The new insights into protein structures by computational methods

by

Ye Zou

B.S., Henan University of Science and Technology, 2009
M.S., Hebei University, 2014

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Abstract

Protein structures have been reported to the Protein Data Bank (PDB) to share the information with scientists and researchers. They provide structural information and help other scientists and researchers to understand the mechanisms of these proteins. However, crystal structures and electron microscopy (EM) structures only show a single conformation of a protein which makes it difficult to understand the protein's flexibility and internal motions. Computational and simulation methods offer alternative ways to study protein structures and understand mechanisms, and have recently been more frequently used for chemical, biochemical, and biological research. This work is going to show new insights into protein structures by integrating different computational methods, solving the issues with proteins and protein complexes, and present protein modelling, ligand docking, and simulations.

Proctolin is a neuropeptide, RYLPT. The proctolin receptor is a G-protein coupled receptor. They are encoded in arthropods, but not in the honey bee. *Varroa* destructor is an ectoparasite of honey bees. Here, we tried to use proctolin to design a novel potential drug to treat these mites and save honey bees. In this work, the homology model of proctolin receptor has been built and used to study proctolin docking by Induced fit docking and QM (quantum mechanics) - polarized ligand docking. For further study, we performed molecular dynamics simulations to unravel the binding mechanism of proctolin. We found and explain that the first and second residues of proctolin form two cation-pi interactions with Tyr99 and Arg111 from proctolin receptor. This shows the first two residues act as an anchor docked into the binding pocket.

Later, we also studied the chromophore behavior of red fluorescent protein. Fluorescent proteins have been improved from natural fluorescent proteins and heavily used in life science as

protein labels, markers of gene expression, and living-cell imaging. The characteristics of fluorescent protein chromophores can give us information about the structure-function relationship with the protein matrix. This can guide us to engineer tuned color variants and broaden the spectral range of useful proteins. We present the behavior of the chromophore with the protein matrix and have a deeper understanding of fluorescent proteins. We performed molecular dynamics simulations on four trans-form fluorescent proteins. All fluorescent proteins with the trans-form chromophores tend to be non-planar, and the residues 67, 92, 143, and 197 are more important sites. These residues interact with the chromophore.

Since the most inspiring thing that happened last year was the achievement of AlphaFold2, we also did an assessment and application of the structures predicted by AlphaFold2 on popular drug targets. It provides and emphasizes a valuable way to apply the AI developed method in drug discovery.

The new insights into protein structures by computational methods

by

Ye Zou

B.S., Henan University of Science and Technology, 2009
M.S., Hebei University, 2014

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Co-Major Professor
Dr. Erika Geisbrecht

Approved by:

Co-Major Professor
Dr. Ho Leung Ng

Copyright

Abstract

Protein structures have been reported to the Protein Data Bank (PDB) to share the information with scientists and researchers. They provide structural information and help other scientists and researchers to understand the mechanisms of these proteins. However, crystal structures and electron microscopy (EM) structures only show a single conformation of a protein which makes it difficult to understand the protein's flexibility and internal motions. Computational and simulation methods offer alternative ways to study protein structures and understand mechanisms, and have recently been more frequently used for chemical, biochemical, and biological research. This work is going to show new insights into protein structures by integrating different computational methods, solving the issues with proteins and protein complexes, and present protein modelling, ligand docking, and simulations.

Proctolin is a neuropeptide, RYLPT. The proctolin receptor is a G-protein coupled receptor. They are encoded in arthropods, but not in the honey bee. *Varroa* destructor is an ectoparasite of honey bees. Here, we tried to use proctolin to design a novel potential drug to treat these mites and save honey bees. In this work, the homology model of proctolin receptor has been built and used to study proctolin docking by Induced fit docking and QM (quantum mechanics) - polarized ligand docking. For further study, we performed molecular dynamics simulations to unravel the binding mechanism of proctolin. We found and explain that the first and second residues of proctolin form two cation-pi interactions with Tyr99 and Arg111 from proctolin receptor. This shows the first two residues act as an anchor docked into the binding pocket.

Later, we also studied the chromophore behavior of red fluorescent protein. Fluorescent proteins have been improved from natural fluorescent proteins and heavily used in life science as

protein labels, markers of gene expression, and living-cell imaging. The characteristics of fluorescent protein chromophores can give us information about the structure-function relationship with the protein matrix. This can guide us to engineer tuned color variants and broaden the spectral range of useful proteins. We present the behavior of the chromophore with the protein matrix and have a deeper understanding of fluorescent proteins. We performed molecular dynamics simulations on four trans-form fluorescent proteins. All fluorescent proteins with the trans-form chromophores tend to be non-planar, and the residues 67, 92, 143, and 197 are more important sites. These residues interact with the chromophore.

Since the most inspiring thing that happened last year was the achievement of AlphaFold2, we also did an assessment and application of the structures predicted by AlphaFold2 on popular drug targets. It provides and emphasizes a valuable way to apply the AI developed method in drug discovery.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First, I want to acknowledge our Biochemistry and Molecular Biophysics department. There are so many outstanding professors, mentors, and teachers. This system offers a unique environment to students and allows us to have a good place to keep following our interesting academic journey. Thank you to Dr. Michal Zolkiewski, you are the first professor from BMB who introduced me to this wonderful department. Thank you to all BMB professors, for your suggestions, recommendations, and comments in all these years.

I want to thank my advisor Dr. Ho Leung Ng for his constant support and encouragement in the past years. He helped me find my interests and brought me into the world of drug discovery, a broad field which allowed me to pursue my curiosity. He was a guide, a mentor, and a pathfinder. His attitude is always positive and affects me and my other lab members. It helps us always be optimistic even in difficult times. He always has a passion for new interesting things and keeps improving himself. He will always be a role model in my life.

I want to thank Dr. Katsura Asano. I was so glad to take your class. The class showed me a different study field and taught me how to use my own specialty in another field of research. He emanates inventiveness, ingenuity, and originality. His passion for research will always affect me.

I want to express my appreciation of Dr. Lawrence Davis. Thanks for your help, not just in gardening but also all your life experience. You are the person who greatly influences students and makes us be better people. You are a life mentor.

I want to thank you to my committee members, Dr. Michael Kanost, and Dr. Erika R. Geisbrecht, for your help, advice, encouragement and mentoring. You taught me how to be a good researcher and scientist. You taught me a great deal. You are my study models and

influenced me. Dr. Erika R. Geisbrecht, you are not just a mentor, also a female role model to students. You affect students to become an excellent scientist like you.

I want to thank my collaborators, Dr. Yoonseong Park, and Moritz Mating. Thank you for bringing such interesting research. This opened my new research scenario, and thank you for the contribution to save honey bees.

Thank you to Dr. John Head, who is a professor from University of Hawaii at Manoa. It was his class which started my computational research life.

I want to thank all my lab mates from Ng's lab, Dr. Samson Souza, Mian Huang, John Ewalt, Cisco Jimenez, Allison Kagawa, Alex Sens and all undergraduate students. Aloha au iā ʻoukou a pau, he ʻohana mākou. Mahalo!

Thank you to Dr. Huafeng Yang and Dr. Weifeng Zheng. Thank you for all your help and support in all these years. Thank you to Doris Li for the crazy time we had.

I want to thank Dr. Miao Li, who is the first person I met at Manhattan. You showed me Manhattan. Thank you for helping me pass through the very hard transition time and thank you for all the fun time you brought to my life, and the hot pot feasts. Thank you to Chenshuang Lu for your kindness and the chats we had. Thank you to Yu Song, and Taihao Yang for having the feasts to warm up the cold spring festivals, mid-autumn festival and all the Chinese festivals. Thank you to Yanan Zhou and Maverick for having fun times and relaxing from the stressful graduate student's life. Thank you to Dr. Xin Xu for the food mapping. Thank you to Linah Alkotami for bringing me to the garden, the fun time at garden and outside lab. Thank you to Nitin Mishra for being a good lab neighbor and friend.

Many thanks for the people who have appeared in my life. All of you taught me how to be a better person, how to have better thinking. I will always remember every moment that happened with you all. These are very precious to me.

When I am writing these words, I realize there are none which can express my appreciation. I appreciate all the people that appeared in my life. Mahalo!

Dedication

To my dear grandpa, Prof. Yusan Tian.

Preface

This study focused on applying computational methods to solve chemical and biochemical questions and trying to reveal the mechanisms of protein interactions with ligands. The first chapter was published as a journal article, Zou, Y., Ewalt, J., Ng, H. Recent insights from molecular dynamics simulations for g protein-coupled receptor drug discovery. It introduced the different computational methods and used G-protein coupled receptors as study cases to show the application of computational methods. This starts the computational scenario of this thesis.

Chapter 2 is a collaboration with Dr. Yoonseong Park to fight against *Varror destructor,* an ectoparasite of honey bee. We studied proctolin receptor shown in arthropods, but not shown in most insects, especially the honey bee. Hence, we were trying to unravel the complex behavior of proctolin and proctolin receptor. Proctolin receptor is a G-protein coupled receptor. We used homology model building, different docking studies, and Molecular Dynamics (MD) simulations to study proctolin and proctolin receptor. It shows the binding interactions between proctolin and proctolin receptor. We found the cation-pi interactions play an important role between proctolin and proctolin receptor. It is shown the reserved residue Tyr 99 acts as a dock to let the anchor - proctolin bind into the binding pocket.

Chapter 3 studied the chromophore behavior of red fluorescent proteins. Fluorescent proteins have been improved from the natural fluorescent proteins. Here, we illustrate the trans-form fluorescent proteins. We found the planar, trans-form chromophores of fluorescent proteins show very interesting conformational changes from their crystal structures. Here, we performed MD simulations to understand the chromophore behavior at 300 K and their interaction with the protein matrix.

In chapter 4, we studied the latest protein modeling method, AlphaFold 2. It achieved the best performance in Critical Assessment of Protein Structure Prediction 14 (CASP14). This is also the biggest, exciting news. AlphaFold 2 also filled the gap between the experimental protein structures from protein data bank and the total known DNA encoded proteins. However, the limitation of this new AI-based method has been neglected. Here, we focused on the assessment and application of the structures predicted by AlphaFold 2 on popular drug targets.

Chapter 1 - Insights from molecular dynamics simulations for G protein-coupled receptor drug discovery

This chapter has been published in International Journal of Molecular Sciences.

Introduction

G protein-coupled receptors (GPCRs) are the largest superfamily of membrane proteins in the human genome (Rosenbaum et al., 2009). They are also the most important and largest collection of pharmacological drug targets (Hauser et al., 2017). These receptors share modest sequence similarity but high structural conservation, all with seven transmembrane-spanning helices, an extracellular N terminus, and an intracellular C terminus. The crystal structure of bovine rhodopsin, one of the most studied GPCRs, shows the seven helices forming a helical cylinder, which is linked by three intracellular and three extracellular loops (Palczewski et al., 2000). There are five main GPCR families: rhodopsin (class A), secretin (class B), glutamate (class C), frizzled/taste (class F), and adhesion (Alexander et al., 2015). GPCRs' typical functions are the translation of extracellular stimulation into intracellular signals via the binding of different ligands, which then cause different conformational changes and downstream effects. Each receptor can activate specific G proteins and regulate unique downstream signals. The GPCR ligands bind to these receptors and stabilize conformations, then regulate and modulate various intracellular transduction processes. GPCR agonist ligands are extremely diverse and include photons, ions, odorants, tastants, small molecule neurotransmitters, amino acids, polypeptides, hormones, nucleotides, and lipids (Allen and Roth, 2011; Roth and Kroeze, 2015). Classical GPCR activation involves an agonist-induced conformational change. This causes the receptor to interact with the Gα subunit, part of the heterotrimeric GTP-binding proteins (G

proteins), which then dissociates from the Gβγ subunits and binds to guanosine diphosphate (GDP). Research has also shown that there are also at least five different activation modes different from classical activations(Wang et al., 2018; Zhou et al., 2017). These are involved in phenomena such as intracellular activation, dimerization activation, transactivation, biphasic activation, and biased activation. Because GPCRs are involved in these activations, which are related to many human diseases, they are common drug targets (Kaczor et al., 2016; Latorraca et al., 2017; Lee et al., 2019; Wacker et al., 2017a). GPCRs are normally activated on the cell surface. GPCRs can also be activated from inside the cell, which is called intracellular activation(Schiaffino et al., 1999). GPCR activation functions also depend on the forms of the GPCRs, monomeric or dimeric, the latter of which is called dimerization activation(Gomes et al., 2001). GPCRs can be activated by ligands and these activated GPCRs can transactivate other proteins, such as receptor tyrosine kinases (RTKs), which then can activate Ras/MAP kinases. This is called transactivation(Daub et al., 1996). GPCRs can also activate two different phases of signaling. This is called biphasic activation(Schorb et al., 1995). The last activation is called biased activation (Figure 1.1). The biased activation (also called biased signaling pathway) involves parallel G protein-independent signaling pathways. Instead of activating G proteins, there are activating β-arrestins, which mediate downstream signaling. The main functions are internalization and desensitization. Activated by agonists, the GPCRs are phosphorylated by GPCR kinases (GRKs) on multiple sites of the C-terminus; arrestins will bind to these phosphorylated sites, and G protein-coupling will be blocked by the arrestin-GPCR complex. This is called biased activation(Alexander *et al.*, 2015; DeWire et al., 2007; Lefkowitz, 2004; Milligan and Kostenis, 2006; Scheerer and Sommer, 2017). In this review article, we focus on the biased-signaling pathway and β-arrestins. There are four classes of arrestins: arrestin 1

(visual arrestin), arrestin 2 (β-arrestin 1), arrestin 3 (β-arrestin-2), and arrestin 4 (cone arrestin). Arrestins are highly conserved, with approximately 50% sequence homology between vertebrates and invertebrate primary structures(Attramadal et al., 1992; Benovic et al., 1987; Craft et al., 1994; Gainetdinov et al., 2004; Smith and Rajagopal, 2016). Because the functions of arrestins are diverse, β-arrestin biased-signaling pathways are of significant pharmaceutical interest. The current understanding of the structural conformations related to biased signaling is sparse.

Recent studies have revealed that GPCRs are dynamic proteins with multiple conformational changes depending on ligand binding, signaling proteins, and the membrane environment(Nygaard et al., 2013; Vaidehi and Bhattacharya, 2016). Several crystals of class A GPCRs in the active state show there are conformational changes in the intracellular domain and transmembrane helices 5, 6, and 7 (TM5, TM6, and TM7) of the receptors(Burg et al., 2015; Huang et al., 2015a; Kruse et al., 2013). Biophysical research has shown that the binding affinity to an agonist is increased by coupling the G protein to the receptor(Yao et al., 2009). After a ligand binds to a receptor, it causes and stabilizes conformational changes(Vaidehi and Kenakin, 2010). Agonists binding to GPCRs induce and mediate different downstream pathways, either through G protein activation or β-arrestin biased signaling. Typically, the primary ligand binding sites, orthosteric sites, are highly conserved. This presents a significant challenge for drug discovery. Ideally, a designed ligand has high selectivity and only activates specific receptors; an orthosteric site that is highly conserved between related receptors increases the difficulty of doing so. Because of the challenges these orthosteric sites present, researchers have been trying to explore new alternative binding sites, which could increase both binding affinity and decrease off-target effects. Allosteric binding sites share less homology compared with orthosteric sites,

so they have become increasingly attractive to researchers(Christopoulos, 2014). Positive allosteric modulators can increase the potency of orthosteric ligands; negative allosteric modulators can decrease the potency of the response to orthosteric ligands(Foster and Conn, 2017). Several crystal structures of GPCRs bound with allosteric modulators have already been solved. Kruse et al. solved a crystal structure of a GPCR (M2 muscarinic receptor) with an allosteric modulator(Kruse *et al.*, 2013). Dore et al. solved a class C GPCR (metabotropic glutamate receptor 5) bound with an allosteric modulator(Doré et al., 2014). Jazayeri et al. solved a class B GPCR (glucagon receptor) bound with an allosteric antagonist(Jazayeri et al., 2016). However, it is also challenging to determine crystal structures of GPCRs with allosteric modulators, and few are in the Protein Data Bank (PDB). Therefore, computational methods are valuable in helping to identify new allosteric binding sites and offer new structural information of GPCRs and their interactions with ligands.

Molecular dynamics simulations are important computational methods widely used in many fields of study. Simulations assist researchers in obtaining structural information, specifically, the conformational states that are hard to capture by experimental methods. Because of rapid technological development, computing is easier and faster than ever. This has also allowed researchers to run long time scale simulations to obtain more detailed mechanistic information(Karplus and McCammon, 2002). In this review, we will discuss the insights on GPCR interactions with ligands revealed by molecular dynamics simulations and enhanced sampling techniques.

New Insights from Molecular Dynamics Simulations

Computational and simulation methods provide other ways to explore complex systems that are difficult to study with current experimental methods. Simulation via molecular dynamics

(MD) is a very powerful and easy to use computational tool. It can help researchers determine receptor-ligand structures, dynamics of binding, and binding kinetics and functions (Cournia et al., 2015; Dror et al., 2012; Khalili-Araghi et al., 2009; Stansfeld and Sansom, 2011). More recently, MD simulations have been used to study macromolecular conformational dynamics on longer time scales, up to a millisecond and beyond (Dror et al., 2009; Dror et al., 2013; Grossfield et al., 2008; Gumbart et al., 2013). Computational and simulation methods can play useful roles in drug discovery as well. These methods have helped improve our understanding of GPCRs' structures and functions (De Vivo et al., 2016; Dror et al., 2011b; Hauser *et al.*, 2017; Marino and Filizola, 2018; Sabbadin and Moro, 2014). There are three main requirements for an MD simulation: the model system, the force field, and the MD simulation software. The most popular MD simulation packages include AMBER (Case et al., 2008), CHARMM (Brooks et al., 1983), GROMACS (Hess et al., 2008), and NAMD (Phillips et al., 2005). These packages are making simulations easier to perform and are quick to adopt new technological and methodological advances.

Using Molecular Dynamics Simulations to Study GPCR-Ligand Binding

Biased signaling generates functional selectivity and has attracted notable drug discovery interest. The structures and mechanisms involved in biased signaling are still not clear as there are very few GPCR crystal structures available.

The μ-opioid receptor (μOR or MOR) is the first GPCR that demonstrated β-arrestin-biased signaling (Violin and Lefkowitz, 2007). The μOR is the primary target for strong analgesics (Vardanyan and Hruby, 2014). The best-known opioid agonists are opiate drugs, which are among the oldest medicines and are analgesics (Stein, 2016). Even though opiates are widely used, they have notorious side effects including addiction, respiratory suppression, and

constipation. β-arrestins act as negative regulators in the μOR signaling pathways(Pasternak and Pan, 2013; Raehal and Bohn, 2005). Recent research has supported a trend in which ligands that are more biased towards the β-arrestin pathway are associated with increased undesired side effects (Conibear and Kelly, 2019; Schmid et al., 2017). Currently, there are two μOR crystal structures and two electron microscopy structures available in the Protein Data Bank (PDB: 4DKL, 5C1M, 6DDE/6DDF) (Huang *et al.*, 2015a; Koehl et al., 2018; Manglik et al., 2012). These high-resolution structures offer the possibilities of using these structures to perform simulations and potentially assist in the discovery of novel drugs with fewer side effects (Kaserer et al., 2016; Latorraca *et al.*, 2017; Manglik et al., 2016).

Crystallographic studies of μOR bound with the potent agonist, fentanyl, involved active and inactive states. The crystal structures used are PDB 5C1M and 4DKL. 4DKL is the structure of inactive μOR with the irreversible morphinan antagonist β-funaltrexamine (β-FNA), which can be seen in Figure 1.2. 5C1M is the structure of active μOR with the agonist BU72. The primary structural difference between active and inactive states is transmembrane 6 (TM6) shifting outwards by 10.3 Å (Huang *et al.*, 2015a).

Fentanyl is an analgesic that is much more potent and dangerous than morphine (Bremer et al., 2016). Unlike morphine, fentanyl can strongly induce β-arrestin biased signaling (Zheng et al., 2011). Both morphine and fentanyl have a protonatable tertiary amine in the piperidine ring. Compared to morphine, fentanyl is more flexible (Figure 1.3). Lipinski et al., using the 4DKL and 5C1M crystal structures and manually docking morphine and fentanyl into the protein, found that the mutation of Ser329$^{7.46}$ to alanine, located in the sodium binding pocket, is sensitive to the N-phenethyl chain of fentanyl (Pil and Tytgat, 2003); mutations involving Trp318$^{7.35}$ and His319$^{7.36}$ to methionine demonstrate similar sensitivities to the N-phenethyl chain of fentanyl

(Ulens et al., 2000). The superscript decimal numbers seen in the previous sentence refer to the Ballesteros–Weinstein numbering scheme, with the number to the left of the decimal referring to which of the seven transmembrane helices the residue is on and the number to the right of the decimal giving the relative position to the most conserved residue on the helix, which is numbered 50 (Isberg et al., 2015).

In the fentanyl binding mode, both active and inactive receptors are stimulated. In the inactive mode, there is a sodium cation in an allosteric site. The simulation results of active and inactive binding interactions are similar. The two ligands both have the amine of the piperidine ring protonated. In the binding mode, the protonated amines interact with residue Asp147$^{3.32}$; hence, the N-phenethyl chain is facing the intracellular side. The piperidine ring plays an important role in this binding mode. It interacts not only with A147$^{3.32}$ but also with Gln124$^{2.60}$. The N-phenethyl chain forms hydrophobic contacts with Try326$^{7.43}$, Ile296$^{6.51}$, Ile322$^{7.39}$, and Trp293$^{6.48}$. In the morphine binding mode, the simulation results of active and inactive binding interactions are also similar. The protonated amine in morphine also interacts with residue Asp147$^{3.32}$. Morphine's binding pocket is similar to fentanyl's. The protonated amine is facing intracellularly. The phenol and the ether group are facing the extracellular side and are exposed to the solvent. The CHARMM-GUI service (Jo et al., 2008) was used to parameterize and prepare the sample for simulation. The receptor was placed in a phosphatidylcholine (POPC) membrane and solvated with the TIP3P water model; the CHARMM36 force-field was used. The simulation runtime was 1.2 μs (Lipiński et al., 2019).

In this research, MD simulations were also used to study a "Trp rotamer toggle switch", which acts as a transmission switch. Trp293$^{6.48}$ was found to be a highly conserved residue, and multiple MD simulations support its central role in conformational change (Marino et al., 2018;

Shang et al., 2014; Sutcliffe et al., 2017). Analysis of MD simulations monitored the movement of Trp293$^{6.48}$ and focused on two different dihedral angles, $\chi 1$ and $\chi 2$. Three rotamers were observed that differed from those seen in the crystal structure. Such rotamers would have been difficult to observe by experimental means.

Using Molecular Dynamics to Predict Arrestin-Biased Ligands

Pursuing biased signaling is an alternative strategy to discover highly selective drugs (Soergel et al., 2014). However, a detailed understanding of the biased signaling pathway is still incomplete. In particular, the crystal structures reported are of the active and inactive states, with very few structures that are relevant to intermediate conformations. MD simulations have assisted in revealing these intermediate conformations and the mechanisms of biased signaling. We discuss an example that used MD simulations to predict arrestin-biased ligands.

Recently, McCorvy et al. reported using the D2 dopamine receptor (D2R) as a model to study GPCR-ligand binding, which involves biased signaling, by MD simulations (McCorvy et al., 2018). D2R is a primary drug target for schizophrenia and Parkinson's disease. The clinical implications of differently biased D2R ligands is a topic of great research interest. The authors describe a new method to use MD simulations to design β-arrestin biased ligands. The crystal structure of the complex of β2 adrenergic receptor (β2AR) bound to epinephrine shows that the ligand forms a hydrogen bond network with the conserved TM5 serine residues (Ring et al., 2013). Three crystal structures of the GPCRs, β2AR bound with epinephrine (4LDO), 5-HT_{2B} serotonin receptor bound with lysergic acid diethylamide (LSD) (5TVN), and β1 adrenergic receptor (β1AR) bound with 4-indole piperazine (3ZPQ), share conserved serine residues, 5.42, 5.43, and 5.46, respectively. The 5-HT_{2B} crystal structure and MD simulations show that

extracellular loop 2 (EL2) plays the role of a lid at the entrance to the binding pocket, slowing LSD's binding kinetics, which can be seen in Figure 1.4 (Wacker et al., 2017b).

There are conserved hydrophobic residues located in EL2, such as isoleucine 184 (I184EL2) in D2R, which may lead to β-arrestin signaling in this receptor. McCorvy et al. also did biophysical experiments to test the structure-functional selectivity relationships of the designed compounds such as indole N-substitutions. They were intended to disrupt interactions with transmembrane 5 (TM5) and confirmed the validity of their simulations. MD simulations used a homology model based on the dopamine D3 receptor with the antagonist eticlopride (3PBL). MD simulations were used to predict EL2 engagement and find the key to biased signaling. The designed compounds contain an indole-piperazine moiety. The MD package AMBER14 was used. MD simulation results show that the conserved $D114^{3.32}$ forms a salt bridge with the protonated nitrogen piperazine ring in the orthosteric site. To better understand the mechanism between G protein activation and β-arrestin signaling, MD simulations were done without dihydroquinoline-2-one and the alkyl linker. These simulations showed that in the head group of Compound 1 (Figure 1.5a), the hydrogen on the indole formed a hydrogen bond with $S193^{5.42}$ and did not interact with I184EL2. Instead of forming a hydrogen bond, the nitrogen on the indole of Compound 2 was tracked by $S193^{5.42}$; the center of the indole is closer to I184EL2 than Compound 1. The two compounds show different poses and interactions with GPCRs. Compound 1 is associated with G protein activation, and Compound 2 (Figure 1.5b) is associated with β-arrestin biased signaling. Based on these critical simulation results, the authors were able to design new arrestin-biased compounds (McCorvy *et al.*, 2018).

Identification of New Ligand Binding Sites and Activation Mechanisms by Accelerated Molecular Dynamics and Metadynamics Simulation

MD simulations have been commonly used to study GPCR activation mechanisms and conformational changes (Dror et al., 2011a; Esguerra et al., 2016; Zeng et al., 2015). However, in the study of conformational dynamics, sampling the extended timescales involved is the greatest challenge for MD simulation studies. To overcome this challenge, computational scientists have invented new methods and algorithms. Continued increases in computing power have also been of assistance. Currently, commonly used advanced hardware includes powerful graphics processing units (GPUs), supercomputers, and cloud computing (Dror *et al.*, 2012; Kohlhoff et al., 2014). Conventional molecular dynamics (cMD) simulations have been widely used to study GPCR activation mechanisms. cMD can reach timescales of microseconds (Bai et al., 2014; Shukla et al., 2015; Vaidehi et al., 2014; Xiao et al., 2015; Xu et al., 2015; Yang et al., 2015b). However, there are many cases of GPCR activation processes that can take longer than the timescale limits of cMD simulation. Two of the most popular computational methods for enhanced sampling of protein molecular dynamics simulations to access longer time scales are accelerated molecular dynamics (aMD) and metadynamics (Maximova et al., 2016). aMD simulation improves conformational space sampling by adding a bias potential into cMD that reduces energy barriers between different states (Hamelberg et al., 2004). Metadynamics simulation parameterizes the model system with collective variables and introduces bias potentials to discourage resampling of explored conformational space (Valsson et al., 2016). As a result, both aMD and metadynamics simulations reduce calculation time and are much faster than regular cMD simulations at the risk of utilizing modified energy landscapes. Here, we discuss an example of using cMD and aMD simulations to find a new ligand-binding site and

activation mechanism. P2Y1R is a purinergic GPCR that is activated by adenosine 5′-diphosphate (ADP). It plays an important role in platelet aggregation and thrombosis formation (Gachet, 2008; Jacobson et al., 2011). The crystal structure of P2Y1R bound with the antagonist MRS2500 is available on the Protein Data Bank (4XNW); however, there is no structure of P2Y1R bound with an agonist. Li et al. used cMD and aMD simulations to find a new agonist-binding site (Li et al., 2017). The crystal structure of P2Y1R bound with MRS2500 (4XNW) was used for simulation. The agonist, 2MeSADP, was docked to the same site as MRS2500. The results showed that the aromatic adenine ring of 2MeSADP has a π-π stacking interaction with Tyr303$^{7.32}$. The pyrophosphates interact with Arg128$^{3.29}$ and Arg310$^{7.39}$. However, these results did not match the experimental results. The experimental results show that the mutagenesis of His132$^{3.33}$, Tyr136$^{3.37}$, and Lys280$^{6.55}$ decreased the binding affinity of 2MeSADP with P2Y1R. So, aMD simulations were used to run the long time-scale simulations necessary to find an alternative binding site, which can be seen in Figure 1.6. In this research, all MD simulations used the particle mesh ewald molecular dynamics (PMEMD) module from AMBER14. The ff99SB force field was used for the receptor, and the general AMBER force field (GAFF) was used for the ligands.

The aMD simulations results show the aromatic adenine ring of 2MeSADP interacting with His132$^{3.33}$ through π-π stacking interactions. The N1 in the adenine forms hydrogen bonds with the Tyr136$^{3.37}$ and Thr221$^{5.42}$. The pyrophosphates interact with Arg128$^{3.29}$, Arg287$^{6.62}$, Arg310$^{7.39}$, Lys280$^{6.55}$, and Tyr306$^{7.35}$. These results match the experimental results. In this research, metadynamics simulations were used to obtain the potential of mean force for helices III-helix VI, which showed that 2MeSADP-P2Y1R has three states: inactive, intermediate, and active. Apo-P2Y1R and MRS2500-P2Y1R only have two states. The aMD simulation allowed

2MeSADP-P2Y1R to pass through the intermediate state and finally reach the active states from the inactive state (Li *et al.*, 2017).

Study of Allosteric Modulation and Dynamics of GPCR-Ligand Binding

Traditionally, research has focused on the orthosteric binding sites of GPCRs. There has been growing recent interest in allosteric sites, especially regarding their advantages for drug discovery. Allosteric ligands can act as positive allosteric modulators, negative allosteric modulators, or neutral allosteric ligands. One of the advantages of allosteric modulators is that they have potentially better selectivity than orthosteric ligands (Wootten et al., 2013).

In studying the mechanisms of allosteric modulators, experiments such as x-ray crystallography, NMR, and systematic mutagenesis experiments have had important roles, and these techniques have already given researchers much conformational information of GPCRs and GPCR-mediated signaling. However, these methods only provide information on static conformational states, with the X-ray crystallography experiments only resolving two conformational states, active and inactive. There is no information about the conformations and transitions between these two states. The transition between inactive and active states is difficult to access by experimental methods. Currently, a more intricate, multi-state model, is being investigated by researchers (Bar-Shavit et al., 2016; Latorraca *et al.*, 2017). Molecular dynamics simulations give the possibility of revealing the intermediates of GPCRs between multiple states (Latorraca *et al.*, 2017; Rajagopal et al., 2010; Violin et al., 2014). In the following section, we will discuss how molecular dynamics simulations provide information on intermediate states and can also identify new allosteric binding sites.

The muscarinic receptors are drug targets for many diseases such as overactive bladder, chronic obstructive pulmonary disease, and neurodegenerative diseases (Felder et al., 2018;

Kruse et al., 2014). Drug discovery for the muscarinic receptors has been hindered due to the challenge of selectively targeting receptor subtypes. The crystal structures of muscarinic M3 and M4 receptors show the antagonist tiotropium (TTP) in the orthosteric site (PDB: 4DAJ, and 5DSG). The crystal structures show that the orthosteric binding site is highly conserved (Kruse et al., 2012; Thal et al., 2016). However, there is no crystal structure of a muscarinic receptor with agonist available yet. Chan et al. used MD simulations to simulate acetylcholine (ACh), the endogenous agonist, binding to M3 and M4 receptors (Chan et al., 2018). Figure 1.7 shows the chemical structures of ACh and TTP. The simulation results show ACh binding to a new, deeper allosteric site near $D^{2.50}$ that is highly conserved (Isberg *et al.*, 2015). To further understand the ACh activation process, an all-atom MD simulation was used to determine the ligand entrance pathway. ACh binds to the orthosteric site, an aromatic cage, quickly (0.1–0.2 μs). ACh was superimposed over TTP in the M3-TPP structure; both of their conformations are the same. After 0.5–0.6 μs into the simulation, the ACh goes deeper into a new binding site, and the pocket size is expanded. In contrast, during the entire simulation, TTP stays at the orthosteric site, as can be seen in Figure 1.8.

The simulation also shows ACh interacting with residues $A112^{2.57}$, $I116^{2.53}$, $A150^{3.35}$, $S154^{3.39}$, and $W503^{6.48}$. In the M4 receptor MD simulation, ACh shows a similar binding behavior with the M3 receptor, with ACh getting into a deeper binding site next to $D78^{2.50}$. The only difference is that there is an ionic interaction between the quaternary nitrogen of ACh and $D113^{2.50}$. In the M4-ACh model, the quaternary nitrogen interacts with $I81^{2.53}$, $V115^{3.35}$, $S116^{3.36}$, and $S119^{3.39}$ because the ACh flips by 180 degrees.

MD simulations are also used to determine free energies. From the extracellular surface to the orthosteric site to the new binding allosteric site, there are energy barriers between the

orthosteric site and the new binding site. This explains why ACh slowly moves from the orthosteric to the new binding sites. These results demonstrate that molecular simulations can be used to determine free energy differences between different binding states, which means that it can also potentially be used to design new allosteric ligands.

In this research, the CHARRMM36 force field was used for the receptor, and the CHARMM CgenFF force field was used for the ligands. The MD package GROMACS was used. The simulation was run longer than 3 μs. Hence, these MD simulations have allowed researchers to see the entire process of ACh binding to the orthosteric and allosteric sites. They also provided unique insights into mechanistic differences induced by the agonist and antagonist. Both TTP and ACh have a quaternary nitrogen. The simulation results suggest that the positively charged nitrogen can be stabilized by forming ionic interactions with the highly conserved Asp2.50 in both M3 and M4. This may help researchers identify and optimize future drug candidates. The successful identification of specific allosteric modulators for muscarinic receptors has already inspired similar approaches targeting other receptors involved in neurodegenerative and psychiatric diseases (Bock et al., 2018).

An interesting interaction between a positively charged ligand nitrogen atom and a negatively charged receptor aspartate is also observed in the simulation studies of μOR. The two opioid ligands, fentanyl and morphine, both have a protonated nitrogen amine in the piperidine ring. Both interact with $Asp147^{3.32}$. In M3 and M4, the positive nitrogen charges form interactions with $Asp^{2.50}$. The positive nitrogen charges in the ligands apparently play a central role in binding and activation.

Insights of Molecular Dynamics to Drug Design

Drug discovery is among the most challenging of scientific enterprises: it is high risk, high cost, and requires a long time to move from the bench to the market (Paul et al., 2010). Early-stage structure-based drug design has the potential to reduce risk and accelerate projects. The recent, tremendous advances in GPCR crystallography have provided new opportunities for structure-based drug design (Jazayeri et al., 2015; Kooistra et al., 2014; Kumari et al., 2015; Shoichet and Kobilka, 2012). Increasing experimental evidence shows that the GPCR-signaling pathway is more complicated than classical signaling (Wacker *et al.*, 2017b). The multiple mechanisms of GPCR activation and regulation offer diverse possibilities for drug discovery (Kaya et al., 2013; Marti-Solano et al., 2015).

GPCRs undergo major conformational changes during their functional cycles. Researchers frequently want to design a ligand that can specifically bind to the target, in addition to wanting the ligand to activate or inactivate a desired signaling pathway with little to no off-target effects. To reach this point, the designed ligand needs to bind to a specific binding site so that it induces certain conformational states of the receptor. Therefore, understanding the structural dynamics and the mechanisms of various signaling pathways of GPCRs is crucial to the design of GPCR-targeted ligands. As described above, GPCR functional mechanisms are extremely complicated. Hence, understanding receptor dynamics is very important for drug design. There is growing interest in using allosteric modulators for GPCR drug discovery (Huang et al., 2015b). It has been difficult thus far to identify new allosteric binding sites from crystal structures, in the rare cases that crystal structures are even available. MD simulations can help researchers solve these problems.

Enhanced sampling methods play an important role in this research. Adaptive biased techniques offer the chance to shorten simulation time, thus decreasing computational cost, and let ligands explore more conformational space by making it easier to escape local minima. aMD, one of the more popular adaptive-biased methods, allows the dynamic processes that are required to have agonists pass some energy barriers and reach new energy states. This should allow the possibility of finding new allosteric sites. Metadynamics allows researchers to observe the two-dimensional energy landscape, thus easily seeing the energy differences in all the states. They can serve as a reliability check on the results that aMD provides (Li *et al.*, 2017). To achieve these enhanced sampling methods, the most popular MD packages currently available include AMBER, NAMD, GROMACS, and CHARMM. To obtain an accurate result, the force field chosen plays a very important role. The ff99SB and ff14SB force fields are those most commonly used with AMBER. ff14SB is similar to ff99SB; however, ff14SB was modified by empirical adjustments of the protein backbone dihedral parameters, ϕ and ψ (Maier et al., 2015). CHARMM36 is a force field commonly used with NAMD and CHARMM, which includes improved refined backbone CMAP (a grid-based correction for the φ-, ψ-angular dependence of the energy) potentials and side-chain dihedral parameters (Huang and MacKerell Jr, 2013). To obtain new and accurate insights from simulations, the inclusion of lipids in the model system is important. Experimental results show that bovine rhodopsin is sensitive to lipid environments (Soubias et al., 2010; Zaitseva et al., 2010). Crozier et al. are the first to report computational insight about rhodopsin-lipid interactions (Crozier et al., 2003). They found differences of lipid accessibility differences for the transmembrane helices of rhodopsin. Later, Lyman et al. reported the adenosine A_{2A} receptor in a cholesterol-free POPC membrane, both with and without the antagonist ZM241385 (Lyman et al., 2009). It showed that there is a gap between

transmembrane 1 (TM1) and transmembrane 2 (TM2), which allows the lipid headgroup into the binding site and causes receptor instability when there is no ligand present. This prediction was proved later (Hanson et al., 2008). This is also a good example of long-time scale simulation, with the time scale being 3 μs. Simulations with longer time scales may give us new insights on slower developing phenomena. Previously, simulations performed by Lipinski et al. were longer than 1.2 μs, from which the "Trp rotamer toggle switch" was found.

The primary goal when designing a GPCR-targeted drug is normally to make a ligand that, in addition to binding the desired target, also creates a specific signaling profile. To achieve the desired signaling profile, the drug needs to be able to stabilize certain conformational states of the receptor. If the desired conformational change involves an agonist creating more stabilized active states compared to inactive states, something that needs to be taken into account is how minor changes in the structure of the binding pocket can be associated with different signaling profiles and different intracellular coupling interface conformations (Latorraca *et al.*, 2017). There are challenges involved in successfully doing this. Small changes in one area of the binding pocket can have larger effects elsewhere in the pocket. Changes in an area completely outside of the pocket itself can also have small or large effects throughout the pocket, and vice versa; the latter of which can be important when an allosteric modulator is introduced after the orthosteric ligand is already in the orthosteric binding site. Further difficulties can arise if one seeks to design a modulator that only affects arrestin signaling when changes to any of the sites mentioned above can have an effect on G protein and/or arrestin signaling (Latorraca *et al.*, 2017).

The potential contributions of simulations include helping identify important interactions the orthosteric ligand can make with the binding pocket or rearrangements of the binding pocket

induced by the ligand. They can also assist with characterizing receptor pocket dynamics, both for known structures based on experiments as well as intermediate or metastable states that are difficult or impossible to currently access experimentally. Another potential use is assessing how ligands affect the pocket and receptor dynamics, and comparing dynamics of closely related GPCRs, which should allow more precise and specific ligand design (Latorraca *et al.*, 2017).

A simulation-based approach was used to design chemical modifications that substantially altered a modulator's allosteric effects on the M2 muscarinic receptor. The modulators the researchers initially studied all partially block the ingress and egress of orthosteric ligands. The allosteric binding site is along the path that the orthosteric ligands take to bind to the orthosteric site, which is curious because the modulators studied also weakened the association and dissociation of the orthosteric ligands. Simulations indicated that the ligand interaction mode was different than initially proposed. The researchers were able to design new modulators that took advantage of information about ligand interactions from the simulation results; the measured affinities of the new modulators were consistent with those predicted by the simulations (Dror *et al.*, 2013).

Summary and Perspectives

Computational methods are essential tools for biomacromolecular structural studies. GPCRs are the largest class of drug targets and have structural flexibility, dynamic structures, and complex biological functions. Recent breakthroughs in GPCR crystallography have enabled accurate and predictive MD simulations. Here, we have reviewed recent works that have used MD simulations and enhanced sampling methods to study interactions with new ligands, characterize unknown active/inactive states, and identify new binding sites. This has allowed researchers to gain insights to study new potential drug candidates and obtain qualitative

structural information in less time. With computers and algorithms continually growing faster, computational methods will be even more effective in helping future researchers reveal the inner mysteries of GPCRs and their ligands.

The ever-increasing number of GPCR structures found by molecular dynamics simulations of crystal structures will provide a growing database from which new ligands and potentially new binding sites can be determined and explored. Because testing in silico is less resource intensive than in vitro or in vivo, this should allow researchers to find new interesting drug targets.

Figures

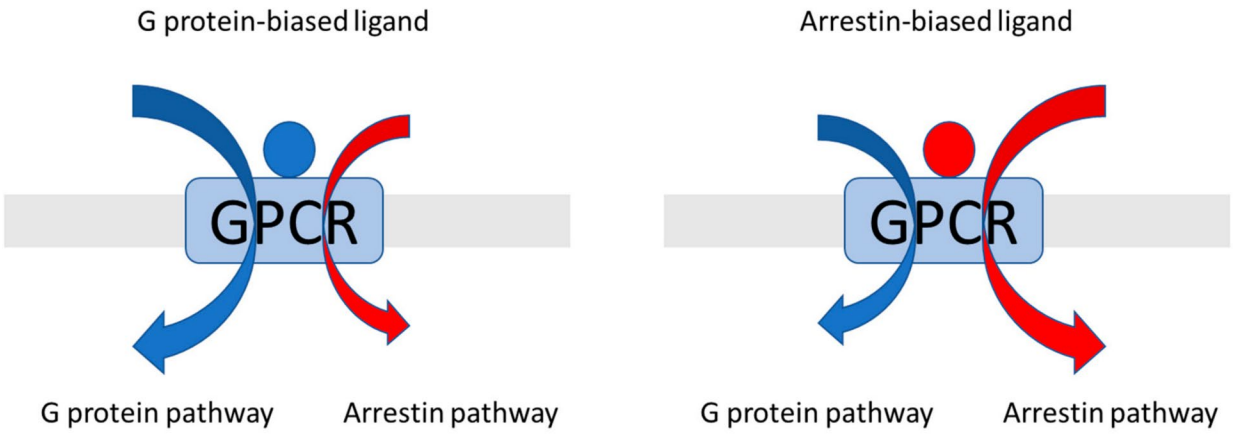


Figure 1.1 Overview of pathway-biased ligand activation for G protein-coupled receptors (GPCRs).

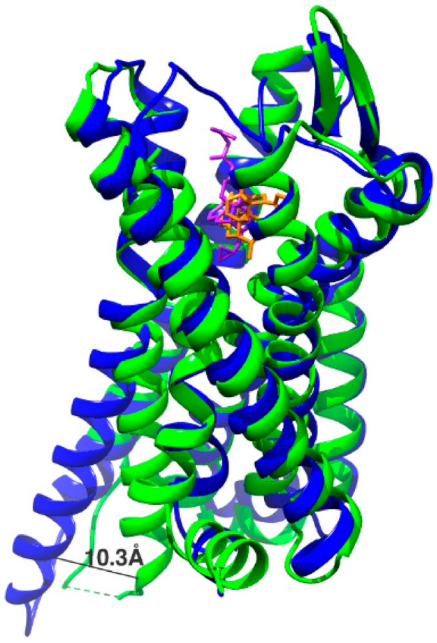


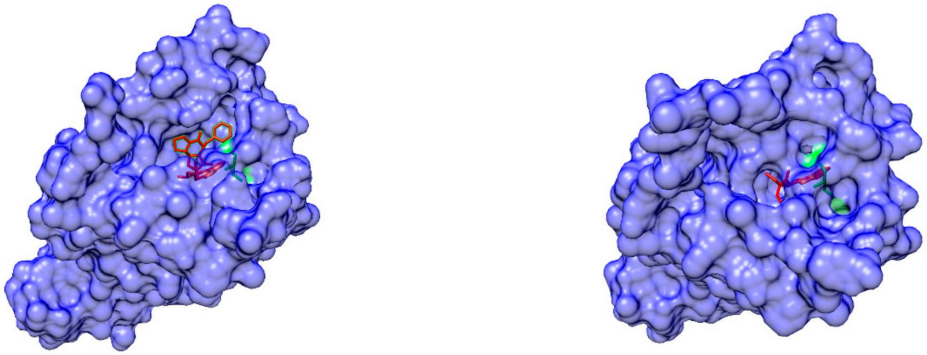
Figure 1.2 Activation of μOR displaces transmembrane 6 (TM6) by 10.3 Å.

The inactive state is in blue and is bound to the antagonist β-FNA (purple). The active state is in green and is bound to the agonist BU72 (orange).

(**a**) (**b**)

Figure 1.3 Chemical structures of μ - opioid receptor agonists.

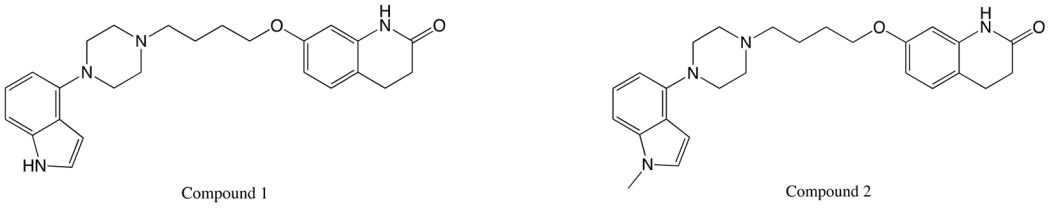
(a) morphine; (b) fentanyl.



(**a**) (**b**)

Figure 1.4 The binding pocket of 5-HT2B

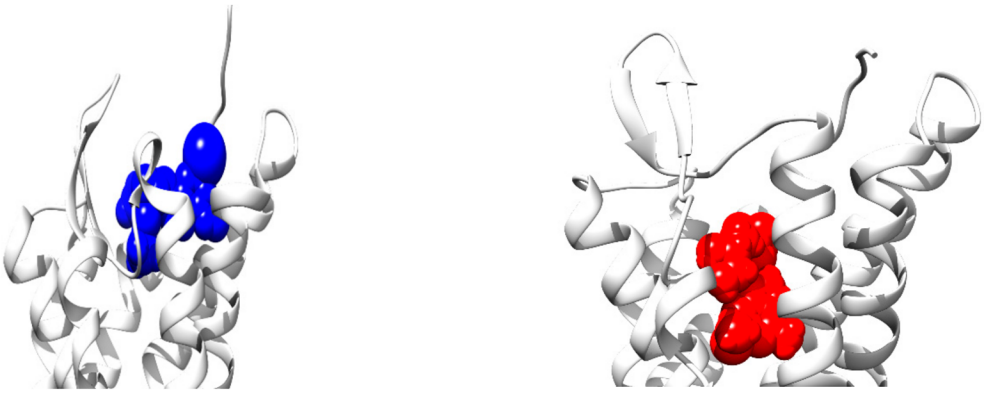
(**a**) open when ergotamine binds and is (**b**) partially closed by movement of $L209^{EL2}$ when LSD binds.



(a) (b)

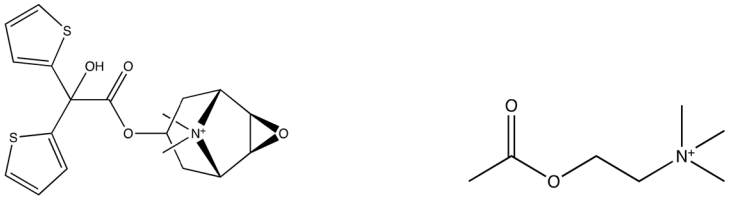
Figure 1.5 Structure of indole-aripiprazole hybrid compounds used to investigate D2R biased signaling.

(a) Compound 1; (b) Compound 2 (McCorvy *et al.*, 2018).



(a) (b)

Figure 1.6 P2Y$_1$R binding sites for (a) MRS2500 and (b) 2MeSADP.



(**a**) (**b**)

Figure 1.7 Chemical structures of muscarinic receptor ligands.

(a) tiotropium (TTP); (b) acetylcholine (Ach).

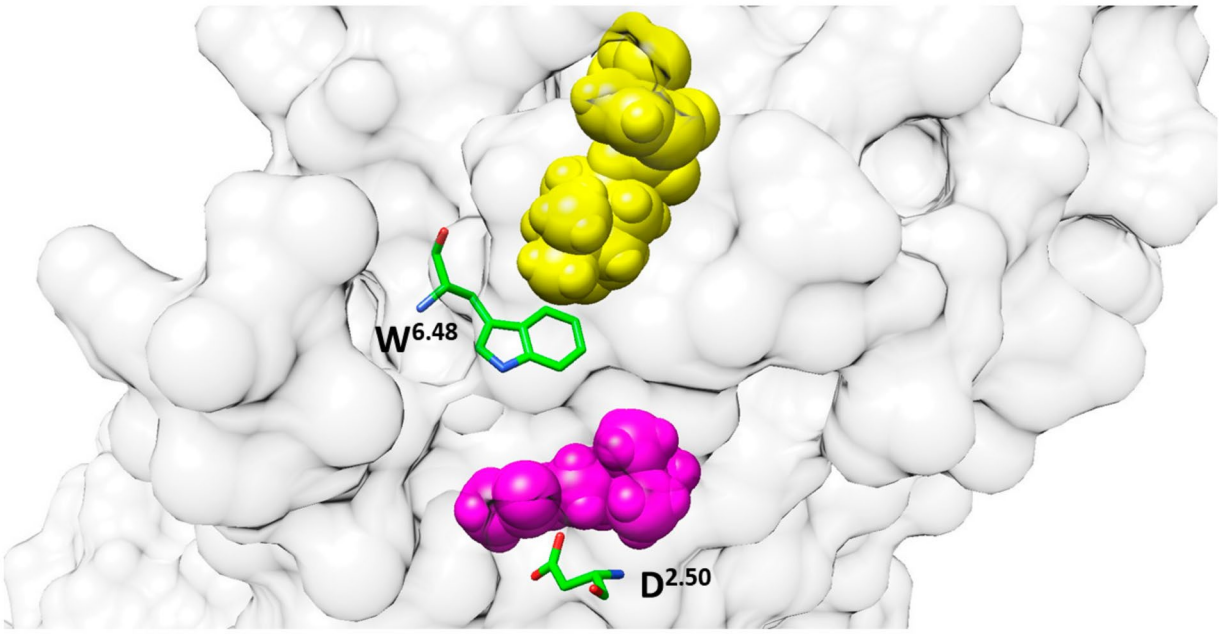


Figure 1.8 The M3 muscarinic receptor orthosteric binding site is near $W^{6.48}$, whereas the new allosteric binding site is near $D^{2.50}$.

Chapter 2 - The study of proctolin and proctolin receptor by computational Methods

Summary

European honey bees are very important to agriculture, however, the population of European honey bees is decreasing. One of the biggest threats is the *Varroa mite,* which is an ectoparasite of the European honey bees that spreads viruses and causes colony collapse disorder (CCD). The purpose of this study is to examine the proctolin receptor from these mites to design potential compounds that can be used as a new generation of acaricides. Proctolin is a unique neuropeptide in a variety of arthropods and most importantly, the proctolin receptor is absent from most species of insects. In this research, we integrated different computational methods to build the proctolin receptor homology model and studied the interactions between proctolin and the proctolin receptor based on experimental data. We found that two cation-pi interactions involving Tyr 99 and Arg 111 mediate important binding interactions between the peptide and the receptor. The first and second residues, Arg 1 and Tyr 2 from proctolin form cation-pi interactions with Tyr 99 and Arg 111 respectively. Kinetics studies also show that the first Arg 1 residue from proctolin forms a cation-pi interaction with Tyr 99 first, followed the second cation-pi interaction. This work shows the interactions between proctolin and proctolin receptor and gives evidence to help design acaricides.

Introduction

Apis mellifera, or the European honey bee, is an important pollinator. This insect is essential to agriculture as over 100 crops are pollinated by honey bees. However, the population of the European honey bees is decreasing. There are many reasons that are causing the decreasing bee population, such as parasites and pathogens (Berthoud et al., 2010; Dainat et al., 2012; Highfield et al., 2009). One of the most critical reasons is *Varroa destructor*. *Varroa destructor,* also called the *Varroa mite,* is a devastating ectoparasite of European honey bees, which can cause honey bee colony collapse disorder (CCD) and can carry viruses (Di Prisco et al., 2011; Highfield *et al.*, 2009; Wang et al., 2013). Acaricides began to be used to treat *Varroa mites,* and *Varroa mite* also very easily generated resistances to general chemical compounds used as acaricides (Hillesheim et al., 1996; Maggi et al., 2009; Milani, 1995; 1999). These acaricides also exhibit toxicity to honey bees (Johnson et al., 2013).

Proctolin and neuropeptides

Proctolin is a neuropeptide, with the sequence Arg-Tyr-Leu-Pro-Thr (shown in Figure 2.5). Neuropeptides are typical peptide-like molecules, and normally act as neurohormones, neurotransmitters, or neuromodulators. Proctolin was the first neuropeptide found and sequenced from insects (Starratt and Brown, 1975). It functions to stimulate visceral and skeletal muscle contraction in insects (Isaac et al., 2004). There are also two other forms, Ala-Tyr-Leu-Pro-Thr and Arg-Tyr-Leu-Met-Thr (Dircksen et al., 2011; Spittaels et al., 1995). There is evidence that proctolin-like immunoreactive neurons are distributed throughout the central nervous system (CNS) of *Rhodnius prolixus*. Proctolin is involved in the stimulation of skeletal and visceral muscles contraction, present in interneurons, and in motoneurons after an infection (Chiang et al., 2010; Lange, 1990; Lange et al., 1988).

Proctolin receptor and neuropeptide receptors

Most neuropeptide receptors have seven-transmembrane helices, which is also called G-protein coupled receptors. The neuropeptide receptor has an extracellular N-terminus, seven-transmembrane helices, and an intracellular C-terminus. When the neuropeptides are agonists, it can activate intracellular G-protein-mediated signaling. The extracellular and the transmembrane domains are more important for ligand binding (Brothers et al., 2003). The proctolin receptor was first identified from the *Drosophila melanogaster* (Johnson et al., 2003). The *Drosophila* proctolin receptor gene encoded by *CG6986*. Proctolin has very high affinity, the IC_{50} is around 4 nM (Johnson *et al.*, 2003).

From available genome sequencing, no genes encoding proctolin or proctolin receptor have been identified in *Apis mellifera* (Veenstra et al., 2012). Proctolin can be a good study model to design new potential peptide-like compounds to treat *Varroa* mites.

To have a better understanding of proctolin and proctolin receptor, we integrated different computational methods (shown in Figure 2.1) to study the binding mechanism of proctolin and proctolin receptor.

Results

The proctolin receptor homology model building

We tried to use docking methods to seek an explanation for the natural peptide exhibiting high potency binding to the receptor. Proctolin receptor is the same as other neuropeptide receptors, having seven-transmembrane helices, and lacking crystal structures due to its high backbone and side-chain flexibility. Hence, we were trying to build the proctolin receptor homology model by GPCR-I-TASSER. The proctolin receptor protein sequence was originally obtained from Dr. Yoonseong Park's lab. Now, the protein sequence was reported to GenBank,

the GenBank ID is QPP11369.1. The crystal structures were used as templates. The PDB IDs are 4N6H, 4BWB, 6ME6, 5ZBH, and 6HLP. Figure 2.2A and 2.2B show the homology model of proctolin receptor built by GPCR-I-TASSER. Figure 2.2A shows the side-view. The N-terminus is on the extracellular side. The seven transmembrane helices are in the membrane. The C-terminus is on the intracellular side. Figure 2.2B shows the top-view of the homology model of the proctolin receptor.

The proctolin docking study

Initial virtual screening

We used regular peptide-protein docking method to dock proctolin to the proctolin receptor. However, there was no rational results. Proctolin is a pentapeptide, RYLPT. There is a proline in this peptide, which increases the molecule's rigidity. As a peptide, proctolin is not as flexible as other peptides. Hence, we prepared and treated proctolin as a small molecule. To easily find the binding pocket, we only kept the extracellular N-terminus and transmembrane domains and removed the intracellular C-terminus. Figure 2.3A shows the remaining receptor and the docking result. It shows the proctolin binding pocket is around the top part of proctolin receptor. Figure 2.3B is the zoomed-in view of the binding pocket. It shows the Thr5 from proctolin forms a hydrogen bond with Arg111. The previous alanine scan results (from my collaborator Dr. Yoonseong Park's lab) show the arginine 1 (Arg, R1) and Tyr2 from proctolin are critical to the binding. However, the docking results do not match the experimental results.

Induced-Fit Docking

GPCRs typically have a flexible conformation. Hence, we chose to use induced-fit docking (IFD) to do the preliminary proctolin docking study in order to increase reliability by allowing for the flexibility of the backbone and sidechain. The IFD method has been successfully

used for other GPCRs (Beuming and Sherman, 2012; McRobb et al., 2010). The induced fit docking allowed flexible receptor conformational changes(Sherman et al., 2006). So, we believed the IFD docking method would help us to obtain better docking results. Figure 2.4 shows the IFD docking result. The first residue of proctolin is arginine (Arg, R), a positive charge rich residue and from the IFD preliminary results there are tyrosine (Tyr, Y) and phenylalanine (Phe, F) around this R1. The second residue is tyrosine (Tyr, Y), an aromatic residue and the IFD preliminary results shows there are arginine (Arg, R) residues around this Y2.

QM (quantum mechanics) - polarized ligand docking

Those residues would indicate binding contributions from cation-π interactions. The cation-π interaction as a noncovalent binding interaction plays an important role in biological receptors and it is a major force for drug and receptor interactions as well (Dougherty, 2013; Mecozzi et al., 1996).

To have more accurate treatment of electrostatic charges like cation-π interactions and to further investigate binding, we used the QM/MM (quantum mechanics/molecular mechanics) based method, QM (quantum mechanics) - polarized ligand docking to continue to process the IFD results. Then, we obtained the docking model.

Figure 2.5 shows the QM-polarized ligand docking result. Figure 2.5A shows the proctolin chemical formula. The two dashed line circles label the two critical residues, Arg1 and Tyr2 from proctolin. Figure 2.5B shows the binding pocket. The Arg 1 from proctolin binds to Tyr99 from proctolin receptor. The Tyr2 from proctolin binds to Arg126 from proctolin receptor (shown in Figure 2.5B, 2.5C, and 2.5D). This docking results show the first residues from proctolin are binding with proctolin receptor, which match with the experimental results.

The Molecular Dynamic Simulations Study of Proctolin Receptor Docking Model

For further study of the docking model, molecular dynamic (MD) simulations were used to further prove the docking model. Figure 2.7A shows the MD result. Figure 2.7B shows the zoomed in view of the binding pocket. In Figure 2.7B, it can be seen that proctolin N terminus residues interact with Tyr 99 and Arg 111 in the receptor, forming cation-π interactions. Cation-π interactions play a unique and important role between protein and ligand interactions, and this interaction has been underestimated. It is a very strong, and noncovalent binding interaction. This interaction requires an aromatic system and a cation to form. In the 20 natural residues, the side chains of phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp) have aromatic systems, which contain only sp^2-hybridized atoms. The side chains of arginine (Arg) and lysine (Lys) are the cations. In our case, Tyr 99 from proctolin receptor offers the aromatic ring, Arg 1 from proctolin offers a cation. The distance of cation-pi between Tyr 99 and Arg 1 is around 3.9 Å. The distance of cation pi between Arg 111 and Tyr 2 is around 3.6 Å. Comparing with experimental results, the alanine scan results show A1's EC50s is 956.5 nM and it is 3921.7-fold higher than the natural proctolin N terminus (from my collaborator Dr. Yoonseong Park's lab). The R1A mutant loses the positive charge in the N terminus to form the cation-π interaction with the tyrosine from receptor. It is the same case with the Y2A mutant, there is no aromatic system like benzene ring to form the cation-π interaction. These docking results match the experimental results for activities of proctolin peptidomimetics on the Varroa mite proctolin receptor. The first two residues of proctolin are like an anchor to dock into the receptor and pull the peptide into the receptor pocket.

Discussion

The results revealed the binding pocket of proctolin, the interactions between proctolin and the proctolin receptor, and the kinetics of the binding interaction. This work also integrates different computational methods to indicate the binding mechanism of proctolin and the proctolin receptor. The workflow of this research can be a model study for binding mechanisms of ligands and unknown structure receptors.

The roles of Tyr 99 and Arg 111

Tyr 99 and Arg 111 from the proctolin receptor play a very important role in the binding pocket, forming a cation-π interaction to stabilize the physical interaction. Two of the residues are conserved sites in the proctolin receptor. Tyr 99 appears more often in different species. In the proctolin receptor, Tyr 99 offers an aromatic system used like an "aromatic dock", the Arg 1 from proctolin is the "cation anchor". The "cation anchor" pulls proctolin into the binding pocket. Even in some other species, Phe replaces the Tyr. The Phe side chain can still offer the aromatic system and form a cation-π interaction with Arg 1 from proctolin. Arg 111 from the proctolin receptor in other species is Tyr 111. This removes the cation in other species making it unable to form the cation-π interaction, however it is still possible to form π-π stacking with Tyr 2 from proctolin. In other species the binding pocket will be slightly different with *Varroa* mite as well. Instead of forming the cation-π interaction, a π-π stacking may have a better fit in other species. In different species, the N-terminus first two residues from proctolin may have interactions with receptor and are critical to the binding to receptor. The differences between proctolin receptors in various species also could allow fine tuning of small molecules to be more targeted towards specific species.

The kinetic study of the two cation-pi interactions

We performed the MD simulations on the complex of proctolin and proctolin receptor. The maximum distance for cation-pi interactions is up to 5 Å. Figure 2.8 shows the distances between Tyr 99 and Arg 1 and Arg 111 and Tyr 2 during the simulations. Figure 2.8 shows the distance between Tyr 99 and Arg 1. This first cation-pi interaction of Tyr 99 and Arg 1 started forming around 480 ns. However, the second cation-pi interaction of Arg 111 and Tyr 2 started forming around 800 ns. This shows the proctolin binding into proctolin receptor has two steps. The first step the Arg 1 binding with the Tyr 99, then the Tyr 2 binding with Arg 111. Eventually, the entirety of proctolin docked into the receptor binding pocket.

The surface study of neuropeptides binding with the neuropeptide receptors

Proctolin and proctolin receptor

Figure 2.9 shows the electrostatic surface of proctolin receptor. The positive charge is shown in blue. The negative charge is shown in red. In Fig 2.9B is the top view of proctolin receptor's binding pocket. It shows the proctolin receptor binding pocket carries more negative charges. The C-terminus Arg 1 from proctolin inserted into the binding pocket and forms the cation-π interaction. Figure 2.9C shows a clip of the electrostatic surface of the binding pocket and the electrostatic surface of proctolin. The bottom of the pocket carries negative charges. The N-terminus of proctolin carries positive charges. Both the receptor pocket and the proctolin match very well. This also shows the reason of high potency with EC_{50} of proctolin.

Neurotensin and the neurotensin receptor

Neurotensin is a neuropeptide. The receptor of neurotensin is neurotensin receptor (NSTR), which is also a neuropeptide receptor. Figure 2.10 shows the electrostatic surface of neurotensin receptor with $neurotensin_{8\text{-}13}$. The PDB ID is 4GRV. The electrostatic surface of the

binding pocket of neurotensin and neurotensin receptor is shown in the Fig. 2.10. In Fig 2.10, the positive charges are shown in blue, the negative charges are shown in red. Figure 2.10B is the top view of neurotensin receptor's binding pocket. It shows the neurotensin receptor binding pocket carries more positive charges. Figure 2.10C is a clip of the electrostatic surface of the binding pocket and shows neurotensin's electrostatic surface. The neurotensin$_{8\text{-}13}$ C-terminus carries the negative charge. Hence, the backbone of the Leu 13 carboxylic group from neurotensin$_{8\text{-}13}$ inserted into the binding pocket. The binding pocket charges of neurotensin receptor is opposed to proctolin receptor binding pocket charges.

Both two neuropeptides and the two neuropeptide receptors have charged dock. Both neuropeptides have the charged residues as well. Even there are different partial charges in the receptor binding pocket, the neuropeptides carry the opposed partial charges with the receptor binding pocket. This allows the neuropeptide to strongly bind to the receptor.

Materials and Methods

Proctolin receptor and proctolin docking

The structure of proctolin receptor was built by using GPCR I-TASSER (Iterative Threading ASSEmbly Refinement) server (Roy et al., 2010; Yang et al., 2015a; Yang and Zhang, 2015; Zhang et al., 2015). The C-terminus of proctolin receptor was removed. The homology model of proctolin receptor was used to dock with proctolin. The Induced Fit Docking (IFD) method (Sherman *et al.*, 2006) was used to obtain the preliminary results. Then the preliminary results from the IFD were used to dock by QM-Polarized Ligand Docking method. All the docking process were implemented in Maestro software (Schrodinger Release 2019-3).

The molecular dynamic simulations study of the proctolin receptor docking model

We used the previous docking model of proctolin and proctolin receptor to continue the simulation study. CHARMM-GUI was used to embed the docking model of proctolin receptor with proctolin into the 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer environment, and the length of X and Y are 80 Å which is perpendicular to the Z-direction, the length of Z is 132 Å (Wu et al., 2014) (shown in Figure 2.6). The Cl^- ions were added to generate a neutral system. Amber17 was used to run the solvated structures by MD simulations and AMBER force field, ff14SB was applied for the MD simulations (Maier *et al.*, 2015). The particle mesh Ewald method was used to calculate the long-range electrostatic interactions *(Petersen, 1995)*. First, the system was minimized in 5000 steps with positional restraints for proteins, ligand, and lipid head groups. The lipid structures were keeping dihedral restraints. Then, the system was equilibrated for 50 ps at 303.15 K (canonical ensemble) and 325 ps (isothermal-isobaric ensemble). The temperature was controlled by using a Langevin thermostat with a friction coefficient 1.0 ps^{-1}, the pressure was controlled by Monte-Carlo barostats. The time step of this simulation was 0.002 ps, the total production time was 1.5 μs.

MD simulations trajectory analysis

VMD software (Humphrey et al., 1996) and CPPTRAJ (Roe and Cheatham III, 2013)was used to analyze the MD trajectory. Xmgrace was used to process and present the simulations distance study of the two cation-pi interactions. Figures were processed and presented by ChimeraX (Goddard et al., 2018; Pettersen et al., 2021).

Figures

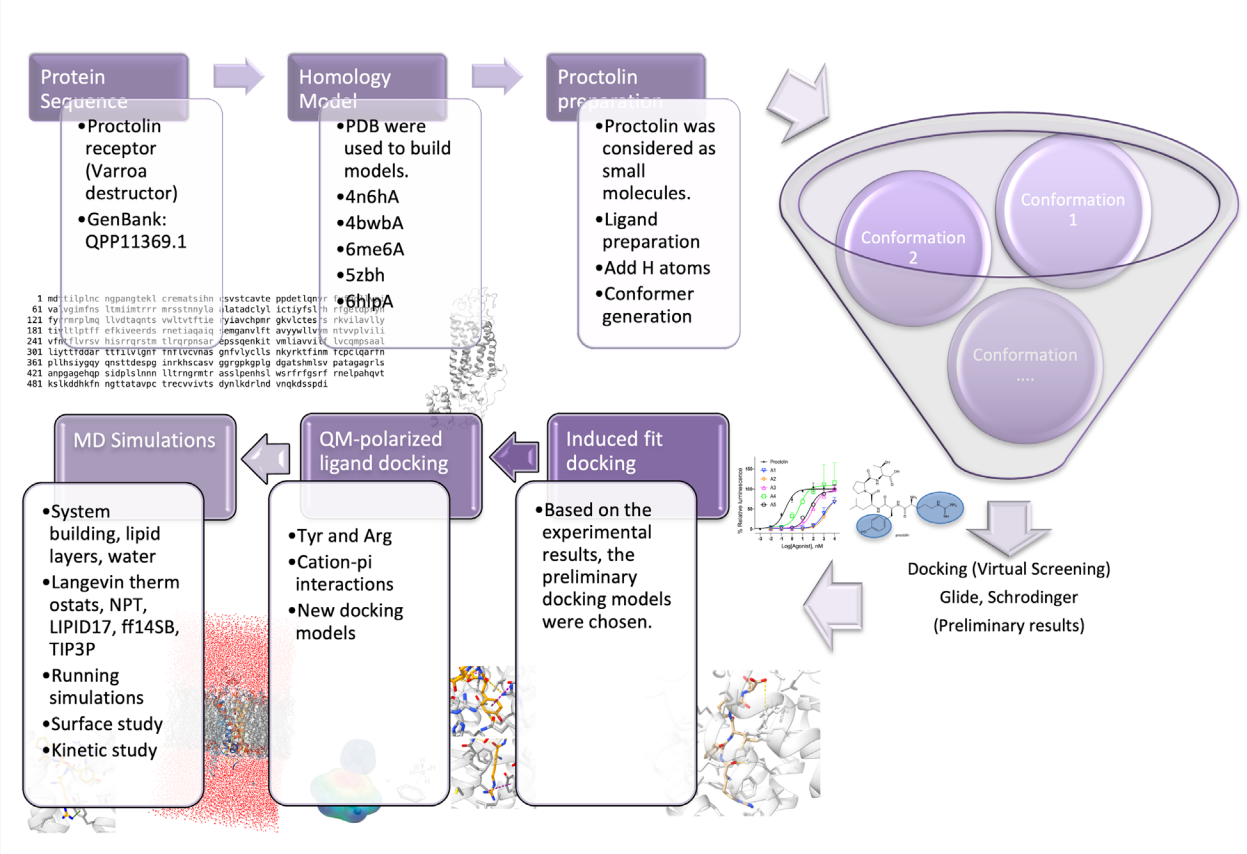


Figure 2.1 The workflow of the proctolin receptor model building, proctolin docking and MD simulations study.

The work started from the homology model building by GPCR-I-TASSER. The proctolin receptor protein sequence was originally obtained from Dr. Park's lab. Now, the protein sequence was reported to GenBank, the GenBank ID is QPP11369.1. The crystal structures were used as templates. The PDB IDs are 4N6H, 4BWB, 6ME6, 5ZBH, and 6HLP. Then, virtual screening, induced fit docking, and QM-polarized ligand docking were performed. The last step was MD simulations to obtain the interactions between proctolin and proctolin receptor.

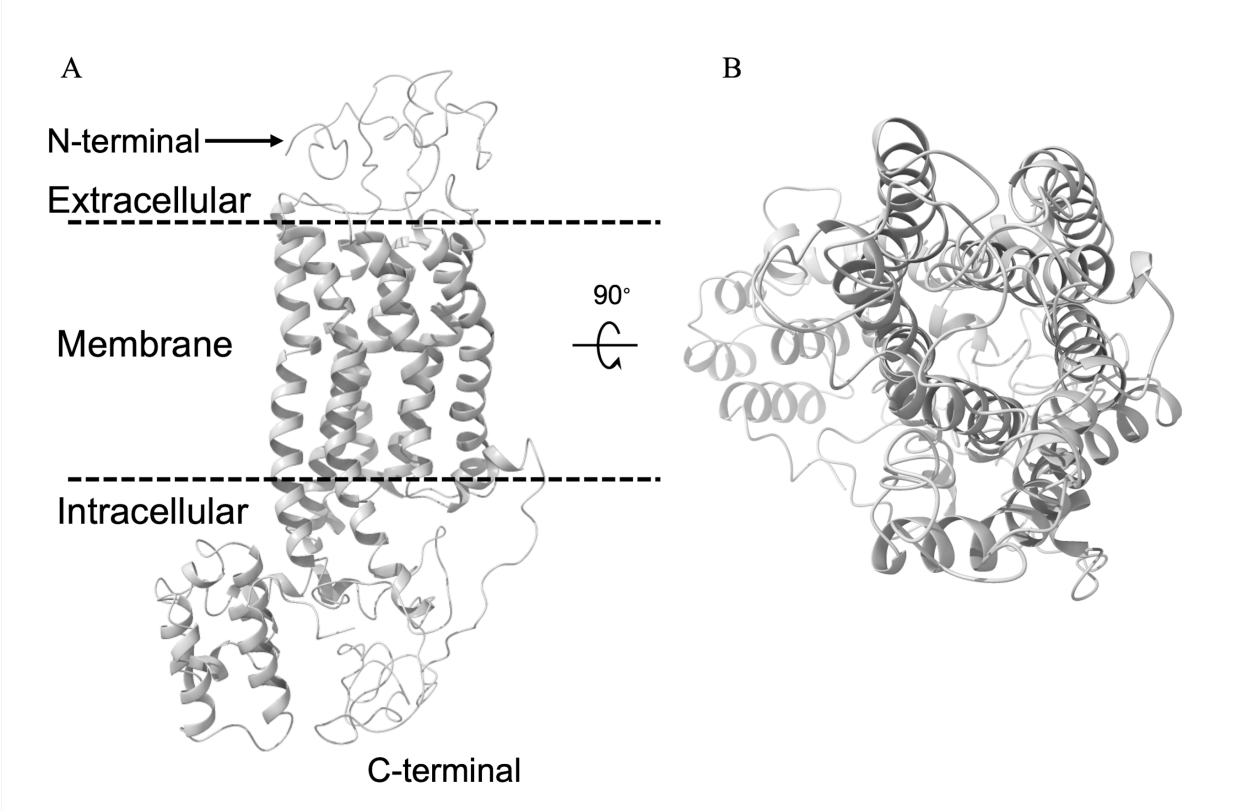


Figure 2.2 The proctolin receptor homology model built by GPCR-I-TASSER.

A. It shows the proctolin receptor homology model. The N-terminus is in the extracellular. The seven transmembrane helices are in the membrane. The C-terminus is in the intracellular. B. shows the top view from the N-terminus of proctolin receptor.

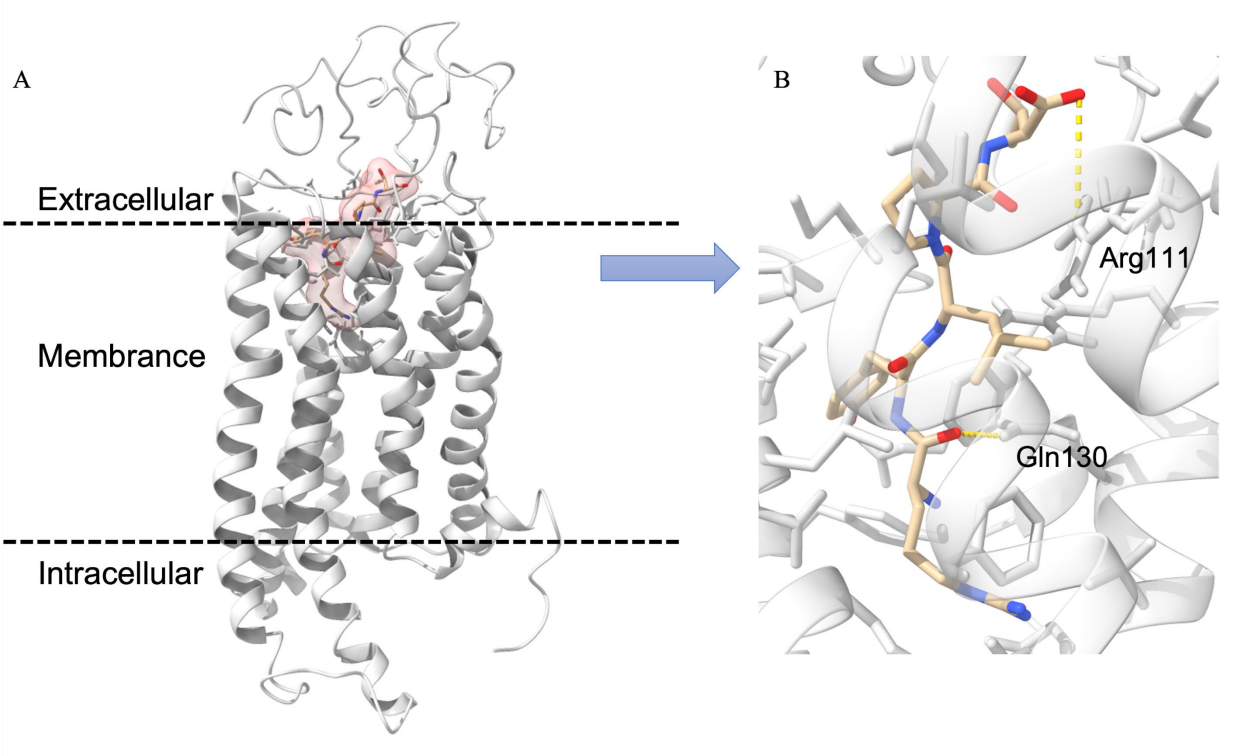


Figure 2.3 The virtual screening results of proctolin and proctolin receptor.

A. This is the preliminary docking results of proctolin and proctolin receptor. The red surface shows the proctolin surface. The tan color shows the proctolin. B. The interactions between proctolin and proctolin receptor do not match with the experimental results.

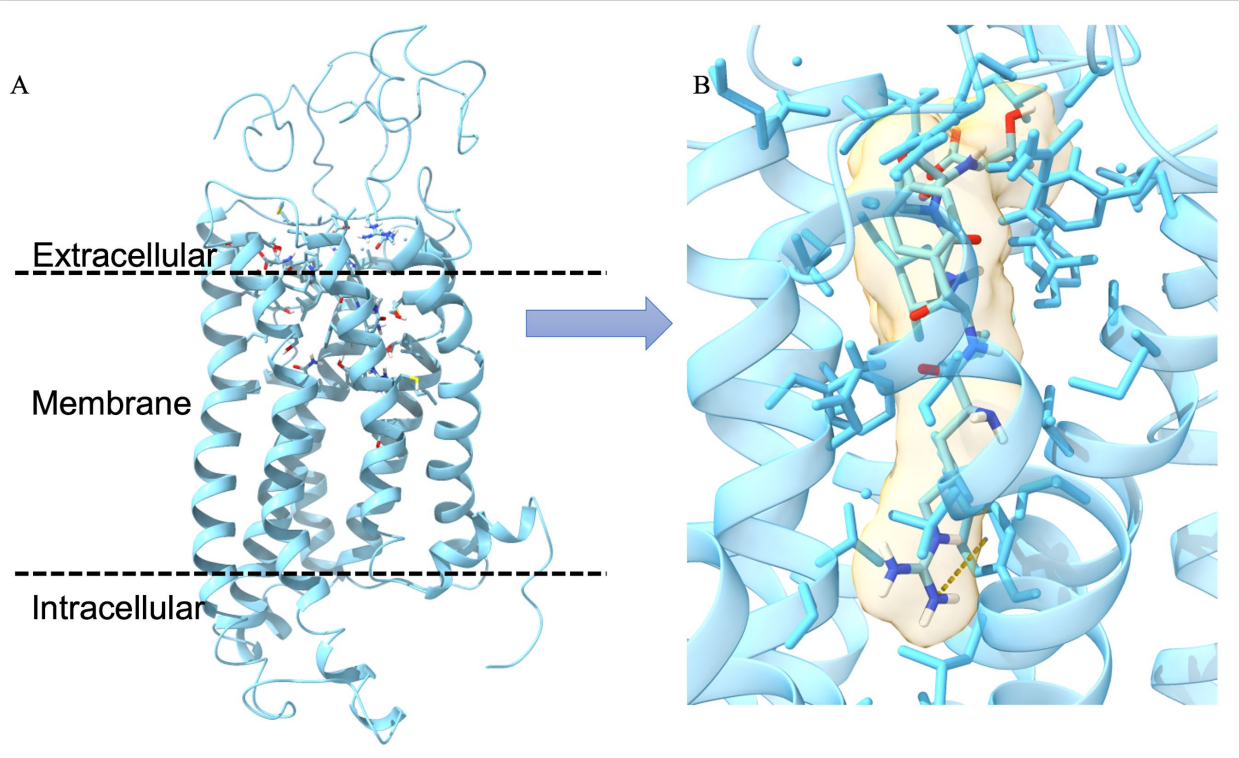


Figure 2.4 The induced fit docking results of proctolin and proctolin receptor.

The light orange surface is showing proctolin surface. The yellow dashed line shows the hydrogen bond.

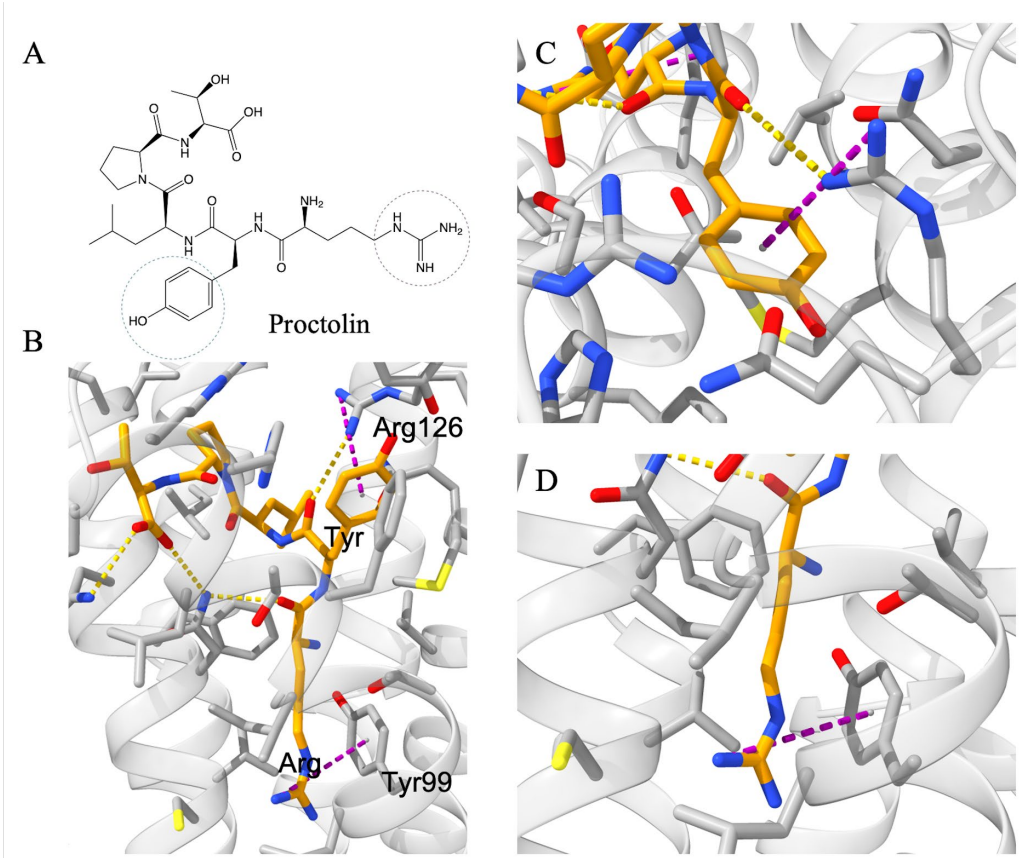


Figure 2.5 The QM-polarized ligand docking results of proctolin and proctolin receptor.

A. shows the chemical formula of proctolin. The two dashed line circles label the first two important residues from proctolin. B. shows the QM-polarized ligand docking result of proctolin and proctolin receptor complex. C. The zoomed-in view of the interaction between Tyr2 and Arg126. D. The zoomed-in view of the interaction between Arg1 and Tyr99. The dark magenta color dashed lines show the cation-pi interactions. The yellow dashed lines show the hydrogen bonds. The hydrogen atoms are hidden.

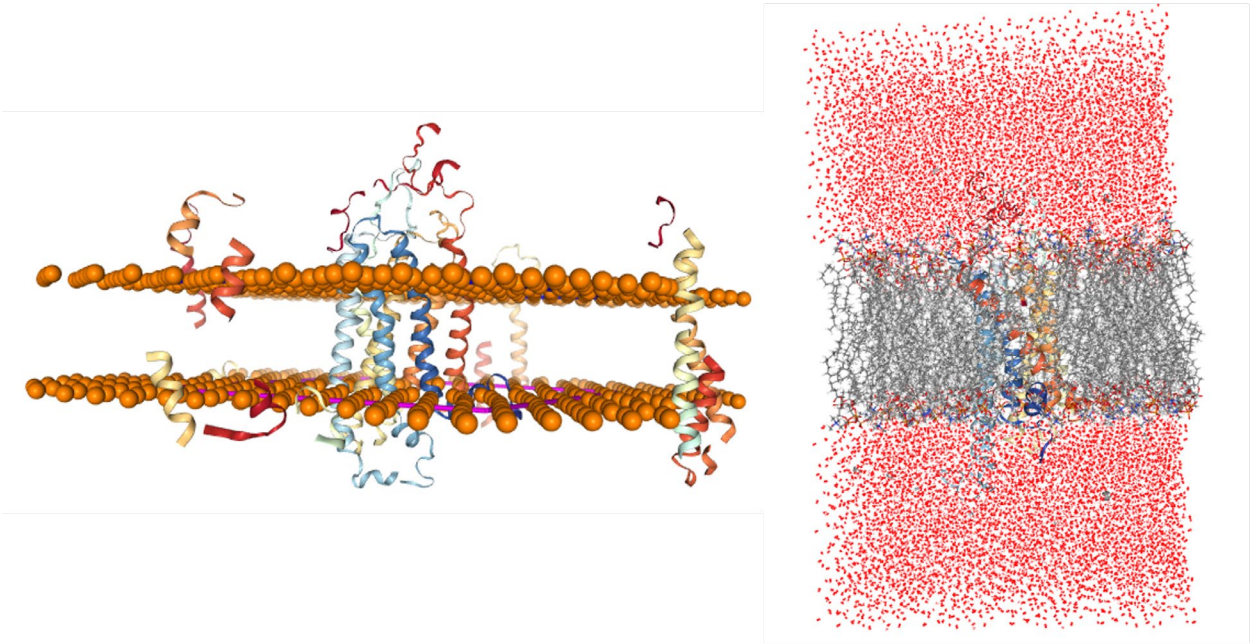


Figure 2.6 The MD simulations system of the proctolin and proctolin receptor complex.

Both figures are showing the bilipid layers, and the complex in the solvent system.

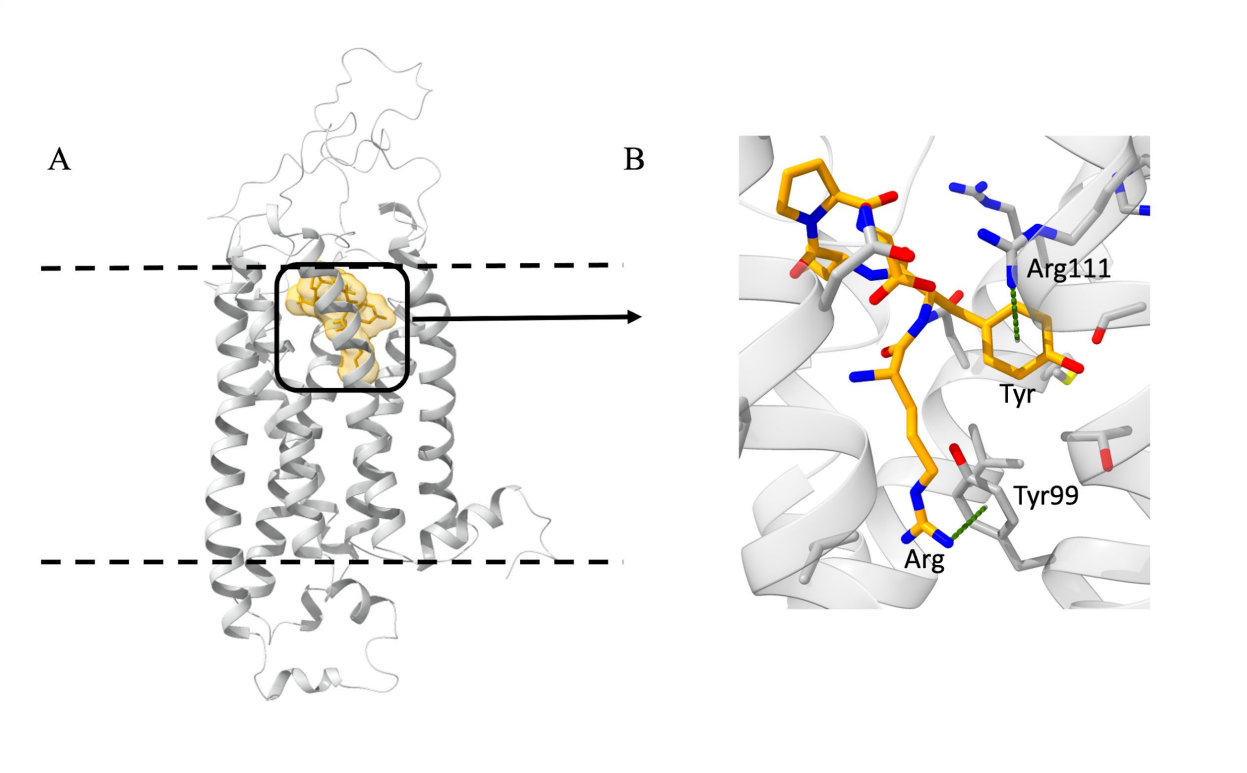


Figure 2.7 The interactions between proctolin and proctolin receptor after running MD simulations.

A. shows the complex of proctolin and proctolin receptor, which is the average structure of 1000ns MD simulations. The orange color shows the proctolin surface. B. shows the zoomed-in view of the binding pocket. The dark green dashed lines show the cation-pi interactions between proctolin and proctolin receptor. The first Arg from proctolin form the first cation-pi interaction with Tyr99 from proctolin receptor. The second Tyr from proctolin forms the second cation-pi interaction with Arg111. The dark yellow sticker shows the proctolin molecule. Nitrogen atoms are shown in blue. Oxygen atoms are shown in red. The dark green dashed lines show the cation-pi interactions.

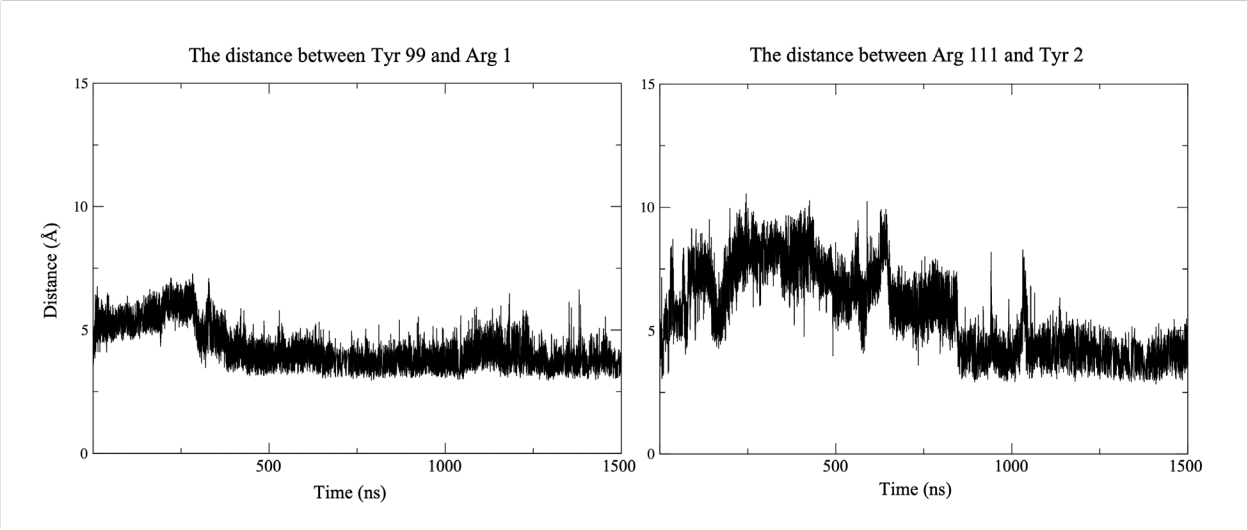


Figure 2.8 The distance between Tyr 99 from proctolin receptor to Arg 1 from proctolin, the distance between Arg 111 from proctolin receptor to Tyr 2 from proctolin.

The distance between cation and pi can be up to 5Å in cation-pi interactions. Left: The figure shows the distance of Tyr 99 from proctolin receptor and Arg 1 from proctolin. Right: The distance Arg 111 from proctolin receptor and Tyr 2 from proctolin. The left figure shows the cation-pi interaction of Tyr 99 and Arg 1 started forming around 480 ns. The right figure shows the cation-pi interaction of Arg 111 and Tyr 2 started forming around 800 ns.

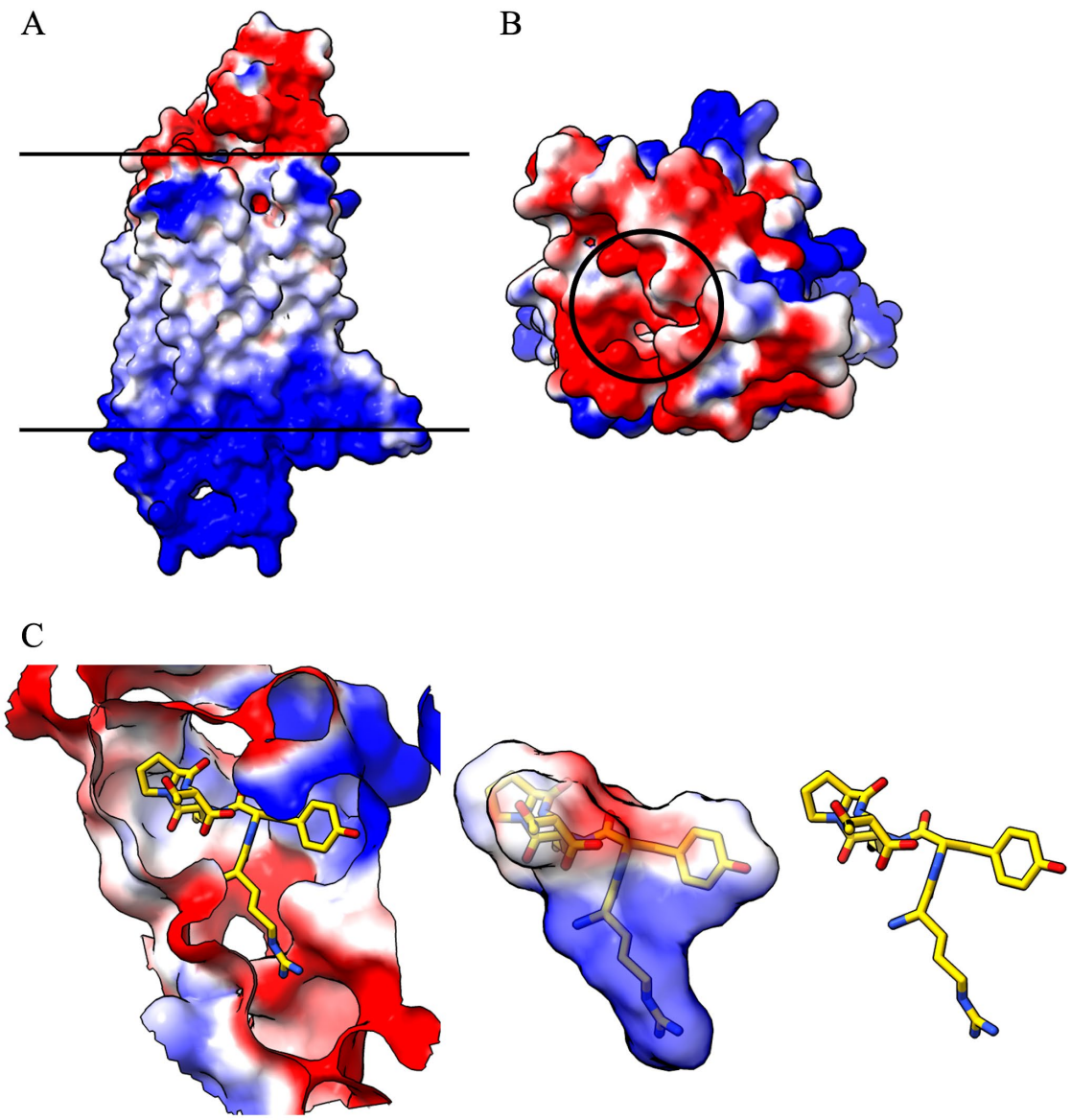


Figure 2.9 The electrostatic surface of proctolin and proctolin receptor.

A. shows the side-view of the surface of the complex of proctolin and proctolin receptor. Blue color shows the positive charges. Red color shows the negative charges. B. shows the top-view of the proctolin receptor. The black circle shows the area of the binding pocket. In the binding pocket, it is carrying negative charges. C. shows the clip view of the binding pocket. Yellow color shows the proctolin in stick form. In the middle of Figure 2.8C is the surface of proctolin. It shows the N-terminus of proctolin carries positive charges. It opposes with the charge of the

binding pocket of proctolin pocket. The proctolin acts like an anchor. The binding pocket of proctolin receptor offers the negative dock and lets the proctolin dock into the pocket and form the interactions.

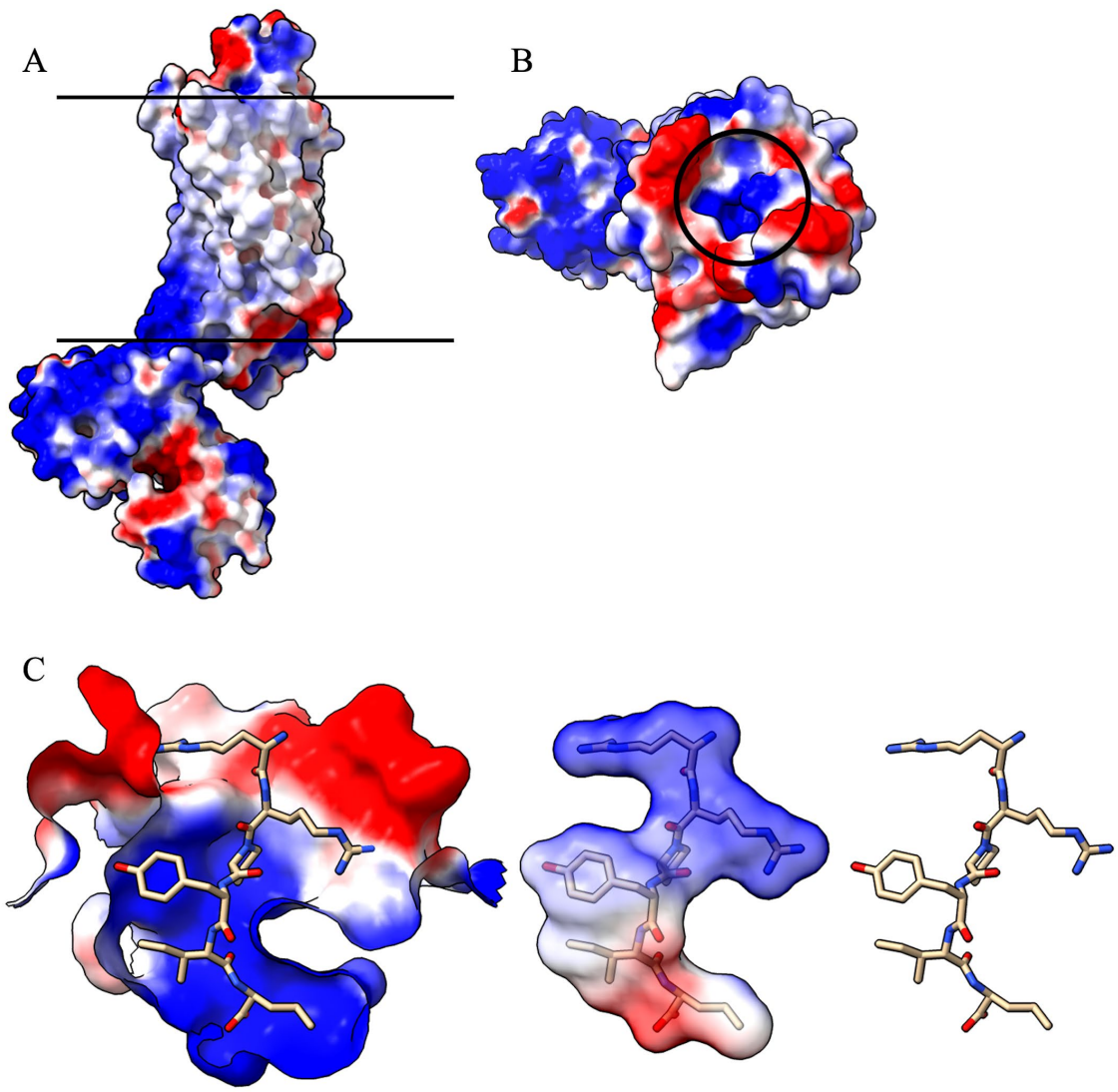


Figure 2.10 The electrostatic surface of the neurotensin receptor binding pocket.

The binding structure of neurotensin$_{8\text{-}13}$ and neurotensin receptor is from the crystal structure of neurotensin receptor. The neurotensin$_{8\text{-}13}$ is N-RRPYIL-C. The PDB ID is 4GRV. In the electrosurface figures, the positive charges are shown in blue, and the negative charges are shown in red. Figure. A shows the electrosurface binding pocket of neurotensin. In the neurotensin receptor binding pocket, the positive charges are inside, the negative charges are around the top part of the pocket. Figure. B shows the electrosurface of neurotensin. The C-

terminus of neurotensin is negative. The N-terminus of neurotensin is positive. Figure. C shows the conformation of neurotensin.

Chapter 3 - The behavior of red fluorescent protein chromophore at room temperature

Summary

Fluorescent proteins have been improved from natural fluorescent proteins and heavily used in life science as protein labels, markers of gene expression, and living-cell imaging. The characteristics of fluorescent protein chromophores can give us information about the structure-function relationship with the protein matrix. This can guide us to engineer tuned color variants and broaden the spectral range of useful proteins. Here, we present the behavior of the chromophore with the protein matrix and develop a deeper understanding of fluorescent proteins. We performed molecular dynamics simulations on four trans-form fluorescent proteins. All fluorescent proteins with the trans-form chromophores tend to be non-planar which has been rarely caught by experimental methods. After the planarity changed, the interactions of chromophore and protein matrix have changed as well. The residues 67, 92, 143, and 197 are identified as more important sites for the fluorescent proteins. These site residues interact with the chromophore.

Introduction

Fluorescent protein (FP) was first found from the jellyfish *Aequoria Victoria*. The color is green. Later, there were aquatic species, such as sea anemones, and coral found displaying different visible fluorescence and colors. The green fluorescent protein (GFP) is monomeric, soluble, stable, and brightly fluorescent. It is used in life science research as protein labels, markers of gene expression, and living-cell imaging, and was hailed as a revolution of fundamental biomedical research (Chalfie et al., 1994; Shaner et al., 2007). Since then, fluorescent proteins have been engineered to meet different requirements. Different engineered GFPs or GFP homologs have been reported which cover almost the entire visible range of emission wavelengths (Gurskaya et al., 2001; Matz et al., 1999; Tsien, 1998). Some other variants have been optimized to fit different bio-conditions such as different pH (Miesenböck et al., 1998), and redox potential (Østergaard et al., 2001).

Engineering fluorescent protein allows for fine-tuning the photophysical properties of blue to yellow variants derived from the *Aequoria Victoria* jellyfish and the improvement of monomeric fluorescent proteins from other organisms, such as the yellow-orange to far-red regions of the visible light spectrum. Table 3.1 shows different fluorescent proteins and their basic information. Specifically, different color fluorescent proteins are very useful to use in multicolor labeling or other experiments. The red-emitting range has been extended with the naturally red fluorescent protein such as DsRed (Disc Red) from the corallimorphian *Discosoma sp* (Matz *et al.*, 1999) and eqFP611 from the sea anemone *Entacmaea quadricolor* (Wiedenmann et al., 2002). However, there is still limitations of DsRed as a fluorescent marker, such as slow chromophore maturation and oligomeric nature. The improvement of DsRed has produced different variants and obtained better fluorescent proteins with better properties (Bevis and

Glick, 2002; Campbell et al., 2002). It is important to understand the interests of the relationship of the structural basis and the spectral properties. Pervious study shows some features that affect the spectral properties, such as the extended conjugated pi-system. The fluorescent protein Rtms5, derived from Motipora efflorescens, exhibits different properties despite having the same chromophore sequence as DsRed due to conformation differences in the chromophore. The phenoxy ring of the chromophore is in a *trans* and non-coplanar conformation (Prescott et al., 2003).

Fluorescent proteins typically have 11-stranded β-barrel and chromophore structures (shown in Figure 3.1A). Different fluorescent proteins have different chromophores, which show different color. The red fluorescent proteins like DsRed, has red tyrosine-based chromophores (shown in Figure 3.1D). The autocatalyzed reaction produces an N-acylimine derivative. This extended the pi-conjugation on the chromophore and shifted the absorbance and emission spectra to red (Yarbrough et al., 2001). Several studies have shown the fundamental origins and manipulation of the emission color (Remington, 2006; Shaner *et al.*, 2007). The different chromophores shown in Figure 3.1B, 3.1C, 3.1D, and 3.1E. Figure 3.1B is green fluorescent protein chromophore, CRO. Figure 3.2C is cyan fluorescent protein chromophore, SWG. Figure 3.1D is red fluorescent protein chromophore, NRQ. Figure 3.1E is yellow fluorescent protein chromophore, CR2. The local environments of the chromophore are very critical to the intensity of the fluorescence, and emission wavelengths, which includes the chromophore interactions with the protein matrix, and the extent of pi conjugation of the chromophore. These variables can cause blue or red spectral shifts in absorption and emission and the maxima shift can reach as much as 40 nm (Shaner *et al.*, 2007).

The characteristics of fluorescent protein chromophore can give us the information about the structure-function relationship with protein matrix. This can guide us to engineer tuned color variants and extending spectral range of proteins. We studied the trans-form chromophore red fluorescent proteins, mRuby, TagBFP, TagRFP, and LSSmKate2, to obtain a deeper understanding the interactions of chromophore affections of these variables to the fluorescence intensity.

Results

TagBFP

TagBFP is one of the blue fluorescent proteins. It is a monomeric fluorescent protein. The excitation peak is 402 nm, and emission peak is 457 nm. It has 0.63 high quantum yield (QY) (shown in Table 3.1). Figure 3.2C and 3.2D show the crystal structure of TagBFP (PDB ID: 3M24). Both Figure 3.2C and 3.2D show the conformation of the chromophore NRP. The trans-form and nonplanar chromophore has been captured. It also has a high quantum yield. Figure 3.2D is the top view of TagBFP. It clearly shows the interactions between the chromophore NRP and the protein matrix. The two light blue dashed lines show the pi-stacking. The dark green dashed line shows the cation-pi interaction. The magenta dashed line shows the salt bridge. Figure 3.3A shows the dihedral angle that was picked to study the chromophore. In the crystal structure, the dihedral angle is around 67.5º. This is different from typical chromophore planar angles. The crystal structure shows the Tyr 197 and Phe 143 form pi-stacking with the chromophore, Arg 92 forms a cation-pi interaction and a salt bridge. We utilized this crystal structure to study this fluorescent protein and gain a better understanding of chromophore behavior. Hence, we perform MD simulations at 300K. Figure 3.3D shows the molecular dynamics simulations. After 2000ns simulation time, the chromophore still forms pi-stacking

with Tyr 197 and Phe 143. The chromophore lost the cation-pi interaction with Arg 92. There is still a salt bridge between Arg 92 and chromophore. There are differences after the simulations. The dihedral angle is rotated. Hence, Lys 67 forms a cation-pi interaction with the chromophore and a salt bridge. Figure 3.4 E shows the planar angles of the chromophore at the last 400 ns simulations, which is between 1600-2000 ns. In this figure, it shows the chromophore keeps rotating, the rotation angle is between 42º and 70º. The most planar angle shown is around 50º. This means the chromophore can form a cation-pi interaction and a salt bridge with Lys67, and it can form a salt bridge with Arg 92 as well. This simulations result shows the conformations of the chromophore tended to be different from the crystal structure. In the TagBFP system, the two pi-stacking interactions extend the pi-conjugation, two salt bridges stabilized the chromophore. All this may increase the quantum yield.

TagRFP

TagRFP is one of the red fluorescent proteins which has a trans-form of chromophore. Figure 3.4B shows the TagRFP crystal structure (PDB ID: 3M22). The excitation peak is 555 nm, the emission peak is 584 nm. The quantum yield is 0.48 (shown in Table 1). Figure 3.4C shows the chromophore NRQ crystal conformation. The planar chromophore was captured. Comparing with TagBFP, TagRFP has a lower quantum yield. Figure 3.2B shows the top view of TagRFP. His 197 forms pi-stacking with the chromophore. Asn 143 and Ser 158 form two hydrogen bonds with the chromophore to stabilize the chromophore (shown in Figure 3.4C). However, in the MD simulations result, the chromophore planarity changed. Figure 3.4D shows the dihedral angle rotates and the chromophore is non-planar. There are three water molecules that form three hydrogen bonds with the phenolic oxygen atom. There are two water molecules that form two hydrogen bonds with the oxygen of the imidazole ring. These hydrogen bonds

stabilized the chromophore. Figure 3.4E shows the planar angle's changes, it is between 26º and 57º which is between 1600-2000 ns simulations.

LSSmKate2

LSSmKate2 is another red fluorescent protein. It also has a trans-form and planar chromophore captured by X-ray crystallology (shown in Figure 3.5A). Figure 3.5B shows the LSSmKate2 crystal structure. Similar with other fluorescent proteins, it has a similar conformation of the beta-barrel. Figure 3.5C shows the interactions of the chromophore and the protein matrix. The Arg 197 forms a cation-pi interaction with the chromophore. The Ser 158 and a water molecule form two hydrogen bonds with the chromophore. It is similar to TagRFP. After MD simulations, the dihedral angle of the chromophore turned. The Arg 197 still forms a cation-pi interaction with the chromophore. Arg 92 forms a hydrogen bond with the phenolic oxygen atom. There are three water molecules that formed hydrogen bonds with the phenolic oxygen atom. Figure 3.5E shows the planar angle of the chromophore during the last 400 ns of the simulations, 1600-2000 ns. The changes of the planar angle are around 38º - 67º.

mRuby

mRuby is one of the red fluorescent proteins, which has been captured as a trans-form and planar chromophore. The excitation peak of this protein is 558, the emission peak is 605. The quantum yield is 0.35. Figure 3.6C shows the interactions between chromophore and protein matrix. His 197 forms pi-stacking with chromophore. Asn 143 and Thr 158 form two hydrogen bonds with the phenolic oxygen atom from the chromophore. Arg 92 forms a hydrogen bond with the oxygen of the imidazole ring from the chromophore. After running MD simulations, the chromophore rotated. Trp 140 forms pi-stacking with the chromophore. Arg 92 forms cation-pi interactions with the chromophore. There are four water molecules that form hydrogen bonds

with the phenolic oxygen atom. Figure 3.6E shows the changes of planar angle during the last 400 ns MD simulations, 1600-2000 ns. The range of the planar angle changes is around 35º - 65º.

Discussion

Since fluorescent proteins were first found, they have been studied to extend the fluorescence ranges and increase the fluorescence intensity. In order to enhance their utility, engineering of the fluorescent proteins has been worked on. The local environments of the chromophores have been shown to have a huge effect on the intensity of the fluorescence, and emission wavelengths (Shaner *et al.*, 2007).

The trans-form chromophore fluorescent proteins are always more interesting than cis-form chromophore fluorescent proteins. Most of the time, the trans-form fluorescent proteins were captured as a planar chromophore by X-ray crystallography. Crystal structures are only showing one conformation and cannot show all the protein's conformations. Hence, we performed MD simulations to study trans-form fluorescent proteins and to identify the properties of local environments and chromophore. From this MD simulation study, we can easily find that the trans-form chromophores tend to be non-planar at 300 K. This is different from the crystal structures. The trans-form chromophores seem to wiggle at certain angles. The chromophore of TagBFP was non-planar when it was captured by X-ray crystallography. The planar angle is 67.5º. Since the chromophore is non-planar, the phenol ring faces to the Tyr 197 and Phe 143 and form the two pi-stacking interactions, and it also makes the chromophore face to Arg 92 to form a cation-pi interaction. These main three interactions also stabilized the chromophore. TagBFP also shows similar results from MD simulations. The differences of the TagBFP chromophore interactions with protein matrix is Lys 67 forms the cation-pi with the chromophore instead of Arg 92. Even though TagBFP chromophore interacts with different

residues, the chromophore still has two pi-stacking and one cation-pi interactions. The two pi-stacking extends the pi-conjugation of the chromophore. This may increase the quantum yield.

TagRFP is another trans-form fluorescent protein. The chromophore was captured planar by X-ray crystallography. His 197 forms pi-stacking with the chromophore which can extend the pi-conjugation. Asn 143 and Ser 158 form two hydrogen bonds with the chromophore. However, the MD simulations results show the planar angle turned and is wiggling between 42º - 58º. Since the chromophore planar angle turned, three water molecules form three hydrogen bonds with the phenolic oxygen atom. After the planar angle turned, the chromophore formed more hydrogen bonds with water, this stabilizes the chromophore.

LSSmKate2 is a long Stokes shift fluorescent protein. The crystal structure shows the Arg 197 forms a cation-pi interaction with the chromophore. Ser 158 and one water molecule forms two hydrogen bonds with the chromophore. After MD simulations, the two ring's dihedral angle of the chromophore turned. Arg 197 still forms a cation-pi with the chromophore. There are three water molecules forming three hydrogen bonds and Arg 92 forms a salt bridge with the chromophore.

mRuby is also a trans-form fluorescent protein. The crystal structure shows His 197 forms pi-stacking, Asn 143 and Thr 138 form two hydrogen bonds with the phenolic oxygen atom. Arg 92 hydrogen bonds with the oxygen from the imidazole ring from the chromophore. After MD simulations, the dihedral angle of the two rings also turned. The interactions of the chromophore and protein matrix are changed. Trp 140 forms pi-stacking and Arg 92 forms a cation-pi interaction with the chromophore. There are three water molecules that form hydrogen bonds.

The planar trans-form fluorescent protein crystal structures are different from MD simulations results. The planar trans-form chromophores tend to be non-planar at 300 K. The non-planar chromophores have different interactions with the protein matrix. This has been ignored. Here, this study pointed out the trans-form chromophores of crystal structures interactions can't present the real interactions of chromophores. Also, there are some important residue sites. Residue site 197 can interact with the chromophore. Histidine can form pi-stacking, Arginine can form cation-pi interaction. Residue 143 can also interact with the chromophore. Residues 67 and 92 also can interact with the chromophore to form either hydrogen bonds or cation-pi interactions. This shows that care must be taken in using the crystal structures of fluorescent proteins and MD simulations can be used to find real interactions and the interaction residue sites. These results can help to engineer new fluorescent proteins with random mutations.

Materials and Methods

The fluorescent proteins' preparation

Protein preparation

The crystal structures of red fluorescent proteins were obtained from Protein Data Bank. The water was removed beyond 3.0 Å. All the proteins were prepared by Schrödinger Protein Preparation Wizard (Sastry et al., 2013).

The molecular dynamic simulations study of the red fluorescent proteins

After the structures were prepared by Schrödinger Protein Preparation Wizard (Sastry *et al.,* 2013), each fluorescent protein system was built by Desmond system builder (Schrödinger Release 2021-4). The TIP3P water molecules were added. Sodium chloride was used to neutralize the system, 0.15M sodium chloride was added into the system. The water box shape

was orthorhombic and the size of the box was 10 × 10 × 10 Å. OPLS4 was used for the force field (Jorgensen and Tirado-Rives, 1988; Lu et al., 2021). The temperature was 300 K. The total simulation time is 2000 ns on each protein.

MD simulations trajectory analysis

The MD trajectory was analyzed by Schrödinger Desmond (Schrödinger Release 2021-4). The planar angles were analyzed by the MD simulations trajectory event study.

All data was processed by Schrödinger Desmond (Schrödinger Release 2021-4). All of the figures were presented by UCSF ChimeraX version is 1.1 (2020-10-07) (Goddard *et al.*, 2018).

Figures

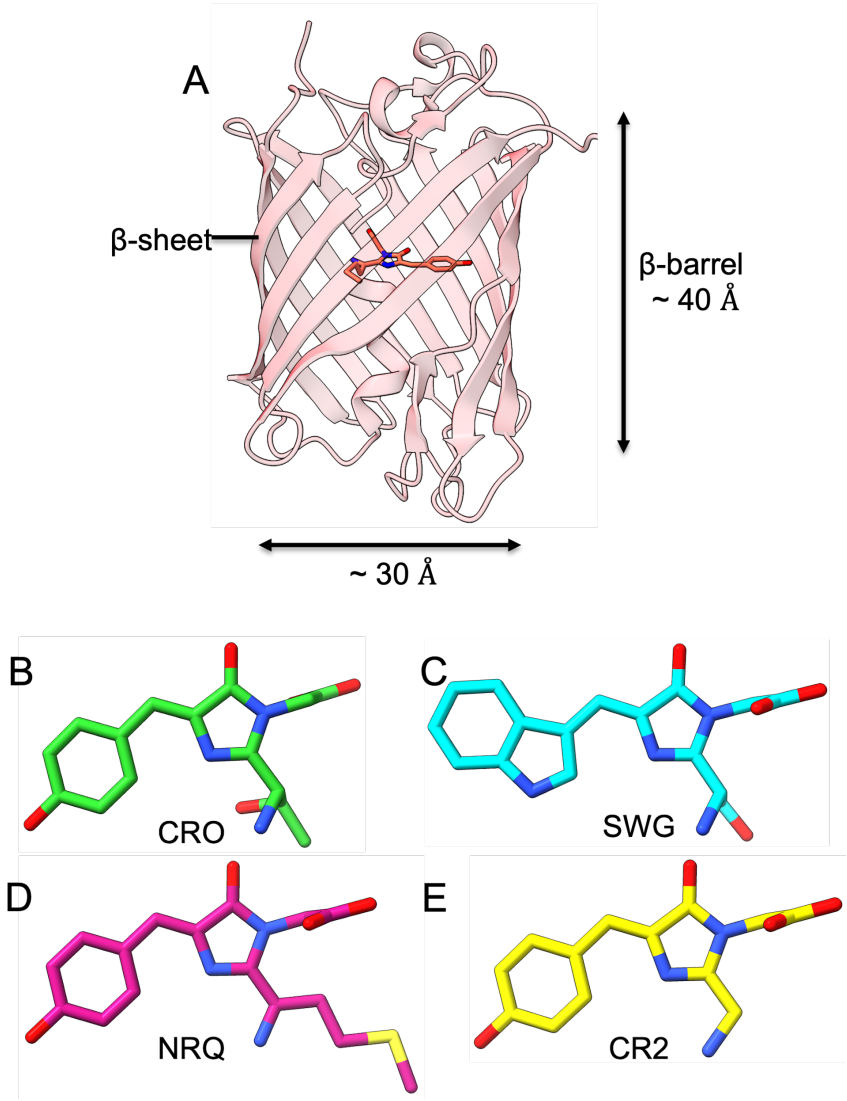


Figure 3.1 The structure of fluorescent protein and different color fluorescent proteins' chromophores.

A. Shows the structure of red fluorescent protein (PDB ID: 3M22). Fluorescent proteins have a β-barrel and a chromophore. Normally, the size of β-barrel is around 30 × 40 Å. B. It shows the typical green fluorescent protein's chromophore, CRO. C. Shows the cyan fluorescent protein's chromophore, SWG. D. Shows the red fluorescent protein's chromophore, NRQ. E. Shows the yellow fluorescent protein's chromophore, CR2.

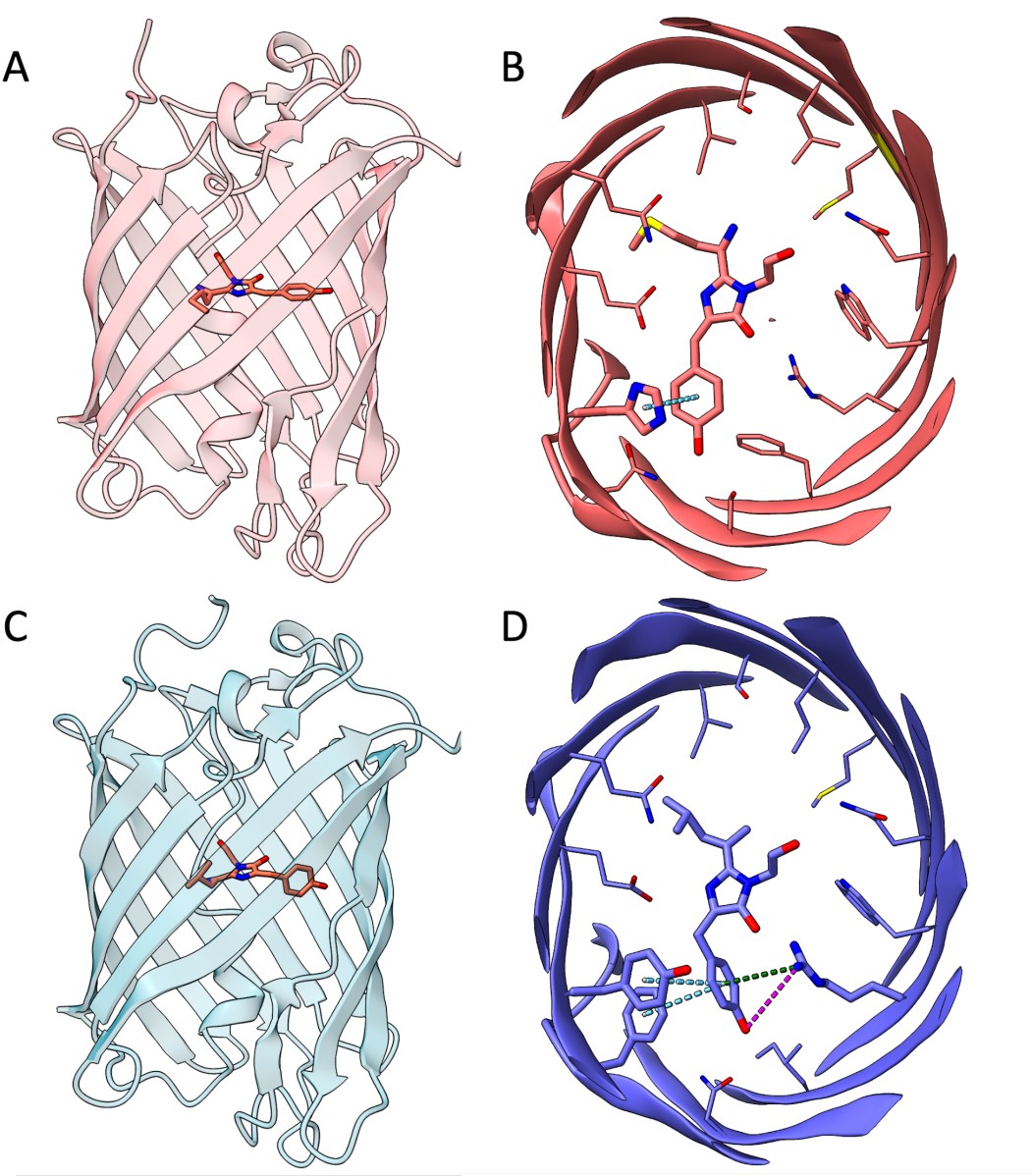


Figure 3.2 The blue and red fluorescent proteins and the chromophore interactions with protein matrix.

Figure 3.2A. and 3.2B. show the red fluorescent protein, TagRFP. The PDB ID is 3M22. Figure 3.2B is the top view of Figure 3.2A. The chromophore of TagRFP is flat. Figures 3.2C. and 3.2D. show the blue fluorescent protein, TagBFP. The PDB ID is 3M24. Figure 3.2 D is the top view of Figure 3.2C. The chromophore of TagBFP is not flat. The light blue dashed lines show

the pi stacking between the chromophore and protein matrix. The dark green dashed line shows the cation-pi interaction between the chromophore and protein matrix. The magenta dashed line shows the salt bridge between the chromophore and protein matrix. The chromophore of blue fluorescent protein has more interactions with protein matrix.

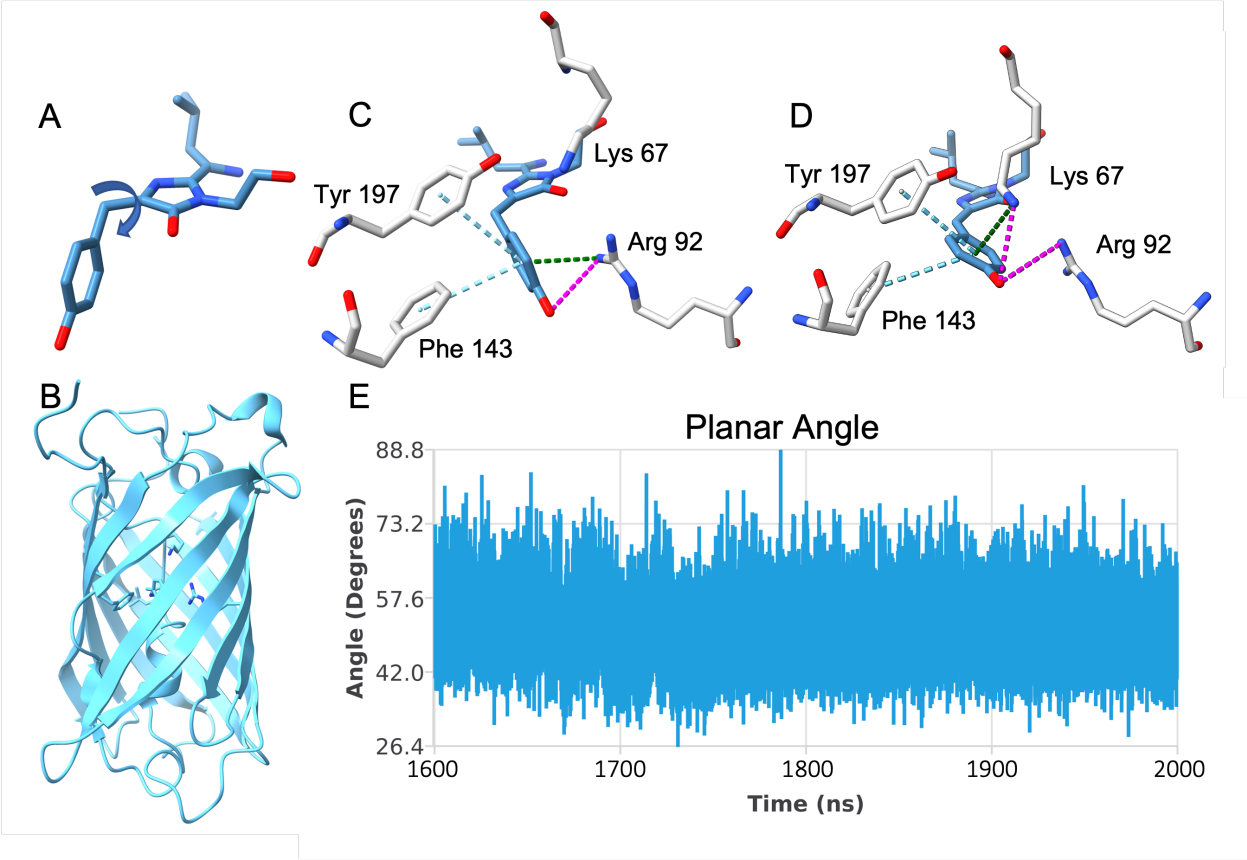


Figure 3.3 The conformations of TagBFP fluorescent protein.

A. The conformation of crystal structure of TagBFP chromophore (PDB ID: 3M24). The blue arrow shows the rotated bond, which results in the dihedral angle changing. B. The crystal structure of TagBFP fluorescent protein. C. The crystal structure of TagBFP chromophore. The light blue dashed lines show the pi stacking. The dark green line shows the cation-pi interaction. The magenta dashed line shows the salt bridge. The Tyr 197 and Phe 143 form two pi-stackings with the chromophore. The Arg 92 forms the cation-pi interaction and the salt bridge with chromophore. D. The most common conformation found in MD simulations. The Tyr 197 and Phe 143 form pi-stacking with the chromophore. Lys 67 forms a cation-pi interaction and a salt bridge interaction with the chromophore. Arg 92 forms a salt bridge with chromophore. E. The planar angle, which is the dihedral angle of the chromophore (shown in 3.4A). In this figure, the

dihedral angle (planar angle) is around 42º - 70º between 1600 ns – 2000 ns. The chromophores two rings dihedral angle prefers certain angles.

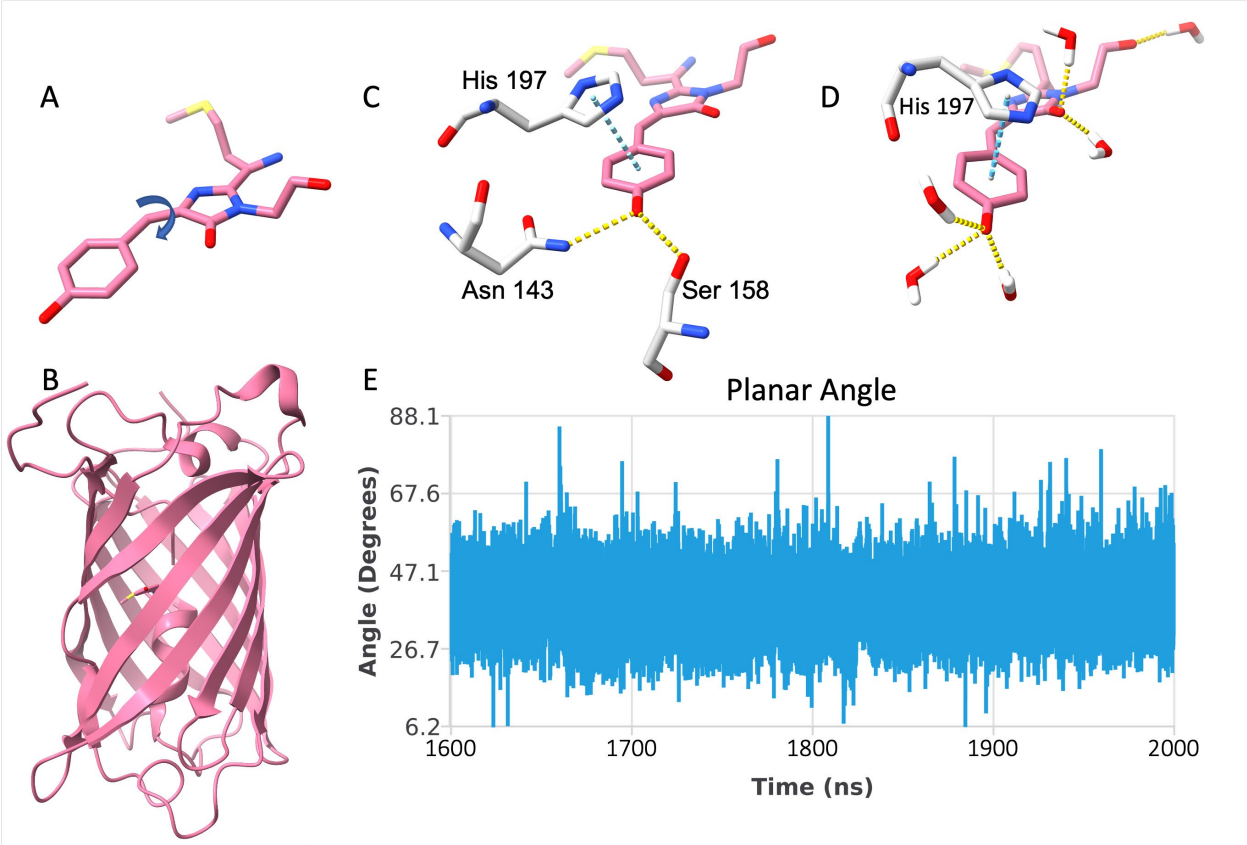


Figure 3.4 The conformations of TagRFP fluorescent protein.

A. The conformation of the crystal structure of TagRFP chromophore (PDB ID: 3M22). The blue arrow shows the rotated bond, which results in the dihedral angle changing. B. The crystal structure of TagRFP fluorescent protein. C. The crystal structure of TagRFP chromophore. The light blue dashed lines show the pi stacking. The yellow lines show the hydrogen bonds. His 197 forms pi-stacking with the chromophore. The Asn 143 and Ser 158 form two hydrogen bonds with the chromophore. D. The most common conformation shown in MD simulations. The His 197 forms pi-stacking with the chromophore. There are three water molecules that form hydrogen bonds with the phenolic oxygen atom from the chromophore. Two water molecules form hydrogen bonds with the oxygen of the imidazole ring. E. Shows the planar angle, which is

the dihedral angle of the chromophore (shown in 3.4A). In this figure, the dihedral angle (planar angle) is around 26º - 57º between 1600 ns – 2000 ns.

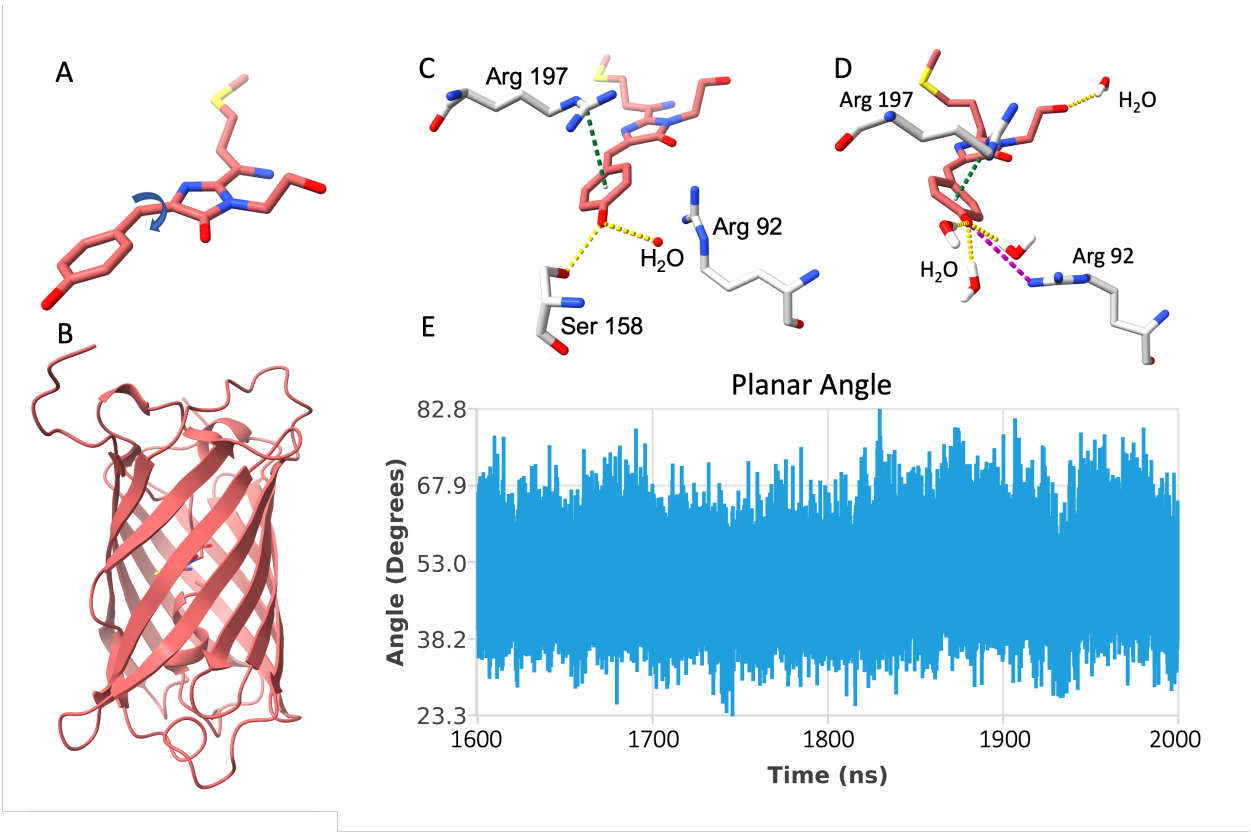


Figure 3.5 The conformations of LSSmKate2 red fluorescent protein.

A. The conformation of the crystal structure of LSSmKate2 chromophore (PDB ID: 3NT3). The blue arrow shows the rotated bond, which results in the dihedral angle changing. B. The crystal structure of LSSmKate2 fluorescent protein. C. The crystal structure of LSSmKate2 chromophore. The dark green dashed lines show the cation-pi interaction. The yellow dashed lines show the hydrogen bonds. Arg 197 forms a cation-pi interaction with chromophore. Ser 158 forms hydrogen bonds with the chromophore. One water molecule forms a hydrogen bond with the oxygen of the phenol from the chromophore. D. The most common conformation found in MD simulations. Arg 197 forms a cation-pi interaction with the chromophore. There are three water molecules that form hydrogen bonds with the oxygen of the phenol from the chromophore. The magenta dashed line indicates a salt bridge Arg 92 forms with the chromophore. E. Shows

the planar angle, which is the dihedral angle of the chromophore (shown in 3.5A). In this figure, the dihedral angle (planar angle) is around 38º - 67º between 1600 ns – 2000 ns.

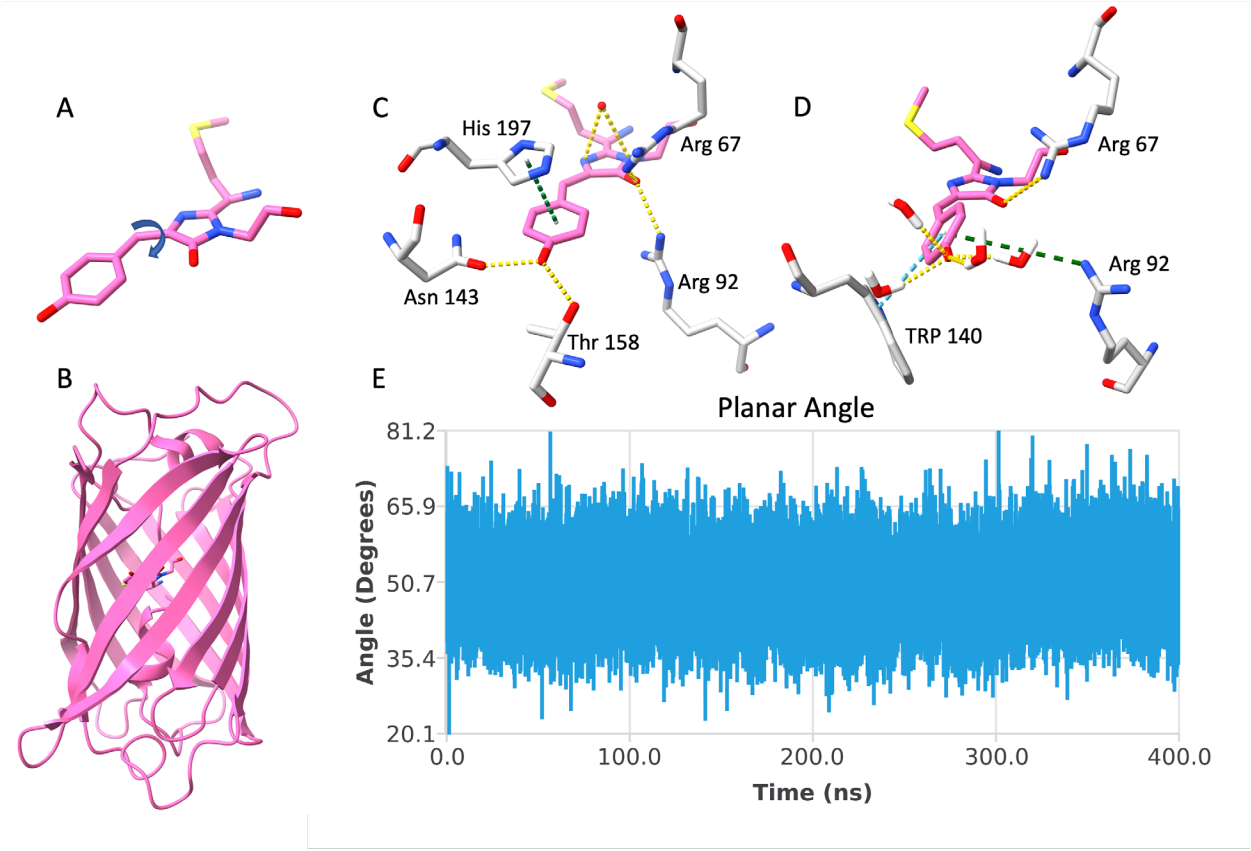


Figure 3.6 The conformations of mRuby fluorescent protein.

A. The conformation of the crystal structure of the mRuby chromophore (PDB ID: 3U0M). The blue arrow shows the rotated bond, which results in the dihedral angle changing. B. The crystal structure of mRuby fluorescent protein. C. The crystal structure of mRuby chromophore. The light blue dashed line shows the pi-stacking. The yellow lines show the hydrogen bonds. His 197 forms pi-stacking with chromophore. Asn 143 and Thr 158 form hydrogen bonds with the oxygen of the phenol from the chromophore. Arg 92 forms a hydrogen bond with the oxygen of the imidazole ring from the chromophore. D. The most common conformation found in MD simulations. Trp 140 forms pi-stacking with the chromophore. There are four water molecules forming hydrogen bonds with the oxygen of the phenol from the chromophore. Arg 92 forms a cation-pi interaction with the chromophore as shown with the dark green dashed line. Arg 67

forms a hydrogen bond with the oxygen of the imidazole ring. E. Shows the planar angle, which is the dihedral angle of the chromophore (shown in 3.6A). In this figure, the dihedral angle (planar angle) is around 35º - 65º between 1600 ns – 2000 ns.

Tables

Table 3.1 Physical properties of red and blue fluorescent proteins

Name	Excitation Peak (nm)	Emission Peak (nm)	Quantum Yield
TagBFP	402	457	0.63
TagRFP	555	584	0.48
DsRed	558	583	0.68
TagRFP-T	555	584	0.41
mRuby3	558	592	0.45
mRuby2	559	600	0.38
mRuby	558	605	0.35
LSS-mKate 2	460	605	0.17
eqFP611	559	611	0.45
LSS-mKate 1	463	624	0.08
Rtms5	592	630	0.004
mKate	588	635	0.33
Neptune	600	650	0.18
mCardinal	604	659	0.19

Chapter 4 - The application and assessment of AlphaFold2 in drug design

Summary

There are 180,000 protein structures available in the Protein Data Bank that have been resolved by experimental methods, however this is still the tip of the iceberg compared to DNA encoded proteins. People have been trying to determine protein structures for the past 50 years, either with experimental or computational methods. The most inspiring thing that happened last year is the achievement of AlphaFold2 in Critical Assessment of Protein Structure Prediction (CASP). After this competition, AlphaFold2 has expanded the protein structures to around 350,000 structures, which is a huge contribution towards understanding numerous biological processes. However, the limitation of this new AI-based method has been neglected. Here, we focused on the assessment and application of the structures predicted by AlphaFold 2 on popular drug targets. It provides and emphasizes a valuable way to apply the AI developed method in drug discovery.

Introduction

Protein structures have been contributing to the elucidation of the mechanisms of metabolisms and drug design. There are over 180,000 protein structures available in the protein data bank (PDB) which contains structures solved by X-ray diffraction, nuclear magnetic resonance (NMR), cryogenic electron microscopy (Cryo-EM), and other experimental methods. The number of experimental structures is still expanding in the protein data bank, however there is still a huge gap between the number of experimentally solved structures and the total known DNA encoded proteins. Last year the biggest news in structural biology is that AlphaFold2 achieved a new level of protein structure prediction in Critical Assessment of Protein Structure Prediction 14 (CASP14). The AlphaFold2 (AF2) paper finally came out in *Nature* in 2021, which explains the algorithm of AI-based structure prediction method (Jumper et al., 2021). After it came out, this method has been applied to multiple different biological and biochemical fields. However, this method has not been applied to the drug discovery field. Computer aided drug design (CADD) is a huge advantage for finding new potential drug compounds as it is faster and lower cost than other approaches. However, it heavily relies on drug target structures. Since there are very limited protein structures available, high confidence predicted structures are very important to drug discovery. A big challenge in drug discovery is the prediction of flexible loops and protein multi-states. Loops always show incredible behavior in protein functions. It is also very hard to obtain the loop structures by experimental methods such cryo-EM or X-ray crystallization. Another big challenge in drug discovery is drug resistance. One of the reasons that causes drug resistance is mutations in the protein target (Fojo and Bates, 2013; Frey et al., 2010; Ghosh et al., 2020; Ito et al., 2001). Researchers are also trying to predict resistance mutations (Frey *et al.*, 2010). To develop drugs for these new mutants, either computational or

crystallization methods can be utilized to determine their structure. However, obtaining structures experimentally takes a long time. These two challenges are critical to drug discovery. The quality and accuracy of loops and mutants' structures is important to drug discovery, and lower quality protein structures can mislead researchers and waste time and resources.

Root-mean-square deviation (RMSD) has been used to evaluate predicted structures. It is a superposition method, aligning the alpha carbons of the predicted structure and reference structures. A lower RMSD value shows the two structures are more similar. This method is a quick and straightforward assessment resulting in a simple evaluation. It is the general standard to evaluate predicted structures and is also heavily used in virtual screening (Kirchmair et al., 2008). There are of course shortcomings in using the RMSD method. The flexibility of proteins is an issue, and especially with the hinges between subdomains which are more likely loops, some of the structures are disordered in apo-form. So, subdomains' orientations are evaluated by RMSD. Sometimes the domain has a low RMSD value, but the overall RMSD is high. To have a better understanding of the predicted structures, AlphaFold2 uses different methods to evaluate the predicted structures by a local superposition-free score (lDDT) and global distance test (GDT). The subdomains' orientations do not affect the lDDT score. (Mariani et al., 2013)

The protein structures used in drug discovery always require very high quality, including the sidechain poses and binding modes. The modes of the AlphaFold2 predicted structures are going to be important to drug design. One of the things that must be taken into account is that the structures predicted by AlphaFold2 from the AlphaFold Protein Structure Database are not considered co-factors and complexes.

Here we used typical popular drug targets to evaluate and more deeply explore the advantages and limitations of this innovative method.

Results

Liver X Receptor beta and Retinoic Acid Receptor alpha

Nuclear hormone receptors (NRs) are transcription factors that are composed of multiple subfamilies of proteins. There are many members of nuclear receptors related to various processes of physiology and pathology(Chawla et al., 2001; Evans, 2005). The study of nuclear receptors has revealed some of the mechanisms of nuclear receptors. It is also shown that many nuclear receptors are related to multiple cancers, hence they have caught people's attention and the members of nuclear receptors became drug targets (Weikum et al., 2018; Zhao et al., 2019). Most of the nuclear receptors have similar structures which contain five domains: N-terminus domain (NTD), DNA binding domain (DBD), hinge region, and ligand binding domain (LBD) (shown in Figure 4.1A) (Lou et al., 2014). Here, we used retinoic acid receptor alpha (RXRα), liver X receptor beta (LXR-β), and RXRα-LXRβ complex to evaluate the quality of predicted structures by AF2.

Liver X receptor beta (LXR-β) is related to regulation of the lipid and cholesterol metabolism, and other physiological processes (Korach-André et al., 2011). LXR-βs are important drug targets in different diseases (Jakobsson et al., 2012). LXR-βs bind with RXRs forming heterodimers which bind to their cognate agonist ligands. Then the complexes conduct the active LBD domain conformations. Understanding the structure of LXR-RXRα complex is essential to reveal the NRs functions, and drug design. The RXRα-LXRβ complex is very interesting as it is different from the other NR family members. The crystal structure of RXRα-LXRβ complex shows the LBD and DBD domains form an X-shaped arrangement (Figure 4.1H). Figure 4.1A shows the crystal structure of retinoic acid receptor alpha (RXR-α) with liver X receptor beta (LXR-β) on DNA (Lou *et al.*, 2014). We used this complex to evaluate the AF2

predicted structure. Figures 4.1B and 4.6A show the AF2 predicted LXR model. After superimposing the predicted structures and the crystal structure, the RMSD is 0.605 Å. The plDDT score indicates the prediction confidence of different areas of the structure. The loops between each domain are shown with low confidence. The same thing happens to the RXR-α predicted model. Figures 4.1C and 4.6B show the AF2 predicted RXRα models. The RMSD is 0.456 Å. The loops between each domain of the RXR-α AF2 predicted model also show low confidence. It is very hard to predict flexible structures and it is also very hard to obtain these regions from crystal structures. Next, we used AF2 to predict the complex. Figures 4.1D and 4.6C show the AF2 predicted RXRα-LXRβ complex. It is similar with the predicted LXRβ and RXR-α structures, the LBD domains from the predicted complex show higher confidence than other domains. Figure 4.1G shows the superimposition of the predicted complex and the crystal structure. From the side view to look at the predicted complex, AF2 failed to generate the DBD domains of RXRα-LXRβ.

The AF2 predicted LBD domains show high confidence. Hence, we did a docking study to deeply evaluate the AF2 predicted structures. Compound 1 and compound 2 shown in Figure 4.1 are LXR and RXR agonists. For better evaluations we used the same protein sequences as the crystal structure to regenerate the AF2 predicted LXR and RXR structures. Figure 4.2A shows the AF2 predicted structure superimposed on to the LXR-RXR crystal structure. Figure 4.2B shows the AF2 predicted structure superimposed on to the LXR-RXR crystal structure. Figures 4.2C and 4.2B show the LXR and RXR structure plDDT scoring and the LXR crystal structure and the RXR crystal structure superimposed. Figure 4.2E shows the binding site of RXR. The crystal structure of RXR and ligand are shown in blue. The predicted structure is shown in green. The docked ligand is shown in pink. Figure 4.2F shows the zoomed-in view of the binding site.

The docking score is -9.420, the glide gscore is -9.425 (the smaller number is better). Using the crystal structure to do docking, the docking score is -9.739 and the glide gscore is -9.743 (Figure 4.7.). AlphaFold2 performed well on the RXR LDB domain. We kept evaluating the LXR LDB domain. The AF2 predicted LXR LBD domain was used to perform docking. However, we can't produce any docking results even though the structure has a low RMSD and is showing high confidence by plDDT. Figure 4.8 explains in detail why this docking failed. Figure 4.8A shows the sidechain of the binding site. Figure 4.8B shows the ligand in the predicted structure after superimposing the two structures. Figure 4.8B shows steric clashing between the sidechain of Phe340, and the ligand circled in black. The red dashed lines show the bad clash that is causing the docking to fail. In Figure 4.8, most of the side chains are quite similar, the biggest difference is Phe340.

Estrogen Receptor alpha

Another popular nuclear receptor as a drug target is estrogen receptor alpha (ERα). Estrogen receptor alpha is related to breast career which is the most common cancer in women. Over half of breast cancers overexpress ERα (Ali and Coombes, 2000). Tamoxifen has been used to treat all stages of estrogen receptor positive breast cancer (Jordan, 2003). Tamoxifen was also the first cancer chemotherapy approved by the FDA. However, clinical evidence is showing drug resistance after extended treatment. A big challenge to predicting protein structures is to predict mutants. For testing AlphaFold2 protein mutant prediction, we used the two common ERα mutants Y537S and D538G to evaluate AF2 prediction performance.

We used AF2 to generate the two mutants and evaluated the predicted wild type ERα, and the two mutants' structures. Figures 4.3A and 4.9 show the predicted full-length wildtype structure. The superimposition of crystal structure and AF2 predicted structure of wildtype ERα

is shown in Figure 4.3A. The crystal structure of wildtype ERα is shown in gold, the ligand is shown in salmon (PDB ID: 1YIM). The AF2 predicted wildtype ERα is shown in purple. The RMSD value between the crystal structure and AF2 predicted structure is 0.514 Å. The black circle shows the steric clash between the ligand and helix 12 (H12). Figure 4.3A shows the ligand clashes with the AF2 predicted structure. Figures 4.3C and 4.10A show the zoomed in view of the binding pocket. The red dashed lines show the clash in Figure 4.10B. This clash causes the docking to fail. We kept studying the other two mutants, Y537S, and D538G. There were similar results as with the wildtype ERα. The RMSD values of Y537S and D538G are 0.534 Å and 0.530 Å. The ligands clash with H12 shown in Figures 4.3B, 4.3D, 4.12B. Figure 4.12A shows the binding site of the crystal structure of mutant Y537S, Figure 4.12B shows the binding site of the AF2 predicted Y537S structure. The red dashed lines show the bad clash between the ligand and the receptor. Figures 4.9, 4.11A, and 4.11B show the wildtype, Y537S, and D538G AF2 predicted structures with plDDT confidence. It shows the AF2 structures with high confidence. Even with high RMSDs and plDDT confidence, the AF2 predicted structures of ERα still fail to generate docking results.

PTP1B

The protein tyrosine kinases (PTKs) and phosphatases (PTPs) are important to the growth, proliferation, differentiation, survival, and apoptosis pathways (Hunter, 2000). Protein tyrosine phosphatase 1B (PTP1B) is one of the protein tyrosine phosphatases (PTPs) family and a potential drug target for the treatments of type 2 diabetes, obesity and HER2-positive breast cancer (Krishnan et al., 2014). PTP1B has been shown to play a critical role in signaling pathways, especially involving insulin and leptin signaling (Zhang and Zhang, 2007). PTP1B as a tyrosine phosphatase catalyzes the hydrolysis of phosphorylated tyrosine residues using a

phosphocysteine (Pannifer et al., 1998). The PTP1B's catalytic site contains the PTP loop, the WPD loop, and the substrate binding loop (Choy et al., 2017). The PTP loop (VHCSXGXGR[T/S]G) includes the catalytic cysteine Cys215, which is the phosphate-binding loop; the WPD loop (WPDXGXP) includes Asp181, which is the proton donor and acceptor during phosphoryl transfer (Andersen et al., 2001). The WPD loop is also highly conserved among PTP domains. The WPD loop switches from the open state to the closed state by the substrate binding (Choy *et al.*, 2017). Figure 4.4A shows the superimposition of crystal structures of the two states of PTP1B (PDB ID: 5K9V – open state; 5K9W – closed state) (Choy *et al.*, 2017). The main difference between the two states is the WPD loop. The open state loop is open towards the outside. After binding the ligand, the WPD loop closes towards the inside of the binding pocket forming the closed state. The ligand is shown in gray. To know which state AF2 predicted for PTP1B, we superimposed the AF2 structure with the open and closed state crystal structures (shown in Figure 4.4B and 4.4C). After comparison we found that the AF2 predicted PTP1B structure is in the closed state which binds to the ligand. Figure 4.4B shows the open state of PTP1B superimposed with the AF2 predicted structure. Figure 4.4C shows the closed state of PTP1B superimposed with the AF2 predicted structure. Figure 4.4D shows the zoomed in view of the superimposed binding pocket of the closed state and the AF2 predicted structure. The sidechain of the AF2 predicted structure is quite similar with the crystal structure. We used the AF2 predicted structure as the receptor and docked the ligand into it. We then used the closed state crystal structure to dock the ligand back into the receptor. The results are shown in Figures 4.4E and 4.4F. The AF2 structure is shown in blue, the docked ligand is shown in dark blue. The crystal structure is shown in pink, the docked ligand is shown in dark pink. Both docked ligands' poses are very similar. The crystal structure's docking score is -9.099, the glide

gscore is -9.123. The AF2 predicted structure's docking score is -8.064, the glide score is -8.088. Both docked ligand's poses look quite similar, the docking scores are around 1 different. Figure 4.4F shows the zoomed in view of the binding pocket. The ligand from the crystal structure is shown in gray. The AF2 structure's binding pocket is quite similar with the crystal structure except for Phe182 which is changing the shape of the binding pocket. The small discrepancy between the two structures leads to the docking error. The AF2 predicted structure docking result is still quite good.

Insulin Receptor

Insulin receptor (IR) is a receptor tyrosine kinase (RTK) and a critical drug target related to glucose metabolism and cell growth. Unlike other RTKs, insulin receptor covalently links and forms a homodimer by multiple disulfide bonds. It contains two chains, α and β. There is the extracellular region, transmembrane region, and the cytoplasmic region. The extracellular region contains various domains, L1, CR, L1, Fibronectin type-III 1 (FnIII-1), Fibronectin type-III 2 (FnIII-2), Fibronectin type-III 3 (FnIII-3) and insert domain (ID-α, ID-β) (Figure 4.5J). In this region it also contains two insulin-binding sites with different binding affinities (Figure 4.5D). The cytoplasmic region contains juxtamembrane (JM), and tyrosine kinase (TK) domains (Figure 4.5J). The cryo-EM structures show the inverted "V" and "T" shape of insulin receptor (Gutmann et al., 2018; Scapin et al., 2018; Weis et al., 2018). Recently the extracellular region structure of insulin receptor was solved by cryo-EM, and it revealed the 'T' shape of the homodimer architecture (Figure 4.5C and Figure 4.5D, PDB ID: 6PXV) (Uchikawa et al., 2019). Here we used the cryo-EM structure of insulin receptor as a reference to evaluate the AF2 predicted model of insulin receptor. This cryo-EM structure resolution is 3.2 Å. Figure 4.5A shows the AF2 predicted structure with confidence evaluated by the lDDT method. It clearly

shows that most of L1, L2, FnII-1, FnIII-2, FnIII-3, and TK domains have relatively high confidence. The CR domain failed to build, and its secondary structures predicted as helices and loops are not packed well in the AF2 model. Due to the failure of the CR domain to pack, the L1 and L2 domains are in the wrong layout. Hence the insulin binding sites also failed to build. The insert domain and transmembrane domain are shown with low confidences. After comparing with the cryo-EM structure, the insert domain is on the wrong side of the protein. The TM domain is folded toward the extracellular region which causes the TK domain to be in the extracellular region. This AF2 predicted structure fails to show each domain layout. The cryo-EM structure revealed insulin 1 bound to L1, FnIII-1, and α-CT (Uchikawa *et al.*, 2019). Since the AF2 predicted model failed to build the overall structure we believe using this structure to study insulin binding or the layout of each domain is not going to be successful. We did not keep evaluating the full-length insulin receptor binding with insulins.

As we know, the flexible region and the connections between the domains are a big challenge to be predicted. These regions are also very hard to obtain by crystallography and cryo-EM structures. Hence, we evaluated the important domains in the insulin receptor.

Insulin receptor is an essential tyrosine kinase, the tyrosine kinase domain is the key to regulation of cell functions. Insulin receptor tyrosine kinase has been crystallized. We used the crystal structure as a reference to evaluate the AlphaFold2 predicted model and used AF2 to rebuild a new model that only contains the TK domain. The crystal structure of insulin receptor tyrosine kinase domain is bound to a small molecule, cis-(R)-7-(3-(azetidin-1-ylmethyl)cyclobutyl)-5-(3-((tetrahydro-2H-pyran-2-yl)methoxy)phenyl)-7H-pyrrolo[2,3-d]pyrimidin-4-amine (Stauffer et al., 2016). The PDB ID is 5HHW shown in Figure 4.5F. We superimposed two structures with the crystal structure. The RMSD of the TK domain and the

crystal structure is 0.776Å, the RMSD of the TK domain from full-length IR and the crystal structure is 0.909Å. The regular superimpose only compares the backbone and uses RMSD value to show the similarity. The smaller RMSD value shows the model is more similar. From the simple backbone comparison, it shows the TK domain predicted model is a little bit better than the TK domain from the full-length IR predicted model. However, the RMSD method is too simple to compare structures' similarity. For deeper investigation of these two predicted models, we looked at the predicted structures shown in plDDT model confidence. The new AF2 predicted TK domain model of insulin receptor is shown in Figures 4.5G and 4.14B. Figures 4.5A and 4.14A show the full-length insulin receptor model with the model confidence, the tyrosine kinase domain region is shown to have relatively high confidence. The figures 4.5G and 4.14B show the new TK domain which is with higher confidence than the TK domain from the full-length IR model. Unfortunately, the binding site of TK domain model shows low confidence (Figures 4.5 and 4.14B). We then performed a docking study of this domain. As expected, the docking results of both predicted models presented poorly (Figures 4.5H). The ligand did not dock into the binding pocket. We also used the crystal structure to generate the grid, then docked the ligand back to the receptor; the docking results are shown in Figures 4.15D and 4.15E. It shows the docked ligand is in the binding pocket. The table in Figure 4.15 shows the docking scores and glide gscores. The crystal structure docking score is -10.346, the glide gscore is -10.349. The docking score of the AF2 predicted structure is -6.062, and the glide gscore is -6.064. Both the scores and the docking pose shows the AF2 result has lower accuracy. We further investigated the AF2 predicted structure to find the reason docking failed. In the binding pocket there is an important loop predicted with low confidence. This loop causes the docking to

fail and blocks the ligand entering into the pocket. In this case AF2 failed to generate a correct loop.

Discussion

AlphaFold Protein Structure Database now is available online. People can easily download the AF2 predicted structure from the Database or through uniprot.org. This allows researchers to easily obtain the protein structures. AlphaFold offers a confidence assessment of the predicted structures by plDDT and expected position error. However, protein structures have multiple states. Sometimes it is very hard to estimate the structure accuracy of the various protein states, and that of sidechains by just looking at the backbone cartoon view. Despite this assessment, our work with AlphaFold2 predicted structures is limited to only a few typical popular drug targets, though it still covers different cases, and the main goal of this paper is to show the performance details of AlphaFold2 and to help people better understand AlphaFold2 and protein structures. It evaluates the performance of AlphaFold2 and points out the confidence of the structures and the details of the predicted protein structures, such as the loops and sidechains which have been neglected in the past. Even with good RMSDs and plDDT values, the predicted structures still need to be reevaluated, especially if there are different states of proteins, and ligand binding modes which is of more interest in drug discovery. Drug discovery requires very high-quality protein structures.

We worked on the assessment of AF2 predicted structures of RXR-α, LXR-β, Erα, PTP1B, insulin receptor, and insulin receptor tyrosine receptor. We found LXR-β has a good docking result. The AF2 predicted LXR-β binding site shows high confidence, and the docking result also shows good results compared with the crystal structure (shown in Figures 4.2D and 4.7). Unfortunately, the AF2 predicted RXR-α structure can't be used to do docking even though

this structure has very high confidence. To study AF2 prediction of mutants, a protein notorious for drug resistance, estrogen receptor alpha was used as an example. After compared to the wildtype, mutant Y537S, and mutant D538G, all AF2 predicted structures failed to generate correct structures on helix 12. PTP1B was studied as a typical protein tyrosine phosphatase. The WPD loop has two states, open and closed states. As drug designers, we are more interested in which state AF2 predicted. In the case of PTP1B, AF2 predicted the closed state, which is the ligand binding mode. This is very important to drug design for studying ligand binding with receptors. The PTP1B docking results show a slight side chain difference which causes the docking results to be different. With a low confidence predicted structure as with the case of insulin receptor, the full-length insulin receptor predicted structure is not as high confidence as other studied cases in this work. As expected, the insulin receptor AF2 predicted structure also did not produce satisfactory results. However, one of the domains were of high confidence, the tyrosine kinase domain. Although the general confidence of the AF2 TK domain structure is quite high, the binding site area shows lower confidence than the other part of the TK domain. Hence, we did not get a decent docking result.

After we evaluated the AF2 predicted structures, we found not all the structures are good enough to apply into drug discovery. There is no rule to make it easy to evaluate the performance of AF2 predicted structures. We still tried to summarize the assessment. First, looking at the plDDT confidence figure we can understand the general AF2 structure performance. If the structure or motif already shows low confidence, this means the predicted structure is not very good to use for docking study or virtual screening. Then if the predicted structure has high confidence, we can look at the details of the AF2 structure. Even if there are a lot of protein structures that have not been solved by either X-crystallography or cryo-EM, we are still able to

find an experimental structure from the same family or other species to look at the binding sites or the conserved residues and verify the state of the predicted structure.

There is a huge contribution to science by AlphaFold2. It offers structural information and is trying to fill the gap between the number of experimental structures and known DNA coded proteins. This helps researchers to have a better understanding of protein functions. This work offers the objective and detailed assessment of using AlphaFold2 predicted structure to study drug discovery and gives a brief guide of how to use AlphaFold2 structures.

Materials and Methods

Predicted Model Building

The full-length human proteins of AlphaFold 2 predicted structures were obtained from AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk) (Jumper *et al.*, 2021). Other proteins of AlphaFold2 predicted structures were obtained by running AlphaFold2 Python code on Google Colab (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb). The crystal structures were obtained from Protein Data Bank (PDB).

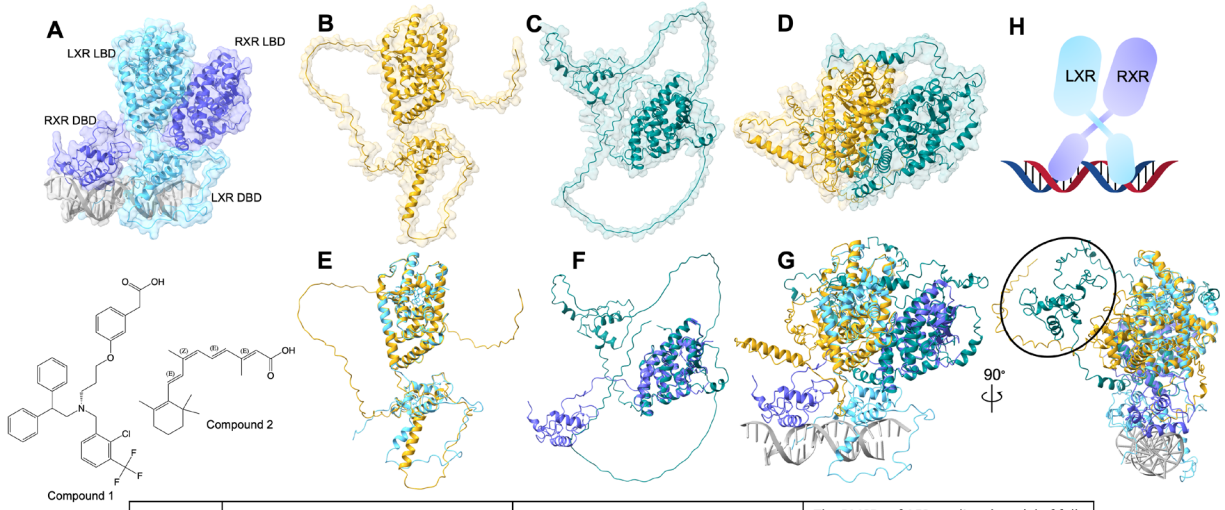
Ligand Docking

Glide was used to study ligand docking. The crystal structures were prepared by Protein Preparation Wizard (Sastry *et al.*, 2013). The ligands were prepared by LigPrep. The receptor grids were generated by Receptor Grid Generation (Schrödinger Release 2021-4). All binding pockets coordinates are obtained from crystal structures. The docking study used Glide (Schrödinger Release 2021-4) (Friesner et al., 2004; Friesner et al., 2006; Halgren et al., 2004).

Structural Assessment

The structural comparisons in this study were performed with different methods, which are RMSD, and plDDT. UCSF ChimeraX was used to align these predicted structures and calculate these RMSD values. The UCSF ChimeraX version is 1.1 (2020-10-07) (Goddard *et al.*, 2018). The values of plDDT were obtained from the AlphaFold2 on Google Colab. The protein structures were visualized by ChimeraX. The chemical formulas were drawn by ChemDraw 20.1.

Figures



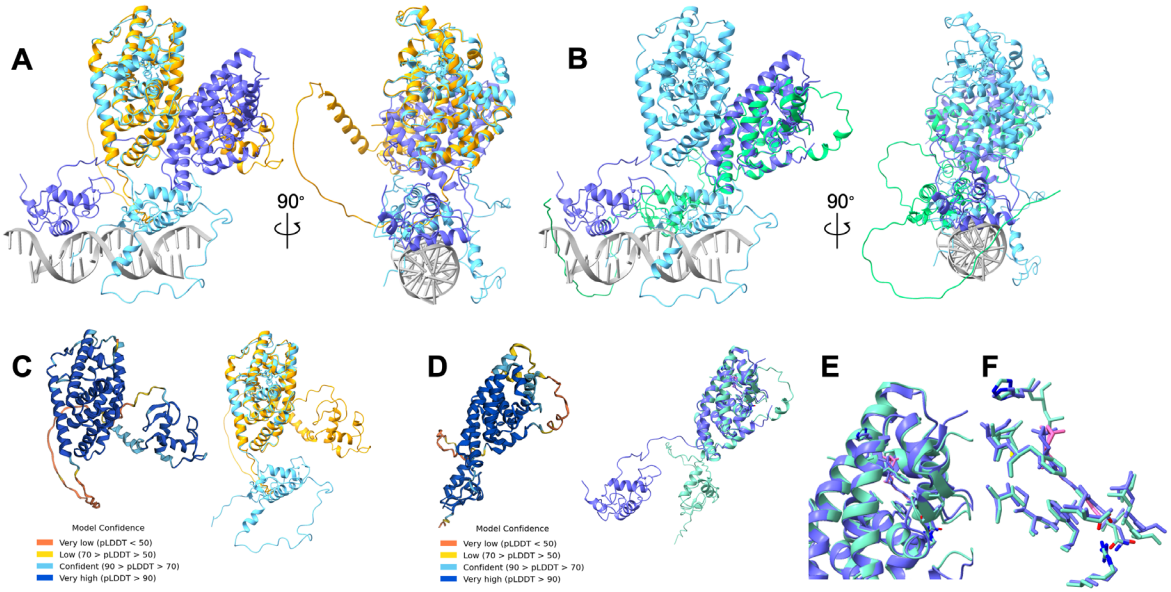
Name	The AF2 predicted model of LXR-β (yellow)	The AF2 predicted model of RXR-α (teal)	The RMSDs of AF2 predicted model of full-length RXRα-LXRβ complex	
			LXR-β (yellow)	RXR-α (teal)
RMSD (Å)*	0.605 between 308 atom pairs. (3.705 cross all 384 pairs)	0.456 between 182 atom pairs. (30.861 cross all 319 pairs)	0.654 between 160 atom pairs. (35.269 cross all 319 pairs)	0.772 between 212 atom pairs. (28.825 cross all 384 pairs)

* The reference structure is the crystal structure of RXRα-LXRβ complex (PDB ID: 4NQA).
The RMSDs were calculated by ChimeraX. The Alignment algorithm is Needleman-Wunsch.

Figure 4.1 The crystal structure and AlphaFold2 predicted retinoid X receptor alpha–liver X receptor beta complex.

A. shows the crystal structure of the retinoid X receptor alpha–liver X receptor beta (RXRα-LXRβ) heterodimer binds to DNA (PDB ID: 4NQA). Most nuclear receptors have a similar structure which mainly contains five subdomains, N-terminus domain (NTD), DNA binding domain (DBD), hinge, and ligand binding domain (LBD). The bottom of the proteins in contact with DNA are the DNA binding domain. The Retinoic acid receptor alpha (RXR-α) is in purple. The liver X nuclear receptor beta (LXR-β) is in cyan. The DNA is in gray. B. shows the AF2 predicted LXR-β structure (yellow). C. shows the AF2 predicted RXR-α structure (teal). D. shows the AF2 predicted full-length RXRα-LXRβ complex structures (RXR-α: teal, LXR-β: yellow). H. shows the layout of the RXRα-LXRβ binding with DNA complex. LXR is in cyan. RXR is in purple. E. shows the superimposition of AF2 predicted full-length RXR-α LXR-β

structure and the crystal structure of LXR- β (AF2 LXR-β: yellow, crystal LXR-β: cyan). F. shows the superimposition of AF2 predicted full-length RXR-α structure and the crystal structure of RXR-α (AF2 RXR-α: teal, crystal RXR-α: purple). G. shows the superimposition of AF2 predicted full-length RXRα-LXRβ complex structure (AF2 LXR-β: yellow, crystal LXR-β: cyan, AF2 RXR-α: teal, crystal RXR-α: purple). The structure turned 90º. The black circle shows the DNA binding domains of AF2 predicted full-length RXRα-LXRβ complex structure. The confidence of the structure is shown in Figure 4.6. The table shows the RMSDs of the AF2 predicted full-length LXR-β structure, the AF2 predicted full-length RXR-α structure, the AF2 predicted full-length RXRα-LXRβ complex structure, with crystal structure. H. shows the schematization of the RXRα-LXRβ complex. Compounds 1 and 2 are the cognate ligand structures which bind with two LBDs.



Name	The AF2 predicted LXR-β (72-460aa) model (color)	The AF2 predicted RXR-α (72-460aa) model (green)
RMSD (Å)*	0.663 between 220 atom pairs. (33.314 cross all 384 pairs)	0.614 between 177 atom pairs. (17.446 cross all 287 pairs)

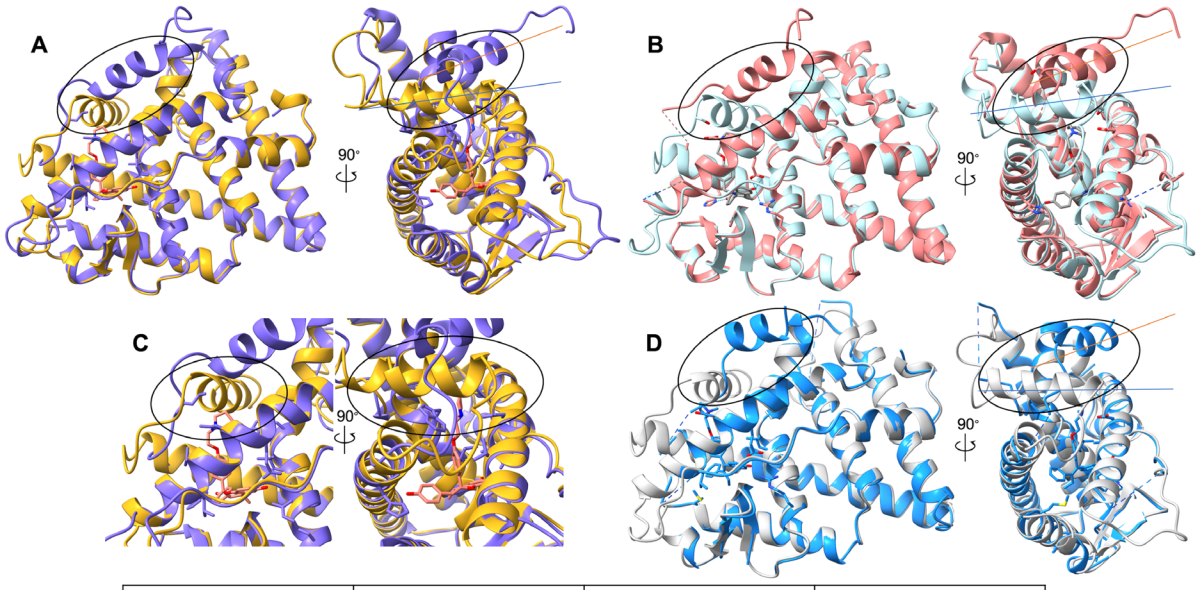
	Docking score	Glide gscore
RXR-α	-9.420	-9.425

* The reference structure is the crystal structure of RXRα-LXRβ complex (PDB ID: 4NQA).
The RMSDs were calculated by ChimeraX. The Alignment algorithm is Needleman-Wunsch.

Figure 4.2 The RXRα and LXRβ AlphaFold2 predicted models and the crystal structure of RXRα-LXRβ complex.

A. shows the superimposition of the RXRα-LXRβ crystal structure and the LXRβ AF2 predicted structure (the protein sequence is the same with the crystal structure). The PDB ID is 4NQA. The crystal structure of the RXRα is shown in purple. The crystal structure of LXRβ is shown in cyan. The LXRβ AF2 model is shown in gold. B. shows the superimposition of the RXRα-LXRβ crystal structure and the RXRα AF2 predicted structure (the protein sequence is the same as the crystal structure). The RXRα AF2 predicted model is shown in green. C. shows the LXRβ AF2 predicted structure. The colors indicate plDDT confidence. The right side of C shows the superimposition of the crystal structure and the AF2 model. D. shows the RXRα AF2 predicted structure. The colors indicate plDDT confidence. The right side of C shows the superimposition

of the crystal structure and the AF2 model. The table shows the RMSDs between the crystal structure and AF2 predicted two models. The ligand binding domains of both RXRα and LXRβ AF2 predicted models are predicted quite well. However, the DNA binding domains of both AF2 predicted models failed to generate. E. shows the superimposition of the ligand binding site of crystal structure of RXRα and the RXRα AF2 predicted model. F. shows the superimposition of the ligand binding site of the RXRα crystal structure and AF2 predicted model. The details of binding results and pocket are shown in Figure 4.7. The docked ligand into the AF2 model is shown in pink. The table shows the docking score and glide gscore. The LXR-β AF2 predicted structure fails to generate a docking result. Figure 4.8 shows the steric clash between the receptor and the ligand.



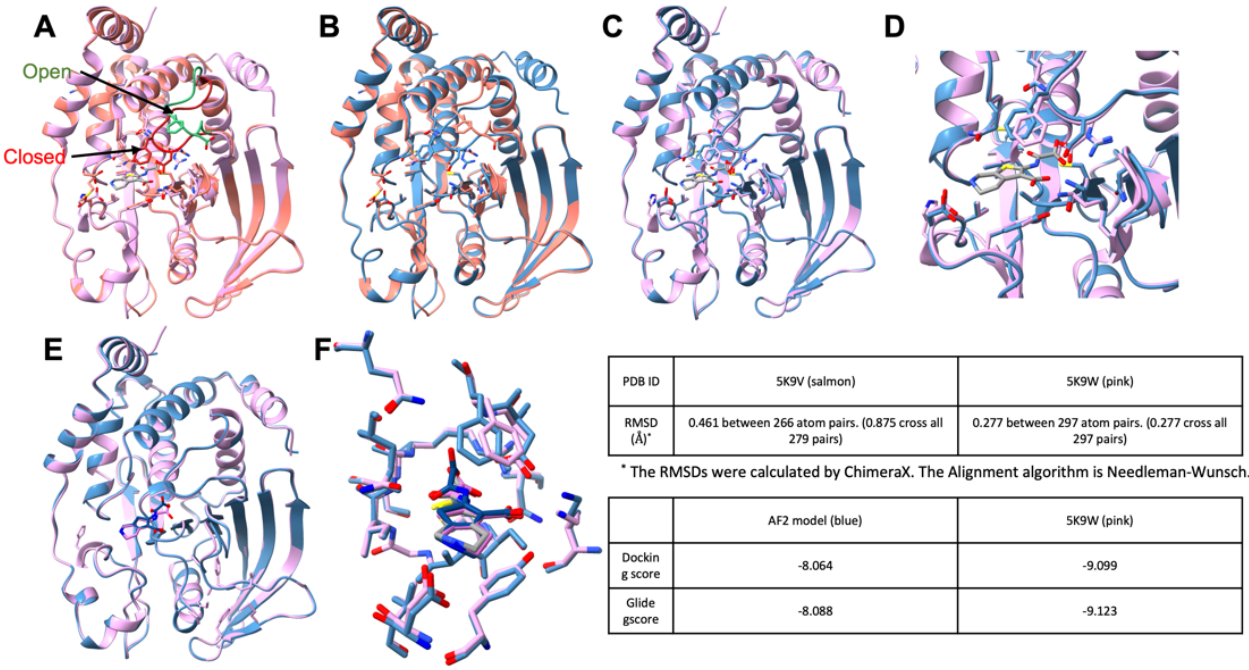
Name	Wildtype (purple)	Y537S (orange)	D538S (sliver)
RMSD (Å)*	0.514 between 202 atom pairs. (6.033 cross all 245 pairs)	0.534 between 202 atom pairs. (5.470 cross all 232 pairs)	0.530 between 189 atom pairs. (3.599 cross all 210 pairs)

* The reference structure is the crystal structure of ERα, ERα (Y537S), and ERα (D538G) (PDB ID: 1YIM, 6PSJ, and 4Q50).
The RMSDs were calculated by ChimeraX. The Alignment algorithm is Needleman-Wunsch.

Figure 4.3 The estrogen receptor alpha crystal structures and AlphaFold2 predicted structures.

A. shows the superimposition of wild type estrogen receptor alpha crystal structure and AF2 predicted structure from two different views. The crystal structure is shown in purple (PDB ID: 1YIM). The AF2 predicted structure is shown in gold. The black circle shows helix 12 (H12). The H12 from the AF2 predicted structure has a different orientation from the crystal structure. This difference in orientation causes the H12 steric clashes with the ligand and changes the ligand binding site. C. shows the two zoomed-in views of the binding site of A. Figure 4.9 shows the AF2 predicted structure's confidence and the superimposition of the crystal structure and the AF2 predicted structure. The black circles show the steric clash of the ligand and the H12 from the AF2 predicted structure. Figure 4.10 also shows the details of wild type ERα binding site of the AF2 predicted structure. B. shows the superimposition of estrogen receptor alpha mutant Y537S crystal structure and the AF2 predicted structure. The crystal structure of the mutant

Y537S (PDB ID: 6PSJ) is in salmon. The AF2 predicted mutant Y537S is in light cyan. Figure 4.11 shows the mutant Y537S AF2 predicted structure's confidence. H12 is indicated by the black circle. There are two different orientations of H12. H12 from the AF2 predicted structure has steric clashing with the ligand as well. The Figure 4.12 shows the details of the AF2 predicted structure steric clashing with the ligand. D. shows the superimposition of estrogen receptor alpha mutant D538G crystal structure and the AF2 predicted structure. The crystal structure of mutant D538G is shown in blue (PDB ID: 4Q50). The AF2 predicted structure is shown in light gray. Figure 4.6B shows the mutant D538G AF2 predicted structure's confidence. H12 from the AF2 predicted structure has steric clashing with the ligand too. Compared to three estrogen receptor alpha, H12 from the AF2 predicted models all have different orientations from each crystal structure. These different orientations cause H12 to steric clash with the ligands. The table shows the RMSDs of crystal structures and AF2 predicted structures.



PDB ID	5K9V (salmon)	5K9W (pink)
RMSD (Å)*	0.461 between 266 atom pairs. (0.875 cross all 279 pairs)	0.277 between 297 atom pairs. (0.277 cross all 297 pairs)

* The RMSDs were calculated by ChimeraX. The Alignment algorithm is Needleman-Wunsch.

	AF2 model (blue)	5K9W (pink)
Docking score	-8.064	-9.099
Glide gscore	-8.088	-9.123

Figure 4.4 The PTP1B crystal structures and AlphaFold2 predicted structure. A. shows two PTP1B crystal structures.

A. shows two PTP1B crystal structures. There are two states, the open and the closed state. The difference between two states is the WPD loop. The 5K9V crystal structure which is the open state is shown in salmon. The open state WPD loop is shown in green. The 5K9W crystal structure which is the closed state is shown in pink. The closed state WPD loop is shown in red. B. shows the superimposing structures of the open state crystal structure 5K9V and the AF2 predicted structure. The AF2 predicted structure is shown in blue. Figures 4.13A and 4.13B shows the AF2 predicted structures' confidences. C. shows the superimposing structures of the closed state crystal structure 5K9W and the AF2 predicted structure. From B and C, it shows the AF2 predicted structure is the closed state. D. shows the zoomed-in view of the Figure 4.4C binding site. E. shows the docking results. The AF2 model's ligand docking result is in dark blue. The closed state crystal structure 5K9W's ligand docking result is in dark pink. F. shows

the zoomed-in view of the binding site. The ligand from crystal structure 5K9W is shown in gray. The tables show the RMSDs and the docking scores. The details of the docking results are also shown in Figures 4.13D, 4.13E, and 4.13F

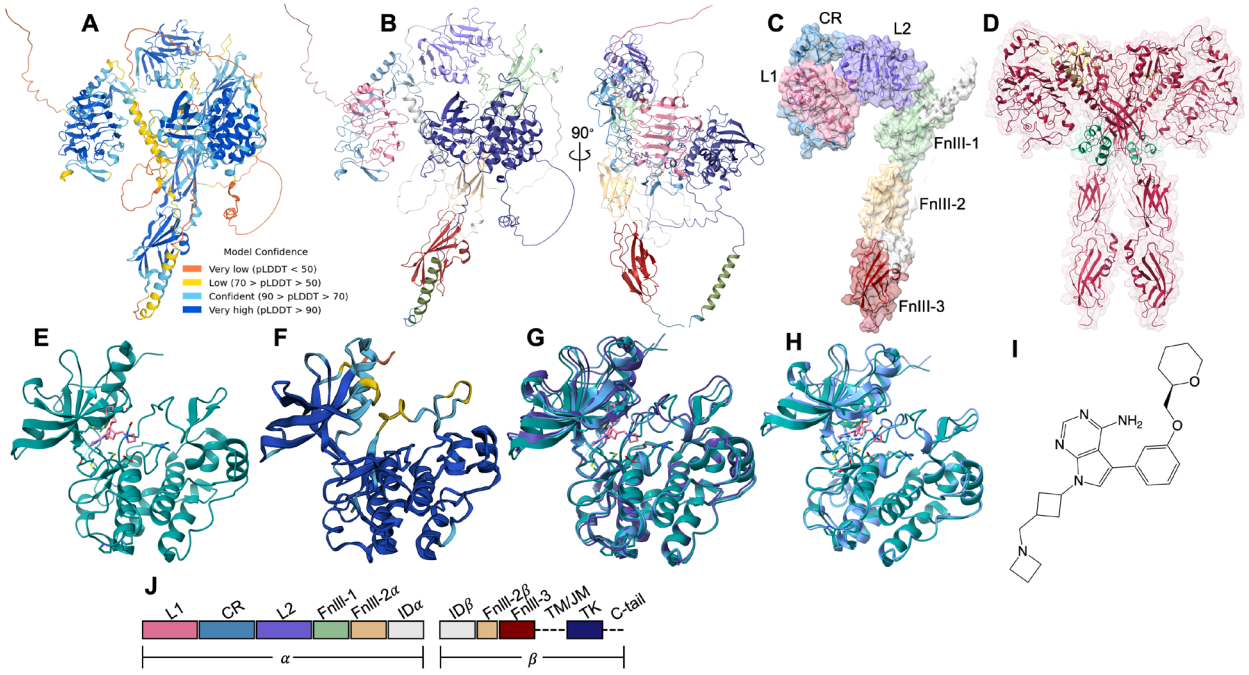


Figure 4.5 The comparison of AlphaFold2 predicted model of full-length human insulin receptor and the cryo-EM structure of extracellular region insulin receptor.

This cryo-EM structure resolution is 3.2 Å, PDB ID is 6PXV.

A. shows the AlphaFold2 predicted model for full-length insulin receptor. The structure is shown in the confidence by plDDT. The dark blue color shows the confidence is beyond 90%. The light blue color shows the confidence between 70% and 90%. The yellow color shows the confidence between 50% and 70%. The orange color shows the confidence below 50%. The loop domains show low confidence. The flexible domains are very hard to predict, even crystal structures and cryo-EM structures are commonly missing these regions' structures. The insert domain and transmembrane domain are shown with low confidences as well. This domain structure appears less commonly in experimental structures.

B. shows the AlphaFold2 predicted model for full-length insulin receptor. The L1 domain is shown in pink. The CR domain is shown in light blue. The L2 domain in shown in light purple.

The Fibronectin type-III 1 domain (FnIII-1) is shown in light green. The Fibronectin type-III 2 domain (FnIII-2) is shown in yellow. The Fibronectin type-III 3 domain (FnIII-3) is shown in red. The transmembrane domain is shown in olive. The insulin receptor tyrosine kinase (cytoplasmic) is shown in dark purple. The insert domain (ID) is shown in light gray. The dark blue is tyrosine kinase domain (TK). The other one shows the AlphaFold2 predicted model rotated 90°.

C. shows the monomer of human insulin receptor. The L1 domain is shown in pink. The CR domain is shown in light blue. The L2 domain in shown in light purple. The FnIII-1 domain is shown in light green. The FnIII-2 domain is shown in yellow. The FnIII-3 is shown in red. The insert domain is shown in light gray.

D. shows the homodimer of human insulin receptor with four insulins. The full-length extracellular insulin receptor is shown in red. The two insulin binding in site 1 and 1' are shown in yellow. The two insulin binding in site 2 and 2' are shown in green.

E. shows the crystal structure of TK domain binding with a small molecule (PDB ID: 5HHW). The crystal structure is in teal. The ligand is in salmon.

F. shows the TK domain from the AlphaFold2 predicted model of the IR TK domain. The structure's colors are shown in plDDT confidence, same with A.

G. shows the superimposed structures of the TK domain from the full-length IR, the TK domain structure built by AF2 and the crystal structure 5HHW. The crystal structure is in teal. The AF2 model is in purple.

H. shows the TK domain from the full-length IR docked with the 5HHW ligand (purple). The binding site coordination setting is the same as the crystal structure.

I. It shows the small molecule, cis-(R)-7-(3-(azetidin-1-ylmethyl)cyclobutyl)-5-(3-((tetrahydro-2H-pyran-2-yl)methoxy)phenyl)-7H-pyrrolo[2,3-d]pyrimidin-4-amine.

J. shows insulin receptor domains. The colors are matched with B. The L1 domain is shown in pink. The CR domain is shown in light blue. The L2 domain is shown in light purple. The FnIII-1 domain is shown in light green. The FnIII-2 alpha and beta domains are shown in wheat. ID alpha and beta domains are shown in light gray. The FnIII-3 domain is shown in red. The dashed lines show the TM/JM domain and C-tail. The TK domain is shown in dark blue.

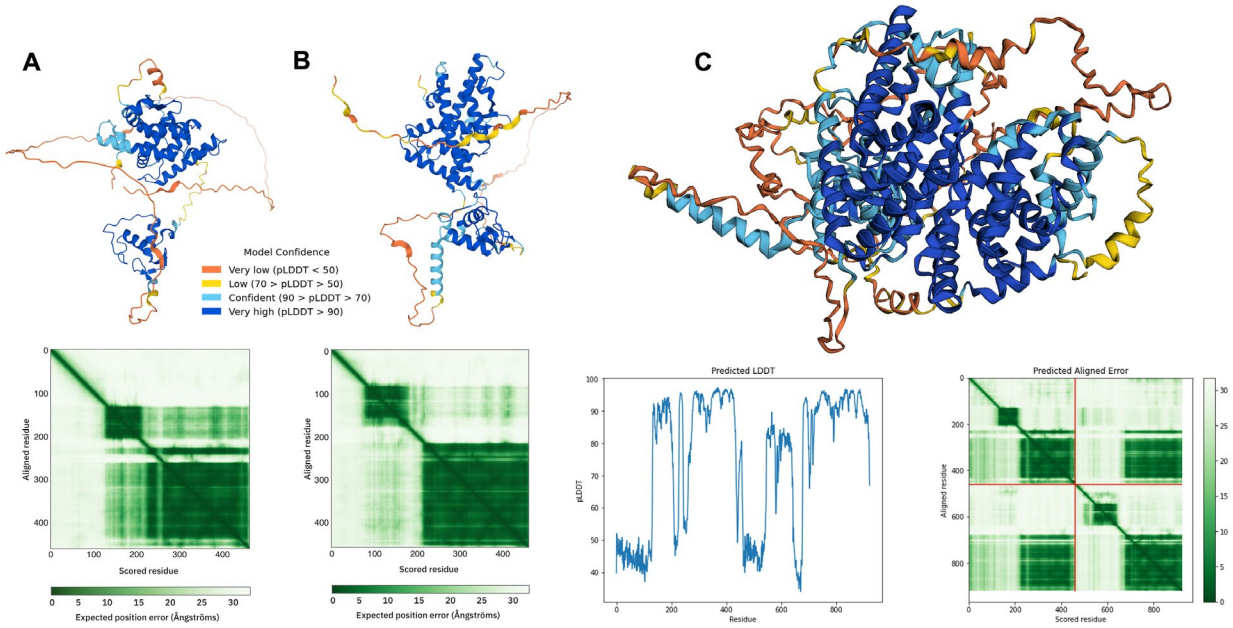
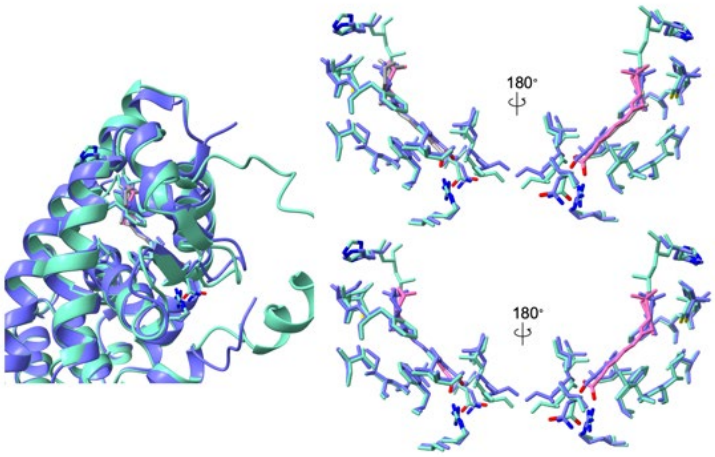


Figure 4.6 The AlphaFold2 predicted structures of RXR-α, LXR-β, and RXRα-LXRβ complex.

A. shows the AF2 predicted model of full-length retinoid X receptor alpha (RXR-α), coloration indicates plDDT confidence. The predicted aligned error graph is under the structure. B. shows the AF2 predicted model of full-length liver X nuclear receptor beta (LXR-β), coloration indicates plDDT confidence. The predicted aligned error graph is under the structure. C. shows the AF2 predicted complex model of the full-length retinoid X receptor α-liver X receptor β (RXRα -LXR β) complex, coloration indicates plDDT confidence. The predicted aligned error graph is under the structure.



	Docking score	Glide gscore
RXR-α (AF2)	-9.420	-9.425
RXR-α (4NQA)	-9.739	-9.743

Figure 4.7 The RXR-α ligand binding domain docking result.

The left figure shows the superimposition of crystal structure and AF2 predicted RXR-α. The crystal structure is shown in purple. The AF2 predicted model is shown in green. The ligand docked into the AF2 model is shown in pink. The right figures show the binding site. The top right shows the superimposition of crystal structure and AF2 model. The ligand docked into the AF2 model is shown in pink. The ligand docked into the crystal structure is shown in gray. The ligand from the crystal structure is shown in purple. The bottom right only shows the docked results. The table shows the docking score and glide gscore. A smaller value shows a better docking result. The AF2 predicted RXR-α model shows a very impressive docking result, and the binding site's sidechains are also showing very good prediction.

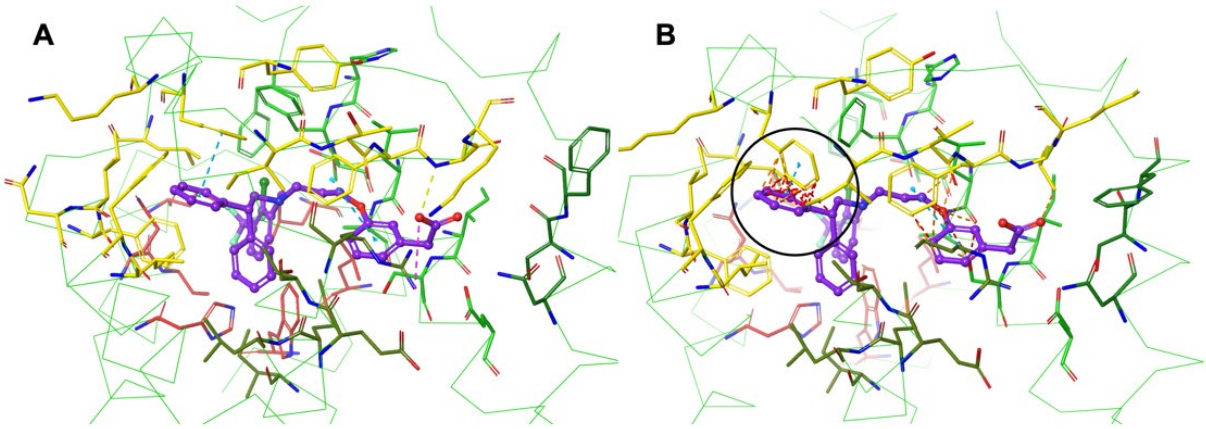


Figure 4.8 The binding site of LXR-β. A. shows the crystal structure of LXR- β.

A. shows the crystal structure of LXR- β (PDB ID: 4NQA). The ligand is in purple. The blue dashed line shows the pi-pi stacking. The yellow dash shows the hydrogen bond. The magenta dashed line shows the salt bridge. B. shows the AF2 predicted LXR-β model. After superimposed the crystal structure and the AF2 model, the ligand lies inside of the AF2 model's binding pocket. The ligand can form similar interactions with the crystal structure. However, there is also a bad clash between the ligand and receptor's sidechains. The red dashed line shows the ugly clash.

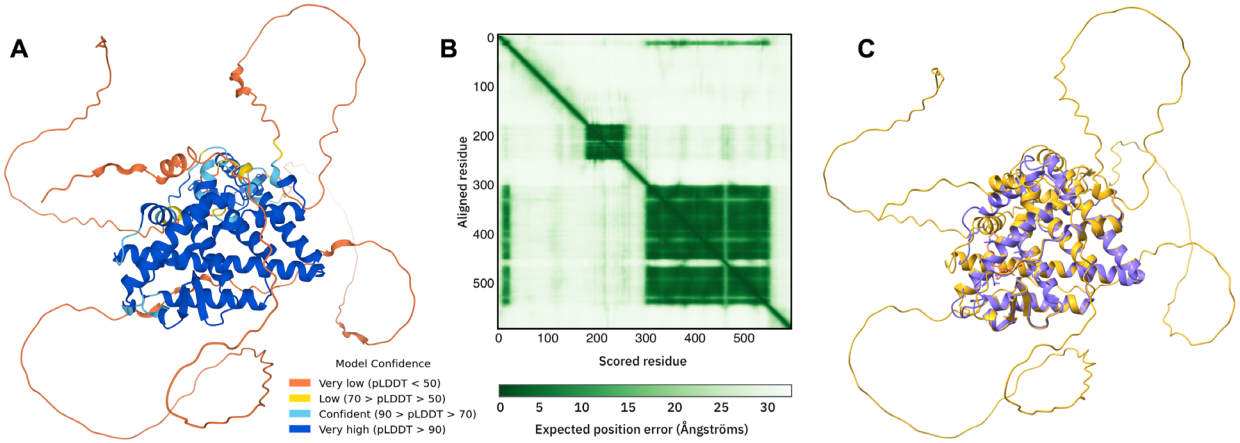


Figure 4.9 The ERα AlphaFold2 predicted model, and the superimposition with the crystal structure of ERα.

A. The structure is the AlphaFold2 predicted model of full-length wildtype ERα. The colors indicate different confidences. B. shows the AF2 predicted expected position error. C. shows the superimposition of the crystal structure and AF2 model. The gold shows the AF2 predicted structure. The purple shows the crystal structure. (PDB ID: 1YIM)

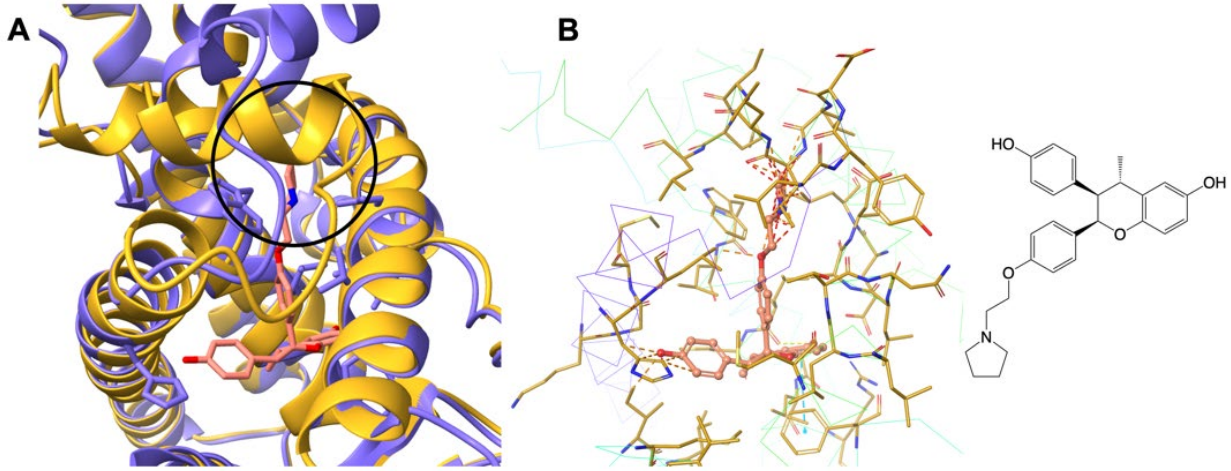


Figure 4.10 The superimposed binding site of wild type estrogen receptor alpha.

The crystal structure (PDB ID:1YIM) is shown in purple. The AF2 predicted wild type full-length estrogen receptor alpha structure is shown in gold. The ligand is in pink. In the black circle, the ligand clashes with the helix from the AF2 predicted structure. B. shows the steric clashes of ligand with the AF2 predicted structure. The color code is the same as A. The red dashed lines show the bad clashes. The chemical formula is for the ligand.

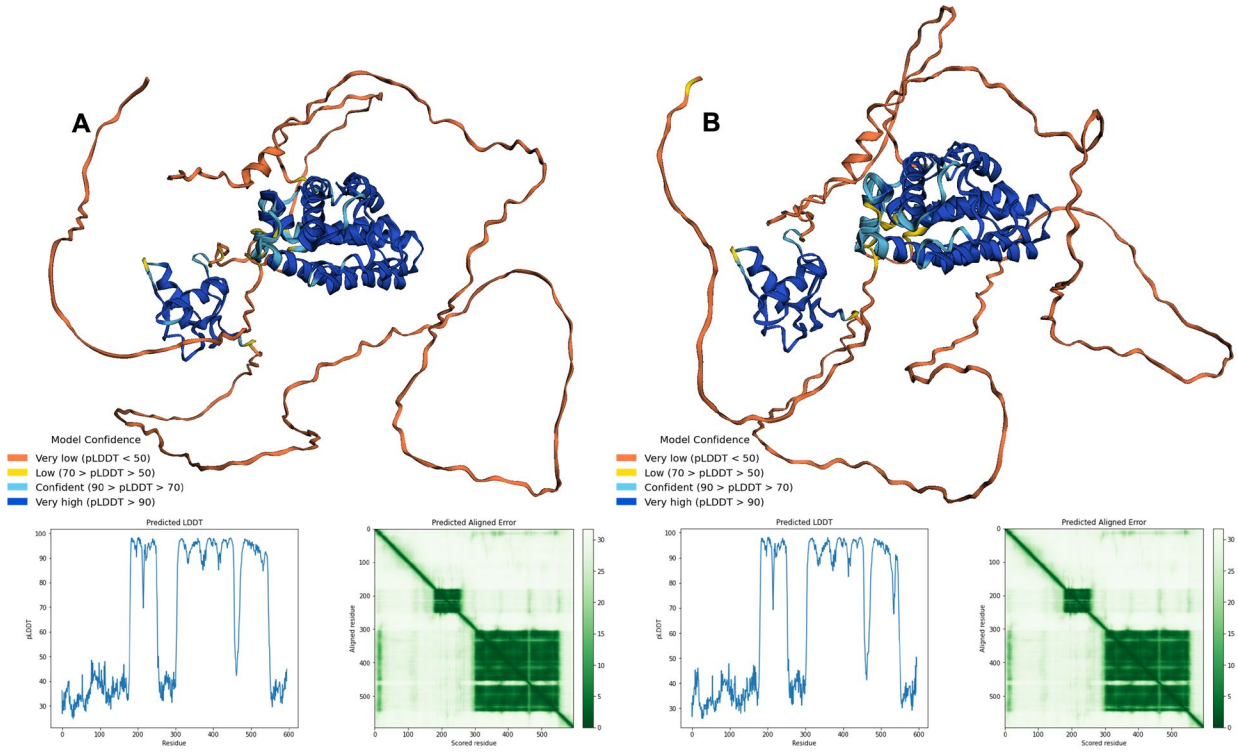


Figure 4.11 The AlphaFold2 predicted estrogen receptor alpha two mutants' structures, Y537S, and D538G.

A. The structure is the AlphaFold2 predicted model of Estrogen Receptor alpha Y537S mutant. The colors indicate different confidences. B. The structure is the AlphaFold2 predicted model of Estrogen Receptor alpha D538G mutant. The colors indicate different confidences. Under each structure, left is the pLDDT of the predicted model of Estrogen Receptor Alpha Mutant Y537A, right is the predicted aligned error of the predicted model.

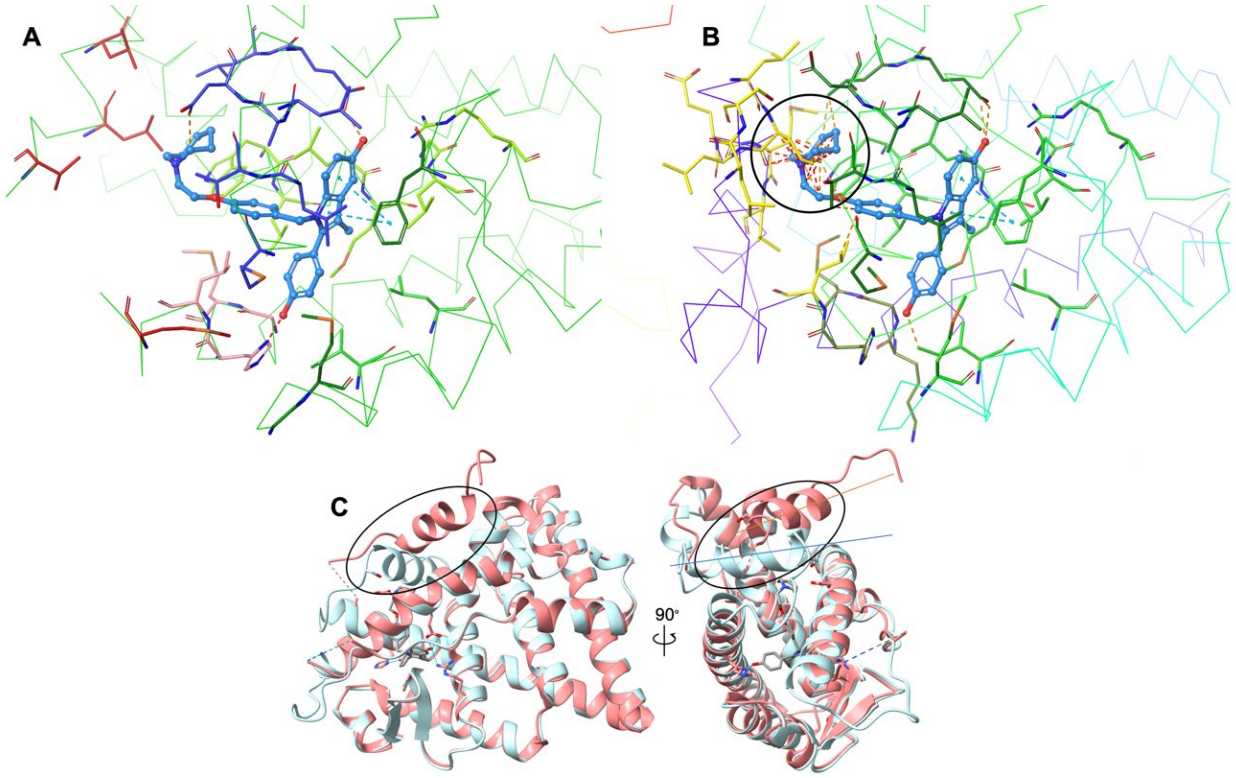


Figure 4.12 The ERa mutant Y537S superimposition of the crystal structure and the AlphaFold2 predicted structure of binding site.

A. shows the binding site of ERa mutant Y537S crystal structure. The blue dashed lines show the pi-pi stacking. The orange dashed line shows the bad clash between the ligand and receptor. The red dashed line shows the ugly clash between the ligand and receptor. The ligand is in blue. Other sticks show the sidechains. B. shows the binding site of full-length ERa mutant Y537S AF2 predicted model. The ligand can't dock into this model. This is the superimposition of the binding site. Most of the sidechains in the binding site of the AF2 predicted model are quite good, which can form the same pi stacking with the ligand. However, there are very bad clashes in the black circle. There is not enough room to let the ligand go into it. This causes the ligand to be unable to dock into this binding site.

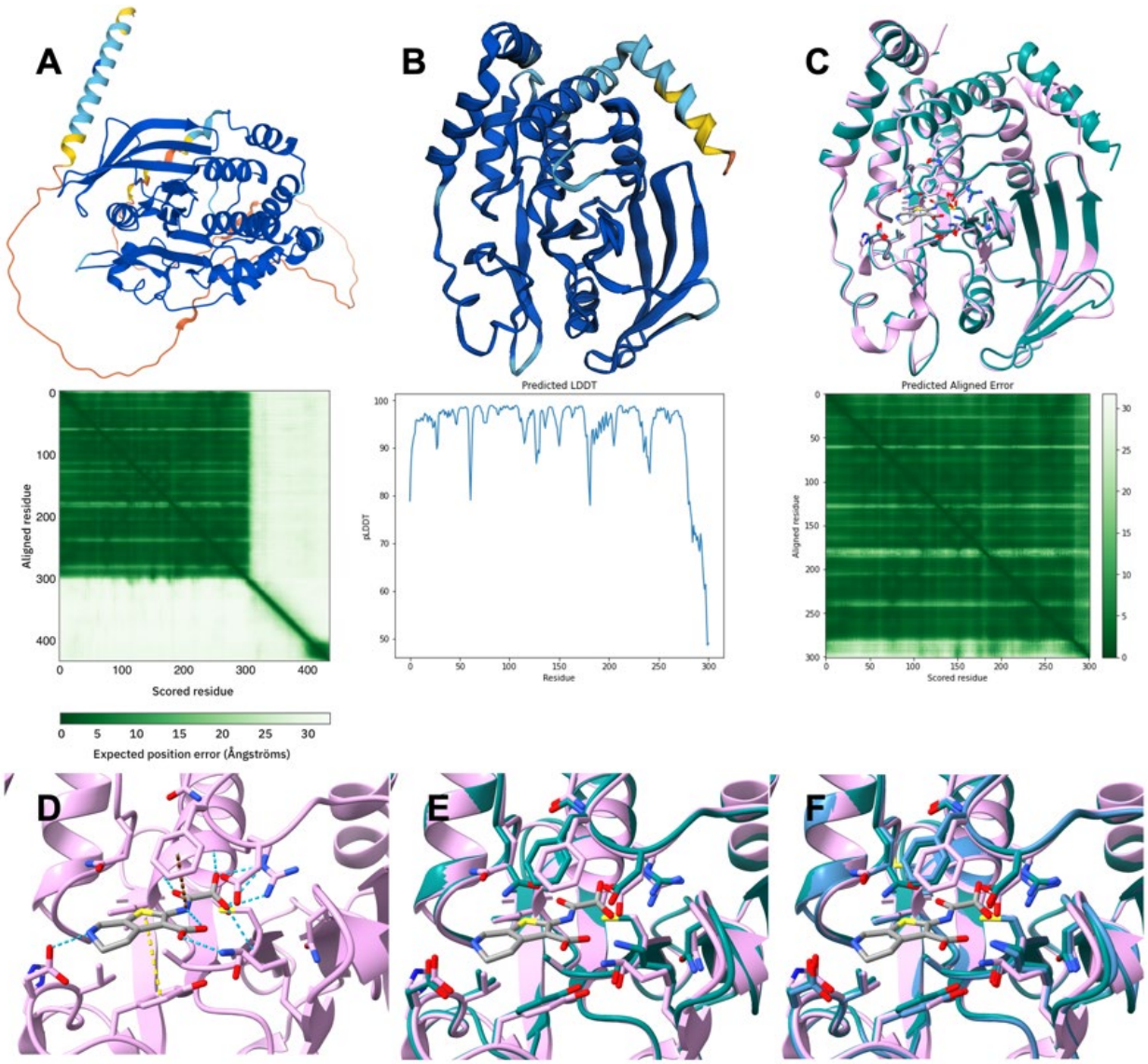
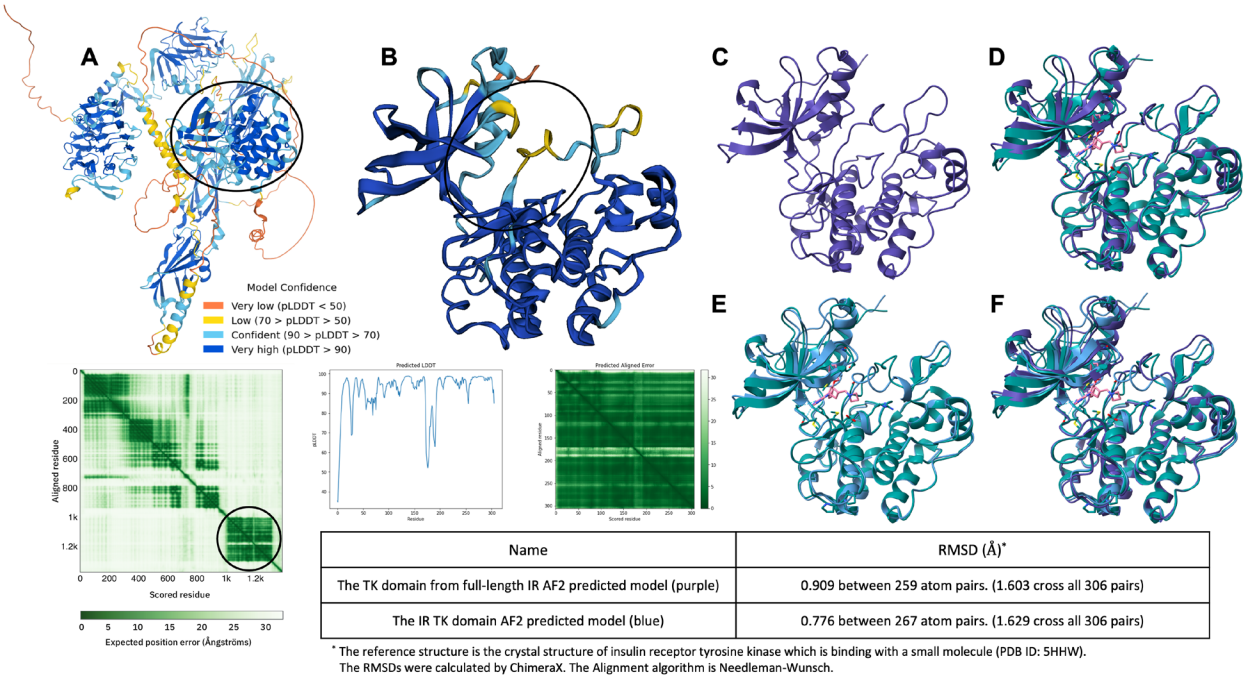


Figure 4.13 The AlphaFold2 predicted PTP1B structure.

A. shows the full-length PTP1B AF2 predicted structure. B. shows the AF2 predicted PTP1B which is the same protein sequence with the crystal structure (PDB ID:5K9W). Under figure A and B are the confidences of the two predicted structures. C. shows the superimposed structures of crystal structure and the AF2 predicted structure. The crystal structure 5K9W is shown in pink. The AF2 predicted 5K9W model is shown in teal. D. shows the zoomed-in view of the ligand binding site. The cyan dashed lines show hydrogen bonds. The yellow dashed line shows

the pi-pi stacking. The brown dashed line shows the cation-pi interaction. E. shows superimposed structures of crystal structure (PDB ID: 5K9W) and the AF2 predicted 5K9W. F. shows superimposed binding site of crystal structure (PDB ID: 5K9W) and the AF2 predicted full-length PTP1B and the 5K9W structures.

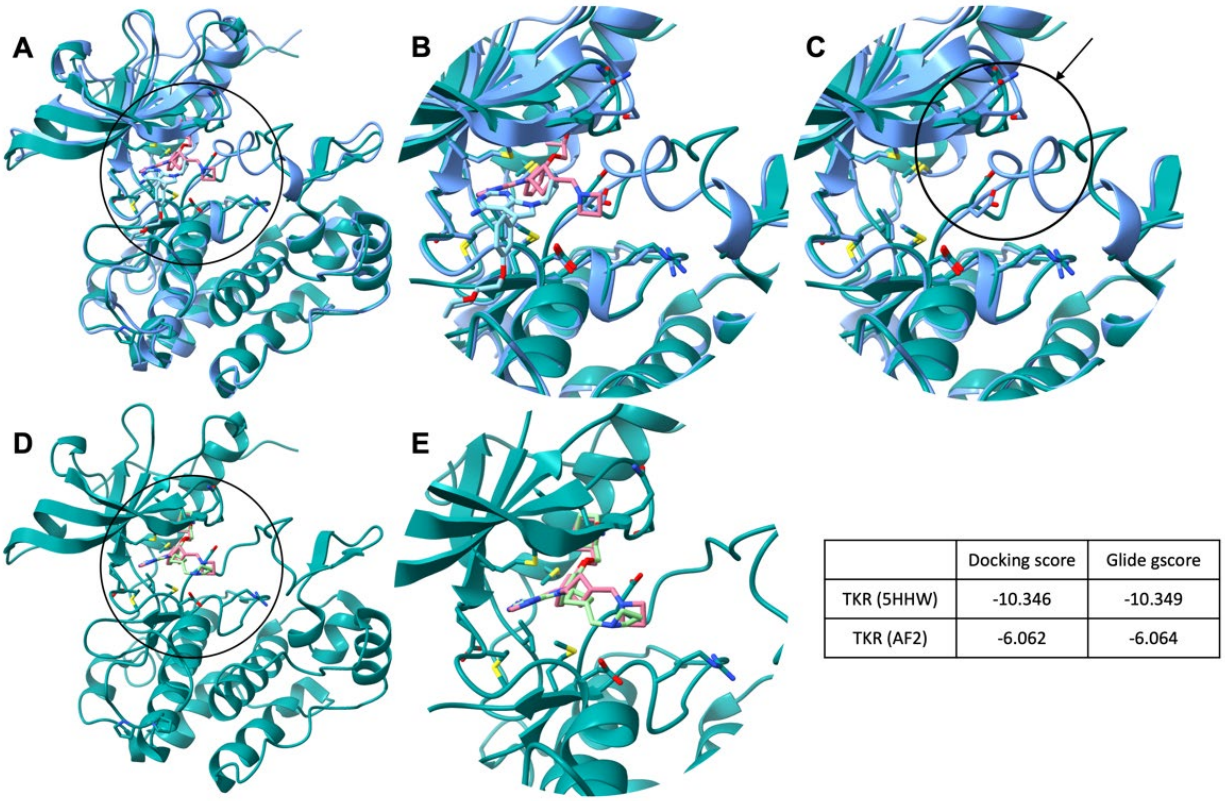


Name	RMSD (Å)*
The TK domain from full-length IR AF2 predicted model (purple)	0.909 between 259 atom pairs. (1.603 cross all 306 pairs)
The IR TK domain AF2 predicted model (blue)	0.776 between 267 atom pairs. (1.629 cross all 306 pairs)

* The reference structure is the crystal structure of insulin receptor tyrosine kinase which is binding with a small molecule (PDB ID: 5HHW).
The RMSDs were calculated by ChimeraX. The Alignment algorithm is Needleman-Wunsch.

Figure 4.14 The AlphaFold2 predicted models and the comparison of AlphaFold2 predicted models with the crystal structures.

A. shows the AF2 predicted model of full-length of insulin receptor. The colors indicate plDDT model confidence and the expected position error figure is below the structure. B. shows the AF2 predicted TK domain model. Under the structure is the plDDT figure and expected position error. From the two expected position error figures, the TK domain AF2 model has better confidence than the AF2 model of TK domain from full-length model. C. shows the AF2 model of TK domain from the full-length IR. D is the superimposed structures of C and the crystal structure 5HHW. E. shows the superimposed structures of B and the crystal structure 5HHW. F. shows the superimposing three structures of B, C, and 5HHW. The table shows the RMSD values.



	Docking score	Glide gscore
TKR (5HHW)	-10.346	-10.349
TKR (AF2)	-6.062	-6.064

Figure 4.15 The docking results of insulin receptor tyrosine kinase receptor.

A. B. and C. show the AF2 predicted TKR domain model docking results. A. shows the superimposition of the AF2 predicted TKR domain model and crystal structure. The AF2 predicted TKR domain model is shown in blue, and the docked ligand is in cyan. The crystal structure is shown in green, and the ligand is in pink. The black circle shows the binding pocket of TKR domain. B. shows the zoomed in view of the binding pocket. C. shows the binding site. The black circle labels the difference of the loop which effects the binding pocket and causes the ligand to be unable to dock properly. D. and E. show the crystal structure and the docked ligand. The black circle shows the binding pocket. The crystal structure ligand is shown in pink. The docked ligand is shown in light green. The table shows the docking scores and glide gscores of using the AF2 predicted structure and crystal structure.

Chapter 5 - Conclusion

This work focused on using computational methods to solve and answer questions about biological systems. Chapter 2 shows how to integrate different computational methods to reveal the interactions between the peptide and protein. Chapter 3 shows the real behavior of red fluorescent protein by MD simulations. Chapter 4 assessed the protein modelling methods which present the predicted models' quality and how to use it for other research and drug design.

The study of proctolin and proctolin receptor by computational methods

In chapter 2, we investigate the integration of different computational methods, such as protein modelling, virtual screening, ligand docking, and MD simulations. We built the proctolin receptor homology model and docked proctolin into the receptor, and also studied kinetics by MD simulations. The workflow (shown in Figure 2.1) of this research can be a study model for ligands or peptides binding with unknown structure receptors.

From this study, we obtained the complex of proctolin and proctolin receptor. By running MD simulations to obtain the binding pose, we revealed the mechanism of the binding, and found that cation-pi interactions play a very important role in the binding. We can get multiple conclusions from the MD simulations results.

First, the results show Tyr 99 and Arg 111 from proctolin receptor are the critical binding sites. Tyr 99 from the receptor binds to the first Arg from proctolin. Arg 111 from the receptor binds to the second Tyr from proctolin. This result also matches with the alanine scan results from Dr. Yoonseong Park's lab. Second, the cation-pi interactions play a very important role in this binding, which is an electrostatic interaction between a cation and a polarizable pi system, such as an aromatic ring. Cation-pi interactions tend to be stronger than hydrogen bonds and pi-stacking interactions (Hunter and Sanders, 1990; Sinnokrot et al., 2002). In this complex, the Arg

1 from proctolin binds to Tyr 99 through a cation-pi interaction. Tyr2 from proctolin binds to Arg 111 through a cation-pi interaction as well. Third, the electrostatic surface study of the proctolin and proctolin receptor complex shows the Arg 1 from proctolin acts as a "cation anchor". This allows the proctolin to dock into the binding pocket and stabilizes the binding. Compared with other neurotensin and neurotensin receptors complex, the neurotensin binding pocket has a positive charge rich pocket. Fourth, the kinetic study of the binding shows the cation-pi interaction between Arg 1 from proctolin and Tyr 99 from the receptor forms first. Then the cation-pi interaction between Tyr2 from proctolin and Arg 111 forms.

This study found the binding pocket and the critical binding sites. These results show the proctolin and proctolin receptor binding mechanism and show the binding mechanism from the kinetic study as well.

The behavior of red fluorescent protein chromophore at room temperature

Fluorescent proteins have been applied to multiple study fields; hence, it has been studied in order to engineer and enhance fluorescent proteins' performance. The local environment of chromophores and the interactions between the chromophores and protein matrix have been shown to have a huge effect on the intensity of fluorescence (Shaner *et al.*, 2007). Obtaining fluorescent proteins' crystal structures is important to help in engineering new fluorescent proteins, however, most crystal structures were obtained at cryo-state. Protein structures are affected by temperatures (Bradford et al., 2021), and a crystal structure is also only one conformation. Most fluorescent protein crystal structures chromophores were captured in a planar conformation. Here, we performed MD simulations on trans-chromophore red fluorescent proteins and were able to come to multiple conclusions. First, the trans-chromophores of red fluorescent proteins are non-planar at 300K. The difference in planarity as seen in crystal

structures causes the interactions between the chromophores and protein matrixes to be different as well. Second, the dihedral of the phenol ring and imidazole ring will "wiggle" at certain angles. This also shows the chromophore is not stable and keeps wiggling at 300K. The wiggling of angles is around 30º. Lastly, residue site 197 interacts with the chromophore and either forms a pi-stacking or a cation-pi interaction to stabilize the chromophore.

The application and assessment of AlphaFold2 in drug design

AlphaFold2 was released late last year. It was the biggest news in structural biology. Predicting protein structures is always a big challenge. After AlphaFold2 was available, this method has been applied to multiple different biological and biochemical fields. In this chapter, we assessed AlphaFold2 and applied it to drug design using some typical drug targets. The results of the prediction of AF2 are complicated. AF2 did not generate a good result for the RXRα-LXRβ complex, specifically the DNA-binding domain. There are also only very limited crystal structures available to use for training. This could be the reason the prediction of the DNA-binding domain was so poor. The next thing we wanted to predict was mutant structures. Hence, estrogen receptor alpha, Y537S, and D538G were tested. Unfortunately, these cannot be used in drug design. Most parts of these proteins were predicted with high confidence. However, the most important helix 12 is in the wrong orientation and blocks the binding site which causes the ligand to be unable to dock into the pocket. PTP1B has two states, open and closed states. The closed state is the ligand binding mode. AF2 successfully predicted the closed state. After the docking study, the docking result is quite close to the crystal structure docking result. The last protein we studied is insulin receptor. It is too hard to predict a protein like insulin receptor. Hence, we also did not see a good prediction. However, we did examine the tyrosine kinase domain. The predicted structure of the tyrosine kinase domain shows high confidence except for the binding domain, though the docking study did not give us a good result.

All the proteins studied in this work already have crystal structures in the Protein Data Bank. Even though there are crystal structures available, it is still very hard to predict the structures using AF2 that meet drug design requirements. When using AF2 to predict protein structures, we need to be very careful. There is still no question about the huge contribution to science by AF2. It is trying to fill the gap between the number of experimental structures and unknown DNA coded proteins. We believe this work can help people have a better understanding of AF2 predicted structures and it is a brief guide of how to use AF2.

References

Alexander, S.P., Davenport, A.P., Kelly, E., Marrion, N., Peters, J.A., Benson, H.E., Faccenda, E., Pawson, A.J., Sharman, J.L., and Southan, C. (2015). The Concise Guide to PHARMACOLOGY 2015/16: G protein‐coupled receptors. British journal of pharmacology *172*, 5744-5869.

Ali, S., and Coombes, R.C. (2000). Estrogen receptor alpha in human breast cancer: occurrence and significance. Journal of mammary gland biology and neoplasia *5*, 271-281.

Allen, J.A., and Roth, B.L. (2011). Strategies to discover unexpected targets for drugs active at G protein–coupled receptors. Annual review of pharmacology and toxicology *51*, 117-144.

Andersen, J.N., Mortensen, O.H., Peters, G.n.H., Drake, P.G., Iversen, L.F., Olsen, O.H., Jansen, P.G., Andersen, H.S., Tonks, N.K., and Møller, N.P.H. (2001). Structural and evolutionary relationships among protein tyrosine phosphatase domains. Molecular and cellular biology *21*, 7117-7136.

Attramadal, H., Arriza, J.L., Aoki, C., Dawson, T.M., Codina, J., Kwatra, M.M., Snyder, S.H., Caron, M.G., and Lefkowitz, R.J. (1992). Beta-arrestin2, a novel member of the arrestin/beta-arrestin gene family. Journal of Biological Chemistry *267*, 17882-17890.

Bai, Q., Shi, D., Zhang, Y., Liu, H., and Yao, X. (2014). Exploration of the antagonist CP-376395 escape pathway for the corticotropin-releasing factor receptor 1 by random acceleration molecular dynamics simulations. Molecular BioSystems *10*, 1958-1967.

Bar-Shavit, R., Maoz, M., Kancharla, A., Nag, J.K., Agranovich, D., Grisaru-Granovsky, S., and Uziely, B. (2016). G protein-coupled receptors in cancer. International journal of molecular sciences *17*, 1320.

Benovic, J., Kühn, H., Weyand, I., Codina, J., Caron, M., and Lefkowitz, R. (1987). Functional desensitization of the isolated beta-adrenergic receptor by the beta-adrenergic receptor kinase: potential role of an analog of the retinal protein arrestin (48-kDa protein). Proceedings of the National Academy of Sciences *84*, 8879-8882.

Berthoud, H., Imdorf, A., Haueter, M., Radloff, S., and Neumann, P. (2010). Virus infections and winter losses of honey bee colonies (Apis mellifera). Journal of Apicultural Research *49*, 60-65.

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

Bevis, B.J., and Glick, B.S. (2002). Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed). Nature biotechnology *20*, 83-87.

Bock, A., Schrage, R., and Mohr, K. (2018). Allosteric modulators targeting CNS muscarinic receptors. Neuropharmacology *136*, 427-437.

Bradford, S.Y., El Khoury, L., Ge, Y., Osato, M., Mobley, D.L., and Fischer, M. (2021). Temperature artifacts in protein structures bias ligand-binding predictions. Chemical science *12*, 11275-11293.

Bremer, P.T., Kimishima, A., Schlosburg, J.E., Zhou, B., Collins, K.C., and Janda, K.D. (2016). Combatting synthetic designer opioids: a conjugate vaccine ablates lethal doses of fentanyl class drugs. Angewandte Chemie *128*, 3836-3839.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S.a., and Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. Journal of computational chemistry *4*, 187-217.

Brothers, S.P., Janovick, J.A., and Conn, P.M. (2003). Unexpected effects of epitope and chimeric tags on gonadotropin-releasing hormone receptors: implications for understanding the molecular etiology of hypogonadotropic hypogonadism. The Journal of Clinical Endocrinology & Metabolism *88*, 6107-6112.

Burg, J.S., Ingram, J.R., Venkatakrishnan, A., Jude, K.M., Dukkipati, A., Feinberg, E.N., Angelini, A., Waghray, D., Dror, R.O., and Ploegh, H.L. (2015). Structural basis for chemokine recognition and activation of a viral G protein–coupled receptor. Science *347*, 1113-1117.

Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S., Zacharias, D.A., and Tsien, R.Y. (2002). A monomeric red fluorescent protein. Proceedings of the National Academy of Sciences *99*, 7877-7882.

Case, D.A., Darden, T.A., Cheatham, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Crowley, M., Walker, R.C., and Zhang, W. (2008). Amber 10. University of California.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-805.

Chan, H.S., Wang, J., Palczewski, K., Filipek, S., Vogel, H., Liu, Z.-J., and Yuan, S. (2018). Exploring a new ligand binding site of G protein-coupled receptors. Chemical science *9*, 6480-6489.

Chawla, A., Repa, J.J., Evans, R.M., and Mangelsdorf, D.J. (2001). Nuclear receptors and lipid physiology: opening the X-files. Science *294*, 1866-1870.

Chiang, R.G., Martens, J.D., and O'DONNELL, M.J. (2010). The vagina muscles of the blood‐sucking insect Rhodnius prolixus as a model for exploring the physiological effects of proctolin. Physiological Entomology *35*, 154-159.

Choy, M.S., Li, Y., Machado, L.E., Kunze, M.B., Connors, C.R., Wei, X., Lindorff-Larsen, K., Page, R., and Peti, W. (2017). Conformational rigidity and protein dynamics at distinct timescales regulate PTP1B activity and allostery. Molecular cell *65*, 644-658. e645.

Christopoulos, A. (2014). Advances in G protein-coupled receptor allostery: from function to structure. Molecular pharmacology *86*, 463-478.

Conibear, A.E., and Kelly, E. (2019). A biased view of μ-opioid receptors? Molecular pharmacology *96*, 542-549.

Cournia, Z., Allen, T.W., Andricioaei, I., Antonny, B., Baum, D., Brannigan, G., Buchete, N.-V., Deckman, J.T., Delemotte, L., and Del Val, C. (2015). Membrane protein structure, function, and dynamics: a perspective from experiments and theory. The Journal of membrane biology *248*, 611-640.

Craft, C.M., Whitmore, D.H., and Wiechmann, A.F. (1994). Cone arrestin identified by targeting expression of a functional family. Journal of Biological Chemistry *269*, 4613-4619.

Crozier, P.S., Stevens, M.J., Forrest, L.R., and Woolf, T.B. (2003). Molecular dynamics simulation of dark-adapted rhodopsin in an explicit membrane bilayer: coupling between local retinal and larger scale conformational change. Journal of molecular biology *333*, 493-514.

Dainat, B., Evans, J.D., Chen, Y.P., Gauthier, L., and Neumann, P. (2012). Predictive markers of honey bee colony collapse. PLoS one *7*, e32151.

Daub, H., Ulrich Weiss, F., Wallasch, C., and Ullrich, A. (1996). Role of transactivation of the EGF receptor in signalling by G-protein-coupled receptors. Nature *379*, 557-560.

De Vivo, M., Masetti, M., Bottegoni, G., and Cavalli, A. (2016). Role of molecular dynamics and related methods in drug discovery. Journal of medicinal chemistry *59*, 4035-4061.

DeWire, S.M., Ahn, S., Lefkowitz, R.J., and Shenoy, S.K. (2007). β-Arrestins and cell signaling. Annu. Rev. Physiol. *69*, 483-510.

Di Prisco, G., Pennacchio, F., Caprio, E., Boncristiani Jr, H.F., Evans, J.D., and Chen, Y. (2011). Varroa destructor is an effective vector of Israeli acute paralysis virus in the honeybee, Apis mellifera. Journal of General Virology *92*, 151-155.

Dircksen, H., Neupert, S., Predel, R., Verleyen, P., Huybrechts, J., Strauss, J., Hauser, F., Stafflinger, E., Schneider, M., and Pauwels, K. (2011). Genomics, transcriptomics, and peptidomics of Daphnia pulex neuropeptides and protein hormones. Journal of proteome research *10*, 4478-4504.

Doré, A.S., Okrasa, K., Patel, J.C., Serrano-Vega, M., Bennett, K., Cooke, R.M., Errey, J.C., Jazayeri, A., Khan, S., and Tehan, B. (2014). Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. Nature *511*, 557-562.

Dror, R.O., Arlow, D.H., Borhani, D.W., Jensen, M.Ø., Piana, S., and Shaw, D.E. (2009). Identification of two distinct inactive conformations of the β2-adrenergic receptor reconciles structural and biochemical observations. Proceedings of the National Academy of Sciences *106*, 4689-4694.

Dror, R.O., Arlow, D.H., Maragakis, P., Mildorf, T.J., Pan, A.C., Xu, H., Borhani, D.W., and Shaw, D.E. (2011a). Activation mechanism of the β2-adrenergic receptor. Proceedings of the National Academy of Sciences *108*, 18684-18689.

Dror, R.O., Dirks, R.M., Grossman, J., Xu, H., and Shaw, D.E. (2012). Biomolecular simulation: a computational microscope for molecular biology. Annual review of biophysics *41*, 429-452.

Dror, R.O., Green, H.F., Valant, C., Borhani, D.W., Valcourt, J.R., Pan, A.C., Arlow, D.H., Canals, M., Lane, J.R., and Rahmani, R. (2013). Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. Nature *503*, 295-299.

Dror, R.O., Pan, A.C., Arlow, D.H., Borhani, D.W., Maragakis, P., Shan, Y., Xu, H., and Shaw, D.E. (2011b). Pathway and mechanism of drug binding to G-protein-coupled receptors. Proceedings of the National Academy of Sciences *108*, 13118-13123.

Esguerra, M., Siretskiy, A., Bello, X., Sallander, J., and Gutiérrez-de-Terán, H. (2016). GPCR-ModSim: A comprehensive web based solution for modeling G-protein coupled receptors. Nucleic acids research *44*, W455-W462.

Evans, R.M. (2005). The nuclear receptor superfamily: a rosetta stone for physiology. Molecular Endocrinology *19*, 1429-1438.

Felder, C.C., Goldsmith, P.J., Jackson, K., Sanger, H.E., Evans, D.A., Mogg, A.J., and Broad, L.M. (2018). Current status of muscarinic M1 and M4 receptors as drug targets for neurodegenerative diseases. Neuropharmacology *136*, 449-458.

Fojo, T., and Bates, S. (2013). Mechanisms of resistance to PARP inhibitors—three and counting. Cancer discovery *3*, 20-23.

Foster, D.J., and Conn, P.J. (2017). Allosteric modulation of GPCRs: new insights and potential utility for treatment of schizophrenia and other CNS disorders. Neuron *94*, 431-446.

Frey, K.M., Georgiev, I., Donald, B.R., and Anderson, A.C. (2010). Predicting resistance mutations using protein design algorithms. Proceedings of the National Academy of Sciences *107*, 13707-13712.

Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., and Perry, J.K. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. Journal of medicinal chemistry *47*, 1739-1749.

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

Gachet, C. (2008). P2 receptors, platelet function and pharmacological implications. Thrombosis and haemostasis *99*, 466-472.

Gainetdinov, R.R., Premont, R.T., Bohn, L.M., Lefkowitz, R.J., and Caron, M.G. (2004). Desensitization of G protein–coupled receptors and neuronal functions. Annu. Rev. Neurosci. *27*, 107-144.

Ghosh, A., Saran, N., and Saha, S. (2020). Survey of drug resistance associated gene mutations in Mycobacterium tuberculosis, ESKAPE and other bacterial species. Scientific reports *10*, 1-11.

Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Science *27*, 14-25.

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

Grossfield, A., Pitman, M.C., Feller, S.E., Soubias, O., and Gawrisch, K. (2008). Internal hydration increases during activation of the G-protein-coupled receptor rhodopsin. Journal of molecular biology *381*, 478-486.

Gumbart, J.C., Teo, I., Roux, B., and Schulten, K. (2013). Reconciling the roles of kinetic and thermodynamic factors in membrane–protein insertion. Journal of the American Chemical Society *135*, 2291-2297.

Gurskaya, N.G., Fradkov, A.F., Terskikh, A., Matz, M.V., Labas, Y.A., Martynov, V.I., Yanushevich, Y.G., Lukyanov, K.A., and Lukyanov, S.A. (2001). GFP‐like chromoproteins as a source of far‐red fluorescent proteins. FEBS letters *507*, 16-20.

Gutmann, T., Kim, K.H., Grzybek, M., Walz, T., and Coskun, Ü. (2018). Visualization of ligand-induced transmembrane signaling in the full-length human insulin receptor. Journal of Cell Biology *217*, 1643-1649.

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

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

Hanson, M.A., Cherezov, V., Griffith, M.T., Roth, C.B., Jaakola, V.-P., Chien, E.Y., Velasquez, J., Kuhn, P., and Stevens, R.C. (2008). A specific cholesterol binding site is established by the 2.8 Å structure of the human β2-adrenergic receptor. Structure *16*, 897-905.

Hauser, A.S., Attwood, M.M., Rask-Andersen, M., Schiöth, H.B., and Gloriam, D.E. (2017). Trends in GPCR drug discovery: new agents, targets and indications. Nature reviews Drug discovery *16*, 829-842.

Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of chemical theory and computation *4*, 435-447.

Highfield, A.C., El Nagar, A., Mackinder, L.C., Noël, L.M.-L., Hall, M.J., Martin, S.J., and Schroeder, D.C. (2009). Deformed wing virus implicated in overwintering honeybee colony losses. Applied and environmental microbiology *75*, 7212-7220.

Hillesheim, E., Ritter, W., and Bassand, D. (1996). First data on resistance mechanisms of Varroa jacobsoni (Oud.) against tau-fluvalinate. Experimental & applied acarology *20*, 283-296.

Huang, J., and MacKerell Jr, A.D. (2013). CHARMM36 all‐atom additive protein force field: Validation based on comparison to NMR data. Journal of computational chemistry *34*, 2135-2145.

Huang, W., Manglik, A., Venkatakrishnan, A., Laeremans, T., Feinberg, E.N., Sanborn, A.L., Kato, H.E., Livingston, K.E., Thorsen, T.S., and Kling, R.C. (2015a). Structural insights into μ-opioid receptor activation. Nature *524*, 315-321.

Huang, X.-P., Karpiak, J., Kroeze, W.K., Zhu, H., Chen, X., Moy, S.S., Saddoris, K.A., Nikolova, V.D., Farrell, M.S., and Wang, S. (2015b). Allosteric ligands for the pharmacologically dark receptors GPR68 and GPR65. Nature *527*, 477-483.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. Journal of molecular graphics *14*, 33-38.

Hunter, C.A., and Sanders, J.K. (1990). The nature of. pi.-. pi. interactions. Journal of the American Chemical Society *112*, 5525-5534.

Hunter, T. (2000). Signaling—2000 and beyond. Cell *100*, 113-127.

Isaac, R.E., Taylor, C.A., Hamasaka, Y., Nässel, D.R., and Shirras, A.D. (2004). Proctolin in the post-genomic era: new insights and challenges. Invertebrate Neuroscience *5*, 51-64.

Isberg, V., de Graaf, C., Bortolato, A., Cherezov, V., Katritch, V., Marshall, F.H., Mordalski, S., Pin, J.-P., Stevens, R.C., and Vriend, G. (2015). Generic GPCR residue numbers–aligning topology maps while minding the gaps. Trends in pharmacological sciences *36*, 22-31.

Ito, K.-i., Olsen, S.L., Qiu, W., Deeley, R.G., and Cole, S.P. (2001). Mutation of a single conserved tryptophan in multidrug resistance protein 1 (MRP1/ABCC1) results in loss of drug resistance and selective loss of organic anion transport. Journal of Biological Chemistry *276*, 15616-15624.

Jacobson, K.A., Deflorian, F., Mishra, S., and Costanzi, S. (2011). Pharmacochemistry of the platelet purinergic receptors. Purinergic signalling *7*, 305-324.

Jakobsson, T., Treuter, E., Gustafsson, J.-Å., and Steffensen, K.R. (2012). Liver X receptor biology and pharmacology: new pathways, challenges and opportunities. Trends in pharmacological sciences *33*, 394-404.

Jazayeri, A., Dias, J.M., and Marshall, F.H. (2015). From G protein-coupled receptor structure resolution to rational drug design. Journal of Biological Chemistry *290*, 19489-19495.

Jazayeri, A., Doré, A.S., Lamb, D., Krishnamurthy, H., Southall, S.M., Baig, A.H., Bortolato, A., Koglin, M., Robertson, N.J., and Errey, J.C. (2016). Extra-helical binding site of a glucagon receptor antagonist. Nature *533*, 274-277.

Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008). CHARMM‐GUI: a web‐based graphical user interface for CHARMM. Journal of computational chemistry *29*, 1859-1865.

Johnson, E.C., Garczynski, S.F., Park, D., Crim, J.W., Nässel, D.R., and Taghert, P.H. (2003). Identification and characterization of a G protein-coupled receptor for the neuropeptide proctolin in Drosophila melanogaster. Proceedings of the National Academy of Sciences *100*, 6198-6203.

Johnson, R.M., Dahlgren, L., Siegfried, B.D., and Ellis, M.D. (2013). Acaricide, fungicide and drug interactions in honey bees (Apis mellifera). PloS one *8*, e54092.

Jordan, V.C. (2003). Tamoxifen: a most unlikely pioneering medicine. Nature reviews Drug discovery *2*, 205-213.

Jorgensen, W.L., and Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. Journal of the American Chemical Society *110*, 1657-1666.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., and Potapenko, A. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589.

Kaczor, A.A., Rutkowska, E., Bartuzi, D., Targowska-Duda, K.M., Matosiuk, D., and Selent, J. (2016). Computational methods for studying G protein-coupled receptors (GPCRs). In Methods in cell biology, (Elsevier), pp. 359-399.

Karplus, M., and McCammon, J.A. (2002). Molecular dynamics simulations of biomolecules. Nature structural biology *9*, 646-652.

Kaserer, T., Lantero, A., Schmidhammer, H., Spetea, M., and Schuster, D. (2016). μ Opioid receptor: novel antagonists and structural modeling. Scientific reports *6*, 1-15.

Kaya, C., Armutlulu, A., Ekesan, S., and Haliloglu, T. (2013). MCPath: Monte Carlo path generation approach to predict likely allosteric pathways and functional residues. Nucleic acids research *41*, W249-W255.

Khalili-Araghi, F., Gumbart, J., Wen, P.-C., Sotomayor, M., Tajkhorshid, E., and Schulten, K. (2009). Molecular dynamics simulations of membrane channels and transporters. Current opinion in structural biology *19*, 128-137.

Kirchmair, J., Markt, P., Distinto, S., Wolber, G., and Langer, T. (2008). Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—what can we learn from earlier mistakes? Journal of computer-aided molecular design *22*, 213-228.

Koehl, A., Hu, H., Maeda, S., Zhang, Y., Qu, Q., Paggi, J.M., Latorraca, N.R., Hilger, D., Dawson, R., and Matile, H. (2018). Structure of the μ-opioid receptor–Gi protein complex. Nature *558*, 547-552.

Kohlhoff, K.J., Shukla, D., Lawrenz, M., Bowman, G.R., Konerding, D.E., Belov, D., Altman, R.B., and Pande, V.S. (2014). Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. Nature chemistry *6*, 15-21.

Kooistra, A.J., Leurs, R., De Esch, I.J., and Graaf, C.d. (2014). From three-dimensional GPCR structure to rational ligand discovery. G Protein-Coupled Receptors-Modeling and Simulation, 129-157.

Korach-André, M., Archer, A., Barros, R.P., Parini, P., and Gustafsson, J.-Å. (2011). Both liver-X receptor (LXR) isoforms control energy expenditure by regulating brown adipose tissue activity. Proceedings of the National Academy of Sciences *108*, 403-408.

Krishnan, N., Koveal, D., Miller, D.H., Xue, B., Akshinthala, S.D., Kragelj, J., Jensen, M.R., Gauss, C.-M., Page, R., and Blackledge, M. (2014). Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. Nature chemical biology *10*, 558-566.

Kruse, A.C., Hu, J., Pan, A.C., Arlow, D.H., Rosenbaum, D.M., Rosemond, E., Green, H.F., Liu, T., Chae, P.S., and Dror, R.O. (2012). Structure and dynamics of the M3 muscarinic acetylcholine receptor. Nature *482*, 552-556.

Kruse, A.C., Kobilka, B.K., Gautam, D., Sexton, P.M., Christopoulos, A., and Wess, J. (2014). Muscarinic acetylcholine receptors: novel opportunities for drug development. Nature reviews Drug discovery *13*, 549-560.

Kruse, A.C., Ring, A.M., Manglik, A., Hu, J., Hu, K., Eitel, K., Hübner, H., Pardon, E., Valant, C., and Sexton, P.M. (2013). Activation and allosteric modulation of a muscarinic acetylcholine receptor. Nature *504*, 101-106.

Kumari, P., Ghosh, E., and Shukla, A.K. (2015). Emerging approaches to GPCR ligand screening for drug discovery. Trends in Molecular Medicine *21*, 687-701.

Lange, A.B. (1990). The presence of proctolin in the reproductive system of Rhodnius prolixus. Journal of insect physiology *36*, 345-351.

Lange, A.B., Orchard, I., and Barrett, F.M. (1988). The presence and distribution of proctolin in the blood-feeding bug, Rhodnius prolixus. Journal of insect physiology *34*, 379-386.

Latorraca, N.R., Venkatakrishnan, A., and Dror, R.O. (2017). GPCR dynamics: structures in motion. Chemical reviews *117*, 139-155.

Lee, Y., Lazim, R., Macalino, S.J.Y., and Choi, S. (2019). Importance of protein dynamics in the structure-based drug discovery of class AG protein-coupled receptors (GPCRs). Current opinion in structural biology *55*, 147-153.

Lefkowitz, R.J. (2004). Historical review: a brief history and personal retrospective of seven-transmembrane receptors. Trends in pharmacological sciences *25*, 413-422.

Li, Y., Yin, C., Liu, P., Li, D., and Lin, J. (2017). Identification of a different agonist-binding site and activation mechanism of the human P2Y1 receptor. Scientific reports *7*, 1-10.

Lipiński, P.F., Jarończyk, M., Dobrowolski, J.C., and Sadlej, J. (2019). Molecular dynamics of fentanyl bound to μ-opioid receptor. Journal of molecular modeling *25*, 1-17.

Lou, X., Toresson, G., Benod, C., Suh, J.H., Philips, K.J., Webb, P., and Gustafsson, J.-A. (2014). Structure of the retinoid X receptor α–liver X receptor β (RXRα–LXRβ) heterodimer on DNA. Nature structural & molecular biology *21*, 277-281.

Lu, C., Wu, C., Ghoreishi, D., Chen, W., Wang, L., Damm, W., Ross, G.A., Dahlgren, M.K., Russell, E., and Von Bargen, C.D. (2021). OPLS4: Improving force field accuracy on challenging regimes of chemical space. Journal of Chemical Theory and Computation *17*, 4291-4300.

Lyman, E., Higgs, C., Kim, B., Lupyan, D., Shelley, J.C., Farid, R., and Voth, G.A. (2009). A role for a specific cholesterol interaction in stabilizing the Apo configuration of the human A2A adenosine receptor. Structure *17*, 1660-1668.

Maggi, M.D., Ruffinengo, S.R., Damiani, N., Sardella, N.H., and Eguaras, M.J. (2009). First detection of Varroa destructor resistance to coumaphos in Argentina. Experimental and Applied Acarology *47*, 317-320.

Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. Journal of chemical theory and computation *11*, 3696-3713.

Manglik, A., Kruse, A.C., Kobilka, T.S., Thian, F.S., Mathiesen, J.M., Sunahara, R.K., Pardo, L., Weis, W.I., Kobilka, B.K., and Granier, S. (2012). Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. Nature *485*, 321-326.

Manglik, A., Lin, H., Aryal, D.K., McCorvy, J.D., Dengler, D., Corder, G., Levit, A., Kling, R.C., Bernat, V., and Hübner, H. (2016). Structure-based discovery of opioid analgesics with reduced side effects. Nature *537*, 185-190.

Mariani, V., Biasini, M., Barbato, A., and Schwede, T. (2013). lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics *29*, 2722-2728.

Marino, K.A., and Filizola, M. (2018). Investigating small-molecule ligand binding to G protein-coupled receptors with biased or unbiased molecular dynamics simulations. In Computational Methods for GPCR Drug Discovery, (Springer), pp. 351-364.

Marino, K.A., Shang, Y., and Filizola, M. (2018). Insights into the function of opioid receptors from molecular dynamics simulations of available crystal structures. British journal of pharmacology *175*, 2834-2845.

Marti-Solano, M., Kaczor, A.A., Guixà-González, R., and Selent, J. (2015). Computational strategies to incorporate gpcr complexity in drug design. In Frontiers in Computational Chemistry, (Elsevier), pp. 3-43.

Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., and Lukyanov, S.A. (1999). Fluorescent proteins from nonbioluminescent Anthozoa species. Nature biotechnology *17*, 969-973.

Maximova, T., Moffatt, R., Ma, B., Nussinov, R., and Shehu, A. (2016). Principles and overview of sampling methods for modeling macromolecular structure and dynamics. PLoS computational biology *12*, e1004619.

McCorvy, J.D., Butler, K.V., Kelly, B., Rechsteiner, K., Karpiak, J., Betz, R.M., Kormos, B.L., Shoichet, B.K., Dror, R.O., and Jin, J. (2018). Structure-inspired design of β-arrestin-biased ligands for aminergic GPCRs. Nature chemical biology *14*, 126-134.

McRobb, F.M., Capuano, B., Crosby, I.T., Chalmers, D.K., and Yuriev, E. (2010). Homology modeling and docking evaluation of aminergic G protein-coupled receptors. Journal of chemical information and modeling *50*, 626-637.

Miesenböck, G., De Angelis, D.A., and Rothman, J.E. (1998). Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. Nature *394*, 192-195.

Milani, N. (1995). The resistance of Varroa jacobsoni Oud to pyrethroids: a laboratory assay. Apidologie *26*, 415-429.

Milani, N. (1999). The resistance of Varroa jacobsoni Oud. to acaricides. Apidologie *30*, 229-234.

Milligan, G., and Kostenis, E. (2006). Heterotrimeric G ‐ proteins: a short history. British journal of pharmacology *147*, S46-S55.

Nygaard, R., Zou, Y., Dror, R.O., Mildorf, T.J., Arlow, D.H., Manglik, A., Pan, A.C., Liu, C.W., Fung, J.J., and Bokoch, M.P. (2013). The dynamic process of β2-adrenergic receptor activation. Cell *152*, 532-542.

Østergaard, H., Henriksen, A., Hansen, F.G., and Winther, J.R. (2001). Shedding light on disulfide bond formation: engineering a redox switch in green fluorescent protein. The EMBO journal *20*, 5853-5862.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Trong, I.L., Teller, D.C., Okada, T., and Stenkamp, R.E. (2000). Crystal structure of rhodopsin: AG protein-coupled receptor. science *289*, 739-745.

Pannifer, A.D., Flint, A.J., Tonks, N.K., and Barford, D. (1998). Visualization of the cysteinyl-phosphate intermediate of a protein-tyrosine phosphatase by X-ray crystallography. Journal of Biological Chemistry *273*, 10454-10462.

Pasternak, G.W., and Pan, Y.-X. (2013). Mu opioids and their receptors: evolution of a concept. Pharmacological reviews *65*, 1257-1317.

Paul, S.M., Mytelka, D.S., Dunwiddie, C.T., Persinger, C.C., Munos, B.H., Lindborg, S.R., and Schacht, A.L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nature reviews Drug discovery *9*, 203-214.

Petersen, H.G. (1995). Accuracy and efficiency of the particle mesh Ewald method. The Journal of chemical physics *103*, 3668-3679.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Science *30*, 70-82.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. Journal of computational chemistry *26*, 1781-1802.

Pil, J., and Tytgat, J. (2003). Serine 329 of the μ-opioid receptor interacts differently with agonists. Journal of Pharmacology and Experimental Therapeutics *304*, 924-930.

Prescott, M., Ling, M., Beddoe, T., Oakley, A.J., Dove, S., Hoegh-Guldberg, O., Devenish, R.J., and Rossjohn, J. (2003). The 2.2 Å crystal structure of a pocilloporin pigment reveals a nonplanar chromophore conformation. Structure *11*, 275-284.

Raehal, K.M., and Bohn, L.M. (2005). Mu opioid receptor regulation and opiate responsiveness. The AAPS journal *7*, E587-E591.

Rajagopal, S., Rajagopal, K., and Lefkowitz, R.J. (2010). Teaching old receptors new tricks: biasing seven-transmembrane receptors. Nature reviews Drug discovery *9*, 373-386.

Remington, S.J. (2006). Fluorescent proteins: maturation, photochemistry and photophysics. Current opinion in structural biology *16*, 714-721.

Ring, A.M., Manglik, A., Kruse, A.C., Enos, M.D., Weis, W.I., Garcia, K.C., and Kobilka, B.K. (2013). Adrenaline-activated structure of β2-adrenoceptor stabilized by an engineered nanobody. Nature *502*, 575-579.

Roe, D.R., and Cheatham III, T.E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. Journal of chemical theory and computation *9*, 3084-3095.

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

Roth, B.L., and Kroeze, W.K. (2015). Integrated approaches for genome-wide interrogation of the druggable non-olfactory G protein-coupled receptor superfamily. Journal of Biological Chemistry *290*, 19471-19477.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols *5*, 725-738. 10.1038/nprot.2010.5.

Sabbadin, D., and Moro, S. (2014). Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR–ligand recognition pathway in a nanosecond time scale. Journal of chemical information and modeling *54*, 372-376.

Sastry, G.M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. Journal of computer-aided molecular design *27*, 221-234.

Scapin, G., Dandey, V.P., Zhang, Z., Prosise, W., Hruza, A., Kelly, T., Mayhood, T., Strickland, C., Potter, C.S., and Carragher, B. (2018). Structure of the insulin receptor–insulin complex by single-particle cryo-EM analysis. Nature *556*, 122-125.

Scheerer, P., and Sommer, M.E. (2017). Structural mechanism of arrestin activation. Current opinion in structural biology *45*, 160-169.

Schiaffino, M.V., d'Addio, M., Alloni, A., Baschirotto, C., Valetti, C., Cortese, K., Puri, C., Bassi, M.T., Colla, C., and De Luca, M. (1999). Ocular albinism: evidence for a defect in an intracellular signal transduction system. Nature genetics *23*, 108-112.

Schmid, C.L., Kennedy, N.M., Ross, N.C., Lovell, K.M., Yue, Z., Morgenweck, J., Cameron, M.D., Bannister, T.D., and Bohn, L.M. (2017). Bias factor and therapeutic window correlate to predict safer opioid analgesics. Cell *171*, 1165-1175. e1113.

Schorb, W., Conrad, K.M., Singer, H.A., Dostal, D.E., and Baker, K.M. (1995). Angiotensin II is a potent stimulator of MAP-kinase activity in neonatal rat cardiac fibroblasts. Journal of molecular and cellular cardiology *27*, 1151-1160.

Shaner, N.C., Patterson, G.H., and Davidson, M.W. (2007). Advances in fluorescent protein technology. Journal of cell science *120*, 4247-4260.

Shang, Y., LeRouzic, V., Schneider, S., Bisignano, P., Pasternak, G.W., and Filizola, M. (2014). Mechanistic insights into the allosteric modulation of opioid receptors by sodium ions. Biochemistry *53*, 5140-5149.

Sherman, W., Day, T., Jacobson, M.P., Friesner, R.A., and Farid, R. (2006). Novel procedure for modeling ligand/receptor induced fit effects. Journal of medicinal chemistry *49*, 534-553.

Shoichet, B.K., and Kobilka, B.K. (2012). Structure-based drug screening for G-protein-coupled receptors. Trends in pharmacological sciences *33*, 268-272.

Shukla, D., Hernández, C.X., Weber, J.K., and Pande, V.S. (2015). Markov state models provide insights into dynamic modulation of protein function. Accounts of chemical research *48*, 414-422.

Sinnokrot, M.O., Valeev, E.F., and Sherrill, C.D. (2002). Estimates of the ab initio limit for π− π interactions: The benzene dimer. Journal of the American Chemical Society *124*, 10887-10893.

Smith, J.S., and Rajagopal, S. (2016). The β-arrestins: multifunctional regulators of G protein-coupled receptors. Journal of Biological Chemistry *291*, 8969-8977.

Soergel, D.G., Subach, R.A., Burnham, N., Lark, M.W., James, I.E., Sadler, B.M., Skobieranda, F., Violin, J.D., and Webster, L.R. (2014). Biased agonism of the μ-opioid receptor by TRV130 increases analgesia and reduces on-target adverse effects versus morphine: a randomized, double-blind, placebo-controlled, crossover study in healthy volunteers. PAIN® *155*, 1829-1835.

Soubias, O., Teague Jr, W.E., Hines, K.G., Mitchell, D.C., and Gawrisch, K. (2010). Contribution of membrane elastic energy to rhodopsin function. Biophysical journal *99*, 817-824.

Spittaels, K., Vankeerberghen, A., Torrekens, S., Devreese, B., Grauwels, L., Van Leuven, F., Hunt, D., Shabanowitz, J., Schoofs, L., and Van Beeumen, J. (1995). Isolation of Ala1-proctolin, the first natural analogue of proctolin, from the brain of the Colorado potato beetle. Molecular and cellular endocrinology *110*, 119-124.

Stansfeld, P.J., and Sansom, M.S. (2011). Molecular simulation approaches to membrane proteins. Structure *19*, 1562-1572.

Starratt, A.N., and Brown, B.E. (1975). Structure of the pentapeptide proctolin, a proposed neurotransmitter in insects. Life sciences *17*, 1253-1256.

Stauffer, F., Cowan-Jacob, S.W., Scheufler, C., and Furet, P. (2016). Identification of a 5-[3-phenyl-(2-cyclic-ether)-methylether]-4-aminopyrrolo [2, 3-d] pyrimidine series of IGF-1R inhibitors. Bioorganic & medicinal chemistry letters *26*, 2065-2067.

Stein, C. (2016). Opioid receptors. Annual review of medicine *67*, 433-451.

Sutcliffe, K.J., Henderson, G., Kelly, E., and Sessions, R.B. (2017). Drug binding poses relate structure with efficacy in the μ opioid receptor. Journal of molecular biology *429*, 1840-1851.

Thal, D.M., Sun, B., Feng, D., Nawaratne, V., Leach, K., Felder, C.C., Bures, M.G., Evans, D.A., Weis, W.I., and Bachhawat, P. (2016). Crystal structures of the M1 and M4 muscarinic acetylcholine receptors. Nature *531*, 335-340.

Tsien, R.Y. (1998). The green fluorescent protein. Annual review of biochemistry *67*, 509-544.

Uchikawa, E., Choi, E., Shang, G., Yu, H., and Bai, X.-c. (2019). Activation mechanism of the insulin receptor revealed by cryo-EM structure of the fully liganded receptor–ligand complex. Elife *8*, e48630.

Ulens, C., Van Boven, M., Daenens, P., and Tytgat, J. (2000). Interaction of p-fluorofentanyl on cloned human opioid receptors and exploration of the role of Trp-318 and His-319 in μ-opioid receptor selectivity. Journal of Pharmacology and Experimental Therapeutics *294*, 1024-1033.

Vaidehi, N., and Bhattacharya, S. (2016). Allosteric communication pipelines in G-protein-coupled receptors. Current opinion in pharmacology *30*, 76-83.

Vaidehi, N., Bhattacharya, S., and Larsen, A.B. (2014). Structure and dynamics of G-protein coupled receptors. G Protein-Coupled Receptors-Modeling and Simulation, 37-54.

Vaidehi, N., and Kenakin, T. (2010). The role of conformational ensembles of seven transmembrane receptors in functional selectivity. Current opinion in pharmacology *10*, 775-781.

Valsson, O., Tiwary, P., and Parrinello, M. (2016). Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. Annual review of physical chemistry *67*, 159-184.

Vardanyan, R.S., and Hruby, V.J. (2014). Fentanyl-related compounds and derivatives: current status and future prospects for pharmaceutical applications. Future medicinal chemistry *6*, 385-412.

Veenstra, J.A., Rombauts, S., and Grbić, M. (2012). In silico cloning of genes encoding neuropeptides, neurohormones and their putative G-protein coupled receptors in a spider mite. Insect biochemistry and molecular biology *42*, 277-295.

Violin, J.D., Crombie, A.L., Soergel, D.G., and Lark, M.W. (2014). Biased ligands at G-protein-coupled receptors: promise and progress. Trends in pharmacological sciences *35*, 308-316.

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

Wacker, D., Stevens, R.C., and Roth, B.L. (2017a). How ligands illuminate GPCR molecular pharmacology. Cell *170*, 414-427.

Wacker, D., Wang, S., McCorvy, J.D., Betz, R.M., Venkatakrishnan, A., Levit, A., Lansu, K., Schools, Z.L., Che, T., and Nichols, D.E. (2017b). Crystal structure of an LSD-bound human serotonin receptor. Cell *168*, 377-389. e312.

Wang, H., Xie, J., Shreeve, T.G., Ma, J., Pallett, D.W., King, L.A., and Possee, R.D. (2013). Sequence recombination and conservation of Varroa destructor virus-1 and deformed wing virus in field collected honey bees (Apis mellifera). PloS one *8*, e74508.

Wang, W., Qiao, Y., and Li, Z. (2018). New insights into modes of GPCR activation. Trends in pharmacological sciences *39*, 367-386.

Weikum, E.R., Liu, X., and Ortlund, E.A. (2018). The nuclear receptor superfamily: A structural perspective. Protein Science *27*, 1876-1892.

Weis, F., Menting, J.G., Margetts, M.B., Chan, S.J., Xu, Y., Tennagels, N., Wohlfart, P., Langer, T., Müller, C.W., and Dreyer, M.K. (2018). The signalling conformation of the insulin receptor ectodomain. Nature communications *9*, 1-10.

Wiedenmann, J., Schenk, A., Röcker, C., Girod, A., Spindler, K.-D., and Nienhaus, G.U. (2002). A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from Entacmaea quadricolor (Anthozoa, Actinaria). Proceedings of the National Academy of Sciences *99*, 11646-11651.

Wootten, D., Christopoulos, A., and Sexton, P.M. (2013). Emerging paradigms in GPCR allostery: implications for drug discovery. Nature reviews Drug discovery *12*, 630-644.

Wu, E.L., Cheng, X., Jo, S., Rui, H., Song, K.C., Dávila‐Contreras, E.M., Qi, Y., Lee, J., Monje‐Galvan, V., and Venable, R.M. (2014). CHARMM‐GUI membrane builder toward realistic biological membrane simulations. Wiley Online Library.

Xiao, X., Zeng, X., Yuan, Y., Gao, N., Guo, Y., Pu, X., and Li, M. (2015). Understanding the conformation transition in the activation pathway of β2 adrenergic receptor via a targeted molecular dynamics simulation. Physical Chemistry Chemical Physics *17*, 2512-2522.

Xu, J., Wang, Z., Liu, P., Li, D., and Lin, J. (2015). An insight into antagonist binding and induced conformational dynamics of class B GPCR corticotropin-releasing factor receptor 1. Molecular BioSystems *11*, 2042-2050.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015a). The I-TASSER Suite: protein structure and function prediction. Nature methods *12*, 7-8.

Yang, J., and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. Nucleic acids research *43*, W174-W181.

Yang, L., Yang, D., De Graaf, C., Moeller, A., West, G.M., Dharmarajan, V., Wang, C., Siu, F.Y., Song, G., and Reedtz-Runge, S. (2015b). Conformational states of the full-length glucagon receptor. Nature communications *6*, 1-13.

Yao, X.J., Ruiz, G.V., Whorton, M.R., Rasmussen, S.G., DeVree, B.T., Deupi, X., Sunahara, R.K., and Kobilka, B. (2009). The effect of ligand efficacy on the formation and stability of a GPCR-G protein complex. Proceedings of the National Academy of Sciences *106*, 9501-9506.

Yarbrough, D., Wachter, R.M., Kallio, K., Matz, M.V., and Remington, S.J. (2001). Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-Å resolution. Proceedings of the National Academy of Sciences *98*, 462-467.

Zaitseva, E., Brown, M.F., and Vogel, R. (2010). Sequential rearrangement of interhelical networks upon rhodopsin activation in membranes: the Meta IIa conformational substate. Journal of the American Chemical Society *132*, 4815-4821.

Zeng, L., Guan, M., Jin, H., Liu, Z., and Zhang, L. (2015). Integrating Pharmacophore into Membrane Molecular Dynamics Simulations to Improve Homology Modeling of G Protein‐coupled Receptors with Ligand Selectivity: A2A Adenosine Receptor as an Example. Chemical Biology & Drug Design *86*, 1438-1450.

Zhang, J., Yang, J., Jang, R., and Zhang, Y. (2015). GPCR-I-TASSER: a hybrid approach to G protein-coupled receptor structure modeling and the application to the human genome. Structure *23*, 1538-1549.

Zhang, S., and Zhang, Z.-Y. (2007). PTP1B as a drug target: recent developments in PTP1B inhibitor discovery. Drug discovery today *12*, 373-381.

Zhao, L., Zhou, S., and Gustafsson, J.-Å. (2019). Nuclear receptors: recent drug discovery for cancer therapies. Endocrine reviews *40*, 1207-1249.

Zheng, H., Chu, J., Zhang, Y., Loh, H.H., and Law, P.-Y. (2011). Modulating μ-Opioid Receptor Phosphorylation Switches Agonist-dependent Signaling as Reflected in PKCϵ Activation and Dendritic Spine Stability. Journal of Biological Chemistry *286*, 12724-12733.

Zhou, X.E., Melcher, K., and Xu, H.E. (2017). Understanding the GPCR biased signaling through G protein and arrestin complex structures. Current opinion in structural biology *45*, 150-159.

Zou, Y., Ewalt, J., and Ng, H.-L. (2019). Recent insights from molecular dynamics simulations for g protein-coupled receptor drug discovery. International Journal of Molecular Sciences *20*, 4237.

Appendix A - Copyright Permissions

Published work – Chapter 1 ((Zou et al., 2019))

Zou, Y., Ewalt, J., and Ng, H.-L. (2019). Recent insights from molecular dynamics simulations for g protein-coupled receptor drug discovery. International Journal of Molecular Sciences 20, 4237.

https://doi.org/10.3390/ijms20174237

For all articles published in MDPI journals, copyright is retained by the authors. Articles are licensed under an open access Creative Commons CC BY 4.0 license, meaning that anyone may download and read the paper for free.

https://www.mdpi.com/authors/rights