Discovery of new selective antagonists of G-protein coupled receptors of therapeutic interest

Cecylia Severin Lupala

Discovery of new selective antagonists of G-Protein coupled receptors of therapeutic interest

Cecylia Severin Lupala

Universitat Politecnica de Catalaunya

Escola Superior d'Agricultura de Barcelona (ESAB)

Departament d'Enginyeria Agroalimentària i Biotecnologia

Discovery of new selective antagonists of G-Protein coupled receptors of therapeutic interest

Cecylia Severin Lupala

This thesis is submitted in partial fulfilment of the degree of Doctor of Philosophy in Agri-food Technology and Biotechnology

Doctorate supervisor

Prof. Juan Jesus Perez

September 2015

Departament d'Enginyeria Agroalimentària i Biotecnologia
Campus del Baix Llobregat,
Edifici D4. C. Esteve Terradas, 8
08860 Castelldefels
Barcelona

Cecylia Severin Lupala

BSc. Molecular biology and Biotechnology   (July 2006- Univerisity of Dar Es Salaam)
MSc. Biotechnological engineering   (July 2011-Universitat Politecnica de Catalunya)

The research was carried out at the Laboratory of molecular engineering, Department of chemical engineering (School of Industrial Engineering of Barcelona - ETSEIB)

Under the supervision of;

Prof. Juan Jesús Pérez González

Laboratori d'Enginyeria Molecular
Departament d'Enginyeria Quimica
Escola Técnica superior d'enginyeria industrial de
Barcelona
Universitat Politecnica de Catalunya
Avda. Diagonal 647 08028 – Barcelona

"Facts are not science…..As the dictionary is not literature".

Martin H. Fischer

To Leon and the next generation, Nathan and Nicole…

"Like an antique ballroom fan your eyelids flutter, Sunlight streams across your eyes through open shutter. Now I think you're ready for the journey. "

Tom Paxton – "I give you the morning"

Abstract

GPCR are integral membrane receptor proteins that are characterized by heptahelical transmembrane (7TM) domains connected by intracellular and extracellular loops, an extracellular N-terminus, and an intracellular C-terminus. GPCRs recognize neurotransmitters, sensory molecules and chemotactic agents and are involved in the control of many aspects of metabolism. Thus, GPCRs play important roles in diverse processes such as pain perception, growth and blood pressure regulation etc. Due to their physiological importance, GPCRs are considered a highly convenient class of proteins for drug discovery, with more than 50% of all drugs regulating GPCR function, and some 30% of these drugs directly target GPCRs. A significant percent of global pharmaceutical sales are realized from drugs targeted against only 40-50 well-characterized GPCRs out of > 1,000 genes encoded in the human genome. Thusly, it is very likely that many more GPCRs remain to be validated as drug targets.

A number of GPCR crystal structures determined in recent years have provided researchers with unprecedented opportunities in investigating how GPCRs function through the conformational changes regulated by their ligands. Unfortunately this number is a small fraction of total number of GPCRs known. Homology modeling, whereby an atomic structure of a target GPCR is constructed from the known crystallographic structure of another GPCR (the template) have been have been used instead. While the general 7TM domain is preserved, what often hampers homology modeling is the low sequence similarity between targets and templates.

To that end, the first study undertaken in this thesis was aimed at describing the effect of template selection. Using molecular dynamic simulations (MD), the primary was aim to develop and model increasingly accurate three-dimensional structures of a target GPCR. In order to test the methodology, a *post factum* comparison between the models with the available

crystallographic structure of the target GPCR. For this purpose several atomistic models of rat M3 muscarinic receptor were constructed by using MOE software from human M2 muscarinic receptor, human histamine 1 receptor and bovine rhodopsin receptor as templates. All of homology models resulted were refined with ligand (Tiotropium) docked inside using GLIDE software. In order to determine the effect of ligand in the simulation system, an extra model of M2 receptor was refined with NMS bound inside and an extra model refined without ligand. The five models system were embedded in a lipid bilayer prior to refinement process using GROMACS. The resulting models were compared with the crystallographic structure of rat M3 muscarinic receptor. Results show the sampling time 500ns is adequate simulation time and molecular dynamics simulation of the protein embedded in a lipid bilayer as a refinement process improves on the homology models. Specifically, the refinement process can correct the length of the TM segment of the target receptor, the accuracy of the model greatly depends on the proximity of the template and the target in the phylogenetic tree and finally, although the presence of a ligand produces a faster equilibration of the system.

With the methodology firmly established, the thesis studies involved an in-depth study of bradykinin receptors B1 and B2 (B1R and B2R respectively) was performed in order to get insight into the stereochemical requirements for binding of their antagonists. Specifically, the goal of the thesis chapters (chapter 4 for B1R and chapter 5 for B2R) was to carry out a structure-activity study of bradykinin antagonism, taking into account the stereochemical features of diverse non-peptide antagonists and the way these features translate into ligand anchoring points to complementary regions of the receptor, through the analysis of the respective ligand-receptor complex. Both receptors are drug targets for several pathophysiological conditions and their antagonists are thoughts to be analgesic, anti-inflammatory agents, etc. Since there are no crystallographic structures for B1R and B2R, atomistic models of each receptor was constructed

and refined using MD.

For B1R was constructed by homology modelling using the chemokine CXC4 and bovine rhodopsin receptors as template. The model was further refined using molecular dynamics with the protein embedded in a POPC bilayer. Simulations were run for 600 ns. From the simulations an average structure was obtained and energy minimized then later used for docking studies using Glide software. Antagonists selected for the docking studies include Compound 11, Compound 12, Chroman28, SSR240612, NPV-SAA164 and PS020990. The results of the docking study show the role of specific receptor residues in ligand binding. These results permitted the definition of a pharmacophore that describes the stereochemical requirements of antagonist binding, and were used in the development of new compounds with varied activity B1R inhibition profile depending on the number of pharmacophoric points fulfilled by the designed antagonistic molecules.

For B2R, a homology model was constructed using the chemokine CXC4 as template and subsequently refined embedded in a lipid bilayer by means of a 600 ns molecular dynamics trajectory. The average structure from the last hundred nanoseconds of the molecular dynamics trajectory was energy minimized and used as model of the receptor for docking studies. For this purpose, a set of compounds with antagonistic profile, covering maximal diversity were selected from the literature. Specifically, the set of compounds include Fasitibant, FR173657, Anatibant, WIN64338, Bradyzide, CHEMBL442294, and JSM10292. Molecules were docked into the B2R model and the corresponding complexes analyzed to understand ligand-receptor interactions. The outcome of this study is summarized in a 3D pharmacophore that explains the observed structure-activity results and provides insight into the design of novel molecules with antagonistic profile.

To prove the validity of the pharmacophoric hypothesis, a virtual screening process was also carried out. The proposed B1R and B2R pharmacophores were used as queries to identify new hits using diverse data bases of molecules. The results of these studies revealed a set of new hits

with structures not connected to the molecules used for pharmacophores development. A few of these structures were purchased and tested. The results of the binding studies show about a 33% success rate with a correlation between the number of pharmacophore points fulfilled and their antagonistic potency. Some of these structures are disclosed in this thesis. Moreover, the B1R and B2R pharmacophores developed were compared and the observed differences permitted to explain the stereochemical requirements for receptor-selective ligands.

The final study explores protein-protein interactions. Protein-protein interactions are important due to the fact that many proteins do not act in isolation, and often an examination of the interaction proteins can reveal a lot about their biological function together. Therefore, GPCR oligomers are also known to be targets of drug discovery. Two class A GPCRs, the serotonin receptor 5-Hydroxytryptamine 1A receptor (5-HT$_{1A}$) and Galanin receptor 1 (GALR1) have been described to heterodimerize triggering an atypical physiological state that would lead to depression. Furthermore, zinc supplementation has been widely reported to improve treatment against major depressive disorder. Hence the main objective of this thesis study was to establish a rational explanation for the role of zinc in the molecular processes associated with protein-protein interactions and its relationship with depression. By means of homology modeling, the atomic models of the receptors were constructed using as templates the crystallographic structures of 5-HT$_{1B}$ for 5-HT$_{1A}$ and κ–opioid receptor, and refined using molecular dynamics simulation. Using MOE GRID program the 5-HT$_{1A}$ was probed for Zinc and the zinc binding sites were annotated. Afterwards, to create the heterodimer of the two receptors, a protein-protein molecular docking was performed using GRAMM software.

The results showed three possible dimer interfaces: the first at TM4 and TM5, the second at TM6 and TM7, and the last at TM1 and TM2. Analysis of the 12 zinc binding sites and the heterodimer interfaces suggests that there is a coincidence between zinc binding sites and

heterodimerization interfaces. The results also establish a rational explanation for the role of zinc in the molecular processes associated with heterodimer interactions and its relationship with depression, in agreement with previously reported evidence for the positive effects of zinc in depression treatment, and the involvement of the dimer in the same disease.

Acknowledgements

One of the joys of completion is to look over the journey past and remember all the friends and family who have helped and supported me along this long but fulfilling road. I would like to express my profound and heartfelt gratitude to Professor Juan Jesús Pérez, my mentor and supervisor. I am grateful to you for many things, for letting me work on a subject matter you have pursued for almost 15 years, even when funding was not easy to come by. Thank you for helping to shape and guide the direction of the work with your careful and instructive comments. This work and all the academic goals I achieved throughout my time in Barcelona would not have been possible without you.

I would also like to thank those who have reviewed this work, who provided encouraging and constructive feedback. It is no easy task, reviewing a thesis, and I am grateful for their thoughtful and detailed comments.

I would like to acknowledge the support from my family throughout my graduate and doctorate studies. I would not have contemplated this road if not for my parents, especially mom who instilled within me a love of creative pursuits and science, all of which finds a place in this thesis. Secondly, to my siblings, my sisters Jane and Anna, and my brothers Joseph and Abel, who have also been the best of friends along this journey: Jane, who never understood why so far away from home but nevertheless encouraged me to go ahead and forge my dreams; Anna, who wisely believes any quest for knowledge must come with an adventure. And my brothers, Jay and Abel I hope this inspire you to go ahead and forge your academic dreams. To my family in the UK, whenever I needed to recharge my batteries, I knew I could count on you. My uncles, Joseph and Sigfrid, and my Aunt Anna, in a time I was beginning to think some dreams and plans were too big, you stepped up. Thank you for continuing to step up whenever I and my siblings need you. And to Dr. Cosam Joseph, for each inspirational, supportive, and patient word of encouragement you said to me and also for guiding me through my bachelor degree. I could not be more proud of my academic

xii

Discovery of new selective antagonists of G-Protein coupled receptors of therapeutic interest

<u>List of publications</u>

Some of the chapters of this thesis have been published in peer reviewed journals. Below are these articles with their publication states.

1. Lupala, C. S., Rasaeifar, B., Gomez-Gutierrez, P., & Perez, J. J. (2015). 193 Effect of template selection on the construction of atomistic models of GPCRs by homology modeling. *Journal of Biomolecular Structure and Dynamics*, 33(sup1), 127-128.(**Published**)

2. Tena-Campos, M., Ramon, E., Lupala, C. S., Pérez, J. J., Koch, K. W., & Garriga, P. (2015). Zinc Is Involved in Depression by Modulating G Protein-Coupled Receptor Heterodimerization. *Molecular neurobiology*, 1-13. (**Published)**

3. Lupala, C. S., Gomez-Gutierrez, P., & Perez, J. J. (2015). New insights into the stereo chemical requirements of the Bradykinin B2R antagonists binding, *Journal of Computer Aided Molecular Design.* (**Accepted with major revision)**

Table of Contents

List of abbreviation, acronyms and symbols

Ala	Alanine		2D	2 Dimensional
Arg	Arginine		3D	3 Dimensional
Asn	Asparagine		GABA	γ-aminobutyric acid
Asp	Aspartic acid		5-HT	5-hydroxytryptamine, serotonin
Cys	Cysteine		EDG1	Endothelial differentiation gene 1
Gln	Glutamine		EDG3	Endothelial differentiation gene 3
Glu	Glutamic acid		GPR26	G Protein-Coupled Receptor 26
Gly	Glycine		FRET	Förster resonance energy transfer
His	Histidine		SPR	Surface plasmon resonance spectroscopy
Ile	Isoleucine		FRAP	Fluorescence recovery after photobleaching
Leu	Leucine			
Lys	Lysine		α	Alpha
Met	Methionine		β	Beta
Phe	Phenylalanine		γ	Gamma
Pro	Proline		κ	Kappa
Ser	Serine		Å	Angstrom (1.0×10^{-10} meters)
Thr	Threonine			
Trp	Tryptophan		ns	Nanosecond (10^{-9} of a second)
Tyr	Tyrosine		fs	Femtosecond (10^{-15} of a 10–15 second)
Val	Valine			

List of figures

List of tables

Preface

I have always been interested in diseases; what causes them, how to characterize them and the therapeutic intervention in treating them. From early age my academic journey has been in service to these interest. With my bachelor studies I was able to gain knowledge on the etiology of many types of diseases. Later in my post-graduation employment I worked in genetic characterization. Learning the role of protein and protein mutations in the development of diseases. Finally in my master's education I was introduced to computer aided drug design, a rational methodology in designing better and selective (where appropriate) drugs. While the subject of my master's thesis involved a smaller class of protein with big therapeutic impact, it is only "rational" that my doctorate thesis subject is the largest class of protein and with even bigger therapeutic impact.

Indeed, the membrane GPCR family contains the largest number of drug targets in the human body, with more than a quarter of the clinically used drugs targeting them. They account for the majority of best-selling drugs and about 40% of all prescription pharmaceuticals on the market and that is from a mere 40-50 well characterized GPCRs which is a fraction of the over 1000 genes encoding for GPCRs. Their importance is also highlighted by the award of the 2012 Nobel Prize in chemistry to Robert Lefkowitz and Brian Kobilka for "characterizing these getaways to the cells". To date there have been at least seven other Nobel Prizes awarded for different aspects of GPCRs mediated signaling.

As it is with most GPCR, the common setback one experience in researching them is the lack of the crystallographic structure of a GPCR of interest. You see, there are only 25 GPCRs crystallographic structures available, with 21 of them belonging to the class A. To put into clear context, there are over 700 class A GPCR and only less than 3% of class A have been crystallized. Due to this, scientists have been using homology modeling to study GPCR. While the general structural characteristics such as the heptahelical transmembrane domain is preserved, low sequence similarity between the target GPCR and template GPCR often hinders the homology modeling

process. Thus scientists have put considerable efforts on ways to improve accuracy of GPCR homology models.

This thesis is aimed at contributing on those efforts. Furthermore, to use the methodology or procedure develop to construct atomistic models of GPCR that can be used in drug discovery process. Specifically for the non-opiate pain mediating receptors bradykinin B1 and B2. Moreover, since GPCRs also signals as oligomers, a procedure is described herein, to models oligomers and study their interfaces either to designs ligands or to get insight into their interfaces as it was used in this thesis. This thesis consists of six main chapters. Chapter 1 provides a general introduction to GPCR, their structures, signaling etc. Chapter 2 also provides general introduction, but here is to the theoretical methodologies used in drug discovery process and molecular dynamic simulation. Chapters 3-6 is deals with the objectives of the thesis.

Specifically, the objectives of this study are

1. To improve on the methodology for obtaining more accurate 3D homology models of the GPCR by molecular dynamic simulation (Chapter 3)

2. To construct an accurate homology model of bradykinin receptor B1 and use it to study the structure-activity-relationship (SAR) of the receptor and the available non-peptidic antagonists of the receptor for the purpose of identifying the stereochemical requirements of the antagonist binding; the pharmacophore (Chapter 4).

3. To construct an accurate homology model of bradykinin receptor B2 and use it to study the SAR of the receptor and the available non-peptidic antagonists of the receptor for the purpose of identifying the pharmacophore for antagonist binding (Chapter 5).

4. Using the pharmacophores developed to explain on the features required for the design of antagonists capable of discriminating between the B1 and B2 bradykinin receptors (Chapter 5)

5.   To construct the heterodimer of GALR1 and 5HT$_{1A}$ in order to study their dimeric nature and explain their proposed link to major depression disorder.

What I do hope for the results of this thesis is to add to the continuing advances in our knowledge of the characteristics of the GPCRs. Specifically, I do hope the procedures developed here can be applied to the study of other GPCRs and further development of new and selective receptor antagonists. GPCR research is not an easy task, and the impact of such research especially in the rational design of drugs effective in the therapeutic manipulation GPCR mediated-diseases will continue to grow. Hence, approaches and tools offered by computer aided drug design can be used to reduce the complexity of this task and, combined with computational modeling, can serve to yield testable predictions for the structural properties of GPCR-ligand complexes and GPCR oligomers.

CHAPTER 1

Introduction

1.1 Membrane proteins and the G protein-coupled receptors

One of the main constituents of cell membranes are bilayers made of lipids that surround the cells and the cell organelles. Lipid bilayers are not smooth because they contain a variety of proteins attached to the surface and embedded in the membrane. These membrane proteins are among the most important proteins from the biological point of view since they allow the cells to communicate with their environment, determine whether the immune system recognizes the cell as foreign or not, control cell adhesion to form tissues, dictate development of plants and animals, and they control important metabolic processes, including salt balance, energy production and transmission, and photosynthesis [1-3].

Membrane proteins are interesting because of their key roles in controlling the processes of life and can be used as targets of drugs. When a membrane protein binds to a chemical such as drug outside of the cell and the binding process causes a chemical response on the inside of a cell it is called a receptor. Most of these receptors exhibit a basal activity in the absence of stimuli, suggesting that part of the receptors on the cell surface are activated. Accordingly, ligands binding on these receptors can act as agonists, antagonists or inverse agonists. An analysis of the human genome indicates that there are about a thousand receptors in the human genome. Thus, these receptors are important and perform critical biological functions [1-2].

There are two types of membrane proteins namely, the peripheral proteins and integral proteins. Peripheral proteins have weaker and temporary connections to the membrane. Some just sit on the surface, anchored with a few ionic bonds while others might have small sections that dip into the hydrophobic section of the bilayer. The integral proteins are permanently connected to the cell membrane. They are hard workers and have large sections embedded in the hydrophobic middle layer of the membrane [1].

G-protein-coupled receptors are integral proteins that cross the membrane and can act as pathways for ions and molecule. GPCR constitute one of the largest families of integral proteins in eukaryotic organisms and are responsible for the primary mechanism for signal transduction from the

cell external environment to its cytoplasm. These receptors are activated by a plethora of different stimuli, including different kind of ligands, like protein hormones, lipids, peptides, biogenic amines, nucleotides or ions and other exogenous stimuli like light, odor or taste. The activation of the G-protein results in inhibition or stimulation of the production of second messengers that eventually elicit the linked cellular response. Abnormalities of signaling by GPCRs are at the root of disorders that affect most tissues and organs in our body, such as color blindness, pain, different types of cancers, obesity, schizophrenia, depression hyper functioning thyroid adenoma, precocious puberty and diabetes [4-7]. The G-protein coupled receptors are also responsible for the control of enzyme activity, ion channels and vesicle transport. Specific ligand binding triggers a conformational change in the receptor that leads to its active state, allowing binding and activation of heterotrimeric G-protein [8-9].

Targeting GPCRs for therapeutic intervention has been fruitful, with >50% of drugs on the market acting as either surrogate activators or inhibitors of the GPCRs that have defined native ligands. However, the majority of GPCRs identified (> 75%) are orphan receptors, which presents a challenge for identifying their native ligands and defining their function [10]. Approximately 9% of global pharmaceutical sales are realized from drugs targeted against only 40-50 well-characterized GPCRs [11-12]. Taking the fact GPCRs are encoded by > 1,000 genes in the human genome [13], it is very likely that there are many more GPCRs that remains to be validated as drug targets. Furthermore, endogenous ligands have been identified for only 200 GPCRs even though the human genome contains many more GPCR genes. Therefore, there are enormous opportunities for further drug discovery in the field of GPCRs [10].

1.2 GPCR topology

A generalized GPCR topology is characterized by seven membrane-spanning alpha helices (7-Transmembrane receptors, 7TM), the receptor amino-terminus (N-terminus) in the extracellular side and consequently with the receptor carboxyl-terminus (C-terminus) in the intracellular side of the cell (Figure 1.1). Typically the 7TM are connected by loops on the intracellular and extracellular side of the cell. The loops are termed according to their occurrence, extracellular loop (ECL1 - ECL3) and intracellular loop (ICL1 - ICL3) [14-15].
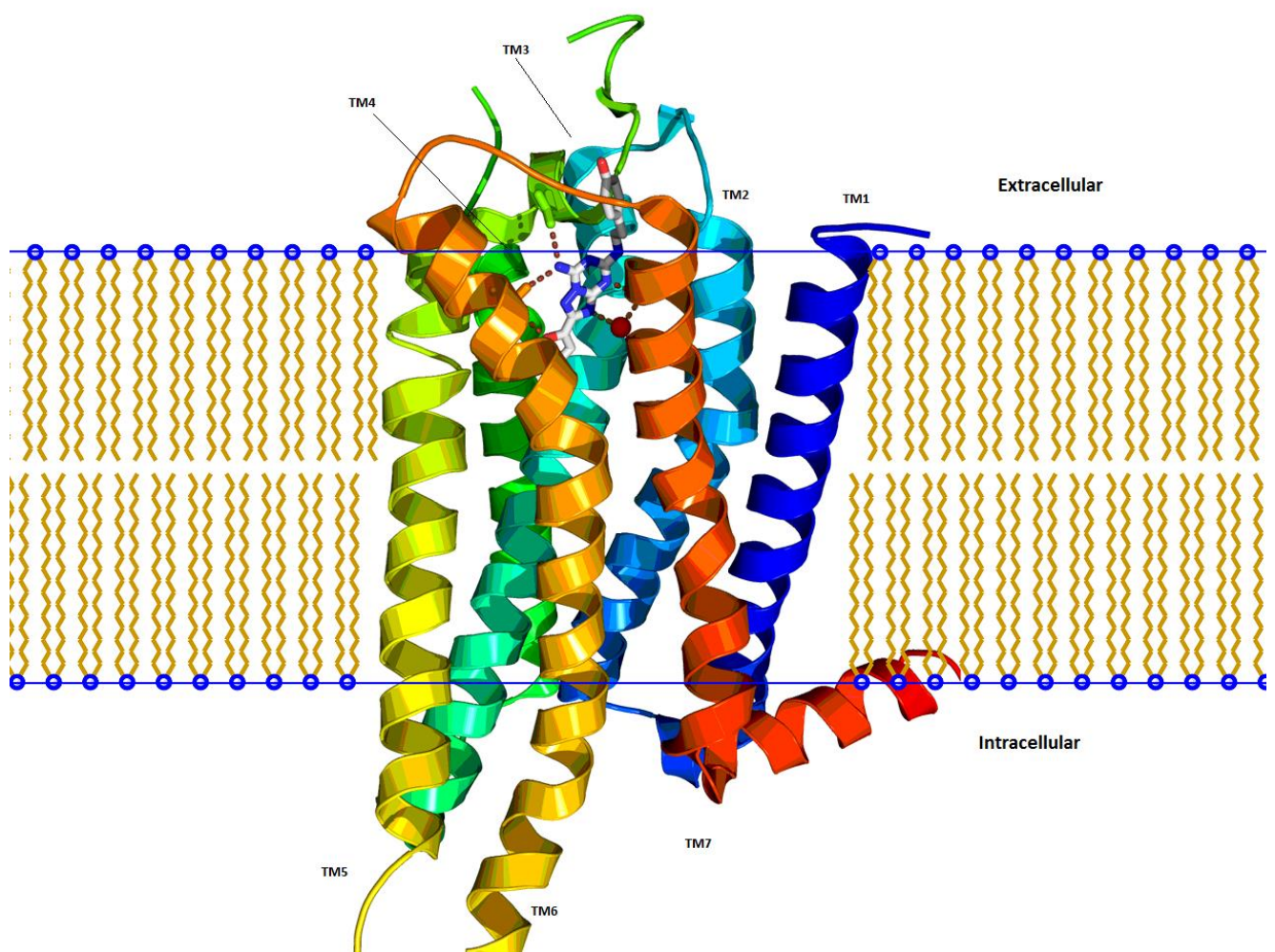
Figure 1.1: The GPCR nestled in a membrane bilayer, showing the transmembrane regions, intracellular side and the extracellular side. Modified from [16].

The 7TMs topology also includes additional similar conserved residues or motifs. Some of these conserved residues includes the conserved asparagine in TM1, aspartic, arginine and tyrosine (DRY motif) in TM3 which is shown in Figure 1.2 as ERY (as it is in rhodopsin receptor), conserved P in TM 4 and 5, CWXP in TM6 and the NPXXY motif in TM7 [14-15].
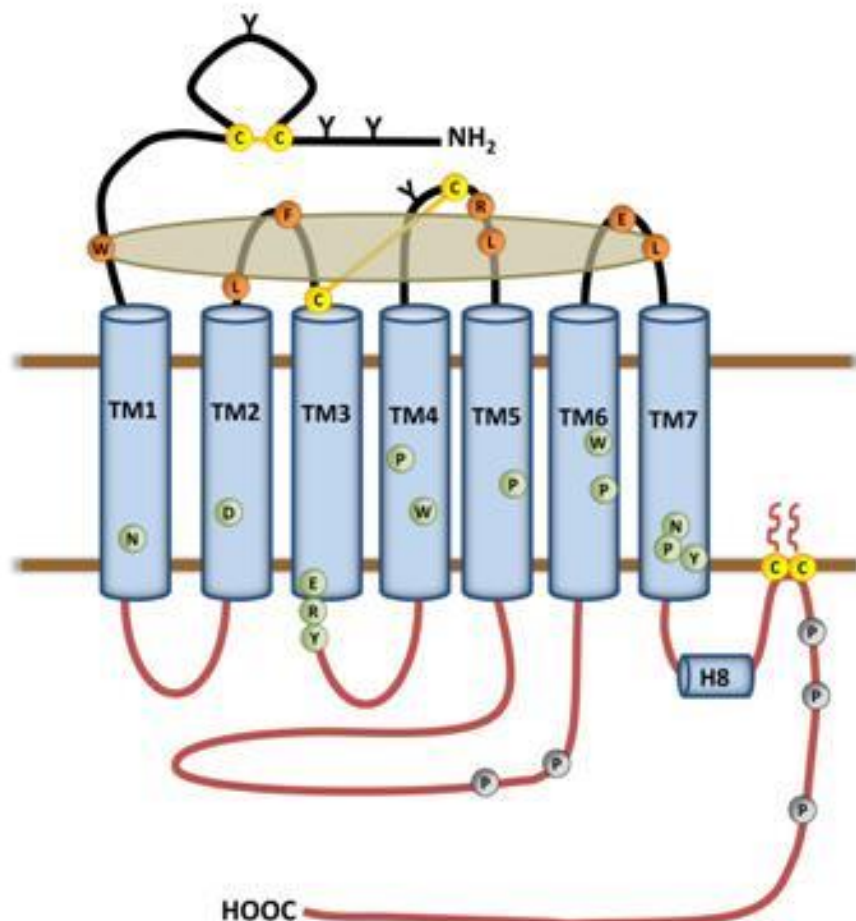
Figure 1.2. A generalized two-dimensional representation of the sequences of the GPCR. Showing Sites of glycosylation (Y) exist on ectodomains, while sites of phosphorylation are present intracellularly (grey 'P' circles). There are two disulfide bonds (connecting the cysteines) within the N-terminal tail and linking the top of TM3 with ECL2. In the C-terminal which undergoes reversible phosphorylation, there is also a site of palmitoylation within (two cysteines anchored) [17].

1.3 Classification of the GPCRs

Despite having a common 7TM motif and the few conserved residues, GPCRs have low overall sequence homology. In fact, GPCRs show a great conservation in their three-dimensional structure but exhibit high functional and sequence diversity. This makes it difficult to develop a comprehensive classification system based on sequence. Previous methods for GPCR classification have included motif-based prediction [18-19], support vector machines [20], as well as machine learning techniques such as Hidden Markov Models [21]. Currently, the most commonly used system for the functional classification of GPCRs (as used by the PSI GPCR network) is the data driven, top-down hierarchical classification GRAFS system which groups the GPCRs into five classes (Figure 1.3), **G**lutamate, **R**hodopsin, **A**dhesion, the **F**rizzled receptors and **S**ecretin class [22-24].
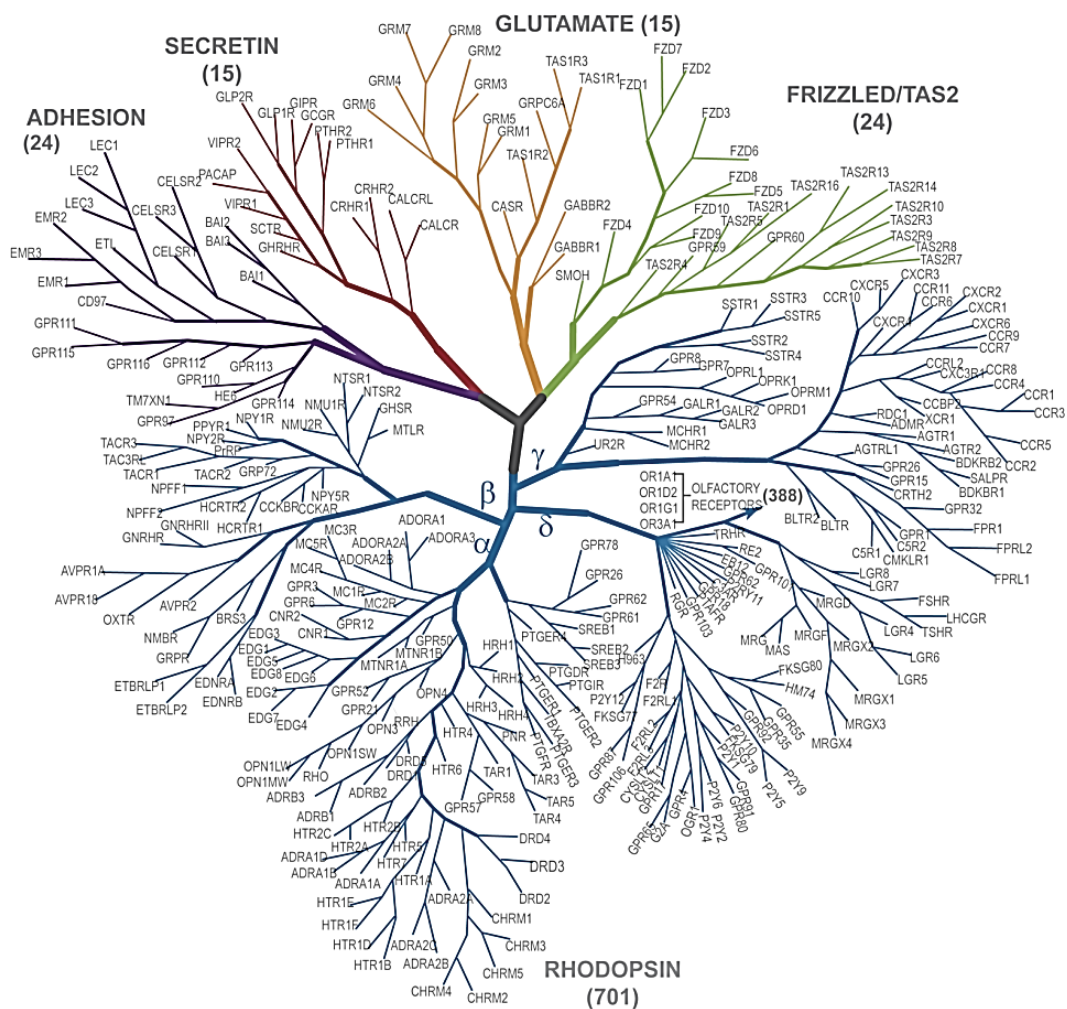
Figure 1.3: A Phylogenetic tree of the GPCRs showing their classification according to GRAFS system.

1.3.1 Rhodopsin-like

The group also known as class A or family I, comprises of about 84% of all GPCRs superfamily. It is the largest subgroup of GPCRs and some of the receptor class members includes receptors for odorants, neurotransmitters (dopamine, serotonin, endocannabinoids etc.) as well as neuropeptides, glycoprotein hormones and chemokines. Rhodopsin-like receptors are characterized by several highly conserved amino acids and a disulphide bridge that connects the first and second extracellular loops.

Most of these receptors also have a palmitoylation site made of cysteine in the C-terminal tail, which serves as an anchor to the membrane. Diversity among members of this group is not found in their N-terminals since most receptors have only a short stretch of amino acids, but within the TM regions. Most Rhodopsin-like receptors are primarily activated by interactions between the ligand and the TM regions and extracellular loops owing to their short N-terminal stretch of amino acids [22, 25-26].

1.3.2 Secretin-like

Also known as Class B or family II, many secretin receptors are regulated by peptide hormones from the glucagon hormone family and are implicated in cAMP-mediated signaling pathways. Members of this class are activated by ligands including secretin, parathyroid hormone, glucagon, calcitonin gene related peptide, adreno-medullin, calcitonin, etc. The binding profile of the Secretin-like receptors can be illustrated mainly by three binding domains consisting of the proximal region and the juxta-membrane region of the N terminus and the extracellular loops together with TM6. The ligand is thought to activate the receptor by bridging the characteristic-long extracellular N-terminal and the TM segments or extracellular loops thereby stabilizing the active conformation of the receptor [26]. While the secretin receptor family GPCRs exist in many animal species, it has not been found in plants, fungi or prokaryotes [21, 27].

1.3.3   Glutamate receptors

Also known as class C or family III, members of this GPCR group are structurally composed of the N-terminal signal sequence whereby the ligand binds, a large hydrophilic extracellular agonist-binding region containing several conserved cysteine residues which could be involved in disulphide bonds, a shorter region containing the 7TM, and a C-terminal cytoplasmic domain of variable length. Members includes metabotropic glutamate receptors, the extracellular calcium-sensing receptors, the gamma-amino-butyric acid (GABA) type B receptors, and the vomer nasal type-2 receptors [22, 28].

1.3.4   Adhesion

The adhesion GPCRs also known as Class D, family IV is comprised of 33 members in humans, including the brain angiogenesis inhibitor 3 (ADGRB3), calcium-independent alpha-latrotoxin receptor 1 (ADGRL1) etc. Ligand profiles and *in vitro* studies indicate that the primary role for adhesion GPCRs relates to the proper positioning of cells in a variety of organ systems (cell adhesion and migration). T he characteristic defining feature of adhesion GPCRs that distinguishes them from other GPCRs is their hybrid molecular structure. The extracellular region of adhesion GPCRs which is linked to the 7TM region by a stalk motif containing a GPCR proteolytic site, can be exceptionally long and contain a variety of structural and functional domains that are known for the ability to facilitate cell and matrix interactions. Most of these receptors have long and diverse N-terminal, which is thought to be highly glycosylated and form a rigid structure that protrudes from the cell surface. Unlike the secretin class, the adhesion class of GPCRs are also found in Fungi [22, 30].

1.3.5   Frizzled/Taste2 family

The frizzled family of receptors includes receptors critically involved in embryogenesis, and there is substantial evidence that members of this family also regulates tissue homeostasis in many different organs in the adult. Some of the members group includes the human smoothened receptor,

the bitter taste receptors etc. Frizzled GPCRS serves as receptors in the *wnt*/β-catenin signaling pathway whose mutations have been demonstrated to lead to a variety of diseases, including breast and prostate cancer, glioblastoma and type II diabetes among others.

Sequence comparisons between the frizzled receptor and secretin receptors revealed resemblance in the extracellular regions and the presence of the well-conserved cysteines in the first and second extracellular loops. The extracellular part of the frizzled GPCRs range from 200 to 320 amino acids in length in which the differences mostly lie in the linker region between the TM part and the extracellular ligand binding domain [21, 31].

1.4 GPCR signaling

It is well known that the signals in and out of a cell are mostly perceived at the level of the membrane and therefore the GPCR TMs are the likely routes for signal generation and transduction between cells [32-34]. First step in the GPCR mediated signaling cascade is binding of an agonist or ligand to GPCRs. Experimental results show that different ligands induce either G-protein dependent or G-protein independent signaling of GPCR via β- arrestins, which might result in functional selectivity [35]. Agonist binding to the GPCR promotes a conformational change in the receptor, specifically in an ionic interchange between TM3 and TM4. This induces coupling of the GPCR to the G-protein, initiating signaling to the cell interior. β-arrestins are well known negative regulators of GPCR signaling. Upon GPCR activation, β-arrestins translocate to the cell membrane and bind to the agonist-occupied receptors. This uncouples these receptors from G-proteins and promotes their internalization, thus causing desensitization. Furthermore, evidences have also indicated that β -arrestins also function as scaffold proteins that interact with several cytoplasmic proteins and link GPCRs to intracellular signaling pathways, such as mitogen activated protein kinase (MAPK) cascades [35-36].

GPCR signaling induces coupling of the ligand-bound receptor to a heteromeric G-protein. These composed of α-, β- and γ- subunits, are also a diverse group of proteins, comprising 17 Gα, 5 Gβ and

12 Gγ subunits at present [37]. When a ligand activates the GPCR, it induces a conformational change in the receptor that allows the receptor to act as a guanine nucleotide exchange factor for the alpha subunit of the G protein, whereby activated receptor promotes the exchange of bound guanine diphosphate (GDP) for guanine triphosphate (GTP) on the Gα subunit, which is the rate-limiting step in G protein activation. In the traditional view of heterotrimeric protein activation, this exchange triggers the dissociation of the Gα subunit, bound to GTP, from the Gβγ dimer and the receptor. The free α- or βγ-subunits 4 then interact with second messengers; the precise nature of which is dependent upon the GPCR type and the G-protein subunits mobilized [38-39].

G-proteins are classified into four major classes namely as the Gs, Gi/o, Gq/11, and G12/13 [40-42]. Stimulation of the Gs subfamily activates adenylyl cyclase, whereas stimulation of the Gi subfamily leads to its inhibition. Stimulation of the Gq subfamily activates phospholipase C, and the G12 family is implicated in the regulation of small GTP binding proteins. It has now become apparent that not only the α-subunits, but also the βγ -subunits can bind to a great variety of effector molecules and regulate their activity [43-45]. Gβγ-subunits mediate signal transduction by interacting with many proteins, including GPCRs, GTPases and various effector molecules. The effector molecules that have been reported to be regulated by Gβγ- subunits include adenylyl cyclase, PLC, inwardly rectifying G-protein-gated potassium channels, voltage-sensitive calcium channels, phosphoinositide 3-kinases (PI3 kinase) and molecules in the MAPK pathway. Recent developments indicate novel levels of complexity in GPCRs functioning [46].

The initial idea of linear signaling pathways, transferring information from the cell membrane to the nucleus, has recently evolved into a complicated network of signaling pathways. Firstly, cross-talk of the GPCRs on signaling pathways is increasingly more evident [37]. Secondly, some GPCRs may be constitutively active, i.e. active in the absence of its ligand. Particularly, the level of constitutive activity may vary in such a profound way between cells and tissues that this could offer new ways of achieving specificity of drug action [46-47]. Thirdly, increasing number of evidence showed that many GPCRs can form oligomeric ensembles [46, 48]. Therefore, regulation of GPCRs at multiple levels causes

emergence of specificity and complexity of GPCRs targeting.

However, this is all a classical paradigm of GPCR signaling, which is rather linear and sequential. Emerging evidence reveals that this is only a part of the complex signaling mediated by GPCR. In the classical model of GPCR signaling, stimulation of the 7TM leads to the activation of heterotrimeric G-proteins, which dissociate into α- and βγ-subunits. These subunits activate effector molecules, which include second messenger generating systems, giving rise to various kinds of cellular, physiological, and biological responses. In contrast to the large number of GPCRs, the number of identified effectors is considerably smaller.

Because many cells express multiple types of GPCRs that signal through limited types of effectors, it is not surprising that cross-regulation occurs in the signaling pathways of GPCRs, thereby leading to diverse physiological responses. Moreover, there is growing evidence that GPCR stimulation modulates upstream and downstream events of other receptor-mediated signaling pathways, which results in complicated and sometimes unpredictable outcomes [37].

1.4.1 GPCR Constitutive Activity

Evidences have demonstrated that GPCRs may exhibit constitutive activity in the absence of their agonists. A two-state, active and inactive receptor state model has been proposed to account for constitutive activity in which GPCRs exist in equilibrium between inactive and active states [49]. An agonist increases the activity of its GPCR above the basal level, presumably through shifting GPCRs into an active state capable of interacting with downstream-signaling G proteins. Thus, agonists stabilize the active state and display positive intrinsic activity by increasing the activity of a receptor above its basal level, resulting in an increase in receptor activity. In contrast, inverse agonists stabilize the inactive state and exhibit negative intrinsic activity by decreasing the receptor activity below the basal level. That is to say an inverse agonist decreases the GPCR activity below its basal level, by stabilizing GPCRs in an inactive state uncoupled from G proteins. Therefore, constitutive activity of

GPCRs can be selectively blocked by ligands that are inverse agonists [47].

A variety of human diseases are attributed to a constitutive activity of GPCRs that is caused by naturally occurring mutations [50]. Consequently, selective inverse agonists open up new therapeutic strategies for these types of human disorders. Since the level of constitutive activity is typically proportional to the number of active receptors, inverse agonism is usually most noticeable under conditions of high receptor expression, such as occurs in over expressed systems [47, 51].

1.4.2 GPCR Oligomerization

For long it was accepted that the mechanism of ligand binding and signal transduction by GPCRs were mediated by monomeric receptors. However, recent evidence reveal that GPCRs may exist as homodimers, or may associate with other GPCRs to form heterodimers [52-53]. This association may alter the function of both receptors, yielding in a distinct functional unit with novel properties [54-56]. Since tissue-selective expression of GPCR heterodimers and their differential activation offer exciting perspectives for the development of tissue- and receptor-subtype-selective drugs, these phenomena have promising potential in both basic and clinical research fields [48, 53].

Currently it is widely accepted that dimerization is fairly common in the GPCR superfamily members and that protein-protein interactions are essential to the organizational structure and function of cell signaling networks. Dimerization is thought to be important for various aspects of GPCR function such as formation of ligand-binding sites, signal transduction, receptor biogenesis, and down-regulation etc. An increasing number of experimental studies have been published showing that various classes of GPCRs form dimers and, or even higher order homo-oligomers [58-63].

The number of experimental reports showing that GPCR oligomerization has pharmacological and functional implications is also growing [56, 64-68]. For example, It was recently reported that inhibition of dimerization in β_2-adrenergic receptor by addition of a synthetic peptide corresponding to TM6 of the receptor limit agonist activation of adenylyl cyclase [69]. Studies on the δ-opioid receptor

dimerization provided evidence that the addition of certain agonists would inhibit or reverse receptor dimerization, meaning the δ-opioid receptor was likely internalized from the cell surface as a monomer in response to agonist challenge [59]. Thus, the study of dimerization has significance since abnormalities of dimer formation may alter the function of the receptors which results in serious disease although specifically for G protein activation and signaling, it is not resolved whether a single GPCR molecule is sufficient or a dimer is necessary [70-71].

Structure wise, various TM domains have been implicated to participate in receptor oligomerization in different receptor systems, suggesting that the structure of oligomers formed in different GPCRs may not be the same. In fact two modes of association of the TM helices of GPCR monomers into dimers have been proposed. Contact dimerization (figure 1.4a.) which shows the packing of two different TM bundles with separate binding sites through interactions at interfaces that otherwise would face the lipid environment [62], and domain-swapped dimerization (figure 1.4 b.),which involves interpenetration of transmembrane bundles whereby the interacting TMs from two different polypeptides appear as interlaced units [72-74].
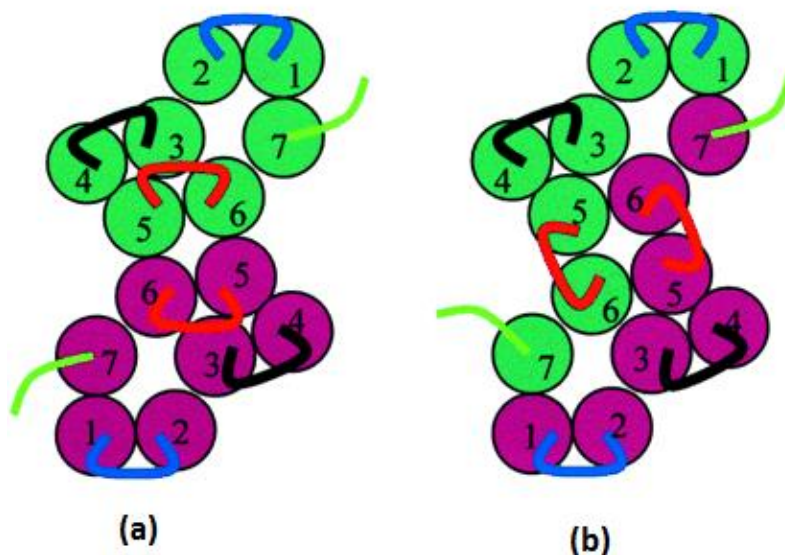
Figure 1.4: The proposed types of GPCR oligomerization, contact dimers **(a)** and domain-swapped dimers **(b).**

The studies on class C receptors provide valuable information that the N-terminal region may also participate into oligomerization as well as transmembrane domains. The N-terminal Venus flytrap (VFT) module is one of main characteristic for class C receptors, which is structurally and functionally homologous to bacterial periplasmic proteins that bind amino acids, sugars and ions [75]. It was reported that VFTs are important for dimerization of this receptor class. Additionally, it is known that cysteine-rich domains (CRDs), localized between VFT and transmembrane domains potentially has a role in dimerization [76-78].

1.5 GPCRs as drug targets

As stated above GPCR represent a significant number of the human genome proteins. Due to their extensive involvement in human pathophysiological processes, development of therapeutically viable agents that counteract the effects these receptors is, therefore, potentially very attractive [79]. Much of the conceptual framework regarding how to study receptor function evolved from pharmacological investigation of drug action. The equations and models that describe the interaction of chemicals with receptors comprise what is generally known as pharmacodynamics. Although largely adapted from enzyme kinetics, the principles necessarily differ from those of enzyme-substrate interactions in that they deal with chemicals that are not changed by the receptor, but rather interact in a reversible manner to produce a change in the state of the receptor that is then transmitted to the cellular host [79-80].

The identification of constitutive, or intrinsic, activity of GPCRs has had major impact on receptor theory. Constitutive activity may be considered to be an equilibrium state of a GPCR between its ground state of no activity and its fully activated state. The identification of agents that inhibit this ligand-independent receptor activity has led, in turn, to the concept of inverse agonism and neutral antagonism. Inverse agonists are able to bind to constitutively active receptor and produce opposite response to that of agonist. A neutral antagonist itself has no effect on the activity of the receptor,

but it can prevent the interaction of agonists or inverse agonists with the receptor, although it does not affect the equilibria of different receptor conformations. Traditionally, antagonists are defined as agents that block receptor-mediated effects elicited by agonists by competing for receptor occupancy. Other antagonists of function mediate their effects by interacting with another allosteric site rather than in the binding pocket of the native agonist, the orthosteric site [80]. Figure 1.5 shows the position of orthosteric and allosteric site in a prototypical rhodopsin-like GPCR.
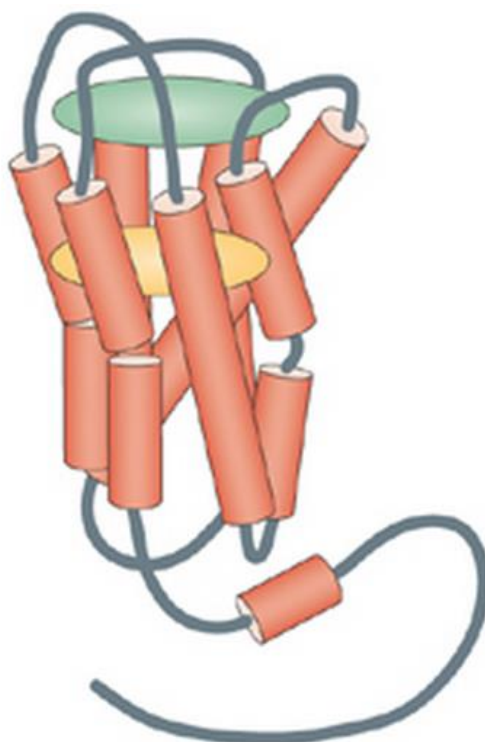
Figure 1.5: The position of orthosteric (in orange) and allosteric (in green) binding site in typical rhodopsin-like GPCR.

In many cases, inverse agonists were originally characterized as antagonists on non-constitutively active GPCRs and their inverse agonism is primarily seen when they interact with constitutively active GPCRs. Thusly, approximately around 85%, of all known GCPR antagonists are in fact inverse agonists [81]. The existence of constitutive activity has required a reconsideration of how we think about GPCR antagonists. Recently, antagonist agents have been observed to have negative intrinsic activity, or behave as inverse agonists, and decrease the basal agonist-independent or

constitutive activity of the receptor. As a result a clear distinction on the concept is employed and neutral antagonists are defined as agents that block both agonist and inverse agonist binding but produce no response on the receptor. Neutral antagonists considered to be relatively rare in pharmacology [80, 82-84].

In the drug discovery research, more success have been found in the identification of non-peptide antagonists for peptide-activated GPCRs [85]. Peptide-based ligand design encounters several major considerations that limit clinical applications such as rapid degradation by many specific or non-specific peptidases under physiological conditions poor absorption and transportation because of their high molecular mass or the lack of specific delivery systems, especially for some peptides which require the passage through the blood-brain-barrier to act in the central nervous system and conformational flexibility which allows a peptide to bind to more than one receptor or receptor subtype leading to undesirable side effects. Thus, to counteract these problems, non-peptide and peptide-mimetic ligands have emerged as an important option and strategy in drug design [85-87].

1.6 Crystallographic structures of the GPCRs

Structures of GPCRs at high resolution are essential for a better understanding of the structure and function of these receptors: from receptor conformational changes upon receptor activation to computational structure-activity studies leading to drug discovery. Indeed, the availability of a high-resolution crystal structure of the drug target receptor is of critical importance in the drug discovery process. Crystallographic structures of GPCRs at high resolution were not available until 2000, when a breakthrough in determining the crystal structure of bovine rhodopsin, a prototype class A GPCR was achieved at the resolution of 2.8 Å [88]. However, due to the difficulties associated in crystallizing the GPCRs such as the intrinsically unstable properties of the GPCRS, the technique used to crystallize rhodopsin was not successful in crystallizing the other GPCRs [89-90].

To improve the stability of membrane proteins, several innovative techniques have been developed and applied to the crystallization of GPCRs. One of the most important innovations is to substitute the flexible region of the protein that hampers the crystallization with a structurally rigid domain to increase the stability of the GPCR [91]. Other innovations include increasing the thermodynamic stability of a GPCR by selected mutations and the use of the lipid cubic phase and new detergents to facilitate GPCR crystallization [92-93].

Through application of the innovative techniques, the structure of the β2 adrenergic receptor was crystallized [94], and similar technique used to crystallize the β2 adrenergic receptor have also been successfully applied to the other GPCRs and have led to a flurry of GPCR structures. To date more than 100 structures of about 25 different GPCRs belonging to four classes have been determined (none of the adhesion class GPCR has yet been crystallized) [95].

All these structures have greatly advanced our knowledge of GPCRs structure and provided insightful information about their function. However, there are still many structures to be solved. In fact, there are more than 1000 GPCR encoding genes and about equally number of GPCRs. Moreover, most of these have been identified as drug targets. To cope with this shortage of structures computational methodologies, such homology modeling can help to provide structures at the atomic resolution at various levels of accuracy after a refinement process. In fact the available crystallographic GPCRs structures have increased the repertoire of templates used for modeling purposes. Moreover, molecular modeling can also help to understand the dynamical behavior of these receptors. Indeed, a crystal structure is a static picture corresponding to the conformation in an energetically favorable state, whereas a GPCR is a dynamic entity that can exist in various states. As a result, little is known about how the conformations changes of GPCRs can be associated with the signaling of the GPCRs or as it is critical in drug design process, how the conformation change is regulated by different ligands. Thus, and molecular dynamics simulation have been employed to understand the dynamical behavior of GPCRs [96-98]. These methodologies and theories governing them are discussed in chapter 2.

The goal of this thesis is to construct accurate models for GPCRs. This has been applied specifically to the structures of the B1 and B2 bradykinin receptors and to understand the heterodimerization of the serotonin-galanin heterodimer. Models of the receptors were constructed by homology modeling. For this purpose we developed and analyzed a refinement technique of the initial models constructed by homology modeling based on molecular dynamics.

1.7 References to chapter 1

1. http://www.britannica.com/EBchecked/topic/101396/cell/37365/The-cell-membrane?Anchor=ref313686

2. Boyd, D., Schierle, C., & Beckwith, J. (1998). How many membrane proteins are there?. *Protein Science*, *7*(1), 201-205.

3. Klein, P., Kanehisa, M., & DeLisi, C. (1985). The detection and classification of membrane-spanning proteins. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *815*(3), 468-476.

4. Drews, J. (2000). Drug discovery: a historical perspective. *Science*, *287*(5460), 1960-1964.

5. Karnik, S. S., Gogonea, C., Patil, S., Saad, Y., & Takezako, T. (2003). Activation of G-protein-coupled receptors: a common molecular mechanism. *Trends in Endocrinology & Metabolism*, *14*(9), 431-437.

6. Insel, P. A., Tang, C. M., Hahntow, I., & Michel, M. C. (2007). Impact of GPCRs in clinical medicine: monogenic diseases, genetic variants and drug targets. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1768*(4), 994-1005.

7. Spiegel, A. M., & Weinstein, L. S. (2004). Inherited Diseases Involving G Proteins and G Protein-Coupled Receptors*. *Annu. Rev. Med.*, *55*, 27-39.

8. Weiss, F. U., Daub, H., & Ullrich, A. (1997). Novel mechanisms of RTK signal generation. *Current opinion in genetics & development*, *7*(1), 80-86.

9. Wess, J. (1998). Molecular basis of receptor/G-protein-coupling selectivity. *Pharmacology & therapeutics*, *80*(3), 231-264.

10. Jacoby, E., Bouhelal, R., Gerspacher, M., & Seuwen, K. (2006). The 7 TM G-protein-coupled receptor target family. *ChemMedChem*, *1*(8), 760-782.

11. Eglen, R. M. (2004, December). Emerging concepts in GPCR function--the influence of cell phenotype on GPCR pharmacology. In *Proceedings of the Western Pharmacology Society* (Vol. 48, pp. 31-34).

12. Dastmalchi, S., Church, W. B., & Morris, M. B. (2008). Modelling the structures of G protein-coupled receptors aided by three-dimensional validation. *BMC bioinformatics*, *9*(Suppl 1), S14.

13. Howard, A. D., McAllister, G., Feighner, S. D., Liu, Q., Nargund, R. P., Van der Ploeg, L. H., & Patchett, A. A. (2001). Orphan G-protein-coupled receptors and natural ligand discovery. *Trends in pharmacological sciences*, *22*(3), 132-140.

14. Mirzadegan, T., Benkö, G., Filipek, S., & Palczewski, K. (2003). Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin.*Biochemistry*, *42*(10), 2759-2767.

15. Ballesteros, J., and H. Weinstein. (1995). Integrated methods for the construction of three-dimensional models of structure-function relations in G protein-coupled receptors. *Methods Neurosci*. 25:366-428.

16. Https://upload.wikimedia.org/wikipedia/commons/1/16/A2A_receptor_bilayer.png

17. Desai, A. J., & Miller, L. J. (2013). Cholecystokinin Type 1 Receptor. *Pancreapedia: The Exocrine Pancreas Knowledge Base*.

18. Flower, D.R. and Attwood, T.K. (2004) 'Integrative bioinformatics for functional genome annotation: trawling for G-Protein-Coupled Receptors', Seminars in Cell and Developmental Biology, Vol. 15, pp.693–701.

19. Holden, N. and Freitas, A.A. (2006) 'Hierarchical classification of G-Protein Coupled Receptors With a PSO/ACO Algorithm', Proc. IEEE Swarm Intelligence Symposium (SIS-06), IEEE Press, Piscataway, NJ, pp.77–84.

20. Karchin, R., Karplus, K., & Haussler, D. (2002). Classifying G-protein coupled receptors with support vector machines. *Bioinformatics*, *18*(1), 147-159.

21. Möller, S., Vilo, J., & Croning, M. D. (2001). Prediction of the coupling specificity of G protein coupled receptors to their G proteins. Bioinformatics,17(suppl 1), S174-S181.

22. Alexander, S. P., Benson, H. E., Faccenda, E., Pawson, A. J., Sharman, J. L., Spedding, M., ... & Harmar, A. J. (2013). The Concise Guide to PHARMACOLOGY 2013/14: G Protein-Coupled Receptors. British journal of pharmacology, 170(8), 1459-1581.

23. Bjarnadóttir, T. K., Gloriam, D. E., Hellstrand, S. H., Kristiansson, H., Fredriksson, R., & Schiöth, H. B. (2006). Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse.*Genomics*, *88*(3), 263-273.

24. Secker, A., Davies, M. N., Freitas, A. A., Clark, E. B., Timmis, J., & Flower, D. R. (2010). Hierarchical classification of G-Protein-Coupled Receptors with data-driven selection of attributes and classifiers. *International journal of data mining and bioinformatics*, *4*(2), 191-210.

25. Attwood, T. K., & Findlay, J. B. C. (1994). Fingerprinting G-protein-coupled receptors. *Protein engineering*, *7*(2), 195-203.

26. Joost, P., & Methner, A. (2002). Phylogenetic analysis of 277 human G-protein-coupled receptors as a tool for the prediction of orphan receptor ligands.*Genome Biol*, *3*(11), 1-16.

27. Hollenstein, K., de Graaf, C., Bortolato, A., Wang, M. W., Marshall, F. H., & Stevens, R. C. (2014). Insights into the structure of class B GPCRs. *Trends in pharmacological sciences*, *35*(1), 12-22.

28. Harmar, A. J. (2001). Family-B G-protein-coupled receptors. *Genome Biol*,*2*(12), 3013-1.

29. Brauner-Osborne, H., Wellendorph, P., & Jensen, A. A. (2007). Structure, pharmacology and therapeutic prospects of family C G-protein coupled receptors. *Current drug targets*, *8*(1), 169-184.

30. Hamann, J., Aust, G., Araç, D., Engel, F. B., Formstone, C., Fredriksson, R., ... & Schiöth, H. B. (2015). International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G Protein–Coupled Receptors.*Pharmacological reviews*, *67*(2), 338-367.

31. Schulte, G., & Bryja, V. (2007). The Frizzled family of unconventional G-protein-coupled receptors. *Trends in pharmacological sciences*, *28*(10), 518-525.

32. Luttrell, L. M. (2006). Transmembrane signaling by G protein-coupled receptors. In *Transmembrane Signaling Protocols* (pp. 3-49). Humana Press.

33. Marinissen, M. J., & Gutkind, J. S. (2001). G-protein-coupled receptors and signaling networks:

emerging paradigms. *Trends in pharmacological sciences*,*22*(7), 368-376.

34. Pierce, K. L., Premont, R. T., & Lefkowitz, R. J. (2002). Seven-transmembrane receptors. *Nature Reviews Molecular Cell Biology*, *3*(9), 639-650.

35. Violin, J. D., & Lefkowitz, R. J. (2007). β-Arrestin-biased ligands at seven-transmembrane receptors. *Trends in pharmacological sciences*, *28*(8), 416-422.

36. Ma, L., & Pei, G. (2007). β-arrestin signaling and regulation of transcription.*Journal of cell science*, *120*(2), 213-218.

37. Hur, E. M., & Kim, K. T. (2002). G protein-coupled receptor signalling and cross-talk: achieving rapidity and specificity. *Cellular signalling*, *14*(5), 397-405.

38. Pitcher, J. A., Freedman, N. J., & Lefkowitz, R. J. (1998). G protein-coupled receptor kinases. *Annual review of biochemistry*, *67*(1), 653-692.

39. Lefkowitz, R. J. (1993). G protein—coupled receptor kinases. *Cell*, *74*(3), 409-412.

40. Neer, E. J. (1995). Heterotrimeric C proteins: Organizers of transmembrane signals. *Cell*, *80*(2), 249-257.

41. Rens-Domiano, S. T. E. P. H. A. N. I. E., & Hamm, H. E. (1995). Structural and functional relationships of heterotrimeric G-proteins. *The FASEB Journal*,*9*(11), 1059-1066.

42. Conklin, B. R., & Bourne, H. R. (1993). Structural elements of Gα subunits that interact with Gβγ, receptors, and effectors. *Cell*, *73*(4), 631-641.

43. Clapham, D. E., & Neer, E. J. (1997). G protein βγ subunits. *Annual review of pharmacology and toxicology*, *37*(1), 167-203.

44. Schwindinger, W. F., & Robishaw, J. D. (2001). Heterotrimeric G-protein bg-dimers in growth and differentiation. *Oncogene*, *20*(13), 1653-1660.

45. Morris, A. J., & Malbon, C. C. (1999). Physiological regulation of G protein-linked signaling. *Physiological Reviews*, *79*(4), 1373-1430.

46. Fredholm, B. B., Hökfelt, T., & Milligan, G. (2007). G-protein-coupled receptors: an update. *Acta physiologica*, *190*(1), 3-7.

47. Milligan, G. (2003). Constitutive activity and inverse agonists of G protein-coupled receptors: a current perspective. *Molecular pharmacology*, *64*(6), 1271-1276.

48. Rozenfeld, R., Décaillot, F. M., IJzerman, A. P., & Devi, L. A. (2007). Heterodimers of G protein-coupled receptors as novel and distinct drug targets.*Drug Discovery Today: Therapeutic Strategies*, *3*(4), 437-443.

49. Costa, T. O. M. M. A. S. O., Ogino, Y. O. S. H. I. O., Munson, P. J., Onaran, H. O., & Rodbard, D. A. V. I. D. (1992). Drug efficacy at guanine nucleotide-binding regulatory protein-linked receptors: thermodynamic interpretation of negative antagonism and of receptor activity in the absence of ligand. *Molecular Pharmacology*, *41*(3), 549-560.

50. Spiegel, A. (2006). Mutations in G proteins and G protein-coupled receptors in human endocrine diseases. In *Insights into Receptor Function and New Drug Development Targets* (pp. 139-150). Springer Berlin Heidelberg.

51. De Ligt, R. A., Kourounakis, A. P., & IJzerman, A. P. (2000). Inverse agonism at G protein-coupled receptors:(patho) physiological relevance and implications for drug discovery. *British journal of pharmacology*, *130*(1), 1-12.

52. Ferré, S., Ciruela, F., Woods, A. S., Lluis, C., & Franco, R. (2007). Functional relevance of neurotransmitter receptor heteromers in the central nervous system. *Trends in neurosciences*, *30*(9), 440-446.

53. Franco, R., Casadó, V., Cortés, A., Ferrada, C., Mallol, J., Woods, A., ... & Ferré, S. (2007). Basic concepts in G-protein-coupled receptor homo-and heterodimerization. *The Scientific World Journal*, *7*, 48-57.

54. Hébert, T. E., & Bouvier, M. (1998). Structural and functional aspects of G protein-coupled receptor oligomerization. *Biochemistry and cell biology*, *76*(1), 1-11.

55. Gomes, I., Jordan, B. A., Gupta, A., Rios, C., Trapaidze, N., & Devi, L. A. (2001). G protein coupled receptor dimerization: implications in modulating receptor function. *Journal of molecular medicine*, *79*(5-6), 226-242.

56. Milligan, G. (2006). G-protein-coupled receptor heterodimers: pharmacology, function and relevance to drug discovery. *Drug discovery today*, *11*(11), 541-549.

57. Franco, R., Casadó, V., Cortés, A., Mallol, J., Ciruela, F., Ferré, S., ... & Canela, E. I. (2008). G-protein-coupled receptor heteromers: function and ligand pharmacology. *British journal of pharmacology*, *153*(S1), S90-S98.

58. Tena-Campos, M., Ramon, E., Lupala, C. S., Pérez, J. J., Koch, K. W., & Garriga, P. (2015). Zinc Is Involved in Depression by Modulating G Protein-Coupled Receptor Heterodimerization. *Molecular neurobiology*, 1-13.

59. Cvejic, S., & Devi, L. A. (1997). Dimerization of the δ opioid receptor: implication for a role in receptor internalization. *Journal of Biological Chemistry*, *272*(43), 26959-26964.

60. Bulenger, S., Marullo, S., & Bouvier, M. (2005). Emerging role of homo-and heterodimerization in G-protein-coupled receptor biosynthesis and maturation. *Trends in pharmacological sciences*, *26*(3), 131-137.

61. Frey, A. J. (2013). Characterization and targeting of thromboxane receptor dimerization: A gateway to novel therapeutic developments.

62. Schulz, A., Grosse, R., Schultz, G., Gudermann, T., & Schöneberg, T. (2000). Structural implication for receptor oligomerization from functional reconstitution studies of mutant V2 vasopressin receptors. *Journal of Biological Chemistry*, *275*(4), 2381-2389.

63. Han, Y., Moreira, I. S., Urizar, E., Weinstein, H., & Javitch, J. A. (2009). Allosteric communication between protomers of dopamine class A GPCR dimers modulates activation. *Nature chemical biology*, *5*(9), 688-695.

64. Breitwieser, G. E. (2004). G Protein–Coupled Receptor Oligomerization Implications for G Protein Activation and Cell Signaling. *Circulation research*, *94*(1), 17-27.

65. Bai, M. (2004). Dimerization of G-protein-coupled receptors: roles in signal transduction. *Cellular signalling*, *16*(2), 175-186.

66. George, S. R., O'Dowd, B. F., & Lee, S. P. (2002). G-protein-coupled receptor oligomerization

and its potential for drug discovery. *Nature Reviews Drug Discovery*, *1*(10), 808-820.

67. Lagerström, M. C., & Schiöth, H. B. (2008). Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nature reviews Drug discovery*, *7*(4), 339-357.

68. Devi, L. A. (2001). Heterodimerization of G-protein-coupled receptors: pharmacology, signaling and trafficking. *Trends in Pharmacological Sciences*,*22*(10), 532-537.

69. Hebert, T. E., Moffett, S., Morello, J. P., Loisel, T. P., Bichet, D. G., Barret, C., & Bouvier, M. (1996). A peptide derived from a β2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation.*Journal of Biological Chemistry*, *271*(27), 16384-16392.

70. Chabre, M., & le Maire, M. (2005). Monomeric G-protein-coupled receptor as a functional unit. *Biochemistry*, *44*(27), 9395-9403.

71. Park, P. S. H., Filipek, S., Wells, J. W., & Palczewski, K. (2004). Oligomerization of G protein-coupled receptors: past, present, and future.*Biochemistry*, *43*(50), 15643-15656.

72. Gouldson, P. R., & Reynolds, C. A. (1997). Simulations on dimeric peptides: evidence for domain swapping in G-protein-coupled receptors?. *Biochemical Society Transactions*, *25*(3), 1066-1071.

73. Gouldson, P. R., Snell, C. R., Bywater, R. P., Higgs, C., & Reynolds, C. A. (1998). Domain swapping in G-protein coupled receptor dimers. *Protein engineering*, *11*(12), 1181-1193.

74. Gouldson, P. R., Higgs, C., Smith, R. E., Dean, M. K., Gkoutos, G. V., & Reynolds, C. A. (2000). Dimerization and domain swapping in G-protein-coupled receptors: a computational study. *Neuropsychopharmacology*, *23*, S60-S77.

75. Kaupmann, K., Malitschek, B., Schuler, V., Heid, J., Froestl, W., Beck, P., ... & Bettler, B. (1998). GABAB-receptor subtypes assemble into functional heteromeric complexes. *Nature*, *396*(6712), 683-687.

76. Jones, K. A., Borowsky, B., Tamm, J. A., Craig, D. A., Durkin, M. M., Dai, M., ... & Gerald, C. (1998). GABAB receptors function as a heteromeric assembly of the subunits GABABR1 and

GABABR2. *Nature*, *396*(6712), 674-679.

77. Pin, J. P., Kniazeff, J., Liu, J., Binet, V., Goudet, C., Rondard, P., & Prézeau, L. (2005). Allosteric functioning of dimeric class CG-protein-coupled receptors.*Febs Journal*, *272*(12), 2947-2955.

78. Kristiansen, K. (2004). Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function.*Pharmacology & therapeutics*, *103*(1), 21-80.

79. Hopkins, A. L., & Groom, C. R. (2002). The druggable genome. *Nature reviews Drug discovery*, *1*(9), 727-730.

80. Kenakin, T. (2004). Principles: receptor theory in pharmacology. *Trends in pharmacological sciences*, *25*(4), 186-192.

81. Greasley, P. J., & Clapham, J. C. (2006). Inverse agonism or neutral antagonism at G-protein coupled receptors: a medicinal chemistry challenge worth pursuing?. *European journal of pharmacology*, *553*(1), 1-9.

82. Eglen, R. M., & Reisine, T. (2011). GPCRs revisited: New insights lead to novel drugs. *Pharmaceuticals*, *4*(2), 244-272.

83. Christopoulos, A., & Kenakin, T. (2002). G protein-coupled receptor allosterism and complexing. *Pharmacological Reviews*, *54*(2), 323-374.

84. Neubig, R. R., Spedding, M., Kenakin, T., & Christopoulos, A. (2003). Update on terms and symbols in quantitative pharmacology. *NC-IUPHAR*, *38*(55), 597-606.

85. Blakeney, J. S., Reid, R. C., Le, G. T., & Fairlie, D. P. (2007). Nonpeptidic ligands for peptide-activated G protein-coupled receptors. *Chemical reviews*,*107*(7), 2960-3041.

86. Beeley, N. R. (2000). Can peptides be mimicked?. *Drug discovery today*, *5*(8), 354-363.

87. Vagner, J., Qu, H., & Hruby, V. J. (2008). Peptidomimetics, a synthetic tool of drug discovery. *Current opinion in chemical biology*, *12*(3), 292-296.

88. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., ... & Miyano, M. (2000). Crystal structure of rhodopsin: AG protein-coupled receptor. *science*, *289*(5480),

739-745.

89. Caffrey, M. (2003). Membrane protein crystallization. *Journal of structural biology*, *142*(1), 108-132.

90. Loll, P. J. (2003). Membrane protein structural biology: the high throughput challenge. *Journal of structural biology*, *142*(1), 144-153.

91. Engel, C. K., Chen, L., & Privé, G. G. (2002). Insertion of carrier proteins into hydrophilic loops of the Escherichia coli lactose permease. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1564*(1), 38-46.

92. Cherezov, V., Clogston, J., Misquitta, Y., Abdel-Gawad, W., & Caffrey, M. (2002). Membrane protein crystallization in meso: lipid type-tailoring of the cubic phase. *Biophysical journal*, *83*(6), 3393-3407.

93. Sarkar, C. A., Dodevski, I., Kenig, M., Dudli, S., Mohr, A., Hermans, E., & Plückthun, A. (2008). Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity. *Proceedings of the National Academy of Sciences*, *105*(39), 14808-14813.

94. Rasmussen, S. G., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., ... & Kobilka, B. K. (2007). Crystal structure of the human &bgr; 2 adrenergic G-protein-coupled receptor. *Nature*, *450*(7168), 383-387.

95. The GPCR Network at http://gpcr.usc.edu/

96. Bissantz, C., Bernard, P., Hibert, M., & Rognan, D. (2003). Protein-based virtual screening of chemical databases. II. Are homology models of g-protein coupled receptors suitable targets?. *Proteins: Structure, Function, and Bioinformatics*, *50*(1), 5-25.

97. Becker, O. M., Shacham, S., Marantz, Y., & Noiman, S. (2003). Modeling the 3D structure of GPCRs: advances and application to drug discovery. *Current opinion in drug discovery & development*, *6*(3), 353-361.

98. Hillisch, A., Pineda, L. F., & Hilgenfeld, R. (2004). Utility of homology models in the drug discovery process. *Drug discovery today*, *9*(15), 659-669.

CHAPTER 2:

Methodology

2.0 Chapter summary

Molecular target (receptor) characterization through computer modeling techniques and application of their results in the drug discovery process is an important aspect of development of new pharmacotherapies. Novel modeling techniques, in particular the way to use theoretical models as a basis for rapidly searching large three dimensional structure activity relationship databases for the discovery of molecules with therapeutic values among previously available compounds, is an area of technological advancement that has rapidly grown in the recent years.

The research studies carried out in this thesis will be done using computational chemistry methods. Computational chemistry is the development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, chemical and behavior of systems. Traditionally, forming large models of systems has been via a mathematical model, which attempts to find analytical solutions to problems and thereby enable the prediction of the behavior of the system from a set of parameters and initial conditions.

This chapter aims to explain the theoretical aspects of the methodologies used to model intermolecular interactions of molecular targets with ligands of interest. There are two levels of modeling molecular interactions. One is based on quantum mechanics, which was not used in this thesis study and thus will not be described in this chapter. The other, molecular mechanics uses empirical forms which was the method of choice for the present study. The specific methods used in each of the particular topics develop in this work are described in each of the corresponding thesis chapters. Henceforth, merely the theoretical and methodological aspects will be discussed here and are valid for chapter 3-6 when nothing else is stated.

2.1 Introduction

As a distinct scientific discipline, drug discovery is a 20th century endeavor that has undergone continuous evolution as the scientific skill bases that support the process have increased in sophistication. From the historical perspective, much of the process of drug discovery through the early 1940s was serendipitous and dependent on plant sources. The technology base was synthetic chemistry, often based on dyes, with qualitative testing of compounds as anti-infective agents, in whole animals, with microbiology and limited biochemistry to support compound evaluation. Among the drugs from this era were the fungal antibiotics [1-4].

Advancements in enzymology and protein biochemistry led to the emergency of a second phase of drug discovery. Many of the biological pathways and processes were identified as enzymologists and pharmacologists discovered new enzymes, receptor ligands, and their functions. Thus, though serendipity was still a major factor, drugs were now directed toward distinct molecular targets. The evolution of this phase was rapid, with increased institutional and commercial support of the biomedical sciences, growing sophistication in computational processing and the availability of the personal computer [3, 5].

The third phase of drug discovery is represented by the popular vision of computer aided discovery and development of molecules that not only treat the symptoms of the disease , such as in the case of asthma, but also lead to finding the cure for the disease. The potential for rigorous targeting of therapeutic agents to molecular targets whose genetics, structure, function, and pathophysiology are well understood is a noble goal [3].

But such a noble goal is hampered by uncertainty. For example, when a molecular target is selected there is still uncertainty as to whether stimulation or inhibition of that target will be effective in a given disease state. This makes the whole process to be hypothesis driven. Thus, to a major extent, the design, synthesis, and testing of new molecules selective for a given target frequently provide the tools by which a hypothesis can be evaluated. If the hypothesis proves to be wrong, information is added to the scientific literature but no drug is developed [6-7].

This is the working approach to drug discovery that is frequently unappreciated and unreported. If drug targets and the design of their ligands were as simple as the application of new technologies, more than a higher percentage of the effort in drug discovery would reach fruition in the identification of clinical candidates. As it stands, 80 percent of the effort is valuable, if ultimately nonproductive, hypothesis testing [7-8].

However there are high hopes in drug discovery research afforded by the development of a high degree of "rationality" (rational drug design). Rationality in drug discovery process represents an intellectually rigorous approach incorporating computer-aided molecular design, limited but sophisticated chemical synthetic effort and highly focused biological assays [9-10]. More often, rational drug design suggests with the knowledge of the three-dimensional (3D) structure and sequence of various drug targets, to design new compounds by iteration, on a computer. Thus, the drug design process may become significantly less of a risk and more quantitavely resource efficient [8, 10].

2.2 Computer-Aided drug discovery

Computer-aided or *in silico* drug discovery consists of the use of computational methodologies based on established chemical and biophysical knowledge and their application in the drug discovery process. The process can be carried out by comparison of the structural features of a set of diverse ligands or by studying the structure of the ligand (drug) - receptor (target) complex with the aim to postulate ligand refinements. Thus, *in silico* methods can be classified as ligand based or structural based methodologies [11-13].

In silico methods primarily aim to rationalize, increase the efficiency, speed and cost-effectiveness of the drug discovery process. Some successful applications of *in silico* methods in drug discovery include identification of hits against variety of target families (including GPCRs, protein kinases, proteases and nuclear hormone receptors. *In silico* methods also boast success in elucidation of

structure-activity relationships of compounds and lead optimization [14-19].

Rapid advancement in cheaper, faster computational power and accessibility coupled with a continuous improvement in biophysical modeling algorithms are the primary reasons that *in silico* methods are now firmly established as complementary methods to experimental drug discovery [20]. *In silico* methods are routinely applied in the development of high-quality 3D protein models (e.g. *homology modeling*) [21-23], identify hits using *molecular docking* or virtual screening methods, optimize lead compounds, predict biological activities using compound-based modeling methods such as quantitative structure activity relationship, QSAR [24], predict the absorption, distribution, metabolism, excretion and toxicity profiles of compounds [25], and design focused chemical libraries for experimental screening design [26].

2.2.1   Ligand-based in silico methods

The underlying principle for ligand-based *in silico* methods is that, molecules with similar chemical structure will have similar biological profiles [27-28]. These methods are chemistry aware, relatively easier to develop and aim to identify compounds based on a small set or even one single active compound. They are further classified into 2D or 3D methods. In 2D methods, molecules are represented as graphs where nodes correspond to atoms and edges to bonds or in a linear notation such as SMILES. To identify similar compounds in a 2D space graph theory methods or fingerprint methods are used to represent a fixed length vector of chemical substructure characteristics [29]. 3D methods such as pharmacophore based approaches use a 3D representation of a molecule and its relevant stereochemical features. This requires steric or electronic features for optimal protein-ligand interaction and subsequent biological activity, to develop a computational model that specifies spatial relationships between the pharmacophoric features. A pharmacophore can be defined as the abstract description of the molecular features which are necessary for molecular recognition of a ligand by a biological macromolecule. Complementary features from the protein may also be incorporated in the model building process. However in case of pharmacophore models the conformational space of

compounds must be determined and a biologically relevant conformation must be selected.

Figure 2.1 illustrates how ligand-based methods are implemented. Ligand based methods are easier to develop and implement, but fall short in the identification of novel compounds for drug discovery or rationalizing bimolecular interactions. The second type of computer-aided drug design i.e. structure-based drug design offers several methods to complement ligand-based computational methods [30-32].
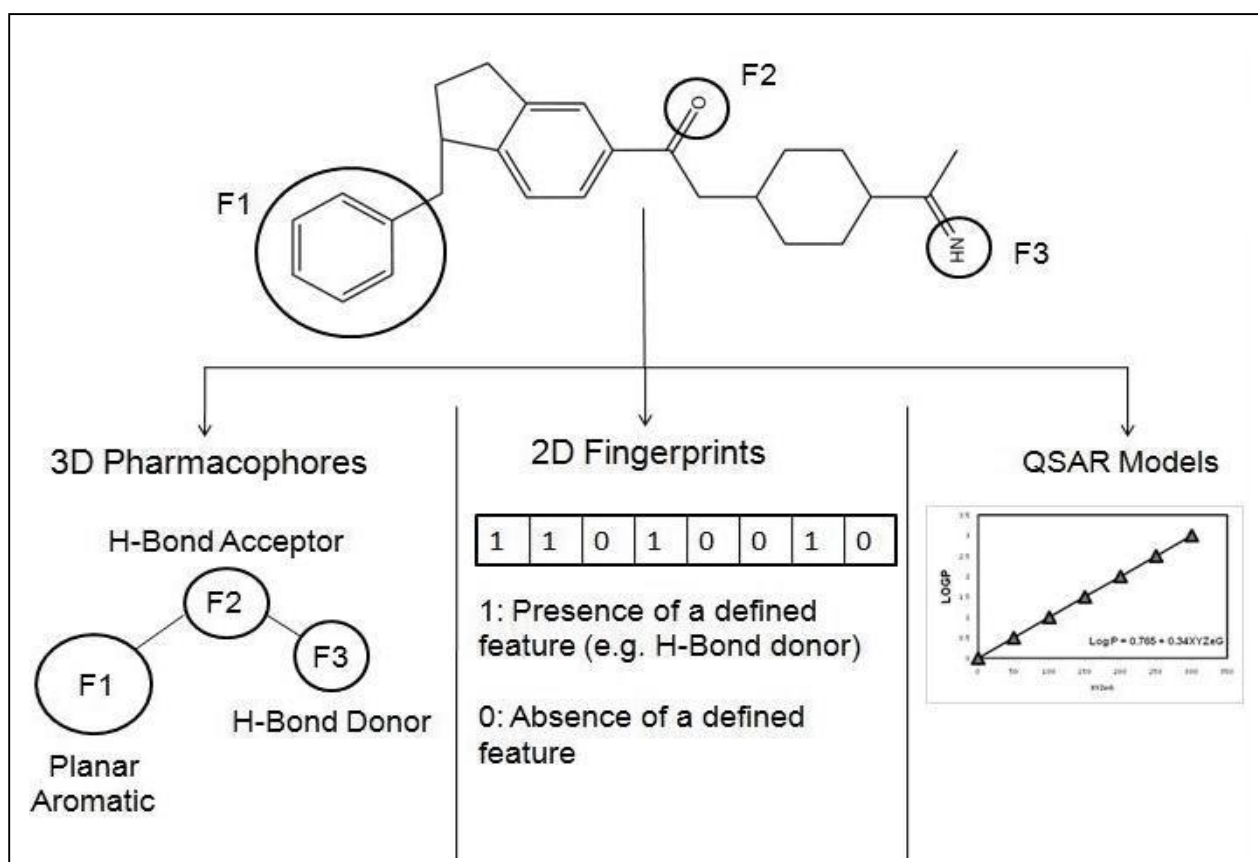
Figure 2.1: The ligand based method scheme, whereby the molecules are either are represented as graphs where nodes correspond to atoms and edges to bonds (3D pharmacophores) or as a fixed-length vector of chemical substructure attributes (2D fingerprint). The QSAR ligand based design correlates compound structural features to its biological activity by a mathematical model [22].

2.2.2   Structure-based in silico methods

In drug discovery process structure-based drug design is a computational process that relies

on the availability of 3D protein structures or models. Molecular docking or high throughput docking

(the computational equivalent of high throughput screening) are some of the most important aspects

of structure-based drug design methods. Structure based design process consists of identifying

potential binding sites where the small molecules (ligand) may interact with the target, a virtual

screening of chemical library, the docking process where compounds are positioned in the binding site

and scoring compounds that estimates the likelihood of binding to the target (scoring and ranking [22,

33-34].
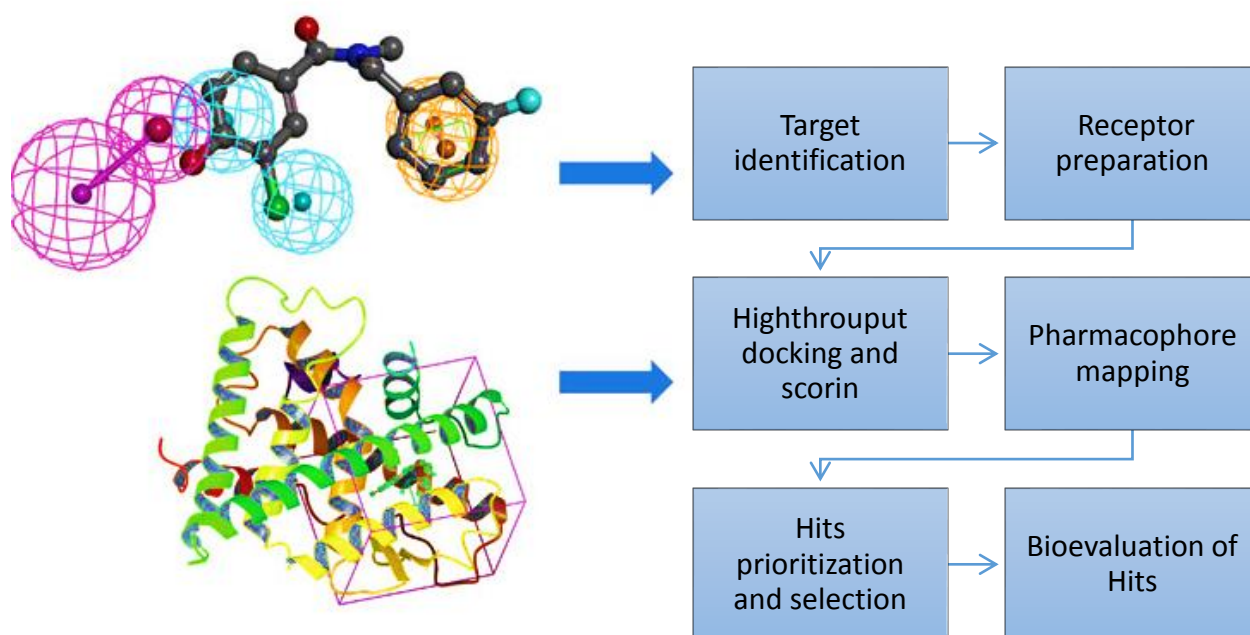
Figure 2.2: The structure based drug discovery scheme as applied in this thesis study.

Molecular docking aims to prioritize a reduced set of compounds that have a higher

probability of being active in subsequent experimental analysis. This methodology has been

successfully applied in several drug discovery stages e.g.: development of target-specific compound libraries, hit identification, and lead optimization. Docking applications have been benefited by the continual development of docking algorithms, cheaper and faster computational resources and an increasing availability of high quality protein structures [35-37]. However, molecular docking and in turn structure-based drug design has two major problems such as 3D crystallographic structures of many therapeutic proteins are currently unavailable [23] and poor handling of receptor or ligand flexibility in docking algorithms [38].

A rational computational drug discovery study requires a high-quality 3D structure of the target receptor. So that, the availability of a high-resolution crystal structure of the target receptor is of critical importance. Crystallographic structures are considered as close as you can get to "reality" of a biochemical event. Where GPCRs are concerned the number of available crystallographic structures is dismal in comparison to the number of GPCR in the human to study (there are more than 30 GPCRs crystallized, while there are approximately 800 GPCRs). As mentioned in previous chapter this is mainly due to the difficulties associated with crystallizing the GPCRs [39-40].

In the absence of crystallographic structures of GPCRs and other therapeutic proteins, computational methods are used to develop 3D protein models at atomic resolution. Computationally modeled proteins may be considered reasonable substitutes until a corresponding crystallographic structure is available. Protein models may provide structural and biochemical functional insights to researchers for various biomedical research needs. However models developed by automated methods have limited accuracy and applicability for drug discovery projects. Henceforth the development of novel and more accurate protein modeling methods is an important problem and an area of active research in the drug discovery domain [20, 23, and 39].

2.3 Homology modeling

The fundamental principle of homology modeling is that proteins with similar sequence will share similar 3D structures. Homology modeling consists of four main steps. The first two steps are the identification of a homologous protein preferably with a high-resolution crystallographic protein structure which will serve as the template and sequence alignment stage. Sequence alignment is primarily a bioinformatics based method in which two or more protein sequences are rearranged such that similar protein residues and regions (motifs) are aligned in a columnar representation. Sequence alignment is particularly important as it serves as the input and basis of homology model building process. The third stage is model construction and the final stage is model assessment.

While the first two steps are often essentially performed together, as the most common methods of identifying templates rely on the production of sequence alignments, the final process can be performed iteratively to improve the quality of the final model [41-44]. The degree of homology and quality of alignment often plays an important role in the quality of the model structure. Generally, if the sequence identity between template-target exceeds 50% highest quality models are generated and if it is less than 25% such models are speculative at best [21]. Figure 2.3 shows the 4 stages of homology modeling and what steps are taken in each stage.

The model construction stage involves transferring of the 3D Cartesian coordinates of the template to the corresponding residues of the target sequence obtained from the sequence alignment stage. Then subject to restraints satisfaction a rough or crude model is constructed. Spatial restraints expressed as probability density functions such as used in the software MODELLER and MOE (used extensively to develop preliminary homology models in this thesis) were developed from extensive statistical analysis of protein families that quantify relationships between alpha carbons (Cα) distances between backbone residues and corresponding main chain dihedral angles. The probabilities are calculated from the range of distances and angles of a particular residue type, conformation and sequence similarities between proteins present in the set [45].
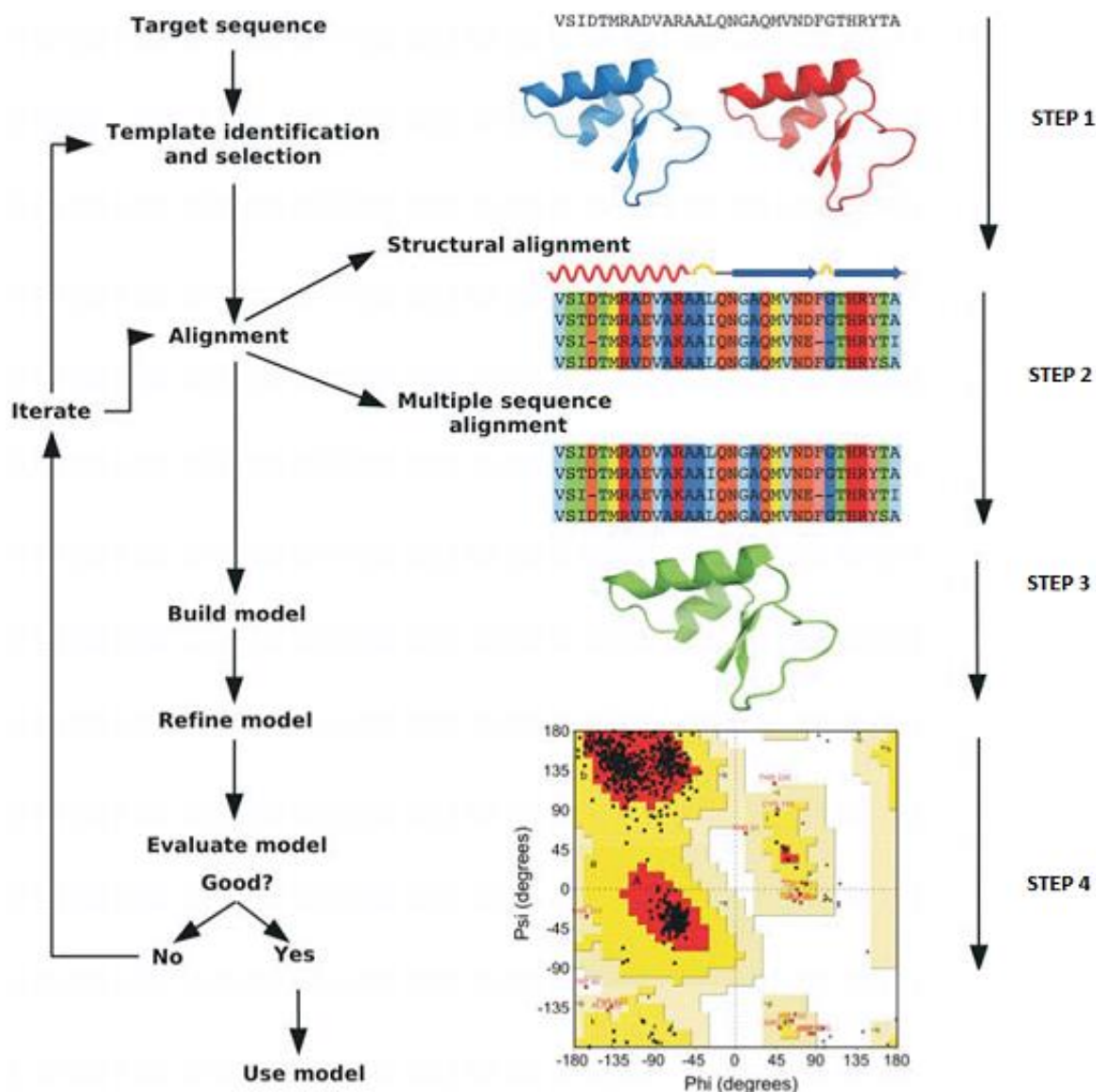
Figure 2.3: A simplified illustration of the homology modeling process, adopted from [46].

Regarding the loops, Modeller builds loops by optimizing a series of probability density functions describing backbone geometry based on amino-acid type, and then refined with an energetic minimization procedure [45]. While for MOE, Loops are modeled first, in random order. For each loop, a contact energy function analyzes the list of candidates collected in the segment searching stage, taking into account all atoms already modeled and any atoms specified by the user as belonging to the model environment (e.g. a ligand bound to the template or structural waters). These energies are

then used to make a Boltzmann-weighted choice from the candidates, the coordinates of which are then copied to the model [47]. Generally, rough models are built to maximize the molecular probability density functions given the sequence alignment between the target and template proteins. The crude model is then iteratively optimized to a low energy state to refine the positions of all heavy (non-hydrogen) atoms. Energy minimization using conjugate gradient method is used to refine the crude models (45).

In the final stage, models are assessed using the discrete optimized protein energy method [48]. They assign a probability or energy score to each possible pairwise interaction between protein amino acids and combine them into a single score for a given model. Typically, several models are built and ranked before selection for advanced refinement procedures or direct applications. The protein geometry of the final model should be inspected with quality evaluation tools such as the Ramachandran plot of the backbone dihedral angles in order to confirm that the model is sterically sound and consistent with typical values found in crystal structures [45, 47-48].

Overall each step of the modeling process can contribute to the errors in the final model. Depending on the degree of sequence identity and the quality of alignment their accuracy as compared to the crystal structure can be up to less than 2Å Cα atom root-mean-square deviation (*rmsd*). Errors of this magnitude, particularly in the binding site, can result in wrong side chain orientations thus misrepresenting receptor-ligand complementarity. This can potentially negate the applicability of the model in a drug discovery project. Thus it is important to develop and apply methods that incorporate receptor flexibility in the modeling process to generate accurate side-chain orientations to improve quality of protein models [49-50].

In the drug discovery process, the use homology modeling has been successfully applied in various stages of the process. For example, homology models have been applied among others to discover novel leads for MCHR-1 receptors for anti-obesity [51, in building cdc25 phosphatase for structural based virtual screening of leads for anti-cancer therapeutics [52], alpha-glucosidase for diabetes [53] and CK1-delta for Alzheimer therapeutics [54]. In 14 some other examples, homology models have

been developed to investigate the role of certain residues in the biological function of the protein via mutagenesis experiments [55-56].

Class A-GPCRs are a special case for homology modeling as they exhibit low sequence identity among members of the family (~20%). However this drawback is compensated by the conserved 7 TM structural characteristic shared by all Class A GPCRs and the conserved motifs found along the TMs [57]. The current consensus of opinion in GPCR homology modeling is that the template selection should be data driven and knowledge-based, incorporating information such as high sequence similarity and a priori assessment of the expected structural features of the receptor to model such as loop morphology, similarity in binding-pocket regions etc. [57-58]. Chapter 3 of this thesis is dedicated solely to investigation of the effects template selection has on the homology modeling.

The modeled receptor is a limited approximation for use in drug discovery since its construction does not consider the rearrangement due to the bound ligand (a molecular docking effect) and the effect that the lipid bilayer may cause to the protein structure. These two effects can be captured in the model thorough molecular dynamics simulations [59].

2.4 Molecular docking

In recent years molecular docking has become an important tool for drug discovery. The focus of molecular docking is to computationally simulate the target-ligand recognition process. The associations between biologically relevant molecules such as proteins, nucleic acids, carbohydrates, and lipids play a central role in signal transduction. Furthermore, the relative orientation of the two interacting partners may affect the type of signal produced (agonism vs. antagonism). Therefore docking is useful for predicting both the strength and type of signal produced [60-61].

The goal of molecular docking is to achieve an optimized conformation for both the protein and ligand and relative orientation between protein and ligand such that the free energy of the overall system is minimized. To understand the ruling principles whereby protein receptors recognize,

interact, and associate with molecular substrates and inhibitors is of paramount importance in drug discovery efforts [60]. Protein-ligand docking aims to predict and rank the structures arising from the association between a given ligand and a target protein of known 3D structure. The fact that proteins are in constant motion between different conformational states with similar energies is still often disregarded in docking studies, even though protein flexibility is known to allow increased affinity to be achieved between a given drug and its target [60, 62].

Furthermore, binding-site location and binding orientation can be greatly influenced by protein flexibility. Combining all these aspects, it is very important to have a good docking program. There have been a number of attempts to compare the strength of different docking programs, where the comparison is often divided into assessment of the search algorithm and the scoring function. Scoring functions are fast approximate mathematical methods used to predict the strength of the non-covalent interaction (also referred to as binding affinity) between two molecules after they have been docked. Each attempt at an unbiased comparison clearly stated that it is very difficult to rank and assess different optimization strategies, particularly with increased algorithms complexity [60, 62-64].

2.4.1 Protein-ligand docking

Predicting the binding modes and affinities of compounds when they interact with a protein-binding site lies at the heart of structure-based drug design. The protein-ligand docking in this thesis study have been done using Genetic Optimization for ligand docking (GOLD) software for rough models [65], and Glide software from Schrodinger suite of programs for refined models [66-67]. Generally speaking, once you have decided, your ligands (drugs) and target protein, the docking program, will "dock" your ligands to the target protein by exploring many biophysically possible ways that the given compound(s) can bind to the target. The combination of position and orientation of a ligand relative to the receptor along with its conformation in docking, is known as a ligand pose, and is assigned a score based on the pre-determined scoring functions. Upon completion of docking all compounds to the target protein, they are ranked and selected for subsequent validations based

on their corresponding scores and binding poses [60].

GOLD uses a genetic search algorithm and allows for full ligand flexibility, as well as rotational flexibility for the protein-receptor polar hydrogens [60, 62, 68-70]. GOLD has one of the most comprehensive test sets from the entire surfeit of docking programs, comprising 100 different protein complexes and has achieved a 71% success rate in identifying the experimental binding models. However, systematic problems in ranking very polar ligands and in ranking general ligands in large cavities have been reported [60, 65, 70]. Like all docking programs, GOLD has three main parts, first, a scoring function to rank different binding modes; the Goldscore function is a molecular mechanics–like function with four terms:

$$GOLD\ Fitness\ =\ s_{hb\_ext}\ +\ s_{vdw\_ext}\ +\ s_{hb\_int}\ +\ s_{vdw\_int}$$

Where by :

- S_{hb_ext} is the protein–ligand hydrogen-bond score

- S_{vdw_ext} is the protein-ligand van der Waals score.

- S_{hb_int} is the contribution to the Fitness due to intramolecular hydrogen bonds in the ligand.

- S_{vdw_int} is the contribution due to intramolecular strain in the ligand.

Second part of GOLD is the mechanism for placing the ligand in the binding site; GOLD uses a unique method to do this, which is based on fitting points; it adds fitting points to hydrogen bonding groups on protein and ligand, and maps acceptor points on the ligand on donor points in the protein and vice versa. Additionally, GOLD generates hydrophobic fitting points in the protein cavity onto which ligand CH groups are mapped. The last part of GOLD software is the search algorithm to explore possible binding modes; GOLD uses a genetic algorithm (GA) in which the parameters such as dihedrals

angles of ligand rotatable bonds, ligand ring geometries (by flipping ring corners), dihedrals angles of

protein -OH groups and NH3 groups and the mappings of the fitting points (the position of the ligand

in the binding site) are modified and optimized [65, 70]. At the start of a docking run, all these variables

are randomized. Whenever GOLD software was used in this thesis chapter, the default GOLD

parameters and Goldscore scoring default parameters were used which includes flipping ring corners,

planar and pyramidal nitrogen, allowing rotatable H_2O molecules, diverse poses during placing and

rotating interacting hydrogen bonds [70].

Glide software on the other end, can run in rigid or flexible docking modes treating the ligand as

a rigid body or generating conformations for the input ligand respectively. The ligand poses that Glide

generates pass through a series of hierarchical filters that evaluate the ligand's interaction with the

receptor. The initial filters test the spatial fit of the ligand to the defined active site, and examine the

complementarity of ligand-receptor interactions using a grid-based method patterned after the

empirical scoring function. Poses that pass these initial screens enter the final stage of the algorithm,

which involves evaluation and minimization of a grid approximation to the all atoms optimized

potentials for liquid simulation (OPLS-AA) non-bonded interaction energy (ligand receptor). Finally,

the minimized poses are re-scored using Glidescore scoring function [66-67, 71].

Glidescore is based on initial empirical scoring function, but includes a steric-clash term, adds

other rewards and penalties such as buried polar terms (devised by Schrödinger to penalize

electrostatic mismatches), amide twist penalties, hydrophobic enclosure terms, excluded volume

penalties and is calculated by;

$$\textbf{\textit{Glidescore}} = (0.05 * vdW) + (0.15 * Coul) + Lipo + Hbond + Metal + Rewards$$

The van der Waals term (*vdW*) calculates repulsion and attraction parameters between the atoms

of the receptor and the ligand. The force is very significant if the contact surface is large and when the

molecules are close. The Coulomb energy term (*Coul*) calculates electrostatic interactions between the atoms of the receptor and the ligand. This force draws molecules closer together or further apart according to their electrical charge. The two terms are calculated with reduced net ionic charges on groups with formal charges (e.g. carboxylates and metals). The lipophilic term (*Lipo*) is calculated for lipophilic receptor- and ligand atoms, rewarding favorable hydrophobic interactions. The Hydrogen bond term (*Hbond*) is calculated for all complementary possibilities of hydrogen bonds between the atoms of the receptor and the atoms of the ligand. Hydrogen bonds give a favorable contribution to affinity. If the ligand is a strong hydrogen bonder the solvent will attract it and make the receptor less favorable. If the receptor has hydrogen bonding qualities this can attract the ligand into binding. The term is separated into three different components depending on the donor and acceptor. Either both are neutral, one is charged and one is neutral or both are charged. The first of these contributions (both neutral) is found to be the most stabilizing and the last (both charged) is the least important. The metal-ligand interaction term (*Metal*) is calculated for metal atoms in the receptor and all acceptor and acceptor/donor atoms in the ligand. Only the interactions with highly polar acceptor atoms or anionic atoms are included. Only if the net metal charge in the apoprotein (enzyme without its prosthetic group) is positive, the preferences for polar or anionic ligands are included . The Site term (*Site*) calculates polar interactions in the active site. Non-hydrogen-bonding polar atoms in a hydrophobic region are rewarded.

The term called *Rewards* provides rewards and penalties for different features, such as a hydrophobic enclosure, buried polar groups and other terms that are not explicitly mentioned. In addition penalties for frozen rotatable bonds (*RotB*) are given. If the ligand is very flexible the equilibrium will be shifted against the solvent because the degree of disorder (ΔS) is greater here, thus causing a lower free energy. The free energy of the system should be as small as possible. A low (negative) energy indicates a stable system, and thus a likely binding interaction [66-67].

In this study, when Glide software was used for molecular docking it was used in flexible mode and scoring done using "extra precision" Glidescore (XP score) scoring method . Flexible mode

automatically generates conformations for each input ligand and the XP scoring method aims to weed out false positives and to provide a better correlation between good poses and good scores. The XP scoring function includes additional terms over the normal Glidescore scoring function. Specifically, XP score rewards occupancy of well-defined hydrophobic pockets by hydrophobic ligand groups. Hydrophobic reward terms are employed in Glidescore in the form of lipophilic-lipophilic pair terms, while other empirical scoring functions use lipophilic surface-area contact terms for this purpose. Research have shown that simple pair terms underestimate hydrophobic effects in certain well-defined cases. The hydrophobic term in XP score was developed to offset this underestimation. The term confers up to several kcal/mol of additional binding energy in favorable cases, and substantially improves enrichment factors. XP score also includes improvements to the scoring of hydrogen bonds as well as detection of buried polar groups, and detection of π-cation and π-π stacking interactions [72].

In a comparison study of protein-ligand docking programs, it was noted that Glide software exhibits excellent docking accuracy, high enrichment across a diverse range of receptor types and reliably finds the correct binding modes for a large set of test cases .i.e. virtual screening. It outperformed other docking programs in achieving lower *rmsd* from native co-crystallized structures especially is [60]. Thusly, for the final, refined model all the molecular docking calculations, either virtual screenings or pharmacophore based was done using Glide software, while GOLD was used for initial molecular docking in the rough models due to the high flexibility exploration the software offers.

2.4.2 Protein –protein docking

Proteins often carry out their functions through the association with other proteins forming multi-protein complexes. Several theoretical algorithms have been designed to predict the 3D structure of protein-protein complexes using docking algorithms for this purpose. These programs as well as Web servers such as PIPER, GRAMM-X, HEX, 3D Garden, SmoothDock, Rosetta and PatchDock aim to assemble two separate protein components into a biologically relevant complex structure, ideally

close to the native structure [73].

The ability to accurately predict protein complexes also opens the potential for rational drug design and protein engineering [74-75]. A comparison study on the diverse protein-protein docking programs concluded that due to its features, Global RAnge Molecular Matching (GRAMM) software is one of the best [76], henceforth the protein-protein docking calculation performed in this study was done using GRAMM (GRAMM docking).

GRAMM docking uses an empirical approach to smoothing the intermolecular energy function by changing the range of the atom-atom potentials. The technique locates the area of the global minimum of intermolecular energy for structures of different accuracy. The quality of the prediction depends on the accuracy of the structures. Thus, the docking of high-resolution structures with small conformational changes yields an accurate prediction, while the docking of ultra-low-resolution structures will give only the gross features of the complex. [73, 77].

Protein docking is based on complementarity between interacting proteins which could be geometric, electrostatic, hydrophobic, or all three. To create different docking models, one protein is fixed in space while the second protein is rotated and translated around the stationary protein. Performing a truly comprehensive global search with the receptor and ligand would require a large investment of computational power which is not practical with current resources. As a result most protein docking prediction programs start with rigid body searches or by geometric matching. In this initial stage GRAMM simplify the protein structure as a rigid body representation on a 3D Cartesian grid then a docks by searching, using Fast Fourier Transform (FFT) based method, for degrees of overlap between the pairs of grids. FFT-based method of searching allows the evaluation of many docked conformations on a grid using a correlation-type scoring function [78-79]. For example for two protein (a and b), projected in 3D grid (N x N x N) points, their discrete functions are presented by;

$$a_{l,m,n} = \begin{cases} 1 & inside\ the\ molecule \\ \\ 0 & outside\ the\ molecule \end{cases}$$

[1a]

And

$$b_{l,m,n} = \begin{cases} 1 & inside\ the\ molecule \\ \\ 0 & outside\ the\ molecule \end{cases}$$

[1b]

Whereby *l*, *m*, and *n* are the indices of the 3D grid (*l*, *m*, *n* = *{1...N})*. Any grid point is considered inside the molecule if there is at least one atom nucleus within a distance r from it, where *r* is of the order of van der Waals atomic radii. To distinguish between the surface and the interior of each molecule, a value of 1 is retained for the grid points along a thin surface layer only and other values are assigned to the internal grid points. Therefore:

$$\bar{a}_{l,m,n} = \begin{cases} 1 & on\ the\ surface\ of\ the\ molecule \\ \rho & inside\ the\ molecule \\ 0 & outside\ the\ molecule \end{cases}$$

[2a]

And

$$\bar{b}_{l,m,n} = \begin{cases} 1 & on\ the\ surface\ of\ the\ molecule \\ \rho & inside\ the\ molecule \\ 0 & outside\ the\ molecule \end{cases}$$

[2b]

Whereby the surface is defined as a boundary layer of finite width between the inside and the outside of the molecule. The parameters p and 8 describe the value of the points inside the molecules, and all points outside are set to zero. GRAMM docking accomplished stage one by matching surface through calculating correlation functions. The correlation between the discrete functions $\bar{a}$ and $\bar{b}$ is defined as:

$$\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \bar{a}_{l,m,n} \cdot \bar{b}_{l+\alpha,m+\beta,n+\gamma,}$$

[3]

Where by α, β and γ are number of grid steps by which molecules **b** is shifted in respect to molecule **a** in every dimension. Figure 2.4 shows the outcome of equation **[3]**. If there is no contact between the two molecules **[3]** = zero, If there is a contact between the surfaces the contribution to the correlation value **[3]** is positive. If one molecule penetrates into the other **[3]** (figure 2.4 **c**). Such penetration is physically forbidden, thus a distinction between surface contact and penetration must be clearly formulated by assigning large negative values to ρ in $\bar{a}$ and small non-negative values to δ in $\bar{b}$. Hence, if the shift vector { α, β, γ} is such that molecule b penetrates molecule a, the multiplication of the negative numbers (ρ) in $\bar{a}$ by the positive numbers (1 or δ) in $\bar{b}$ results in a negative contribution to the overall correlation value [80].

Positive correlation values are obtained when the contribution from surface contact outweighs that from penetration. Thus, a good geometric match (as in figure 2.4 d) is represented by a high positive peak, and low values reflect a poor match between the molecules. Simply put, correlation value for each displacement is simply the score for overlapping surfaces corrected by the penalty for penetration [80].
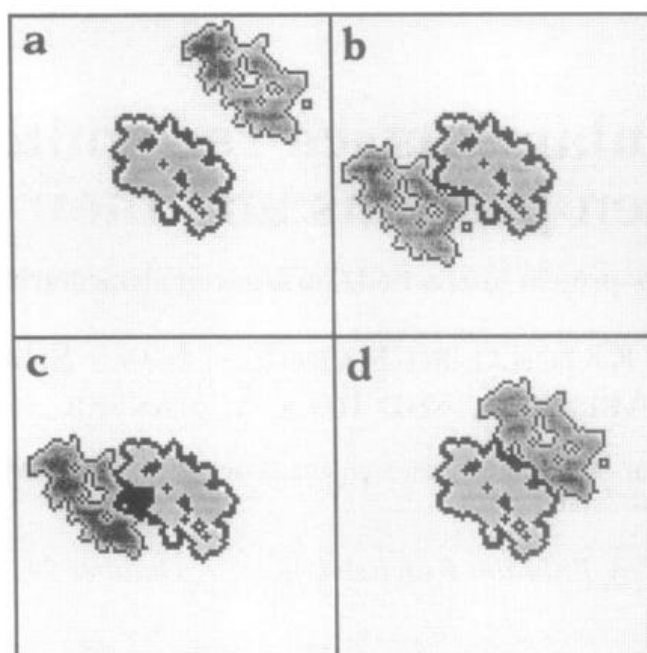
Figure 2.4: Different relative positions of 2 molecules. **a**) No contact, **b**) limited contact, **c**) penetration (in black), and **d**) good geometric match [80].

Since the initial docking stage will generate multiple docked model complexes, they will need to be filtered to select the best conformations based on their scores [78]. GRAMM re-ranks the models resulting from initial stage by running a local minimization with soft van der Waals interaction, clustering of predictions within the same local minima, and rescoring with the target function combining Lennard-Jones [79]. Finally, the models can be refined by incorporating flexibility into the protein sidechains as well as the backbone [80]. GRAMM-X displays their top scoring models based on soft Lennard-Jones potential, evolutionary conservation of predicted interface, statistical residue-residue preference, empirical binding free energy and atomic contact [79, 81]. At this moment, the models are ready for the utilization.

2.5 Molecular Dynamics Simulations

Despite the fact that they are normally represented as static structures, molecules such as protein are in fact dynamic. In this thesis, study chapters 3-6 uses techniques of classical molecular dynamics simulation (MD simulation) to refine the structures resulted from homology modeling. This point aims to give a basic introduction to the equations that the method of MD relies on.

All the atoms of a molecule are subject to a harmonic movement because of forces, which result from the interactions between atoms. And although the main purpose of molecular dynamics simulations is modeling of molecular motions, molecular dynamics simulations are also applied for structural optimization. MD simulation calculates the time dependent behavior of a molecular system and provides detailed information on the fluctuations and conformational changes of proteins [82-83]. Thus one can study both the thermodynamic properties and/or time dependent (kinetic) phenomena.

One way to investigate the range of accessible configurations is to simulate the motions or dynamics of a molecule numerically. This can be done by computing a trajectory, a series of molecular configurations as a function of time, by the simultaneous integration of Newton's equations of motion. If the force on each atom is known, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. From this trajectory, the average values of properties can be determined. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past. [82-85].

Starting from a coordinate set taken from a starting structure, the interactions among atoms and the position are defined and velocities calculated and assigned to each atom (typically from a Boltzmann distribution at a given temperature) using a numerical method such as the Verlet method [86]. The successive coordinates and velocities are obtained by integrating the Newton's equation for

the motion in each coordinate direction according to the forces derived from the gradients of the

interaction potentials involving the atom. The equations of motion;

$$\frac{d\vec{r}_i(t)}{dt} = \vec{v}_i(t) \qquad [4]$$

and

$$\frac{d\vec{v}_i(t)}{dt} = \frac{\vec{F}_i(t)}{m_i} \qquad [5]$$

for all atoms (**i** = 1, 2 ,...,N) of the molecular system. The atomic coordinates, *r*, and the velocity V, of

atom, **i,** with mass, m_i, thus become functions of time. The force **F$_i$** exerted on atom **i** by the other

atoms in the system is given by the negative gradient of the potential energy function **V** which in turn

depends on the coordinates of all **N** atoms in the system:

$$F_i(t) = -\frac{\partial V(\vec{r}_1(t), \vec{r}_2(t), \dots \vec{r}_N(t))}{\partial \vec{r}_1(t)} \qquad [6]$$

For small time steps **δt**, **[5]** can be approximated by:

$$\vec{v}_i(t + \Delta t / 2) = \vec{v}_i(t - \Delta t / 2) + \frac{\vec{F}_i(t)}{m_i} \Delta t \qquad [7]$$

And [**4**] likewise by:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t + \Delta t / 2)\Delta t \qquad [8]$$

[**7**] and [**8**] form the so-called leap-frog scheme for integrating Newton's equations of motion. The leap-frog algorithm is used due to the fact that a large number of particles are interacting with each other [85].

The numeric integration is carried out step-wisely and is limited to the fastest motion in the system. Typically a time step of 1 to 10 fs is used for molecular systems. Thus a 100ps (10^{-10} seconds) molecular dynamics simulation involves 10^5 to 10^4 integration steps. The process is repeated at each discrete time step. The trajectory of each individual object can be tracked by connecting its states to form a time series. The time evolution of the entire system can be viewed as a fiber bundle of time series in the phase space. Even using the fastest computers only very rapid molecular processes can be simulated at an atomic level [83, 85]. Therefore, for atomistic simulations the step size is usually 1 fs, or 2 fs if restraining bond lengths, generally done using SHAKE [87] and LINCS [88] algorithms. As with any aspect of modelling, the accuracy of the predicted dynamics will depend on the validity of the underlying assumptions of the model. In this case the model is essentially defined by the force field that is used [84].

Generally, factors that govern the outcome of MD simulations includes the starting configuration of the system, how you will treat temperature and pressure (statistical ensemble), boundary conditions, solvation effect, choice of the degrees of freedom, how you treat the non-bonded interactions, integration time step (the bigger the time step the higher the computing power/cost), and the force field parameters (how the atoms are described in your system) [83, 85].

As with any aspect of modeling, the accuracy of the predicted dynamics will depend on the validity of the underlying assumptions of the model. In this case this is essentially defined by the model for the intermolecular interactions (or potential energy) used. That model is a mathematical function (force field) that describes how the value for the potential energy depends on the spatial arrangement of all the atoms [85].

2.6 Atomic Force fields

The calculation of the interaction energy within a classical description of a molecular system requires a force field. A force field is built up from two distinct components, a set of equations used to generate the potential energies (and their derivatives, the forces) and the parameters used in these equations. There are four main force fields in common use for simulating biological macromolecules AMBER [89], CHARMM [90-92], GROMOS [87] and OPLS [93]. The force field used in describing all the components of the molecular system, such as the protein, water molecules, and ions is OPLS-AA.

Atoms basically interact with each other through van der Waals forces and electrostatic forces. When they are covalently bonded to others, strong forces hold them together as stable chemical groups. The intramolecular potential energy **[9]** is typically represented by harmonic oscillators for bond stretching and angle bending, a Fourier series for each torsional angle, and Coulomb and Lennard–Jones (LJ) accounting for the interactions between atoms separated by three or more bonds. The latter two terms -referred together as non-bonded terms- are evaluated between all atom pairs in the system to yield the intermolecular energy. Such force fields compute the energy as a sum of terms representing bond elongation, angle and dihedral deformation, and non-bonded interactions with the following general form;

$$E = \sum_{bonds} V^{srt} + \sum_{angles} V^{bend} + \sum_{torsions} V^{tor} + \sum_{i,j} V_{ij}^{LJ} + \sum_{i,j} V_{ij}^{Coulomb}$$

[9]

Whereby the three first summations correspond the bonded terms (that include atoms connected up to three consecutive bonds) and the last two refer to the non-bonded ones. All summations can be easily calculated from the coordinates of the system at a given time. For each pair of bonded atoms (*i* and *j*), the stretching term is computed as;

$$V_{ij}^{srt}(r_{ij}) = k_{ij}^{r}(r_{ij} - r_{ij}^{o})^2$$

[9.1]

Whereby K_{ij}^r is the stretching force constant and r_{ij} and r_{ij}^o are the distance between atoms and its equilibrium bond length, respectively. For every group of three bonded atoms (*i, j* and *k*), the angle term is described as;

$$V_{ijk}^{bend}(\theta_{ijk}) = k_{ijk}^{\theta}(\theta_{ijk} - \theta_{ij}^{o})^2$$

[9.2]

Whereby K_{ijk}^{ϑ} is the bending force constant, ϑ_{ijk} and ϑ_{ijk}^{o} are the angle between atoms and its equilibrium value, respectively. For every group of four bonded atoms (*i, j, j, l*), the dihedral term is often represented as a cosine expansion;

$$V_{ijkl}^{tor}(\vartheta_{ijkl}) = k_{ijkl}^{\vartheta}[1 + \cos(n\vartheta_{ijkl} - \vartheta_{ijkl}^{o})]$$

[9.3]

Whereby k_{ijkl}^{φ} is a dihedral constant affecting the barrier height, n is number of minima in a 360º rotation, and φ_{ijkl} and φ_{ijkl}^{o} are the dihedral angle and the equilibrium value according the biochemical convention (trans ɸ=180º, cis ɸ=0º and gauche ɸ=60º/300º), respectively. The first non-bonded term is often represented by a 6-12 Lennard-Jones potential, a simple mathematical model [94], that accounts for two distinct forces (an attractive and a repulsive) that neutral atoms and molecules are subject to;

$$V_{ij}^{LJ}(r_{ij}) = \left[\left(\frac{A_{ij}}{r_{ij}}\right)^{12} - \left(\frac{B_{ij}}{r_{ij}}\right)^{6}\right]$$

[9.4]

Whereby A_{ij} and B_{ij} are parameters that depend from on each pair of atoms. The first term accounts for the attractive forces at long range (van der Waals or dispersion) and the other for the repulsive forces at short range, resulting from the overlap between electronic orbitals. In MD simulations, it is the non-bonded interaction that are more important than the bonded interactions. It is the non-bonded interactions among the atoms of a macromolecule that affect its secondary structure. It is the non-bonded interactions among the atoms of different molecules that organize them into crystals, complexes and other assemblies.

Finally, the last term in the force field equation **[9]** is a Coulombic potential describing the electrostatic interactions;

$$V_{ij}^{Coulomb}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}}$$

[9.5]

Where q_i and q_j are the charges of atoms *i* and j, r_{ij} the relative distance between them and ε_o the vacuum permittivity.

2.7 Statistical ensemble

In order to relate the microscopic features of individual atoms and molecules to the macroscopic or bulk thermodynamic properties of materials one needs to use statistical mechanics methods. In this context, an ensemble formalizes the notion that, repeating an experiment again and again, a physicist may expect to observe a range of different outcomes under the same macroscopic conditions, but unable to control microscopic details. Different macroscopic environmental constraints lead to different types of ensembles, with particular statistical characteristics. The three most important ensembles used in MD simulations of biological systems includes Micro-canonical ensemble (NVE), Canonical ensemble (NVT) and the Isothermal–isobaric (NPT) ensemble.

The NVE requires keeping the total energy of the system constant (e.g. system thermally isolated). The system is assumed to be isolated in the sense that the system cannot exchange energy or particles with its environment, so that (by conservation of energy) the energy of the system remains exactly known as time goes on. The initial N and V account for a constant number of atoms and volume such as a fixed box shape. Although the NVE is confined to around a constant-energy surface, it is not necessarily uniformly distributed over that surface: if the gradient of energy in phase space varies, then the NVE is more concentrated in some parts of the surface than others.

NVT is the statistical ensemble that represents the possible states of a mechanical system in thermal equilibrium with a heat bath at a fixed temperature. It shares its energy with a large heat reservoir or heat bath allowing the system to exchange energy assuming that the heat capacity of the reservoir is so large as to maintain a fixed temperature for the coupled system [99]. Like in NVE, the initial N and V account for a constant number of atoms and volume (i.e. a fixed box shape).

The NPT ensemble is analogous to the NVT one, but imposes a barostat in the same spirit as the temperature coupling that maintains a fixed pressure instead of considering a constant volume. It corresponds most closely to laboratory conditions with a flask open to ambient temperature and pressure [99].

For lipid bilayers, pressure control occurs under constant membrane area (NPAT) or constant surface tension "gamma" (NPγT). Thus, the NPT ensemble will be used in this study. As previous stated, all the molecular components of the MD systems in this thesis study will be described using the OPLS-all atom force field parameters. This includes for the water model TIP3P, TIP3P water model has been shown to provide a good compromise between quality and computational cost and thus will be used the water model for the simulations in this study [95]. Ions and lipids palmitoyl-oleyl phosphatidylcholine (POPC) whose force field were taken from OPLS for its use in the GROMACS package 4.6 [96-98]. As an added advantage, OPLS all-atom force field includes explicit representation of all atoms, including non-polar hydrogens [93].

2.8 Conclusions to chapter 2

Once a decision has been made to target a particular therapeutic area (such as analgesic, antidepressants, anti-infective etc.), it is very clear the road to drug discovery will depend on a number of things such as the existence of potential molecular targets, existing technologies and scientific expertise among others. While others factors such as marketability of the drug play important role, it is the scientific side of the process that will determine the important aspects of the drug discovery process (e.g. Duration, costs).

If the molecular targets do exist in a researchable manner, for example crystallographic structures, the drug discovery process can proceed via ligand based or structure based method needing both in house and commercial application to be used in process such as molecular docking, high-throughput virtual screening, pharmacophore design etc. The drug discovery process will become more costly both in material/methodology and time consuming if there is no molecular targets available in researchable manner. At this stage atomistic homology models of the molecular targets will have to be constructed and refined or optimized. This will involves the time-consuming and often resource consuming (computing powers) methods such as MD simulation to prepare the molecular targets ready for the drug discovery process.

This chapter dealt with the discussion of the theoretical and methodological aspects of the whole process of preparing the molecular targets ready for the drug discovery process albeit one should consider these preparatory process is also part of the drug discovery process as it set a tone and have a considerable impact on the outcomes of the drug discovery process. That is to say, good methodology makes better drug discovery tools and leads to good drug.

2.9 References to chapter 2

1. Lewis, J. G. (1986). Drug Discovery; the Evolution of Modern Medicines. *Postgraduate medical journal*, *62*(729), 704.

2. Sneader, W. (1990). The prehistory of psychotherapeutic agents. *Journal of Psychopharmacology*, *4*(3), 115-119.

3. Sneader, W. E. (2005). *Drug Discovery (The History)*. John Wiley & Sons, Inc.

4. Strenio Jr, A. J. (1996). The Aspirin Wars. *Journal of Public Policy & Marketing*, *15*(2), 319.

5. Augen, J. (2002). The evolving role of information technology in the drug discovery process. *Drug Discovery Today*, *7*(5), 315-323.

6. Warmuth, M. K., Liao, J., Rätsch, G., Mathieson, M., Putta, S., & Lemmen, C. (2003). Active learning with support vector machines in the drug discovery process. *Journal of Chemical Information and Computer Sciences*, *43*(2), 667-673.

7. Sams-Dodd, F. (2005). Target-based drug discovery: is something wrong?. *Drug discovery today*, *10*(2), 139-147.

8. Butcher, E. C. (2005). Can cell systems biology rescue drug discovery?. *Nature Reviews Drug Discovery*, *4*(6), 461-467.

9. Eddershaw, P. J., Beresford, A. P., & Bayliss, M. K. (2000). ADME/PK as part of a rational approach to drug discovery. *Drug Discovery Today*, *5*(9), 409-414.

10. Henderson, R., & Cockburn, I. (1996). Scale, scope, and spillovers: the determinants of research productivity in drug discovery. *The Rand journal of economics*, 32-59.

11. Terstappen, G. C., & Reggiani, A. (2001). In silico research in drug discovery.*Trends in pharmacological sciences*, *22*(1), 23-26.

12. Butcher, E. C., Berg, E. L., & Kunkel, E. J. (2004). Systems biology in drug discovery. *Nature biotechnology*, *22*(10), 1253-1259.

13. Kapetanovic, I. M. (2008). Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. *Chemico-biological interactions*, *171*(2), 165-176.

14. Stoermer, M. J. (2006). Current status of virtual screening as analysed by target class. *Med Chem* 2:89-112.

15. Sharma, H., Cheng, X., & Buolamwini, J. K. (2012). Homology model-guided 3D-QSAR studies of HIV-1 integrase inhibitors. *Journal of chemical information and modeling*, *52*(2), 515-544.

16. Edink, E., Akdemir, A., Jansen, C., van Elk, R., Zuiderveld, O., de Kanter, F. J., ... & de Esch, I. J. (2012). Structure-based design, synthesis and structure–activity relationships of dibenzosuberyl- and benzoate-substituted tropines as ligands for acetylcholine-binding protein. *Bioorganic & medicinal chemistry letters*, *22*(3), 1448-1454.

17. Nguyen, H. N., Bregman, H., Buchanan, J. L., Du, B., Feric, E., Huang, L., ... & DiMauro, E. F. (2012). Discovery and optimization of aminopyrimidinones as potent and state-dependent Nav1. 7 antagonists. *Bioorganic & medicinal chemistry letters*, *22*(2), 1055-1060.

18. Poulsen, A., Williams, M., Nagaraj, H. M., William, A. D., Wang, H., Soh, C. K., ... & Dymock, B. (2012). Structure-based optimization of morpholino-triazines as PI3K and mTOR inhibitors. *Bioorganic & medicinal chemistry letters*, *22*(2), 1009-1013.

19. Reader, J. C., Matthews, T. P., Klair, S., Cheung, K. M. J., Scanlon, J., Proisy, N., ... & Collins, I. (2011). Structure-guided evolution of potent and selective CHK1 inhibitors through scaffold morphing. *Journal of medicinal chemistry*, *54*(24), 8328-8342.

20. Becker, O. M., Shacham, S., Marantz, Y., & Noiman, S. (2003). Modeling the 3D structure of GPCRs: advances and application to drug discovery. *Current opinion in drug discovery & development*, *6*(3), 353-361.

21. Hillisch, A., Pineda, L. F., & Hilgenfeld, R. (2004). Utility of homology models in the drug discovery process. *Drug discovery today*, *9*(15), 659-669.

22. Phatak, S. S., Stephan, C. C., & Cavasotto, C. N. (2009). High-throughput and *in silico* screenings in drug discovery.

23. Cavasotto, C. N., & Phatak, S. S. (2009). Homology modeling in drug discovery: current trends and applications. *Drug discovery today*, *14*(13), 676-683.

24. Guha, R. (2008). On the interpretation and interpretability of quantitative structure–activity relationship models. *Journal of computer-aided molecular design*, *22*(12), 857-871.

25. Vedani, A., & Smiesko, M. (2009). In silico toxicology in drug discovery-concepts based on three-dimensional models. *Alternatives to laboratory animals: ATLA*, *37*(5), 477-496.

26. Reynès, C., Host, H., Camproux, A. C., Laconde, G., Leroux, F., Mazars, A., . & Sperandio, O. (2010). Designing focused chemical libraries enriched in protein-protein interaction inhibitors using machine-learning methods. *PLoS Comput Biol*, *6*(3), e1000695-e1000695.

27. Bacilieri, M., and S. Moro. (2006). Ligand-based drug design methodologies in drug discovery process: an overview. *Curr Drug Discov Technol* 3:155-165.

28. Guido, R. V., G. Oliva, and A. D. Andricopulo. (2008). Virtual screening and its integration with modern drug design technologies. *Curr Med Chem* 15:37-46.

29. Hattori, M., Tanaka, N., Kanehisa, M., & Goto, S. (2010). SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. *Nucleic acids research*, gkq367.

30. Wolber, G., Seidel, T., Bendix, F., & Langer, T. (2008). Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug discovery today*, *13*(1), 23-29.

31. Yang, S. Y. (2010). Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug discovery today*, *15*(11), 444-450.

32. Leach, A. R., Gillet, V. J., Lewis, R. A., & Taylor, R. (2010). Three-dimensional pharmacophore methods in drug discovery. *J. Med. Chem*, *53*(2), 539-558.

33. Sun, H., & Scott, D. O. (2010). Structure-based Drug Metabolism Predictions for Drug Design. *Chemical biology & drug design*, *75*(1), 3-17.

34. Ekins, S., Mestres, J., & Testa, B. (2007). In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British journal of pharmacology*, *152*(1), 9-20.

35. Klebe, G. (2006). Virtual ligand screening: strategies, perspectives and limitations. *Drug discovery today*, *11*(13), 580-594.

36. Leach, A. R., Shoichet, B. K., & Peishoff, C. E. (2006). Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *Journal of medicinal chemistry*, *49*(20), 5851-5855.

37. Lundstrom, K. (2007). Structural genomics and drug discovery. *Journal of cellular and molecular medicine*, *11*(2), 224-238.

38. Chandrika, B. R., Subramanian, J., & Sharma, S. D. (2009). Managing protein flexibility in docking and its applications. *Drug discovery today*, *14*(7), 394-400.

39. Anderson, A. C. (2003). The process of structure-based drug design. *Chemistry & biology*, *10*(9), 787-797.

40. Bajorath, J. (2001). Rational drug discovery revisited: interfacing experimental programs with bio- and chemo-informatics. *Drug discovery today*, *6*(19), 989-995.

41. Krieger, E., Nabuurs, S. B., & Vriend, G. (2003). Homology modeling. *Methods of biochemical analysis*, *44*, 509-524.

42. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., ... & Sali, A. (2006). Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics*, 5-6.

43. Fiser, A., & Šali, A. (2003). Modeller: generation and refinement of homology-based protein structure models. *Methods in enzymology*, *374*, 461-491.

44. Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Šali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual review of biophysics and biomolecular structure*, *29*(1), 291-325.

45. Šali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*, *234*(3), 779-815.

46. Bishop, A., De Beer, T. A., & Joubert, F. (2008). Protein homology modelling and its use in South Africa. *South African Journal of Science*, *104*(1-2), 2-6.

47. Molecular Operating Environment; MOE,(2015). 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.

48. Shen, M. Y., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein science*, *15*(11), 2507-2524.

49. Burley, S. K., Joachimiak, A., Montelione, G. T., & Wilson, I. A. (2008). Contributions to the NIH-NIGMS protein structure initiative from the PSI production centers. *Structure*, *16*(1), 5-11.

50. Ginalski, K. (2006). Comparative modeling for protein structure prediction.*Current opinion in structural biology*, *16*(2), 172-177.

51. Cavasotto, C. N., Orry, A. J., Murgolo, N. J., Czarniecki, M. F., Kocsi, S. A., Hawes, B. E., ... & Monsma Jr, F. J. (2008). Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *Journal of medicinal chemistry*, *51*(3), 581-588.

52. Park, H., Bahn, Y. J., Jung, S. K., Jeong, D. G., Lee, S. H., Seo, I., ... & Ryu, S. E. (2008). Discovery of novel Cdc25 phosphatase inhibitors with micromolar activity based on the structure-based virtual screening. *Journal of medicinal chemistry*, *51*(18), 5533-5541.

53. Park, H., Hwang, K. Y., Oh, K. H., Kim, Y. H., Lee, J. Y., & Kim, K. (2008). Discovery of novel α-glucosidase inhibitors based on the virtual screening with the homology-modeled protein structure. *Bioorganic & medicinal chemistry*,*16*(1), 284-292.

54. Cozza, G., Gianoncelli, A., Montopoli, M., Caparrotta, L., Venerando, A., Meggio, F., ... & Moro, S. (2008). Identification of novel protein kinase CK1 delta (CK1δ) inhibitors through structure-based virtual screening. *Bioorganic & medicinal chemistry letters*, *18*(20), 5672-5675.

55. Luther, K. B., Schindelin, H., & Haltiwanger, R. S. (2009). Structural and mechanistic insights into lunatic fringe from a kinetic analysis of enzyme mutants. *Journal of Biological Chemistry*, *284*(5), 3294-3305.

56. Gagnidze, K., Rozenfeld, R., Mezei, M., Zhou, M. M., & Devi, L. A. (2008). Homology modeling and site-directed mutagenesis to identify selective inhibitors of endothelin-converting enzyme-2. *Journal of medicinal chemistry*,*51*(12), 3378-3387.

57. Ballesteros, J. A., & Weinstein, H. (1995). [19] Integrated methods for the construction of three-

dimensional models and computational probing of structure-function relations in G protein-coupled receptors.

58. Mobarec, J. C., Sanchez, R., & Filizola, M. (2009). Modern homology modeling of G-protein coupled receptors: which structural template to use?. *Journal of medicinal chemistry*, *52*(16), 5207-5216.

59. Topiol, S., & Sabio, M. (2008). Use of the X-ray structure of the Beta2-adrenergic receptor for drug discovery. *Bioorganic & medicinal chemistry letters*, *18*(5), 1598-1602.

60. Sousa, S. F., Fernandes, P. A., & Ramos, M. J. (2006). Protein–ligand docking: current status and future challenges. *Proteins: Structure, Function, and Bioinformatics*, *65*(1), 15-26.

61. Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design*, *7*(2), 146.

62. Bissantz, C., Folkers, G., & Rognan, D. (2000). Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations.*Journal of medicinal chemistry*, *43*(25), 4759-4767.

63. Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2002). A review of protein-small molecule docking methods. *Journal of computer-aided molecular design*, *16*(3), 151-166.

64. Cole, J. C., Murray, C. W., Nissink, J. W. M., Taylor, R. D., & Taylor, R. (2005). Comparing protein–ligand docking programs is difficult. *Proteins: Structure, Function, and Bioinformatics*, *60*(3), 325-332.

65. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, *52*(4), 609-623.

66. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., ... & Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy.*Journal of medicinal chemistry*, *47*(7), 1739-1749.

67. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in

database screening. *Journal of medicinal chemistry*, *47*(7), 1750-1759.

68. Jones, G., Willett, P., & Glen, R. C. (1995). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of molecular biology*, *245*(1), 43-53.

69. Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology*, *267*(3), 727-748.

70. Verdonk, M. L., Chessari, G., Cole, J. C., Hartshorn, M. J., Murray, C. W., Nissink, J. W. M., ... & Taylor, R. (2005). Modeling water molecules in protein-ligand docking using GOLD. *Journal of medicinal chemistry*, *48*(20), 6504-6515.

71. Induced Fit Docking protocol 2013-2, Glide version 5.9, Prime version 3.2, Schrödinger, LLC

72. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., ... & Mainz, D. T. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of medicinal chemistry*, *49*(21), 6177-6196.

73. Moreira, I. S., Fernandes, P. A., & Ramos, M. J. (2010). Protein–protein docking dealing with the unknown. *Journal of computational chemistry*, *31*(2), 317-342.

74. Grosdidier, S., Totrov, M., & Fernández-Recio, J. (2009). Computer applications for prediction of protein–protein interactions and rational drug design. *Advances and applications in bioinformatics and chemistry: AABC*, *2*, 101.

75. Grosdidier, S., & Fernandez-Recio, J. (2012). Protein-protein docking and hot-spot prediction for drug discovery. *Current pharmaceutical design*, *18*(30), 4607-4618.

76. Ly, N. (2011). Evaluating Protein-Protein Docking Web Servers. *biochem218.* stanford.edu

77. Huang, S. Y. (2014). Search strategies and evaluation in protein–protein docking: principles, advances and challenges. *Drug discovery today*, *19*(8), 1081-1096.

78. Vajda, S., & Kozakov, D. (2009). Convergence and combination of methods in protein–protein docking. *Current opinion in structural biology*, *19*(2), 164-170.

79. Tovchigrechko, A., & Vakser, I. A. (2006). GRAMM-X public web server for protein–protein

docking. *Nucleic acids research*, *34*(suppl 2), W310-W314.

80. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., & Vakser, I. A. (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proceedings of the National Academy of Sciences*, *89*(6), 2195-2199.

81. Tovchigrechko, A., and Vakser, I.A. (2005). Development and testing of an automated approach to protein docking. *Proteins: Structure, Function, and Bioinformatics* 60, 296-301.

82. Xie, Q., and Tinker, R. (2006). Molecular Dynamics Simulations of Chemical Reactions for Use in Education, *Journal of Chemical Education*, **83**, 77-83.

83. Xie, C. Molecular Dynamics for Everyone: A Technical Introduc-tion to the Molecular Workbench Software.

84. Zhang, Z. H., & Linn, M. C. (2011). Can generating representations enhance learning with dynamic visualizations?. *Journal of Research in Science Teaching*, *48*(10), 1177-1198.

85. Ercolessi, F. (1997). A molecular dynamics primer. *Spring college in computational physics, ICTP, Trieste*, *19*.

86. Grubmüller, H., Heller, H., Windemuth, A., & Schulten, K. (1991). Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation*, *6*(1-3), 121-142.

87. Van Gunsteren, W. F., & Berendsen, H. J. C. (1977). Algorithms for macromolecular dynamics and constraint dynamics. *Molecular Physics*, *34*(5), 1311-1327.

88. Hess, B., Bekker, H., Berendsen, H. J., & Fraaije, J. G. (1997). LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry*,*18*(12), 1463-1472.

89. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., ... & Kollman, P. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, *91*(1), 1-41.

90. Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics

calculations. *J Comput Chem*, *4*, 187217.

91. MacKerell, A. D., Bashford, D., Bellott, M. L. D. R., Dunbrack, R. L., Evanseck, J. D., Field, M. J., ... & Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, *102*(18), 3586-3616.

92. MacKerell, A. D., Feig, M., & Brooks, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of computational chemistry*, *25*(11), 1400-1415.

93. Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, *118*(45), 11225-11236.

94. Lennard-Jones, J. E. (1931). Cohesion. *Proceedings of the Physical Society*, *43*(5), 461.

95. Mark, P., & Nilsson, L. (2001). Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A*, *105*(43), 9954-9960.

96. Lindahl, E., Hess, B., & Van Der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, *7*(8), 306-317.

97. Berendsen, H. J., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*(1), 43-56.

98. Berger, O., Edholm, O., & Jähnig, F. (1997). Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical journal*, *72*(5), 2002.

99. Berendsen, H. J., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. R. H. J., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, *81*(8), 3684-3690.

CHAPTER 3

Effect of template selection on the modeling of GPCR structures by homology modeling

3.0 Chapter summary

Analysis of the diverse available crystallographic structures of GPCRs suggests that although they share a common seven helix bundle, there are specific features that differentiate them that might be relevant for ligand design. This underlines the interest in solving novel GPCR crystallographic structures. However, despite the fact the number of crystallographic structures is steadily increasing, there are still challenges that hamper the availability of new ones. In the absence of a crystallographic structure for a specific receptor, homology modeling remains one of the most important techniques for constructing 3D models of proteins.

In this chapter, the results of an investigation devoted to understand the use of molecular dynamics as refinement process for models of GPCRs produced by homology modeling are described. Specifically, several atomistic models of rat M3 muscarinic receptor were constructed using diverse templates, including the human M2 muscarinic receptor, the human histamine 1 receptor and the bovine rhodopsin receptor using the MOE software. These homology models were refined with tiotropium docked in the orthosteric site by means of the GLIDE software. Furthermore, in order to determine the effect of ligand in the refinement process, two additional models of the M3 receptor were constructed and refined. The first had N-methyl scopolamine bound inside and the second did not include any ligand. The five models were embedded in a lipid bilayer prior to refinement process and molecular dynamics simulations were carried out using GROMACS. Finally, the resulting models were compared with the crystallographic structure of rat M3 muscarinic receptor.

The results show that molecular dynamics simulations of the protein embedded in a lipid bilayer improve the starting homology models and a sampling time of 500ns is adequate for the refinement process. Specifically, the refinement process can correct the length of the TM segment of the target receptor, the accuracy of the model greatly depends on the proximity of the template and the target in the phylogenetic tree and finally, although the presence of a ligand produces a faster equilibration of the system.

3.1 Introduction

Prediction of 3D structures of GPCRs is motivated by their importance in biological systems and the difficulties associated with their structure determination by experimental procedures. Knowledge of the 3D structure of GPCRs is important for understanding the molecular mechanisms underlying diseases and syndromes caused by mutations in these receptors, as well as for the structure-based design of small molecules acting as therapeutic treatments [1-3]. In fact, GPCRs account for the majority of best-selling drugs and about 40% of all prescription pharmaceuticals on the market [3-5].

Despite the number of crystallographic structures of GPCRs available has increased steadily during the last few years, there are still challenges that hamper the availability of new ones, including their low-expression yields, low receptor stability after detergent extraction from native membranes, and high conformational heterogeneity. With so little structural information available, homology modeling has become a widespread technique to get insights into the structure-function relationships of the receptors and facilitate the discovery of chemicals capable of modulating their activity [6-13]. Thus, homology modeling is used to construct atomic resolution models that could be used in virtual screening studies, to accurate predict ligand binding pocket, to study drug-receptor recognition as well as well as the clarification of experimental data [12-14]. However, reliable modeling of GPCR-ligand complexes for the vast majority of receptors with unknown structure remains to be one of the most challenging goals for the rational drug discovery process [15-16].

The publication of the crystallographic structure of rhodopsin at atomic resolution provided for the first time a template to construct realistic models of GPCRs by homology modeling [11, 17]. This structure has been widely used as template to construct models of diverse members of the rhodopsin-like family [11-12, 16, 18]. Although sequence identity with other GPCRs is small (about 20% on average) conservation of specific features among all the members of the rhodopsin-like family provides support for its use as template [11, 19]. In spite of the success, the suitability of the rhodopsin template to accurately predict the structure of other GPCRs was soon questioned [14, 20-21]. A critical point was made regards its low sequence identity with other GPCRs and also, the high diversity of ligand-binding

features observed among different GPCRs [20]. The publication of the crystallographic structures of diverse ligand-GPCR complexes a few years later, made these drawbacks more evident, as inferred from the results of the diverse community-wide GPCR-Dock assessment competitions. In these studies it was established a 35%–40% sequence identity between target and template as an empirical cutoff for reliable homology based prediction of ligand-receptor interactions [12 14].

Presently, homology modeling approaches of GPCRs achieve close-to-experimental accuracy for small rigid orthosteric ligands when templates with high sequence identity are used [15]. However, there still many issues that need to be addressed to improve the quality of the models of GPCRs constructed by homology modeling. Diverse reports published recently address issues like template selection [22-24]; inclusion of knowledge-based constraints into the modeling process [24-25]; modeling of the extracellular loops [26-27] and on structure refinement [10, 28-30].

In this chapter, both the effect of the template selection and the use of molecular dynamics simulations for the refinement of GPCRs models constructed by homology modeling were investigated. Specifically, the use of templates located at diverse distances from target receptor in the GPCR phylogenetic tree; the use of MD for the refinement process as well as the effect of including a ligand bound onto the orthosteric site. To reach the aforementioned objective, a "semi-blind" homology modeling study of the rat M3 muscarinic receptor whose crystallographic structure is available, was carried out using the human M2 muscarinic receptor with and without an antagonist bound to it, the human H1 histamine receptor and bovine rhodopsin as templates. The study is semi-blind since the crystallographic structure of the target is available, hence provides an evaluation opportunity for both the final models and applied methods used in the construction of the models.

3.2 Methodology

Several computational chemistry methodologies are applied in the constructions of atomistic

3D models of GPCR. The notion that amongst homologues, protein structures are more conserved than protein sequences is the widely accepted underlying principle when constructing an unknown atomic-resolution model of the "target" protein from its amino acid sequence and a template an experimental 3D structure of a related homologous protein [6, 9-10, 14-16].

While the overall architecture between the target and template GPCR are conserved, for example the 7TM motif, there are number of factors that need to addressed and corrected or refined in order to get a final model. The noted differences includes, the length and orientation of the TM helices, orientation of sidechains, differences in the extracellular and intracellular loops , the presence and extent of helical distortions (kinks and bulges). Thus, a good refinement model is needed to correct the noted differences and help annotated the expected structure and features of the final model [14-16].

3.2.1 Constructing GPCR homology model by homology modeling

Homology models of the rat muscarinic M3 receptor were constructed from diverse templates including the human muscarinic M2 receptor [31], the human H1 histamine receptor [32], and bovine rhodopsin [17] using comparative modeling by the Molecular operating Environment (MOE) software [33]. The sequence of the rat M3 muscarinic receptor was retrieved from the Uniprot database (ID: P08484), whereas those of the templates were retrieved from the corresponding crystallographic structures (PDB entry code: 3UON for the human muscarinic M2, 3RZE for the human histamine H1 and 1GZM for bovine rhodopsin). Residues of the T4 Lysozyme were removed from the respective templates and sequence alignment was done between the target and the templates using MOE-Align tool.

All muscarinic receptors subtypes contain a long ICL3, where by only the N and C terminal portion are essential and of needed coupling specificity of G-protein. A large portion of ICL3 can be deleted without interfering with the general functions of the receptor [34]. Thus, muscarinic receptors need just the N terminus and C-terminus residues of the ICL3 for general receptor functions.

Accordingly, in order to simplify the construction of the atomistic model of the rat M3 receptor without compromising its accuracy, a major part of the intracellular loop (ICL3) was not included in the modeling, using instead a stretch of three residues at each end of helices TM5 and TM6, respectively that were subsequently linked. Thus, the crystallographic structure of the M2 muscarinic receptor had over 90% of the ICL3 residues removed and accordingly, based on comparative sequence alignment these residues were removed from the target residues and realignment was commenced. For rhodopsin and histamine based models a single comparative alignment was performed with the target ICL3 residues exclusion done solely based on those not aligning with template N and C terminals of the ICL3. Once the realignment was completed and checked for inspection of unwanted gaps in the TM regions, homology models were constructed.

Models generated were validated visualizing the bonds or clashes in between atoms in the models and by computing the values of the backbone torsional angles and plotting them in a Ramachandran map to show the distribution of angles observed in a single structure.

3.2.2 Molecular docking

Once the structure validation process was completed, the highest ranking starting models for each of the three templates was selected for refinement. In a step forward two ligands (Figure 3.1), the inverse agonist Tiotropium that was also co-crystalized with the published M3 structure [35], and the non-selective antagonist N-methyl scopolamine (NMS) were docked onto the orthosteric site of the starting models and guided by the structural information available on GPCRs. The docking process was carried out using the Glide software from the Schrodinger programs [36-37]. The resulted poses were rank ordered by binding and docking score, poses were rank ordered using the XP score function of GLIDE [38]. Final poses of each of the compounds was selected based on their ranking and similarity to the pose previously reported [18, 35]. This permitted continuation to the next step, molecular dynamics simulation to refine the "rough" homology models.

Tiotropium

N-methyl scopolamine (NMS)
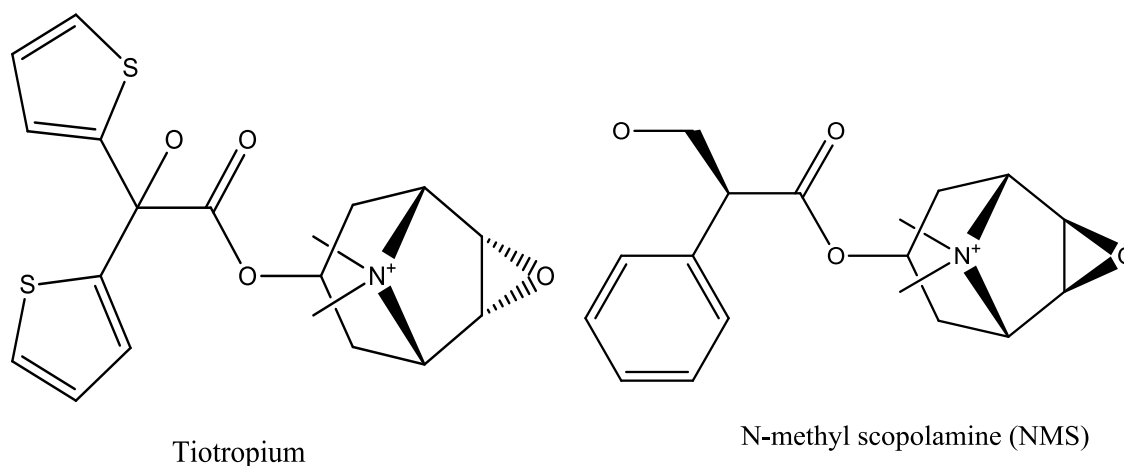
Figure 3.1: The structures of the Tiotropium and N-methyl scopolamine (NMS), the ligands used in this study

Henceforth, four protein-ligand complexes of M3 muscarinic receptor, the first based on M2 muscarinic receptor and bound to Tiotropium, second also constructed using M2 receptor as a template with NMS bound to it, and the last two complexes based on the human histamine 1 receptor (bound with NMS) and bovine rhodopsin receptor (bound with NMS) as templates were refined using molecular dynamic simulations.

3.2.3 Model refinement

Models were refined using molecular dynamics (MD) simulations. For each of the models, the starting structure was embedded into a box consisting of equilibrated 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids and water molecules and equilibrated according to the procedure described previously elsewhere [39-40]. The box had an initial size of 10.3 x 8.0 x 10.2 nm^3 (XYZ), organized in such a way that the bilayer plane was oriented on the XY plane. Before protein insertion, the box contained 256 lipids (corresponding to an area per lipid of 0.64 nm^2) and circa 17,000 water molecules. The protein was placed in the center of the box, and the overlapping molecules were

removed. Specifically, all water molecules with oxygen atoms closer than 0.40 nm to a non-hydrogen atom of the protein, as well as all lipid molecules with at least one atom closer than 0.25 nm to a non-hydrogen atom of the protein were removed. This resulted in a final system containing 188 lipids and circa 14,655 water molecules. Removal of these atoms introduced small voids between the protein and water or lipid molecules that disappeared during the first part of the MD simulation, in which a progressive adjustment of the lipid bilayer and water molecules to the protein takes place.

Next, 105 randomly selected water molecules were replaced by 45 sodium and 60 chloride ions, providing a neutral system with a concentration approximately 0.2 M on sodium chloride. This concentration is fairly similar to that found in biological organisms, although they exhibit different intra- and extra-cellular ion concentrations. Both the protein, ligand, ions and TIP3P water model parameters were describes using OPLSAA [41], while sampling was carried out for 500ns using GROMACS package 4.6 [42-43]. It should be noted that the same box (size and features) were used for each of the five models refinement. While the number of the water and lipid molecules displaced by the addition of the protein in the box were slight different depending on the size of the protein, it bears no effect of the final concentration of ions. Hence the parameters and condition employed in the refinement process were similar to all models, barring their inherent differences i.e. protein size, type of ligand included etc.

3.3 Results and discussion

As detailed in the previous section, the procedure to construct atomistic models of GPCRs by homology modeling involves the selection of a template; sequence alignment between target and

template sequences; construction of a starting model and its verification; a refinement process. And as prior stated the goal of this thesis chapter study is to investigate the performance of molecular dynamics simulations when used in the refinement process. For this purpose, the relevance of template selection, ligand inclusion, as well as the length of the simulation on the quality of the models of GPCRs constructed were investigated.

In the present study atomistic models of M3 receptor were constructed by homology modeling using diverse templates and compared with the crystallographic structure of the rat muscarinic M3 receptor. Templates selected for the present study include the human muscarinic M2, the human histamine H1 more distant and bovine rhodopsin as a very distant template. The selection of a template in this study was based on their proximity to the target M3 receptor in the GPCRs phylogenetic tree. The effect of the ligand in the refinement process was also investigated. Therefore molecular dynamic simulations were conducted using the M2 muscarinic as template with Tiotropium or NMS docked in the orthosteric site and the results obtained were compared with the simulation of M2 muscarinic template based modeled without ligand inclusion.

In order to understand the performance of molecular dynamics in the refinement process the starting structure of each model must be considered. Starting models were constructed by using the sequence of the target receptor, arranging them on the corresponding positions to the sequence of crystallographic structure of the template as determined by the prior sequence alignment. In fact, starting models present different distances to the target M3 crystallographic structure that parallel to those of the corresponding starting crystallographic structures. Thus, the muscarinic M2 receptor exhibits a root-mean-square deviation (*rmsd*) of all the alpha carbons of 1.6 Å, whereas this value rises to 1.9 Å for the histamine H1 receptor and to 2.3 Å for rhodopsin and are in accordance to their distances in the phylogenetic tree. These differences will be also shown in the respective starting models and consequently the models constructed will have to undertake changes of the same order to provide an accurate model of the target receptor.

Pairwise sequence alignments were carried out as described in the methods section. Sequence

alignment of the rat muscarinic M3 receptor with the human muscarinic M2, the human histamine H1, and bovine rhodopsin show sequence identities between of 60%, 34% and 17%, respectively. These values increase to 79%, 41% and 23% when only the transmembrane (TM) regions are considered. Interestingly, the best score of sequence identity is found between TM3 and the lowest between TM1 in all the cases (Figure 3.2).

As mentioned above the intracellular loop ICL3 is very long and had to be removed for modeling purposes. This deletion bears no consequences to the overall structure of the GPCR or the orthosteric binding pocket. In the case of the M2 muscarinic receptor based models this was easily facilitated by using the alignment of the sequence of target and template and removal of the aligned residues missing in the sequence of the human M2 crystallographic structure. Since both the human histamine H1 and rhodopsin receptors have relatively short ICL3, the removal was simply done on the unaligned sequences between sequences of templates TM5 and TM6. This had a major effect on the length and in some extent the conformation of both TM4 and TM5 of histamine H1 and rhodopsin based models.

Sequence alignments were used to construct the models of the rat M3 muscarinic receptor using the diverse templates selected for the present study. For each of the templates alternative models were generated using template backbone coordinates with alternative side chain conformations using an extended rotamer library implemented in MOE. Following this process a number of independent models based on loop and side chain placements scored by a contact energy function can be generated. From these a model is selected for refinement based on the score and the least number of steric violations.

```
4DAJ_A|R.norvegicus|CHRM3    ----DPLGGHTIWQVVFIAFLTGFLALVTIIGNILVIVAFKVNKQLKTVN
3UON_A|H.sapiens|CHRM2       ---LALTSPYKTFEVVFIVLVAGSLSLVTIIGNILVMVSIKVNRHLQTVN
3RZE_A|Human                 -------TTMASPQLMPLVVVLSTICLVTVGLNLLVLYAVRSERKLHTVG      ━━  TM1
1GZM_A|Bovine                PFEAPQYYLAEPWQFSMLAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPL
                                  :.   :.      : ::  .  *:*.:      .::*:*      ━━  TM2

4DAJ_A|R.norvegicus|CHRM3    NYFLLSLACADLIIGVISMNLFTTYIIMNRWALGNLACDLWLSIDYVASN      ━━  TM3
3UON_A|H.sapiens|CHRM2       NYFLFSLACADLIIGVFSMNLYTLYTVIGYWPLGPVVCDLWLALDYVVSN
3RZE_A|Human                 NLYIVSLSVADLIVGAVVMPMNILYLLMSKWSLGRPLCLFWLSMDYVAST
1GZM_A|Bovine                NYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLGGE
                             *  :..*: ***::    :          : . : :*   *:   :  : .

4DAJ_A|R.norvegicus|CHRM3    ASVMNLLVISFDRYFSITRPLTYRAKRTTKRAGVMIGLAWVISFVLWAPA
3UON_A|H.sapiens|CHRM2       ASVMNLLIISFDRYFCVTKPLTYPVKRTTKMAGMMIAAAWVLSFILWAPA
3RZE_A|Human                 ASIFSVFILCIDRYRSVQQPLRYLKYRTKTRASATILGAWFLSFLWVIPI
1GZM_A|Bovine                IALWSLVVLAIERYVVVCKPMSNFRFG-ENHAIMGVAFTWVMALACAAPP
                                :: .:.::.::**  : :*:       . *   :  :*.::   *

4DAJ_A|R.norvegicus|CHRM3    ILFWQYFVGKRTVP----PGECFIQFLSEPTITFGTAIAAFYMPVTIMTI
3UON_A|H.sapiens|CHRM2       ILFWQFIVGVRTVE----DGECYIQFFSNAAVTFGTAIAAFYLPVIIMTV
3RZE_A|Human                 LGWNHFMQQTSVRR----EDKCETDFYDVTWFKVMTAIINFYLPTLLMLW
1GZM_A|Bovine                LVGWSRYIPEGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFF
                             :                 :  .   ..  :   *  :*  ::

4DAJ_A|R.norvegicus|CHRM3    LYWRIYKETEKMLIK----------EKKAAQTLSAILLAFIITWTPYN-I
3UON_A|H.sapiens|CHRM2       LYWHISRASKSRIPPP-------SREKKVTRTILAILLAFIITWAPYN-V
3RZE_A|Human                 FYAKIYKAVRQHCLHM-------NRERKAAKQLGFIMAAFILCWIPYF-I
1GZM_A|Bovine                CYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYAGV
                             *  ::      ..             *::.:: :   ::  **:: *  **   :

4DAJ_A|R.norvegicus|CHRM3    MVLVNTFCDSCIPKTYWNLGYWLCYINSTVNPVCYALCNKTFRTTFKTLL
3UON_A|H.sapiens|CHRM2       MVLINTFCAPCIPNTVWTIGYWLCYINSTINPACYALCNATFKKTFKHLL
3RZE_A|Human                 FFMVIAFCKNCCNEHLHMFTIWLGYINSTLNPLIYPLCNENFKKTFKRIL
1GZM_A|Bovine                AFYIFTHQGSDFGPIFMTIPAFFAKTSAVYNPVIYIMMNKQFRNCMVTTL
                             . : :.        :  ::    .:. **  *  : *  *:. :     *

4DAJ_A|R.norvegicus|CHRM3    LCQCDKRKRRKQQYQQRQSVIFHKRVPEQAL
3UON_A|H.sapiens|CHRM2       MCHYKNIGATR--------------------
3RZE_A|Human                 HIRSGENLYFQ--------------------
1GZM_A|Bovine                CCGKNPLGDDEASTTVSKTETSQVAPA----
```

Figure 3.2: Sequence alignment of target and the templates used annotating TM3 (in orange) and TM1 (in green) as area with highest and lowest similarity, respectively.

3.3.1 The length of the refinement process

The refinement process aims to relax the structure of the target protein in an environment close to the native one. This was carried out in the present work through the use of MD simulations of the protein embedded in a lipid environment. The final model of the target protein is produced as the energy minimized structure obtained from the average of a set of configurations sampled during the

production run. Accordingly, in order for the procedure to provide reliable results simulation times need to be long enough to have a reasonable sampling after equilibration [39].

Monitoring of the refinement process can be carried out by inspection of the time evolution of the *rmsd* in regard to the starting structure. As can be seen in Figure 3.3 for both the whole protein (a) as well as the transmembrane regions (b), the equilibration process takes around 250-300 ns, except for the model constructed from the bovine rhodopsin that equilibration is achieved short after 200ns. More specifically, rmsd time evolution of the whole protein (a) or the transmembrane region (b) follow the same trends although values of the latter are smaller, suggesting that the loops provide an important contribution to the rmsd. Furthermore, as can be seen the models of M3 constructed using the M2 muscarinic receptor exhibit a rmsd of around 2.3 Å in regard to the starting structure, smaller than the model constructed from the histamine H1 receptor of around 2.7 Å and lower than the rmsd observed with the model constructed from rhodopsin that reaches around 3.0 Å. These results correlate well to the rmsd of the templates suggesting that rhodopsin and histamine suffer larger structural changes than the M2 muscarinic to get transformed into the M3 muscarinic receptor as can be expected from the distance of the different templates to the modelled receptor in the phylogenetic tree.

(a)

(b)
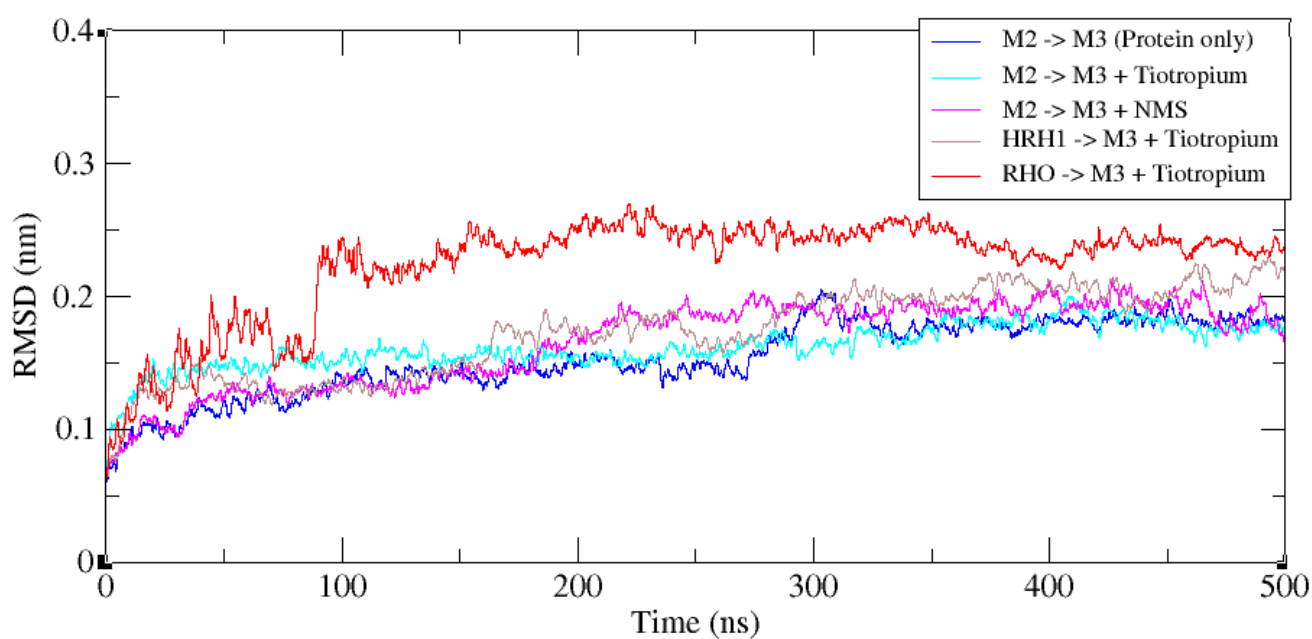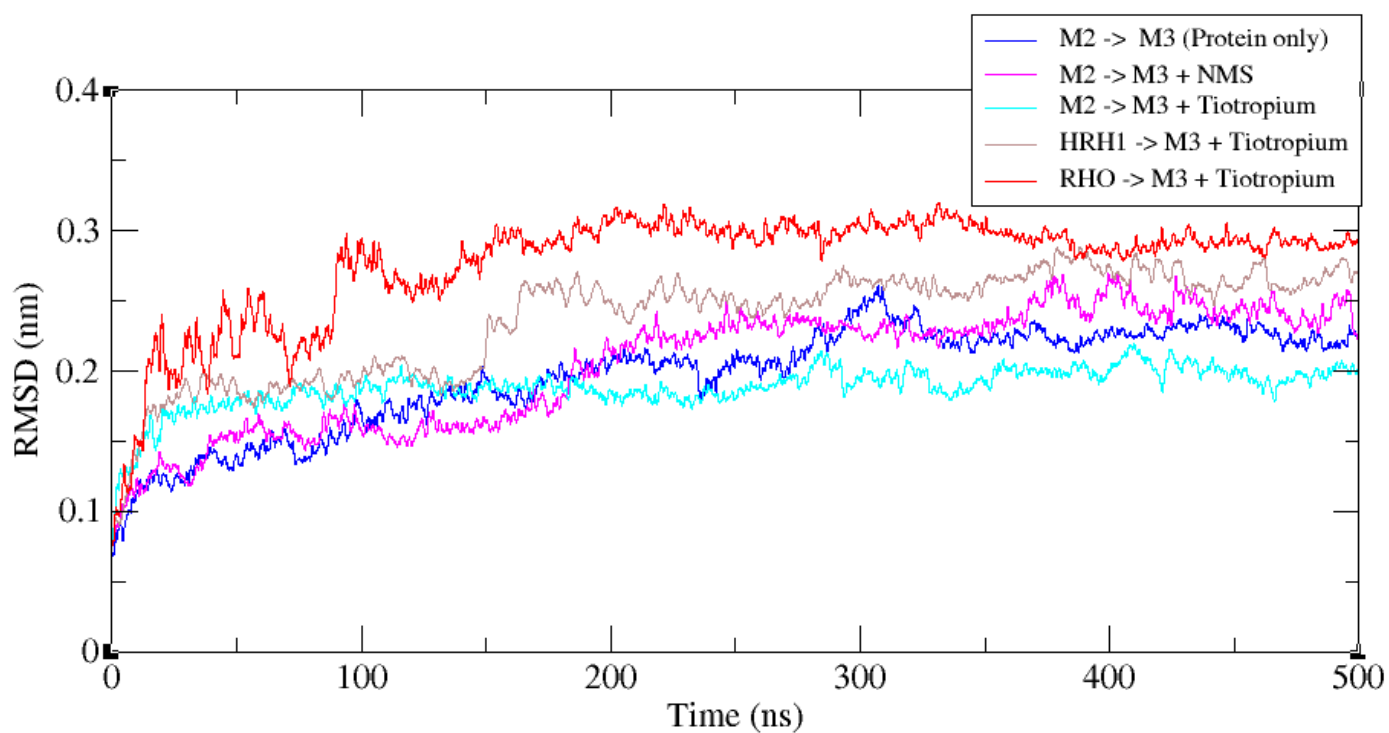
Figures 3.3: The time evolution of the *rmsd* considering the alpha carbons of the whole protein **(a)** and when considering only the alpha carbons of the TM bundle residues **(b)**, for each of the model simulations.

3.3.2 The average structures

Analysis of the trajectories permitted the best conditions to construct a working model of the target receptor. From Figure 3.3 the time evolution of the *rmsd* for each of the models constructed, exhibits fluctuations up to approximately halfway of the trajectory and then begin to be stable. Therefore, for each trajectory of the model refinement process at least 200ns after stabilization of the simulation can be considered for analysis. Accordingly, working models for further analysis were selected using the average structure of the last 50 ns of the trajectory.

The working models or refined model structures presented here are simply minimized average structure of the simulated "rough" homology model. That is to say, after running molecular simulation of a system containing the rough receptor /receptor complex, an average structure of a chosen receptor trajectory is calculated and then minimized *in vacuo*. An average structure is obtained after the root mean square fluctuations (*rmsf*) of atomic positions in the trajectory are calculated then-after fit to a reference initial frame (in this case the beginning of a chosen stable trajectory timeframe). The *rmsf* values of the average structures are then converted and saved as average coordinate B-factor values.

Analysis of the (*rmsf*) of each of the residues along the refinement process is shown in Figures 3.4. The Figure shows the fluctuations of the models along the last 50 ns. Inspection of Figure shows that the highest fluctuation are in the loops region both intracellular and extracellular, as expected. Specifically, ECL1 and ICL3 loops in rhodopsin based model, ICL2 in the M2 based model that was refined with NMS, ECL2 of the human histamine base model as well as ICL3 of all the models shows the largest fluctuations.
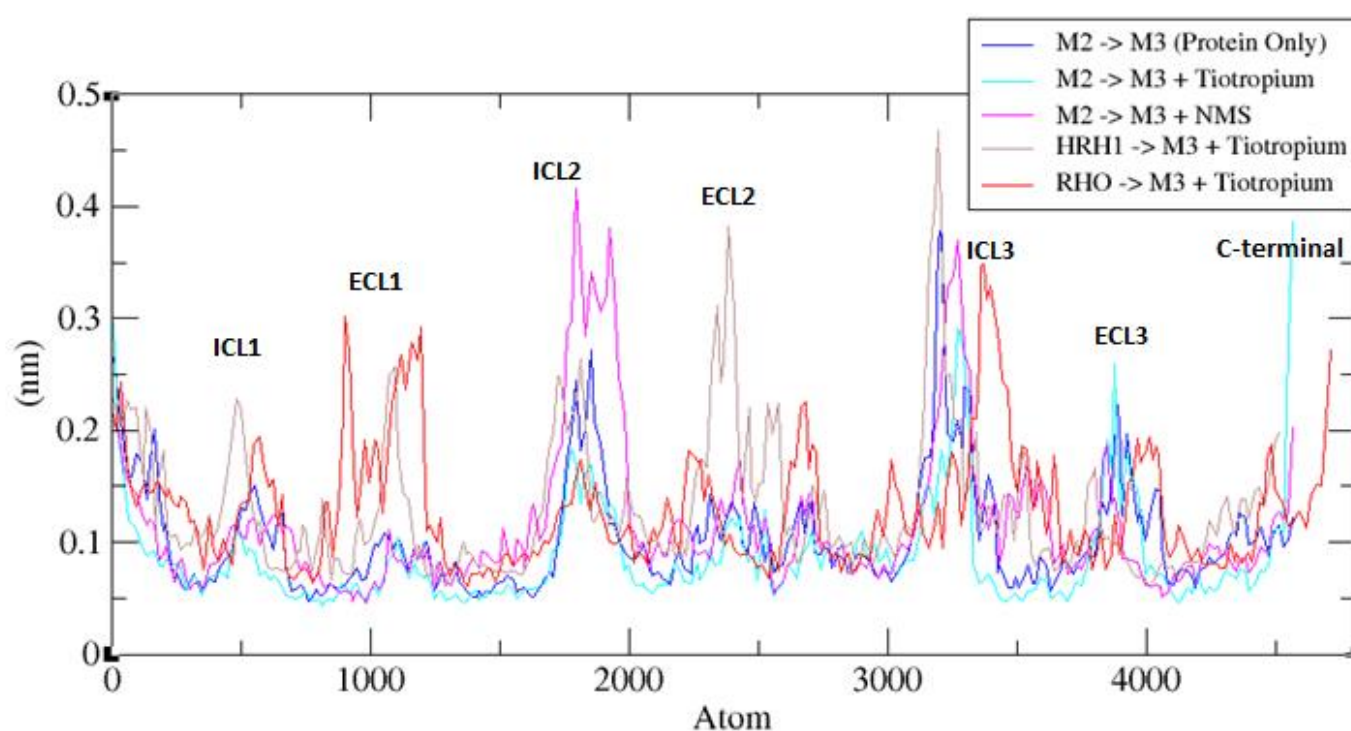
Figures 3.4: The *rmsf* of the refined models during the last 50ns of the molecular simulations.

3.4 Effect of template selection

The selection of a suitable template is of paramount importance in homology modeling. While a correct sequence alignment is key and a good refinement process can help to correct mistakes of the modeling process, our results show that the choice of the template is important for model accuracy.

First, lets analyze the differences in the length of the loops and TM region between the constructed rough models, refined models (average structure) and the crystallographic structure. Table 3.1 lists the residues involved in each of the TM regions for the homology models, refined models (in green) and the crystallographic structure of M3 muscarinic receptor.

Table 3.1: Comparison of residues in TMs for the homology models, refined models (in green) and the crystallographic

structure of M3 muscarinic receptor.

TM region	M3 Crystallographic structure	M3$_{M2}$	M3$_{M2}$ NMS	M3$_{M2}$ with Tiotropium	M3$_{HRH1}$ with Tiotropium	M3$_{RHO}$ with Tiotropium
TM1	W66 - V95	W66-V95	W66-V95	W66-V95	F70-K92	I64-F93
		W66-V95	W66-V95	W66-V95	F70-F93	V65-F93
TM2	V102 - M131	V102-M131	V102-M131	V102-M131	V102-I130	I103-I130
		V102-M131	V102-M131	V102-M131	V102-I130	T101-Y128
TM3	L139 – T171	G137-Y176	G137-Y176	G137-Y176	G137-Y176	N138-I170
		N138-I170	N138-T171	N138-T171	A140-T171	L139-Y167
TM4	T182 - V211	T182-V211	T182-V211	T182 - V211	T182-A203	T182-L205
		T182-V211	T182-V211	T182 - V211	K183-Q208	R180-W200
TM5	P229 – K256	P229-K256	P229-K256	P229-K256	P229-E257	E228-K260
		P229-K256	P229-K256	P229-K256	P229-E257	P229-E259
TM6	L492 – T514	K488-T515	K488-T515	K488-T515	I484-C517	A489-T514
		A490-T515	Q491-T515	Q491-T515	K488-C517	I484-T514
TM7	K522 – C546	K522-C546	K522-C546	K522-C546	K522-L546	K522-C546
		K522-C546	K522-C546	K522-C546	K522-L546	K522-C546

Regarding the differences in the TM length between the rough and refined models, it is very encouraging that the refinement process is capable to correct the length of the TM segments. For example in the starting model TM3 is five residues longer in the rough models based M2 muscarinic receptor than in crystallographic structure of M3 receptor. However, during the refinement process the length of the helices from the M2 based model is shortened, ending only one residue shorter than the crystallographic structure. In histamine receptor based model, the same TM3 ending is also five residue longer and completely corrected to be equal with M3 crystallographic structure after refinement process. The correction is afforded albeit not completely but the effort presented is significant enough to recommend the methodology for other GPCR homology modeling.

While the length of the intracellular and extracellular loops can be gauged from the above table 3.1, the shape and orientation adopted by those loops, especially the extracellular loops (Figure 3.5) is of significant importance in homology modeling especially when the constructed models is to be used in the study ligand allosterism or in designing ligand of interest.
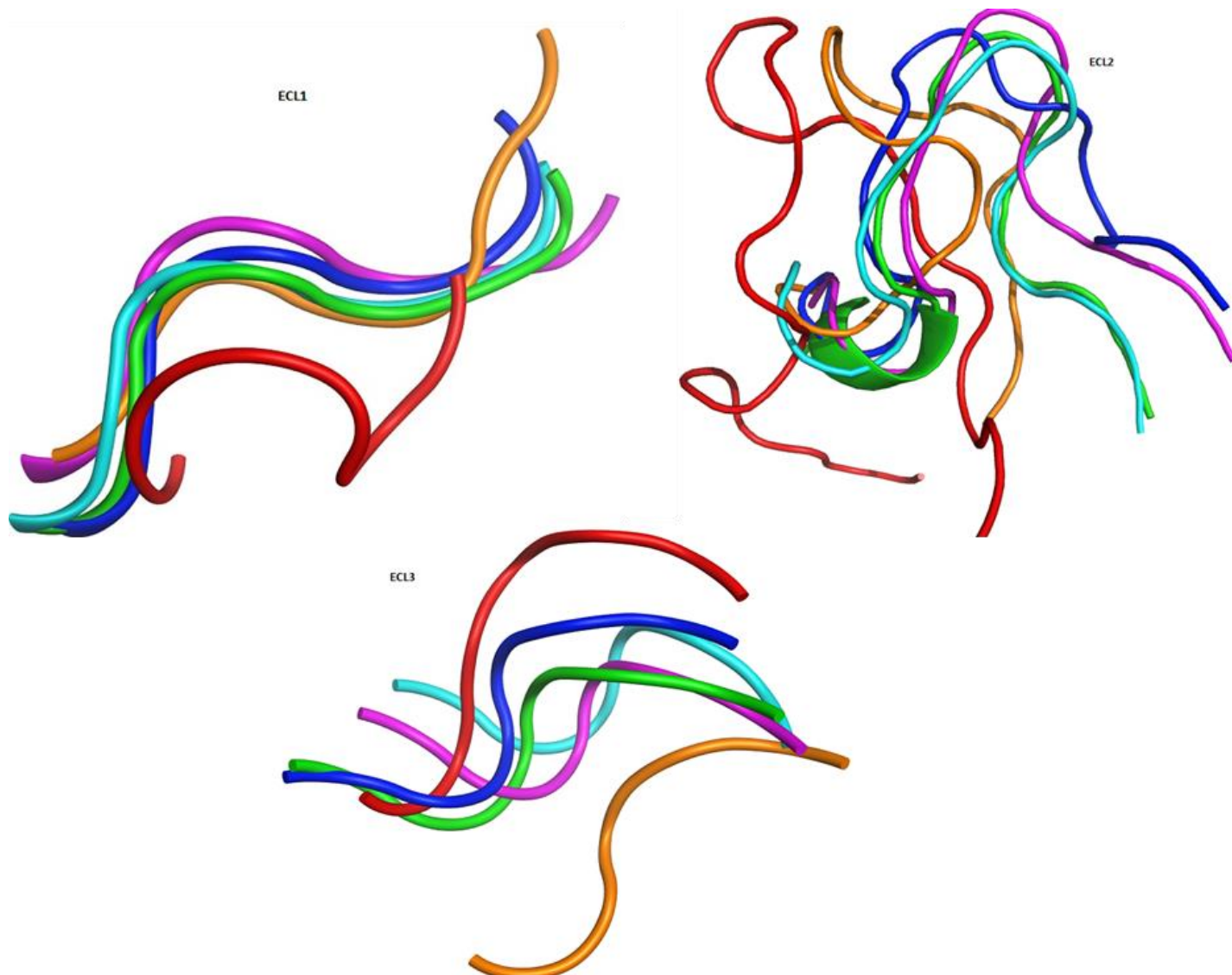
Figure 3.5: Comparison between the loops M3 crystallographic structure (green) and the refined M3 models made from M2 refined without ligand (blue), M2 refined with ligand NMS (magenta), M2 refined with ligand Tiotropium (cyan), histamine receptor H1 (brown) and Rhodopsin receptor (Red).

As shown is Figure 3.5, the extracellular loop regions in the constructed models are very

different from those of the M3 crystallographic structure. For example in rhodopsin based model ECL2

is so different in shape and morphology, length and the occupying space. While the same ECL2 in those

modeled from M2 muscarinic as template occupies the same space, adopt a somewhat similar shape

and structural features, with the highest similarity observed in the model refined with ligand

Tiotropium bound while that without ligand showing the higher deviation from the ECL2 of the

crystallographic structure. Table 3.2 shows the dihedral angles for residues in the ECL region of both

the crystallographic structure in comparison to those of the refined average structures.

Table 3.2: Dihedral angle (in degrees) for the ECL regions of M3 crystallographic structure Vs refined average structures (ECL1 residue # 131-136, ECL2 residue # 214-222 and ECL3 residue # 515-520)

Dihedral	Dihedral residues	Crystallographic structure	M2->M3	M2->M3 + NMS	M2->M3 + Tiotropium	Histamine-> M3 + Tiotropium	Rhodopsin-> M3 + Tiotropium
ECL1-1	N-R-W-A	-107.1	-89.137	-100.208	-113.438	-57.709	97.597
ECL1-2	R-W-A-L	174.5	136.118	138.438	104.12	-14.663	2.466
ECL1-3	W-A-L-G	76.0	77.527	27.671	78.351	-122.438	4.853
ECL2-1	T-V- P-P	-155.3	-112.204	-107.778	-120.051	-109.049	-121.127
ECL2-2	V- P-P-G	-86.3	-107.187	-101.638	-105.511	-79.735	-152.381
ECL2-3	P-P-G-E	55.0	148.892	7.041	43.588	31.452	-77.230
ECL2-4	P-G-E-C	47.3	34.782	83.542	64.672	95.754	-86.521
ECL2-5	G-E-C-F	-149.3	-139.267	-34.945	-153.601	-94.472	107.164
ECL2-6	E-C-F-I	123.6	138.713	-156.9	131.849	-66.969	60.848
ECL3-1	F-C-D-S	-140.7	-1.271	-139.917	-142.147	47.916	15.743
ECL3-2	C-D-S-C	33.9	46.808	3.294	0.656	48.326	4.000
ECL3-3	D-S-C-I	105.3	77.83	97.969	114.569	50.473	-115.992

Clearly, the resulting ECL are also impacted by the choice of the template used, and how the

model was refined. However, the *rmsd* of the between the alpha carbons of the constructed models

and the crystallographic structure in general is tied to this. In particular explaining the loops that seem

to occupy different space as such in the case of ECL2 of rhodopsin based and ECL3 of the histamine

based constructed model.

Second, comparison of the distance between the rough model, the refined model and the target

crystallographic M3 structure shows interesting results. Tables 3.3-3.7 shows the rmsd of the whole structures (a) and the TM region (b) for all the models constructed, comparing the template selected, the constructed rough model, 3 different types of refined model and the crystallographic structure of M3 Muscarinic receptor (the target). The 3 different types of refined average structure depicted here are calculated from using different dielectric constant (ratio of the permittivity of a substance to the permittivity of free space) during the process of *in vacuo* minimization for the average structure.

Table 3.3: The comparative *rmsd* between the whole structures **(a)** and only TM bundle when M2 muscarinic receptor was used as template and refined without ligand. D= Dielectric constant during *in vacuo* minimization.

(a)

All residues	1	2	3	4	5	6
1. M2 crystallographic	0.00	1.559	0.531	2.047	1.905	1.885
2. M3 crystallographic	1.559	0.00	1.590	1.906	1.900	1.947
3. Rough M3 model	0.531	1.590	0.00	2.068	1.988	1.985
4. Average D=1	2.047	1.906	2,068	0.00	0.797	1.008
5. Average D=2	1.905	1.900	1.988	0.797	0.00	0.621
6. Average D=4	1.885	1.947	1.985	1.008	0.621	0.00

(b)

TM residues	1	2	3	4	5	6
1. M2 crystallographic	0.00	0.946	0.473	1.801	1.728	1.762
2. M3 crystallographic	0.946	0.00	1.164	1.904	1.819	1.861
3. Rough M3 model	0.473	1.164	0.00	2.050	1.978	1.980
4. Average D=1	1.801	1.904	2.068	0.00	0.795	1.004
5. Average D=2	1.728	1.819	1.978	0.795	0.00	0.620
6. Average D=4	1.762	1.861	1.980	1.004	0.620	0.00

Table 3.4: The comparative *rmsd* between the whole structures **(a)** and only TM bundle when M2 muscarinic receptor was used as template and refined with ligand NMS. D= Dielectric constant during *in vacuo* minimization.

(a)

All residues	1	2	3	4	5	6
1. M2 crystallographic	0.00	1.559	0.531	1.911	2.072	2.075
2. M3 crystallographic	1.559	0.00	1.590	1.997	1.798	2.113
3. Rough M3 model	0.531	1.590	0.00	1.840	2.074	2.157
4. Average D=1	1.911	1.997	1.840	0.00	1.165	1.201
5. Average D=2	2.068	1.798	2.074	1.165	0.00	0.951
6. Average D=4	2.075	2.113	2.157	1.201	0.951	0.00

(b)

TM residues	1	2	3	4	5	6
1. M2 crystallographic	0.00	0.946	0.471	1.701	1.756	1.834
2. M3 crystallographic	0.946	0.00	1.158	1.765	1.696	1.902
3. Rough M3 model	0.471	1.158	0.00	1.836	2.070	2.152
4. Average D=1	1.701	1.765	1.836	0.00	1.160	1.200
5. Average D=2	1.756	1.696	2.070	1.160	0.00	0.950
6. Average D=4	1.834	1.902	2.152	1.200	0.950	0.00

Table 3.5: The comparative *rmsd* between the whole structures **(a)** and only TM bundle when M2 muscarinic receptor was used as template and refined with ligand Tiotropium. D= Dielectric constant during *in vacuo* minimization.

(a)

All residues	1	2	3	4	5	6
1. M2 crystallographic	0.00	1.559	0.531	2.198	2.171	2.228
2. M3 crystallographic	1.559	0.00	1.590	2.001	1.742	1.944
3. Rough M3 model	0.531	1.590	0.00	2.174	2.061	2.122
4. Average D=1	2.198	2.001	2.174	0.00	0.882	1.230
5. Average D=2	2.171	1.742	2.061	0.882	0.00	0.861
6. Average D=4	2.228	1.944	2.122	1.230	0.861	0.00

(b)

TM residues	1	2	3	4	5	6
1. M2 crystallographic	0.00	0.946	0.531	2.057	1.986	2.054
2. M3 crystallographic	0.946	0.00	1.157	1.940	1.641	1.853
3. Rough M3 model	0.531	1.157	0.00	2.170	2.055	2.118
4. Average D=1	2.057	1.940	2.170	0.00	0.876	1.224
5. Average D=2	1.986	1.641	2.055	0.876	0.00	0.859
6. Average D=4	2.054	1.853	2.118	1.224	0.859	0.00

Table 3.6: The comparative *rmsd* between the whole structures **(a)** and only TM bundle when human histamine receptor was used as template and refined with ligand Tiotropium. D= Dielectric constant during *in vacuo* minimization.

(a)

All residues	1	2	3	4	5	6
1. HIS crystallographic	0.00		1.885	0.344	2.622	2.046	2.132
2. M3 crystallographic	1.885	0.00	2.121	2.556	2.369	2.453
3. Rough M3his model	0.344	2.121	0.00	2.654	2.045	2.179
4. Average D=1	2.622	2.556	2.654	0.00	1.113	1.281
5. Average D=2	2.046	2.369	2.045	1.113	0.00	0.845
6. Average D=4	2.132	2.453	2.179	1.281	0.845	0.00

(b)

TM residues	1	2	3	4	5	6
1. HIS crystallographic	0.00		1.527	0.331	2.375	1.996	2.086
2. M3 crystallographic	1.527	0.00	1.801	2.552	2.368	2.439
3. Rough M3 model	0.331	1.801	0.00	2.650	2.041	2.173
4. Average D=1	2.375	2.552	2.650	0.00	1.108	1.275
5. Average D=2	1.996	2.368	2.041	1.108	0.00	0.835
6. Average D=4	2.086	2.439	2.173	1.275	0.835	0.00

Table 3.7: The comparative *rmsd* between the whole structures **(a)** and only TM bundle when bovine rhodopsin receptor was used as template and refined with ligand Tiotropium. D= Dielectric constant during *in vacuo* minimization.

(a)

All residues	1	2	3	4	5	6
1. RHO crystallographic	0.00	2.318	0.744	3.173	3.371	3.277
2. M3 crystallographic	2.318	0.00	2.308	3.707	3.183	3.002
3. Rough M3 model	0.744	2.308	0.00	2.906	2.927	2.982
4. Average D=1	3.173	3.707	2.906	0.00	0.986	1.257
5. Average D=2	3.371	3.183	2.927	0.986	0.00	1.007
6. Average D=4	3.277	3.002	2.982	1.257	1.007	0.00

(b)

| TM residues	1	2	3	4	5	6
1. RHO crystallographic	0.00	2.297	0.734	2.701	3.063	2.927
2. M3 crystallographic	2.297	0.00	2.239	3.666	3.090	2.912
3. Rough M3 model	0.734	2.239	0.00	2.890	2.910	2.970
4. Average D=1	2.701	3.666	2.890	0.00	0.980	1.249
5. Average D=2	3.063	3.090	2.910	0.980	0.00	1.000
6. Average D=4	2.927	2.912	2.970	1.249	1.000	0.00

While in some cases the rough model seemed to have much better rmsd when compared with the crystal structure, it should be noted that the rough model are not refined (may contain structural kinks) and there are trade off afforded by the refinement process such as correction of the length of TM segments of the constructed models and relaxing the structure kinks, and refining the binding pocket. Clearly, the dielectric constant used in the minimization of the receptors plays a significant role. Implicit methods for modeling protein electrostatics require dielectric properties of the system to be known, in particular, the value of the dielectric constant of protein. However, the protein dielectric constant is not a constant but is a complex function reflecting the properties of the protein

structure and sequence [44]. By comparing the *rmsd* value obtained when dielectric of 1, 2 or 4 was used, the *rmsd* values obtained when dielectric constant of 2 was used are much closer to the *rmsd* M3 crystallographic structure than 1 or 4. Hence all average structures (the refined models) presented here were obtained after *in vacuo* minimization at a dielectric constant of 2.

3.5 Analysis of the refinement process

In order to understand the role played by molecular dynamics as refinement method of the initial models generated by homology modeling, the evolution of the distance (*rmsd* of the C^α) between the different models and the crystallographic structure of the target along the refinement process was analyzed. Figure 3.6 (a) shows the evolution during the refinement process of the residues of the TM region for the different M3 models generated in the present work and the M3 crystallographic structure and similarly, Figure 3.6 (b) shows the rmsd evolution considering the whole protein. As can be seen the Figures clear stress the marked differences of the starting points that correlates well with the distance of the difference templates in the phylogenetic tree. The most interesting result is the negative slope exhibited by the rmsd vs time, suggesting that the refinement process improves moderately the quality of the models. Improvements measured on the a rmsd are between 0.5-0.8 Å. In contrast, no improvement is observed when all the atoms of the protein are considered, probably due to the fluctuations of the loops along the trajectory.

Interestingly, the differences observed in these Figures differ slightly from those of Table 3.7 because there comparison is made with the average structure and because of the TM regions used here for comparison are those as defined in the rough model at the beginning, while in Table 3.7 was calculated using TM as defined in refined model.
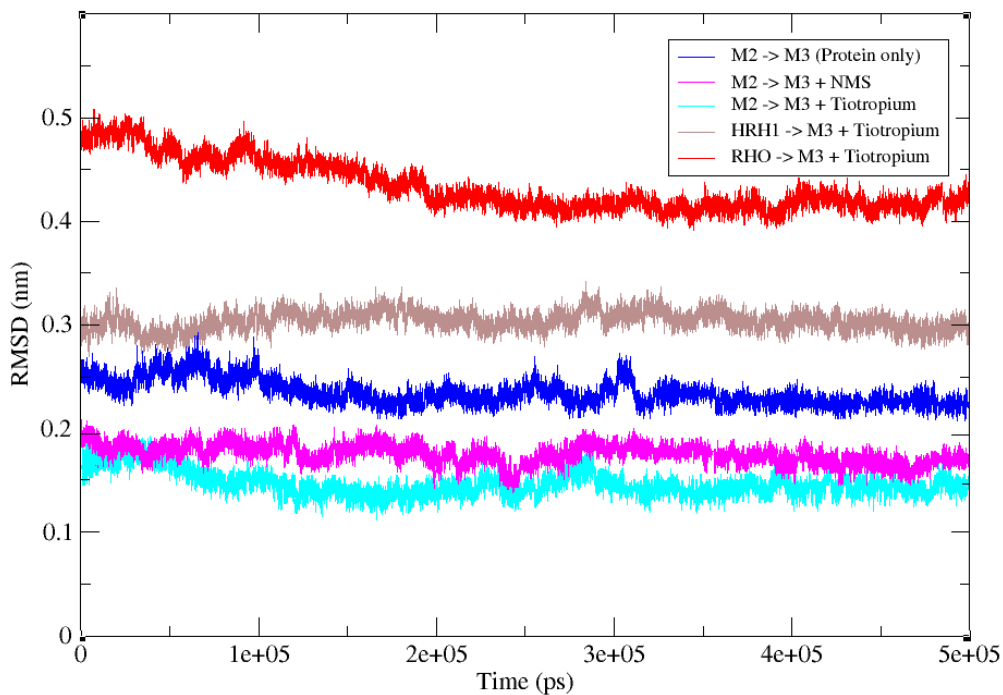
(a)
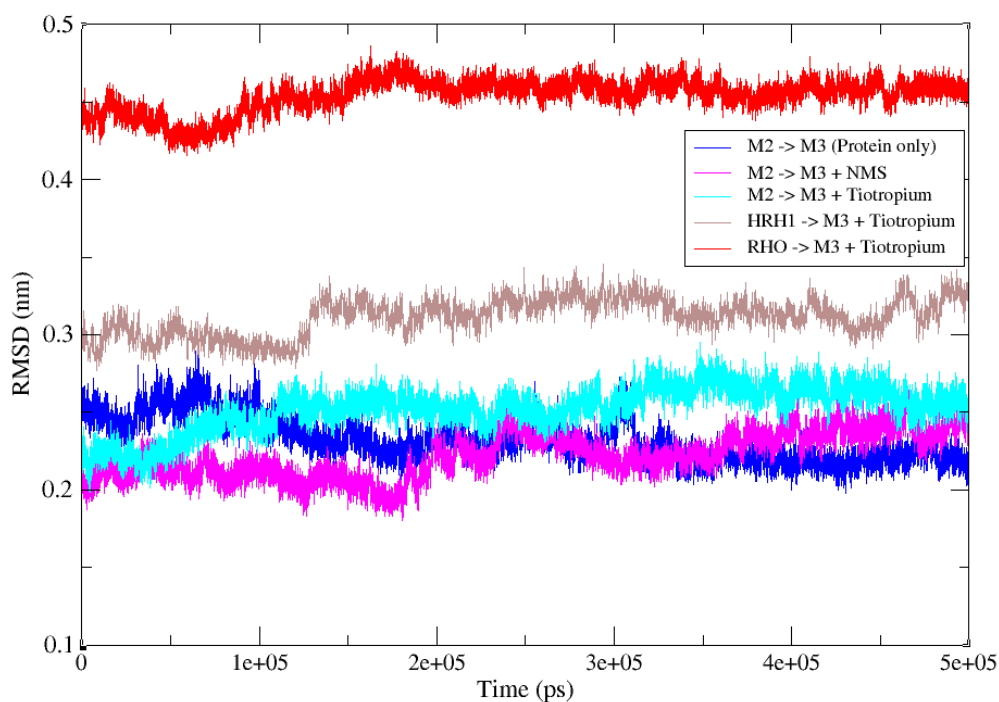
(b)

Figure 3.6. (a) *rmsd* evolution along the trajectory of the distance between the TM regions of the different M3 models computed in this work and the M3 crystallographic structure computed using the alpha carbons. (b) same as (a) considering the whole protein.

## 3.6 The effect of ligand addition in simulation process

The addition of a ligand in the refinement process provides significant advantages to the constructed atomistic models during the refinement process. For example, as shown in Figure 3.4 (a), the convergence of models with a ligand bound is achieved relatively faster when ligand is included in the refinement process. Notably, it took 25 ns for M2 based model receptor refined with Tiotropium bound system and 50ns for that containing NMS to equilibrate, while about 300ns were required for the M2 based receptor model that was refined without any ligand to equilibrate. Interestingly, shorter equilibration times were observed for Histamine H1 and Rhodopsin based models since they were refined with a ligand inside. Which further prove the addition of a ligand in the refinement process makes the process more robust since convergence is observed at earlier times in models refined with ligand bound in the system rather than those without ligand. It should also be taken into account that much shorter times can be observed if only the helical section of the receptors is considered Figure 3.3 (a).

These results show a comparative and that has also experimental credence to the stability that addition of a ligand offers to the protein system. For example, several studies including the T4L fusion protein crystallography strategy, and other experimental studies have provided with evidence that supports the idea that ligand (both agonist and antagonists) binding stabilize a receptor. Also an increasing number of ligand bound crystallographic structure of GPCR give credence to stabilization properties of ligands [45-49].
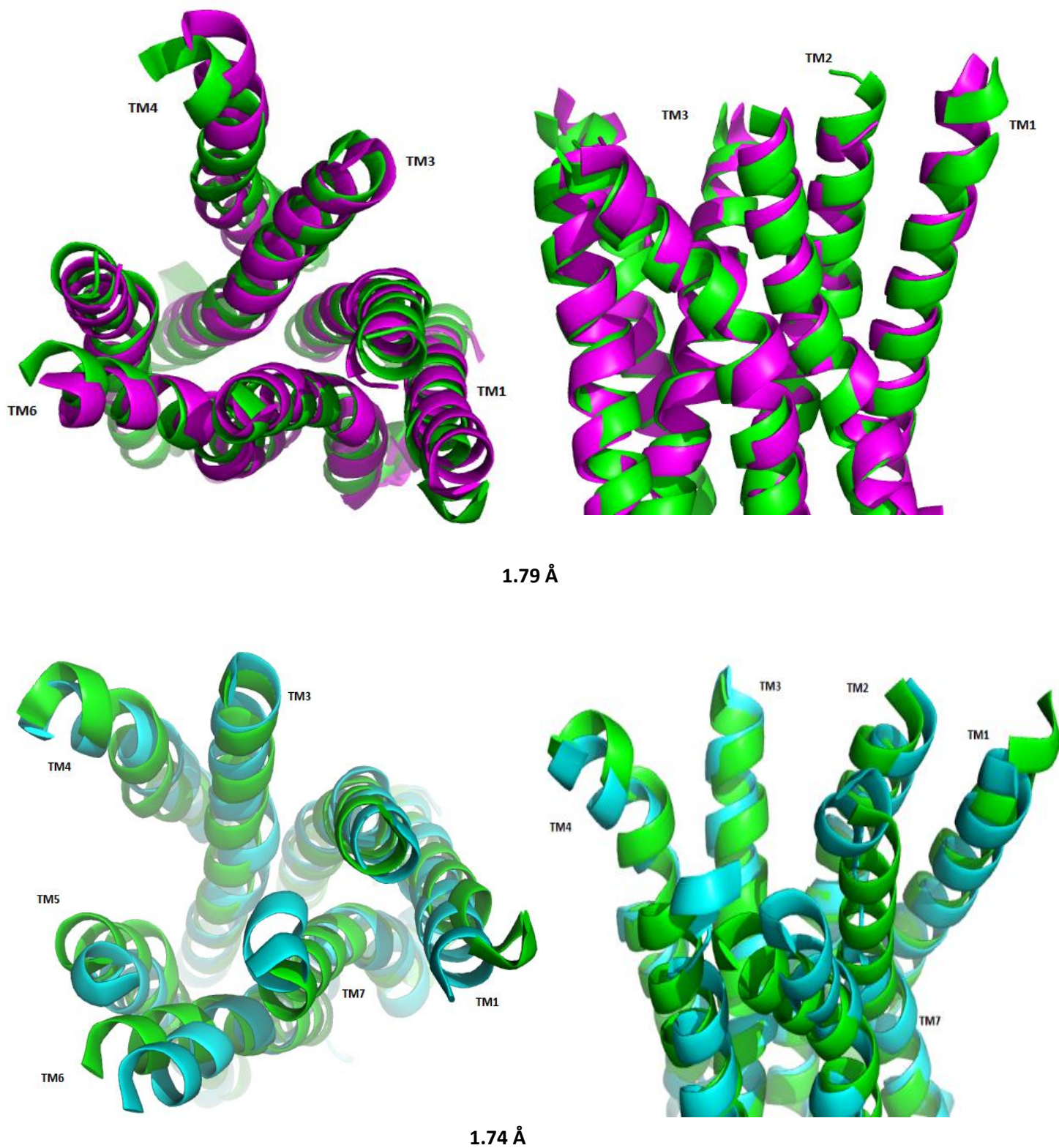
3.7 Comparison between the constructed models and the M3 muscarinic crystallographic structure

3.7.1 Crystallographic structure Vs Modeled structure (M2 muscarinic receptor template)

As explained above, three different models that use the M2 muscarinic receptor as template were constructed and refined. The addition of the ligand in the refinement process does not only have effect on the time it takes the receptor to equilibrate, it also affects the structure features of the final refined model. Looking from the extracellular side the orthosteric binding site appears more compact (Figure 3.7: a-c) as in the crystallographic structure while in rough models it is more significant vacuous (data not shown but can be gauged from the difference in the rmsd between rough and refined models (table 3.3-3.5).Furthermore, the outward bend seen at the extracellular end of TM4 is present in both the refined models and the crystallographic structure of the M3 receptor.

1.90 Å

1.79 Å

1.74 Å
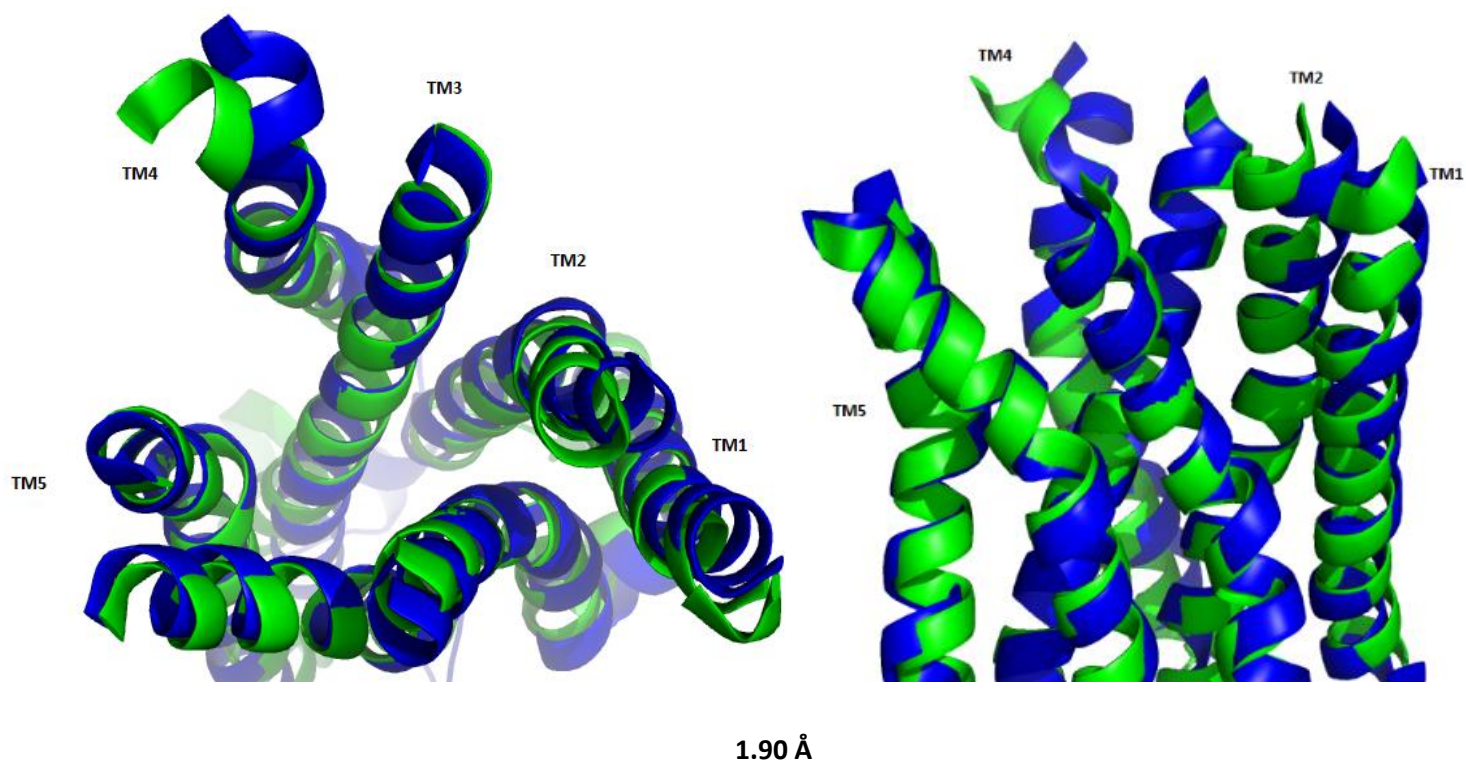
Figure 3.7: Comparison between the crystallographic structure of M3 muscarinic receptor and refined models constructed using M2 receptor refined without ligand **(a)**, with ligand NMS **(b)** and with ligand Tiotropium **(c).** Their comparative *rmsd* is also noted.

Another difference between the crystallographic structure and the models constructed from the M2 muscarinic receptor as a template is observed in the length of the TM regions, specifically in the TM3 and TM6. The refinement process attempted to correct the length and was successful in TM3 for the models refined ligand and for that refined without a ligand the differences was residues to one residues (Table 3.1). For TM6, the refinement process was able to reduce the difference in length from four to two residues in the structure refined without ligand and four to one residue in both of the structure refined with a ligand.

3.7.2 Crystallographic structure Vs Modeled structure (human histamine 1 and bovine rhodopsin receptor as template)

Comparison of the models generated form human H1 and bovine rhodopsin point to the absence of the pronounced outward bend at the extracellular end of TM4 found in the crystallographic structures of the M2 and M3 muscarinic receptors (Figure 3.8 (a)and (b)). In the model of M3 constructed using the histamine H1 receptor with bound Tiotropium yields a ECL3 extracellular loop that does not fit well with the structure of the target (Figure 3.5).

In regard to the length of the TM segments the refinement process reduces slightly the differences found in the starting structure (Table 3.1), but differences are still remarkable. For example for the model constructed using as template the human histamine 1 receptor, TM1 is seven residues shorter; TM2 and TM3 are one residue shorter; TM4 four residues shorter catching up part the eight residues shorter of the starting model; TM5 is remains one residue longer; TM6 is seven residues longer correcting in part the eleven residues of the starting structure and TM7 is corrected to the correct TM length.

The differences in the intracellular end of TM5 and thusly the ICL3 where lysozyme is attached in the crystallographic structure of histamine 1 receptor can be attributed to the resulting differences in length for TM3. This also affects or is affected by the residues in the target ICL3 that had to be significantly reduced. As previously mentioned ICL3s of muscarinic receptor subtypes are long and can

be reduced without affecting the overall function of the receptor.

In the model of M3 muscarinic receptor constructed using the crystallographic structure of bovine rhodopsin receptor and refined with Tiotropium bound, the extracellular loop ECL2 does not fit well with the structure of the target ECL2 it appears to occupy a slight different space. In regard to the length of the TM segments the refinement process reduces slightly the differences found in the starting structure (table 3.1), but differences are here are the most highest recorded. With the template being also the furthest from the target in the phylogenetic tree, and having the lowest sequence similarity between template and target, all the noted can be thusly attributed. To annotate, TM1 is one residue shorter; TM2 is two residues; TM3 is four residues shorter; TM4 is reduced to eight residues shorter by the refinement process; TM5 is three residues longer and TM6 is seven residues shorter while TM7 is perfectly corrected by the refinement process.

(a)

2.36 Å

(b)

3.66 Å
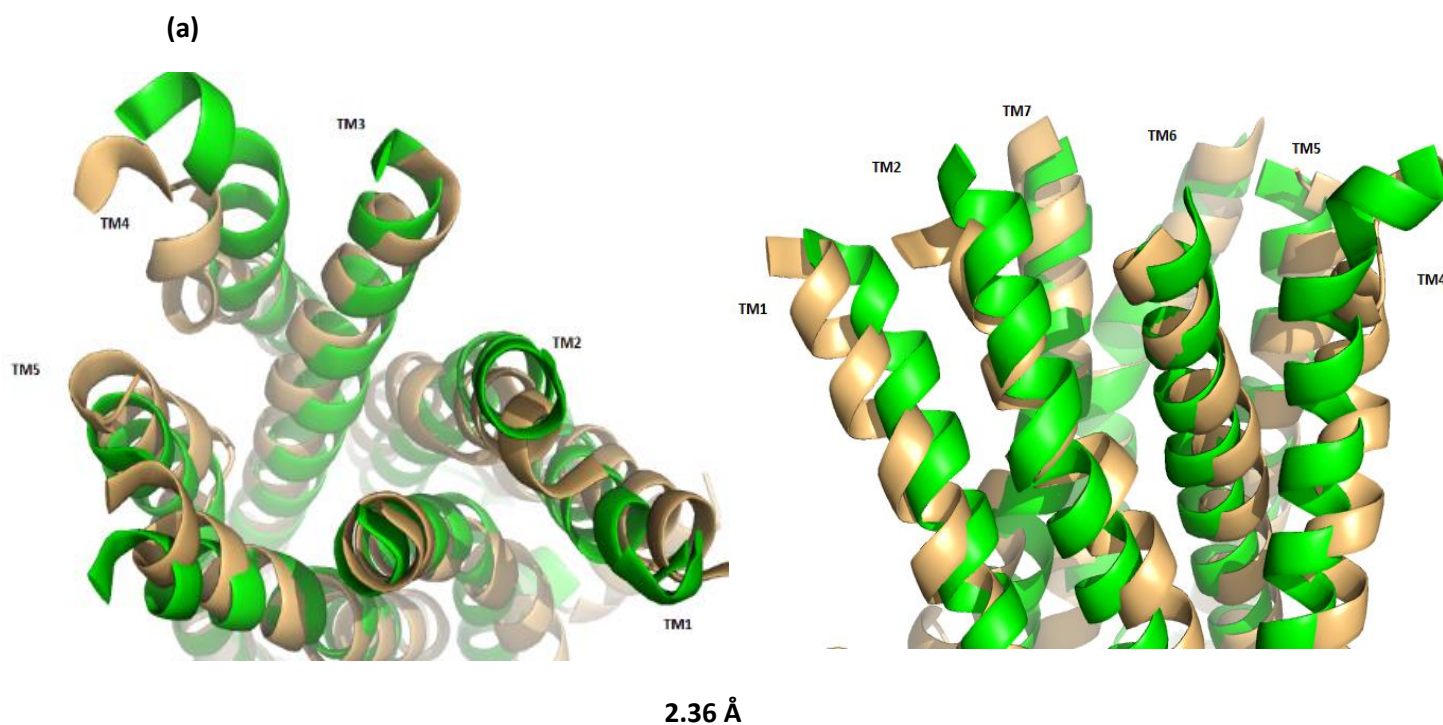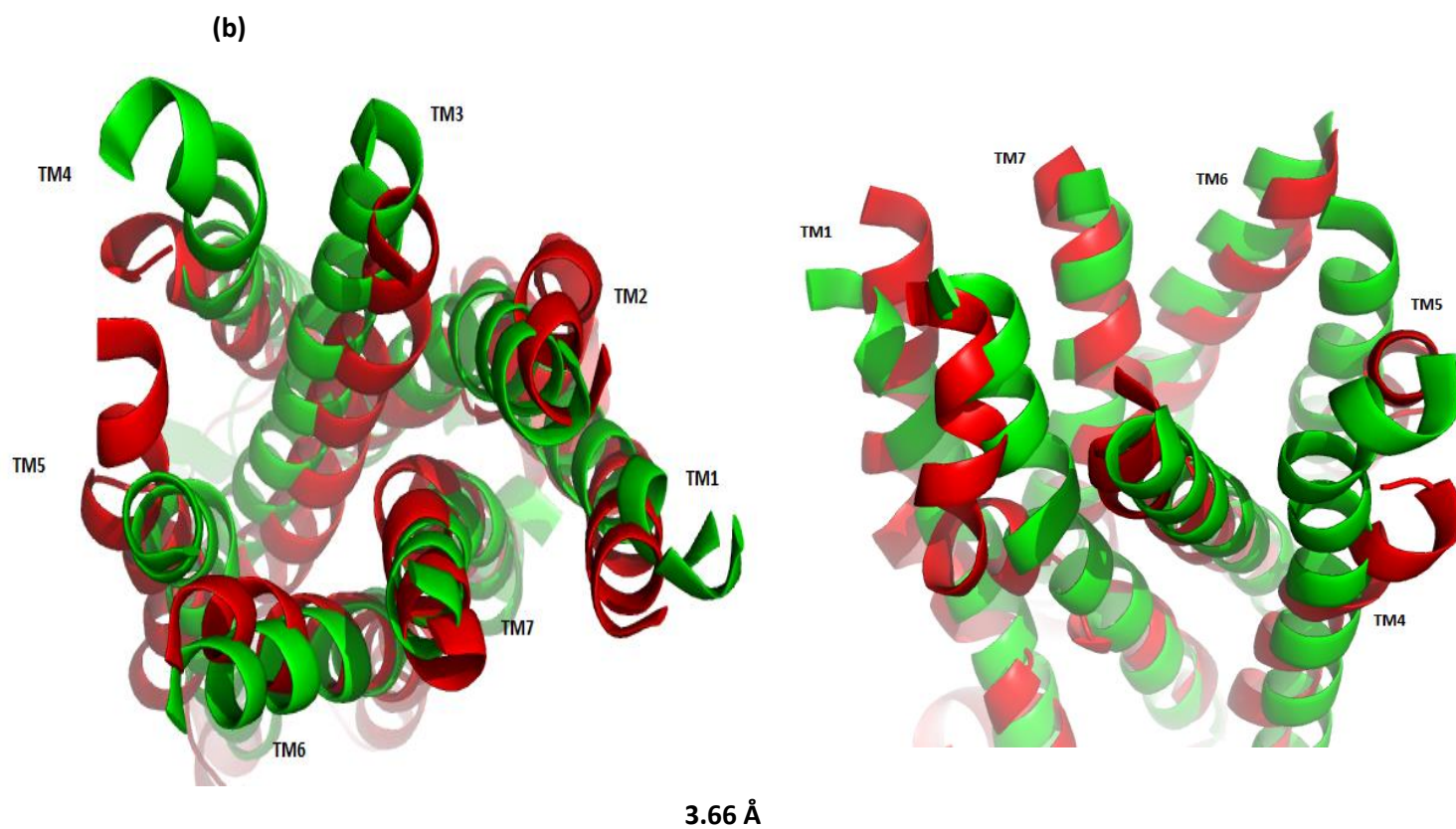
Figure 3.8: Comparison between the crystallographic structure of M3 muscarinic receptor and refined models constructed using human histamine 1 receptor and refined with Tiotropium **(a)**, and that constructed using bovine rhodopsin receptor and refined with Tiotropium **(b).** Their comparative *rmsd* is also noted.

3.8 Conclusions to chapter 3

In validating both the constructed models and the methodology used to construct them, both the choice of the template as well as the length and manner (with ligand Vs without ligand) were observed. Clearly, the simulation time of 500 ns is an adequate simulation time to evaluate and make conclusion on the objectives. The refinement process is also able to correct the length of the TM helices from the starting rough models.

While the constructed atomistic models capture most of the features of the M3 receptor, the *rmsd* and general structure conformation observes here shows that the higher the sequence homology between the target and the template the better the results. By comparing phylogenetically distant (bovine rhodopsin receptor), to the closest evolutionary templates (M2 muscarinic receptor), a clear effect of evolutionary distance between the templates and our target receptor was observed. Hence, the higher the sequence homology between the target and the template, the more accurate the models constructed and the accuracy correlates with the distance between the template and the target receptors in the phylogenetic tree.

Moreover, inclusion of a ligand on the modeling makes the refinement process more robust since the convergence is observed at earlier times in model refined with ligand bound rather than in those refined without ligand. This approach is more suitable in such occasions where time and resources are limited in modeling target receptors. Unfortunately this cannot overcome much the hurdle of choosing unsuitable receptor template. Analysis of both the *rmsd*, the *rmsf* and the equilibration times during the trajectory in the model refinement process support this conclusion. At 17% and 34 % sequence homology with our target refinement with ligand did not improve the quality of the resulting model, it did however have effect on equilibration time. While this in itself is not a deterrent for general virtual screening purpose, for those receptor-ligand whose some of the interaction are mediated by residues in the loops, more attention must be paid in designing loops separately for example chimeric, in order to elevate the quality of the results.

However, since the addition of a ligand affect the length of the sampling time and the template selection the structural quality of the resulting models, all these factors needs to be preferably applied upon when constructing homology models.

3.9 References to chapter 3

1. Insel, P. A., Tang, C. M., Hahntow, I., & Michel, M. C. (2007). Impact of GPCRs in clinical medicine: monogenic diseases, genetic variants and drug targets. *Biochimica et Biophysica Acta (BBA)- Biomembranes*, *1768*(4), 994-1005.

2. Hopkins, A. L., & Groom, C. R. (2002). The druggable genome. *Nature reviews Drug discovery*, *1*(9), 727-730.

3. Filmore, D. (2004). Cell-based screening assays and structural studies are fueling g-protein coupled receptors as one of the most popular classes of investigational drug targets. *Modern Drug Discovery*, *7*(11).

4. Jacobson, K. A., & Costanzi, S. (2012). New insights for drug design from the X-ray crystallographic structures of G-protein-coupled receptors. *Molecular pharmacology*, *82*(3), 361-371.

5. Mason, J. S., Bortolato, A., Congreve, M., & Marshall, F. H. (2012). New insights from structural biology into the druggability of G protein-coupled receptors. *Trends in pharmacological sciences*, *33*(5), 249-260.

6. Costanzi, S. (2010). Modeling G protein-coupled receptors: a concrete possibility. *Chimica oggi*, *28*(3), 26.

7. Carlsson, J., Coleman, R. G., Setola, V., Irwin, J. J., Fan, H., Schlessinger, A., ... & Shoichet, B. K. (2011). Ligand discovery from a dopamine D3 receptor homology model and crystal structure. *Nature chemical biology*, *7*(11), 769-778.

8. Bissantz, C., Bernard, P., Hibert, M., & Rognan, D. (2003). Protein-based virtual screening of chemical databases. II. Are homology models of g-protein coupled receptors suitable targets?. *Proteins: Structure, Function, and Bioinformatics*, *50*(1), 5-25.

9. Costanzi, S. (2012). Homology modeling of class a g protein-coupled receptors. In *Homology Modeling* (pp. 259-279). Humana Press.

10. Koehler Leman, J., Ulmschneider, M. B., & Gray, J. J. (2015). Computational modeling of membrane proteins. *Proteins: Structure, Function, and Bioinformatics*, *83*(1), 1-24.

11. Costanzi, S., Siegel, J., Tikhonova, I. G., & Jacobson, K. A. (2009). Rhodopsin and the others: a historical perspective on structural studies of G protein-coupled receptors. *Current pharmaceutical design*, *15*(35), 3994.

12. Kufareva, I., Rueda, M., Katritch, V., Dock, G. P. C. R., Stevens, R. C., & Abagyan, R. (2011). Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure*, *19*(8), 1108-1126.

13. Katritch, V., & Abagyan, R. (2011). GPCR agonist binding revealed by modeling and crystallography. *Trends in pharmacological sciences*, *32*(11), 637-643.

14. Michino, M., Abola, E., Participants, G. A., Brooks III, C. L., Dixon, J. S., Moult, J., & Stevens, R. C. (2009). Community-wide assessment of GPCR structure modeling and docking understanding. *Nature reviews. Drug discovery*,*8*(6), 455.

15. Beuming, T., & Sherman, W. (2012). Current assessment of docking into GPCR crystal structures and homology models: successes, challenges, and guidelines. *Journal of chemical information and modeling*, *52*(12), 3263-3277.

16. Kufareva, I., Katritch, V., Stevens, R. C., & Abagyan, R. (2014). Advances in GPCR modeling evaluated by the GPCR dock 2013 assessment: meeting new challenges. *Structure*, *22*(8), 1120-1139.

17. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., ... & Miyano, M. (2000). Crystal structure of rhodopsin: AG protein-coupled receptor. *science*, *289*(5480), 739-745.

18. Martinez-Archundia, M., Cordomi, A., Garriga, P., & Perez, J. J. (2012). Molecular modeling of the M3 acetylcholine muscarinic receptor and its binding site. *BioMed Research International*, *2012*.

19. Becker, O. M., Shacham, S., Marantz, Y., & Noiman, S. (2003). Modeling the 3D structure of GPCRs: advances and application to drug discovery. *Current opinion in drug discovery & development*, *6*(3), 353-361.

20. Archer, E., Maigret, B., Escrieut, C., Pradayrol, L., & Fourmy, D. (2003). Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors?. *Trends in pharmacological sciences*, *24*(1), 36-40.

21. Bywater, R. P. (2005), Location and nature of the residues important for ligand recognition in G-protein coupled receptors. J. Mol. Recognit., 18: 60–72. doi: 10.1002/jmr.685

22. Forrest, L. R., Tang, C. L., & Honig, B. (2006). On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins.*Biophysical journal*, *91*(2), 508-517.

23. Mobarec, J. C., Sanchez, R., & Filizola, M. (2009). Modern homology modeling of G-protein coupled receptors: which structural template to use?. *Journal of medicinal chemistry*, *52*(16), 5207-5216.

24. Liu, T., W Tang, G., & Capriotti, E. (2011). Comparative modeling: the state of the art and protein drug target structure prediction. *Combinatorial chemistry & high throughput screening*, *14*(6), 532-547.

25. Worth, C. L., Kleinau, G., & Krause, G. (2009). Comparative sequence and structural analyses of G-protein-coupled receptor crystal structures and implications for molecular models. *PloS one*, *4*(9), e7011.

26. Goldfeld, D. A., Zhu, K., Beuming, T., & Friesner, R. A. (2013). Loop prediction for a GPCR homology model: algorithms and results. *Proteins: Structure, Function, and Bioinformatics*, *81*(2), 214-228.

27. Latek, D., Pasznik, P., Carlomagno, T., & Filipek, S. (2013). Towards improved quality of GPCR models by usage of multiple templates and profile-profile comparison. *PloS one*, *8*(2), e56742.

28. Kimura, S. R., Tebben, A. J., & Langley, D. R. (2008). Expanding GPCR homology model binding sites via a balloon potential: A molecular dynamics refinement approach. *Proteins: Structure, Function, and Bioinformatics*, *71*(4), 1919-1929.

29. Zhang, Y., Sham, Y. Y., Rajamani, R., Gao, J., & Portoghese, P. S. (2005). Homology modeling and molecular dynamics simulations of the mu opioid receptor in a membrane–aqueous system. *Chembiochem*, *6*(5), 853-859.

30. Platania, C. B., Salomone, S., Leggio, G. M., Drago, F., & Bucolo, C. (2012). Homology modeling of dopamine D2 and D3 receptors: molecular dynamics refinement and docking evaluation. *PLoS One*, *7*(9), e44316.

31. Haga, K., Kruse, A. C., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., Zhang, C., ... & Kobayashi, T. (2012). Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature*, *482*(7386), 547-551.

32. Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katritch, V., ... & Iwata, S. (2011). Structure of the human histamine H1 receptor complex with doxepin. *Nature*, *475*(7354), 65-70.

33. Molecular Operating Environment; MOE,(2015). 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.

34. Wess, J., Liu, J., Blin, N., Yun, J., Lerche, C., & Kostenis, E. (1997). Structural basis of receptor/G protein coupling selectivity studied with muscarinic receptors as model systems. *Life sciences*, *60*(13), 1007-1014.

35. Kruse, A. C., Hu, J., Pan, A. C., Arlow, D. H., Rosenbaum, D. M., Rosemond, E., ... & Kobilka, B. K. (2012). Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature*, *482*(7386), 552-556.

36. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., ... & Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy.Journal of medicinal chemistry, 47(7), 1739-1749.

37. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. Journal of medicinal chemistry, 47(7), 1750-1759.

38. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., ... & Mainz, D. T. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. Journal of medicinal chemistry, 49(21),

6177-6196.

39. Cordomi A, Edholm O, Perez JJ. (2007) "Effect of different treatments of long-range interactions and sampling conditions in molecular dynamic simulations of rhodopsin embedded in a dipalmitoyl phosphatidylcholine bilayer." J Comput Chem;28:1017–30

40. Lindahl E, Edholm O. (2000) "Mesoscopic undulations and thickness fluctuations in lipid bilayers from molecular dynamics simulations". Biophys J;79:426–33

41. Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*,*118*(45), 11225-11236.

42. Lindahl, E., Hess, B., & Van Der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. Journal of Molecular Modeling,7(8), 306-317.

43. Berendsen, H. J., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. Computer Physics Communications, 91(1), 43-56.

44. Li, L., Li, C., Zhang, Z., & Alexov, E. (2013). On the dielectric "constant" of proteins: smooth dielectric function for macromolecular modeling and its implementation in Delphi. *Journal of chemical theory and computation*, *9*(4), 2126-2136.

45. Lin, S., Gether, U., & Kobilka, B. K. (1996). Ligand stabilization of the β2 adrenergic receptor: Effect of DTT on receptor conformation monitored by circular dichroism and fluorescence spectroscopy. *Biochemistry*, *35*(46), 14445-14451.

46. Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., ... & Kobilka, B. K. (2007). GPCR engineering yields high-resolution structural insights into β2-adrenergic receptor function. *science*,*318*(5854), 1266-1273.

47. Miller-Gallacher, J. L., Nehme, R., Warne, T., Edwards, P. C., Schertler, G. F., Leslie, A. G., & Tate, C. G. (2014). The 2.1 Å resolution structure of cyanopindolol-bound β1-adrenoceptor identifies an intramembrane Na+ ion that stabilises the ligand-free receptor. *PloS one*, *9*(3), e92727.

48. Lebon, G., Warne, T., Edwards, P. C., Bennett, K., Langmead, C. J., Leslie, A. G., & Tate, C. G. (2011). Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature*, *474*(7352), 521-525.

49. Wu, B., Chien, E. Y., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., ... & Stevens, R. C. (2010). Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science*, *330*(6007), 1066-1071.

CHAPTER 4

Insight into the stereochemical requirements of the bradykinin B$_1$ receptor 1 antagonists binding

4.0 Chapter summary

Bradykinin (BK) is a nonapeptide involved in several physiological and pathophysiological conditions including among others, septic and hemorrhagic shock, anaphylaxis, arthritis, rhinitis, asthma, inflammatory bowel disease. Accordingly, BK antagonists have long been sought after for therapeutic intervention. Action of BK is mediated through two different G-protein coupled receptors known as B1R and B2R. Although there are several B1 antagonists reported in literature, their pharmacological profile is not yet optimal, thus new molecules need to be described.

Knowledge of the stereochemical requirements of the ligand-receptor at this receptor could help in this direction. The absence of a crystallographic structure of B1R prompted a construction of an atomistic model of the B1R. The model was constructed by homology modelling using the chemokine CXC4 and bovine rhodopsin receptors as template. The model was further refined using molecular dynamics with the protein embedded in a POPC bilayer. Simulations were run for 600 ns.

From the simulations, an average structure was obtained that later used for docking studies using Glide software. Antagonists selected for the docking studies include Compound 11, Compound 12, Chroman-28, SSR240612, NPV-SAA164 and PS020990. The results of the docking study show the role of specific receptor residues in ligand binding. These results permitted the definition of a pharmacophore that describes the stereochemical requirements of antagonist binding, and were used in the development of new compounds with varied activity B1R inhibition profile depending on the number of pharmacophoric points fulfilled by the designed antagonistic molecules.

4.1 Introduction

Bradykinin (BK) is a member of the kinin family, a group of structurally related polypeptides that act locally to induce vasodilation and smooth muscle contraction. Kinins are liberated from the kininogens through the action of a set of serine proteases called Kallikreins.   There are two types of kininogens, high-molecular weight kininogen and low molecular weight kininogen; cleavage of high-molecular weight kininogen by kallikreins produces BK (Figure 4.1) and cleavage of low-molecular weight kininogen produce Kallidin (a decapeptide that has the same amino acid sequence as BK with the addition of a Lysine at the N-terminus) [1-2]. Kallidin is sometimes referred to as Lys-BK, and will thusly so be referred as from this point onwards where appropriate.

The kinins are converted into vasoactive peptides and the major physiological role appears to be the local response to tissue injury and as an agent to aid in tissue repair. Therefore, both tissue and plasma Kallikreins found are activated to produce the kinins by a number of physiological traumas. Once produced, BK and Lys-BK have a number of pharmacological effects which are virtually identical for the two peptides. Despite their similar pharmacological profiles, BK has by far been the most extensively studied of the two peptides [1, 3-4].

It should be noted that BK is also a very powerful vasodilator that increases capillary permeability and in addition constricts smooth muscle and stimulate pain receptors. Thus, BK is involved in several pathophysiological conditions including among others septic and haemorrhagic shock, anaphylaxis, arthritis rheumatoid, rhinitis, asthma, inflammatory bowel disease etc. [1-4].
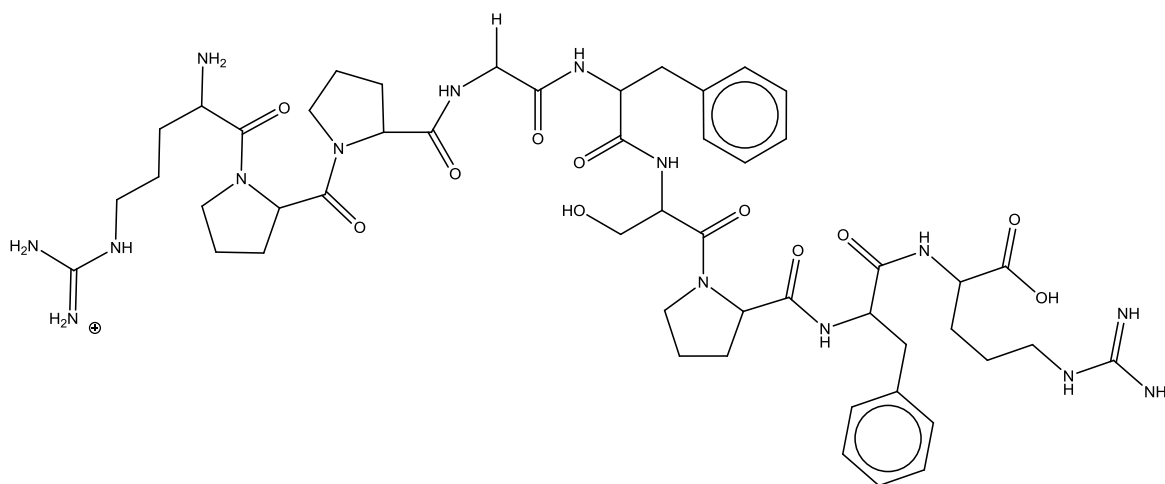
Figure 4.1: Chemical structure of the peptide bradykinin (BK)

Most physiological and pathophysiological effects of the kinins are believed to be mediated by the activation of two GPCRs: BK receptor 1 (B1R) and BK receptor 2 (B2R). The physiological ligands acting on B2R in humans are BK and Lys-BK. In contrast, BK is believed to be the only kinin acting on the B2R in rat and mouse. However, a decade ago a Kallidin-like peptide (KLP=Arg-BK) was isolated from plasma and urine of rats. The naturally occurring metabolites of Kallidin and BK, des-Arg10-Kallidin (desArg10-Kallidin) and des-Arg9-BK) are binding with high affinity to the B1R [5].

The pharmacological properties of the B1R in binding the endogenous kinin peptides are known to differ across species. Molecular cloning has revealed that these pharmacological differences arise from the diversity within the BDKRB gene [6]. The amino acid sequences of both receptors show an overall homology of only 36%, however, with higher areas of homology in the transmembrane regions. Despite their relative low homology, the receptors appear to share similar signal transduction pathways, such as activation of phospholipase C that leads to the release of inositol phosphates and increase in intracellular $[Ca^{2+}]$ levels, in response to binding of agonists [7].

In separate binding assays, the potency order of agonists were established (Table 1) and demonstrated that the removal of the C-terminal arginine is essential for high affinity on the B1R and detrimental for affinity toward the B2R. This holds true for Kallidin and des-Arg[10]-Kallidin, which have

higher affinities for the human and rabbit B1Rs compared with BK and des-Arg[9]-BK, respectively. However, the pharmacology of the mouse receptor is unique in that it possesses a 2-3-fold selectivity for des-Arg[9] -BK over the agonist des-Arg[10]-Kallidin [8].

While B2R are constitutively expressed in tissues, B1R molecules are not normally expressed in tissues but behave as key inducible molecules. In fact B1Rs are normally absent and may be highly upregulated following inflammatory stimuli (bacterial endotoxin, LPS-12h, ischemia, IL1, TNFαβ). This is an unusual characteristic of G-protein-coupled receptors. In fact, the pattern of induction of B1R resembles more closely to tyrosine-kinase-linked receptors. Kinin B1R seem to be upregulated under the same conditions described for the inducible pro-inflammatory enzymes (COX-2, iNOS). More recent reports indicate the existence of a crosstalk between B1R and B2R.

Table 4.1: Affinity Estimation (Ki) of kinins obtained using radio-ligand binding assay for recombinant kinin receptors, adopted from [3].

Peptide	Humans		Mice		Rabbits	
	B1R	B2R	B1R	B2R	B1R	B2R
Bradykinin	> 10,000	0,54	200	0,48	>5000	4.5
Kallidin (Lys-BK)	2.54	0.63	510	0.52	19	2
des-Arg9 -Bradykinin	1930	8100	0.7	6400	32	>1000
Des-Arg10-Kallidin	0.12	>30,000	1.7	25,000	0.23	>1000

Evidence also shows that persistent stimulation of B2R may result in B1R receptor upregulation. It is worth noting that stimulation of B2R receptors leads to transient increases in Ca^{2+}

concentration and to the fast desensitisation of these receptors. For the B1Rs, they do not seem to be susceptible to desensitisation mechanisms and their stimulation results in sustained elevations of Ca^{2+} concentration [9]. Both receptor subtypes for kinins can be expressed by the same cell types: vascular cells (endothelial, smooth muscle), fibroblasts, epithelial cells, nervous cells, and various tumor cells [3].Accordingly, BK antagonists have long been sought after for therapeutic intervention. However since the B2R is constitutively expressed on the membranes of most cells and most of the physiological and pathological processes of BK are mediated through this receptor, most attempts to develop antagonists of BK have been directed at the B2R. As it can be deduced, the fundamental distinction for ligand binding between the two receptors is that the B2R requires the whole nonapeptide for effective binding and activation, while the B1R selectively binds to the desArg[9] metabolites of BK and Lys-BK [1-4].

While stride has been made in the antagonism of B2R with the market release of Icatibant a peptidic antagonist of the receptor, there is no drug currently in the market designed to antagonize the B1 bradykinin receptor [10-11]. The B1R is not normally expressed in tissues, but has recently received a greater amount of attention as pre-clinical studies has implicated the expression of B1R in establishing and maintaining the signaling of chronic and inflammatory pain [3,12-13]. Furthermore, studies have also indicated B1R to be involved in hyperalgesia and leucocyte infiltration through the activation of a cytokine network. Thusly effective and selective and B1R antagonist are desirable as therapeutic agents for the treatment of both pain and inflammation [14-18].

The goal of the present work is to carry out a structure-activity study of B1R antagonism, taking into account the stereochemical features of diverse non-peptide antagonists and the way these features translate into ligand anchoring points of complementary regions in the receptor, through the analysis of the respective ligand-receptor complexes.

The compounds selected for the present study include: Compound 11 (**1**) [19-21], Compound 12 (**2**) [22-23], NPV-SAA164 (**3**) [24-25], Chroman-28 (**4**) [26-27], SSR240612 (**5**) [28] and PS020990 (**6**) [29]

shown in Figure 4.2. Compounds were docked into a refined model of the B1R constructed by homology modeling following the procedure explained in the methods section and the complexes were subsequently analyzed for their ligand-receptor interactions. The outcome of this study is summarized on a 3D pharmacophore that explains the observed structure-activity results and provides insight into the design of novel molecules with antagonistic profile.

1

Compound 11

2

Compound 12

3

NVP-SAA164

4

Chroman28

5

SSR240612

6
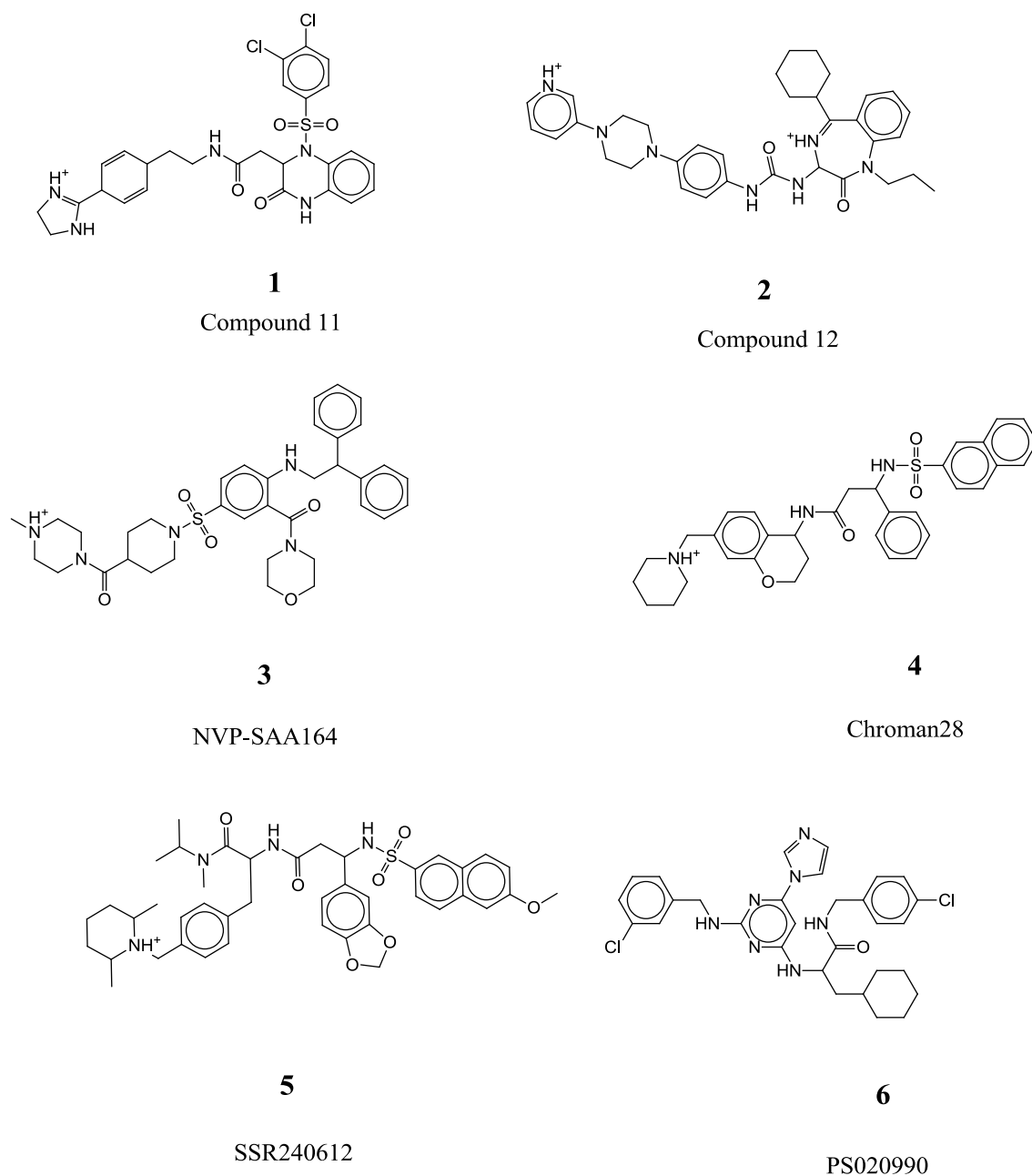
PS020990

Figure 4.2: Chemical structures of the B1R antagonists studied in the present work. Compound 11 (**1**), Compound 12 (**2**), NPV-SAA164 (**3**), Chroman-28 (**4**), SSR240612 (**5**) and PS020990 (**6**).

4.2 Homology modeling of the B1R

A starting model of the human B1R was constructed by homology modeling using the chemokine CXCR4 [30] as template that shares about 25% overall sequence similarity with the B1R. The template was selected due to its proximity to B1R in the GPCRs phylogenetic tree among those GPCRs whose crystallographic structure is known. The sequences of the two receptors were aligned, taking into account the conserved motifs found in all GPCRs, as well as the location of the six disulfide bridges. Since the chemokine CXCR4 lacks the helical TM8, residues of B1R known to be in the TM7 of all class GPCR starting from the highly conserved NPXXY motif [31] were modeled from the crystal structure of the bovine rhodopsin pdb entry code: 1GZM [32]. These motifs, together with salt bridges are important factors in constraining the conformation of the extracellular and transmembrane domains of the B1R. From the aligned sequences a starting model of the receptor was constructed using the Modeller 9 version 8 (9v8) software [33].

Model validation was carried out using the Molecular operating Environment (MOE) program [34]. In a subsequent step the B1 antagonist Compound 11 (**1** in Figure 4.2) was docked into the starting model using the GLIDE software [35]. The choice of this ligand was due to the abundant information available from the effect of mutations on its binding affinity [19-20]. Finally, the ligand-receptor complex was refined using molecular dynamics of the system embedded in a lipid bilayer.

Specifically, the protein was embedded in a box consisting in a 1-palmitoyl-2- oleoyl-sn-glycero-3-phosphocholine (POPC) lipids and water molecules generated and equilibrated according to the procedure described previously. The box had an initial size of 9.5 x 9.7 x 15.6 nm3 (XYZ), organized in such a way that the bilayer plane was oriented on the XY plane. Before protein insertion, the box contained 256 lipids (corresponding to an area per lipid of 0.64 nm2) and about 15,000 water molecules. The protein was placed in the center of the box, and the overlapping molecules were removed. Specifically, all water molecules with oxygen atoms closer than 0.40 nm to a non-hydrogen atom of the protein, as well as all lipid molecules with at least one atom closer than 0.25 nm to a non-hydrogen atom of the protein were removed. This resulted in a final system containing 242 lipids and

approximately 11,000 water molecules. Removal of these atoms produced small voids between the protein and water or lipid molecules that disappeared during the first part of the MD simulation, in which a progressive adjustment of the lipid bilayer and water molecules to the protein takes place. Next, 172 water molecules were randomly selected and replaced by 79 sodium and 93 chloride ions, providing a neutral system with a concentration approximately 0.2 M on sodium chloride. This concentration is fairly similar to that found in biological organisms, although they exhibit different intra- and extra-cellular ion concentrations. Sampling was carried out for 600ns using GROMACS package 4.6 [31].

The final model of the B1R was generated from the average structure of the molecular dynamics trajectory using the last 100 ns. The average structure of the receptor was subsequently minimized in a two-step process using the steepest descent method with a distance dependent dielectric constant of 2. First, side chains are optimized with the backbone atoms constrained to be subsequently released in a second minimization. This structure was used for further docking studies using the GLIDE [35] software. Due to the flexibility of the ligands, the docking process was carried out using a set of unique conformations resulted from thorough conformational searches. Poses were rank ordered using the XP score function of GLIDE [37]. Final poses of the compounds were decided based on their ranking and fulfillment of site directed mutagenesis information available.

4.3 Results and discussion

As mentioned above, in the absence of a crystallographic structure, an atomistic model of B1R was constructed by means of homology modelling. As explained in chapter 3, several experimental challenges hampers the availability of crystallographic structures of GPCRs, including their low-expression yields, low receptor stability after detergent extraction from native membranes, and high conformational heterogeneity. Although the number of the crystallographic structures available has steadily increased during the last few years, homology modelling remains to be one of the important

techniques to construct 3D models of proteins. However, in order for the models constructed to be as accurate as possible, the procedure requires a careful choice of the template and a robust refinement procedure [39-41]. This is important because from the analysis of the diverse known structures, although they share a common seven helix bundle, each structure exhibits specific features that might be relevant for ligand design [42].

The N terminal of the B1R as well as TM1-TM6 and part of TM7 was constructed using CXCR4 chemokine receptor (PDB entry code 3ODU) [30] as a template, while the part of TM7 starting from the conserved motif NPXXY was constructed using bovine rhodopsin as a template (pdb entry code 1GZM) [32]. Figure 4.3 shows the alignment of the sequences of the CXCR4 and B1Rs carried out taking into account the conserved motifs found among GPCRs, as explained in the methods section. This procedure is crucial for the assignment of the transmembrane regions. This information is then given as input to the Modeller software that produces a rank order set of models based on a scoring function. The final model selected for refinement was the one with the least steric conflicts from those that incorporated all the specified constraints considered to be conserved among GPCRs.

Previous experience in GPCR homology modeling [39] suggests that the presence of the ligand permits a faster equilibration of the system. Accordingly, Compound 11 was docked into the initial model. Several docking attempts were carried out using diverse conformations that were generated automatically as explained in the methods section. The final complex considered for refinement was selected based on the degree of fulfilment of diverse site directed mutagenesis studies. The B1R – Compound 11 complex was embedded into a pre-equilibrated bilayer of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and water and subjected to a molecular dynamics simulation. Time evolution of the root mean square deviation (rmsd) of the alpha carbons of the protein as well as those of its helical bundle subset is shown in Figure 4.4. Inspection of the Figure indicates that when all the alpha carbons of the protein are considered equilibration is reached after 350 ns, whereas when the helical bundle subset is used equilibration is reached about 100 ns earlier.
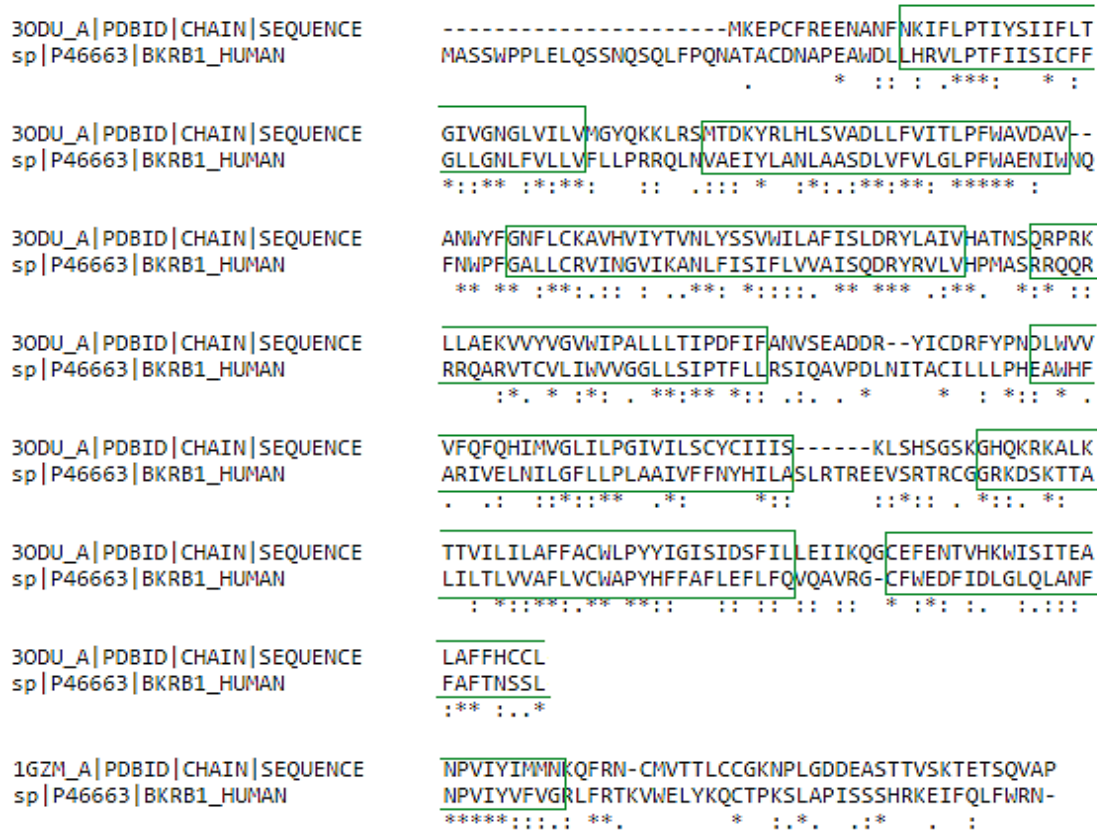
```
3ODU_A|PDBID|CHAIN|SEQUENCE        ---------------------MKEPCFREENANFNKIFLPTIYSIIFLT
sp|P46663|BKRB1_HUMAN              MASSWPPLELQSSNQSQLFPQNATACDNAPEAWDLLHRVLPTFIISICFF
                                                      .       *   :: : .***:   * :

3ODU_A|PDBID|CHAIN|SEQUENCE        GIVGNGLVILVMGYQKKLRSMTDKYRLHLSVADLLFVITLPFWAVDAV--
sp|P46663|BKRB1_HUMAN              GLLGNLFVLLVFLLPRRQLNVAEIYLANLAASDLVFVLGLPFWAENIWNQ
                                   *::** :*:**:    ::  .:::  *   :*:.:**:**: ***** :

3ODU_A|PDBID|CHAIN|SEQUENCE        ANWYFGNFLCKAVHVIYTVNLYSSVWILAFISLDRYLAIVHATNSQRPRK
sp|P46663|BKRB1_HUMAN              FNWPFGALLCRVINGVIKANLFISIFLVVAISQDRYRVLVHPMASRRQQR
                                    ** ** :**:.:: :  ..**: *::::. ** ***  .:**.  *:* ::

3ODU_A|PDBID|CHAIN|SEQUENCE        LLAEKVVYVGVWIPALLLTIPDFIFANVSEADDR--YICDRFYPNDLWVV
sp|P46663|BKRB1_HUMAN              RRQARVTCVLIWVVGGLLSIPTFLLRSIQAVPDLNITACILLLPHEAWHF
                                    :*. * :*: . **:** *:: .:. . *    *   : *:: * .

3ODU_A|PDBID|CHAIN|SEQUENCE        VFQFQHIMVGLILPGIVILSCYCIIIS------KLSHSGSKGHQKRKALK
sp|P46663|BKRB1_HUMAN              ARIVELNILGFLLPLAAIVFFNYHILASLRTREEVSRTRCGGRKDSKTTA
                                    .  .:  ::*::**  .*:    *:      ::*::  . *::. *:

3ODU_A|PDBID|CHAIN|SEQUENCE        TTVILILAFFACWLPYYIGISIDSFILLEIIKQGCEFENTVHKWISITEA
sp|P46663|BKRB1_HUMAN              LILTLVVAFLVCWAPYHFFAFLEFLFQVQAVRG-CFWEDFIDLGLQLANF
                                    : *::**:.** **::    :: :: :: ::  * :*: :.  :.:::

3ODU_A|PDBID|CHAIN|SEQUENCE        LAFFHCCL
sp|P46663|BKRB1_HUMAN              FAFTNSSL
                                   :** :..*

1GZM_A|PDBID|CHAIN|SEQUENCE        NPVIYIMMNKQFRN-CMVTTLCCGKNPLGDDEASTTVSKTETSQVAP
sp|P46663|BKRB1_HUMAN              NPVIYVFVGRLFRTKVWELYKQCTPKSLAPISSSHRKEIFQLFWRN-
                                   *****::..: **.           *  :.*.  .:*    .  :
```

Figure 4.3: Sequence alignment of human B1R, human CXC4 chemokine receptor and bovine rhodopsin receptor (TM8 only).

These results support the choice made in the present work of using the last 100 ns of the refinement process for the generation of an atomistic model of the B1R. As explained in the chapter 3, the last 100 ns segment of the molecular dynamics trajectory was used to generate an average structure that was subsequently minimized in a two-step process using the steepest descent method with a distance dependent dielectric constant of 2.
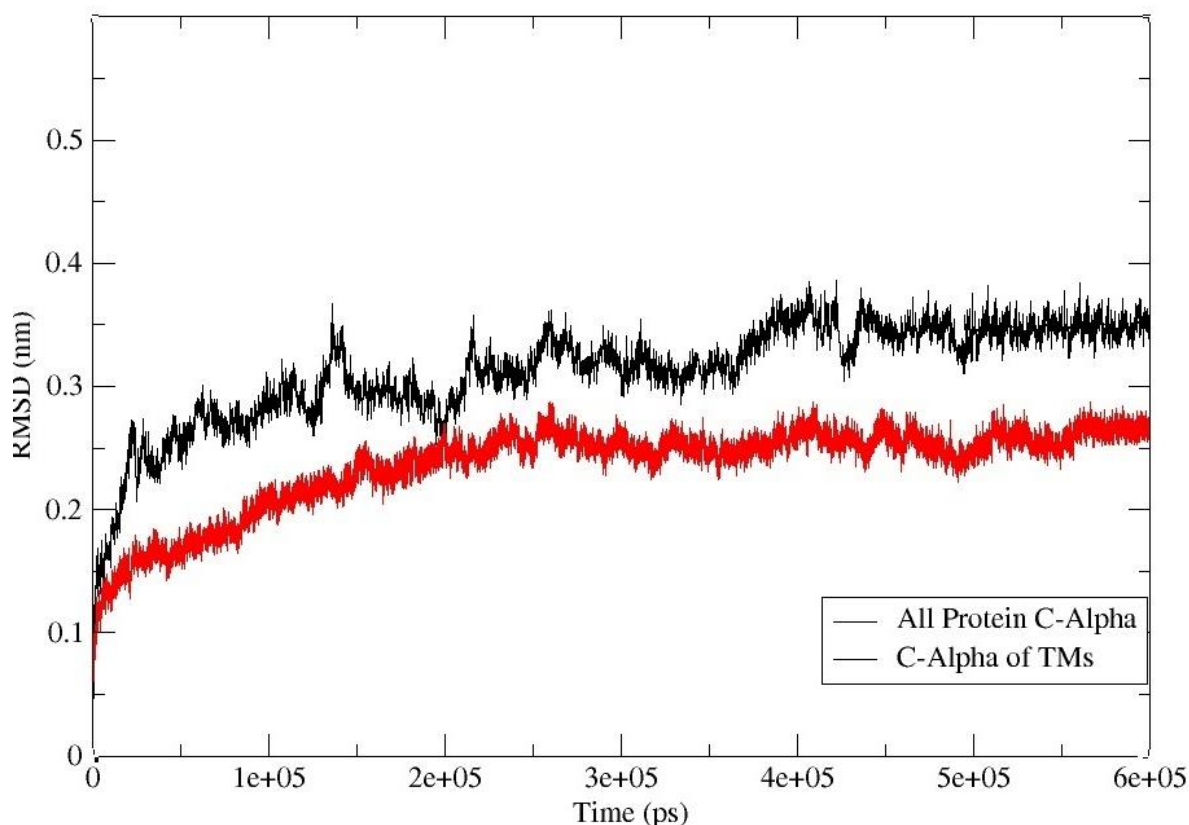
Figure 4.4: Time evolution of the *rmsd* of the B1R during the refinement process. In black is the *rmsd* of the alpha carbons of the protein and in red the rmsd of the alpha carbons of the helix bundle subset.

The orthosteric site of the B1R can be described as a large highly hydrophobic pocket formed by TM2, ECL1 and TM7 including residues like Trp[93], Ile[97], Phe[101], Phe[299], Phe[302] and a on the bottom left of the binding pocket. Interestingly most interactions at this side of the pocket involving Trp[93], Phe[101] and Phe[299] are coupled through quadrupole-quadrupole interactions. On the bottom right of the binding both polar and hydrophobic residues makes the binding site. Residues such as Asn[114] in TM3, Arg[202] of TM5 and Tyr[266] of TM6 are found on this binding pocket side. On the other hand, at the top and mouth of the binding site there are several polar residues including, Glu[196] and His[199] of TM5, Glu[273] and Gln[277] of TM6 and Glu[287], Asp[288], Asp[291] of TM7. Figure 4.5 shows the orthosteric pocket of the receptor. Described below are the bound conformation of each of the diverse antagonists used in this study, obtained from docking studies on the refined model.

(a)   (b)
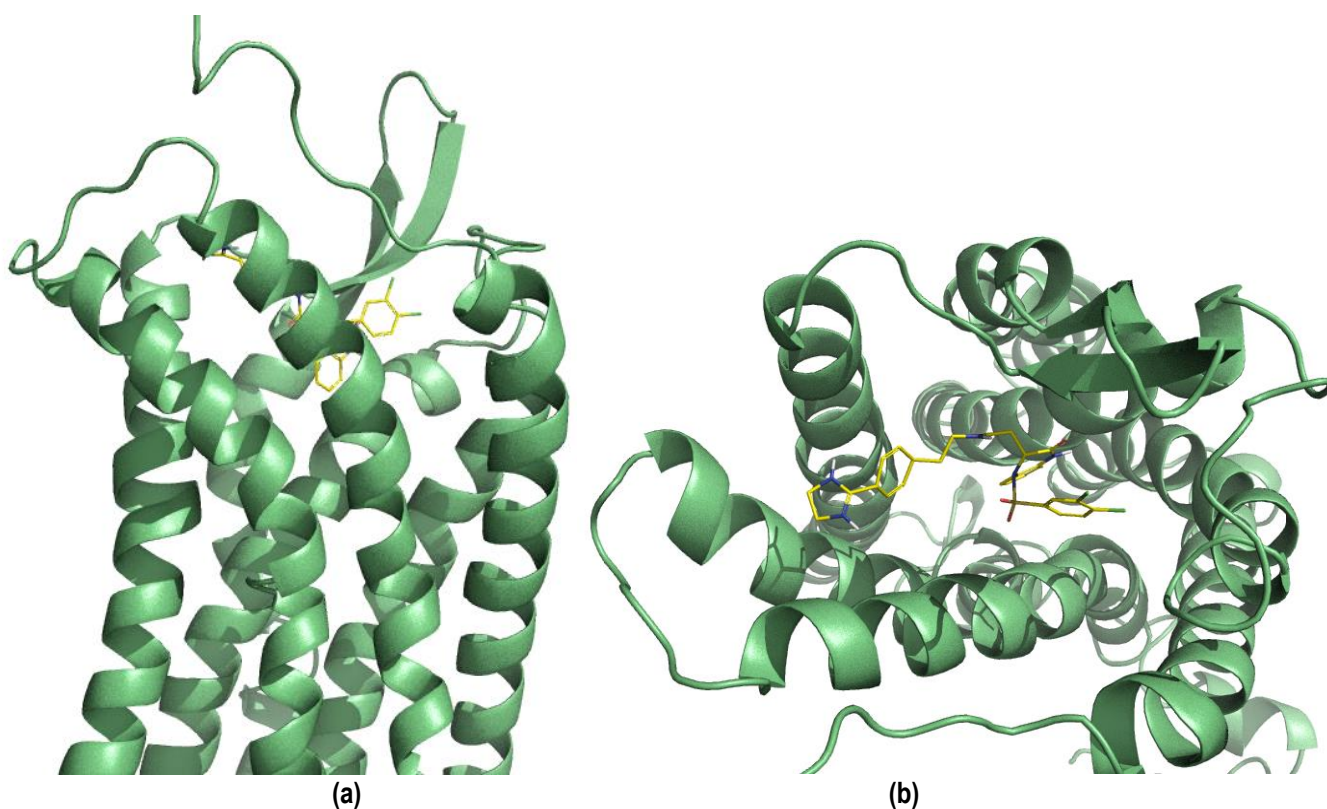
Figure 4.5: (a) Lateral view of the orthosteric binding pocket of the human B1R modeled in complex with compound 11, **(b)** same as (a) viewed from the extracellular side (aerial).

4.4 **Antagonist of B1Rs**

Compound 11

Developed by Merck research laboratories, compound 11 exhibits notable binding affinity for B1Rs in various animal species. In humans the Ki value of the compound is 0.034nM and decreases to 0.05nM in rabbit, 1.28nM in dogs and 62nM in rats [21]. The compound was discovered through an initial high-throughput screening, which led to the development of dihydroquinoxalinone as lead compound. Through homology modeling studies and structure-activity relationship studies (SAR), the affinity and potency of the lead molecules was optimized and leading to compound 11 [19]. Compound 11 has shown to antagonize both des-Arg[9]-BK and Lys-des-Arg[9]-BK induced contraction of the rabbit isolated aorta [18].

As can be seen in Figure 4.6, our B1R-compound 11 complex shows both the dihydroquinoxalinone moiety and the aromatic ring on the substituted sulfonamide (phenyl sulfonamide) added at the second nitrogen of the lead molecule occupying the hydrophobic pocket formed Try[103], Ile[97] among other at the bottom of the binding pocket. A poor interaction can be noted between the Asn[114] and the dihydroquinoxalinone moiety N1, but substitution of N1 with methyl group diminished the potency thus it was deduced sulfonamide was the key to further modification.

Similar results were obtained by Merck researchers from a homology modeling study using the bovine rhodopsin crystal structure as template in the process of lead optimization of compound 11. Mutagenesis studies shows that mutations at residues Ile[97], Trp[98], Asn[114], Gln[295], Asp[291] and Glu[273] affect the binding affinity at various degrees while mutation in Asn298 has little impact on the binding affinity [20], implying that the residues Ile[97], Trp[98], Asn[114], Gln[295], Asp[291] and Glu[273] are among those to interact directly with the Compound 11.
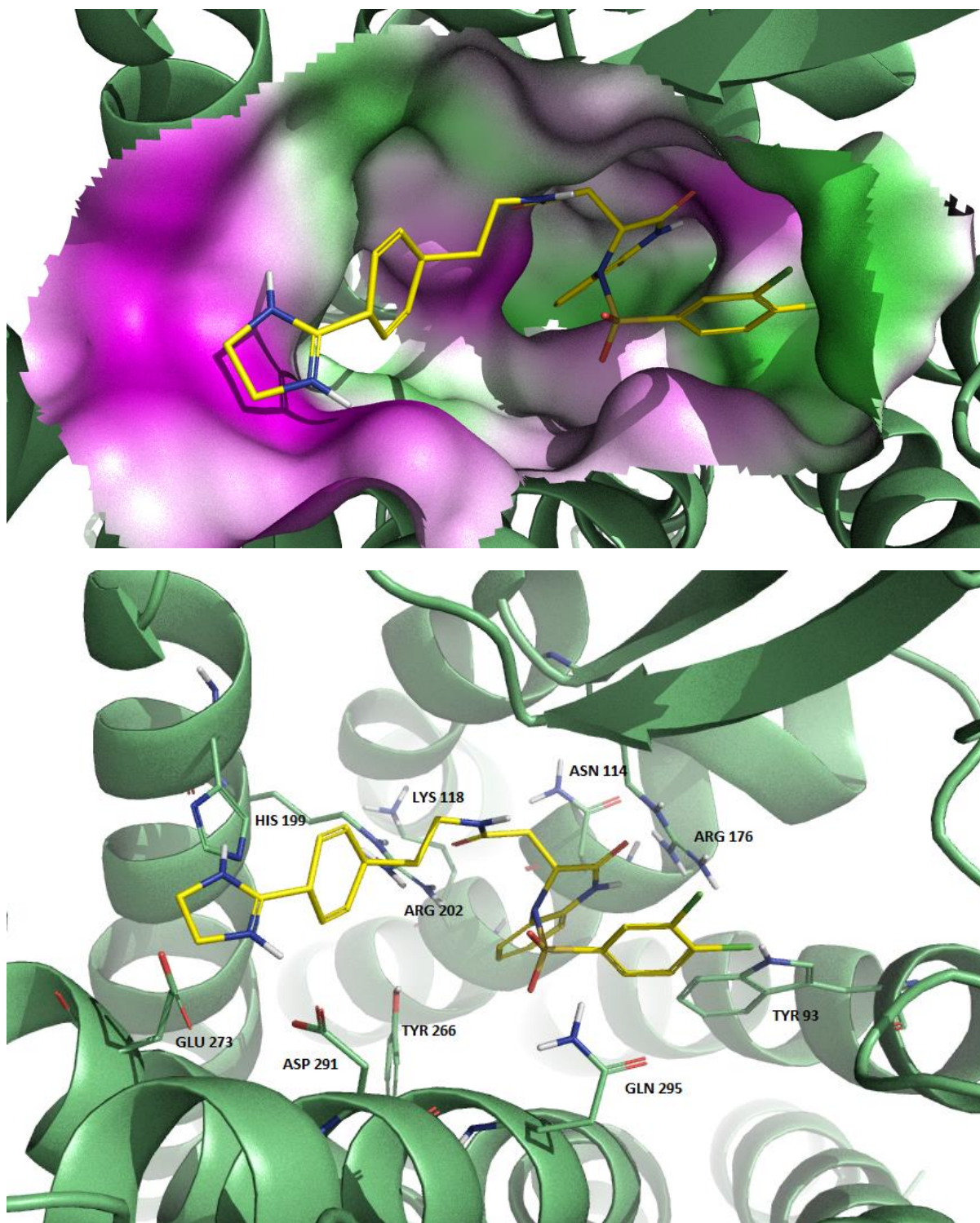
Figure 4.6: Pictorial view of B1R antagonist Compound 11 bound to the B1R.

Via the imidazole ring which has been noted to boost affinity to subnanomolar affinity, both TM6 residue Glu[273] and TM7 Asp[291] sidechains interacts with Compound 11 by forming hydrogen bonds. Hydrogen bonds are also observed between sulfonyl oxygen of the sulfonamide, Nitrogen and carboxyl group of the dihydroquinoxalinone and sidechains of Gln[295] in TM7, Asn[114] in TM3 and Arg[176] at the N terminus of ECL2 respectively. Comparatively in our model, Trp[98] is seen at the end of TM2 (N terminus of ECL1) also the mouth of the hydrophobic pocket of the binding site and its aromatic sidechain is positioned away from the binding pocket.

It can be speculated that Trp[98] contributes to facilitate the passing of the ligand to the pocket rather than on binding itself. Instead the dichlorophenyl moiety is situated at a suitable distance for hydrophobic interactions with Ile[97] and π-π interactions with Trp[93] of TM2. In TM6, hydrophobic interaction between Tyr[266] and the benzyl side of the dihydroquinoxalinone group is observed. Another hydrophobic interaction between the sidechain of His[199] in TM5 and the phenyl group attached to the imidazole is observed. Asn[298] in TM7 is observed to be close enough to interact with the sulfonyl oxygen but the conformation is unsuitable for hydrogen bond formation. Mutagenesis data indicate that Asn[298] has little effect in the binding of the antagonists. Indeed, mutagenesis studies suggest that any hydrophobic residue at this position is suitable for binding. These findings are in agreement with aforementioned SAR and mutagenesis studies results.

Compound 12

Like Compound 11, compound 12 was also developed by Merck research laboratories through optimization of novel benzodiazepines found to be B1R antagonists through in-house screening [22]. Replacement of dimethylamine group at the terminus end of the lead compound with hydrogen rendered the resulting compound unable to bind to B1R, but modification of the chemical group to piperidine led to an increase in activity and doubled the affinity. Further modification at piperidinyl with an addition on nitrogen, making a pyridine-piperazine group had a desirable effect on the affinity of the compound albeit with dependence to the position of the added nitrogen. At position 2 or 3

marked a notable decrease in affinity, so was the deletion of the pyridine nitrogen. Thusly a 4-pyridinylpiperazinyl group was deemed important and reduced basicity at this position was attributed to a decrease in the binding affinity. On the other end ,at the core benzodiazepine C-5 position an addition of phenethyl group rather than shorter alkyl group c-hexyl enhanced affinity by 4-fold at human and 9-fold at rat B1R [22].

This SAR information agrees with the bound conformation obtained when Compound 12 was docked to the refined human B1R model. Present results (Figure 4.7) show that the 4-pyridinylpiperazinyl interacts with several residues at the hydrophilic mouth of the binding pocket. Specifically, the 4-pyridnyl- end forms a hydrogen bond with the side chain of Asp[291] in TM7 while the piperazine group interacts with His[199] in TM5. An interaction with Gln[295] in TM7 is mediated through the carbonyl group found in the middle of the molecule. In addition, the ligand forms a hydrogen bond between the carbonyl of the diazepine ring and Asn[114]. Aromatic residues at the bottom left of the binding pocket (Trp[93] in TM2, and Phe[299] in TM7) are shown to be π-π stacked by the benzyl group while phenethyl moiety is seen to interact with both the Trp[93] in TM2 and the Phe[101] of the ECL1. In the absence of mutagenesis studies information, our results are in accord with the SAR studies information.
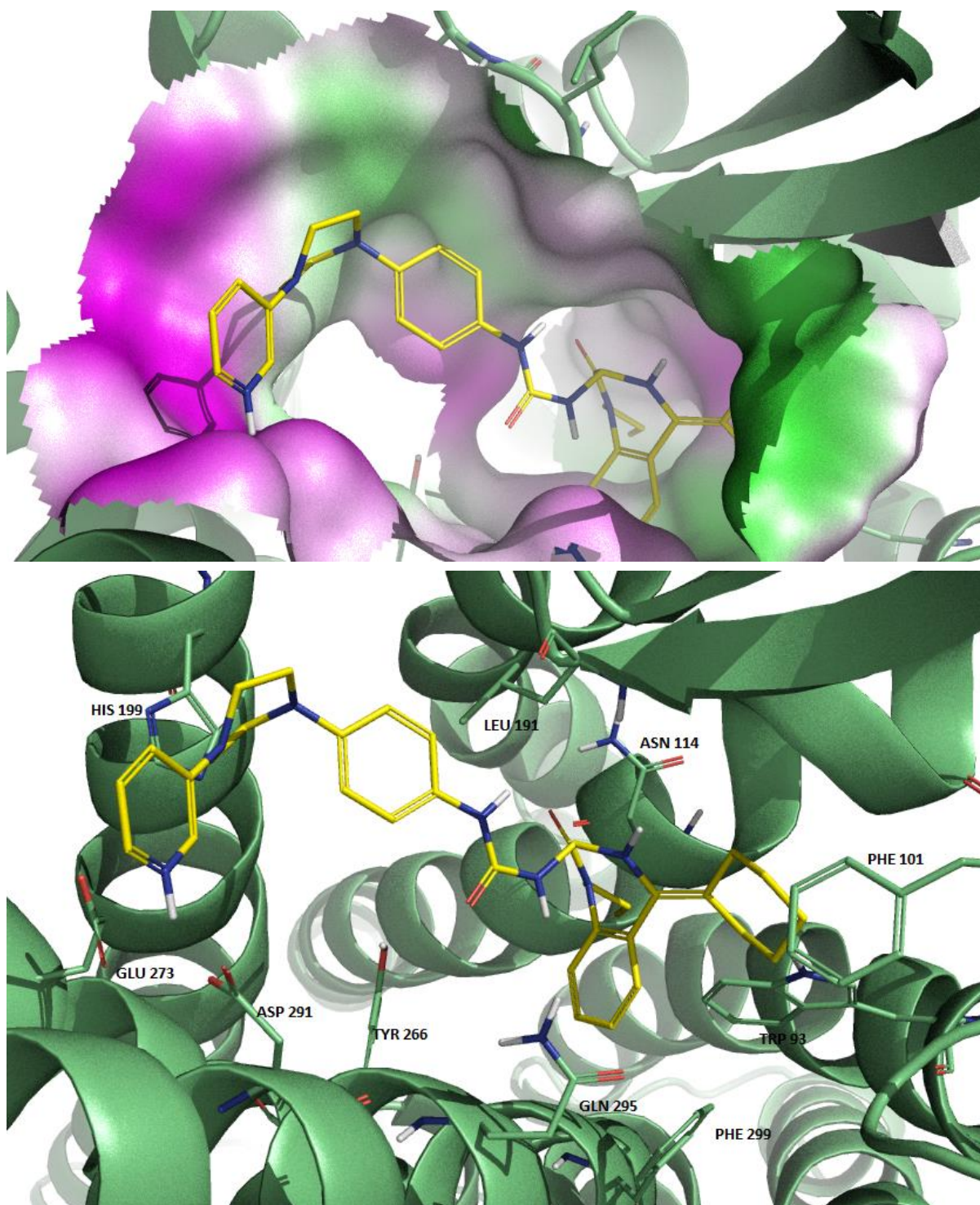
Figure 4.7: Pictorial view of B1R antagonist Compound 12 bound to the B1R.

NPV-SAA164

Developed by Novartis, NPV-SAA164 is an orally bioavailable antihyperalgesic that acts as a B1 antagonist with Ki of 8nM in HEK293 cells expressing human B1R and no activity in rat B1R [24-25]. Through in-house successful screening of non-peptidic B2 antagonists [38], binding affinity to B1 was assayed and led to the development of lead molecule 2-alkylamino-5-sulfomoybenzamide. Further investigation of the moieties in the lead molecules and their SAR optimization yielded NPV-SAA164. Although there are no mutagenesis studies available that can help to understand the effect of diverse mutations on the binding of NPV-SAA164 to B1R, but there is information about the effect of diverse chemical substitutions on the lead molecule [23] that can be analyzed through view of the complex model (Figure 4.8). For example, chemical replacement of the morpholine group to primary or secondary amide rendered the molecule inactive and to chlorine resulted to in lower affinity while those to tertiary amide improved the activity suggesting that at this position an ability to form both a hydrogen bond (both accepting and donating) is favored. In our B1R - NPV-SAA164 complex model, the morpholine group can be seen interacting with Arg[202] in TM5 as well as Tyr[266] in TM6. Asn[114] can also be seen exposed to the moiety.

The two phenyl groups in the lead compound are observed to bind in the aromatic pocket of our B1R model, the groups adopt a branched out conformation that allows NPV-SAA164 to interact with both the Trp[93] at the bottom of the aromatic pocket in TM2 and both Phe[101] in ECL1 and Leu[292] in TM7, at the top of the aromatic pocket. At the mouth of the binding pocket, the methyl-piperazine group form hydrogen bonds with the side chain of Asp[291] in TM7. Furthermore the piperidine moiety is positioned to interacts with the aromatic imidazole ring of His[199] in TM5 and the carbonyl group linking morpholine and phenyl-sulfonamide can be seen directly interacts with Gln[295] via a hydrogen bond.
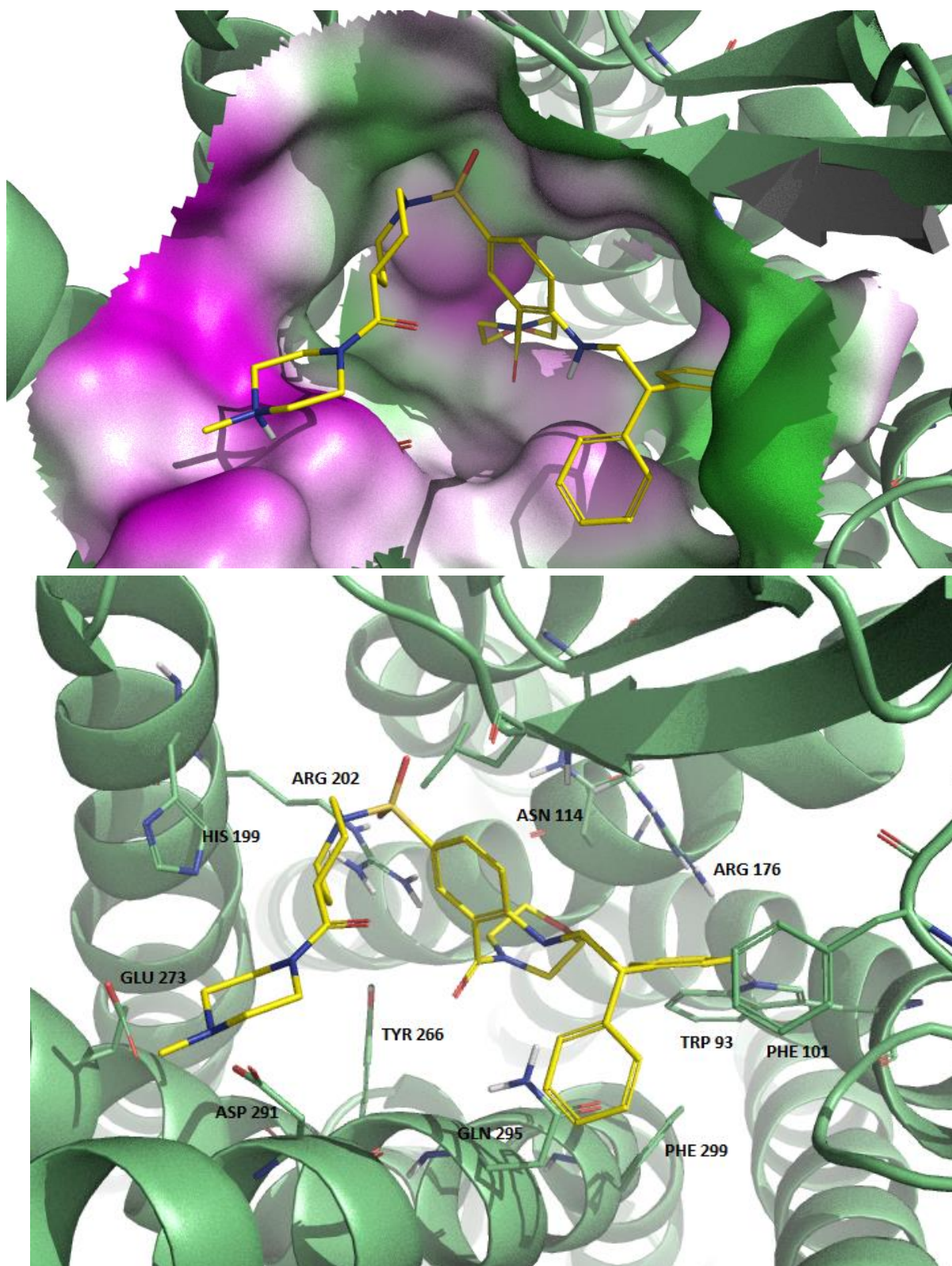
Figure 4.8: Pictorial view of B1R antagonist NPV-SAA164 bound to the B1R.

Chroman-28

Chroman-28 developed by Amgen Inc., is a potent and selective antagonist of human, primate, rat, and rabbit B1Rs with anti-inflammatory activity and Ki between 0.7 nM at human B1Rs [26]. Chroman-28 was developed through efforts to optimize spatial placement of lipophilic sulfonamide and basic amine moieties previous noted to be important for binding affinity and activity as B1R antagonist [39]. The earlier resulting compound had moderate activity against B1R and selectivity over B2. Addition of one methylene group in order to extend the distance between the two groups yielded another lead but did not improve the activity. Further modification on the second lead aimed at decreasing the flexibility of the compound so as to minimize the entropic penalty associated with adopting the preferential binding conformation led Chroman-28 and its analogues with improved potency [26, 45].

The SAR results for Chroman-28 and analogues at the piperidine group position showed that a hydroxyl group or a small amine NH2 decreases the pharmacokinetic properties of the lead compound, a pyrrolidine group or butyl amine improves the activity while a methylated piperidine decreases the activity. Furthermore changing the Chroman group to tetrahydro-naphthalene also had negative impact of the activity while modification at the sulfonamide to a bulkier aryl sulfonamide (naphthysulfonamide) led to an improvement in both activity and cellular potency [41]

Figure 4.9 shows the adopted conformation of Chroman-28 in our refined human B1R model. Hydrogen bonds are also observed between sulfonyl oxygen of the sulfonamide and residue Gln[295] in TM7, oxygen of the Chroman group and His[199] (including a hydrogen − aromatic ring interaction), Asp[291] in TM7 and the piperidine nitrogen, and carbonyl group of the amide group and Arg[202] both of TM5. Furthermore the sulfonyl oxygen also interacts with Arg[176] at the N terminus of ECL2 via a hydrogen bond. The naphthalene group in aromatic pocket can be seen quadruple- quadruple interact with Trp[93] in TM2, as well as Phe[101]. Phe[299] in TM7 and Ile[97] in TM2, a notable important residue in B1 antagonism, are in the vicinity to interact with the naphthalene group. Tyr[266] residue at TM6 can be seen to interact with the phenyl group, albeit at a bit longer distance.
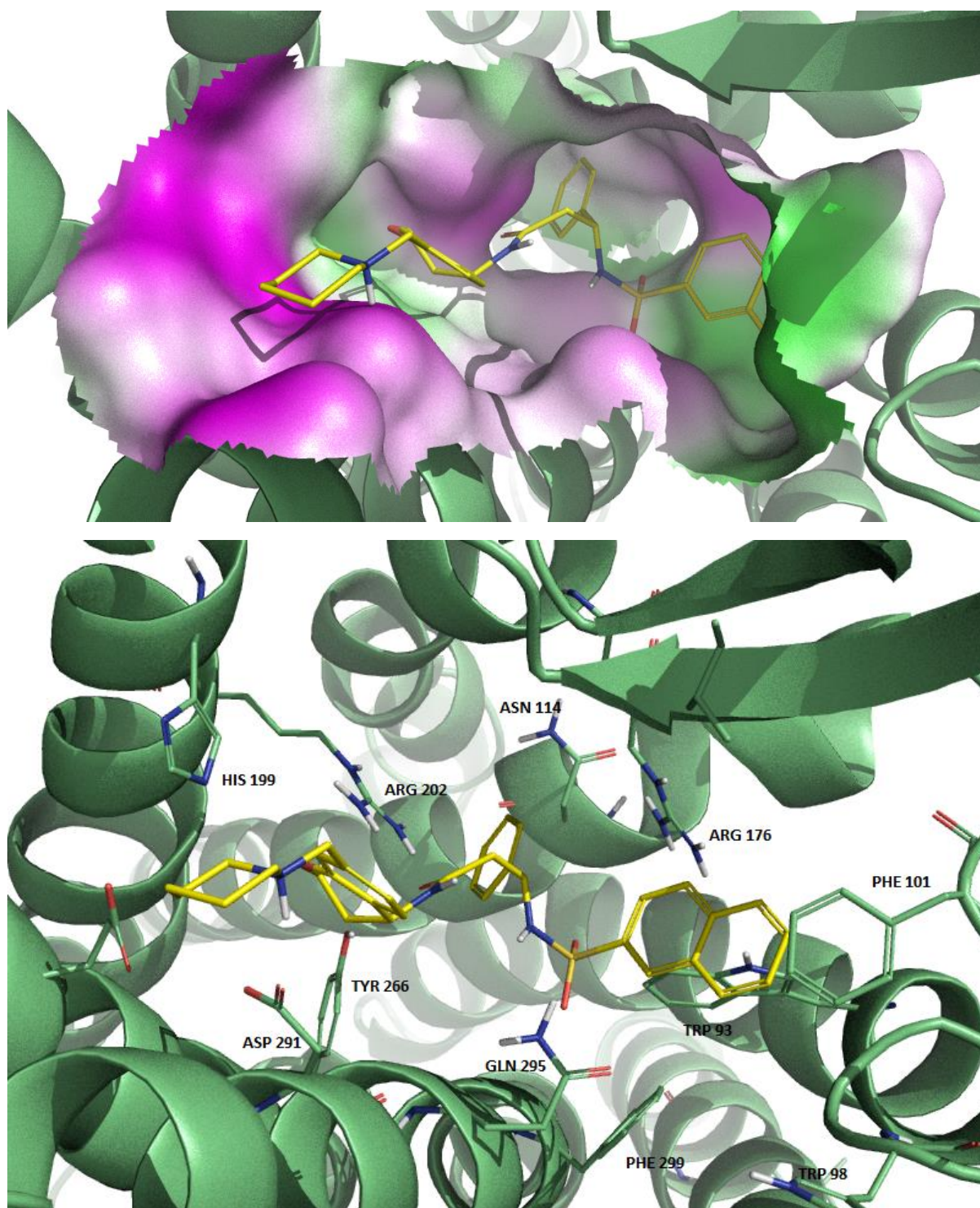
Figure 4.9: Pictorial view of B1R antagonist Chroman-28 bound to the B1R.

SSR240612

SSR240612, exhibits an inhibitory profile of Lys-des-Arg[9] binding in HEK cells expressing human B1R and MRC5 human lung fibroblast B1 expressing cells. Developed in 1997 by Sanofi research laboratories in France, SSR240612 was the first B1R antagonist with proven efficacy and oral bioavailability in models of pain and inflammation [28, 47]. In published literature, there are no reports on SAR nor site directed mutagenesis studies conducted on the compound. Taking in account the XP glide scores for the docking calculation performed with this molecule as well as bound conformation of other non-peptide B1R antagonists, Figure 4.10 shows a putative bound conformation of SSR240612.

An amide carbonyl at the middle of the molecule is observed to interact with Gln[295] at TM7 via hydrogen bond. On the other hand, a hydrogen bond is observed between the sulfonyl oxygen of sulfonamide and the side chains of residues Arg[174] at TM4 and Asn[114] in TM3. Apparently, the sulfonamide behaves as a linker of two binding pockets. Specifically, the methoxy-naphthalene group on the right side interacting with residue Tyr[266] in TM6 and Asn[298] in TM7 and on the left side, the benzyl group in benzodioxole moiety interacting with aromatic ring of Trp[93] sidechain while the oxygen in diaxole interacts with Asn[96] via hydrogen bond. Like in Chroman-28, the piperidine nitrogen (dimethyl-piperidine in SSR240612) also forms a hydrogen bond with Asp[291] in TM7.
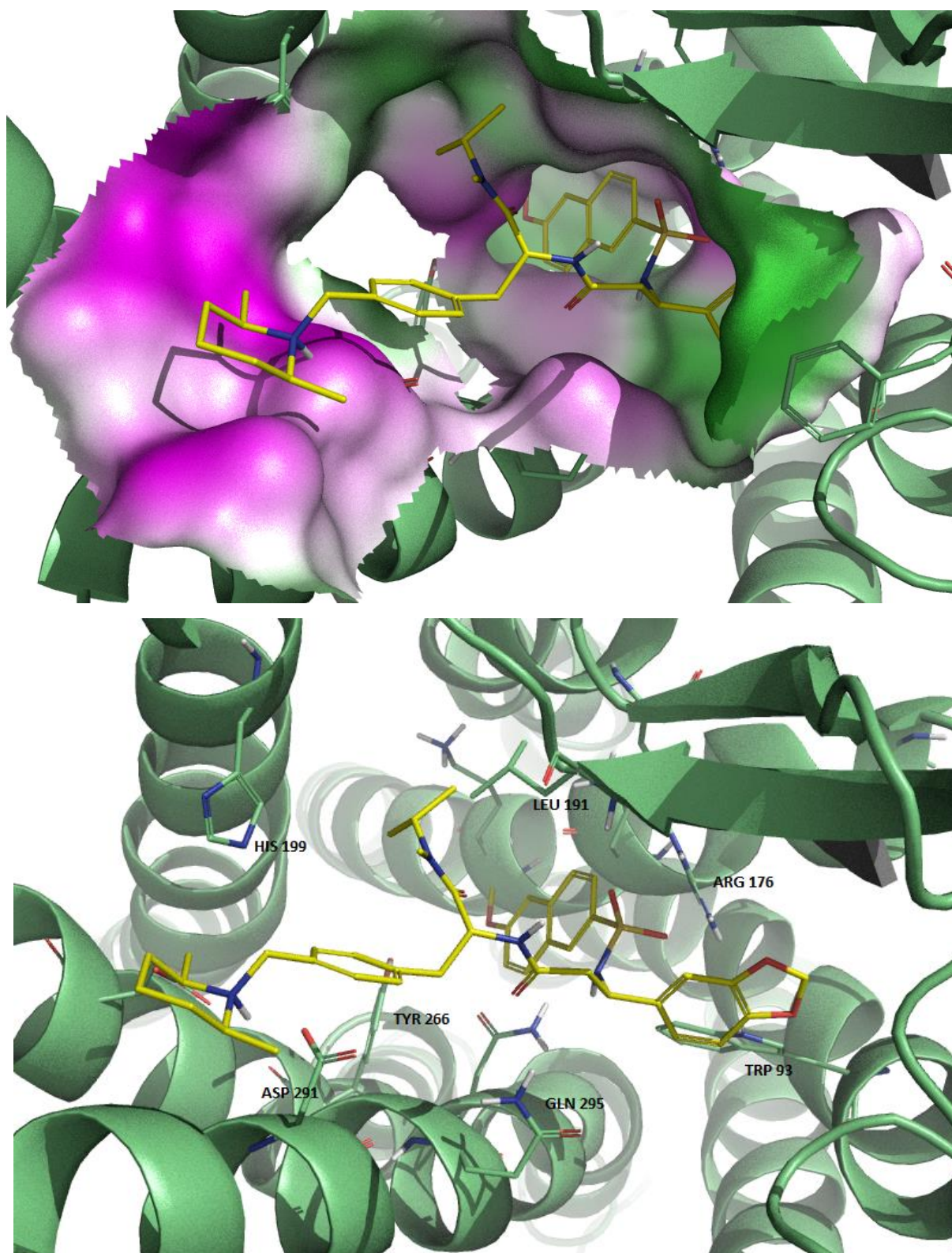
Figure 4.10: Pictorial view of B1R antagonist SSR240612 bound to the B1R.

PSO20990

Discovered by Pharmacopeia (Princeton, NJ) from a throughput screening program using

encoded combinatorial libraries, PS020990 exhibits a potent antagonist profile on human B1R, inhibiting ^{3}H-Des-Arg^{10}-Kalidin at IMR-90 cells with a K1 of 1nM. A set of active compounds permitted to generate an initial structure activity relationship which was later used to develop additional compounds with improved potency. When PS020990 and several compounds in the same lead series was assessed for inhibition at B2R, PS020990 and the related compounds were found to be inactive or with little activity , hence suggesting PS020990 to be a high selective B1R antagonists at reported a >1000-fold over B2R [29].

Present docking studies reveal (Figure 4.11) that the imidazole ring of PS020990 sits at the polar top end of the binding pocket interacting with the His^{199} in TM5 with its charged nitrogen also at a suitable distance to bond with Asp^{291} in TM7 via hydrogen bond. The oxygen atom of the cyclohexyl-propanamide moiety also forms a hydrogen bond with Gln^{295} in TM7, while one of the pyrimidine nitrogens interact with Arg^{202} and hydroxyl group of the Tyr^{266}. In the bottom of the binding site, one of the chlorophenyl group adjacent to the cyclohexyl-propanamide rests at the center of the aromatic site, coordinating the π-π interactions with Trp^{93} of TM2 that also involves Phe^{299} of TM7. Finally, the remaining chlorophenyl sits close to the backbone of Asn^{114}.
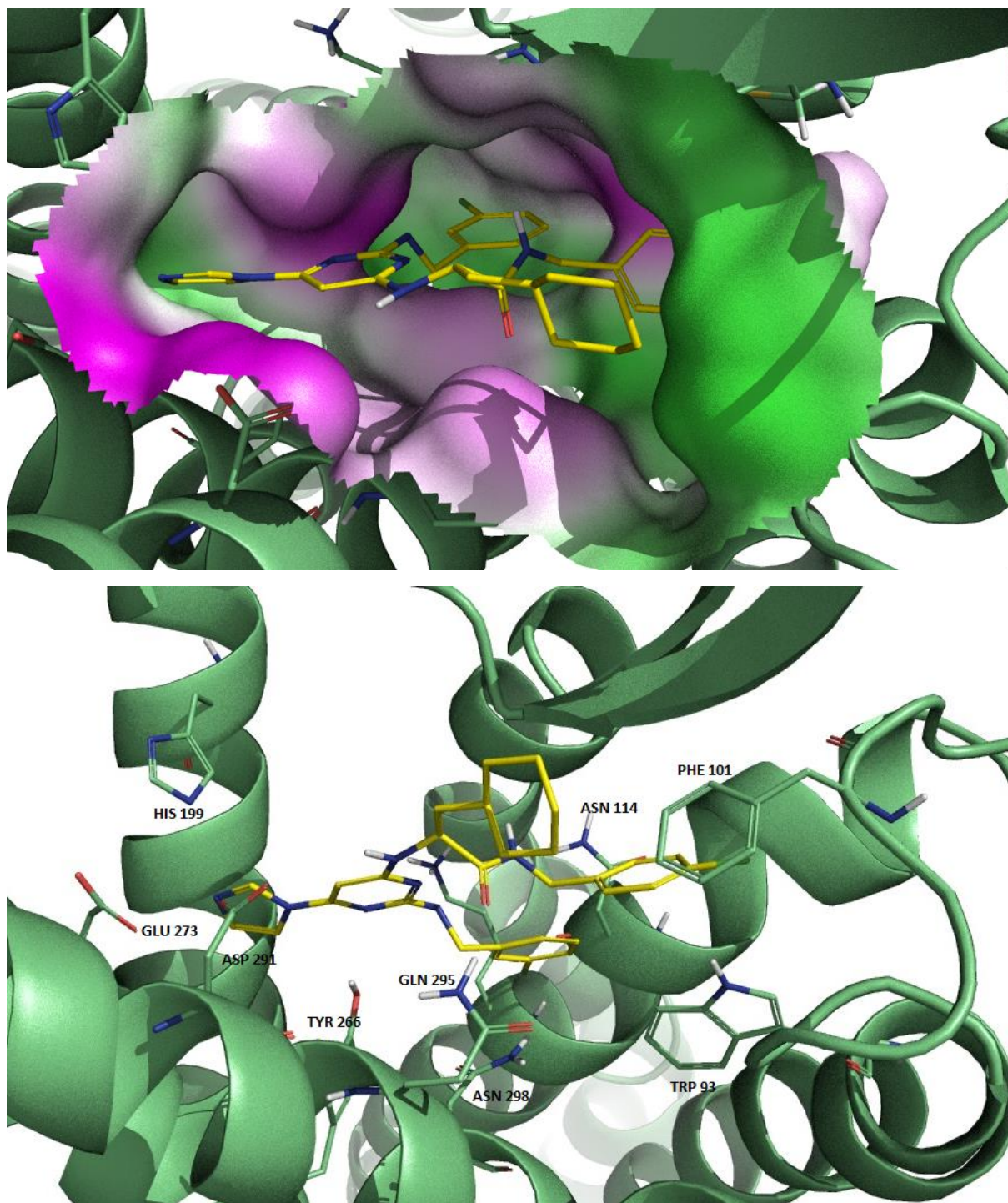
Figure 4.11: Pictorial view of B1R antagonist PS020990 bound to the B1R.

4.5 Designing pharmacophore of BKRB1 antagonists

Comparison of the ligand-receptor complexes of the diverse antagonists used for the present study allows the definition of a pharmacophore that explains the observed structure-activity relationship observed. The proposed pharmacophore is shown in Figure 4.12 and it is defined considering geometrical constraints on ligand moieties. It consists of five pharmacophoric points that all the ligands studied in the present study fulfill.
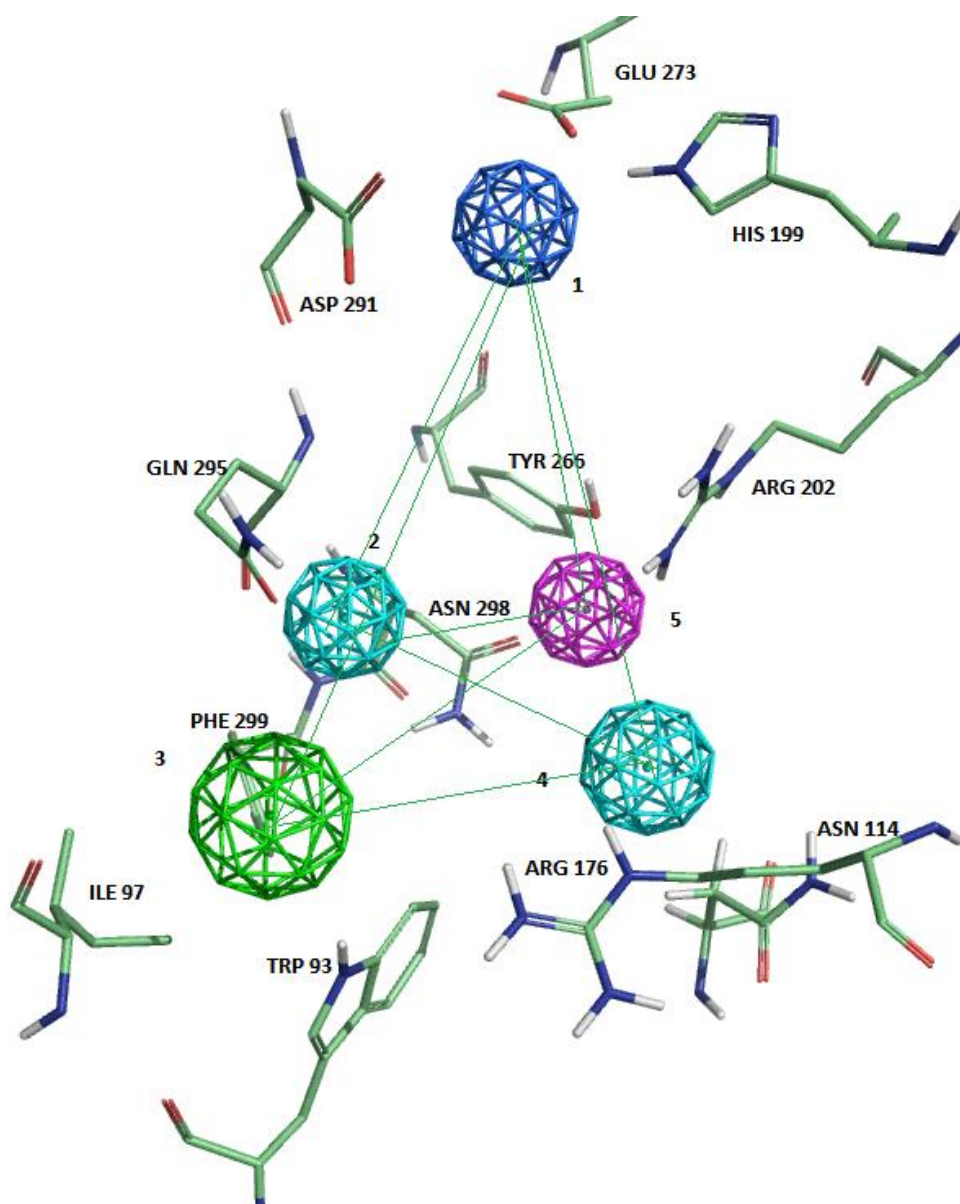
Figure 4.12: Proposed pharmacophore for the B1R antagonism. Distance between pharmacophoric points are: d(1,2)= 9.2 A; d(1,3)=13.1; d(1,4)= 1.5; d(1,5)=9.5; d(2,3)=5.1; d(2,4)=7.6.; d(2,5)=9.3; d(3,4)=7.8; d(3,5)=9.5; d(4,5)=5.7.

The first point, at the polar mouth of the binding pocket, is represented by a proton donor center that complements an interaction with Asp[291] and/or Glu[273] and /His[199]; point 2 is either a proton donor/proton acceptor center that will interact with either Gln[295]; point 3 is a hydrophobic ring that will interact with Trp[93] and/or Phe[299]; point 4 is a proton accepting center that will interact with Asn114 and/or Arg[176]; point 5 is a proton donor and /or proton acceptor site that will interact with Arg[202], Asn[298] and the side chain of Tyr[266].

Thus, Compound 11 fulfils point 1 by means of the 2-phenyldihydroimidazole; point 2 by the sulfonamide oxygen; point 3 by means of the chlorophenyl moiety; point 4 by the oxygen of the dihydroquinoxalinone group and point 5 by means of by the oxygen of the amide linker group. For Compound 12, point 1 is fulfilled by the 4-pyridinepiperazine moiety; point 2 is fulfilled by means of the amide linker moiety; point 4 by both the cyclohexyl group and the benzyl moiety of the benzodiazepine and point 4 and 5 are fulfilled by the carbonyl group of the diazepine moiety. In the case of the NPV-SAA164 point 1 is fulfilled by the methyl-piperazine group nitrogen; point 2 by means of the carbonyl group linking to the morpholine group that fulfills point 4; point 3 by both the phenyl group providing maximum coverage of the pocket side, point 5 by means of the sulfonyl oxygens of the sulfonamide. Chroman-28 covers the pharmacophore point 1 by using the chroman-7-methylpiperidine, point 2 and point 4 by means of the oxygens of the sulfonamide group; point 3 is fulfilled by the naphthalene rings and point 5 by the linker amide carbonyl group. In SSR240612 point 1 is fulfilled by the dimethyl-piperidine group; point 2 is fulfilled by the carbonyl group of the amide linker; point 3 is fulfilled by the benzodioxole moiety; while the sulphonyl oxygen fulfils point 4 and point 5 is fulfilled by the methyl-propanamide group. Finally, in the case of compound PS020990 point 1 is fulfilled by the imidazole group; point 2 by the carbonyl group of the cyclohexyl propanamide moiety; point 3 by means of the chlorophenyl group adjacent to the cyclohexyl propanamide moiety and point 5 by means of the nitrogen of the pyrimidine ring.

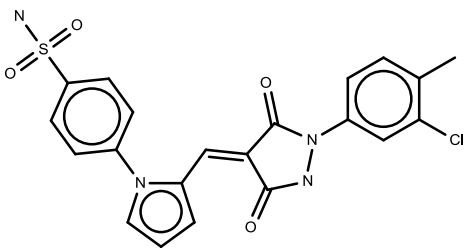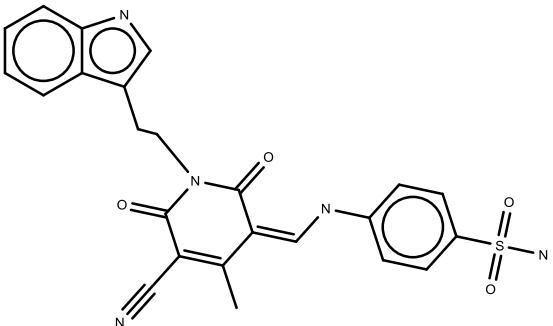### 4.6 Designing novel B1R antagonists: Proof of concept
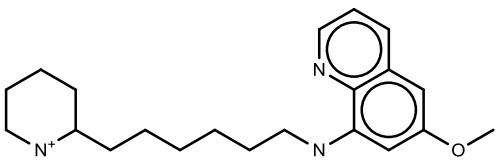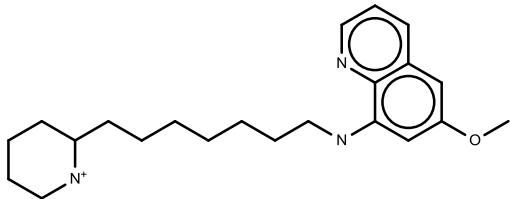
The pharmacophore described above was later used in the discovery of novel structures with antagonistic activity for B1R by virtual screening. For this purpose a search was made in few databases of 3D structures of compounds including the Available Chemical Directory (ACD), the Derwent World Drug Index, the National Cancer Institute (NCI) and Maybridge for approximately 500,000 compounds. A separated searches for compounds fulfilling five, four and three pharmacophoric points yielding a large set of compounds was carried out. A selected group of them were purchased and tested for their B1 antagonistic activity.

Biological assays permitted to identify new highly diverse hits with structures that do not resemble those used for pharmacophore development. The success rate achieved to identify new hits with an antagonistic activity by virtual screening was approximately one third of the selected molecules, as previously found by other authors [48]. Table 4.2 shows the structures as well as the antagonistic activity to the human bradykinin B1R of a selected group of hits that are disclosed to give support to the pharmacophoric hypothesis developed in this work.

These molecules were docked onto the receptor model and inspected for fulfillment of the pharmacophore. Information regarding the number of pharmacophoric points fulfilled by each of the hits is also included in Table 4.2. Interestingly, the antagonistic activity observed experimentally correlates well with the number of pharmacophoric points fulfilled by these molecules.

As an example, Figure 4.13 shows compound #3 bound to the B1R showing the fulfillment of the pharmacophoric points. The ligand is shown to fulfill 3 pharmacophoric points 3, 4 and 5. While it does have a charged amine moiety that can interacts with the Asp^{291}, and thus fulfil point 1, the amine is positioned at a long distance unsuitable for a proper interaction. This emphasize on the reason why most of the B1R antagonist published are big as they need to span out to all pharmacophoric points for suitable activity.

Table 4.2. Structures of the new hits discovered together with their antagonistic effect towards the human B1R and the number of pharmacophore points fulfilled.

Structure	Inhibition	Pharmacophore points
Compound #1	31% @ 50 μM	2
Compound #2	46% @ 50 μM	2
Compound #3	31% @ 10 μM	3
Compound #4	32% @ 10 μM	3
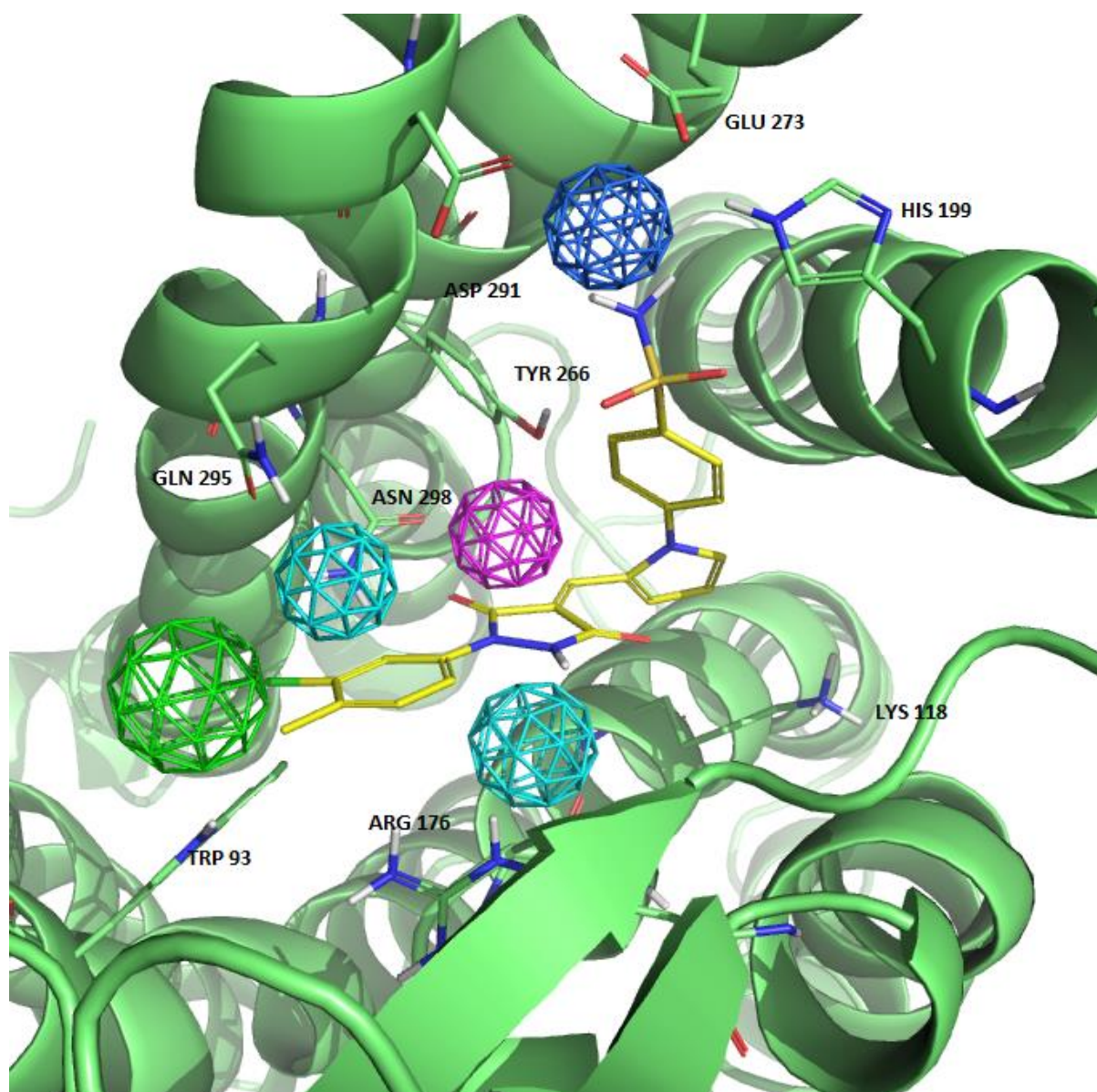
Figure 4.13: Pictorial view of compound #3 of Table 4.2 bound to the B1R with the pharmacophore points represented as spheres of different colors

4.7 **Conclusions to chapter 4**

Models of the bound conformation of diverse non-peptide human B1R antagonists were constructed and the stereochemical features of the complexes were analyzed with the aim to find common trends. To accomplish this, an atomistic model of the receptor was constructed by homology modeling, using the human CXC4 chemokine and bovine rhodopsin receptor as template (chimeric template). Antagonists selected for the present study included Compound 11, Compound 12, NPV-SAA164, Chroman-28, SSR240612 and PS020990, covering the maximum possible diversity. Complexes with the bound conformation of each of the antagonists were constructed by docking the molecules into the receptor. Due to the flexibility of the ligands and the size of the orthosteric site of the receptor, several docking attempts were carried out for each of the molecules. The final conformation was selected by the scoring function and the results of both SAR and site directed mutagenesis studies available in the published literature.

Our results suggest that there are certain anchoring points that are found in more than one compound permitting the definition of a common pharmacophore. This consist of five points that defined on the features of the ligand include proton donor (point 1); a proton acceptor/proton donor center (point 2); a hydrophobic ring (point 3); proton acceptor/proton donor center (point 4) and proton donor/proton acceptor.

The pharmacophore was used in a subsequent study to guide a virtual screening process. The results permitted to identify a set of compounds some of which were purchased and *in vitro* tested for their capability to antagonize the B1R. Consequently, a subset of compounds with highly diverse structures that give support to the validity of the pharmacophoric hypothesis described in this work, was also disclosed.

4.8 References to chapter 4

1. Hall, J. M. (1992). Bradykinin receptors: pharmacological properties and biological roles. Pharmacology & therapeutics, 56(2), 131-190.

2. Regoli D, Barabe J (1980) Pharmacology of bradykinin and related kinins. Pharmacol Rev 32: 1–46.

3. Leeb-Lundberg L M F, Marceau F, Muller-Esterl W, Pettibone D J, Zuraw B L (2005) International Union of Pharmacology. XLV. Classification of the Kinin Receptor Family: from Molecular Mechanisms to Pathophysiological Consequences. Pharmacol Rev 57: 27–77.

4. Stewart, J. (1995) Bradykinin Antagonists: Development and Applications. Biopolymers (Peptide Science) 37; 143-155.

5. Hilgenfeldt U, Stannek C, Lukasova M, Schnölzer M and Lewicka S. (2005): Rat tissue kallikrein releases a kallidin-like peptide from rat low-molecular-weight kininogen. Brit J Pharmacol 146: 958-963.

6. Hess JF, Hey PJ, Chen TB, Pettibone DJ and Chang RS (2002): Molecular and pharmacological diversity of the kinin B1 receptor. Int Immunopharmacol 2:1747-1754.

7. Faussner A, Bathon JM, Proud D (1999): Comparison of the responses of B1 and B2 kinin receptors to agonist stimulation. Immunopharmacol 45: 13-20.

8. Marceau, F., Hess, J. F., & Bachvarov, D. R. (1998). The B1 receptors for kinins. *Pharmacological reviews*, *50*(3), 357-386.

9. Calixto JB, Medeiros R, Fernandes ES, Ferreira J, Cabrini DA and Campos MM (2004): Kinin B1 receptors: key G-protein-coupled receptors and their role in inflammatory and painful processes. Br J Pharmacol 143: 803-818.

10. Hock F J, Wirth K, Albus U, Linz W, Gerhards G H, Wiemer G, Henke S, Breipohl G, Konig W, Knolle J, Scholkens B A (1991) Hoe 140 a New Potent and Long-Acting Bradykinin-Antagonist: in vitro Studies. Br J Pharmacol 102: 769-773.

11. Cole, S. W., & Lundquist, L. M. (2013). Icatibant for the treatment of hereditary

angioedema. Annals of Pharmacotherapy, 47(1), 49-55.

12. Marceau, F. (1995) Kinin B1Rs: A Review. Immunopharmacol. 30; 1-26.

13. Ahluwalia, A., & Perretti, M. (1999). B 1 receptors as a new inflammatory target. Could this B the 1?. Trends in pharmacological sciences, 20(3), 100-104.

14. Dray, A. (1995). Inflammatory mediators of pain. British Journal of Anaesthesia, 75(2), 125-131.

15. Biswas, K., Aya, T., Qian, W., Peterkin, T. A., Chen, J. J., Human, J., ... & Askew, B. C. (2008). Aryl sulfones as novel B1R antagonists for treatment of chronic pain. Bioorganic & medicinal chemistry letters, 18(17), 4764-4769

16. Wong, C. T., Rowlands, D. K., Wong, C. H., Lo, T. W., Nguyen, G. K., Li, H. Y., & Tam, J. P. (2012). Orally active peptidic B1R antagonists engineered from a cyclotide scaffold for inflammatory pain treatment. Angewandte Chemie, 124(23), 5718-5722.

17. Chen, J. J., & Johnson, E. J. (2007). Targeting the B1R to reduce pain.

18. Campos, M. M., Leal, P. C., Yunes, R. A., & Calixto, J. B. (2006). Non-peptide antagonists for kinin B 1 receptors: new insights into their therapeutic potential for the management of inflammation and pain. Trends in pharmacological sciences, 27(12), 646-651.

19. Su, D. S., Markowitz, M. K., DiPardo, R. M., Murphy, K. L., Harrell, C. M., O'Malley, S. S., ... & Bock, M. G. (2003). Discovery of a potent, non-peptide B1R antagonist. Journal of the American Chemical Society,125(25), 7516-7517.

20. Ha, S. N., Hey, P. J., Ransom, R. W., Harrell, C. M., Murphy, K. L., Chang, R.,. & Hess, F. J. (2005). Binding modes of dihydroquinoxalinones in a homology model of bradykinin receptor 1. Biochemical and biophysical research communications, 331(1), 159-166.

21. Morissette, G., Fortin, J. P., Otis, S., Bouthillier, J., & Marceau, F. (2004). A novel nonpeptide antagonist of the kinin B1R: effects at the rabbit receptor. Journal of Pharmacology and Experimental Therapeutics, 311(3), 1121-1130.

22. Wood, M. R., Kim, J. J., Han, W., Dorsey, B. D., Homnick, C. F., DiPardo, R. M., ... & Bock, M. G. (2003). Benzodiazepines as potent and selective bradykinin B1 antagonists. Journal of medicinal

chemistry, 46(10), 1803-1806.

23. Kuduk, S. D., Ng, C., Feng, D. M., Wai, J. M. C., Chang, R. S., Harrell, C. M,.. & Bock, M. G. (2004). 2, 3-Diaminopyridine B1R antagonists. Journal of medicinal chemistry, 47(26), 6439-6442.

24. Ritchie, T. J., Dziadulewicz, E. K., Culshaw, A. J., Müller, W., Burgess, G. M., Bloomfield, G. C., ... & Yaqoob, M. (2004). Potent and orally bioavailable non-peptide antagonists at the human B1R based on a 2-alkylamino-5-sulfamoylbenzamide core. Journal of medicinal chemistry, 47(19), 4642-4644.

25. Fox, A., Kaur, S., Li, B., Panesar, M., Saha, U., Davis, C., ... & McIntyre, P. (2005). Antihyperalgesic activity of a novel nonpeptide B1R antagonist in transgenic mice expressing the human B1R. British journal of pharmacology, 144(7), 889-899.

26. D'Amico, D. C., Aya, T., Human, J., Fotsch, C., Chen, J. J., Biswas, K., ... & Ng, G. (2007). Identification of a nonpeptidic and conformationally restricted B1R antagonist with anti-inflammatory activity. Journal of medicinal chemistry, 50(4), 607-610.

27. Chen, J. J., Qian, W., Biswas, K., Viswanadhan, V. N., Askew, B. C., Hitchcock, S., ... & Johnson, E. (2008). Discovery of dihydroquinoxalinone acetamides containing bicyclic amines as potent B1R antagonists. Bioorganic & medicinal chemistry letters, 18(16), 4477-4481.

28. Ferrari, B., Sarran, L., Planchenault, C., Poncelet, M., Maruani, J., Alonso, R., ... & Le Fur, G. (2004). SSR240612, a new non-peptide antagonist of the B1R. Biochemical and pharmacological characterization. Journal of Pharmacology and Experimental Therapeutics.

29. Horlick, R. A., Ohlmeyer, M. H., Stroke, I. L., Strohl, B., Pan, G., Schilling, A. E., ... & Sigal, N. H. (1999). Small molecule antagonists of the B1R. Immunopharmacology, 43(2), 169-177.

30. Wu, B., Chien, E. Y., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., ... & Stevens, R. C. (2010). Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science, 330(6007), 1066-1071.

31. Rosenbaum, D. M., Rasmussen, S. G., & Kobilka, B. K. (2009). The structure and function of G-protein-coupled receptors. Nature, 459(7245), 356-363.

32. Li, J., Edwards, P. C., Burghammer, M., Villa, C., & Schertler, G. F. (2004). Structure of bovine rhodopsin in a trigonal crystal form. Journal of molecular biology, 343(5), 1409-1438.

33. Sali A, Blundell T L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234: 779–815.

34. Molecular Operating Environment (MOE). (Chemical Computing Group Inc. 1010 Sherbooke St. West Suite #910 Montreal QC Canada H3A 2R7.

35. Friesner R A, Banks J L, Murphy R B, Halgren T A, Klicic J J, Mainz D T, Repasky M P, Knoll E H, Shaw D E, Shelley M, Perry J K, Francis P, Shenkin P S. (2004) Glide: A New Approach for Rapid Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 47: 1739–1749. 20

36. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark A E, Berendsen H J. (2005) GROMACS: fast flexible and free. J Comp Chem 26: 1701–1718.

37. Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T.,(2006) Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes," J Med Chem 49, 6177–6196.

38. Jones, C., Phillips, E., Davis, C., Arbuckle, J., Yaqoob, M., Burgess, G. M, .. & McIntyre, P. (1999). Molecular characterisation of cloned bradykinin B 1 receptors from rat and human. European journal of pharmacology, 374(3), 423-433.

39. Lupala, C. S., Rasaeifar, B., Gomez-Gutierrez, P., & Perez, J. J. (2015). 193 Effect of template selection on the construction of atomistic models of GPCRs by homology modeling. Journal of Biomolecular Structure and Dynamics,33 (sup1), 127-128.

40. Mobarec, J. C., Sanchez, R., & Filizola, M. (2009). Modern homology modeling of G-protein coupled receptors: which structural template to use?. Journal of medicinal chemistry, 52(16), 5207-5216.

41. Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., ... & Karplus, K. (2009). Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four

approaches that performed well in CASP8. Proteins: Structure, Function, and Bioinformatics, 77(S9), 114-122.

42. Katritch V, Cherezov V, Stevens R C. (2012) Diversity and modularity of G protein-coupled receptor structures. Trends Pharmacol Sci 33: 17-27.

43. Dziadulewicz E K, Ritchie T J, Hallett A, Snell C R, Davies J W, Wrigglesworth R, Dunstan A R, Bloomfield G C, Drake G S, McIntyre P, Brown M C, Burgess G. M, Lee W, Davis C, Yaqoob M, Phagoo S B, Phillips E, Perkins M N, Campbell E A, Davis A J, Rang H P. (2002) Nonpeptide Bradykinin B2R Antagonists: Conversion of Rodent-Selective Bradyzide Analogues into Potent Orally-Active Human Bradykinin B2R Antagonists. J Med Chem 45: 2160-2172.

44. Marceau, F. (2005). A possible common pharmacophore in the non-peptide antagonists of the B1R. Drugs, 59, 1-8.

45. Cannon, J. G. (1995). Analog design. In Burger's Medicinal Chemistry and Drug DiscoVery, 5th ed.; Wolff, M. E., Ed.; John Wiley and Sons Inc.: New York,; Vol. 1: Principles and Practice, pp 788-791.

46. Biswas, K., Li, A., Chen, J. J., D'Amico, D. C., Fotsch, C., Han, N., ... & Johnson, E. (2007). Potent nonpeptide antagonists of the B1R: structure-activity relationship studies with novel diaminochroman carboxamides. Journal of medicinal chemistry, 50(9), 2200-2212.

47. Pietrovski, E. F., Otuki, M. F., Regoli, D., Bader, M., Pesquero, J. B., Cabrini, D. A., & Zampronio, A. R. (2009). The non-peptide kinin receptor antagonists FR 173657 and SSR 240612: preclinical evidence for the treatment of skin inflammation. Regulatory peptides, 152(1), 67-72.

48. Doman, T. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R., Stallings, W. C., ... & Shoichet, B. K. (2002). Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B.Journal of medicinal chemistry, 45(11), 2213-2221.

CHAPTER 5

New insights into the stereo chemical requirements of the Bradykinin B2R antagonists binding

5.0 Chapter summary

Bradykinin (BK) is a member of the kinin family, released in response to inflammation, trauma, burns, shock, allergy and some cardiovascular diseases, provoking vasodilatation and increased vascular permeability among other effects. Their actions are mediated through at least two G-protein coupled receptors B1R, a receptor up-regulated during inflammation episodes or tissue trauma and B2R that is constitutively expressed in a variety of cell types.

The goal of the thesis chapter was to carry out a structure-activity study of B2R antagonism, taking into account the stereochemical features of diverse non-peptide antagonists and the way these features translate into ligand anchoring points to complementary regions of the receptor, through the analysis of the respective ligand-receptor complex. Hence, an atomistic model of the B2R was built by homology modeling and subsequently refined embedded in a lipid bilayer by means of a 600ns molecular dynamics trajectory. The average structure from the last hundred nanoseconds of the molecular dynamics trajectory was energy minimized and used as model of the receptor for docking studies. For this purpose, a set of compounds with antagonistic profile, covering maximal diversity were selected from the literature. Specifically, the set of compounds include Fasitibant, FR173657, Anatibant, WIN64338, Bradyzide, CHEMBL442294, and JSM10292. Molecules were docked into the B2R model and the corresponding complexes analyzed to understand ligand-receptor interactions. The outcome of this study is summarized in a 3D pharmacophore that explains the observed structure-activity results and provides insight into the design of novel molecules with antagonistic profile.

To prove the validity of the pharmacophore hypothesized, a virtual screening process was also carried out. The pharmacophore was used as query to identify new hits using diverse data bases of molecules. The results of this study revealed a set of new hits with structures not connected to the molecules used for pharmacophore development. A few of these structures were purchased and tested. The results of the binding studies show about a 33% success rate with a correlation between the number of pharmacophore points fulfilled and their antagonistic potency.

5.1 Introduction

As explained in the previous chapter, the pharmacological actions of kinins are mediated by at least two GPCRs: B1R and B2R. While the former is up-regulated during inflammation episodes or tissue trauma, the latter is constitutively expressed in a variety of cell types. Thus, the majority of kinins effects in vivo have been attributed to the B2R. For the B2R, agonists bradykinin (BK) and Lys-BK exhibit much higher affinity than the des-Arg[9]-BK metabolites which bind only to the B1R, being Lys-desArg[9]-BK a potent B1R agonist. Due to their diverse roles, such as in mediating pain and inflammation there has been a remarkable interest for identifying potent B2R antagonists for therapeutically intervention. Coincidentally, among the two kinins receptors, B2R is the most studied [1- 4].

Since the chemical synthesis of BK for the first time in the early 60s [5] diverse analogs with agonistic activity have been reported, providing key information about the relevance of every residue for ligand activity. Analogs with antagonistic activity were not available until 1985, when Pro[7] in BK was replaced by an aromatic D-amino acid [6]. This led to the first generation of antagonists with the synthesis of potent analogs, including D-Arg-[Hyp[3], D-Phe[7]]-BK (NPC-567) (Hyp = hydroxyproline); D-Arg-[Hyp[3],Thi[5,8],D-Phe[7]]-BK (NPC-349) (Thi = thienylalanine); or D-Arg-[Hyp[3],D-Phe[7],Leu[8]]-BK, among the most active compounds reported [7,8]. Although this first generation of antagonists was useful to understand the involvement of BK in many pathophysiological processes, these compounds exhibit drawbacks that prevent them to be used for therapeutical intervention. Specifically, they exhibit low affinity for the B2R compared to BK itself and are not selective, showing higher affinity for B2R. Interestingly, removal of their C-terminal arginine by carboxypeptidases results in a decrease of affinity for the B2R, turning them selective B1R antagonists.

A second generation of antagonists with improved pharmacological profile was designed on the basis that the C-terminus of BK adopts a β-turn when bound to the receptor, as had been suggested from spectroscopic and molecular modeling studies [9] and confirmed more recently, from solid state NMR experiments [10]. Thus, with the help of conformationally constrained unnatural amino acids, diverse analogs designed to mimic the secondary structural motif of BK at the C-terminus were synthesized. These studies resulted in the discovery of several potent antagonists, including icatibant (formerly known

as HOE-140) with sequence D-Arg[0]-[Hyp[3]-Thi5-D-Tic[7]-Oic8]-BK (Tic = tetrahydroisoquinoline; Oic = octahydroindole carboxylic acid) [11] or NPC17731 (D-Arg-[Hyp[3], D-HypE(trans-proyl)[7], Oic[8]]-BK) [12]. In a parallel effort, the search for the shortest peptide sequence retaining antagonistic activity lead to conclude that adoption of a β-turn conformation at the C-terminus is a necessary condition for high affinity to the B2R, but not sufficient. This conclusion came from the analysis of the binding affinity of diverse cyclic peptides inspired on the C-terminus of icatibant. Thus for example compounds like the cyclo-(Gly-Thi-D-Tic-Oic-Arg) [13] or cyclo-(Pro-Orn-D-Tic-Oic-Arg) [14] show poor antagonistic affinity for the B2R.

Accordingly, the affinity of icatibant and analogs was rationalized in terms of the interactions of the compound with the receptor, such that the β-turn at the C-terminus was thought to occupy a hydrophobic region on the orthosteric pocket, whereas the N-terminal arginine were thought to interact with the negatively charged residues Asp[266] and Asp[284] putatively located at the mouth of the receptor [15]. As an indirect proof of concept, the high affinity peptide D-Arg[0]-Arg[1]-Pro[2]-Hyp[3]-Gly[4]-Thi[5]-cyclo[Dab[6]-D-Tic[7]-Oic[8]-Arg[9]] (Dab = diaminobutyric acid) (MEN11270) exhibits a cyclic structure at the C-terminus mimicking the β-turn secondary structure and preserves the N-terminal segment of icatibant [16].

The second generation of B$_2$ antagonists represented an improvement in regard to the first one. Thus, in addition to have designed antagonist with high affinity for the B2R, these compounds are highly selective and exhibit an improved pharmacokinetic profile due to their higher resistance to enzymatic degradation. However, they exhibit a limited oral bioavailability. Thus Icatibant, the first B2R antagonist to reach the market and currently used for the symptomatic treatment of acute attacks of hereditary angioedema in adults with C1-esterase-inhibitor deficiency, needs to be administered via subcutaneous injection [17]. In order to improve the oral bioavailability, research efforts were put forward to design non-peptide B$_2$ selective antagonists.

This third generation of B2R antagonists includes diverse molecules disclosed during the 90s and the beginning of the XXI century [18, 19]. Specifically, WIN64338 developed at Sterling Winthrop was the first B2R non-peptide antagonist disclosed (**4** in Figure 5.1) [20]. Other compounds were disclosed in the

following years, including a series of compounds developed at Fujisawa, like FR173657 (**2** in Figure 5.1) [21]; bradyzide developed at Novartis (**5** in Figure 5.1) [22]; anatibant developed at Fournier (**3** in Figure 5.1) [23] or fasitibant developed at Menarini (**1** in Figure 5.1) [24]. These compounds are high affinity B2R selective antagonists with limited oral bioavailability. The major drawback of these compounds is their high molecular mass, which ranges between 500 and 600. Aimed at finding compounds with lower molecular mass, scientists at Jerini carried out a medicinal chemistry optimization process, using the 8-benzyloxy-2-methyl-quinoline moiety that is the common scaffold of several of the non-peptide antagonists listed above, as starting structure. Their study led to the design and synthesis of JSM10292 a potent B2R antagonist with similar affinity and selectivity to the previous compounds, but with lower molecular mass [25].

The objective of this chapter is to carry out a structure-activity study of B2R antagonism, taking into account the stereochemical features of diverse non-peptide antagonists and the way these features translate into ligand anchoring points to complementary regions of the receptor, through the analysis of the respective ligand-receptor complex. For this purpose, a set of compounds from the literature covering maximal diversity were selected. The compounds selected for the present study include Fasitibant (**1**) [24], FR173657 (**2**) [21], Anatibant (**4**) [23], WIN64338 (**5**) [20], Bradyzide (**6**) [22], CHEMBL442294 (**7**) [25], and JSM10292 (**8**) [26], shown in Figure 5.1. Compounds were docked into a refined model of the B2R constructed by homology modeling, following the procedure explained in the methods section and the complexes were further analyzed for their ligand-receptor interactions. The outcome of this study is summarized on a 3D pharmacophore that explains the observed structure-activity results and provides insight into the design of novel molecules with antagonistic profile.
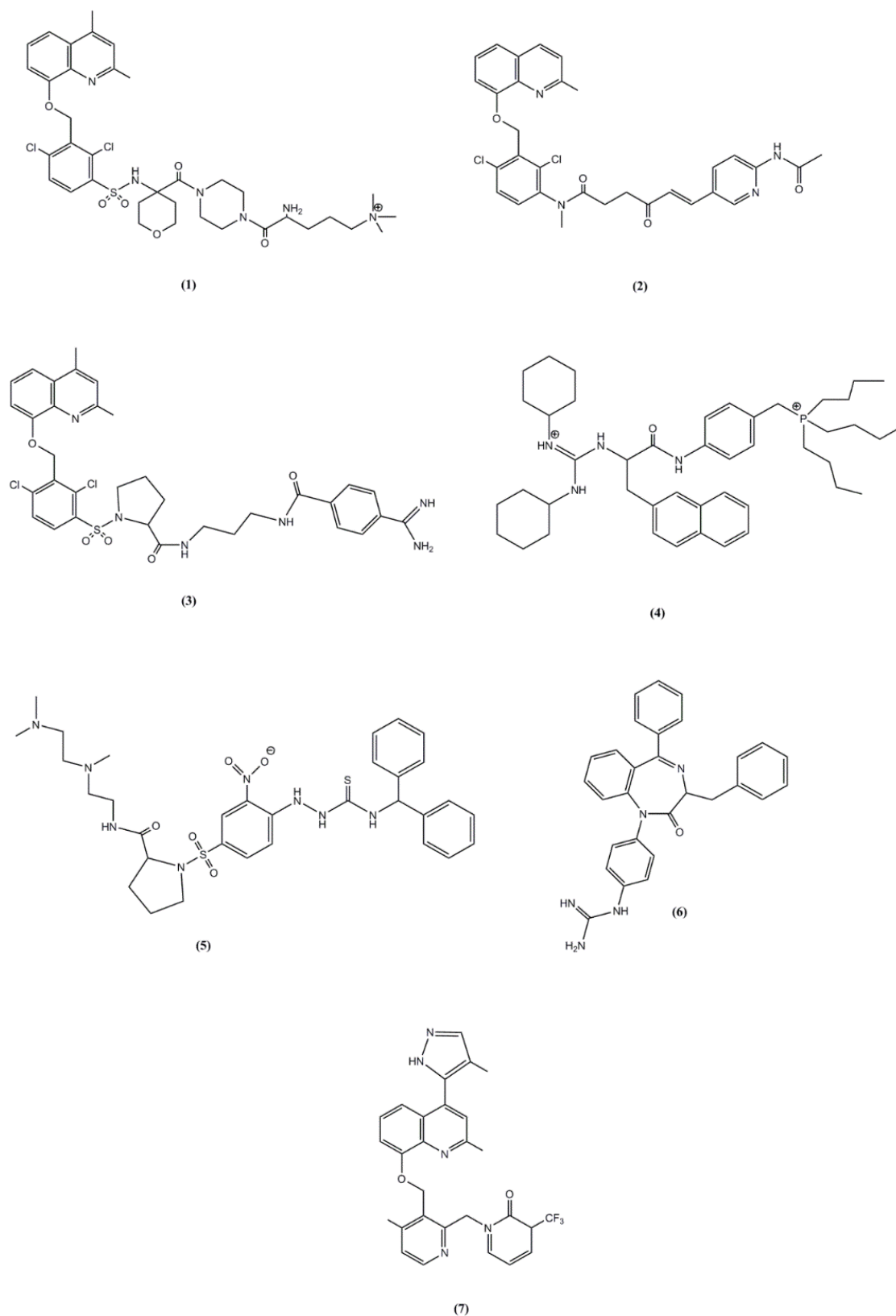
Figure 5.1. Chemical structures of the B2R antagonists studied in the present work. Fasitibant (1), FR173657 (2), Anatibant (3); WIN64338 (4); Bradyzide (5); CHEMBL442294 (6); JSM10292 (7)

5.2 Modeling of Bradykinin B2R

A starting model of the human B2R was constructed by homology modeling using the chemokine CXCR4 receptor as template. The template was selected due to its proximity to B2R in the GPCRs phylogenetic tree among those GPCRs whose crystallographic structure is known. The sequences of the two receptors were aligned, taking into account the conserved motifs found in all GPCRs, as well as the location of the disulfide bridges. These motifs, together with salt bridges are important factors in constraining the conformation of the extracellular and transmembrane domains of the B2R. From the aligned sequences a starting model of the receptor was constructed using the Modeller 9 version 8 (9v8) software [27]. Model validation was carried out using the Molecular operating Environment (MOE) program [28]. In a subsequent step the B2R antagonist Fasitibant (compound 7 in Figure 5.1) was docked into the starting model using the GLIDE software [29]. The choice of this ligand was due to the abundant information available from site directed mutagenesis experiments [30].

Finally, the ligand-receptor complex was embedded in a lipid bilayer and refined using molecular dynamics. Specifically, the protein was embedded in a box consisting in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids and water molecules previously equilibrated according to the procedure described elsewhere [31]. The box had an initial size of 9.5 x 9.7 x 15.6 nm3 (XYZ), organized in such a way that the bilayer plane was oriented on the XY plane. Before protein insertion, the box contained 256 lipids (corresponding to an area per lipid of 0.64 nm^2) and about 15,000 water molecules. The protein was placed in the center of the box, and the overlapping molecules were removed. More specifically, all water molecules with oxygen atoms closer than 0.40 nm to a non-hydrogen atom of the protein, as well as all lipid molecules with at least one atom closer than 0.25 nm to a non-hydrogen atom of the protein were removed. This resulted in a final box containing 228 lipids and circa 11,000 water molecules. Removal of these atoms introduced small voids between the protein and water or lipid molecules that disappeared during the first part of the MD simulation, in which a progressive adjustment of the lipid bilayer and water molecules to the protein takes place. Next, 171 randomly selected water molecules were replaced by 81 sodium and 90 chloride ions, providing a neutral system with a concentration

approximately 0.2 M on sodium chloride. This concentration is fairly similar to that found in biological organisms, although they exhibit different intra- and extra-cellular ion concentrations.

```
---------DYKDDDDAGAPEGISIYTSDNYTEEMGSGDYDSMKEPCFRE
MFSPWKISMFLSVREDSVPTTASFSADMLNVTLQGPTLNGTFAQSKCPQV
        .  :*:  .. .          * * :   : :     :. * :

ENANFNKIFLPTIYSIIFLTGIVGNGLVILVMGYQKKLRSMTDKYRLHLS
EWLGWLNTIQPPFLWVLFVLATLENIFVLSVFCLHKSSCTVAEIYLGNLA
*    .: : : *.:    ::*: . : * :*: *:   :*.  :::: *   :*:

VADLLFVITLPFWAVDAVAN--WYFGNFLCKAVHVIYTVNLYSSVWILAF
AADLILACGLPFWAITISNNFDWLFGETLCRVVNAIISMNLYSSICFLML
.***::.   *****:    *   * **: **:.*:.* ::*****: :* :

ISLDRYLAIVHATNSQRPRKLLAEKVVYVGVWIPALLLTIPDFIFANVSE
VSIDRYLALVKTMSMGRMRGVRWAKLYSLVIWGCTLLLSSPMLVFRTMKE
:*:*****:*:: .  * * :    *:  : :*  :***: * ::* .:.*

ADDR----YICDRFYPNDLWVVVFQFQHIMVGLILPGIVILSCYCIIISK
YSDEGHNVTACVISYPSLIWEVFTNMLLNVVGFLLPLSVITFCTMQIMQV
 .*.        *    **. :* *. ::   :**::** ** *    *:.

LSHS------KGHQKRKALKTTVILILAFFACWLPYYIGISIDSFILLEI
LRNNEMQKFKEIQTERRATVLVLVVLLLFIICWLPFQISTFLDTLHRLGI
*  :.    : : :*:*   .:::* *: ****: *.   :*::   * *

IKQGCEFENTVHKWISITEALAFFHCCLNPILYAFLGAKFKTS-------
LSS-CQDERIIDVITQIASFMAYSNSCLNPLVYVIVGKRFRKKSWEVYQG
:.. *: *. :.     .*:. :*: :.****::*..:* :*:..

-AQHALTSGRPLEVLFQ----------------------
VCQKGGCRSEPIQMENSMGTLRTSISVERQIHKLQDWAGSRQ
 .*:.   ..*::: .
```

Figure 5.2. Sequence alignment of human B2R (bottom) and CXC4 receptor. Transmembrane segments are inserted in boxes and sequence Identities are colored in blue.

Sampling was carried out for 600ns using the GROMACS package 4.6 [32]. The refined model of the B2R was generated from the average structure of the last 100 ns of the molecular dynamics trajectory. The structure was subsequently minimized in a two-step process using the steepest descent method with a distance dependent dielectric constant of 2. First, side chains are optimized with the backbone atoms constrained to be subsequently released in a second minimization. This structure was used for further docking studies using the GLIDE [29] software. Due to the flexibility of the ligands, the docking process

was carried out using a set of unique conformations resulted from thorough conformational analysis. Poses were rank ordered using the XP score function of GLIDE. Final poses of the compounds were decided based on their ranking and fulfillment of site directed mutagenesis information available.

5.3 Results and discussion

Challenges such as low-expression yields, low receptor stability after detergent extraction from native membranes, and high conformational heterogeneity hampers the availability of new crystallographic structures of GPCR. In the absence of a crystallographic structure of the B2R, an atomistic model of B2R was constructed by homology modelling for the present study. Under these circumstances homology modelling remains one of the important techniques aimed at constructing 3D models of proteins, however in order for the models constructed to be as accurate as possible the procedure requires a careful choice of the template and a robust refinement procedure. This is important because from the analysis of the diverse known structures, although they share a common seven helix bundle, each structure exhibits specific features that might be relevant for ligand design [34]. The CXCR4 chemokine receptor (pdb entry code 3ODU) [35] was selected as template for the present study due to its proximity with BK B_2 in the phylogenetic tree of the class A family of GPCRs.

Figure 5.2 shows the alignment of the sequences of the CXCR4 and B2Rs carried out taking into account the conserved motifs found among GPCRs, as explained in the methods section. This procedure is crucial for the assignment of the transmembrane regions. This information is then given as input to the Modeller software that produces a rank order set of models based on a scoring function. The final model selected for the refinement process was the one with the least steric conflicts from those that incorporated all the specified constraints considered to be conserved among GPCRs.

Before proceeding to the refinement process, Fasitibant was docked into the initial model. Due to its flexible structure several docking attempts were carried out using diverse conformations that were generated automatically as explained in the methods section. The final complex considered for

refinement was selected based on the degree of fulfilment of diverse site directed mutagenesis studies. Special attention was given to residues Trp[86], Ile[110], Trp[256], Asp[266] and Tyr[295] [30].

The complex Fasitibant bound-receptor was embedded into a pre-equilibrated bilayer of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and water and subjected to a molecular dynamics simulation. Previous experience in GPCR homology modeling (data not shown), suggest that the presence of the ligand permits a faster equilibration of the system.

Time evolution of the root mean square deviation (*rmsd*) of the alpha carbons of the protein as well as those of its helical bundle subset is show in Figure 5.3. Inspection of the Figure indicates that when all the alpha carbons of the protein are considered equilibration is reached after 300ns, whereas when the helical bundle subset is used equilibration is reached about 50ns earlier. These results support the choice made in the present work of using the last 100 ns of the refinement process for the generation of an atomistic model of the B2R. As mentioned in the methods section the last 100 ns segment of the molecular dynamics trajectory was used to generate an average structure that was subsequently minimized in a two-step process using the steepest descent method with a distance dependent dielectric constant of 2. Figure 5.4 shows the root mean square fluctuations (r*msf*) of the average structure extracted.

The orthosteric site of the B2R can be described as two hydrophobic pockets, a lager one formed between TM3, TM4, TM5 and TM6 including residues like Ile[110], Met[165], Leu[201], Trp[256], Phe[259] and a smaller one formed between TM2, TM3 and TM7 including residues like Trp[86] and Tyr[295]. Interestingly, the aromatic side chains Trp[86], Trp[256] and Tyr[295] are coupled through quadrupole-quadrupole interactions. On the other hand, at the mouth of the site there are several polar residues including, Glu[24], Gln[33], Asp[266], Asp[284] and Gln[288]. Figure 5.5 shows the orthosteric pocket of the receptor. Below, the bound conformation of the diverse antagonists selected for the present study obtained from docking studies on the refined model are described.
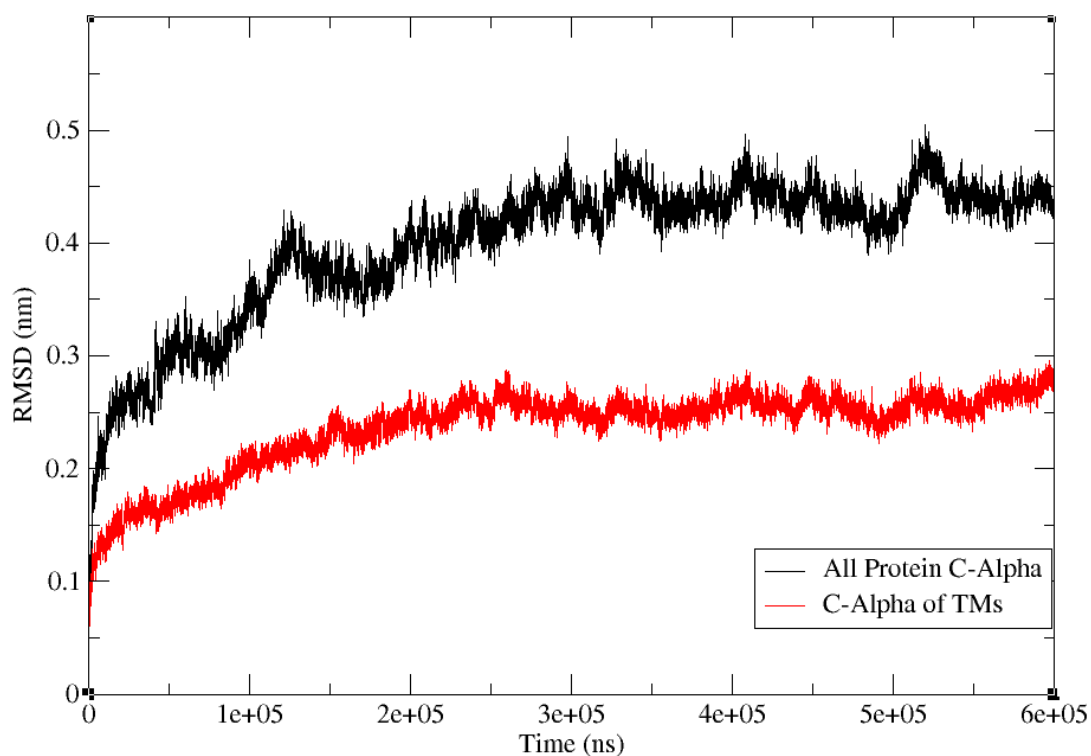
Figure 5.3. Time evolution of the root mean square deviation (rmsd) of the bradykinin B2R during the refinement process. In black is the rmsd of the alpha carbons of the protein and in red the rmsd of the alpha carbons of the helix bundle subset.
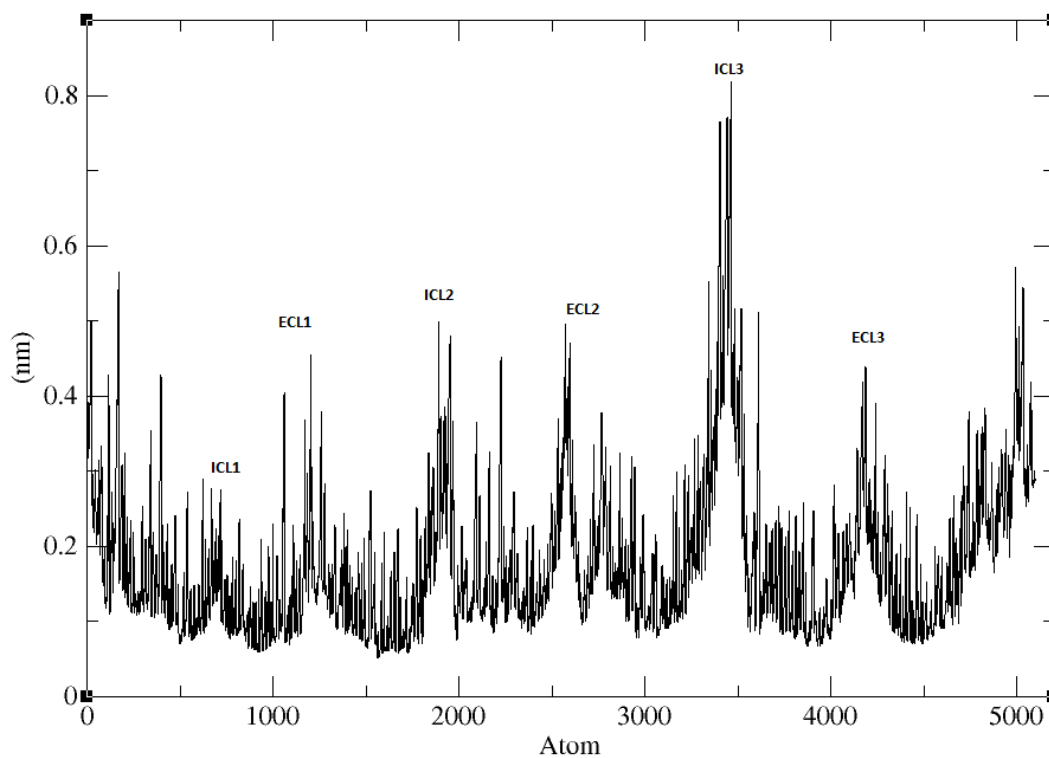
Figure 5.4. The rmsf of the average structure of B2R
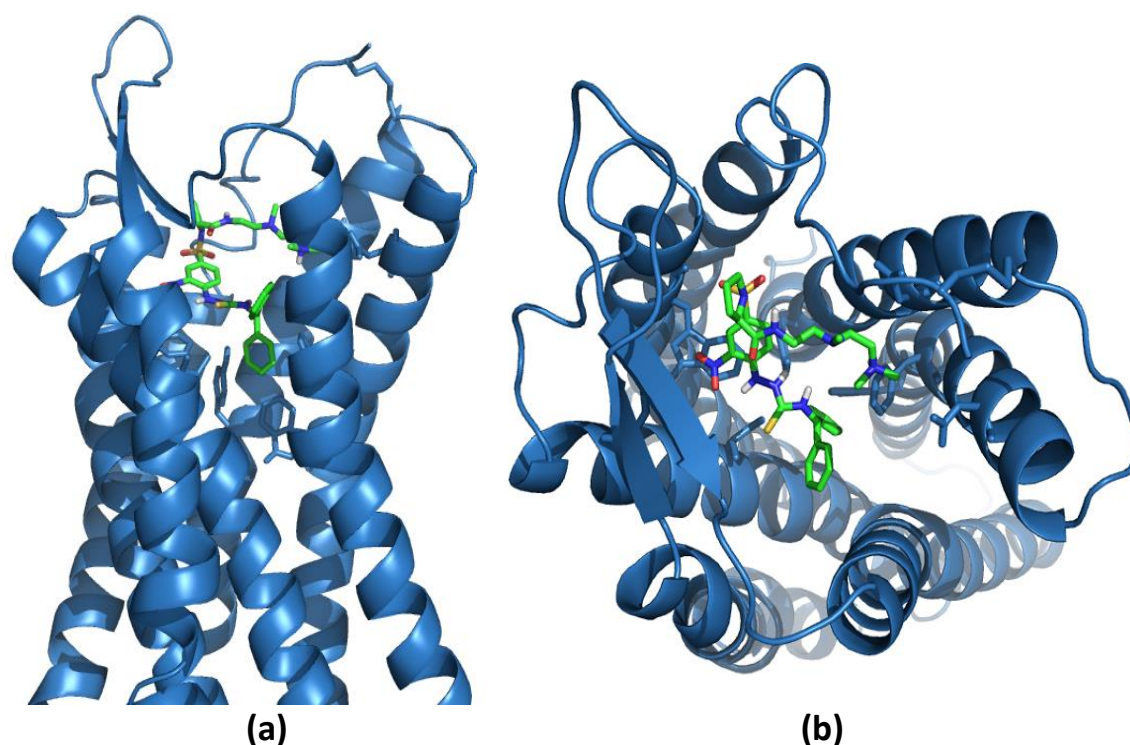
Figure 5.5. **a)** Lateral view of the orthosteric binding pocket of the human B2R with antagonist Bradyzide bound. **b)** Same as **a)** viewed from the extracellular side (top).

5.4 **Antagonist of Bradykinin** B2R**s**

5.4.1 Fasitibant

Fasitibant (MEN16132) (**1,** in Figure 5.1) is a potent antagonist of the B2R with a K_i of 0.09 nM [30]. The structure of Fasitibant bound to the model receptor from the present docking study is shown in Figure 5.5. In accordance with previous docking studies [30], the quaternary terminal amine interacts with Asp[266] and Asp[284] at the mouth of the receptor, whereas the hydrophobic 8-benzyloxy-2-methyl-quinoline moiety gets deep into the orthosteric pocket. The involvement of the two aspartates in ligand binding is supported by site directed mutagenesis analysis [36]. Inspection of Figure 5.6 shows the tetrahydropyranyl moiety sitting in a hydrophobic region on top of Trp[86], with a hydrogen bond

interaction between the heterocycle oxygen and the indole nitrogen of the triptophan side chain. This result is consistent with the fact that binding of Fasitibant to the Trp[86]Ala mutant is about 1200 times lower [30] and can be explained on the basis that the quadrupole-quadrupole interaction between the two rings and a hydrogen bond are lost in the mutant. Analyzing other functional groups of the ligand, one of the oxygens of the sulfonamide appears forming a hydrogen bond with Arg[169]; one of the chlorines of the dichlorobenzyloxy moiety forms a polar interaction with Arg[169] and Asn[107], whereas the other sits in a hydrophobic environment flanked by Phe[259], Phe[292] and possibly with Tyr[295].

Unfortunately, there is no mutagenesis data available about the involvement of Arg[169] or Phe[259] in the affinity of the ligand to give support to these interactions. At the molecule ending, the quinoline moiety sits in the vicinity of Ile[110], Trp[256] and Tyr[295] interacting with the latter two through quadrupole-quadrupole interactions. Moreover, the interaction with Tyr[295] is enhanced by the presence of a hydrogen bond between the quinoline amine group of the forms and the hydroxyl group of Tyr[295]. These results are consistent with diverse site directed mutagenesis data available. Thus, the mutant Ile[100]Ala reduces 300 times the affinity of Fasitibant and furthermore, the mutation Tyr[295]Phe reduces the affinity 100 times, whereas the mutation Tyr[295]Ala 755 times, suggesting a dual role as aromatic/hydrophobic residue and as hydrogen bond donor/acceptor for Tyr[295] [30].
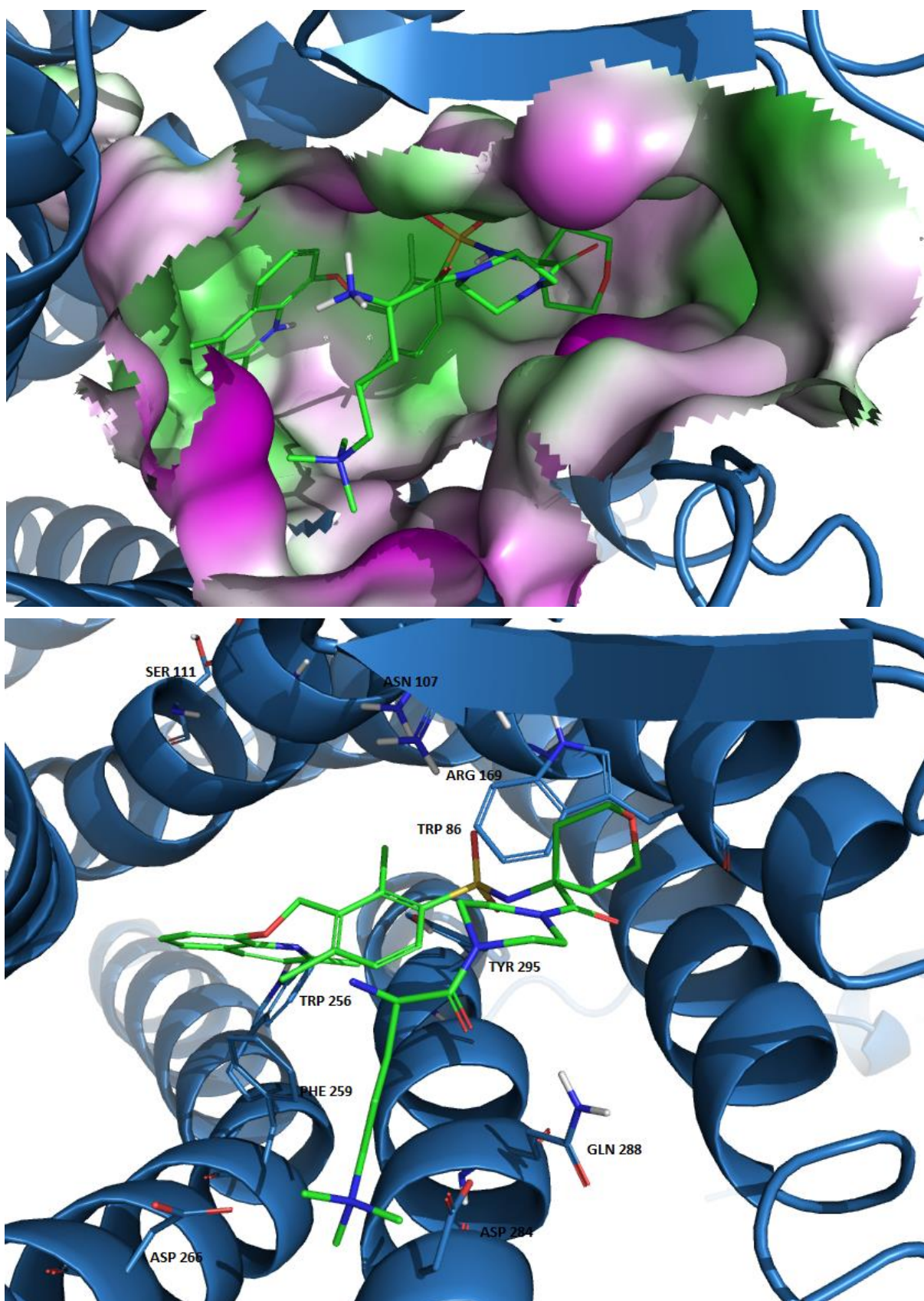
Figure 5.6. Pictorial view of B2R antagonist Fasitibant bound to the B2R.

5.4.2 FR173657

This Fujisawa compound (**2** in Figure 5.1) is a potent B2R antagonist with an IC_{50} of 1.4 nM at the human receptor [21]. This compound is the result of an extensive medicinal chemistry program aimed at designing B2R selective antagonists from a hit found in a screening program for angiotensin II AT_1 antagonists. As can be seen the compound shares with Fasitibant the 8-benzyloxy-2-methyl-quinoline moiety and consequently, it could be thought that the compound binds in a similar manner, however site directed mutagenesis studies suggest that the bound conformation is different. Thus, for example the affinity of FR173657 is not affected by the mutation of Asp^{266} and Asp^{284} [36] or Ser^{111} and Trp^{256} [37], suggesting that these residues do not act as anchoring points in the bound conformation of the compound. In contrast, Trp^{86} and Tyr^{295} must be actively involved since their mutation to alanine decreases the affinity of the compound about 500 times [30].

Bearing these results in mind, several docking attempts were carried out, obtaining diverse alternative poses. Analysis of the results in view of the site directed mutagenesis results available suggests that the ligand binds according to the ligand-receptor complex shown in Figure 5.7. Specifically, the molecule adopts a L-shape with the dichlorobenzyloxyl moiety found on top of Trp^{86}, nicely interacting through a parallel π-π stacking in such a way that both chlorines in addition to sit in a hydrophobic environment, interact through a hydrogen bond with Asn^{107} and Gln^{288} respectively.

Unfortunately, there are not mutagenesis results of the role of these two residues, but structure activity studies of different analogs of FR173657 suggest that replacing the chlorines for methyl groups decreases the affinity about five times, supporting the existence of the hydrogen bond interaction [38]. The position of the dichlorobenzyloxyl permits to direct the quinoline moiety -similar as it is found in fasitibant- in the vicinity of residues Ile^{110} and Phe^{256} whose mutation is known to affect significantly ligand binding. An additional piece of information to support the position of the quinoline moiety comes from a CoMFA analysis described in reference 36.
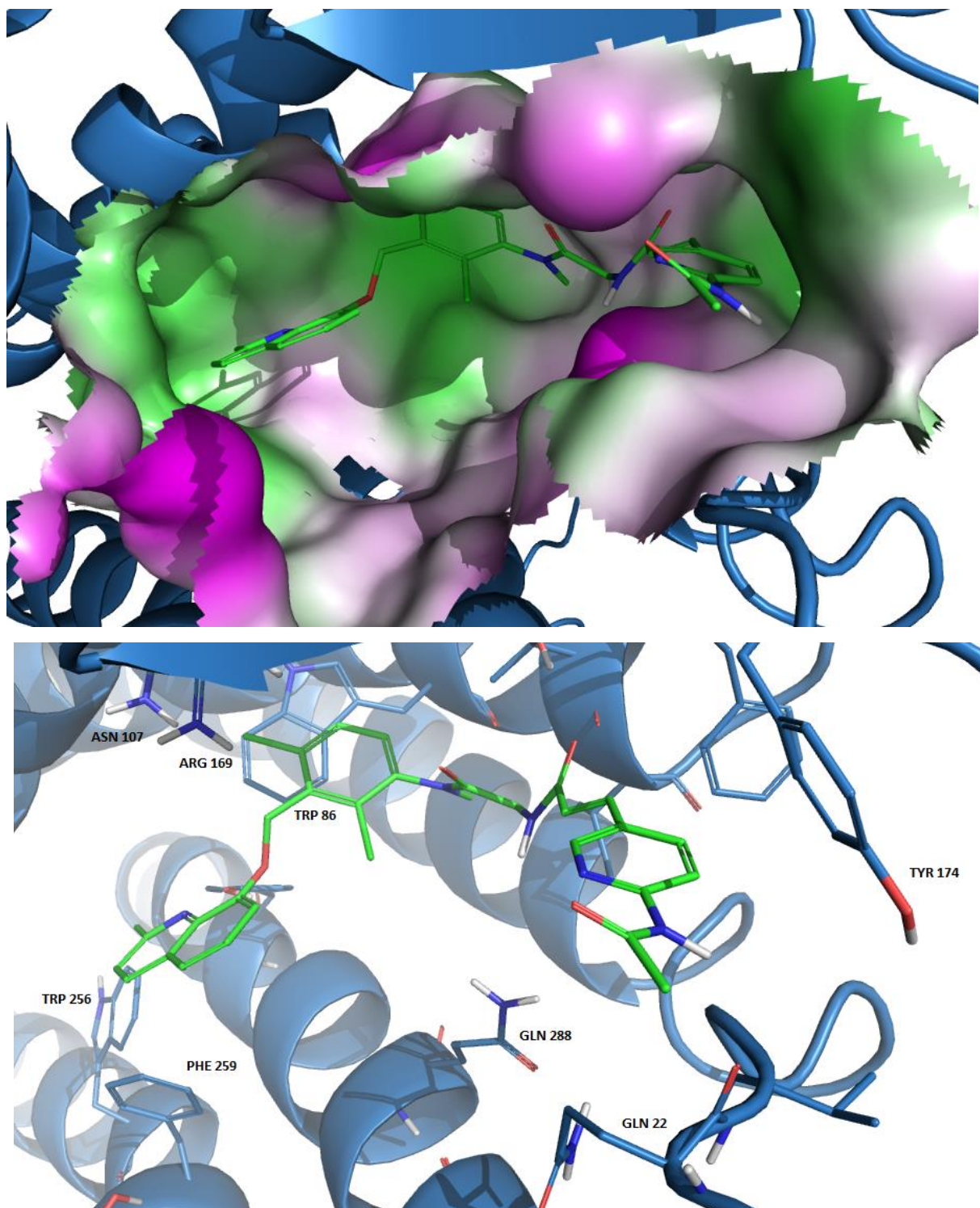
Figure 5.7: Pictorial view of B2R antagonist FR173657 bound to the B2R.

Specifically, the authors proposed constraints for favorable groups to improve the affinity of the

ligand by extension of the molecule and these fit well with the positions of residues Trp[256], Phe[259] and

Asn[198] in the present ligand-receptor complex. On the other side of the molecule, structure activity studies

of diverse analogs suggest the importance of this part of the molecule for obtaining good antagonists at

the human B2R. Analysis of ligand-receptor complex shown in Figure 5.7 suggests that the amide groups

interact with polar residues like Glu[24] or Thr[89] whereas the pyridine ring interacts with Phe[94] and Tyr[174]

through quadrupole-quadrupole interactions. These results agree well with the structure-activity studies

carried out on these compound series. Specifically, the introduction of a phenylurea moiety induces an

increase in the affinity one order of magnitude in guinea pig ileum membrane preparations and nearly

two orders in A-431 cells that express the human receptor due to the favorable interaction with the side

chains of Phe[94] and Tyr[174].

5.4.3 Anatibant

Previously known as LF 16-0687 (**3** in Figure 5.1), Anatibant is a potent B2R antagonist of Fournier

with an IC_{50} of 0.67 nM at the human receptor [23]. The compound shares the 8-benzyloxy-2-methyl-

quinoline moiety with the two antagonists described above, and in this case linked to a pyrrolidine

sulfonamide with a 4-amidinofenil moiety as charged terminal group. Analysis of the different poses

found in our docking studies in view of the site directed mutagenesis results available suggest the

molecule sits inside the orthosteric site as shown in Figure 5.8. As can be seen the bound conformation

of Anatibant shows similarities with that of Fasitibant. Indeed, the common 8-dichlorobenzyloxy-2-

methyl-quinoline substructure sits in a similar manner inside the receptor. However the stereochemical

differences on the other side of the molecules force them to exhibit differential poses as discussed below.
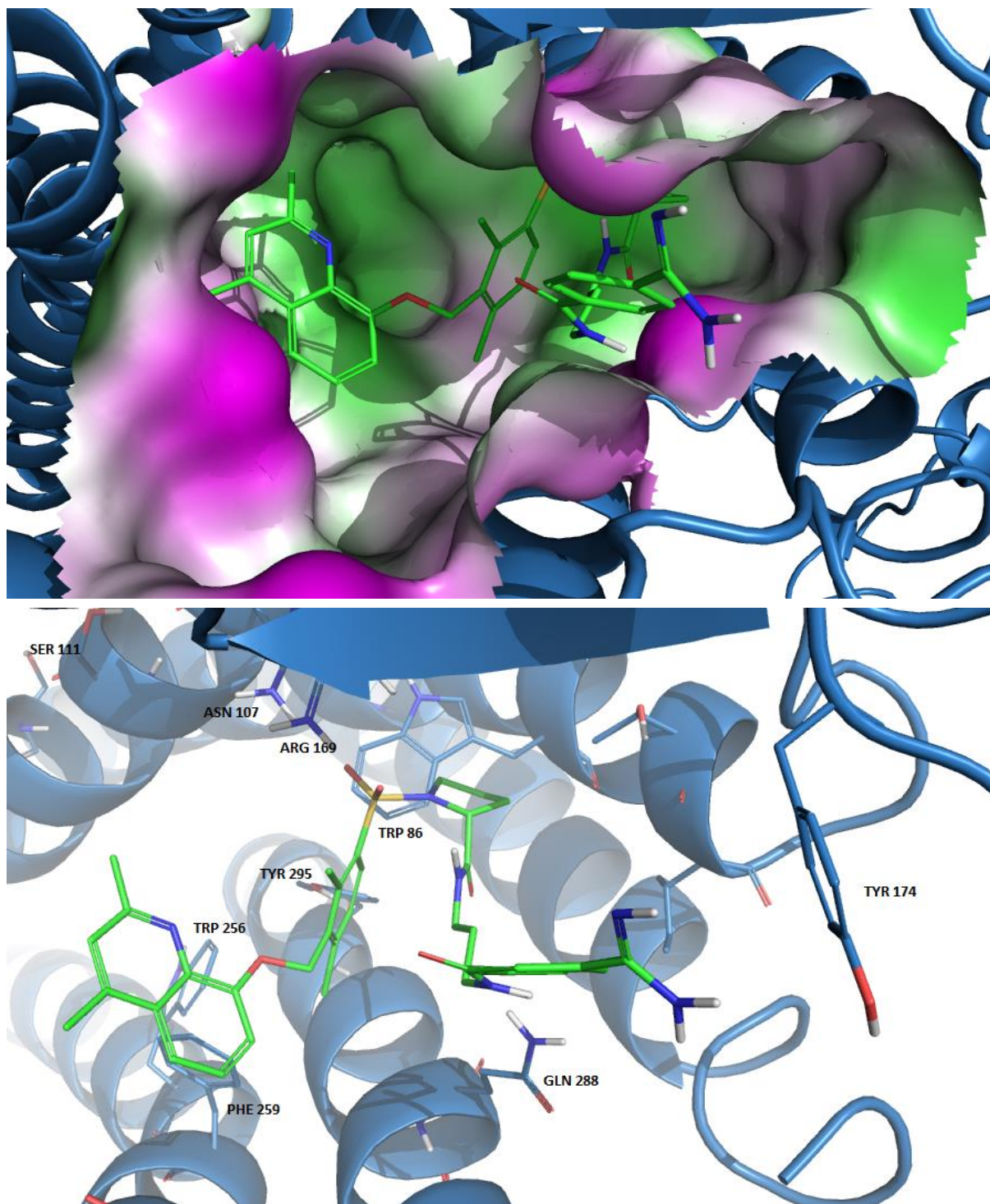
Figure 5.8: Pictorial view of B2R antagonist Anatibant bound to the B2R.

The dichlorobenzyloxyl moiety of Anatibant binds in the same region of the receptor as Fasitibant, consequently the other moieties attached to it and shared between the two molecules, including the quinoline and the sulfonamide access similar regions of the receptor. Specifically, the sulfonyl groups exhibit hydrogen bonds with Asn[107] and Arg[169], respectively. Moreover, the side chain of the latter shows an additional hydrogen bond with one of the chlorides of the dichlorophenoxyl moiety. On the other hand, the quinoline moiety like in Fasitibant sits close to Tyr[295] and interacts through a hydrogen bond and the quadrupole-quadrupole interaction between the two aromatic rings. Moreover, our docking studies also show the quinolone moiety interacting with Ile[110], Trp[256] and Phe[259]. These results are supported by mutagenesis studies, since the mutation of Tyr[295] to Phe[295] reduces the affinity one order of magnitude and to two orders of magnitude when mutated to Ala[295] [39]. Similarly, the mutation of Trp[256] to Ala[256] decreases the affinity one order of magnitude. Moreover, this idea also explains the observed effect of the mutation Asn[297]Ala in close contact with Trp[256] [39].

There are no results available on the mutation of Ile[110] or Phe[259] on the affinity of the ligand, but according to the present modeling study it is expected one order of magnitude decrease as shown in Fasitibant [30]. On the other side of the molecule Anatibant and Fasitibant exhibit differential stereochemical features that force them to bind in a differential way. This idea is supported by mutagenesis studies, since the mutation of Asp[266] or Asp[284] to alanine decreases dramatically the binding affinity of Fasitibant but does not alter that of Anatibant [36]. As can be seen in Figure 5.8 the pyrrolidine ring sits perpendicular to Trp[86] interacting through a quadrupole-quadrupole interaction, being its position reinforced by a hydrogen bond with the side chain of Arg[169]. The ring plays a fundamental role in distributing the rest of the molecule and this justifies the dramatic effect observed in the affinity of the compound when the residue is mutated to alanine [40]. From the pyrrolidine, the molecule extends towards the mouth of the receptor with the aminoiminophenyl moiety surrounded by polar residues including, Tyr[174] of the second extracellular loop or Gln[22] located at the N-terminus. Moreover, there are diverse polar residues that interact with the amide groups of the molecule, including Glu[24], Gln[33] and Gln[288]. The importance of the latter has been shown by mutagenesis studies [39].

5.4.4 WIN64338

This was the first non-peptide antagonist disclosed in the literature with a Ki of 64 nM in human IMR90 fibroblasts [20]. The antagonist was designed using a simple pharmacophore defined by two charged groups separated by a distance of about 10 Å – mimicking the distance between the two terminal arginines of BK in its bioactive conformation- linked by lipophilic groups [20]. From a hit obtained, subsequent medicinal chemistry efforts yielded the compound WIN64338 (**4** in Figure 5.1). As can be seen the compound exhibits two charged groups at both ends surrounded by bulky hydrophobic groups, together with a naphtyl moiety. The results of our docking study of the compound onto the B2R model is shown in Figure 5.9. WIN64338 is much shorter than the rest of the compounds described so far and is consequently, expected to cover a smaller region of the binding pocket. Although there are no reports on directed mutagenesis studies conducted with the compound, structure activity studies underline the importance of the two charges for high affinity.

As can be seen, the positively charged phosphine group nicely sits surrounded by two negatively charged residues Asp[266] and Asp[284], although the model suggest that is the latter that is involved in a charge-charge interaction. The other charge in the model interacts with the backbone of Cys[184]. In regard to the bulky cyclohexyl groups, one sits in the proximity of Trp[86], whereas the other points toward the solvent in a region surrounded by Ala[183] and Phe[94] that provide a hydrophobic environment. Finally, the naphtyl group sits in the hydrophobic pocket in the proximity of Trp[256], Phe[259] and Ile[110] similarly as does the quinoline moiety in the previous antagonists described.
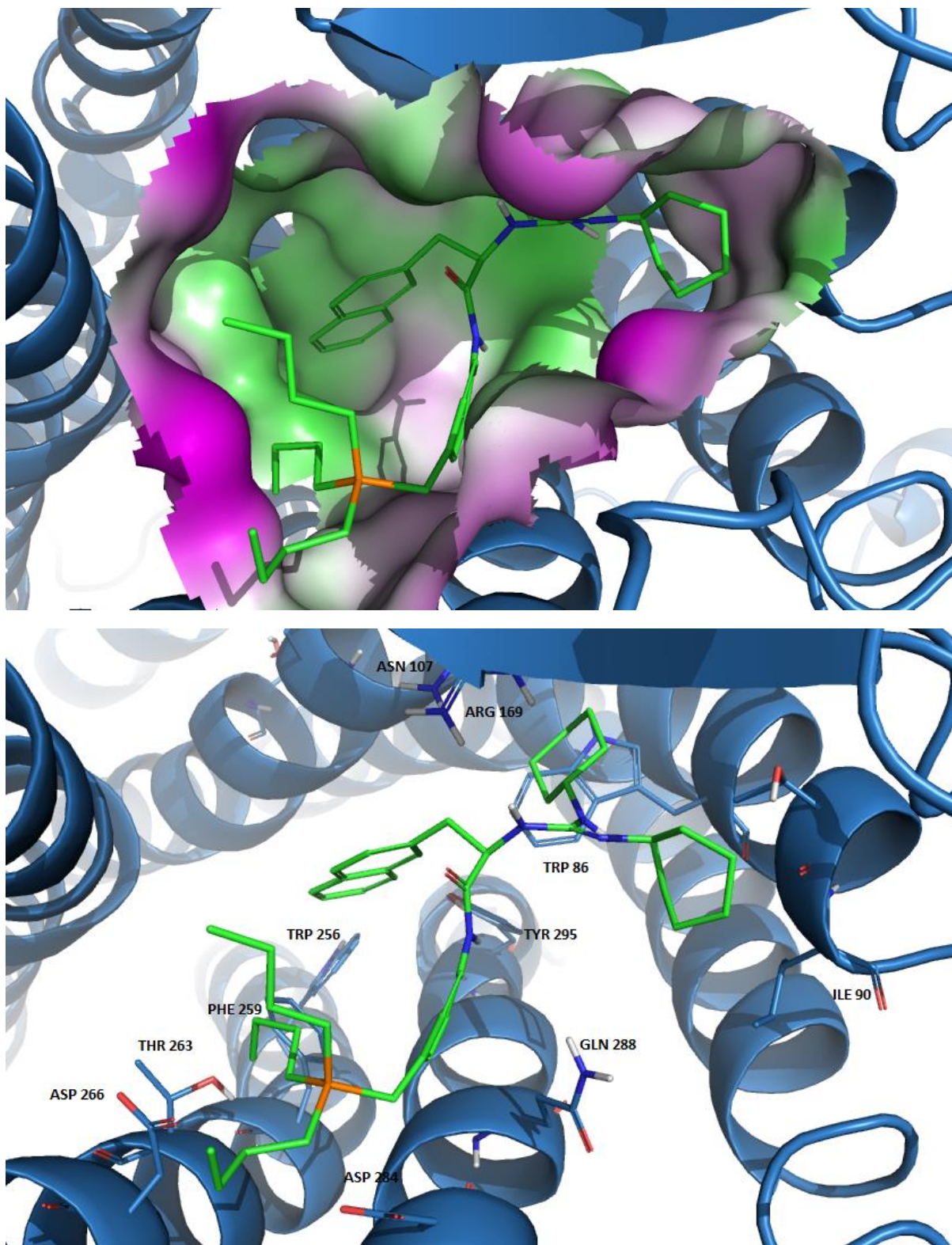
Figure 5.9: Pictorial view of B2R antagonist WIN64338 bound to the B2R.

5.4.5 Bradyzide

Bradyzide (**5**, in Figure 5.1) is a potent, rat-selective B_2 antagonist that causes a long lasting reversal of inflammatory hyperalgesia [22]. The compound is the result of a lead optimization from a hit discovered by random screening. Interestingly, the compound exhibits high affinity for the rat B2R (0.5 nM) expressed in NG108-15 cell membranes, but much lower in human BK B_2 expressed in Cos-7 cells (772 nM). The result of the docking study of this compound onto the B2R is shown in Figure 5.10. There are no mutagenesis studies available that can help to understand the effect of diverse mutations on the binding of Bradyzide to the B2R, but there is information about the effect of diverse chemical substitutions on the molecule [41] that can be analyzed through view of the complex model.

Thus, the terminal charged amino nitrogen located at the end of the diaminoalkyl chain in Bradyzide sits at the mouth of the receptor, interacting with Asp^{266} in TM6 and Asp^{284} in TM7, although the model actually suggests that is the latter that is involved in a charge-charge interaction. Structure-activity studies support this result by since elimination of this chain decreases the affinity about 100 times [41]. Furthermore, these studies also point to the basicity of the nitrogen as important feature to get better affinities, confirming the role of the terminal nitrogen in a charge-charge interaction. In the model the sulfone group provides an anchoring point to ligand through a hydrogen bond with Gln^{33} in TM1. Further down the ligand, the aromatic ring of the nitrobenzene moiety interacts with Trp^{86} in TM2 with the nitro group interacting with Asn^{107} via a hydrogen bond. The sulfur of the thiosemicarbazyl moiety establishes a hydrogen bond with Asn^{107} as has been previously suggested and though a hydrophobic interaction with Ile^{110}; finally, the phenyl groups close to the thiosemicarbazyl moiety sit well in the aromatic region Thr^{256}, Phe^{259}, Tyr^{295}.
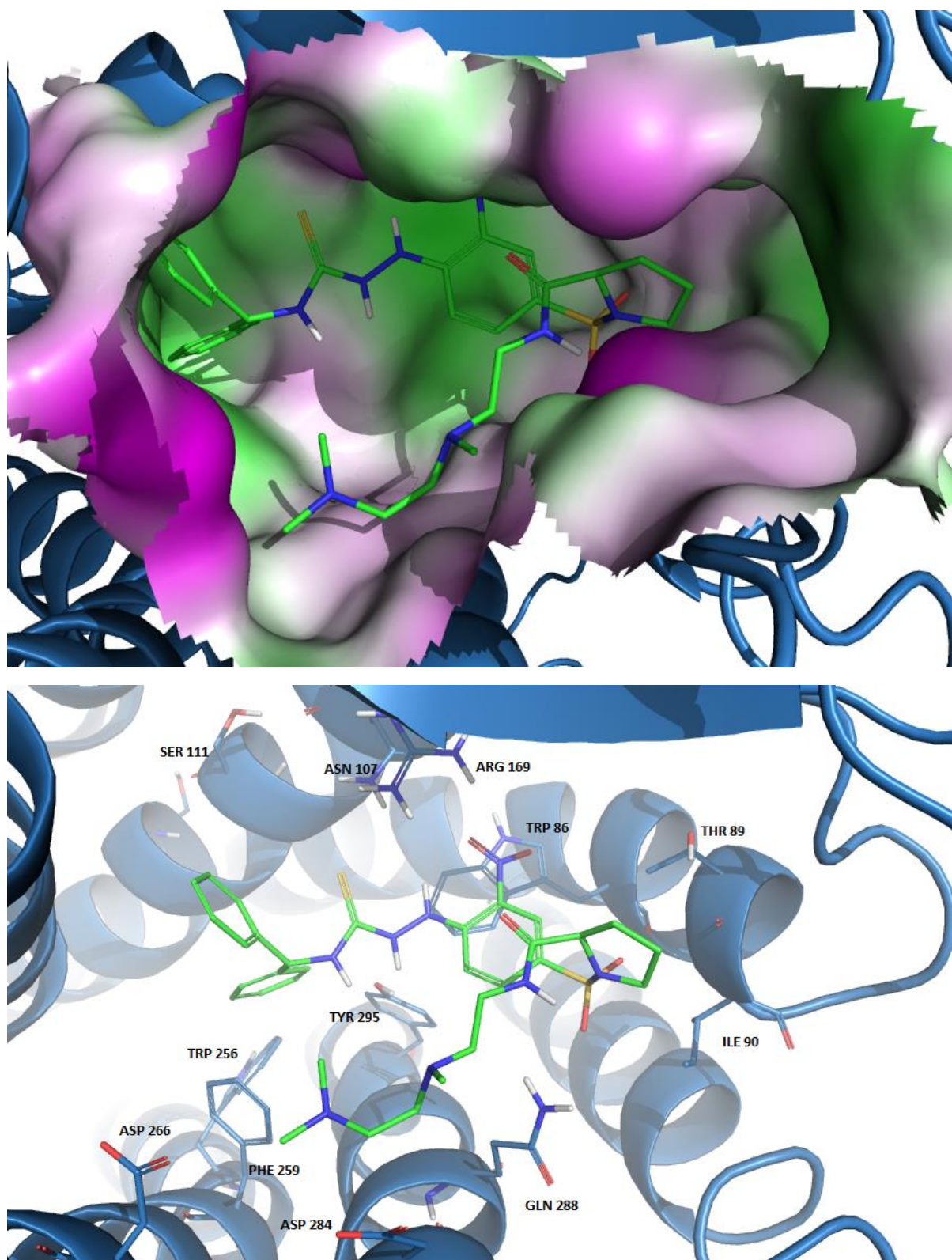
Figure 5.10: Pictorial view of B2R antagonist Bradyzide bound to the B2R.

5.4.6 CHEMBL442294

Designed by Novartis, ccompound (**6**, in Figure 5.1) was the most active compound of a series of benzodiazepines designed to mimic the β–turn adopted by BK in its bioactive form [25]. This peptidomimetic exhibits a binding affinity for the B2R in the micromolar range, result that is consistent with the low affinity exhibited by a series of cyclic peptides designed to mimic the C-terminus of BK [13, 14]. These results indicate that mimicking the C-terminus of the peptide is necessary condition to get good binding affinity, but not sufficient. Docking of the molecule into the receptor generated diverse poses that were analyzed in order to understand the features of the ligand-receptor interaction Figure 5.11 shows the ligand-receptor complex with the best docking score.

As can be seen the ligand adopts a pose covering a number of residues identified in the docking in previous ligands. In this case, the guanidinium moiety binds to Asp[284] and Gln[288] in such a way that allows the interaction of an aromatic ring with Trp[86] and another with Trp[256] and Phe[259] in TM6. In addition, the ligand exhibits a hydrogen bond between the carbonyl of the diazepine ring and Arg[169].

5.4.7 JSM10292

This compound (7, in Figure 5.1) is the result of an optimization process aimed to find low molecular mass non-peptide B2 antagonists based on the structures of previously disclosed compounds [26]. Specifically, inspection of the structure of diverse antagonist including FR173657 [21], compound 8d [38], Anatibant [23] and Fasitibant [30] shows that these molecules share a 8-benzyloxy-2-methyl-quinoline as common structural feature. Thus, the authors used this moiety as starting structure to follow a medicinal chemistry approach, leading to the compound JSM10292 that exhibits an IC50 of 8.7 nM in the human B2R, expressed in HEK293 cells [26]. Although the molecule shares a common substructure with those molecules used for its design, JSM10292 binds in different way. Analysis of different poses obtained during the docking process and analyzed according with the mutagenesis results available [42], it was selected as putative bound conformation the one shown in Figure 5.12.
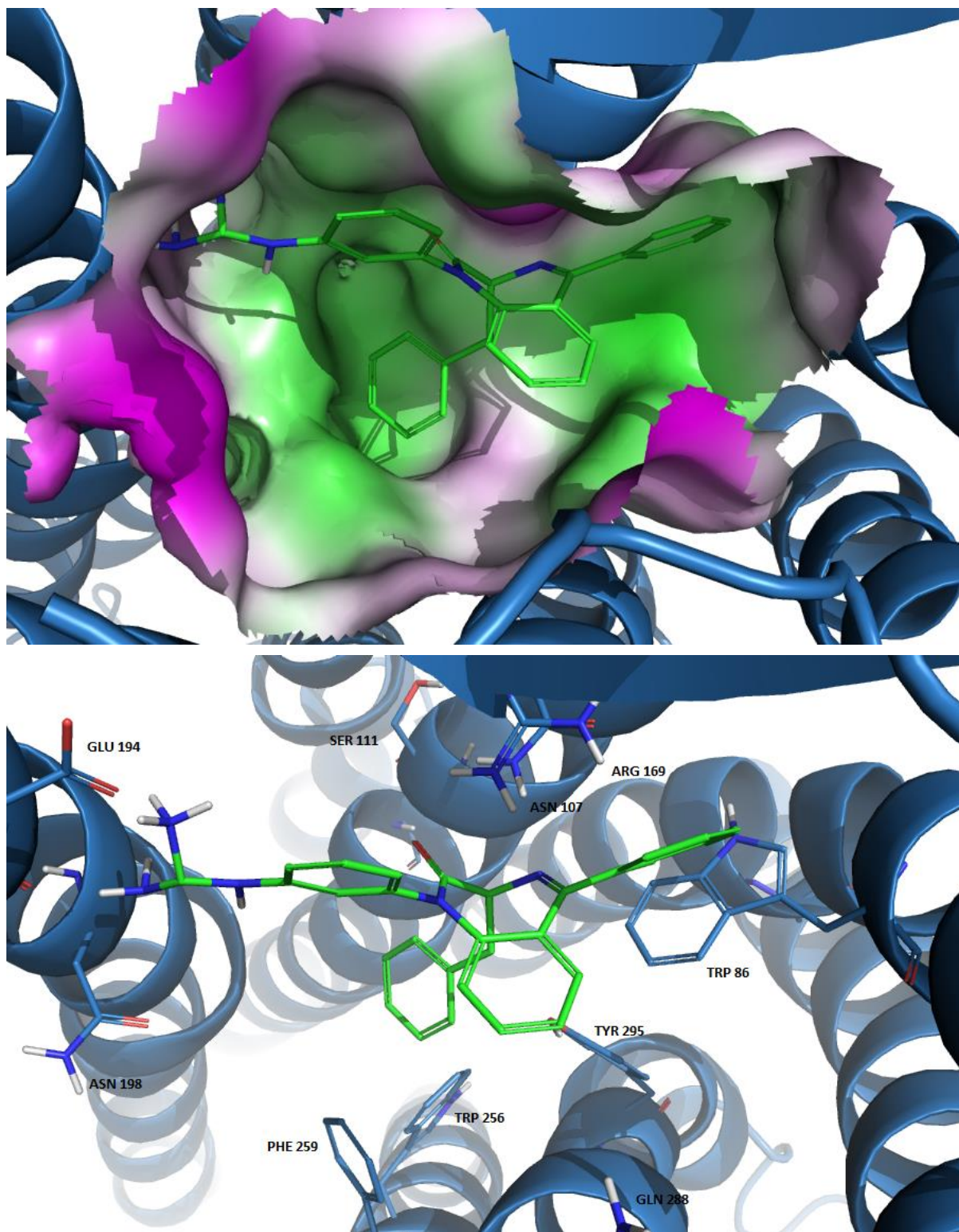
Figure 5.11: Pictorial view of B2R antagonist *CHEMBL442294* bound to the B2R.
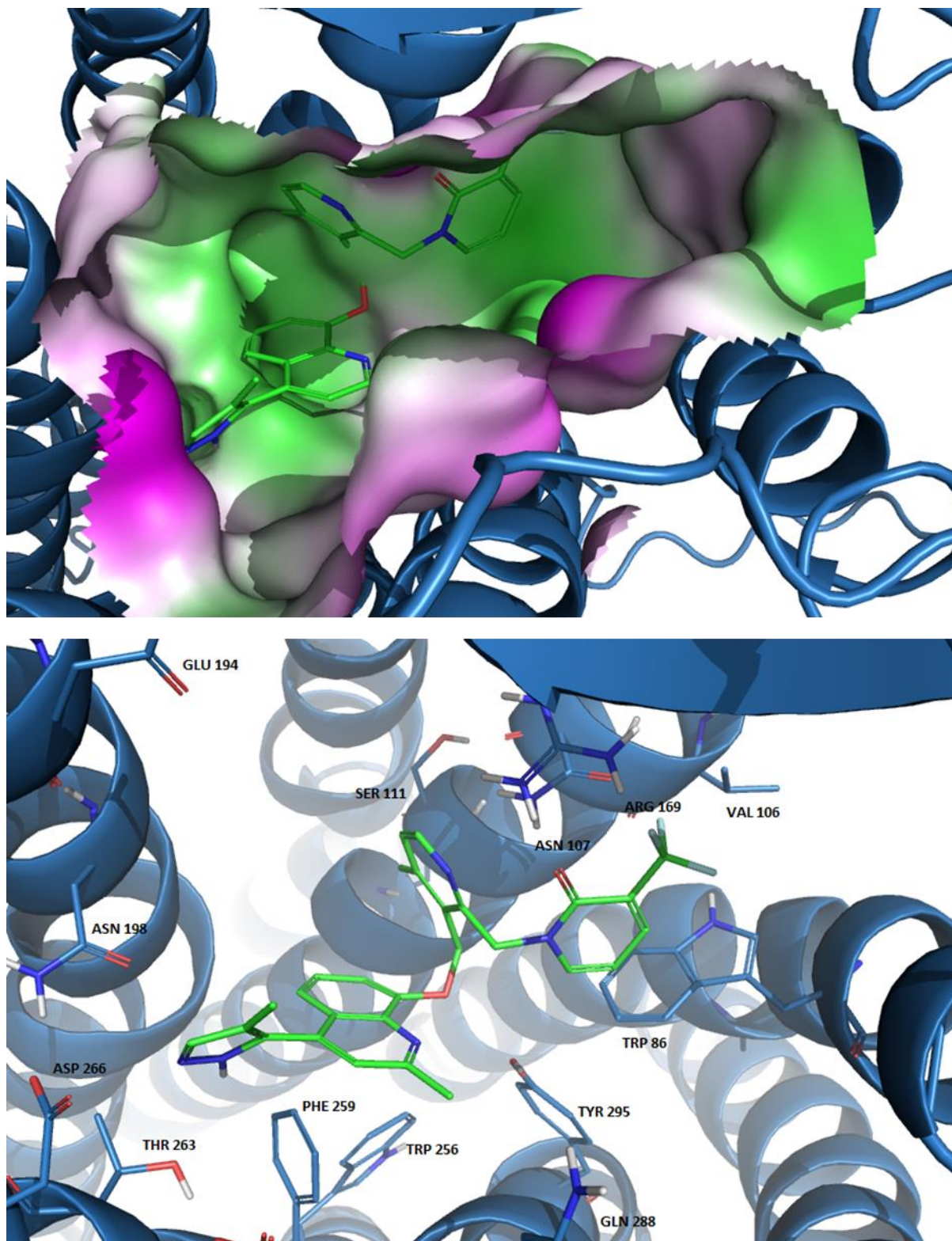
Figure 5.12: Pictorial view of B2R antagonist JSM10292 bound to the B2R.

Indeed, the pyrazole substituent to the quinoline group sits close to Phe259 with one of the pyrazole nitrogens acting as proton acceptor in a hydrogen bond interaction with Thr263. Although there are not mutagenesis results on the role of the latter, the results underline the important role of Phe259 acting as anchoring point of the bound conformation. The nitrogen of the quinolone ring and the oxygen atom of the benzyloxy moiety interact via a hydrogen bond with Tyr[295]; the nitrogen of the pyridine interacts with the side chain of Arg[169], whereas the aromatic ring and its methyl group sit close to Ile[110]; the carbonyl oxygen of the trifluoropyridone moiety also interacts with Arg[169] and the pyridone ring interacts with Trp[86] whereas the trifluromethyl group interacts with Asn[107]. The bound conformation of JSM10292 found in the present study differs slightly of the one described in reference 40. The two models actually differ in the conformation of the ligands. In our model the ligand gets an extended conformation, whereas in their model the ligand adopts a conformation the trifluoropyridone ring folds back towards the pyridine ring. As result of the different conformation, the trifluromoiety interacts in the present model with Asn[107] whereas in the other model gets close to Ser[111]. In fact, there are not mutagenesis results of the mutation of serine to alanine; however the replacement to lysine provokes a great loss of affinity [42]. However, due to the size of the lysine side chain in both models a steric hindrance is generated to perturb the binding of the ligand and consequently does not allow discriminating between the two models.

5.5 Designing pharmacophore of BKRB2 antagonists

Comparison of the complex ligand-receptor of the different antagonists used for the present study suggests the definition of a pharmacophore that explains the observed structure-activity. It consists of five pharmacophoric points that not all the ligands studied in the present study fulfill. The pharmacophore proposed is shown in Figure 5.12 and it is defined considering geometrical constraints on ligand moieties. Thus, point 1 is a proton donor center that will interact with Asp[84] and/or Asp[266]; point 2 is either a proton accepting/proton donor center that will interact with either Gln[32] or Gln[288]; point 3 is a hydrophobic ring

that will interact with Trp[86]; point 4 is a proton accepting center that will interact with Asn[107] and/or Arg[169]; point 5 is an aromatic moiety that will interact with Trp[256], Phe[259] and Tyr[295].

Thus, Fasitibant fulfils point 1 by means of the terminal amine; point 3 by means of the dichlorophenoxyl moiety; point 4 by means of one of the sulfonyl oxygens as well as one of the chlorines and point 5 by means of the quinoline moiety. In the case of FR173657 point 2 is fulfilled by means of one of the amide groups; point 3 is fulfilled by means of the dichlorophenoxyl moiety that also fulfills point 4 by means of one of the chlorine atoms; point 5 is fulfilled by means of the quinoline moiety. In the case of Anatibant point 2 is fulfilled by means of the pyrrodiline amide; point 3 by means of the pyrrolidine ring; point 4 by means of the sulfonyl groups and point 5 by means of the quinoline moiety. In the case of WIN64338 point 1 of the pharmacophore is fulfilled by the phosphine group; point 3 is fulfilled by one of the cyclohexanes; point 4 by the amide carbonyl and point 5 by means of the naphtyl moiety. In Bradyzide point 1 is fulfilled by the terminal amine; point 2 is fulfilled by one of the sulphonyl oxygens, although in this case the interaction is more likely with Gln[33] in the vicinity of Gln[288]; point 3 is fulfilled by means of the nitrobenzene moiety and point 5 by means of the phenyl terminal groups. For CHEMBL442294 point 1 is fulfilled by the guanidinium moiety; point 3 by means of the phenyl substituent of the benzodiazepine scaffold and point 5 by means of the benzyl moiety. In the case of JSM10292 point 1 is fulfilled by means of the pyrazole moiety; point 3 is fulfilled by means of the pyrazone moiety that also fulfils point 4 by means of the carbonyl group. Finally, point 5 is fulfilled by means of the quinoline moiety.
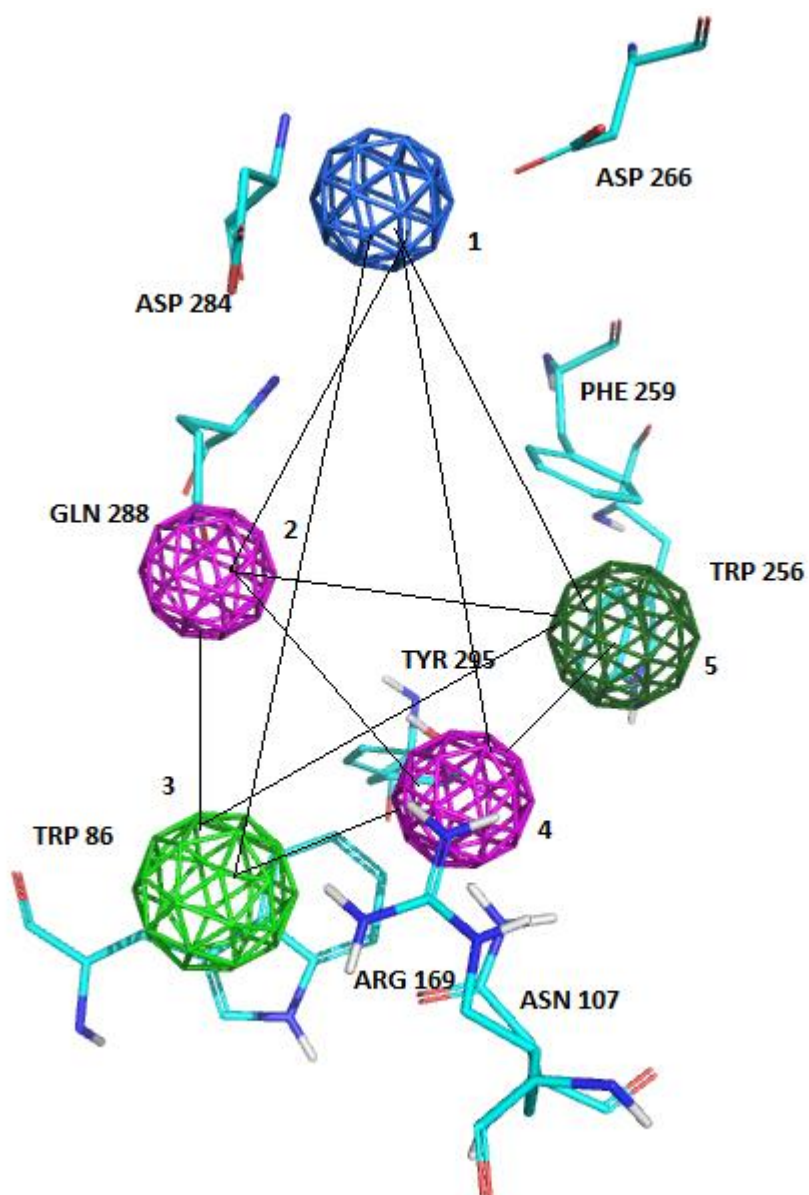
Figure 5.13: Proposed pharmacophore for the BK B2 antagonism. Distance between pharmacophoric points are: d(1,2)=8.8 A; d(1,3)=17.4; d(1,4)=15.5; d(1,5)=12.2; d(2,3)=12.1; d(2,4)=12.2; d(2,5)=11.6; d(3,4)=7.2; d(3,5)=8.8; d(4,5)=10.3

Present pharmacophore is in agreement for most part with previously B2R pharmacophore described in the literature. Specifically, the simple pharmacophore proposed by Salvino et al. [43]

consisting in two charges separated 10 Å that gave rise to the discovery of WIN64338, includes points 1 and 4 of the present pharmacophore. On the other hand, in order to mimic the C-terminus of BK after a thorough exploration of the conformational space of the five bradykinin analogues, it was suggested a few years ago a partial pharmacophore for BK antagonism [44, 45]. This includes an ionizable positive charge, a hydrophobic group and an aromatic/hydrophobic group in a specific spatial arrangement of 4.5-7.5 Å, 5.5-8-5 Å and 8.5Å. This pharmacophore is included in the more general pharmacophore described in the present work. Distances are not directly comparable since in the present pharmacophore distances are defined on the side chains of receptor residues and not on the chemical moieties. Thus, the ionizable positive charge is the moiety facing point 2 of the present pharmacophore; the hydrophobic group is the one facing point 5 of the present pharmacophore and the aromatic ring is the moiety interacting with point 3 in the present pharmacophore.

Proof of concept

The pharmacophore described above was later used in the discovery of new structures with antagonistic activity for the B2 bradykinin receptor by virtual screening. For this purpose a search was made in few databases of 3D structures of compounds including the Available Chemical Directory (ACD), the Derwent World Drug Index, the National Cancer Institute (NCI) and Maybridge for approximately 500,000 compounds. We carried out separated searches for compounds fulfilling five, four and three pharmacophoric points yielding a large set of compounds. A selected group of them were purchased and tested for their B2 antagonistic activity.

Biological assays permitted to identify new highly diverse hits with structures that do not resemble those used for pharmacophore development. The success rate achieved to identify new hits with an antagonistic activity by virtual screening was approximately one third of the selected molecules, as previously found by other authors [46]. Table 1 shows the structures as well as the antagonistic activity to the human bradykinin B2R of a selected group of hits that are disclosed to give support to the pharmacophoric hypothesis developed in this work.
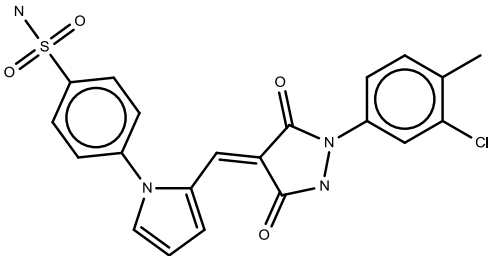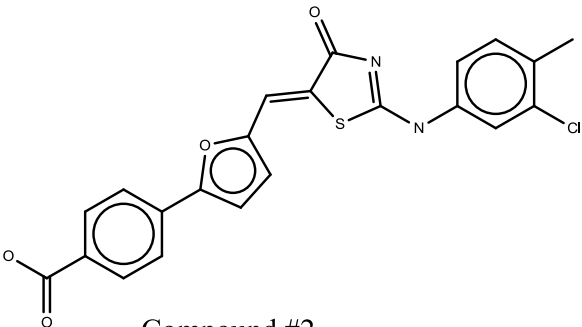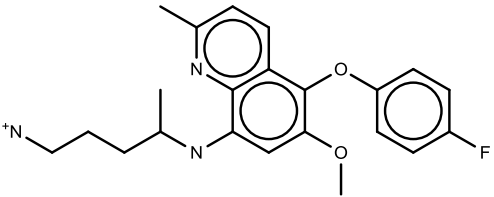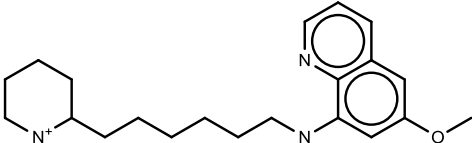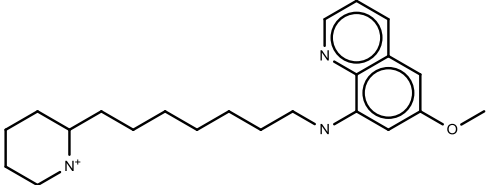
These molecules were docked onto the receptor model and inspected for fulfillment of the pharmacophore. Information regarding the number of pharmacophoric points fulfilled by each of the hits is also included in Figure 5.11. As an example, Figure 5.13 shows compound #1 bound to the B2R showing the fulfillment of the pharmacophoric points. Interestingly, the antagonistic activity observed experimentally correlates well with the number of pharmacophoric points fulfilled by these molecules.
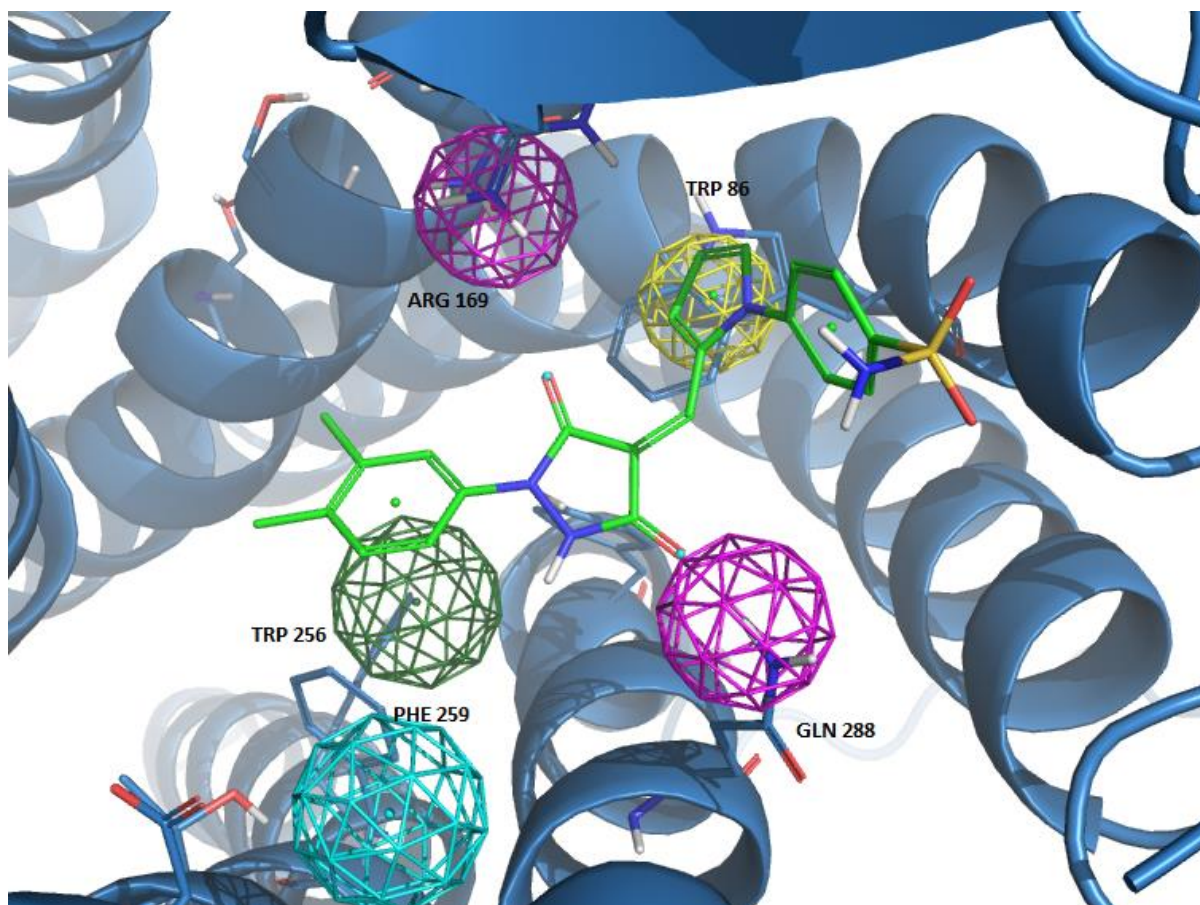
Figure 5.14: Pictorial view of compound #1 of Table 1 bound to the B2R with the pharmacophore points represented as spheres of different colors.

The chloromethybenzyl group interacts in the aromatic pocket Trp[256], Phe[259] and thus fulfilling pharmacophoric point 5, while the dioxopyrazollidine group fulfils point 2 via interactions with Gln[288] and the pyrrole group fulfills point 3 by interacting with Trp86. The hydrogens in benzylsulfonamide moiety do not position correctly for an interaction with point 1.

Table 5.1. Structures of the new hits discovered together with their antagonistic effect towards the human B2 bradykinin receptor and the number of pharmacophore points fulfilled.

Structure	Inhibition	Pharmacophore points
Compound #1	41% @ 10 μM	3
Compound #2	33% @ 10 μM	3
Compound #3	45% @ 50 μM	2
Compound #4	36% @ 50 μM	2
Compound #5	66% @ 50 μM	2

5.6 **Antagonistic selectivity of B2R over B1R**

Most of the antagonists studied in this chapter as well as those studied in chapter 4 are bulky compounds with high molecular weight covering a large volume of the binding site. These characteristics can be clearly seen and explained from the analysis of the receptor-ligand models developed in this work. These results permitted the development of the pharmacophoric hypothesis of B1R and B2R antagonism explained above. Comparison of these pharmacophores can be used in a step forward to understand ligand selectivity to the two bradykinin receptors.
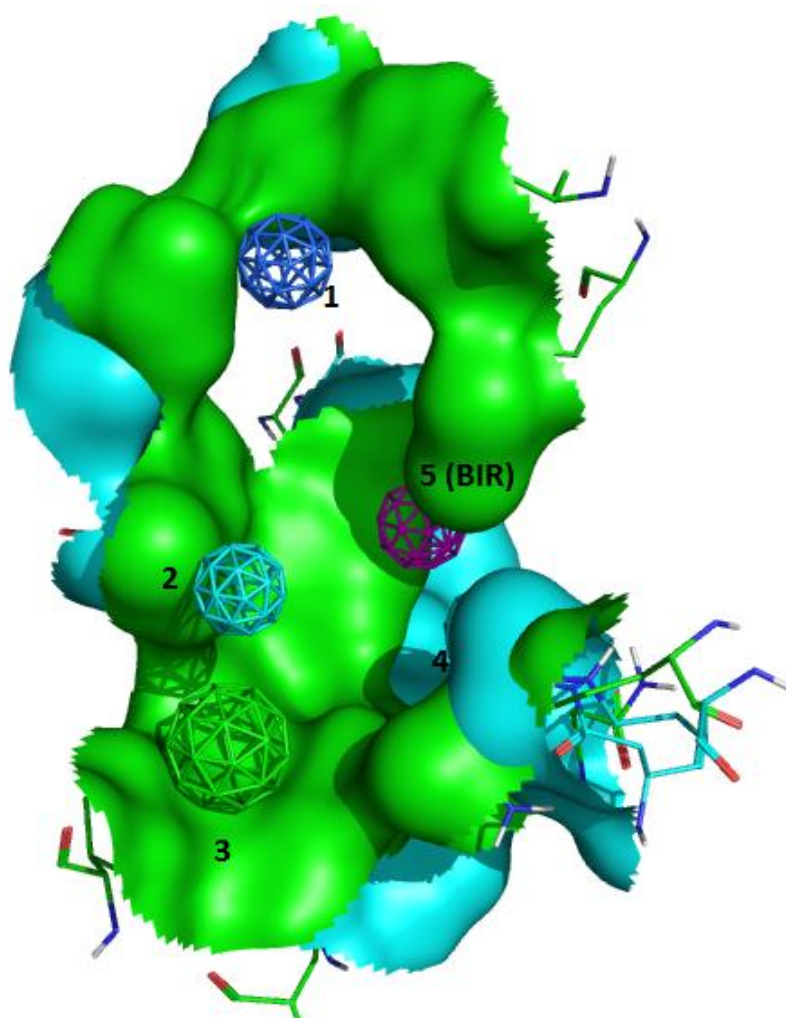
Figure 5.15: A superposition of the pharmacophores for B1R (in green) and B2R (in blue) showing similar pharmacophoric points 1-4.
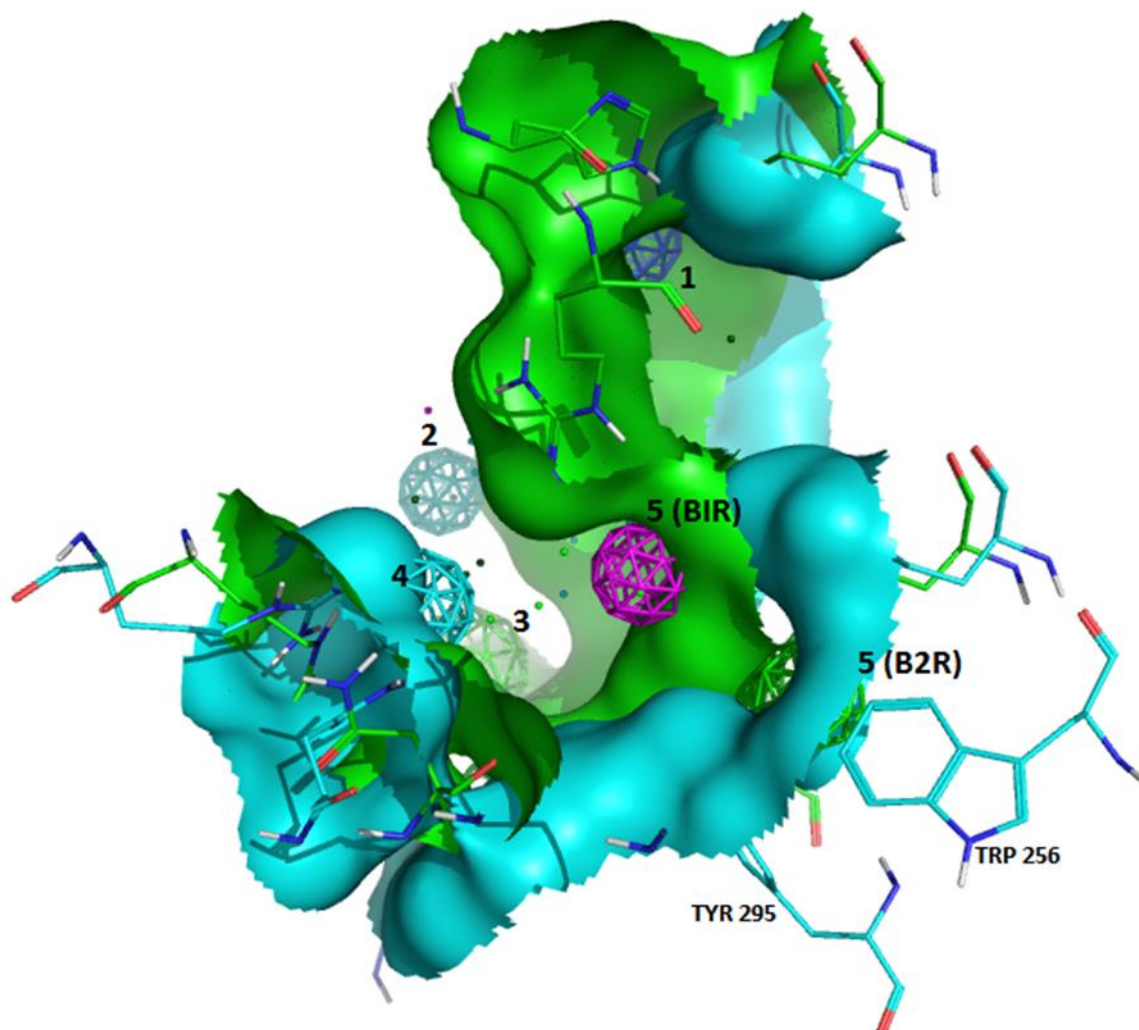
Figure 5.16: A superposition of the pharmacophores for B1R (in green) and B2R (in blue) highlighting the pharmacophoric points 5 which differs in B1R and B2R.

Certainly by looking at the proposed pharmacophores of the two receptors, it can be seen that the majority of the residues of the binding pocket involved and hence, the pharmacophores obtained are quite similar. In fact, pharmacophoric points 1-4 (an ionizable positive charge at point 1; a proton donor/ acceptor group at point 2; an aromatic/saturated ring at point 3 and another proton donor or acceptor group at point 4) are the same as shown in Figure 5.15. So, the main difference is found in pharmacophore point 5 (Figure 5.16) that involves a polar moiety in B1R and an aromatic ring in B2R. Moreover, point 5 in the B1 pharmacophore is located at a closer distance to point 1 than in the B2 pharmacophore, where point 5 in the located deep in the binding pocket encompassing the large cleft

made by aromatic residues (Figure 5.17). More Specifically, in the B1R point 5 is defined as a proton donor or acceptor group that interacts with Arg[202], Tyr[266] and the side chain of Asn[298] in TM5, TM6 and TM7 respectively, whereas in B2R this point is defined by an aromatic ring that interacts with residues of a hydrophobic cleft comprised by Trp[256], Phe[259] in TM6 and Tyr[295] in TM7.
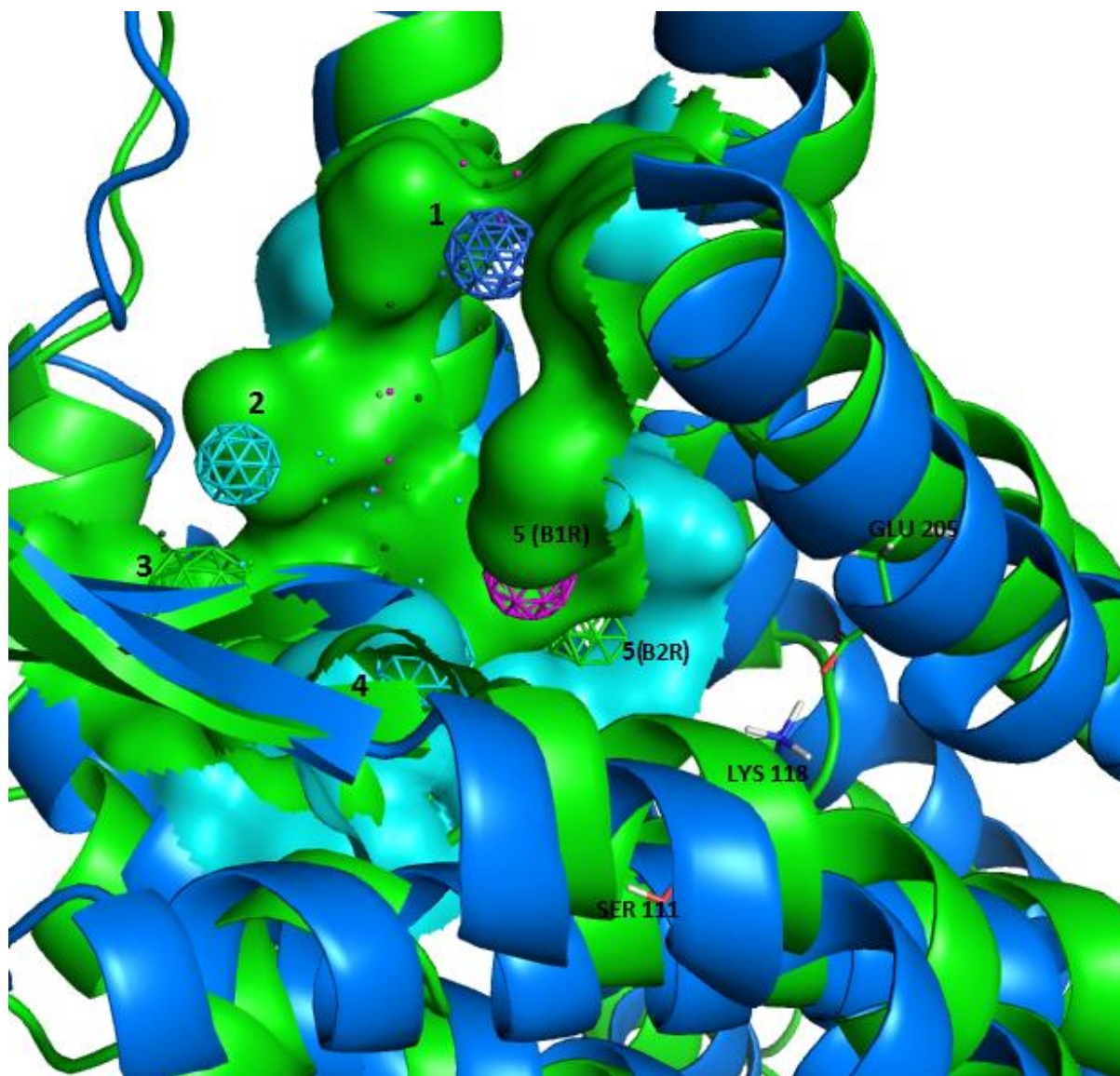

Figure 5.17: A superposition of the pharmacophores for B1R (in green) and B2R (in blue) as they appear in receptor.

The reason of the different location and nature of point 5 in the two pharmacophores can be attributed to a specific TM3 residue: Lys[118] in B1R and the corresponding Ser[111] in B2R. According to

the models described in this work the backbone of either of these residues is bonded to the backbone of a conserved Asn[114/107] in TM3, a prominent residue making up the pharmacophore point 4, in such a way that the sidechain of residue Lys[118] forms a hydrogen bond with the sidechain of residue Glu[205] in TM5, preventing the equivalent residues of point 5 of B2R in B1R from facing the binding pocket. In B2R the sidechain of Ser[111] is short and does not interact with TM5, thus making the area vacuous and allows the orientation of the aromatic rings of Trp[256], Phe[259] in TM6 and Tyr[295]   in TM7 towards the binding pocket. This explains the results of a selectivity study of peptide agonist and antagonist towards B1R and B2R that found that the mutation of Lys[118] for Serine B1R and vice versa in B2R, were responsible for discrimination of either receptor [47]. Specifically, mutating Ser[111] to Lys[111] in B2R led to the decrease of the affinity in BK and the peptide antagonist NPC17731, while increased the affinity of B1R selective antagonist, thus effectively turning the receptor, at least in terms of activity, from B2R to B1R. The change was attributed to the ability of the positive charge in the lysine to repel the C-terminal arginine in B2R selective peptides such as BK and NPC17731 and to attract the negative charge of the C-terminal of B1R selective peptides which lacks the C-terminal arginine [47].

This hypothesis can also explain the behavior of compound#3 in Table 1, found to be a specific binding inhibition (45%) in B2R but inactive in B1R. Figure 5.18 shows the bound conformation of compound#3 to the B2R. Considering the similarities in the binding pocket between B1R and B2R and consequently the pharmacophores, in B1R the fluorophenyl moiety would be oriented towards Lys[118] and could not properly fit because of steric hindrance. On the other hand, since the sidechain of Ser[111] in B2R orients away from the binding pocket, the fluorophenyl group comfortably rests on the hydrophobic cleft and make interactions with residues in the region such as Trp[256].
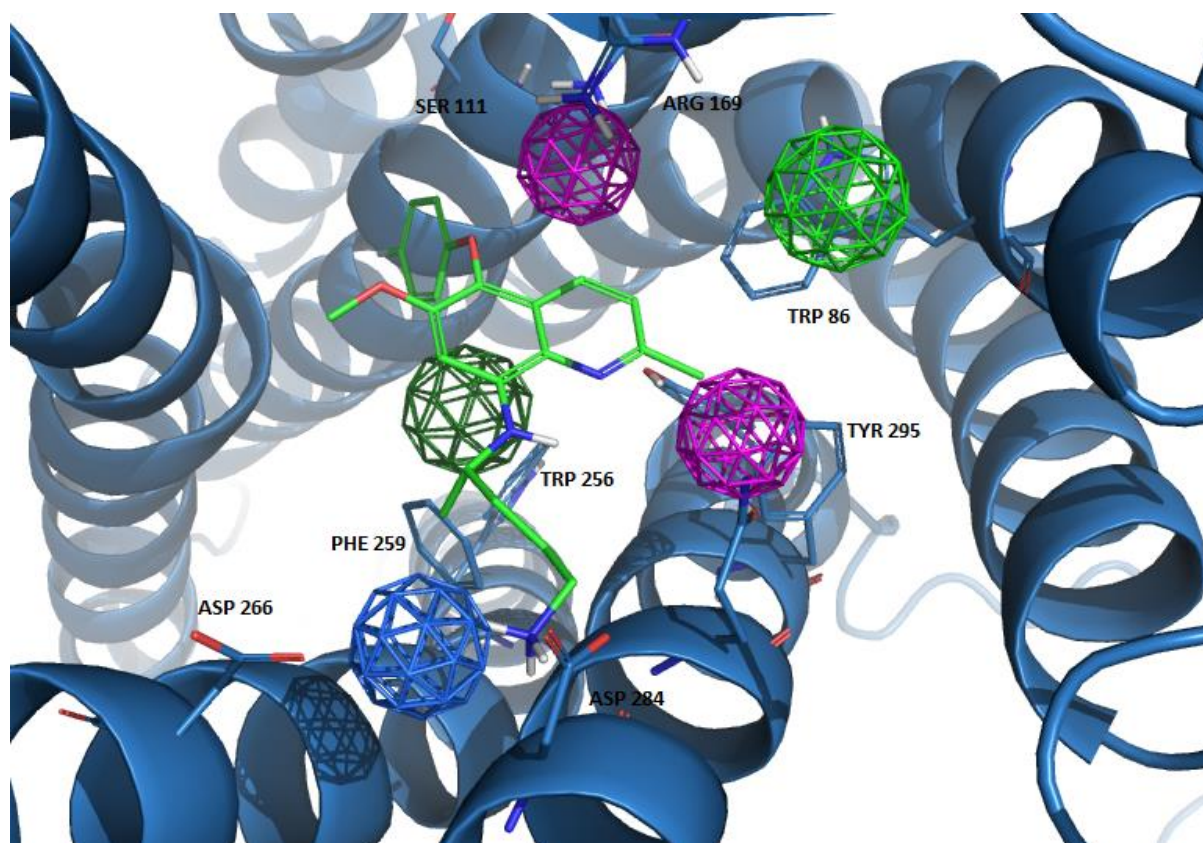
Figure 5.18: Pictorial view of compound #3 of Table 1 bound to the B2R with the pharmacophore points represented as spheres of different colors.

To conclude, both receptors require large molecules to fulfill the five proposed pharmacophoric points. Moreover, in regard to receptor selective ligands, a particular attention must be paid to the moiety fulfilling pharmacophoric point 5.

5.7 **Conclusions to chapter** 5

We have constructed models of the bound conformation(s) of diverse non-peptide B$_2$ bradykinin antagonist and analyzed the stereochemical features of the complexes with the aim to find common trends. For this purpose an atomistic model of the B2R was constructed by homology modeling, using the CXC4 chemokine receptor as template. Antagonists selected for the present study include Fasitibant, FR173657, Anatibant, WIN64338, Bradyzide, CHEMBL442294 and JSM10292, encompassing the maximum possible diversity. Complexes with the bound conformation of each of the antagonists were constructed by docking the molecules into the receptor. Due to the flexibility of the ligands and the size of the orthosteric site of the receptor, several docking attempts were carried out for each of the molecules. The final conformation was selected by the scoring function and the results on site directed mutagenesis studies available.

Our results suggest that there are certain anchoring points that are found in more than one compound permitting the definition of a common pharmacophore. This consist of five points that defined on the features of the ligand include a proton donor/positive charge (point 1), a proton acceptor/proton donor (point 2), an aromatic/cyclic moiety (point 3), a proton acceptor/proton donor (point 4) and a hydrophobic/aromatic moiety (point 5).

The pharmacophore was used in a subsequent study to guide a virtual screening process. The results permitted to identify a set of compounds some of which were purchased and *in vitro* tested for their capability to antagonize the bradykinin B2R. In the present work a subset of the hits compounds are disclosed with highly diverse structures that give support to the validity of the pharmacophoric hypothesis described in this work.

Finally, in order to design selective ligands, comparison of the pharmacophores of B1R and B2R suggest that specific attention must be paid to the moiety fulfilling pharmacophoric point 5.

5.8 References to chapter 5

1. Regoli D, Barabe J (1980) Pharmacology of bradykinin and related kinins. Pharmacol Rev 32: 1–46.

2. Leeb-Lundberg L M F, Marceau F, Muller-Esterl W, Pettibone D J, Zuraw B L (2005) International Union of Pharmacology. XLV. Classification of the Kinin Receptor Family: from Molecular Mechanisms to Pathophysiological Consequences. Pharmacol Rev 57: 27–77.

3. Marceau F, Regoli D. (2004) Bradykinin receptor ligands: therapeutic perspectives. Nat. Rev. Drug Discov. 3: 845-852

4. Blaes N, Girolami J-P (2013) Targeting the 'Janus face' of the B2-bradykinin receptor. Expert Opin Ther. Targets 17: 1145-1166.

5. Boissonnas R A, Guttmann S, Jaquenoud P A (1960) Synthese de la L-raginyl-L-prolyl-L-prolyl-gycyl-L-phenylalanyl-L-seryl-L-prolyl-L-phenylalanyl-L-ariginine un nonapeptide presentant les proprietes de la bradykinine. Helv. Chim. Acta 43: 1349-1358.

6. Vavrek R J, Stewart J M (1985) Competitive Antagonists of Bradykinin. Peptides 6: 161-164.

7. Steranka L R, Farmer S G, Burch R M (1989) Antagonists of B2 Bradykinin Receptors. FASEB J 3: 2019-2025.

8. Stewart J M (2004) Bradykinin antagonists: discovery and development. Peptides 25: 527–532.

9. Kyle D J, Blake P R, Smithwick D, Green L M, Summers M F, Martin J A, Sinsko J A (1993) NMR and computational evidence that high-affinity bradykinin receptor antagonists adopt C-terminal β-turns. J Med Chem 36: 1450–1460.

10. Lopez J J, Shukla A K, Reinhart C, Schwalbe H, Michel H, Glaubitz C. (2008) The Structure of the Neuropeptide Bradykinin Bound to the Human G-Protein Coupled Receptor Bradykinin B2 as Determined by Solid-State NMR Spectroscopy. Angew Chem Int Ed 47: 1668–1671.

11. Hock F J, Wirth K, Albus U, Linz W, Gerhards G H, Wiemer G, Henke S, Breipohl G, Konig W, Knolle J, Scholkens B A (1991) Hoe 140 a New Potent and Long-Acting Bradykinin-Antagonist: in vitro Studies. Br J Pharmacol 102: 769-773.

12. Kyle D J, Burch R M (1992) Recent advances toward novel bradykinin antagonists. Drugs Future 17: 305-312.

13. Thurieau C, Feleton M, Hennig P, Raimbaud E, Canet E, Fauchere J-L (1996) Design and Synthesis of New Linear and Cyclic Bradykinin Antagonists. J Med Chem 39: 2095-2101.

14. Monteagudo E S, Calvani F, Catrambone F, Fincham C I, Madami A, Meini S, Terracciano R (2001) New conformationally homogeneous beta-turn antagonists of the human B-2 kinin receptor. J Pept Sci 7: 270-283

15. Kyle D J, Chakravarty S, Sinsko J A, Stormann T M (1994) A Proposed Model of Bradykinin Bound to the Rat B2R and its Utility for Drug Design. J Med Chem 37: 1347-1354.

16. Meini S, Quartara L, Rizzi A, Patacchini R, Cucchi P, Giolitti A, Calo G, Regoli D, Criscuoli M, Maggi C A (1999) MEN 11270 A novel selective constrained peptide antagonist with high affinity at the human B2 kinin receptor. J Pharmacol Exp Ther 289: 1250-1256.

17. Bork K, Yasothan U, Kirkpatrick P. Icatibant. (2008) Nat Rev Drug Discov 7: 801-802

18. Heitsch H. (2002) Non-peptide antagonists and agonists of the bradykinin B2R. Curr Med Chem 9: 913-928.

19. Dziadulewicz E D (2005) Non-peptide ligands for bradykinin receptors 1995-2004. Expert Opin Ther Patents 15: 829-859.

20. Sawutz D G, Salvino J M, Dolle R E, Casiano F, Ward S J, Houck W T, Faunce D M, Douty B D, Baizman E, Awad M M A, Marceau F, Seoane P. (1994) The Nonpeptide WIN 64338 is a Bradykinin B2R Antagonist. Proc Natl Acad Sci USA 91: 4693-4697.

21. Asano M, Inamura N, Hatori C, Sawai H, Fujiwara T, Katayama A, Kayakiri H, Satoh S, Abe Y, Inoue T, Sawada Y, Nakahara K, Oku T, Okuhara M. (1997) The identification of an orally active nonpeptide bradykinin B2R antagonist FR173657. Br J Pharmacol 120: 617–624.

22. Burgess G M, Perkins M N, Rang H P, Campbell E A, Brown M C, McIntyre P, Urban L, Dziadulewicz E K, Ritchie T J, Hallett A, Snell C R, Wrigglesworth R, Lee W, Davis C, Phagoo S B, Davis J, Phillips E, Drake G S, Hughes G A, Dunstan A, Bloomfield G C. (2000) Bradyzide a potent

non-peptide B2 bradykinin receptor antagonist with long-lasting oral activity in animal models of inflammatory hyperalgesia. British J Pharmacol 129: 77-86.

23. Pruneau D, Paquet J L, Luccarini J M, Defrene E, Fouchet C, Franck R M, Loillier B, Robert C, Belichard P, Duclos H, Cremers B, Dodey P. (1999) Pharmacological profile of LF 16-0687 a new potent non-peptide bradykinin B2R antagonist. Immunopharmacol 43: 187-194.

24. Cucchi P, Meini S, Bressan A, Catalani C, Bellucci F, Santicioli P, Lecci A, Faiella A, Rotondaro L, Giuliani S, Giolitti A, Quartara L, Maggi C A. (2005) MEN16132 a novel potent and selective nonpeptide antagonist for the human bradykinin B2R. In vitro pharmacology and molecular characterization. Eur J Pharmacol 528: 7-16.

25. Dziadulewicz E K, Brown M C, Dunstan A R, Lee W, Said N B, Garratt P J. (1999) The design of non-peptide human bradykinin B2R antagonists employing the benzodiazepine peptidomimetic scaffold. Bioorg Med Chem Lett 9: 463-468.

26. Gibson C, Schnatbaum K, Pfeifer J R, Locardi E, Paschke M, Reimer U, Richter U, Scharn D, Faussner A, Tradler T. (2009) Novel Small Molecule Bradykinin B2R Antagonists. J Med Chem 52: 4370–4379.

27. Sali A, Blundell T L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234: 779–815.

28. Molecular Operating Environment (MOE). Chemical Computing Group Inc. 1010 Sherbooke St. West Suite #910 Montreal QC Canada H3A 2R7. 2013

29. Friesner R A, Banks J L, Murphy R B, Halgren T A, Klicic J J, Mainz D T, Repasky M P, Knoll E H, Shaw D E, Shelley M, Perry J K, Francis P, Shenkin P S. (2004) Glide: A New Approach for Rapid Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 47: 1739–1749.

30. Meini S, Bellucci F, Catalani C, Cucchi P, Giolitti A, Giuliani S, Quartara L, Rotondaro L, Zappitelli S, Maggi C A. (2011) Comparison of the molecular interactions of two antagonists MEN16132 or icatibant at the human kinin B2R. British J Pharmacol 162: 1202-1212.

31. Cordomi A, Edholm O, Perez J J. (2007) Effect of different treatments of long-range interactions and sampling conditions in molecular dynamic simulations of rhodopsin embedded in a dipalmitoyl phosphatidylcholine bilayer. J Comput Chem 28: 1017-1030.

32. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark A E, Berendsen H J. (2005) GROMACS: fast flexible and free. J Comp Chem 26: 1701–1718

33. Pruneau D, Luccarini J-M, Fouchet C, Defrene E, Franck R-M, Loillier B, Duclos H, Robert C, Cremers B, Belichard P, Paquet J-L. (1998) LF 16.0335, a novel potent and selective nonpeptide antagonist of the human bradykinin B2R. British J Pharmacol 125: 365-372.

34. Katritch V, Cherezov V, Stevens R C. (2012) Diversity and modularity of G protein-coupled receptor structures. Trends Pharmacol Sci 33: 17-27.

35. Wu B, Mol C D, Han G W, Katritch V, Chien E Y T, Liu W, Cherezov V, Stevens R C (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science 330: 1066-1071.

36. Bellucci F, Meini S, Cucchi P, Catalani C, Giuliani S, Zappitelli S, Rotondaro L, Quartara L, Giolitti A, Maggi A C. (2004) The N-terminal of Icatibant and bradykinin interact with the same Asp residues in the human B2R. Eur J Pharmacol 491: 121–125.

37. Meini S, Cucchi P, Belluccia F, Catalania C, Faiella A, Rotondaro L, Quartara L, Giolitti A, Maggi C A. (2004) Site-directed mutagenesis at the human B2R and molecular modelling to define the pharmacophore of non-peptide bradykinin receptor antagonists. Biochem Pharmacol 67: 601–609.

38. Sawada Y, Kayakiri H, Abe Y, Imai K, Mizutani T, Inamura N, Asano M, Aramori I, Hatori C, Katayama A, Oku T, Tanaka H. (2004) A New Series of Highly Potent Non-Peptide Bradykinin B2R Antagonists Incorporating the 4-Heteroarylquinoline Framework. Improvement of Aqueous Solubility and New Insights into Species Difference. J Med Chem 47: 1617-1630.

39. Marie J, Richard E, Pruneau D, Paquet J-L, Siatka C, Larguier R, Ponce C, Vassault P, Groblewski T, Maigret B, Bonnafous J-C (2001) Control of Conformational Equilibria in the Human B2

Bradykinin Receptor. J Biol Chem 276: 41100–41111.

40. Meini S, Meinihi P, Zappitelli S, Rotondaro L, Quartara L, Giolitti A, Maggi C A. (2002) Preliminary mutational analysis of the human kinin B2R for nonpeptide antagonist ligands recognition. Can J Physiol Pharmacol 80: 303–309.

41. Dziadulewicz E K, Ritchie T J, Hallett A, Snell C R, Davies J W, Wrigglesworth R, Dunstan A R, Bloomfield G C, Drake G S, McIntyre P, Brown M C, Burgess G. M, Lee W, Davis C, Yaqoob M, Phagoo S B, Phillips E, Perkins M N, Campbell E A, Davis A J, Rang H P. (2002) Nonpeptide Bradykinin B2R Antagonists: Conversion of Rodent-Selective Bradyzide Analogues into Potent Orally-Active Human Bradykinin B2R Antagonists. J Med Chem 45: 2160-2172.

42. Faussner A, Schüssler S, Feierler J, Bermudez M, Pfeifer J, Schnatbaum K, Tradler T, Jochum M, Wolber G, Gibson C. (2012) Binding characteristics of [3H]-JSM10292: a new cell membrane-permeant non-peptide bradykinin B2R antagonist. British J Pharmacol 167: 839–853.

43. Salvino J M, Seoane P R, Douty B D, Awad M M A, Dolle R E, Houck W T, Faunce D M, Sawutz D G. (1993) Design of Potent Non-Peptide Competitive Antagonists of the Human Bradykinin B2R. J Med Chem 36: 2683-2584.

44. Filizola M, Llorens O, Carteni-Farina M, Perez J J (1998) New insights into the conformational requirements of B2 bradykinin antagonism. Bioorg. Med. Chem. Lett. 6: 1491-1500.

45. Pineda L F, Hassan M (1998) Formulation of 3D pharmacophore models for bradykinin B2-receptor Antagonists. Lett Peptide Sci 5: 189–192.

46. Krishna S, Singh D K, Meena S, Datta D, Siddiqi M I, Banerjee D. (2014) Pharmacophore-Based Screening and Identification of Novel Human Ligase I Inhibitors with Potential Anticancer Activity. J. Chem. Inf. Model. 54: 781–792

47. Fathy, D. B., Mathis, S. A., Leeb, T., & Leeb-Lundberg, L. F. (1998). A single position in the third transmembrane domains of the human B1 and B2 bradykinin receptors is adjacent to and discriminates between the C-terminal residues of subtype-selective ligands. *Journal of Biological Chemistry*, *273*(20), 12210-12218.

CHAPTER 6:

Dimerization of Serotonin receptor

1A and Galanin Receptor 1

6.0 Chapter Summary

Currently, the number of reports in the literature citing specific examples of GPCR homodimers, heterodimers and oligomers are increasing. GPCR dimer formation has been reported to be responsible of aspects of GPCR regulation. However, the functional relevance and the presence of GPCR dimers in vivo is still a subject of great debate. The 5-Hydroxytryptamine 1A receptor and the galanin receptor 1, both belong to the rhodopsin-like family, both are known to interact with the same G protein subtype: G_i/G_0 albeit in different manner and, have been described to heterodimerize, triggering an anomalous physiological state that eventually cause depression. On the other hand, zinc supplementation has been widely reported to improve treatment against major depressive disorders. In this chapter protein-protein interactions between the two receptors have been studied to establish a rational explanation for the role of zinc in the molecular processes associated with protein-protein interactions and the relationship with depression. Using homology modelling, atomic models of the receptors were constructed and refined using molecular dynamics simulations. Moreover, using MOE GRID program the 5-Hydroxytryptamine 1A receptor was explored with a Zinc probe to identify prospective binding sites. Finally a heterodimer of the two receptors was constructed using GRAMM software and concluded that some of the Zn binding sites rest on the dimer interface preventing its formation.

The results show three possible dimer interfaces: the first at TM4 and TM5, the second at TM6 and TM7, and the last at TM1 and TM2. Analysis of the zinc sites and the heterodimer interfaces suggests that there is a coincidence between zinc binding sites and heterodimerization interfaces. The results also establish a rational explanation for the role of zinc in the molecular processes associated with heterodimer interactions and its relationship with depression, in agreement with previously reported evidence for the positive effects of zinc in depression treatment, and the involvement of dimer formation in the same disease.

6.1 Introduction

Protein-protein interactions are key to carry out many of the functions of proteins including signal transduction, transport across membranes, cell metabolism or muscle contraction. Moreover diverse essential molecular processes within a cell are carried out by molecular machines that are built from a large number of protein components, organized by their protein-protein interactions. Thus, many proteins do not act in isolation and often, examination of the interaction between proteins provides plenty of information on their biological function together [1].

In recent years a vast number of studies have demonstrated the capacity of many GPCRs to dimerize into homodimers or heterodimers [2-11]. Evidence for this ability has unambiguous been demonstrated for the glutamate class, including the calcium-sensing receptors, metabotropic glutamate receptors and taste-1 receptors that are stabilized by disulfide bonds between their extra-cellular Venus flytrap motifs [12-14], and moreover the mutations that prevent the formation of these disulfide bonds also disrupt receptor dimerization [15]. In fact, for some of these receptors, dimerization has been shown to increase the affinity to their agonist and hence, dimerization can be regarded as the important aspect of optimal receptor function [16-17]. Moreover, the potential for dimeric interactions between rhodopsin-like GPCRs have been investigated extensively [18-25]. Such interactions can result in alterations in ligand pharmacology, the nature of the signals generated (Figure 1), and cellular trafficking of the complexes [9-11].

The 5-Hydroxytryptamine 1A receptor (5-HT1A), one of the receptor subtypes for the neurotransmitter serotonin (5HT) and the galanin receptor 1 (GALR1), a receptor subtype for the neuropeptide galanin, both belonging to the rhodopsin-like family, have been described to heterodimerize, triggering an anomalous physiological state that essentially would cause depression [26-27]. The two receptors also are known to interact with the same G protein subtype, the heterotrimeric $G_{i/o}$. The $G_{i/o}$ inhibits the production of cAMP from ATP [28].

The 5-HT_{1A} receptor is the most widespread among all the serotonin receptors. In the central nervous system, 5-HT_{1A} is expressed presynaptically in the brain on the soma and dendrites of 5-HT

neurons located in the dorsal and medial raphe nuclei. Furthermore 5-HT$_{1A}$ is also expressed postsynaptically, particularly in the hippocampus, lateral septum, cortex, and amygdala area of the brain [29-30]. On the other hand, GALR1 is widely expressed throughout the human body such as in the brain and spinal cord, as well as in peripheral sites such as the small intestine and heart [31]. Because of lack of reliable antibodies for either of these receptor subtypes, definitive information about the subcellular location of the receptor protein is still sparse [32].
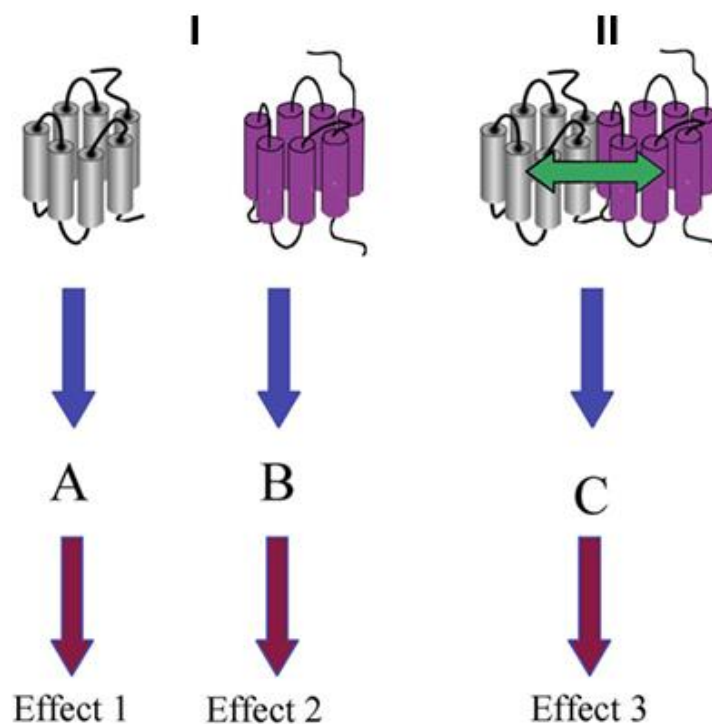
Figure 6.1: The schematic of the signal type(s) produced by GPCR dimers. Monomers that forms a dimer produce different signals individually, a different signal when the monomers dimerize.

In a normal body system the functions of 5-HT (Figure 2), produced in serotonergic neurons of the central nervous system includes the regulation of mood, appetite, and sleep. 5HT also has some cognitive functions, including memory and learning. Modulation of 5-HT release and activity in the synapses is thought to be mode of action of several classes of pharmacological antidepressant [33-35]. In contrast, in the nervous system galanin primarily displays predominantly an inhibitory,

hyperpolarizing and modulatory role and as such inhibits release of neurotransmitters [36].
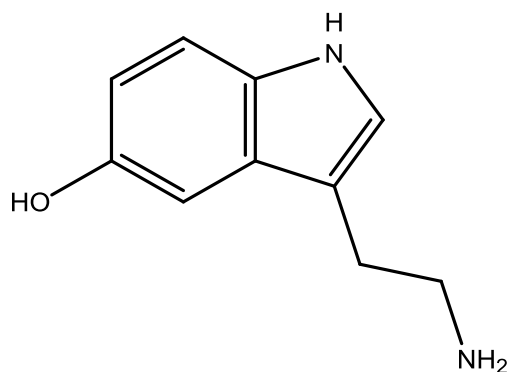

Figure 6.2: The chemical structure of the neurotransmitter serotonin.

Several neurochemical and molecular studies on the *in vivo* effects of the neuropeptide galanin on central 5-HT neurotransmission and on 5-HT1A receptor-mediated responses have shown that galanin (Figure 3) affects noradrenalin and 5HT release in locus coeruleus and dorsal raphe nucleus regions [37-38], the two neurotransmitter systems known to be important in mood disorders. Galanin is also known to downregulate 5-HT1A since intraventricular administration of galanin was shown to reduce the efficacy of antidepressants [39], indicating that galanin has important inhibitory actions on central 5HT neurotransmission and on 5-HT$_{1A}$ receptor-mediated activities [37-40]. Furthermore, segments of galanin such as the fifteen residues of the N-terminus, are especially potent in antagonizing 5-HT1A receptor signaling in limbic regions and this effect is blocked by known GALR1 antagonists [24]. In the dorsal raphe, similar antagonistic interactions have been reported to exist, on the contrary here 5-HT$_{1A}$ is involved [41].

These experimental results were the first indications that heterodimers of GALR1 and 5-HT$_{1A}$ do form in which GALR1 antagonize 5-HT1A recognition and signaling via a protein-protein interactions postulated to involve allosteric mechanisms [24, 26-27, 42]. Indeed, FRET experiments have shown 5-HT$_{1A}$ to heterodimerize with galanin, furthermore studies have also found that the 5-HT$_{1A}$ receptor can be co-immuno-precipitated with a wide range of different GPCRs, including 5-HT$_{1B}$, 5-HT$_{1D}$, EDG1, EDG3, GPR26 and GABAB2 receptors [43]. Thusly, 5-HT$_{1A}$ has propensity to

homo/heterodimerize.
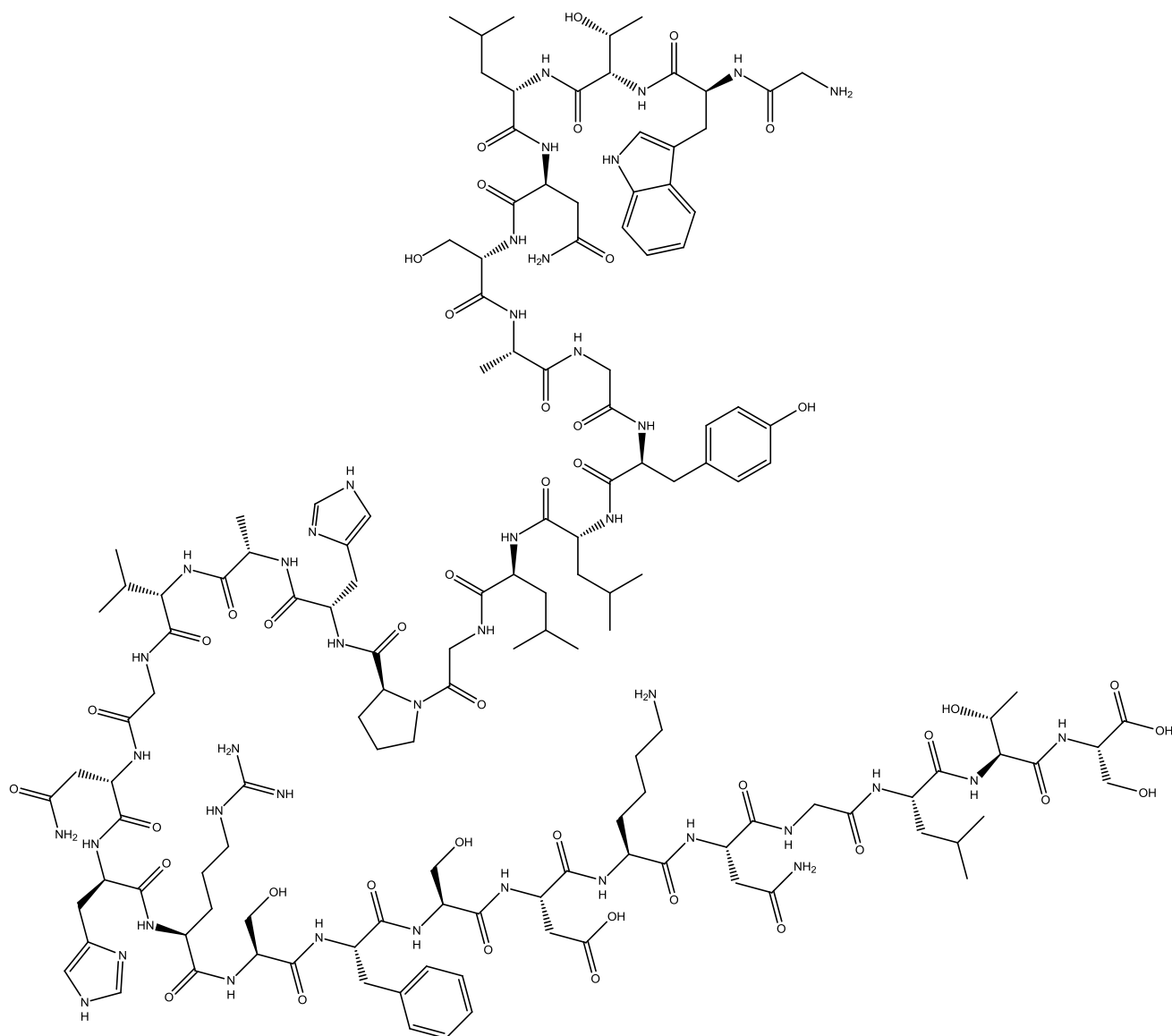
Figure 6.3: The structure of the 30 AA human neuropeptide galanin (WTLNSAGYLLGPHAVGNHRSFSDKNGLTS).

Metal ions such as zinc are known to be physiologically relevant agonist or modulators of GPCRs [44-50]. For example, zinc ions specifically, has been reported to be the agonist for the GPR39 receptor which has been linked with depressive disorders. Furthermore, in diverse types of binding assays reported zinc was found to interact allosterically with 5-HT1A receptors. More specifically, it was concluded that the zinc ion is an insurmountable antagonist of 5-HT1A receptor [46]. Finally,

clinical studies have shown that zinc supplementation increases the efficiency of the 5-HT1A receptor ligands citalopram and imipramine used in depression treatment [51-52].

Based on these well-established links between zinc and 5-HT1A, and of 5-HT1A with GalR1, it was hypothesized that zinc could be involved in altering the heterodimerization process between 5-HT1A and GALR1. Accordingly experimental studies using FRET and SPR results were conducted. The results of that experimental study confirmed proved the hypothesis that zinc antagonize formation of GALR1 and 5-HT1A heterodimer [53]. This chapter aims to provide molecular explanation to the conclusion reached in that study.

6.2 Methodology

In order to establish a rational explanation for the role of zinc in the molecular processes associated with the interaction of GALR1 and 5-HT1A heterodimer and its relationship with depression, diverse computational methods were employed. Homology modeling was used to construct atomistic models of both monomers: GALR1 and 5-HT1A that were subsequently refined using molecular dynamics simulation. In addition, MOE GRID program was used to scan a zinc probe all over the surface of the 5-Hydroxytryptamine 1A receptor to identify putative zinc binding sites. Finally, the protein-protein molecular docking GRAMM software was used to create initial heterodimers for further refinement.

6.3 Homology modeling of GARL1 and 5-HT$_{1A}$

A starting model of the human GARL1 and of 5-HT$_{1A}$ receptor was constructed by homology modeling using the crystal structure of kappa –opioid receptor (PDB entry code 4DJH) and 5-HT$_{1B}$ receptor (PDB entry code 4IAR) as templates, respectively [54-55]. In each case, sequences of the two receptors were aligned, taking into account the conserved motifs found in all GPCRs, as well as the location of the disulfide bridges. These motifs, together with salt bridges are important factors in

constraining the conformation of the extracellular and transmembrane domains of the 5-HT$_{1A}$ receptor.

From the aligned sequences, a starting model of each of the receptors was constructed using the Modeller 9 version 8 (9v8) software [56]. Model validation was carried out using the molecular operating environment (MOE) program. In a subsequent step, the selective antagonist Rec 15/3079 [57], was docked into the starting model using the GLIDE software [58]. Little information is known on the non-peptidic, selective antagonists of GALR1, therefore no molecular docking was performed on the GALR1 model.

Finally, both the GALR1 model and 5-HT1A - Rec 15/3079 complex were refined using molecular dynamics simulation with the individual systems embedded in a lipid bilayer. Specifically, the both GALR1 protein and 5-HT1A - Rec 15/3079 was each embedded in separate boxes consisting in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids and water molecules generated and equilibrated according to a previously described procedure [59]. Each of the boxes had an initial size of 10.3 x 8.0 x 10.2 nm^3 (XYZ), organized in such a way that the bilayer plane was oriented on the XY plane. Before protein insertion, the box contained 256 lipids (corresponding to an area per lipid of 0.64 nm^2) and circa 17000 water molecules. The proteins were placed in the center of the box, and the overlapping molecules were removed. Particularly, all water molecules with oxygen atoms closer than 0.40 nm to a non-hydrogen atom of the protein, as well as all lipid molecules with at least one atom closer than 0.25 nm to a non-hydrogen atom of the protein, were removed. This resulted in a 5-HT1A final system containing 197 lipids and approximately 16000 water molecules and GALR1 final system containing 189 lipids and around 15000 water molecules. In order to obtain a neutral system with approximately 0.2 M NaCl concentration, randomly selected water molecules were replaced with sodium and chloride ions. For 5-HT$_{1A}$ system, 103 water molecules were replaced by 41 sodium and 62 chloride ions. A 0.2 M NaCl concentration is fairly similar to that found in biological organisms, although they exhibit different intra and extracellular ion concentrations. The molecular dynamic sampling was carried out for 600ns using GROMACS package 4.6 [60].

6.4 Heterodimerization of GARL1 and 5-HT$_{1A}$

The GARL1 and 5-HT$_{1A}$ docking was performed using GRAMM-X [61]. GRAMM-X docking involves a simplification of the structures that are considered as a rigid body representation on a 3D Cartesian grid. The program then performs an exhaustive sampling of the translation or rotation space with a softened Lennard–Jones potential to model conformational changes. The program employs a Fast Fourier Transform based conformational search method and docking involves searching for degrees of overlap between the pairs of grid. The use of GRAMM-X to study protein-protein interactions has been extensively published and validated over the years [62-65].

Model dimers resulted from the docking stage were ranked order and subject to minimization using a soft van der Waals interaction potential. Dimers were then rescored and refined by incorporating flexibility into the protein sidechains as well as the backbones. Top scoring models obtained showed three different interaction interfaces: TM4 -TM5, TM6 -TM7 and TM1-TM2. Finally, the heterodimer complex with the highest score for each of the three groups was selected for refinement. For this purpose a 600 ns molecular dynamics simulation of the heterodimer embedded in a lipid bilayer was performed on the three dimeric systems, following the procedure protocol explained in section 6.3.

6.5 Results and discussion

Time evolution of the root mean square deviation (rmsd) of the alpha carbons of the both GALR1 and 5-HT1A as well as those of their TMs bundle subset during the refinement process are shown in Figure 4. In accordance to the discussion in the previous sections only the last 100ns segment of the molecular dynamics trajectory was used to generate an average structure that was subsequently minimized in a two-step process using the steepest descent method with a distance dependent dielectric constant of 2. The average protein structure generated for each system was annotated as the refined model of GARL and of 5-HT1A.
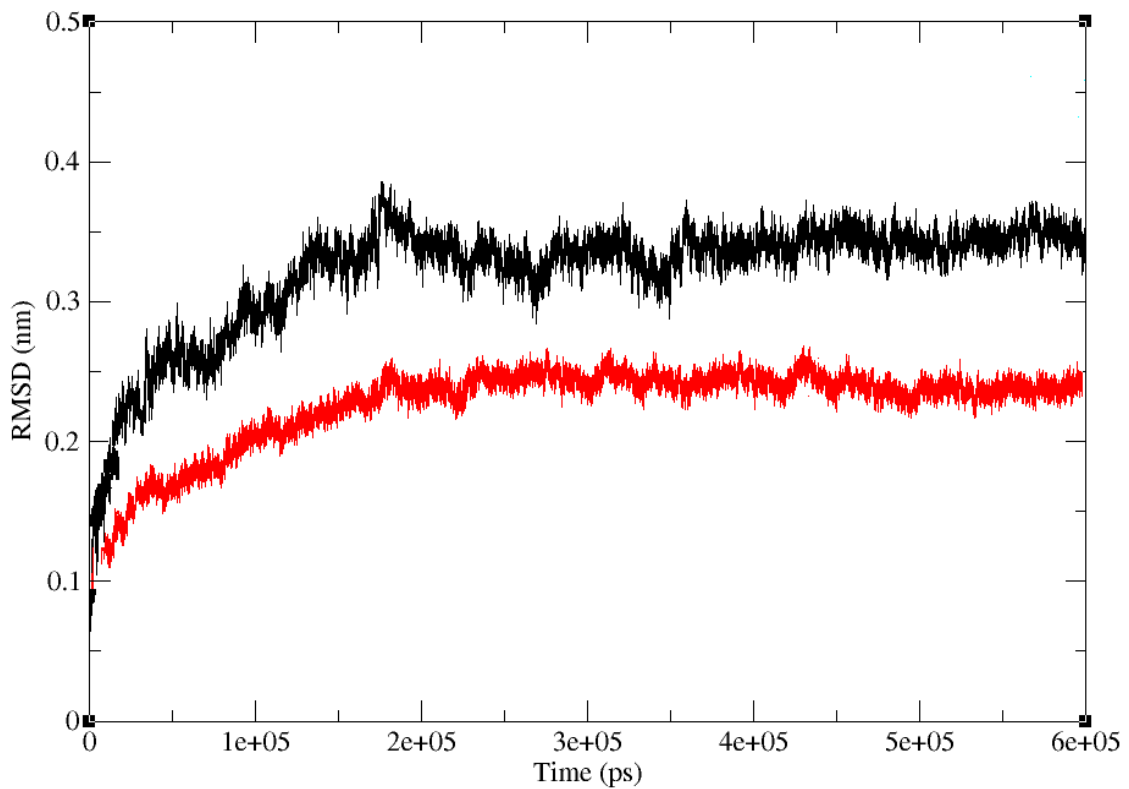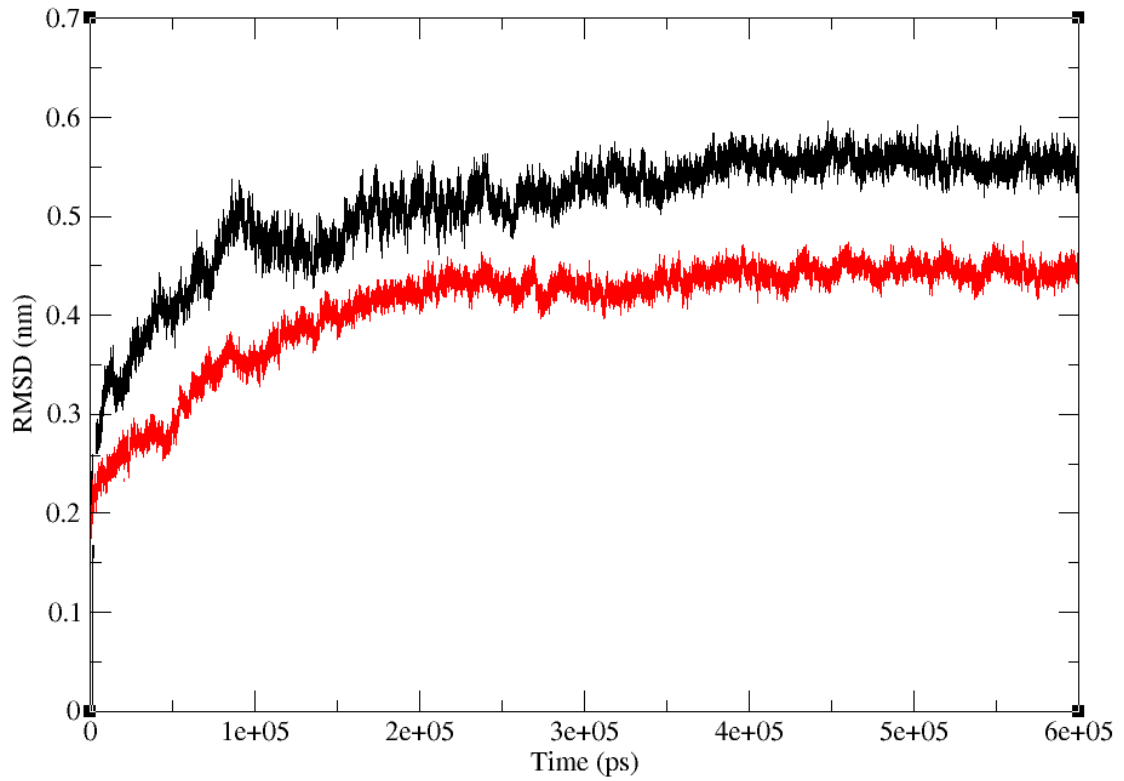
(a)

Figure 6.4: The *rmsd* of the GALR1 (a) and 5-HT$_{1A}$ (b) during the refinement process. In black is the *rmsd* of the alpha carbons of the protein and in red the *rmsd* of the alpha carbons of the helix bundle subset.

Since experimental results indicate that heterodimerization is prevented by zinc and based on previous results, zinc can directly interact with the 5-HT$_{1A}$ receptor [51-52] it is likely to consider that binding sites coincide with the heterodimerization interface [53]. Accordingly, we investigated prospective sites for zinc binding. For this purpose, a zinc probe was rolled over the surface of the refined model of 5-HT$_{1A}$ using the GRID as implemented by MOE program [66]. This permits to identify energetically favorable sites for zinc ion binding.

6.6 Determination of Zinc sites

Figure 5 shows putative sites for zinc binding on the 5-HT$_{1A}$ receptor as green spheres that are also labelled. As can be seen, most of the putative binding sites are located on the extracellular and intracellular loops, although some sites are located on the transmembrane regions including TM1; on the interface TM4 -TM5; on the interface TM6-TM7 and TM7-H8. In addition, there is another putative binding site on the inside of TM3. This site (Figure 6i) is located beneath the orthosteric binding site of the 5-HT$_{1A}$ receptor.

Experimental results aimed at characterizing receptor heterodimerization using SPR spectroscopy reported that a relative low level of receptor binding when 5-HT$_{1A}$ receptor was exposed to ZnCl$_2$ and injected to GALR1 immobilized on SPR sensorchip, compared to the control sample that was not exposed to Zncl$_2$ and furthermore, a complete receptor removal was achieved upon washing with buffer. Together these results suggest a transient protein-protein interaction much less stable than the one observed in the absence of zinc. Thus, presence of zinc ions impairs heterodimerization. At a structural level, this also suggests that zinc binding sites coincide with key protein-protein interactions sites [53].
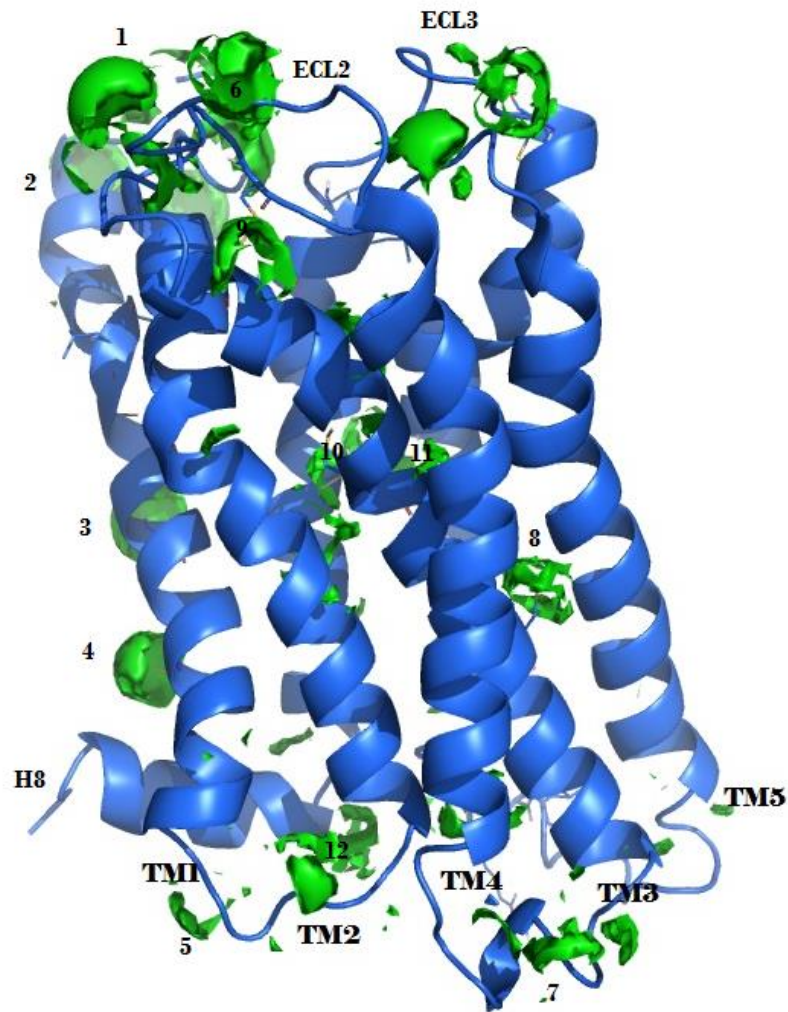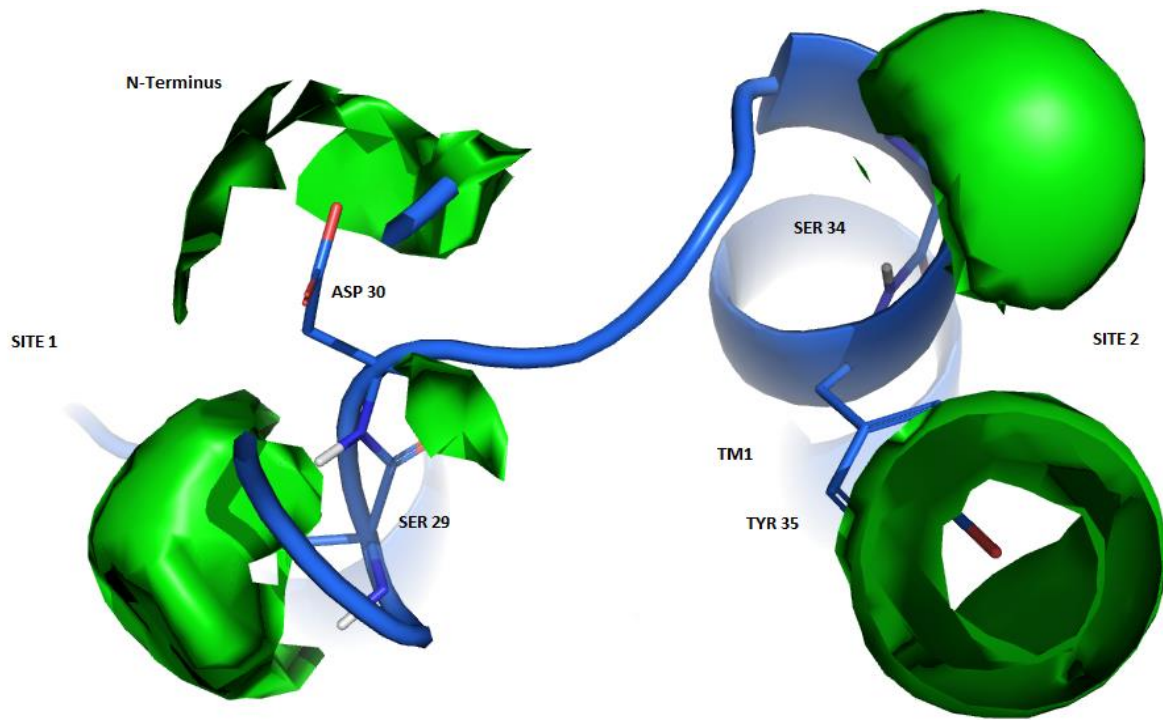
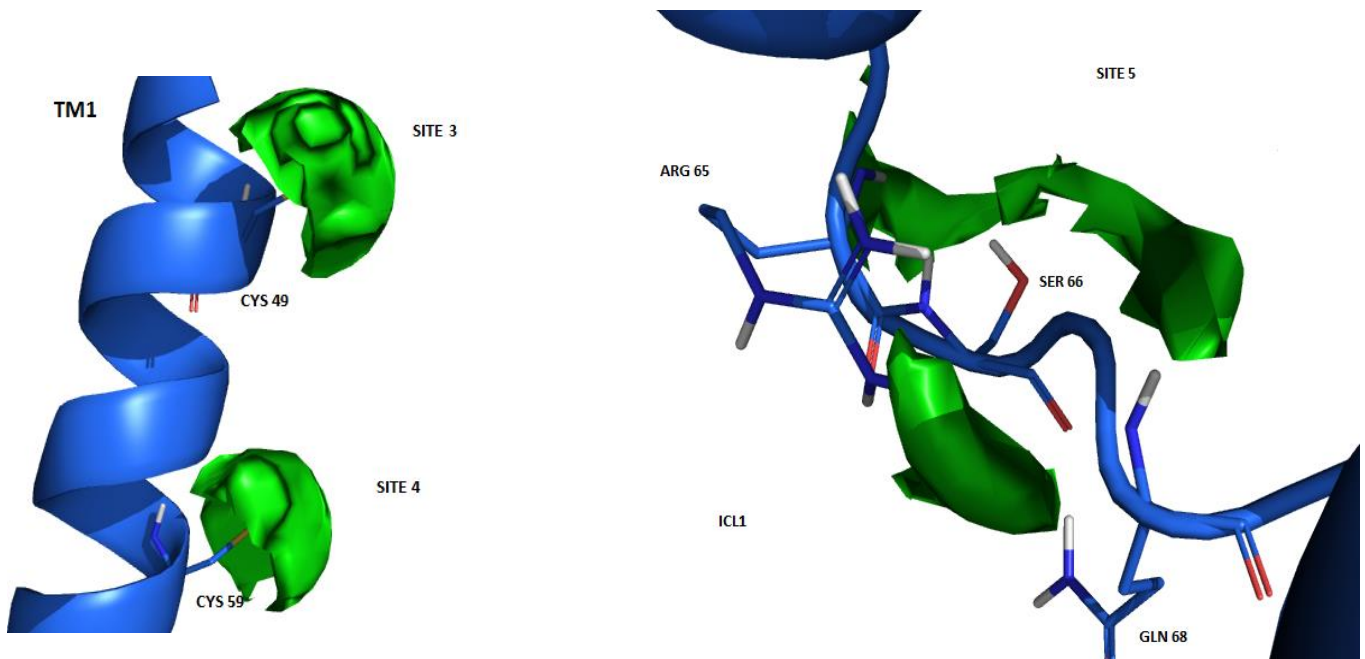**Figure 6.5**: Putative zinc binding sites in the 5-HT$_{1A}$.

It is known that amino acids such as glutamic acid, aspartic acid, histidine etc. are excellent coordinating ligands for metals such zinc, manganese, nickel etc. This characteristic makes them the prototypical metal chelators that form stable dative covalent bonds [67-68]. The capability of the 5-HT$_{1A}$ receptor to bind zinc can explain the transiency of the formed dimers and the complete removal. In other words, the existence of multiple zinc binding sites in different receptor interfaces suggests that the heterodimer between 5-HT$_{1A}$ and GALR1 will be unable to be formed in the presence of zinc. Furthermore, in case that few of the sites are populated with zinc, the heterodimer will be formed transiently or less stable either because of the location of the site or nature of the binding site free to interact with GALR1. Figure 6 shows each of the 12 putative zinc binding sites in

detail. As it can be seen, most of residues found in these sites are aspartic acids, serine, threonine
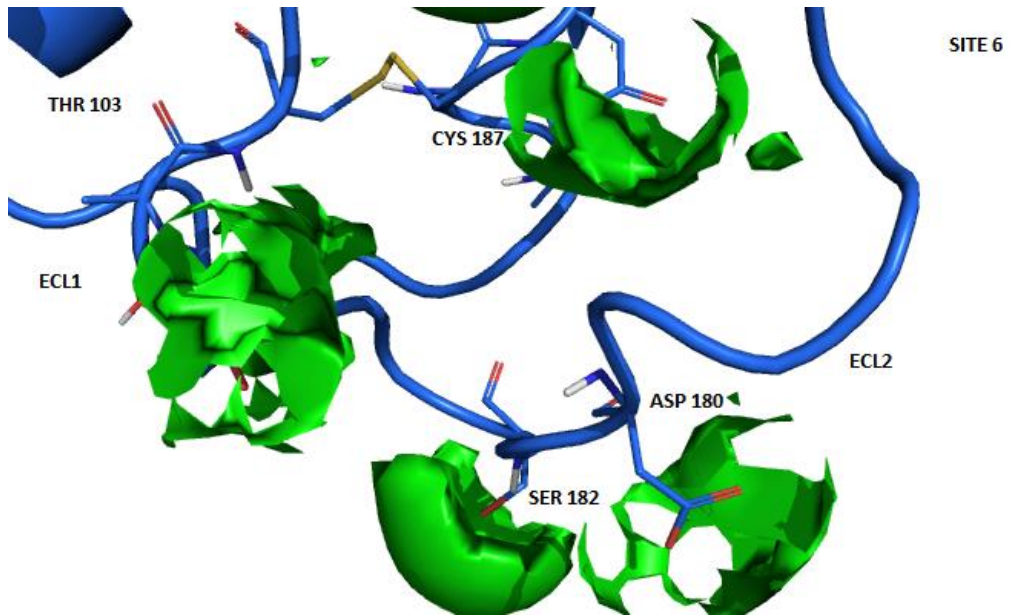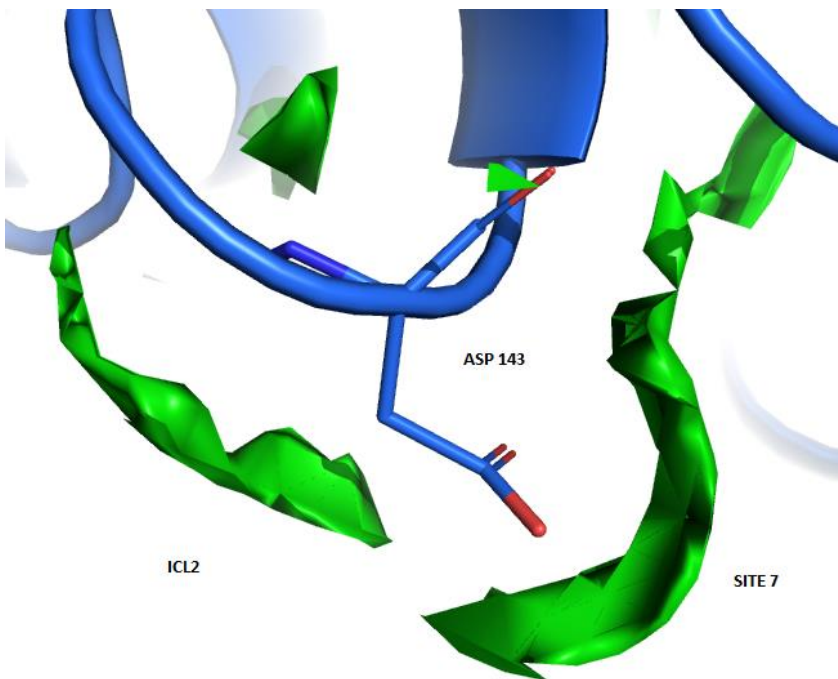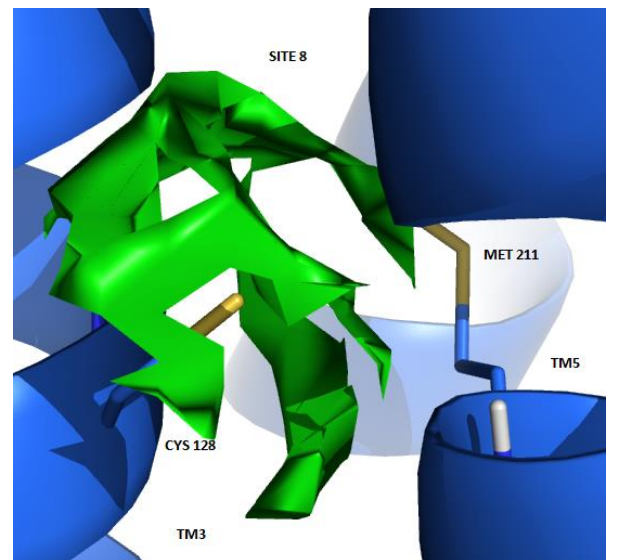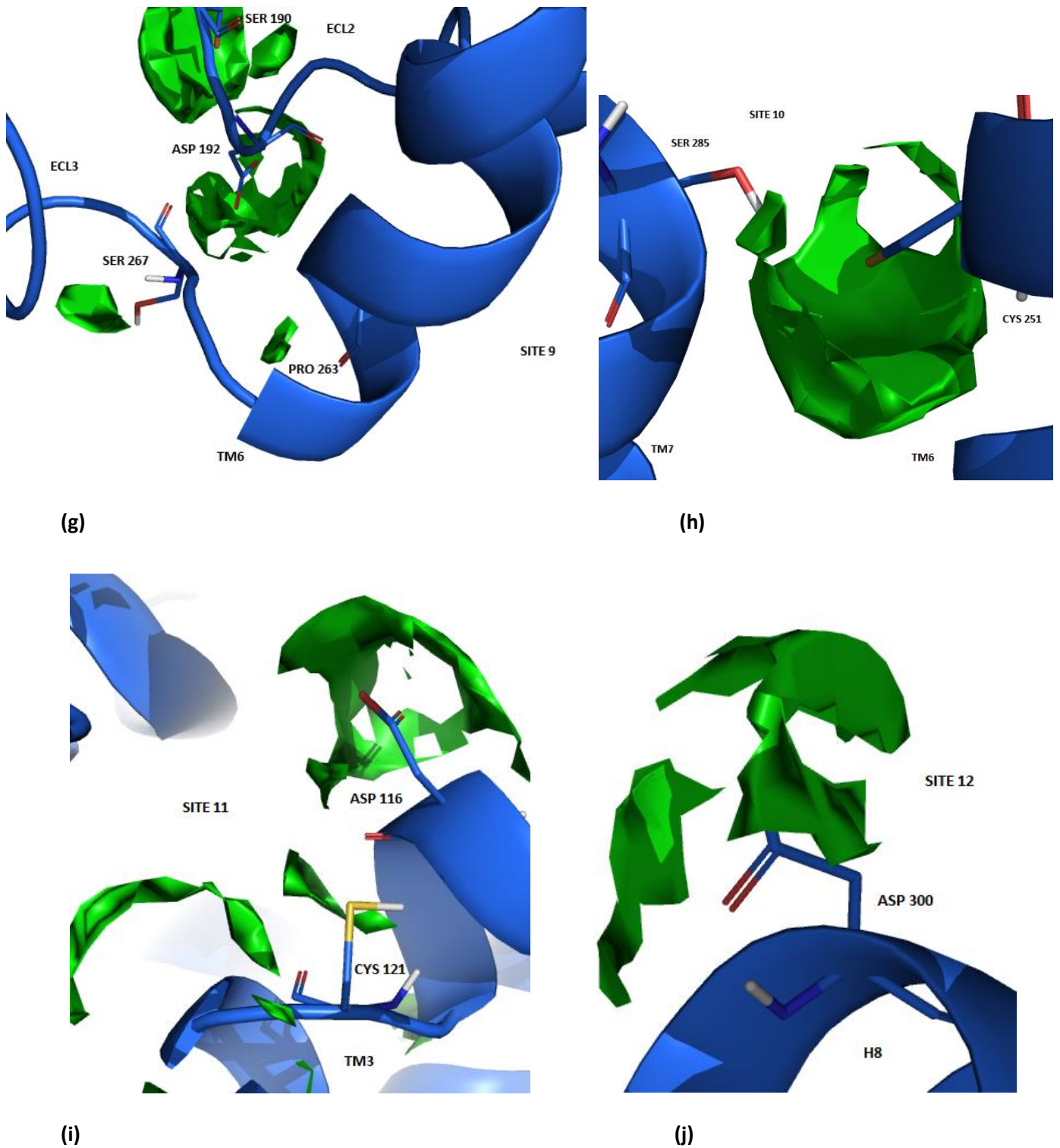
etc.

(a)

(b)

(C)

(d)

(e)

(f)

Figure 6.6: Putative zinc binding sites in the 5-HT$_{1A}$. **(a)** Site 1 on the receptor N-terminus and site 2 on the N terminal of TM1, **(b)** site 3 and 4 both on TM1, **(c)** site 5 on ICL1, **(d)** site 6 on ECL1 and ECL2, **(e)** site 7 on ICL2, **(f)** site 8 in the interface of TM3 and TM5, **(g)** site 9 on the ECL2, ECL3 and the N-terminal of TM6 ,**(h)** site 10 in the interface of TM6 and TM7 **(i)** site 11 in the TM3, and **(j)** site 12 on the N-terminal of helix 8.

6.7 Oligomerization interfaces

Identification of the helices involved in the ligand-ligand interaction is a very intriguing problem. In the case of homodimers, information provided by high-resolution crystallographic structures points to the existence of different dimer interfaces. However, some of the observed interfaces may reflect artifacts of crystal packing and the conditions employed for crystallization [69]. Moreover, they may also depend on the ligand-specific conformational states of receptors may adopt. With these considerations in mind, analysis of the available crystallographic structures reveals that TM5 and TM6 residues constitute the main interfaces for chemokine CXCR4 [70], and μ-opioid receptors [71]. Apart from this interface, crystallized chemokine CXCR4 dimers also show contacts at the intracellular ends of TM3 and TM4 and μ-opioid dimers also show a second, less prominent symmetric interface, involving TM1, TM2, and H8. However, the latter interface is also found in crystals of κ-opioid receptor dimers [54], rhodopsin, [72-73], opsin (ligand-free rhodopsin), [74], and β1-adrenoceptor [75]. Moreover, an additional interface involving TM4 and TM5 can also be found in the β1-adrenergic and smoothened receptors [76].

The use of interfering synthetic peptides and cysteine cross-linking techniques also has provided additional information. Thus, the use of interfering synthetic peptides pointed the involvement of TM6 in the formation of β2-adrenoceptor homodimers. Moreover, the involvement of TM5 in the homodimerization of dopamine D2, muscarinic M3, and serotonin 5-HT$_{2C}$ receptors has been suggested from cysteine cross-linking experiments [77-79]. Similar experiments support the two crystallographic interfaces of the β1-adrenoceptor (TM1–TM2–H8 and TM4–TM5) as physiologically relevant [75], and the involvement of a TM1–TM2–H8 interface in rhodopsin quaternary organization [80].

The case of the heterodimers is even more intriguing. Experimental results suggest that spite evidence of hetero-dimers in class C of GPCRs, there is no evidence of the existence of such a quaternary structure in class A GPCRs. Instead, it is thought that hetero-oligomers are formed from two different homodimers [69]. Reported computer modelling studies designed to investigate

potential GPCR dimer interfaces have concluded that there is no energy barrier to the formation of a TM1-H8 or a TM4-TM5 interface, but predict that a symmetric TM1-H8 interface is favored energetically than a TM4-TM5 interface [81-83]. Similar conclusions were reached in saturation FRET studies of the β2 adrenergic receptors [82], and from cysteine cross-linking and FRAP studies of dopamine D2 receptors [22, 77], suggesting the TM1-H8 interface to be more stable than TM4-TM5 interface, while the later potentially responsible for transient association of dimers into tetramers. On the contrary, similar dimerization studies on 5-HT receptors subtypes have concluded that TM4-TM5 is the preferred dimer interface. Through FRET and computer modelling, specific residues Trp^{175}, Tyr^{198}, Arg^{151}, and Arg^{152} in TM4 were identified and conformed by mutagenesis as being essential for maintaining the homodimer structure [83].

An extensive computational study on the homo- and hetero-dimerization of opioid receptors it was found that only selected interfaces in the diverse homo- and heterodimeric systems were formed. Specifically, they include TM1–TM2–H8/TM1–TM2–H8; TM1–TM2/ TM4-TM5 (also TM4-TM5/ TM1–TM2 for hetero-dimers); and TM1–TM2/TM5-TM6 (also TM5-TM6 / TM1–TM2 for hetero-dimers). Other interfaces scarcely were found along the simulations [84]. The results suggest that the dimer interface from simulations that is closest to one inferred from crystal structures is the TM1–TM2–H8/TM1–TM2–H8 interface.

Although it is not complete clear that class A GPCRs do not form heterodimers, computer simulations of such a system provide information on the favorite interfaces that could be found either in heterodimers or hetero-oligomers. As mentioned in the methods section in the present work we have carried out three different simulations of the heterodimer GARL1 and 5-HT$_{1A}$ in three different possible arrangements: the TM4-TM5/TM4-TM5; TM6-TM7/ TM6-TM7 and TM1-TM2-H8/ TM1-TM2-H8 interface.

### 6.7.1. Dimer interface 1:TM4-TM5

The TM4-TM5 dimer interface (Figure 7), has been reported to be the preferred interface for the homodimers of 5-HT$_{1A}$ and 5-HT receptors subtypes [85], however dimerization studies on other rhodopsin class GPCR [22, 81-83], interactions at this TM regions is deemed weak or unstable. Incorporating the experimental studies supporting this work [53], and the putative zinc binding site disclosed in the previous subchapter, it is very likely that at least for the heterodimer of GARL1 and of 5-HT$_{1A}$, the interactions are transient. Consider the putative zinc binding sites, at TM4-TM5 only site 8, Figure 6(f) and site 11 in Figure 6 (i) will be obscured for interactions with zinc. However the other 10 putative sites will bind with zinc and impair heterodimerization.
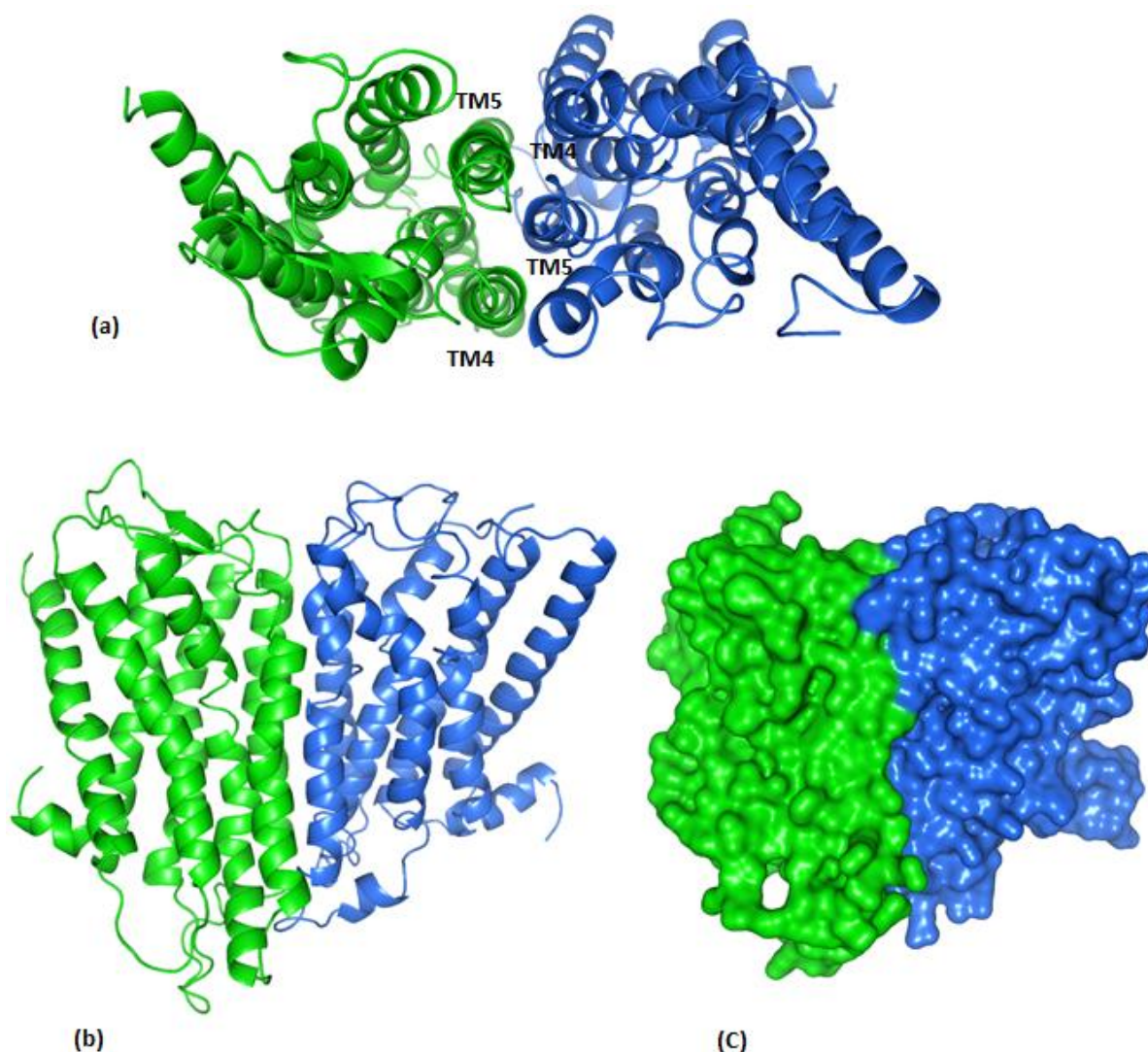
Figure 6.7: **(a)** The aerial representation of dimer 1 (TM4-TM5 interface), side view **(b)** and **(c)** molecular surface. GALR1 is shown in green and 5-HT$_{1A}$ in blue.

6.7.2. Dimer interface 2: TM6-TM7

Unlike TM4-TM5 dimer interface, TM6-TM7 (Figure 8) offer compounding credence to various reported dimerization interfaces, specifically active conformation of 5-HT$_{1B}$ as it pertaining to the angle of rotation of TM6 [55, 85], dimerization studies on other rhodopsin class GPCRs [81-83], which indicated the involvement of TM1-H8. While TM1 is not directly involved here, H8 is involved and the fact that the interfacing TM6 and TM7 here are parallel organized exposed the TM1 to interfacing region and favoring it. Furthermore, the putative zinc binding sites obscured by dimerization at this interface includes site 2, site 3, site 4, site 9, site 10, site 12 and part of site 1. Figure 8(a) shows that N-terminal of TM1 of 5-HT$_{1A}$ is oriented towards the interfacing TMs.
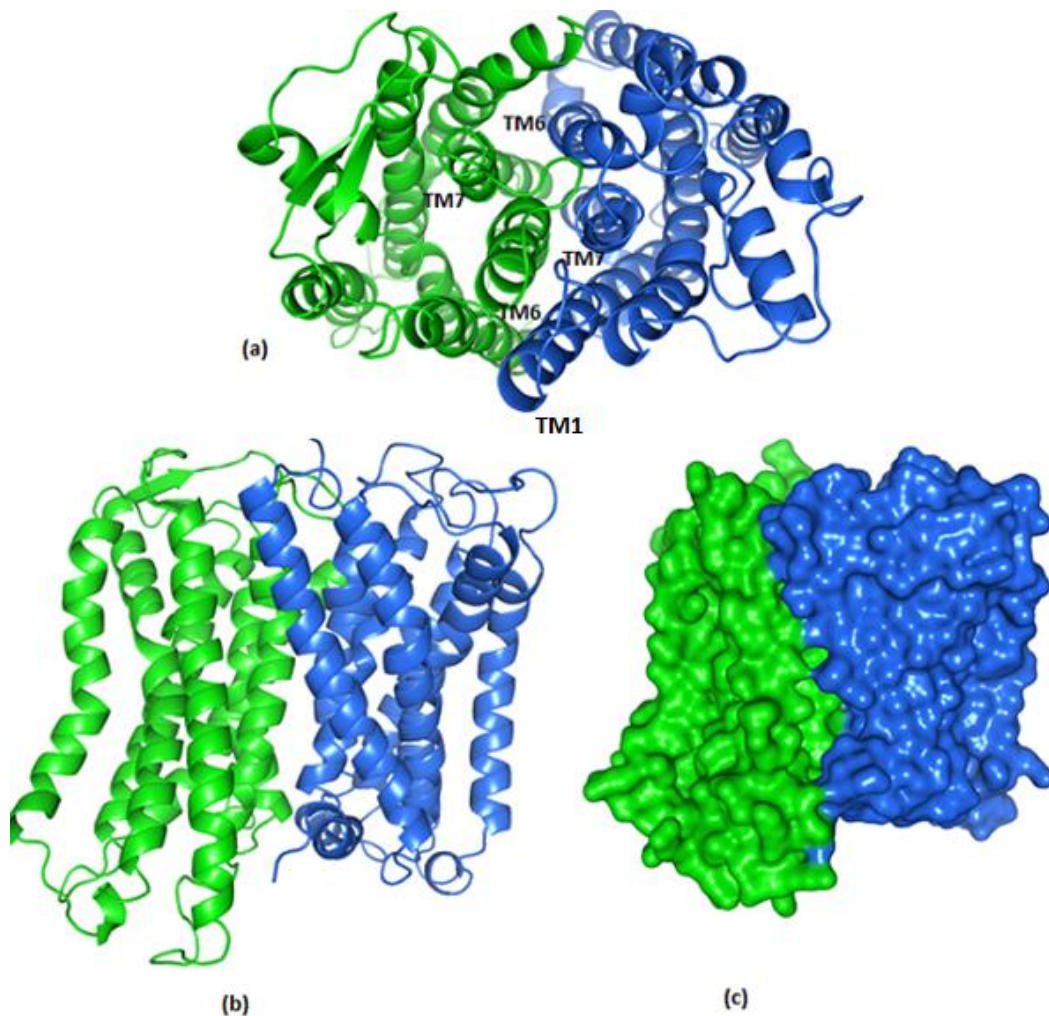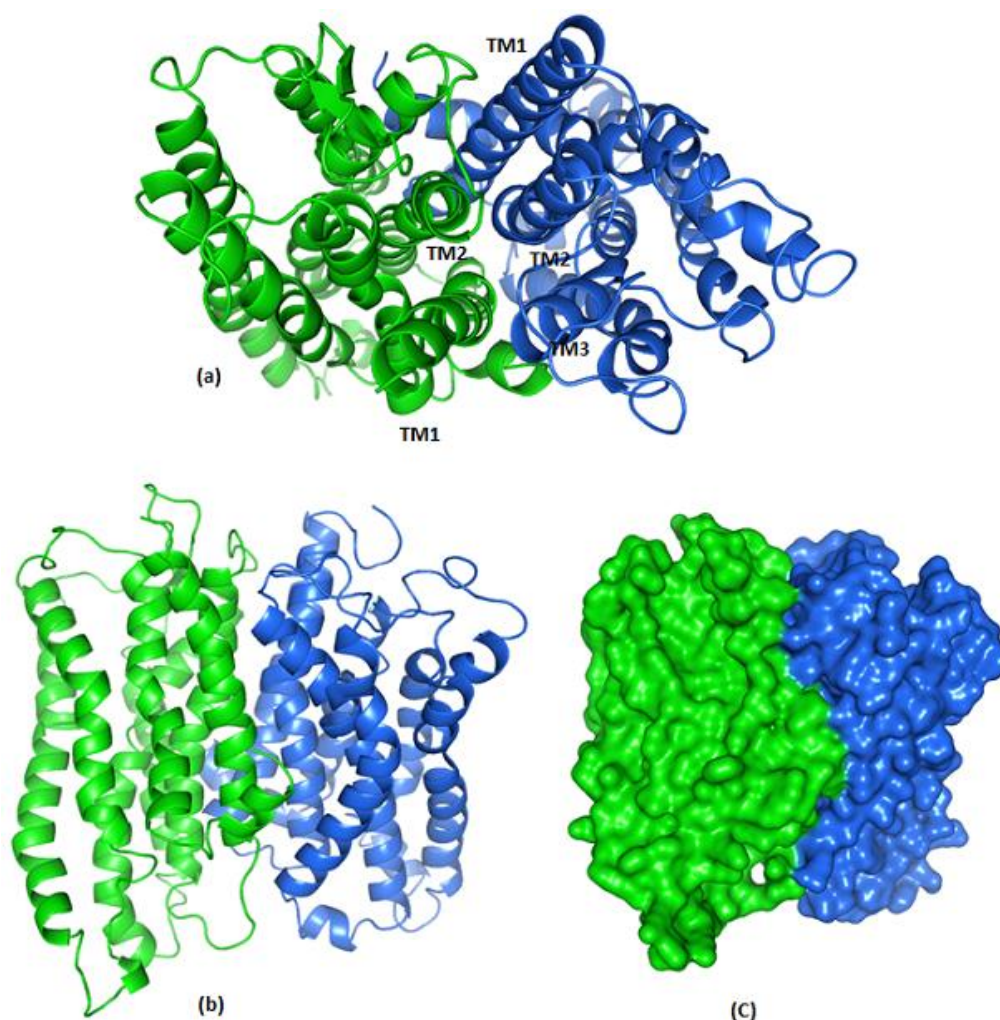

Figure 6.8: **(a)** The aerial representation of dimer 1 (TM6-TM7 interface), side view **(b)** and **(c)** molecular surface. GALR1 is shown in green and 5-HT$_{1A}$ in blue.

Thusly, in the absence of zinc, the protein-protein interaction at this interface maybe be more stable than at TM4-TM5 interface. This is more in accord with the suggested dimer intefaces of other rhodopsin class GPCRs and not with suggested homodimer interfaces of 5-HT$_{1A}$ and other 5-HT receptors subtypes.

6.7.3. Dimer interface 3: TM1-TM2

TM2 in most cases has not been considered to be a possible dimer interface, however TM1 has been strongly implicated [81-82]. With regard to putative zinc binding sites on the 5-HT1A receptor, dimerization at TM1-TM2 (Figure 9) interface will obscure site 2, site 3 and site 4.


Figure 6.9: (a) The aerial representation of dimer 1 (TM1-TM2 interface), side view (b) and (c) molecular surface. GALR1 is shown in green and 5-HT1A in blue. The TM3 of 5-HT1A receptor is also seen involved in the dimerization due to the slanted positioning of the receptor.

Furthermore, dimerization at TM1-TM2 interface present a peculiar arrangement of the TMs. Specifically TM3 as can be seen in figure 9(a), is skewed at an angle that allows the TM3 of 5-HT$_{1A}$ receptor to be exposed to towards the interfacing region but the corresponding TM3 in GALR1 is not involved. In most of contact dimers, there is parallel arrangement of the interfaces, such that as an example TM4 of monomer A, will interface with TM5 of monomer B and accordingly TM5 of monomer A will interface with TM4 of monomer B [22, 81-83, 85].

6.8 The implication of GARL1 and 5-HT$_{1A}$ heterodimer formation

As it can be deduced from both the computer modeling results presented here, and the published FRET and SPR sensorchip experimental studies, zinc impairs the heterodimerization of GARL1 and 5-HT$_{1A}$. Considering the concentration of zinc used in the FRET experiments is on par with the reported physiological concentration of zinc, a disease model was disclosed (Figure 10). Where by in a healthy cell environment where by zinc is present at a normal physiological concentration of 50µM, there is no heterodimer formation. However in a diseased cell environment such as in depressed patients where there is a zinc deficiency, heterodimers will be formed predominantly, hence the disease phenotype [53].
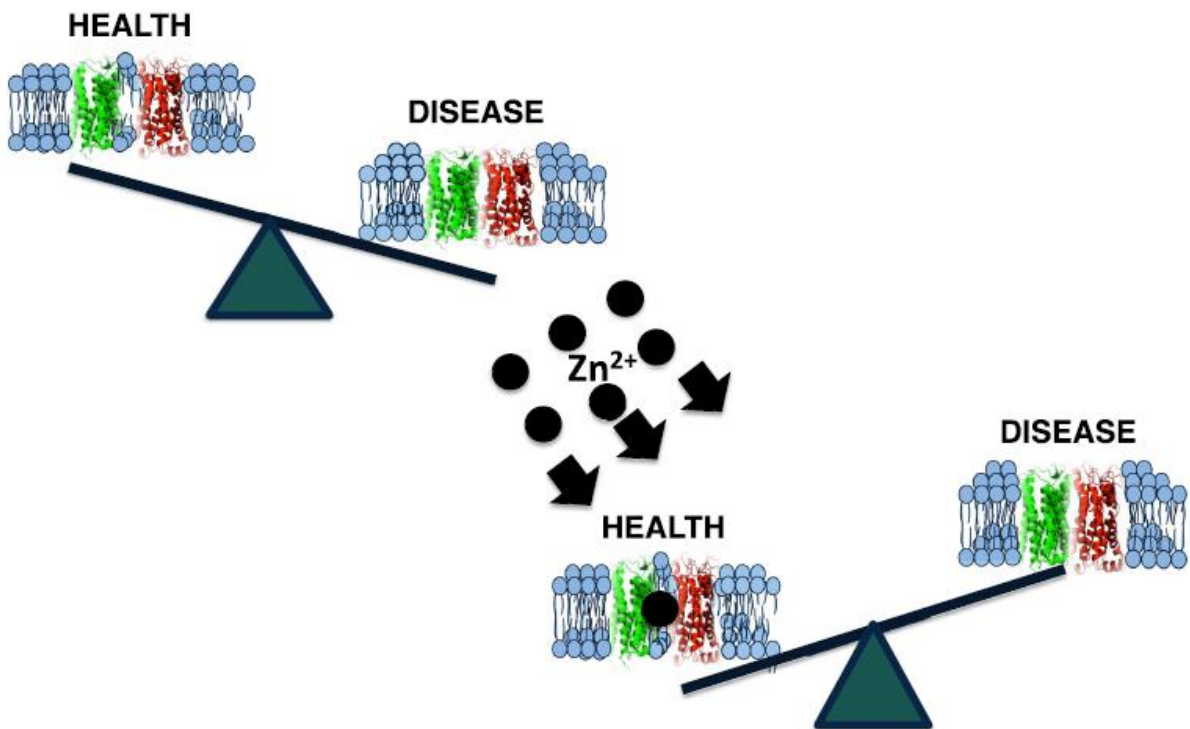
Figure 6.10: Scheme of zinc inhibition of heterodimer formation. Whereby under physiological zinc deficient conditions, the heterodimer would be the predominant form, causing a disease phenotype, while in optimal zinc condition, heterodimer formation is disrupted [53].

6.9 Conclusions to chapter 6

Considering the available extensive literature on both the existence and relevance of GPCR dimers as the therapeutical targets, it is only natural the relevance of GPCR heterodimers are explored. From the literature, the 5-HT$_{1A}$ receptor is known to be involved with major depressive disorders, moreover GARL1 receptor is known to be involved in major depressive through its modulation of the 5-HT$_{1A}$ receptor function via heterodimerization. Thusly atomistic heterodimers were created and disclosed in this chapter. The heterodimers interfaces are TM4-TM5, TM6- TM7, and TM1- TM2.

Furthermore, zinc supplementation is known to improve treatment of major depressive disorder specifically by interacting with 5-HT$_{1A}$ receptor. Hence the 5-HT$_{1A}$ receptor was probed for zinc prior to heterodimerization and 12 sites were identified. Analysis of the zinc sites and the heterodimer interfaces suggests that there is a coincidence between zinc binding sites and heterodimerization interfaces. Further information on the stability and propensity of the heterodimer interfaces was gauged. While the number of zinc sites blocked by heterodimer formation is not enough to predict the stability or transiency of the heterodimer formation, it provides a starting point for experimental studies of mutation at these putative zinc sites.

The results presented here also establish a rational explanation for the role of zinc in the molecular processes associated with heterodimer interactions and its relationship with depression, in agreement with previously reported evidence for the positive effects of zinc in depression treatment, and the involvement of the heterodimer in the same disease.

6.10   References to chapter 6

1. Pal, D. (2006). On gene ontology and function annotation. *Bioinformation*, *1*(3), 97.

2. Bouvier, M. (2001). Oligomerization of G-protein-coupled transmitter receptors. *Nature Reviews Neuroscience*, *2*(4), 274-286.

3. Milligan, G., Ramsay, D., Pascal, G., & Carrillo, J. J. (2003). GPCR dimerisation. *Life sciences*, *74*(2), 181-188.

4. Kroeger, K. M., Pfleger, K. D., & Eidne, K. A. (2003). G-protein coupled receptor oligomerization in neuroendocrine pathways. *Frontiers in neuroendocrinology*, *24*(4), 254-278.

5. Terrillon, S., & Bouvier, M. (2004). Roles of G-protein-coupled receptor dimerization. *EMBO reports*, *5*(1), 30-34.

6. Milligan, G., & Bouvier, M. (2005). Methods to monitor the quaternary structure of G protein-coupled receptors. *Febs Journal*, *272*(12), 2914-2925.

7. Milligan, G. (2006). G-protein-coupled receptor heterodimers: pharmacology, function and relevance to drug discovery. *Drug discovery today*, *11*(11), 541-549.

8. Minneman, K. P. (2007). Heterodimerization and surface localization of G protein coupled receptors. *Biochemical pharmacology*, *73*(8), 1043-1050.

9. Panetta, R., & Greenwood, M. T. (2008). Physiological relevance of GPCR oligomerization and its impact on drug discovery. *Drug discovery today*, *13*(23), 1059-1066.

10. Milligan, G. (2009). G protein-coupled receptor hetero-dimerization: contribution to pharmacology and function. *British journal of pharmacology*, *158*(1), 5-14.

11. Rozenfeld, R., & Devi, L. (2011). Exploring a role for heteromerization in GPCR signalling specificity. *Biochem. J*, *433*, 11-18.

12. Nelson, G., Chandrashekar, J., Hoon, M. A., Feng, L., Zhao, G., Ryba, N. J., & Zuker, C. S. (2002). An amino-acid taste receptor. *Nature*, *416*(6877), 199-202.

13. Pin, J. P., Kniazeff, J., Liu, J., Binet, V., Goudet, C., Rondard, P., & Prézeau, L. (2005). Allosteric functioning of dimeric class CG-protein-coupled receptors.*Febs Journal*, *272*(12), 2947-2955.

14. Romano, C., Yang, W. L., & O'Malley, K. L. (1996). Metabotropic glutamate receptor 5 is a disulfide-linked dimer. *Journal of Biological Chemistry*, *271*(45), 28612-28616.

15. Thompson, M. D., Percy, M. E., Burnham, W. M., & Cole, D. E. (2008). G protein-coupled receptors disrupted in human genetic disease. In*Pharmacogenomics in drug discovery and development* (pp. 109-137). Humana Press.

16. White, J. H., Wise, A., Main, M. J., Green, A., Fraser, N. J., Disney, G. H., ... & Marshall, F. H. (1998). Heterodimerization is required for the formation of a functional GABAB receptor. *Nature*, *396*(6712), 679-682.

17. Kuner, R., Köhr, G., Grünewald, S., Eisenhardt, G., Bach, A., & Kornau, H. C. (1999). Role of heteromer formation in GABAB receptor function. *Science*,*283*(5398), 74-77.

18. Fotiadis, D., Liang, Y., Filipek, S., Saperstein, D. A., Engel, A., & Palczewski, K. (2003). Atomic-force microscopy: rhodopsin dimers in native disc membranes. *Nature*, *421*(6919), 127-128.

19. Gurevich, V. V., & Gurevich, E. V. (2008). How and why do GPCRs dimerize?.*Trends in pharmacological sciences*, *29*(5), 234-240.

20. Berthouze, M., Ayoub, M., Russo, O., Rivail, L., Sicsic, S., Fischmeister, R., ... & Lezoualc'h, F. (2005). Constitutive dimerization of human serotonin 5-HT 4 receptors in living cells. *FEBS letters*, *579*(14), 2973-2980.

21. Filizola, M., & Weinstein, H. (2005). The study of G-protein coupled receptor oligomerization with computational modeling and bioinformatics. *Febs Journal*,*272*(12), 2926-2938.

22. Fonseca, J. M., & Lambert, N. A. (2009). Instability of a class AG protein-coupled receptor oligomer interface. *Molecular pharmacology*, *75*(6), 1296-1299.

23. Albizu, L., Holloway, T., González-Maeso, J., & Sealfon, S. C. (2011). Functional crosstalk and heteromerization of serotonin 5-HT 2A and dopamine D 2 receptors. *Neuropharmacology*, *61*(4), 770-777.

24. Fuxe, K., Borroto-Escuela, D. O., Romero-Fernandez, W., Tarakanov, A. O., Calvo, F., Garriga, P., ... & Díaz-Cabiale, Z. (2012). On the existence and function of galanin receptor heteromers

in the central nervous system. *Frontiers in endocrinology*, *3*.

25. Herrick-Davis, K. (2013). Functional significance of serotonin receptor dimerization. *Experimental brain research*, *230*(4), 375-386.

26. Fuxe, K., Marcellino, D., Rivera, A., Diaz-Cabiale, Z., Filip, M., Gago, B., ... & Agnati, L. F. (2008). Receptor–receptor interactions within receptor mosaics. Impact on neuropsychopharmacology. *Brain research reviews*, *58*(2), 415-452.

27. Borroto-Escuela, D. O., Narvaez, M., Marcellino, D., Parrado, C., Narvaez, J. A., Tarakanov, A. O., ... & Fuxe, K. (2010). Galanin receptor-1 modulates 5-hydroxtryptamine-1A signaling via heterodimerization. *Biochemical and biophysical research communications*, *393*(4), 767-772.

28. Birnbaumer, L. (2007). Expansion of signal transduction by G proteins: The second 15 years or so: From 3 to 16 α subunits plus βγ dimers. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1768*(4), 772-793.

29. Ito, H., Halldin, C., & Farde, L. (1999). Localization of 5-HTIA receptors in the living human brain using [carbonyl-" C] WAY-100635: PET with anatomic standardization technique. *J Nucl Med*, *40*, 102-109.

30. de Almeida, J., & Mengod, G. (2008). Serotonin 1A receptors in human and monkey prefrontal cortex are mainly expressed in pyramidal neurons and in a GABAergic interneuron subpopulation: implications for schizophrenia and its treatment. *Journal of neurochemistry*, *107*(2), 488-496.

31. Sullivan, K. A., Shiao, L. L., & Cascieri, M. A. (1997). Pharmacological characterization and tissue distribution of the human and rat GALR1 receptors. *Biochemical and biophysical research communications*, *233*(3), 823-828.

32. Lu, X., & Bartfai, T. (2009). Analyzing the validity of GalR1 and GalR2 antibodies using knockout mice. *Naunyn-Schmiedeberg's archives of pharmacology*, *379*(4), 417-420.

33. Lam, D. D., Garfield, A. S., Marston, O. J., Shaw, J., & Heisler, L. K. (2010). Brain serotonin system in the coordination of food intake and body weight. *Pharmacology Biochemistry and*

Behavior, *97*(1), 84-91.

34. Basky, G. (2000). Suicide linked to serotonin gene. *Canadian Medical Association Journal*, *162*(9), 1343-1343.

35. Caspi, N., Modai, I., Barak, P., Waisbourd, A., Zbarsky, H., Hirschmann, S., & Ritsner, M. (2001). Pindolol augmentation in aggressive schizophrenic patients: a double-blind crossover randomized study. *International clinical psychopharmacology*, *16*(2), 111-115.

36. Ito, M. (2009). Functional roles of neuropeptides in cerebellar circuits.*Neuroscience*, *162*(3), 666-672.

37. Razani, H., Diaz-Cabiale, Z., Fuxe, K., & Ögren, S. O. (2000). Intraventricular galanin produces a timedependent modulation of 5-HT1A receptors in the dorsal raphe of the rat. *Neuroreport*, *11*(18), 3943-3948.

38. Kehr, J., Yoshitake, T., Wang, F. H., Razani, H., Gimenez-Llort, L., Jansson, A., ... & Ögren, S. O. (2002). Galanin is a potent in vivo modulator of mesencephalic serotonergic neurotransmission. *Neuropsychopharmacology*,*27*(3), 341-356.

39. Yoshitake, T., Yoshitake, S., Yamaguchi, M., Ögren, S. O., & Kehr, J. (2003). Activation of 5-HT 1A autoreceptors enhances the inhibitory effect of galanin on hippocampal 5-HT release in vivo. *Neuropharmacology*, *44*(2), 206-213.

40. Razani, H., Díaz-Cabiale, Z., Misane, I., Wang, F. H., Fuxe, K., & Ögren, S. O. (2001). Prolonged effects of intraventricular galanin on a 5-hydroxytryptamine 1A receptor mediated function in the rat. *Neuroscience letters*, *299*(1), 145-149.

41. Razani, H., Hedlund, P. B., Hansson, A. C., Fuxe, K., & Ögren, S. O. (1997). 5-HT1A receptor activation: short-term effects on the mRNA expression of the 5-HT1A receptor and galanin in the raphe nuclei. *Neuroreport*, *8*(16), 3565-3569.

42. Fuxe, K., Hedlund, P., Von Euler, G., & Lundgren, K. (1991). Galanin/5-HT interactions in the rat central nervous system. Relevance for depression. In *WENNER GREN CENTER INTERNATIONAL SYMPOSIUM SERIES* (pp. 221-221). MACMILLAN ACADEMIC AND PROFESSIONAL LTD.

43. Salim, K., Fenton, T., Bacha, J., Urien-Rodriguez, H., Bonnert, T., Skynner, H. A., ... & Guest, P. C. (2002). Oligomerization of G-protein-coupled receptors shown by selective co-immunoprecipitation. *Journal of Biological Chemistry*, *277*(18), 15482-15485.

44. Sánchez-Blázquez, P., Rodríguez-Muñoz, M., Bailón, C., & Garzón, J. (2012). GPCRs promote the release of zinc ions mediated by nNOS/NO and the redox transducer RGSZ2 protein. *Antioxidants & redox signaling*, *17*(9), 1163-1177.

45. Lagerström, M. C., Klovins, J., Fredriksson, R., Fridmanis, D., Haitina, T., Ling, M. K., ... & Schiöth, H. B. (2003). High affinity agonistic metal ion binding sites within the melanocortin 4 receptor illustrate conformational change of transmembrane region 3. *Journal of Biological Chemistry*, *278*(51), 51521-51526.

46. Barrondo, S., & Sallés, J. (2009). Allosteric modulation of 5-HT 1A receptors by zinc: Binding studies. *Neuropharmacology*, *56*(2), 455-462.

47. Liu, W., Chun, E., Thompson, A. A., Chubukov, P., Xu, F., Katritch, V., ... & Stevens, R. C. (2012). Structural basis for allosteric regulation of GPCRs by sodium ions. *Science*, *337*(6091), 232-236.

48. Besser, L., Chorin, E., Sekler, I., Silverman, W. F., Atkin, S., Russell, J. T., & Hershfinkel, M. (2009). Synaptically released zinc triggers metabotropic signaling via a zinc-sensing receptor in the hippocampus. *The Journal of Neuroscience*, *29*(9), 2890-2901.

49. Hershfinkel, M., Silverman, W. F., & Sekler, I. (2007). The zinc sensing receptor, a link between zinc and cell signaling. *Molecular Medicine*, *13*(7-8), 331.

50. Parkel, S., Näsman, J., & Rinken, A. (2009). Enhancement of agonist binding to 5-HT 1A receptors in rat brain membranes by millimolar Mn 2+. *Neuroscience letters*, *457*(1), 32-35.

51. Siwek, M., Dudek, D., Paul, I. A., Sowa-Kućma, M., Zięba, A., Popik, P., ... & Nowak, G. (2009). Zinc supplementation augments efficacy of imipramine in treatment resistant patients: a double blind, placebo-controlled study. *Journal of affective disorders*, *118*(1), 187-195.

52. Siwek, M., Dudek, D., Schlegel-Zawadzka, M., Morawska, A., Piekoszewski, W., Opoka, W., ... & Nowak, G. (2010). Serum zinc level in depressed patients during zinc supplementation of

imipramine treatment. *Journal of affective disorders*, *126*(3), 447-452.

53. Tena-Campos, M., Ramon, E., Lupala, C. S., Pérez, J. J., Koch, K. W., & Garriga, P. (2015). Zinc is involved in depression by modulating G-protein coupled receptor Heterodimerization. *Molecular neurobiology*, 1-13.

54. Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G. W., Vardy, E., ... & Stevens, R. C. (2012). Structure of the human [kgr]-opioid receptor in complex with JDTic. *Nature*, *485*(7398), 327-332.

55. Wang, C., Jiang, Y., Ma, J., Wu, H., Wacker, D., Katritch, V., ... & Xu, H. E. (2013). Structural basis for molecular recognition at serotonin receptors. *Science*, *340*(6132), 610-614.

56. Šali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*, *234*(3), 779-815.

57. Leonardi, A., Guarneri, L., Poggesi, E., Angelico, P., Velasco, C., Cilia, A., & Testa, R. (2001). N-[2-[4-(2-methoxyphenyl)-1-piperazinyl] ethyl]-N-(2-nitrophenyl) cyclohexanecarboxamide: a novel pre-and postsynaptic 5-hydroxytryptamine1A receptor antagonist active on the lower urinary tract.*Journal of Pharmacology and Experimental Therapeutics*, *299*(3), 1027-1037.

58. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of medicinal chemistry*, *47*(7), 1750-1759.

59. Cordomí, A., & Perez, J. J. (2007). Molecular dynamics simulations of rhodopsin in different one-component lipid bilayers. *The Journal of Physical Chemistry B*, *111*(25), 7052-7063.

60. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *Journal of computational chemistry*, *26*(16), 1701-1718.

61. Tovchigrechko, A., & Vakser, I. A. (2005). Development and testing of an automated approach to protein docking. *Proteins: Structure, Function, and Bioinformatics*, *60*(2), 296-301.

62. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., & Vakser, I. A. (1992). Molecular surface recognition: determination of geometric fit between proteins and their

ligands by correlation techniques. *Proceedings of the National Academy of Sciences*, *89*(6), 2195-2199.

63. Vakser, I. A., Matar, O. G., & Lam, C. F. (1999). A systematic study of low-resolution recognition in protein–protein complexes. *Proceedings of the National Academy of Sciences*, *96*(15), 8477-8482.

64. Ly, N. (2011). Evaluating Protein-Protein Docking Web Servers. biochem218. stanford.edu

65. Tovchigrechko, A., & Vakser, I. A. (2006). GRAMM-X public web server for protein–protein docking. Nucleic acids research, 34(suppl 2), W310-W314.

66. *Molecular Operating Environment (MOE)*, 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2015.

67. Krämer, U., Cotter-Howells, J. D., Charnock, J. M., Baker, A. J., & Smith, J. A. C. (1996). Free histidine as a metal chelator in plants that accumulate nickel.

68. Hall, J. L. (2002). Cellular mechanisms for heavy metal detoxification and tolerance. Journal of experimental botany, 53(366), 1-11.

69. Ferré, S., Casadó, V., Devi, L. A., Filizola, M., Jockers, R., Lohse, M. J., ... & Guitart, X. (2014). G protein–coupled receptor oligomerization revisited: functional and pharmacological perspectives. *Pharmacological reviews*, *66*(2), 413-434.

70. Wu, B., Chien, E. Y., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., ... & Stevens, R. C. (2010). Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science*, *330*(6007), 1066-1071.

71. Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Mathiesen, J. M., Sunahara, R. K., ... & Granier, S. (2012). Crystal structure of the [micro]-opioid receptor bound to a morphinan antagonist. *Nature*, *485*(7398), 321-326.

72. Ruprecht, J. J., Mielke, T., Vogel, R., Villa, C., & Schertler, G. F. (2004). Electron crystallography reveals the structure of metarhodopsin I. *The EMBO journal*, *23*(18), 3609-3620.

73. Salom, D., Lodowski, D. T., Stenkamp, R. E., Le Trong, I., Golczak, M., Jastrzebska, B., ... &

Palczewski, K. (2006). Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *Proceedings of the National Academy of Sciences*, *103*(44), 16123-16128.

74. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W., & Ernst, O. P. (2008). Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature*,*454*(7201), 183-187.

75. Huang, J., Chen, S., Zhang, J. J., & Huang, X. Y. (2013). Crystal structure of oligomeric β1-adrenergic G protein–coupled receptors in ligand-free basal state.*Nature structural & molecular biology*, *20*(4), 419-425.

76. Wang, C., Wu, H., Katritch, V., Han, G. W., Huang, X. P., Liu, W., ... & Stevens, R. C. (2013). Structure of the human smoothened receptor bound to an antitumour agent. *Nature*, *497*(7449), 338-343.

77. Guo, W., Urizar, E., Kralikova, M., Mobarec, J. C., Shi, L., Filizola, M., & Javitch, J. A. (2008). Dopamine D2 receptors form higher order oligomers at physiological expression levels. *The EMBO journal*, *27*(17), 2293-2304.

78. Mancia, F., Assur, Z., Herman, A. G., Siegel, R., & Hendrickson, W. A. (2008). Ligand sensitivity in dimeric associations of the serotonin 5HT2c receptor.*EMBO reports*, *9*(4), 363-369.

79. Kruse, A. C., Hu, J., Pan, A. C., Arlow, D. H., Rosenbaum, D. M., Rosemond, E., ... & Kobilka, B. K. (2012). Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature*, *482*(7386), 552-556.

80. Periole, X., Knepp, A. M., Sakmar, T. P., Marrink, S. J., & Huber, T. (2012). Structural determinants of the supramolecular organization of G protein-coupled receptors in bilayers. *Journal of the American Chemical Society*, *134*(26), 10959-10965.

81. Johnston, J. M., Wang, H., Provasi, D., & Filizola, M. (2012). Assessing the relative stability of dimer interfaces in g protein-coupled receptors. *PLoS Comput Biol*, *8*(8), e1002649-e1002649.

82. Fung, J. J., Deupi, X., Pardo, L., Yao, X. J., Velez-Ruiz, G. A., DeVree, B. T., ... & Kobilka, B. K. (2009). Ligand-regulated oligomerization of β2-adrenoceptors in a model lipid bilayer. *The EMBO journal*, *28*(21), 3315-3328.

83. Gorinski, N., Kowalsman, N., Renner, U., Wirth, A., Reinartz, M. T., Seifert, R., ... & Niv, M. Y. (2012). Computational and experimental analysis of the transmembrane domain 4/5 dimerization interface of the serotonin 5-HT1A receptor. Molecular pharmacology, 82(3), 448-463.

84. Provasi, D., Boz, M. B., Johnston, J. M., & Filizola, M. (2015). Preferred Supramolecular Organization and Dimer Interfaces of Opioid Receptors from Simulated Self-Association. *PLoS computational biology*, *11*(3), e1004148-e1004148.

85. Wacker, D., Wang, C., Katritch, V., Han, G. W., Huang, X. P., Vardy, E., ... & Stevens, R. C. (2013). Structural features for functional selectivity at serotonin receptors. *Science*, *340*(6132), 615-619.

86. Cordomí, A., & Perez, J. J. (2009). Structural rearrangements of rhodopsin subunits in a dimer complex: a molecular dynamics simulation study. *Journal of Biomolecular Structure and Dynamics*, *27*(2), 127-147.

General conclusions

1. The results of this work support that molecular dynamics in an adequate methodology to refine the structures of GPCRs constructed by homology modeling. Specifically, the rmsd of the transmembrane region of constructed models compared to the target receptor improves with the refinement process. Moreover, higher accuracy is obtained when the template used is closer to the target in the phylogenetic tree in the sampling times used in this work.

2. The use of a ligand in the molecular dynamics refinement process makes convergence time to be shorter and yields more accurate models for the transmembrane region than in the absence of a ligand.

3. The model constructed in this work of the bradykinin B1 receptor together with the docking studies of several selective antagonist described in the literature permitted the definition of a pharmacophore. This consists of five points that defined on ligand features correspond to a proton donor (point 1); a proton acceptor/donor (point 2); a hydrophobic ring (point 3); a proton acceptor/ donor (point 4) and proton donor/ acceptor (point 5).

4. The pharmacophore was used to guide a virtual screening process. The results permitted to identify a set of compounds some of which were purchased and *in vitro* tested for their capability to antagonize the B1R. These results give support to the validity of the bradykinin B1 receptor pharmacophoric hypothesis described in this work.

5. The model constructed in this work of the bradykinin B2 receptor together with the docking studies of several selective antagonist described in the literature permitted the definition of a pharmacophore. This consists of five points that defined on ligand features correspond to a proton donor (point 1); a proton acceptor/donor (point 2); a hydrophobic ring (point 3); a proton acceptor/ donor (point 4) and hydrophobic/aromatic moiety (point 5).

6. The pharmacophore was used to guide a virtual screening process. The results permitted to identify a set of compounds some of which were purchased and in vitro tested for their capability to antagonize the B2R. These results give support to the validity of the bradykinin B2 receptor pharmacophoric hypothesis described in this work.

7. Analysis of the pharmacophores for B1R and B2R permitted to get insight on the features responsible for receptor selectivity. Specifically, the key difference lies on pharmacophore point 5 that exhibits different nature on both receptors.

8. The model constructed in this work of the serotonin 5-HT$_{1A}$ receptor permitted the identification of 12 prospective sites for zinc binding. Most of the putative binding sites are located on the extracellular and intracellular loops, although some sites are located on the transmembrane regions including TM1; on the interface TM4 -TM5; on the interface TM6-TM7 and TM7-H8.

9. The models constructed in this work of the serotonin 5-HT$_{1A}$ and GAL1 receptors were used to build heterodimers with diverse possible interfaces. Analysis of the models together with the results of zinc binding to the serotonin 5-HT$_{1A}$ receptor permitted to identify the interface TM6-TM7 as most likely for heterodimerization.

Annex

8.1 Sequence Alignments

8.1.1 Sequence alignments

```
4DAJ_A|R.norvegicus|CHRM3    ----DPLGGHTIWQVVFIAFLTGFLALVTIIGNILVIVAFKVNKQLKTVN
sp|H.sapiens|CHRM3           ----DPLGGHTVWQVVFIAFLTGILALVTIIGNILVIVSFKVNKQLKTVN
3UON_A|H.sapiens|CHRM2       ---LALTSPYKTFEVVFIVLVAGSLSLVTIIGNILVMVSIKVNRHLQTVN
3RZE_A|Histamine             -------TTMASPQLMPLVVVLSTICLVTVGLNLLVLYAVRSERKLHTVG
1GZM_A|RHO                   PFEAPQYYLAEPWQFSMLAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPL
                                 :.   :.        : ::  .  *:*.:      .::*:*

4DAJ_A|R.norvegicus|CHRM3    NYFLLSLACADLIIGVISMNLFTTYIIMNRWALGNLACDLWLSIDYVASN
sp|H.sapiens|CHRM3           NYFLLSLACADLIIGVISMNLFTTYIIMNRWALGNLACDLWLAIDYVASN
3UON_A|H.sapiens|CHRM2       NYFLFSLACADLIIGVFSMNLYTLYTVIGYWPLGPVVCDLWLALDYVVSN
3RZE_A|Histamine             NLYIVSLSVADLIVGAVVMPMNILYLLMSKWSLGRPLCLFWLSMDYVAST
1GZM_A|RHO                   NYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLGGE
                             *   :..*: ***::       :          : .: :*   * :   : : .

4DAJ_A|R.norvegicus|CHRM3    ASVMNLLVISFDRYFSITRPLTYRAKRTTKRAGVMIGLAWVISFVLWAPA
sp|H.sapiens|CHRM3           ASVMNLLVISFDRYFSITRPLTYRAKRTTKRAGVMIGLAWVISFVLWAPA
3UON_A|H.sapiens|CHRM2       ASVMNLLIISFDRYFCVTKPLTYPVKRTTKMAGMMIAAAWVLSFILWAPA
3RZE_A|Histamine             ASIFSVFILCIDRYRSVQQPLRYLKYRTKTRASATILGAWFLSFLWVIPI
1GZM_A|RHO                   IALWSLVVLAIERYVVVCKPMSNFRFG-ENHAIMGVAFTWVMALACAAPP
                              ::  .:.::.::**  :  :*:         . *   :   :*.:::    *

4DAJ_A|R.norvegicus|CHRM3    ILFWQYFVGKRTVPPGECFIQFLS------EPTITFGTAIAAFYMPVTIM
sp|H.sapiens|CHRM3           ILFWQYFVGKRTVPPGECFIQFLS------EPTITFGTAIAAFYMPVTIM
3UON_A|H.sapiens|CHRM2       ILFWQFIVGVRTVEDGECYIQFFS------NAAVTFGTAIAAFYLPVIIM
3RZE_A|Histamine             LGWNHFMQQTSVRREDKCETDFYD------VTWFKVMTAIINFYLPTLLM
1GZM_A|RHO                   LVGWSRYIPEGMQCS--CGIDYYTPHEETNNESFVIYMFVVHFIIPLIVI
                             :                *  ::            . .   :   * :*  ::

4DAJ_A|R.norvegicus|CHRM3    TILYWRIYKETEKMLIK-----------EKKAAQTLSAILLAFIITWTPYN
sp|H.sapiens|CHRM3           TILYWRIYKETEKSLVK-----------EKKAAQTLSAILLAFIITWTPYN
3UON_A|H.sapiens|CHRM2       TVLYWHISRASKSRIPPP-------SREKKVTRTILAILLAFIITWAPYN
3RZE_A|Histamine             LWFYAKIYKAVRQHCLHM-------NRERKAAKQLGFIMAAFILCWIPYF
1GZM_A|RHO                   FFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYA
                                * ::     ..                  *::.:: :   :: **:: * **

4DAJ_A|R.norvegicus|CHRM3    -IMVLVNTFCDSCIPKTYWNLGYWLCYINSTVNPVCYALCNKTFRTTFKT
sp|H.sapiens|CHRM3           -IMVLVNTFCDSCIPKTFWNLGYWLCYINSTVNPVCYALCNKTFRTTFKM
3UON_A|H.sapiens|CHRM2       -VMVLINTFCAPCIPNTVWTIGYWLCYINSTINPACYALCNATFKKTFKH
3RZE_A|Histamine             -IFFMVIAFCKNCCNEHLHMFTIWLGYINSTLNPLIYPLCNENFKKTFKR
1GZM_A|RHO                   GVAFYIFTHQGSDFGPIFMTIPAFFAKTSAVYNPVIYIMMNKQFRNCMVT
                              :  .: :.          :  ::   .:. **  * : *   *:. :
```

Figure 8.1: Sequence alignment of the target selected templates (muscarinic M2, Histamine H1 and Rhodopsin receptor), and the target muscarinic M3 receptor.

```
4DAJ_A|R.norvegicus|CHRM3    DPLGGHTIWQVVFIAFLTGFLALVTIIGNILVIVAFKVNKQLKTVNNYFL
sp|H.sapiens|CHRM3           DPLGGHTVWQVVFIAFLTGILALVTIIGNILVIVSFKVNKQLKTVNNYFL
                            *******:************:***************:*************

4DAJ_A|R.norvegicus|CHRM3    LSLACADLIIGVISMNLFTTYIIMNRWALGNLACDLWLSIDYVASNASVM
sp|H.sapiens|CHRM3           LSLACADLIIGVISMNLFTTYIIMNRWALGNLACDLWLAIDYVASNASVM
                            *************************************:***********

4DAJ_A|R.norvegicus|CHRM3    NLLVISFDRYFSITRPLTYRAKRTTKRAGVMIGLAWVISFVLWAPAILFW
sp|H.sapiens|CHRM3           NLLVISFDRYFSITRPLTYRAKRTTKRAGVMIGLAWVISFVLWAPAILFW
                            **************************************************

4DAJ_A|R.norvegicus|CHRM3    QYFVGKRTVPPGECFIQFLSEPTITFGTAIAAFYMPVTIMTILYWRIYKE
sp|H.sapiens|CHRM3           QYFVGKRTVPPGECFIQFLSEPTITFGTAIAAFYMPVTIMTILYWRIYKE
                            **************************************************

4DAJ_A|R.norvegicus|CHRM3    TEKMLIKEKKAAQTLSAILLAFIITWTPYNIMVLVNTFCDSCIPKTYWNL
sp|H.sapiens|CHRM3           TEKSLVKEKKAAQTLSAILLAFIITWTPYNIMVLVNTFCDSCIPKTFWNL
                            *** *:*************************************:.***

4DAJ_A|R.norvegicus|CHRM3    GYWLCYINSTVNPVCYALCNKTFRTTFKTLLLCQCDKRKRRKQQYQQRQS
sp|H.sapiens|CHRM3           GYWLCYINSTVNPVCYALCNKTFRTTFKMLLLCQCDKKKRRKQQYQQRQS
                            ***************************.********:.*************

4DAJ_A|R.norvegicus|CHRM3    VIFHKRVPEQAL
sp|H.sapiens|CHRM3           VIFHKRAPEQAL
                            ******.*****
```

Figure 8.2: Sequence alignment between rat and human CHRM3.

```
4IAR_A|PDBID|CHAIN|SEQUENCE      ----------------GGTCSAKDYIYQDSIS LPWKVLLVMLLALITLAT
sp|P08908|5HT1A_HUMAN            MDVLSPGQGNNTTSPPAPFETGGNTTGISDVT VSYQVITSLLLGTLIFCA
                                           .    :. :     ..::::.::*:  :**. : :.:

4IAR_A|PDBID|CHAIN|SEQUENCE      TLSNAFVIATVYR TRKLHT PANYLIASLAVTDLLVSILVMPISTMYTVT G
sp|P08908|5HT1A_HUMAN            VLGNACVVAAIAL ERSLQN VANYLIGSLAVTDLMVSVLVLPMAALYQVL N
                                 .*.** *:*::    *.*:. ****.*******:**:**:*::::* * .

4IAR_A|PDBID|CHAIN|SEQUENCE      RWTL GQVVCDFWLSSDITCCTASIWHLCVIALDRYWAITDAVEYS AKRTP
sp|P08908|5HT1A_HUMAN            KWTL GQVTCDLFIALDVLCCTSSILHLCAIALDRYWAITDPIDYV NKRTP
                                 :******.**:::: *: ***:** ***.***********..:*   ****

4IAR_A|PDBID|CHAIN|SEQUENCE      KRAAVMIALVWVFSISISLPPFFWRQAK AEEEVSECVVNTDHI LYTVYST
sp|P08908|5HT1A_HUMAN            RRAAALISLTWLIGFLISIPPMLGWRTP EDRSDPDACTISKDH GYTIYST
                                 :***.:*:*.*:::: **:**::   ::  :.. .:. . :..  **:***

4IAR_A|PDBID|CHAIN|SEQUENCE      VGAFYFPTLLLIALYGRIYVEARS --------------------------
sp|P08908|5HT1A_HUMAN            FGAFYIPLLLMLVLYGRIFRAARF RIRKTVKKVEKTGADTRHGASPAPQP
                                 .****:* **::.*****:   **

4IAR_A|PDBID|CHAIN|SEQUENCE      --------------------------------------------------
sp|P08908|5HT1A_HUMAN            KKSVNGESGSRNWRLGVESKAGGALCANGAVRQGDDGAALEVIEVHRVGN

4IAR_A|PDBID|CHAIN|SEQUENCE      --------------------------------RIAARER KATKTLGII
sp|P08908|5HT1A_HUMAN            SKEHLPLPSEAGPTPCAPASFERKNERNAEAKRKMALARER KTVKTLGII
                                                                          *****:.******

4IAR_A|PDBID|CHAIN|SEQUENCE      LGAFIVCWLPFFIISLV MPICKDACWFHL AIFDFFTWLGYLNSLINPIIY
sp|P08908|5HT1A_HUMAN            MGTFILCWLPFFIVALV LPFCESSCHMPT LLGAIINWLGYSNSLLNPVIY
                                 :*:**:*******::**:*:*:.:* :   :   ::.**** ***:**:**

4IAR_A|PDBID|CHAIN|SEQUENCE      TMS NEDFKQAFHKLIRFKCTS-
sp|P08908|5HT1A_HUMAN            AYF NKDFQNAFKKIIKCKFCRQ
                                 :   *:**::**:*:*: *
```

Figure 8.3: Sequence alignment between the template, crystallographic structure of 5HT$_{1B}$ and the target 5HT$_{1A}$ showing an overall sequence homology of 47%. The blue blocks represent the TM residues

```
4DJH_A|PDBID|CHAIN|SEQUENCE      --------------GGTTMGSEDAQLEPAHISPAIPVIITAVYSVVFVVG
sp|P47211|GALR1_HUMAN            MELAVGNLSEGNASWPEPPAPEPGPLFGIGVENFVTLVVFGLIFALGVLG
                                          . ..* . *       :.  :.:::  .:   .: *:*

4DJH_A|PDBID|CHAIN|SEQUENCE      LVGNSLVMFVIIRYTKMKTATNIYIFNLALADALVTTTMPFQSTVYLMNS
sp|P47211|GALR1_HUMAN            NSLVITVLARSKPGKPRSTTNLFILNLSIADLAYLLFCIPFQATVYALPT
                                        *:      .   .*:. : :        * :   :***:*** : :

4DJH_A|PDBID|CHAIN|SEQUENCE      WPFGDVLCKIVLSIDYYNMFTSIFTLTMMSVDRYIAVCHPVKALDFRTPL
sp|P47211|GALR1_HUMAN            WVLGAFICKFIHYFFTVSMLVSIFTLAAMSVDRYVAIVHSRRSSSLRVSR
                                 * :* .:.:**::  :    .*:.*****: ******:*: *. :: .:*..

4DJH_A|PDBID|CHAIN|SEQUENCE      KAKIINICIWLLSSSVGISAIVLGGTKVREDVDVIECSLQFPDDDYSWWD
sp|P47211|GALR1_HUMAN            NALLGVGCIWALSIAMASPVAYHQGLFHPRASNQTFCWEQWPDP---RHK
                                 :* :    *** ** ::. ..     *     . :    *   *:.**   .

4DJH_A|PDBID|CHAIN|SEQUENCE      LFMKICVFIFAFVIPVLIIIVCYTLMILRLKSVRLLSGREKDRNLRRITR
sp|P47211|GALR1_HUMAN            KAYVVCTFVFGYLLPLLLICFCYAKVLNHLHKKLKNMSKKSEASKKKTAQ
                                     :*.*:*.:::*:*:* .**: :: :*:.    .::.: . :: ::

4DJH_A|PDBID|CHAIN|SEQUENCE      LVLVVVAVFVVCWTPIHIFILVEALGSTSHSTAALSSYYFCIALGYTNSS
sp|P47211|GALR1_HUMAN            TVLVVVVVFGISWLPHHIIHLWAEFGVFPLTPASFLFRITAHCLAYSNSS
                                  *****.** :.* * **: *     :*  . .:.*::     . .*.*:***

4DJH_A|PDBID|CHAIN|SEQUENCE      LNPILYAFLDENFKRCFRDFCFPLKMRMERQSTS----------------
sp|P47211|GALR1_HUMAN            VNPIIYAFLSENFRKAYKQVFKCHIRKDSHLSDTKESKSRIDTPPSTNCT
                                 :***:****.***::.:::.      : .: * :

4DJH_A|PDBID|CHAIN|SEQUENCE      --
sp|P47211|GALR1_HUMAN            HV
```

Figure 8.4: Sequence alignment between the template, crystallographic structure of κ-opioid receptor and the target GALR1 showing an overall sequence homology of 28%. The green blocks represent the TM residues.

8.2 Radio-ligand binding assays

8.2.1 BKRB1 radio-ligand Binding assays

B1R antagonism assays were carried out following a protocol described elsewhere [1]. Specifically, compounds were tested on human recombinant B1R expressed in CHO cells. Saturation isotherms were obtained with [^{3}H] des-Arg^{10}- (0.35 nM) incubated for 60 minutes at room temperature. Nonspecific binding was evaluated by adding bradykinin at 1 mM. Antagonism was measured by the capacity of a specific compound to displace the radio-ligand using the scintillation counting method. Scintillation counting refers to the the measurement of activity of a sample of radioactive material which uses the technique of mixing the active material with a liquid scintillator, and counting the resultant photon emissions.

8.2.2 BKRB2 radio-ligand Binding assays

B2R antagonism assays were carried out following a protocol described elsewhere [2]. Specifically, compounds were tested on human recombinant B2R expressed in CHO cells. Saturation isotherms were obtained with [^{3}H]-bradykinin (0.30 nM) incubated for 60 minutes at room temperature. Nonspecific binding was evaluated by adding bradykinin at 1 mM. Antagonism was measured by the capacity of a specific compound to displace the radio-ligand using the scintillation counting method.

1. Jones, C., Phillips, E., Davis, C., Arbuckle, J., Yaqoob, M., Burgess, G. M., ... & McIntyre, P. (1999). Molecular characterisation of cloned bradykinin B 1 receptors from rat and human. European journal of pharmacology, 374(3), 423-433.

2. Pruneau D, Luccarini J-M, Fouchet C, Defrene E, Franck R-M, Loillier B, Duclos H, Robert C, Cremers B, Belichard P, Paquet J-L. (1998) LF 16.0335, a novel potent and selective nonpeptide antagonist of the human bradykinin B2 receptor. British J Pharmacol 125: 365-372.

8.3 Analysis of the MD trajectories

8.3.1 Dihedral angles

Dihedral angle for the ECL regions of M3 models along their trajectories. The residues are identified according to the M3 muscarinic crystal structure, thus below pictures depict the dihedral angles of the residue in the models that makes up the ECL region in the M3 crystal structure.

- ECL1 residue # 131-136 (ECL1-1: N-R-W-A, ECL1-2: R-W-A-L and ECL1-3: W-A-L-G)

- ECL2 residue # 214-222 (ECL2-1 : T-V- P-P, ECL2-2: V- P-P-G, ECL2-3: P-P-G-E, ECL2-4: P-G-E-C, ECL2-5: G-E-C-F, and ECL2-6: E-C-F-I)

- ECL3 residue # 515-520 (ECL3-1: F-C-D-S, ECL3-2: C-D-S-C and ECL3-3: D-S-C-I)
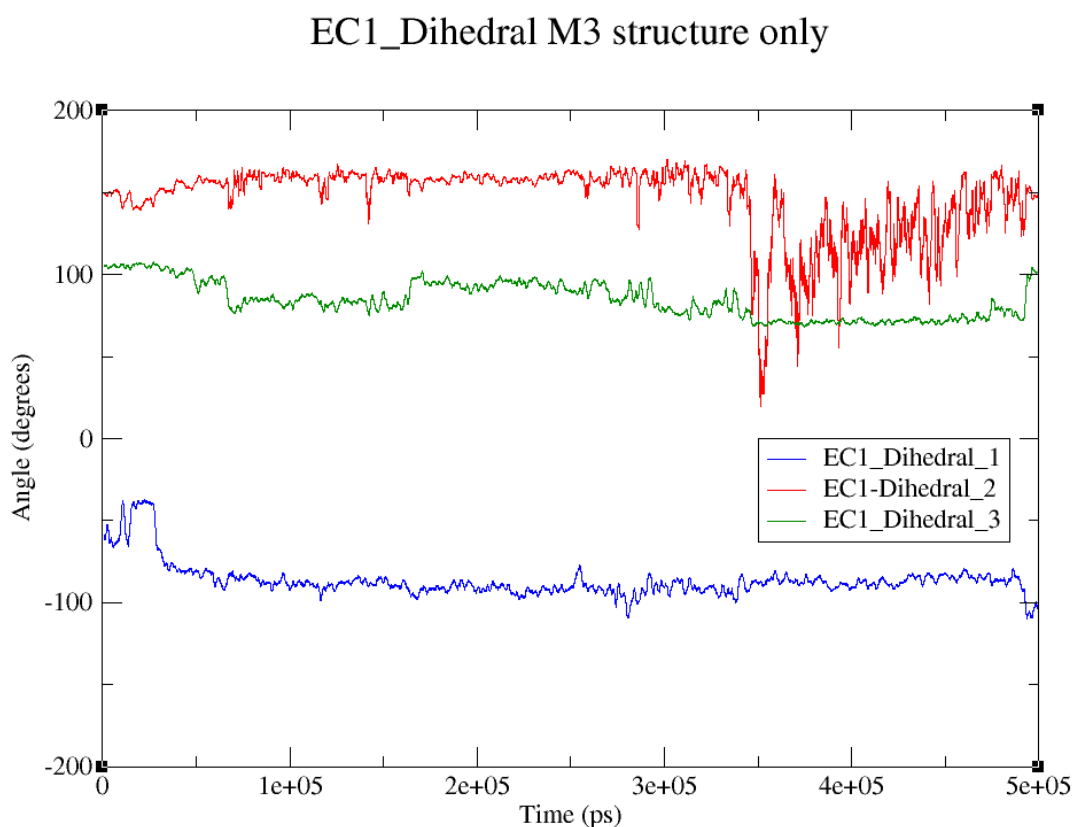

(a)

EC1_Dihedrals M3-NMS
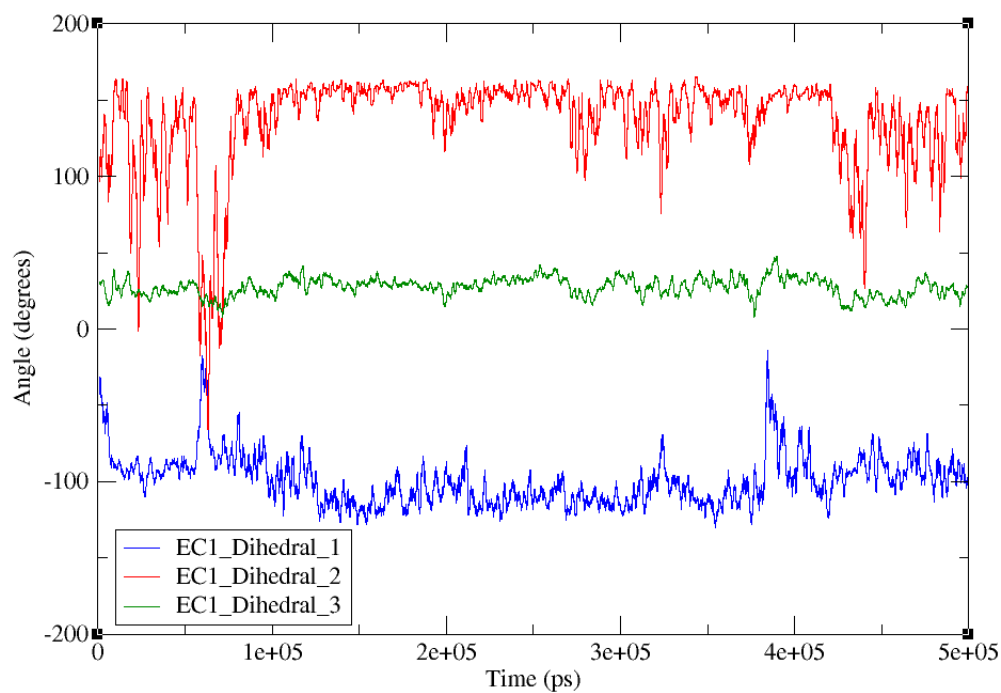
(b)

EC1_Dihedral M3-Tiotropium
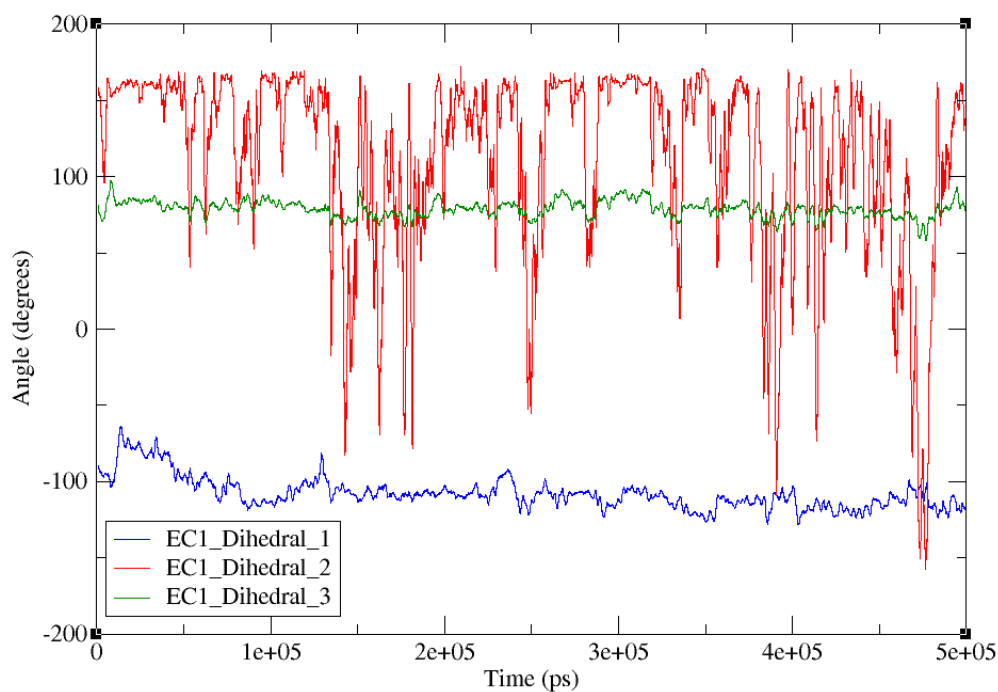
(c)

EC1 Dihedrals M3-Histamine H1-Tio

(d)
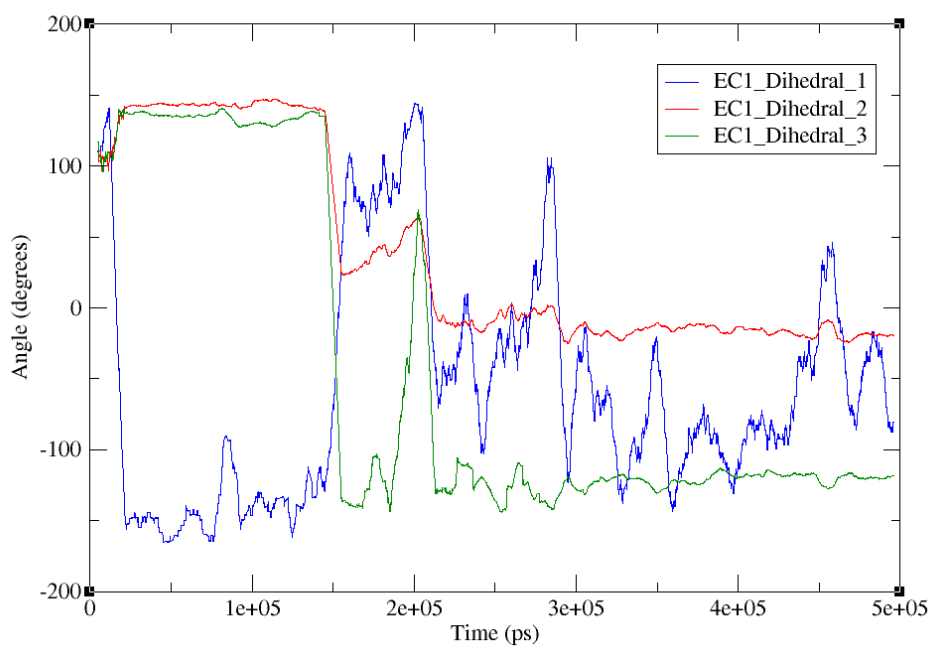
ECI Dihedrals M3-Rhodopsin-Tio
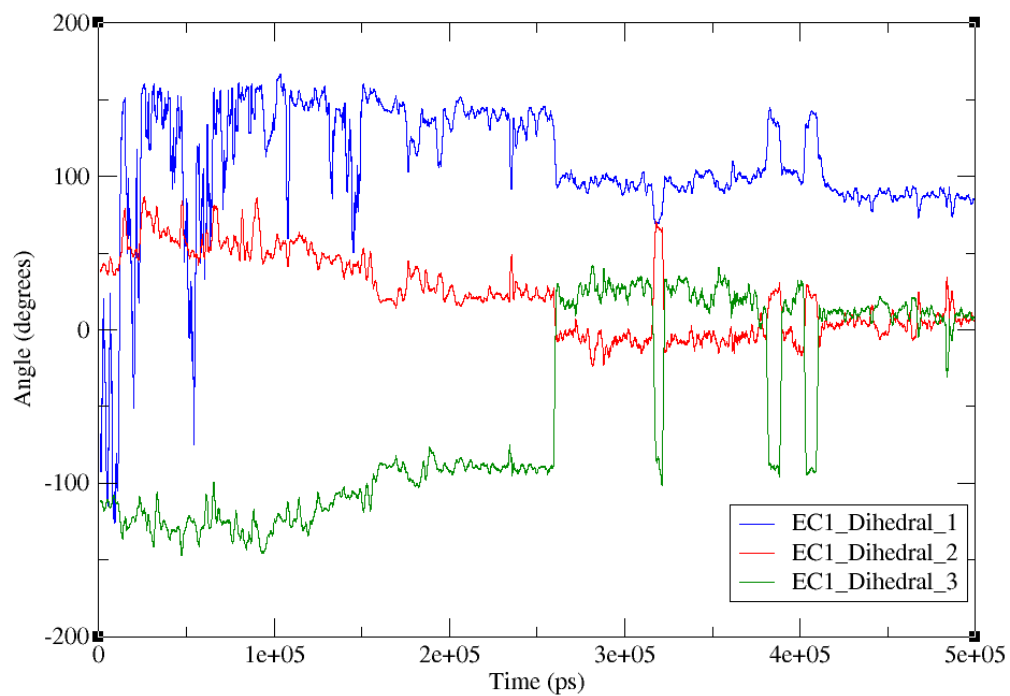
(e)

EC2 Dihedral M3 structure only

(f)

EC2 Dihedrals M3-NMS

(g)

EC2 Dihedrals M3-Tiotropium

(h)

EC2 Dihedrals M3 Histamine H1-Tio

(i)

EC2 Dihedrals M3-Rhodopsin-Tio

(j)

EC3_Dihedral M3 structure Only
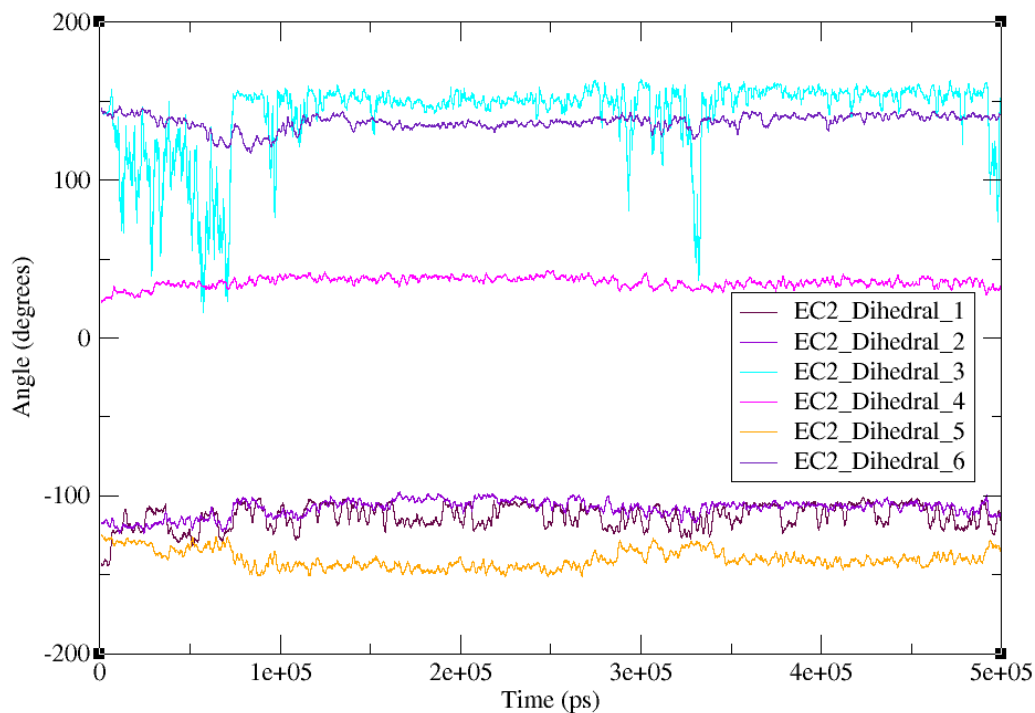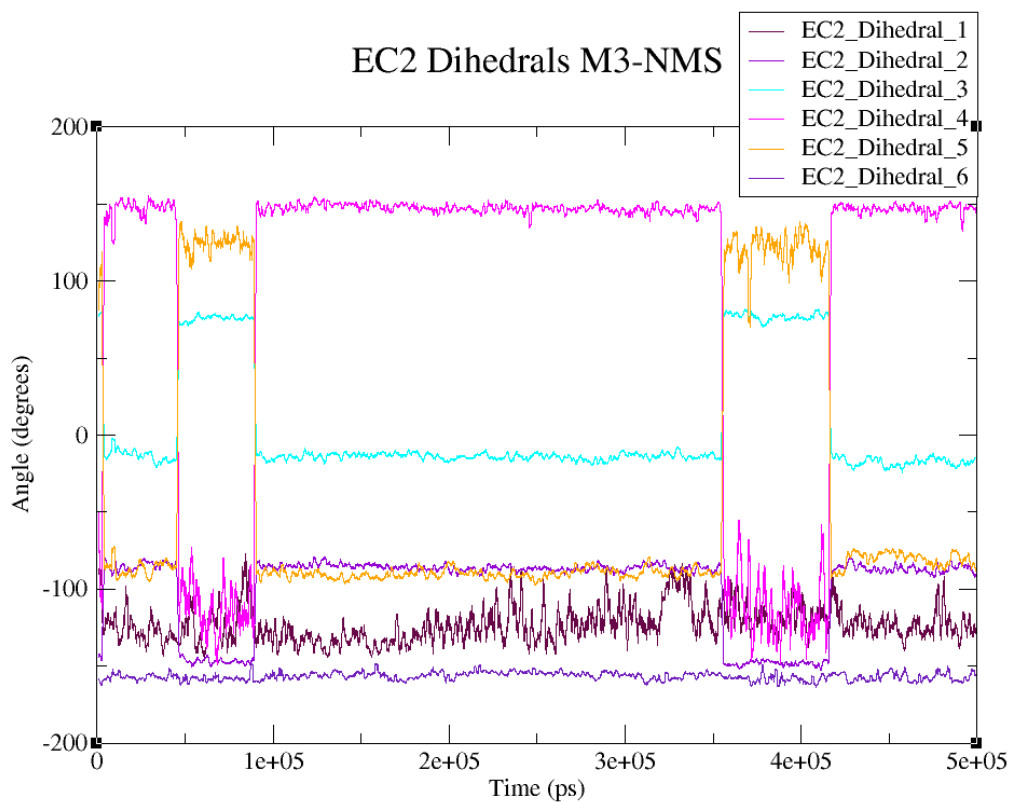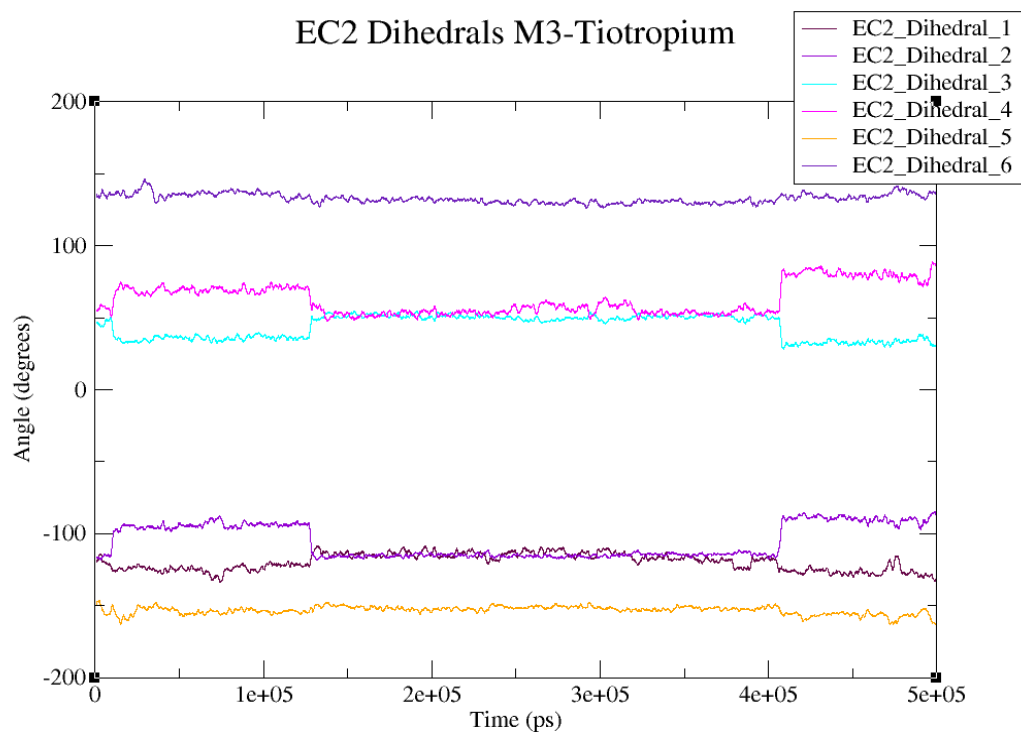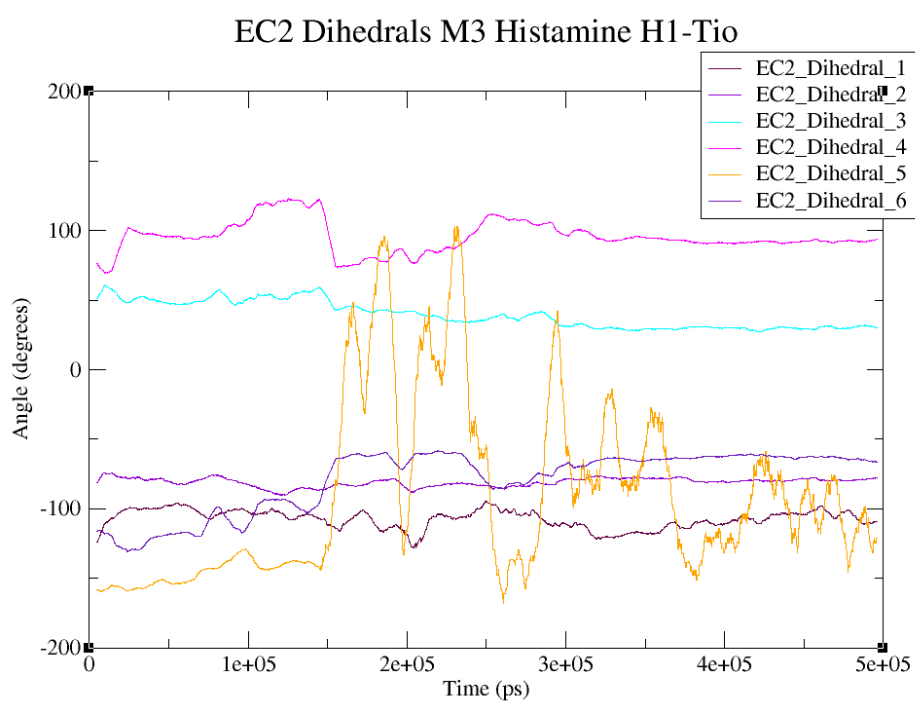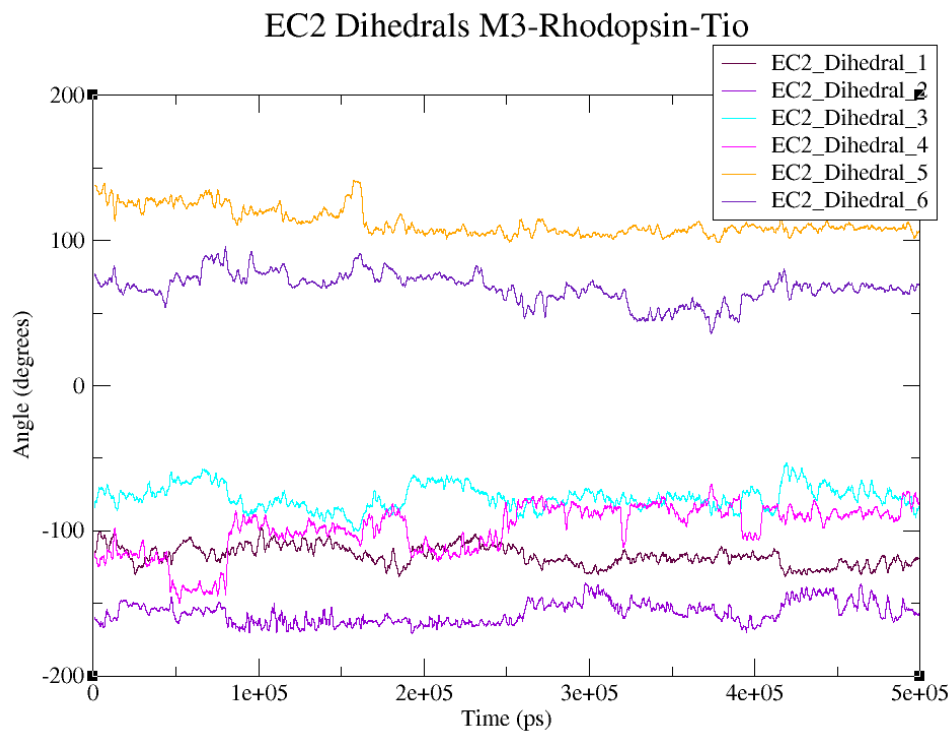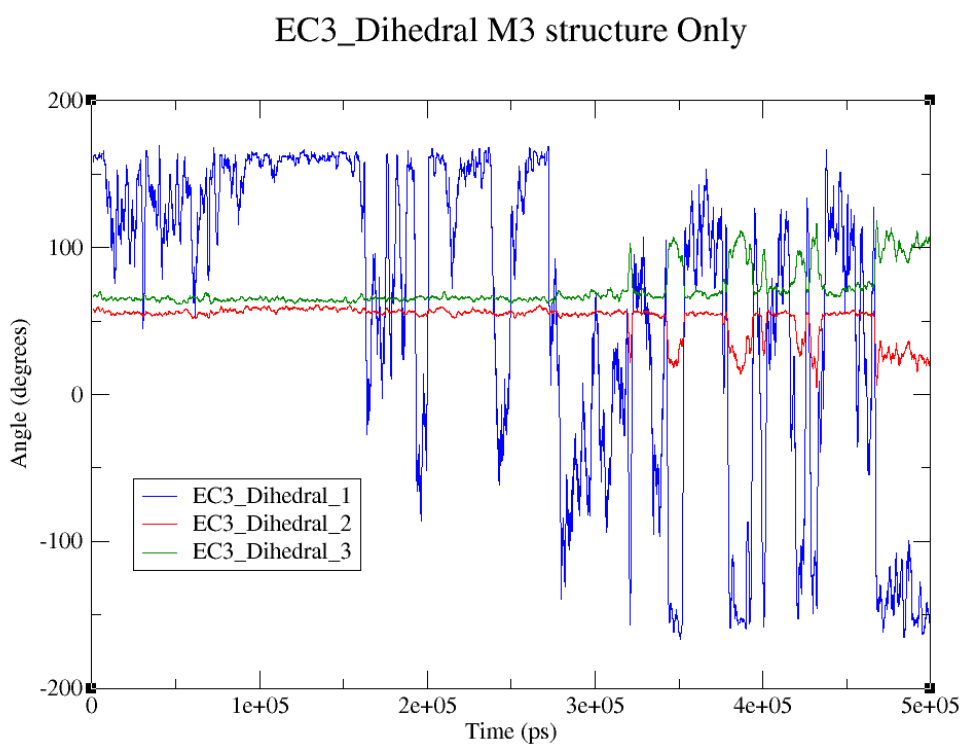
(k)

EC3 Dihedrals M3-NMS
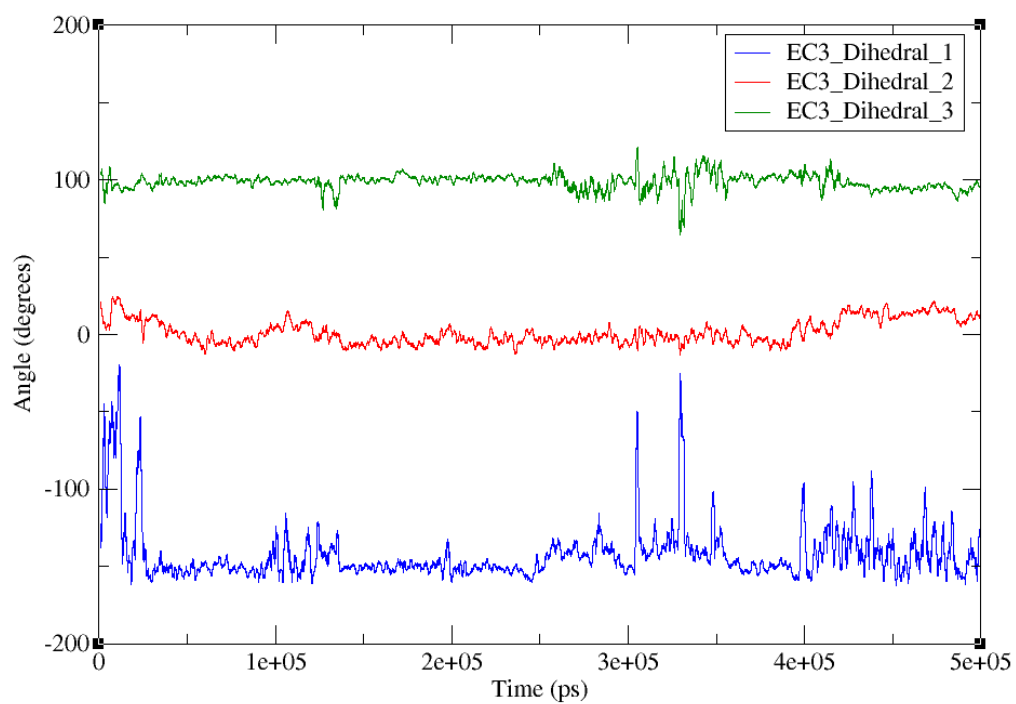
(l)

EC3_Dihedral M3-Tiotropium
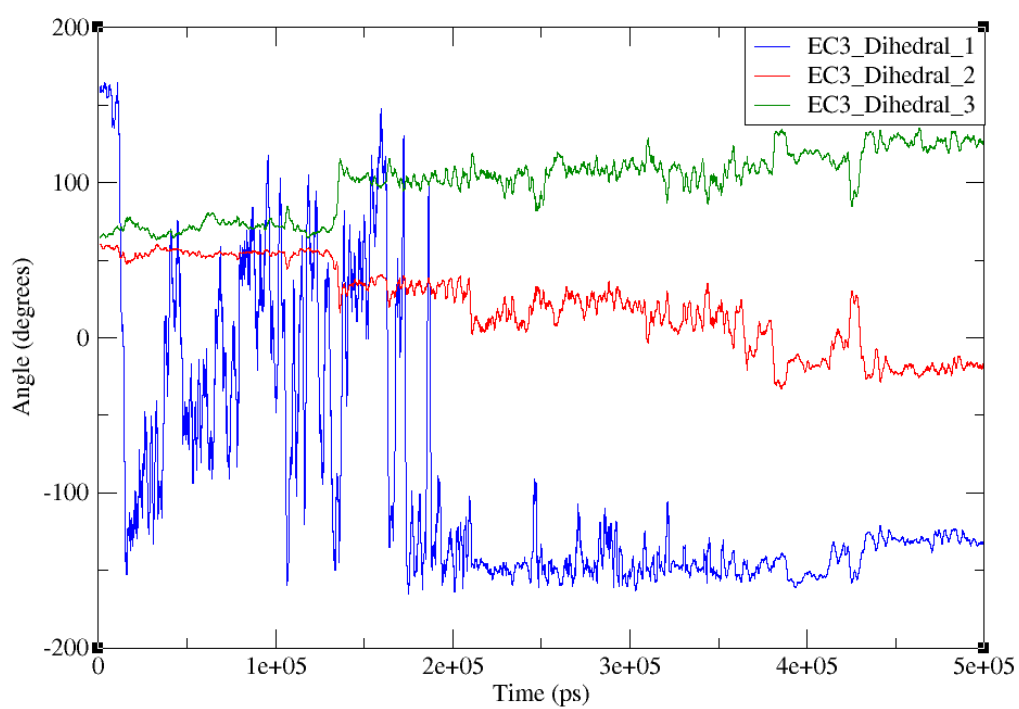
(m)

EC3 Dihedrals M3 Histamine H1-Tio
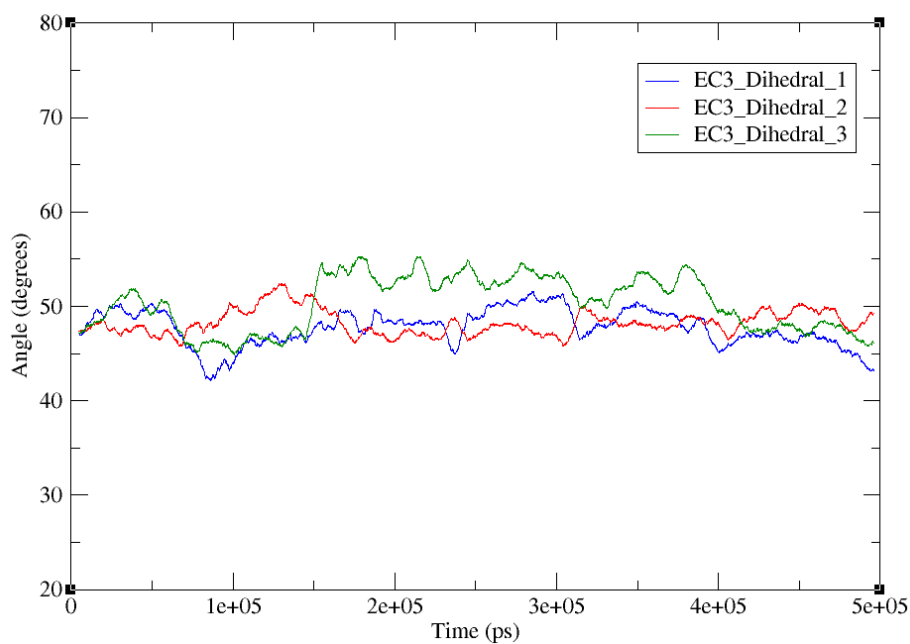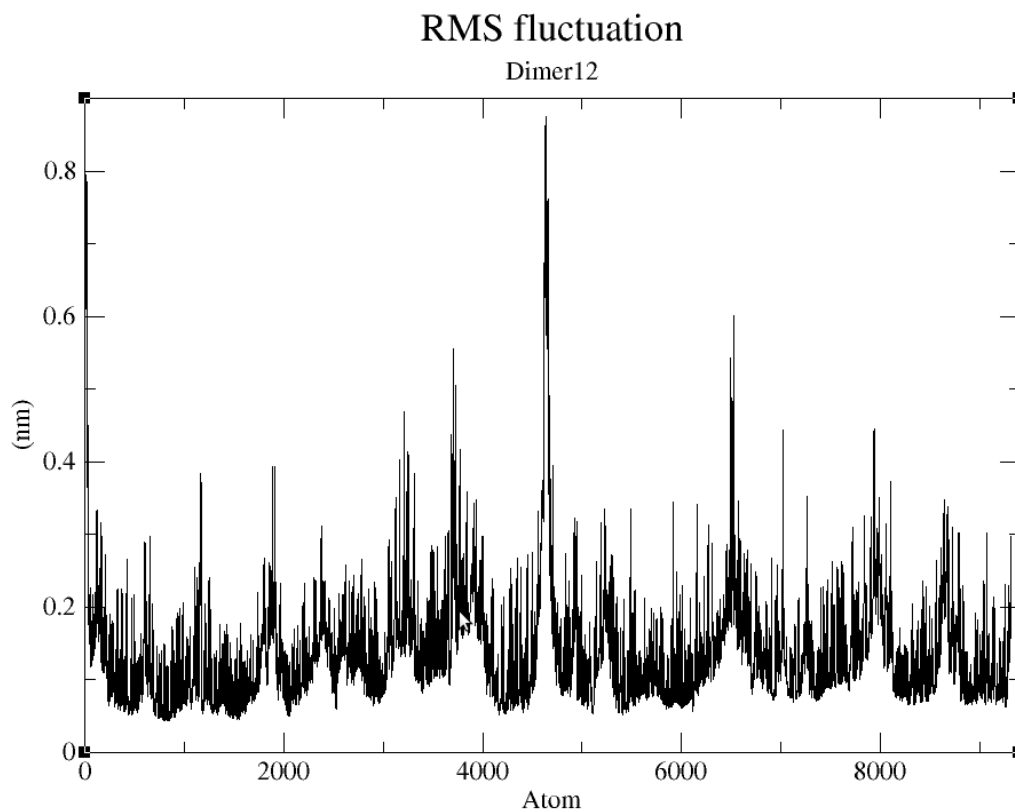

(n)

EC3 Dihedrals M3-Rhodopsin-Tio
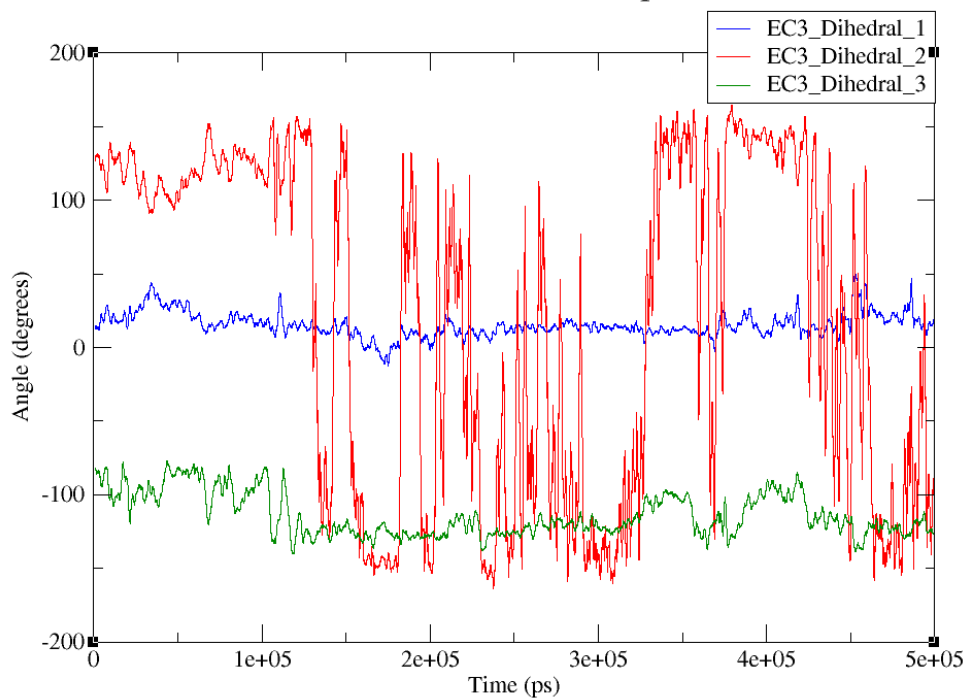
(o)

Figure 8.5: (**a-o**), the evolution of the extracellular dihedral angle for each of the receptor model during the refinement process.
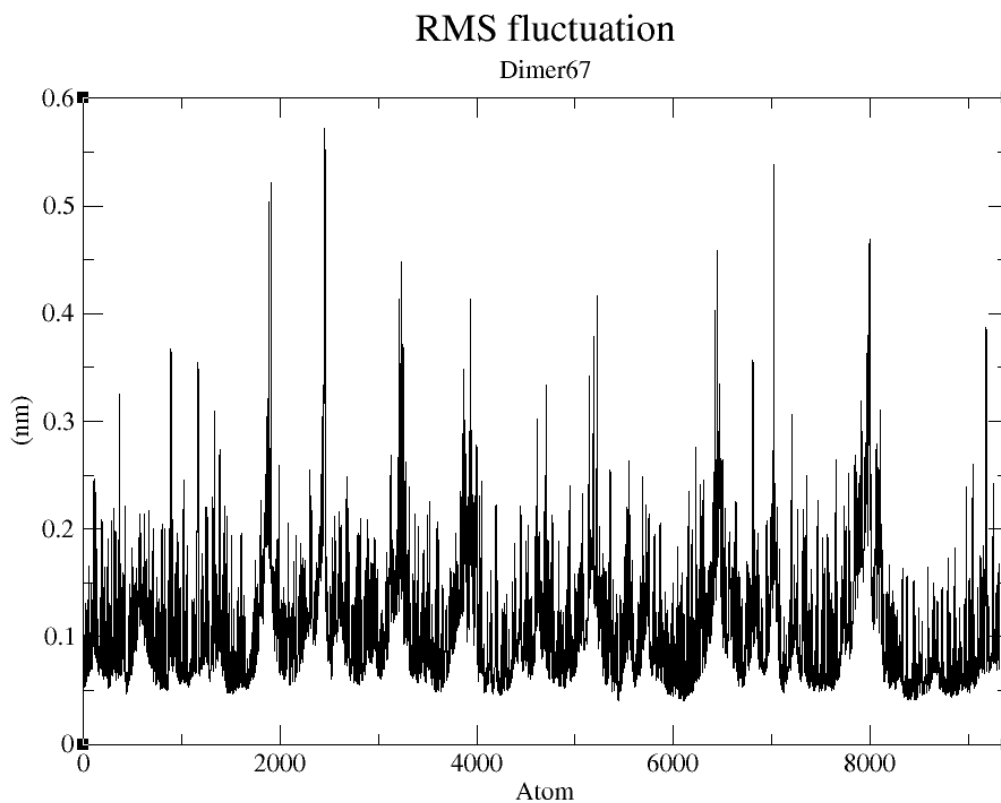
8.3.2 Root mean square fluctuations
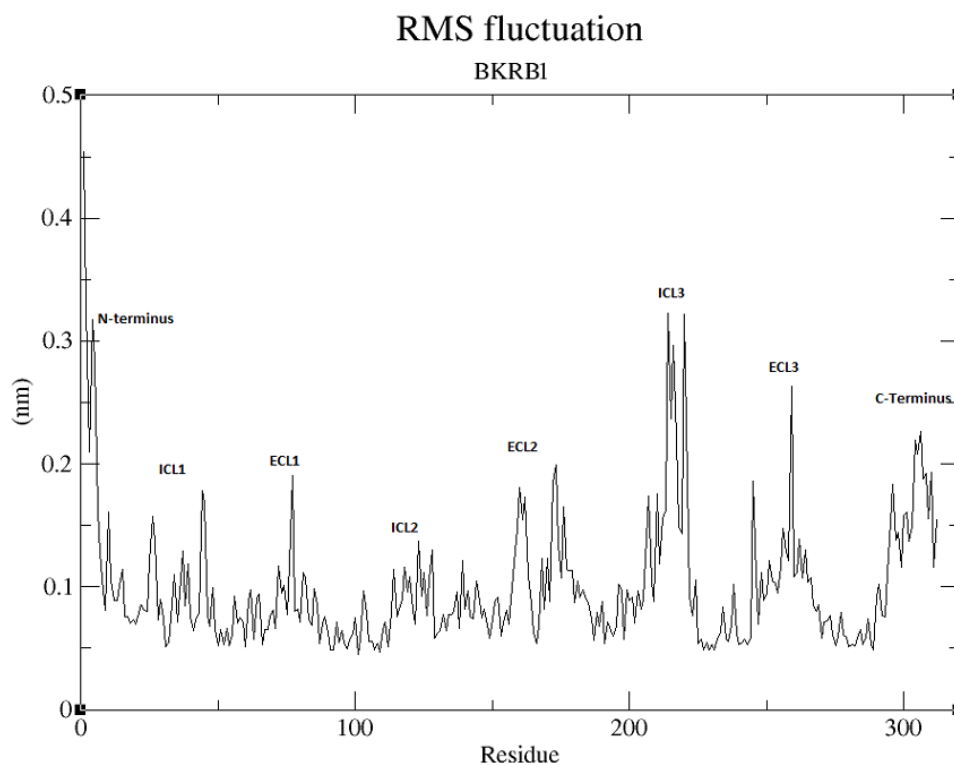
Figure 8.6: The *rmsf* of the B1R average structure (at 450-500ns)
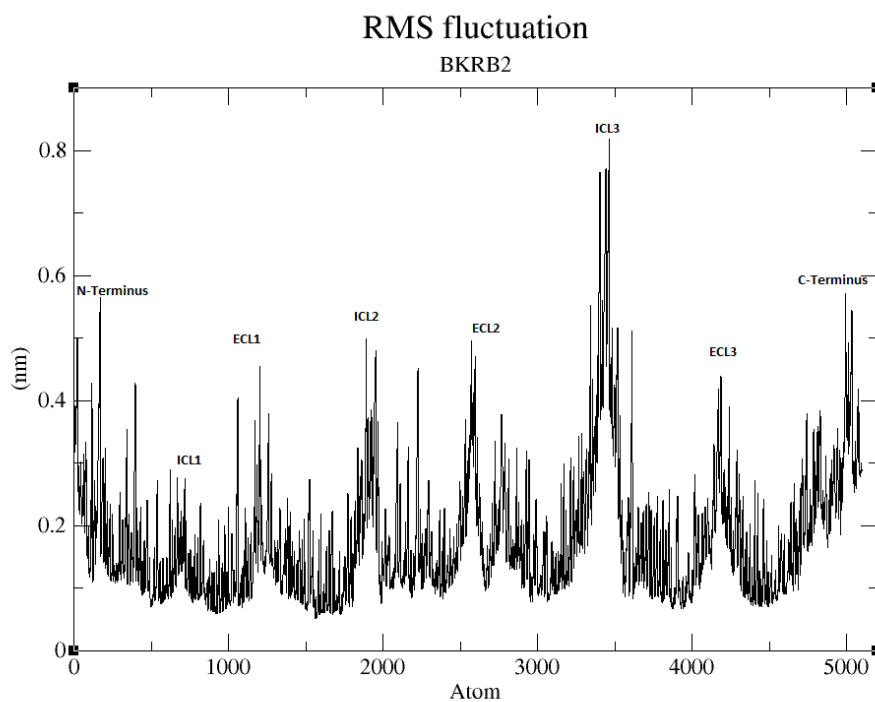
Figure 8.7: The *rmsf* of the B2R average structure (at 450-500ns)

RMS fluctuation
Dimer12

Figure 8.8: The *rmsf* of the Dimer interfacing at TM1-TM2 average structure (at 450-500ns)

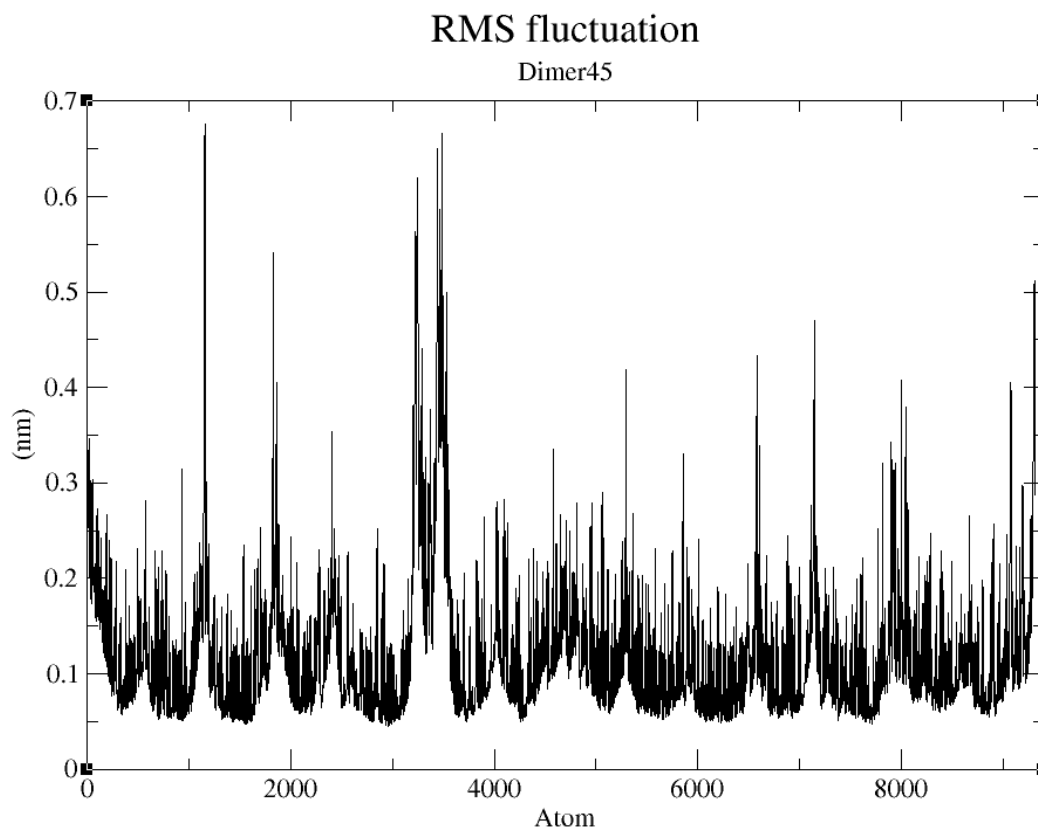## RMS fluctuation
Dimer67

Figure 8.8: The *rmsf* of the Dimer interfacing at TM6-TM7 average structure (at 450-500ns)

Figure 8.9: The *rmsf* of the Dimer interfacing at TM4-TM5 average structure (at 450-500ns)