

Innovative high-performance computing approaches for molecular neuromedicine

Von der Fakultät für Mathematik und Naturwissenschaften der RWTH Aachen University
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Alessandro Marchetto, M. Sc.

aus

Legnago (VR), Italien

Berichter: Univ.-Prof. Dr. phil. nat. Gabriele Pradel
Univ.-Prof. Dr. Giulia Rossetti, Ph.D.
Univ.-Prof. Paolo Carloni, Ph.D.

Tag der mündlichen Prüfung: 21.04.2026

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek verfügbar.

Eidesstattliche Erklärung

Alessandro Marchetto erklärt hiermit, dass diese Dissertation und die darin dargelegten Inhalte die eigenen sind und selbstständig, als Ergebnis der eigenen originären Forschung, generiert wurden. Hiermit erkläre ich an Eides statt:

- a) Diese Arbeit wurde vollständig oder größtenteils in der Phase als Doktorand dieser Fakultät und Universität angefertigt.
- b) Sofern irgendein Bestandteil dieser Dissertation zuvor für einen akademischen Abschluss oder eine andere Qualifikation an dieser oder einer anderen Institution verwendet wurde, wurde dies klar angezeigt.
- c) Wenn immer andere eigene oder Veröffentlichungen Dritter herangezogen wurden, wurden diese klar benannt.
- d) Wenn aus anderen eigenen oder Veröffentlichungen Dritter zitiert wurde, wurde stets die Quelle hierfür angegeben. Diese Dissertation ist vollständig meine eigene Arbeit, mit der Ausnahme solcher Zitate.
- e) Alle wesentlichen Quellen von Unterstützung wurden benannt.
- f) Wenn immer ein Teil dieser Dissertation auf der Zusammenarbeit mit anderen basiert, wurde von mir klar gekennzeichnet, was von anderen und was von mir selbst erarbeitet wurde.
- g) Ein Teil oder Teile dieser Arbeit wurden zuvor veröffentlicht, wie unter *List of Publications* aufgeführt.

Ort, Datum: Jülich, October 21, 2025

Unterschrift:

List of Publications

Parts of this thesis have been published or were submitted for publication in accordance with §5(4) in the following papers:

1. Kjølbbye, L. R., Pereira, G. P., Bartocci, A., Pannuzzo, M., Albani, S., **Marchetto, A.**, ..., Souza, P. C. (2022). Towards design of drugs and delivery systems with the Martini coarse-grained model. *QRB Discovery*, 3, e19.
2. **Marchetto, A.**, Si Chaib, Z., Rossi, C. A., Ribeiro, R., Pantano, S., Rossetti, G., Giorgetti, A. (2020). CGMD platform: integrated web servers for the preparation, running, and analysis of coarse-grained molecular dynamics simulations. *Molecules*, 25(24), 5934.

The perspective section of the Martini 3 force field (above **1.**) is included in Chapter 2; Chapter 3 is based on **2.**.

My contribution to these articles in accordance with §5(6) is the following:

1. I reviewed the manuscript and contributed to the writing throughout the preparation process.
2. I designed and implemented the entire platform architecture (PHP, JavaScript, and HTML), executed multiple use-cases (available in the CGMD Platform), analyzed and tested the systems with Dr. Zeineb Si Chaib under the supervision of Prof. Dr. Giulia Rossetti, Prof. Alejandro Giorgetti, Prof. Sergio Pantano, Dr. Rui Pedro Fernandes Ribeiro and Prof. Dr. Paolo Carloni, and contributed to the manuscript writing. CSS stylesheets were developed in collaboration with Carlo Alberto Rossi.

Abstract

G protein-coupled receptors (GPCRs) are integral membrane protein characterized by a seven trans-membrane scaffold. They are involved in the regulation of several cellular downstream pathways, and they are the target of more than one third of the current available drugs on the market. Among GPCRs, dopamine D2 represents the main target of both first- and second-generation antipsychotic drugs used to treat several neurological disorders, such as schizophrenia and Parkinson's disease. In cells, D2 can exist as a monomer, dimer, or higher-order oligomer. The presence of agonists or antagonists affects dimer formation; in particular, the proportion of D2 dimers increases from approximately 30% to 45% upon agonist activation. To date, no experimental structural information on D2 dimers has been available.

Here, I developed a computational protocol that combines coarse-grained molecular dynamics simulations based on the Martini 3 force field combined with well-tempered metadynamics enhanced sampling to predict the structural determinants of D2 dimerization. My simulations are consistent with experimental observations and show that the protocol is sensitive not only to the activation state of the receptors, but also to the lipid membrane composition, which plays a crucial role in receptor signaling and membrane transport.

This approach can naturally be extended to D2 hetero-oligomers and other membrane protein complexes, such as those involving the adenosine A2 receptor, with which it shares an intimate interaction.

Overall, my results provide a structural framework and a computational strategy that may serve as a solid basis for structure-based drug design targeting membrane receptor complexes.

Kurzfassung

G-Protein-gekoppelte Rezeptoren (GPCRs) sind integrale Membranproteine, die durch ein Gerüst aus sieben Transmembrandomänen charakterisiert sind. Sie sind an der Regulation zahlreicher intrazellulärer Signalwege beteiligt und stellen das Ziel von mehr als einem Drittel der derzeit verfügbaren Medikamente auf dem Markt dar. Unter den GPCRs ist der Dopamin-D2-Rezeptor das Hauptziel sowohl der Antipsychotika der ersten als auch der zweiten Generation, die zur Behandlung verschiedener neurologischer Erkrankungen wie Schizophrenie und Morbus Parkinson eingesetzt werden. In Zellen kann D2 als Monomer, Dimer oder höhergeordnetes Oligomer vorkommen. Das Vorhandensein von Agonisten oder Antagonisten beeinflusst die Dimerbildung; insbesondere steigt der Anteil der D2-Dimere bei Agonistenaktivierung von etwa 30% auf 45%. Bisher lagen keine experimentellen Strukturdaten zu D2-Dimeren vor.

In dieser Arbeit habe ich ein rechnergestütztes Protokoll entwickelt, das grobkörnige Molekulardynamik-Simulationen auf der Grundlage des Martini-3-Kraftfelds mit Well-Tempered-Metadynamics-Enhanced-Sampling kombiniert, um die strukturellen Determinanten der D2-Dimerisierung vorherzusagen. Meine Simulationen stimmen mit experimentellen Beobachtungen überein und zeigen, dass das Protokoll nicht nur gegenüber dem Aktivierungszustand der Rezeptoren empfindlich ist, sondern auch gegenüber der Zusammensetzung der Lipidmembran, die eine entscheidende Rolle bei der Rezeptorsignalisierung und dem Membrantransport spielt.

Dieser Ansatz lässt sich auf D2-Heterooligomere und andere Membranproteinkomplexe, wie etwa solche mit dem Adenosin-A2-Rezeptor, mit dem D2 eng interagiert, natürlich erweitern.

Insgesamt liefern meine Ergebnisse einen strukturellen Rahmen und eine rechnergestützte Strategie, die als solide Grundlage für das strukturbasierte Wirkstoffdesign gegen Membranrezeptorkomplexe dienen kann.

*To my parents,
who have always supported me...*

Index

Eidesstattliche Erklärung	iii
List of Publications	v
Abstract	vi
Kurzfassung	vii
Index	ix
List of Figures	xiii
List of Tables	xv
Acknowledgements	xvii
Introduction	xix
1 Biological background	3
1.1 G-protein Coupled receptor: introduction and classification according to the GRAFS system	3
1.1.1 Class A	4
1.1.2 Class B (B1 and B2)	5
1.1.3 Class C	5
1.1.4 Class F	5
1.2 Role of GPCRs, activation/deactivation mechanisms, and drug targeting	5
1.3 The Dopamine D2 receptor	6
1.4 Structural differences between the inactive and activate state of the D2	8
1.5 The Role of the Sodium Ion in the stabilization of the inactive configuration	10
1.6 The Adenosine A2 receptor	11
1.7 Homodimerization of the D2 and A2 receptors	12
1.8 Heterodimerization of the D2 and A2 receptors	13
1.9 The G_s and G_i : canonical and allosteric interactions	15
1.10 Residence Time	17
1.10.1 Mechanisms of Ligand-Receptor Interaction	18
1.10.2 Drug potency and the relationship with the residence time	19
1.10.3 Techniques to evaluate drug lifetime and the lifetime of D2 drugs	19
2 Methods	23
2.1 Computational methods to reconstruct 3D protein structures	23
2.1.1 Model Building and Loop Modeling	24
2.1.2 Modeling Validation	25
2.2 Molecular Dynamics	25
2.3 General expression of a force field	26
2.4 Statistical mechanics	27
2.4.1 Ergodic Hypothesis	27
2.4.2 Equations of motion	28
2.4.3 Numerical integration of Newton Equation	28

2.5	Thermostats and Barostats	29
2.6	Periodic Boundary Conditions	31
2.7	LINCS Algorithm	32
2.8	Enhanced Sampling Techniques	32
2.8.1	Metadynamics and Well-Tempered Metadynamics	32
2.8.2	The choice of the optimal CVs	34
2.8.3	Plumed 2	35
2.9	The Martini 3 force field	36
2.9.1	Beads radii and beads interaction	37
2.9.2	Elastic network	37
2.9.3	Bonded and non bonded parameters in Martini 3	39
2.9.4	Insane	40
2.10	MD Simulation Protocol setup	40
2.10.1	Backmapping	41
2.10.2	The Martini 3 Force Field: Future Perspectives on the Horizon	42
2.11	Contribution of this Chapter	44
3	Coarse-Grained Molecular Dynamics (CGMD) Platform	47
3.1	CGMD Platform documentation	48
3.2	CGMD Platform: computational details	49
3.3	CGMD Platform pipelines	49
3.3.1	The SIRAH Pipeline	49
3.3.2	Martini 2.2 with membrane	50
3.3.3	Backmapping procedure	50
3.4	CGMD Platform as part of the Human Brain Project (HBP)	51
3.5	CGMD Platform: Application Cases	51
3.5.1	Martini 2.2 with Membrane	51
3.5.2	Water Martini	51
3.5.3	The SIRAH Force field	52
3.6	Future Perspectives of the CGMD Platform	52
3.7	Conclusions	52
3.8	Contribution of this Chapter	53
4	GPCRs oligomerization protocol	55
4.1	First module: the Choice of the Collective Variables (CVs) for the Well-tempered Metadynamics	56
4.1.1	Analysis of the first CV: the distance of centers of mass	57
4.1.2	Analysis of the second CV: the transmembrane atoms	57
4.1.3	Well-tempered Metadynamics	58
4.1.4	Clusterization of the structures in the basins	59
4.1.5	Determination of the population of each cluster and evaluation of the oligomeric interfaces	60
4.2	Second module: Unbiased Simulation	60
4.2.1	Analysis of contacts	60
4.2.2	Analysis of the orthosteric and vestibular binding pockets	61
4.2.3	Analysis of putative CHOL binding spots	62
4.3	Membrane environments: from the CNS towards the CHOL Enriched membrane	62
4.3.1	Central Nervous System (CNS) membrane	63
4.3.2	CHOL Enriched Membrane	64
4.4	Current Limitations of the protocol: optimizing the modeling in the next future	64
4.5	Contribution of this Chapter	65
5	Results	67
5.1	Free Energy Profiles	67
5.1.1	Free Energy Surface: D2 homodimer in its inactive state in the CNS Membrane	67
5.1.2	Free Energy Surface: D2 homodimer in its active state in the CNS Membrane	68
5.2	Analysis of the D2 dimer in its active state	69
5.2.1	Analysis of interactions	69
5.2.2	D2 Dimer: my Protocol vs. AlphaFold 3 Predictions	71

5.3	Analysis of CHOL binding spots in the dimer	72
5.3.1	CHOL binding spots found with all-atom simulations	74
5.4	Analysis of CHOL binding spots in the monomer	75
5.4.1	Volume changes upon D2 dimer formation in the active state	75
5.5	Analysis of the D2 dimer in its inactive state	78
5.6	Analysis of the D2 dimer in its active state in the CHOL Enriched membrane	80
5.7	Analysis of the D2 dimer in its inactive state in the CHOL Enriched membrane	82
5.8	Conclusions	83
5.9	Contribution to this Chapter	84
6	Conclusions and Future Outlooks	85
7	Appendix	89
Appendix		89
Appendix A		89
Calculation of CHOL lipid count		89
Visualization of the interaction count		91
Normality plots		94
Bibliography		114

List of Figures

1.1	GPCR architecture	3
1.2	GPCRs classes	4
1.3	Canonical activation Mechanism of a GPCR	6
1.4	Topology of the Dopamine D2 receptor	7
1.5	The hydrophobic cage that entraps the antagonist within the binding cavity	8
1.6	The antipsychotic risperidone bound to the D2 in its inactive state	9
1.7	Alignment D2 active state - D2 inactive state	9
1.8	Structural similarity of D2-like receptors	10
1.9	Topology of the Adenosine A2 receptor	11
1.10	Receptor association within the membrane	12
1.11	Dynamic equilibrium monomer-to-dimer D2	12
1.12	Allosterism between A2 and D2	14
1.13	Allosteric and canonical pathway of the A2-D2 heterotetramer	16
1.14	Configuration of the A2-D2 heterotetramer in complex with the AC5	17
2.1	Sequence alignment zones in homology modeling.	24
2.2	Protein Boundary Condition (PBC) definition. PBCs mimic an infinite system using a finite simulation box. Instead of simulating a massive number of particles, one can simulate in a smaller box that is conceptually repeated infinitely in all directions.	31
2.3	The Well-tempered Metadynamics Technique scheme	34
2.4	The scheme of the DEEP-LDA algorithm	35
2.5	The scheme of the CG mapping	36
2.6	The Elastic network scheme	38
2.7	The scheme of the Martini 3 CG backmapping	41
2.8	TEM-1 β lactamase complexed with a high-temperature-destabilizing allosteric inhibitor bound within its cryptic binding pocket	42
2.9	Example of nanoparticle shuttle system carrying a drug, in this example Risperidone.	43
3.1	Membrane Composition	48
3.2	Configuration of the SIRAH Pipeline embedded in the web platform	49
3.3	Configuration of the Martini 2.2 Pipeline embedded in the web platform	50
3.4	Backmapping workflow	51
3.5	Set of functionalities offered by CGMD Platform	52
4.1	Set-up of the first CV	56
4.2	Set-up of the second CV	57
4.3	D2 monomer and the region considered for the MDPocket Analysis	62
4.4	Schematic view of the cellular membrane	63
5.1	Free Energy Surface of the D2 homodimer in its inactive state	68
5.2	Free Energy Surface of the D2 homodimer in its active state	68
5.3	Contacts within the interface for the D2 in its active state	69
5.4	Contacts for TM4 and TM5 in the D2 in its active state	70
5.5	Contacts for TM4 and TM5 in the D2 in its active state depicted every 5 μs	71
5.6	Superimposition of the D2 structure found with my protocol and the D2 structure predicted by AlphaFold 3 multimer	72
5.7	CHOL Lipid Count - Dimer D2 Active	73

5.8	CHOL Lipid Count - Dimer D2 Active	73
5.9	CHOL Hotspots found averaged after 1 μs all-atom simulation	74
5.10	CHOL Lipid Count - Monomer D2 Active	75
5.11	Boxplot showing the Volume and SASA of the orthosteric and vestibular binding cavities of the monomer and the D2 dimer in its active state	75
5.12	Boxplot showing the Volume and SASA of the orthosteric binding cavities of the monomer and the D2 dimer in its active state	76
5.13	RMSD Value	76
5.14	U test conducted on the properties of Volume and SASA of the monomer and dimer binding cavities	77
5.15	U test conducted on the properties of Volume and SASA of the monomer and dimer binding cavities	77
5.16	Contacts within the interface for the D2 in its inactive state	79
5.17	Radar plot monitoring the contacts within the interface	79
5.18	Lipid rafts in GPCRs	80
5.19	Free Energy Surface of the D2 homodimer in its active state in the CHOL Enriched membrane	81
5.20	Density distribution of the five lipids of the CHOL Enriched Membrane	81
5.21	Free Energy Surface of the D2 homodimer in its inactive state in the CHOL Enriched membrane	82
5.22	Density distribution of the five lipids of the CHOL Enriched Membrane	83
1	Histogram and Q-Q plot to check the normality of the distribution of SASA values (pock_asa) of the dimer OBC 1-Vest	94
2	Histogram and Q-Q plot to check the normality of the distribution of Volume values (pock_volume) of the dimer OBC 1-Vest	94
3	Histogram and Q-Q plot to check the normality of the distribution of SASA values (pock_asa) of the dimer OBC 2-Vest	94
4	Histogram and Q-Q plot to check the normality of the distribution of Volume values (pock_volume) of the dimer OBC 2-Vest	95
5	Histogram and Q-Q plot to check the normality of the distribution of SASA values (pock_asa) of the Monomer OBC-Vest	95
6	Histogram and Q-Q plot to check the normality of the distribution of Volume values (pock_volume) of the Monomer-OBC-Vest	95
7	Histogram and Q-Q plot to check the normality of the distribution of SASA values (pock_asa) of the dimer OBC 1	96
8	Histogram and Q-Q plot to check the normality of the distribution of Volume values (pock_volume) of the dimer OBC 1	96
9	Histogram and Q-Q plot to check the normality of the distribution of SASA values (pock_asa) of the dimer OBC 2	96
10	Histogram and Q-Q plot to check the normality of the distribution of Volume values (pock_volume) of the dimer OBC 2	97
11	Histogram and Q-Q plot to check the normality of the distribution of SASA values (pock_asa) of the Monomer OBC	97
12	Histogram and Q-Q plot to check the normality of the distribution of Volume values (pock_volume) of the Monomer-OBC	97

List of Tables

1.1	Summary of the mechanism of action, propensities and adverse effects for FGA and SGA. Table adapted from[41].	8
1.2	Table of Molecule Affinities with their dissociation half-life $t_{1/2}$	19
2.1	Bead-Type,Size and vdw Radii	37
2.2	Bead-Bead Interaction and Lennard Jones potential	37
4.1	Lipid composition of the Upper and Lower Leaflets of the CNS membrane	64
4.2	Lipid composition of the Upper and Lower Leaflets of the CHOL Enriched membrane	64
5.1	Contacts among TM4 and TM5 helices and their opposite monomers	71

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Dr. Dr. Giulia Rossetti, for giving me the opportunity to pursue my Ph.D., for her invaluable guidance, and for presenting me with the intriguing research question explored in this thesis. I am also profoundly grateful for her unwavering support and assistance throughout my Ph.D. journey, regardless of the challenges encountered.

I would further like to extend my sincere thanks to Prof. Dr. Paolo Carloni and Prof. Dr. Alejandro Giorgetti, at Forschungszentrum Jülich and the University of Verona, respectively, for their exceptional guidance and support. I am also immensely grateful to Prof. Dr. phil. nat. Gabriele Pradel for being my supervisor at the Biology Department of RWTH Aachen University.

My gratitude also goes to Prof. Dr. Mark Spehr, Dr. Jennifer Spehr, and Oke Nommensen for including me in the Research Training Group (RTG) 2416 MultiSenses–MultiScales, and for providing invaluable assistance, stimulating discussions, and enjoyable retreats.

I have greatly enjoyed working in a collegial and supportive environment. For this reason, I extend my heartfelt thanks to all members of the IAS-5/INM-9 institute, especially Alessia Piergentili, Simone Albani, Simone Mariani, Matteo Bersani, Giuseppe Silvestri, Giacomo Salvadori, Dr. Benjamin Joseph, Verena Weber, Marta Devodier, and Federico Natale. I am particularly grateful to Dr. Emiliano Ippoliti for his outstanding IT support and for sharing the joy of watching Formula 1[©] together; his reliability has been invaluable to the functioning of the institute. I also wish to thank Dr. Vania Calandrini for her moral support and insightful guidance throughout my Ph.D. I am further indebted to Petra Rott, Sabrina Zander, and Elisa Polese for their exceptional administrative support over the years. Additionally, I would like to thank my dear friend Andrea for accompanying me on various travels and for always being there for me.

Finally, I would like to thank my girlfriend, Alessia Visigalli, for sharing both the challenging moments and the joys of life with me.

Introduction

G-protein-coupled receptors (GPCRs) are integral membrane proteins composed of seven transmembrane helices. Class A GPCRs are known to exert their own function as monomers, but emerging evidence suggests that GPCRs can form also transient dimers or higher-order oligomers, possibly impacting on cellular functions. This thesis is a contribution to understand the homodimerization mechanisms of one neuroreceptors, Dopamine D2 (D2) - protein highly expressed in the brain that plays pivotal roles in neurodegenerative diseases like Parkinson's disease and schizophrenia. Using advanced molecular dynamics and enhanced sampling techniques, this work aims to capture potential dimer configurations, providing structures and the ground basis for future experimental validation, drug discovery studies, and kinetic and thermodynamic studies.

Chapter 1 begins with an overview of the GPCRs GRAFS classification system and the classical activation pathway of class A GPCRs. The second section focuses on the dopaminergic system (encompassing the D1, D2, D3, D4, and D5, with the D1-D5 belonging to the D1-like class, and the rest to the D2-like class) in the human brain, emphasizing the importance of D2 as a main target of antipsychotics. The third section examines A2 and D2 homodimerization, highlighting how the activation of the receptor induced by a molecule (generally a drug) can modify oligomerization and impact the dimer's stability in the brain. Finally, the chapter explores the heterodimerization of D2 and A2, highlighting its influence on regulating canonical and allosteric pathways of these GPCRs.

Chapter 2 starts explaining the methods employed in this thesis. The first section focuses on homology modeling, outlining the steps from sequence alignment and template selection to model refinement and quality assessment for D2 and A2 receptors. This is followed by an explanation of classical molecular dynamics principles, such as periodic boundary conditions, thermostats, barostats, and the numerical integration of Newton's equations of motion. The next section explains the enhanced sampling methodology used in this study to describe and capture association and dissociation events, specifically well-tempered metadynamics. The final sections of the chapter examine the recently developed coarsegrained (CG) Martini 3 force field, which was utilized for simulations and structural detection of the dimer. This section outlines the benefit of using the CG Martini 3 force-field, and also this chapter discusses bead radii and interactions, elastic networks, the use of insane for embedding protein in the membrane, bonded and non-bonded parameters, and the backmapping procedure to recover atomistic structures. Lastly, the chapter highlights the future applications of the Martini 3 force-field, such as detecting cryptic pockets, targeting protein-protein interactions (PPIs), describing soft nanoparticles, and identifying protein binding pockets embedded within membrane bilayers.

Chapter 3 describes the Coarse-grained Molecular Dynamics (CGMD) Platform. This platform is a web-server that aims to prepare and describe CG simulations via different type of analysis, such as Root-Mean Square Deviation (RMSD), temperature, pressure, volume, by using either of the following two force-fields: the Martini 2.2 and the Southamerican Initiative for a Rapid and Accurate Hamiltonian (SIRAH). The CGMD Platform possesses a on-the-fly documentation through which the user can be documented at each step of the preparation of the system. To the best of my knowledge, this is the first web platform that embrace these two force-field in one practical web-interface. The protocol presented in Chapter 4 will be implemented in the next future on this web platform and extended to all users around the globe.

Chapter 4 presents the protocol utilized in this thesis. Unlike other computational protocols found in the literature, the approach developed here leverages on two main features: **it does not rely on prior knowledge of the target final complex - namely it is a "blind protocol"**, and **it simultaneously accounts for the membrane environment, taking into account interactions formed between the protein complexes and the membrane itself**. The protocol can be divided in two different macro modules: Well-tempered metadynamics and unbiased simulations modules. The first module exhibits the tuning used for the two CVs used to push on the slow modes of the system,

namely the distance of the center of masses between the two subunits on the membrane and the number of contacts of the transmembrane region of the helices of the GPCRs. The first module continues by showing the process of clusterization, detailing how structures are selected and the multiple criteria exploited to retrieve the most appropriate configurations. The second module examines properties evaluated on the structures collected from the first module. To evaluate the stability of the quaternary complex of the formed structure, analysis of the contacts at its interface and surrounding annular lipid are performed. Also calculation of Volume and Solvent Accessible Surface Area (SASA) are conducted on the orthosteric binding cavities and vestibular binding cavities of the dimer and monomer to evaluate the impact of dimerization: this analysis is important to determine the ground basis for future pharmaphocore design, as well as kinetic and thermodynamic calculations. All these analysis have been conducted on the unbiased 40 μ s CG simulations on the structures extracted after the first module. Finally, this chapter concludes by showing the two membranes used in my simulations, namely the Central Nervous System (CNS) and the cholesterol Enriched (CHOL Enriched) membrane.

Chapter 5 shows the results obtained by my protocol run on the D2 receptor homodimers in the two different membrane environments explained in the previous chapter. **The most significant aspect of this work is demonstrating that this protocol can generate structures with specific structural determinants, featuring, in particular, TM4 and TM6 as signature helices, consistent with experimental findings across two distinct membrane environments.** Moreover, lipid impact on the structures is evaluated, highlighting the impact of CHOL and other lipids surrounding the subunits of the dimer.

Chapter 1

Biological background

1.1 G-protein Coupled receptor: introduction and classification according to the GRAFS system

G-protein coupled receptors (GPCRs) are membrane proteins characterized by a seven transmembrane spanning architecture, with the N-terminus facing the extracellular space and the C-terminus oriented towards the cytosol (Figure 1.1). They represent one of the largest protein families in the mammalian genome[1, 2, 3, 4]. GPCRs form the largest family of membrane proteins and play a key role in mediating a myriad of cellular responses to hormones and neurotransmitters[5, 6]. They are also essential for functions such as vision, smell, and taste[5, 6]. GPCRs are associated with a plethora of cardiovascular and neurological diseases, such as Alzheimer's disease (AD), type 2 diabetes, obesity, depression, and cancer[2].

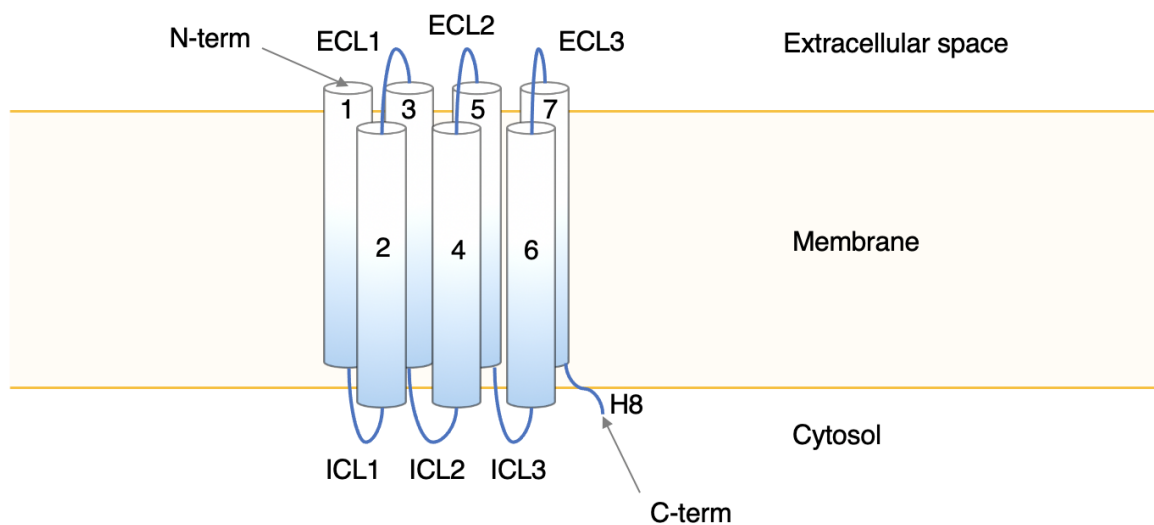


Figure 1.1: GPCR seven transmembrane spanning architecture connected by three extracellular (ECL1, ECL2, and ECL3) and intracellular (ICL1, ICL2, and ICL3) loops. The N-terminus is exposed in the extracellular space, while the C-terminus is located in the cytosol. The H8 helix has an amphipathic character. Numbering scheme refers to the alpha-helix. Membrane (in light yellow) separates the Extracellular space from the Intracellular space (cytosol).

According to the most recent classification system - GRAFS[7] - GPCRs can be divided in five classes: A (Rhodopsin-like), B1 (Secretin), B2 (Adhesion), C (Glutamate), and F (Frizzled) (See Fig 1.2)[2].

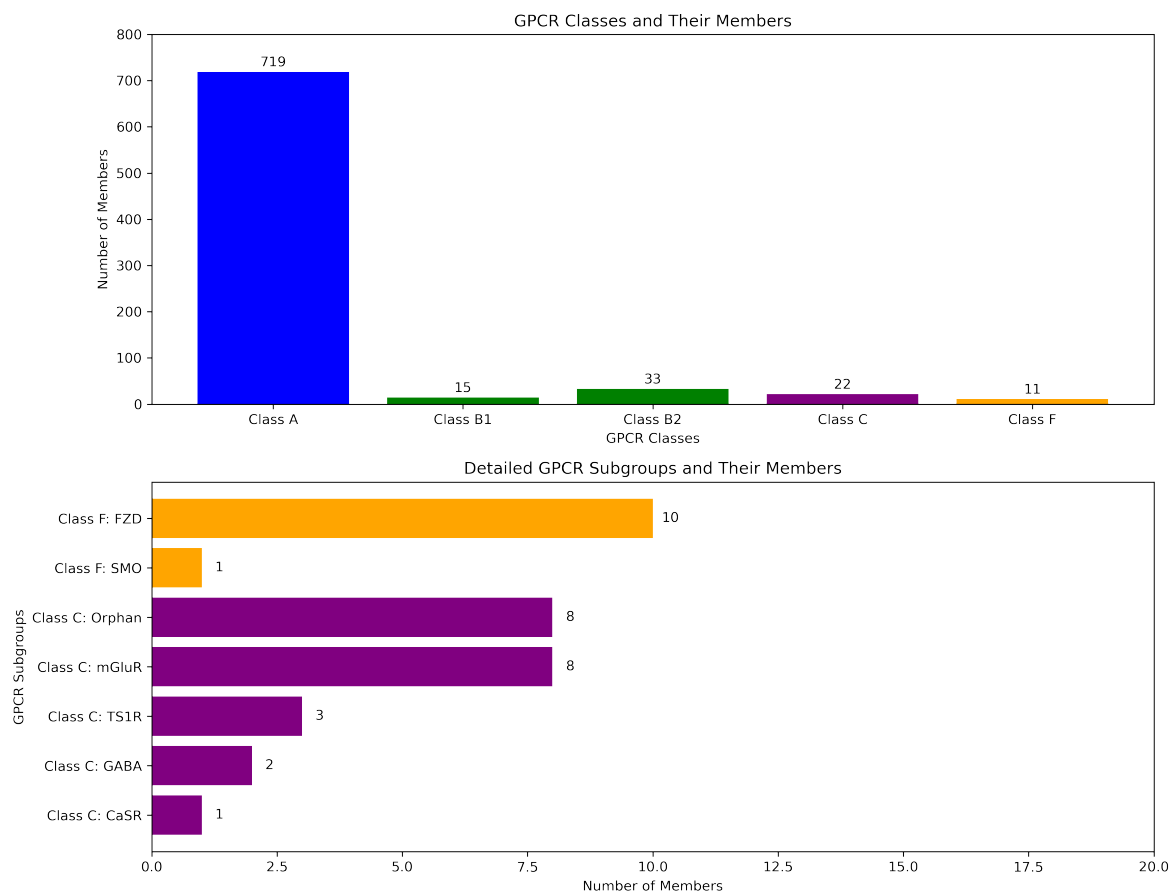


Figure 1.2: GPCRs classes classified accordingly to the GRAFS system[7]. In the upper plot, the number and different classes of GPCRs are indicated; in the lower plot, an overview is provided of the subgroups into which classes F and C are divided.

1.1.1 Class A

Class A GPCRs, constituting the largest group among GPCRs with 719 members are also referred to as the "rhodopsin-like family"[2, 8]. Class A is subdivided into multiple categories, including peptides, proteins, orphan receptors, lipids, steroids, aminergic receptors, sensory receptors, melatonin receptors, nucleotide receptors, and aliphatic acid receptors[2]. They are constituted by an extracellular N-terminal, a common seven transmembrane helices pattern and an additional eighth amphiphilic helix with a palmitoylated Cys at the C-terminal[8]. The C-ter Cys is an important post-translational modification implicated in various roles, including dimerization[9]. Around 700 drugs targeting class A GPCRs are currently available on the market, and many of them display multiple receptor target[10]. These drugs are used against a wide range of disorders, such as glaucoma, analgesics, allergies, cardiovascular diseases, schizophrenia, hypertension, pulmonary diseases, depression, migraine, and Parkinson's disease (PD)[2]. It is also important to underline that sensory and alicarboxylic acid receptors (which represents a minority of this class, only the 6% of it) possess a significant therapeutic potential[2]. Notably, prostanoid, chemokine, and melanocortin receptors collectively entail over 8% of clinical trial targets within class A[2].

1.1.2 Class B (B1 and B2)

This class of GPCRs is differentiated into two distinct subfamilies: secretin (B1) and adhesion (B2), containing 15 and 33 members, respectively[7]. The first subfamily is characterized by a large N-terminal domain protruded to the extracellular side and binds a different variety of ligands: peptides, hormones, and growth factors. Cryo-EM technique have so far resolved the B1 receptor structures in their agonist-bound states, all stabilized through approaches such as Nb35, dominant-negative G_s protein, or NanoBiT tethering[11, 12, 13]. The only exception is represented by the glucagon receptor (GCGR)- G_i protein complex[11]. The B2 subfamily is identified by its adhesion and migration properties and consists of nine subgroups distinguished by a singular N-terminal motif, such as epidermal growth factor, cadherin, and immunoglobulin domains[2]. Additionally, the B2 subfamily is peculiar for its autoproteolysis-inducing domain and proteolysis site, both mediating the activation via a Stachel sequence, resulting in the production of N-terminal fragment (NTF) and C-terminal fragment (CTF)[2].

1.1.3 Class C

Class C (also known as glutamate receptor class) encompasses 22 receptors divided into 5 subfamilies: 8 metabotropic glutamate receptors (mGluR1–8), 8 orphan GPCRs, 3 taste 1 receptors (TS1R1–3), 2 gamma-aminobutyric acid (GABA) type B receptors (GABAB1 and GABAB2), and 1 calcium-sensing receptor (CaSR). The glutamate subfamily presents a large extracellular domain (ECD) and they require dimer constitution to initiate receptor activation[2, 14]. Two conserved structural determinants belong to the ECD, namely the venus fly trap (VFT) and cysteine-rich domain (CRD) on the ligand binding site[2]. Key examples of class C GPCRs include mGluRs and GABA_B receptors, which modulate synaptic transmission in the CNS, crucial for learning, memory, and neuroplasticity[15]. CaSR plays a vital role in maintaining calcium balance and regulating parathyroid hormone (PTH) secretion[16]. Moreover, taste receptors detect sweet and umami flavors, contributing to metabolic regulation and taste perception[17]. These receptors are also involved for therapeutic applications due to their involvement in crucial physiological pathways[18].

1.1.4 Class F

Ten members of this class are Frizzled receptors (FZD1–10), which facilitate Wnt signaling and are the key for both embryonic development and the function of adult organisms[2]. The eleventh member of this class, namely the smoothened (SMO) receptor, has been approved as a drug target, with its small molecule antagonists authorized as anti-neoplastic agents[2]. Class F GPCRs regulate several processes, such as cellular polarity, migration, and adult tissue homeostasis[19].

1.2 Role of GPCRs, activation/deactivation mechanisms, and drug targeting

GPCRs are crucial in mediating the majority of physiological responses to a plethora of molecules: neurotransmitters, amines, carbohydrates, lipids, odors, photon, hormones, and other stimuli coming from outside[2, 5, 20, 21]. The general mechanism of activation of a GPCR can be delineated into three key stages: (i) interaction with an agonist¹, (ii) induction of a receptor conformational change (also known as receptor dynamics[21]), and (iii) subsequent activation of a heterotrimeric G-protein. The GPCRs ligand-bound complexes promote and catalyze the exchange of GTP for GDP from the G_α subunit[23]. When bound to GTP (present within the cell in high concentrations), G_α dissociates from $G_{\beta\gamma}$, allowing G_α and $G_{\beta\gamma}$ to independently trigger intracellular signaling processes[5, 20, 21, 24] (Figure 1.3). At this point, activated GPCRs would transmit signals downstream in the cell without any interruption[24]. Therefore, the mechanism interrupting this signaling process is driven by a C-terminal phosphorylation event, induced by G-protein-coupled receptor kinases (GRKs). As a result, this phosphorylation regulates the binding affinity of β -arrestin, leading to receptor desensitization via a plethora of mechanisms, such as clathrin-mediated endocytosis, steric hindrance, and receptor ubiquitination[24]. The GPCR-arrestin bound complex operates as a platform for over twenty different

¹Agonists are molecules characterized by both affinity and intrinsic efficacy, which induce a physiological response downstream in the cell. In contrast, antagonists molecules exhibit affinity for receptors but lack intrinsic efficacy, thus failing to elicit a biological response (22).

kinases, including mitogen-activated protein MAP kinases (MAPK), ERK1/2, c-Jun N-terminal and p38 kinases: this complex triggers activating G-protein-independent signaling pathways[24].

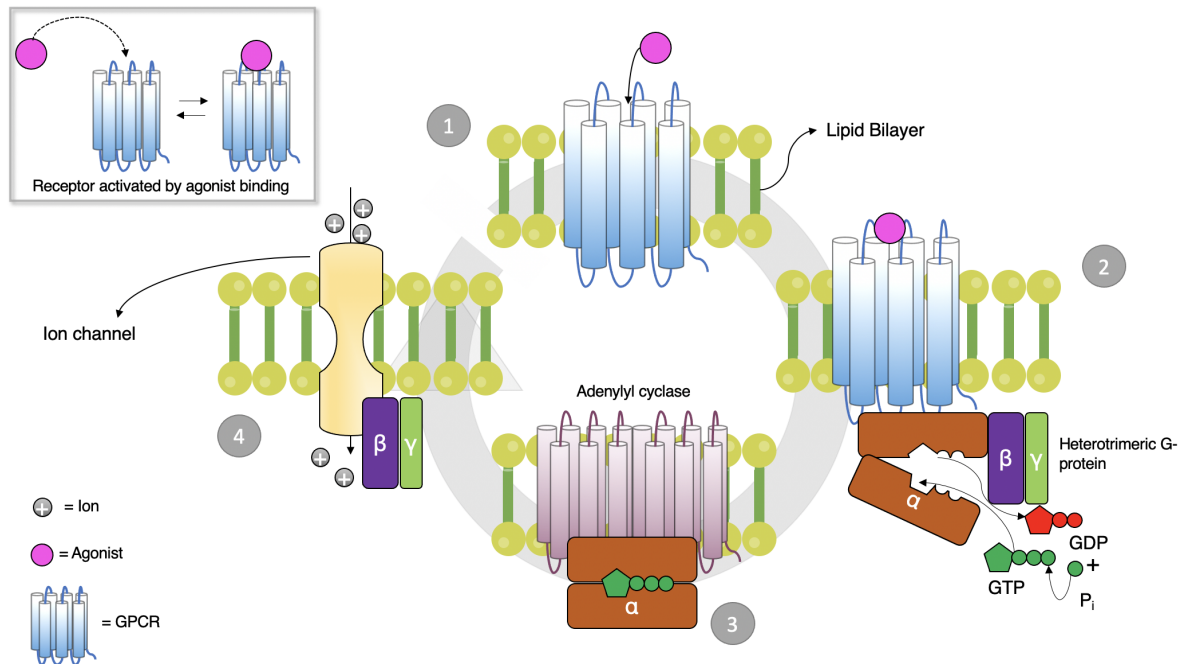


Figure 1.3: Classical activation pathway mechanism of a GPCR. When an agonist binds the canonical binding site (box and 1) of the GPCR, the protein undergoes a conformational change that promotes the recruitment of the G-protein (2). The activation of the receptor triggers the allosteric activation of the G-protein, fostering the exchange from GDP to GTP via phosphorylation within the Gα subunit (2). At this point, Gα dissociates from the βγ: the Gα interacts with the Adenylyl cyclase (3), the Gβγ migrates and modulates other proteins, such as an ion channel (4)[20, 24].

About 35%-50% of the currently available drugs on the market target GPCRs[3, 6, 24, 25, 26]. The European Medicines Agency (EMA) and the Food and Drug Administration (FDA) have approved more than 500 drugs targeting 110-130 distinct GPCRs, predominantly class A GPCRs, such as aminergic and opioid receptors[25]. It is not surprising that class A GPCRs, among which the Dopamine D2 and Adenosine A2 receptors presented in the section 1.3 and 1.6, respectively, are the primary targets for pharmaceutical companies, considering their accessibility and pivotal role in a large variety of disorders[27]. Indeed, several pharmacological drugs target the dopaminergic system, as it has been associated with several neuronal diseases (ND), such as PD, Huntington's disease (HD), and schizophrenia[28, 29].

1.3 The Dopamine D2 receptor

In the human brain and peripheral nervous system, five dopamine receptors (D1, D2, D3, D4, and D5) have been discovered[29, 30]. These receptors are essential for regulating all physiological functions involving the catecholamine dopamine[29]. Dopamine is an endogenous monoamine neurotransmitter produced in different areas of the brain and is involved in various brain functions, such as reward, metabolism, and hormonal secretion[28, 30, 31, 32, 33]. Dopaminergic system is also associated with the pathogenesis of several ND, such as PD, schizophrenia, depression, and attention deficit hyperactivity disorder (ADHD)[28, 34]. Dopamine receptors can be further differentiated in two classes, accordingly to the signalling downstream, position and pharmacology[30]: the D1-class (encompassing D1-D5) and the D2-class (entailing the D2, D3, and D4 instead)[29].

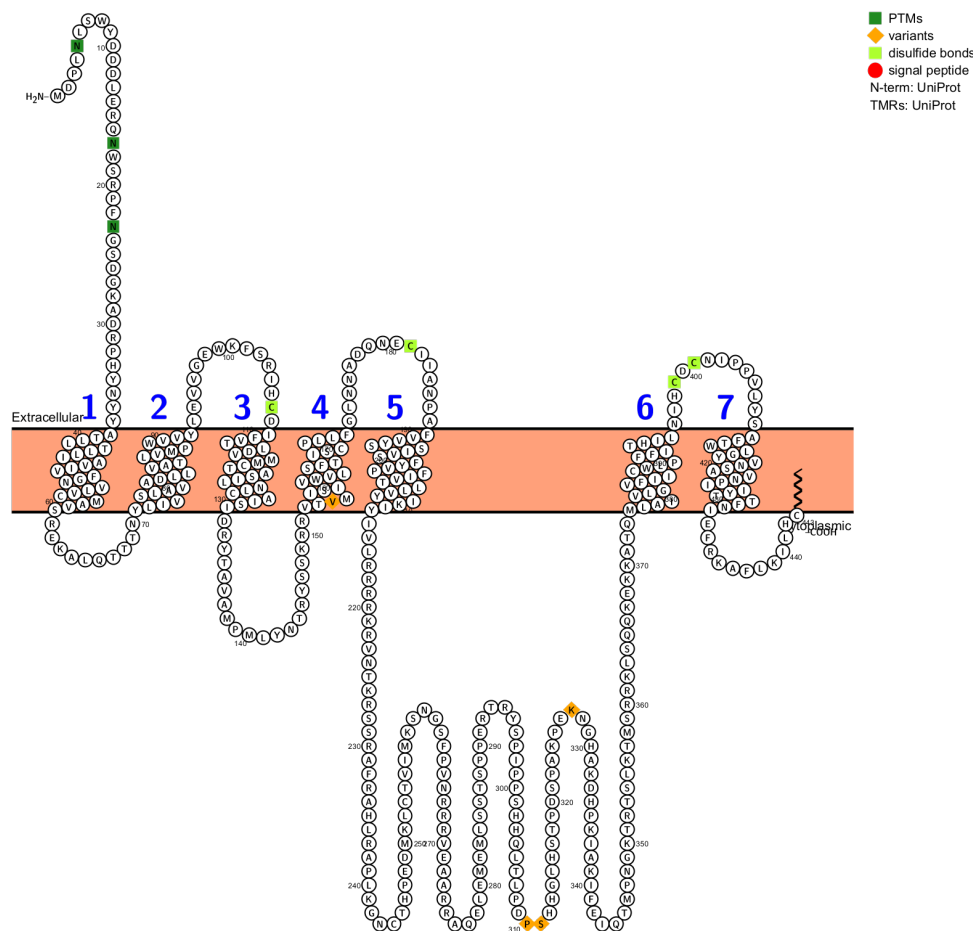


Figure 1.4: Topology of the Dopamine D2 receptor. The scheme was obtained by Protter[35].

The D2 receptor is a ~ 50 KDa single chain protein located in different areas of the midbrain, such as the *Substantia nigra pars compacta* or ventral tegmental area, but also in other regions such as cerebral cortex and nucleus accumbens[30, 33, 36]. D2 can also be expressed in other tissues, such as kidney or heart[36]. D2 exhibits three isoforms: two D2L and one D2S²[37]. The first isoform (see topology in Fig 1.4) sequence accounts for 443 amino acids with the C-terminus CYS443³ palmitoylated. D2 presents a long intracellular loop 3 (ICL3) that extends in the cytosol from ILE214 to GLN373, for a total of 159 residues. D2 contains also two intracellular disulfide bonds. The first bond measures 2.03 Å and it is established between CYS107 on extracellular loop 1 (ECL1) and CYS182 on ECL2. The second bond measures 2.04 Å and connects CYS399 and CYS401, both located on ECL3[28, 37]. D2 is the main target of the antipsychotic drugs[28].

Antipsychotics can be divided in two main categories: typical and atypical[28]. Typical antipsychotics or first-generation antipsychotic drugs (FGA) primarily target the D2, blocking it at the nigrostriatal and mesolimbic levels[38]. While very successful in treating ND like schizophrenia by blocking the dopamine D2 pathways, FGA drugs are associated with numerous side effects, including EPS[30, 38]. These side effects rise because of the excessive blockade of the D2[38]. On the other hand, with the aim of mitigating the negative symptoms associated with first-generation antipsychotics (FGAs), atypical antipsychotics, also known as second-generation antipsychotic drugs (SGAs), exert antagonistic activity not only at D2 receptors but also at other neurotransmitter receptors, such as the serotonin 5-HT_{2A} receptor (Table 1.1)[30, 39].

SGA drugs help reduce the EPS induced by typical antipsychotics; however, other adverse effects may arise due to their non-specific binding to various class A aminergic receptors, such as the serotonin 5-HT_{2C} receptor [30]. For the sake of clarity, it is worth briefly mentioning third-generation antipsychotics (TGAs), whose main innovation lies in their mechanism of action as partial agonists, such as

²L = long; S = short.

³I will use the three-letter amino acids code in this thesis.

lurasidone, cariprazine, brexpiprazole, and aripiprazole[40]. Ricci and colleagues emphasize that while TGAs are promising for psychotic disorders, they may not adequately address cognitive and compulsive symptoms often seen in drug abuse cases. Additionally, these TGAs may have limited effectiveness and cause several side effects, such as metabolic dysfunction and tardive dyskinesia. This uncertainty is due to their very recent introduction[40].

First Generation Antipsychotics	Second Generation Antipsychotics
D ₂ antagonism	5-HT _{2A} /D ₂ antagonism, Fast K_{off}
Reduction of positive symptoms	Reduction of EPS
Increase Risk of neurodiseases	Increase risk of metabolic diseases

Table 1.1: Summary of the mechanism of action, propensities and adverse effects for FGA and SGA. Table adapted from[41].

1.4 Structural differences between the inactive and activate state of the D2

The inactive structure (RCSB PDB: 6CM4[28]) of the D2 is a crystal structure (resolution 2.89 Å) bound to the risperidone, a SGA drug. The tetrahydropyridopyrimidinone moiety of risperidone (Figure 1.6) is oriented approximately perpendicular within the D2 main pocket due to the interaction with three hydrophobic residues, two located in the extracellular loops and the other one at level of helix 2⁴: Trp100^{ECL1}, Ile184^{ECL2}, and Leu94^{2.64} (Figure 1.5)[28]. The peculiar embedded configuration of Trp100^{ECL1} onto the extended binding pocket (EBP) is governed by the electronic density generated by the hydrophobic patch of these three amino acids[28].

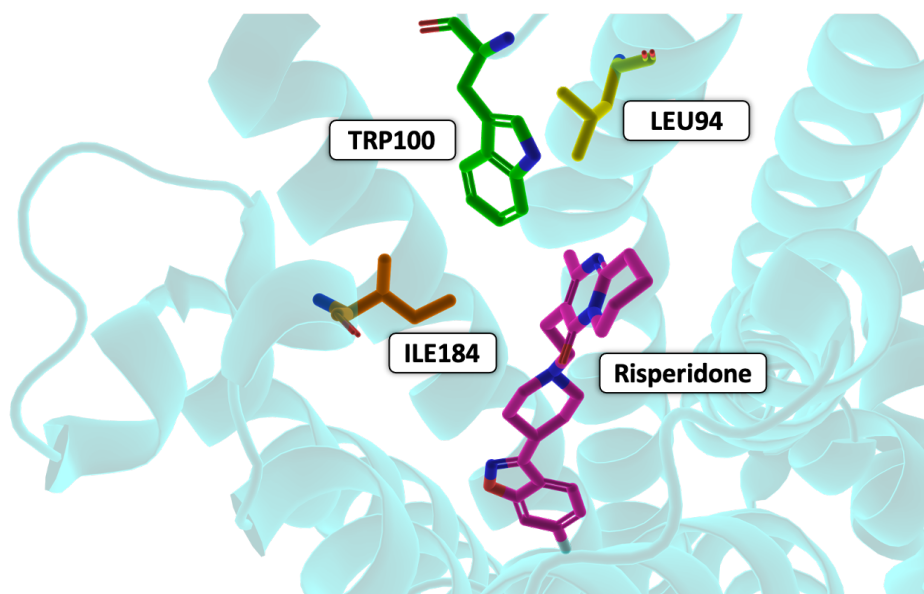


Figure 1.5: The hydrophobic cage that entraps the antagonist within the binding cavity interacting with the tetrahydropyridopyrimidinone moiety[28].

Secondly, the tertiary amine is positively charged, and interacts with the two oxygens of Asp114^{3.32}, forming an salt bridge that stabilizes the interaction in the binary complex risperidone-receptor (Figure 1.6)[28]. Eventually, the benzisoxazole moiety of risperidone is profoundly embedded at the bottom of the orthosteric binding cavity and it forges direct hydrophobic interactions with the toggle switch residue Trp386^{6.48} besides other residues, such as Cys118^{3.36}, Thr119^{3.37}, Ser197^{5.46}, Phe198^{5.47}, Phe382^{6.44}, and Phe390^{6.52}. This engagement network restricts the downward motion of Trp386^{6.48}, resulting in the constraining of the receptor in its inactive state[28, 43].

⁴The following nomenclature follows the Ballesteros–Weinstein numbering scheme (42).

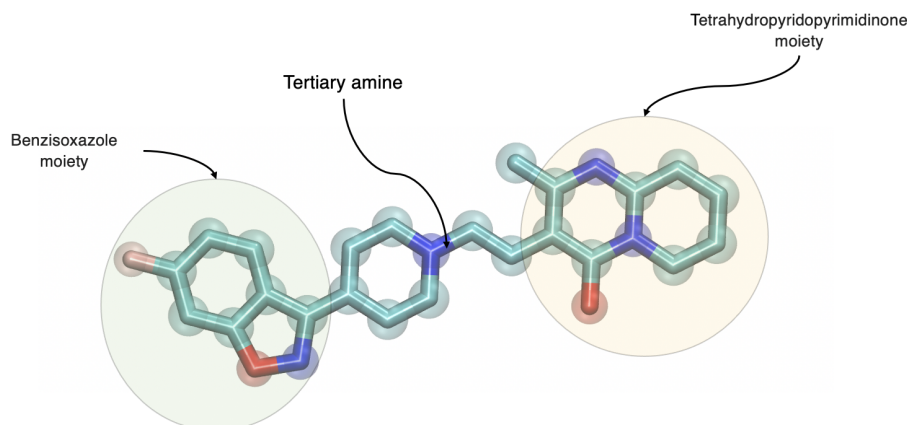


Figure 1.6: The antipsychotic risperidone bound to the D2 in its inactive state[28]. Risperidone can be largely divided in two big moieties, namely the tetrahydropyridopyrimidinone moiety (pale yellow) and the benzisoxazole moiety (pale green).

The activated structure (RCSB PDB: 7JVR[43]) is a 2.80 Å cryo-EM structure bound to a non-selective full agonist: the bromocriptine. The combined hydrophobic interactions between the bromocriptine and Trp386^{6.48} induce a downward movement of the activation switch residue Trp386^{6.48}. This motion propagates three conformational changes: one involving Arg132^{3.50} in the DRY motif, another involving Trp403^{6.44} in the PIF motif, and eventually a third inward motion of TM7[43]. These changes lead to the disruption of the ionic lock interaction established between Arg132^{3.50} and Asp389^{6.30}, particularly known for stabilizing the inactive state of the receptor. This cleavage, in turn, cause an outward shift of 9 Å in the intracellular part of TM6[43]. This orchestrated movement promotes the receptor to an activated state upon insertion of the α helix of $G_{\alpha i}$ into the transmembrane cavity (Figure 1.7)[43]. This activation pattern mechanism mirrors that observed in other class A GPCRs[43].

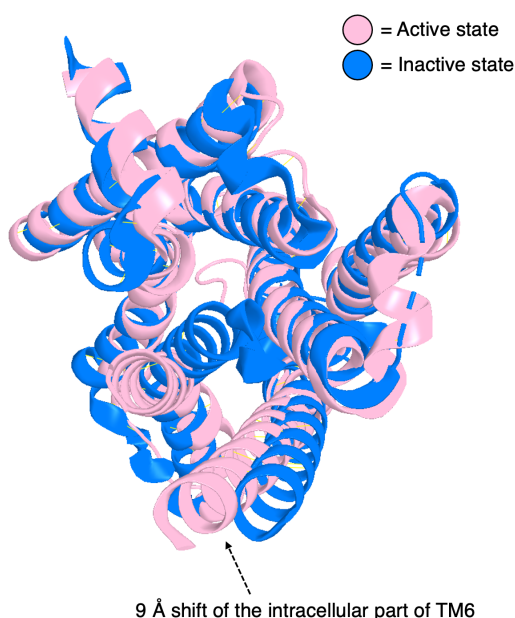


Figure 1.7: Alignment of the D2 active state (depicted in pink) and the D2 inactive state (depicted in blue). All ligands and crystallographic waters have been removed from the picture. The Root-Mean square deviation (RMSD) of the alignment resulted in 1.65 Å. The alignment and the RMSD value have been calculated with PyMOL[44].

1.5 The Role of the Sodium Ion in the stabilization of the inactive configuration

The effect of sodium ions has been documented since last century. In class A GPCRs, these ions induce stabilization of the receptor's inactive conformation[45]. This phenomenon is far from being isolated and indeed it has been observed across a wide array of class A GPCRs and is highly conserved owing to the Asp^{2.50} residue, which, in tandem with the Ser^{3.39}, coordinates the sodium ion[45]. This concerted action forms the so-called *ionic lock*. The first crystallographic evidence came out with the structure of the A2A adenosine receptor, and subsequent structures of the β_1 -adrenergic have revealed the same ion in the exact position, corroborating its role of "stabilizer" in the antagonistic state. However, the pocket that accommodates the sodium ion is not identical across all class A GPCRs. Structural differences can be observed, for example, among the A2A adenosine, β_1 -adrenergic receptors and the opioid receptors, resulting in distinct patterns for sodium coordination[45]. The key residue assortment involved in Na⁺ binding for class A GPCR that do not belong to the olfactory receptors pool are largely determined by seven conserved polar residues: Asp^{2.50} (90% conserved), Ser^{3.39} (75%), Asn^{1.50} (97%), Ser^{7.46} (66% as Ser and 75% as any polar side chain), Asn^{7.45} (70% as Asn or 90% as any polar side chain), Asn^{7.49} (75% as Asn, 20% as Asp), and Tyr^{7.53} (89% as Tyr)[45]. The polarity of these residues critically influences both sodium affinity and coordination mode[45]. To preserve the inactive conformation of a class A GPCR, an antagonist typically binds to the orthosteric site. Ligands resolved by X-ray crystallography - such as risperidone, eticlopride, and nemonapride - exhibit limited selectivity among the closely related D2-like receptor subtypes (D2, D3, and D4, respectively)[46]. Although the D2-like receptor subfamily shares a high degree of sequence and structural homology, the variations observed in their inactive conformations are more likely due to the distinct structural characteristics of the antagonists employed to stabilize each receptor (Figure 1.8)[46].

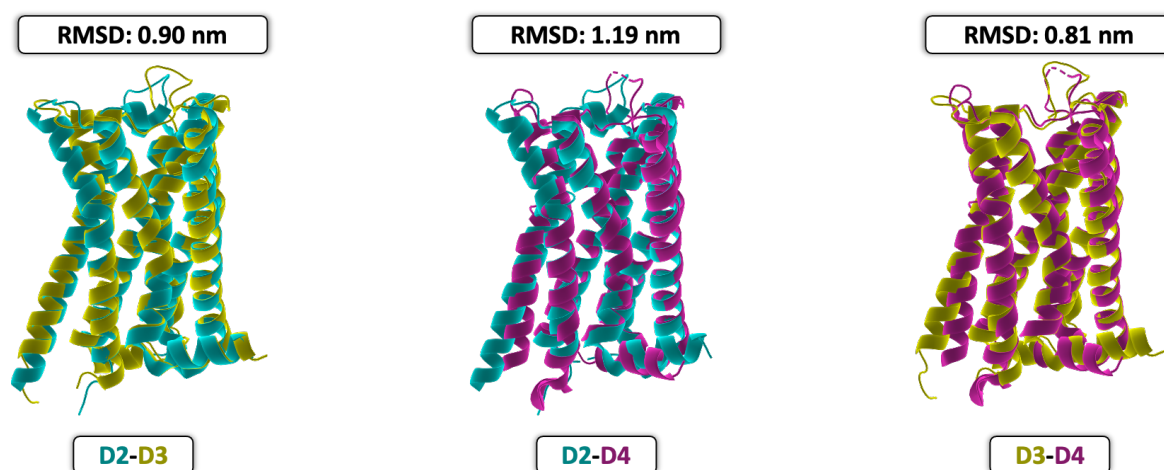


Figure 1.8: The D2-like receptors exhibit high structural similarity. From left to right, the RMSD between D2 and D3 receptors is 0.90 nm, between D2 and D4 is 1.19 nm, and between D3 and D4 is 0.81 nm. In the structural representations, the D2, D3, and D4 receptors are shown in cyan, yellow, and magenta, respectively. The alignment and the RMSD value have been calculated with PyMOL[44].

In the case of risperidone, it is well established in the literature that the ligand interacts with several residues located in the lower portion of the D2 receptor's orthosteric binding cavity, including Phe^{6.44}, Phe^{5.47}, Phe^{6.52}, Thr^{3.37}, Cys^{3.36}, Ser^{5.46}, and Ile^{3.40}[28]. Although the crystal structure reported by Wang in 2018 identifies an Ala instead of Ile at position 3.40, this mutation reveals minimal impact on risperidone binding[28, 46]. However, the occupancy of the Ile122^{3.40} sub-pocket by risperidone, combined with the neighbour Ser^{3.39}, one of the key residues involved in Na⁺ coordination, appears to render the binding of risperidone insensitive to the presence of sodium[46]. This phenomenon does not extend to all antagonists, for example, eticlopride and raclopride exhibit sodium sensitivity, suggesting that the ligand's scaffold plays a critical role in modulating this effect[46]. These findings have been experimentally validated using BRET assays to monitor G_o protein activity following D2 receptor activation[46]. Thus, depending on the occupancy of Ile122^{3.40}, both sodium sensitivity and the nature of inverse agonism can vary accordingly to the ligand[46]. Notably, in the structure reported by Wang,

no sodium ion is located in the above indicated coordination region, which supports the hypothesis that the occupancy of the Ile122^{3.40} sub-pocket by risperidone is sufficient to stabilize the receptor in its inactive conformation, even in the absence of Na⁺[28, 46]. Lastly, in my molecular dynamics simulations, elastic network restraints were applied. By their intrinsic nature, these restraints maintain the receptor in a fixed conformation - in this case, the inactive state - effectively rigidifying the D2 receptor tertiary structure. I kept the same scaffold (with the decorations explained in the next chapter) extracted from the X-ray structure published in 2018 by Wang.

1.6 The Adenosine A2 receptor

In the human brain, there are four Adenosine receptor subtypes: A1, A2A, A2B, and A3[47]. As the name suggests, they are the main target of the endogenous ligand adenosine, a nucleoside that plays a role in various physio/pathological processes by influencing numerous neuronal functions, including astrocytic activity, memory, synaptic plasticity, motor function, and aging[47, 48]. A1 and A2A possess higher affinity for adenosine[47].

The A2A (from this point simply A2) is a ~ 45 KDa protein mainly expressed in the striatum, but also in other different regions of the brain, such hippocampus, cerebral cortex, nucleus tractus solitarius, and glial cells[49]. A2 receptors play a relevant role in the pathogenesis of ND, such as PD and HD[50]. At level of primary sequence, A2 is constituted by 412 residues. A2 presents a long intracellular intrinsically disordered C-terminus and four pairs of disulfide bonds. The first disulfide bond connects CYS71^{ECL1} with CYS159^{ECL2} (length 2.05 Å); the second bond connects CYS74^{ECL1} with CYS146^{ECL2} (length 2.02 Å); the third bond takes place between CYS77^{ECL1} and CYS166^{ECL2} (length 2.03 Å); the latter bond is established between CYS259^{ECL3} and CYS262^{ECL3} (length 2.04 Å).

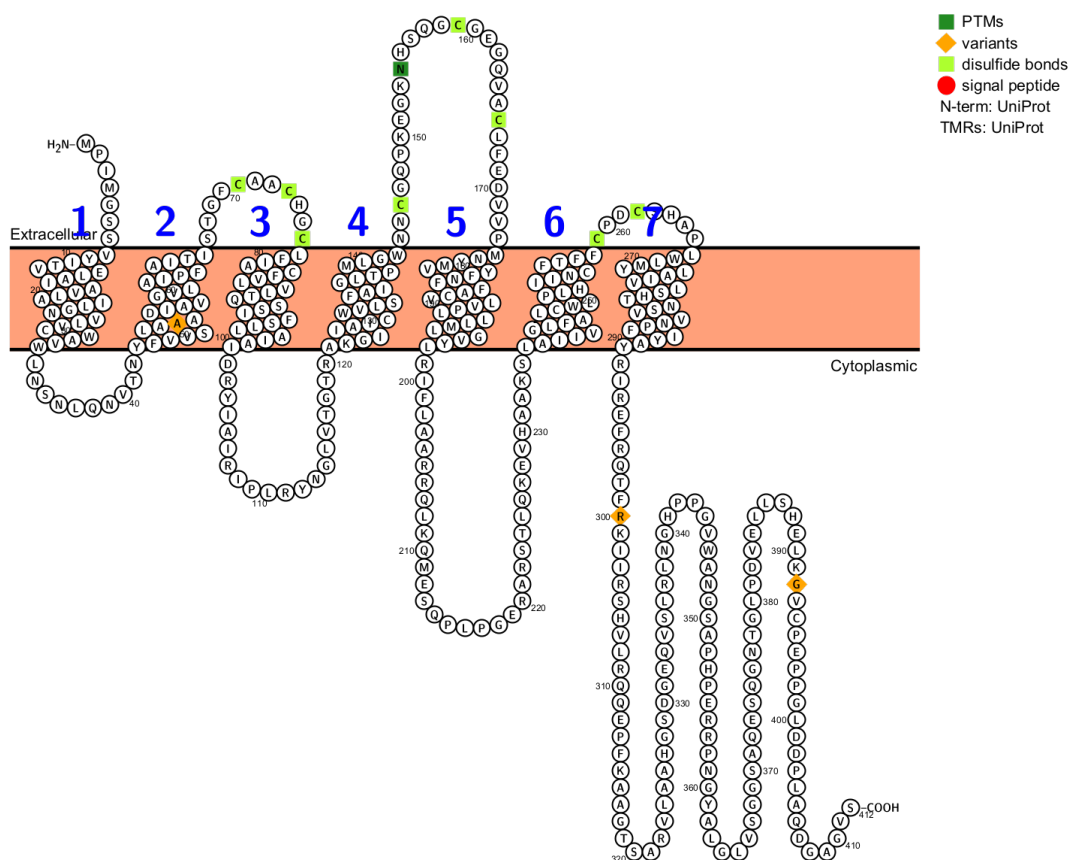


Figure 1.9: Topology of the Adenosine A2 receptor. The scheme was obtained by Protter[35].

1.7 Homodimerization of the D2 and A2 receptors

Initially, GPCRs were presumed to function only as standalone monomeric entities[51]. However, further investigations over the years unravelled that these membrane receptors were capable of forming homo- and heterodimers, as well as higher-order oligomeric complexes[51, 52]. This supramolecular organization - also known as *oligomerization* or, more in general, *protein-protein interaction* (PPI) - envisages that two or more proteins can allosterically interact with each other (Figure 1.10). This interaction is *not* covalent, and receptors within the membrane can interconvert from monomers to higher oligomeric forms and vice versa[36]. This crosstalk amplifies the functionality of the GPCRs landscape and the repertoire of recognition of several molecules by GPCRs, yielding the field of drug design more fertile than ever[52]: the surprising fact is that these newly formed entities exhibit *distinct* pharmacokinetic properties from their monomer counterpart, which can lead to a broader pharmacological spectrum[52, 53]. It is well-known from literature that class C GPCRs form dimers within the plasma membrane, such as the metabotropic glutamate receptors[54, 55], the calcium-sensing receptor[55, 56], or the GABAB receptor[55, 57]. But also class A GPCRs, such as Dopamine D2 and Adenosine A2 receptors, exhibit dimers and/or high order oligomers formation[36, 53, 55, 58, 59, 60]. To characterize and detect homo/heteromers in the membrane in living cells, Bioluminescence Resonance Energy Transfer (BRET) and Fluorescence Resonance Energy Transfer (FRET) have been extensively exploited[61, 62, 63]. However, these techniques does not monitor neither the dynamic spatial nor the lifespan behaviour of oligomerization at individual molecular level[61]. Therefore, latter researches suggested to focus more on the Total Internal Reflection Fluorescence (TIRF) Microscopy, as it allows to visualize single GPCR entities in real time in living cells[61, 64]. Via TIRF Microscopy, the diffusion and the supramolecular interaction of muscarinic acetylcholine M1, M2 receptors have been already observed and quantified[61, 65, 66]. Also the β_1 and β_2 receptors have been detected to form dimers within the membrane environment, and the ratio of β_1 and β_2 dimers is directly proportional to the level of the protein expression[61, 66, 67, 68]. Despite that, the role of ligands - whether agonists or antagonists - in regulating the dimerization process remains rather unclear for these class A GPCRs[67].

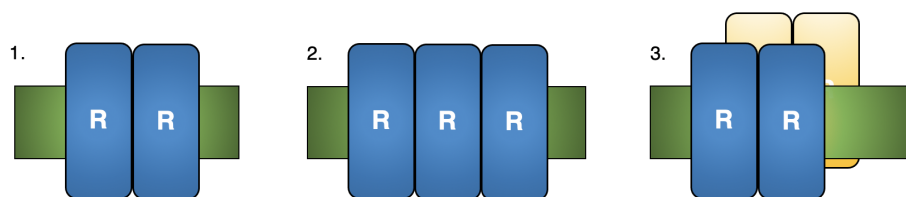


Figure 1.10: In the context of receptor association within the membrane: in (1.), when two copies of the same receptor are embedded in the membrane and interact, this is referred to as a (homo)dimer; in (2.), when three copies of the same receptor interact with each other simultaneously, this is termed a trimer; and in (3.), when different types of receptors interact with each other, forming a complex, this is known as a heterotetramer (in this case formed by two pair of homodimers).

Instead, the addition of a ligand influences the dynamics of interaction of the D2 (Figure 1.11).

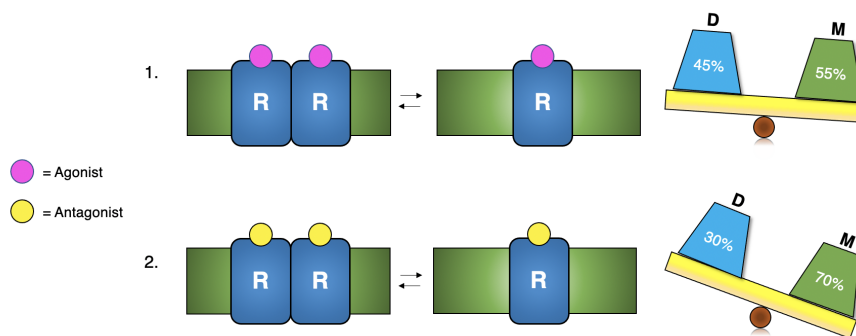


Figure 1.11: Dynamic equilibrium monomer(M)-to-dimer(D) of the D2. When the D2 is bound to an antagonist (in yellow) the equilibrium is biased towards the monomeric state; conversely, with an agonist, the equilibrium between populations is almost 50%-50%[67].

In (1., Figure 1.11), when an agonist (in magenta) is bound to the receptor, the monomer(**M**)-to-dimer(**D**) population is slightly skewed towards to the monomeric state (55% over 45%); in (2., Figure 1.11), when an antagonist (in yellow) is bound to the D2, the **M** population is favoured over the **D** population (70% over 30%)[61, 67]. This dynamic equilibrium between populations highlights the concept of *transient dimerization*[36, 67, 69]. The lifespan of the D2 transient dimer determined by Tabor and colleagues is approximately 0.5 s at a temperature of 24°C[61, 67]; in contrast, the lifespan estimated by Kasai *et al.* is about 70 ms at a temperature of 37°C[36], approximately 7.5 times less. This last approximation is more accurate as the experiments were conducted at body temperature[36]. Therefore it is crucial to correctly determine the lifespan of the dimer in order to identify potential drugs that are most suitable and selective for that dimer. Moreover, the addition of an agonist but *not* of a neutral antagonist stabilizes even more the D2 transient dimer, from about 70 ms to about 0.1 s. This impact of the agonist reflects the tendency observed in (1., Figure 1.11), and it is also corroborates the key role of transient dimerization in pharmacology[36]. This suggests that transient homodimerization appears to be a general, temperature- and ligand-sensitive feature of class A GPCRs with probable implication for receptor signalling and regulation[36].

Nonetheless, D2 can be further found as constitutive homodimer, trimer, tetramer, and pentamer within the membrane[36, 55, 70].

Besides, other GPCRs like A2 homodimerization has been well reported and documented[71, 72]. To investigate the A2 homodimerization, BRET and FRET techniques have been utilized in cells transfected with a variety of receptor fusion proteins tagged to GFP2, YFP, or Renilla luciferase (Rluc)[71]. Receptor functionality has been further confirmed by measuring the cAMP levels downstream in response to the A2 specific agonist CGS21680, which induced similar responses in all constructs[71]. Moreover, this ligand stimulation did not cause any significant changes in the BRET₅₀ signal[71]. This observation is consistent with findings by Kamiya and colleagues, suggesting that receptor activation *does not* influence its capacity of oligomerization, and this, in turn, implies that A2 homodimers are likely constitutively preformed within the cellular environment[73].

Additionally, an experimental study using immunofluorescence revealed that the intrinsically disordered region (IDR) of the A2 C-terminus plays a relevant role in protein aggregation and facilitates protein-protein homodimer interactions[72]. Firstly, residue Cys394 within the C-terminal region of A2 has been showed to reduce A2 oligomer levels when mutated[72, 74]. Indeed, the replacement of Cys with different hydrophobic amino acids (Ala, Leu, Met, and Val) resulted in a reduction of the dimer population[72]. This indicates that Cys394 plays a critical role in A2 homodimerization through disulfide bond formation. Secondly, Size Exclusion Chromatography (SEC) experiments conducted on A2 truncated at Gln372 (A2-Gln372ΔC) suggest that while Cys394 significantly contributes to dimer stabilization via disulfide bonds, there are also other interactions, both disulfide bonds and others, that play a role in maintaining the architecture of these dimers[72]. This multifaceted stabilization mechanism suggests a complex interplay of various molecular interactions driving A2 dimer/oligomerization[72].

Moreover, to investigate the role of the C-terminus in A2 oligomerization and dimerization, mutagenesis analysis was performed by truncating eight specific sites within this region[72]. SEC analysis revealed that as the C-terminus was progressively shortened, there was a notable decrease in the populations of dimers and oligomers[72]. This reduction was especially pronounced when the hydrophilic patch containing residues P354 to N359 was chopped off, leading to a nearly complete absence of oligomerization when the entire C-terminal IDR was truncated. These findings exhibit and emphasize even more the critical role of the C-terminal region in A2 oligomerization, highlighting a plethora of interaction hotspot sites along its length that contribute to manage this process[72].

1.8 Heterodimerization of the D2 and A2 receptors

GPCRs can form both homomers and heteromers within the plasma membrane[75]. When GPCRs are activated by the action of an agonist, conformational changes occur, impacting on their pharmacological and signalling fingerprint[75]. The activation pathway of class A GPCRs has been extensively discussed in this paper and will not be further addressed in this thesis[76]. These allosteric modifications gained via PPI serve as a primary mechanism for receptor regulation[75]. Additionally, heteromeric complexes are naturally distributed in a scattered manner within the plasma membrane and often aggregate into clusters, consistent with the fluid mosaic model theory proposed in the 1970s[77, 78]. The interactions between these heteromeric complexes within the clusters further amplify the regulatory signal[75, 79, 80, 81].

More generally, a heteromer can be considered as such if it satisfies one or more of the following criteria[80, 81]:

- **The single protein entities interacting within the heteromer framework should coexist in the same location** (i.e. in proximity) **and interact physically**;
- **The pharmacology properties of the new formed heteromeric complexes must be unique from those of the single protomers**, thereby amplifying their biochemical spectrum[75, 79, 80, 81];
- **Disrupting the heteromer should result in the loss of properties peculiar to that heteromer**: this would invalidate the aforementioned last two criteria[81].

Additionally, another very important aspect to consider is the presence of lipid rafts within the plasma membrane[82, 83]. These lipid rafts are microdomains particularly enriched in sphingomyelins, phospholipids, glycopospholipids, and cholesterol, and they act as active hotspots for GPCRs[82, 84]. Some GPCRs may associate at level of these microdomains in which lipids appear more densely packed[85], while others may utilize preferential pathways to associate, for example involving those with lower friction levels within the plasma membrane[79].

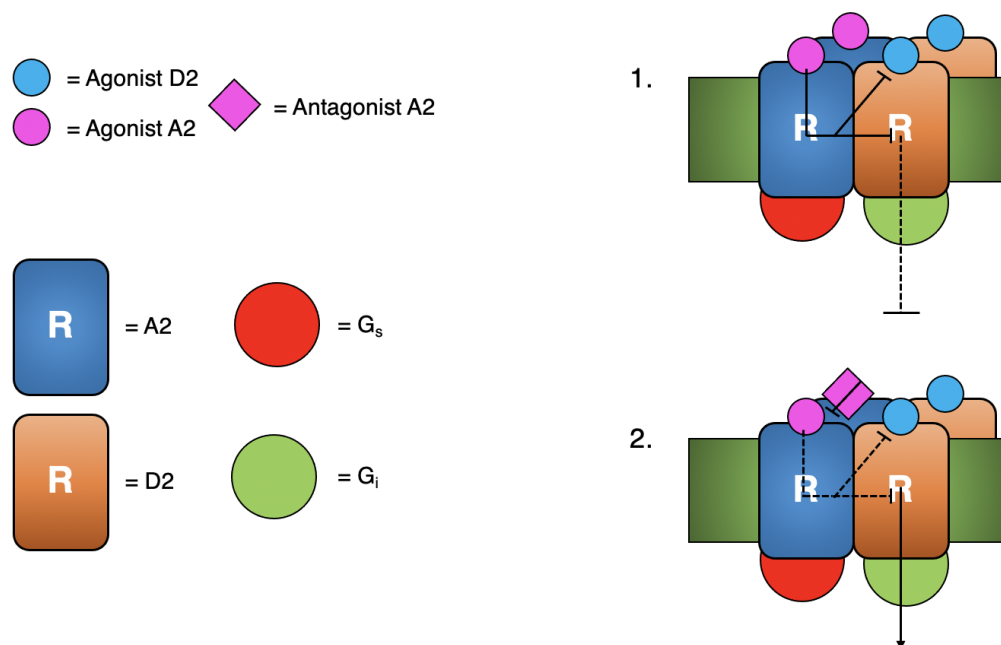


Figure 1.12: Allosterism between A2 and D2. Figure inspired by [58].

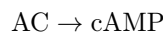
Of particular interest is the allosteric crosstalk between the A2 and D2 receptors[58, 59, 79]. When an agonist of the A2 receptor binds to the receptor, the affinity and efficacy of agonists at the D2 receptor and the intrinsic efficacy of D2 receptor antagonists decrease[59]. This same allosteric modulation also occurs when an antagonist of the A2 receptor binds to the A2 receptor while the D2 receptor is bound to a D2 antagonist (1., Figure 1.12). However, the *co-administration* of an A2 agonist and an A2 antagonist blocks this modulation. This concurrent occupancy prevents the conformational change in the A2-D2 receptor complex, thereby restoring the functionality of the D2 receptor, which can then transmit signals intracellularly (2., Figure 1.12). This apparent paradoxical mechanism can only be explained if the cooperativity between A2 and D2 receptors involves four entities, namely a heterotetramer composed of A2 and D2 homodimers[58, 59]. This framework is compatible as the binding of an agonist (or antagonist) to the A2 receptor induces a conformational change in the A2 homodimer that would not be achievable with the concurrent administration of both an A2 agonist **and** an A2 antagonist (Figure 1.12)[59]. This scenario has been confirmed by BRET and Bimolecular fluorescence and bioluminescence complementation assays[59]. This modulation induced by any A2 ligand requires the heteromer to maintain its proper quaternary conformation. Evidence from experiments in transfected mammalian cells and striatal tissue indicates that disrupting the heteromer via the use of

peptides or mutations abolishes the cooperative interaction and equilibrium between A2 and D2 in the heterotetramer complex[59]. This suggests that the correct electrostatic allosteric interaction among protomers must be identified and preserved. Furthermore, it has been postulated that the allosterism between the two receptors in the complex can shed light on both the depressant and stimulating effects induced by A2 agonists and A2 antagonists (such as the well-known caffeine), respectively[59]. This biochemical profile appears to be the structural determinant for a wide spectrum of NDs, including PD, and it also sustains the administration of both selective and potent A2 antagonists in treating such severe neurological conditions involved in the D2 pathway[59]. Also, the problem is determining the architecture of the quaternary complex by ensuring proper interactions to avoid steric hindrance between G-proteins bound to the complex - G_i with D2 and G_s with A2[86]. This issue will be discussed in Chapter 5. **For this reason, the protocol outlined below poses particular emphasis on determining the structure of the complex, a key factor underpinning several fundamental biological processes**, like the one just shown here. Eventually, new allosteric sites may emerge during the formation of the heterotetramer, which could represent further putative targets for ligands and other (small) molecules[80, 87]. These sites could potentially be exploited to modulate the complex for further therapeutic purposes[80, 87]. Collectively, these evidences expand the concept of GPCRs' plasticity.

1.9 The G_s and G_i : canonical and allosteric interactions

Not only does the A2-D2 heterotetramer meet all the criteria listed above[80, 81] but, as previously mentioned in section 1.8, the allosteric antagonism at the A2-D2 heterotetramer level can be harnessed to uncover therapeutic targets in ways not previously attainable[58]. Specifically, high dosage of A2 antagonists can be used to treat NDs that particularly involve the D2[58, 59]. Moreover, to further corroborate the allosteric cooperativity between these two receptors, it has also been demonstrated in mammalian transfected cells and striatum cells that D2 agonists are capable of blocking A2 agonist-mediated activation of AC[58, 80, 86, 88]. But how do A2 and D2 interact with each other? Are they still involved in their respective cellular pathways?

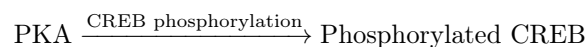
An answer to these questions starts with the activation of A2 via adenosine binding, which triggers a signaling cascade that commences with cyclic AMP (cAMP), a second messenger generated by adenylyl cyclase (AC) through the G_α subunit of the G_s protein.:



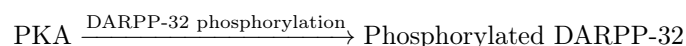
The cAMP activates the protein kinase A (PKA), a serine-threonine cAMP-dependent protein kinase[89]:



PKA activation results in phosphorylation of cAMP-response element binding protein (CREB), enhancing gene transcription:



Additionally, PKA phosphorylates 'dopamine- and cAMP-regulated phosphoprotein (DARPP-32), amplifying the signaling cascade:

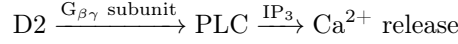


A2-mediated PKA activation also phosphorylates voltage-dependent Ca^{2+} channels (VDCC), NMDA, and AMPA receptors:

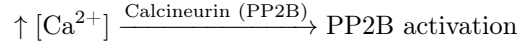


These phosphorylation events regulate the activity of these proteins, influencing in turn the excitability of striatopallidal neurons[86]. This modulation is especially sensitive in pathological conditions, emphasizing the significant role of A2 in the neurons[86].

Normally D2 signals through phospholipase C (PLC) employing a mechanism dependent on the $G_{\beta\gamma}$ subunit[29, 86]. PLC cleaves the phospholipid phosphatidylinositol 4,5-bisphosphate (PIP2) into diacyl glycerol (DAG) and inositol (1,4,5)-triphosphate (IP_3)[86, 90]. IP_3 induces the release of Ca^{2+} in the cytosol:



The increased $[Ca^{2+}]$ activates Ca^{2+} /calmodulin-dependent protein phosphatase calcineurin (PP2B):



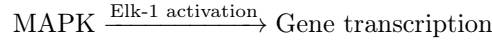
When phosphorylated, VDCC, NMDA, and AMPA receptors, along with DARPP-32, become primary targets of PP2B. These results reduce neuronal excitability, shedding light on the downstream mechanism of D2-mediated inhibition of A2 signaling[86]:



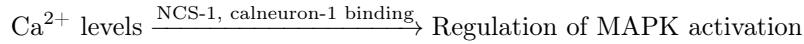
A2 and D2 activation can also finely regulate gene expression through G protein-dependent and -independent Mitogen-Activated Protein Kinase (MAPK) activation:



MAPK activation transmits signals from synapse to nucleus by activating the transcription factor Elk-1:



Intracellular Ca^{2+} levels modulate the simultaneous activation of A2 and D2 on MAPK by influencing the binding of the neuronal Ca^{2+} -binding proteins NCS-1 and calneuron-1:



NCS-1 and calneuron-1 associated with the A2-D2 heteromer under low and high Ca^{2+} concentrations, respectively, alter A2 ligand allosteric modulation of D2-ligand-mediated MAPK activation[86]. This highlights a orchestrated mechanism whereby simultaneous activation of A2 and D2 in the heterotetramer can either potentiate or attenuate MAPK signaling, depending on varying Ca^{2+} conditions[86].

Therefore, A2 and D2 receptors interact within striatopallidal neurons through both allosteric and canonical mechanisms (Figure 1.13)[86].

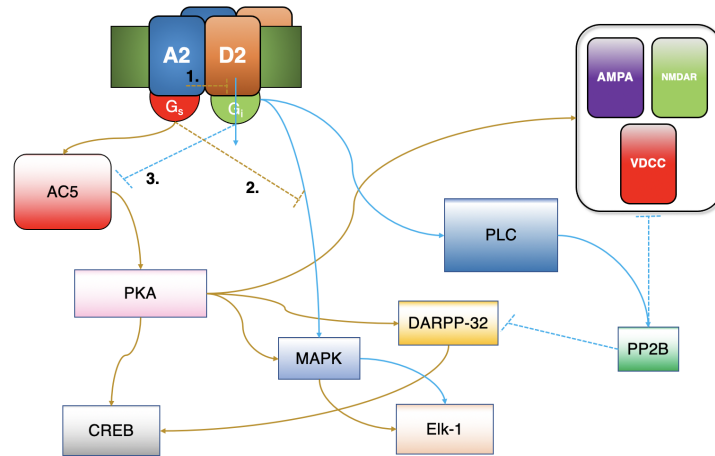


Figure 1.13: Allosteric and canonical pathway of the A2-D2 heterotetramer. The A2-D2 heterotetramer quaternary structure facilitates complex interactions between A2 and D2 exogenous and endogenous ligands. Any A2 ligands is capable of attenuating of the G protein-dependent (1., dashed line) and independent (2., dashed line) signaling by endogenous dopamine or other D2 ligands via allosteric interactions. Vice versa, D2 agonists can suppress A2-mediated activation of AC5 via antagonistic G_s - G_i interactions at the AC5 level (3., dashed line). A2 activates AC5 and in turn PKA, leading to phosphorylation of DARPP-32, which enhances striatopallidal neuron excitability and supports PPI (not shown in the figure). Instead, D2 activates both PLC and PP2B, which dephosphorylates PKA substrates, reducing neuron excitability and fostering psychomotor responses (not shown in the figure). The allosteric cooperativity induced by both receptors modulates gene expression through MAPK and the Elk-1 transcription factor activation[86]. Cyan lines refer to the D2 pathway; conversely, ivory lines indicate the A2 pathway.

Initially, it was proposed that A2 subunits either form complexes with D2 receptors, thereby influencing signaling, or function independently to regulate cAMP levels[86]. However, more recent theories suggest that A2-D2 heterotetramers are capable of supporting both modes of interaction simultaneously. This statement is further supported by the pivotal role played by A2-D2 heterotetramer complexes, which consist of interacting A2 and D2 homodimers[59, 91]. However, not all interfaces within the heterotetramer can accommodate the simultaneous presence of two heterotrimeric G proteins. Specifically, this requires that the interfaces between the two homodimers adopt a peculiar configuration, avoiding steric hinderance that could impair the complex's stability. The arrangement that maximizes the number of associated heterotrimeric G proteins ($G_{i/o}$ for D2 and $G_{s/olf}$ for A2) involves interactions between TM6 helices at the homodimer interfaces and TM4-TM5 helices at the heterodimer interfaces, as suggested by Navarro and colleagues[92]. Also, to get a broader picture, these complexes are crucial in orchestrating signaling pathways that involve *both* $G_{s/olf}$ and $G_{i/o}$ proteins, but also an additional interaction partner - AC5⁵, in the striatum should be considered (Figure 1.14)[86, 91, 92]. Indeed, experimental evidence based on mammalian transfected cells confirms the presence of supramolecular three object complexes made of of A2/D2/AC5. Synthetic peptides with primary sequences mimicking the TM domains of A2 and D2, as well as the potential TM domains of AC5, were employed in these experiments to confirm this theory[86, 92, 93]. Therefore, this complex machinery underlines the functional significance in cellular downstream pathways in detecting the proper interactions[86, 92].

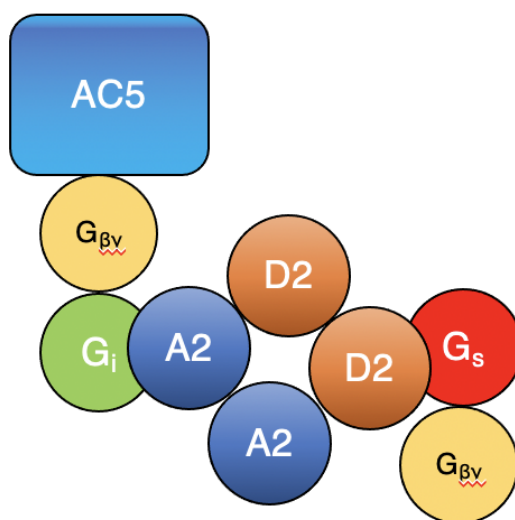


Figure 1.14: Arrangement of the A2-D2 heterotetramer in association with AC5, viewed from the extracellular side.

1.10 Residence Time

Biochemical and cellular assays measure drug–target binding affinity under thermodynamic equilibrium conditions (in closed systems)[94]. Accordingly to this statement, metrics like the half-maximal inhibitory concentration (IC_{50}), the concentration of effector yielding half-maximal response (EC_{50}), the equilibrium dissociation constant (K_d), and the inhibition constant (K_i) remain robust[94]. However, when dealing with living organisms, the processes of absorption, distribution, metabolism, and excretion (ADME), along with pharmacokinetics (PK), lead to fluctuations in the concentration of administered drugs within the bloodstream. These physiological mechanisms affect the drug's *bioavailability*, resulting in dynamic changes in its effective levels over time[94]. As a result, the previous equilibrium metrics fail to accurately capture the complex dynamics observed in *in vivo* pharmacology, thereby reducing the predictive power of these equilibrium-based measurements[94, 95].

For this reason, the lifetime, or *residence time* (RT), rather than binding affinity, more significantly influences the *in vivo* pharmacological activity[94].

Residence time is calculated as follows:

⁵AC type 5

$$\tau = \frac{1}{K_{off}},$$

where,

K_{off} is the dissociation constant.

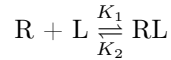
The underlying idea is straightforward: pharmacological activity remains as long as the binary drug-receptor complex (DRC) is established[94]. This concept emphasizes how crucial the DRC is in influencing both the efficacy of the ligand and the duration of their interaction within this complex[94, 96]. Several drugs in the market possess large lifetime within the body, and they remain active for an extended period of time[94, 96]. This is the case of numerous drugs targeting GPCRs, such as antagonists of M3 receptors or agonists of β_2 -adrenoceptor[96]. However, for other receptors such as GluN2B or D2, a short RT might be effective in avoiding unwanted side effects, such as toxicity (or prolongation of EPS, in the case of D2)[94, 96, 97]. Thus, the ideal strategy would aim to enhance target potency while concurrently minimizing its lifetime[94].

1.10.1 Mechanisms of Ligand-Receptor Interaction

K_{off} relates to the mechanisms involving the drug-receptor complex[39, 94, 96]. Three mechanisms (mainly taken from[39]) can be summarized as follows[39]:

1. Simple One-Step Association and Dissociation mechanism

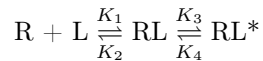
- **Process:** The receptor (R) and ligand (L) establish a binary complex (RL).
- **Rates:** Association rate k_1 (K_{on}) and dissociation rate k_2 (K_{off}).
- **Kinetics:**



- **Scenario:** Mostly encompassed in GPCRs binding kinetics[39].

2. Ligand-Receptor Induced-Fit Mechanism

- **Process:** The ligand binds to the receptor to form an initial complex (RL). The complex RL experiences a conformational change towards a higher affinity state (RL*).
- **Rates:**
 - Initial binding and dissociation: k_1 (K_{on}) and k_2 (K_{off}).
 - Induced conformational change: k_3 (forward) and k_4 (reverse).
- **Kinetics:**



- **Dissociation Rate:**

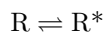
$$K_{off} = \frac{k_2 k_4}{k_2 + k_3 + k_4}$$

- **Scenario:** Typical in high-affinity interactions: enzyme inhibitors and receptor antagonists[39].

3. Conformation-Selection Mechanism

- **Process:** The receptor is expressed in two conformational states (R and R*), both in equilibrium. The ligand binds specifically to one state (R*) to form the complex R*L.
- **Rates:** Interconversion between R and R* represents the rate-limiting step to ligand binding.

- **Kinetics:**



- **Dissociation Rate:**

$$K_{\text{off}} = k_4$$

- **Scenario:** Not commonly observed in ligand-receptor interactions[39, 98].

1.10.2 Drug potency and the relationship with the residence time

Drug-target binding potency is measured by its K_d . Accordingly to the model **1.** (see Section 1.10.1), K_d is formulated as follows:

$$K_d = \frac{K_{\text{off}}}{K_{\text{on}}}. \quad (1.1)$$

For **1.** (see Section 1.10.1), K_d is directly proportional to K_{off} and it is inversely related to both residence time and K_{on} . This assumption is correct because both the dissociation and association rates are managed in one single step[94].

However, calculating only K_d is insufficient if the binding interactions require multiple steps, such as in the case of the two-step induced-fit process (see **2.**, Section 1.10.1). In other words, since this mechanism involves distinct binding events and a plethora of spacial rearrangements, it is *not* correct to assume that all compounds with a specific K_{on} will influence all four related rate constants equally (see **2.**, Section 1.10.1). In particular, a compound can bind to the same target with the same K_d values but can exhibit significantly different lifetimes due to variations in different compound structures and stereochemistry that affect the K_3 and K_4 constants[94]. Interestingly, even if the binding process follows the "Simple One-Step Association and Dissociation mechanism" (see **1.**, Section 1.10.1), structural changes in the compounds may lead to contingent changes in the K_{on} and K_{off} values.

This suggests that, for effective drug discovery, it is essential to measure both K_d and τ [94].

1.10.3 Techniques to evaluate drug lifetime and the lifetime of D2 drugs

Several experimental techniques have been utilized for calculating the drug RT, wherein we found: Direct kinetic radioligand binding assay, Indirect kinetic radioligand binding assay, Competition association assay, Dual-point competition association assay Dual-point IC_{50}/K_i value determination, Functional reversibility assay, and Label-Free Kinetic Measurements[39]. These techniques will be not discussed further in this thesis. Drug RT can be expressed also in relationship with the dissociation half-life ($t_{1/2}$)[39]. Indeed:

$$t_{1/2} = \frac{\ln 2}{k_{\text{off}}} \implies RT = \frac{t_{1/2}}{\ln 2}. \quad (1.2)$$

The factor $\ln 2$ in Equation 1.2 arises because $t_{1/2}$ represents the time at which half of the complexes have dissociated, which corresponds to one half-life in exponential decay processes[39].

Molecule	Affinity (K_i , nM)	$t_{1/2}$	References
Clozapine	137	14.5 s	[39, 99]
JNJ-37822681 ⁶	158	6.5 s	[39, 99]
Risperidone ⁷	3.3	~ 162 s	[28, 100]
Nemonapride	0.025 (in rats)	355 min (in rats)	[39, 101]
Spiperone	0.1 (in rats)	200 min (in rats)	[39, 101]

Table 1.2: Table of Molecule Affinities with their dissociation half-life $t_{1/2}$. In Red, first-generation antipsychotics drugs.

As highlighted in Table 1.2, typical antipsychotics present higher values of $t_{1/2}$, and they are therefore characterized by blocking the D2 receptors for hours. This suggests that atypical antipsychotics

function mainly due to their high K_{off} values for the D2, which cause them to dissociate faster than typical antipsychotics[39].

SGAs are known for their polypharmacological profile, i.e. their ability to interact with multiple molecular targets, which contributes to both their therapeutic and adverse effects[39, 96, 102]. This polypharmacology largely stems from the high homology that characterizes the D1/D2-like receptors with other aminergic receptors[43]. Unlike typical antipsychotics that primarily act on dopamine D2 receptors[39], SGAs generally show high binding affinity for both dopamine D2 and other off-targets, such as serotonin 5-HT_{2A} receptors[39]. Notably, despite their full occupancy of 5-HT_{2A} receptors, SGAs are associated with a lower incidence of EPS compared to typical antipsychotics - a discriminant factor contributing to their broad clinical adoption in hospital and pharmaceutical therapies. Their polypharmacological fingerprint supports the hypothesis that their atypical nature is more likely attributable to their *short residence times* when forming binary complexes with D2 receptors, rather than solely to their antagonistic action at 5-HT_{2A} receptors[39].

A corollary of this theory is that off-target side effects should be mitigated as much as possible. To achieve this for the D2 receptor, the gold standard would be to shorten the RT to avoid excessive blockade, while keeping high the affinity for the intended receptor. One last, but no less important, factor to consider is how the residence time of a given ligand may change when the receptor is in an oligomeric state. One hypothesis is that, upon oligomerization, the receptors interact and the interface between the two protomers alters the volume of the orthosteric cavity where ligands typically bind, thereby changing their residence time. The data presented in the previous table refer to the D2 receptor in its monomeric form, but, as seen in the previous sections, it may adopt supramolecular structures such as homodimers or, more in general, oligomers. **The protocol that will be presented in the following chapters is capable of producing structures that could potentially serve as a foundation for studies of this kind.**

⁶The N-[1-(3,4-difluorobenzyl)piperidin-4-yl]-6-(trifluoromethyl)pyridazin-3-amine is a promising atypical antipsychotic under clinical trials.

⁷ $t_{1/2}$ calculated considering the RT of risperidone of 233 min found by Wang *et al.* [28] and using (1.2).

Chapter 2

Methods

In this chapter, I will describe all theoretical background and methods applied in the thesis. Firstly, I will describe the tool I employed for modeling the missing parts of the structures used in the projects. Secondly, I will outline the basic principles of molecular dynamics simulation.

2.1 Computational methods to reconstruct 3D protein structures

As of 2023, UniProtKB features an expertly manually curated collection of around 570,000 reviewed UniProtKB/Swiss-Prot protein sequence entries. Additionally, it contains almost 230 million unreviewed UniProtKB/TrEMBL entries that are annotated using automated systems[103]. Through decades of extensive research, approximately 200,000 protein structures have been determined experimentally[104, 105], among which about 140,000 proteins are included in the world wide Protein Data Bank (wwPDB)[106]. However, it is not always possible to determine the 3D structure of a protein due to the inherent difficulties of the techniques used for different reasons: the protein might be too large for NMR analysis, unsuitable for cryo-EM studies, or unable to be crystallized for X-ray diffraction.

Therefore, computational tools are required to address this issue. Computational protein structure prediction methods can be differentiated into four categories: fold recognition, *ab initio*, homology modeling, and machine learning.

- (i) Fold recognition, or fold threading methods, are used when comparative modeling methods are ineffective. Unlike comparative modeling, fold recognition can identify sequences with similar protein folds even when they have low sequence similarity. They constitute a category of techniques designed to predict the three-dimensional folded structure when comparative modeling methods fail to provide dependable predictions. The basic assumption on which fold recognition methods rely upon is that the structure is more conserved than primary sequence from an evolutionary point of view[107];
- (ii) Instead, *Ab initio* methods are referred for structure prediction that do not exploit known structures experimentally determined and only rely upon the primary sequence, without the exploitation of any template[107];
- (iii) Machine learning (ML) approaches have emerged as a viable strategy for efficiently predicting three dimensional structures. Among these techniques, we found AlphaFold, a machine learning tool used to predict 3D structures with the requirement of two information as inputs: the primary sequence and aligned sequences of homologues[108, 109]. The resulting MSA is fundamental to get a proper final 3D prediction[108, 109];
- (iv) Eventually, homology modeling is a powerful tool to predict three-dimensional structures from its own primary sequence based on its template[106]. Homology modeling is considered one of the most accurate 3D protein prediction tool, despite the new breakthroughs achieved with AlphaFold[106, 109]. The rationale behind homology modelling can be found in two main categories: the three dimensional structures undergo slower evolutionary changes than the primary sequence[106, 110, 111] (Chothia and Lesk); the three dimensional structure depends on its own primary sequence[106]. The reliability of this method relies upon the choice of both the template

and the target sequence and the length of the sequence coverage. The alignment between the template and the target sequence is pivotal.

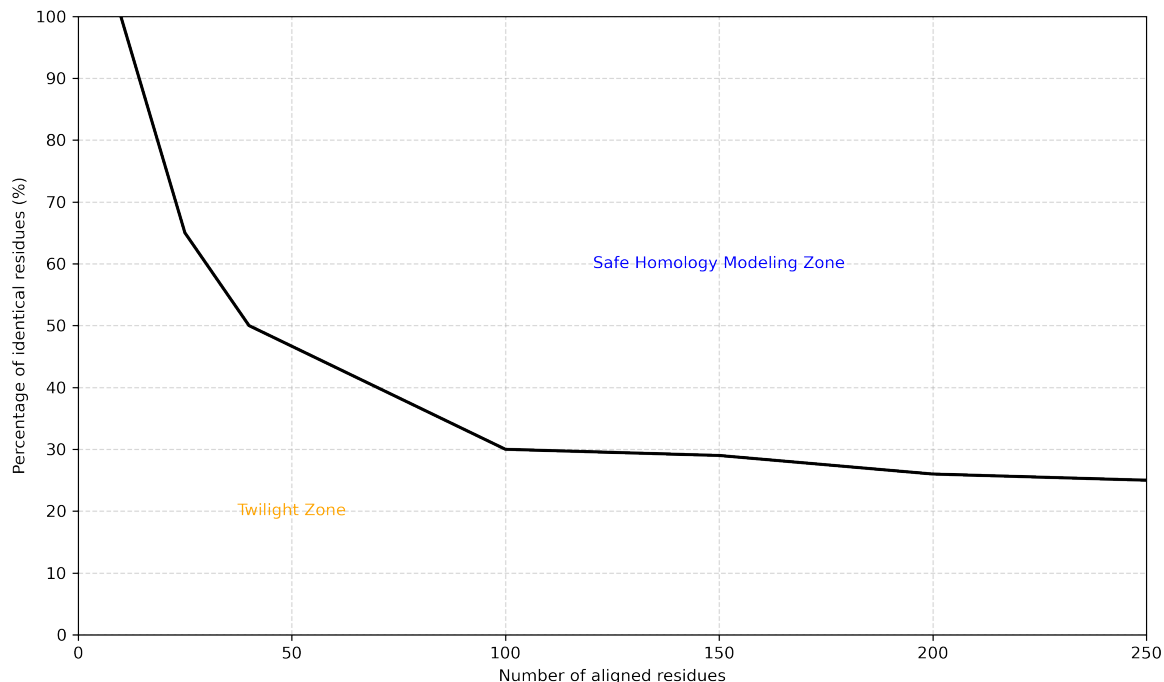


Figure 2.1: Sequence alignment zones in homology modeling.

According to this plot, only the region above the black curve (Safe Homology Modeling Zone) represents the safer region to model the target by the template.

Homology modelling procedure can be differentiated in seven steps[106]:

1. Identification of templates;
2. Sequence Alignment;
3. Model Building;
4. Loop Modeling;
5. Side Chain Modeling;
6. Modeling Optimization, and
7. Modeling Validation.

In this thesis, experimentally X-ray structures were already stored in the RCSB Protein Data Bank (PDB) database[112, 113]. Missing residues and loops were added to the final models. Additionally, chimeric endolysin and G_i/G_s G-proteins were removed from the final models via homology modeling. Thus, not all the aforementioned seven steps were followed, and only the one that have been used are explained in detail in the following subsections.

2.1.1 Model Building and Loop Modeling

To build the missing loops and missing residues of the D2, I utilized MODELLER[114, 115]. Modeller remains one of the state-of-the-art software tools for homology modeling, despite the recent introduction of AlphaFold. MODELLER functions exploiting the spatial restraints, which interatomic distances and angles in the target structure closely resemble those in the template. Therefore, having a template with high sequence identity and a valid alignment are essential for accurate modeling and getting a valuable final 3D structure. To get a proper alignment, bioinformatic programs such CLUSTALW or MUSCLE

can be utilized[116, 117]. Subsequently, structural restraints are extracted from the template and then employed to construct the model for the target structure[115]. In the initial stage, the protein backbone structure is established by satisfying spatial restraints, which originate from the optimization of probability density functions that encode three main classes of structural restraints[115]. These include spatial details, such as main chain and sidechain dihedral angles, C_α - C_α distances, and hydrogen bonds; secondly, stereochemical restraints covering angles and bond lengths; thirdly, statistical preferences for non-bonded interatomic distances and dihedral angles[115]. The combination of these probability density functions covers the spatial restraints of the entire protein structure. The preliminary 3D protein model is finally improved by optimizing this probability density function through minimization of restraint violations[118]. Each model undergoes optimization using an extensive variable target function method (VTFM), which utilizes conjugate gradients with simulated annealing to improve model quality, as detailed here[115]. This ultimate refinement ensures that the final 3D structure satisfies both stereochemical and energetic restraints derived from the template used in input.

For the D2 receptor in its inactive state (RCSB PDB ID: 6CM4[28]), four residues belonging to the intracellular loop 2 (ICL2) were added to the final model. Additionally, the C-terminus Cys in the amphiphilic H8 was added to the same structure. For the D2 in its active state (RCSB PDB ID: 7JVR[43]), the N-terminus and C-terminus residues were reconstructed. For both active and inactive structures, loop connecting TM5 and 6 (i.e. the intracellular loop 3 - ICL3) is absent and therefore it has been included by adding four residues following the Uniprot Canonical sequence[37].

2.1.2 Modeling Validation

Modeling validation can be computed in different ways. Usually, models are evaluated via Discrete optimized protein energy (DOPE) score[119]. DOPE score is implemented in MODELLER[114]. The more negative the DOPE score is, the better the resulting model structure should be. Moreover, to check the stereochemistry quality parameters of a protein model, PROCHECK program is commonly used[120, 121]. I used the combination of both DOPE score and PROCHECK for validating the final D2 structures, both in their active and inactive states.

2.2 Molecular Dynamics

Molecular dynamics (MD) simulations represent a valuable approach in the study of biological systems. They offer insight into molecular processes that are often difficult or even impossible to capture experimentally. By enabling the observation of structural dynamics over femtosecond to milliseconds timescales, MD allow researchers to investigate conformational changes, molecular interactions, and dynamic equilibria at a resolution beyond that of most experimental techniques. In addition to providing detailed mechanistic understanding, MD offers a cost-effective and time-efficient alternative to laboratory experimentation, particularly when screening multiple molecular variants or testing conditions that may be technically challenging, hazardous, or resource-intensive to replicate *in vitro*, and can also contribute to reducing the reliance on animal models in early-stage research. Furthermore, simulations serve as a powerful complement to experimental data, supporting the interpretation of experiments. In certain contexts, especially where experimental access is limited or unfeasible, MD may provide the only viable means of probing the molecular basis of biological function.

In particular, MD is a computational method that aims to numerically integrate the *equations of motion*, which dictate the time evolution of a system of classical particles. Classical mechanics allows determining positions $x(t)$ and velocities of a particle system over time, knowing particle masses and the forces acting on them. To determine the $x(t)$, is sufficient to apply the Newton's second law. For conservative systems (in which forces are conserved) and for a given initial position, $x(t)$, and initial velocities, this law can be always determined[122].

Nevertheless, molecular mechanics (as coarse-grained force field based one, see Section 2.9) is not capable of deriving properties depending on the electronic distribution, such as observing a chemical reaction. In other words, the energy of a system is directly calculated solely from the positions and velocities of atoms[123]. Instead, quantum mechanical approaches entail explicitly electrons in the calculation, thus they are capable to describe the processes like bond formation or proton transfer in a more detailed manner[123]. In quantum mechanics we are facing the same problem as in classical mechanics, i.e. the time evolution of the system, but with different actors with different laws. Here, in fact, a particle is described by the so-called wavefunction $\psi(x, t)$, which evolves in time according to the Schrödinger equation[122]:

$$i\hbar \frac{\partial \psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m} \cdot \frac{\partial^2 \psi(x,t)}{\partial x^2} + V(x,t)\psi(x,t), \quad (2.1)$$

where,

$$i \text{ is } \sqrt{-1},$$

$$\hbar = \frac{h}{2\pi} = 1.054573 \times 10^{-34} \text{ J} \cdot \text{s}, \text{ with } h \text{ the Planck constant.}$$

Nevertheless, solving equation 2.1 is challenging, if not impossible, when the system consists of thousands of molecules containing a large number of neutrons, protons (both in the nucleus), and electrons. Indeed, these particles would dramatically increase the number of degrees of freedom of the system, and the analytical solution of the Schrödinger equation would require computational power that increases exponentially with the degrees of freedom of the system. This interaction is described by the potential term V . The problem is exceedingly complex, and solving the Schrödinger equation for such systems exactly, even with the help of current supercomputers, is an unfeasible task. Therefore, one solution is dictated by the Born-Oppenheimer (BO) approximation, that simplifies the resolution of Schrödinger's equation[124]. BO approximation is based on two factors:

- **Difference in mass between electrons and nuclei:** electrons are 1836 times lighter (in term of mass m) than protons, and therefore the nuclei move slower than electrons;
- **Different motion between electrons and nuclei:** since the m of the electrons is much smaller than the one of the nuclei, in most of the cases we can assume that the motion of the electrons and the motion of the nuclei can be separated even if coupled.

Therefore, the Schrödinger equation can be split into two parts, i.e. one describing the motion of electrons with fixed position of the nuclei (electronic equation) and one describing the motion of nuclei (nuclear equation). Consequently, the solutions of the electronic equation depend on the position of the nuclei and provide the potential energy surfaces in which nuclei move in[122]. Henceforth, these surfaces are in turn used in the nuclear equation to describe the motion of the nuclei[122, 124]. This simplifies the calculation of the motion of molecules within the system[122]. To describe electrons and their evolution, one would need to employ quantum mechanics or hybrid methods[125]. In the present work, I do not consider the behavior of electrons explicitly. Instead, I focus solely on the motion of the nuclei. To describe their dynamics, we solve Newton's equations of motion, which provide a classical framework suitable for capturing the trajectories of atomic nuclei in the systems under investigation.

2.3 General expression of a force field

The basic structure of potential energy $V(x_1, x_2, x_3, \dots, x_N)$ for modeling molecular systems includes the combination of intramolecular interactions for atoms bonded together and non-bonded interactions accounting for long-range/short-range forces like electrostatic and van der Waals (vdW) forces. This is commonly referred to as a *force field*. The functional form and the parametrization of these terms are contingent upon the specific force field employed; however, a generic expression for the total energy can be represented as follows:

$$V_{total} = V_{bonded} + V_{nonbonded} \quad (2.2)$$

V_{bonded} component can be further breakdown as follows:

$$V_{bonded} = V_{angles} + V_{dihedral} + V_{improper\ dihedral} + V_{bond}. \quad (2.3)$$

More in detail we have:

$$V_{angles} = \sum_i^{angles} = \frac{f_i}{2} (\psi_i - \psi_i^0)^2, \quad (2.4)$$

$$V_{dihedral} = \sum_i^{dihedral} = \frac{v_i}{2} [1 + \cos(\theta_i - \theta_i^0)], \quad (2.5)$$

$$V_{\text{improper dihedral}} = \sum_i^{\text{improper dihedral}} = \kappa_i (\epsilon_i - \epsilon_i^0)^2, \quad (2.6)$$

$$V_{\text{bond}} = \sum_i^{\text{bond}} = \frac{\kappa_i}{2} (l_i - l_i^0)^2. \quad (2.7)$$

The constants f_i , κ_i , and v_i are determined experimentally and/or by quantum mechanical calculations. The variables l_i , ψ_i , ϵ_i , and θ_i represent the distance, angle, improper dihedral, and proper dihedral, respectively. l_i^0 is the equilibrium value for the bond length, ψ_i^0 , θ_i^0 , and ϵ_i^0 are equilibrium values for valence between angles[104].

Non-bonded interactions include short-range attractive and repulsive dispersion (i.e. Lennard-Jones potential) and long-range interactions (Coulomb potential):

$$V_{\text{nonbonded}} = \sum_i^N \sum_{j \neq i}^{N-1} \{ 4\epsilon_{ij} [(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6] + \frac{q_i q_j}{r_{ij}} \} \quad (2.8)$$

where,

- q_i, q_j are the magnitudes of the i -th and j -th point charges,
- ϵ_{ij} is the depth of the potential well,
- σ_{ij} is the finite distance at which the inter-particle potential is zero,
- r_{ij} is the distance between the point charges.

- **Short-range interactions:** Van der Waals interactions are responsible for interactions between atoms at short-range distances and they are computed by using the Lennard-Jones (LJ) potential;
- **Long-range interactions:** attraction or repulsion of charged atoms are represented by the Coulomb potential.

2.4 Statistical mechanics

A classical mechanical system is formed by set of particles. Each particle is any entity that can be characterized by a position ($\mathbf{q}$) and a momentum ($\mathbf{p}$) within a three-dimensional Euclidean space. More in general, in a system composed of N particles, the position and momentum of each particle are specified. All the positions ($\mathbf{q}$) and momenta ($\mathbf{p}$) constitutes a set of $6N$ real variables, referred as the microstate of the system. This microstate can be represented as a point in a $6N$ -dimensional space known as phase space. A statistical ensemble refers to a set of distinct microstates that collectively describe a system in thermodynamic equilibrium. In statistical mechanics, a physical (macroscopic) quantity O over all the microstates compatible with the corresponding statistical ensemble can be defined as follows:

$$\langle O \rangle_{\text{ensemble}} = \int \int O(q_N, p_N) \rho(q_N, p_N) dq_N dp_N, \quad (2.9)$$

where $\rho(q_N, p_N)$ represents the probability density function of that statistical ensemble. The evolution of the system from one microstate to another describes the *trajectory* in the phase space.

2.4.1 Ergodic Hypothesis

MD does not allow for the direct extraction of $\langle O \rangle_{\text{ensemble}}$, which is the quantity measured in experiments. Instead, MD provides access to the values of O over time. Ergodicity is the property of a dynamical system that allow connecting $\langle O \rangle_{\text{ensemble}}$ with the time average of O ($\langle O \rangle_{\text{time}}$)[126]. Theorists are only able to prove the ergodicity of dynamical systems when they are extremely simple. Nevertheless, for particle systems relevant to biology, such as the ones discussed in this thesis, the complexity makes it impossible to mathematically demonstrate their ergodicity. However, there is a general consensus that highly complex systems tend to be ergodic. This is known as the Ergodic Hypothesis, and it can be formulated as follows:

$$\langle O \rangle_{ensemble} = \langle O \rangle_{time} = \frac{1}{\Delta t} \int_{t_0}^{t_0 + \Delta t} O(q(t), p(t)) dt, \quad (2.10)$$

where $O(q(t), p(t))$ represents the observable as a function of the coordinates $q(t)$ and momenta $p(t)$ over time[126].

2.4.2 Equations of motion

In the Hamiltonian formalism the motion equations read:

$$\frac{d\mathbf{q}_i}{dt} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}_i}; \quad \frac{d\mathbf{p}_i}{dt} = -\frac{\partial \mathcal{H}}{\partial \mathbf{q}_i}. \quad (2.11)$$

where $i = 1, \dots, N$ the i -th particle. The Hamiltonian $\mathcal{H}$ represents the sum of potential energy V and the kinetic T of a system. For a finite system of atoms, $\mathcal{H}$ can be written as:

$$\mathcal{H}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n), \quad (2.12)$$

where,

m_i is the mass of i -th atom,

p_i is the momentum,

$V(q_1, q_2, \dots, q_N)$ is the the system potential energy.

Combining the expressions for momentum and coordinates $q = (q_1, q_2, \dots, q_n)$ from (2.12) yields these two following two equations:

$$\dot{\mathbf{q}}_i(t) = \frac{p(t)}{m_i} \equiv v_i(t) \implies \ddot{q}_i(t) = \frac{\dot{p}_i(t)}{m_i} \quad (2.13)$$

$$\dot{p}_i(t) = -\frac{\partial V(q(t))}{\partial q_i}. \quad (2.14)$$

$\dot{q}_i(t)$ and $\ddot{q}_i(t)$ specify the first and the second derivative of $q_i(t)$, respectively. Rewriting the above equation, we have:

$$\vec{F}_i = m_i \ddot{q}_i(t) \equiv -\vec{\nabla} V(q) = \left(\frac{\partial V}{\partial x_i}, \frac{\partial V}{\partial y_i}, \frac{\partial V}{\partial z_i} \right), \quad (2.15)$$

where $\vec{F}_i$, the force acting on the i -th atom in the system, is proportional to the mass m_i of the particle and the second derivative of its position with respect to time, $\ddot{r}_i$. This corresponds to the change in its momentum in accordance with the Newton's second law.

2.4.3 Numerical integration of Newton Equation

Except for very small systems or under specific conditions, Equation (2.15) cannot be solved numerically, as systems, including coarse-grained systems simulated in this thesis, comprise more than $150-200 \times 10^3$ particles, among lipids, protein, solvent, and counterions. For this reason, in MD, velocities and positions of particles are discretized and calculated at each predefined time step, denoted as Δt . The typical approach to intertwining positions and velocities at a specific moment in time t with their previous values is using an integration algorithm. One of the algorithm mostly used is the Leap-Frog algorithm[127]. It is based on a Taylor expansion:

$$v(t + \frac{\Delta t}{2}) = v(t) + \frac{dv}{dt} \frac{\Delta t}{2} = v(t) + \frac{1}{2} \frac{d^2 v}{dt^2} \frac{\Delta t^2}{4} + o(\Delta t^3), \quad (2.16)$$

and

$$v(t - \frac{\Delta t}{2}) = v(t) - \frac{dv}{dt} \frac{\Delta t}{2} = v(t) + \frac{1}{2} \frac{d^2 v}{dt^2} \frac{\Delta t^2}{4} - o(\Delta t^3). \quad (2.17)$$

Subtracting 2.16 and 2.17 equations, we obtain:

$$v(t + \frac{\Delta t}{2}) = v(t - \frac{\Delta t}{2}) + \frac{dv}{dt} \frac{\Delta t}{2} + o(\Delta t^3). \quad (2.18)$$

Also positions are computed at each Δt . Thereby, we can calculate an expression also for $q(t + \Delta t)$:

$$q(t + \Delta t) = q(t) + v(t + \frac{\Delta t}{2})\Delta t + o(\Delta t^3). \quad (2.19)$$

(2.18) and (2.19) represent the Leap-frog algorithm. Other MD integrators, such as Euler methods and Velocity-Verlet are implemented in GROMACS[128, 129]. I used the Leap-frog algorithm in all my coarse-grained Martini 3 simulations.

2.5 Thermostats and Barostats

Managing the temperature and evaluating the quality of generated trajectories are essential aspects of any MD simulation. Traditionally, MD simulations produce a microcanonical ensemble - also known as *NVE* - due to the conservation laws of Hamilton's equations, where the number of particles, volume, and energy remain constant. Nevertheless, this approach is risky: when lower potential energy conformations are reached over the course of a simulation, they are accompanied by a corresponding rise in kinetic energy. Because the total energy is conserved in this ensemble, the total energy must remain constant; therefore, the system temperature can increase dramatically, potentially causing box instability or implosion. This is where the thermostat comes into play, serving to maintain the system temperature by adjusting the particle velocities within the simulation environment. This is the idea introduced in the so-called canonical ensemble or *NVT*.

Specifically, if we consider the problem from a physical point of view, we can define the relationship between Kinetic Energy and momentum as:

$$V_k = \frac{p^2}{2m} \quad (2.20)$$

where,

V_k is the kinetic energy,
 p is the momentum,
 m is the mass of the object.

If we now describe the relationship between the temperature T and the Boltzmann's constant k_B , we found that:

$$\langle V_k \rangle = \frac{3}{2} k_B T, \quad (2.21)$$

where,

$\langle V_k \rangle$ is the average kinetic energy per particle,
 k_B is the Boltzmann constant,
 T is the temperature.

When we combine 2.20 and 2.21, we have that:

$$T = \frac{1}{3Nk_B} \sum_{i=1}^N \frac{p_i^2}{m_i}. \quad (2.22)$$

As mentioned before, the concept of a thermostat involves adjusting the velocities by a scaling factor λ to maintain a consistent overall temperature in the system over the course of the simulation. Therefore:

$$v_i^{new} = \lambda v_i^{old}. \quad (2.23)$$

If we consider the case of the Berendsen thermostat, we have that[130]:

$$\frac{d}{dt}T = \frac{1}{\tau}(T_0 - T). \quad (2.24)$$

In this method, the system interacts with a conceptual heat bath with temperature T_0 . In particular, at each simulation step, the velocities are adjusted based on the time derivative of the temperature, which depends on the difference between the heat bath temperature and that of the system[130]. The coupling strength is represented by the constant τ [130]. The integration of the equation 2.24 demonstrates that the system's temperature decreases exponentially to match the heat bath temperature. The scaling factor for the Berendsen thermostat is denoted as follows:

$$\lambda_{ber} = \sqrt{\left(1 - \frac{\delta t}{\tau} \left(\frac{T_0}{T(t)} - 1\right)\right)}. \quad (2.25)$$

τ for my simulations has been set to 0.1 *ps*, in accordance with the coarse-grained Martini parameters published by de Jong and colleagues[131].

One major drawback with Berendsen thermostat is that it lacks a conserved quantity, thereby is not associated with a well-defined ensemble, except in very limiting cases, very far from biological problems[132]. To compensate this issue, Bussi *et al.* defined an evolution of the Berendsen thermostat by adding a stochastic term related to velocities' rescaling to permit the proper reproduction of the *NVT* ensemble[132]. In this thesis, I used the Berendsen thermostat enhanced with the *v-rescale* option implemented in GROMACS[133].

The *NVT* ensemble aims to keep the temperature of the system constant over time and equilibrate the surrounding solvent, but also the pressure of the system should be kept constant over time. To maintain the pressure fluctuations, the *V* in the canonical ensemble is replaced by constant pressure *P*, yielding to the so-called isothermal-isobaric ensemble or *NPT*. For my simulations, I exploited the Parrinello-Rahman barostat[134]. Parrinello-Rahman barostat represents an extension of the Andersen Barostat[135]. Briefly, Andersen's barostat maintains the system's pressure by allowing its volume to change dynamically over the course of the simulation[135]. To achieve this, Andersen introduced a method in which the system is connected to an external volume, symbolized by *V*, which mimics a piston affecting the system's volume[135]. Instead of directly using the original coordinates r_i , Andersen adjusted them to scaled coordinates ρ in constant pressure MD as follows:

$$\rho_i = \frac{q_i}{\sqrt[3]{V}}. \quad (2.26)$$

Andersen introduced then a Lagrangian $\mathcal{L}$, i.e. the difference between the potential energy and the kinetic energy for the scaled systems:

$$\mathcal{L}(\rho^N, \dot{\rho}^N, V, \dot{V}) = - \sum_{i=1}^N \sum_{j=1, j \neq i}^N U(\sqrt[3]{V} \rho_{ij}^2) + \frac{1}{2} M \dot{V}^2 - p_0 V + \frac{1}{2} \sqrt[3]{V}^2 \sum_{i=1}^N m_i \rho_i^2. \quad (2.27)$$

The initial terms of the Lagrangian depict the potential energy and the kinetic energy, respectively of the unscaled system; *V* refers to the system's volume. The dynamics of the system's volume *V* are outlined in the subsequent term, while the potential energy associated with *V* is articulated in the final term. Pressure regulation is illustrated by envisioning the motion of a piston with mass *M*, exerting compression on both the system and its volume *V*[135]. For the sake of clarity, in Andersen's method, the volume changes dynamically, while simultaneously scaling the coordinates, preserving the box's isotropic shape[135].

Contrarily, in Parrinello-Rahman barostat, the *size of the box* changes dynamically[134]. If *x*, *y*, and *z* are the vector of the cell having any shape, its volume *V* can be calculated as follows:

$$V = \det(h) = x \cdot (y \times z), \quad (2.28)$$

where,

h is the 3×3 matrix determined by the vectors *x*, *y*, and *z*.

Given a column vector s_i and its components ξ_i , η_i , and ϕ_i , the particle position q_i can be defined as follows:

$$q_i = h s_i = \xi_i + \eta_i + \phi_i. \quad (2.29)$$

The squared distance between particles i and j can be determined by the expression considering $0 \leq \xi_i, \eta_i, \phi_i \leq 1$, and employing the metric tensor G , such as:

$$q_{ij}^2 = s_{ij}^T G s_{ij}, \quad (2.30)$$

where,

$$G \text{ is } h^T h.$$

The aforementioned formulae can be utilized to describe a new Lagrangian $\mathcal{L}$ that governs how the shape and size of the MD simulation cell evolve over time[134]. Therefore, the Lagrangian $\mathcal{L}$ can be defined and summarized as follows:

$$\mathcal{L}(q, \dot{s}, \dot{h}) = - \sum_{i=1}^N \sum_{j=1, j \neq i}^N U(q_{ij}) + \frac{1}{2} m_i \dot{s}_i^T G \dot{s}_i + \frac{1}{2} M \text{Tr}(\dot{h}^T \dot{h}) - p_0 V. \quad (2.31)$$

In my systems, I used the Parrinello-Rahman barostat with a semiisotropic coupling: the x and y coordinates are kept isotropic, instead the z coordinate is decoupled from the others. This is the typical convention adopted for membrane systems[133].

2.6 Periodic Boundary Conditions

Periodic Boundary Conditions (PBCs) envelop the simulation box by replicating it in every conceivable direction, forming therefore an infinite lattice (Figure 2.2)[126]. The underlying principle of PBC is that when an atom exits in the box, a periodic image enters from the opposite side, thus preserving a constant number of atoms in the simulation box (Figure 2.2). PBC enables the examination of the bulk properties of the system. It is worth noting that atoms within the box can interact with those positioned in the periodic image, potentially leading to unwanted artifacts. However, if the system is large enough, like in my simulations, its thermodynamic equilibrium is preserved[136].

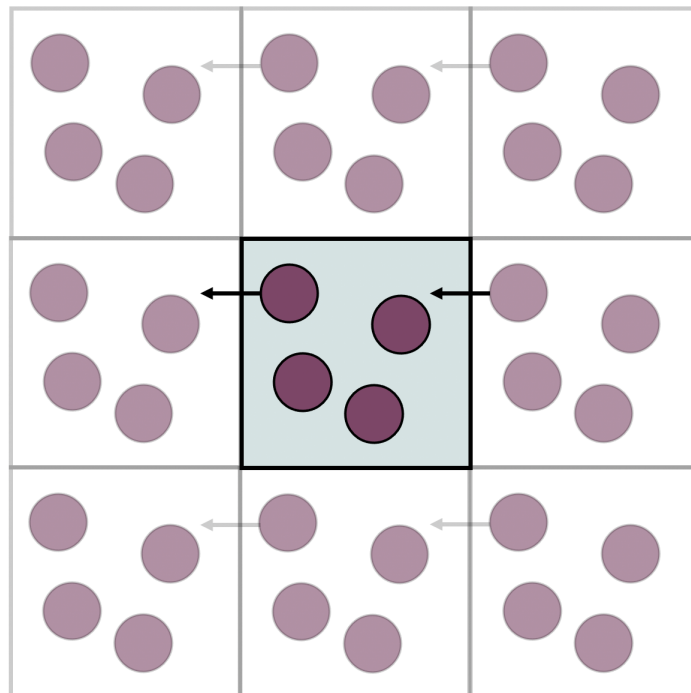


Figure 2.2: Protein Boundary Condition (PBC) definition. PBCs mimic an infinite system using a finite simulation box. Instead of simulating a massive number of particles, one can simulate in a smaller box that is conceptually repeated infinitely in all directions.

2.7 LINCS Algorithm

The LINear Constraint Solver (LINCS) is an algorithm designed to restore bond lengths after an unconstrained update[137]. It works in just two steps and it avoids costly (and useless) matrix-matrix multiplications, although imatrix operations are exploited. This makes it both faster and more stable than the SHAKE algorithm[137, 138]. Due to its stability, it is particularly well-suited for Brownian dynamics simulations[137]. The algorithm relies upon two adjustable parameters. In my simulations, default options provided by GROMACS are kept. LINCS is limited to handling bond constraints and certain isolated angle constraints[137].

2.8 Enhanced Sampling Techniques

MD is crucial for studying conformational changes in molecules, dimerization, protein folding, and/or ligand binding events[139]. However, one of the main difficulties when dealing with biomolecular systems or large scale sampling stems from the complexity of their energy landscapes, which becomes increasingly problematic as the system size consists of hundreds or thousands of atoms. The energetic profile is therefore characterized by numerous local minima separated by energy barriers higher than $k_B T$. Classical MD simulations often struggle to accurately depict various biological events, such as significant conformational changes or receptor dimerization within complex membrane environments like the ones presented in this thesis, due to the considerable time required for these processes[140]. To tackle these challenges, enhanced sampling techniques have been developed. Here, in the subsection 2.8.1, I will discuss the enhance technique employed in this thesis: the Well-tempered Metadynamics[141].

2.8.1 Metadynamics and Well-Tempered Metadynamics

This subsection is mainly taken from[141]. Calculating free-energy disparities is pivotal in MD simulations[141]. The first critical step is defining a set of collective variables (CVs) or reaction coordinates that offers a rough portrayal of the slow modes of the system[141]. Also, describing a way to compute the corresponding free-energy surface (FES) when performing MD is relevant. However, generating such a surface of these rare events is very often impeded by high energetic barriers[141]. Therefore, to address this issue, enhanced sampling techniques have been developed to allow the observation of rare events that would otherwise be almost impossible to sample within the timescales accessible to plain MD simulations.

In Metadynamics, the system's progression is influenced by a potential that depends on its past, constructed by accumulating Gaussian functions along the trajectory in the defined CVs space[141]. In the first implementation of Metadynamics[142], a bias that depends on the system's history is introduced by intermittently incorporating Gaussian functions into the potential as follows:

$$V(\vec{s}, t) = \sum_{k\tau < t} W(k\tau) \exp \left(- \sum_{i=1}^d \frac{(s_i - s_i^{(0)}(k\tau))^2}{2\sigma_i^2} \right), \quad (2.32)$$

where,

$\vec{s}$ the CVs as a function of coordinates q .

The potential $V(\vec{s}, t)$ described in Equation (2.32) steers the system away from kinetic traps in the potential energy surface, directing it towards unexplored regions of the energy landscape[142]. This dissuades the system from revisiting the same areas of phase space again[139].

Nevertheless, the first arising problem with this implementation lies in determining when to interrupt the Metadynamics run. In complex systems, the free energy does not converge to the true value immediately, but tend to oscillate around the correct result[141, 143, 144]. Yet, lowering this deposition rate inevitably prolongs the time needed to fill the FES, because the system may remain stuck in a minimum, without escaping easily from it. As a consequence, from a biological perspective, continuing the run risks irreversibly pushing the system into regions that completely lack physical relevance, creating non-biological artifacts[141].

With the introduction of the Well-Tempered Metadynamics, a significant effort has been done in this direction and to achieve an estimate of the FES that converges to the exact result over time[141,

145]. Firstly, let us indicate a system having a temperature T constituted by a set of microscopic coordinates q and a potential energy $U(q)$, obtained during the evolution of a MD[141]. **The final scope is to determine the free-energy profile with respect to the set of CVs, $s(q)$.** First and foremost, the FES can be described by the following equation:

$$F(s) = -T \lim_{t \rightarrow \infty} \ln N(s, t) \quad (2.33)$$

where,

$$N(s, t) = \int_0^t \delta_{s, s(t')} dt'$$

is the histogram associated with the CVs s collected from an unbiased simulation (see subsection 2.8.2 on how I identified my CVs), with $N(s, 0) = 0$ and the first time derivative $\dot{N}(s, t) = \delta_{s, s(t)}$ [141].

By adding a bias to the MD a **history-dependent potential**, we have that:

$$V(s, t) = \Delta T \ln \left(1 + \frac{\omega N(s, t)}{\Delta T} \right), \quad (2.34)$$

where,

ω rate that possess the same dimension of an energy,

$N(s, t)$ is the histogram calculated from the biased simulation,

ΔT is the difference of the temperature.

In other words, the potential $V(s, t)$ disincentivizes the system to explore already-seen configurations in the phase space, preventing the system to be permanently kinetically trapped[139, 141]. At this point, a question arises spontaneously: how does this rate ω change during a MD simulation? The rate through which the potential $V(s, t)$ commutes is dictated by:

$$\dot{V}(s, t) = \frac{\omega \Delta T \delta_{s, s(t)}}{\Delta T + \omega N(s, t)} = \omega \exp\left(-\frac{V(s, t)}{\Delta T}\right) \delta_{s, s(t)}. \quad (2.35)$$

The height of the Gaussian is scaled-up in this way:

$$w = \omega \exp\left(-\frac{V(s, t)}{\Delta T}\right) \tau_G, \quad (2.36)$$

where,

τ_G represents the time interval of the Gaussian's deposition.

Two corollaries can be deduced from the previous equations. Firstly, as $N(s, t)$ grows linearly over the course of the simulation, the $\dot{V}(s, t)$ decays to 0 with a $\sim \frac{1}{t}$. Secondly, V is not uniform in the s space, as the rate is inversely proportional to the time spent in that s [141]. Therefore, if we conduct simulations over a sufficient period of time to explore the entire free-energy landscape, the potential $V(s, t)$ undergoes minimal change. In other words, when this exploration occurs, the converge of a simulation is reached. Assuming that the microscopic coordinates q reach equilibrium, the probability distribution function $P(s, t)$ can be described as follows:

$$P(s, t) ds \sim \exp\left(\frac{-F(s) - V(s, t)}{T}\right) ds \quad (2.37)$$

and

$$\begin{aligned} \dot{V}(s, t) &= \omega \exp\left(-\frac{V(s, t)}{\Delta T}\right) \times P(s, t) = \\ &= \omega \exp\left(-\frac{V(s, t)}{\Delta T}\right) \times \frac{\exp\left(-\frac{[F(s) - V(s, t)]}{T}\right)}{\int ds \exp\left(-\frac{[F(s) - V(s, t)]}{T}\right)}. \end{aligned} \quad (2.38)$$

If we consider infinite time ($t \rightarrow \infty$), we will have:

$$V(s, t \rightarrow \infty) = -\frac{\Delta T}{\Delta T + T} F(s). \quad (2.39)$$

For $t \rightarrow \infty$, we can redefine the probability distribution $P(s, t \rightarrow \infty)$ in:

$$P(s, t \rightarrow \infty) ds \sim \exp \left[-\frac{F(s)}{T + \Delta T} \right] ds \quad (2.40)$$

By using equation (2.34) we can eventually estimate the free-energy surface:

$$\tilde{F}(s, t) = -\frac{T + \Delta T}{\Delta T} V(s, t) = -(T + \Delta T) \ln \left(1 + \frac{\omega N(s, t)}{\Delta T} \right). \quad (2.41)$$

Well-tempered Metadynamics effectively addresses the convergence issues found in traditional Metadynamics, exploring solely physically relevant regions of the conformational space, as long as the selected CVs fall within the set of **optimal** CVs[141, 142]. Figure 2.3 depicts the Well-tempered Metadynamics technique in four stages:

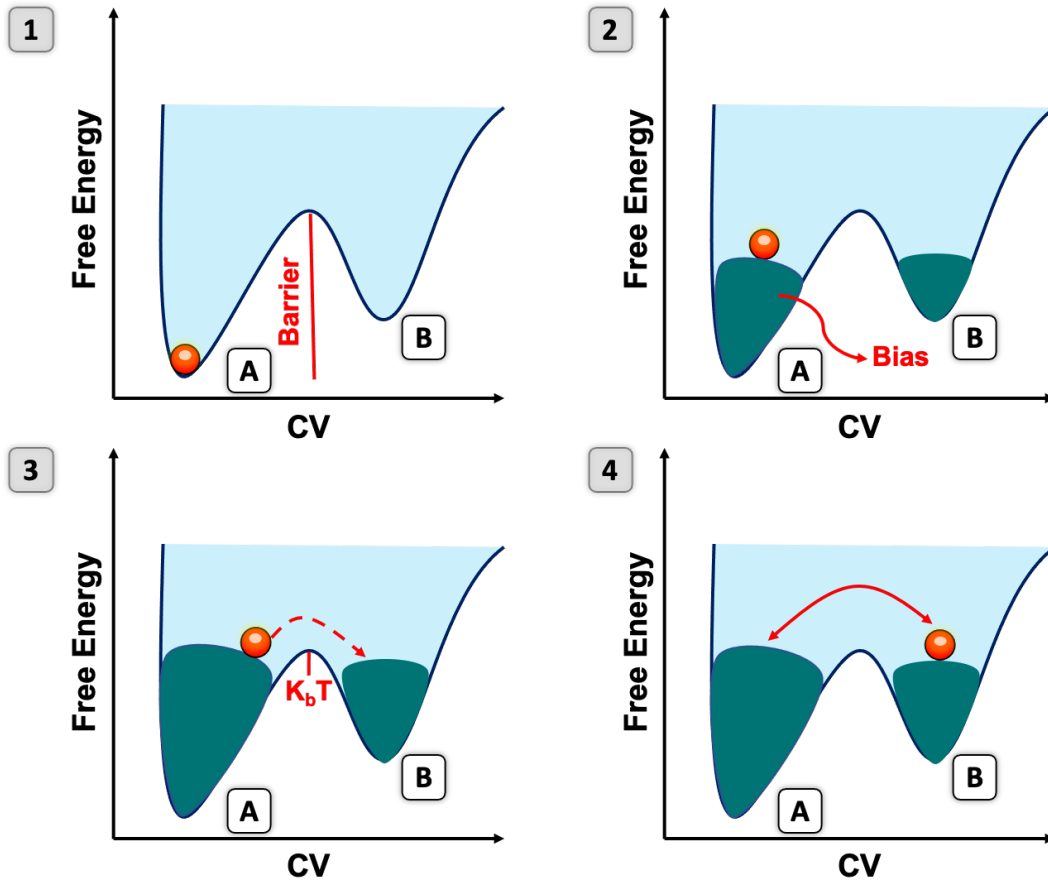


Figure 2.3: The scheme of the Well-tempered Metadynamics Technique. In **1.**, the system (represented by a red ball) is confined in a state **A**. It cannot reach state **B** due to the high energy barrier separating them. To overcome this, Well-Tempered Metadynamics introduces a bias potential (**2.**, in skobeloff) by depositing Gaussian along the trajectory, in function of the CVs. This gradually flattens the free energy wells, enabling the system to overcome barriers in the $k_B T$ range (**3.**). As a result, the system can now move freely between **A** and **B** (**4.**).

2.8.2 The choice of the optimal CVs

The identification of optimal and suitable CVs that accurately represent a system's essential dynamics remains a complex and non-trivial task. Poor CV selection (i.e. bad CVs selection or even sub-optimal CVs) - in particular those that fail to capture slow degrees of freedom or relevant transition pathways,

can lead to several issues, including the exploration of unphysical regions of configuration space, lack of convergence in a realistic time, ineffective biasing that results in trapping within metastable states, and ultimately the generation of FES that lack physical meaning[141, 146, 147]. To address these challenges, I conducted an exhaustive assessment of CV selection for my CG systems. As part of this effort, I employed the Deep Linear Discriminant Analysis (Deep-LDA) framework - a data-driven method designed to construct optimal CVs capable of distinguishing between biologically and physically distinct states in a high-dimensional descriptor space[146]. In this approach, a set of physically meaningful descriptors, typically derived from molecular structural features or biological experience (or, better, a combination of both), is used as input to a neural network composed of multiple hidden layers[146]. The network is meant to perform a nonlinear transformation of these descriptors, capturing complex, higher-order correlations from the original variables[146]. In other words, the neural network decreases the dimensionality of the input data via a series of continuous nonlinear transformations[146]. In the final layer, Linear Discriminant Analysis (LDA) is applied to the transformed data to identify the direction in feature space that best separates the predefined states - for example, conformational states **A** and **B**, which in my case correspond to "separate" (when the dimer is not formed) and "not separate" (when the dimer is formed), respectively. This direction defines the Deep-LDA CV, namely a low-dimensional yet highly informative variable that captures the key transition between those two states, thereby enhancing the efficiency and accuracy of free energy sampling[146]. A schematic overview on how the DEEP-LDA algorithm works is depicted in Figure 2.4.

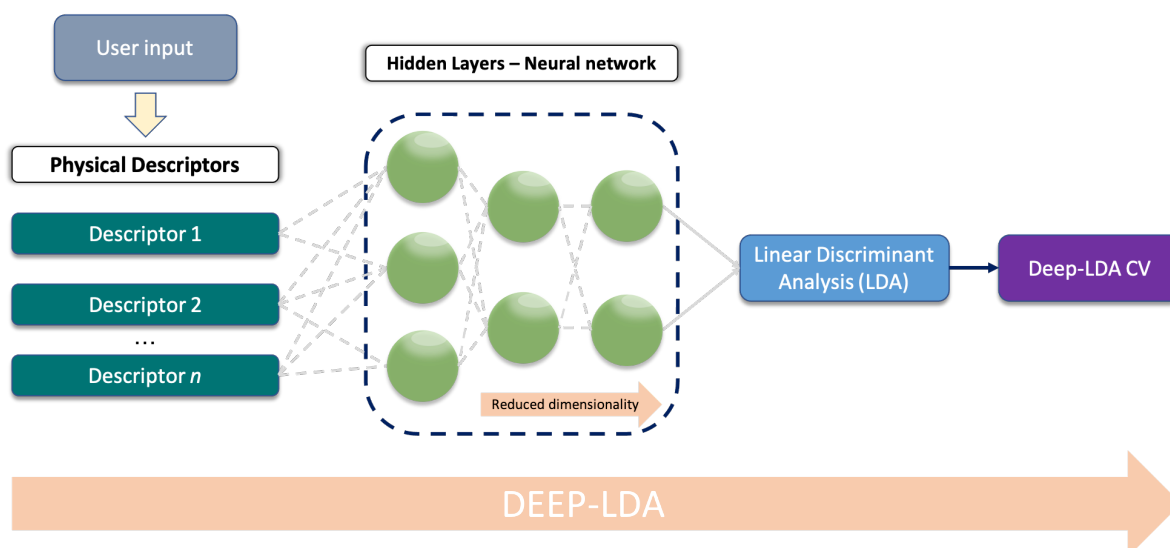


Figure 2.4: The user provides a set of n physically meaningful descriptors that characterize the molecular system. These descriptors are processed by a neural network composed of hidden layers, which perform a nonlinear transformation to extract complex correlations among the input variables, reducing the dimensionality of the problem (in this simplified case, $3 \times 2 \times 2$)[146]. At the final stage, Linear Discriminant Analysis (LDA) is applied in the transformed feature space to identify the direction that best discriminates the two conformational states, **A** and **B**. The resulting Deep-LDA CV effectively captures the most relevant transitions between these states[146]. State **A** corresponds to the "separate" state, whereas **B** to the "not separate" state.

2.8.3 Plumed 2

PLUMED 2 is an open-source plug-in that offers implementations of various enhanced sampling algorithms and CVs for some of the most popular MD codes, such as GROMACS[133, 148]. The utility of PLUMED lies in its ability to compute CVs from atomic positions. In this thesis, two CVs have been utilized and corroborated by DEEP-LDA (see section 2.8.2). The first CV (**DISTANCE**) involves the distance between the centers of mass of two transmembrane receptors; the second CV (**COORDINATION**) employed in my simulations is the count of transmembrane atoms within the two receptors (i.e. the number of contacts). Further details will be provided in the Chapter 4.

2.9 The Martini 3 force field

A significant constraint in MD is dictated by the restricted accessible time and length scales to tackle several biochemical issues, such as protein-protein interaction, ligand binding and receptor activation[149, 150]. To address this limitation, coarse-grained (CG) models, have been extensively adopted in different fields, such as biomedicine or biology[149, 151, 152, 153]. This model represents groups of atoms entailed in a fundamental unit called **bead** (Figure 2.5): a 4-to-1 mapping scheme, in which, on average, four heavy atoms along with their associated hydrogens are represented by a single CG bead[149]. This model has been parameterized using a top-down approach, i.e. considering thermodynamic partitioning data[149, 151]. The new Martini 3 force field architecture includes four main classes of CG bead types: P, Q, C, and N, featuring polar, charged, nonpolar, and intermediately polar chemical groups, respectively[149]. Charged beads and neutral beads interaction are modelled by the LJ potential and Coulombic forces, respectively[149]. These beads can be further refined to map the entire chemical space in terms of polarity, and hydrogen bond accepting and donating properties, fundamental descriptors of biology[149, 154]. This is one of the critical advancement compared to the Martini 2 force field[149, 151].

The main advantages of a CG force field can be summarized in three main steps. The following strengths of the CG approach are mainly taken from[155]:

1. **Faster Dynamics and Reduced Degrees of Freedom (DOFs):** The CG process reduces the number of atomistic degrees of freedom, which smooths the potential energy surface and lowers friction. This allows the CG system to sample phase space more rapidly within the same simulation time. Additionally, Martini models aggregate multiple water molecules into a single CG bead, significantly reducing solvent DOFs. This reduction yields a computational speedup roughly proportional to the square of the reduction factor, n^2 , enabling simulations on longer timescales. Depending on the mapping and system, CG simulations can be approximately 100 times faster than classical all-atom simulations[156]. Simulations in the absence of solvent (e.g., in vacuum) benefit from even greater speedups, despite they are unfeasible in biological backgrounds.
2. **Time Step:** The smoother energy surface in CG models permits larger integration time steps, typically 20 – 40 fs for dissipative particle dynamics; conversely, a typical all-atom integration step is 1 – 2 fs;
3. **Short-Range Interactions:** Most CG models compute only short-range interactions, typically with a cutoff around 1 nm/1.1 nm. This allows for the elimination of expensive particle mesh Ewald (PME) methods for electrostatic interactions[155, 157]. Many CG models also shift potentials to zero at the cutoff, allowing less frequent pairlist updates. This approach speeds up by at least an order of magnitude.

The following Figure (Figure 2.5) illustrates the overview of the mapping from an all-atom representation toward a CG model.

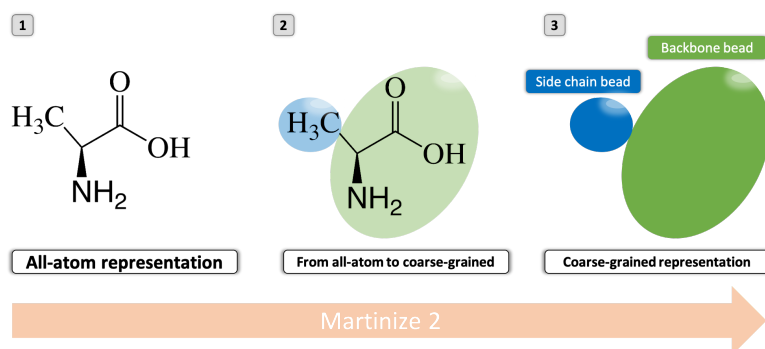


Figure 2.5: The scheme of the CG mapping from an all-atom model[149, 151]. The mapping from an all-atom structure to a CG representation is described in the following three steps. In **1**, the alanine residue is decomposed into its constituent moieties: the backbone and the side chain. In **2**, each moiety is mapped onto a single CG bead, resulting in two distinct beads, corresponding to the backbone (green) and side chain (blue) identified in **1**. In the final step (**3**), these two beads are assembled into a CG topology: the amino acid alanine is now represented by two CG beads. This mapping procedure is carried out using the *Martinize 2* tool[149].

2.9.1 Beads radii and beads interaction

Martini 3 beads vary in radii: *R* beads are classified as *regular beads* (4-to-1 model), *S* beads as *small beads* (3-to-1 model), which are particularly suitable for ring-like ligands, and *T* beads as *tiny beads* (2-to-1 model), used specifically for stacking and hydrogen-bonding distances between nucleotides[149]¹. Beads exhibit variations in size, and the potential well of the Lennard-Jones between distinct beads is adjusted accordingly to the type of the bead:

Bead Type	Bead's Size (nm)	vdW Radii (nm)
<i>R</i>	0.47	0.264
<i>S</i>	0.41	0.230
<i>T</i>	0.34	0.191

Table 2.1: Table depicting, Bead Type, size and their relative van der Waals radii.

The vdW radii for the Martini 3 are calculated as follows:

$$r_{vdW} = r_{min} = \frac{d_{min}}{2} = \frac{\sqrt[6]{2}}{2}\sigma, \quad (2.42)$$

where,

σ represents the LJ potential of the CG bead.

Consequently, the resizing of the coarse-grained beads underwent to a revision of the interactions between the Martini 3 beads from the previous version[149]:

Bead-Bead Interaction	σ_{ij} of LJ potential (nm)
<i>R</i> $\leftrightarrow$ <i>S</i>	0.430
<i>S</i> $\leftrightarrow$ <i>T</i>	0.365
<i>T</i> $\leftrightarrow$ <i>R</i>	0.395

Table 2.2: Table depicting the Bead-Bead interaction and their σ of Lennard-Jones potential.

The choice of σ_{ij} for *S* and *T* beads is determined by the aim of reproducing the symmetry of the molecules[149]. For self-interaction, σ_{SS} has been set to 0.41, σ_{TT} to 0.34. Instead, σ_{ij} is expressed by the arithmetic average of the cross-interaction (*R* $\leftrightarrow$ *S*, *S* $\leftrightarrow$ *T*, and *T* $\leftrightarrow$ *R*) as follows:

$$\sigma_{ij} = \left(\frac{\sigma_{ii} + \sigma_{jj}}{2} \right) - s_{ij}, \quad (2.43)$$

where,

$s_{ij} = 0.01$ nm represents the softness parameter for non-water beads.

This parameter is dictated by the principle that larger beads ought to possess a softer nature in comparison to the smaller ones[149]. This combination of rules suited for beads of different sizes effectively alleviates the caveat of artificial free energy barriers noted in dimerization potential of mean force (PMF) profiles of rings within the Martini 2 framework, as observed in these studies[149, 154, 158].

2.9.2 Elastic network

In Martini 3, non-bonded interactions between connected beads are excluded, also in the Elastic network of proteins[149]. An Elastic network employs harmonic potentials between non-bonded beads within a certain cut-off to keep the overall architecture and shape of a protein(Figure 2.6). They are used in conjunction with a CG model[149, 159]. Alessandri et al. demonstrated that weak bond force constants (below 700 *kJ/mol*) should not be used as they can promote excessive interactions between beads and

¹X-to-1 model specifies the number of atoms mapped in one single Martini bead.

environment, leading to artifacts, such as immoderate aggregation[149, 158]. That is why in thesis I chose this cut-off as force constant.

Within the Elastic network framework, local transient structural alterations are however permitted throughout the simulations[159]. Nevertheless, it is important to acknowledge that Elastic network model is based on the atomistic reference configuration given as input. Thus, the CG model cannot undergo to large conformational displacements from the reference structure and a correct 3D model or structure given in input is of paramount importance[159]. For this reason I put particular effort into finding representative starting models.

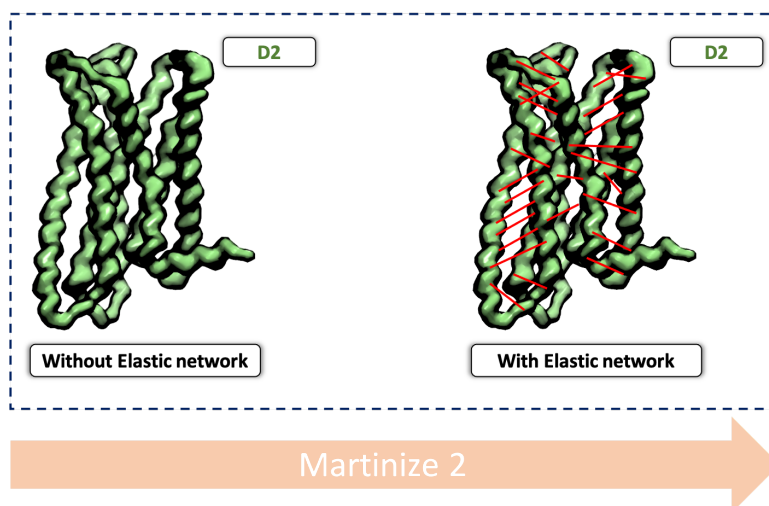


Figure 2.6: The Elastic network scheme applied to the D2[149, 151]. A simplified Elastic network pattern is depicted with red lines. The Elastic network framework is carried out using the *Martinize 2* tool[149].

As a result, conformational changes between active and inactive states and vice versa cannot be achieved in my simulations due to the Elastic network framework. Therefore, independent simulations in two distinct states, active and inactive, were conducted. The pipeline used for my CG simulations, integrating Elastic networks, exploits *martinize2*² as follows[160]:

Pipeline martinize 2

```
martinize2 -f atomistic_model.pdb -dssp /path/to/mkdssp -nt -cys auto -p
backbone -sep -ff martini3001 -elastic -scfix -eu 0.9 -el 0.5 -ea 0 -ep 0 -ef
700.0 -o topology-Martini.top -x Martini-model.gro
```

where,

- nt indicates the neutral termini applied at the N-terminus and C-terminus,
- cys auto applies disulphide bonds within the coarse-grained model automatically within a default cut-off,
- dssp applies the Dictionary of Protein Secondary Structure (DSSP)[161] to get secondary structure information, specifying if a particular domain is a helix, beta-turn or a loop,
- ff martini3001 -elastic sets the Elastic network in conjunction with the Martini 3 force field,

²<https://github.com/marrink-lab/vermouth-martinize>

- scfix improves the secondary structure of the coarse-grained model by using the information of the atomistic structure, adding extra dihedrals between side chains and backbone beads,
- eu 0.9 and -el 0.5 set the upper and lower cut-offs of the Elastic networks, respectively,
- ea 0 and -ep 0 set to 0 the Elastic bond decay factor and decay power of the Elastic networks, respectively,
- ef 700 sets to 700 kJ/mol/nm^2 the force constant.

2.9.3 Bonded and non bonded parameters in Martini 3

Bonded parameters

- The V_{bond} term of the functional implemented within the Martini 3 framework is described by a harmonic potential, such that:

$$V_{\text{bond}}(R) = \frac{1}{2} K_{\text{bond}} (R - R_{\text{bond}})^2, \quad (2.44)$$

where,

R_{bond} is the equilibrium distance set to 0.47 nm,

K_{bond} is the force constant tuned to $1250 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$.

- Also the angle is described by a harmonic potential V_{angle} :

$$V_{\text{angle}} = \frac{1}{2} K_{\text{angle}} (\cos \theta - \cos \theta_0)^2, \quad (2.45)$$

where,

θ_0 is the equilibrium bond angle set to 180° ,

K_{angle} is the force constant tuned to 25 mol^{-1}

for aliphatic chains[151]. For cis-unsaturated bonds, $K_{\text{angle}} = 45 \text{ mol}^{-1}$, with $\theta_0 = 120^\circ$; for trans-unsaturated bonds $K_{\text{angle}} = 45 \text{ mol}^{-1}$, with $\theta_0 = 180^\circ$ instead.

- Eventually, the improper dihedral angle potential $V_{\text{improper dihedral}}$ is described as follows[151]:

$$V_{\text{improper dihedral}}(\Theta) = k_{id} (\Theta - \Theta_{id})^2, \quad (2.46)$$

where,

Θ_{id} is the equilibrium angle,

Θ is the angle between i,j,k and j,k,l,

K_d is the force constant.

Non bonded interactions

The functional forms for both bonded and non-bonded interactions of the Martini 3 force field are kept unchanged from Martini 2[149, 151]. Following parameters promoted by de Jong and colleagues[131], the optimal cut-off range for non-bonded interactions results in 1.1 nm. The primary concern is related to the significantly restricted range of the LJ potential in the liquid phase, as discussed in this perspective review[162]. Consequently, the free energy of vaporization (Δ_{VAP}), free energy of solvation (Δ_{SOL}), and gas-liquid surface tension cannot be accurately determined yet[149].

Instead, properties involving $\Delta\Delta G$ calculations, like the excess free energy of mixing or the oil/water transfer free energy, can be easily retrieved from the CG framework[149]. The non bonded interactions defined within the Martini 3 force field are described by the sum of the LJ potential (V_{LJ}) and the long range Coulombic interaction (V_C)³, as follows:

³In my CG simulations, reaction-field treatment is used.

$$V_{LJ}(r_1, r_2, \dots, r_n) + V_C(r_1, r_2, \dots, r_n) = \sum 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i,j} \frac{q1 \cdot q2}{4\pi\epsilon_0\epsilon_r r_{ij}}. \quad (2.47)$$

2.9.4 Insane

One of the most critical steps when performing membrane simulations is ensuring both the correct positioning and integrity of the lipid bilayer. There are various methods to achieve this, such as using GROMACS or CHARMM-GUI[133, 163]. Nevertheless, the simplest and also more accurate approach when dealing with CG Martini simulations is utilizing *insane.py* (an acronym for INsErT membrANE) tool[164]. Using this tool, researchers can not only define any desired lipid composition - from the simplest to the most complex, depending on the specific goals - but also customize the simulation box geometry, solvate the system, and add counterions to ensure overall charge neutrality[164]. Using *insane.py*, it is also possible to reorient the protein structure perpendicularly to the membrane's plane. However, to correctly position the CG protein along the z -plane of the membrane, I employed the PPM 2.0 Server[165]. The pipeline used for my CG simulations exploits *insane.py* as follows[164]:

Pipeline insane

```
python2.7 insane.py -f Martini-model.gro -p topology-Martini.top -pbc cubic -x
xdim -y ydim -z zdim -l [lower leaflet membrane composition] -u [upper leaflet
membrane composition] -sol W -d 1.5 -o insane-Martini.gro -center
```

where,

- pbc cubic indicates the pbc box shape, in this case cubic,
- x, -y, and -z applies the box dimension in the x, y, and z dimension, in nm, respectively,
- l, -u sets the lipid composition in the lower and upper leaflet, respectively,
- sol W applies the CG water model to solvate the box. The total number of W CG beads inserted into the box is determined by the available free volume of the box,
- d 1.5 sets the minimum distance to the periodic images, in nm,
- center centers the protein in the membrane.

2.10 MD Simulation Protocol setup

So far, we have described the preliminary phases of the protocol, including the mapping from the all-atom system to Martini 3 and the embedding of monomers into the membrane, along with the adopted specifications. The protocol majorly follows the parameters of de Jong and colleagues[131]. After these preliminary steps, the protocol proceeds with an energy minimization step using the steepest descent algorithm for 50,000 steps[166]. This step is carried out to eliminate steric clashes and ensure an energetically favorable protein conformation prior to subsequent simulations. The cutoff scheme employed is Verlet, and the reaction-field treatment for long-range electrostatic interactions is used[167, 168]. The minimization phase is followed by a canonical ensemble (NVT) simulation, in which temperature and volume are kept constant, for 50 ns. Constraints were applied using the LINCS algorithm, and the temperature was maintained at 315 K using the V-rescale thermostat[132, 137]. Subsequently, an NPT simulation was performed using the Parrinello-Rahman barostat with a reference pressure (ref-p) of 1 bar, compressibility of 3×10^{-4} and τ -p of 12 ps[134]. Pressure coupling was applied semi-isotropically. Both the NVT and NPT simulations were conducted with backbone restraints. The NPT phase was run for 100 ns. The production phase was then carried out without restraints for 1 μ s. From the equilibrated portion of the trajectory (identified as the plateau region of the RMSD plot), clustering was performed, and the most representative configuration was selected for subsequent enhanced sampling simulations.

2.10.1 Backmapping: from the coarse-grained to the atomistic structure

Various biological events, such as dimerization, ligand binding/unbinding and protein-lipid interactions, are characterized by a slow process, making them difficult to reproduce computationally[150, 169, 170, 171, 172]. To overcome this challenge, CG models are employed to speed-up these events[172]. However, once the desired structure or interaction has been obtained, it is often necessary to reconstruct the atomistic detail in order to achieve higher resolution and more accurate structural information. This step is essential, for instance, in docking studies or MD reproducibility[149]. This impelling need arises because, due to the reduced degrees of freedom resulting from atom-to-bead mapping in CG models, certain molecular properties, like hydrogen bonding, may be represented inaccurately or not captured at all[149, 150]. This "translation" process is best known as *backmapping* (Figure 2.7). Backmapping can be performed in three distinct modalities[173]: using a geometrical, a machine learning-based, or a fragment-based approach[172, 174, 175]. The backmapping procedure I used in this thesis is the latter approach and it is implemented within the *CG2AT2* framework[172].

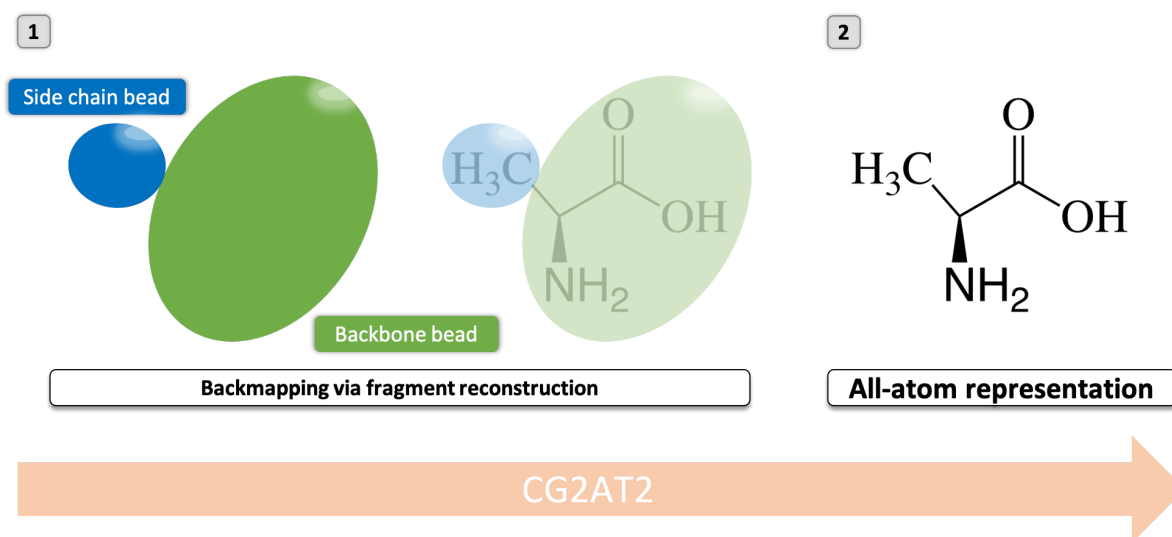


Figure 2.7: The scheme of the CG backmapping from CG model[149, 151]. The mapping from a CG structure to an all-atom representation is described in the following two steps. In **1**, the alanine residue is represented by two beads: the backbone (green) and the side chain (blue). In **2**, each bead is backmapped into the corresponding atomic fragment composing the alanine residue, thereby reconstructing the all-atom representation of the residue[176]. The algorithm iterates over the entire CG model presented in input.

The *CG2AT2* workflow can be divided in four major steps[172]:

- **Conversion:** Atomistic fragments are mapped and aligned to the center of mass of the CG bead. These atomistic fragments contain all the necessary chemistry and stereochemistry information to convert the model to an atomistic structure[172]. *CG2AT2* provides atomistic fragments details for AMBER, CHARMM, OPLS, and GROMOS[177, 178, 179, 180]. Additionally, conversion can be performed *de novo* (nude and crude reconstruction of the atomistic structure using only the CG beads information) or by providing an initial atomistic structure. The latter method is preferable, since it allows the correct repositioning of amine bonds within the correct plane, minimizing initial steric clashes[172]. This was also done to reduce the timing for the minimization step explained after. For my backmapping procedure, I used atomistic fragments from CHARMM36m providing the initial atomistic structure as input. At this stage, the quality of the model becomes paramount, as its energetics - regardless of the algorithm used to obtain it - may differ significantly from those of its atomistic counterpart[150].
- **Minimization:** bond lengths and angles are adjusted and any steric clashes are corrected. *CG2AT2* exploits the steepest descent algorithm to perform minimization[166].
- **Integration:** After rebuilding, the system components are reintegrated and energy-minimized. Occasionally, artifacts may arise: lipid tails might insert through the ring-like structures of aromatic residues. Such configurations must be avoided to elude unwanted lipid-protein interaction,

and *CG2AT2* can detect them by identifying abnormal bond lengths[172]. These issues are resolved by translating the affected atom groups, followed by an additional minimization and geometrical optimization step to eliminate any remaining potential distortions that could introduce artifacts. This correction process is repeated iteratively for each detected inconsistency[172].

- **Equilibration:** *CG2AT2* involves relaxing the system through a *NVT* simulation. Position restraints are applied to the backbone C_α atoms of the protein, if present[172].

The following pipeline was used for my coarse-grained-to-atomistic reconstruction solely for my protein structures and not for lipids:

Pipeline CG2AT2

```
cg2at2 -ff charmm36-mar2019-updated -fg charmm36 -w tip3p -c CG_input.gro -a
AT_input.pdb
```

2.10.2 The Martini 3 Force Field: Future Perspectives on the Horizon

Recent advances with the Martini 3 CG force field have markedly accelerated studies, enabling investigations that range from GPCR oligomerization and detailed small-molecule parametrization to intricate lipid-protein interactions[169, 181, 182]. Despite this progress, open challenges remain: protocols for molecule parametrization still need fine-tuning, protein flexibility is only partly captured, and further work is required to map allosteric, cryptic or protein-protein interfaces[176]. The next sections outline how forthcoming extensions of Martini 3 could tackle these gaps and broaden its impact on drug-discovery pipelines[176].

Cryptic pockets

In general, drug-binding sites on target proteins are accessible cavities exposed even in the absence of a ligand, as observed in GPCRs[183]. Yet, due to the inherently dynamic nature of protein structures, additional transient cavities - commonly known as cryptic pockets - can emerge upon ligand interaction, driven by conformational rearrangements[183, 184]. These pockets are typically not detectable in static, ligand-free protein conformations (Figure 2.8).

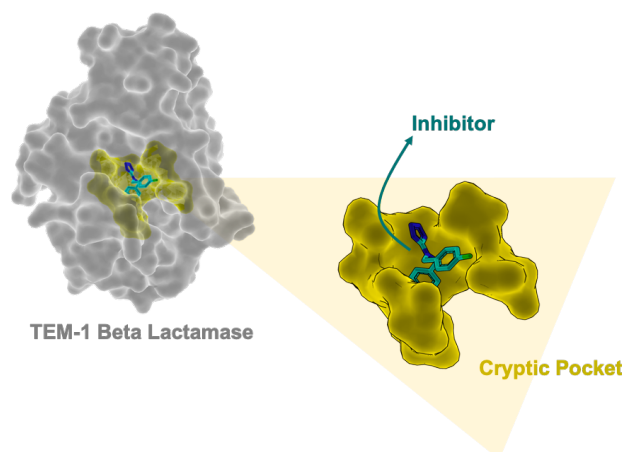


Figure 2.8: TEM-1 β lactamase (light gray) complexed with a high-temperature-destabilizing allosteric inhibitor bound within its cryptic binding pocket (yellow)[185].

Cryptic pockets are not *permanently* viable and they manifest sporadically in response to a ligand action[176]. Therefore, these novel allosteric transient cryptic binding sites could serve as putative targets for drug development for various diseases, including therapeutic drugs for cancer treatment[176, 186] or SARS-CoV-2 infections[176, 187].

Cryptic binding sites in proteins possess an opening time in the range up to millisecond, and for this fast exposure they are not often captured by state-of-the-art experimental methods[176]. Although

there have been recent advancements in supercomputers, accessing these timescales for MD simulations remains largely unfeasible[187, 188]. As mentioned in section 2.8, to address this issue, enhance sampling techniques may be exploited, such as mixed-solvent MD, Metadynamics/Well-tempered Metadynamics or flooding MD[141, 142, 176]. However, using these approaches remain time consuming, and they still require significant efforts to open cryptic pockets[176, 189, 190].

Another potential caveat of using enhanced sampling techniques lies in the selection of CVs (see section 2.8.2); a poor choice can hinder convergence and reduce the effectiveness of the method[147, 176, 191]. Thereby, a combination of CGMD and enhance sampling techniques may accelerate the findings of cryptic pockets[176].

A further limitation of using CG approaches lies in accurately describing protein conformational flexibility, which is crucial for biological function and therapeutic scopes[149, 176]. To avoid any collapse of the protein's scaffold, Martini-based CG approaches exploit harmonic potential restraints on BB beads, also defined as elastic networks[159]. This strategy can induce bias results[149, 176]. Strategies like localized distance restraints, Gō-like models, or adaptive Elastic networks, and, in the next future also artificial intelligence (AI), may help to counteract this issue[176, 192, 193].

Drugs targeting PPI

PPIs are emerging drug targets with three main classes of modulators: small molecules, antibodies, and peptides. Each class has its strengths and bottlenecks: these aspects have been extensively discussed and summarised in recent reviews (as reported in this perspective review[176]). One possible strategy to mitigate these limits is the use of Proteolysis Targeting Chimeras (PROTACs), which function catalytically to induce target degradation[176]. Their design and optimisation can be significantly aided by computational approaches, including protein-protein docking and MD simulations, particularly when structural data are lacking. CGMD such as Martini 3 offer an efficient framework to investigate large molecular systems and assist in the rational development of PROTACs (as reviewed here: [176]).

Soft nanoparticles with Martini 3

Certain drugs, when administered orally, exhibit low bioavailability in the blood stream, thereby reducing their efficacy and necessitating multiple administrations during the therapeutic regime. In this regard, drug shuttle systems play a vital role in increasing drug efficacy by overcoming challenges such as low aqueous solubility, like the blood brain barrier and rapid degradation of many approved drugs (Figure 2.9)[176].

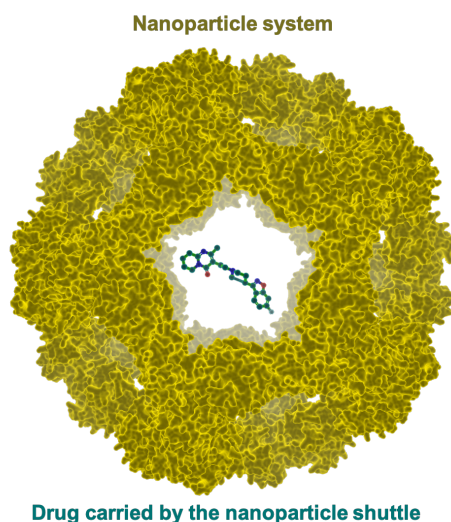


Figure 2.9: Example of nanoparticle shuttle system carrying a drug, in this example Risperidone.

Indeed, Soft nanoparticles (SNs), including lipid, polymer, and hybrid carriers, are properly engineered to counteract these pitfalls, optimizing drug solubility, targeting efficiency, and stability[176]. In this context, CG modeling allows for the computational study of SN formation, morphology, and drug distribution across biological membranes[176]. As highlighted in this chapter, the expanded chemical

space of the Martini 3 force field, compared to Martini 2, offers greater potential for unlocking deeper insights into SN behavior, construction and interactions[149, 176].

Protein binding pockets embedded in the membrane bilayer

CG simulations under the Martini 3 force field framework represent a critical advancement in understanding protein-ligand interactions within membranes, crucial for drug discovery and membrane biology research[176]. Future directions aim to expand simulation capabilities with Martini 3 force field enhancements, facilitating automatic parameterization for diverse ligands and enabling high-throughput screening[176]. This promises to further accelerate drug discovery efforts and deepen our understanding of membrane biology, highlighting the pivotal role of CG simulations in bridging experimental challenges and computational insights in complex biological systems[176].

In conclusion, the Martini CG infrastructure has emerged as a powerful tool for *in silico* studies targeting druggable cryptic binding sites, optimizing protein-ligand interactions, and designing drug delivery systems[176, 194]. Eventually, the automated parameterization of ligands already furnished by Bereau and colleagues[195], as well as tools like Swarm-CG[196] and PyCGTOOL[197], are embracing the Martini 3 framework[149, 176]. Next subsection will briefly illustrates the new advancement in Martini 3 ligand parametrization.

CG Ligand parametrization

CGMD simulations with Martini 3 enable the investigation of large protein complexes on millisecond timescales, but still lack an automated approach for generating small-molecule models. Although Auto-Martini 2 was a widely adopted tool for rapid ligand parametrization, the expanded chemical space and improved accuracy of Martini 3 in reproducing protein-ligand interactions have rendered it unsuitable[195, 198]. To bridge this gap, Auto-MartiniM3 represents a valid alternative that automates the CG of small molecules for Martini 3 *without the need* for AA input trajectories[198]. This tool only relies on Voronoi graph partitioning of atoms into beads, assign partial SMILES (SMILE code is the only input needed for this tool) into beads, assign bead type according to their polarity and then bond interaction. Auto-MartiniM3 offers a fast and robust solution for high-throughput generation of CG models.

Protein-ligand binding

As treated in the previous paragraphs, the investigation of protein-ligand interactions remains of paramount importance[176, 194]. Although determining the binding free energies between a receptor and a molecule (whether agonist or antagonist) is quite resource-intensive and difficult to setup, it is a crucial step for the successful development of pharmaceuticals and lead compounds[194]. In this regards, the application of the Martini 3 force field can therefore accelerate this process, and several results have already been published recently[194, 199, 200, 201].

2.11 Contribution of this Chapter

My contribution to paragraph 2.10.2 is based on an in-depth review of the literature concerning potential future outcomes of the Martini 3 force field, providing insights as outlined in this last subsection. In addition, I contributed to revising and refining the manuscript to ensure its suitability for publication in the journal, actively collaborating with members of my research group at the Forschungszentrum Jülich and multiple groups around Europe and World.

Paragraph 2.10.2 is an extract published in accordance with §5(6) of the doctoral regulations in the following peer-review Journal:

Lisbeth R. Kjølbye , Gilberto P. Pereira, Alessio Bartocci, Martina Pannuzzo, Simone Albani, Alessandro Marchetto, Brian Jiménez-García, Juliette Martin, Giulia Rossetti, Marco Cecchini, Luca Monticelli and Paulo C. T. Souza. "Towards design of drugs and delivery systems with the Martini coarse-grained model." QRB discovery 3 (2022): e19.

Chapter 3

Coarse-Grained Molecular Dynamics (CGMD) Platform

CG models are simplified representations of AA models, in which essential molecular features are only retained. This approximation enables faster sampling by decreasing the degrees of freedom[149]. Consequently, CG simulations can span longer timescales, providing insights into the dynamics of macromolecules that are beyond the scope of classical MD simulations. CG models have been extensively utilized to study a multitude of biological problems, such as protein folding mechanisms or protein-membrane interactions[156]. The development of CG protein models is constantly expansion. Numerous studies have exploited CG force fields, such as Martini[149, 151], SIRAH[202], OPEP[203], Shinoda-Devane-Klein (SDK)[204], or ELBA[205].

The **Coarse-Grained Molecular Dynamics (CGMD) Platform** is a web service¹ dedicated to prepare and analyze CG simulations through two independent force fields[156]: the Martini 2.2 and the Southamerican Initiative for a Rapid and Accurate Hamiltonian (SIRAH)[151, 202]. CGMD Platform represents an evolution of the already existing web server Martini coarsE gRained MembrAne proteIn Dynamics - MERMAID[206]. As the previous version, the user can change several components of the simulation, from the molecular dynamics parameters (mdp) to the lipid composition. However, in this version, several new features are introduced:

- **New web interface:** CGMD Platform exploits a new graphical engine by using a combination of:
 1. **Javascript** (<https://www.javascript.com/>)[207];
 2. **jQuery 3.4.1** (<https://jquery.com/>);
 3. **Bootstrap 4.4.1** (<https://getbootstrap.com/>);
 4. **PHP 7.4.3** (<https://www.php.net/>)[208].

It is also conceivable to interact and inspect the 3D protein structure model directly on the browser via NGL Viewer[209]. It is also possible visualizing trajectories on the browser[209]. The implementation of the interface was done in collaboration with Carlo Alberto Rossi from University of Verona, Italy.

- **SIRAH force-field:** The SIRAH 2.2 force field is implemented within CGMD Platform. Currently, only simulations in explicit solvent are available in this platform. Here an input *pqr* file must be provided. A pqr file contains charges and radii in addition to the canonical RCSB PDB information. Users can obtain this file, for example, via APBS Server[210]. All the necessary information to run and apply this tool were acquired from Sergio Pantano from Institut Pasteur de Montevideo, Uruguay.
- **Martini simulations in membrane environment:** This represents the core functionality of the CGMD Platform. The web interface incorporates all features of the previous published MERMAID tool[206], along with additional configurable parameters (for further details, see the following link: <https://molsim.sci.univr.it/mermaid/begin.php>). A direct integration with the

¹<https://molsim.sci.univr.it/mermaid/begin.php>

GOMoDo web server allows for the retrieval of GPCR models[211]. The connection with GOMoDo web server was done in Collaboration with Dr. Rui Ribeiro. Given the fact that GPCRs share the conserved 7TM helical architecture, retrieved models are structurally aligned to the Nociceptin receptor in its inactive conformation (PDB ID: 4EA3), as downloaded from the OPM database[212, 165]. This receptor serves as a reference to achieve proper perpendicular orientation within the lipid bilayer, a requirement for realistic membrane insertion of the protein complexes. The structural alignment is performed using the LovoAlign tool[213]. To obtain detailed secondary structure information, a local version of the Dictionary of Protein Secondary Structure (DSSP) has been employed[161]. This web platform covers a broad user base. Advanced users are granted the flexibility to modify default parameters provided through the graphical interface and may upload their own "home-made parameters" for MD simulations. The force fields Martini2.2 and Elnedyn22 - two variants of Martini 2.2 - are currently implemented in the system[151]. After selecting the preferred force field, users can navigate through several configuration panels, each offering extensive customization options. Among these is support for the insane.py script[164], which facilitates the embedding of protein complexes into lipid membranes. Users can also choose from a variety of counterions, Na^+ , Ca^{2+} , choline (NC3^+), and Cl^- , to neutralize the system. Furthermore, the platform provides access to over 55 different lipid species, enabling detailed customization of membrane composition. Each leaflet of the bilayer can be individually tailored, including the option to introduce membrane asymmetry. In addition to manual selection, users may select one of the four predefined membrane compositions (main lipids are present in the next figure) available through the CGMD Platform interface (Figure 3.1)[214]:

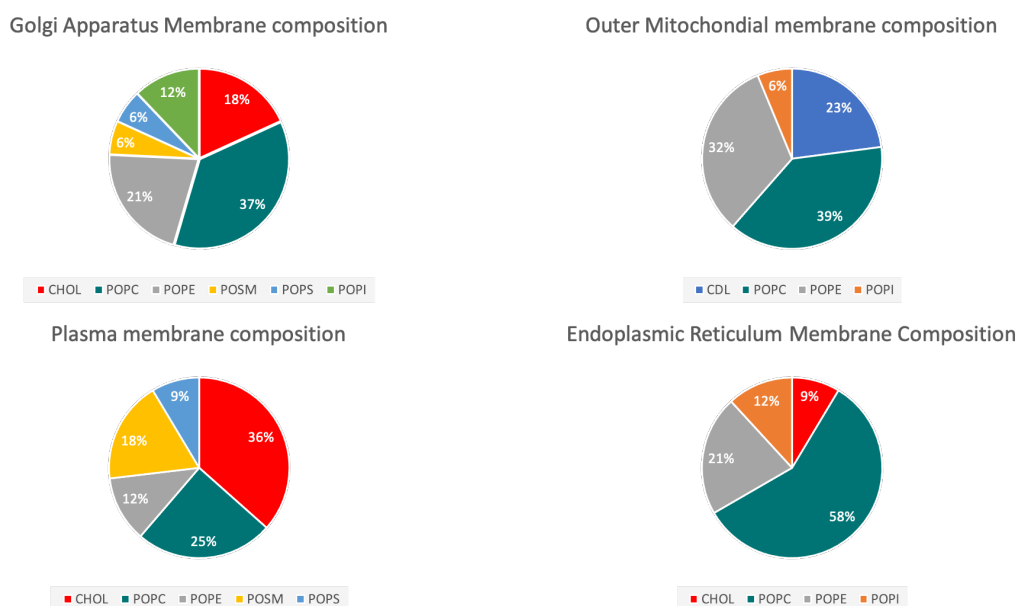


Figure 3.1: Main lipids membrane composition implemented within this platform for: Golgi Apparatus, Mitochondrial membrane, Plasma membrane and Endoplasmic Reticulum.

- **Martini simulations in water environment:** this tool allows the simulation of soluble proteins, in addition to membrane simulations, with the Martini 2.2 force field. This tool replicates the procedural workflow of the previous step, except for the insertion in the membrane step. Additionally, it should be noted that proteins cannot be obtained from GOMoDo[156].

3.1 CGMD Platform documentation

Dynamic pop-ups, generated via locally installed Bootstrap Tour tools (<https://bootstraptour.com/>, see 3.2), offer step-by-step documentation for each parameter and action. This ensures users receive real-time, contextual guidance at any step without the need to browse separate pages or verbose manuals. Additionally, tutorials include pre-calculated systems that illustrate all computational features available within the CGMD platform[156].

3.2 CGMD Platform: computational details

All the CG simulations in the CGMD Platform are run on a Linux Ubuntu Server 20.04.1 LTS machine with 12 cores and 32GB of RAM. GROMACS 2019.3 is installed for running the simulations[133]. To model the system, Modeller 9.25 is utilized[115]. Users must insert their Modeller key to run this feature. All jobs are stored on the server for future retrieval.

3.3 CGMD Platform pipelines

Three tools are implemented within the CGMD Platform: a tool for simulations with the SIRAH force field, a tool for running simulations with the Martini 2.2 force field, and a backmapping tool *backward.py* that exploits a geometrical approach (more information about backmapping here: 2.10.1) to reconstruct the atomic structures. The following pipelines are utilized:

3.3.1 The SIRAH Pipeline

The SIRAH-implemented architecture is structured into five sequential layers. The first layer involves user input. The second focuses on the addition of solutes. The third encompasses energy minimization, system equilibration, and the production run. The fourth layer pertains to data analysis, and finally, the fifth layer facilitates downloading the results to the local machine (see Figure 3.2). Specifically, in the first layer, the user is responsible for uploading a PQR file containing one or more protein complexes[210]. This file format, an extension of the PDB, includes not only structural information but also the partial charges assigned to each amino acid. These charges are essential, for instance, in accurately mapping the polarity of the beads during the CG mapping process. As soon as the PQR file is uploaded, each amino acid is automatically converted into its corresponding CG representation.

Subsequently, the server generates the simulation box by adding ions to neutralize the charge of the box and solvent. The water model employed by SIRAH corresponds to the tetrahedral WT4 model.

In the third layer, the various steps of the molecular dynamics protocol are carried out, starting with two minimization stages (with and without restraints), followed by NVT and NPT equilibration phases, and finally the production run. Each step is accompanied by a series of analyses aimed at assessing the stability of both the system and the simulation, such as RMSD analysis, pressure and temperature variability. This third layer is directly connected to the subsequent one, the analysis layer.

The final step consists of the user downloading the data to a local cluster or workstation. The primary objective of this tool is to facilitate the setup and initiation of CG molecular dynamics simulations, providing the end user with an efficient means to streamline and manage the simulation workflow. More information and a detailed explanation can be found here[156].

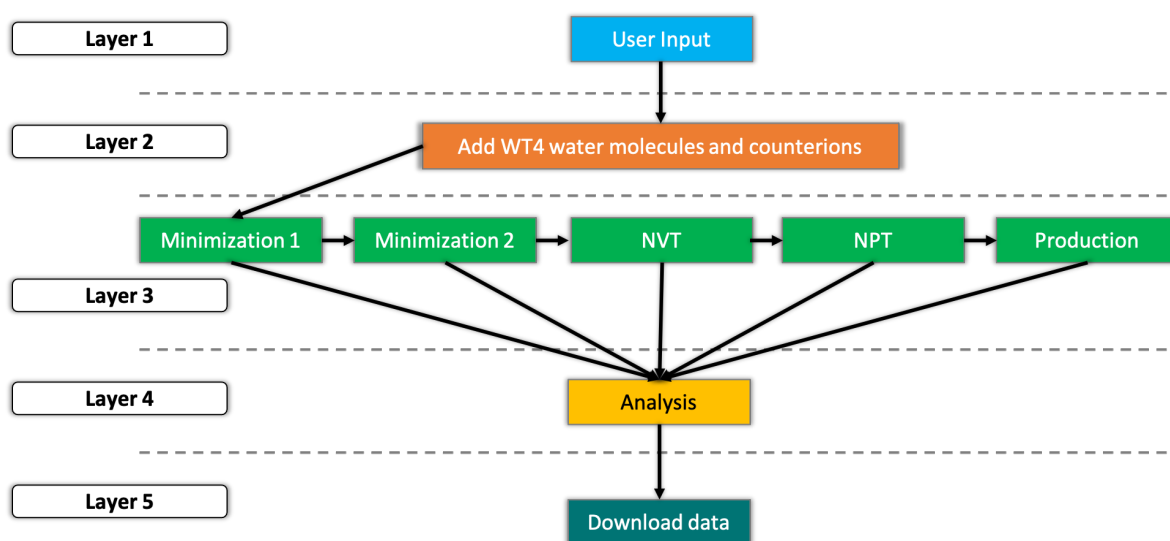


Figure 3.2: Configuration of the SIRAH Pipeline embedded in the web platform.

3.3.2 Martini 2.2 with membrane

The architecture implemented in Martini 2.2 is organized into five sequential layers, each corresponding to a distinct stage of the simulation pipeline (see Figure 3.2). The first layer handles user input, where the user uploads a PDB file containing one or more protein complexes. Upon upload, each amino acid residue is automatically converted into its corresponding CG representation according to the Martini 2.2 force field topology.

The second layer manages system setup by adding solutes, neutralizing ions, and solvent molecules to construct the simulation box. Unlike the previous tool, this stage includes a membrane embedding procedure, allowing the protein complexes to be inserted into a customizable lipid bilayer. Users can choose from over 55 lipid types, thereby enabling the exploration of potential protein–lipid interactions, regardless of their level of expertise.

The third layer covers the same simulation protocol seen before. It begins with an energy minimization step, followed by NVT and NPT equilibration phases, and culminates in the production run. Each stage is accompanied by automated stability assessments to monitor the behavior of the system. This layer is directly integrated with the analysis stage (Layer 4).

The fourth layer is dedicated to post-simulation analysis, offering users a suite of tools to interpret and visualize the simulation results.

Finally, the fifth layer enables users to download all output files to a local machine, cluster, or workstation for further analysis. Additional details and a comprehensive description of this pipeline are available in Marchetto *et al.* [156].

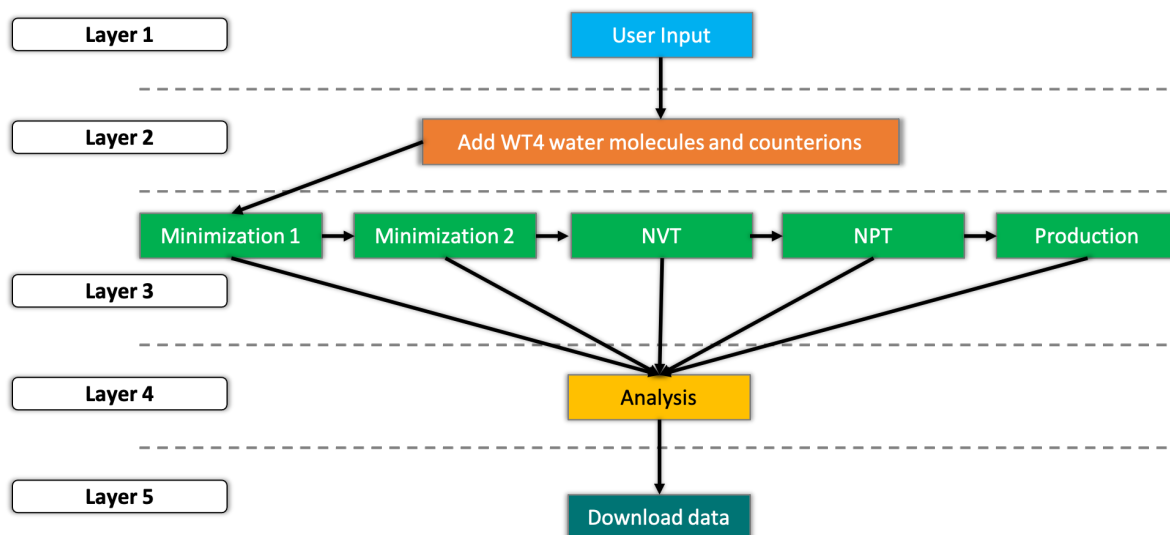


Figure 3.3: Configuration of the Martini 2.2 Pipeline embedded in the web platform[156].

As previously stated, the schematic representation of the "Martini water" section will not be included in this thesis, as it is identical to the one already presented - except for the step involving the insertion of the protein into the membrane.

3.3.3 Backmapping procedure

The reduction in resolution in CG representations poses a limitation for interpreting detailed aspects of simulations, such as specific contacts in a quaternary complex or contacts between ligand and surrounding residues[156]. Atomistic-level details are therefore an essential instrument for understanding molecular recognition and precise intramolecular and intermolecular interactions. The process of recovering atomistic details from a CG representation is known as **backmapping**, and it has been widely discussed in Section 2.10.1. The backmapping procedure via the *backward.py* algorithm is implemented in the CGMD Platform[156, 174]. The four-layers backmapping pipeline is only available for the Martini Models. With this tool, users can backmap a soluble protein from the Martini force field to different force fields, such as Charmm36[215], Amber[216], or Gromos[217]. For membrane systems, there are nine supported lipids: DPPC, POPC, CHOL, DOPG, DOPC, DOPE, DOPS, POPG, and POPE. Slipids force field topologies are utilized to reconstruct the atomistic details for both protein

and lipid, and, therefore, the associated membrane protein is backmapped to the Amber force field[156]. Follows the backmapping pipeline:

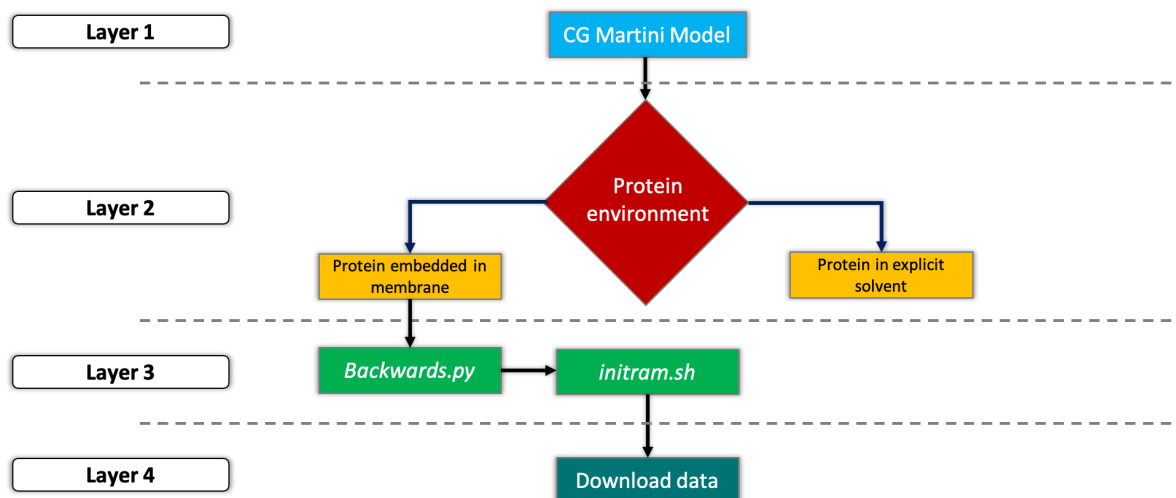


Figure 3.4: Simplified workflow diagram illustrating the backmapping procedure of a CG Martini model within the CGMD Platform[156]. *Initram.sh* is a shell script that accepts parameters for the backmapping[156, 174].

3.4 CGMD Platform as part of the Human Brain Project (HBP)

The Human Brain Project (HBP) provides researchers with advanced tools and services for sustainable brain simulation studies. These resources allow scientists to create, simulate, validate models, and share results efficiently. Simulation engines are available for different levels of brain complexity, from molecular and cellular to network and whole-brain levels. They also support interaction between multiple simulation engines, facilitating the simultaneous simulation of detailed models within broader, more abstract frameworks. CGMD Platform is part of HBP: <https://wiki.ebrains.eu/bin/view/Collabs/molecular-tools-cgmd-platform/https%3A/molsim.sci.univr.it/mermaid/begin.php>.

3.5 CGMD Platform: Application Cases

In this section, we will briefly expose three powerful application cases implemented within the CGMD Platform[156]. These tools offer unique functionalities that cater to a variety of needs, making them invaluable assets for both personal and professional use. Files for running test case are freely available in this Github repository: https://github.com/JavaScript92/CGMD_Platform.

3.5.1 Martini 2.2 with Membrane

This first tool describes the preparation and running of a CGMD simulation for the *Rhodobacter sphaeroides* translocator protein (RsTSPO), highlighting its roles and the effect of cholesterol (CHOL) on its structure[153]. The simulation, accessible online, was conducted in a 70% Phosphatidylglycerol (POPG) and 30% CHOL membrane model with Na^+ and Cl^- as counterions. The process involved the classical MD steps shown in Figure 3.3, showing stable results after ~ 20 ns. Users can track progress via a status bar and access the final structure and reference files through a public GitHub repository (https://github.com/JavaScript92/CGMD_Platform). A demo case option allows automatic PDB upload and simulation execution. All results can be downloaded as a zip archive.

3.5.2 Water Martini

This second tool describes the simulation of the hen egg-white lysozyme (PDB: 1AKI) using a 15 nm^3 cubic box[218]. The system, neutralized with Na^+ and Cl^- ions, was simulated via the classical four MD steps: minimization, NVT, NPT, and production. The structure was equilibrated after ~ 2

ns, with Root-mean Square deviation (RMSD) and radius of gyration plots indicating general system stability. Results, analysis, structures, and other different data can be downloaded as a zip archive.

3.5.3 The SIRAH Force field

This last tool describes the Binary Complex of Restriction Endonuclease HinP1I with its Cognate DNA (PDB: 2FL3) using the SIRAH force field[202, 219]. The system, prepared as outlined in Figure 3.2, was simulated in an octahedron box with K^+ and Cl^- as counterions to achieve physiological ionic strength. Also here, users can track the simulation via a progress bar and perform various analyses, including dynamic plots and 3D visualizations. Results, molecular structures, and properties can be downloaded as a zip archive.

3.6 Future Perspectives of the CGMD Platform

In forthcoming releases, the CGMD Platform will be upgraded to support Martini 3, whose bead definitions cover a broader chemical space than those of Martini 2.2, as explained in the previous chapter. By incorporating this force field, the platform will enable users to probe an even wider spectrum of biologically relevant questions with greater molecular detail. Backmapping capabilities will, at the same time, be changed towards more "state-of-the-art" tools. Current *backwards.py* tool will be replaced by more advanced solutions, such as CG2AT2 and the emerging HEroBM, offering higher accuracy and efficiency when converting CG trajectories to atomistic resolution topologies[150, 172]. To extend the scope of accessible phenomena, the pipeline will integrate enhanced-sampling protocols, such as OPES (and its variants) and well-tempered metadynamics, along with dedicated workflows for ligand parametrization[141, 142, 147]. These additions will facilitate the exploration of rare events and support the characterization of receptor–ligand interactions, at a CG level. Finally, the current PHP/JavaScript web-server architecture will be migrated to a fully Python-based framework. This transition will improve code portability, simplify maintenance and debugging, and foster more rapid development of new features.

3.7 Conclusions

CGMD platform entails a multitude of functionalities, by offering (Figure 3.5):

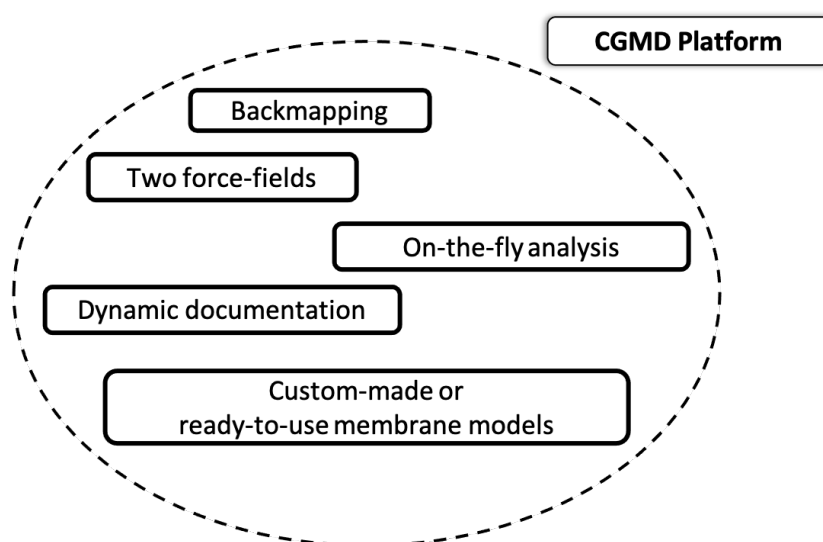


Figure 3.5: Set of functionalities offered by CGMD Platform: a backmapping service, options for Martini or SIRAH CG force fields, on-the-fly analysis of simulation parameters like RMSD, temperature, pressure, and energy. Dynamic help pop-ups for user guidance, and a choice of more than 55 lipids for custom membranes or ready-to-use models for common physiological membranes are available[156].

The CGMD Platform is particularly tailored for investigating a wide range of biologically relevant processes, such as protein-protein interaction or lipid-protein interactions, which play critical roles in

membrane organization, signal transduction, and cellular homeostasis[156]. One of the key strengths of the CGMD Platform lies in its support for comparative studies. **Users can easily set up and simulate multiple systems with a wide range of options, enabling systematic exploration of critical biological determinants such as lipid count, lipid flip-flop, and the clusterization of lipids alongside helices or structural domains.** These comparative analyses can be performed once the simulation trajectories are downloaded from the platform. CGMD platform also facilitates reproducibility and standardization in CG simulations by integrating pre-defined protocols: advanced users can still adjust all essential parameters to meet their specific research goals. Furthermore, the web interface is designed to be intuitive and accessible (also for color blind people), ensuring that both novice and experienced users can benefit from the platform. Comprehensive "on-the-fly" pop-up documentation eradicates the need for manual script editing or advanced computational skills. To support downstream analysis, the platform incorporates also a robust analysis layer that offers visualizations and structural stability metrics, at any step of the MD simulation. This allows researchers to monitor and interpret stability-related insights directly within the web interface, even before the simulation is fully completed. Once finished, the data can be downloaded for more detailed, local analysis. In summary, our CGMD platform is a powerful, user-friendly, and fully web-based tool that simplifies the preparation, running, and analysis of CGMD simulations. **As such, it serves as an invaluable resource for studying membrane proteins, macromolecular assemblies, and their interactions with lipid environments in biologically meaningful contexts.**

3.8 Contribution of this Chapter

The structure of this chapter is intended to describe how the computational platform known as the CGMD Platform is organized. I contributed to several aspects of this web server.

My primary contribution involved the setup of the graphical user interface, developed using programming tools such as PHP and JavaScript. In addition, I assisted in building the source code required to run the underlying algorithms in the background. These algorithms enable users, through various tools, to set up and execute coarse-grained molecular dynamics simulations, as well as to reconstruct all-atom structures from coarse-grained representations.

The connection between GOMoDo web server and this platform was done by me and Rui Ribeiro[211].

Furthermore, I participated in the revision and writing of the manuscript and in the collection of data from the scientific literature. The benchmarking was carried out by me in collaboration with several members of Forschungszentrum Jülich, Germany (Zeineb Si Chaib, Giulia Rossetti), Institut Pasteur de Montevideo, Uruguay (Sergio Pantano) and the University of Verona, Italy (Carlo Alberto Rossi, Rui Ribeiro and Alejandro Giorgetti).

This chapter is an extract published in accordance with §5(6) of the doctoral regulations in the following peer-review Journal:

Alessandro Marchetto, Zeineb Si Chaib*, Carlo Alberto Rossi*, Rui Ribeiro, Sergio Pantano, Giulia Rossetti and Alejandro Giorgetti "CGMD Platform: Integrated Web Servers for the Preparation, Running, and Analysis of Coarse-Grained Molecular Dynamics Simulations." Molecules 25.24 (2020): 5934.*

Chapter 4

GPCRs oligomerization protocol

PPIs are fundamental for cellular processes, but understanding their exact structures and mechanisms is difficult[220]. Classical MD simulations provide valuable insights into PPIs, but their effectiveness in studying reversible association and dissociation processes is hindered by slow kinetics, and represent them is challenging[169, 220]. Addressing these issues is crucial for gaining atomic-level understanding of protein oligomerization pathways. To expedite these processes, CGMD simulations have been utilized[149, 151]. As discussed in Chapter 2, CGMD simulation could represent the gateway to overcome limitations in time and length scales[149]. Using this approach, it is still possible to retrieve molecular details for the protein-protein oligomer with high accuracy[149]. However, representing all metastable states and transitions between them remains enormously challenging[220]. Therefore, a common strategy is to combine CGMD simulations with enhanced sampling techniques, which further accelerates the exploration of phase space. For my simulations, I used a combination approach of CGMD and Well-tempered Metadynamics techniques[141, 149]. The GPCRs oligomerization protocol consists of two modules:

- **Well-Tempered Metadynamics:** This first step runs Well-tempered Metadynamics using two CVs. The scope of this module is to reproduce the association/dissociation mechanism typical of GPCRs in complex membranes, finding interfaces potentially in line with experiments. Free-energy calculations are also performed to analyze the convergence of the simulation and evaluate the minima extracted.
- **Unbiased Simulation:** The scope of this module is to evaluate the degree of stability of the quaternary structure of the complex. Here, the selected structure undergoes 40 μ s of simulations, and different properties are evaluated, such as: contact analysis matrix, solvent accessible surface area (SASA) of the orthosteric binding cavity or cavities of the GPCRs, and volume of the orthosteric binding cavity or cavities of the GPCRs. This module ultimately aims to evaluate the impact of the membrane on one or more helices within the GPCRs structure.

This protocol, compared to others in the literature[169, 220], leverages two distinct features: it does not require prior knowledge of the complex to be formed (therefore, the interaction sites that determine receptor-receptor interactions are not specified in the definition of the descriptors of the CVs), and it also aims to take into account the membrane environment surrounding the receptors. The ultimate goal of this protocol is to determine structures/complexes with the correct interface that can subsequently be exploited to perform kinetic and thermodynamic calculations for drugs binding to that specific complex.

4.1 First module: the Choice of the Collective Variables (CVs) for the Well-tempered Metadynamics

To accurately describe the association/dissociation pathway, is important to investigate the correct CVs to push on the slow modes of the system and, thereby, describe the entire free energy landscape (see subsection 2.8.2).

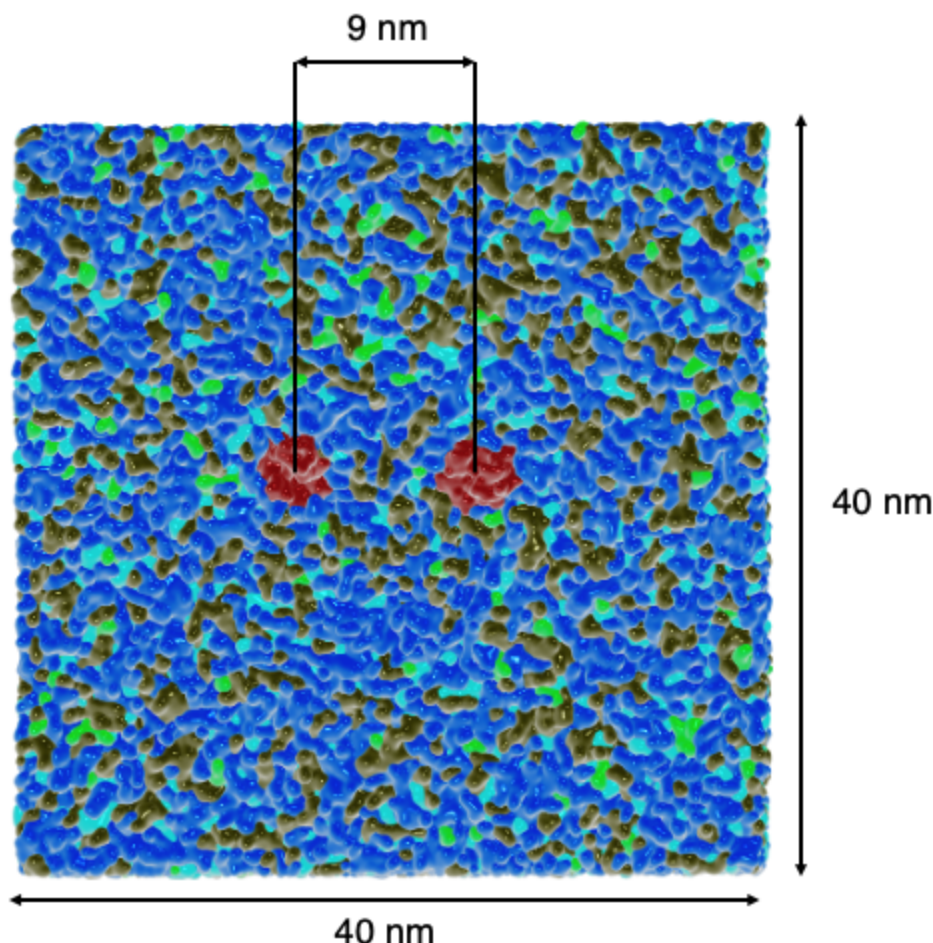


Figure 4.1: Set-up of the first CV. Proteins are depicted in red.

To do so for my simulations, I used two reaction coordinates, namely the centers of mass (COM) of the two monomer embedded in membrane as a first CV and, as second CV, the transmembrane atoms (COORDINATION) of the two embedded systems. All the simulations were run with the support of Plumed 2 plug-in patched with Gromacs[133, 148]. All the Well-tempered Metadynamics simulations consist of 10 μs , except where indicated. Additionally, as depicted in the image (Figure 4.1), the membrane measures 40 nm along the x-axis and 40 nm along the y-axis (12 nm along the z-axis). These seemingly large dimensions were necessary to prevent self-interactions among adjacent images (PBC) and to allow large diffusion of the monomers within the lipid environment. Unless specified otherwise, the lipid environment is referred to the neuronal membrane (here called CNS membrane)[221].

4.1.1 Analysis of the first CV: the distance of centers of mass

Two monomers have been placed 9 *nm* apart within a complex membrane at the start of the simulation (Figure 4.1). The distance of the centers of mass between the two membrane entities was then set up. The decision to position the monomers at this distance at the beginning of the simulation was made to avoid any potential long-range interactions between the two monomers from the beginning of the equilibration phase.

The definition of the first CV is the following:

Pipeline first CV: distance of centers of mass

```
monomerA: COM ATOMS=first_atom-last_atom
monomerB: COM ATOMS=first_atom-last_atom
comMonomers: DISTANCE ATOMS=monomerA,monomerB

UPPER_WALLS ...

ARG=comMonomers
LABEL=restraint
AT=9
KAPPA=1000

... UPPER_WALLS
```

where,

monomerA and monomerB	refer to definition of the center of mass (COM) of monomer A and monomer B, respectively. first_atom and last_atom refer to the first and last atom of the monomer, respectively,
comMonomers	indicates the distance of the COM,
UPPER WALLS	puts an upper wall on the comMonomers: the two monomers are restrained within a maximum distance of 9 nm with a harmonic potential having a constant KAPPA = 1000.

4.1.2 Analysis of the second CV: the transmembrane atoms

Upon the selection of the first CV, the selection of the second CV, i.e. transmembrane atoms, was determined using a combination of three different sources: PDBTM, Uniprot, and visual inspection via VMD[37, 222, 223]. I chose to combine these three factors to gain greater detail and accuracy in identifying the transmembrane atoms, needed for this protocol (Figure 4.2). The representation shown in Figure 4.2 displays only one helix for the sake of clarity; however, the selection was performed on all transmembrane helices of the GPCRs. Helix H8 was not included as part of this CV.

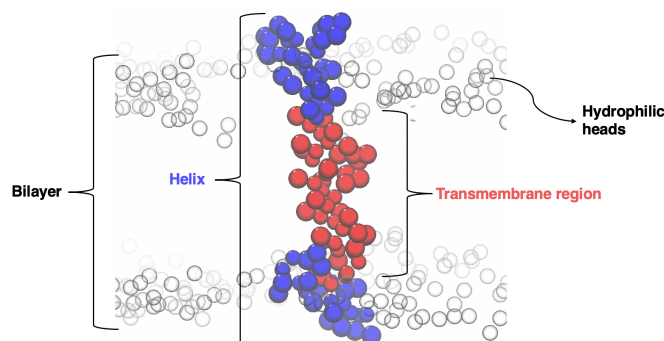


Figure 4.2: In red, the region of the membrane selected as descriptors for the CVs is shown. The remaining part of the helix (in blue) is not included. As the name suggests, the transmembrane region (in red) is the portion between the upper layer and the lower layer.

The definition of the second CV is the following:

Pipeline second CV: transmembrane atoms

```
COORDINATION ...

LABEL=coordination
NN=6
MM=12
R_0=0.9
D_0=0.0

GROUPA=H1_TM_range,H2_TM_range,H3_TM_range,H4_TM_range,H5_TM_range,
H6_TM_range,H7_TM_range
GROUPB=H1_TM_range,H2_TM_range,H3_TM_range,H4_TM_range,H5_TM_range,
H6_TM_range,H7_TM_range

... COORDINATION
```

where,

NN=6, MM=12 R_0=0.9, and D_0=0.0 refer to the parameter of the switching function,
 GROUPA indicates the transmembrane atoms of the first monomer,
 GROUPB indicates the transmembrane atoms of the second monomer.

The switching function s_{ij} obeys to the following formula:

$$s_{ij} = \frac{1 - \left(\frac{\mathbf{r}_{ij} - d_0}{r_0} \right)^n}{1 - \left(\frac{\mathbf{r}_{ij} - d_0}{r_0} \right)^m}. \quad (4.1)$$

4.1 guarantees that the COORDINATION CV has continuous derivatives. This CV has been meticulously analyzed using the DEEP-LDA algorithm[146].

4.1.3 Well-tempered Metadynamics

The Well-Tempered Metadynamics protocol I proposed utilizes the two CVs mentioned above. To calculate the value of sigma, I referred to the following formula:

$$\sigma = \frac{1}{2 * \text{unbiased range of variable}}. \quad (4.2)$$

Of course, there are several other criteria to describe sigma values, such as selecting sigma corresponding to one-third of the standard deviation of the basin during the unbiased simulation, than equation 4.2; however, values very similar to these have been observed with this criterion and/or values that were too small were unable to overcome the various barriers to explore different energy minima. Therefore, the values I chose, represent the compromise adopted in my simulations. The Well-Tempered Metadynamics protocol is the following:

Pipeline second CV: transmembrane atoms

```

METAD ...

LABEL=metad
ARG=comMonomers,coordination
SIGMA=0.58,1.203
GRID_MIN=-3,-3
GRID_MAX=10,3000
PACE=500
HEIGHT=1.2
FILE=HILLS
TEMP=315
BIASFACTOR=30
CALC_RCT
RESTART=YES

... METAD

```

where,

SIGMA	is the width of the Gaussian,
GRID_MIN	indicates the lower bound of the grid,
GRID_MAX	indicates the upper bound of the grid,
PACE	is the deposition rate of the Gaussian,
HEIGHT	is the height of the Gaussian,
TEMP	is the temperature of the system in K,
BIASFACTOR	is the bias to overcome minima, how rapidly the Gaussian height is reducing,
CALC_RCT	calculates the <i>rct</i> value necessary for the reweighting,
RESTART=YES	indicates that the simulation must restart from the point it stopped.

In this protocol, the CVs are kept *intentionally* generic to target the system's slow modes. As a result, the approach can be described as "blind" because it does not focus on specific interactions between monomers within the membrane. Instead, it accounts for a broad range of possible interactions. The rearrangement of oligomers in the membrane is influenced by several factors, including the primary sequence of the proteins, the composition of the membrane - which may affect oligomerization due to lipid diversity - the local arrangement of monomer helices, or a combination of these elements.

4.1.4 Clusterization of the structures in the basins

Once the free-energy profile has been determined, it is necessary to select the structures of a particular minimum. The selection of a minimum cannot be determined from a single frame of the simulation; instead, it must be based on a collection of frames, as each frame contributes with a different statistical weight to that minimum. To proceed with clustering, it is essential to determine an algorithm and a clustering criterion. I used the GROMACS *gmx cluster* tool to cluster trajectories using the *gromos* method, with RMSD as the criterion for including frames in a specific cluster[224]. The cut-off for including frames in the neighborhood ranges from 3.5 Å to 4.5 Å, depending on the number of output structures[133, 224]. To extract the most populous clusters, I utilized the combination of the following three criteria:

1. Minimizing the number of clusters that exclusively contain a single element;
2. The collection of frames extracted from the minima are included in a 10 or few clusters;
3. The first three clusters should include at least 70% of the extracted structures.

The first three top entries are the three most populous clusters. Following the used pipeline:

Pipeline clustering protocol

```
gmx cluster -f traj.xtc -s topol.tpr -pbc yes -cutoff 0.35 -g -sz -cl -o
-method gromos
```

where,

- pbc yes takes into account protein boundary conditions during calculation,
- g indicates to print log file to get information on the clusters,
- sz indicates the cluster size in output,
- cl outputs the centroid of each cluster as a pdb structure. The centroid is the structure with the smallest average RMSD from all other structures within the cluster,
- method gromos is the method used for clustering[224].

4.1.5 Determination of the population of each cluster and evaluation of the oligomeric interfaces

Once the frames were collected using the chosen clustering algorithm, I attempted to assess each cluster in terms of its percentage relative to the number of structures extracted in that basin. For example, if 1000 structures are isolated in the selected basin and 700 structures appear in the first cluster, that cluster alone will contain 70% of the frames in that minimum. Thus, the minimum, consisting of 1000 frames (structures), is represented by 70% in that cluster. This process was repeated for the other two clusters considered. Furthermore, for each of the three clusters, I endeavored to determine which interfaces were selected by the clustering protocol in those clusters. For each cluster, the centroid was taken and its interface was evaluated. The evaluation of the interfaces was done by using VMD[223].

4.2 Second module: Unbiased Simulation

In the second module of the protocol, I examined a variety of properties on the structures collected from the previous step. To evaluate the stability of the quaternary structure of the complex and the impact of oligomerization compared to monomers, different analyses have been conducted. These include examining contacts at the interface, annular lipid interactions, and the SASA and volume of the orthosteric/vestibular binding cavities of the GPCRs in both monomeric and dimeric states[225]. I included the vestibular binding cavity in the analysis due to its contiguous nature as an allosteric extracellular binding cavity directly connected to the orthosteric pocket of the GPCR[225]. Additionally, ligands may interact within this region, making it appealing for further comprehensive analysis, such as docking and virtual screening studies[226, 227]. All the CG unbiased simulations consist of 40 μ s: I simulated the D2 dimer in its active and inactive state, but also the D2 monomer in its active and inactive state to compare the impact of dimerization in changing SASA and volume properties. The total simulation time for the unbiased simulations is 0.16 *ms*. All the unbiased simulations were run only on the CNS membrane (see Section 4.3.1).

4.2.1 Analysis of contacts

The interactions between two membrane entities were calculated by considering all residues belonging to each helix of a given monomer and the other monomer. For example, when selecting helix 1, all transmembrane residues of this helix from a specific monomer were examined, and the distance between each of these residues and all the residues of the other monomer was calculated. This approach highlights how a particular helix belonging to a monomer interacts not with a specific helix, but rather with multiple regions of the other monomer. This relies on the fact that PPIs are unlikely to occur solely through interactions between single pairs of helices, as demonstrated by multiple papers[228, 229, 230, 231]. Instead, it is more probable that multiple regions among monomers or protomers interact with each other. Residues in the transmembrane region were chosen following the Uniprot canonical primary sequence[37]. The pipeline to calculate the contacts between monomers is the following:

Pipeline contacts between monomers

```
gmx mindist -f traj.xtc -s topol.tpr -on numcount.xvg -or mindistres.xvg -b
150000 -d 0.6 -pbc yes -printresname yes index.ndx -respertime yes
```

where,

- on numcount.xvg specifies the number of contacts between a reference (helix) and a selection (the other monomer),
- or mindistres.xvg outputs the contacts matrix: residues and the relative distance over the course of the trajectory are depicted in a matrix of contacts
- d 0.6 indicates the minimum distance to consider the contact as true, in *nm*
- respertime yes outputs the three letter amino acid code in the header of the contacts matrix.

I started calculating the contacts matrix from 150 *ns*, discarding therefore the first frames of the simulations and keeping only the equilibrated part of the trajectory.

4.2.2 Analysis of the orthosteric and vestibular binding pockets

There are several methods to identify binding pockets and their various properties: energy-based, sequence-based, and, finally, geometry-based cavity approaches[232]. The approach I used in this thesis to determine the binding pockets and to calculate the various properties related to them is the latter. For this detection scheme, I utilized MDPocket[232]. The choice of this approach and algorithm is based on the fact that membrane proteins act as dynamic bodies that adapt to various cellular stimuli[233]. Thus, the dimerization of two membrane receptors might affect the arrangement of the transmembrane helices as the monomers approaching, and as a result, the orthosteric and vestibular cavities formed by these helices may undergo modification. As described in paragraph 2.10.1, to retrieve atomistic information for a protein from a Martini CG model, it is necessary employing a backmapping algorithm[172]. After backmapping the structures, I utilized the subsequent procedure to determine properties SASA and Volume of the orthosteric and vestibular binding properties:

- **Run MDPocket:** this operation determines the binding pockets and transient binding pockets of the entire structure. At this stage, several cavities can be detected, and therefore, to remove this background "noise", a further filtering step is required. All the frames of the trajectory have been superimposed to a reference structure (first frame) to avoid artifacts in the calculations;
- **Filtering of binding pockets:** this operation circumvents the inclusion of unwanted pockets in the calculation. MDPocket provides for the *extractISOPdb.py* script (<https://github.com/Discngine/fpocket/blob/master/scripts/extractISOPdb.py>) to remove the introduction of these pockets[232]. The isovalue chosen for this script is 3;
- **Re-run MDPocket on the selected pockets:** this run MDPocket again for the second time to calculate a variety descriptors for the selected pockets.

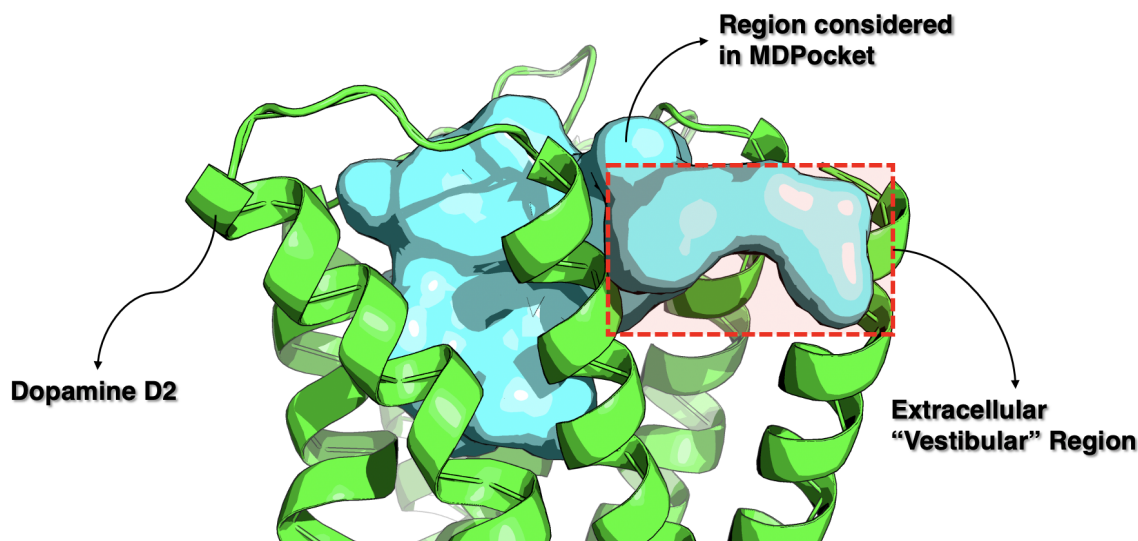


Figure 4.3: D2 monomer (in green) and the surface region considered for the MDPocket Analysis (in cyan). The red-dashed rectangle depicts the extracellular allosteric vestibular region of the D2 (see 4.2.2).

4.2.3 Analysis of putative CHOL binding spots

CHOL plays an important role in eukaryotic membranes[234]. It regulates a plethora of properties, such as fluidity, physiological functions, and, more in general, thickness and undulation[235, 236, 237]. Until 2021, 560 GPCR structures were deposited in the PDB[112]. Out of these, 152 structures have bound CHOL molecules. These 152 structures correspond to 40 unique GPCRs[238]. Some GPCRs have multiple structures with CHOL bound: the A2 possesses the highest number of CHOL-bound structures, with 39 representatives[238]. Moreover, the number of CHOL molecules varies consistently among CHOL-bound GPCRs, from 1 (the most common with 37% of the structures) to 16, as the case of GABA_B receptor[238, 239]. One of the most studied GPCRs is the β_2 receptor[238]. It has been found co-crystallized with two molecules of CHOL, which interact with four different residues of the structure: Tyr70^{2.41}, Arg151^{4.43}, Ile154^{4.46}, and Trp158^{4.50}[237]¹. This four residues represent the so-called CHOL consensus motif (CCM) for the β_2 receptor[237]. According to the CCM pattern, positions from 4.39 to 4.43 must contain a positively charged residue (either Arg or Lys), position 4.46 must have a hydrophobic residue (Ile, Val, or Leu), and position 4.50 must include a hydrophobic Trp. Additionally, the position at 2.41 must contain an aromatic residue, namely either a Phe or a Tyr[237]. The conserved nature of these interactions suggests that other class A GPCRs may share a CCM similar to the one observed for the β_2 receptor. In particular, D2 meets these four criteria, and CHOL may putatively bind to D2 at the same location as β_2 [237]. To calculate the annular CHOL around D2 in its active and inactive state, both monomer and dimer, I used MDAnalysis Python toolkit[240]. The code I implemented for running this analysis is in the Appendix in Chapter 7.

4.3 Membrane environments: from the CNS towards the CHOL Enriched membrane

One of the most critical biological roles is played by membranes, such as the maintenance of structural scaffolding, regulation of different hormones/drugs, active transport and, also, cell communication[241, 242]. The lipid bilayer membrane is a crucial self-assembled structure serving as a boundary between the extracellular environment and cytosol (Figure 4.4), but also delineates the separation among cellular compartments[242, 243]. A typical plasma membrane entails several distinct lipid species, which are modulated by the cellular machinery. Depending on the interplay within the membrane environment,

¹The second CHOL molecule presents a minor extent of interactions with the β_2 structure.

diverse lipid compositions can regulate a wide range of protein functions through shell lipid-protein interactions or modulate bilayer properties[244].

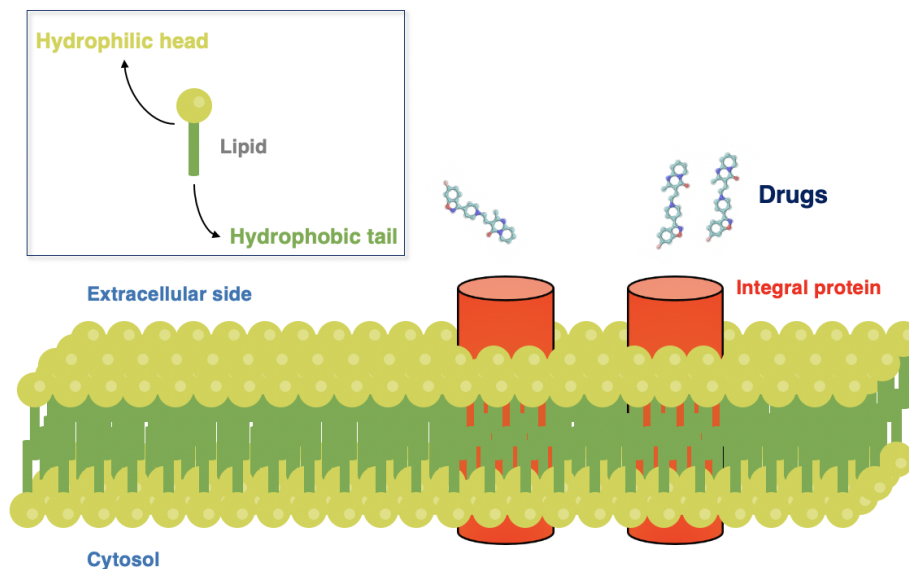


Figure 4.4: The cellular membrane is schematically composed of lipids, which consist of hydrophilic heads and hydrophobic tails. The bilayer model features an upper and lower leaflet that separate the extracellular environment from the intracellular side (cytosol)[78]. Embedded within the membrane are numerous effector proteins. Drugs bind to receptors and induce a conformational change, triggering signals within the cell.

CG simulations have become a valuable tool for studying membrane dynamics and investigating the interactions between membranes and embedded integral membrane proteins. As discussed in Chapter 2, CG simulations simplify the system by grouping multiple atoms into beads, which reduces the number of particles and consequently smooths the energy landscape[149]. This enables larger integration time steps, often resulting in faster dynamics[244]. The lipid bilayer model follows the one initially proposed by Singer and Nicolson in 1972[77, 78]. This membrane model is the key component in pharmaceuticals and biotechnology and for studying lipid dynamics, protein function, and their diffusion[243]. Here, as we will see in Chapter 5, I investigated the dimerization mechanism of D2 in two distinct membrane environments to understand how these membrane compositions can help explain different experimental outcomes. Following this idea, I computationally identified different interfaces consistent with experimental findings[245].

4.3.1 Central Nervous System (CNS) membrane

The first membrane used in my simulations contains five components: CHOL, POPC, POPE, POPS, and POPG. This membrane represents an approximation of the central nervous system (CNS) membrane as proposed in the study[221]. I approximated the lipid composition of the CNS membrane present in this study for two reasons. Firstly, not all lipids referenced in the study by Fitzner *et al.* [221] have been parameterized within the Martini 3 framework[149, 246]. Secondly, cellular membranes contain a wide variety of phospholipids and sterols[155], and representing all of them would have made the simulations exceedingly complex to manage in terms of time scale, also for a CG model and supercomputers. Following the neuronal membrane composition defined as **CNS Membrane** in this thesis:

Lipid name	Upper Leaflet (%)	Lower Leaflet (%)
POPC	36.0	36.0
CHOL	31.0	31.0
POPE	22.0	22.0
POPS	8.0	8.0
POPG	3.0	3.0

Table 4.1: Lipid composition of the Upper and Lower Leaflets of the CNS membrane.

4.3.2 CHOL Enriched Membrane

CHOL is an ubiquitous component in eukaryotic membranes present to a large extent within the plasma membrane[155, 247, 248]. It also represents a main component of myelin in the brain[248]. The percentage of CHOL within the human brain can fluctuate significantly, ranging from 20% to 80% in synaptic vesicles[155, 249]. Moreover, alteration in brain CHOL metabolism have been linked to pathological conditions[250]. CHOL is also important because it can influence a plethora of processes, such as receptor dimerization, ligand binding, and signal transduction[169, 251]. Besides, as mentioned in Chapter 1 and in the next Chapter 5, CHOL is one of the key participants in lipid rafts[82]. Within these lipid rafts, CHOL content can range from three to five times higher than in surrounding membranes and this might have an impact on several biological mechanisms[248]. Collectively, for these reasons, I decided to simulate some of the systems in the following membrane composition, defined as **CHOL Enriched**:

Lipid name	Upper Leaflet (%)	Lower Leaflet (%)
POPC	15.2	29.3
PAPC	11.0	10.2
PAPE	22.9	15.9
PAPS	16.2	-
CHOL	44.5	44.5

Table 4.2: Lipid composition of the Upper and Lower Leaflets of the CHOL Enriched membrane.

4.4 Current Limitations of the protocol: optimizing the modeling in the next future

One of the most debated topics in the field of MD concerning the enhanced sampling techniques is the convergence. The convergence of a simulation can be assessed via two key factors. The first is the rapid diffusion of the system across its phase space: according to the estimated slow motions (the reaction coordinates the researcher intend to push the system on), the system should be able to reach state A (the initial state) and state B (the final state) multiple times during the simulation, thereby ensuring efficient sampling. The second factor involves reconstructing the total FES in a block-wise manner, the so-called block-average analysis: in detail, the simulation is divided into predefined equal blocks, and the deviation among the resulting FES reconstructions is eventually evaluated. The underlying assumption is that the final reconstructed curve should display a hyperbolic trend with a plateau at convergence. However, these criteria are generally evaluated under highly controlled conditions, such as studies on very simple systems like alanine dipeptide in water environment or small proteins such as chignolin[252, 253]. In more complex scenarios, such as the present case of a biological system composed of two monomers embedded in a large membrane in this thesis, achieving convergence remains extremely highly challenging due to the astronomical variability of configurations of the single elements involved during the simulation (protein, solvent, membrane lipids, ...). Consequently, the profiles presented in the following chapter should be regarded as *partial descriptions* that have not yet fully reached convergence.

To address this limitation, together with my colleagues Simone Mariani, Simone Albani and Federico Natale, we are working to optimize the dimerization process computationally in order to observe convergence and obtain a more reliable free-energy profile. Our current strategy involves employing 10 multiple walkers per simulation and considering hybrid approaches - for example, using umbrella sampling to enforce specific configurations of the system (such as at an interface experimentally found)

and subsequently applying another enhanced sampling technique, such as Well-Tempered Metadynamics or the more recent OPES method, to achieve the desired free-energy landscape[141, 147]. These simulations will be extended beyond 10 μ s, possibly reaching the millisecond scale[169]. The results of this work will be detailed in our forthcoming publications.

4.5 Contribution of this Chapter

My contribution to this chapter consisted of setting up the computational protocol whose results are presented in the following chapter. In this work, I first aimed to identify the most appropriate CVs to describe the system's transition from an associated to a dissociated state. To achieve this, I employed a data-driven classifier, Deep-LDA, which, given a set of descriptors derived from literature data and prior knowledge of the system, is able to identify the direction in feature space that best discriminates between relevant conformational states[146]. Based on this approach, I extracted two CVs from the model: the distance between the centers of mass of the two monomers and the number of transmembrane contacts. Once the collective variables were defined, I performed Well-Tempered Metadynamics simulations on the high-performance computing infrastructure at Forschungszentrum Jülich in order to investigate the mechanism of dimer association and dissociation, and to identify the most stable dimeric state(s). To support the interpretation of these results, I also interacted with several leading researchers at the Istituto Italiano di Tecnologia (IIT), including Michele Parrinello, Andrea Rizzi, and Valerio Rizzi. After identifying the relevant structures, I extended the simulations for an additional 40 μ s to assess the stability of the resulting complex. In addition to analyzing intermolecular contacts, I evaluated key structural properties such as the solvent-accessible surface area (SASA) and the pocket volume of the cavities within each monomer in the dimer. In parallel, together with members of the INM-9 group at Forschungszentrum Jülich, I contributed to the design and construction of a membrane model intended to reproduce the dimerization mechanism in a lipid-raft-mimicking environment. In particular, I was actively involved in the setup of this membrane system, ensuring that its lipid composition and physical properties were appropriate to realistically capture biologically relevant oligomerization behavior. Overall, this work represents my main contribution to the development, implementation, and application of the computational protocol for studying the oligomerization process, together with the collaborative efforts that supported its refinement and validation.

Chapter 5

Results

This part of the thesis presents the results obtained with the GPCRs oligomerization protocol seen in Chapter 4. As discussed in Chapter 1, investigating PPIs, particularly the process of dimerization, is critically important in various biological processes, including signal transduction, enzyme activation, and the recruitment of transcriptional cofactors[254, 255]. These interactions play a pivotal role in cellular functions and pathogenic pathways[254]. Research shows that abnormal PPIs are linked to several diseases, including cancer, infectious diseases, and also NDs. As a result, targeting PPIs has emerged as a central strategy in drug development[256]. For this reason, identifying the correct oligomer is of paramount importance for accurate characterization, and my protocol is specifically designed to accomplish this. Gaining deeper insights into the molecular mechanisms that drive protein dimerization could unveil new horizons for drug design, offering a new dawn for innovative therapies[228, 254, 256].

5.1 Free Energy Profiles

To illustrate the dimerization mechanism of the D2 and A2 receptors, I used the WTMETAD enhanced sampling technique[141, 142] employing the two CVs discussed in Chapter 4, namely the distance of center of mass (d_{COM}) and the number of transmembrane helices contacts (COORDINATION), to push on the slow modes of the system. Once the system has reached convergence, the free energy as a function of the CVs can be derived from the bias potential accumulated during the WTMETAD run [141, 142, 257]. The conformations sampled throughout the simulation offer valuable insights into the unbiased state of the system, enabling the calculation of unbiased observables through a procedure known as the *reweighting process*. Additionally, to better characterize the associated and dissociated states, the free energy landscape has been remapped using an alternative CV instead of the one used for the biased simulations, specifically the Root-Mean Square Deviation (RMSD). This was done to determine a better visualization and clear distinction between associate and dissociate states. The RMSD was calculated using the structure from the most populous cluster extracted after the 10 μs WTMETAD simulation as a reference structure. In the upcoming plots depicting the Free Energy Surface (FES), lower absolute values of Free Energy correspond to a more energetically favorable state.

5.1.1 Free Energy Surface: D2 homodimer in its inactive state in the CNS Membrane

From the FES profile, three distinct states can be identified, which I have designated as *A*, *B*, and *C*. In state *A*, the two monomers are dissociated, with a center-of-mass distance (d_{COM}) of approximately 9 nm (around 8 kJ/mol). At this separation, long-range interactions between the monomers are negligible, and their interaction is primarily limited with the surrounding membrane environment and solvent. In state *B* (around 48 kJ/mol), the monomers begin to approach each other, initiating dimer formation. Finally, in state *C* (around 48 kJ/mol), the TM6-TM6 interface is fully established. The structure obtained from the clustering analysis, as detailed in Chapter 4, aligns with the experimental data reported by Pulido *et al.*, representing *one* of the possible D2 interface assortment corresponding to the inactive state of the monomers[53, 92]. As noted from the FES profile, it is evident that state *A* is energetically more favorable than the other two states *B* and *C*. This observation is not unexpected, as explained in the paper by Carli and colleagues[67], where it is stated that when the D2 receptor binds to an antagonist, i.e., when the receptor is inactivated, the proportion of the D2 population in the

monomeric state is favored over the dimeric state (with a ratio of 7:3 in favour of the monomeric state). The dynamic equilibrium between the associate and dissociate states illustrated via my simulations, corroborates the idea of the transient dimerization discussed in Chapter 1.

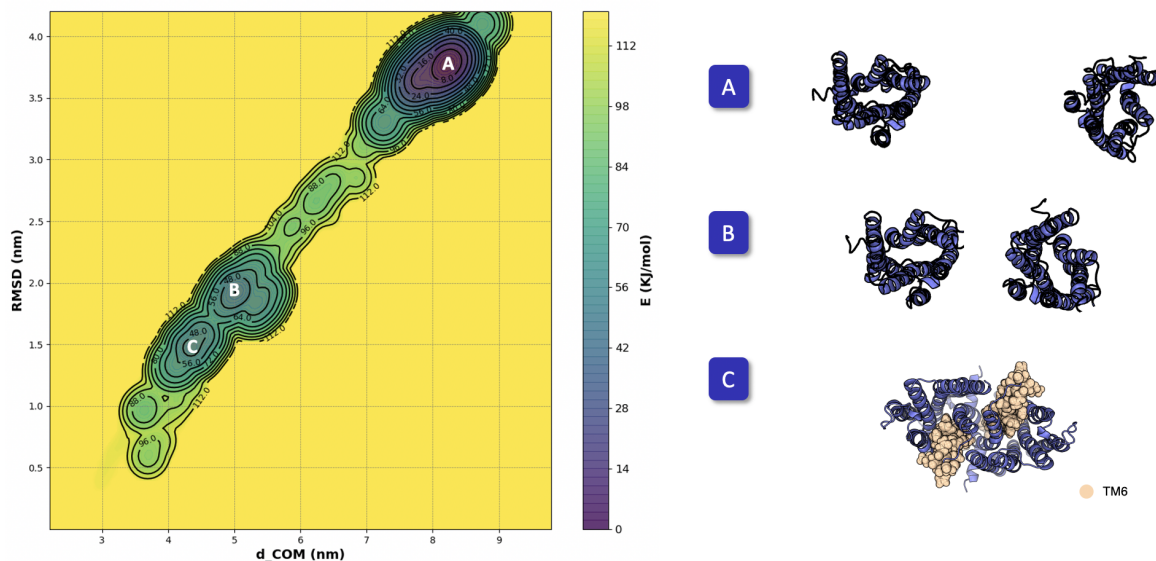


Figure 5.1: FES of the D2 homodimer in its inactive state.

5.1.2 Free Energy Surface: D2 homodimer in its active state in the CNS Membrane

From the FES profile, three distinct states can be identified, which I have labeled as *A*, *B*, and *C*. In states *A* and *B* (about 20 kJ/mol), the two monomers are dissociated, with center-of-mass distances (d_{COM}) of approximately 9 nm and 8 nm, respectively. These two states are equally energetically favorable. Eventually, in state *C* (about 60 kJ/mol), the TM4-TM4 interface is fully formed, consistent with experimental findings[55, 258, 245]. As observed in the FES profile, states *A* and *B* are more energetically favorable than state *C*. This is not surprising, as noted in the studies by Carli and colleagues[67] and Kasai *et al.*[36], where they suggest that when the D2 receptor binds to an agonist, i.e., when the receptor is activated, the dimer tends to be more stable than its inactive counterpart, resulting in a ratio of almost 50%-50% (5.5:4.5 in favour of the monomeric state).

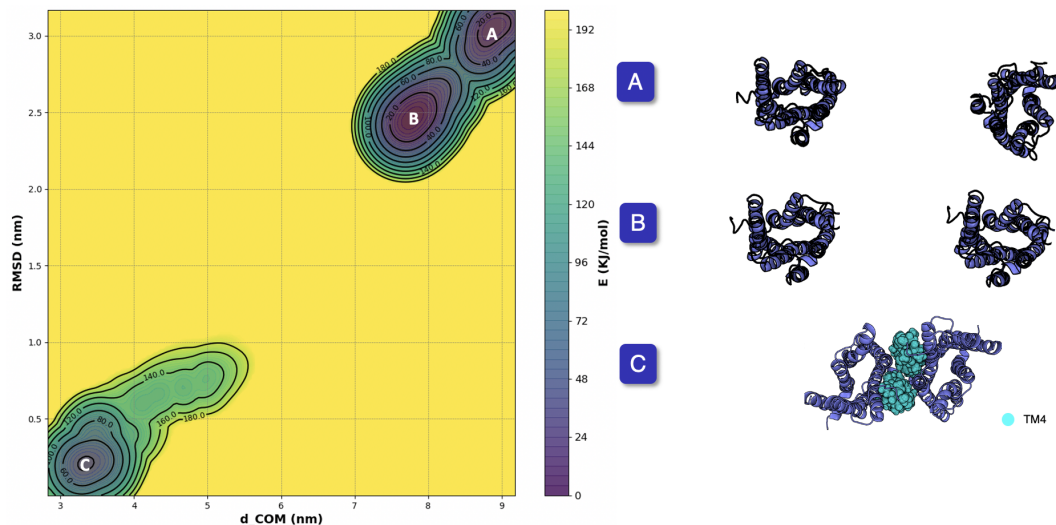


Figure 5.2: FES of the D2 homodimer in its active state.

5.2 Analysis of the D2 dimer in its active state

In this section, I will analyze the stability of the interface during the 40 μs unbiased simulation and the contacts within the dimer interface. Determining the correct interface and assessing their stability is crucial for two main reasons. First, it significantly enhances the cellular signaling system by amplifying various pathways[228]. Second, identifying the correct receptor PPI allosterism is essential for a deeper structural and mechanical understanding, and this is the baseline for advance future drug design studies[228]. The computational evaluation of the stability of the receptor-receptor interface is conducted through three distinct analyses: the examination of interactions between the single helix with the opposing monomer, the assessment of the potential interface using a contact matrix, and visual inspection utilizing VMD[223]. The analysis will be divided for the D2 in its active and inactive state.

5.2.1 Analysis of interactions

For the calculation of contacts for the D2 dimer in its active state I followed the pipeline in Section 4.2.1.

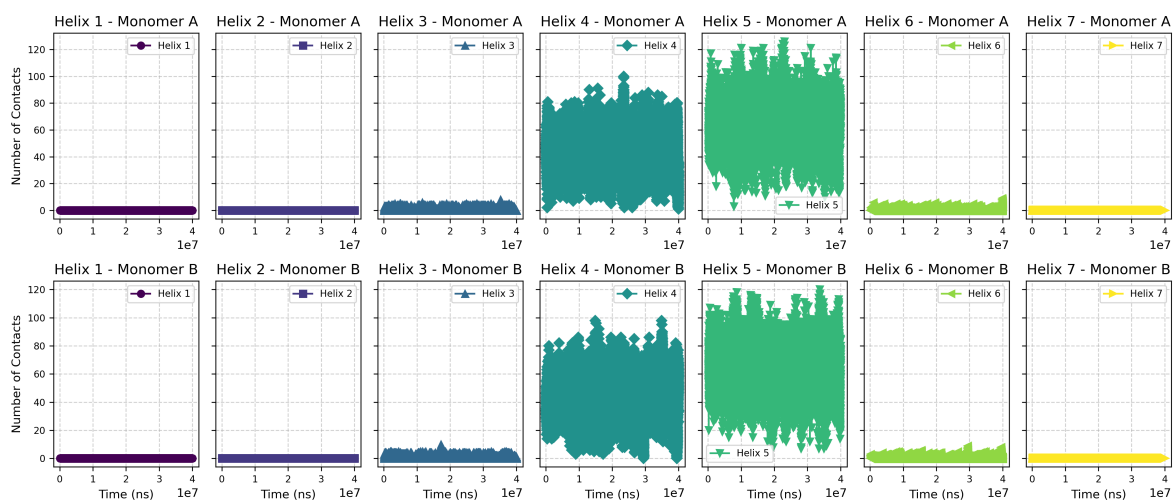


Figure 5.3: Contacts within the interface for the D2 in its active state.

As highlighted in Figures 5.3 (previous page) contacts at the TM4 and TM5 helices remain stable throughout the trajectory. Other interactions are formed - such as the one between TM3 and TM6 - but they are negligible.

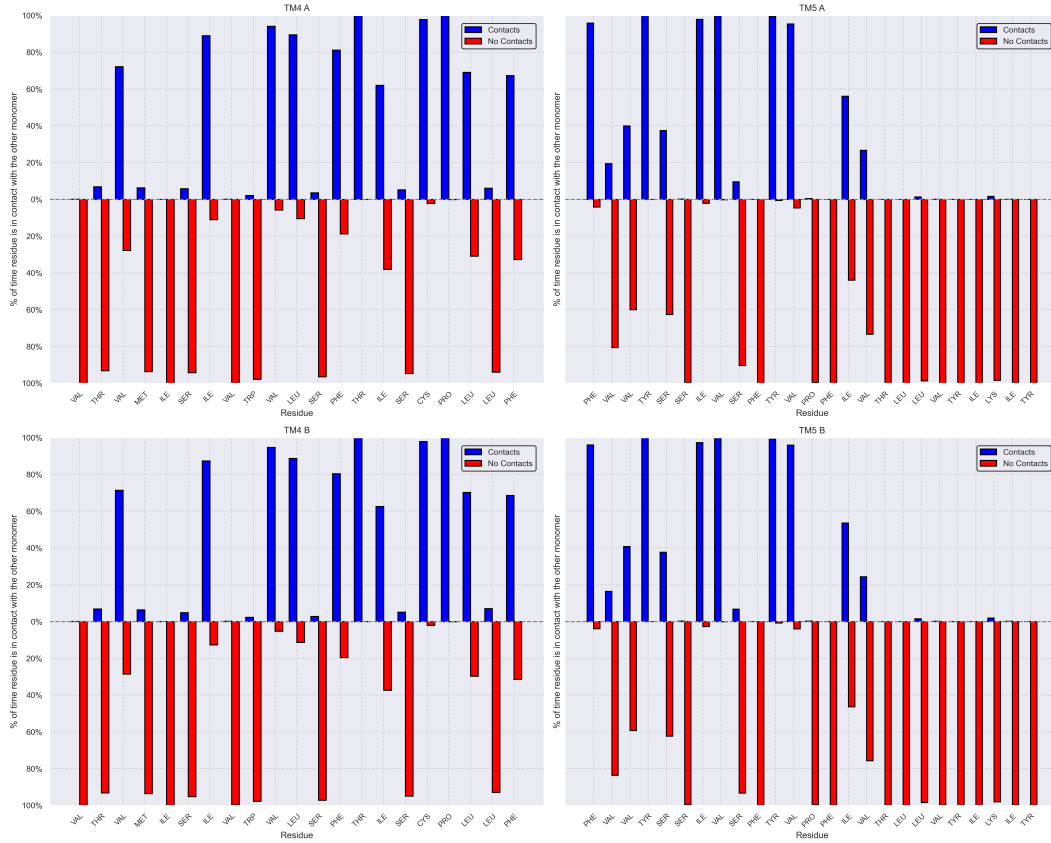


Figure 5.4: Distribution of contacts/no contacts for TM4 and TM5 in the D2 in its active state. Contacts with a cutoff of ≤ 0.6 nm are represented in blue, while those above this threshold are depicted in red.

Figure 5.4 shows the distribution of contacts across the helices TM4 and TM5, the two main characters in the D2 dimer interface. On the x-axis, residues belonging to the transmembrane region of the helix are depicted; on the y-axis, the % of the time the specific residue on the x-axis belonging to one helix spent in contact with the opposite monomer. A residue may gain or lose contact with its opposite partner throughout the course of the simulation. This analysis delineates a crucial aspect as it identifies potential actors involved in the stabilization of the complex. Contacts between TM4/5 belonging to monomer A and the opposite monomer (denoted as Residue TM4/5_A - monomer B) and TM4/5 belonging to monomer B and the opposite partner (denoted as Residue TM4/5_B - monomer A) are summarized in table 5.1. As evidenced by the table 5.1, the preceding plots 5.3, 5.4, and the image 5.5 in next page, contacts not only remain constant throughout the overall 40 μ s trajectory, but they also exhibit near symmetry. Collectively, these results indicate that the structure remains stable over the course of the simulation. This behaviour is expected when the structure of the homodimer consists of two active state monomers[36, 61, 67]. Additionally, the residues computationally identified within the interface by Bueschbell et al.[259] through the calculation of Δ SASA¹ were also detected by my protocol. Their analysis was based on the 6VMS Cryo-EM structure (resolution 3.8 Å)², whereas my simulations employed the 7JVR structure[43]. Specifically, the residues Val154^{4.44}, Ser157^{4.47}, Ile158^{4.48}, Val161^{4.51}, Phe164^{4.54}, Thr165^{4.55}, Pro169^{4.59}, Tyr192^{5.41}, Ile195^{5.44}, Val196^{5.45}, Val200^{5.49}, Val204^{5.53}, and Leu207^{5.56} were consistently captured[259]. In this array of residues, amino acids belonging to loops were excluded. Eventually, contacts between TM4 helices and the opposite monomers are distributed across the transmembrane region of TM4. In contrast, for TM5, the contacts are more localized at the N-terminal portion of the transmembrane helix. This suggests the crucial role of TM4 in forming and maintaining the scaffolding of the homodimer, specifically through the TM4-TM5/TM4-TM5 interface, as shown in various papers[55, 245, 259, 260].

¹ Δ SASA is defined as the difference in solvent-accessible surface area between the monomeric and dimeric receptor states.

²The 6VMS Cryo-EM structure has a resolution of 3.8 Å, approximately 1 Å lower than the 7JVR structure used in this work.

Residue TM4_A - monomer B	Note	Residue TM4_B - monomer A	Note	Residue TM5_A - monomer B	Note	Residue TM5_B - monomer A	Note
Thr153 ^{4.43}	★	Thr153 ^{4.43}	★	Phe189 ^{5.38}	◆	Phe189 ^{5.38}	◆
Val154 ^{4.44}	◆	Val154 ^{4.44}	◆	Val190 ^{5.39}	★	Val190 ^{5.39}	★
Met155 ^{4.45}	★	Met155 ^{4.45}	★	Val191 ^{5.40}	◆	Val191 ^{5.40}	◆
Ser157 ^{4.47}	★	Ser157 ^{4.47}	★	Tyr192 ^{5.41}	★	Tyr192 ^{5.41}	★
Ile158 ^{4.48}	◆	Ile158 ^{4.48}	◆	Ser193 ^{5.42}	★	Tyr193 ^{5.42}	★
Trp160 ^{4.50}	★	Val159 ^{4.49}	★	Ile195 ^{5.44}	★	Ile195 ^{5.44}	★
Val161 ^{4.51}	◆	Trp160 ^{4.50}	★	Val196 ^{5.45}	★	Val196 ^{5.45}	★
Leu162 ^{4.52}	◆	Val161 ^{4.51}	◆	Ser197 ^{5.46}	◆	Ser197 ^{5.46}	◆
Ser163 ^{4.53}	★	Leu162 ^{4.52}	◆	Tyr199 ^{5.48}	★	Tyr199 ^{5.48}	★
Phe164 ^{4.54}	◆	Ser163 ^{4.53}	★	Val200 ^{5.49}	★	Val200 ^{5.49}	★
Thr165 ^{4.55}	◆	Phe164 ^{4.54}	◆	Ile203 ^{5.52}	★	Ile203 ^{5.52}	★
Ile166 ^{4.56}	◆	Thr165 ^{4.55}	◆	Val204 ^{5.53}	★	Val204 ^{5.53}	★
Ser167 ^{4.57}	★	Ile166 ^{4.56}	◆	Leu207 ^{5.56}	◆	Leu207 ^{5.56}	◆
Cys168 ^{4.58}	◆	Ser167 ^{4.57}	★	Lys211 ^{5.60}	◆	Lys211 ^{5.60}	◆
Pro169 ^{4.59}	◆	Cys168 ^{4.58}	◆				
Leu170 ^{4.60}	◆	Pro169 ^{4.59}	◆				
Leu171 ^{4.61}	◆	Leu170 ^{4.60}	◆				
Phe172 ^{4.62}	◆	Leu171 ^{4.61}	◆				
		Phe172 ^{4.62}	◆				

Table 5.1: Contacts among TM4 and TM5 helices and their opposite monomers. ★ denotes a frequency $\leq$ than 20%, i.e. the time spent by a specific residue in contact with the opposite monomer during the simulation; ◆ otherwise. Only the amino acids that engage in interactions with the opposite partner are present in this table. Loops are excluded in this analysis.

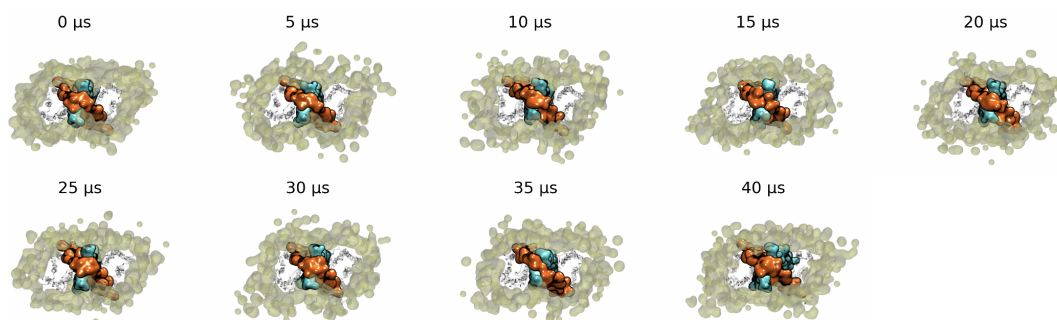


Figure 5.5: Contacts for TM4 and TM5 in the D2 in its active state depicted every 5 μ s. TM4 is depicted with a cyan surface, TM5 with an orange surface instead. The yellow surface represents the surrounding 1 nm annular lipid environment. Backbone structure is depicted in white. Only TM4 and TM5 helices are shown for the sake of clarity.

5.2.2 D2 Dimer: my Protocol vs. AlphaFold 3 Predictions

Once the D2 dimer structure in its active state was determined, I compared it with the structure predicted by AlphaFold 3[109]. AlphaFold 3 consistently predicts the same dimer interface TM4-TM5/TM4-TM5 as identified through my protocol. After superimposing the two model structures, the resulting RMSD value is less than 2.3 Å. The primary differences between the two structures are observed in the extra/intracellular loop regions and the terminal part of helix 6 (Figure 5.6). This difference is likely due to the fact that the training dataset used for AlphaFold’s neural network largely consists of inactive structures[261].



Figure 5.6: Superimposition of the D2 structure found with my protocol (dark lime green) and the D2 structure predicted by AlphaFold 3 (dark magenta)[109].

5.3 Analysis of CHOL binding spots in the dimer

Through this analysis, I quantified the CHOL lipid count surrounding the dimer. The cut-off used for taking into account these lipid interactions is 5 Å. Interactions were collected every 100 *ns*. All the CHOL molecules below the cut-off range, at each iteration, was counted for the lipid count calculation. The analysis, with its code found in the Appendix in chapter 7, aims to physically count the presence of CHOL (all its moieties are taken into account) around a specific amino acid. Therefore, an amino acid can be in contact with multiple CHOL moieties and, also, with multiple CHOL molecules simultaneously. Subsequently, at the end of the calculation, all interactions are summed, providing the heatmap in Figure 5.7. The range of helices is the same as that selected for the analysis conducted in Section 5.2.1. From this analysis, it is evident that TM1 is in greater contact with the sterol during the trajectory. This behavior is not unexpected, as TM1 appears to be more kinked towards the membranous environment compared to the other helices, such as TM4, helix known to be part of the D2 interface[55]. Additionally, helices TM4 and TM7 exhibit the second and third highest sterol counts, respectively. It is surprising how TM4, in particular, being involved at the interface, still maintains a higher sterol count despite the interactions with the other partner.

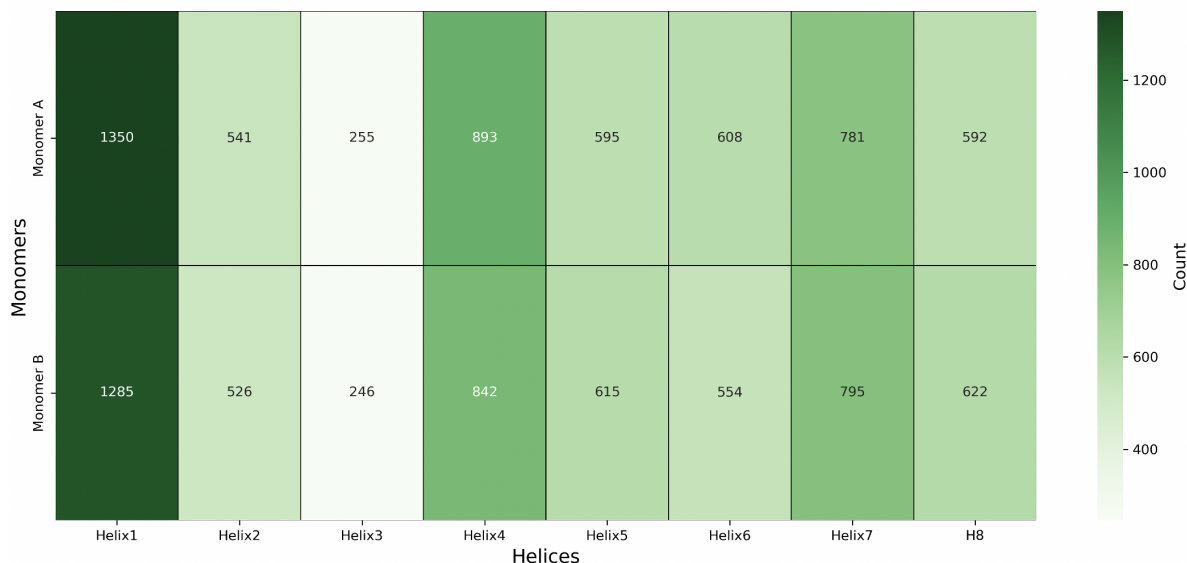


Figure 5.7: CHOL Lipid count for the two monomers in the D2 dimer in its active state.

If we examine in detail the amino acids interacting on average with CHOL during the simulation, we can observe that various interactions are established in TM1, TM4, and TM7 - the helices that exhibit the highest lipid count and are thus represented in the plot in figure 5.8 - as well as in other regions of the dimer.

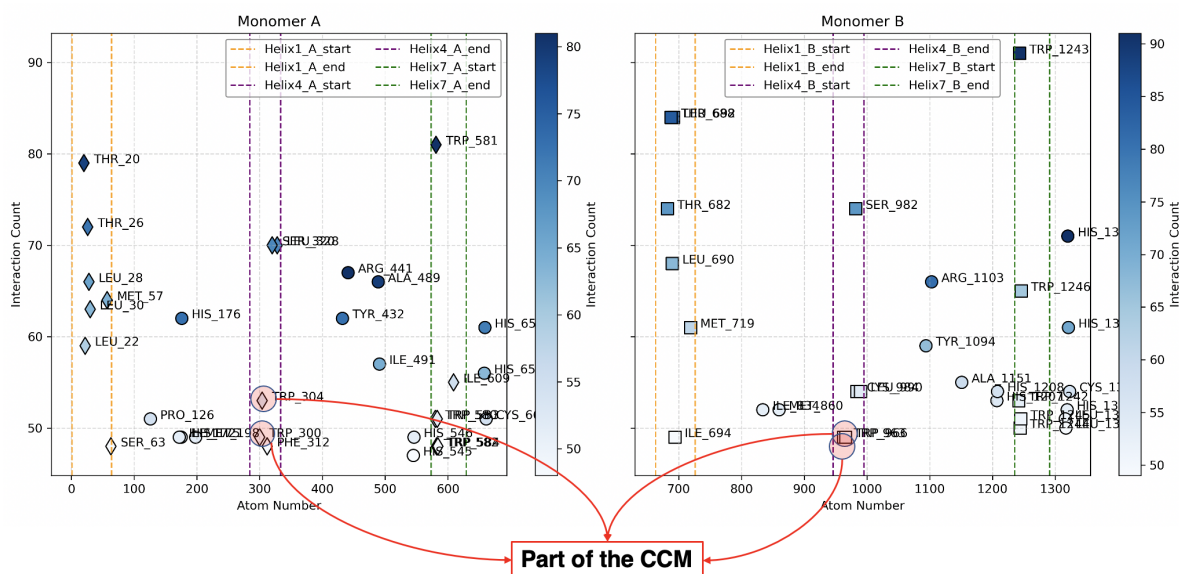


Figure 5.8: CHOL Lipid count for the D2 dimer in its active state. Red circles denote the Trp160^{4.50}, the residue belonging to the CCM. The analysis is divided for Monomer A (left) and Monomer B (right).

The interactions between monomer A and monomer B are asymmetrical, consistently with the intrinsic motion of lipids within the membrane. Regarding monomer A, the most prominent interactions occur at the levels of TM1, TM4, TM5, TM6, TM7, and H8, and the same behavior is reflected in monomer B. Furthermore, as mentioned in section 4.2.3, among the interactions observed at the level of TM4, we also find Trp160^{4.50}, a residue that is part of the CCM of D2. No other residues participating in the CCM are detected. In plot 5.8 diamonds and squares denote atoms belonging to helices TM1, TM4, and TM7, in monomer A and monomer B, respectively.

The most representative dimer structure was eventually backmapped via CG2AT2 tool[172] from the 40 μ s CG simulation. CHARMM-GUI was utilized to prepare the all-atom system, run the mini-

mization, and conduct the NVT and NPT ensembles, as well as the production run[163]. CHARMM force field was employed[178].

5.3.1 CHOL binding spots found with all-atom simulations

To confirm CHOL hotspots found in the CG simulation, I conducted 1 μs all-atom simulation of the D2 receptor dimer in its active state, using the identical CNS membrane composition employed in the CG simulations.

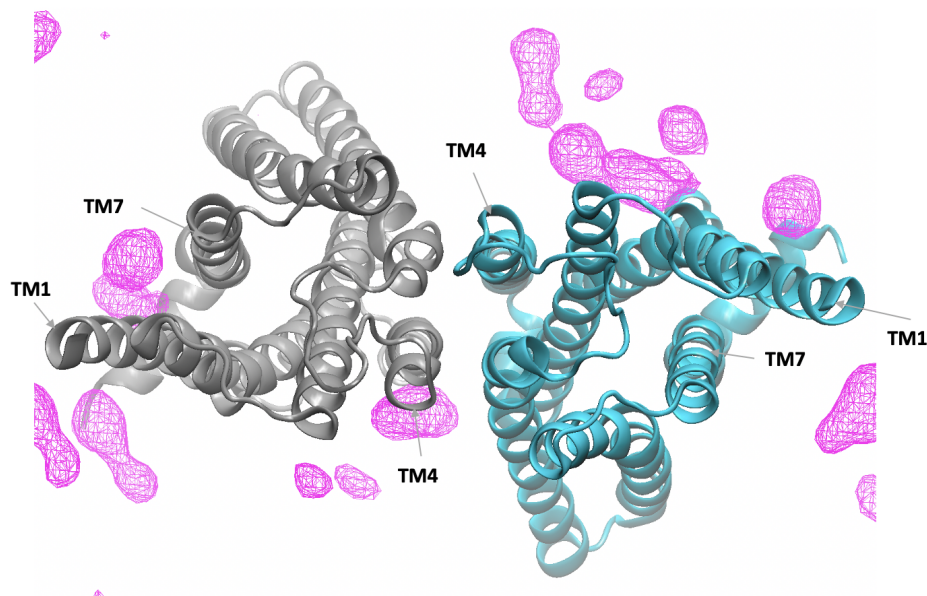


Figure 5.9: Average CHOL hotspots (magenta) identified after 1 μs all-atom simulation. Monomer A is shown in light cyan, and Monomer B in silver.

Figure 5.9 reveals distinct CHOL hotspot in proximity (5 Å) to TM1/TM2, and TM4 of monomer A (in light cyan). Additionally, hotspots were identified along TM4 in monomer B (in silver). Hotspots were also observed around TM7 and upon the amphiphilic H8 helices for both protomers. Collectively, simulations conducted using both CG and all-atom suggest that multiple putative CHOL allosteric binding sites may be located in correspondance of the TM1, TM4, and TM7 regions (Figure 5.9). Nevertheless, since no X-Ray or Cryo-EM structure has been experimentally determined so far for the D2 receptor bound to a CHOL molecule in the dimer state (as well for the D2 receptor in its monomer state), the role of CHOL in D2 can be only computationally postulated. In my simulations, the dimer in its active state, the presence of CHOL alongside TM1, TM4, and TM7 helices might prevent the protomers undergoing a rotation and thereby keeping stable the scaffolding over the course of the simulations (both CG and all-atom). Further experimental results are required to confirm the role of CHOL in the D2 dimer, and how impactful the depletion of CHOL might be in the stabilization and regulation of the dimeric structure in the dimerization process. Experimentally, determining the function of CHOL in GPCRs still poses a significant challenge, though the role of it in some class A GPCRs has been delineated[251, 262]. For example, the oligomerization of the 5HT1_A receptor has been extensively investigated[262]. Time-resolved fluorescence anisotropy studies in live cells have shown that 5HT1_A receptors form constitutive oligomers independently from the activation of the receptor driven by agonist modulation[262]. Interestingly, CHOL depletion was observed to increase the number of 5HT1_A oligomers[251, 262]. However, employing homo-FRET and FLIM techniques[251, 263], ascertained the existence of constitutive oligomers, but at the same time found that simultaneous CHOL depletion and antagonist treatment reduce the oligomer population, while agonist stimulation increases it[251]. These discrepancies may stem from different techniques capturing distinct populations of receptor oligomers[251]. CHOL has also been shown to influence the oligomerization and signaling of other receptors, such as chemokine receptors, particularly when they are embedded in lipid rafts microdomains[251].

5.4 Analysis of CHOL binding spots in the monomer

In contrast, for the D2 monomer in its active state, despite the kinking of TM1 towards the membrane environment, the lipid count in my simulations is higher for TM4. This might indicate that TM4 possess a greater propensity to interact with CHOL (Figure 5.10), given its high hydrophobicity (71.43%, sequence TM4: **VTVMISIVWVLSFTISCPLLF**[37]) content.

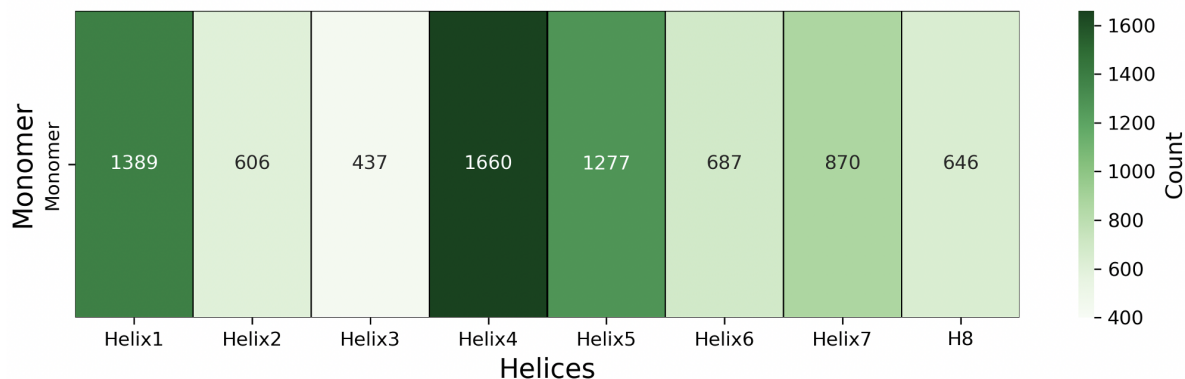


Figure 5.10: CHOL Lipid count for the D2 monomer in its active state.

In the monomeric state, the TM4 and TM5 helices acquire novel interactions in contrast with the dimer as they no longer interact with the other partner. The TM4 helix, in particular, gains several additional interactions, and here, as occurred for the dimer, the interaction with one residue belonging to the CCM (Trp160^{4.50}) is observed. No other residues participating in the CCM are detected. Also TM7 and H8 helices exhibit a similar pattern of contacts to that of the dimeric state: as they are not involved in dimerization, this behavior is expected.

5.4.1 Volume changes upon D2 dimer formation in the active state

To computationally assess the impact of dimerization on the identified model at structural level, I calculated the volume of the orthosteric cavity as well as the combined volume of the orthosteric and vestibular cavities. A similar analysis was performed for the SASA (Figures 5.11 and 5.12). This analysis has been conducted for both monomer and dimer in the region indicated in cyan surface in Figure 4.3.

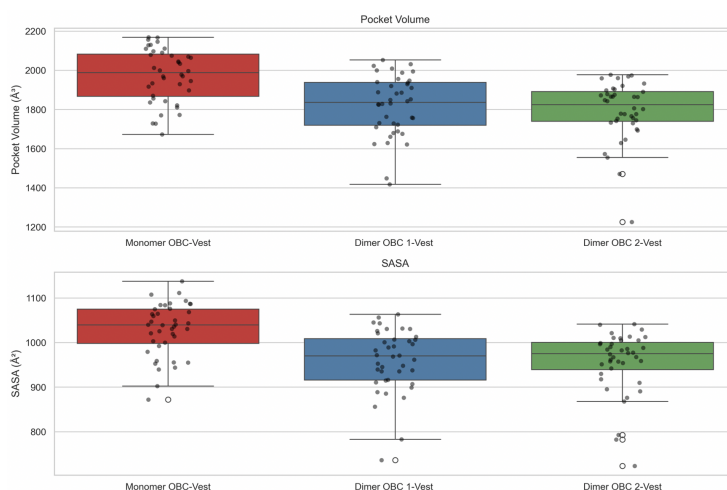


Figure 5.11: Boxplot showing the Volume and SASA of the orthosteric and vestibular binding cavities of the monomer and the D2 dimer in its active state. "Monomer OBC-Vest" refers to the monomer orthosteric binding cavity combined with the vestibular binding cavity. "Dimer OBC 1-Vest" and "Dimer OBC 2-Vest" refer to the dimer orthosteric binding cavity combined with the vestibular binding cavity for cavity 1 and cavity 2, respectively. Dots represent values calculated using MDPocket[232].

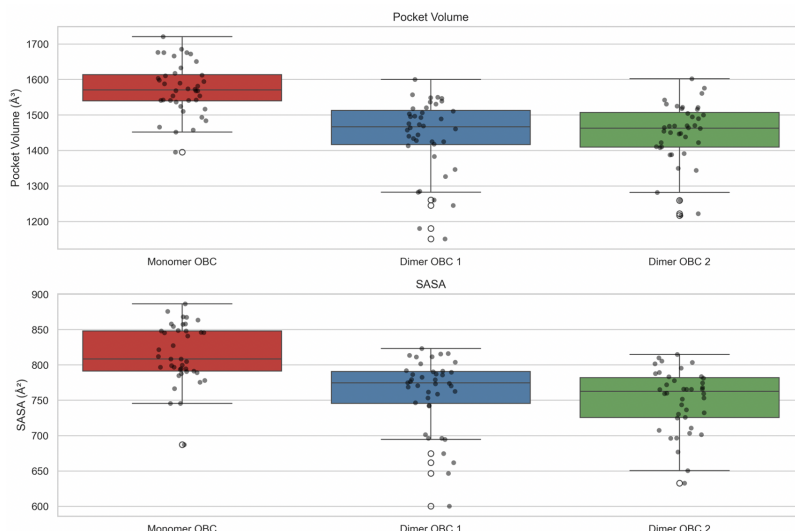


Figure 5.12: Boxplot showing the Volume and SASA of the orthosteric binding cavities of the monomer and the D2 dimer in its active state. "Monomer OBC" refers to the monomer orthosteric binding cavity. "Dimer OBC 1" and "Dimer OBC 2" refer to the dimer orthosteric binding cavities for cavity 1 and cavity 2, respectively. Dots represent values calculated using MDPocket[232].

The data presented in graphs 5.11 and 5.12 show that the monomer (in red) exhibits higher values for both volume and SASA compared to the individual cavities of the dimers (in blue and green). This difference can be attributed to the fact that hydrophobic interactions between the helices at the interface lead to a contraction of the cavities, resulting in smaller values of both Volume and SASA. This contraction might significantly affect drug binding and future design of pharmacophores, as it reduces the available space for identifying compounds that can bind to the D2 receptor in that pocket. Data were collected over 40 μs of trajectory, with each microsecond being analyzed. Each of the 40 structures was backmapped using CG2AT2[172] and evaluated with MDPocket[232]. The backmapped structures were aligned to the atomistic references structure. Monomer is not showed, but the same procedure was followed. This alignment has been utilized to accurately reposition amine bonds within the correct plane (Figure 5.13)[172].

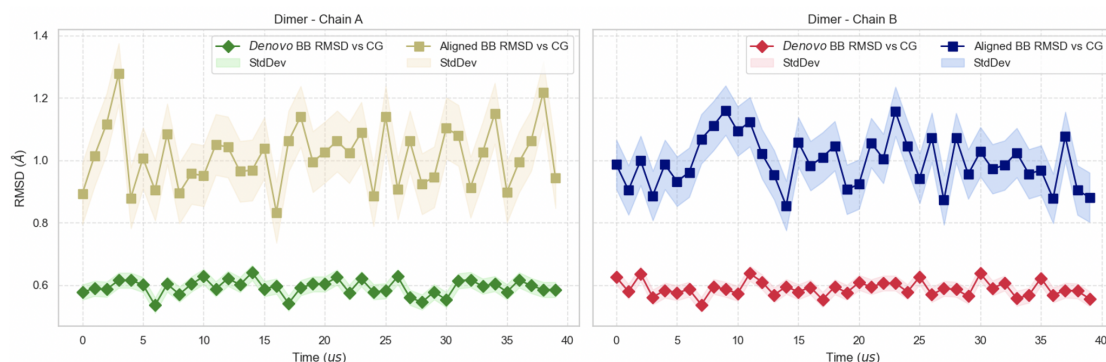


Figure 5.13: RMSD Value calculated between the *de novo* BB backmapped atomistic structure and the CG structure and the Aligned BB backmapped atomistic structure and the CG structure.

To determine if there are significant differences between the distributions of the two independent groups (monomer and dimer simulations), I used The Mann-Whitney (U) test, a non-parametric test[264]. I used this non-parametric test because it does not assume that the data follows a normal distribution, as many biological data[265, 266, 267]. Instead, it assesses whether one group generally exhibits higher or lower values compared to the other[264].

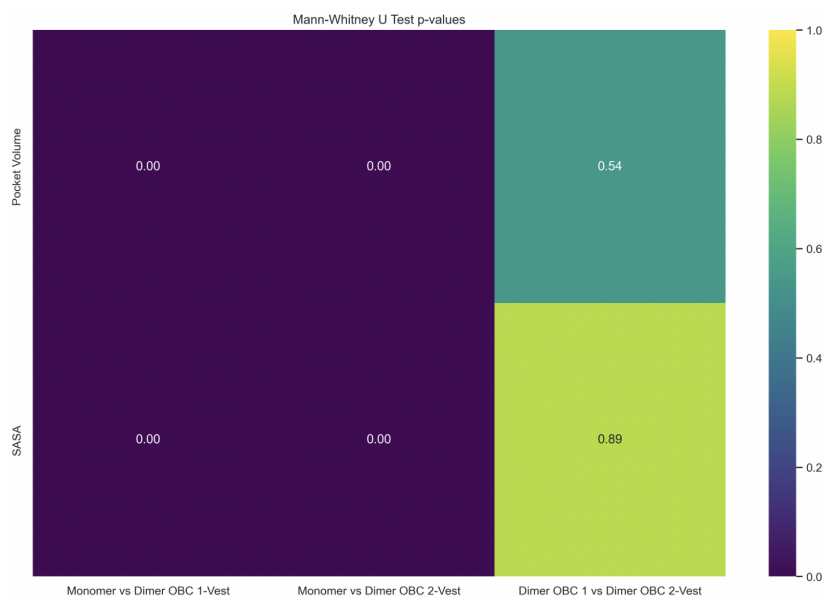


Figure 5.14: Mann-Whitney U test conducted on the properties of Volume and SASA of the monomer and dimer binding cavities. Both the orthosteric and vestibular binding cavities have been analyzed. P-values are annotated in the heatmap.

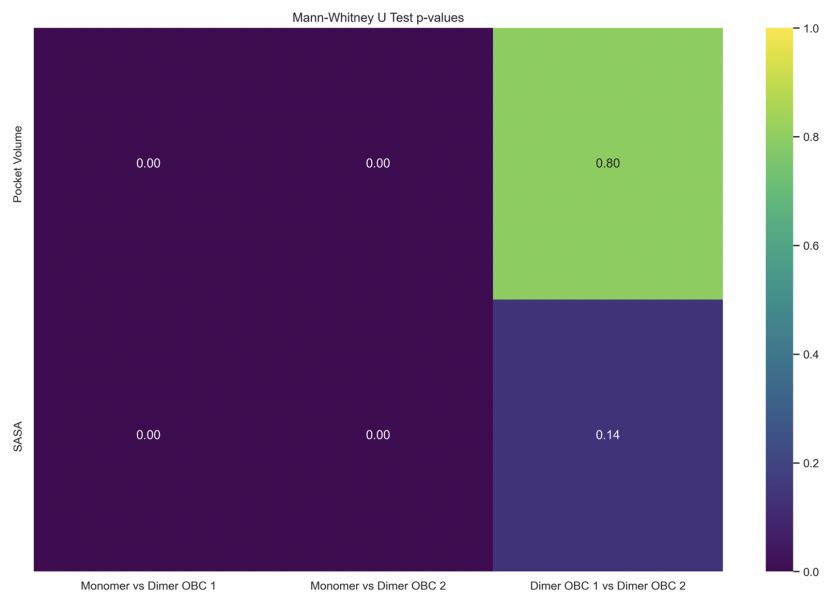


Figure 5.15: Mann-Whitney U test conducted on the Volume and SASA properties of the monomer and dimer binding cavities. Only the orthosteric binding cavity has been analyzed. P-values are annotated in the heatmap.

Several conclusions can be drawn from plots 5.14 and 5.15:

OBC Values

- **Monomer vs Dimer OBC 1:**

- Volume: U-statistic = 1397.00, p-value = 0.00

Conclusion: There is a statistically significant difference in Volume between Monomers and Dimer OBC 1.

- SASA: U-statistic = 1305.00, p-value = 0.00
Conclusion: There is a statistically significant difference in SASA between Monomers and Dimer OBC 1.

- **Monomer vs Dimer OBC 2:**

- Volume: U-statistic = 1413.00, p-value = 0.00
Conclusion: There is a statistically significant difference in Volume between Monomers and Dimer OBC 2.
- SASA: U-statistic = 1395.00, p-value = 0.00
Conclusion: There is a statistically significant difference in SASA between Monomers and Dimer OBC 2.

- **Dimer OBC 1 vs Dimer OBC 2:**

- Volume: U-statistic = 827.00, p-value = 0.80
Conclusion: There is no statistically significant difference in Volume between Dimer OBC 1 and Dimer OBC 2.
- SASA: U-statistic = 954.00, p-value = 0.14
*Conclusion: **There is no** statistically significant difference in SASA between Dimer OBC 1 and Dimer OBC 2.*

OBC-Vest Values

- **Monomer OBC-Vest vs Dimer OBC 1-Vest:**

- Volume: U-statistic = 1237.00, p-value = 0.00
Conclusion: There is a statistically significant difference in Volume between Monomer OBC-Vest and Dimer OBC 1-Vest.
- SASA: U-statistic = 1261.50, p-value = 0.00
Conclusion: There is a statistically significant difference in SASA between Monomer OBC-Vest and Dimer OBC 1-Vest .

- **Monomer OBC-Vest vs Dimer OBC 2-Vest:**

- Volume: U-statistic = 1302.00, p-value = 0.00
Conclusion: There is a statistically significant difference in Volume between Monomer and Dimer OBC 2-Vest.
- SASA: U-statistic = 1288.50, p-value = 0.00
Conclusion: There is a statistically significant difference in SASA between Monomer OBC-Vest and Dimer OBC 2-Vest.

- **Dimer OBC 1-Vest vs Dimer OBC 2-Vest:**

- Volume: U-statistic = 864.00, p-value = 0.54
Conclusion: There is no statistically significant difference in Volume between Dimer OBC 1-Vest and Dimer OBC 2-Vest.
- SASA: U-statistic = 815.50, p-value = 0.89
*Conclusion: **There is no** statistically significant difference in SASA between Dimer OBC 1-Vest and Dimer OBC 2-Vest.*

The threshold for determining statistical significance is set at 0.05 (5%).

5.5 Analysis of the D2 dimer in its inactive state

In this section, I will analyze the stability of the interface during the 40 μs unbiased simulation conducted for the D2 dimer in its inactive state. This evaluation follows the previously mentioned three separate analyses: the examination of interactions between the helix and the opposing monomer, the assessment of the potential interface using a contact matrix, and visual inspection utilizing VMD[223]. From this contact analysis, it can be observed that after approximately 1 μs , the number of contacts

is no longer stable. The TM1 of monomer A loses contact, while TM1 of monomer B gains contacts. The TM2 of both monomers maintain a low amount of contacts, as does the TM3 of monomer A. The TM4 of monomer A, similar to TM1, gains contacts. The most dramatic changes, however, occur in the remaining helices, namely TM5, TM6, and TM7. In TM5 of monomer A, the contacts increase drastically, while they are nearly zero in monomer B. TM6 loses contacts in both monomers, and TM7 in monomer A loses all contacts, whereas in monomer B, the contacts oscillate significantly. This unstable interface suggests that the initial TM6-TM6 interface is not stable during the 40 μs of unbiased simulation[53]. This finding is consistent with the observations made by Carli and colleagues[67], where the population of monomers is higher than that of dimers when D2 is inactive[36, 67]. Therefore, the preferential state for the D2 in its inactive state is the monomeric[67]. The fact that D2 receptors can exist in these two different states supports the hypothesis that class A G protein-coupled receptors may form transient dimers[36, 61]. This behavior appears clear in both the radar plot and the table shown in Figure 5.17: the average contacts during the simulation fluctuate, leading to an interface that is not only distinct from the initial structure but also asymmetric. This observation aligns with the "rolling dimer theory", which suggests that GPCR monomers in the dimer architecture can rotate relative to one another, enabling a range of configurations throughout the dimer's lifetime, establishing interfaces that might co-exist[268]. At a highly speculative level, this behavior could be explained in terms of energy, potentially minimizing ATP consumption within the cellular machinery[268]. Of course, experimental models would play a pivotal role in confirming this statement, and additional experimental results are therefore necessary to validate this theory. Nevertheless, these models elucidate the interchanging of interfaces upon dimer formation, illustrating that these are dynamic and capable of adopting multiple favored arrangements[268]. Naturally, the concept would need to be expanded if the receptor concentration in the membrane were greater than just two monomers and if other effectors were simultaneously present in unison in the membrane.

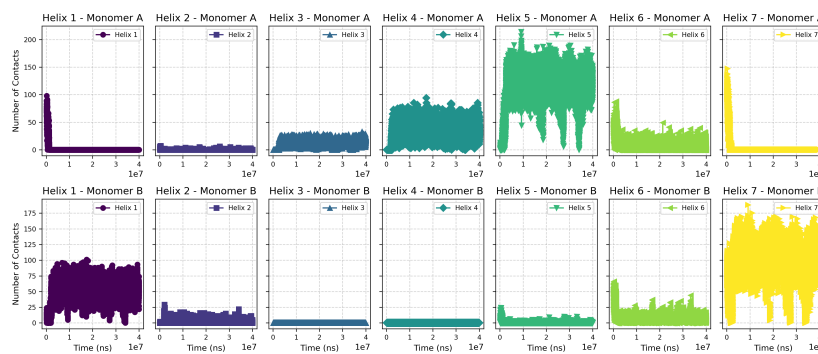


Figure 5.16: Contacts within the interface for the D2 in its inactive state.

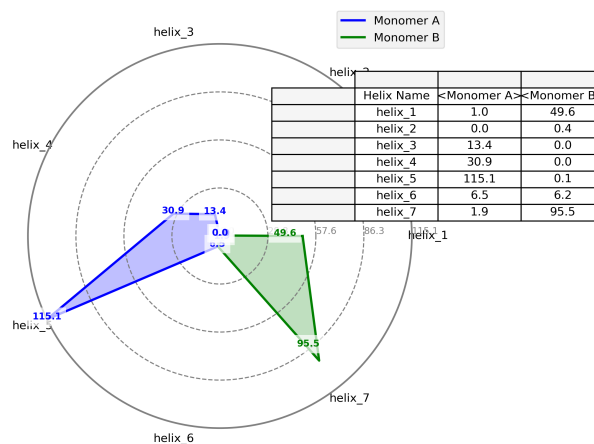


Figure 5.17: Radar plot monitoring the contacts within the interface. Angular brackets (" $<$ " and " $>$ ") denote the average of contacts for the specific helix.

5.6 Analysis of the D2 dimer in its active state in the CHOL Enriched membrane

I ran the protocol also in the CHOL Enriched Membrane to observe if the protocol was sensitive to the change of the membrane environment in finding the proper D2 dimer interface. The main reason is that GPCRs can be found in lipids rafts (see Chapter 1, Figure 5.18), fullfledged biological active hotspot microdomains for the cells, particularly enriched in sphingomyelins, phospholipids, glycopospholipids, and CHOL[82, 84].

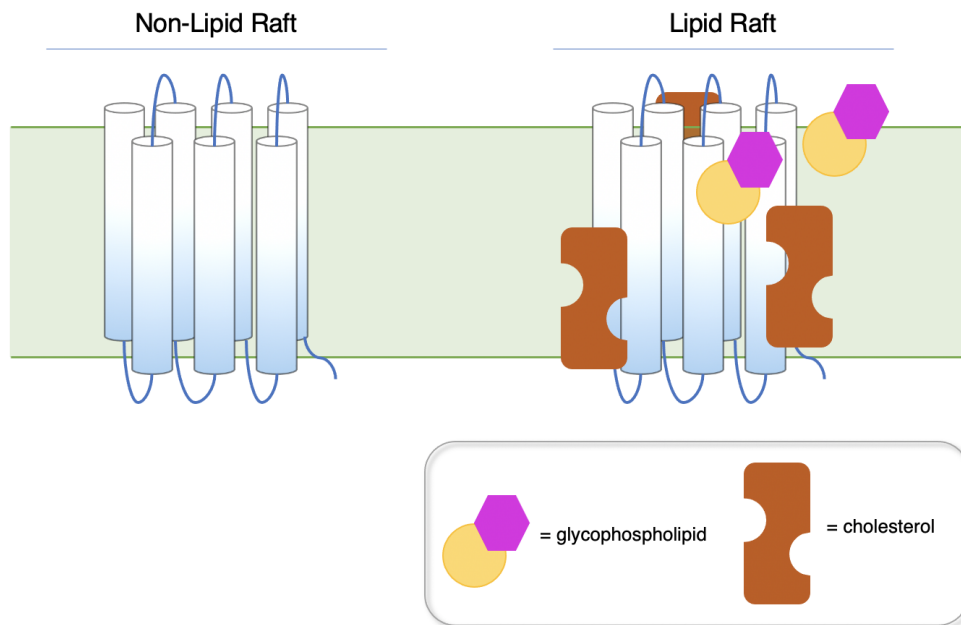


Figure 5.18: Lipid rafts in GPCRs.

Studying GPCRs in lipid rafts and their downstream pathways implies analyzing their localization and functions, but the main bottlenecks are due to the lack of standardized methods and strong limitations related to existing techniques[82].

Despite the limited information available in the literature regarding lipid rafts and their interplay with receptors[82], it is well-established that CHOL acts as an allosteric modulator of class A GPCRs[251]. CHOL can influence these receptors through two crucial mechanisms. The first is an *indirect* mechanism in which CHOL modifies the fluidity and organization of membrane lipid components and effectors[251]. These alterations, in turn, impact receptor diffusion within the membrane, and this can affect signal kinetics and efficacy[251]. The second mechanism is the *direct*. This mechanism involves CHOL interacting in proximity to the receptor(s), either by engaging at or alongside the interface or by binding to specific sites located around the receptor[251]. CHOL does not possess universal or conserved binding sites across GPCRs. Instead, it exhibits preferential binding regions that may vary among different GPCRs. The interaction of CHOL with receptors within the membrane is mainly driven by structural and dynamic factors, which are unique to each receptor[269].

From the FES profile, two states can be identified, which I have labeled as *A* and *B*. In states *A* (about 8 kJ/mol), the two monomers are partially dissociated, with center-of-mass distances (d_{COM}) of approximately 6.7 nm (Figure 5.19). The N-terminal part of the helices TM5 and TM6 of the two subunits interact, as well the two ECL3s that connect the two helices. In state *B* (about 48 kJ/mol), the TM4-TM4 interface is fully formed, consistent with experimental findings (Figure 5.19)[55, 245, 258]. Differently from the FES profile of the D2 in its active state when it is embedded in the CNS membrane, there is not a clear separation of the two subunits, and this is probably due to the interplay of lipids with the CHOL Enriched membrane.

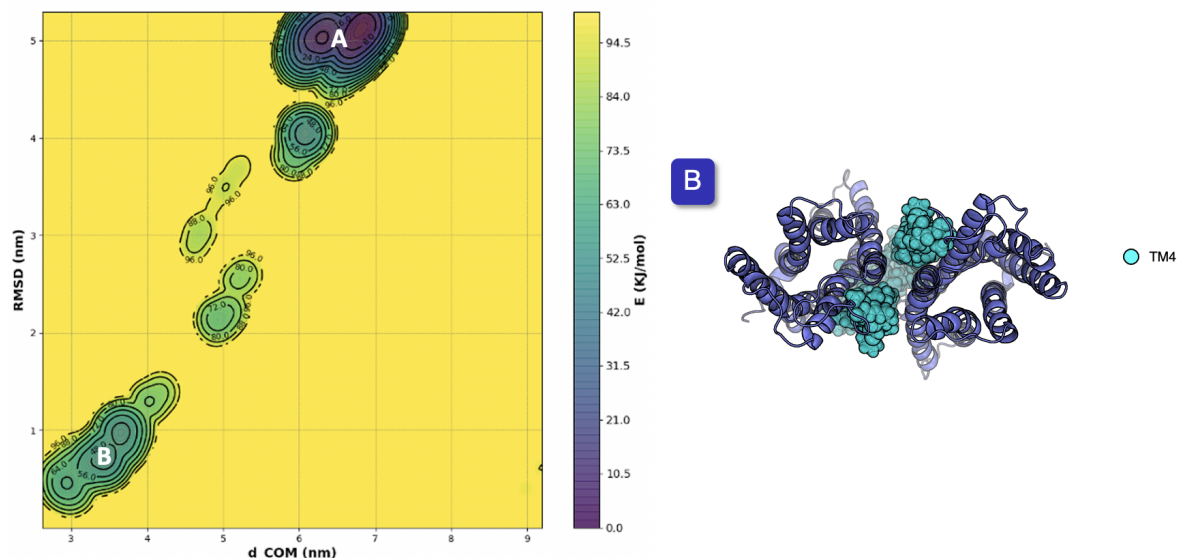


Figure 5.19: Free Energy Surface of the D2 homodimer in its active state in the CHOL Enriched membrane.

If we analyze state *A* more in detail, we can notice some differences at level of lipid interaction. Here, I calculated the density of the five lipids that constitute the CHOL Enriched membrane around the dimer receptor in its active state (Figure 5.20). The larger the surface area, the denser the lipid at that specific point. The density analysis was conducted for each of the five lipids constituting the membrane, with different colors assigned to each: green for CHOL, magenta for PAPC, cyan for PAPE, pink for PAPS, and red for POPC. The analysis was performed considering the entire lipid molecule - and not just its hydrophilic head - as the aim was to identify the density of the entire lipid rather than a specific moiety of it.

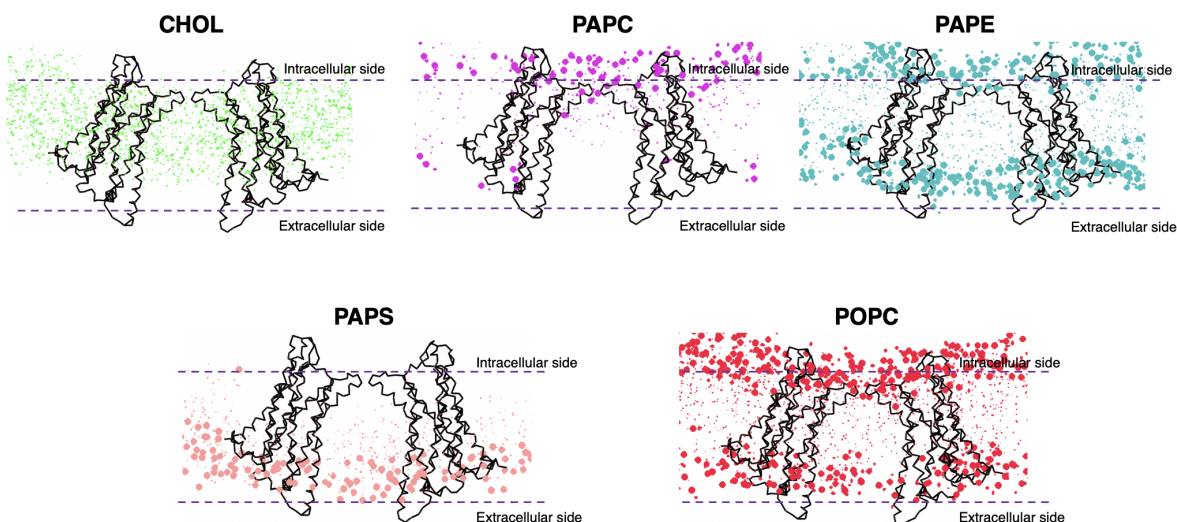


Figure 5.20: Density distribution of the five lipids of the CHOL Enriched Membrane. Dashed lines confine the membrane plane, identifying the upper part of the upper leaflet and the bottom part of the lower leaflet.

As shown in Figure 5.20, POPC and PAPE appear distributed around the N-terminal and C-terminal regions of the receptor helices, corresponding to the upper and lower parts of the two membrane leaflets. PAPC, on the other hand, seems to localize predominantly in the upper leaflet, while PAPS is confined to the lower leaflet (inherent distribution due to the membrane composition). CHOL displays a uniform density distribution (albeit lower than that of POPC, PAPC, and PAPE) specifically in the

central region of the two leaflets, referred to as the midplane. This indicates that CHOL, accordingly to this lipid composition of the membrane, exhibits a distribution pattern distinct from that of the other lipids, interacting preferentially in the midplane region.

5.7 Analysis of the D2 dimer in its inactive state in the CHOL Enriched membrane

From the FES profile, multiple states can be identified, but only two of them, labeled as *A* and *B*, have been analyzed. Other intermediates states are not considered because they possess a higher RMSD value compared to state *B* and/or are energetically equivalent. In states *A* (about 8 *kJ/mol*), the two monomers are dissociated, with center-of-mass distances (d_{COM}) of approximately 5.5 nm (Figure 5.21). In state *B* (about 32 *kJ/mol*), the TM4-TM4 interface is fully formed, consistent with experimental findings (Figure 5.21)[55, 245, 258]. Differently from the structure extracted for the D2 in its inactive state when it is embedded in the CNS membrane, here a different interface involving TM4-TM4 appears: conversely than before, the two subunits of the dimer undergo a rotation, thereby configuring a different interface from the one claimed by Pulido and colleagues and presented in Figure 5.1[53]. **This suggests that our protocol is sensitive *not only* to the receptor's activation state, but also to the surrounding membrane environment, providing a computational pipeline for understanding the differences already explored in experiments conducted using various techniques[270].**

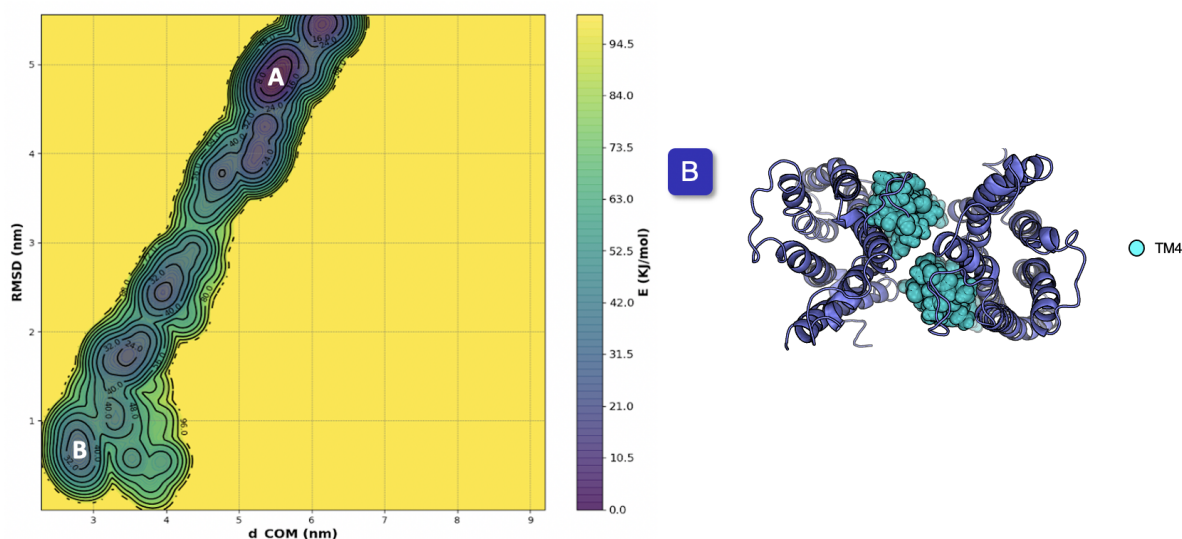


Figure 5.21: Free Energy Surface of the D2 homodimer in its inactive state in the CHOL Enriched membrane.

If once again we take a look at state *A* more in detail, we can still notice some differences at level of lipid interplay with the dimer receptor (Figure 5.22). Indeed, comparing these densities distributions with that of Figure 5.20, we notice that, the two monomers are arranged parallel to each other within the membrane, rather than being kinked relative to one another. This parallel configuration eliminates interactions involving ECL3 and TM5-TM6 in the N-terminal region. This lack of interaction might lead to an increase in the densities of PAPC, PAPE and POPC for the D2 in its inactive state in the upper and lower layer with respect of the densities in the D2 in its activate state. For CHOL and PAPS, only minor increases are detected: for CHOL, this is likely due to the midplane remaining substantially unchanged, and for PAPS, because it is exclusively present in the lower leaflet. For both Figures 5.20 and 5.22, the representative structure coming from the most representative cluster in that minimum was extracted and analyzed.

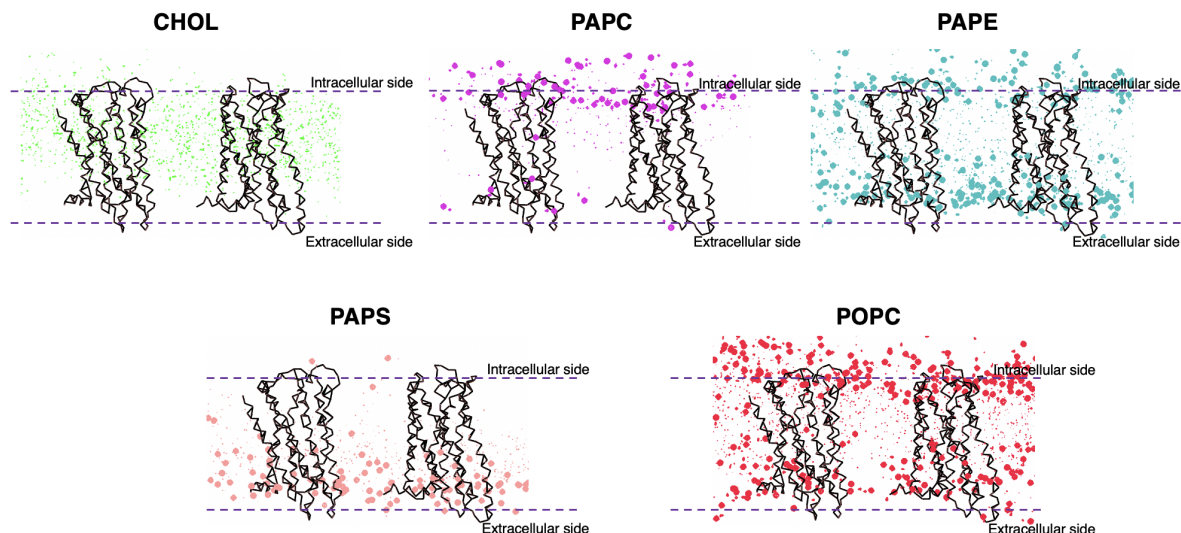


Figure 5.22: Density distribution of the five lipids of the CHOL Enriched Membrane. Dashed lines confine the membrane plane, identifying the upper part of the upper leaflet and the bottom part of the lower leaflet.

5.8 Conclusions

Two key aspects clearly emerge from these simulations. The first concerns the activation state of the D2 receptor. When this receptor is in its active state (i.e. bound to an agonist), TM6 undergoes an outward movement of approximately 9 Å, accompanied by a rearrangement of TM5, TM3, and TM7[43]. This movement is essential for accommodating the trimeric protein in the intracellular space and enabling downstream functions, and could also disfavor TM6-TM6 interactions at protein-protein interfaces. In my simulations, we observed that after 40 μ s of unbiased simulation, the TM4-TM4 and TM5-TM5 (with less extent) regions represented the preferred interaction interface when the receptor is in its active state. Lipids play a key role in stabilizing this interface, particularly CHOL, which interacts with TM4 residues within the CCM (identified in the monomer). **My hypothesis is that CHOL stabilizes the dimer in this configuration.**

The situation is different, however, in the inactive state. When D2 is in this state, the dynamic equilibrium between dimer and monomer shifts from approximately 50%-50% to 70% monomer-30% dimer. The inward movement of TM6 described above, when reversed, could explain this shift and the loss of interface stability. This hypothesis is also supported by the "Rolling Dimer Theory", according to which multiple dimers can co-exist simultaneously[268]. It is clear, however, that further experimental data are needed to corroborate these findings also in D2 receptors, specifically. **The protocol, based on what has been observed so far, has proven sensitive to the GPCR's activation state.** Naturally, we cannot reproduce exactly the percentage of dynamic equilibrium involved for the activate and inactive state of the GPCR, but *we can* reproduce the trend, and this is visible from the energy profiles.

The second key aspect concerns membrane interaction and its topology. GPCRs can be localized within lipid raft domains, micro environments particularly rich in CHOL (and other lipids)[82, 84]. For this reason, I decided to simulate D2 in two different membrane environments to better understand their potential differences, namely in CNS and CHOL Enriched membranes.

The results show that the energy profile varies depending on the membrane type. As mentioned previously, the models will require further optimization, **but this provides an excellent starting point for determining that, in addition to being sensitive to the receptor's activation state, the protocol is also highly dependent on the membrane environment.** Furthermore, the identified dimers have been used as a starting point for other projects by colleagues, such as Simone Mariani, Simone Albani and Federico Natale, for calculating the residence time (described in the introductory chapter). **This protocol therefore represents a fundamental starting point for drug design.**

Nevertheless, enhanced sampling techniques poses challenges related to convergence[141, 142]. Briefly,

convergence requires both rapid diffusion of the system along the chosen CVs and a consistent reconstruction of the FES, as discussed more in detail in the conclusions of the previous chapter. Despite the use of DEEP-LDA derived CVs, the usage of CG force field and extended simulation time scales, full convergence was not achieved[146, 149].

To overcome this limitation, always in collaboration with my colleagues Simone Mariani, Simone Albani and Federico Natale, we are optimizing the computational dimerization protocol to obtain a more reliable free-energy profile. Our current approach employs ten multiple walkers per simulation and foresees hybrid strategies, i.e. it combines umbrella sampling to restrain specific configurations (e.g., at experimentally observed interfaces[53, 55, 92, 260]) with other enhanced sampling methods, such as Well-Tempered Metadynamics or the new OPES[141, 147]. These simulations are ongoing and being extended beyond 10 μ s, potentially achieving the millisecond timescale, as demonstrated by Di Marino *et al.* [169].

Finally, an extra consideration for future work is the adoption of a more complex membrane model. The most recent release of the Martini 3 force field entails improved parametrizations for cholesterol and an expansion of its lipidome[182, 271]. As the membrane can be considered as fully fledged second messenger, also this "new version" of the membrane will be incorporated into the forthcoming computational publications.

5.9 Contribution to this Chapter

In this chapter, I calculated the energy profiles of D2 in both its active and inactive states, considering its behavior in the CNS membrane as well as in the previously defined CHOL Enriched membrane. As evident from the resulting energy landscapes, the two CVs have been redefined to enhance their visual clarity, allowing for a more precise, practical and unambiguous identification of energy minima.

Before this analysis, all free energy profiles underwent a procedure commonly referred to as *reweighting*. In brief, this statistical technique enables the recovery of the system's original equilibrium properties after it has been subjected to biased simulations, like in my case, having used the WTMETAD. The reweighting process and the improvement of the visualization and rationalization of the minima in all the simulations were carried out by me in collaboration with Dr. Davide Mandelli at the Forschungszentrum Jülich. The remainder of this chapter is entirely my own work.

I made extensive use of various computational tools, all of which are properly cited. In addition, the algorithms presented in Appendix A.1 and A.2, developed independently by me, were designed to compute the Lipid Count and Interaction Count, and were further employed to generate their corresponding graphical representations. Finally, I personally conducted the statistical analyses (Appendix A.3) to provide a stronger quantitative result for the data derived from the calculation of Pocket Volume and SASA, giving greater robustness and reliability of the calculated computational results.

Chapter 6

Conclusions and Future Outlooks

Membrane proteins are pivotal components of eukaryotic cells and play crucial roles in cellular communication, transport, and molecular recognition[272]. Due to their essential biological functions, they represent one of the most important classes of pharmacological targets[2, 272, 273]. Among membrane proteins, G protein-coupled receptors (GPCRs) constitute one of the largest receptor superfamilies that are responsible for transducing extracellular stimuli derived from exogenous or endogenous molecules into intracellular signaling responses.

Beyond their canonical signaling function, GPCRs oligomerization introduces an additional aspect of functional complexity, as these receptors do not necessarily operate exclusively as monomers, but can form homo- or hetero-oligomeric entities within the plasma membrane environment with distinct pharmacological and signaling properties compared to their monomeric counterparts[58, 59]. In particular, this oligomerization-mediated allosteric mechanisms can significantly impact ligand binding by modulating receptor affinity and cooperativity between protomer[274]. In several receptor systems, including dopamine and adenosine receptors, allosteric interactions between receptor units have been shown to alter receptor pharmacology and downstream signaling pathways[59]. Likewise, in GABAB receptor, oligomerization mechanism has been reported to regulate the accessibility and simultaneous occupancy of orthosteric binding sites, further emphasizing the functional relevance of protein-protein interactions[275]. Besides, GPCR oligomerization can alter downstream signaling by modulating G-protein coupling selectivity, β -arrestin recruitment, and receptor trafficking, spotlighting its important role in regulating GPCR-mediated cellular responses[276].

A plethora of experimental techniques allow the detection of GPCR oligomers[277]. Biochemical methods, such as co-immunoprecipitation, in situ hybridization or immunohistochemistry, are helpful to demonstrate receptor colocalization, despite providing limited information on protein dynamics and may introduce experimental artefacts[277]. Resonance energy transfer (RET)-based techniques, such as FRET and BRET, enable the investigation of receptor interactions in living cells with high sensitivity. However, FRET measurements may affect protein function due to the steric hindrance introduced by fluorescent protein tags, while BRET does not clearly differentiate between receptors located at the cell surface and from those located intracellularly[67, 277]. Moreover, these techniques do not allow direct quantification of receptor dimerization or the resolution of single-molecule dynamics[67]. Advanced single-molecule imaging techniques, such as TIRFM, facilitate real-time analysis of GPCR oligomerization at single-molecule resolution, but their performance is often constrained by low signal intensity and unstable background fluorescence[270, 277]. In contrast, X-ray crystallography provides atomic-level structural resolution, although it captures static and potentially non-physiological conformations[277]. Overall, each technique has intrinsic limitations, and none is sufficient alone to fully describe oligomerization[277].

A further layer of complexity emerges from the highly dynamic nature of GPCR oligomerization. The oligomeric interactions are often transient, specifically for class A GPCRs, and the monomer-to-dimer equilibria is modulated by several factors, including receptor expression levels, membrane composition, lipid microdomains (lipid rafts), and ligand binding (agonists or antagonists)[67]. This complex interplay makes it challenging to discriminate physiologically relevant oligomers from experimental artefacts[278].

Due to these limitations, the structural characterization of GPCR oligomeric complexes in native environments remains particularly complicated. This has therefore led the researchers to the development of computational approaches aimed at complementing experimental data for the modelling

and prediction of GPCR oligomerization. The scope of these approaches is to identify detailed atomic determinants of protein-protein interfaces/interactions. Several methods have been proposed so far in the literature to reconstruct biological assemblies and derive their corresponding interaction interfaces, including protein-protein docking and molecular dynamics (MD) simulations[169, 220, 259].

In this context, the protocol described in Chapter 4 of this thesis offers an additional approach for the identification of GPCR oligomers within a membrane environment. Compared to existing computational methods, our pipeline differs in two key aspects: first, it is a blind protocol, relying on generic collective variables (CVs) to describe the receptor association-dissociation process; second, it explicitly accounts for the role of the membrane environment, which can be key in modulating receptor-receptor interactions[251]. By employing the Martini 3 force field to speed-up simulations and well-tempered metadynamics to enhance sampling, we were able to identify the interfaces involved in D2 receptor dimerization, in agreement with experimental data[53, 55, 258]. Specifically, the approach developed in this thesis was utilized to probe the conformational dynamics of the D2 receptor in both its active and inactive states across two membrane environments (see Chapter 4): one CNS-mimicking membrane and one cholesterol-enriched lipid raft membrane, where many GPCRs are known to localize and coordinate signaling regulation[82]. Given its blind design, this pipeline is not limited to the D2 and can be naturally extended to other membrane proteins.

As a future outlook of this thesis, the protocol will be expanded to other membrane receptors, initially to GPCRs such as the A2 receptor, whose dimerization interfaces have already been experimentally characterized, and subsequently to other classes of integral membrane proteins[59, 92]. Moreover, increasing evidence suggests that GPCR oligomerization is strongly modulated by ligand binding[67]. Agonists, antagonists, and allosteric modulators may alter receptor conformational states, thus impacting oligomer stability, signaling, association-dissociation equilibria, and interface formation[67]. The inclusion of ligands and, potentially, G proteins or β -arrestins in computational frameworks such as the one presented here will represent a pivotal extension toward a more realistic description of membrane protein oligomerization closer to physiological conditions[67]. As an extra perspective, membrane modeling will be further refined through the adoption of recently developed state-of-the-art sterols and phospholipids parameterization[182, 271], alongside optimized CVs and hybrid enhanced-sampling strategies designed in collaboration with my colleagues Federico Natale and Simone Albani at the Computational Biomedicine Institute (INM-9) at Forschungszentrum Jülich. Overall, these advancements will enable a more accurate characterization of the association-dissociation mechanism and support the investigation of increasingly complex biological questions. Additionally, the pipeline developed in this work will be integrated into the CGMD Platform (Chapter 3), thereby expanding its accessibility and use within the scientific community[156]. The platform currently implements a previous version of the Martini force field[151] and does not yet include the protocol presented here. Its future integration will therefore represent a substantial methodological advancement.

Beyond the specific implementation within web platforms like the CGMD web-server, the availability of shared dimerization protocols provides a valuable framework for studying receptor-receptor interactions across membrane systems such as GPCRs, ion channels, and transporters, where these interactions play a key role in regulating biological activities[279]. In this context, the structures generated by the methodology presented here may provide reliable starting points for future studies of membrane protein function and drug design. More broadly, such computational pipelines aim to overcome and complement the limitation of experimental approaches providing key information on the impact of dimerization on the structure and dynamics of membrane proteins.

The entire methodology described above was developed through two complementary parts of my Ph.D. The first part was conducted within the Helmholtz European Partnering (HEP) framework at the Computational Biomedicine Institute (INM-9) at Forschungszentrum Jülich in collaboration with the Italian Institute of Technology (IIT). During this period, I focused on the development of computational workflows for neuroscience applications, with a specific focus on the implementation and application of algorithms for the definition and selection of CVs, developed jointly with IIT. Following this first part of my Ph.D., the second part was conducted within the RTG framework, where I developed a deeper understanding of multiscale approaches in GPCR biology. GPCR function manifests across multiple levels, from molecular structure to physiological responses. At the molecular level, membrane protein oligomerization and consequent signaling properties are driven by tailored interactions and conformational dynamics within the surrounding lipid milieu. These molecular processes extend to cellular signaling and functional outcomes, illustrating how biological mechanisms propagate across different scales. Accordingly, computational approaches such as those developed in this work are crucial for enabling a multiscale description of these processes.

Chapter 7

Appendix

A.1

This appendix shows the code used within this thesis. All codes are run with Python 3[280]. This code aims to calculate the CHOL lipid count (all the beads of CHOL are taken into account) within 5 nm from the CG protein model, defined with beads *BB*, *SC1*, *SC2*, *SC3*, *SC4*, and *SC5*. Trajectory is read every 100 ns.

Code 1: calculation of CHOL lipid count

```

1  import csv
2  from collections import Counter
3  import MDAnalysis as mda
4  from MDAnalysis.analysis import distances
5
6  topology = "topology.gro"
7  trajectory = "trj.xtc"
8
9  u = mda.Universe(topology, trajectory)
10
11 lip = "CHOL"
12
13 lipid_selection = "resname " + lip
14 lipid_atoms = u.select_atoms(lipid_selection)
15 cg_residues = "name BB SC1 SC2 SC3 SC4 SC5"
16 cg_residues_atoms = u.select_atoms(cg_residues)
17
18 cut_off = 5
19 skip = 100
20
21 total_res_counts = Counter()
22
23 with open('output_CHOL.csv', 'w', newline='') as csvfile:
24     fieldnames = ['Frame', 'Residue Name', 'Residue ID', 'Count']
25     writer = csv.DictWriter(csvfile, fieldnames=fieldnames)
26
27     writer.writeheader()
28
29     for i, ts in enumerate(u.trajectory[::skip]):
30         interactions = []
31         for res in cg_residues_atoms:
32             for lipid in lipid_atoms:
33                 distances_array = distances.distance_array(res.position, lipid.
                     position)

```

```
34
35 # Check if the distance is within the cutoff
36 if all(d <= cut_off for d in distances_array):
37     interactions.append([res.resname, res.id, lipid.id])
38
39 # Count the occurrences of res.id in the interactions list
40 res_counts = Counter((res_name, res_id) for res_name, res_id, _ in
41     interactions)
42
43 # Update the total_res_counts with counts from this frame
44 total_res_counts.update(res_counts)
45 all_res = res_counts.most_common() #All frames with a lipid count > 1
46 are written in the csv file.
47 top_10_res = res_counts.most_common(10)
48
49 print(f"Frame = {ts.frame}")
50 print("")
51 for res, count in top_10_res:
52     print(f"Residue Name: {res[0]}, Residue ID: {res[1]}, Count: {count}")
53     print("")
54
55 for res, count in all_res:
56     writer.writerow({'Frame': ts.frame, 'Residue Name': res[0], 'Residue
57         ID': res[1], 'Count': count})
58
59 print("\n*****Aggregated counts across all frames*****")
60 for res, count in total_res_counts.most_common(10):
61     print(f"Residue Name: {res[0]}, Residue ID: {res[1]}, Count: {count}")
```

A.2

The following code aims to view the result calculated with Code 1.

Code 2: visualization of the interaction count

```

1      import pandas as pd
2      import matplotlib.pyplot as plt
3      import matplotlib.colors as mcolors
4      import numpy as np
5
6      # Load the CSV file into a DataFrame
7      df = pd.read_csv('output_all_CHOL_more_specific_monomer.csv')
8
9      # Define the ranges for TM helices and H8 for both monomers
10     ranges_a = {
11         'Helix1_A': (1, 64),
12         'Helix2_A': (87, 142),
13         'Helix3_A': (182, 227),
14         'Helix4_A': (284, 333),
15         'Helix5_A': (365, 433),
16         'Helix6_A': (485, 539),
17         'Helix7_A': (573, 629),
18         'H8_A': (630, 662)
19     }
20
21     ranges_b = {
22         'Helix1_B': (663, 726),
23         'Helix2_B': (749, 804),
24         'Helix3_B': (844, 889),
25         'Helix4_B': (946, 995),
26         'Helix5_B': (1027, 1095),
27         'Helix6_B': (1147, 1201),
28         'Helix7_B': (1235, 1291),
29         'H8_B': (1292, 1324)
30     }
31
32     # Define a custom color palette for TM helices
33     custom_palette = {
34         'Helix1': 'dodgerblue', # blue
35         'Helix2': '#ff7f0e', # orange
36         'Helix3': 'forestgreen', # green
37         'Helix4': 'crimson', # red
38         'Helix5': '#9467bd', # purple
39         'Helix6': '#8c564b', # brown
40         'Helix7': '#e377c2', # pink
41         'H8': '#7f7f7f', # gray
42     }
43
44     # Initialize dictionaries to store lipid counts for each helix
45     # for both monomers
46     lipid_counts_a = {helix: [] for helix in ranges_a}
47     lipid_counts_b = {helix: [] for helix in ranges_b}
48
49     for index, row in df.iterrows():
50         for helix, (start, end) in ranges_a.items():
51             if start <= row['Residue ID'] <= end:
52                 lipid_counts_a[helix].append(row['Count'])

```

```

52
53     for helix, (start, end) in ranges_b.items():
54         if start <= row['Residue ID'] <= end:
55             lipid_counts_b[helix].append(row['Count'])
56
57
58     helices_a = list(lipid_counts_a.keys())
59     lipid_count_freq_a = [len(counts) for counts in lipid_counts_a
60                           .values()]
61     helices_b = list(lipid_counts_b.keys())
62     lipid_count_freq_b = [len(counts) for counts in lipid_counts_b
63                           .values()]
64
65     helices_combined = list(zip(helices_a, helices_b))
66     lipid_count_freq_combined = list(zip(lipid_count_freq_a,
67                                          lipid_count_freq_b))
68
69     # Use a custom color palette
70     colors_a = [custom_palette[helix.split('_')[0]] for helix in
71                helices_a]
72     colors_b = [custom_palette[helix.split('_')[0]] for helix in
73                helices_b]
74
75     # Plot the distribution of lipid count frequencies for both
76     # monomers combined
77     plt.figure(figsize=(12, 6))
78     bars_combined = []
79     width = 0.35
80     ind = np.arange(len(helices_combined))
81
82     for i, (helix_a, helix_b) in enumerate(helices_combined):
83         freq_a, freq_b = lipid_count_freq_combined[i]
84         bar_a = plt.bar(ind[i] - width/2, freq_a, width,
85                         color=colors_a[i], edgecolor="black", label=f'{helix_a[:-2]} (
86                             Monomer A)')
87         bars_combined.extend([bar_a])
88
89     plt.ylabel('Interaction Count')
90     #plt.title('CHOL count for the helices and H8 of the D2 active
91     #dimer in CNS membrane')
92     plt.xticks(ind, [f'{pair[0][:-2]}' for pair in
93                     helices_combined], rotation=45, ha='right')
94     plt.grid(axis='y', linestyle='--', alpha=0.5)
95
96     # Create legend entries for Monomer A
97     legend_handles_a = [plt.Rectangle((0,0),1,1, color=color) for
98                         color in colors_a]
99     legend_labels_a = [f'{helix.split("_")[0]}' for helix in
100                       helices_a]
101
102     # Combine legend entries for Monomer A
103     legend_handles = legend_handles_a
104     legend_labels = legend_labels_a
105
106     plt.legend(legend_handles, legend_labels, title='Helices', loc
107               ='upper right',
108               fancybox=True, ncol=2, frameon=True, edgecolor='gray')

```

```
98
99     plt.tight_layout()
100    plt.show()
```

A.3

The first six analysis are conducted on the orthosteric and vestibular binding sites of the GPCR. The next six only on the orthosteric binding sites.

Normality plots - Orthosteric and Vestibular binding pockets

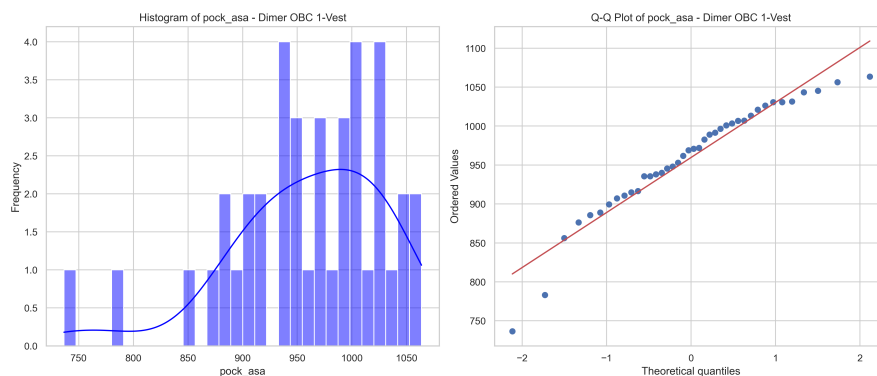


Figure 1: Histogram and Q-Q plot to check the normality of the distribution of SASA values (pock_asa) of the dimer OBC 1-Vest.



Figure 2: Histogram and Q-Q plot to check the normality of the distribution of Volume values (pock_volume) of the dimer OBC 1-Vest.

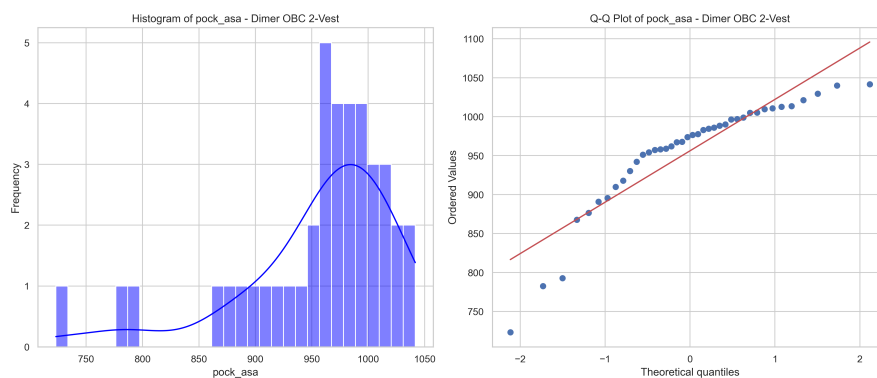


Figure 3: Histogram and Q-Q plot to check the normality of the distribution of SASA values (pock_asa) of the dimer OBC 2-Vest.

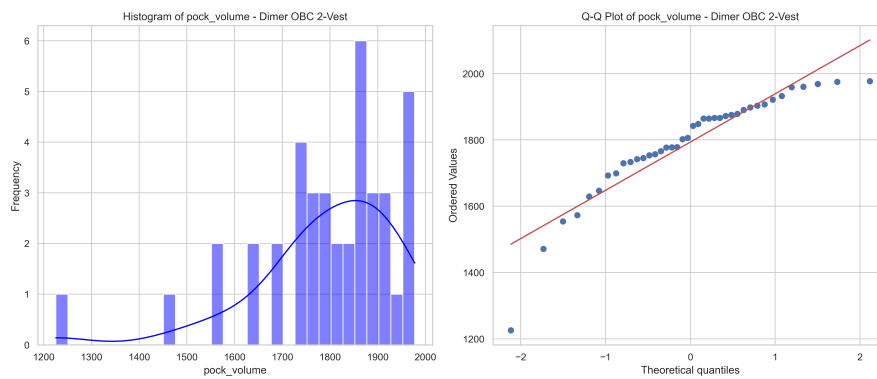


Figure 4: Histogram and Q-Q plot to check the normality of the distribution of Volume values (pock_volume) of the dimer OBC 2-Vest.

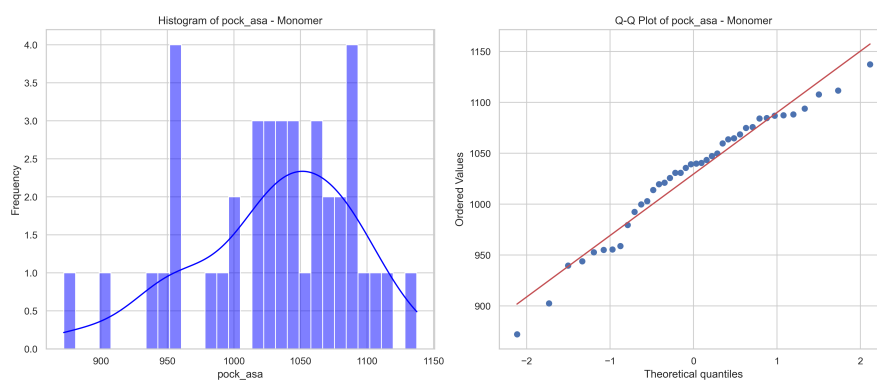


Figure 5: Histogram and Q-Q plot to check the normality of the distribution of SASA values (pock_asa) of the Monomer OBC-Vest.

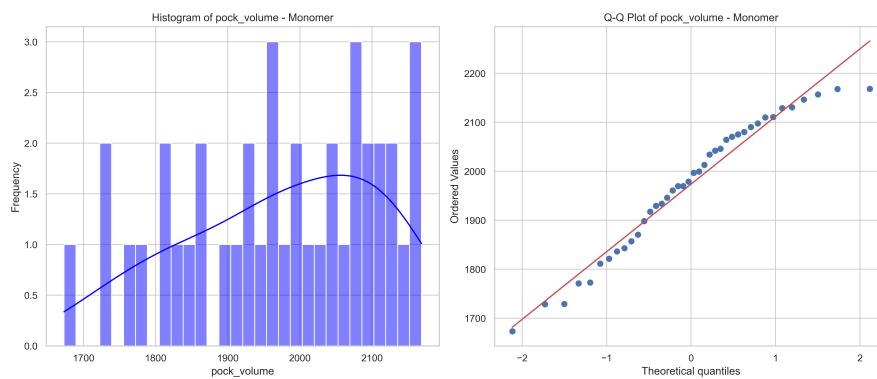


Figure 6: Histogram and Q-Q plot to check the normality of the distribution of Volume values (pock_volume) of the Monomer-OBC-Vest.

Normality plots - Orthosteric binding pockets

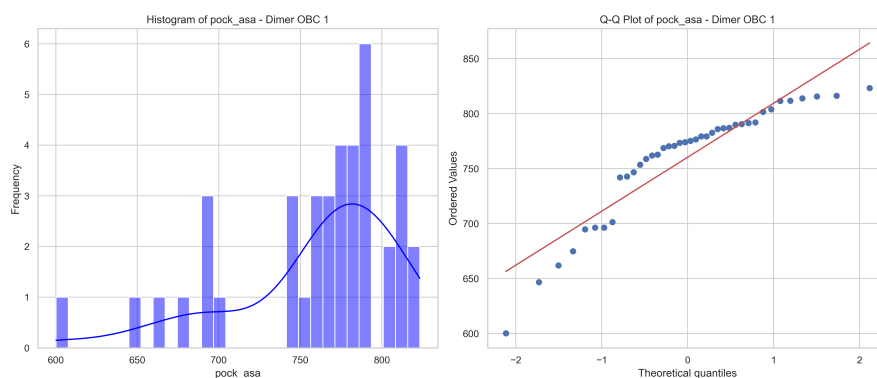


Figure 7: Histogram and Q-Q plot to check the normality of the distribution of SASA values (pock_asa) of the dimer OBC 1.

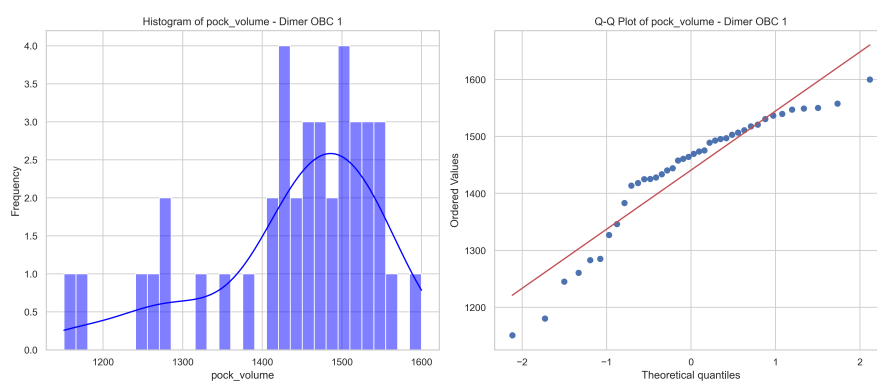


Figure 8: Histogram and Q-Q plot to check the normality of the distribution of Volume values (pock_volume) of the dimer OBC 1.

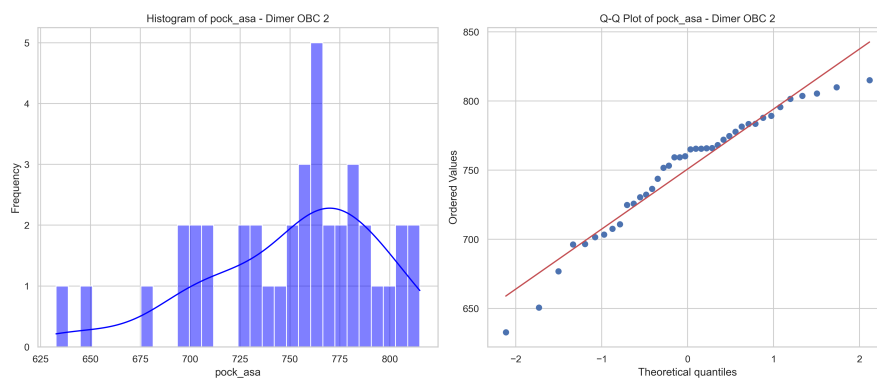


Figure 9: Histogram and Q-Q plot to check the normality of the distribution of SASA values (pock_asa) of the dimer OBC 2.

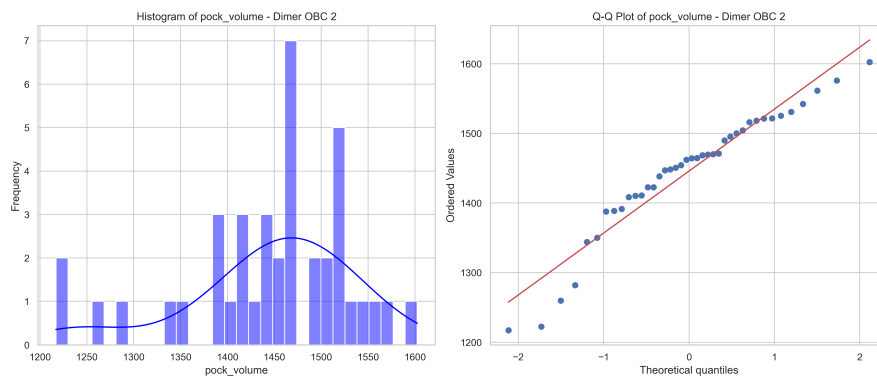


Figure 10: Histogram and Q-Q plot to check the normality of the distribution of Volume values (pock_volume) of the dimer OBC 2.

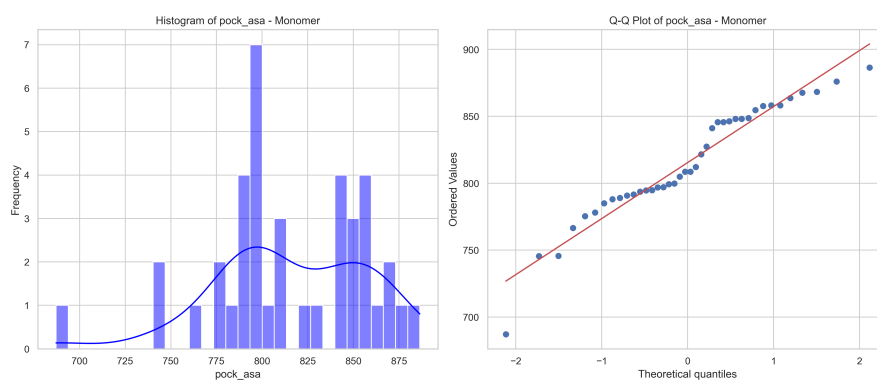


Figure 11: Histogram and Q-Q plot to check the normality of the distribution of SASA values (pock_asa) of the Monomer OBC.

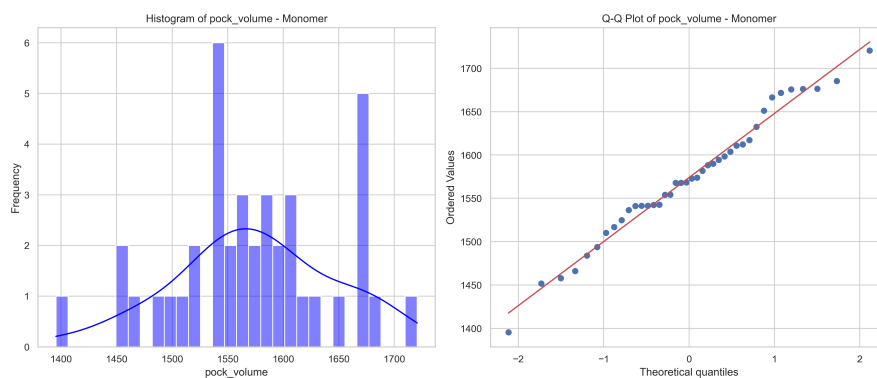


Figure 12: Histogram and Q-Q plot to check the normality of the distribution of Volume values (pock_volume) of the Monomer OBC.

Bibliography

- [1] Kazimierz Kochman. “Superfamily of G-protein coupled receptors (GPCRs)—extraordinary and outstanding success of evolution”. In: *Advances in Hygiene and Experimental Medicine* 68 (2014), pp. 1225–1237.
- [2] Dehua Yang, Qingtong Zhou, Viktorija Labroska, Shanshan Qin, Sanaz Darbalaei, Yiran Wu, Elita Yuliantie, Linshan Xie, Houchao Tao, Jianjun Cheng, et al. “G protein-coupled receptors: structure-and function-based drug discovery”. In: *Signal transduction and targeted therapy* 6.1 (2021), p. 7.
- [3] Paul A Insel, Krishna Sriram, Matthew W Gorr, Shu Z Wiley, Alexander Michkov, Cristina Salmerón, and Amy M Chinn. “GPCRomics: an approach to discover GPCR drug targets”. In: *Trends in pharmacological sciences* 40.6 (2019), pp. 378–387.
- [4] Davide Calebiro, Tamara Miljus, and Shannon O’Brien. “Endomembrane GPCR signaling: 15 years on, the quest continues”. In: *Trends in Biochemical Sciences* (2024).
- [5] Daniel M Rosenbaum, Søren GF Rasmussen, and Brian K Kobilka. “The structure and function of G-protein-coupled receptors”. In: *Nature* 459.7245 (2009), pp. 356–363.
- [6] Lin Cheng, Fan Xia, Ziyang Li, Chenglong Shen, Zhiqian Yang, Hanlin Hou, Suyue Sun, Yuying Feng, Xihao Yong, Xiaowen Tian, et al. “Structure, function and drug discovery of GPCR signaling”. In: *Molecular Biomedicine* 4.1 (2023), p. 46.
- [7] Helgi B Schiöth and Robert Fredriksson. “The GRAFS classification system of G-protein coupled receptors in comparative perspective”. In: *General and comparative endocrinology* 142.1-2 (2005), pp. 94–101.
- [8] Geng-Ming Hu, Te-Lun Mai, and Chi-Ming Chen. “Visualizing the GPCR network: classification and evolution”. In: *Scientific reports* 7.1 (2017), p. 15495.
- [9] Alan D Goddard and Anthony Watts. “Regulation of G protein-coupled receptors by palmitoylation and cholesterol”. In: *BMC biology* 10 (2012), pp. 1–3.
- [10] Krishna Sriram and Paul A Insel. “G protein-coupled receptors as targets for approved drugs: how many targets and how many drugs?” In: *Molecular pharmacology* 93.4 (2018), pp. 251–258.
- [11] Zhaotong Cong, Yi-Lynn Liang, Qingtong Zhou, Sanaz Darbalaei, Fenghui Zhao, Wenbo Feng, Lihua Zhao, H Eric Xu, Dehua Yang, and Ming-Wei Wang. “Structural perspective of class B1 GPCR signaling”. In: *Trends in Pharmacological Sciences* 43.4 (2022), pp. 321–334.
- [12] Anna Qiao, Shuo Han, Xinmei Li, Zhixin Li, Peishen Zhao, Antao Dai, Rulve Chang, Linhua Tai, Qiuxiang Tan, Xiaojing Chu, et al. “Structural basis of Gs and Gi recognition by the human glucagon receptor”. In: *Science* 367.6484 (2020), pp. 1346–1352.
- [13] Yi-Lynn Liang, Peishen Zhao, Christopher Draper-Joyce, Jo-Anne Baltos, Alisa Glukhova, Tin T Truong, Lauren T May, Arthur Christopoulos, Denise Wootten, Patrick M Sexton, et al. “Dominant negative G proteins enhance formation and purification of agonist-GPCR-G protein complexes for structure determination”. In: *ACS pharmacology & translational science* 1.1 (2018), pp. 12–20.
- [14] J-P Pin, Julie Kniazeff, Jianfeng Liu, Virginie Binet, Cyril Goudet, Philippe Rondard, and Laurent Prézeau. “Allosteric functioning of dimeric class CG-protein-coupled receptors”. In: *The FEBS journal* 272.12 (2005), pp. 2947–2955.
- [15] Jean-Philippe Pin and Bernhard Bettler. “Organization and functions of mGlu and GABAB receptor complexes”. In: *Nature* 540.7631 (2016), pp. 60–68.

- [16] Edward M Brown and Robert J MacLeod. “Extracellular calcium sensing and extracellular calcium signaling”. In: *Physiological reviews* 81.1 (2001), pp. 239–297.
- [17] Anne Brockhoff, Maik Behrens, and Wolfgang Meyerhof. “Structural requirements of bitter taste receptor activation”. In: *Proceedings of the National Academy of Sciences* 107.24 (2010), pp. 11110–11115.
- [18] Hans Bräuner-Osborne, Patrik Wellendorph, and Anders A Jensen. “Structure, pharmacology and therapeutic prospects of family C G-protein coupled receptors”. In: *Current drug targets* 8.1 (2007), pp. 169–184.
- [19] Gunnar Schulte and Paweł Koźmielewicz. “Structural insight into Class F receptors—What have we learnt regarding agonist-induced activation?” In: *Basic & Clinical Pharmacology & Toxicology* 126 (2020), pp. 17–24.
- [20] William I Weis and Brian K Kobilka. “The molecular basis of G protein-coupled receptor activation”. In: *Annual review of biochemistry* 87 (2018), pp. 897–919.
- [21] Marta Lopez-Balastegui, Tomasz Maciej Stepniewski, Małgorzata M Kogut-Günthel, Antonella Di Pizio, Mette Marie Rosenkilde, Jiafei Mao, and Jana Selent. “Relevance of G protein-coupled receptor (GPCR) dynamics for receptor activation, signalling bias and allosteric modulation”. In: *British Journal of Pharmacology* (2024).
- [22] Kelly A Berg and William P Clarke. “Making sense of pharmacology: inverse agonism and functional selectivity”. In: *International Journal of Neuropsychopharmacology* 21.10 (2018), pp. 962–977.
- [23] Ikuo Masuho, Ryoji Kise, Pablo Gainza, Ee Von Moo, Xiaona Li, Ryosuke Tany, Hideko Wakasugi-Masuho, Bruno E Correia, and Kirill A Martemyanov. “Rules and mechanisms governing G protein coupling selectivity of GPCRs”. In: *Cell reports* 42.10 (2023).
- [24] Mingyang Zhang, Ting Chen, Xun Lu, Xiaobing Lan, Ziqiang Chen, and Shaoyong Lu. “G protein-coupled receptors (GPCRs): advances in structures, mechanisms, and drug discovery”. In: *Signal Transduction and Targeted Therapy* 9.1 (2024), p. 88.
- [25] Vicent Casadó and Verònica Casadó-Anguera. “What are the current trends in G protein-coupled receptor targeted drug discovery?” In: *Expert opinion on drug discovery* 18.8 (2023), pp. 815–820.
- [26] Xin Cai, Dexiu Wang, Rumin Zhang, Yanchun Chen, and Jing Chen. “The transmembrane domains of GPCR dimers as targets for drug development”. In: *Drug Discovery Today* 28.1 (2023), p. 103419.
- [27] Shaherin Basith, Minghua Cui, Stephani JY Macalino, Jongmi Park, Nina AB Clavio, Soosung Kang, and Sun Choi. “Exploring G protein-coupled receptors (GPCRs) ligand space via cheminformatics approaches: impact on rational drug design”. In: *Frontiers in pharmacology* 9 (2018), p. 338587.
- [28] Sheng Wang, Tao Che, Anat Levit, Brian K Shoichet, Daniel Wacker, and Bryan L Roth. “Structure of the D2 dopamine receptor bound to the atypical antipsychotic drug risperidone”. In: *Nature* 555.7695 (2018), pp. 269–273.
- [29] Jean-Martin Beaulieu and Raul R Gainetdinov. “The physiology, signaling, and pharmacology of dopamine receptors”. In: *Pharmacological reviews* 63.1 (2011), pp. 182–217.
- [30] Dohyun Im, Asuka Inoue, Takaaki Fujiwara, Takanori Nakane, Yasuaki Yamanaka, Tomoko Uemura, Chihiro Mori, Yuki Shiimura, Kanako Terakado Kimura, Hidetsugu Asada, et al. “Structure of the dopamine D2 receptor in complex with the antipsychotic drug spiperone”. In: *Nature communications* 11.1 (2020), p. 6442.
- [31] Hugo Juárez Olgun, David Calderón Guzmán, Ernestina Hernández Garcia, and Gerardo Barragán Mejía. “The role of dopamine and its dysfunction as a consequence of oxidative stress”. In: *Oxidative medicine and cellular longevity* 2016.1 (2016), p. 9730467.
- [32] Mathias Uhlén, Linn Fagerberg, Björn M Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt, Anna Asplund, et al. “Tissue-based map of the human proteome”. In: *Science* 347.6220 (2015), p. 1260419.
- [33] Oliver D Howes, Bernard R Bukala, and Katherine Beck. “Schizophrenia: from neurochemistry to circuits, symptoms and treatments”. In: *Nature Reviews Neurology* 20.1 (2024), pp. 22–35.

- [34] Fulvio Lauretani, Francesco Giallauria, Crescenzo Testa, Claudia Zinni, Beatrice Lorenzi, Irene Zucchini, Marco Salvi, Raffaele Napoli, and Marcello Giuseppe Maggio. “Dopamine Pharmacodynamics: New Insights”. In: *International Journal of Molecular Sciences* 25.10 (2024), p. 5293.
- [35] Ulrich Omasits, Christian H Ahrens, Sebastian Müller, and Bernd Wollscheid. “Protter: interactive protein feature visualization and integration with experimental proteomic data”. In: *Bioinformatics* 30.6 (2014), pp. 884–886.
- [36] Rinshi S Kasai, Shuichi V Ito, Ryo M Awane, Takahiro K Fujiwara, and Akihiro Kusumi. “The class-A GPCR dopamine D2 receptor forms transient dimers stabilized by agonists: detection by single-molecule tracking”. In: *Cell biochemistry and biophysics* 76 (2018), pp. 29–37.
- [37] “UniProt: the universal protein knowledgebase in 2023”. In: *Nucleic acids research* 51.D1 (2023), pp. D523–D531.
- [38] Arvid Carlsson. “Antipsychotic drugs, neurotransmitters, and schizophrenia”. In: *American Journal of Psychiatry* 135.2 (1978), pp. 164–173.
- [39] Dong Guo, Julia M Hillger, Adriaan P IJzerman, and Laura H Heitman. “Drug-target residence time—a case for G protein-coupled receptors”. In: *Medicinal research reviews* 34.4 (2014), pp. 856–892.
- [40] Valerio Ricci, Domenico De Berardis, and Giuseppe Maina. “Third-generation antipsychotics and lurasidone in the treatment of substance-induced psychoses: A narrative review”. In: *Healthcare*. Vol. 12. 3. MDPI. 2024, p. 339.
- [41] Magdalena Markowicz-Piasecka, Marcin Kubisiak, Katarzyna Asendrych-Wicik, Michał Kołodziejczyk, Joanna Grzelińska, Małgorzata Fabijańska, and Tomasz Pietrzak. “Long-acting injectable antipsychotics—a review on formulation and in vitro dissolution”. In: *Pharmaceutics* 16.1 (2023), p. 28.
- [42] Juan A Ballesteros and Harel Weinstein. “[19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors”. In: *Methods in neurosciences*. Vol. 25. Elsevier, 1995, pp. 366–428.
- [43] Youwen Zhuang, Peiyu Xu, Chunyou Mao, Lei Wang, Brian Krumm, X Edward Zhou, Sijie Huang, Heng Liu, Xi Cheng, Xi-Ping Huang, et al. “Structural insights into the human D1 and D2 dopamine receptor signaling complexes”. In: *Cell* 184.4 (2021), pp. 931–942.
- [44] LLC Schrodinger. “The PyMOL molecular graphics system”. In: *Version 1* (2015), p. 8.
- [45] Vsevolod Katritch, Gustavo Fenalti, Enrique E Abola, Bryan L Roth, Vadim Cherezov, and Raymond C Stevens. “Allosteric sodium in class A GPCR signaling”. In: *Trends in biochemical sciences* 39.5 (2014), pp. 233–244.
- [46] J Robert Lane, Ara M Abramyan, Pramisha Adhikari, Alastair C Keen, Kuo-Hao Lee, Julie Sanchez, Ravi Kumar Verma, Herman D Lim, Hideaki Yano, Jonathan A Javitch, et al. “Distinct inactive conformations of the dopamine D2 and D3 receptors correspond to different extents of inverse agonism”. In: *Elife* 9 (2020), e52189.
- [47] Sandeep Sheth, Rafael Brito, Debashree Mukherjee, Leonard P Rybak, and Vickram Ramkumar. “Adenosine receptors: expression, function and regulation”. In: *International journal of molecular sciences* 15.2 (2014), pp. 2024–2052.
- [48] Mercedes Garcia-Gil, Marcella Camici, Simone Allegrini, Rossana Pesi, and Maria Grazia Tozzi. “Metabolic aspects of adenosine functions in the brain”. In: *Frontiers in pharmacology* 12 (2021), p. 672182.
- [49] Hai-Ying Shen and Jiang-Fan Chen. “Adenosine A2A receptors in psychopharmacology: modulators of behavior, mood and cognition”. In: *Current neuropharmacology* 7.3 (2009), pp. 195–206.
- [50] Annalisa Pinna, Jadwiga Wardas, Maria R Domenici, Patrizia Popoli, Giovanni Cossu, and Micaela Morelli. “Control of Motor Function by Adenosine A2A Receptors in Parkinson’s and Huntington’s Disease”. In: *Adenosine Receptors in Neurodegenerative Diseases*. Elsevier, 2017, pp. 187–213.
- [51] Agata Faron-Górecka, Marta Szlachta, Magdalena Kolasa, Joanna Solich, Andrzej Górecki, Maciej Kuśmider, Dariusz Żurawek, and Marta Dziedzicka-Wasylewska. “Understanding GPCR dimerization”. In: *Methods in Cell Biology*. Vol. 149. Elsevier, 2019, pp. 155–178.

- [52] Raphael Rozenfeld and Lakshmi A Devi. “Receptor heteromerization and drug discovery”. In: *Trends in pharmacological sciences* 31.3 (2010), pp. 124–130.
- [53] Daniel Pulido, Veronica Casado-Anguera, Laura Perez-Benito, Estefania Moreno, Arnau Cordomí, Laura Lopez, Antoni Cortes, Sergi Ferre, Leonardo Pardo, Vicent Casado, et al. “Design of a true bivalent ligand with picomolar binding affinity for a G protein-coupled receptor homodimer”. In: *Journal of medicinal chemistry* 61.20 (2018), pp. 9335–9346.
- [54] Carmelo Romano, Wan-Lin Yang, and Karen L O’Malley. “Metabotropic glutamate receptor 5 is a disulfide-linked dimer”. In: *Journal of Biological Chemistry* 271.45 (1996), pp. 28612–28616.
- [55] Wen Guo, Lei Shi, and Jonathan A Javitch. “The fourth transmembrane segment forms the interface of the dopamine D2 receptor homodimer”. In: *Journal of Biological Chemistry* 278.7 (2003), pp. 4385–4388.
- [56] Mei Bai, Sunita Trivedi, and Edward M Brown. “Dimerization of the extracellular calcium-sensing receptor (CaR) on the cell surface of CaR-transfected HEK293 cells”. In: *Journal of Biological Chemistry* 273.36 (1998), pp. 23605–23610.
- [57] Kenneth A Jones, Beth Borowsky, Joe A Tamm, Douglas A Craig, Margaret M Durkin, Meng Dai, Wen-Jeng Yao, Mary Johnson, Caryn Gunwaldsen, Ling-Yan Huang, et al. “GABAB receptors function as a heteromeric assembly of the subunits GABABR1 and GABABR2”. In: *Nature* 396.6712 (1998), pp. 674–679.
- [58] Sergi Ferré, Jordi Bonaventura, Dardo Tomasi, Gemma Navarro, Estefania Moreno, Antonio Cortés, Carme Llus, Vicent Casadó, and Nora D Volkow. “Allosteric mechanisms within the adenosine A2A–dopamine D2 receptor heterotetramer”. In: *Neuropharmacology* 104 (2016), pp. 154–160.
- [59] Jordi Bonaventura, Gemma Navarro, Verònica Casadó-Anguera, Karima Azdad, William Rea, Estefania Moreno, Marc Brugarolas, Josefa Mallol, Enric I Canela, Carme Llus, et al. “Allosteric interactions between agonists and antagonists within the adenosine A2A receptor-dopamine D2 receptor heterotetramer”. In: *Proceedings of the National Academy of Sciences* 112.27 (2015), E3609–E3618.
- [60] Damien Thevenin, Tzvetana Lazarova, Matthew F Roberts, and Clifford R Robinson. “Oligomerization of the fifth transmembrane domain from the adenosine A2A receptor”. In: *Protein science* 14.8 (2005), pp. 2177–2186.
- [61] Alina Tabor, Siegfried Weisenburger, Ashutosh Banerjee, Nirupam Purkayastha, Jonas M Kaindl, Harald Hübner, Luxi Wei, Teja W Grömer, Johannes Kornhuber, Nuska Tschammer, et al. “Visualization and ligand-induced modulation of dopamine receptor dimerization at the single molecule level”. In: *Scientific reports* 6.1 (2016), p. 33233.
- [62] Rahul Roy, Sungchul Hohng, and Taekjip Ha. “A practical guide to single-molecule FRET”. In: *Nature methods* 5.6 (2008), pp. 507–516.
- [63] Kevin DG Pfleger and Karin A Eidne. “Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET)”. In: *Nature methods* 3.3 (2006), pp. 165–174.
- [64] Davide Calebiro, Finn Rieken, Julia Wagner, Titiwat Sungkaworn, Ulrike Zabel, Alfio Borzi, Emanuele Cocucci, Alexander Zürn, and Martin J Lohse. “Single-molecule analysis of fluorescently labeled G-protein-coupled receptors reveals complexes with distinct dynamics and organization”. In: *Proceedings of the National Academy of Sciences* 110.2 (2013), pp. 743–748.
- [65] Jonathan A Hern, Asma H Baig, Gregory I Mashanov, Berry Birdsall, John ET Corrie, Sebastian Lazareno, Justin E Molloy, and Nigel JM Birdsall. “Formation and dissociation of M1 muscarinic receptor dimers seen by total internal reflection fluorescence imaging of single molecules”. In: *Proceedings of the National Academy of Sciences* 107.6 (2010), pp. 2693–2698.
- [66] Rinshi S Kasai, Kenichi GN Suzuki, Eric R Prossnitz, Ikuko Koyama-Honda, Chieko Nakada, Takahiro K Fujiwara, and Akihiro Kusumi. “Full characterization of GPCR monomer–dimer dynamic equilibrium by single molecule imaging”. In: *Journal of Cell Biology* 192.3 (2011), pp. 463–480.
- [67] Marco Carli, Shivakumar Kolachalam, Stefano Aringhieri, Mario Rossi, Luca Giovannini, Roberto Maggio, and Marco Scarselli. “Dopamine D2 receptors dimers: how can we pharmacologically target them?” In: *Current neuropharmacology* 16.2 (2018), pp. 222–230.

- [68] Jean-François Mercier, Ali Salahpour, Stéphane Angers, Andreas Breit, and Michel Bouvier. “Quantitative assessment of β 1-and β 2-adrenergic receptor homo-and heterodimerization by bioluminescence resonance energy transfer”. In: *Journal of Biological Chemistry* 277.47 (2002), pp. 44925–44931.
- [69] Tamara Ullmann, Marie Gienger, Julian Budzinski, Jan Hellmann, Harald Hübner, Peter Gmeiner, and Dorothee Weikert. “Homobivalent dopamine D2 receptor ligands modulate the dynamic equilibrium of D2 monomers and homo-and heterodimers”. In: *ACS Chemical Biology* 16.2 (2021), pp. 371–379.
- [70] Paul Zawarynski, Teresa Tallerico, Philip Seeman, Samuel P Lee, Brian F O’Dowd, and Susan R George. “Dopamine D2 receptor dimers in human and rat brain”. In: *FEBS letters* 441.3 (1998), pp. 383–386.
- [71] Meritxell Canals, Javier Burgueño, Daniel Marcellino, Núria Cabello, Enric I Canela, Josefa Mallol, Luigi Agnati, Sergi Ferré, Michel Bouvier, Kjell Fuxe, et al. “Homodimerization of adenosine A2A receptors: qualitative and quantitative assessment by fluorescence and bioluminescence energy transfer”. In: *Journal of neurochemistry* 88.3 (2004), pp. 726–734.
- [72] Khanh Dinh Quoc Nguyen, Michael Vigers, Eric Sefah, Susanna Seppälä, Jennifer Paige Hoover, Nicole Star Schonenbach, Blake Mertz, Michelle Ann O’Malley, and Songi Han. “Homo-oligomerization of the human adenosine A2A receptor is driven by the intrinsically disordered C-terminus”. In: *Elife* 10 (2021), e66662.
- [73] Toshio Kamiya, Osamu Saitoh, Kazuaki Yoshioka, and Hiroyasu Nakata. “Oligomerization of adenosine A2A and dopamine D2 receptors in living cells”. In: *Biochemical and biophysical research communications* 306.2 (2003), pp. 544–549.
- [74] Nicole S Schonenbach, Monica D Rieth, Songi Han, and Michelle A O’Malley. “Adenosine A2a receptors form distinct oligomers in protein detergent complexes”. In: *FEBS letters* 590.18 (2016), pp. 3295–3306.
- [75] Rafael Franco, Meritxell Canals, Daniel Marcellino, Sergi Ferré, Luigi Agnati, Josefa Mallol, Vicent Casadó, Francisco Ciruela, Kjell Fuxe, Carmen Lluís, et al. “Regulation of heptaspanning-membrane-receptor function by dimerization and clustering”. In: *Trends in biochemical sciences* 28.5 (2003), pp. 238–243.
- [76] Qingtong Zhou, Dehua Yang, Meng Wu, Yu Guo, Wanjing Guo, Li Zhong, Xiaoqing Cai, Antao Dai, Wonjo Jang, Eugene I Shakhnovich, et al. “Common activation mechanism of class A GPCRs”. In: *Elife* 8 (2019), e50279.
- [77] S Jonathan Singer and Garth L Nicolson. “The Fluid Mosaic Model of the Structure of Cell Membranes: Cell membranes are viewed as two-dimensional solutions of oriented globular proteins and lipids.” In: *Science* 175.4023 (1972), pp. 720–731.
- [78] Garth L Nicolson. “The Fluid—Mosaic Model of Membrane Structure: Still relevant to understanding the structure, function and dynamics of biological membranes after more than 40 years”. In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1838.6 (2014), pp. 1451–1466.
- [79] Meritxell Canals, Daniel Marcellino, Francesca Fanelli, Francisco Ciruela, Piero De Benedetti, Steven R Goldberg, Kim Neve, Kjell Fuxe, Luigi F Agnati, Amina S Woods, et al. “Adenosine A2A-dopamine D2 receptor-receptor heteromerization: qualitative and quantitative assessment by fluorescence and bioluminescence energy transfer”. In: *Journal of Biological Chemistry* 278.47 (2003), pp. 46741–46749.
- [80] Kavya Prasad, Erik FJ de Vries, Philip H Elsinga, Rudi AJO Dierckx, and Aren van Waarde. “Allosteric interactions between adenosine A2A and dopamine D2 receptors in heteromeric complexes: biochemical and pharmacological characteristics, and opportunities for PET imaging”. In: *International Journal of Molecular Sciences* 22.4 (2021), p. 1719.
- [81] Ivone Gomes, Mohammed Akli Ayoub, Wakako Fujita, Werner C Jaeger, Kevin DG Pflieger, and Lakshmi A Devi. “G protein-coupled receptor heteromers”. In: *Annual review of pharmacology and toxicology* 56.1 (2016), pp. 403–425.
- [82] M Villar Van Anthony, Santiago Cuevas, Xiaoxu Zheng, and Pedro A Jose. “Localization and signaling of GPCRs in lipid rafts”. In: *Methods in cell biology*. Vol. 132. Elsevier, 2016, pp. 3–23.
- [83] Linda J Pike. “Lipid rafts: bringing order to chaos”. In: *Journal of lipid research* 44.4 (2003), pp. 655–667.

- [84] Mohammad Fallahi-Sichani and Jennifer J Linderman. “Lipid raft-mediated regulation of G-protein coupled receptor signaling by ligands which influence receptor dimerization: a computational study”. In: *PloS one* 4.8 (2009), e6604.
- [85] Rim Baccouch, Estelle Rascol, Kaja Stoklosa, and Isabel D Alves. “The role of the lipid environment in the activity of G protein coupled receptors”. In: *Biophysical chemistry* 285 (2022), p. 106794.
- [86] Sergi Ferré, Jordi Bonaventura, Wendy Zhu, Candice Hatcher-Solis, Jaume Taura, César Quiroz, Ning-Sheng Cai, Estefania Moreno, Veronica Casado-Anguera, Alexxai V Kravitz, et al. “Essential control of the function of the striatopallidal neuron by pre-coupled complexes of adenosine A2A-dopamine D2 receptor heterotetramers and adenylyl cyclase”. In: *Frontiers in pharmacology* 9 (2018), p. 243.
- [87] Luigi Francesco Agnati, Giuseppina Leo, Susanna Genedani, Nicola Andreoli, D Marcellino, A Woods, L Piron, Diego Guidolin, and K Fuxe. “Structural plasticity in G-protein coupled receptors as demonstrated by the allosteric actions of homocysteine and computer-assisted analysis of disordered domains”. In: *Brain research reviews* 58.2 (2008), pp. 459–474.
- [88] Christa E Müller and Sergi Ferré. “Blocking striatal adenosine A2A receptors: a new strategy for basal ganglia disorders.” In: *Frontiers in CNS Drug Discov* 1 (2010), pp. 304–341.
- [89] Rigney E Turnham and John D Scott. “Protein kinase A catalytic subunit isoform PRKACA; History, function and physiology”. In: *Gene* 577.2 (2016), pp. 101–108.
- [90] Ganesh Kadamur and Elliott M Ross. “Mammalian phospholipase C”. In: *Annual review of physiology* 75 (2013), pp. 127–154.
- [91] Sergi Ferré, Vicent Casadó, Lakshmi A Devi, Marta Filizola, Ralf Jockers, Martin J Lohse, Graeme Milligan, Jean-Philippe Pin, and Xavier Guitart. “G protein-coupled receptor oligomerization revisited: functional and pharmacological perspectives”. In: *Pharmacological reviews* 66.2 (2014), pp. 413–434.
- [92] Gemma Navarro, Arnau Cordero, Verónica Casadó-Anguera, Estefania Moreno, Ning-Sheng Cai, Antoni Cortés, Enric I Canela, Carmen W Dessauer, Vicent Casadó, Leonardo Pardo, et al. “Evidence for functional pre-coupled complexes of receptor heteromers and adenylyl cyclase”. In: *Nature communications* 9.1 (2018), p. 1242.
- [93] Ko-Woon Lee, Jang-Hee Hong, In Young Choi, Yongzhe Che, Ja-Kyeong Lee, Sung-Don Yang, Chang-Woo Song, Ho Sung Kang, Jae-Heun Lee, Jai Sung Noh, et al. “Impaired D2 dopamine receptor function in mice lacking type 5 adenylyl cyclase”. In: *Journal of Neuroscience* 22.18 (2002), pp. 7931–7940.
- [94] Robert A Copeland. “The drug-target residence time model: a 10-year retrospective”. In: *Nature Reviews Drug Discovery* 15.2 (2016), pp. 87–95.
- [95] Ali S Faqi. *A comprehensive guide to toxicology in nonclinical drug development*. Academic Press, 2016.
- [96] Christofer S Tautermann. “Impact, determination and prediction of drug-receptor residence times for GPCRs”. In: *Current Opinion in Pharmacology* 30 (2016), pp. 22–26.
- [97] David C Swinney, Brad A Haubrich, Isabelle Van Liefde, and Georges Vauquelin. “The role of binding kinetics in GPCR drug discovery”. In: *Current Topics in Medicinal Chemistry* 15.24 (2015), pp. 2504–2522.
- [98] Peter J Tummino and Robert A Copeland. “Residence time of receptor- ligand complexes and its effect on biological function”. In: *Biochemistry* 47.20 (2008), pp. 5481–5492.
- [99] Xavier Langlois, Anton Megens, Hilde Lavreysen, John Atack, Miroslav Cik, Paula te Riele, Luc Peeters, Ria Wouters, Jef Vermeire, Herman Hendrickx, et al. “Pharmacology of JNJ-37822681, a specific and fast-dissociating D2 antagonist for the treatment of schizophrenia”. In: *Journal of Pharmacology and Experimental Therapeutics* 342.1 (2012), pp. 91–105.
- [100] JE Leysen, W Gommeren, A Eens, D De Chaffoy De Courcelles, JC Stoof, and PA Janssen. “Biochemical profile of risperidone, a new antipsychotic.” In: *Journal of Pharmacology and Experimental Therapeutics* 247.2 (1988), pp. 661–670.
- [101] Shitij Kapur and Philip Seeman. “Antipsychotic agents differ in how fast they come off the dopamine D2 receptors. Implications for atypical antipsychotic action.” In: *Journal of Psychiatry and Neuroscience* 25.2 (2000), p. 161.

- [102] Herbert Y Meltzer. “Multireceptor atypical antipsychotic drugs”. In: *Atypical antipsychotics* (2000), pp. 191–213.
- [103] Elisabeth Coudert, Sebastien Gehant, Edouard de Castro, Monica Pozzato, Delphine Baratin, Teresa Neto, Christian JA Sigrist, Nicole Redaschi, and Alan Bridge. “Annotation of biologically relevant ligands in UniProtKB using ChEBI”. In: *Bioinformatics* 39.1 (2023), btac793.
- [104] Fang-Yu Lin and Alexander D MacKerell. “Force fields for small molecules”. In: *Biomolecular simulations: Methods and protocols* (2019), pp. 21–54.
- [105] Mariusz Jaskolski, Zbigniew Dauter, and Alexander Wlodawer. “A brief history of macromolecular crystallography, illustrated by a family tree and its N obel fruits”. In: *The FEBS journal* 281.18 (2014), pp. 3985–4009.
- [106] Muhammed Tilahun Muhammed and Esin Aki-Yalcin. “Homology modeling in drug discovery: Overview, current applications, and future perspectives”. In: *Chemical biology & drug design* 93.1 (2019), pp. 12–20.
- [107] CA Floudas, HK Fung, SR McAllister, M Mönnigmann, and R Rajgaria. “Advances in protein structure prediction and de novo protein design: A review”. In: *Chemical Engineering Science* 61.3 (2006), pp. 966–988.
- [108] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Židek, Anna Potapenko, et al. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596.7873 (2021), pp. 583–589.
- [109] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. “Accurate structure prediction of biomolecular interactions with AlphaFold 3”. In: *Nature* (2024), pp. 1–3.
- [110] Jenny Gu and Philip E Bourne. *Structural bioinformatics*. Vol. 44. John Wiley & Sons, 2009.
- [111] Arthur M Lesk and Cyrus Chothia. “How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins”. In: *Journal of molecular biology* 136.3 (1980), pp. 225–270.
- [112] Stephen K Burley, Helen M Berman, Charmi Bhikadiya, Chunxiao Bi, Li Chen, Luigi Di Costanzo, Cole Christie, Ken Dalenberg, Jose M Duarte, Shuchismita Dutta, et al. “RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy”. In: *Nucleic acids research* 47.D1 (2019), pp. D464–D474.
- [113] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. “The protein data bank”. In: *Nucleic acids research* 28.1 (2000), pp. 235–242.
- [114] Benjamin Webb and Andrej Šali. “Comparative protein structure modeling using MODELLER”. In: *Current protocols in bioinformatics* 54.1 (2016), pp. 5–6.
- [115] Andrej Šali and Tom L Blundell. “Comparative protein modelling by satisfaction of spatial restraints”. In: *Journal of molecular biology* 234.3 (1993), pp. 779–815.
- [116] Robert C Edgar. “MUSCLE: multiple sequence alignment with high accuracy and high throughput”. In: *Nucleic acids research* 32.5 (2004), pp. 1792–1797.
- [117] Fabian Sievers and Desmond G Higgins. “Clustal omega”. In: *Current protocols in bioinformatics* 48.1 (2014), pp. 3–13.
- [118] Nadira Khatoon, Sushanth Adusumilli, Poulomi Dey, Rachita Sharma, Pradeepti Kampani, Jayasha Shandilya, and Tapan K Nayak. “Protein engineering and design in ion channels and receptors”. In: *Methods in Cell Biology*. Vol. 169. Elsevier, 2022, pp. 143–168.
- [119] Min-yi Shen and Andrej Šali. “Statistical potential for assessment and prediction of protein structures”. In: *Protein science* 15.11 (2006), pp. 2507–2524.
- [120] Anne Louise Morris, Malcolm W MacArthur, E Gail Hutchinson, and Janet M Thornton. “Stereochemical quality of protein structure coordinates”. In: *Proteins: Structure, Function, and Bioinformatics* 12.4 (1992), pp. 345–364.

- [121] Roman A Laskowski, Malcolm W MacArthur, David S Moss, and Janet M Thornton. "PROCHECK: a program to check the stereochemical quality of protein structures". In: *Journal of applied crystallography* 26.2 (1993), pp. 283–291.
- [122] David J Griffiths and Darrell F Schroeter. *Introduction to quantum mechanics*. Cambridge university press, 2018.
- [123] Andrew R Leach. *Molecular modelling: principles and applications*. Pearson education, 2001.
- [124] Max Born and W Heisenberg. "Zur quantentheorie der molekeln". In: *Original Scientific Papers Wissenschaftliche Originalarbeiten* (1985), pp. 216–246.
- [125] Hans Martin Senn and Walter Thiel. "QM/MM methods for biomolecular systems". In: *Angewandte Chemie International Edition* 48.7 (2009), pp. 1198–1229.
- [126] Normand M Laurendeau. *Statistical thermodynamics: fundamentals and applications*. Cambridge University Press, 2005.
- [127] Michel A Cuendet and Wilfred F van Gunsteren. "On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm". In: *The Journal of chemical physics* 127.18 (2007).
- [128] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Vol. 1. Academic Press San Diego, 2002.
- [129] P Bauer, B Hess, and E Lindahl. "GROMACS 2022.4 Manual". In: *November* 16 (2022), p. 2022.
- [130] Herman JC Berendsen, JPM van Postma, Wilfred F Van Gunsteren, ARHJ DiNola, and Jan R Haak. "Molecular dynamics with coupling to an external bath". In: *The Journal of chemical physics* 81.8 (1984), pp. 3684–3690.
- [131] Djurre H De Jong, Svetlana Baoukina, Helgi I Ingólfsson, and Siewert J Marrink. "Martini straight: Boosting performance using a shorter cutoff and GPUs". In: *Computer Physics Communications* 199 (2016), pp. 1–7.
- [132] Giovanni Bussi, Davide Donadio, and Michele Parrinello. "Canonical sampling through velocity rescaling". In: *The Journal of chemical physics* 126.1 (2007).
- [133] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". In: *SoftwareX* 1 (2015), pp. 19–25.
- [134] Michele Parrinello and Aneesur Rahman. "Polymorphic transitions in single crystals: A new molecular dynamics method". In: *Journal of Applied physics* 52.12 (1981), pp. 7182–7190.
- [135] Hans C Andersen. "Molecular dynamics simulations at constant pressure and/or temperature". In: *The Journal of chemical physics* 72.4 (1980), pp. 2384–2393.
- [136] Michael P Allen and Dominic J Tildesley. *Computer simulation of liquids*. Oxford university press, 2017.
- [137] Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. "LINCS: A linear constraint solver for molecular simulations". In: *Journal of computational chemistry* 18.12 (1997), pp. 1463–1472.
- [138] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *Journal of computational physics* 23.3 (1977), pp. 327–341.
- [139] Scott A Hollingsworth and Ron O Dror. "Molecular dynamics simulation for all". In: *Neuron* 99.6 (2018), pp. 1129–1143.
- [140] Rafael C Bernardi, Marcelo CR Melo, and Klaus Schulten. "Enhanced sampling techniques in molecular dynamics simulations of biological systems". In: *Biochimica et Biophysica Acta (BBA)-General Subjects* 1850.5 (2015), pp. 872–877.
- [141] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. "Well-tempered metadynamics: a smoothly converging and tunable free-energy method". In: *Physical review letters* 100.2 (2008), p. 020603.
- [142] Alessandro Laio and Michele Parrinello. "Escaping free-energy minima". In: *Proceedings of the national academy of sciences* 99.20 (2002), pp. 12562–12566.

- [143] Giovanni Bussi, Alessandro Laio, and Michele Parrinello. "Equilibrium free energies from nonequilibrium metadynamics". In: *Physical review letters* 96.9 (2006), p. 090601.
- [144] Alessandro Laio, Antonio Rodriguez-Fortea, Francesco Luigi Gervasio, Matteo Ceccarelli, and Michele Parrinello. "Assessing the accuracy of metadynamics". In: *The journal of physical chemistry B* 109.14 (2005), pp. 6714–6721.
- [145] James F Dama, Michele Parrinello, and Gregory A Voth. "Well-tempered metadynamics converges asymptotically". In: *Physical review letters* 112.24 (2014), p. 240602.
- [146] Luigi Bonati, Valerio Rizzi, and Michele Parrinello. "Data-driven collective variables for enhanced sampling". In: *The journal of physical chemistry letters* 11.8 (2020), pp. 2998–3004.
- [147] Michele Invernizzi and Michele Parrinello. "Exploration vs convergence speed in adaptive-bias enhanced sampling". In: *Journal of Chemical Theory and Computation* 18.6 (2022), pp. 3988–3996.
- [148] Gareth A Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. "PLUMED 2: New feathers for an old bird". In: *Computer physics communications* 185.2 (2014), pp. 604–613.
- [149] Paulo CT Souza, Riccardo Alessandri, Jonathan Barnoud, Sebastian Thallmair, Ignacio Faustino, Fabian Grünewald, Ilias Patmanidis, Haleh Abdizadeh, Bart MH Bruininks, Tsjerk A Wassenaar, et al. "Martini 3: a general purpose force field for coarse-grained molecular dynamics". In: *Nature methods* 18.4 (2021), pp. 382–388.
- [150] Daniele Angioletti, Stefano Raniolo, and Vittorio Limongelli. "HEroBM: A deep equivariant graph neural network for high-fidelity backmapping from coarse-grained to all-atom structures". In: *The Journal of Chemical Physics* 163.7 (2025).
- [151] Siewert J Marrink, H Jelger Risselada, Serge Yefimov, D Peter Tieleman, and Alex H De Vries. "The MARTINI force field: coarse grained model for biomolecular simulations". In: *The journal of physical chemistry B* 111.27 (2007), pp. 7812–7824.
- [152] MT Jeena, L Palanikumar, Eun Min Go, Inhye Kim, Myoung Gyun Kang, Seonik Lee, Sooham Park, Huyeon Choi, Chaekyu Kim, Seon-Mi Jin, et al. "Mitochondria localization induced self-assembly of peptide amphiphiles for cellular dysfunction". In: *Nature communications* 8.1 (2017), p. 26.
- [153] Zeineb Si Chaib, Alessandro Marchetto, Klevia Dishnica, Paolo Carloni, Alejandro Giorgetti, and Giulia Rossetti. "Impact of cholesterol on the stability of monomeric and dimeric forms of the translocator protein TSPO: A molecular simulation study". In: *Molecules* 25.18 (2020), p. 4299.
- [154] Jaakko J Uusitalo, Helgi I Ingólfsson, Parisa Akhshi, D Peter Tieleman, and Siewert J Marrink. "Martini coarse-grained force field: extension to DNA". In: *Journal of chemical theory and computation* 11.8 (2015), pp. 3932–3945.
- [155] Helgi I Ingólfsson, Cesar A Lopez, Jaakko J Uusitalo, Djurre H de Jong, Srinivasa M Gopal, Xavier Periole, and Siewert J Marrink. "The power of coarse graining in biomolecular simulations". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 4.3 (2014), pp. 225–248.
- [156] Alessandro Marchetto, Zeineb Si Chaib, Carlo Alberto Rossi, Rui Ribeiro, Sergio Pantano, Giulia Rossetti, and Alejandro Giorgetti. "CGMD platform: integrated web servers for the preparation, running, and analysis of coarse-grained molecular dynamics simulations". In: *Molecules* 25.24 (2020), p. 5934.
- [157] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. "A smooth particle mesh Ewald method". In: *The Journal of chemical physics* 103.19 (1995), pp. 8577–8593.
- [158] Riccardo Alessandri, Paulo CT Souza, Sebastian Thallmair, Manuel N Melo, Alex H De Vries, and Siewert J Marrink. "Pitfalls of the Martini model". In: *Journal of chemical theory and computation* 15.10 (2019), pp. 5448–5460.
- [159] Xavier Periole, Marco Cavalli, Siewert-Jan Marrink, and Marco A Ceruso. "Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition". In: *Journal of chemical theory and computation* 5.9 (2009), pp. 2531–2543.

- [160] PC Kroon, F Grunewald, J Barnoud, Mv Tilburg, PCT Souza, TA Wassenaar, and SJ Marrink. "Martinize2 and vermouth: unified framework for topology generation. Elife 12". In: *RP90627* (2023).
- [161] Wolfgang Kabsch and Christian Sander. "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features". In: *Biopolymers: Original Research on Biomolecules* 22.12 (1983), pp. 2577–2637.
- [162] Siewert J Marrink and D Peter Tieleman. "Perspective on the Martini model". In: *Chemical Society Reviews* 42.16 (2013), pp. 6801–6822.
- [163] Sunhwan Jo, Taehoon Kim, Vidyashankara G Iyer, and Wonpil Im. "CHARMM-GUI: a web-based graphical user interface for CHARMM". In: *Journal of computational chemistry* 29.11 (2008), pp. 1859–1865.
- [164] Tsjerk A Wassenaar, Helgi I Ingólfsson, Rainer A Bockmann, D Peter Tieleman, and Siewert J Marrink. "Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations". In: *Journal of chemical theory and computation* 11.5 (2015), pp. 2144–2155.
- [165] Mikhail A Lomize, Irina D Pogozheva, Hyeon Joo, Henry I Mosberg, and Andrei L Lomize. "OPM database and PPM web server: resources for positioning of proteins in membranes". In: *Nucleic acids research* 40.D1 (2012), pp. D370–D376.
- [166] EJ Haug, JS Arora, and K Matsui. "A steepest-descent method for optimization of mechanical systems". In: *Journal of Optimization Theory and Applications* 19 (1976), pp. 401–424.
- [167] Loup Verlet. "Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules". In: *Physical review* 159.1 (1967), p. 98.
- [168] John A Barker and Robert O Watts. "Monte Carlo studies of the dielectric properties of water-like models". In: *Molecular Physics* 26.3 (1973), pp. 789–792.
- [169] Daniele Di Marino, Paolo Conflitti, Stefano Motta, and Vittorio Limongelli. "Structural basis of dimerization of chemokine receptors CCR5 and CXCR4". In: *Nature Communications* 14.1 (2023), p. 6439.
- [170] Robin A Corey, Owen N Vickery, Mark SP Sansom, and Phillip J Stansfeld. "Insights into membrane protein–lipid interactions from free energy calculations". In: *Journal of chemical theory and computation* 15.10 (2019), pp. 5727–5736.
- [171] Valentina Corradi, Eduardo Mendez-Villuendas, Helgi I Ingólfsson, Ruo-Xu Gu, Iwona Siuda, Manuel N Melo, Anastassia Moussatova, Lucien J DeGagné, Besian I Sejdiu, Gurpreet Singh, et al. "Lipid–protein interactions are unique fingerprints for membrane proteins". In: *ACS central science* 4.6 (2018), pp. 709–717.
- [172] Owen N Vickery and Phillip J Stansfeld. "CG2AT2: an enhanced fragment-based approach for serial multi-scale molecular dynamics simulations". In: *Journal of chemical theory and computation* 17.10 (2021), pp. 6472–6482.
- [173] Siewert J Marrink, Luca Monticelli, Manuel N Melo, Riccardo Alessandri, D Peter Tieleman, and Paulo CT Souza. "Two decades of Martini: Better beads, broader scope". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 13.1 (2023), e1620.
- [174] Tsjerk A Wassenaar, Kristyna Pluhackova, Rainer A Bockmann, Siewert J Marrink, and D Peter Tieleman. "Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models". In: *Journal of chemical theory and computation* 10.2 (2014), pp. 676–690.
- [175] Keverne A Louison, Ian L Dryden, and Charles A Laughton. "GLIMPS: a machine learning approach to resolution transformation for multiscale modeling". In: *Journal of Chemical Theory and Computation* 17.12 (2021), pp. 7930–7937.
- [176] Lisbeth R Kjølbye, Gilberto P Pereira, Alessio Bartocci, Martina Pannuzzo, Simone Albani, Alessandro Marchetto, Brian Jiménez-García, Juliette Martin, Giulia Rossetti, Marco Cecchini, et al. "Towards design of drugs and delivery systems with the Martini coarse-grained model". In: *QRB discovery* 3 (2022), e19.
- [177] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. "Improved side-chain torsion potentials for the Amber ff99SB protein force field". In: *Proteins: Structure, Function, and Bioinformatics* 78.8 (2010), pp. 1950–1958.

- [178] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L De Groot, Helmut Grubmüller, and Alexander D MacKerell Jr. “CHARMM36m: an improved force field for folded and intrinsically disordered proteins”. In: *Nature methods* 14.1 (2017), pp. 71–73.
- [179] George A Kaminski, Richard A Friesner, Julian Tirado-Rives, and William L Jorgensen. “Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides”. In: *The Journal of Physical Chemistry B* 105.28 (2001), pp. 6474–6487.
- [180] Chris Oostenbrink, Alessandra Villa, Alan E Mark, and Wilfred F Van Gunsteren. “A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6”. In: *Journal of computational chemistry* 25.13 (2004), pp. 1656–1676.
- [181] Riccardo Alessandri, Jonathan Barnoud, Anders S Gertsen, Ilias Patmanidis, Alex H de Vries, Paulo CT Souza, and Siewert J Marrink. “Martini 3 coarse-grained force field: small molecules”. In: *Advanced Theory and Simulations* 5.1 (2022), p. 2100391.
- [182] Luis Borges-Araújo, Ana C Borges-Araújo, Tugba Nur Ozturk, Daniel P Ramirez-Echemendia, Balázs Fábián, Timothy S Carpenter, Sebastian Thallmair, Jonathan Barnoud, Helgi I Ingólfsson, Gerhard Hummer, et al. “Martini 3 coarse-grained force field for cholesterol”. In: *Journal of Chemical Theory and Computation* 19.20 (2023), pp. 7387–7404.
- [183] Sandor Vajda, Dmitri Beglov, Amanda E Wakefield, Megan Egbert, and Adrian Whitty. “Cryptic binding sites on proteins: definition, detection, and druggability”. In: *Current opinion in chemical biology* 44 (2018), pp. 1–8.
- [184] Vladimiras Oleinikovas, Giorgio Saladino, Benjamin P Cossins, and Francesco L Gervasio. “Understanding cryptic pocket formation in protein targets by enhanced sampling simulations”. In: *Journal of the American Chemical Society* 138.43 (2016), pp. 14257–14263.
- [185] James R Horn and Brian K Shoichet. “Allosteric inhibition through core disruption”. In: *Journal of molecular biology* 336.5 (2004), pp. 1283–1291.
- [186] Lei Zhong, Yueshan Li, Liang Xiong, Wenjing Wang, Ming Wu, Ting Yuan, Wei Yang, Chenyu Tian, Zhuang Miao, Tianqi Wang, et al. “Small molecules in targeted cancer therapy: advances, challenges, and future perspectives”. In: *Signal transduction and targeted therapy* 6.1 (2021), pp. 1–48.
- [187] MI Zimmerman, JR Porter, MD Ward, S Singh, N Vithani, A Meller, and UL Mallimadugula. “989 Kuhn, CE; Borowsky, JH; Wiewiora, RP, SARS-CoV-2 simulations go exascale to predict dramatic 990 spike opening and cryptic pockets across the proteome”. In: *Nat Chem* 13 (2021), pp. 651–659.
- [188] David E Shaw, Peter J Adams, Asaph Azaria, Joseph A Bank, Brannon Batson, Alistair Bell, Michael Bergdorf, Jhanvi Bhatt, J Adam Butts, Timothy Correia, et al. “Anton 3: twenty microseconds of molecular dynamics simulation before lunch”. In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. 2021, pp. 1–11.
- [189] Geoffrey M Gray, Ning Ma, Carl E Wagner, and Arjan van der Vaart. “Molecular dynamics simulations and molecular flooding studies of the retinoid X-receptor ligand binding domain”. In: *Journal of molecular modeling* 23 (2017), pp. 1–9.
- [190] Rommie E Amaro and Wilfred W Li. “Emerging methods for ensemble-based virtual screening”. In: *Current topics in medicinal chemistry* 10.1 (2010), pp. 3–13.
- [191] Antonija Kuzmanic, Gregory R Bowman, Jordi Juarez-Jimenez, Julien Michel, and Francesco L Gervasio. “Investigating cryptic binding sites by molecular dynamics simulations”. In: *Accounts of chemical research* 53.3 (2020), pp. 654–661.
- [192] Adolfo B Poma, Marek Cieplak, and Panagiotis E Theodorakis. “Combining the MARTINI and structure-based coarse-grained approaches for the molecular dynamics studies of conformational transitions in proteins”. In: *Journal of Chemical Theory and Computation* 13.3 (2017), pp. 1366–1374.

- [193] Ryo Kanada, Kei Terayama, Atsushi Tokuhisa, Shigeyuki Matsumoto, and Yasushi Okuno. “Enhanced conformational sampling with an adaptive coarse-grained elastic network model using short-time all-atom molecular dynamics”. In: *Journal of Chemical Theory and Computation* 18.4 (2022), pp. 2062–2074.
- [194] Luis Borges-Araújo, Gilberto P Pereira, Mariana Valério, and Paulo CT Souza. “Assessing the martini 3 protein model: A review of its path and potential”. In: *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* (2024), p. 141014.
- [195] Tristan Bereau and Kurt Kremer. “Automated parametrization of the coarse-grained Martini force field for small organic molecules”. In: *Journal of chemical theory and computation* 11.6 (2015), pp. 2783–2791.
- [196] Charly Empereur-Mot, Luca Pesce, Giovanni Doni, Davide Bochicchio, Riccardo Capelli, Claudio Perego, and Giovanni M Pavan. “Swarm-CG: automatic parametrization of bonded terms in MARTINI-based coarse-grained models of simple to complex molecules via fuzzy self-tuning particle swarm optimization”. In: *ACS omega* 5.50 (2020), pp. 32823–32843.
- [197] James A Graham, Jonathan W Essex, and Syma Khalid. “PyCGTOOL: automated generation of coarse-grained molecular dynamics models from atomistic trajectories”. In: *Journal of chemical information and modeling* 57.4 (2017), pp. 650–656.
- [198] Magdalena Szczuka, Gilberto P Pereira, Luis J Walter, Marc Gueroult, Pierre Poulain, Tristan Bereau, Paulo CT Souza, and Matthieu Chavent. “Fast parameterization of Martini3 models for fragments and small molecules”. In: *bioRxiv* (2025), pp. 2025–07.
- [199] Alessio Bartocci, Andrea Grazi, Nour Awad, Pierre-Jean Corringer, Paulo CT Souza, and Marco Cecchini. “A millisecond coarse-grained simulation approach to decipher allosteric cannabinoid binding at the glycine receptor α 1”. In: *Nature Communications* 15.1 (2024), p. 9040.
- [200] Valery Lutsyk, Pawel Wolski, and Wojciech Plazinski. “Extending the Martini 3 coarse-grained force field to carbohydrates”. In: *Journal of Chemical Theory and Computation* 18.8 (2022), pp. 5089–5107.
- [201] Barbora Wacławiková, Paulo Cesar Telles de Souza, Markus Schwalbe, Constantinos G Neochoritis, Warner Hoornenborg, Sieger A Nelemans, Siewert J Marrink, and Sahar El Aidy. “Potential binding modes of the gut bacterial metabolite, 5-hydroxyindole, to the intestinal L-type calcium channels and its impact on the microbiota in rats”. In: *Gut microbes* 15.1 (2023), p. 2154544.
- [202] Matias R Machado, Exequiel E Barrera, Florencia Klein, Martin Sónora, Steffano Silva, and Sergio Pantano. “The SIRAH 2.0 force field: altius, fortius, citius”. In: *Journal of chemical theory and computation* 15.4 (2019), pp. 2719–2733.
- [203] Fernando Luis Barroso da Silva, Fabio Sterpone, and Philippe Derreumaux. “OPEP6: A new constant-pH molecular dynamics simulation scheme with OPEP coarse-grained force field”. In: *Journal of Chemical Theory and Computation* 15.6 (2019), pp. 3875–3888.
- [204] Wataru Shinoda, Russell DeVane, and Michael L Klein. “Zwitterionic lipid assemblies: molecular dynamics studies of monolayers, bilayers, and vesicles using a new coarse grain force field”. In: *The Journal of Physical Chemistry B* 114.20 (2010), pp. 6836–6849.
- [205] Mario Orsi and Jonathan W Essex. “The ELBA force field for coarse-grain modeling of lipid membranes”. In: *PloS one* 6.12 (2011), e28637.
- [206] Mangesh Damre, Alessandro Marchetto, and Alejandro Giorgetti. “MERMAID: dedicated web server to prepare and run coarse-grained membrane protein dynamics”. In: *Nucleic Acids Research* 47.W1 (2019), W456–W461.
- [207] Brendan Eich and C Rand McKinney. “JavaScript Language Specification”. In: *Techn. Ber. Netscape Communications, Nov* (1996), pp. 96–002.
- [208] Luke Welling and Laura Thomson. *PHP and MySQL Web development*. Sams publishing, 2003.
- [209] Alexander S Rose and Peter W Hildebrand. “NGL Viewer: a web application for molecular visualization”. In: *Nucleic acids research* 43.W1 (2015), W576–W579.
- [210] Elizabeth Jurrus, Dave Engel, Keith Star, Kyle Monson, Juan Brandi, Lisa E Felberg, David H Brookes, Leighton Wilson, Jiahui Chen, Karina Liles, et al. “Improvements to the APBS biomolecular solvation software suite”. In: *Protein Science* 27.1 (2018), pp. 112–128.

- [211] Massimo Sandal, Tran Phuoc Duy, Matteo Cona, Hoang Zung, Paolo Carloni, Francesco Musiani, and Alejandro Giorgetti. “GOMoDo: a GPCRs online modeling and docking webserver”. In: *PloS one* 8.9 (2013), e74092.
- [212] Aaron A Thompson, Wei Liu, Eugene Chun, Vsevolod Katritch, Huixian Wu, Eyal Vardy, Xi-Ping Huang, Claudio Trapella, Remo Guerrini, Girolamo Calo, et al. “Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic”. In: *Nature* 485.7398 (2012), pp. 395–399.
- [213] Leandro Martinez, Roberto Andreani, and José Mario Martinez. “Convergent algorithms for protein structural alignment”. In: *BMC bioinformatics* 8 (2007), pp. 1–15.
- [214] Doralicia Casares, Pablo V Escribá, and Catalina Ana Rosselló. “Membrane lipid composition: effect on membrane and organelle structure, function and compartmentalization and therapeutic avenues”. In: *International journal of molecular sciences* 20.9 (2019), p. 2167.
- [215] Robert B Best, Xiao Zhu, Jihyun Shim, Pedro EM Lopes, Jeetain Mittal, Michael Feig, and Alexander D MacKerell Jr. “Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles”. In: *Journal of chemical theory and computation* 8.9 (2012), pp. 3257–3273.
- [216] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. “Comparison of multiple Amber force fields and development of improved protein backbone parameters”. In: *Proteins: Structure, Function, and Bioinformatics* 65.3 (2006), pp. 712–725.
- [217] Wilfred F van Gunsteren, SR Billeter, AA Eising, PH Hünenberger, PKHC Krüger, AE Mark, WRP Scott, and IG Tironi. “Biomolecular simulation: the GROMOS96 manual and user guide”. In: *Vdf Hochschulverlag AG an der ETH Zürich, Zürich* 86 (1996), pp. 1–1044.
- [218] PJ Artymiuk, CCF Blake, DW Rice, and KS Wilson. “The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme at 6 Å resolution”. In: *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* 38.3 (1982), pp. 778–783.
- [219] John R Horton, Xing Zhang, Robert Maunus, Zhe Yang, Geoffrey G Wilson, Richard J Roberts, and Xiaodong Cheng. “DNA nicking by HinP1I endonuclease: bending, base flipping and minor groove expansion”. In: *Nucleic acids research* 34.3 (2006), pp. 939–948.
- [220] Christos Lamprakis, Ioannis Andreadelis, John Manchester, Camilo Velez-Vega, José S Duca, and Zoe Cournia. “Evaluating the efficiency of the Martini force field to study protein dimerization in aqueous and membrane environments”. In: *Journal of Chemical Theory and Computation* 17.5 (2021), pp. 3088–3102.
- [221] Dirk Fitzner, Jakob M Bader, Horst Penkert, Caroline G Bergner, Minhui Su, Marie-Theres Weil, Michal A Surma, Matthias Mann, Christian Klose, and Mikael Simons. “Cell-type-and brain-region-resolved mouse brain lipidome”. In: *Cell reports* 32.11 (2020).
- [222] Daniel Kozma, Istvan Simon, and Gabor E Tusnady. “PDBTM: Protein Data Bank of transmembrane proteins after 8 years”. In: *Nucleic acids research* 41.D1 (2012), pp. D524–D529.
- [223] William Humphrey, Andrew Dalke, and Klaus Schulten. “VMD: visual molecular dynamics”. In: *Journal of molecular graphics* 14.1 (1996), pp. 33–38.
- [224] Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F Van Gunsteren, and Alan E Mark. “Peptide folding: when simulation meets experiment”. In: *Angewandte Chemie International Edition* 38.1-2 (1999), pp. 236–240.
- [225] Mercedes Alfonso-Prieto, Luciano Navarini, and Paolo Carloni. “Understanding ligand binding to G-protein coupled receptors using multiscale simulations”. In: *Frontiers in molecular biosciences* 6 (2019), p. 29.
- [226] Nandan Haloi, Emelia Karlsson, Rebecca J Howard, and Erik R Lindahl. “Discovering cryptic pocket opening and ligand binding in a vestibular site of the 5-HT3A receptor”. In: *Biophysical Journal* 123.3 (2024), 394a.
- [227] Flavio Ballante, Albert J Kooistra, Stefanie Kampen, Chris de Graaf, and Jens Carlsson. “Structure-based virtual screening for ligands of G protein-coupled receptors: what can molecular docking do for you?” In: *Pharmacological Reviews* 73.4 (2021), pp. 1698–1736.

- [228] Dasiel O Borroto-Escuela, David Rodriguez, Wilber Romero-Fernandez, Jon Kapla, Mariama Jaiteh, Anirudh Ranganathan, Tzvetana Lazarova, Kjell Fuxe, and Jens Carlsson. "Mapping the interface of a GPCR dimer: a structural model of the A2A adenosine and D2 dopamine receptor heteromer". In: *Frontiers in pharmacology* 9 (2018), p. 829.
- [229] Jennifer M Johnston, Hao Wang, Davide Provati, and Marta Filizola. "Assessing the relative stability of dimer interfaces in G protein-coupled receptors". In: (2012).
- [230] Raudah Lazim, Donghyuk Suh, Jai Woo Lee, Thi Ngoc Lan Vu, Sanghee Yoon, and Sun Choi. "Structural characterization of receptor-receptor interactions in the allosteric modulation of G protein-coupled receptor (Gpcr) dimers". In: *International Journal of Molecular Sciences* 22.6 (2021), p. 3241.
- [231] Diego Guidolin, Manuela Marcoli, Cinzia Tortorella, Guido Maura, and Luigi F Agnati. "Receptor-receptor interactions as a widespread phenomenon: novel targets for drug development?" In: *Frontiers in endocrinology* 10 (2019), p. 53.
- [232] Peter Schmidtke, Axel Bidon-Chanal, F Javier Luque, and Xavier Barril. "MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories". In: *Bioinformatics* 27.23 (2011), pp. 3276–3285.
- [233] Francesca Spyraakis, Serena Faggiano, Stefania Abbruzzetti, Paola Dominici, Elena Cacciatori, Alessandra Astegno, Enrica Droghetti, Alessandro Feis, Giulietta Smulevich, Stefano Bruno, et al. "Histidine E7 dynamics modulates ligand exchange between distal pocket and solvent in AHB1 from *Arabidopsis thaliana*". In: *The Journal of Physical Chemistry B* 115.14 (2011), pp. 4138–4146.
- [234] Witold K Subczynski, Marta Pasenkiewicz-Gierula, Justyna Widomska, Laxman Mainali, and Marija Raguz. "High cholesterol/low cholesterol: effects in biological membranes: a review". In: *Cell biochemistry and biophysics* 75.3 (2017), pp. 369–385.
- [235] Karel Paukner, Ivana Králová Lesná, and Rudolf Poledne. "Cholesterol in the cell membrane—an emerging player in atherogenesis". In: *International Journal of Molecular Sciences* 23.1 (2022), p. 533.
- [236] Sung-Tae Yang, Alex JB Kreutzberger, Jinwoo Lee, Volker Kiessling, and Lukas K Tamm. "The role of cholesterol in membrane fusion". In: *Chemistry and physics of lipids* 199 (2016), pp. 136–143.
- [237] Michael A Hanson, Vadim Cherezov, Mark T Griffith, Christopher B Roth, Veli-Pekka Jaakola, Ellen YT Chien, Jeffrey Velasquez, Peter Kuhn, and Raymond C Stevens. "A specific cholesterol binding site is established by the 2.8 Å structure of the human β 2-adrenergic receptor". In: *Structure* 16.6 (2008), pp. 897–905.
- [238] Parijat Sarkar and Amitabha Chattopadhyay. "Cholesterol in GPCR structures: Prevalence and relevance". In: *The Journal of Membrane Biology* 255.1 (2022), pp. 99–106.
- [239] Yoojoong Kim, Eunyoung Jeong, Ji-Hong Jeong, Youngjin Kim, and Yunje Cho. "Structural basis for activation of the heterodimeric GABAB receptor". In: *Journal of molecular biology* 432.22 (2020), pp. 5966–5984.
- [240] Naveen Michaud-Agrawal, Elizabeth J Denning, Thomas B Woolf, and Oliver Beckstein. "MDAnalysis: a toolkit for the analysis of molecular dynamics simulations". In: *Journal of computational chemistry* 32.10 (2011), pp. 2319–2327.
- [241] Małgorzata Jurak, Monika Golabek, Lucyna Holysz, and Emil Chibowski. "Properties of Langmuir and solid supported lipid films with sphingomyelin". In: *Advances in Colloid and Interface Science* 222 (2015), pp. 385–397.
- [242] Helen Watson. "Biological membranes". In: *Essays in biochemistry* 59 (2015), pp. 43–69.
- [243] Carine Sebaaly, Hélène Greige-Gerges, and Catherine Charcosset. "Lipid membrane models for biomembrane properties' investigation". In: *Current trends and future developments on (bio-) membranes*. Elsevier, 2019, pp. 311–340.
- [244] Helgi I Ingólfsson, Timothy S Carpenter, Harsh Bhatia, Peer-Timo Bremer, Siewert J Marrink, and Felice C Lightstone. "Computational lipidomics of the neuronal plasma membrane". In: *Biophysical journal* 113.10 (2017), pp. 2271–2280.

- [245] Elise Wouters, Adrián Ricarte Marín, James Andrew Rupert Dalton, Jesús Giraldo, and Christophe Stove. “Distinct dopamine D2 receptor antagonists differentially impact D2 receptor oligomerization”. In: *International journal of molecular sciences* 20.7 (2019), p. 1686.
- [246] Fabian Keller, Azadeh Alavizargar, Roland Wedlich-Söldner, and Andreas Heuer. “The impact of bilayer composition on the dimerization properties of the Slg1 stress sensor TMD from a multiscale analysis”. In: *Physical Chemistry Chemical Physics* 25.2 (2023), pp. 1299–1309.
- [247] Ole G Mouritsen and Martin J Zuckermann. “What’s so special about cholesterol?” In: *Lipids* 39.11 (2004), pp. 1101–1113.
- [248] Laura Anchisi, Sandra Dessì, Alessandra Pani, and Antonella Mandas. “Cholesterol homeostasis: a key to prevent or slow down neurodegeneration”. In: *Frontiers in physiology* 3 (2013), p. 486.
- [249] Satyan Sharma, Brian N Kim, Phillip J Stansfeld, Mark SP Sansom, and Manfred Lindau. “A coarse grained model for a lipid membrane with physiological composition and leaflet asymmetry”. In: *PLoS One* 10.12 (2015), e0144814.
- [250] Mauricio G Martín, Frank Pfrieger, and Carlos G Dotti. “Cholesterol in brain disease: sometimes determinant and frequently implicated”. In: *EMBO reports* 15.10 (2014), pp. 1036–1052.
- [251] Jan Jakubik and Esam E El-Fakahany. “Allosteric modulation of GPCRs of class A by cholesterol”. In: *International journal of molecular sciences* 22.4 (2021), p. 1953.
- [252] Sathish Dasari and Bhabani S Mallik. “Conformational free-energy landscapes of alanine dipeptide in hydrated ionic liquids from enhanced sampling methods”. In: *The Journal of Physical Chemistry B* 124.31 (2020), pp. 6728–6737.
- [253] Patrick Shaffer, Omar Valsson, and Michele Parrinello. “Enhanced, targeted sampling of high-dimensional free-energy landscapes using variationally enhanced sampling, with an application to chignolin”. In: *Proceedings of the National Academy of Sciences* 113.5 (2016), pp. 1150–1155.
- [254] Dung Thanh Dang. “Molecular approaches to protein dimerization: opportunities for supramolecular chemistry”. In: *Frontiers in Chemistry* 10 (2022), p. 829312.
- [255] David J Austin, Stuart L Schreiber, and Gerald R Crabtree. “Proximity versus allostery: the role of regulated protein dimerization in biology”. In: *Chemistry & Biology* 1.3 (1994), pp. 131–136.
- [256] Haiying Lu, Qiaodan Zhou, Jun He, Zhongliang Jiang, Cheng Peng, Rongsheng Tong, and Jianyou Shi. “Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials”. In: *Signal transduction and targeted therapy* 5.1 (2020), p. 213.
- [257] Timo M Schäfer and Giovanni Settanni. “Data reweighting in metadynamics simulations”. In: *Journal of chemical theory and computation* 16.4 (2020), pp. 2042–2052.
- [258] Samuel P Lee, Brian F O’Dowd, Ryan D Rajaram, Tuan Nguyen, and Susan R George. “D2 dopamine receptor homodimerization is mediated by multiple sites of interaction, including an intermolecular interaction involving transmembrane domain 4”. In: *Biochemistry* 42.37 (2003), pp. 11023–11031.
- [259] Beatriz Bueschbell, Pedro R Magalhães, Carlos AV Barreto, Rita Melo, Anke C Schiedel, Miguel Machuqueiro, and Irina S Moreira. “The World of GPCR dimers—Mapping dopamine receptor D2 homodimers in different activation states and configuration arrangements”. In: *Computational and Structural Biotechnology Journal* 21 (2023), pp. 4336–4353.
- [260] Wen Guo, Lei Shi, Marta Filizola, Harel Weinstein, and Jonathan A Javitch. “Crosstalk in G protein-coupled receptors: changes at the transmembrane homodimer interface determine activation”. In: *Proceedings of the National Academy of Sciences* 102.48 (2005), pp. 17495–17500.
- [261] Davide Sala, Peter W Hildebrand, and Jens Meiler. “Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties”. In: *Frontiers in Molecular Biosciences* 10 (2023), p. 1121962.
- [262] Yamuna Devi Paila, Mamata Kombrabail, G Krishnamoorthy, and Amitabha Chattopadhyay. “Oligomerization of the serotonin1A receptor in live cells: a time-resolved fluorescence anisotropy approach”. In: *The Journal of Physical Chemistry B* 115.39 (2011), pp. 11439–11447.
- [263] Arjen N Bader, Erik G Hofman, Jarno Voortman, Paul MP van Bergen en Henegouwen, and Hans C Gerritsen. “Homo-FRET imaging enables quantification of protein cluster sizes with subcellular resolution”. In: *Biophysical journal* 97.9 (2009), pp. 2613–2622.

- [264] Patrick E McKnight and Julius Najab. “Mann-Whitney U Test”. In: *The Corsini encyclopedia of psychology* (2010), pp. 1–1.
- [265] Roser Bono, Maria J Blanca, Jaume Arnau, and Juana Gómez-Benito. “Non-normal distributions commonly used in health, education, and social sciences: A systematic review”. In: *Frontiers in psychology* 8 (2017), p. 1602.
- [266] Jessica C Mar. “The rise of the distributions: why non-normality is important for understanding the transcriptome and beyond”. In: *Biophysical reviews* 11.1 (2019), pp. 89–94.
- [267] Yuliya V Karpievitch, Alan R Dabney, and Richard D Smith. “Normalization and missing value imputation for label-free LC-MS analysis”. In: *BMC bioinformatics* 13 (2012), pp. 1–9.
- [268] Patricia M Dijkman, Oliver K Castell, Alan D Goddard, Juan C Munoz-Garcia, Chris De Graaf, Mark I Wallace, and Anthony Watts. “Dynamic tuneable G protein-coupled receptor monomer-dimer populations”. In: *Nature communications* 9.1 (2018), p. 1710.
- [269] Geoffrey J Taghon, Jacob B Rowe, Nicholas J Kapolka, and Daniel G Isom. “Predictable cholesterol binding sites in GPCRs lack consensus motifs”. In: *Structure* 29.5 (2021), pp. 499–506.
- [270] Rinshi S Kasai and Akihiro Kusumi. “Single-molecule imaging revealed dynamic GPCR dimerization”. In: *Current opinion in cell biology* 27 (2014), pp. 78–86.
- [271] Kasper B Pedersen, Helgi I Ingólfsson, Daniel P Ramirez-Echemendia, Llus Borges-Araújo, Mikkel D Andreasen, Charly Empereur-Mot, Josef Melcr, Tugba N Ozturk, WF Drew Bennett, Lisbeth R Kjølbye, et al. “The Martini 3 lipidome: expanded and refined parameters improve lipid phase behavior”. In: *ACS central science* (2025).
- [272] Daniel M McKenzie, Daniel Wirth, Taras V Pogorelov, and Kalina Hristova. “Utility of FRET in studies of membrane protein oligomerization: The concept of the effective dissociation constant”. In: *Biophysical journal* 122.20 (2023), pp. 4113–4120.
- [273] Ana B Caniceiro, Beatriz Bueschbell, Anke C Schiedel, and Irina S Moreira. “Class A and C GPCR dimers in neurodegenerative diseases”. In: *Current Neuropharmacology* 20.11 (2022), pp. 2081–2141.
- [274] Sergi Ferré. “The GPCR heterotetramer: challenging classical pharmacology”. In: *Trends in pharmacological sciences* 36.3 (2015), pp. 145–152.
- [275] Gregory D Stewart, Laëtitia Comps-Agrar, Lenea Blanc Nørskov-Lauritsen, Jean-Philippe Pin, and Julie Kniazeff. “Allosteric interactions between GABAB1 subunits control orthosteric binding sites occupancy within GABAB oligomers”. In: *Neuropharmacology* 136 (2018), pp. 92–101.
- [276] Katie L Sharrocks, Francesca Fanelli, Yewei Liu, Annabelle J Milner, Wu Yining, Bernadette Byrne, and Aylin C Hanyaloglu. “Stabilized D2R G protein-coupled receptor oligomers identify multi-state β -arrestin complexes”. In: *Nature Communications* 16.1 (2025), p. 8768.
- [277] Hui Guo, Su An, Richard Ward, Yang Yang, Ying Liu, Xiao-Xi Guo, Qian Hao, and Tian-Rui Xu. “Methods used to study the oligomeric structure of G-protein-coupled receptors”. In: *Bioscience Reports* 37.2 (2017), BSR20160547.
- [278] Victor Fernández-Dueñas, Jordi Bonaventura, Ester Aso, Rafael Luján, Sergi Ferré, and Francisco Ciruela. “Overcoming the challenges of detecting GPCR oligomerization in the brain”. In: *Current neuropharmacology* 20.6 (2022), pp. 1035–1045.
- [279] Yulin Luo, Liping Sun, and Yao Peng. “The structural basis of the G protein-coupled receptor and ion channel axis”. In: *Current Research in Structural Biology* 9 (2025), p. 100165.
- [280] Mark Summerfield. *Programming in Python 3: a complete introduction to the Python language*. Addison-Wesley Professional, 2009.