T.C

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL

THE DEPARTMENT OF BIOENGINEERING

IN SILICO MODELING AND MOLECULAR DYNAMICS SIMULATIONS OF A cpGFP-DERIVED FLUORESCENT OXYTOCIN RECEPTOR BIOSENSOR

MASTER'S THESIS

BOUTHEINA ZENDER

ISTANBUL 2024

T.C
BAHCESEHIR UNIVERSITY
GRADUATED SCHOOL
THE DEPARTMENT OF BIOENGINEERING

IN SILICO MODELING AND MOLECULAR DYNAMICS SIMULATIONS OF A cpGFP-DERIVED FLUORESCENT OXYTOCIN RECEPTOR BIOSENSOR

MASTER'S THESIS

THESIS ADVISOR:
Assist. Prof. Dr. ALPER DEVRIM OZKAN

ISTANBUL 2024

THE REPUBLIC OF TURKEY
BAHCESEHIR UNIVERSITY
GRADUATE SCHOOL

19/04/2024

MASTER THESIS APPROVAL FORM

Program Name:	Bioengineering (Thesis, English)
Student's Name and Surname:	BOUTHEINA ZENDER
Name of The Thesis:	*In Silico* Modeling And Molecular Dynamic Simulations Of A cpGFP-Derived Fluorescent Oxytocin Receptor Biosensor
Thesis Defense Date:	April 19th, 2024

This thesis has been approved by the Graduate School, which has fulfilled the necessary conditions as a Master thesis.

Associate Prof.

Institute Director

This thesis was read by us, quality, and content as a Master's thesis have been seen and accepted as sufficient.

	Title/Name	**Signature**
Thesis Advisor's	Asst. Prof. Dr. Alper Devrim ÖZKAN	
Member	Asst. Prof. Dr. Pınar SİYAH	
Member	Asst. Prof. Dr. Ebuzer KALYONCU	

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: **BOUTHEINA ZENDER**

Signature:

ABSTRACT

BOUTHEINA ZENDER

***IN SILICO* MODELING AND MOLECULAR DYNAMICS SIMULATIONS OF A cpGFP-DERIVED FLUORESCENT OXYTOCIN RECEPTOR BIOSENSOR**

Bioengineering Master Program

Thesis Advisor: Asst. Prof. Dr. Alper Devrim OZKAN

April 2024, 55 Pages

Oxytocin receptor (OXTR) is a G protein-coupled receptor (GPCR) that interconnects with the oxytocin hormone and heterometric G proteins to trigger downstream signaling cascades. This complex has been associated with a variety of physiological and behavioral processes and attracted much interest as a potential drug target. This *in silico* study aims to produce a structural-dynamic profile of a genetically encoded fluorescent reporter (GEFR), tentatively named *iXsov*, to enable real-time assessment of OXTR activity in response to fluctuating oxytocin levels. The design of the optical sensor was performed by the integration of cyclic permuted green fluorescent protein (cpGFP) within the third intracellular loop of OXTR. A series of molecular dynamics simulations were performed to determine whether cpGFP insertion interferes with OXTR function in the presence of oxytocin and the $G\alpha_O$ ras-associated domain. The tertiary structure of iXsov was predicted via Alphafold2 and placed in a lipid membrane via CHARMM-GUI. Subsequent molecular dynamics simulations were performed using GROMACS 2023.2 software with CHARMM36m force field and TIP3P water model to determine the stability of iXsov with and without its associated ligand and G protein. Six distinct systems were modeled based on three templates (PDB ID: 6TPK, 7RYC, and 7QVM) representing active and inactive states for both iXsov and WT OXTR. The root means square deviation of backbone atoms and root mean square fluctuation of residues were calculated to highlight comparative changes in the system dynamics of iXsov. Our computational study serves as a proof-of-concept for the feasibility of the biosensor design, informing the planning of *in vitro* studies in the near future. In addition, we emphasize the importance of using customized predicted structures to ensure accurate and reliable MD simulation outcomes. These findings hold promise in informing the future design and optimization of iXsov, as well as advancing research in drug screening applications.

Keywords: GPCR, cpGFP, MD simulation, Alphafold2, GROMACS.

ÖZ

BOUTHEINA ZENDER

cpGFP'den TÜRETİLMİŞ FLORESAN OKSİTOSİN RESEPTÖRÜ BİYOSENSÖRÜNÜN *In-Silico* MOLEKÜLER MODELLEME VE MOLEKÜLER DİNAMİK SİMÜLASYONLARININ ARAŞTIRILMASI

Biyomühendislik Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi Alper Devrim ÖZKAN

Nisan 2024, 55 Sayfa

Oksitosin reseptörü (OXTR), oksitosin hormonu ile etkileşen ve heterometrik G proteinini aktive eden bir G proteini kenetli reseptör (GPCR) olup, aşağı yöndeki sinyalleşme yolağının basamaklarını tetikler. Bu kompleks çeşitli fizyolojik ve davranışsal süreçlere katkıda bulunurken, potansiyel bir ilaç hedefi olarak büyük ilgi çekmektedir. Bu *in silico* çalışma, dalgalanan oksitosin seviyelerine yanıt olarak OXTR aktivitesinin gerçek zamanlı değerlendirmesini sağlamak amacıyla tasarlanan ve burada geçici olarak iXsov olarak adlandırılan genetik olarak kodlanmış floresan raporlayıcının (GEFR) yapısal dinamik profilini çizmeyi amaçlamaktadır. Optik sensörün tasarımı, döngüsel permütasyonlu Yeşil Floresan Protein (cpGFP)'nin OXTR'nin üçüncü hücre içi döngüsüne entegrasyonuyla gerçekleştirildi. cpGFP eklenmesinin OXTR fonksiyonuna oksitosin ve $G\alpha_O$ proteininin Ras ile ilişkili bölgesinin varlığında müdahale edip etmediğini belirlemek için bir dizi moleküler dinamik simülasyon gerçekleştirildi. iXsov'un üçüncül yapısı AlphaFold2 aracılığıyla tahmin edildi ve CHARMM-GUI aracılığıyla bir lipid membrana yerleştirildi. Ardışık Moleküler Dinamik simülasyonlar, ligand ve G protein iXsov'a bağlıyken ve bağlı değilken iXsov'un stabilitesini belirlemek için CHARMM36m Kuvvet Alanı ve TIP3P su modeli ile GROMACS 2023.2 yazılımı kullanılarak gerçekleştirildi. Aktif ve inaktif durumlar için, cpGFP'li ve cpGFP'siz olarak üç şablon (PDB Kimliği: 6TPK, 7RYC ve 7QVM) temel alınarak altı farklı sistem modellenmiştir. iXsov'un sistem dinamiklerindeki karşılaştırmalı değişiklikleri vurgulamak için omurga atomlarının karekök ortalamalı sapması ve kalıntılarının karekök ortalamalı dalgalanması hesaplandı. Hesaplamalı çalışmamız, biyosensör tasarımının fizibilitesine yönelik bir kavram kanıtı olarak hizmet etmekte ve yakın gelecekte *in* vitro çalışmaların

planlanması konusunda bilgi vermektedir. Bu bulgular, özellikle ilaç tarama uygulamalarında araştırma deneylerinin ilerletilmesinde umut vaat ediyor.

Anahtar kelimeler: GPCR, cpGFP, MD simülasyonu, Alphafold2, GROMACS.

ACKNOWLEDGEMENT

I dedicate this to my family and friends.

I acknowledge my advisors, Dr. Alper Devrim Ozkan and Dr. Ahmet Emin Topal who continuously guided me in my Master's study. I want to thank them for their professional attitude and the theoretical skills that I have learned from them during this research. I also want to dedicate a warm thanks to Dr. Ismail Erol for the computational support, for his patience, and for sharing his knowledge and codes. I am grateful to interconnect and network with the pharmacy faculty members including Dr. Pınar Siyah for her help and psychological support.

I am forever grateful to my family member my father, my mother, and my sisters who have always supported me and believed in my capacities. No matter how far they are I can always feel their love and support. I thank my friends and roommates for their patience and psychological support.

Last but not least, I thank myself for the courage, love, support, care, and patience that I had during this challenge. I am extremely proud of myself.

With love

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AI	Artificial intelligence
OXTR	Oxytocin receptor
iXsov	Indicator of OXTR signaling by Oxytocin variation
OT	Oxytocin hormone
MD	Molecular dynamics
cpGFP	circularly permuted green fluorescent protein
AF2	AlphaFold2
GPCR	G-protein coupled receptor
ASD	Autism spectrum disorder
ICL3	Third intracellular loop
RMSD	Root means square deviation
RMSF	Root means square fluctuation
GEFR	Genetically encoded fluorescent reporter
MSA	Multiple sequence alignment
VMD	Visual Molecular Dynamics
$G\alpha_O$	Go protein alpha subunit
SNc	Substantia nigra pars compacta
PDB	Protein Data Bank

Chapter 1
Introduction

1.1 Theoretical framework

A range of psychiatric and psychological phenotypes are associated with GPCRs signaling dysfunction, such as schizophrenia, major depressive disorder, bipolar disorder, and autism spectrum disorders (ASDs) (Rokicki et al., 2022). GPCRs, or G protein-coupled receptors, are hepta-helical transmembrane (7TM) proteins and represent approximately 3% of the human genome. They facilitate vital cell signaling processes by relaying intra- and extracellular chemical signals through interactions with heterotrimeric G protein complexes. GPCR plays a role in mediating the effects of over 50% of drugs available on the market (Sanchez-Reyes et al., 2023). Oxytocin receptor (OXTR) belongs to the Class A of GPCRs and has been suggested as a potential drug target (Moerkerke et al., 2021).

GPCRs undergo structural transformations resulting from the binding of extracellular ligands to facilitate the activation of heterotrimeric G proteins, which in turn play a crucial step in many intracellular signaling cascades (Sun et al., n.d.). Oxytocin (OT) is a neuropeptide hormone that regulates social behavior through binding to the OXTR receptors found throughout the brain and peripheral nervous system. However, the exact relationship between oxytocin levels, OXTR, and the neurobiological mechanism of the OXTR gene expression pattern remains unknown, representing a significant gap in our knowledge (Rokicki et al., 2022). The complex neurobiology of the oxytocin signaling system in various brain regions has been investigated in the past decad, using advanced neuro-technologies such as optogenetics, chemogenetics, and calcium imaging. However, these approaches pose their own challenges and are not commercially available for drug screening efforts. As such, the lack of commonly available, high-throughput, and easy-to-implement methods for the detection of OXTR activation is a current challenge for drug development. To address similar limitations in other GPCRs, recent researchers have engineered various genetic tools to visualize and manipulate GPCR-expressing cells

in vitro and in mouse brains. These tools include engineered fluorescent reporters, which have emerged as a successful strategy to visualize GPCR protein expression by themselves and in combination with real-time super-resolution fluorescence imaging (Inoue et al., 2022). While these GPCR biosensors have been explored experimentally, the structural dynamics of the complex remain unknown at the atomic level. Molecular dynamic simulations (MD) are routinely used to bridge this gap in knowledge by providing interaction data at resolutions not available in experiments.

Here, our *in silico* approach aims to observe the intra- and intermolecular dynamics in active and inactive OXTR systems in the presence and absence of the cyclically permuted green fluorescent protein (cpGFP) motif. These observations will inform the final design and cloning of a new fluorescent reporter that we tentatively call iXsov, by integrating the cpGFP in the third intracellular loop of the oxytocin receptor. We modeled six models based on three PDB templates (ID : 6TPK, 7RYC, 7QVM) with the updated version AlphaFold2 (AF2), then built the lipid membrane using CHARMM-GUI. Molecular dynamics simulations were then performed with GROMACS 2023.2 to determine the potential impact of cpGFP insertion on OXTR function and downstream signaling in various biological contexts (Lin & Hsu, 2018).

1.2 Aim

This study aims to produce a structural dynamics profile of a genetically encoded fluorescent reporter. The MD simulation study is performed to highlight the changes that occur in the structure of inactive and active states of OXTR coupled to cpGFP (iXsov). Five key steps have been performed to this end:

1. ***In silico* design of the fluorescent reporter's sequence**. This involves integrating cpGFP into the third intracellular loop (ICL3) of OXTR
2. **Homology modeling of the 3D structures using Alphafold2.** We generated six models using PDB customized modeling**: Inactive OXTR | Inactive iXsov | OXTR-OT-Galpha protein (AF2-7RYC & AF2-7QVM) | iXsov-OT-Galpha protein (AF2-7RYC & AF2-7QVM).**
3. **Building the bilayer lipid membrane using CHARMM-GUI.**
4. **Molecular Dynamic simulation using GROMACS 2023.2.**

5. **Data analysis.** In particular, root mean square deviation (**RMSD**), root mean square fluctuation (**RMSF**), and **MM-PBSA decomposition** analysis.

Chapter 2
Literature Review

2.1 The Oxytocin Hormone (OT)

Oxytocin (OT) is a neuromodulator and a nine-amino acid hormone with the sequence Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-(NH2). Its molecular weight is 10007.2 daltons (Inoue et al., 2022). OT contains one active disulfate bridge between the first and sixth cysteine residues, forming a hexapeptide ring and a C-terminal alpha amidated tripeptide residue tail (3, 6), as shown in Figure 1. Oxytocin is produced by hypothalamic neurons in the supraoptic and paraventricular nucleus (PVN) of the brain and released into the blood circulation of the periphery and central nervous system as peptide hormone, in response to stimuli such as stress, aggression, and anxiety (Jurek & Meyer, 2020). Oxytocin hormone has a significant impact in general health, development, adaptation, reproduction, stress, and the immune system (Moerkerke et al., 2021a). Oxytocin acts as an anti-anxiety, anti-inflammatory, and antioxidant molecule, and has protective effects against adversity and trauma. Animal studies indicated that oxytocin perturbation affects various neurotransmitter systems such as GABA, dopamine, and serotonin, leading to long-term consequences on the maturation of the social brain (Friedlander et al., 2019). Moreover, low oxytocin levels is a main factor in several psychiatric diseases. As such, it has been investigated in the treatment of various disorders including autism spectrum disorders, anxiety, post-traumatic stress disorders, and schizophrenia (Sato et al., 2023). These properties led to oxytocin and its receptor becoming potential targets for developing potential therapeutic drugs for many mental and health disorders. Consequently, understanding the complexities of oxytocin signaling is essential for future drug discovery efforts. The oxytocin receptor (OXTR) belongs to G protein-coupled receptor (GPCR) family, and is found in peripheral tissues like the kidney, ovary, testis, heart vascular endothelium, various cancer cells, as well as in the periphery and central nervous system (Antobreh et al., 2017). Like all GPCRs, the oxytocin-OXTR binding interaction generates intracellular signaling pathways, through G-protein subunits (Borroto-Escuela et al., 2022).

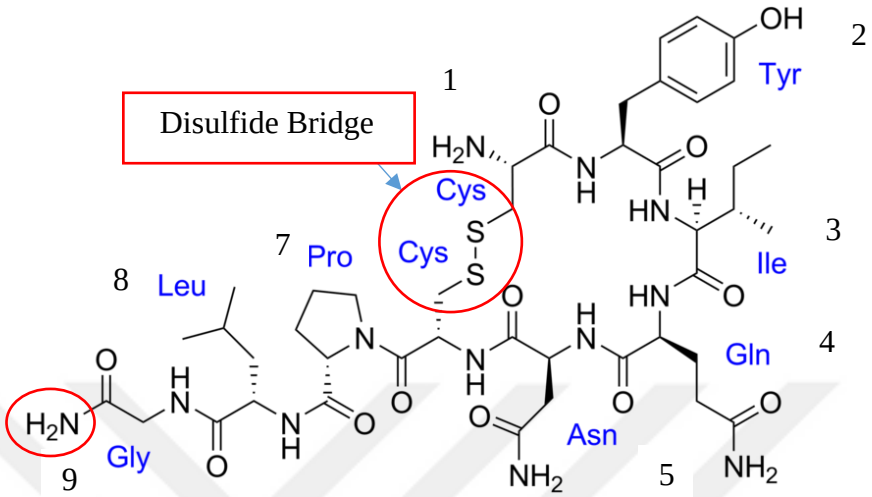


Figure 1. Structure and amino acid sequence of oxytocin.

2.2 G Protein-Coupled Receptors (GPCRs)

G protein-coupled receptors represent the largest family of membrane proteins that modulate physiology across human tissues in response to external stimuli such as hormones and neurotransmitters. GPCRs have a crucial role in the transmission of chemical signals from the extracellular matrix into the internal cell environment through complex signaling networks. Over 800 human GPCRs and 16 different G-alpha genes exist (Flock et al., 2015). The interaction of GPCRs with specific signaling molecules and ligands induces conformational changes leading to the activation of signaling pathways and the generation of a cascade of neural and intracellular events. GPCR family is divided according to their amino acid sequences and structural characteristics into six families: class A (rhodopsin), B (secretin), C (metabotropic glutamate), D (fungus pheromone), E (cAMP) and F (frizzled/smoothened). Class A is the largest and most well-studied class of GPCRs and includes OXTR as well as other receptors for odorants and small ligands (Tuteja, 2009). This class contains typical GPCRs with a seven-helix transmembrane structure (7TM α-helices) flanked by extracellular amino-terminal and intracellular carboxy-

terminal segments. GPCRs vary in length between 311 and 1490 amino acids, giving them remarkable adaptability in spanning the cellular membrane and flexibility in transmitting signals. These properties make GPCRs an attractive target for the development of new therapeutic drugs (Patrone et al., 2020). GPCRs are involved in a wide range of diseases, including diabetes type 2, obesity, depression, cancer, and autism, through both well-studied and more obscure cellular signaling pathways. Cellular responses to type-A GPCRs typically include cyclic adenosine 3,5-monophosphate (cAMP) production, calcium mobilization, and phosphorylation of extracellular regulated protein kinases ½ (pERK1/2) (Yang et al., 2021).

2.3 Trimeric G-proteins

The GPCR signal transmission mechanism is achieved by the interaction of GPCRs with heterotrimeric G-proteins, leading to the activation of downstream GPCR pathways, such as adenylate cyclase (AC), cAMPs, and protein kinases (PKA). These pathways are responsible for the regulation of a wide array of physiological processes such as reproduction, feeding behaviors, stress, growth, and development (N. Liu et al., 2021). Trimeric G proteins are composed of three subunits; Gα, Gβ and Gγ. Once activated, Gα and the Gβ/γ dimer are independently able to convert external stimuli to various signaling pathways (Tuteja, 2009). They are classified based on their downstream signal and sequence similarities, including at least 16 Gα subunits, 5 Gβ subunits, and 12 Gγ subunits (Escribá et al., 2008). The Gα (G-alpha) subunit is recognized as the primary driver of GPCR signaling. The 16 different Gα subunits found in the human genome are classified into four subclasses: Gs, Gi/o, Gq/11, and G12/13 (Okashah et al., 2019). $G\alpha_o$ proteins are highly abundant and mainly expressed in the brain, they have crucial role in linking receptors to potassium and calcium channels in neurons, they are also involved in the membrane trafficking processes, which maintain cellular communication and functionality within cells (Greif et al., 2000). The G protein subclasses initiate different cellular responses in a process called G protein engagement (Kise et al., 2021). The association of GPCRs with ligands creates an allosteric communication that triggers changes in the conformation of the Gα subunit leading to the exchange of guanosine diphosphate (GDP) to guanosine triphosphate (GTP), facilitating transition to an “active” mode characterized by the separation of the GTP-bound Gα subunit and the Gβγ complex from each other and

from their receptor (Figure 2). This process enables the activation of the downstream cellular effectors, which could be stimulatory or inhibitory depending on the type of Gα protein (Maruta et al., 2021).

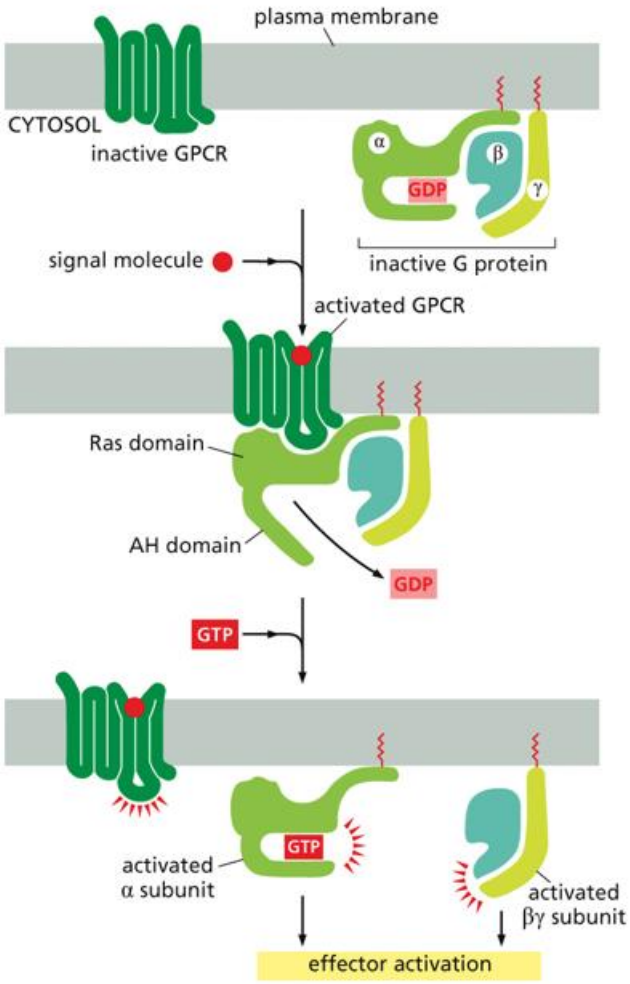


Figure 2. Dynamic interactions in heterotrimeric G protein signaling, ligand-induced activation, and the GTP hydrolysis cascade. Replicated with permission from (Maruta et al., 2021).

2.4 Oxytocin Receptor (OXTR)

Oxytocin receptor (OXTR) is class A/1 GPCR (Jacob et al., 2007). OXTR is found abundantly in the brain with crucial functions related to brain activity and signaling (Meyer et al., 2022). It is present both presynaptically and postsynaptically in excitatory and inhibitory synapses, influencing the modulation of synaptic transmission at both levels (Borroto-Escuela et al., 2022). OXTR interacts with

different G-proteins leading to multiple intracellular signaling cascades and ionic mechanisms in neurons. These include calcium ion (Ca^{2+}), protein kinase C, mitrogen-activated protein kinase (MAPK) and MAPK/ERK kinase (MEK1/2) signaling pathways (Meyer et al., 2023). The oxytocin receptor gene, known also as *OXTR* (italicized per convention) is responsible for producing the oxytocin receptor protein (Inoue et al., 2022). It is located on chromosome 3 3p25-3p26.2 in the human genome. The gene is composed of 3 introns and 4 exons. Exons 3 and 4 encode the amino acids that comprise the oxytocin receptor protein. Exons 1 and 2 are non-coding regions, they play important regulatory roles in gene expression and function such as initiating transcription and regulate gene activity. The OXTR gene is approximately 17 kilobases (kb) long (Moerkerke et al., 2021).

2.4.1 Structure of OXTR

The OXTR protein consists of 388 amino acids and has 7 highly conserved hydrophobic transmembrane alpha helices (7TM) spanning the cell membrane, as illustrated in (Figure 3). The 7TM is connected by intra- and extracellular loops. The three intracellular loops (ICL1, ICL2, ICL3) and the cytoplasmic C-terminal domain are involved in G-protein attachment, initiating intracellular and downstream signaling after ligand binding (Antobreh et al., 2017). In contrast, the three extracellular loops (ECL1, ECL2, ECL3), and the extracellular N-terminal domain are responsible for recognizing and binding the ligand oxytocin hormone. The loops are flexible and play crucial roles in OXTR receptor activation and signaling. These 7 transmembrane domains can interact with oxytocin outside the cell and transmit signals inside the cell to initiate cellular responses. Experimentally identifying OXTR expression levels and understanding the OXTR function provide insights into ligand-receptor interactions, affirming OXTR's crucial role in regulating social behavior. In addition, it allows researchers to realize further studies on various social and physiological disease processes and drug development (Moerkerke et al., 2021).

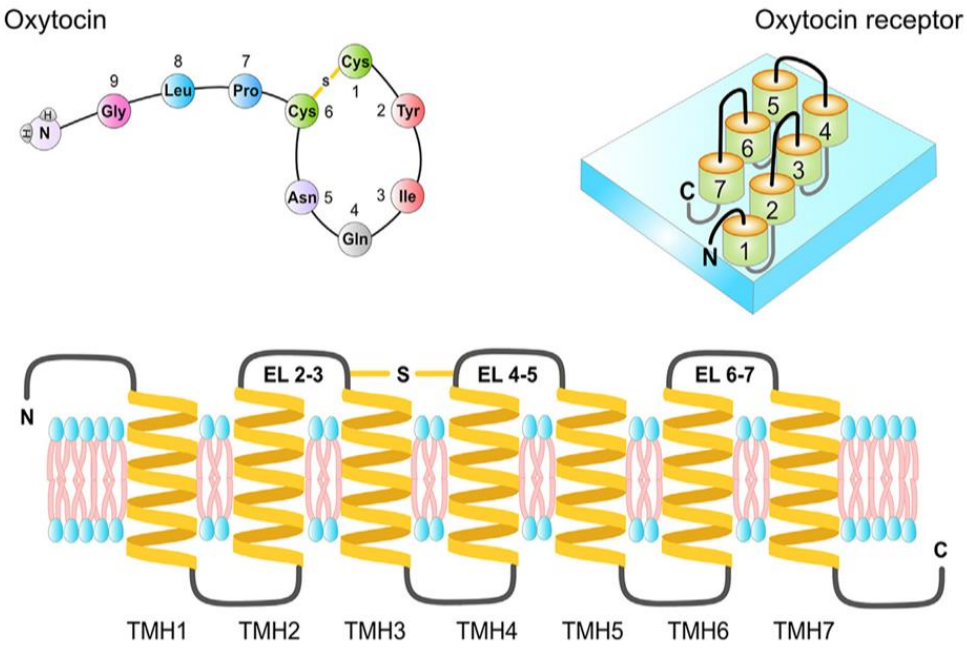


Figure 3. Illustration of the structure of the oxytocin receptor within the cell membrane and the oxytocin hormone. Replicated with permission from (Uvnäs-Moberg, 2024).

2.5 OXTR Involvement In Autism Spectrum Disorders: A Case Study

Autism spectrum disorders (ASDs) are a group of prevalent developmental disorders that affect up to 1 in 55 children globally. ASDs are characterized by a triad of abnormalities including delayed learning, a lack of communicative abilities, and acute anxiety. Despite decades of research, the specific origins and underlying mechanisms of ASD are yet to be elucidated, but it is believed to be a complex condition resulting from a combination of various factors. Finding a treatment to manage or reduce the social symptoms of ASD has been a goal for numerous researchers looking to increase the quality of life for individuals with ASD over the past few decades. Although the prevalence of autism differs between sexes, previous studies suggest that both females and males with autism have significantly lower levels of OXTR density in the substantia nigra pars compacta (SNc) compared to unaffected controls (Frehner et al., 2022). A recent study by Horie et al. concluded that OXTR mutations had an impact on behavior related to repetitive actions

(obsessive-compulsive disorder) when induced in praire voles by CRISPR/Cas9 gene editing. These mutants also showed reduced preference for new social stimuli or individuals (Horie et al., 2019). Similar contributions have been made by Leonzino et al. to explore the effects of oxytocin receptors genetically and pharmacologically inactivated in male mice. These mutations resulted in alterations to the responses of affected mice to the novel object recognition test. However, learning performances and memory have not been disrupted by OXTR inactivation (Leonzino et al., 2019). Some authors also have investigated oxytocin concentration as an ASD biomarker, and results demonstrate a relationship between genetic variations in the OXTR gene and increased levels of OXT in ASD patients (Al-Ali et al., 2022). Moreover, DNA methylation levels in OXTR have been implicated in post-trauma, anxiety, and social factor disorders. Moerkerke et al. conclude that hypermethylation of OXTR may inhibit sensitivity to social cues and contribute to the extreme social characteristics seen in individuals with ASD. On the other hand, hypomethylated OXTR may enhance sensitivity to social cues (oversensitivity) leading to social anxiety disorder (SAD) (Moerkerke et al., 2021).

2.6 Genetically Engineered Fluorescent Reporters (GEFR)

Studying neuropeptide functions *in vivo* presents a significant challenge, because of the lack of suitable tools. In previous studies, neuromodulator actions were observed indirectly using techniques such as Ca^{2+} imaging and electrophysiological recording, which are able to detect activity changes at the level of a single cell or population without identifying the exact origins. While fast-scan cyclic voltammetry (FSCV) can measure neuromodulator release levels at the nanomolar scale and in real-time with rapid temporal resolution, it fails in its capability to detect neuropeptides such as oxytocin or vasopressin. To overcome these challenges, researchers have developed multiple genetically encoded neuromodulator sensors designed for detecting neuromodulator expression and activation (Mignocchi et al., 2020). These genetically encoded sensors are based on fluorescent proteins that enable precise access to specific cell types, such as dopaminergic and oxytocinergic neurons, through a combination of genetic targeting and fluorescence microscopy techniques (Yun et al., 2022). Mignocchi et al. developed an optogenetic gene expression system named OXTR-iTango2 that can selectively detect oxytocin-sensitive neurons in animals

when exposed to blue light. This toolset has a crucial role in investigating oxytocin's effects during specific behaviors with great precision in terms of timing (Mignocchi et al., 2020). Jonathan et al. have also successfully engineered an intensity-based GABA-sensing fluorescent reporter named iGABASnFR. This sensor has demonstrated a capacity to detect real-time tracking of GABA release by measuring fluorescence enhancement *in vivo* in mouse and zebrafish models (Marvin et al., 2019).

2.6.1 Circularly Permutated Green Fluorescent Protein (cpGFP)

Insertion of circularly permuted green fluorescent protein (cpGFP) within receptor sequences typically results in an attenuated fluorophore that increases in fluorescence intensity in response to receptor activation. Sensors of this type were first described by Baird et al., and subsequently codon-optimized to improve efficiency (Arnold et al., 2021). The structure of cpGFP is based on the original GFP structure and consists of 11 strands of beta-sheets organized in cylindric arrangement ("beta-barrel" structure). Circular polarization of GFP entailed the creation of new N- and C-termini, with their old counterparts fused together and moved to the middle of the sequence. The chromophore that absorbs and emits light is responsible for the fluorescence of the protein. It is protected and stabilized by the specific arrangement of beta-barrel forms, alpha-helices, and loops within cpGFP. The sequence rearrangement in cpGFP serves to quench the fluorophore in its native configuration, but facilitates higher fluorescence efficiency in response to changes in the secondary structure of adjacent regions. cpGFP can fluoresce when properly folded, enabling the visualization of protein localization and dynamics in living cells. Tagging proteins with cpGFP is particularly important for understanding biological processes and disease mechanisms. (W. Liu et al., 2022). Hsu et al. have successfully developed a novel optical biosensor for measuring membrane tension by introducing cpGFP into the *E. coli* mechanosensitive channel MscL. This alteration in protein's conformation led to variations in fluorescent intensity when exposed to increased membrane tension (Hsu et al., 2023). Arnold et al. have also designed a biosensor that combines phospholamban (PLB), cpGFP and SERCA2a to produce a reporter for PLB-

SERCA2a interactions that play a critical role in heart infarctions, potentially allowing the screening of novel drugs for heart disease (Arnold et al., 2021).

2.6.2 iXsov (Indicator of OXTR Signaling by Oxytocin Variation)

Here, we propose the design of a novel genetically engineered optical reporter called **iXsov,** an acronym for Indicator of OXTR signaling by oxytocin variation. iXsov is designed to optically measure changes in oxytocin concentration by coupling the conformational changes induced by oxytocin binding in the human oxytocin receptor (OXTR) to change in the fluorescence intensity of circularly permuted green fluorescent protein (cpGFP). The initial version of iXsov was created by inserting a cpGFP module from the genetically encoded calcium indicator GCaMP6 (Patriarchi et al., 2018) with a specific linker sequence in the third intracellular loop (ICL3) between two amino acid positions (Alanine 249 & Glycine 250) of the oxytocin receptor (OXTR). The genetically encoded sensor is intended to be expressed in the target neurons to detect changes in oxytocin parameters by producing changes in its fluorescence properties. The changes can be quantified and analyzed for a deeper understanding of the functions and dynamics of these neurons in various physiological and pathological contexts. The genetically encoded oxytocin indicator provides a tool for future researchers to measure oxytocin variation with high sensitivity and specificity. However, as cpGFP insertions are known to interfere with GPCR function and result in inert receptors at times, a proof-of-concept confirming iXsov functionality *in silico* is necessary prior to cloning and *in vitro* experiments. The present thesis describes the process of these *in silico* confirmation efforts.

2.7 AlphaFold2 (DeepMind)

Artificial Intelligence (AI) is applied with increasing frequency to solve fundamental problems in science. DeepMind is an artificial intelligence research lab composed of an interdisciplinary team to work on protein folding. Predicting the 3D structure of proteins from its amino acid sequence is recognized as one of the most challenging problems in biology. AlphaFold2 (AF2) was created to solve this problem, especially for newly engineered and/or poorly investigated proteins. AlphaFold uses two main data references: an array of the number of sequences related to the target

representing the **MSA (Multiple Sequence Alignment)** and an array of numbers representing all pairs of residues in the target protein (**residue pairs pairwise distance plot)**. In addition, it uses **PDB structures as a template** (template mode: Custom) which are expected to be extremely similar across class-A GPCRs (Figure 4). For evaluation metrics, the deep learning algorithm uses evaluation metrics like CASP14 (critical assessment of structure prediction 14) to evaluate AlphaFold's performance. **LDDT** (local distance difference test) is another **evaluation metric** used to assess local correctness of prediction, while **pLDDT** (Per-Residue Local Distance Difference Test) indicates **confidence in prediction** in the relative positions of residues, crucial for interpreting domain positions in multi-domain proteins (Jumper, Evans, Pritzel, Green, Figurnov, Ronneberger, Tunyasuvunakool, Bates, Žídek, Potapenko, Bridgland, Meyer, Kohl, Ballard, Cowie, Romera-Paredes, Nikolov, Jain, Adler, Back, Petersen, Reiman, Clancy, Zielinski, Steinegger, Pacholska, Berghammer, Bodenstein, et al., 2021). Lastly, the **PAE plot** predicts alignment error. AlphaFold2 excels in predicting novel protein structures and complexes in addition to well-established, PDB-supported models based on cryoEM, crystallography and NMR results. It provides guidance for protein construct design, oligomerization predictions (dimers, tetramers etc., using pTM_{max}). Five ranked structures are generated in each run depending on the average of pLDDT or pTM. The limitation of Alphafold is its inability to predict complex proteins and biological models, which include membranes (Sala et al., 2023). AlphaFold2 is a crucial tool for protein engineering and design, enabling innovative approaches across various research domains. In a study by Kreitz et al (2023), AF2 guided the re-modeling structure of a bacterial "syringe" molecule designed for delivering therapeutic proteins into human cells, despite the lack of experimental structural data (Kreitz et al., 2023). Additionally, Wicky et al. (2022) utilized AF2 to engineer symmetric protein assemblies not found in nature (Wicky et al., 2022). Furthermore, Ennist et al (2022) also employed AF2 to design an engineered protein to mimic a compact photosynthetic reaction for capturing solar energy (Ennist et al., 2022). These applications confirm the role of AF2 in guiding and validating diverse newly engineered proteins.

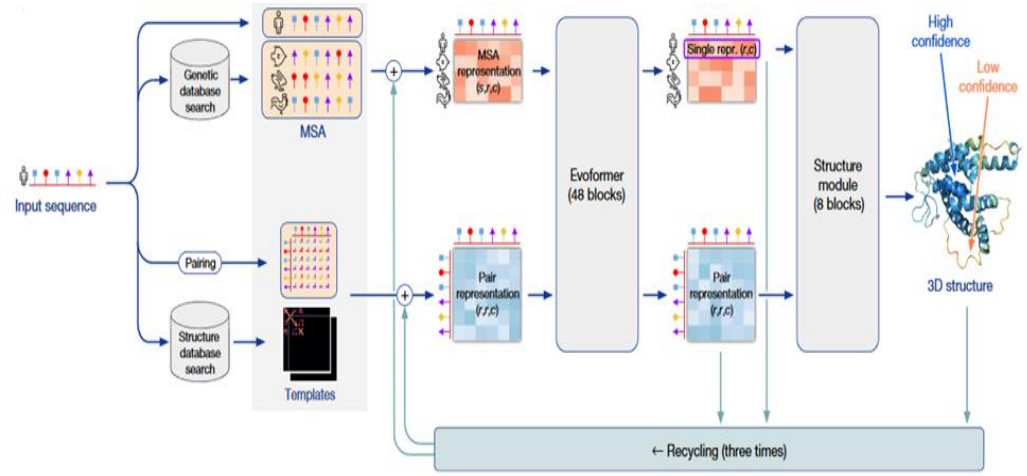


Figure 4. Diagram of AlphaFold2 as published in the official Nature paper in July 2021. Replicated with permission from Jumper et al., 2021

2.8 CHARMM-GUI Membrane Builder

As GPCRs only assume their native configuration when embedded in a lipid bilayer, their simulation is only feasible through the use of a membrane builder. CHARMM-GUI is a web-based graphic user interface for modeling and preparing complex systems and generating files for running simulations using CHARMM, AMBER, Desmond, GROMACS, LAMMPS, NAMD, OpenMM, and TINKER, and covers most frequently used force fields. CHARMM-GUI is used in the simulation of many membrane-related biological processes to ensure the accuracy of the analyzed systems. Building a bilayer membrane in CHARMM-GUI requires five steps. In step 1 the PDB files of the protein structure are uploaded to check specific modifications like the number of chains, disulfide bond, pH, and protonation state. In step 2, the orientation of the protein should be adjusted based on the OPM database or manually oriented based on user input. In step 3 there is a determination of system size and the lipid composition of the membrane. Step 4 is for the generation of the system components, which are lipids, solvents, and ions. In this stage, there is a checkpoint of lipid penetration to the ring structure and the whole protein surface (parameters can be changed if necessary). All components (the protein, membrane, solvent and ions) are assembled in step 5. The system topology is generated in this step and used as input for equilibration (step 6) and simulation production (step 7). The resulting files are used to run subsequent molecular dynamic simulations (Feng et al., 2023).

2.9 Molecular Dynamics Simulations

2.9.1 Definition and History

Molecular dynamics simulations entail the use of a computer to mimic and model the dynamic response of a system, where the interaction is described mathematically in the form of a force field created by Newton's laws using potential energy calculations. MD methods were created for the first time in the mid-1950s. In 1957 Alder and Wainwright performed the first MD simulation using hard-sphere models. Starting from 1970 and with the development of computers, MD simulation has become more and more applicable to different systems such as peptides, proteins, DNA, and macromolecules. The first MD simulation of a folded protein was published in 1977 (Martin & McCammon, 2002). The importance of molecular dynamics simulation in chemistry and biological systems has been greatly highlighted after the 2013 Nobel Prize in Chemistry, which was awarded to Martin Karplus, Michael Levitt, and Ariech Warshel for the development of multiscale models for complex chemical systems (Chmiela et al., 2018). Due to its *in silico* nature, MD simulations have important advantages in terms of speed and cost-effectiveness (but, importantly, not accuracy) over experimental work, allowing computational approaches to serve as an effective initial substrate for later work (Hollingsworth & Dror, 2018).

2.9.2 The Principles of MD Simulations

Molecular dynamics must begin with the initial positions and velocities of each atom in the system. New positions will be computed depending on the timestep, which should be small enough with respect to the rapidity of the motions of system components (Schlick et al., 1997). New velocities and accelerations for atoms will be calculated using Newton's equations of motion. The properties of the system including temperature and pressure should be maintained and kept constant around a specific value in the NPT ensemble according to the Verlet algorithm during each simulation step (Scott Shell., 2016(*COARSE-GRAINING WITH THE RELATIVE ENTROPY - Shell - 2016 - Advances in Chemical Physics - Wiley Online Library*, n.d.). All stages of the process are repeated until equilibrium is reached (the equilibrium state is a stable configuration at a specific energy level). The atomic coordination at equilibrium is saved at regular intervals, forming the trajectory. Analysis of the trajectory, in turn, provides information about the system (Kimber et al., 2011).

2.9.3 CHARMM and Other Force Fields

The force field (FF) is a mathematical model describing parameters required to determine the calculation of atomic interactions. The force field is set depending on the types of molecules present in the research study and consists of three components: the list of atom types, a set of energy of functional forms, and a set of parameters suitable for the experimental data. The forces are a combination of the potential energy terms and functions (Bonded and Non-Bonded interactions) as given in Equation 1.

$$E_{Total} = \underbrace{V_{Bond} + V_{Angle} + V_{Dihedral}}_{E_{Bonded}} + \underbrace{V_{Electrostatic} + V_{Vanderwaals}}_{E_{Non-Bonded}}$$

The origin of the force field comes from experimental sources (X-ray diffraction) and theoretical studies (quantum mechanical (QM) calculations). A literature review is typically performed to decide the FFs suitable for a given study. The most common molecular mechanics force fields in computational chemistry are AMBER (Assisted Model Building with Energy Refinement), GROMOS (GROenigen Molecular Simulation), OPLS (Optimized Parameters for Large-scale Simulation), and CHARMM (Chemistry at Harvard Macromolecular Mechanics).

CHARMM is a computer software developed by Martin Karplus in 1983. It has been successfully used in the realization of molecular dynamics simulation especially for models within membrane lipids (Nagle & Tristram-Nagle, 2000). It includes a wide range of biological databases such as proteins, peptides, small molecules (ligands), lipids, nucleic acids, and carbohydrates. The CHARMM force field has different versions for each type of molecule; for example, CHARMM36 is optimized for proteins, while CHARMM22 and 27 are for lipids, DNA, and RNA (Terteci-Popescu & Beu, 2022).

2.9.4 GROMACS

GROMACS is the most used open-source and free software code in biomolecular dynamic simulations. It provides a wide set of calculation parameters and analysis tools supported by enhanced and optimized algorithms (Abraham et al., 2015). GROMACS supports heterogeneous acceleration (CPU-GPU), and all widely used molecular mechanics force fields are validated and can be used (Kutzner et al., 2022). Simulation using GROMACS has the flexibility to incorporate various geometric constraints (shapes and boundaries) and utilizes both explicit and implicit

solvents. Simulation can represent a system at atomistic or coarse-grained (larger group of atoms) levels of details. The mdrun tool in GROMACS is capable to multirun simulations within single executable file. Moreover, it allows advanced simulation techniques such as replica exchange (exchange of information between multiple simulations to enhance sampling efficiency) (Hsueh et al., 2022) and studying nonequilibrium processes through umbrella sampling as well as free energy calculations (Macchiagodena et al., 2021), exploring thermodynamic properties and molecular motions (Mitsutake et al., 2001). GROMACS is not a Graphical User Interface (GUI), therefore simulation files are plugged directly through VMD (Visual Molecular Dynamics), a popular software packaging used to visualize molecular structures and dynamics (Fernandes et al., 2019).

2.9.5 Protein-Ligand Simulation Using GROMACS

The *in silico* observation of protein-ligand interactions is one of the most widespread applications of all-atom molecular simulations. Especially pertinent for drug discovery efforts is the determination of factors such as ligand binding affinity, complex stability, impact of the binding on conformation changes, and the energetic variation of the system. Moreover, the simulation identifies precisely the key residues involved in the interaction and the binding approaches that may exist. This is a crucial step that comes after docking in pharmaceutical research pipelines to computationally improve the binding affinity and pharmacokinetic properties of drug candidates prior to *in vitro* and *in vivo* work. In our research, we aim to ensure that the engineered fluorescent protein does not interfere with the binding affinity of the oxytocin ligand, which is crucial for the induction of a quantifiable fluorescent signal. To this end, we evaluate the stability and specificity of oxytocin-OXTR interactions using the above approaches, in addition to other parameters such as the free energy binding in the protein-ligand complex (calculated using g_mmpbsa tool). These parameters provide quantitative insights into the system's conformational stability (Kumari et al., 2014).

2.9.6 The Role of Molecular Dynamic Simulations in Study Improvement

It is worth noting that approximately 30% of market drugs target GPCRs. About 350 GPCRs are known targets for treating human diseases, many others with known ligands and functions (such as oxytocin receptors) still await further exploration. Besides an impressive list of experimental discoveries, computational

approaches have provided a trove of information regarding GPCR structure and function, allowing the design and discovery potential therapeutic ligands that target this class. Homology modeling is a cost-efficient alternative to the experimental determination of structures, which still remains a slow and expensive (albeit necessary) process. MD simulation could significantly improve homology modeling approaches. Many activation processes of GPCRs have been investigated at the atomic level using MD simulations. Moreover, they play a crucial role in pinpointing binding sites to enhance the interaction between receptors, ligands, and modulators such as G-proteins (Tautermann et al., 2015). Recent advances in MD simulation techniques have significantly contributed to understanding the signaling mechanisms that occur in GPCRs and their dynamic changes. (Zou et al., 2019). MD simulation is a very powerful tool used to capture structural information that is hard to obtain with current experimental methods (Klepeis et al., 2009). Altogether, *in silico* methods represent promising new opportunities for drug discovery and development, although it is always important to underline the fact that all computational data should be validated through later cycles of *in vitro* and (where applicable) *in vivo* work.

Chapter 3
Methodology

3.1 Compilation of Datasets

The protein sequence of the human oxytocin receptor (**OXYR_HUMAN** oxytocin receptor) was extracted from **UniProt** with access number **P30559**. cpGFP sequence was extracted from **pActP-GcaMP6m** plasmid **(#64714)** from Addgene using the **Benchling** Bioinformatic tool. The **PROSITE Database** of protein families and domains was used to identify the conserved motifs of OXTR to predict protein function and structure relationships. **Protein Data Bank (PDB)** was used to extract possible homologous proteins with experimentally determined 3D structures as templates. The updated version of **AlphaFold2** was used to predict the three-dimensional structure of iXsov from its provisional amino acid sequence **(**https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb). **GROMACS 2023.2** was used to perform MD simulations.

3.2 In Silico Design of iXsov

3.2.1 Amino Acid Sequence and Site of Insertion Choice

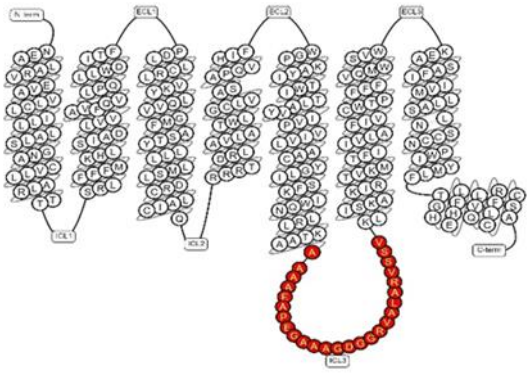
iXsov was designed based on the prior design of a genetically encoded GPCR dopamine sensor called dLight1.2. (Patriarchi et al., 2018). iXsov is created by fusing the cpGFP sequence of 241 residues into a conformationally sensitive site of the third intracellular loop (ICL3) of the human OXTR receptor. The insertion site of cpGFP was determined through the alignment of sequences of OXTR, DRD1, and GPR68 performed using NCBI Basic Local Alignment Search Tool, enabling a precise identification and placement of cpGFP in ICL3 between Alanine 249 and Glycine 250 of OXTR with genetically optimized linker sequences (N-terminal: LSSLI & C-terminal: NHDQL) derived from iGlow sensor of the GPR68 which were previously screened by Patriarchi et al. to enhance the fluorescence response (Ozkan et al., 2021). The final product is a single genetically encoded OXTR-cpGFP construct (iXsov) represented as **N-terminus-OXTR-LSSLI-cpGFP-NHDQL-OXTR-C-terminus** (Figure 5). Its fluorescence changes are assumed to be dependent on OXTR activation,

pending confirmation by *in vitro* studies. To streamline the computational process and reduce time costs in MD simulation, we truncated our structures at residue 30 and 353 for models which do not incorporate the cpGFP structure. However for models integrating cpGFP we truncated at residue 30 and 604. This strategy allowed us to focus on computational analysis of essential structural elements.

A

MEGALAANWSAEAANASAAPPGAEGNRTAGPPRRNEALARVEVAVLCLILLLALSGNACVLLALRTTRQKH
SRLFFFMKHLSIADLVVAVFQVLPQLLWDITFRFYGPDLLCRLVKYLQVVGMFASTYLLLLMSLDRCLAICQPL
RSLRRRTDRLAVLATWLGCLVASAPQVHIFSLREVADGVFDCWAVFIQPWGPKAYITWITLAVYIVPVIVLAA
CYGLISFKIWQNLRLKTAAAAAAEAPEGAAALSSLINVYIKADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNT
PIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPIL
VELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAM
PEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNHDQLGDGGRVALARVSSV
KLISKAKIRTVKMTFIIVLAFIVCWTPFFFVQMWSVWDANAPKEASAFIIVMLLASLNSCCNPWIYMLFTGHL
FHELVQRFLCCSASYLKGRRLGETSASKKSNSSSFVLSHRSSSQRSCSQPSTA

B



C

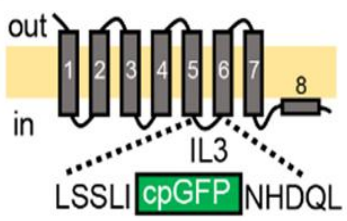


Figure 5. (A) Amino acids sequence of iXsov. OXTR (black); linkers (bleu); cpGFP(green). (B) schematic diagram of the oxytocin receptor highlighting the intracellular loop 3 (ICL3). Image credit: GPCR database. (C) iXsov was designed by inserting cpGFP (green) into the ICL3 of OXTR between helices 5 and 6.

3.3 Modeling Approaches: Alphafold2 3D Structure Prediction.

In this work, we engineered six complex structures representing OXTR complexes through custom prediction options available in the updated version of AlphaFold (AlphaFold2). The PROSITE Database was used to identify conserved motifs within OXTR using other class-A GPCRs as a reference (Figure 6). Hits of PS00237 on PDB 3D structures x-referenced by P30559 are: 6TPK, 7QVM, 7RYC. 6TPK was used as a template in AlphaFold2 to predict the three-dimensional structure of the inactive iXsov using the above-mentioned amino acid sequence. 7QVM and 7RYC were used as a template for predicting active OXTR in the presence of the ligand oxytocin (OT), G-alpha subunit, and cpGFP. Structures obtained from the experimental work and uploaded in PDB (such as inactive OXTR) may still have missing residue and atoms. Therefore, our six models have been re-folded through their amino acid sequences. Re-folding all structures is also a step required to remove any possibility of clashes between atoms during the MD simulation. The models' predictions were generated through their amino acid sequences using a custom template mode and custom Multi-Sequence Alignment (MSA) in unpaired mode which processes separate MSAs for each chain to allow precise prediction of protein structures. As the oxytocin receptor primarily functions as a monomeric protein in its active state, the auto model type was selected for monomer prediction and the number of recycles was set at 20, which enhances the ability to determine structure features not readily evident in initial predictions. The pairing strategy was set to greedy, which pairs any taxonomically matching subsets. All parameters used are listed in (Table 1). The iXsov modeled structure is shown in (Figure 7).

Tool Prosite

P30559 OXYR_HUMAN (389 aa)

Found: 2 hits in 1 sequence

hits by profiles: [1 hit (by 1 profile) on 1 sequence]

```
MEGALAANWSAEAANASAAPPGAEGNRTAGPPRRNEALARVEVAVLCLILLLALSGNACVLLALRT
TRQKHSRLFFFMKHLSIADLVVAVFQVLPQLLWDITFRFYGPDLLCRLVKYLQVVGMFASTYLLLL
MSLDRCLAICQPLRSLRRRTDRLAVLATWLGCLVASAPQVHIFSLREVADGVFDCWAVFIQPWGPK
AYITWITLAVYIVPVIVLAACYGLISFKIWQNLRLKTAAAAAAEAPEGAAAGDGGRVALARVSSVK
LISKAKIRTVKMTFIIVLAFIVCWTPFFFVQMWSVWDANAPKEASAFIIVMLLASLNSCCNPWIYM
LFTGHLFHELVQRFLCCSASYLKGRRLGETSASKKSNSSSFVLSHRSSSQRSCSQPSTA
```

PS50262 G_PROTEIN_RECEP_F1_2 *G-protein coupled receptors family*
1 profile : Hits of PS50262 on PDB 3D structures x-referenced by P30559: 6TPK_A, 7QVM_R, 7RYC_O

hits by patterns: [1 hit (by 1 pattern) on 1 sequence]

PS00237 G_PROTEIN_RECEP_F1_1 *G-protein coupled receptors family 1 signature* : Hits of PS00237 on PDB 3D structures x-referenced by P30559: 6TPK_A, 7QVM_R, 7RYC_O

Figure 6. ScanProsite results for OXTR_human (P30559). Three PDB structures referenced by P30559 are 6TPK, 7QVM, 7RYC

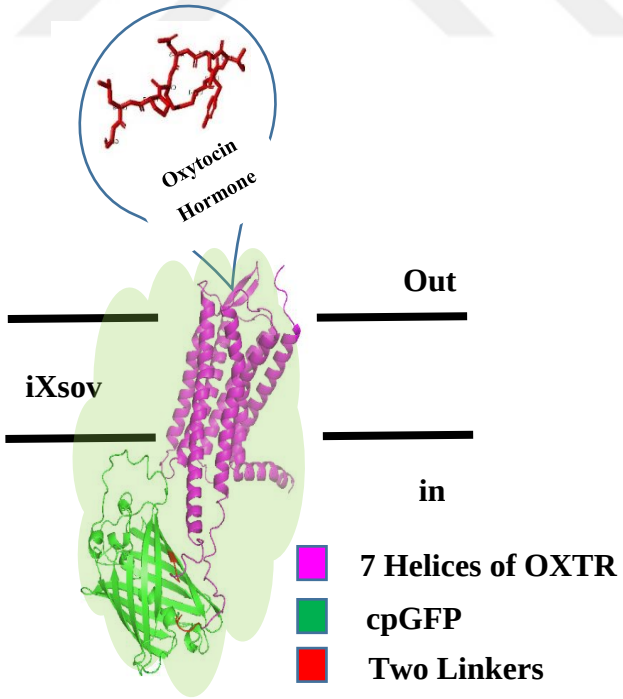


Figure 7. iXsov was designed by inserting cpGFP (green) into the third intracellular loop (ICL3) of OXTR (purple), attached with two optimized linkers (red).

Table 1.

AlphaFold2 structure prediction parameters.

num_relax	0
template_mode	custom
msa_mode	mmseqs2_uniref_env
Pair_mode	unpaired
model_type	auto
num_recycles	3
relax_max_iretation	200
pairing_strategy	greedy
max_msa	auto
num_seeds	1
runk_num	1
Color	IDDT

3.4 Membrane Bilayer Modeling

The lipid membrane was built using CHARMM-GUI. The PDB structure of our models was oriented in a lipid bilayer with the OPM webserver (http://opm.phar.umich.edu) to mimic the plasmid membrane with a heterogeneous lipid composition of 1:20 ratio of neutral CHS (cholesteryl hemisuccinate, protonated) to POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine) (Table 2). The Simulations were performed with a hexagonal periodic box in an explicit 3 points (TIP3P) water model and 0.15 M NaCl ions were added to neutralize the system (Meyerowitz et al., 2022). The simulation was performed using the CHARMM36m force field, known for its credibility in modeling protein-lipid interaction (Aimeur et al., 2024). WYF parameters were incorporated specifically for cation-pi interactions. Equilibration procedures included the automatic generation of grid information for Particle Mesh Edwald (PME) and Fast Fourier Transform (FFT). The NPT ensemble was employed with Gibbs equilibration to stabilize the temperature at 310K by the canonical ensemble. At the end of step six, the bilayer membrane simulation files were generated from CHARMM-GUI and used for MD simulation. The GROMACS files include topology and PDB files, minimization, equilibration, and production

parameters files. The result of the CHARMM-GUI lipid membrane simulation is shown in (Figure 8).

Table 2.

Lipid composition and water thickness parameters used in the four models molecular dynamics simulations to build the bilayer membrane.

Models Name	POPC	Neutral CHS	Water Thickness
Inactive OXTR	200	10	22.5
Inactive iXsov	200	10	22.5
OXTR-OT-alpha	200	10	22.5
iXsov-OT-alpha	400	20	25

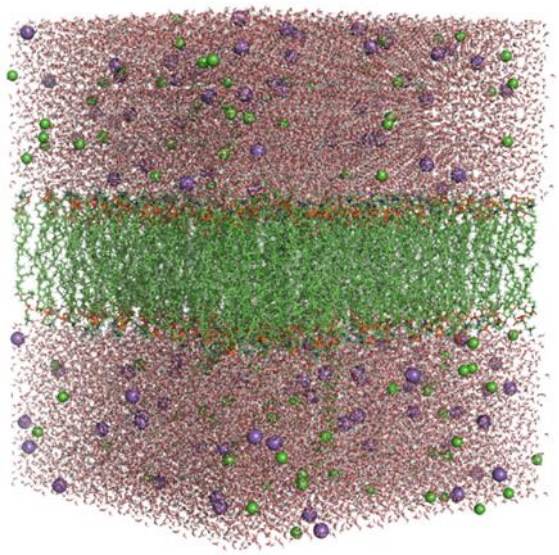


Figure 8. Three-dimensional Bilayer lipid membrane constructed by CHARMM-GUI membrane builder. Water molecules are shown in red. POPC and CHS are shown in green

3.5 Molecular Dynamic Simulations

Optical sensors are commonly used in molecular biology and cell biology research due to their capacity to report and measure protein activities in real-time. The

design of our engineered reporter involves fusing circularly permuted green fluorescent protein (cpGFP) into OXTR, our protein of interest. When OXTR interacts with its target molecules or ligands (including its native ligand oxytocin), it triggers conformational changes leading to the activation of cpGFP, which emits light at a specific wavelength (532 nm in this case). This technology enables researchers to study gene expression, protein-ligand and protein-protein interactions, signal transduction, and intracellular trafficking in real time with high temporal resolution.

MD simulations were performed in GROMACS 2023.2 version to set up, perform, and analyze MD simulation and suites under the CHARMM36m force-field (Zou et al., 2019). Moreover, it supports serial and parallel CPU and GPU simulation. I used this software in combination with Visual Molecular Dynamic (VMD) to set up and analyze the four models and explore their dynamics. Each MD simulation differs in the number of atoms and simulation time including the number of steps, and timestep. These parameters were optimized to further investigate the movement of particles caused by their interactions and decide which improvement could be made to reach targeted stability (Table 3).

The four systems were minimized for 5000 steps at absolute zero temperature to ensure the minimum potential energy (Acharya et al., 2021). The particle mesh Ewald (PME) method was used to treat long-range electrostatic interactions between pairs of atoms (Coulomb type) at 1.2 nm. Van der Waals (VDW) interactions were directed using a cut-off at 1.2 nm and a force-switch modifier at 1.0 nm. The LINCS algorithm was applied for hydrogen bond constraints. The four systems were maintained at thermodynamic ensemble in equilibration steps with the temperature and pressure set up at 310 K and 1.0 bar, respectively, to replicate to the environment of the human body. All equilibration steps were run for 15 ns with time steps of 0.001 ps and 0.002 ps to ensure the stabilization of systems before the production phase. Production simulation was conducted in the NPT ensemble. Each system was performed for 115 ns simulation time.

Table 3.

Molecular Dynamic Simulation Systems parameters. Information was extracted from CHARMM-GUI and GROMACS.

Simulated System	Inactive OXTR	Inactive iXsov	OXTR-OT-Galpha (7RYC)	iXsov-OT-Galpha (7RYC)	OXTR-OT-Galpha (7QVM)	iXsov-OT-Galpha (7QVM)
No. of residues	389	640	626	877	626	877
No. of residues after truncation	324	575	561	812	561	812
Box dimension (A)	12x12x13	12x12x16	12x12x15	17x17x16	12x12x14	17x17x17
Total no. of atoms	198116	230680	216996	459634	205335	470341
Frames	1001	1001	1001	1001	1001	1001
No. of water molecules	137316	55304	50815	111843	46931	115407
No. of lipids: upper/Lower leaflet	CHS: 10/10 POPC: 200/200	CHS: 10/10 POPC: 200/200	CHS: 15/15 POPC: 300/300	CHS: 20/20 POPC: 400/400	CHS: 15/15 POPC: 300/300	CHS: 20/20 POPC: 400/400
No. of Na+/Cl- ions [ ]	0.15 M	0.15 M	0.15 M	0.15 M	0.15 M	0.15 M
Simulation time (ns)	115	115	115	115	115	115

3.6 Binding Free Energy Calculations (MM-PBSA & Binding Energy Decomposition Analysis)

The binding free energy is an essential parameter for determining the strength of protein-ligand interactions. Binding affinity and binding energy are obtained from docking, representing the stability of ligands within the binding pocket of the protein and the strength of their interaction. In addition, binding free energy is also a crucial indicator of the change in the potential energy surface in the docked state, providing insight into the thermodynamic stability of the complex. The solvation-free energy, molecular mechanical energy, and conformational entropy changes are employed for the calculation of the free binding energy (Hall et al., 2020). Calculations are made through the GROMACS MM-PBSA package using the ligand-protein complex for 1000 frames. The *g_mmpbsa* tool developed by Kumari *et al.* is used with a trajectory file (trr or xtr), a topology parameter file (tpr), an index file (ndx), and a file for $G_{solvation}$ parameters (mdp) as input (Kumari et al., 2014). The equation to calculate ΔG_{bind} is given below:

$$\Delta G_{bind} = \Delta G_{solv} + \Delta E_{MM} + \Delta G_{SA}$$

Where ΔG_{solv} represents the difference in PBSA solvation energy between the protein-ligand complex, ΔE_{MM} is the difference in minimized energies between the protein-ligand complex and the sum of energies for the unligated protein and ligand, and ΔG_{SA} indicates the difference in surface area energies between the complex and the sum of surface area energies for the unligated protein and ligand. These parameters are incorporated in the calculation of binding free energy. Through MD simulations coupled with MM-PBSA decomposition analysis, we obtained information about energetically stable conformations to understand molecular interactions in our models.

The MM-PBSA calculations produce two files as output. The first file contains the following average energies: VDWAALS (wan der Waals contribution from MM), EEL (electrostatic energy), EGB (the electrostatic contribution to the solvation-free energy calculated by GB), ESURF (nonpolar contribution to the solvation free energy calculated by an empirical model) and DELTA $G_{binding}$ (the final estimated binding free energy). The second output file is FINAL_DECOMP_MMPBSA.dat, which contains information about the interaction energy between every residue with the rest of the system, presented in several sections.

Chapter 4
Results

4.1 Protein Modeling with AlphaFold2

Each step of the computational procedure used to obtain models for iXsov is detailed in the methodology section (Chapter 3). While crystal structures for OXTR are available in the Protein Data Bank (PDB), they contain gaps due to missing residues and unresolved loops. Therefore, homology modeling was required for conducting MD simulations. Two homology models were folded for OXTR in its inactive form with and without cpGFP, and four homology models were folded for OXTR in active form with and without cpGFP. AlphaFold2 was used for folding calculations using the crystal structures (PDB ID: 6TPK, 7QVM, and 7RYC) as templates. This approach helps overcome experimental limitations of missing residues in structures presented in PDB. Refolding these structures recovers missing parts to avoid collisions in MD simulation. Moreover, AF2 facilitates the folding of *de novo* proteins that are not present in PDB, such as our iXsov chimera. The structural modeling of the complete oxytocin receptor complex including G protein alpha, beta, and gamma subunits couldn't be achieved using the AF2 free version due to the high number of residues (>1000) implicated in each complex. In addition, all-atom simulation of the entire complex was deemed too computationally intensive for a preliminary study. As such, focus was given to the $G\alpha_O$ subunit that directly interacts with OXTR.

The percentage of correctly predicted inter-atomic distances (pLDDT) of OXTR in its inactive form is ~80-90, suggesting high confidence in structural accuracy (Figure 9A). However, pLDDT values for iXsov is lower ($60<pLDDT<69$) (Figure 9B). The prediction of OXTR-Gα structures is in the 70-79 range, which suggests that the backbone is likely to be accurate (Figure 9C-F). The general score is impacted by the lower prediction of loops and the N and C termini, which are highly motile. Despite this, the 7 helices are all predicted in high quality (pLDDT>90). In order to reduce the time cost of our MD simulations, we implemented a truncation strategy on the amino

acids sequences of our models in the C and N-termini, which were poorly predicted to start with. This strategy also has the advantage of optimizing the prediction score. Additionally, helix 8, which stabilizes the C-terminal end of the structure against the membrane, has a pLDDT ranging between 60 and 70. The pLDDT of iXsov overall reflects high confidence in the predicted structure and ranges between 80 and 90 (Figure 9H). However, loops in general and particularly the third loop have notably low prediction confidence with pLDDT <50, classifying them as disordered regions.

The major difference between the AF2-7RYC and AF2-7QVM models is the location of the Gα subunit in the N-terminus of OXTR. AF2-7RYC was unable to insert the helix correctly, as reflected by its low prediction confidence (50<pLDDT<60, Figure 9C). This model shows a significant deviation in its orientation in intracellular regions, while we expected to see a similar position observed in prior literature and the AF2-7QVM template. Compared to the three other structures of active OXTR, the structure model based on the 7RYC template in the absence of cpGFP was found to exhibit differences in the Gα binding region, which may not truly correspond to OXTR in its active form. As such, the AF2-7QVM model appears to be the more accurate template for the activate OXTR folding models. In other AF2-generated models, however the cpGFP domain adopts a clear position in the intracellular side of the lipid membrane and does not interfere with the Gα protein domain, suggesting the absence or reduction of the troubleshooting clashes in the structure. With these results, we provide an approximate full structure of the active and inactive configuration of the OXTR-encoded fluorescent reporter structures (iXsov) for subsequent MD simulations.

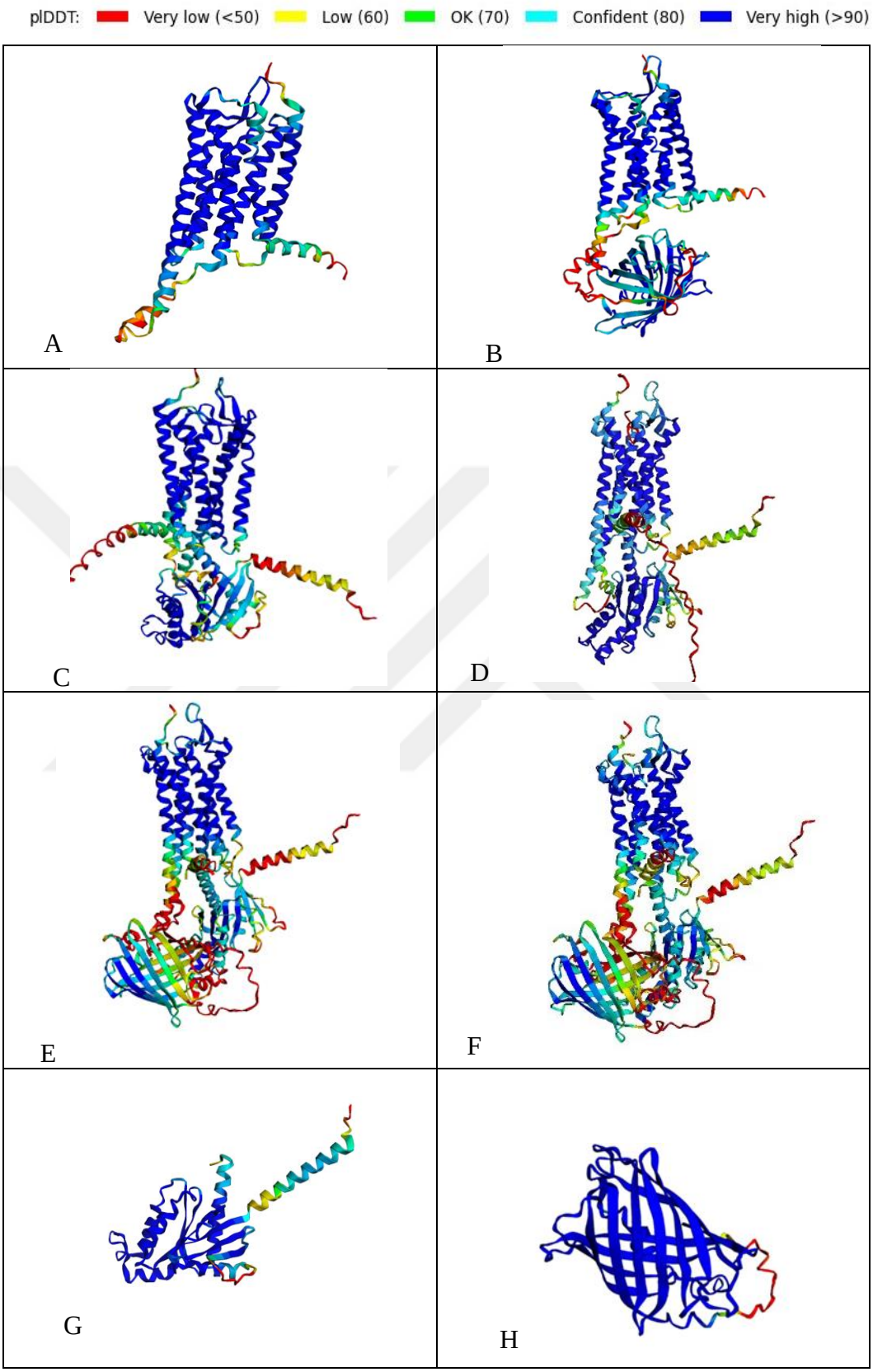


Figure 9. AlphaFold2 structure predictions. (A) OXTR, inactive form. (B) iXsov, inactive form. (C) OXTR-Gα (AF2-7RYC). (D) iXsov-Gα (AF2-7RYC). (E) OXTR-Gα (AF2-7QVM). (F) iXsov-Gα (AF2-7QVM). (G) Gα protein subunit. (H) cpGFP domain.

4.2 MD Simulation Results and Discussion

GPCRs has been thoroughly explored using MD simulation, highlighting their stability, dynamics, and interactions with ligands and with G protein subunits (Bruno & Costantino, 2012). This approach offers a comprehensive view of the contribution of the seven alpha-helical TM in protein allostery, understanding how protein-ligand binding triggers specific conformation changes. This study is therefore crucial for confirming the suitability of iXsov as a potential assay for discovering OXTR-modulating drugs. MD simulation also provides essential data to investigate trajectories and individual residues.

4.2.1 Total Energy

Total energy analysis plays a crucial role in providing insight into the stability and dynamics of protein-ligand systems. It also provides information about the selection of an appropriate force field. The comparison of the calculated total average energy between different systems (Table 4) remains steady and conserved throughout the simulation in both types of models (OXTR and OXTR-encoded fluorescent reporter), providing validation to the simulation and confirming the accuracy and reliability of the chosen force field (CHARMM36m). This analysis reflects how well the CHARMM36m force field parameters capture the interactions within the system, especially in modeling protein-lipid interactions – a fact previously noted in the literature (Aimeur et al., 2024).

The dynamic range refers to a significant conformation change within the systems depending on the active or inactive state, ligand binding, and the presence or absence of cpGFP within the structure. In dynamic range analysis, we observe a significant conformation alteration of >10 Å in the oxytocin receptor when coupled to cpGFP in its inactive state. Deviations of 21 Å and 25 Å were observed when OXTR is modeled based on 7QVM and 7RYC, respectively. The changes are further accentuated in structures lacking incorporated cpGFP, as we observe dynamic alteration of 5 Å, 7 Å, and 13 Å in inactive OXTR, OXTR-7QVM model, and OXTR-7RYC model, respectively (Table 4).

Table 4.

Total energy profile of the four systems at 310 K and dynamic range in Å.

Studies Systems	Total average energy	Dynamic range in Å
Inactive OXTR	-1.41 x 10^6	5 Å
Inactive iXsov	-1.75 x 10^6	10 Å
OXTR-OT-Galpha (7ryc)	-1.60 x 10^6	13 Å
OXTR-OT-Galpha (7qvm)	-1.46 x 10^6	7 Å
iXsov-OT-Galpha (7ryc)	-3.54 x 10^6	25 Å
iXsov-OT-Galpha (7qvm)	-3.66 x 10^6	21 Å

4.2.2 Root of Mean Square Deviation (RMSD)

4.2.2.1 Inactive OXTR and Inactive iXsov

The deviations of inactive OXTR and inactive iXsov were evaluated throughout 100 ns according to the RMSD of the backbone of all atoms with respect to the starting structure coordinates of each system. MD calculations are plotted in (Figure 10). The first 50 ns of the two independent runs for each model show different dynamic levels. However, in the second 50 ns simulation time we observe equilibration of inactive OXTR and inactive OXTR with incorporated cpGFP (iXsov).

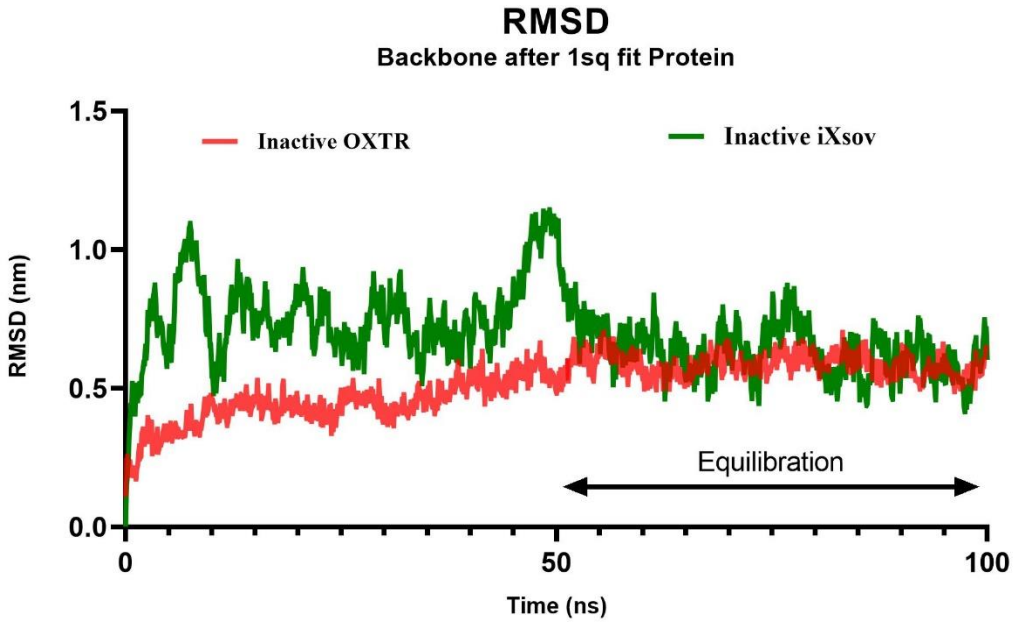


Figure 10. RMSD analysis for inactive OXTR (red) and iXsov (green).

4.2.2.2 RMSD comparison of inactive and active states of OXTR

Previous studies have demonstrated that spontaneous fluctuations are essential for the functional capabilities of GPCRs and other receptors (Changeux & Edelstein, 2011). In our analysis, we examined the root mean square deviation (RMSD) during the molecular dynamic simulation of the inactive OXTR, inactive iXsov (AF2-6TPK), OXTR receptor bound to oxytocin and Gα (OXTR-OT-Gα) and iXsov-OT-Gα complexes (AF2-7QVM). RMSD results suggest that OXTR is structurally stable after 50 ns of simulation regardless of the presence or absence of cpGFP, suggesting that the insertion does not disrupt the structure of the receptor or lead to unfolding in the inactive form (Figure 12).

We then ran simulations involving the OXTR in the presence of its ligand (oxytocin) to stimulate its activity. The RMSD analysis of the OXTR-OT-Gα (AF2-7QVM) complex suggests that the structure is also stable, confirming that ligand and Gα binding do not significantly alter backbone stability. In contrast, iXsov-OT-Gα (AF2-7QVM) exhibited major fluctuations and high dynamic motion, as shown in Figures 11 and 12, potentially corresponding to the activation of cpGFP in response to oxytocin-induced conformational changes.

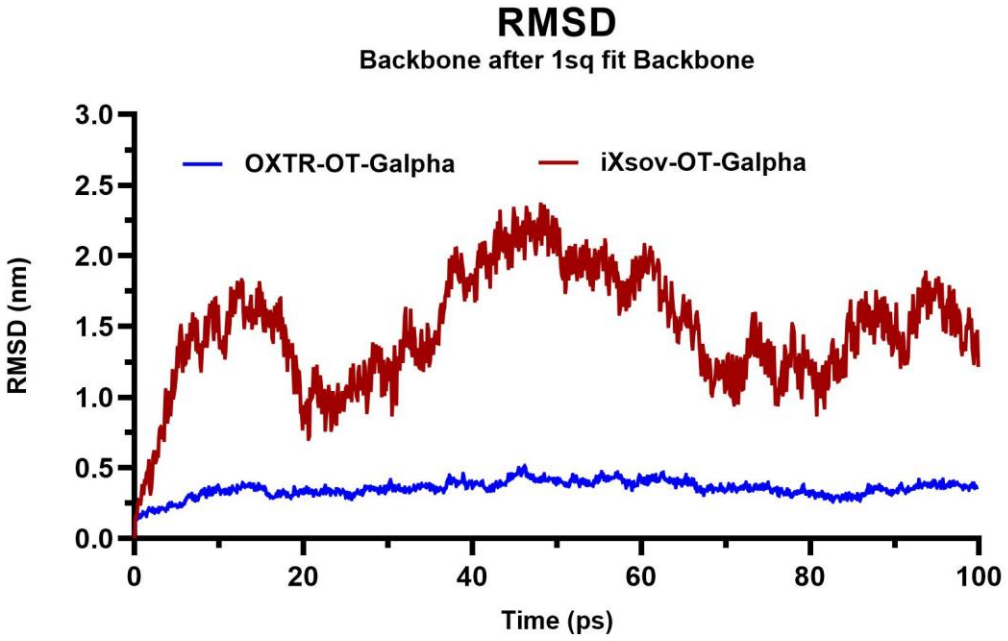


Figure 11. Comparative RMSD analysis of OXTR and iXsov (AF2-7QVM) in active states with ligand and Galpha protein interaction.

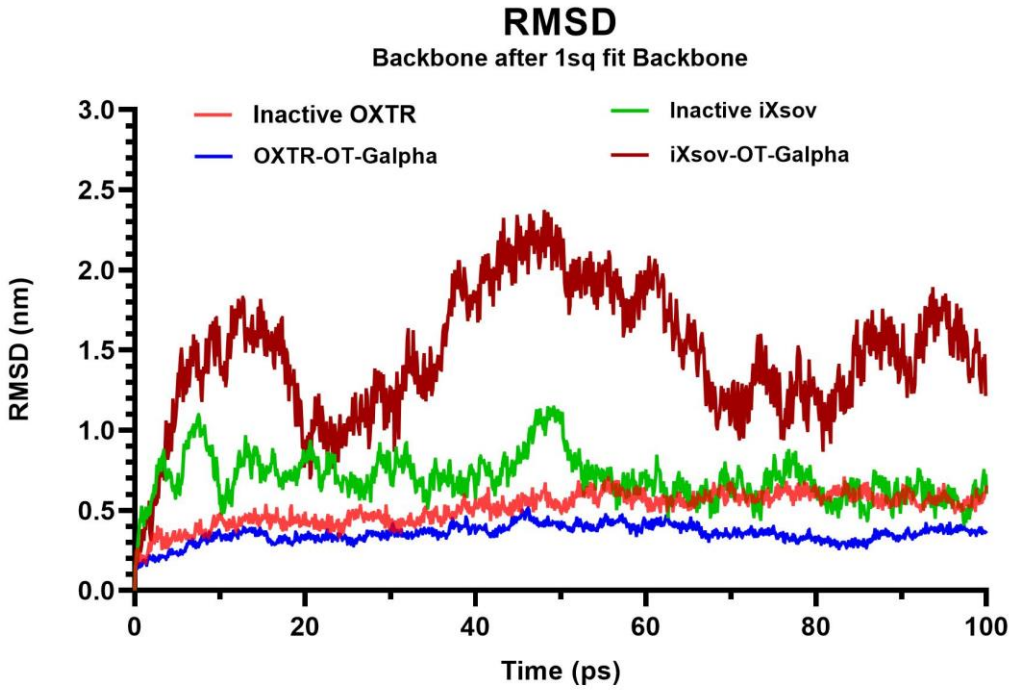


Figure 12. Comparative RMSD analysis for active (AF2-7QVM) and inactive (AF2-6TPK) structures of OXTR and iXsov

4.2.2.3 RMSD comparison of AF2-generated OXTR model based on 7RYC and 7QVM templates

RMSD analysis provides an important dynamic view of protein complexes in their active state. Our study was intended to provide information on the behavior of four OXTR complexes binding to their ligand (oxytocin) and Gαo. However, we observed distinct structural dynamic behaviors among these complexes, primarily attributed to specific template used for homology-based modeling. As such, a more detailed analysis was conducted to determine the precise cause of these differences, treating the three component molecules of the complex (OXTR, oxytocin and the Gαo protein) as separate entities for RMSD analysis.

4.2.2.3.1 Oxytocin receptor

The oxytocin receptor of the AF2-7RYC and AF2-7QVM gave an RMSD of 6 Å and 4 Å, respectively (Figure 13). However, the dynamic motion of the complex of the OXTR AF2-7RYC remains a challenge, compared with complex AF2-7QVM, considering the fact that the assembly of the complex OXTR-7RYC was different in the N terminus of G-alpha protein subunit. Unfortunately, this lowest folding

confidence in the complex OXTR-OT-Galpha (AF2-7RYC) is demonstrated in a higher dynamic range in RMSD analysis (Figure 13). iXsov also show differences between models based on 7RYC and 7QVM templates. In iXsov-OT-Galpha complex AF2-7QVM based template is more stabilized after 80 ns of simulation compared to the same complex (7RYC-based template) in which we observe less stability and an increase in dynamic motion (Figure 14). However, compared with the wild type of OXTR in the active form (4 Å <RMSD <6 Å), we still confirm the remarkable changes in structure obtained in the presence of cpGFP (5 Å<RMSD<10 Å).

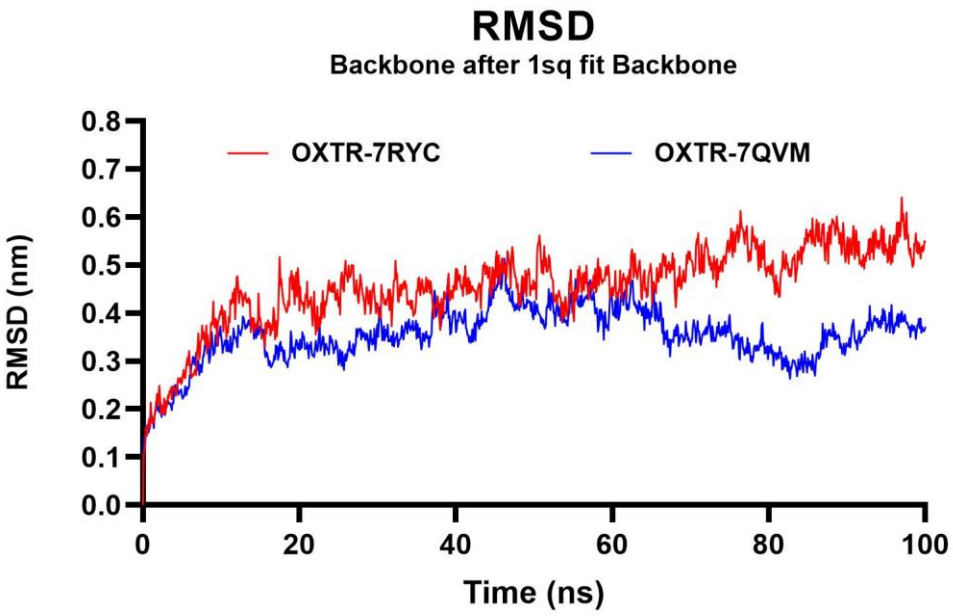


Figure 13. Comparative RMSD analysis for the active structure of OXTR AF2-PDB based template (7RYC-Red) and (7QVM-Blue)

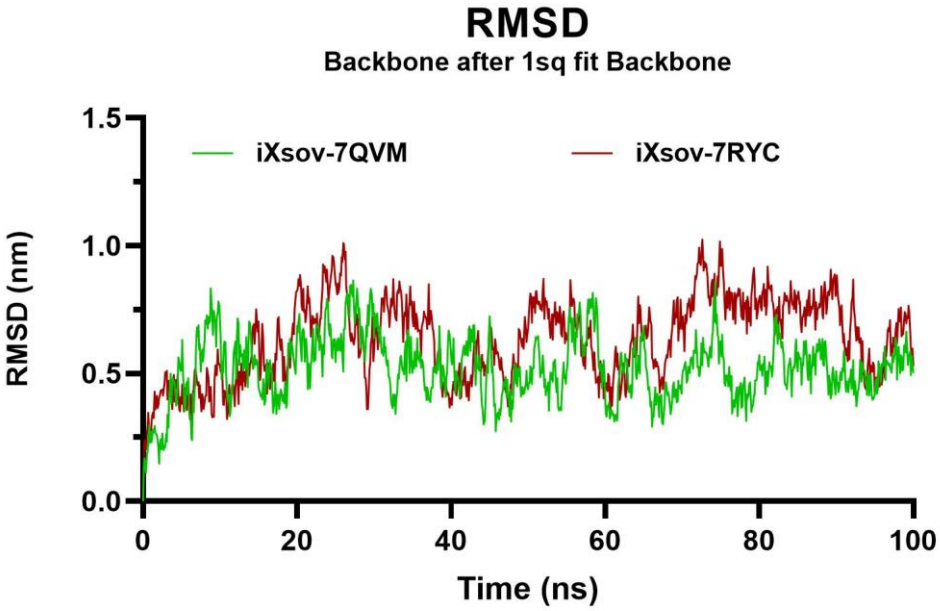


Figure 14. Comparative RMSD analysis for the active structure of iXsov-7QVM (green) and iXsov-7RYC (brown).

4.2.2.3.2 G-alpha protein

The N-terminus helices of iXsov AF2-7QVM/7RYC and OXTR AF2-7QVM structures are structurally aligned. However, the N-terminus of OXTR AF2-7RYC is rotated away (Figure 15). This deviation is due to the low structural prediction quality observed in AlphaFold results (pLDDT<50). Unlike the low prediction of the N-terminus of OXTR AF2-7RYC, the prediction of the Gα N-terminus in other complexes (OXTR and iXsov) ranged between (50<pLDDT<70). This deviation and low prediction quality of the N-terminus (OXTR AF2-7RYC), imply a higher dynamic motion in the RMSD analysis of G- alpha, compared to AF2-7QVM models (Figure 16), indicating the influence of the AF2 prediction on RMSD analysis. On another hand, the RMSD analysis of iXsov and WT OXTR was compared to observe the impact of cpGFP on the dynamic motion of the G-alpha subunit. The G-alpha subunits in the iXsov AF2-7QVM/7RYC and OXTR AF2-7QVM structures show a similar approximate range of motion (Figure 17).

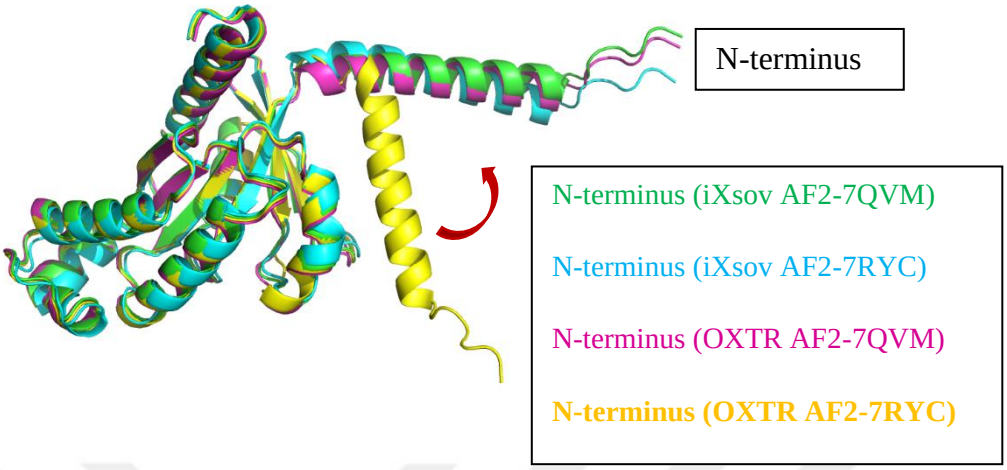


Figure 15. N-terminal view of the G-alpha Ras-like domai) in OXTR and iXsov structures

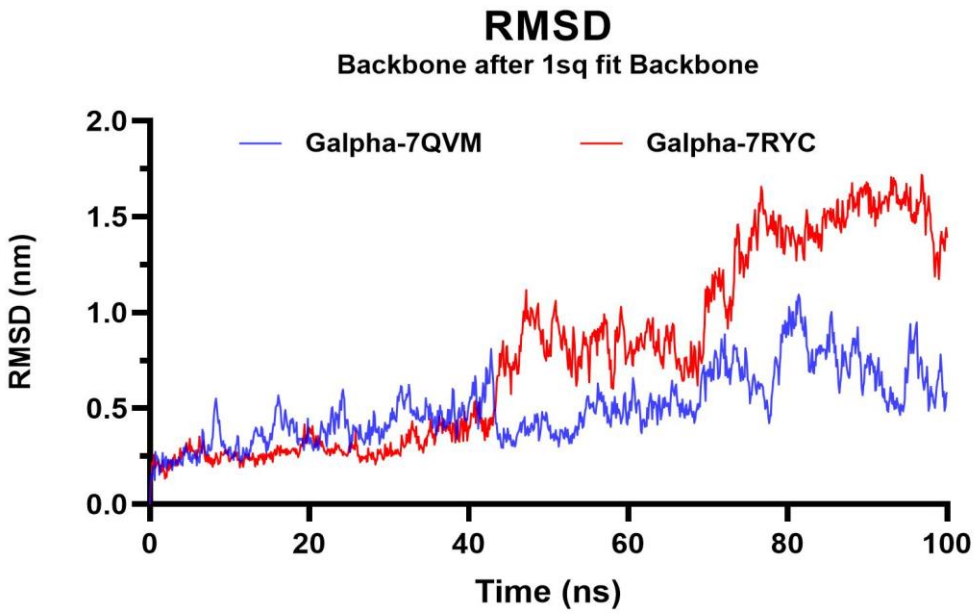


Figure 16. Comparative RMSD analysis of the G-alpha protein subunit PDB-based templates AF2-7QVM (blue) and AF2-7RYC (red)

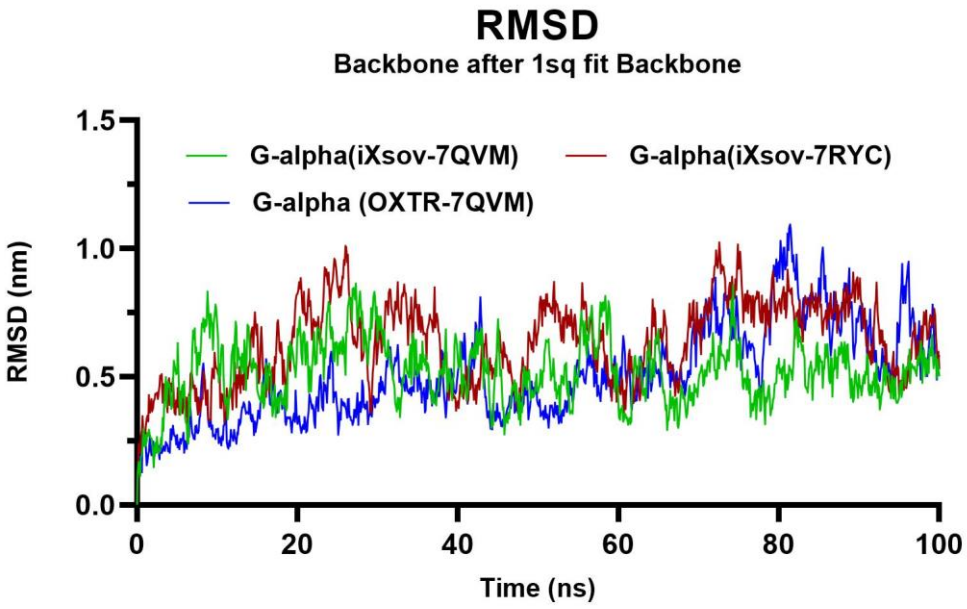


Figure 17. Comparative RMSD analysis of the G-alpha protein subunit of iXsov structures AF2-7QVM (green) and AF2-7RYC (brown) and the G-alpha protein of OXTR AF2-7QVM (blue).

4.2.3 Root mean square fluctuation (RMSF) analysis

4.2.3.1 RMS fluctuations in ICL3

Strong RMS fluctuations were observed in the third intracellular loop (ICL3) of OXTR in the absence of cpGFP, with deviations reaching 8 Å. In contrast, other loops exhibited fluctuations around 3 Å, indicating relatively stable regions in contrast to the flexible and dynamic nature of ICL3 (Figure 18).

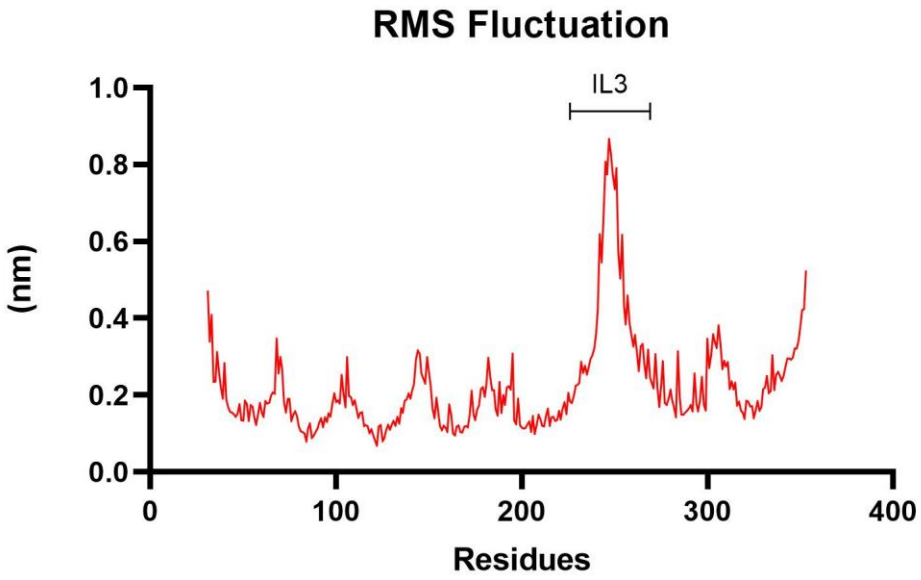


Figure 18. Residue RMSF analysis for OXTR and its intracellular third loop (ICL3)

4.2.3.2 RMS fluctuation in oxytocin receptor residues

In addition to gaining insights into the conformational changes adopted by the oxytocin receptor through RMSD trajectory analysis, we explored each residue's flexibility using root mean square fluctuation (RMSF) analysis. Since the experimental structures have some missing residues, we utilized AlphaFold2-generated models folded in the presence of the ligand to derive RMSF data.

The wild-type oxytocin receptor showed a fluctuation range of approximately 4 Å for both model-based templates (7QVM & 7RYC). However, when incorporating cpGFP into the oxytocin receptor (iXsov), we observe significantly higher fluctuation in the cpGFP domain, exceeding 15 Å, while the residues of the OXTR showed lower fluctuation, around 7 Å (Figure 19). Analyzing the RMSF of each residue during molecular dynamic simulation for iXsov in interaction with oxytocin revealed increased fluctuation, particularly in the cpGFP domain. These fluctuations may reflect the activation of cpGFP domain, predicting a spike in fluorescence in response to oxytocin in in vitro studies. On the other hand, the G-alpha protein domain did not present much fluctuation differences across the three complexes, except for OXTR AF2-7RYC due to its low prediction confidence and instability (Figure 20). From

these results, we can conclude the incorporation of cpGFP does not significantly alter residue stability outside ICL3 in OXTR, suggesting that its impact on protein function will be minimal.

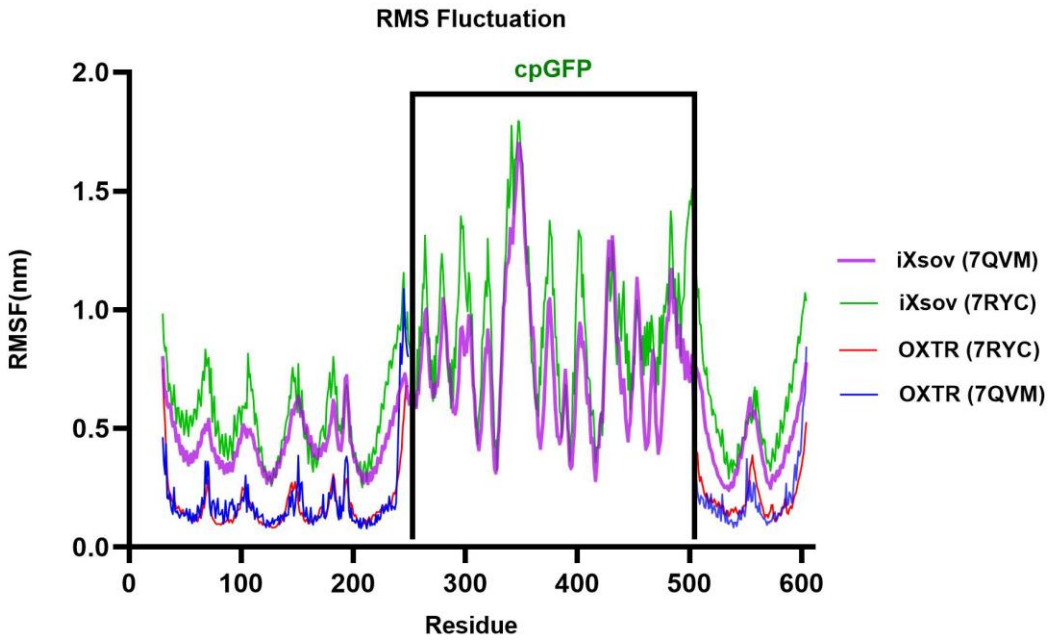


Figure 19. Comparative RMSF analysis for the OXTR and iXsov for both templates (7QVM and 7RYC)

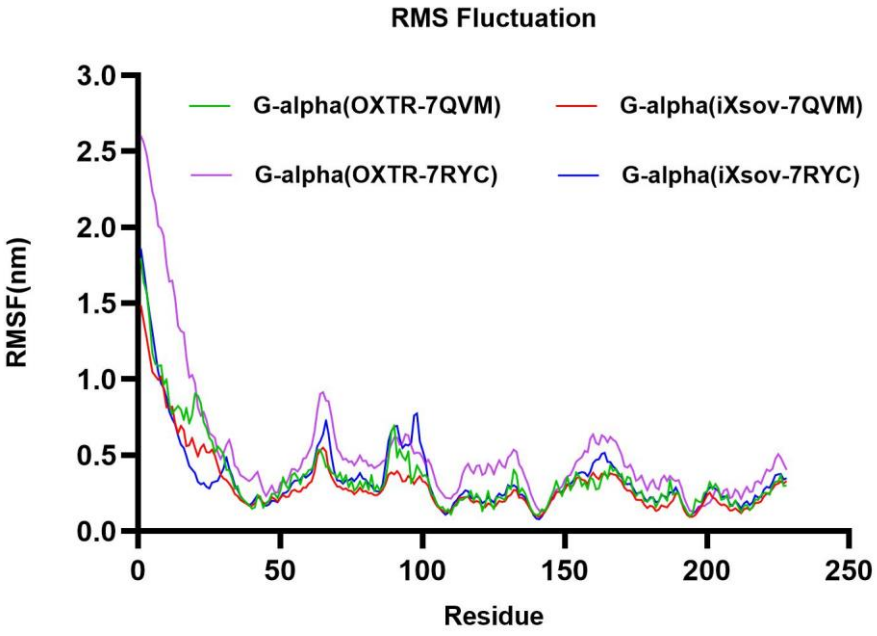


Figure 20. Comparative RMSF analysis for the G-alpha protein subunit in OXTR and iXsov with AF2-7RYC & 7QVM templates

4.2.4 Hydrogen Bonds

Protein structures typically maintain their stability through hydrogen bonding. As such, hydrogen bond analysis was performed as a supplement to the previously-described backbone fluctuation data. While all four structures were generally stable, the first 50 ns of the OXTR AF2-7RYC simulation showed some instability, as indicated by higher fluctuation and the formation of 15 additional hydrogen bonds. However, this structure also stabilizes in the second 50 ns of simulation, suggesting a more energetically favorable state with a reduced number of hydrogen bonds (10). On the other hand, OXTR AF2-QVM and iXsov maintained stability during 100 ns simulation with around 10 hydrogen bonds formed (Figure 21).

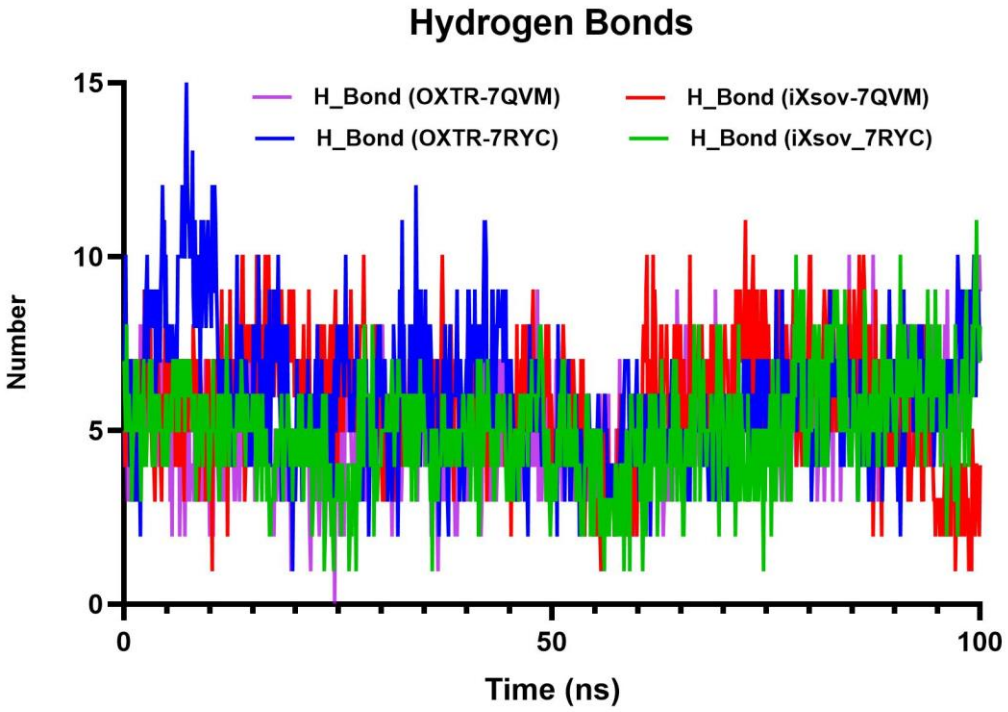


Figure 21. Comparative hydrogen bond analysis throughout 100 ns of simulation in OXTR and iXsov using AF2-7QVM and 7RYC templates

4.2.5 Binding Free Energy Calculation (MM-PBSA & Energy decomposition analysis)

4.2.5.1 Molecular Mechanics Poisson-Boltzmann surface area (MM-PBSA)

In our study, we utilize the gmx_MMPBSA decomposition analysis package developed by Kumari et al. to calculate binding free energy (Kumari et al., 2014). This analysis predicts the total free binding energy of the protein-ligand complex and the energy contribution of each residue, enhancing our understanding of the molecular interactions involved. The total free binding energy values of the OXTR-OT-Galpha and iXsov-OT-Galpha complexes based on different templates are given in Figure 22.

From our findings, it is clear that the complexes folded based on the 7QVM template have a lower free binding energy of -51.73 kcal/mol and -51.27 kcal/mol for OXTR and iXsov, respectively. In contrast, complexes folded based on the 7RYC

template showed higher free energy binding values of -41.68 kcal/mol and -50.40 kcal/mol for iXsov-OT-Galpha and OXTR-OT-Galpha, respectively.

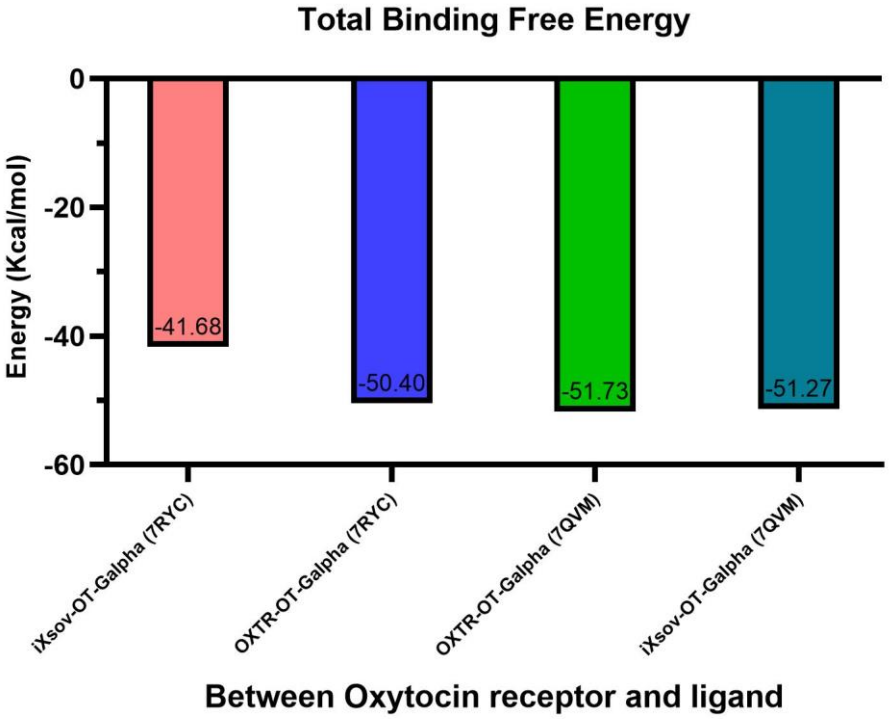


Figure 22. Total binding free energy between OXTR and oxytocin for WT-OXTR and iXsov complexes.

Furthermore, we analyzed the interaction between each residue and the receptor in the active state to identify their individual level contribution to the binding affinity and stability.

4.2.5.2 Decomposition Analysis

4.2.5.2.1 OXTR-OT binding mode

In this MM-PBSA decomposition analysis, the binding energy is decomposed per residue. This process allows the calculation of the relative importance of each residue on the protein-ligand binding interaction. G_mmpbsa decomposes total energy into the individual contribution of each residue, enabling relative comparisons to be made. Here we calculated the 20 highest-contributing residues in the OXTR-OT and iXsov-OT systems (Table 5) based on both templates (7QVM and 7RYC). The results demonstrate a similarity in the residues and energy contribution profile for the four complexes. The residues that similary interact with OT are : GLN92, GLN96 LYS116

GLN119 ILE201 PHE295, GLN295 (OXTR) /PHE542, GLN546, (iXsov), ILE201, ILE204, TRP188.

In the MMPBSA decomposition analysis of the oxytocin ligand, we observe that all nine amino acids contribute to OXTR-OT and iXsov binding. The residues most implicated in the binding are CYS1, TYR2, ILE3, and GLN4 with an energy range of -3.88 to -8.60.

Table 5.

Comparative MM-PBSA decomposition analysis results

	OXTR-OT(7RYC)		iXsov-OT (7RYC)		OXTR-OT(7QVM)		iXsov-OT (7QVM)	
R	**PRO:32**	-0.95	**PRO:32**	-0.72	**LEU:38**	-0.90	**ARG:34**	-1.28
R	**ARG:40**	0.94	**ARG:33**	-0.55	**GLU:42**	-2.97	**ASN:35**	-0.72
R	**GLN:92**	-1.93	**ARG:34**	-1.13	**GLN:92**	-1.01	**LEU:38**	-0.74
R	**LYS:116**	1.41	**GLN:92**	-1.69	**GLN:96**	-1.19	**GLN:92**	-0.62
R	**GLN:119**	-0.62	**GLN:96**	-0.61	**TRP:99**	-0.63	**GLN:96**	-1.15
R	**PHE:175**	-1.25	**LYS:116**	2.55	**ASP:100**	2.17	**LYS:116**	2.07
R	**TRP:188**	-2.48	**GLN:119**	-0.61	**LYS:116**	2.14	**GLN:119**	-0.95
R	**ALA:189**	-0.53	**MET:123**	-0.77	**GLN:119**	-0.90	**VAL:120**	-0.74
R	**PHE:191**	-0.73	**PHE:175**	-1.22	**VAL:120**	-0.62	**MET:132**	-0.92
R	**PRO:197**	-0.73	**TRP:188**	-3.37	**MET:123**	-0.76	**GLN:171**	-0.52
R	**ILE:201**	-1.18	**ALA:189**	-2.18	**PHE:175**	-1.00	**PHE:175**	-1.38
R	**ILE:204**	-0.59	**PHE:191**	-0.77	**TRP:188**	-1.96	**TRP:188**	-2.39
R	**PHE:291**	-1.33	**TYR:200**	-0.98	**ALA:189**	-1.53	**ALA:189**	-1.77
R	**GLN:295**	-1.75	**ILE:201**	-0.94	**PHE:191**	-0.65	**PHE:191**	-0.67
R	**SER:298**	-1.64	**ILE:204**	-0.59	**TYR:200**	-1.43	**TYR:200**	-0.64
R	**VAL:299**	-0.60	**PHE:542**	-0.98	**ILE:201**	-0.93	**ILE:201**	-0.95
R	**LYS:306**	-0.82	**GLN:546**	-0.67	**PHE:291**	-0.87	**ILE:204**	-0.52
R	**PHE:311**	-0.56	**ILE:563**	-0.84	**GLN:295**	-1.95	**PHE:542**	-1.19
R	**ILE:312**	-1.21	**MET:566**	-1.68	**VAL:299**	-0.69	**GLN:546**	-3.35
R	**LEU:316**	-3.20	**LEU:567**	-0.99	**ILE:316**	-0.65	**VAL:567**	-0.57
L	**CYS:1**	-5.51	**CYS:1**	-6.84	**CYS:1**	-5.79	**CYS:1**	-4.23
L	**TYR:2**	-5.99	**TYR:2**	-8.60	**TYR:2**	-7.72	**TYR:2**	-6.61
L	**ILE:3**	-6.77	**ILE:3**	-5.80	**ILE:3**	-6.24	**ILE:3**	-5.65
L	**GLN:4**	-3.29	**GLN:4**	-2.29	**GLN:4**	-3.88	**GLN:4**	-3.50

L	ASN:5	-2.84	ASN:5	-2.69	ASN:5	-2.29	ASN:5	-2.63
L	CYS:6	-1.90	CYS:6	-1.06	CYS:6	-0.98	CYS:6	-1.32
L	PRO:7	-3.17	PRO:7	-2.06	PRO:7	-2.71	PRO:7	-3.55
L	LEU:8	-2.96	LEU:8	-3.31	LEU:8	-3.45	LEU:8	-3.47
L	GLY:9	-3.45	GLY:9	-1.88	GLY:9	-1.69	GLY:9	-2.51

R: Receptor L: Ligand G: G protein alpha

G protein Interaction

	OXTR-OT (7RYC)		iXsov-OT (7RYC)		OXTR-OT(7QVM)		iXsov-OT (7QVM)	
R	ARG:73	-1.47	ARG:73	-1.07	SER:72	-1.57	SER:72	-1.58
R	LEU:74	-1.53	LEU:74	-1.53	ARG:73	-12.45	ARG:73	-4.07
R	ASP:136	-0.56	ARG:137	-2.56	LEU:74	-1.58	LEU:74	-1.99
R	ARG:137	-2.22	ALA:140	-1.22	ASP:136	0.84	ARG:137	-3.57
R	ALA:140	-1.63	ILE:141	-2.63	ARG :137	-2.37	ALA:140	-1.83
R	ILE:141	-2.43	PRO:144	-2.33	ALA :140	-1.04	ILE:141	-2.61
R	PRO:144	-2.55	LEU:145	-2.54	ILE :141	-2.63	PRO:144	-2.79
R	LEU:145	-2.44	LEU:148	-1.34	PRO :144	-2.08	LEU:145	-3.32
R	SER:147	-1.54	ILE:227	-1.34	LEU:145	-2.25	LEU:148	-1.40
R	LEU:148	-2.79	LYS:234	-0.79	SER :147	-0.99	ILE:227	-0.86
R	ARG:150	-0.76	PHE:329	-3.67	LEU :148	-2.68	LYS:234	-9.26
R	ILE:227	-1.54	THR:331	-0.54	ARG:150	-0.67	ALA:248	-1.19
R	LYS:234	-6.40	ARG:510	-6.33	ARG :151	-1.07	ARG:510	-10.80
R	ARG:272	-3.73	SER:512	-1.71	ILE:227	-0.90	VAL:511	-5.92
R	PHE:329	-3.03	LEU:516	-1.73	ARG :272	-4.37	SER:512	-1.17
G	ALA:31	0.51	LYS:32	0.51	ALA :31	-0.53	LEU:79	-0.93
G	LYS:32	-0.77	LEU:79	-0.97	LYS :32	-0.55	GLU:164	0.93
G	LEU:79	-1.85	PRO:163	-1.55	LEU :79	-1.23	THR:166	-1.06
G	GLU:164	0.92	GLU:164	0.81	GLU :172	0.66	GLN:179	-0.73
G	ASP:173	0.55	ASP:173	0.77	GLU :183	0.92	PRO:189	-0.56
G	ALA:175	-0.43	ALA:175	-0.74	GLU :192	-0.63	ASN:190	-0.87
G	GLN:179	-1.49	ALA:176	-1.42	ILE :193	-0.78	GLU:192	-3.21
G	THR:214	-1.11	GLN:179	-1.09	THR :214	-1.00	THR:214	-0.65
G	ASP:215	-1.77	THR:214	-0.77	ASP :215	-1.83	ASP:215	-2.15
G	ILE:217	-2.86	ASP:215	-1.16	ILE :217	-2.56	ILE:217	-2.36
G	LEU:218	-2.97	ILE:217	-2.17	LEU :218	-3.15	LEU:218	-3.18
G	ASN:221	-2.98	LEU:218	-2.88	GLN :219	-0.56	ASN:221	-2.68

G	**LEU:222**	-2.67	**ASN:221**	-1.69	**ASN :221**	-3.90	**LEU:222**	-2.48
G	**ARG:223**	-2.49	**LEU:222**	-2.49	**LEU :222**	-2.73	**ARG:223**	0.7
G	**GLU:224**	-1.78	**GLU:224**	-2.72	**GLU224**	-3.12	**GLU:224**	-2.67
G	**TYR:225**	-4.51	**TYR:225**	-4.11	**TYR :225**	-6.04	**TYR:225**	-8.05
G	**ASN:226**	-2.96	**ASN:226**	-2.83	**ASN :226**	-3.45	**ASN:226**	-1.15
G	**LEU:227**	-3.64	**LEU:227**	-3.82	**LEU :227**	-3.68	**LEU:227**	-4.08
G	**VAL:228**	-1.49	**VAL:228**	-0.72	**VAL:228**	-1.00	**VAL:228**	-0.53

4.2.5.2.2 OXTR Gα$_O$ protein interaction

In the active OXTR signaling complex, helix 5 of Go/q alpha protein binds within a crack formed by helices II, III, V, VI, and VII on the intracellular side of the OXTR receptor as shown in Figure 24-C. In Figure 24-B we observe a precise structural alignment of the individual Gα helix 5 for both OXTR and iXsov complexes and also for both templates (7QVM and 7RYC).

The MM-PBSA tool was used to predict the residues involved in the interaction between OXTR Gα$_O$ in of the four complexes including OXTR and iXsov modeled based on AF2-7QVM/7RYC templates. The analysis aimed to compare our prediction with experimental conformational analysis.

Our results revealed that the four proposed models shared preserved residues contribution in G alpha helix 5 and in OXTR specificaly in Helices 1, 3, 5,6 and 7. The following OXTR residues were observed to contribute significantly in the four complexes: ARG73 and LEU74 (TM1), ARG137, ALA140, ILE141, and PRO144 (TM3), ILE227 (TM5), ARG272 (OXTR) / ARG510 (iXsov) (TM6). In addition to the residues from OXTR, those from G alpha o helix 5 also play a significant role in the interaction of OXTR and G protein α5 helix. The contribution of residues is consistently observed across the four complexes, including, VAL228, LEU227, ASN226, TYR225, GLU224, LEU222, ASN221, LEU218, ILE217, ASP215, THR214 (Figure 23).

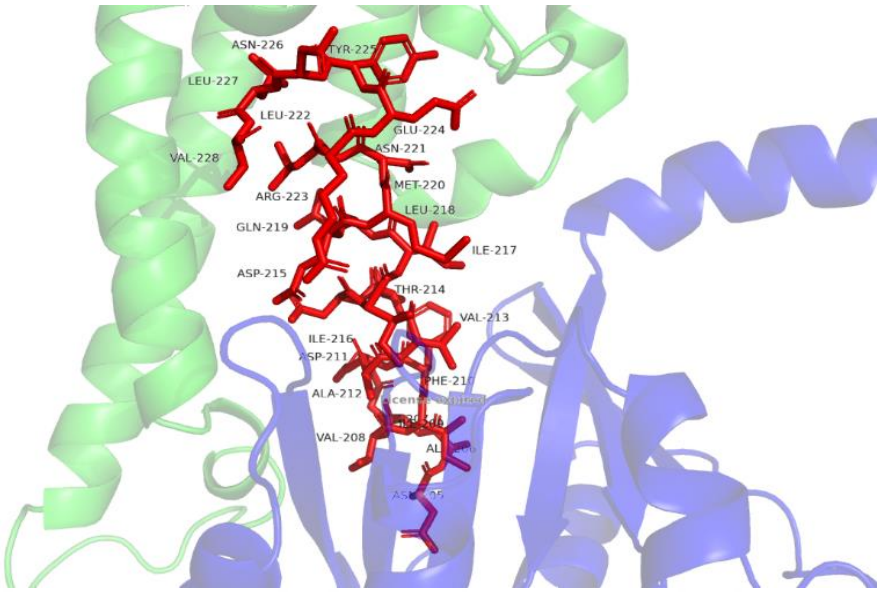


Figure 23. The amino acid of Gαo protein α5 helix (red), OXTR helices (green), Go alpha ras-like domain (Bleu)

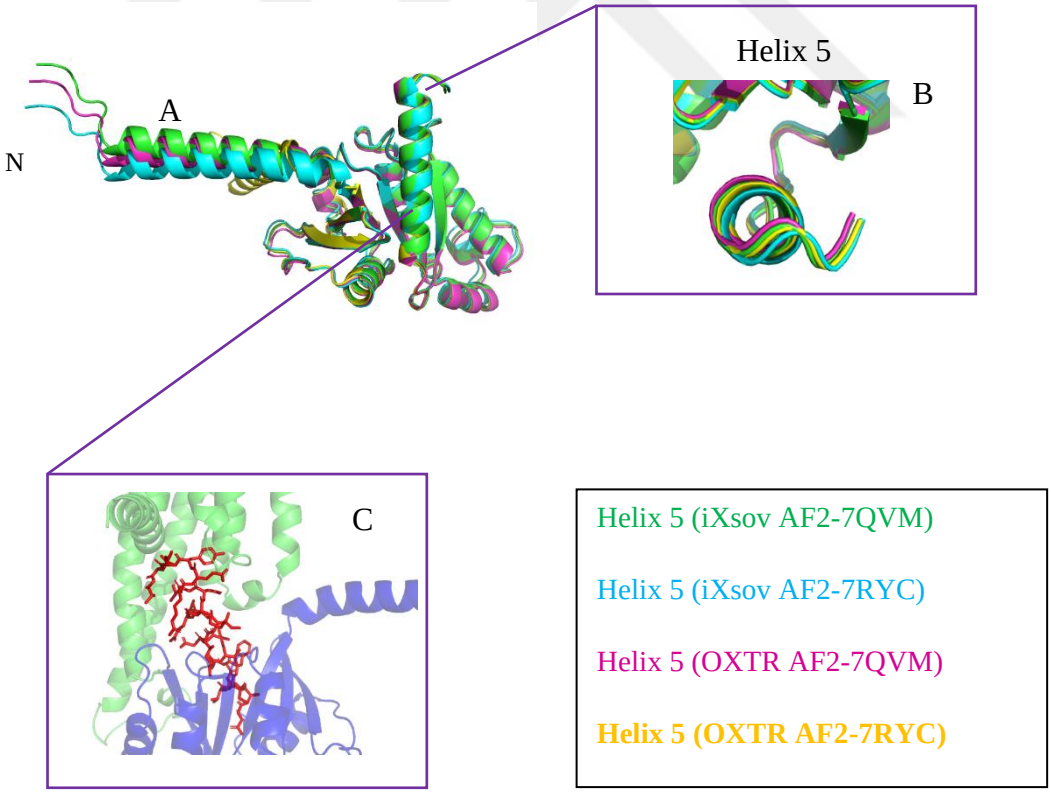


Figure 24. A-Superposition of GαO protein Ras-like domain; B-superposition of α5 helix; C-view showing the crack formed by α5 helix and helices of OXTR.

Chapter 5
Discussion and Conclusion

5.1 Discussion

AlphaFold2 represents the updated version of the AlphaFold system developed by DeepMind to enter the CASP14 competition (Flower & Hurley, 2021). It serves as an end-to-end solution designed to predict protein folding based on the amino acid sequence of the protein (*Groups Analysis: Zscores - CASP14*, n.d.). Despite its recent development, AlphaFold2 has already made a substantial impact on the industry by offering faster and more accessible protein folding prediction methods, as well as accelerating protein research discoveries (Jumper, et al., 2021). In AlphaFold2 the 7 helices of six GPCRs models show a very high level of prediction accuracy (pLDDT>90) per position, except for the N and C terminal and loops, which present lower prediction confidences (pLDDT<50) that could be attributed to high residue flexibility (Guo et al., 2022). This observation aligns with experimental evidence suggesting that loops remain unresolved due to their flexibility, reduced conservation and higher structural variance. It's also known that inactive GPCR structures are more common in the AF2 training set than active ones (Sala et al., 2023).

Despite its efficiency, AlphaFold2 has some limitations and drawbacks related to multiple sequence alignment (MSA). This alignment is limited by existing and available sequences in datasets. It can estimate structures that are similar to known or predicted structures based directly on what is observed or known (Skolnick et al., 2021). Another limitation is related to Protein Data Base (PDB). The AF2 algorithm predicts folding based only on those structures present in PDB, but may not correctly predict structures that occur under specific conditions, such as interactions with other proteins and domains (Lyu et al., 2023). This limitation is particularly evident for proteins with multiple native structures such as our case study (OXTR binding to cpGFP and G-alpha protein subunit). Moreover, the computational method's accuracy could not include new configurations and arrangements of amino acids that have not already been observed before in protein structure space (Marcu et al., 2022). Jumper et al. also confirmed that some predictions made by AlphaFold could not reach a high

level of accuracy. They observed that a poor prediction could be caused by the absence of total intra-domain structure together in some regions with very high oligomerization states (Jumper, et al., 2021). As such, the misorientation of the N-terminus helix in the G alpha protein within the OXTR AF2-7RYR model has impacted the outcomes of our MD simulation, resulting in few differences when compared to other models.

RMSF of the third intracellular loop (ICL3) presents large fluctuations, indicating its flexibility and involvement in dynamic interactions. Several studies suggested that significant changes in fluorescence are expected when cpGFP is inserted in ICL3 between TM5 and TM6 of GPCR through its N- and C- terminal. The cpGFP was inserted into the ICL3 region of the DRD biosensor (dLight & GRAB-DA) and GPR68 biosensor (iGlow) (Kim et al., 2022). This particular region of GPCRs is known to undergo substantial conformational changes upon activation triggered by a stimulus, as demonstrated in a study by Rasmussen et al. in 2011 and Ozkan et al. in 2021 (Rasmussen et al., 2011) (Ozkan et al., 2021). In Previous research, it has been proven experimentally that the insertion of cpGFP in ICL3 maximizes the interaction of ligand-receptor inducing conformational changes leading to an increase in the fluorescent intensity of the inserted cpGFP (Labouesse et al., 2020). Our sensor was designed based on these guiding principles.

RMSD analysis of the inactive state showed the conformation stability of the engineered fluorescent oxytocin receptor (iXsov), which was comparable to OXTR in its inactive state. This result suggests that cpGFP is not disrupting OXTR's structure in its rest state. However, recent studies have revealed that GPCRs are dynamic proteins with multiple conformational changes depending on ligand binding, signaling proteins, and the membrane environment (Wingler & Lefkowitz, 2020). These conformation changes demonstrate the allosteric communication that occurs as a result of oxytocin receptor binding to oxytocin hormone (Zou et al., 2019). Computational methods are valuable in helping to understand the new structure conformation of GPCRs and their interactions with ligands in the presence of different domains that are hard to extract by experimental methods (Srivastava et al., 2018). This study presents a pioneering exploration of the dynamic structure profile of cpGFP-derived fluorescent oxytocin receptor biosensor. Here, we employed MD simulation to describe the influence of the cpGFP domain on the structural conformation and the interaction dynamics between the OXTR and oxytocin hormone, particularly in the

presence of cpGFP. The analysis of the modeling process and the results of MD simulation is based on two factors: first, the selection of the template for homology modeling can significantly influence the quality of the outcomes and the conclusion derived therefrom (Tautermann et al., 2015). The unusual helix orientation shown in the OXTR AF2-7RYC model demonstrated changes observed in the MD simulation results compared to OXTR AF2-7QVM, emphasizing the importance of a suitable template for accurate modeling followed by credible analysis. The second factor is the limitation of MD simulation in isolating specific binding modes. The MD simulation can provide insight into the general dynamic of structures, conformational change, ligand binding, and protein folding. Particulary MD simulation can also predict how a structure will respond at an atomic level to perturbations such as cpGFP domains. However, it may not determine the relevant binding mode without additional experimental validation (Hollingsworth & Dror, 2018).

The activation mechanism of GPCRs is common knowledge. However, receptors exhibiting large-scale movements or occluding residues (as would be induced by the insertion of cpGFP) may have special alterations in G-protein-binding, which can be observed by MD simulations. Recent experimental studies showed that GPCRs can exist in a dynamic equilibrium and transit spontaneously between inactive and active conformation even in the absence of ligand, confirming that GPCRs are flexible and capable of adopting different structural states. The same study highlights that GPCRs could stabilize or trigger specific conformational changes depending on the presence of ligands and its efficiency in activating signaling pathways (Bhattacharya et al., 2008). Other genetically encoded fluorescent GPCR biosensors were experimentally shown to exhibit structural alterations between TM5 and TM6 of the biosensor after the ligand binding, leading to the rearrangement of the cyclic permuted fluorescent protein (cpFP) and subsequently causing significant conformation changes in the entire GPCR upon activation. These conformation changes in turn are able to increase the fluorescent intensity of the inserted cpGFP in the biosensor (Kim et al., 2022). Significant conformational changes have been identified through RMSD analysis in our OXTR-fluorescent biosensor (iXsov). Moreover, high RMS fluctuation analysis was observed in the region of the cpGFP domain within the ICL3 between helices TM5 and TM6 of the OXTR. The conformational changes observed in iXsov provide insight into the specific

mechanism and expand our understanding of the activation of fluorescent GPCR. Interestingly, we highlight the relationship between ligand binding, structural dynamics, and fluorescence modulation in GPCR biosensor design. Zou et al. confirmed that changes occurring in or outside the pocket or any area in the protein could lead to small or large effects within the protein (Zou et al., 2019). Once the oxytocin hormone is attached to the receptor, it creates allosteric communication that shapes and active changes that trigger an intracellular signal cascade. The heterotrimeric G protein also plays a crucial role in these downstream by switching the signal from the extracellular to the intracellular environment, and the prevention of its binding could be one possible hypothesis for the absence of downstream signaling in GPCR biosensors with cpGFP inserts (Y. Liu et al., 2021).

Yann et al found that complexes involving OXTR and the frequently used engineered mini-$G_{s/q}$ were not stable and disassembled upon rapid freezing. Despite OTR primarily activating Gq-based signaling pathways, previous research has shown interactions with Go and Gi, although not with Gs. This observation led the researchers to hypothesize that the instability of the OTR:mini-Gs/q complex might stem from unfavorable interactions between OTR and the Gs domain, with the interactions of the Gq α5 helix insufficient to stabilize the complex. To address this issue and maximize stability in the active state of OTR, they proposed designing a G protein chimera using mini-Go as a foundation to enhance the stability and functionality of the OTR-G protein complex, offering insights into potential mechanisms governing receptor-G protein interactions (Waltenspühl et al., 2022).

MM-PBSA Binding energy decomposition analysis has been calculated to determine the energy contribution of each residue to the overall binding energy. This analysis can extract the crucial residues contributing to the OXTR-OT and iXsov-OT. All nine amino acids of OT, participate in ligand binding as previously proved by (Waltenspühl et al., 2022) highlighting the crucial role of these residues in mediating the binding process. Specific interactions between residues were observed in the OXTR-OT binding mode, drawing a comparison with findings from other research studies by Waltenspühl et al., 2022, reinforcing the validity of our observations. For instance, the interaction of GLN4 in OT with GLN295 stabilizes the cyclic ring position via a hydrogen bond, results published in (Meyerowitz et al., 2022) and (Waltenspühl et al., 2022), a relationship also noted in the OXTR AF2-7QVM and

AF2-7RYC structures, as well as in iXsov AF2-7QVM and AF2-7RYC (GLN546), highlighting their consistent role in binding across different structural contexts. Furthermore, TYR2 was found to interact deeply within the pocket, forming hydrophobic interactions and hydrogen bonds with specific residues such as GLN171, GLN92, PHE291, and L316, as referenced in Waltenspühl et al., 2022 research, and ILE3 occupied a hydrophobic pocket formed by TYR200, ILE201 and ILE204 and TRP188 of ECL2 capping the ligand (Meyerowitz et al., 2022). These residues contribute to both wild-type OXTR and iXsov, further emphasizing their importance in ligand binding. Specifically, the OT interacts with the residue PHE291 at the bottom of the binding pocket. Cys1 also played a crucial role in stabilizing the OT ring conformation through a disulfide bond with Cys6 and interacting with a polar cluster of residues (GLN96, LYS116, and GLN119) (Waltenspühl et al., 2022), similar residue contributions are observed in both wild-type OXTR and iXsov, indicating the stability and functionality of iXsov. Interestingly, the analysis shows that the cpGFP may not perturb the binding pocket, and the same residues continue to play a crucial role in the binding process, suggesting the stability and functionality of iXsov.

Similar to previous research on the G protein complex structures, in the OXTR-G protein binding (Waltenspühl et al., 2022), CCR6-Go complex (Wasilko et al., 2020) and α_2 adrenergic receptors coupled to Go protein (Yuan et al., 2020), the C terminus of the α5 helix of Go adopt a loop confirmation situated within a crevice formed by specific helices on the intracellular side of the activated receptor, where key hydrogen bonding interactions occur between G protein helix 5 C-terminal residues and receptor residues. This arrangement is notably reported and observed in our four OXTR/iXsov-Gα_O protein complexes. The MM-PBSA results predict residue contributions in the OXTR-Gα_O interaction across the four complexes highlighting preserved residues, particularly in Gα_O. helix 5 and OXTR helices 1, 3, 5, 6, and 7 (Waltenspühl et al., 2022). Moreover, The interactions between α_{2B}AR and Go are distributed extensively throughout ICL1, ICL2, TM5, and TM6 (Yuan et al., 2020). These results aligned with the experimental study conducted by Waltenspühl et al in 2022 in the schematic drawing of direct interactions between OXTR and the α5 helix of Gα_O. This interaction include Go α5 helix (VAL354, LEU353, ASN352, TYR351, GLU350, LEU348, ASN347, GLY345, LEU344 and ILE343) and OXTR residues (ARG73 and LEU74 (TM1) , ASP136, ARG137, ALA140, ILE141 and PRO144

(TM3), ILE227 and ASN230 (TM5), ARG272 (TM6) and HIS335 (TM7)). The arrangement observed in the four OXTR/iXsov: $G\alpha_O$ protein complexes and its alignment with experimental results indicates a consistent mode of interaction between OXTR helices and the α5 helix of $G\alpha_O$, suggesting a stable and reliable binding interface. Overall, this analysis provides valuable insights into the molecular mechanisms underlying OXTR- $G\alpha_O$ protein interactions, offering a basis for understanding the structural and functional aspects of receptor-G protein signaling pathways.

The absence of significant differences in The RMSD, RMSF H-bonds, and MM-PBSA decomposition analysis results between the G alpha dynamic structure in the OXTR and iXsov, suggests that the presence of the cpGFP domain does not impact significantly the conformation and the binding mode of the G-alpha protein within the GPCR complex. Moreover, the G-alpha domain does not interfere with the cpGFP domain (absence of steric clash) during the MD simulation movie observed using VMD. This finding serves to validate the strategic design of the engineered protein reporter. We are aware that studying the whole mechanism of the heterotrimeric G protein activation, including the three subunits (α, β, and γ) would be more representative. However, the limitation of Alphafold2 and the computational cost of our simulation prevent us. Several previous computational research investigated how G-alpha is activated. Our model serves as the first research that targets G-alpha in the presence of cpGFR within a GPCR structure (genetically engineered fluorescent reporter). This simulation is able to suggest new experimental testable hypotheses to contribute the the progress of experimental work, including the loss of downstream signal, previously experimentally confirmed by Ozkan et al 2021.

The stabilization observed in AF2-7QVM models during the simulation suggests a more favorable and energetically stable conformation in the presence of the ligand. The results of the calculation of the binding free energy using the g_mmpbsa tool between the protein (OXTR) and ligand (OT), showed the distinct stability levels between the two templates, concluding that the stability of models based on the 7QVM complex is higher compared to those based on 7RYC template. In addition, it is worth noting that the inclusion of cpGFP in the OXTR structure (iXsov) does not significantly alter the stability state at the energetic level when compared to the wild

type (OXTR). This conclusion will improve computational predicted models for future protein-ligand interactions.

In summary, we have produced a design candidate for a new real-time fluorescent biosensor, designed an appropriate sequence and created a homology modeling of its structure based on templates, and performed MD simulations by taking into account both homology models as the initial condition to understand the importance of correct modeling.

Genetically encoded fluorescent reporters are an effective means to visualize the activity of GPCR in real-time and in a stimulus-dependent manner. The true dynamic behavior of this fluorescent reporter in contact with their stimulus requires follow-up studies to supplement molecular modeling and MD simulations. However, *in silico* results have demonstrated that the design is likely to be comparable to previously reported GPCR biosensors in function, acting as an effective proof of concept. In addition, our integrated approach is essential for deriving general rules that can guide the future application of MD simulations in the design of GPCR-fluorescent reporters.

5.2 Conclusion

The oxytocin receptor (OXTR) has received significant interest due to its impact on social behavior, particularly in the context of autism spectrum disorders (ASD). The development of engineered fluorescent reporters, such as the OXTR-derived fluorescent biosensor iXsov, could open new and exciting possibilities for examining biological processes at the molecular level. The genetic expression of these sensors facilitates real-time monitoring of cell signaling, neural activity, and enzyme function across a wide spectrum of living systems ranging from cultured cells to entire organisms. In this study, we conducted a comprehensive molecular dynamic simulation of the OXTR-cpGFP biosensor (iXsov), offering for the first time a dynamic structure description of this GPCR-encoded fluorescent reporter. Through simulations, we gained insights into the interactions of the transmembrane receptor with the lipid bilayer, cpGFP, and G-alpha protein subunit in both active and inactive states, predicting molecular collisions with solvent molecules and providing theoretical hypotheses for experimental observations in fluorescence emissions and downstream signaling pathways. Our research highlights the importance of computational tools like AlphaFold2 and CHARMM-GUI in folding protein targets and building lipid membranes for MD simulations in drug screening and discovery efforts. The structural dynamic profiles extracted from the MD simulation of iXsov in simulated lipid membranes provide a foundation for further studies in pharmacological and physiological research, including investigating the role of genetic variations in OXTR related to ASD at the organism level. Moving forward, our approach holds promise for studying genetically engineered fluorescent reporters using other GPCRs, enabling further exploration of conformational changes, optimization of fluorescent reporter design strategies, enhancing fluorescence sensitivity, and discovering the reason behind unusual downstream signaling behaviors. Thus, the combination of computational simulations and experimental validations remains crucial for advancing our understanding of GPCR-mediated signaling and developing innovative biosensors in diverse biomedical applications.

REFERENCES

Acharya, K., Koirala, R. Prasad, & Pantha, N. (2021). Diffusion of oxytocin in water: A molecular dynamics study. *BIBECHANA*, *18*, 108–117. https://doi.org/10.3126/bibechana.v18i1.29316

Aimeur, S., Fas, B. A., Serfaty, X., Santuz, H., Sacquin-Mora, S., Bizouarn, T., Taly, A., & Baciou, L. (2024). *Combining computational biology and experimental knowledge to draw a structural profile of the active membrane-assembled NADPH oxidase complex* (p. 2024.02.19.579638). bioRxiv. https://doi.org/10.1101/2024.02.19.579638

Al-Ali, Z., Yasseen, A. A., Al-Dujailli, A., Al-Karaqully, A. J., McAllister, K. A., & Jumaah, A. S. (2022). The oxytocin receptor gene polymorphism rs2268491 and serum oxytocin alterations are indicative of autism spectrum disorder: A case-control paediatric study in Iraq with personalized medicine implications. *PloS One*, *17*(3), e0265217. https://doi.org/10.1371/journal.pone.0265217

Antobreh, G., Enyedy, I., & Ravna, A. W. (2017, December 13). *Molecular modeling and docking studies of the oxytocin receptor* (London, UK) [Research-article]. Https://Doi.Org/10.4155/Fmc-2017-0078; Future Science Ltd London, UK. https://doi.org/10.4155/fmc-2017-0078

Arnold, M. E., Dostmann, W. R., Martin, J., Previs, M. J., Palmer, B., LeWinter, M., & Meyer, M. (2021). SERCA2a-phospholamban interaction monitored by an interposed circularly permutated green fluorescent protein. *American Journal of Physiology. Heart and Circulatory Physiology*, *320*(6), H2188–H2200. https://doi.org/10.1152/ajpheart.00858.2020

Bhattacharya, S., Hall, S. E., Li, H., & Vaidehi, N. (2008). Ligand-Stabilized Conformational States of Human β2 Adrenergic Receptor: Insight into G-Protein-Coupled Receptor Activation. *Biophysical Journal*, *94*(6), 2027–2042. https://doi.org/10.1529/biophysj.107.117648

Borroto-Escuela, D. O., Cuesta-Marti, C., Lopez-Salas, A., Chruścicka-Smaga, B., Crespo-Ramírez, M., Tesoro-Cruz, E., Palacios-Lagunas, D. A., Perez de la Mora, M., Schellekens, H., & Fuxe, K. (2022). The oxytocin receptor represents a key hub in the GPCR heteroreceptor network: Potential relevance

for brain and behavior. *Frontiers in Molecular Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389/fnmol.2022.1055344

Bruno, A., & Costantino, G. (2012). Molecular Dynamics Simulations of G Protein-Coupled Receptors. *Molecular Informatics*, *31*(3–4), 222–230. https://doi.org/10.1002/minf.201100138

Changeux, J.-P., & Edelstein, S. (2011). Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biology Reports*, *3*, 19. https://doi.org/10.3410/B3-19

Chmiela, S., Sauceda, H. E., Müller, K.-R., & Tkatchenko, A. (2018). Towards exact molecular dynamics simulations with machine-learned force fields. *Nature Communications*, *9*(1), 3887. https://doi.org/10.1038/s41467-018-06169-2

COARSE-GRAINING WITH THE RELATIVE ENTROPY - Shell—2016—Advances in Chemical Physics—Wiley Online Library. (n.d.). Retrieved March 6, 2024, from https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119290971.ch5

Ennist, N. M., Stayrook, S. E., Dutton, P. L., & Moser, C. C. (2022). Rational design of photosynthetic reaction center protein maquettes. *Frontiers in Molecular Biosciences*, *9*. https://doi.org/10.3389/fmolb.2022.997295

Escribá, P. V., González-Ros, J. M., Goñi, F. M., Kinnunen, P. K. J., Vigh, L., Sánchez-Magraner, L., Fernández, A. M., Busquets, X., Horváth, I., & Barceló-Coblijn, G. (2008). Membranes: A meeting point for lipids, proteins and therapies. *Journal of Cellular and Molecular Medicine*, *12*(3), 829–875. https://doi.org/10.1111/j.1582-4934.2008.00281.x

Feng, S., Park, S., Choi, Y. K., & Im, W. (2023). CHARMM-GUI Membrane Builder: Past, Current, and Future Developments and Applications. *Journal of Chemical Theory and Computation*, *19*(8), 2161–2185. https://doi.org/10.1021/acs.jctc.2c01246

Fernandes, H. S., Sousa, S. F., & Cerqueira, N. M. F. S. A. (2019). VMD Store-A VMD Plugin to Browse, Discover, and Install VMD Extensions. *Journal of Chemical Information and Modeling*, *59*(11), 4519–4523. https://doi.org/10.1021/acs.jcim.9b00739

Flock, T., Ravarani, C. N. J., Sun, D., Venkatakrishnan, A. J., Kayikci, M., Tate, C. G., Veprintsev, D. B., & Babu, M. M. (2015). Universal allosteric mechanism for Gα activation by GPCRs. *Nature*, *524*(7564), Article 7564. https://doi.org/10.1038/nature14663

Flower, T. G., & Hurley, J. H. (2021). Crystallographic molecular replacement using an in silico-generated search model of SARS-CoV-2 ORF8. *Protein Science: A Publication of the Protein Society*, *30*(4), 728–734. https://doi.org/10.1002/pro.4050

Frehner, S. S., Dooley, K. T., Palumbo, M. C., Smith, A. L., Goodman, M. M., Bales, K. L., & Freeman, S. M. (2022). Effect of sex and autism spectrum disorder on oxytocin receptor binding and mRNA expression in the dopaminergic pars compacta of the human substantia nigra. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *377*(1858), 20210118. https://doi.org/10.1098/rstb.2021.0118

Friedlander, E., Yirmiya, N., Laiba, E., Harel- Gadassi, A., Yaari, M., Feldstein, O., Mankuta, D., & Israel, S. (2019). Cumulative Risk of the Oxytocin Receptor Gene Interacts with Prenatal Exposure to Oxytocin Receptor Antagonist to Predict Children's Social Communication Development. *Autism Research*, *12*. https://doi.org/10.1002/aur.2111

Greif, G. J., Sodickson, D. L., Bean, B. P., Neer, E. J., & Mende, U. (2000). Altered regulation of potassium and calcium channels by GABA(B) and adenosine receptors in hippocampal neurons from mice lacking Galpha(o). *Journal of Neurophysiology*, *83*(2), 1010–1018. https://doi.org/10.1152/jn.2000.83.2.1010

Groups Analysis: Zscores—CASP14. (n.d.). Retrieved March 25, 2024, from https://predictioncenter.org/casp14/zscores_final.cgi

Guo, H.-B., Perminov, A., Bekele, S., Kedziora, G., Farajollahi, S., Varaljay, V., Hinkle, K., Molinero, V., Meister, K., Hung, C., Dennis, P., Kelley-Loughnane, N., & Berry, R. (2022). AlphaFold2 models indicate that protein sequence determines both structure and dynamics. *Scientific Reports*, *12*(1), 10696. https://doi.org/10.1038/s41598-022-14382-9

Hall, R., Dixon, T., & Dickson, A. (2020). On Calculating Free Energy Differences Using Ensembles of Transition Paths. *Frontiers in Molecular Biosciences*, *7*. https://doi.org/10.3389/fmolb.2020.00106

Hollingsworth, S. A., & Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron*, *99*(6), 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011

Horie, K., Inoue, K., Suzuki, S., Adachi, S., Yada, S., Hirayama, T., Hidema, S., Young, L. J., & Nishimori, K. (2019). Oxytocin receptor knockout prairie

voles generated by CRISPR/Cas9 editing show reduced preference for social novelty and exaggerated repetitive behaviors. *Hormones and Behavior*, *111*, 60–69. https://doi.org/10.1016/j.yhbeh.2018.10.011

Hsu, Y.-Y., Resto Irizarry, A. M., Fu, J., & Liu, A. P. (2023). Mechanosensitive Channel-Based Optical Membrane Tension Reporter. *ACS Sensors*, *8*(1), 12–18. https://doi.org/10.1021/acssensors.2c01921

Inoue, Y. U., Miwa, H., Hori, K., Kaneko, R., Morimoto, Y., Koike, E., Asami, J., Kamijo, S., Yamada, M., Hoshino, M., & Inoue, T. (2022). Targeting Neurons with Functional Oxytocin Receptors: A Novel Set of Simple Knock-In Mouse Lines for Oxytocin Receptor Visualization and Manipulation. *eNeuro*, *9*(1), ENEURO.0423-21.2022. https://doi.org/10.1523/ENEURO.0423-21.2022

Jacob, S., Brune, C. W., Carter, C. S., Leventhal, B. L., Lord, C., & Cook, E. H. (2007). Association of the oxytocin receptor gene (OXTR) in Caucasian children and adolescents with autism. *Neuroscience Letters*, *417*(1), 6–9. https://doi.org/10.1016/j.neulet.2007.02.001

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Applying and improving AlphaFold at CASP14. *Proteins*, *89*(12), 1711–1721. https://doi.org/10.1002/prot.26257

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), Article 7873. https://doi.org/10.1038/s41586-021-03819-2

Jurek, B., & Meyer, M. (2020). Anxiolytic and Anxiogenic? How the Transcription Factor MEF2 Might Explain the Manifold Behavioral Effects of Oxytocin. *Frontiers in Endocrinology*, *11*, 528401. https://doi.org/10.3389/fendo.2020.00186

Kim, H., Baek, I.-Y., & Seong, J. (2022). Genetically encoded fluorescent biosensors for GPCR research. *Frontiers in Cell and Developmental Biology*, *10*. https://doi.org/10.3389/fcell.2022.1007893

Kise, R., Ono, Y., Kawakami, K., & Inoue, A. (2021). Toward understanding the role of G-protein signaling. *Current Opinion in Endocrine and Metabolic Research*, *16*, 51–55. https://doi.org/10.1016/j.coemr.2020.08.006

Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O., & Shaw, D. E. (2009). Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology*, *19*(2), 120–127. https://doi.org/10.1016/j.sbi.2009.03.004

Kreitz, J., Friedrich, M. J., Guru, A., Lash, B., Saito, M., Macrae, R. K., & Zhang, F. (2023). Programmable protein delivery with a bacterial contractile injection system. *Nature*, *616*(7956), 357–364. https://doi.org/10.1038/s41586-023-05870-7

Kumari, R., Kumar, R., Open Source Drug Discovery Consortium, & Lynn, A. (2014). g_mmpbsa—A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, *54*(7), 1951–1962. https://doi.org/10.1021/ci500020m

Kutzner, C., Kniep, C., Cherian, A., Nordstrom, L., Grubmüller, H., de Groot, B. L., & Gapsys, V. (2022). GROMACS in the Cloud: A Global Supercomputer to Speed Up Alchemical Drug Design. *Journal of Chemical Information and Modeling*, *62*(7), 1691–1711. https://doi.org/10.1021/acs.jcim.2c00044

Labouesse, M. A., Cola, R. B., & Patriarchi, T. (2020). GPCR-Based Dopamine Sensors—A Detailed Guide to Inform Sensor Choice for In Vivo Imaging. *International Journal of Molecular Sciences*, *21*(21), Article 21. https://doi.org/10.3390/ijms21218048

Leonzino, M., Ponzoni, L., Braida, D., Gigliucci, V., Busnelli, M., Ceresini, I., Duque-Wilckens, N., Nishimori, K., Trainor, B. C., Sala, M., & Chini, B. (2019). Impaired approach to novelty and striatal alterations in the oxytocin receptor deficient mouse model of autism. *Hormones and Behavior*, *114*, 104543. https://doi.org/10.1016/j.yhbeh.2019.06.007

Lin, Y.-T., & Hsu, K.-S. (2018). Oxytocin receptor signaling in the hippocampus: Role in regulating neuronal excitability, network oscillatory activity, synaptic plasticity and social memory. *Progress in Neurobiology*, *171*, 1–14. https://doi.org/10.1016/j.pneurobio.2018.10.003

Liu, N., Wang, Y., Li, T., & Feng, X. (2021). G-Protein Coupled Receptors (GPCRs): Signaling Pathways, Characterization, and Functions in Insect Physiology and

Toxicology. *International Journal of Molecular Sciences*, *22*(10), Article 10. https://doi.org/10.3390/ijms22105260

Liu, W., Liu, C., Ren, P.-G., Chu, J., & Wang, L. (2022). An Improved Genetically Encoded Fluorescent cAMP Indicator for Sensitive cAMP Imaging and Fast Drug Screening. *Frontiers in Pharmacology*, *13*, 902290. https://doi.org/10.3389/fphar.2022.902290

Liu, Y., Wang, X., Dong, D., Guo, L., Dong, X., Leng, J., Zhao, B., Guo, Y.-D., & Zhang, N. (2021). Research Advances in Heterotrimeric G-Protein α Subunits and Uncanonical G-Protein Coupled Receptors in Plants. *International Journal of Molecular Sciences*, *22*(16), 8678. https://doi.org/10.3390/ijms22168678

Lyu, J., Kapolka, N., Gumpper, R., Alon, A., Wang, L., Jain, M. K., Barros-Álvarez, X., Sakamoto, K., Kim, Y., DiBerto, J., Kim, K., Tummino, T. A., Huang, S., Irwin, J. J., Tarkhanova, O. O., Moroz, Y., Skiniotis, G., Kruse, A. C., Shoichet, B. K., & Roth, B. L. (2023). *AlphaFold2 structures template ligand discovery* (p. 2023.12.20.572662). bioRxiv. https://doi.org/10.1101/2023.12.20.572662

Marcu, Ş.-B., Tăbîrcă, S., & Tangney, M. (2022). An Overview of Alphafold's Breakthrough. *Frontiers in Artificial Intelligence*, *5*. https://doi.org/10.3389/frai.2022.875587

Maruta, N., Trusov, Y., Jones, A. M., & Botella, J. R. (2021). Heterotrimeric G Proteins in Plants: Canonical and Atypical Gα Subunits. *International Journal of Molecular Sciences*, *22*(21), 11841. https://doi.org/10.3390/ijms222111841

Marvin, J. S., Shimoda, Y., Magloire, V., Leite, M., Kawashima, T., Jensen, T. P., Kolb, I., Knott, E. L., Novak, O., Podgorski, K., Leidenheimer, N. J., Rusakov, D. A., Ahrens, M. B., Kullmann, D. M., & Looger, L. L. (2019). A genetically encoded fluorescent sensor for in vivo imaging of GABA. *Nature Methods*, *16*(8), Article 8. https://doi.org/10.1038/s41592-019-0471-2

Meyer, M., Jurek, B., Alfonso-Prieto, M., Ribeiro, R., Milenkovic, V. M., Winter, J., Hoffmann, P., Wetzel, C. H., Giorgetti, A., Carloni, P., & Neumann, I. D. (2022). Structure-function relationships of the disease-linked A218T oxytocin receptor variant. *Molecular Psychiatry*, *27*(2), Article 2. https://doi.org/10.1038/s41380-021-01241-8

Meyerowitz, J. G., Robertson, M. J., Barros-Álvarez, X., Panova, O., Nwokonko, R. M., Gao, Y., & Skiniotis, G. (2022). The oxytocin signaling complex reveals a molecular switch for cation dependence. *Nature Structural & Molecular Biology*, *29*(3), Article 3. https://doi.org/10.1038/s41594-022-00728-4

Mignocchi, N., Krüssel, S., Jung, K., Lee, D., & Kwon, H.-B. (2020). *Development of a genetically-encoded oxytocin sensor* (p. 2020.07.14.202598). bioRxiv. https://doi.org/10.1101/2020.07.14.202598

Moerkerke, M., Bonte, M.-L., Daniels, N., Chubar, V., Alaerts, K., Steyaert, J., & Boets, B. (2021a). Oxytocin receptor gene (OXTR) DNA methylation is associated with autism and related social traits – A systematic review. *Research in Autism Spectrum Disorders*, *85*, 101785. https://doi.org/10.1016/j.rasd.2021.101785

Moerkerke, M., Bonte, M.-L., Daniels, N., Chubar, V., Alaerts, K., Steyaert, J., & Boets, B. (2021b). Oxytocin receptor gene (OXTR) DNA methylation is associated with autism and related social traits – A systematic review. *Research in Autism Spectrum Disorders*, *85*, 101785. https://doi.org/10.1016/j.rasd.2021.101785

Nagle, J. F., & Tristram-Nagle, S. (2000). Structure of lipid bilayers. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes*, *1469*(3), 159–195. https://doi.org/10.1016/S0304-4157(00)00016-2

Okashah, N., Wan, Q., Ghosh, S., Sandhu, M., Inoue, A., Vaidehi, N., & Lambert, N. A. (2019). Variable G protein determinants of GPCR coupling selectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(24), 12054–12059. https://doi.org/10.1073/pnas.1905993116

Patriarchi, T., Cho, J. R., Merten, K., Howe, M. W., Marley, A., Xiong, W.-H., Folk, R. W., Broussard, G. J., Liang, R., Jang, M. J., Zhong, H., Dombeck, D., von Zastrow, M., Nimmerjahn, A., Gradinaru, V., Williams, J. T., & Tian, L. (2018). Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science*, *360*(6396), eaat4422. https://doi.org/10.1126/science.aat4422

Patrone, M., Cammarota, E., Berno, V., Tornaghi, P., Mazza, D., & Degano, M. (2020). Combinatorial allosteric modulation of agonist response in a self-interacting G-protein coupled receptor. *Communications Biology*, *3*(1), Article 1. https://doi.org/10.1038/s42003-020-0752-4

Rokicki, J., Kaufmann, T., de Lange, A.-M. G., van der Meer, D., Bahrami, S., Sartorius, A. M., Haukvik, U. K., Steen, N. E., Schwarz, E., Stein, D. J., Nærland, T., Andreassen, O. A., Westlye, L. T., & Quintana, D. S. (2022). Oxytocin receptor expression patterns in the human brain across development. *Neuropsychopharmacology*, *47*(8), Article 8. https://doi.org/10.1038/s41386-022-01305-5

Sala, D., Hildebrand, P. W., & Meiler, J. (2023). Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties. *Frontiers in Molecular Biosciences*, *10*, 1121962. https://doi.org/10.3389/fmolb.2023.1121962

Sanchez-Reyes, O. B., Zilberg, G., McCorvy, J. D., & Wacker, D. (2023). Molecular insights into GPCR mechanisms for drugs of abuse. *Journal of Biological Chemistry*, *299*(9). https://doi.org/10.1016/j.jbc.2023.105176

Sato, M., Nakai, N., Fujima, S., Choe, K. Y., & Takumi, T. (2023). Social circuits and their dysfunction in autism spectrum disorder. *Molecular Psychiatry*, 1–13. https://doi.org/10.1038/s41380-023-02201-0

Schlick, T., Barth, E., & Mandziuk, M. (1997). Biomolecular dynamics at long timesteps: Bridging the timescale gap between simulation and experimentation. *Annual Review of Biophysics and Biomolecular Structure*, *26*, 181–222. https://doi.org/10.1146/annurev.biophys.26.1.181

Skolnick, J., Gao, M., Zhou, H., & Singh, S. (2021). AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function. *Journal of Chemical Information and Modeling*, *61*. https://doi.org/10.1021/acs.jcim.1c01114

Srivastava, A., Nagai, T., Srivastava, A., Miyashita, O., & Tama, F. (2018). Role of Computational Methods in Going beyond X-ray Crystallography to Explore Protein Structure and Dynamics. *International Journal of Molecular Sciences*, *19*(11), 3401. https://doi.org/10.3390/ijms19113401

Sun, X., Singh, S., Blumer, K. J., & Bowman, G. R. (n.d.). Simulation of spontaneous G protein activation reveals a new intermediate driving GDP unbinding. *eLife*, *7*, e38465. https://doi.org/10.7554/eLife.38465

Tautermann, C. S., Seeliger, D., & Kriegl, J. M. (2015). What can we learn from molecular dynamics simulations for GPCR drug design? *Computational and*

Structural Biotechnology Journal, *13*, 111–121. https://doi.org/10.1016/j.csbj.2014.12.002

Terteci-Popescu, A.-E., & Beu, T. A. (2022). Branched polyethyleneimine: CHARMM force field and molecular dynamics simulations. *Journal of Computational Chemistry*, *43*(31), 2072–2083. https://doi.org/10.1002/jcc.27005

Tuteja, N. (2009). Signaling through G protein coupled receptors. *Plant Signaling & Behavior*, *4*(10), 942–947. https://doi.org/10.4161/psb.4.10.9530

Uvnäs-Moberg, K. (2024). The physiology and pharmacology of oxytocin in labor and in the peripartum period. *American Journal of Obstetrics and Gynecology*, *230*(3, Supplement), S740–S758. https://doi.org/10.1016/j.ajog.2023.04.011

Waltenspühl, Y., Ehrenmann, J., Vacca, S., Thom, C., Medalia, O., & Plückthun, A. (2022). Structural basis for the activation and ligand recognition of the human oxytocin receptor. *Nature Communications*, *13*(1), 4153. https://doi.org/10.1038/s41467-022-31325-0

Wasilko, D. J., Johnson, Z. L., Ammirati, M., Che, Y., Griffor, M. C., Han, S., & Wu, H. (2020). Structural basis for chemokine receptor CCR6 activation by the endogenous protein ligand CCL20. *Nature Communications*, *11*(1), 3031. https://doi.org/10.1038/s41467-020-16820-6

Wicky, B. I. M., Milles, L. F., Courbet, A., Ragotte, R. J., Dauparas, J., Kinfu, E., Tipps, S., Kibler, R. D., Baek, M., DiMaio, F., Li, X., Carter, L., Kang, A., Nguyen, H., Bera, A. K., & Baker, D. (2022). Hallucinating symmetric protein assemblies. *Science*, *378*(6615), 56–61. https://doi.org/10.1126/science.add1964

Wingler, L. M., & Lefkowitz, R. J. (2020). Conformational basis of G protein-coupled receptor signaling versatility. *Trends in Cell Biology*, *30*(9), 736–747. https://doi.org/10.1016/j.tcb.2020.06.002

Yang, D., Zhou, Q., Labroska, V., Qin, S., Darbalaei, S., Wu, Y., Yuliantie, E., Xie, L., Tao, H., Cheng, J., Liu, Q., Zhao, S., Shui, W., Jiang, Y., & Wang, M.-W. (2021). G protein-coupled receptors: Structure- and function-based drug discovery. *Signal Transduction and Targeted Therapy*, *6*(1), Article 1. https://doi.org/10.1038/s41392-020-00435-w

Yuan, D., Liu, Z., Kaindl, J., Maeda, S., Zhao, J., Sun, X., Xu, J., Gmeiner, P., Wang, H.-W., & Kobilka, B. K. (2020). Activation of the α2B adrenoceptor by the

sedative sympatholytic dexmedetomidine. *Nature Chemical Biology*, *16*(5), 507–512. https://doi.org/10.1038/s41589-020-0492-2

Yun, J., Baldini, M., Chowdhury, R., & Mukherjee, A. (2022). Designing Protein-Based Probes for Sensing Biological Analytes with Magnetic Resonance Imaging. *Analysis & Sensing*, *2*(5), e202200019. https://doi.org/10.1002/anse.202200019

Zou, Y., Ewalt, J., & Ng, H.-L. (2019). Recent Insights from Molecular Dynamics Simulations for G Protein-Coupled Receptor Drug Discovery. *International Journal of Molecular Sciences*, *20*(17), 4237. https://doi.org/10.3390/ijms20174237