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

KING'S College LONDON

COMPUTATIONAL MODELLING OF ACTIVATION MECHANISMS IN LIGAND-GATED ION CHANNELS

Comitani, Federico

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.

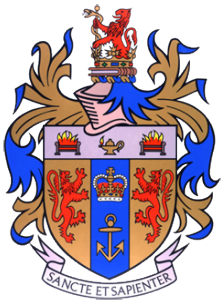
Take down policy

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: 08. Oct. 2023

King's College London

SCHOOL OF NATURAL & MATHEMATICAL SCIENCES
Department of Physics



COMPUTATIONAL MODELLING OF ACTIVATION MECHANISMS IN LIGAND-GATED ION CHANNELS

Federico Comitani

A thesis submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy

First Supervisor:
Prof. CARLA MOLTENI

Second Supervisor:
Dr. CHRIS LORENZ

February 2017

Sometimes science is more art than science, Morty.
A lot of people don't get that.
- Rick Sanchez C-137

Abstract

Pentameric ligand-gated ion channels (pLGICs) are a family of membrane proteins that mediate the fast synaptic communication between neurons. The delicate role played by these neuroreceptors makes them major therapeutic targets for a variety of neuronal disorders, including Alzheimer's disease, which can be attributed to flaws in neuronal signalling. pLGICs are formed by five subunits arranged around an ion permeable pore and are activated by the binding of small molecules, the neurotransmitters, which triggers a wave of conformational changes culminating with the opening (gating) of the ion channel. In spite of their importance, we still have a very limited understanding of their behaviour and only in recent years the availability of structural information from experiments and progress in computational methods are opening new possibilities to tackle their complexity at the atomic level.

In this project, we investigated the activation mechanisms of prototypical pLGICs with the aid of a number of state-of-the-art and novel computational techniques. In particular, molecular dynamics and metadynamics, a powerful enhanced sampling method to accelerate rare events, were employed to study the binding process of the neurotransmitter GABA to the insect RDL receptor, complementing mutagenesis electrophysiology experiments; binding free energy landscapes and affinities were evaluated both for the wild-type and selected mutants and plausible binding paths were identified.

The role of the *trans-cis* isomerisation of a highly conserved proline on the gating of the serotonin-activated 5-HT_3 receptor was also explored with metadynamics. Experiments found this single process to be crucial for the correct functioning of the gate and our simulations showed a correlation between the isomerisation of this proline and conformational changes in the helices lining the pore.

Understanding the conformational preference of protein building blocks may thus hold the key to fully comprehend the behaviour of whole receptors. Hence, a novel protocol, based on metadynamics and the dimensionality reduction algorithm sketch-map, was developed to map the conformational free energy of amino acids. As an example, it was applied here to aspartic acid, for which a number of stable conformers, including those experimentally observed with rotational spectroscopy, were identified.

The results obtained in this thesis contribute to the fundamental understanding

of pLGICs by helping unravel the complexity of the neurotransmitter binding and channel gating processes. Given the number of experimental structures recently made avalable, they are particlarly timely in paving the way for future studies on this important family of proteins.

Acknowledgements

This work would have not been the same without the contribution of many people.

First and foremost, I would like to express my deepest gratitude to my supervisor Carla Molteni. Her invaluable guidance, extraordinary support, and constant advice were fundamental throughout these years.

I would like to congratulate with my examiners Syma Khalid and Fabrizio Cleri for managing to read the entire thesis without falling asleep. Thank you for the useful comments and interesting discussion.

I would like also to thank my second supervisor Chris Lorenz for his constant patience and support.

I greatly appreciate the feedback received during these years by the members of the computational biophysics community, my friends and colleagues Max, CJ, Sarah, Kevin, Thras and all the people in the physics and chemistry department at King's College London, who gave me the chance to exchange ideas and improve my work.

My gratitude goes also to all our collaborators from both sides of the ocean, Sarah Lummis, Vittorio Limongelli, Michele Ceriotti, Maria Sanz and Andreas Mershin, for the insightful discussions and suggestions which were essential to my understanding of the chemistry and biology behind our projects.

Last but not least, I would like to express a special thank you to all the friends and family who were close to me in these years.

Contents

List of Figures **v**

List of Tables **x**

1 Introduction **1**
- 1.1 Ligand-Gated Ion Channels . . . 3
- 1.2 The Ligand Binding Mechanism . . . 8
  - 1.2.1 The Resistant to Dieldrin GABA Receptor . . . 10
- 1.3 The Gating Mechanism . . . 12
  - 1.3.1 The 5-HT_3 Receptor . . . 16
  - 1.3.2 The Conformational Landscape of Amino Acids . . . 19

2 Methods **22**
- 2.1 Molecular Dynamics . . . 23
- 2.2 Metadynamics . . . 41
  - 2.2.1 Well-Tempered Metadynamics . . . 44
  - 2.2.2 The Reweighting Algorithm . . . 45
  - 2.2.3 Funnel Metadynamics . . . 46
- 2.3 Sketch-map . . . 49

3 Results **52**
- 3.1 A Comparison of Models for the ECD of the RDL receptor . . . 53
- 3.2 The Binding of GABA to the RDL receptor with Metadynamics . . . 68
- 3.3 Proline isomerisation in the 5-HT_3 Receptor . . . 84
- 3.4 Conformational Analysis of Amino Acids . . . 116

4 Conclusions **146**

Appendix A Notes on the funnel restraining potential **151**

Bibliography **155**

List of Figures

1.1 Drawing of a neuron in the Cerebral Motor Cortex by Santiago Ramón y Cajal 1
1.2 Schematic representation of a neuron cell and its components 2
1.3 The sequence of activation of a ligand-gated ion channel 3
1.4 An X-ray crystallography structure of nAChR 5
1.5 Comparison of the transmembrane domains of the open GLIC and closed ELIC structures 7
1.6 Schematic representation of a hydrogen bond and cation-π interactions, electrostatic potential of the aromatic rings of tyrosine and phenylalanine calculated with DFT and the representation of a cation-π interaction between a portion of serotonin (5-HT) and a tyrosine in the 5-HT_3 receptor 9
1.7 Drosophila Melanogaster and chemical structure of the pesticide dieldrin 11
1.8 The neurotransmitter GABA in its zwitterionic form 12
1.9 Two adjacent subunits of the 5-HT_3 receptor, where the structures involved in the activation mechanism are highlighted 13
1.10 Proposed movements of the TMD secondary structures during the gating process of pLGICs 14
1.11 Trans and *cis* conformations of the proline amino acid 15
1.12 The serotonin neurotransmitter 16
1.13 X-ray crystallography structure of the mouse serotonin 5-HT_3 receptor 17
1.14 Closed and open conformations of the 5-HT_3R subunits according to the model proposed in ref. [1] 18
1.15 L-aspartic acid in its neutral and zwitterionic forms 20

2.1 Flow diagram of the Velocity Verlet integrator 26
2.2 Lennard-Jones pairwise potential 27
2.3 Geometry of a simple chain molecule, illustrating the definition of interatomic distance, bending angle and torsional angle 29
2.4 Schematic representation of periodic boundary conditions 34
2.5 Schematic representation of the TIP3P and TIP4P water models . . 36

2.6 Flow diagram of the preparation steps of an MD simulation 37
2.7 Comparison between the timescale of the typical processes of interest in biochemistry and that accessible to different scientific methods . . 39
2.8 Schematic representation of the effects of MD and Metadynamics on a one dimensional free energy landscape 45
2.9 The concept of funnel metadynamics 48
2.10 Example of the efficacy of sketch-map 51

3.1 Sequence alignments of the RDL receptor with AChBP, ELIC and GluCl 55
3.2 The selected homology models for the pentameric extracellular domain of the RDL receptor, after the structural optimisation 57
3.3 The GABA binding sites consisting of residues Phe146, Glu204, Phe206, Tyr254, Tyr109, Arg111 and Ser176, in the optimised homology models before GABA docking 58
3.4 Volume box used during the docking procedure of GABA to the RDL receptor ECD 59
3.5 Example of GABA docked in the extracellular domain of the RDL receptor (RDL-GluCl1 model), with important residues in the binding site 60
3.6 The dihedral angles θ and ψ of GABA 61
3.7 Time evolution of the dihedral angles θ and ψ which characterise the GABA conformers, and conformer relative occupancy during MD simulations in water, RDL-AChBP-5, RDL-ELIC-5, RDL-GluCl1-5 and RDL-GluCl2-5 62
3.8 Exploratory free energy map of GABA in the RDL-GluCl1 model as a function of the distance of the GABA amino group from Glu204 side chain (CV_{Glu}) and of the GABA carboxyl group from the Arg111 side chain (CV_{Arg}) 66
3.9 Principal and complementary subunits of the RDL receptor ECD and GABA in the binding site with the seven residues identified by experiments as important for binding and Thr251 in loop C 71
3.10 Free energy maps as a function of the biased CVs for GABA binding to the wild-type RDL receptor and the mutated Arg111Ala and Glu204Ala RDL receptors, their projected free energy profiles as a function of the position z and the evolution of the position z of GABA centre of mass along the funnel axis during the wild-type receptor metadynamics 73

3.11 Comparison of the projected free energy profile of the wild-type RDL receptor as a function of the position z along the funnel axis with or without the funnel correction 73
3.12 Free energy maps of the Glu204Ala-mutated RDL receptor, the wild-type RDL receptor and the Arg111Ala-mutated RDL receptor reweighted as a function of the distances between GABA charged groups and the C_αs of the receptor 111 and 204 residues 75
3.13 Binding free energy maps in the wild-type RDL receptor projected onto the funnel sections and typical conformations from the basins . 76
3.14 Snapshots of a favourable binding path observed in the metadynamics simulations for the wild-type RDL receptor 78
3.15 Evolution of the angle describing the orientation of loop C and its RMSD during GABA first unbinding event in the wild-type RDL receptor; the structure of the loop and the centre of mass of GABA are shown every 2 ns. 80
3.16 Metastable conformations observed during the binding of GABA to the mutant forms of RDL. 81
3.17 Two non-consecutive subunits taken from the original 5-HT_3R crystallographic structure 84
3.18 Section of the 5-HT_3R model used for this project and zoom of its interface between ECD and TMD 85
3.19 A single subunit of the unmodified X-ray structure of the serotonin mouse receptor 5-HT_3R and its amino acid sequence 88
3.20 Proline dipeptide with the atoms relevant to define the metadynamics CVs labelled 89
3.21 The subsequent *trans-cis* isomerisations of Pro281 in the five subunits of the 5-HT_3 receptor described through a sigmoid function 91
3.22 RMSD of the secondary structures of the five subunits of the 5-HT_3R during the 5Trans, 3Cis and 5Cis simulations 92
3.23 Average density of the lipid heads and tails during the three 250 ns long MD simulations of the 5-HT_3R 93
3.24 POPC molecule wedged between the M2 and M3 helices of the TMD of 5-HT_3R 94
3.25 Number of hydrogen bonds observed between the residues belonging to the interface loops and water molecules surrounding them during the MD simulations of the 5-HT_3R 96
3.26 Tables showing the presence of interloop and intraloop hydrogen bonds between the residues of the M2-M3 loop and the residues of the loops belonging to the ECD during the MD simulations of the three 5-HT_3 receptor models 97

3.27 Snapshots of the M2-M3 loop section taken from the trajectories of the three 5-HT_3 receptor models 99
3.28 Number of hydrogen bonds formed between Thr280 in the M2-M3 loop and the phosphatidylcholine heads during the MD simulations in the three 5-HT_3R models and typical conformation assumed by Thr280 and Pro281 when the proline is either in a *trans* state or in a *cis* state 101
3.29 Free energy surfaces of the solvated proline dipeptide and of the proline in the 5-HT_3R environment as a function of the torsional angle ξ and ψ and corresponding potentials of mean force 103
3.30 Polar and twist angle of the M2 helices with respect to the main axis of the 5-HT_3 channel, monitored during the three MD simulations, for the 5-HT_3R models 105
3.31 Polar and twist angle of M2 helices with respect to their respective M3 helices of the 5-HT_3 channel, monitored during the three MD simulations, in the three models 106
3.32 The position of water molecules and charged ions along the central pore of the 5-HT_3 receptor during the three MD simulations, and a snapshot of a section of the 5Trans receptor and its inner pore surface 108
3.33 Distribution of the position of the C_α in the side chains of Leu260 during 250 ns simulations for the three 5-HT_3R models and a view of the M2 helix 109
3.34 Free energy landscape from metadynamics for the 5-HT_3 receptor reweighted as a function of the proline isomerisation angle ξ and the M2 helix bend angle 110
3.35 The dynamics of the M2 helices group projected onto the space of the first two PCA eigenvectors of 5Trans and a representation of such vectors 112
3.36 Neutral and zwitterionic aspartic acid, with relevant atoms and torsional angles labelled 117
3.37 Projected free energy maps from metadynamics simulations for neutral Asp in the gas phase as a function of pairs of torsional angles calculated with the CHARMM36 force field 122
3.38 Sketch-map of neutral Asp in the gas phase obtained by a dimensional reduction of the free energy as a function of the β, γ and ζ CVs . . . 124
3.39 Free energy maps for neutral Asp in the gas phase from metadynamics reweighted as a function of the neglected torsional angles α, γ and ϵ against those who were chosen as CVs (β, δ and ζ) with the CHARMM36 force field 127

3.40 Sketch-map of neutral Asp in the gas phase obtained from the five-dimensional space of the torsional angles α, β, γ, δ and ζ 128
3.41 Free energy surface of zwitterionic Asp in water as a function of β and δ 131
3.42 Sketch-map of zwitterionic Asp in water and its neutral version in the gas phase obtained starting from the three-dimensional space of the torsional angles α, β and γ 134
3.43 Comparison of the free energy sketch-map of neutral Asp in water with the corresponding sketch-map in the gas phase and of the free energy map from metadynamics as a function of β and δ of zwitterionic Asp in the gas phase and the corresponding map in water 136
3.44 Average numbers of water-mediated intramolecular hydrogen bonds per free energy basin in the metadynamics simulation of the zwitterionic form of Asp in water with the CHARMM36 force field 138
3.45 Average number of water-mediated intramolecular hydrogen bonds per basin during the metadynamics simulation of the neutral form of Asp in water with the CHARMM36 force field 139
3.46 Projected free energy maps from metadynamics for neutral Asp in the gas phase as a function of pairs of torsional angles calculated with the AMBER ff99SB-ildn force field and the OPLS-AA force field 142
3.47 Free energy sketch-maps for neutral Asp in the gas phase as a function of pairs of torsional angles calculated with the AMBER ff99SB-ildn and OPLS-AA force field and their comparison with the maps obtained with CHARMM36 143

List of Tables

3.1 Average number of hydrogen bonds formed by GABA in the RDL-models and in water solution 63
3.2 Average number of cation-π interactions formed by GABA in the RDL models 65
3.3 Enthalpic (ΔH) and entropic ($T\Delta S$) contributions and free energy of binding (ΔG) of GABA to the RDL receptor models calculated with the MM/GBSA and MM/PBSA methods 65
3.4 Free energies of neutral Asp *in vacuo* stable conformers relative to their absolute minimum calculated with the CHARMM36 force field and metadynamics simulations, DFT and MP2 125
3.5 Free energies of the conformers of neutral Asp in the gas phase, identified in the sketch-map of Fig. 3.40, relative to their absolute minimum, calculated with the CHARMM36 force field and metadynamics simulations, DFT and MP2 129
3.6 The free energies for the stable conformers of zwitterionic Asp in water, calculated with metadynamics and the CHARMM36 force field, DFT and MP2 (with implicit solvent), with respect to their respective absolute minima 132
3.7 Free energies of neutral Asp conformers *in vacuo* relative to their absolute minimum calculated with the CHARMM36, AMBER99-ildn and OPLS-AA force fields from the sketch-map with three CVs . . . 143

Chapter 1

Introduction

In the human body, an impressive range of functions is controlled by a single organ, the brain. It manages the heart rate and other voluntary and involuntary muscular movements, acquires and processes information from external stimuli and designs adequate responses to the environment; moreover, it is involved in learning, dreaming, memory and emotions. In order to operate this extensive apparatus, the brain is provided with about 10^{12} nerve cells, which can create $10^3 - 10^4$ connections each, making it the most complex living structure known to date [2, 3].

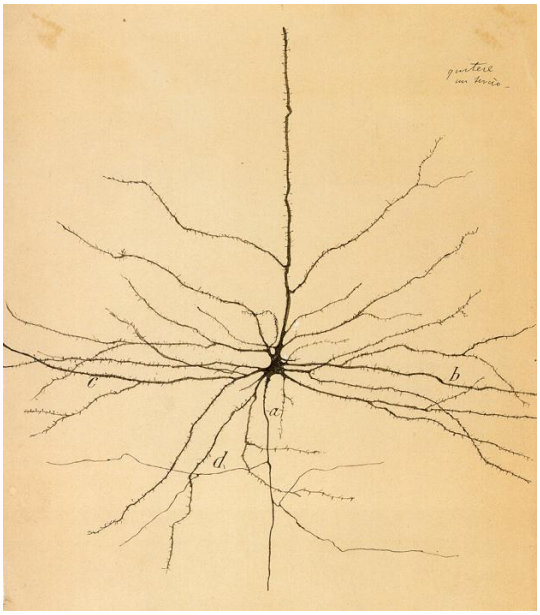


Figure 1.1: Drawing of a neuron in the Cerebral Motor Cortex by Santiago Ramón y Cajal (1852-1934), Spanish neuroscientist and Nobel laureate famous for its pioneering work in describing and depict the complexity of the neuronal structures. Here, the body, the axon and the network of dendrites are clearly visible.

These basic working units are known as neurons: they are very specialised cells, composed of a main body, dendrites, and an axon and they are supported externally

and protected by the glial cells. While the body contains all the substances necessary for the functioning of the cell, i.e. the nucleus and the cytoplasm, the axon is an electrically excitable part which extends into many smaller branches before ending at nerve terminals. On the contrary, dendrites are devised to receive signals from other neurons. The terminal part of the axon can form contact points, the synapses, on the dendrites and the cell body of other neurons. Here, chemical or electrical signals are exchanged. Almost $10^{12} - 10^{14}$ of these junctions are present in the human brain: an extensive network which is characterised by a degree of plasticity, as it is able to modify the connection paths to better cope with problems and new situations [2–4]. In Fig. 1.2 a schematic representation of a neuron is shown.
Neuronal communication relies on a wide set of proteins with different structural and functional properties, the neuroreceptors. While some of these proteins can be activated by differences in the electrical potential on the two sides of the neuronal membrane, some of them are activated by the interaction with small molecules, the ligands or neurotransmitters, which act as carriers of the information outside the neuronal cell.

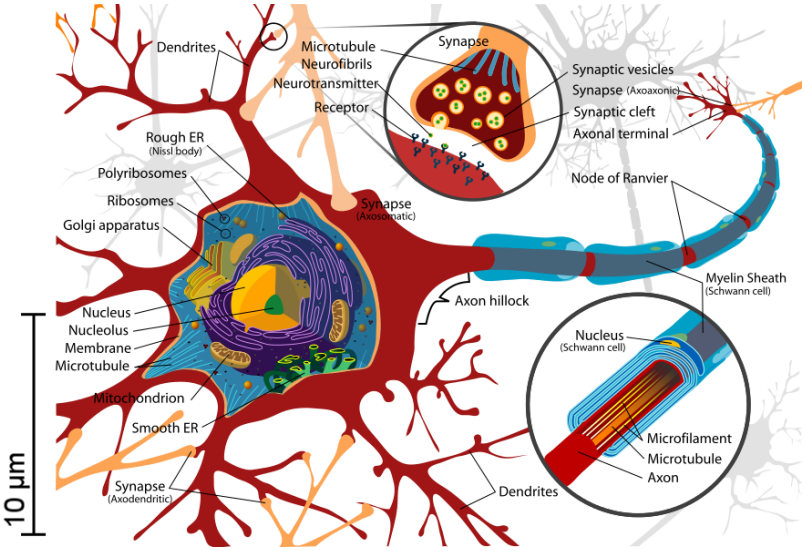


Figure 1.2: Schematic representation of a neuron cell and its components. On the top, a close view of a synapse is shown. (Image from Wikimedia Commons, Mariana Ruiz Villarreal, 2007)

Neurotransmitter activated receptors can be divided in two groups with radically different structure and function: G-protein coupled receptors (GPCRs) [5] and ligand-gated ion channels (LGICs). Signal transduction of GPCRs is metabotropic: when activated, these proteins transfer the information via a chain of events involving other compounds, i.e. through secondary messengers. On the contrary, LGICs are ionotropic: they feature an ion permeable pore whose opening and closure is key to the fast synaptic transmission between neurons [6]. They are multi-domain proteins

embedded in the membrane of nerve cells and have been the topic of study in this thesis.

1.1 Ligand-Gated Ion Channels

LGICs are composed of an extracellular domain, which is located outside the cell, a transmembrane domain, spanning the membrane, and an intracellular domain, located in the cytosol. These proteins work as bridges between what is inside and what is outside the cell, interacting and responding to external stimuli and modifying the difference of electric potential between the internal and the external environments. During the synaptic transmission, the information travels within the cell via electrical signals, while the exchange of information outside the membrane is mediated by the neurotransmitters, which diffuse across a synaptic cleft from the presynaptic to the postsynaptic neuron. These chemical messengers can open pathways to the signals by binding to ion channels, which in turn act as selective filters, allowing specific ions to flow into the cell. In Fig. 1.3 a representation of this mechanism is shown.

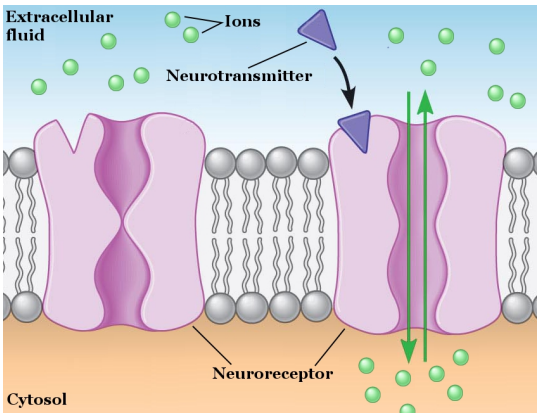


Figure 1.3: The sequence of activation of a ligand-gated ion channel (adapted from ref. [7]), where it is possible to observe the binding of the ligand which triggers the opening of the channel and the consequent flow of ions.

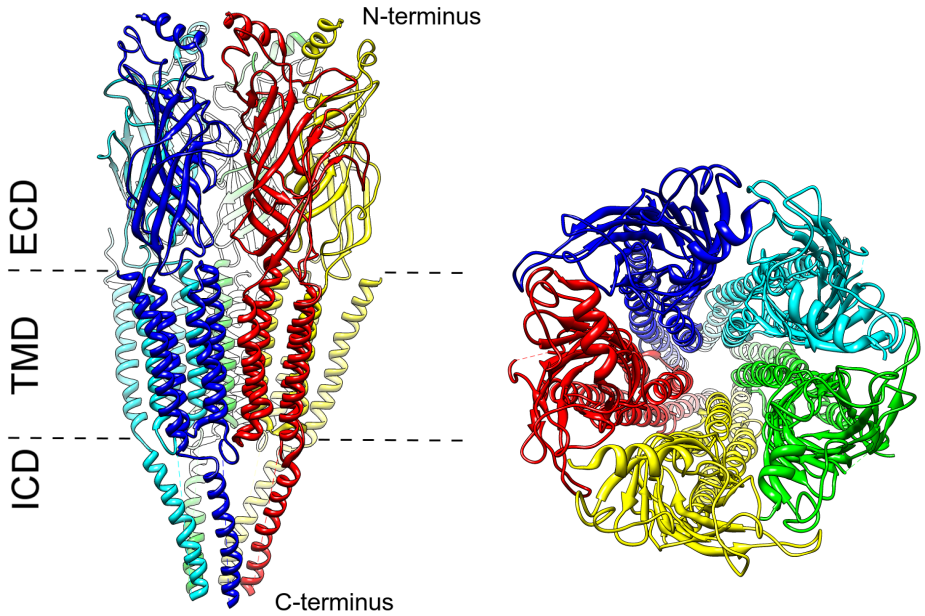
The change in internal electric potential generates (or inhibits) the electric signal allowing (or preventing) it to travel through the neuron body and axon. When this so-called action potential reaches the end of the axon, it triggers the release of neurotransmitters in the next cleft, propagating the message across the environment. Neurotransmitters can be excitatory, when they activate channels permeable to potassium (K^+) and sodium (Na^+) ions and stimulate the generation of the electric potential, or inhibitory, when they allow negative chlorine (Cl^-) ions to flow inside the cell membrane reducing it instead [7]. This simple mechanism can generate complex behaviours: at each stage particular reactions in the receiving cell can be switched on or off and the reactions chain can then culminate in a body response such as the

contraction of a muscle or the production of specific enzymes [2, 7].
There are two theories that try to describe the activation of LGICs. The first one is an induced-fit theory that states that conformational changes in the protein, which lead to the opening (gating) of the ion-conducting pore, are driven by the selective binding of ligands to a specific location in the protein (Fig. 1.3) [8].
The alternative explanation, known as the Monod-Wyman-Changeux (MWC) theory, postulates instead that open and closed conformations are both present and interconvert spontaneously at equilibrium: in this picture, each neurotransmitter binding event promotes the additive stabilisation of the open conformation, which takes effect as a concerted motion of all subunits, while the absence of ligands favours the closed state [9, 10].
The fundamental role of these proteins as information filters within the central nervous system has made them targets of numerous projects of drug design. As a matter of fact, malfunctioning channels can lead to serious neural diseases, such as Alzheimer's disease, Parkinson's disease, insomnia, epilepsy, anxiety, addictions and many others [11, 12]. Research for therapeutics aims at engineering artificial ligands, which can compete with or substitute the natural neurotransmitters and push the protein toward the desired response [13].

The Cys-loop superfamily

Most of the eukaryotic LGICs have highly conserved disulphide bridges, covalent bonds between two sulphur atoms belonging to different Cysteines (CYS) found at a distance of 15 residues in the extracellular domain [7, 14]. These structural motifs give the name to the Cys-loop receptors superfamily, whose members are then divided in two categories: the cation-selective (excitatory) receptors, like the nicotinic acetylcholine receptors (nAChR), in Fig. 1.5, and the serotonin 5-HT_3 receptors, and the anion-selective (inhibitory) members, such as the γ-aminobutyric acid ($GABA_{A/C}$) receptors, glycine receptors (GlyR) and glutamate-gated chloride channels (GluCl) [4, 11].
The Cys-loop family has specific conformational features: its members are pentameric LGICs (pLGICs), namely they are made of five subunits surrounding the ion permeable pore; each one of these subunits spans the cellular membrane and contains the disulphide bond in its N-terminal ECD, where also the binding site for their natural neurotransmitters is found, at the interface between two subunits. The TMD is characterised by a series of helices, composed of mostly hydrophobic residues in their central section and polar residues at their upper and lower ends. This facilitates the positioning of the protein within the phospholipid bilayer, thanks to polar interactions with the lipid heads at the membrane surface and hydrophobic interactions with the tails within its core. In Fig. 1.5 the alleged position of the lipid membrane is shown with dashed lines for the nACh receptor. Because of their

excitatory and inhibitory roles across the central nervous system, this family is therapeutic target for benzodiazepines, barbiturates, general anaesthetics and other sedative drugs [14, 15].



(a) (b)

Figure 1.4: An X-ray crystallography structure of nAChR, entry 2BG9 [16] of the Protein Data Bank (PDB). Each subunit is colored independently from red to blue in anticlockwise order and the lipid membrane boundaries are shown with a dashed line.

Crystal Structures and Homology Models

The first challenge when working with membrane proteins such as pLGICs is the limited number of atomistic structures available in protein databases. This information is usually obtained through X-ray crystallography, but the hydrophobicity of the transmembrane section makes them particularly difficult to crystallise. Alternative techniques, such as Nuclear Magnetic Resonance (NMR) and Cryo-Electron Microscopy (CryoEM) are also available, but they still present challenges: they are expensive, they may be limited in the size of the samples and the resulting data is generally lower in resolution when compared to that of X-ray crystallography. To date, less than 3% of the structures found in the Protein Data Bank (PDB) database belong to membrane proteins, despite the fact that they represent between 20% to 30% of most organisms' proteins and account for more than 40% of drug design targets. [17, 18] While techniques to increase the efficacy of these structural investigation methods are being developed, one has to rely on the available structures, complementing with homology modeling and structure refinement methods.

One of the first significant structures made available was the nicotinic acetylcholine receptor, a non-specific cationic channel (i.e. it is permeable to many kinds of positive ions), that was the first pLGIC to be discovered [19] and was mapped at 4 Å resolution by means of electron cryomicroscopy [16, 20, 21]. The nAChR structure, shown in Fig. 1.5, can be formed by five identical or different subunits in various combinations; its associated transmitter is acetylcholine, which is involved in the activation of motor neurons and, thus, in the muscular movements. nAChR can also be activated by nicotine, an addictive stimulant, which acts as agonist ligand.
The term agonist is commonly used to define a ligand that fully (or partially) activates the neuronal receptor to which it attaches, while antagonist defines a ligand whose role is to block the action of an agonist, i.e. binding to the receptor in its place, but without necessarily activating it.

As mentioned before, to bypass the lack of structural information for the Cys-loop family, it is possible to exploit the homology modelling technique, which relies on the idea that proteins sharing evolutionary paths show similar folding patterns even with relatively low amino-acid sequence identity [22, 23]. However, an identity of more than 20% is required to obtain reliable models. The idea is then to use a known structure as template to produce a three dimensional model of a protein whose sequence is known but whose spatial arrangement is not. The starting object has to be related and homologous to the final product: the higher is the sequence identity, the better is the expected result.
In the case of Cys-loop superfamily, the first high resolution available template was the acetylcholine binding protein (AChBP), a globular protein which shares 15-20% of its amino acid sequence with the extracellular domain of Cys-loop pLGICs. Found in the glial cells of the garden snail *Lymnaea Stagnalis*, the structure of this protein was determined with a resolution of 2.2 Å in 2001 [24]. As a globular protein, the AChBP has no transmembrane or intracellular domain, making it easier to crystallise, but also limiting its use to studies of the extracellular domain of pLGICs, for instance to investigate the orthosteric binding site.
More recently, in 2008, another interesting receptor, extracted from the *Erwinia Chrysantemi* bacterium and known as ELIC, was crystallized. It is a pentameric pLGIC and its structure resolved at a 3.3 Å by means of X-rays studies [25]. The structure has a transmembrane and no intracellular domain, making it an adequate template for both the ECD and the TMD, but it does not belong to the Cys-loop superfamily, so it is less reliable when modelling the region around the disulphide bridges. The narrowing of just a few Angstroms in the partly hydrophobic pore of this structure with respect to other pLGIC models, suggests that the channel is in a closed conformation. In the following year, an open channel structure, GLIC, was crystallised from the bacterium *Gloeobacter violaceus* first at 3.1 Å and then at 2.9 Å

of resolution. To capture one of the first X-ray open channel structures, both groups, Hilf et al. and Bocquet et al., exploited the fact that GLIC is permanently activated by protons without observable desensitisation, namely without a re-closure of the channel after the binding and the activation process are completed, and crystallised the structure at low pH. Similarly to ELIC, extracellular and transmembrane domains are present in this case, but no Cys-loop [26, 27].

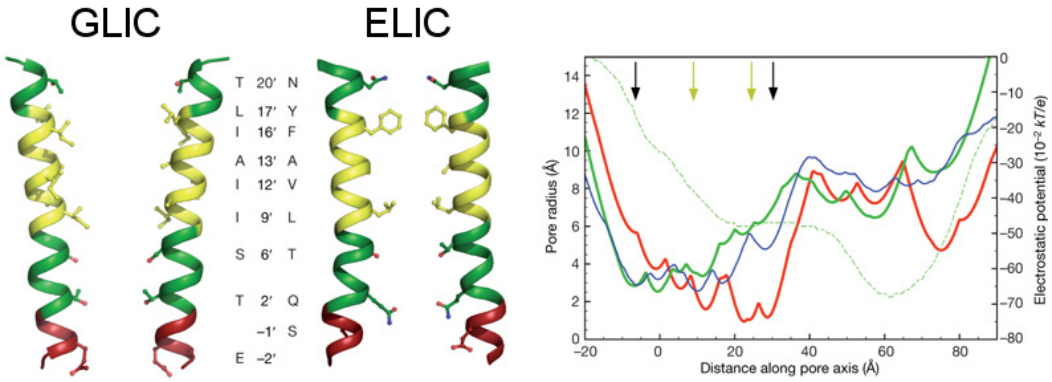


Figure 1.5: A comparison of the TMD helices for the open structure of GLIC and the closed ELIC. On the left, the M2 helices lining the pore are shown for two subunits. On the right, the pore radius for 1-nAChR-Tm (not discussed here, in blue), ELIC (in red) and GLIC (in green) along the channel axis is compared, and the electrostatic potential is represented as a dotted green line. Black and yellow arrows indicate the limits of the membrane and the hydrophobic rings, respectively. Adapted from Ref. [27]

In 2011 an X-ray structure of a Cys-loop receptor was resolved. The Glutamate-gated chloride channel (GluCl) is a homopentameric receptor found in *Caenorhabditis Elegans*, a nematode, or roundworm, commonly used as model organism. The structure is available at 3.3 Å of resolution: the extracellular domain is fairly similar to ELIC, with the exception of an additional helix at the N terminus, typical of proteins like AChBP and nAChR, while the transmembrane structure folds similarly to nAChR [15].

In 2014 the structure of a mouse homopentameric 5-HT_3 receptor was crystallised by Hassaine et al. [28] with a 3.5 Å resolution. While the ECD is similar to that that of the ELIC, GLIC and GluCl, this is the first structure to contain part of the conical intracellular domain. Together with the human $GABA_A$ receptor [29], the human glycine receptor-α3 [30], and the zebrafish *alpha*1-GlyR [31], the first two from X-ray spectroscopy and the latter from cryomicroscopy respectively, this remains one of the few structures belonging to Cys-loop family which has been successfully resolved.

1.2 The Ligand Binding Mechanism

Understanding movements and interactions underlying the binding of a small molecule to its target site within a protein is one of the main challenges when working with ligand-activated receptors. It is the first crucial step in the chain of events which characterise the activation process of pLGICs. Whether the aim is to control the channel response, by opening it when needed with agonist ligands, or preventing its activation with antagonists, or to regulate the binding of specific neuroreceptors, the knowledge of details at the atomic level can be highly beneficial to pharmacological studies and may help the design of new and better drugs.
Ligand binding studies usually focus on the interplay of hydrogen bonds and cation-π interactions and on the evaluation of binding free energies, which are indispensable to quantify the strength of the binding, yet extremely difficult to calculate.
Hydrogen bonds are attractive interactions between polar atoms involving hydrogen. As schematically shown in Fig. 1.6 (a), the attraction arises when a hydrogen atom (H) is bound to an highly electronegative atom, such as oxygen (O), nitrogen (N) or fluorine (F): the charge is displaced, causing a polarisation of the hydrogen which is then attracted to other highly electronegative atoms. The atom attached to the hydrogen is commonly known as donor, while the second electronegative atom, with whom the hydrogen bond is created, is defined as acceptor. The nature of these bonds, which depends on the charge displacement within the molecules, makes them distance and angle dependent, with maximum strength observed when the three atoms involved are in a linear arrangement. This strength can vary from 5 to 30 kJ/mol, placing them between weak van der Waals interactions and much stronger ionic and covalent bonds.
Hydrogen bonds are very frequent in both organic and inorganic systems, especially when water is involved [32]. When water is pervasive, as it is the case in most biological systems, hydrogen bonds mediated by water are also observed. Water molecules, whose OH groups can act both as acceptors and donors, can place themselves between other molecular acceptors and donors, bridging the interaction.
Similarly to hydrogen bonds, cation-π interactions, also in Fig. 1.6 (b), are frequently observed in biological systems: electron rich aromatic rings can attract cations or other positively charged amino acid moieties. They are distance and angle dependent and the maximum strength is observed when the cation lays on the axis normal to the plane of the carbon ring. Although generally comparable to that of hydrogen bonds, their strength is highly variable and it generally depends on the polarity of the solvent [33].

All neurotransmitters are characterised by a positively charged ammonium moiety, which allows for a number of hydrogen bond interactions with the negatively

charged residues found in what is known as the principal subunit, while interactions with the complementary subunit are less consistent among different pLGICs. The binding pocket is also typically provided with a number of aromatic residues which, caging the ligand, further stabilise its ammonium group [10, 11].

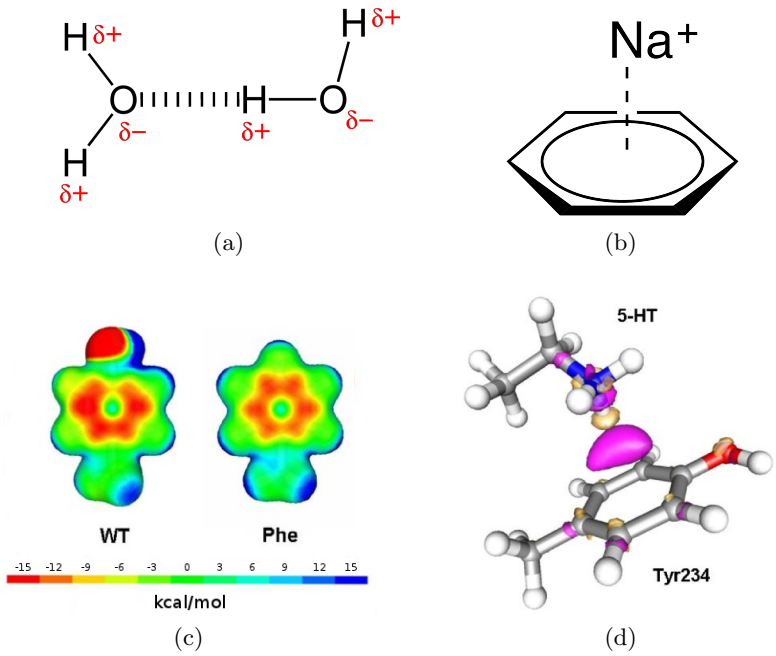


Figure 1.6: In (a), schematic representation of a hydrogen bond between two water molecules, where the difference in charge are shown. In (b), representation of a cation-π interaction between a generic aromatic ring and a positively charged sodium ion. In (c), electrostatic potential of the aromatic rings of tyrosine and phenylalanine calculated with density functional theory and mapped onto an electronic density isosurface of 0.01 $e/\mathring{A}^3$. In (d), representation of a cation-π interaction between a portion of serotonin (5-HT) and a tyrosine in the 5-HT_3 receptor; the electronic density changes due to the interaction are shown: in purple is the isosurface of electronic charge gain, while in orange is the isosurface of electronic charge depletion (adapted from ref. [34]).

Mutagenesis electrophysiology experiments are widely used to understand the role of specific amino acids in the functioning of pLGICs, but they are limited in the level of detail they can obtain. In site-directed mutagenesis, specific strands of DNA are modified to induce mutations localised on single residues. These experiments seek to observe the effect that properties, such us charge or shape, of specific residues have on the interactions and dynamics of the protein. Typical is the case where polar interactions are "deactivated" by substituting charged residues (e.g. Arg or Glu) with non-polar residues (e.g. Ala) or aromatic residues are substituted with amino acids lacking of carbon rings, thus inhibiting cation-π interactions.
The mutated proteins are then expressed in controlled environments, such as frog's eggs, and electrophysiology is applied to verify the correct functioning of the expressed proteins. This method requires to apply a current to the system and measure the output at different ligand concentrations. The value of the EC_{50}, i.e. the ligand

concentration needed to get half the maximum current output, is compared with that obtained with the wild-type receptor, namely the original non-mutated protein, or when other mutations are expressed. An increase of the EC_{50} means that a higher concentration of ligand is required to enforce the activation of the protein, probably due to a less favorable binding or less efficient gating induced by the mutation, while, on the other hand, a lower EC_{50} means that the mutation brought a more favourable environment to the ligand or a most efficient gating mechanisms. In some cases, mutations can cause the receptor to be unresponsive and the current output to drop to zero for the tested ligand concentrations: this may be the case when "deactivating" residues that are crucial for the opening of the channel, or when the new ligand-protein network of interactions does not suffice in maintaining a stable binding, thus preventing the flow of ions through the cell membrane. [35–37]
However these methods cannot tell us why these residues are important (or fundamental) to the activation process of the neuroreceptors, which specific bonds are disrupted or created upon mutation and how the binding free energy landscape changes. Computational simulations provide the tools to answer these and further questions, overcoming the limits of experiments in describing details at the atomistic level.

To date, a wide variety of computational methods to evaluate binding free energies are available, but they are mostly based on simplified static models and are limited in accuracy. Together with hydrogen bonds and cation-π interactions, polar and/or hydrophobic interactions and conformational entropies need to be also included when calculating this critical quantity: it is also clear that an accurate evaluation of the binding free energy cannot overlook the dynamics of the system and efforts are being made to develop new and reliable alternatives. In the first part of this thesis, we applied one of these novel techniques, the funnel metadynamics scheme, to the study of the ligand-neuroreceptor binding mechanism in an insect pLGIC. The aim was to complement experimental data on the system and improve the understanding of the very first stages in the activation mechanism of the RDL receptor, and more in general of the Cys-loop family. At the same time, a thorough assessment of the efficiency and efficacy of commonly used computational methods, such as protein-ligand docking, MD, molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) and molecular mechanics/generalised Born surface area (MM/GBSA) was carried out: this allowed to define a methodical protocol for the study of the ligand binding dynamics in pLGICs.

1.2.1 The Resistant to Dieldrin GABA Receptor

The first part of this thesis focuses on the study of the resistant to dieldrin (RDL) receptor, a GABA-gated ion channel of the Cys-loop superfamily involved in olfactory

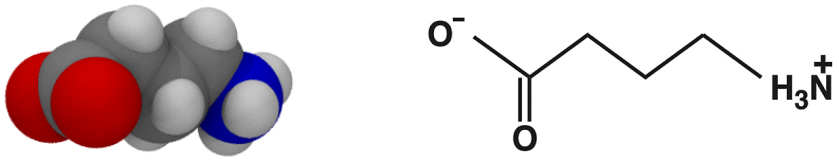
learning and insecticide resistance in the fruit fly *Drosophila Melanogaster* [38–42]. In Fig. 1.7 (b), the structure of dieldrin, a potent insecticide which gives the name to the receptor, is shown: this compound, which was once largely used as an alternative to dichlorodiphenyltrichloroethane (DDT), has proven toxic to animals and humans and is now banned worldwide [43].
RDL is relatively easy to express and readily available, making it an ideal candidate for experimental studies complemented by computational simulations. The work here presented is widely based on experimental research performed by the group led by Sarah C. R. Lummis in the Biochemistry Department at Cambridge University.

(a) (b)

Figure 1.7: Drosophila Melanogaster (Courtesy of Max Westby, licence Creative Commons 2.0, 2005) (left) and chemical structure of the pesticide dieldrin (from ref. [43]) (right).

γ-aminobutyric acid (GABA) is the main inhibitory neurotransmitter in the human central nervous system. It acts on ion channels and G protein-coupled receptors causing a negative change in the transmembrane potential, known as hyperpolarisation, and decreasing the chance for an action potential to occur. Its structure is shown in Fig. 1.8.
Among others, its action affects proteins belonging to the Cys-loop superfamily [44–46], the $GABA_A$ and $GABA_C$ neuroreceptors, chloride channels which are widely diffused in mammalian nervous systems and are involved in most brain functions. In particular, $GABA_C$ are a subclass of $GABA_A$, with which they share the anion selectivity of the channel pore; however, $GABA_C$ response is slower and display different features when treated pharmacologically. $GABA_C$ related receptors are mainly found in retinal neurons and their malfunction can bring problems in retinal signal processing. The RDL receptor bears resemblance to both $GABA_A$ and $GABA_C$ receptors and this study may be thus beneficial to their understanding [4, 47, 48].
The conformations and polarity of GABA depend on the environment, with the zwitterionic form commonly favoured in solution. In this state, the molecule presents opposite charges at its ends, as shown in Figure 1.8. The positive charge is localised on the amino group (NH_3^+), while the negative charge is found on the carboxyl moiety (COO^-).
The zwitterionic polarity grants the ability to form hydrogen bonds and cation-π

interactions with residues of the RDL receptor. The displacement of opposite charges at the GABA ends imposes strict requirements when looking at possible binding sites. The carboxyl group can bind to positively charged residues like arginine (Arg) and lysine (Lys), while the amino group is more likely to be found near negatively charged glutamate (Glu) or aspartate (Asp) [48]. Similarly, the position of the amino group can be influenced by the presence of aromatic residues, such as tyrosine (Tyr), phenylalanine (Phe), tryptophan (Trp) or histidine (His). Boxes or "cages" of aromatic residues, capable of surrounding and trapping cations or positive fragments, have been observed in some pLGICs binding pockets. [48–50].



(a) (b)

Figure 1.8: The neurotransmitter GABA in its zwitterionic form. In (a), atom types are shown in different colours: in red the oxygens, in blue the nitrogen, in grey the carbons and in white the hydrogens. The positive charge in the amino group is given by an extra hydrogen, while the negative charge in the carboxyl group is due to a missing hydrogen.

As in other pLGICs, the binding site of the RDL receptor is localised in the ECD, at the interface between two subunits. The pocket is partially covered by a flexible loop, known as loop C, whose actual role in the binding process is still elusive, in spite of its prominent position [51]. This neurotransmitter is homopentameric, i.e. is formed by five identical subunits, hence providing five possible binding pockets for GABA. It is still not clear how many ligands are required to fully activate the receptor, although studies suggest that at least two or three molecules may be necessary [14, 52].

1.3 The Gating Mechanism

Once the ligand has entered the binding pocket, the information has to travel through the receptor and reach the residues lining the ion permeable pore. It is still unclear how the activation signal is transmitted, leading to critical conformational changes in regions of the protein that can be more than 50 Å apart within the ms timescale. The extracellular domain is mainly composed of beta-sheets, while the transmembrane domain is characterised by four helices, M1-M4, with the M2 helix lining the ion permeable pore. Despite being topologically distinct, these domains are allosterically coupled and can interact through a number of highly flexible loops.

In the ECD, the Cys-loop, the $\beta 1 - \beta 2$ loop and the $\beta 8 - \beta 9$ loop (also known as F loop) are three well-conserved structures at the interface with the TMD. Some of these loops, shown in Fig. 1.9 for the 5-HT_3 receptor as an example, are likely to play a role in the transmission of the activation signal: localised in the same area is, in fact, the M2-M3 loop, a fragment of the TMD bridging two transmembrane helices (M2 and M3) with which they can interact. The role of these residues has also been confirmed by mutagenesis and electrophysiology experiments [53].

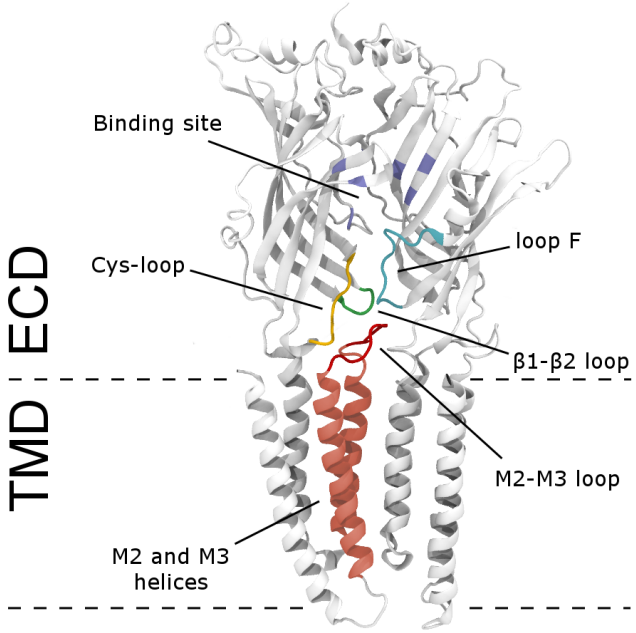


Figure 1.9: Two adjacent subunits of the 5-HT_3 receptor [28], where the structures involved in the activation mechanism are highlighted: in shades of red are the M2 and M3 helices and the M2-M3 loop joining them, in cyan is the $\beta 1 - \beta 2$ loop, in green is the F loop, in yellow the Cys-loop, and in purple some of the residues involved in the serotonin binding. The position of the lipid membrane is shown with dashed lines.

Position and orientation of the pore lining M2 helix can then, in turn, affect the opening or the closure of the central channel. To prevent the flow of ions during the resting and the desensitised state, different mechanisms of occlusion have been hypothesised; in most cases, however, they involve specific hydrophobic or polar residues lining the pore. In most pLGICs, rings of polar residues surrounding the pore have been observed at top of the M2 helices, acting as first charge filters, followed further down by rings of hydrophobic amino acids. Close to the middle of the helices are highly conserved leucine (Leu), or isoleucine (Ile) in a few exceptions, pointing toward the central axis and narrowing the channel to a hydrophobic constriction (or "hydrophobic gate"). These residues are believed to be the principal cause of what was proposed as a liquid-vapor transition in the region, the dewetting of the pore,

which acts as a barrier against ion conduction [54]. Down along the channel, within the M1-M2, loop highly variable rings of polar amino acids are expected to act as further ion selectivity filters during the actual permeation [11].
Various movements of the helices, such as bending or tilt and rotation around their axis in Fig. 1.10, have been identified during the opening process in different LGICs [55–57]. Although some of these studies suggest that the helices move as a rigid body [1], it is more plausible to assume that, starting with the binding of the neurotransmitter, a chain reaction or "wave" of small conformational fluctuations may cross the receptor, culminating with a movement of the hydrophobic gate. Even relatively small fluctuations in the residues lining the central pore can be sufficient to create enough space to trigger the ion and water flow [10].

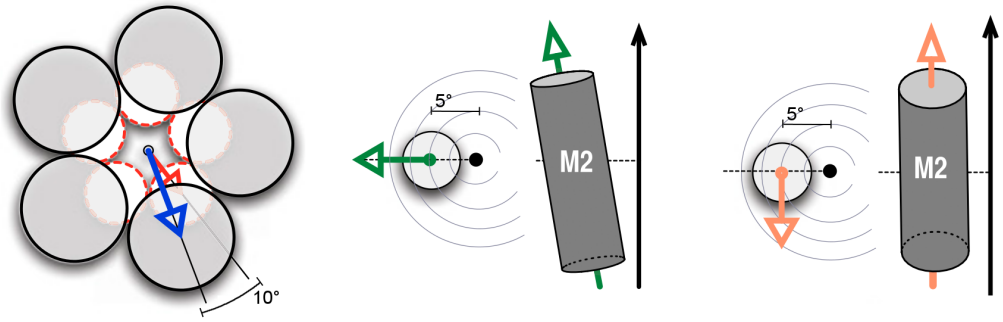


Figure 1.10: Proposed movements of the TMD secondary structures during the gating process of pLGICs. From left to right: twist of the five M2 helices (in blue) around the ion channel axis, tilt of the polar angle (in green) and tilt of the azimuthal angle (in orange) with respect to the channel axis(in black). Adapted from Ref. [57].

Whether acting as an hinge for a rigid body translation or as mediator for the transmission of small conformational rearrangements, the M2-M3 loop is expected to play a crucial role in this mechanism. Its sequence is in fact characterised by the presence of at least one proline (Pro), highly conserved across pLGICs.
Proline is a unique amino acid: its side chain forms a covalent bond with the amino group of the backbone giving rise to a characteristic cyclic structure. This distinctive intramolecular bond strips Pro of a hydrogen in the amino group, preventing it from forming hydrogen bonds. Indeed, this residue is known to disrupt hydrogen bonding networks in α-helices and β-sheets, making it ideal at the apex of loops and kinks, around which other secondary structures can then move [4, 58, 59].
Another feature that discriminates Pro from the other amino acids is its ability to access both the *trans* and *cis* conformations at ambient temperature. Amino acid peptides can assume different spatial arrangements, e.g. rotating around specific torsional angles, without change in their chemical formula, giving rise to what is known as geometric isomerism. In the case of peptide chains, the *cis* conformation, or

isomer, is defined when the C_αs of two consecutive residues are on the same side with respect to the peptide bond that joins them, while the *trans* isomer requires these two carbons to be on opposite sides of the bond, as shown in Fig. 1.11 in the case of proline. In most amino acids, the *cis* isomer is populated for less than $\sim 1\%$. The *cis* isomer of Pro is instead more energetically accessible and it has been observed that $\sim 5-6\%$ of all prolines are in this conformational state. The environment surrounding the proline can influence substantially its propensity to isomerise: polar solvents have been observed to increase the population of the *cis* isomer over non polar counterparts. The natural rate of interconversion between the two states is, however, extremely low, due to a high energy barrier of 20 kcal/mol [4, 60, 61].

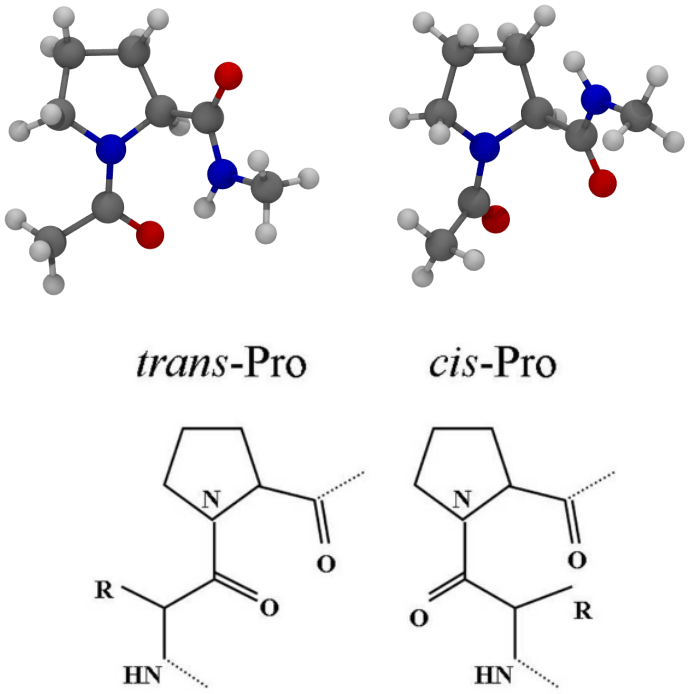


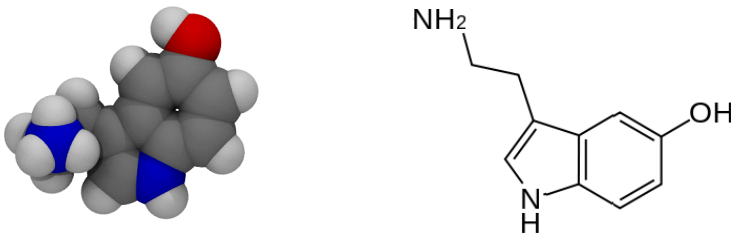
Figure 1.11: Trans and *cis* conformations of the proline amino acid.

Its strategical position in loops and kinks and the possibility of catalysing a trans-cis isomerisation are at the base of the hypothesis of the so called Pro8*, an highly conserved proline found at the top of the M2-M3 loop, as a possible molecular switch for channel gating in pLGICs [1, 56, 62]. Although this idea is based on the unlikely assumption that the M2 helix moves as a rigid body during the gating process, it is undeniable that these prolines, and more in general the M2-M3 loop, play a fundamental role in the opening of the receptor pore.
In the second part of this thesis, the behaviour of the loops at the interface between ECD and TMD, their interactions and structure under different conditions, was thus investigated to shed light on their role as carrier of the activation information in pLGICs. The work was carried out on a serotonin 5-HT_3 receptor, analysing the

structural features of its TMD and paying special care to the behaviour of Pro8*. The effects of proline isomerisation on the dynamics of the nearby secondary structures were also investigated.

1.3.1 The 5-HT$_3$ Receptor

The 5-hydroxytryptamine (5-HT) or serotonin receptors are found in most animals; they modulate the release of both inhibitory and excitatory neurotransmitters such as GABA, glutamate, dopamine and acetylcholine, and more in general they are involved in regulating a wide range of neurophysiological processes, such as anxiety, emotional and cognitive control, learning and memory but also biological processes such as appetite, sleep or sexual behaviour. At least fifteen 5-HT receptors have been discovered so far, grouped in seven families in relation to the signalling pathways they are involved in. Because of these multiple roles, they are subject to many studies of drug design in both the psychiatric and the medical fields [11, 63, 64]. The 5-HT$_3$ receptor (in Fig. 1.13) represents an exception, being the only pLGIC among 5-HT receptors, whereas the other members of this group are G-protein coupled receptors. This particular protein is linked to neurological disorders such as schizophrenia and drug abuse, and is the target of nausea suppressant drugs, designed to alleviate this unpleasant side effects commonly experienced after treatments with anesthetics or chemotherapy. [4, 64]



(a) (b)

Figure 1.12: The serotonin neurotransmitter. In (a), atoms are shown with different colours, while in (b) the Lewis formula is shown.

Serotonin (Fig. 1.12) is a highly versatile compound which is ubiquitous in the central nervous system. It is a derivative of tryptophan, an aromatic amino acid, and has been observed in most animals, in fungi and plants. This molecule is also found outside of the nervous system, where it participates in a variety of biological functions, like digestion, bone metabolism and organs growth and is consequently implicated in numerous diversified pathologies, such as mood disorders, migraine and irritable bowel syndrome (IBS). [63, 65]
The active form of serotonin is characterised by a double ring and a flexible tail

terminating in a positively charged amino group NH^+. This group can form hydrogen bonds with negatively charged residues and can interact with aromatic residues via cation-π interactions. An amino group NH found in the double ring can also form polar interactions and hydrogen bonds. Water molecules and other polar residues have also been observed interacting with the hydroxyl group at the other end of the molecule. As for the RDL receptor, it is not clear how many neurotransmitters are required to fully activate the 5-HT_3 receptor, but experimental studies seem to suggest that at least two binding sites need to be occupied for maximal activation [66].

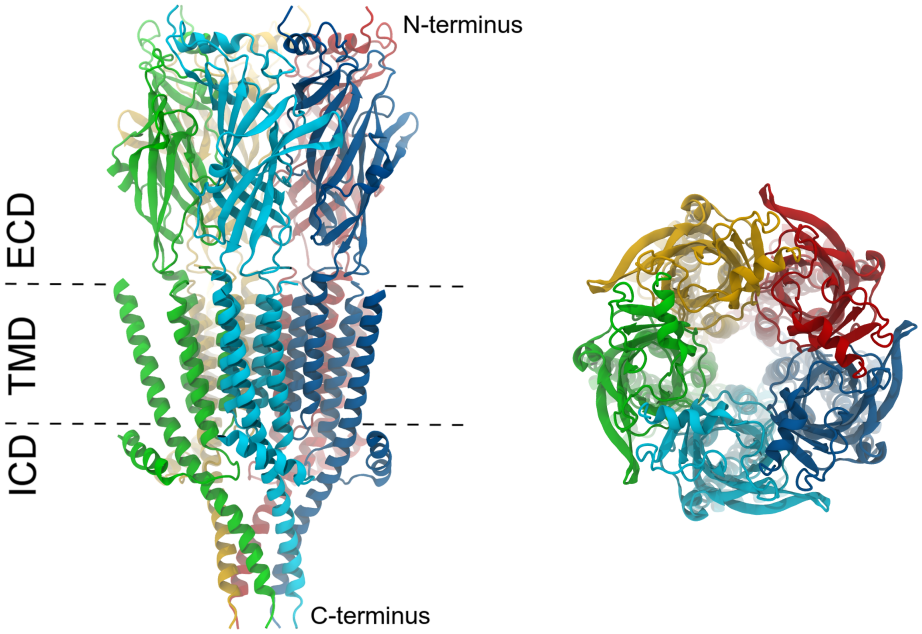


Figure 1.13: X-ray crystallography structure of the mouse serotonin 5-HT_3 receptor at 3.5Å resolution, PDB entry 4PIR [28]. The subunits are independently colored from red to blue in anticlockwise order, while the position of the lipid membrane is indicated with a dashed line.

There is great interest around the 5-HT_3R in the scientific community, in the light of the importance and versatility of functions that it regulates, and for this reason, it is subject to numerous experimental studies. Regarding his activation mechanisms, an interesting model of the gating has been proposed and tested with mutagenesis electrophysiology experiments, where Pro8* was mutated with proline analogues which favor or hinder its isomerisation. The results showed that when the isomerisation was suppressed, the channel were not responsive. From this outcome, an hypothetical mechanism was drawn, as shown in Fig. 1.14: the binding of serotonin agonists induces a rotation of the extracellular domain around the channel axis. This movement allows the $\beta 1 - \beta 2$ loop in the ECD and the M2-M3 loop in the TMD to come into contact. The latter receives the torsional energy from the ECD,

which is then transferred through the isomerisation of Pro8* to the M2 helix, the secondary structure lining the lower part of the ion permeable pore. The model assumes at this point a rigid body movement of the helix, which clears the pore from any hydrophobic residues occluding it [1].

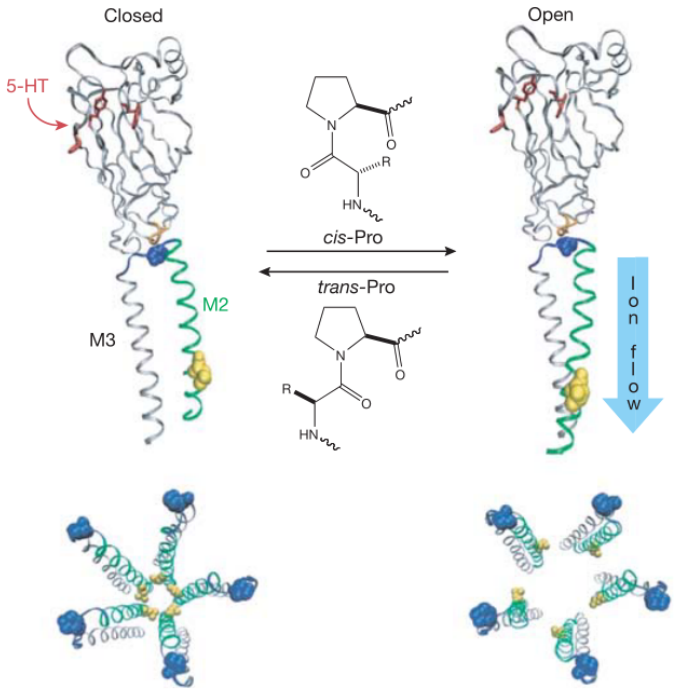


Figure 1.14: Closed and open conformations of the 5-HT_3R subunits according to the model proposed in ref. [1]. The binding site is shown in red, in orange is the $\beta 1 - \beta 2$ loop and in blue the M2-M3 loop, with the highly conserved proline highlighted. In yellow is leucine Leu286, the hydrophobic amino acid responsible for the occlusion of the channel.

Atomistic simulations can supply details on the dynamics of the M2 and M3 helices and the residues belonging to their connecting loop that may confirm or invalidate these hypotheses. However, until fairly recently, no X-ray structure of 5-HT_3 receptors was available: computational studies, generally aimed at investigating ligand binding or the gating hypothesis, were bound to the use of homology modelling to recover the atomic coordinates, reducing significantly the level of detail that could be achieved [4, 48]. To partly bypass this limitation, a 5-HT_3R chimera was experimentally build from a model of the acetylcholine-binding protein of *Aplysia Californica* (Ac-AChBP): the engineering exploited the structural similarities of the two proteins, apparent in particular in the overlap of ligand recognition mechanisms, which respond to both serotonin and acetylcholine agonists in invertebrates. Five 5-HT_3 binding sites were then inserted at the interface of the five subunits of Ac-AChBP. An X-ray crystallographic structure of this chimera at 2.4 Å was then made available [64]. Being the Ac-AChBP a globular protein, the 5-HT_3-AChBP chimera lacked a TMD and an ICD, leaving the problem unsolved for gating studies. On the other hand, it

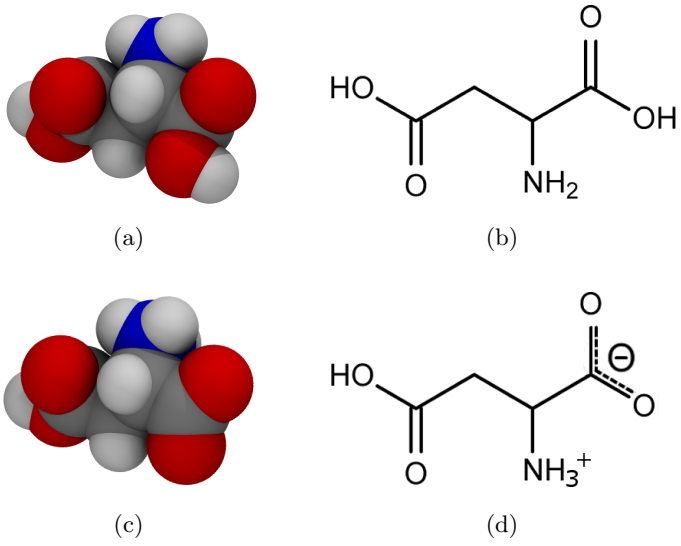
proved useful in works focusing on ligand binding, e.g. to investigate the affinity of varenicline, a potent nicotine partial agonist, whose side effect is to cause nausea in patients undergoing smoking cessation therapies, due to its ability to bind precisely to the 5-HT_3 receptor [67, 68].
The recently resolved structure of a mouse homopentameric 5-HT_3 receptor [28], shown in Fig. 1.13, includes the extracellular and transmembrane domains (and partially the intracellular domain), opening the possibility to model the whole protein embedded in a lipid membrane and thus greatly increasing the accuracy of the simulated environment. Although a mixture of different lipid species and cholesterol is normally found in biological membranes, in this work a simple model build from 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC) lipids was chosen. POPC lipids are widely used both in experiments and simulations to model eukaryotic cell membranes, and previous works on the same protein showed that the channel properties of the receptor are properly reproduced in such simplified environment [28, 69].

1.3.2 The Conformational Landscape of Amino Acids

The study of proline isomerisation can be framed within the wider field of the conformational analysis of amino acids and polypeptides. The case of proline in the M2-M3 loop as possible mediator of the gating process is a good example of how the behaviour of single amino acids can influence large biomolecules. Despite being the basic building blocks of living matter, the varied chemistry of side chains can give rise to extremely complex interactions, which in turn define macroscopic properties, such us the protein folding and their biological functions [70].
Experimental studies have thus focused on the identification of conformational properties of amino acids, both *in vacuo* and in aqueous solution. Typical experiments *in vacuo* require first to vaporise the amino acids by means of laser ablation, where the fast transfer of energy associated with short and intense laser pulses allow to separate molecules from the substrate without the risk of thermal damage and fragmentation. The resulting molecules are then mixed with inert gas and moved to a vacuum chamber, where they are polarised through pulses of microwave radiation. Treating the consequent relaxation decay time with Fourier transform allows to recover the spectra of the energies of transitions between rotational states, and from there the rotational constants and consequently the molecular geometry of the amino acids can be determined. In the condensed phase, instead, NMR and X-ray crystallography are generally used. Despite the progress in techniques and methodologies of recent years, the design and realisation of these experiments still represent a challenge and the results are often partial [71, 72]. Computational modelling provides thus the opportunity to complete the mapping of these conformations, thanks to advanced sampling techniques that allow an exhaustive exploration of the free energy landscape with reasonable computational resources. Even for simple molecules, such landscape

is characterised by a complex network of interconnected basins, which arise from the interplay of intramolecular interactions between polar groups and intermolecular bonds with solvent and other residues in the environment. Although it can be challenging to disentangle and understand, this map contains a wealth of information on conformers, favoured isomerisation paths and kinetics.

As a prototypical example, aspartic acid, in Fig. 1.15, was here chosen: this molecule has been extensively studied through rotational spectroscopy in its neutral form, stable in the gas phase, by our collaborators in the group of M. Eugenia Sanz in the Chemistry Department at King's College London, while X-ray diffraction and Nuclear Magnetic Resonance studies investigated its zwitterionic form, normally found in most biological environments. These gas phase experiments, however, were able to identify only six low energy isomers, whose stability was subsequently confirmed by *ab initio* calculations, but higher energy structures, outside the range detectable by the experimental apparatus, were not observed [72, 73].



(a) (b)

(c) (d)

Figure 1.15: L-aspartic acid in its neutral (top) and zwitterionic (bottom) forms. In (a) and (c), atom types are marked with different colours while in (b) and (d), the chemical structures are shown.

Among the natural acidic amino acids, aspartate has the simplest structure. It is an important mediator in the generation of other amino acids and large biomolecules, but it was also found to play a role in protein active sites and proton transfer phenomena. It can be found in two enantiomers, two forms which are mirror images, but while the L-aspartic acid is abundant in the human body, is found in proteins and is involved in regulating the levels of ammonia in the body, its chiral counterpart, D-aspartic acid, has limited biological functions [74]. In proteins, the side chain of aspartate is characterised by the presence of a negatively charged COO^- carboxyl

group and a positive NH_3^+. This zwitterionic form is stabilised by interactions with solvent molecules, but falls in the gas phase to a neutral form, through a proton transfer from the amino group to the carboxyl. The possibility to form intermolecular hydrogen bonds between polar or charged portions of the molecule is crucial in defining the shape of the conformational free energy surface.
In the last part of this thesis, novel computational techniques allowed us to exhaustively map the conformational free energy landscape of both neutral and zwitterionic aspartate under different conditions, recovering the conformers identified by rotational spectroscopy and complementing them with higher free energy structures. The protocol developed here could be then easily transferred to other compounds, amino acids or peptide chains, and would provide an efficient scheme to extract information on the role of their conformational dynamics in affecting the behaviour of bigger biological complexes.

Chapter 2

Methods

In the last decades, an amazing improvement in processing power, technologies and computational methods have been observed, breaking limits that were thought insurmountable before. The increase in power and efficiency has made computational modelling ubiquitous in science; it is now crucial both in solving more abstract and theoretical problems and in the applied industries and engineering, for the design of novel materials, drugs or chemicals.

The advantages introduced by these techniques to complement and sometimes replace experiments have now surpassed their cost and limitations. Experimental setups are often expensive and time consuming; as an example, storing for long periods of time chemical compounds or complex biological molecules can be difficult, if not impossible. Besides being a relatively cheap and efficient alternative, computational simulations can extract information on structures, dynamics and energies with molecular or atomic details, while also allowing to easily explore extreme or even unphysical environment and conditions, normally out of reach for experimental techniques. Computational modelling is now commonly used alongside with experiments, to test hypotheses and complement their data. However, experimental information still plays an essential role in the validation of the results; systematic errors, such as those introduced by the use of empirical potentials, are difficult to quantify theoretically and experiments provide a way to assess the reliability of simulations and improve the methodology [75–77]. Two are the main limitations which still affect atomistic simulations: the accessible scale (both in space and time) and the need for approximations. While the increase in raw computational power and the development of novel insightful techniques to accelerate rare events and to simulate increasingly larger systems are pushing these boundaries, a number of relatively substantial approximations are still required to boost computations of proteins, membrane patches and other complexes, where hundreds of thousands of atoms come into play.

Atomistic molecular dynamics, based on classical force fields, defines the middle ground between more accurate, but expensive, quantum mechanical calculations and

approximate, but efficient, coarse-grained and general models. Information from first principle calculations and/or empirical data are taken into account within the force fields in an attempt to reproduce physical properties and dynamics of materials and biological processes at an accessible computational cost. This family of techniques has proved successful in countless cases and has been chosen here to shed light on the activation mechanisms of pLGICs.

2.1 Molecular Dynamics

Molecular dynamics is a computational technique commonly used to study systems in materials, chemical and biological sciences. It was introduced in the late 1950's by Alder and Wainwright as a mean to investigate the interactions of simple many-body systems, such as hard spheres, but it took about twenty year for the scientific community to realise its potential. By then, its efficacy in evaluating the properties of liquids had became apparent and MD was finally applied to the study of a protein for the first time in 1977 [78–83].

The idea behind MD is simple: coordinates and velocities of particles which compose the system, may they represent single atoms or groups of atoms as in coarse grained simulations, are obtained as a function of time by integrating Newton's equations of motion in discrete steps. These particles are treated as classical objects, an approximation that limits the information that can be extracted form the system, but also decreases considerably the cost of the simulations. MD allows to explore the phase space, monitor ensemble averages and the dynamical behaviour of relatively large systems, composed of hundreds of thousand of bodies, such as proteins or lipid membranes, on timescales that are simply not accessible to more accurate *ab initio* methods, such as density functional theory (DFT) [84, 85], which instead work at the quantum level [86, 87].

Although a certain degree or randomness may be introduced, e.g. activating a stochastic thermostat, MD simulations are deterministic and complementary to Monte Carlo (MC). The latter probe the configurational space with trial stochastic moves of its particles; the acceptance or rejection of each move is then regulated through different algorithms. The most used of these is the Metropolis algorithm, which relates the acceptance ratio to the energy differences between starting and final configurations, according to a Boltzmann statistic [88, 89]. Both the MD and MC methods are aimed at generating an ensemble of representative configurations with certain thermodynamical properties; from the microscopic information contained in these ensembles, macroscopic observables, such as pressure or free energy, and probabilities can be extracted. From MD trajectories, it is straightforward to compute temporal averages of thermodynamical observables and, thanks to the ergodicity principle, their ensemble averages as well. This principle, although never formally

proven (and sometimes not applicable) for MD trajectories, states that in the limit of infinite simulation time, the time average of a given observable over a trajectory is identical to its ensemble average, such as that obtained with MC. What makes conventional MC attractive is that only the potential energy is needed to generate these configurations and no forces need to be evaluated, making the algorithm much more efficient when compared to MD. On the other hand, the exploration of the configurational space is inefficient with MC methods, especially when high energy barriers are present as in the case of organic macromolecules. In addition, MC cannot provide information on the time evolution of the system [87, 90].

The Molecular Dynamics Algorithm

Molecular mechanics (MM) models, such us those implemented in MD simulations, typically describe the system in terms of simple potential functions that govern the behaviour of spherical atoms connected by springs. This simplified scheme allows to quickly integrate Newton's equations of motion for point-like particles, whose internal interaction forces are modelled on basic pairwise functions. This assumption is justified in terms of the Born-Oppenheimer approximation [91], which allows to decouple electronic and nuclear motions, the latter being much slower. In this framework, MD trajectory integrators work only with nuclear coordinates, incorporating the electronic behaviour in the simple interaction potential description. The equation of motion for a system of N particles is

$$m_i \ddot{\mathbf{r}}_i = -\frac{\partial}{\partial \mathbf{r}_i} U(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N), \tag{2.1}$$

where m_i is the mass of the i-th particle and $\mathbf{r}_i$ its position.

An analytic solution of Eq.2.1 for every particle at all times is simply not feasible, as the forces vary with any geometrical change. In order to observe the system evolving dynamically over a finite length of time, these equations have thus to be integrated numerically and over discrete time steps δt. The length of such steps has to be carefully chosen, since the stability of the simulation is dependent on this choice. In order to speed up the calculations and save computer time, one would ideally choose a large time step, but this would probably lead to the loss of information on those motions acting on timescales smaller than the step itself. Ignoring these motions may introduce errors in the observed thermodynamics, structural or dynamical properties of the system under investigation [87, 92, 93]. The highest frequency motion in classical simulations of biological molecules pertains bonds containing hydrogen atoms, which require a time step no larger than 1 fs in order to be properly sampled [94]. A number of algorithms, such as SHAKE [95] or RATTLE [96], which are implemented in the AMBER [97] and NAMD [98] MD

packages respectively, have been thus developed in order to keep these bonds fixed during the simulations through Lagrange multipliers added to the equation of motion, allowing for an increase of the time step up to 2 fs [90].

Both SHAKE and RATTLE are based on the Velocity Verlet algorithm, a widely used integration method. It is a variation of the Verlet algorithm, which separates the calculation in sub-steps, where the velocity is explicitly considered and calculated together with the coordinates at the same point of the time variable. The assumption is that, when sub-dividing the time in small finite intervals, the force can be considered constant within each step [99].
In Eq. 2.2 the basic form of the algorithm is shown

$$\mathbf{p}_i(t + \frac{1}{2}\delta t) = \mathbf{p}_i(t) + \frac{1}{2}\delta t \mathbf{f}_i(t) \tag{2.2a}$$

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \delta t \mathbf{p}_i(t + \frac{1}{2}\delta t)/m_i \tag{2.2b}$$

$$\mathbf{p}_i(t + \delta t) = \mathbf{p}_i(t + \frac{1}{2}\delta t) + \frac{1}{2}\delta t \mathbf{f}_i(t + \delta t) \tag{2.2c}$$

Momenta $\mathbf{p}_i$ and forces $\mathbf{f}_i$ at a time t are used to evaluate the momenta half a time step δt later. These are then used to evaluate the coordinates $\mathbf{r}_i$ at the next time step $t + \delta t$. The positions allow, in turn, to calculate the forces at the new time value and, from there, it is possible to calculate the momenta at $t + \delta t$. Repeating these operations at each time step, the dynamics of the system is reproduced [86, 87, 92]. This implementation solves one of the major drawbacks of the original Verlet algorithm, in which the velocities are not calculated explicitly by the integrator at each steps, but are obtained instead through a mean value approach of the kind $\mathbf{v}_i(t + \delta t) = \frac{\mathbf{x}_i(t+\delta t) - \mathbf{x}_i(t+\delta t)}{2\delta t} + \mathcal{O}(\delta t^2)$. Calculating certain physical quantities dependent on the velocity, such as kinetic energy and instantaneous temperature, can become challenging with this approach: beside the errors introduced by the use of an approximate equation, the value of such quantities at a specific timestep cannot be calculated for the system until the position at the next timestep are known.
The success of Velocity Verlet is also due to a number of properties which are crucial for a correct representation of a physical system: it respects in fact the time reversibility of the Newton's equations and the conservation of energy, which, although trivial, is still problematic for many algorithms, and is symplectic (namely it preserves the infinitesimal volume in the phase space $d\mathbf{r}d\mathbf{p}$ during the time trajectory, a property, also known as Liouville's theorem, which as a consequence gives global stability to coordinates and energy) [90]. Moreover, a good algorithm should be both quick and stable, possibly avoiding to calculate forces too often and allowing longer time step choices, usually done by storing high order derivatives of positions and velocities,

while avoiding high memory usage.
In spite of being increasingly accurate and efficient, it is still unrealistic to expect numerical methods to strictly follow the true trajectory of the system for very long times, mainly because of the limits inherent to finite time step integration, and the consequent accumulation of numerical errors. As an example, the Verlet algorithm is a fourth-order integration method, meaning that the terms in the Taylor expansion of the coordinates $r(t + \Delta t)$ beyond Δt^4 are truncated. As small as these errors may seem, on a long trajectory they can lead to serious artifacts, such as the loss of conservation of total energy or total momentum. However, when adequate tools and methods that do not deviate significantly from the path of the analytic trajectory are implemented, MD simulations allow to obtain invaluable information on the studied system [86, 87, 92].

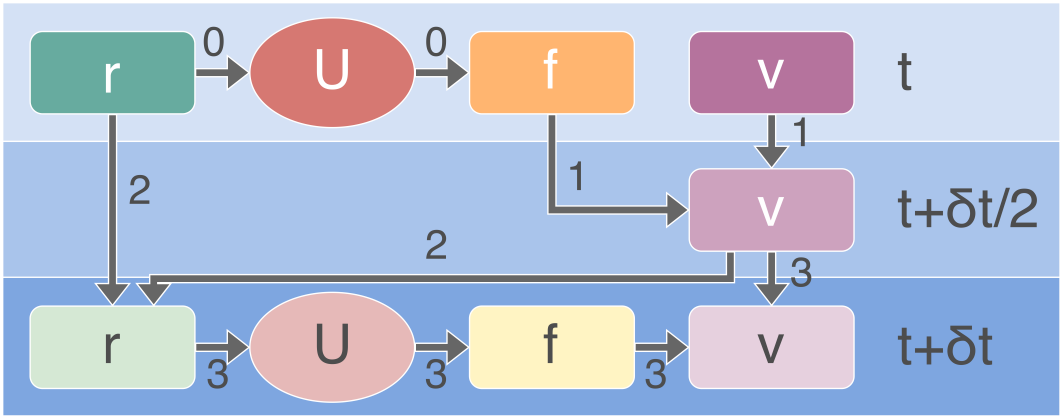


Figure 2.1: Flow diagram showing the sequence of steps executed by the Velocity Verlet integrator: each step is numbered sequentially. 0) Forces (f) are derived from the potential (U) dependent on the positions (r). 1) forces and velocities (v) at timestep t are used to calculate the velocities half a timestep later. 2) These are then used to obtain the positions at the next timestep. 3) The new forces can be evaluated and, together with the half timestep velocities, they can be used to obtain the new velocities.

The Force Field

The potential energy that defines the interactions among the particles of the system, $U(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$ in Eq. 2.1, can be expressed in different forms: in order to ease the amount of computational time and the power required when working with large systems, simplified schemes are commonly preferred. These functional forms of the potential are usually purely addictive and take into account only pairwise interactions between atoms; neglecting higher order terms improves the scalability up to $\mathcal{O}(N^2)$, where N is the number of particles in the system, at the cost of a reasonable approximation. Range and strength of interactions between different atom types can be obtained experimentally or by quantum mechanical calculations and are then introduced in the algorithm through a set of parameters. The functional

form and the chosen set of parameters are known as empirical force field [86, 87]. In the most common form of force field, known as Class I and found in all major MD engines including AMBER and NAMD, bond length, angles and torsions are considered independent and the magnitude of atomic charges is fixed. At variance, Class II force fields incorporate cross or higher order terms to fit vibrational spectra accurately, but their use is not widespread. Here, long- and short-range interactions are accounted for in two separate terms [92, 93].

$$U_{forcefield} = U_{bonded} + U_{non-bonded} \tag{2.3}$$

Non-bonded interaction are applied to all atoms but those connected through up to three covalent bonds. They are modelled as a Coulombic term plus a Lennard-Jones potential, the latter (Eq. 2.4) being a widely used potential which includes both the excluded volume repulsion, represented by the r^{-12} term, and the van der Waals attractive component, that goes instead as r^{-6}, where r is here used in place of the interparticle distance $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$.

$$v^{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{2.4}$$

The Lennard-Jones potential can be tuned to reproduce the interaction of specific atomic species through two parameters: the energy ϵ, which represents the well depth, and σ, which defines the length scale. Figure 2.2 shows the analytic behaviour of this potential as a function of the characteristic length.

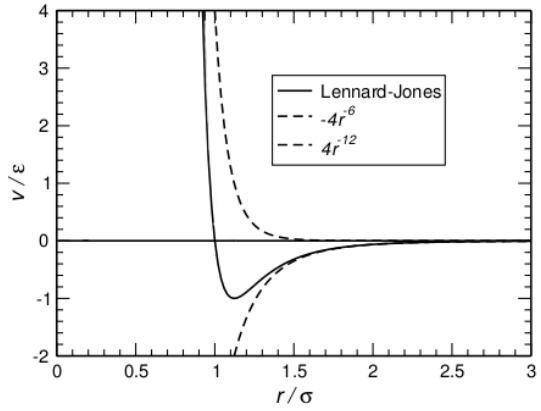


Figure 2.2: Lennard-Jones pairwise potential, with the typical r^{-12} and r^{-6} contributions. Adapted from ref. [86].

To further speed up the calculations, the Lennard-Jones potential function is usually truncated at a fixed cut-off value. This choice introduces discontinuity in forces and energy and consequently large errors. To avoid these discontinuities, a switching function is commonly used to smoothly deactivate the potential after the

cut-off radius with the aid of a cubic spline.
A Coulomb potential is added to model the behaviour of charged particles

$$v^{Coulomb}(r) = \frac{q_1 q_2}{4\pi\epsilon_0 r} \tag{2.5}$$

with q_1 and q_2 as the charges of the two-body interaction and ϵ_0 the permittivity of free space [86]. The Lennard-Jones and the Coulomb potential are spherically symmetric, i.e. they depend on the modulus of the distance between pairs of atoms $|\mathbf{r}_{ij}|$. Their total contribution is then obtained by summing over all these pairs.
It is necessary to comment here on the fact that the Coulomb potential represents the most problematic term in the force field calculations. It is a long-range interaction and decays only as $\frac{1}{r}$, making its computation highly expensive. To lessen the costs, it is possible to introduce a cutoff distance around each atom and naively take into account only the interactions that fall within this arbitrary radius. This approximation is, however, too drastic in most cases, especially when considering the importance that electrostatic interactions have in biological systems, and can introduce energetic discontinuities that lead to unstable simulations. The Ewald summation is, in contrast, an high accuracy method: the Coulomb term is decomposed into a short-range component, treated in the real space, and a long-range one, which is summed in the Fourier space. The idea is to exploit the rapid convergence of long range sums in the Fourier space, when compared to their real space convergence. This method is extremely efficient and is now commonly used in its Particle Mesh Ewald (PME) implementation, in calculations involving periodic boundary conditions [87, 92].

The basic scheme for intramolecular interactions (namely interactions between different parts of the same molecules, i.e. the $U_{non-bonded}$ term in Eq. 2.3) is characterised instead by harmonic potentials that generate energetic penalties when the system deviates from reference positions of distances or angles, obtained by *ab initio* calculations or measured experimentally. More sophisticated force fields have also been developed, which introduce additional terms of increasing detail and cost, such as the Urey-Bradley non-bonded cross term and the improper dihedral angles, but the basic structure remains the same with three main terms accounting for stretching bond distances, bending angles and torsional angles [87, 93].

$$
\begin{aligned}
U_{bonded} &= \frac{1}{2}\sum_{bonds} k_{ij}^{\mathbf{r}}(\mathbf{r}_{ij}-\mathbf{r}_{eq})^2 \\
&+ \frac{1}{2}\sum_{bending} k_{ijk}^{\theta}(\theta_{ij}-\theta_{eq})^2 \\
&+ \frac{1}{2}\sum_{torsional}\sum_{m} k_{ijkl}^{\phi}(1+cos(m\phi_{ijkl}-\gamma))
\end{aligned}
\tag{2.6}
$$

All three terms in Eq. 2.7 share the same harmonic form, featuring a strength parameter k and an equilibrium value.
In the first component, the bond distance between adjacent pairs of atoms $\mathbf{r}_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the monitored variable, while the bending angle θ_{ijk} between successive bond vectors $\mathbf{r}_{ij}$ and $\mathbf{r}_{jk}$ is followed in the second term. The torsional angle ϕ_{ijkl} is the angle between the two planes defined by three connected bonds, $\mathbf{r}_{ij}$, $\mathbf{r}_{jk}$ and $\mathbf{r}_{kl}$. Its part of the potential involves an expansion in series of the cosine, with m minima as the bond completes a rotation of 360°; γ is in this case the equilibrium value [86, 87, 93]. These three components can be visualised in Figure 2.3.

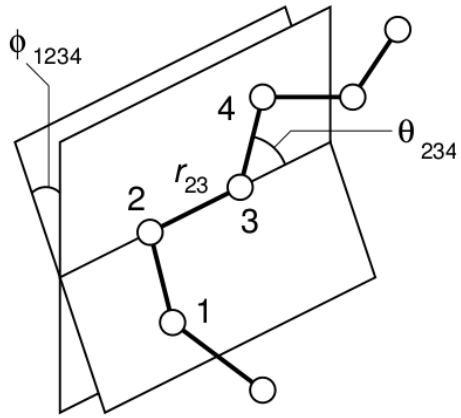


Figure 2.3: Geometry of a simple chain molecule, illustrating the definition of interatomic distance r_{23}, bending angle θ_{234} and torsional angle ϕ_{1234}. (From ref [86]).

Charge, mass, or more in general the atomic species of the atoms involved, together with those of nearby atoms, influence these potential through the set of parameters described above. Differences between the available force fields mainly arise from the different approaches to fit these parameters. Moreover, extra components optimised for different MD software packages or to reproduce specific conditions, like crystalline materials versus more flexible biomolecules, may be found in these parameters, diminishing their transferability. A constant work of refinement is being carried out by the developers of these force fields, making them more and more accurate and transferable at each distribution. A general consensus among the force

fields is commonly sought after, together with a systematic validation by experiments. However, it is important to keep in mind that state-of-the-art MD codes still rely on approximations, and a careful choice, weighting strengths and limitations of each force field, should be a priority when designing computational simulations.

The choice of using a simplified force field description is not always appropriate. A lot of interesting features, such as anharmonic motions, polarizability, quantum effects, or bond breaking and forming, are in fact lost in exchange for a drastic reduction of the computational cost. In cases where these elements are crucial for the correct reproduction of biological phenomena and cannot be simply ignored (e.g. when transition metals are involved in binding), first principles simulations are preferred [100]. If, however, the system is too complex and *ab initio* methods are still too expensive to be considered alone, hybrid techniques come into play.
A wide variety of mixed techniques have been developed through the years. As an example, in Quantum Mechanics/Molecular Mechanics (QM/MM) simulations, a limited area surrounding the region of interest is modelled within first principles methods, typically with DFT [82, 101], while the rest of the system is described at the classical atomistic level. This method is well suited to reproduce chemical reactions, charge or proton transfer and other quantum mechanical events within larger biological structures, but again, in spite of the growing popularity, the accessible time scales are reduced when compared to a fully classical MD, limiting their applicability.
Moreover, QM/MM is not suitable when the interest lays in properties of the entire system (e.g. macroscopic changes in conformation), rather than focusing on a limited active site: in these cases, a better strategy is to modify the MD algorithm, adding extra features or higher order terms to the calculation of the forces. This is the case of polarizable force fields, where the static point partial charges model is abandoned and the change in charge distribution of molecules in response to their environment is calculated on-the-fly or induced with dipole models [87, 102]. Similarly, different groups have been working on the development of reactive force fields, that allow for the breaking and forming of chemical bonds [103, 104]. In spite of the remarkable efforts put in the development of these models, the problem of finding appropriate approximations, together with issues related to their practical implementation, still hinder their progress. The lack of transferability, a problem which affects force fields in general, prevents their widespread use in the scientific community and restricts them to niche problems. The choice of the most adequate model for the problem of interest should not be underestimated: adding detail comes at the price of increasing computational requirements, so there is an inevitable trade-off between accuracy and efficiency which affects computational simulations.
It is worth mentioning here also the role played by force fields for lipid bilayers,

extensively used in the work presented in this thesis. As in the case of proteins, these parametrisations rely on QM mechanical calculations trying to reproduce properties, such as the are per lipid, observed by NMR, X-ray and neutron diffraction experiments. The amphiphilic structure of lipids (i.e. they are formed by an hydrophilic head and an hydrophobic tail) modulates their temperature and solvent sensitivity, and their characteristic phase as liquid crystals at room temperature. This, together with the wide variety of heads and tail species found in nature, gives the membranes a remarkable complexity and makes their modelling particularly challenging. A number of accurate force fields are nowadays available for both coarse-grained and all-atom simulations, but the choice of the most adequate parametrisation has to be carried out with care. Not only direct lipid-protein interactions, but also structural properties of the membrane itself, such as the surface tension or the bilayer thickness, can influence the behaviour of proteins and the accuracy of their description should be carefully taken into account [105–107].

Reproducing the environment

In MD simulations, the aim is to accurately reproduce the natural evolution of classical systems under different physical conditions: the total energy E is commonly a conserved quantity, and if the volume V is fixed and so is the number of particles N, the time average of physical quantities along the simulation are equivalent to ensemble averages of a microcanonical NVE state. The NVE ensemble, however, corresponds to a close, isolated state; when validation of the MD results through experiments is demanded, the simulated environment is expected to properly reproduce the conditions in the laboratory by means of a more adequate statistical ensemble. [87, 92].

Experiments are usually performed in the canonical ensemble, NPT: under these regime, the system is kept in thermal equilibrium with a heat bath. As in the case of the microcanonical ensemble, number of particles N and volume V are also constant. While N and V are both extensive quantities, temperature is intensive, but it can be calculated from an extensive counterpart, the kinetic energy [87, 93].

$$\frac{< K >_{NVT}}{n} = \frac{3}{2} k_B T \tag{2.7}$$

Equation 2.7 takes into account the time average of the kinetic energy K for an unconstrained system. Here k_B is the Boltzmann constant and n the number of degrees of freedom.

To simulate the effect of the heat bath and control the temperature, thermostats are added to the simulations. Different implementations exist, the basic one being just a periodic rescaling of the velocities following a Boltzmann distribution around the desired temperature [108]. Although extremely efficient, such algorithms deviate from

the correct NVT trajectories, especially for small systems, although the calculated physical quantities are roughly correct. The stochastic thermostats are the most commonly used subset of methods: the Andersen thermostat [109] belongs to this group and imposes collisions with virtual particles. At variance, the Nosé-Hoover thermostat is deterministic and introduces the energy fluctuations extending directly the Lagrangian [110, 111].
In this project, the Langevin thermostat was chosen for its ability to reproduce correctly the canonical ensemble while maintaining a stable dynamics. The assumption behind it is that the simulated atoms are embedded in a bath of virtual particles, which, as in the case of the Anderson thermostat, influence the solute with stochastic collisions. This effect is accounted for by the addition of a friction drag term (with γ as the friction constant) and a Gaussian random force $\mathbf{R}(t)$ in the equations of motion

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial U}{\partial \mathbf{r}_i} - \gamma \mathbf{p}_i + \mathbf{R}(t). \tag{2.8}$$

Equation 2.12 is known as Langevin stochastic differential equation and describes the motion of a particle affected by the thermal agitation of the heath bath [112].
The drawback of these stochastic methods is that the random component makes impossible the calculation of the real diffusion coefficient and other dynamical properties, mainly due to the introduction of an artificial transfer of momentum to the system.

The NVT ensemble is usually chosen when working with biological phenomena, as it correctly emulates the environment found within a living organism. However, when trying to reproduce chemical or biological reactions open to the atmosphere, as in the case of an experimental laboratory, an isothermal-isobaric ensemble (NPT) is the most appropriate choice. In place of energy, the new conserved quantity is enthalpy $H = U + pV$ and as a consequence of this relation, volume has to be allowed to vary while pressure is fixed.
In atomistic simulations, pressure can be calculated through what is known as the virial equation, whose terms are all easily accessible:

$$P = \frac{k_B T N}{V} + \frac{1}{Vd}\left\langle \sum_{i<j} \mathbf{f}_{ij} \cdot \mathbf{r}_{ij} \right\rangle. \tag{2.9}$$

Here, d is the dimensionality of the system and $\mathbf{f}_{ij}$ is the internal force (due to the force field) exerted on particle i by particle j, while $\mathbf{r}_{ij}$ is the vector joining their positions. The function is composed of an ideal gas term and of an interaction term, the ensemble average, here denoted by the brackets, over forces and distances.
As in the case of thermostats, different approaches have been developed, but the

simplest one achieves a periodic change in volume by scaling isotropically the particles' positions. In the case of the more sophisticated Berendsen barostat, the system is coupled with a pressure reservoir with pressure P_{bath} and coupling constant τ_p as in Eq. 2.10 [108].

$$\dot{V} = \frac{\kappa}{\tau_P}(P(t) - P_{bath})V \tag{2.10}$$

Here, κ is the isothermal compressibility

$$\kappa = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \tag{2.11}$$

On the same line of its thermostat counterpart, the Nosé-Hoover barostat adds an external piston as additional degree of freedom to the Lagrangian [113]. This concept has been then extended by the Parrinello-Rahman barostat, in which each unit vector of the unit-cell is made independent, allowing for dynamic changes in the shape of the simulation box [114, 115]. These approaches are more accurate then the Berendsen's one, but also slower to compute.
A modified version of the Nosé-Hoover barostat, known as Langevin piston, where the control of the pressure piston fluctuations are implemented using a Langevin dynamics [116], has been used for the most part of this project. In this framework, the Langevin equation is rewritten as

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} + \frac{1}{3}\frac{\dot{V}}{V}\mathbf{r}_i, \tag{2.12a}$$

$$\dot{\mathbf{p}}_i = \mathbf{f}_i - \frac{1}{3}\frac{\dot{V}}{V}\mathbf{p}_i, \tag{2.12b}$$

$$\ddot{V} = \frac{1}{W}(P(t) - P_{bath}) - \gamma\dot{V} + R(t). \tag{2.12c}$$

W is the fictitious "mass" of the piston, of dimensions $(mass \times length^{-4})$, γ is a collision frequency and $R(t)$ is, as in the case of the thermostat, a random force taken from a Gaussian distribution. As before, the latter has zero mean, but variance

$$\langle R(0)R(t)\rangle = \frac{2\gamma k_b T\delta(t)}{W}. \tag{2.13}$$

When this implementation was not available in the chosen MD engine, the Berendsen barostat was used instead.

Boundary Conditions

As already mentioned, the cost of computing the forces within a classical force field framework grows as $\sim \mathcal{O}(N^2)$, An easy way to extend the size of the simulated system,

without directly increasing the number of molecules and thus slowing excessively the computation, is to apply periodic boundary conditions (PBC).
Typically, PBC are imposed to reproduce the properties of fluid bulk and build a proper solvent environment for proteins and other solutes. The simulation volume is thus surrounded by replicas of itself: when an atom leaves the simulation box, an incoming image is created on the opposite side as in Figure 2.3. As shown schematically in Fig. 2.4, every atom is replicated in the nearby boxes and can interact with the nearby particles or their replica images, depending on which ones fall within the interacting potential cutoff. The minimum image convention guarantees that these interactions are not duplicated between different copies of atoms. This technique allows to simulate a more realistic environment and avoid surface effects without requiring excessive resources [86, 87, 90].

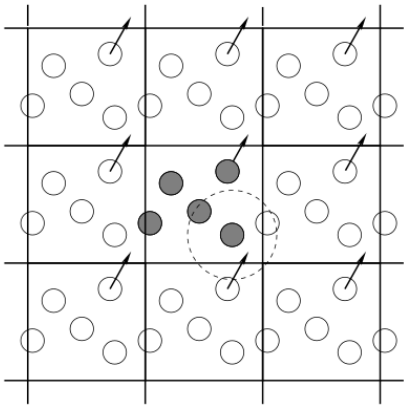


Figure 2.4: Schematic representation of periodic boundary conditions. As a particle moves out of the simulation box, an image particle moves in to replace it. In calculating particle interactions within the cutoff range, both real and image neighbours are included.

Despite the introduction of a cutoff for the long-range pairwise interactions, the evaluation of the distance of all pair of atoms can be still time consuming. A simple strategy to bypass this issue is to introduce a second cutoff that prevents repetitive calculations of the interatomic distances on the entire replica structure. For each particle, a list of neighbour atoms, known as Verlet list, is constructed and then periodically updated. This list contains the names of all the objects falling within the cutoff distance from the reference atom, on which the interactions will be calculated. The list is updated automatically when the distance travelled by its atoms, from the time of the last update, overcomes a defined value. The computational cost of this method varies with the cutoff choice: larger cutoff distances means larger lists that can be reconstructed less frequently. A compromise has to be found between the length of the lists and the cost of their update.
As an alternative, it is sometimes more efficient to simply dissect the space in three

dimensional cells with characteristic length larger than the cutoff distance and and build a separate list of atoms for each of these boxes. The search for interacting neighbours is then much quicker, as it requires to look only at atoms inside the same cell or in nearest neighbour cells. The spatial decomposition is also useful to distribute the system on multiple processors while reducing the need of communication among them, as only neighbour cells need to share information [117]; this method is now the standard choice in MD packages such as those used in this project.
Different box geometries can be used, including an orthorhombic box and a truncated octahedron, to optimize the number of solvent molecules required. If Coulomb interactions are involved, the total neutrality of the system with PBC must be imposed, to avoid the divergence of the total charge due to its endless replication in the infinite replica.

It is important to comment on the fact that the imposed artificial periodicity can influence quantities dependent from long-range correlation properties. As in the case of thermostats and barostats, PBC parameters have to be carefully chosen to avoid introducing unphysical effects and contaminate the simulations. As an example, when working with solvated systems in bulk, the water buffer has to be large enough to surround adequately the protein and isolate it from direct interactions with its periodic images [86, 90].

Solvation

Until a decade ago, a fully inclusive representation of a protein's physiological environment, with explicit solvent molecules and ions was outside the capabilities of commonly available computational resources. To speed up the simulations and avoid calculating solvent-solvent interactions, implicit solvent models, based on classical electrostatic theories such as the Generalized Born (GB) and Poisson-Boltzmann (PB) schemes, have been often used. Salinity effects can also be modelled implicitly. These models perform reasonably well and are computationally much cheaper than explicit solvents; however, they fall short when trying to reproduce the dielectric screening effects and specific solvent-protein interactions, often crucial in mediating conformational changes or biological functions. This is especially important in free energy calculations where the location of water molecules may play an important role (e.g. in the case of water mediated protein-ligand interactions) [87, 118, 119]. There is a wide variety of water models available, each one with features and approximations adapted to the reproduction of specific properties of water: as an example, TIP3P and TIP4P are preferred in MD simulations. While in the former the molecules are kept completely rigid, consistently with SHAKE, the latter adds off center point charges, increasing the accuracy of electrostatics and their cost [120] (see Fig. 2.6). Thermal properties such as the enthalpy of vaporisation and more in general the

temperature dependence of water density are not described very well, but these models are able to recover most of the solvent behaviour, including entropic effects and hydrophobicity. With the advent of parallel computing and the increase in power of modern processors, and given the accuracy and detail they can provide in contrast to an implicit solvent description, the benefits surpassed the cost and explicit solvents have now become the most common choice.

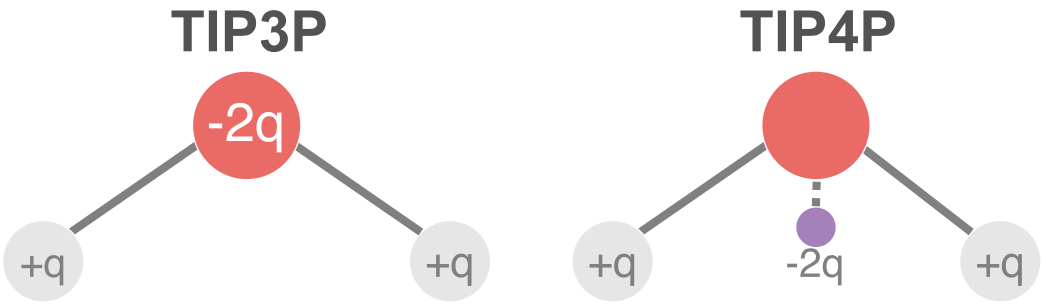


Figure 2.5: Schematic representation of the TIP3P and TIP4P water models, where oxygens are shown in red and hydrogens in grey. The main difference is the displacement of the negative point charge (-2q, in purple) with respect to the oxygen in TIP4P.

Energy minimisation

The crystallisation process required to extract X-ray information of the structures of proteins may distort the positions of the constituent atoms. The value of bond lengths and angles may be far from their natural relaxed state and steric clashes and overlaps between different particles may also be present, e.g. due to the subsequent addition of hydrogen atoms. Unrefined molecular structures need thus to be subject to a thorough energy minimisation, before being used in MD simulations. Optimising a nonlinear function like the potential is an extremely complex problem: many different minimisation methods exist, but although efficient in refining these structures and removing gross inconsistencies, they are inadequate for fully sampling the conformational space as they can only reach the closest local minimum. The most popular methods, such as steepest descend and conjugate gradient, make use of first-order derivatives, while alternative techniques, such as the Newton-Raphson method, require the calculation of second-order derivatives [87, 90, 121]. The simplest of these methods, steepest descent, exploits the gradient of the potential-energy surface to guide the search for the nearest energy minimum. An iterative relationship like $\mathbf{r}(k) = \mathbf{r}(k-1) + \lambda(k)\mathbf{F}(k)$ is applied, where k is the iteration step, $\mathbf{r}$ are the selected coordinates, λ is the length of the minimisation step and $\mathbf{F}$ is the so called force vector, containing the forces derived by the gradient of the potential energy and applied to the atoms at each iteration. λ is adjusted at each iteration to adequately follow the curvature of the potential surface and avoid overshooting the

local minimum [87].

Set-up protocol

Before starting a molecular dynamics study, a number of preparation steps have to be carried out on the system to avoid instability and errors during the simulation. This protocol is model dependent, nevertheless it is possible to outline the main points of an general procedure.

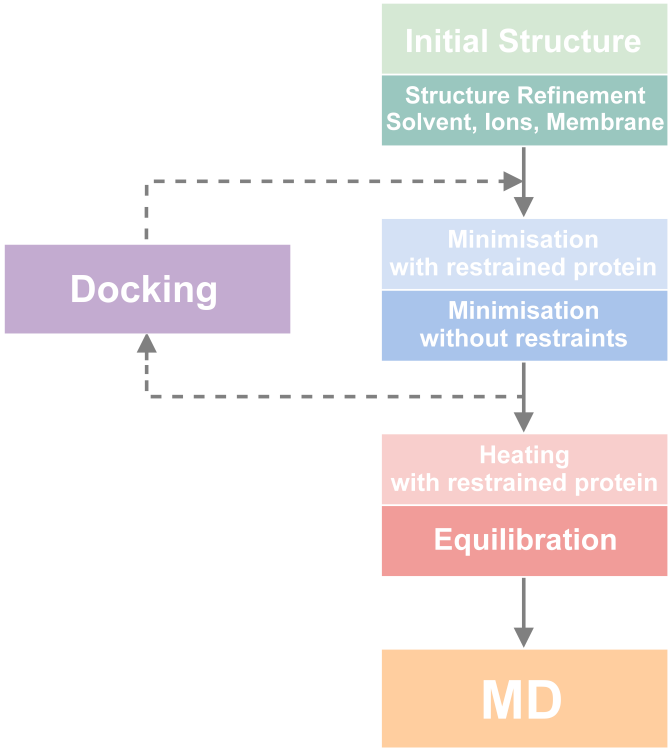


Figure 2.6: Flow diagram of the preparation steps of an MD simulation.

Once the coordinates of the protein of interest have been obtained, e.g. from a crystallographic structure or homology modelling, it is important to run a preliminary analysis to assess the reliability of the model and fix inconsistencies. A number of tools are available to check that the position of amino acids, their rotameric state or backbone angles, are satisfactory and to fix them where needed. This procedure should be carried out with a special attention for parts of the protein that were missing and were subsequently reconstructed, since an incorrect positioning of even a few side chains may be sufficient to make the system unstable.
If other compounds, such us ligands or lipids, are required, they can be added at this point. Solvent and ions, to reproduce the desired salinity and ensure the neutrality of the system, are also added and the periodic boundary cell is built around the protein. Particular care has to be given when building a system embedded in a lipid bilayer. Membrane proteins are characterised by areas with a high number of

hydrophobic residues, which should come in contact with the membrane tails. In the case of ion channels the transmembrane domain is characterised by a series of helices with polar residues at their ends, which are expected to be in contact with the charged heads of the phospholipids, and with hydrophobic amino acids in the middle, at the level of the aliphatic chains. This is particularly helpful when positioning the membrane during the model construction phase, which can be carried out through appropriate software or manually; the following equilibration stages will take care of adjusting minor errors in the relative position of protein and membrane. As an alternative, the membrane can be left to self-assembly around the TMD, starting from a mixture of solvent and lipids, although the timescales necessary to observe this kind of behaviour require the use of coarse-grained models [18]
Once the system has been built, a first minimisation to remove steric clashes and major structural errors is carried out. This is usually divided in stages: first solvent and ions are let relax around the constrained protein and membrane (if present), then the minimisation is repeated for all atoms.
When working with bound ligands, the docking procedure should be applied to the minimised structure and, once an adequate conformation has been identified, the preparation and minimisation steps should be repeated.
Before starting the MD production run, however, the system should be set to the room or physiological temperature (300 and 310 K respectively) by means of a gradual heating, where backbone or the C_α atoms of the protein are usually restrained. This prevents instabilities and clashes in the structure that could be otherwise observed by simply resetting the velocities according to a Boltzmann distribution around the target temperature. The restraints are then slowly decreased, leaving the tertiary structure of the protein time to adapt to the new temperature and to relax to its native conformation.
Finally, if using NVT for the actual study, it is good norm to run a preliminary short NPT simulation to let the periodic boundary box adjust its size and reach the target pressure, and then extract the average box size from this trajectory to be used in the next stages.
Details on the specific protocol applied to each system here studied can be found in their respective results sections in Chapter 3.

Enhanced sampling techniques

The more is known about proteins and other biological systems, the more is clear that their behaviour is strictly co-dependent on the environment. When working with complex substances, like in the case of protein embedded in lipid membranes or multi-protein structures, approximating or even ignoring environmental interactions may lead to inaccurate results. However, even the most efficient MD algorithm scales with the square of the number of particles and soon the weight of evaluating the

forces on all the added atoms becomes unmanageable. As much as limiting the size of the model, the computational cost also restricts the time scale accessible through MD simulations. Typical atomistic MD runs last few hundreds nanoseconds of real time, rarely extending over the microsecond regime. However, many interesting biological phenomena require much larger times to be observed: as an example, while major conformational rearrangements of secondary structures, such as those observed in the gating of ion channels, span timescales from micro- to milliseconds, ligand-protein binding events can require times well over the milliseconds.
Recent developments in parallel computing and the building of high performance computing (HPC) platforms dedicated to material and chemical calculations opened new possibilities in the field. Most of the major MD codes are now fully compatible with the message passing interface (MPI) making them suitable for multi-core processors, and can now run even on massively parallelizable graphic processing units (GPUs). The tendency has been thus to straightforwardly increase the size and length of the MD simulated systems, pushing toward the macroscopic level and hoping that the size and timescale limits will sort themselves out with time. However, a part from a few exceptions where huge resources are available, a brute-force approach may still not be the smartest choice or may be simply not enough.

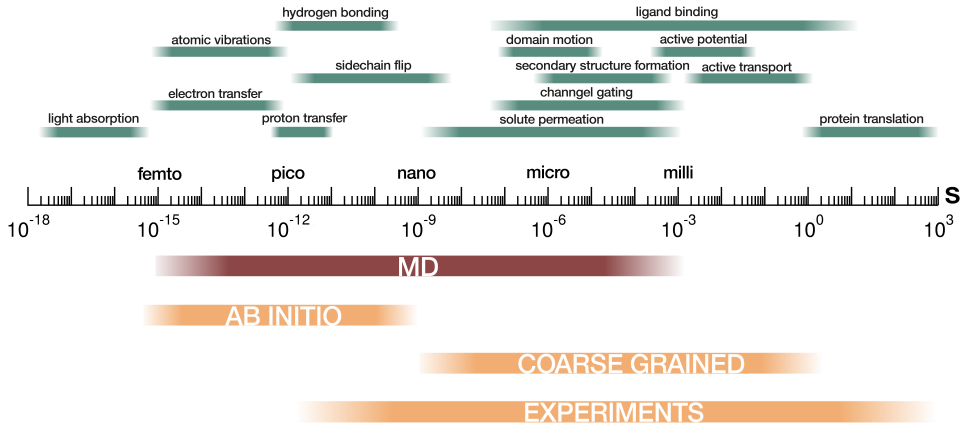


Figure 2.7: Comparison between the timescale of the typical processes of interest in biochemistry and that accessible to different scientific methods (adapted from ref. [122]).

Various groups have focused on the development of ingenious techniques to bypass the size and time scale problems. Coarse-grained models, where sets of atoms are grouped under single objects with appropriate parameters and force field descriptions, are an example of this. Their aim is to allow for larger systems to be studied by reducing the number of particles and interactions that the MD engine has to handle. They are commonly used to study lipidic structures and self assembly

properties [123–125]. In parallel, several enhanced sampling techniques have been developed to accelerate the dynamics of the system, improving the exploration of the free energy landscape and helping access the so called rare events [87, 93, 126]. Even the simplest biological system is characterised by complex free energy maps, often separated in high-probability basins connected by a network of low-probability transition regions. Because of the corrugated shape of these landscapes, ergodicity is rarely attainable with plain MD simulations. Enhanced sampling techniques exploit different approaches to ensure that the relevant regions of the configurational space are visited, while paying attention that the samples generated respect the underlying statistical probability: two common strategies are based either on tempering or on biasing the potential [126, 127].

The first class of methods is based on artificially increasing the temperature of the system, relying on the fact that the rate of barrier crossing events is temperature dependent. In simulated annealing (SA) [128], the simulation is initially run at a temperature sufficiently high to overcome high free energy barriers and is slowly decreased until the system reaches a local minimum. It is generally run as a Markov-chain Monte Carlo rather than an MD and its efficiency is strongly dependent on the cooling speed parameter. SA is often used as an optimisation method to escape from an initial metastable configuration an hopefully reach a more physiologically relevant structure to be used in MD simulations. As an alternative, parallel tempering is a replica exchange (RE or sometimes REMD) method [129, 130], where several replicas of the MD simulation are run at the same time. The different simulations have different temperatures which are then periodically swapped between two random replicas with Metropolis style acceptance probability based on potential energies and temperatures. Starting from a system at standard room or physiological temperature, the subsequent replicas are simulated with increasing temperatures, making the potential barriers easier to overcome and alternative conformations more accessible. This technique is very versatile; in fact the property exchanged among the replicas is not necessarily restricted to be the temperature and different parameters, including modified forms of the Hamiltonian as in the case of generalized RE [131], can be exchanged.

The main criticism of tempering methods is that MM force fields are not designed to work outside physiological temperatures and may thus induce unphysical behaviour and artifacts when the temperature is increased significantly. Also they are based on the assumption that ergodicity can be reached in the modified ensemble, which may not be the case in the presence of entropic barriers, whose contribution would be enhanced by the increase in temperature [126, 127]. Moreover, running multiple replicas of the same simulation increases dramatically the computational effort required, making this techniques suitable only for systems of limited size.

Another family of techniques focuses instead on the potential energy itself, through

the calculation of mean forces (the gradient of the free energy as measured at specific locations) or with the addition of an external potential. Thermodynamical integration (TI) [132] is an approach that allows recovery of the free energy from the integration of these local mean forces, and is often used in conjunction with methods that introduce external forces to drift the system toward desired conformations, such as targeted MD [133] and steered MD [134]. Belonging to this group is also the adaptive biasing force (ABF) algorithm [135], which tries to achieve a uniform sampling by adding external forces built on-the-fly to counterbalance the mean forces felt locally by the system along an MD trajectory.
Methods that enhance sampling by adding bias potential are also known as non-Boltzmann sampling methods, and umbrella sampling [136] is the best known example. In this framework, a number of neighbour and overlapping windows are defined by the introduction of what is usually a simple harmonic spring potential. These windows span sections of the free energy space and force the system to explore both hard-to-sample and low-energy configurations alike in that localised region. Information from each window can then be composed to reconstruct unbiased probabilities and free energies. Along with umbrella sampling, metadynamics [137] is another non-Boltzmann sampling method that have been more recently developed. The elegant idea behind it and its efficient implementation in the most common MD codes makes it the ideal candidate to study rare events in biomolecules; for this reason it was chosen for this project to enhance the conformational sampling of portions of pLGICs during binding and gating events.

2.2 Metadynamics

Metadynamics is a powerful and reliable enhanced sampling method; it has proven successful for the study of many systems, from materials science to biology [138–142]. As most non-Boltzmann sampling methods it relies on a dimensionality reduction of the conformational space: instead of working with all the Cartesian coordinates $\mathbf{r}(t)$ along the entire trajectory, the bias is applied only to a limited set of M so-called collective variables $\mathbf{s}(t)$ (CVs), or reaction coordinates. CVs are multidimensional functions of the atomic coordinates $\mathbf{r}$ and allow to map the $3N$ dimensional free energy space to a reduced dimensionality CV space $\mathbf{s}$, with dimensionality $M << 3N$. The free energy in this new space can be easily recovered from the probability of observing the system in $\mathbf{s}$, the weight of all configurations $\mathbf{r}$ that map onto $\mathbf{s}$, which is

$$P(\mathbf{s}) = \langle \delta[\mathbf{s}(\mathbf{r}) - \mathbf{s}] \rangle \tag{2.14}$$

The angle brackets represent the Boltzmann ensemble average, which for the

system described by a set of coordinates $\mathbf{r}(t)$ and subject to a potential $U(\mathbf{r}, t)$, leads to a free energy of the form

$$F(\mathbf{s}) = -\frac{1}{\beta} \log \left(\int \exp(-\beta U(\mathbf{r}, t)) \delta(\mathbf{s} - \mathbf{s}(\mathbf{r})) d\mathbf{r} \right) \tag{2.15}$$

where $\beta = \frac{1}{k_B T}$, k_B is the Boltzmann constant as before, and T is the temperature. In order to improve the exploration of the CVs space, metadynamics uses a simple end elegant approach: a history-dependent potential is built on-the-fly by accumulating Gaussian repulsive terms deposited along the trajectory followed by the CVs. This bias potential hinders the exploration of places that were already visited, and iteratively fills the underlying free energy basins, allowing the system to escape from any energy well and overcome the high energy barriers through saddle points. During plain MD simulations, the system would be normally trapped in the starting potential minimum, but the added bias allows it instead to visit other stable and metastable states, otherwise unknown. In this manner, the transitions to different states, previously defined as "rare events", are accelerated with efficiency and great flexibility.
Equation 2.16 illustrates the bias potential as a function of $\mathbf{s}(\mathbf{r}, t)$, the value of the collective variable at the position $\mathbf{r}$ at the time t and $\mathbf{s}(\mathbf{r}, t')$ the values at past times $t' < t$.

$$U_G(\mathbf{s}, t) = w \sum_{t'=\tau, 2\tau \ldots < t} exp \left(-\frac{(\mathbf{s}(\mathbf{r}, t) - \mathbf{s}(\mathbf{r}, t'))^2}{2\sigma^2} \right) \tag{2.16}$$

Here, w is the Gaussian height, σ is the Gaussian width and τ is the time step of deposition, which defines how often a Gaussian is added.
Once convergence is reached, the biased potential provides an unbiased estimate of the underlying free energy in the limit

$$\lim_{t \to \infty} U_G(\mathbf{s}, t) \sim -F(\mathbf{s}). \tag{2.17}$$

This relation implies that an equilibrium quantity, the free energy, can be estimated by non-equilibrium dynamics, which changes during time with the addition of Gaussians. This property was postulated heuristically and then verified empirically. However, it can be demonstrated rigorously for a system evolving under the action of a Langevin dynamics [137, 143].

The choice of the Gaussian parameters defines the trade-off between accuracy and efficacy in the free energy surface (FES) exploration. Gaussians with a large value of height will fill the basins rapidly, but the consequently coarse grained reconstruction of the surface will be affected by bigger errors. On the contrary, smaller Gaussians

lead to an accurate profile, but the convergence may become expensive in terms of time and resources. A similar argument relates to the frequency of deposition: the addition of extra bias perturbs the system, which in turn needs time to recover equilibrium. A low deposition rate gives to the system the chance to relax between each metadynamics step, but lengthens the convergence time. On the contrary, a higher rate can speed up the simulation, but with the risk of affecting the dynamics. A thoughtful choice of these parameters needs to take into account the properties of the unbiased system and adapt. As a rule of thumb, a fraction ($\sim 1/2$ or $1/3$) of the fluctuations along the selected CV is taken as width σ of the Gaussians, and similarly their height should be a fraction ($\sim 1/10$) of the expected height of the free energy barriers. This choice should lead to a sufficiently accurate sampling of the free energy surface [137, 143, 144].

It is important here to briefly comment on the the choice of CVs. These variables are usually derived from simple functions of the atomic positions (e.g. distance between atoms, angles, torsional angles, but also the number of hydrogen bonds or the coordination number) and have to be chosen carefully: although this choice is not unique and strongly depends on the process one wants to observe, there are rules of thumbs that can help. It is required that such variables unambiguously identify and separate the different stable and metastable states that define the reaction of interest. This choice is not trivial: if selected incorrectly, the presence of hidden barriers may lead to an incomplete exploration and to an incorrect reconstruction of the free energy. Increasing the dimensionality of the CVs space would reduce the likelihood of this to happen, but, given the added difficulty to manage many variables and the decreased efficiency of metadynamics, it is not an ideal solution in most cases. It is thus generally advised to keep the number of CVs small, as the efficiency of the method scales exponentially with the number of variables involved. Intuitively, this can be understood by considering that the time required to escape from a local minimum is dependent on the amount of Gaussians needed to fill the well. This quantity is proportional to $(1/\sigma)^D$, where D is the number of dimensions (namely the number of CVs). When increasing D, the Gaussians have also to increase in size to maintain the efficiency. Soon, the Gaussians reach a size comparable to that of the wells they are supposed to fill, leading to significant errors in the FES reconstruction. With few particularly interesting exceptions [145], up to four CVs are typically chosen. Hence, although in principle metadynamics does not require the knowledge of the free energy path that one wants to explore *a priori*, knowledge of the system and its behaviour is indispensable. Extensive testing, often through simple trial end error, has to be carried out in order to identify the most adequate CVs for the selected problem. The best selection of CVs is not always clear and is often a matter of experience and taste [126].

To further improve the sampling, metadynamics can be combined with other approaches, as in the case of bias exchange metadynamics [146], where multiple replicas are run in parallel and biased along different CVs that can then be swapped as in other RE methods.
The the well-tempered version of metadynamics, described in the following, was chosen for this project, as implemented in PLUMED 1.3 [147] and PLUMED 2.0 [148] plug-ins within both the AMBER and NAMD packages.

2.2.1 Well-Tempered Metadynamics

To determine the convergence of a metadynamics run, one should ideally monitor the behaviour of the CVs and make sure that a diffusive regime has been reached. Once the FES is completely filled by the repulsive potential and all barriers and wells have been flattened, the system can freely move in the reduced space of the CVs as all the conformations are equiprobable. Assuming that CVs and Gaussian parameters have been appropriately chosen, the bias potential is the inverse cast of the underlying free energy landscape with a precision equal to the height of the Gaussian. Once the profile has been completely filled, the algorithm keeps adding layers of bias until interrupted. Depending on when the simulation is terminated, the final reconstruction may vary, due to the not uniform deposition within a finite simulation time. There are different methods to tackle this issue, the most trivial one being to calculate the average of a number of profiles extracted at discrete intervals after convergence [149].
A more sophisticated method to increase the accuracy of metadynamics is to decrease adaptively the deposition rate (the Gaussian height per unit time) of the bias with its well-tempered version [143]. This method automatically rescales the Gaussian height w during the simulation, according to the equation

$$w = w_0 e^{-\frac{U(\mathbf{s},t)}{k_B \Delta T}} \tag{2.18}$$

where w_0 is the starting Gaussian height, and ΔT is a tunable parameter with the dimensions of a temperature. The bias factor $\frac{T+\Delta T}{T}$, where $T + \Delta T$ is known as the fictitious temperature, is a parameter that regulates how fast the addition of Gaussians decreases with the simulation time and can determine the extent of the exploration. This parameter has to be chosen in such a way as to improve the accuracy of the reconstructed free energy, as shown in Figure 2.8, without slowing the convergence too much.
Thanks to this approach, the bias potential converges to the FES as in Eq. 2.19, avoiding overfilling and oscillations that can occur when the height of Gaussians is fixed [143, 147].

$$U(\mathbf{s}, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(\mathbf{s}) + const. \tag{2.19}$$

With well-tempered metadynamics, the effective sampling of the CVs is at the higher temperature $T + \Delta T$, which can be intuitively understood by looking at the two extreme cases, $\Delta T \to 0$ which leads back to classical MD and $\Delta T \to \infty$ which deactivates the lowering of the Gaussians height and recovers normal metadynamics.

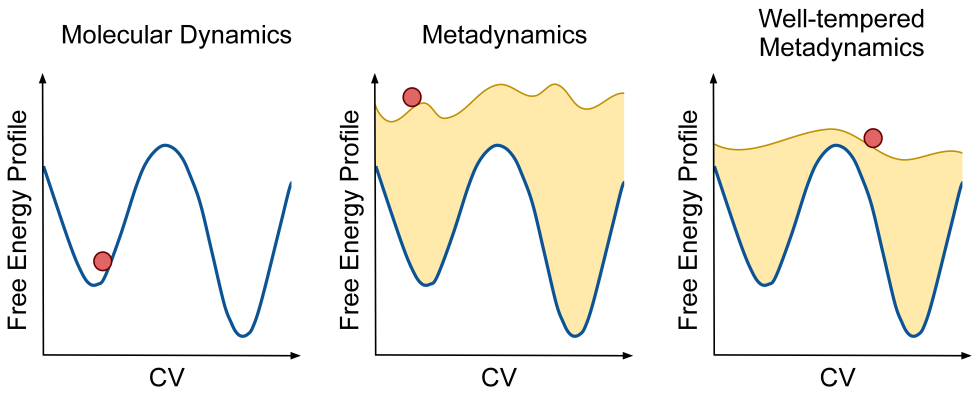


Figure 2.8: Schematic representation of the effects of MD and Metadynamics on a one dimensional free energy landscape. In blue is the true profile, in red is the position of the system and in yellow is the bias potential. On the right, the Well-tempered algorithm is able to prevent overfilling and decrease the error in the free energy cast.

2.2.2 The Reweighting Algorithm

While the probability distribution of the set of variables chosen as CVs can be correctly reproduced, the introduction of the bias potential may distort the distribution of the other degrees of freedom. Unfortunately, the reconstruction of the full Boltzmann distribution from a Metadynamics run is more difficult than in the case of other methods (e.g umbrella sampling), because of the time dependence of the bias. A reweighting algorithm has been thus developed with the aim to extract the unbiased probability distribution of any variable of interest from a well-tempered metadynamics run [150].
This technique exploits the relation between the biased probability distribution and the Boltzmann one, which can be expressed by the equation

$$P(\mathbf{r}, t) = e^{-\beta(U(\mathbf{s}(\mathbf{r}),t)+c(t))} \cdot P_0(\mathbf{r}) \tag{2.20}$$

where $P_0(\mathbf{r})$ is the Boltzmann distribution $e^{-\beta U(\mathbf{x})}/Z$ and $c(t)$ is the time dependent bias offset, defined as

$$c(t) = -\frac{1}{\beta}\log\left(\frac{\int d\mathbf{s}e^{-\beta F(\mathbf{s})}}{\int d\mathbf{s}e^{-\beta(F(\mathbf{s})+U(\mathbf{s},t))}}\right). \tag{2.21}$$

The reweighting algorithm has been used extensively in this project to reconstruct the free energy surface as function of different quantities not belonging to the chosen set of Metadynamics CVs. However, the results obtained by such algorithm need to be properly analysed to make sure that the unbiased CVs chosen for the re-weighting have been adequately sampled during the simulation. When this is not the case, partial or incorrect reconstructions may occur.

2.2.3 Funnel Metadynamics

An accurate estimate of the ligand affinity to its target is of great interest to pharmacological studies, where the drug design pipeline would widely benefit from an increase in speed and efficiency. The most commonly used methods to evaluate this quantity however rely on a number of substantial approximations that may undermine their reliability. The dynamical model, that accounts for conformational changes in both ligand and receptor during the binding process, is nowadays commonly accepted [151]; however some of these computational methods still rely to different degrees on the old "lock-and-key" theory of ligand binding [152], where a small molecules is expected to interact in the proper binding site of a static macromolecule. Structure-based drug design is the norm in the search for novel drugs by the pharmaceutical industry; docking methods usually rely on scoring function extracted by static conformations and novel dynamical techniques are sometimes used in a support role [100, 119].
The free energy is a state function, meaning that the difference between the energy of two states (such us a ligand bound to its receptor versus the respective unbound state) is independent on the path taken from the initial to the final event. This path influences the kinetics, the rate at which the reactions happens, but not the value of its free energy [100]. Among the end-point methods, which extract the free energy from the initial and final states only, the molecular mechanics/Poisson-Boltzmann surface area and the molecular mechanics/generalised Born surface area schemes are two of the most widely used. In these algorithms, the difference in free energy between the bound and unbound states is obtained through the computation of a molecular mechanics term, a solvation free energy, a conformational entropic contribution, and an electrostatic contribution calculated either with GB or PB, measured on a few static snapshots extracted from the dynamics of the bound state [153, 154]. These quantities are then calculated for the equivalent system deprived of the ligand, but no conformational change in the protein is taken into account at all. The output free energies are generally far from the real value, but they are still useful in qualitative

comparisons of different systems or to rank the affinities of different ligands to the same biomolecule.
Despite the fact that they do not require direct information of the transition paths, end-point methods are still affected by limited sampling. Conformational changes in the protein that may be decisive for the binding need to be known *a priori* and taken into account during these calculations, but the slow degrees of freedom involved in such changes are in most cases not accessible to plain MD or algorithm used to generate ensembles. The lack of sampling is probably the major drawback that prevents all these techniques to acquire reputation against cheaper structure-based docking [119].
Alternative strategies based on thermodynamic integration, free energy perturbation [155] or alchemical transformations [156] are in principle more accurate, but computationally more intense. In these cases the ligand-protein interactions is gradually switched off at variance to enhances sampling techniques, such as steered molecular dynamics and umbrella sampling, where the molecule is physically removed by its receptor. Some of these methods have been successfully used to measure the binding free energy, but they still require *a priori* knowledge of the binding path, limiting their transferability [119, 157].
This is not the case for metadynamics, which allows to extract information on the binding pocket, bound poses, pre-binding sites, etc. without necessarily requiring knowledge of the system in advance. Given the possibility to accelerate the dynamics with metadynamics, the idea of applying this technique to the study of ligand binding processes and to map its free energy landscape is attractive, nevertheless, the complexity of the binding problem can hinder the convergence of the simulations. Limited runs can still be useful for qualitatively exploring the interaction environment inside or in proximity of the binding pocket, but they may not be enough to extract an accurate quantitative value of the binding free energy [158, 159]. To improve on this limitation and allow an accurate evaluation of such quantity at a relatively affordable computational cost, a metadynamics-based method, called funnel metadynamics (FM), was thus designed [160].

The idea behind FM is to speed up the evaluation of the binding free energy and observe multiple binding and unbinding events in a relatively short computational time scale, by forcing the ligand to avoid the major part of the possible solvated states that lay outside the binding site. The movements of the ligand are limited inside a funnel-shaped restraining potential, so that the space explored in the unbound state is greatly reduced, while the configurations inside the binding pocked remain unaffected.
The shape of the funnel is defined by three parameters as shown in Fig. 2.9, and its effect on the binding has to be taken into account by adding a corrective term in

the computation of the energy. Once a statistically significant number of recrossing events between the bound and the unbound states has been observed, the absolute binding free energy can be evaluated by

$$\Delta G_b^0 = -\frac{1}{\beta}\log(K_b) \tag{2.22}$$

where K_b is the equilibrium binding constant. The latter, according to the approach used in Ref. [160] assumes, in the case of a funnel-shaped potential, the form

$$K_b = \pi R_{cyl}^2 C^0 \int_{z_{min}}^{z_{max}} dz\ e^{-\beta(w(z)-w_{ref})} \tag{2.23}$$

where R_{cyl} is the radius of the cylindrical region of the potential, $C^0 = 1/1660$ Å^{-3} is the standard concentration and z_{min} and z_{max} are respectively the lower and the upper limit along the axis of the funnel of the region within which the binding is considered. $w(z)$ is here the calculated one-dimensional potential of mean force, obtained as a projection of the free energy as a function of the position z along the funnel axis only, and w_{ref} is its value at a reference unbound state. The derivation of Eq. 2.23 can be found in Appendix A. As already mentioned, this formulation takes into account the effect of the cylindrical restraint on the solvated conformations, which are measured to be less energetically favourable than in reality due to the smaller volume available.

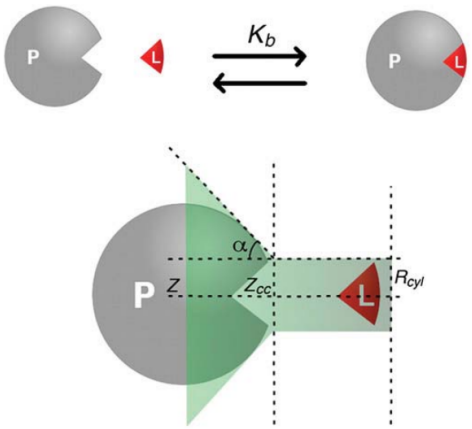


Figure 2.9: The concept of FM. The shape of the funnel, in green, can be adapted to the system by tuning few parameters: Z is the axis defining the exit-binding path of the ligand, Z_{cc} is the switching distance from a cone shape into a cylinder, while α is the angle of amplitude of the cone and R_{cyl} is the radius of the cylindrical section (from Ref. [160]).

2.3 Sketch-map

One of the main features of the Metadynamics approach is the reduction of the problem to a subset of collective variables describing the biological phenomena of interest. In a multi-dimensional free energy landscape the identification of the stable/metastable conformers and their connection network may not be trivial; even after this dimensionality reduction, however, a straightforward representation of the data may still pose a problem when three or more dimensions are involved. Simple projections to two-dimensions by integrating out the other CVs may result in unclear or incomplete representations and lead to a loss of precious information. It is sometimes possible to work around this issue, by exploiting as an example experiments or other sources of extra information on the system to guide the analysis, but when these data are not available the interpretation of the results is less obvious. Sketch-map is a new approach developed within this framework, aimed at mapping trajectory points extracted from the high-dimensionality CV space of Metadynamics or other atomistic simulations onto a bi-dimensional representation. This algorithm has been specifically designed to apply non-linear dimensionality reduction to atomistic simulation data characterized by a set of basins interconnected by transition paths, while dealing at the same time with the thermal fluctuation noise around these basins and the limited sampling of the transition pathways [161–163].
The core implementation of this algorithm is based on metric multi-dimensional scaling (MDS) [164] and requires, as first step, to select a limited number of high-dimensionality landmarks. This allows to speed up the calculation which would otherwise scale quadratically with the length of the simulation. The choice of landmarks is essential and it has to be representative of the underlying distribution of points in the multi-dimensional space. Random points are initially chosen by maximizing their relative distance through a farthest point sampling (FPS) algorithm to ensure that all configurations visited during the trajectory are represented. For each one of these points, a Voronoy polyhedron is built and probabilities, weighted according to the number of points falling in such polyhedra, are calculated. The remaining landmarks are then randomly selected according to this probabilities. This two-step procedure ensures that the Boltzmann distribution is adequately represented in the set of landmarks. The selection is followed by a multi-step minimisation of a stress function. This function is responsible for mapping the distances to the reduced space and, trivially, could assume the form:

$$\chi^2 = \sum_{i \neq j} (\mathbf{R}_{ij} - \mathbf{r}_{ij})^2 \tag{2.24}$$

where R_{ij} is a measure of the separation between the high-dimensional points and r_{ij} is the Euclidean distance between their projections. Although intuitive, these

Euclidean distances are not the ideal choice when working with MD trajectories because of their inherent non linear distribution in the CV space. For this reason, sigmoid filter functions have been introduced in place of the Euclidean distances. In Eq. 2.25a and 2.25b, the sigmoid functions for the high-dimensionality and low-dimensionality spaces respectively are shown

$$F(\mathbf{R}) = 1 - (1 + (2^{A/B} - 1)(\mathbf{R}/\sigma)^A)^{-B/A} \tag{2.25a}$$

$$f(\mathbf{r}) = 1 - (1 + (2^{a/b} - 1)(\mathbf{r}/\sigma)^a)^{-b/a}. \tag{2.25b}$$

These objects map the distances that are less than a reference value σ to be closer to zero, while those greater than σ are transformed into values closer to one. This choice guarantees that points closer to σ are projected close together, while those far from it are projected far apart. A, B, a, b and σ are all tunable parameters, which determine the shape of the sigmoid functions. However, the choice of σ is what really determines the projection: length scales lower than σ will be systematically ignored when projecting and one should carefully select which short-range features should be neglected. The Eq. 2.24 can be thus rewritten as

$$\chi^2 = \sum_{i \neq j} \left[F(\mathbf{R}_{ij}) - f(\mathbf{r}_{ij}) \right]^2 . \tag{2.26}$$

The remaining points are then projected so to maintain their proximity to the landmarks in both high- and low-dimensional spaces. If the simulation is properly converged and the choice of sketch-map parameters is adequate, one should obtain a projection of points on a plane, from which a two dimensional free energy map, with well separated basins and connection paths can be extracted in post-processing, as shown as an example in Fig. 2.10 [163].

Sketch-map has been successfully applied to map the conformations of Lennard-Jones clusters [163] and to distinguish helix-like and sheet-like secondary structures in metadynamics simulations of polyalanine chains [161]. However, the algorithm is not free from flaws, one for all the intrinsic necessity to sever connection paths when working with periodic data as shown in Fig. 2.10. In the figure, the three dimensional free energy landscape of a toy model with periodicity in the three directions is shown (a); the volume is characterised by low free energy basins connected to each other by transition paths, as in a typical MD simulation. Projecting an hypothetical trajectory from three dimensions (b) to two dimensions (c) causes the loss of information on a number of connections (in orange in the example), which are severed. Similarly, although all the basins are adequately identified, the algorithm cannot reproduce the correct distance relationship for all of them, and a few minima that should be mapped in close proximity are positioned at the opposite ends of the map (the orange

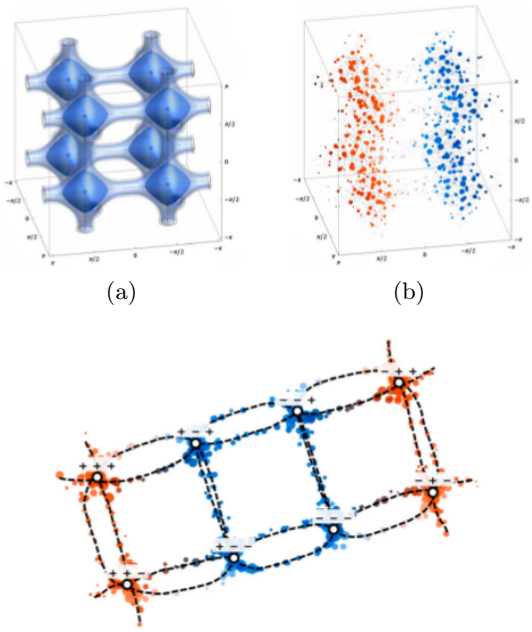


Figure 2.10: Example of the efficacy of sketch-map. In (a) a sample of free energy map in three dimension, where the basins and the connection paths are visible, in (b) a set of landmarks selected in the high-dimensionality space and finally in (c) the low-dimensional projection of the same landmarks produced with sketch-map. (From Ref. [161]).

basins in the figure). Moreover, the choice of the parameters can sometimes be counterintuitive; it requires long testing times and a meticulous refining procedure; finally the resulting map is not necessarily simple to understand, but hopefully will help in the interpretation of the results. When properly used, however, this algorithm can be a powerful tool, allowing to disentangle and reconstruct a bi-dimensional representation of the multi-dimensional free energy landscape, complete of its network of basins and paths, from which an identification of the minimum free energy conformer populations is straightforward.

Chapter 3

Results

This chapter is subdivided in four independent sections, where the results of different studies to understand the behaviour of pLGICs upon their activation by neurotransmitters are presented.
In Sec. 3.1 preliminary work to assess the reliability of commonly used methods such us homology modelling, docking and classical MD, was carried out on the ECD of the GABA-bound RDL neuroreceptor; most of these techniques require choices that, depending on the user choices and numerous other factors, can often affect significantly the outcome of the simulations. The results of this work led to the publication of Ref. [158], and allowed to select the best homology model for the following work with metadynamics.
As a matter of fact, the evaluation of quantities such us the binding free energy, crucial to the development of effective drugs and biologically active compounds require methods beyond the limitations of MD and conventional docking. Funnel metadynamics was thus introduced in Sec. 3.2 to accelerate binding and unbinding events and measure the binding free energy with statistical significance for the wild-type and two RDL receptor mutants. The results here obtained were published in the work of Ref. [165].
While the first two sections focus on the ligand binding problem in pLGICs, the second part of this chapter tries to shed light on the gating mechanism, by analysing interactions and dynamics of the loops at the interface between ECD and TMD of the 5-HT_3 serotonin receptor. In Sec. 3.3, the role a highly conserved proline and the effects of its *trans-cis* isomerisation on the structural and dynamical properties of the 5-HT_3R TMD have been investigated. The choice of working with this structure, as well as with the RDL receptor, was mainly dictated by the abundance of experimental data available to complement and compare these results. Moreover, their similarity to other Cys-loop receptors made them the ideal prototypes for such studies and increased range of applicability of methods and results here described.
Similarly, the availability of experimental results led to the choice of aspartic acid in

the study described in Sec. 3.4. A metadynamics and sketch-map based approach to the classification of its isomers in the gas phase and aqueous solution has been tested here. As confirmed by the study on the proline isomerisation in pLGICs, the conformational preferences of single residues can influence the behaviour of larger structures: with metadynamics it was possible to identify all major conformers, where experiments and conventional computational methods gave only partial results. This work will be published in the paper of Ref. [166], which is under preparation.

3.1 A Comparison of Models for the ECD of the RDL receptor

To elucidate the details of the neuroreceptor-neurotransmitter binding mechanism in the RDL receptor, conventional methods such us homology modelling, ligand docking and classical MD can be useful and informative. Data from mutagenesis electrophysiology experiments [50, 167, 168] was here successfully complemented with new computational information on the flexibility of the ligand, on the dynamical bond networks formed upon binding and on the role played by water molecules within the binding pocket. Both homology modelling and ligand-protein docking are however not unique procedures: the first significantly depends on the starting template and the alignment between the amino acids sequences, while docking results are not always reliable, since they are based on static models of the receptor and the scoring functions on which the selection of the poses is based are heavily approximate. An exploratory metadynamics on the binding integrated the results of this analysis and showed how enhanced sampling techniques can overcome the limitations of conventional techniques and mitigate the dependence on such choices. In order to assess systematically the reliability of these commonly used methods and to select the best model for a more thorough metadynamics study, an extensive comparison of models of the GABA-bound RDL receptor was carried out in this section.

Preparing the models

Three templates were considered for homology modelling, chosen among a limited number of available homologous proteins described in Sec.1.1: AChBP from the garden snail *Lymnaea Stagnalis* (entry 1I9B of the protein data bank [24]), ELIC from the bacterium *Erwinia Chrysantemi* (entry 2VL0 [25]) and GluCl from the nematode *Caenorhabditis Elegans* (entry 3RHW [15]) were chosen because of their different features and amino acids sequence identity against the RDL receptor. While both ELIC and GluCl have a transmembrane domain, AChBP can only be used to reconstruct the extracellular domain, being a globular protein rather than a membrane channel. Despite the fact that the interplay between ECD and TMD is

crucial during the transmission of the activation signal at later stages, the TMD and the surrounding membrane do not play a direct role in the ligand binding process to the orthosteric site: the binding pocket is quite far away from them and the slow conformational changes initiated by the neurotransmitter binding and then transmitted to the TMD are not sampled in these simulations. Moreover, the two domains are fairly independent: studies have shown that some isolated ECDs are able to fold correctly and even self-assemble [11, 169–171]. Models with only ECD are commonly used to study the binding process in ligand-gated ion channels [48, 167], also due to the limited structural information for the complete receptors. The reduced models allow to afford longer time scales while retaining all the important features for the binding. Nevertheless, the C_αs of the five terminal residues of each subunit were here restrained during the MD runs to simulate the effect of the TMD on the ECD.

In the case of AChBP and ELIC, where more than one PDB entry was present at the time of the work, the choice of a particular X-ray structure as template was based on the analysis of a number of scoring functions which helped assess the quality of the model, and of the sequence alignments and the structures produced out of them. FUGUE [172] is a server specialised in alignment and is able to recognise homologous proteins by comparing sequences and structures. It ranked 1I9B as a suitable template together with other AChBP PDB models. Nevertheless, 1I9B produced an alignment with higher sequence identity ($\sim$ 19 %) and was thus preferred. Similarly, 2VL0 was ranked as suitable homologue among the ELIC models and was then chosen mainly because of its high resolution (3.30 Å). The alignment in this case reached $\sim$ 24 % of sequence identity. Finally, GluCl 3RHW obtained the highest absolute score in the FUGUE ranking and was thus expected to produce the most reliable homology model. Two different alignments, one with CLUSTAL [173] and one with FUGUE, were then produced, with $\sim$ 39 % and $\sim$ 38 % of sequence identity (indicated as GluCl1 and GluCl2) respectively. This was aimed at comparing the influence that different alignments, even from the same template, could have on the simulation results. In some of these cases, alignment was also slightly manipulated manually, in order to account for the available experimental information and improve the reproduction of the most important regions (e.g. the binding sites), while other regions, like unstructured loops, which, being highly flexible, were expected to adapt to their natural shape during the MD simulation, were left untouched with respect to the automated alignment procedure.

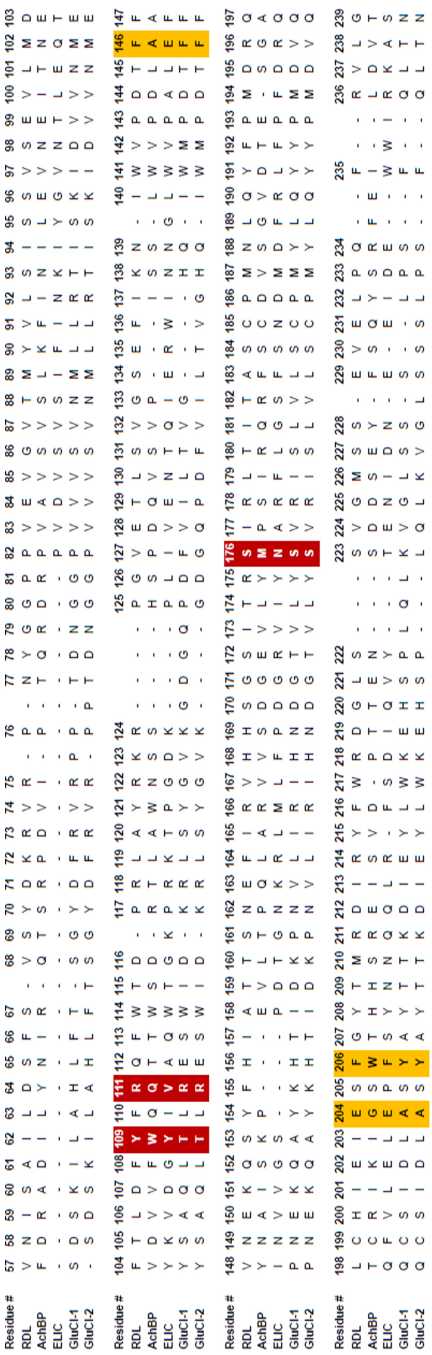


Figure 3.1: Sequence alignments of the insect RDL receptor ECD with AChBP (∼ 19% sequence identity), ELIC (∼24 %) and GluCl (∼ 39% for GluCl1 and ∼ 38% for GluCl2) templates. The residues identified as relevant for binding by mutagenesis experiments are highlighted (principal subunit in yellow: Phe146, Glu204, Phe206 and Tyr254; complementary subunit in red: Tyr109; Arg111 and Ser176). From Ref. [158].

The RDL receptor sequence here used is from *Drosophila Melanogaster* (NCBI accession: $NM_168321.1$, residues $57 - 266$). RDL receptor subunits can occur as different genetic splice variants, which have minor changes in their D and F loops, but the "ac" variant used in this study is considered the canonical isoform [174]. MODELLER v9.8 [175] was used to build the homopentameric RDL structures from this sequence. The structures built from each template and alignment were then labelled RDL-AChBP, RDL-ELIC, RDL-GluCl1 and RDL-GluCl2. During the building of these models, structural motifs (disulphide bridges typical for the Cys-loops) and symmetries among the five subunits were imposed, with the exception of RDL-GluCl2, where no symmetry was induced, although no actual difference was later observed in the stability of the simulation. For each template, 100 models were produced and then ranked according to GA431 and Normalised DOPE scores [176–178]. These quality indicators assess the structural properties of the model, defining statistical potentials composed of distant-dependent terms or solvent accessibility to discriminate if a structure is correctly folded [179]. Additional analysis was carried out with tools from the SWISS MODEL server [180] such as the QMEAN [181], a scoring function accounting for the major geometrical aspects of a protein backbone. Ramachandran plots, diagram of the possible torsional angle conformations of the protein backbone, were produced and analysed with PROCHECK [182] to exclude the substantial presence of residues in non-allowed regions [183].

The structures selected were then treated according to a standard protocol: standard amino acid protonation at neutral pH was applied, followed by solvation with 12 Å buffer of TIP3P water and addition of Na^+ and Cl^- ions to reach a physiological saline concentration of 0.15 M. Counterions were added to neutralise the structure and optimisation runs were applied to eliminate steric clashes. Using the AMBER 11 package and the AMBER ff03 force field [184], the structures were then minimised in stages, with progressive release of harmonic restraints on the position of atoms. The use of other more recent force fields within the AMBER family, which improve on some of the limitations of ff03, is unlikely to significantly affect the results for the RDL extracellular domain, as improvements are mostly related to helical propensity which is of limited relevance to the RDL receptor ECD [77, 185]. The standard AMBER minimisation uses first the steepest descent algorithm and then the conjugate gradient method. The whole protein was initially fixed to let water and ion relax, then only the C_α of the backbone were restrained and finally all the restraints were removed. The four optimised pentameric models are shown in Fig. 3.2, where relative differences of the backbone, with respect to that of RDL-GluCl2, have been highlighted. Small variations in the five subunits of each model, especially in the unstructured loops, can be observed due to the random initial positions of the water molecules and ions; however, these variations were not big enough to make the five interfaces non-equivalent. All structures are characterised by very high

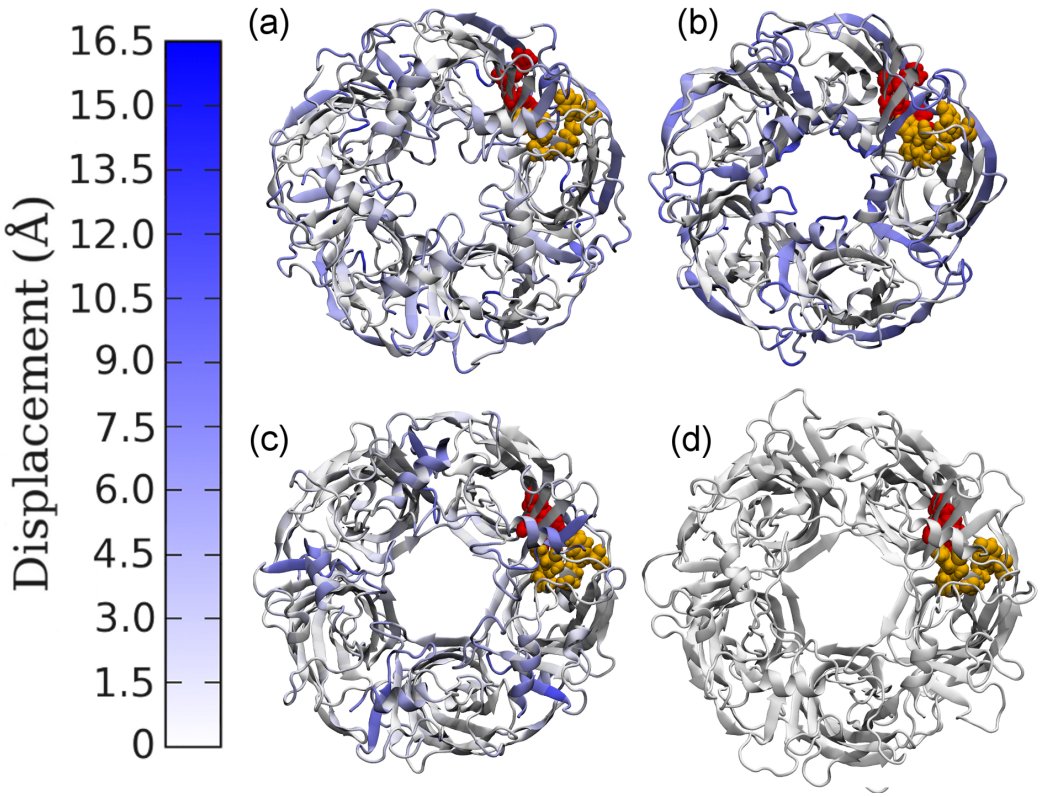


Figure 3.2: The selected homology models for the pentameric extracellular domain of the RDL receptor, after the structural optimisation: (a) RDL-AChBP; (b) RDL-ELIC; (c) RDL-GluCl1; (d) RDL-GluCl2. The seven residues experimentally identified as important for binding, Phe146, Glu204, Phe206 and Tyr254 in the principal subunit and Tyr109, Arg111 and Ser176 in the complementary subunit, are explicitly shown in red and yellow respectively. The displacement of the protein backbone from that of the RDL-GluCl2 model is highlighted. From Ref. [158].

percentages of residues in allowed regions of the Ramachandran plot (specifically 99.4 % for RDL-AChBP, 98.7 % for RDL-ELIC, 98.5 % for RDL-GluCl-1 and 98.6 % for RDL-GluCl2). Structural quality indicators pointed to better quality for the models built on the GluCl templates (which have a similar value of QMEAN) followed by that on AChBP and then ELIC.

Seven residues were identified by mutagenesis electrophysiology experiments as important for the binding of GABA in its zwitterionic form to the RDL receptor. These residues shown in both Fig. 3.3 (where the relative differences are highlighted) and 3.5 are Arg111, Glu204, Tyr109, Tyr254, Phe206, Phe146 and Ser174. Studies on other pLGICs [48] suggest that the negatively charged carboxyl group of GABA could interact with an arginine residue, while the positively charged amine could interact *via* hydrogen bonds with polar residues such as glutamic acid. Furthermore, a cation-π cage, formed by a number of aromatic residues such as phenylalanine and tyrosine, is expected to confine the amino group [49, 50, 167, 168]. Despite the

difference in the starting templates and alignment, the binding site was overall well conserved in all models, validating the choices made during the alignments and in the homology modelling. The main displacement is observed in Fig. 3.3 for Arg111. Neither AChBP nor ELIC have an arginine residue in the corresponding position in the alignment with the RDL receptor, at variance from GluCl. Hence, RDL-AChBP and RDL-ELIC might have had more difficulties in representing realistically Arg111. All the other residues show a relatively small displacement from the reference models, confined to the side chains which could also readjust during MD simulations. In particular Tyr109 was in the same rotameric state in the two GluCl based models, and behaved similarly during the corresponding MD simulations.

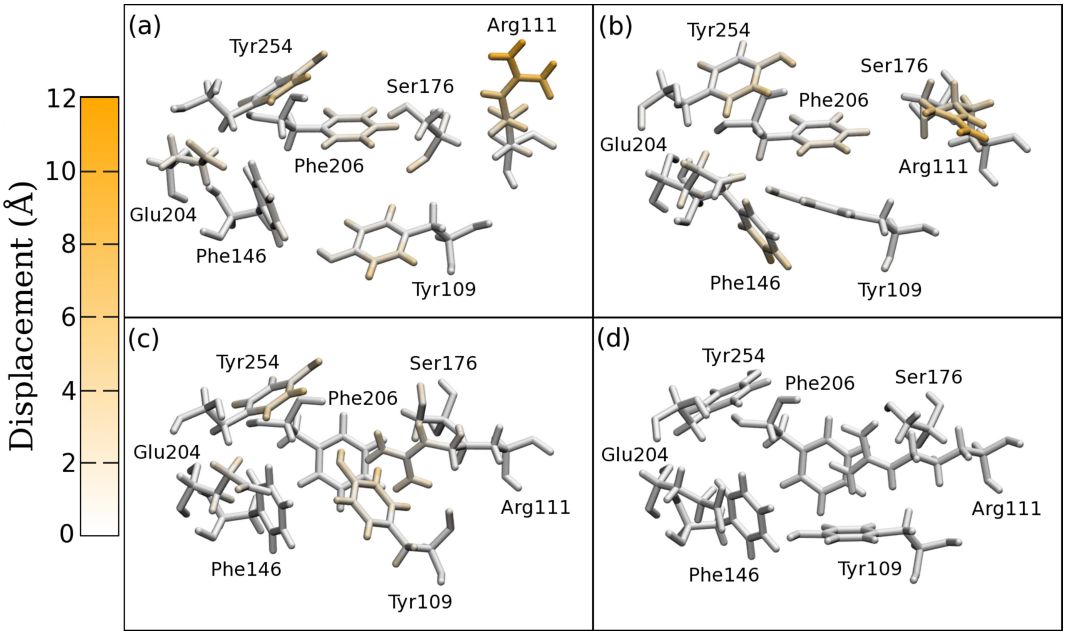


Figure 3.3: The GABA binding sites consisting of residues Phe146, Glu204, Phe206, Tyr254, Tyr109, Arg111 and Ser176, in the optimised homology models before GABA docking: (a) RDL-AChBP;(b) RDL-ELIC; (c) RDL-GluCl1; (d) RDL-GluCl2. The displacement of each atom from its position in the RDL-GluCl2 model is highlighted. From Ref. [158].

The experimental data were integrated in the docking process, where possible binding poses of the ligand were identified inside a user defined region, large enough to include the seven residues identified experimentally, as shown in Fig. 3.4 These were then ranked according to number and kind of interactions formed with the residues lining the pocket, to their roughly calculated binding energy and to conformational similarity among themselves. Docking was carried out with the AUTODOCK 4.2 suite [186]: GABA and the side chains of the first five residues listed before were allowed to be flexible during the random attempts to place the ligand within a user defined box-shaped region of search. A different procedure, where the region within 10 Å radius from Phe206 was imposed as the only restraint, was carried out with

GOLD 4.0 [187] on the RDL-GluCl2 model.
GABA is not a standard amino acid of ff03 force field and was thus necessary to calculate charges and parameters. ESP partial charges (which are fitted to reproduce the electrostatic potential) were evaluated within density functional theory with the CPMD code [188], using a plane wave basis set with a kinetic energy cutoff of 70 Ry, Martins-Troullier pseudopotentials [189] and the Perdew-Burke-Erzenhof (PBE) exchange and correlation functional [190]. The ESP partial charges were averaged over eight structures, in order to account somehow for molecule flexibility during the simulation.

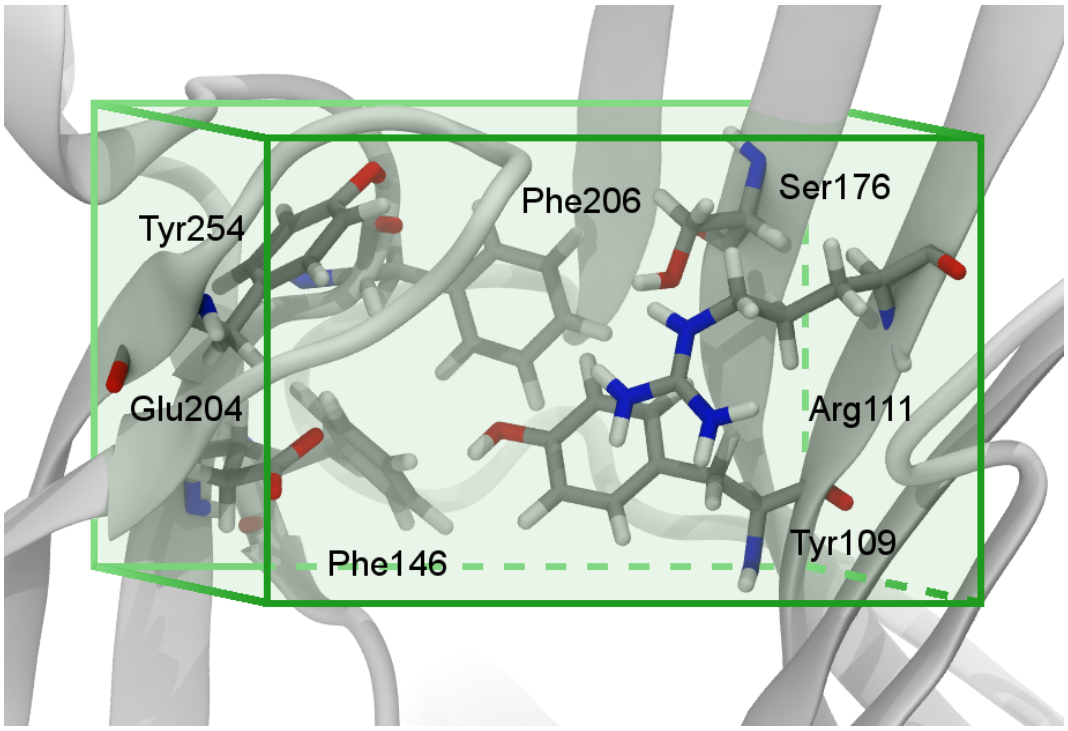


Figure 3.4: Volume box used to find the best docking poses of GABA to the RDL receptor ECD. The side chains of the seven residues identified experimentally are here shown and were left flexible during the search procedure.

A single GABA was docked to one of the equivalent interfaces for each model, although the current hypothesis suggests that at least two agonist bound to two non-consecutive interfaces are required for the optimal activation of pLGICs. The choice of working with a single GABA was justified by the assumption that non-consecutive interfaces are independent. The accuracy of this approximation was demonstrated in previous studies [167] and was confirmed by the results on further models, RDL-AChBP-5, RDL-ELIC-5, RDL-GluCl1-5 and RDL-GluCl2-5, where the docking poses were replicated at the five interfaces, improving the statistics.

Again, the selected structures were solvated with a 12 Å buffer of TIP3P waters in form of truncated octahedral cell with periodic boundary conditions. Na^+ and Cl^-

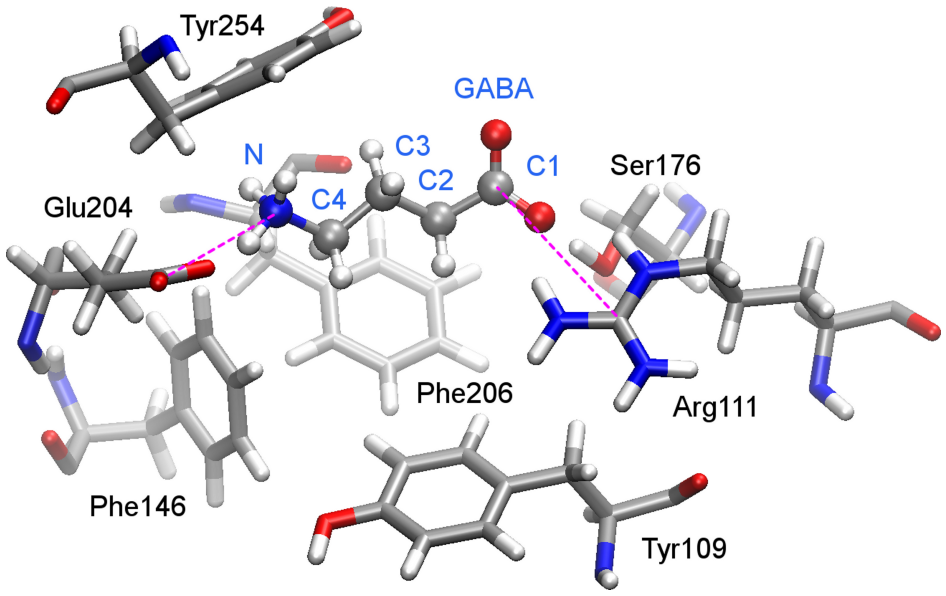


Figure 3.5: Example of GABA docked in the extracellular domain of the RDL receptor (RDL-GluCl1 model), with important residues in the binding site. The dashed lines represent the collective variables used in the Metadynamics simulation. From Ref. [158].

counterions were added to ensure charge neutrality and reach the physiological conditions of saline concentration at 0.15 M. After few stages of structural optimisation with gradually decreasing harmonic restraints on atomic positions and thermalisation, the RDL models were used as starting point for 25 ns of classical molecular dynamics. The vibrational motion of bonds containing hydrogen atoms was removed with the SHAKE algorithm [95] allowing a time step of 2 fs. The Particle Mesh Ewald method was applied to the long range electrostatic interactions and a cutoff of 10 Å was used for non-bonded interactions. During the 25 ns of MD, the temperature was maintained at 300 K with a Langevin thermostat with collision frequency γ of 1 ps^{-1}, while a Berendsen barostat with relaxation time τ of 2 ps kept the pressure around 1 atm. The first 3 ns were characterised by the presence of restraints on the C_α which were gradually removed, with the exception of the five terminal amino acid of each subunit, which were kept restrained to mimic the effect of the transmembrane domain during the whole run. Statistical averages were calculated over the last 12 ns when all models were equilibrated and stabilized; RDL-ELIC and RDL- ELIC-5 were the slowest models to equilibrate. Average quantities evaluated over longer times (e.g. 18 ns for the RDL-GluCl2 models) or over the five binding sites did not change the overall picture.

Comparison Results

In this section, an extensive comparison of structural stability, binding interactions and free energies among the four RDL models is presented. The models

built from GluCl templates show the most promising results, with higher number of hydrogen bonds and cation-π interactions between ligand and protein, and with a binding pocket less permeable to water. A preliminary metadynamics simulation confirmed the presence of two binding modes where the interaction with Glu204 can be either direct or solvent-mediated.

From the RMSD of the backbone atoms, all models were stable, given their complexity, and GABA remained in the binding sites in all the MD simulations. The average RMSD, calculated on the statistical sample, were 3.17 ± 0.11 Å for RDL-AChBP, 3.35 ± 0.09 Å for RDL-AChBP-5, 3.64 ± 0.13 Å for RDL-ELIC, 3.56 ± 0.16 Å for RDL-ELIC-5, 2.98 ± 0.09 Å for RDL-GluCl1, 3.19 ± 0.09 Å for RDL-GluCl1-5, 2.79 ± 0.12 Å for RDL-GluCl2 and 2.53 ± 0.06 Åfor RDL-GluCl2-5.

GABA Flexibility

In spite of being steadily bound to the receptor in all the MD simulations, GABA was characterised by an interesting degree of flexibility, as shown in Fig. 3.7 for the models with five ligands. In this image, the values of θ and ψ, the two torsional angles defined by the atoms C1-C2-C3-C4 and C2-C3-C4-N (see Fig. 3.5) respectively, are monitored: there are 9 possible conformers, as the preferred values are 60°, 180° and 300° for each torsional angle. Because of the force field, no hydrogen transfer between the GABA extremities was possible. The zwitterionic form should be favoured both in solution and in the receptor, due to its enhanced capability of interaction with solvent molecules and protein residues.

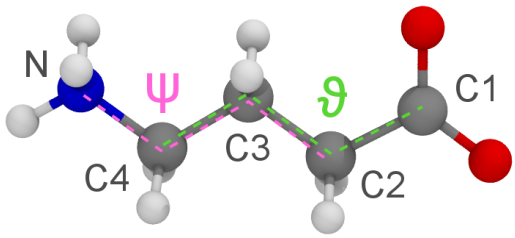


Figure 3.6: The dihedral angles θ (in green) and ψ (in magenta) of GABA.

GABA is in the extended conformation when both θ and ψ are 180° (center of each panel in Fig. 3.7), which was expected to be the most favourable [167]. It can potentially form intramolecular hydrogen bonds between its carboxylate and amino group when in the conformations corresponding to the quadrants at the four corners, but in competition with intermolecular bonds with either the solvent or the receptor. All conformers were easily accessible in water solution (panel (a)), with a preference for the extended configuration and low occupancies for the "corner"

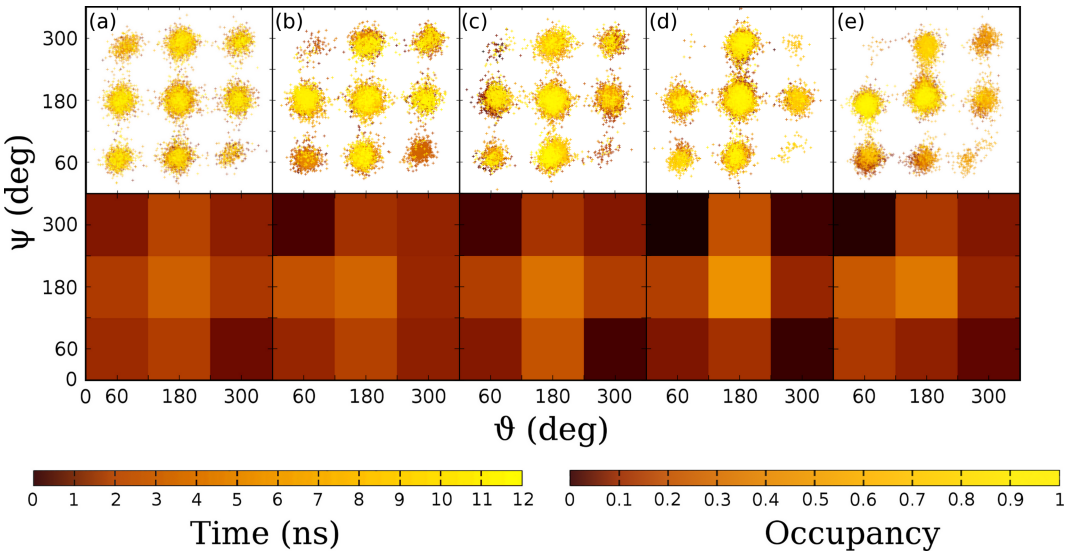


Figure 3.7: Time evolution (top) of the dihedral angles θ and ψ which characterise the GABA conformers, and conformer relative occupancy (bottom) during MD simulations in (a) water, (b) RDL-AChBP-5, (c) RDL-ELIC-5, (d) RDL-GluCl1-5, (e) RDL-GluCl2-5. The time evolutions of the five ligands are superimposed and the occupancies are averaged over the five binding sites. Adapted from Ref. [158].

conformers. Comparing to the case of GABA alone in water solution, there were more restrictions for the conformers in the RDL receptor models, where the top left structures were hardly populated. The extended conformer was overall preferred, without excluding the others: GABA can thus show a significant level of flexibility and still stay bound to the RDL receptor models.

Hydrogen Bonds Interactions

The hydrogen bonds analysis also gave some interesting insight into the interaction network formed inside the RDL receptor binding pocket. A bond was counted as present when donor and acceptor atoms were closer than 3.5 Å and when the donor-H$\cdots$acceptor angle was larger than 120°. The results are summarised in Table 3.1, where the number of hydrogen bonds averaged over the simulation time used to collect statistics is shown. Here, only the data for residues forming at least 0.10 hydrogen bonds on average are shown and will be here discussed for the models with five ligands, which were consistent with the single GABA-bound counterparts, but provided a significant improvement in statistics. Transient bonds were excluded from this analysis since they are not statistically significant and don't add information on the behaviour of the ligand and the binding site residues.

In all models bonding interactions with Arg111 were present. The positively charged arginine formed two or more salt bridges with the negatively charged carboxyl group of GABA in RDL-GluCl1-5 and RDL-GluCl2-5, while it formed on average less

	RDL-AChBP-5	RDL-ELIC-5	RDL-GluCl1-5	RDL-GluCl2-5	Water
Arg111	0.56 ± 0.67	0.71 ± 0.34	2.08 ± 0.65	2.16 ± 0.79	
H2O-mediated	0.46 ± 0.15	0.35 ± 0.20	0.17 ± 0.13		
Arg166		0.46 ± 0.50		0.55 ± 1.07	
H2O-mediated	0.28 ± 0.21	0.72 ± 0.56		0.17 ± 0.26	
Glu204	0.65 ± 0.34	0.44 ± 0.33	0.44 ± 0.36	0.60 ± 0.36	
H2O-mediated	0.49 ± 0.23	0.96 ± 0.11	0.61 ± 0.18	0.57 ± 0.40	
Glu246		0.17 ± 0.33			
Phe206	0.12 ± 0.15	0.12 ± 0.20	0.32 ± 0.24	0.22 ± 0.26	
Ser176	0.64 ± 0.44	0.58 ± 0.40	0.89 ± 0.19	0.94 ± 0.32	
H2O-mediated	0.20 ± 0.12	0.33 ± 0.15	0.14 ± 0.16	0.16 ± 0.19	
Ser205	0.62 ± 0.32	0.28 ± 0.33	0.57 ± 0.39	0.61 ± 0.30	
H2O-mediated	0.18 ± 0.13	0.31 ± 0.22	0.28 ± 0.30	0.24 ± 0.30	
Thr250	0.27 ± 0.47	0.33 ± 0.44			
Thr251			0.42 ± 0.37	0.53 ± 0.45	
Tyr109	0.54 ± 0.58	0.34 ± 0.35	0.27 ± 0.37		
Tyr254	0.12 ± 0.23	0.31 ± 0.39			
Protein	3.63 ± 0.67 (2.56)	3.80 ± 1.22 (4.23)	5.06 ± 0.62 (5.24)	5.72 ± 1.05 (6.79)	
H2O-mediated	1.88 ± 0.67 (2.82)	2.92 ± 0.51 (2.09)	1.53 ± 0.38 (1.05)	1.27 ± 0.96 (0.27)	
H2O	3.99 ± 0.59 (6.41)	4.32 ± 0.95 (3.77)	2.78 ± 0.68 (2.83)	1.97 ± 1.10 (0.27)	9.17
Intra-molecular	0.09 ± 0.13 (0.04)	0.02 ± 0.02 (0.05)	0.01 ± 0.02 (0.12)	0.04 ± 0.05 (0.0)	0.09
Total	7.71 ± 0.33 (9.01)	8.14 ± 0.30 (8.05)	7.85 ± 0.45 (8.19)	7.73 ± 0.26 (7.73)	9.28

Table 3.1: Average number of hydrogen bonds formed by GABA in the RDL-models and in water solution. The residues identified by experiments as important for receptor activation are in boldface. For interactions with specific residues, only average values equal or larger than 0.10 are shown; total values include all contributions. The data for the models with five ligands are averaged over the five binding sites; the data for the corresponding models with one ligand are shown in brackets. From Ref. [158].

than one direct hydrogen bond in RDL-AChBP-5 and RDL-ELIC-5, where however it also participated in water-mediated hydrogen bonds. This discrepancy can be due to the positional difference of Arg111 in the RDL-AChBP model and RDL-ELIC models with respect to the models based on the GluCl templates, as discussed previously. A second arginine, Arg166, was also bonding in RDL-AChBP-5, RDL-ELIC-5 and RDL-GluCl2-5; this interaction was also observed with low frequency in Ref. [167]. Glu204 formed salt bridges in all models through its negatively charged carboxyl group interacting with the positively charged amine of GABA. A significant percentage of this interaction was mediated by a molecule of water. On the right hand side of Fig. 3.8 two snapshots from the following exploratory metadynamics simulation of RDL-GluCl1 are shown, where the two possible states, with (top) or without (bottom) the mediation of water are shown. The polar residue Ser176 interacted with GABA through hydrogen bonds in all models consistently with experimental evidence; also Ser205 formed direct or water mediated hydrogen bonds. Thr250 interacted through hydrogen bonding with GABA in RDL-AChBP and in RDL-ELIC, but insignificantly in the other two models, where Thr251 formed hydrogen bonds

instead. The aromatic residues Phe206, Tyr109 and Tyr254 formed fairly infrequent hydrogen bonds with GABA, the most populated of which occurred with Try254 in RDL-AChBP-5; they were however involved in cation-p interactions as described below. Interestingly, GABA formed more hydrogen bonds directly with the receptor in models built with the GluCl template, where more than five hydrogen bonds were observed in all cases. A part from the mediation of the Glu204 interaction, GABA also formed hydrogen bonds with other water molecules in the binding site. The percentage of intramolecular hydrogen bonds between the GABA carboxylate and the amino groups was lower than 10% in all cases, and linked to the occupancy of the "corner" conformers of Fig. 3.7. The total number of hydrogen bonds formed by GABA in the MD simulations both with the receptor and water is lower than that found in water solution, ~ 9.3, of which ~ 6 are with the carboxyl group and 3 with the amino group, exhausting all the possibilities to form hydrogen bonds.

Cation-π Interactions

Data from different pLGICs seem to confirm the hypothesis that aromatic residues in the binding pocket help stabilise the positive moiety of the ligand, by means of an aromatic cage [34, 48]. As observed in Fig. 3.5, this is the case also for the RDL receptor, where a few amino acids with aromatic rings are found surrounding the amino group of GABA. This time an interaction was counted when the distance between the positively charged atom and the center of the carbon ring was lower than 6 Å and the angle between the vector joining the cation and the ring center, and the vector normal to the ring plane was lower than 60°. The average number of cation-π interactions, counted during the simulations and shown in Table 3.2, ranged from less than one on average in the AChBP based models, mostly due to Tyr254, to more than two in the GluCl based models, where GABA was observed interacting with Phe206, Tyr109 and Tyr254 and also Phe164, with intermediate values for the models based on ELIC. No evidence of a direct interaction between GABA and Phe146 was found. This is consistent with mutagenesis data showing that a mutation of this residue with Ala produced only a moderate increase in EC50, indicating that there is no requirement for an aromatic residue in this position and that Phe146 does not contribute to cation-π interactions and is not critical for binding [50, 167].

Binding Free Energies

As mentioned in Chapter 2, the MM/PBSA and the MM/GBSA methods are two of the most commonly used techniques to evaluate the binding free energy of a ligand-receptor complex. Although the idea is very simple, they are often not very accurate and the results cannot be considered reliable from a quantitative point of view. As an example, all water are considered implicitly, despite the hydrogen

	RDL-AChBP-5	**RDL-ELIC-5**	**RDL-GluCl1-5**	**RDL-GluCl2-5**
Phe164		0.12 ± 0.24	0.30 ± 0.3	0.13 ± 0.17
Phe206	0.12 ± 0.24	0.35 ± 0.34	0.64 ± 0.22	0.78 ± 0.24
Tyr109	0.05 ± 0.10	0.31 ± 0.35	0.22 ± 0.26	0.27 ± 0.23
Tyr254	0.80 ± 0.34	0.52 ± 0.34	0.97 ± 0.03	0.99 ± 0.01
Total	0.97 ± 0.15 (0.65)	1.33 ± 0.94 (1.53)	2.16 ± 0.68 (2.48)	2.18 ± 0.49 (2.46)

Table 3.2: Average number of cation-π interactions formed by GABA in the RDL models. The residues identified by experiments as important for receptor activation are in boldface. The data are averaged over the five binding sites for the models with five ligands, while the corresponding results for the structures with one ligand are shown in brackets. From Ref. [158].

bonds analysis demonstrated their role in actively mediating important interactions. Nevertheless, these techniques are very useful in ranking similar systems: in Table 3.3 the binding free energies evaluated on the models bound to a single ligand are shown, broken down into the enthalpic and entropic contributions obtained with normal mode analysis.

	ΔH **(kJ/mol) MM/GBSA**	ΔH **(kJ/mol) MM/PBSA**	$T\Delta S$ **(kJ/mol)**	ΔG **(kJ/mol) MM/GBSA**	ΔG **(kJ/mol) MM/GBSA**
RDL-AChBP	-34.73 ± 21.34	-30.12 ± 25.10	-60.67 ± 45.19	26.36 ± 50.21	30.96 ± 50.79
RDL-ELIC	-78.66 ± 13.39	-76.57 ± 21.76	-40.58 ± 23.01	-38.07 ± 26.36	-35.98 ± 30.96
RDL-GluCl1	-98.74 ± 15.90	-91.63 ± 26.78	-49.37 ± 27.20	-48.95 ± 31.38	-42.26 ± 38.07
RDL-GluCl2	-144.35 ± 12.55	-161.50 ± 16.74	-58.16 ± 31.80	-119.66 ± 33.89	-103.34 ± 35.98

Table 3.3: Enthalpic (ΔH) and entropic ($T\Delta S$) contributions and free energy of binding (ΔG) of GABA to the RDL receptor models calculated with the MM/GBSA and MM/PBSA methods. From Ref. [158]

MM/PBSA is more accurate and more computationally expensive then MM/GBSA, but the results are consistent and the trends are in line with the hydrogen bonds and cation-π analysis, showing the weakest binding in the case of RDL-AChBP, where water-mediated interactions contributed significantly, and the strongest one in RDL-GluCl2, were the water played a minimal role.

Preliminary Metadynamics

In order to obtain a quantitatively accurate evaluation of the free energies, the limited sampling accessible to these standard techniques needs to be improved, e.g.

with the introduction of enhanced sampling methods such as metadynamics. A preliminary free energy exploration of the RDL-GluCl1 binding pocket was thus carried out here with metadynamics, to test the efficacy of this method. The FES was partially mapped as a function of two CVs, representing the distance between the carbon of the GABA carboxyl group and the carbon connected to the three nitrogens in the Arg111 side chain (CV_{Arg}) and the distance between the nitrogen of the GABA amino group and the carbon of the Glu104 carboxyl group (CV_{Glu}), as shown with dashed lines in Fig. 3.5. Confining walls at 9 Å were introduced to avoid the sampling of regions too far away from the binding site. The well-tempered metadynamics simulation was started from the equilibrated MD structures, using the PLUMED 1.2 plugin coupled to the AMBER 11 package. A bias factor of 10 was chosen, while Gaussians with a 0.3 Å width and an initial height of 0.4 kJ/mol were deposited every ps for a total of 50 ns.

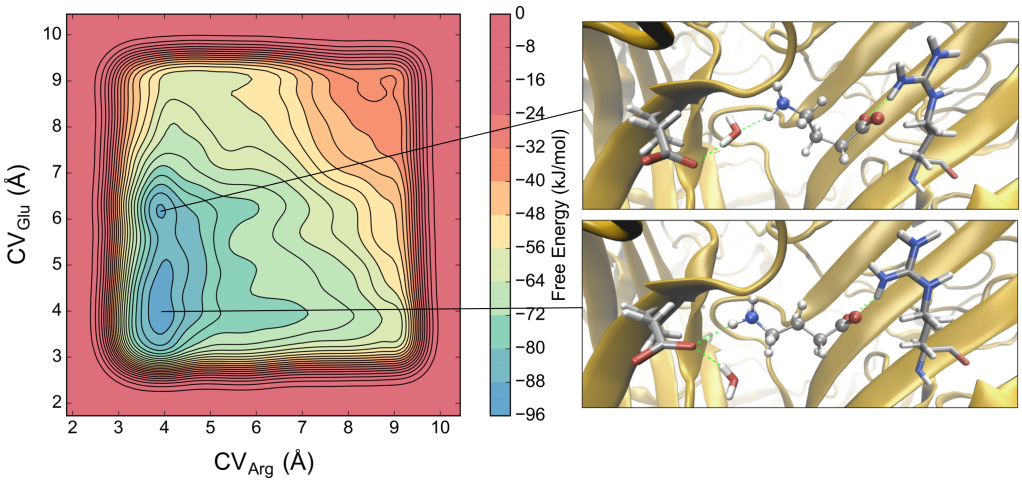


Figure 3.8: Exploratory free energy map of GABA in the RDL-GluCl1 model as a function of the distance of the GABA amino group from Glu204 side chain (CV_{Glu}) and of the GABA carboxyl group from the Arg111 side chain (CV_{Arg}). Contour lines are every 8 kJ/mol. The binding arrangements corresponding to the free energy minima are shown. Adapted from Ref. [158].

The resulting free energy map is shown in Fig. 3.8. It is possible to identify an elongated basin of attraction corresponding to a range of relative distances between the amine of Glu204, while the GABA carboxylate was clearly pinned to Arg111. Two minima separated by a low barrier can also be identified, related to the binding configurations shown on the right hand side: in one case (the bottom minimum) the amino group of GABA, in the extended conformer, is closer to Glu204 and forms a direct hydrogen bond, while in the other case (top minimum) a bent conformer is observed forming a water-mediated interactions with the same residue.

Conclusions

In summary, five stable models of the RDL receptor were built from homology modelling and studied with the use of Molecular Dynamics after docking GABA in them; this allowed to investigate in detail the interactions underlying the binding of the ligand to the neuroreceptor. The network of interactions that GABA formed inside the five RDL binding pockets showed quantitatively common features and model independent informations were extracted. The role of residues highlighted by experiments as important was confirmed and clarified, while new residues, not studied experimentally before (e.g. Ser205), were suggested as potential interaction partners. Also, experiments have a limited capacity to probe the role of water molecules in such systems: water mediated interactions were here observed and analysed, providing new insights into alternative binding conformations, although the importance of water was showed to be dependent also on the quality of the model. Models built from templates with lower sequence identity were less stable and were characterised by less direct interactions and a more water permeable binding site. The presence of more solvent molecules that could interact with the ligand while bound decreased its interactions with the polar residues of RDL, causing an overall weaker binding. GluCl, instead, was confirmed to be the best template and provided stable and reliable models.

Different binding arrangements were observed in the MD simulations. They are not mutually exclusive, instead they may coexist or exist at different times and are consistent with the available experimental data. The proposal that there is a degree of flexibility in the binding arrangements is supported by the exploratory data on the binding of GABA to the RDL receptor obtained with metadynamics, which identified two minima separated by a small barrier, related to the binding configurations in the presence and absence of a water molecule which mediates the interaction between the GABA amine and Glu204.

Extending these preliminary MD would bring little improvement, as the results would be still affected by the limited time scale, by the not negligible influence of the starting docking pose and by the approximate evaluation of the binding free energy with post-processing techniques such as the MM/PBSA and MM/GBSA methods. Given this first promising run, metadynamics was thus chosen to simulate the entire binding process, to achieve a more thorough exploration of the orthosteric pocket, and to finally predict with accuracy the binding free energy. These results, together with the work described in the next section, outline a simulation protocol for the investigation of neurotransmitter-neuroreceptor binding in pLGICs.

3.2 The Binding of GABA to the RDL receptor with Metadynamics

Metadynamics has been successfully used in studies at the interface between physics, chemistry and biology for its ability to improve the exploration of the phase space and, once this exploration is complete, to efficiently reconstruct through the added bias the free energy landscape as a function of selected CVs. However intuitive the idea to apply it to estimate binding free energies may be, its implementation is far from trivial. To obtain a reliable evaluation of such quantity, the observation of a statistically significant number of forth-back events between the ligand bound and unbound states is necessary. While a ligand can be easily driven out of its protein binding site with metadynamics, the huge number of unbound conformations in the solvent dramatically decreases the probability to observe a binding event in a reasonable computational time.
A simple and elegant solution to this problem is the introduction of a funnel-shaped restraining potential to limit the exploration of the solvated states outside the binding pocket in a scheme which as has been named "funnel metadynamics" [160]. This method has been recently applied to a few biological systems and proved effective in enhancing the sampling of bound and unbound states and evaluate with accuracy the free energy differences in complex biological systems [191–193].

In the present work, funnel metadynamics was applied for the first time to investigate the free energy landscape of a neurotransmitter binding to a pLGICs and assess how such a landscape is affected by selected mutations, which experimentally produce non-functional channels. Rationalising the different binding mechanisms and affinities of a ligand toward the wild-type and mutant forms of its molecular target is, in general, a crucial issue in the chemical and pharmacological industries. The picture stemming from experiments and previous simulations (as discussed in Sec. 3.1) of the binding process of GABA to the RDL receptor and its non-functional Arg11Ala and Glu204Ala mutants was completed. With μ-sec FM simulations, a large number of binding and unbinding events that lead to a quantitatively well-characterised free energy landscape and accurate estimate of the ligand binding free energy were reproduced. Furthermore, this work provided structural insights into the ligand binding mechanism with atomistic resolution, identifying crucial protein motions like the dynamics of loop C which partially protects the binding site from the solvent and is expected to play an important role in the activation processes of Cys-loop receptors. [51, 194–196]
Since the RDL receptor is structurally similar to other pLGICs, this work represents a point of reference for further investigations on the binding of neurotransmitters and ligands to pLGICs beyond conventional docking and MD-based techniques.

Simulation Details

Among the models tested in the previous session and in Ref. [158], RDL-GluCl2 resulted in better stability of the simulation and improved consistency with experimental information. This was the motivation behind the choice of this specific model as starting point to proceed with an accurate study of the binding properties of GABA and the evaluation of the binding free energy with metadynamics. This work concentrates on the binding and unbinding of a single ligand, since MD simulations with one or five ligands did not provide significant differences in the binding features [158]. Together with the ECD of the wild-type RDL receptor, two mutants where the charged residues Arg111 and Glu204 were substituted in turn with non polar alanine (named here Arg111Ala and Glu204Ala respectively) were prepared according to the procedure described in Sec. 3.1. In these models, as in previous works, the TMD and lipid membrane were not included, since they do not play an active role during the binding process. The C_αs of the five terminal residues of each subunit were restrained to mimic the effect of the TMD on the ECD instead, as before.
The system was prepared with a 12 Å buffer of TIP3P waters and 0.15 M of (NaCl) salinity to reproduce physiological conditions in a periodically repeated truncated octahedral supercell, and counter ions were added to neutralise the total charge. The solvated model contained about 83,000 atoms. The AMBER ff12SB [97] force field was used in this case: this is an improved force field with respect to the ff03 previously employed [48, 158] and, with respect to the widely used ff99SB-ildn [197] which we also tested, and results in a improved stability on long time scales for the β-hairpin of the loop C, which is critical for the binding process. Restrained ESP charges were calculated for GABA at the Hartree Fock level of theory, with a 6-31G* basis set, on the molecular geometry optimised with density functional theory and the B3LYP exchange and correlation functional using Gaussian 09 [198] and AmberTools [97]. All metadynamics simulations were run with NAMD 2.9 patched with the PLUMED 1.3 metadynamics plug-in. To enforce ambient conditions (T=300 K and P=1 bar), a Langevin piston barostat with oscillation period of 200 fs and damping timescale of 100 fs was introduced in place of the Berendsen barostat, while the remaining parameters (for the Langevin thermostat, Particle-Mesh Ewald and SHAKE) were kept the same as in the preliminary MD simulations (see Sec. 3.1).
After the minimisation and equilibration procedures, 100 ns of MD were carried out, during which the stability of the structure was monitored through the root mean square deviation (RMSD) of the backbone atoms of the secondary structures with respect to the minimised structure, before starting the metadynamics simulations.

Funnel-metadynamics simulations

A number of exploratory metadynamics simulations were first performed without the restraining potential to get an idea of the preferred path for unbinding with statistical significance. Gaussians with a height of 0.4 kJ/mol and a width of 0.15 Å were deposited every 2 ps along the trajectory of the selected CV: the distance between the centre of mass of GABA and the centre of mass of the C_αs of the seven residues identified by experiments as important for binding. This procedure allowed us to also to observe the behaviour of the secondary structures surrounding the binding site upon unbinding. Monitoring slow motions in proximity of the regions of interest is fundamental to detect structural features that may influence the binding. Particular attention had to be paid, for example, to loop C, highlighted in orange in Fig. 3.9 a), which partially protects the binding site from the solvent and, as already mentioned, is expected to be actively involved in the binding process of Cys-loop receptors. Indeed, we observed that loop C is able to open and close within a few nanoseconds following the passage of the ligand, when described with the ff12SB force field and thus it does not represent an obstacle to the binding of the ligand and did not require a specific CV to describe its motion. Conversely, the β-hairpin at the edge of this loop tends to lose its structure when ff99SB-ildn force field is used, decreasing the potency of the binding. Indeed, we observed that loop C is able to open and close within a few nanoseconds following the passage of the ligand, when described with the ff12SB force field and thus it does not represent an obstacle to the binding of the ligand and did not require a specific CV to describe its motion. Conversely, the β-hairpin at the edge of this loop tends to lose its structure when ff99SB-ildn force field is used, decreasing the potency of the binding.
The information gathered from these preliminary runs allowed to position the funnel-shaped restraining potential for GABA centre of mass in a meaningful way as shown in Fig. 3.9 c): the conical region surrounding the binding site was positioned so not to interfere with the binding and unbinding processes, while its dimensions were chosen as reduced as possible to speed up convergence, but large enough to avoid artifacts in the sampling. The cone angle α was set to 25°. The zero of the coordinate system was set around the minimum free energy for the wild-type receptor, with the conical region between $z = -5$ Å and $z_{cc} = 20$ Å, the point where the potential switched to a cylindrical shape with radius $R_{cyl} = 1$ Å and length 10 Å. Soft harmonic restraining walls were applied at both ends of the funnel axis to limit the exploration of the ligand inside the region enclosed by the funnel potential.

Consistently with what was observed in the preliminary docking and MD simulations of Ref. [158], the zwitterionic form of the ligand and the presence of residues of opposite charge in the binding site (i.e. Glu204 and Arg111 in the principal and complementary subunits respectively) support the picture that the only plausible

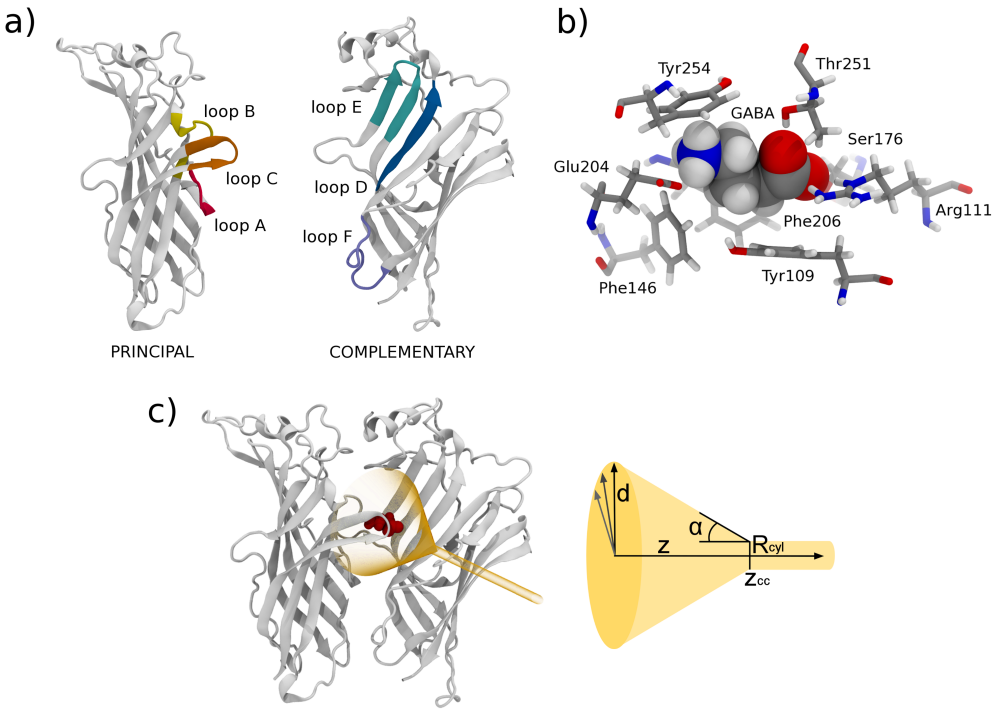


Figure 3.9: a) Principal and complementary subunits of the RDL receptor ECD, with highlighted secondary structures. b) GABA in the binding site with the seven residues identified by experiments as important for binding and Thr251 in loop C. c) Left, principal and complementary subunits with GABA bound in red and the funnel restraining potential. Right, geometric parameters of the funnel restraining potential. Adapted from Ref. [165].

binding poses should all have a similar orientation with respect to the binding pocket, with GABA carboxyl group interacting with Arg111 and the amino group interacting with Glu204. No major rotations or change in orientation were observed when the ligand was inside the binding pocket, making it unnecessary to use a CV describing the orientation of GABA. This is consistent with the unbinding tests, where no major rotations or change in orientation were observed when the ligand was inside the binding pocket. Hence, the position along the funnel axis (z) and the distance from it (d), as shown in Fig. 3.9 c), were chosen: these proved to be an appropriate selection as they allowed a complete exploration of the space inside the funnel potential without producing hysteresis in the binding and unbinding paths or other artifacts. Although these CVs easily discriminate between bound and unbound states, the reweighting algorithm illustrated in Sec. 2.2 was applied to the trajectories in order to remap the free energy as a function of alternative CVs and gain more detailed insights, such as revealing bound minima which are overlying in the maps corresponding to the initial variables choice.

Under well-tempered metadynamics, Gaussians were deposited every 2 ps. The initial height was 2.0 kJ/mol, the target temperature 300 K and the bias factor 15;

the simulations were carried out in the NVT ensemble starting with the average volume obtained in the equilibrated MD. The calculations were considered converged when a statistically significant number of re-crossing events between the bound and unbound states had been observed. Once this regime is reached, the difference in free energy between the two states oscillates around a constant value: from this oscillations, one can then easily extract the value of binding free energy and its error. In the wild-type receptor the metadynamics simulation were carried out for 1.25 μs. The same protocol was then applied to the two Arg111Ala and Glu204Ala mutants, with the goal to compare the resulting free energy landscapes. The simulations were carried out for 0.75 μs for the Arg111Ala RDL receptor and for 0.95 μs for the Glu204Ala RDL receptor, where the binding is weaker and the ligand easier to drive out of the binding site. The free energy landscapes and binding free energies were averaged over the last 0.25 μs in all models, to reduce the metadynamics error and increase the accuracy of the reconstructed profiles [137, 149]. Binding free energies were evaluated following the approach described in Ref. [160] and reported in Sec. 2.2.3, which accounts for the effects of the cylindrical restraint on the solvated states.

Wild-type RDL Receptor Results and Discussion

In this section, the binding free energy landscape of the wild-type RDL receptor is presented. Two equiprobable binding modes are identified, together with a pre-binding conformation. The flexibility of loop C plays a crucial role in allowing the ligand to enter and exit the binding pocket. A favourable binding path is also described in detail.

During the metadynamics simulations, we monitored the conformational stability of the RDL receptor models by calculating the RMSD of the secondary structure backbone atoms with respect to the corresponding initial minimised structures. The average RMSD was 2.3 $\pm$ 0.2 Å for the wild-type RDL receptor, 2.6 $\pm$ 0.2 Å for the Arg111Ala mutant and 2.6 $\pm$ 0.1 Å for the Glu204Ala mutant. Hence the overall structure of the receptor was preserved throughout the μ-sec long simulations, during which the loops and the binding site residues showed the necessary flexibility to allow the binding and unbinding of the neurotransmitter. In the free energy surfaces as a function of the biased CVs in Fig. 3.10 a) it is clear that the selected mutations induce drastic changes in the free energy landscape. This is even more evident in the one-dimensional free energy profiles projected onto the funnel axis in Fig. 3.10 b): an energetically favourable bound state for GABA is present only in the wild-type receptor.

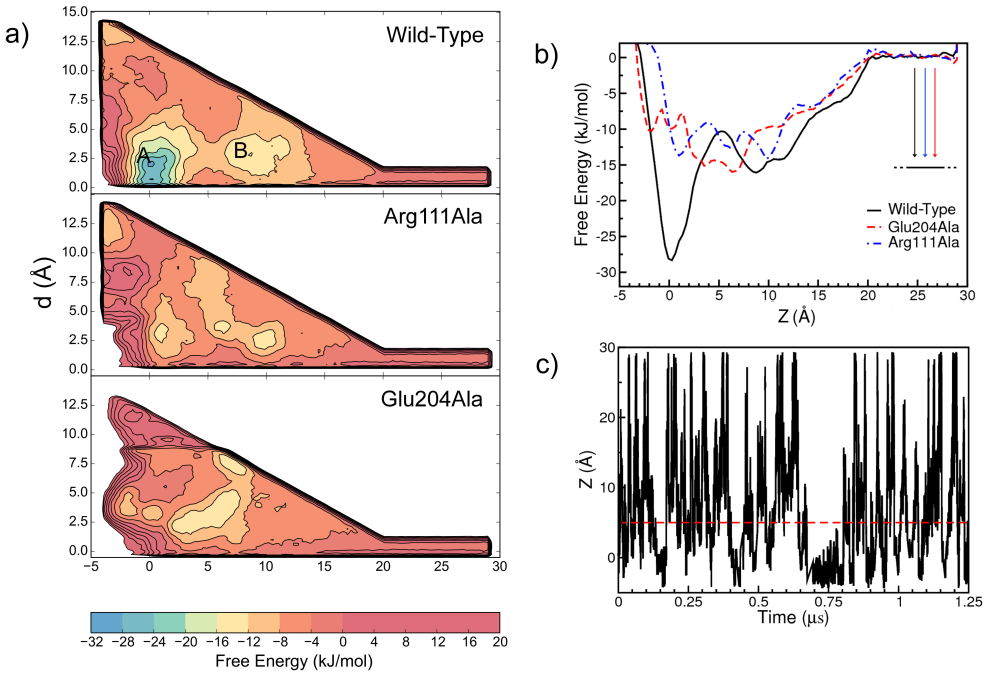


Figure 3.10: a) Free energy maps as a function of the biased CVs for GABA binding to the wild-type RDL receptor and the mutated Arg111Ala and Glu204Ala RDL receptors. The free energy of the unbound state is set as reference to zero. Contour lines are every 4 kJ/mol and all values of free energy above zero are shown in dark red. b) Projected free energy profiles as a function of the position z along the funnel. The correction due to the cylindrical part of the funnel potential, which lowers the free energy of the unbound states, is shown and indicated with vertical arrows. c) Evolution of the position z of GABA centre of mass along the funnel axis during the wild-type receptor metadynamics. The red dashed line indicates the separation between bound and unbound states. From Ref. [165].

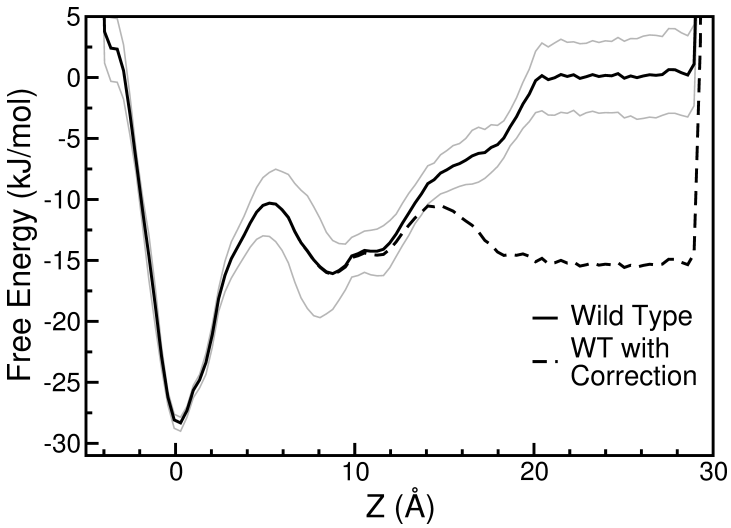


Figure 3.11: Comparison of the projected free energy profile of the wild-type RDL receptor as a function of the position z along the funnel axis with or without the funnel correction. In solid line is the original profile extracted from the simulations, while in dashed line is the corrected profile accounting for the restraining potential. In grey are the error bars for the uncorrected profile.

This lower dimensionality profiles are obtained by integrating the free energy in each point along the remaining CVs, according to the simple formula:

$$F(x) = -k_B T \; log \int_{y_{min}}^{y_{max}} dy \; e^{\frac{-F(x,y)}{k_b T}}, \tag{3.1}$$

where $F(x)$, the free energy as a function of only the variable x, $F(x, y)$ is the same free energy as a function of both x and y variables, y_{min} and y_{max} are the minimum and maximum values assumed by y, k_b is the Boltzmann constant and T is the temperature.
In the wild-type receptor the lowest free energy conformation corresponds to the minimum free energy basin A in the FES of Fig. 3.10 at z $\simeq$ 0 Å and 0 < d < 2 Å and to the global minimum at z $\simeq$ 0 Å in the one-dimensional free energy profile. Panel c) in Fig. 3.10 shows the evolution of the projection of the centre of mass of GABA onto the funnel axis during the FM simulation. Several binding (z < 5 Å) and unbinding events were reproduced, ensuring an exhaustive exploration of the phase space and a quantitatively well-characterised FES. The absolute ligand binding affinity was estimated by computing the free energy difference between the bound state in basin A and the isoenergetic states in the unbound region (with z > 20 Å). The calculated value is -29.7 ± 2.9 kJ/mol, which reduces to -14.4 ± 2.9 kJ/mol when considering the effects of the restraining potential on the unbound states. The entity of the correction to the unbound states due to the cylindrical part of the funnel potential is shown in Fig. 3.10 b) and Fig. 3.11.

To provide further structural details of the ligand binding, the FES was remapped as a function of different CVs with respect to those originally biased in the FM simulations, using the reweighting algorithm. In Fig. 3.12 the wild-type RDL receptor FES is shown as a function of the distance between the carbon of GABA carboxyl group and the C_α of Arg111 (indicated as CV_{Arg111}) and of the distance between the nitrogen of GABA amino moiety and the C_α of Glu204 (CV_{Glu204}). These CVs are somehow similar to those chosen for the exploratory metadynamics in the previous work shown in Fig. 3.8. The reweighted FES reveals two distinct energy minima in the bound region corresponding to two different binding modes, here labelled as A_1 and A_2 respectively, which are overlapped in the same basin A in the FES representation of Fig. 3.10. The evolution of (CV_{Arg111}) and (CV_{Glu204}) was monitored during the simulation, showing that two basins of minimal free energy were visited several times. The two minima are also observed in the reweighted FES as a function of the position of GABA with respect to the x,y,z Cartesian axes in Fig. 3.13, where z and x are the axes of the funnel and of the channel respectively. Here the two minima A_1 and A_2 have coordinates x $\simeq -1$ Å, y $\simeq -2$ Å, z $\simeq$ 4 Å and x $\simeq$ 0 Å, y $\simeq$ 0 Å, z $\simeq$ 4 Å respectively.

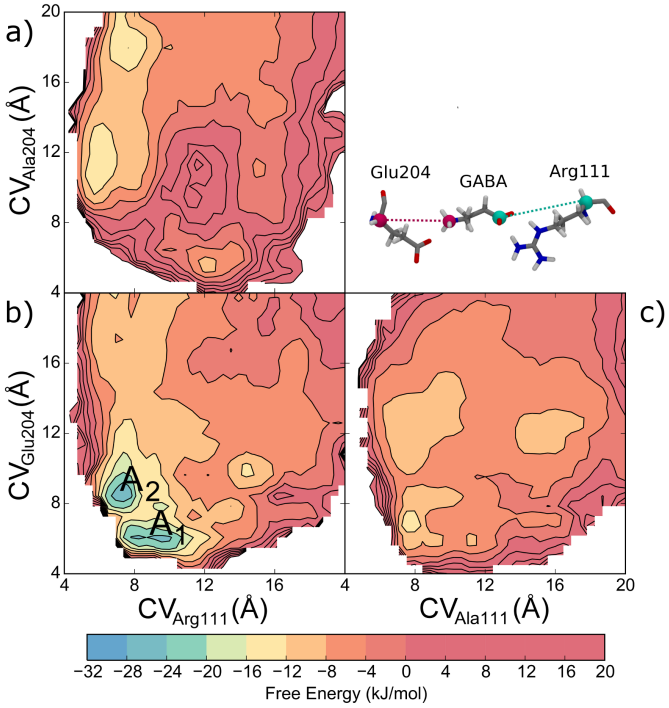


Figure 3.12: Free energy maps of a) the Glu204Ala-mutated RDL receptor, b) the wild-type RDL receptor and c) the Arg111Ala-mutated RDL receptor reweighted as a function of the distances between GABA charged groups and the C_αs of the receptor 111 and 204 residues. These distances are sketched in the top right corner in cyan and magenta. From Ref. [165].

A conformational cluster analysis of the poses representing the two minimal free energy basins A_1 and A_2 was performed. In A_1, the most populated family corresponds to the binding mode where GABA forms strong direct hydrogen bonds with both Arg111 and Glu204 as in Fig. 3.13. Specifically, the negatively charged carboxylate moiety of GABA forms salt bridges with Arg111 in loop D and interacts through multiple hydrogen bonds with Ser176 in loop E and Thr251 in loop C. The latter loop was conformationally rather flexible during the simulation, fluctuating between a closed and open conformation every few nanoseconds and allowing the access and the exit of the ligand to and from the binding pocket through its closure and opening motion. GABA amine is bound, through hydrogen bonds, to the residues of loop B Glu204 and Ser205, and can also form hydrogen bonds with the hydroxyl groups of Tyr109 in loop D and Tyr254 in loop C. The amine is buried in an aromatic cage formed by these two tyrosines and Phe206 in loop B with which it forms cation-π interactions. Consistently with this picture, in Fig. 3.12 (b), basin A_1 is located at $CV_{Arg111} \simeq 8$ Å and $CV_{Glu204} \simeq 6$ Å, with GABA tightly bound to both the primary and the complementary subunit.

The A_2 basin is observed at $CV_{Arg111} \simeq 7$ Å and $CV_{Glu204} \simeq 8$ Å in Fig. 3.12 (b). The cluster analysis of the poses corresponding to this basin showed, in the

most populated family, GABA bound to the residues of the complementary unit, i.e. Arg111, Ser176, and to Thr251 in loop C. Similarly to conformation A_1, GABA amino group is held in place by cation-π interactions with residues of the aromatic cage formed by Tyr109, Tyr254 and Phe360 and by hydrogen bonds with the hydroxyl group of the two tyrosines and Ser205. However, at variance from A_1, in this basin the interaction with Glu204 is typically mediated by a water molecule present in the binding site, and coloured in green in Fig. 3.13. As a consequence, GABA assumes a slightly different position within the binding pocket with respect to the A_1 basin, which allows the concurrent presence of a water molecule without altering the main interactions formed by the ligand with both subunits. This result is in line with what was observed in the previous MD study. Again, these simulations highlight the key role played by water molecules during the binding process, confirming the necessity of explicitly including water molecules in the simulations: water mediated interactions have been suggested for other Cys-loop receptors and analogous systems [199, 200].

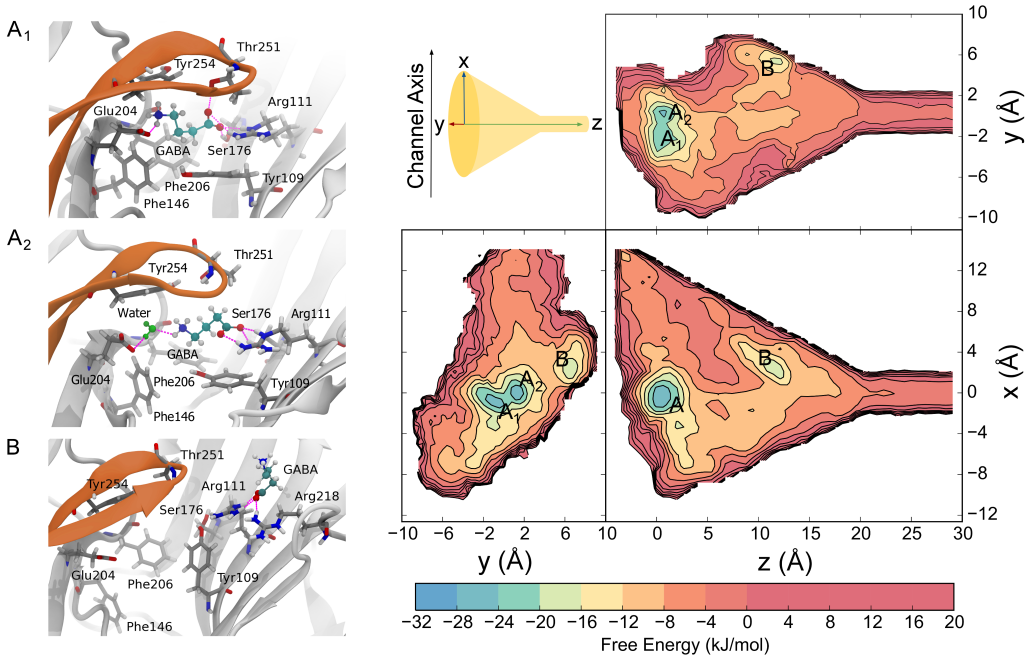


Figure 3.13: Right, binding free energy maps in the wild-type RDL receptor projected onto the funnel sections. Left, typical conformations from the basins: GABA is represented in CPK style, the C-loop is in orange and hydrogen bonds are indicated in magenta. In A_2, the water molecule that mediates the hydrogen bond interaction between GABA and Glu204 is represented in CPK style and in green. Adapted from Ref. [165].

In this representation, the difference in free energy between the two minima is $\simeq$ 1 kJ/mol, with the A_1 minimum slightly favoured: the two states are easily interchangeable at physiological conditions.

A shallow energy minimum, B, is found in the FES in Fig. 3.10 at 7.5 Å$< z <$ 12

Å and 1 Å $<$ d $<$ 5 Å, farther from the binding site and closer to the cylinder region of the funnel restraining potential. This minimum represents a prebinding pose, as also suggested by the one-dimensional free energy profile in Fig. 3.10. It is important to stress that this free energy minimum is fully within the cone section of the funnel, where no external potential is applied to the system and therefore it does not represent an artifact, as it can be seen in Fig. 3.11. The conformational cluster analysis of the poses populating this basin shows a number of different states where GABA is almost out of the binding pocket, but still able to interact with the receptor through its carboxylate moiety. Here the GABA carboxyl group forms hydrogen bonds with Arg111, whose side chain is tilted outward with respect to the channel axis, and Arg218, while its amino group points toward the solvent. This is the first time that Arg218 is found to take part in the ligand binding process and further investigations such as mutagenesis experiments might provide more details on its functional role for the receptor activation. As it can be seen in the left bottom panel of Fig. 3.13, in this state loop C is slightly open and no interaction is observed between this loop and the complementary subunit. Basin B is 12.4 kJ/mol higher in free energy than basin A. From this conformation the ligand can reach the lowest energy state A, which corresponds to the final binding mode, crossing a small energy barrier as evident in Fig. 3.10. During the transition from the prebinding mode B to the binding mode A, the ligand interacts with the residues of loop C: the subsequent closure of the loop locks it in the binding pocket.

Binding Path

The sequence of events characterising the binding path of GABA is shown in Fig. 3.14. For the sake of clarity, the RDL receptor ECD is here displayed perpendicular to the membrane surface. In the wild-type RDL receptor, the first protein residue that interacts with GABA when it is still almost fully solvated, is Arg218 in the complementary subunit. Its conformational flexibility and the positive charge of its side chain are crucial to capture the negatively charged carboxyl group of GABA from the solvent. The GABA-Arg218 interaction is maintained until the formation of hydrogen bonds between GABA and the side chain of Arg111, which also belongs to the complementary subunit. Thus, the GABA carboxylate is exchanged from Arg218 to Arg111, which acts as the first attractor of the ligand into the binding site thanks to the flexibility and charge of its side chain. These interactions with the complementary subunit are instrumental in optimizing the orientation of GABA with respect to the receptor binding pocket. In this phase, loop C has a fairly open conformation that allows the ligand to approach the binding pocket. The entry of the ligand into the binding site is also facilitated by a hydrogen bond interaction formed by the GABA carboxylate with the hydroxyl group of Thr251 in loop C.

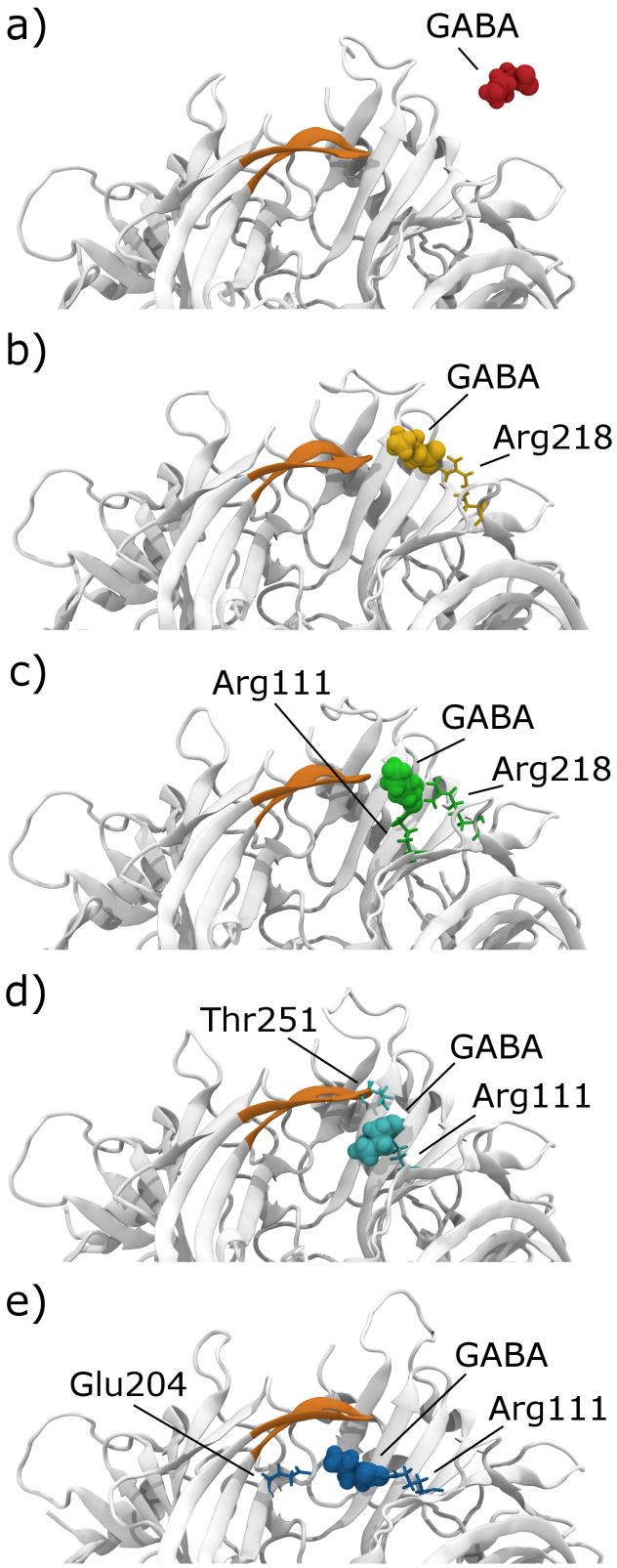


Figure 3.14: Snapshots of a favourable binding path observed in the metadynamics simulations for the wild-type RDL receptor, here seen from the membrane orthogonally to the channel axis. GABA and the residues interacting with it through hydrogen bonds are shown in colours that go from red (at the relative metadynamics time of 0 ns, a), through yellow (1.4 ns, b), green (1.8 ns, c) and cyan (2.7 ns, d), to blue (9.3 ns, e). Loop C is in orange. From Ref. [165].

The carboxyl group of GABA is the main player during this preliminary phase of the binding process, while the amino group does not actively take part in the action, remaining in an almost fully solvated state. To reach the final binding state, GABA passes under loop C, preserving, however, its interactions with Arg111 and Thr251. Upon the ligand passage, the loop closes, dragged by the ligand contact with Thr251, while the GABA amine moiety is positioned so to optimally interact with Glu204, through either direct or water-mediated hydrogen bonds, and with the surrounding aromatic residues, through cation-π interactions.

The conformational changes of loop C play a key role also in the unbinding of the ligand from the binding site. In fact, the opening of loop C favours the release of GABA followed by a partial reclosure of the loop after the ligand unbinding. This mechanism is clearly demonstrated by the time evolution of the angle (θ_{loopC}) defined by the axis of loop C and of the RMSD of the 246-255 residues C_αs (around the apex of loop C) at the start and during the simulation. Their running averages over subsets of one thousand points and the corresponding standard deviations are shown in Fig. 3.16 a) and b) respectively for the first unbinding event observed in the FM simulation; panel c) shows the motion of the loop and the relative position of GABA during the unbinding. The full closure of loop C occurs only when GABA is in the binding pocket and interacts with Thr251, as previously described.

Mutant RDL Receptor Results and Discussion

In this section, the binding free energy landscapes of two mutants of the RDL receptor are studied. Replacing Glu204 or Arg111 with non polar alanines disrupts the interaction network between GABA and the protein, preventing the ligand from maintaining a stable conformation within the binding pocket.

For the Arg111Ala-mutated RDL receptor, a few shallow minima appear in the region $0 < z < 12$ Å and $0 < d < 5$ Å of the FES and in the corresponding one-dimensional free energy profile in Fig. 3.10. The lower depth of these minima with respect to those of the wild-type receptor is also evident in the FES representation of Fig. 3.12 a) and c). No deep free energy minimum is present, thus suggesting a poor affinity of GABA for this mutant form. The calculated binding free energy, when considering the entropic cost of using the funnel potential, is close to 0 kJ/mol, indicating weak or no binding, in agreement with mutagenesis experiments that recorded non-responsive channels for this mutant form [50, 167]. From the trajectory analysis, it was observed that, when the ligand approaches the Arg111Ala-mutated RDL receptor, the first point of interaction is still Arg218, but the absence of Arg111 prevents GABA from optimally entering the binding pocket and forming strong interactions with its residues. Nevertheless, the ligand manages to form hydrogen

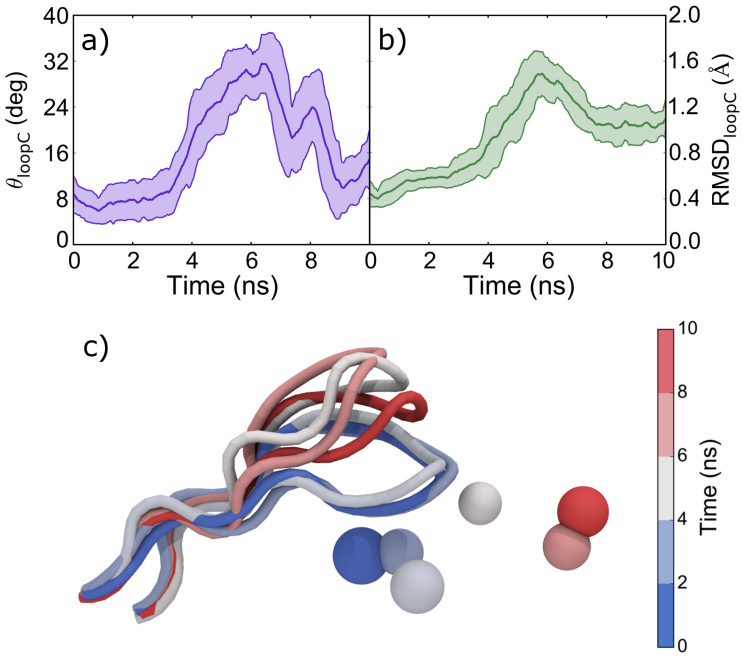


Figure 3.15: Evolution of a) the angle describing the orientation of loop C and b) of the RMSD of loop C during GABA first unbinding event in the wild-type RDL receptor calculated over the C_αs of residues 246-255. In panel c) the structure of loop C and the centre of mass of GABA, represented as a sphere, are shown every 2 ns with blue indicating the bound state and red the unbound state. From Ref. [165].

bonds through its carboxyl group with the hydroxyl group of Thr88, and through its amino group with Thr250 and Thr251 in loop C, but no cation-π interactions. These interactions give rise to the shallow minimum at $5 < z < 12$ Å in the FES of Fig. 3.10. Another shallow energy minimum is found around $z \simeq 1.5$ Å and $d \simeq 3.0$ Å. Here GABA is closer to the wild-type binding site, but the reduced and weaker interactions with the protein makes this state less stable than the corresponding conformation in the wild-type receptor. In this basin, Glu204 represents the anchor point of GABA in the principal subunit, where the amino group forms cation-π interactions with Phe206 and Tyr254. The latter is also able to engage in sporadic hydrogen bonding with the ligand. In the complementary subunit, Ser176 and Tyr109 interact with the GABA carboxylate, but these residues are farther apart from Glu204 than position 111, thus forming weaker interactions. Loop C is slightly more flexible if compared with the wild-type receptor and is unable to close completely and to lock the ligand into the binding pocket.

Similarly to what was observed for the Arg111Ala mutant, the mutation of Glu204 in loop B with alanine causes significant changes in the binding mechanism of GABA with altered interactions with the protein. The FES in Fig. 3.10 shows no energy minimum around $z \simeq 0$ Å and $d \simeq 0.5$ Å, thus indicating that in this form GABA is not able to bind in the binding pocket identified in the wild-type. A wide shallow free

energy basin is observed at $3.0 < z < 7.5$ Å in Fig. 3.10 and its depth is consistent with very weak or no binding. In this basin the GABA carboxylate is bound to the residues of the complementary subunit Arg111 and Arg218, while the amino group does not form any persistent interaction. The ligand forms hydrogen bonds in the principal subunit with Thr251 in loop C and less frequently with Tyr254. In all these states GABA remains outside the binding pocket identified in the wild-type receptor and does not form any cation-π interaction. These data suggest that Glu204 is crucial to pin the GABA amino group in the most appropriate position to engage in the cation-π interactions typical of Cys-loop pLGICs.

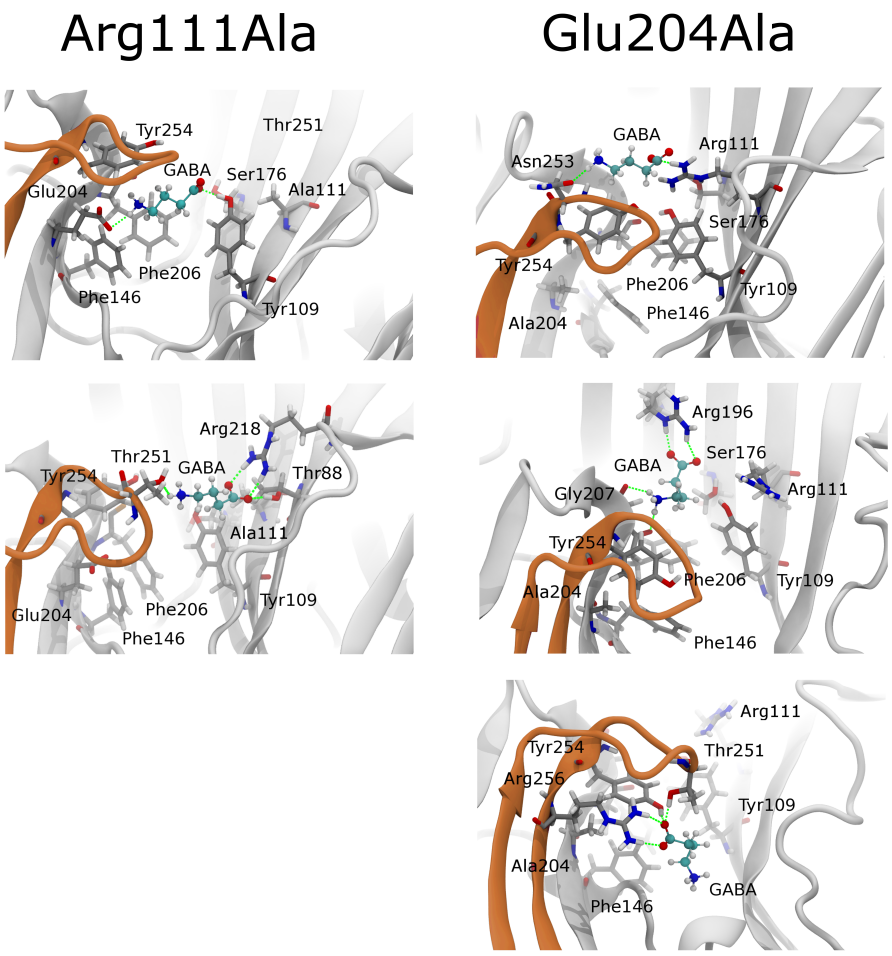


Figure 3.16: Metastable conformations observed during the binding of GABA to the mutant forms of RDL.

Conclusions

The results here presented on the GABA interacting network within the RDL receptor are in line with the picture stemming from mutagenesis electrophysiology experiments and preliminary MD simulations. Thanks to metadynamics, however, it was possible to improve this picture by extending the simulation timescale, accelerating the binding process and evaluating the binding free energy with accuracy. The sequence of events that lead the ligand to its optimal binding conformation was unveiled, starting with the capture from the solvent of its negatively charged carboxyl group by Arg218, followed by the establishment of the crucial interaction with Arg111. The active role of Arg218 as first interactor during the binding of GABA was not observed before: mutagenesis experiments could thus provide a decisive proof of the importance of this residue, e.g. by analysing the impact that substituting this arginine with a non polar alanine have on the receptor activity. Once the carboxyl group is bound to the receptor through Arg111, the positively charged amino group is able to engage with Glu204 by means of a direct or water-mediated hydrogen bond, and with the surrounding aromatic residues through cation-π interactions. Moreover, it was possible to investigate the role played by the solvent and the motion of secondary structures during the binding. In particular, the functional conformational changes of loop C, which actively participates in the binding process by opening when the ligand approaches the protein and closing after the binding was studied: this mechanism locks the ligand in the competent binding conformation for the activation of the neuroreceptor. It would be interesting to observe if tuning experimentally the flexibility of this loop, by means of selective mutations on residues maintaining its β-hairpin structure, could lead to observable changes in the binding affinity.

In the two mutant forms Arg111Ala and Glu204Ala an alteration of crucial ligand-protein interactions necessary for the binding to the wild-type receptor was observed, resulting in significant changes in the free energy landscapes and the ligand binding mechanisms. These changes explain the inactivity of the channel in the mutant forms, in agreement with experimental findings. Specifically, the mutation Arg111Ala affects the first stages of ligand binding, preventing GABA from finding a stable anchor point in the complementary subunit and entering the binding pocket with optimal orientation. While alternative interactions with residues of the complementary subunit are observed, they are not sufficient for a stable binding across the two subunits. On the other hand, in the Glu204Ala mutant form, GABA is able to initially follow the sequence of binding events found in the wild-type RDL receptor, including the interactions with Arg218 and Arg111. However, the absence of the negatively charged Glu204, which anchors the GABA amino group in the wild-type receptor, prevents the completion of the binding process. As a consequence,

the neurotransmitter interacts through its carboxylate with Arg111 and the nearby residues of the complementary subunit, while its amine remains unbound without finding suitable interaction partners.
The possibility of cooperative binding to some or all the five subunit interfaces in the RDL receptor ECD was not accounted for in this work, where the binding of a single ligand to one interface was instead investigated. As previously mentioned, the binding of two or three ligands may be necessary to activate pLGICs [11, 14, 52]. A cooperative binding mechanism may result into an additive effect for the ligand binding free energy evaluated for each subunit and potential allosteric effects among the subunits that could be worth investigating in the future.

This study represents the first example where the power of funnel metadynamics simulations was exploited to explore the different binding mechanisms of a ligand toward the wild-type and mutant forms of its molecular target. The wealth and accuracy of structural and free energy information provided by this investigation are fundamental to complement and interpret the available experimental data.
The growing availability of structural information from experiments and the improvement and reliability of computational methods like metadynamics with respect to conventional MD make these studies timely since they have the potential to greatly contribute to the detailed understanding of the activation mechanisms of these complex and important neuroreceptors. Similar simulation protocols may be applied to other pLGICs, whose structures have been recently experimentally resolved [15, 28, 30], and that share structural and functional features with the RDL receptor.
Given its efficacy, metadynamics has been thus applied in the next section to study the effects of proline isomerisation on the TMD of a more recent structure, the mouse 5-HT_3 serotonin receptor, in the context of its gating process.

3.3 Proline isomerisation in the 5-HT$_3$ Receptor

In 2014, Hassaine et al. [28] crystallised a structure of the mouse serotonin 5-HT$_3$ receptor, including the extracellular, the transmembrane and part of the intracellular domain, at 3.5 Å of resolution, opening the way for more accurate simulations of pLGICs than those based on homology models. Part of the original structure is shown in Fig. 3.17. From the X-ray coordinates, a model of the full receptor, solvated and embedded in a lipid bilayer mimicking the cellular membrane, was built here and used for a thorough study of the structural and dynamical properties of the TMD regions involved in the gating mechanism and is shown in Fig. 3.18.

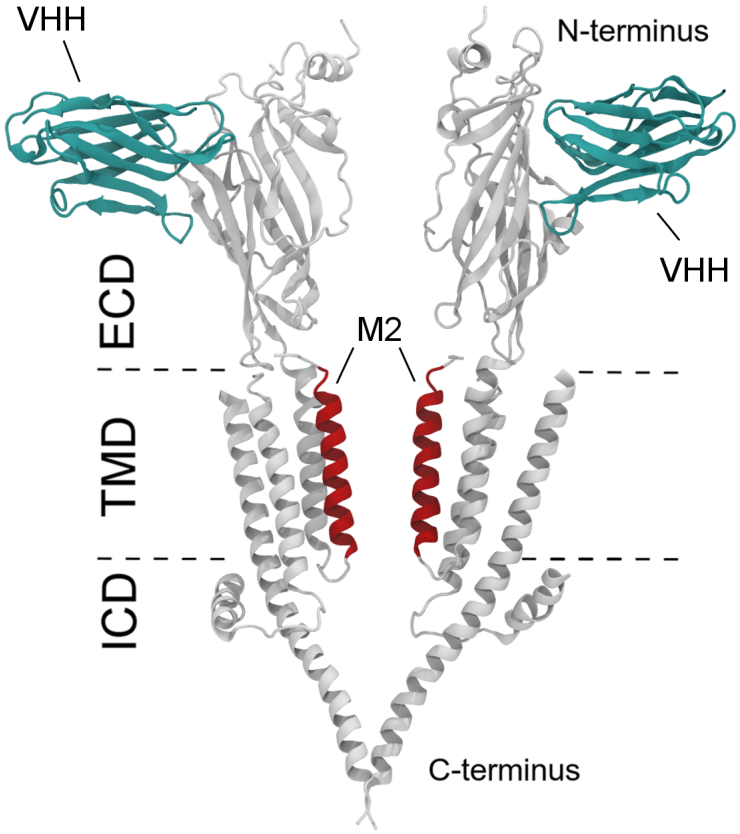


Figure 3.17: Two non-consecutive subunits taken from the original 5-HT$_3$R crystallographic structure. Here, two VHH chaperon proteins are also shown in cyan, while the M2 helices lining the ion permeable pore are highlighted in red. The alleged position of the membrane is marked with dashed lines.

The structure has cylindrical extracellular and transmembrane domain with an architecture similar to ELIC, GLIC and GluCl. Llama-derived single-chain antibodies (or VHH), were used as chaperons to stabilise the structure during the crystallisation process: these molecules were positioned radially, bound at the interface of two subunits at the level of the serotonin binding pocket. Being crystallised in the presence VHH, which act as inhibitors, the structure is be expected to be in a closed

state; however, the M2 helices lining the pore in the TMD are similar to those observed in allegedly open GLIC and GluCl structures. External compounds used to help crystallisation, such as invermectine, have been shown to induce desensitised states [11] and too many dissimilarities from GLIC, GluCl and other pLGICs structures in the TMD make it difficult to identify unequivocally the state of the receptor.

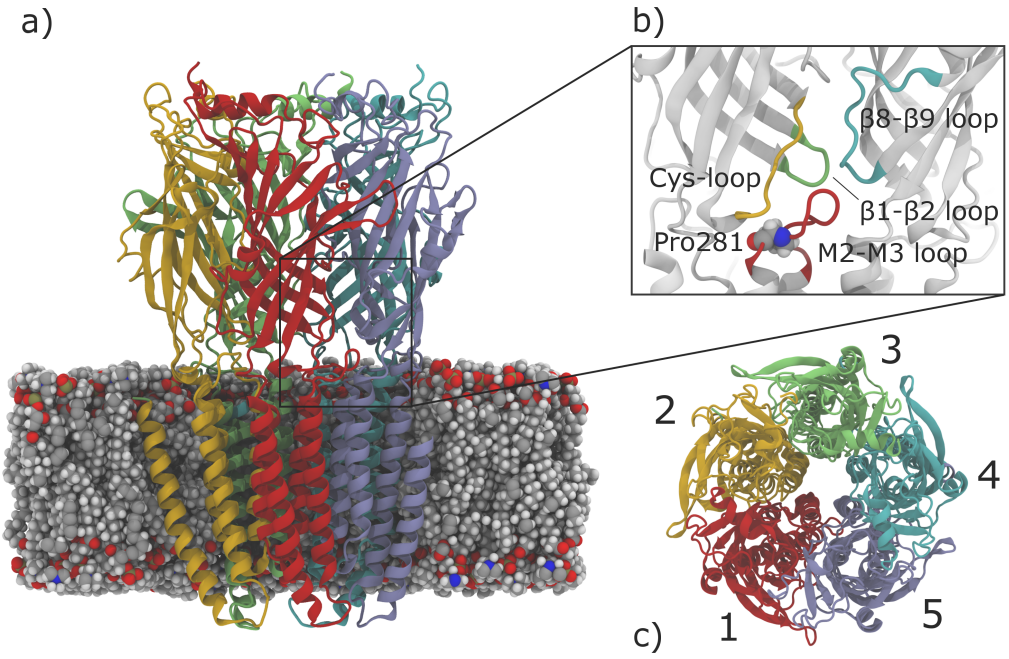


Figure 3.18: a) Section of the 5-HT$_3$R model used for this project. The receptor is represented in colors from red to blue embedded in the POPC lipid bilayer, shown in Van der Waals representation; to simplify the representation, the orthorhombic solvation box has been omitted. b) Zoom of the interface, where the loops responsible for the information transfer are highlighted in colors: in yellow is the Cys-loop, in green is the $\beta 1 - \beta 2$ loop, in blue is the $\beta 8 - \beta 9$ loop (or loop F) and in red is the M2-M3 loop. The highly conserved Pro281 is shown in Van der Waals. c) View of the receptor model from top, where the anticlockwise labelling of the subunits is shown.

The interface between ECD and TMD is characterised by the presence of a number of loops, both in the principal and in the complementary subunit, which are expected to play a major role in the transmission of the activation signal from the ECD, where the binding site sits, to the pore-lining helices of the TMD. These loops can communicate with direct hydrogen bonds or electrostatic interactions, or indirectly through interactions with the surrounding water molecules and lipids. Despite their undeniable role, how they communicate is still unclear. As shown in Fig. 3.18 b) and in Fig. 3.19, among these structures are the Cys-loop, which contains the disulphide bridge that gives the name to the superfamily of these proteins, and the $\beta 1 - \beta 2$ loop, both in the ECD of the principal subunit, the $\beta 8 - \beta 9$ loop (also known as F loop) in the ECD of the complementary subunit and the M2-M3 loop, in the TMD of the principal subunit. The M2-M3 linker connects the M3 helix, which is partially exposed to the lipid membrane, and the M2 helix, which lines the ion

permeable pore. It is thus a crucial element in the gating process as confirmed by the presence of highly conserved residues in its sequence. Mutagenesis experiments on these residues produced partially- or non-functional channels, whose binding properties were however untouched; these studies helped define the hypothesis of Pro8* *trans-cis* isomerisation as a potential molecular switch for the 5-HT$_3$R gating. Through unnatural amino acid mutagenesis, the so-called Pro8*, that sits at the top of the M2-M3 loop, was substituted with a set of proline analogues that favoured the *cis* isomer to varying degrees. As a result, non-functional channels were produced when the proline was substituted with residues strongly favouring the *trans* conformer [1]. Moreover, a strong correlation between the *cis-trans* energy difference of the analogues and the activity of the channels was observed among functional proteins and confirmed by metadynamics studies on proline analogue dipeptides [62]. From these result it is straightforward to suggest the molecular rearrangement of Pro8* as the key mechanism that opens and closes the pore. This theory however has not been spared from criticism, mainly focusing on the improbability that a relatively small change in a single amino acid can lead to larger tertiary movements, such us the rigid translation of the M2 helices. It is more plausible to assume that the *cis-trans* isomerisation is just one, but crucial, element in a more complex action within the M2-M3 linker.

Through the amino acids of the M2-M3 loop, the activation signal is likely to be translated into conformational changes that affect the M2 α-helix structures lining the ion permeable pore, where residues like Leu260 (at position 9' in the M2 helix) are responsible for hydrophobically occluding the pore and hindering the flux of ions and water molecules through the TMD in the resting state.

In the work presented in this chapter, a set of three exploratory molecular dynamics simulations of the 5-HT$_3$R model were carried out for 250 ns with the aim to investigate the *trans-cis* isomerisation of Pro8* (Pro281 in our model) within the receptor environment and its effects on the structure and dynamics of the M2 helices. Metadynamics was exploited to induce the *trans-cis* isomerisation of Pro8* at various stages in different subunits of the receptor, allowing us to examine the resulting effects in the TMD. Furthermore, 500 ns of metadynamics were performed in order to map the free energy landscape of the proline isomerisation in the M2-M3 loop and compare it with a control system of a proline dipeptide in aqueous solution, so to assess how different environments affect the conformer populations and the isomerisation propensity. This is one of the first complete models of the 5-HT$_3$R reported in literature, and the introduction of the environment represents a major step up from previous MD and metadynamics studies, where reduced models of proline dipeptides and M2-M3 helices complexes were used instead [48, 62].

Simulation Details

The 5-HT$_3$ receptor X-ray structure found in the Protein Data Bank (PDB entry: 4PIR) is missing four residues located at the top of the M2-M3 linker. The reason behind this is the high flexibility of loops, which makes them difficult to reconstruct from crystallographic data. The missing residues (Thr-Ala-Ile-Gly) are highlighted in the sequence of Fig. 3.19 and precede the highly conserved Pro281; they were manually added with AmberTools 14 and adjusted with a quick spatial relaxation, their rotamers were then checked with NQ-flipper, a web based server for the analysis and refinement of rotameric states [201, 202]. The structure was cleaned of all the stabilising agents and not native fragments; moreover, the terminal MA and MX helices were removed. According to Ref. [28] these two structures make up the intracellular domain of each subunit; however a wide loop section connecting these two helices was removed by proteolysis, a procedure applied to increase the resolution of the diffracting X-rays by breaking down the protein into smaller polypeptides, and was not resolved. ICDs are commonly removed during the expression of pLGICs as this does not seem to affect their structure, though their actual role during protein activation is still unclear. Most studies seem to indicate that they are involved in late stages: a ring of highly varying polar residues observed in open-channel structures is believed to act as a final selectivity filter for the ions, which then leave the pore through lateral windows [11, 203–205]. Adding this missing sequence would have required a meticulous reconstruction, with the risk of very approximate results. Hence, we chose to remove the incomplete ICD in its entirety, focusing instead on the interplay between ECD and TMD. M3 and M4 were thus connected directly through a short loop, with sequence shown in Fig. 3.19, in accordance with what observed in ICD-lacking structures like GluCl and ELIC [202].

The reconstructed protein was subject to standard protonation, inserted in a homogeneous 1-palmitoyl-2oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer membrane with the help of CHARMM-GUI membrane builder [206] and solvated with an orthorhombic 12 Å buffer of TIP3P water and Na^+ and Cl^- ions to reproduce physiological conditions with salinity of 0.15 M, following the protocol used for the RDL receptors in Sec. 3.1. The complete system without solvent is shown in Fig. 3.18: it contains 183,360 atoms and was simulated with NAMD 2.9, the AMBER ff14SB force field [207] and LIPID14 [208]. The latter is a recently developed force field specifically designed to simulate lipids consistently with the AMBER force field parametrisation. With respect to previous versions, a re-fit of charges and torsional parameters brought a number of improvements that now allow to simulate membranes without the need to apply an external tension: LIPID14 properly reproduces properties such as area per lipid, volume per lipid, bilayer thickness and give results consistent with experimental data [208]. As in the case of proteins, the choice of

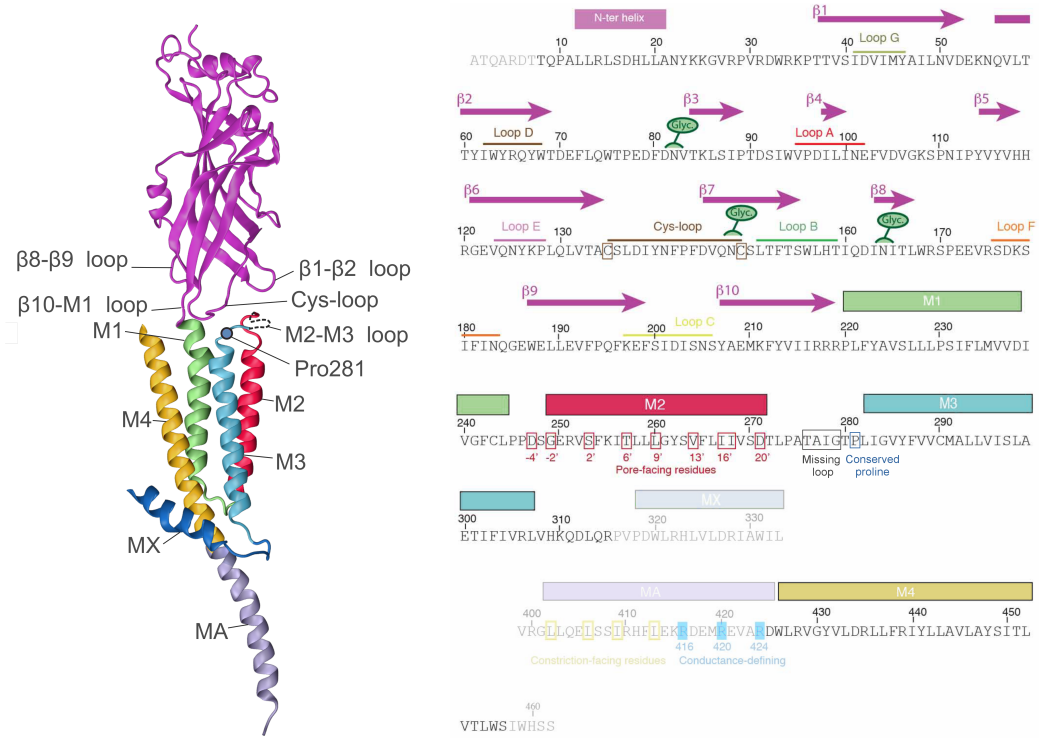


Figure 3.19: On the left, a single subunit of the unmodified X-ray structure of the serotonin mouse receptor 5-HT$_3$R (4PIR entry in the Protein Data Bank). Here the different helices of the transmembrane and intracellular domain are highlighted. On the right, the amino acid sequence of the same structure: the components that were subsequently eliminated in the model used in the simulations are greyed. (adapted from Ref. [28])

an appropriate lipid force field is crucial; in spite of many improvements, studies have shown that a force field dependence of structural and dynamical properties of membranes is still present even for the most widely used lipids such as POPC [106]. In the case of 5-HT$_3$R, LIPID14 was chosen for consistency with the AMBER force field used to describe the receptor. The same protocol used with the RDL receptor was followed for the molecular dynamics parameters: the time step was kept at 2 fs thanks to the use of SHAKE to constrain the bonds containing hydrogen, Particle Mesh Ewald was employed for the electrostatic interactions and a cut off of 10 Å was introduced for non-bonded interactions. A series of MD simulations were carried out within the isothermal-isobaric NPT ensemble: the temperature was maintained at 300 K with a Langevin thermostat with a collision frequency of 1.0 ps^{-1}, while the pressure was kept at 1 atm with a Langevin piston barostat with an oscillation period of 200 fs and a damping timescale of 100 fs. A structure optimisation procedure was followed by a slow heating and a partially restrained equilibration, which lasted for 3.5 ns [202].

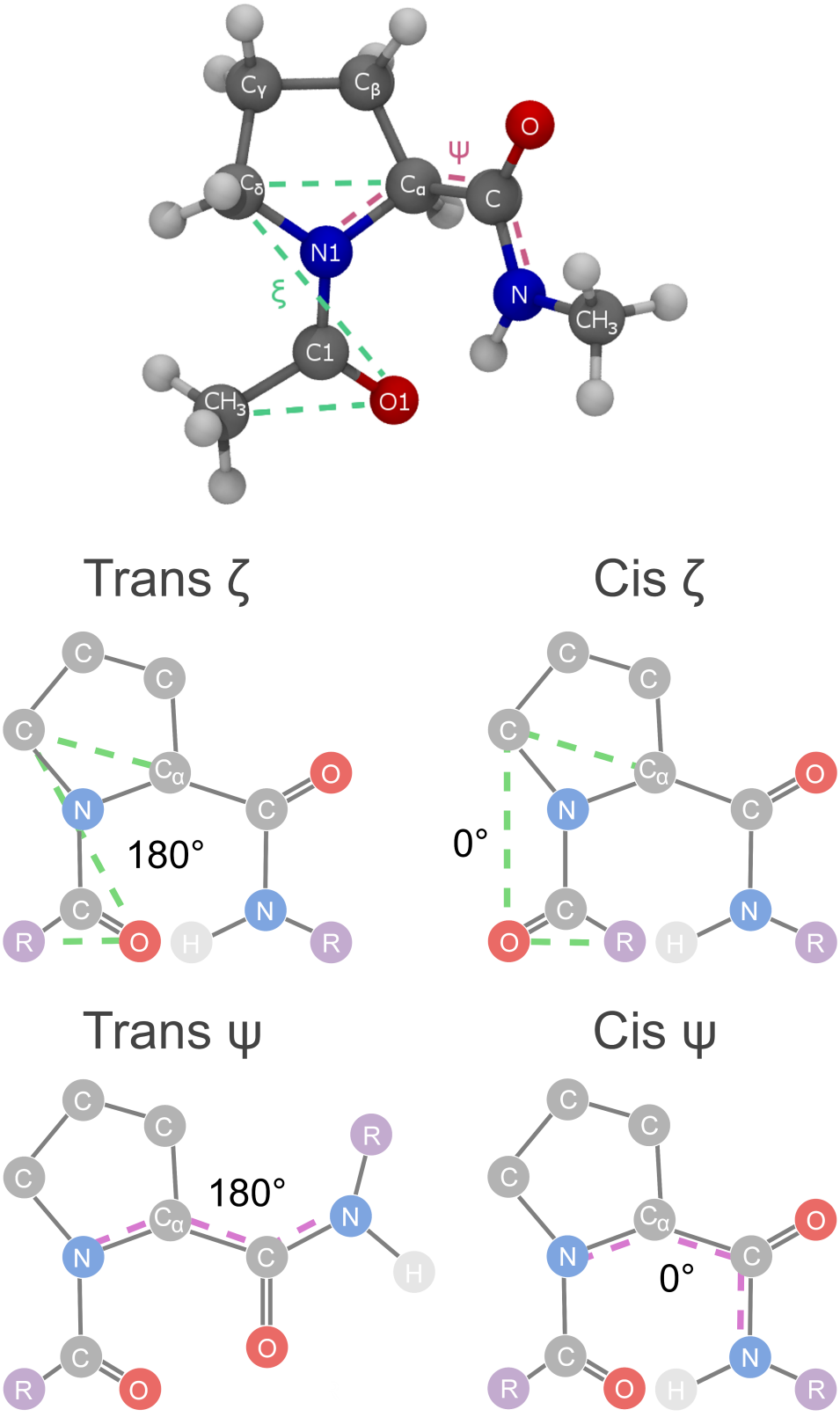


Figure 3.20: Top, proline dipeptide with the atoms relevant to define the metadynamics CVs labelled. The torsional angles ξ and ψ are highlighted in green and magenta respectively. Bottom, schematic representation of *trans* and *cis* angles for both ξ and ψ. Here, carbons are in dark grey, hydrogens in light grey, oxygens in red, nitrogens in blue and in violet are the backbone connections with the rest of the protein (R).

Metadynamics simulations were also run with the same NPT ensemble to study the isomerisation of the highly conserved Pro281 at in the M2-M3 loop: Plumed 2.0 plugin was used in this case together with NAMD 2.9. As already mentioned, the choice of working with this specific proline is based on previous mutagenesis electrophysiology experiments and computational studies to assess the validity of the molecular switch hypothesis [1, 62].
In order to drive the isomerisation of the proline, two angles are necessary: the improper dihedral angle ξ (defined by the atoms CH_3-O1-C_δ-C_α) and the torsional angle ψ (defined by N1-C_α-C-N) as in Fig. 3.20. While ξ accounts for the *cis-trans* isomerisation and the pyramidalisation (that is the distortion of a trigonal planar geometry toward a tetrahedral geometry) of the imide nitrogen (N1), ψ is required to control the C-terminal amide orientation, which may affect the rate of transition between the *trans* and *cis* conformations [62]. These metadynamics results in the receptor environment were compared to those of a control system: a proline dipeptide in water, which was modelled in a truncated octahedron periodic cell with a 12 Å of solvent buffer. The well-tempered regime was used, the initial Gaussian height was 0.8 kJ/mol and the width was 9°, while the reference temperature was 300 K and the bias factor was set to 10. The frequency of deposition was a Gaussian every 1 ps. While only 40 ns were needed to converge the free energy map of proline dipeptide in water, a 500 ns long metadynamics simulation was carried out in the case of the 5-HT$_3$ receptor.

Environmental Effects of Proline Trans-Cis Isomerisation: Results and Discussion

In this section, the effects of Pro281 isomerisation on the surrounding residues are analysed. Conformational changes in the M2-M3 loop have been observed upon isomerisation, leading to an increase in interactions between lipids heads and Thr280. This suggests an active role of the membrane in the gating process. Moreover, the isomerisation free energy landscape shows how, within the serotonin receptor environment, the barrier between cis and trans conformers decreases and a preference for an anti-clockwise isomerisation arises, thanks to the obstruction of residues from loops of the ECD.

To understand the effects of Pro281 isomerisation on the TMD of 5-HT$_3$R and especially on the helices lining the pore, a set of MD simulations were carried out under different conditions. Together with an MD of the original structure with all five Pro281 in the *trans* conformation (here called 5Trans), two other MD simulations, with the *cis* isomers of Pro281 in three non consecutive subunits (3Cis) and in all five of them (5Cis) respectively were performed, as it is still unclear how many ligands

(and consequent conformational changes in the respective subunits) are required to fully activate the opening of the serotonin receptor pore. The isomerisations were induced by subsequently biasing the ψ and ξ CVs of the selected prolines in the different subunits with metadynamics by using fixed height Gaussians at 0.8 kJ/mol deposited every 1 ps. To speed up the transition, and since there was no interest in recovering the free energy landscape at this stage, the well-tempered regime was not applied. After an isomerisation event was observed and the *cis* conformation was stable enough, the bias was removed and the procedure was repeated for the next subunit. To monitor the transition of the prolines the sigmoid function of Eq. 3.2 was devised to clearly distinguish the *cis* and the *trans* conformations.

$$s(\xi) = \frac{1}{1 + e^{\sigma(|\xi| - \xi_0)}} \tag{3.2}$$

The parameter ξ_0 defines the origin, the value of the angle at which the function assumes half value (0.5) and was here set at 90°, corresponding to the position of the barriers separating the isomers. On the other hand, σ affects the slope of the function and was chosen to be 0.15. The sigmoid function was applied to ξ as the primary angle of the isomerisation, and assumed a 0.0 value when the amino acid was in the *trans* state and 1.0 when in *cis*.

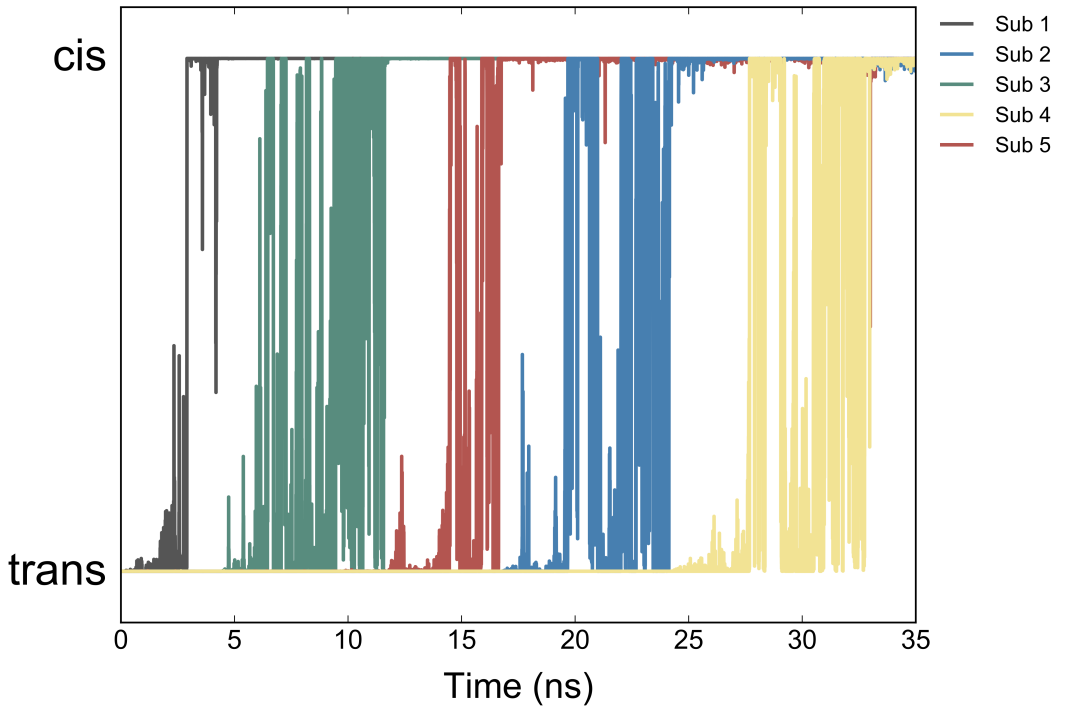


Figure 3.21: The subsequent *trans-cis* isomerisations of Pro281 in the five subunits of the 5-HT$_3$ receptor described through the sigmoid function of Eq. 3.2

A window of at least 5 ns was given to each metadynamics run, to assure that the system overcame the barrier and stabilised properly in the *cis* minimum. According to the counterclockwise numbering (see Fig. 3.18 b)), the subunits number 1, 3 and 5 were first biased in this order to build the 3Cis system, followed by subunit number 2 and finally 4 in the 5Cis protein. Within 35 ns all five subunits underwent the isomerisation as shown in Figure 3.21.
The three systems were then simulated for a 250 ns production run under identical conditions, the only difference being the *trans-cis* conformation of the five Pro281. These relatively short simulations can give us a good idea of the behaviour of the protein upon isomerisation, but in order to support the results here presented, it will be important to increase the statistics, e.g. by repeating independent simulations of these three systems a number of times. The RMSD of the backbone atoms in the secondary structures was calculated with respect to the initial minimised coordinates, and is shown for each distinct subunit in Fig. 3.22, where 5Trans includes the initial 3.5 ns of equilibration. In this case, after a few nanoseconds, the RMSD stabilises around a fixed value and assumes a flat profile which is then maintained across all the simulations.

Structural Stability and Membrane Density

The average RMSD per subunit is 1.7 ± 0.2 Å for 5Trans, 1.8 ± 0.2 Å for 3Cis, and 1.9 ± 0.2 Å for the 5Cis system, where a minor increase is observed. This is mainly due to the subunit 5, whose single RMSD increased from 1.8 ± 0.1 Å for 5Trans, to 2.1 ± 0.1 Å for 3Cis and finally to 2.3 ± 0.1 Å for 5Cis. The tertiary structure of the protein is however well maintained, especially in the secondary structures, which do not undergo significant conformational changes throughout the simulations.

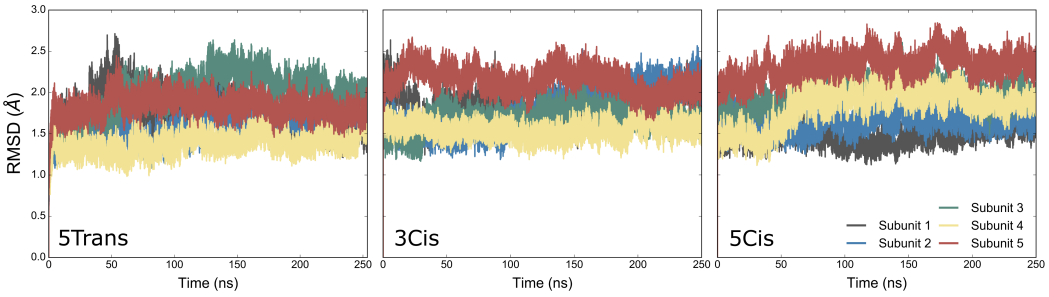


Figure 3.22: Root Mean Square Displacement of the secondary structures of the five subunits of the 5-HT$_3$ receptor during the 5Trans (a), 3Cis (b) and 5Cis (c) simulations. In all three cases, the initial minimised structure is kept as reference.

The structural integrity is also demonstrated in Fig. 3.23, where the average density of the lipids surrounding the TMD is shown for the three simulations. The results have been divided for the bilayer's hydrophilic phosphatidylcholine heads (on the left) and the hydrophobic tails (on the right). While at first sight no major

differences in the lipid distribution are observed, and the shape of the protein in the centre is well maintained among models, a second read can give an interesting insight in the properties of this protein's transmembrane domain.

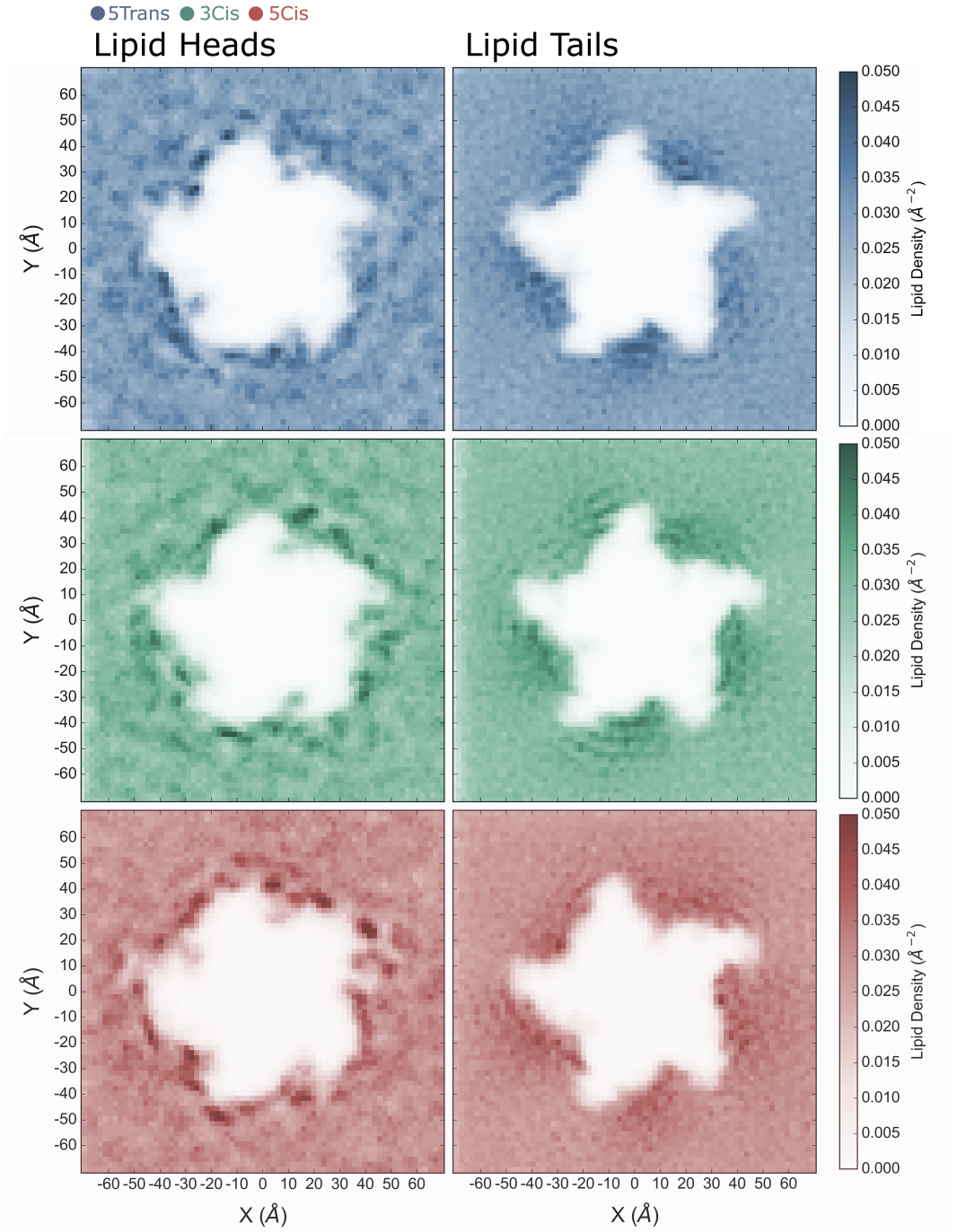


Figure 3.23: Average density of the lipid heads (left) and tails (right) during the three 250 ns long MD simulations of the 5-HT$_3$ receptor. Here XY is the plane of the membrane bilayer. For each system, the centre of mass of the carbons in the tails and of the heavy atoms in the heads of the lipids surrounding the 5-HT$_3$ receptor were used in the bi-dimensional histogram.

On the right side, no significant differences are observed in the density histogram of their hydrophobic tails for the three simulations. Interestingly, their behaviour is similar to that typically observed in the case of liquid solvents, with "ripples" associated with coordination number shells in the vicinity of the protein, slowly dampened to a uniform distribution farther from the molecule. The region of highest density, which can reach up to 0.05 Å^{-2} are found near the corners between the M1 and M3 helices in the complementary and primary subunits respectively. This can be easily understood in terms of the amino acid sequences of these two structures, whose sides facing outward, toward the membrane, are characterised by strongly hydrophobic residues, mostly Leu, Ile, Val and less commonly Phe. Similarly, the side chains of the hydrophobic residues of the M4 helices, which reach the farthest points from the channel centre in the TMD, forming the "tips" at the star shaped holes in Fig. 3.23 (right), are mostly directed at the sides, rather than outward, adding to these higher density regions. These results are in line with observations found in literature, where in a few cases POPC hydrophobic tails were found wedged between the M1 and M3 helices of crystallised Cys-loop receptors [15, 209]; this discovery led to the formulation of the hypothesis that this behaviour may actually influence the shape of the receptor up to a point that ensues an easier binding in the ECD. In Fig. 3.24 an example of POPC lipid between the M1 and M3 helices in the 5-HT$_3$R system is shown.

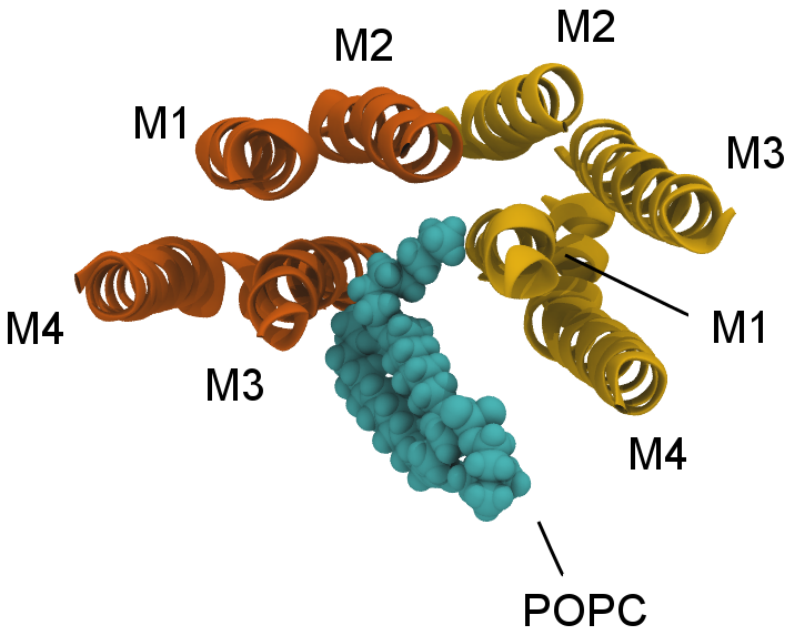


Figure 3.24: POPC molecule wedged between the M2 and M3 helices of the TMD of 5-HT$_3$R as seen from the top of the protein. In cyan is the lipid molecule, in orange are the four helices of the principal subunit and in yellow are the helices of the complementary subunit. The snapshot was taken from the trajectory of the 5Cis model.

Looking at the results on the left side of Fig. 3.23, a slight increase in the density of the lipid heads in proximity of the protein is observed in the 3Cis model and even more in the 5Cis one. The hydrophilicity of the phosphatidylcholine heads derives from the presence of a negatively charged phosphate and a positively charged choline group. Their polarity also allow for interactions with the residues at the top and and bottom of the TMD helices and even within the loops connecting them. While Ser and Thr are found at the top end of M4 helix, a series of arginines, whose side chains are facing outward, are found in its lower section. Similarly, Cys, Arg, Asp and His residues are found at the bottom of the M1 and M3 helices, while at the other extreme, the lipid heads are more likely to interact with residues found in their M1-M2 and M3-M4 connecting loops. This distribution of hydrophilic and hydrophobic residues is in line with what observed in other transmembrane proteins and assures the correct positioning of the structure within the lipid bilayer, but beyond that, it is important to consider whether interactions with the bilayer heads can influence the electrostatic environment of the M2-M3 loop and possibly affect the propensity of proline isomerisation. As already mentioned, an increase in the number of spots with up to 0.05 Å^{-2} in lipid head density is observed for the 3Cis and and the 5Cis system respectively. In the latter case case, the protein seem to be consistently bordered by such higher density regions. Although the magnitude of this increase is relatively small, the possibility that more interactions with the polar heads are formed due to a rearrangement of the loop residues in the presence of *cis* Pro281 conformers should not be excluded.

ECD-TMD Interface Interactions

To further assess this point, an analysis of the interactions between protein and lipids and/or waters in the ECD-TMD interfacial region was carried out, together with an intraresidues hydrogen bond analysis. In the resulting graphs, hydrogen bonds were defined as having a donor-acceptor distance smaller than 3.5 Å and an donor-hydrogen-acceptor angle larger than 120°.

In Fig. 3.25 the average number of hydrogen bonds formed between the residues of the interface loops and the solvent in the three systems investigated is compared. No notable change in these interactions is observed for the three loops of the ECD, $\beta1-\beta2$ loop (in green), Cys-loop (in yellow) and $\beta8-\beta9$ loop (in blue), where the main interacting residues are negatively charged amino acids such as Glu53, Asp138, Asp177 or Glu186, which maintain a good solvation throughout the simulations and form on average between 2.7 and 4.9 hydrogen bonds per simulation. Notable, although relatively small, differences are observed instead in the M2-M3 linker, which is directly affected by the isomerisation of Pro281. The number of hydrogen bonds formed by Asp271, the strongest interacting residue in these loops, decreases from an average of 5.4 in 5Trans and 3Cis to 4.4 in 5Cis. Much smaller is the magnitude

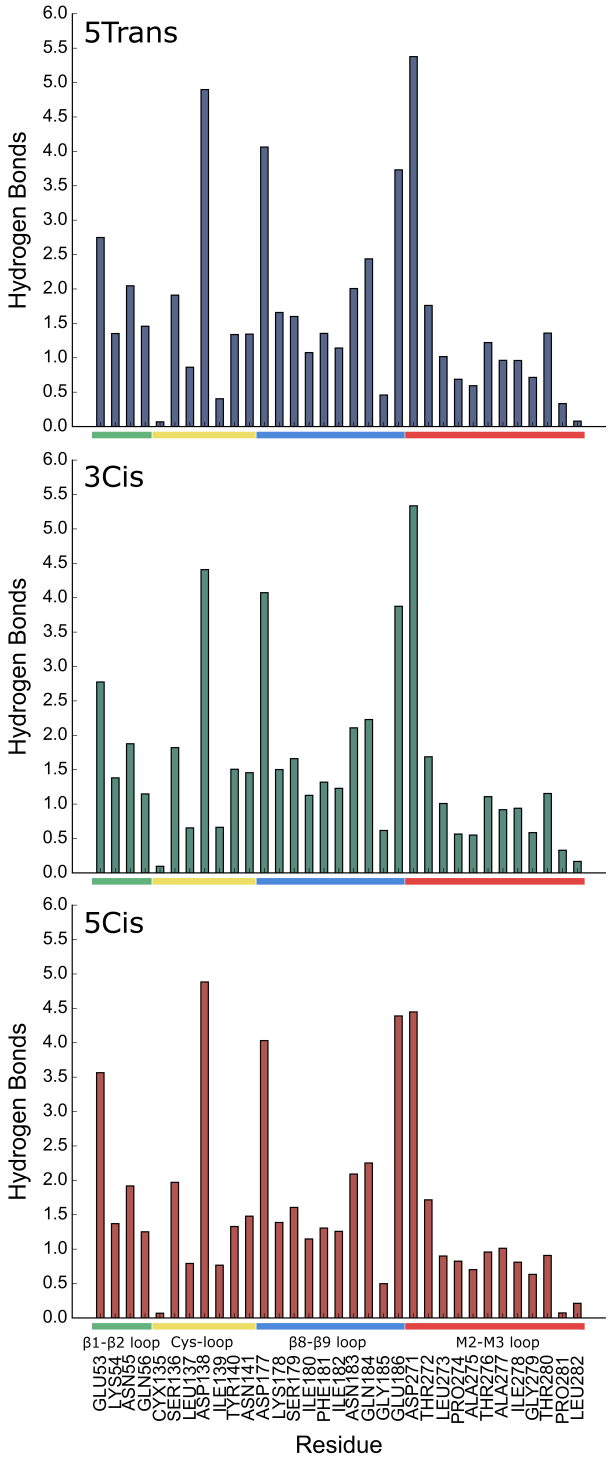


Figure 3.25: Number of hydrogen bonds observed between the residues belonging to the interface loops and water molecules surrounding them during the MD simulations of the 5-HT$_3$ receptor. The results are averaged over the five subunits and grouped for the three models 5Trans (dark blue, top panel), 3Cis (dark green, middle panel) and 5Cis (dark red, bottom panel). Residues belonging to different loops are underlined in different colors: $\beta 1 - \beta 2$ loop in green, Cys-loop in yellow, $\beta 8 - \beta 9$ loop in blue and M2-M3 loop in red.

of changes in the rest of the M2-M3 loop amino acids, where almost no differences are observed between the 5Trans and the 3Cis models, with changes in the average hydrogen bond numbers smaller than 0.5 moving to the 5Cis system. Almost no hydrogen bonds are formed by Pro281, which forms a single interaction through its backbone oxygen with an occurrence of 30% of the simulation time in the 5Trans model, decreasing to less then 10% in the 5Cis model. Hence, conformational changes in the M2-M3 loop that may be induced by the isomerisation do not seem to affect the solvent exposure of the interface by a significant degree.

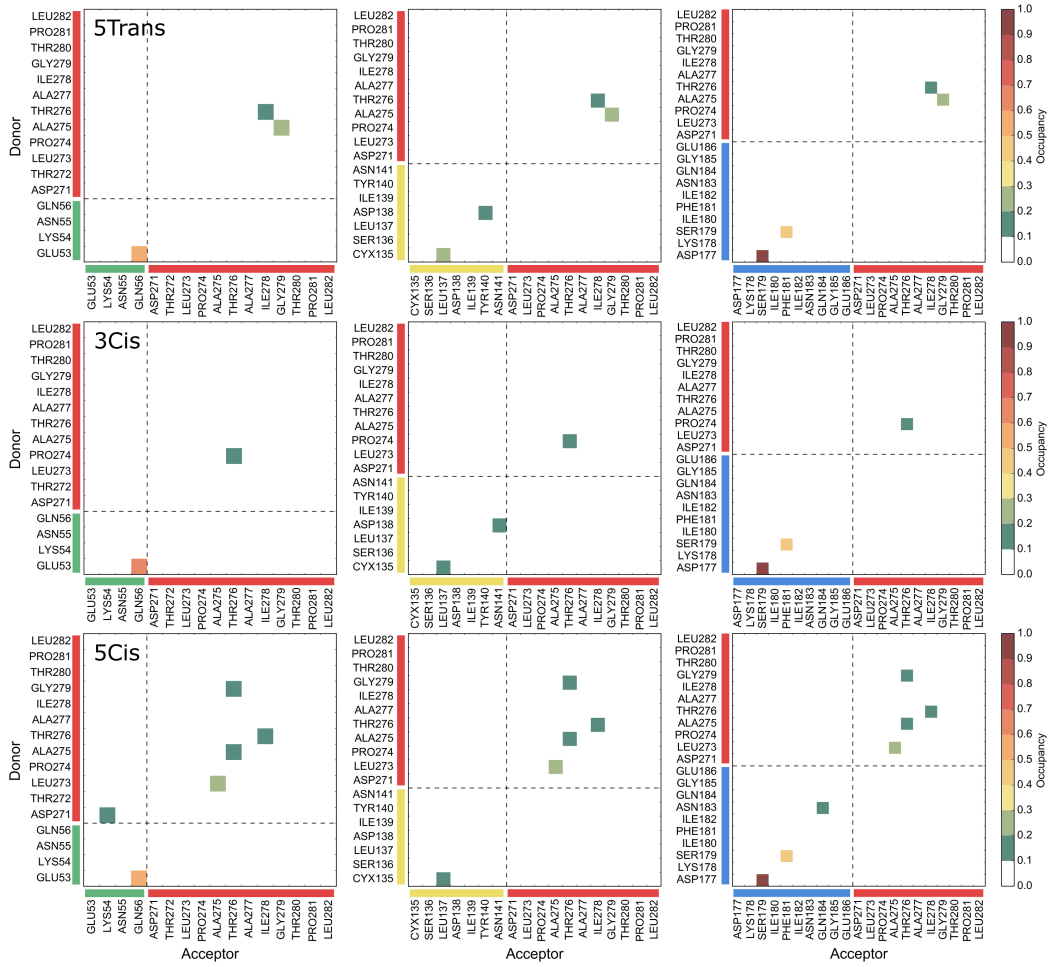


Figure 3.26: Tables showing the presence of interloop and intraloop hydrogen bonds between the residues of the M2-M3 loop (underlined in red) and the residues of the loops belonging to the ECD, namely the $\beta1-\beta2$ loop (left tables, underlined in green), the Cys-loop (central tables, in yellow) and the $\beta8-\beta9$ loop (right tables, in blue) during the MD simulations of the 5Trans model (top row), the 3Cis model (central row) and the 5Cis model (bottom row). The occupancy goes from green (0.1 to 0.2) to dark red (0.9 to 1.0); interactions with an occupancy of less than 0.1 are not shown. The results are averaged over the five subunits of each model.

Similarly, no significant change in hydrogen bond interactions between residues of the interface loops are observed. In Fig. 3.26, tables showing the occurrence (from 0 to 1) of hydrogen bonds between the M2-M3 loop and the three ECD loops are

shown, while averaged over the five subunits of the three models here investigated. The first information that catches the eye is the almost complete absence of hydrogen bond interactions between different loops. With the exception of a bond between the carboxyl group of Asp271 in the M2-M3 linker and the amino group of Lys54 in the $\beta 1 - \beta 2$ loop in the case of the 5Cis model (bottom left corner of Fig. 3.26), no other interloop interaction is present for more than 10% of the time on average: the conformational rearrangement of the M2-M3 loop following the *cis-trans* isomerisation of Pro281 does not affect these numbers.
Intraloop interactions among residues belonging to the ECD do not change: these bonds, which for the most part play a structural role, help maintaining the shape of the loops. An hydrogen bond between Glu53 and Gln56, present between 50% and 70% of the time, keeps the curvature at the top of the $\beta 1 - \beta 2$ hairpin. Similarly, the Cys-loop is characterised by the weak presence of hydrogen bonds between some side chains and backbone atoms of a few residues, such as Cys135 (also responsible for the disulphide bond in this protein) and Leu137 or, less consistently, Asp138 and Tyr140, although these interactions never overcome 30% of occurrence. The only interaction which is consistently present during the whole simulations is formed between the negatively charged Asp177 and the hydroxyl of Ser179 at the beginning of the $\beta 8 - \beta 9$ loop. Ser179 also acts as donor when interacting with the backbone of Phe181, forming a bond which is maintained between 50% and 60% of the simulation time. Sporadically, Asn183 and Gln184 have also been observed forming hydrogen bonds.
Relatively small are also the differences observed in the M2-M3 intraloop interactions, which can give an insight in the conformational changes that this loop undergoes. In the 5Trans model, the loop is characterised by a curvature in its middle, maintained by weak backbone interactions between Thr276 and Ile278 and between Ala275 and Gln279. A similar behaviour is observed in the 3Cis simulations: in this case Thr276 weakly interacts with the backbone atoms of Pro274. In the 5Cis model, Thr276 bonds with both Ala275 and Gly279, moving the curvature down the loop, and with Ile278 as acceptor. Less consistent bonds are formed between Ala275 and Leu273.

Conformational Changes in the M2-M3 loop and Lipid-Protein Interactions

In Fig 3.27 the highly flexible M2-M3 loop is shown in exemplary snapshots taken from the three MD simulations. Here the conformational change that the loop undergoes during the isomerisation of Pro281 is clear. Moreover, the residues are coloured according to the difference from the average Root Mean Square Fluctuations (RMSF) of the loop backbone, allowing to highlight which residues move the most and which instead maintain a more rigid structure. This study draws a significant comparison between the 5Trans and the 5Cis models. Although less meaningful,

the case of 3Cis is also shown for completeness. It is however clear that the RMSF results for this system, here projected onto a loop with a Pro281 *cis* conformer, are intermediate between the two other models, as it is the case for any other analysis where quantities have been averaged over the five subunits. Although the value of the differences in the RMSF between residues of the same model is limited to 0.4 Å, the most notable change in flexibility is observed in residues such as Asp271, Thr272 and Leu273, which can still be considered part of the helical structure of M2. This increase flexibility is felt up to Pro274 and Ala275, and a few differences are also observable in the conformation of this area, where the top of M2 seems to be drawn toward the M3 helix. At the other end of the spectrum, Thr276, responsible for most of the intraloop interactions and positioned at the first bend of the loop, is subject to a decrease in relative flexibility: it is the most flexible section in 5Trans, but slightly stiffens when the *cis* isomer of Pro281 is present, probably due to the increased number of hydrogen bonds as observed in the tables of Fig. 3.26. A similar decrease, although less evident, is observed in the residues closer to the isomerising proline, among which Ile278 and Gly279 which bind with Thr276.

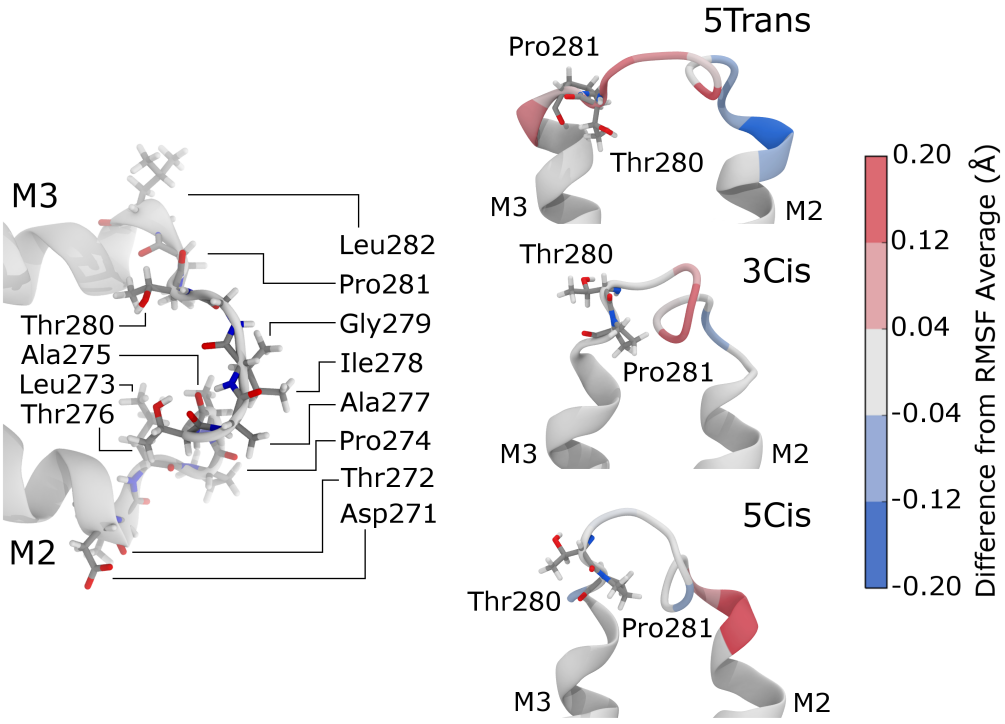


Figure 3.27: Left, the residues of the M2-M3 loop. Right, snapshots of the M2-M3 loop section taken from the trajectories of the three models 5Trans (top), 3Cis (middle) and 5Cis (bottom). The residues from 271 to 282 are coloured according to their difference in RMSF with respect to the loop average.

The most notable feature of this analysis is the evident shape shift toward a more "closed" structure of the loop. It is however unclear if the bonds with Thr276 are a cause or a consequence of this conformational change. The position of most

of the residues' side chains and rotamers do not seem to be affected by the Pro281 isomerisation, in line with the results of the hydrogen bond analysis of the three models. This raises the question if such isomerisation can affect a highly flexible loop like M2-M3 more significantly than with irrelevant modifications of its transient turns and bends. However, the behaviour of Pro281 itself and its flanking residue Thr280, directly perturbed by the metadynamics bias through the ψ angle CV, seems to suggest a different interpretation. While initially facing internally, toward the channel, the *cis* conformation pushes the side chain of Thr280 outward. In this way, the threonine can more easily interact with elements of the environment, such as waters or lipids or the ECD loops. This is confirmed by the analysis of the hydrogen bonds between the M2-M3 linker amino acids and the phosphatidylcholine heads of the lipid bilayer. In fact, among the few residues interacting with the lipid heads, the polar Thr280 is the only one forming hydrogen bonds consistently throughout the simulations of all three the models.

On the left hand side of Fig. 3.28 the average number of hydrogen bonds during the MD simulations are shown for the three models. A hydrogen bond between the hydroxyl group of the threonine and the phosphate of the phosphatidylcholine is formed only for 33% of the time on average in the 5Trans model. The occurrence of this interaction is dramatically increased in the two other models, where it is observed for 63% and 73% in 3Cis and 5Cis respectively. It is not clear if this conformational change, i.e. the rotation of the OH group of Thr280 that exposes it to the lipid bilayer, is solely induced by the proline isomerisation, which is technically a rotation of the peptide bond between the proline and the threonine. The *cis* conformers is characterised by the presence of the cyclic side chain of proline and the peptidic oxygen of threonine on the same side with respect to the backbone, while the *trans* conformation has them on opposite sides, as shown in Fig. 1.11 in Sec. 1.3. What seems plausible is that the lipid environment, together with the M2-M3 loop residues flanking Pro281, favours a reorientation of both the threonine hydroxyl, which reorientates across the loop, and of the cyclic side chain of proline, which, together with the peptide oxygen of Thr280 rotates to face downward, toward the helical structures. This interaction between the M2-M3 linker and the phosphatidylcholine heads is consistent with the slight increase in number density observed in Fig. 3.23. It is thus reasonable to suggest that lipids may be more than a simple scaffold for the membrane and may play a support role in mediating the gating process, e.g. by modifying the electrostatic environment of the Pro-Thr peptide bond in such a way to facilitate a rearrangement of the loop upon isomerisation. Although no interactions between such threonine and elements of the ECD loops are here observed for more than 10 % in any of the simulations, the possibility that different rearrangements of these loops in serotonine-bound proteins may also bond with Thr280 and with it affect the proline isomerisation should not be excluded.

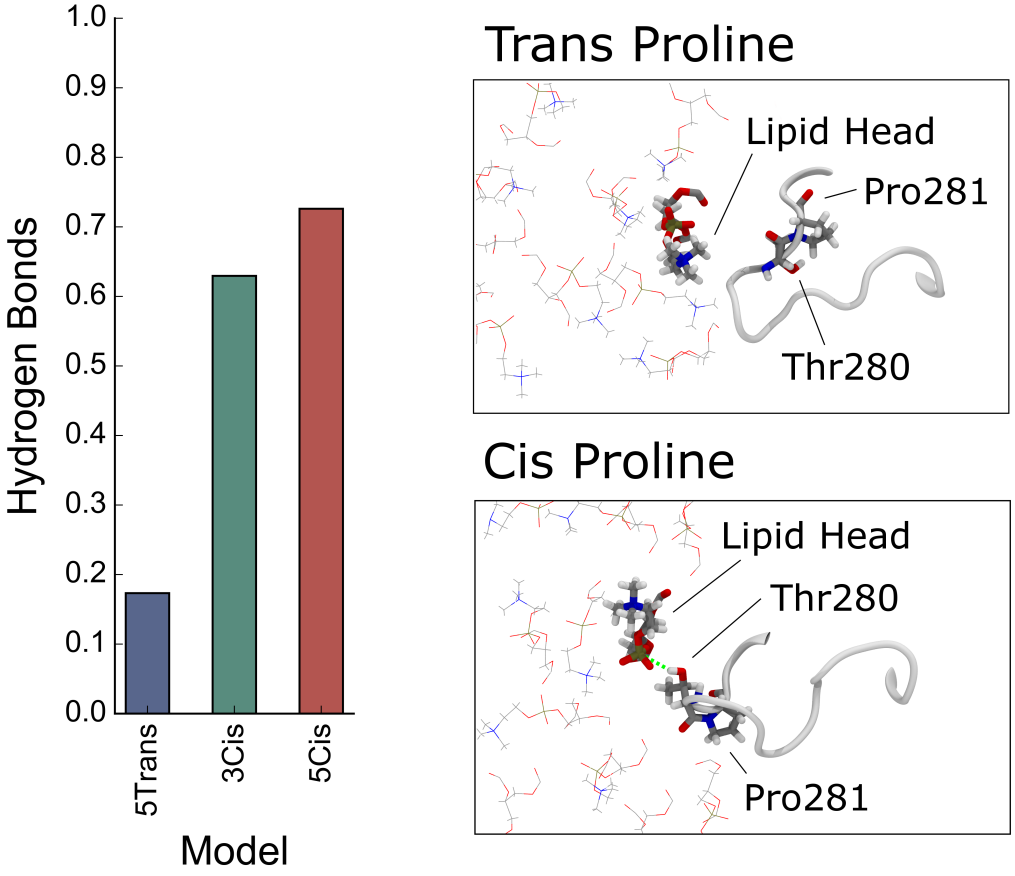


Figure 3.28: Left, number of hydrogen bonds formed between Thr280 in the M2-M3 loop and the phosphatidylcholine heads during the MD simulations in the three 5-HT$_3$R models: 5Trans (in blue), 3Cis (in green) and 5Cis (in red). The results are averaged over the 5 subunits. Right, typical conformation assumed by Thr280 and Pro281 when the proline is either in a *trans* state (top) or in a *cis* state (bottom) as seen from the top of the channel axis. The hydrogen bond interaction formed by the threonine and the phosphate of the lipid heads is highlighted in green in the *cis* case. Pro281, Thr280 and one of the interacting heads are shown explicitly, while the heads of the surrounding lipids are represented as lines.

Proline Isomerisation Free Energy Landscape

The free energy landscape of the isomerisation process has been investigated with the help of metadynamics. The exploration has been carried out for Pro281 in the M2-M3 loop in the receptor model and for a control system of proline dipeptide in water, with the aim to highlight how the surrounding environment affects the isomerisation within a 5-HT$_3$ receptor. While the metadynamics of the dipeptide in water took only 40 ns to converge, the high flexibility of the M2-M3 structure and the presence of complex interactions with lipids, water and the surrounding residues made the convergence of the Pro281 system more challenging, requiring to extend the simulation to at least 500 ns. To reduce the error on the reconstructed profiles, the free energy landscapes were averaged over the last 10 ns and 100 ns for the dipeptide and the M2-M3 proline respectively. The maps are shown in Fig. 3.29

as a function of the two torsional angles ξ and ψ, together with their projections along the ξ and ψ angles individually.
Some notable differences are present in the two maps. First of all the shape of the minima is much less defined in the M2-M3 loop, where the high flexibility of the surrounding residues and the presence of lipids interfere with the isomerisation process. The right hand side barriers are markedly increased up to about 116 ± 0.6 kJ/mol, while they never overcame 80 ± 0.1 kJ/mol in the dipeptide, and three out of four saddle points separating the *cis* and the *trans* conformers along the ξ torsion are increased by about 20-30 kJ/mol. As a result, there is a preferential path for the isomerisation, connecting the absolute minimum at $\xi = \pm 180°$ and $\psi = \pm 180°$ to the more shallow *cis* minimum at $\xi = 0°$ and $\psi = \pm 180°$. In fact, the orientation of the peptidic amide is crucial in catalysing the isomerisation, lowering the saddle point by about 10 kJ/mol when the hydrogen is in a *cis* arrangement with respect to the cyclic side chain.

Although containing integrated information, the profiles of the potential of mean force, namely the equivalent of the free energy as a function of a partial number of CVs and integrated over the other remaining degrees of freedom (d.o.f.), can help simplify the description and give a rough estimate of the energetic differences between the *cis* and *trans* minima in the two systems. In the profile as a function of ξ only, while for the solvated dipeptide the two conformations have similar free energy, with a difference of only 4.2 ± 0.4 kJ/mol in favour of the *trans* conformer, this difference increases significantly to 22.2 ± 2.4 kJ/mol in the 5-HT$_3$R system. This seems to suggest that the protein environment as studied in our simulations disfavours the isomerisation. Interesting differences are also observed in the high free energy regions separating the minima: the symmetry observed in the dipeptide, where two roughly equivalent barriers, at about 65 ± 0.1 kJ/mol are observed, is lost in the M2-M3 linker. In the latter case, one barrier reaches only about 54 ± 0.5 kJ/mol, while the other is about 78 ± 0.7 kJ/mol high. Hence, this system displays a non negligible preference for an anticlockwise isomerisation, which requires the Pro281 cyclic side chain and the Thr280 side chain to move under the M2-M3 loop. This may be simply the result of the presence of Leu282 side chain oriented outward the top of the M2-M3 loop and of Tyr140 aromatic ring in the Cys-loop, which comes in contact with the M2-M3 linker from the top and may obstruct the rotation in the opposite direction. This hypothesis is supported by the fact that the orientation of these two residues is consistent across the subunits, but the possibility of other elements at play should not be excluded. The lowering of the left barrier is consistent with the idea that the 5-HT$_3$ environment helps the transition between the isomers along a favourable path, although at equilibrium, the higher free energy of the *cis* isomer makes its less populated. Nevertheless, after being induced with metadynamics, the *cis* conformation was maintained during the entire length of the MD simulations,

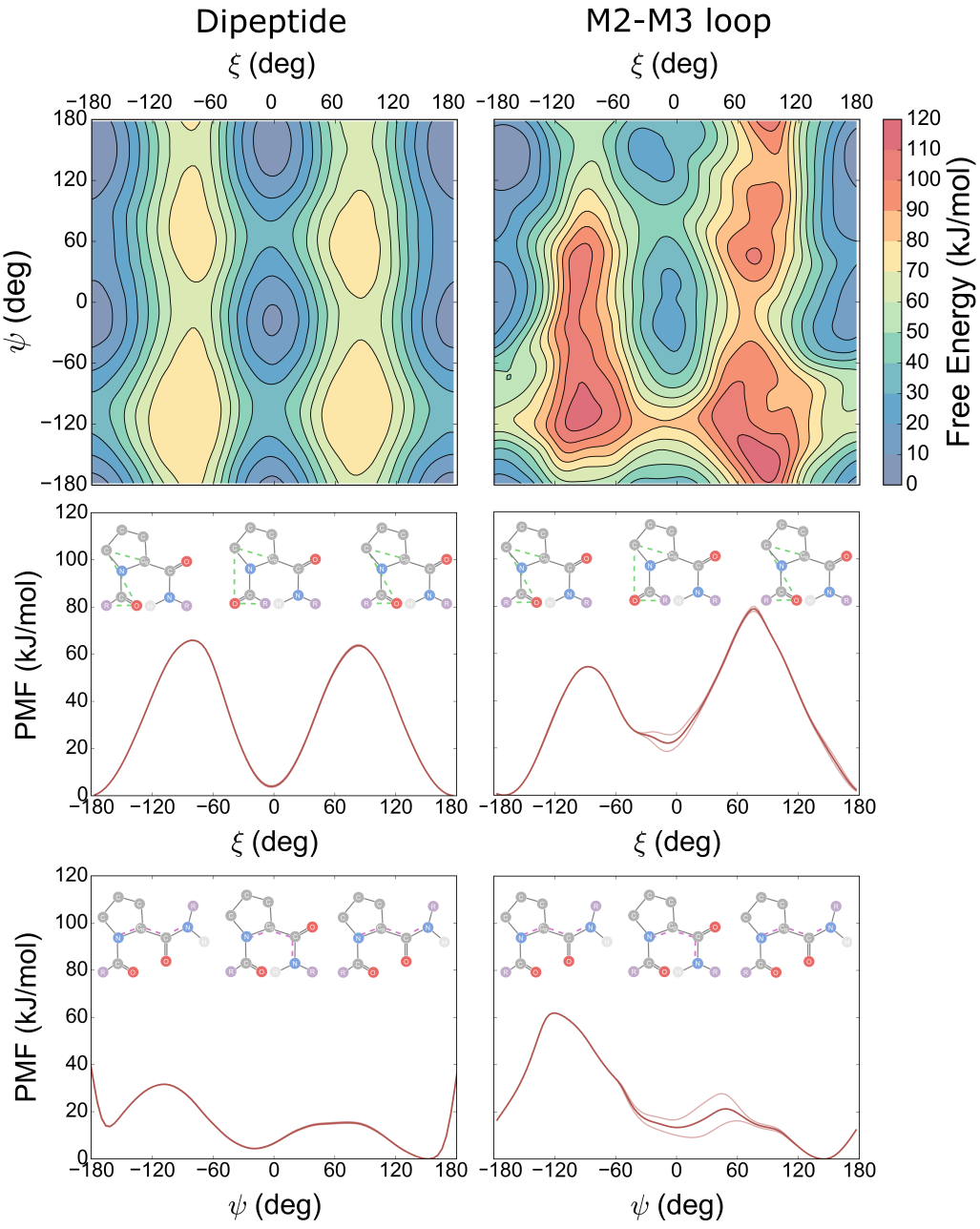


Figure 3.29: At the top, free energy surfaces of the solvated proline dipeptide (left) and of the proline in the 5-HT$_3$R environment (right) as a function of the torsional angle ξ and ψ. The absolute minimum has been set to zero and contour lines are every 10 kJ/mol. Bottom, the corresponding potentials of mean force (PMF) as a function of ζ only. The maps are averaged over the last 10 ns and 100 ns of simulation respectively. The free energy of the absolute minimum has been set to zero and the error is shown in light red lines.

confirming it as a metastable state. The difference of only 11 kJ/mol for free energy of the *cis* minima between the M2-M3 proline and the dipeptide in water is however rather small to draw definitive conclusions. It is not completely clear if the crystallised structure is in a closed, open or desensitised state and it is plausible to assume that

a bound protein may be characterised by an even lower barrier, maybe with the aid of interactions between the M2-M3 loop and the ECD loops that here are still scarce.

Structural Motions: Results and Discussion

In this section, the effects of Pro281 isomerisation on the secondary structures of the 5-HT$_3$ receptor are shown. In particular, the isomerisation induces a straightening of the upper part of the M2 helix, lining the ion permeable pore and which is key to the gating in pLGICs. This result suggests that Pro281 conformational state can influence the arrangement of the M2 helices and possibly initiate the opening of the channel.

The increase in free energy difference between the *trans* and *cis* conformers in the M2-M3 loop seems to suggest that the natural environment of apo-5-HT$_3$R, namely the protein in the absence of serotonin, disfavours the population of *cis* isomers, although the slightly decrease in energy of the barrier may help the exploration of such metastable state. Thanks to metadynamics, however, it was possible to induce the isomerisation in different subunits and study the effects that such conformational change has on the receptor secondary structures. The M2 and M3 helices belonging to the transmembrane domain have been proposed to be involved in the opening and closing of the ion permeable channel through a series of movements and conformational rearrangements [57]. Hence, a thorough analysis of the movements of such helices has been carried out on the MD trajectories of the three 5Trans, 3Cis and 5Cis systems, in order to identify to which degree the Pro281 isomerisation may affect them.

A first simple structural analysis has been carried out for the TMD helices. In Fig. 3.30 the polar and twist angles of the M2 helices principal axis with respect to the protein channel axis (as schematically represented) are monitored during the 250 ns of MD simulation for the three models. No major movements are observed, with an average polar angle which slightly decreases from $9.6 \pm 1.5^\circ$ for the 5Trans system, to $9.2 \pm 1.4^\circ$ in the two subunits of 3Cis where the proline is in the *trans* conformation, to $9.3 \pm 1.5^\circ$ for three remaining *cis* subunits and to $8.6 \pm 1.4^\circ$ for the 5Cis simulation. Similarly, the average twist angle gives a good indication of the magnitude of the rotational fluctuations that these helices undergo with respect to the ECD. In this case, $-0.5 \pm 2.1^\circ$ is the average angle for 5Trans, $1.0 \pm 2.1^\circ$ and $-3.0 \pm 2.2^\circ$ is observed for the *trans* and *cis* subunits of 3Cis respectively, and $-1.1 \pm 1.7^\circ$ in 5Cis. While the 5Trans model angle trend is mostly flat, introducing *cis* conformers seems to affect it: a slight anticlockwise twist is observed in the *cis* induced subunits of 3Cis, while this effect is dampened in the 5Cis simulation.

This result supports the idea that the system of five M2 helices do not overall significantly modify their conformation with respect to the rest of the protein.

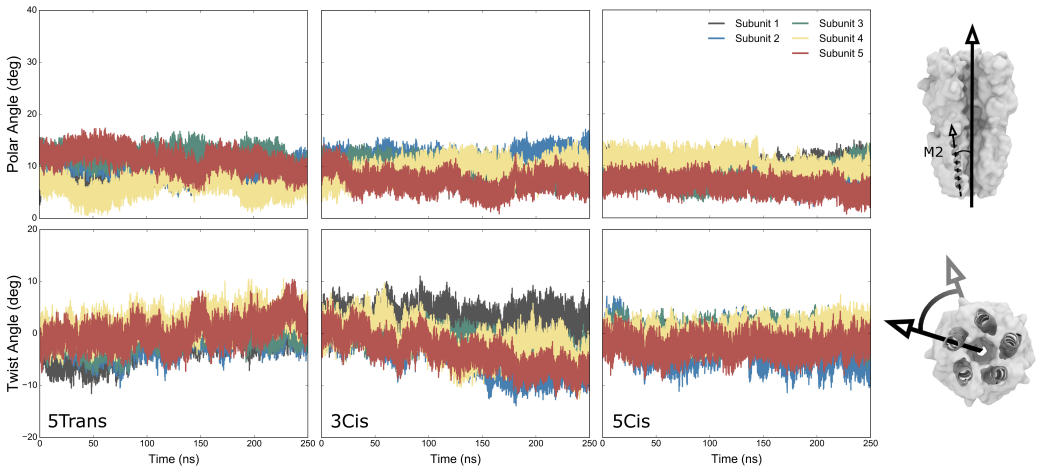


Figure 3.30: Polar angle (top row) and the twist angle (bottom row) of the M2 helices with respect to the main axis of the 5-HT$_3$ channel, monitored during the three MD simulations, for the 5Trans (left), 3Cis (center) and 5Cis (right) models.

Macroscopic movements cannot be expected to be induced in timescales no longer than a few hundreds of nanoseconds by simply tweaking the isomers of prolines in the interfacial region while leaving the binding site untouched.

Similarly, the relative polar and twist angles of the M2 helix axis with respect to the M3 helix axis have been monitored: the results are shown in Fig. 3.31, where again no major changes in conformation are observed. It is interesting to note the natural asymmetry of this model, as highlighted by the different values of polar angle for each subunit. While the two helices do not twist with respect to each other during the simulations (the average is $0.2 \pm 1.8°$ for 5Trans, $-0.4 \pm 1.6°$ for the *trans* subunits of 3Cis, $-1.3 \pm 2.0°$ for its *cis* subunits, and $-1.1 \pm 1.3°$ for 5Cis respectively), subunits 3 and 5 are characterised by a larger polar angle than the remaining three subunits. The average values of the polar angle are $13.0 \pm 1.9°$ for the 5Trans model, only $7.5 \pm 2.0°$ for the two *trans* subunits of 3Cis, $14.9 \pm 1.9°$ for the remaining three, and finally $13.1 \pm 2.0°$ for the 5Cis. According to a recent proposal, supported by mutagenesis experiments, interactions among residues of these two helices may be the key to regulate the desensitisation process [210]; in all the three cases investigated however, the M2-M3 structure is stable, the monitored properties are flat and the relative coordinates of the helices see no major change.

The last row of Fig. 3.31, however gives an interesting insight. In this case, variations in the angle defined by the extremes of the M2 helix and the C_α of Leu260, the hydrophobic residues occluding the ion permeable pore, which also splits the helix in half were observed. This angle was chosen so that an ideally straight helix would give 180°. The averages in this case are $153.0 \pm 4.0°$ for the 5Trans, $153.2 \pm 3.4°$ for the *trans* subunits in the 3Cis model, $159.3 \pm 4.0°$, the maximum value for the three subunits of 3Cis where the *cis* conformers were induced and and $158.4 \pm 5.4°$

for the 5Cis system. Beside the increase of about 5° observed in the average of 5Cis, this trajectory is characterised by temporary conformational changes in subunit 3, between 75 and 100ns of simulation, and a much more defined "jump" in the helix of subunit 4, toward the second half of the simulation. This behaviour is associated with minor conformational changes in the upper part of the helices, which straighten, as well as less important movements the lower part. These results seem to confirm that the increased RMSF in the residues of the M2-M3 linker closer to the M2 helix, discussed previously within the interface analysis, is indicative of a conformational change in the region. Moreover, the movements here observed seem to be maximized in the three subunits of 3Cis where the isomerisation was induced, with respect to the 5Cis model: although these changes are relatively small, this may be an indication that not all five subunits need to be bound to activate the ligand.

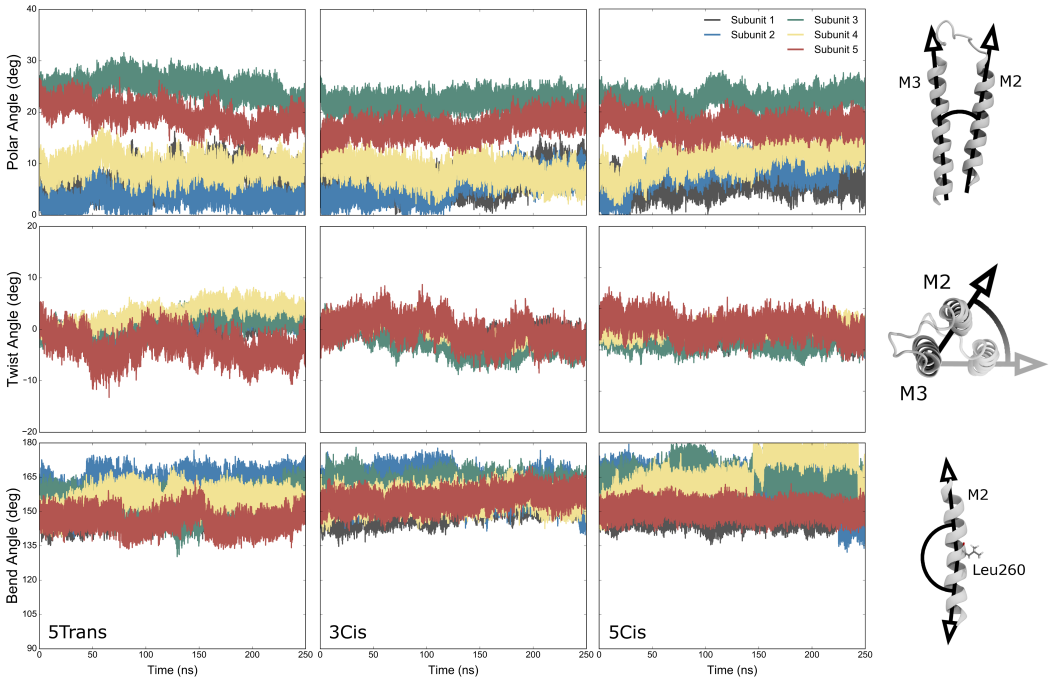


Figure 3.31: Polar angle (top row) and the twist angle (central row) of M2 helices with respect to their respective M3 helices of the 5-HT$_3$ channel, monitored during the three MD simulations, in the 5Trans (left), 3Cis (center) and 5Cis (right) models. The bottom row shows the behaviour of the bending angle of M2 where Leu260 is considered as the pivotal point during the same simulations.

Pore Diameter and Permeability

The analysis of the pore width profiles, shown on the right in Fig. 3.32, helps highlight this conformational change. In these graphs, the position of Leu260 is set to zero. From the 5Trans to the *cis* systems, a gradual narrowing of the pore on average in the transmembrane section is observed: the upper part of the helices narrows from 4.4 Å on average, to 2.7 Å and 2.3 Å at about +15 Å along the channel, where widest movements are observed, while the rest of the structure is fairly stable. The

narrowest point remains at the Leu260 ring, with 1.3 Å of radius on average, while the radius increases back again at about 2.8 Å on average at the bottom of the helices. These values are within the range that prevents a flux of ions and waters. In general it is possible to observe an increase in the fluctuations of the transmembrane M2 helices of 5Cis, which is reflected in the increased error bars of the profiles between -15 and $+15$ Å.

The rest of the protein remains unaffected. As observed in recent literature [211], a first narrowing in the ECD at $+48$ Å to a average of 3 Å acts as charge filter and keeps the negatively charged chloride ions out of the channel core; here a ring formed by the positively charged side chains of Lys108 constrict the pore, catching Cl^- ions and preventing their entrance, with few exceptions. According to the same study, this ring may however also partially hinder the entrance of Na^+ ions, which would instead prefer lateral cavities formed between the subunits as the main path to enter the vestibule. The ions and water molecules are then gathered in a vestibule, with the help of the negatively charged Asp271 and polar Thr272 at the top of the M2 helix, in position 20' and 21' respectively and whose side chains are facing the pore. These residues are followed by a set of mostly hydrophobic residues: in particular Ile268 and Val264, whose side chains are oriented toward the pore and help create further restrictions before Leu260 (Fig. 3.33, right). Lower down the helices, in positions -2' to -3', are residues which act as further polar filter, preventing Cl^- residues in the cytosol to enter the channel from below. They also seem to be involved in the desensitisation stage, in a process which would see this region decrease sensibly its radius [210]. However, it is difficult to draw conclusions about these specific residues, since this region in the model has been manipulated by removing the terminal MA and MX helices of the crystal structure. The hydrophobic area spans from -5 and to $+10$ Å along the channel axis. Thanks to the evidence of a number of recently crystallised structures, the hypothesis of an hydrophobic gating has been proposed for pLGICs: according to this picture, ion conduction is prevented by means of an energetic barrier generated by the absence of water molecules in the pore cavity; such liquid-vapour water transitions, or "dewetting" events, are caused by the hydrophobicity of the residues lining the pore. This mechanism has already been extensively validated for other kind of membrane channels, such as nanopores [54].

In Fig. 3.32 (left), fluctuations in the distribution of water and ions in the pore are observed in all three systems. A partial dewetting event, a decrease in water density, is observed between 50 and 100 ns in the ECD and in the lower TMD sections of the 5Trans system, proof that such events can spontaneously produce, probably due to stochastic fluctuations in the side chains of residues lining the channel and more in general in the structure of the protein. No particular conformational change was associated with this behaviour in the simulation trajectories.

On the other hand, longer although still partial dewetting events in the ECD and lower TMD are observed in both 3Cis and 5Cis trajectories at about 100 and 125 ns, as a consequence of the decrease in radius of the TMD channel cavity found in both profiles.

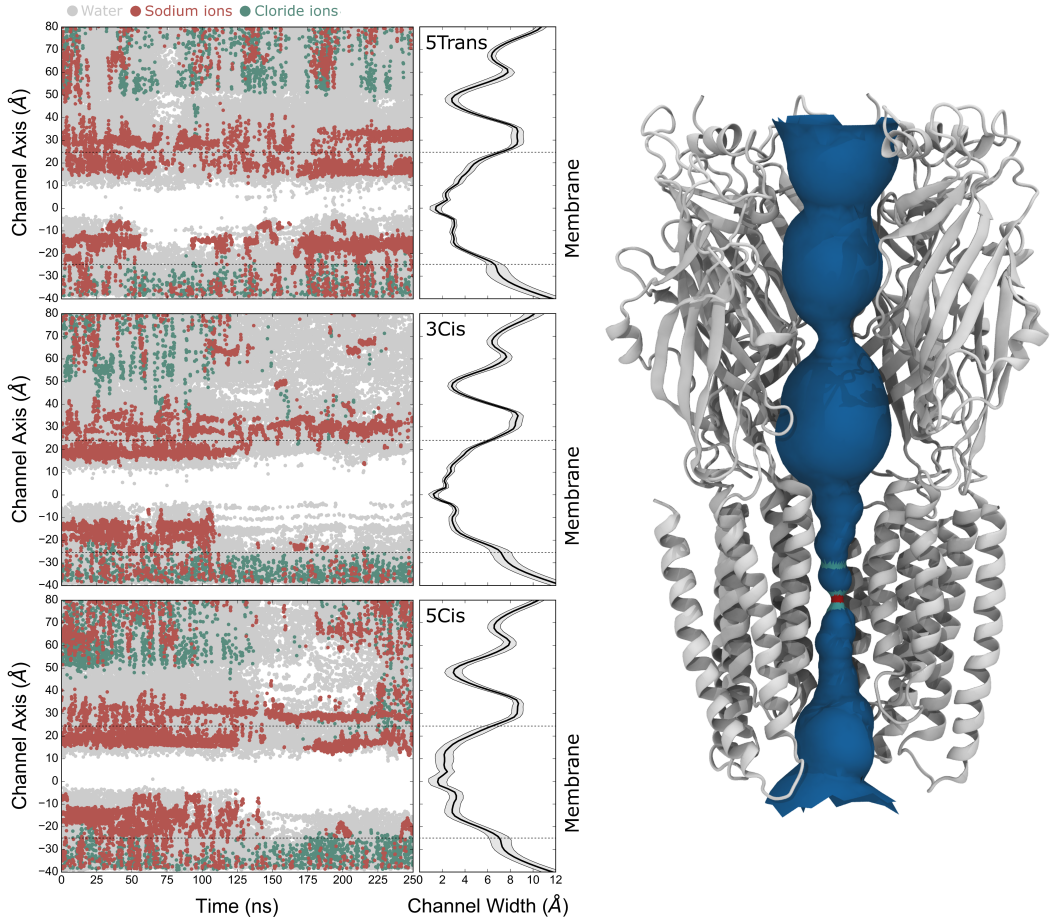


Figure 3.32: On the left, the position of water molecules and charged ions along the central pore of the 5-HT$_3$ receptor during the three MD simulations, 5Trans (top), 3Cis (middle) and 5Cis (bottom). In the central section, the profile of the pore of the three systems calculated with HOLE 2.0 [212] and averaged over 250 ns. In both graph, Leu260 position is used as the zero reference. On the right, snapshot of a section of the 5Trans receptor and its inner pore surface. In red are the regions where the radius is lower than 1.5 Å, the minimum radius to accommodate a water molecule, in blue are the regions with radius higher than 2.5 Å, sufficient to fit two water molecules, while in cyan are the areas in between.

The straightening of the M2 helices does not seem however to affect the hydrophobic restriction at the level of the Leu260 ring within the simulation time. The graphs in Fig. 3.33 shows no remarkable differences among the positions of the Leu260 C$_\alpha$ in all three models: both average position and standard deviation radius are similar in all five subunits. Nevertheless, it is plausible to suggest that straightening of M2 may be a first step toward the opening of the ion permeable pore, which would be followed, on longer scales, by the withdrawal of the hydrophobic side chains of

Leu260 from the central axis. The flexibility of the side chains in the residues facing the pore has been observed to play a crucial role in managing the permeability of the pore in computational MD and Monte Carlo studies of Cys-loop receptors [213].

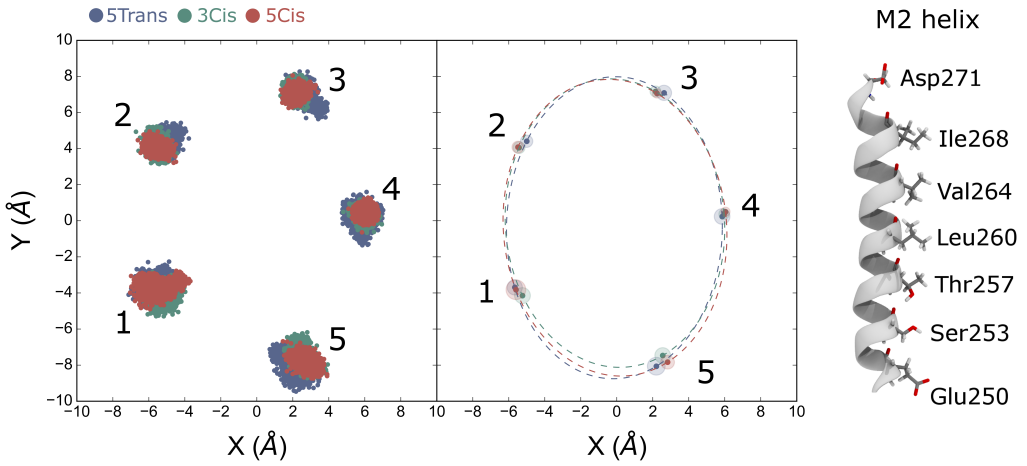


Figure 3.33: On the left, distribution of the position of the C_α in the side chains of Leu260 during 250 ns simulations for the 5Trans (blue), 3Cis (green) and 5Cis (red) models, as seen from the top of the channel axis (here set as the Z axis). In the central panel, the position of each residues is represented as circles centered around the average position and with the standard deviation as radius. The width of the pore is approximated to an ellipse (dashed line). The numbers refer to the subunits to which each residue belongs. On the right, a view of the M2 helix where the residues whose side chain are facing the ion permeable pore are explicitly shown.

M2 Helix Bending

To further assess the relation between the isomerisation of proline and these conformational rearrangement, the metadynamics bias accumulated over 500 ns can be reweighted as a function of alternative degrees of freedom of the M2 helix that were not selected as collective variables. In Fig. 3.34 the free energy landscape reweighted as a function of the isomerisation angle ξ and of the bend angle of helix M2 is shown. The characteristic time of major structural rearrangements such us the movements of M2 helix is much slower than the time required to the metadynamics to sample the isomerisation, preventing a complete and accurate reweighting of this mechanism. While the values of the free energy in the various minima may be inaccurate, the resulting graph can, however, qualitatively suggest possible effects on the helix in the presence of the different proline isomers. An increase in the bend angle when transitioning from *trans* to *cis* conformers can in fact be observed. While a deep basin extends between 145° and 160° for the *trans* isomer ($\xi = \pm 180°$), for the *cis* isomer ($\xi = \pm 0°$) the borders of the minimum are pushed up by at least five degrees, with the basin extending between 150° and 165°, reaching in a few cases regions at 170°.

Fluctuations in the helix conformation ensure the presence of conformations between

145° and 160° favourable both in *cis* and *trans* basins. This suggests that more than an abrupt switch in conformation, the isomerisation may push smoothly the helix toward higher bend angles. An analysis of the conformation belonging to the two basins here identified confirms that most of the difference is concentrated in the upper body of the helix, toward the M2-M3 loop, as seen on the right hand side of Fig. 3.34.

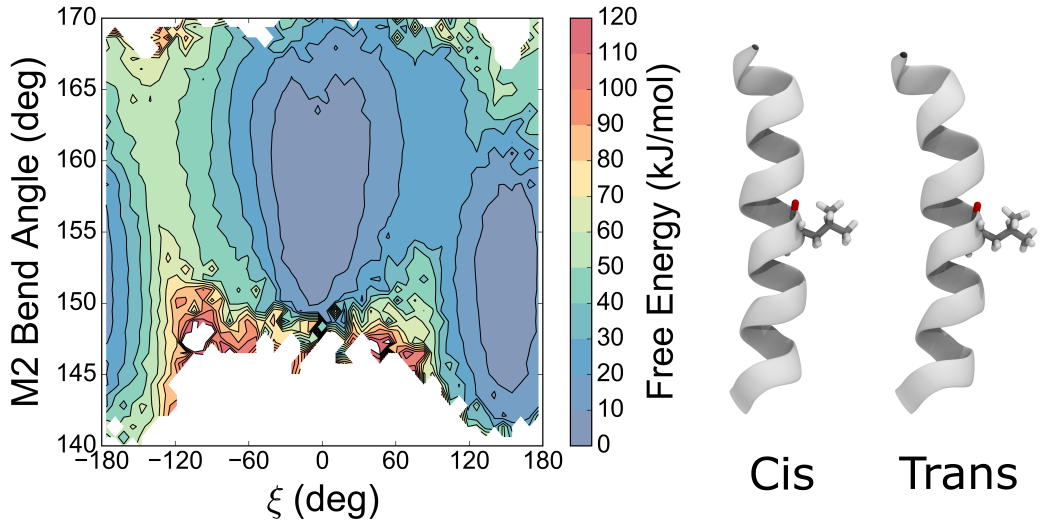


Figure 3.34: Free energy landscape from metadynamics for the 5-HT$_3$ receptor reweighted as a function of the proline isomerisation angle ξ and the M2 helix bend angle. The free energy of the lowest minimum has been set to zero.

Essential Dynamics Analysis

Although relatively small, the induction of *cis* conformers in the M2-M3 loop prolines at position 281 seems to subtly modify the dynamics of the system: an Essential Dynamics Analysis (EDA) on the trajectories of the five M2 helices, can help highlighting if these differences are also found in major functional motions of the three models. EDA, also known as Principal Component Analysis (PCA), is a powerful method to extract and analyse the collective motions that may be hidden in MD trajectories. E.g. major collective vibrational modes, which are generally assumed to define the function of the protein, can be filtered and separated from fast degrees of freedom, usually related to local fluctuations [214]. The application of this method is straightforward: a variance-covariance matrix of the coordinate fluctuations is first built as in

$$\mathbf{C} = < (\mathbf{x}(\mathbf{t}) - < \mathbf{x} >)(\mathbf{x}(\mathbf{t}) - < \mathbf{x} >)^T > \tag{3.3}$$

where the angle brackets represent the ensemble average and ^{T} is the transpose vector. $\mathbf{C}$ is a symmetric matrix and can then be diagonalised with an orthogonal

coordinate transformation $\mathbf{C} = \mathbf{V}\Lambda\mathbf{V}^T$. In $\mathbf{V}$ are the eigenvectors as columns, while on the diagonal of the Λ matrix are the eigenvalues (the mean square of the eigenvector coordinate fluctuations). Only the motions of highest intensity (highest eigenvalue) are usually taken in consideration as they represent the major collective modes [214].
Differently from other, more approximate normal modes methods, EDA does not assume harmonicity in the potential; in fact, being based directly on the trajectories, anharmonicity is naturally included in the protein motions.
The EDA was performed on the C_αs atoms of the five M2 helices; in Fig. 3.35 (left), their trajectories during the three simulations are projected onto the space defined by the first two eigenvectors of the 5Trans model. The displacement of the C_αs atoms of the five helices for such modes are shown on the right hand side of the same figure.

Although a significant overlap among the three systems is observed in the trajectory projections, a few regions are exclusive of a specific simulation. As expected, these are characterised by relatively small differences in the position of the top part of the M2 helix. This can be also seen in the right panel of Fig. 3.35, where the amplitude of the displacement, here represented by the length of the black arrows, increases for the C_αs of the top portion. These modes thus relate mostly to a deformation of the top half of the M2 helices, which may be at the base of the functional opening and closure mechanism of the hydrophobic gate. 5Trans is able to visit most of the space and is thus characterised by a good flexibility, 5Cis seems to be more rigidly limited to the region between -0.2 and 0.2 for both the eigenvectors. In the case of 3Cis, a region at the lower left extreme, where the first PCA eigenvector assumes values between -0.8 and -0.6, is visited more often than in the other two cases.
EDA gives an interesting insight as it allows to separate these relatively small fluctuation differences and identify different regions of the eigenvectors space explored during the simulations. These results strongly support the picture that previous analysis had given. The *cis-trans* isomerisation of the M2-M3 linker prolines does not lead to strong conformational rearrangements in the TMD within the time sampled time scale, but rather to some subtle modifications in its helical structures and dynamics. Speculating that the straightening of the top section of the M2 helices, the closer part to the loop, may be the first step of the gating mechanism may be far fetched; however, it is plausible that relatively small movements of this kind in the helices lining the central pore may induce rearrangements in the side chain of hydrophobic residues such as Leu260, big enough to start a selective flow of water molecules and ions.

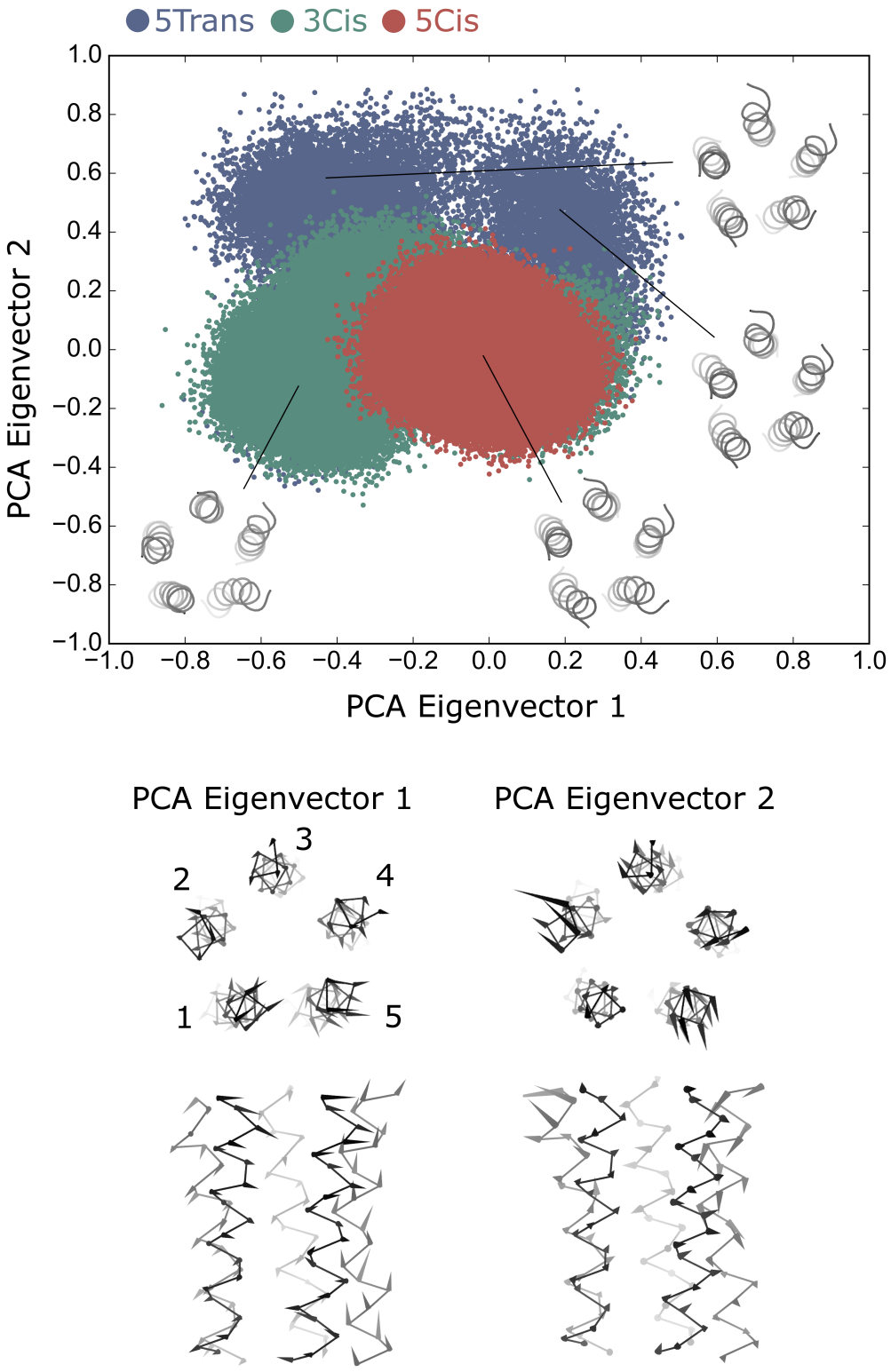


Figure 3.35: On the left, the dynamics of the M2 helices group projected onto the space of the first two PCA eigenvectors of 5Trans. Here the trajectory of the 5Trans model is in blue, that of the 3Cis model is in green and that of the 5Cis model is in red. Conformations assumed by the five M2 helices in the different regions of the plot are shown from the top of the channel. On the right, a representation of the two lowest PCA vectors of the 5Trans simulation. In black are the traces of the M2 helices C_αs from the top of the channel and from the side; arrows indicate direction and relative magnitude of the displacements.

Conclusions

In this chapter we studied a model of the 5-HT$_3$ receptor, based on a recent X-ray structure, including the extracellular and transmembrane domain embedded in a lipid bilayer to mimic the cell membrane. We concentrated on the interface between ECD and TMD and, in particular, on Pro281, a highly conserved proline in the M2-M3 loop. This proline has been proposed, with the support of electrophysiology mutagenesis experiments, as a potential molecular switch to open the ion channel. Specifically, we exploited the power of metadynamics to investigate the effects that the *trans-cis* isomerisation of Pro281 has on the transmembrane domain of the 5-HT$_3$ receptor. In order to evaluate these effects, three MD simulations of the 5-HT$_3$ receptor with all *trans*, three *cis* and all *cis* Pro281 isomers were carried out. This allowed us to analyse in detail the repercussions on the rest of the receptor, at least on the timescale sampled by the MD, and shed light on the structural role that this proline plays within the M2-M3 structure. Relatively small, but significant differences were observed in the conformations assumed by the M2 helices, which line the channel and whose role in the gating process has been widely debated, for the models containing metastable *cis* prolines.

Mapping the free energy landscape as a function of the torsional angles that define the isomerisation process showed that the environment of the M2-M3 linker in this receptor model does not favour the *cis* conformer, whose free energy relative to the *trans* conformation is about 18 kJ/mol smaller in the proline dipeptide in water. The presence of obstructing amino acid side chains from neighbouring loops and within the M2-M3 loop itself gives rise to a clear preference for a directional anticlockwise isomerisation path: for this, the barrier is lower than what observed in aqueous solution, suggesting an easier transition in such direction. Although debatable, the similarities with other crystallised pLGICs suggest that the 5-HT$_3$ structure here studied is in a closed or resting state, which could explain the preference of the *trans* conformer. A 2-3 Å radius at the narrowest point of the channel is in line with the hypothesis of a resting/closed structure, since experiments identified minimal pore radii for the ionic flux between 3.7 and 4.2 Å in cation selective receptors such as 5-HT$_3$R and nAChR [213]. It is plausible to assume that in the presence of a serotonine-bound ECD, conformational changes, such as the twisting of the domain, may modify the contacts between residues flanking the proline and amino acids in the Cys-loop, the $\beta1-\beta2$ loop or $\beta8-\beta9$ loop, further lowering the free energy barrier and the *cis* basin and increasing the rate of isomerisation. Thr280, which flanks Pro281, has been here highlighted as a good candidate to participate in this process: this is supported by its interactions with phosphatidylcholine lipid heads which substantially increases for *cis* Pro281, and possibly by the role of its peptide oxygen acting as an electrostatically sensitive handle to induce the isomerisation.

An interesting work by Camilloni et al. [215] highlighted how the electrostatic field within the active site of cyclophilin A can catalyse the *cis-trans* isomerisation of proline by acting on the N-C-O peptide bonds, an electrostatic lever, between proline and glycine on a model substrate. In 5-HT$_3$R a similar mechanism may be suggested. The chain of events started by the binding of serotonin to the ECD binding site may lead to conformational rearrangements in the interfacial loops that modify the electrostatic field surrounding Pro281 and Thr280. Helped by the environment created by the phosphatidylcholine heads, favourable to the rotation of Thr280, this event may favour the isomerisation.

To confirm this hypothesis, molecular dynamics and metadynamics simulations with the holo-protein may focus on variations in the ECD loops conformation, hydrogen bond and electrostatic interactions with the residues of the M2-M3 loops and the surrounding lipids. Similarly, mutagenesis experiments could be designed to confirm the role of surrounding residues in assisting or hindering the proline isomerisation. Together with Thr280, Tyr140 and Thr276 are possible candidates for relevant roles: Thr276 was here observed taking part in a number of interactions with other residues of the M2-M3 loop, the basis for the rearrangement to a "closed" loop conformation in the presence of the *cis* isomer, which forces the M2 helix to straighten. On the other hand, Tyr140, in the Cys-loop, ensured the preference for an anticlockwise isomerisation, by occluding the movements of the proline cyclic side chain in the opposite direction. Studies on GluCl crystallised with invermectine proposed as crucial for channel gating a movement of the conserved proline which pushes against a valine in the $\beta1 - \beta2$ loop, keeping the M2 helices in their open state, and finally passes under it thanks to a concerted twisting rotation of the TMD against the ECD which allows for the closure of the channel [57, 209]. A parallel can be drawn here: although in the 5-HT$_3$ receptor structure the arrangement of the M2-M3 linker and the $\beta1 - \beta2$ one does not allow the proline to approach this specific ECD loop, the twisting of the ECD with respect to the TMD could still induce sufficient changes in other loops, for example allowing Tyr140 to move out of the way, thus decreasing sensibly the barrier for a clockwise isomerisation of the proline and facilitating the transition.

After the isomerisation has been induced, the M2-M3 loop adjusts consequently, drawing the top of M2 helix to a straighter angle with respect to it axis. Movements at the level of the hydrophobic Leu260 may ensue from this transition and cause the enlargement of the channel just enough to allow sodium ions and water molecules to cross the membrane. A few works on a locally closed GLIC receptor (where the M2 helices are in a closed arrangement, while the rest of the architecture resembles an open state) proposed the intriguing idea that conformational rearrangements during opening events pertain only the upper part of the M2 helix, while the lower end remains immobile [216, 217].

It is still unclear how many bound ligands to their respective subunits are required to fully activate the receptor. However, the effect on the pore radius may be additive and the flow of ions may start once enough space has been made by the side chains of Leu260 and other hydrophobic residues close by. At variance to the more commonly accepted idea of a concerted motion of the entire TMD, an asymmetric opening of the channel pore, following the binding to a partial number of subunits has been suggested for other Cys-loop receptors and should also be taken into consideration for the serotonin receptor [11, 218]. To assess more conclusively the role of proline isomerisation in these structural and dynamical changes observed in both the M2-M3 linker and the M2 helical structure, metadynamics simulations taking into account slowly varying degrees of freedom of the 5TH$_3$ receptor TMD explicitly may prove extremely useful.

In conclusion, this work shed light on the direct effects of Pro281 isomerisation on the structural and dynamical properties of the M2 helices lining the ion permeable pore. A relation between the presence of *cis* conformers and rearrangements in the upper region of the helical structures have been observed which may represent the first step of the ion permeable pore opening. These results open the way to further studies on this receptor, which should focus on the role of interfacial loops in transmitting the signal from the ligand bound ECD to the M2-M3 loop here investigated. Understanding how the presence of serotonin can influence conformations and interactions in this region and affect the isomerisation properties of Pro281 could finally lead to a complete understanding of the chain of events that characterise the gating mechanism in the 5-HT$_3$ receptor.

3.4 Conformational Analysis of Amino Acids

The study of proline isomerisation in the 5-HT_3 receptor highlighted how structure and dynamics of short peptides, or even single amino acids, can influence the behaviour of secondary structures and proteins on a much larger scale. In spite of their simplicity, however, understanding the conformational preferences of such systems, which is key to understand their role and behaviour in larger biomolecules, can be a daunting task. Even a limited number of atoms can give rise to an exceeding number of different conformations and an automated way to explore, extract and evaluate all these structures would be very useful. In this chapter state-of-the-art computational techniques have been thus used to define a protocol to obtain a relatively quick and complete exploration of the conformational free energy landscape of small biological fragments.

We chose to work with aspartic acid (Asp), a proteogenic amino acid involved in a wide variety of biological functions, as an example of the complexity that the flexible degrees of freedom of even relatively small molecules produce. Aspartic acid is an non-essential α-amino acid involved in a number of metabolic functions, from the synthesis of proteins and biomolecules to the gluconeogenesis and the activation of NMDA glutamate receptors. It is characterised by a carboxyl group as side chain, usually deprotonated and negatively charged under physiological conditions, which justifies its presence in numerous protein active sites [74]. In nature, this compound is found in the neutral form in the gas phase and in the zwitterionic form in aqueous solution and in crystals. In the latter form the hydrogen of the Cα-COOH group is donated to the nitrogen to form a NH_3 amino group, as shown in Fig. 3.36, resulting into opposite charges for these two moieties [219]. There are many stable/metastable conformers for both these forms, which can be exhaustively characterised by the six torsional angles labelled in Fig. 3.36: three along the backbone, i.e. α (O_6-C_1-C_2-C_3), β (C_1-C_2-C_3-C_4) and γ (C_3-C_3-C_4-O_8); two to describe the relative position of hydrogen atoms in the carboxyl groups, i.e. δ (C_2-C_1-O_6-H_{O6}) and ζ (C_3-C_4-O_8-H_{O8}); and one for the rotation of the amino group, i.e. ϵ (C_2-C_3-N-H_N). ζ is not defined in the zwitterionic form, because of the absence of the hydroxyl, while γ loses relevance due to the consequent symmetry of the remaining carboxyl group.
Even for a basic amino acid such as Asp, the flexibility of its torsional angles makes a complete exploration of the conformational space far from trivial. In general, the larger is the molecule, the more complex this problem becomes; however, a thorough knowledge of the conformational preferences, and thus of the behaviour, of the building blocks of proteins represents a first crucial step toward the understanding of larger biomolecules.

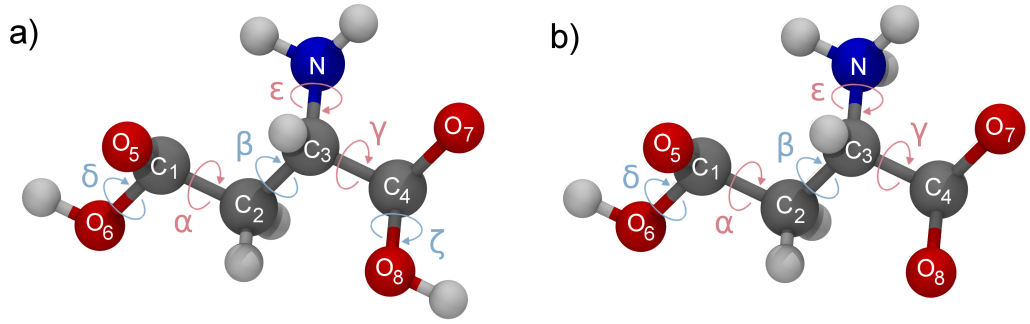


Figure 3.36: Neutral (left, a) and zwitterionic (right, b) aspartic acid, with relevant atoms and torsional angles labelled.

Experimental information on low energy conformers can be obtained by spectroscopic methods. The most powerful technique for conformational identification in the gas phase is rotational spectroscopy, which has been successfully applied in combination with laser ablation to determine the conformations of many natural amino acids [220–222]. In the condensed phases, methods such as NMR and X-ray crystallography are also widely used. Theoretically, stable/metastable conformers can be identified by scanning the torsional angles at small intervals, calculating the potential energy for each conformation with force fields or more accurate methods like density functional theory or second order Møller-Plesset perturbation theory (MP2), and correcting it *a posteriori* for temperature effects and entropic contributions from normal mode analysis to estimate the free energy [223]. Rotational constants and quadrupolar moments can then be calculated to interpret experiments. However, this protocol is cumbersome even for few torsional angles, it does not calculate the free energy directly at finite temperature, e.g. it disregards anharmonic effects, and, in aqueous solution, it does not consider the solvent explicitly.
Hence, we suggest an alternative approach to explore the conformational landscape of Asp to benchmark our theoretical protocol. Experimental information (in the gas phase) derived from rotational spectroscopy is available for this amino acid, which has a tractable size, making it ideal for this study [72]. The proposed approach is based on the calculation of the free energy landscape at room temperature by means of the enhanced sampling metadynamics technique, as a function of selected torsional angles. The exploration was performed with classical force fields, to ease the computational cost with respect to *ab initio* calculations, and results obtained with different parametrisations were compared to assess the influence of the force field choice on the free energy surface. When needed, the dimensionality reduction algorithm sketch-map [161] was employed to simplify the interpretation of the data and facilitate the discrimination of free energy basins and their interconnecting paths. Once minimal free energy conformers were identified, their stability and energies were evaluated at the DFT and MP2 levels of theory. Besides the low energy conformers in

the gas phase accessible to rotational spectroscopy experiments, high energy isomers and conformers in solution could also be identified.
The promising results here obtained demonstrate how the coupling of force field based metadynamics, sketch-map analysis and *ab initio* calculations is able to produce an exhaustive conformational exploration to complement the experimental data. Following this work, the protocol could then be applied to disentangle complex conformational landscapes of larger peptides. A paper containing these results is currently under preparation [166].

Simulation Details

Preliminary tests showed that the paths separating some of the conformers are too high in free energy to allow transitions between minima in the limited timescale accessible to MD, and thus required simulation techniques to accelerate rare events. Given its efficiency in enhancing the sampling of conventional MD, and the quality of results obtained with the RDL and the 5-HT_3 receptors, metadynamics was here chosen to speed up the exploration of the conformational landscape.
As metadynamics efficiency decreases with the number of biased CVs, the minimal CV set providing meaningful results should be chosen. To map the conformers of Asp, the three backbone torsional angles α, β and γ are an obvious, but not adequate choice, as demonstrated by preliminary calculations. In fact, the free energy as a function of α and γ is characterised by minima separated by barriers of few kJ/mol, which can be adequately sampled by conventional MD. However much higher barriers were observed between the minima as a function of δ and ζ, which define the *trans* and the *cis* isomers of the carboxyl groups. These two angles are involved in the formation of intramolecular hydrogen bonds, with the amino group acting as donor or acceptor when interacting with the hydroxyl and/or the oxygen of the carboxyl groups at the opposite ends. These interactions can influence the position of the carboxyl groups (e.g. restraining their mobility when the hydrogen bond is present) creating a correlation between the two angles, α and γ, which define the orientation of the groups, and the two angles which describe the position of the carboxylic hydrogens, δ and ζ, respectively. Hence, the most relevant torsional angles driving conformational changes in Asp are β, δ and ζ, which were used as CVs in the metadynamics simulations. Finally, low free energy barriers, which can be quickly sampled by MD, characterise the torsion about ϵ, which determines the orientation of the amino group, justifying its exclusion from the CV selection. For the zwitterionic form, the CVs were limited to β and δ, because the symmetry of the carboxyl group makes γ negligible, while ζ is not defined.
Asp was simulated with NAMD 2.9 and the CHARMM36 force field [224], one of the most commonly used force fields in biomolecular simulations, under different conditions: the neutral form in the gas phase, used as reference and for comparison

with rotational spectroscopy experiments, and the zwitterionic form in water solution. Two non-physical control systems, i.e. neutral Asp in water and zwitterionic Asp in the gas phase were also studied to better understand the effect of environment on the force field parameters. The two latter cases can be simulated only thanks to the pre-defined unbreakable bonds in the force field, which prevent the proton transfer between the amino and carboxyl group that would naturally occur. To further assess the impact of the choice of the force field, the calculations for neutral Asp *in vacuo* were repeated with two other commonly used parametrisations, i.e. AMBER ff99SB-ildn and OPLS-AA [225, 226], and the resulting free energy maps were compared to identify shared features and differences.
Restrained ESP charges were calculated at the Hartree Fock level of theory with a 6-31G* basis set on a molecular geometry optimised with DFT and the B3LYP exchange and correlation functional, using GAUSSIAN 09.
In the gas phase, the molecule was first equilibrated at 300 K with no periodic boundary conditions, while a 12 Å buffer of TIP3P water molecules was added in the solvated systems in a periodically repeated truncated octahedral cell and equilibrated at 300 K and 1 atm. A Langevin thermostat with a collision frequency of 1 ps^{-1} and a Langevin piston with period of 0.1 ps and decay time of 0.05 ps were applied. Long-range interactions were cut at 10 Å and particle mesh Ewald was employed for the calculation of long range electrostatic interactions. No restraints were applied to the vibration of bonds containing hydrogens. A time step of 0.5 fs was used. 100 ns of well-tempered metadynamics simulations were then performed with PLUMED 1.3 in the NVT ensemble, using, for the solvated systems, the average volume from the equilibrated NPT dynamics. All others parameters were kept the same as those used in the MD simulations. Gaussians with an initial height of 0.8 kJ/mol and width of 9° were deposited every 0.25 ps to bias the torsional angles, with a bias factor of 5. Free energy landscapes were reconstructed after 100 ns, namely after 400,000 Gaussians were deposited and the CVs showed diffusive behavior.

In a multi-dimensional free energy landscape the identification of the stable/metastable conformers and their connection network may be non-trivial, even when the dimensions are limited to three; projections to two-dimensions by integrating out the third or more CVs may lose significant information. To overcome these limitations (which becomes severe when increasing the number of variables), we adopted sketch-map [161–163], a recent method based on metric multi-dimensional scaling, [164] to disentangle the conformational landscape of Asp by mapping it from the high-dimensional space of the metadynamics CVs and other unbiased torsional angles to a bi-dimensional plane. This dimensionality reduction algorithm has been specifically designed to deal with atomistic simulation data, usually characterised by a set of well separated basins interconnected by transition paths. Thermal fluctuation noise around the basins and

the limited sampling of transition pathways are further features of this kind of data. The sketch-map result is an intuitive representation of the data network, in our case of the conformational landscape of the molecule of interest, with transition states mapped as continuous paths. Although the relative position of the basins is partially lost and the barriers may not be realistic, the free energies and the basin connectivity are preserved where possible.

The sketch-map algorithm requires first the choice of which degrees of freedom to map; then a limited set of landmark points has to be selected, taking care that they respect the underlaying Boltzmann distribution of the trajectory points in the multi-dimensional space. Working with a subset of the trajectory allows to speed up the calculation, and the remaining point can be then quickly mapped based on their initial distance form such landmarks. In our case, 500 landmark points were selected according to this criterion and mapped through a multi-step minimisation of a sigmoid function. This function was chosen instead of a naive Euclidean distance for its efficacy in maintaining the distance relation between points during the projection to a lower-dimensionality space: points that are far or close together in the initial high-dimensionality space are projected far or close together in the final map, at the cost of losing the quantitative value of their distances. The specific parameters of the sigmoid function has to be chosen so to to clearly separates the different conformational basins; the resulting map is not unique [163]. We used sketch-map to reduce to bi-dimensional maps the free energy landscape as a function of the three metadynamics CVs β, δ and ζ and of all five torsional angles, namely including also α and γ.

Once the data had been mapped and the bi-dimensional free energy network calculated, the characteristic conformations belonging to each basin were extracted: a RMSD based cluster analysis on the structures within the trajectories showed that, once cleaned of noise, a single strongly favoured conformation was present in each basin in most cases. When mapping the three-dimensional space of the metadynamics CVs, in a few cases more than one conformer was associated to a single identified minimal free energy basin, as information on the α and γ angles had been partially lost because of the choice of the mapped degrees of freedom. The cluster analysis helped disentangle the degenerate conformers within the same basin as the free energy dependence on α and γ is strongly influenced by the presence of intramolecular hydrogen bonds, making it easy to extrapolate data on all the six torsional angles and partially recover the information hidden by the selection of degrees of freedom to map. A full discrimination of the conformers, where needed, was obtained by mapping the free energy landscape as a function of (up to) five torsional angles, e.g. excluding only the fast degree of freedom described by ϵ.

In order to distinguish the identified conformers, we adopted the following nomenclature, loosely following Ref. [72]. The free energy profile of the central torsional angle

β is characterised by three minima: $a \simeq \pm 180°$, $b \simeq -60°$ and $c \simeq 60°$. These have the strongest impact on the sketch-map, which is circularly arranged around the values of β. The δ and ζ profiles have two minima related to the *cis* and *trans* isomers of the COOH group: δ and ζ, $C \simeq \pm 180°$ correspond to the *cis* isomer, and $T \simeq 0°$ to the *trans* isomer. We labelled each conformer according to the order δ, β and ζ. Moreover, although slightly redundant, we added extra information at both ends to describe the intramolecular hydrogen bonds that the specific conformer forms, and to make the different interactions identifiable at a glance. Intra-molecular hydrogen bonds were here counted when the donor-acceptor distance was less than a 3.5 Å cut-off and when the donor-H$\cdots$acceptor angle was larger than 100°. Specifically, we labelled type 1 a hydrogen bond where the amino group acted as a donor to the carbonyl oxygen, as in N-H$\cdots$O-C, and type 2 a hydrogen bond where the amino group acted as acceptor and the O-H group as donor, as in O-H$\cdots$N. In type 3 hydrogen bonds, the amino group acted as a donor, but the acceptor was the hydrogenated oxygen of the carboxyl, i.e. N-H$\cdots$O-H. Finally, when the two carboxylic groups were interacting, as in O-H$\cdots$O, the hydrogen bond was of type 4: in this case, the hydrogen bond label was attributed to the donor to avoid double counting. In the few cases where the same group formed multiple hydrogen bonds, both types of bond were noted.
The stable conformers identified with the CHARMM36 force field were then optimized both with DFT and the B3LYP exchange and correlation hybrid functional and with MP2, with a 6-311++ G(d,p) basis set. The corresponding free energies were estimated by adding temperature and entropic effects within thermochemistry algorithms in GAUSSIAN09 [198]. DFT and MP2 calculations in water solution were carried out using a polarizable continuum model. This allowed us to confirm the stability of such isomers and compare energies and ranking obtained with empirical force fields to those obtained with quantum mechanical calculations.

Neutral Asp in Gas Phase: Results and Discussion

In the following section, the three-dimensional free energy landscape of the neutral form of aspartic acid in the gas phase is presented, and two-dimensional maps obtained with sketch-map are also discussed. A total of 32 stable conformers, among which are those identified by experiments, are found, analysed and ranked according to their free energy.

In Fig. 3.37 are the projections of the CHARMM36 free energy maps for neutral Asp as a function of pairs of torsional angles. Although these projections give a rough idea of the arrangement of basins, details regarding the free energy of individual minima and the complexity of the network of paths and barriers among them are lost in the integration procedure over the third variable. To complement these pro-

jected maps and disentangle the network of interconnected basins, a bi-dimensional sketck-map of the free energy landscape as a function of β, δ and ζ was produced with sigmoid function parameters $\sigma = 2.0$, $A = 8.0$, $B = 2.0$, $a = 6.0$, $b = 2.0$ (see Eq. 2.25a and 2.25b). The result is shown in Fig. 3.38. Here, 12 minimal free energy basins are quickly identified and labelled with capital letters in order of increasing free energy; the corresponding labelled structures are also shown.

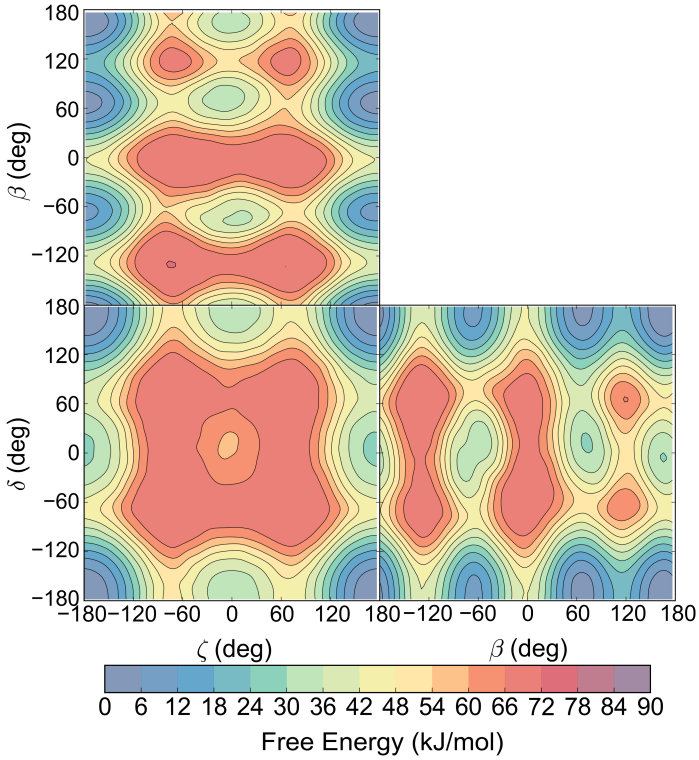


Figure 3.37: Projected free energy maps from metadynamics simulations for neutral Asp in the gas phase as a function of pairs of torsional angles calculated with the CHARMM36 force field. Contour lines are every 6 kJ/mol.

A typical weakness of non-linear dimensionality reduction methods is the difficulty to reproduce the space in presence of topological obstructions without losing information on the distances between points and introducing artifacts, such as tearing, in the connectivity [227]. However, here the sketch-map algorithm is remarkably able to preserve the periodicity of the primary torsional angle β and displays the bi-dimensional sketch-map circularly around it, showing tearing only in the two other torsional variables and making the interpretation of the data and the association of conformers to the minima very intuitive. The map is indeed circular around the value of β, which defines the main backbone variations in the Asp conformers. This is also shown schematically in the right panel of Fig. 3.38; in the central panel the angle distributions onto the sketch-map demonstrate that the specific choice of

sketch-map parameters is able to separate conformers characterised by the three different torsional angles. In particular, the bottom region with respect to the central hole, is mainly characterised by conformers with $\beta \simeq \pm 180°$, the top left region with $\beta \simeq -60°$ and the top right region with $\beta \simeq 60°$. The exception is represented by minima I which has $\beta \simeq 60°$, but is disjointed from the conformers with a similar β value.

Three minima A, B, and C have significant lower free energies than the rest of the map and are characterised by both δ and ζ in the *cis* isomer of the carboxyl group. Transitions between the A and B basins are favoured with respect to transitions to the C basin. The *cis* isomers for both carboxyl groups make the formation of type 2 hydrogen bonds impossible, while type 1 hydrogen bonds are frequent and more favourable with respect to type 3 or 4, which are not observed in these basins. In fact, it is a characteristic of type 2 hydrogen bonds to have the COOH group in a *trans* arrangement; the establishment of a hydrogen bond compensates the higher-in-energy COOH *trans* configuration with respect to the *cis*. The conformer with $\beta \simeq \pm 180°$, i.e. with a fully extended backbone, is the most favourable and its free energy is set as the zero reference. A second concentric tier of several higher free energy minima is found around the three main ones. These contain conformers where either δ and ζ are in the *trans* isomerisation state, suggesting that *trans* isomers are overall disfavoured. This is confirmed also by the presence of a third scarcely populated tier, with only three high free energy minima, with both δ and ζ in the *trans* state; no connection to other basins appears because of the high free energy and consequently insufficient sampling. In this tier, the free energy is up to 55.3 kJ/mol with respect to the zero reference, and few stable conformers can be extracted. The free energy of the minima belonging to the second tier, from D to I is between 30 and 34 kJ/mol and again, type 1 hydrogen bonds are the most frequent, although the presence of hydrogen bond of type 2, 3 and even 4 is observed.

From these 12 basins, 18 distinct stable conformers were extracted, the increased number being due to the free energy degeneracy of conformers with different α or γ angles: each structure is shown in Fig. 3.38, linked to the respective basin. Exemplary is the case of the minimum B, where no intramolecular hydrogen bonds are present and thus two similar conformers, with complementary orientations of α, are observed. Similarly, rotations about the angle ϵ, that defines the position of the amino group hydrogens should introduce degeneracy. However, isomers with different values of ϵ, e.g. with hydrogen bond interactions with either one or the other hydrogen of the amine, show little or no difference in terms of free energy. For the sake of clarity, we decided to disregard this extra degree of freedom which gave no useful information.

In Table 3.4 the free energy of the conformers extracted with metadynamics and sketch-map (on β, δ and ζ) for neutral Asp with respect to their respective

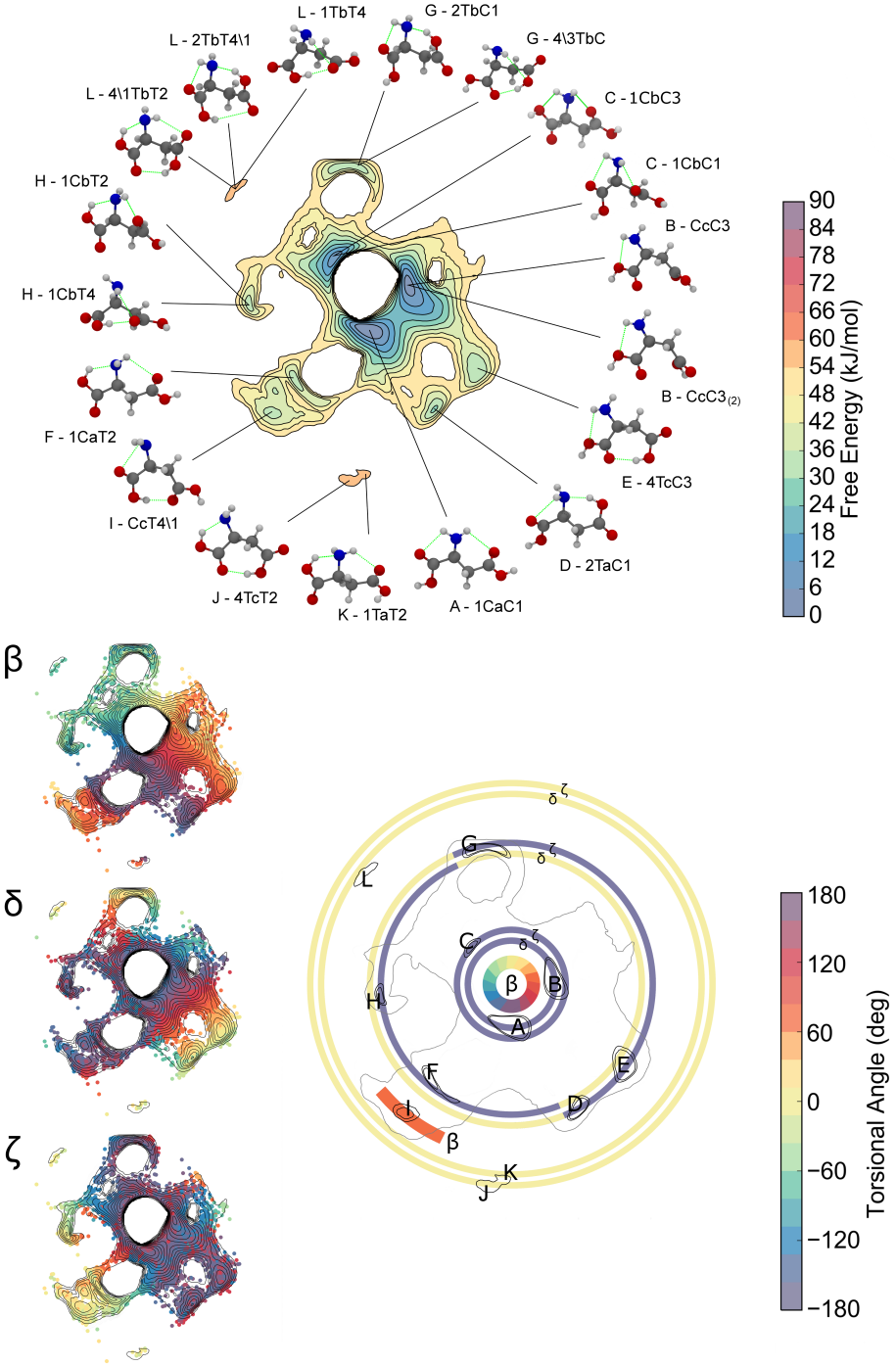


Figure 3.38: Top: Sketch-map of neutral Asp in the gas phase obtained by a dimensional reduction of the free energy as a function of the β, γ and ζ CVs. As before, contour lines are every 6 kJ/mol. The minimal free energy conformers are shown and labelled, with the front letter indicating the name of the basin they belong to; intramolecular hydrogen bonds are in green. Bottom left: distribution of the CV values onto the sketch-map. Bottom right: schematic distributions of the CV values for the minima onto the sketch-map. β values as displayed in the inner circle, with the exception of the minimum I. δ and ζ values are shown by pair of circles at different radii corresponding to the minima locations (δ internal and ζ external circle).

		Free Energy (kJ/mol)		
Basin	Conformer	FF	DFT	MP2
A	1CaC1*	0.0±0.4	2.9	3.8
B	CcC3	2.9±0.8	11.3	9.2
B	$CcC3_{(2)}$	2.9±0.8	17.2	14.7
C	1CbC1*	8.8±1.3	3.3	0.0
C	1CbC3*	8.8±1.3	4.1	1.7
D	2TaC1*	28.9±7.5	3.8	5.0
E	4TcC3	30.6±6.3	25.1	26.4
F	1CaT2*	31.0±9.2	0.0	2.9
G	2TbC1*	31.4±9.2	10.0	8.4
G	4\3TbC	31.4±9.2	24.3	26.0
H	1CbT2	32.7±9.6	1.7	1.7
H	1CbT4	32.7±9.6	18.8	21.4
I	CcT4\1	33.9±12.1	10.5	13.8
J	4TcT2	50.2±10.0	16.7	22.2
K	1TaT2	50.7±8.4	25.1	26.8
L	4\1TbT2	50.7±9.2	18.8	18.4
L	2TbT4\1	50.7±9.2	34.3	32.7
L	1TbT4	50.7±9.2	39.8	42.7

Table 3.4: Free energies of neutral Asp *in vacuo* stable conformers relative to their absolute minimum calculated with the CHARMM36 force field and metadynamics simulations, DFT and MP2. An asterisk has been added to the conformers identified experimentally.

absolute minimum are shown. The list includes the Boltzmann averages and errors of free energies calculated with metadynamics and the CHARMM36 force field (FF) and those obtained from *ab initio* thermochemistry calculations of the optimized structures, with DFT-B3LYP and MP2. Errors on the metadynamics based averages are weighted over the relative number of conformations in each basin to account for the fact that the reproduction of higher energy basins in the map is less accurate when less points are available. *Ab initio* calculations confirm that the identified conformers are all stable. However, there are some discrepancies between the free energies obtained with the force field and those obtained with *ab initio* calculations in the free energy ranking; DFT and MP2 results are overall in good agreement, with differences mostly within chemical accuracy (< 4 kJ/mol). According to these calculations, 1CbC1 and 1CaT2 found in the basins C and F are the absolute minima for MP2 and DFT-B3LYP respectively, closely followed by 1CaC1 in basin A. Nevertheless, when making these comparisons, one should bear in mind the differences between the adopted methodologies; while the force field metadynamics simulations were carried out at finite temperature and the free energies were calculated directly on the basins, the B3LYP-DFT and MP2 calculations were performed over the minimal potential energy structure at absolute zero, with temperature effects and entropy corrections added *a posteriori*, disregarding e.g. anharmonic effects.

Experimentally, six conformers were observed with rotational spectroscopy. The results are summarised in the work of Ref. [72], where the estimated relative abundances did not fully correlate with *ab initio* free energy calculations at the MP2 level. During these experiments, conformational relaxation of higher energy conformers to the lowest energy ones can happen if the transition paths between the two basins are sufficiently low. As a consequence, only the most populated conformers can be resolved and their populations are affected by the presences of these relaxed structures. Using our nomenclature, the relative abundances estimated in experiments follow the ranking 1CaC1 > 1CbC1 ≈ 2TaC1 > 1CaT2 ≈ 2TbC1 > 1CbC3. Correlating relative free energies and abundances, this trend is in overall good agreement with the results obtained with metadynamics and the CHARMM26 force field, with the exception of 1CbC3, which belongs to basin C and is thus lower in free energy than expected. The choice of mapping only the metadynamics CVs makes this conformer degenerate with 1CbC1, from which it differs only for the orientation in α of the carboxyl group.

Sketch-map with 5 Collective Variables

A consequence of mapping the free energy as a function of only three dihedral angles is the aggregation of different of conformers within the same basins. If such structures are sufficiently populated, the cluster analysis allows to discriminate them and highlight the degeneracy of the minimum. Higher free energy conformers, which may have been visited less during the simulation, however, could be hidden in lower energy basins by the presence of more abundantly populated conformers. In order to identify other potentially missing conformers and eliminate all degeneracies, the free energy was remapped starting from the full space of the torsional angles. In Fig. 3.39, the free energy maps reweighted as a function of the neglected angles α, γ and ϵ can be found. These maps, which were produced with the aid of an algorithm explained in Sec. 2.2 and developed in Ref. [150], show that these alternative torsional angles are adequately sampled during the metadynamics runs, confirming the choice of CVs as appropriate and allowing for a remapping of the free energy as a function of such angles. As already mentioned, we chose to neglect ϵ that defines the rotation of the amino group to simplify the representation. This angle gives no additional useful information, as its presence would just duplicate the map symmetrically according to which of the two hydrogens in the amine is forming hydrogen bonds. We thus worked with the remaining five angles, with sketch-map parameters $\sigma = 2.0$, $A = 6.0$, $B = 2.0$, $a = 4.0$, $b = 2.0$.

The resulting sketch-map is shown in Fig. 3.40. The resolution of the first two layers is increased and conformers that were degenerate in the previous map are now localised in different minima, although information on the highest free energy conformers, namely those belonging to basins J,K and L, and characterised by both

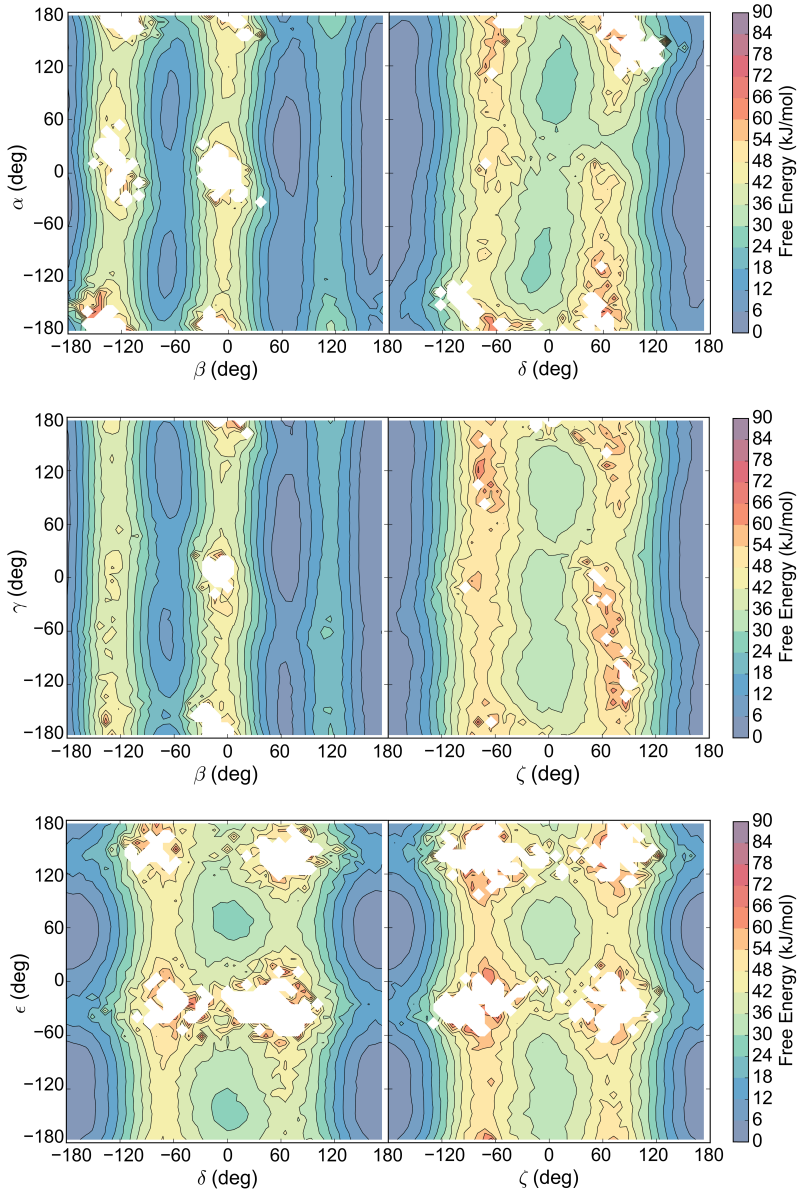


Figure 3.39: Free energy maps for neutral Asp in the gas phase from metadynamics reweighted as a function of the neglected torsional angles α (top), γ (center) and ϵ (bottom) against those who were chosen as CVs (β, δ and ζ) with the CHARMM36 force field. The zero of the free energy scale corresponds to the absolute minimum for each map. Contour lines are every 6 kJ/mol.

δ and ζ in the *trans* conformer, is lost at the borders of the map. The range of free energies was consequently reduced in Fig. 3.39 to identify more easily different basins. This map also helps understand how the choice of the mapped variables is crucial in defining at which level the separation of the basins will be carried out. Working with more variables increases drastically the number of lower free energy minima that the algorithm is trying to correctly map, resulting in a loss of information on the highest free energy minima, where less landmarks are found and which consequently weight less in the minimisation procedure.

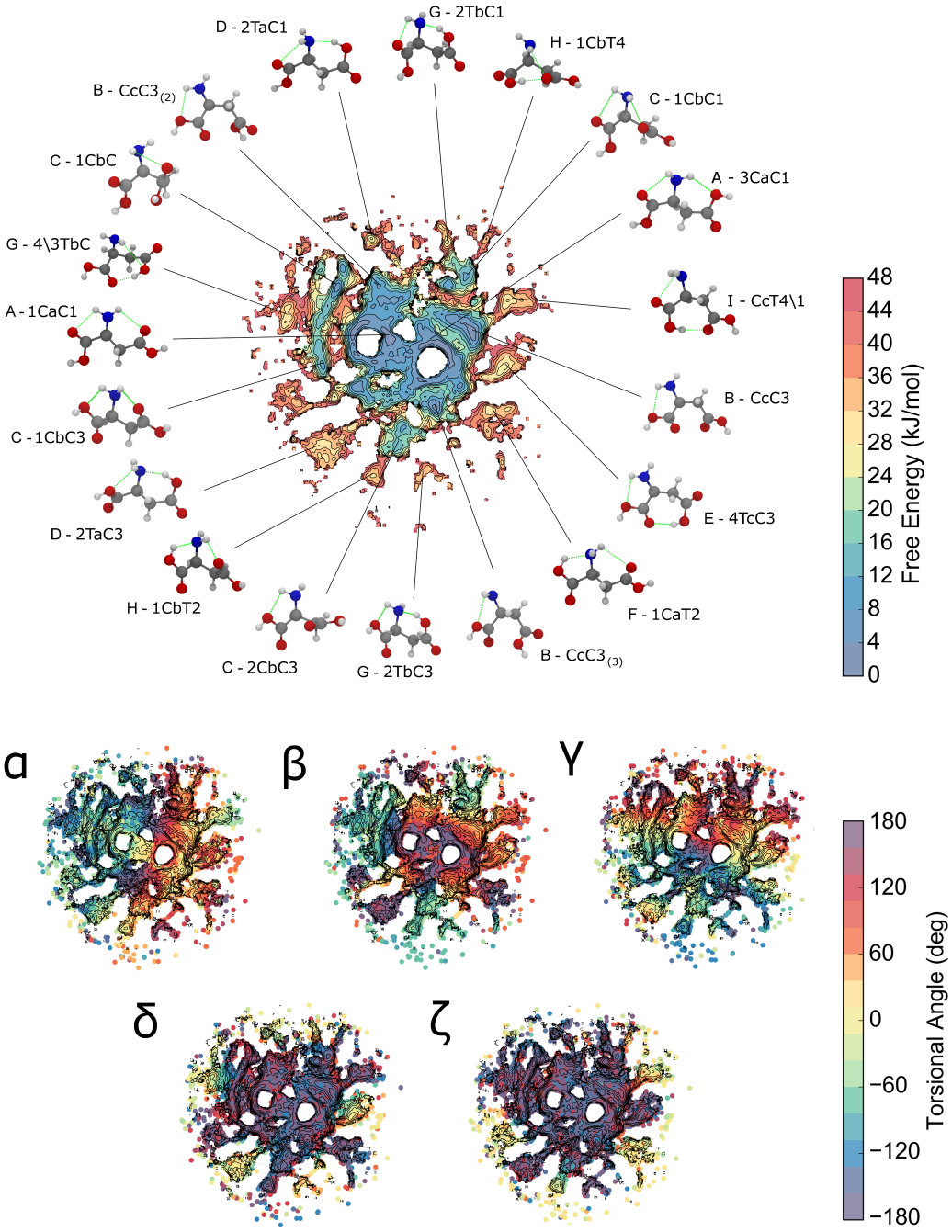


Figure 3.40: Top: Sketch-map of neutral Asp in the gas phase obtained from the five-dimensional space of the torsional angles α, β, γ, δ and ζ. The minimal free energy conformers are shown and labelled, with the front letter indicating the name of the original basin they belonged to in the previous map; intramolecular hydrogen bonds are in green, while the colour range has been reduced to improve clarity. Bottom: distribution of the values of α, β, γ, δ and ζ onto the sketch-map.

A total of 19 stable conformations were characterised thanks to this sketch-map, some of which were not visible in the previous representation, hidden by degeneracies or distortions in the mapped basins. To avoid confusion with the basin names of the previous sketch-map, no labels have been used in this case, but the previous nomenclature was maintained.

		Free Energy (kJ/mol)		
Basin	Conformer	FF	DFT	MP2
A	**3CaC1**	0.0 ± 0.8	12.6	12.6
A	1CaC1*	0.8 ± 0.8	2.9	3.8
B	CcC3	2.1 ± 0.8	11.3	9.2
B	**CcC3$_{(3)}$**	2.9 ± 2.5	11.3	9.2
C	**1CbC**	6.3 ± 1.7	13.8	10.5
B	CcC3$_{(2)}$	6.7 ± 3.3	17.2	14.2
C	1CbC3*	10.9 ± 5.0	4.1	1.7
C	1CbC1*	11.7 ± 3.3	3.3	0.0
C	2CbC3	12.9 ± 10.0	4.1	12.6
D	2TaC1	24.7 ± 10.0	3.8	5.0
E	4TcC3	25.5 ± 3.3	25.1	26.4
G	2TbC1*	27.6 ± 8.0	10.0	8.4
G	4\3TbC	28.9 ± 9.2	24.3	26.0
I	CcT4\1	28.9 ± 5.0	10.5	13.8
D	**2TaC3**	30.6 ± 13.4	10.5	11.3
F	1CaT2*	30.6 ± 6.3	0.0	2.9
H	1CbT2	32.7 ± 8.0	1.7	1.7
H	1CbT4	33.1 ± 8.0	18.8	21.3
G	**2TbC3**	33.5 ± 9.2	16.7	14.7

Table 3.5: Free energies of the conformers of neutral Asp in the gas phase, identified in the sketch-map of Fig. 3.40, relative to their absolute minimum, calculated with the CHARMM36 force field and metadynamics simulations, DFT and MP2. An asterisk has been added to the conformers identified experimentally, while the conformers hidden in the previous 3CVs sketch-map are in boldface.

As before, the map can be separated in a central region, where lower free energy basins with both δ and ζ $\pm 180°$ are found, surrounded by higher free energy basins with δ and ζ alternatively in the *cis* or *trans* conformation scattered at the border. The centre is characterised by two circular holes, around which the sketch-map algorithm arranged the periodicity of the γ angle, making positive values fall on the top part of the map and negative values at the bottom, while α is arranged horizontally, with values closer to 90° on the right hand side and about $-90°$ on the left hand side. These circles are surrounded by a layer of basins containing the conformers characterised by β $\pm 180°$, previously belonging to basin A in Fig. 3.38. Here different conformers are arranged according to the α and γ angle, although the difference in free energy is almost negligible, making them interchangeable at room temperature. Farther apart, but still closely interconnected are the conformers belonging to basin B of the 3CVs sketch-map, and even more those belonging to C,

whose transition paths are mostly lost due to the necessary distortion of this reduced representation. Thanks to the separation of α and γ previous degeneracies are lifted, and conformers that were hidden before by other, more favourable structures are now easily identified. The case of 1CbC3 and 1CbC1 is exemplar: the two conformers, which differ only for the value of α, $-70°$ and $0°$ respectively, were previously both found in basin C and are now pushed in two opposite regions of the map. The choice of using five variables instead of three allows thus to increase the resolution on the lowest energy conformers and easily discriminate between different structures, at the price, however, of losing the highest free energy basins.
In Table 3.5 the complete list of conformers identified in this new sketch-map, with their respective force field and *ab initio* free energies, is shown. To simplify the comparison, at each conformer is associated a basins label according to the previous 3CVs map in the first column. The number of previously unknown low energy conformers extracted from this second map highlights how the choice of different variables in the mapping of the free energy landscape can affect the efficiency of the structures extraction and should be thus carried out with extreme care.

Zwitterionic Asp in Water: Results and Discussion

In this section, the free energy landscape of the zwitterionic form of aspartic acid in water is presented and 10 stable conformers are identified. The results are then compared with the data obtained from the neutral form in the gas phase.

The free energy surface obtained with metadynamics of the solvated Asp in its zwitterionic form is shown in Fig. 3.41 (top panel) as a function of β and δ; here, six minima can be observed. The symmetry of the COO^- carboxylate moiety allows to reduce the number of CVs to two (β and δ) without losing crucial information on the system; ζ is non existent in this system, due to the transfer of the carboxyl oxygen to the nitrogen group. There is in principle no need to use sketch-map in this case because the map is already bi-dimensional; however, in order to clarify the process of deformation involved in the application of the algorithm, a two CV sketch-map, computed with sigmoid parameters $\sigma = 2.0$, $A = 8.0$, $B = 2.0$, $a = 6.0$, $b = 2.0$, is also shown in Fig. 3.41. This sketch-map transforms the toroidal surface of the two periodic CVs into a map in the Euclidean plane.

10 stable conformers, which account for the free energy degeneracy with respect to the α torsion, were extracted with a structural cluster analysis and associated to their respective minima. A further step to remap the free energy starting from the four-dimensional space of the α, β, γ and δ angles to eliminate degeneracies and focus on the lower energy basins, however, is not necessary here as tests confirmed that this list exhausts all the favourable isomers of zwitterionic Asp. The only two remaining conformers, obtained from the rotation of the α angle in basins D and E, require the

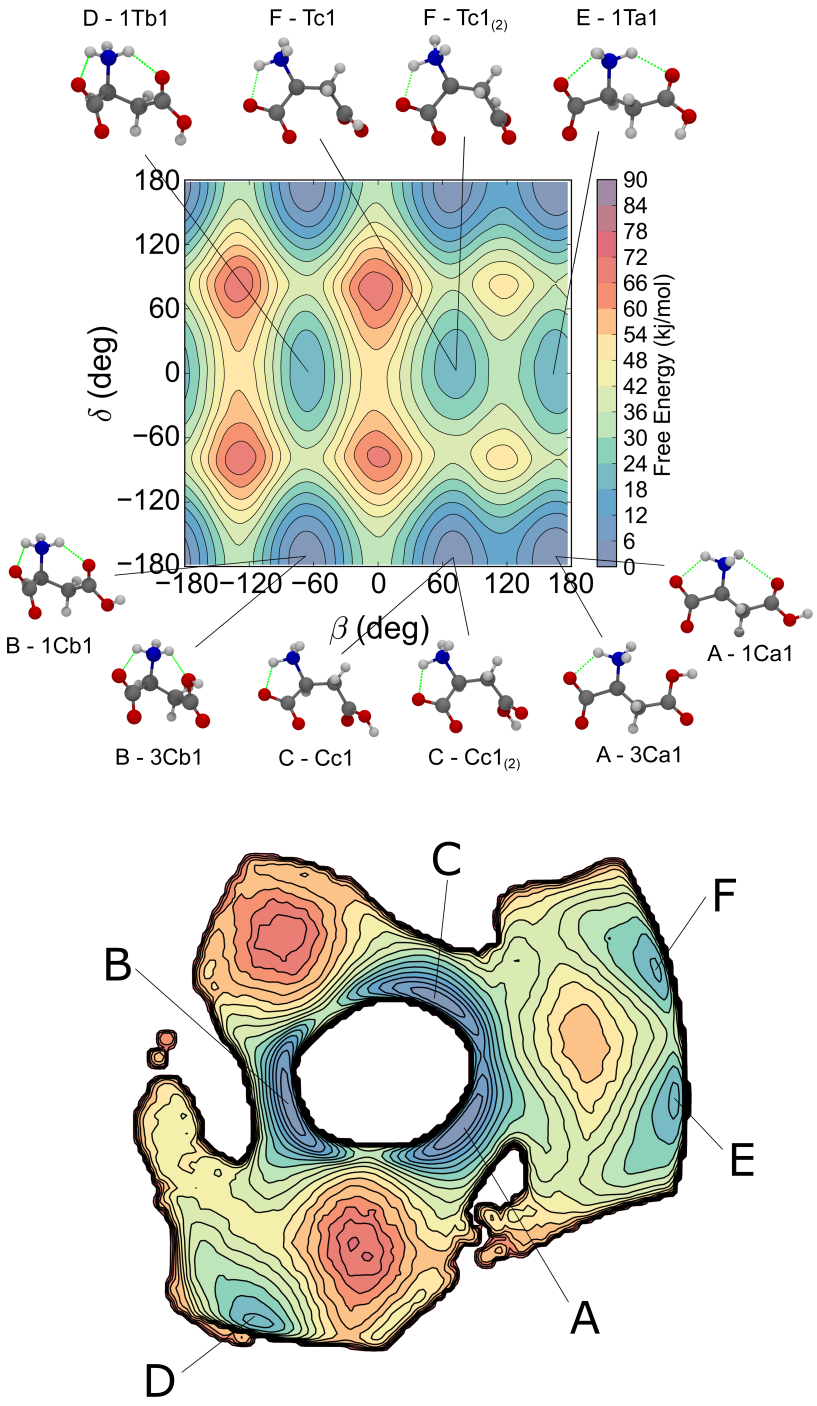


Figure 3.41: Top: Free energy surface of zwitterionic Asp in water as a function of β and δ. The conformers extracted are shown and labelled with the front letter indicating the basin they belong to. Intra-molecular hydrogen bonds are shown in green. Bottom: bi-dimensional representation of the free energy obtained with sketch-map.

carboxyl hydrogen to move closer to the positive amino group, in highly unfavourable conformations. In all the stable isomers, a hydrogen bond between the positively charged NH_3 moiety and the negatively COO^- group is present. This interaction is stabilized in water and does not prevent the formation of further hydrogen bonds with the surrounding water molecules.

		Free Energy (kJ/mol)		
Basin	Conformer	FF	DFT	MP2
A	3Ca1	0.0 ± 2.1	0.8	3.8
A	1Ca1	0.0 ± 2.1	6.3	7.5
B	1Cb1	1.7 ± 2.1	0.0	0.0
B	3Cb1	1.7 ± 2.1	6.7	5.4
C	Cc1	2.1 ± 2.1	8.4	10.9
C	$Cc1_{(2)}$	2.1 ± 2.1	10.0	11.7
D	1Tb1	18.8 ± 2.1	10.5	9.6
E	1Tc1	19.3 ± 2.1	18.4	22.2
F	Ta1	20.1 ± 2.1	13.0	15.1
F	$Ta1_{(2)}$	20.1 ± 2.1	13.0	15.1

Table 3.6: The free energies for the stable conformers of zwitterionic Asp in water, calculated with metadynamics and the CHARMM36 force field, DFT and MP2 (with implicit solvent), with respect to their respective absolute minima.

When comparing the metadynamics map in Fig. 3.41, with the projection of the three-dimensional free energy landscape of neutral Asp in the gas phase onto the β-δ plane in Fig. 3.37 (right), interesting similarities emerge. The main change seems to be the decreased free energy difference between *trans* and *cis* isomers in solution with respect to the gas phase. The free energy has two regions corresponding to the *trans* and *cis* values of δ and, as in the gas phase, the *cis* isomers seem favoured. In the three lowest free energy basins A, B and C, the carboxyl group is indeed in the *cis* isomers, while β has values of $\pm 180°$, $\simeq -60°$ and $60°$ respectively. All three minima have similar free energies (within 2.1 kJ/mol) and again are arranged circularly in the sketch-map. Basin A contains conformers with $\beta \simeq \pm 180°$ with the possibility to form a type 1 or a type 3 hydrogen bond involving the COOH group. Similarly, both B and C comprised degenerate conformers characterised by the two possible orientations of α. Type 3 hydrogen bond was observed in B, while in C no hydrogen bond interaction between the amino group and the carboxylic group can occur. This similarity in the free energies of the A, B and C basins may be facilitated by the presence of water molecules, which provide a favourable environment for Asp through numerous hydrogen bond interactions with both the carboxyl groups and the amine. For the *trans* conformers in δ (D and E), the double degeneracy in α is lost as orienting the OH group near the NH_3 moiety would draw the hydrogens of both groups too close: the repulsion of their positive charges would prevail making these conformers unstable and only type 1 hydrogen bonds were thus observed. This

is not the case for basin F, where, as for basin C, the configuration of β prevents the NH_3 and the COOH groups to interact; hence the double-degeneracy with respect to the α-torsion is preserved.
In Table 3.6 the free energies for zwitterionic Asp in a water solution are shown for the 10 distinct conformers obtained from metadynamics and *ab initio* calculations with implicit solvent. The free energies calculated with the different methods are more similar in water solution than *in vacuo*. This may be due to the force field parametrization, which is primarily developed for use within a solved biomolecular environment rather than in the gas phase, and may be therefore more appropriate in this case.

Sketch-map as a Function of α, β and γ

In Fig. 3.42 the free energies of both the zwitterionic system in water and the neutral system in the gas phase are mapped via sketch-map starting from the three-dimensional space of α, β and γ. While the absence of ζ in the zwitterionic simulations prevented a direct comparison of the previous sketch-map, working with these angles allows us to observe how the free energy is modified by the protonation state of the molecule and the presence of water. Neglecting δ and ζ, whose isomerisation heavily affects the free energy landscape, however, causes a not indifferent loss of information, as most basins are now hidden by lower energy minima previously labelled in both systems as A, B and C. This makes it more difficult to distinguish quantitatively among different conformations and requires to decrease the range of free energy in Fig. 3.42.

The most evident effect of the proton transfer is the absence of half of the map in the zwitterionic simulation with respect to the vertical axis. This is due to the formation of a stable hydrogen bond between the amino group and the negative carboxyl group. γ, the angle defining the rotation of such group, is bound to explore only values around 120° , as observed in Fig. 3.42 (center) since it was not biased during the metadynamics and couldn't escape its starting free energy minimum. The symmetry of this moiety, however, guarantees that the second half of the map would be identically mirrored horizontally. This is intuitively not the case for the neutral map, where differences between the two halves arise from the presence of the hydroxyl hydrogen.
While γ is mapped horizontally, α is instead arranged vertically, with angles toward negative values mapped on the top half, while conformations with positive values of α are mapped on the lower half. The basins belonging to different values of β are also easily discernible.
Another interesting feature is the lowering of certain regions of the sketch-map. In particular, at the top, a basin containing 1Cb1 and 1Tb1, previously belonging to basins A and D, in the middle, the basin containing $Cc1_{(2)}$ and $Tc1_{(2)}$, previously

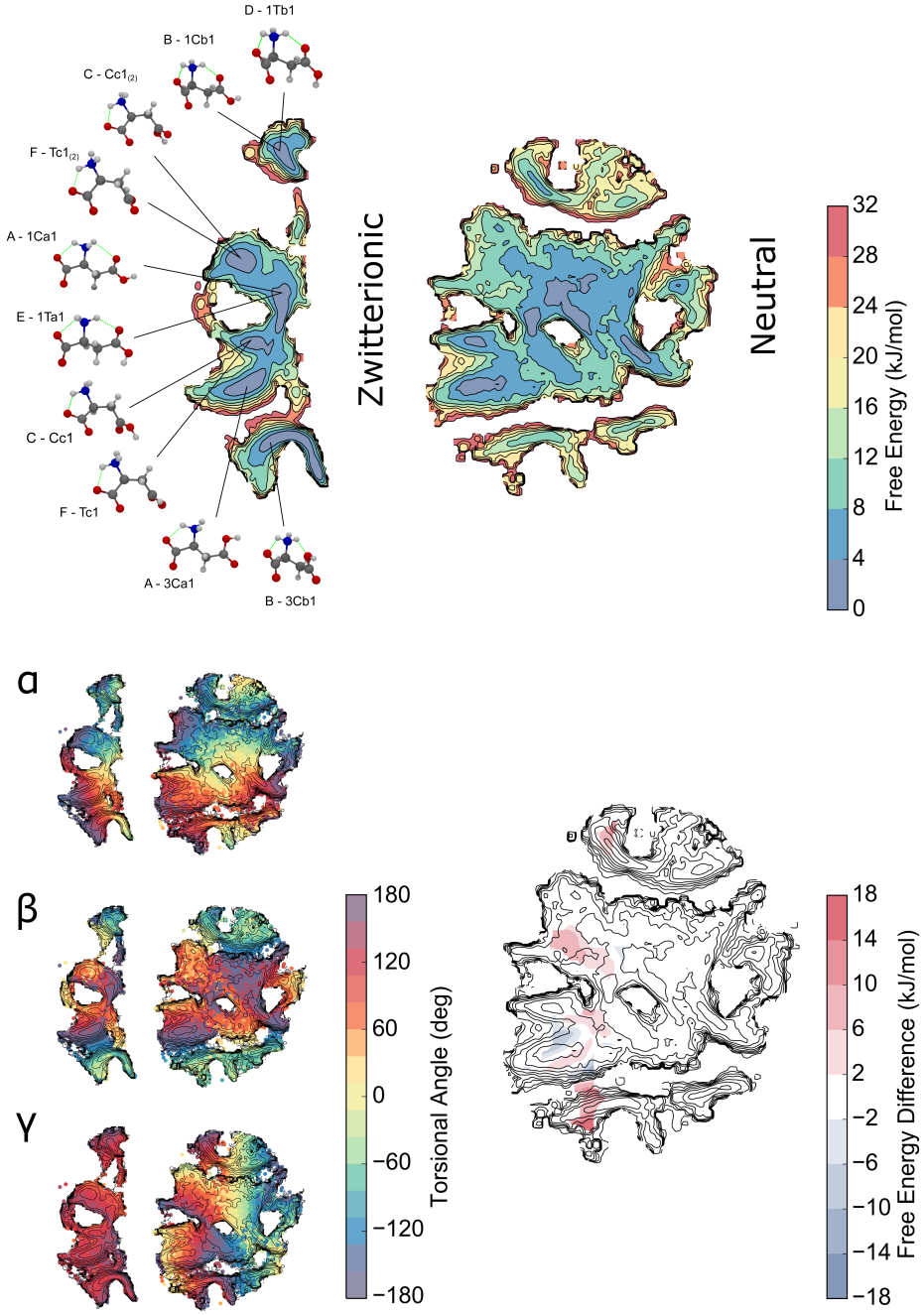


Figure 3.42: Top: Sketch-map of zwitterionic Asp in water and its neutral version in the gas phase obtained starting from the three-dimensional space of the torsional angles α, β and γ. The minimal free energy conformers obtained in the zwitterionic form are shown and labelled to help the comparison; intramolecular hydrogen bonds are in green, while the colour range has been reduced to improve clarity. Bottom left: distribution of the values of α, β and γ onto the sketch-maps. Bottom right: Free Energy difference between the neutral Asp *in vacuo* sketch-map and the solvated zwitterionic sketch-map, calculated as $\Delta F_{Neutral} - \Delta F_{Zwitterionic}$.

in C and F respectively, and finally at the bottom of the map, a basin containing 3Cb1, previously in B. It is safe to assume that the lowering of such minima is due to the presence of 1Cb1, $Cc1_{(2)}$ and 3Cb1, rather than the other three conformers, as these are the most populated conformations and their contribution dominates the free energy. In this map all minima are quite close in free energy, as already observed in the analysis of the bi-dimensional map from metadynamics of zwitterionic Asp, in line with the idea that the presence of water increases the flexibility of aspartic acid, thanks to the numerous favourable intermolecular hydrogen bonds it can form, and allows it it to easily access basins that would otherwise be less populated. A more in-depth analysis of the role of water molecules can be found in a later section.

Unphysical Systems: Results and Discussion

Here, the free energy maps obtained from simulations of two unphysical systems, the neutral aspartic acid in solvent and the zwitterionic form in the gas phase, are presented and compared with the physical systems analysed before.

The preset unbreakable bonds of the force field allow us to study extremely unfavourable environments: two control simulations of unphysical systems, neutral Asp in water and zwitterionic Asp in the gas phase, have been thus carried out (using the same simulation protocol as before), to analyse the behaviour of this amino acid from a different perspective. It is interesting to analyze the behaviour of Asp in such extreme conditions as they can provide information on the transformation between the neutral and the zwitterionic form.

The map for solvated neutral Asp can be seen at the top of Fig. 3.43 and, when compared with the sketch-map obtained in the gas phase, it allows us to understand the effect of the water environment. The positions of the minima are consistent with those of the gas phase and so is their ranking in most cases, but the relative free energy differences between *trans* and *cis* conformers (in δ and ζ) are smaller. Indeed, the energy difference of the *cis-trans* isomerisation (in δ or ζ) products is about 16 kJ/mol, with respect to 30 kJ/mol *in vacuo*, in line with the effect that water on the *trans* isomers observed in the previous section for zwitterionic Asp in solution. This is likely to be due to the occurrence of hydrogen bonds interactions between the carboxyl groups and the water molecules, which weaken the intramolecular hydrogen bond network, previously seen to define the morphology of this map in the gas phase, and diminish its effect on the Asp free energy. As a consequence new minima of relatively high free energy, which were not visited in the gas phase scenario, are sampled at the bottom and at the periphery of the sketch-map.
The transition paths connecting the basins of the second layer are also lowered by this effect, allowing for higher flexibility of the δ and ζ angles. The *cis-trans* transition of these two angles is thus more favoured, although the hydrogens still needs to

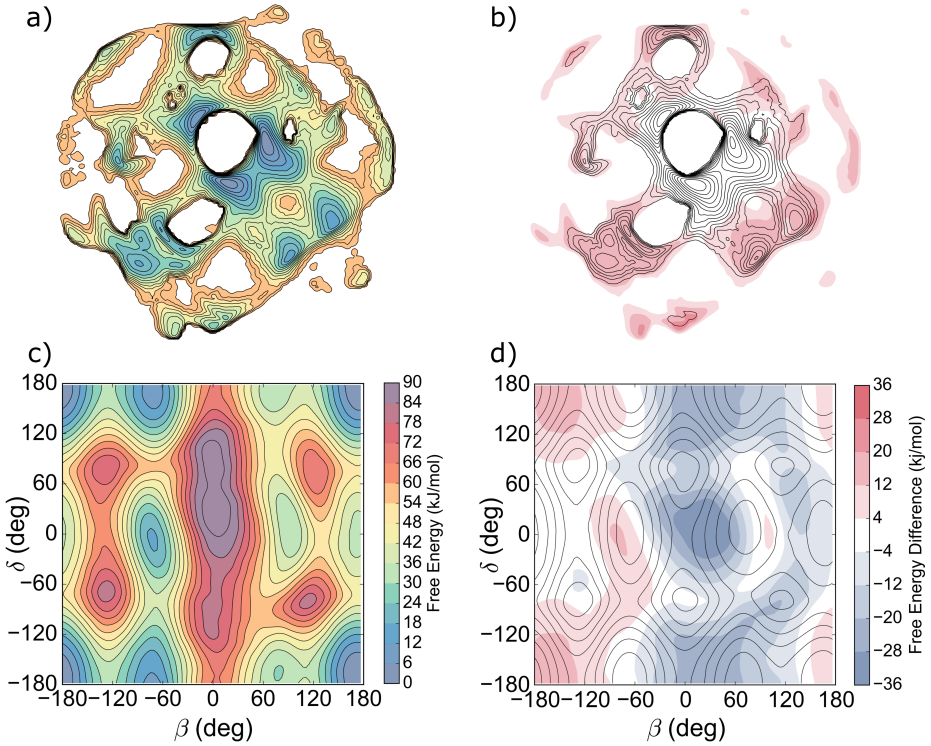


Figure 3.43: Top: (a) free energy sketch-map of neutral Asp in water and (b) its comparison with the corresponding sketch-map in the gas phase $\Delta F_{gas-phase} - \Delta F_{water}$. Bottom: (c) free energy map from metadynamics as a function of β and δ of zwitterionic Asp in the gas phase and (d) its comparison with the corresponding map in water $\Delta F_{water} - \Delta F_{gas-phase}$. The color range goes from blue, where the original physical systems have lower free energy, to white and red, where the physical systems were higher in free energy.

overcome a 25 kJ/mol barrier to change position. The *trans* conformation of δ is expected to facilitate a direct proton transfer to the nitrogen group during the formation of the zwitterionic molecule. One could sensibly speculate that this is the favoured proton transfer mechanism, however, conformers characterised by direct intramolecular hydrogen bonds between the amino moiety, acting as acceptor, and any of the hydroxyl groups, as donor, are the minima with the highest energy.
An alternative description reconciles the proton transfer with this free energy ranking, hypothesising the intervention of chains of water molecules as proton carriers. Being the rate of isomerisation too slow, the carboxyl group could donate the hydrogen to the solvent, which could then transport and give it to the amine. A study of the incidence of such water chains, in the following section, aims at clarifying this point.

In the map for zwitterionic Asp *in vacuo* at the bottom of Fig. 3.43, differences with respect to the case in water can be clearly seen. Although the location of the minimal free energy basins remains the same and the absolute minimum still corresponds to basin A, the basins are deformed and there is a change in their relative free

energy and ranking. While B was before the second deepest basin, it is here the fourth, with C and D which become instead the second and third lowest free energy minima. In particular, the conformers with $\beta \simeq 60°$ are destabilized with respect to the water solution. The map loses symmetry because of the presence of a wide barrier at $\beta \simeq 0°$ affecting especially the basins at $\delta \simeq 0°$, which are now narrower but higher in free energy for $\beta \simeq 60°$ and 180°. Conformational changes involving β torsions between $\simeq 60°$ and $\simeq -60°$ are hindered by this barrier. All the minimal free energy conformers present a direct interaction between the COO^- group and the NH_3 moiety, which would lead, in unconstrained conditions, to a direct proton transfer without barrier as easily demonstrated with *ab initio* methods.

Solvation Effects

In this section, the solvation preferences of the conformers identified for both the zwitterionic and neutral molecules are investigated. The results suggest that the proton transfer mechanism between the neutral and zwitterionic forms does not rely on the mediation of chains of water molecules, but it is more likely to happen as a direct exchange between the donor and the acceptor parts of the molecule.

With the help of enhanced sampling methods, new possibilities to study the solvation properties of small peptides are now open. While first principle methods can simulate the forming and breaking of bonds and describe explicitly the proton transfer, their cost often represents a problem. To date, it is still unclear how many water molecules are necessary to favour the proton transfer from the carboxyl group to the amine, thus transforming Asp from neutral into its zwitterionic form, and which proton transfer mechanism is activated when such critical quantity is reached.
Given the height of the barriers separating different isomers, if conformational changes are involved in this mechanism, the timescale accessible to *ab initio* calculations may not be sufficient to describe the whole process. Preliminary DFT calculations seem to agree with this hypothesis: the occurrence of the proton transfer was observed to be dependent on the initial configuration of both Asp and the first surrounding water shell. On the other hand, MD and metadynamics simulations can easily accelerate the structural rearrangement that the molecule undergoes during longer time ranges, while missing on the actual proton transfer if performed with conventional non-reactive force fields. A comparison of the results here obtained between the zwitterionic form and the unphysical neutral Asp in water, however, can give information on favourable arrangements of both solvent and solute and clarify to what extent the water molecules may be involved in the deprotonation.
In Fig. 3.44, the relative occupancy of water mediated hydrogen bonds observed during the simulation are shown. The results are grouped according to the different basins identified by metadynamics in the map of Fig. 3.41, and they give a

glimpse of the solvation properties of Asp under these specific conditions. A 3.5 Å donor-acceptor distance and 120°donor-hydrogen-acceptor angle cut-offs were used to define the intermolecular hydrogen bonds and hence the chains of water molecules connecting a donor atom to its acceptor. All conformers are characterised by a relatively high number of water chains connecting the nitrogen to the first oxygen O7 of the deprotonated carboxyl group, which also forms a direct interaction. This suggests that the preferred path for the exchange of proton is between these two moieties.

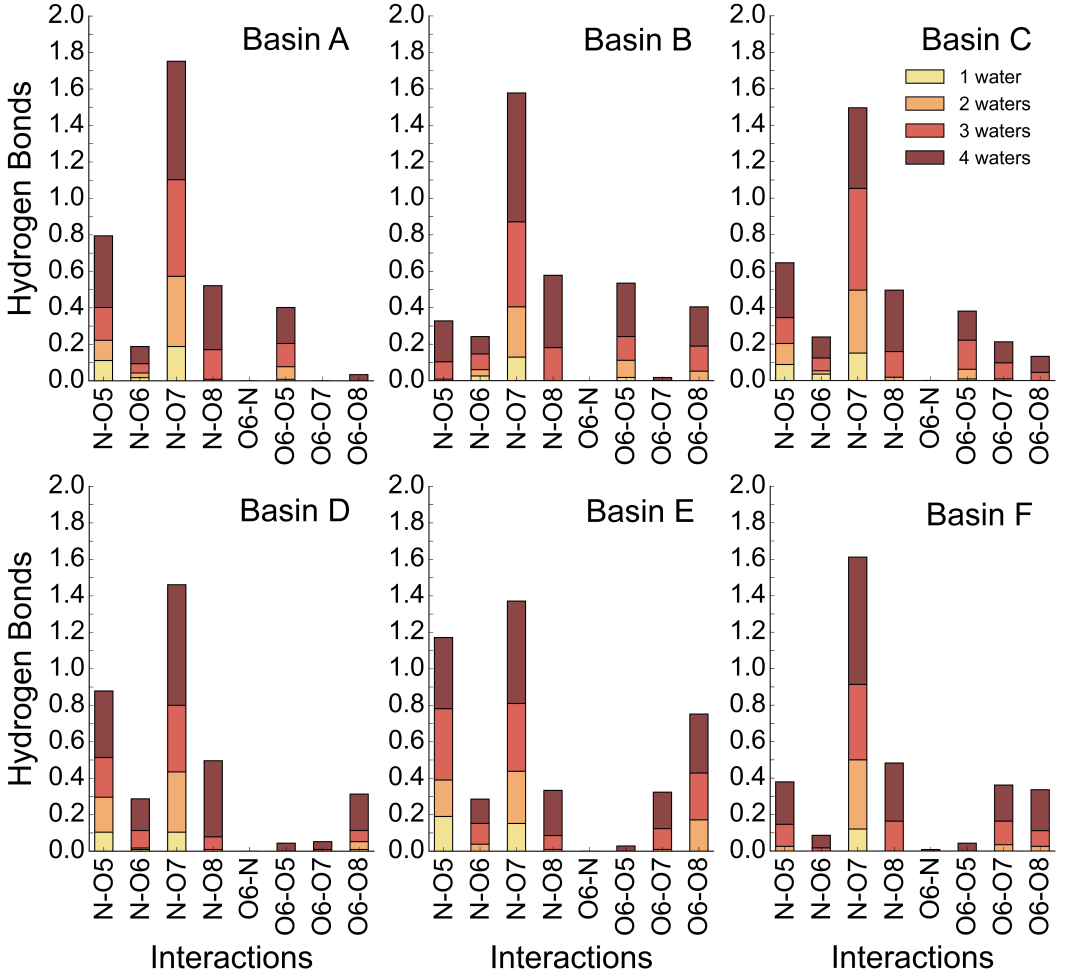


Figure 3.44: Average numbers of water-mediated intramolecular hydrogen bonds per free energy basin in the metadynamics simulation of the zwitterionic form of Asp in water with the CHARMM36 force field. The results include chains formed by up to four water molecules.

Interestingly, the chains connecting the nitrogen to the second carboxyl oxygen O8 are scarce and require a high number of water molecules. This is in line with the idea that the rotation of the γ angle is in this case disfavoured by the presence of strong interactions between the amino group and the carboxyl group. In our simulations of zwitterionic Asp, the rotation about γ was not observed, but

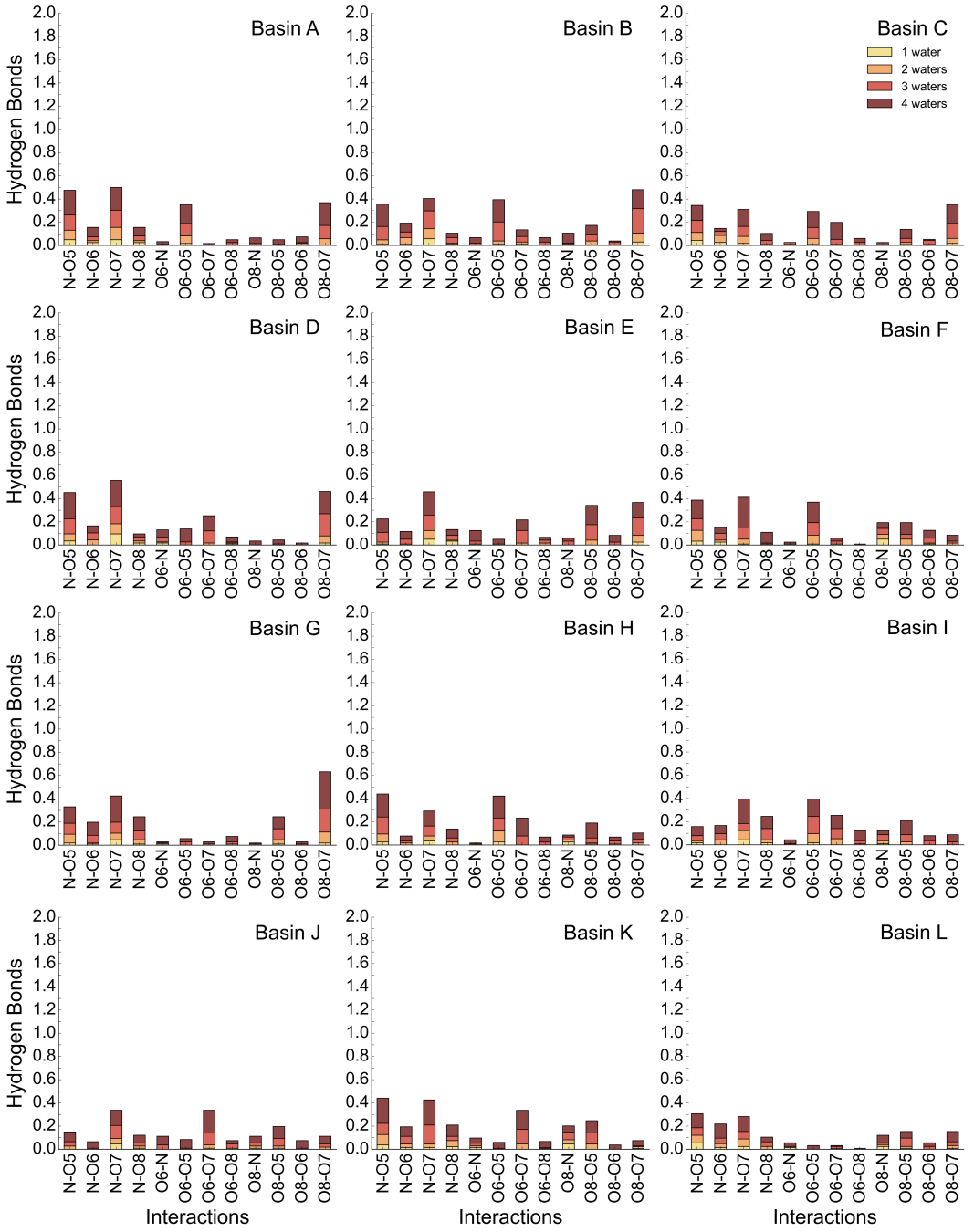


Figure 3.45: Average number of water-mediated intramolecular hydrogen bonds per basin during the metadynamics simulation of the neutral form of Asp in water with the CHARMM36 force field. The results include chains formed by up to four water molecules.

its symmetry assures that no information about low energy conformers was lost. Interactions with the second, protonated, carboxyl group show a slight preference in forming water mediated interactions with the unprotonated oxygen O5 rather than the protonated O6 in most basins, in some cases with the aid of only one water molecule. This preference becomes sharper in the case of *trans* conformers, due to the positioning of the oxygen externally to the carboxyl group, which makes it difficult to form donor-acceptor interactions with the amino group even with the aid of water molecules. The exceptions are basin B and F, where the bent conformation of the backbone (due to the value of β at 60°) makes it difficult to interact with both O5 and O6, decreasing, as a consequence, both occupancies. Isomers with water chains between the two carboxylic groups are in all cases scarcely populated and interactions between oxygens O5 and O6 of the protonated carboxyl moiety are mediated with difficulty, especially when the hydrogen is external, i.e. in the *trans* conformation.

The neutral form of aspartic acid disrupts the network of hydrogen bonds in the surrounding water molecules when put in an unphysical solvated environment. This is reflected in the substantially lowered number of water mediated hydrogen bonds, as it can be observed in Fig. 3.45. In this case, even the more favourite interaction chains are rarely populated for more than 50% of the basin conformers, with hydrogen bonds mediated by a single water molecule for less than 10% in all cases. Interestingly, the path connecting the protonated carboxyl oxygen O8 to the nitrogen, the donor-acceptor couple that is expected to transfer the proton during the transformation into zwitterion, is one of the less populated by water chains, both in the basins containing conformers with that carboxyl group in *cis* and in *trans* configuration. This seems to suggest that water chains as carriers of the proton may be generally unlikely with respect to direct proton transfer, independently on the conformation assumed by the molecule.
These speculations could only be confirmed by a complete set of first principle calculations describing proton transfer and taking into account different isomers including the *cis-trans* isomerisation of δ, which is clearly the key process to understand the proton transfer. A set of initial structures, with different water shell arrangements, may be extracted by the simulations presented here, as starting point for such simulations.

Force field assessment

In the following section, the free energy landscapes of the neutral aspartic acid in the gas phase obtained with three different force field parametrisations are compared. The results show consistency among the force fields in the identification of the basins, but discrepancy in their free energy values and ranking.

As already observed, in the sketch-map of neutral Asp in the gas phase in Fig. 3.38 it is notable the arrangement of basins in tiers, which reflect an increase of about 25 kJ/mol per tier in the free energies of Tab. 3.4. From this data, one could argue that the CHARMM36 force field penalizes the *trans* isomer of the carboxyl COOH groups with respect to the *cis* by such an amount, increasing the free energy of the basin by $\simeq$ 25 kJ/mol whenever one of the two carboxyl groups assumes the *trans* configuration or by $\simeq$ 50 kJ/mol when both do, as in the case of the external tier. The force field parametrization is, however, optimized for solvated proteins rather than gas phase simulations; hence, this penalty may arise from its use *in vacuo*, where the absence of networks of hydrogen bonds with the surrounding water molecules may affect the simulation. In fact, the equivalent results for zwitterionic Asp in water seems to confirm this, as the difference between basins with the δ CV in *cis* or *trans* configuration is lowered to around 16 kJ/mol.

As a final assessment of the selected parametrisation, the neutral Asp *in vacuo* was simulated with two other common force fields, AMBER ff99SB-ildn and OPLS-AA. In spite of the differences in the relative free energies, all three force fields were consistent in maintaining the locations of the principal minima and maxima. This can be observed in Fig. 3.47, where the sketch-maps obtained with AMBER ff99SB-ildn and OPLS-AA are compared with that of CHARMM36. It is evident that the energy penalty of the *trans* with respect to the *cis* isomer is decreased by AMBER, but it is sensibly increased by OPLS-AA. This may suggest that the AMBER force field parameters may be overfitted to reproduce the behaviour in water under any circumstance. The penalty for the isomerisation of the carboxylic hydrogens is reduced consistently, similarly to what was observed in the case of solvated systems with CHARMM36, which modifies the ranking of the extracted conformers, although a significant discrepancy with the *ab initio* results is still present. The previously lowest free energy minimum A has its free energy increased to 14.2 kJ/mol, which is higher than those of minima like C and I, thus making B the new absolute minimum. As before, minima belonging to the outer tier are the least favoured, but with a free energy now lowered at about 18.8 kJ/mol.

On the opposite side of the spectrum, OPLS-AA may downplay the role of hydrogen bonds over the penalties of *cis*-*trans* isomerisation. This force field reproduces the lowest *cis* conformers similarly to the others, but its sampling of the higher free energy *trans* basins is too sparse due to their increased energy.
In conclusion the choice of the force field does not affect the identification of the structures of the minimal free energy conformers, but only their relative free energies. CHARMM36 produces results for Asp in the gas phase in between the other two tested force fields, and is able to reproduce reasonably well the estimated abundances of the experimentally identified conformers. Hence we can consider it an adequate choice, at least at the qualitative level.

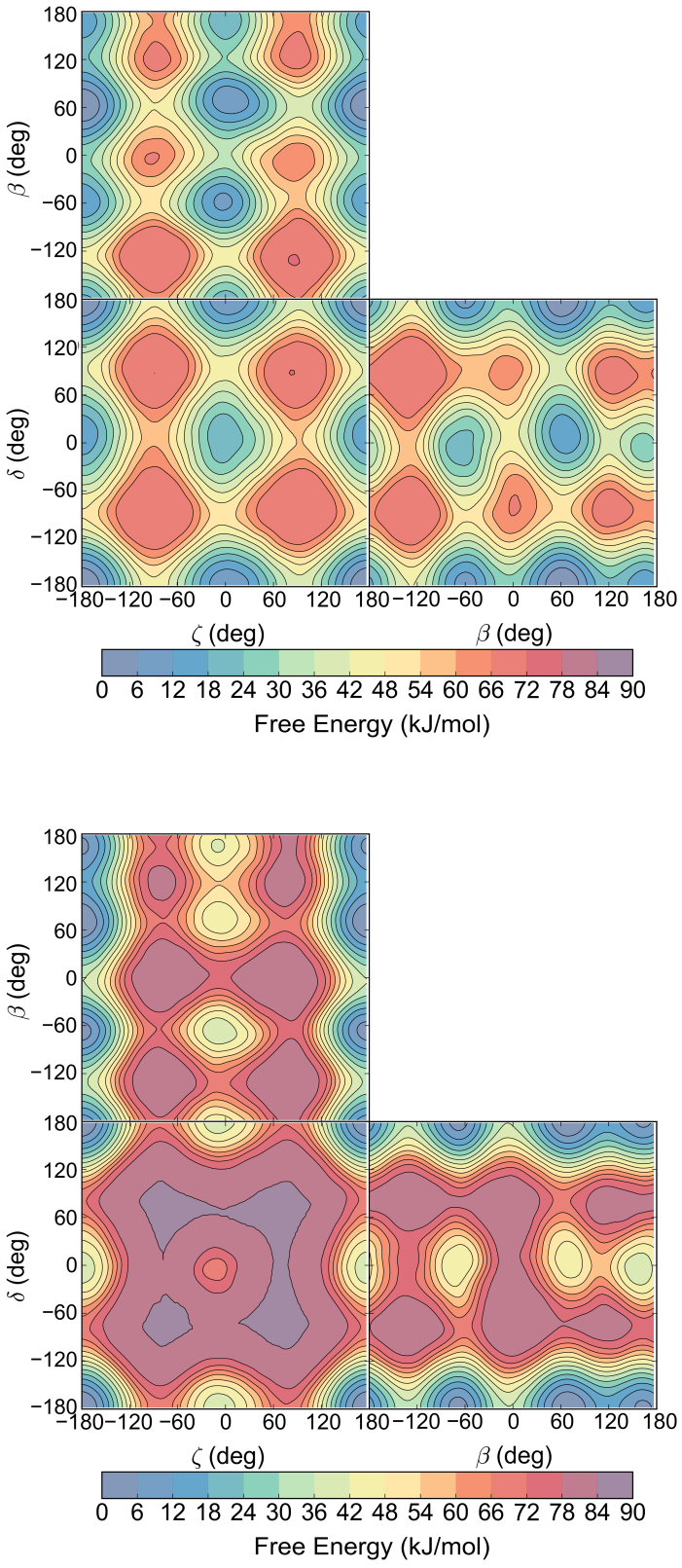


Figure 3.46: Projected free energy maps from metadynamics for neutral Asp in the gas phase as a function of pairs of torsional angles calculated with the AMBER ff99SB-ildn force field (top) and the OPLS-AA force field (bottom). The zero of the free energy scale corresponds to the absolute minimum for each map. Contour lines are every 6 kJ/mol.

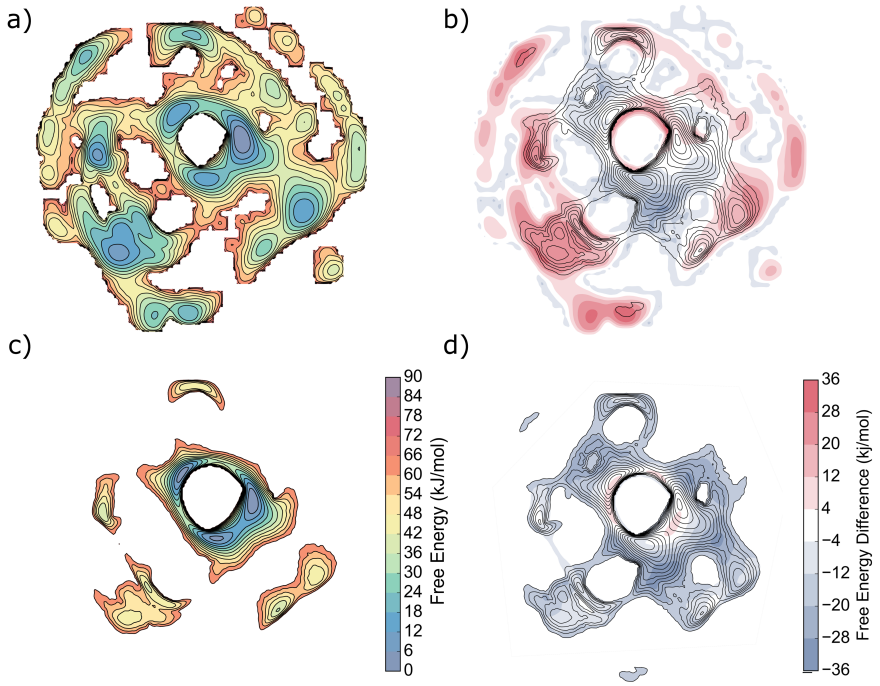


Figure 3.47: On the left, Free energy sketch-maps for neutral Asp in the gas phase as a function of pairs of torsional angles calculated with the AMBER ff99SB-ildn (top) and OPLS-AA (bottom) force field. On the right, the difference in free energy with respect to the maps obtained with CHARMM36.

	Free Energy (kJ/mol)		
Basin	CHARMM	AMBER	OPLS
A	0.0±0.4	14.2 ± 1.7	4.1 ± 0.4
B	2.9±0.8	0.0 ± 0.8	0.0 ± 0.4
C	8.8±1.3	12.1 ± 1.7	2.7 ± 0.4
D	28.9±7.5	30.0 ± 50.0	35.6 ± 4.1
E	30.6±6.3	13.0 ± 1.7	41.9 ± 13.4
F	31.0±9.2	15.1 ± 3.8	39.8 ± 10.9
G	31.4±9.2	22.2 ± 4.1	42.3 ± 14.2
H	32.7±9.6	10.0 ± 1.3	36.0 ± 6.7
I	33.9±12.1	8.8 ± 1.7	44.0 ± 16.7
J	50.2±10.0	18.4 ± 2.1	
K	50.7±9.2	18.4 ± 3.8	
L	50.7±9.2	21.8 ± 3.8	67.4 ± 26.4

Table 3.7: Free energies of neutral Asp conformers *in vacuo* relative to their absolute minimum calculated with the CHARMM36, AMBER99-ildn and OPLS-AA force fields from the and sketch-map with three CVs.

Conclusions

A thorough investigation of the conformational properties of biological systems can be cumbersome even for relatively small molecules. Rotational spectroscopy experiments are able to extract information only for the most populated low energy conformers in the gas phase and discrepancies with single point *ab initio* calculations may occur, where the calculated free energies and the measured relative populations disagree.

In this work we tested a protocol, which combines classical MD simulations enhanced by metadynamics and the dimensional reduction algorithm sketch-map and allowed us to efficiently explore at finite temperature and at a relatively low computational cost the conformational free energy landscape of a typical amino acid, Asp, in the neutral and zwitterionic form, in the gas phase and aqueous solution. This scheme allowed us to recover the conformers that were identified with rotational spectroscopy experiments, together with a number of structures that were not previously observed experimentally because of their high free energy. The use of metadynamics to enhance the sampling of the conformational space with respect to conventional MD helped us overcome energy barriers, accessing stable conformers within a wide range of free energies. The dimensionality reduction algorithm sketch-map proved useful in simplifying the representation of the multidimensional free energy, disentangling basins and their network of interconversion paths. The use of a classical force field, which is imposed by the computational cost, has some limitations, especially in the gas phase when using force fields developed for the wet biomolecular environment. Qualitative consistency, however, was found among the three different force field here tested, which produced the same conformers albeit with discrepancies in the relative free energies.

Differences in the energy ranking were also observed with the *ab initio* data, more prominently in the gas phase than in solution. It should however be noted that these calculations are also approximated: they consist of single-point potential energy calculations corrected *a posteriori* and, at variance from directly free energy calculations, do not include anharmonic entropic effects.

Moreover, an analysis of the solvation properties of the different conformers identified with metadynamics, helped shed some light on the proton transfer mechanism, which characterises the transformation from neutral to zwitterionic Asp when the amino acid is solvated. The results support the idea that the transfer from the carboxyl group to the amino group through chains of water molecules is unlikely. A direct transfer appears more probable, although the rate of isomerisation of the ζ angle, a necessary preliminary step, appears to be slow within the force field accuracy. The different conformations of Asp here identified could be then exploited to overcome the limited timescale of *ab initio* simulations and simulate the proton transfer under

different initial circumstances.

In conclusion, we outlined a powerful protocol for mapping amino acid and, in general, peptide conformers at finite temperature that can be potentially applied to other more complex cases, such as polypeptide chains. Results can be improved with the use of novel and accurate force fields, e.g. based on machine learning algorithms on high quality training sets, presently under development.
These studies would ultimately contribute to the understanding of the conformational complexity of the building blocks of biological matter, which is important in a wide variety of problems and fields, from drug design based on molecular recognition mechanisms to the identification of organic compounds and amino acids in the interstellar environment. One example that particularly fits this project would be the investigation of the conformational properties of the M2-M3 linker or other interfacial loops in pLGICs to shed light on their role in the transmission of the activation signal from the ECD to the TMD.

Chapter 4

Conclusions

The use of metadynamics in this thesis to accelerate rare events involved in the activation mechanisms of pLGICs led to promising results and improved our understanding of different aspects of such complex phenomena. The results here presented are a step towards a complete description of the activation signalling in Cys-loop receptors, from the initial binding of a small ligand in the extracellular domain, to the opening movements of the helices lining the ion permeable pore.
Exploratory MD simulations were first performed on a number of homology models (based on AChBP, ELIC and GluCl) of the RDL GABA receptor ECD, in order to evaluate the influence that different templates, alignments and docking poses, can have on the final outcome of MD simulations. Stability and structural properties of the models were monitored, together with the persistence of hydrogen bonds and cation-π interactions within the binding site.
Comparing the differences among these structures allowed us to extract model-independent information on the binding pocket, which were compared to experimental data obtained by mutagenesis electrophysiology experiments on the same protein: key interactions with the charged Arg111 and Glu204 residues were confirmed to act at the opposite ends of the zwitterionic GABA, while a cage of surrounding aromatic residues forming cation-π interactions with the ligand helped pinning its positively charged amino group in place. Moreover, this preliminary work also raised questions on the role of water molecules within the binding pocket: a previously unknown binding conformation, where a water molecule bridges the interaction between the carboxyl group of GABA and positive Arg111, was observed.
The results of this initial study revealed in a non-trivial way that the percentage of sequence identity is crucial in the process of building a structure from a homologous template: models from pLGICs templates provided stronger binding and more numerous interaction opportunities for the ligand. Similarly, the initial docking conformation can affect the quantity and quality of bonds within the relatively short timescale accessible to classical MD. Ideally, a metadynamics based docking protocol

could improve on this limitation, by accelerating the exploration of the binding site and by helping identify all the possible binding conformations. This was confirmed by a preliminary metadynamics run on one of the two homology models built from the GluCl template, which provided the strongest binding and consistency with experimental data, and was identified as the best candidate for further studies on the RDL receptor.

The power of metadynamics was thus exploited during μs simulations of the chosen GluCl-based RDL model and two mutants to gain significant insights into the actual binding process. The relative position of the ligand from the alleged binding site was biased, forcing it to access and leave the pocket multiple times: with the help of a funnel-shaped restraining potential, exploration of the solvated states was limited to a small cylindrical volume, guaranteeing the convergence of the simulations within an affordable computational time. The binding conformations, with or without the mediation of the hydrogen bond with Glu204 by a water molecule, and an alleged pre-binding conformation were characterised in detail and their free energy was evaluated.

Other than mapping the free energy landscape of the RDL binding pocket with increased statistical accuracy, this simulation allowed us to speculate on the steps that guide the ligand into the binding when approaching the protein. The flexible side chain of Arg218 is the first point of contact with the RDL receptor for GABA and is responsible for catching its carboxyl group and pass it to Arg111. Subsequently, when fluctuations of loop C, a highly flexible structure that covers the binding site, leave enough space, Arg111 leads the ligand inside the pocket, where it is stabilised by interactions of its amino group with Glu204 and the surrounding aromatic cage. The funnel-metadynamics simulation and the analysis were repeated for two mutants, where Arg111 and Glu204 were substituted with non polar alanines. The consequent disruption of the binding site free energy landscape confirmed their crucial role in the binding of GABA and proved that the lack of activity in equivalent mutant channels in mutagenesis electrophysiology experiments is due to the absence of ligand binding at the ECD level.

Following the study of the binding properties of pLGICs, exemplified by the investigation of GABA binding to the RDL receptor, our focus moved to the gating problem. Working with a recently crystallised structure of the 5-HT_3 serotonin receptor, we analysed the effects that the *trans-cis* isomerisation of a highly conserved proline (in this system Pro281) in the M2-M3 loop at the interface between ECD and TMD, has on the position and dynamics of the helices lining the ion permeable pore and on the interactions among surrounding loops. Mutagenesis experiments identified this proline as essential for the correct functioning of the channels, suggesting that its *trans-cis* isomerisation may act as a molecular switch to open the channel. We compared the outcome of 250 ns MD simulations of three models of

the 5-HT_3 receptor where the *trans-cis* isomerisation was induced in none, three and five subunits: no hydrogen bonds were consistently formed or broken among the interfacial loops, although an increase in interactions between Thr280 and the negatively charged moieties of lipid heads mimicking the membrane was observed. This specific interaction was suggested to play a key role in favouring the *trans* to *cis* transition, where the Thr280 backbone atoms may act as a handle and influence the free energy ratio between the two isomers. A 500 ns metadynamics confirmed that the protein environment affects the free energy landscape when compared to that of a proline dipeptide in water: although the *cis* basin is higher in free energy, the transition from the *trans* state is characterised by a lower barrier and may be more probable. A clear directional preference for the isomerisation in the receptor, due to the presence of obstructing residues such as Tyr140 belonging to the Cys-loop in the ECD, was also highlighted in this simulation. We thus proposed that movements in the ECD following the binding of a ligand may lead to the displacement of such residues and further favour the isomerisation.

Increased flexibility and structural rearrangements were also observed in other amino acids along the M2-M3 loop in the TMD, where the Pro281 is found, which led to the straightening of the top part of the M2 helix. Movements of this secondary structure, which lines the central pore, could be at the basis of the final gating step: this is expected to involve the side chain of Leu260, in this helix, which was suggested to be responsible for the hydrophobic occlusion in previous studies. Although with caution, we advanced the idea that the straightening of the upper part of this helix may be the first step in a chain of conformational rearrangements culminating in the displacement of Leu260. The consequent increase in pore radius may be sufficient to allow the flow of water molecules and ions through the channel.

In light of the substantial effects that the isomerisation of a single proline can have on structure and dynamics of the surrounding environment, it is clear that conformational preferences of small peptides or even single residues can play a decisive role in defining the behaviour of proteins and bigger biomolecules. In the final chapter of this thesis, we thus explored a new protocol to obtain a exhaustive exploration of the conformational landscape of relatively small but biologically interesting systems, with the aid of metadynamics and the dimensionality reduction algorithm sketch-map.

Metadynamics simulations of aspartic acid under different conditions were run until convergence, where only a subset of the torsional angles defining the full space of the conformations were biased. Sketch-map helped disentangle the resulting complex network of basins and transition paths, by mapping the free energy surfaces obtained by this simulations to a two-dimensional representation. From here, conformers were extracted and refined with *ab initio* methods. The study was first carried out for the neutral form of aspartic acid in gas phase and a total of 24 conformers were characterised, where rotational spectroscopy experiments were able to identify only

six as the free energy range of the observable structures is relatively limited. A similar study was then carried out for the zwitterionic form of aspartic acid in water and then again for the unphysical systems of the neutral and zwitterionic in water and in gas phase respectively. A comparison among these reduced maps allowed to evaluate the effects that solvation and the protonation state of the molecule have on the free energy landscape and gave a few interesting suggestions about the potential proton transfer mechanism that from the neutral Asp leads to its zwitterionic form in presence of water molecules. An analysis of the chains of waters bridging intramolecular bonds also confirmed that a direct transfer from the carboxyl to the amino group may be the most probable mechanism, even when conformational rearrangements of the peptide are required.
The choice of force field was finally assessed by repeating the simulations with the most commonly used parametrisations: although all of them were consistent in identifying the low energy basins, the relative free energies proved to be strongly dependent on this choice.

Due to their complexity and limited experimental information, to date it is still not possible to achieve a complete computational description of the activation mechanisms of pLGICs. However, the continuous improvement in power and efficiency of computers and the development of smart techniques to overcome the limitations of older, more commonly used methods, allow now to simulate and study what was unreachable just a few years ago. In this project, we exploited the power of some of these techniques, such as MD and metadynamics, to improve our knowledge of the activation mechanism of pLGICs, focusing on some of its crucial steps, i.e. the binding of the ligand and the conformational rearrangements of the gating stage. A lot of work, however, still needs to be done to fill the gaps.
The work on the RDL receptor demonstrated the capabilities of MD and funnel-metadynamics to characterise the ligand binding and may easily be adapted to similar systems, such as the serotonin bound 5-HT_3 receptor for which an X-ray structure is now available. This new procedure would allow to map the free energy landscape, evaluate binding affinities and identify potential binding paths with more accuracy and efficiency.
The binding path of GABA to the RDL receptor proposed in this thesis is based on a number of binding and unbinding events observed during the metadynamics simulations, and experimental information could be extremely helpful in assessing its validity. For example, the influence of the positively charged Arg218 on the binding may be tested by mutating it with a non-polar or negative residues. Similarly, experimenting on the loop C, by tuning its flexibility and its propensity to form a β-hairpin structure could decrease or increase the binding affinity and affect the activity of the channel. Furthermore, a thorough study of the interface between the ECD and the

TMD of the 5-HT_3 would help in understanding how the activation signal travels from the binding site to the M2 helices; while extending the statistics on the simulations here provided would be beneficial, it would be also important to add further systems for comparison, for example working with a serotonin bound receptor, to observe how the presence of the ligand affects Pro281 and other residues of the interface. Mutagenesis simulations and experiments could also be designed to confirm the roles of such amino acids in the gating: besides testing residues that could increase and decrease the interaction network between ECD and TMD loops, substituting the side chain of Tyr140 in the Cys-loop with shorter compounds could help remove its steric occlusion to the clockwise isomerisation of Pro281, while mutations on Thr276 may be able to confirm its role in transmitting the conformational change from the isomerising proline to the M2 helix.

These studies are particularly timely, since an increasing number of pLGICs structures has been recently resolved. Computational techniques beyond conventional MD could thus exploit this new wealth of information, interpreting and guiding the experiments.

Finally, the protocol here defined for a thorough exploration of the conformational free energy landscape of small molecules with metadynamics and sketch-map could be easily applied to other amino acids or peptides chains. If properly used in conjunction with experiments, it could prove useful in shedding light on the behaviour of more complex biological structures starting from their basic building blocks.

Appendix A

Notes on the funnel restraining potential

In the following, the procedure to obtain the binding free energy of a system in presence of a funnel restraining potential such as that described in Sec. 2.2.3 and applied in Sec. 3.2 is here briefly outlined. It is mostly based on the works of Ref. [160, 228]

Given a system, composed by a ligand with coordinates $\mathbf{r}$, and a protein with which it may interact, the potential of mean force (PMF) W of the ligand only can be made independent on the coordinates of any other body in the system (e.g. substrate or solvent) $\mathbf{X}$ thorough the the relation:

$$e^{-\beta W(\mathbf{r})} = \int d\mathbf{X} e^{-\beta U(\mathbf{r},\mathbf{X})} \tag{A.1}$$

where $U(\mathbf{r},\mathbf{X})$ is the potential of the entire system, $\beta = 1/k_b T$ with k_b the Boltzmann constant and T the temperature. Here, all the degrees of freedom that do not pertain the ligand are simply integrated out. Consequently, the probability of the ligand to be found in a certain area is equal, according to Boltzmann probability, to the integral of these exponential weights over the area itself, divided by the integral of all possible states:

$$p_{area} = \frac{\int_{area} d\mathbf{r} \int d\mathbf{X} e^{-\beta U(\mathbf{r},\mathbf{X})}}{\int d\mathbf{r} \int d\mathbf{X} e^{-\beta U(\mathbf{r},\mathbf{X})}} = \frac{\int_{area} d\mathbf{r} e^{-\beta W(\mathbf{r})}}{\int d\mathbf{r} e^{-\beta W(\mathbf{r})}} \tag{A.2}$$

The equilibrium constant K_b (opposite of the more commonly used dissociation constant $K_d = \frac{1}{K_b}$) between a ligand L and a substrate S is defined from the bound and unbound concentrations $[LS]$ and $[L]$, $[S]$ of the ligand and the substrate respectively as:

$$K_b = [LS]/[L][S] \tag{A.3}$$

This can be rewritten in terms of probabilities, being p_{bound} the probability to find a protein with the ligand bound and $p_{unbound}$ the probability to find the protein with the ligand unbound:

$$K_b[L] = K_b = \frac{p_{bound}}{p_{unbound}} = \frac{\int_{site} d\mathbf{r} e^{-\beta W(\mathbf{r})}}{\int_{bulk} d\mathbf{r} e^{-\beta W(\mathbf{r})}} \tag{A.4}$$

Where the diluted limit $[L] \to 1$ was applied. At this point, it is assumed that the PMF in bulk is constant (and usually set equal to zero). A reference point in the bulk $\mathbf{r}_{ref}$ can then be taken, such that $W(\mathbf{r}_{ref}) = W_{ref}$ is the reference value of the PMF in the unbound state. Integrating the denominator in Eq. A.4, one gets V, the volume accessible to the ligand in bulk, which can be taken as the inverse of the standard concentration $C^0 = 1/V = 1/1660$ Å^{-3}. This maintains K_b as the dimensionless object required to calculate the binding free energy.

$$K_b = \frac{\int_{site} d\mathbf{r} e^{-\beta W(\mathbf{r})}}{e^{-\beta W_{ref}} \int_{bulk} d\mathbf{r}} = \frac{\int_{site} d\mathbf{r} e^{-\beta W(\mathbf{r})}}{V e^{-\beta W_{ref}}} = \frac{C^0 \int_{site} d\mathbf{r} e^{-\beta W(\mathbf{r})}}{e^{-\beta W_{ref}}} \tag{A.5}$$

The Gibbs free energy of a state A is defined from its Boltzmann probability as:

$$G_A = -\frac{1}{\beta} ln(p_A) = -\frac{1}{\beta} ln\Big(\frac{\int_A d\mathbf{r} e^{-\beta W(\mathbf{r})}}{Z}\Big) \tag{A.6}$$

where here we use Z, the partition function to indicate the integral over all possible states. This is of limited use as the the absolute value of G is arbitrary. More significant is instead the difference in Gibbs free energy between states:

$$\Delta G_{AB} = -\frac{1}{\beta} ln(p_A) + \frac{1}{\beta} ln(p_B) = -\frac{1}{\beta} ln\Big(\frac{p_A}{p_B}\Big) = -\frac{1}{\beta} ln\Big(\frac{\int_A d\mathbf{r} e^{-\beta W(\mathbf{r})}}{\int_B d\mathbf{r} e^{-\beta W(\mathbf{r})}}\Big) \tag{A.7}$$

To calculate the binding free energy, the difference in free energy between the bound and the unbound state, one gets

$$\Delta G_{binding} = -\frac{1}{\beta} ln\Big(\frac{p_{bound}}{p_{unbound}}\Big) = -\frac{1}{\beta} ln(K_b) \tag{A.8}$$

The presence of a restraining potential, however, must be taken into account when passing from a 3D to a more simple 1D PMF. The degrees of freedom of the ligand $\mathbf{r}$ can be explicitly rewritten as a function of the Cartesian coordinates (x, y, z), and assuming as an example that the major axis of the restraining potential is aligned with the z axis, one can obtain the 1D PMF integrating the other remaining degrees

of freedom (x, y). The introduction of this restriction, the potential $u(x, y)$, also helps keeping the integrals well behaved and converging. We get that

$$C \int dz\ e^{-\beta W(z)} = \int dxdydz\ e^{-\beta(W(x,y,z)+u(x,y))} \tag{A.9}$$

Where C is a constant, result of the integration along (x, y). The restraining potential may then assume any shape, but the procedure doesn't change. It is possible to take into account its effect rewriting K_b in the following way: first we add extra integrals including the potential

$$\begin{aligned} K_b = \frac{C^0 \int_{site} dxdydz\ e^{-\beta W(x,y,z)}}{e^{-\beta W_{ref}}} \\ \frac{\int_{site} dxdydz\ e^{-\beta(W(x,y,z)+u(x,y))}}{\int_{site} dxdydz\ e^{-\beta(W(x,y,z)+u(x,y))}} \\ \frac{\int_{bulk} dxdydz\ e^{-\beta(W(x',y',z')+u(x',y'))}}{\int_{bulk} dxdydz\ e^{-\beta(W(x',y',z')+u(x',y'))}} \end{aligned} \tag{A.10}$$

where $r_{ref} = (x', y', z')$ is the reference position in the unbound state in that gives $W(x', y', z') = W_{ref}$. Then, rearranging Eq. A.10, one gets

$$\begin{aligned} K_b = \frac{C^0 \int_{site} dxdydz\ e^{-\beta W(x,y,z)}}{\int_{site} dxdydz\ e^{-\beta(W(x,y,z)+u(x,y))}} \\ \frac{\int_{site} dxdydz\ e^{-\beta(W(x,y,z)+u(x,y))}}{\int_{bulk} dxdydz\ e^{-\beta(W(x',y',z')+u(x',y'))}} \\ \frac{\int_{bulk} dxdydz\ e^{-\beta(W(x',y',z')+u(x',y'))}}{e^{-\beta W_{ref}}} \end{aligned} \tag{A.11}$$

Now, the first term is simply the Boltzmann weight of the difference ΔG_{site} in free energy induced by the restraining potential when the ligand is in the binding site. At the same time, being $e^{W_{ref}}$ a constant, it can be taken out from the third term integral as before, and one obtains

$$\begin{aligned} K_b = C^0 e^{-\beta \Delta G_{site}} \\ \frac{\int_{site} dxdydz\ e^{-\beta(W(x,y,z)+u(x,y))}}{\int_{bulk} dxdydz\ e^{-\beta(W(x',y',z')+u(x',y'))}} \\ \frac{e^{-\beta W_{ref}}}{e^{-\beta W_{ref}}} \int dxdy\ e^{-\beta u(x',y')} \end{aligned} \tag{A.12}$$

The middle term is simply the integral over the difference in the PMF w between bound and unbound states measured from the biased trajectory, that can be integrated

to the monodimensional $w(z)$. The last term instead is just an integral in (x, y) over the potential shape in the unbound region. The result is the cross section of the potential, or surface $S_{unbound}$. From this, one obtains

$$K_b = C^0 e^{-\beta \Delta G_{site}} \int_{site} dz \; e^{-\beta(w(z)-w_{ref})} S_{unbound} \tag{A.13}$$

ΔG_{site} goes to zero if the potential is large enough in the region of the site not to influence the binding, as it is in the case of the funnel potential. Similarly, $S_{unbound} = \pi R_{cyl}^2$, where R_{cyl} is the radius of the cylindrical restraining potential acting on the solvated states. This gives the unbiased estimator of the binding rate for the funnel potential

$$K_b = \pi R_{cyl}^2 C^0 \int_{site} dz \; e^{-\beta(w(z)-w_{ref})} \tag{A.14}$$

Finally, substituting this result in the Gibbs free energy of Eq. A.8 and splitting the logarithm, we get

$$\begin{aligned} \Delta G_{binding} = -\frac{1}{\beta} ln\Big(\pi R_{cyl}^2 C^0 \int_{site} dz \; e^{-\beta(w(z)-w_{ref})}\Big) = \\ -\frac{1}{\beta} ln\Big(\int_{site} dz \; e^{-\beta(w(z)-w_{ref})}\Big) - \frac{1}{\beta} ln(\pi R_{cyl}^2 C^0) = \\ \Delta G_{measured} - \frac{1}{\beta} ln(\pi R_{cyl}^2 C^0). \end{aligned} \tag{A.15}$$

Bibliography

[1] S. C. R. Lummis, D. L. Beene, L. W. Lee, H. A. Lester, R. W. Broadhurst, and D. A. Dougherty, "Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel," *Nature*, vol. 438, pp. 248–52, Nov. 2005.

[2] Society for Neuroscience, *Brain Facts*. Society for Neuroscience, 2008.

[3] Scientific American, *The Scientific American Book of the Brain*. Lyons Press, 2001.

[4] C. Melis, *Mutagenesis Computer Experiments on Ligand-gated Ion Channels*. Phd thesis, King's College London, 2007.

[5] V. Katritch, V. Cherezov, and R. C. Stevens, "Structure-function of the G protein-coupled receptor superfamily," *Annual Review of Pharmacologicy and Toxicology*, vol. 53, pp. 531–56, Nov. 2013.

[6] D. Lemoine, R. Jiang, A. Taly, T. Chataigneau, A. Specht, and T. Grutter, "Ligand-gated ion channels: new insights into neurological disorders and ligand recognition," *Chemical Reviews*, vol. 112, pp. 6285–318, Dec. 2012.

[7] J. B. Reece, *Biology*. Pearson, Benjamin Cummings, 2008.

[8] D. E. Koshland, "Enzyme flexibility and enzyme action," *Journal of Cellular and Comparative Physiology*, vol. 54, pp. 245–58, Dec. 1959.

[9] J. Monod, J. Wyman, and J. P. Changeux, "On the nature of allosteric transitions: a plausible model," *Journal of Molecular Biology*, vol. 12, pp. 88–118, May 1965.

[10] P.-J. Corringer, F. Poitevin, M. S. Prevost, L. Sauguet, M. Delarue, and J.-P. Changeux, "Structure and pharmacology of pentameric receptor channels: from bacteria to brain," *Structure*, vol. 20, pp. 941–56, June 2012.

[11] Á. Nemecz, M. S. Prevost, A. Menny, and P.-J. Corringer, "Emerging Molecular Mechanisms of Signal Transduction in Pentameric Ligand-Gated Ion Channels," *Neuron*, vol. 90, pp. 452–70, May 2016.

[12] F. M. Ashcroft, *Ion Channels and Disease: Channelopathies*. Academic Press, 2000.

[13] G. Fernández-Ballester, A. Fernández-Carvajal, J. M. González-Ros, and A. Ferrer-Montiel, "Ionic channels as targets for drug design: a review on computational methods," *Pharmaceutics*, vol. 3, pp. 932–53, Jan. 2011.

[14] F. Hucho and C. Weise, "Ligand-Gated Ion Channels," *Angewandte Chemie International Edition*, vol. 40, pp. 3100–16, Sept. 2001.

[15] R. E. Hibbs and E. Gouaux, "Principles of activation and permeation in an anion-selective Cys-loop receptor," *Nature*, vol. 474, pp. 54–60, June 2011.

[16] N. Unwin, "Refined structure of the nicotinic acetylcholine receptor at 4A resolution," *Journal of Molecular Biology*, vol. 346, pp. 967–89, Mar. 2005.

[17] M. Caffrey, "Membrane protein crystallization," *Journal of Structural Biology*, vol. 142, pp. 108–32, Apr. 2003.

[18] E. P. Carpenter, K. Beis, A. D. Cameron, and S. Iwata, "Overcoming the challenges of membrane protein crystallography," *Current Opinion in Structural Biology*, vol. 18, pp. 581–6, Oct. 2008.

[19] J. N. Langley, "On the reaction of cells and of nerve-endings to certain poisons, chiefly as regards the reaction of striated muscle to nicotine and to curari," *The Journal of Physiology*, vol. 33, pp. 374–413, Dec. 1905.

[20] A. Miyazawa, Y. Fujiyoshi, M. Stowell, and N. Unwin, "Nicotinic acetylcholine receptor at 4.6 Å resolution: transverse tunnels in the channel wall," *Journal of Molecular Biology*, vol. 288, pp. 765–86, May 1999.

[21] J.-P. Changeux, "The nicotinic acetylcholine receptor: the founding father of the pentameric ligand-gated ion channel superfamily," *Journal of Biological Chemistry*, vol. 287, pp. 40207–15, Oct. 2012.

[22] C. Chothia and A. M. Lesk, "The relation between the divergence of sequence and structure in proteins," *The EMBO Journal*, vol. 5, pp. 823–6, Apr. 1986.

[23] S. Kaczanowski and P. Zielenkiewicz, "Why similar protein sequences encode similar three-dimensional structures?," *Theoretical Chemistry Accounts*, vol. 125, pp. 643–50, Oct. 2009.

[24] K. Brejc, W. J. van Dijk, R. V. Klaassen, M. Schuurmans, J. van Der Oost, A. B. Smit, and T. K. Sixma, "Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors," *Nature*, vol. 411, pp. 269–76, May 2001.

[25] R. J. C. Hilf and R. Dutzler, "X-ray structure of a prokaryotic pentameric ligand-gated ion channel," *Nature*, vol. 452, pp. 375–9, Mar. 2008.

[26] R. J. C. Hilf and R. Dutzler, "Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel," *Nature*, vol. 457, pp. 115–8, Jan. 2009.

[27] N. Bocquet, H. Nury, M. Baaden, C. Le Poupon, J.-P. Changeux, M. Delarue, and P.-J. Corringer, "X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation," *Nature*, vol. 457, pp. 111–4, Jan. 2009.

[28] G. Hassaine, C. Deluz, L. Grasso, R. Wyss, M. B. Tol, R. Hovius, A. Graff, H. Stahlberg, T. Tomizaki, A. Desmyter, C. Moreau, X.-D. Li, F. Poitevin, H. Vogel, and H. Nury, "X-ray structure of the mouse serotonin 5-HT_3 receptor," *Nature*, vol. 512, pp. 276–81, Aug. 2014.

[29] P. S. Miller and A. R. Aricescu, "Crystal structure of a human $GABA_A$ receptor," *Nature*, vol. 512, pp. 270–75, June 2014.

[30] X. Huang, H. Chen, K. Michelsen, S. Schneider, and P. L. Shaffer, "Crystal structure of human glycine receptor-α3 bound to antagonist strychnine," *Nature*, vol. 526, pp. 277–80, Sept. 2015.

[31] J. Du, W. Lü, S. Wu, Y. Cheng, and E. Gouaux, "Glycine receptor mechanism elucidated by electron cryo-microscopy," *Nature*, vol. 526, pp. 224–9, Sept. 2015.

[32] G. A. Jeffrey, *An introduction to hydrogen bonding*. Oxford University Press, 1997.

[33] 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*, vol. 96, no. 17, pp. 9459–64, 1999.

[34] C. Melis, P.-L. Chau, K. L. Price, S. C. R. Lummis, and C. Molteni, "Exploring the binding of serotonin to the 5-HT_3 receptor by density functional theory," *The Journal of Physical Chemistry B*, vol. 110, pp. 26313–9, Dec. 2006.

[35] E. O. McCullum, B. a. R. Williams, J. Zhang, and J. C. Chaput, "In vitro mutagenesis protocols," *Methods in Molecular Biology*, vol. 634, no. 3, pp. 103–9, 2010.

[36] A. Finkel and R. Bookman, *The electrophysiology setup*. Current Protocols in Neuroscience, Aug. 2001.

[37] J. L. Sebaugh, "Guidelines for accurate EC50/IC50 estimation," *Pharmaceutical Statistics*, vol. 10, no. 2, pp. 128–34, 2011.

[38] X. Liu, W. C. Krause, and R. L. Davis, "$GABA_A$ receptor RDL inhibits Drosophila olfactory associative learning," *Neuron*, vol. 56, pp. 1090–102, Dec. 2007.

[39] K. Sato, M. Pellegrino, T. Nakagawa, T. Nakagawa, L. B. Vosshall, and K. Touhara, "Insect olfactory receptors are heteromeric ligand-gated ion channels," *Nature*, vol. 452, pp. 1002–6, Apr. 2008.

[40] M. M. Rahman, G. Liu, K. Furuta, F. Ozoe, and Y. Ozoe, "Synthesis of 1,3-di- and 1,3,4-trisubstituted 1,6-dihydro-6-iminopyridazines as competitive antagonists of insect GABA receptors," *Journal of Pesticide Science*, vol. 39, pp. 133–43, Aug. 2014.

[41] G. Liu, F. Ozoe, K. Furuta, and Y. Ozoe, "4,5-Substituted 3-Isoxazolols with insecticidal activity act as competitive antagonists of housefly GABA Receptors," *Journal of Agricultural and Food Chemistry*, vol. 63, pp. 6304–12, July 2015.

[42] G. Liu, B. Frølund, F. Ozoe, and Y. Ozoe, "Differential interactions of 5-(4-piperidyl)-3-isoxazolol analogues with insect γ-aminobutyric acid receptors leading to functional selectivity," *Insect Biochemistry and Molecular Biology*, vol. 66, pp. 64–71, Oct. 2015.

[43] S. Kegley, B. Hill, S. Orme, and A. Choi, *PAN pesticide database*. http://www.pesticideinfo.org [Accessed: 02-09-2015].

[44] A. J. Thompson, H. A. Lester, and S. C. R. Lummis, "The structural basis of function in Cys-loop receptors," *Quarterly Reviews of Biophysics*, vol. 43, pp. 449–99, Nov. 2010.

[45] M. Nys, D. Kesters, and C. Ulens, "Structural insights into Cys-loop receptor function and ligand recognition," *Biochemical Pharmacology*, vol. 86, pp. 1042–53, Oct. 2013.

[46] T. Lynagh and S. A. Pless, "Principles of agonist recognition in Cys-loop receptors," *Frontiers in Physiology*, vol. 5, p. 160, Jan. 2014.

[47] C. B. Nemeroff, vol. 161. American Journal of Psychiatry, 8th ed., 2004.

[48] C. Melis, S. C. R. Lummis, and C. Molteni, "Molecular dynamics simulations of GABA binding to the $GABA_C$ receptor: the role of Arg104," *Biophysical Journal*, vol. 95, pp. 4115–23, Nov. 2008.

[49] S. C. R. Lummis, D. L Beene, N. J. Harrison, H. A. Lester, and D. A. Dougherty, "A cation-π binding interaction with a tyrosine in the binding site of the $GABA_C$ receptor," *Chemical Biology*, vol. 12, pp. 993–7, Sept. 2005.

[50] S. C. R. Lummis, I. McGonigle, J. A. Ashby, and D. A. Dougherty, "Two amino acid residues contribute to a cation-π binding interaction in the binding site of an insect GABA receptor," *The Journal of Neuroscience*, vol. 31, pp. 12371–6, Aug. 2011.

[51] X. Cheng, H. Wang, B. Grant, S. M. Sine, and J. A. McCammon, "Targeted molecular dynamics study of C-loop closure and channel gating in nicotinic receptors," *PLoS Computational Biology*, vol. 2, p. e134, Sept. 2006.

[52] S. J. Edelstein and J. P. Changeux, "Allosteric transitions of the acetylcholine receptor," *Advances in Protein Chemistry*, vol. 51, pp. 121–84, 1998.

[53] E. X. Albuquerque, E. F. R. Pereira, M. Alkondon, and S. W. Rogers, "Mammalian nicotinic acetylcholine receptors: from structure to function," *Physiological Reviews*, vol. 89, pp. 73–120, jan 2009.

[54] P. Aryal, M. S. P. Sansom, and S. J. Tucker, "Hydrophobic gating in ion channels," *Journal of Molecular Biology*, vol. 427, pp. 121–30, Jan. 2015.

[55] A. Miyazawa, Y. Fujiyoshi, and N. Unwin, "Structure and gating mechanism of the acetylcholine receptor pore," *Nature*, vol. 423, pp. 949–55, June 2003.

[56] D. A. Doyle, "Structural changes during ion channel gating," *Trends in Neurosciences*, vol. 27, pp. 298–302, June 2004.

[57] N. Calimet, M. Simoes, J.-P. Changeux, M. Karplus, A. Taly, and M. Cecchini, "A gating mechanism of pentameric ligand-gated ion channels," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. E3987–96, 2013.

[58] R. Sankararamakrishnan and S. Vishveshwara, "Geometry of proline-containing α-helices in proteins," *International Journal of Peptide and Protein Research*, vol. 39, pp. 356–63, Apr. 1992.

[59] D. P. Tieleman, I. H. Shrivastava, M. R. Ulmschneider, and M. S. Sansom, "Proline-induced hinges in transmembrane helices: possible roles in ion channel gating," *Proteins*, vol. 44, pp. 63–72, Aug. 2001.

[60] K. A. Williams and C. M. Deber, "Proline residues in transmembrane helixes: structural or dynamic role?," *Biochemistry*, vol. 30, pp. 8919–23, Sept. 1991.

[61] W. J. Wedemeyer, E. Welker, and H. A. Scheraga, "Proline cis-trans isomerization and protein folding," *Biochemistry*, vol. 41, pp. 14637–44, Sept. 2002.

[62] C. Melis, G. Bussi, S. C. R. Lummis, and C. Molteni, "Trans-cis switching mechanisms in proline analogues and their relevance for the gating of the 5-ht$_3$ receptor," *The Journal of Physical Chemistry B*, vol. 113, pp. 12148–53, Sept. 2009.

[63] M. Berger, J. a. Gray, and B. L. Roth, "The expanded biology of serotonin," *Annual Review of Medicine*, vol. 60, pp. 355–66, Jan. 2009.

[64] D. Kesters, A. J. Thompson, M. Brams, R. van Elk, R. Spurny, M. Geitmann, J. M. Villalgordo, A. Guskov, U. H. Danielson, S. C. R. Lummis, A. B. Smit, and C. Ulens, "Structural basis of ligand recognition in 5-HT$_3$ receptors," *EMBO reports*, vol. 14, pp. 49–56, Jan. 2013.

[65] L. F. Mohammad-Zadeh, L. Moses, and S. M. Gwaltney-Brant, "Serotonin: a review," *Journal of Veterinary Pharmacology and Therapeutics*, vol. 31, pp. 187–99, June 2008.

[66] M. B. Jackson and J. L. Yakel, "The 5-HT$_3$ receptor channel," *Annual Review of Physiology*, vol. 57, pp. 447–68, Jan. 1995.

[67] S. C. R. Lummis, A. J. Thompson, M. Bencherif, and H. A. Lester, "Varenicline is a potent agonist of the human 5-hydroxytryptamine3 receptor," *The Journal of Pharmacology and Experimental Therapeutics*, vol. 339, pp. 125–31, oct 2011.

[68] K. L. Price, R. K. Lillestol, C. Ulens, and S. C. R. Lummis, "Varenicline interactions at the 5-HT$_3$ receptor ligand Binding site are revealed by 5-HTBP," *ACS Chemical Neuroscience*, vol. 6, pp. 1151–7, jul 2015.

[69] S. Yuan, S. Filipek, and H. Vogel, "A Gating Mechanism of the Serotonin 5-HT3 Receptor," *Structure*, vol. 24, pp. 816–825, may 2016.

[70] M. Levitt, "Conformational preferences of amino acids in globular proteins," *Biochemistry*, vol. 17, no. 20, pp. 4277–85, 1978.

[71] I. P. S. Mata, M. E. Sanz, V. Vaquero, C. Cabezas, C. Perez, S. Blanco, J. C. López, and J. L. Alonso, "Gas phase structure of amino acids: LA-MB-FTMW studies," in *64th International Symposium on Molecular Spectroscopy*, Ohio State University, 2009.

[72] M. E. Sanz, J. C. López, and J. L. Alonso, "Six conformers of neutral aspartic acid identified in the gas phase," *Physical Chemistry Chemical Physics*, vol. 12, pp. 3573–8, Apr. 2010.

[73] L. Polgár, "The catalytic triad of serine peptidases," *Cellular and Molecular Life Sciences*, vol. 62, pp. 2161–72, Oct. 2005.

[74] D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*. Freeman, W. H., 6th ed., 2013.

[75] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nature structural biology*, vol. 9, pp. 646–52, Sept. 2002.

[76] W. F. van Gunsteren, J. Dolenc, and A. E. Mark, "Molecular simulation as an aid to experimentalists," *Current Opinion in Structural Biology*, vol. 18, pp. 149–53, Apr. 2008.

[77] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, "Systematic validation of protein force fields against experimental data," *PLoS ONE*, vol. 7, p. e32131, Feb. 2012.

[78] B. J. Alder and T. E. Wainwright, "Phase Transition for a Hard Sphere System," *The Journal of Chemical Physics*, vol. 27, p. 1208, Aug. 1957.

[79] B. J. Alder and T. E. Wainwright, "Studies in molecular dynamics. I. General method," *The Journal of Chemical Physics*, vol. 31, p. 459, Aug. 1959.

[80] A. Rahman, "Correlations in the motion of atoms in liquid argon," *Physical Review*, vol. 136, pp. A405–A411, Oct. 1964.

[81] F. H. Stillinger, "Improved simulation of liquid water by molecular dynamics," *The Journal of Chemical Physics*, vol. 60, p. 1545, Aug. 1974.

[82] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," *Journal of Molecular Biology*, vol. 103, pp. 227–49, May 1976.

[83] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, pp. 585–590, June 1977.

[84] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Challenges for density functional theory," *Chemical Reviews*, vol. 112, pp. 289–320, Jan. 2012.

[85] K. Burke and L. O. Wagner, "DFT in a nutshell," *International Journal of Quantum Chemistry*, vol. 113, pp. 96–101, Jan. 2013.

[86] M. P. Allen, "Introduction to molecular dynamics simulation," in *Computational Soft Matter: From Synthetic Polymers to Proteins*, John von Neumann Institute for Computing, Julich, 2004.

[87] S. A. Adcock and J. A. McCammon, "Molecular dynamics: survey of methods for simulating the activity of proteins," *Chemical Reviews*, vol. 106, pp. 1589–1615, May 2006.

[88] K. A. Fichthorn and W. H. Weinberg, "Theoretical foundations of dynamical Monte Carlo simulations," *The Journal of Chemical Physics*, vol. 95, p. 1090, July 1991.

[89] N. Metropolis and S. Ulam, "The Monte Carlo method," *Journal of the American Statistical Association*, vol. 44, pp. 335–41, 1949.

[90] E. Paquet and H. L. Viktor, "Molecular dynamics, Monte Carlo simulations, and Langevin dynamics: a computational review," *BioMed Research International*, vol. 2015, pp. 1–18, Nov. 2015.

[91] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Annalen der Physik*, vol. 389, no. 20, pp. 457–84, 1927.

[92] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 1996.

[93] A. R. Leach, *Molecular Modelling: Principles and Applications*. Prentice Hall, 2001.

[94] R. O. Dror, M. Ø. Jensen, D. W. Borhani, and D. E. Shaw, "Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations," *The Journal of General Physiology*, vol. 135, pp. 555–62, June 2010.

[95] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *Journal of Computational Physics*, vol. 23, pp. 327–41, Mar. 1977.

[96] H. C. Andersen, "Rattle: a "velocity" version of the shake algorithm for molecular dynamics calculations," *Journal of Computational Physics*, vol. 52, pp. 24–34, Oct. 1983.

[97] R. D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, G. S. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, J. L. J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, G. X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, T. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, Luchko, S. Gusarov, A. Kovalenko, and P. Kollman, "AMBER 12," 2012.

[98] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, "Scalable molecular dynamics with NAMD," *Journal of Computational Chemistry*, vol. 26, pp. 1781–802, Dec. 2005.

[99] W. C. Swope, "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters," *The Journal of Chemical Physics*, vol. 76, p. 637, Jan. 1982.

[100] J. D. Durrant and J. A. McCammon, "Molecular dynamics simulations and drug discovery," *BMC biology*, vol. 9, p. 71, Jan. 2011.

[101] E. Brunk and U. Rothlisberger, "Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states," *Chemical Reviews*, vol. 115, pp. 6217–63, June 2015.

[102] C. M. Baker, "Polarizable force fields for molecular dynamics simulations of biomolecules," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 5, pp. 241–54, Mar. 2015.

[103] S. W. Rick, S. J. Stuart, and B. J. Berne, "Dynamical fluctuating charge force fields: application to liquid water," *The Journal of Chemical Physics*, vol. 101, p. 6141, Oct. 1994.

[104] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, "ReaxFF: a reactive force field for hydrocarbons," *The Journal of Physical Chemistry A*, vol. 105, pp. 9396–409, Oct. 2001.

[105] R. W. Pastor and A. D. MacKerell, "Development of the CHARMM Force Field for Lipids," *The Journal of Physical Chemistry Letters*, vol. 2, pp. 1526–1532, jul 2011.

[106] T. J. Piggot, Á. Piñeiro, and S. Khalid, "Molecular dynamics simulations of phosphatidylcholine membranes: a comparative force field study," *Journal of Chemical Theory and Computation*, vol. 8, pp. 4593–609, Nov. 2012.

[107] A. P. Lyubartsev and A. L. Rabinovich, "Force Field Development for Lipid Membrane Simulations," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1858, no. 10, pp. 2483–2497, 2016.

[108] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, p. 3684, Oct. 1984.

[109] H. C. Andersen, "Molecular dynamics simulations at constant pressure and/or temperature," *The Journal of Chemical Physics*, vol. 72, p. 2384, July 1980.

[110] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *The Journal of Chemical Physics*, vol. 81, p. 511, July 1984.

[111] W. Hoover, "Canonical dynamics: equilibrium phase-space distributions," *Physical Review A*, vol. 31, pp. 1695–97, Mar. 1985.

[112] D. S. Lemons, "Paul Langevin's 1908 paper "On the Theory of Brownian Motion" ["Sur la théorie du mouvement brownien", C. R. Acad. Sci. (Paris) 146, 530-3 (1908)]," *American Journal of Physics*, vol. 65, p. 1079, Nov. 1997.

[113] S. Nosé and M. Klein, "Constant pressure molecular dynamics for molecular systems," *Molecular Physics*, vol. 50, pp. 1055–76, Aug. 2006.

[114] M. Parrinello and A. Rahman, "Crystal structure and pair potentials: a molecular dynamics study," *Physical Review Letters*, vol. 45, pp. 1196–9, Oct. 1980.

[115] M. Parrinello, "Polymorphic transitions in single crystals: a new molecular dynamics method," *Journal of Applied Physics*, vol. 52, p. 7182, Dec. 1981.

[116] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, "Constant pressure molecular dynamics simulation: the Langevin piston method," *The Journal of Chemical Physics*, vol. 103, p. 4613, Sept. 1995.

[117] P. Larsson, B. Hess, and E. Lindahl, "Algorithm improvements for molecular dynamics simulations," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, pp. 93–108, Jan. 2011.

[118] B. Roux and T. Simonson, "Implicit solvent models," *Biophysical Chemistry*, vol. 78, pp. 1–20, Apr. 1999.

[119] J. Michel and J. W. Essex, "Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations," *Journal of Computer-Aided Molecular Design*, vol. 24, pp. 639–58, Aug. 2010.

[120] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, no. 2, p. 926, 1983.

[121] J. Nocedal and S. J. Wright, *Numerical Optimisation.* New York, NY: Springer, 2006.

[122] H. Ode, M. Nakashima, S. Kitamura, W. Sugiura, and H. Sato, "Molecular dynamics simulation in virus research," *Frontiers in Microbiology*, vol. 3, p. 258, Jan. 2012.

[123] S. C. Kamerlin, S. Vicatos, A. Dryga, and A. Warshel, "Coarse-Grained (Multiscale) Simulations in Studies of Biophysical and Chemical Systems," *Annual Review of Physical Chemistry*, vol. 62, pp. 41–64, may 2011.

[124] P. J. Bond, J. Holyoake, A. Ivetac, S. Khalid, and M. S. P. Sansom, "Coarse-grained molecular dynamics simulations of membrane proteins and peptides," *Journal of Structural Biology*, vol. 157, pp. 593–605, Mar. 2007.

[125] M. Orozco, L. Orellana, A. Hospital, A. N. Naganathan, A. Emperador, O. Carrillo, and J. L. Gelpí, "Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings," *Advances in Protein Chemistry and Structural Biology*, vol. 85, pp. 183–215, Mar. 2011.

[126] C. Abrams and G. Bussi, "Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration," *Entropy*, vol. 16, Dec. 2014.

[127] R. C. Bernardi, M. C. R. Melo, and K. Schulten, "Enhanced sampling techniques in molecular dynamics simulations of biological systems," *Biochimica et biophysica acta*, vol. 1850, pp. 872–7, May 2015.

[128] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–80, May 1983.

[129] U. H. E. Hansmann, "Parallel tempering algorithm for conformational studies of biological molecules," *Chemical Physics Letters*, vol. 281, pp. 140–50, Dec. 1997.

[130] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chemical Physics Letters*, vol. 314, pp. 141–151, Nov. 1999.

[131] Y. Sugita and Y. Okamoto, "Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape," *Chemical Physics Letters*, vol. 329, pp. 261–70, Oct. 2000.

[132] J. G. Kirkwood, "Statistical Mechanics of Fluid Mixtures," *The Journal of Chemical Physics*, vol. 3, no. 5, p. 300, 1935.

[133] J. Schlitter, M. Engels, and P. Krüger, "Targeted molecular dynamics: a new approach for searching pathways of conformational transitions," *Journal of molecular graphics*, vol. 12, pp. 84–9, June 1994.

[134] E. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, "Constrained reaction coordinate dynamics for the simulation of rare events," *Chemical Physics Letters*, vol. 156, pp. 472–77, Apr. 1989.

[135] E. Darve, D. Rodríguez-Gómez, and A. Pohorille, "Adaptive biasing force method for scalar and vector free energy calculations," *The Journal of Chemical Physics*, vol. 128, p. 144120, Apr. 2008.

[136] G. Torrie and J. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling," *Journal of Computational Physics*, vol. 23, pp. 187–199, Feb. 1977.

[137] A. Laio and F. L. Gervasio, "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Reports on Progress in Physics*, vol. 71, p. 126601, Dec. 2008.

[138] B. Ensing, M. De Vivo, Z. Liu, P. Moore, and M. L. Klein, "Metadynamics as a tool for exploring free energy landscapes of chemical reactions," *Accounts of Chemical Research*, vol. 39, pp. 73–81, Feb. 2006.

[139] C. Bealing, R. Martonák, and C. Molteni, "Pressure-induced structural phase transitions in CdSe: a metadynamics study," *The Journal of Chemical Physics*, vol. 130, p. 124712, Mar. 2009.

[140] V. Leone, G. Lattanzi, C. Molteni, and P. Carloni, "Mechanism of action of cyclophilin a explored by metadynamics simulations," *PLoS Computational Biology*, vol. 5, p. e1000309, Mar. 2009.

[141] A. Cavalli, A. Spitaleri, G. Saladino, and F. L. Gervasio, "Investigating drug-target association and dissociation mechanisms using metadynamics-based algorithms," *Accounts of Chemical Research*, vol. 48, pp. 277–85, Feb. 2015.

[142] F. Comitani, C. Melis, and C. Molteni, "Elucidating ligand binding and channel gating mechanisms in pentameric ligand-gated ion channels by atomistic simulations," *Biochemical Society transactions*, vol. 43, pp. 151–6, Apr. 2015.

[143] A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: a smoothly converging and tunable free-energy method," *Physical Review Letters*, vol. 100, p. 020603, Jan. 2008.

[144] A. Laio, "Advanced sampling techniques for numerical simulations: lecture notes," 2008.

[145] G. A. Tribello, M. Ceriotti, and M. Parrinello, "A self-learning algorithm for biased molecular dynamics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 17509–14, Oct. 2010.

[146] S. Piana and A. Laio, "A bias-exchange approach to protein folding," *The Journal of Physical Chemistry B*, vol. 111, pp. 4553–4559, May 2007.

[147] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello, "PLUMED: a portable plugin for free-energy calculations with molecular dynamics," *Computer Physics Communications*, vol. 180, no. 10, pp. 1961–72, 2009.

[148] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "PLUMED 2: New feathers for an old bird," *Computer Physics Communications*, vol. 185, pp. 604–13, Feb. 2014.

[149] C. Micheletti, A. Laio, and M. Parrinello, "Reconstructing the density of states by history-dependent metadynamics," *Physical Review Letters*, vol. 92, p. 170601, Apr. 2004.

[150] M. Bonomi, A. Barducci, and M. Parrinello, "Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics," *Journal of Computational Chemistry*, vol. 30, pp. 1615–21, Aug. 2009.

[151] S. J. Teague, "Implications of protein flexibility for drug discovery," *Nature reviews: drug discovery*, vol. 2, pp. 527–41, July 2003.

[152] E. Fischer, "Einfluss der Configuration auf die Wirkung der Enzyme," *Berichte der deutschen chemischen Gesellschaft*, vol. 27, pp. 2985–93, Oct. 1894.

[153] T. Hou, J. Wang, Y. Li, and W. Wang, "Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations," *Journal of Chemical Information and Modeling*, vol. 51, pp. 69–82, Jan. 2011.

[154] B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg, "MMPBSA.py : an efficient program for end-state free energy calculations," *Journal of Chemical Theory and Computation*, vol. 8, pp. 3314–21, sep 2012.

[155] J. T. Kim, A. D. Hamilton, C. M. Bailey, R. A. Domaoal, R. A. Domoal, L. Wang, K. S. Anderson, and W. L. Jorgensen, "FEP-guided selection of bicyclic heterocycles in lead optimization for non-nucleoside inhibitors of HIV-1 reverse transcriptase," *Journal of the American Chemical Society*, vol. 128, pp. 15372–3, Dec. 2006.

[156] B. L. Tembre and J. Mc Cammon, "Ligand-receptor interactions," *Computers & Chemistry*, vol. 8, pp. 281–83, Jan. 1984.

[157] H.-J. Woo and B. Roux, "Calculation of absolute protein-ligand binding free energy from computer simulations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 6825–30, May 2005.

[158] F. Comitani, N. Cohen, J. Ashby, D. Botten, S. C. R. Lummis, and C. 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*, vol. 28, pp. 35–48, Jan. 2014.

[159] F. L. Gervasio, A. Laio, and M. Parrinello, "Flexible docking in solution using metadynamics," *Journal of the American Chemical Society*, vol. 127, pp. 2600–7, Mar. 2005.

[160] V. Limongelli, M. Bonomi, and M. Parrinello, "Funnel metadynamics as accurate binding free-energy method," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 6358–63, Apr. 2013.

[161] M. Ceriotti, G. A. Tribello, and M. Parrinello, "Simplifying the representation of complex free-energy landscapes using sketch-map," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 13023–8, Aug. 2011.

[162] G. A. Tribello, M. Ceriotti, and M. Parrinello, "Using sketch-map coordinates to analyze and bias molecular dynamics simulations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 5196–201, Apr. 2012.

[163] M. Ceriotti, G. A. Tribello, and M. Parrinello, "Demonstrating the Transferability and the Descriptive Power of Sketch-Map," *Journal of Chemical Theory and Computation*, vol. 9, pp. 1521–1532, Mar. 2013.

[164] I. Borg and P. J. F. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. Springer Science & Business Media, 2005.

[165] F. Comitani, V. Limongelli, and C. 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*, vol. 12, pp. 3398–3406, July 2016.

[166] F. Comitani, K. Rossi, M. Ceriotti, M. E. Sanz, and C. Molteni, "Mapping the conformational free energy of aspartic acid in gas phase and aqueous solution," *in preparation*.

[167] J. A. Ashby, I. V. McGonigle, K. L. Price, N. Cohen, F. Comitani, D. a. Dougherty, C. Molteni, and S. C. R. Lummis, "GABA binding to an insect

GABA receptor: a molecular dynamics and mutagenesis study," *Biophysical Journal*, vol. 103, pp. 2071–81, Nov. 2012.

[168] I. McGonigle and S. C. R. Lummis, "Molecular characterization of agonists that bind to an insect GABA receptor," *Biochemistry*, vol. 49, pp. 2897–902, Apr. 2010.

[169] C. D. Dellisanti, Y. Yao, J. C. Stroud, Z.-Z. Wang, and L. Chen, "Crystal structure of the extracellular domain of nAChR α1 bound to α-bungarotoxin at 1.94 Å resolution," *Nature Neuroscience*, vol. 10, pp. 953–62, Aug. 2007.

[170] H. Nury, N. Bocquet, C. Le Poupon, B. Raynal, A. Haouz, P.-J. Corringer, and M. Delarue, "Crystal structure of the extracellular domain of a bacterial ligand-gated ion channel," *Journal of Molecular Biology*, vol. 395, pp. 1114–27, Feb. 2010.

[171] M. Zouridakis, P. Giastas, E. Zarkadas, D. Chroni-Tzartou, P. Bregestovski, and S. J. Tzartos, "Crystal structures of free and antagonist-bound states of human α9 nicotinic receptor extracellular domain," *Nature Structural & Molecular Biology*, vol. 21, pp. 976–80, Nov. 2014.

[172] J. Shi, T. L. Blundell, and K. Mizuguchi, "FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties," *Journal of Molecular Biology*, vol. 310, pp. 243–57, June 2001.

[173] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. López, J. D. Thompson, T. J. Gibson, and D. G. Higgins, "Clustal W and Clustal X version 2.0," *Bioinformatics*, vol. 23, pp. 2947–8, Nov. 2007.

[174] S. D. Buckingham, P. C. Biggin, B. M. Sattelle, L. A. Brown, and D. B. Sattelle, "Insect GABA receptors: splicing, editing, and targeting by antiparasitics and insecticides," *Molecular Pharmacology*, vol. 68, pp. 942–51, Oct. 2005.

[175] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper, and A. Sali, "Comparative protein structure modeling using MODELLER," *Current Protocols in Protein Science*, Nov. 2007.

[176] F. Melo, R. Sánchez, and A. Sali, "Statistical potentials for fold assessment," *Protein Science*, vol. 11, pp. 430–448, Apr. 2009.

[177] B. John and A. Sali, "Comparative protein structure modeling by iterative alignment, model building and model assessment," *Nucleic Acids Research*, vol. 31, pp. 3982–92, July 2003.

[178] M.-Y. Shen and A. Sali, "Statistical potential for assessment and prediction of protein structures," *Protein Science*, vol. 15, pp. 2507–24, Nov. 2006.

[179] A. Sali, *Modeval, Model Evaluation Server*. https://modbase.compbio.ucsf.edu/evaluation/ [Accessed: 04-03-2012].

[180] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede, "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling," *Bioinformatics*, vol. 22, pp. 195–201, Jan. 2006.

[181] P. Benkert, S. C. E. Tosatto, and D. Schomburg, "QMEAN: A comprehensive scoring function for model quality assessment," *Proteins*, vol. 71, pp. 261–77, Apr. 2008.

[182] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, "PROCHECK: a program to check the stereochemical quality of protein structures," *Journal of Applied Crystallography*, vol. 26, pp. 283–91, Apr. 1993.

[183] R. A. Laskowski, V. V. Chistyakov, and J. M. Thornton, "PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids," *Nucleic Acids Research*, vol. 33, pp. D266–8, Jan. 2005.

[184] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations," *Journal of Computational Chemistry*, vol. 24, pp. 1999–2012, Dec. 2003.

[185] K. A. Beauchamp, Y.-S. Lin, R. Das, and V. S. Pande, "Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements," *Journal of Chemical Theory and Computation*, vol. 8, pp. 1409–14, Apr. 2012.

[186] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility," *Journal of Computational Chemistry*, vol. 30, pp. 2785–91, Dec. 2009.

[187] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor, "Improved protein-ligand docking using GOLD," *Proteins*, vol. 52, pp. 609–23, Sept. 2003.

[188] J. Hutter and M. Iannuzzi, "CPMD: Car-Parrinello molecular dynamics," *Zeitschrift für Kristallographie - Crystalline Materials*, vol. 220, pp. 549–1, Jan. 2005.

[189] N. Troullier and J. L. Martins, "Efficient pseudopotentials for plane-wave calculations," *Physical Review B*, vol. 43, pp. 1993–2006, Jan. 1991.

[190] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Physical Review Letters*, vol. 77, pp. 3865–8, Oct. 1996.

[191] A. Bruno, M. Scrima, E. Novellino, G. D'Errico, A. M. D'Ursi, and V. Limongelli, "The glycan role in the glycopeptide immunogenicity revealed by atomistic simulations and spectroscopic experiments on the multiple sclerosis biomarker CSF114(Glc)," *Scientific Reports*, vol. 5, p. 9200, Jan. 2015.

[192] Y.-W. Hsiao and P. Söderhjelm, "Prediction of SAMPL4 host-guest binding affinities using funnel metadynamics," *Journal of Computer-Aided Molecular Design*, vol. 28, pp. 443–54, Apr. 2014.

[193] L. Troussicot, F. Guillière, V. Limongelli, O. Walker, and J.-M. Lancelin, "Funnel-metadynamics and solution NMR to estimate protein−ligand affinities," *Journal of the American Chemical Society*, vol. 137, pp. 1273–81, Jan. 2015.

[194] W. Y. Lee, C. R. Free, and S. M. Sine, "Binding to gating transduction in nicotinic receptors: Cys-loop energetically couples to pre-M1 and M2-M3 regions," *The Journal of Neuroscience*, vol. 29, pp. 3189–99, Mar. 2009.

[195] J. Yakel, "Advances and hold-ups in the study of structure, function and regulation of Cys-loop ligand-gated ion channels and receptors," *The Journal of Physiology*, vol. 588, pp. 555–6, Feb. 2010.

[196] J. Zheng and M. C. Trudeau, *Handbook of ion channels*. Taylor & Francis, 2015.

[197] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the Amber ff99SB protein force field," *Proteins*, vol. 78, no. 8, pp. 1950–1958, 2010.

[198] G. M. J. Frisch, 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, and J. L. Sonnenberg, "Gaussian 09, Revision C.01," 2009.

[199] S. Amiri, M. S. P. Sansom, and P. C. Biggin, "Molecular dynamics studies of AChBP with nicotine and carbamylcholine: the role of water in the binding pocket," *Protein Engineering, Design & Selection*, vol. 20, pp. 353–9, July 2007.

[200] A. P. Blum, H. A. Lester, and D. A. Dougherty, "Nicotinic pharmacophore: the pyridine N of nicotine and carbonyl of acetylcholine hydrogen bond across a subunit interface to a backbone NH," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 13206–11, June 2010.

[201] C. X. Weichenberger and M. J. Sippl, "NQ-Flipper: recognition and correction of erroneous asparagine and glutamine side-chain rotamers in protein structures," *Nucleic Acids Research*, vol. 35, pp. W403–6, May 2007.

[202] W. Hester, *Exploring the Interface Between the Extracellular and the Transmembrane Domain of the 5-HT$_3$ Receptor: a Molecular Dynamics and Metadynamics Study.* Bachelor's degree thesis, King's College London, 2015.

[203] P. J. Corringer, S. Bertrand, J. L. Galzi, A. Devillers-Thiéry, J. P. Changeux, and D. Bertrand, "Mutational analysis of the charge selectivity filter of the alpha7 nicotinic acetylcholine receptor," *Neuron*, vol. 22, pp. 831–43, Apr. 1999.

[204] A. Keramidas, A. J. Moorhouse, P. R. Schofield, and P. H. Barry, "Ligand-gated ion channels: mechanisms underlying ion selectivity," *Progress in Biophysics and Molecular Biology*, vol. 86, pp. 161–204, Oct. 2004.

[205] J. E. Carland, M. A. Cooper, M. R. Livesey, T. G. Hales, J. A. Peters, and J. J. Lambert, "Mutagenic analysis of the intracellular portals of the human 5-HT$_3$A receptor," *The Journal of Biological Chemistry*, vol. 288, pp. 31592–601, Nov. 2013.

[206] S. Jo, T. Kim, V. G. Iyer, and W. Im, "CHARMM-GUI: A web-based graphical user interface for CHARMM," *Journal of Computational Chemistry*, vol. 29, pp. 1859–65, Aug. 2008.

[207] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. Hauser, and C. Simmerling, "ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB," *Journal of Chemical Theory and Computation*, vol. 11, pp. 3696–3713, July 2015.

[208] C. J. Dickson, B. D. Madej, A. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould, and R. C. Walker, "Lipid14: the Amber lipid force field," *Journal of Chemical Theory and Computation*, vol. 10, pp. 865–79, Feb. 2014.

[209] T. Althoff, R. E. Hibbs, S. Banerjee, and E. Gouaux, "X-ray structures of GluCl in apo states reveal a gating mechanism of Cys-loop receptors," *Nature*, vol. 512, pp. 333–7, Aug. 2014.

[210] M. Gielen, P. Thomas, and T. G. Smart, "The desensitization gate of inhibitory Cys-loop receptors," *Nature Communications*, vol. 6, p. 6829, Apr. 2015.

[211] D. Di Maio, B. Chandramouli, and G. Brancato, "Pathways and barriers for ion translocation through the 5-HT_3A receptor channel," *PloS One*, vol. 10, p. e0140258, Jan. 2015.

[212] O. S. Smart, J. M. Goodfellow, and B. A. Wallace, "The pore dimensions of gramicidin A," *Biophysical Journal*, vol. 65, pp. 2455–60, Dec. 1993.

[213] A. V. Rossokhin and B. S. Zhorov, "Side chain flexibility and the pore dimensions in the GABAA receptor," *Journal of Computer-Aided Molecular Design*, vol. 30, pp. 559–67, July 2016.

[214] A. Amadei, A. B. Linssen, and H. J. Berendsen, "Essential dynamics of proteins," *Proteins*, vol. 17, pp. 412–25, Dec. 1993.

[215] C. Camilloni, A. B. Sahakyan, M. J. Holliday, N. G. Isern, F. Zhang, E. Z. Eisenmesser, and M. Vendruscolo, "Cyclophilin A catalyzes proline isomerization by an electrostatic handle mechanism," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, pp. 10203–10208, July 2014.

[216] P. Velisetty, S. V. Chalamalasetti, and S. Chakrapani, "Conformational transitions underlying pore opening and desensitization in membrane-embedded Gloeobacter violaceus ligand-gated ion channel (GLIC)," *The Journal of Biological Chemistry*, vol. 287, pp. 36864–72, Oct. 2012.

[217] L. Sauguet, A. Shahsavar, F. Poitevin, C. Huon, A. Menny, A. Nemecz, A. Haouz, J.-P. Changeux, P.-J. Corringer, and M. 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*, vol. 111, pp. 966–71, Jan. 2014.

[218] A. Marabelli, R. Lape, and L. Sivilotti, "Mechanism of activation of the prokaryotic channel ELIC by propylamine: a single-channel study," *The Journal of General Physiology*, vol. 145, pp. 23–45, Jan. 2015.

[219] K. C. Timberlake, *General, Organic, and Biological Chemistry: Structures of Life*. Pearson Benjamin Cummings, 2006.

[220] A. Lesarri, S. Mata, J. C. López, and J. L. Alonso, "A laser-ablation molecular-beam Fourier-transform microwave spectrometer: the rotational spectrum of organic solids," *Review of Scientific Instruments*, vol. 74, p. 4799, Apr. 2003.

[221] J. C. López, M. Isabel Peña, M. Eugenia Sanz, and J. L. Alonso, "Probing thymine with laser ablation molecular beam Fourier transform microwave spectroscopy," *The Journal of Chemical Physics*, vol. 126, no. 19, p. 191103, 2007.

[222] S. L. Stephens, D. P. Tew, N. R. Walker, and A. C. Legon, "H3P· · ·AgI: generation by laser-ablation and characterization by rotational spectroscopy and ab initio calculations," *Physical Chemistry Chemical Physics*, vol. 18, pp. 18971–18977, June 2016.

[223] D. A. McQuarrie and J. D. Simon, *Molecular Thermodynamics*. University Science Books, 1999.

[224] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. Mackerell, "Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles," *Journal of Chemical Theory and Computation*, vol. 8, pp. 3257–73, Sept. 2012.

[225] W. L. Jorgensen and J. Tirado-Rives, "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin," *Journal of the American Chemical Society*, vol. 110, pp. 1657–66, Mar. 1988.

[226] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids," *Journal of the American Chemical Society*, vol. 118, pp. 11225–36, Jan. 1996.

[227] B. Hashemian and M. Arroyo, "Topological obstructions in the way of data-driven collective variables," *The Journal of Chemical Physics*, vol. 142, p. 044102, Jan. 2015.

[228] B. Roux, O. S. Andersen, and T. W. Allen, "Comment on "Free energy simulations of single and double ion occupancy in gramicidin A" [J. Chem. Phys. 126, 105103 (2007)]," *The Journal of Chemical Physics*, vol. 128, p. 227101; author reply 227102, June 2008.