – 57, 2001.

[67] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12562–6, 2002.

[68] Alessandro Laio and Francesco Luigi Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics*, 71(12):126601, 2008.

[69] Sanghyun Park, Fatemeh Khalili-Araghi, Emad Tajkhorshid, and Klaus Schulten. Free energy calculation from steered molecular dynamics simulations using jarzynskis equality. *The Journal of Chemical Physics*, 119(6):3559–3566, 2003.

[70] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Physical Review B*, 66:1–4, 2002.

[71] Albert C. Pan, Deniz Sezer, and Benoit Roux. Finding transition pathways using the string method with swarms of trajectories. *The Journal of Physical Chemistry B*, 112(11):3432–3440, 2008.

[72] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical Review Letters*, 100(2):020603, 2008.

[73] M. Bonomi, A. Barducci, and M. Parrinello. Reconstructing the equilibrium boltzmann distribution from well-tempered metadynamics. *Journal of Computational Chemistry*, 30(11):1615–1621, 2009.

[74] Debabrata Pramanik, Zachary Smith, Adam Kells, and Pratyush Tiwary. Can one trust kinetic and thermodynamic observables from biased metadynamics simulations: detailed quantitative benchmarks on millimolar drug fragment dissociation. *bioRxiv*, 2019.

[75] Frank J. Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[76] L.H. Miller. Table of percentage points of kolmogorov statistics. *Journal of the American Statistical Association*, 51:111–121, 1956.

[77] Matteo Salvalaglio, Pratyush Tiwary, and Michele Parrinello. Assessing the reliability of the dynamics reconstructed from metadynamics. *Journal of Chemical Theory and Computation*, 10(4):1420–1425, 2014.

[78] Pratyush Tiwary and Michele Parrinello. From metadynamics to dynamics. *Physical Review Letters*, 111:230602, 2013.

[79] Yong Wang, Omar Valsson, Pratyush Tiwary, Michele Parrinello, and Kresten Lindorff-Larsen. Frequency adaptive metadynamics for the calculation of rare-event kinetics. *The Journal of Chemical Physics*, 149(7):072309, 2018.

[80] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–802, 2005.

[81] Claudio Melis, Giovanni Bussi, Sarah C R Lummis, and Carla Molteni. Trans-cis switching mechanisms in proline analogues and their relevance for the gating of the 5-HT3 receptor. *The Journal of Physical Chemistry B*, 113(35):12148–53, 2009.

[82] Vanessa Leone, Gianluca Lattanzi, Carla Molteni, and Paolo Carloni. Mechanism of action of cyclophilin a explored by metadynamics simulations. *PLoS Computational Biology*, 5(3):e1000309, 2009.

[83] Federico Comitani, Claudio Melis, and Carla Molteni. Elucidating ligand binding and channel gating mechanisms in pentameric ligand-gated ion channels by atomistic simulations. *Biochemical Society Transactions*, 43(2):151–6, 2015.

[84] Carlo Camilloni, Aleksandr B Sahakyan, Michael J Holliday, Nancy G Isern, Fengli Zhang, Elan Z Eisenmesser, and Michele Vendruscolo. Cyclophilin A catalyzes proline isomerization by an electrostatic handle mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 111(28):10203–10208, 2014.

[85] Junchao Xia and Ronald M. Levy. Molecular dynamics of the proline switch and its role in crk signaling. *The Journal of Physical Chemistry B*, 118(17):4535–4545, 2014.

[86] Keiko Shinoda and Hideaki Fujitani. Initiation of prolyl cis-trans isomerisation in the cdr-h3 loop of an antibody in response to antigen binding. *Scientific Reports*, 7(16964):1–12, 2017.

[87] Maria Celeste Maschio, Jacopo Fregoni, Carla Molteni, and Stefano Corni. Proline isomerization effects in the amyloidogenic protein β2-microglobulin. *Physical Chemistry Chemical Physics*, 23:356–367, 2021.

[88] Franziska Zosel, Davide Mercadante, Daniel Nettels, and Benjamin Schuler. A proline switch explains kinetic heterogeneity in a coupled folding and binding reaction. *Nature Communications*, 9(1):3332, 2018.

[89] Sarah C. R. Lummis, Darren L. Beene, Lori W. Lee, Henry A. Lester, R. William Broadhurst, and Dennis A. Dougherty. Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature*, 438(7065):248–252, 2005.

[90] Isabelle M. Paulsen, Ian L. Martin, and Susan M. J. Dunn. Isomerization of the proline in the M2-M3 linker is not required for activation of the human 5-HT3A receptor. *Journal of Neurochemistry*, 110(3):870–878, 2009.

[91] Richard Mosesso, Dennis A. Dougherty, and Sarah C. R. Lummis. Proline Residues in the Transmembrane/Extracellular Domain Interface Loops Have Different Behaviors in 5-HT3 and nACh Receptors. *ACS Chemical Neuroscience*, 10(7):3327–3333, 2019.

[92] Won Yong Lee, Chris R. Free, and Steven M. Sine. Nicotinic receptor interloop proline anchors β1-β2 and Cys-loops in coupling agonist binding to channel gating. *The Journal of General Physiology*, 132(2):265–278, 2008.

[93] Richard Mosesso, Dennis A. Dougherty, and Sarah C. R. Lummis. Probing proline residues in the prokaryotic ligand-gated ion channel, elic. *Biochemistry*, 57(27):4036–4043, 2018.

[94] Gustav Akk and Anthony Auerbach. Activation of muscle nicotinic acetylcholine receptor channels by nicotinic and muscarinic agonists. *British Journal of Pharmacology*, 128:1467–1476, 1999.

[95] Lie Min, Bruce D. Fulton, and Amy H. Andreotti. A case study of proline isomerization in cell signaling. *British Journal of Pharmacology*, 10:385–397, 2005.

[96] Christian X. Weichenberger and Manfred J. Sippl. NQ-Flipper: recognition and correction of erroneous asparagine and glutamine side-chain rotamers in protein structures. *Nucleic acids research*, 35(Web Server issue):W403–W406, 2007.

[97] Angelo Keramidas, Andrew J Moorhouse, Peter R Schofield, and Peter H Barry. Ligand-gated ion channels: mechanisms underlying ion selectivity. *Progress in Biophysics and Molecular Biology*, 86(2):161–204, 2004.

[98] Jane E Carland, Michelle A Cooper, Matthew R Livesey, Tim G Hales, John A Peters, and Jeremy J Lambert. Mutagenic analysis of the intracellular portals of the human 5-HT$_{3A}$ receptor. *The Journal of Biological Chemistry*, 288(44):31592–601, 2013.

[99] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11):1859–65, 2008.

[100] Stefan Fischer, Roland L. Jr. Dunbrack, and Martin Karplus. Cis-trans imide isomerization of the proline dipeptide. *Journal of the American Chemical Society*, 116:11931–11937, 1994.

[101] Gareth A. Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. PLUMED 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604–13, 2014.

[102] Nicolas Calimet, Manuel Simoes, Jean-Pierre Changeux, Martin Karplus, Antoine Taly, and Marco Cecchini. A gating mechanism of pentameric ligand-gated ion channels. *Proceedings of the National Academy of Sciences of the United States of America*, 110(42):E3987–E3996, 2013.

[103] Aleksij Aksimentiev and Klaus Schulten. Imaging α-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophysical Journal*, 88(6):3745–3761, 2005.

[104] Alessandro Pandini, Arianna Fornili, Franca Fraternali, and Jens Kleinjung. Detection of allosteric signal transmission by information-theoretic analysis of protein dynamics. *The FASEB Journal*, 26(2):868–881, 2012.

[105] Valentina Corradi, Besian I. Sejdiu, Haydee Mesa-Galloso, Haleh Abdizadeh, Sergei Yu. Noskov, Siewert J. Marrink, and D. Peter Tieleman. Emerging diversity in lipidprotein interactions. *Chemical Reviews*, 119(9):5775–5848, 2019.

[106] Mackenzie J. Thompson and John E. Baenziger. Structural basis for the modulation of pentameric ligand-gated ion channel function by lipids. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, page 183304, 2020.

[107] Carlos J. Baier, Jacques Fantini, and Francisco J. Barrantes. Disclosure of cholesterol recognition motifs in transmembrane domains of the human nicotinic acetylcholine receptor. *Scientific Reports*, 1(1):69, 2011.

[108] F.J. Barrantes. Structural basis for lipid modulation of nicotinic acetylcholine receptor function. *Brain Research Reviews*, 47:71 – 95, 2004.

[109] Grace Brannigan, Jérôme Hénin, Richard Law, Roderic Eckenhoff, and Michael L. Klein. Embedded cholesterol in the nicotinic acetylcholine receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 105(38):14418–14423, 2008.

[110] John E. Baenziger, Jaimee A. Domville, and J.P. Daniel Therien. Chapter four - the role of cholesterol in the activation of nicotinic acetylcholine receptors. In Irena Levitan, editor, *Sterol Regulation of Ion Channels*, volume 80 of *Current Topics in Membranes*, pages 95 – 137. Academic Press, 2017.

[111] Francisco J. Barrantes. Cholesterol effects on nicotinic acetylcholine receptor. *Journal of Neurochemistry*, 103(s1):72–80, 2007.

[112] Corrie J. B. daCosta and John E. Baenziger. A lipid-dependent uncoupled conformation of the acetylcholine receptor. *The Journal of Biological Chemistry*, 284(26):17819–17825, 2009.

[113] Corrie J. B. daCosta, Lopamudra Dey, J. P. Daniel Therien, and John E. Baenziger. A distinct mechanism for activating uncoupled nicotinic acetylcholine receptors. *Nature Chemical Biology*, 9(11):701–707, 2013.

[114] Jerome Henin, Reza Salari, Sruthi Murlidaran, and Grace Brannigan. A Predicted Binding Site for Cholesterol on the GABA A Receptor. *Biophysical Journal*, 106(9):1938 – 1949, 2014.

[115] Marc A. Dämgen and Philip C. Biggin. State-dependent protein-lipid interactions of a pentameric ligand-gated ion channel in a neuronal membrane. *bioRxiv*, 2020.

[116] Pramod Kumar, Yuhang Wang, Zhening Zhang, Zhiyu Zhao, Gisela D. Cymes, Emad Tajkhorshid, and Claudio Grosman. Cryo-em structures of a lipid-sensitive pentameric ligand-gated ion channel embedded in a phosphatidylcholine-only bilayer. *Proceedings of the National Academy of Sciences of the United States of America*, 117(3):1788–1798, 2020.

[117] Ailing Tong, John T. 2nd Petroff, Fong-Fu Hsu, Philipp Am Schmidpeter, Crina M. Nimigean, Liam Sharp, Grace Brannigan, and Wayland Wl Cheng. Direct binding of phosphatidylglycerol at specific sites modulates desensitization of a ligand-gated ion channel. *eLife*, 8:e50766, 2019.

[118] Camille M. Hénault, Cedric Govaerts, Radovan Spurny, Marijke Brams, Argel Estrada-Mondragon, Joseph Lynch, Daniel Bertrand, Els Pardon, Genevieve L. Evans, Kristen Woods, Benjamin W. Elberson, Luis G. Cuello, Grace Brannigan, Hugues Nury, Jan Steyaert, John E.

Baenziger, and Chris Ulens. A lipid site shapes the agonist response of a pentameric ligand-gated ion channel. *Nature Chemical Biology*, 15(12):1156–1164, 2019.

[119] Sandip Basak, Nicolaus Schmandt, Yvonne Gicheru, and Sudha Chakrapani. Crystal structure and dynamics of a lipid-induced potential desensitized-state of a pentameric ligand-gated channel. *eLife*, 6:e23886, 2017.

[120] Stephanie A. Heusser, Marie Lycksell, Xueqing Wang, Sarah E. McComas, Rebecca J. Howard, and Erik Lindahl. Allosteric potentiation of a ligand-gated ion channel is mediated by access to a deep membrane-facing cavity. *Proceedings of the National Academy of Sciences*, 115(42):10672–10677, 2018.

[121] Swarna M. Patra, Sudip Chakraborty, Ganesh Shahane, Xavier Prasanna, Durba Sengupta, Prabal K. Maiti, and Amitabha Chattopadhyay. Differential dynamics of the serotonin1A receptor in membrane bilayers of varying cholesterol content revealed by all atom molecular dynamics simulation. *Molecular Membrane Biology*, 32(4):127–137, 2015.

[122] Iqbal Mahmood, Xinli Liu, Saburo Neya, and Tyuji Hoshino. Influence of Lipid Composition on the Structural Stability of G-Protein Coupled Receptor. *Chemical and Pharmaceutical Bulletin*, 61(4):426–437, 2013.

[123] Md. Jafurulla, Bhagyashree D. Rao, Sugunan Sreedevi, Jean-Marie Ruysschaert, Douglas F. Covey, and Amitabha Chattopadhyay. Stereospecific requirement of cholesterol in the function of the serotonin1A receptor. *Biochimica et biophysica acta*, 1838(1 Pt B):158–163, 2014.

[124] Rosie Dawaliby, Cataldo Trubbia, Cédric Delporte, Matthieu Masureel, Pierre Van Antwerpen, Brian K. Kobilka, and Cédric Govaerts. Allosteric regulation of G protein-coupled receptor activity by phospholipids. *Nature chemical biology*, (1):35–39, 2016.

[125] Donald E. Elmore and Dennis A. Dougherty. Investigating lipid composition effects on the mechanosensitive channel of large conductance (mscl) using molecular dynamics simulations. *Biophysical journal*, 85(3):1512–1524, 2003.

[126] Laura Dominguez, Leigh Foster, John E. Straub, and D. Thirumalai. Impact of membrane lipid composition on the structure and stability of the transmembrane domain of amyloid

precursor protein. *Proceedings of the National Academy of Sciences of the United States of America*, 113(36):E5281–E5287, 2016.

[127] P. V. Escribá, A. Ozaita, C. Ribas, A. Miralles, E. Fodor, T. Farkas, and J. A. García-Sevilla. Role of lipid polymorphism in G protein-membrane interactions: nonlamellar-prone phospholipids and peripheral protein binding to membranes. *Proceedings of the National Academy of Sciences of the United States of America*, 94(21):11375–11380, 1997.

[128] Helgi I. Inglfsson, Manuel N. Melo, Floris J. van Eerden, Clement Arnarez, Cesar A. Lopez, Tsjerk A. Wassenaar, Xavier Periole, Alex H. de Vries, D. Peter Tieleman, and Siewert J. Marrink. Lipid organization of the plasma membrane. *Journal of the American Chemical Society*, 136(41):14554–14559, 2014.

[129] Siewert J. Marrink, Valentina Corradi, Paulo C.T. Souza, Helgi I. Inglfsson, D. Peter Tieleman, and Mark S.P. Sansom. Computational modeling of realistic cell membranes. *Chemical Reviews*, 119(9):6184–6226, 2019.

[130] Jos A. Poveda, A. Marcela Giudici, M. Lourdes Renart, Oscar Millet, Andrs Morales, Jos M. Gonzlez-Ros, Victoria Oakes, Simone Furini, and Carmen Domene. Modulation of the potassium channel kcsa by anionic phospholipids: Role of arginines at the non-annular lipid binding sites. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1861(10):183029, 2019.

[131] Victoria Oakes, Simone Furini, and Carmen Domene. Effect of anionic lipids on ion permeation through the kcsa k+-channel. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1862(11):183406, 2020.

[132] Ruyin Cao, Giulia Rossetti, Andreas Bauer, and Paolo CarIoni. Binding of the Antagonist Caffeine to the Human Adenosine Receptor hA2AR in Nearly Physiological Conditions. *PLOS ONE*, 10(5):1–13, 2015.

[133] Victoria Oakes and Carmen Domene. Influence of cholesterol and its stereoisomers on members of the serotonin receptor family. *Journal of Molecular Biology*, 431(8):1633 – 1649, 2019.

[134] Robin B. Chan, Tiago G. Oliveira, Etty P. Cortes, Lawrence S. Honig, Karen E. Duff, Scott A. Small, Markus R. Wenk, Guanghou Shui, and Gilbert Di Paolo. Comparative lipidomic analysis of mouse and human brain with alzheimer disease. *The Journal of biological chemistry*, 287(4):2678–2688, 2012.

[135] Frank W Pfrieger. Role of cholesterol in synapse formation and function. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1610(2):271 – 280, 2003.

[136] Andrey Filippov, Greger Orädd, and Göran Lindblom. The effect of cholesterol on the lateral diffusion of phospholipids in oriented bilayers. *Biophysical journal*, 84(5):3079–3086, 2003.

[137] Hua Li and Vassilios Papadopoulos. Peripheral-Type Benzodiazepine Receptor Function in Cholesterol Transport. Identification of a Putative Cholesterol Recognition/Interaction Amino Acid Sequence and Consensus Pattern. *Endocrinology*, 139(12):4991–4997, 1998.

[138] Jacques Fantini and Francisco J. Barrantes. How cholesterol interacts with membrane proteins: an exploration of cholesterol-binding sites including CRAC, CARC, and tilted domains. *Frontiers in physiology*, 4:31–31, 2013.

[139] Jacques Fantini, Coralie Di Scala, Luke S. Evans, Philip T. F. Williamson, and Francisco J. Barrantes. A mirror code for protein-cholesterol interactions in the two leaflets of biological membranes. *Scientific reports*, 6:21907–21907, 2016.

[140] Jufang Shan, George Khelashvili, Sayan Mondal, Ernest L. Mehler, and Harel Weinstein. Ligand-dependent conformations and dynamics of the serotonin 5-HT2A receptor determine its activation and membrane-driven oligomerization properties. *PLoS computational biology*, 8(4):e1002473–e1002473, 2012.

[141] David J. Smith, Jeffery B. Klauda, and Alexander J. Sodt. Simulation best practices for lipid membranes. *Living Journal of Computational Molecular Science*, 1(1):5966, 2018.

[142] Christian Kandt, Walter L. Ash, and D. Peter Tieleman. Setting up and running molecular dynamics simulations of membrane proteins. *Methods*, 41(4):475 – 488, 2007.

[143] William V Kraske and Donald B Mountcastle. Effects of cholesterol and temperature on the permeability of dimyristoylphosphatidylcholine bilayers near the chain melting phase transition. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1514(2):159 – 164, 2001.

[144] J. R. Silvius. Thermotropic phase transitions of pure lipids in model membranes and their modifications by membrane proteins. *Lipid-Protein Interactions*, pages 239–281, 1982.

[145] Daniel R. Roe and Thomas E. Cheatham. Ptraj and cpptraj: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, 9(7):3084–3095, 2013.

[146] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein. Mdanalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32:2319–2327, 2011.

[147] R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, D. L. Dotson, J. Domanski, S. Buchoux, I. M. Kenney, , and O. Beckstein. Mdanalysis: A python package for the rapid analysis of molecular dynamics simulations. *Proceedings of the 15th Python in Science Conference*, pages 98–105, 2016.

[148] Sajad Moradi, Amin Nowroozi, and Mohsen Shahlaei. Shedding light on the structural properties of lipid bilayers using molecular dynamics simulation: a review study. *RSC Advances*, 9:4644–4658, 2019.

[149] Sivaramakrishnan Ramadurai, Ria Duurkens, Victor V. Krasnikov, and Bert Poolman. Lateral diffusion of membrane proteins: consequences of hydrophobic mismatch and lipid composition. *Biophysical journal*, 99(5):1482–1489, 2010.

[150] Oliver S. Smart, Joseph G. Neduvelil, Xiaonan Wang, B.A. Wallace, and Mark S.P. Sansom. Hole: A program for the analysis of the pore dimensions of ion channel structural models. *Journal of Molecular Graphics*, 14(6):354 – 360, 1996.

[151] Zhong-Shan Wu, Hao Cheng, Yi Jiang, Karsten Melcher, and H. Eric Xu. Ion channels gated by acetylcholine and serotonin: structures, biology, and drug discovery. *Acta Pharmacologica Sinica*, 36(8):895–907, 2015.

[152] I-Shan Chen and Yoshihiro Kubo. Ivermectin and its target molecules: shared and unique modulation mechanisms of ion channels and receptors by ivermectin. *The Journal of Physiology*, 596(10):1833–1845, 2018.

[153] Sundeep S. Deol, Peter J. Bond, Carmen Domene, and Mark S.P. Sansom. Lipid-protein interactions of integral membrane proteins: A comparative simulation study. *Biophysical Journal*, 87(6):3737 – 3749, 2004.

[154] Javier Iglesias-Fernandez, Peter J. Quinn, Richard J. Naftalin, and Carmen Domene. Membrane phase-dependent occlusion of intramolecular glut1 cavities demonstrated by simulations. *Biophysical Journal*, 112(6):1176–1184, 2017.

[155] Gianni Klesse, Shanlin Rao, Stephen J. Tucker, and Mark S.P. Sansom. Induced Polarization in Molecular Dynamics Simulations of the 5-HT3 Receptor Channel. *Journal of the American Chemical Society*, 142(20):9415–9427, 2020.

[156] Robin A. Corey, Phillip J. Stansfeld, and Mark S.P. Sansom. The energetics of proteinlipid interactions as viewed by molecular simulations. *Biochemical Society Transactions*, 48(1):25–37, 2019.

[157] Bijun Tang, Steven O. Devenish, and Sarah C.R. Lummis. Identification of Novel Functionally Important Aromatic Residue Interactions in the Extracellular Domain of the Glycine Receptor. *Biochemistry*, 57(27):4029–4035, 2018.

[158] James Cory-Wright, Mona Alqazzaz, Francesca Wroe, Jenny Jeffreys, Lu Zhou, and Sarah C. R. Lummis. Aromatic residues in the fourth transmembrane-spanning helix M4 are important for GABA ρ receptor function. *ACS Chemical Neuroscience*, 9(2):284–290, 2018.

[159] Susanne Mesoy, Jennifer Jeffreys, and Sarah C. R. Lummis. Characterization of Residues in the 5-HT3 Receptor M4 Region That Contribute to Function. *ACS Chemical Neuroscience*, 10(7):3167–3172, 2019.

[160] Ana R.G.M. da Costa Couto, Kerry L. Price, Susanne Mesoy, Emily Capes, and Sarah C. R. Lummis. The M4 helix is involved in α7 nach receptor function. *ACS Chemical Neuroscience*, 11(10):1406–1412, 2020.

[161] Susanne M. Mesoy and Sarah C. R. Lummis. M4, the outermost helix, is extensively involved in opening of the $\alpha4\beta2$ nACh receptor. *ACS Chemical Neuroscience*, 12(1):133–139, 2020.

[162] Mackenzie J. Thompson, Jaimee A. Domville, and John E. Baenziger. The functional role of the m4 transmembrane helix in the muscle nicotinic acetylcholine receptor probed through mutagenesis and coevolutionary analyses. *Journal of Biological Chemistry*, 295(32):11056–11067, 2020.

[163] Svenja Haeger, Dmitry Kuzmin, Silvia Detro-Dassen, Niklas Lang, Michael Kilb, Victor Tsetlin, Heinrich Betz, Bodo Laube, and Günther Schmalzing. An intramembrane aromatic network determines pentameric assembly of Cys-loop receptors. *Nature Structural and Molecular Biology*, 17(1):90–99, 2010.

[164] Camille M. Hénault, Peter F. Juranka, and John E. Baenziger. The M4 transmembrane α-helix contributes differently to both the maturation and function of two prokaryotic pentameric ligand-gated ion channels. *Journal of Biological Chemistry*, 290(41):25118–25128, 2015.

[165] Constanza Alcaino, Maria Musgaard, Teresa Minguez, Simone Mazzaferro, Manuel Faundez, Patricio Iturriaga-Vasquez, Philip C. Biggin, and Isabel Bermudez. Role of the Cys-loop and transmembrane domain in the allosteric modulation of $\alpha4\beta2$ nicotinic acetylcholine receptors. *Journal of Biological Chemistry*, 292(2):551–562, 2017.

[166] Takamasa Tobimatsu, Yoshihiko Fujita, Kazuhiko Fukuda, Ken ichi Tanaka, Yasuo Mori, Takashi Konno, Masayoshi Mishina, and Shosaku Numa. Effects of substitution of putative transmembrane segments on nicotinic acetylcholine receptor function. *FEBS Letters*, 222(1):56–62, 1987.

[167] S. Pons, J. Sallette, J. P. Bourgeois, A. Taly, J. P. Changeux, and A. Devillers-Thiery. Critical role of the C-terminal segment in the maturation and export to the cell surface of the homopentameric $\alpha7$-5HT3A receptor. *European Journal of Neuroscience*, 20(8):2022–2030, 2004.

[168] Amy S. Butler, Sarah A. Lindesay, Terri J. Dover, Matthew D. Kennedy, Valerie B. Patchell, Barry A. Levine, Anthony G. Hope, and Nicholas M. Barnes. Importance of the C-terminus of the human 5-HT3A receptor subunit. *Neuropharmacology*, 56(1):292–302, 2009.

[169] Jaimee A. Domville and John E. Baenziger. An allosteric link connecting the lipid-protein interface to the gating of the nicotinic acetylcholine receptor. *Scientific Reports*, 8(1):3898, 2018.

[170] Aixin Cheng, Neil A McDonald, and Christopher N Connolly. Cell surface expression of 5-hydroxytryptamine type 3 receptors is promoted by RIC-3. *The Journal of biological chemistry*, 280(23):22502–7, 2005.

[171] M. Opekarová and W. Tanner. Specific lipid requirements of membrane proteins - A putative bottleneck in heterologous expression. *Biochimica et Biophysica Acta - Biomembranes*, 1610(1):11–22, 2003.

[172] Rosie Dawaliby, Cataldo Trubbia, Cdric Delporte, Caroline Noyon, Jean-Marie Ruysschaert, Pierre Van Antwerpen, and Cdric Govaerts. Phosphatidylethanolamine is a key regulator of membrane fluidity in eukaryotic cells. *Journal of Biological Chemistry*, 291(7):3658–3667, 2016.

[173] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. Charmm-gui: A web-based graphical user interface for charmm. *Journal of Computational Chemistry*, 29(11):1859–1865, 2008.

[174] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. 2015.

[175] Junmei Wang, Wei Wang, Peter a Kollman, and David a Case. Antechamber, An Accessory Software Package For Molecular Mechanical Calculations. *Journal of the American Chemical Society*, 222(2):U403, 2001.

[176] Axel D. Becke. A new mixing of hartreefock and local densityfunctional theories. *The Journal of Chemical Physics*, 98(2):1372–1377, 1993.

[177] P.H. Huenenberger, A.E. Mark, and W.F. van Gunsteren. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *Journal of Molecular Biology*, 252(4):492 – 503, 1995.

[178] Tung Ming Fong and Mark G. McNamee. Correlation between Acetylcholine Receptor Function and Structural Properties of Membranes. *Biochemistry*, 25(4):830–840, 1986.

[179] John E. Baenziger, Mary Louise Morris, Tim E. Darsaut, and Stephen E. Ryan. Effect of membrane lipid composition on the conformational equilibria of the nicotinic acetylcholine receptor. *Journal of Biological Chemistry*, 275(2):777–784, 2000.

[180] Robert A. Copeland, David L. Pompliano, and Thomas D. Meek. Drug–target residence time and its implications for lead optimization. *Nature Reviews Drug Discovery*, 5(9):730–739, 2006.

[181] Robert A. Copeland. The drug–target residence time model: a 10-year retrospective. *Nature Reviews Drug Discovery*, 15(2):87–95, 2016.

[182] Grant K Walkup, Zhiping You, Philip L Ross, Eleanor K H Allen, Fereidoon Daryaee, Michael R Hale, John O'Donnell, David E Ehmann, Virna J A Schuck, Ed T Buurman, Allison L Choy, Laurel Hajec, Kerry Murphy-Benenato, Valerie Marone, Sara A Patey, Lena A Grosser, Michele Johnstone, Stephen G Walker, Peter J Tonge, and Stewart L Fisher. Translating slow-binding inhibition kinetics into cellular and in vivo effects. *Nature chemical biology*, 11(6):416423, 2015.

[183] Rodrigo Casasnovas, Vittorio Limongelli, Pratyush Tiwary, Paolo Carloni, and Michele Parrinello. Unbinding kinetics of a p38 map kinase type ii inhibitor from metadynamics simulations. *Journal of the American Chemical Society*, 139(13):4780–4788, 2017.

[184] Nuria Plattner and Frank Noé. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and markov models. *Nature Communications*, 6(1):7653, 2015.

[185] Pratyush Tiwary, Vittorio Limongelli, Matteo Salvalaglio, and Michele Parrinello. Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proceedings of the National Academy of Sciences of the United States of America*, 112(5):E386–E391, 2015.

[186] Stefano Raniolo and Vittorio Limongelli. Ligand binding free-energy calculations with funnel metadynamics. *Nature Protocols*, 15(9):2837–2866, 2020.

[187] Pratyush Tiwary, Jagannath Mondal, and B. J. Berne. How and when does an anticancer drug leave its binding site? *Science Advances*, 3(5):1–8, 2017.

[188] Davide Branduardi, Francesco Luigi Gervasio, and Michele Parrinello. From A to B in free energy space. *The Journal of Chemical Physics*, 126(5):054103, 2007.

[189] Z. Faidon Brotzakis, Vittorio Limongelli, and Michele Parrinello. Accelerating the calculation of protein–ligand binding free energy and residence times using dynamically optimized collective variables. *Journal of Chemical Theory and Computation*, 15(1):743–750, 2019.

[190] Mattia Bernetti, Matteo Masetti, Maurizio Recanatini, Rommie E. Amaro, and Andrea Cavalli. An integrated markov state model and path metadynamics approach to characterize drug binding processes. *Journal of Chemical Theory and Computation*, 15(10):5689–5702, 2019.

[191] Riccardo Capelli, Anna Bochicchio, Giovannimaria Piccini, Rodrigo Casasnovas, Paolo Carloni, and Michele Parrinello. Chasing the Full Free Energy Landscape of Neuroreceptor/Ligand Unbinding by Metadynamics Simulations. *Journal of Chemical Theory and Computation*, 15(5):3354–3361, 2019.

[192] Riccardo Capelli, Paolo Carloni, and Michele Parrinello. Exhaustive Search of Ligand Binding Pathways via Volume-Based Metadynamics. *Journal of Physical Chemistry Letters*, 10(12):3495–3499, 2019.

[193] Susanta Haldar, Federico Comitani, Giorgio Saladino, Christopher Woods, Marc W. Van Der Kamp, Adrian J. Mulholland, and Francesco Luigi Gervasio. A Multiscale Simulation Approach to Modeling Drug-Protein Binding Kinetics. *Journal of Chemical Theory and Computation*, 14(11):6093–6101, 2018.

[194] Shenggao Zhou, R. Gregor Weiß, Li-Tien Cheng, Joachim Dzubiella, J. Andrew McCammon, and Bo Li. Variational implicit-solvent predictions of the dry–wet transition pathways for ligand–receptor binding and unbinding kinetics. *Proceedings of the National Academy of Sciences*, 116(30):14989–14994, 2019.

[195] Kerry L. Price, Kiowa S. Bower, Andrew J. Thompson, Henry A. Lester, Dennis A. Dougherty, and Sarah C. R. Lummis. A Hydrogen Bond in Loop A Is Critical for the Binding and Function of the 5-HT3 Receptor. *Biochemistry*, 47(24):6370–6377, 2008.

[196] Avron D. Spier and Sarah C. R. Lummis. The role of tryptophan residues in the 5-hydroxytryptamine3 receptor ligand binding domain. *Journal of Biological Chemistry*, 275(8):5620–5625, 2000.

[197] Darren L. Beene, Gabriel S. Brandt, Wenge Zhong, Niki M. Zacharias, Henry A. Lester, and Dennis A. Dougherty. Cation-π Interactions in Ligand Recognition by Serotonergic (5-HT3A) and Nicotinic Acetylcholine Receptors: The Anomalous Binding Properties of Nicotine. *Biochemistry*, 41(32):10262–10269, 2002.

[198] Henry A. Lester, Mohammed I. Dibas, David S. Dahan, John F. Leite, and Dennis A. Dougherty. Cys-loop receptors: new twists and turns. *Trends in Neurosciences*, 27(6):329–336, 2004.

[199] Sarah C R Lummis. 5-HT3 receptors. *The Journal of biological chemistry*, 287(48):40239–45, 2012.

[200] Dong Yan and Michael M. White. Spatial Orientation of the Antagonist Granisetron in the Ligand-Binding Site of the 5-HT3 Receptor. *Molecular Pharmacology*, 68(2):365–371, 2005.

[201] Andrew J Thompson, Claire L Padgett, and Sarah C R Lummis. Mutagenesis and molecular modeling reveal the importance of the 5-HT3 receptor F-loop. *The Journal of biological chemistry*, 281(24):1657616582, 2006.

[202] Kerry L. Price, Reidun K. Lillestol, Chris Ulens, and Sarah C. R. Lummis. Palonosetron-5-HT3 Receptor Interactions As Shown by a Binding Protein Cocrystal Structure. *ACS Chemical Neuroscience*, 7(12):1641–1646, 2016.

[203] David C Reeves, Muhammed F R Sayed, Pak-Lee Chau, Kerry L Price, and Sarah C R Lummis. Prediction of 5-HT3 receptor agonist-binding residues using homology modeling. *Biophysical journal*, 84(4):2338–44, 2003.

[204] Claudio Melis, P. L. Chau, Kerry L. Price, Sarah C. R. Lummis, and Carla Molteni. Exploring the binding of serotonin to the 5-HT3 receptor by density functional theory. *The Journal of Physical Chemistry B*, 110(51):26313–26319, 2006.

[205] Eleftherios Zarkadas, Hong Zhang, Wensheng Cai, Gregory Effantin, Jonathan Perot, Jacques Neyton, Christophe Chipot, Guy Schoehn, Francois Dehez, and Hugues Nury. The binding of palonosetron and other antiemetic drugs to the serotonin 5-HT3 receptor. *bioRxiv*, 2020.

[206] Xiaolin Cheng, Hailong Wang, Barry Grant, Steven M Sine, and J Andrew McCammon. Targeted molecular dynamics study of C-loop closure and channel gating in nicotinic receptors. *PLOS Computational Biology*, 2(9):e134, 2006.

[207] Federico Comitani, Netta Cohen, Jamie Ashby, Dominic Botten, Sarah C. R. Lummis, and Carla Molteni. Insights into the binding of gaba to the insect rdl receptor from atomistic simulations: a comparison of models. *Journal of computer-aided molecular design*, 28(1):35–48, 2014.

[208] Jingyi Wang and Jon Lindstrom. Orthosteric and allosteric potentiation of heteromeric neuronal nicotinic acetylcholine receptors. *British Journal of Pharmacology*, 175(11):1805–1821, 2018.

[209] Leela S. Dodda, Julian Tirado-Rives, and William L. Jorgensen. Unbinding Dynamics of Non-Nucleoside Inhibitors from HIV-1 Reverse Transcriptase. *The journal of physical chemistry. B*, 123(8):1741–1748, 2019.

[210] Mohammad M. Sultan and Vijay S. Pande. tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *Journal of Chemical Theory and Computation*, 13(6):2440–2447, 2017.

[211] Luigi Bonati, Yue-Yu Zhang, and Michele Parrinello. Neural networks-based variationally enhanced sampling. *Proceedings of the National Academy of Sciences*, 116(36):17641–17647, 2019.

[212] Mutasem Omar Sinnokrot, Edward F. Valeev, and C. David Sherrill. Estimates of the Ab Initio Limit for π-π Interactions: The Benzene Dimer. *Journal of the American Chemical Society*, 124(36):10887–10893, 2002.

[213] J. P. Gallivan and D. A. Dougherty. Cation-π interactions in structural biology. *Proceedings of the National Academy of Sciences of the United States of America*, 96(17):9459–9464, 1999.

[214] Jamie A Ashby, Ian V McGonigle, Kerry L Price, Netta Cohen, Federico Comitani, Dennis A Dougherty, Carla Molteni, and Sarah C R Lummis. GABA binding to an insect GABA receptor: a molecular dynamics and mutagenesis study. *Biophysical Journal*, 103(10):2071–81, 2012.

[215] Claudio Melis, Sarah C. R. Lummis, and Carla Molteni. Molecular dynamics simulations of GABA binding to the GABA C receptor: The Role of Arg104. *Biophysical Journal*, 95(9):4115–4123, 2008.

[216] Xavier Lucas, Antonio Bauz, Antonio Frontera, and David Quionero. A thorough anionπ interaction study in biomolecules: on the importance of cooperativity effects. *Chemical Science*, 7:1038–1050, 2016.