Understanding the dynamics and molecular interactions of the human cannabinoid receptor 1

by

Maya Petgrave

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Pharmacy

Waterloo, Ontario, Canada 2022

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

ABSTRACT

Cannabinoid receptor 1 (CB_1), one of the most populated G-protein coupled receptors (GPCR) present within the brain, is involved in various physiological and homeostatic processes. These pathways include metabolism and hunger, locomotion, memory processing, and the onset of psychoactive effects. In recent decades, it has been uncovered that (both endogenous and exogenous) ligands and protein-CB_1 interactions can elicit specific cellular pathways via biased intracellular signalling. Biased signalling is commenced by inducing various receptor conformations through protein/ligand-based interactions. These receptor conformations promote CB_1 coupling with different G-proteins. Coupling to various G-proteins mediates potential therapeutic or adverse side effects. However, the full details on how CB_1 controls these pathways and produces its psychotropic effects (upon cannabis consumption) are not fully understood. Numerous attempts have been conducted to comprehend the CB_1 receptor and its signalling mechanism. Unfortunately, due to the complexity and novelty of the field, many knowledge gaps remain.

Another prominent issue in the cannabinoid field is the lack of diverse and complete crystallographic structures. To date, the isolation of a complete human crystallographic CB_1 structure—free from alterations (i.e., truncations, mutations, and insertion of stabilizing domains)—has not been solved. Moreover, there's a limited crystallographic database on the reported CB_1 interacting proteins, thus further restricting the accuracy of computational studies. For these studies, the ligand-receptor and protein-protein binding interactions and their adopted conformations under the influence of the complete receptor are unknown. These aspects further shroud the cannabinoid field in mystery and limit its clinical utility.

However, advancements in crystallographic and computational techniques it has facilitated a novel avenue in the generation of more accurate three-dimensional (3D) computational models and results. This MSc thesis project addressed two major aims – **i)** developing a comprehensive atomistic model of the human CB_1 receptor and investigating its effects on ligand binding, in addition to **ii)** modelling the interactions of the human CB_1 receptor with one of its protein binding partners using computational tools.

To investigate the impacts of ligand-CB_1 binding interactions, the first complete atomistic model of the human CB_1 receptor was built. Three chemically similar ligands were docked to the receptor and subjected to classical molecular dynamic (MD) simulations and binding-free energy

calculations. The role of the N-terminus in orthosteric ligand binding was observed and determined for each ligand. These examinations postulated the potential importance of incorporating the complete receptor in computer-aided drug design (CADD) studies.

Moreover, the recently published first mammalian cannabinoid interacting protein 1a (CRIP1a) crystallographic structure provided a productive template (~96% identity with that of the human CRIP1a protein) for this study. Construction of the complete activated human CB_1 receptor and CRIP1a models' interactions were investigated utilizing docking approaches. The complexes were subjected to classical MD simulations and binding-free energy calculations that allowed for subsequent ranking of the best complex. Residues involved in the molecular recognition process between human CB_1 and CRIP1a were all confined within the experimentally determined binding region, thus validating the model.

Overall, this thesis supports previous experimental and computational findings on CRIP1a-CB_1 binding interactions and the role of the N-terminus in protein-ligand binding. These results demonstrate the significance of modelling the complete human CB_1 receptor to comprehend the protein-protein and protein-ligand interactions within the receptor. Outcomes we can extract from this thesis are **i)** the entire N-terminal segment is necessary for CADD, docking, and affinity-binding predictions studies. Furthermore, we concluded that **ii)** the residues involved in the molecular recognition process between that of CRIP1a and CB_1 are all within the confines of the experimentally determined regions (residues 34-110 for CRIP1a, and residues within the distal nine amino acids in CB_1). Outcomes from this thesis can aid when driving future mutational *in vitro* studies or designing small molecules to target the CRIP1a-CB_1 interface. These small molecules can be significant in the generation of anti-nociceptive therapeutical drugs.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors Dr. Aravindhan Ganesan and Dr.Praveen Nekkar Rao, for offering me advice and supporting me throughout my Masters research. Thank you for spending time to teach me different valuable aspects that I will take away for the rest of my life, beyond my Master's degree, and for your continued time and patience spent helping me when conducting my research. This thesis wouldn't have been possible without the both of you. I would also like to thank my supervisor committee members Dr. Micheal Beazely for sending me valuable resources and for your continued support throughout my degree, in addition to Dr. William Wong, who has helped me reassess my methodology, through his advice, helping me to think more critically, and become a better researcher.

I would also like to thank all the current and previous members of ArGan's lab, who have helped me throughout conducting my research and have given me continued support, you have been a valuable aspect in my growth as teacher, researcher, and a student. This thesis wouldn't have been possible without the input and lessons taught from all of you!

I also could not have completed this thesis without the love and support of my family: Mom, Keneisha, Chantal, and my beloved brother Miguel. You all have been there for me, helping me out through all the hard times, and have provided a lot for during my years of study. I honestly wouldn't have been who I am without your encouragement and support, thank you! I would also like to thank all my close friends who have been there for me as well! Your support, gave me the confidence and strength I needed to push on and complete my research, thank you as well!

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

LIST OF TABLES

Chapter 3:

Chapter 4:

LIST OF ABBREVIATIONS

PDB	Protein databank
2-AG	2 -Arachidonoylglycerol
AEA	Anandamide
MAGL	Monoacylglycerol lipase
FAAH	Fatty acid amide hydrolase
MD	Molecular dynamics
PCA	Principal component analysis
CB_1	Cannabinoid (receptor) 1
CB_2	Cannabinoid (receptor) 2
MMPB(GB)SA	Molecular Mechanics Poisson-Boltzmann (generalized Born) Surface Area
GPCRs	G-protein coupled receptors
MPR	Membrane proximal region
CBD	Cannabidiol
Δ^9-THC	Δ-^{9}-Tetrahydrocannabinol
AMBER	Assisted Model Building with Energy Refinement
NAMD	Nanoscale Molecular Dynamics
ITASSER	Iterative Threading ASSEmbly Refinement
DOPE	Discrete Optimized Protein Energy
VMD	Visual Molecular Dynamics
pSF	pseudo-F statistic
DBI	Davies-Bouldin Index
PC	Principal component
CaFE	Calculation of free energy
CBRs	Cannabinoids receptors
ERK1/2	Extracellular signal -regulated kinases type 1 and 2

GABA	γ-aminobutyric acid
TRPV1	Transient receptor potential vanilloid 1
SR141716A	Rimonabant
DSI/DSE	depolarization-induced suppression of inhibition /excitation
DGLα/ DGLβ	*sn*-1-specific diacylglycerol lipase-α/β
NAPE-PLD	*N*-acyl-phosphatidylethanolamine-hydrolysing phospholipase D
LTD	long-term depression
TM	Transmembrane
ECL	Extracellular loop
ICL	Intracellular loop
NTH	N-Terminal Helix
MAPK	Mitogen-activated protein kinase
cAMP	Cyclic AMP
ps/ns	Picoseconds/Nanoseconds

CHAPTER 1

Introduction to the CB_1 Receptor and its Physiological and Pathophysiological Roles

1.1 The Endocannabinoid System

The psychedelic effects mediated by *Cannabis sativa* are largely due to Δ^9-tetrahydrocannabinol (Δ^9-THC), the major psychoactive constituent within the plant, which mediates its effects within the endocannabinoid system. [1–3] Numerous attempts for the isolation of the active constituents in cannabis were carried out for over a century.[4] It wasn't until 1964[1], that Δ^9-THC was finally extracted, due to the advance in modern separation techniques.[5] With the discovery and elucidation of the first chemical constituents of cannabis, Δ^9-THC, it became widely available for research in the mid-1960s, when it was believed that the medicinal activity of the plant could be separated from its psychotropic effects.[4,5] For these reasons, most efforts were dedicated to understanding the mechanism of action of Δ^9-THC, and its potential toxicological and addictive effects.[5–9] Studies on Δ^9-THC[11–13] used its structure as a prototype in the discovery of various synthetic analogues and novel compounds such as CP-55,940[14], which then led to the detection of the main constituents of the endocannabinoid system, the cannabinoid receptors (CBRs): cannabinoid receptor 1 (CB_1) and cannabinoid receptor 2 (CB_2), within the brain[15] and spleen[16], respectively. The identification of the CBRs and their chemically derived chemical modulators suggested that endogenous molecules stimulate (agonists) or inhibit (antagonists) the receptors present in the body, thus leading to the isolation of the naturally occurring cannabinoids anandamide (AEA) and 2-arachidonoylglycerol (2-AG), later termed the endocannabinoids.[5,17–19] The endocannabinoid system, is composed of the G-protein coupled CBRs, the endocannabinoids anandamide (AEA) and 2-arachidonoylglycerol (2-AG), as well as the enzymes that synthesize and degrade endocannabinoids. The enzymes that are part of the endocannabinoid system include *N*-acyl-phosphatidylethanolamine-hydrolysing phospholipase D (NAPE-PLD), *sn*-1-specific diacylglycerol lipase-α (DGLα), DGLβ, fatty acid amide hydrolase 1 (FAAH) and monoacylglycerol lipase (MAGL; also known as MGL). Therefore, the efforts to understand the pharmacological actions of Δ^9-THC led to a wealth of insights into the physiological roles of the endocannabinoid system.[5] Mainly, the endocannabinoid system was found to be involved in the homeostatic balance of the nervous and immune systems, as well as many other organ systems within the body.[3] Endocannabinoids were found to be responsible for regulating the neurotransmitter release through an interplay between the different components within the endocannabinoid system, thus allowing for the system to be conceived as a drug target for

regulating pathogenic diseases and conditions.[20] Throughout various *in vitro* and *in vivo* models, the endocannabinoid system was confirmed to play a role in various homeostatic roles such as memory processing[21,22], hunger[23], locomotor coordination[24,25] and the maintain a role in the inception of various psychological[26,27] and physiological[28–30] disorders. Alterations in endocannabinoid signalling, are caused due to changes in the expression and function of the CBRs, as well as the concentration of endocannabinoid enzymes. These changes are associated with various pathological conditions.[5] Thus, therapies that exploit or correct such alterations, may be developed from the modulators of the CBRs.[5] The road to the clinical development of safer synthetic cannabinoids has yet to be fruitful in its discoveries, due to the lack of understanding of the receptor and atomistic interactions with its ligands.[5]

1.2 Cannabinoid Receptor Type-1

The unearthing of the CB_1 and CB_2 receptors in 1990[15] and 1993[16], respectively, led to the uncovering of their effects within various parts of the body. CB_2 is located within the peripheral cells, particularly immune cells, and involves in immunomodulation. Conversely, CB_1 is predominantly located within the central nervous system (CNS) and thus is associated with the psychedelic effects of Δ^9-THC. Uniquely, a low concentration of CB_1 is also expressed within various peripheral organs, where it partakes in a wealth of roles including in gastro-intestinal mobility[31], prenatal development[32–34], and metabolism[23]. Likewise, the type-2 receptor (i.e., CB_2) was also found to be expressed in the CNS in small numbers, where it involves the regulation of neuroinflammation and increases the instances of brain injury.[35] However, due to the predominant roles of CB_1 in CNS and its association with Δ^9-THC-mediated psychoactive effects, more studies have focused on understanding the structure and regulation of CB_1.

CB_1, a 472 amino acid long transmembrane (TM) receptor, encoded by the CNR1 gene, with a 48% sequence identity with that of its counterpart, CB_2, is the most abundant GPCR within the brain.[36] CB_1 maintains its highest density within the basal ganglia, substantia nigra, globus pallidus, cerebellum, and hippocampus of rodent brains.[4] It displays moderate expression in the cerebral cortex, septum, amygdala, hypothalamus, and parts of the brainstem and dorsal horn of the spinal cord, where it can regulate the sensation of hunger. Whereas regions such as the thalamus and ventral horn of the spinal cord have low expression of CB_1. The CB_1 receptor is also expressed to a lower extent in astrocytes, oligodendrocytes, and microglia, where it has been shown to mediate

synaptic transmission. These regions of high abundancy are presented within the sensory and motor regions which are consistent with the principal role of CB_1 receptors in motivation and cognition.[4,37] Also present within the peripheral nervous system (PNS), CB_1 is mostly located in the sympathetic nerve terminal, while also being observed in the trigeminal ganglion, dorsal root ganglion and dermic nerve endings, where it regulates nociception from afferent nerve fibres.[37] CB_1 is also located in the GI tract where it modulates the mobility of the intestines, and the secretion of gastric acids, fluids, neurotransmitters, and hormones. In conjunction with the hypothalamus, CB_1 participates in the regulation of energy balance and metabolism. Given the diverse physiological roles associated with CB_1 processes, it remains an attractive drug target and, consequently, there is a growing interest in therapeutic benefits of cannabinoids that modulate CB_1 signalling.[38]

1.3 Localization and Cellular Signaling Effects of CB_1

Electrical activity in the CNS is regulated by neurotransmitter release at synapses (**Figure 1.1**). Endocannabinoids, naturally occurring cannabinoid ligands, are responsible for regulating neurotransmitter release through an interplay between the different components within the endocannabinoid system.[20]

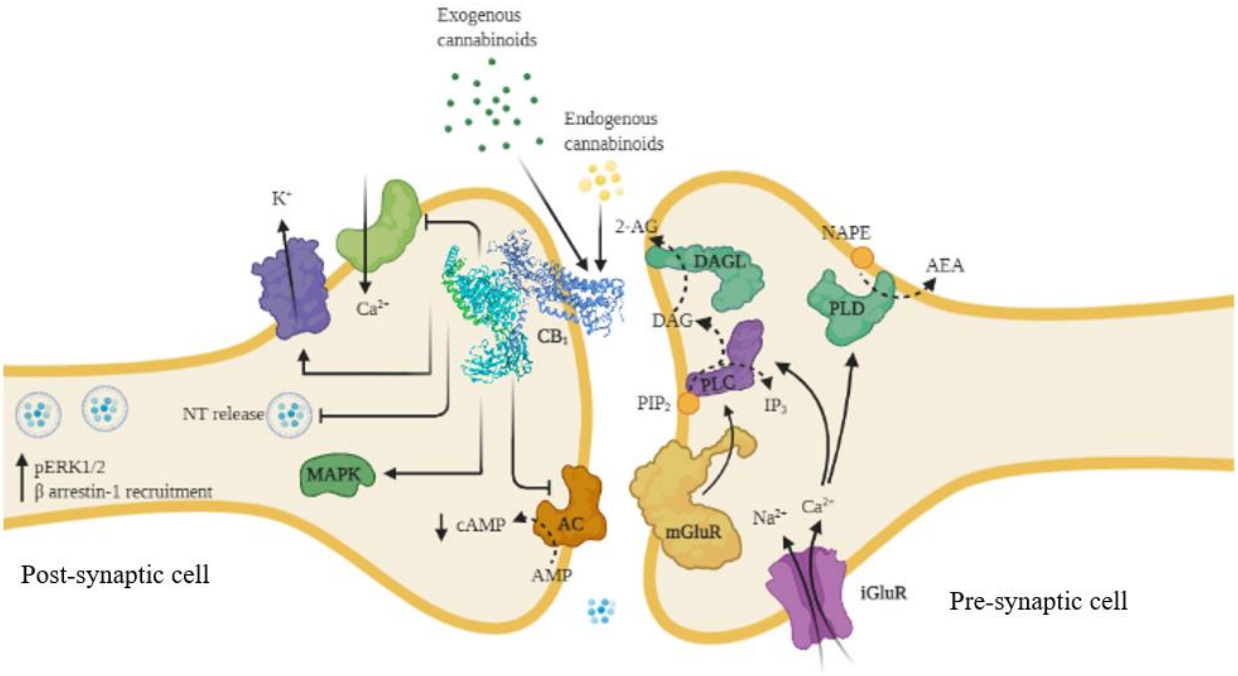


Figure 1.1: CB_1's Signalling mechanisms when activated by endogenous or exogenous cannabinoids. Events within the pre-synaptic and post-synaptic cell are visualized. Within the pre-synaptic cell, enzymes synthesize the endogenous cannabinoid ligands, which then travel in a retrograde manner to activate the CB_1 receptors on the post-synaptic cell. (Created with BioRender.com)

AEA, one of the endocannabinoids, was discovered to be a high-affinity, partial-agonist of CB_1 while displaying predominate inactivity at CB_2.[24] In contrast, 2-AG, the second endocannabinoid, has been reported to act as a full agonist at both receptors. [24] Despite their differences in receptor selectivity, these endocannabinoids are produced on demand in response to increased intracellular Ca^{2+} concentration, thus affecting the neurotransmission process. For instance, depolarization-induced suppression of inhibition (DSI) and excitation (DSE) in the GABAergic and glutamatergic neurons, respectively, was the first conclusive evidence to support endocannabinoid signalling.[39] This phenomenon, occurs from the release of retrograde endocannabinoids (from the post-synaptic cell) binding and activation to the CBRs, to then reversibly inhibit afferent synapses.[39] Later it was discovered that the endocannabinoid system was not only involved in short-term depression but long-term depression (LTD) at excitatory and inhibitory synapses.[39] In

response to increased Ca^{2+} concentrations, 2-AG, synthesized at the post-synaptic cleft, serves as a retrograde synaptic messenger where it binds to CB_1 at the presynaptic terminal and suppresses the release of neurotransmitters and their systems in two ways: 1) inhibition of voltage-gated Ca^{2+} channels thus reducing Ca^{2+} influx, and 2) inhibition of adenylyl cyclase and the subsequent cAMP/PKA pathway that is involved in LTD.[4] CB_1 is also one of the most important receptors involved in the neuro-regulation of γ-aminobutyric acid (GABA) and glutamate, due to its location on the GABAergic and glutamatergic central and pre-synaptic neurons.[40,41] The localization of these receptors aids in one of their major functions of the endocannabinoid system and the receptors, to inhibit neurotransmitter release. The termination of the signalling requires the degradation of the cannabinoid ligands.[37]

From the initial discovery and precipitation of the CBRs, considerable research has been carried out to elucidate and understand their molecular signalling pathways using a variety of endogenous and exogenous ligands.[38] Upon activation, CB_1 signals through the pertussis toxin (PTX)-sensitive $G\alpha_{i/o}$-type G protein to inhibit the production of 3', 5' adenosine triphosphate (cAMP) by adenylate cyclase.[38] This activation promotes the Gβγ subunit dissociating from the Gα subunit, which subsequently promotes the opening of inwardly rectifying potassium channels and the inhibition of N-type and P/Q type calcium channels. Besides that, of the canonical $G_{i/o}$ receptor, CB_1 has also been found to associate with non-canonical receptors such as the G_s and $G_{q/11}$ subunit, which can increase in association with exposure to certain synthetic cannabinoid ligands such as AB-CHMINACA, as well as in a cell-type dependent manner, through functional selectivity or biased agonism.[38,42] Biased agonism describes the ability of ligands to elicit differential activation of signalling pathways through stabilization of distinct receptor conformations and promoting various G-protein couplings which drive different downstream signalling cascades. For instance, AB-CHMINACA was able to increase cAMP concentrations dose-dependently in HEK cells, attributing AB-CHMINACA's ability to regulate CB_1-mediated activation of the G_s pathway more efficiently than in comparison to CP 55,940 and WIN55,212-2 which inhibited elevated cAMP concentrations under the same conditions.[38,43] Moreover, another instance of cannabinoid non-canonical bias in G-protein coupling was observed with WIN 55,212-2 and HU-210, which displayed equipotent coupling of the CB_1 receptor to the $G_{\alpha i}$ and $G_{\alpha s}$ subunit, while CP55,940 favourably promoted CB_1 coupling to $G_{\alpha i}$ protein.[44] Due to the plethora of diverse classes of synthetic compounds, it is unclear how their structural diversity may induce a range of

pharmacological effects. Therefore, understanding the binding mode and interactions of the cannabinoid ligands against CB_1 may provide insights into these effects.

1.4 Introduction of CRIP1a and its *in vitro* Implications on CB_1

In recent years, it was discovered that CB_1 mediates a wide variety of protein-protein interactions through its C-terminal tail which is important for CB_1-intracellularization, cellular signalling, as well as its overall functional state (**Figure 1.2**). One example of such protein-protein interactions is the cannabinoid receptor interacting protein 1a (CRIP1a). CRIP1a arrives from the alternative splicing of the CRIP1 gene located on the human chromosome 2,[20] generating two isoforms: CRIP1a and CRIP1b.

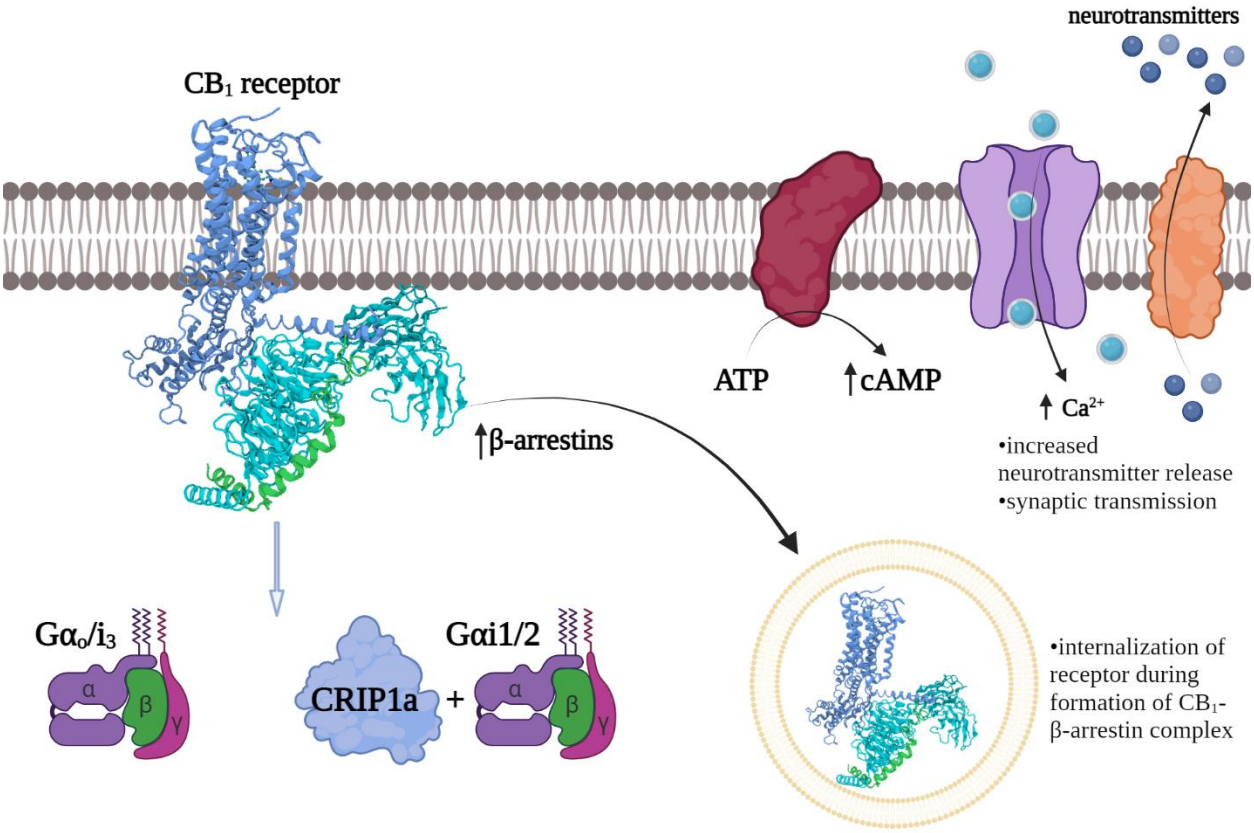


Figure 1.2: Cellular transduction alterations induced by CRIP1a-CB_1 interactions. CRIP1a upregulates the association with the Gα½ subunit while decreasing the CB_1 association with the Gαo/i_3 subunit. Increased CRIP1a expression, also increases Ca^{2+} influx into the cell, which in turn increases the cellular neurotransmitter release. CRIP1a also promotes increased cAMP cellular

concentrations. In contrast, heightened association with β-arrestins promotes CB_1 internalization. (Created with BioRender.com)

CRIP1a is highly expressed in the central nervous system and is co-localized with CB_1 in pre-synaptic compartments of excitatory glutamatergic and inhibitory GABAergic neurons.[20] Previous experiments report that the last nine C-terminal amino acids of CB_1 are responsible for binding with CRIP1a, and this molecular recognition alters the G-α protein coupling of CB_1 and its intracellular signalling effects.[45] CRIP1a promotes the selectivity of the $G\alpha_i$ subtype activation with the $G\alpha_{1/2}$ subunit while discouraging CB_1 coupling to that of the Gi_3 and G_o subunit[46]. G_{i3} and G_o activation are associated with the common CB_1 agonist-induced intracellular signalling such as cAMP inhibition, Ca^{2+} channel inhibition, MAPK activation, and reduced glutamate or neurotransmitter release into the presynaptic cleft.[20,46,47] In contrast, overexpression of CRIP1a suppresses CB_1-mediated tonic inhibition of N-type voltage-gated Ca^{2+} channels in superior cervical ganglion neurons, increasing glutamate release and decreasing CB_1-stimulated [^{35}S]GTPγS binding to $Gi_{3/o}$.[20,48] Thus, the CRIP1a association promotes the commonly associated inactivating cellular effects induced by CB_1, which alters agonist-induced receptor internalization through competitive interactions with β-arrestin proteins.[48] With the constant over-activation of the receptor, there is a recruitment of β-arrestins (β-arrestin 1 and 2) to the phosphorylated C-terminus by GPCR kinases.[20,38] Interaction with β-arrestins has modulatory effects on downstream signalling and regulation (desensitization and internalization) of the GPCRs. Competition with *β*-arrestin provides a functional interference with the clathrin- and dynamin protein mechanisms that aid in the internalization process.[48] Phosphorylation of CB_1 C-terminal peptides through affinity pull-down studies identified that phosphorylated threonine-468 at the CB_1 distal C-terminus reduced CB_1-CRIP1a association but increased the competition for the CB_1 association with *β*-arrestins.[48] Moreover, mutation of potential phosphorylation sites in the distal CB_1 C-terminus, as well as five distal residues with the C-terminus (D466, T467, S468 and A471, L472) that are associated with internalization, reduced receptor endocytosis.[20] Thus, competitive binding to the critical sites for internalization in CB_1, serves to functionally attenuate agonist-mediated *β*-arrestins recruitment, postulating the functional influence of CRIP1a on agonist-driven CB_1 internalization cellular signalling efficacy, surface density and tolerance.[48] CRIP1a has many

pharmacological benefits[20,47,49] such as mitigating some of the pathological side effects induced by synthetic agonists, for example, increased tolerance, abstinent from inducing toxicological or psychoactive effects in the body, as well as being a more natural drug target, and thus may avoid the off-target side and psychoactive side effects.[20,45,47,49] Due to the inactivated state that this protein-protein interaction entails, it aids in treating certain pathological states such as inflammatory diseases and conditions, like rimonabant or other cannabinoid inactivators. For instance, CRIP1a-CB_1 interactions can aid those experiencing obesity, by decreasing the stimulation of hunger through the endocannabinoid system's role in metabolism and regulating leptin concentrations.[23] Overactivation of the endocannabinoid system may contribute to the development of diabetes through increased feeding, which leads to a higher energy intake and storage.[50] These events impair glucose and lipid metabolism, by exerting pro-apoptotic effects in pancreatic β-cells and inflammation in pancreatic islets.[50] Therefore, the association of these two proteins may prove beneficial in treating diabetes. CRIP1a provides a new offset drug target that can indirectly control the signalling transduction capabilities of the human CB_1 receptor, thus generating a new approach that pharmaceutical companies can take when mitigating or enhancing the effects of the CB_1 receptor.

1.5 Pathological Roles of CB_1

Due to the many expressed locations of CB_1, it can participate in a wide range of roles and functions within the body., which allows CB_1 to possess a direct involvement and therapeutic role in many pathological processes. In particular, CB_1 has been discovered to play a role in pain and nausea[51–55], seizures, cancer[17,56–60],inflammatory diseases[61–63],gastro-intestinal[31,64–68] and cardiovascular disorders[28,29,63,69,70], as well as infertility [32,33,71–73]. Thus, targeting CB_1 provides therapeutic potential in a plethora of various conditions.

CB_1 regulates energy homeostasis through the central and peripheral mechanisms, where inactivation of the receptor in the periphery ameliorates the development of visceral obesity and diabetic-induced metabolic complications.[74] Within the hypothalamus CB_1 regulates the concentration of leptin through a negative feedback loop with the endocannabinoid levels.[23] High levels of endocannabinoids reduce the level of leptin, thus decreasing the number of anorexigenic hormones throughout the body.[23] In contrast, the inactivation of the CB_1 receptor, allows for elevated leptin levels, thus reducing the sensation of hunger and feeding.[23] Generally, the presence

of CB_1 in the periphery is low but was reported to be elevated under pathological conditions such as obesity and diabetes, under which conditions increased CB_1 was observed in adipose tissues.[75] Blockade of the CB_1 receptor, attenuates lipid metabolism downregulation in visceral adipose tissue[76] while enhancing insulin receptor signalling thus increasing β-cell proliferation[77]. Inactivation of the CB_1 receptor is also associated with improved obesity-related kidney injury.[78] Obesity-promoted renal inflammation and oxidative stress can lead to kidney failure, glomerulosclerosis and tubulointerstitial fibrosis.[74,78,79] Exposure to CB_1-inactivating molecules such as rimonabant or AM251 were able to delay the development of proteinuria and glomerular and tubulointerstitial lesion in obese rats.[80,81]

The blockage of the CB_1 receptor was also reported to be of potential therapeutic value in cardiovascular disorders such as myocardial infarction, hemorrhagic and septic shock.[63] Elevated endocannabinoid levels were observed under such a cardiovascular pathological state.[63] Endocannabinoids were reported to stimulate the vasodilatory effects through blockage of norepinephrine from perivascular nerves in the sympathetic nervous system.[70] For instance, Pacher and co-workers[28,69] observed that the hypotensive effects mediated by AEA, an endocannabinoid agonist, were elicited through activation of the CB_1 receptor during myocardial function.[28,82] Thus inactivation of the CB_1 receptor may be able to mitigate the effects of reduced heart contractility and hypotension in patients with cardiovascular disorders.[29]

The peripheral CB_1 system is an important therapeutic and prognostic tool concerning various metabolic and cardiovascular diseases. Reported agonistic and antagonistic peripheral therapeutic effects at the CB_1 receptor [83–86] have facilitated a complex network of therapeutic roles that the receptor could be involved in, depending on the inactivation or activation of the signalling transductions of the receptor. Thus, further research is necessary to understand the contributing roles that the endocannabinoid system plays a part in within certain pathological diseases so that clinical utilization of cannabinoid antagonists/agonists can be made plausible.

1.6 Progress in the Discovery and Development of CB_1 Receptor Modulators as Therapeutics

Since CB_1 is linked with various physiological and pathological processes in humans, several ligands (or cannabinoids) have been developed and pursued since the 1940's[38] to mediate different functional states of this receptor. However, the pharmaceutical applications of these cannabinoids have been hindered due to their adverse effects. For instance, synthetic agonists and partial

agonists of the CB_1 receptor (**Figure 1.3**) were developed to explore the therapeutic benefits of CBR activation, such as its anti- nociceptive[51] and anti-emesis[87] effects. However, more recently, concerns surrounding the abuse of the compounds for recreational use as well as incidences of toxicity have prompted their withdrawal or abstinent from clinical use. Examples include synthetic agonists such as JWH compounds—a subclass of synthetic cannabinoid agonists called naphthoylindoles, which include compounds like JWH-018[88] and JWH-133[89] —, WIN55,212-2, and CP55,940, Δ^9-THC (a phyto-cannabinoid), as well as endocannabinoids (2-AG and AEA).

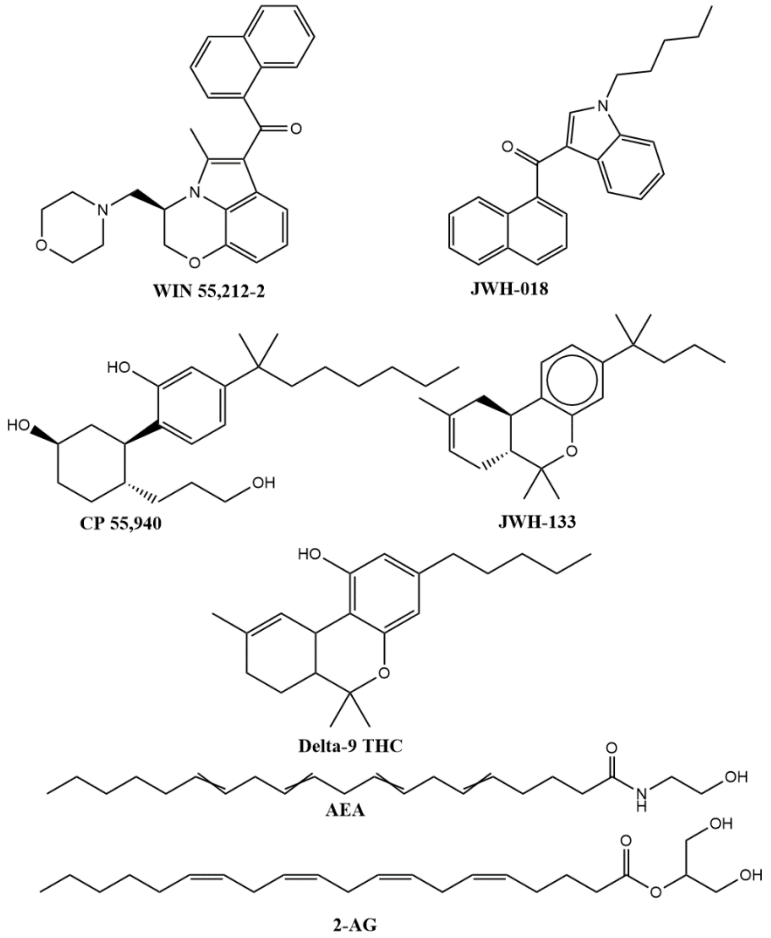


Figure 1.3: Common cannabinoid ligands. WIN 55,212-2, JWH compounds (JWH-018 and JWH-133), and CP 55,940 are agonist ligands (compounds which activate the receptor). Δ-9 THC is a partial agonist (a compound that is only able to activate the receptor sometimes). Lastly, AEA and 2-AG are examples of the two endocannabinoids within our biological systems that activate the receptor.

On the other hand, cannabinoid antagonists and adverse agonists (i.e., ligands that decrease the basal signalling rate of the receptor) were also developed to inactivate CB_1 for therapeutic benefits against inflammatory-based conditions and obesity, for example.[61,90] However, these cannabinoids also triggered adverse effects such as enhanced progenesis of mood disorders, as well as gastrointestinal concerns such as the onset of increased diarrhea.[91] The most well-known example of an antagonist of CB_1 is Rimonabant (also known as SR141716) which underwent various clinical trials and was market-approved for anti-obesity treatment in Europe in 2006-2008.[92–95] As a drug, rimonabant was found to be effective at reducing food intake, BMI, waist circumference as well as improving obesity-induced insulin and leptin resistance, restoring glucose homeostasis and dyslipidemias, as well as reducing hepatic fat accumulation in obese individuals, improvement in cardiovascular risk factors (improved HbA1c &TG, HDL-C, increased glucose tolerance and insulin levels).[30,90,96,97] However, in 2008, the European Medicines Agency (EMEA) suspended the market approval of Rimonabant due to its severe side effects and its role in the onset of mood disorders.[94,95] Likewise, Rimonabant also failed to secure market approval by the US Food and Drug Administration (FDA) based on the same grounds of psychological side effects, as well as the high dropout rate during the North American Rimonabant in Obesity (RIO) clinical trial.[93,97]

Rimonabant in Obesity (RIO)— RIO Europe, RIO Lipids, RIO Diabetes and RIO North America—were carried out for a total of 1-2 years[97]. Due to CB_1's abundancy on the presynaptic nerve terminals, it was not only involved in controlling food intake and energy expenditure but also reward-related responses, thus explaining the well-documented adverse effects amongst cross RIO studies such as nausea, vomiting, diarrhea, headaches, dizziness, anxiety, depressed mood as well as depression, anxiety, and suicidal ideation. [74,94,97] Despite, many of the RIO trials making their way into Phase 3, increased suicidal tendencies within patients ultimately resulted in its failure. Two patients in the clinical trials had committed suicide—one in the RIO-North American trial taking 5 mg of rimonabant and another in the STRADIVARIUS study for taking 20 mg.[93,97] The US FDA committee has expressed concerns over the increased risk of psychiatric adverse effects, which the RIO trials had not initially considered when conducting their studies. The results with rimonabant not only led to the withdrawal of the drug from the market but the subsequent discontinuation of the development of CB_1 inactivators by many pharmaceutical companies, fearing the therapeutic potential vs. harm of these classes of compounds.[74,98] Evidently, the outcomes of the RIO trials have highlighted the increased need in understanding the interplay

between the synthetic cannabinoid binding interactions and the mediated effects within the body. Potentially synthesizing a ligand that can bind in such a way, to mediate the therapeutic effects within the body without the adverse effects would be most optimal. Another therapeutic potential could be choosing different off-target drug targets that can alter the signalling of the receptor indirectly, through allosteric modulators or protein-protein interactions, to help mitigate the potent orthosteric enhanced efficacy effects mediated by synthetic ligands.

CHAPTER 2

Objectives and Hypothesis

Objectives and Hypothesis 2.1

CB_1, (**Figure 2.1**) a class A GPCR, composed of 472 amino acids, contains seven transmembrane segments, which are connected through subsequent extracellular (ECL) and intracellular loops (ICL). ECL2 is the largest ECL (composed of 17 amino acids) in CB_1 that contributes to orthosteric-ligand binding and stability, while the ICL3 is the largest ICL (composed of 44 amino acids) which has a role in G-protein coupling and potential CRIP1a interactions[99]. CB_1 also contains a large N-terminus (116 residues long), and a C-terminal domain (72 residues long) which consists of two short helices, termed helix 8 and 9. The publication of crystallographic structures has revealed the structural morphology of primarily the transmembrane segment of the CBRs to scientists.

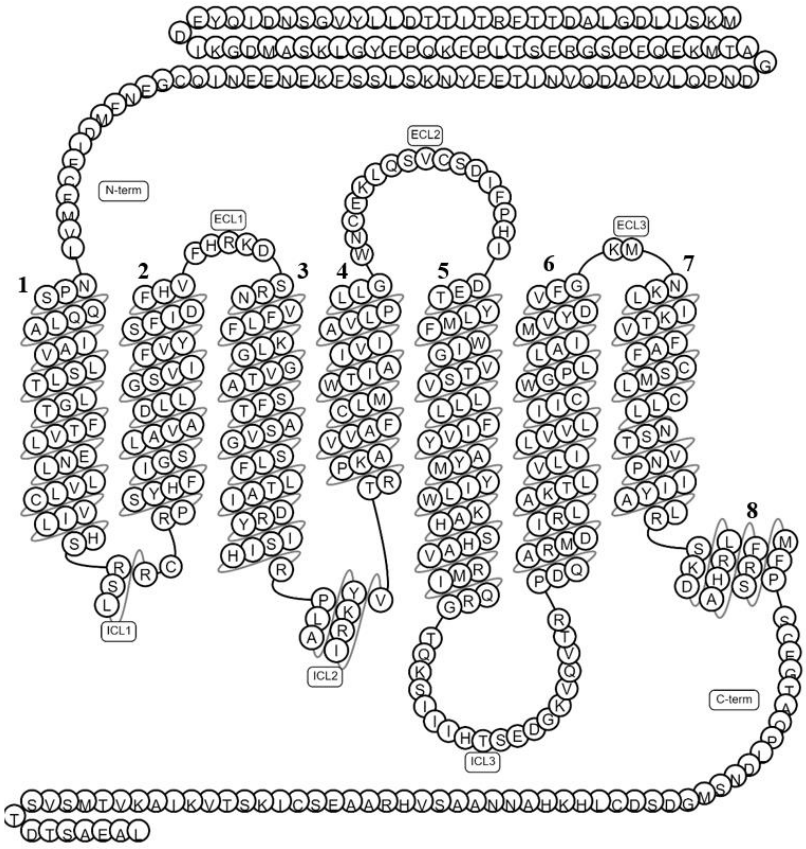


Figure 2.1: Snake diagram of cannabinoid receptor 1. Visualization of the seven TM helices, as well as the long N-/C- termini and the ICL3.[100,101]

Since 2016, various crystallographic CBR structures are being published, which have provided an understanding of the conformational states adopted by the receptor and at its orthosteric and allosteric sites, before, and during ligand binding via computational, structural, and biochemical methods. Understanding the dynamical adopted conformations of the receptor and how these modes implicate and are altered under protein-protein and protein-ligand interactions, are fundamental to understanding the biological system. Ligand binding is not a static process and involves a wealth of dynamic transitions that a protein accommodates to facilitate ligand binding, these concepts are of great importance when contributing to the fields of cannabinoid medicinal and pharmaceutical-based chemistry. The acquired conformation induced by each cannabinoid ligand in addition to their intra-molecular mediated interactions within the receptor is unique. Therefore, concerns arise, when the crystallographic structures available on the PDB database provide a static image of the receptor under the influence of a single ligand. In essence, these static images represent a single time-point of the receptor, leaving most of its assumed conformations, in addition to adopted conformations for other ligands and protein interactions unaccounted for. These same crystallographic structures are subjected to the introduction of various modifications— such as truncations, point mutations and insertion of stabilizing domains— which are necessary to increase the stability of the receptor in the crystallization process. These modifications alter the conformational state of the receptor and generate a modified but stabilized protein or protein complex. Consequently, the larger segments in CB_1 crystallographic structures, such as the ICL3, the N- and C- termini are often subjected to severe modifications, where there has yet to be isolation and crystallization of the entirety of these segments. To date, only portions of the ICL3 (PDB accession codes: 1LVR, and 1LVQ), the membrane-proximal region (MPR) (residues 90-110) of the N-terminus and the helical 8 segment of the C-terminus (residues 400-411), have been crystallized. Thus, crystallographic structures deliver only a partial story of the receptor's dynamics and structure. For these reasons, additional steps are crucial in the re-building of the disordered and large regions within the system, in addition to utilizing appropriate force fields, to generate the greatest accuracy of the protein and its conditions. By doing so, reproducibility of the receptor dynamics to that of *in vivo* proteins can be obtained, therefore allowing for an increased understanding of, and accurate analysis of the biological system.

In addition to the altercations within the receptor, there exists a limitation within the diversity of available cannabinoid structures, further shrouding the CBRs and their modes of interaction in

mystery. To date, only three crystallographic structures contain allosteric modulators (PDB accession code: 7FEE,7WV9 and 6KQI) negative allosteric modulator (NAM) ORG27569, and positive allosteric modulator (PAM) ZCZ011, while four structures exist with the $G_{i\alpha}$ protein subunit (PDB accession codes: 6KPG,6KPF,7WV9,6N4B). Moreover, there has yet to be the crystallization of any cannabinergic protein-protein complexes, other than that of the $G_{i\alpha}$-$\beta\gamma$ subunit, thus leaving most cannabinergic protein-protein interactions—and the understanding of their intra-residue interactions involved in the molecular recognition process, the binding mode conformations and overall structural morphology and orientation of the proteins within the complex—to remain elusive. Crystal structures, of a singularized protein, do not define nor distinguish protein-protein interfaces, and therefore the prediction of the protein-protein interfaces is often not straightforward.[102] Despite the existence of computational approaches for distinguishing these protein-protein interfaces, these are often subjected to errors and artifacts leading to data misprepresentation.[102] Therefore, experimental data, in addition to computational approaches, when defining the protein-protein contact surface is of great importance when outlining protein-protein contact surfaces. Protein-protein interactions are essential for all cellular processes, in a which high-resolution crystallographic structure can reveal the details of their association and function. These details provide a basis to help drive many computational and experimental approaches. The generation of biological to non-biological contacts and the construction of an accurate in-tact complex is often challenging. However, even upon successful extraction of crystallographic protein-protein complexes, these often provide limited information on the biological recognitions between the two structures. Macromolecular X-ray crystallography, a popular method to obtain high-resolution structural information on protein-protein complexes, does not define the entire contacts formed in solution, therefore generating an inconclusive story.[102] Crystal protein-protein structures also introduce crystal contacts or crystal-packing interfaces, that get introduced during the process of crystallization, which is not a part of the biological contacts.[103] This creates a fundamental difficulty in differentiating between biological and crystal contacts, where more crystal contacts have been shown to have larger interface areas than biological associations.[102] Additionally, PDB's files (BIOMT records) contain information on generating the biological complex, however, often this information is error protein being at least 9%.[104] Computational methods, together with experimental data, predict and improve the reliability of the biological complex formation beyond that of other methods.[104] This is especially

true for protein-protein complexes such as CRIP1a-CB_1 in addition to β-arrestins and CB_1 C-terminal tail complex formation. Due to the lack of crystallographic data on these complexes, computational biologists need to predict the binding mode interactions through various docking programs utilizing structural and energy constraints in addition to experimental data. The combination of computational and experimental methods allows for the understanding of plausible modes of interaction, the generation of more accurate models and development of a deeper comprehensive understanding of these mediated interactions.

The full cognizance of how the assumed conformations by various cannabinoid ligands, allosteric modulators and protein interactions affect CB_1 cellular signalling, the onset of adverse effects (i.e., psychoactive effects), promotion of Gα-CB_1 biased coupling, are some of the knowledge gaps on the cannabinergic field, which can be further unearthed using computational methods. MD and computational protein analyses, provide the initial building blocks to generate a more comprehensive understanding of the conformational poses CB_1 adopts during intra-molecular interactions, how these adopted conformations affect the activated state of the receptor, and how the different structural components and residue present within CB_1, interact, and stabilize the CB_1 protein-protein or protein-ligand interactions, beyond that of the static crystallographic snapshots.

For instance, *in silico* studies[46,47,105] conducted by Singh and co-workers in 2019 and 2017, as well as Ahmed *et al*., (2014)[47] had looked at the binding morphology between that of CB_1 and CRIP1a, to gain a deeper understanding of the residues involved within the molecular recognition and binding process between the two proteins. By utilizing classical MD trajectories and binding affinity calculations, further information and dynamics of these protein contacts were able to be extracted. However, despite these two teams' efforts, their models were based on a low-quality template, potentially introducing inaccuracies in their results. Exemplifying the importance of generating accurate models and the limitations of computational models without the presence of experimental structural data. Works conducted by Jakowiecki and co-workers (2021)[106] and Chung *et al*., (2019)[107] are examples of successful computational studies that have further explained the molecular-level interactions and dynamics present between receptor-ligand and allosteric interactions, utilizing experimental structural information data conducted by Laprairie *et al*., (2015)[108]. The *in vitro* study by Laprairie and colleagues[108] concluded that the MPR was important in the binding to allosteric modulator CBD, which has been proposed to bind around the orthosteric

pocket of the receptor. Utilizing binding-free energy calculations along with classical MD simulations, Chung *et al*., (2019)[107] and Jakowiecki and co-workers (2021)[106] were able to show the importance of the presence of the MPR and entire N-terminus, respectively, in addition to showing a network of important contacts that are formed between the ligand, allosteric modulator, the N-terminal segment and residues within the receptor. Results from these computational studies demonstrate the utility of the extent of information which can be retrieved through *in silico* methods, which is unable to be captured through experimental studies alone.

With the quick advancements in the cannabinergic field, more and more structural information and crystallizations are being published allowing for increased accuracy and analyses for future computational and experimental studies. Therefore, we hypothesize that the combination of computational techniques with previous recently published experimental structural data should help to provide novel structural and dynamics insights that are sometimes difficult to be captured by experiments alone and their usefulness in safe-therapeutics design.

Ultimately, within this thesis we aimed to *1) build the complete atomistic models of the human CB_1 receptor (both activated and inactivated conformations), in addition to the human CRIP1a protein. Utilizing these computational models, we then 2) analyzed the binding-affinity relationships between inverse-agonist under the presence of the complete CB_1 receptor and how the presence of the N-terminus affects the stability and binding affinity of the orthosteric ligand in the receptor complex. Lastly, we focused on the 3) residues involved in the CRIP1a-CB_1 protein-protein binding interactions, their modes of association, and the newly adopted dynamical interplay between both proteins during their formation in comparison to their isolated systems.*

CHAPTER 3

Ligand Binding Interactions and Dynamic Interplay with the Complete Human Cannabinoid Receptor 1 (CB_1)

3.1 Introduction

CB_1 contains one of the longest N-terminal domains (116 amino acids-long segment) amongst the class A GPCRs. Its structure and role in ligand binding are not clear and are shrouded in mystery. Given its larger size and highly dynamic nature, it is often difficult to characterize it using structure determination techniques such as X-ray crystallography or cryo-EM technology. For these reasons, only a small portion, commonly known as the membrane-proximal region (MPR), of the N-terminus has been resolved in the known experimental structures of CB_1.[109,110] The MPR, encompassing residues ~90-110, has been shown to participate in an important role in ligand stability and solvent exclusion, within inverse-agonist/antagonist- crystallographic structures (PDB accession codes: 5TGZ[109] and 5U09[110]). Particularly, the MPR has been observed to form a 'V-shaped' plug into the orthosteric pocket, shielding the lipophilic-based core from solvent molecules and limiting entry.[110–112] Additionally, the MPR has also been identified to mediate interactions with the antagonists binding to the orthosteric site in CB_1, thus playing a role in inverse-agonist/antagonist binding.[110] A mutational study conducted by Laprairie *et al.*,[108] has highlighted the MPR's ability to impair CB_1's capacity to bind ligands due to the existence of a disulfide bond Cys_{98}-Cys_{107} within the confines of the MPR. This disulfide bond was also shown to play a role in modulating ligand binding to orthosteric and allosteric sites of the receptor.[108] Both *in silico*[106] and experimental[108] studies have outlined the importance of the entire N-terminus and MPR segments, respectively, in context to allosteric ligand binding, particularly to that of CBD, Org 27569 and PSNCBAM-1.[113] Thus, this begs the question: if the entire N-terminus also plays a role in the binding toward inverse agonists and antagonists modulators beyond that of the MPR.

Typically, the importance of the N-terminus has been underestimated in terms of its significance in protein binding and dynamics, as most previous simulation-based efforts were only focused on the TM model ignoring the terminal segments.[107,114–116] Jakowiecki and co-workers[106] were the first to report the full N-terminal domain *in silico* using the replica exchange molecular dynamics with solute tempering (REST2) approach to their model. Within their work, they studied the impacts of the N-terminal domain on negative allosteric modulator (CBD) and partial-agonist binding. It was observed that the N-terminus was more stabilized under the influence of duo ligands: a negative allosteric modulator (CBD) and a partial agonist (Δ^9-THC) bound to CB_1. In

contrast, the N-terminus was more dynamic in the presence of Δ^9-THC alone, where it had undergone a structural rearrangement in terms of the lengths of its secondary structure loops. Overall, the presence of the N-terminus was observed to have a great impact on the stability of ligand binding via interactions with various structural segments along the orthosteric pocket.[106] Furthermore, the current knowledge also highlighted the accommodating nature of the N-terminus, where it can adopt different conformations depending on the ligand(s) bound to them. The structural conformations of the entire N-terminal region in the presence of orthosteric antagonists/inverse-agonists, beyond that of the MPR, remain elusive. This chapter aims to address this gap using atomistic modelling and simulation of a complete inactivated CB_1 structure in a complex with orthosteric inactivating ligands. For this purpose, one inverse agonist AM251, and two antagonists AM6538 and MJ15 were considered. AM6538 and AM251 are analogs of rimonabant (or SR141716A), a well-known antagonist of the human CB_1 receptor (as discussed in Chapter 1). AM6538 (K_i =0.038 nM)[109,117] is, a derivative of AM251, and widely used as a rimonabant analogue, is a functionally irreversible antagonist of CB_1 *in vitro* and *in vivo*.[118] Compared to AM251, AM6538, contains an acetylenic chain system incorporating four carbons and a substitute at the omega carbon, in replacement of the iodo group at the para position of the 5-phenyl ring in AM251.[109] Unlike AM6538, which was used to stabilize the CB_1 receptor during its crystallization process, ZDG is a modified version of the antagonist visualized in the PDB structure. ZDG lacks the nitrate (ONO_2) group in its structural morphology, due to its higher degree of flexibility in comparison to the other atoms in AM6538, resulting in the absence of electron density, and thus its exclusion only in the visualization aspect.[109] On the other hand, AM251(K_d= 0.23-0.8 nM)[119–122], the precursor of AM6538, is used as a pharmacological standard as a CB_1-selective antagonist[109]. In AM251, the *p*-chloro group attached to the phenyl substituent at C5 of the pyrazole (originally present in rimonabant) is replaced with a *p*-iodo group, giving AM251 a better binding affinity than in comparison to rimonabant.[123] On the other hand, MJ15 is a potent and selective CB_1 antagonist (K_i = 27.2 pM against rat CB_1).[124] To date, two X-ray crystallographic structures of the inactivated CB_1 receptor have been reported in the PDB database: a structure of human CB_1 in complex with AM6538 (PDB accession code: 5TGZ[109]) and CB_1 in complex with taranabant (PDB: 5U09[110]). Initially, we constructed a complete model of the human CB_1 receptor using the X-ray crystallographic structure of CB_1 in a complex with AM6538 (PDB: 5TGZ). Subsequently, the ligand molecules were docked to the CB_1 receptor, to

which the dynamics of the complexes were probed using extensive classical MD simulations and binding free energy calculations. This chapter will extend insights into the structural basis of the ligand-bound inactivated CB_1 complex under physiological conditions.

3.2 Methods:

3.2.1 Model Building

The first CB_1 structure was solved in 2016, using X-ray diffraction crystallography. To date, there are nine CB_1 structures in the PDB database, with the newest structure being published in 2022, these crystallographic model's resolution range between 2.6 to 2.95 angstrom (Å) (PDB accession code: 5TGZ[109], 5U09[110],5XR8[111],5XRA[111],6N4B[114],6KQI[125], 7V3Z[126], 7WV9[127] and 7FEE[127]).

Table 3.1: Crystallographic information of CB_1 receptors. Updated crystallographic information on the CB_1 receptors, including their activated status, crystallized ligands, resolution, and publishers.

PDB accession code	**Ligand(s)**	**Activated(A)/Inactivated (I)**	**Resolution (Å)**	**Citation**
5TGZ	ZDG (AM6538)	I	2.80	Hua *et al*., (2016)[109]
5U09	(Taranabant)	I	2.60	Shao *et al*., (2016)[110]
5XR8	8D0(AM841)	A	2.95	Hua *et al*., (2017)[111]
5XRA	8D3(AM11542)	A	2.80	Hua *et al*., (2017)[111]
6N4B	KCA(FUB)	A	3.00	Kumar *et al*., (2019)[114]
6KQI	9GF(CP55,940): orthosteric ligand	A	3.25	Shao *et al*., (2019)[125]

	9GL(ORG27569): allosteric modulator			
7V3Z	9GF(CP55,940)	A	3.29	Wang et al., (2021)[126]
7WV9	9GF(CP55,940): orthosteric ligand 7IC(ZCZ011): allosteric modulator	D	3.36	Qin *et al*., (2022)[127]
7FEE	9GF(CP55,940): orthosteric ligand 7IC(ZCZ011): allosteric modulator	A	2.70	Yang *et al*., (2022)[127]

To induce crystallization, deletion of residues Val 306 to Pro332 encompassing the highly variable ICL3, were substituted with the flavodoxin stabilizing protein (residues 1002-1148). These crystallographic structures are also often subjected to the truncation of their N- and C-terminal domains, due to their large nature and flexibility. The availability of these crystallographic structures helped to understand the structures of CB_1 and support the computational investigation of ligand binding using techniques such as molecular docking and MD simulation with a high degree of accuracy. In our study, we selected an inactivated CB_1 structure co-crystallized with AM6538 from the PDB[128] database (PDB accession code: 5TGZ) as a building block to construct the complete CB_1 model. Initial preparation of the CB_1 structure concurred through the elimination of associated ligands, the deletion of flavodoxin (the stabilizing domain in the crystal structure), the addition of hydrogen atoms, and the re-construction of missing loops and atom backbones. All these initial processing steps of the structure were carried out utilizing the software package UCSF Chimera version 1.13[129]. Specifically, modelling the missing loops and segments such as the ICL3,

as well as the N- and C- termini were conducted by Modeller(v10.1)[130] and the online server Threading ASSEmbly Refinement (I-TASSER)[131], respectively. Models in I-TASSER were picked based on the quality of their template, which is heavily reliant on the C-score. The C-score in I-TASSER is defined as:

$$C - score = \ln\left(\frac{M}{M_{tot}} \cdot \frac{1}{\langle RMSD \rangle} \cdot \frac{\prod_{i=1}^{4} Z(i)}{\prod_{i=1}^{4} Z_0(i)}\right) \quad \textbf{(3.1)}$$

M is the multiplicity of structures in the cluster, M_{tot} is the total number of I-TASSER structures used in the clustering, <*RMSD*> is the average RMSD of the structures to the cluster centroid. The terms $\frac{M}{M_{tot}}$ and $\frac{1}{\langle RMSD \rangle}$ described the degree to which the structure has converged during the process of clustering. This is also correlated to the consistency of the external restraints and the I-TASSER potential implemented within the program. The next terms in the equation $\frac{\prod_{i=1}^{4} Z(i)}{\prod_{i=1}^{4} Z_0(i)}$ are explained by the Z-score value. In essence, *Z(i)* is the highest Z-score (the energy to mean in the unit of standard deviation) of the templates generated by the *ith* profile-profile threading alignment threading program. This program threads target sequences through a representative PDB library to search for possible folds. Fragments excised from the threading aligned regions are then used to generate full models, while the threading unaligned regions are built through *ab inito* modelling. Lastly, the $Z_0(i)$ term, is the program Z-score cut-off for distinguishing between the good quality and bad quality templates. The standard of the template can be described where $Z_0(1) = 1.0$, $Z_0(2) = 8.5$, $Z_0(3) = 8.0$, $Z_0(4) = 10.5$, ordering from high to low-quality threading templates.[131]

3.2.2 Ligand Preparation and Molecular Docking Calculations

Molecular docking of the selected ligands (i.e., AM6538, AM251, and MJ15) against the complete human CB_1 model was carried out using AutoDock-VINA[132], an efficient and rapid docking program. Autodock-VINA employs a docking method that uses a simplified empirical scoring function (for ranking of poses) and a gradient-optimization algorithm for conformational searches during docking.[133] The target structure and the ligands were prepared in AutoDockTools version 1.5.6[134], utilizing energy bonds, with the addition of hydrogen bonds (so they could be accounted for during docking calculations). Ligand structures were retrieved from the PubChem[135] database in the SMILES format, which was then converted to the PDB format using OpenBable. PDB files

of the ligands were then converted to the AutoDock-VINA format and Gasteiger charges were assigned to the ligand atoms using the ligand preparation script from AutoDockTools. Subsequently, the ligands were docked into the pre-determined docking grid (size) centred on the orthosteric binding site of the CB_1 model. The molecules were ranked based on their docking score, which is based on effectiveness and free binding energy value. The results from docking were then assessed through a visual comparison of the ligand binding pose, in each of the docked complexes, against that of ZDG in the crystal structure (5TGZ). Since docking programs are known to be able to reproduce crystallographic binding poses, but not necessarily rank them correctly, binding poses that showed a good agreement with the crystal structure, demonstrated by the overlap of their cores, and heads were selected for further analyses of classical MD simulation and MM/PBSA-based binding free energy calculations[136]. For visual analyses of the docking results, the PyMOL[137] program was used. Since the experimental binding data were available for all the ligands considered in this work, we then used the reported K_i (inhibition constant) data and the resultant experimental binding free energies (ΔG) against the values predicted from our calculations.

K_i values were converted into binding free energies using the following equations:

$$\Delta G_{bind} = RTlnK_i \quad \textbf{(3.2)}$$

$$lnK_b = -lnK_d \quad lnK_b = -lnK_i \quad \textbf{(3.3)}$$

Here, R is the gas constant (0.0831446261815324 $L{\cdot}bar{\cdot}K^{-1}{\cdot}mol^{-1}$), and T is 310K in Kelvin.[136]

3.2.3 System Building

Utilizing the Membrane Builder program in CHARMM-GUI[138], the ligand-bound CB_1 complexes were constructed in a physiological environment of the membrane bilayer, solvent (water molecules) and ions. In the first step, the TM positioning of the CB_1 receptor was determined using the Position of proteins in the membrane (PPM)[139] program and the orientations of proteins and membranes (OPM) database[139]. Subsequently, each of the ligand-protein complexes was embedded in a pre-equilibrated lipid bilayer made up of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) molecules and solvated with TIP3P water molecules. The ion

concentration for each system was set to 0.15M NaCl. The overall sizes of the assembled systems were 150Å x 150Å x 143Å; 150Å x 150Å x 143Å and 150Å x 150Å x 144Å for the AM6538, AM251 and MJ15 complexes, respectively. The parameter and topology files for the protein and the ligand were generated for further MD simulation.[140]

3.2.4 Classical Molecular Dynamic Simulation and Analyses

All MD simulations were run using the NAMD version 2.11[141] simulation package with CUDA support, with the CHARMM36m forcefield[140] for describing the structures of the protein, and the CHARMM general force field (CGenFF)[140] for the bound ligands in this study. Each system underwent multiple stages of energy minimization utilizing conjugate gradient, 15 ns long equilibration and 50ns long production MD simulation. MD equilibration and production were carried out at a preset temperature of 310K with an atmospheric pressure of 1.01325 bar. In addition, Particle Mesh Ewald (PME) electrostatics were implemented, allowing for infinite electrostatics (long-range electrostatic interactions) without truncation. The van der Waals force-based switching distance of 10 Å was applied to the system to ensure smoothing of the function for the electrostatics and van der Waals forces, and to avoid abrupt truncations in the force, thus conserving energy. A cut-off of 12Å was set for local electrostatic and van der Waals interaction distance calculations. All bonds involving hydrogens were set as rigid, to avoid high dynamics, and Langevin dynamics were applied to the system. Harmonic, planar, and dihedral restraints (**Table 3**) were applied to the protein, lipids, and the ligand within the system, in which beta-coupling (temperature-coupling) was used for the harmonic constraint force constant. Harmonic restraints and planar restraints were gradually reduced from 500 and 5 to 0, respectively, over equilibration and minimization. Production was run restraint free with rigidity added to the hydrogen atoms and bonds within the system.

Table 3.2: Stepwise information on the protocol for equilibration. Minimization was conducted together with equilibration in step 6.1. Release of planar and harmonic (dihedral restraints) throughout the simulation are shown within each step.

Equilibration step	Restraints (Kcal/mol):	
	Planar	**Harmonic (dihedral)**
6.1	5	500
6.2	5	200
6.3	2	100
6.4	1	100
6.5	0.2	50
6.6	0	0

The stability and flexibility of the ligand-protein complexes during the MD simulations were assessed based on the evolution of the backbone root mean square deviations (RMSDs) and per-residue root mean square fluctuations (RMSF). The RMSD and RMSF plots were computed from the MD trajectories of the complexes and using the CPPTRAJ[142] software in the AMBER18[143] package. All other trajectory visualization and analyses were carried out using PyMOL and VMD.

3.2.5 Binding-Free Energy Calculations

Accurate estimation of binding free energy calculations is an important aspect of computer-aided drug design (CADD). Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) methods provide an efficient, low-cost means of estimating the binding-free energies of ligand-protein complexes.[144] In this study, we used the MM-PBSA, which employs the Poisson-Boltzmann model, to compute the relative binding free energies of ligand-CB_1 complexes. Respectively, then in comparison to the computationally expensive free-energy perturbation method. MM-PBSA method can be described by the following equation:

$$\Delta G_{bind} = \Delta H - T\Delta S = \langle \Delta E_{gas} + \Delta G^{polar}_{solvation} + \Delta G^{nonpolar}_{solvation} - T\Delta S \rangle \qquad \textbf{(3.4)}$$

The binding free energy (ΔG_{bind}) is equated to the entropy (ΔS) (the disorder within the system, which is dependent on temperature) and the enthalpy (ΔH) (the thermodynamic and the sum of the internal energy contained within the complex). These two terms can be further divided into the gas-phase binding free energy (ΔE_{gas}) as well as the polar solvation free energy of the complex($\Delta G^{polar}_{solvation}$) relative to that of the inverse-agonists and CB_1.

$$\Delta G_{bind} = \Delta G_{gas} + (\Delta G_{sol})$$

The gas-phase binding free energy (ΔE_{gas}) can be divided into the averaged CB_1-ligand interactions, which include the potential energy or molecular mechanics energy term E_{MM} and the contribution of entropy ($-T\Delta S$). E_{MM} can be further composed into the non-bonded and bonded energy, in which the non-bonded energy is composed of three terms: **1)** an aggregate of electrostatic (E_{es}), **2)** van der Waals (E_{vdW}), and **3)** internal energy (E_{int}):

$$E_{MM} = E_{bond} + E_{angle} + E_{torsion} \text{ and } E_{non-bonded} = E_{vdW} + E_{es} + E_{int}$$

The solvation energy of the complex can also be comprised into two terms:

$$\Delta G^{polar}_{solvation} = \Delta G_{pb} + \Delta G^{nonpolar}_{solvation}$$

$$\Delta G^{nonpolar}_{solvation} = \gamma SASA + b$$

ΔG_{pb} is the polar solvation free energy contribution calculated by solving the nonlinear Poison-Boltzmann (PB) equation.[145] On the other hand the nonpolar solvation free energy, is estimated using the solvent-accessible surface area (SA) term.

All MM-PBSA calculations were carried out utilizing the Calculation of Free Energy (CaFE) VMD plugin for NAMD.[146] CaFE predicts binding affinities from MD trajectories and using an end-point approach. The receptor-ligand binding event can be split into two stages: association in the gas-phase evaluated with a classical forcefield, and in the aqueous phase using implicit solvent and the difference between the two phases results in the ΔG. CaFE conducts these steps by utilizing the MD trajectory files for end-point analysis. The end-point approach in energy calculations is one in which the complex, receptor and ligand conformations are extracted from the MD trajectory files. In terms of the energy calculations, three energetic components were calculated, the first being the gas-phase energy difference between the complex, the isolated receptor and the ligand.

Second, the polar solvation free energy is calculated by calling the VMD plugin APBS[147], which calculates the solvated PB equation implemented in the program. Lastly, the difference in the solvent accessible area (SASA) was considered, along with the calculation of the nonpolar solvation free energy through its linear relationship with the SASA.[146]

The binding free energy is averaged throughout the frames, in which the entropic term (ΔS) was ignored due to its high computational costs and frequent generation of inaccurate results in entropic calculations. In addition, the entropy term can be ignored because the estimation of the binding free energies was conducted to determine the relative binding strength of structurally similar compounds in a complex with CB_1. All MM-PBSA calculations were carried out with an internal dielectric constant (indi) set to 2.0, while the external dielectric constant (exdi) was set to 80.0. 300 snapshots were used for the last 30ns, in which 100 used snapshots per 1000 frames, with a stride set to 10.

3.2.6 Principal Component Analysis

Principal component analysis (PCA) is a statistical method that is efficient in reducing the dimensionality of a complex system while extracting essential information from the principal modes of sampled motion.[148,149] PCA represents a linear transformation that diagonalizes the covariance matrix and removes instantaneous linear correlations among the variables.[148] By ordering the eigenvalues of the transformed matrix in decreasing order, it allows for the analysis of the relationship between different sampled conformations during the trajectory, as well as the principal components (orthogonal eigenvectors) describing the maximal variance on the distribution of the structures.[150,148] The percentage of the total mean square displacement (or variance) of atomic positional fluctuations captured in each dimension are characterized by their corresponding eigenvalue. Usually, 3-5 dimensions are sufficient to capture over 70% of the total variance within a given MD trajectory. [150] Moreover, the utilization of dihedral angles or atomic coordinates for α-carbon atoms is a more valuable component to measure coordinates in PCA, than in comparison to other internal coordinates such as bond lengths, and bond angles which do not undergo large changes in amplitude.[148] In this PCA analysis, the x, y and z coordinates of the C-α atoms were used for trajectory frame superimposition and PCA analysis.[148] The time-dependent motions (k) are described as a co-variance matrix (C_{ij}) between the atomic Cartesian positions of our residues (i,j), which can be described as follows

$$C_{ij} = \frac{1}{N}\sum_{k=1}^{N}\left(x_i^k - \overline{x}_i\right)\left(x_j^k - \overline{x}_j\right) \quad \textbf{(3.5)}$$

$$C_{ij} = \frac{1}{N}\sum_{k=1}^{N}(\boldsymbol{w^T}x^k - \boldsymbol{w^T}\overline{x})^2 \quad \textbf{(3.6)}$$

$$C_{ij} = \boldsymbol{w^T}\left[\frac{1}{N}\sum_{k=1}^{N}\left(x_i^k - \overline{x}_i\right)\left(x_j^k - \overline{x}_j\right)^T\right]\boldsymbol{w} \quad \textbf{(3.7)}$$

$$\boldsymbol{W^T C W} = \boldsymbol{\lambda} \quad \textbf{(3.8)}$$

To derive the principal components for our data, reduction of the dimensionality of the system is necessary by finding the projections which maximize the variance or minimize the total mean square displacement. In essence, the first principal component is the vector that has the largest variance, while the second, is the direction in space that maximizes the variance orthogonal to the first component's direction.[151] To formulate the equation of the covariance matrix the composition of x_i^k, x_j^k and their means $\overline{x}_i, \overline{x}_j$ need to be incorporated. x_i^k, x_j^k encompass a pair of elements (*i*,*j*) of vector x^k , which includes any of the x, y, and z Cartesian coordinates of the $C_\propto$ atoms for the protein at time step **k**. On the other hand, $\overline{x}_i, \overline{x}_j$ are the mean values of those Cartesian coordinates calculated from N structures within the MD simulation. Ensuring the maximization of the variance on the projected vector, the variance of the projections along the vectors, is taken by squaring the sum of the magnitude of the vector,$\boldsymbol{w^T}x^k$ for *i* and *j*, subtracted by the mean of all the projected vectors (of *i* and *j*), $\boldsymbol{w^T}\overline{x}$, where $\boldsymbol{w}$ is the unit vector and T is transformed matrix (**Equation 3.6**). In essence, equation 3.7 computes the closed form of the covariance matrix (**Equation 3.8**) or $\boldsymbol{w^T C w = w}$**,** where **C** is the closed form of the covariance matrix, and needed to be maximized. Using, Lagrangian multipliers, while incorporating the constraint that the unit vector needs to be 1, generates the final form of $\boldsymbol{\lambda = W^T C W}$ where $\boldsymbol{W}$ is transformation matrix whose columns are eigenvectors of the motions, while the diagonal elements of the, $\boldsymbol{\lambda}$, are the associated eigenvalues. Thus, using the largest eigenvalue, generates the highest variance, and thus the highest degree of fluctuation within the system. The eigenvectors show the direction and magnitude of motion of the backbone and the associated eigenvalues are the frequency or

amplitude of that motion.[149] All PCA analyses in this work were conducted using Bio3D[150], a package in R used for analyzing protein structures and MD trajectories.

3.2.7 Cross-Correlation Plots

Dynamic cross-correlation extracts essential information about the fluctuations on a macromolecule spanning hundreds of nanoseconds (ns) in MD simulations. Since correlating the dynamics between different atomistic segments is often difficult, cross-correlation plots ease this process by quantifying the correlation coefficients of motions between atoms. These cross-correlation maps help to analyze the effects that various residues and or atoms may have on each other[152]. Dynamic cross-correlation can provide insights into the correlative motion of atoms by analyzing the displacement of determining average coordinates on the backbone fluctuations and domain motions within the C-α atoms[153]. The equation of this method can be described as:

$$DCC(i,j) = \frac{\langle \Delta r_i(t) \cdot \Delta r_j(t) \rangle_t}{\sqrt{\langle ||\Delta r_i(t)||^2 \rangle_t} \sqrt{\langle ||\Delta r_j(t)||^2 \rangle_t}} \quad \textbf{(3.9)}$$

The displacement of the atom's C-α vector coordinates with respect to time is described by, $r_i(t)$, which is the vector of the *i*th atom's C-α Cartesian coordinates in any of the x, y and z plane as a function of time, while $\Delta r_i(t)$, expressed as $r_i(t) - \langle r_i(t) \rangle_t$,is the vector's *i*th C-α atoms coordinates subtracted from the average of the vector's coordinates throughout each MD time snapshot[153]. The cross-correlation plot for the isolated proteins and their complexes was also calculated using the Bio3d module[150].

3.2.8 Cluster analyses of MD trajectories

MD trajectories often involve various conformational changes of the system under study. By performing a cluster analysis, it is possible to identify the most dominant conformational states sampled during the simulation. In this work, we performed cluster analyses of the MD trajectories of the ligand-CB_1 complexes using the K-means algorithm and by employing backbone RMSD as a distance metric. K-means clustering, a popular unsupervised machine learning algorithm, partitions datasets into non-overlapping clusters{C1, C2…., C*n*} without the presumption that all

conformations are hierarchical.[154] For protein clustering, K-means aims to cluster the three-dimensional Cartesian coordinates of all the atoms within the simulated system throughout each time-evolved snapshot, based on their similarity to each other [154,152]. The similarity parameter that characterizes the formation of each cluster indoctrinates a geometric constraint, being the RMSD distance between pairs of MD conformations. Particularly, the RMSD of each conformational snapshot is calculated based on the protein atoms (ex) heavy atoms or C-α atoms) in their Cartesian (x, y, and z) coordinates. Thus, each cluster can aid in describing similarities between pairs of conformations, and facilitating the ease of trajectory analysis.[154]

$$min \sum_{i=1}^{K} \sum_{x \in C_i} \|x - \mu_i\|^2 \qquad \textbf{(3.10)}^{154}$$

Defining the algorithm; x denotes the MD conformations in the *i*th cluster, C_i ,while μ_i describes the cluster's center (geometric mean)[154]. Initially, a hyperparameter of *K* conformations is picked as the initial number of cluster centers, to which all other conformations are assigned to their closest centers (also termed nearest mean or cluster centroid). By doing this it ensures that the total within-cluster sum of squares of distances of the conformations or the Euclidean distances and their corresponding cluster centers (geometric means) are minimized[154]. In the iteration step, the re-calculation of the geometric means of each cluster is computed, in addition to the distance of all other conformations to their closest centers. The algorithm is terminated once the system reaches convergence, and the amount of *K* conformations is satisfied by the initially given hyperparameter. The average runtime complexity for K-means is *O(Knl)*, where *n* is the number of conformations in the MD dataset, K is the desired number of clusters, and l is the number of iteration steps.[154]

All cluster analyses in this work were carried out on the last 30 ns of MD trajectories of the ligand-CB_1 complexes using the CPPTRAJ[142] tool available within AMBER18[143]. To find the most optimal number of clusters, clustering was performed on the RMSD of the protein backbone atoms, which is the prime indicator of elucidated variances, to which 50 clusters were generated. The cluster with the lowest Davies-Bouldin Index (DBI) score, a relatively high pseudo-F statistic (pSF) value and reached a plateau on the SSR/SST was chosen to be the more optimal cluster. In essence, the DBI[155] measures the sum over all clusters within the current cluster to the between

cluster separation or essentially measures the average similarity of each cluster with its most similar cluster.[155,156] In this case, the similarity is the ratio of within-cluster distances to between-cluster distances, therefore clusters that are farther apart and less dispersed will result in a greater DBI score.[155] In contrast, the pSF indicates the 'tightness' of the clusters and is a ratio of the mean sum of square (variation) between groups to the mean sum of squares (variation) within the group, or in essence the similarity within the group to the variation between the group. Generally, a lower DBI and a higher pSF value indicate a higher quality cluster. The SSR/SST value also known as the R-squared value (R^2) or the coefficient of determination, is an indication of how well the percentage of the variation in the response variable (or output) can be explained by the hyper-parameter or the predictor variable[157].

3.3 Results

3.3.1 Generation of the Physiological CB_1 Model

3.3.1.1 C-terminal Building

Implementing data from various experimental and computational studies to solve the structural components comprising the entire C-terminus [46,158,159]; were used as a guide to constructing the complete C-terminal tail in I-TASSER. Outlined locations for helix $9^{\text{C-terminus}}$ and helix $8^{\text{C-terminus}}$ were used as pre-defined structural constraints in the generation of the models. The structural location and composition of helix 8 have already been solved through various CB_1 crystallographic structures, which span S400 to P412. In terms of helix 8's purpose within the receptor, it has been investigated to be involved in G-protein coupling, as well as performing in the interactions with various ICLs, which become disrupted upon activation. [159] Despite little information surrounding helix 9, NMR studies conducted by Ahn and co-workers[158], examining the entire CB_1 carboxyl-terminal tail, confirmed the existence of the location helix 8, in addition to a helix 9, located near the terminal end of the tail.[158,159] The function of helix 9 has still yet to be determined. Specifically, helix $9^{\text{C-terminus}}$ was alluded to be confined within the regions of residues 441-463 in the *in-silico* paper by Singh *et al*.,[46], and A440-M461 in the *in-vitro* study conducted by Ahn *et al*.,[158]. Utilizing these literatures, the C-terminal tail was submitted to the I-TASSER online server under the structural constraints of helix 8 encompassing residues 400-412 and helix 9 containing residues 440-463. I-TASSER[131], generated 5 models. Models with low C-scores (the scoring function in I-TASSER), which abstained from incorporating our structural restrictions for helices 8 and 9, were

omitted. The C-score ranges from values of -2 to 5, in which a higher C-score depicts a better model.[131] The chosen structure encompassed most of the pre-defined structural constraints: 404-413 for helix 8, and 441-456 for helix 9, out of all generated structures, in addition to maintaining a moderate C-score (~3).

3.3.1.2 N-terminal Building

The N-terminal section was modelled with an absence of folding restrictions due to the lack of experimental studies outlining its structural composition at the time when building the N-terminal segment. The only structural component taken into consideration was the presence of the disulfide bond between $Cys107^{N\text{-terminus}}$ and $Cys98^{N\text{-terminus}}$; previously discovered in the experimental studies with cannabidiol (CBD) by Blume and colleagues[48]. Generated models of the N-terminus by I-TASSER were visually inspected to contain the disulfide bond, where structures missing this component were disregarded. Recently, an *in silico* model for the CB_1 N-terminal domain was published by Jakowiecki and colleagues.[106] This recent paper analyzed the impacts of the N-terminus in the presence of partial agonist Δ^9-THC and negative allosteric modulator CBD. Filtered structures from I-TASSER were compared to the model by Jakowiecki and colleagues[106], as an additional structural criterion for picking the best model. The chosen model contained the N-terminal helices (NTH) 1 and 3 within Jakowiecki and colleagues'[106] initial model (**NTH1**:30-35, **NTH2**:61-69, **NTH3**:80-94), with an absence of NTH2. It is important to note, however, that throughout simulation Jakowiecki *et al.*, (2021)[106] notified that NTH2 (**NTH2**: 61-69) had disappeared with the subsequent lengthening of NTH1 in the agonist-only Δ^9-THC model (**NTH3**: 78-92, **NTH1**:29-35), which more accurately aligned with our generated model. None of the β-sheets were present in our N-terminal model. Overall, our structure contained a short helix at residues 37-39(**NTH1**), a helix at amino acids 77-84(**NTH2**), and 86-90(**NTH3**). To date, the exact entire functionality of the N-terminus is unknown.

3.3.1.3 Validation of Models

Due to the plethora of loops and disordered regions present within both extremity structures, both N- and C-termini models' structural qualities were checked using the online structural validation servers ERRAT[160,161] and VERIFY3D[162,163]. ERRAT is a structural quality program that analyzes the statistics of non-bonded interactions between different atom types to maintain the three-dimensional structure(3D).[151] On the other hand, VERIFY3D assesses the compatibility of an

atomic model (3D) with its amino acid sequence (1D).[154] Both of these programs work together to measure how well our pre-defined structural restraints in conjunction with the amino acid sequence fits the generated model. In essence, these specific validation criteria were used together to help confirm the 3D structure of each model. Cut-offs <50% were applied to both quality checks, in which both structures had passed, thus deeming them to be of suitable quality. Specifically, the N-terminal segment was assessed with a score of 66.67% and 92.24%, while the C-terminal segment had scores of 86.15% and 90.41%, for ERRAT and VERIFY3D, respectively. The discrepancy in the lower N-terminal's ERRAT score can be explained by the lack of a solved CB_1 N-terminal crystal structure (in contrast to that of the partially solved C-terminal segment) in addition to a large number of disordered regions present within the domain. These two aspects added to the lower ERRAT score, for the scoring function uses homology-based comparisons to crystallographic structures when assessing the quality of the model, therefore creating a higher level of uncertainty in its score.[161] However, the ERRAT scoring function certifies any score >50% to be deemed a high-quality model, to which our structure adhered.[161,164] In addition, our N-terminus also contained the disulfide bond ($Cys107^{N\text{-terminus}}$ and $Cys98^{N\text{-terminus}}$), solved through *in vitro* studies, further validating our model. Likewise, our C-terminal segment contained most of the residues within the confines of the experimentally determined regions of helix 8 and 9, further adding to our validation criteria. Therefore, both models were then stitched onto the modified crystal structure's TM segment using Chimera version 1.14[129].

3.3.2 Molecular Docking of Ligands against CB_1

Ligands were individually docked within a grid box centred (x = 43.44 Å; y = 27.17 Å; and z = 318.99 Å) on the orthosteric binding site of the human CB_1 model.

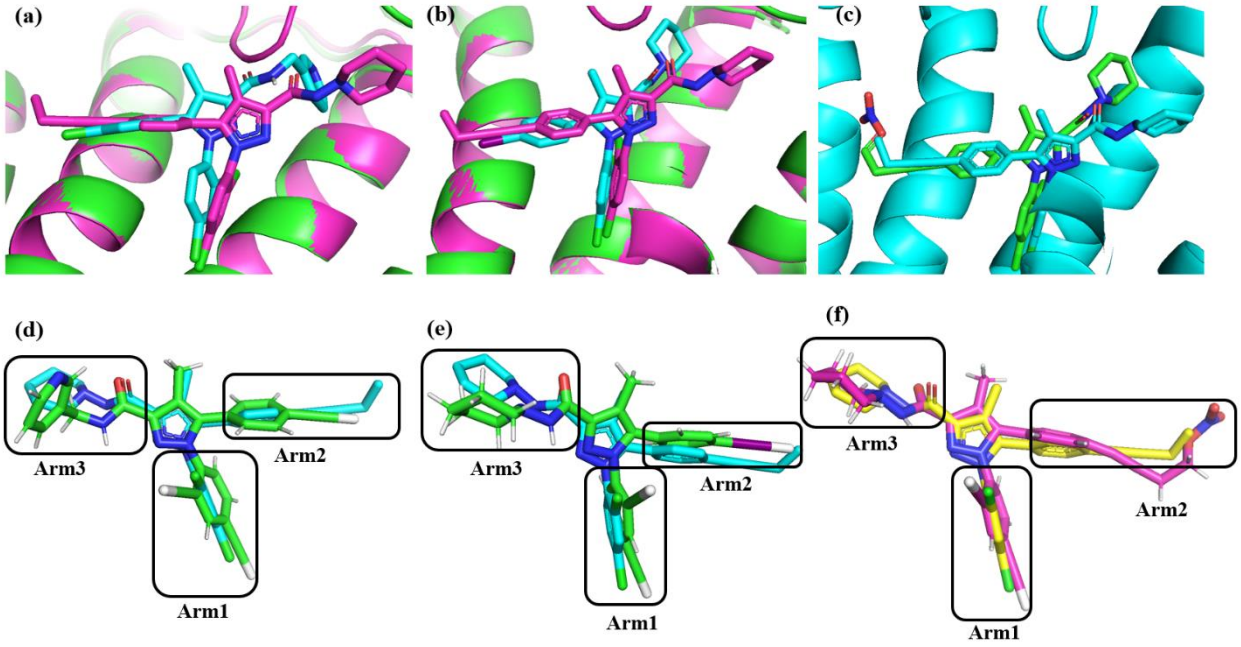


Figure 3.1: Docking poses for cannabinoid inverse-agonists. Visualization of docked **(a)** MJ15 (blue) with crystallographic ligand ZDG (magenta) **(b)** AM251 (blue) with crystallographic ligand ZDG (magenta) and **(c)** AM6538 (green) with crystallographic ligand ZDG (blue). Comparison of inverse-agonist ligand poses to the crystallographic ligand ZDG after subjection to production runs for **(d)** MJ15 (green) and **(e)** AM251 (green) and **(f)** AM6538 (pink).

The structural models of the docked complexes are shown in **Figure 3.1** All the selected ligands shared a similar chemical structural architecture, in which they contained three fragments (named arms 1-3) that are connected to the central pyrazole core (see **Figure 3.2**).

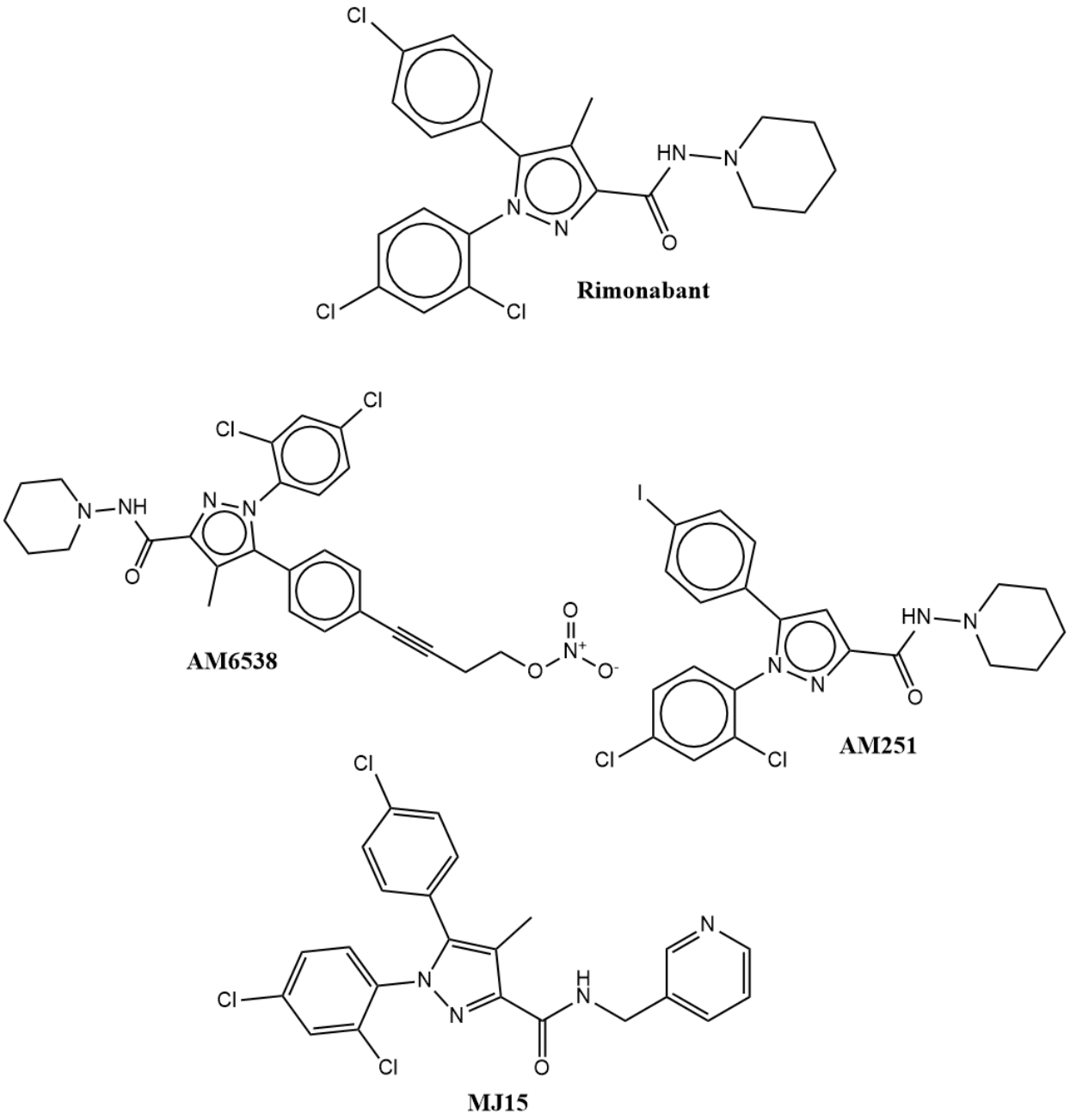


Figure 3.2: Inverse agonist MJ15 and Rimonabant antagonist analogues AM251 and AM6538. Chemical composition of various molecules.

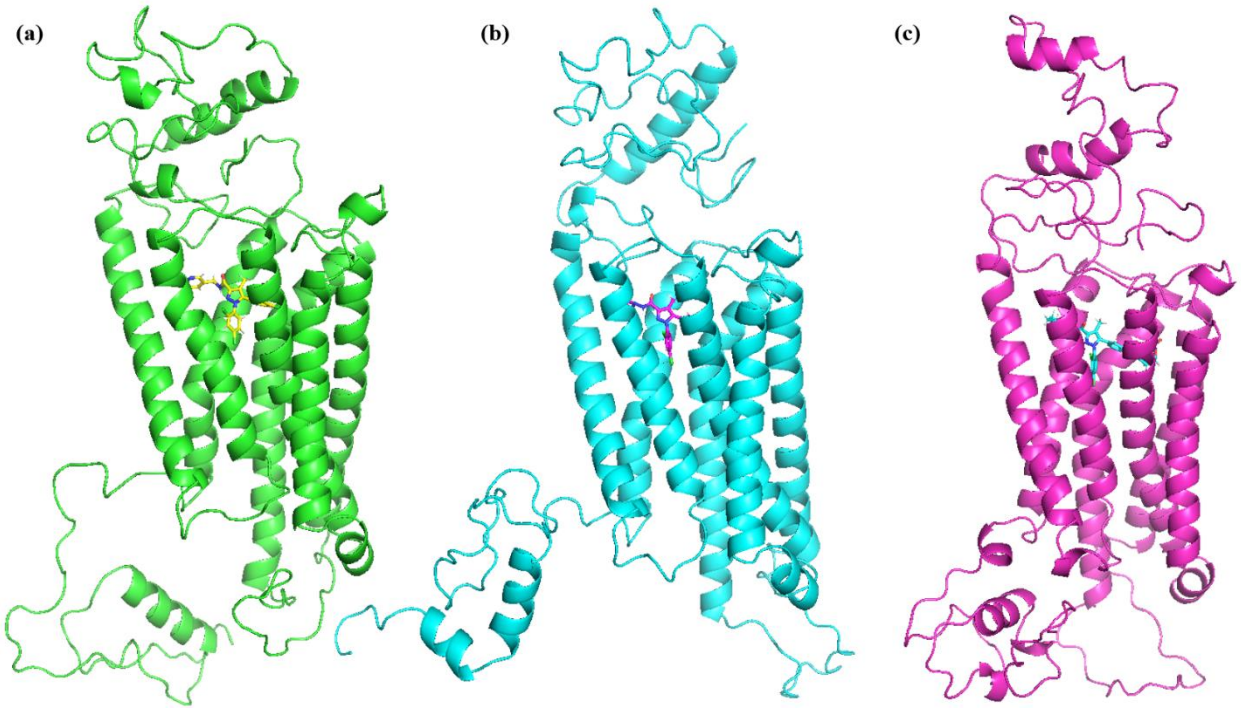


Figure 3.3: Computational models of inactivated CB_1 in complex with three ligands. Visualization of CB_1 (PDB code: 5TGZ) in complex with **(a)** MJ15 **(b)** AM251 **(c)** AM6538

For each ligand-protein complex, several poses were generated and scored; however, as described in the methods section, only the pose that showed good agreement with the binding mode of the ligand in the crystal structure (used as a reference) was selected. This is a reasonable assumption given the high degree of similarities in the chemical structures of the ligands selected for this work. As shown in Figure 3.1, all three ligands displayed a binding pose that closely resembled the reference pose. For instance, (**Figure 3.4**) AM6538's arm 1 lies in a side pocket formed by helices II, III, VI and VII. Within the docked structure, arm 1 was observed to form electrostatic interactions with Val196$^{3.32}$, Cys386$^{7.42}$, Leu387$^{7.43}$ and Gly166$^{2.53}$ (bond distance = 3.36Å), which agree with the crystallographic structure interactions. Arm 1 was also seen forming additional electrostatic interactions, beyond that of the crystallographic structure, with the hydroxyl group in Ser390$^{7.47}$ and Ser173$^{2.60}$ (bond distance = 2.39Å, 2.93 Å, respectively). Moreover, Trp356$^{6.48}$ and Phe170$^{2.57}$ were seen mediating π-π interactions with arm 1. However, unlike the crystallographic structure, our docked AM6538 was not observed to form π-π and hydrophobic interactions with Ser167$^{2.54}$ and Met103$^{N\text{-term}}$. In terms of arm 2, it extended to a narrow channel formed by helices III, V, VI and ECL2.[109] The phenyl group in arm2 established π π interactions with Phe268^{ECL2},

and Trp356$^{6.48}$; in addition to electrostatic interactions formed with several residues, including Leu193$^{3.29}$, Val196$^{3.32}$, Thr197$^{3.33}$, Leu359$^{6.51}$, and Met363$^{6.55}$.[109] The nitrate group was observed forming hydrogen bonding interactions with Thr197$^{3.33}$ (hydrogen bond distance = 3.40Å) and π-π interactions with Tyr275$^{5.39}$. However, in arm2, our system was not observed forming interactions with Phe102$^{N\text{-term}}$ nor is our nitrate forming any hydrogen bond interactions with Tyr275$^{5.39}$ and Trp279$^{5.43}$, which are all interactions reported in the crystallographic structure.[109] Lastly, arm 3 had extended to a gap created by helices I, II, and VII and was capped by the N-terminal loop. Arm 3 formed a combination of π-π and electrostatic interactions with His178$^{2.65}$, Phe177$^{2.64}$, Phe174$^{2.61}$ and Asn101$^{N\text{-term}}$. However, unlike the crystallographic structure, our AM6538 was not seen mediating interactions with the hydrophobic residues Met103$^{N\text{-term}}$, Ile105$^{N\text{-term}}$, Ile119$^{1.35}$, Ser123$^{1.39}$, Phe170$^{2.57}$, Ser383$^{7.39}$ and Met384$^{7.40}$.[109]

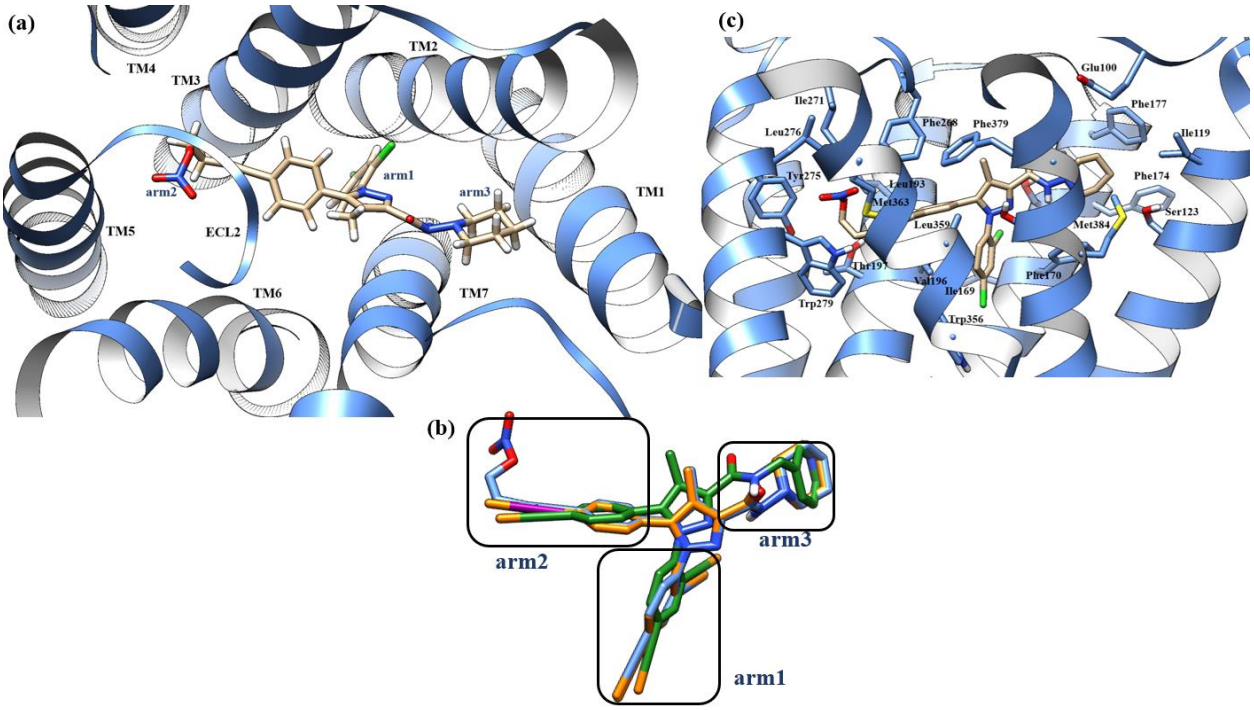


Figure 3.4: Binding location of the inactivating molecules AM6538, AM251 and MJ15. **(a)** Location of AM6538's arms within the CB_1 receptor TMH. **(b)** AM6538 (blue), MJ15 (green) and AM251 (orange) superimposed on each other to demonstrate the similar localization of their arms. **(c)** AM6538's residue interactions formed within the orthosteric pocket, which are similar mediated amongst MJ15 and AM251 ligands.

MJ15 and AM251 were also mediating similar interactions as AM6538, in addition to maintaining the same localized regions for their arm placements. Regarding MJ15's arm 1, it was observed forming π-π interactions with Trp356$^{6.48}$ and Phe170$^{2.57}$ in addition to establishing further hydrogen bonding interactions with the hydroxyl group in Ser390$^{7.47}$, and with residues Ile169$^{2.56}$, Val196$^{3.32}$, Cys386$^{7.42}$. Arm 2 was observed forming hydrogen bonding and π-π interactions with Trp279$^{5.43}$, Met363$^{6.55}$, Thr197$^{3.33}$, Trp3566$^{.48}$, Val196$^{3.32}$, Phe268^{ECL2}, Leu193$^{3.29}$. Regarding arm3, its docked position was seen mediating hydrogen bonding and π-π interactions with Asn101$^{N\text{-}term}$, His178$^{2.65}$, Phe177$^{2.64}$ and Phe174$^{2.61}$. Similarly, AM251's arm 1 was also observed forming π-π interactions with Trp356$^{6.48}$ and Phe170$^{2.57}$, with additional hydrogen bonding interactions formed with the hydroxyl group in Ser390$^{7.47}$ and Ser199$^{3.35}$, and with residues Ile169$^{2.56}$, Val196$^{3.32}$, Cys386$^{7.42}$. Arm2 was also observed forming hydrogen bonding and π-π interactions with Trp279$^{5.43}$, Met363$^{6.55}$, Thr197$^{3.33}$, Trp356$^{6.48}$, Val196$^{3.32}$, Phe268^{ECL2}, Leu193$^{3.29}$, and Leu359$^{6.51}$. Arm3 was also forming hydrogen bonding interactions and π-π interactions with Asn101$^{N\text{-}term}$, His178$^{2.65}$, Phe177$^{2.64}$, Phe174$^{2.61}$, Ser173$^{2.60}$. [109]

The binding energy for the docked complexes of AM251, AM6538, and MJ15 against CB_1 are -8.2 kcal/mol, -8.3 kcal/mol, and -7.9 kcal/mol, respectively. However, it should be noted that the docking scores are only based on the interactions within the individual static poses and do not account for conformational flexibility upon complex formation. Therefore, we optimized the ligand-receptor complexes in their physiological environment of a lipid bilayer and solvents using classical MD simulation.

3.3.3 MD Simulation and Structural Stability Analyses of the Ligand-Bound CB_1 Complexes

Each complex was run for 50 ns of classical MD simulation with NAMD to optimize the system and assess its dynamic behaviour and molecular interactions. The RMSD (**Figure 3.5**) was calculated over the course of the entire simulation for the protein backbone to probe the stability

of the systems during simulation.

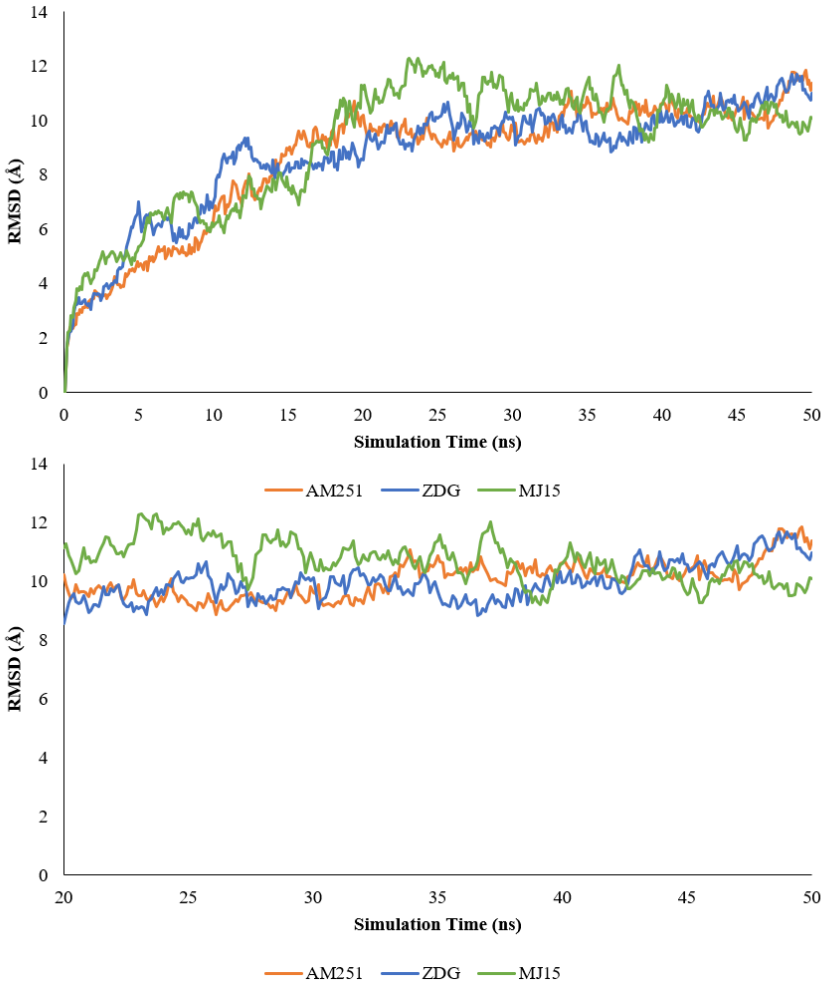


Figure 3.5: RMSD graph for the 50 ns classical MD simulation for AM251, ZDG and MJ15. The last 30 ns were chosen based on the stability of the systems at this time. RMSD was plotted based on the protein backbone atoms: carbon, nitrogen, oxygen, and α-carbon.

Results from the RMSD plots of the receptor backbone revealed that all the complexes underwent some structural changes during the initial 20 ns of MD simulation which is reflected in the linear increase in the RMSD values in this stage. Nevertheless, the RMSD values reached a plateau after 20 ns and remained stable until the end of the simulation, which is an indication of the system's achieved equilibrium. This is apparent as the average deviation for the MJ15, AM251 and AM6538 complexes during the last 30 ns of MD simulation were observed to be 0.55Å, 0.50Å and 0.50Å, respectively. It was also noted that the AM6538- CB_1 complex reached stability at ~ 9.98Å from

the starting conformation; whereas the AM251- CB_1 complex and the MJ15-CB_1 complex stabilized at 10.04 Å and 10.70Å from their respective starting structures. Therefore, the AM6538-CB_1 complex displayed the most stable behaviour.

To observe specific residue fluctuations and regions of high dynamics across the trajectory, RMSF plots (**Figure 3**) were calculated based on the protein backbone.

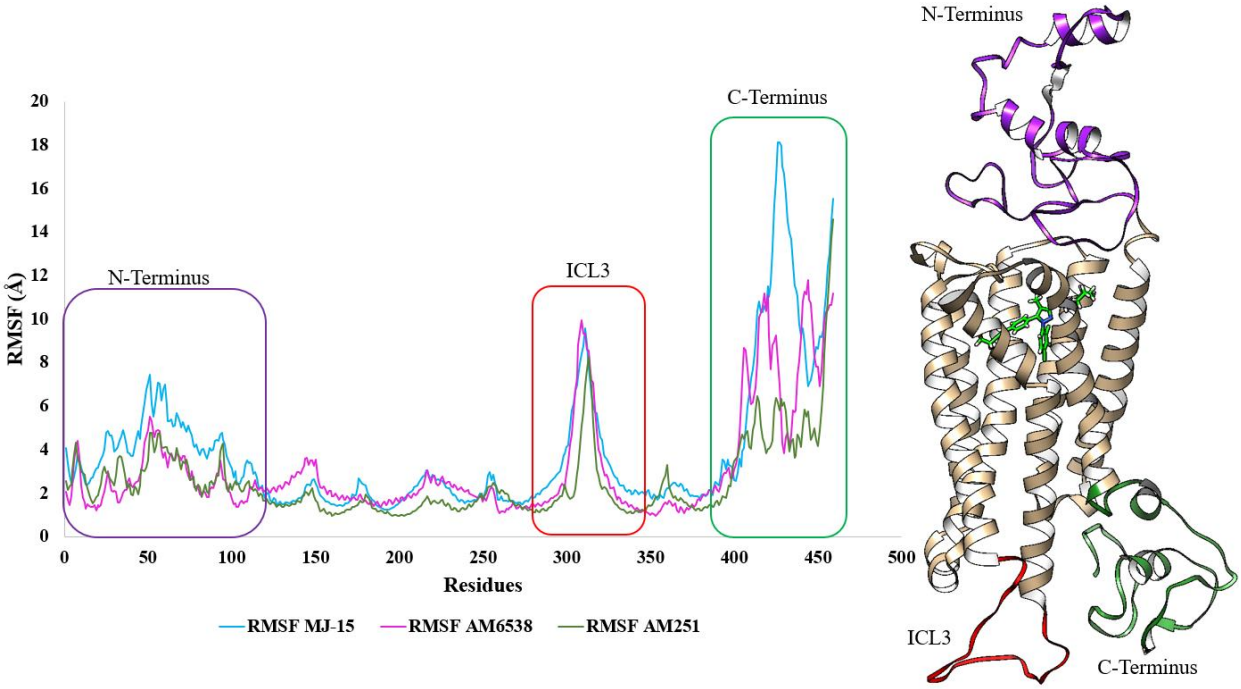


Figure 3.6: RMSF plot of all three inactivating systems, and their respective residue fluctuations. Regions experiencing high dynamics are all indicated in a square. Dynamic regions in the CB_1 protein (visualized with co-crystalized ligand AM6538) are shown.

Plots for each complex were generated and compared for the last 30ns of the classical MD trajectory. Regions of high fluctuations were comparable with other reported *in silico* studies [106,107,145]. The highly fluctuating regions in all three systems were mainly observed in the N-terminus (residues M1-Q116), the ICL3 (residues K300-T344) and the C-terminal (residues R400-L472) regions. In addition, other intracellular and extracellular loops also displayed some variations. These are not surprising as the loops are often flexible and prone to change, whereas the N-terminal and C-terminal regions included more disordered regions. Comparison between the different systems revealed that the MJ15 complex was most dynamic in its N- and C-termini, whereas the other two ligands exhibited a slightly lesser degree of fluctuations in these regions.

Nevertheless, the dynamics of ICL3 were almost similar in all the ligand-bound systems with a fluctuation of ~10 Å. This indicates that the nature of the ligand probably renders the least impact on this large ICL3. Unsurprisingly, all systems maintained stable dynamics in their TM regions (between residues ~117-400 in **Figure 3.6**).

Despite there being many unstructured segments in the N-terminus, the RMSF values corresponding to this N-terminus across all systems, experienced fewer fluctuations (~10 Å to ~4 Å reduction) than in comparison to the C-terminal region. To access the reason for this, further exploration of the N-terminus and its impact was analyzed throughout the simulation. Interactions introducing stability were observed and compared to a previous work reported by Jakowiecki *et al*., [106] to identify any conserved atomic interactions. Independent of the structural composition of the N-terminus, each complex's N-terminal domain adopted a similar positioning in respective to the orthosteric pocket (**Figure 3.7 and Figure 3.8**).

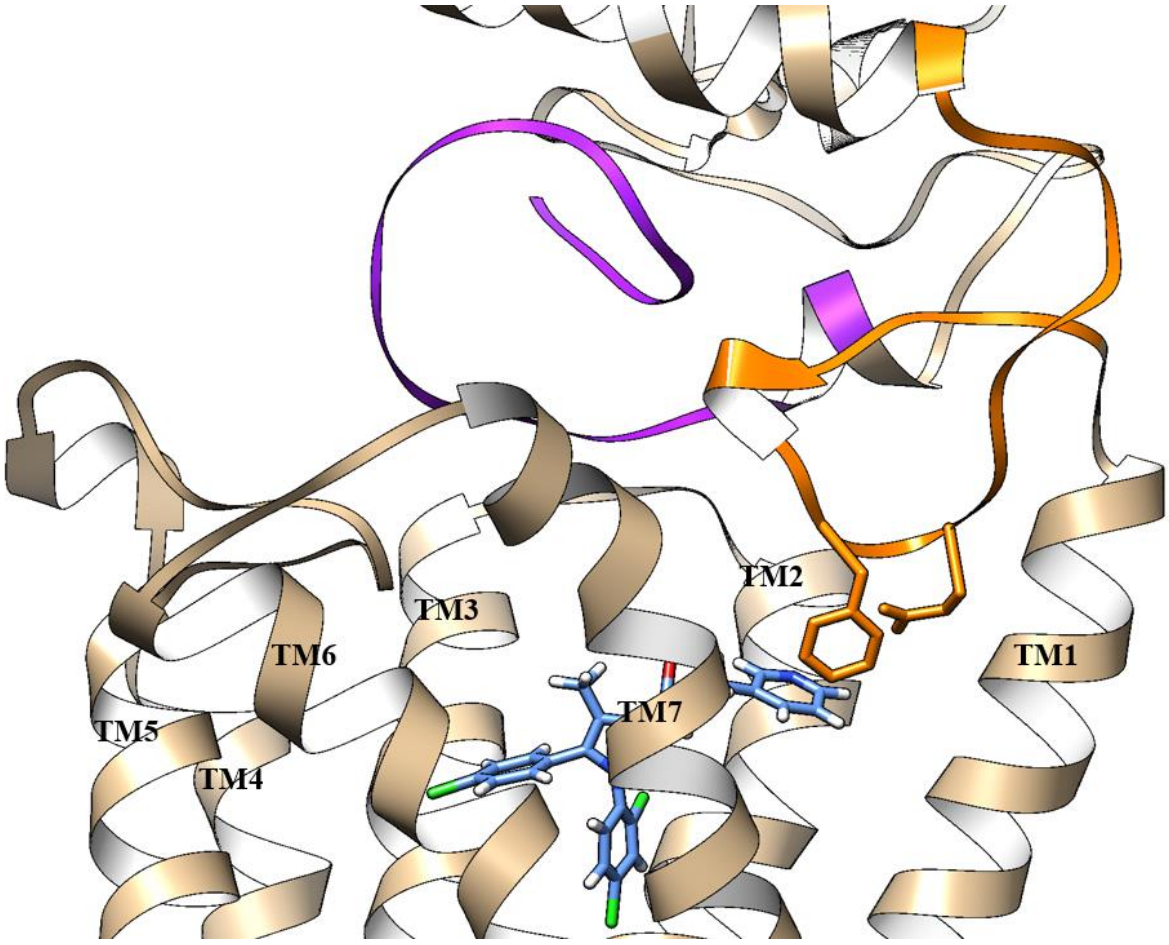


Figure 3.7: N-terminal domain placement in relation to the orthosteric site and the ECLs. MJ15 is visualized in the orthosteric pocket, where the MPR (orange; residues 90-110) are forming a 'V-shape' into the pocket, and the N-terminal cap (purple; residues 1-20) is situated over the orthosteric pocket forming stabilized interactions with the ECLs.

For all three compounds (MJ15, AM6538 and AM251), a portion of the N-terminus (Gln97 to Leu111) was projecting into the gap between TM1 and TM7, which has also been proposed numerous times to be the gating channel to the orthosteric pocket from the extracellular solvent. This suggests that the N-terminus may play a role in the access channel for ligand binding to the orthosteric site. The N-terminus region spanning residues Met1-Thr18 was also seen projecting itself near the orthosteric pocket, similar to the findings by Jakowiecki *et al.*, [106]. This conformation of the N-terminal region potentially acts as a method to extract the solvent from the binding pocket.[106] A network of salt bridge interactions forming within the N-terminus and between the

ECLs, were observed in all systems. The presence of these interactions is proposed to help orient the N-terminus near the orthosteric pocket and support the stability of ECLs and the N-terminus.

In the AM251 system, intra- N-terminal electrostatic interactions were seen forming between Glu31-Lys34, and Asp32-Lys2, in addition to Glu93-Lys90 which were mediating sporadic salt bridge interactions for a shorter amount of time during the simulation. Whereas the salt bridge interactions conserved between Asp6-Lys373, Asp19-Lys40, Asp27-Lys183, Asp36-Lys40, and Glu106-Lys376 were more stabilized and were seen for most of the trajectory.

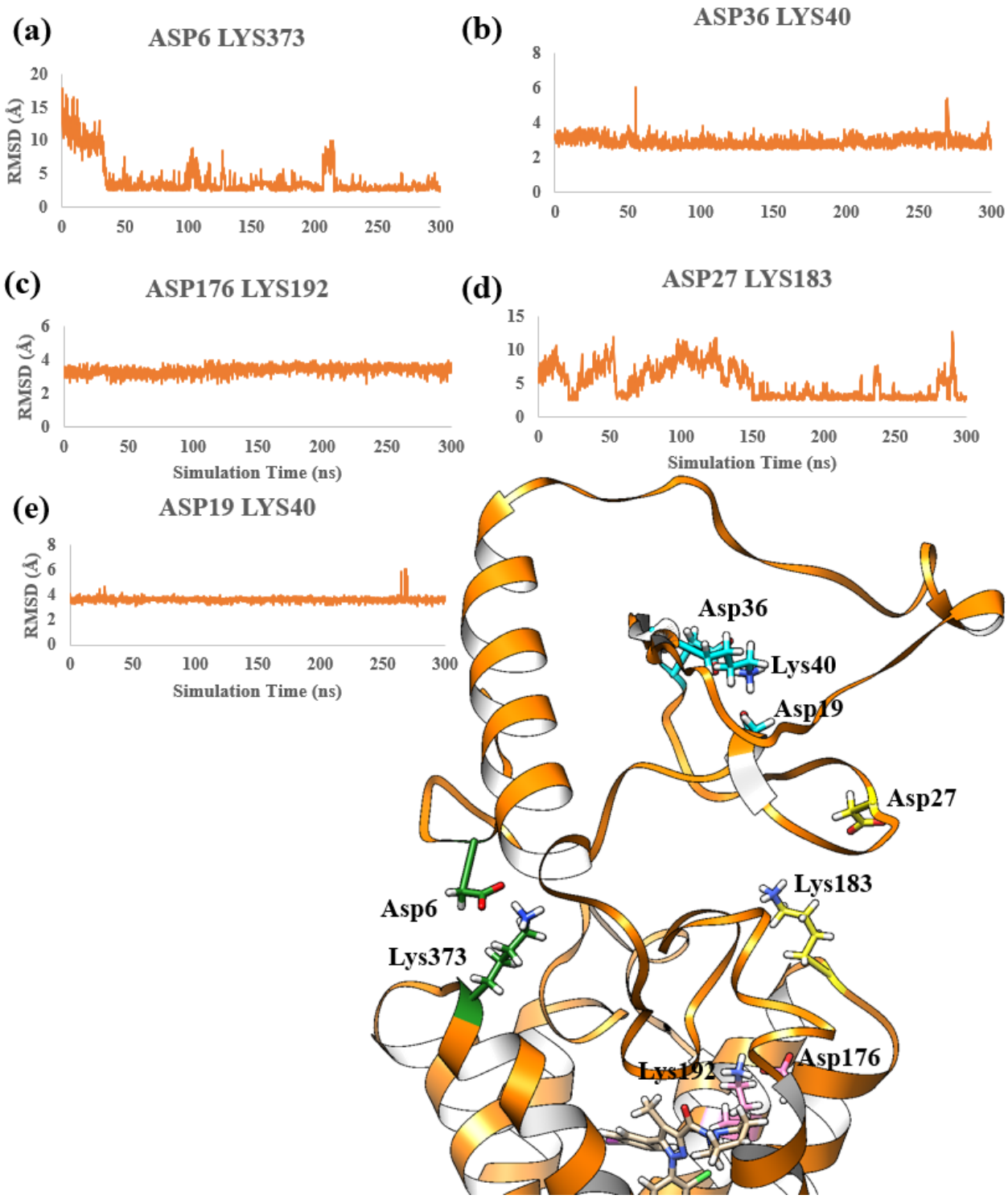


Figure 3.8: Salt bridge formation throughout the N-terminus adding to the structural component's stability for MJ15. The last 30 ns of trajectory are visualized as a trajectorial evolution plot (left), and bond formation in the N-terminus (right). N and O distances are plotted

(orange) between the two residues. **(a)** Asp 6-Lys3 (green), **(b)** Asp36-Lys40(cyan), **(c)**Asp176-Lys192(pink), **(d)** Asp27-Lys183(yellow) **(d)** Asp19-Lys40 (cyan).

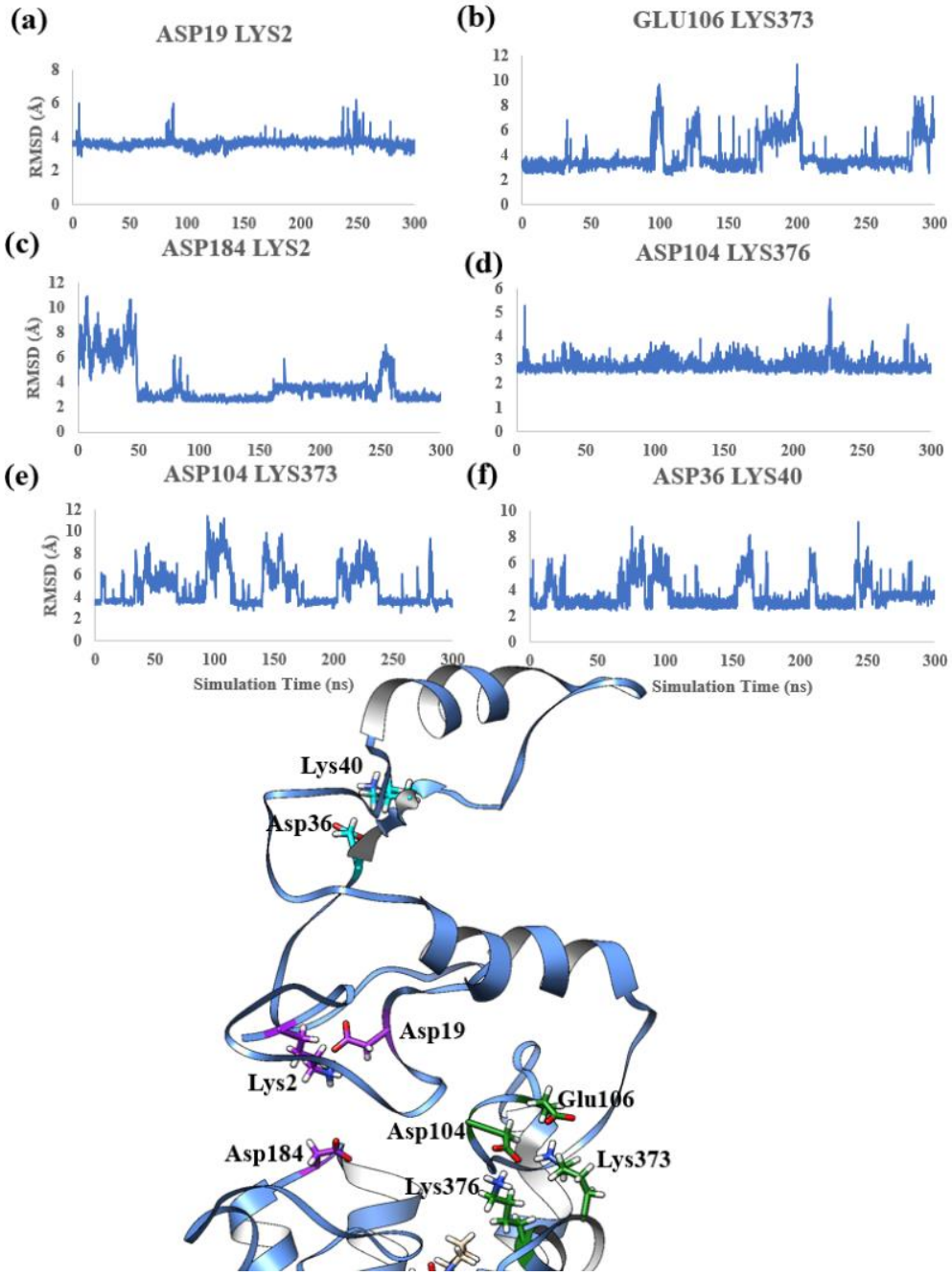


Figure 3.9: Salt bridge formation throughout the N-terminus adding to the structural component's stability for AM6538. The last 30 ns of trajectory are visualized as a trajectorial evolution plot (left), and bond formation in the N-terminus (right). N and O distances are plotted

(blue) between the two residues. **(a)** Asp19 – Lys 2 (purple). **(b)** Glu 106- Lys 373 (green) **(c)** Asp184-Lys2 **(d)** Asp 104-Lys 376 (green) **(e)** Asp104-Lys373 (green), **(f)** Asp36-Lys40 (cyan).

For the CB_1-AM6538 complex, Asp19-Lys183, and Glu93-Lys90 were mediating shorter salt-bridge interactions, whereas Asp19-Lys2, Asp36-Lys40, Asp104-Lys373, Asp104-Lys376, Asp184-Lys2 and Glu106-Lys373 exhibited stronger interactions encompassing most of the trajectory. The MJ15 system was observed to incorporate shorter formed salt-bridge interactions Asp19-Lys183, Asp104-Lys376, and Glu31-Lys40 with each other. Asp-Lys373, Asp19-Lys40, Glu31-Lys183, Glu80-Lys84, Glu93-Lys90, and Glu106-Lys373 were having more stabilized salt-bridge interactions spanning the majority of the MD trajectory.

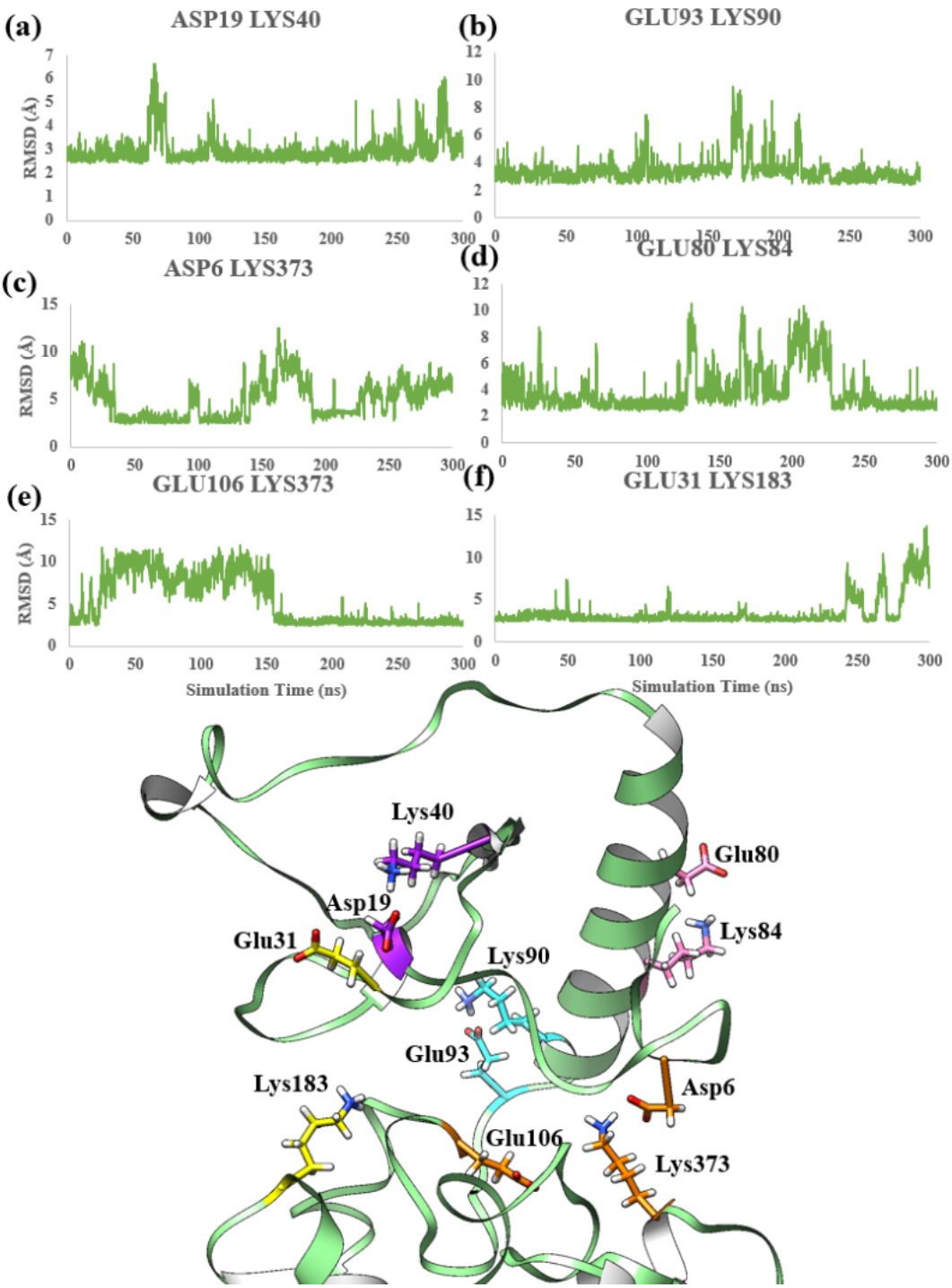


Figure 3.10: Salt bridge formation throughout the N-terminus adding to the structural component's stability for MJ15. The last 30 ns of trajectory are visualized as a trajectorial evolution plot (left), and bond formation in the N-terminus (right). N and O distances are plotted between the two residues. **(a)** Asp19-Lys40 (purple) **(b)** Glu93-Lys90 (cyan) **(c)** Asp6-Lys373(orange) **(d)** Glu80-Lys84 (pink) **I** Glu106-Lys373(orange) **(f)** Glu31-Lys183 (yellow)

The average distance between the strong salt-bridge contacts from the last 30 ns MD simulation averaged 3.07 Å, 3.77 Å, and 4.05 Å for CB_1 receptors bound to AM251, AM6538 and MJ15, respectively. This explains the superior stability observed for the N-terminal domain in the AM251-CB_1 complex.

3.3.4 Structural features of the N- and C-terminal domain of CB_1

Since the N-terminal and C-terminal domains were separately modelled and attached to the inactivated TM state of human CB_1, we analyzed the secondary structural composition of these regions before and after MD simulation to observe their evolution. Both the terminal segments were mostly unstructured with some levels of helical contents. In the initial structures of the complexes, at the N-terminal end, all the ligand-bound models exhibited two helices (dubbed here as NTH2 and NTH3). NTH2 was formed by ~7 (±3) residues spanning Ser56- Met62, whereas NTH3 was composed of ~15 (±3) residues involving Asn77-Glu91. In addition, we found an extra four residues-long helix (dubbed NTH1 based on the sequence numbering) in the complexes of MJ15 and AM6538; but the position of this helix was not conserved in the two complexes. In the MJ15 complex, the helix was formed by residues Thr18-Tyr22, whereas this helix encompassed residues 37-41 in the AM6538 system. The presence of the helices was in good agreement with a previous *in silico* work by Jakowiecki and colleagues[106] on CBD and Δ^9-THC bound CB_1 models that did not include the C-terminal domain. While the positioning of NTH2 and NTH3 reported in this work were in good agreement with our models, the NTH1 in the earlier work was seen to be present at residue position 30-35 which is consistent with the observation in our AM6538 system. Nevertheless, the previous work[106] reported that NTH2 became unstructured during the MD simulation. We observed the same behaviour in our simulations as well. While the NTH3 was maintained throughout the simulation, NTH2 disappeared and the NTH1 helix mostly remained stable only in the AM6538 complex. Thus, our simulation suggests the structural roles of NTH3 in the dynamics of the CB_1 N-terminal domain.

On the C-terminal end, we found the presence of a slightly larger helix (helix 9) that was approximately formed by residues Val422 to Lys455 in the ligand-bound complexes in this work except for the AM251 complex, in which this helix was formed by residues Lys455 to Val463. Helix 9 remained mostly stable in our models during the MD simulation. Earlier, Ahn *et al*., (2009)[158] employed circular dichroism spectroscopy and characterized the residues A440-M461

in human CB_1 formed an amphipathic helix. In another study by Singh *et al*., (2019)[46] modelled the segment spanning residues 441-455 in human CB_1 and characterized it as a helix. Thus, the key helical content observed in our models agrees with previous reports. Cumulatively, our model represents the first full-length inactivated CB_1 structure that extends molecular-level insights.

3.3.5 Binding Free Energy Calculations

The binding-free energies of the ligand-protein complexes (**Table 3.3**) were calculated and averaged over the last 30 ns timeframe of their respective MD trajectory.

Table 3.3: Conversion of antagonist/inverse agonist ligand's K_i values into Kcal/mol. Calculation of experimentally determined K_i values for AM6538, MJ15 and AM251 (put the assay) to binding free energy (ΔG_{bind}) in kcal/mol. Temperature (T) was set to 310K, and R, the gas constant, was set to 0.0831446261815324 $L \cdot bar \cdot K^{-1} \cdot mol^{-1}$.

Ligand Molecule	**K_i/K_d value(experimental)**	**Kcal/mol (calculation)**	**ΔG from MM-PBSA calculations (kcal/mol)**	**Standard Deviation (kcal/mol)**	**Reference**
AM6538	K_i = 0.038 nM	-84.29	-52.2	3.68	[109,117]
MJ15	K_i = 0.0272 nM	-92.01	-41.04	3.77	[124]
AM251	K_d = 0.23nM-0.8nM	-37.88 to -5.75	-37.38	4.84	[119,120,121,122]

The results indicated that AM6538 exhibited the strongest binding affinity (with a ΔG value of -52.2 kcal/mol) among the three ligands in this study. MJ15 exhibited a binding affinity of ~41 kcal/mol toward our CB_1 model, whereas AM251 displayed the weakest affinity against the receptor. Experimental data suggest that AM6538 and MJ15 exhibited almost similar inhibition kinetics against CB_1; while the AM251 shows the weakest affinity, which fairly correlates with our MM-PBSA values. Although the binding pose and affinity of the AM6538-CB_1 complex are in good agreement with the experimental data (**Table 3.3**), it should be acknowledged that the calculation might have been influenced by the fact that the X-ray crystal structure used to build

CB_1 in this work was originally co-crystallized with AM6538. Previous studies by Jung *et al*., (2018)[145] and Loo et al., (2019)[116] stated that the crystallographic structures introduce a bias, where the positioning of their structural components is more favourable for their co-crystallized ligands, as well as molecules that are within the same class of the co-crystallized ligand. Limitations in the crystallographic resolution, create this bias whereby the crystal structures represent a conformation that is primarily favourable for the binding and co-crystallization of a single ligand. Nevertheless, our approach was still able to discern the strong and weak affinity binders amongst the three molecules selected for this study.

To identify the nature of interactions that contributed to the binding free energies of the complexes, we computed the non-bonded interaction energies between ligand and the surrounding amino acid residues in the orthosteric binding site of CB_1 (**Figure 3.11**; for residue-ligand interactions see **Figure 3.4**). These calculations were carried out using the NAMD Energy plugin from VMD.[165] Results across each inverse-agonist system revealed that residues Phe170, Lys192, Val196, Phe268, and Ser383 are involved in strong non-bonding contacts with the bound ligand, in which most of these interactions are electrostatic-hydrogen bonding interactions. Residues Phe170, Phe174, and Phe177 were also observed forming π-π stacking with the ligands. In terms of the N-terminal residues situated within the MPR, residues such as Glu100, Asn101 Phe102 and Met103 were observed mediating π-π interactions with some of the ligand in our dataset. More favourable MPR interactions were displayed within the AM251 and MJ15 systems, due to their positioning of the N-terminal domain, which was orientated closer to the orthosteric pocket. These results show the importance and impact of the MPR in the binding of antagonists or inverse agonists. In essence, the MPR contribution to ligand binding is highly dependent on the conformation of the ligand-bound at the receptor and within the orthosteric pocket. From these results, we can conclude that modelling the entire MPR is important when analyzing all residue-ligand interactions. Thus, incorporation of the N-terminal domain in CADD and high through-put virtual screening may be necessary.

(a)

Residues	Elec	VdW	Nonbond
100	18.55017	-1.9312	16.61898
101	-0.33673	-1.33097	-1.33097
102	0.639319	-1.50074	-0.86142
103	0.059427	-0.0301	0.029324
119	-0.2695	-0.85647	-1.12597
120	-0.31561	-0.94958	-1.26519
123	-0.34593	-1.20823	-1.55416
166	-1.92126	-1.20276	-3.12402
169	0.06198	-1.78939	-1.72741
170	-0.55841	-6.1299	-6.68831
173	-2.952	-3.26125	-6.21325
174	0.117938	-2.38178	-2.26384
177	0.187494	-2.05844	-1.87095
192	-1.96412	-2.19955	-4.16366
193	-0.144	-1.53991	-1.68391
196	0.347182	-3.93055	-3.58337
268	-0.07743	-3.48918	-3.56661
271	-0.14392	-0.3326	-0.47652
276	-0.03141	-0.08497	-0.11637
279	0.145544	-0.66189	-0.51635
356	-0.26499	-0.80108	-1.06608
359	0.156268	-1.19565	-1.03938
363	-0.0902	-0.97147	-1.06167
379	-0.98694	-1.62008	-2.60703
383	-2.30062	-2.44537	-4.74599
384	-0.51918	-0.51142	-1.0306
386	1.322904	-1.53528	-0.21238
387	1.034856	-1.75186	-0.717
390	0.062861	-0.17299	-0.11012

(b)

Residue	Elec	VdW	Nonbond
100	-0.12782	-0.22743	-0.35525
101	-0.62986	-0.56095	-1.19081
102	0.143845	-0.07054	0.073306
103	-0.01331	-0.005	-0.01831
119	-0.10319	-0.07366	-0.17684
120	-0.14152	-0.09626	-0.23778
123	-0.40701	-0.37095	-0.77795
166	-0.60225	-0.78097	-1.38323
169	0.382418	-1.35972	-0.9773
170	-0.44133	-4.996	-5.43733
173	0.750395	-2.71454	-1.96414
174	0.173776	-2.51354	-2.33976
177	-0.11272	-2.29842	-2.41114
192	-6.81486	-2.18257	-8.99743
193	0.169321	-1.71796	-1.54864
196	-0.06079	-3.91093	-3.97172
197	0.065278	-0.44527	-0.37999
268	-0.07283	-1.59838	-1.6712
271	-0.00487	-0.0056	-0.01047
275	-0.0005	-0.01776	-0.01826
276	-0.01493	-0.04567	-0.0606
279	-0.0121	-0.39069	-0.40279
356	-0.48672	-1.04136	-1.52808
359	0.308949	-1.23643	-0.92749
360	9.4E-05	-0.11469	-0.11459
363	0.00314	-0.72627	-0.72313
379	-1.27478	-1.67716	-1.67716
380	-0.8626	-0.21745	-1.08005
383	-1.47878	-2.35324	-3.83201
384	-0.41931	-0.43543	-0.85474
386	1.23917	-1.0878	0.151366
387	0.933193	-1.32829	-1.32829
390	0.117337	-0.25193	-0.13459

(c)

Residue	Elec	VdW	Nonbond
100	-0.12782	-0.22743	-0.35525
101	-0.62986	-0.56095	-1.19081
102	0.143845	-0.07054	0.073306
103	-0.01331	-0.005	-0.01831
119	-0.10319	-0.07366	-0.17684
120	-0.14152	-0.09626	-0.23778
123	-0.40701	-0.37095	-0.77795
166	-0.60225	-0.78097	-1.38323
169	0.382418	-1.35972	-0.9773
170	-0.44133	-4.996	-5.43733
173	0.750395	-2.71454	-1.96414
174	0.173776	-2.51354	-2.33976
177	-0.11272	-2.29842	-2.41114
192	-6.81486	-2.18257	-8.99743
193	0.169321	-1.71796	-1.54864
196	-0.06079	-3.91093	-3.97172
197	0.065278	-0.44527	-0.37999
268	-0.07283	-1.59838	-1.6712
271	-0.00487	-0.0056	-0.01047
275	-0.0005	-0.01776	-0.01826
276	-0.01493	-0.04567	-0.0606
279	-0.0121	-0.39069	-0.40279
356	-0.48672	-1.04136	-1.52808
359	0.308949	-1.23643	-0.92749
360	9.4E-05	-0.11469	-0.11459
363	0.00314	-0.72627	-0.72313
379	-1.27478	-1.67716	-1.67716
380	-0.8626	-0.21745	-1.08005
383	-1.47878	-2.35324	-3.83201
384	-0.41931	-0.43543	-0.85474
386	1.23917	-1.0878	0.151366
387	0.933193	-1.32829	-1.32829
390	0.117337	-0.25193	-0.13459

Figure 3.11: Electrostatic and van der Waals interactions mediated with CB_1 receptor modulators. Inter-atomic ligand-receptor interactions mediated with **(a)** AM251 **(b)** MJ15 and **(c)** AM6538.

Overall, the residues observed interacting with each ligand, agree with the same mediated interactions in the paper for the crystallographic structure 5TGZ by Hua *et al.*,[109] helping to validate our models.

3.3.6 K-Means Trajectorial Clustering Algorithm

To understand the dominant structural conformations adopted throughout the trajectory, while eliminating noise or less prominent/ higher energy adopted conformations within the trajectory, the K-means clustering algorithm was applied to the last 30 ns of the classical MD trajectory. 50 clusters were generated for each system utilizing 49 hyperparameters with a stepwise index of 1 from an initial cluster size of 2. The best clusters were picked based on a low DBI and a plateau SSR/SST(R^2) value. Based on the criterion (**Table 3.4**, and **Figure 3.12**), clusters 38, 24 and 37 were picked to be the most optimal clusters for ligand complex MJ15, AM6538 and AM251.

Table 3.4: Associated statistic clustering results for each protein-ligand system. Complexes and their associated best clusters sizes, as well as the corresponding cluster size DBI, pSF value(s) and SSR/SST score.

Ligand	Cluster # chosen	DBI	psF	SSR/SST (R^2)
MJ15	38	1.303014	749.138875	0.903455
AM6538	24	1.349322	1074.238504	0.892499
AM251	37	1.465873	510.422871	0.861141

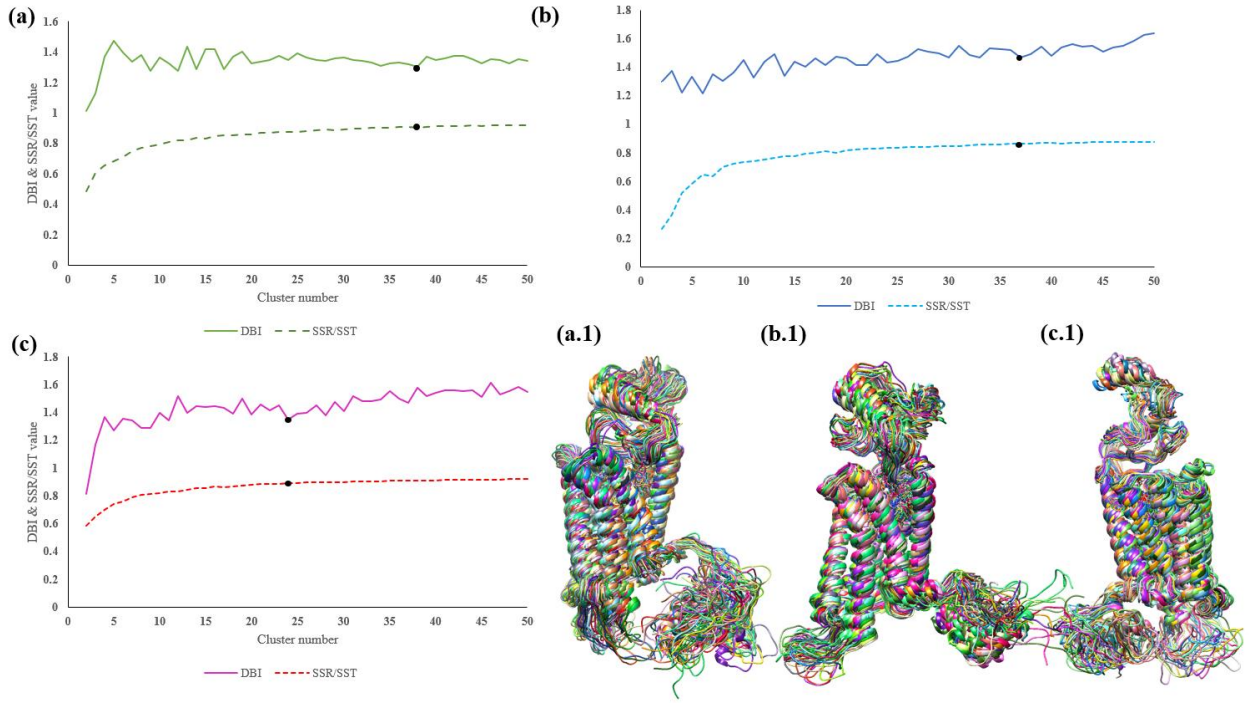


Figure 3.12: Results from K-means clustering algorithm for the last 30ns of the trajectory. Visualization of the DBI and R^2 plot as well as the corresponding clusters for CB_1 ligand systems **(a), (a.1)** MJ15 **(b), (b.1)** AM251 and **(c), (c.1)** AM6538, respectively.

The top three dominant conformations produced by each ligand highlighted the predominant conformations formed between the inverse-agonists/antagonists and the N-terminus, as well as the residues in the orthosteric pocket as seen in **Figure 3.7**, which could be further analyzed in the trajectories. In essence, these interactions present within these dominant conformations further explained the stability of the complex through a static image presenting the network of predominantly formed inter- and intra-molecular interactions.

The frames encompassing the best representative structures for each complex, AM6538, AM251 and MJ15, were between frames: 1521 to 1829 (152-182 ns); 692-873 (69-87 ns), and 329-670 (32-67 ns), respectively. AM251 and MJ51 best representative structure is generated from the beginning of the simulation, whereas AM6538 is extracted from the middle of the last 30 ns from the MD simulation.

3.3.7 Principal Component Analysis

To assess the fluctuations of the system, each system was subjected to PCA for the last 30 ns of the trajectory (**Figure 3.13**). The degree of dynamics captured in each principal component (PC) space, as well as the protein backbone fluctuations for the corresponding eigenvalues, were assessed.

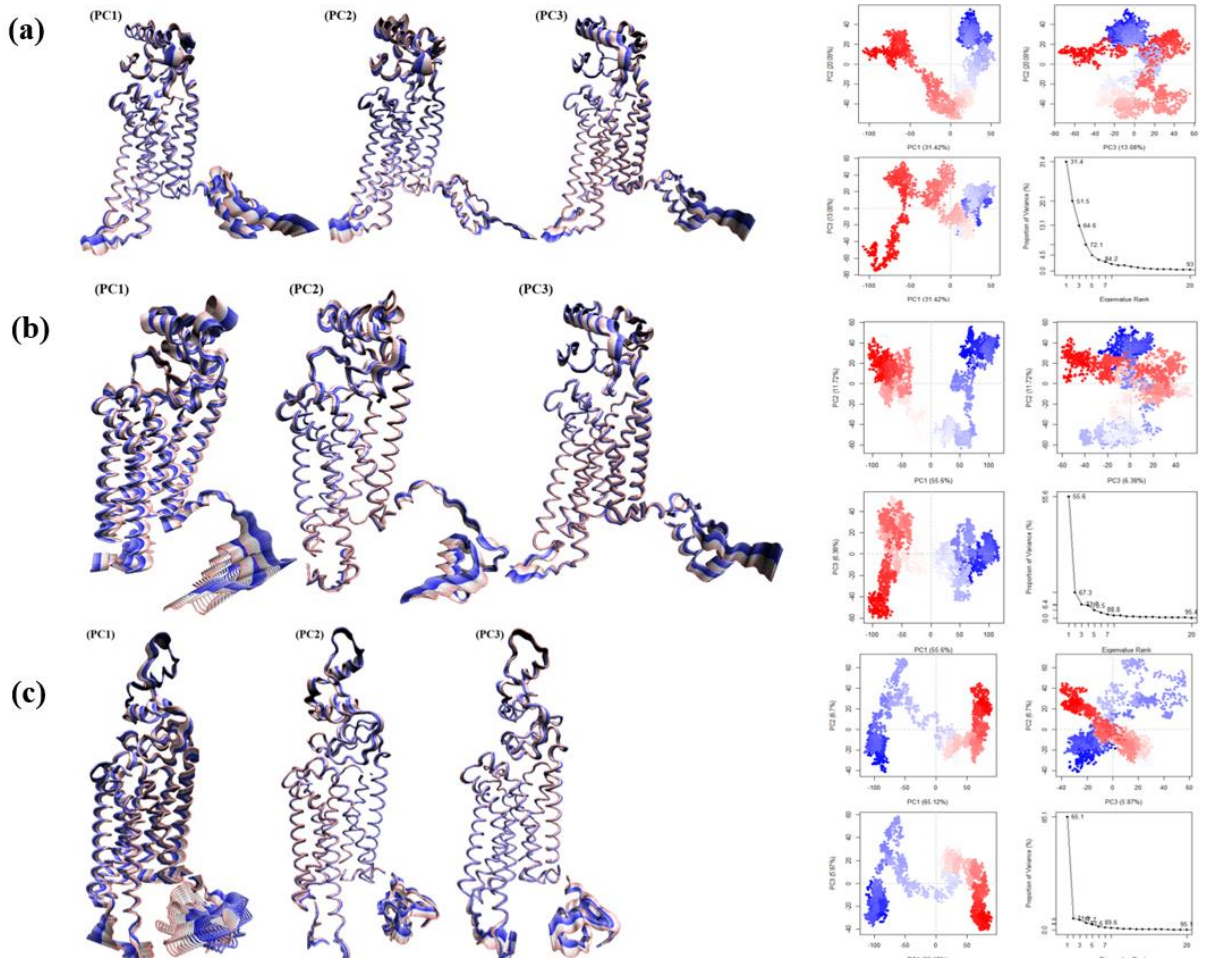


Figure 3.13: PCA analysis and the fluctuations in the backbones, for each corresponding PC space explored. Visualizations of the dynamics captured in the protein backbone for each PC, in addition to their corresponding PCA plot, are depicted for systems containing the ligands **(a)** AM251 **(b)** MJ15 and **(c)** AM6538

When analyzing the dynamics captured in the system MJ15, the 1st eigenvalue contained over half of the trajectorial fluctuations within the system (55.6%), allowing for half of the fluctuations in the system to be explained by the first PC. The 3rd eigenvalue, however, encompassed 73.3% of the system's total proportional variance, thus capturing most of the dynamic conformations adopted in the protein-ligand complex, allowing for a deeper understanding of the system. In contrast, complex AM251's 1st eigenvalue embodied a small amount of the system's variance, capturing only 31.4% of the total proportional variance within the system. The first four eigenvalues, however, adopted most of the protein's conformational exploration within the trajectory, making up 72.1% of the total variance. Overall, it is evident that AM251's system

explores more conformations unrelated to one another, thus capturing fewer dynamics in each of its subsequent eigenvalues. For the complex AM6538, similar to the MJ15 system, the 1st eigenvalue encompasses most of the dynamics formed throughout the trajectory, capturing 65.1% of the total proportional variance. Due to most of the proportional variance being contained within that first eigenvalue, this means that most explored conformations are happening in the first PC space. As a result, the following eigenvalues capture a meniscal number of explored conformations, to which the 3rd eigenvalue contains 71.8% of the total proportional trajectorial variance, which is only 6.7% more captured variance than the 1st eigenvalue.

To get a better understanding of the types of fluctuations adopted in each eigenvalue, the backbone of the complex was visualized for the first three principal components (**Figure 3.13**). For the MJ15 system, the protein backbone for PC1 was seen mediating most of its dynamics in the C-terminus (residues Glu-Leu472), the ICL3(Ala305-Ala335) and the N-terminus (Met1-Phe44), in which the most of the proteins' dynamics in the TM region, the ICL3, N- and C-termini are all captured. PC2 encompassed fewer fluctuations than in comparison to PC1, where most of the dynamics are mediated in the extracellular and intracellular segments, such as the N-terminal region (residues Met1-Leu111), the ECL2(residues Trp255-Ile267), the ICL3(Ala305-Pro332) as well as the C-terminal domain (residues Ala407-Leu472). ECL2, large fluctuations are due to it being the largest ECL domain in the CB_1 protein. In addition, the ECL2 has a large role in ligand binding, in which a high degree of flexibility is adopted in this region when interacting with ligand MJ15. PC3 possesses the least number of fluctuations, where most of the dynamics in this eigenvalue are between the C-terminal region (Pro402-Leu472), the and ICL3(Arg307-Val329) and the N-terminal helices (Asp36-Asn95). When analyzing the protein backbone fluctuations in the system containing the ligand AM251, all three of its PCs were mediating the same level of dynamics relative to each other. PC1 was having most of its dynamics within the C-terminus (Ala407-Leu472), the ICL3 (Ala305-Arg331), and the unstructured regions of the N-terminus (Met1-Gln116). PC2 fluctuations were within the C-terminus (Ala407-Leu472), ICL3 (Thr313-Arg331), and the N-terminus (Met1-Gln116), in which the N-terminal helices' dynamics were mostly being captured. In contrast, the C-terminal domain dynamics were captured to a lesser degree than in the previous PCs. PC3 fluctuations were mediating most of its dynamics in the C-terminal region (Ala407-Leu472), the ICL3(Gly312-Arg331) and mainly within the helices of the N-terminus (Met1-Glu100). For the system bound to AM6538, PC1 was encompassing most of the dynamics

in the protein backbone, similar to that of the CB_1-MJ15 bound complex, in which the entire protein's dynamics in the TM region, the ICL3, N- and C-termini dynamics are all captured. However, the regions mediating the most dynamics were apparent in the C-terminus (Ala407-Leu472), and the ICL3(Met308-Arg335), where the N-terminal dynamics were empowered with the TM segment. The dynamics in PC2 were significantly reduced, however, the most dynamic regions were observed within the C- (Ala407-Leu472) and N-termini (Met1-Ala118), whereas the TM region was very stabilized. PC3 was similar to PC2, where very few fluctuations were being captured. Likewise, in this eigenvalue, the C-terminus (Ala407-Leu472), the N-terminus (Asp36-Asn77), and the ICL3(Ala305-Asp333), had experienced the highest fluctuations, whereas the TM segment was very stabilized. Overall, across all systems, most fluctuations were within the N- and C- termini, as well as the ICL3, and only smaller variations were captured in the TM region. Understandably, the TM segment exhibited a higher degree of stability due to the formation of rigid secondary structures supported by inter-atomic interactions between TM helices. The stability of the TM was also further enhanced because of the bound ligand.

3.3.8 Residue Dynamics Correlation

To extract the dynamic relationships between the different structural components within the CB_1 receptor, dynamic cross-correlation plots were generated for each complex. The cross-correlation plots revealed the links between the external and internal segments with the TM component. In terms of the CB_1 bound MJ15 system's (**Figure 3.14**) cross-correlation plot communicated the differences in atomic fluctuations adopted at different points in the N-terminus.

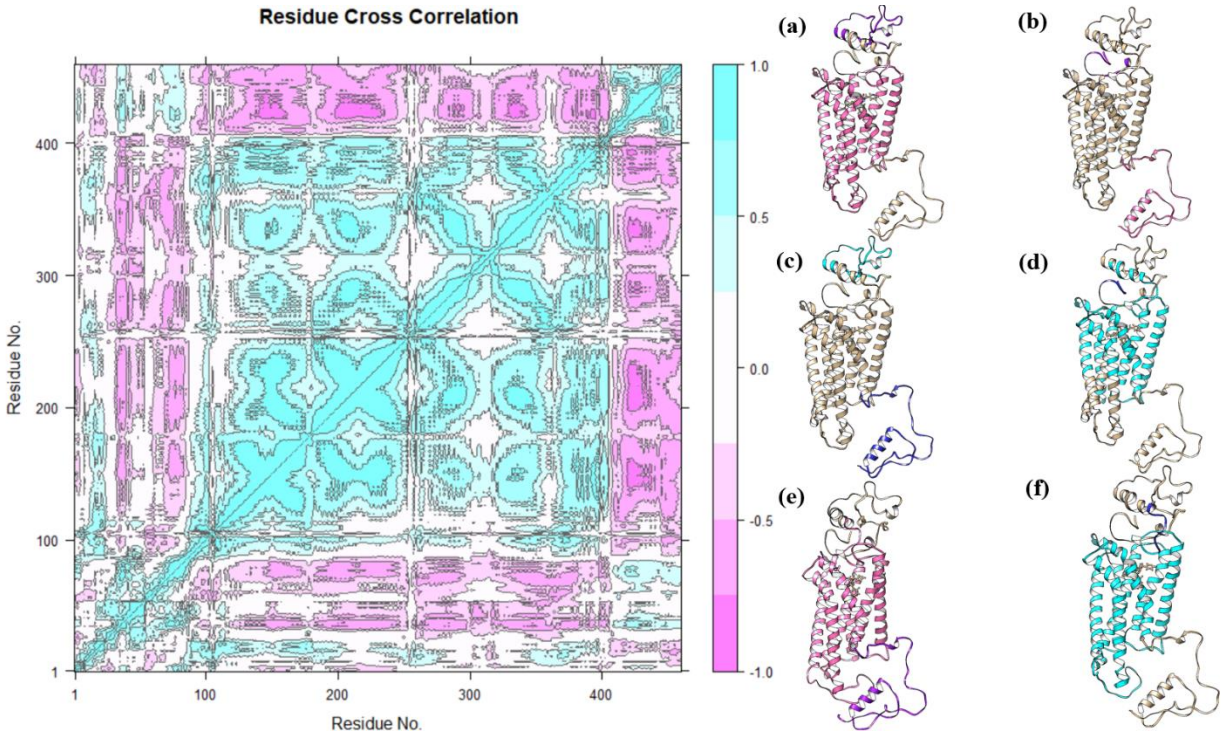


Figure 3.14: Cross-correlation plot for MJ15. **(a)** Residues 95-400 encompassing the MPR, and the TM segments negatively correlate with residues at the beginning of the N-terminal region (residues 30-80) **(b)** The C-terminus (encompassing ~residues 400-472) is negatively correlated to the helices and unstructured loops near the beginning of the N-terminal region (residues 1-20). **(c)** In contrast, the C-terminal region is positively correlated with the more flexible helices of the N-terminus (residues 30-80). **(d)** The helices in the N-terminus near the MPR to the beginning of TM3; TM4 -TM 5; the beginning of TM6 and all of TM7 (residues 80-200; 230-290; 350-400) are all positively correlated with the beginning of the N-terminus (residues 1-10) **(e)** The C-terminal region (residues 400-472) is negatively dynamically correlated with The MPR and C-terminal segment (residues 90-400) **(f)** TM7,6,5, ECL2 (residues 250-400) positively correlate to the MPR and TM1-4 (residues 100-250), which are both stable regions

For instance, the N-terminal helices facing the extracellular solvent in the complex were subjected to more fluctuations than in comparison to the N-terminal cap (residues 1-20) and the MPR (residues 90-110) that formed stable interactions near the orthosteric pocket. This was also confirmed by the RMSF analysis described above. RMSF and PCA analyses together

demonstrated the degree of fluctuations in these different sections. As a result, the stable TM domain of CB_1 exhibited negative dynamic correlations with those of the N- terminus and the variable C-terminal region (**Figure 3.14**). In contrast, the N-terminal cap (residues 1-20) mediated positive dynamic correlations with the TM and MPR regions in CB_1. Likewise, other positive dynamic correlations were seen between the highly dynamic C-terminal region and the highly fluctuating helical region in the N-terminus.

Likewise, the AM251-bound complex (**Figure 3.15**), displayed the most dynamic correlations between the more variable extracellular/intracellular regions with the highly stable TM helices.

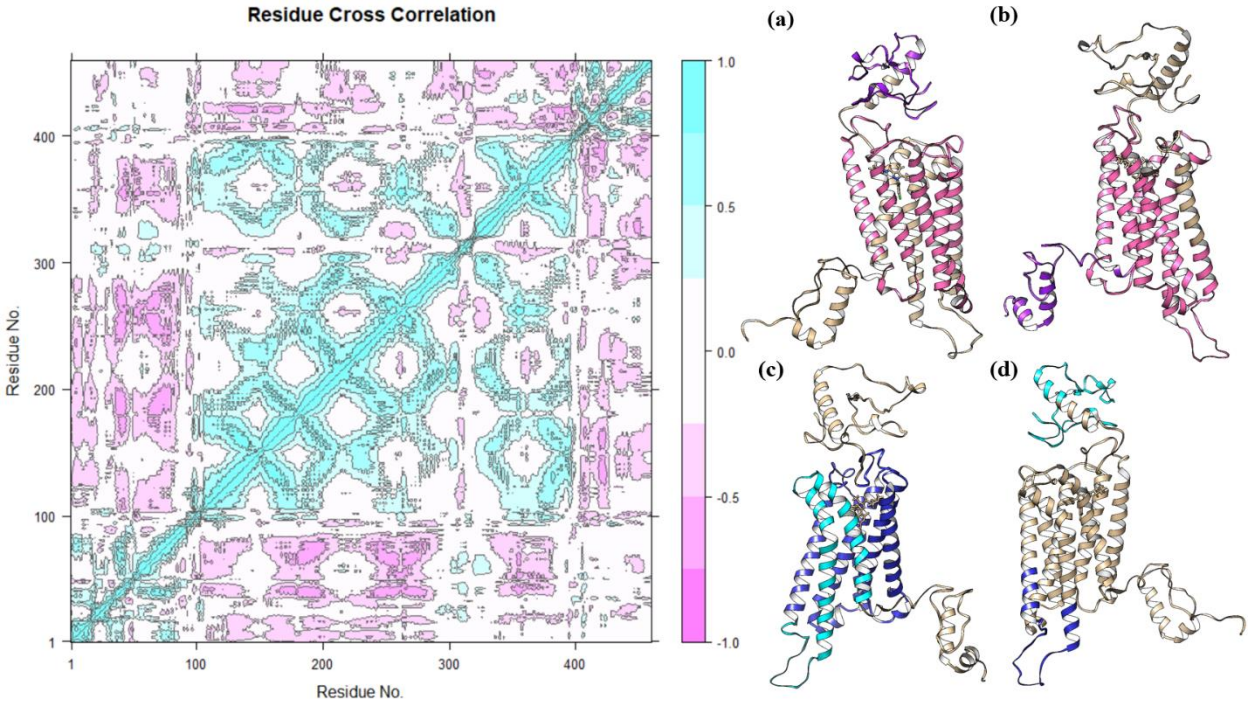


Figure 3.15: Cross-correlation plot for AM251. **(a)** The stabilizing MPR-TM5 (residues 100-290; 350-472) are mostly eliciting negative and neutral correlations with the highly dynamic N-terminus (residues 1-80). **(b)** The C-terminal region (residues 400-472) is mostly having negative and neutral dynamic correlations with the TM helices: TM1-TM4 and TM5-TM7 (encompassing residues 100-250; 290-400) **(c)** TM7 and TM6 (residues 300-400) are having positive and neutral correlations with TM1-5(residues 100-300) **(d)** The ICL3 (residues 290-340) is mediating positive and neutral correlations with the N-terminus (residues 1-80).

Most negative dynamic correlations were observed between the highly dynamic N- and C-termini extremities, and the stabilizing TM region. As the N- and C- termini contain more unstructured regions and loops, thus adding to their high instability and flexibility. In contrast, the TM region is mostly cultivated of helices that form interactions between and within each other, thus aiding in the adoption of a more stabilized region. In terms of the positive dynamic correlations, these were specifically observed between the TM helices and MPR, which are highly stable regions due to their roles in ligand-binding. Another positive interaction was displayed between the ICL3 and the N-terminus, which both adopt a high level of flexibility due to their large, unstructured segments.

For the CB_1 system containing ligand AM6538 (**Figure 3.16**), correlations between the highly dynamic helical and stabilizing unstructured segments within the N-terminus, to the rest of the CB_1 protein are also apparent, as within the MJ15 complex.

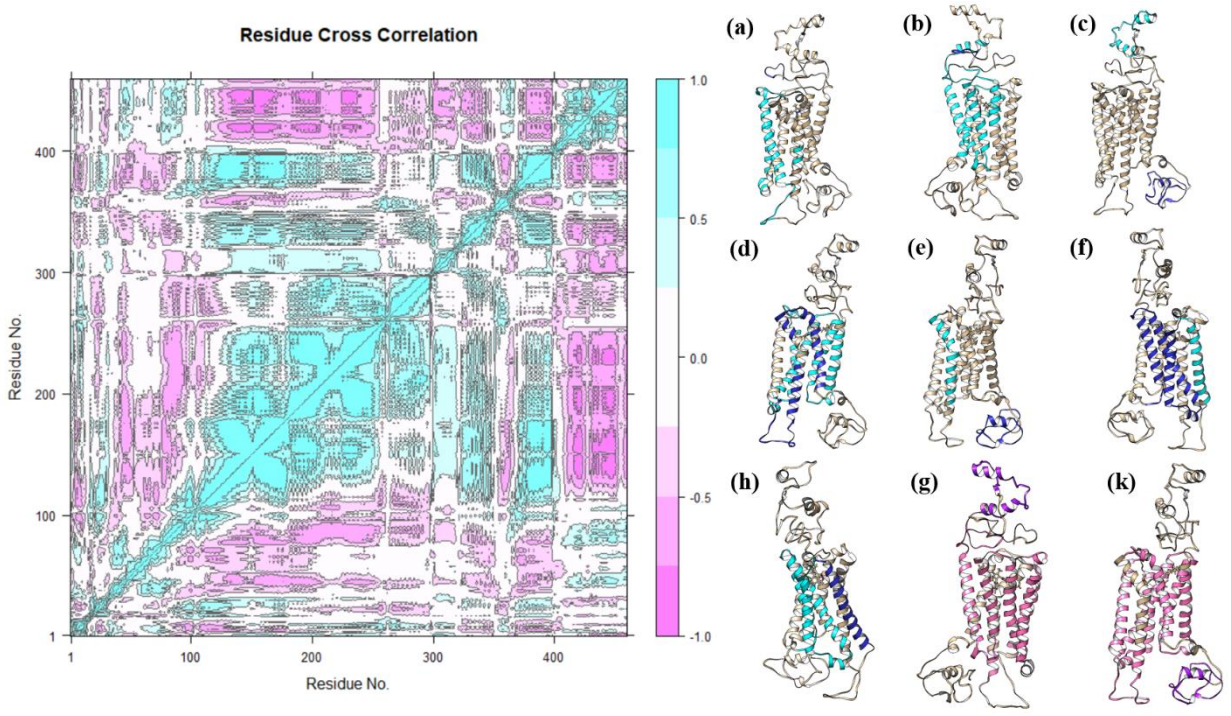


Figure 3.16: Cross-correlation plot for AM6538. **(a)** TM 4 and 5, in addition to TM 6 (residues 200-320;345-350, respectively) are mediating positive and neutral dynamic correlations with the beginning of the N-terminus which is much more stabilized (residues 1-10) **(b)** TM1, 2 and beginning of TM3 as well as the beginning of TM6 and TM7 (residues 80-200;360-400) are having

positive and neutral dynamic correlations with the unstructured loop region near the beginning of the N-terminus(residues 20-30). **(c)** The flexible C-terminus (encompassing residues 420-472) is having positive and neutral correlations with the highly dynamic helical loops in the N-terminus (residues 40-90). **(d)** TM1-4 and the ECL2 (residues 120-270) are having positive and neutral dynamic correlations with ICL3 and TM6-7 (residues 300-350;360-400, respectively) **(e)** TM6 (residues 340-370) and the C-terminus (residues 420-472) are having a slightly positive and neutral fluctuating dynamic correlation with each other. **(f)** TM5 and ECL2 (residues 260-300), as well as the end of TM1-TM4 (residues 130-260), are mediating positive highly dynamic correlations with each other. **(h)** ECL1, TM3 and 4 (residues 150-260) with is having positive correlations with TM5 (residues 270-300) **(g)** The TM segments unstructured region of the N-terminus incorporating and succeeding the MPR, in addition to TM1-5; the end of TM6 (residues 90-300; 330-420) are mediating negative and neutral correlations with the N-terminus (residues 40-90). **(k)** The C-terminus (residues 420-472) is having mostly negative correlations with the TM segment 1-5 and TM7 (residues 120-340;370-400).

Given this connection, the stabilizing TM segments are positively correlated with each other, and with the low fluctuating N-terminal cap. Both the level of dynamics between the TM segments and the N-terminal cap are within similar degrees of fluctuations, which is also visible in the RMSF plot (**Figure 3.6**). Other positive dynamic correlations were observed between the N- and C-termini, which are both experiencing a level of high fluctuations, as visualized in the backbone of the protein for eigenvalue ranks 1 to 3. In contrast, the highly dynamic helices in the N-terminus and helix 9 in the C-terminus were experiencing a negative dynamic correlation with most of the stabilized TM helices.

Overall, our results highlight the dynamics between the TM regions and the extra- and intra-cellular segments, as well as the dynamics within different areas in the N-terminus, such as the unstructured region and helices in the N-terminus to the rest of the CB_1 receptor. These results also show how the N-terminal cap, the MPR and the TM segments, which all mediate a high degree of stability and are positioned within proximity to the binding pocket, may contribute to the stability and binding affinity of the ligand. These results give a deeper insight into the level and types of dynamics adopted in the system.

3.4 Discussion

The rationale for this study was inspired by the CB_1's crystallographic structure's work (PDB code: 5U09) by Shao *et al*.,[110]. In the literature, Shao and co-workers[110] acknowledged how the lack of clear density and model ambiguity for the N-terminal domain in the AM6538's bound structure may limit the utility of the crystal structure to predict the binding modes of other ligands. Through carrying out protein-ligand docking with taranabant, a CB_1 inverse agonist, it was revealed by Shao and colleagues[110] that arm1 and arm2 (chlorophenyl and cyanophenyl groups) in taranabant were swapped, relative to their experimentally determined binding positions, highlighting the importance of including the N-terminal domain and MPR in docking predictions. In addition, previous experimental[108] and computational[106] studies, have highlighted the importance of MPR in allosteric binding, specifically for the allosteric binding of CBD. These studies together insist on the importance of accounting for the effects of N- and C- terminal regions on ligand binding to the orthosteric site of CB_1. Nevertheless, the complete structure of human CB_1 has not been reported either through any experiments or computational modelling methods. This motivated us to pursue this research, where we attempted to address the existing knowledge gap.

In this study, we built the complete atomistic model of the inactivated CB_1 receptor, by using a crystallographic structure of an inactive TM domain of human CB_1 (5TGZ) as a building block. We separately modelled the N-terminal and C-terminal domains of CB_1 and affixed them to the existing TM structure to construct a comprehensive model of the receptor. Using this model, we assessed the binding interactions of three different inverse agonists, AM251, AM6538 and MJ15, to the CB_1 receptor.

Our results from the analyses of the dynamical relationships of the inactivated receptor under the influence of the N-terminus revealed that the N-terminus radiates different degrees of stability, based on the N-terminal region's proximity to the orthosteric pocket. Specifically, the N-terminal cap and the MPR were observed mediating interactions with the TM, the extracellular regions, and the ligand, thus all adding to its stability. In addition, the MPR was seen forming a plug into the orthosteric pocket, while the N-terminal cap was observed to orient itself over the pocket, potentially further aiding in the process of solvent exclusion from the binding domain. In contrast, the N-terminal helical segments, orientated towards the extracellular solvent, were observed to

experience the greatest number of dynamics. Despite the presence of N-terminal helices' structured region, their lack of stable interactions near the TM region and orthosteric pocket contributed to their high dynamics. Our results on the dynamics of the receptor, give further insights into the various fluctuation within the N-terminal domain, as well as how its presence aids in the process of ligand binding and potential solvent exclusion from the orthosteric site.

To assess the accuracy and efficiency of our model, binding-free energy calculations were conducted for each protein-ligand complex. We achieved a good agreement between our predicted binding-free energies for the complexes and they're respective *in vitro* binding data from the literature[109,110]. This confirmed that our model was able to distinguish between the strong and weak affinity binders of CB_1 in this study. Assessing the nonbonded interactions between the inverse agonists and the orthosteric pocket, revealed that the N-terminus plays a role in the binding interactions with the ligand, thus potentially contributing to its binding-free energy and stability. These results may explain the influence that the N-terminus may have on the accuracy of the calculation and ranking of MMPBSA binding-free energies. The model from this study can be implemented in future works to analyze the impacts of the N-terminus on the process of cannabinoid-based binding through the orthosteric ligand entry domain, as well as assess the role of the N-terminus in the binding to other allosteric modulators. Overall, the results from our study exemplified the importance of modelling the N-terminus, and MPR for CADD, and in docking predictions. Finally, the complete inactivated CB_1 structure from this chapter can be used as a model system for any future virtual screening and drug discovery campaigns.

CHAPTER 4

Binding Interaction Interplay Between the Human Cannabinoid Receptor 1 (CB_1) and the Human Cannabinoid Interacting Protein 1a (CRIP1a)

4.1 Introduction

Protein-protein interactions by accessory molecules are an important means of GPCR regulation in cellular transduction and intracellular trafficking processes.[45] For these reasons, the emergence of GPCR interacting proteins as important modulators of ligand specificity, signalling, cell surface expression and trafficking has opened a new avenue in their investigation.[48] As discussed in Chapter 1, the human CB_1 receptor has been known to bind different protein partners that regulate its coupling with specific G-proteins. In particular, the C-terminal region of CB_1 is an important interaction site for various proteins including β-arrestins, and CRIP1 (isoforms a and b). Through these associations, certain regulatory pathways such as desensitization and internalization by β-arrestins can be controlled.[45] β-arrestins bind to the phosphorylated C-terminal tail of the CB_1 receptor. Upon their association, subsequent recruitment of a clathrin dynamin complex is formed to promote the endocytosis or internalization of the receptor. Outcomes of CB_1 internalization increase the organisms' cannabinoid tolerance which is a common occurrence seen in active cannabis users. Thus novel insights into CB_1 accessory proteins may help to uncover certain mechanisms involved in the regulation of its signalling.[48] In this work, we focused on constructing the complete atomistic model of the agonist-activated human CB_1 and human CRIP1a, while describing the dynamic interactions between both proteins.

Highly expressed in the central nervous system, CRIP1a co-localizes with CB_1 in pre-synaptic compartments of excitatory glutamatergic and inhibitory GABAergic neurons, where it acts as a binding partner to the CB_1 receptor.[20] CRIP1a plays a significant role in protein trafficking and signal transduction.[46–48] By interacting with CB1, CRIP1a alters its selectivity for $G\alpha_i$ subtype activation with the $G\alpha_{1/2}$ subunit, through the reduction of Gi_3 and G_o coupling[46]. In particular, G_{i3} and G_o activation are associated with the common CB_1 agonist-induced intracellular signalling, such as cAMP inhibition Ca^{2+} channel inhibition, MAPK activation, and reduction of glutamate or neurotransmitter release into the presynaptic cleft.[20,46,47] In contrast, CRIP1a overexpression has shown to be capable of suppressing CB_1-mediated tonic inhibition of N-type voltage-gated Ca^{2+} channels in superior cervical ganglion neurons, increasing glutamate release and decreasing CB_1-stimulated [^{35}S]GTPγS binding to $Gi_{3/o}$.[20,48] Due to CRIP1a's effects on receptor-mediated signalling transduction, efficacy, and neurotransmitter release, within the nervous system CRIP1a has been sought to be a potential novel therapeutic target to control neuropathic diseases, illnesses,

and symptoms[47], as well as treating disorders such as chronic pain, obesity, epilepsy as well as psychological disorders.[45]

Composed of 164 amino acids that form a 10- antiparallel stranded β-barrel with an interior hydrophobic core, CRIP1a encompasses a multitude of loops present at the bottom of the protein with a short helical cap on the top to exclude solvent.[166] Between β-strands β-8 and β-10 a gap exists which is held together by a hydrogen-bonding interaction between Trp121 and Tyr145 in addition to a hydrogen bonding network that includes Thr119 and Glu161 and water molecules.[166]

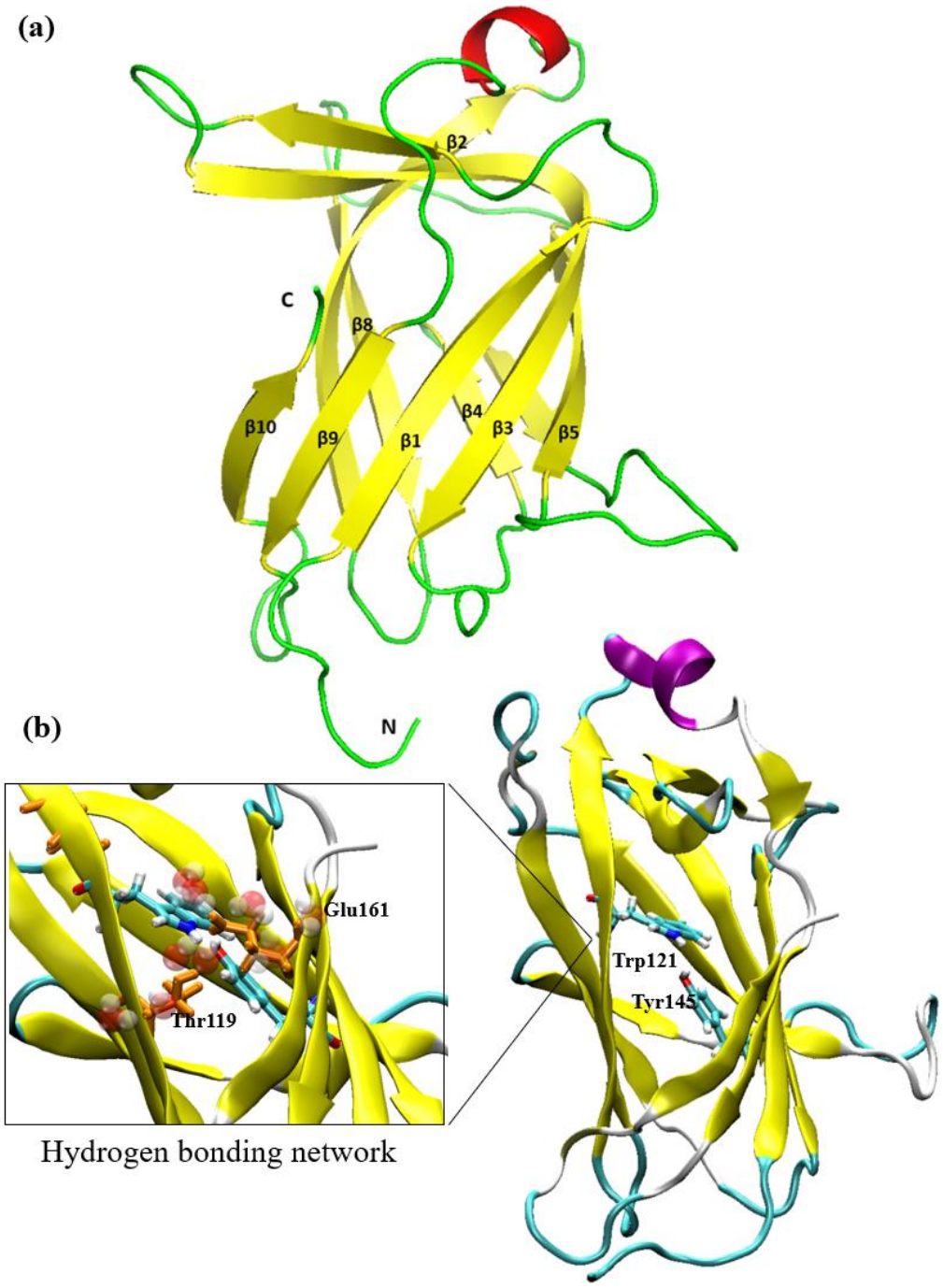


Figure 4.1: Structure composition of the human CRIP1a structure. **(a)** Visualization of its 10-β sheets as well as its N- and C-terminal loops, which aid to exclude solvent from the hydrophobic core of the protein. **(b)** The hydrogen bonding network is present between β sheets 8 and 10, which helps to exclude solvent and maintain the structure of the protein.

CRIP1a selectively binds to the last nine C-terminal residues for CB_1. Blume and colleagues[48] also described additional binding sites central to the C-terminal, Asp430 and Thr418, which offers a competitive factor for the CB_1-CRIP1a interaction.[46] In contrast, for CRIP1a the minimal binding domain for effective CRIP1a-CB_1 association is through residues 34-110[20,45](exons 1 and 2). Residues 34-110 are common amongst CRIP1a/b and thus are assumed to be more critical for binding interactions.[47] Deletion of residues 34 to 110 was shown to remove the N-terminal and C-terminal loops, thus exposing the hydrophobic interior core to solvent. [166] However, data from *in* vitro studies[20,45], suggests that residues 34-110 must be able to achieve stability when bound to another protein partner like CB_1. Maintenance of exons 1 and 2 was also proved to be sufficient in inhibiting endogenous CB_1 endocytosis similar to that of the wildtype CRIP1a.[20]

CRIP1a is known to participate in competitive interactions with *β*-arrestin proteins to bind to the distal C-terminus, thus affecting agonist-mediated internalization of CB_1, the efficiency of CB_1 signalling and the receptor surface density.[48,20]In essence, these competitive effects alter the clathrin- and dynamin-dependent internalization.[48] In contrast, a previous study based on affinity pull-down experiments revealed that phosphorylation at threonine-468 in the CB_1 C-terminal peptides, reduced CB_1-CRIP1a association while increasing CB_1-*β*-arrestin interactions.[48] Whereas mutation of the potential phosphorylation sites in the distal CB_1 C-terminus, as well as five distal residues with the C-terminus (D466, T467, S468 and A471, L472) that are associated with internalization, reduced receptor endocytosis.[20] Thus, competitive binding of CRIP1a to the critical sites for internalization in CB_1, serves to functionally attenuate agonist-mediated *β*-arrestin recruitment and thereby increase tolerance and cellular signalling efficacy.[48]

Currently (June 2022), there are no comprehensive structural models (either from experiments or through computations) describing the CRIP1a-CB_1 complex. Earlier attempts[46,47,105] to computationally model the CB_1 association with CRIP1a were built using homology template hematopoietic Rho-GDI2 (Rho-GDIβ or Ly/D4GDI), which shares a low (15.9%) sequence identity and functional homology with CRIP1a. In addition, Booth *et al*., [166] highlighted that the superimposing of the two crystal structures of CRIP1a and Rho-GDI2 illustrates significant dissimilarity between the two proteins (RMSD of 4.6 Å over 96 residues).[166] For these described reasons, it creates a large degree of uncertainty in the accuracy of the structure and thus binding results from the studies.[166] In addition to the low sequence identity, these studies did not include

the entire CB_1 structure, with the activated protein, and failed to subject their system to extensive molecular dynamic simulations for system convergence and relaxation. Computationally, these structural aspects may play a role in the interatomic protein-protein interactions between CRIP1a and CB_1. In essence, these missing structural aspects may limit the extracted data from these computational studies.

Recently, the first mammalian CRIP1a structure[166] (PDB accession code: 6WSK) was crystallized at a resolution of 1.55Å, allowing for the investigation of a more concrete computational analysis of CB_1-CRIP1a interactions. The crystallographic rat CRIP1a (rCRIP1a) structure, provides a more accurate template (**Figure 4.2**), that shares a 96% sequence identity with the human CRIP1a, generating a greater certainty in the computational model. In conjunction with a fully built activated CB_1 structure, and a complete human CRIP1a model (using homology modelling) we assessed the inter-atomic structural interactions between the world's first most comprehensive reported structure of the human CRIP1a-CB_1 complex. Finally, we employed MD simulations to gain a deeper understanding of their interplay of protein-protein association and binding.

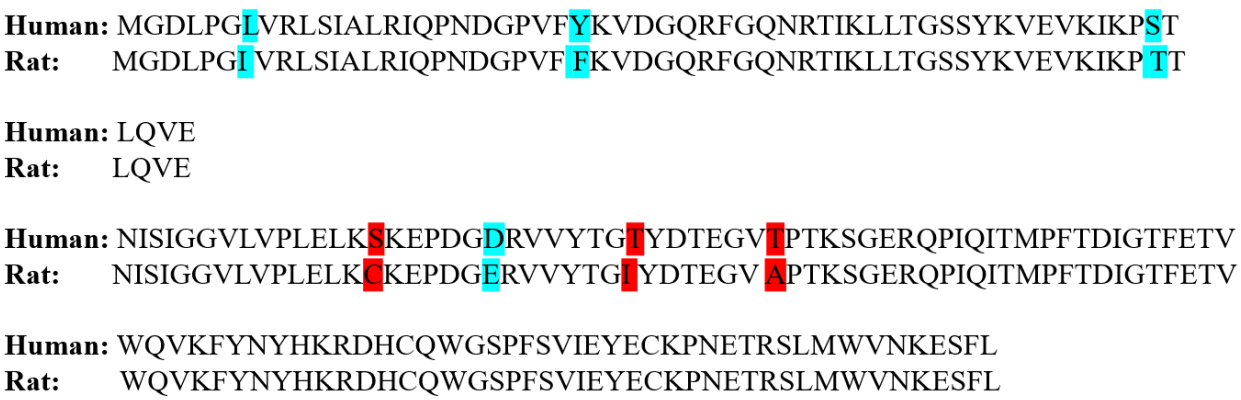


Figure 4.2: Sequence alignment between the human and rat CRIP1a protein. Red and turquoise highlight residues without any and with chemically similar properties, respectively. Sequence alignment encompasses a 95.73% percent identity.

4.2 Methods:

4.2.1 Model Building

The representative crystallographic structure of the active CB_1 receptor (accession code: 5XR8[111], resolution 2.95 Å) was selected from the Protein Data Bank (PDB) database[128]. Initially, the

protein was prepared by the addition of hydrogen atoms, the deletion of flavodoxin stabilizing protein, and the subsequent reconstruction of any missing loops and atom backbones, utilizing the software package UCSF Chimera version 1.14[129]. Methods of the construction of the N- and C-termini can be seen in Chapter 3, "Inverse-Agonist Binding Interactions and Dynamic Interplay with the Complete Human Cannabinoid Receptor 1 (CB_1)" of this thesis.

For building the CRIP1a structure, the recently published crystallographic CRIP1a (PDB accession code:6WSK[166], resolution 1.55 Å) structure obtained from the PDB database and was used as a template for homology modelling of the complete human CRIP1a protein in the online server Robetta[167].

4.2.2 Physiological System Setup and Ligand Preparation

The Membrane Builder in CHARMM-GUI[138] was used for modelling the physiological environment of the CB_1 receptor and the CRIP1a protein. For membrane embedding of CB_1, we initially determined the position of its TM domain using the PPM[139] server and the orientations of proteins and membranes (OPM)[168] database. The membrane was built using a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer. Since the CRIP1a is not an integral protein, there was no membrane-embedding procedure for this protein. CRIP1a and membrane-bound CB_1 were solvated with TIP3P water molecules. The solvated systems were neutralized using a 0.15 M concentration of Na^+ and Cl^- ions. The total system sizes of CB_1 and CRIP1a were ~100Å x 100Å x 160Å. For the CB_1 receptor, we extracted the agonist ligand (AM841) that was co-crystallized in the template TM model (PDB accession code: 5XR8[111]), which was used to build the complete structure of activated CB1 in this work. The parameter and topology files of the ligand (AM841) were prepared in CHARMM-GUI's PDB reader module.

4.2.3 Molecular Dynamic Simulations and Analysis

All classical MD simulations were performed using the AMBER18[143] MD simulation software in conjunction with CHARMM36M[169] forcefields for all atoms within the system. Each system was minimized under 2,500 steps of steepest descent minimization, followed by 2,500 steps of conjugate gradient minimization. Constant volume periodic boundary conditions, in addition to harmonic restraints (250 kcal/mol) were applied to the system. The heating step was applied to the isolated CRIP1a protein, in which the system was gradually heated from 0K to 310K. To avoid

substantial fluctuations in the solute, a 25 kcal/mol restraint was applied to the system. Equilibration was carried out for a total of 7 ns under a constant pressure periodic (NPT) ensemble with additional added positional (10 kcal/mol to 0 kcal/mol) and harmonic restraints (100kcal/mol to 0 kcal/mol). The temperature was maintained at 310K using Langevin dynamics with a collision frequency (γ) of 1.0 ps^{-1}. The production runs for the isolated CB_1 receptor had a slight (0.1 kcal/mol) positional restraint on the backbone atoms within the TM segment for the first 50 ns. After 50ns, the CB_1 system was run restraint-free, whereas the entire production run for the isolated CRIP1a protein was run without restraints. Isolated proteins were subjected to a total of 250 ns of production runs, while final complexes were run for a total of 100 ns. The particle-mesh Ewald (PME) method was applied for long-range electrostatic interactions. Short-ranged non-bonded van der Waals interactions had a distance cut off 10 Å, with a force-base switching set to 10Å. The SHAKE algorithm was employed to constrain the bond lengths of hydrogen atoms.

4.2.4 Root-Mean Square Fluctuations and Root-Mean Square Deviation

RMSD is a commonly used method to compare or analyze a structure's stability. RMSD measures the distance or dissimilarity between molecular conformations at one reference state. The trajectory conformations are used to measure the dissimilarity of the trajectory to the reference. Therefore, the fewer fluctuations experienced in the structure correlate to a higher degree of similarity between the trajectorial ensemble to the reference structure, and therefore stability in the system. The position (*x*) of the atom (*i*) in the structure, is calculated and, summed throughout the trajectory between the coordinate arrays x_i and x_i^{ref}, exemplified in the equation (**4.1**) below[170]:

$$RMSD\ (x, x^{ref}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|x_i - x_i^{ref}\right|^2} \quad \textbf{4.1}$$

$$RMSD\ (x, x^{ref}) = \frac{min}{R, t}\sqrt{\frac{1}{N}\sum_{i=1}^{n}\left[(R \cdot x_i(t) + \boldsymbol{t}) - x_i^{ref}\right]^2} \quad \textbf{4.2}$$

Where *n*/*N* is the total number of positions or frames. As atoms displace from their original position, the structure atom *x* is translated by a vector t and rotated by a matrix *R* to align with the reference atom (usually at frame 0 or the first frame), so that the RMSD can be minimized (as depicted by **Equation 4.2**). When analyzing the RMSD of proteins, typically not all the coordinates in the

structure are used in the analysis, and instead only the backbone atoms are measured. Similarly, RMSF measures the residues in a protein that contribute to its motion. However, unlike RMSD which calculates the difference in position between the protein's backbone atoms over time, RMSF calculates the individual residue flexibility (or how much it moved) during the simulation. [170] Therefore, residues that have more dynamics throughout the trajectory will have higher RMSF values and vice versa.

The RMSD and RMSF for isolated CB_1 and CRIP1a structures, as well as for the CB_1-CRIP1a complexes were calculated using the backbone atoms C, Cα, N and O atoms of the proteins. RMSD and RMSF values were calculated to assess the system's stability, and to analyze the segments within the protein experiencing a high amount of dynamics, respectively. RMSF can also generate further biomolecular and structural details about the protein within its environment, helping in the understanding of its nature and function. All trajectories were pre-aligned to the center and fit the first conformational frame as a reference structure, utilizing the CPPTRAJ tool from AMBER18[143] package.

4.2.5 Unsupervised Machine Learning-Based Clustering

The K-means algorithm was carried out using the CPPTRAJ [142] tool from the AMBER18[143] package which was used to cluster the last 100 ns of the CRIP1a and CB_1 systems, due to their achieved convergence and stability within the region. To find the most optimal cluster hyperparameter, the algorithm was run clustering the data based on the RMSD of the backbone, for 99 clusters, starting at a cluster size of 2. The cluster with the lowest DBI score, a relatively high pSF value, and reached a plateau on the SSR/SST line was chosen to be the most optimal cluster size. For more information on the K-means clustering algorithm please see Chapter 3.

4.2.6 Protein-Protein Docking

The dominant conformations of CB_1 and CRIP1a identified from clustering were used to perform protein-protein docking calculations using HADDOCK[171171]. Clustered complexes were first modified by removing solvent and extra lipid molecules from the PDB files, before submission to Haddock[171]. To assess the validity of our predicted docking results for CRIP1a, we submitted two docking protocols. These protocols were termed **1)** guided docking – an approach where certain

active interacting residues of CRIP1a from the literature knowledge were used to guide the docking; and **2)** blind docking, where no restraints on CRIP1a were used.

For guided docking, the CRIP1a was submitted with a docking region encompassing residues 34-110, while the additional surrounding area was coined as passive residues. In terms of CB_1, the minimal binding region, being that of the last 9 C-terminal amino acids (residues 464-472) were labelled as active (or docking) residues, with the surrounding residues being labelled as passive.

For the blind dockings, the last 9 amino acids (residues 464-472) were also set to be the docking region within the CB_1 receptor, while the entire CRIP1a protein (residues 1-164) was labelled as active (or docking) residues. In the case of the blind docking protocol, the entire CRIP1a protein (residues 1-164) was docked to the last 9 C-terminal amino acids of the CB_1 receptor (residues 464-472). This procedure provides an opportunity to sample the interactions of the C-terminal chain of CB_1 with any of the surface regions of CRIP1a and not just the residues 34-110 as in guided docking. The Haddock score was modified where 10% of the restraint energy was subtracted from the Haddock total. This was carried out because the high restraint energy value was affecting the viability of the Haddock score. Each complex from the Haddock docking results was ranked based on its modified Haddock score (a.u). Complexes were visually inspected to see their alignment with the membrane. Complexes that were going into the membrane were omitted. CRIP1a-CB_1 complexes were subjected to the same MD protocol as the isolated CB_1 receptor, with the exception that the entire production run for the CRIP1a-CB_1 complex was run restraint free.

4.2.7 Binding-Free Energy Calculations

Molecular Mechanics Generalized Born Surface Area (MM-GBSA)[172] method was used to examine, and rank the binding-free energies employed within each complex. It was also used to generate energy decomposition plots, so that the per-residue binding affinity contributions, involved in the formation of the complex, could be assessed. MMPBSA and MMGBSA provide an alternate method to predict the binding energy of molecules. These methods introduce a compromise in accuracy and speed by lowering the computational costs, when compared to the highly accurate but computationally expensive free energy perturbation (FEP) methods.[144]Evidently MMPBSA and MMGBSA outperform, the fast but less accurate empirical scoring functions implemented in docking programs.[144]Thus, these binding-free energies help to

discriminate the distinct energy profiles of the ligands and to understand the binding mode interactions between that of CRIP1a and CB_1. MM-P(G)BSA are widely used for capturing electrostatic energies and forces in implicit solvent modelling.[145] For our system, the last 40ns of the MD trajectory were used for the final CRIP1a-CB_1 complex, equating to a total of 4000 extracted snapshots. The binding free energy of each complex is computed by the following equation:

$$\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand})$$

$G_{complex}$, $G_{protein}$, G_{ligand} represent the free energy of the CB_1-CRIP1a complex, the isolated CB_1 receptor and CRIP1a protein in the solvent, respectively. The binding free energy (ΔG_{bind}) contains the gas-phase binding free energy (ΔG_{gas}) as well as the solvation free energy of the complex(ΔG_{sol}) relative to that of CB_1 and CRIP1a.

$$\Delta G_{bind} = \Delta G_{gas} + (\Delta G_{sol})$$

The gas-phase binding free energy (ΔG_{gas}) can be divided into the averaged CB_1-CRIP1a interactions, which include the potential energy or molecular mechanics energy term E_{MM} and the contribution of entropy ($-T\Delta S$)

$$\Delta G_{gas} = E_{MM} + (\Delta G_{sol}) \ - T\Delta S$$

E_{MM} can be further composed into three terms, which are: **1)** an aggregate of electrostatic (E_{es}), **2)** van der Waals (E_{vdW}), and **3)** internal energy (E_{int}):

$$E_{MM} = E_{bonded} - E_{non-bonded}$$

$E_{MM} = E_{bond} + E_{angle} + E_{torsion}$ and $E_{non-bonded} = E_{vdW} + E_{es} + E_{int}$

While the solvation energy ΔG_{sol} is composed of two terms:

$$\Delta G_{sol} = \Delta G_{pb/gb} + \Delta G_{nsp}$$

$$G_{nps} = \gamma SASA + b$$

$\Delta G_{pb/gb}$is the polar solvation free energy contribution calculated by solving the nonlinear Poison-Boltzmann (PB) and GB equation(s).[145] ΔG_{nsp} is the nonpolar solvation free energy, which is

estimated using the solvent-accessible surface area (SA) term. The values for the solute embedded in a membrane (interior dielectric constant) and solvent (external dielectric constant) dielectric constants were set to 1.0 and 80 for MMGBSA. The nonpolar solvation free energy term G_{nps}, was estimated by the solvent-accessible surface area, using a water probe radius of 4.0 Å.

A previous study conducted by Hou *et al*.,[144] demonstrated that the MM-GBSA approach was able to provide accurate relative binding free energies of the biological systems that are usually sufficient to discern the strong-affinity complexes from the weak-affinity ones. Thus, when considering computational efficiency, MMGBSA can serve as a powerful tool for ranking the complexes in this work. All MMGBSA calculations were run using AMBER21[173] using the mmpbsa.py[172] script.

4.2.8 Principal Component Analysis and Cross-Correlation Plots

Principal component analysis (PCA) is a statistical method that is efficient in reducing the dimensionality of a complex system while extracting essential information from the principal modes of sampled motion.[148,149] PCA analyzes the relationship between different sampled conformations during the trajectory while describing the maximal variance in the distribution of the structures.[150,148] The variance of the atomic positional fluctuations captured in each dimension are characterized by their corresponding eigenvalue. Usually, 3-5 dimensions are sufficient to capture over 70% of the total variance within a given MD trajectory.[150] When calculating the PCA of biomolecules, utilization of dihedral angles or atomic coordinates for α-carbon atoms, is sufficient to capture the dynamics.[148] In the PCA and cross-correlation analysis, the x, y and z Cartesian coordinates of the C-α atoms were used for the trajectory frame superimposition.[148] All PCA and cross-correlation plots were conducted using the Bio3d[150] package in R and were throughout over the course of the last 100ns.

4.2.9 Electrostatic Interactions

We employed electrostatic potential maps as a tool to understand the charge complementarity between CB_1 and CRIP1a in each of the complexes. Electrostatic maps help in identifying the most likely complex model that shares both shape and electrostatic complementarities, which are important factors in driving molecular recognition. The electrostatic maps of the CRIP1a-CB_1 complexes were generated using PyMOL[137]. CRIP1a-CB_1 complexes from protein-protein

docking were aligned to every representative structure generated from the K-means clustering output for both CRIP1a and CB_1 isolated proteins. The isolated CRIP1a and CB_1 representative structure with the lowest RMSD fit to the CRIP1a-CB_1 complex were used as a one-to-one residue comparison between structures. APBS Electrostatics[147] and PDB2PQR[174] methods in PyMOL was used to generate the electrostatic maps of the two proteins that were sampled at the beginning and the end of classical MD simulations.

4.3 Results

4.3.1 Generation of Physiological CB_1 and CRIP1a Models

4.3.1.1 Cannabinoid Receptor 1

Although the 3D structures of CB_1 were resolved using experimental techniques, the available structures are incomplete as they do not describe the intra- and extracellular loops (e.g., the functionally important ICL3 loop) and their N- and C-terminal segments. As a result, we employed a combination of loop building, and threading-based methods to construct the complete human CB_1 model in its agonist-bound activated state. Preparation of the missing loops, particularly the ICL3 from the CB_1 crystallographic structure (PDB accession code:5XR8[111]), was built using the Modeller9v8 Python script in the software package UCSF Chimera version 1.14[129], to use the program Modeller (v10.1)[130]. The module generated a total of five different ICL3 conformations, and the one with the lowest discrete optimized protein energy (DOPE) score was selected. DOPE, a pairwise atomistic statistical potential, utilizing energy comparison between models, is designed to select the best structure from a collection of models.[175] The lower the DOPE score, the better the model. On the other hand, the protein extremities, including the N- and C- termini, were modelled using homology and threading techniques from the online server I-TASSER[131].

4.3.1.2 Cannabinoid receptor Interacting Protein 1a

To assess the productivity of the recently published crystallographic rat CRIP1a structure (PDB accession code: 6WSK[166]), as a template in homology modelling for the human CRIP1a protein, the crystallographic structure's sequence was aligned to that of the human CRIP1a's sequence using the online server BLAST. Results from BLAST identified a ~96% sequence identity between both mammalian CRIP1a proteins, ensuring a productive template. Utilizing the online server Robetta[167], the CRIP1a PDB code (6WSK) was given as a reference structure for building the

complete 3D computational model of the human CRIP1a protein. Robetta generated 5 different models, each with varying RMSD scores, and the model with the overall lowest RMSD value was chosen as the best structure. The chosen 3D atomistic model was run through various structural quality check programs such as ERRAT[160], verify3D[163], PROCHECK[176], and WHATCHECK[177] to assess the stereochemical quality and validity of the model. For definitions of ERRAT and Verify3D please see Chapter 3 of this thesis. In terms of the quality assessment program PROCHECK, it analyzes the stereochemical quality of a protein structure by investigating the residue and overall structure's geometry which is indicated through a Ramachandran plot. Results from PROCHECK deemed our protein to be of good quality except for only three residues (Lys76, Ser75, Lys26) which were present within the loops regions and not within the PDB crystal structure, adding to their low assessment score. Likewise, WHATCHECK assesses the stereochemical properties of the structure. Results generated from this program highlighted two errors (1) average B-factor error, and (2) the backbone conformation in comparison with other database proteins showed that the backbone fold in the structure was unusual. Since our model was in the early stages, these unfavourable atomic positions were able to be minimized throughout classical MD simulation steps such as minimization, equilibration, and production. By conducting these steps, it ensures structure optimization at a later point. Regarding, the quality check programs ERRAT and VERIFY3D, they gave scores of 83.66% and 92.07%, respectively. Overall, the high scores achieved from each program helped in validating our model.

4.3.2 RMSD and RMSF: CB_1Receptor CRIP1a

To analyze the dynamics of each protein, and optimize the models, each isolated protein, CRIP1a and CB_1, were embedded in their physiological environments using the CHARMM-GUI[138] online server and were run for a total of 250ns (**Figure 4.1** and **Figure 4.2**) of classical MD production runs. Throughout the trajectory, both systems had reached stability after ~150 ns, allowing for the last 100ns to be taken for further structural analysis. To analyze the fluctuations between the residues, the RMSF was calculated for all residues within the isolated CRIP1a and CB_1 models. The RMSF results of CB_1 had accurately aligned with previous computational studies[106,145], where the N-, C- termini, as well as the ICL3, exhibited the highest amount of dynamicity, out of other regions. The higher number of dynamics adopted by the extremities can be attributed to the disordered (or loop) regions. The lack of secondary structures within these regions allows for the

segments to adopt more conformations adding to the segment's flexibility. Moreover, since these regions face the solvent, it also increases the number of solvent-protein interactions, which often can be short-lived, thus further increasing the variability in these regions. It has been previously postulated that ICL3 high dynamicity allows it to interact with other intracellular regions in the CB_1 receptor, as well as potentially having a role in CRIP1a binding.[99] Other note-worthy fluctuations are the ECL2, as well as subsequent loops that connect the TM helices (ICL1, ECL1, ICL2, ECL3), in which many of these ECL regions partake in ligand binding.[109,110] The ECL2, in particular, is the largest ECL in the receptor, and thus like ICL3, it, too, experiences more dynamics due to its large nature. Likewise, CRIP1a shares a similar narrative, where its structured regions experience fewer fluctuations in comparison to its loop segments. For instance, in CRIP1a the β-sheets maintained more stability, in comparison to the disordered regions or segments surrounding the loop, such as the N-terminal cap, which were more dynamic. On the other hand, the increased stability present in the β-sheets can be attributed to ensuring that the core of the protein remains hydrophobic, in addition to the hydrogen bonding network present within the protein, which adds to its stability and structural formation.[166]

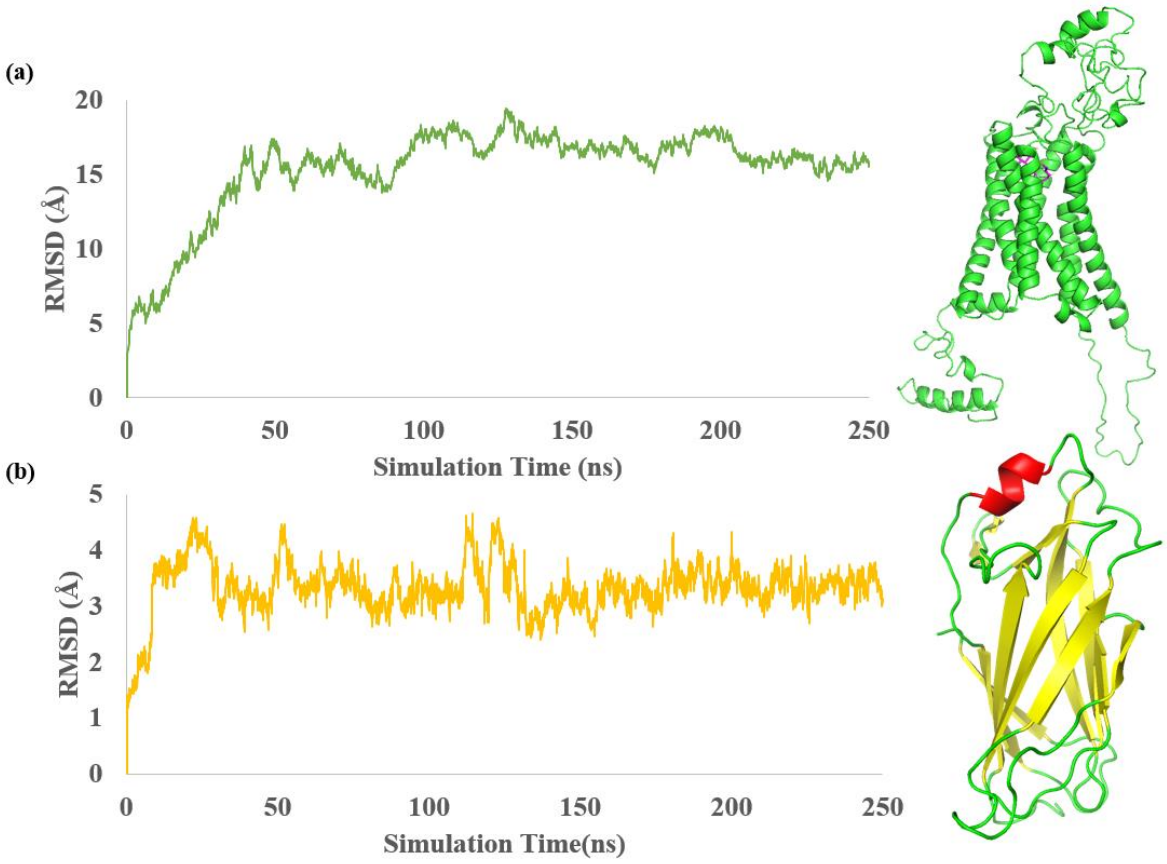


Figure 4.3: RMSD graphs of isolated CRIP1a and human CB_1 receptor, with their corresponding protein structures. Each system was run for 250 ns of classical MD, to which the last 100 ns were used for subsequent trajectorial analysis, due to the achieved stability in both systems. Both trajectories for the **(a)** human CB_1 receptor and **(b)** CRIP1a protein maintain an overall high degree of stability.

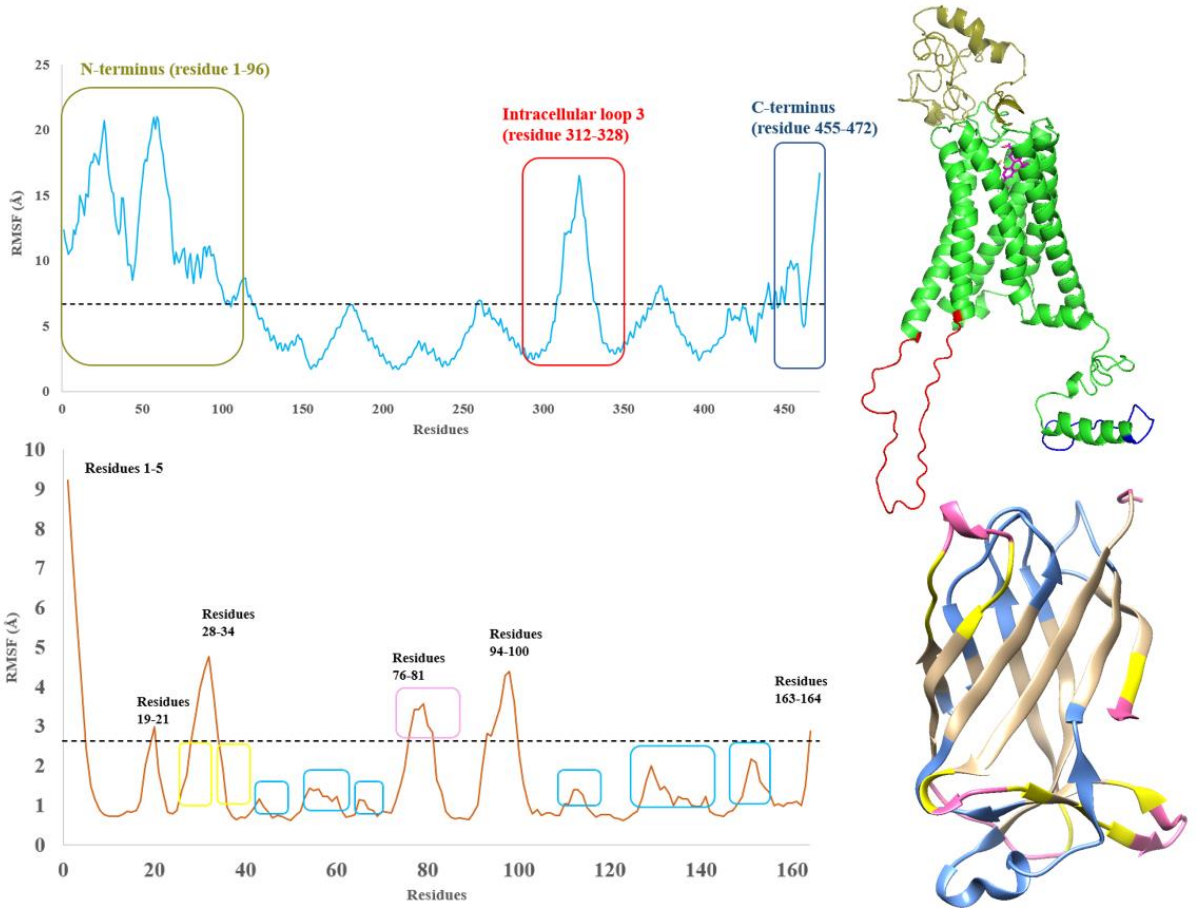


Figure 4.4: RMSF plot of isolated human CB_1 (PDB: 5XR8) and CRIP1a protein. **(a)** Crystallographic CB_1 (PDB accession code: 5XR8), with ligand AM841 (magenta). >7.5Å indicate areas of high fluctuation, which are evident in the ICL3 (red), C-terminus (blue), and N-terminus (olive). Lower areas of high fluctuation correspond to the ECLs and ICLs. **(b)** Human CRIP1a structure (based on PDB accession code: 6WSK). Areas of high pink fluctuations (>2.5Å higher areas of fluctuation), regions that become before the areas of high fluctuation (<2.5) yellow and areas of low fluctuation blue fluctuation are visualized within their corresponding colours. Most dynamic areas contain most of the loops and connecting β –sheets, while more stabilized regions are the β–sheets (visualized in brown).

4.3.3 K-Means Trajectorial Clustering Algorithm

Due to various structural conformations adopted throughout the classical MD trajectory, analysis and generation of the most dominant conformations were necessary, while eliminating noise or

less prominent conformations. We employed the K-Means algorithm to perform cluster analyses of MD trajectories of both CB1 and CRIP1a. Due to the high degree of stability that was achieved within the last 100 ns of both CRIP1a and CB_1, the trajectorial consensus region of 150-250ns was used in the generation of the clustering data (**Figure 4.3**).

Throughout generating 99 clusters by utilizing 99 hyperparameters with a stepwise index of 1 from an initial cluster size of 2, the best clusters were picked based on a low DBI and a plateau SSR/SST(R^2) value. Based on the criterion, (**Figure 4.5**) cluster 20 was to be the most optimal cluster for out generated data for CB_1, due to its low DBI (1.359083) score, and the plateauing of the R^2 value (0.853807), as well as the associated high psF value (301.542676).

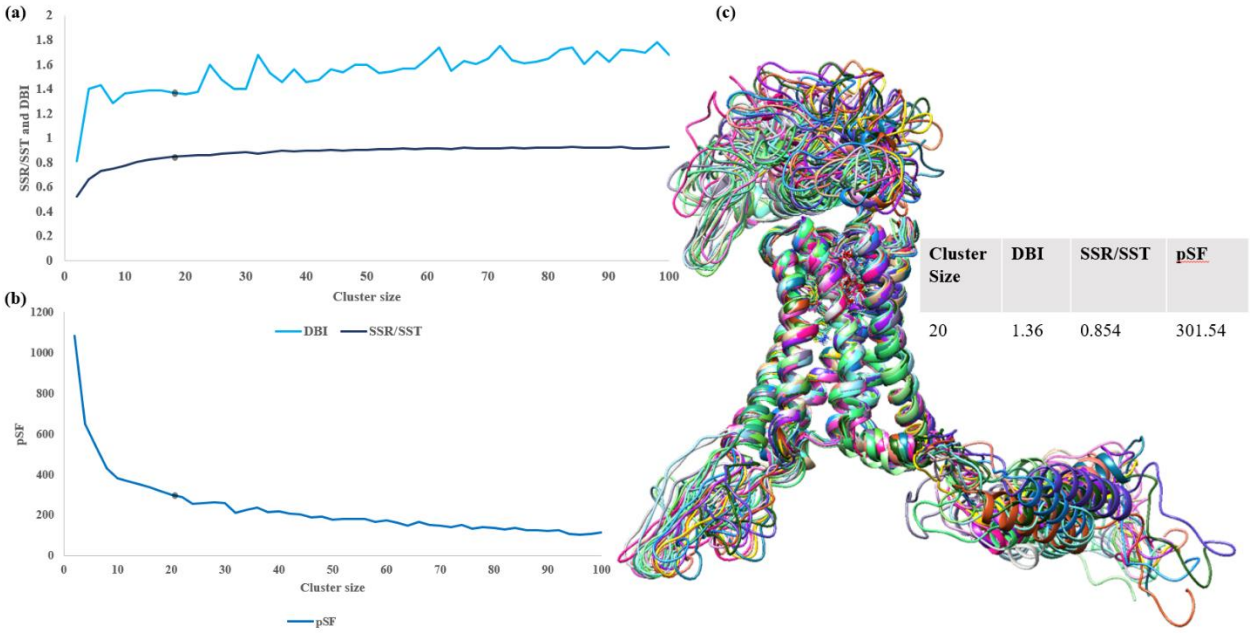


Figure 4.5: Results of CB_1 K-means clustering. Graphical and statistical results for the K-means clustering of the last 100ns of the 250ns trajectory. Hyperparameters from 2-100 were chosen for the algorithm. Cluster 20 was deemed the most optimal cluster based on its DBI and SSR/SST score. **(a)** DBI and SSR/SST graph **(b)** psF graphs values, cluster 20 is marked on both graphs (black dot) **(c)** visualization of the 20 structures from cluster size 20, with incorporated statistical information corresponding to cluster 20.

The majority of the frames incorporated in cluster 0 (the best cluster) were chosen from 40-55ns (400-550 frames or 190-220ns from the total trajectory) from the last 100ns of the trajectory, outlining their similarities in conformations of the protein during that timestamp.

In terms of CRIP1a, (**Figure 4.6**) a cluster size of 32 was picked, largely due to the plateauing of the R^2 value (0.452620), and the sloping of the DBI value (1.914726) as well as the large pSF score (25.846756).

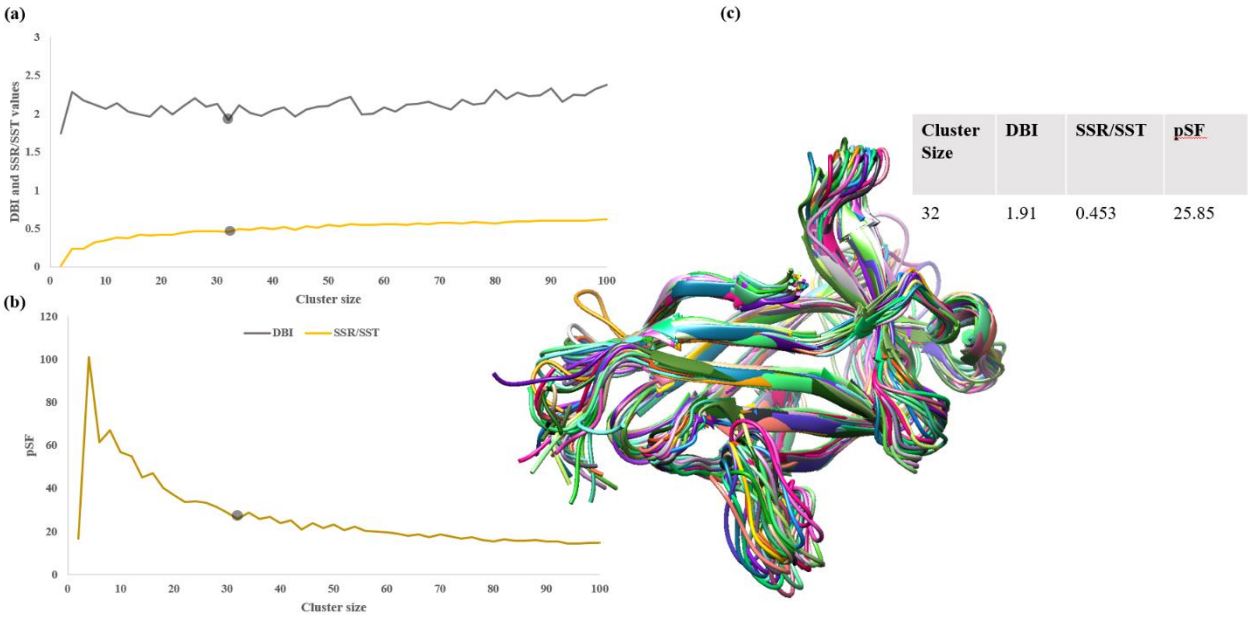


Figure 4.6: Results of CRIP1a K-means clustering. Graphical and statistical results for the K-means clustering for the last 100 ns of the 250 ns trajectory. Hyperparameters from 2-100 were chosen for the algorithm. Cluster 32 was deemed the most optimal cluster based on its DBI and SSR/SST score. **(a)** DBI and SSR/SST graph **(b)** psF graphs values, cluster 32 is marked on both graphs **(c)** visualization of the 32 structures from cluster size 32, with incorporated statistical information corresponding to cluster 32.

Despite the R^2 value generating a low confidence level, we believe this to be because of the more randomized fluctuations (**Figure 4.3**) within the unstructured regions within the protein, producing more noise than comparison to the stabilized CB_1. The consideration of our cut-off of 100 K-clusters was to avoid a large dataset of protein conformations, thus cluster 32 was deemed as the

best cluster for CRIP1a, which is also evident by the variation of incorporated frames seen in **Figure 4.7**. The best representative cluster (cluster 0), is encompassed within the 35-78 ns or 3000-7800 frame region, where the CRIP1a protein geometric distance conformations adopted during these ns were more similar to each other.

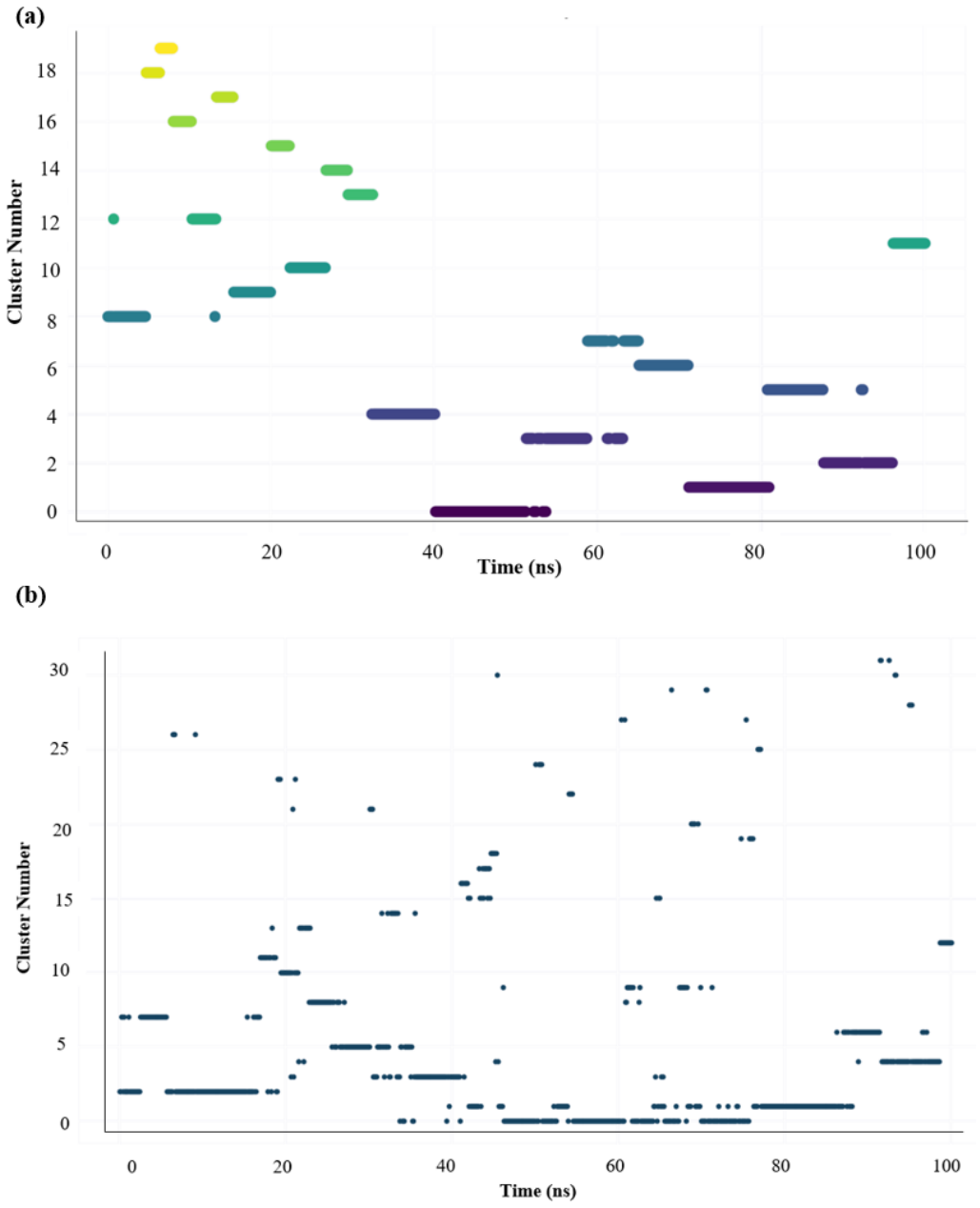


Figure 4.7: Frames or simulation time that encompass the representative cluster number. The last 100 ns of the 250 ns trajectory was used for the K-means cluster analysis. Visualization results for the hyperparameters 20 and 32 are depicted for **(a)** isolated CB_1 and **(b)** isolated CRIP1a,

respectively. Cluster number 0 makes up the best representative clusters and encompasses the greatest number of consecutive frames.

4.3.4 Protein-Protein Docking

The dominant conformations of agonist-bound human CB_1 (20 structures) and CRIP1a (32 structures) from the MD simulation were employed to perform protein-protein docking calculations using the HADDOCK[171] server. Our blind docking protocol was defined by docking the entire CRIP1a protein (residues 1-164) to the last 9 amino acids in the CB1's C-terminal tail. In contrast, our guided docking was based on previous *in vitro* data, in addition to the results produced from computational servers that predict residues participating in protein-protein interactions. Regarding the experimentally determined region, it was determined from the *in vitro* deletion studies[20,45] (residues 34-110 in CRIP1a) to be the minimal distance for CRIP1a-CB_1 association. The predictions from the computational servers increased our docking region from residues 34-110 to 34-138 (in CRIP1a), which was then docked to the last 9 amino acids in CB_1. The results from Haddock (**Table 4.1** and **Figure 4.8**) ended up generating a total of 41 clusters, in which a cut-off of -60a.u was applied to both the blind docking and restricted dockings complexes. 1 complex was discarded as the CRIP1a in this model was penetrating the membrane region thus causing steric hindrance. Therefore, a total of 11 structures from the blind docking (6 structures) and the guided docking (5 structures) were selected for further MD-refinement and rescoring based on their binding-free energies.

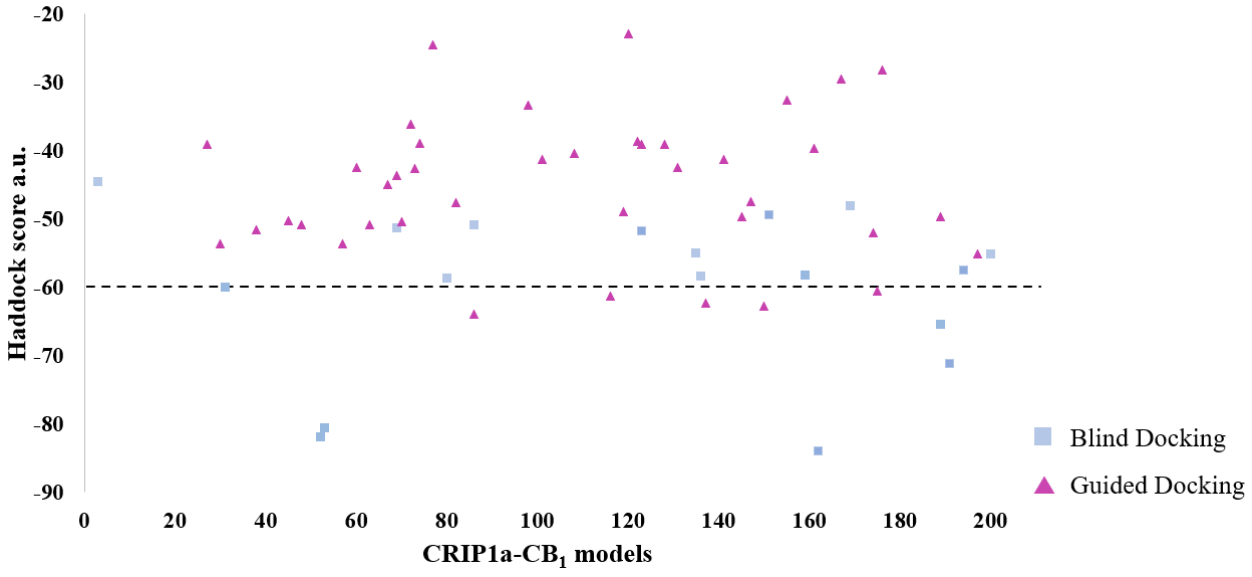


Figure 4.8: Haddock results of CRIP1a-CB_1 blind and guided docking protocol generated complexes. Our cut-off was established to be -60 a.u. (or values that were above the 75% percentile from the highest ranked complex). Complexes that adhered to the cut-off were used for further analysis.

Table 4.1: Blind and Guided docking complexes and their associated haddock scores.

Complex	Protocol: Guided Docking (G)/ Blind docking (B)	Haddock score (a.u.)
150	G	-62.765
86	G	-63.984
116	G	-61.242
137	G	-62.356
175	G	-60.567
31	B	-60.03
189	B	-65.363
191	B	-71.111
162	B	-83.888
52	B	-81.934
53	B	-80.538

Classical MD for Top Complexes

Since docking was performed without the membrane environment, we rescored the top 11 complexes from protein-protein docking using MD simulations in their respective lipid-based environments. To do this, each of the agonist-bound CB_1-CRIP1a complexes (i.e., CB_1[AM841]-CRIP1a complexes), were embedded in their physiological environments: the TM segment of CB_1 was submerged into the POPC lipid bilayer. Each system was solvated with TIP3P water molecules and in a 0.15M concentration of NaCl. The solvated

systems were subjected to a 30 ns long MD simulation each, to which the stability of the backbone (RMSD) was assessed (**Figure 4.9**).

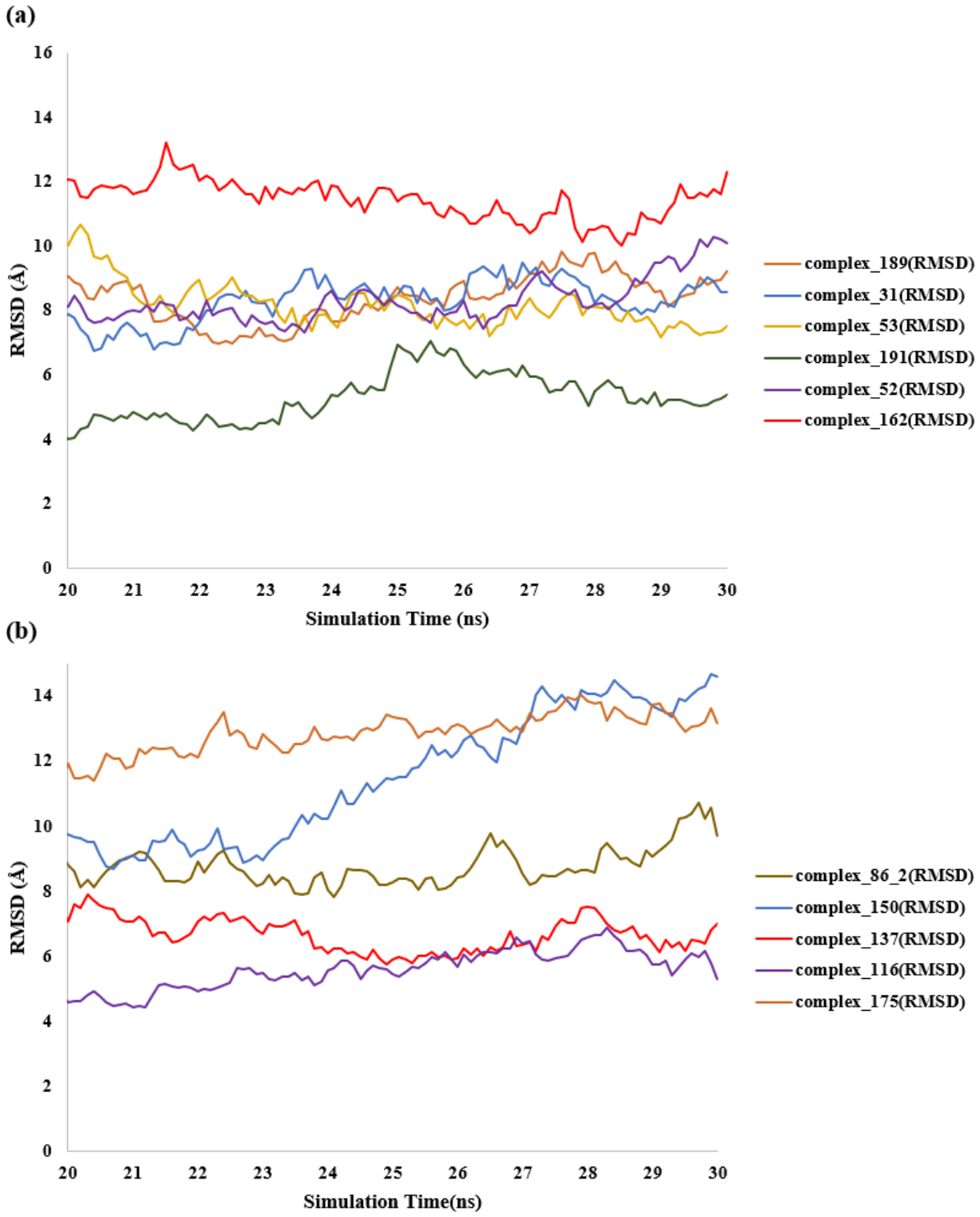


Figure 4.9: RMSD graphs of CRIP1a-CB$_1$ complexes. RMSD graphs of 30 ns of classical MD for **(a)** blind and **(b)** guided docking complexes. The inability of complex

150 (guided docking protocol) is evident in the lack of stability incorporated in its RMSD plot.

4.3.5 Binding-free Energy Calculations and Extended MD

Out of 11 complexes, we noted that 10 of them remained intact and exhibited stable behaviour. However, one of the complexes (complex 150) fell apart suggesting its weak interactions. Thus, this complex was dropped from further analyses. The binding affinities of the rest of the 10 agonist-bound CB_1:CRIP1a complexes were computed using the MM-GBSA method. To assess the reproducibility of the binding free energies, we computed the free energies by sampling snapshots at different timepoints from MD simulation such as the last 5 ns, 7.5 ns,10 ns and 15 ns from the 30 ns trajectories. The comparison of the MM-GBSA scores of all the 10 complexes at different timepoints of MD trajectories in **Figure 4.10** demonstrated that two complexes, Complex 116 from guided docking and complex 191 from blind docking, were consistently ranked in the top.

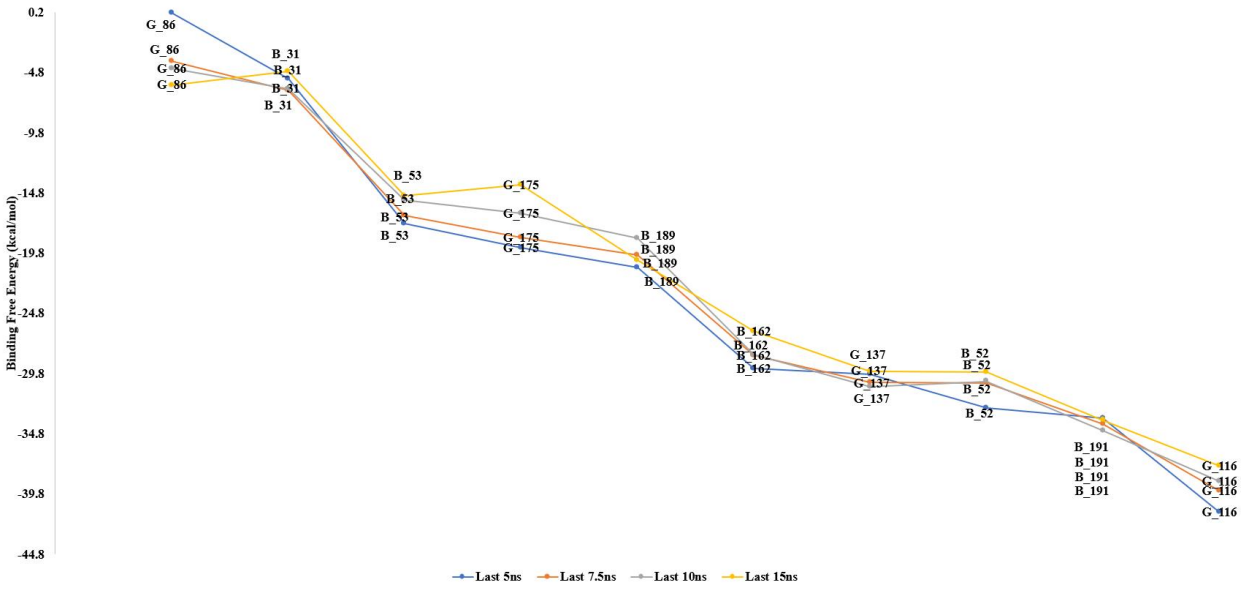
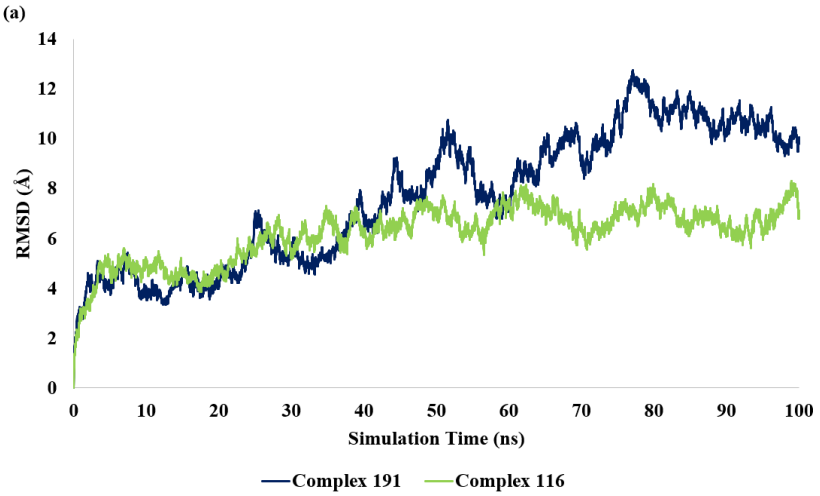


Figure 4.10: MMGBSA binding-free energy calculations for the blind(B) and guided(G) docking complexes. Each system was run for various intervals from the 30ns classical MD trajectory. The best complex was guided docking complex G_116.

Since there was a small energy gap (ΔG = 3.8 kcal/mol) between these two complexes, we subjected both systems to an extended 70 ns long MD simulation each to re-assess their stability

and binding free energies (**Figure 4.11**). As seen in **Figure 4.11a**, the backbone RMSD of complex 116 exhibited a better stable behaviour than that of complex 191. This is consistent with their binding free energies computed from the last 40 ns of their extended MD trajectories. Complex 116 exhibited binding free energy of -43.20 kcal/mol, which is much better than that of complex 119 (ΔG = -28.84kcal/mol). Thus, the free energy gap between the two complexes became larger from the extended MD simulation, resulting in complex 116 being favoured over the other model. Thus, complex 116 was selected as the best model to describe the human CB_1-CRIP1a complex interactions.



(b)

Binding-Free Energy Final Results		
Complex	191	116
Binding Energy (Kcal/mol)	-28.8459	-43.2069

Figure 4.11: RMSD graphs and binding free energy re-calculations for blind and restricted docking complexes. **(a)** RMSD graphs for the extended trajectory for complex 116(guided docking protocol) and complex 191(blind docking protocol), **(b)** Final MMGBSA binding-free energy calculations for the final 40 ns of the 100 ns trajectory. Complex 116 mediates a higher degree of stability within its RMSD fluctuations in comparison to complex 191, where it maintains a lower average (complex 116: 6.38 Å; complex 191: 8.70 Å) and average deviation values (complex 116: 0.72 Å; complex 191: 2.20Å).

4.3.6 Electrostatic and MMGBSA Decomposition Maps

To understand the nature of interactions between CB_1 and CRIP1a in complexes 119 and 116, we assessed the electrostatic complementarity and per-residue contacts between the two proteins in the two complexes. Results from the electrostatic maps (**Figure 4.12**) showed that complex 116 displayed more complementary regions of high negative and positive interactions. Further analysis

of the residues in proximity, within complex 116, participating in favourable positive and negative interactions were analyzed. Residues in proximity were Lys47(CRIP1a)-Glu634(CB_1), Lys74(CRIP1a)-Glu634, and Glu118(CRIP1a)-Lys622(CB_1), Lys619(CB_1)-Glu60(CRIP1a), Lys47(CRIP1a)-Leu636 (CB_1). In contrast, the blind docking protocol complex 191, mediated less favourable interactions, where only polar residues Asp630(CB_1)-Arg36(CRIP1a) and Glu615(CB_1)-Asp28(CRIP1a) were observed to be in proximity. In addition, electrostatic maps exhibited less complementary polar regions in comparison to guided docking complex 116. MMGBSA-generated decomposition plots additionally supported the data from the electrostatic maps, where the number of residues participating in strong interactions was more abundant in the guided docking complex 116 than in comparison complex 191. Residues 47(CRIP1a)-472(CB_1) and Leu455(CB_1)-Glu60(CRIP1a) were observed to form strong electrostatic hydrogen bonding interactions, whereas Thr88(CRIP1a)-Leu472(CB_1) and Ala469 (CB_1), Leu68(CRIP1a)-Lys458(CB_1), Thr108(CRIP1a)-Lys458(CB_1), were observed to mediate weaker hydrogen bonds while participating in mainly van der Waals interactions with surrounding residues. Met461(CB_1)-Thr108(CRIP1a) and Lys458(CB_1)-Gln106(CRIP1a), Met461(CB_1)-Leu68(CRIP1a),471(CB_1)-Pro70(CRIP1a) and Thr88(CRIP1a) formed mostly non-polar interactions. In addition, residues (**Figure 4.13)** participating in stronger interactions between CRIP1a and CB_1 in complex 116 was all present within the experimentally predicted minimal binding region (**Figure 4.14**) of 34^{CRIP1a}-110^{CRIP1a}. Despite our docking protocol containing more residues beyond that of 110, the strength of the interactions beyond the experimentally predicted region were greatly less. The generated data closely aligns with previous experimental results[20,45,99], thus giving deeper insights as to why the CRIP1a region 34-110 is important in the binding interactions between CB_1 and CRIP1a. Opposite results were obtained in complex 191, where most of the residues participating in strong intra-interactions were mediated at the extremities, thus agreeing less with experimentally derived results. Likewise, in CB_1, more residues in complex 116 were participating in stronger protein-protein binding interactions with CRIP1a. More distal-central amino acids such as those in helix 9 were observed forming interactions with CRIP1a. This data supports the previous experimental findings from Blume *et al*.,[48], that state that CRIP1a interacts with residues outside of the distal 9 C-terminal amino acids. In contrast, complex 191, had encompassed fewer residues participating in the protein-protein interactions, within the confines of the last nine C-terminal amino acids, thus contributing to its lower binding-free energy.

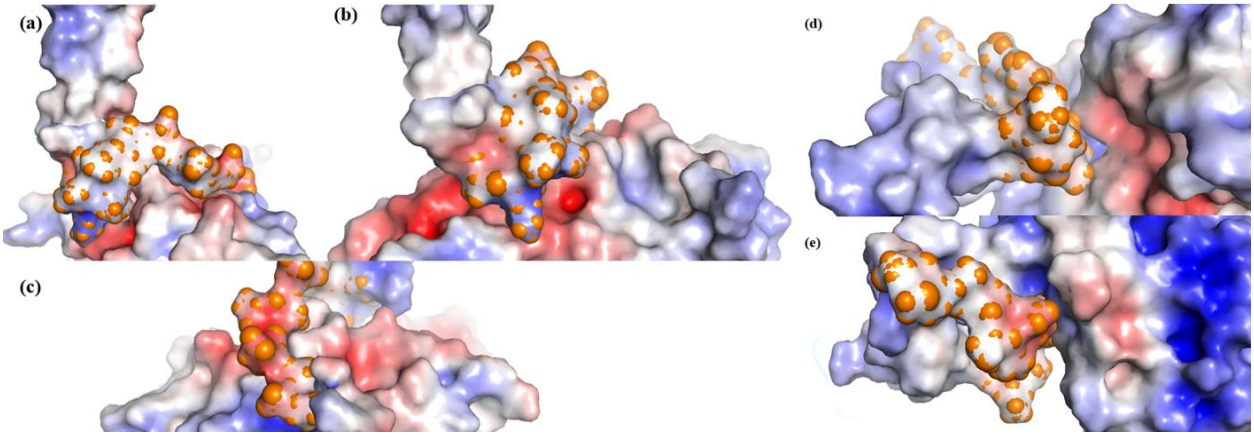


Figure 4.12: Electrostatic map of human CRIP1a and CB_1 complex interactions. The last 9 CB_1 C-terminal residues are visualized as orange spheres. **(a)-(c)** depict the electrostatic interactions in guided docking complex 116, while **(d) and (e)** show electrostatic interactions between blind docking complex 191.

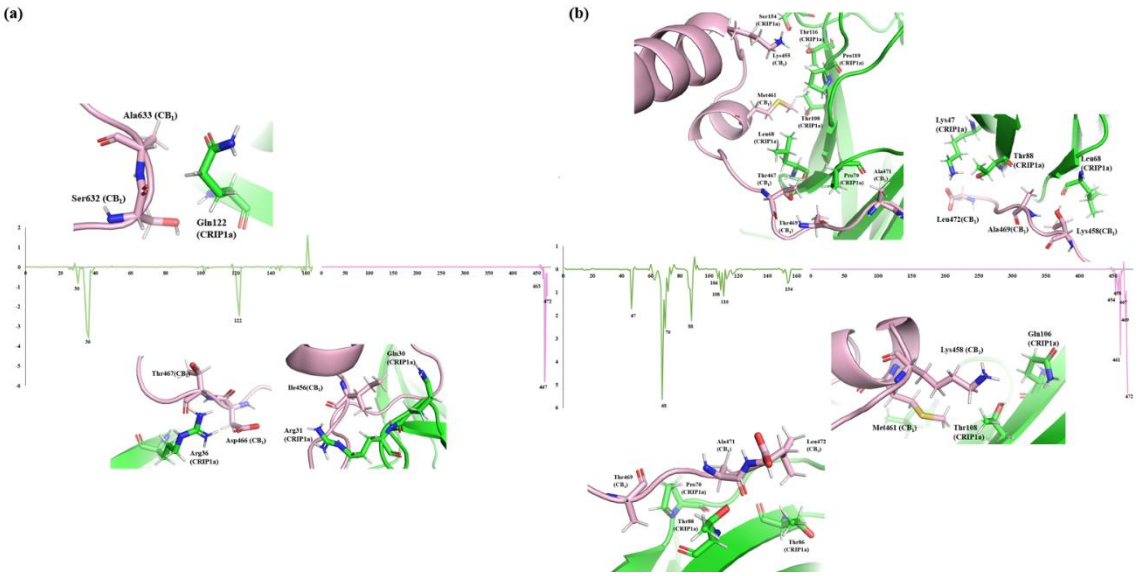


Figure 4.13: MMGBSA residue decomposition plot for blind and guided docking protocols. (a) complex 191 and **(b)** complex 116, residues participating in van der Waals and electrostatic interactions between CRIP1a and CB_1 protein-protein interactions.

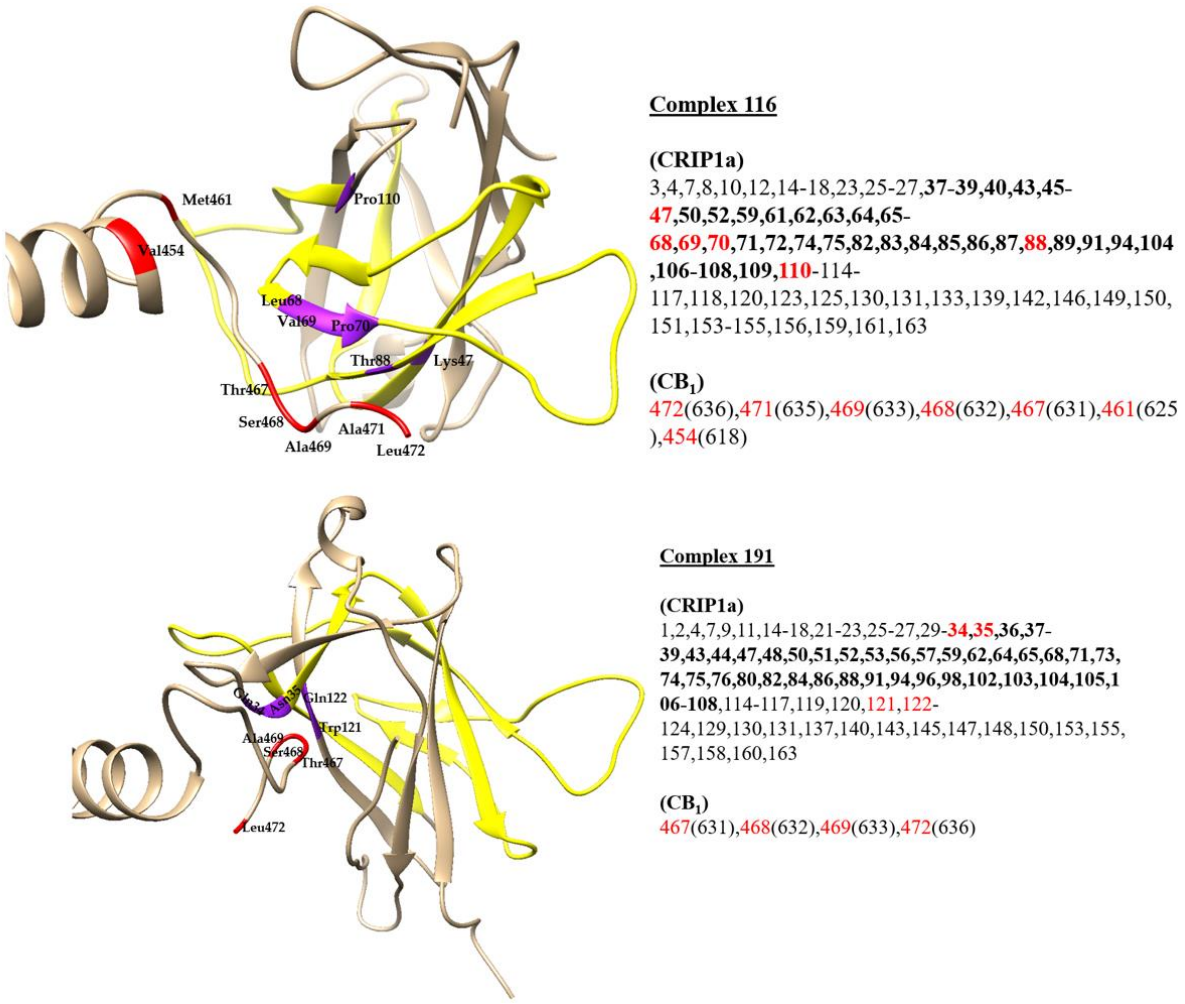


Figure 4.14: Residue interactions in guided (complex 116) and blind (complex 191) docking. Residues participating in strong average interactions are visualized in red (CB_1) and purple (CRIP1a). Residues participating in protein-protein binding interactions between CB_1 and CRIP1a for complex 116 and 191 are written to the left. Bold numbers (in text) and yellow region (in visualization) indicate binding region 34-110 from experimental studies[20,45]

4.3.7 Differences in the isolated protein vs. CRIP1a-CB_1 complex

To assess the impacts that the protein-protein interactions have on the dynamics of each protein, analysis of PCA, RMSF, and cross-correlation analysis was conducted on complex 116 in relation to the isolated CRIP1a and CB_1 systems.

4.3.7.1 Principal Component Analysis

The PCA was analyzed for the isolated CB_1 and CRIP1a proteins to compare the influence that the complex has on their dynamics. In terms of the CB_1 isolated protein, the last 100ns of the trajectory were analyzed using PCA due to the stability achieved within this region (**Figure 4.3**). A high degree of the proportional variance (**Figure 4.15**) was encompassed within the first eigenvalue (61.8%), which subsequently decreased in the level of significance after the second (9.7%) and third (6.3%) eigenvalues. The first three eigenvalues incorporated a total of 77.8% of the total proportional variance. Comparing the first, second and third principal component (PC) spaces, the degree of explored conformations is vast. To identify the patterns of motion in the CB_1 model, the projected eigenvectors (on the C-α backbone atoms) of the first three eigenvalues, which capture most of the maximal variances, were identified. From the ensemble for the first three PCs, we can see that the level of motion differs from each other. A common feature is that the N- (residues M1-Q116) and C- (residues R400-L472) termini, as well as the ICL3 (residues K300-T344), are having a large degree of fluctuations, where the first eigenvalue is capturing most of the dynamics, the number of captured fluctuations decreases as the PCs go on.

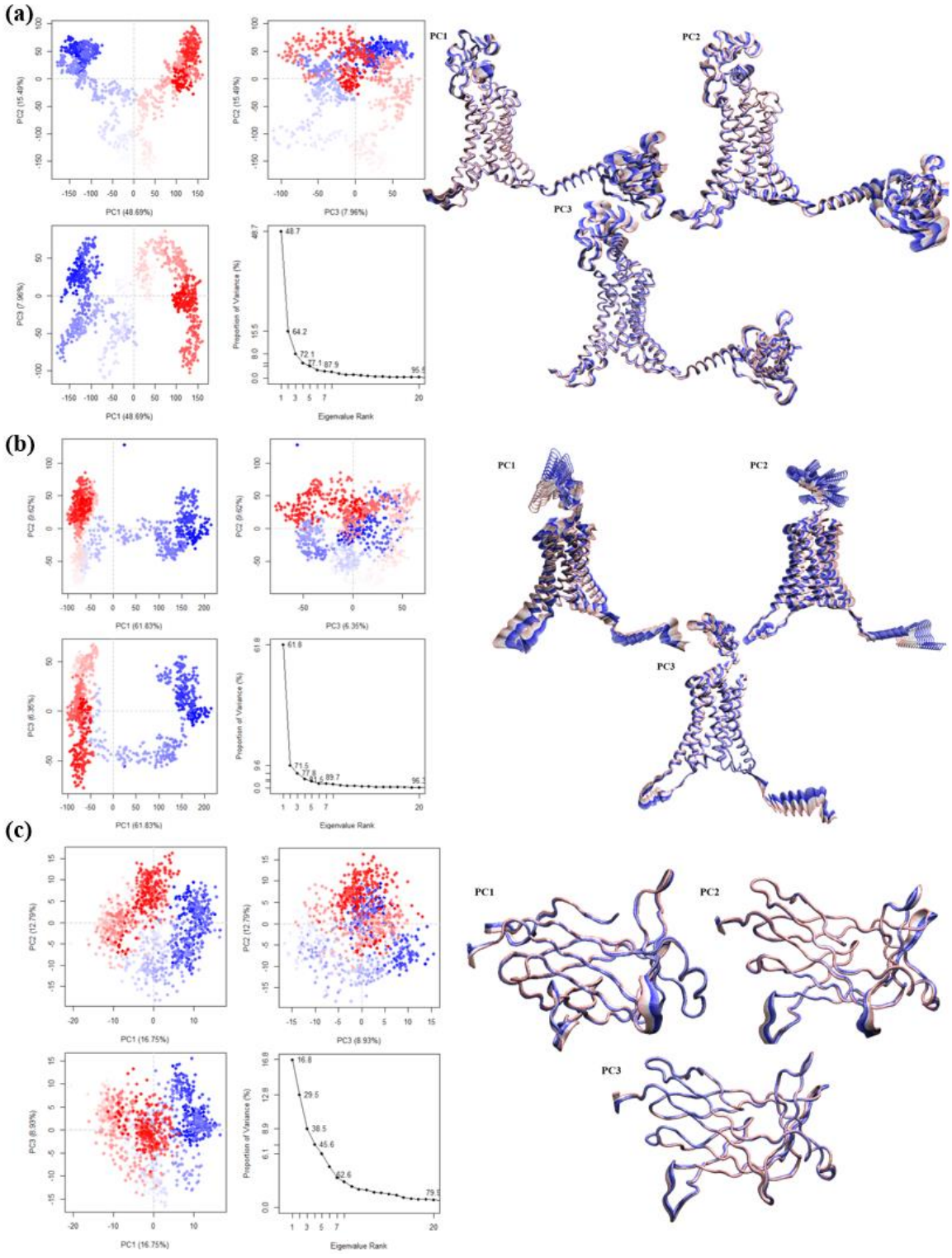


Figure 4.15: PCA plots and protein backbone fluctuations in each PC space. PCA plot and backbone fluctuations for **(a)** CB_1-CRIP1a complex **(b)** CB_1 isolated protein and **(c)** isolated CRIP1a protein for eigenvalue ranks 1, 2 and 3.

In comparison to CB_1, (**Figure 4.16)**, only a small proportion of dynamics (~16.8%) was captured by the PC1 in CRIP1a. Even the first three eigenvalues cumulatively described only 38.5% of the variations. Up to 20 PC spaces together were able to capture ~80% of variations suggesting the high dynamics of CRIP1a. In contrast, in the case of CB_1, 96.3% of the proportional variance was captured within the first 20 eigenvalues, while the top 3 PCs together described ~80% of variations. Upon visual inspection of CRIP1a for PC1, the main dynamic changes in the protein are seen at the N- and C-termini, as well as between residues: Ile16-Val23; Lys26-Arg36; Leu73-Tyr85; Thr91-Arg102; Lys148-Val158; $Met1^{N\text{-}terminus}$-$Pro5^{N\text{-}terminus}$. These areas were seen to have a great number of dynamics, where the fluctuations in the protein had gradually decreased throughout each subsequent PC. In PC2 the main fluctuations were within the same region: Gln17-Val23; Asp28-Gly33; Thr91-Gly100; Leu73-Val83; $Met1^{N\text{-}terminus}$-$Pro5^{N\text{-}terminus}$, whereas for PC3 the dynamic changes were mostly present within the regions: Lys74-Val84; Thr91-Gly100; $Met1^{N\text{-}terminus}$-$Leu4^{N\text{-}terminus}$.

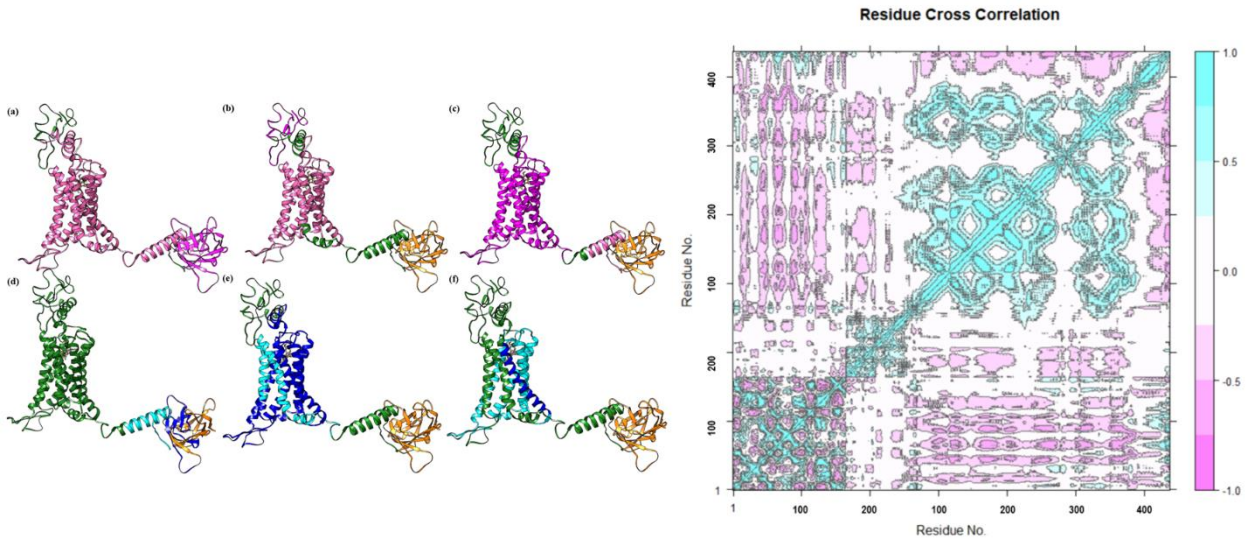


Figure 4.16: Cross-correlation plot of CB_1-CRIP1a complex. The plot indicated the negative dynamic and positive interactions between CRIP1a and CB_1 (CRIP1a values come first on the plot from 100-200, whereas CB_1 values preceded afterwards from 100-400). **(a)** corresponds to residues 66^{CB1}-466^{CB1} and 1^{CB1}-164^{CB1}, which display negative and neutral dynamic correlations **(b)** correspond to the MPR as well as the TM segments (residues 96^{CB1}-396^{CB1}) and the beginning unstructured region of CB1's N-terminal region (residues 1^{CB1}-66^{CB1}) which mediate mainly negative and neutral correlations between each other. **(c)** encompasses CB_1's C-terminal region

(residues 436^{CB1}-472^{CB1}) which has mainly negative correlations with the TM segment (residues 96^{CB1}-426^{CB1}). **(d)** depicts positive and neutral correlated interactions between residues 436^{CB1}-636^{CB1} and 50^{CRIP1a}-80^{CRIP1a},90^{CRIP1a}-130^{CRIP1a} which are the residues participating in the protein-protein interactions. **(e)** incorporates TM segments 1-5(residues 86^{CB1}-346^{CB1}) positive and neutral dynamic correlation with TM 7 and 6 (residues 326^{CB1}-426^{CB1}). Lastly **(f)** corresponds to TM 2-5 (residues 147^{CB1}-326^{CB1}) positive correlations with that of TM 1(residues 116^{CB1}-146^{CB1}).

4.3.7.2 CRIP1a-CB_1 complex

The amount of proportional variance explored within the complex's consecutive eigenvalues (**Figure 4.16)** was smaller than that of the isolated CB_1 protein but larger than that of the isolated CRIP1a system. This can be partially ascribed to the fact that due to the incorporation of two proteins, there is a greater threshold of exploration with a higher subset of residues than in the isolated protein systems. However, the presence of the protein-protein interactions introduces a stabilizing effect thus decreasing the total variability in the system, then in comparison to the highly variable isolated CRIP1a protein. This phenomenon was observed between the proportional variance in captured dynamics between the first (32.4%), second (21.9%) and third (11.1%) eigenvalues, equating to a total of 65.4% explored variance within the first three eigenvalues. There was also an absence of a drastic difference in the number of residual fluctuations captured between each eigenvalue. Overall, the amount of conformational exploration in the 20^{th} PC was the same as within the isolated CB_1 system, (92.1% of total conformational variance), demonstrating that the complex mediates a moderate degree of dynamic exportability. To get a better understanding of the types of fluctuations mediated within the protein-protein association, the backbone of the complex was visualized for the first three PCs. The first ranked eigenvalue captured a lot of the variability in the N- and C-termini regions of CB_1, while capturing fewer fluctuations in the TM region of the CB_1 receptor. In contrast, for CRIP1a, the entire protein's dynamics were encapsulated, with a heavier focus on the top (N-terminal) and central regions of the protein. These large fluctuations in PC1 are seen in residues: Ser425-Leu472($CB_1^{C\text{-terminus}}$);1-164(CRIP1a); HSD304-Ile339(CB_1^{ICL3}); Met1-Ile105($CB_1^{N\text{-terminus}}$). The second-ranked eigenvalue encapsulated most of the fluctuations within the entire CRIP1a protein, with a heavier emphasis on the bottom (C-terminal) and central region of the protein. In terms of CB_1, the C-

terminus' dynamics were mostly captured while a lower emphasis on the dynamics within the ICL3 and N-terminal regions. Specifically, the residues involved in the high dynamics are Ser410-Leu472($CB_1^{C\text{-terminus}}$);1-164(CRIP1a); HSD302-Leu341(CB_1^{ICL3}); and Met1-Phe102 ($CB_1^{N\text{-terminus}}$). The 3rd ranked PC, mostly captured the dynamics of the CB_1's N- and C- termini in addition to the ICL3 and a bit of the entire CRIP1a protein as: Val454-Leu472($CB_1^{C\text{-terminus}}$);1^{CRIP1a}-164^{CRIP1a}; Arg340-Gln314(CB_1^{ICL3}); Met1-Ile96 ($CB_1^{N\text{-terminus}}$).

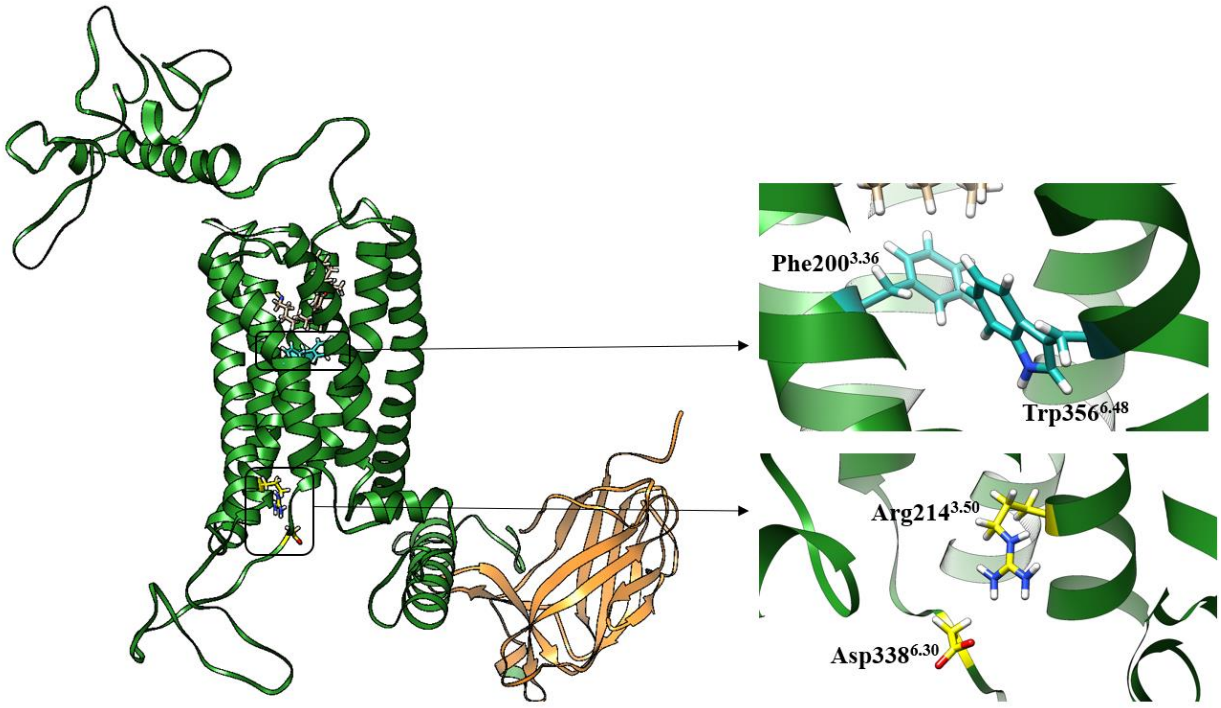


Figure 4.17: CRIP1a (orange) and CB_1 (forest green) complex formation. Visualization of conserved GPCR activated conformational characteristics. Breaking of the ionic lock between residues Arg214$^{3.50}$ and Asp338$^{6.30}$ (visualized in yellow) and mediated T-shaped π-π interactions amongst toggle switch residues Phe200$^{3.36}$ and Trp356$^{6.48}$ (visualized in blue).

4.3.7.3 Cross-Correlation Plot

In terms of CB_1's dynamic cross-correlation plot, it revealed the relationship between the external and internal segments with the TM component. In particular, the N- (residues M1-Q116) and C- (residues R400-L472) termini, as well as the ICL3 (residues K300-T344), were all negatively dynamically correlated to the TM segments of the protein. The segmental extremities (N- and C-

termini), as well as the large loop (ICL3), would fluctuate to a high degree, in contrast, the TM region was observed to encapsulate the most amount of stability over the entire trajectory. These observations are also in unison with the RMSF plots for CB_1. In terms of CRIP1a, an absence of concrete correlations could be drawn from the trajectory, where degrees of motion were more randomized. The only stabilizing segments in CRIP1a were observed within the central β-sheet regions, which displayed a more positive dynamic correlation with each other while mediating a negative dynamic relationship with the more disordered and highly variable loop regions within the protein.

4.3.7.4 CRIP1a-CB_1 Complex Cross-Correlation Plot

The cross-correlation plot for the CRIP1a-CB_1 complex shows the stabilizing effects emitted by the presence of the protein-protein interactions. Particularly, residues 66^{CB1}-426^{CB1} (CB_1) corresponding to the stable TM region of CB_1 were displayed to have negative as well as neutral correlations to most of the highly dynamic residues within the CRIP1a (residues 1^{CRIP1a}-164^{CRIP1a}) protein. Other negative and neutral correlations were observed between the highly dynamic unstructured loops within CB_1's N-terminus (residues1^{CB1}-66^{CB1}) and its very stabilized TM segments (residues 96^{CB1}-396^{CB1}). Mostly negative dynamic correlations were present between the highly variable CB_1's C-terminus (residues 436^{CB1}-472^{CB1}) and TM segment (residues 96^{CB1}-426^{CB1}).

Positive and neutral dynamic correlations expressed within the CRIP1a-CB_1 complex were observed to be between the C-terminus (helix 9 and terminal loops) of CB_1 (residues436^{CB1}-472^{CB1}) and the region of CRIP1a that performs in strong interactions with CB_1 (residues 40-80^{CRIP1a} and 90-130^{CRIP1a}). This observation demonstrates the positive dynamic impact that the binding interactions induce within the CRIP1a backbone to mediate these interactions. Other positive and neutral dynamic correlations were observed between TM 6 and 7 (residues 326^{CB1}-426^{CB1}) and TM 1-5 (residues 86^{CB1}-346^{CB1}). TM 1 (residues 116^{CB1}-146^{CB1}) was also observed to elicit mostly positive dynamic correlations with TM 2-5 (residues147^{CB1}-326^{CB1}), which are the TM segments involved in the conformational changes during inactivation and activation as well as within ligand binding. Specifically, TM6 undergoes an outwards helical rotation to accommodate G-protein binding, while TM1 and 2 change in conformation to accommodate the ligand binding and the size of the orthosteric pocket in the inactivate or activated state. TM3 and

6 encompass the toggle switch residues F200$^{3.36}$ and W356$^{6.48}$ which mediate π-π interactions (T-shaped in the activated state, and face-stacking in the inactivated state), as well as the ionic lock residues Asp338$^{6.30}$ and Arg214$^{3.50}$ which break upon activation (superscripts are indicative of the Ballesteros–Weinstein numbering[178]). In addition, in TM 5 and 7, there the inward movements of Tyr294$^{5.58}$ and 397$^{7.53}$ during GPCR activation.[114,125,179]

4.3.7.5 RMSF CRIP1a-CB$_1$ Complex

In comparison to the isolated CRIP1a and CB_1 proteins (**Figure 4.18**), the CRIP1a-CB_1 complex maintained a more stabilized CB_1, where the N- and C- terminal region, TM segments as well as the ICL3 had noticeable lower degrees of fluctuation, mediating an RMSF ranging from ~2 Å to 11 Å less than in the isolated complex. In contrast, the CRIP1a in the complex experienced more fluctuations in the backbone within the residues 36-120 surrounding and participating in intra-residue interactions, than in comparison to the isolated protein. The increased number of dynamics in the backbone between the residues participating in the protein-protein interactions can be ascribed to the highly dynamic nature of the C-terminal tail. Due to many disordered regions present in the C-terminal domain, thus causes for the backbone atoms in CRIP1a also become more flexible to accommodate and mediate interactions with the C-terminus. This phenomenon was also observed in the paper by Fayos *et al*., (2003)[180], where protein-protein interactions induced a higher degree of flexibility in the backbone atoms. Outside of these binding regions, residues 1^{CRIP1a}-35^{CRIP1a} and 121^{CRIP1a}-164 mediated a similar degree of fluctuation or an increased stabilizing effect in the case of residues 25-35.

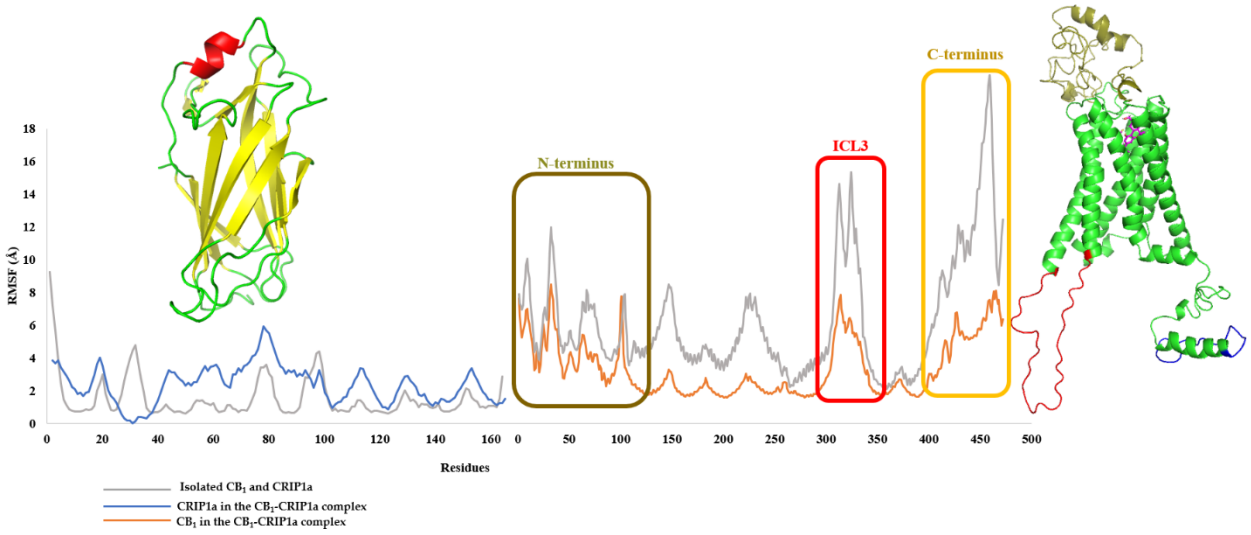


Figure 4.18: RMSF stabilizing protein-protein interaction effect in the presence of a protein complex, and without. RMSF comparison between the isolated CRIP1a and CB_1 proteins (grey), and the CRIP1a in complex (blue), along with the CB_1 in complex (orange).

4.4 Discussion

In this study, we built the first comprehensive model of the human CB_1-CRIP1a complex. The complete model of human CB_1 was constructed by building the N- and C-terminal domains of the protein separately using a threading-based approach and stitched them to an activated TM structure of the protein from PDB (5XR8[111]). The 3D model of CRIP1a was built using the homology modelling approach by using the recently published crystallographic rat CRIP1a structure (PDB: 6WSK[166]) as a template. The dominant conformations sampled from the MD trajectories of the two proteins were used to perform rigorous protein-protein docking and MD-based binding free energy calculations to predict the most likely model of the human CB_1-CRIP1a complex. Our model is in good agreement with the previously reported *in vitro* experimental data[20,45]which highlighted the regions that promote the CB_1-CRIP1a contacts. While previous modelling efforts[48] attempted to understand these protein-protein interactions exist, they were mostly focused on a reduced model of the receptor. Particularly, these studies had only used a portion of CB_1's C-terminal tail, and a CRIP1a model was constructed based on a low-homology template. Thus, there was a need for a comprehensive model to understand the molecules involved in the recognition process in the CB_1-CRIP1a complex; and this study addressed this gap. By application of a

combination of computational tools, we present the first comprehensive atomistic model that best describes the CB_1-CRIP1a associations and dynamics under physiological conditions.

Our research gives novel insights into how CRIP1a interacts with the activated CB_1 complex. The results generated from this study can be useful for understanding the competitive binding interactions between β-arrestin and CRIP1a to the distal region in CB_1's C-terminus, as well to design drugs to target CRIP1a. Overall, the data from these results can be used for future *in vitro* mutagenesis experiments to study the binding interactions between CRIP1a and CB_1 in more depth.

CHAPTER 5

General Discussion

5.1 General Discussion

This dissertation investigated the protein-protein and protein-ligand interactions of human CB_1 protein that modulates its signalling and physiological responses. Understanding the inactivated state of CB_1 has been of particular interest due to its role in different pathological conditions including inflammatory conditions[75], seizures[48], and metabolic disorders[74]. The inactivated state of the human CB_1 receptor has proved to be beneficial in reducing some of the pathological and toxicological side effects that high-affinity binding agonists induce. Instances in reduced cannabinoid-tolerance[181], decreased food-intake[23], and increased vasodilatory effects in cardiovascular disorders patients[29], are all beneficial mediated outcomes during the inactivation of the receptor. Despite the affluence in therapeutic outcomes retrieved from the inactivation of the receptor, limitations in the clinical utility of inactivating molecules have seen their share of controversy. Unlike the psychedelic adverse effects experienced during the activation of the receptor, inverse-agonists mediate a plethora of negative outcomes, observed amongst patients in the RIO-clinical trials.[93] Unwanted and harmful side effects such as decreased gastrointestinal mobility, the increased onset of mood disorders such as depression and anxiety, in addition to suicidal ideation, were some of the reported outcomes from ingestion of the cannabinoid antagonist, rimonabant.[64,94,182] Outcomes from these trials highlighted the importance of understanding the clinical utility- with the indoctrinated side effects – from the main therapeutical utility. Since the RIO trials, the ability to synthesize antagonists on the grounds of safety has remained elusive. Consequently, scientists have attempted to understand more about the protein-ligand conformations, binding interactions, and their roles in biased-signalling transduction, throughout decades. Unearthing more of the mysteries and uncertainties of the endocannabinoid system, and the ligand-binding interactions, were believed to generate more knowledge, so that synthesis of a safer ligand could be generated.

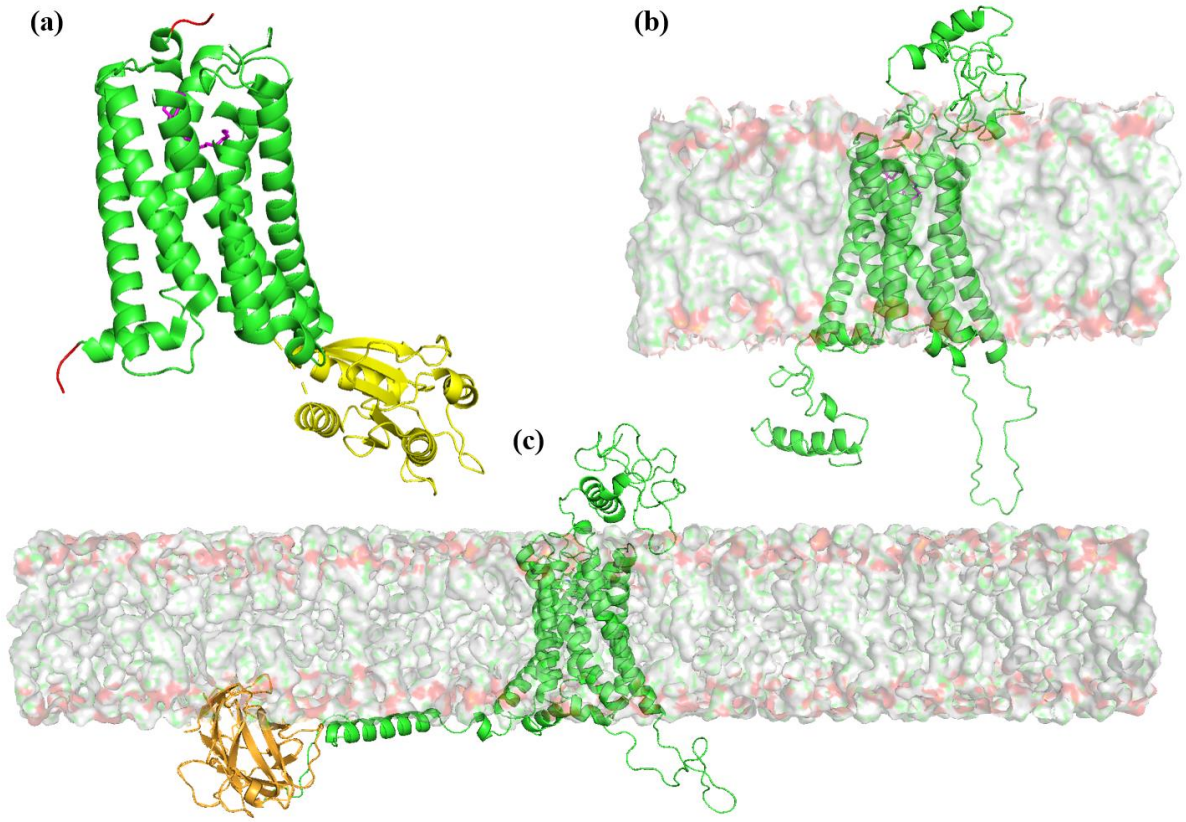


Figure 5.1: Construction process of the CB_1 receptor (a) 5XR8 crystallographic structure truncation of N- and C-termini (red), and insertion of stabilizing domain (yellow). **(b)** Complete structure with built N- and C- termini as well as ICL3 embedded in the POPC membrane. **(c)** Complete ribbon structure of the human CB_1 receptor (green) in association with the human CRIP1a protein (orange)

Another topic of discussion is the outlook on other binding targets, other than the traditional orthosteric pocket. In this regard, over the past couple of decades, researchers have made efforts at studying allosteric [125,183] and protein-protein associations[20,48,166], in addition to analyzing their binding sites and modes of interactions[46,105]. It is believed that targeting another site, other than the orthosteric pocket, may aid in reducing the toxicological side effects that orthosteric ligands promote. In effect, other non-traditional binding sites may provide safer targeting options, where they may offer more specificity for the receptor subtypes in comparison to the non-discriminatory cannabinergic ligands. It has been reported that orthosteric cannabinoids not only bind to CB_1 and

CB_2 but to other orphan CBRs, such as transient receptor potential cation channel subfamily V member 1 (TrpV1), thus creating more off-target side effects.[184] For these pre-described reasons, within this thesis we specifically looked at the influence of the N-terminal domain in the process of inverse-agonist binding, as well as the residues that play an important role during the molecular recognition process for CRIP1a-CB_1 protein-protein interactions in Chapters 2 and 3, respectively.

This thesis aimed to gain a deeper understanding of the inactivating protein-protein and protein-ligand interactions that the CB_1 receptor partakes in. For Chapter 3, we aimed to extract the importance of the N-terminus, in mediating protein-ligand associations, whereas Chapter 4, was focused on extracting the residues in CRIP1a and CB_1 that were participating in high-affinity binding interactions.

5.1.1 The N-terminus in ligand binding

In Chapter 3, the importance of the N-terminal domain's contribution to the receptor-inverse-agonist binding interactions was investigated by computational studies. The N-terminal domain of the human CB_1 receptor is a very large structural segment encompassing 116 residues, in comparison with its counterpart, CB_2, which contains 33 amino acids. As a result of its large length, numerous attempts to deduce the importance of the N-terminal domain in its relations to protein stability, ligand binding and overall functionality, have been reported in the past recent years. One of the first instances to outline the potential functionality and role of the N-terminal domain was first reported by Laprairie *et al*.,[108]. Within their study, they demonstrated that CBD reduced the affinity and efficacy of partial-agonists Δ^9-THC and 2-AG. It was proposed that CBD was bound close to the disulfide bond formation between residues C98 and C107, where mutation of these residues to alanine (C98A and C107A) led to a 50% reduction in CBD negative allosteric modulator activity than in comparison with the wild-type receptor. Computational studies conducted by Sabatucci *et al*.,[115], Jakowiecki *et al*.,[106] and Chung *et al*.,[107] had further investigated the allosteric binding of CBD-CB_1, in which Jakowiecki and co-workers[106] had assessed the influence of the whole N-terminal domain on the allosteric binding of CBD, under the influence of the partial agonist, Δ^9-THC. In their study, they evaluated the relevance of the CB_1's N-terminus within the presence of CBD, through the introduction of truncation sites at Δ-98, Δ-88, and Δ-8. [106] Through each subsequent truncation there was an increase in the ligand binding energy, reduced stability in the system, and disorientation of the CBD binding pose, compared to the wild-type

receptor. It was also reported that truncation of the N-terminus had resulted in the exiting of CBD from the allosteric site completely. Besides CBD, the disulfide bond located in the MPR has also been reported to affect allosteric modulators ORG27569 and PSNCBAM-1 binding to the CB_1 receptor.

When assessing the dynamics of the N-terminus Jakowiecki and co-workers[106] evaluated the variability of the N-terminal domain under different environments. They analyzed the stability of the N-terminus under the influence of CBD - Δ^9-THC, and Δ^9-THC alone. In their findings, they reported a conformational shift in the N-terminus during the Δ^9-THC only simulation, where the presence of NTH2(residues 61-69) disappeared, leaving behind a more stabilized NTH1(residues 29-35), NTH3(residues 78-92), and N-terminal cap (residues 1-20). In contrast, under the influence of the allosteric modulator, CBD, the N-terminus adopted three helical structures NTH1 (residues 30-35) NTH2 (residues 61-69), and NTH3 (residues 80-94), with a slightly different structural composition than in the Δ^9-THC only run. Despite unearthing the role and dynamics of the N-terminus in CBD and partial-agonist binding, Jakowiecki and co-workers[106], and Laprairie *et al.*,[108], had failed to analyze the impacts and stability of the N-terminal domain under the influence of inverse-agonist binding. For these reasons, we conducted the docking of different inverse-agonists to the CB_1 receptor.

Our results, from Chapter 3, more closely aligned with that of the Δ^9-THC only run[106], in which the N-terminal structure composition under the influence of inverse-agonists AM251, AM6538 and MJ15, had generated a stabilized NTH3 across systems. In contrast, a more variable NTH2 and NTH1 were apparent, in which their formation was highly dependent on the type of ligand binding. Likewise, in our simulations, we too observed the presence of a very stabilized N-terminal cap (residues 1-20) mediating interactions with the ECLs and TM regions of CB_1. Further results generated from Chapter 3 showcased the variability in the dynamics mediated within the N-terminal domain while giving further insights into its adopted conformation throughout the classical MD trajectory. In essence, these adopted conformations elucidated how the dynamics, and acquired N-terminal conformations play a role in ligand binding and receptor functionality. Our results demonstrated that the MPR and N-terminal cap encompass most of the dire interactions with the ligand, resulting in their high level of achieved stability. Moreover, these two stable regions throughout our trajectory were observed to cover the orthosteric pocket, potentially acting

as a shield from external solvent accessing the hydrophobic ligand binding site. In contrast, we observed that the N-terminal regions furthest from the ligand, such as NTH3 had experienced the highest degree of dynamics, where the stability of these regions was less dire, as they do not contribute to the ligand binding interactions.

The importance of studying the entire N-terminus and its participation in the ligand binding effects uncovers novel insights into the entire functionality of the receptor. Furthermore, it allows for further inclusion and analysis of an additional structural component that may play a role in the orthosteric and receptor-ligand induced conformations. It is these adopted conformations which are the main root cause of biased agonism and selective cellular signalling transduction. One of the first reported inverse-agonist and N-terminal functionalities was analyzed in the crystallographic paper by Shao *et al.*,[110] (PDB accession code:5U09[110]). The functionality of the N-terminus in the binding of taranabant was specifically observed within the confines of the MPR. The MPR was highlighted to have a potential role in solvent exclusion from the orthosteric pocket in CB_1 inactivated system. It has been noted in various papers[109,110,112] that the MPR forms a 'V-shaped' plug projecting itself into the ligand binding pocket, acting as a barrier from the extracellular solvent interacting with the inverse-agonist or antagonist. As also outlined in the crystallographic paper, by Shao and co-workers[110], the importance of the N-terminal domain in the predication of ligand-binding modes was deemed to be of utmost importance. Decreased resolution in the N-terminal domain had resulted in the inaccurate docking prediction of the inverse-agonist taranabant.[110] In essence, the importance of the MPR in inverse-agonist and allosteric systems have been discussed[108,185], however, the specific impacts of the entire N-terminus in inverse-agonist binding, remained unresolved. In our results, binding interactions mediated between the N-terminus and the inverse agonists, were primarily within the confines of the MPR, in which residues Glu100, Asn101, and Phe103 were observed mediating π-π and van der Waals interactions with the ligands. However, the strength of their interactions was largely dependent on the type of inverse-agonist binding and the conformation they induced at the orthosteric site. Therefore, concluding that the presence of MPR and inverse-agonists interactions are also ligand-dependent. In summary, our atomistic model describing the complete inactivated form of human CB_1 will be useful for understanding the interplay between different regions of the receptor with that of ligand binding to both orthosteric and allosteric sites.

5.1.2 CRIP1a interactions with the CB_1 receptor

The specific residues participating in the binding interactions between CRIP1a and CB_1 have been a more recent topic of discussion. *In vitro* deletion-based studies conducted by Niehaus *et al*.,[45] and Mascia *et al*.,[20] revealed the minimal binding domains for both CRIP1a and CB_1 to mediate their interactions. It was revealed that the last 9 amino acids in CB_1's C-terminal tail were found to be the minimal reported domain forming productive protein-protein interactions with the CRIP1a/b protein. Mutation of the terminal 9 amino acids in CB_1 was revealed to induce weak binding interactions between CB_1-CRIP1a/b. These results suggested that the distal region in the receptor (residues 464-472) is necessary to mediate strong interactions with CRIP1a/b. In an *in-vitro* binding assay study conducted by Blume *et al*.,[48] had analyzed for potential regions, other than the distal 9 amino acids in CB_1, involved in the binding of CRIP1a. Utilizing affinity pull-down studies; coimmunoprecipitation of CB_1 complexes with CRIP1a and β-arrestins, demonstrated that the central or distal C-terminal peptides had competed for association with CRIP1a. In addition, the presence of a phosphorylated Thr468 greatly reduced the CRIP1a-CB_1 association. Furthermore, the minimal number of residues involved in CRIP1a-CB_1 interactions to reduce endocytosis induced by β-arrestins was discovered to be D466, T467, S468, A471 and L472, by Mascia and colleagues[20]. Ultimately, these results conclude that the central and distal regions in the C-terminus play a role in forming interactions with the CRIP1a, whereas the five residues located in the distal 7 amino acids are required to stunt the process of CB_1 endocytosis.

When assessing the regions within CRIP1a important in the molecular recognition process with CB_1; using *in vitro* mutagenesis Niehaus *et al*.,[45] suggested that the region necessary for CRIP1a interaction with the receptor included exons 1 and 2 or amino acids 34-110. Further validating the experimental outcomes from Niehaus and colleagues' study, Mascia *et al*.,[20], had also discovered that residues 34-110 were sufficient to promote these protein-protein interactions. Despite these experimental studies highlighting the region of CRIP1a-CB_1 interactions, these results produced a large consensus binding region in CRIP1a where the specific residues predominantly inducing these interactions remained elusive. For these reasons, a few computational studies in the past have tried to isolate the main residues participating in these protein-protein interactions, by analyzing the specific residues participating in the CRIP1a-CB_1 associations.

Nevertheless, these computational studies[46,47,105,166] were not comprehensive, as they incorporated a low-quality template in their homology template to model the human CRIP1a protein (i.e. the model was based on the Rho-GDIβ structure (PDB accession code: 1DS6), which contains a 15.9% sequence identity and a low functional homology to the human CRIP1a protein).[46,47,105,166] In addition, these *in silico* models[46,47,105] had only used a small portion of the C-terminus, in their binding studies, without considering the entire CB_1 receptor and its C-terminal tail, in its interactions with CRIP1a. In essence, due to the low-quality template, and utilization of an incomplete C-terminal model, we concluded that their results leave room for many discrepancies and inaccuracies. These previous papers also did not consider the influence of the activated membrane-embedded receptor and the dynamics it induces at the C-terminal domain, which may play a part in the binding of CRIP1a, thus creating many uncertainties in their findings.

Chapter 4 of this thesis addressed this gap by constructing the first comprehensive atomistic model of the human CB_1-CRIP1a complex in its physiological environment, utilizing the first mammalian CRIP1a crystallographic structure[166]. Our model was consistent with earlier experimental findings[20,45] and was able to deduce the key molecular interactions that stabilize these complexes. Our model also demonstrated little complimentary to previous *in silico* studies in terms of formed interactions between that of CRIP1a and CB_1.[46,47,105] Instead, our results adopted a new set of residues partaking in the molecular-level recognition process between CRIP1a and CB_1. Given our results, we observed that the residues playing a large role in the molecular recognition process between the proteins are all within the experimentally determined regions[20,45] of 34-110 for the CRIP1a protein. Specifically, residues such as $Lys47^{CRIP1a}$, $Leu68^{CRIP1a}$, $Val69^{CRIP1a}$, $Pro70^{CRIP1a}$, $Thr88^{CRIP1a}$, and $Pro110^{CRIP1a}$, in CRIP1a engaged in strong binding interactions with the residues from CB_1 that included $Val151^{CB1}$, $Met461^{CB1}$, $Thr467^{CB1}$-$Ala469^{CB1}$, $Ala471^{CB1}$ and $Leu472^{CB1}$ in CB_1. Most of the residues involved in the formation of the protein-protein complex were found to mediate stronger electrostatic and/or hydrogen bond interactions. In terms of the CB_1 receptor, residues outside of the 9-distal C-terminal residues, such as helix 9, were also observed to support CB_1 interactions with CRIP1a. Dynamics elicited by the presence of the protein-protein interactions induced a high level of stability in the CB_1 receptor. All previous dynamic interaction regions in the N-terminus and the ICL3 became very stabilized, similar to the RMSF fluctuation values depicted with the inactivated receptor in Chapter 3. Interestingly, we found that the residues participating in the binding interactions between CRIP1a and the CB_1 receptor had induced higher

dynamics within the protein backbone, encompassing residues 34-110. This can be explained by the higher levels of fluctuations, and instability in the unstructured region within the C-terminal domain. Thus, to accommodate the increased flexibility of the C-terminus, CRIP1a must, too, adopt a higher degree of flexibility to mediate these interactions. Overall, our results from Chapter 4 give more detailed molecular-level insights into the inter-protein interactions stabilizing the human CB_1-CRIP1a complex.

5.2 Future Directions

The result of this present study provides further knowledge of the dynamic interplay between protein-protein and protein-ligand interactions with the CB_1 receptor, as well as the residues that cultivate these interactions. Further research is required to address certain topics this thesis did not cover such as CRIP1a alterations in CB_1-G-protein coupling. CRIP1a specifically regulates the signalling transduction capabilities of the CB_1 receptor through its alterations in G_α protein selectivity[99,166]. CRIP1a promotes the receptor coalition with the Gα½ subunit while impeding the formation with the $G\alpha_o$ and i_3 subunits.[99] Alterations in the G-protein coupling result in the downstream signalling events that are typically associated with CB_1 inactivation. Thus, more research is needed to address the specific residues, or modes of interaction that CRIP1a possesses in its decreased selectivity for the $G\alpha_o$ and i_3 subunits, in comparison to that of the Gα½ protein.

Another limitation of our study was the disregard for other factors impacting the CRIP1a-CB_1 association. CB_1 is also involved in protein-protein interactions with β-arrestins which promote the formation of a clathrin-dynamin complex when undergoing the process of CB_1-endocytosis.[99,166] Constant exposure to agonists, results in the phosphorylation of the C-terminal tail at two sites.[99] This phosphorylation promotes the association with the β-arrestins, and a reduction in the number of CB_1 receptors present at the membrane surface.[48,166] These events, increase tolerance that is typically seen in active users of marijuana.[161] Through competitive interactions, β-arrestins and CRIP1a compete for the association at the C-terminal tail.[48,166] However, during the presence of phosphorylation, association with β-arrestins is greatly favoured, reducing the binding interactions with CRIP1a.[48] To further generate a deeper analysis of the reduction in CRIP1a interactions under the presence of phosphorylation, the introduction of the two phosphoserines, S425 and S429, and their impacts on CRIP1a binding is necessary, to gain a deeper understanding of the three-protein binding interplay.

Unclear from the present investigation is how the presence of differing degrees of N-terminal lengths impacts the stability and binding affinity interactions with inverse-agonists to the CB_1 receptor. As evident in the study conducted by Jakowiecki *et al.*,[106], truncation of the N-terminus at different sites, following a re-calculation of the ligand binding energy, can give a more definitive analysis of the importance of the entire N-terminal domain during inactivation studies. Due to the ligand bias presented in crystallographic structures, the model used in this thesis favours the association of inactivating cannabinoid molecules. Thus, utilization of our model in the future virtual screening of inverse-agonists, antagonists and allosteric modulators can further address the impacts that the entire N-terminus has on ligand binding.

Overall, the goal of this thesis was to analyze the influence that a complete and accurate 3D computational model had during protein-protein and protein-ligand interactions. We established that modelling the complete CB_1 receptor is dire when generating a full picture of the dynamics, stability and formation of the protein and ligand contacts. Despite this thesis answering some questions pertaining to these different associations, certain topics such as CRIP1a's influence on G-protein coupling, and N-terminal truncation outcomes on inverse-agonist binding, were not covered. Therefore, continued research on the different binding interactions mediated at the CB_1 receptor is necessary. Understanding the interplay between CB_1 ligand-binding and protein-protein interactions will help to create a new avenue in the generation of future cannabinoid-based pharmaceuticals and analysis of novel potential drug targets.

5.3 Conclusions and Implications

Results extracted from our studies, highlight the importance of incorporating complete and accurate 3D-computational models during structural and inter-molecular analysis. Our findings suggest that the presence of the CB_1 N- and C-terminal domains could impact the affinity and interactions of the ligand within the orthosteric site. In particular, the presence of segments such as MPR is important when assessing and ranking inverse-agonists and antagonists' binding-affinity predictions and rankings, as well as generating accurate docking modes, as seen within our study. [106,110,116,145] In addition to generating more accurate binding affinity predictions, we have also established a more comprehensive network of receptor-ligand binding contacts, that extend beyond the TM and ECL region. These findings are in agreement with the experimental study conducted by Shao *et al.*,[110] which also highlighted the importance of the CB_1's N-terminal

segment. Thus, our model presents an opportunity to facilitate a more accurate, structure-based virtual screening of compounds. Given that the number of cannabinoid-based ligands (both synthetic and plant-based) is steadily increasing, the complete model of human CB_1 presented in this thesis can be a vital toolkit for understanding the molecular pharmacology of these ligands.

Further, we extended our research into building the first comprehensive atomistic model of the human CB_1-CRIP1a complex. In essence, the full protein dynamics, stability and binding associations have been shown in our results to be affected by the presence of the complete receptor. Through this model, we were able to delineate the key residues that facilitate the molecular recognition between CB_1 and CRIP1a. This model also provided an important hypothesis that some residues beyond the C-terminal tail of CB_1 could play a role in its binding to CRIP1a.

The significance of our results demonstrates the importance of the inclusion of the entire receptor—N- and C- termini and connecting loops— in CADD, as well as the residues involved in the molecular recognition process between CRIP1a and CB_1. Insights from these conclusions can be used to drive future *in-vitro* studies, such as mutational studies—mutating the residues that play a large role in the CRIP1a-CB_1 associations—, or certain applications in medicinal chemistry— to design drugs that can target the residues mediating high-affinity interactions between the two proteins— to therefore alter their binding association. Outcomes from these *in vitro* experiments can also have various applications in generating anti- nociceptive and emesis effects. Moreover, evaluations from our N-terminal and inverse-agonist/antagonist studies have also displayed the necessity of the N-terminus in drug discovery and the structural analysis of the receptor. The N-terminal's association with the protein-ligand complex, and its impact on allosteric binding (as proven by other computational studies[106,186]), have highlighted the impact and importance of the N-terminus. Conclusions from these results, give rise to the generation of more accurate protein-ligand binding interactions, and affinities, as well as docking and simulation analysis, which can be pivotal aspects in the field of computational drug discovery such as within high-throughput virtual screening.

Given the data from this thesis, scientists specializing in crystallography should attempt to extract more residues within the confines of the MPR and at a higher resolution. As outlined by Shao *et al.*,[110], the generation of higher quality resolutions of the MPR can aid in producing more accurate and reproducible docking results, specifically in the case of antagonists/inverse-agonists. In

addition, scientists within the computational field should look towards incorporating full models when assessing the impacts of protein-protein interactions, to get a full comprehension of the dynamics and binding influence under the presence of two proteins.

Overall, the overarching aim of this thesis was to add to the growing literature on the various interactions that CB_1 participates in, and how we can use these interactions as future potential drug targets or when synthesizing inverse-agonists or antagonists. Despite the existence of multiple experimental[20,45,48] and *in-silico*[46,47,105,106] studies analyzing the various dynamic and binding interactions of CB_1, often these studies do not go into specificities nor answer all outstanding questions within the field. Another discerning factor with previous computational studies is potential inaccuracies in older developed models. With the fast advancement of the cannabinoid field, new information is constantly being published, where the establishment of novel crystallographic structures gives rise to the generation of more productive templates to be used in computational studies. This subjects older models at a disadvantage, especially if there was a decreased amount of known information or crystallographic tools made available at the time when the study was conducted. Often, certain factors are over-looked, such as the inclusion of the entire CB_1 receptor or the N-terminal domain during protein-protein and protein-ligand studies, but, as we have shown, the presence of the entire receptor's structure may have dire roles in ligand stability, solvent exclusion, and protein-protein formation. As further information is made available about the CB_1 receptor and its various interactions, awareness of the implications of accurate and complete models in computational studies is increasingly important. Continued research of the dynamic interplay of CB_1 binding partners and their modes of interactions is required to fully elucidate the complete understanding of the mediated effects at the receptor.

REFERENCES

1. Gaoni, Y. & Mechoulam, R. Isolation, Structure, and Partial Synthesis of an Active Constituent of Hashish. *J. Am. Chem. Soc.* (1964) doi:10.1021/ja01062a046.

2. Mechoulam, R., Shani, A., Edery, H. & Grunfeld, Y. Chemical basis of hashish activity. *Science (80-.).* **169**, 611–612 (1970).

3. Silver, R. J. The endocannabinoid system of animals. *Animals* **9**, (2019).

4. Mechoulam, R. & Parker, L. A. The endocannabinoid system and the brain. *Annu. Rev. Psychol.* **64**, 21–47 (2013).

5. Di Marzo, V. New approaches and challenges to targeting the endocannabinoid system. *Nat. Rev. Drug Discov.* **17**, 623–639 (2018).

6. Silva, M. T. A., Carlini, E. A., Claussen, U. & Korte, F. Lack of cross-tolerance in rats among (-) Δ9-trans-tetrahydrocannabinol (Δ9-THC), cannabis extract, mescaline and lysergic acid diethylamide (LSD-25). *Psychopharmacologia* **13**, 332–340 (1968).

7. Sofia, R. D., Nalepa, S. D., Harakal, J. J. & Vassar, H. B. Anti-edema and analgesic properties of Δ9-tetrahydrocannabinol (THC). *J. Pharmacol. Exp. Ther.* **186**, 646–655 (1973).

8. Fernandes, M., Schabarek, A., Coper, H. & Hill, R. Modification of Δ9-THC-actions by cannabinol and cannabidiol in the rat. *Psychopharmacologia* **38**, 329–338 (1974).

9. Koh, W. S., Crawford, R. B. & Kaminski, N. E. Inhibition of protein kinase A and cyclic AMP response element (CRE)-specific transcription factor binding by Δ9-tetrahydrocannabinol (Δ9-THC): A putative mechanism of cannabinoid-induced immune modulation. *Biochem. Pharmacol.* **53**, 1477–1484 (1997).

10. Loewe, S. Active principals of the cannabis and the pharmacology of the cannabinols. *Naunyn. Schmiedebergs. Arch. Exp. Pathol. Pharmakol.* **211**, 175–193 (1950).

11. Bloemendal, V. R. L. J., van Hest, J. C. M. & Rutjes, F. P. J. T. Synthetic pathways to tetrahydrocannabinol (THC): an overview. *Org. \& Biomol. Chem.* **18**, 3203–3215 (2020).

12. Domino, E. F. Neuropsychopharmacologic studies of marijuana: Some synthetic and

natural THC derivatives in animals and man. (1971).

13. Barth, F. & Rinaldi-Carmona, M. The development of cannabinoid antagonists. *Curr. Med. Chem.* **6**, 745–756 (1999).

14. Rinaldi-Carmona, M., Calandra, B., Shire, D., Bouaboula, M., Oustric, D., Barth, F., Casellas, P., Ferrara, P. & Le Fur, G. Characterization of two cloned human CB1 cannabinoid receptor isoforms. *J. Pharmacol. Exp. Ther.* **278**, 871–878 (1996).

15. Matsuda, L. A., Lolait, S. J., Brownstein, M. J., Young, A. C. & Bonner, T. I. Structure of a cannabinoid receptor and functional expression of the cloned cDNA. *Nature* **346**, 561–564 (1990).

16. Munro, S., Thomas, K. L. & Abu-Shaar, M. Molecular characterization of a peripheral receptor for cannabinoids. *Nature* **365**, 61–65 (1993).

17. De Petrocellis, L., Melck, D., Palmisano, A., Bisogno, T., Laezza, C., Bifulco, M. & Di Marzo, V. The endogenous cannabinoid anandamide inhibits human breast cancer cell proliferation. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8375–8380 (1998).

18. Eissenstat, M. A., Bell, M. R., D'Ambra, T. E., Alexander, E. J., Daum, S. J., Ackerman, J. H., Gruett, M. D., Kumar, V., Estep, K. G., Olefirowicz, E. M., Wetzel, J. R., Alexander, M. D., Weaver, J. D., Haycock, D. A., Luttinger, D. A., Casiano, F. M., Chippari, S. M., Kuster, J. E., Stevenson, J. I., Ward, S J., Aminoalkylindoles: Structure-Activity Relationships of Novel Cannabinoid Mimetics. *J. Med. Chem.* **38**, 3094–3105 (1995).

19. Shao, Z., Yan, W., Chapman, K., Ramesh, K., Ferrell, A. J., Yin, J., Wang, X., Xu, Q. & Rosenbaum, D. M. Structure of an allosteric modulator bound to the CB1 cannabinoid receptor. *Nat. Chem. Biol.* **15**, 1199–1205 (2019).

20. Mascia, F., Klotz, L., Lerch, J., Ahmed, M. H., Zhang, Y. & Enz, R. CRIP1a inhibits endocytosis of G-protein coupled receptors activated by endocannabinoids and glutamate by a common molecular mechanism. *J. Neurochem.* **141**, 577–591 (2017).

21. Hebert-Chatelain, E., Desprez, T., Serrat, R., Bellocchio, L., Soria-Gomez, E., Busquets-Garcia, A., Pagano-Zottola., A. C., Delamarre, A., Cannich, A., Vincent, P., Varilh, M., Robin, L. M., Terral, G., García-Fernández, M. D., Colavita, M., Mazier, W., Drago, F.,

Puente, N., Reguero, L., Elezgarai, I., Dupuy, J. W., Cota, D., Lopez-Rodriguez, M. L., Barreda-Gómez, G., Massa, F., Grandes, P., Bénard, G., Marsicano, G., A cannabinoid link between mitochondria and memory. *Nature* **539**, 555–559 (2016).

22. Andres-Mach, M., Haratym-Maj, A., Zagaja, M., Rola, R., Maj, M., Chrościńska-Krawczyk, M. & Luszczki, J. J. ACEA (a highly selective cannabinoid CB1 receptor agonist) stimulates hippocampal neurogenesis in mice treated with antiepileptic drugs. *Brain Res.* **1624**, 86–94 (2015).

23. Di Marzo, V., Goparaju, S. K., Wang, L., Liu, J., Bátkai, S., Járai, Z., Fezza, F., Miura, G. I., Palmiter, R. D., Sugiura, T. & Kunos, G. Leptin-regulated endocannabinoids are involved in maintaining food intake. *Nature* **410**, 822–825 (2001).

24. Gerdeman, G. L., Ronesi, J. & Lovinger, D. M. Postsynaptic endocannabinoid release is critical to long-term depression in the striatum. *Nat. Neurosci.* **5**, 446–451 (2002).

25. De Giacomo, V., Ruehle, S., Lutz, B., Häring, M. & Remmers, F. Differential glutamatergic and GABAergic contributions to the tetrad effects of $Δ$9-tetrahydrocannabinol revealed by cell-type-specific reconstitution of the CB1 receptor. *Neuropharmacology* **179**, 108287 (2020).

26. Black, N., Stockings, E., Campbell, G., Tran, L. T., Zagic, D., Hall, W. D., Farrell, M. & Degenhardt, L. Cannabinoids for the treatment of mental disorders and symptoms of mental disorders: a systematic review and meta-analysis. *The Lancet Psychiatry* (2019) doi:10.1016/S2215-0366(19)30401-8.

27. Hungund, B. L., Vinod, K. Y., Kassir, S. A., Basavarajappa, B. S., Yalamanchili, R., Cooper, T. B., Mann, J. J. & Arango, V. Upregulation of CB1 receptors and agonist-stimulated [35S]GTPγS binding in the prefrontal cortex of depressed suicide victims. *Mol. Psychiatry* (2004) doi:10.1038/sj.mp.4001376.

28. Pacher, P., Bátkai, S. & Kunos, G. Haemodynamic profile and responsiveness to anandamide of TRPV1 receptor knock-out mice. *J. Physiol.* **558**, 647–657 (2004).

29. Eid, B. G. Cannabinoids for treating cardiovascular disorders: putting together a complex puzzle. *J. Microsc. Ultrastruct.* **6**, 171 (2018).

30. Rosenstock, J., Hollander, P., Chevalier, S. & Iranmanesh, A. SERENADE: The study evaluating rimonabant efficacy in drug-naive diabetic patients: Effects of monotherapy with rimonabant, the first selective CB 1 receptor antagonist, on glycemic control, body weight, and lipid profile in drug-naive type 2 diabetes. *Diabetes Care* **31**, 2169–2176 (2008).

31. Camilleri, M. Cannabinoids and gastrointestinal motility: Pharmacology, clinical effects, and potential therapeutics in humans. *Neurogastroenterol. Motil.* **30**, 1–10 (2018).

32. El-Talatini, M. R., Taylor, A. H., Elson, J. C., Brown, L., Davidson, A. C. & Konje, J. C. Localisation and function of the endocannabinoid system in the human ovary. *PLoS One* **4**, (2009).

33. Cecconi, S., Rapino, C., Di Nisio, V., Rossi, G. & Maccarrone, M. The (endo)cannabinoid signaling in female reproduction: What are the latest advances? *Prog. Lipid Res.* **77**, 101019 (2020).

34. Park, B., McPartland, J. M. & Glass, M. Cannabis, cannabinoids and reproduction. *Prostaglandins Leukot. Essent. Fat. Acids* **70**, 189–197 (2004).

35. Amenta, P. S., Jallo, J. I., Tuma, R. F., Craig Hooper, D. & Elliott, M. B. Cannabinoid receptor type-2 stimulation, blockade, and deletion alter the vascular inflammatory responses to traumatic brain injury. *J. Neuroinflammation* **11**, 1–10 (2014).

36. Fletcher-Jones, A., Hildick, K. L., Evans, A. J., Nakamura, Y., Henley, J. M. & Wilkinson, K. A. Protein Interactors and Trafficking Pathways That Regulate the Cannabinoid Type 1 Receptor (CB1R). *Front. Mol. Neurosci.* **13**, (2020).

37. Zou, S. & Kumar, U. Cannabinoid receptors and the endocannabinoid system: Signaling and function in the central nervous system. *Int. J. Mol. Sci.* **19**, (2018).

38. Patel, M., Finlay, D. B. & Glass, M. Biased agonism at the cannabinoid receptors – Evidence from synthetic cannabinoid receptor agonists. *Cell. Signal.* **78**, 109865 (2021).

39. Diana, M. A. & Marty, A. Endocannabinoid-mediated short-term synaptic plasticity: Depolarization-induced suppression of inhibition (DSI) and depolarization- induced suppression of excitation (DSE). *Br. J. Pharmacol.* **142**, 9–19 (2004).

40. Childers, S. R., Pacheco, M. A., Bennett, B. A., Edwards, T. A., Hampson, R. E., Mu, J. & Deadwyler, S. A. Cannabinoid receptors: G-protein-mediated signal transduction mechanisms. *Biochem. Soc. Symp.* **59**, 27–50 (1993).

41. Ferré, S., Karcz-Kubicha, M., Hope, B. T., Popoli, P., Burgueño, J., Gutiérrez, M. A., Casadó, V., Fuxe, K., Goldberg, S. R., Lluis, C., Franco, R. & Ciruela, F. Synergistic interaction between adenosine A2A and glutamate mGlu5 receptors: Implications for striatal neuronal function. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11940–11945 (2002).

42. Sachdev, S., Banister, S. D., Santiago, M., Bladen, C., Kassiou, M. & Connor, M. Differential activation of G protein-mediated signaling by synthetic cannabinoid receptor agonists. *Pharmacol. Res. Perspect.* **8**, (2020).

43. Patel, M., Manning, J. J., Finlay, D. B., Javitch, J. A., Banister, S. D., Grimsey, N. L. & Glass, M. Signalling profiles of a structurally diverse panel of synthetic cannabinoid receptor agonists. *Biochem. Pharmacol.* **175**, 113871 (2020).

44. Bonhaus, D. W., Chang, L. K., Kwan, J. & Martin, G. R. Dual activation and inhibition of adenylyl cyclase by cannabinoid receptor agonists: Evidence for agonist-specific trafficking of intracellular responses. *J. Pharmacol. Exp. Ther.* **287**, 884–888 (1998).

45. Niehaus, J. L., Liu, Y., Wallis, K. T., Egertová, M., Bhartur, S. G., Mukhopadhyay, S., Shi, S., He, H., Selley, D. E., Howlett, A. C., Elphick, M. R. & Lewis, D. L. CB1 cannabinoid receptor activity is modulated by the cannabinoid receptor interacting protein CRIP 1a. *Mol. Pharmacol.* **72**, 1557–1566 (2007).

46. Singh, P., Ganjiwale, A., Howlett, A. C. & Cowsik, S. M. Molecular Interaction between Distal C-Terminal Domain of the CB1 Cannabinoid Receptor and Cannabinoid Receptor Interacting Proteins (CRIP1a/CRIP1b). *J. Chem. Inf. Model.* **59**, 5294–5303 (2019).

47. Ahmed, M. H., Kellogg, G. E., Selley, D. E., Safo, M. K. & Zhang, Y. Predicting the molecular interactions of CRIP1a-cannabinoid 1 receptor with integrated molecular modeling approaches. *Bioorganic Med. Chem. Lett.* **24**, 1158–1165 (2014).

48. Blume, L. C., Patten, T., Eldeeb, K., Leone-Kabler, S., Ilyasov, A. A., Keegan, B. M., O'Neal, J. E., Bass, C. E., Hantgan, R. R., Lowther, W. T., Selley, D. E. & Howlett, A. C.

Cannabinoid receptor interacting protein 1a competition with β-arrestin for CB1 receptor binding sites. *Mol. Pharmacol.* **91**, 75–86 (2017).

49. Guggenhuber, S., Alpar, A., Chen, R., Schmitz, N., Wickert, M., Mattheus, T., Harasta, A. E., Purrio, M., Kaiser, N., Elphick, M. R., Monory, K., Kilb, W., Luhmann, H. J., Harkany, T., Lutz, B. & Klugmann, M. Cannabinoid receptor-interacting protein Crip1a modulates CB1 receptor signaling in mouse hippocampus. *Brain Struct. Funct.* **221**, 2061–2074 (2016).

50. Gruden, G., Barutta, F., Kunos, G. & Pacher, P. Role of the endocannabinoid system in diabetes and diabetic complications. *Br. J. Pharmacol.* **173**, 1116–1127 (2016).

51. Davis, M. P. Cannabinoids in pain management: CB1, CB2 and non-classic receptor ligands. *Expert Opin. Investig. Drugs* **23**, 1123–1140 (2014).

52. Grinspoon, Lester and Bakalar, J. B. Marihuana: the forbidden medicine. *Am. J. Psychiatry* **151**, 1091–1091 (1994).

53. Bishay, P., Schmidt, H., Marian, C., Häussler, A., Wijnvoord, N., Ziebell, S., Metzner, J., Koch, M., Myrczek, T., Bechmann, I., Kuner, R., Costigan, M., Dehghani, F., Geisslinger, G. & Tegeder, I. R-flurbiprofen reduces neuropathic pain in rodents by restoring endogenous cannabinoids. *PLoS One* **5**, (2010).

54. Blankman, J. L., Simon, G. M. & Cravatt, B. F. A Comprehensive Profile of Brain Enzymes that Hydrolyze the Endocannabinoid 2-Arachidonoylglycerol. *Chem. Biol.* **14**, 1347–1356 (2007).

55. Long, J. Z., Li, W., Booker, L., Burston, J. J., Kinsey, S. G., Schlosburg, J. E., Pavón, F. J., Serrano, A. M., Selley, D. E., Parsons, L. H., Lichtman, A. H. & Cravatt, B. F. Selective blockade of 2-arachidonoylglycerol hydrolysis produces cannabinoid behavioral effects. *Nat. Chem. Biol.* **5**, 37–44 (2009).

56. Caffarel, M. M., Andradas, C., Mira, E., Pérez-Gómez, E., Cerutti, C., Moreno-Bueno, G., Flores, J. M., García-Real, I., Palacios, J., Mañes, S., Guzmán, M. & Sánchez, C. Cannabinoids reduce ErbB2-driven breast cancer progression through Akt inhibition. *Mol. Cancer* **9**, 1–11 (2010).

57. Soliman, E. & Van Dross, R. Anandamide-induced endoplasmic reticulum stress and apoptosis are mediated by oxidative stress in non-melanoma skin cancer: Receptor-independent endocannabinoid signaling. *Mol. Carcinog.* **55**, 1807–1821 (2016).

58. Morales, P. & Reggio, P. H. CBD: A New Hope? *ACS Med. Chem. Lett.* **10**, 694–695 (2019).

59. Galve-Roperh, I., Sánchez, C., Cortés, M. L., Del Pulgar, T. G., Izquierdo, M. & Guzmán, M. Anti-tumoral action of cannabinoids: Involvement of sustained ceramide accumulation and extracellular signal-regulated kinase activation. *Nat. Med.* **6**, 313–319 (2000).

60. Fiore, D., Ramesh, P., Proto, M. C., Piscopo, C., Franceschelli, S., Anzelmo, S., Medema, J. P., Bifulco, M. & Gazzerro, P. Rimonabant kills colon cancer stem cells without inducing toxicity in normal colon organoids. *Front. Pharmacol.* **8**, 1–15 (2018).

61. Janero, D. R. & Makriyannis, A. Cannabinoid receptor antagonists: Pharmacological opportunities, clinical experience, and translational prognosis. *Expert Opin. Emerg. Drugs* **14**, 43–65 (2009).

62. Marzo, V. Di, Goparaju, S. K., Wang, L. & Liu, J. Are Involved in Maintaining Food Intake. *Animals* **410**, 822–825 (2001).

63. Di Marzo, V., Bifulco, M. & De Petrocellis, L. The endocannabinoid system and its therapeutic exploitation. *Nat. Rev. Drug Discov.* **3**, 771–784 (2004).

64. Izzo, A. A. & Camilleri, M. Emerging role of cannabinoids in gastrointestinal and liver diseases: Basic and clinical aspects. *Gut* **57**, 1140–1155 (2008).

65. Izzo, A. A., Fezza, F., Capasso, R., Bisogno, T., Pinto, L., Iuvone, T., Esposito, G., Mascolo, N., Di Marzo, V. & Capasso, F. Cannabinoid CB1-receptor mediated regulation of gastrointestinal motility in mice in a model of intestinal inflammation. *Br. J. Pharmacol.* **134**, 563–570 (2001).

66. Mascolo, N., Izzo, A. A., Ligresti, A., Costagliola, A., Pinto, L., Cascio, M. G., Maffia, P., Cecio, A., Capasso, F. & Di Marzo, V. The endocannabinoid system and the molecular basis of paralytic ileus in mice. *FASEB J.* **16**, 1973–1975 (2002).

67. Massa, F., Marsicano, G., Hermana, H., Cannich, A., Monory, K., Cravatt, B. F., Ferri, G. L., Sibaev, A., Storr, M. & Lutz, B. The endogenous cannabinoid system protects against colonic inflammation. *J. Clin. Invest.* **113**, 1202–1209 (2004).

68. Izzo, A. A., Mascolo, N., Pinto, L., Capasso, R. & Capasso, F. The role of cannabinoid receptors in intestinal motility, defaecation and diarrhoea in rats. *Eur. J. Pharmacol.* **384**, 37–42 (1999).

69. Pacher, P., Bátkai, S. & Kunos, G. The endocannabinoid system as an emerging target of pharmacotherapy. *Pharmacol. Rev.* **58**, 389–462 (2006).

70. Schultheiß, T., Flau, K., Kathmann, M., Göthert, M. & Schlicker, E. Cannabinoid CB1 receptor-mediated inhibition of noradrenaline release in guinea-pig vessels, but not in rat and mouse aorta. *Naunyn. Schmiedebergs. Arch. Pharmacol.* **372**, 139–146 (2005).

71. Merlob, Paul and Stahl, Bracha and Klinger, G. For debate: does cannabis use by the pregnant mother affect the fetus and newborn? *Pediatr. Endocrinol. Rev.* **15**, 4–7 (2017).

72. Metz, T. D., Allshouse, A. A., Hogue, C. J., Goldenberg, R. L., Dudley, D. J., Varner, M. W., Conway, D. L., Saade, G. R. & Silver, R. M. Maternal marijuana use, adverse pregnancy outcomes, and neonatal morbidity. *Am. J. Obstet. Gynecol.* **217**, 478.e1-478.e8 (2017).

73. Yang, Y., Kanno, C., Sakaguchi, K., Katagiri, S., Yanagawa, Y. & Nagano, M. Theca cells can support bovine oocyte growth in vitro without the addition of steroid hormones. *Theriogenology* **142**, 41–47 (2020).

74. Tam, J., Hinden, L., Drori, A., Udi, S., Azar, S. & Baraghithy, S. The therapeutic potential of targeting the peripheral endocannabinoid/CB1 receptor system. *Eur. J. Intern. Med.* **49**, 23–29 (2018).

75. Bensaid, M., Gary-Bobo, M., Esclangon, A., Maffrand, J. P., Le Fur, G., Oury-donat, F. & Soubrié, P. The cannabinoid CB1 receptor antagonist SR141716 increases Acrp30 mRNA expression in adipose tissue of obese fa/fa rats and in cultured adipocyte cells. *Mol. Pharmacol.* **63**, 908–914 (2003).

76. Jourdan, T., Djaouti, L., Demizieux, L., Gresti, J., Vergès, B. & Degrace, P. CB1

antagonism exerts specific molecular effects on visceral and subcutaneous fat and reverses liver steatosis in diet-induced obese mice. *Diabetes* **59**, 926–934 (2010).

77. Nakata, M. & Yada, T. Cannabinoids inhibit insulin secretion and cytosolic Ca2+ oscillation in islet β-cells via CB1 receptors. *Regul. Pept.* **145**, 49–53 (2008).

78. Hsu, C. Y., McCulloch, C. E., Iribarren, C., Darbinian, J. & Go, A. S. Body mass index and risk for end-stage renal disease. *Ann. Intern. Med.* **144**, 21–28 (2006).

79. Coimbra, T. M., Janssen, U., Gröne, H. J., Ostendorf, T., Kunter, U., Schmidt, H., Brabant, G. & Floege, J. Early events leading to renal injury in obese Zucker (fatty) rats with type II diabetes. *Kidney Int.* **57**, 167–182 (2000).

80. Janiak, P., Poirier, B., Bidouard, J. P., Cadrouvele, C., Pierre, F., Gouraud, L., Barbosa, I., Dedio, J., Maffrand, J. P., Le Fur, G., O'Connor, S. & Herbert, J. M. Blockade of cannabinoid CB1 receptors improves renal function, metabolic profile, and increased survival of obese Zucker rats. *Kidney Int.* **72**, 1345–1357 (2007).

81. Jenkin, K. A., O'Keefe, L., Simcocks, A. C., Grinfeld, E., Mathai, M. L., McAinch, A. J. & Hryciw, D. H. Chronic administration of AM251 improves albuminuria and renal tubular structure in obese rats. *J. Endocrinol.* **225**, 113–124 (2015).

82. Starowicz, K., Nigam, S. & Di Marzo, V. Biochemistry and pharmacology of endovanilloids. *Pharmacol. Ther.* **114**, 13–33 (2007).

83. Brotchie, J. M. CB1 cannabinoid receptor signalling in Parkinson's disease. *Curr. Opin. Pharmacol.* **3**, 54–61 (2003).

84. Siegling, A., Hofmann, H. A., Denzer, D., Mauler, F. & De Vry, J. Cannabinoid CB1 receptor upregulation in a rat model of chronic neuropathic pain. *Eur. J. Pharmacol.* **415**, R5 (2001).

85. Ramírez-Orozco, R. E., García-Ruiz, R., Morales, P., Villalón, C. M., Villafán-Bernal, J. R. & Marichal-Cancino, B. A. Potential metabolic and behavioural roles of the putative endocannabinoid receptors GPR18, GPR55 and GPR119 in feeding. *Curr. Neuropharmacol.* **17**, 947–960 (2019).

86. Cota, D. CB1 receptors: emerging evidence for central and peripheral mechanisms that regulate energy balance, metabolism, and cardiovascular health. *Diabetes. Metab. Res. Rev.* **23**, 507–517 (2007).

87. Dariš, B., Verboten, M. T., Knez, Ž. & Ferk, P. Cannabinoids in cancer treatment: Therapeutic potential and legislation. *Bosn. J. Basic Med. Sci.* **19**, 14 (2019).

88. Elmore, J. S. & Baumann, M. H. Repeated exposure to the "spice" cannabinoid JWH-018 induces tolerance and enhances responsiveness to 5-HT1A receptor stimulation in male rats. *Front. psychiatry* 55 (2018).

89. Caffarel, M. M., Andradas, C., Mira, E., Pérez-Gómez, E., Cerutti, C., Moreno-Bueno, G., Flores, J. M., García-Real, I., Palacios, J., Mañes, S., Guzmán, M. & Sánchez, C. Cannabinoids reduce ErbB2-driven breast cancer progression through Akt inhibition. *Mol. Cancer* **9**, 1–11 (2010).

90. Pi-sunyer, F. X., Aronne, L. J., Devin, J. & Rosenstock, J. Effect of Rimonabant , a Cannabinoid-1. *Jama* **295**, 761–776 (2007).

91. Croci, T. & Zarini, E. Effect of the cannabinoid CB 1 receptor antagonist rimonabant on nociceptive responses and adjuvant-induced arthritis in obese and lean rats. *Br. J. Pharmacol.* **150**, 559–566 (2007).

92. Erdozain, A. M., Diez-Alarcia, R., Meana, J. J. & Callado, L. F. The inverse agonist effect of rimonabant on G protein activation is not mediated by the cannabinoid CB1 receptor: Evidence from postmortem human brain. *Biochem. Pharmacol.* **83**, 260–268 (2012).

93. Sam, A. H., Salem, V. & Ghatei, M. A. Rimonabant: From RIO to Ban. *J. Obes.* **2011**, (2011).

94. NIH. *Rimonabant in Obesity Over a 2-Year Duration (RIO-Europe).* https://clinicaltrials.gov/ct2/show/NCT00386061 (2004).

95. Clinical Trials Arena. Acomplia (Rimonabant) – Investigational Agent for the Management of Obesity. (2021).

96. Després, J. P., Ross, R., Boka, G., Alméras, N. & Lemieux, I. Effect of rimonabant on the

high-triglyceride/low-HDL-cholesterol dyslipidemia, intraabdominal adiposity, and liver fat the ADAGIO-lipids trial. *Arterioscler. Thromb. Vasc. Biol.* **29**, 416–423 (2009).

97. Egan, A. *FDA - Clinical Review of NDA 21888 (rimonabant)*. https://www.fda.gov/media/78285/download (2007).

98. Bifulco, M. & Pisanti, S. End of the line for cannabinoid receptor 1 as an anti-obesity target? *Nat. Rev. Drug Discov.* **7**, 961–962 (2008).

99. Booth, W. T., Walker, N. B., Lowther, W. T. & Howlett, A. C. Cannabinoid receptor interacting protein 1a (CRIP1a): Function and structure. *Molecules* **24**, 1–9 (2019).

100. Kooistra, A. J., Mordalski, S., Pándy-Szekeres, G., Esguerra, M., Mamyrbekov, A., Munk, C., Keserű, G. M. & Gloriam, D. E. GPCRdb in 2021: Integrating GPCR sequence, structure and function. *Nucleic Acids Res.* **49**, D335–D343 (2021).

101. Munk, C., Mutt, E., Isberg, V., Nikolajsen, L. F., Bibbe, J. M., Flock, T., Hanson, M. A., Stevens, R. C., Deupi, X. & Gloriam, D. E. An online resource for GPCR structure determination and analysis. *Nat. Methods* **16**, 151–162 (2019).

102. Kobe, B., Guncar, G., Buchholz, R., Huber, T., Maco, B., Cowieson, N., Martin, J. L., Marfori, M. & Forwood, J. K. Crystallography and protein--protein interactions: biological interfaces and crystal contacts. *Biochem. Soc. Trans.* **36**, 1438–1441 (2008).

103. Luo, J., Liu, Z., Guo, Y. & Li, M. A structural dissection of large protein-protein crystal packing contacts. *Sci. Rep.* **5**, 1–13 (2015).

104. Bordner, A. J. & Gorin, A. A. Comprehensive inventory of protein complexes in the Protein Data Bank from consistent classification of interfaces. *BMC Bioinformatics* **9**, 1–11 (2008).

105. Singh, P., Ganjiwale, A., Howlett, A. C. & Cowsik, S. M. In silico interaction analysis of cannabinoid receptor interacting protein 1b (CRIP1b) – CB1 cannabinoid receptor. *J. Mol. Graph. Model.* **77**, 311–321 (2017).

106. Jakowiecki, J., Abel, R., Orzeł, U., Pasznik, P., Preissner, R. & Filipek, S. Allosteric modulation of the cb1 cannabinoid receptor by cannabidiol—a molecular modeling study

of the n-terminal domain and the allosteric-orthosteric coupling. *Molecules* **26**, 2456 (2021).

107. Chung, H., Fierro, A. & David Pessoa-Mahana, C. Cannabidiol binding and negative allosteric modulation at the cannabinoid type 1 receptor in the presence of delta-9tetrahydrocannabinol: An in Silico study. *Public Libr. Sci. San Fr. CA USA* **14**, e0220025 (2019).

108. Laprairie, R. B., Bagher, A. M., Kelly, M. E. M. & Denovan-Wright, E. M. Cannabidiol is a negative allosteric modulator of the cannabinoid CB1 receptor. *Br. J. Pharmacol.* **172**, 4790–4805 (2015).

109. Hua, T. Vemuri, K., Pu, M., Qu, L., Han, G-W., Wu, Y., Zhao, S., Shui, W., Li, S., Korde, A., Laprairie, R. B., Stahl, E. L., Ho, J. H., Zvonok, N., Zhou, H., Kufareva, I., Wu, B., Zhao, Q., Hanson, M. A., Bohn, L. M., Makriyannis, A., Stevens, R. C., Liu, Z. J., Crystal Structure of the Human Cannabinoid Receptor CB1. *Cell* **167**, 750-762.e14 (2016).

110. Shao, Z., Yin, J., Chapman, K., Grzemska, M., Clark, L., Wang, J. & Rosenbaum, D. M. High-resolution crystal structure of the human CB1 cannabinoid receptor. *Nature* **540**, 602–606 (2016).

111. Hua, T., Vemuri, K., Nikas, S. P., Laprairie, R. B., Wu, Y., Qu, L., Pu, M., Korde, A., Jiang, S., Ho, J. H., Han, G. W., Ding, K., Li, X., Liu, H., Hanson, M. A., Zhao, S., Bohn, L. M., Makriyannis, A., Stevens, R. C., Liu, Z. J., Crystal structures of agonist-bound human cannabinoid receptor CB 1. *Nature* **547**, 468–471 (2017).

112. Li, X., Hua, T., Vemuri, K., Ho, J. H., Wu, Y., Wu, L., Popov, P., Benchama, O., Zvonok, N., Locke, K., Qu, L., Han, G. W., Iyer, M. R., Cinar, R., Coffey, N. J., Wang, J., Wu, M., Katritch, V., Zhao, S., Kunos, G., Bohn, L. M., Makriyannis, A., Stevens, R. C., Liu, Z. J., Crystal Structure of the Human Cannabinoid Receptor CB2. *Cell* **176**, 459-467.e13 (2019).

113. Fay, J. F. & Farrens, D. L. N-terminus can allosterically modulate ligand affinity. *Biochemistry* **52**, 8286–8294 (2014).

114. Krishna Kumar, K., Shalev-Benami, M., Robertson, M. J., Hu, H., Banister, S. D.,

Hollingsworth, S. A., Latorraca, N. R., Kato, H. E., Hilger, D., Maeda, S., Weis, W. I., Farrens, D. L., Dror, R. O., Malhotra, S. V., Kobilka, B. K. & Skiniotis, G. Structure of a Signaling Cannabinoid Receptor 1-G Protein Complex. *Cell* **176**, 448-458.e12 (2019).

115. Sabatucci, A., Tortolani, D., Dainese, E. & Maccarrone, M. In silico mapping of allosteric ligand binding sites in type-1 cannabinoid receptor. *Biotechnol. Appl. Biochem.* **65**, 21–28 (2018).

116. Loo, J. S. E., Emtage, A. L., Murali, L., Lee, S. S., Kueh, A. L. W. & Alexander, S. P. H. Ligand discrimination during virtual screening of the CB1 cannabinoid receptor crystal structures following cross-docking and microsecond molecular dynamics simulations. *RSC Adv.* **9**, 15949–15956 (2019).

117. Shahbazi, F., Grandi, V., Banerjee, A. & Trant, J. F. Cannabinoids and Cannabinoid Receptors: The Story so Far. *iScience* **23**, 101301 (2020).

118. Laprairie, R. B., Vemuri, K., Stahl, E. L., Korde, A., Ho, J.-H., Grim, T. W., Hua, T., Wu, Y., Stevens, R. C., Liu, Z.-J. & others. Probing the CB1 cannabinoid receptor binding pocket with AM6538, a high-affinity irreversible antagonist. *Mol. Pharmacol.* **96**, 619–628 (2019).

119. J Kuder, K. & Kiec-Kononowicz, K. Fluorescent GPCR ligands as new tools in pharmacology-update, years 2008-early 2014. *Curr. Med. Chem.* **21**, 3962–3975 (2014).

120. Casteels, C., Bormans, G. & Van Laere, K. Brain Imaging of Cannabinoid Receptors. *Imaging Hum. Brain Heal. Dis.* 37–79 (2014).

121. Gatley, S. J., Lan, R., Pyatt, B., Gifford, A. N., Volkow, N. D. & Makriyannis, A. Binding of the non-classical cannabinoid CP 55,940, and the diarylpyrazole AM251 to rodent brain cannabinoid receptors. *Life Sci.* **61**, PL191--PL197 (1997).

122. Seely, K. A., Brents, L. K., Franks, L. N., Rajasekaran, M., Zimmerman, S. M., Fantegrossi, W. E. & Prather, P. L. AM-251 and rimonabant act as direct antagonists at mu-opioid receptors: implications for opioid/cannabinoid interaction studies. *Neuropharmacology* **63**, 905–915 (2012).

123. Lan, R., Liu, Q., Fan, P., Lin, S., Fernando, S. R., McCallion, D., Pertwee, R. &

Makriyannis, A. Structure- activity relationships of pyrazole derivatives as cannabinoid receptor antagonists. *J. Med. Chem.* **42**, 769–776 (1999).

124. Chen, W., Tang, H., Liu, H., Long, L., Gong, Z., Zheng, J., Chi, M., Xie, Y., Zheng, Z., Li, S. & Wang, L. Novel selective antagonist of the cannabinoid CB1 receptor, MJ15, with prominent anti-obesity effect in rodent models. *Eur. J. Pharmacol.* **637**, 178–185 (2010).

125. Shao, Z., Yan, W., Chapman, K., Ramesh, K., Ferrell, A. J., Yin, J., Wang, X., Xu, Q. & Rosenbaum, D. M. Structure of an allosteric modulator bound to the CB1 cannabinoid receptor. *Nat. Chem. Biol.* **15**, 1199–1205 (2019).

126. Wang, X., Liu, D., Shen, L., Li, F., Li, Y., Yang, L., Xu, T., Tao, H., Yao, D., Wu, L. & others. A genetically encoded F-19 NMR probe reveals the allosteric modulation mechanism of cannabinoid receptor 1. *J. Am. Chem. Soc.* **143**, 16320–16325 (2021).

127. Yang, X., Wang, X., Xu, Z., Wu, C., Zhou, Y., Wang, Y., Lin, G., Li, K., Wu, M., Xia, A. & others. Molecular mechanism of allosteric modulation for the cannabinoid receptor CB1. *Nat. Chem. Biol.* 1–10 (2022).

128. Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O. & Abola, E. E. Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr. Sect. D* **54**, 1078–1084 (1998).

129. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, F. TE. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

130. Helgesen, Ø., Glavee-Geo, R., Mustafa, G., Nesset, E. & Rice, P. *Modeller*. *Modeller* (2019). doi:10.18261/9788215034393-2019.

131. Yang, J. & Zhang, Y. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* **43**, W174–W181 (2015).

132. Oleg, T. & Arthur J., O. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).

133. Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S. & Olson, A. J. Computational protein--ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **11**, 905–919 (2016).

134. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. *J. Comput. Chem.* **30**, 2785–2791 (2009).

135. Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A. & others. PubChem substance and compound databases. *Nucleic Acids Res.* **44**, D1202--D1213 (2016).

136. Yau, M. Q., Emtage, A. L., Chan, N. J. Y., Doughty, S. W. & Loo, J. S. E. Evaluating the performance of MM/PBSA for binding affinity prediction using class A GPCR crystal structures. *J. Comput. Aided. Mol. Des.* **33**, 487–496 (2019).

137. DeLano, W. PyMOL. *DeLano Scientific, San Carlos, CA* 1–15 at http://virology.wisc.edu/acp/Classes/DropFolders/Drop660_lectures/2013_660/L01_PyMOL_2013r.pdf (2002).

138. S. Jo, T. Kim, V.G. Iyer, and W. I. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).

139. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, 370–376 (2012).

140. Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I. & others. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).

141. Phillips, J. C., Schulten, K., Bhatele, A., Mei, C., Sun, Y., Bohm, E. J. & Kale, L. V. Scalable molecular dynamics with NAMD. *Parallel Sci. Eng. Appl. Charm++ Approach* **26**, 60–76 (2016).

142. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and

analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3085–3095 (2013).

143. DA, C., Ben-Shalom, S. R., Brozell, D. S., Cerutti III, T. E., Cheatham III, V. W. D., Cruzeiro, T. A., Darden, R. E., Duke, D. G., Gilson, M. K., Gohlke, H. & others. AMBER 18. *University of California, San Francisco* at (2018).

144. Hou, T., Wang, J., Li, Y. & Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **51**, 69–82 (2011).

145. Jung, S. W., Cho, A. E. & Yu, W. Exploring the Ligand Efficacy of Cannabinoid Receptor 1 (CB1) using Molecular Dynamics Simulations. *Sci. Rep.* **8**, 1–11 (2018).

146. Liu, H. & Hou, T. CaFE: A tool for binding affinity prediction using end-point free energy methods. *Bioinformatics* **32**, 2216–2218 (2016).

147. Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G. W., Holst, M. J., McCammon, J. A., Baker, N. A., Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).

148. Altis, A., Nguyen, P. H., Hegger, R. & Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Chem. Phys.* **126**, (2007).

149. Majumder, S. & Giri, K. An insight into the binding mechanism of Viprinin and its morpholine and piperidine derivatives with HIV-1 Vpr: molecular dynamics simulation, principal component analysis and binding free energy calculation study. *J. Biomol. Struct. Dyn.* **0**, 1–13 (2021).

150. Grant, Barry J and Rodrigues, Ana PC and ElSawy, Karim M and McCammon, J Andrew and Caves, L. S. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **22**, 2695–2696 (2006).

151. Shalizi, C. R. Advanced data analysis from from an Elementary Point of View. 1–583 (2013).

152. McCammon, J. A. Protein dynamics. *Reports Prog. Phys.* **47**, 1–46 (1984).

153. Kasahara, K., Fukuda, I. & Nakamura, H. A novel approach of dynamic cross correlation analysis on molecular dynamics simulations and its application to Ets1 dimer-DNA complex. *PLoS One* **9**, (2014).

154. Peng, J. H., Wang, W., Yu, Y. Q., Gu, H. L. & Huang, X. Clustering algorithms to analyze molecular dynamics simulation trajectories for complex chemical and biological systems. *Chinese J. Chem. Phys.* **31**, 404–420 (2018).

155. Davies, David L.; Bouldin, D. W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-1 (2)**, 224–227 (1979).

156. Cheatham, T. E. AMBER-Hub. https://amberhub.chpc.utah.edu/cluster/ (2021).

157. Altman, N. & Krzywinski, M. Points of Significance: Simple linear regression. *Nat. Methods* **12**, 999–1000 (2015).

158. Ahn, K. H., Nishiyama, A., Mierke, D. F. & Kendall, D. A. Hydrophobic residues in helix 8 of cannabinoid receptor 1 are critical for structural and functional Properties. *Biochemistry* **49**, 502–511 (2010).

159. Stadel, R., Ahn, K. H. & Kendall, D. A. The cannabinoid type-1 receptor carboxyl-terminus, more than just a tail. *J. Neurochem.* **117**, 1–18 (2011).

160. Dym, O., Eisenberg, D. & Yeates, T. O. ERRAT. at (2012).

161. Colovos, C. & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–1519 (1993).

162. Bowie, J. U., Lüthy, R. & Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science (80-.).* **253**, 164–170 (1991).

163. Dym, O., Eisenberg, D. & Yeates, T. O. VERIFY3D. at (2006).

164. Messaoudi, A., Belguith, H. & Ben Hamida, J. Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β-lactamase. *Theor. Biol. Med. Model.* **10**, 1–10 (2013).

165. Humphrey, W., Dalke, A. and Schulten, K. VMD - Visual Molecular Dynamics. *J. Molec. Graph.* **14**, 33–38 (1996).

166. Booth, W. T., Clodfelter, J. E., Leone-Kabler, S., Hughes, E. K., Eldeeb, K., Howlett, A. C. & Lowther, W. T. Cannabinoid receptor interacting protein 1a interacts with myristoylated Gαi N terminus via a unique gapped β-barrel structure. *J. Biol. Chem.* **297**, 101099 (2021).

167. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, 526–531 (2004).

168. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. OPM: Orientations of proteins in membranes database. *Bioinformatics* **22**, 623–625 (2006).

169. J. Lee, M. Hitzenberger, M. Rieger, N.R. Kern, M. Zacharias, and W. I. CHARMM-GUI supports the Amber force fields. *J. Chem. Phys.* **153**, 035103 (2020).

170. Lily Wang, Irfan Alibay, Rocco Meli, Paul Smith, Mieczyslaw Torchala, Yuxuan Zhuang, Richard J. Gowers, O. B. MDAnalysis. 1 at https://userguide.mdanalysis.org/2.0.0-dev0/examples/analysis/alignment_and_rms/rmsd.html (2019).

171. De Vries, S. J., Van Dijk, M. & Bonvin, A. M. J. J. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883–897 (2010).

172. Miller, B. R., McGee, T. D., Swails, J. M., Homeyer, N., Gohlke, H. & Roitberg, A. E. MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* (2012) doi:10.1021/ct300418h.

173. Case, D. A., Aktulga, H. M., Belfon, K., Ben-Shalom, I., Brozell, S. R., Cerutti, D., Cheatham, T., Cruzeiro, V., Wilian D., Darden, T., Duke., R. E., Amber 2021. at (2021).

174. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, 665–667 (2004).

175. Shen, M. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524 (2006).

176. Laskowski, Roman A., Malcolm W. MacArthur, David S. Moss, and J. M. T. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, (1993).

177. Dunbrack, R. L. Whatcheck. *Dict. Bioinforma. Comput. Biol.* (2004).

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

179. Shim, J.-Y., Ahn, K. H. & Kendall, D. A. Molecular basis of cannabinoid CB1 receptor coupling to the G protein heterotrimer Gαiβγ. *J. Biol. Chem.* **288**, 32449–32465 (2013).

180. Fayos, R., Melacini, G., Newlon, M. G., Burns, L., Scott, J. D. & Jennings, P. A. Induction of flexibility through protein-protein interactions. *J. Biol. Chem.* **278**, 18581–18587 (2003).

181. Hsieh, C., Brown, S., Derleth, C. & Mackie, K. Internalization and recycling of the CB1 cannabinoid receptor. *J. Neurochem.* **73**, 493–501 (1999).

182. Moreira, F. A. & Crippa, J. A. S. The psychiatric side-effects of rimonabant. *Rev. Bras. Psiquiatr.* **31**, 145–153 (2009).

183. Kenakin, T. & Strachan, R. T. PAM-antagonists: a better way to block pathological receptor signaling? *Trends Pharmacol. Sci.* **39**, 748–765 (2018).

184. Begg, M., Pacher, P., Bátkai, S., Osei-Hyiaman, D., Offertáler, L., Fong, M. M., Liu, J. & Kunos, G. Evidence for novel cannabinoid receptors. *Pharmacol. Ther.* **106**, 133–145 (2005).

185. Jakowiecki, J., Orzeł, U., Chawananon, S., Miszta, P. & Filipek, S. The hydrophobic ligands entry and exit from the GPCR binding site-SMD and SUMD simulations. *Molecules* **25**, (2020).

186. Chung, H., Fierro, A. & Pessoa-Mahana, C. D. Cannabidiol binding and negative allosteric modulation at the cannabinoid type 1 receptor in the presence of delta-9-tetrahydrocannabinol: An In Silico study. *PLoS One* **14**, e0220025 (2019).

APPENDIX

Table 5.1: Docking results for structure AM251, AM6538 and MJ15. Mode 7,3 and 19 were picked as the structures for inverse agonist AM251, AM6538 and MJ15, respectively. Picked structures, were filtered based on a structural constraint selection. Docked complexes were compared through visual inspection to the crystallographic structure. Indications of closeness in rotation and distance to the three arms and pyrazole core between the docked structure and the crystallographic structure AM6538 (PDB code: ZDG), were used as the structural constraint.

System	**AM251**	**AM6538**	**MJ15**
Mode	Affinity (kcal/mol)	Affinity (kcal/mol)	Affinity (kcal/mol)
1	-9.6	-9.5	-10.8
2	-8.9	-8.5	-10.6
3	-8.9	-8.3	-10.3
4	-8.7	-7.8	-10.1
5	-8.6	-7.1	-10.0
6	-8.3	-7.1	-9.3
7	-8.2	-6.9	-9.3
8	-8.0	-6.9	-9.2
9	-7.0	-6.9	-9.1
10	-6.9	-6.9	-9.1
11	-6.8	-6.8	-9.0
12	-6.8	-6.8	-9.0
13	-6.7	-6.7	-8.9
14	-6.7	-6.6	-8.6
15	-6.7	-6.6	-8.5
16	-6.6	-6.5	-8.2
17	-6.6	ND	-8.2
18	ND	ND	-8.1
19	ND	ND	-7.9

ND – Not Determined