

Characterizing the binding region of GPCRs based on molecular simulation and energy decomposition

by

Siyi Wang

Bachelor of Science, Shanxi Medical University, 2018

Submitted to the Graduate Faculty of
School of Pharmacy in partial fulfillment
of the requirement for the degree of
Master of Science

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF PHARMACY

This thesis/dissertation was presented

by

Siyi Wang

It was defended on

March 18, 2020

And approved by

Xiang-Qun (Sean) Xie, PhD, EMBA, Professor, Department of Pharmaceutical Sciences

Levent Kirisci, PhD, Professor, Department of Pharmaceutical Sciences

Junmei Wang, PhD, Associate Professor, Department of Pharmaceutical Sciences

Zhiwei Feng, PhD, Assistant Professor, Department of Pharmaceutical Sciences

Thesis Advisor/Dissertation Director:

Zhiwei Feng, PhD, Assistant Professor, Department of Pharmaceutical Sciences

Copyright © by Siyi Wang

2020

Characterizing the binding region of GPCRs based on molecular simulation and energy decomposition

Siyi Wang, B.S

University of Pittsburgh, 2020

ABSTRACT

G protein coupled receptors (GPCRs) are one of the largest and most important families of membrane proteins in humans. The binding of GPCRs with their orthosteric ligands activates the internal signal transduction pathways and sets off a cascade of reactions to regulate biological processes. 30-40% drugs currently on the market target the orthosteric binding pocket of GPCRs, however, increasing attention has been devoted to the allosteric binding pockets for its unique advantages, such as high selectivity and less side effects. Comprehensive characterization of protein is invaluable to infer its evolutionary processes and biological functions to achieve the desired substrate specificity and to design a drug with coveted selectivity. However, protein is complicated by its different functional regions or domains that can bind to some of its protein partner(s), substrate(s), orthosteric ligand(s), or allosteric modulator(s). Unlike a small molecule that can be easily characterized by its fingerprints, there are no cogent methods to comprehensively characterize the features of an entire protein or its substructure. In the present work, a scoring function-based computing protocol Molecular Complex Characterizing System (MCCS) was applied to help characterize the GPCRs. I first quantitatively calculate the energy contribution of each individual residue based on the receptor-ligand/modulator interactions, and the quantitated energy contribution of the residues involved in the binding pocket was used to represent a pattern

of the ligand recognition of a receptor. It was found that residue energy contribution can not only identify the residues that contribute commonly to the binding of agonist and antagonist, but also distinguish the selective ones for either agonist or antagonist. Then, I applied multiple molecular simulation method to predict the CB1/CB2 allosteric binding mode and study the structure-activity relationship (SAR) of CB1/CB2 modulators. The present work comprehensively characterizes the interaction between GPCR with their ligands based on a new approach, which may aid in the facilitation of rational drug development.

Key words: Molecular Simulation, molecular docking, molecular dynamic simulation, GPCR, cannabinoid receptor

Table of Contents

PREFACE.....	xiii
1.0 INTRODUCTION.....	1
1.1 G protein-coupled receptors.....	1
1.2 Traditional GPCRs' orthosteric binding pocket.....	5
1.3 Allosteric modulation of GPCRs.....	6
1.4 Cannabinoid receptors.....	8
1.5 Generic GPCR residue numbering scheme	9
1.6 MCCS: a scoring-function based characterization protocol for protein	9
2.0 METHODS	12
2.1 Protein-ligand complexes.....	12
2.2 Docking.....	12
2.2.1 MCCS	12
2.2.2 Glide docking	20
2.3 Similarity and clustering matching	22
2.4 Molecular Dynamics- (MD) Simulation	23
2.5 CB1/CB2 pocket prediction and analysis.....	24
2.6 Statistics.....	25
3.0 RESULTS	26
3.1 Characterization of traditional orthosteric binding pocket of GPCRs.....	26
3.1.1 Receptor-Ligand Dataset of 249 Crystallized GPCRs	26
3.1.2 Ligand-binding pocket recognition patterns of GPCRs	31

3.1.3 The key residues of class A GPCRs orthosteric binding pocket	35
3.1.4 Protein Similarity and Clustering	41
3.2 Characterization of the allosteric binding region of class A GPCRs	43
3.2.1 Reported Allosteric Binding Pockets of Class A GPCRs.....	43
3.2.2 Characterization of three typical allosteric binding region of class A GPCRs.....	47
3.3 Cannabinoid receptor 2	52
3.3.1 CB2 orthosteric binding pattern	52
3.3.2 CB1/CB2 allosteric binding mode predictions and validations.....	54
3.3.2 The most possible binding pocket for CB2.....	59
3.3.3 Validate the CB1 allosteric binding prediction using Molecular dynamic simulations.....	67
3.3.4 Exploration of the CB2 allosteric regulation using MD simulation.	75
4.0 SUMMARY AND CONCLUSIONS	78
APPENDIX.....	80
BIBLIOGRAPHY	83

List of Tables

Table 1 The role of residue at the same position on ECL2 for GPCR-ligand complex.....	39
Table 2 Summarization of all reported allosteric regulation structures.	45
Table 3 Key residues involved in the Site C binding site are conserved	51
Table 4 Parameters of ten potential CB1 binding cavity.	57
Table 5 Parameters of ten potential CB2 binding cavity.	58
Table 6 Allosteric modulators of CB1	60
Table 7 Conservation of W4.50 among class-A GPCRs.....	64

List of Figures

Figure 1 Overview of 7 TM architecture structure	2
Figure 2 Schematic diagram of GPCR signaling transduction	4
Figure 3 Activation ability of four types of agonists and antagonists	6
Figure 4 Workflow for the MCCS analysis	14
Figure 5 The Revised Docking Workflow of idock.	15
Figure 6 The Score Only Workflow of jdock.	16
Figure 7 Variations of each energy term with surface distance	18
Figure 8 Variation of total atomic energy with surface distance.....	19
Figure 9 Overall workflow of glide docking.	21
Figure 10 Different state conformations of ADRB2.	28
Figure 11 Radar map of 249 crystallized GPCRs.....	30
Figure 12 The Pattern of residue energy contribution and its energy decomposition.	32
Figure 13 Representative residues involved in the binding pocket of each class GPCRs.	34
Figure 14 The role of residue 6.48 for distinguishing the inactive status from the active one.	36
Figure 15 Multiple roles of residue 3.32 for ligand recognition.	37
Figure 16 The role of residue at the same position on ECL2 for GPCR-ligand complex.	40
Figure 17 Heatmap demonstrating a correlation between score vectors of GPCRs.	42
Figure 18 The summarization of allosteric binding pockets of class A GPCRs.....	44
Figure 19 Detailed interaction in allosteric binding pocket of ACM2 (Site A).	48

Figure 20 Interaction pattern in allosteric binding Site B.	49
Figure 21 Interaction pattern in allosteric binding Site C.	50
Figure 22 Selectivity pattern between CB1 and CB2 agonists.....	53
Figure 23 All potential binding cavity of CB1 structure.	56
Figure 24 All potential binding cavity of CB2 structure.	56
Figure 25 The predicted allosteric binding pocket of CB1 and CB2.	59
Figure 26 Alignment of crystallized and docked org27569	62
Figure 27 Detail interaction and energy decomposition of PAM and NAM with CB1.	65
Figure 28 Comparison of docking pose and energy decomposition between CB1 and CB2 modulator.....	66
Figure 29 MD simulation of CB1-Org27569 complex.	69
Figure 30 MD simulation for PSNCBAM-1.	70
Figure 31 MD simulation for GAT211.....	71
Figure 32 MD simulation for Lipoxin A4.	72
Figure 33 MD simulation for RTI-371.	72
Figure 34 MD simulation for ZCZ011.	73
Figure 35 MD simulation for Fenofibrate.	73
Figure 36 MD simulation of Cannabidiol.	74
Figure 37 MD simulation of Pregnenolone.....	75
Figure 38 Exploration of the CB2 allosteric regulation using MD simulation.....	77
Figure 39 Correlation between score vectors of GPCRs calculated using Cosine Similarity	80
Figure 40 Detail binding interaction of crystlaized CB1-org2759 complex.....	81

Figure 41 Ligand interaction diagram of PAM and NAM with CB1.	82
---	-----------

List of Equations

Equation 1. Gauss₁	17
Equation 2. Gauss₂	17
Equation 3. Repulsion	17
Equation 4. Hydrophobic	17
Equation 5. Hbonding	17
Equation 6. Atomic total score	19
Equation 7. Residue total score	20
Equation 8. Cosine similarity	22
Equation 9. Pearson correlation coefficient	23

PREFACE

I sincerely appreciate my co-advisor Dr. Zhiwei Feng. He met me twice a week in the past two years and spent a lot of time discussing with me about my ongoing projects. He gave me a lot of detailed and useful suggestions, not only for research but also the life advice. He provided me a lot of opportunities and motivated me to work hard, which benefited me to immerse myself in diverse interesting disciplines and topics. Importantly, he trained me a lot in scientific writing and research design, which cultivate my independent thinking style and it would be important to my future research.

Also, I would like to express my gratefulness to Dr. Xiangqun Xie. As the director of our CCGS center, he gave me much more instructions, supports, and also encouragement during my entire master study.

Meanwhile, I would like to thank my committee member, Dr. Junmei Wang and Dr. Levent Kirisci. Dr. Junmei Wang is an expert on force field developing and molecular simulation. I learned much from him, not only his expertise and knowledge, but also his enthusiasm and attitude toward study and life. Dr. Levent Kirisci is my instructor at *Advanced Statistical* class, he hierarchically construct my statistical knowledge system, which benefit me a lot in my research study.

At the same time, let me express my sincere acknowledgement to all the team members at Dr. Xiangqun Xie lab and Dr. Junmei Wang lab. Especially thanks Maozi Chen, who is a poet of programming, he wrote the *MCCS* program and provide me a lot of technical supports. I'd like to thanks Dr. Lirong Wang, Dr. Terence McGuire, and Dr. Jaden Jun, as professors and team leaders at our *CCGS* center, they instructed me their research knowledge and experience, provide me valuable advice. Also, thanks to Yankang Jing, Mingzhe Shen, Shuhan Liu and all other peers, they support, assist, and accompany me during my entire two-year study.

Lastly, I would like to mention my parents. Definitely, the word "thank you" are far less enough to express my emotion to them at any moment. It is hard to find the words to say how much they mean to me, but I want them know that I'm grateful each and every day.

I offer my warmest regards and sincere blessings to all of those who helped me during the completion of this two-year study in any respect.

1.0 INTRODUCTION

1.1 G protein-coupled receptors

The most important biological process of multicellular organisms is the communications between cells and their extracellular environment. Transmembrane (TM) receptors are receptors that are embedded in the membrane of cells or some subcellular compartments and organelles. They dedicated to cell signaling transduction by recognizing and receiving extracellular molecules. There are three general families of TM receptors: ion channel-linked receptors, enzyme-linked receptors, and G protein-coupled receptors (GPCRs).

GPCRs, also known as 7 transmembrane domain receptors, are the largest superfamily of TM proteins in our body, with more than 800 members[1]. All GPCRs characterized by a common architecture of 7 TM helices along with some highly conserved motifs. GPCRs canonical 7 TM α -helical bundle (helices I–VII) start from N-terminus at the extracellular side, connected by three extracellular loops (ECL 1-3) and three intracellular loops (ICL 1-3), and finally end in C-terminus at the intracellular side[2], as shown in Figure 1.

GPCRs are important for many essential physiological processes: a broad spectrum of extracellular signals need to be transported through GPCRs to intracellular side, including hormones, neurotransmitters, growth and developmental factors, amino-acid residues, nucleotides, peptides and proteins, as well as light, odors, and gustative molecules[3].

The mechanisms behind signaling transduction through GPCRs are shown in Figure 2. The inactivated GPCRs can be activated by binding of an agonist molecules. Afterwards, it will undergo a conformational change which activate coupled heterotrimeric guanine nucleotide-

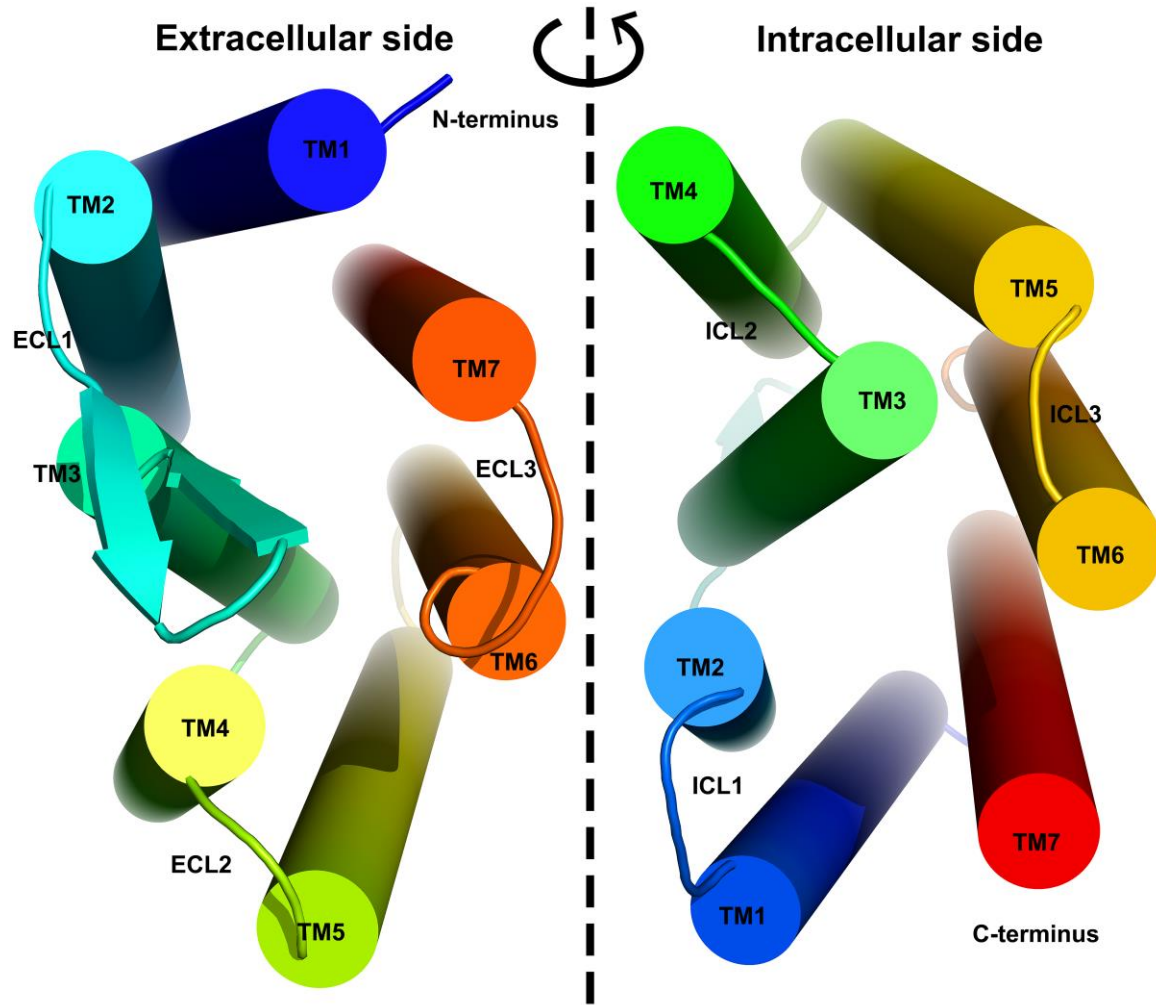


Figure 1 Overview of 7 TM architecture structure

bind proteins (G protein). This interaction will induce the exchange of guanosine diphosphate (GDP)/ guanosine triphosphate (GTP) associated with the $G\alpha$ subunit. As a result, $G\beta\gamma$ dimer dissociate from $G\alpha$ subunit and thus multiple G protein subunits were activated and bound to various effectors. Those effectors in turn switch on or off in different systems and trigger cascades of cellular and physiological responses. Due to the intrinsic GTPase activity of $G\alpha$ subunit, the bound GTP will be gradually hydrolyzed to GDP, followed by the reassociation of $G\beta\gamma$ dimer with $G\alpha$ subunit and the inactivation of G protein [1, 4, 5]. Due to their broadly distributions and multiple functions, GPCRs act as targets of about 30% of marketed drugs[6]. Yet only a limited number of GPCRs are currently targeted by marketed drugs, a plenty of opportunities for targeting to many other GPCRs remained.

Despite the similar 7 TM helical structure, GPCRs are diverse in sequence and function, and based on these similarity and diversity, the human GPCRs are usually classified into 5 major classes: Rhodopsin (class A); Secretin and Adhesion (class B); Glutamate (class C); and Frizzled/taste receptor 2 (class F), these classes can be further divided into subclasses based on the sequence similarity. Among them, class A is the largest superfamily which consists of four major subclasses with 13 sub-branches [1, 7].

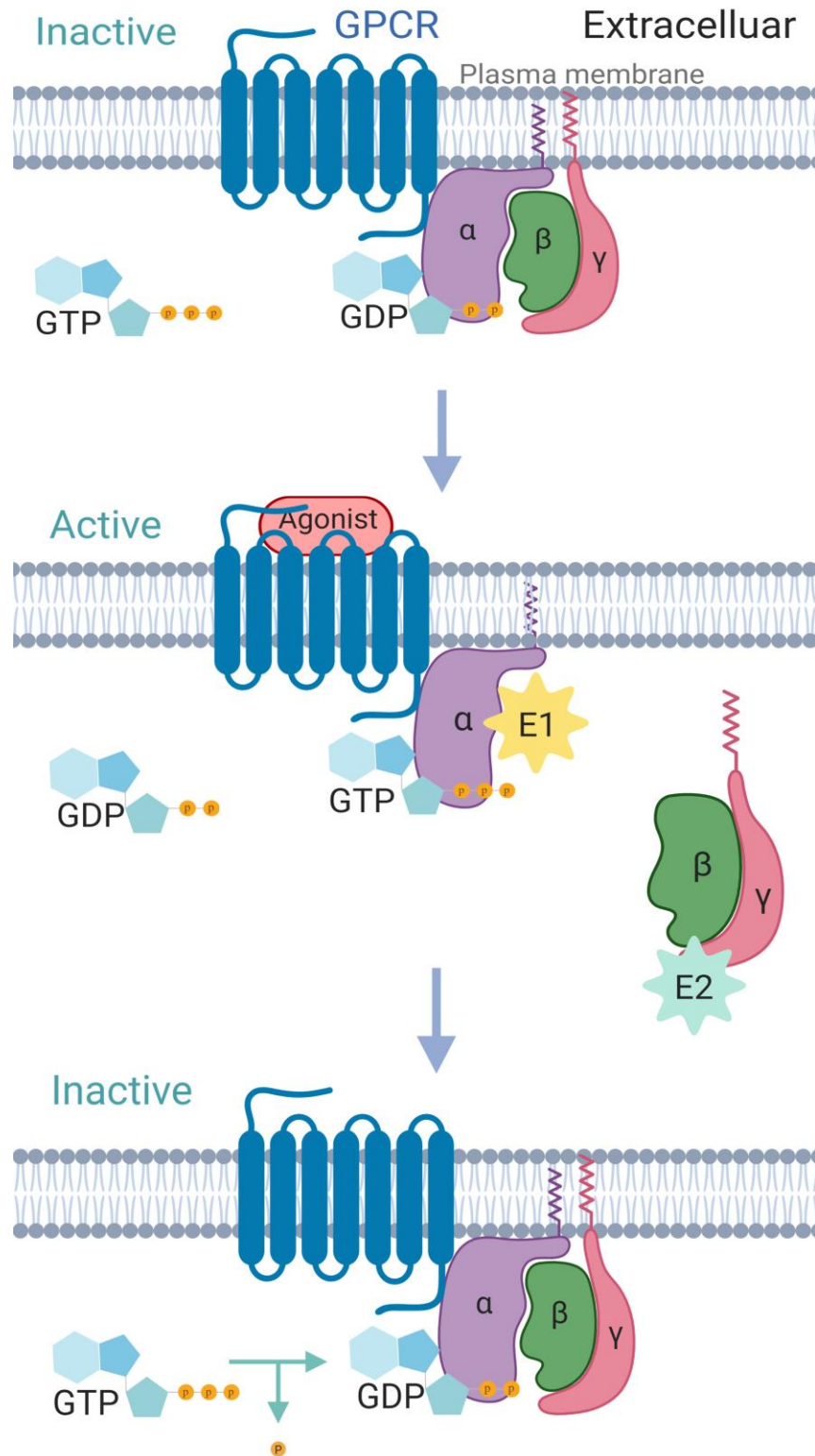


Figure 2 Schematic diagram of GPCR signaling transduction

1.2 Traditional GPCRs' orthosteric binding pocket

The GPCRs structural biology study started with two-dimensional crystallography in 1990s[8], then stepped steps into its renaissance era as the first three-dimensional (3D) structure of rhodopsin has been is crystallized and reported in 2000[9]. With the help of innovation technologies such as X-ray crystallography and transmission electron cryo-microscopy (cryo-EM), more and more 3D structures with high resolution are reported during the past two decades. These structures advanced our understanding on the architecture and actions of GPCRs.

According to the classical models of GPCRs, the states of GPCRs included 'fully inactive', 'partially active', and 'fully active'. Under normal physiological condition, GPCRs enable to present variable degrees of basal activity and induce active downstream signaling even in the absence of ligands. Typically, in order to reach maximal signaling, GPCRs require agonists to stabilize fully active states. As shown in Figure 3, full agonists can maximally activate their cognate GPCRs, while the partial agonists hard to induce 100% activation. Antagonists, on the other hand, are compounds that block GPCRs activation, they are of two types: inverse agonists and neutral antagonists. Inverse agonists can decrease the basal activity of GPCRs, while neutral antagonists act on inhibiting agonist effects exert little influence on basal activity [10].

Both agonists and antagonists targeted traditional orthosteric site, which share the same binding site for endogenous agonists activating the GPCRs. The orthosteric binding site of each class GPCRs locates on distinct position. Under the highly pressure of evolution, though GPCRs from same class bind with ligands on a conserved position, their binding pockets still vary greatly regarding to sequence involved in the pocket, the shape and size of the pocket. On the other hand, members from the same sub-families, such as serotonin receptors family, possess more conserved orthosteric binding pocket[8].

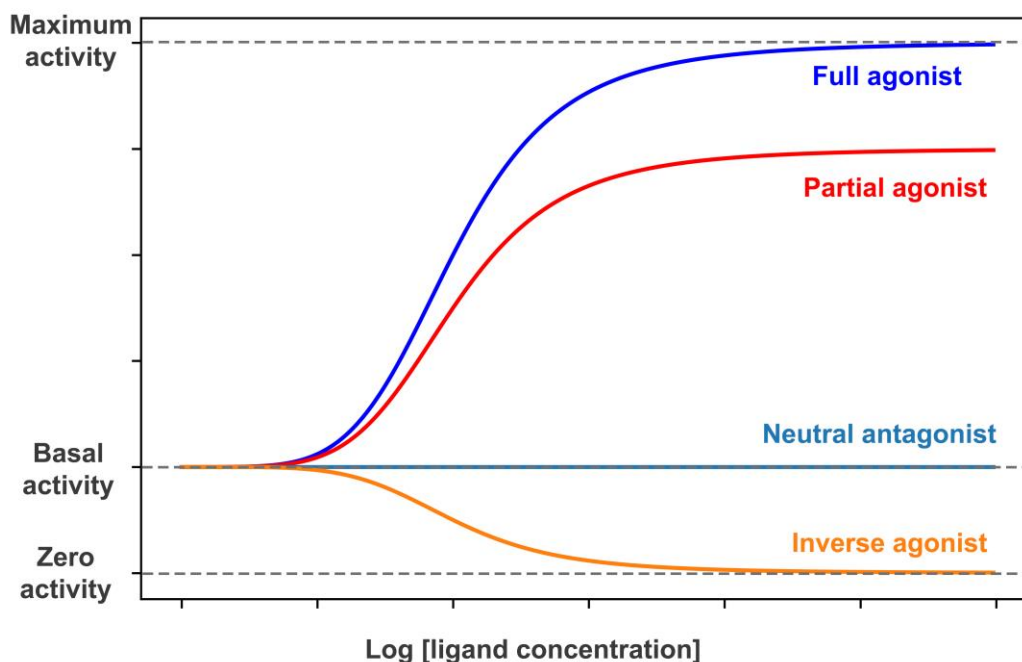


Figure 3 Activation ability of four types of agonists and antagonists

1.3 Allosteric modulation of GPCRs

The most well researched type of binding site for GPCRs is their orthosteric binding sites. Most of the FDA-approved drugs targeting GPCRs by interacting with the orthosteric binding site. However, this situation might be changed as more and more attention has been devoted to the study of the allosteric binding site of GPCRs.

Allosteric regulation is the interaction of receptors with modulators at a site other than the orthosteric site of the protein. The binding between modulators at allosteric site often results in a conformational change, which may affect the binding affinity of the ligands to the orthosteric site of the protein in one of the three following way: (1) positive allosteric modulators (PAMs)

potentiate the response of protein to ligands; (2) negative allosteric modulators (NAMs) decrease the ligand-mediated protein response; (3) neutral allosteric ligands (NALs) occupy the allosteric site while inducing no functional effect, which can block the action of PAMs or NAMs. Unlike the orthosteric site, which showed more evolutionary conservation, the allosteric binding pocket of GPCRs showed greater divergence across subtypes of receptors from given family based on amino acid sequence and pocket position. This mechanism allows the potential of designing small molecules with ‘absolute subtype selectivity’. An additional advantage of allosteric regulation is that the modulators can only exert effects in the presence of the endogenous agonist. Thus, the allosteric modulators showed both spatial and temporal selectivity. The third advantage of allosteric modulators is that they lowered the potential toxicity due to the ‘ceiling’ effect. Because the allosteric modulators exert their effect depend on the endogenous orthosteric agonist, the receptor activation won’t increase when reach to a certain ceiling even with an extremely high concentration of potentiator[11].

Despite the boom of GPCR structure biology over the past two decades, the general trend of the researchers highlights the importance of orthosteric modulators for rational drug design in recent years, thus limited GPCR receptor allosteric regulation pattern has been revealed. To date, there are 11 PDB files available from 10 individual class A GPCRs revealing the interaction pattern between the modulator and the receptors in the allosteric binding site. The majority of the class-A GPCRs receptors lacked testified receptor-modulator interaction, and this opaque understanding limits the potential of GPCR structure-based rational drug design.

1.4 Cannabinoid receptors

Cannabinoid receptors, which belong to the class A GPCRs, are a part of the endocannabinoid system. Currently, there are two subtypes of cannabinoid receptors: cannabinoid receptor type 1 (CB1) and cannabinoid receptor type 2 (CB2). CB1 is distributed mostly in the central nervous system (CNS). They influence a wide range of essential physiological functions such as nociception, motor coordination, and the regulation of appetite, mood, and memory[12]. On the other hand, CB2 locates in peripheral tissues such as the spleen, thymus, mast cells, and blood cells[13, 14]. These receptors are involved in modulatory functions like immune system regulation, cell apoptosis, and migration[15]. Both CB1 and CB2 implicated in a substantial number of clinical disorders and had been shown to promising therapeutic prospects. However, traditional small molecule ligand for CB1 and CB2 only achieve very limited success due to unfavorable side effects. For a therapeutic target, CB2 possesses conspicuous advantages comparing with CB1, in that the CB2 is mainly expressed in peripheral tissues which avoid crippling CNS side effects. Therefore, CB2 highly selective ligands were considered as a promising treatment of a lot of endocannabinoid system-related disorders without severe psychiatric side effects. In addition, the nature of allosteric modulation allows the designing of allosteric modulator become another promising avenue for fighting endocannabinoid system-related disorders.

1.5 Generic GPCR residue numbering scheme

All GPCRs share a core 7TM structure and functioning signal transduction based on similar mechanism. Thanks to the conserved 7 TM scaffold, alignment and comparison of sequences or structures among all GPCRs can be applied by indexing with a generic number.

The best numbering scheme of GPCRs is the Ballesteros-Weinstein numbering (BW numbering) scheme, which created for numbering all class A GPCRs. BW numbering consists of two numbers, the first represent the 1-7 helix, and the second denotes the residue's position relative to the most conserved residue in this helix, which is numbered as 50[16].

As the increasing of 3D structure of other class GPCRs, the numbering scheme of class B, C and F have also been established recently. The class B GPCR Wootten scheme based on the secretin class B1 GPCRs, but the reference residue of each helix is also highly conserved among adhesion class B2 GPCRs[17]. Same as class A and B, there were also class C GPCR Pin numbering scheme[18] and class F GPCR Wang scheme[19] published.

1.6 MCCS: a scoring-function based characterization protocol for protein

It is well-known that proteins are critical for most of the essential physiological processes, and they are the most important drug targets in the human body. However, unlike a small molecule which is relatively simple to characterize, process, store, and compare with molecular fingerprints [20], it is complicated to find a cogent way to comprehensively characterize proteins' features. Thus, the development of such a representation method for proteins is a high priority.

During the biological binding process of a small molecule, the key residues involved in the corresponding binding region play important roles in the recognition of a protein. To mimic this important physiological process, different kinds of *in silico* docking algorithms [21-31] have been developed. Briefly, molecular docking is a computational method used to predict the binding pose and binding affinity of a small molecule within a protein, using the technologies of conformational search and scoring function. The more reliable the binding pose is, the more accurate the docking score will be. However, the importance and contribution of key binding residues were always overlooked during the selection of binding pose. Thus, the information of binding residues should be considered during docking simulations.

Over the years, increasing numbers of 3D protein structures with high resolution have been determined by X-ray crystallography or transmission electron cryo-microscopy (cryo-EM) [32-34]. The X-ray crystal or cryo-EM structures of a protein complexed with a small molecule(s) provide an atomic snapshot of the ligand binding process [35]. Ligand binding always involves the interactions with key binding residues, the changes in the features of the ligand binding site, the interesting rearrangements of the protein structure, and so on [35]. Thus, various important information can be explored quantitatively in the binding regions of a crystal/cryo-EM structure, including the involved key residues, residue energy contribution, energy term, etc.

Inspired by the evaluation of conformations and binding strength of a ligand via docking algorithms, a residue vector-based computing program has been constructed to analyze the individual residue energy contribution and the corresponding energy terms for protein characterization. Essentially, this algorithm is a scoring function-based method and the outcome vector computing results are calculated on basis of: (i) 3D X-ray crystal or cryo-EM structure of a ligand-protein complex (without docking) reported, or (ii) the predicted binding pose of the ligand

within the protein (with docking) if experimental 3D structure is unavailable. For a resolved or modeled protein complexed with diverse small molecules of the same binding region, the energy contribution and energy term of each binding residue may vary from virtually identical to significantly different. Such plasticity helps to distinguish binding modes. The end, it constructed a feature vector consisting of a series of aggregated energy scores for each residue for a specific region of a protein. Due to the diversity of small molecules, there may be multiple score sequences or functions for each binding region. Such sequences or functions may reflect the diversities of key residues and their contributions to the recognition of different kinds of ligands. Although such sequences or functions are neither as simple as molecular fingerprints nor unique for a specific protein region, regularities still may be identified. Therefore, proteins may be characterized to an extent, and the outcome of docking or high-throughput virtual screening may be improved.

2.0 METHODS

2.1 Protein-ligand complexes

The X-ray crystal and cryo-EM structures were collected from Protein Data Bank <https://www.rcsb.org>[36, 37] and GPCRs database <https://www.gpcrdb.org>[38]. Then manual filtering, residue renumbering, and redundant chains/ions removal were performed with the help of our established utility programs.

2.2 Docking

2.2.1 MCCA

The entire MCCA workflow was applied to our collected and prepared protein structures. Figure 4 shows a high-level flowchart on top of classic docking workflow. Initially, each input PDB structure was split into a receptor PDB and a ligand PDB. A sequence of operations was carried out on the split PDBs. For the receptor, Chimera (Version 1.13.1)[39] was first applied to repair the residues with an incomplete side chain. Briefly, Chimera first scanned all the residues in a target protein and reported the residues with incomplete side chains; then Chimera replaced each truncated side chain with a complete sidechain of the same residue type using Dunbrack rotamer library.[40] For the ligand, PROPKA (version 3.1)[41] was used to predict the corresponding pKa values so as to determine whether the tertiary (3°) amide of the small molecule

should be protonated. Then using VEGA,[42] polar hydrogens were added, Vina force fields and Gasteiger charges were applied, and the file format was transformed into PDBQT for both PDBs. Especially for the ligand PDB, torsions were defined by VEGA prior to the transformation into PDBQT. The output PDBQT files and the pKa file of the ligand form the input of the next step in MCCS – jdock.

Three different modes are implemented in jdock - the docking mode (Figure 5), the “score only” mode (Figure 6), and the “scoring & docking” mode. In the search of conformational structures in docking mode, a great number of candidates need to be evaluated and ranked, and the results are ordered ascendingly by their total scores, each of which is the receptor-wide accumulation of scores that are computed on a per-atom pair basis. Unlike that, in the “score only” mode (without docking), scores of all receptor-ligand atom pairs are directly computed and summed to form the overall score, in which the input coordinates or conformation of ligand are considered in the optimal pose to the receptor (e.g., ligand obtained from crystal or cryo-EM structure). In idock, although not optimal due to the redundant grid maps creation, a “--score_only” argument can be passed to perform direct computation with input coordinates. In Vina, the “score only” mode is not officially documented or mentioned anywhere, but its command-line binary does have such a hidden parameter which works efficiently. In our research of protein characterization, an established implementation of the scoring function is a good starting point in which we can build our core method. More precisely, we adopted and refactored the score-only mode of idock to achieve the efficiency of that of Vina as well as to support our per residue energy (score) contribution calculation. The “score & docking” mode virtually performs the regular docking and “score only” calculation in a single run.

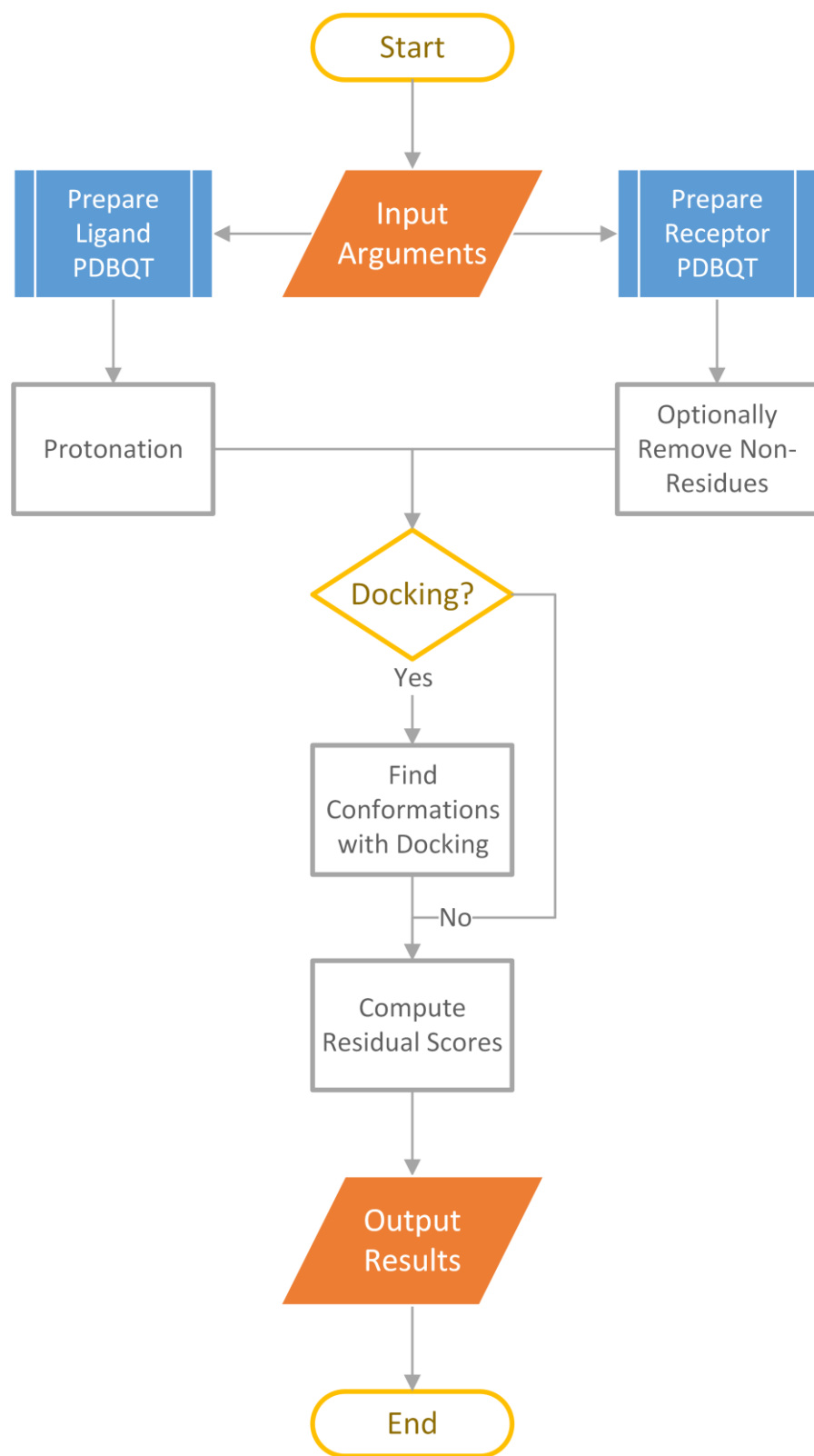


Figure 4 Workflow for the MCCS analysis

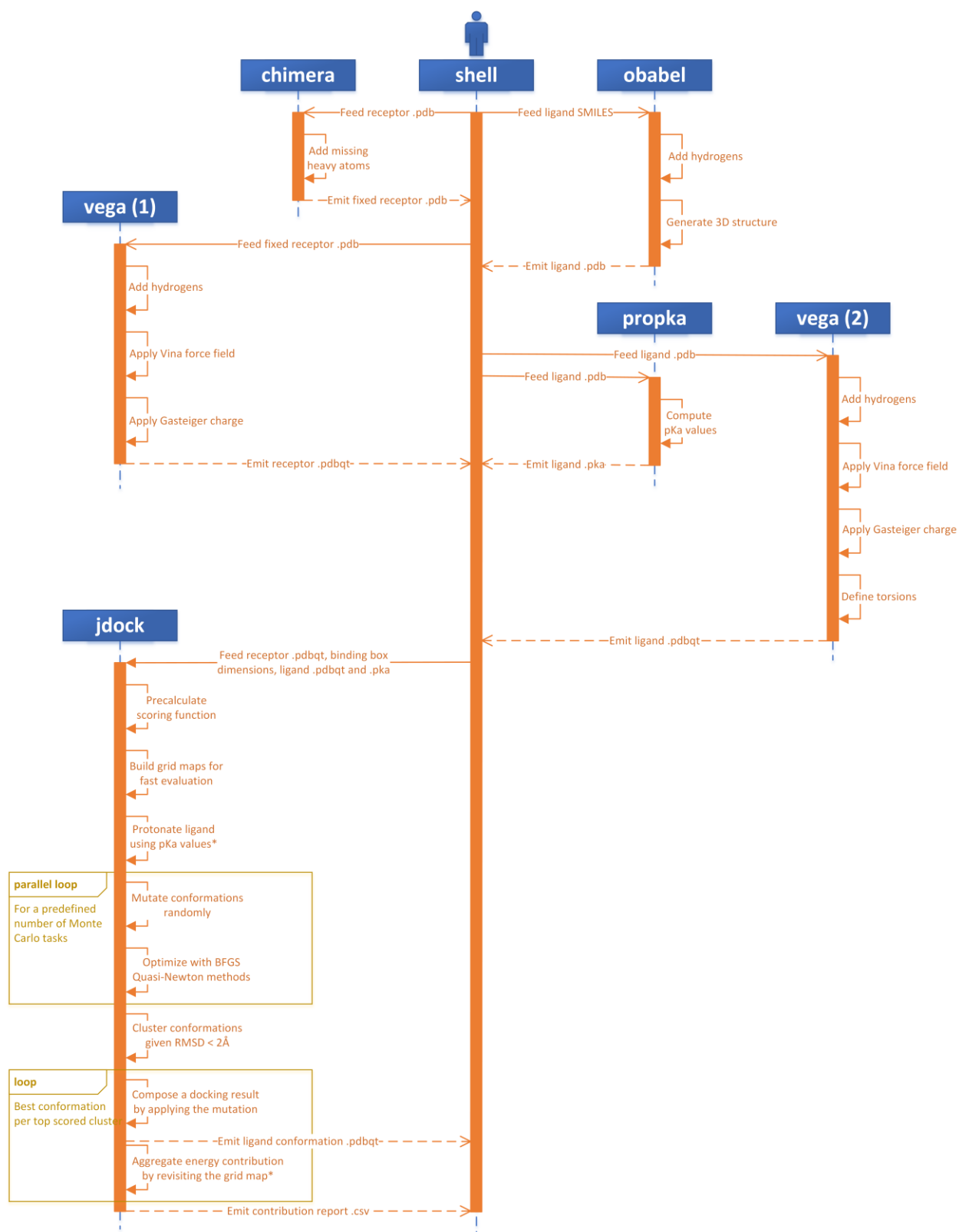


Figure 5 The Revised Docking Workflow of idock.

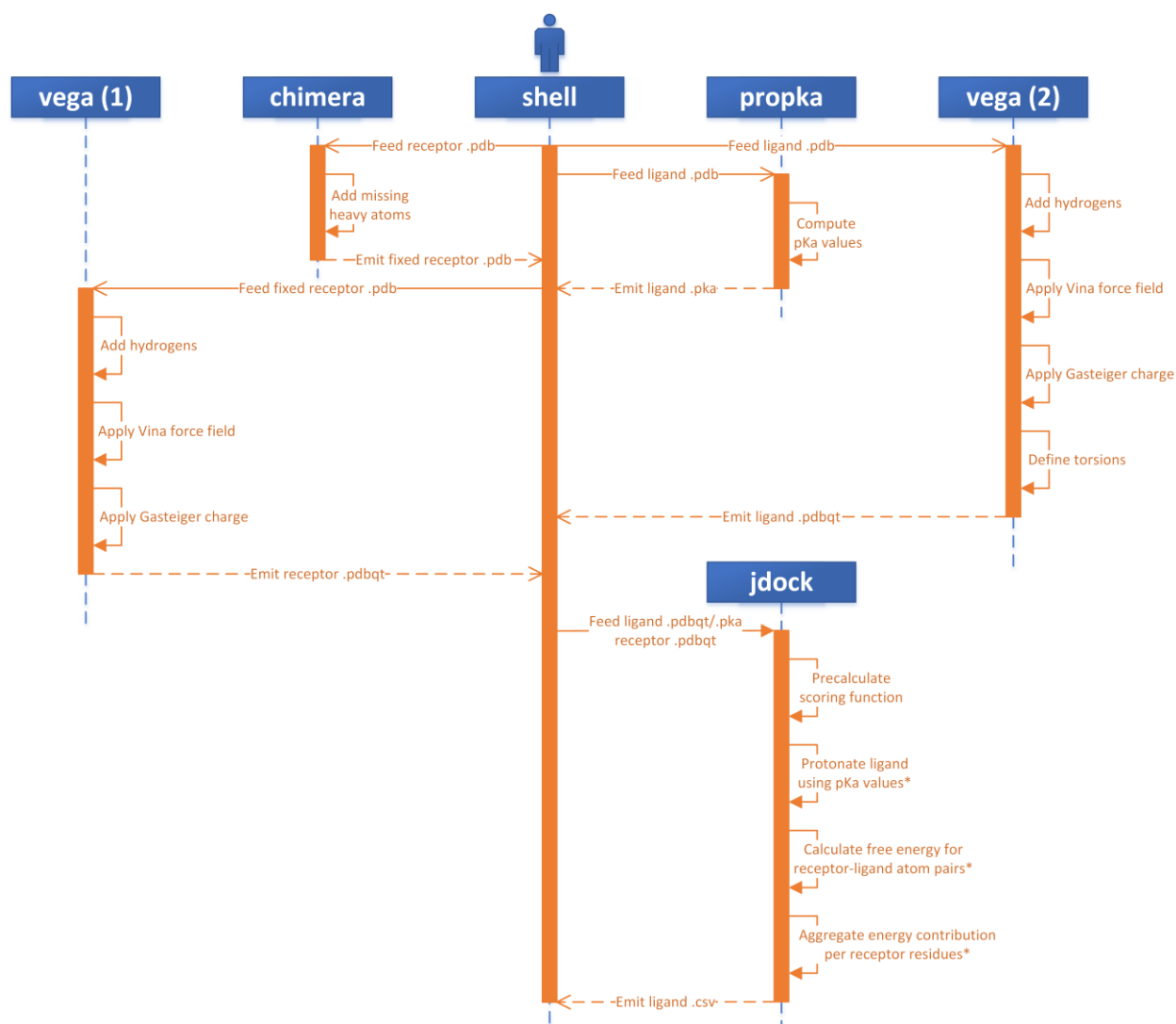


Figure 6 The Score Only Workflow of jdock.

Jdock is a variant and successor of idock[31] built for M CCS. It adopts the exact same 5-term scoring function invented by AutoDock Vina[42] and could generate a vector of residual free energy from the conformation either predicted from Monte Carlo based docking algorithm or determined by X-ray crystallography or cryo-EM. Equation 1-5 are the five core terms equation used by the scoring function in Vina and idock.

Equation 1. Gauss₁

$$\text{gauss}_1(d_{ij}) = e^{-\left(\frac{d_{ij}}{0.5}\right)^2}$$

Equation 2. Gauss₂

$$\text{gauss}_2(d_{ij}) = e^{-\left(\frac{d_{ij}-3}{2}\right)^2}$$

Equation 3. Repulsion

$$\text{repulsion}(d_{ij}) = \begin{cases} d_{ij}^2, & \text{if } d_{ij} < 0 \\ 0, & \text{if } d_{ij} \geq 0 \end{cases}$$

Equation 4. Hydrophobic

$$\text{hydrophobic}(d_{ij}) = \begin{cases} 1, & \text{if hydrophobic and } d_{ij} \leq 0.5 \\ 1.5 - d_{ij}, & \text{if hydrophobic and } 0.5 < d_{ij} < 1.5 \\ 0, & \text{if not hydrophobic or } d_{ij} \geq 1.5 \end{cases}$$

Equation 5. Hbonding

$$\text{hbonding}(d_{ij}) = \begin{cases} 1, & \text{if hbonding and } d_{ij} \leq -0.7 \\ d_{ij}/(-0.7), & \text{if hbonding and } -0.7 < d_{ij} < 0 \\ 0, & \text{if not hbonding or } d_{ij} \geq 0 \end{cases}$$

In a typical docking process, the atom pair consists of one atom from the protein and the other from the small molecule. Essentially the energy terms are related to three variables: the distance between and the van der Waals radii of the two interacting atoms. By introducing a new variable d_{ij} to represent the surface distance between the interacting atoms, the functions can be reduced to unary functions of d_{ij} . In this model, atom interactions are divided into three kinds, i.e. regular interaction, hydrophobic interaction or hydrogen bond interaction. The hydrophobic(d_{ij}) and the hbonding(d_{ij}) function evaluate to a non-zero value only if the atom pair is of the hydrophobic interaction or the hydrogen bond, respectively. Figure 7 depicted the relationship between these energy terms and the distance between the surfaces of the two interacting atoms.

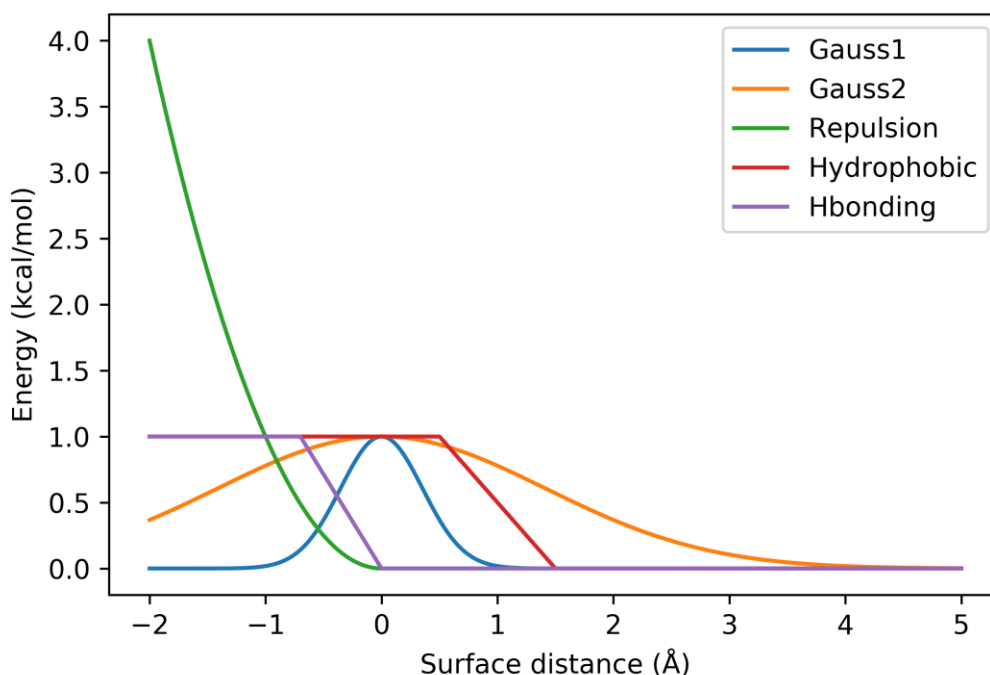


Figure 7 Variations of each energy term with surface distance

The weighted sum of the five terms forms the total score (as visualized in Figure 8 and detailed in Equation 6), where the coefficients are also given in the original Vina literature.

Equation 6. Atomic total score

$$\begin{aligned} \text{score_atomic}(d_{ij}) = & (-0.035579) * \text{gauss}_1(d_{ij}) \\ & + (-0.005156) * \text{gauss}_2(d_{ij}) \\ & + (+0.840245) * \text{repulsion}(d_{ij}) \\ & + (-0.035069) * \text{hydrophobic}(d_{ij}) \\ & + (-0.587439) * \text{hbonding}(d_{ij}) \end{aligned}$$

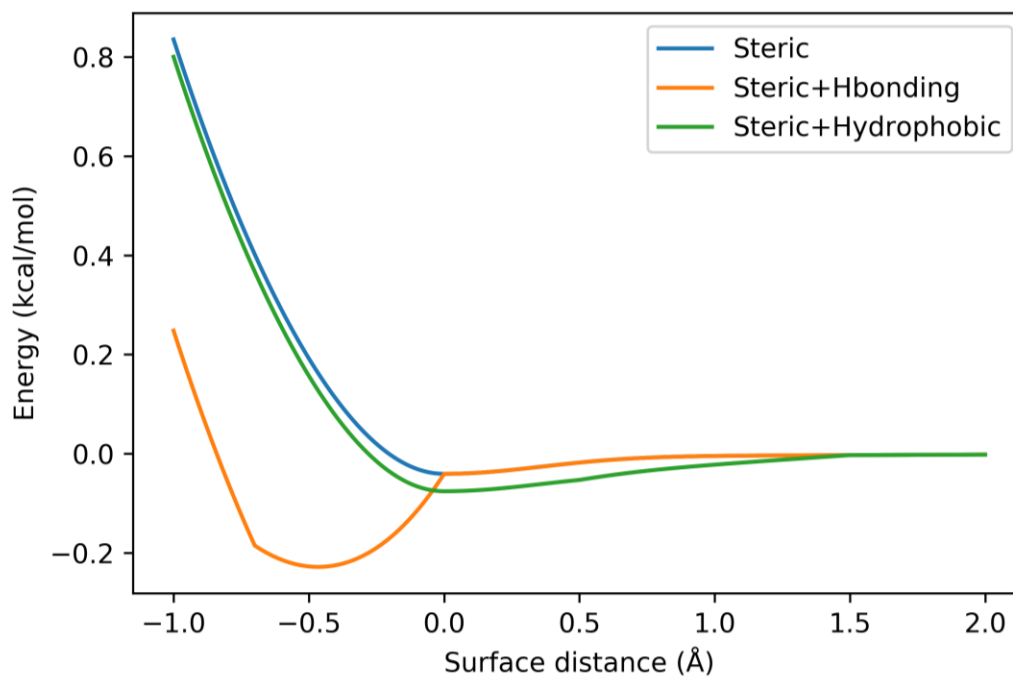


Figure 8 Variation of total atomic energy with surface distance

The residue energy contribution is calculated as Equation 7, where R is the set of atoms in the residue being considered and L is the set of atoms in the ligand whose coordinates are either computed from Monte Carlo based docking or determined by X-ray crystallography or cryo-EM. As the hydrophobic and hydrogen bond interactions are both one term more than a regular interaction, the additional terms are the key to making a residue prominent.

Equation 7. Residue total score

$$\text{score_residue} = \sum_{\substack{i \in R \\ j \in L}} \text{score_atomic}(d_{ij})$$

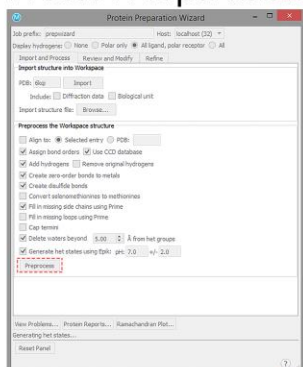
In our study, the “score only” mode in jdock was used to directly compute the residue energy vector for the characterization of the published 3D GPCRs structure and applied docking with scoring mode for CB1/2 modulators within the predicted binding cavity.

2.2.2 Glide docking

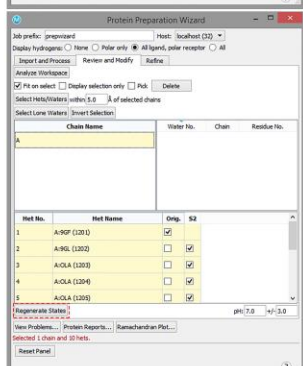
When it comes to predicting CB1/CB2–modulators binding modes, the simple docking algorithms meet difficulties as the half of the binding sites are composed by cell membrane, and even worse is that the pockets are too wide and shallow to constrain the modulator small molecules. Therefore, the flexible glide docking with constraints was applied in this case.

The glide docking was ran by Maestro 11.2[43], structures were directly got from PDB database by Maestro 11.2 (CB1: 6kqi[44], CB2: 6PT0[45]), all small molecules were got from pubchem, and the protein preparation and ligand preparation were followed. The protein structures were undergone preprocess, modify to proper states, and refine; the ligand structures were

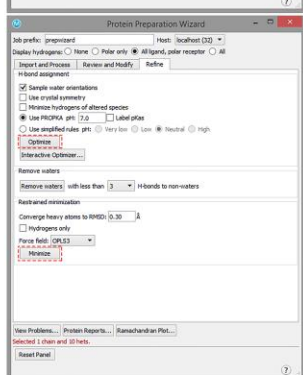
Step 1:



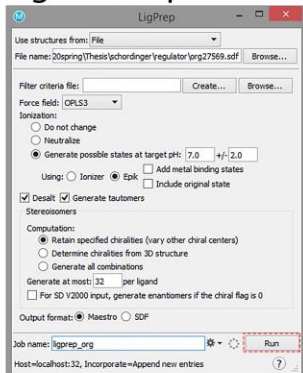
Step 2:



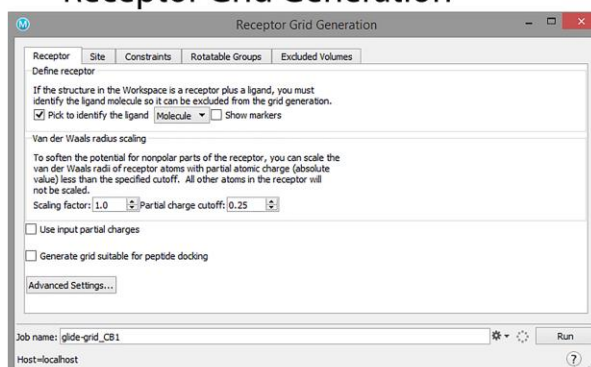
Step 3:



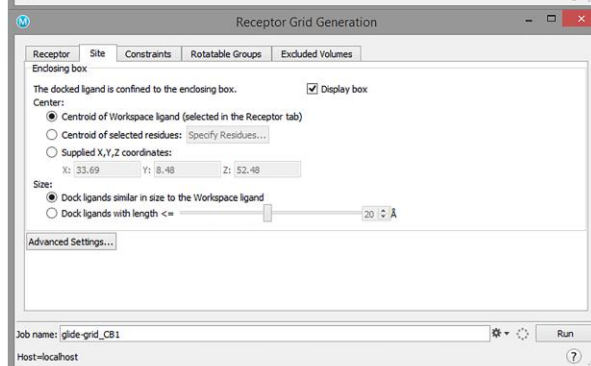
Step 4:



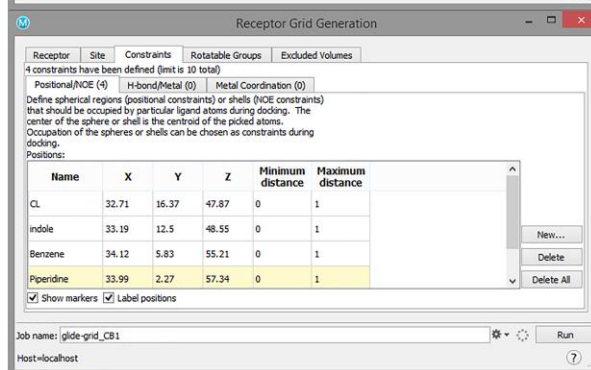
Step 5:



Step 6:



Step 7:



Step 8:

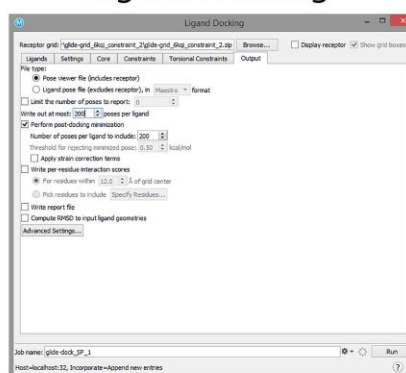


Figure 9 Overall workflow of glide docking.

undergone ionization and stereoisomers. Next, based on the binding information supported by crystalized CB1-modulator complex (PDB: 6kqi), CB1 and CB2 grids with constraints were generated. Flexible docking with constraints were carried out finally, detailed parameters were shown in Figure 9.

2.3 Similarity and clustering matching

To quantify the similarity between two vectors, a real-valued similarity function is used in statistics and related fields. Such similarity measures include cosine similarity, the Pearson correlation coefficient (the PCC), Euclidean distance, etc. Among all the various similarity measures, cosine similarity and the PCC are commonly used for real-valued vectors, which are suitable in our scenario.

Given two vectors of residue energy contribution, \mathbf{u} and \mathbf{v} , the cosine similarity, $\cos(\theta)$, is represented using a dot product and vector length as Equation 8. θ is the angle between the two vectors and the subscripts i refers to the residue numbering under the selected scheme. The function gives a positive value (up to one) for similar vectors and either zero or a negative value for dissimilar vectors.

Equation 8. Cosine similarity

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} = \frac{\sum_{i=1}^n u_i v_i}{\sqrt{\sum_{i=1}^n u_i^2} \sqrt{\sum_{i=1}^n v_i^2}}$$

The PCC is defined in the same form but subtracts the mean from every vector element (as shown in Equation 9).

Equation 9. Pearson correlation coefficient

$$\rho(\mathbf{u}, \mathbf{v}) = \frac{\text{cov}(\mathbf{u}, \mathbf{v})}{\sigma_u \sigma_v} = \frac{\sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^n (u_i - \bar{u})^2} \sqrt{\sum_{i=1}^n (v_i - \bar{v})^2}}$$

With any similarity measure, protein clustering can be carried out in such a way that proteins in the same group are more alike, in terms of binding mode, to each other than to that of other groups. In the visualization, a heatmap of an (n, n) -sized similarity matrix is commonly used for showing similarities among a set of n vectors, with a dendrogram for demonstration of the clustering. The grids of the heatmap use a color scale to display a color mapped from its numeric value which represents the similarity between two vectors. The clustering can be carried out directly with the n vectors or with the similarity rows or columns of the n vectors.

2.4 Molecular Dynamics- (MD) Simulation

The docking poses of CB1/CB2 with their modulators were used to perform the MD simulation. Each system consisted of one copy of CB1/CB2 structure, one modulator, 240 POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) molecules, taranabant, Na⁺, Cl⁻, and water molecules. MD simulation was carried out using the AMBER18 [46] software package, periodic boundary condition (PBC) were used in entire simulations, which enabled isothermal-isobaric ensembles generated by the modified PMEMD.CUDA program in AMBER 18. Temperature was

regulated with a collision frequency of 5 ps, and the pressure relaxation time was set to 1.0 ps. The MD system was first relaxed by a set of minimizations, which removed possible steric clashes. Integration of the equations of motion was conducted at a time step of 1 fs for the relaxation phase and 2 fs for the equilibrium and sampling phases. The subsequent MD simulation step consists of three phases, namely, the relaxation phase, the equilibrium phase, and the sampling phase. In the relaxation phase, the simulation system was heated up progressively from 50 K to 250 K at steps of 50 K. In the following equilibrium phase, the system was equilibrated at same temperature until stable. Finally, a 100-nanosecond MD simulation was performed to produce NTP (constant temperature and pressure) ensembles, longer MD simulation may be performed if needed. The r.m.s.d. value support that the CB1/CB2-modulator complexes were very stable.

2.5 CB1/CB2 pocket prediction and analysis

Allosteric binding pockets of CB1 / CB2 were predicted by CavityPlus Server (<http://www.pkumdl.cn:8000/cavityplus/index.php>)[47], Protein Allosteric and Regulatory Sites (PARS) (<http://bioinf.uab.cat/cgi-bin/pars-cgi/pars.pl>)[48], and Sybyl-X. All predicted pockets from these methods were undergone further pocket analysis by CavityPlus Server. The ligandability, Druggability, CavityScore, and the correction with the motions of orthosteric sites were taken into account when screened the possible allosteric pocket. The ligandability value represents the possibility of designing small ligands with high binding affinities to a certain cavity. The druggability value reflects the possibility of a cavity being a good target for binding drug-like molecules. The CavityScore is influenced by cavity volume, pocket lip size, hydrophobic volume, cavity surface area, and hydrogen-bond-forming surface area. It uses the correlation with the

motions of orthosteric sites, which come from the hypothesis that the motions of orthosteric and allosteric sites are highly correlated. The correlation was calculated using the Gaussian network model, the Gaussian network model is a minimalist Normal-Mode Analysis model, and then the calculated correlations were normalized by the Z-score.

2.6 Statistics

Data analyses were performed using SPSS for Windows, release 26.0.0.0 (SPSS, Chicago, IL, USA). The *Paired t-test* was used for analyzing the mean residue energy contributions on active conformation and inactive conformation of selected GPCRs. In order to analyze the average residue energy contributions on all GPCRs and the exact residue energy contributions on each PDB files of specific GPCRs, an *independent sample t-test* was used in the present work. A p-value < 0.05 was considered significant.

3.0 RESULTS

3.1 Characterization of traditional orthosteric binding pocket of GPCRs

3.1.1 Receptor-Ligand Dataset of 249 Crystallized GPCRs

In order to characterize the traditional orthosteric binding pocket of each class GPCRs using MCCS, a list of receptor-ligand complexes of G-protein coupled receptors (GPCRs) was retrieved and downloaded from public databases, including GPCRs database <https://www.gpcrdb.org> [38] and Protein Data Bank (PDB) <https://www.rcsb.org> [49].

Currently there are 276 resolved PDB files belonging to 63 individual GPCRs available. 249 PDB files (90.2%) from 53 individual GPCRs (84.1%) were selected for further studies, all of which provided effective receptor-small molecule interactions. As shown in Figure 11, the pink sector indicated the Class A (Rhodopsin) GPCRs, class B1 (Secretin) GPCRs in the yellow sector, class C (Glutamate) in the green sector, and class F (Frizzled) in the blue sector. Overall, two individual proteins from class C (Glutamate, 6 PDB files) and class F (Frizzled, 9 PDB files) were included in our present work respectively. They are GRM1 (human, 1 PDB file), GRM5 (human, 5 PDB file), SMO (XENLA) (African clawed frog, 1 PDB file), and SMO (human, 8 PDB files). Although there are seven individual GPCRs of class B1 being crystallized, I selected CRFR1 (human, 2 PDB files) which is the only one that resolved the orthosteric ligand to conduct the calculation. No protein in class B2 (Adhesion) had been crystallized, so none of them were selected in our work. Finally, 48 individual GPCRs belonging to class A (Rhodopsin) with 232 PDB files were selected to build our dataset. Except ACM3 (rat), ADRB1 (common turkey), NTR1 (rat),

OPRD (human & mouse), OPRM (mouse), OPSD (bovine & Japanese flying squid), LPAR6A (zebrafish), and LTB4R (guinea pig), all other selected structures of class A (Rhodopsin) belong to the human species. Additionally, three different states of GPCRs can be observed, including inactive (black wedges), intermediate (red wedges), and active states (blue wedges) [50]. Moreover, most of the collected GPCRs in the present work are in inactive conformation (171 PDB files, 68.7%) that stabilized by inverse agonist or antagonist, a few GPCRs (53 PDB files, 21.3%) are in active state that induced by agonist or partial agonist, while the rest (25 PDB files, 10.0%) are in the intermediate state. Figure 10 depicted the conformational variation between different structure states of ADRB2. As the GPCRs structure was activated, the cavity formed by its intracellular half of the bundle become slightly wider. To be more specifically, its TM6 moved outward greatly while the TM5 and TM7 moved inward slightly. Therefore, a cavity formed by TM3, TM5, and TM6 was created in which the G protein can bind[51]. And intermediate structure only underwent slighter movement and haven't reach to active state.

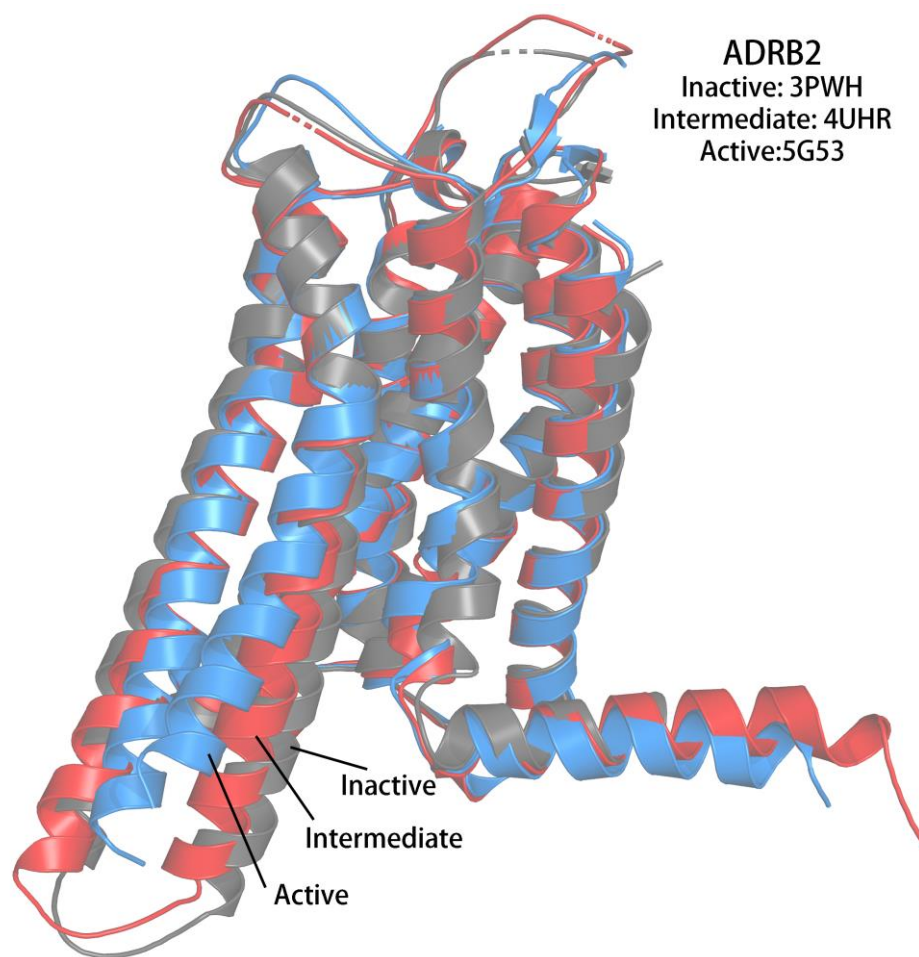


Figure 10 Different state conformations of ADRB2.

Figure 11 also showed the diverse orthosteric binding pocket between each class GPCRs. The orthosteric binding site of class A GPCRs is positioned in the extracellular half of the bundle and is formed mainly by residues from TM3, 6 and 7. In some cases, residues from other helices and the extracellular loops also contribute to the interaction between ligands and GPCRs.

Compared with class A GPCRs, there are limited number of 3D structure of other class GPCRs have been reported. Therefore, their orthosteric binding pockets haven't been understood completely up to now. Based on the limited reported structure, the small molecule ligands binding site of class B GPCRs is located deep within the intracellular half of the 7TM domain. For class F GPCRs, its special extracellular linker domain formed a small opening which enable their long and narrow binding pocket to connect to extracellular aqueous environment. Unlike other classes of GPCRs, the orthosteric binding site of class C GPCRs is at a Venus flytrap (VFT) domain, and its allosteric binding pocket partially overlaps with the orthosteric binding pocket of class A GPCRs, which has been reported and also demonstrated greater benefits than the orthosteric one in drug design [52]. Therefore, the allosteric binding pocket of class C GPCRs was selected and studied in this work.

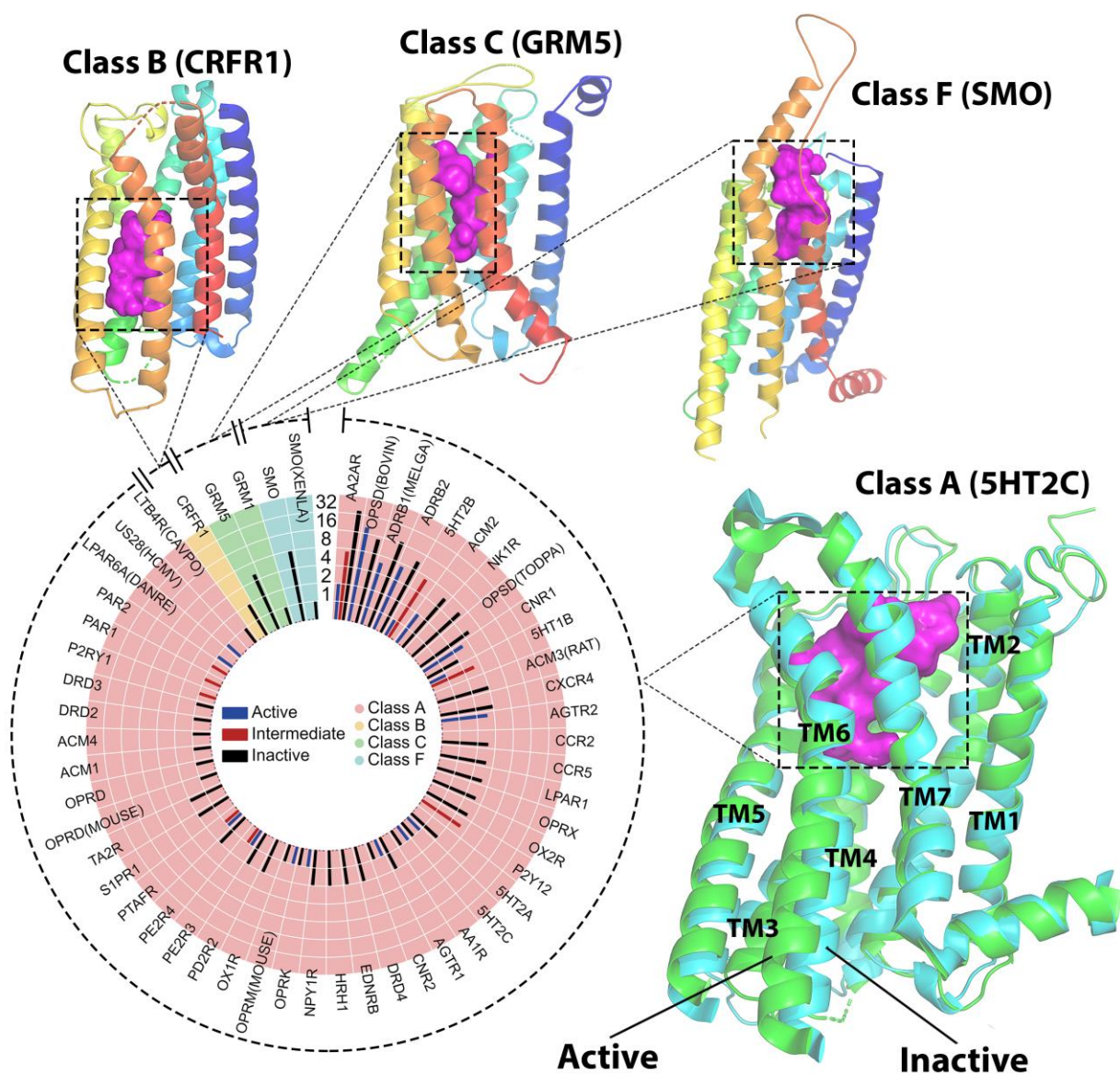


Figure 11 Radar map of 249 crystallized GPCRs.

3.1.2 Ligand-binding pocket recognition patterns of GPCRs

Using MCCC protocol, 249 PDB files of 53 individual GPCRs were prepared to calculate the residue energy contribution. Since each GPCR has different numbers of receptor-ligand complexes, the average residue energy contribution was used to sort ascendingly. As shown in Figure 12a, I plotted the heatmap of residue energy contribution in the Generic GPCR residue numbering scheme [16] for 53 individual GPCRs. X-axis represented 53 individual GPCRs, y-axis represented the binding residues of GPCRs. The residues were sorted by the average residue energy contribution ascendingly. Different box (red/yellow/green/blue) was used to highlight individual class (A/B/C/F) GPCRs. The results showed that the pattern of residue energy contribution in each class of GPCRs can distinguish itself from others, due to the different pocket of each class of GPCRs (see below).

To further investigate the contribution of residues, we selected and mapped out the decomposition of 15 representative residues, as shown in Figure 12b, where five different energy terms were calculated, the red wedges represented the hydrogen bonding component, blue wedges represented the hydrophobic component, green wedges represented the gauss1 component, magenta wedges represented the gauss2 component, and black wedges represented the repulsion component. Our results showed that three residues contributed greatly to the recognition of ligand through hydrogen bonding, including residues at position 3.32, 7.39, and 6.55 with the average energy contribution of -0.238, -0.147, and -0.171 kcal/mol. Moreover, these residues also displayed higher repulsion interactions with the average energy of 0.402, 0.185, and 0.176 kcal/mol. In addition, three other residues played an important role for the ligand recognition via hydrophobic interactions, including residues at 6.51, 3.33, and 6.48, with the average energy of -0.235, -0.174, and -0.203 kcal/mol.

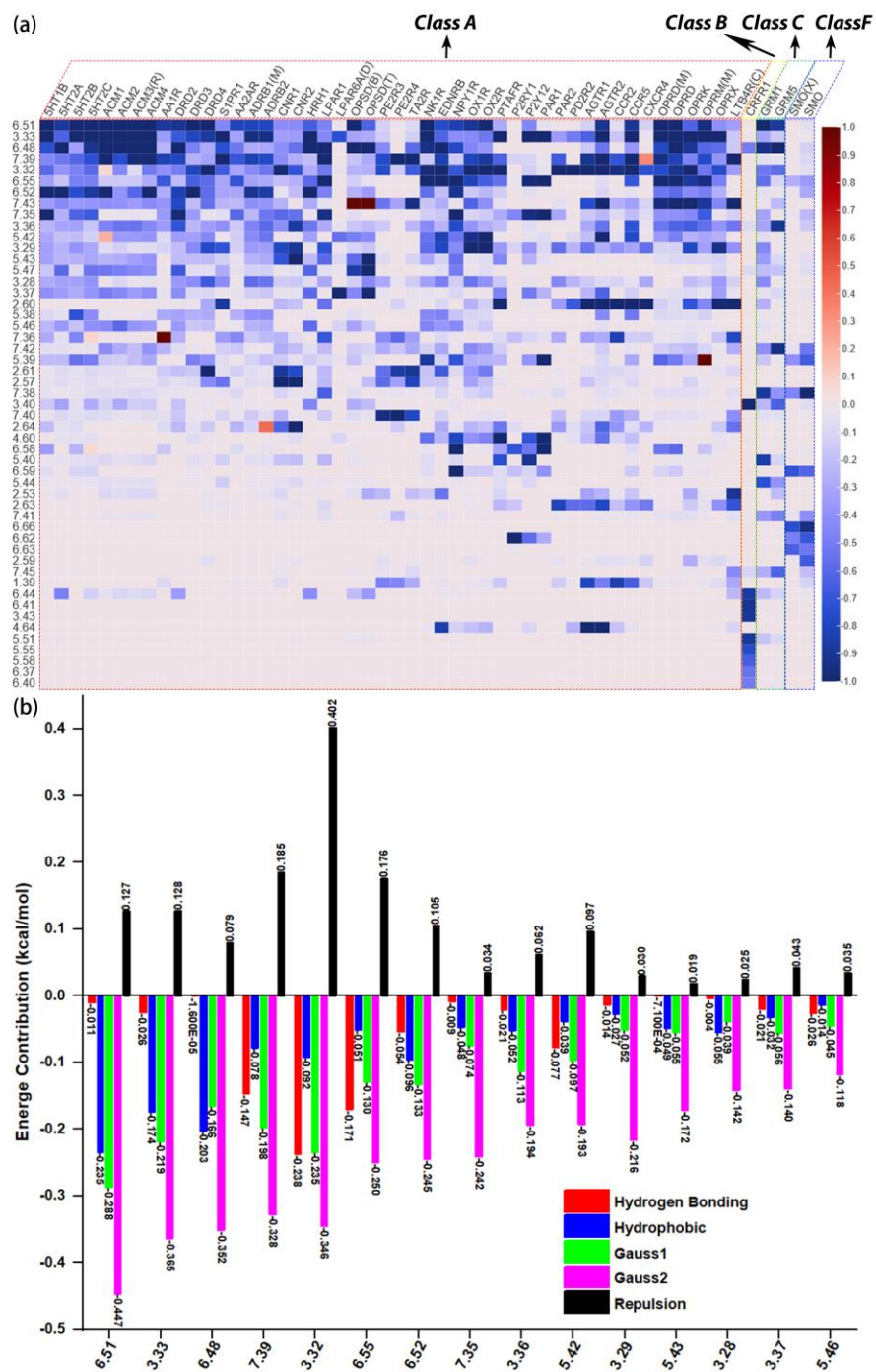


Figure 12 The Pattern of residue energy contribution and its energy decomposition.

To characterize the binding pocket in each class of GPCRs, we selected representative binding residues for further analyses as shown in Figure 13, the binding pockets were highlighted in the magenta surface. Based on the calculations of class A GPCRs (49 individual proteins with 233 PDB files), we noted that the key residues were mainly from TM3/TM6/TM7 Figure 13a, including residues at 3.33 (-0.816 kcal/mol), 6.51 (-0.708 kcal/mol), 3.32 (-0.688 kcal/mol), 7.39 (-0.643 kcal/mol), 6.48 (-0.606 kcal/mol), 6.55 (-0.481 kcal/mol), 6.52 (-0.413 kcal/mol), 2.60 (-0.391 kcal/mol), 7.35 (-0.367 kcal/mol). Due to the orthosteric binding pocket of class B1 (1 individual protein with 2 PDB files) is much deeper and closer to TM3/TM5/TM6 than that of class A, ten key residues were in the position of 3.40 (-1.013 kcal/mol), 6.41 (-0.922 kcal/mol), 3.43 (-0.898 kcal/mol), 5.51 (-0.884 kcal/mol), 6.44 (-0.870 kcal/mol), 5.55 (-0.752 kcal/mol), 5.58 (-0.652 kcal/mol), 6.37 (-0.543 kcal/mol), 6.40 (-0.506 kcal/mol), and 6.48 (-0.495 kcal/mol) (Figure 13b). Unlike other class GPCRs, the allosteric binding pocket of class C GPCRs (2 individual proteins with 6 PDB files) was selected and studied in our work. As shown in Figure 13c, although the recognition pattern of the allosteric binding site in class C shared large similarity with that of class A, six distinct residues contributed greatly in class C GPCRs, including residues at position 7.42 (-0.668 kcal/mol), 7.38 (-0.634 kcal/mol), 5.44 (-0.557 kcal/mol), 5.40 (-0.470 kcal/mol), 7.45 (-0.444 kcal/mol), and 3.40 (-0.433 kcal/mol). Finally, as shown in Figure 13d, ten key residues in the orthosteric pocket of class F (2 individual proteins with 9 PDB files) included those at position 6.66 (-0.847 kcal/mol), 7.38 (-0.808 kcal/mol), 6.59 (-0.660 kcal/mol), 6.62 (-0.591 kcal/mol), 6.63 (-0.583 kcal/mol), 5.39 (-0.530 kcal/mol), 2.59 (-0.339 kcal/mol), 6.55 (-0.313 kcal/mol), 7.42 (-0.278 kcal/mol), and 7.41 (-0.218 kcal/mol).

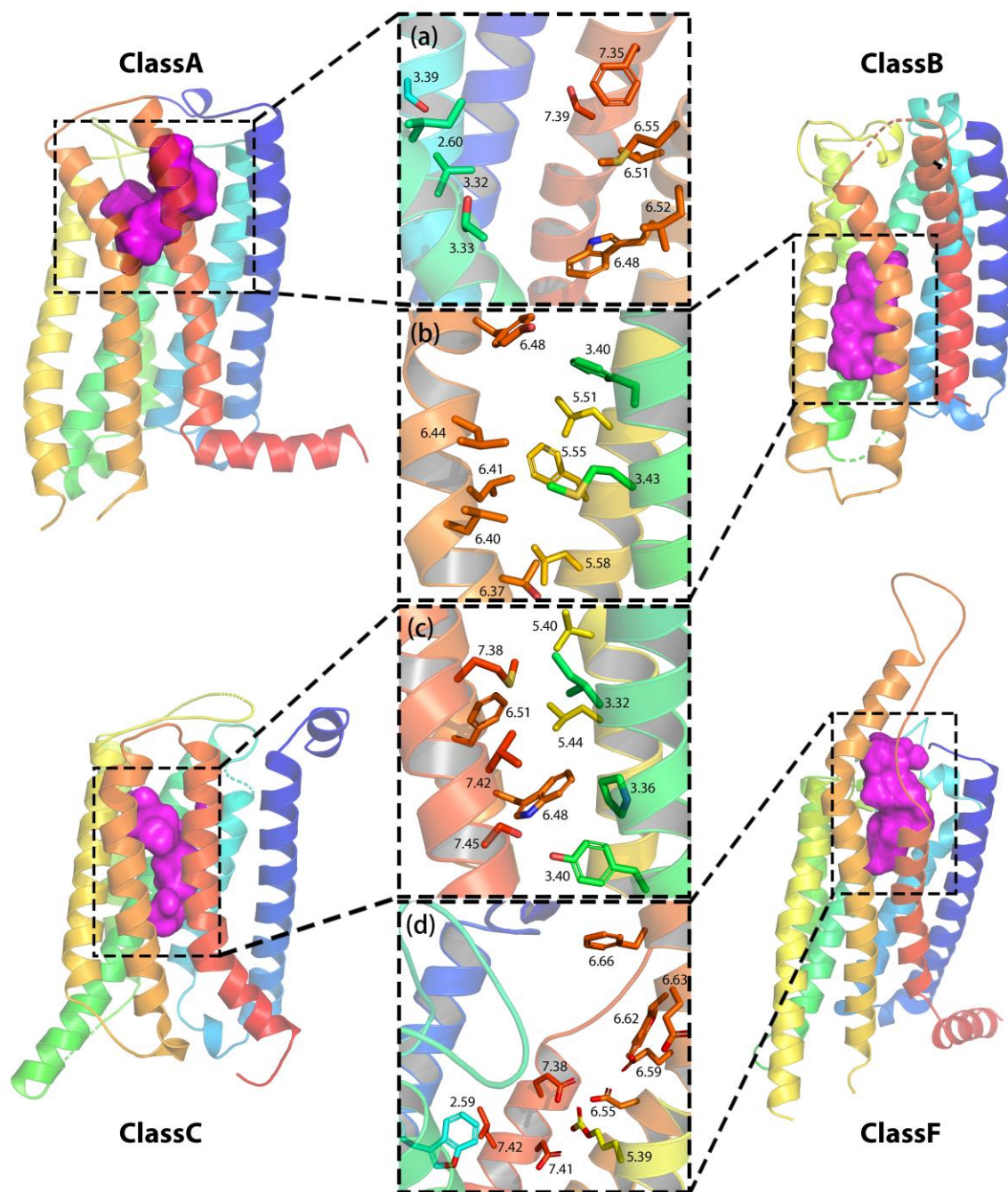


Figure 13 Representative residues involved in the binding pocket of each class GPCRs.

3.1.3 The key residues of class A GPCRs orthosteric binding pocket

In this section, we selected three most important residues from class A GPCRs structures (from 48 individual proteins) to further discuss their roles based on their residue energy contribution, including residues at position 6.48, 3.32, and residue on ECL2.

First, in order to investigate the difference of residue energy contributions between the active and the inactive state, we conducted a paired t-test and an independent sample t-test. For the paired t-test, we selected 14 individual proteins from class A GPCRs. Among them, AA2AR has inactive, intermediate, and active conformations. Ten GPCRs own both active and inactive conformations, including OPSD (BOVIN), ADRB1 (MELGA), ADRB2, ACM2, CNR1, 5HT2C, AA1R, CNR2, OPRK, and OPRM (MOUSE). Three other GPCRs including 5HT2B, 5HT1B, and PE2R3 have active and intermediate conformations being crystallized. We also carried out an independent sample t-test for (1) the average energy contribution on active and inactive structures of class A GPCRs respectively, and (2) five selected GPCRs that have multiple PDB files for each conformation, including AA2AR (active: 2 PDB files, inactive: 36 PDB files), OPSD (BOVIN, active: 20 PDB files, inactive: 14 PDB files), ADRB2 (active: 7 PDB files, inactive: 11 PDB files), ACM2 (active: 2 PDB files, inactive: 6 PDB files), and CNR1 (active: 2 PDB files, inactive: 5 PDB files).

Based on our results, the mean energy contribution of residue at position 6.48 (inactive: -0.875 kcal/mol, active: -0.422 kcal/mol, p-value = 0.008) (Figure 14a) showed a significant difference between active and inactive conformation when we conducted a paired t-test. Moreover, for the independent sample t-test between active and inactive structures, the average contribution of residue 6.48 showed a similar result (inactive: -0.779 kcal/mol, active: -0.372 kcal/mol, p-value = 0.004) (Figure 14b). Based on the results of the independent sample t-test for five selected

GPCRs, we found that except on AA2AR (inactive: -0.609 kcal/mol, active: -0.562 kcal/mol), residue 6.48 on the other four GPCRs showed a significant contribution on inactive conformations rather than on the active ones (Figure 14b). Regarding AA2AR, its anomaly may be attributed to its limited conformational changes of Trp6.48 (“Toggle Switch”) between inactive and active states (Figure 14c). Finally, we investigated the binding pockets of 48 selected class A GPCRs and found that 42 out of 48 residues (87.5%) at position 6.48 were Tryptophan (Trp), which Trp6.48 was highly conserved among class A GPCRs. This results are consistent with other’s experiments, Lin and Sakmar have reported that Trp6.48 has a crucial role in class A GPCR activation.

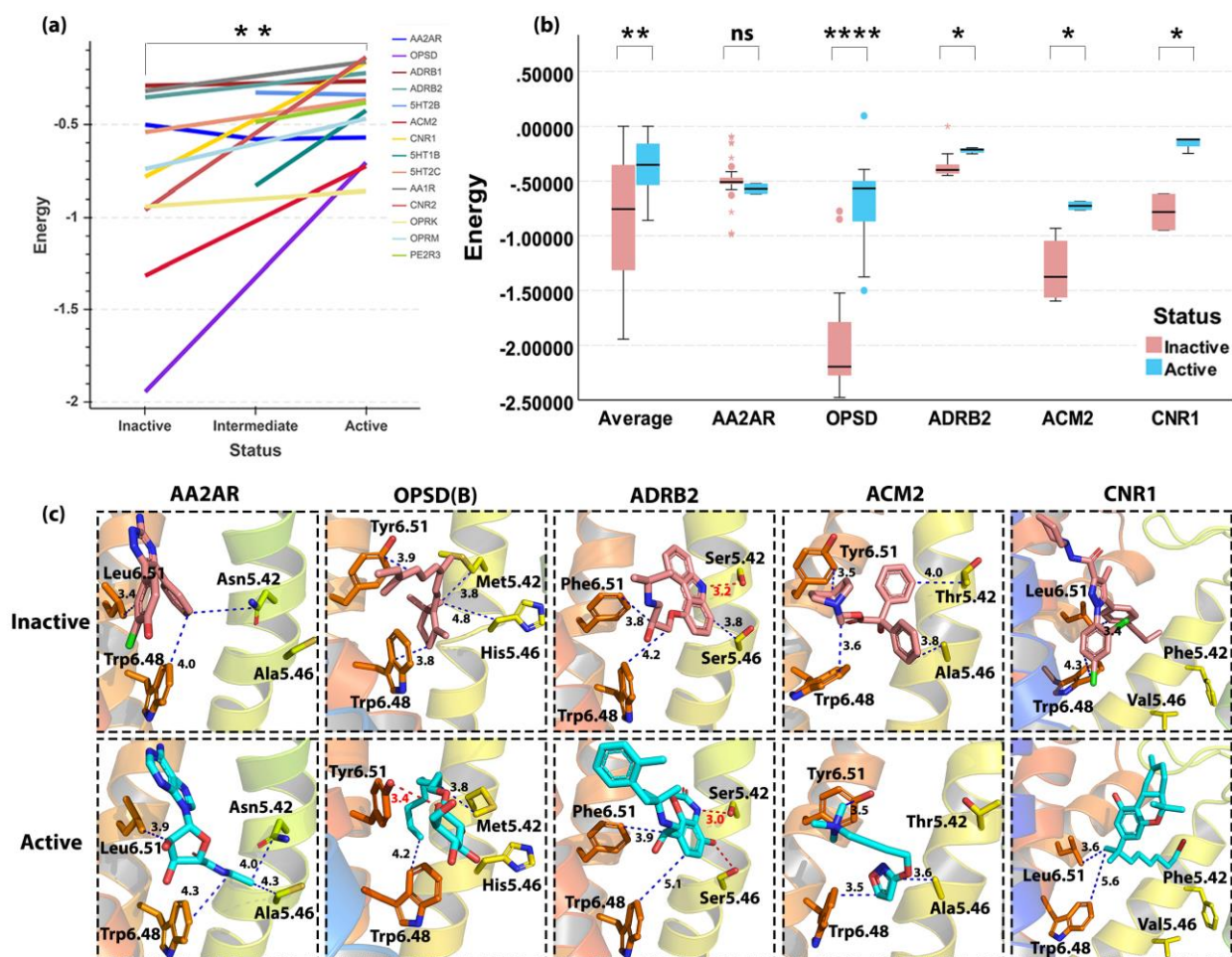


Figure 14 The role of residue 6.48 for distinguishing the inactive status from the active one.

Sequentially, we investigated the role of residue 3.32 in class A GPCRs, in which the 3.32 position is reported to have an important role in ligand recognition. As shown in Figure 15a, we found that aspartic acid (Asp) residue is the most conserved residue at position 3.32, with the number of 19 (~40%) out of 48 individual class A GPCRs. Moreover, 15 out of 19 Asp3.32 (~79%) contributed greatly to the ligand binding via hydrogen bonding interaction (Figure 15b), with an average energy contribution of -0.801 kcal/mol. Moreover, 4 glutamine acid (Gln) residues (~8%) were found in 48 individual GPCRs, and two of them (in EDNRB and NPY1R) contributed a moderate energy of hydrogen bonding (~0.215 kcal/mol, Figure 15b) to ligand's recognition. Finally, other residues at position 3.32 were mainly hydrophobic, including phenylalanine acid residue (Phe, 7, ~15%), valine acid residue (Val, 5, 11%), and so on (Figure 15).

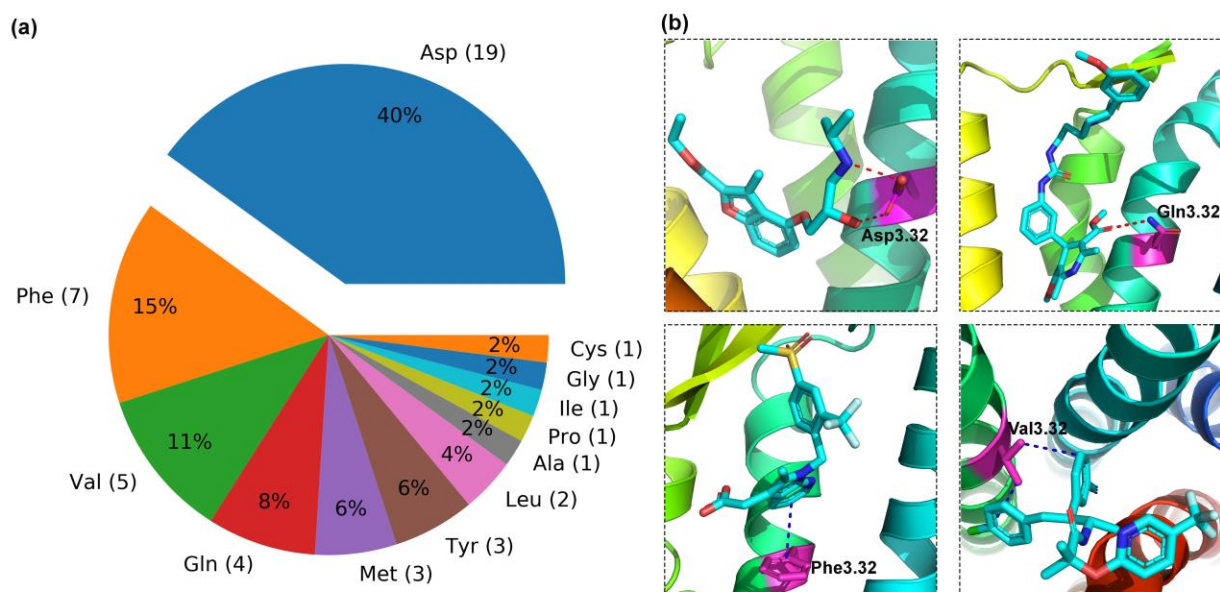


Figure 15 Multiple roles of residue 3.32 for ligand recognition.

Additionally, I also analyzed the role of residues on ECL2. By analyzing the 3D structures of GPCRs, I noted that extracellular loop 2 (ECL2) plays an important role in recognizing the binding of ligands. Here I selected 23 individual GPCRs and focused on the analyses of one of its most important residues at the same position on ECL2. Figure 16 showed the overlap of residues at the same position on ECL2, and the Table 1 showed the energy contributions of these residues in different GPCRs. First, we found that these residues directly interacted with the small molecule in the orthosteric ligand binding site of GPCRs. Moreover, most of these residues contributed greatly to the binding of ligand within a receptor through hydrophobic interactions, including Phe (9, 39%), Val (3, 13%), Ile (3, 13%), Leu (2, 8%), and so on. Final, we found that residues in ECL2 may contribute to GPCRs' selectivity. Taking dopamine receptors as an example, we found that the residues on DRD2 and DRD3 are the hydrophobic residues (Ile184 and Ile183), while the corresponding residue on DRD4 is a charged amino acid (Arg186), which may relate to the selectivity among dopamine receptors.

All the information about the key residues discussed above provides new insight into the binding pocket and will facilitate the drug design of GPCRs.

Table 1 The role of residue at the same position on ECL2 for GPCR-ligand complex.

Protein	Residue	Energy contribution
5HT1B	Val200	-0.571 kcal/mol
5HT2A	Leu228	-0.380 kcal/mol
5HT2B	Val208	-0.489 kcal/mol
5HT2C	Val208	-0.831 kcal/mol
AA1R	Phe171	-0.942 kcal/mol
AA2AR	Phe168	-0.574 kcal/mol
ADRB1	Phe201	-1.096 kcal/mol
ADRB2	Phe193	-1.213 kcal/mol
CNR1	Phe268	-1.573 kcal/mol
CNR2	Phe183	-1.643 kcal/mol
DRD2	Ile184	-0.265 kcal/mol
DRD3	Ile183	-0.772 kcal/mol
DRD4	Arg186	-0.338 kcal/mol
LPAR1	Met198	-0.749 kcal/mol
OPSD	Ile189	-0.417 kcal/mol
P2RY1	Asp204	-0.556 kcal/mol
PAR1	His255	-0.726 kcal/mol
PAR2	His227	-1.217 kcal/mol
PD2R2	Tyr183	-1.723 kcal/mol
PE2R3	Phe209	-0.259 kcal/mol
PTAFR	Phe174	-1.330 kcal/mol

Table 1 (continued)		
S1PR1	Leu195	-0.467 kcal/mol
TA2R	Phe184	-0.223 kcal/mol

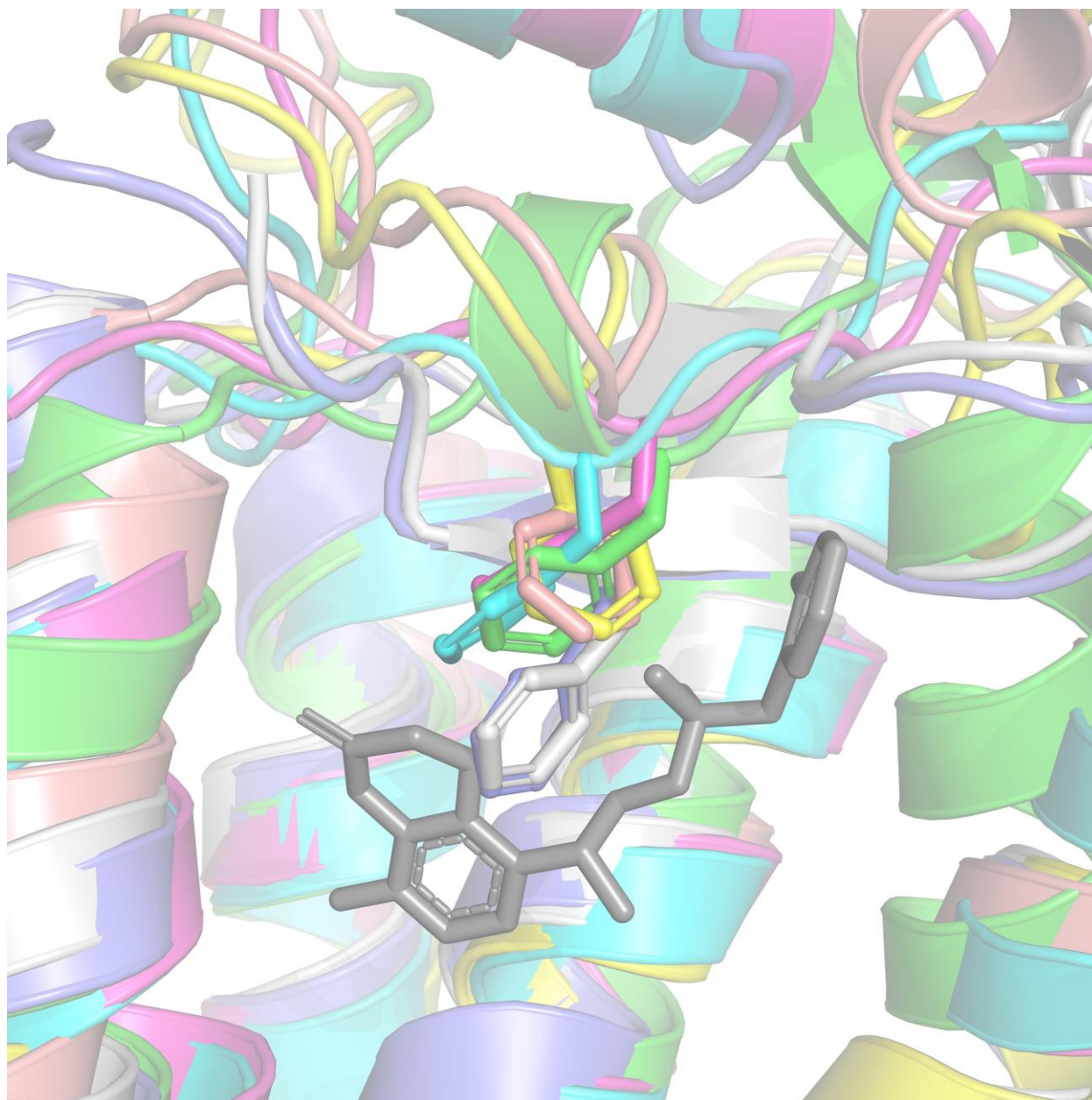


Figure 16 The role of residue at the same position on ECL2 for GPCR-ligand complex.

3.1.4 Protein Similarity and Clustering

In this section, I tried to construct a feature vector consisting of a series of aggregated energy scores for each residue for a specific region in a protein to some extent describe the property or characteristic of a protein. The outcome feature vectors of proteins can be easily compared and used to cluster similar proteins.

In Figure 17, the grids were calculated using the Pearson Correlation Coefficient where -1 in dark blue is a total negative linear correlation, 0 in white is no linear correlation, and 1 in dark red is a total positive linear correlation. The dendrogram to the top of the grids demonstrates the clustering of the similarity columns using Pearson's distance (1 minus the PCC). The dendrogram to the left of the grids demonstrates the clustering of the similarity rows using cosine distance (1 minus the cosine similarity). Both clusters use the Farthest Point Algorithm (the Voor Hees Algorithm) in determining the linkage between two clusters. As shown in Figure 17, the selection of Pearson's distance or cosine distance does not significantly affect the clustering outcome. Indeed, in most of our tests, both measures deliver like results with a minor amount of exceptions. For example, we observed that the score vector of class B (SMO and SMO(X), yellow sector) and class F (CRFR1, blue sector) were dramatically different from that of most class A and C. The score vector of P2RY1 (subclass: A-delta) was similar to that of class B, although it did show a correlation to limited GPCRs that included PAR1 (subclass: A-delta), NPY1R (subclass: A-beta), and P2Y12 (subclass: A-delta). Moreover, the score vector of CRFR1 (class F) was similar to OPSD (B) and OPSD (T). Here, although OPSD(B) and OPSD(T) were clustered with class A/C GPCRs, they located the outermost layer, indicating they showed low correlation with some class A/C GPCRs. As a result, P2RY1, SMO, SMO(X), CRFR1, OPSD (B), and OPSD (T) were clustered together as shown in Figure 17. Interestingly, we observed that the score vector of class

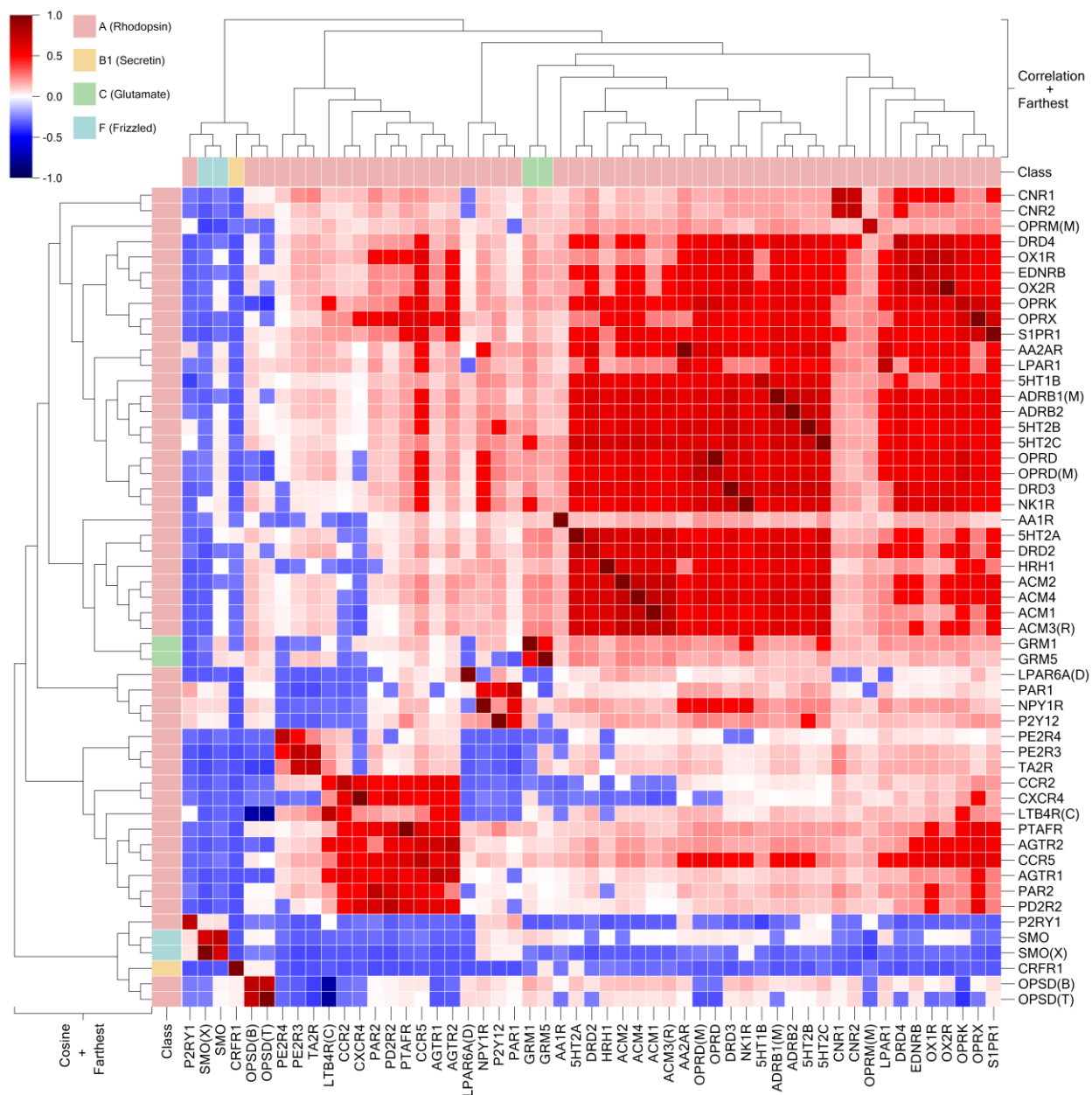


Figure 17 Heatmap demonstrating a correlation between score vectors of GPCRs.

C (GRM1 and GRM5, green sector) were similar to many class A GPCRs, so class C GPCRs were clustered with class A GPCRs together, as shown in Figure 17. Importantly, we found that proteins that shared large sequence identity or similarity were clustered together or close, indicating our approach is reasonable for protein classification or clustering. A similar heatmap with cosine similarity grids was attached as Figure 30 (Appendix), which supported our observations.

3.2 Characterization of the allosteric binding region of class A GPCRs

3.2.1 Reported Allosteric Binding Pockets of Class A GPCRs.

Unlike traditional orthosteric binding pocket, which highly conserved among sub-classes, the allosteric binding pockets of GPCRs hold huge diversity even within sub-family. Therefore, the characterization of allosteric binding pocket of GPCRs is a case by case study.

I first summarized all 11 PDB files (Table 2) for the reported class A GPCRs with effective allosteric modulator-receptor interaction. As shown in Figure 18, a total of 6 allosteric binding regions have been identified. Though the crystallized protein-molecule complex structures revealed multiple allosteric binding sites for class A GPCRs, three sites (Site A, B and C) are more typical than others, thus we focused our research on these regions.

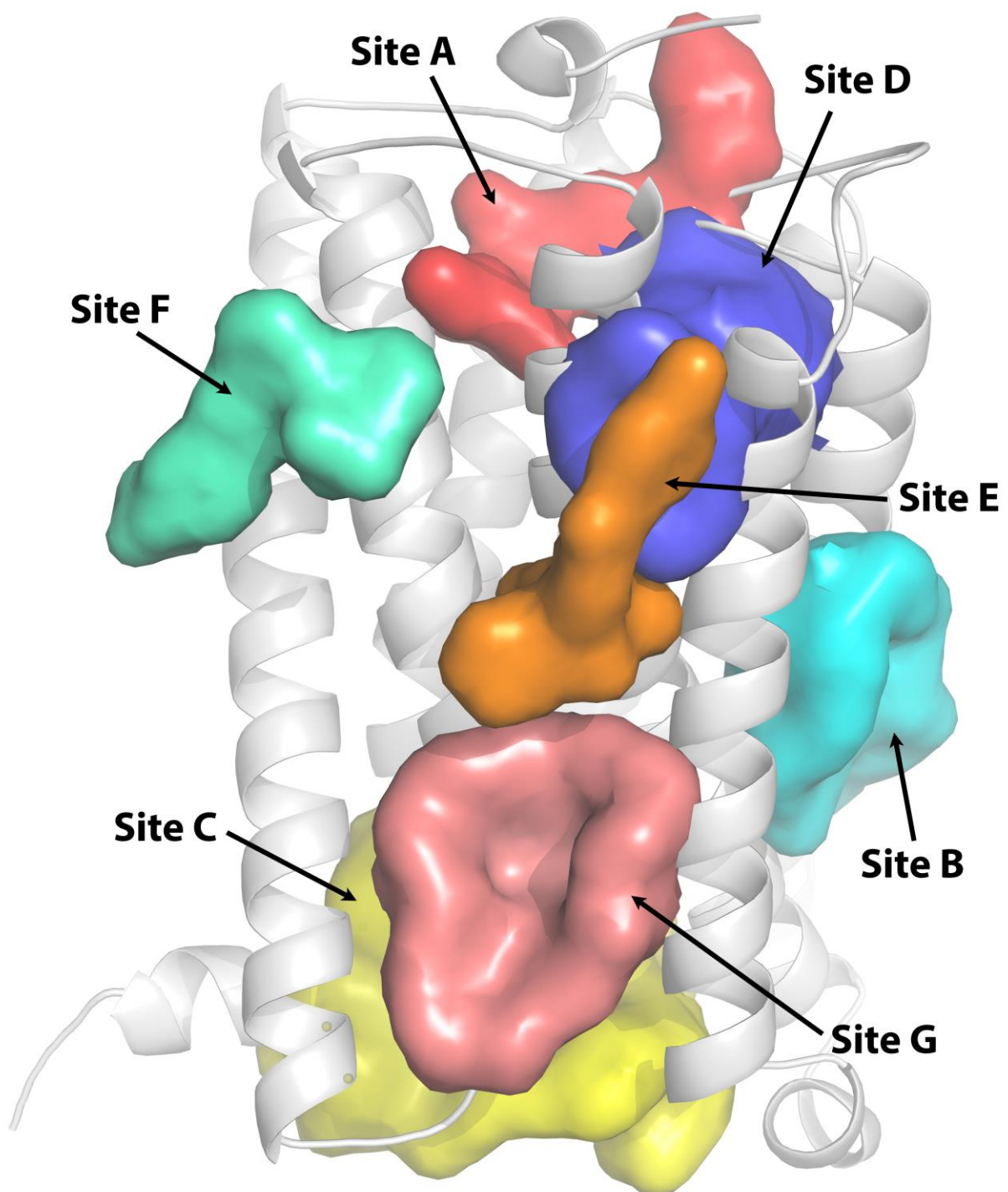


Figure 18 The summarization of allosteric binding pockets of class A GPCRs.

Table 2 Summarization of all reported allosteric regulation structures.

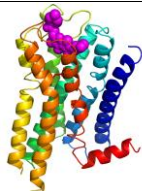






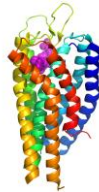

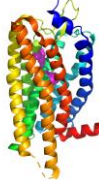
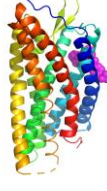

Site	Receptor	PDB code	Structure	Reference
A	ACM2	4MQT		[53]
	PAR2	5NDD		[54]
B	C5AR1	5O9H		[55]
	FFAR1	5TZY		[56]
C	CCR2	5T1A		[57]
	CCR9	5LWE		[58]

Table 2 (continued)

	ADRB2	5X7D		[59]
D	FFAR1	4PHU		[60]
		5TZR		[56]
E	PAR2	5NDZ		[54]
F	P2Y1	4XNV		[61]
G	CNR1	6KQI		[44]

3.2.2 Characterization of three typical allosteric binding region of class A GPCRs

As shown in Figure 19, Site A allosteric pocket locates in the extracellular vestibule, which is surrounded by extracellular loops (ECL) and the classic orthosteric binding region of class A GPCRs. Muscarinic acetylcholine receptor M2 (ACM2) is a typical receptor whose allosteric binding region locates in this extracellular vestibule. To further investigate the mechanisms of allosteric regulation in Site A, I analyze the small molecule-receptor recognition pattern based on MCCS protocol using the ACM2 crystallized structure (PDB:4MQT) [53]. As shown in Figure19, Iperoxo is the orthosteric ligand, while LY2119620 is the allosteric modulator. Results showed that Trp422^{7.35}, Tyr177^{ECL2}, Asn410^{6.58}, Tyr83^{2.64}, and Tyr80^{2.61} contribute greatly to the binding of LY2119620 in ACM2. Although the distance between Trp422^{7.35} and LY219620 is 4.1 Å, it contributes the greatest energy with a total energy contribution of -2.04 kcal/mol. The Tyr177^{ECL2} contributed -1.55 kcal/mol to the binding of LY219620. With a distance of 2.6 Å, Tyr80^{2.61} forms strong hydrogen bonding (HBonding energy: -0.59 kcal/mol) with LY2119620 but also shown strong repulsion (Steric energy: 0.26 kcal/mol). Other residues like Asn410^{6.58} and Tyr83^{2.64} contribute -0.67 and -0.51 kcal/mol respectively for the allosteric binding between LY2119620 and ACM2.

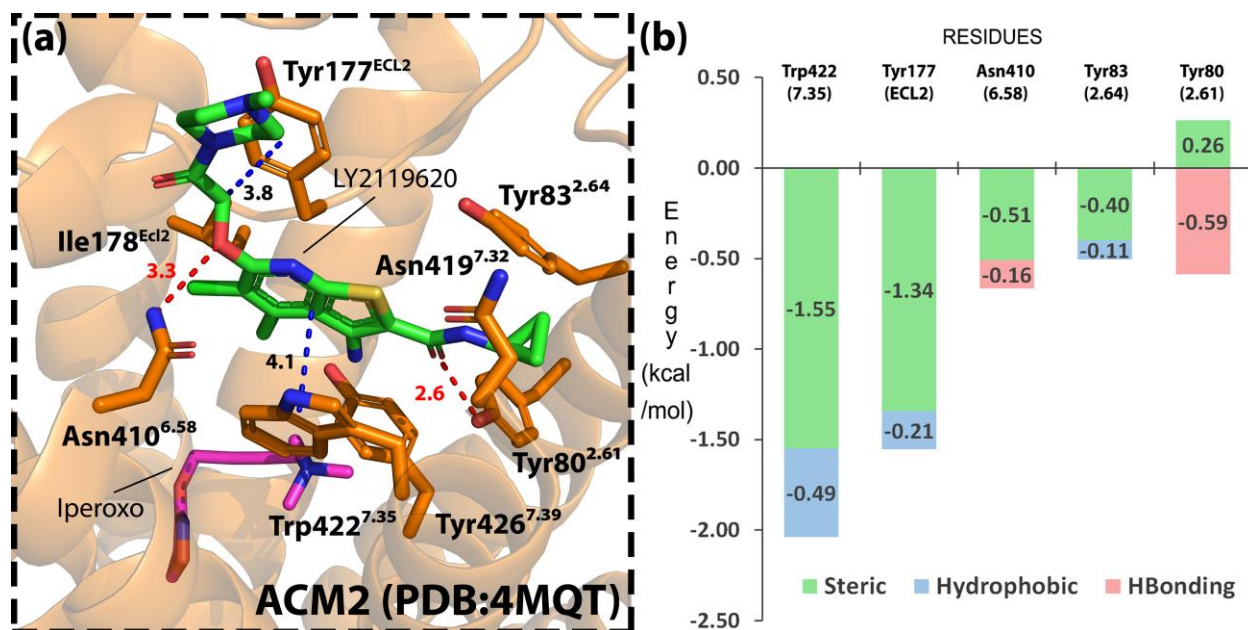


Figure 19 Detailed interaction in allosteric binding pocket of ACM2 (Site A).

As shown in Figure 20, Site B allosteric pocket locates near the TM4/5 region outside the 7TM domain. Though this binding site has great diversity in binding behavior, both proteins form strong hydrophobic interactions with the allosteric modulators during binding and share important residues including 5.49, 4.48, and 5.45. However, they have different energy contribution patterns. As shown in Figure 20a and 20b, for the C5a anaphylatoxin chemotactic receptor 1 (C5AR1)-NDT9513727 complex[55], Trp213^{5.49} contributed the greatest to the total binding energy (-1.85 kcal/mol) and formed strong hydrogen bonds with the modulator. Leu163^{4.48} contributed -0.67 kcal/mol and Leu209^{5.45} contributed -0.48 kcal/mol during the binding process. Other important residues, including Leu125^{3.37}, Ala128^{3.40}, Pro214^{5.50}, and Ile124^{3.36} all formed hydrophobic interactions with NDT9513727. As to the free fatty acid receptor 1 (FFAR1) [56], among the three common residues, Leu193^{5.49} and Ala129^{4.48} contribute equally (-0.45 kcal/mol) for the binding of allosteric modulator (MK-866), while the energy contribution of Leu189^{5.45} is -0.38 kcal/mol. Besides these common residues, the most dominant residues are Val126^{4.45} and Tyr44^{2.42}

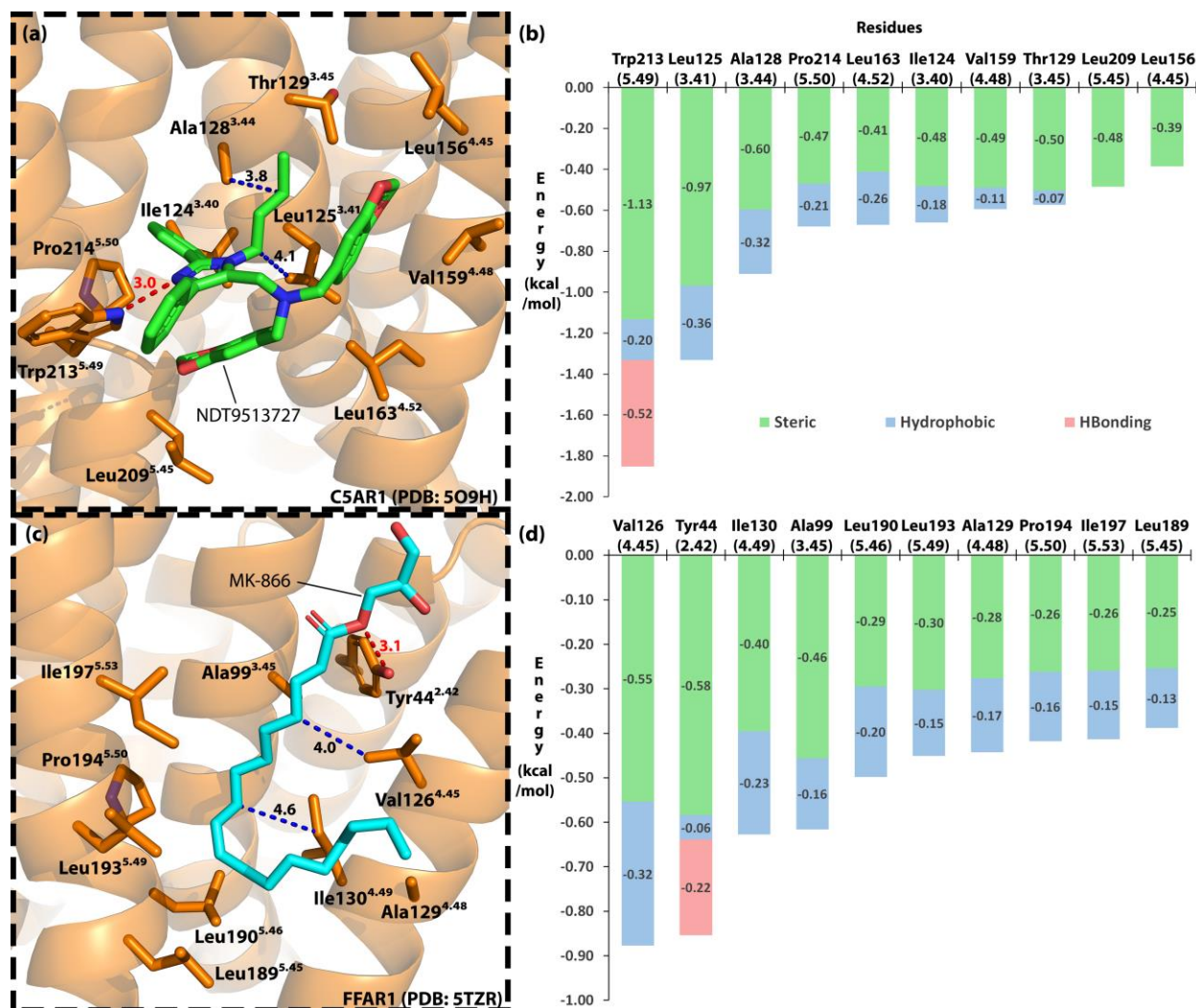


Figure 20 Interaction pattern in allosteric binding Site B.

As shown in Figure 21, Site C locates within the helical bundle of the receptor but open to the cytoplasm and spatially overlaps with the G-protein binding site. The important residues in this binding pocket include 8.50, 7.53, 8.49, 2.43, 6.36, 8.48, 8.47, 2.40, and 2.39. Among these residues, 8.50 and 7.53 is highly conserved among all the class A GPCRs (Table 3). The three reported ligands share some similarities when binding with their respective receptors. We summarized the binding poses based on the three available PDB files with ligand binding at this allosteric binding site (Figure 21a, 21b, 20c) [57] [58] [59]. A comparison of the energy

contribution of each residue is shown in Figure 21d. Among the highest energy contribution residues, F8.50 and D/K/R8.49 can form hydrophobic interactions or hydrogen bonds with all the three reported modulators. As to 7.53, hydrophobic interactions can be observed with CMPD-15PA and CCR2-RA-[R] while not for the vercirnon. The absence of the hydrophobic interaction is probably due to the near vertical spatial relationship between the residue and the benzene ring of vercirnon. Though for other residues, the energy contribution pattern is diverse and limited conservative, it may contribute to the designing of highly selective ligands for the specific receptors.

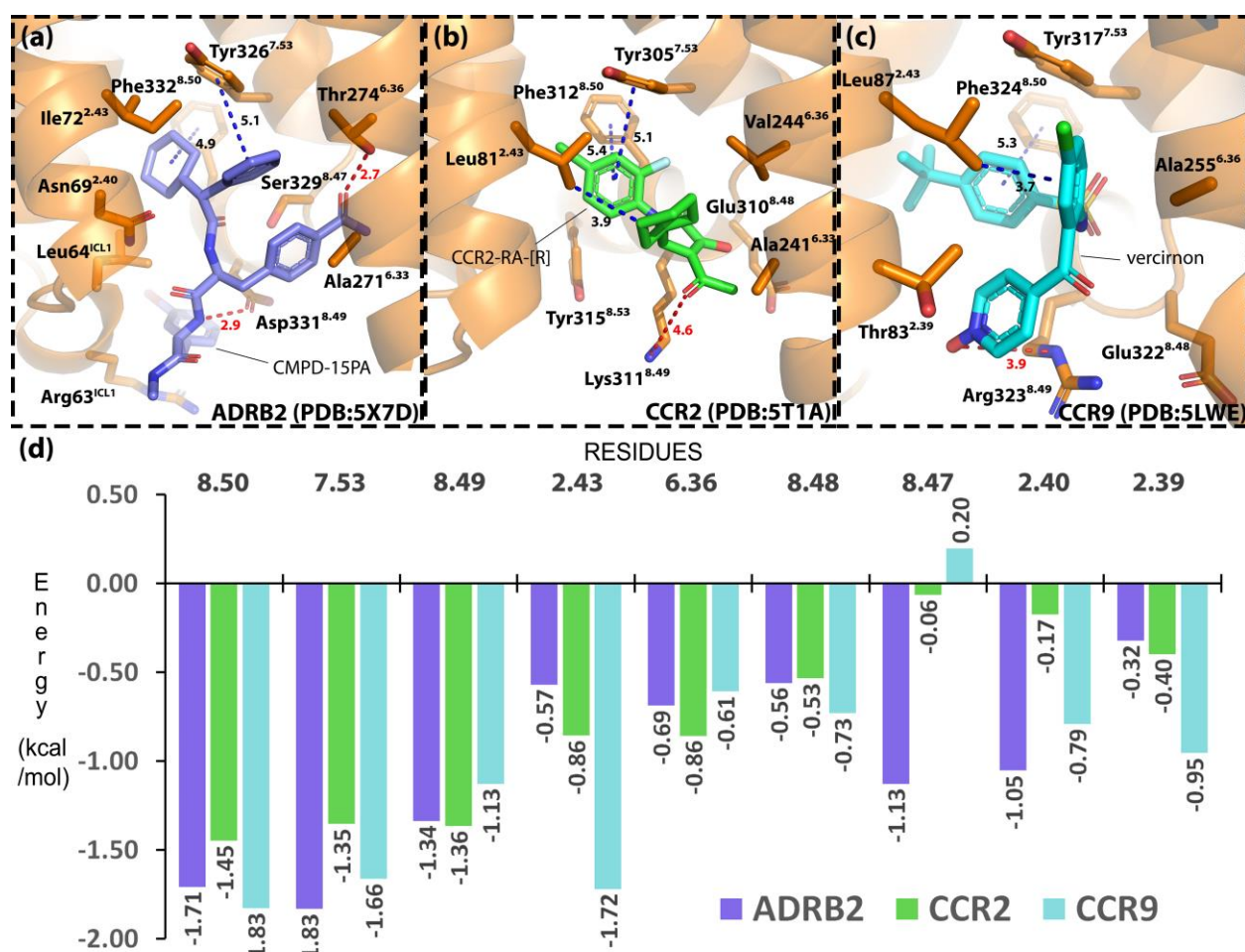


Figure 21 Interaction pattern in allosteric binding Site C.

Table 3 Key residues involved in the Site C binding site are conserved

	2.39	2.40	2.43	6.36	7.53	8.47	8.48	8.49	8.50
ADRB2	T68	N69	I72	T274	Y326	S329	P330	D331	F332
DRD2	T69	N70	I73	M374	Y426	N430	I431	E432	F433
DRD3	T64	N65	V68	M330	Y383	N387	I388	E389	F390
DRD4	T69	N70	I73	V395	Y448	N452	A453	E454	F455
ACM1	N60	N61	L64	T366	Y418	N422	K423	A424	F425
ACM2	N58	N59	L62	T388	Y440	N444	A445	T446	F447
ACM4	N67	N68	L71	T401	Y453	N457	A458	T459	F460
AA1R	T44	F45	I48	S235	Y288	I292	Q293	K294	F295
AA2AR	T41	N42	V45	S234	Y288	I292	R293	E294	F295
P2RY1	I86	S87	M90	L261	Y324	G328	D329	T330	F331
CCR2	T77	D78	L81	V244	Y305	G309	E310	K311	F312
CCR5	T65	D66	L69	L236	Y297	G301	E302	K303	F304
CCR9	T83	D84	L87	V255	Y317	G321	E322	R323	F324
5HT1B	A84	N85	I88	T315	Y369	N373	E374	D375	F376
5HT2A	T109	N110	L113	V324	Y380	N384	K385	T386	Y387
OPRD	T84	N85	I86	M262	Y318	D322	E323	N324	F325
OPRK	T94	N95	I98	L275	Y330	D334	E335	N336	F337
CXCR4	T73	D74	R77	T240	Y302	G306	A307	K308	F309
CB1	S152	Y153	I156	T344	Y397	S401	K402	D403	L404
CB2	S69	Y70	I73	T246	Y299	S303	G304	E305	I306

3.3 Cannabinoid receptor 2

3.3.1 CB2 orthosteric binding pattern

Up to date, both active [45] and inactive [62] CB2 structures have been published and can provide effective agonist and antagonist interaction relationships. To figure out the CB2 ligands binding mechanism enable illuminate the path of CB2 rational drug design. According to Xing, Changrui[45] and Hua, Tian[63], the antagonists extend deeper into the binding pocket compared with agonists, which lead to a $>60^\circ$ clockwise rotation of the ‘toggle switch’ residue Trp258^{6,48}, which is consistent with the above-mentioned results about 6.48 (as shown in Figure 14).

The antagonist-binding pockets in both CB1 and CB2 are varying greatly, which reasonable explained their highly selectivity. However, the agonist-binding pockets of both structures are hard to differentiate. Herein, to quantitatively analyze the interactions between key residues and ligands is necessary for structure-based selective agonist design. As shown in Figure 22, we analyzed and compared the residue energy contributions between active CB1 and CB2. Figure 22a, b detailed the binding interaction between CB1, CB2 with their agonists. Figure 21c mapped out the energy contribution of key residues involved in the binding pockets of active CB1 and CB2. Green bars, calculated energy contributions of key residues based on the active CB1 crystal structure (PDB: 5XRA)[62]; yellow bars, calculated energy contributions of key residues using the active CB2 cryo-EM structure[45]. Most of the key residues contribute similar at two binding pockets, however, four residues, W5.43, F2.57, F2.61, and F3.25, were found to be

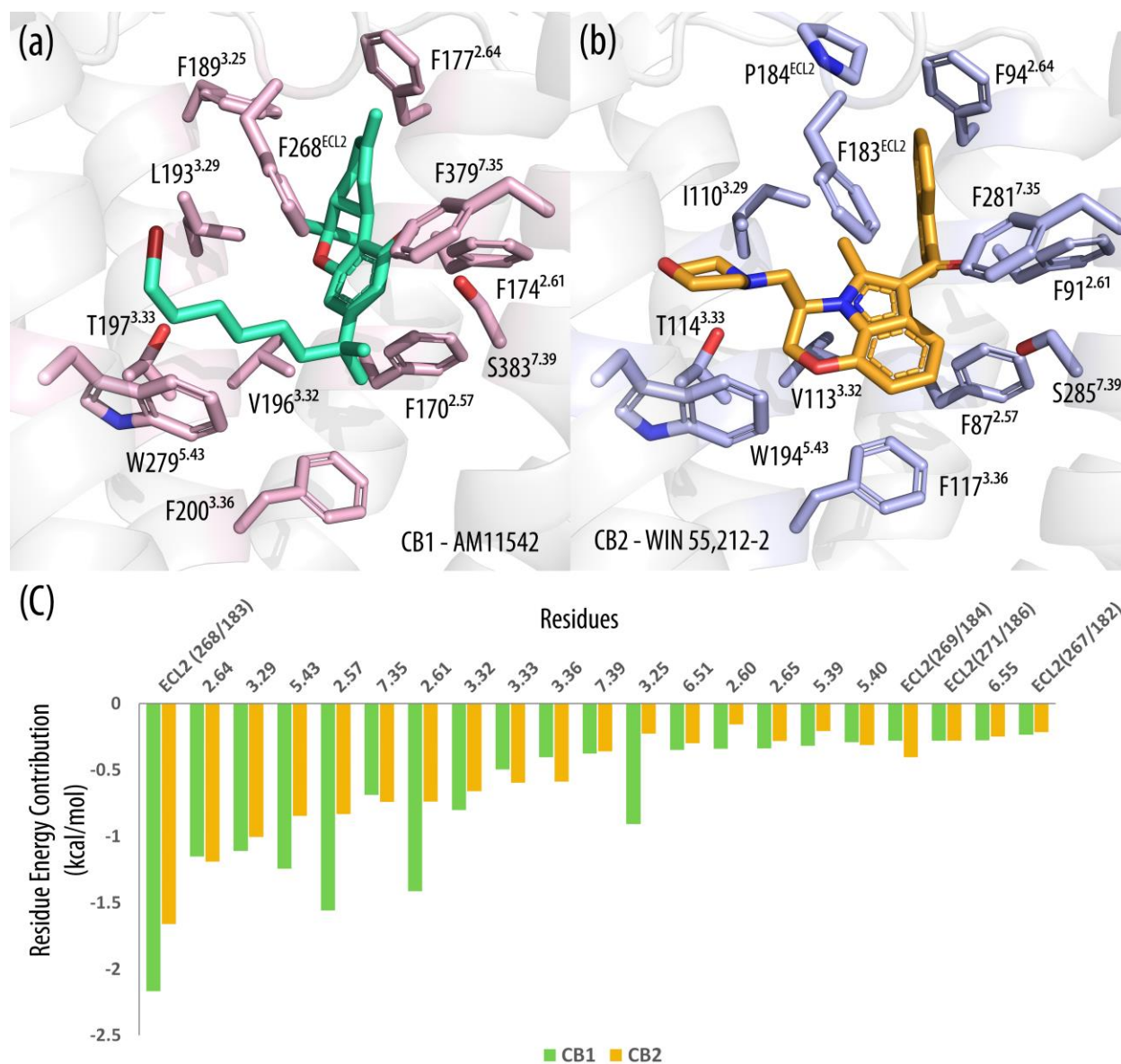


Figure 22 Selectivity pattern between CB1 and CB2 agonists.

potentially contribute to the selectivity of two binding pockets. All four residues are aromatic amino acids that present the most hydrophobic-at-interface in proteins. In this case, all four residues contribute great hydrophobic bonding at CB1 binding pocket (-0.56 kcal/mol, -0.55 kcal/mol, -0.35 kcal/mol, -0.38 kcal/mol, respectively) but not at CB2's (-0.02 kcal/mol, -0.26 kcal/mol, -0.17 kcal/mol, -0.05 kcal/mol, respectively). Therefore, the non-polar functional groups may necessary when designing CB1 highly selective agonists.

3.3.2 CB1/CB2 allosteric binding mode predictions and validations

Till now, the crystal structure of CB1 with Org-27569 has been reported, which revealed a novel extrahelical site that within the inner leaflet of the cell membrane. Org-27569, an indole-containing small molecule, was the first allosteric modulators reported for the CB1 receptor. Unlike other modulators, Org-27569 behaves an unusual pharmacological profile: it increases the binding of CP55940, a potent orthosteric agonist of CB1 receptor; meanwhile it reduces $G_{i/o}$ coupling and signaling, thus favoring an overall inactive state for G protein coupling. Another CB1 allosteric modulator, PSNCBAM-1, displays similar complex effects on CB1 receptor. However, plenty of small molecules have been reported that behave pure PAM effect (ZCZ011, Lipoxin A4, RTI-371, GAT211, JWH007) or NAM effect (Pregnenolone, Cannabidiol, Fenofibrate), indicating a diverse interaction pattern of CB1 allosteric regulation (detail compound information is listed in Table 6). Considering the different pharmacological effect, the pure CB1 PAM and NAM may possess diverse binding mode comparing with the Org-27569.

In addition, CB1 and CB2 are highly similar in both sequence and structure. Many works of literatures reported that they are also similar in allosteric regulation mode, for example, cannabidiol (CBD) is an allosteric modulator of CB1[64], and some studies suggested that CBD

also regulate CB2 in an allosteric manner[65]; pepcan-12 can allosteric regulate both CB1 and CB2[66, 67]; the newly developed CB2 synthetic modulator also positively affect the binding of CB1 with its agonist[68]. Therefore, we propose that CB2 may possess a similar allosteric binding region as CB1. In the virtue of the advantages of allosteric regulation, comprehension of the binding mode of CB1 PAM and NAM may illuminate the rational drug design of CB1 allosteric modulators and provides clues for CB2 allosteric regulation pattern.

Here, we applied several algorithms included CavityPlus, PARS, and Sybyl to identify the potential allosteric binding pockets for CB2, same methods were also used for CB1 to demonstrate the reliability of our study. After scanning the whole 3D structures of CB1 and CB2, as shown in Figure 21 and Figure 22, both of CB1 and CB2 possess ten cavities that potentially able to provide binding site for small molecule allosteric modulators. All of these cavities were analyzed, the related parameters, including predict pKd, drug score, and collection with orthosteric pocket, were listed in Table 4, 5. After excluding cavities with poor ligandability, drugability, or less motion correlation with known orthosteric sites, four sites on CB1 and three sites on CB2 remained, and three of them shared the same topological position (As shown in Figure 23, 24). Taken the similarity of allosteric regulation into consideration, three common binding cavities were selected and used for further analysis in this study (As shown in Figure 25).

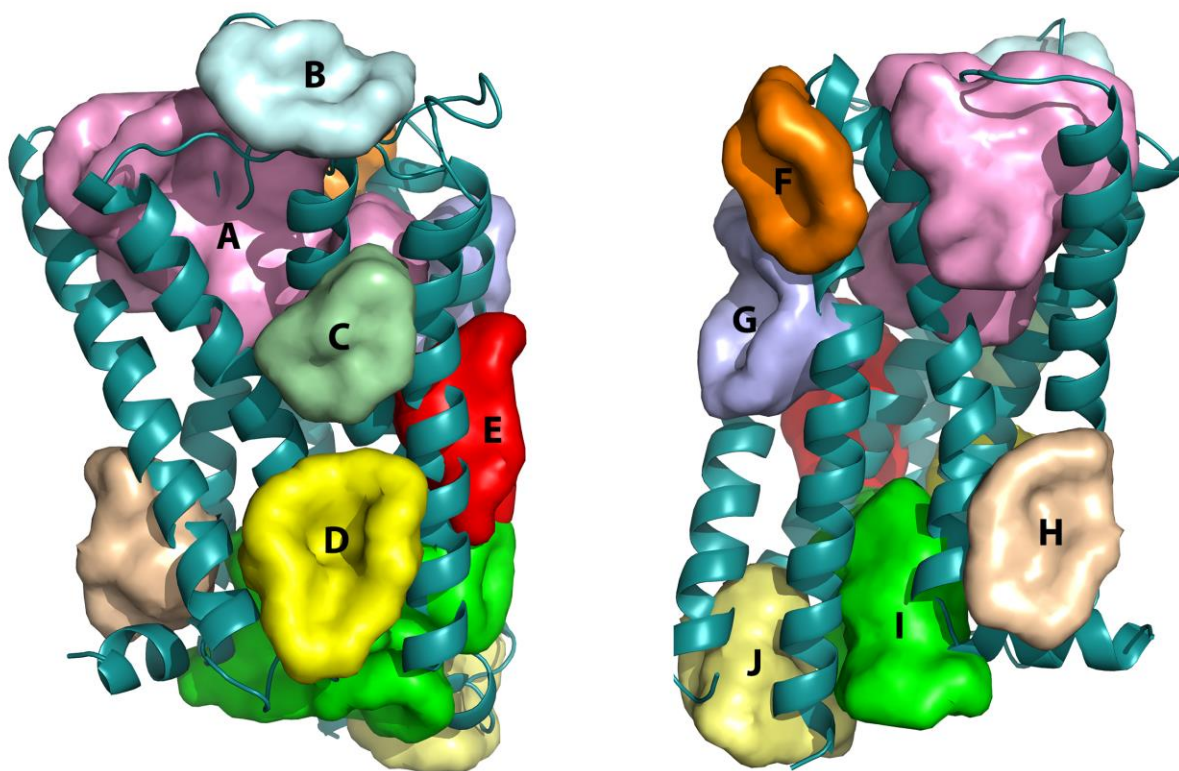


Figure 23 All potential binding cavity of CB1 structure.

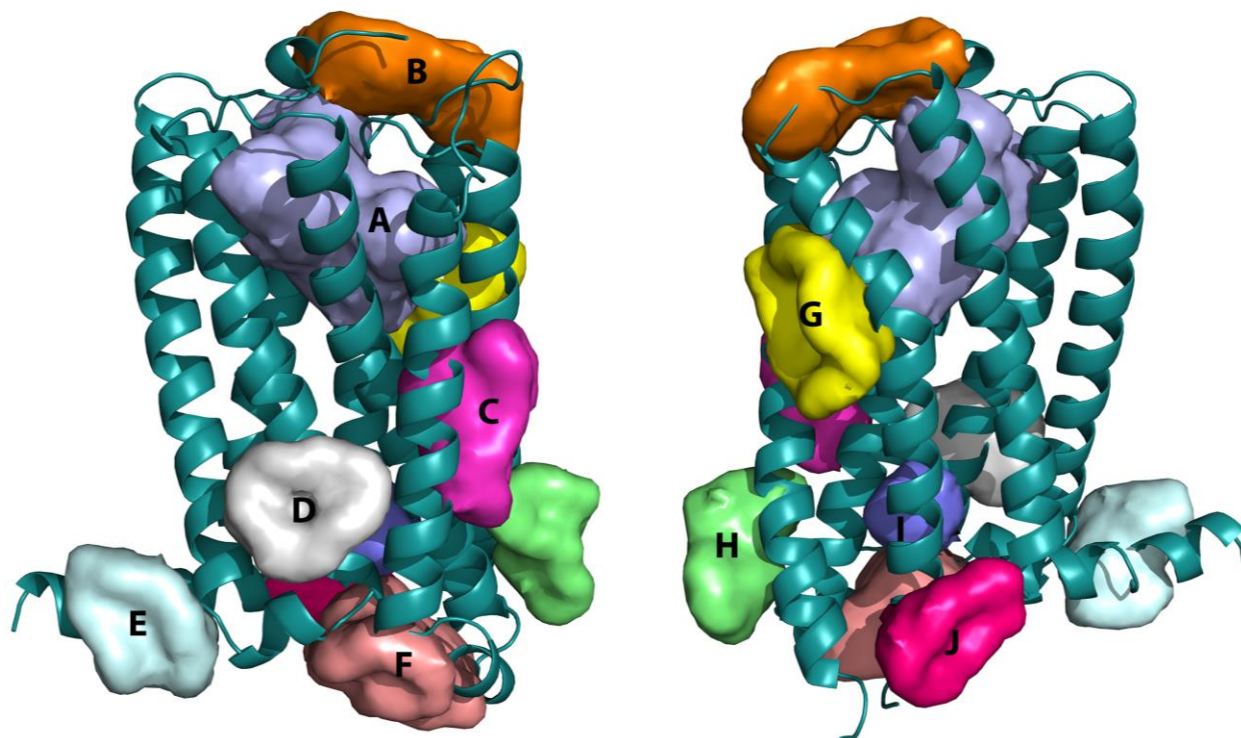


Figure 24 All potential binding cavity of CB2 structure.

Table 4 Parameters of ten potential CB1 binding cavity.

Pocket	Pred. pKd	Drug Score	Allosites
A	6.35	4075	Ortho
H	6.07	296	0.43
D	6.04	228	0.87
E	5.98	672	0.14
I	5.88	1414	2.09
C	5.86	41	-1.56
G	5.76	251	-0.17
J	5.65	-542	0.54
F	4.88	-1006	-1.39
B	4.66	-1168	-1.41

Table 5 Parameters of ten potential CB2 binding cavity.

Pocket	Pred. pKd	Drug Score	Allosites
A	6.95	3745	Ortho
C	6.17	1011	0.66
G	6.08	188	-1.43
F	5.70	-675	2.14
D	5.49	-18	0.15
I	5.02	179	0.68
B	4.78	-1226	-0.95
H	4.61	-1370	-0.24
E	4.60	-1186	-0.45
J	4.57	-1353	-0.56

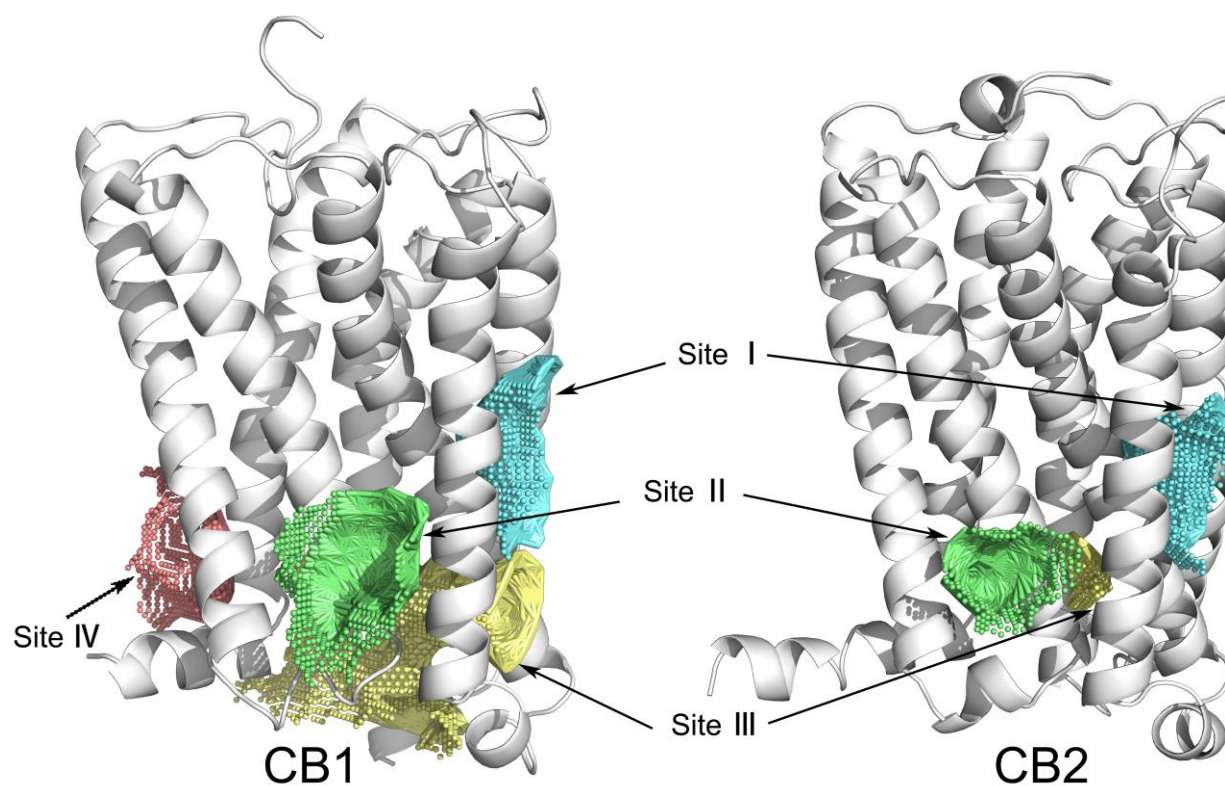
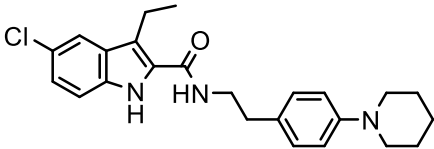
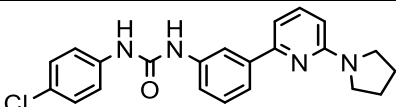
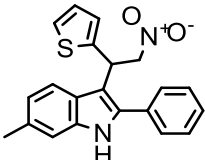
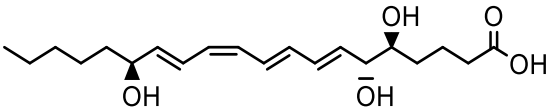
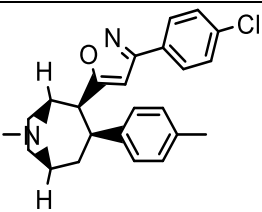
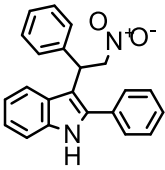
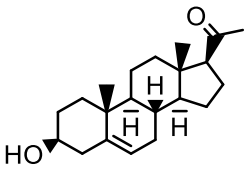
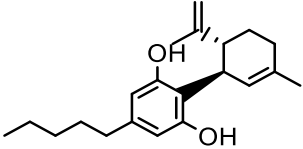


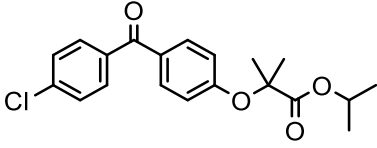
Figure 25 The predicted allosteric binding pocket of CB1 and CB2.

3.3.2 The most possible binding pocket for CB2.

In order to further investigate the allosteric binding pocket of CB1, we docked eight typical allosteric modulators of CB1 (as listed in Table 6) into Site I , Site II , and Site III of CB1.

Table 6 Allosteric modulators of CB1

Name	Effect	LogP	Structure	Reference
Org27569	Regulator	5.27		[69]
PSNCBAM-1	Regulator	5.15		[70]
ZCZ011	PAM	5.184		[71]
Lipoxin A4	PAM	3.03		[72]
RTI-371	PAM	5.64		[73]
GAT211	PAM	5.03		[74]
Pregnenolone	NAM	3.98		[75]
Cannabidiol	NAM	5.91		[64]

Fenofibrate	NAM	4.75		[76]
-------------	-----	------	--	------

According to the cavity analysis, the CB1 and CB2 structure share a similar potential binding cavity, which further supported their similar allosteric regulation mode. Next, docking study was applied to testify which cavity can be bound with the known CB1 and CB2 allosteric modulator.

The predicted Site I is similar to the allosteric binding region of C5AR1 and FFAR1, which owned higher diversity among different receptors. We docked the eight known modulators into Site I, however, no common binding pattern was observed.

Another predicted binding site, Site III, is also a known allosteric binding region for some other class A GPCRs, including ADRB2, CCR2, and CCR9. As stated above, this binding region possesses some highly conserved key residues, such as 8.50 and 7.53 which play an indispensable role in the recognition and interaction between modulators and receptors. Nevertheless, CB1 and CB2, as shown in Table 3, have variant at residue 8.50, which obstruct the formation of stable and consistent interaction with the known modulators based on docking study.

The last predicted common binding site, Site II, is the binding cavity of org27569 but without reported 3D structure among all other class A GPCRs. Because half of this site was surrounded by cell membrane, the simple docking algorithms can't mimic this kind system, glide constraint docking was applied in this situation, and human visual evaluation was followed. Surprisingly, after the docking study, a clear and consistent binding pattern was noted for the eight known modulators classified by three kinds of action. We also docked org27569 back to Site II

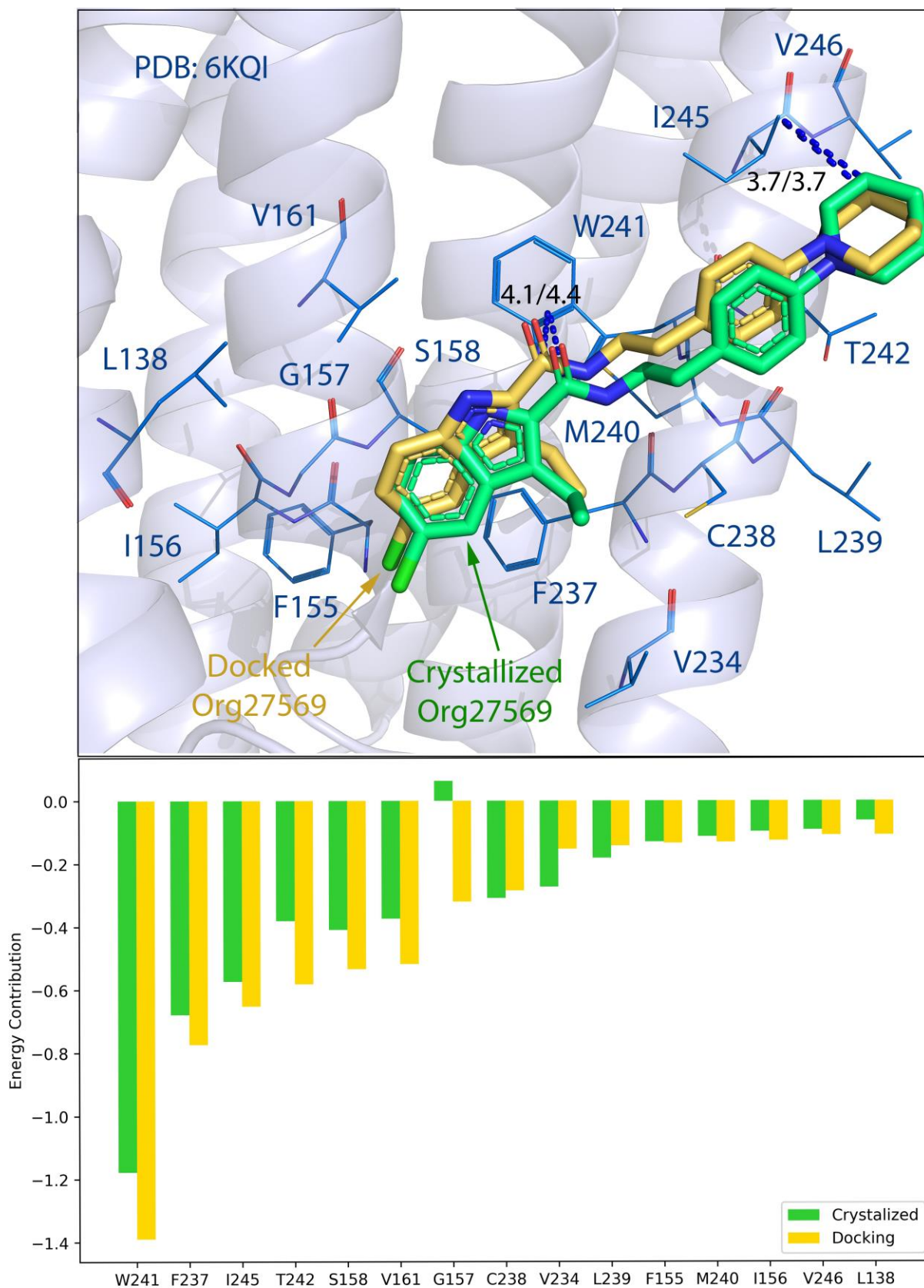


Figure 26 Alignment of crystallized and docked org27569

was only 1.1406Å, which demonstrated that this docking method enable provide reliable docking result and provide effective interact relationship. Besides, the docked ligands formed similar interaction with key residues involved in this binding region, for example, Trp241^{4.50} can form strong hydrophobic interaction with modulators (with distance 4.1 and 4.4 Å), which is responsible for the majority of small-molecule modulators recognition and binding site interaction. Besides Trp241^{4.50}, other key residues, including Phe237^{4.46}, Ile245^{4.54}, Tyr242^{4.51}, Ser158^{2.45}, Val161^{2.48}, Gly157^{2.44}, and Cys238^{4.47} also contributed on increasing stability of the binding conformation of the small molecules. Moreover, after analyzing the residue energy contribution, two distinguishable binding patterns can be found for PAM and NAM. The end of NAMs structure can extend deeply and wedge into the middle of TM1, TM2, and intracellular loop 1, which distinguish themselves from the NAMs. Unlike NAMs, PAMs had shorter structure but stretch out a side group, which forms interactions with residues at the external side of TM4 (as shown in Figure 27, 28).

In that both CB1 and CB2 are transmembrane receptor, the predicted Site II is surrounded by cell membrane that block most of small molecules get into this binding site, only small hydrophobic molecules can partition into the bilayer and binding at this predicted Site II. In order to explore if CB1 modulators can exert effect at this predicted Site II, most of the representative CB1 modulators' log P were listed in Table 6, all of them shows high log P value. Accordingly, these representative modulators possess ability to partitioning into the bilayer and bind at the predicted Site II. Therefore, we hypothesized that Site II is the highly possible allosteric binding pocket of CB1.

Table 7 Conservation of W4.50 among class-A GPCRs.

	1.57	2.40	2.41	2.44	2.45	2.48	4.46	4.47	4.50	4.51
CB1	I141	Y153	H154	G157	S158	V161	F237	C238	W241	T242
CB2	I58	Y70	L71	G74	S75	G78	L154	G155	W158	V159
D2R	V59	N70	N71	V74	S75	V78	I156	S157	W160	V161
D3R	V54	N65	N66	V69	S70	V73	I154	T155	W158	V159
M1	F50	N61	Y62	L65	S66	C69	I146	G147	W150	L151
M2	I48	N59	Y60	F63	S64	C67	I144	A145	W148	V149
CCR1	L59	S70	I71	I74	N75	I78	S154	I155	W158	A159
CCR5	L55	D66	I67	L70	N71	I74	S149	V150	W153	V154
5HT1A	I61	N72	Y73	G76	S77	V80	I157	S158	W161	L162
5HT1B	V74	N85	Y86	A89	S90	V93	I170	A171	W174	V175
A1A	V34	F45	C46	V49	S50	V53	I128	A129	W132	I133
A2A	V31	N42	N43	V46	S47	A50	I125	A126	W129	V130
μ OR	I95	N106	I107	F110	N111	L114	N190	V191	W194	I195
δ OR	I74	N85	I86	F89	N90	L93	N169	I170	W173	V174
κ OR	I84	N95	I96	F99	N100	L103	N179	I180	W183	L184

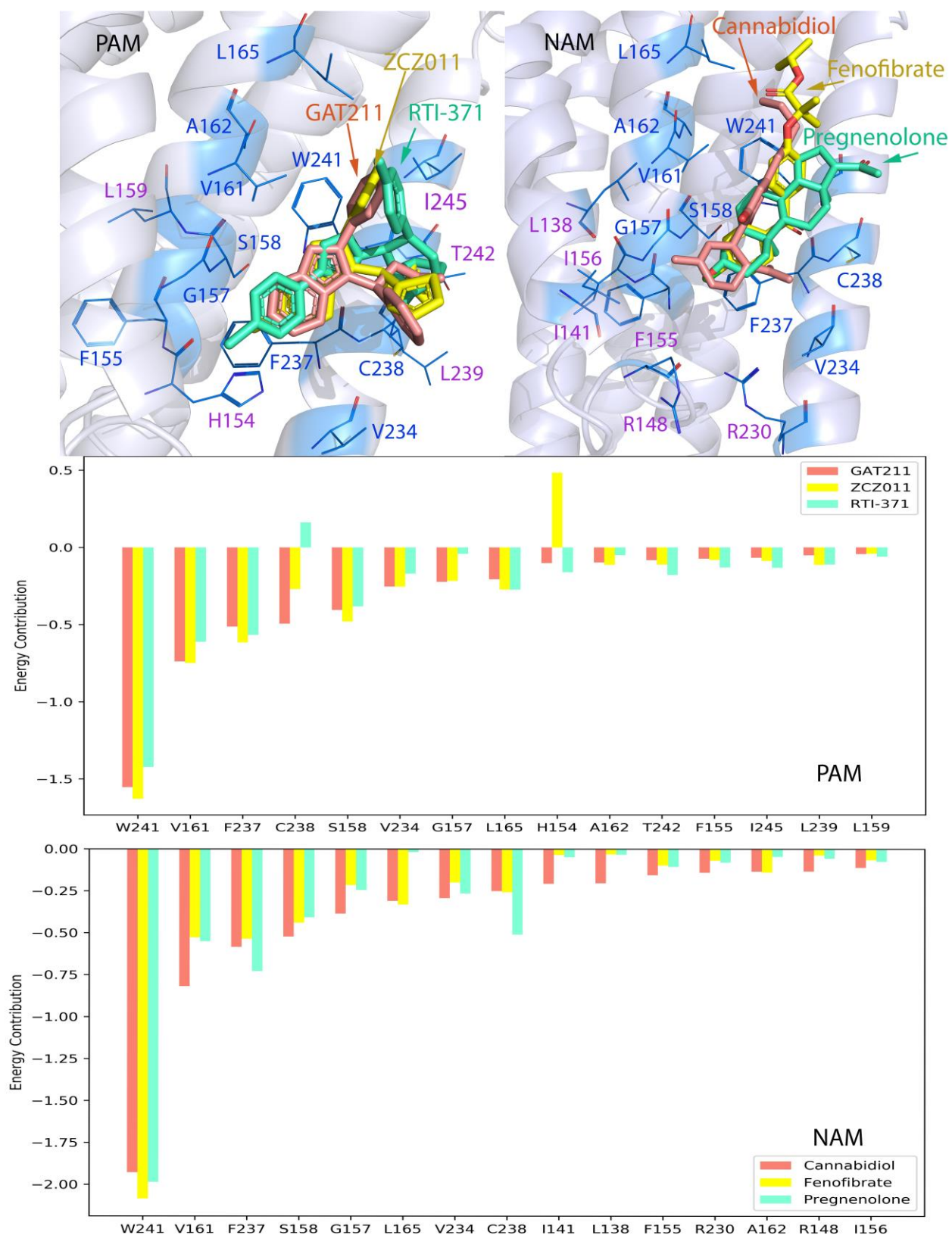


Figure 27 Detail interaction and energy decomposition of PAM and NAM with CB1.

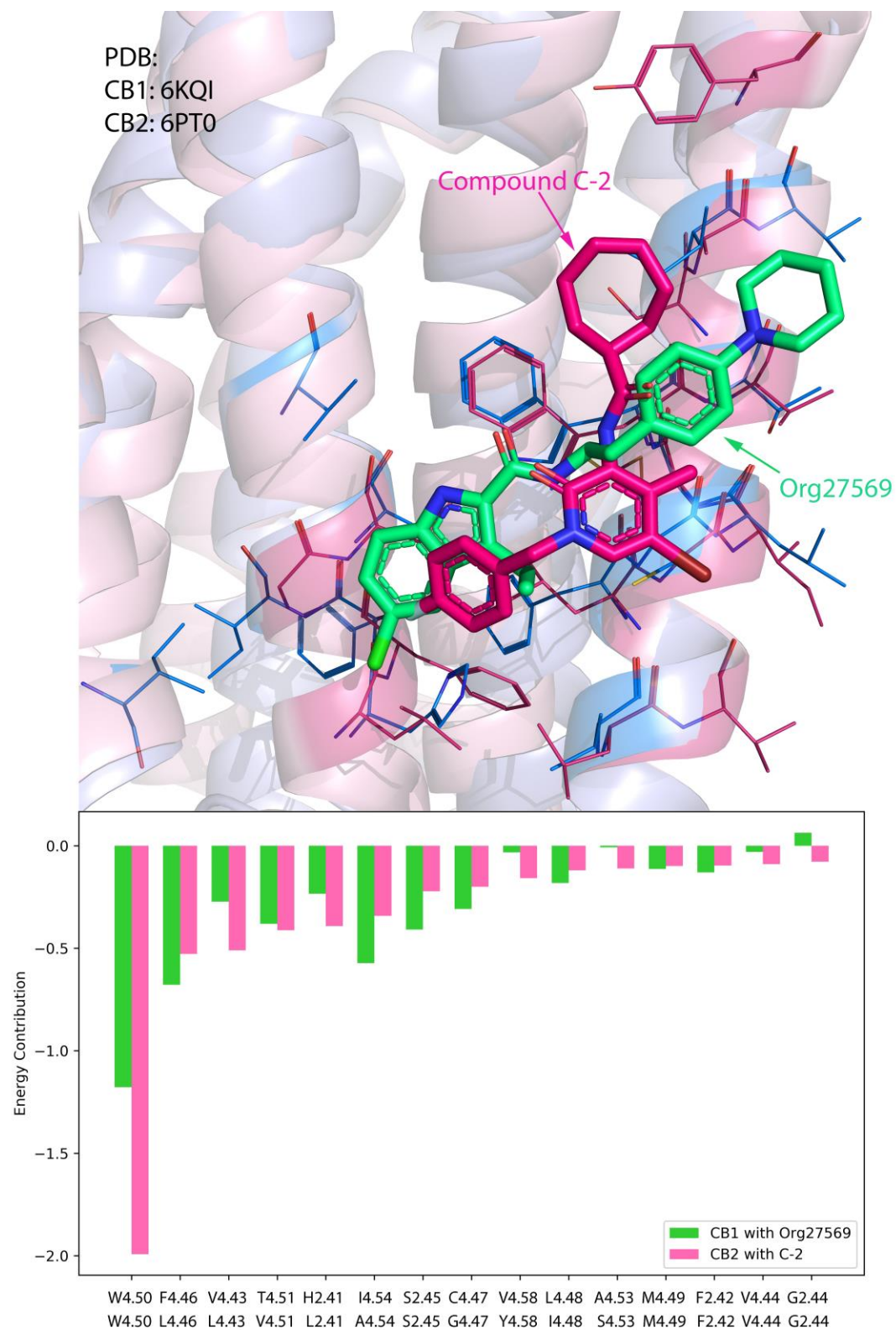


Figure 28 Comparison of docking pose and energy decomposition between CB1 and CB2 modulator.

To date, Ec21a, also called C-2[68], is the only one reported CB2 allosteric modulator. Based on the previous regular pattern that we summarized from CB1 and its allosteric modulators, we docked the known modulator back to CB2. During the binding of ligand and CB1, their recognition pattern is similar as CB1's. However, the Ec21a can form more stable interaction with residues at TM4, such as W4.50, L4.43, Y4.58, and S4.53. Those difference mostly result from the variation of many residues involved in this region between CB1 and CB2, even though the most important key residues, W4.50, kept most conservative role among class A GPCRs. Those residue variations may provide insight for the future development of novel allosteric modulators with high CB1/CB2 selectivity to achieve specificity and fewer side effects.

3.3.3 Validate the CB1 allosteric binding prediction using Molecular dynamic simulations.

The model of CB1/CB2 allosteric binding cavity was proposed through the molecular docking studies in a static manner. But the docking study can not mimic a real binding environment as half of this cavity was surrounded by cell membrane. This bilipid membrane which can not build model by static docking study renders difficulties for accurately binding pose predicting. Therefore, MD simulation was further performed in order to get the stable binding modes of the modulators with CB1/CB2 proposed allosteric binding cavity. Fluctuations of small molecular modulators-receptor complex conformation are described by the RMSD ~ Simulation Time (RST) plots. For the receptors, only mainchain atoms within 7TM domain were considered for least-square (LS) fitting; while for small molecules, all heavy atoms were considered. To measure the dynamics of a ligand in the binding pocket, non-LS fitting RMSDs were calculated for the ligand after the receptor was aligned to the reference structure. After MD simulation, snapshots will be captured every 10-picosecond as record. Next, an average conformation during the whole sampling

phase will be generated, then the coordinate of a snapshot that has lowest RMSD with average conformation will be selected as representative binding conformation for studying the potential interactions and the detailed binding mode.

For the sake of demonstrating the reliability of our docking protocol and to testify the MD simulation system and parameters, the CB1 complexed docking Org27569 conformation was firstly applied a 150nm MD simulation. The CB1-org27569 complex system was first undergo a - 15ns equilibration, then kept stable afterwards. Both the CB1 (7TM) domain and LS fitting ligand RMSD stay around 1Å, while non-LS fitting RMSDs were always larger than the corresponding LS fitting RMSDs, as a result of the non-LS RMSDs takes not only the conformational change, but also the small molecule's translational and rotational movements into account, but the LS RMSDs only consider the conformational change. Unlike classical binding site, this allosteric binding site shaped wide and shallow, and the surrounded bilipid membrane fluctuated more than the residues, all of these lead to reasonable violent movements of the small molecule. Since the org27569 binding site has been testified by crystallized structure, we can conclude that this kind of fluctuation is acceptable for this special binding environment as long as the small molecule stay in this binding site. From Figure 29a, during the MD simulation, the org27569 wedged deeper into the cavity, and form more interactions with residues at TM1 and TM2. The indole ring and the connected ethyl group was flipped upward during the MD simulation. The representative conformation was shown in Figure 29b, the important residues for this binding pose were labeled by lines, hydrophobic residues were colored by marine, polar residues were colored by green, and the charged residues was colored by salmon. Apparently, majority of important residues are hydrophobic, which composed a strong hydrophobic interaction surface.

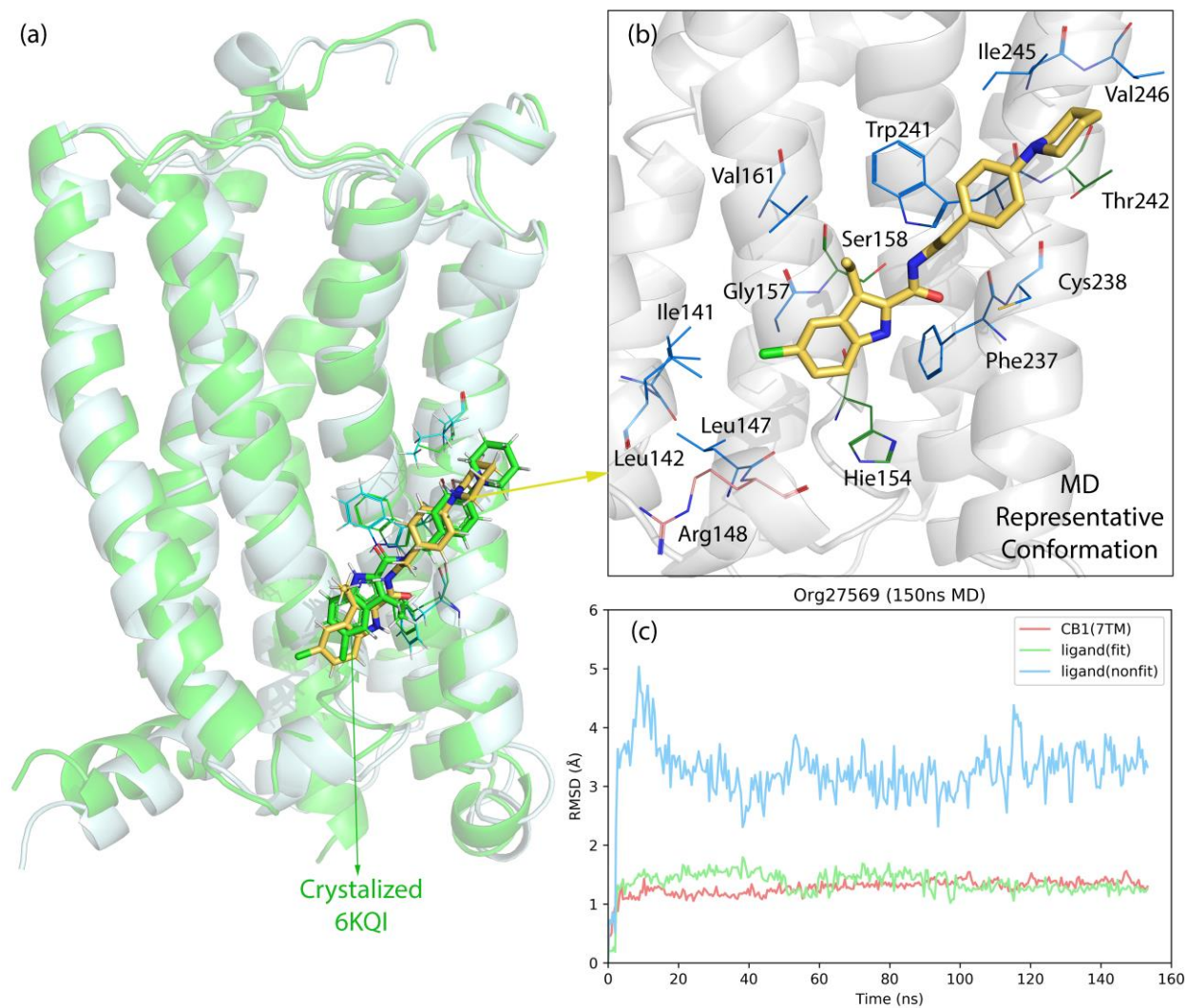


Figure 29 MD simulation of CB1-Org27569 complex.

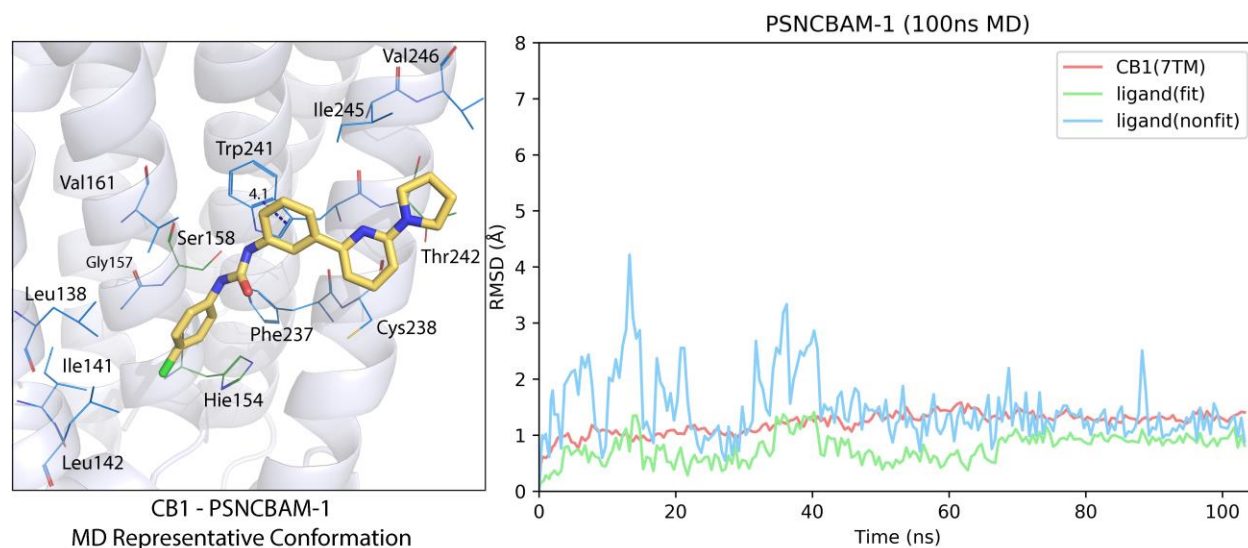


Figure 30 MD simulation for PSNCBAM-1.

Another CB1 regulator, PSNCBAM-1, displays similar pharmacological profiles as Org27569. It increases the binding of CB1 agonists, while reduces the Gi signaling pathway, overall, it negatively regulates CB1 activation. As shown in Figure 30, during the MD simulation, the complex system equilibrated about 40 ns, then kept stable afterwards. An average structure was generated from 40ns to 100ns as the sampling phase. The coordinate that most close to the average structure was selected as representative conformation for detail interaction study. During the MD simulation, the benzene ring in the middle of PSNCBAM-1 structure switch to face the indole group of Trp241, formed a face to face pi-pi stacking, which favors a more stable conformation for this complex.

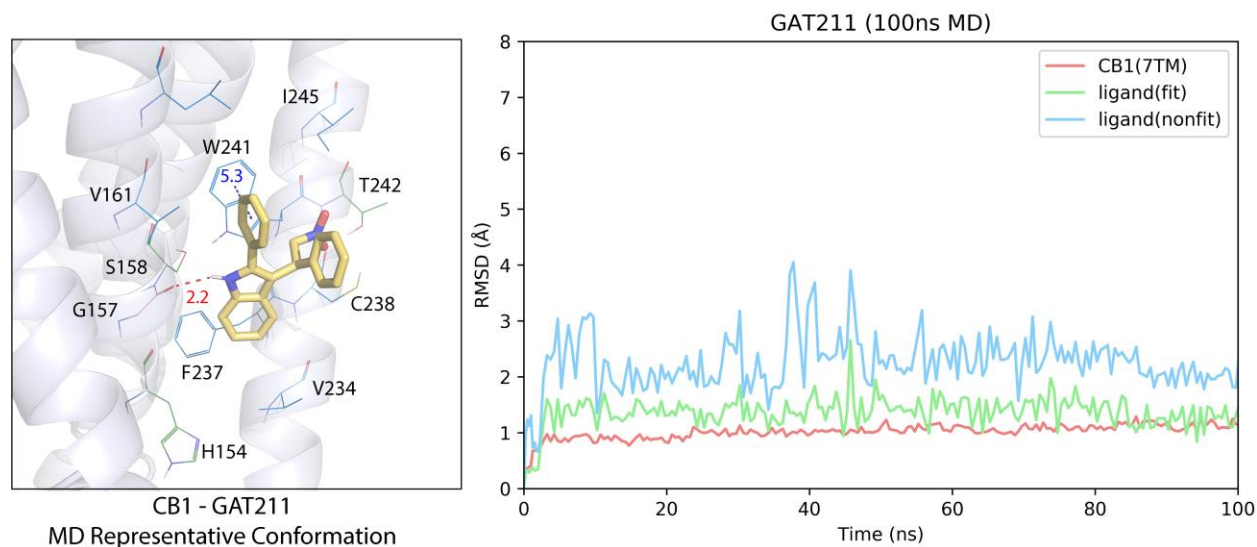


Figure 31 MD simulation for GAT211.

Four CB1 positive allosteric modulator were performed MD simulation. Taken GAT211 as an example, during the MD simulation, even though it underwent large translational or conformational movements, but it always stayed inside this binding site. Even though it faced a huge fluctuation at around 40ns, it equilibrated back to the stable conformation after 50ns. Look into the representative conformation, its benzene ring formed a face to edge pi-pi stacking, and its nitrogen atom in the indole ring can form strong hydrogen bond with the hydroxyl group of Ser158, an important polar residue involved in this binding region (Figure 31). Other three CB1 PAM molecules, Lipoxin A4, RTI-371, and ZCZ011, also reach stable conformation during MD simulation. All four PAMs formed stable interaction with residues S158, V161, V234, F237, C238, W241, and I245 (Figure 32, 33, and 34).

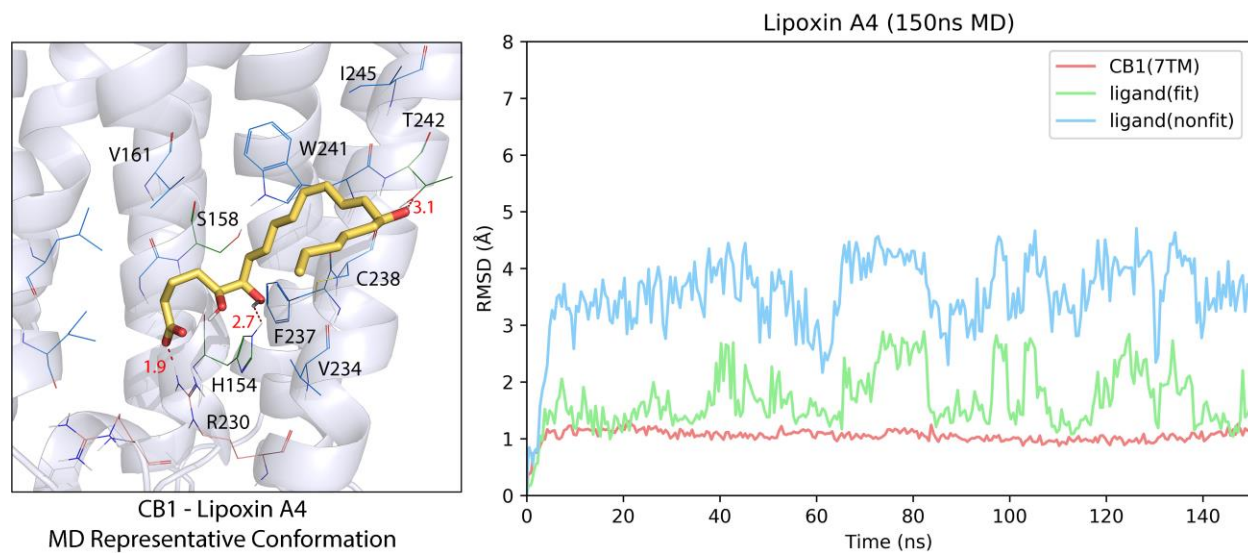


Figure 32 MD simulation for Lipoxin A4.

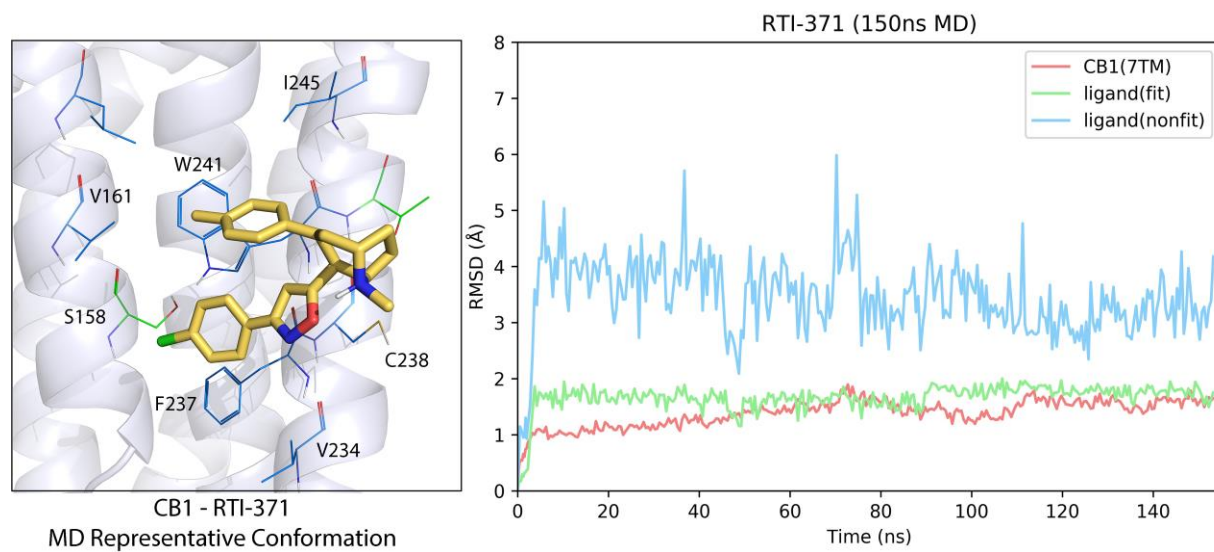


Figure 33 MD simulation for RTI-371.

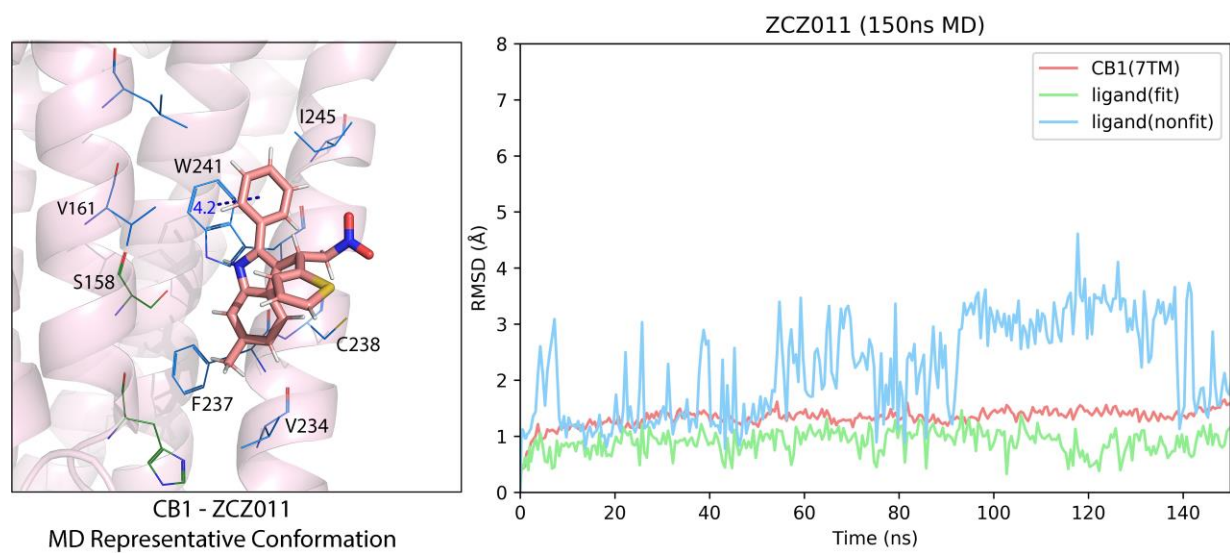


Figure 34 MD simulation for ZCZ011.

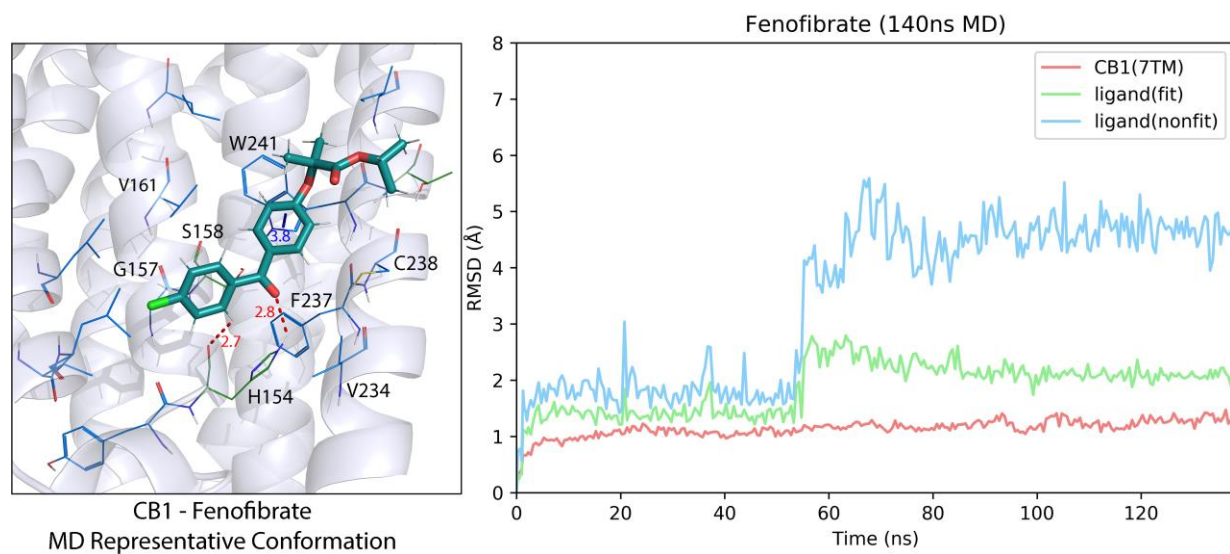


Figure 35 MD simulation for Fenofibrate.

Three CB1 negative allosteric modulators were also performed MD simulation. Taken Fenofibrate as an example, it reached a stable conformation with RMSD around 2 Å at beginning, however, at 55ns – 60ns its chlorobenzene group wedged deeper between TM1 and TM2, which favoring a face to face pi-pi stacking between the benzene ring of Fenofibrate and the indole ring of W241. Overall, this conformational movement lead to a more stable interaction and this conformation kept stable during the afterward 90ns MD simulation (as shown in Figure 35). The detail MD information of the other two NAMs, Cannabidiol and Pregnenolone, can be found in Figure 36 and Figure 37. The common key residues for NAM binding are H154, G157, S158, V161, V234, F237, C238, and W241. Comparing with PAMs, the NAMs interact more with residues at intracellular side of TM2, thus favoring an inward movement of intracellular side of CB1, which consistent with the CB1 structure inactivation movement.

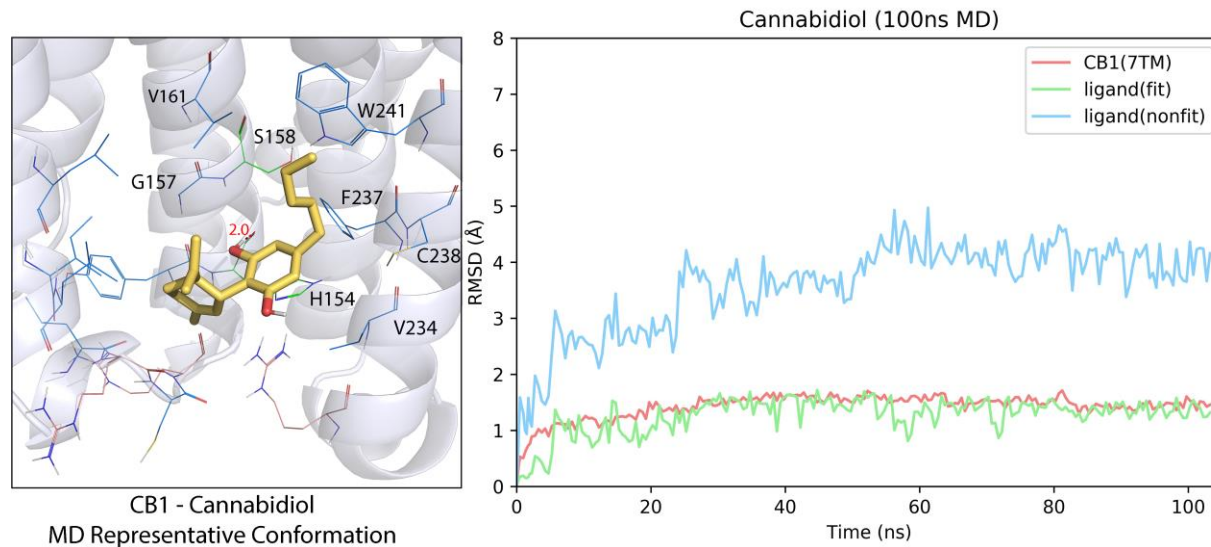


Figure 36 MD simulation of Cannabidiol.

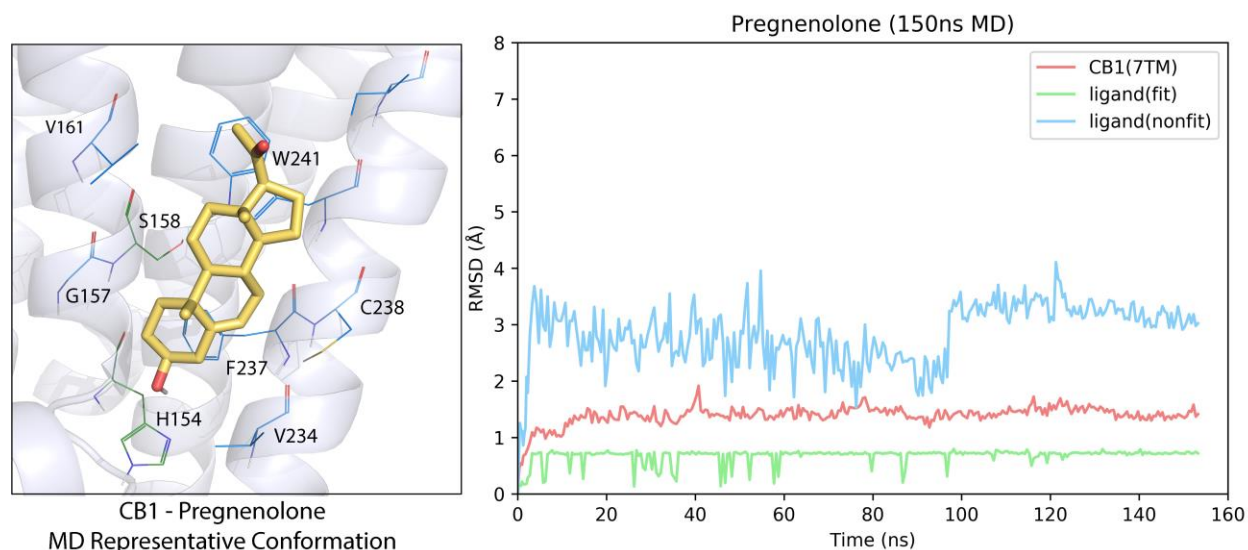


Figure 37 MD simulation of Pregnenolone.

3.3.4 Exploration of the CB2 allosteric regulation using MD simulation.

Till now, there are no CB2-modulator complexed structure have been reported, and Ec2la is the first and only one the human designed, synthesized, and testified CB2 positive allosteric modulator, all of these rendering an opaque mask for CB2 allosteric modulator discovery and rational design. However, some clues, such as the Ec2la also displayed PAM profile for CB1, or Pepcan-12, a peptide that can allosterically regulate CB1, also been demonstrated can exert similar allosteric pharmacological function for CB2, pointed that the CB2 may share a similar allosteric binding pattern as CB1. Herein, the docked Ec2la at the predict binding site, complexed with the first cryo-EM active status CB2 structure were performed MD simulation. In order to see if the Ec2la will impact the interaction between CB2 and its agonist, the Win 55212-2, a selective CB2 agonist also included in this MD system, the initial pose come from the cryo-EM structure. And the CB2-Win 55212-2 without Ec2la complexed was also performed MD simulation as a negative

control. As shown in Figure 38c, the Ec2la endured equilibration about 80 ns, and then kept stable RMSD around 4 Å. After 120 ns MD simulation, the CB2 with Ec2la structure displayed difference, even though they have exactly the same initial conformation. In more detail, the vertical view (Figure 38b) showed that the CB2 with Ec2la structure gathered its extracellular side, which tightened the orthosteric binding pocket and trapped the agonist in this pocket. What's more, the intracellular side also went through a conformational change that favoring a stable activation status, in other word, the intracellular side move outward, which benefit the downward G protein activation (Figure 38a). Accordingly, the MD fluctuation of Win 55212-2 in system with Ec2la showed more stable RMSD ~ time plot comparing with Win 55212-2 only system. In all, the Ec2la can form stable interaction at the predict allosteric binding site and render a favorable influence in both the CB2 activation and CB2 agonists binding.

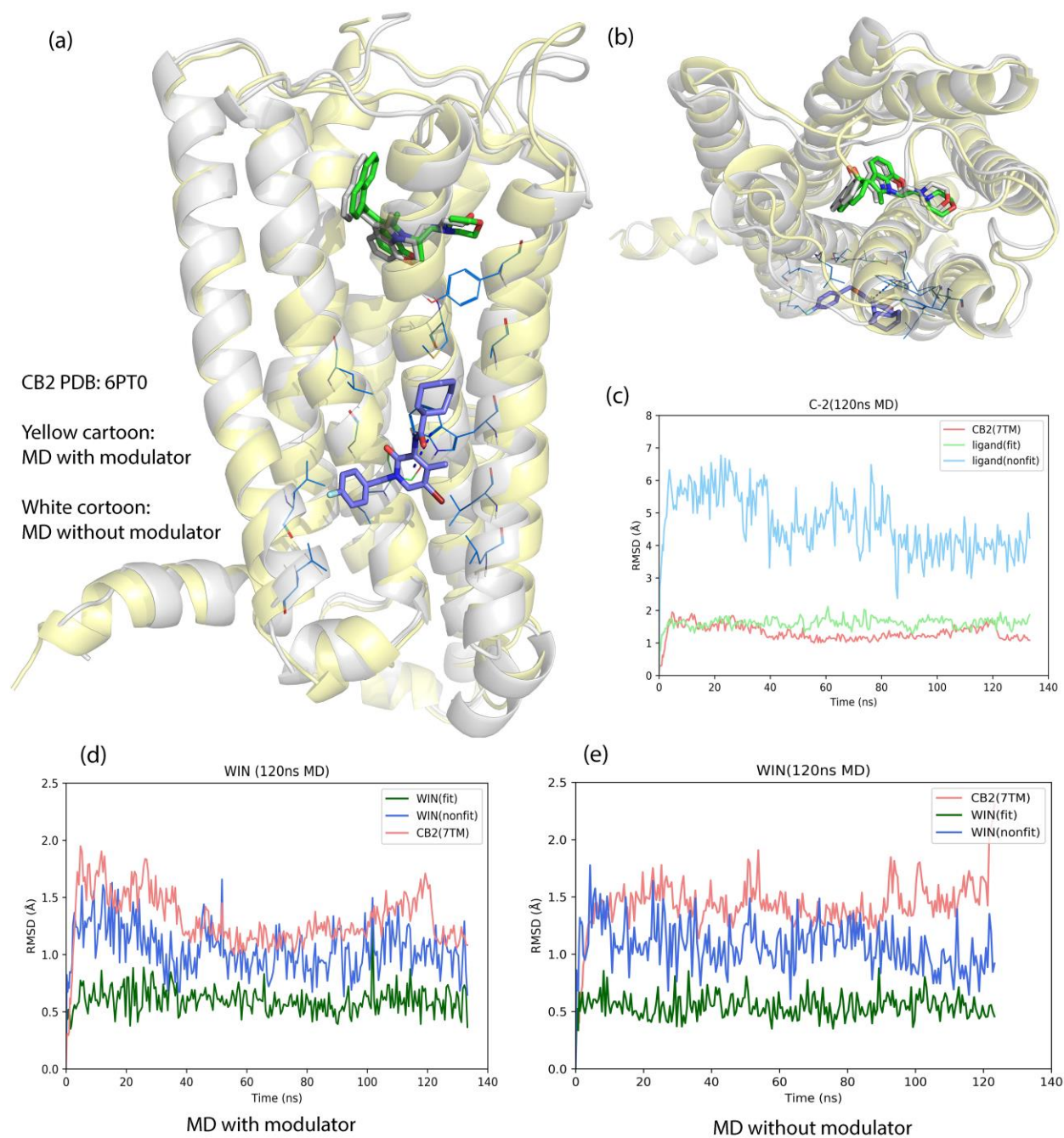


Figure 38 Exploration of the CB2 allosteric regulation using MD simulation.

4.0 SUMMARY AND CONCLUSIONS

Delineating the fingerprint or feature vector of a receptor/protein will facilitate the structural and biological studies, as well as the rational design and development of drugs with high affinities and selectivity.

In the present work, we borrowed the idea of scoring function and designed a method (MCCS) to characterize a protein by totaling per residue scores in a receptor-ligand complex model which contains 3D conformational coordinates about a ligand found at the binding site of a receptor under an electron microscope. To experiment with our idea, we extended idock, revised dedicatedly and eventually developed jdock as a part of the MCCS implementation while remaining the backward compatibility. Unlike other molecular fingerprints e.g. MACCS, ECFP2, etc. which are a sequence of summarized descriptors of small molecules, the method we are proposing is a way to characterize a protein.

The devised “residue energy contribution” term enable to determine how residues actually contribute to the binding of ligand in its receptor. The feasibility and accuracy of residue energy contribution were further supported by our recent work [77]. Using a receptor-ligand dataset of 249 crystallized GPCRs, we sequentially characterized the features of the classical orthosteric binding pockets in GPCRs and produced heatmaps for various classes of GPCRs. Finally, we created a feature vector for each protein for further studies of protein similarity and clustering.

Despite the high similarity of 7TM architecture structure and ligand binding pocket of Class-A GPCRs, their allosteric modulators can function in very distinct modes. In the second part, the newly developed MCCS program was applied to systematically analyze and characterized the known typical allosteric binding pockets of class-A GPCRs. Furthermore, based on the result of

docking study and MD simulation, a potential allosteric modulator binding mode of CB1/CB2 is revealed in this work. It is conceded that the MCCA algorithms still have far way to go for analyzing and simulate some complicated environment, but the

In summary, the idea to quantitatively analyze the residue energy contribution can help for recognition and interaction pattern conclusion of small-molecule modulators by receptor proteins to characterize the binding regions. All the information provides new insight into the protein features and will facilitate the characterization of GPCR.

APPENDIX

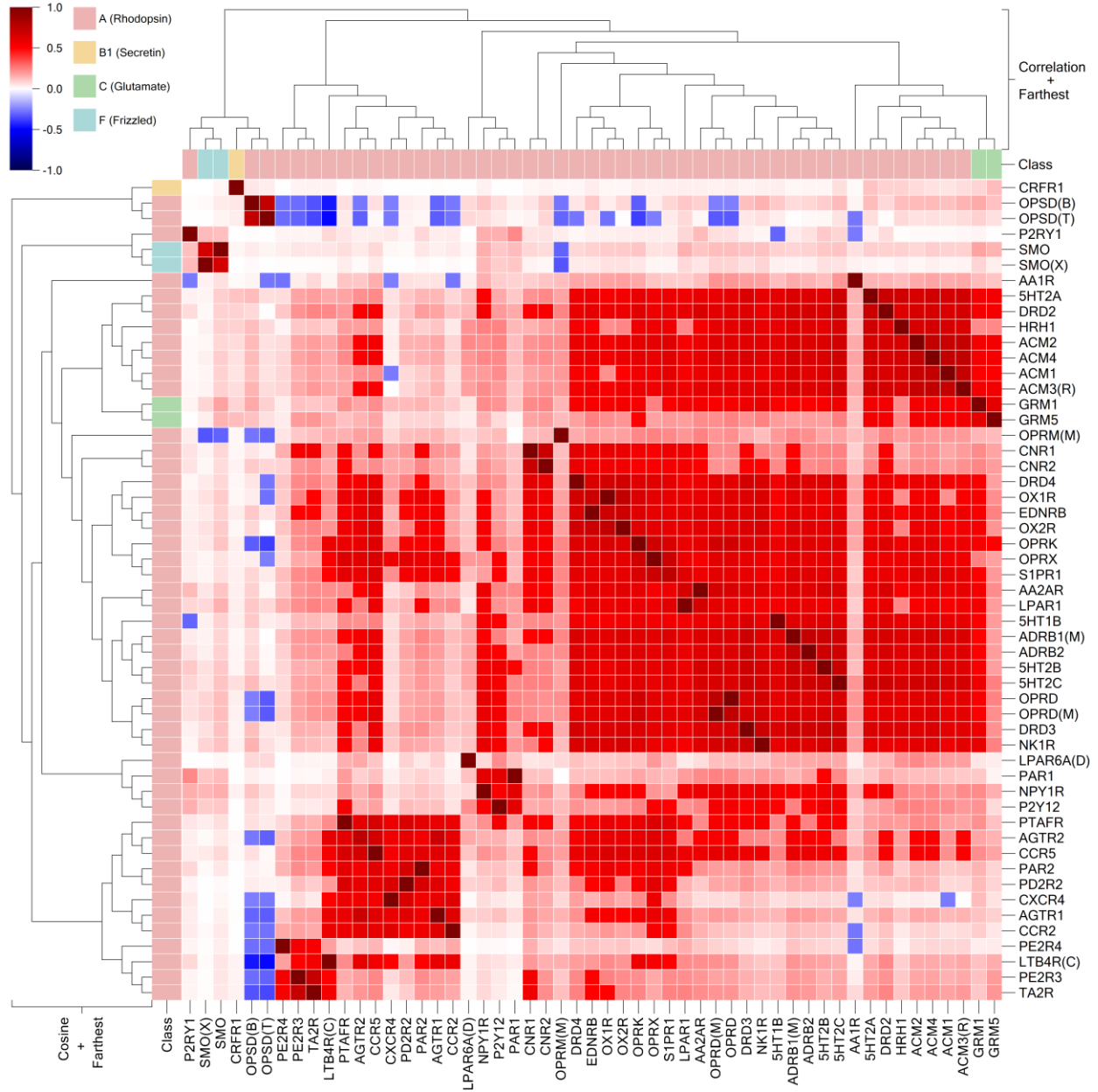


Figure 39 Correlation between score vectors of GPCRs calculated using Cosine Similarity

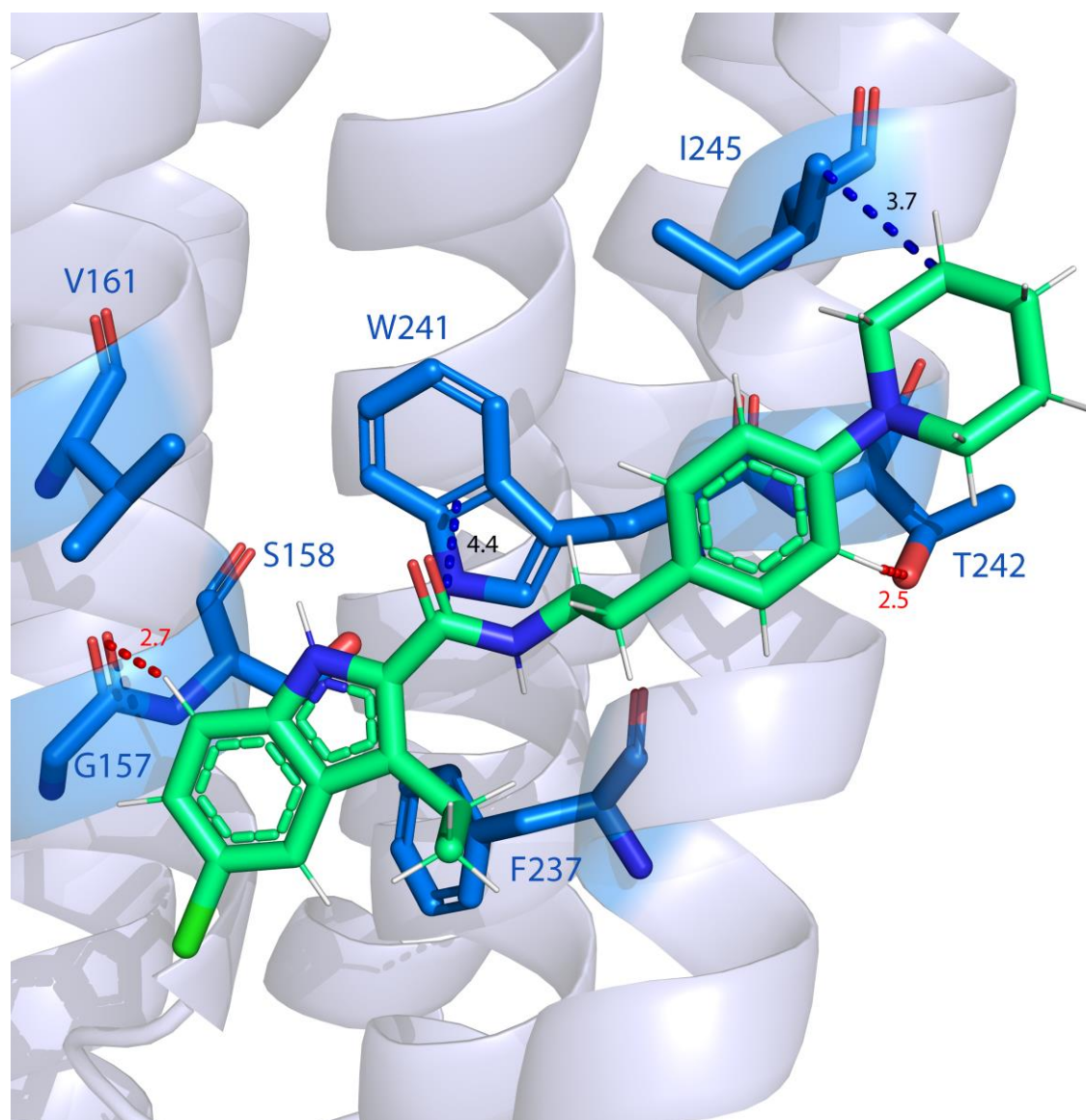


Figure 40 Detail binding interaction of crystallized CB1-ORG2759 complex

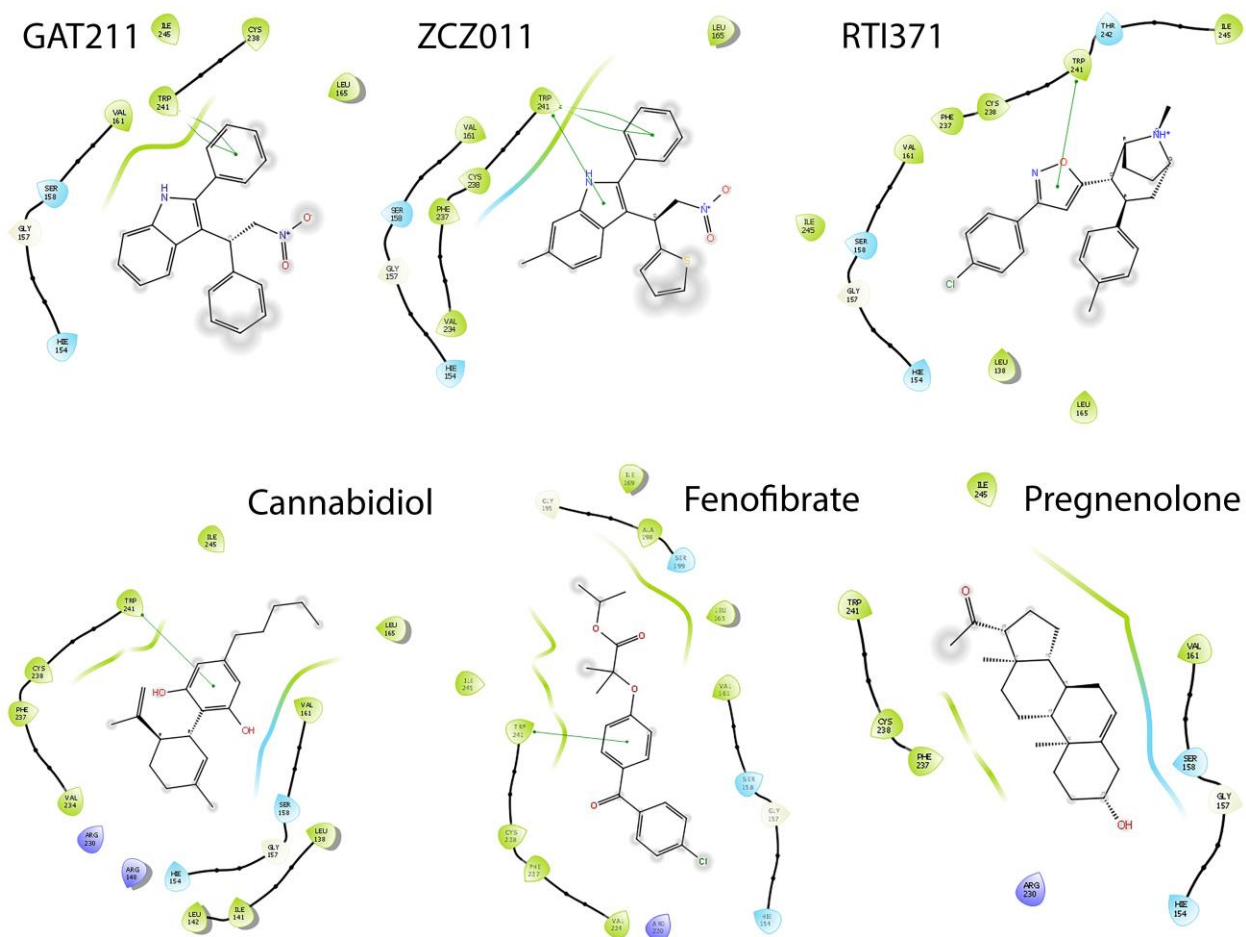


Figure 41 Ligand interaction diagram of PAM and NAM with CB1.

BIBLIOGRAPHY

1. Stevens, R.C., et al., *The GPCR Network: a large-scale collaboration to determine human GPCR structure and function*. Nature Reviews Drug Discovery, 2013. **12**(1): p. 25-34.
2. Lu, M. and B. Wu, *Structural studies of G protein-coupled receptors*. IUBMB Life, 2016. **68**(11): p. 894-903.
3. Bockaert, J., *Molecular tinkering of G protein-coupled receptors: an evolutionary success*. The EMBO Journal, 1999. **18**(7): p. 1723-1729.
4. Hamm, H.E., *The many faces of G protein signaling*. J Biol Chem, 1998. **273**(2): p. 669-72.
5. Tuteja, N., *Signaling through G protein coupled receptors*. Plant Signaling & Behavior, 2009. **4**(10): p. 942-947.
6. Hauser, A.S., et al., *Trends in GPCR drug discovery: new agents, targets and indications*. Nature Reviews Drug Discovery, 2017. **16**(12): p. 829-842.
7. Kolakowski, L.F., *GCRDb: a G-protein-coupled receptor database*. Receptors & channels, 1994. **2**(1): p. 1-7.
8. Granier, S. and B. Kobilka, *A new era of GPCR structural and chemical biology*. Nature Chemical Biology, 2012. **8**(8): p. 670-673.
9. Palczewski, K., et al., *Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor*. Science, 2000. **289**(5480): p. 739-745.
10. Wacker, D., R.C. Stevens, and B.L. Roth, *How Ligands Illuminate GPCR Molecular Pharmacology*. Cell, 2017. **170**(3): p. 414-427.
11. Bridges, T.M. and C.W. Lindsley, *G-Protein-Coupled Receptors: From Classical Modes of Modulation to Allosteric Mechanisms*. ACS Chemical Biology, 2008. **3**(9): p. 530-541.
12. Aizpurua-Olaizola, O., et al., *Targeting the endocannabinoid system: future therapeutic strategies*. Drug Discov Today, 2017. **22**(1): p. 105-110.
13. Munro, S., K.L. Thomas, and M. Abu-Shaar, *Molecular characterization of a peripheral receptor for cannabinoids*. Nature, 1993. **365**(6441): p. 61-65.
14. Wilson, R.I. and R.A. Nicoll, *Endogenous cannabinoids mediate retrograde signalling at hippocampal synapses*. Nature, 2001. **410**(6828): p. 588-592.
15. Basu, S., A. Ray, and B.N. Dittel, *Cannabinoid Receptor 2 Is Critical for the Homing and Retention of Marginal Zone B Lineage Cells and for Efficient T-Independent Immune Responses*. The Journal of Immunology, 2011. **187**(11): p. 5720-5732.
16. Ballesteros, J.A. and H. Weinstein, *[19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors*, in *Methods in neurosciences*. 1995, Elsevier. p. 366-428.
17. Wootten, D., et al., *Polar transmembrane interactions drive formation of ligand-specific and signal pathway-biased family B G protein-coupled receptor conformations*. Proceedings of the National Academy of Sciences, 2013. **110**(13): p. 5211-5216.
18. Pin, J.-P., T. Galvez, and L. Prézeau, *Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors*. Pharmacology & Therapeutics, 2003. **98**(3): p. 325-354.

19. Wang, C., et al., *Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs*. Nature Communications, 2014. **5**(1): p. 4355.
20. Axen, S.D., et al., *A simple representation of three-dimensional molecular structure*. Journal of medicinal chemistry, 2017. **60**(17): p. 7393-7409.
21. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading*. Journal of computational chemistry, 2010. **31**(2): p. 455-461.
22. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility*. Journal of computational chemistry, 2009. **30**(16): p. 2785-2791.
23. Friesner, R.A., et al., *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy*. Journal of medicinal chemistry, 2004. **47**(7): p. 1739-1749.
24. Jain, A.N., *Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search*. Journal of computer-aided molecular design, 2007. **21**(5): p. 281-306.
25. Kramer, B., M. Rarey, and T. Lengauer, *Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking*. Proteins: Structure, Function, Bioinformatics, 1999. **37**(2): p. 228-241.
26. Verdonk, M.L., et al., *Improved protein-ligand docking using GOLD*. Proteins: Structure, Function, Bioinformatics, 2003. **52**(4): p. 609-623.
27. Venkatachalam, C.M., et al., *LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites*. Journal of Molecular Graphics Modelling, 2003. **21**(4): p. 289-307.
28. Ewing, T.J., et al., *DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases*. Journal of computer-aided molecular design, 2001. **15**(5): p. 411-428.
29. Rohl, C.A., et al., *Protein structure prediction using Rosetta*. Methods in enzymology, 2004. **383**: p. 66-93.
30. Koes, D.R., M.P. Baumgartner, and C.J. Camacho, *Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise*. Journal of chemical information modeling, 2013. **53**(8): p. 1893-1904.
31. Li, H., K.-S. Leung, and M.-H. Wong, *idock: A multithreaded virtual screening tool for flexible ligand docking*. 2012 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2012: p. 77-84.
32. Chen, M., et al., *DAKB-GPCRs: An Integrated Computational Platform for Drug Abuse Related GPCRs*. Journal of chemical information modeling, 2019. **59**(4): p. 1283-1289.
33. Lee, J.Y., et al., *Molecular Basis for the Functionality of γ -Secretase Inferred From Structure-Based Modeling and Druggability Simulations*. Biophysical Journal, 2018. **114**(3): p. 232a.
34. Feng, Z., et al., *Modeling, molecular dynamics simulation, and mutation validation for structure of cannabinoid receptor 2 based on known crystal structures of GPCRs*. Journal of chemical information, 2014. **54**(9): p. 2483-2499.
35. Deller, M.C. and B. Rupp, *Models of protein-ligand crystal structures: trust, but verify*. Journal of computer-aided molecular design, 2015. **29**(9): p. 817-836.
36. Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Res, 2000. **28**(1): p. 235-42.

37. Burley, S.K., et al., *RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy*. Nucleic Acids Res, 2019. **47**(D1): p. D464-d474.
38. Pándy-Szekeres, G., et al., *GPCRdb in 2018: adding GPCR structure models and ligands*. Nucleic Acids Research, 2017. **46**(D1): p. D440-D446.
39. Pettersen, E.F., et al., *UCSF Chimera—a visualization system for exploratory research and analysis*. Journal of computational chemistry, 2004. **25**(13): p. 1605-1612.
40. Shapovalov, M.V. and R.L. Dunbrack Jr, *A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions*. Structure, 2011. **19**(6): p. 844-858.
41. Olsson, M.H., et al., *PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions*. Journal of chemical theory, 2011. **7**(2): p. 525-537.
42. Pedretti, A., L. Villa, and G. Vistoli, *VEGA—an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming*. Journal of computer-aided molecular design, 2004. **18**(3): p. 167-173.
43. Repasky, M.P., M. Shelley, and R.A. Friesner, *Flexible Ligand Docking with Glide*. 2007. **18**(1): p. 8.12.1-8.12.36.
44. Shao, Z., et al., *Structure of an allosteric modulator bound to the CB1 cannabinoid receptor*. Nat Chem Biol, 2019. **15**(12): p. 1199-1205.
45. Xing, C., et al., *Cryo-EM Structure of the Human Cannabinoid Receptor CB2-Gi Signaling Complex*. Cell, 2020.
46. Case, D.A., et al., *AMBER 2018*. University of California, San Francisco, 2018.
47. Xu, Y., et al., *CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction*. Nucleic Acids Research, 2018. **46**(W1): p. W374-W379.
48. Panjkovich, A. and X. Daura, *PARS: a web server for the prediction of Protein Allosteric and Regulatory Sites*. Bioinformatics, 2014. **30**(9): p. 1314-1315.
49. Berman, H.M., et al., *The protein data bank*. Nucleic acids research, 2000. **28**(1): p. 235-242.
50. Bouvier, M., *Unraveling the structural basis of GPCR activation and inactivation*. Nature Structural & Molecular Biology, 2013. **20**: p. 539.
51. Rosenbaum, D.M., S.G.F. Rasmussen, and B.K. Kobilka, *The structure and function of G-protein-coupled receptors*. Nature, 2009. **459**(7245): p. 356-363.
52. Chun, L., W.-h. Zhang, and J.-f. Liu, *Structure and ligand recognition of class C GPCRs*. Acta pharmacologica Sinica, 2012. **33**(3): p. 312.
53. Kruse, A.C., et al., *Activation and allosteric modulation of a muscarinic acetylcholine receptor*. Nature, 2013. **504**(7478): p. 101-106.
54. Cheng, R.K.Y., et al., *Structural insight into allosteric modulation of protease-activated receptor 2*. Nature, 2017. **545**(7652): p. 112-115.
55. Robertson, N., et al., *Structure of the complement C5a receptor bound to the extra-helical antagonist NDT9513727*. Nature, 2018. **553**(7686): p. 111-114.
56. Lu, J., et al., *Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40*. Nature Structural & Molecular Biology, 2017. **24**(7): p. 570-577.
57. Zheng, Y., et al., *Structure of CC chemokine receptor 2 with orthosteric and allosteric antagonists*. Nature, 2016. **540**(7633): p. 458-461.

58. Oswald, C., et al., *Intracellular allosteric antagonism of the CCR9 receptor*. Nature, 2016. **540**(7633): p. 462-465.
59. Liu, X., et al., *Mechanism of intracellular allosteric β 2AR antagonist revealed by X-ray crystal structure*. Nature, 2017. **548**(7668): p. 480-484.
60. Srivastava, A., et al., *High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875*. Nature, 2014. **513**(7516): p. 124-127.
61. Zhang, D., et al., *Two disparate ligand-binding sites in the human P2Y1 receptor*. Nature, 2015. **520**(7547): p. 317-321.
62. Li, X., et al., *Crystal Structure of the Human Cannabinoid Receptor CB2*. Cell, 2019. **176**(3): p. 459-467.e13.
63. Hua, T., et al., *Activation and Signaling Mechanism Revealed by Cannabinoid Receptor-Gi Complex Structures*. Cell, 2020.
64. Laprairie, R.B., et al., *Cannabidiol is a negative allosteric modulator of the cannabinoid CB1 receptor*. British Journal of Pharmacology, 2015. **172**(20): p. 4790-4805.
65. Martínez-Pinilla, E., et al., *Binding and Signaling Studies Disclose a Potential Allosteric Site for Cannabidiol in Cannabinoid CB2 Receptors*. Frontiers in Pharmacology, 2017. **8**(744).
66. Petrucci, V., et al., *Pepcan-12 (RVD-hemopressin) is a CB2 receptor positive allosteric modulator constitutively secreted by adrenals and in liver upon tissue damage*. Scientific Reports, 2017. **7**(1): p. 9560.
67. *Allosteric Modulators of the CB1 Cannabinoid Receptor: A Structural Update Review*. Cannabis and Cannabinoid Research, 2016. **1**(1): p. 22-30.
68. Gado, F., et al., *Identification of the First Synthetic Allosteric Modulator of the CB2 Receptors and Evidence of Its Efficacy for Neuropathic Pain Relief*. Journal of Medicinal Chemistry, 2019. **62**(1): p. 276-287.
69. Price, M.R., et al., *Allosteric Modulation of the Cannabinoid CB₁ Receptor*. 2005. **68**(5): p. 1484-1495.
70. Horswill, J.G., et al., *PSNCBAM-1, a novel allosteric antagonist at cannabinoid CB1 receptors with hypophagic effects in rats*. British Journal of Pharmacology, 2007. **152**(5): p. 805-814.
71. Ignatowska-Jankowska, B.M., et al., *A Cannabinoid CB1 Receptor-Positive Allosteric Modulator Reduces Neuropathic Pain in the Mouse with No Psychoactive Effects*. Neuropsychopharmacology, 2015. **40**(13): p. 2948-2959.
72. Pamplona, F.A., et al., *Anti-inflammatory lipoxin A₄ is an endogenous allosteric enhancer of CB₁ cannabinoid receptor*. Proceedings of the National Academy of Sciences, 2012. **109**(51): p. 21134-21139.
73. Navarro, H.A., et al., *Positive allosteric modulation of the human cannabinoid (CB1) receptor by RTI-371, a selective inhibitor of the dopamine transporter*. 2009. **156**(7): p. 1178-1184.
74. Laprairie, R.B., et al., *Enantiospecific Allosteric Modulation of Cannabinoid 1 Receptor*. ACS Chemical Neuroscience, 2017. **8**(6): p. 1188-1203.
75. Vallée, M., et al., *Pregnenolone Can Protect the Brain from Cannabis Intoxication*. Science, 2014. **343**(6166): p. 94-98.
76. Priestley, R.S., et al., *A potential role for cannabinoid receptors in the therapeutic action of fenofibrate*. 2015. **29**(4): p. 1446-1455.

77. Xing, C., et al., *Cryo-EM Structure of the Human Cannabinoid Receptor CB2-Gi Signaling Complex*. Cell, 2020. **180**(4): p. 645-654.e13.