Title

Exploring and Characterizing the Conformational Space of the A2AAR and the S1PR1 for the Treatment of Immunological Disorders

Permalink

https://escholarship.org/uc/item/0bw8d9k7

Author

Caliman, Alisha Danielle

Publication Date

2017

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Exploring and Characterizing the Conformational Space of the A_{2A}AR and the S1PR$_1$ for the Treatment of Immunological Disorders

A dissertation submitted in partial satisfaction of the requirements for the degree of
Doctor of Philosophy

in

Biomedical Sciences

by

Alisha Danielle Caliman

Committee in Charge:

Professor J. Andrew McCammon, Chair
Professor Rommie E. Amaro
Professor Joan Heller Brown
Professor Michael K. Gilson
Professor JoAnn Trejo

2017

The dissertation of Alisha Danielle Caliman is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2017

DEDICATION

For my mother: This would not have been possible without you. I love you for all the late-night conversations, for all the tears, and for all the laughter.

For my grandmother, granny and gaga: I wish you could have been here to see this. You started the journey with me, and I know you would be proud of me. I miss you all so very much.

Philippians 4:13.

EPIGRAPH

Our deepest fear is not that we are inadequate.
Our deepest fear is that we are powerful beyond measure.
It is our light, not our darkness that most frightens us.
We ask ourselves, who am I to be brilliant, gorgeous, talented, fabulous?
Actually, who are you *not* to be?
You are a child of God.
Your playing small does not serve the world.
There is nothing enlightened about shrinking so that other people won't feel insecure around you.
We are all meant to shine, as children do.
We were born to make manifest the glory of God that is within us.
It's not just in some of us; it's in everyone.
And as we let our own light shine, we unconsciously give other people permission to do the same.
As we are liberated from our own fear, our presence automatically liberates others.

Marianne Williamson

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ρ – Probability Distribution
β_1AR – β_1 Adrenergic Receptor
β_2AR – β_2 Adrenergic Receptor
2-AG - 2-Arachidonoylglycerol
Å – Angstrom
A – Alanine
A_{2A}AR – Adenosine A_{2A} Receptor
aMD – Accelerated Molecular Dynamics
Antag – Antagonist
AP-1 – Activator Protein 1
Apo – Ligand Free
Asp – Aspartic Acid
BBB – Blood Brain Barrier
C – Cysteine
C-Terminus – Carboxyl Terminus
C-Term – Carboxyl Terminus
CADD – Computer-Aided Drug Design
cAMP – Cyclic Adenosine Monophosphate
CB2 – Cannabinoid Receptor Type 2
CD4 – Cluster of Differentiation 4
CD8 – Cluster of Differentiation 8
CHARMM – Chemistry at HARvard Macromolecular Mechanics
CHMBID – Chembridge ID
cMD –Conventional Molecular Dynamics
CNS – Central Nervous System
CS – Crystal Structure
CWxP – Cysteine, Tryptophan, (any amino acid), Proline
D – Aspartic Acid
DRY – Aspartic acid, Arginine, Tyrosine
DS_{avg} – Average Docking Score
DS_{min} – Minimum Docking Score
E – Glutamic Acid
EAE – Experimental Autoimmune Encephalomyelitis
ECL – Extracellular Loop
EF – Enrichment Factor
F – Phenylalanine
fs – Femtosecond
G – Glycine
GABA – Gamma-Aminobutyric Acid
GPCR – G Protein-Coupled Receptor
H – Histidine
His – Histidine
I – Isoleucine
ICL– Intracellular Loop
IFN-γ – Interferon Gamma
IL23 – Interleukin 23

IL6 – Interleukin 6
IL8 – Interleukin 8
IL10 – Interleukin 10
K – Lysine
kcal/mol – Kilocalorie/Mole
L – Leucine
M – Methionine
MD – Molecular Dynamics
MI – Metarhodopsin I
MII – Metarhodopsin II
ms – Millisecond
MS – Multiple Sclerosis
N – Asparagine
N-Terminus – Amine Terminus
N-Term – Amine Terminus
NaCl – Sodium Chloride
NAM – Negative Allosteric Modulators
NAMD – Nanoscale Molecular Dynamics
NCI – National Cancer Institute
NF-κB – Nuclear Factor Kappa-Light-Chain-Enhancer of Activated B Cells
NKT – Natural Killer T cells
NPT – Constant Number, Pressure, Temperature
NPxxY – Asparagine, Proline, (any amino acid), (any amino acid), Tyrosine
ns – Nanosecond
NSC – National Service Center
NVT – Constant Number, Volume, Temperature
P – Proline
PAM – Positive Allosteric Modulators
PC – Principal Component
PCA – Principal Component Analysis
PDB – Protein Data Bank
pH – Potential of Hydrogen
PKA – Protein Kinase A
PMF – Potential of Mean Force
POPC – 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE – 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPG – 1-Palmitoyl-2-oleoyl-sn-glycero-3-phospho-rac-glycerol
ps – Picosecond
Q – Glutamine
R – Arginine
RMSD – Root Mean-Square Deviation
RMSF – Root Mean-Square Fluctuation
Res_{avg} – Average Water Residency Time
S – Serine
S1P – Sphingosine-1-Phosphate
S1PR – Sphingosine-1-Phosphate Receptor
$S1PR_1$ – Sphingosine-1-Phosphate Receptor Subtype 1

S1PR$_4$ – Sphingosine-1-Phosphate Receptor Subtype 4
S1PR$_5$ – Sphingosine-1-Phosphate Receptor Subtype 5
T – Threonine
TM – Transmembrane Helix
TNF-α – Tumor Necrosis Factor Alpha
V – Valine
VMD – Visual Molecular Dynamics
W – Tryptophan
Y – Tyrosine
µs – Microseconds
χ_1 – Chi1 Side Chain Dihedral
χ_2 – Chi2 Side Chain Dihedral

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Prof. J. Andrew McCammon for his continuous support during my PhD matriculation. He has shown me unwavering kindness, patience, and motivation. Throughout my time at UCSD, Andy has made me feel like I have unlimited potential and that all things are possible. I could not have chosen a better mentor.

I would also like to thank the former and present members of the McCammon group, particularly Yinglong Miao, Sarah Swift, and Christopher Lee. I thank Yinglong for our weekly discussions. He always answers any question that I have; in fact, his “very-serious-thinking-face” is burnt into my memory. Yinglong is a brilliant mentor, and has positively changed my research. I thank Sarah for her help transitioning into the field of computational chemistry. She is always willing to meet with me, even after her departure. I thank Christopher for being a great office mate. He is always available for a great discussion on our research or future careers.

The McCammon Lab would not work without Patti Craft. I will sincerely miss Patti, and in so many ways, she has become my friend. Patti is always available for help with lab issues, whether it be reimbursement, funding, losing lab equipment, losing lab keys, or breaking personal heaters. More importantly, she is always available for a chat about the future, or the past. I have heard so many amazing stories from Patti, and often hope that I will have as many adventures as she has. Thank you, Patti.

Finally, I would like to thank my family. I would not be here if my mother hadn’t sacrificed so much. I would also like to thank my dad, my sisters, Rose and Aryn, my cousins, Lisa and April, my aunts, Karen, Meredith, and Jan, my uncle, Neil, and my grandparents, Mildred, Ferris, Marjorie, and Oscar.

Chapter 2 is a minimally modified reprint of the material as it appears in Alisha D. Caliman, Sara E. Swift, Yi Wang, Yinglong Miao, and J. Andrew McCammon, “Investigation of the conformational dynamics of the apo A_{2A} adenosine receptor,” Protein Science, 2015. The dissertation author was the primary investigator and author of this paper.

Chapter 3 is a modified reprint of the material as it will appear in Alisha D. Caliman, Yinglong Miao, and J. Andrew McCammon, “Mapping the Allosteric Sites of the A_{2A} Adenosine Receptor,” Chemical Biology & Drug Design, 2017. The dissertation author was the primary investigator and author of this paper.

Chapter 4 is a minimally modified reprint of the material as it appears in Alisha D. Caliman, Yinglong Miao, and J. Andrew McCammon, “Activation Mechanisms of the First Sphingosine-1-Phosphate Receptor,” Protein Science, 2017. The dissertation author was the primary investigator and author of this paper.

Chapter 5 is a modified reprint of the material that will be submitted in Alisha D. Caliman, Yinglong Miao, and J. Andrew McCammon, “The Identification of Non-Orthosteric Sites and Potential Modulators for the First Sphingosine-1-Phosphate Receptor.” The dissertation author was the primary investigator and author of this paper.

VITA

2011	B.S. in Biology Spelman College, Atlanta, GA
2017	PhD in Biomedical Sciences University of California, San Diego, La Jolla, CA

Publications:

Caliman, A.D., S.E. Swift, Y. Wang, Y. Miao. J.A. McCammon. Investigation of the conformational dynamics of the apo A2A adenosine receptor. Protein Science. 24(6) pp 1004-1012, (2015)

Miao, Y., A.D. Caliman, J.A. McCammon. Allosteric Effects of Sodium Ion Binding on Activation of the M3 Muscarinic G-Protein Coupled Receptor. Biophysical Journal. 8(7) pp 1796-1806, (2015)

Caliman, A.D., Y. Miao, J.A. McCammon. Activation Mechanisms of the First Sphingosine-1-Phosphate Receptor. Protein Science, *in press* (2017)

Caliman, A.D., Y. Miao, J.A. McCammon. Mapping the Allosteric Sites of the A_{2A} Adenosine Receptor. Chemical Biology & Drug Design (pending minor revisions)

ABSTRACT OF THE DISSERTATION

Exploring and Characterizing the Conformational Spaces of the A_{2A}AR and the S1PR$_1$ for the Treatment of Immunological Diseases

by

Alisha Danielle Caliman

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2017

Professor J. Andrew McCammon

Sepsis and multiple sclerosis are immunological disorders defined by an over-response of the immune system to the body’s own tissues. Several G protein-coupled receptors (GPCRs), including the adenosine A_{2A} receptor (A_{2A}AR) and the first sphingosine-1 phosphate receptor (S1PR$_1$), play a role in the immune system. Activation of the A_{2A}AR on lymphocytes is anti-inflammatory, because the activation of the receptor inhibits pro-inflammatory cytokines, such as TNF-α, while the S1PR$_1$ is pro-inflammatory by promoting lymphocyte circulation. In this dissertation, I applied a computational approach to explore the activation pathways of the A_{2A}AR and the

$S1PR_1$, and a structural-based drug design approach to identify potential modulators for the treatment of immunological disorders. The deactivation process of the $A_{2A}AR$ was characterized from long time-scale molecular dynamics simulations. Four conformers of the $A_{2A}AR$ during deactivation were identified, including the active, intermediate 1, intermediate 2, and inactive. The non-orthosteric binding sites in each conformer of the receptor were identified. Activation of the $S1PR_1$, initiated by a rotameric switch of residue $W268^{6.48}$, was captured with accelerated molecular dynamics simulations. Eight non-orthosteric binding sites were identified on the $S1PR_1$, and 39 potential orthosteric and allosteric modulators were discovered. These modulators are being tested for activity on the $S1PR_1$.

CHAPTER 1

Exploring and Characterizing the Conformational Space of the $A_{2A}AR$ and the $S1PR_1$ for the Treatment of Immunological Disorders

Sepsis is one of the leading causes of death around the world [1-3]. It is caused by an infection, either viral or bacterial, that triggers the immune response [2, 3]. The initial infection causes a signaling cascade and leads to the activation of pro-inflammatory transcription factors such as NF-κB, AP-1, c-Fos, and c-Jun [1, 3]. These transcription factors activate pro-inflammatory cytokines, such as TNF-α and IL6 [3]. The immune response is then heightened by the recruitment of CD4+ T cells and the subsequent activation of macrophages. In a normal case, this immune response would fight the initial infection and cease. In sepsis, this response becomes generalized to normal tissues and leads to a period of cytokine-mediated hyper-inflammation followed by a period of immunodeficiency [2-4]. Hyper-inflammation causes the activation of the coagulation system and an increase in capillary permeability [1-3]. Despite the role of cytokines in sepsis, compounds that directly inhibit pro-inflammatory cytokines, like TNF-α, are not effective treatment options [3, 4].

Multiple sclerosis (MS) affects approximately 2.5 million individuals worldwide [5]. It is characterized by lesions of demyelinated neurons, or neurons with destroyed myelin fibers [6]. The initial causes of MS are not known [7], but the periods of neuronal inflammation are thought to be triggered by activated CD4+ and CD8+ T cells that enter the central nervous system (CNS) after being released from secondary lymphoid organs [7, 8]. These cells cross the blood brain barrier (BBB) and secrete pro-inflammatory cytokines and chemokines including, IL8, that activate microglia and astrocytes and recruit lymphocytes [7-9]. The complete immune response against neuronal cells ultimately leads to apoptosis of oligodendrocytes, demyelination, and axonal damage [7, 8, 10].

The role of the adenosine receptors in the immune system has been studied extensively. The adenosine receptors, subtypes A_1, A_{2A}, A_{2B} and A_3, are a class of G protein-coupled receptor (GPCR) that bind to the purine, adenosine [11]. Upon activation, the A_{2A}AR couples to the G_s protein in the peripheral tissues and the G_{olf} protein in the brain to stimulate the cAMP-PKA pathway [3, 12, 13]. In T cells, this activation is anti-inflammatory and inhibits NF-κB and the pro-inflammation cytokines, TNF-α and IFN-γ [14-16]. Activation of the A_{2A}AR leads to the production of anti-inflammatory cytokine, IL-10, in monocytes, and inhibition of TNF-α in neutrophils [17]. Because of the anti-inflammatory activities of the A_{2A}AR, modulators of this receptor are being studied for treatments of sepsis [3].

The anti-inflammatory activities of the A_{2A}AR also play a role in MS. Mice with a global knockout of the A_{2A}AR have a more severe type of experimental autoimmune encephalomyelitis (EAE), the mouse model of MS, due to an increase in IFN-γ production and an increase in lymphocyte proliferation [18]. Additionally, treatment with an A_{2A}AR agonist prevents EAE in mice [19].

Sphingosine-1-phosphate's (S1P) role in the immune response has also been of keen interest over the past decade. This zwitterionic agonist with a polar head group and nonpolar tail binds to the five subtypes of the sphingosine-1-phosphate receptors ($S1PR_{1-5}$) with high affinity. The $S1PR_1$, specifically, is known to be involved in leukocyte egress from secondary lymphoid tissues by acting as a chemoattractant receptor. In secondary lymphoid tissues, there is an S1P gradient. S1P levels are low where T cells enter the lymphoid tissue and high in the exit region, or medullary sinus, and in the blood [20]. Newly generated T cells express the $S1PR_1$ and follow the gradient leading to the rapid release of T cells to the blood stream and the internalization of the receptor [20]. A conditional T cell knockouts of the $S1PR_1$ in mice prevent T cell traffic from lymphoid organs [21], and internalization of the $S1PR_1$ with an agonist in EAE induces disease remission because of T cell sequestration [22]. Due to its importance in lymphocyte traffic, $S1PR_1$ is a target for MS treatment and the treatment of other autoimmune diseases including lupus, sepsis, and psoriasis [23].

Like all class A GPCRs, the $A_{2A}AR$ and $S1PR_1$ structures are composed of seven transmembrane helices (TM), three extracellular loops (ECL1-3), three intracellular loops (ICL1-3), an extracellular N-terminus, and an intracellular C-terminus [24-34]. GPCRs are inherently dynamic, and often transition between the active and inactive states with or without a ligand. X-ray structures often only represent one low-energy state of the receptor. The $A_{2A}AR$ has been co-crystallized with agonists [24-27], antagonists [28-33], coupled to a G protein [27].The $S1PR_1$ has only been co-crystallized with an antagonist [34].

In this study, I apply molecular dynamics, fragment mapping and a computer aided drug design (CADD) techniques to the $A_{2A}AR$ and the $S1PR_1$. I investigate the flexibility and dynamics of the $A_{2A}AR$ and the $S1PR_1$, identified non-orthosteric (allosteric) binding sites on both the A_{2A}AR and the S1PR$_1$, and discovered potential orthosteric and allosteric modulators for the S1PR$_1$.

Dynamic Properties of the A_{2A}AR and the S1PR$_1$

In chapter 2, I describe the transition states between the active and inactive conformers of the A_{2A}AR. I performed conventional molecular dynamics (cMD) simulations on the ligand-free form of the A_{2A}AR after removing the agonist/antagonist. In molecular dynamics (MD), Newton's equation of motion is solved at each time step based on the classical force field, which considers the bond, angle, dihedral, van der Waals, and electrostatic forces acting on each atom [35]. The enables the physical movement of a protein to be studied with atomistic details. To our knowledge, this is the first study to identify four key transition steps in the deactivation process of the A_{2A}AR.

In chapter 4, I describe the activation pathway of the ligand-free S1PR$_1$, by applying accelerated molecular dynamics (aMD) to the system. aMD enhances the sampling space of biomolecules (e.g. the S1PR$_1$) by applying a boost potential to the system when the potential energy falls below a threshold. This prevents the receptor from remaining in one low-energy state [36, 37]. To our knowledge, this is the first study to capture the basal activation process of the S1PR$_1$.

Allosteric Sites

In chapter 3 and chapter 5, I identify allosteric binding pockets on the A_{2A}AR and the S1PR$_1$. Representative conformers of the A_{2A}AR and S1PR$_1$ from the previously described dynamics studies were screened using the fragment-mapping software, FTMap [38]. FTMap can reveal the druggable orthosteric and allosteric sites on the receptor. Allosteric modulators either enhance or inhibit the activation of the receptor in the presence of its endogenous ligand; so, without the ligand, most modulators have no activity [39]. Additionally, targeting pockets of the receptor outside of the conserved orthosteric site can provide a mechanism to achieve subtype specificity. Five novel allosteric modulators are classified in the A_{2A}AR and 8 novel allosteric sites in the S1PR$_1$ are evaluated.

Modulators for the Treatment of Immunological Disorders

In chapter 5, I applied a CADD approach on the S1PR$_1$, screening a total of 13,000 compounds on the orthosteric site and four non-orthosteric sites on the S1PR$_1$. In CADD, multiple conformers of biomolecules (e.g. S1PR$_1$), are screened with various compound libraries. I select 39 compounds as potential modulators of the S1PR$_1$.

The results of my studies will advance the knowledge of A_{2A}AR and S1PR$_1$ activation, and facilitate the discovery of novel A_{2A}AR and S1PR$_1$ modulators for the treatment of immunological disorders.

CHAPTER 2

Investigation of the Conformational Dynamics of the Apo A_{2A} Adenosine Receptor

Abstract

The activation/deactivation processes of G protein-coupled receptors (GPCRs) have been computationally studied for several different classes, including rhodopsin, the β_2 adrenergic receptor, and the M_2 muscarinic receptor. Despite determined co-crystal structures of the adenosine A_{2A} receptor (A_{2A}AR) in complex with antagonists, agonists and an antibody, the deactivation process of this GPCR is not completely understood. In this study, we investigate the convergence of two apo simulations, one starting with an agonist-bound conformation (PDB: 3QAK) and the other starting with an antagonist-bound conformation (PDB: 3EML). Despite the two simulations not completely converging, we identified distinct intermediate steps of the deactivation process characterized by the movement of Y288$^{7.53}$ in the NPxxY motif. We find that Y288$^{7.53}$ contributes to the process by forming hydrogen bonds to residues in transmembrane helices 2 and 7 and breaking these interactions upon full deactivation. Y197$^{5.58}$ also plays a role in the process by forming a hydrogen bond once the residue side chain moves from the lipid interface to the helical bundle.

Introduction

Adenosine receptors are a class of G protein-coupled receptors (GPCRs) that mediate several cellular processes by binding endogenously to adenosine [12]. Upon binding of adenosine or other agonists, the A_{2A}AR subtype of the adenosine receptor undergoes a conformational change and couples to G_s in the peripheral tissues and G_{olf} in the brain to stimulate the cAMP-PKA pathway. In the immune system, activation leads to immunosuppression by inhibiting pro-inflammation cytokines, including TNF-α and IFN-γ [13-16]. In the brain, the A_{2A}AR is present in the dopamine-rich areas, such as the globus pallidus, and works by inhibiting dopaminergic activity by increasing gamma-aminobutyric acid (GABA) or directly interacting with the D_2 dopamine receptor [13, 40]. Thus, antagonists of A_{2A}AR, like caffeine, increase dopaminergic activity, making such inhibitors possible therapeutics for Parkinson's disease and Alzheimer's disease [13, 40, 41]. Additionally, antagonism of A_{2A}AR is thought to decrease lymphocyte migration across the blood brain barrier (BBB) to the central nervous system, making such antagonists potential candidates for treating BBB diseases such as multiple sclerosis [18].

The X-ray structure of A_{2A}AR has been solved in complex with several different ligands including antagonists [29, 30], an inverse agonist ZM241385 [28], an antibody [31], and agonists [24, 25] including UK-432097 [25]. Each of these structures reveals an extracellular N-terminus, seven transmembrane (TM1-7) alpha helices, three extracellular loops (ECL1-3), three intracellular loops (ICL1-3) and an intracellular C-terminus. Compared with the inverse-agonist bound structure [28], the UK-432097 agonist bound structure [25] shows the side-chain dihedral switch of residues F201$^{5.62}$ and Y197$^{5.58}$ from inside the helical bundle to outside the helical bundle. Furthermore, there is a χ_1 side-chain dihedral switch of residue Y288$^{7.53}$ from trans in the inverse-agonist bound structure to gauche in the antagonist bound structure. These structural features cause the distance between the hydroxyl oxygen of Y197$^{5.58}$ and Y288$^{7.53}$ to be 15.7 Å in the agonist-bound structure and 9.58 Å in the inverse-agonist bound structure (Figure 2.1). In the active structure, there is an outward tilt of residue W246$^{6.48}$ causing the displacement from the intracellular end to be ~3 Å. As described below, these differences vary from other active GPCRs.

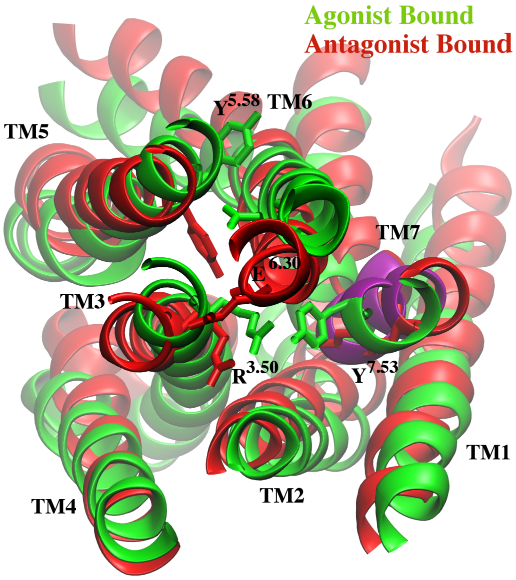

Figure 2.1: A cartoon representation of the starting antagonist-bound apo (red) and the agonist-bound apo (green) structures of the A$_{2A}$AR. Key residues Y197$^{5.58}$, Y228$^{7.53}$, E228$^{6.30}$, and R102$^{3.50}$ are highlighted as sticks. The NPxxY motif is shown in purple.

Activation of several GPCRs is characterized by a structural rearrangement of TM5, TM6, and TM7 to accommodate coupling of the G protein. Upon activation, a salt bridge between residues R$^{3.50}$ and E$^{6.30}$, often called the “ionic lock”, is disrupted and the χ_1 dihedral angle of W$^{6.48}$ switches from the gauche to trans conformation, which breaks interactions between residues D$^{2.50}$, N$^{7.49}$, and S$^{7.45}$. This rearrangement facilitates the side-chain relocation of Y$^{5.58}$ and Y$^{7.53}$, allowing the two tyrosines to interact through either a direct hydrogen bond or through water molecules. In the currently available active structures, the distance between the hydroxyl oxygen of Y$^{5.58}$ and Y$^{7.53}$ is 4.28 Å in β_2 adrenergic receptor (β_2AR) [42], 5.43 Å in rhodopsin [43], and

4.21 Å in the M_2 muscarinic receptor [44]. This is coupled with the displacement of the intracellular end of TM6, which can vary from 7 Å in rhodopsin [43, 45], to 10.4 Å in the M_2 receptor [44, 46], and 14 Å in the β_2AR [42, 47].

Previous computational studies have investigated the activation/deactivation process of GPCRs, and the role of side-chain dihedral switches in it [48-51]. In a study by Dror et al, the deactivation process of β_2AR was characterized from several µs of conventional molecular dynamics (cMD) simulations on the Anton supercomputer. The group started with an agonist-bound structure and was able to identify distinct conformations during the course of deactivation [48]. During the first step of deactivation, TM7 adopts the inactive conformation followed by the movement of both TM5 and TM6. In another study by Miao et al, accelerated molecular dynamics (aMD) was used to identify the pathway from the inactive M_2 receptor to the active receptor. Using the enhanced sampling, the group was able to identify two intermediate conformations on the pathway and observe the direct interaction between $Y^{5.58}$ and $Y^{7.53}$ in the active structures [51]. In a third study by Li et al, the deactivation process of the A_{2A}AR was found to involve the separation of TM4, TM5 and TM3 and a rearrangement of TM6, and deactivation caused the helices to bundle together. Additionally, three separate conformations for the ‘toggle switch,’ $W246^{6.48}$, were identified for apo, active, and inactive structures [49].

One consistent feature of the active structures of the β_2AR, the M_2 receptor and rhodopsin is the interaction between $Y^{5.58}$ and $Y^{7.53}$. This interaction is absent in the A_{2A}AR agonist-bound structure however. Moreover, $Y197^{5.58}$ is facing the lipid interface, causing the distance between the two residues to be greater than 15 Å. For the A_{2A}AR receptor, open questions include what is the role of these groups in the activation/deactivation process. In this study, we performed MD simulations on the apo form of two starting structures of the $A_{2A}AR$, the agonist-bound (PDB: 3QAK) and the antagonist-bound (PDB: 3EML), and examined the convergence of these two simulations over ~1.6 µs. We found the two sampled ensembles of each of these two structures do not converge, although the agonist-bound apo structure is more dynamic and approaches inactivation more significantly. By studying these simulations, we were able to identify two intermediate steps between active and inactive structures based on the dynamics of the NPxxY motif. Interestingly, the antagonist-bound apo structure remains in the inactive conformation, and eventually forms the "ionic lock," which was observed to be broken in the inactive crystal structures [28-31].

Results

Simulation Convergence

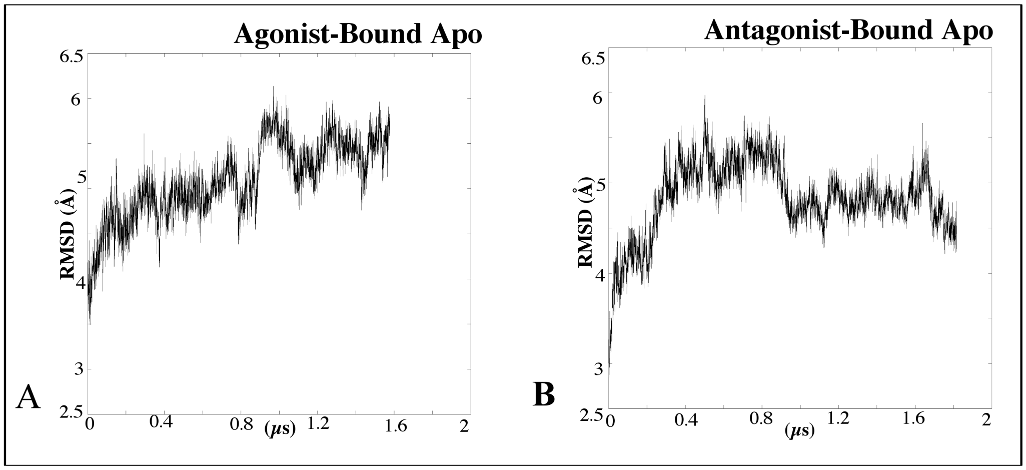

Figure 2.2: The RMSD of Cα atoms to the respective starting structures of the $A_{2A}AR$. (A) shows the agonist-bound apo simulation, while (B) shows the initial antagonist-bound apo simulation.

Starting with the agonist-bound (PDB: 3QAK) and the antagonist-bound (PDB: 3EML) structures of $A_{2A}AR$ with both ligands removed, sufficiently long simulations should generate converged ensembles, given that both structures have the same sequence and are modeled in the same environment. With the final length of the Anton simulations, ~1.6 µs starting from the agonist-bound conformation and ~1.8 µs starting from the antagonist-bound conformation, convergence did not occur. Both simulations deviate from their starting structures, with the initial antagonist-bound apo simulation stabilizing at ~1 µs and the initial agonist-bound apo simulation continuously sampling different phase space (Figure 2.2A and Figure 2.2B).

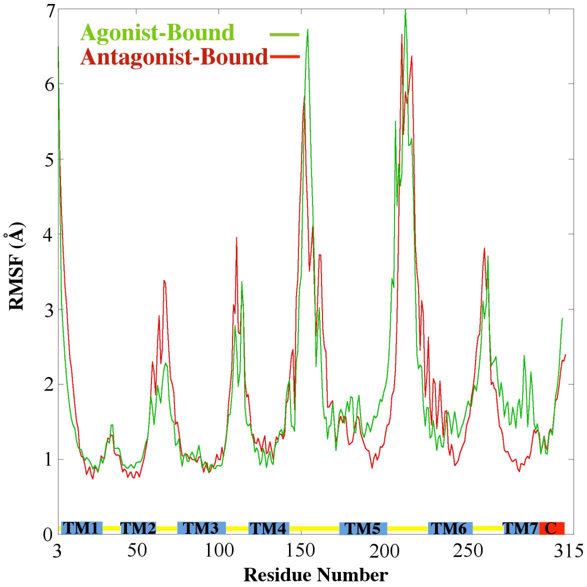

Figure 2.3: The RMSF of the Cα atoms. The initial antagonist-bound apo simulation is shown in red and the initial agonist-bound apo simulation in green. The blue bars represent the location of the TM regions, while the yellow bars represent the location of the loops. The C-Terminus is shown as a red bar.

The most flexible regions for both simulations are ECL2 and ICL3, which is in agreement with several other GPCR computational studies [50, 52]. The intracellular end of TM5 is slightly more flexible in the initial agonist-bound apo simulation compared to the initial antagonist-bound apo simulation with the root mean squared deviation (RMSF) being 1.5 Å and 1 Å, respectively (Figure 2.3). TM7 is also more flexible in the initial agonist-bound apo simulation, with the RMSF being ~2 Å and ~1 Å, respectively.

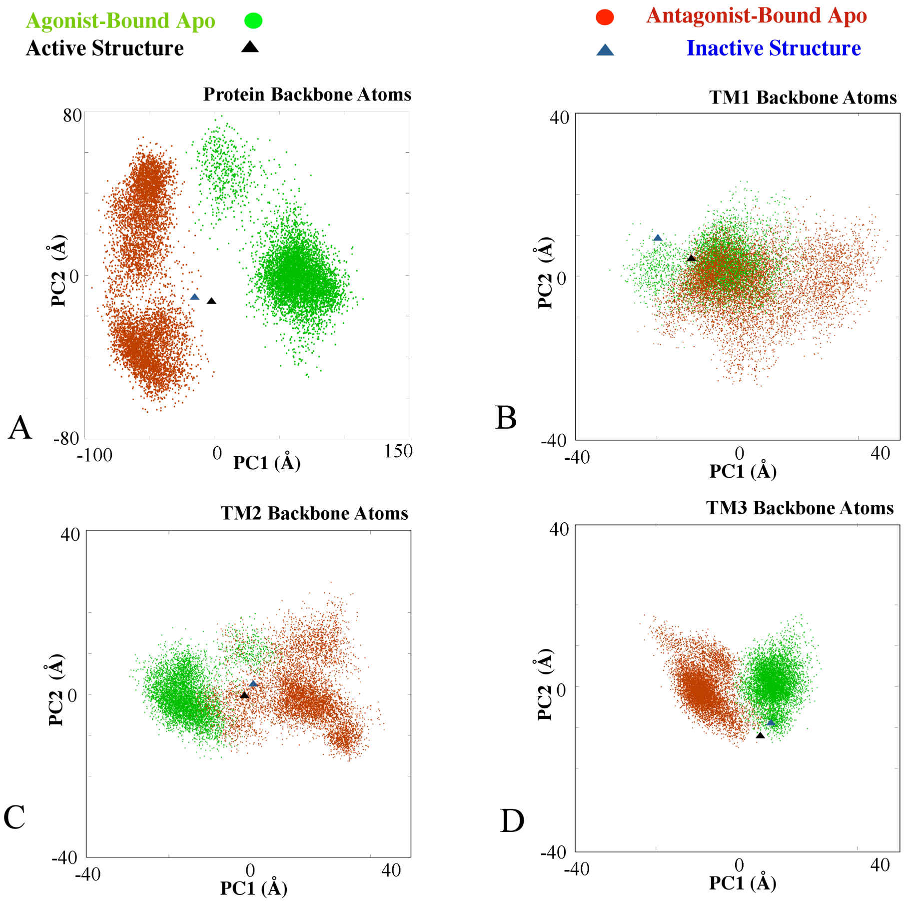

Figure 2.4: The principal component analysis of all the backbone atoms (A), the backbone atoms of TM1 (B), the backbone atoms of TM2 (C), and the backbone atoms of TM3 (D), for both simulations is plotted. The starting active structure is shown as a black triangle, and the starting inactive structure as a blue triangle. The initial antagonist-bound apo simulation is shown as red dots, and the initial agonist-bound apo simulation is shown as green dots.

To further characterize the convergence of the ensembles, we used principal component analysis (PCA) to reduce the dimensionality of the atomic motion and extrapolate correlations from the simulations. We calculated the first two components, PC1 and PC2, from all of the backbone atoms and observed that these two components account for only ~40% of the variance of the simulations. These projections reveal some overlap between the antagonist-bound apo starting structure and the subsequent simulation, and no overlap between the agonist-bound apo starting structure and that subsequent simulation. Additionally, there is no overlap between the two simulations (Figure 2.4A). The first two principal components of all of the backbone atoms contain information about motions of the extracellular and intracellular loops.

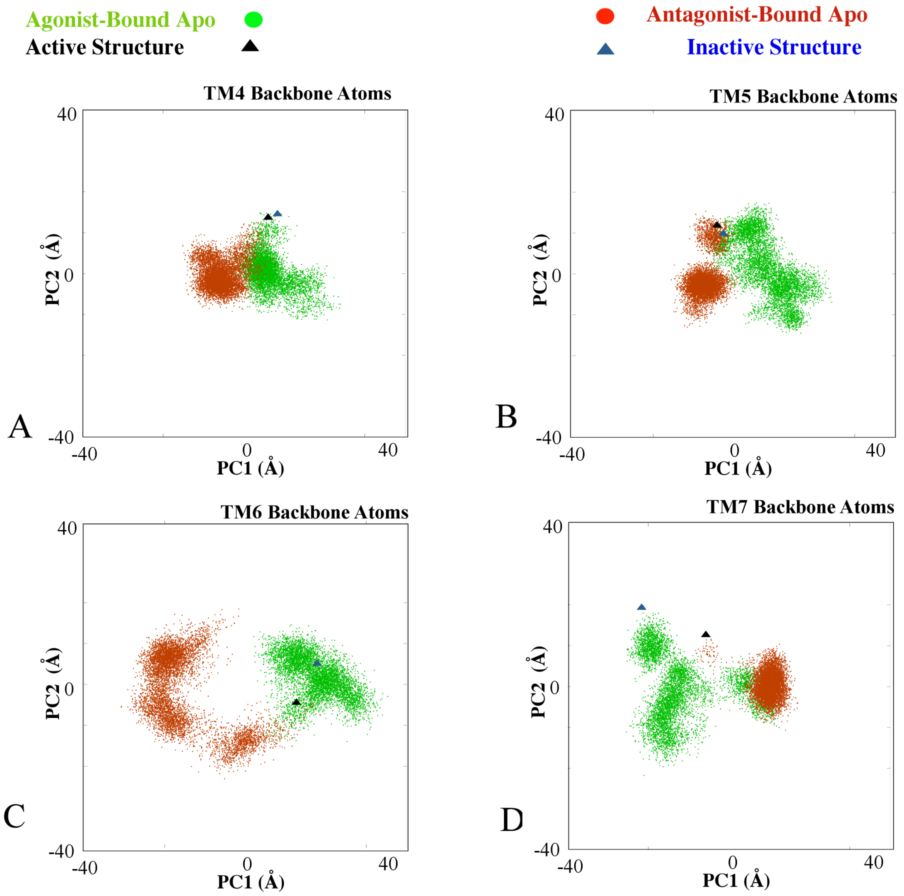

Figure 2.5: The principal component analysis of the backbone atoms of TM4 (A), the backbone atoms of TM5 (B), the backbone atoms of TM6 (C) and the backbone atoms of TM6 (D) for both simulations is plotted. The starting active structure is shown as a black triangle, and the starting inactive structure as a blue triangle. The initial antagonist-bound apo simulation is shown as red dots, and the initial agonist-bound apo simulation is shown as green dots.

To examine convergence of the TM helices, we performed additional PC analysis from the backbone atoms of specific helices. Both TM1 and TM2 show large sampling areas, and overlap between both simulations (Figure 2.4B and Figure 2.4C). TM3 samples less conformational space than TM1 and TM2 and there is no overlap between both simulations, suggesting the movements from this helix do not converge (Figure 2.4D), and the movement of TM4 is also stable with the two simulations slightly converged (Figure 2.5A). The largest differences are between the two simulations are between TM5, TM6, and TM7 (Figure 2.5B, Figure 2.5C, and Figure 2.5D). The antagonist-bound apo simulation shows two distinct populations in TM5, while the agonist-bound apo simulation shows large motions, which were also observed from characterizing the RMSF of this helix (Figure 2.5B). In TM6, there is no overlap between the antagonist-bound apo starting structure and the subsequent simulation, and there is a large motion for the antagonist-bound simulation, while the TM6 in agonist-bound simulation is more stable (Figure 2.5C). In TM7, there is no overlap between the agonist-bound apo starting structure and the subsequent simulation. For the agonist-bound apo simulation, there are three distinct conformations sampled, which drift toward those of the antagonist-bound apo simulation. The antagonist-bound apo simulation is very stable, sampling only one conformation, consistent with the RMSF results (Figure 2.5D). These data suggest that there are large differences between TM5, TM6 and TM7 in the near agonist-bound ensemble compared to the near antagonist-bound ensemble.

Reaction Coordinates

Although the ensembles of the two simulations do not converge completely, key steps can be identified from the evolution of the initial agonist-bound apo simulation towards the inactive structure. A representative coordinate of the inactive structure is the distance between R102$^{3.50}$ and E228$^{6.30}$, the “ionic lock” [51, 52]. The root mean squared deviation (RMSD) of Y288$^{7.53}$ in the NPxxY motif from the starting structure and the distance between the ionic lock side-chains identify three different conformations evolving from the initial agonist-bound apo simulation. A final conformation can be identified from the initial antagonist-bound apo simulation. A potential of mean force (PMF) profile was calculated with these reaction coordinates to isolate all four conformations from the two simulations (Figure 2.6A and Figure 2.6B). From this profile, we discern four main states, and furthermore suggest a hypothesis for deactivation. Because these simulations are not fully converged, the PMF is only qualitatively meaningful, and is used here to help locate the active, intermediate, and inactive conformations.

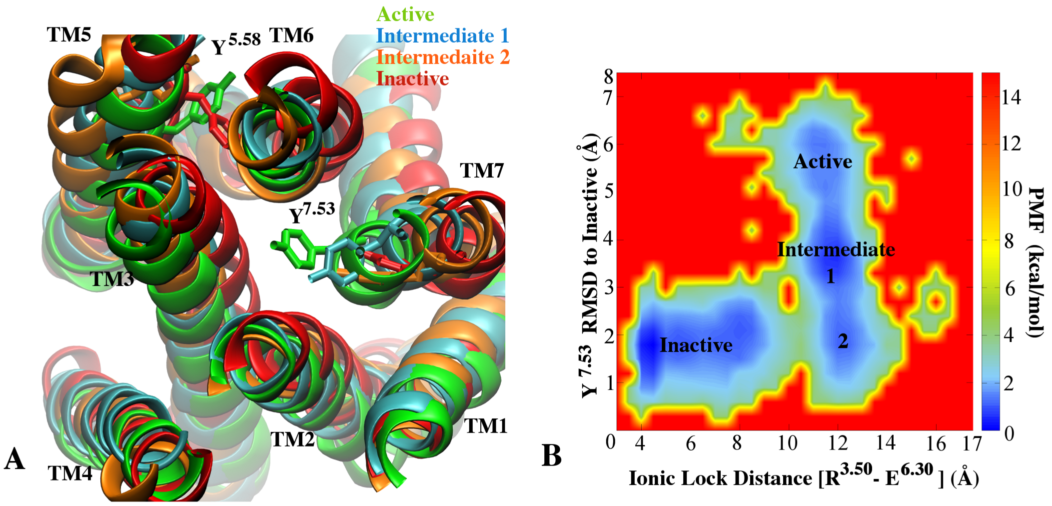

Figure 2.6: Deactivation of the A_{2A}AR from the initial agonist-bound apo conformation to the initial antagonist-bound apo conformation. (A)The active (green), inactive (red), intermediate one (blue) and intermediate two (orange) structures identified during the deactivation process are aligned. Residues Y197$^{5.58}$ and Y288$^{7.53}$ are shown as sticks. (B) The potential of mean force (PMF) is calculated as a function of the distance between the ionic lock (R102$^{3.50}$ - E228$^{6.30}$) and the RMSD of Y288$^{7.53}$ in the NPxxY motif relative to the inactive structure.

The profile for the active and inactive conformations is reproduced in shorter MD simulations originating from the starting agonist-bound apo and antagonist-bound apo structures. The RMSD of Y288$^{7.53}$ and the distance between the ionic lock is nearly identical in the shorter 200 ns antagonist-bound apo simulation and four separate 50 ns agonist-bound apo simulations compared to the longer ~1.8 μs antagonist-bound apo simulation and ~1.6 μs agonist-bound apo simulation, respectively (Figure 2.7).

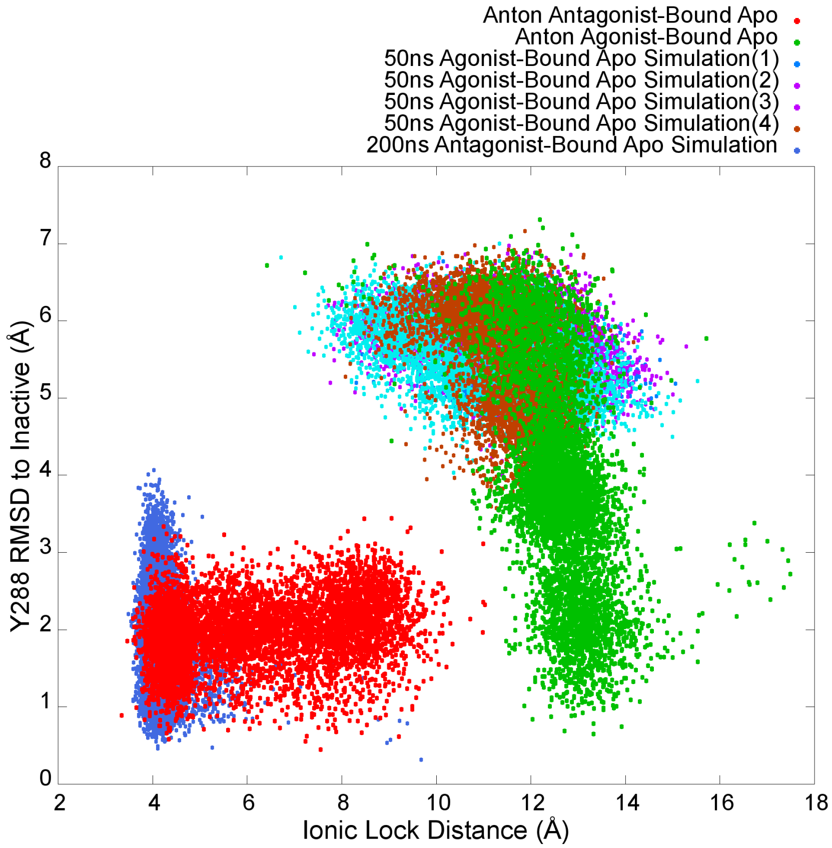

Figure 2.7: The distance between the "ionic lock" (R102$^{3.50}$ - E228$^{6.30}$) and the RMSD of Y288$^{7.53}$ in the NPxxY motif relative to the inactive structure is plotted for the five control simulations, one 200 ns antagonist-bound apo simulation (dark blue) and four 50 ns agonist-bound apo simulations (light blue, purple, turquoise, brown), and the Anton antagonist-bound apo simulation (red) and agonist-bound apo simulation (green).

Deactivation

When the receptor transitions from the active to the intermediate 1 structures, there is a formation of hydrogen bonds with Y288$^{7.53}$ (Figure 2.8). This hydrogen bond network is highly dynamic. During the first 250 ns of the agonist-bound apo simulation, Y288$^{7.53}$ forms no hydrogen bonds, and this time corresponds to the active conformations.

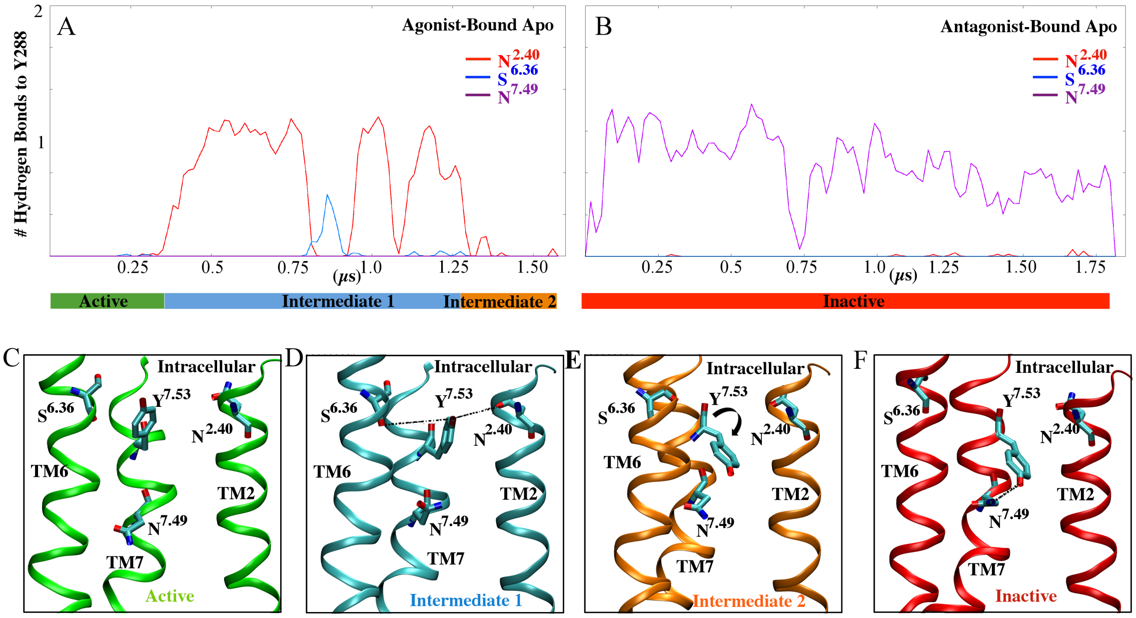

Figure 2.8: The hydrogen bond network of Y288$^{7.53}$ during A$_{2A}$AR deactivation. (A) shows the number of hydrogen bonds formed with Y288$^{7.53}$ during the initial agonist-bound apo simulation. A bar below the x-axis shows the time evolution of the active, intermediate one and intermediate two conformations in green, blue and orange, respectively. (B) shows the number of hydrogen bonds formed with Y288$^{7.53}$ during the initial antagonist-bound apo structure with the inactive conformation shown in red. (C) is a cartoon representation of the hydrogen bonds identified in the active trajectory. Y288$^{7.53}$ is shown as a stick. (D) is a cartoon representation of the hydrogen bonds identified in intermediate one. Residues S234$^{6.36}$, Y288$^{7.53}$ and, N$^{2.40}$ are shown as sticks. (E) is a cartoon representation of the hydrogen bonds in intermediate two. I292$^{C\text{-}TERM}$ and Y288$^{7.53}$ are shown as sticks. (F) is a cartoon representations of hydrogen bonds in inactive trajectories. N284$^{7.49}$, Y288$^{7.53}$ and I292$^{C\text{-}TERM}$ are shown as sticks.

This differs from other active structures, such as rhodopsin, M$_2$ muscarinic and β$_2$AR, in that Y288$^{7.53}$ forms hydrogen bonds with Y197$^{5.58}$ either directly or through a water-mediated bond [48, 51-53]. In the intermediate 1 conformation, Y288$^{7.53}$

alternates hydrogen bonds with residues S234$^{6.36}$, N284$^{7.49}$ of the NPxxY motif, and N42$^{2.40}$. In intermediate 2, the Y288$^{7.53}$ side chain dihedral switches from trans to gauche, which causes a break in hydrogen bonds from residues in TM2 and TM6 (Figure 2.8A and Figure 2.8E). During the initial antagonist-bound apo simulation, Y288$^{7.53}$ forms a hydrogen bond with N284$^{7.49}$ during the entire simulation (Figure 2.8B and Figure 2.8F). These hydrogen bonds reduce the tilt of the backbone atoms of TM7 ~4 Å. While the first step of ‘deactivation’ is the formation of hydrogen bonds, the next step is the side-chain dihedral switch of Y288$^{7.53}$ from trans to gauche (Figure 2.9).

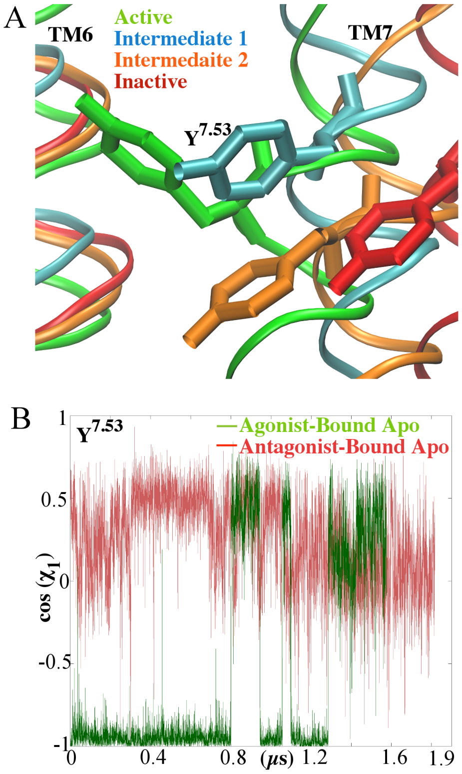

Figure 2.9: Time course of the χ_1 side-chain dihedral of residue Y288$^{7.53}$. (A) is a cartoon representation of the conformation of Y288$^{7.53}$. The aligned active (green), inactive (red), intermediate one (blue) and intermediate two (orange) structures are shown as ribbons. (B) The initial agonist-bound apo simulation is shown in dark green and the initial antagonist-bound simulation is shown in red.

This switch contributes to the movement of TM7, which tilts outwards ~3 Å compared to the starting structure. The gauche conformation of Y288$^{7.53}$ is maintained throughout the initial antagonist-bound apo simulation. During the initial agonist-bound apo simulation, Y197$^{5.58}$ remains in the lipid interface, similar to the inactive conformation of rhodopsin [54], and in the trans side-chain dihedral conformation (Figure 2.10).

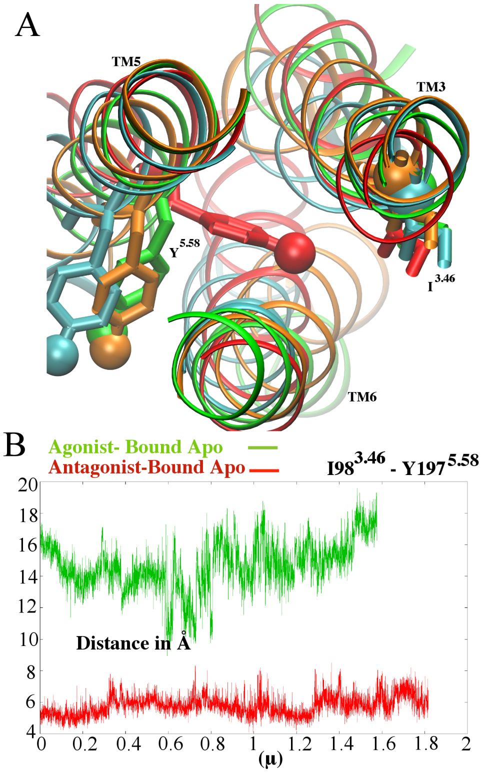

Figure 2.10: Conformation of Y197$^{5.58}$ during initial agonist-bound apo and initial antagonist-bound apo simulations. (A) is a cartoon representation of the conformation of Y197$^{5.58}$. Y197$^{5.58}$ and I98$^{3.46}$ are shown as sticks, and the atoms of Y197$^{5.58}$ (hydroxyl oxygen) and I98$^{3.46}$ (Cα atom) used to calculate distance are shown as balls. The aligned active (green), inactive (red), intermediate one (blue) and intermediate two (orange) structures are shown as cartoon. The calculated distance between I98$^{3.46}$ and Y179$^{5.58}$ in the initial antagonist-bound apo simulation (red), and the initial agonist-bound apo simulation (green) is shown in (B).

In the initial antagonist-bound apo simulation, Y197$^{5.58}$ faces the helical bundle and remains in the gauche conformation, similar to the active conformations of M$_2$, rhodopsin and β$_2$AR [43, 44, 54]. This conformation prevents the hydrogen bond between Y288$^{7.53}$ and Y197$^{5.58}$ from forming. The distance between the hydroxyl oxygen of Y288$^{7.53}$ and Y197$^{5.58}$ remains above 9 Å throughout the entire agonist-bound apo simulation (Figure 2.11). In the antagonist-bound apo simulation, due to the conformational switch of Y197$^{5.58}$ the distance between these two residues is decreased.

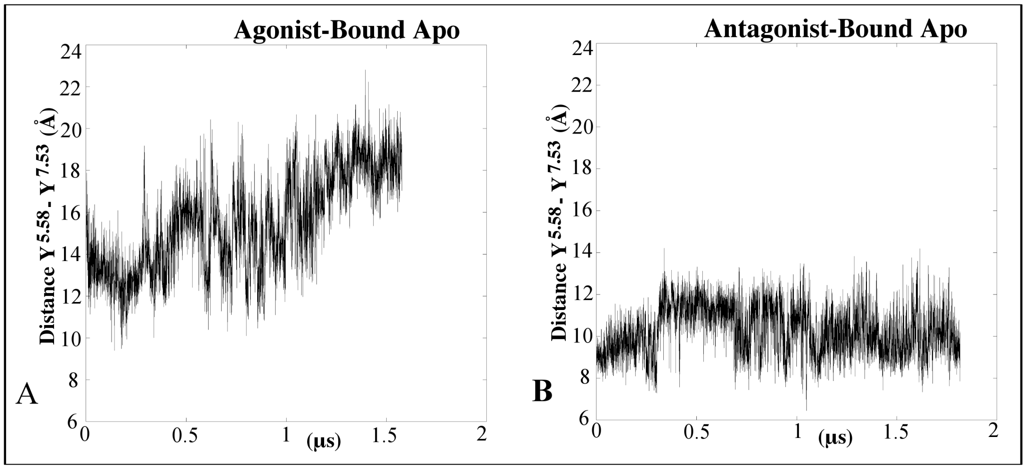

Figure 2.11: The distance between the hydroxyl atoms for Y197$^{5.58}$ and Y288$^{7.53}$ for the agonist-bound simulation (A) and the antagonist bound simulation (B) is plotted.

Figure 2.10 shows the distance between Y197$^{5.58}$ and I98$^{3.46}$, which represents the helical bundle. This distance is greater than 10 Å for most the initial agonist-bound apo simulation, while Y197$^{5.58}$ remains in the lipid interface. In the initial antagonist-bound apo simulation, the distance stays larger than 8 Å, while Y197$^{5.58}$ remains near the helical bundle. Additionally, the movement of Y197$^{5.58}$ is highly coordinated with F201$^{5.62}$ (Figure 2.12A and Figure 2.12B).

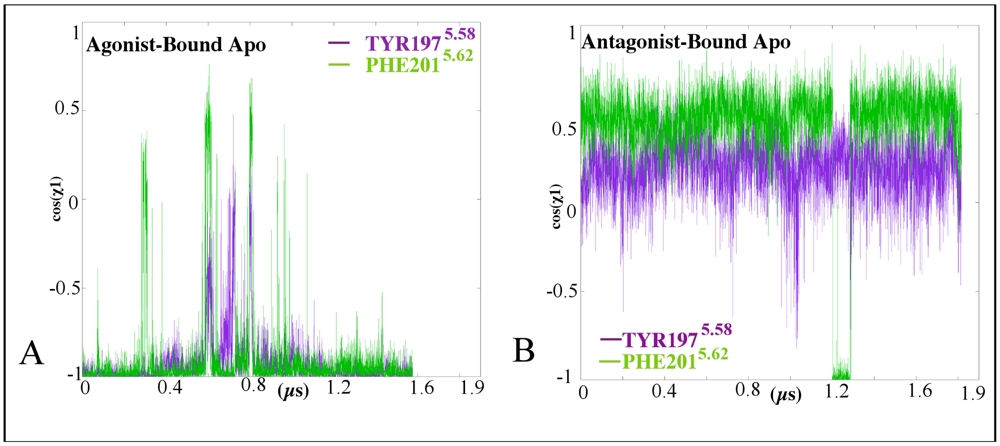

Figure 2.12: The side-chain dihedral χ_1 of Y197$^{5.58}$ and F201$^{5.62}$ (A) in the initial antagonist bound apo simulation and (B) the initial agonist-bound apo simulation is plotted.

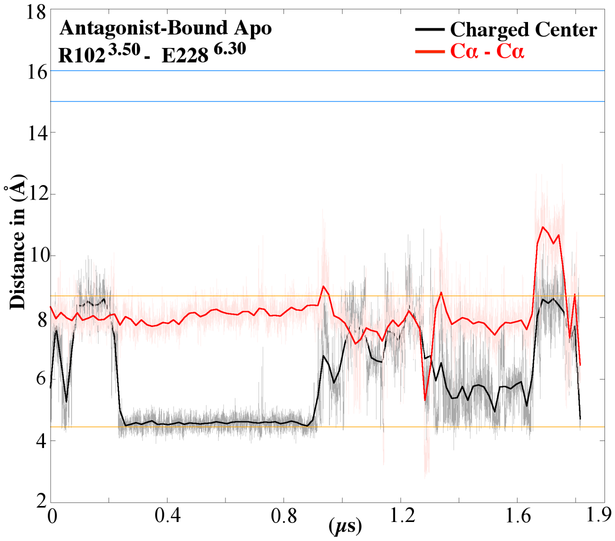

Figure 2.13: Distance between the R102$^{3.50}$ – E228$^{6.30}$ "ionic lock" groups in the initial antagonist-bound apo simulation. The smoothed distance between charged centers of R102$^{3.50}$ and E228$^{6.30}$ is plotted in black and the distance between the Cα atoms in red, while the raw data in shown in gray and pink, respectively. The distances between the charged centers and the Cα atoms (4.5 Å and 8.6 Å, respectively) for inactive rhodopsin (PDB ID: 1F88) [45] are shown as orange lines. The distances between the charged centers and the Cα atoms (16 Å and 14.6 Å, respectively) for active rhodopsin (PDB ID: 3CAP) [55] are shown as blue lines.

During the initial antagonist-bound apo simulation, the ionic lock between R102$^{3.50}$ and E228$^{6.30}$ forms for ~500 ns (Figure 2.13), but it does not form in the agonist-bound apo simulation (Figure 2.14). This lock is characteristic of complete inactivation of GPCRs [53] (Figure 2.13).

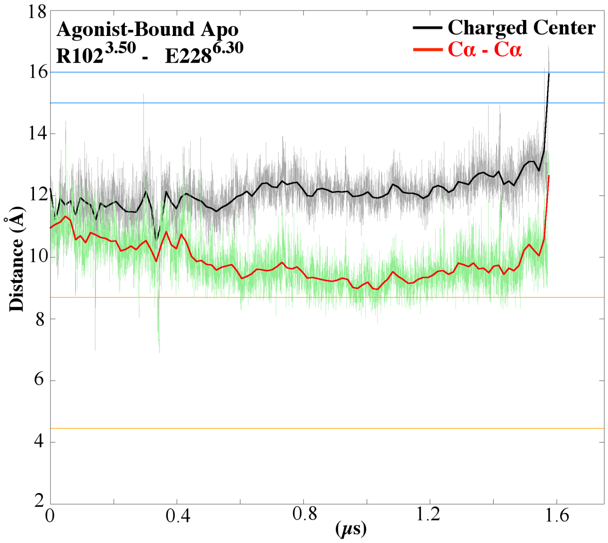

Figure 2.14: Distance between the R102$^{3.50}$ – E228$^{6.30}$ "ionic lock" groups in the initial agonist-bound apo simulation. The smoothed distance between charged centers of R102$^{3.50}$ and E228$^{6.30}$ is plotted in black and the distance between the Cα atoms in red, while the raw data in shown in gray and green, respectively. The distances between the charged centers and the Cα atoms (4.5 Å and 8.6 Å, respectively) for inactive rhodopsin (PDB ID: 1F88) [45] are shown as orange lines. The distances between the charged centers and the Cα atoms (16 Å and 14.6 Å, respectively) for active rhodopsin (PDB ID: 3CAP) [55] are shown as blue lines.

Discussion

Molecular dynamics simulations allow us to visualize the time evolution of molecular motion. In this study, we simulated two A_{2A}AR structures, an agonist-bound and an antagonist-bound both with the ligands removed. Two partial lipids and five stearic acid molecules were crystallized with the agonist-bound [25] and antagonist-bound [28] structures, respectively, but were also removed for the simulations. Lipid composition influences the fluidity of the membrane and can play a part in GPCR stability and function. In rhodopsin, the changes between the MI and MII states are dependent on the lipid environment [56]. Additionally, a recent study compared the molecular dynamics of $A_{2A}AR$ in a POPC bilayer and a POPE bilayer and showed large differences in the interhelical motions depending on the environment [57]. This suggests that changing the lipid environment may influence the rearrangement of TM5, TM6, and TM7 and the time scale of deactivation.

Given the same system starting in different areas of phase space, one would expect convergence of phase space sampling from both simulations given a long time-scale. In a study by Dror et al, the deactivation process of β_2AR took between 400 ns to 4.5 μs to fully complete [48], and experimental data suggests that it takes approximately 40 ms for the intracellular displacement of TM6 to occur [58]. After an investigation of the agonist-bound apo and the antagonist-bound apo simulations, it was determined that the given timescale, ~1.6 μs for each, was not long enough to observe convergence of the ensembles yet interesting phenomena were observed with respect to deactivation upon which hypotheses can be suggested and tested.

The largest differences between the simulations occurred in TM5, TM6, and TM7. A PMF was constructed to determine the approximate probability distribution of the simulations using the RMSD of $Y288^{7.53}$ and the distance between the “ionic lock” groups as reaction coordinates. From this analysis, we can see three different conformations from the agonist-bound simulation and one from the antagonist-bound simulation. In the agonist-bound simulation, $Y288^{7.53}$ participates in a change in the hydrogen bond network, eventually reducing the tilt of TM7 by ~4 Å. Additionally, $Y197^{5.58}$ faces the lipid interface during the agonist-bound simulation, and faces the helical bundle during the antagonist-bound simulation. The conformation of $Y197^{5.58}$

prevents an interaction between Y197$^{5.58}$ and Y288$^{7.53}$, which is key for activation. Based on these results, it is likely that the starting crystal structure of the agonist-bound A_{2A}AR is not in the full active conformation.

The changes in the agonist-bound apo simulation only occur after a sodium ion interacts with D52$^{2.50}$. This residue had an estimated pKA of 8.91 in the agonist-bound structure, but was left deprotonated because it was in solvent in the antagonist-bound structure. In the Anton agonist-bound apo simulation, a sodium ion enters the sodium-binding site, which includes D52$^{2.50}$, S91$^{3.39}$, and N280$^{7.46}$, initially at 44 ns, again at 405 ns, and stably remains in this for 74% of the simulation. The rearrangement of TM7 after sodium binding is consistent with a previous study [59].

Materials and Methods

The residues were numbered using the Ballesteros and Weinstein format of X.YY, where X is the transmembrane helix number, one through seven, and YY is the relative number from the most conserved residue of the transmembrane helix, labeled 50 [60].

System Setup

The starting structure for the agonist-bound simulation of the A_{2A}AR was solved in complex with the agonist UK-432097 (PDB: 3QAK) [25] at 2.71 Å resolution, and the antagonist-bound structure of A_{2A}AR was determined in complex with ZM241385 (PDB: 3EML) [28] at 2.6 Å resolution. Two partial lipids were crystallized with 3QAK and five stearic acid molecules were crystallized with 3EML. Both structures were crystallized with a T4 lysozyme fused to the ICL3. The T4 lysozyme, lipid molecules and the ligands were removed and the missing ECL2 and ICL3 loops and missing residues were modeled using Modeller [61]. All hydrogens were added using the psfgen plugin on VMD [62] and the internal water molecules were added with Dowser [63]. PROPKA was used to predict the pKA of all titratable residues [64]. All titratable residues were left in their dominant protonation state at pH 7.0, except Asp52 and His155, which were deprotonated.

The four disulfide bonds resolved in the crystal structures, Cys71$^{2.69}$ – Cys159$^{5.20}$, Cys74$^{3.22}$ – Cys146$^{4.67}$, Cys77$^{3.25}$ – Cys166$^{5.27}$, and Cys259^{ECL3} – Cys262^{ECL3}, were maintained. The psfgen plugin on VMD [62] was used to generate the system topology in the CHARMM format [65], the palmitoyl–oleoyl-phosphatidyl-choline (POPC) lipid molecules were added using the membrane plugin in VMD [62], and the system was solvated using the Solvate plugin [62]. The net charge of the receptor structures was neutralized and 0.15 M NaCl was added. The final initial agonist-bound apo system, used for five MD simulations, had a total of 170 POPC lipid molecules, 23 sodium ions, 32 chloride ions and 5,392 water molecules with a total of 76,216 atoms. The initial antagonist-bound apo system, used for two MD simulations, had 169 POPC molecules, 23 sodium ions, 33 chloride ions and 5,410 water molecules with a total of 77,289 atoms. Periodic boundary conditions were applied to both simulation systems.

Molecular Dynamic Simulations

All molecular dynamics simulations were initially performed using NAMD 2.8b3 [66]. The CHARMM27 with CMAP parameter set was used for the protein [67, 68], and the CHARMM36 for the POPC lipid molecules [65]. The cutoff distance for the van der Waals and short range electrostatics was 12 Å, and the particle mesh Ewald method was applied for the calculation of long range electrostatic interactions [69]. A 2 fs time step and a multiple-time-stepping algorithm [66] were used with bonded and short range non bonded interactions computed for each time step, and long range electrostatics every two-time steps. The SHAKE algorithm was used to constrain bond lengths to hydrogen atoms [70].

Initially, the lipid tails were minimized for 5,000 steps and simulated for 3 ns with NVT at 300 K. Next, the protein atoms were relaxed for 5 ns with NPT conditions with 10 kcal/(mol*Å^2) harmonic restraints applied. All structures were simulated for an additional 5 ns with only the Cα atoms restrained with 5 kcal/(mol*Å^2) harmonic restraints. One antagonist-bound apo simulation and one agonist-bound apo simulation were simulated for 100 ns with everything released in NPT conditions. The production runs on Anton [71] were initiated from the final structures of the these two MD runs.

The coordinate and velocity files were converted from the NAMD format to Anton using a script (Appendix). The initial agonist-bound apo structure and the antagonist-bound apo structure were then simulated on Anton for 1.57 μs and 1.75 μs, respectively. M-SHAKE was applied to the hydrogen-containing bonds [72], and a simulation time step of 2 fs was used. The cutoff distance for the van der Waals and short range electrostatics was 13.5 Å and the Gaussian Split Ewald method [73] with a 64 × 64 × 64 grid, $\sigma = 2.51$ Å and $\sigma_s = 1.77$ Å, was applied for the calculation of long range electrostatic interactions. Trajectory snapshots were saved every 240 ps for analysis.

Five control simulations with everything released in NPT conditions, including one 200 ns antagonist-bound apo simulation and four 50 ns agonist-bound apo simulations were performed using NAMD [66].

Calculation of Potential of Mean Force

The potential of mean force (PMF) was used to study changes in free energy based on the sampling space of both Anton simulations as a function of directed reaction coordinates. To identify key changes in deactivation, the distance between the ionic lock groups and the RMSD of Y288$^{7.53}$ were chosen as reaction coordinates due to their role in activation. The PMF obtained was calculated using the following equation (1) [74]:

$$\mathrm{A}(\xi_J, \xi_I) = -k_B T \ln\left(\rho\left(\xi_J, \xi_I\right)\right) \tag{1}$$

where ξ_J and ξ_I are the reaction coordinates, k_B is the Boltzmann constant, T is the temperature, and ρ is the probability distribution.

Acknowledgements

We thank Aaron Freidman and Dahlia Goldfeld for valuable discussion. Computing time was provided on the Anton supercomputer, courtesy of D.E. Shaw Research. A.D.C is supported by the NIH Pharmaceutical Sciences Training Grant at UCSD. Additional support is provided by NSF, NIH, HHMI, NBCR, and the NSF supercomputer centers.

Chapter 2 is a minimally modified reprint of the material as it appears in Alisha D. Caliman, Sara E. Swift, Yi Wang, Yinglong Miao, and J. Andrew McCammon, “Investigation of the conformational dynamics of the apo A_{2A} adenosine receptor,” Protein Science, 2015. The dissertation author was the primary investigator and author of this paper.

CHAPTER 3

Mapping the Allosteric Sites of the A_{2A} Adenosine Receptor

Abstract

The A_{2A} adenosine receptor (A_{2A}AR) is a G protein-coupled receptor that is pharmacologically targeted for the treatment of inflammation, sepsis, cancer, neurodegeneration, and Parkinson's disease. Recently, we applied long-timescale molecular dynamics simulations on two ligand-free receptor conformations, starting from the agonist-bound (PDB ID:3QAK) and antagonist-bound (PDB ID:3EML) X-ray structures. This analysis revealed four distinct conformers of the A_{2A}AR: the active, intermediate 1, intermediate 2, and inactive. In this study, we apply the fragment-based mapping algorithm, FTMap, on these receptor conformations to uncover five non-orthosteric sites on the A_{2A}AR. Two sites that are identified in the active conformation are located in the intracellular region of the transmembrane helices (TM) 3/TM4 and the G protein-binding site in the intracellular region between TM2/TM3/TM6/TM7. Three sites are identified in the intermediate 1 and intermediate 2 conformations, annexing a site in the lipid interface of TM5/TM6. Five sites are identified in the inactive conformation, comprising of a site in the intracellular region of TM1/TM7, and in the extracellular region of TM3/TM4 of the A_{2A}AR. We postulate that these sites on the A_{2A}AR be screened for allosteric modulators for the treatment of inflammatory and neurological diseases.

Introduction

The adenosine receptors are a class of four G protein-coupled receptors, A_1, A_{2A}, A_{2B}, and A_3, whose endogenous ligand is the purine adenosine [3, 17, 75]. The adenosine A_{2A} receptor (A_{2A}AR) is expressed in leukocytes, such as T cells, monocytes, macrophages, and natural killer cells (NKT) [3, 13, 17], and in the brain, specifically the GABAergic neurons in the basal ganglia that are responsible for voluntary movement [13, 76]. The receptor couples to the G_s protein in the periphery and the G_{olf} protein in the brain [3, 13]. The G protein stimulates the adenylyl cyclase pathway, causing a surge in cAMP, and the activation of the protein kinase A (PKA) signaling cascade [3, 17, 75, 76].

In leukocytes, the activation of PKA by the A_{2A}AR produces anti-inflammatory effects [3, 17]. Furthermore, a knockout of the A_{2A}AR in mice causes an accumulation of pro-inflammatory cytokines and leads to tissue damage [77]. When the A_{2A}AR is activated on T cells, the adenylyl cyclase/cAMP/PKA pathway leads to the inhibition of NF-κB, a pro-inflammatory transcription factor [17]. Additionally, the pathway activates anti-inflammatory transcription factors, which leads to the decrease in pro-inflammatory cytokines [3]. Because of its role in the immune response, A_{2A}AR agonists are being developed and studied for the treatment of sepsis, which is caused by hyper-inflammation and characterized by multiple organ failure, and multiple sclerosis, where leukocytes attack the myelin on neurons and cause lesions in the brain [3, 75, 78]. Moreover, the anti-inflammatory role of the A_{2A}AR in the immune system is protective to tumors, suggesting that antagonists of the receptor can be developed to treat cancer [79].

Antagonists of the A_{2A}AR is a treatment option for Parkinson's disease. Parkinson's disease is a motor disorder caused by the degradation of the dopamine pathway. In the GABAergic neurons, the activation of the A_{2A}AR acts counter to the D_2 dopamine signaling pathway, and suppresses dopamine release [40].

Overall, A_{2A}AR antagonists can treat Parkinson's disease [40] and cancer [79]. A_{2A}AR agonists can treat sepsis [3] and inflammation [75, 78]. Allosteric modulators bind to sites spatially different from the orthosteric binding site [80, 81]. Because allosteric regions are not as conserved as the orthosteric ones across receptor subtypes, modulators that bind to non-orthosteric sites can maintain subtype specificity. Allosteric modulators alter the association or disassociation rate of the endogenous ligand [80]. They include positive allosteric modulators (PAM), which increase the effect of the orthosteric ligand and negative allosteric modulators (NAM), which decrease the effect of the orthosteric ligand [80].

The structure of the A_{2A}AR consists of 7 transmembrane helices (TM1-TM7) connected by three intracellular loops (ICL1-ICL3) and three extracellular loops (ECL1-ECL3) [24-33]. The receptor has been crystallized in complex with agonists, including adenosine [24-27], antagonists [28-33], and coupled to a G protein fragment [27]. The A_{2A}AR has also been crystallized with interacting lipids [25, 26, 28, 31-33] and sodium ions [32, 33].

Recently, we applied molecular dynamics (MD) to the ligand–free (apo) form of two conformations of the A_{2A}AR X-ray structures, agonist-bound (PDB ID: 3QAK) [25] and antagonist-bound (PDB ID: 3EML) [28]. We uncovered the deactivation process of the A_{2A}AR from the active conformation of 3QAK to the inactive crystal structure 3EML [82]. We identified four distinct conformations of the receptor, the active, intermediate

1, intermediate 2, and inactive, which were characterized particularly by the motions of the Y288$^{7.53}$ residue in the NPxxY motif. The number '288' denotes the residue number, '7' denotes the transmembrane helix number, and '53' represents the location of the residue in the transmembrane helix relative to the most conserved residue (denoted 50) [60]. In the active conformation, the χ_1 angle of Y288$^{7.53}$ remained in the trans conformation and the side-chain faced the intracellular region of the receptor. In the intermediate 1 conformation, Y288$^{7.53}$ interacted with residues in TM2, TM6, and TM7. In the intermediate 2 conformation, the χ_1 angle of Y288$^{7.53}$ switched from trans to gauche conformation, causing a break in the interactions between TM2 and TM6. Finally, in the simulation of the inactive A_{2A}AR, a salt bridge was formed between the D101$^{3.50}$ and E228$^{6.30}$ residues, or the 'ionic-lock' [82].

In this study, we obtain representative conformations of the A_{2A}AR from the previous MD simulations. We identify non-orthosteric sites on the active, intermediate 1, intermediate 2, and inactive conformers, and 20 A_{2A}AR crystal structures using FTMap [38]. FTMap is fragment-based mapping algorithm, which screens the surface of the receptor with small probes, such as ethanol and isopropanol, and identifies hot spots based on regions of the receptor where multiple probes bind at low energy. This tool was previously used to identify five non-orthosteric sites on the β_2 adrenergic receptor (β_2AR), four non-orthosteric sites on the β_1 adrenergic receptor (β_1AR) [83], and seven non-orthosteric sites on the M_2 muscarinic receptor [84]. We classify two non-orthosteric sites on the active conformations, three on the intermediate 1 and intermediate 2 conformations, and five non-orthosteric sites on the inactive conformations of the A_{2A}AR.

Materials and Methods

Simulations

In a previous study [82], we performed two simulations on the ligand-free $A_{2A}AR$ starting from the agonist-bound (PDB ID: 3QAK) [25] and the antagonist-bound (PDB ID:3EML) [28] X-ray structures. 3QAK was crystallized in complex with the agonist UK-432097 at a resolution of 2.71 Å [25]. 3EML was crystallized in complex with the antagonist ZM241385 at a resolution of 2.6 Å [28]. The ligands were removed from each crystal structure. In this study, the term "agonist-bound apo" denotes the ligand-free simulation starting from the 3QAK X-ray structure. The term "antagonist-bound apo" denotes the ligand-free simulation starting from the 3EML X-ray structure. The agonist-bound apo and antagonist-bound apo simulations were performed using all-atom MD simulations on the Anton supercomputer [71] for 1.57 µs and 1.75 µs, respectively, as previously described in Caliman, AD, et al [82].

FTMap

FTMap is a fragment-based site mapping algorithm that identifies potential binding sites on a receptor. The program docks 16 molecules, ethanol, isopropanol, isobutanol, acetone, acetaldehyde, dimethyl ether, cyclohexane, ethane, acetonitrile, urea, methylamine, phenol, benzaldehyde, benzene, acetamide, and N,N-dimethylformamide, to the surface of the receptor [38]. To use this algorithm, PDB files are uploaded to the online server, http://FTMap.bu.edu . The output of FTMap is a PDB file with the coordinates of the receptor and the probes found in low-energy hot spots.

Clustering of the Receptor

In the agonist-bound apo simulation, the receptor transitioned from the active to the intermediate 1 conformation at 250 ns. The receptor transitioned from intermediate 1 to intermediate 2 at 1.25 µs. The antagonist-bound apo simulation remained in the inactive conformation for the entire 1.75 µs [82]. The receptor trajectories were aligned on the backbone atoms of the transmembrane helices, and RMSD-based clustering of the receptor was performed with a 1.5 Å cutoff to identify representative structures in the active, intermediate 1, intermediate 2, and inactive conformations. There were a total of 16 receptor clusters obtained from the agonist-bound apo simulation and 14 receptor clusters from the antagonist-bound apo simulation. Three receptor clusters were found in the active conformer. Nine clusters were found in the intermediate 1 conformer. Four receptor clusters were found in the intermediate 2 conformer, and 14 receptor clusters were found in the inactive conformer [82]. All 30 receptor clusters were submitted to the FTMap server for analysis.

Crystal Structures

Twenty crystal structures were analyzed using the FTMap server. The antagonist-bound crystal structures included are shown in Table 3.1 and the agonist-bound crystal structures are shown in Table 3.2. The crystal structures were prepared with Schrödinger's Protein Preparation Wizard, where missing side chains, residues, and loops were added [85].

Table 3.1. Antagonist-bound X-ray structures used for site-mapping. The PDB ID, ligand name, and resolution, interesting features, and references are shown.

PDB ID	Ligand	Resolution in Å	Interesting Features	Reference
3EML	ZMA	2.6	Lipid Interactions	[28]
3PWH	ZMA	3.3		[29]
3REY	XAC	3.31		[29]
3RFM	Caffeine	3.6		[29]
3UZA	T4G	3.27		[30]
3UZC	T4E	3.34		[30]
3VG9	ZMA	2.7	Lipid Interactions	[31]
4EIY	ZMA	1.8	Lipid Interactions, Sodium	[32]
5IU4	ZMA	1.72	Lipid Interactions, Sodium	[33]
5IU7	6DY	1.9	Lipid Interactions, Sodium	[33]
5IUA	6DX	2.2	Lipid Interactions, Sodium	[33]
5IUB	6DV	2.1	Lipid Interactions, Sodium	[33]
5IU8	6DZ	2	Lipid Interactions, Sodium	[33]

Table 3.2. Agonist-bound X-ray structures used for site-mapping. The PDB ID, ligand name, and resolution, interesting features, and references are shown.

PDB ID	Ligand	Resolution in Å	Interesting Features	Reference
2YDO	Adenosine	3.0		[24]
2YDV	NECA-agonist	2.6		[24]
3QAK	UKA-agonist	2.71	Lipid Interactions	[25]
4UG2 – Subunit A	NGI - agonist	2.6	Lipid Interactions, Two Subunits Crystallized	[26]
4UG2 – Subunit B	NGI - agonist	2.6	Lipid Interactions, Two Subunits Crystallized	[26]
5G53 – Subunit A	NECA + G protein	3.4	Two Subunits Crystallized	[27]
5G53 – Subunit B	NECA + G protein	3.4	Two Subunits Crystallized	[27]

Probe Occupancy Analysis

The total number of probes within 5 Å of the Cα atom of each residue was calculated for every receptor cluster and X-ray structure. The probe occupancy designates the probability that a probe was found within 5 Å of a residue. The probe occupancy was calculated by:

$$Occupancy_{\,i} = \frac{1}{P}\sum_{j=1}^{N_c} P_{ij}\ , \qquad i=1, \ldots\ldots N_{res} \tag{1}$$

where P_j is the number of probes within 5 Å of a residue i in receptor cluster j , N_c is the total number of receptor clusters, N_{res} is the total number of residues, and P is the total number of possible probes that could interact with that residue, or

$$P = 16 \times N_c\ , \tag{2}$$

where 16 is the total number of probes and N_c is the total number of receptor clusters.

The binding site residues plotted in the probe occupancy figures include I66$^{2.64}$, V84$^{3.32}$, L85$^{3.33}$, T88$^{3.36}$, Q89$^{3.37}$, I92$^{3.40}$, L167^{ECL2}, F168^{ECL2}, E169^{ECL2}, M177$^{5.38}$, N181$^{5.42}$, W246$^{6.48}$, L249$^{6.51}$, H250$^{6.52}$, N253$^{6.55}$, T256$^{6.58}$, H264, L267$^{6.32}$, M270$^{7.35}$, Y271$^{7.36}$, I274$^{7.39}$, S277$^{7.42}$, and H278$^{7.43}$ (Figure 3.1 and Figure 3.2).

Center-of-Mass

If the distance between the center-of-mass for one hot spot was within 5 Å of the center-of-mass of another hot spot, they were considered a part of one larger site, and identified as a non-orthosteric site.

Figures 3.3 - 3.13 were generated with VMD [62].

Results and Discussion

Probe Distribution in X-ray Structures

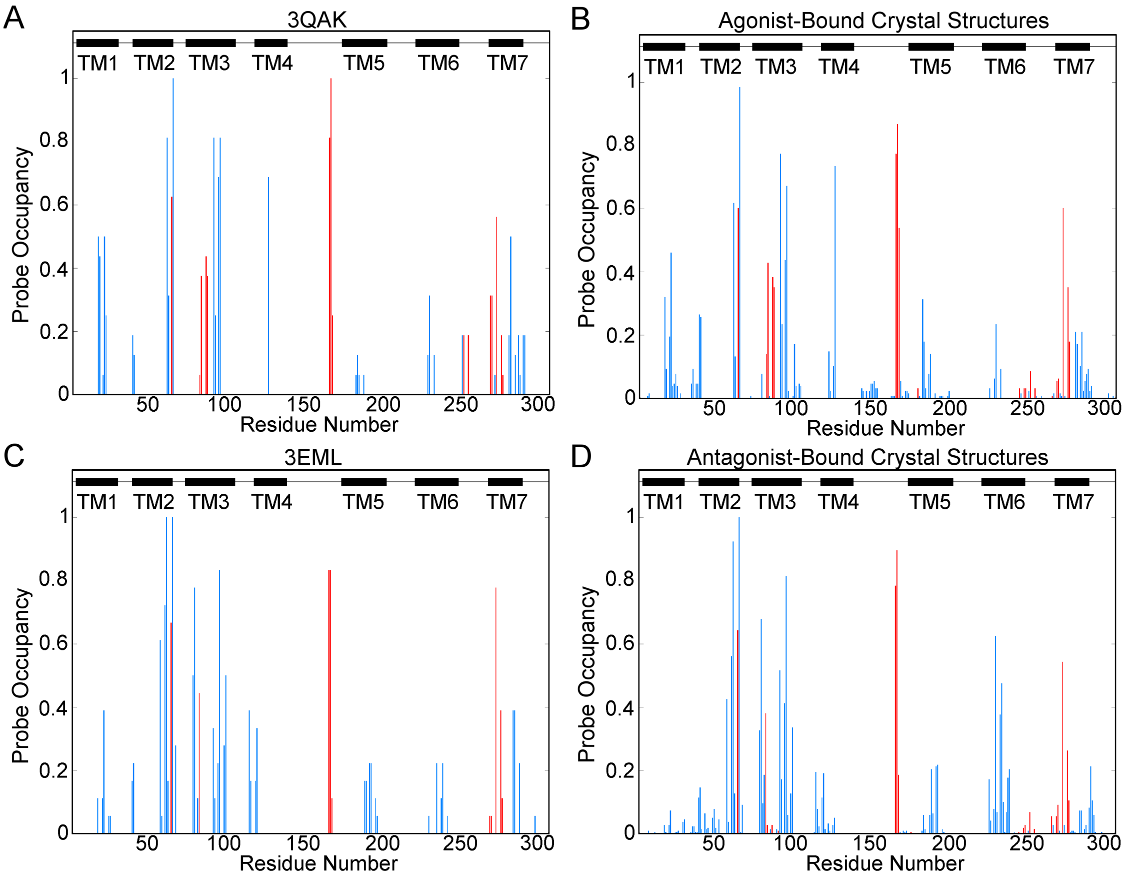

Figure 3.1: Probe Occupancy of A$_{2A}$AR X-ray Structures. Probe Occupancies for the starting agonist-bound X-ray structure, 3QAK, [A], all agonist-bound X-ray structures, [B], the starting antagonist-bound X-ray structure, 3EML, [C], and all antagonist-bound X-ray structures, [D]. Binding site residues are shown in red, the remaining residues are shown in blue, and black boxes represent the transmembrane regions. ICL1 is located in between TM1 and TM2. ECL1 is located between TM2 and TM3. ICL2 is located between TM3 and TM4. ECL2 is located between TM4 and TM5. ICL3 is located between TM5 and TM6, and ECL3 is located between TM6 and TM7.

The overall distribution of probes in a receptor obtained from the FTMap analysis can help us to identify specific residues within the receptor with a propensity for interacting with drug fragments. The probe occupancy for the simulation starting agonist-bound (PDB ID: 3QAK) and simulation starting antagonist-bound (PDB ID: 3EML) X-ray structures were calculated (Figure 3.1A and Figure 3.1C). Additionally, the probe occupancy for all antagonist-bound and all agonist-bound X-ray structures as listed in Table 3.1 and Table 3.2, respectively, were calculated (Figure 3.1B and Figure 3.1D).

In the simulation starting agonist-bound X-ray structure, probes interact with residues in the TM, ICL1, and ECL2 regions (Figure 3.1A). In all agonist-bound X-ray structures, the distribution of the probes to the receptor is identical to the distribution of the simulation starting X-ray structure, but the number of residues in TM1, ICL1, TM3, ECL2, and TM6 within 5 Å of a probe increases (Figure 3.1B).

In the simulation starting antagonist-bound X-ray structure, probes interact with the TM, ECL1, ECL2, and ICL2 regions of the receptor (Figure 3.1C). In the antagonist-bound X-ray structures, there are fewer residues within 5 Å of a probe in TM1, while there are more residues in the ICL1, ECL1, and TM7 regions that are within 5 Å of a probe (Figure 3.1D).

Given that there are only slight differences in the probe distribution between the X-ray structures (Figure 3.1), and that X-ray structures represent static snapshots of the receptor, an FTMap analysis on MD structures will provide a more complete description of the probe distribution, and reveal transient sites in different conformers of the receptor.

Residues with the Highest Probe Occupancies

In the MD representative structures, probes interact primarily with the transmembrane regions, with residues with the highest probe occupancies being found in the TM2, TM3, ECL2, TM5, and TM7 regions in the agonist-bound apo simulation (Figure 3.2A), and in the TM2, TM3, TM6, and TM7 regions in the antagonist-bound apo simulation (Figure 3.2B).

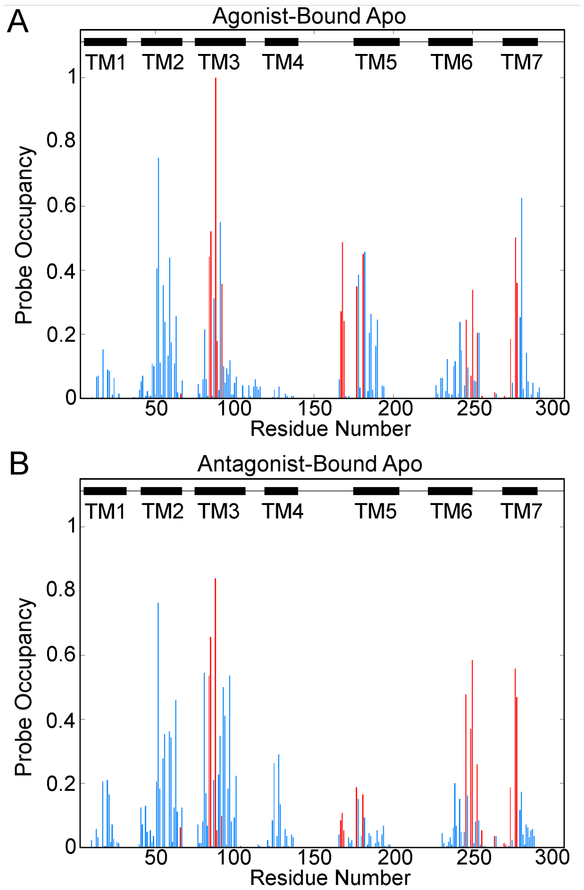

Figure 3.2: Probe Occupancy Per Residue during the Molecular Dynamic Simulations. Probe occupancies for the agonist-bound apo simulation [A], and the antagonist-bound apo simulation [B] are shown. Binding site residues are shown in red, non-orthosteric residues are shown in blue, and black boxes represent the transmembrane regions. ICL1 is located in between TM1 and TM2. ECL1 is located between TM2 and TM3. ICL2 is located between TM3 and TM4. ECL2 is located between TM4 and TM5. ICL3 is located between TM5 and TM6, and ECL3 is located between TM6 and TM7.

In the X-ray structures, the residues with the highest probe occupancies are A63$^{2.61}$, S67$^{2.65}$, A97$^{3.45}$, L167^{ECL2}, and F168^{ECL2} (Figure 3.1 and Table 3.3). The residues with the highest probe occupancies in both simulations are T88$^{3.36}$, D52$^{2.50}$, and L85$^{3.33}$ (Figure 3.1, Figure 3.2, and Table 3.3).

Table 3.3: Probe Occupancies of Key Residues. The probe occupancies of key residues in the starting agonist-bound X-ray structures, 3QAK, the starting antagonist-bound X-ray structure, 3EML, all of the agonist-bound X-ray structures, all of the antagonist-bound X-ray structures, the agonist-bound apo simulation, and the antagonist-bound apo simulation are listed.

Residues	3QAK	3EML	Agonist-Bound X-ray	Antagonist-Bound X-ray	Agonist-Bound Apo	Antagonist-Bound Apo
Highest Probe Occupancies in the X-ray Structures						
A63$^{2.61}$	0.81	1	0.62	0.92	0.26	0.46
S67$^{2.65}$	1	1	0.98	1	0.06	0.13
A97$^{3.45}$	0.81	0.83	0.67	0.81	0.12	0.54
L167^{ECL2}	0.81	0.83	0.77	0.78	0.27	0.08
F168^{ECL2}	1	0.83	0.87	0.90	0.49	0.11
Highest Probe Occupancies in the Simulations						
T88$^{3.36}$	0.44	0	0.38	0.03	1	0.84
D52$^{2.50}$	0	0	0	0.02	0.75	0.76
L85$^{3.33}$	0.34	0	0.43	0.03	0.52	0.66
Probe Occupancies that are Higher in the Agonist-Bound X-Ray Structures						
C128$^{4.49}$	0.69	0	0.73	0.05	0.04	0.29
F93$^{3.41}$	0.81	0.33	0.77	0.52	0.1	0.5
V283$^{7.48}$	0.5	0	0.17	0	0	0.01
Probe Occupancies that are Higher in the Agonist-Bound Apo Simulation						
S91$^{3.39}$	0	0	0	0.01	0.55	0.35
I92$^{3.40}$	0	0	0	0.01	0.36	0.10
S281$^{7.46}$	0	0	0	0.01	0.62	0.17
Probe Occupancies that are Higher in the Antagonist-Bound X-Ray Structures						
A81$^{3.29}$	0	0.78	0.08	0.68	0.22	0.54
D101$^{3.49}$	0	0.5	0.01	0.33	0.07	0.22
Probe Occupancies that are Higher in the Antagonist-Bound Apo Simulation						
W246$^{6.48}$	0	0	0.03	0	0.25	0.48

Residue T88$^{3.36}$ is located in the sodium ion binding site [59] and the orthosteric binding site in the X-ray structures of the agonist-bound A$_{2A}$AR [24-26]. The probe occupancy surrounding T88$^{3.36}$ is higher in the agonist-bound X-ray structures (0.38) compared with the antagonist-bound X-ray structures (0.03) (Table 3.3). This is consistent with previous findings that mutating T88$^{3.36}$ affects agonist-binding, but not antagonist binding [24, 25]. In the agonist-bound apo simulation, the probe occupancy for T88$^{3.36}$ is 1.0, and in the antagonist-bound apo simulation, the probe occupancy is 0.84 (Table 3.3). The increase in probe occupancies in the agonist-bound apo and antagonist-bound apo simulations compared to the X-ray structures is due to the entrance of a sodium ion into the sodium ion binding pocket during both simulations (Figure 3.3)

The probe occupancy of D52$^{2.50}$ is 0.75 in the agonist-bound apo simulation and 0.76 in the antagonist-bound apo simulation (Table 3.3). D52$^{2.50}$ is a conserved residue located in the sodium ion-binding site [32, 33, 59]. A sodium ion is not found in either the simulation starting agonist-bound X-ray structure [25] or the simulation starting antagonist-bound X-ray structure [28]. Appropriately, the probe occupancy surrounding D52$^{2.50}$ is 0.0 in the simulation starting agonist-bound and antagonist-bound X-ray structures, as well as in most A_{2A}AR X-ray structures (Table 3.3). It has been suggested that sodium binding to D$^{2.50}$ triggers deactivation of the A_{2A}AR and other GPCRs [32, 59, 86]. In the agonist-bound apo simulation, a sodium ion enters the site at 405 ns (Figure 3.3B), after the receptor has transitioned from the active state to intermediate 1 [82]. In the antagonist-bound apo simulation, a sodium ion enters the site after 34 ns (Figure 3.3C).

The probe occupancy surrounding L85$^{3.33}$ is 0.52 in the agonist-bound apo simulation and 0.66 in the antagonist-bound apo simulation. L85$^{3.33}$ is a residue in the orthosteric binding site in the agonist-bound and antagonist-bound X-ray structures [25, 28]. However, the probe occupancy in the simulation starting agonist-bound X-ray structure is much higher (0.34) than in the simulation starting antagonist-bound (0.0) X-ray structure (Table 3.3). The probe occupancy is 0.43 in the all agonist-bound and 0.03 in all antagonist-bound X-ray structures (Table 3.3).

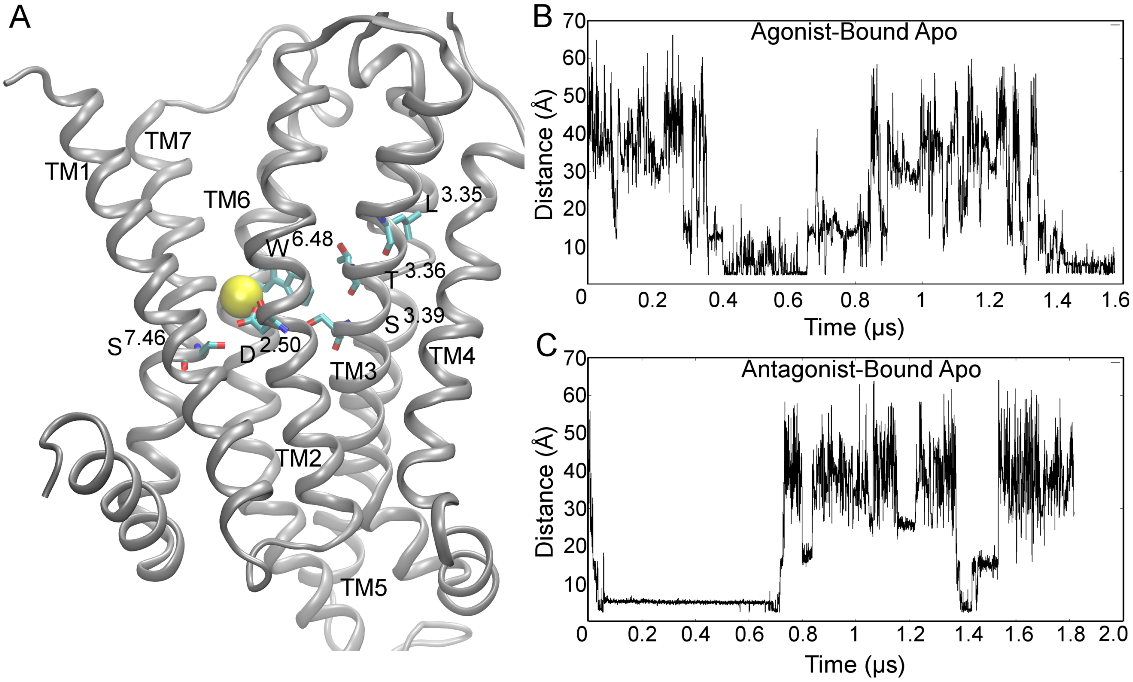

Figure 3.3: Sodium Ion Binding Site. The sodium ion binding site, consisting of residues D52$^{2.50}$, L85$^{3.33}$, T88$^{3.36}$, S91$^{3.39}$, S281$^{7.46}$, and W256$^{6.48}$, is shown in [A]. The sodium ion is represented as a yellow sphere. The distance of the closest sodium ion to the Cα of D52$^{2.50}$ is shown in the agonist-bound [B] and antagonist-bound [C] simulations.

Higher Probe Occupancies Observed in MD Simulations of the A_{2A}AR

Residues C128$^{4.49}$, F93$^{3.41}$, and V283$^{7.48}$ have higher probe occupancies in the agonist-bound X-ray structures compared to the antagonist- bound X-ray structures (Figure 3.1B, Figure 3.1D, and Table 3.3). The TM5, ECL2 and residues S91$^{3.39}$, I92$^{3.40}$, and S281$^{7.46}$ have higher probe occupancies in the agonist-bound apo simulation structures compared to the antagonist-bound apo simulation structures (Figure 3.2A and Table 3.3). The conformation of TM5 is different in the agonist-bound apo simulation compared to the antagonist-bound apo simulation. TM5 is more flexible in the agonist-bound apo simulation than in the agonist-bound apo simulation [82]. Additionally, during receptor deactivation, residue Y197$^{5.58}$ rotates the side chain towards the lipid interface, while in the antagonist-bound apo simulation, Y197$^{5.58}$ faces the helical bundle [82]. Residues S91$^{3.39}$ and S281$^{7.46}$ interact with a sodium ion [32, 59], which enters the sodium binding site in both simulations (Figure 3.3). S91$^{3.39}$ has a probe occupancy of 0.55 in the agonist-bound apo simulation and 0.35 in the antagonist bound apo simulation, compared to 0.0 in the X-ray structures (Table 3.3). S281$^{7.46}$ has a probe occupancy of 0.62 in the agonist-bound apo simulation and 0.17 in the antagonist-bound apo simulation, compared to 0.0 in the X-ray structures (Table 3.3).

TM2, TM3, A81$^{3.29}$, and D101$^{3.49}$ have higher probe occupancies in the antagonist-bound X-ray structures than in the agonist-bound X-ray structures (Figure 3.1D and Figure 3.1B). TM4 and TM6 (including residue W246$^{6.48}$) have higher probe occupancies in the antagonist-bound apo simulation, compared to the agonist-bound simulation. W$^{6.48}$ is located in the conserved CWxP site, a rotameric trigger for GPCR activation [87], and residue in the sodium ion binding site (Figure 3.3A) [59]. W246$^{6.48}$ is also a key residue identified in the intrinsic water pathway in A$_{2A}$AR [88]. During the antagonist-bound apo simulation, intracellular water molecules enter into the receptor and interact with W246$^{6.48}$ (Figure 3.4B).

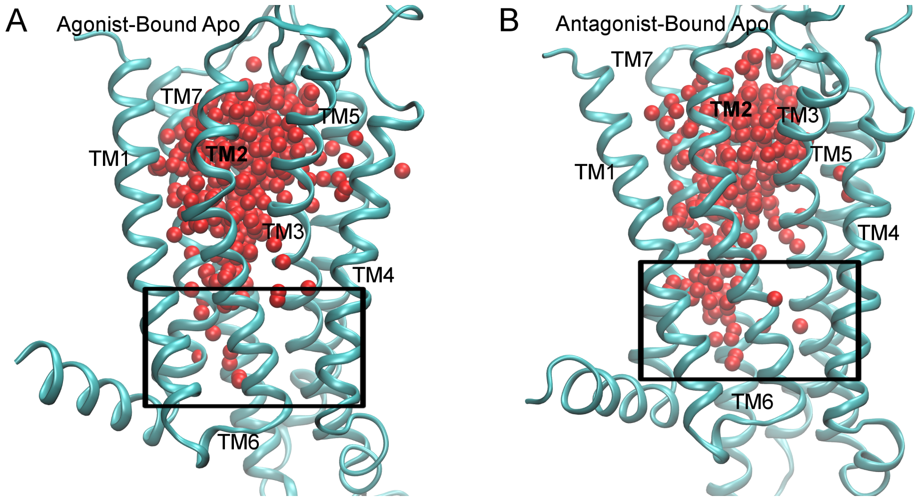

Figure 3.4: Internal Water Channel. A cluster of internal water molecules in the agonist-bound [A], and antagonist-bound, [B] simulations is shown.

Hot Spots of Probe Molecules Identified in the A_{2A}AR during Deactivation

Different conformers of the A_{2A}AR were identified from the long time-scale molecular dynamics simulations: the active, intermediate 1, intermediate 2, and inactive [82]. FTMap analysis revealed sites and relevant key residues on each of these conformers (Table 3.4 and Figure 3.5). The inactive conformer has more non-orthosteric sites compared with the active conformer (Figure 3.5), consistent with findings on the simulation starting agonist and antagonist-bound X-ray structures (Figure 3.6).

Each conformer and the X-ray structures have an orthosteric site (Figure 3.5 and Figure 3.6). Residues in the orthosteric site include I66$^{2.64}$, V84$^{3.32}$, L85$^{3.33}$, T88$^{3.36}$, Q89$^{3.37}$, I92$^{3.40}$, L167^{ECL2}, F168^{ECL2}, E169^{ECL2}, M177$^{5.38}$, N181$^{5.42}$, W246$^{6.48}$, L249$^{6.51}$, H250$^{6.52}$, N253$^{6.55}$, T256$^{6.58}$, H264^{ECL3}, L267$^{7.32}$, M270$^{7.35}$, Y271$^{7.36}$, I274$^{7.39}$, S277$^{7.42}$, and H278$^{7.43}$. This site is smaller in the inactive conformer and the simulation starting antagonist-bound X-ray structure than in the active conformer and simulation starting agonist-bound X-ray structure (Figure 3.5A, Figure 3.5D, and Figure 3.6).

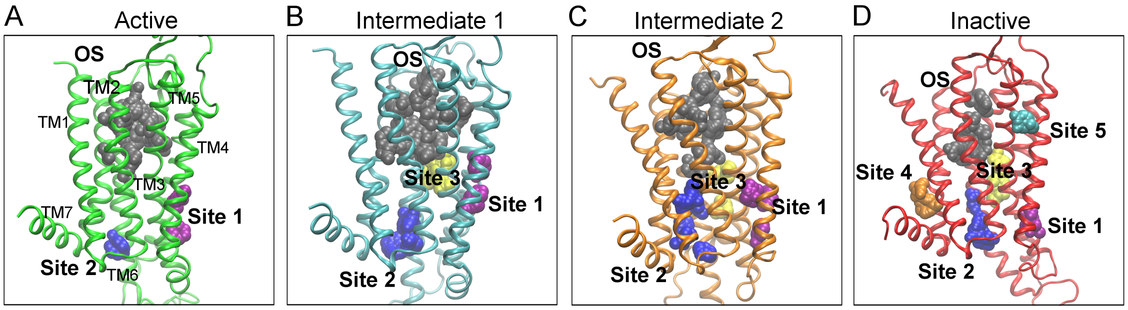

Figure 3.5: The Hot Spots on the A_{2A}AR. The hot spots for probe-binding in the active (green) [A], intermediate 1 (cyan) [B], intermediate 2 (orange) [C], and inactive (red) [D] conformers are shown. The orthosteric site (OS) is shown in gray, site 1 in purple, site 2 in blue, site 3 yellow, site 4 in orange, and site 5 in cyan.

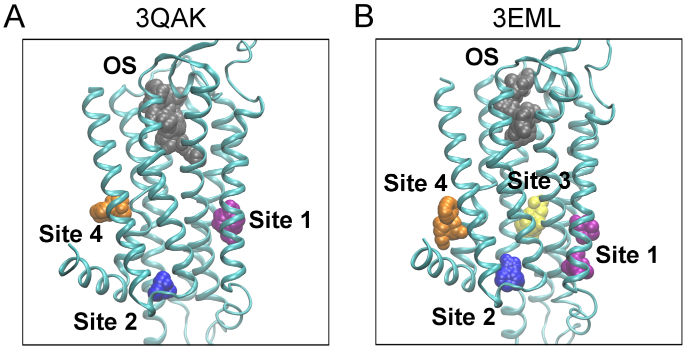

Figure 3.6: Receptor Hotspots in Starting Structures. The hotspots in the 3QAK [A] and 3EML [B] X-ray structures are shown. The orthosteric site (OS) is shown in gray, site 1 in purple, site 2 is in blue, site 3 in yellow, and site 4 in orange.

Non-orthosteric site 1 is located in the intracellular crevice between TM3, TM4, and TM5 (Figure 3.5 and Figure 3.7). Non-orthosteric site 2 is located in the G protein-coupling site of TM2, TM3, TM6, and TM7 (Figure 3.5 and Figure 3.8). These sites are found in the active, intermediate 1, intermediate 2, and inactive conformers (Figure 3.5), as well as the simulation starting agonist-bound and antagonist-bound X-ray structures (Figure 3.6).

Table 3.4: Non-orthosteric sites in the $A_{2A}AR$ conformers are listed. The site number, location, transmembrane helices and residues are shown.

Site Number	Location	Regions	Residues
1	Intracellular Crevice	TM3/TM4/TM5	$I^{3.40}$, $F^{3.41}$, $L^{3.44}$, $A^{3.45}$, $D^{3.49}$ (TM3) $I^{4.45}$, $I^{4.48}$, $C^{4.49}$ (TM4) Y112 (ICL1) $C^{5.46}$, $P^{5.50}$ (TM5)
2	G Protein-Coupling Site	TM2/TM3/TM6/TM7	N39 (ICL1) $T^{2.39}$, $N^{2.40}$ (TM2) $D^{3.49}$, $R^{3.50}$ (TM3) $H^{6.32}$, $S^{6.36}$, $F^{6.44}$ (TM6) $Y^{7.53}$ (TM7) I292 (C-Term)
3	The Lipid Interface	TM5/TM6	$P^{5.50}$, $M^{5.54}$ (TM5) $V^{6.41,}$ $F^{6.44}$, $W^{6.48}$ (TM6)
4	C-Terminus Cleft	TM1/TM7	$L^{1.45}$, $G^{1.49}$, $L^{1.52}$ (TM1) $V^{7.47}$, $P^{7.50}$, $F^{7.51}$ (TM7)
5	Extracellular Cleft	TM3/TM4	$C^{3.30}$, $F^{3.31}$, $V^{3.34}$ (TM3) $L^{4.58}$, $G^{4.57}$, $F^{4.54}$ (TM4)

Non-orthosteric site 3 is located on the lipid interface of TM5 and TM6 (Figure 3.6 and Figure 3.9). This site is found in the intermediate 1, intermediate 2, and inactive conformers (Figure 3.5B, Figure 3.5C, and Figure 3.5D), and in the simulation starting antagonist-bound X-ray structures (Figure 3.6B). Non-orthosteric site 4 is located on the intracellular end of TM1 and TM7, near the C-terminus (Figure 3.5D and Figure 3.11). It is found in the inactive conformer (Figure 3.5D), the simulation starting agonist-bound X-ray, and the simulation starting antagonist-bound X-ray structures

(Figure 3.6). Non-orthosteric site 5 is located between the extracellular ends of TM3 and TM4 (Figure 3.5D and Figure 3.13). This site is found in the inactive conformer (Figure 3.5D).

The key residues in each of these sites are listed in Table 3.4.

Site 1: The Intracellular Crevice

The “intracellular crevice” is located in between TM3, TM4, and TM5. The three regions in this site are denoted 1.1, 1.2, and 1.3 (Figure 3.7), and correspond to the different probe clusters in the receptor. Site 1.1 is positioned closest to the intracellular region and involves residues Y112^{ICL2}, D101$^{3.49}$ and A97$^{3.45}$, and I124$^{4.45}$ (Figure 3.7A and Figure 3.7B). D$^{3.49}$ is a conserved residue of the DRY motif and Y112^{ICL2} is a key residue in the G protein-coupling site [27]. This region of the intracellular crevice is located in the active, intermediate 2, and inactive conformers (Figure 3.5A, Figure 3.5C, and Figure 3.5D), and the simulation starting antagonist-bound X-ray structure (Figure 3.6B). The aromatic and hydrophobic probes that interact with this region are cyclohexane and benzene.

Site 1.2 is present in every structure (Figure 3.5 and Figure 3.6) and encompasses residues I92$^{3.40}$, C185$^{5.46}$, and P189$^{5.50}$ (Figure 3.7D). Hydrophobic cyclohexane, benzene, and phenol, and hydrophilic isobutanol interact with this site. Site 1.3 is only located in the intermediate 1 conformer (Figure 3.5B), and comprises of residues I92$^{3.40}$, C185$^{5.46}$, and P189$^{5.50}$ (Figure 3.7D). Probes in this region include benzaldehyde, benzene, cyclohexane, acetone, and phenol. Site 1.1 and 1.2 are separated by I124$^{4.45}$ and A97$^{3.45}$, while site 1.2 and 1.3 are separated by F93$^{3.41}$ (Figure 3.7B and Figure 3.7D).

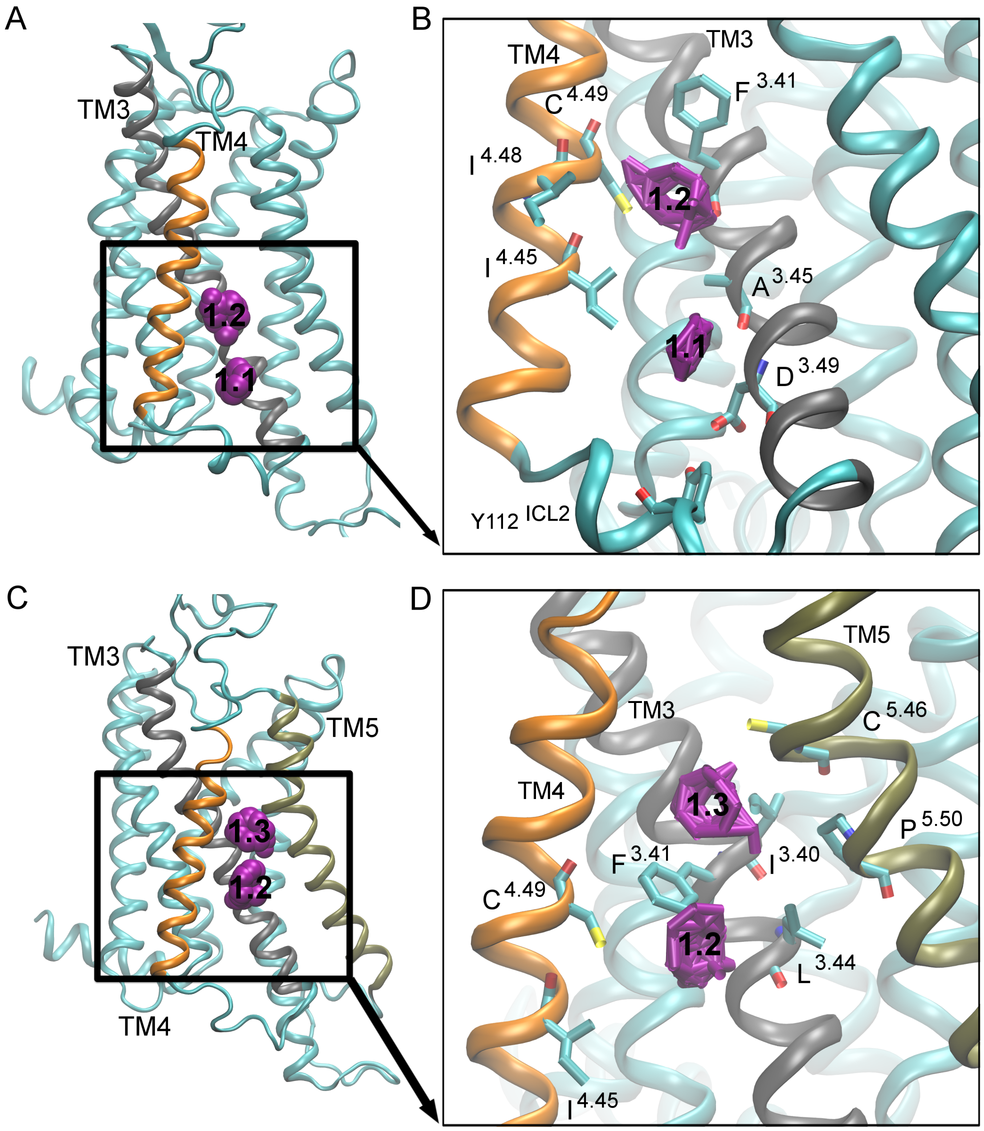

Figure 3.7: Hot spot 1: 'The Intracellular Crevice' is located in the intracellular region between TM3 (gray), TM4 (orange) and TM5 (tan). The probes in the active [A] and the intermediate 1 conformers [C] are shown as purple spheres, with the numbers '1.1', '1.2' and '1.3' denoting different regions of the pocket. The key interacting residues for the active clusters [B] and intermediate 1 conformers [D] are shown as bonds, and the probes that bind to this site are shown as purple bonds.

The intracellular crevice site is also found in the β_1AR, β_2AR, denoted sites 3 and 5 in Ref. [83], and the M_2 muscarinic receptor, denoted site 4 in Ref. [84]. In the M_2 receptor, there are two components to this site, 4.1 and 4.2. Site 4.1, located closest to the intracellular region, can be found in the inactive and intermediate 2 conformers. In ref. [84], the conformers of the M_2 receptor include the inactive, intermediate 1, intermediate 2 and the active, where the intermediate 1 is found between inactive and intermediate 2, and the intermediate 2 is found between intermediate 1 and active. Site 4.2 is identified in the intermediate 1 and active states of the M_2 receptor.

Site 2: G Protein-Coupling Site

The “G protein-coupling site” is located between the intracellular ends of the TM2, TM3, TM6, and TM7 (Figure 3.8). Allosteric modulators designed to bind to this site could prevent or enhance the binding of the G protein, attenuating the signaling cascade. In the active conformer, residues encompassing the G protein-coupling site include $H230^{6.32}$, $S234^{6.36}$, $Y288^{7.53}$, $R102^{3.50}$, and $I292^{C\text{-}TERM}$, and the probes that interact with this site are cyclohexane, N, N-dimethylformamide, and phenol (Figure 3.8B). In the inactive conformer, the interacting residues are $D101^{3.49}$, $R102^{3.50}$, $T41^{2.39}$, $F242^{6.44}$, $I287^{7.52}$, $Y288^{7.53}$, $N42^{2.40}$, and $N39^{ICL1}$ (Figure 3.8D).

Probes found in this site include methylamine, acetone, cyclohexane, ethanol, urea, benzene, N, N-dimethylformamide, acetamide, and benzaldehyde. The G protein coupling site is smaller in the active conformer and the simulation starting agonist-bound structure (Figure 3.5A, Figure 3.6A, and Figure 3.8A), than in the intermediate 1, intermediate 2 and inactive conformers, and simulation starting antagonist-bound X-ray structure (Figure 3.5B, Figure 3.5C, Figure 3.5D, Figure 3.6B, and Figure 3.8C). In the antagonist-bound apo simulation, an internal water channel floods the intracellular region of the receptor (Figure 3.4B). In the agonist-bound apo simulation, fewer water molecules enter the receptor (Figure 3.4A).

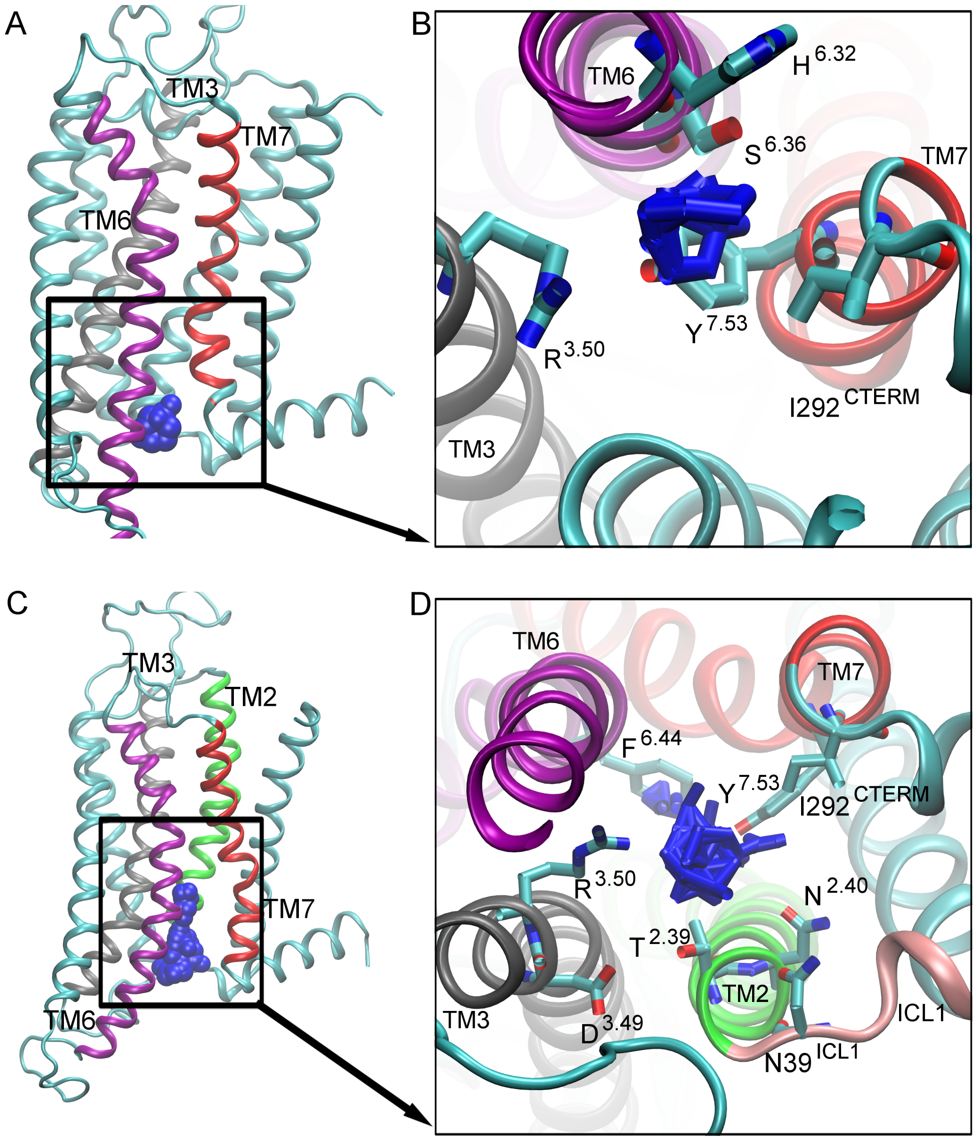

Figure 3.8: Hot spot 2: 'The G Protein-Coupling Site' is located between TM2 (green), TM3 (gray), TM6 (purple), and TM7 (red). The probes in the active [A] and inactive [C] conformers are shown as blue spheres. The key interacting residues for the active conformers [B] and inactive conformers [D] are shown as bonds, and the probes that bind to this site are shown as blue bonds.

The residues in this site are key to deactivation of the A_{2A}AR [82]. Residue Y288$^{7.53}$ is oriented toward the intracellular region of the receptor in the active conformation. In the intermediate 1, Y288$^{7.53}$ forms hydrogen bonds with N42$^{2.40}$ and S234$^{6.36}$. Next, the side chain of Y288$^{7.53}$ switches from the gauche to the trans conformation. In the inactive conformation, a salt bridge is formed between R102$^{3.50}$ and E228$^{6.30}$. Additionally, F242$^{6.44}$ is a residue in the sodium ion binding site [89]. R$^{3.50}$ is a key residue in the G protein binding site [27].

This G protein-coupling site is found in the β_1AR, β_2AR, denoted site 4 in ref. [83], and the M_2 receptor, denoted site 7 in ref. [84]. In the M_2 receptor, this site is identified only in the active state.

Site 3: The Lipid Interface

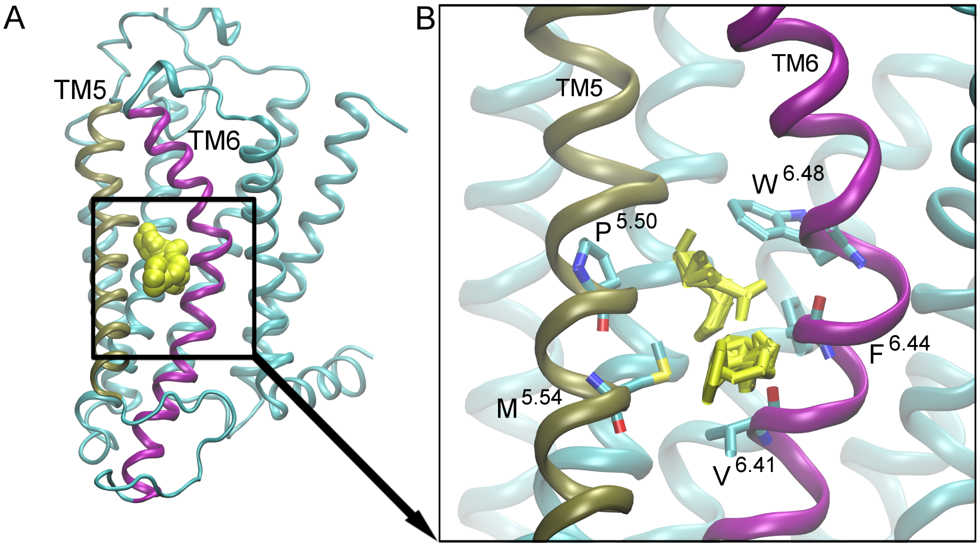

Figure 3.9: Hot spot 3: 'The Lipid interface' is located between TM5 (tan) and TM6 (purple). The probes are shown as yellow spheres [A]. The key interacting residues are shown as bonds [B], and the probes that bind to this site are shown as yellow bonds.

The "lipid interface" site is located on the lipid exposed region of TM5 and TM6 (Figure 3.9). It is present in the intermediate 1, intermediate 2, and inactive conformers (Figure 3.5B, Figure 3.5C, and Figure 3.5D), and the simulation starting antagonist-bound X-ray structure of the A_{2A}AR (Figure 3.6B). The residues P189$^{5.50}$, W246$^{6.48}$, F242$^{6.44}$, M193$^{5.54}$, and V239$^{6.41}$ interact with probes, such as acetone, benzene, ethane, acetaldehyde, acetamide, isobutanol, isopropanol, urea, ethanol, dimethyl ether, and acetonitrile (Figure 3.9B).

The conformations of P189$^{5.50}$ and F242$^{6.44}$ determine whether this site is present in the conformers (Figure 3.9 and Figure 3.10). In the active conformer, the side chains of key residues in this site, P189$^{5.50}$ and F242$^{6.44}$, face towards each other. The distance between the Cα atoms of these two residues is smaller than 10 Å (Figure 3.10B). When the receptor transitions to the intermediate 1 conformation in the agonist-bound apo simulation, the distance between P189$^{5.50}$ and F242$^{6.44}$ increases to an average of 12 Å (Figure 3.10B). This distances increases to 15 Å in the intermediate 2 conformer (Figure 3.10B). In the antagonist-bound apo simulation, the distance between these two residues is consistently above 15 Å (Figure 3.10C).

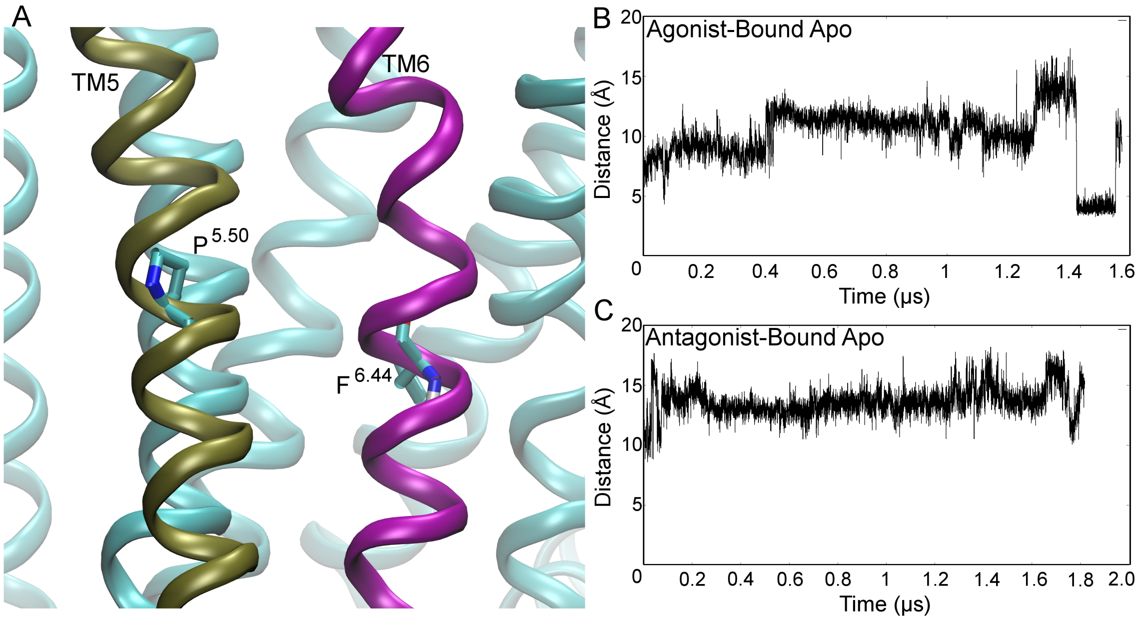

Figure 3.10: Conformation of P189$^{5.50}$ and F242$^{6.44}$ in Site 3. A representation of P189$^{5.50}$ and F242$^{6.44}$ in the inactive cluster is shown [A]. The distance between the Cα atoms of P189$^{5.50}$ and F242$^{6.44}$ is shown in the agonist-bound apo [B] and the antagonist-bound apo [C] simulations.

While this site is not present in the β$_1$AR or β$_2$AR receptors [83], it is present in the inactive M$_2$ muscarinic receptor, denoted site 2 in ref. [84]. Allosteric modulators designed to bind to site 3 could lock the protein in the inactive or intermediate conformers, by interacting with residues P189$^{5.50}$, F242$^{6.44}$, and W246$^{6.48}$ (the “toggle switch” for GPCR activation) [87].

Site 4: The C-Terminus Cleft and Site 5: Extracellular Cleft

The “C-terminus cleft” is located in the intracellular end of TM1 and TM7 (Figure 3.11). It is present in the inactive conformer (Figure 3.5D) and the simulation starting antagonist-bound X-ray structure (Figure 3.6B). The residues located in the C-terminus cleft include V282$^{7.47}$, F286$^{7.51}$, P285$^{7.50}$, L26$^{1.52}$, G23$^{1.49}$, L22$^{1.48}$, and L19$^{1.45}$. These residues interact with benzene, cyclohexane, phenol, N, N-dimethylformamide, acetone, cyclohexane, isobutanol, and benzaldehyde (Figure 3.11B).

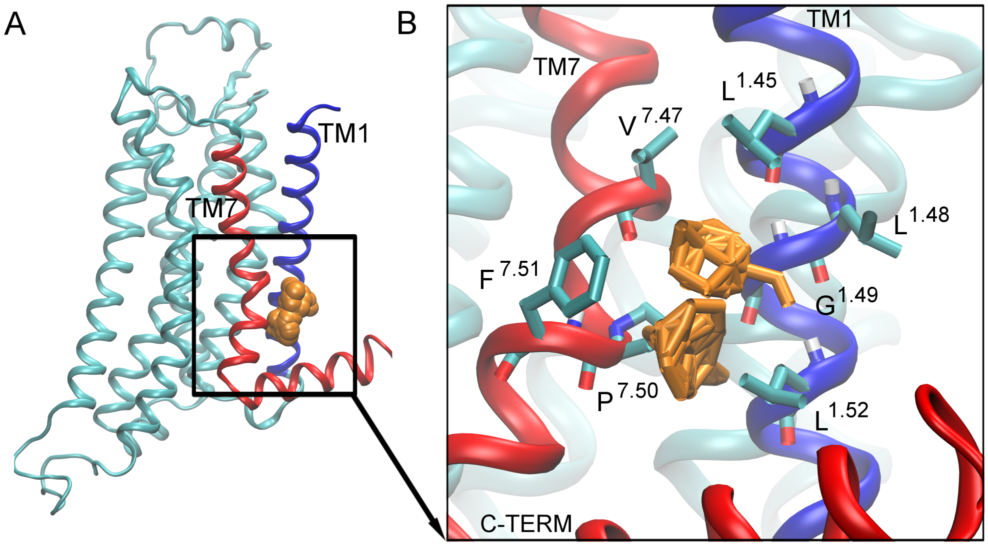

Figure 3.11: Hot spot 4: ‘The C-Terminus Cleft’ is located between TM1 (blue), TM7 (red), and TM8 (red). The probes are shown as orange spheres [A]. The key interacting residues are shown as bonds [B], and the probes that bind to this site are shown as orange bonds.

This cleft is caused by a tilt in TM1 in the inactive conformer relative to the active conformer (Figure 3.12). The side chains of residues V282$^{7.47}$ and L19$^{1.45}$ move towards each other with a distance of 4.9 Å, while the side chain of residues F286$^{7.51}$ and L26$^{1.52}$ move away from each other with a distance of 10.0 Å, forming a cleft (Figure 3.12).

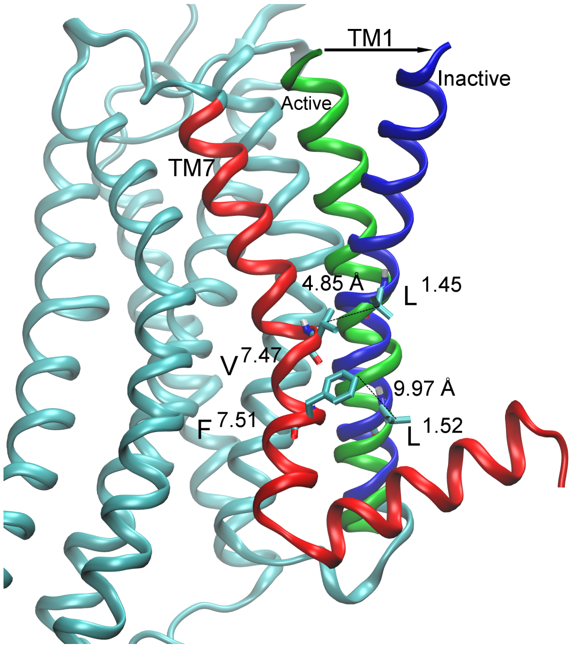

Figure 3.12: Conformation of TM1 and TM7 in Site 4. A representative of TM7 in the active and inactive conformers is shown in red. TM1 in the active conformers is shown in green and TM1 in the inactive is shown in blue. L26$^{1.52}$, L19$^{1.45}$, F286$^{7.51}$, and V282$^{7.47}$ are shown as bonds. The distance between the Cγ of L19$^{1.45}$ and the Cβ of V282$^{7.47}$ in the inactive cluster is 4.85 Å, while the distance between Cγ of L26$^{1.52}$ and the Cζ of F286$^{7.51}$ in the inactive cluster is 9.97 Å.

The "extracellular cleft" is located between the extracellular ends of TM3 and TM4 (Figure 3.13). It is present in the inactive conformer (Figure 3.5D). The residues located in the extracellular cleft site include F83$^{3.31}$, C82$^{3.30}$, L137$^{4.58}$, G136$^{4.57}$, F133$^{4.54}$, and V86$^{3.34}$ (Figure 3.13B). This site interacts with acetone, benzene, dimethyl ether, ethane, and phenol.

The extracellular cleft is only present in the inactive conformer of the A_{2A}AR (Figure 3.5D), but in the M_2 receptor, the extracellular cleft is found in intermediate 2 and active conformers, denoted as site 3.2 in ref [84].

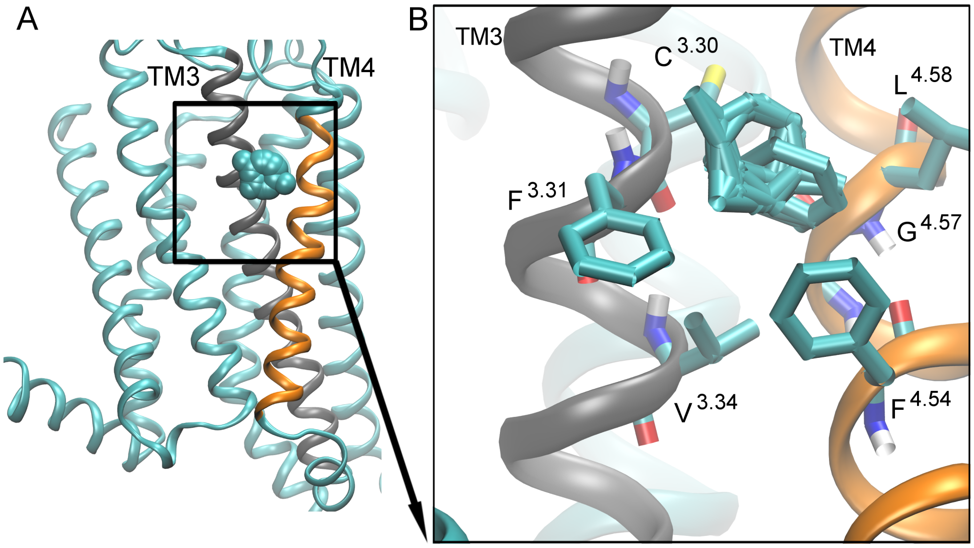

Figure 3.13: Hot spot 5: 'The Extracellular Cleft' is located between TM3 (gray), and TM4 (orange) helices. The probes are shown as cyan spheres [A]. The key interacting residues are shown as bonds [B], and the probes that bind to this site are shown as cyan bonds.

Conclusion

The A_{2A}AR is a GPCR that plays a key role in the immune and neurological systems [17, 40]. In the immune system, the receptor is anti-inflammatory, and activation of the A_{2A}AR leads to an inhibition of pro-inflammatory cytokines and a secretion of anti-inflammatory cytokines [3, 17, 77]. In the neurological disorders, such as Parkinson's Disease, A_{2A}AR activation acts counter to the D_2 dopamine receptor [40]. Given these roles, A_{2A}AR agonists are being developed for the treatment of immunological diseases, such as sepsis, and A_{2A}AR antagonists are being developed for the treatment of PD [76].

Allosteric modulators bind to regions outside of the orthosteric site. Some of these modulators do not have activity on their own, but either increase or decrease the activity of the endogenous ligand [80]. In this study, we applied FTMap, a fragment mapping technique, to analyze representative receptor structures obtained from previous agonist-bound apo and antagonist-bound apo MD simulations [82] and the 20 X-ray structures of the A_{2A}AR to identify allosteric sites on the receptor.

First, we calculated the probe occupancy per residue. In the MD simulation representative clusters and the X-ray structures, probes mostly interacted with residues in the TM regions. Many residues with high probe occupancies, I92$^{3.40}$, S281$^{7.46}$, T88$^{3.36}$, D52$^{2.50}$, L85$^{3.33}$, and W246$^{6.48}$, are involved in sodium binding. Sodium cannot bind to the active state of the receptor. Previous A_{2A}AR studies have shown that a sodium ion can bind to when the receptor is in complex with an antagonist, but cannot bind when the receptor is bound to an agonist [59]. Additionally, studies of the M_3 muscarinic receptor reveal that sodium binding locks the receptor in the inactive conformation [86]. The propensity of these probes to bind to residues of the sodium ion binding site suggest that negative allosteric modulators could be developed to target this region of the receptor.

In the agonist-bound apo and antagonist-bound apo MD simulations, four key receptor conformers, active, intermediate 1, intermediate 2, and inactive, were identified during deactivation of the A_{2A}AR [82]. We identified five non-orthosteric sites on the A_{2A}AR that can be targeted for designing non-orthosteric modulators. Overall, the inactive conformer and antagonist-bound X-ray structures exhibited more non-orthosteric sites than the active conformer and agonist-bound X-ray structures.

The intracellular crevice, present on all conformers, is located between the intracellular ends of TM3/TM4/TM5. This site is also present in the β_1AR, β_2AR, and M_2 muscarinic receptor. Key interacting residues include D101$^{3.49}$ of the conserved DRY motif. The G protein-coupling site, present on all conformers, is located in the intracellular mouth of TM2/TM3/TM6/TM7. This site is also present in the β_1AR, β_2AR, and M_2 muscarinic receptor. Key interacting residues include Y288$^{7.53}$ of the NPxxY motif, and R102$^{3.50}$ of the conserved DRY motif. This site is larger in the inactive conformer than in the active conformer due the intracellular influx of water molecules. Given the role of two key residues involved in GPCR activation, Y288$^{7.53}$ and R102$^{3.50}$ of the ionic lock, either PAMS or NAMs could be developed to bind this site.

The lipid interface site, present on the intermediate 1, intermediate 2 and inactive conformers, is located on the lipid interface between TM5/TM6. This site is present in the M_2 muscarinic receptor, and includes key residues including W246$^{6.48}$. The C-terminus cleft is located in the intracellular end of TM1/TM7. The extracellular cleft is present on the extracellular region of TM3/TM4. These sites are not found in the active conformer, and thus could be targeted for designing novel NAMs.

Overall, these sites provide an array of available non-orthosteric sites on the active, intermediate 1, intermediate 2, and inactive A_{2A}AR conformers. These sites can be screened for novel PAMs for the treatment of sepsis and novel NAMs for the treatment of PD and cancer.

Acknowledgements

Computing time was provided on the Anton supercomputer, courtesy of D.E. Shaw Research. Additional support is provided by the National Science Foundation

(NSF grant: MCB1020765), National Institutes of Health (NIH grant: GM31749), Howard Hughes Medical Institute, National Biomedical Computation Resource (NBCR), and the National Institute of Health Training Grant (NIH Training Grant: T32GM007752).

Chapter 3 is a modified reprint of the material as it will appear in Alisha D. Caliman, Yinglong Miao, and J. Andrew McCammon, “Mapping the Allosteric Sites of the A_{2A} Adenosine Receptor,” Chemical Biology & Drug Design, 2017. The dissertation author was the primary investigator and author of this paper.

CHAPTER 4

Activation Mechanisms of the First Sphingosime-1-Phosphate Receptor

Abstract

Activation of the first sphingosine-1-phosphate receptor ($S1PR_1$) promotes astrocyte and neuronal protection, and lymphocyte egress from secondary lymphoid tissues. Although an agonist often activates the $S1PR_1$, the receptor exhibits high levels of basal activity. In this study, we performed long-timescale molecular dynamics and accelerated molecular dynamics (aMD) simulations to investigate activation mechanisms of the ligand-free (apo) $S1PR_1$. In the aMD enhanced sampling simulations, we observed four independent events of activation, which is characterized by close interaction between $Y311^{7.53}$ and $Y221^{5.58}$ and increased distance between the intracellular ends of transmembrane helices (TM) 3 and 6. Although TM3, TM6, TM5 and, TM7 are associated with GPCR activation, we discovered that their movements are not necessarily correlated during receptor activation. Instead, TM5 showed a decreased correlation with each of these regions during activation. During activation of the apo receptor, $Y221^{5.58}$ and $Y311^{7.53}$ became more solvated, because a water channel formed in the intracellular pocket. Additionally, a lipid molecule repeatedly entered the receptor between the extracellular ends of TM1 and TM7, providing important insights into the pathway of ligand entry into the $S1PR_1$.

Introduction

Sphingosine-1-phosphate (S1P) is a zwitterionic ligand derived by the degradation of ceramide and the subsequent phosphorylation of sphingosine by either of two kinases, sphingosine kinase 1 or 2 [90]. S1P has low solubility in water and inserts itself into the plasma membrane to signal through five G protein-coupled receptors (GPCRs), S1PR$_{1-5}$ [90]. S1P signaling influences a number of physiological processes in the cardiovascular, renal, and lymphatic systems [90].

Activation of the S1PR$_1$ by S1P promotes astrocyte and neuronal protection, and lymphocyte egress from secondary lymphoid tissues [20]. During lymphocyte egress, the S1PR$_1$ acts as a chemoattractant receptor. In secondary lymphoid tissues, there is an S1P gradient. S1P levels are low where T cells enter the lymphoid tissue. In comparison, they are high in the exit region, or the medullary sinus, and the blood [20]. When the S1P concentration is high, the S1PR$_1$ expression levels on the plasma membrane are low [91-93]. When the S1PR$_1$ is activated, it couples to Gα_i and ultimately leads to the internalization and recycling of the receptor [90].

The inactive structure of the S1PR$_1$ was crystalized while fused to a T4 lysozyme in complex with the antagonist ML056 [94]. As with all class A GPCRs, the structure consists of seven transmembrane helices (TM1-7) that are connected by 3 extracellular loops (ECL1-3) and three intracellular loops (ICL1-3). One novel feature in this structure is that the receptor-ligand binding pocket is occluded by an alpha helical N-terminus [94].

X-ray structures have been determined for several GPCRs in the active form, including the β$_2$ adrenergic receptor (β$_2$AR) [42], rhodopsin [43], the M$_2$ muscarinic receptor [46], and the μ-opioid receptor [95]. Activation of these GPCRs is characterized by a rearrangement of TM5, TM6 and TM7, to accommodate the G protein or mimetic nanobody in the intracellular pocket [42-44, 95, 96]. Several micro-switches are involved in this rearrangement, including a dihedral switch of the conserved $W^{6.48}$, the inward movement of the NPxxY motif, and a conformational change of the DRY motif [42-44, 95-97]. The rearrangement of TM5, TM6, and TM7 leads to close interaction between the $Y^{5.58}$ and $Y^{7.53,}$ either through a direct or water-bridged hydrogen bond, and a separation of the TM3 and TM6 cytoplasmic ends compared to the inactive structures.

Computational tools have been used to determine the activation and deactivation pathways of several GPCRs [48, 51, 82, 98-101], and have correctly predicted the features of active structure of GPCRs [44, 51]. The activation of the $S1PR_1$ bound to the S1P agonist was investigated in two previous computational studies using conventional molecular dynamics (cMD) [98, 101]. In both cases, activation of the $S1PR_1$ is initiated by a rotameric conformational change in $W269^{6.48}$ from the gauche to the trans conformation due to agonist binding [98, 101]. This switch leads to the rearrangements of several residues including conserved $D91^{2.50}$ and $N307^{7.49}$ of the NPxxY motif [101], and interactions between the TM1 – TM4 and TM2 – TM7 helices [98]. Particularly, the rearrangement of the NPxxY motif facilitates influx of the cytoplasmic waters, which precedes the intracellular rearrangement of TM5, TM6, and TM7 [101].

The $S1PR_1$ exhibits high levels of basal activity [102], but activation of the receptor in the ligand –free (apo) form has not been investigated. Activation of a GPCR without ligand binding has yet to be observed using cMD methods. In a study by Miao et al. [51], the activation pathway of the apo M_2 muscarinic receptor was captured through accelerated molecular dynamics (aMD) simulations. aMD enhances conformational sampling of the protein by applying a boost potential to the system when the potential energy falls below a threshold [103, 104]. This reduces the energy barrier and thus accelerates protein conformational transitions between different low-energy states. The aMD simulations revealed significant conformational changes of the M_2 muscarinic receptor upon activation, notably close interaction between $Y^{5.58}$ and $Y^{7.53}$ and outward tilting of the TM6 intracellular end [51]. The subsequent release of the active X-ray M_2 receptor supported these predictions [46, 105].

In this study, we perform extensive cMD and aMD simulations to explore the dynamics of the apo and antagonist-bound forms of the $S1PR_1$. We uncovered the basal activation pathway of the $S1PR_1$. Interactions between $Y311^{7.53}$ and $Y221^{5.58}$ and water molecules during receptor activation are discussed. Additionally, our simulations reveal the entry of a lipid molecule, which may provide important insights into the ligand-binding pathway.

Results

System Dynamics

The $S1PR_1$ crystal structure (PDB: 3V2Y) with the antagonist ML056 bound and unbound (apo) were used as a starting structure for six cMD simulations. The final structures from the first cMD simulation of the antagonist-bound and the first cMD simulation of the apo receptor were used as starting structures for six total aMD simulations (Table 4.1).

Table 4.1: The List of Molecular Dynamics Simulations. The label, system, simulation type starting structure, and duration are all included.

ID	System	Simulation	Starting Structure	Duration (ns)
cMD Apo 1	Apo	cMD	X-Ray without ligand	200
cMD Apo 2	Apo	cMD	X-Ray without ligand	200
cMD Apo 3	Apo	cMD	X-Ray without ligand	200
cMD Holo 1	Antag-Bound	cMD	X-Ray	200
cMD Holo 2	Antag-Bound	cMD	X-Ray	200
cMD Holo 3	Antag-Bound	cMD	X-Ray	200
aMD Apo 1	Apo	DB aMD	Final Structure of cMD Apo 1	384
aMD Apo 2	Apo	DB aMD	Final Structure of cMD Apo 1	200
aMD Apo 3	Apo	DB aMD	Final Structure of cMD Apo 1	200
aMD Holo 1	Antag-Bound	DB aMD	Final Structure of cMD Holo 1	335
aMD Holo 2	Antag-Bound	DB aMD	Final Structure of cMD Holo 1	200
aMD Holo 3	Antag-Bound	DB aMD	Final Structure of cMD Holo 1	200

In the cMD simulations, the root-mean-square deviations (RMSDs) of all atoms of the $S1PR_1$ relative to the starting structure leveled off to ~ 4 Å after 20 ns for both the antagonist-bound and apo forms (Figure 4.1). In comparison, the aMD simulations sampled a larger conformational change of the receptor. The protein RMSDs leveled off to 9-12 Å after 125 ns for the antagonist-bound simulations (Figure 4.1A). The protein RMSDs for the aMD simulations of the apo receptor leveled off to 8-10 Å after 125 ns for the second and third simulations and after 200 ns for the first aMD simulation (Figure 4.1B).

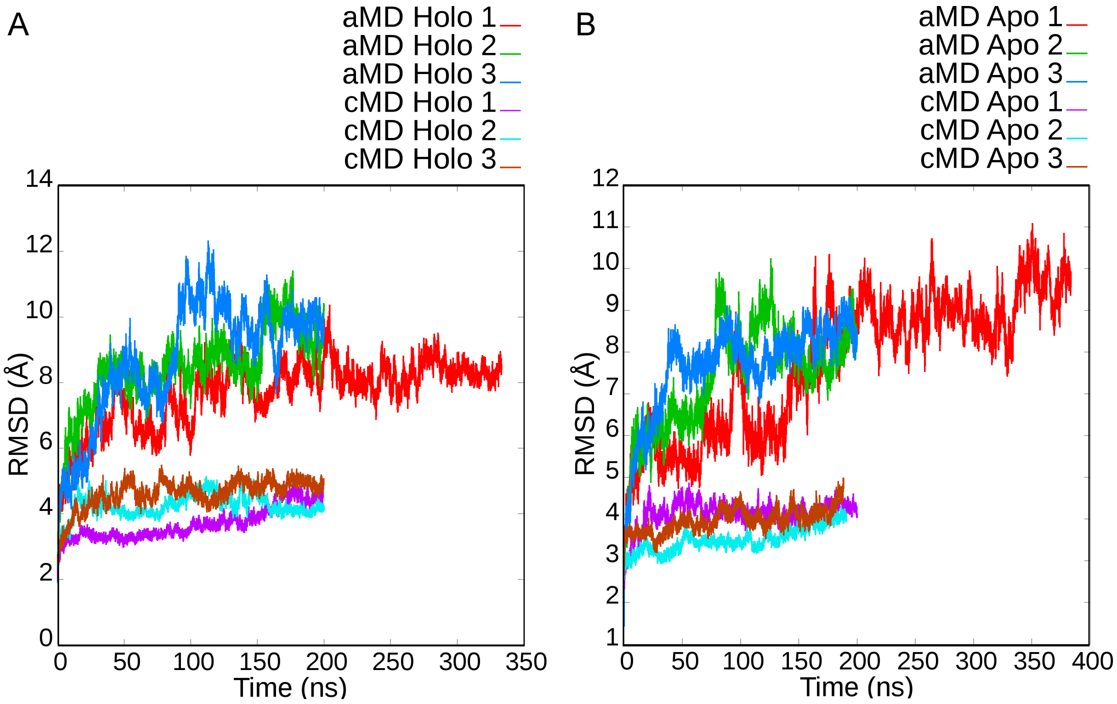

Figure 4.1: Root-Mean-Square Deviation of the Antagonist-Bound (A) and Apo (B) Simulations. The cMD simulations are shown in turquoise, purple, and burgundy and the aMD simulations are shown in red, green, and blue.

Considering the root-mean-square fluctuations (RMSFs), the most flexible regions of the protein were found in the loops, especially the third intracellular loop (ICL3), and the N- and C-termini (Figure 4.2) in the cMD and aMD simulations. The TM2-7 helices were less flexible. TM1, connected to the N-terminal helix, was flexible in the aMD simulations. Overall, applying aMD to the S1PR$_1$ enhanced the molecular dynamics of the receptor.

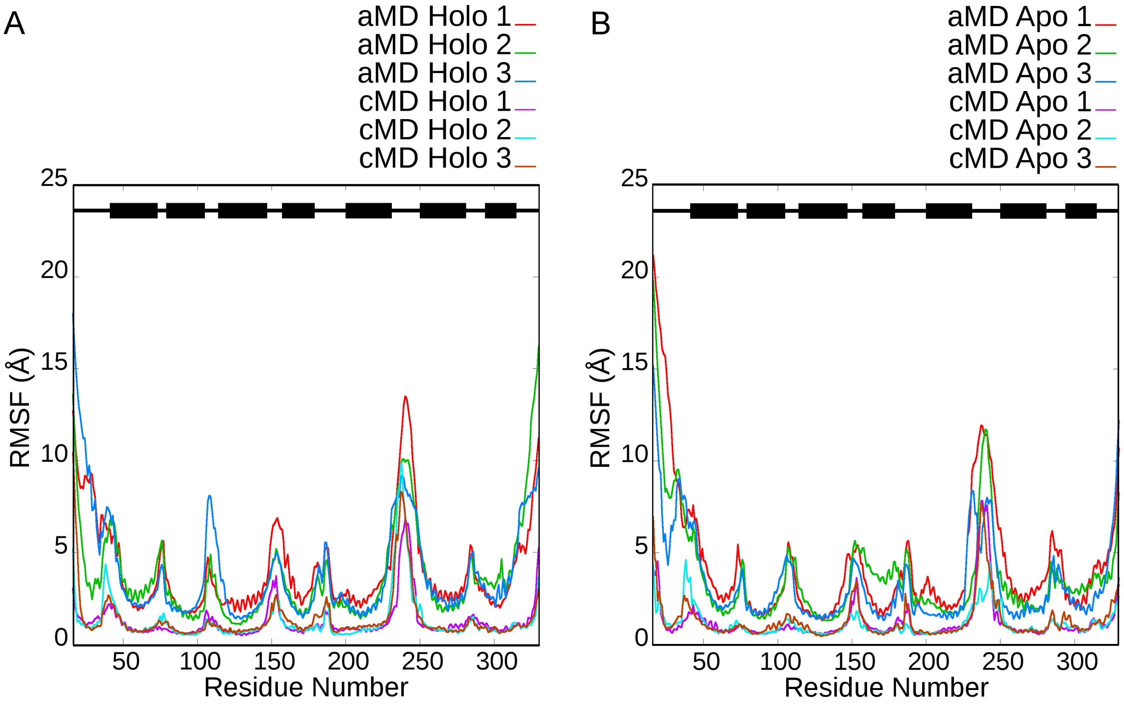

Figure 4.2: Root-Mean-Square Fluctuation of the Antagonist-Bound (A) and Apo (B) Simulations. The cMD simulations are shown in turquoise, purple, and burgundy and the aMD simulations are shown in red, green, and blue.

Due to inherent flexibility of the loop regions, we performed the principal component analysis (PCA) using backbone atoms of the TM helices. We generated the principal component (PC) space by combining the trajectories from all the cMD and aMD simulations with a sampling frequency of 20 frames for a total of 134,746 data points. The individual simulations were then projected onto this PC space (Figure 4.3). Notably, the aMD simulations sampled significantly larger conformational space than the cMD simulations. While the three cMD simulations of the apo and antagonist-bound receptor forms were restricted to locally clustered regions, the aMD simulations sampled unique and widely spread regions (Figure 4.3A and Figure 4.3B).

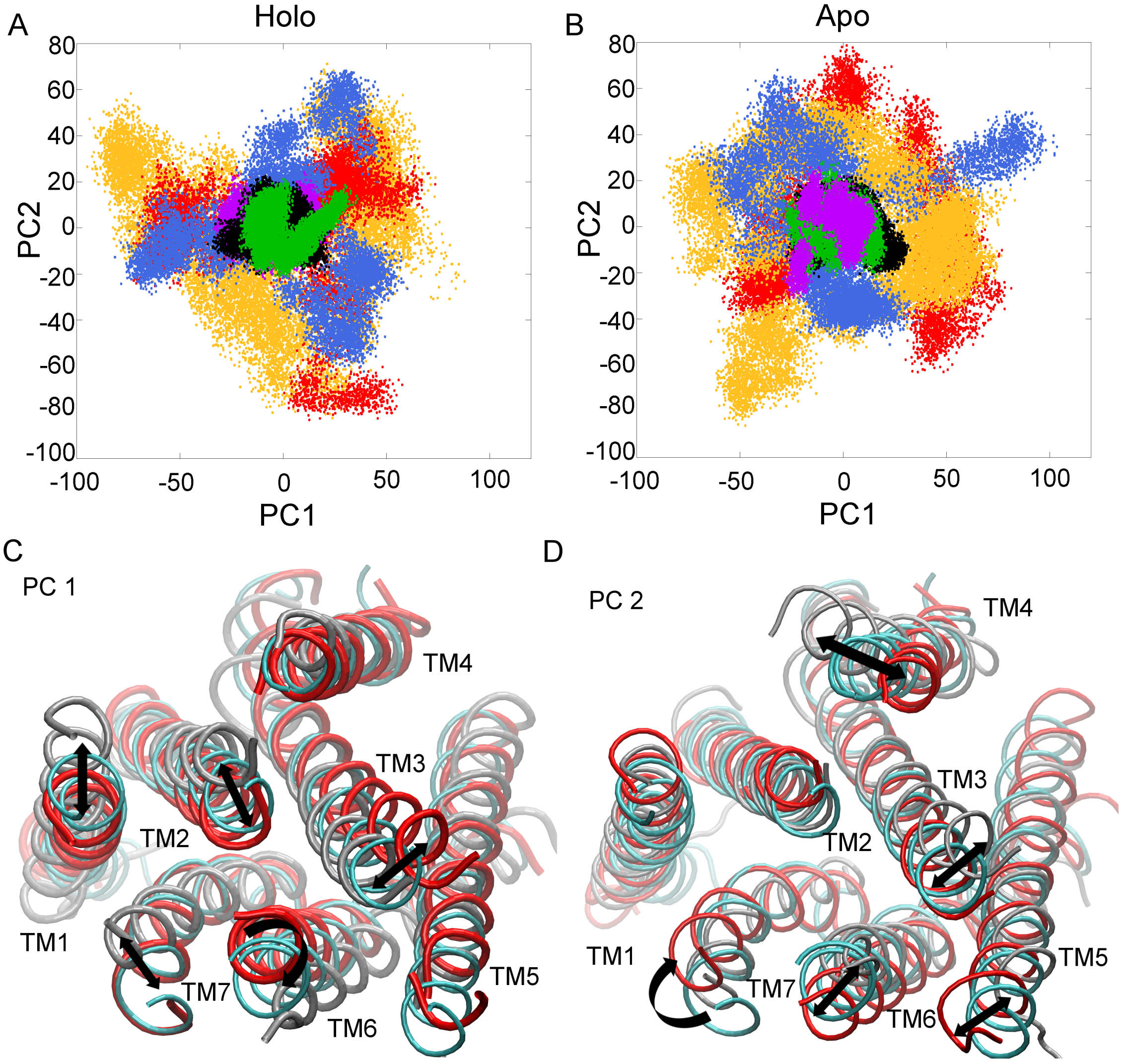

Figure 4.3: Principal Component Analysis (PCA) of the Antagonist-Bound [A] and Apo [B] Molecular Dynamics Simulations of the $S1PR_1$. The subsequent representations of principal component 1 (PC1) (C) and 2 (PC2) (D) are shown. The triplicate conventional molecular dynamics (cMD) simulations for both antagonist-bound (A) and apo (B) are shown in purple, green, and black. The triplicate accelerated molecular dynamics (aMD) simulations for both antagonist-bound (A) and apo (B) are shown in gold, red, and blue. The crystal structure (cyan) and two extreme representations (gray and red) of PC1 (C) and PC2 (D) are shown. Black arrows represent the main movements for each PC.

Relative to the crystal structure, projection of aMD simulation 1 and 2 of the apo receptor along PC1 corresponded to rotation of the TM7 helix and intracellular titling of the TM1, TM2, TM3, and TM6 helices (Figure 4.3C). Moreover, PC2 accounted for extracellular tilting of the TM1 helix (largely due to high fluctuations of that region), the intracellular tilting of the TM1, TM3, TM4, TM5, and TM6 helices, and rotation of the TM7 helix (Figure 4.3D).

Activation of the S1PR$_1$ receptor

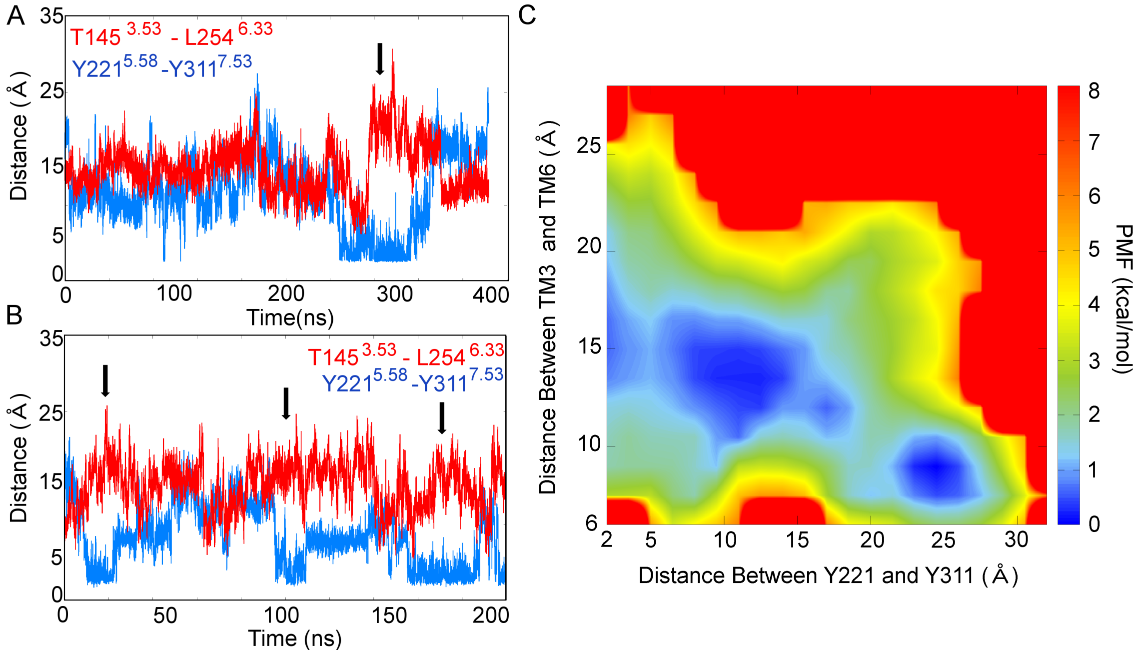

Figure 4.4: Activation during aMD Simulations of the Apo S1PR$_1$. The distances between the hydroxyl oxygens of Y311$^{7.53}$- Y221$^{5.58}$ are shown in blue and between the intracellular ends of TM3 - TM6 (Cα atoms of T145$^{3.53}$ and L254$^{6.33}$) are shown in red for aMD apo 1 (A) and aMD apo 2 (B). The arrows represent the initiation of activation. The potential of mean force (C) is calculated for all aMD simulations of the apo receptor, using the distance between Y311$^{7.53}$- Y221$^{5.58}$ and the distance between the Cα atoms of T145$^{3.53}$ and L254$^{6.33}$ as reaction coordinates.

Because the S1PR$_1$ receptor exhibits high levels of basal activity experimentally [102], we examined whether the apo receptor became active during our simulations. Previous studies of the apo M$_2$ receptor have demonstrated that enhanced sampling methods can accurately predict the activation of a GPCR [51]. In this case, we defined activation as close interaction between Y311$^{7.53}$ of the NPxxY and Y221$^{5.58}$, and an increased distance between the TM3 and TM6 intracellular ends (Figure 4.4).

Using the two reaction coordinates, we calculated the potential of mean force (PMF) for the aMD simulations of the apo receptor (Figure 4.4C) and the aMD simulations of the antagonist-bound receptor, to determine the qualitative transition states between the inactive and active receptors (Figure 4.5). Three low-energy wells were identified from the aMD simulations of the apo receptor, corresponding to the active, intermediate, and inactive states of the receptor. The inactive well exhibited an average TM3-TM6 distance of 7 Å and an Y311$^{7.53}$ - Y221$^{5.58}$ distance of 25 Å. The intermediate well was large and had an average tyrosine distance of 12 Å and an average TM distance of 14 Å. The active well had an average TM distance of 14 Å and a tyrosine distance of 3 Å (Figure 4.4C).

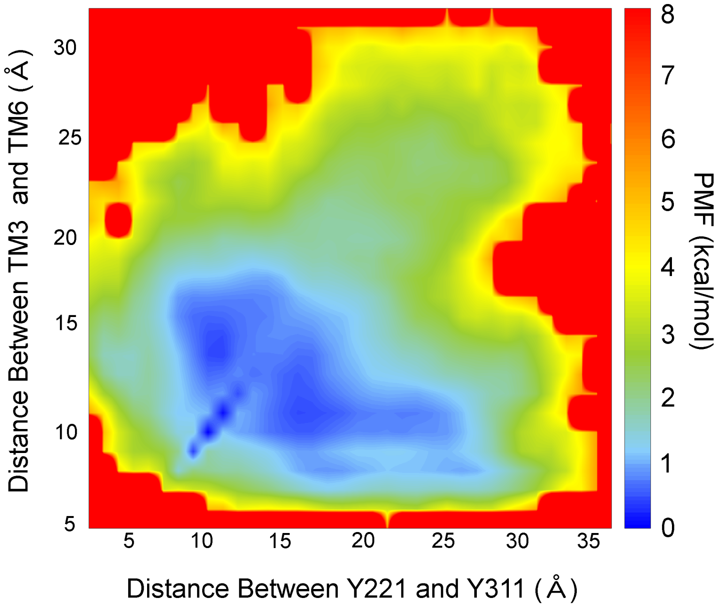

Figure 4.5: Potential of Mean Force for the Antagonist-Bound Simulations. The potential of mean force is calculated of all aMD simulations for the antagonist-bound receptor, using the distances between Y311$^{7.53}$- Y221$^{5.58}$ and the distance between the intracellular ends of TM3 - TM6 as reaction coordinates.

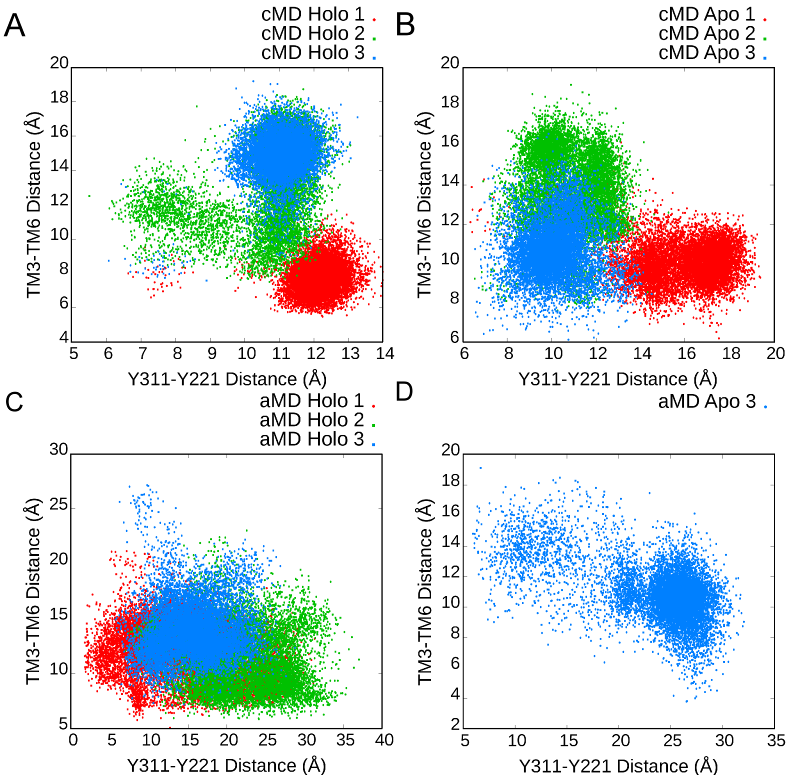

Figure 4.6: Distance Between Y311$^{7.53}$-Y221$^{5.58}$ and the Intracellular Ends of TM3 and TM6. The distance between Y311$^{7.53}$-Y221$^{5.58}$ and the intracellular ends of TM3 and TM6 for the cMD antagonist-bound (A), cMD apo (B), cMD antagonist-bound (C), and aMD apo 3 (D) simulations are plotted.

One large low-energy well was recovered from the aMD simulations of the antagonist-bound receptor. This well, in part, overlapped with the intermediate well of the aMD simulations of the apo receptor, but showed a wider range of the TM3-TM6 distance from 8 Å to 17 Å. The distance between Y311$^{7.53}$ and Y221$^{5.58}$ ranged from 7 Å to 28 Å (Figure 4.5). This low-energy region also overlapped with low energy wells in the cMD simulations of the antagonist-bound and apo receptors (Figure 4.5, Figure 4.6A, and Figure 4.6B). Ultimately, a close interaction between Y311$^{7.53}$ and Y221$^{5.58}$ indicated activation of the receptor.

Correlated Movements During Activation

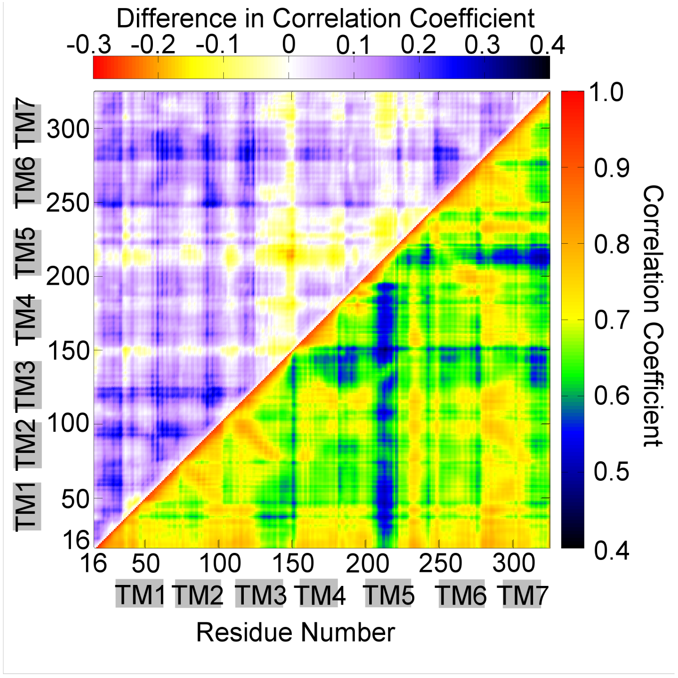

Figure 4.7: Cross Correlation of the Active Trajectories. The top triangle is the difference between correlation of the active trajectories compared to the cMD antagonist-bound control. The lower triangle is the cross correlation values of the active trajectories.

Next, we performed a cross correlation analysis using the simulation trajectories that captured the S1PR$_1$ activation (Figure 4.7, Lower Triangle). Additionally, we calculated the difference between the correlations of the receptor activation trajectories and the control (cMD simulation of the antagonist-bound receptor) to isolate the correlations that were specific to activation (Figure 4.7, Upper Triangle). This analysis determined which regions or specific residues of the receptor were most or least correlated during receptor activation. In general, TM3, TM5, TM6 and TM7 are the transmembrane regions that are mostly associated with GPCR activation, but this study showed that their movements are not necessarily correlated with each other. The motions of TM3 during activation were associated with the intracellular and extracellular ends of TM6, and the extracellular and central ends of

TM7, but not with TM5. The middle region of TM5 had a low correlation with each region of the receptor, except TM6 (Figure 4.7, Lower Triangle). This was also confirmed in the correlation difference, which showed a negative difference compared to the control (Figure 4.7, Upper Triangle). In comparison, TM5 was correlated with TM6 in the aMD simulations (Figure 4.8, Lower Triangle). The TM regions exhibited low correlations in the cMD simulations (Figure 4.8, Upper Triangle), compared to those in the aMD simulations (Figure 4.8, Lower Triangle) and the active trajectories (Figure 4.7, Lower Triangle).

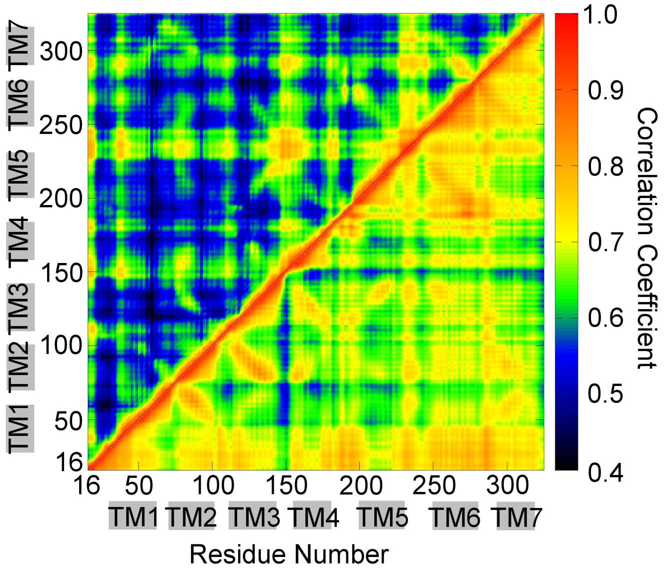

Figure 4.8: Cross Correlation of the aMD and cMD Trajectories. The top triangle is the cross correlation of the cMD simulations. The lower triangle is the cross correlation values of the aMD simulations.

During activation, strong hydrophobic interactions were formed between TM3 and TM2, and TM2 and TM1, as shown with previous study [97, 98]. The correlations between these respective TM regions increased during activation (Figure 4.7, Upper

Triangle). The N-terminus and TM2 were correlated with each region of the receptor, except TM5 and ICL2, suggesting that the N-terminus has a role in activation.

Motions of the Toggle Switch, W269$^{6.48}$

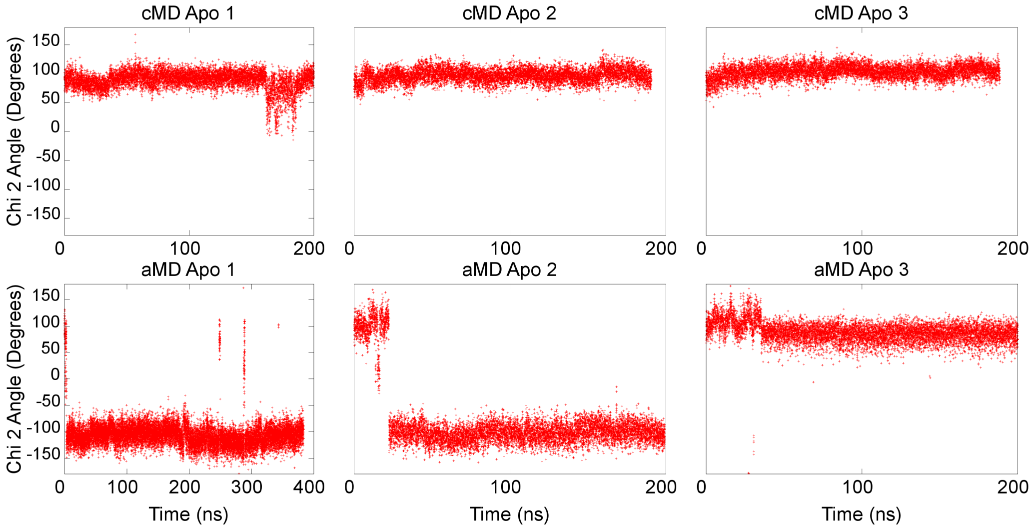

Figure 4.9: χ_2 Dihedral Angle of W269$^{6.48.}$ The χ_2 dihedral angle is shown for the six apo simulations.

According to the cross correlation analysis, W269$^{6.48}$ was not explicitly associated with activation. Previous studies, however, showed that the χ_2 dihedral angle represents a rotameric switch that occurs during activation [96, 98, 101, 106]. In all of the cMD apo simulations, during which the receptor adopts the intermediate conformation, W269$^{6.48}$ maintained a gauche (-) dihedral conformation ($0 \leq \chi_2 \leq 120$). This was also true for the third aMD apo simulation (Figure 4.9). Thus, activation did not occur in these simulations. In the first and second aMD apo simulations, the χ_2 dihedral switched from the gauche (-) to the trans ($-180 \leq \chi_2 \leq -120$) conformation, and maintained these conformations for the remainder of the simulations (Figure 4.9). This suggests that W269$^{6.48}$ must be in the trans position for activation to occur (Figure 4.9).

The motions of W269$^{6.48}$ were highly correlated with M124$^{3.32}$, R120$^{3.28}$, and E121$^{3.29}$, and the correlations increased during activation (Figure 4.7). This suggests that even in the absence of a ligand, the binding pocket plays a central role in the activation process.

Water Channel

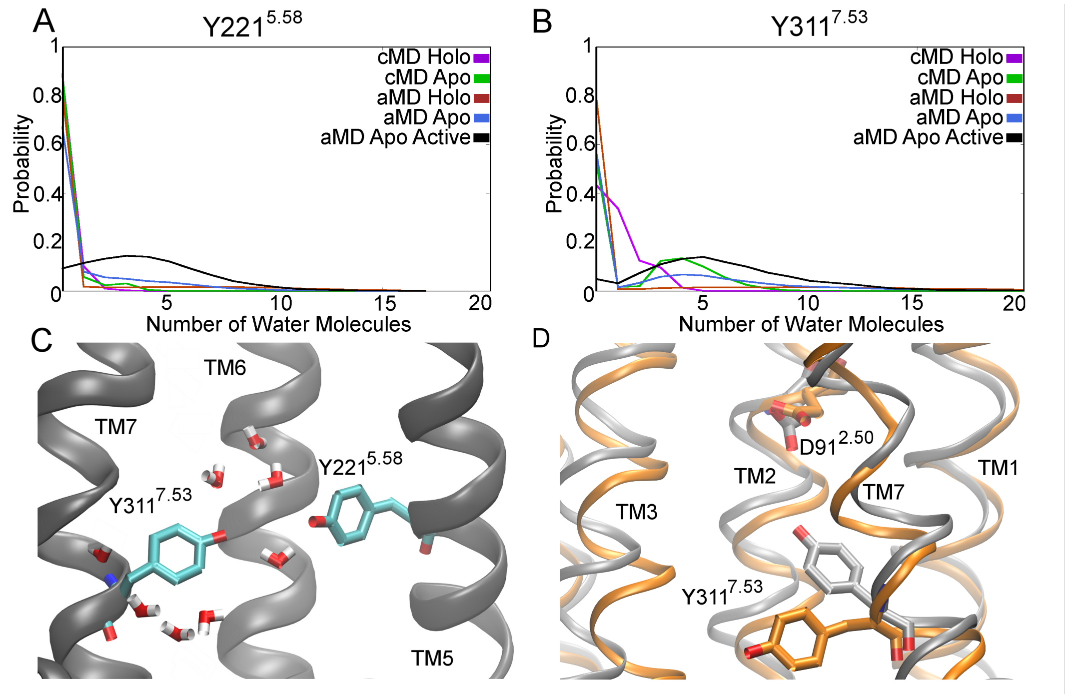

Figure 4.10: Number of Water Molecules in the Water Channel During the Activation Process. The probabilities that zero to twenty waters molecules are present near Y221$^{5.58}$ (A), and Y311$^{7.53}$ (B) are shown for the cMD antagonist-bound (purple), cMD apo (green), aMD antagonist-bound (red), aMD apo (blue), and aMD apo active (black) trajectories. A representative snapshot from the cMD apo simulations is shown (C). The conformational difference of Y311$^{7.53}$ between apo and antagonist-bound simulations for both D91$^{2.50}$ and Y311$^{7.53}$ are shown as bonds (D), with the antagonist-bound representative structure shown in gray, and the apo representative shown in orange (D).

Water dynamics have been shown to be an important aspect of GPCR activation [99-101, 107]. In this study, we calculated both the total number of water molecules surrounding Y221$^{5.58}$ and Y311$^{7.53}$ (Figure 4.10), and the hydrogen bond interactions between these water molecules and the protein residues (Figure 4.11), for the cMD apo, aMD apo, cMD holo, aMD holo, and active trajectories separately. The active trajectories are defined as 239 ns – 330 ns in the aMD simulation 1 of the apo receptor and 8 ns – 26 ns, 94 ns – 112 ns, and 144 ns – 190 ns in the aMD simulation 2 of the apo receptor.

The number of water molecules that entered the intracellular pocket of the S1PR$_1$ was dependent on the receptor activation. While the number of water molecules surrounding Y221$^{5.58}$ peaked at zero for each of the receptor simulations, four water molecules were found in this pocket on average when the receptor adopted active conformation in aMD simulation of the apo receptor (Figure 4.10A). Moreover, the number of water molecules surrounding Y311$^{7.53}$ peaked at zero for all of the simulations (Figure 4.10B). However, the cMD and aMD simulations of the apo receptor had a second peak of four water molecules in this intracellular pocket surrounding Y311$^{7.53}$. When the apo receptor adopted the active conformation in the aMD simulations, it exhibited a low probability (0.04) of having zero water molecules in the Y311$^{7.53}$ pocket, and had an average of five water molecules surrounding this residue, suggesting that more water molecules entered this intracellular region during receptor activation.

The residency time of a water molecule near a particular region or pocket of the receptor can determine if water has an allosteric effect on the receptor [108]. The residency time of each water molecule near Y221$^{5.58}$ and Y311$^{7.53}$ was dependent upon the receptor conformational states (Table 4.2 and Table 4.3). The average residency time of water molecules surrounding Y211$^{5.58}$ was 0.16 ns in the active, 0.15 ns in the intermediate, and 0.11 ns of the inactive trajectories. Although the average residency time is higher in the active trajectories, water molecules with long residency times (≥ 10 ns) were found in the intermediate trajectories (Table 4.2). The average residency time of water molecules surround Y311$^{7.53}$ was 0.17 ns in the active, 0.20 ns in the intermediate, and 0.13 ns in the inactive, and the water molecules with long residency times were found in the intermediate and inactive trajectories (Table 4.3). Water appears more dynamic in the receptor active conformer than in the intermediate and inactive receptor conformers, suggesting that water may stabilize the receptor in the intermediate and inactive conformations.

Table 4.2: Residency Times of Water Molecules near Y221$^{5.58}$ in the Active, Intermediate and Inactive conformers of the S1PR$_1$. The average overall residency times, and standard deviation times are shown. The residency times of the top 5 longest water molecules are listed.

	Active	Intermediate	Inactive
Average	0.16 ns	0.15 ns	0.11 ns
Standard Deviation	0.23 ns	0.48 ns	0.13 ns
Top 5	2.74 ns	35.54 ns	1.28 ns
	2.56 ns	27.96 ns	0.96 ns
	2.36 ns	19.2 ns	0.8 ns
	2.08 ns	8.2 ns	0.7 ns
	2.06 ns	4.98 ns	0.68 ns

Table 4.3: Residency Times of Water Molecules near Y311$^{7.53}$ in the Active, Intermediate and Inactive conformers of the S1PR$_1$. The average overall residency times, and standard deviation times are shown. The residency times of the top 5 longest water molecules are listed.

	Active	Intermediate	Inactive
Average	0.17 ns	0.20 ns	0.13 ns
Standard Deviation	0.29 ns	1.55 ns	0.30 ns
Top 5	4.52 ns	104.8 ns	3.62 ns
	4.2 ns	52.86 ns	3.26 ns
	4.16 ns	42.48 ns	3.1 ns
	3.74 ns	41.29 ns	2.98 ns
	3.46 ns	38.64 ns	2.9 ns

Further analysis showed that in the apo receptor, Y311$^{7.53}$ reoriented its side chain towards TM5 and thus can be solvated from the intracellular side (Figure 4.10C). In contrast, in the antagonist-bound receptor, the hydroxyl oxygen of Y311$^{7.53}$ formed a hydrogen bond with the side chain of D91$^{2.50}$, which buried Y311$^{7.53}$ from intracellular solvation (Figure 4.10D). Therefore, Y221$^{5.58}$ and Y311$^{7.53}$ became more solvated with significant structural rearrangements during the receptor activation.

While more water molecules were found surrounding Y221$^{5.58}$ and Y311$^{7.53}$ upon activation of the apo receptor, hydrogen bond formation between these residues and the water molecules was not closely correlated with receptor activation (Figure 4.11). The probability that a water molecule did not form a hydrogen bond with Y221$^{5.58}$ was greater than ~80% for all the simulations (Figure 4.11A). Y311$^{7.53}$ did not form any hydrogen bonds with water molecule in most simulations, except the cMD simulations of the apo receptor, where the receptor formed one hydrogen bond with a water molecule (Figure 4.11B). Therefore, although an intracellular water channel opened up during activation of the apo receptor, Y221$^{5.58}$ and Y311$^{7.53}$ did not form stable hydrogen bonds with water molecules, largely due to highly dynamic fluctuations of the water molecules (Table 4.2 and Table 4.3).

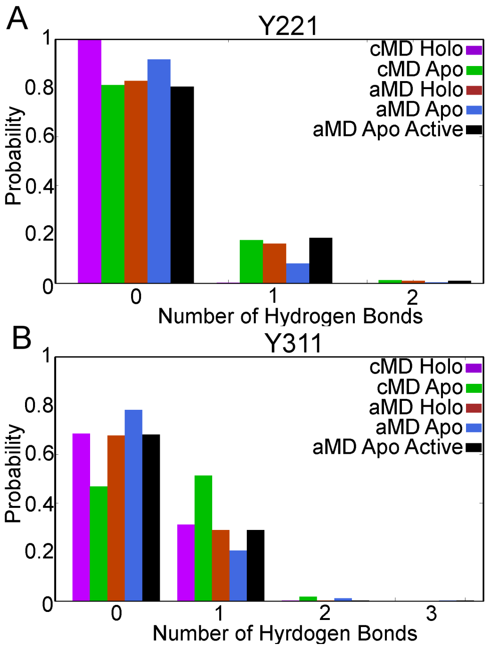

Figure 4.11: Hydrogen Bonds in the Water Channel During the Activation Process. The probability that water molecules form zero to five hydrogen bonds with Y221$^{5.58}$ (A) and Y311$^{7.53}$ (B) are shown for the cMD antagonist-bound (purple), cMD apo (green), aMD antagonist-bound (red), aMD apo (blue), and aMD apo active (black) simulations.

Lipid Entry

Surprisingly, during every simulation, except the first aMD simulation of the antagonist-bound receptor, a palmitoyl–oleoyl-phosphatidyl-choline POPC lipid molecule was observed to interact closely with the extracellular ends of TM1 and TM7 of the receptor and, at certain points to enter the receptor (Figure 4.12 and Figure 4.13). This region is the proposed ligand binding site for S1PR$_1$ [94]. The distance between the extracellular end of TM1 and TM7 is 13.5 Å in the crystal structure, but this distance increased during the simulations (Figure 4.13). The distance ranged from 12 Å to 22 Å in the cMD simulations of the antagonist-bound receptor (Figure 4.13A), 14 Å to 22 Å in the cMD simulations of the apo receptor (Figure 4.13B), 6 Å to 32 Å in the aMD simulations of the antagonist-bound receptor (Figure 4.13C), and 7 Å to 25 Å in the aMD simulations of the apo receptor (Figure 4.13D). The increased distance between TM1-TM7 led to opening of the receptor, which allowed entry of the POPC molecule (Figure 4.12A).

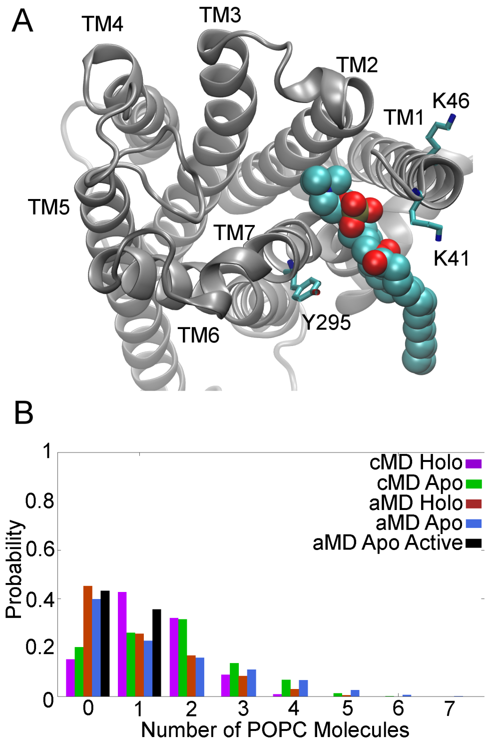

Figure 4.12: Number of POPC Molecules Interacting with the Receptor. A POPC molecule enters the receptor between TM1 and TM7 (A). The total number of POPC molecules that interact with this region at a given time is shown for the cMD antagonist-bound (purple), cMD apo (green), aMD antagonist-bound (red), aMD apo (blue), and aMD active (black) trajectories.

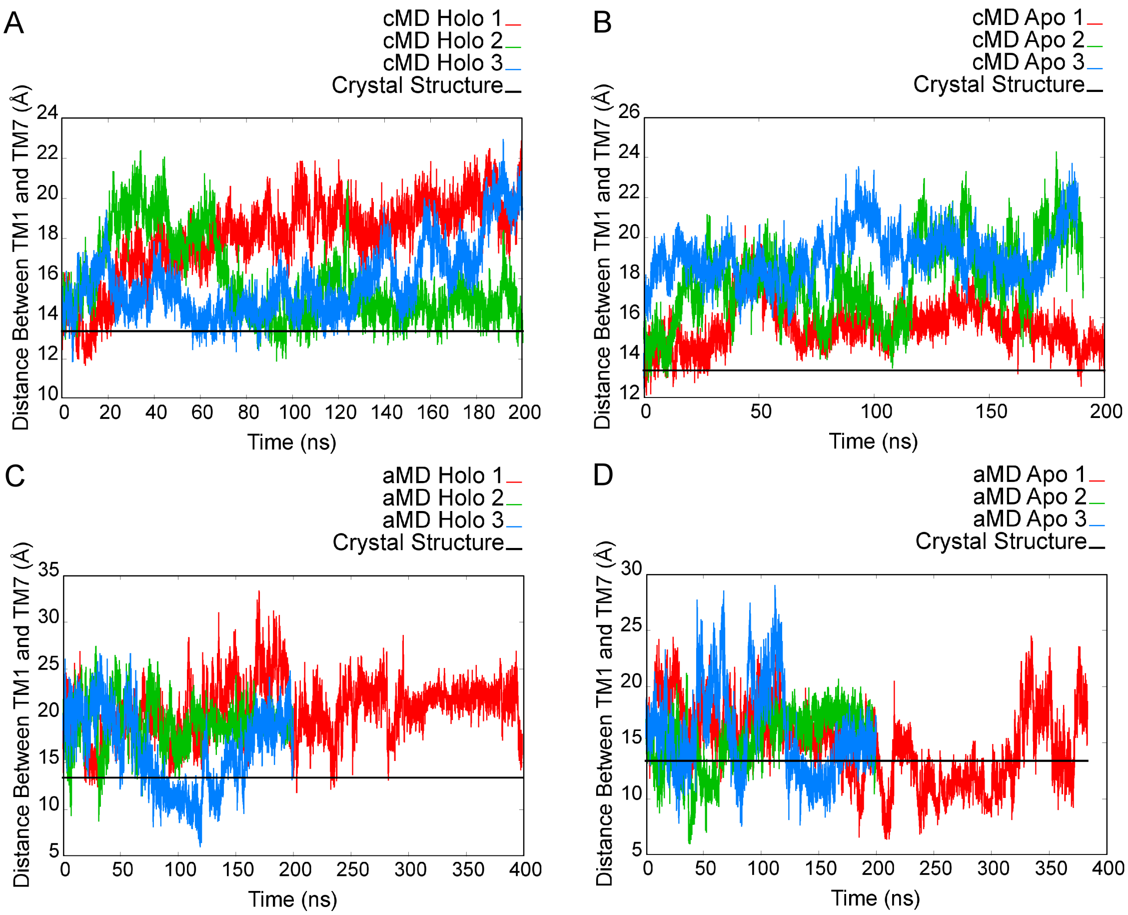

Figure 4.13: Distance Between the Extracellular Ends of TM1 and TM7. The distance is shown for the cMD antagonist-bound (A), cMD apo (B), aMD antagonist-bound (C), and aMD apo (D). The distance of the crystal structure is shown as a black line at 13.5 Å.

In the cMD simulations of the apo and antagonist-bound receptor, there was likely one or more POPC molecules in the receptor (Figure 4.12B). The POPC molecules interacted with the receptor with direct hydrogen bonds (Figure 4.14). In the aMD simulations of the apo and antagonist-bound receptor, there were fewer POPC molecules found in this region (Figure 4.12B and Figure 4.14). The number of hydrogen bonds, particularly formed between POPC and residues K46$^{1.33}$, K41$^{\text{N-TERM}}$ and Y295$^{7.37}$, peaked at one or two (Figure 4.12A and Figure 4.14). In the active state, a maximum of one POPC molecule entered the site. Comparatively, in the aMD simulations of the apo receptor, a maximum of five POPC molecules were seen interacting with this region. This suggested that activation of S1PR$_1$ reduces the affinity for lipid molecules in this region of the receptor.

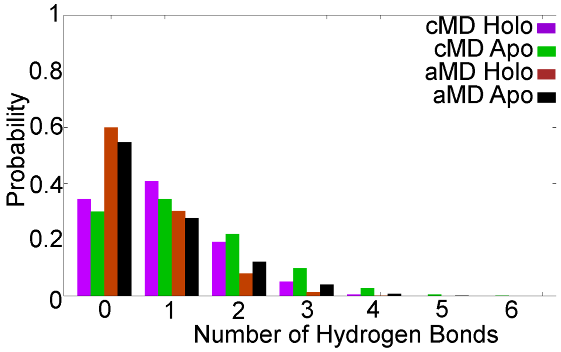

Figure 4.14: Hydrogen Bond Interactions with POPC Molecules. A POPC molecule forms zero to six hydrogen bonds with K46$^{1.33}$, K41$^{\text{N-TERM}}$, and Y295$^{7.37}$ for the cMD antagonist-bound (purple), cMD apo (green), aMD antagonist-bound (red), and aMD apo (black).

Comparison to Other Known Structures of Active GPCRs

Next, we decided to compare our simulation-derived active S1PR$_1$ with known X-ray structures of active GPCRs. We aligned the backbone atoms of the TM regions using active β$_2$AR in complex with the heterotrimeric G protein as the reference structure [42]. We aligned active structures of rhodopsin [43], the M$_2$ muscarinic receptor [46], the A$_{2A}$AR [24], μ-opioid [95], and a representative aMD snapshot of the active S1PR$_1$. We calculated the RMSDs of each of the receptors compared to the active β$_2$AR. The structure most similar to β$_2$AR was the active M$_2$ muscarinic receptor with and RMSD of 1.36 Å compared to the β$_2$AR [44]. The active M$_2$ muscarinic receptor has a RMSD of 1.36 Å compared to the active β$_2$AR. The active μ-opioid receptor has a RMSD of 1.88 Å compared to the active β_2AR. The active opsin has a RMSD 1.75 Å compared to the active β_2AR. The A_{2A}AR had the largest RMSD compared to the β_2AR, at 2.64 Å [24] (Figure 4.15). Our simulation-derived active S1PR$_1$ had an RMSD of 2.64 Å compared to the active β_2AR, being similar to other active GPCRs.

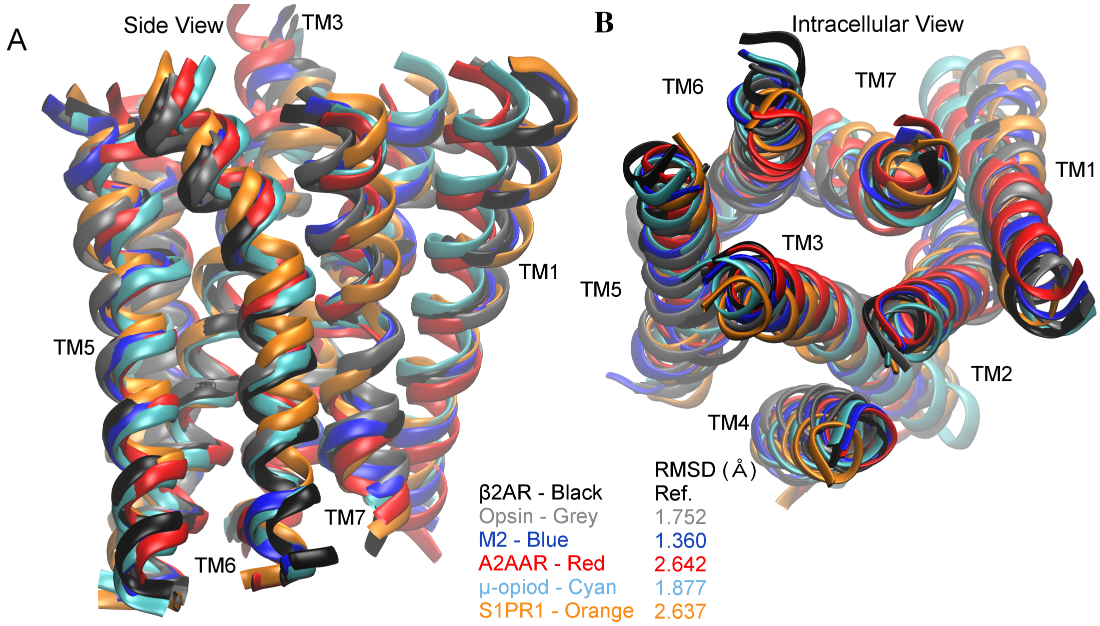

Figure 4.15: Comparison Between Known Active GPCRs. RMSD calculation of the backbone atoms of the TM regions. Opsin (grey), M_2 (blue), A_{2A}AR (red), mu-opioid (cyan), and a representative active S1PR$_1$ (Orange) were aligned against β_2AR (black). The side view is shown in A, and the intracellular view is shown in B. The RMSD values are shown in the caption.

Discussion

Similar to several other GPCRs, S1PR$_1$ exhibits high levels of basal activity [102]. In this study, we characterize activation of the apo (ligand-free) receptor from the aMD enhanced sampling simulations. We identify residues that correlate with activation, explore the water network that occurs in basal activation and examine the interactions between lipid molecules.

A previous study of the apo M_2 receptor demonstrates that aMD can predict the activation of a GPCR, when compared to the active crystal structure [44, 51]. The aMD simulation of the M_2 receptor predicted the relocation of the $Y^{7.53}$ and $Y^{5.58}$ side chains towards each other [51]. The aMD simulation of the M_2 receptor also predicted a flipping of $Y^{7.43}$ and relocation of $W^{6.48}$, which were not seen in the crystal structure [44, 51].

Activation in GPCRs is characterized by a structural rearrangement of TM5, TM6 and TM7 helices [42-44, 95, 96]. During GPCR activation, the DRY motif is rearranged and the cytoplasmic end of TM6 moves away from the TM3. A dihedral switch of conserved $W^{6.48}$ and rearrangement of the NPxxY motif also occur [42-44, 95-97]. The rearrangement of the NPxxY motif leads to close interaction between $Y^{5.58}$ and $Y^{7.53}$, either through a direct or water-bridged hydrogen bond, and a separation of TM3 and TM6 compared to the inactive structures.

In this study, the activation of the $S1PR_1$ occurs four separate times, once during the first aMD simulation of the apo receptor and three times during the second aMD simulation of the apo receptor. Unlike previous MD studies of the $S1PR_1$ [98, 101], the metric for activation in this current study is defined as a small distance (< 5 Å) between $Y311^{7.53}$ and $Y221^{5.58}$, (rather than the RMSD of the NPxxY motif) and a large distance (>12 Å) between the intracellular ends of TM3 and TM6. $Y221^{5.58}$ and $Y311^{7.53}$ interact through a direct hydrogen bond or through water molecules in several active GPCRs, including β_2AR, with a distance of 4.28 Å [42], rhodopsin with a distance of 5.43 Å [43], the M_2 muscarinic receptor, with a distance of 4.21 Å [44] and the μ-opioid receptor, with a distance of 4.57 Å [95]. The distance between the TM3 and TM6 varies for GPCRs. This distance is 14 Å in the active β_2AR [42, 47], 7 Å in the active rhodopsin [43, 45], 10.4 Å in the active M_2 receptor [44, 46], and 14.23 Å in the active μ-opioid receptor [42, 47]. This distance reaches 25 Å in the present study, although previous studies of the dynamics of the $S1PR_1$ suggest that the TM3-TM6 distance is flexible and more indicative of an intermediate stage rather than a fully active receptor [98].

In the intermediate state, the most flexible state of the receptor, the distance between the $Y221^{5.58}$ and $Y311^{7.53}$ ranged from 7 Å to 18 Å, and the distance between the TM3 and TM6 range of 10 Å to 17 Å. Large differences between the intermediate and active conformations include the χ_2 position of $W269^{6.48}$, and the water residency times. A key step in the activation of $S1PR_1$ is the χ_2 switch of $W269^{6.48}$ from the gauche (-) conformation to the trans position. In the intermediate well, which includes trajectories from the cMD apo simulations, this switch does not occur, suggesting that although those regions are flexible, this switch must occur before the receptor can exhibit basal activation. Additionally, the average water residency time of waters near $Y221^{5.58}$ fall directly between the active and the inactive. The average water residency time near $Y311^{7.53}$ is higher in the intermediate state than in the inactive and active conformations. Moreover, there are more residues with long residency times in this state, with the water molecules with the longest residency times being 35.54 ns in residue $Y221^{5.58}$, and 104.8 ns in residue $Y311^{7.53}$. Water molecules can stabilize a protein or facilitate the conformational transitions of a protein [108]. For the $S1PR_1$, the water molecules near $Y311^{7.53}$ and $Y221^{5.58}$ are possibly stabilizing the receptor in the intermediate conformation. Overall, the intermediate step of activation of the $S1PR_1$ needs to be presented for further studies.

After isolating the active trajectories, we decided to explore the correlation between the residues of the protein. The cross correlation analysis reveals interhelical interactions between TM1 and TM2, and TM2 and TM3 that have been seen in previous studies [96-98]. The cross correlation analysis also reveals the involvement of the residues of the binding site in activation. M124$^{3.32}$, R120$^{3.28}$, and E121$^{3.29}$ exhibit increased correlations with W269$^{6.48}$. W$^{6.48}$ is known as a GPCR toggle switch, which triggers the rearrangement of TM6 [98, 101, 106]. In this study, W269$^{6.48}$ maintained a χ_2 gauche (-) conformation in the three cMD apo simulations and the third aMD apo simulation. Before activation, W269$^{6.48}$ switched from a χ_2 gauche conformation to a χ_2 trans position, which is similar to a previous S1PR$_1$ activation study [98]. In comparison, in the S1PR$_1$-S1P activation study by Yuan et al, W269$^{6.48}$ maintained a χ_2 gauche (-) position in the apo simulation, while the W269$^{6.48}$ in both antagonist and agonist bound simulations fluctuated between a gauche (-) position and a trans position [101].

Several studies have examined the water channel of GPCRs during activation [99-101, 107]. All of the previous S1PR$_1$ activation studies simulated activation with an agonist bound to the receptor. In this study, we explore the water channel of the receptor in the apo form, and find a similar correlation between water dynamics and receptor activation. During activation of the apo receptor, Y311$^{7.53}$ breaks a stable hydrogen bond that is formed with D91$^{2.50}$ in the antagonist-bound receptor and moves towards Y221$^{5.58}$ in the TM5 helix. This leads to opening of the intracellular pocket of the receptor and increases the number of water molecules surrounding Y221$^{5.58}$ and Y311$^{7.53}$.

During 11 out of 12 MD simulations, a zwitterionic POPC molecule interacts with residues surrounding the extracellular end of TM1 and TM7. This is consistent with previous findings. A previous computational study showed that 2-arachidonoylglycerol (2-AG) enters the cannabinoid CB2 receptor between TM6 and TM7, and can partition through a POPC lipid bilayer to do so [109]. Additionally, 11-cis-retinal enters opsin between TM5 and TM6 and exits the receptor between TM1 and TM7 [110]. Lipid molecules interact with S1PR$_1$ regardless of ligand entry. However, the molecules are less likely to interact with this site during the active state of the receptor during the aMD simulations of the apo receptor. This is unlike previous computational studies that showed that lipid molecules in the bilayer can stabilize the receptor in the active state [111]. A computational study showed that both the zwitterionic POPC and anionic palmitoyl–oleoyl-phosphotidyl-glycerol (POPG) entered the [51] β$_2$AR receptor between TM6 and TM7, except during the deactivation process [111].

Based on the findings in this study, apo activation is initiated by a rotameric switch of W269$^{6.48}$ to the trans position. The rotameric switch of W269$^{6.48}$ has been previously shown to be initiated by agonist binding [98, 101], but can stably occur without the presence of a ligand. The switch is also correlated with the movements of M124$^{3.32}$, D279$^{6.58}$, R120$^{3.28}$ and E121$^{3.29}$ of the binding site. Next, Y311$^{7.53}$, which previously interacted with D52$^{2.50}$, relocates towards TM5 and Y221$^{5.58}$. This movement opens up the internal water channel, which leads to an increase in the TM3-TM6 distance, making space for G protein binding.

Methods

The Ballesteros and Weinstein numbering format of GPCRs, X.YY, was applied to label the relative location of specific residues in S1PR$_1$. 'X' denotes the transmembrane helix number and 'YY' represents the location of the residue in the transmembrane helix related to the most conserved residue (denoted 50) [60].

System Setup

The S1PR$_1$ X-ray structure determined at a 2.80 Å resolution was used for the simulations. The receptor was in complex with the antagonist ML056 and fused to a T4 lysozyme in place of ICL3 (PDB:3V2Y) [94]. The T4 lysozyme was removed and the missing residues, including the ICL3, were added using Prime in the Schrodinger package [85]. For the apo state, the ligand was removed to obtain the starting structure.

The system topologies were generated using the *psfgen* plugin in VMD [112]. The CHARMM36 force field parameters were used for the protein and lipid molecules [65]. CHARMM's ParamChem generated the force field parameters for the antagonist, ML056 [113]. The two disulfide bonds, Cys184 - Cys191 and Cys282 – Cys287, that were resolved in the crystal structure were maintained for the simulations. The receptor was embedded into a POPC lipid bilayer. The TIP3P water molecules were added using the *solvate* plugin in VMD [112] to solvate the membrane-protein system. The net charge of the system was neutralized with 0.15 M NaCl. The resulting ML056-bound S1PR$_1$ system had a total of 170 POPC lipid molecules, 32 sodium ions, 39 chloride ions and 5,392 water molecules with a total of 91,503 atoms. The apo system had 169 POPC molecules, 49 sodium ions, 53 chloride ions and 5,410 water molecules with a total of 88,094 atoms. Periodic boundary conditions were applied to both simulation systems.

Molecular Dynamics Simulations

Three independent cMD simulations were initially performed for each starting system, ML056-bound and apo (Table 4.1), using NAMD 2.9 [66]. The cutoff distance for the van der Waals and short range electrostatics was 12 Å, and the particle mesh Ewald method was applied for the calculation of long range electrostatic interactions [69]. A 2 fs time step and a multiple-time-stepping algorithm [66] were used with bonded and short range non bonded interactions computed for each time step, and long range electrostatics every two-time steps. The SHAKE algorithm was applied to the hydrogen-containing bonds [70].

Initially, the lipid tails for each simulation were minimized for 1,000 steps and simulated for 0.5 ns with NVT at 300 K. Next, the protein atoms were relaxed for 0.5 ns with NPT conditions with 2 kcal/(mol*Å^2) harmonic restraints applied. Each system was equilibrated for 0.5 ns with everything released in NPT conditions at 310 K. The production runs for all six simulations were run for 200 ns each in NPT conditions.

Accelerated Molecular Dynamics

Three dual-boost aMD simulations of the antagonist-bound receptor were initiated with different velocity assignments from the final structure of the first cMD antagonist-bound simulation, and three dual-boost simulations were similarly initiated from the final structure of the first cMD simulation of the apo receptor [103, 104]. In aMD, a boost potential, $\Delta V(r)$, is added when the potential energy, $V(r)$, of a system falls below a user-defined threshold, E.

$$\Delta V^*(r) = V(r), \quad V(r) \geq E \tag{1}$$

$$V^*(r) = V(r) + \Delta V(r), \qquad V(r) < E \tag{2}$$

The boost potential is defined as:

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + E - V(r)} \tag{3}$$

where α is the acceleration factor. In dual-boost aMD, a boost is applied to the dihedrals of the system and to the total energy of the system [103, 104]. The energy threshold is determined by [103, 104]:

$$E_{dihedral} = V_{dihed\_avg} + (\lambda \times V_{dihed\_avg}) \tag{4}$$

$$E_{total} = V_{total\_avg} + (0.2\frac{kcal}{mol} atoms^{-1} \times N_{atoms}) \tag{5}$$

The acceleration factor is determined by [103, 104]:

$$\alpha_{dihedral} = \lambda \times \frac{V_{dihed\_avg}}{5} \tag{6}$$

$$\alpha_{total} = 0.2\frac{kcal}{mol} atoms^{-1} \times N_{atoms} \tag{7}$$

For these simulations, λ was 0.3. The first antagonist-bound and apo aMD simulations ran for 384 ns and 335 ns, respectively, while the other two antagonist-bound and apo simulations ran for 200 ns each (Table 4.1).

Analysis

Unless otherwise stated, all analysis on the cMD holo, aMD holo, cMD apo, aMD holo, and aMD active simulations were calculated with a sampling frequency every 20 frames, or 0.2 ns.

Potential of Mean Force

The PMF was calculated from the aMD simulations, which enables us to determine the low-energy conformations of the receptor. This analysis is qualitative without energetic reweighting and thus not quantitative. The aMD simulations of the antagonist bound and apo receptors were analyzed separately. Every frame was sampled for a total of 392,564 data points in the antagonist-bound and 393,931 data points in the apo simulations. The reaction coordinates, distances between the intracellular ends of TM3 and TM6 and distance between Y311$^{7.53}$ and Y221$^{5.58}$, were chosen to characterize activation of the receptor. The PMF was calculated with the following equation [74]:

$$A(\xi_J, \xi_I) = -k_B T \ln(\rho(\xi_J, \xi_I)) \tag{8}$$

where ξ_J and ξ_I are the reaction coordinates, k_B is the Boltzmann constant, T is the temperature, and ρ is the probability distribution. The grid size for this analysis was 1 by 1 Å (Figure 4.4C and Figure 4.5). If there were no data points in the certain grids, the corresponding PMF was considered infinity. To verify convergence, we calculated the PMF with a grid size of 3 Å by 3 Å (Figure 4.16). This plot showed the same general pattern, seen in the original PMF (Figure 4.4C).

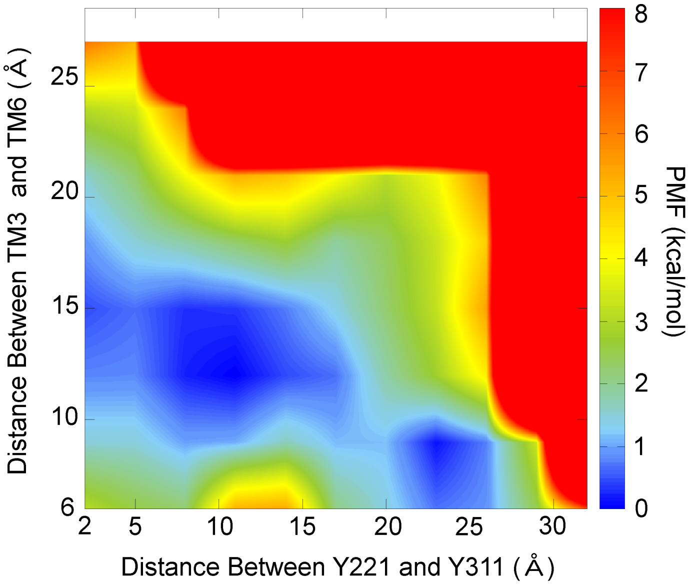

Figure 4.16: Convergence of the Apo Potential of Mean Force. The potential of mean force is calculated of all aMD simulations for the ligand-free receptor, using the distances between $Y311^{7.53}$- $Y221^{5.58}$ and the distance between the intracellular ends of TM3 - TM6 as reaction coordinates, and a box size of 3 Å by 3 Å.

Hydrogen Bonds

Hydrogen bonds between the protein and water molecules or the protein and POPC molecules were calculated using the VMD *HBond* plugin [112]. An interaction was considered a hydrogen bond if the donor-acceptor distance was less than 3 Å with an angle cutoff of 20°.

Generalized Cross Correlation

Generalized cross correlation [114] values were calculated using the *g_correlation* module in GROMACS [115]. Generalized cross correlation can identify linear and nonlinear correlated motions of protein residues. Only the Cα atoms in protein residues were included in the analysis.

The difference in cross correlation was calculated by subtracting the correlation value of the cMD antagonist-bound simulations at a sampling frequency of 20, from the values during the activation events, defined as 239 ns – 330 ns in the aMD simulation 1 of the apo receptor and 8 ns – 26 ns, 94 ns – 112 ns, and 144 ns – 190 ns in the aMD simulation 2 of the apo receptor.

Water Residency

The water residency times for each water molecule within 5 Å of Y221$^{5.58}$ and Y311$^{7.53}$ were calculated with the following equation:

$$Res_i = \frac{1}{50}\sum_{n=1}^{N} T_{n,i} \,, \qquad T_{n,i} \geq 0.04\, ns \tag{9}$$

where Res_i is the residency time in *ns* for each water molecule *i*, *N* is the total number of frames, $T_{n,i}$ is 0.02 ns for each consecutive frame (above 0.04 ns) where water molecule *i* is within 5 Å of the residues, and 50 is the number of frames/ns. Each water residency time for a water molecule is counted. The average residency time is calculated with:

$$Res_{avg} = \frac{1}{N_w}\sum_{i=0}^{N_w} Res_i \tag{10}$$

where Res_{avg} is the average residency time, N_w is the total number of water molecules, and Res_i is the residency time in ns for each water molecule, *i*.

Acknowledgements

This work was supported in part by the NSF, NIH, HHMI, NBCR and the NSF Supercomputing Centers.

Chapter 4 is a minimally modified reprint of the material as it appears in Alisha D. Caliman, Yinglong Miao, and J. Andrew McCammon, “Activation Mechanisms of the First Sphingosine-1-Phosphate Receptor,” Protein Science, 2017. The dissertation author was the primary investigator and author of this paper.

CHAPTER 5

The Identification of Non-Orthosteric Sites and Potential Modulators for the First Sphingosine-1-Phosphate Receptor

Abstract

The first sphingosine-1-phosphate receptor (S1PR$_1$) is a G protein-coupled receptor (GPCR) that plays a key role in lymphocyte circulation, neuronal myelination, and cardiac inflammation. Recently, we applied conventional and accelerated molecular dynamics to investigate the structural dynamics of the S1PR$_1$. In the accelerated molecular dynamics simulations, we observed the activation of the S1PR$_1$ as characterized by the rotameric switch of residue W269$^{6.48}$ and an intracellular rearrangement of transmembrane helix (TM) 5 and TM7, which led to a close interaction between residues Y221$^{5.58}$ and Y311$^{7.53}$. In this study, we identify eight non-orthosteric sites on the S1PR$_1$ by applying a fragment-mapping algorithm, FTMap, to representative structures of the receptor. Non-orthosteric site 1 is located between TM2, TM3, TM6, and TM7. Non-orthosteric site 2 is located between the extracellular ends of TM1 and TM7, and non-orthosteric site 3 is located on the intracellular end of TM1, TM2, and TM4. The orthosteric site and the three previously described non-orthosteric sites were screened with two compound libraries, the NCI Diversity Set IV and the Chembridge GPCR targeted library. The orthosteric site docked favorably with 21 compounds. 18 of those 21 compounds share the same chemical core, which interacts with residues F125$^{3.33}$, L128$^{3.36}$, W269$^{6.48}$, L195^{ECL2}, and M124$^{3.32}$. We also

identify six top-ranked compounds for each non-orthosteric site. We recommend that these compounds be experimentally tested for activity on the $S1PR_1$.

Introduction

The first sphingosine-1-phosphate receptor ($S1PR_1$) is a G protein-coupled receptor (GPCR) that binds to the zwitterion, S1P [20, 90, 92]. The $S1PR_1$ is ubiquitously expressed in human tissues, and the S1P/$S1PR_1$ axis plays a key role in T cell and B cell migration [20, 21, 116, 117], demyelination of neuronal cells [118], cardiac inflammation [119], and estrogen receptor positive (ER+) breast cancer [120].

Ligands that act as functional antagonists of the $S1PR_1$ can treat autoimmune diseases, such as sepsis [121] and multiple sclerosis (MS) [91, 117, 118, 122], and heart disease [119]. Activation of the $S1PR_1$ promotes lymphocyte circulation, and antagonism of the S1P gradient [92] or the ability of the $S1PR_1$ to recognize the gradient leads to the sequestration of lymphocytes [117]. The $S1PR_1$ is located on the surface of newly generated T cells and B cells. These lymphocytes follow the gradient of the ligand, S1P, from low in the secondary lymphoid tissues to high in the blood, thus facilitating lymphocyte egress [20, 21, 116]. In neuronal cells, modulation of the $S1PR_1$ by a functional antagonist prevents demyelination of neurons by inhibiting the release of chemokines, including CXCL5 [118], and promotes remyelination [122]. In cardiac cells, antagonism of $S1PR_1$ after a myocardial infarction blocks cardiac inflammation by inhibiting pro-inflammatory cytokines like TNF-α and IL-6 [119]. The $S1PR_1$ thus provides an important drug target for treating immunological disorders.

The $S1PR_1$ structure was crystallized in complex with the antagonist, ML056. The receptor consists of seven transmembrane helices (TM1-7) embedded in a lipid bilayer, three extracellular loops, three intracellular loops, an N-terminal helix, and a C-terminus [94]. In a previous study, the antagonist-bound and ligand-free conformers of the S1PR$_1$ were simulated by conventional molecular dynamics (cMD) and accelerated molecular dynamics (aMD) techniques [123]. The aMD simulations enhanced the conformational sampling of the receptor. The aMD simulations of the ligand-free receptor captured basal activation of the S1PR$_1$. Activation of the S1PR$_1$ is triggered by a rotameric switch of residue W269$^{6.48}$ in the side chain from gauche to trans [98, 101, 123], and characterized by rearrangement of TM5 and TM7 culminating in a close interaction between Y311$^{7.53}$ and Y221$^{5.58}$ [123].

In this study, we apply a structure-based drug discovery approach [124] to identify novel orthosteric and allosteric ligands for the S1PR$_1$. In this approach, we utilize previous molecular dynamics simulations to fully evaluate the structural conformations of the receptor and obtain representative structures [123]. Druggable binding sites are identified and characterized from the representative structures through fragment mapping, and high-throughput virtual screening (HTVS) and induced-fit docking (IFD) protocols are applied to the desired binding sites, where the compounds are ranked based on the intermolecular interactions between the compound and binding site [124].

Representative conformations from previous simulations were submitted to FTMap for site-mapping which reveals ligand binding sites on the S1PR$_1$. FTMap has identified non-orthosteric sites on the β_1 adrenergic receptor (β_1AR), β_2 adrenergic receptor (β_2AR) [83], M$_2$ muscarinic receptor [84], and the A$_{2A}$ adenosine receptor (A$_{2A}$AR) [125]. Eight non-orthosteric sites are found on the S1PR$_1$. They are distributed in the intracellular, lipid interface, and extracellular regions of the receptor.

Modulators that bind to allosteric sites can alter the signaling effect of the endogenous ligand [80]. Allosteric modulators can also promote subtype specificity, because the regions outside of the orthosteric site are often not as conserved. Three non-orthosteric sites on the $S1PR_1$ were found in most of the representative conformations of the receptor. Non-orthosteric site 1 of the $S1PR_1$ is located between TM2, TM3, TM6, and TM7. Non-orthosteric site 2 is located between the extracellular ends of TM1 and TM7, and non-orthosteric site 3 is located on the intracellular end of TM1, TM2, and TM4. The orthosteric site and these three non-orthosteric sites were screened with the NCI Diversity Set IV and the Chembridge GPCR targeted library for modulators using Schrödinger's HTVS protocol [126]. The top 100 compounds obtained from HTVS for each site were re-ranked with IFD [127-130].

HTVS has been applied to other GPCRs, including the M_2 muscarinic receptor [105, 131] and a GPCR from the fungus, *Fusarium graminearum* [132]. In the M_2 muscarinic receptor, 38 top-ranked compounds were selected from NCI Diversity Set IV with the HTVS and IFD approach. Half of these compounds were confirmed as allosteric modulators after experimental testing [131]. In the present study, 21 potential modulators are identified for the for the orthosteric site and 18 for the three non-orthosteric sites. These compounds can be experimentally evaluated and studied for the treatment of sepsis [121], multiple sclerosis [91, 117, 118, 122], and heart disease [119].

Materials and Methods

Ballesteros and Weinstein Residue Numbering of GPCRs

The numbering format for GPCR residues is X.YY. ‘X’ is the transmembrane helix number (1-7) and ‘YY’ is the relative location of the residue compared to the most conserved residue (labeled 50) [60].

Simulations

In a previous study [123], we performed a total of 12 cMD and aMD simulations of the S1PR$_1$. The starting S1PR$_1$ X-ray structure was determined at 2.80 Å in complex with the antagonist ML056 [94]. Three cMD simulations were performed on the S1PR$_1$ in complex with the antagonist for 200 ns each. Additionally, three cMD simulations were performed on the S1PR$_1$ with the ligand removed (apo) for 200 ns each. Three aMD simulations of the antagonist-bound and apo conformations of the S1PR$_1$ were run from the final step of the cMD simulations for the antagonist-bound and apo forms of the receptors, respectively. Two aMD simulations of the antagonist-bound receptor were run for 200 ns, while one simulation was run for 335 ns. Two aMD simulations of the apo S1PR$_1$ were run for 200 ns, while one simulation lasted for 384 ns. aMD enhances conformational sampling of the S1PR$_1$ by applying a boost of potential energy to smooth the potential energy surface [103, 104]. This results in enhanced sampling because the energy barriers between stable states are reduced.

Identification of Non-Orthosteric Sites on the S1PR$_1$

FTMap is an online sever that identifies potential binding sites on the receptor by screening the surface of the receptor with 16 small probes: ethanol, isopropanol, isobutanol, acetone, acetaldehyde, dimethyl ether, cyclohexane, ethane, acetonitrile, urea, methylamine, phenol, benzaldehyde, benzene, acetamide and N,N-dimethylformamide. The output of FTMap is a PDB file with the coordinates of the probes that interact with the receptor organized by hotspots. The non-orthosteric sites on the S1PR$_1$ were identified by analyzing representative conformations of the S1PR$_1$ obtained from the cMD and aMD simulations with FTMap [38].

The representative receptor conformations from the cMD and aMD simulations were identified from root-mean-square deviation (RMSD)-based structural clustering. The trajectories from the cMD simulations were aligned on the non-hydrogen atoms of the TM helices. A total of 192 receptor clusters were obtained with a RMSD cutoff of 2.0 Å. The first 75 clusters represented 90% of the total trajectories, and were submitted to FTMap for analysis. The trajectories from the aMD simulation were aligned on the non-hydrogen atoms of the TM helices. A total of 519 receptor clusters were obtained with a RMSD cutoff of 2.5 Å. The top 125 receptor clusters represented 90% of the aMD simulation snapshots, and were submitted to FTMap for analysis. The X-ray structure of the S1PR$_1$ was prepared with Schrödinger's Protein Preparation Wizard, where the missing side chains, residues and loops were added [85]. The refined structure was also submitted to FTMap for analysis.

Structural Clustering of the Non-Orthosteric Sites for Docking

FTMap identified 8 non-orthosteric sites on the S1PR$_1$. Three of these sites were selected for molecular docking based on their presence in multiple structural clusters and the X-ray crystal structure of the receptor. The orthosteric site and three non-orthosteric sites were re-clustered based on the RMSD of residues found in the respective sites to generate receptor structural clusters for HTVS screening. The orthosteric site was clustered on residues Y29$^{\mathrm{NTERM}}$, L34$^{\mathrm{NTERM}}$, S105$^{\mathrm{ECL1}}$, R120$^{3.28}$, N101$^{2.60}$, Y98$^{2.57}$, M124$^{3.32}$, L272$^{6.51}$, W269$^{6.48}$, L128$^{3.36}$, L276$^{6.55}$, F273$^{6.52}$, T207$^{5.44}$, F210$^{5.47}$, F125$^{3.33}$, L297$^{7.39}$, L195$^{\mathrm{ECL2}}$, and E121$^{3.39}$. A total of 14 receptor clusters were identified from the cMD simulations with a RMSD cutoff of 2.0 Å. In the aMD simulations, the orthosteric site was clustered with a cutoff of 2.5 Å, leading to a total of 518 receptor clusters. The top 30 receptor clusters accounted for 60% of the trajectory snapshots.

Non-orthosteric site 1 was clustered on residues I59$^{1.46}$, N63$^{1.59}$, L87$^{2.46}$, S90$^{2.49}$, D91$^{2.50}$, A94$^{2.53}$, G95$^{2.54}$, Y98$^{2.57}$, M124$^{3.32}$, F125$^{3.33}$, A127$^{3.35}$, S128$^{3.37}$, S131$^{3.39}$, L135$^{3.43}$, V261$^{6.40}$, F265$^{6.44}$, W269$^{6.48}$, L272$^{6.51}$, L297$^{7.39}$, A300$^{7.42}$, V301$^{7.43}$, N303$^{7.45}$, S304$^{7.46}$, N307$^{7.49}$, and Y311$^{7.53}$. A total of 17 receptor clusters were identified from the cMD simulations with a cutoff of 2.0 Å. In the aMD simulations, non-orthosteric site 1 was clustered with a cutoff of 2.5 Å, generating a total of 77 receptor clusters. The top 30 receptor clusters accounted for 98.9% of the trajectory snapshots.

Non-orthosteric site 2 was clustered on residues I25$^{\mathrm{NTERM}}$, H28$^{\mathrm{NTERM}}$, Y29$^{\mathrm{NTERM}}$, T32$^{\mathrm{NTERM}}$, K34$^{\mathrm{NTERM}}$, L35$^{\mathrm{NTERM}}$, N36$^{\mathrm{NTERM}}$, I37$^{\mathrm{NTERM}}$, S38$^{\mathrm{NTERM}}$, A39$^{\mathrm{NTERM}}$, D40$^{\mathrm{NTERM}}$, K41$^{\mathrm{NTERM}}$, E42$^{1.29}$, S 44$^{1.31}$, I45$^{1.32}$, T48$^{1.35}$, S49$^{1.36}$, F52$^{1.39}$, Y98$^{2.57}$, N101$^{2.60}$, L102$^{2.61}$, L104$^{2.63}$, S105$^{\mathrm{ECL1}}$, G106$^{\mathrm{ECL1}}$, A107$^{\mathrm{ECL1}}$, T109$^{\mathrm{ECL1}}$, Y110$^{\mathrm{ECL1}}$, L290$^{7.32}$, F291$^{7.33}$, R292$^{7.34}$, A293$^{7.35}$, E294$^{7.36}$, Y295$^{7.37}$, F296$^{7.38}$, L297$^{7.39}$, V298$^{7.40}$, and V301$^{7.43}$. A total of 318 receptor clusters were identified in the cMD simulations with a RMSD cutoff of 2.0 Å. The top 30 receptor clusters accounted for 72% of the trajectories. In the aMD simulations, the non-orthosteric site 2 was clustered with a cutoff of 2.5 Å, generating a total of 116 receptor clusters. The top 30 receptor clusters accounted for 96% of the trajectory snapshots.

Non orthosteric site 3 was clustered on residues V66$^{1.53}$, L67$^{1.54}$, I70$^{1.57}$, W71$^{1.58}$, T73$^{1.60}$, K74$^{\mathrm{ICL1}}$, K76$^{\mathrm{ICL1}}$, F77$^{\mathrm{ICL1}}$, R78$^{2.37}$, P79$^{2.38}$, Y81$^{2.40}$, Y82$^{2.41}$, F83$^{2.42}$, G85$^{2.44}$, N86$^{2.45}$, L89$^{2.48}$, N157$^{4.39}$, L160$^{4.42}$, F161$^{4.43}$, I164$^{4.46}$, S165$^{4.47}$, and W168$^{4.50}$. A total of 12 receptor clusters were identified in the cMD simulations with a cutoff of 2.0

Å. In the aMD simulations, non-orthosteric site 3 was clustered with a cutoff of 2.5 Å, generating a total of 887 receptor clusters. The top 30 receptor clusters accounted for 49.7% of the trajectories.

Enrichment Factor

The enrichment factor (EF) scores were calculated to verify the HTVS protocol for the S1PR$_1$. The library used to calculate the EF scores was generated by combining known S1PR$_1$ ligands identified from the GLASS database [133] and the decoy dataset from Schrödinger [134-136]. There were a total of 2523 active ligands and 1067 decoy compounds. The compounds were prepared with Schrödinger's LigPrep [137].

The EF compound library was docked into the 14 receptor clusters of the orthosteric site from the cMD simulations, the top 30 receptor clusters of the orthosteric site from the aMD simulations, and the X-ray structure. The top 33% of compounds from the HTVS precision were re-docked with the Glide SP precision [126, 138, 139]. The top 33% of compounds from Glide SP precision were re-docked with Glide XP precision [140], and the top 33% of compounds from the Glide XP precision were retained and ranked based on the minimum binding energy for (DS$_{min}$). The enrichment factor was calculated by:

$$EF = \frac{(Ligands_{sampled}/N_{sampled})}{(Ligands_{total}/N_{total}\ )} \tag{1}$$

where Ligands$_{sampled}$ represents the number of active compounds above the user set cutoff, e.g. 0.5%, 1%, 2%, 5%, 10% or 15% (Table 5.3). N$_{sampled}$ represents the total number of compounds, decoy and active, above the user set cutoff. Ligands$_{total}$ is the total number of known actives (or 2523) and N$_{total}$ is the total number of compounds in the library (3590). Because of the high ratio of active ligands to decoy compounds, an EF score of 0.42 means that the HTVS protocol does as well as random selection. A value above 0.42 implies the protocol does better than random selection, and a value less than 0.42 means the protocol does worse than random selection.

High-Throughput Virtual Screening

Four sites on the receptor were screened with the NCI Diversity Set IV and Chembridge's GPCR targeted library. The NCI Diversity Set IV has 1596 compounds, and Chembridge's GPCR targeted library has 11,614 compounds. The compounds were prepared with Schrödinger's LigPrep [137]. For the NCI Diversity Set IV, the top 33% of compounds from HTVS precision were re-docked with the Glide SP precision [126, 138, 139]. The top 33% of compounds from the Glide SP precision were re-docked with Glide XP precision [140], and the output was the top 33% of compounds from the Glide XP precision. In Chembridge's GPCR targeted library, the top 10% of compounds from the HTVS precision were re-docked with the Glide SP precision [126, 138, 139]. The top 20% of compounds from the Glide SP precision were re-docked with Glide XP precision [140], and the output was the top 25% of compounds from the Glide XP precision. The compound libraries were docked into the 14, 17, top 30, and 12 structural clusters of the orthosteric site, non-orthosteric site 1, non-orthosteric site 2, and non-orthosteric site 3, respectively.

Induced Fit Docking

The top 50 compounds from the NCI Diversity Set IV and the top 50 compounds from the GPCR targeted library for each binding, determined by the minimum docking score (DS_{min}), were re-docked into the top 14 cMD clusters of the orthosteric site, 17 cMD clusters of non-orthosteric site 1, the top 30 cMD clusters of non-orthosteric site 2, and 12 cMD clusters of non-orthosteric site 3. The DS_{min} was calculated for each compound by:

$$DS_{\min} = Min(DS_1, DS_2, DS_3 ... DS_{Nc}) \tag{2}$$

where N_c is the number of clusters. After applying the IFD protocol, the compounds were r-ranked based on the average docking score (DS_{avg}), which was calculated by considering the weight of each receptor cluster:

$$DS_{avg} = \frac{1}{N}\sum_{i=1}^{Nc} N_i DS_{\min\_i} \tag{3}$$

where N_c is the total number of clusters, N_i is the number of trajectories snapshots in structural cluster *i*, DS_{min_i} is the lowest docking score for that compound, and N is the total number of trajectory snapshots.

The top 50 compounds from the NCI Diversity Set IV and the top 50 compounds from Chembridge's GPCR targeted library were also re-docked into the top 30 aMD receptor clusters for each site, and the X-ray structure using Schrödinger's IFD protocol [126, 138, 140].

Results

Non-Orthosteric Sites on the $S1PR_1$

Six non-orthosteric sites were identified from cMD simulations of the $S1PR_1$ (non-orthosteric site 1-6) (Figure 5.1A), and 7 non-orthosteric sites were identified from the aMD simulation (non-orthosteric sites 1-4, 6-8) (Figure 5.1B). Non-orthosteric site 1 is located in between TM2, TM3, TM6, and TM7 (Figure 5.2, Figure 5.8, and Figure 5.9). Non-orthosteric site 2 is located between the extracellular end of the N-terminus, TM1, TM2, ECL1, TM3, and TM7 (Figure 5.2, Figure 5.11, and Figure 5.12). Non-orthosteric site 3 is located on the intracellular ends of TM1, ICL1, TM2, and TM4 (Figure 5.2 and Figure 5.14). Non-orthosteric site 4 is located on the intracellular ends of TM1, TM7, and the C-terminus (Figure 5.1). Non-orthosteric site 5 is located between the intracellular regions of TM3, TM5, and TM6 (Figure 5.1A). Non-orthosteric site 6 is located between TM2, TM3 and TM4 (Figure 5.1). Non-orthosteric site 7 is located on the lipid interface of TM3, TM4, and TM5 (Figure 5.1B), and non-orthosteric site 8 is located on the lipid interface of TM2, TM3, and TM4 (Figure 5.1B). The key residues in each of these sites are listed in Table 5.1.

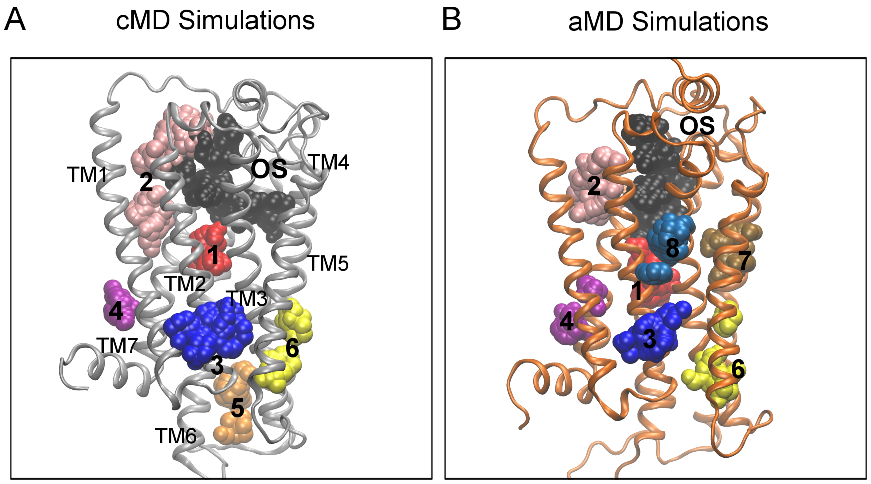

Figure 5.1: All Hotspots on the S1PR$_1$. Several hotspots for probe binding in the cMD simulations and aMD simulations are shown. A representative conformation from the cMD simulations is shown in gray [A]. A representative conformation from the aMD simulations is shown in orange [B]. Non-orthosteric site 1 is shown in red, site 2 in pink, site 3 in blue, site 4 in purple, site 5 in orange, site 6 in yellow, site 7 in tan, and site 8 in dark blue.

Table 5.1: Orthosteric and non-orthosteric sites on the S1PR$_1$. The site number, location, receptor regions, and residues are listed.

Site Number	Site:	Regions	Residues
OS	Orthosteric Site	N-Term/TM2/TM3/TM6/TM7	Y29, K34 (N-Term) Y98$^{2.57}$, N101$^{2.60}$ (TM2) R120$^{3.28}$, E121$^{3.29}$, M124$^{3.32}$, F125$^{3.33}$, L128$^{3.36}$ (TM3) L195 (ECL2) W269$^{6.48}$, L272$^{6.51}$, L276$^{6.55}$ (TM6) L297$^{7.39}$, A300$^{7.42}$ (TM7)
1	Sodium Ion Binding Site	TM2/TM3/TM6/TM7	D91$^{2.50}$, Y98$^{2.57}$ (TM2) M124$^{3.32}$, L128$^{3.36}$, S131$^{3.39}$ (TM3) F265$^{6.44}$, W269$^{6.48}$ (TM6) V301$^{7.43}$, N307$^{7.49}$, Y311$^{7.53}$ (TM7)
2	Ligand Entry Site	N-Term/TM1/TM2/ECL1/TM3/TM7	Y29, K34 (N-Term) S49$^{1.36}$, F52$^{1.39}$ (TM1) N101$^{2.60}$ (TM2) S105 (ECL1) R120$^{3.28}$ (TM3) E296$^{7.36}$, V301$^{7.43}$ (TM7)
3	Intracellular Crevice	TM1/ICL1/TM2/TM4	W71$^{1.58}$ (TM1) F77 (ICL1) Y81$^{2.40}$, N86$^{2.45}$ (TM2) N157$^{4.39}$, F161$^{4.43}$, W168$^{4.50}$ (TM4)
4	Intracellular Regions	TM1/TM7/C-Term	F65$^{1.52}$ (TM1) I309$^{7.51}$ (TM7) F322$^{\text{C-TERM}}$ (C-Term)
5		TM3/TM5/TM6	M146$^{3.54}$ (TM3) Y221$^{5.58}$ (TM5) K256$^{6.35}$ (TM6)
6		TM2/TM3/TM4	F83$^{2.42}$ (TM2) F133$^{3.41}$, E141$^{3.49}$ (TM3) C167$^{4.49}$ (TM4)
7	Lipid Interface	TM3/TM4/TM5	F125$^{3.33}$, A130$^{3.38}$, F133$^{3.41}$ (TM3) L174$^{4.56}$ (TM4) F205$^{5.44}$, L213$^{5.50}$ (TM5)
8		TM2/TM3	V96$^{2.55}$ (TM2) S123$^{3.31}$, V126$^{3.34}$ (TM3) W168$^{4.50}$ (TM4)

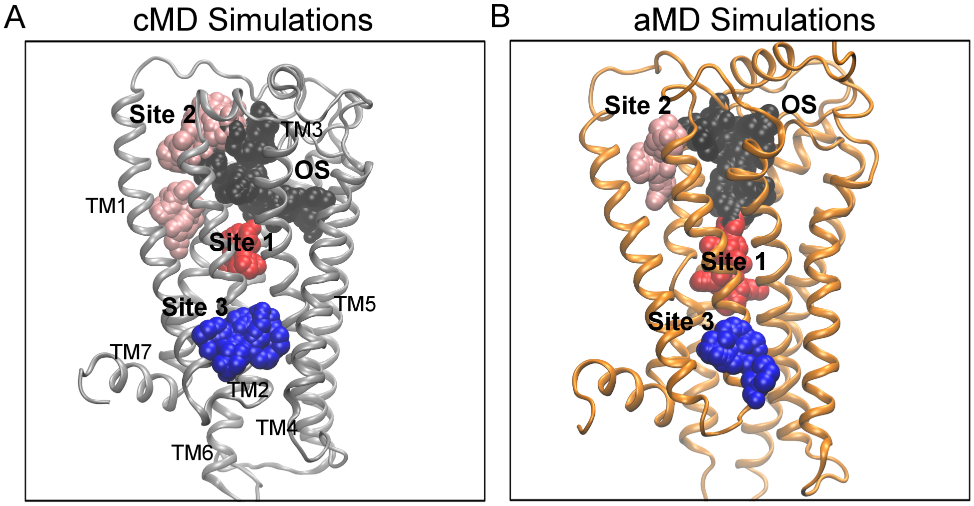

Figure 5.2: The Screened Binding Sites on the S1PR$_1$. Four hotspots for probe binding that were screened with HTVS methods in the cMD simulations and the aMD simulations include the orthosteric site, the sodium ion binding site (site 1), the ligand entry site (site 2), and the intracellular crevice (site 3). A representative conformation from the cMD simulations is shown in gray [A]. A representative conformation from the aMD simulations is shown in orange [B]. The orthosteric site (OS) is shown as black spheres, site 1 as red spheres, site 2 as pink spheres, and site 3 as blue spheres.

In this study, the orthosteric site and three non-orthosteric sites on the S1PR$_1$, non-orthosteric sites 1-3, were screened for potential modulators (Figure 5.2). These sites were present in most of the receptor structural clusters obtained from the cMD and aMD simulations. Non-orthosteric site 1, located directly under the GPCR orthosteric site, includes residues in the sodium ion binding site, the G protein-binding site, and the intracellular water pathway of the receptor (Figure 5.2). Site 2 is located at the ligand entry site (Figure 5.2). Site 3 is located on the intracellular ends of TM1, TM2, and TM4 (Figure 5.2).

Self-Docking of the Co-Crystallized Ligand to the Orthosteric Site of the S1PR$_1$

The orthosteric site of the S1PR$_1$ involves residues in the N-terminus, TM2, TM3, ECL2, TM6, and TM7 regions. Key residues include W269$^{6.48}$, which is involved in the intracellular water pathway [101]. W269$^{6.48}$ is known as the GPCR toggle switch, and in previous studies of the S1PR$_1$ activation pathway, the χ_2 conformational change initiated the transition of the receptor from inactive to active [101, 123].

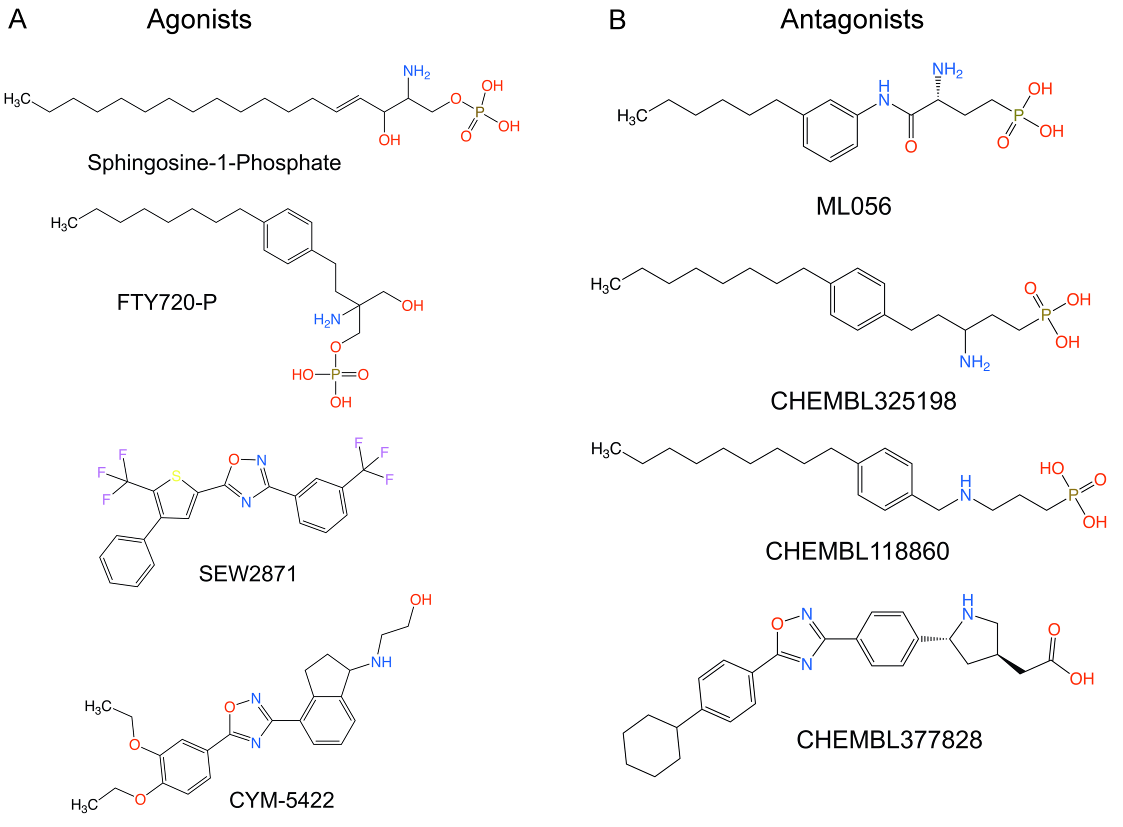

Figure 5.3: Agonists and Antagonists for the S1PR$_1$. The structure of the known agonists, sphingosine-1-phosphate, FTY720-P, SEW2871, and CYM-5422 are shown [A]. The structure of the known antagonists, ML056, CHEMBL325198, CHEMBL118860, and CHEMBL377828 are shown [B].

Self-docking of the co-crystallized ligand, ML056, to the S1PR$_1$ can validate the docking method. ML056 has a charged head group and an alkyl tail (Figure 5.3B). The crystal structure (CS) and lowest energy pose from the IFD results reveal that the head group of ML056 forms hydrogen bonds with residues K34$^{\text{N-TERM}}$, R120$^{3.28}$, E121$^{3.28}$, and N101$^{2.60}$, and the non-polar tail group has hydrophobic interactions with residues

F125$^{3.33}$, L195^{ECL2}, L128$^{3.36}$, M124$^{3.32}$, and L297$^{7.39}$ (Figure 5.3B and Figure 5.4A). The minimum docking score (DS$_{min}$) of ML056 to the crystal structure of the S1PR$_1$ is -9.62 kcal/mol. Comparatively, the average docking score (DS$_{avg}$) of ML056 to the cMD and aMD receptor clusters is -8.57 kcal/mol and -6.47 kcal/mol, respectively (Table 5.2).

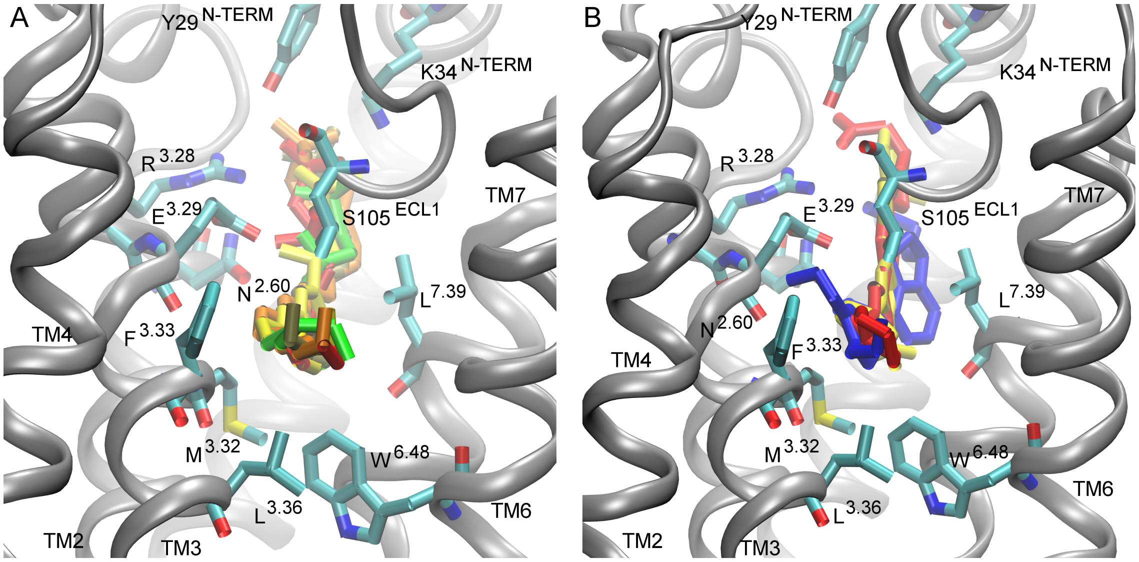

Figure 5.4: Docking Poses of Known Ligands. The ligands with a charged head group and alkyl tail are shown in A. These ligands include sphingosine-1-phosphate (orange), FTY720-P (red), ML056 (yellow), CHEMBL325198 (green), and CHEMBL118860 (tan). The known ligands with a oxadiazole core are shown in B. These ligands include SEW2871 (yellow), CYM-5422 (blue), and CHEMBL377828 (red).

Cross-Docking of Known Agonists and Antagonists to the S1PR$_1$

Cross-docking known agonists and antagonists to the S1PR$_1$ can validate the docking method and provide a range of average docking scores for known actives. The agonists include the endogenous ligand (S1P), FTY720-P, CYM-5422, and SEW2871. The antagonists include CHEMBL325198, CHEMBL377828, and CHEMBL118860 (Figure 5.3). Agonists S1P and FTY720-P, and antagonists CHEMBL325198 and CHEMBL118860 have a charged head group, and an alkyl chain tail (Figure 5.3). The lowest energy poses from IFD docking these known ligands to the S1PR$_1$ reveal that the head groups of S1P, FTY720-P, CHEMBL325198, and CHEMBL118860 form hydrogen bond interactions with residues Y29$^{\text{N-TERM}}$, K34$^{\text{N-TERM}}$, and R120$^{3.28}$, and the alkyl chain interacts with L195$^{\text{ECL2}}$, L128$^{3.36}$, M124$^{3.32}$, and L297$^{7.39}$ (Figure 5.3 and Figure 5.4). FTY720-P also has hydrophobic interactions with W269$^{6.48}$. The docking poses of these compounds are similar to the docking pose of ML056, which has been validated with the crystal structure [94]. The DS$_{\text{min}}$ of S1P to the crystal structure of the S1PR$_1$ is -7.96 kcal/mol. The DS$_{\text{avg}}$ of S1P to the cMD receptor clusters is -7.89 kcal/mol and -5.16 kcal/mol in the aMD receptor clusters (Table 5.2). The DS$_{\text{min}}$ of agonist FTY720-P to the crystal structure of the S1PR$_1$ is -7.47 kcal/mol. The DS$_{\text{avg}}$ of FTY720-P to the cMD receptor clusters is -9.75 kcal/mol and -6.33 kcal/mol in the aMD receptor clusters (Table 5.2). The antagonist CHEMBL325198 does not bind to the crystal structure of the S1PR$_1$, but the DS$_{\text{avg}}$ of CHEMBL325198 to the cMD receptor clusters is -8.20 kcal/mol and -4.34 kcal/mol in the aMD receptor clusters (Table 5.2). The DS$_{\text{min}}$ of antagonist CHEMBL118860 to the crystal structure of the S1PR$_1$ is -7.48 kcal/mol. The DS$_{\text{avg}}$ of CHEMBL118860 to the cMD receptor clusters is -7.42 kcal/mol and -4.35 kcal/mol in the aMD receptor clusters (Table 5.2).

Table 5.2: Known agonists, S1P, FTY720-P, SEW2871, and CYM-5422, and known antagonists, ML056, CHEMBL325198, CHEMBL118860, and CHEMBL377828, were docked against cMD and aMD receptor clusters, and the crystal structure (CS). The average docking score (denoted DS_{avg}) of these ligands to the cMD and aMD receptor clusters are listed. The minimum docking score (denoted DS_{min}) of these ligands to the CS is also listed.

$S1PR_1$ Ligand	Affinity	cMD DS_{avg} (kcal/mol)	aMD DS_{avg} (kcal/mol)	CS DS_{min} (kcal/mol
Agonists				
S1P	EC50: 0.027nM	-7.89	-5.16	-7.96
FTY720 -P	EC50: 0.3nM	-9.75	-6.33	-7.47
CYM-5422	EC50: 1.23nM	-6.61	-5.52	n/a
SEW2871	EC50: 12.5nM	-6.5	-5.04	-7.01
Antagonists				
ML056	Ki = 18nM	-8.57	-6.47	-9.62
CHEMBL325198	IC50: 8.4nM	-8.20	-4.34	n/a
CHEMBL377828	IC50: 1.4nM	-8.69	-3.81	n/a
CHEMBL118860	IC50: 0.2nM	-7.42	-4.35	-7.48

The chemical structures of agonists SEW2871 and CYM-5422, and antagonist CHEMBL377828 include an oxadiazole core (Figure 5.3). In SEW2871, the oxadiazole core forms π-cation interactions with residue $R120^{3.28}$. The trifluoromethyl group attached to a benzene ring in SEW2871, the ethanol group on CYM-5422, and the acetic acid functional group of CHEMBL377828 form interactions with $K34^{N\text{-}TERM}$, $S105^{ECL1}$, and $Y29^{N\text{-}TERM}$. The phenyl group of SEW2871, the diethyl phenyl functional group of CYM-5422, and the cyclohexane group on CHEMBL377828 form hydrophobic interactions with residues $F125^{3.33}$, $L195^{ECL2}$, and $M124^{3.32}$ (Figure 5.3 and Figure 5.4B). The DS_{min} of the agonist SEW2871 to the crystal structure of the $S1PR_1$ is -7.01 kcal/mol (Table 5.2). The DS_{avg} of SEW2871 to the cMD receptor clusters is -6.5

kcal/mol and -5.04 kcal/mol in the aMD receptor clusters. The agonist CYM-5422 does not bind to the crystal structure, but the DS_{avg} of CYM-5422 to the cMD receptor clusters is -6.61 kcal/mol and -5.52 kcal/mol in the aMD receptor clusters (Table 5.2). The crystal structure does not bind to the antagonist CHEMBL377828, but the DS_{avg} of CHEMBL377828 to the cMD receptor clusters is -8.69 kcal/mol and -3.81 kcal/mol in the aMD clusters (Table 5.2).

The known agonists and antagonists have more favorable docking scores to the crystal structure and the cMD receptor clusters than the aMD receptor clusters. In the crystal structure, several ligands, including CYM5422, CHEMBL325198, and CHEMBL377828 do not bind, reinforcing previous studies that suggest that relaxation of the receptor structure is necessary to capture the binding of all known ligands [131, 141]. On average, the cMD receptor clusters have more favorable docking scores for the antagonists compared to the agonists, while the aMD clusters dock more favorably to the agonists, compared to the antagonists (Table 5.2).

Verification of the HTVS Protocol

The enrichment factor (EF) score is an analytical metric used to evaluate the ability of Schrödinger's HTVS protocol [126, 134, 136, 138-140] to distinguish between the known ligands and decoys [135, 136]. Retrospective docking can determine whether the HTVS protocol ranks active compounds better than random selection. In this analysis, an EF score of 0.42 is the value calculated when the protocol does as well as random selection, and an EF score of 1.42 indicates that all of the compounds above the cutoff are known ligands, not decoy compounds. At a cutoff of 0.5%, 1%, 2%, 5%, 10%, and 15%, the HTVS consistently ranks the known ligands above the decoy compounds in the cMD clusters, aMD clusters, and crystal structure, suggesting that this protocol is effective at ranking active ligands above decoy compounds (Table 5.3).

Table 5.3: The enrichment factor (EF) score obtained from retrospective docking of known ligands against the cMD, aMD receptor clusters, and crystal structure (CS) are listed.

Percentage	cMD EF (DSmin)	aMD EF (DSmin)	CS (DSmin)
0.5%	1.42	1.42	1.42
1%	1.42	1.42	1.42
2%	1.42	1.42	1.42
5%	1.42	1.42	1.42
10%	1.42	1.42	1.42
15%	1.42	1.42	0.95

Top-Ranked Compounds that Bind to the Orthosteric Site

The orthosteric site docks favorably to 18 compounds from Chembridge's GPCR targeted library and three compounds from the NCI Diversity set IV. The top 3 compounds from Chembridge's GPCR targeted library are CHEMBID-45291808 (DS_{avg}: -10.19 kcal/mol), CHMBID-42911631 (DS_{avg}: -10.03 kcal/mol), and CHMBID-50250081 (DS_{avg}: -10.02 kcal/mol) (Figure 5.5, Figure 5.6, and Table 5.4). The top 3 compounds from the NCI diversity set IV are NSC-319990 (DS_{avg}: -9.39 kcal/mol), NSC-34875 (DS_{avg}: -9.09 kcal/mol), and NSC-305780 (DS_{avg}: -8.98 kcal/mol) (Figure 5.5, Figure 5.6, and Table 5.4). 18 additional compounds with low binding energies from the Chembridge GPCR targeted library were also identified (Table 5.5 and Figure 5.7).

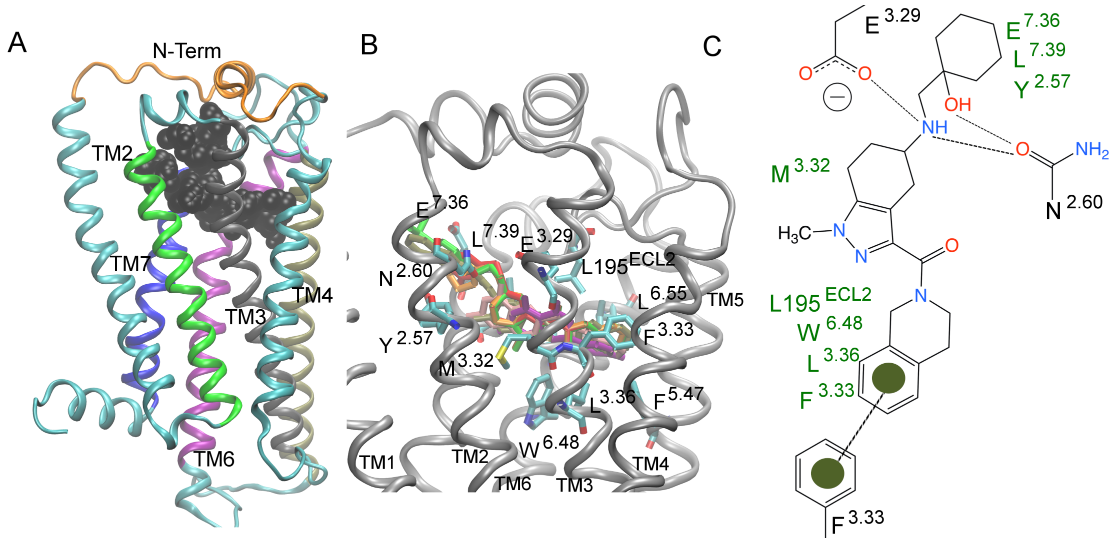

Figure 5.5: The Orthosteric Site. The orthosteric site is located between the N-terminus (orange), TM2 (green), TM3 (gray), TM4 (tan), TM6 (purple), TM7 (blue). The probes in the orthosteric site are shown as black spheres [A]. The top six ligands, CHMBID-45291808 (orange), CHMBID-42911631 (green), CHMBID-50250081 (red), NSC-319990 (tan), NSC-34875 (purple), NSC-305780 (pink) are docked to the orthosteric site [B]. The binding pose of CHMBID-45291808 is shown in complex with the receptor [C]. Residues in green have hydrophobic interactions with the ligand. The dotted lines represent hydrogen bonds between the residues and ligand. The green dots represent π-π interactions between the residues and ligand.

Table 5.4: The top-ranking compounds from the Chembridge GPCR targeted library and the NCI diversity set IV that bind to the orthosteric site are shown. The average docking score (denoted DS_{avg}) is listed for the cMD and aMD receptor clusters. The minimum docking score (denoted DS_{min}) for the crystal structure (CS) is listed.

Compound ID	cMD DS_{avg} (kcal/mol)	aMD DS_{avg} (kcal/mol)	CS DS_{min} (kcal/mol)
Chembridge GPCR Library			
45291808	-10.19	-5.71	-8.32
42911631	-10.03	-5.03	-7.94
50250081	-10.02	-5.94	-8.24
NCI Diversity Set IV			
319990	-9.39	-5.97	-7.08
34875	-9.09	-4.53	-6.84
305780	-8.98	-5.77	-7.0

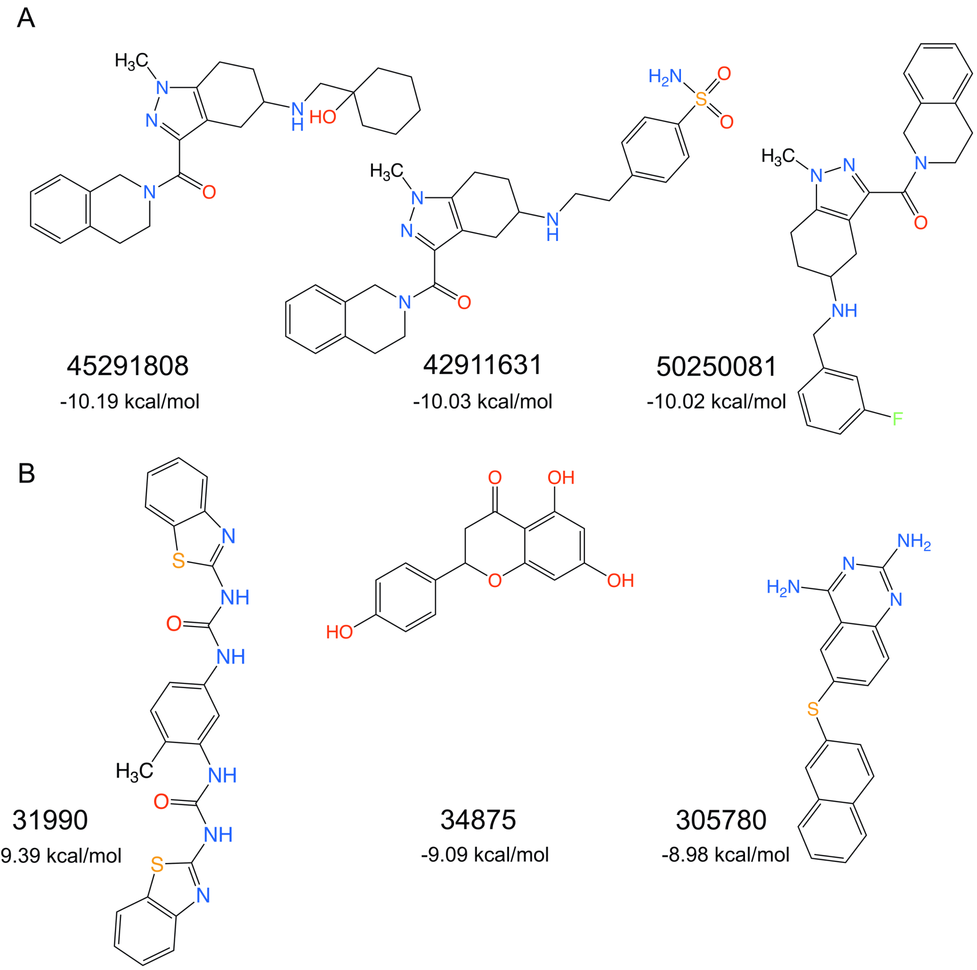

Figure 5.6: Compounds that Bind to the Orthosteric Site. The compound structures of the top three ligands that bind to the orthosteric site from the Chembridge GPCR Library [A], (CHMBID-45291808, CHMBID-42911631, CHMBID-50250081), and the top three ligands from the NCI Diversity Set IV [B] (NSC-319990, NSC-34875, NSC-305780) are shown. The DS_{avg} for each compound to the cMD receptor clusters is listed.

The compounds that dock to the orthosteric site interact with residues in TM1, TM2, TM3, TM5, TM6, and TM7, but not the N-terminus, as seen with the known orthosteric ligands (Figure 5.5B and Figure 5.4). The compounds from the Chembridge GPCR targeted library have a common chemical core, [3-(3,4-dihydro-2(1H)-isoquinolinylcarbonyl)-1-methyl-4,5,6,7-tetrahydro-1H-indazol-5-yl]amino (Figure 5.6 and Figure 5.7). The core forms π -π interactions with $F125^{3.33}$, and hydrophobic interactions with residues L128$^{3.36}$, W269$^{6.48}$, L195^{ECL2}, and M124$^{3.32}$ (Figure 5.5C). These residues also interact with the alkyl chain of the ligands, S1P, ML056, CHEMBL325198, and CHEMBL118860 (Figure 5.4A), and the phenyl group, diethyl phenyl group, and cyclohexane group on SEW2871, CYM-5422, and CHEMBL377828, respectively. Residues N101$^{2.60}$ and E121$^{3.29}$ form hydrogen bonds with the amine group in the common chemical core (Figure 5.5C). Residue N101$^{2.60}$ also forms a hydrogen bond with the oxygen atom in the cyclohexanol group of CHEMBID-45291808, and residues Y98$^{2.57}$, L297$^{7.39}$, and E294$^{7.36}$ have hydrophobic interactions with the cyclohexanol group of CHEMBID-45291808.

In compound CHMBID-42911631, the sulfonamide functional group forms a hydrogen bond with E294$^{7.36}$ (Figure 5.5B and Figure 5.6A). In compound CHMBID-50250081, the fluorobenzene functional group has hydrophobic interactions with residue L297$^{7.39}$ (Figure 5.5B and Figure 5.6A). In the remaining Chembridge compounds that bind with high favorability to the orthosteric site, the functional groups interact with residues N101$^{2.60}$, E121$^{3.29}$, Y98$^{2.57}$, L297$^{7.39}$, and E294$^{7.36}$ (Figure 5.5B and Figure 5.6).

In compound NSC-34875 from the NCI Diversity Set IV, the phenol group forms π-π interactions with F125$^{3.33}$, and hydrophobic interactions with residues L128$^{3.36}$ and F125$^{3.33}$. The oxygen atom on the chromane group forms a hydrogen bond with residue L297$^{7.39}$ (Figure 5.5B and Figure 5.6B). In compound NSC-305780, the naphthalene group forms π-π interactions with residue F125$^{3.33}$ and hydrophobic interactions with residues L128$^{3.36}$, L276$^{6.55}$, L272$^{6.51}$, M124$^{3.32}$, L195^{ECL2}, W269$^{6.48}$, F125$^{3.33}$, F210$^{5.47}$, and L297$^{7.39}$. The amine groups on NSC-305780 form hydrogen bonds with residues N101$^{2.60}$ and L297$^{7.39}$ (Figure 5.5B and Figure 5.6B). In compound NSC-319990, a benzothiazole functional group forms π-π interactions with F125$^{3.33}$. The phenylene group on NSC-305780 has hydrophobic interactions with residues M124$^{3.32}$, L195^{ECL2}, and L297$^{7.39}$, and the adjacent urea functional group forms a hydrogen bond with residue N101$^{2.60}$ (Figure 5.5B and Figure 5.6B).

Table 5.5: The top compounds that bind to the orthosteric site from the Chembridge GPCR targeted library are listed. The average docking score (denoted DS_{avg}) is listed for the cMD and aMD receptor clusters. The minimum docking score (denoted DS_{min}) is listed for the crystal structure (CS).

Chembridge ID	cMD DS_{avg} (kcal/mol)	aMD DS_{avg} (kcal/mol)	CS DS_{min} (kcal/mol)
45291808	-10.19	-5.7	-8.32
42911631	-10.03	-5.03	-7.94
50250081	-10.02	-5.94	-8.24
38225912	-10.02	-5.08	-7.31
20544582	-9.92	-5.9	-7.71
74523809	-9.83	-6.10	-7.73
93406380	-9.77	-4.78	-7.49
50234778	-9.75	-5.46	-6.38
36974671	-9.72	-5.59	-7.6
60591631	-9.70	-4.79	-6.12
29805761	-9.69	-5.49	-7.28
18786700	-9.67	-5.73	-7.90
65462646	-9.64	-4.61	-6.73
47736738	-9.63	-5.79	-7.81
44744047	-9.62	-4.78	-5.14
19939660	-9.58	-5.84	-8.20
35726876	-9.58	-5.47	-7.35
19999002	-9.56	-5.58	-6.60

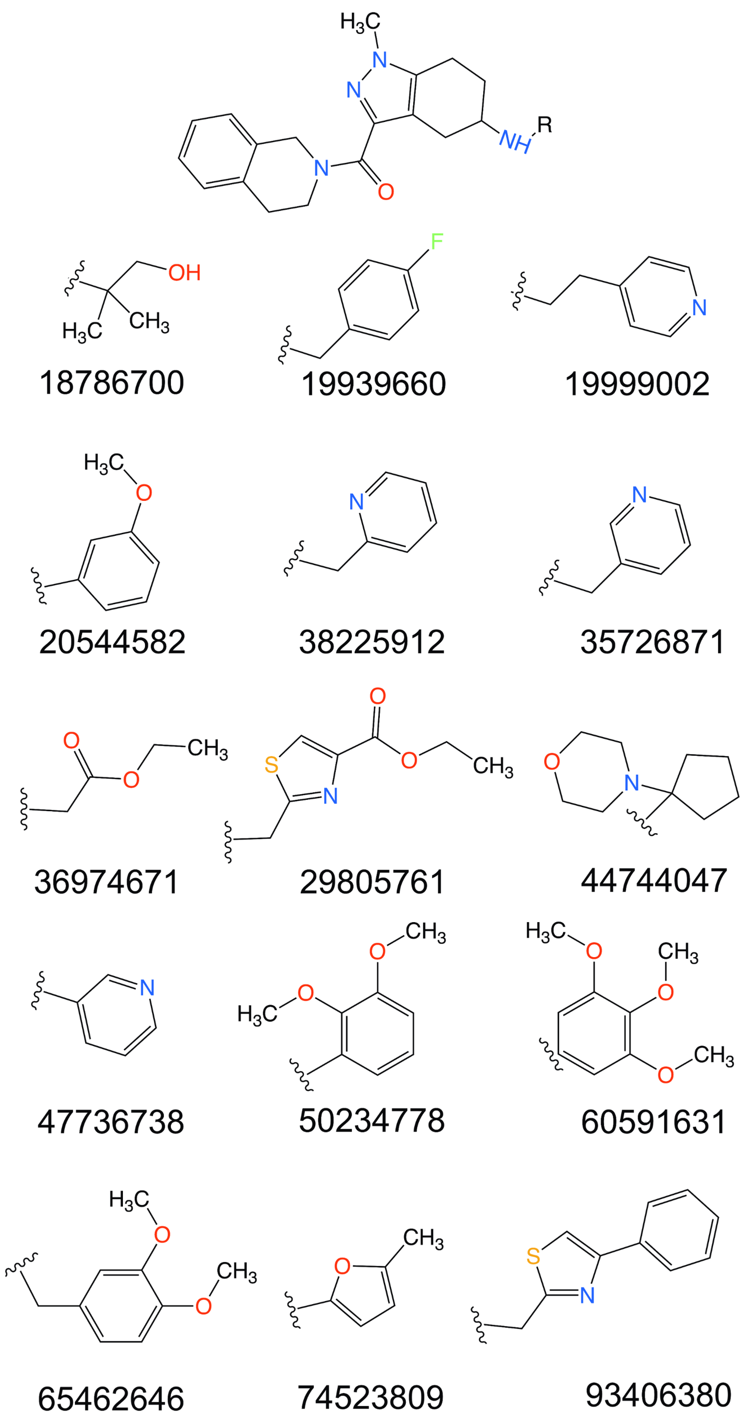

Figure 5.7: The compound structures for the remaining 15 compounds from Chembridge's GPCR targeted library that bind favorably to the orthosteric site are shown. The common compound core is shown at the top of the figure. The functional groups are shown below. The curved line on the functional groups denote where the functional group attaches to the common compound core. The top 3 compounds (CHMBID-45291808, CHMBID-42911631, CHMBID-50250081) are not shown.

Non-Orthosteric Site 1: The Sodium Ion Binding Site

The “sodium ion binding site” is located in between TM2, TM3, TM6, and TM7 (Figure 5.8 and Figure 5.9). In the cMD simulations, residues encompassing the sodium ion binding site include conserved D91$^{2.50}$, N307$^{7.49}$ in the conserved NPxxY motif, as well as residues L128$^{3.36}$, W269$^{6.48}$, and V301$^{7.43}$ (Figure 5.8B). The probes that interact with this site include acetaldehyde, acetonitrile, ethane, acetamide, urea, ethanol, methylamine, and dimethyl ether (Figure 5.8B). The sodium ion binding site is larger in the aMD simulations than in the cMD simulations (Figure 5.9A). In the aMD simulations, the key residues include conserved residue D91$^{2.50}$, the toggle switch W269$^{6.48}$, N307$^{7.49}$ and Y311$^{7.53}$ of the NPxxY motif, S131$^{3.39}$, M124$^{3.32}$, Y98$^{2.57}$, and F265$^{6.44}$ (Figure 5.9B). The probes that interact with the sodium ion binding site in the aMD representative receptor clusters are dimethyl ether, ethane, ethanol, methylamine, acetaldehyde, urea, acetonitrile, and benzene (Figure 5.9B).

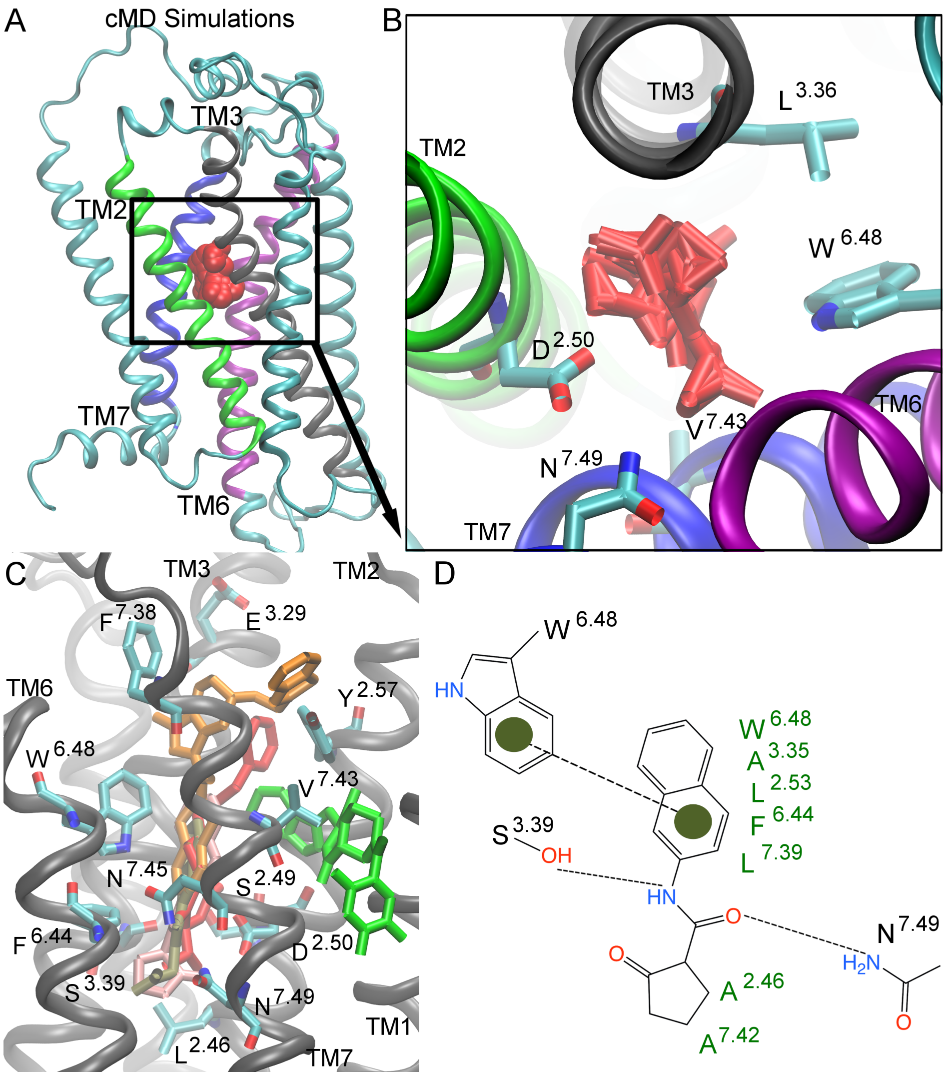

Figure 5.8: The Sodium Ion Binding Site. The sodium ion binding site in the cMD simulations is located between TM2 (green), TM3 (gray), TM6 (purple), and TM7 (blue). The probes in the sodium ion binding site are shown as red spheres [A]. The key interacting residues in the cMD simulations are shown as bonds, and the probes that bind to this site are shown as red bonds [B]. The top six ligands, CHMBID-57373503 (red), CHMBID-67337572 (orange), CHMBID-98659294 (green), NSC-88795 (purple), NSC-106506 (tan), and NSC-89759 (pink) are docked to this site [C]. The binding pose of NSC-88795 is shown in complex with the receptor [D]. Residues in green have hydrophobic interactions with the ligand. The dotted lines represent hydrogen bonds between the residues and ligand. The green dots represent π-π interactions between the residues and ligand.

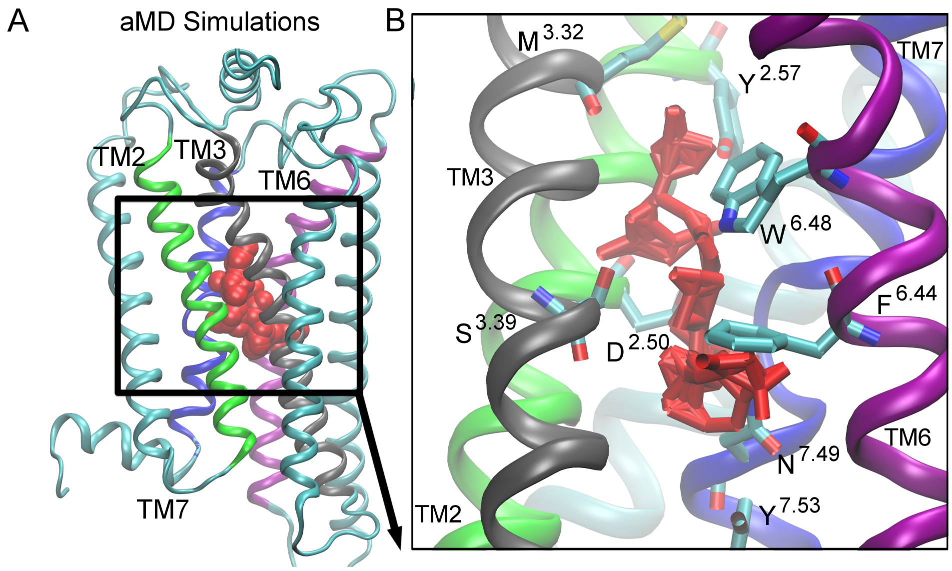

Figure 5.9: The Sodium Ion Binding Site in the aMD Simulations. The sodium ion binding site in the aMD simulations is also located between TM2 (green), TM3 (gray), TM6 (purple), and TM7 (blue). The probes in the sodium ion binding site are shown as red spheres [A]. The key interacting residues in the aMD simulations are shown as bonds, and the probes that bind to this site are shown as red bonds [B].

Top-Ranked Compounds that Bind to the Sodium Ion Binding Site

CHMBID-57373503 (DS_{avg}: -7.44 kcal/mol), CHMBID-67337572 (DS_{avg}: -7.44 kcal/mol), and CHMBID-98659294 (DS_{avg}: -6.72 kcal/mol) are the top 3 compounds from the Chembridge GPCR targeted library that docked to the sodium ion binding site with high favorability. The top 3 compounds from the NCI diversity set IV are NSC-88795 (DS_{avg}: -7.52 kcal/mol), NSC-106506 (DS_{avg}: -7.48 kcal/mol), and NSC-89759 (DS_{avg}: -7.24 kcal/mol) (Figure 5.10 and Table 5.6). These compounds interact with residues in TM2, TM3, TM6 and TM7 (Figure 5.8C). In compound NSC-88795 of the NCI Diversity Set IV, the naphthalene group forms π-π interactions with residue $W269^{6.48}$ (Figure 5.8B and Figure 5.10B). The naphthalene group also has hydrophobic interactions with residues $W269^{6.48}$, $A127^{3.35}$, $L87^{2.53}$, $F265^{6.44}$, and $L297^{7.39}$ (Figure 5.8B and Figure 5.10B). The nitrogen atom of the amide group forms a hydrogen bond with residue S131$^{3.39}$, and the amide oxygen atom forms a hydrogen bond with residue N307$^{7.49}$ of the NPxxY motif. The oxycyclopentane group has hydrophobic interactions with A94$^{2.46}$ and A300$^{7.42}$ (Figure 5.8D and Figure 5.10B).

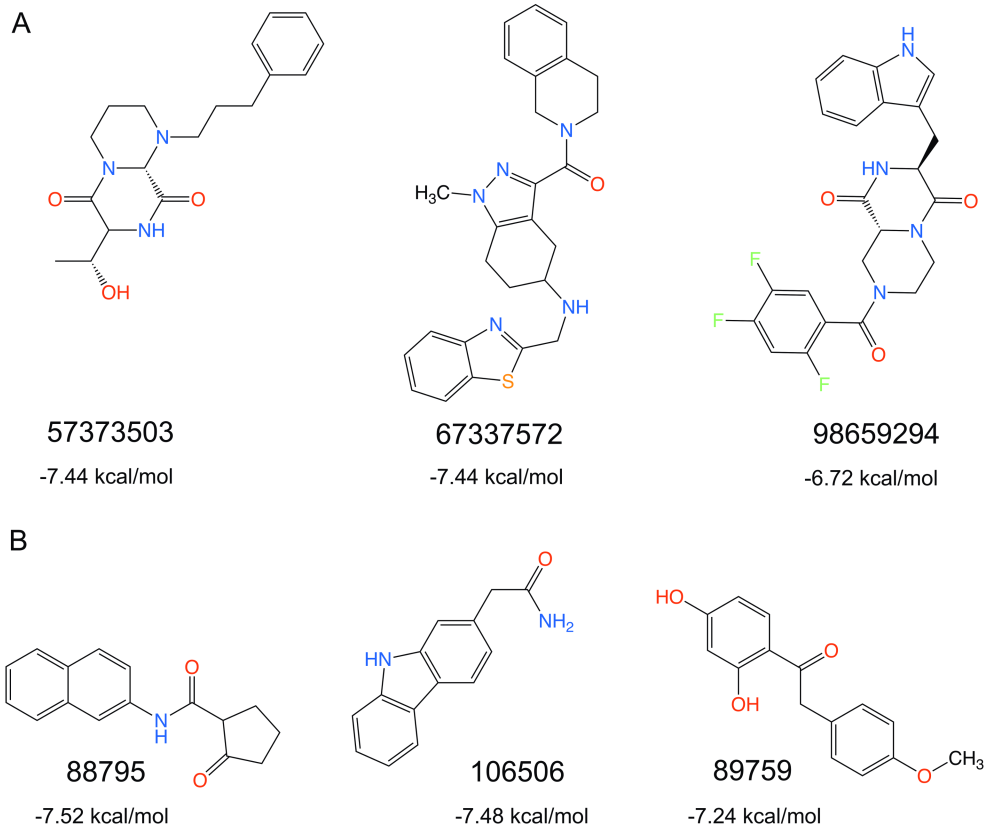

Figure 5.10: Compounds that Bind to the Sodium Ion Binding Site. The compound structures of the top three ligands that bind to the sodium ion binding site from the Chembridge GPCR targeted library [A] (CHMBID-57373503, CHMBID-67337572, CHMBID-98659294), and the top three ligands from the NCI Diversity Set IV [B] (NSC-88795, NSC-106506, and NSC-89759) are shown. The DS_{avg} for each compound to the cMD receptor clusters is listed.

In compound NSC-89759, the oxygen atoms in the dihydroxyphenyl group form hydrogen bonds with E121$^{3.29}$. Both phenyl groups have hydrophobic interactions with residues W269$^{6.48}$ and L297$^{7.39}$. The methoxyphenyl group forms π-π interactions with residue F125$^{3.33}$, and has hydrophobic interactions with residues F125$^{3.33}$ and L195^{ECL2} (Figure 5.8C and Figure 5.10B). In compound NSC-106506, the oxygen in the amide group forms a hydrogen bond with residue N307$^{7.49}$, and has hydrophobic interactions with residues V261$^{6.40}$ and L135$^{3.43}$. The carbazole group forms π-π interactions with residues F265$^{6.44}$ and W269$^{6.48}$, and hydrophobic interactions with residues N303$^{7.45}$ and D91$^{2.50}$ (Figure 5.8C and Figure 5.10B).

Table 5.6: The top-ranking compounds from the Chembridge GPCR targeted library and the NCI diversity set IV that bind to the sodium ion binding site are shown. The average docking score (denoted DS_{avg}) is listed for the cMD and aMD receptor clusters. The minimum docking score (denoted DS_{min}) for the crystal structure (CS) is listed.

Compound ID	cMD DS_{avg} (kcal/mol)	aMD DS_{avg} (kcal/mol)	CS DS_{min} (kcal/mol)
Chembridge GPCR Library			
57373503	-7.44	-1.30	n/a
67337572	-7.44	-3.69	n/a
98659294	-6.72	-3.57	n/a
NCI Diversity Set IV			
88795	-7.52	-4.50	-5.98
106506	-7.48	-4.53	-7.21
89759	-7.24	-3.70	-7.84

In compound CHMBID-57373503 from the Chembridge GPCR targeted library, the phenyl group has hydrophobic interactions with residues Y98$^{2.57}$, N101$^{2.60}$, and M124$^{3.32}$. The oxygen atom in the hydroxyl group forms a hydrogen bond with residue L87$^{2.46}$. The oxygen atom from the amide group forms a hydrogen bond with residue S90$^{2.49}$, and the pyrazino[1,2-a]pyrazine group has hydrophobic interactions with residues N303$^{7.45}$, F265$^{6.44}$, and N307$^{7.49}$ (Figure 5.8C and Figure 5.10A). In compound CHMBID-67337572, the isoquinolinylcarbonyl group forms π-π interactions with W269$^{6.48}$, and has hydrophobic interactions with residues D91$^{2.50}$, N303$^{7.45}$, and N307$^{7.49}$. The benzothiazole group has hydrophobic interactions with residue E294$^{7.36}$. The indazol group forms π-π interactions with residue F296$^{7.38}$ (Figure 5.8C and Figure

5.10A). In compound CHMBID-98659294, the pyrazino[1,2-a]pyrazine group forms π-π interactions with residue $Y98^{2.57}$, and has hydrophobic interactions with residues $I55^{1.42}$, $V301^{7.43}$, $A94^{2.53}$, $Y98^{2.57}$, $S304^{7.46}$, and $I59^{1.46}$. The nitrogen atom in the amide groups forms a hydrogen bond with $V301^{7.43}$ (Figure 5.8C and Figure 5.10A).

Non-Orthosteric Site 2: The Ligand Entry Site

The "ligand entry site" is located between the extracellular ends of TM1 and TM7 (Figure 5.11 and Figure 5.12). S1P enters the orthosteric site of the $S1PR_1$ through the lipid interface, between TM1 and TM7. TM1 is the most flexible TM of the receptor [123]. The distance between the extracellular ends of TM1 and TM7 reaches 23 Å in the cMD receptor clusters, and 33.4 Å in the aMD receptor clusters [123]. This distance creates the channel needed for the binding of the endogenous ligand [94] (Figure 5.12). Allosteric modulators that bind to this site could block the endogenous ligand from entering the site, or could prevent the endogenous ligand from dissociation. Residues in the ligand entry site are found in the N-terminus, TM1, TM2, TM3, and TM7 (Figure 5.11B). The key residues in this site are $V301^{7.43}$, $N101^{2.60}$, $E294^{7.36}$, $S105^{ECL1}$, $K34^{N\text{-}TERM}$, $Y29^{N\text{-}TERM}$, $R120^{3.28}$, $S49^{1.36}$, and $F52^{1.39}$ (Figure 5.11B). These residues interact with probes including methylamine, isopropanol, urea, cyclohexane, acetonitrile, N,N-dimethylformamide, acetaldehyde, benzaldehyde, isobutanol, acetamide, ethane, ethanol, phenol, benzene, dimethyl ether, and acetone (Figure 5.11B).

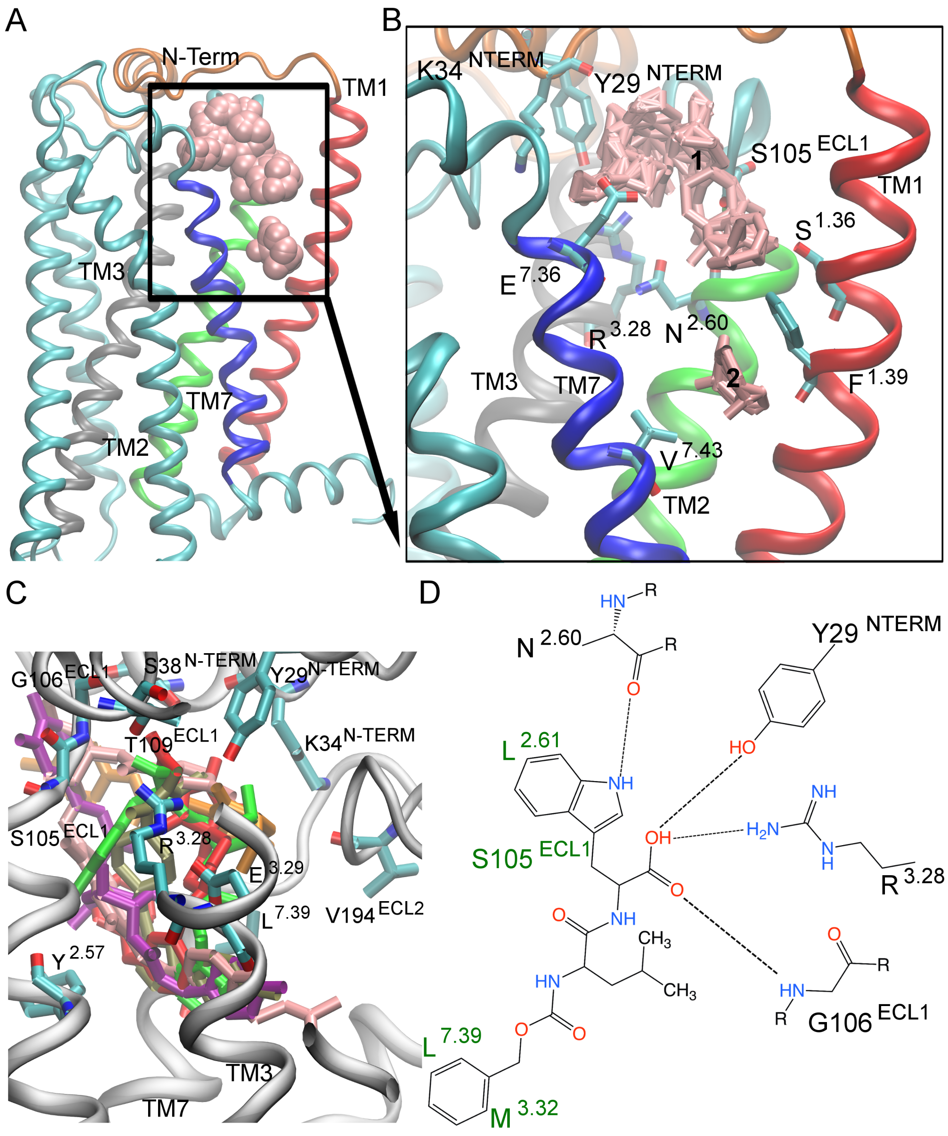

Figure 5.11: The Ligand Entry Site. The ligand entry site is located between the N-terminus (orange) and the extracellular ends of TM1 (red), TM2 (green), TM3 (gray), and TM7 (blue). The probes in the ligand entry site are shown as pink spheres [A]. The key interacting residues are shown as bonds, and the probes that bind to this site are shown as pink bonds [B]. The top six ligands, CHMBID-51157124 (red), CHMBID-98577748 (orange), CHMBID-66353657 (tan), NSC-335979 (green), NSC-91529 (purple), and NSC-227186 (pink) are docked to this site [C]. The binding pose of NSC-335979 is shown in complex with the receptor [D]. Residues in green have hydrophobic interactions with the ligand. The dotted lines represent hydrogen bonds between the residues and ligand.

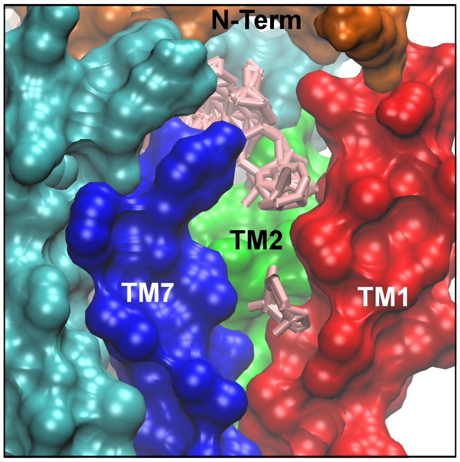

Figure 5.12: The ligand entry site between the N-terminus (orange), TM1 (red), TM2 (green), and TM7 (blue) as a surface representation is shown.

Top-Ranked Compounds that Bind to the Ligand Entry Site

The top 3 compounds from the Chembridge GPCR targeted library that dock to the ligand entry site are CHMBID-51157124 (DS_{avg}: -6.59 kcal/mol), CHMBID-98577748 (DS_{avg}: -6.58 kcal/mol), and CHMBID-66353657 (DS_{avg}: -6.58 kcal/mol). The top 3 compounds from the NCI diversity set IV are NSC-335979 (DS_{avg}: -7.68 kcal/mol), NSC-91529 (DS_{avg}: -7.61 kcal/mol), and NSC-27186 (DS_{avg}: -7.37 kcal/mol) (Figure 5.13 and Table 5.7). These compounds interact with residues in the N-terminus, TM1, TM2, TM3, and TM7 (Figure 5.11C).

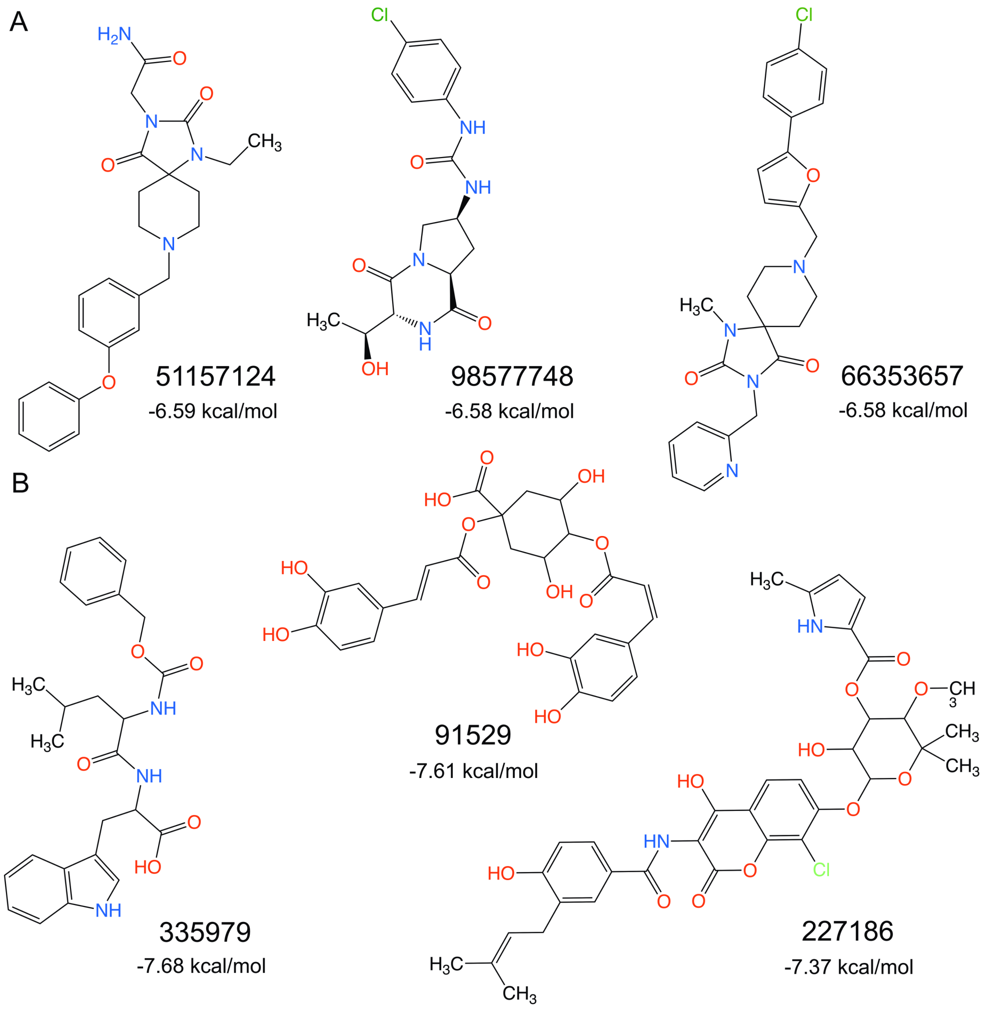

Figure 5.13: Compounds that Bind to the Ligand Entry Site. The compound structures of the top three ligands that bind to the ligand entry site from the Chembridge GPCR Library [A] (CHMBID-51157124, CHMBID-98577748, CHMBID-66353657), and the top three ligands from the NCI Diversity Set IV [B] (NSC-335979, NSC-91529, and NSC-227186) are shown. The DS$_{avg}$ for each compound to the cMD clusters are listed.

In compound NSC-335979 from the NCI Diversity Set IV, the nitrogen atom of the indol group forms a hydrogen bond with residue N101$^{2.60}$. The indol group also has hydrophobic interactions with residues L102$^{2.61}$ and S105^{ECL1}. The hydroxyl oxygen of the propanoic acid group forms hydrogen bonds with residues Y29$^{N\text{-}TERM}$ and R120$^{3.28}$. The carbonyl oxygen on NSC-335979 forms a hydrogen bond with residue G106^{ECL1}.

The phenyl group has hydrophobic interactions with residues M124$^{3.34}$ and L297$^{3.39}$ (Figure. 5.11D and Figure 5.13B). In compound NSC-91529, one carboxylic acid forms a hydrogen bond with residue Y98$^{2.57}$. There are two dihydroxyphenyl acryloyl oxy groups on the compound. One hydroxyl oxygen on a dihydroxyphenyl group forms a hydrogen bond with residue S105^{ECL1}, while another hydroxyl oxygen group on the second dihydroxyphenyl group forms a hydrogen bond with residue E121$^{3.29}$ and has hydrophobic interactions with residues M124$^{3.32}$ and L297$^{7.39}$. The carbonyl oxygen on the second dihydroxyphenyl acryloyl oxy group forms a hydrogen bond with N101$^{2.60}$ (Figure 5.11C and Figure 5.13B). In compound NSC-227168, the carbonyl oxygen of the pyrrole carboxylic acid group forms a hydrogen bond with residue G106^{ECL1}, and the phenol group has hydrophobic interactions with residue M124$^{3.32}$ and L297$^{7.39}$ (Figure 5.11C and Figure 5.13B).

Table 5.7: The top-ranking compounds from the Chembridge GPCR targeted library and the NCI diversity set IV that bind to the ligand entry site are listed. The average docking score (denoted DS_{avg}) is listed for the cMD and aMD receptor clusters. The minimum docking score (denoted DS_{min}) for the crystal structure (CS) is listed.

Compound ID	cMD DS_{avg} (kcal/mol)	aMD DS_{avg} (kcal/mol)	CS DS_{min} (kcal/mol)
Chembridge GPCR Library			
51157124	-6.59	-5.63	-6.17
98577748	-6.58	-4.79	-5.03
66353657	-6.58	-4.95	-4.65
NCI Diversity Set IV			
335979	-7.68	-5.96	-7.67
91529	-7.61	-6.61	-7.2
227186	-7.37	-6.49	-8.27

In compound CHMBID-51157124 from the Chembridge GPCR targeted library, the nitrogen atom of the amide group forms hydrogen bonds with both T109^{ECL1} and

S38$^{\text{N-TERM}}$. The carbonyl oxygen of the amide group forms hydrogen bonds with residues Y29$^{\text{N-TERM}}$ and R120$^{3.28}$. The phenyloxybenzyl group has hydrophobic interactions with residue Y98$^{2.57}$, E294$^{7.36}$, L297$^{7.39}$, and V298$^{7.40}$ (Figure 5.11C and Figure 5.13A). In compound CHMBID-98577748, the protonated nitrogen atom in the piperidine-2,5-dione group forms a hydrogen bond with residue V194$^{\text{ECL2}}$, and a carbonyl oxygen on the piperidine-2,5-dione group forms a hydrogen bond with residue Y29$^{\text{N-TERM}}$ (Figure 5.11C and Figure 5.13A). In compound CHMBID-66353657, the oxygen atoms on the imidazolidin-2-one group form hydrogen bond interactions with residues Y29$^{\text{N-TERM}}$ and R120$^{3.28}$. The pyridine group has hydrophobic interactions with residue K34$^{\text{N-TERM}}$, and the chlorobenzene group has hydrophobic interactions with residues L297$^{7.39}$, V301$^{7.43}$, and M124$^{3.32}$ (Figure 5.11C and Figure 5.13A).

Non-Orthosteric Site 3: The Intracellular Crevice

The “intracellular crevice” is located on the intracellular ends of TM1, TM2, and TM4 in the S1PR$_1$ (Figure 5.14). Non-orthosteric modulators that bind to this site could modulate the intracellular signaling pathway. W71$^{1.58}$, F77$^{\text{ICL1}}$, Y81$^{2.40}$, N86$^{2.45}$, W168$^{4.50}$, F161$^{4.43}$, and N157$^{4.39}$ are key residues involved in the intracellular crevice (Figure 5.14B). Urea, acetaldehyde, ethanol, cyclohexane, ethane, benzaldehyde, dimethyl ether, benzene, acetone, isobutanol, N,N-dimethylformamide, isopropanol, phenol, methylamine, acetonitrile, and acetamide are the probes that interact with the intracellular crevice (Figure 5.14B).

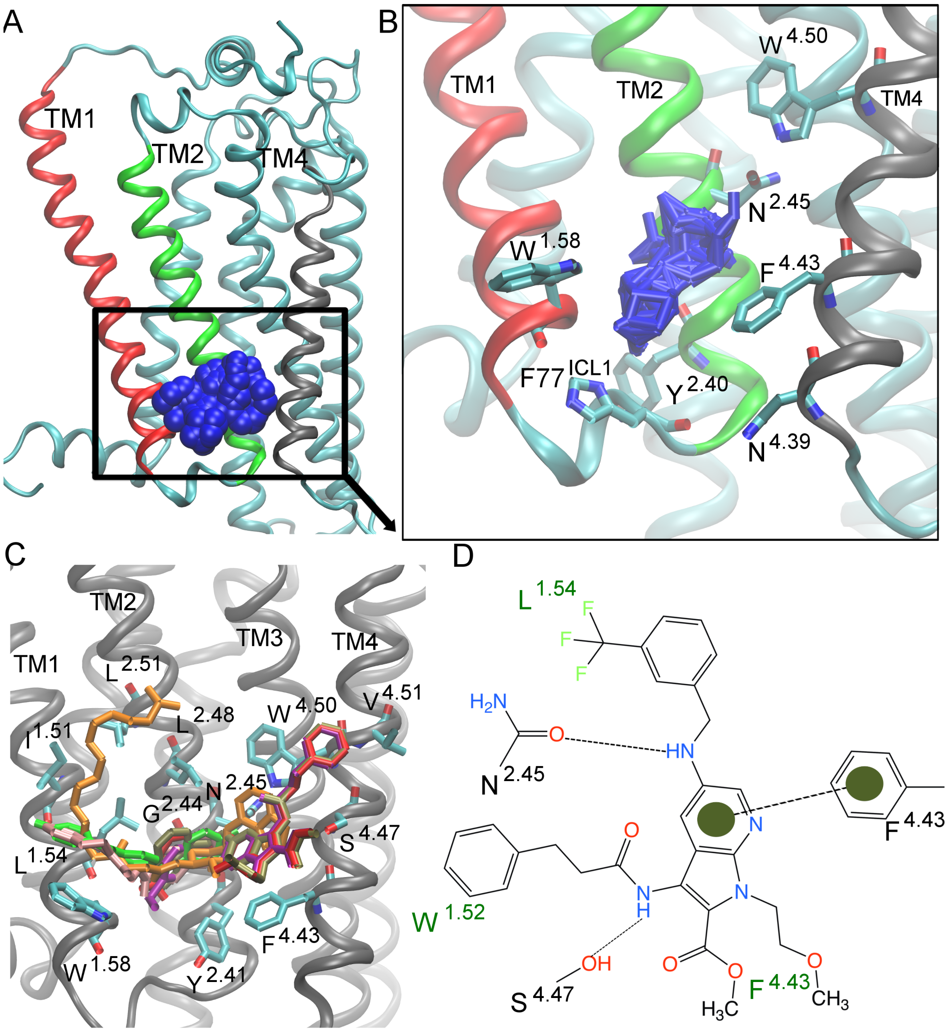

Figure 5.14: The Intracellular Crevice. The intracellular crevice is located on the lipid interface of TM1 (red), TM2 (green), and TM4 (gray). The probes in the intracellular crevice are shown as blue spheres [A]. The key interacting residues are shown as bonds, and the probes that bind to this site are shown as blue bonds [B]. The top six ligands, CHMBID-16073365 (red), CHMBID-37926360 (tan), CHMBID-37562562 (purple), NSC-87838 (pink), NSC-268251 (orange), and NSC-326757 (green) are docked to this site [C]. The binding pose of CHMBID-16073365 is shown in complex with the receptor [D]. Residues in green have hydrophobic interactions with the ligand. The dotted lines represent hydrogen bonds between the residues and ligand. The green dots represent a π-π interaction between the residue and ligand

Top-Ranked Compounds that Bind to the Intracellular Crevice

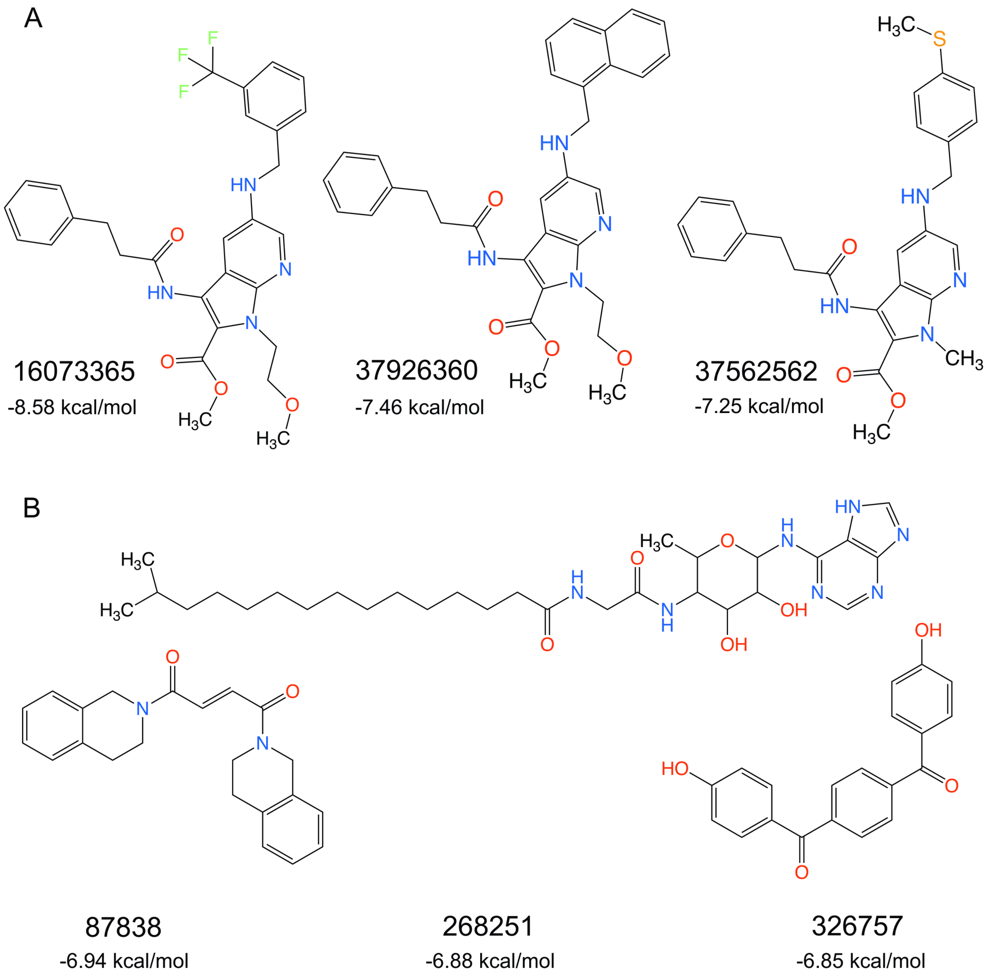

Figure 5.15: Compounds that Bind to the Intracellular Crevice. The compound structures of the top three ligands that bind to the intracellular crevice from the Chembridge GPCR targeted library [A] (CHMBID-16073365, CHMBID-37926360, CHMBID-37562562), and the top three ligands from the NCI diversity set IV [B] (NSC-87838, NSC-268251, and NSC-326757) are shown. The DS_{avg} for each compound to the cMD receptor clusters are listed.

The top 3 compounds from the Chembridge GPCR targeted library that bind to the intracellular crevice are CHMBID-16073365 (DS_{avg}: -8.58 kcal/mol), CHMBID-37926360 (DS_{avg}: -7.46 kcal/mol), and CHMBID-37562562 (DS_{avg}: -7.25 kcal/mol). The top 3 compounds from the NCI diversity set IV are NSC-87838 (DS_{avg}: -6.94 kcal/mol), NSC-268251 (DS_{avg}: -6.88 kcal/mol), and NSC-326757 (DS_{avg}: -6.85 kcal/mol) (Figure

5.15 and Table 5.8). These compounds interact with residues in the TM1, TM2, and TM4. (Figure 5.14C).

Table 5.8: The top-ranking compounds from the Chembridge GPCR targeted library and the NCI diversity set IV that bind to the intracellular crevice are shown. The average docking score (denoted DS_{avg}) is listed for the cMD and aMD receptor clusters. The minimum docking score (denoted DS_{min}) for the crystal structure (CS) is listed.

Compound ID	cMD DS_{avg} (kcal/mol)	aMD DS_{avg} (kcal/mol)	CS DS_{min} (kcal/mol)
Chembridge GPCR Library			
16073365	-8.58	-6.23	-4.74
37926360	-7.46	-5.82	-5.73
37562562	-7.25	-6.11	-5.62
NCI Diversity Set IV			
87838	-6.94	-6.59	-4.98
268251	-6.88	-6.69	-5.60
326757	-6.85	-6.38	-5.14

Compounds CHMBID-16073365, CHMBID-37926360, and CHMBID-37562562 share a common chemical core of *N*-(5-(methylamino)-3a*H*-pyrrolo[2,3-*b*]pyridin-3-yl)-3-phenylpropanamide. The pyrrolo[2,3-b]pyridine group forms π-π interactions with residue F161$^{4.43}$. The phenyl group forms hydrophobic interactions with residue W168$^{4.50}$. The nitrogen atom in the amide group of the common chemical core forms a hydrogen bond with residue S165$^{4.47}$, and the nitrogen atom in the benzylamine group forms a hydrogen bond with residue N86$^{2.45}$ (Figure 5.14D and Figure 5.15A). In compound CHMBID-16073365 from the Chembridge GPCR targeted library, the triflouromethyl group has hydrophobic interactions with residue L67$^{1.54}$. Residue F161$^{4.43}$ also has hydrophobic interactions with the carboxylate groups (Figure 5.14D and Figure 5.15A). In compound CHMBID-37926360, the methoxyethane group has hydrophobic interactions with residue F161$^{4.43}$, and the naphthalene group has hydrophobic interactions with residues L67$^{1.54}$, L89$^{2.48}$, and G85$^{2.44}$ (Figure 5.14C and Figure 5.15A). In compound CHMBID-37562562, the pyrrolo[2,3-b]pyridine also has hydrophobic interactions with residue F161$^{4.43}$. The methyl(phenyl)sulfane group has hydrophobic interactions with residue L67$^{1.54}$ (Figure 5.14C and Figure 5.15A).

In compound NSC-87838 from the NCI Diversity Set IV, the dihydroisoquinoline group has hydrophobic interactions with residues L67$^{1.54}$, N86$^{2.45}$, F161$^{4.43}$, and W168$^{4.50}$ (Figure 5.14C and Figure 5.15B). In compound NSC-268251, the purine group has hydrophobic interactions with residue I64$^{1.51}$, and a hydrogen bond with residue S165$^{4.47}$. The hydroxyl oxygen in the compound forms a hydrogen bond with residue Y82$^{2.41}$, and the hydrophobic tail group has hydrophobic interactions with residues L89$^{2.48}$, I64$^{1.51}$, L67$^{1.54}$, and L92$^{2.51}$ (Figure 5.14C and Figure 5.15B). In compound NSC-326757, one phenol oxygen forms a hydrogen bond with residue N86$^{2.45}$, and the remainder of the compound has hydrophobic interactions with residues W71$^{1.58}$ and L67$^{1.54}$ (Figure 5.14C and Figure 5.15B).

Discussion

To our knowledge, this is the first study that identified non-orthosteric sites on the receptor and screened these sites for potential orthosteric and allosteric modulators. Allosteric modulators can either enhance or diminish the activity of the endogenous ligand, by binding to sites spatially different from the orthosteric binding site [80, 81]. Additionally, allosteric sites are not as conserved as the orthosteric site and can provide the added benefit of subtype specificity within GPCR families [142, 143]. In class A GPCRs, allosteric sites have been identified in the N-terminus, ECLs, lipid bilayer, TM domains, and intracellular regions [142]. Moreover, allosteric modulators have been identified in both the $S1PR_2$ and the $S1PR_3$ [142], which are receptors in the same family as the $S1PR_1$. We identify 8 non-orthosteric sites near the N-terminus, TM domain, lipid bilayer, and the intracellular regions on the $S1PR_1$. We also identify potential modulators for the orthosteric site and non-orthosteric sites in the N-terminus, TM domain, and intracellular region.

There are two groups of known agonists and antagonists of the $S1PR_1$ mentioned in this study. One group, including S1P, ML056, FTY720-P, CHEMBL325198, and CHEMBL118860, has a charged head group and an alkyl tail, while the other group including SEW2871, CYM-5422, and CHEMBL377828, has an oxadiazol core with two attached functional groups. These ligands interact with residues in the N-terminus, TM2, ECL1, TM3, TM6, and TM7 regions. Some of the known ligands do not dock to the rigid crystal structure, suggesting a structural-based drug design approach benefits from using molecular dynamics simulations to account for receptor flexibility. Additionally, the cMD receptor clusters have more favorable docking scores to the known ligands compared to the aMD receptor clusters. aMD may allow the discovery of chemically diverse ligands that differ from known ligands.

Overall, docking the known ligands to the crystal structure, cMD receptor clusters, and aMD receptor clusters provided a range of DS_{avg} for known ligands. To the cMD receptor clusters, the DS_{avg} of the agonists ranged from -9.75 kcal/mol to -6.5 kcal/mol, while the antagonists ranged from -8.69 kcal/mol to -7.42 kcal/mol. Compared to experimental data, this range may be low. The endogenous ligand, S1P, has a dissociation constant (K_D) of 0.39, and an estimated binding energy of -12.96 kcal/mol. The DS_{min} of S1P to the crystal structure is -7.96, and the DS_{avg} of S1P to the cMD receptor clusters and the aMD receptor clusters is -7.89 kcal/mol and -5.16 kcal/mol, respectably. However, the DS_{min} of S1P to the cMD receptor clusters is -9.57 kcal/mol, and -7.37 kcal/mol in the aMD receptor clusters, which is much closer to the estimated binding energy.

A total number of 21 novel compounds were identified as binders to the orthosteric binding site. Most of the compounds had DS_{avg} scores higher than the known agonists and antagonists. These compounds interacted with TM2, TM3, TM6, and TM7, but did not interact with the N-terminus or ECLs, like the known agonists and antagonists. Eighteen of the top 21 compounds have the same chemical core, which interacted with TM3, TM6, and TM6, or the deeper regions of the binding site. These compounds had a variety of functional groups including, fluorobenzene, propylpyridine, ethylpyridine, methoxy-benzene groups, which interacted with the upper regions of the binding site, like TM2 and TM6. The diversity of these ligands is promising, and suggests that this chemical core is a stable binder to the orthosteric site.

The first non-orthosteric binding site, the sodium ion binding site, is larger in the aMD clusters than the cMD clusters. Key residues involved in the sodium ion binding site include $D91^{2.50}$. Although sodium did not bind in the previous $S1PR_1$ study [101], $D^{2.50}$ is a conserved residue found in 93 % of class A GPCRS [96]. In previous M_3 muscarinic receptor [86] and $A_{2A}AR$ simulations [125], sodium ion was observed to binds to this site. In the M_3 muscarinic receptor, sodium ion binding triggers deactivation of the receptor, and sodium cannot bind to this site when the receptor is in the active state [86, 89]. It is possible that the ligands that bind to this site could stabilize the receptor by preventing sodium binding, and promote activation of the receptor. For example, amilorides, which bind to the sodium ion binding site of the $A_{2A}AR$, accelerated the dissociation of a known inverse agonist to the receptor, promoting activation [142, 144]. N307$^{7.49}$ is a residue in the conserved NPxxY motif. In a previous S1PR$_1$ computational study, rearrangement of N307$^{7.49}$ was observed during receptor activation, and facilitated the opening of the intracellular end of the receptor [101]. In our aMD simulations of the ligand-free receptor, activation is observed. During activation of the S1PR$_1$, an internal water channel that surrounds residues Y221$^{5.58}$ and Y311$^{7.53}$ opens it to accommodate G protein binding. This is due to a conformational switch of Y311$^{7.53}$ which has interactions with D91$^{2.50}$ when the receptor is not active, and breaks these interactions to form ones with Y221$^{5.58}$ [123]. Because the sodium ion binding site opens up specifically during receptor activation, allosteric modulators that bind to this site can significantly intensify or attenuate the S1PR$_1$ signaling cascade.

Non-orthosteric site 2, the ligand entry site, is located between the extracellular region of TM1 and TM7. In the molecular dynamics simulations, the distance between the extracellular ends of these two TM reached 33.4 Å in the simulations [123]. Modulators that interact with this site have the potential to block S1P binding, or work as a biopic ligand and prolong S1P interactions with the receptor. In the IFD studies, compounds that bind to this site interact with residues in the N-terminus, and the EC loops, as well as residues in the orthosteric binding site.

Non-orthosteric site 3 is located between the intracellular ends of TM1, TM2, and TM4. Modulators that bind to this site could affect the intracellular signaling pathway of the receptor, and can be tested in both in-vitro binding and in-vivo experimental assays.

Conclusion

Using structural clusters obtained from the cMD and aMD simulations, we identified 8 non-orthosteric sites of the $S1PR_1$. The orthosteric site and three non-orthosteric sites were screened with the NCI diversity set IV and Chembridge's GPCR targeted library. Twenty-one compounds interacted with the orthosteric site of the receptor with favorable docking scores. The top six compounds for each non-orthosteric site were also identified. These compounds will serve as potential modulators of the $S1PR_1$, and we recommend that these compounds be tested for activity in the $S1PR_1$. This work can provide a valuable pathway for computer-aided drug discovery of $S1PR_1$ and other GPCRs.

Acknowledgements

This work was supported in part by the National Science Foundation (NSF grant: MCB1020765), National Institutes of Health (NIH grant: GM31749), Howard Hughes Medical Institute, National Biomedical Computation Resource (NBCR), and the National Institute of Health Training Grant (NIH Training Grant: T32GM007752).

Chapter 5 is a modified reprint of the material that will be submitted in Alisha D. Caliman, Yinglong Miao, and J. Andrew McCammon, “The Identification of Non-Orthosteric Sites and Potential Modulators for the First Sphingosine-1-Phosphate Receptor.” The dissertation author was the primary investigator and author of this paper.

APPENDIX

Convert Namd2Maestro

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
####################################################
#
# this script converts a snapshot from a NAMD trajectory
# to a Maestro input file.
#
# (C) 2010 Markus Dittrich, NRBSC, PSC, CMU
#
# call with:
#
# vmd -dispdev text -python -e convertNAMDtoMaestro.py \
#     -args -p ubq_wb.psf \
#     -c ubq_wb_eq.restart.coor -v ubq_wb_eq.restart.vel \
#     -x ubq_wb_eq.xsc -o outfile -s -S "protein"
#
# This program is distributed in the hope that it will be useful,
# but WITHOUT ANY WARRANTY; without even the implied warranty of
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
#
####################################################

import sys
import optparse
from atomsel import *
from AtomSel import AtomSel
from Molecule import Molecule
from VMD import evaltcl


def parse_cmdline(cmdlineArgs):
   """
   This function initializes the command line parser.
   """

   parser = optparse.OptionParser("Usage: vmdt -python -e "
         "readVelocities.py -args [options]")

   parser.add_option("-p", "--psffile", action="store", dest="psfFile")
   parser.add_option("-c", "--coorfile", action="store", dest="coorFile")
   parser.add_option("-v", "--velfile", action="store", dest="velFile")
```

```
    parser.add_option("-x", "--xscfile", action="store", dest="xscFile")
    parser.add_option("-o", "--outputfile", action="store", dest="outFile")
    parser.add_option("-s", "--centerSystem", action="store_true", dest="doCenter")
    parser.add_option("-S", "--centerSelection", action="store", dest="centerSel")

    parser.set_defaults(doCenter = False, centerSel = "all")

    opts, args = parser.parse_args(cmdlineArgs)
    psfFile   = opts.psfFile
    coorFile  = opts.coorFile
    velFile   = opts.velFile
    xscFile   = opts.xscFile
    outFile   = opts.outFile
    doCenter  = opts.doCenter
    centerSel = opts.centerSel

    # all filenames are required
    if (psfFile == None) or (coorFile == None) or (velFile == None) \
       or (xscFile == None) or (outFile == None):

                parser.print_help()
                exit()

    return psfFile, coorFile, velFile, xscFile, outFile, doCenter, \
           centerSel

def load_velocities(psfFile, velFile):
    """
    Load the binary velocity file and extract velocities.
    """

    mol = Molecule()
    mol.load(psfFile)
    mol.load(velFile, "namdbin")

    allVelocities = atomsel('all')
    xVel = allVelocities.get('x')
    yVel = allVelocities.get('y')
    zVel = allVelocities.get('z')

    # conversion from binvel units to A/ps
    convFactor = 20.4582651391
    xVel = [v * convFactor for v in xVel]
    yVel = [v * convFactor for v in yVel]
    zVel = [v * convFactor for v in zVel]
```

```
    mol.delete()
    return xVel, yVel, zVel

def load_system(psfFile, coorFile):
    """
    Load the main system.
    """

    mol = Molecule()
    mol.load(psfFile)
    mol.load(coorFile, "namdbin")
    return mol

def set_velocities(mol, xVel, yVel, zVel):
    """
    Add the molecule velocities to the system.
    """

    allAtoms = atomsel("all")
    allAtoms.set("vx", xVel)
    allAtoms.set("vy", yVel)
    allAtoms.set("vz", zVel)

def save_mol_as_maestro(mol, fileName):
    """
    Save the current molecule as maestro file.
    """

    mol.save(fileName + ".mae")

def set_pbc(xscFile):
    """
    Sets the systems periodic boundaries."
    """
    xscFile = open(xscFile,"r")

    for line in xscFile:
        continue

    items = line.split()
    xDim  = items[1]
    yDim  = items[5]
    zDim  = items[9]

    #set pbd
    pbcCommand = ("package require pbctools; pbc set { %s %s %s }"
            % (xDim, yDim, zDim))
```

```
    evaltcl(pbcCommand)

    xscFile.close()

def center_system(selection):
    """
    Center the system around the selection.
    """

    centerSel = atomsel(selection)
    center = centerSel.center()
    negCenter = [-1.0 * item for item in center]

    moveSel = atomsel("all")
    moveSel.moveby(negCenter)

def remove_tip3p_hh_bond():
    """
    This removes the bond between hydrogen atoms in
    TIP3P water if present since viparr will introduce
    the proper constraint.
    """

    # it looks like atomsel doesn't support set/getbonds
    # so we have to use the deprecated AtomSel for now
    oh2Sel = AtomSel("resname TIP3 and name OH2", 1)
    h1Sel  = AtomSel("resname TIP3 and name H1", 1)

    oh2Indices = oh2Sel.get("index")
    bondlist = []
    for i in oh2Indices:
        bondlist.append([i])
    h1Sel.setbonds(bondlist)

#####################################################
# main routine
#####################################################
if __name__ == "__main__":

    # parse the command line
    psfFile, coorFile, velFile, xscFile, outfile, doCenter, \
        centerSel = parse_cmdline(sys.argv[1:])

    # transform NAMD to Maestro
    vx, vy, vz = load_velocities(psfFile, velFile)
    mol = load_system(psfFile, coorFile)
```

```
if doCenter:
   center_system(centerSel)

set_velocities(mol, vx, vy, vz)
set_pbc(xscFile)
remove_tip3p_hh_bond()
save_mol_as_maestro(mol, outfile)

exit()
```

REFERENCES

1. Boomer, J.S., J.M. Green, and R.S. Hotchkiss, *The changing immune system in sepsis: is individualized immuno-modulatory therapy the answer?* Virulence, 2014. **5**(1): p. 45-56.

2. van der Poll, T. and S.M. Opal, *Host-pathogen interactions in sepsis.* Lancet Infectious Disease, 2008. **8**(1): p. 32-43.

3. Sivak, K.V., A.V. Vasin, V.V. Egorov, V.B. Tsevtkov, N.N. Kuzmich, V.A. Savina, and O.I. Kiselev, *Adenosine A2A receptor as a drug target for treatment of sepsis.* Molecular Biology, 2016. **50**(2): p. 200-212.

4. Davidson, A. and B. Diamond, *Autoimmune Diseases.* New England Journal of Medicine, 2001. **345**(5): p. 340-350.

5. Browning, V., M. Joseph, and M. Sedrak, *Multiple sclerosis: A comprehensive review for the physician assistant.* Journal of the American Academy of Physician Assistants, 2012. **25**(8): p. 24-29.

6. Trapp, B.D. and K.A. Nave, *Multiple sclerosis: an immune or neurodegenerative disorder?* Annual Review of Neuroscience, 2008. **31**: p. 247-69.

7. Dendrou, C.A., L. Fugger, and M.A. Friese, *Immunopathology of multiple sclerosis.* Nature Reviews Immunology, 2015. **15**(9): p. 545-58.

8. Sospedra, M. and R. Martin, *Immunology of multiple sclerosis.* Annual Review of Immunology, 2005. **23**: p. 683-747.

9. Spiegel, S. and S. Milstien, *The outs and the ins of sphingosine-1-phosphate in immunity.* Nature Reviews Immunology, 2011. **11**(6): p. 403-15.

10. van Noort, J.M., D. Baker, and S. Amor, *Mechanisms in the Development of Multiple Sclerosis Lesions: Reconciling Autoimmune and Neurodegenerative Factors.* CNS & Neurological Disorders- Drug Targets, 2012. **11**(5): p. 556-569.

11. Chen, J.F., H.K. Eltzschig, and B.B. Fredholm, *Adenosine receptors as drug targets--what are the challenges?* Nature Reviews Drug Discovery, 2013. **12**(4): p. 265-86.

12. Jacobson, K.A., *Introduction to adenosine receptors as therapeutic targets*, in *Adenosine Receptors in Health and Disease*, C.N. Wilson and S.J. Mustafa, Editors. 2009, Springer Berlin Heidelberg.

13. Chen, J.-F., H.K. Eltzschig, and B.B. Fredholm, *Adenosine receptors as drug targets — what are the challenges?* Nature Reviews Drug Discovery, 2013. **12**(4): p. 265-286.

14. Mills, J.H., D.G. Kim, A. Krenz, J.F. Chen, and M.S. Bynoe, *A2A adenosine receptor signaling in lymphocytes and the central nervous system regulates inflammation during experimental autoimmune encephalomyelitis.* Journal of Immunology, 2012. **188**(11): p. 5713-5722.

15. Hasko, G., J. Linden, B. Cronstein, and P. Pacher, *Adenosine receptors: therapeutic aspects for inflammatory and immune diseases.* Nature Reviews Drug Discovery, 2008. **7**(9): p. 759-770.

16. Vincenzi, F., C. Corciulo, M. Targa, S. Merighi, S. Gessi, I. Casetta, M. Gentile, E. Granieri, P.A. Borea, and K. Varani, *Multiple sclerosis lymphocytes upregulate A2A adenosine receptors that are antiinflammatory when stimulated.* European Journal of Immunology, 2013. **43**(8): p. 2206-2216.

17. Cekic, C. and J. Linden, *Purinergic regulation of the immune system.* Nature Reviews Immunology, 2016. **16**(3): p. 177-192.

18. Mills, J.H., L.F. Thompson, C. Mueller, A.T. Waickman, S. Jalkanen, J. Niemela, L. Airas, and M.S. Bynoe, *CD73 is required for efficient entry of lymphocytes into the central nervous system during experimental autoimmune encephalomyelitis.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(27): p. 9325-9330.

19. Ingwersen, J., B. Wingerath, J. Graf, K. Lepka, M. Hofrichter, F. Schroter, F. Wedekind, A. Bauer, J. Schrader, H.P. Hartung, T. Prozorovski, and O. Aktas, *Dual roles of the adenosine A2a receptor in autoimmune neuroinflammation.* Journal of Neuroinflammation, 2016. **13**: p. 48.

20. Cyster, J.G. and S.R. Schwab, *Sphingosine-1-phosphate and lymphocyte egress from lymphoid organs.* Annual Review of Immunology, 2012. **30**: p. 69-94.

21. Matloubian, M., C.G. Lo, G. Cinamon, M.J. Lesneski, Y. Xu, V. Brinkmann, M.L. Allende, R.L. Proia, and J.G. Cyster, *Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1.* Nature, 2004. **427**(6972): p. 355-360.

22. Kataoka, H., K. Sugahara, K. Shimano, K. Teshima, M. Koyama, A. Fukunari, and K. Chiba, *FTY720, sphingosine 1-phosphate receptor modulator, ameliorates experimental autoimmune encephalomyelitis by inhibition of T cell infiltration.* Cellular and Molecular Immunology, 2005. **2**(6): p. 439-448.

23. Gonzalez-Cabrera, P.J., S. Brown, S.M. Studer, and H. Rosen, *S1P signaling: new therapies and opportunities.* F1000Prime Reports, 2014. **6**(109).

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

25. Xu, F., H. Wu, V. Katritch, G.W. Han, K.A. Jacobson, Z.G. Gao, V. Cherezov, and R.C. Stevens, *Structure of an agonist-bound human A2A adenosine receptor.* Science, 2011. **332**(6027): p. 322-327.

26. Lebon, G., P.C. Edwards, A.G. Leslie, and C.G. Tate, *Molecular Determinants of CGS21680 Binding to the Human Adenosine A2A Receptor.* Molecular Pharmacology, 2015. **87**(6): p. 907-915.

27. Carpenter, B., R. Nehme, T. Warne, A.G. Leslie, and C.G. Tate, *Structure of the adenosine A(2A) receptor bound to an engineered G protein.* Nature, 2016. **536**(7614): p. 104-7.

28. Jaakola, V.-P., M.T. Griffith, M.A. Hanson, V. Cherezov, E.Y.T. Chien, J.R. Lane, Adriaan P. Ijzerman, and Raymond C. Stevens, *The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist.* Science, 2008. **322**(5905): p. 1211-1217.

29. Doré, Andrew S., N. Robertson, James C. Errey, I. Ng, K. Hollenstein, B. Tehan, E. Hurrell, K. Bennett, M. Congreve, F. Magnani, Christopher G. Tate, M. Weir, and Fiona H. Marshall, *Structure of the adenosine A2A receptor in complex with ZM241385 and the xanthines XAC and caffeine.* Structure, 2011. **19**(9): p. 1283-1293.

30. Congreve, M., S.P. Andrews, A.S. Doré, K. Hollenstein, E. Hurrell, C.J. Langmead, J.S. Mason, I.W. Ng, B. Tehan, A. Zhukov, M. Weir, and F.H. Marshall, *Discovery of 1,2,4-triazine derivatives as adenosine A2A antagonists using structure based drug design.* Journal of Medicinal Chemistry, 2012. **55**(5): p. 1898-1903.

31. Hino, T., T. Arakawa, H. Iwanari, T. Yurugi-Kobayashi, C. Ikeda-Suno, Y. Nakada-Nakura, O. Kusano-Arai, S. Weyand, T. Shimamura, N. Nomura, A.D. Cameron, T. Kobayashi, T. Hamakubo, S. Iwata, and T. Murata, *G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody.* Nature, 2012. **482**(7384): p. 237-240.

32. Liu, W., E. Chun, A.A. Thompson, P. Chubukov, F. Xu, V. Katritch, G.W. Han, C.B. Roth, L.H. Heitman, A.P. Ijzerman, V. Cherezov, and Raymond C. Stevens, *Structural Basis for Allosteric Regulation of GPCRs by Sodium Ions.* Science, 2012. **337**(6091): p. 232-236.

33. Segala, E., D. Guo, R.K. Cheng, A. Bortolato, F. Deflorian, A.S. Dore, J.C. Errey, L.H. Heitman, I.J. AP, F.H. Marshall, and R.M. Cooke, *Controlling the Dissociation of Ligands from the Adenosine A2A Receptor through Modulation*

of Salt Bridge Strength. Journal of Medicinal Chemistry, 2016. **59**(13): p. 6470-6479.

34. Hanson, M.A., C.B. Roth, E. Jo, M.T. Griffith, F.L. Scott, G. Reinhart, H. Desale, B. Clemons, S.M. Cahalan, S.C. Schuerer, M.G. Sanna, G.W. Han, P. Kuhn, H. Rosen, and R.C. Stevens, *Crystal Structure of a Lipid G Protein-Coupled Receptor.* Science, 2012. **335**(6070): p. 851-855.

35. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nature Structural & Molecular Biology, 2002. **9**(9): p. 647-652.

36. Hamelberg, D., J. Mongan, and J.A. McCammon, *Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules.* The Journal of Chemical Physics, 2004. **120**(24): p. 11919-29.

37. Wereszczynski, J. and J.A. McCammon, *Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition.* Quarterly Reviews of Biophysics, 2012. **45**(1): p. 1-25.

38. Brenke, R., D. Kozakov, G.Y. Chuang, D. Beglov, D. Hall, M.R. Landon, C. Mattos, and S. Vajda, *Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques.* Bioinformatics, 2009. **25**(5): p. 621-627.

39. Schwartz, T.W. and B. Holst, *Allosteric enhancers, allosteric agonists and ago-allosteric modulators: where do they bind and how do they act?* Trends in Pharmacological Sciences, 2007. **28**(8): p. 366-73.

40. Morelli, M., A.R. Carta, and P. Jenner, *Adenosine A2A receptors and Parkinson’s Disease*, in *Adenosine Receptors in Health and Disease*, C.N. Wilson and S.J. Mustafa, Editors. 2009, Springer Berlin Heidelberg. p. 589-615.

41. Chen, J.-F. and Y. Chern, *Impacts of methylxanthines and adenosine receptors on neurodegeneration: human and experimental studies*, in *Methylaxanthines*, B.B. Fredholm, Editor. 2011, Springer Berlin Heidelberg. p. 267-310.

42. Rasmussen, S.G.F., B.T. DeVree, Y. Zou, A.C. Kruse, K.Y. Chung, T.S. Kobilka, F.S. Thian, P.S. Chae, E. Pardon, D. Calinski, J.M. Mathiesen, S.T.A. Shah, J.A. Lyons, M. Caffrey, S.H. Gellman, J. Steyaert, G. Skiniotis, W.I. Weis, R.K. Sunahara, and B.K. Kobilka, *Crystal structure of the β2 adrenergic receptor–Gs protein complex.* Nature, 2011. **477**(7366): p. 549-555.

43. Scheerer, P., J.H. Park, P.W. Hildebrand, Y.J. Kim, N. Krauß, H.-W. Choe, K.P. Hofmann, and O.P. Ernst, *Crystal structure of opsin in its G-protein-interacting conformation.* Nature, 2008. **455**(7212): p. 497-502.

44. Kruse, A.C., A.M. Ring, A. Manglik, J. Hu, K. Hu, K. Eitel, H. Hubner, E. Pardon, C. Valant, P.M. Sexton, A. Christopoulos, C.C. Felder, P. Gmeiner, J. Steyaert, W.I. Weis, K.C. Garcia, J. Wess, and B.K. Kobilka, *Activation and allosteric modulation of a muscarinic acetylcholine receptor.* Nature, 2013. **504**(7478): p. 101-106.

45. Palczweski, K., T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamota, and M. Miyano, *Crystal structure of rhodopsin: a G protein-coupled receptor.* Science, 2000. **289**(5480): p. 739-745.

46. Haga, K., A.C. Kruse, H. Asada, T. Yurugi-Kobayashi, M. Shiroishi, C. Zhang, W.I. Weis, T. Okada, B.K. Kobilka, T. Haga, and T. Kobayashi, *Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist.* Nature, 2012. **482**(7386): p. 547-551.

47. Cherezov, V., D.M. Rosenbaum, M.A. Hanson, S.G. Rasmussen, F.S. Thian, T.S. Kobilka, H.J. Choi, P. Kuhn, W.I. Weis, B.K. Kobilka, and R.C. Stevens, *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science, 2007. **318**(5854): p. 1258-1265.

48. Dror, R.O., D.H. Arlow, P. Maragakis, T.J. Mildorf, A.C. Pan, H. Xu, D.W. Borhani, and D.E. Shaw, *Activation mechanism of the B2-adrenergic receptor.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(46): p. 18684-18689.

49. Li, J., A.L. Jonsson, T. Beuming, J.C. Shelley, and G.A. Voth, *Ligand-dependent activation and deactivation of the human adenosine A2A receptor.* Journal of the American Chemical Society, 2013. **135**(23): p. 8749-8759.

50. Pang, X., M. Yang, and K. Han, *Antagonist binding and induced conformational dynamics of GPCR A2A adenosine receptor.* Proteins, 2013. **81**(8): p. 1399-1410.

51. Miao, Y., S.E. Nichols, P.M. Gasper, V.T. Metzger, and J.A. McCammon, *Activation and dynamic network of the M2 muscarinic receptor.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(27): p. 10982-10989.

52. Leioatts, N., P. Suresh, T.D. Romo, and A. Grossfield, *Structure-based simulations reveal concerted dynamics of GPCR activation.* Proteins, 2014. **82**(10): p. 2538-2551.

53. Tehan, B.G., A. Bortolato, F.E. Blaney, M.P. Weir, and J.S. Mason, *Unifying family A GPCR theories of activation.* Pharmacol Ther, 2014. **143**(1): p. 51-60.

54. George, L., C. Arnau, and P. Leonardo, *The G-protein coupled receptor family: actors with many faces.* Current Pharmaceutical Design, 2012. **18**(2): p. 175-185.

55. Park, J.H., P. Scheerer, K.P. Hofmann, H.-W. Choe, and O.P. Ernst, *Crystal structure of the ligand-free G-protein-coupled receptor opsin.* Nature, 2008. **454**(7201): p. 183-187.

56. Oates, J. and A. Watts, *Uncovering the intimate relationship between lipids, cholesterol and GPCR activation.* Current Opinion in Structural Biology, 2011. **21**(6): p. 802-807.

57. Ng, H.W., C.A. Laughton, and S.W. Doughty, *Molecular dynamics simulations of the adenosine A2a receptor in POPC and POPE lipid bilayers: effects of membrane on protein behavior.* Journal of Chemical Information and Modeling, 2014. **54**(2): p. 573-581.

58. Vilardaga, J.-P., M. Bunemann, C. Krasel, M. Castro, and M.J. Lohse, *Measurement of the millisecond activation switch of G protein-coupled receptors in living cells.* Nature Biotechnology, 2003. **21**(7): p. 807-812.

59. Gutierrez-de-Teran, H., A. Massink, D. Rodriguez, W. Liu, G.W. Han, J.S. Joseph, I. Katritch, L.H. Heitman, L. Xia, A.P. Ijzerman, V. Cherezov, V. Katritch, and R.C. Stevens, *The role of a sodium ion binding site in the allosteric modulation of the A(2A) adenosine G protein-coupled receptor.* Structure, 2013. **21**(12): p. 2175-2185.

60. Ballesteros, J.A. and H. Weinstein, *Integrated methods for the construction of three-dimensional models and computational probing of structure-dunction relations in G protein-coupled receptor.* Methods in Neurosciences, 1995. **25**: p. 366-428.

61. Eswar, N., B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.-y. Shen, U. Pieper, and A. Sali, *Comparative protein structure modeling using MODELLER.* Current Protocols in Bioinformatics, 2007.

62. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* The Journal of Molecular Graphics and Modelling, 1996. **14**(1): p. 33-38.

63. Gumbart, J., L.G. Trabuco, E. Schreiner, E. Villa, and K. Schulten, *Regulation of the protein-conducting channel by a bound ribosome.* Structure, 2009. **17**(11): p. 1453-1464.

64. Li, H., A.D. Robertson, and J.H. Jensen, *Very fast empirical prediction and rationalization of protein pKa values.* Proteins, 2005. **61**(4): p. 704-721.

65. Klauda, J.B., R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell, and R.W. Pastor, *Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types.* The Journal of Physical Chemistry B, 2010. **114**(23): p. 7830-7843.

66. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, and K. Schulten, *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry, 2005. **26**(16): p. 1781-1802.

67. MacKerell, A.D., M. Feig, and C.L. Brooks, 3rd, *Improved treatment of the protein backbone in empirical force fields.* Journal of the American Chemical Society, 2004. **126**(3): p. 398-699.

68. Mackerell, A.D., B. D., M. Bellott, D.R.L. Jr., J.D. Evanseck, M.H. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, S. J., M. Watanabe, J. Wiorkieqicz-Kuczera, D. Yin, and M. Karplus, *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* The Journal of Chemical Physics, 1998. **102**(18): p. 3586-3616.

69. Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen, *A smooth particle mesh Ewald method.* The Journal of Chemical Physics, 1995. **103**(19): p. 8577-8577.

70. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* The Journal of Computational Physics, 1997. **23**(3): p. 327-341.

71. Shaw, D.E., R.O. Dror, J.K. Salmon, J.P. Grossman, K.M. Mackenzie, J.A. Bank, C. Young, M.M. Deneroff, B. Baston, K.K. Bowers, E. Chow, M.P. Eastwood, D.J. Ierardi, J.L. Klepeis, J.S. Kuskin, R.H. Larson, K. Lindorff-Larsen, P. Maragakis, M.A. Moraes, S. Piana, Y. Shan, and B. Towles. *Millisecond-scale molecular dynamics simulations on Anton*. in *In: Proceedings of the 2009 ACM/IEEE Conference on Supercomputing (SC09)*. 2009. Portland, OR: ACM Press.

72. Krautles, V., W.F. Van Gunsteren, and P.H. Hunenberger, *A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamic Simulations.* Journal of Computational Chemistry, 2001. **22**(5): p. 501-508.

73. Shan, Y., J.L. Klepeis, M.P. Eastwood, R.O. Dror, and D.E. Shaw, *Gaussian split Ewald: A fast Ewald mesh method for molecular simulation.* The Journal of Chemical Physics, 2005. **122**(5): p. 54101.

74. Roux, B., *The calculation of the potential of mean force using computer simulations.* Computer Physics Communications, 1995. **91**(1-3): p. 275-282.

75. Thiel, M., C.C. Caldwell, and M.V. Sitkovsky, *The critical role of adenosine A2A receptors in downregulation of inflammation and immunity in the pathogenesis of infectious diseases.* Microbes and Infection, 2003. **5**(6): p. 515-526.

76. Pinna, A., *Adenosine A2A receptor antagonists in Parkinson's disease: progress in clinical trials from the newly approved istradefylline to drugs in early development and those already discontinued.* CNS Drugs, 2014. **28**(5): p. 455-474.

77. Ohta, A. and M. Sitkovsky, *Role of G-protein-coupled adenosine receptors in downregulation of inflamation and protetion from tissue damage.* Nature, 2001. **414**(6866): p. 916-920.

78. Jacobson, K.A. and Z.G. Gao, *Adenosine receptors as therapeutic targets.* Nature Reviews Drug Discovery, 2006. **5**(3): p. 247-264.

79. Ohta, A., E. Gorelik, S.J. Prasad, F. Ronchese, D. Lukashev, M.K.K. Wong, X. Huang, S. Caldwell, K. Liu, P. Smith, J.-F. Chen, E.K. Jackson, S. Apasov, S. Abrams, and M. Sitkovsky, *A2A adenosine receptor protects tumors from antitumor T cells.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(35): p. 13132-13137.

80. Conn, P.J., A. Christopoulos, and C.W. Lindsley, *Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders.* Nature Reviews Drug Discovery, 2009. **8**(1): p. 41-54.

81. Goblyos, A. and A.P. Ijzerman, *Allosteric modulation of adenosine receptors.* Biochimica et Biophysica Acta, 2011. **1808**(5): p. 1309-1318.

82. Caliman, A.D., S.E. Swift, Y. Wang, Y. Miao, and J.A. McCammon, *Investigation of the conformational dynamics of the apo A2A adenosine receptor.* Protein Science, 2015. **24**(6): p. 1004-1012.

83. Ivetac, A. and J.A. McCammon, *Mapping the druggable allosteric space of G-protein coupled receptors: a fragment-based molecular dynamics approach.* Chemical Biology & Drug Design, 2010. **76**(3): p. 201-217.

84. Miao, Y., S.E. Nichols, and J.A. McCammon, *Mapping of allosteric druggable sites in activation-associated conformers of the M2 muscarinic receptor.* Chemical Biology & Drug Design, 2014. **83**(2): p. 237-246.

85. Jacobson, M.P., D.L. Pincus, C.S. Rapp, T.J. Day, B. Honig, D.E. Shaw, and R.A. Friesner, *A hierarchical approach to all-atom protein loop prediction.* Proteins, 2004. **55**(2): p. 351-367.

86. Miao, Y., A.D. Caliman, and J.A. McCammon, *Allosteric effects of sodium ion binding on activation of the m3 muscarinic g-protein-coupled receptor.* Biophysical Journal, 2015. **108**(7): p. 1796-1806.

87. Olivella, M., G. Caltabianco, and A. Cordomí, *The role of Cysteine 6.47 in class A GPCRs.* BMC Structural Biology, 2013. **13**(3): p. 1-10.

88. Yuan, S., J. Hu, S. Filipek, and H. Vogel, *W246(6.48) Opens a Gate for a Continuous Intrinsic Water Pathway during Activation of the Adenosine A2A Receptor.* Angewandte Chemie International Edition, 2015. **54**(2): p. 556-559.

89. Katritch, V., G. Fenalti, E.E. Abola, B.L. Roth, V. Cherezov, and R.C. Stevens, *Allosteric sodium in class A GPCR signaling.* Trends in Biochemical Sciences, 2014. **39**(5): p. 233-244.

90. Rosen, H., M. Germana Sanna, P.J. Gonzalez-Cabrera, and E. Roberts, *The organization of the sphingosine 1-phosphate signaling system.* Current Topics in Microbiology and Immunology, 2014. **378**: p. 1-21.

91. Brinkmann, V., *FTY720 (fingolimod) in Multiple Sclerosis: therapeutic effects in the immune and the central nervous system.* British Journal of Pharmacology, 2009. **158**(5): p. 1173-1182.

92. Schwab, S.R., J.P. Pereira, M. Matloubian, Y. Xu, Y. Huang, and J.G. Cyster, *Lymphocyte Sequestration Through S1P Lyase Inhibition and Disruption of S1P Gradients.* Science, 2005. **309**(5741): p. 1735-1739.

93. Lo, C.G., Y. Xu, R.L. Proia, and J.G. Cyster, *Cyclical modulation of sphingosine-1-phosphate receptor 1 surface expression during lymphocyte recirculation and relationship to lymphoid organ transit.* The Journal of Experimental Medicine, 2005. **201**(2): p. 291-301.

94. Hanson, M.A., C. Roth, E. Jo, M. Griffith, F. Scott, G. Reinhart, H. Desale, B. Clemons, S. Cahalan, S. Schuerer, M. Sanna, G. Han, P. Kuhn, H. Rosen, and R.C. Stevens, *Crystal Structure of a Lipid G Protein-Coupled Receptor.* Science, 2012. **335**(6070): p. 851-855.

95. Huang, W., A. Manglik, A.J. Venkatakrishnan, T. Laeremans, E.N. Feinberg, A.L. Sanborn, H.E. Kato, K.E. Livingston, T.S. Thorsen, R.C. Kling, S. Granier, P. Gmeiner, S.M. Husbands, J.R. Traynor, W.I. Weis, J. Steyaert, R.O. Dror, and B.K. Kobilka, *Structural insights into mu-opioid receptor activation.* Nature, 2015. **524**(7565): p. 315-321.

96. Tehan, B.G., A. Bortolato, F.E. Blaney, M.P. Weir, and J.S. Mason, *Unifying family A GPCR theories of activation.* Pharmacology & Therapeutics, 2014. **143**: p. 51- 60.

97. Cvicek, V., W.A. Goddard, 3rd, and R. Abrol, *Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications.* PLoS Computational Biology, 2016. **12**(3): p. e1004805.

98. Omotuyi, O. and H. Ueda, *Inter-helical Electrostatic Interaction and Rotameric Signatures in Activated Sphingosine-1-Phosphate Receptor 1.* Chemical Informatics, 2014. **1**(1:2).

99. Yuan, S., S. Filipek, K. Palczewski, and H. Vogel, *Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway.* Nature Communications, 2014. **5**: p. 4733.

100. Yuan, S., Z. Hu, S. Filipek, and H. Vogel, *W246(6.48) opens a gate for a continuous intrinsic water pathway during activation of the adenosine A2A receptor.* Angewandte Chemie International Edition, 2015. **54**(2): p. 556-559.

101. Yuan, S., R. Wu, D. Latek, B. Trzaskowski, and S. Filipek, *Lipid Receptor S1P1 Activation Scheme Concluded from Microsecond All-Atom Molecular Dynamics Simulations.* PLoS Computational Biology, 2013. **9**(10): p. e1003261.

102. Waters, C.M., J. Long, I. Gorshkova, Y. Fujiwara, M. Connell, K.E. Belmonte, G. Tigyi, V. Natarajan, S. Pyne, and N.J. Pyne, *Cell migration activated by platelet-derived growth factor receptor is blocked by an inverse agonist of the sphingosine 1-phosphate receptor-1.* The FASEB Journal, 2006. **20**(3): p. 509-511.

103. Hamelberg, D., J. Mongan, and J.A. McCammon, *Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules.* The Journal of Chemical Physics, 2004. **120**(24): p. 11919-11929.

104. Wang, Y., C.B. Harrison, K. Schulten, and J.A. McCammon, *Implementation of Accelerated Molecular Dynamics in NAMD.* Computational Science & Discovery, 2011. **4**(1): p. 015002.

105. Miao, Y. and J.A. McCammon, *G-protein coupled receptors: advances in simulation and drug discovery.* Current Opinion in Structural Biology, 2016. **41**: p. 83-89.

106. Nygaard, R., T.M. Frimurer, B. Holst, M.M. Rosenkilde, and T.W. Schwartz, *Ligand binding and micro-switches in 7TM receptor structures.* Trends in Pharmacological Sciences, 2009. **30**(5): p. 249-259.

107. Angel, T.E., S. Gupta, B. Jastrzebska, K. Palczewski, and M.R. Chance, *Structural waters define a functional channel mediating activation of the GPCR,*

rhodopsin. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(34): p. 14367-14372.

108. Prakash, P., A. Sayyed-Ahmad, and A.A. Gorfe, *The Role of Conserved Waters in Conformational Transitions of Q61H K-ras.* PLoS Computational Biology, 2012. **8**(2): p. e1002394.

109. Hurst, D.P., A. Grossfield, D.L. Lynch, S. Feller, T.D. Romo, K. Gawrisch, M.C. Pitman, and P.H. Reggio, *A lipid pathway for ligand binding is necessary for a cannabinoid G protein-coupled receptor.* The Journal of Biological Chemistry, 2010. **285**(23): p. 17954-17964.

110. Hildebrand, P.W., P. Scheerer, J.H. Park, H.W. Choe, R. Piechnick, O.P. Ernst, K.P. Hofmann, and M. Heck, *A ligand channel through the G protein coupled receptor opsin.* PLoS One, 2009. **4**(2): p. e4382.

111. Neale, C., H.D. Herce, R. Pomes, and A.E. Garcia, *Can Specific Protein-Lipid Interactions Stabilize an Active State of the Beta 2 Adrenergic Receptor?* Biophysical Journal, 2015. **109**(8): p. 1652-1662.

112. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics and Modelling, 1996. **14**(1): p. 33-38.

113. Vanommeslaeghe, K., E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A.D. Mackerell, Jr., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.* Journal of Computational Chemistry, 2010. **31**(4): p. 671-690.

114. Lange, O.F. and H. Grubmuller, *Generalized correlation for biomolecular dynamics.* Proteins, 2006. **62**(4): p. 1053-1061.

115. Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J. Berendsen, *GROMACS: fast, flexible, and free.* Journal of Computational Chemistry, 2005. **26**(16): p. 1701-1718.

116. Pereira, J.P., J.G. Cyster, and Y. Xu, *A Role for S1P and S1P1 in Immature-B Cell Egress from Mouse Bone Marrow.* PLoS One, 2009. **5**(2): p. e9277.

117. Cahalan, S.M., P.J. Gonzalez-Cabrera, N. Nguyen, M. Guerrero, E.A. Cisar, N.B. Leaf, S.J. Brown, E. Roberts, and H. Rosen, *Sphingosine 1-phosphate receptor 1 (S1P(1)) upregulation and amelioration of experimental autoimmune encephalomyelitis by an S1P(1) antagonist.* Molecular Pharmacology, 2013. **83**(2): p. 316-321.

118. Sheridan, G.K. and K.K. Dev, *S1P1 receptor subtype inhibits demyelination and regulates chemokine release in cerebellar slice cultures.* Glia, 2012. **60**(3): p. 382-392.

119. Zhang, F., Y. Xia, W. Yan, H. Zhang, F. Zhou, S. Zhao, W. Wang, D. Zhu, C. Xin, Y. Lee, L. Zhang, Y. He, E. Gao, and L. Tao, *Sphingosine 1-phosphate signaling contributes to cardiac inflammation, dysfunction, and remodeling following myocardial infarction.* American Journal of Physiology Heart and Circulatory Physiology, 2016. **310**(2): p. H250-H261.

120. Watson, C., J.S. Long, C. Orange, C.L. Tannahill, E. Mallon, L.M. McGlynn, S. Pyne, N.J. Pyne, and J. Edwards, *High expression of sphingosine 1-phosphate receptors, S1P1 and S1P3, sphingosine kinase 1, and extracellular signal-regulated kinase-1/2 is associated with development of tamoxifen resistance in estrogen receptor-positive breast cancer patients.* American Journal of Pathology, 2010. **177**(5): p. 2205-2215.

121. Graler, M.H., *The role of sphingosine 1-phosphate in immunity and sepsis.* American Journal of Clinical and Experimental Immunology, 2012. **1**(2): p. 90-100.

122. Jackson, S.J., G. Giovannoni, and D. Baker, *Fingolimod modulates microglial activation to augment markers of remyelination.* Journal of Neuroinflammation, 2011. **8**(76).

123. Caliman, A.D., Y. Miao, and J.A. McCammon, *Activation Mechanisms of the First Sphingosine-1-Phosphate Receptor.* Protein Science, 2017(In Press).

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

125. Caliman, A.D., Y. Miao, and J.A. McCammon, *Mapping the Allosteric Sites of the A2A Adenosine Receptor.* Chemical Biology & Drug Design, 2017. **Accepted Upon Minor Revisions**.

126. Friesner, R.A., J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Farancis, and P.S. Shenkin, *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy.* Journal of Medicinal Chemistry, 2004. **47**(7): p. 1739-1749.

127. Sherman, W., H.S. Beard, and R. Farid, *Use of an induced fit receptor structure in virtual screening.* Chemical Biology & Drug Design, 2006. **67**(1): p. 83-84.

128. *Induced Fit Docking Protocol*, in *Glide*. Schrödinger, LLC: New York, NY.

129. Sherman, W., T. Day, M.P. Jacobson, R.A. Friesner, and R. Farid, *Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects.* Journal of Medicinal Chemistry, 2006. **49**(2): p. 534-553.

130. Farid, R., T. Day, R.A. Friesner, and R.A. Pearlstein, *New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies.* Bioorganic & Medicinal Chemistry, 2006. **14**(9): p. 3160-3173.

131. Miao, Y., D. Goldfeld, E. Moo, P.M. Sexton, A. Christopoulos, J.A. McCammon, and C. Valant, *Accelerated structure-based design of chemically diverse allosteric modulators of a muscarinic G protein-coupled receptor.* Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(38): p. E5675–E5684.

132. Bresso, E., R. Togawa, K. Hammond-Kosack, M. Urban, B. Maigret, and N.F. Martins, *GPCRs from fusarium graminearum detection, modeling and virtual screening - the search for new routes to control head blight disease.* BMC Bioinformatics, 2016. **17**(Supplement 18): p. 463.

133. Chan, W.K., H. Zhang, J. Yang, J.R. Brender, J. Hur, A. Ozgur, and Y. Zhang, *GLASS: a comprehensive database for experimentally validated GPCR-ligand associations.* Bioinformatics, 2015. **31**(18): p. 3035-3042.

134. Repasky, M.P., R.B. Murphy, J.L. Banks, J.R. Greenwood, I. Tubert-Brohman, S. Bhat, and R.A. Friesner, *Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide.* Journal of Computer-Aided Molecular Design, 2012. **26**(6): p. 787-799.

135. Huang, N., B.K. Shoichet, and J.J. Irwin, *Benchmarking Sets for Molecular Docking.* Journal of Medicinal Chemistry, 2006. **49**(23): p. 6789-6801.

136. Mysinger, M.M., M. Carchia, J.J. Irwin, and B.K. Shoichet, *Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking.* Journal of Medicinal Chemistry, 2012. **55**(14): p. 6582-94.

137. *LigPrep*. Schrödinger, LLC: New York.

138. Halgren, T.A., R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard, and J.L. Banks, *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening.* Journal of Medicinal Chemistry, 2004. **47**(7): p. 1750-1759.

139. *Glide*. Schrödinger, LLC: New York.

140. Friesner, R.A., R.B. Murphy, M.P. Repasky, L.L. Frye, J.R. Greenwood, T.A. Halgren, P.C. Sanschagrin, and D.T. Mainz, *Extra Precision Glide: Docking and*

Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. Journal of Medicinal Chemistry, 2006. **49**(21): p. 6177-6196.

141. Nichols, S.E., R. Baron, and J.A. McCammon, *On the use of molecular dynamics receptor conformations for virtual screening*, in *Computational Drug Discovery and Design*, R. Baron, Editor. 2012. p. 93-103.

142. Gentry, P.R., P.M. Sexton, and A. Christopoulos, *Novel Allosteric Modulators of G Protein-coupled Receptors.* Journal of Biological Chemistry, 2015. **290**(32): p. 19478-19488.

143. Changeux, J.P. and A. Christopoulos, *Allosteric Modulation as a Unifying Mechanism for Receptor Function and Regulation.* Cell, 2016. **166**(5): p. 1084-1102.

144. Gao, Z. and A. Ijzerman, *Allosteric Modulation of A2A Adenosine Receptors by Amiloride Analogues and Sodium Ions.* Biochemical Pharmacology, 2000. **60**(5): p. 669-676.