University of West Attica

School of Engineering

Department of Biomedical Engineering

MSc program "Biomedical Engineering and Technology"

Study of the orphan GPR21 towards therapeutics for metabolic disorders

SALOMI SKARMALIORAKI

Registration Number: bmet04

Supervisor

Minos Matsoukas, Assistant Professor, Department of Biomedical Engineering, University of West Attica

Athens 14/ 10 / 2024

The Three-Member Examination Committee

Supervisor		
Minos Matsoukas Assistant Professor, Department of Biomedical Engineering, University of West Attica	Manolis Athanasiadis Assistant Professor, Department of Biomedical Engineering, University of West Attica	Spiros Kostopoulos Associate Professor, Department of Biomedical Engineering, University of West Attica
MINOS TIMOTHEOS MATSOUKAS Digitally signed by MINOS TIMOTHEOS MATSOUKAS Date: 2024.10.15 13:08:22 +03'00'	EMMANOUIL ATHANASIADIS 2024.10.16 17:53:54 +03'00'	Digitally signed by Spyridon Kostopoulos Date: 2024.10.17 01:00:17 +03'00'
Signature	Signature	Signature

DECLARATION BY THE AUTHOR OF THE DIPLOMA THESIS

The signatory Salomi Skarmalioraki of Konstantinos and Mina, with registration number bmet04, student of the MSc Program “Biomedical Engineering and Technology” of the University of West Attica, I declare responsibly that:

"I am the author of this Diploma Thesis and any help I had for its preparation is fully recognized and referenced. Also, any sources from which I have used data, ideas, or words, whether exact or paraphrased, are listed in their entirety, with full reference to the authors, the publisher, or the journal, including any sources that may have been used by the internet. I also certify that this work has been written exclusively by me and is a product of intellectual property of both myself and the University of West Attica.

Violation of my above academic responsibility is an essential reason for the revocation of my diploma ".

Date: **14**/ **10** / **2024**

Signature

CONTENTS

Abstract

Since G-Protein Coupled Receptors (GPCRs) are engaged in a variety of signal transduction pathways in many pathophysiological situations, they are important targets for the development of novel drugs. GPR21 is a constitutively active orphan GPCR, which means that it remains in an active state even in the absence of an intrinsic ligand. GPR21 is involved in insulin sensitivity regulation. This suggests that it has potential as a therapeutic target for metabolic diseases, including type 2 diabetes. The small molecules GRA2 and UCSF924 have been found to interact with GPR21, however their binding mechanisms still remain unclear. Here we propose interaction scenarios between these ligands and GPR21, using homology modelling and molecular docking methods. Our results show that models with refined extracellular loop 2 (ECL2) regions, especially the ones where the whole loop or immersed region (residues 169-178) is refined, display enhanced druggability and better docking outcomes with GRA2 and UCSF924. In the immersed region and whole loop refined models we observed that the residue $H174^{ECL2}$, which is important for the receptor activation, is displaced from the orthosteric pocket. Additionally, AlphaFold predictions suggest a side pocket between TM1, TM2, and TM7, similar to that of GPR52, where both ligands could form stable hydrophobic interactions. These findings indicate that GRA2 and UCSF924 are likely to bind either within the orthosteric pocket or the side pocket, reducing GPR21's constitutive activity and therefore act as GPR21 inhibitors. Our models provide a basis for further molecular dynamics simulations and in vitro studies to confirm these interactions and explore the therapeutic potential of targeting GPR21.

Keywords: GPR21, GPCR, G-Protein Coupled Receptor, Molecular Docking, Homology Modelling, GRA2, UCSF924

Purpose of the study

The aim of this work is to investigate the mechanisms of interaction between GRA2 and UCSF924 and the G protein-coupled receptor 21 (GPR21) using structural analysis, docking simulations and homology modeling. GPR21 is a member of the GPCR family, which is involved in many physiological and pathological processes and is essential for cellular signaling. GPR21 was found to be constitutively active. This activity, in the absence of an intrinsic ligand, has been implicated in several diseases, including neuroinflammatory and metabolic disorders. GRA2 and UCSF924 have been shown to interact with the receptor, but how they interact with GPR21 is not well understood.

We are refining models of GPR21, focusing on extracellular loop 2 (ECL2), and investigating whether the two ligands are likely to interact with the receptor in its orthosteric pocket. By comparing the druggability and docking scores of different refined models, we aim to identify which structural configurations of GPR21 are most likely to effectively accommodate GRA2 and UCSF924.

In addition, we are investigating the existence of an alternative binding site, such as a side pocket, where the ligands could interact with GPR21. For this purpose, models of GPR21 were generated using Alphafold2, with the main goal of finding a side pocket similar to the one found in GPR52 due to their high sequence homology. Based on this, we propose interaction scenarios for GRA2 and UCSF924 with GPR21, which could help to understand their effects on the activity of the receptor.

This study suggests two possible ways that GRA2 and UCSF924 could interact with GPR21: either by binding to the main pocket of the receptor to alter its high activity, or by binding to a side pocket to disrupt its constant signaling.

The goal of the study is to deepen the understanding of the structure and function of GPR21, which could help in the development of new treatments targeting this receptor. By learning more about how GPR21 interacts with specific molecules, this research could have a significant impact on drug discovery for diseases related to GPCR pathways.

1. Acknowledgments

First I would like to express my gratitude to my supervisor, Minos Matsoukas, for his guidance and support throughout this project. His insights and expertise have greatly contributed to this thesis, and I am very fortunate to have worked with him.

A special thank you goes to my thesis committee members, Manolis Athanasiadis and Spiros Kostopoulos, for their time and comments.

I would also like to thank the members of the CODDIA group, Vasilis Panagiotopoulos and Marios Giatro, for their advice and help. Throughout this period, they have been more than willing to help and provide feedback that has shaped the progress of this thesis.

On a personal note, I am grateful to my family and friends for their love and encouragement. To Karlyn, your support, understanding, and constant belief in my abilities have been a source motivation during the most challenging times.

Finally, I would like to thank everyone who, in one way or another, contributed to the completion of this thesis. Your contributions, both big and small, are deeply appreciated.

List of Abbreviations

Abbreviation	Definition
GPCR	G-Protein Coupled Receptor
TM	Transmembrane
ECL	Extracellular Loop
ICL	Intracellular Loop
GTP	Guanosine-5'-triphosphate
GDP	Guanosine diphosphate
RGS	Regulator of G Protein Signaling
CNS	Central Nervous System
oGPCR	Orphan G-Protein Coupled Receptor
FDA	Food and Drug Administration
MC1R	Melanocortin-1 Receptor
LPA	Lysophosphatidic Acid
S1P	Sphingosine-1-Phosphate
TSH receptor	Thyroid-stimulating hormone receptor
LHR	Luteinizing Hormone Receptor
KO	Knock Out
PCR	Polymerase Chain Reaction
cAMP	Cyclic adenosine monophosphate
MCP-1	Monocyte Chemoattractant Protein-1
CCR2	C-C Motif Chemokine Receptor 2
ATMs	Adipose Tissue Macrophages
HGP	Hepatic Glucose Production
BRET assay	Bioluminescence Resonance Energy Transfer Assay
WT	Wild Type
MAPK	Mitogen-Activated Protein Kinase
Cryo- EM	Cryogenic Electron Microscopy
PDB	Protein Data Bank
DOPE score	Discrete Optimized Protein Energy
SF	Scoring Function
PAE	Predicted Alignment Error
plDDT	Predicted local Distance Difference Test

List of Tables

List of Figures

2. Introduction

2.1 G-Protein Coupled Receptors (GPCRs)

With 4% of all protein-coding genes belonging to the G protein-coupled receptors (GPCRs), they are the biggest receptor superfamily [1][2][3]. GPCRs mediate the effects of two-thirds of hormones and neurotransmitters. Their influence on human physiology is reflected in medicine, where GPCRs are the target of 34% of marketed drugs. As a result, a sizable community of researchers studying receptors and drug discovery works with GPCRs [4][5].

The main structural characteristic that is shared among GPCRs is the seven hydrophobic transmembrane (TM) segments, with an external amino terminus and an intracellular carboxyl terminus. Great homology is found within the TM segments among GPCRs. On the contrary, areas such as the carboxyl terminus, intracellular loop (ICL) between TM5 and TM6 and the amino terminus present great variability [6].

The GPCR superfamily is divided into 6 classes, based on functional similarities and amino acid sequence [7] [8]:

- Class-A: “Rhodopsin-like” GPCRs. 80% of GPCRs
- Class-B: Secretin receptor family
- Class-C: Metabotropic glutamate receptors
- Class-D: Fungal mating pheromone receptors
- Class-E: Cyclic AMP (cAMP) receptors
- Class-F: Frizzled/Smoothened

2.2 Class A GPCRs and activation mechanism

Seven alpha-helical transmembrane (TM1–7) segments, an eighth intracellular helix (H8), three intracellular loop regions (ICL1-3), three extracellular loop regions (ECL1-3), an extracellular N-terminus, and an intracellular C-terminus make up the structural components of class A GPCR. Ligands that activate the class A GPCRs bind to an orthosteric site created by the TM domains, which is located at the extracellular side of the receptor. Variations in binding pocket size, shape, and electrostatics affect receptor-ligand selectivity [9][10].

The receptor undergoes a conformational shift upon agonist engagement that causes intrahelical residue interactions to reorganize. Activation of Class-A GPCRs results in the movement of the intracellular ends of TM5, TM6 and TM7. Therefore, contacts between residues of TM3 and TM6 are eliminated, whereas new ones are being

formed between TM3 and TM7, along with the structural repacking of TM5 and TM6. Since TM3 has state-specific contacts to other TM domains, it frequently experiences helical rotation upon receptor activation and operates as a hub for activation state stabilization [9]. The outward movement of transmembrane helix 6 (TM6) has been characterized as the hallmark of class A GPCR activation [10]. (Fig.1) [10]. (Fig.1)

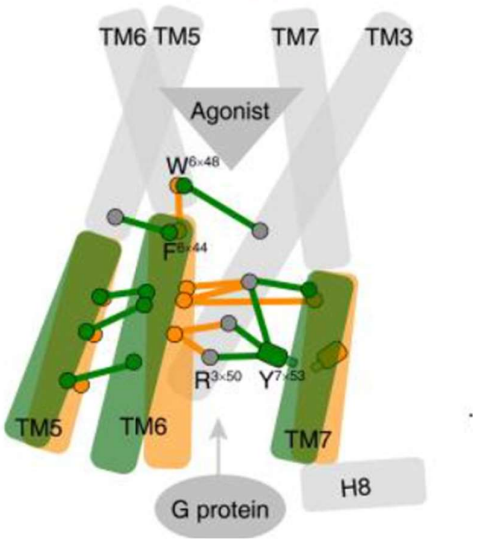

Figure 1: Common activation model of class A GPCRs (adapted from [11]*)*

This conformational shift, stimulates the activation of the G protein located to the intracellular side of the Class-A receptor. On the Gα subunit of the heterotrimeric G protein, guanosine triphosphate (GTP) replaces guanosine diphosphate (GDP) following G protein activation. The G_{α} and $G_{\beta\gamma}$ subunits of the G protein separate when GDP is exchanged for GTP, and both subunits can subsequently alter the function of effectors downstream in cell signaling pathways. Binding of the regulator of G protein signaling protein (RGS) promotes GTPase activity, which aids in the restoration of the basal state. Receptor affinity for the ligand is decreased and the complex returns to its basal state when GTP is hydrolyzed to GDP (Fig.2) [9].

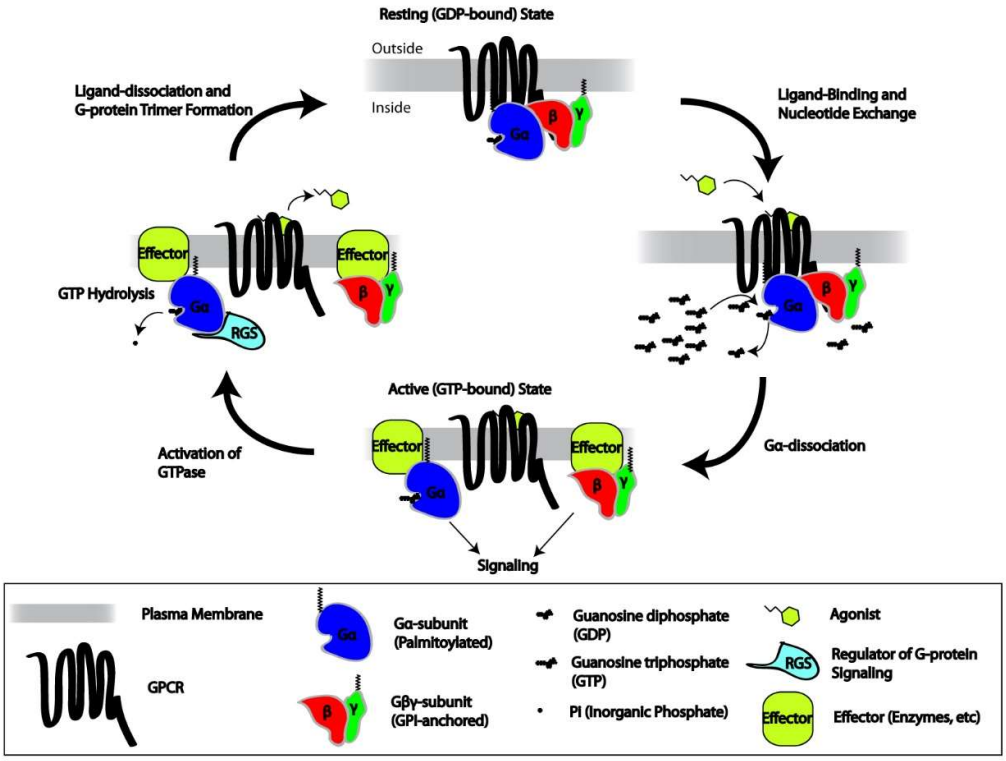

Figure 2: Class A GPCR activation cycle

2.2.1 GPCRs as drug targets

GPCRs are highly effective and promising targets in pharmaceuticals [12][13][14][15]. For decades they have been the subject of intense study by the pharmaceutical industry, because of their physiological significance and their small molecule targeting accessibility [15]. These receptors are essential to how the body reacts to a wide variety of external or endogenous stimuli, which includes proteins, peptides or macromolecules, small molecules, photons and even ions. This wide variety of input, depicts accurately the amount of different functions that GPCRs are involved in. GPCR signaling controls immunological response, cell organization, and cognition [15]. The receptors are key targets for small-molecule therapies used in treatments for cancer, viral infections, inflammatory conditions, metabolic disorders, and Central Nervous System (CNS) disorders [12][15].

GPCRs are located in the cell membrane which allows pharmacological access to them. Many GPCR binding pockets' topology, location, and physical and chemical characteristics have assisted the identification of small molecules that have been used in medicine for many years [15].

Over 35% of newly introduced medications target GPCRs, a number that represents more than 27% of the global pharmaceutical market. Only about 13% of GPCRs (103 members) are effectively targeted by FDA-approved medications. Of these druggable GPCRs, only half have known endogenous ligands [12]. More than 400 of them have the potential to be targeted for therapeutic treatments, whereas for roughly 120 orphan GPCRs (oGPCRs) there are no known endogenous ligands and their functional roles are still to be discovered [16][17][18].

According to GPCRdb 94% of approved drugs are targeting Class-A GPCRs, 4% of them target Class-B, and Classes C and F occupy 1% respectively. (Fig. 3)

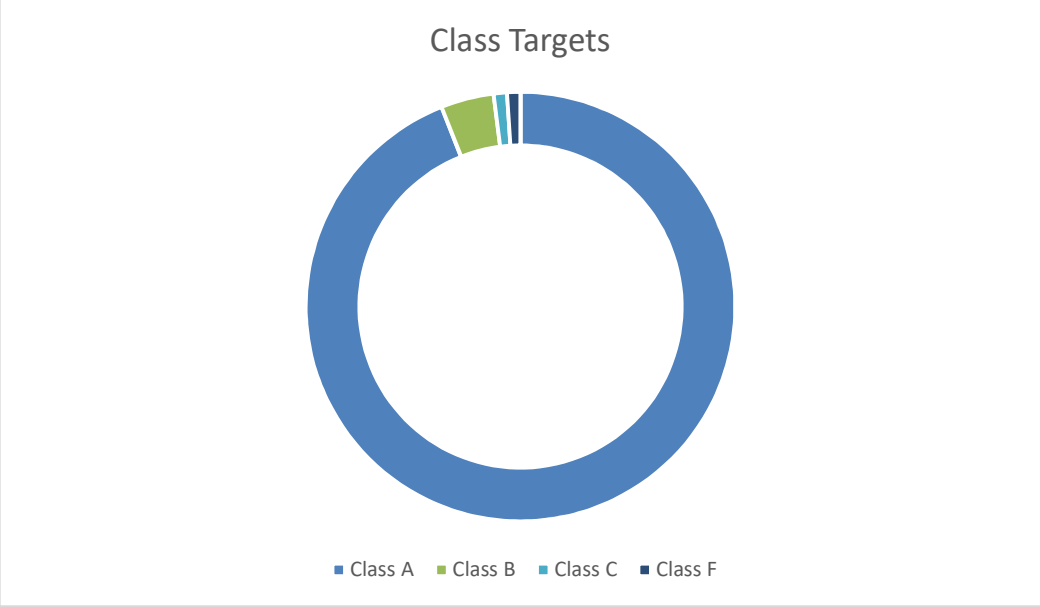

Figure 3: Class targets of approved and distinct drugs. Data was extracted from https://gpcrdb.org/drugs/drugstatistics

There are many reasons regarding the low number of FDA-approved drugs targeting GPCRs. First, many GPCRs that could be potentially targeted by drugs, especially oGPCRs, have biological roles that are still not fully understood [12][15]. Also, even in cases of GPCRs that are already linked to specific diseases, traditional drugs (agonists or antagonists) that bind into the highly conserved orthosteric pocket have not always found clinical application, because they cause off-target effects [15][19][20].

Despite the clear interest in GPCRs as potential drug targets, another reason why they remain understudied is due to technical and handling difficulties. Many of the understudied receptors are challenging to analyze in assay settings because of their low cell expression levels, their unique and complex behaviors and their involvement

in pathways that require more costly, time-consuming methods. Additionally, lack of detailed structural information and specific ligands further complicates research in silico. In vivo studies also face challenges because of the differences between animal models and humans in receptor subtypes and signaling mechanisms [21].

2.3 Diseases linked to GPCRs

GPCRs are involved in a wide range of cellular and physiological processes. Therefore, their abnormal activity or expression is connected to many disorders.[22] In Figure 4 there are listed some of the most common disease indications of approved drugs that are targeting GPCRs according to GPCRdb.

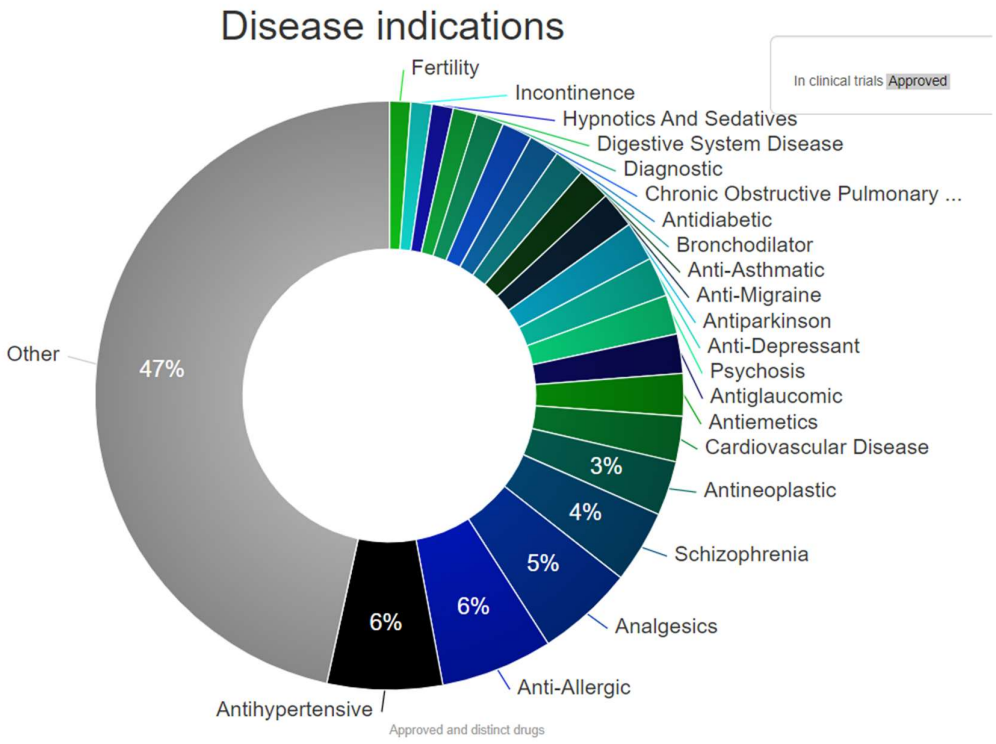

Figure 4: Disease indications of approved drugs linked to GPCRs. Data extracted from https://gpcrdb.org/drugs/drugstatistics [23]

More and more studies are finding that abnormal GPCR expression and activation are linked to various types of human cancers. For example, some GPCRs are found in high levels in different tumors, and certain GPCR variants can increase the risk of cancer [24]. A study has shown that alterations in the melanocortin-1 receptor (MC1R) gene are connected to an increased risk of skin cancer development.[25] Abnormal activation of GPCRs due to exposure to high concentrations of specific chemicals, such as LPA (lysophosphatidic acid), S1P (sphingosine-1-phosphate) (bioactive lipids), and

chemokines, has been linked with processes such as drug resistance, cancer spread, blood vessel development, cell transformation, and growth [24].

GPCRs play important roles in metabolism and energy homeostasis. Some have been linked to metabolic diseases like obesity and type 2 diabetes [26]. Studies have found that some GPCRs n chare involved in processes like nutrient sensing, appetite control, and the metabolism of glucose and fatty acids and respond to endogenous metabolites (fatty acids, sugars, amino acids) that are usually dysregulated in metabolic diseases [26][27].

Over 30 diseases are linked to GPCR loss or gain, which can be caused by mutations. Apart for cancer, these mutations can activate GPCRs constitutively, potentially leading to cellular transformation as well. For instance, mutations in α1B-adrenergic receptor can enhance receptor/G protein coupling even without agonist ligands. Diseases like autosomal dominant non-autoimmune hyperthyroidism and familial male-limited precocious puberty are caused by activating mutations in TSH receptor and LHR genes, respectively. Conversely, loss-of-function mutations in GPCRs are associated with diseases like nephrogenic diabetes insipidus syndrome and retinitis pigmentosa, disorders that are both caused by gene mutations [28].

Over 800 GPCRs exist in the body, and mutations in different ones lead to conditions like hypothyroidism, hyperthyroidism, nephrogenic diabetes insipidus, and fertility issues [28].

2.4 Orphan GPCRs (oGPCRs)

Most GPCRs were initially discovered as "orphan" receptors [29]. That means that their ligands were still unknown [29][30]. It was believed that they belonged to a supergene family and subsequently they would share sequence similarities. Researchers proceeded the search for new GPCRs using techniques like homology screening and low stringency hybridization. However, the discovery of receptors through homology screening faces the serious challenge of not knowing their pharmacological role and their involvement in various functions, and therefore not having any information about their endogenous ligands [29].

The GPCR superfamily consists of more than 800 members [28]. 49% of them are olfactory receptors, most of which are orphans, 30% are natural ligand receptors, 15% are orphan GPCRs and 6% are used are identified drug target receptors and are used in clinical applications [31]. (Figure 5)

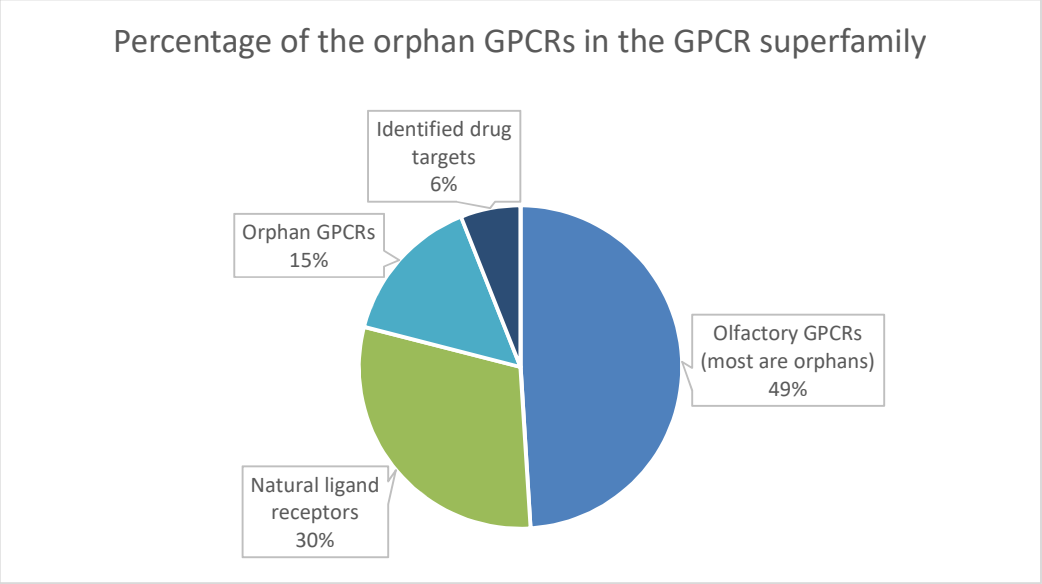

Figure 5: Percentage of the orphan GPCRs in the GPCR superfamily. Graph adapted from [31]

oGPCRs are discovered to be abundant in tissues associated to immunity and metabolism, as well as the brain, and they have a clear tissue-specific distribution. Many physiological and pathological processes, such as cell growth and survival, metabolism, immunological response, allergic reaction, cognition, movement control, and reward, are mediated by oGPCRs and the downstream proteins that are associated with them [12]. One major area impacted by orphan GPCR research is sleep, and in recent years it has been shown than GPCR systems are involved in many aspects of sleep and the circadian rhythm as well. Other physiological responses that are affected by GPCRs is appetite and food intake, and stress-related anxiety [29].

On the other hand, malfunctions of known GPCRs and oGPCRs may result in cancer, neuropsychiatric disorders, metabolic disorders, immunological diseases, and other human disorders. For this reason, the creation of small compounds that specifically target oGPCRs offers a valuable and unexplored field for medication research and development [12].

2.5 Deorphanization approaches

The process of identifying ligands that are highly selective to orphan GPCRs is called deorphanization [31].

Before deorphanization, it is crucial to determine the clinical relevance of the orphan GPCR. The GPCR's physiological role is explored by phenotypic characterization of knockout (KO) mouse models, and in situ hybridization which both assist in evaluating the receptor's potential as a therapeutic target. Reliable screening techniques, good ligands, and sufficient receptor expression are necessary for successful reverse pharmacology, which is the first method that was used to deorphanize GPCRs [1]. It is a widely used method, where the orphan GPCR is first expressed in eukaryotic cells so it can be studied. Then the receptor is tested in various functional assays to screen for potential ligands. Once a specific ligand is identified, it is used to explore and understand the biological and physiological roles of the receptor [30].

The first successful deorphanizations in 1988 involved serotonin 5-HT1A [32] and dopamine D2 receptors [33]. GPCR discovery involved low-stringency hybridization for identifying GPCR subtypes [34]and PCR-based methods for discovering novel orphan GPCRs [1][29].

GPCR deorphanization peaked in the late 1990s and early 2000s, with about 10 deorphanization reports per year, fueled by industrial funding, high-throughput reverse pharmacology, and human genome sequencing [1]. In recent years, the number of newly discovered orphan GPCRs has grown, however, research into their functions has been slowed down by the challenge of identifying their ligands and by the unique structures of these receptors [31] [35].

2.6 GPR21

GPR21 is a class A orphan GPCR with a potential role in type 2 diabetes [7][36]. In humans *Gpr21* gene is located in chromosome 9 (chromosome 2 in mice) [37]. Its amino acid sequence is shown in Figure 6. GPR21 forms a stable complex with Gαs and is essential for transducing intracellular signals via cAMP [7]. It is broadly expressed, including in macrophages and some brain regions, especially the hypothalamus [36][37][38]. Knockout mice lacking GPR21 display improved glucose tolerance and increased insulin response, suggesting a role in regulating body weight and glucose metabolism [36]. Other studies have mentioned that inhibition of GPR21 increases glucose uptake in liver cells, indicating a negative effect on glucose uptake [39].

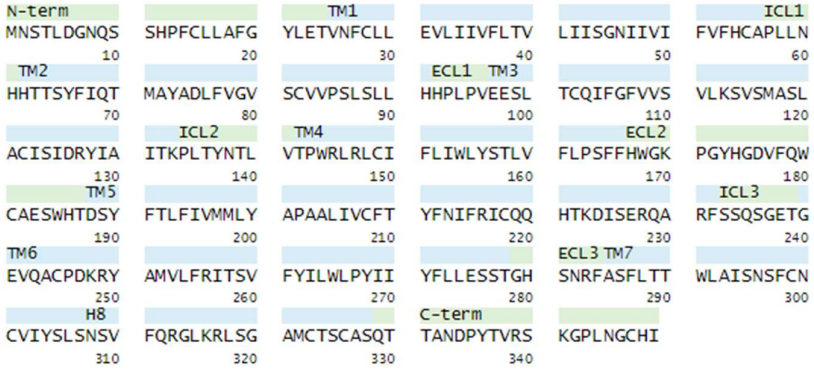

Figure 6: GPR21 amino acid sequence

Mouse *Gpr21* gene is located within the *Rabgap1* gene on chromosome 2, and therefore it was questioned whether these metabolic improvements were due to the *Rabgap1* disruption. In order to further examine the question that occurred, researchers developed CRISPR-Cas9 knockout mouse models, to ensure that the *Rabgap1* gene remained intact. They found that removing the *Gpr21* gene in human and mouse monocytes reduces their movement towards MCP-1, an inflammatory signal, without changing the level of its receptor, CCR2. Monocytes lacking GPR21 show delayed movement in response to the inflammatory signal MCP-1. This delay is linked to lower activity in certain genes responsible for cell movement and adhesion, which might explain their reduced migration. This change in monocyte behavior suggests that GPR21 plays a role in controlling inflammation, which can impact both glucose control and inflammatory responses. Mice without GPR21 showed better glucose tolerance and less inflammation, while higher levels of GPR21 are linked to type 2 diabetes [40].

Targeting GPR21 could be a new way to treat diseases like atherosclerosis, where the inflammatory signals MCP-1 and CCR2 play harmful roles [40]. Additionally, GPR21 is a potential target for type 2 diabetes treatment, with its overexpression attenuating insulin signaling [41].

2.6.1 GPR21 and Type 2 Diabetes

Obesity occurs, leading to chronic low-grade inflammation [42][43][44]. The body responds to obesity with an inflammatory response causing macrophages and other immune cells infiltrate adipose tissue and the liver. Adipose Tissue Macrophages (ATMs) accumulate in increased numbers within adipose tissue leading to the activation of proinflammatory pathways and the secretion of proinflammatory

cytokines. These cytokines impair insulin action in adipose tissue, leading to insulin resistance [45][46]. At the same time, macrophages are also recruited to the liver [47], where they activate proinflammatory pathways, resulting in decreased insulin-mediated suppression of hepatic glucose production (HGP) [48][49][50][51]. The circulation of proinflammatory cytokines in blood, leads to insulin resistance in other tissues- apart from liver (Fig. 7). It has been found that GPR21 is highly expressed in temperature-sensing hypothalamic neurons and in macrophages. Therefore it poses as a potential target for new treatments for obesity and type 2 diabetes [38].

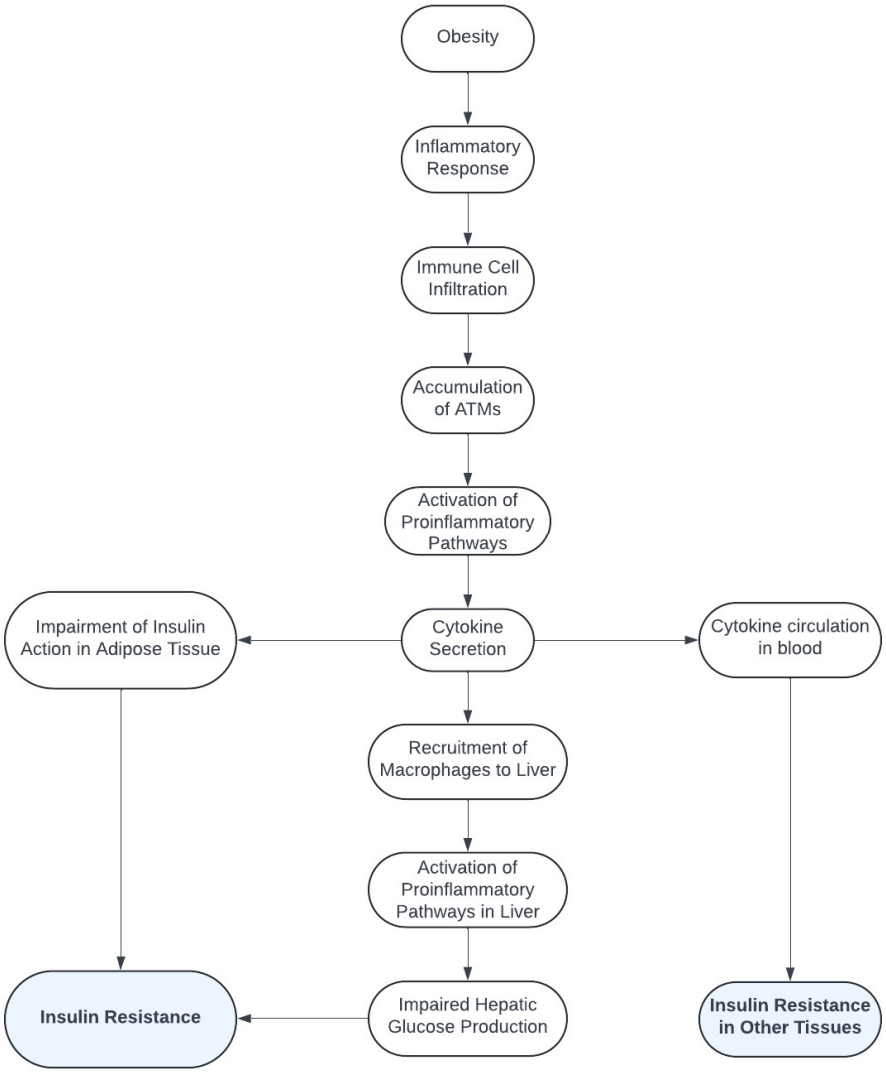

Figure 7: The process of the development of insulin resistance as a result of obesity

2.7 Structural properties of GPR21

As other Class-A GPCR members, GPR21 is comprised from seven transmembrane a-helixes. These helixes are connected with three intracellular loops, responsible for the communication with the G-Protein and three extracellular loops. The N-terminus of GPR21 is located extracellularly, and the C-terminus is intracellular.

In 2023, the 3D structures of GPR21 were determined by two different teams, with the use of Cryo-EM. The structures of the human GPR21 show the receptor coupled with different G-proteins [7][52]. In general, the transition from inactive to active state class A GPCRs is usually accompanied by structural conformation changes. Wong et. al compared Gs-coupled GPR21 with apo-GPR52 structures and mini-Gs-coupled GPR52, due to the high structural similarity that the two receptors show [7]. The GPR21 structure showed great conformational similarity to the mini-Gs-coupled GPR52, which is in an active state. Also, the outward movement of TM6 was observed (Fig. 8), which is a characteristic feature of Class-A GPCR activation. These observations strongly suggested that the GPR21 structure indicates that the receptor is in an active state [7].

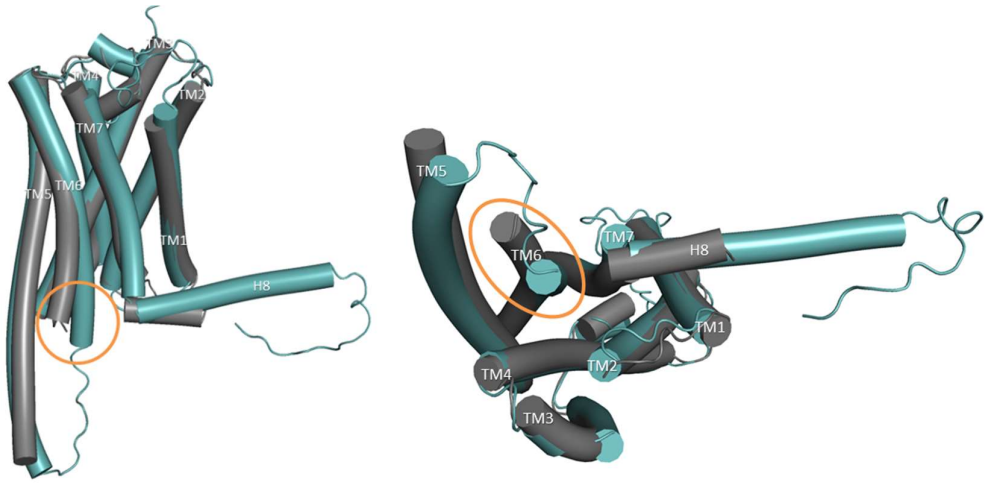

Figure 8: Superposition of GPR21-active state structure 8HMV (gray) and GPR21-inactive state model (downloaded from GPCRdb) (teal). The characteristic outward movement of TM6 is shown in orange.

Figure 9A shows the W265$^{6.48}$ rotation within the CWxP motif on TM6 upon ligand activation. This switch is crucial as it leads to the rotation of the side chains of F261$^{6.44}$, in the PMF domain and W265$^{6.48}$, causing TM6 to move closer to TM3 (Fig.9B). Regarding the sodium pocket, GPR21 side chains maintain a conformation that resembles more the GPR52 inactive state (Fig. 9C). Within the NPxxY motif, Y304$^{7.53}$ appears to have obtained a conformation indicating more similar to the active GPR52

(Fig. 9D). There is a salt bridge between D126$^{3.49}$ and R127$^{3.50}$ of the DRY motif, which keeps GPCRs in the inactive state. In the GPR21 structure, a rotameric switch is observed in the side chain of R127$^{3.50}$, breaking the ionic lock (Fig. 9E). This allows change is important for the receptor to be able to couple with the G-protein. These observations further support that the GPR21 structure is in an active state [7] .

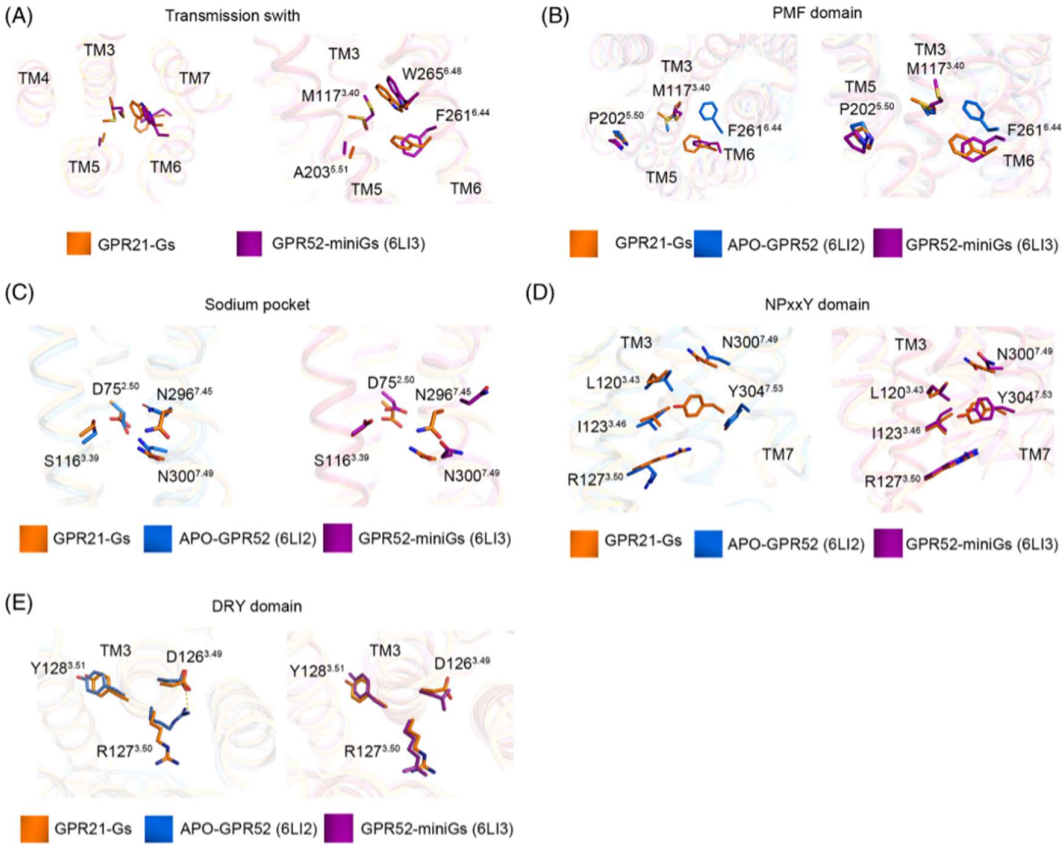

Figure 9: Comparison of conserved motifs between GPR21 (8HMV), apo- GPR52 (6LI2) and miniGs-GPR52 complex (6LI3). A) transmission switch, B)PMF domain, C) Na+ Pocket, D) NpxxY domain, E) DRY domain (Adapted from[7])

Studies have shown that ECL2 acts as a built-in agonist, promoting high basal signaling and stabilizing GPR21-G protein complexes. ECL2 contains a conserved 22-residue sequence, similar to GPR52 where a constitutive activity has also been observed. First half of ECL2 (residues 169–178) are found buried in the pocket (immersed region),

while the second half (residues 179–186) protrudes to the extracellular surface (cap region) (Fig. 10) [7][52].

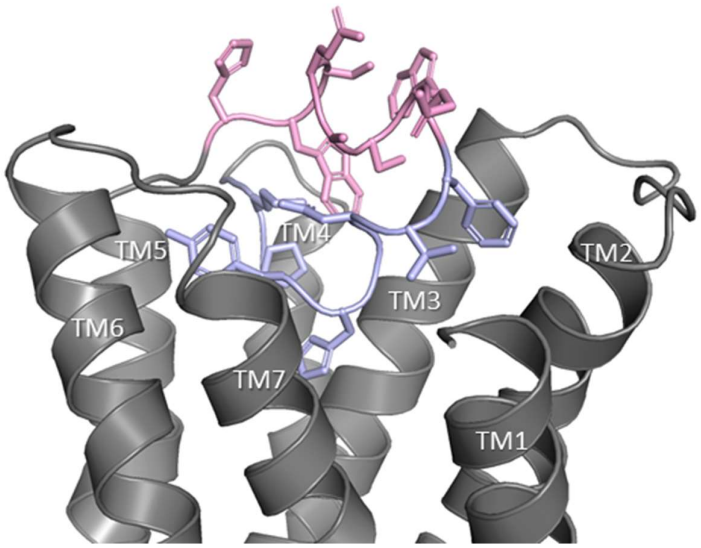

Figure 10: Regions of ECL2: cap region is represented with pink, and immersed region is represented with blue

The placement of ECL2 is characterized by a few stabilizing interactions (Fig. 11):

- K170 forms a key salt bridge with D176
- Two π-π interactions: H174-Y268 and F178-F105
- Conserved disulfide bond between C181 and C102 in TM3

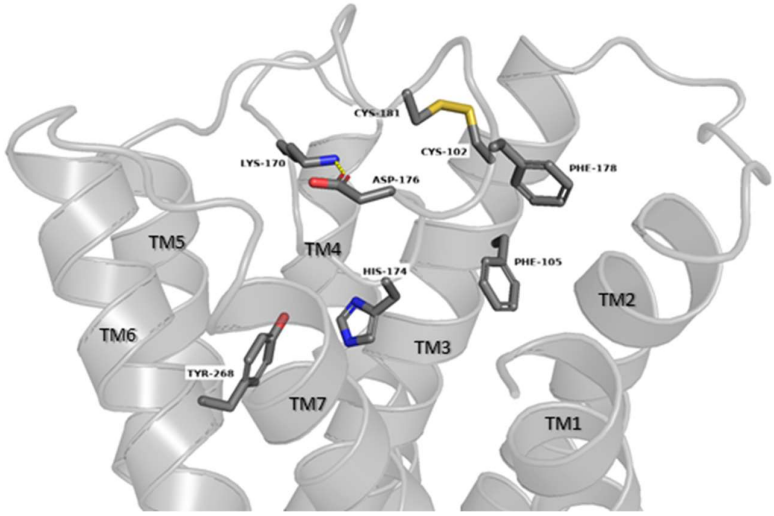

Figure 11: ECL2 stabilizing interactions

2.8 Constitutive activity and mechanism

It has been shown that disrupting ECL2 conformation, by introducing mutations like replacing C181ECL2 to Alanine or K170 to Glutamic Acid, significantly reduces GPR21 basal activity, which refers to the intrinsic receptor activity in the absence of a ligand. The deletion of ECL2 prevents GPR21 from forming a GPCR-G protein complex, explaining the reduction of GPR21 basal activity in both Gs and G15 cases. Replacing parts of ECL2 (residues 170-178, and residues 179-186 respectively) with a 6-residue GS linker (GGSGGS) also results in a drop of the receptors basal activity (Fig. 12) [52].

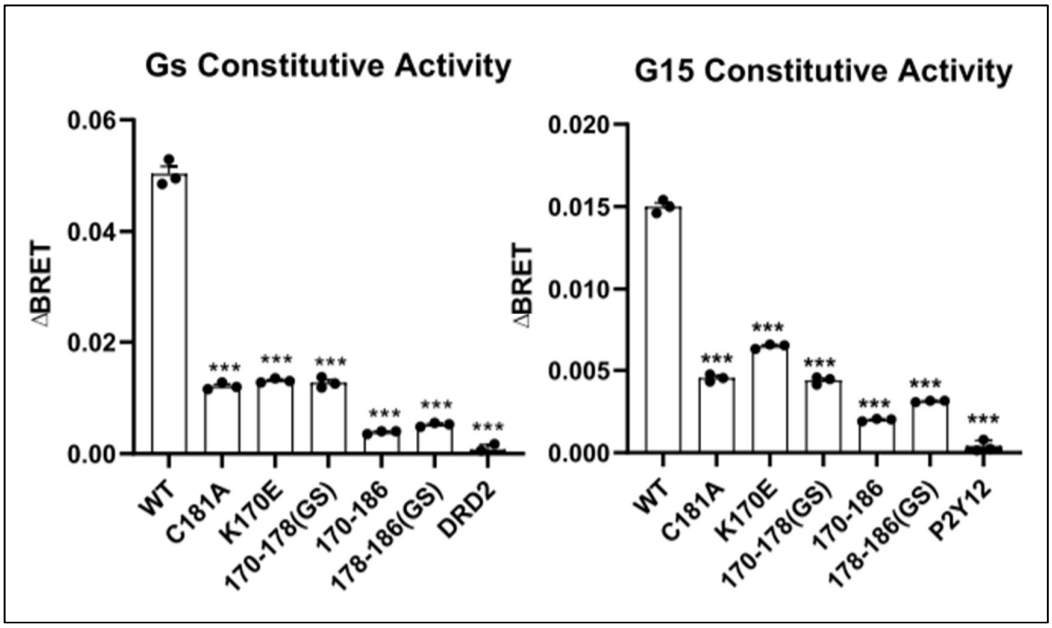

Figure 12: Constitutive activities of WT GPR21 and GPR21 ECL2 mutants in Gs (left) and G15 (right) signaling were assessed using the BRET (C181A: Cysteine to Alanine mutation, K170E: Lysine to Glutamic Acid mutation, 170-186: deleting these residues from WT GPR21, 170-178(GS) and 178-186(GS): replacing each region of WT GPR21 with 6-residue linker (GGSGGS). DRD2 and P2Y12: negative controls for Gs and G15 signaling (adapted from [52])

Additionally, replacing ECL2 with a flexible GS linker reduces cAMP accumulation mediated by Gαs protein. Three scenarios were explored, where two different sections of ECL2 (residues 169-178, and residues 179-186 respectively), and the whole ECL2 were replaced by a 6-residue GS linker (GGSGGS). In every scenario the cAMP accumulation was significantly reduced, whereas the overall expression of the receptor was not affected by any of these mutations (Figure 13) [7].

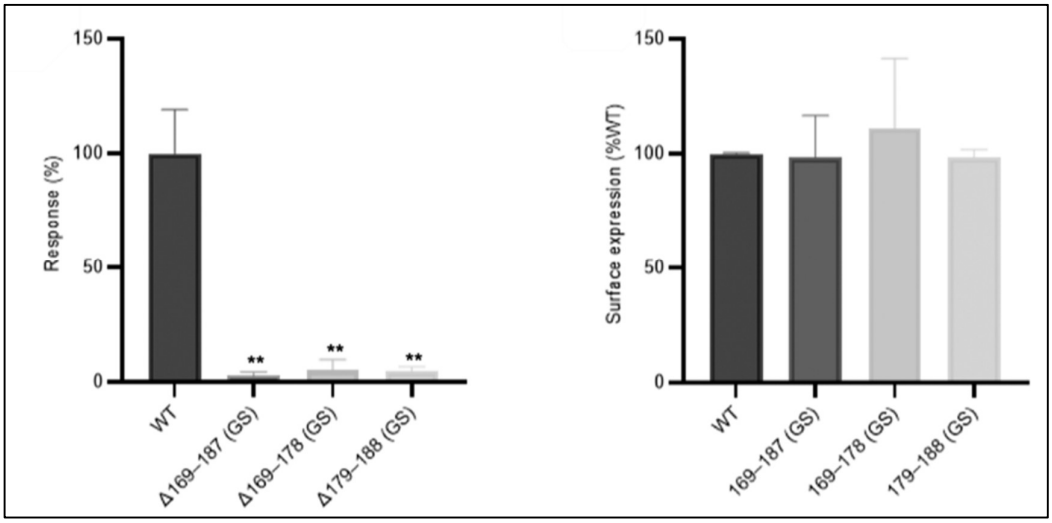

Figure 13:The effect of replacing different regions of ECL2 to cAMP accumulation compared to WT GPR21 (left), while the receptor expression remained unchanged by those mutations(right). (Adapted from [7])

When combined, these findings demonstrate how crucial ECL2 is for the baseline activation of the receptor in the absence of a ligand [7][52].

2.9 Ligands

Even though GPR21 is an orphan receptor, there are two compounds that have been found to interact with the receptor.

GRA2

The effect of GRA2 (Fig. 14) has been studied in macrophages, M1 (pro-inflammatory phenotype) and M2 (anti-inflammatory phenotype) macrophages and it was found that it can reduce GPR21 basal activity.

Additionally, the same study shown that GPR21 caused MAPKs activation and negatively affected insulin signaling, results that have been linked to the promotion of macrophage concentration in the tissue [39], [41]. Since GRA2 is an inverse agonist of GPR21, it can prevent receptor activation and therefore prevent these events. Its properties are shown in TABLE 1.

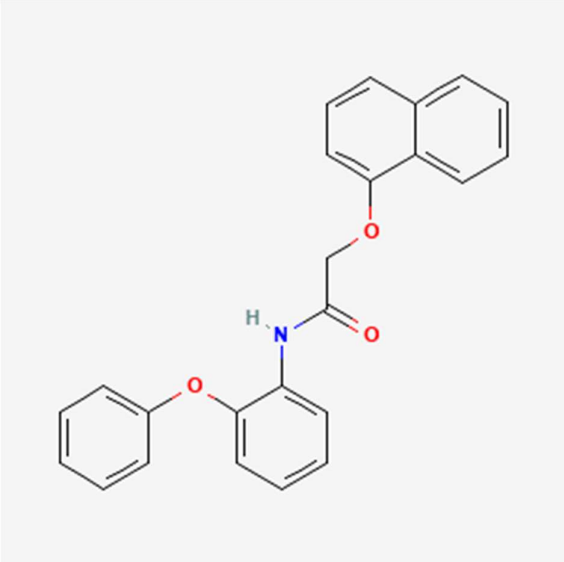

Figure 14: 2-naphthalen-1-yloxy-N-(2-phenoxyphenyl)acetamide (GRA2)

TABLE 1: PROPERTIES OF GRA2

Property Name	Property Value
Molecular Weight	369.4 g/mol
Hydrogen Bond Donor Count	1
Hydrogen Bond Acceptor Count	3
Rotatable Bond Count	6

UCSF924

UCSF924 (Fig. 15) (properties shown in TABLE 2) is included in the ChEMBL database as a compound evaluated against GPR21 as it was part in a high-throughput screening using PRESTO-Tango assay. The purpose of the PRESTO-Tango system is to quantify the important GPCR signaling event known as β-arrestin2 recruitment. With the help of this assay, substances can be profiled across a variety of GPCRs to determine whether they have the potential to be agonists or antagonists [53]. Since UCSF924 is the only listed compound linked to GPR21 in ChEMBL, we used molecular docking to investigate more of its possible interactions with the receptor.

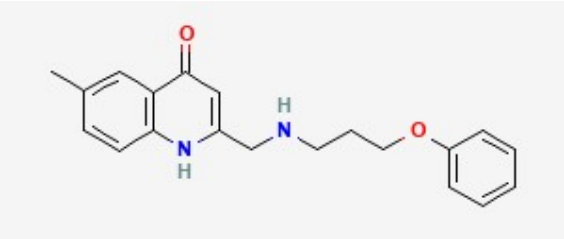

Figure 15: 6-Methyl-2-{[(3-phenoxypropyl)amino]methyl}quinolin-4-OL (UCSF924)

TABLE 2: PROPERTIES OF UCSF924

Property Name	Property Value
Molecular Weight	322.4 g/mol
Hydrogen Bond Donor Count	2
Hydrogen Bond Acceptor Count	4
Rotatable Bond Count	7

2.10 Comparison with GPR52

The *Gpr52* gene was discovered through homology searches of high-throughput genome databases [54]. Studies have linked the *Gpr52* gene with the risk of schizophrenia development [55][56]. GPR52 is another oGPCR, with a promising potential for being a druggable target. About 70% of this receptor's expression in human tissues is in striatal brain regions [12][57].

It has been found that GPR21 has a high structural similarity to GPR52 (Fig.17) and the two receptors share 71% sequence homology (Fig16) [7].

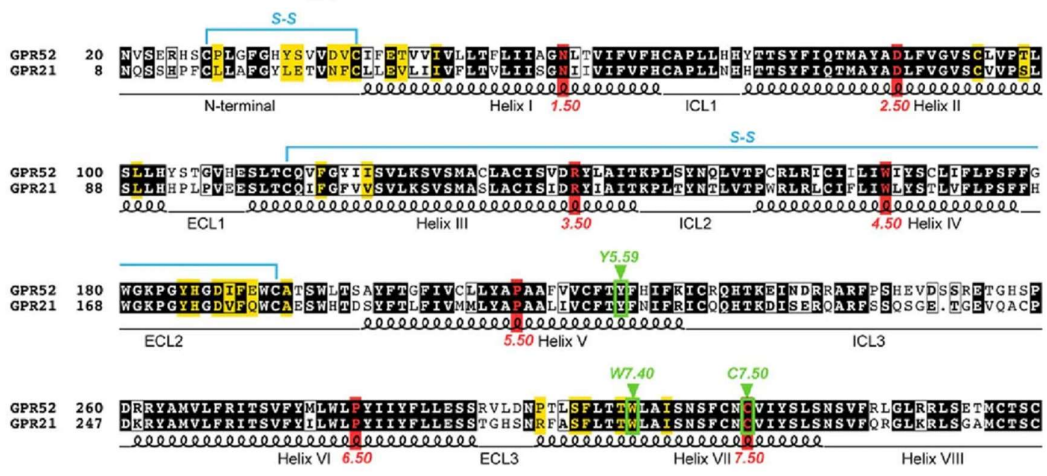

Figure 16: Sequence alignment of GPR52 and GPR21: green denotes essential residues for structural characteristics; yellow, less than 5.0 Å to ligand

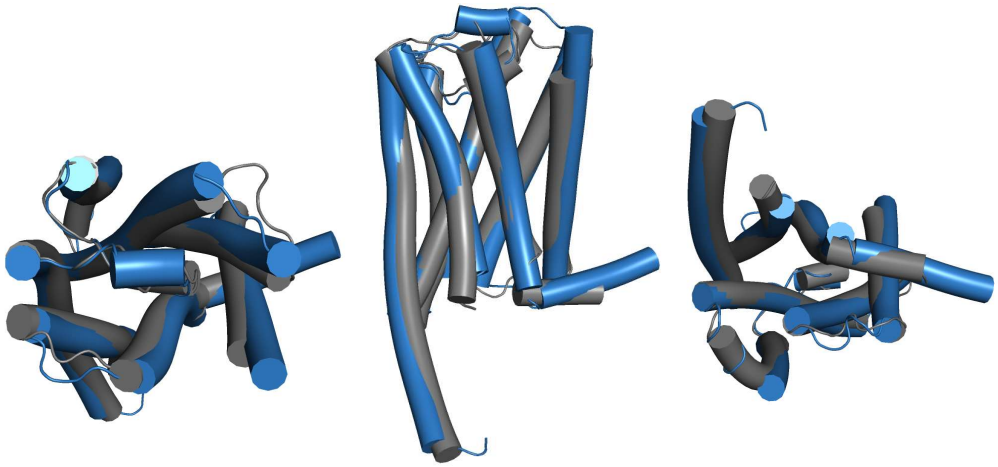

Figure 17: Superposition of GPR21 (8HMV-grey) and GPR52 (6LI3-blue). Extracellular side is on the left, and the intracellular side view is on the right.

Like in GPR21, the ECL2 domain is found inside the orthosteric pocket, resulting in a high basal activity in the absence of an intrinsic ligand [12]. Figure 18 shows the side chains of the amino acids that are deeper into the orthosteric pocket of the two receptor, and they have very similar conformations in both.

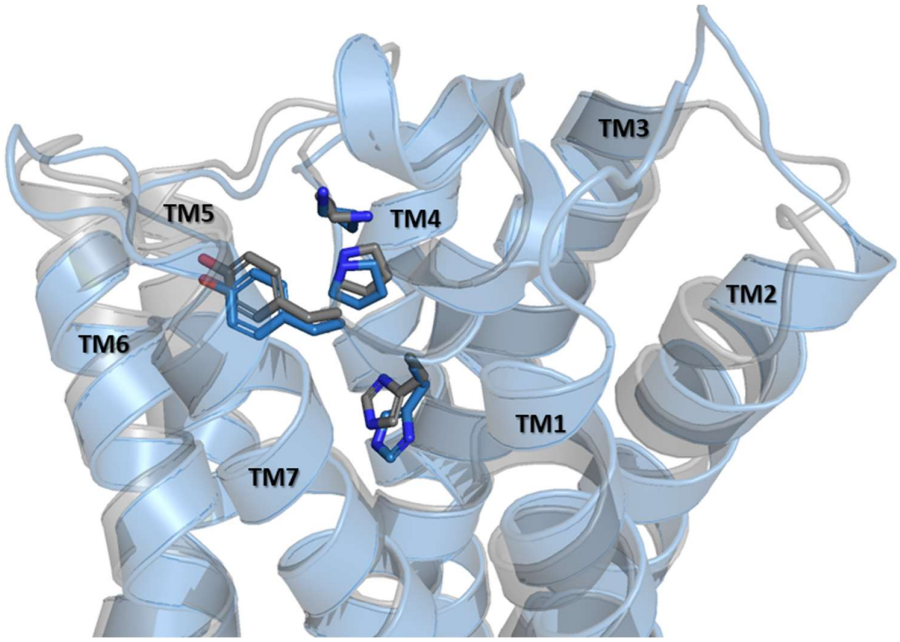

Figure 18: Superposition of GPR52 (blue) and GPR21 (gray). The amino acids that are deeper in the orthosteric pocket are shown as sticks

While ECL2 occupies the orthosteric binding site of GPR52, a side pocket has been found, surrounded by TM1, TM2 and TM7 (Fig. 19) [52]. Studies have shown that even though the similarity between the two receptors in the side pocket region is quite high

(59% identity), neither of the GPR52 agonists (c17, m7), are capable of increasing the basal activity of GPR21 as well [58][59][60].

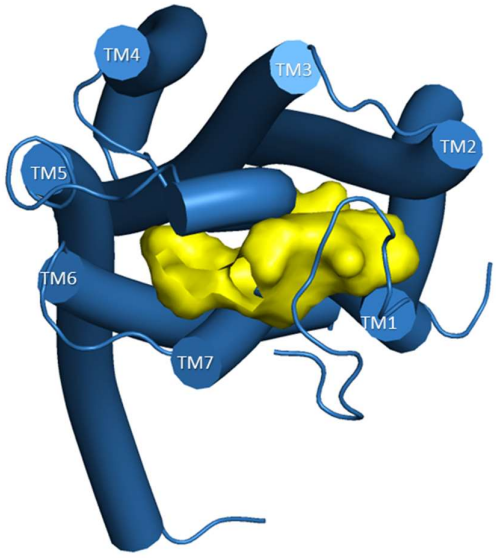

Figure 19: Side pocket of GPR52 (yellow)

One study explored whether GPR21 has a similar side pocket with GPR52. It was determined that the N-terminal region is crucial, since GPR52 agonists could activate GPR21 signaling when the N-terminal region of GPR21 was replaced with that of GPR52 [52]. Mutations in the transmembrane domain did not enable GPR21 to bind GPR52 agonists. Finally, the single residue L16 in GPR21 (corresponding to P28 in GPR52) is of great importance. The absence of proline at this position in GPR21 leads to the loss of binding capability for GPR52 ligands [52].

2.11 Advances in Drug Discovery

In 2015, more than 120 structures from 32 different receptors were available [61][62]. Now, according to GPCRdb, there are a total of 249 entries referring to representative structures (state and receptor) of human receptors. 164 of them were solved by cryo-electron microscopy (cryo-EM) and 85 by X-ray crystallography [4][63].

2.11.1 Structural biology methods

X-Ray Crystallography

X-ray crystallography was first reported in the early 1930s [64][65], while the atomic structure of myoglobin was published nearly 30 years later [66]. In this method, a

single crystal is exposed to a focused beam electromagnetic waves, the X-rays. When the X-rays hit the crystal, they create a diffraction pattern that is captured as 2D images. By rotating the crystal and taking multiple 2D images from different angles, in order to combine them into a detailed, atomic-level picture of the crystal [67].

It was not until 1976 that the concept was introduced that knowing the structure of a protein could greatly aid in the design of specific ligands [68]. This idea led to the "rational drug design cycle", where knowing the structure of a drug-target complex can help optimize the drug.[69] Now, X-ray Crystallography is commonly used for drug characterization, with the modern techniques such as crystallization automation faster detectors, and automated structure solutions. The benefit of these developments is that macromolecular structures may now be determined quickly, sometimes in a matter of days [70][71].

Cryo- EM

Cryo-EM is a method of creating microscopic images of individual molecules, which does not require crystals. This is achieved by freezing the solution of the protein (or other molecule) and then exposing it to an electron beam. These help to reconstruct the 3D structure of the molecule. These structures are helpful in understanding how proteins function, how they might be affected by disease, and how they might be targeted by drugs.

For many years, X-ray crystallography was the preferred method for determining protein structures. It can produce high quality images, but it comes with the limitation of some proteins being difficult or impossible to crystallize. Cryo-EM was dismissed by some scientists because it produced low resolution images. In 2012–2013, advancements in hardware and software led to the development of more sensitive electron microscopes and better software for processing images into clearer molecular structures [72].

2.11.2 Computational methods assisting drug discovery

Protein structure prediction

3D protein models can be predicted by following one of two approaches:

- Physical Interaction methods that uses kinetic and thermodynamic models to simulate molecular physics.

- Evolutionary history approach, which has benefited by from the growing of PDB, deep learning advances and genomic sequencing. These methods use bioinformatics to analyze evolutionary history, homology to already known structures and pairwise correlations.

Both approaches have produced unreliable results, with their accuracy being far from the experiments [73].

Homology Modelling

Homology modelling is a method that predicts the 3D structure of the protein from its amino acid sequence [74]. It is based on two major observations: The amino acid sequence of a protein specifically dictates its three-dimensional structure and during evolution the structure tends to change in a much slower rate and remain more conserved, compared to the sequence [75].

The process of homology modelling (Fig. 20) starts with identifying and selecting the template structure from PDB of a protein that presents significant amino acid sequence similarity to the protein whose 3D structure is to be predicted. Usually tools like BLAST, are used to find proteins with high sequence homology. The next step is to perform alignment of the query sequence with the selected template. Then the 3D model of the target protein can be built, using methods like rigid-body assembly, segment matching, spatial restraints, or artificial evolution. The less conserved regions, the loops, are later added to the model, after comparing with known structures (database search) or after conformational search. Conformational search includes the generation of random conformations of the loop, from which the one with the lowest energy is selected. Side chains are typically modeled using rotamers, which are preferred conformations stored in rotamer libraries. Then the model is optimized with adjustments that reduce steric clashes to minimize energy. For the optimization, molecular dynamics and Monte Carlo simulations are often used in order to find the most stable and energetically favorable structure. After constructing the 3D model, it must be validated to ensure it is biologically relevant. This involves comparing the model to experimental structures and using evaluation to assess the model's stereochemistry, physical parameters, and overall quality [75].

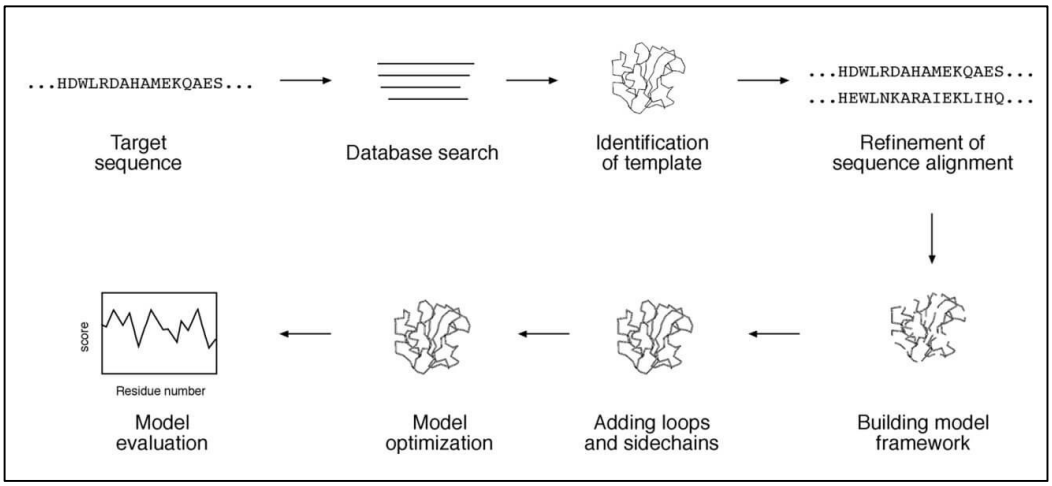

Figure 20: The steps of homology modelling

MODELLER (among others) is a type of software that is used for homology modelling. It creates models of 3D structures of proteins complying with the restraints on the spatial structure of the amino acid sequence. MODELLER automatically takes these restraints from the already known structures and their alignment with the sequence of interest. It is also possible to produce models with only parts of the sequence refined. MODELLER also provides the option to evaluate the produced models, by DOPE (Discrete Optimized Protein Energy) score assessment and determine the quality of a structure model as a whole [76].

Protein- Ligand interaction prediction

Molecular docking

Molecular docking is a method that examines how two structures, for example a protein and a ligand, interact with each other (Fig. 21) [77]. A molecular docking algorithm was first developed in the 1980s [78]. The method revolves around two classical models of molecular recognition: Lock-and-key, and Induced fit. The lock-and-key model suggests that the receptor will recognize only a specific ligand. On the contrary, the induced fit model suggests that the receptor can change its conformation when binding with the ligand [79].

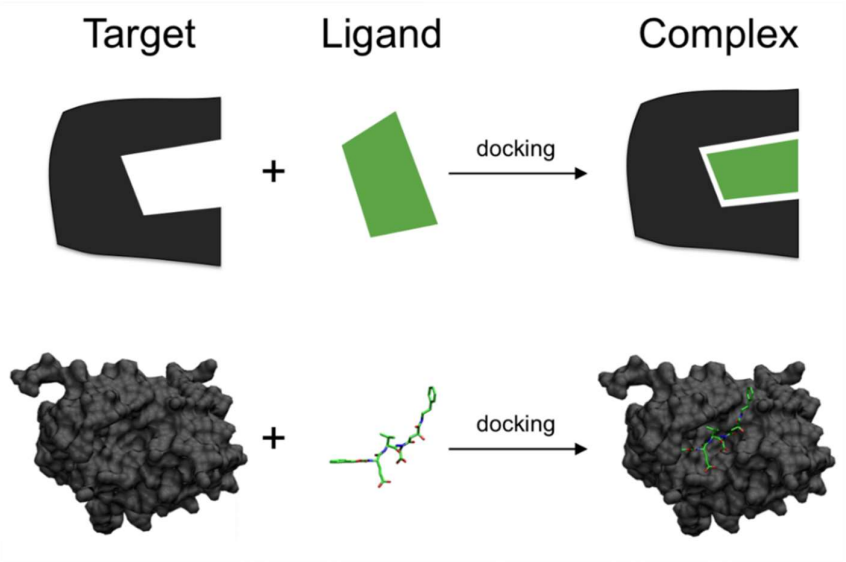

Figure 21: Schematic illustration of docking a small molecule ligand (green) to a protein target (black) forming a protein-ligand complex. (By Scigenis - Own work, CC BY-SA 4.0)

Types of Docking

The docking methods are divided according to the flexibility that the receptor and/ or ligand have during the process. In rigid docking, like the receptor and ligand are considered static, and no movement being taken into account. For example, the first docking algorithm belongs to the rigid docking method. The receptor was modeled by filling its surface openings (binding sites) with spheres that represent the shape of the binding site. The ligand was also modeled by spheres that represent its structure. A search algorithm was used to find the best alignment between the spheres of the ligand and those of the receptor binding site. This aims to determine how well the ligand fits into the receptor's binding site based on their steric (spatial) compatibility. In the flexible docking method, both the receptor and the ligand are considered flexible, and not static. Lastly, there is the semi-flexible docking method, where only the ligand is flexible, and the protein remains rigid [78].

Types of algorithms

Search algorithms are used to pose the ligand in the area of interest. They can be divided in three basic categories that are systematic, stochastic and simulation methods. Systematic algorithms explore all the possible rotations, translations, bond angles, etc. in a molecule during docking. They often run into the issue of an overwhelming number of possibilities to consider [80]. The stochastic algorithms work by making random changes to the ligand, which are evaluated to determine if they resulted in a better conformation. Some widely used stochastic algorithms are Monte Carlo and Genetic algorithms [80]. The simulation approaches aim to simulate how

the receptor and ligand interact. The most popular simulation method at the moment is Molecular Dynamics [78][80].

Scoring functions (SF) play a very important role in molecular docking since they evaluate and rank the predicted conformations of the protein-ligand complex [80]. There are three types of scoring functions:

- Force-field based SF: it is used to approximate the potential energy of the protein-ligand system, involving both intramolecular and intermolecular interactions [78][80].
- Empirical SF: it is usually the sum of empirical term such as van der Waals, electrostatic, hydrogen bond, desolvation, entropy, hydrophobicity, etc. Each term is weighed by a coefficient and are optimized through least squares fitting to match binding affinity data from a training set of known complexes [78][80].
- Knowledge-based SF: they operate on the assumption that ligand-protein interactions that are statistically more frequent have a correlation to favorable interactions [78][80]

Docking Software

Docking software are computational tools that are used to predict interactions between two molecules. Usually a small molecule, which is a potential drug candidate, and a larger molecule like a protein. They aim to predict where and how the two molecules interact with each other. The docking software predicts the orientation and conformation that the small molecule is most likely to have, when it binds to the protein. The different potential conformations that the ligand can have, are called binding poses. Some programs allow for the protein and/or the ligand to have some flexibility during the docking process [81].

After the docking poses have been produced, the molecular docking software use a scoring function (SF), which is another computational tool, in order to evaluate and rank them. The SF calculates the binding energy of the ligand-protein complex, after taking into consideration the strength and stability of their interaction. Lower binding energy indicates a stronger interaction [81].

Autodock Vina, one of the engines of AutoDock Suite, is one of the most successful and widely used docking software. It is a molecular docking program that is used to forecast the affinities and patterns of binding of small molecules, or ligands, to a target protein. It uses a scoring function along with an effective search algorithm to forecast the ligands' binding affinities and modes of binding. It can be used as a command-line tool, but there are also graphical interfaces available, such as PyMOL plugin

DockingPie. DockingPie 1.2 is a platform made especially for the investigation of molecular docking. It incorporates other software that enable the user to prepare the protein structures and ligands, add hydrogens to perform docking. Its goal is to make docking data from molecular docking simulations easier to visualize and analyze. The plugin offers a graphical user interface in PyMOL that also enables users to load and examine docking poses produced by docking software [82].

Modular Synthon-Based Approaches

The V-SYNTHES algorithm creates molecules from a small set of fragments and uses a modular, hierarchical approach to docking-based screening. It reduces the requirement for specialized synthesis. This method is able to achieve screening large chemical regions with great efficiency. Another method that relies on docking individual fragments before combining them into forming the final molecule is Chemical Space docking [83].

Hybrid Computational Approaches

Another approach to drug discovery is combining both physics-based and also data driven methods, in order to take advantage of the unique strengths of both. An example of this hybrid approach, is using results that have occurred from docking, to train machine learning models for screening [83].

In 2021, AlphaFold2 was introduced. It is a neural network, which is like a network of virtual nodes that are connected by links that can be made stronger or weaker. It is able to predict protein structures, and in most cases its results' accuracy is close to the experimental ones. This network improves the 3D structure prediction accuracy because it uses network architectures and training methods that are based on evolutionary, physical and geometric constraints. Alphafold2 has many layers of virtual nodes and it is considered a "deep learning" algorithm. The data that was used to train the network came from PDB [73].

The network in Alphafold2 functions in two main stages. First, the input data is being processed repeatedly by Evoformer blocks, which are a set of layers and operations that process the protein sequence. At this stage, the network starts forming a rough idea of the protein's structure and refines it continuously. In the second stage, a 3D structure is added, by rotating and positioning each part of the protein [73].

Shortly after the release of Alphafold, RoseTTAfold was released by the Baker Lab. RoseTTAFold is a "three-track" neural network that analyzes protein structures by

looking at three things at once: patterns in protein sequences, how the amino acids interact, and the possible 3D shape of the protein. This design lets information move between the one-, two-, and three-dimensional levels, helping the network understand how the chemical components of a protein relate to its final folded structure [84]. Earlier this year, AlphaFold3 was presented, with new capabilities such as prediction of complexes of proteins, nucleic acids, small molecules, ions, and modified residues [85].

3. Methods and Materials

3.1 PyMOL - Visualization

PyMOL is an open source molecular visualization tool. It can produce high-quality 3D images of small molecules and biological macromolecules, such as proteins [86]. We used this software (version 2.5) to visualize the different receptor refined models, ligands and also, explore the rotameric conformations of residues by using the Mutagenesis Wizard.

We studied the structure of GPR21 in active state, using Protein Data Bank (PDB) code 8HMV (Fig. 22). The model contains the GPR21-Gs complex without a ligand present, produced by Cryo-EM (resolution: 2.91 Å- best one, compared to the rest available structures).

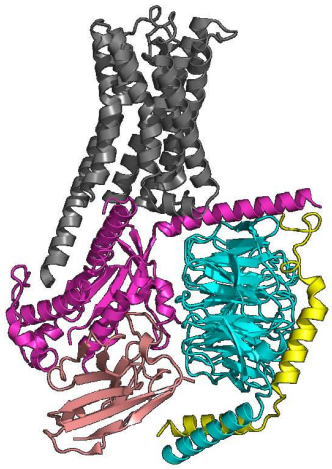

Figure 22: 8HMV - Structure of GPR21-Gs complex

In this structure, we included the disulfide bond between C102$^{3.25}$ and C181^{ECL2} (Fig. 23) which was mentioned in the relevant paper [7] but not shown in the 8HMV model.

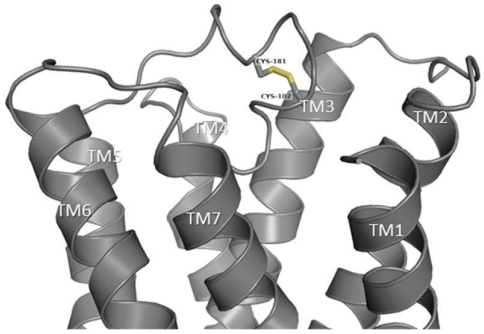

Figure 23: Disulfide bond between CYS102-CYS181

The 3D structures for GRA2 and UCSF924 (Fig 24), were downloaded from PubChem.

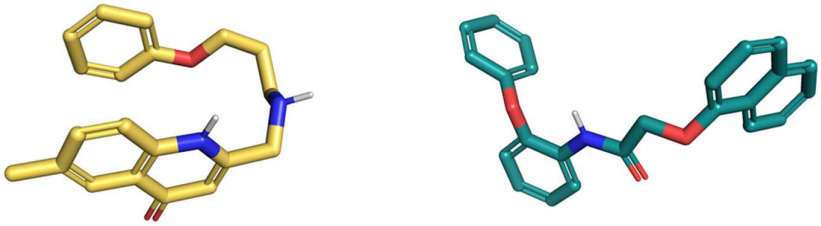

Figure 24: 3D structures of UCSF924 (left-yellow) and GRA2 (right- teal)

3.2 MODELLER – Loop Refinement

With MODELLER, we explored the refinement of three different regions of ECL2:

- Immersed region 169-178
- Cap region 179-186
- Whole loop 169-186

A python script was used for each region (Immersed- Cap- Whole), where the desired amino acid range was specified, respectively, while maintaining the disulfide bond between C102$^{3.25}$ and C181^{ECL2}. We produced twenty refined models for each region, and with a second python script we calculated the Discrete Optimized Protein Energy (DOPE) scores for each one for later evaluation. DOPE is a statistical potential that depends on atomic distance. It calculates the energy of a protein structure based on atomic distances and compares them to a reference state. This state assumes that the protein consists of non-interacting atoms inside a uniform spherical space, with the sphere's radius matching the protein's size. This helps account for the natural shape and size of proteins when evaluating how accurate or stable a given protein model is. The models with the lowest DOPE scores are considered to be better [87]. TABLES 3 and 4 present the scripts for immersed region as an example. (See Appendix 1 for the other two regions).

TABLE 3: Script for loop refinement of immersed region of GPR21, while maintaining the disulfide bond. Highlighted (gray) parts are modified accordingly to produce the desired refined region

```
from modeller import *
from modeller.automodel import *

env = environ()
```

```
class MyLoop(loopmodel):
    def select_loop_atoms(self):
        return selection(self.residue_range('169:A', '178:A'))
    def special_patches(self, aln):
        self.patch(residue_type='DISU',
                   residues=(self.residues['102:A'],
                             self.residues['181:A']))

m = MyLoop(env,
           inimodel='8hmva_disu.pdb',
           sequence='8hmva_disu_im')
m.loop.starting_model = 1
m.loop.ending_model = 20
m.loop.md_level = refine.very_fast
m.make()
```

TABLE 4: Script for DOPE score regarding immersed region refined models. Highlighted (gray) parts are modified according to the refined region

```
from modeller import *
from modeller.scripts import complete_pdb

# Create a MODELLER environment
env = Environ()

# Load the topology and parameters
env.libs.topology.read(file='$(LIB)/top_heav.lib')
env.libs.parameters.read(file='$(LIB)/par.lib')

# List of PDB filenames
pdb_filenames = [
    "8hmva_disu_im.BL00010001.pdb",
    …
    "8hmva_disu_im.BL00200001.pdb"
]

# Create a list to store DOPE scores
dope_scores = []

# Loop through and score each PDB file
for pdb_filename in pdb_filenames:
    # Create a model for scoring
    mdl = complete_pdb(env, pdb_filename)

    # Select all atoms in the first chain
    atmsel = selection(mdl.chains[0])

    # Calculate DOPE score
    score = atmsel.assess_dope()

    # Append the score to the list
    dope_scores.append((pdb_filename, score))

# Output the results to a text file
```

```
#output_filename = "dope_scores.txt"
#with open(output_filename, "w") as f:
 #    for filename, score in dope_scores:
  #       f.write(f"File: {filename}, DOPE Score: {score}\n")


# Finalize and clean up
env.io.atom_files_directory = ['./']  # Set the directory where
your PDB files are located
env.io.hetatm = True  # If your PDB files contain HETATM records
env.io.water = True   # If your PDB files contain water molecules
env.io.stop_at_pdb = True

env.cleanup()
```

3.3 Cavity search

CavityPlus is a web server that allows the user to search for cavities within protein structures. The tool CAVITY uses the protein 3D structure as input, in order to identify potential binding sites on the surface of the protein, and then ranks them. It is possible to achieve further analysis of the discovered cavities by 3 other submodules, CavPharmer, CorrSite, and CovCys. CavPharmer uses a receptor-based pharmacophore modeling application to extract the pharmacophore properties of the cavities. Allosteric binding sites can be found by using CorrSite, while CovCys can automatically find druggable cysteine residues [88].

Each refined model was loaded on the web server CavityPlus as a pbd file to search for potential binding sites. We used the tool CAVITY with default settings (Fig. 25).

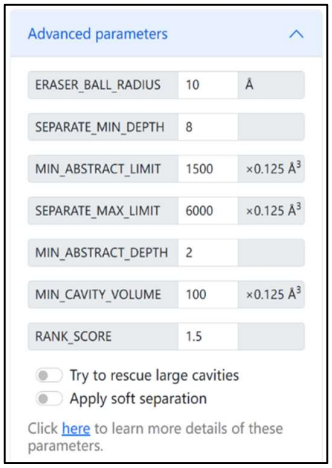

Figure 25: Default settings for CAVITY

This tool is used to find cavities (or pockets) in proteins where molecules might bind. First it maps the protein onto a 3D grid with points spaced 0.5 Å apart. A sphere that

is representing a large molecule, moves around the protein surface. If the sphere touches a grid point, that point is marked as "outside", and all the points that were not touched by the sphere are considered potential cavity spaces (vacant points). Vacant points are grouped into layers based on how deep they are inside the protein, starting with layer 1 which is the closest to the outside, followed by layer 2 that is deeper, and so on. The layers are then gradually erased from the outside in, which can split cavities into smaller sub-cavities. This is the shrinking stage. After that, the cavities are expanded back to their original size, but the sub-cavities remain separate. This is called the expanding stage. All of the parameters that can be adjusted by the user are listed in TABLE 5 [89].

TABLE 5: CAVITY tool parameter explanation

Parameter Name	**Explanation**
ERASER_BALL_RADIUS	The radius of the sphere used to simulate molecule binding. It influences the shape of the cavity's external surface. Larger radius results in a flatter surface.
MIN_ABSTRACT_DEPTH (MAD)	Determines the depth at which the shrinking process stops. Higher values stop shrinking earlier, preserving deeper layers.
MIN_ABSTRACT_DEPTH (MAD)	Determines the depth at which the shrinking process stops. Higher values stop shrinking earlier, preserving deeper layers.
SEPARATE_MIN_DEPTH (SMD)	Controls the minimum number of shrinking steps. It also influences whether a cavity is divided into sub-cavities. Lower values stop shrinking sooner.
MIN_ABSTRACT_LIMIT (MAL)	Ensures that the shrinking process continues until the cavity is smaller than this set value. Higher values stop shrinking sooner if the cavity is still large.
SEPARATE_MAX_LIMIT (SML)	Similar to MAL, but it controls the maximum size of the final cavity after all shrinking and expansion steps. Higher values continue shrinking if the final cavity is still too big.
MIN_CAVITY_VOLUME	Filters out cavities that are smaller than the specified volume, ensuring that only cavities larger than this value are considered.
RANK_SCORE	Only cavities with a score higher than this threshold are output, helping to filter out less significant cavities.
Rescue Large Cavities	An option to recover large, valid cavities that might have been mistakenly excluded during processing.
Apply Soft Separation	An option to recover parts of a cavity that might have been separated and discarded incorrectly during the shrinking process.

The CAVITY tool presents cavities and ranks them according to ligandability and druggability scores. Ligandability stands for the possibility of designing a small molecule that has high affinity for a specific cavity. How good of a target the specific cavity might be, is described by the druggability value. The CavityScore is influenced by cavity surface area and volume, pocket size, the surface that is covered by hydrophobic residues, and lastly hydrogen-bond-forming surface area [88].

3.4 Docking

To perform molecular docking we used the PyMOL plugin DockingPie (Version: 1.2.1) [82]. DockingPie is a platform that integrates four external docking tools: Vina, Smina, RxDock, and ADFR along with several chemoinformatics python modules like AutoDockTools, Openbabel, sPyRMSD and other external tools like sdsorter. Here we will use Vina. For every model we followed the process described below:

Protein and Ligand Preparation

The pbd files are loaded in PyMOL and then the action “remove waters” was performed. The receptor and ligand tabs in the plugin are similar and the preparation steps are shown in Figure 26. First, we import the object from PyMOL (steps 2, 3), select it (step 4), then add hydrogens (step 5). Then the receptor and ligands are generated and set in the “docking” tab (steps 6, 7, 8).

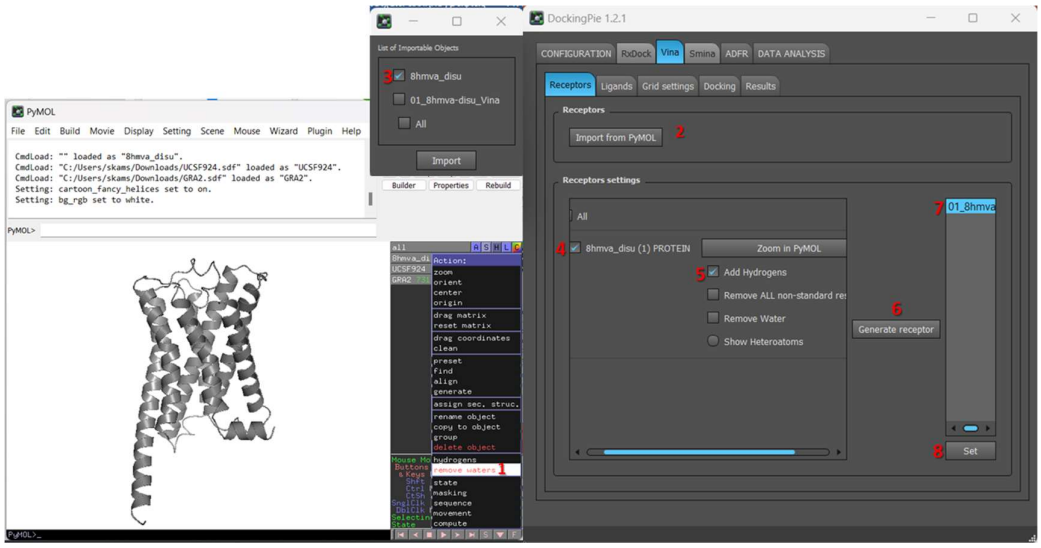

Figure 26: Import from PyMOL and receptor/ligand preparation steps for docking. In this figure the “Receptors” tab (similar to the “Ligands” tab) is shown.

Grid box settings

We used H174 as the center of the grid box (x:156, y:150, z:115), and set grid dimensions to x:20 y:20 z:20 for the docking sessions that were performed by

DockingPie (Fig. 27), and x:30 y:30 z:30 for sessions that were performed by AutoDock Vina. In every session we set the number of desired poses to 10, exhaustiveness to 8 and energy range to 3.

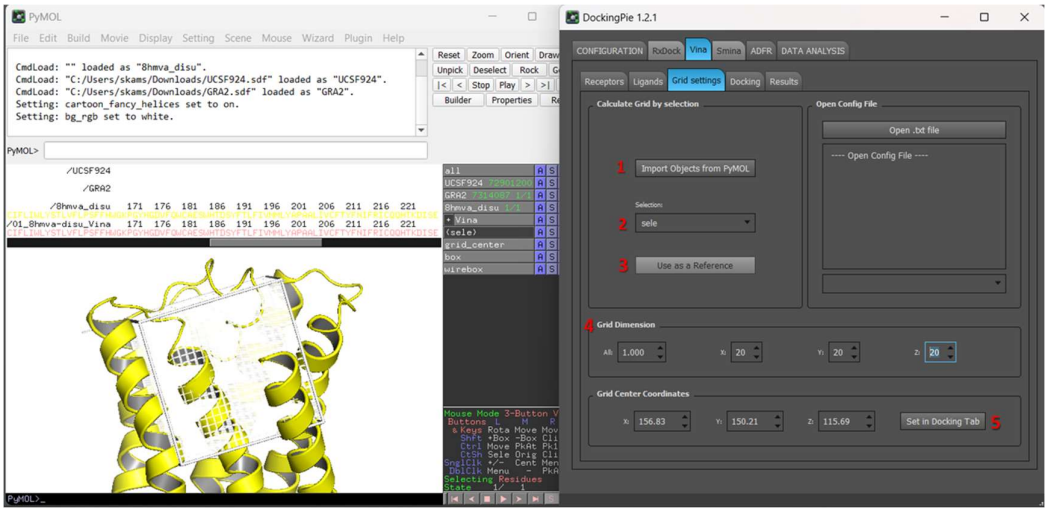

Figure 27: Grid settings for docking. (1) Importing the reference object from PyMOL, (2) Make selection from the drop down list, (3)Use the objects coordinates for the center, (4) Set the grid dimensions, (5) Set it in the docking tab

3.5 Alphafold Models

We explored three different alternatives by using ColabFold v1.5.5: AlphaFold2 using MMseqs2 [90] Then, we proceeded to explore for cavities in a similar manner to the refined models that was described above, and performed molecular docking. In this case, the grid box settings were defined by the CavityPlus results, which included the box size and box center for each cavity of interest.

First, we only used the full sequence of GPR21 without a template sequence. Then we used the same full sequence of GPR21, but this time we used GPR52 structure as a custom template. For the last run, we used the GPR21 without N-Terminus sequence with GPR52 structure as a custom template.

- Full GPR21 Sequence used:
  MNSTLDGNQSSHPFCLLAFGYLETVNFCLLEVLIIVFLTVLIISGNIIVIFVFHCAPLLNHHT
  TSYFIQTMAYADLFVGVSCVVPSLSLLHHPLPVEESLTCQIFGFVVSVLKSVSMASLACISI
  DRYIAITKPLTYNTLVTPWRLRLCIFLIWLYSTLVFLPSFFHWGKPGYHGDVFQWCAES
  WHTDSYFTLFIVMMLYAPAALIVCFTYFNIFRICQQHTKDISERQARFSSQSGETGEVQ
  ACPDKRYAMVLFRITSVFYILWLPYIIYFLLESSTGHSNRFASFLTTWLAISNSFCNCVIYSL
  SNSVFQRGLKRLSGAMCTSCASQTTANDPYTVRSKGPLNGCHI

- GPR21 without N-Terminus sequence:
  VNFCLLEVLIIVFLTVLIISGNIIVIFVFHCAPLLNHHTTSYFIQTMAYADLFVGVSCVVPSLS
  LLHHPLPVEESLTCQIFGFVVSVLKSVSMASLACISIDRYIAITKPLTYNTLVTPWRLRLCIF
  LIWLYSTLVFLPSFFHWGKPGYHGDVFQWCAESWHTDSYFTLFIVMMLYAPAALIVCF
  TYFNIFRICQQHTKDISERQARFSSQSGETGEVQACPDKRYAMVLFRITSVFYILWLPYII
  YFLLESSTGHSNRFASFLTTWLAISNSFCNCVIYSLSNSVFQRGLKRLSGAMCTSCASQTT
  ANDPYTVRSKGPLNGCHI

4. Results

Recently, structures of orphan GPR21-G-protein complexes were presented. A section of the ECL2 of GPR21 reaches deep in the orthosteric pocket, causing receptor activation. This observation explains the high basal activity of GPR21, without the presence of a ligand. Consequently, it was an area of interest to investigate the different conformations of ECL2, in order to explore the different pockets formed by each one. We started with the 8HMV structure, and proceeded with three loop refinement scenarios:

- The immersed region of ECL2 (Fig. 28A), referring to the amino acids 169-178, that reach in the orthosteric pocket, was refined, in order to explore the possibility of forming a cavity where GRA2 and/or UCSF924 could interact with the receptor, disrupting its self-activation.
- the cap region, referring to the residues 179-186 (Fig. 28B)
- The entire ECL2 (Fig. 28C)

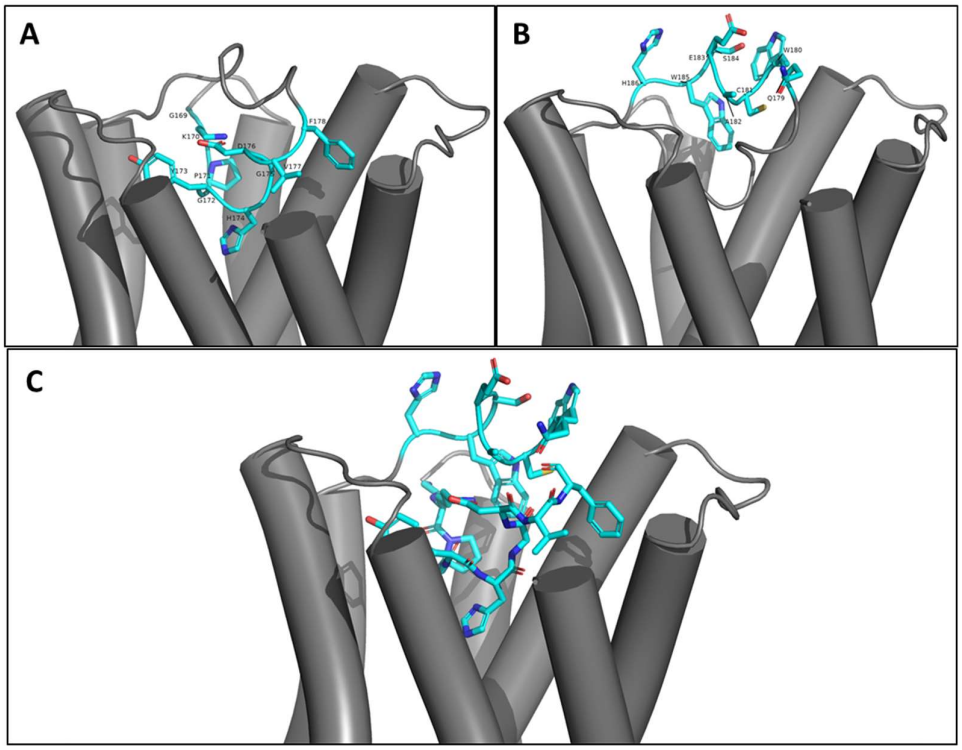

Figure 28: Regions of ECL2 that are going to be refined in each scenario are highlighted with light blue. (A) Immersed region, (B) Cap region, (C) Whole loop

4.1 Refined Models and Evaluation

With MODELLER we produced 20 models for each refinement scenario. We explored the different pockets formed by different loop conformation with CavityPlus, and ranked the refined models. They were evaluated by their DOPE scores produced by MODELLER, and by the druggability and DrugScore of their cavities, which were calculated by CavityPlus. Out of the 60 refined models that were produced by MODELLER, we selected the top 5 per refined region based on DOPE Score, DrugScore and Druggability score, and therefore we had 5 different conformations of the immersed region (amino acids 169-178) (Fig. 29), 5 of the cap region (amino acids 179-186) (Fig. 30), and 5 of the whole ECL2 (amino acids 169-189) (Fig. 31).

$H174^{ECL2}$ is the amino acid that reaches the deepest into the orthosteric pocket of GPR21 interacting with surrounding amino acid side chains, and therefore keeping ECL2 inside the pocket. We chose to produce refined models of the immersed region to explore the possibility of having a model where the ECL2 is positioned in a way that leaves enough space for small molecules like GRA2 and UCSF924, to enter in the orthosteric pocket. This could displace ECL2, and disrupt the receptor's activation.

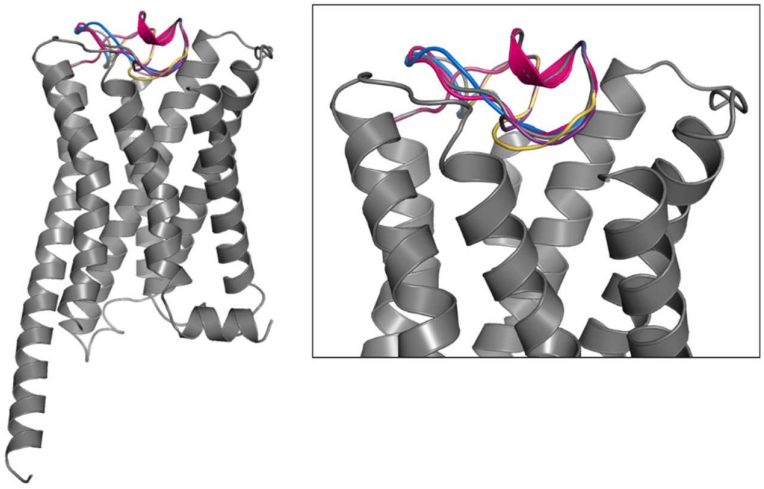

Figure 29: Top 5 refined immersed region models

The refined cap region models (Fig. 30) were produced in order to examine if there could be a pocket on the top side of the receptor, where GRA2 and UCSF924 could bind.

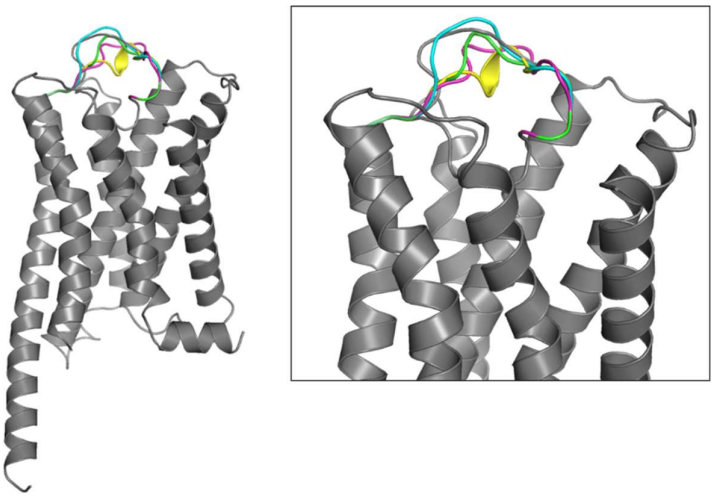

Figure 30: Top 5 refined cap region models

Finally, the third scenario of the whole ECL2 refinement aimed to explore different loop conformations and potential cavities, while maintaining the conserved disulfide bond between C181 and C102 in TM3.

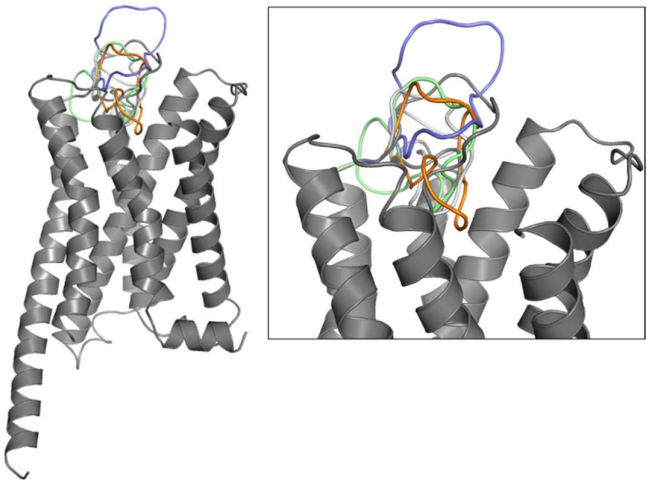

Figure 31: Top 5 refined whole loop models

Tables 6-8 show the top 5 model for each refined region (Immersed- Cap- Whole loop). For each model there is information about the DOPE Score that was calculated by MODELLER. It is used to assess energy of the models, so the ones with lower DOPE scores are considered better. Furthermore, there is information about the druggability (weak/medium/strong) of the cavity that was detected from CavityPlus. In the last columns there is the DrugScore of each cavity, which was also calculated by CavityPlus.

The immersed region and whole loop refined models show cavities with strong druggability and higher DrugScore compared to the cap region refined ones.

TABLE 6 DOPE Scores/ Druggability/ DrugScore/ for models with refined immersed region

Model	Immersed Region (169-178)		
	DOPE Score	Druggability	DrugScore
immersed 1	-32475.77	Strong	1643.00
immersed 14	-32458.23	Strong	1640.00
immersed 20	-32394.85	Strong	1687.00
immersed 5	-32356.52	Strong	1500.00
immersed 12	-32348.17	Strong	1124.00

TABLE 7 DOPE Scores/ Druggability/ DrugScore/ for models with refined cap region

Model	Cap Region (179- 186)		
	DOPE Score	Druggability	DrugScore
cap 2	-32969.88	Medium	-61.00
cap 10	-32944.01	Medium	158.00
cap 4	-32863.50	Medium	55.00
cap 19	-32825.12	Medium	11.00
cap 3	-32818.45	Medium	72.00

TABLE 8 DOPE Scores/ Druggability/ DrugScore/ for models with refined whole loop

Model	Whole Loop (169-186)		
	DOPE Score	Druggability	DrugScore
whole 13	-32635.56	Strong	612.00
whole 19	-32446.96	Strong	1700.00
whole 18	-32194.00	Strong	1125.00
whole 4	-32069.14	Strong	880.00
whole 10	-31786.29	Strong	2559.00

4.2 Docking Scores and Top Poses

After determining the top 5 models for each refined region, we performed molecular docking with GRA2 and UCSF924. All the poses of both ligands were visualized in PyMOL to examine the ligands' interactions with the surrounding amino acid side chains.

4.2.1 Immersed Region

For the immersed region refined models, we performed docking with GRA2 and UCSF924 and the docking scores are presented for each ligand separately, per model

and per pose. For GRA2, we found that the best docking scores were achieved with model 1 (poses 1-4) and model 14 (pose 1) (Fig. 32).

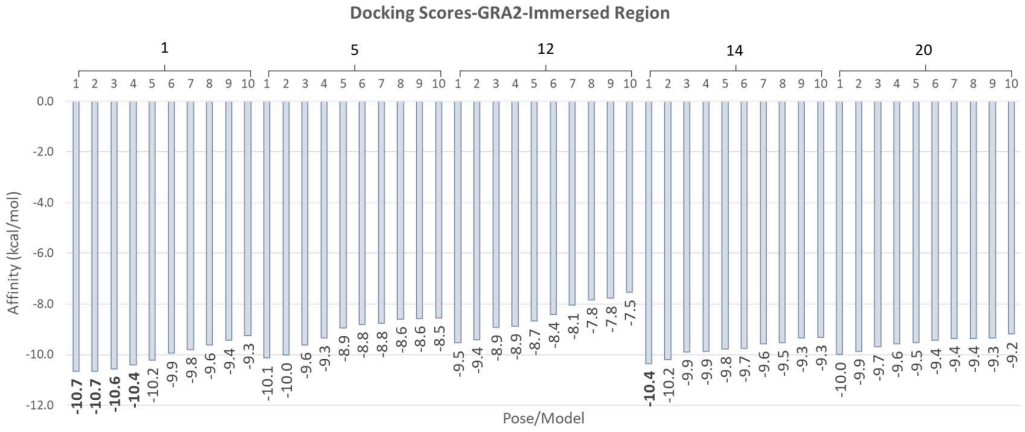

Figure 32: Docking scores for immersed region refined models with GRA2. There are 10 poses for each model, the top scores are in bold.

In pose 3 (Fig. 33), we found that there are two H-bonds between GRA2 and the side chains of Y268$^{6.51}$ and S110$^{3.33}$. There is also π-π interaction between two aromatic rings of GRA2 and F191$^{5.39}$.

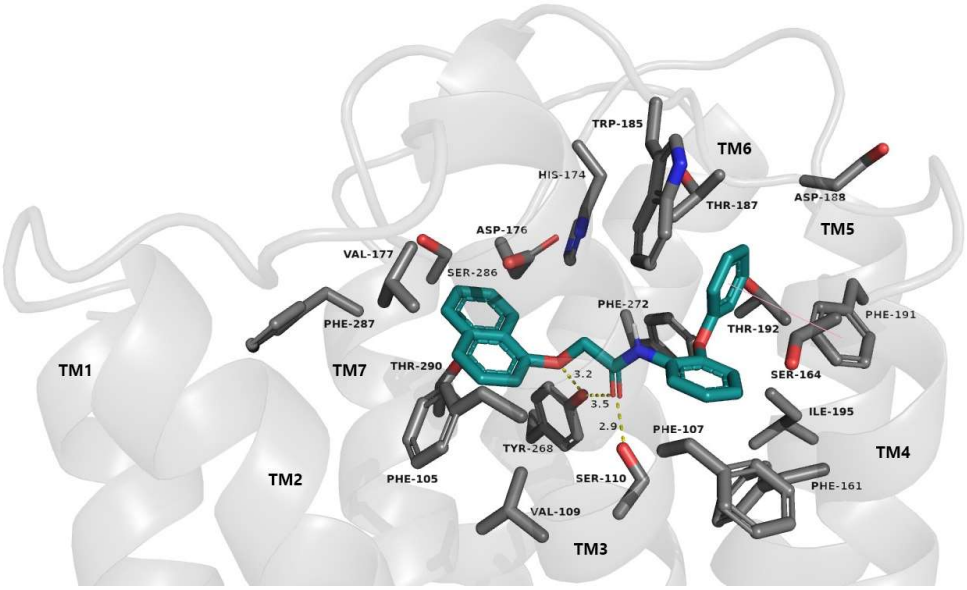

Figure 33: Pose 3 for immersed region refined model 1. Hydrogen bonds are shown with yellow dotted lines, π- π interactions are represented by a pink line, and GRA2 is shown in teal.

However, for UCSF924 models 5 (poses 1-4) and 12 (pose1) produced the best scores (Fig. 34).

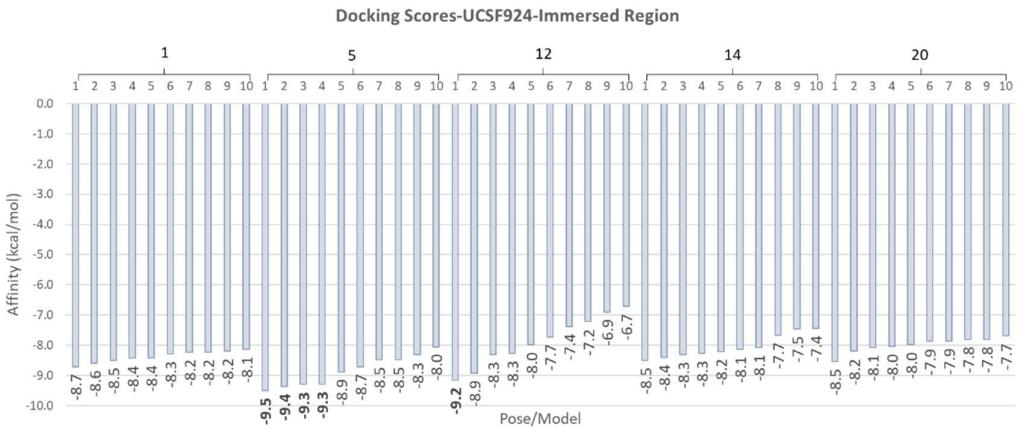

Figure 34: Docking scores for immersed region refined models with UCSF924. There are 10 poses for each model, the top scores are in bold.

For the first pose of UCSF924 with immersed region refined model 12 (Fig. 35), we observed H-bonds between the O atom of UCSF924's bicyclic group and K113$^{3.36}$ and Y200$^{5.48}$ side chains. There is another hydrogen bond interaction between H174^{ECL2} side chain and NH of the bicyclic group of UCSF924. In addition, we observed π-π interaction between the ligands aromatic ring and W185^{185}.

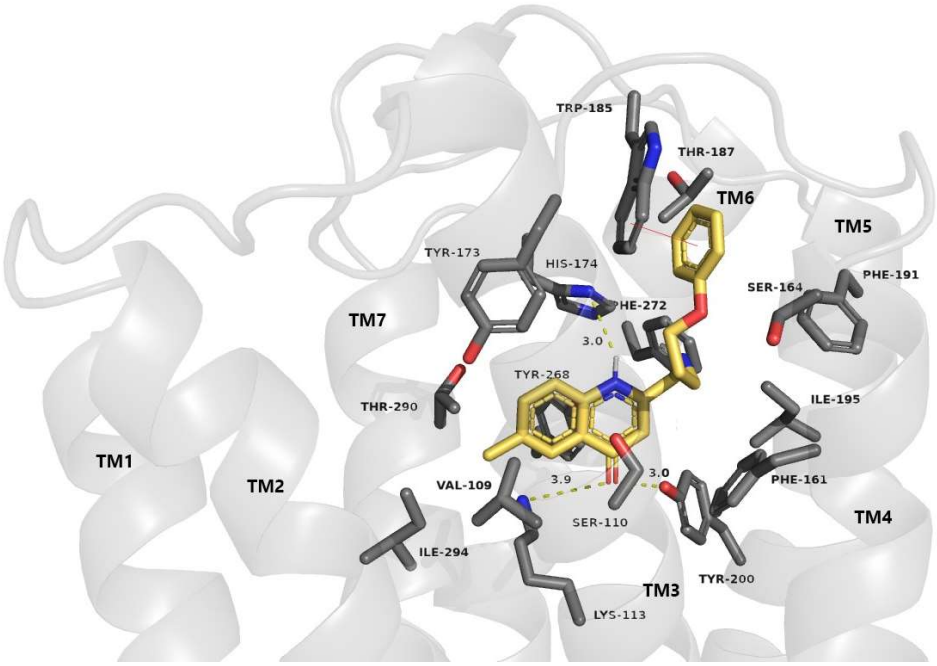

Figure 35: Pose 1 for immersed region refined model 12. Hydrogen bonds are shown with yellow dotted lines. UCSF924 is shown in yellow.

4.2.2 Whole Loop

Model 19, from the whole loop refined models, gave the best docking scores for GRA2 with poses 1-5 (Fig. 36), while for UCSF924 we received better scores from model 4 (poses 1-4), followed by the first pose of model 19 (Fig. 38).

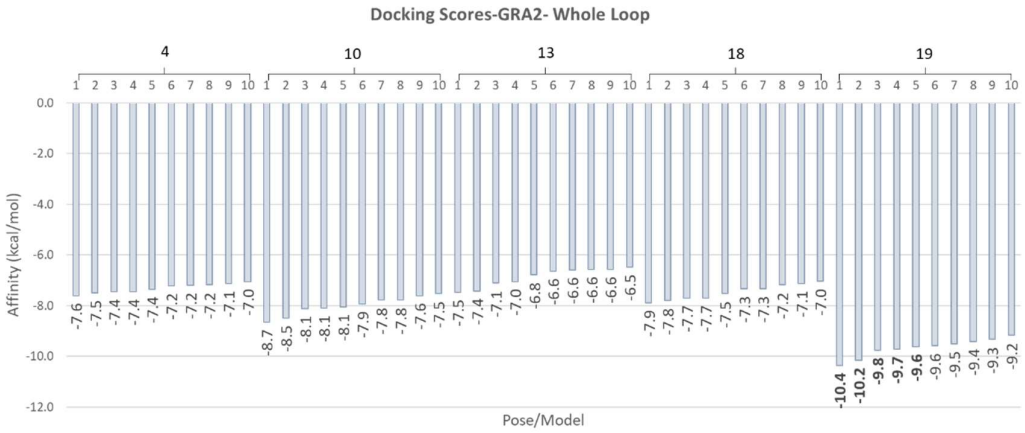

Figure 36: Docking scores for whole loop refined models with GRA2. There are 10 poses for each model, the top scores are in bold

In the 2nd pose of GRA2 in the whole loop refined model 19 (Fig. 37), we see that the aromatic ring and the bicyclic group of the ligand are in hydrophobic environments, created by hydrophobic amino acids V84$^{2.59}$, P85$^{2.60}$, F105$^{3.28}$, V109$^{3.32}$, V177^{ECL2}, F178^{ECL2}, W291$^{7.40}$, I294$^{7.43}$. The amino acid W291$^{7.40}$ also creates a hydrogen bond with GRA2.

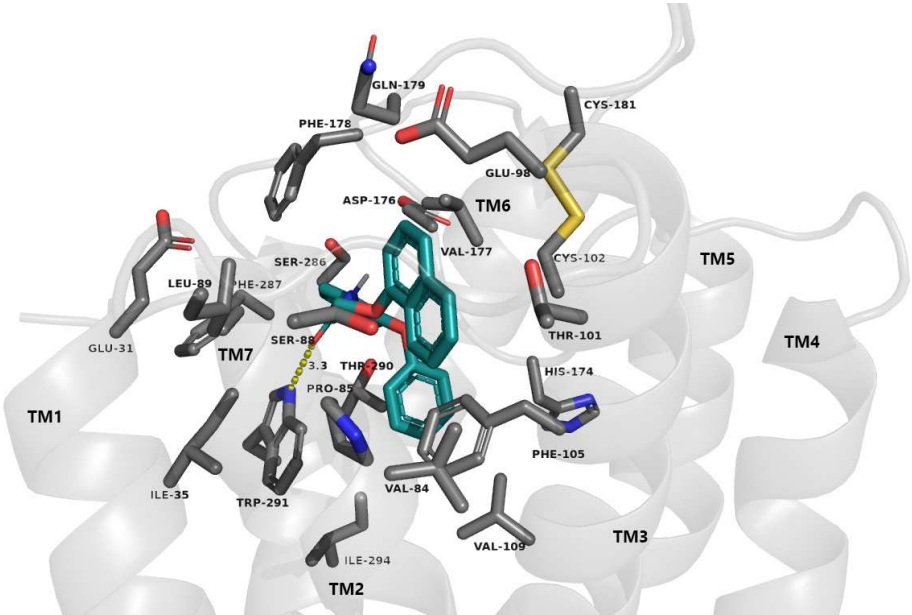

Figure 37: Pose 2 for whole loop refined model 19. Hydrogen bonds are represented with yellow dotted lines. GRA2 is shown in teal.

Docking scores for the whole loop refined models are shown in Figure 38. Refined models 4 (poses 1-4) and 9 (pose 1) had the best scores.

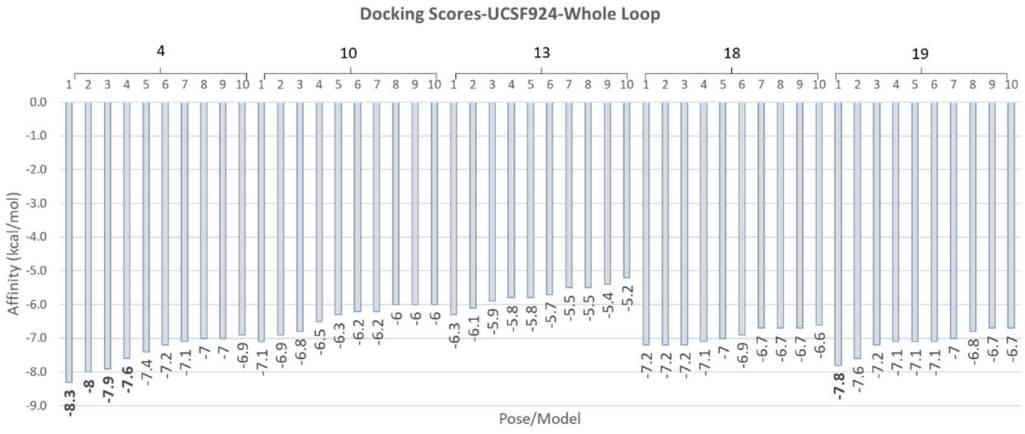

Figure 38: Docking scores for whole loop refined models with UCSF924. There are 10 poses for each model, the top scores are in bold

In the pose shown in Figure 39, UCSF924 is forming a hydrogen bond with the side chain of T290$^{7.39}$. Hydrophobic interactions are observed between the aromatic ring of the ligand and Y271$^{6.54}$, F272$^{6.55}$, F178^{ECL2}.

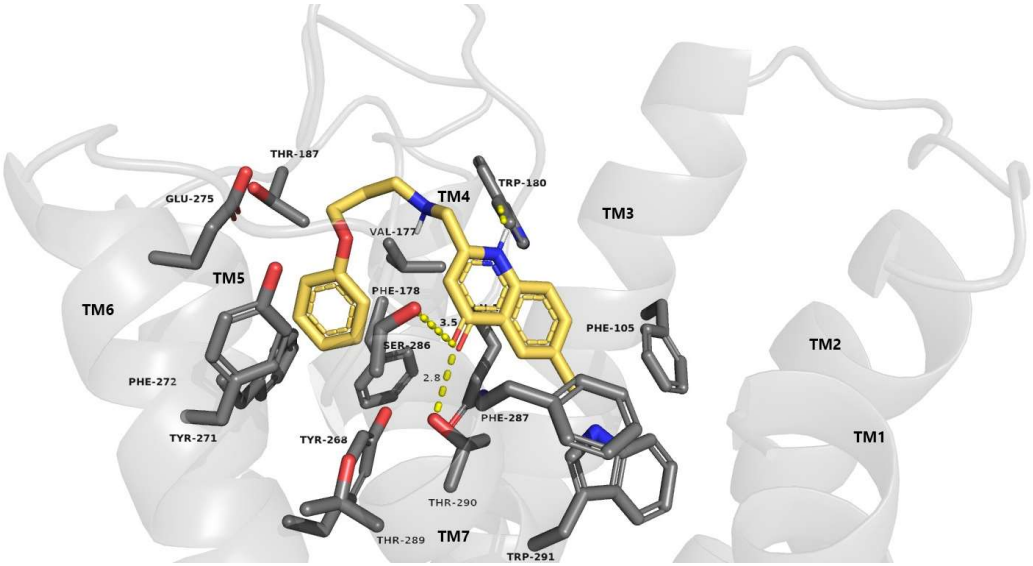

Figure 39: Pose 1 for whole loop refined model 4. Hydrogen bonds are shown with yellow dotted lines. UCSF924 is shown in yellow.

4.2.3 Cap Region

When examining the docking results for the cap region (Fig. 40, Fig. 42), we found that in many of the produced poses, the ligand was located outside of the desired cavity (bars shown in white). From the poses where the ligand is found in the desired location (bars shown in blue), model 2 (pose 9) (Fig. 41) gave the best score for GRA2, and model 4 (poses 7 (Fig. 43) and 8) for UCSF924.

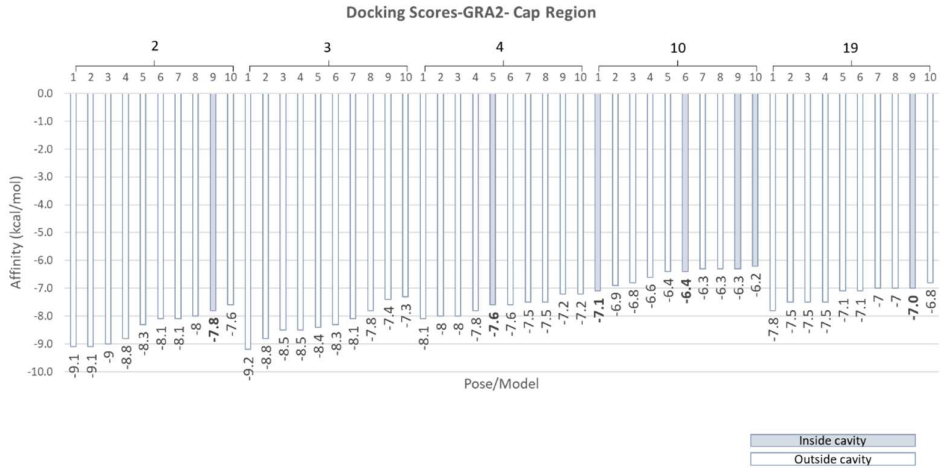

Figure 40: Docking scores for cap region refined models with GRA2. There are 10 poses for each model, the top scores are in bold

For the cap region refined model 2, GRA2 is forming a hydrogen bond with the side chain of R283$^{7.32}$. The aromatic ring of the ligand forms hydrophobic interactions with L89$^{2.64}$, F178^{ECL2}, and the bicyclic group of GRA2 with V177^{ECL2}, F287$^{7.36}$, W180^{ECL2} (Fig. 41).

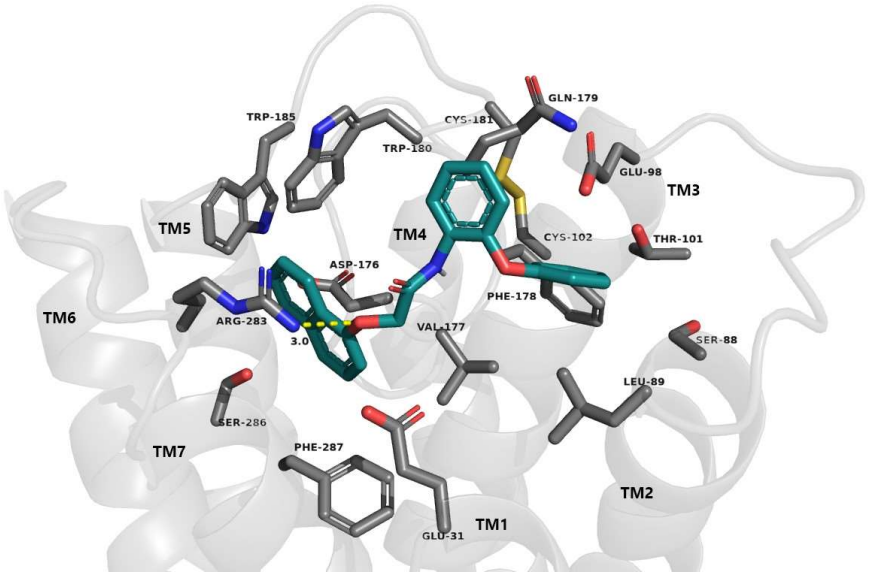

Figure 41: Top pose (9) for cap region refined model 2. Hydrogen bonds are represented with yellow dotted lines. GRA2 is shown in teal.

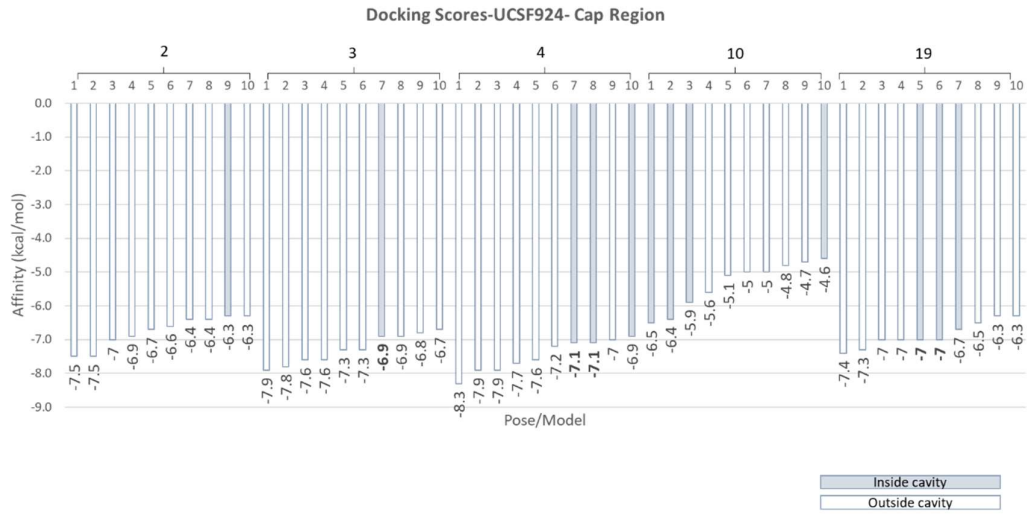

Figure 42: Docking scores for cap region refined models with UCSF924. There are 10 poses for each model, the top scores are in bold

There are H-bonds between the O atom of UCSF924's bicyclic group and D176^{ECL2} and H186^{ECL2}. The side chain of R283$^{7.32}$ is forming another hydrogen bond with the ligand's carbonyl group (Fig. 43).

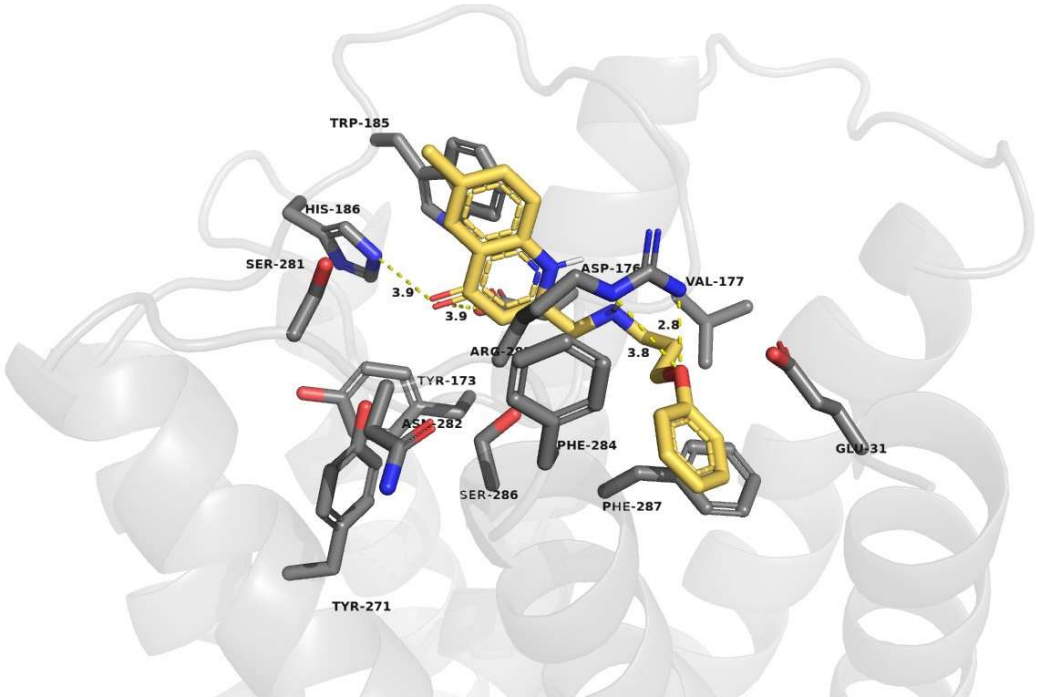

Figure 43: Top pose (7) for whole loop refined model 4. Hydrogen bonds are shown with yellow dotted lines. UCSF924 is shown in yellow.

4.3 ARG283 Rotamers – Docking

GPR21 has a great structural similarity to GPR52. Since GPR52 was found to have a side pocket close to the receptor surface that could interact with ligands, we examined GPR21 for a similarly located cavity. We found that the R283$^{7.32}$ side chain in the 8HMV structure was shifted inward, leaving less space for a cavity. Therefore we looked into the different rotational conformations of this side chain, to determine whether a "better" cavity could be formed in order to perform molecular docking with the two potential ligands.

We used CavityPlus to explore the different cavities that may form on the surface of GPR21 (8HMV structure), for different rotamers of R283$^{7.32}$ since it presents a high structural similarity to GPR52 which has a side pocket that is surrounded by TM1, TM2 and TM7. In the cavity shown below (Fig. 44), we noticed that the side chain of R283$^{7.32}$ is shifted inwards, closing the space of the cavity which resulted in a weak druggability score. Therefore, we examined the different rotational conformations of this side chain.

Figure 44: GPR21 (green), Cavity (pink) and R283$^{7.32}$ (white)

With PyMOL Mutagenesis wizard we found the conformations of R283$^{7.32}$ and examined all of them to investigate if there are any changes in the druggability and DrugScore of the cavity shown in Figure 44. One of the rotational conformations (Fig. 45) of R283$^{7.32}$ gave us a slightly more open cavity with medium Druggability, with which we proceeded to perform molecular docking, with both GRA2 and UCSF924.

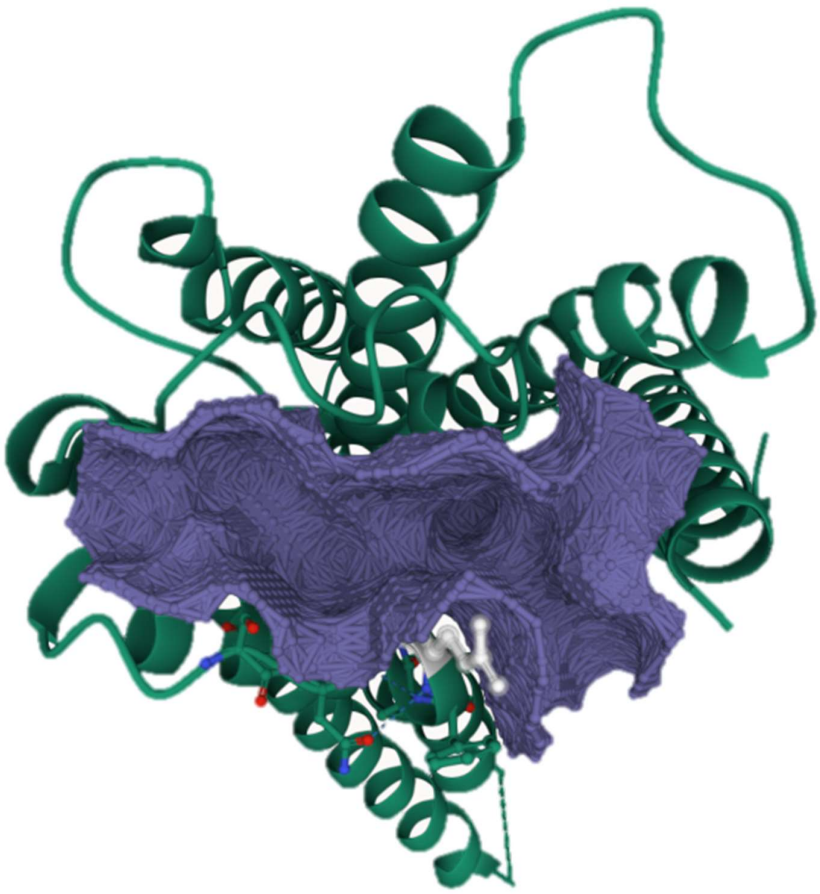

Figure 45: GPR21 (green), new cavity (purple) and shifted R283$^{7.32}$ (white)

The docking scores are presented in Figure 46 even though they are not better than the ones produced by the refined models. Both GRA2 and UCSF924 gave similar docking scores to this cavity.

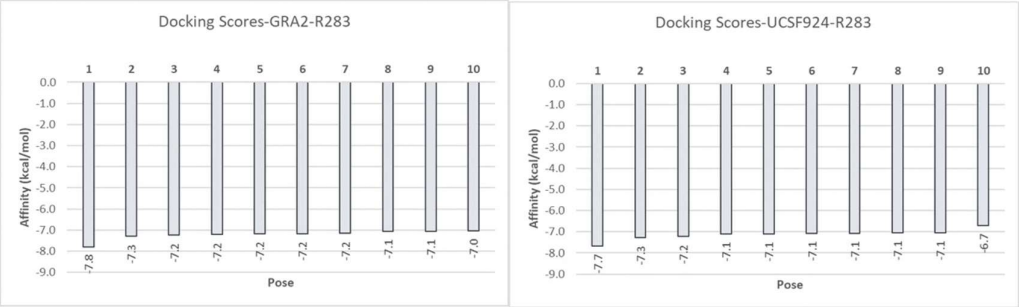

Figure 46: Docking scores for the cavity created by slightly shifting R283 side chain. GRA2 results are on the left and UCSF924 on the right

4.4 Alphafold Models - Docking

Finally, we looked at Alphafold's ability to create models of GPR21 with a cavity resembling that of the GPR52 side pocket in order to perform docking with GRA2 and UCSF924 and explore the possibility of them binding in this pocket.

Apart from the predicted models, Alphafold produces three types of graphs. The first is a Predicted Alignment Error plot that refers to the confidence in the relative positions of different residues within a predicted protein structure. Specifically, PAE indicates the expected positional error between two residues (X and Y) in the predicted structure, expressed in Angstroms (Å). This error estimate is based on how well AlphaFold2 expects the structure to match the actual protein when aligned on one of those residues. Low PAE is a high confidence indicator, while high PAE shows low confidence in the predicted positions [91].

The sequence coverage graph visualizes how well sequences from a database align with the input sequence. The red to blue color scale represents the identity score, with red indicating low, and blue indicating high sequence similarity. The sequences are positioned from bottom to top according to their similarity, low to high, respectively. The white sections represent uncovered regions, and the black line represents the relative coverage of the input sequence to the total number of aligned sequences from the database [91].

Last is the predicted local distance difference test (pLDDT). It gives a confidence score per-residue, indicating the likelihood of the predicted position to be the correct one. The scores scale from 0 (low confidence) to 100 (high confidence). Scores above 90 are very high, 70-90 is moderate, 50-70 is low, and below 50 the scores are considered to be very low [91].

The Alphafold results of GPR21 models without using another structure as a template are shown in Figure 47. There were 5 models that were produced, but none of them

presented a cavity similar, or close to where we would expect to see the GPR52 side pocket.

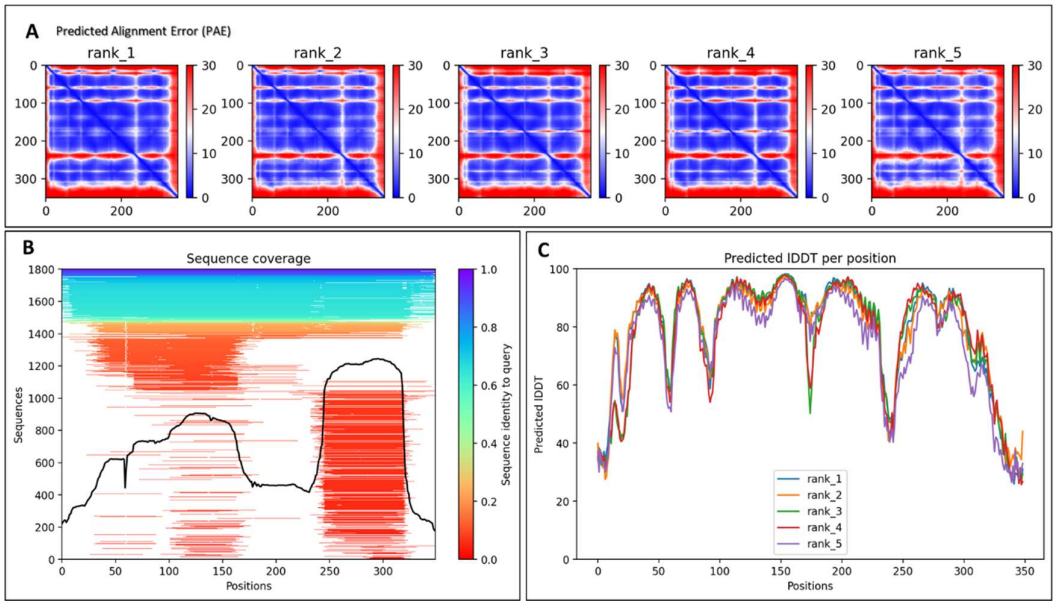

Figure 47: Alphafold results for GPR21 models without a template sequence. A) PAE Plot for each one of the 5 predicted models (rank 1-5). B) Sequence coverage graph. C) Predicted lDDT for each one of the 5 models, rank 1-5 presented with different color lines

In Figure 48, we present the results for the predicted models of GPR21 with the use of GPR52 sequence as template. They seem to be very similar to the results that were produced by the previous run, without GPR52 structure as template.

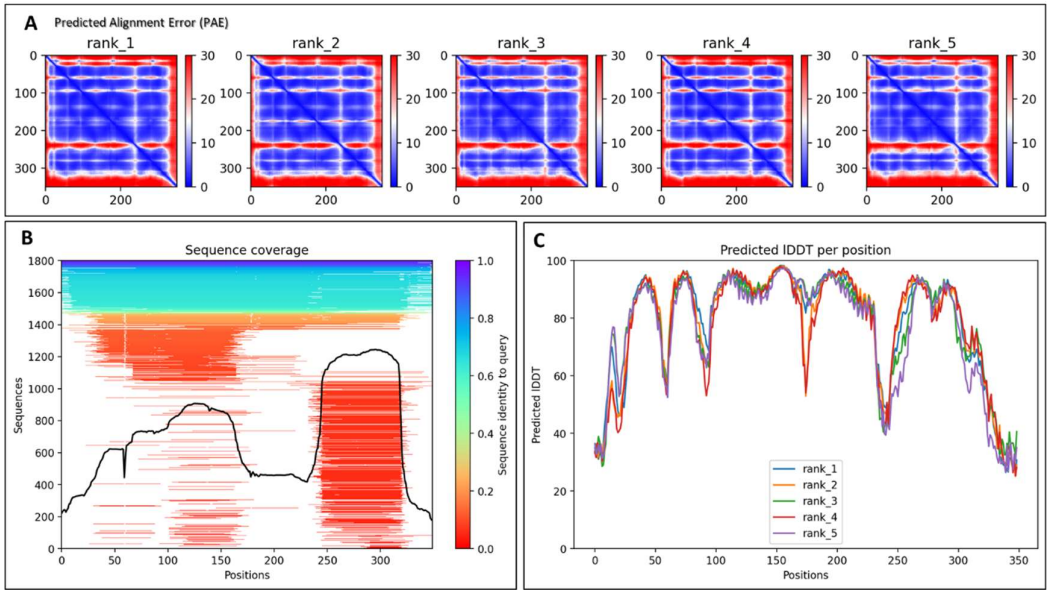

Figure 48: Alphafold results for GPR21 models with the GPR52 sequence as template. . A) PAE Plot for each one of the 5 predicted models (rank 1-5). B) Sequence coverage graph. C) Predicted lDDT for each one of the 5 models, rank 1-5 presented with different color lines

After visualizing the predicted structures (Fig. 49), we notice that a small section of the N-terminus is folded between TM1, TM2 and TM7, covering some of the space that we would anticipate to find the side pocket that is present in GPR52.

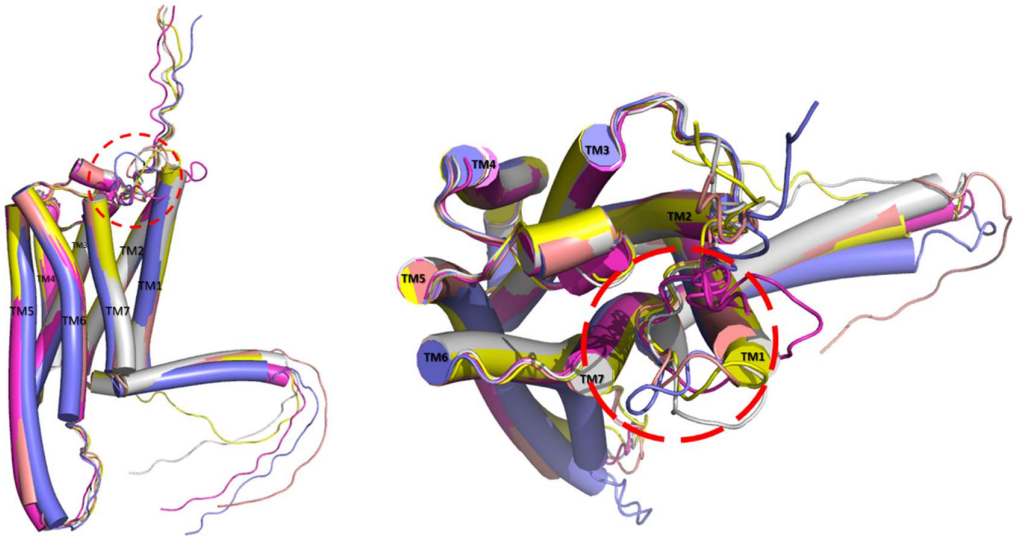

Figure 49: Superposition of the 5 predicted models of GPR21 with the GPR52 sequence as template. The red dotted circle is highlighting the small part of the N-terminus that is shifted inwards.

Therefore, we proceeded with exploring Alphafold predictions of GPR21 models without the N-terminus, using the GPR52 sequence as template (Fig. 50). Even though they still look quite similar, we notice in that the predicted models are show higher confidence, since the N-terminus, an area of low confidence is eliminated (Fig. 50A, Fig. 50C)

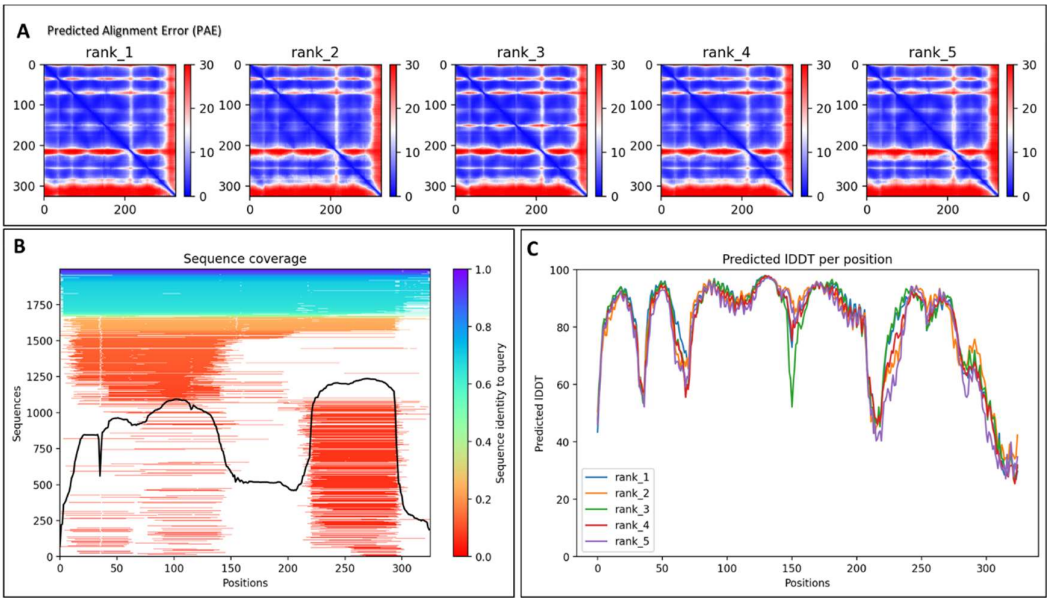

Figure 50: Alphafold results for GPR21 models without the N-terminus, using the GPR52 sequence as template. A) PAE Plot for each one of the 5 predicted models (rank 1-5). B) Sequence coverage graph. C) Predicted lDDT for each one of the 5 models, rank 1-5 presented with different color lines

Out of the 5 models that were produced, rank2 was the one that included a cavity in the area of interest (Fig. 51). The cavity has medium druggability and a DrugScore of 472. The cavity's surface area is 625.75 Å2, and the volume is 648.88 Å3. The box size (Å) is x=17.5, y=14.5, z=17.5 and the center (Å) x=7.25, y=1.75, z=-20.25.

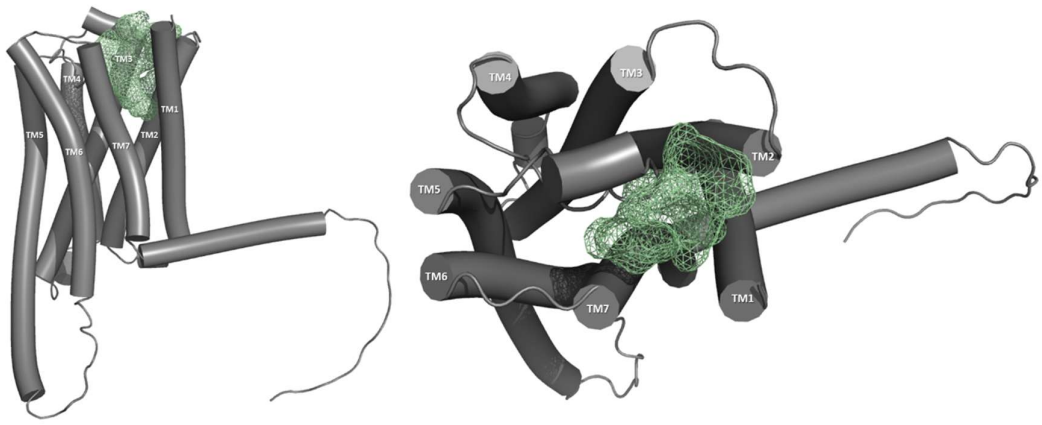

Figure 51: Predicted GPR21 model-rank2 with cavity in the area of interest

We performed molecular docking in this cavity and the scores are presented in Figure 52.

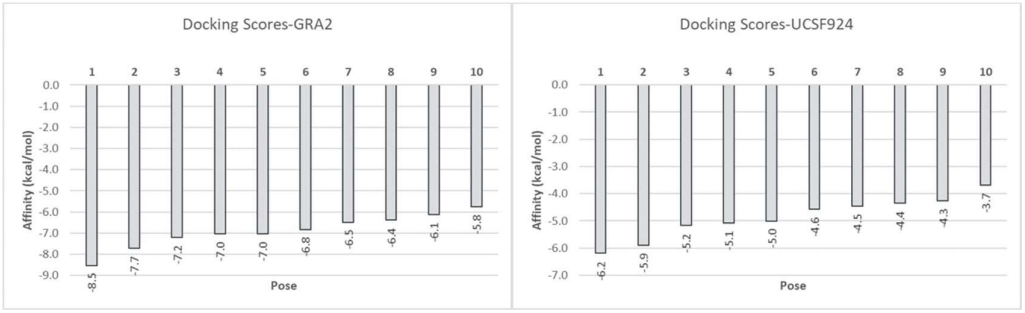

Figure 52: Docking scores for GRA2 (left) and UCSF924 (right) for predicted model-rank 2

In both cases, the ligand was found to be inside of the cavity, forming strong hydrophobic interactions with surrounding amino acids (Fig. 53, Fig. 54).

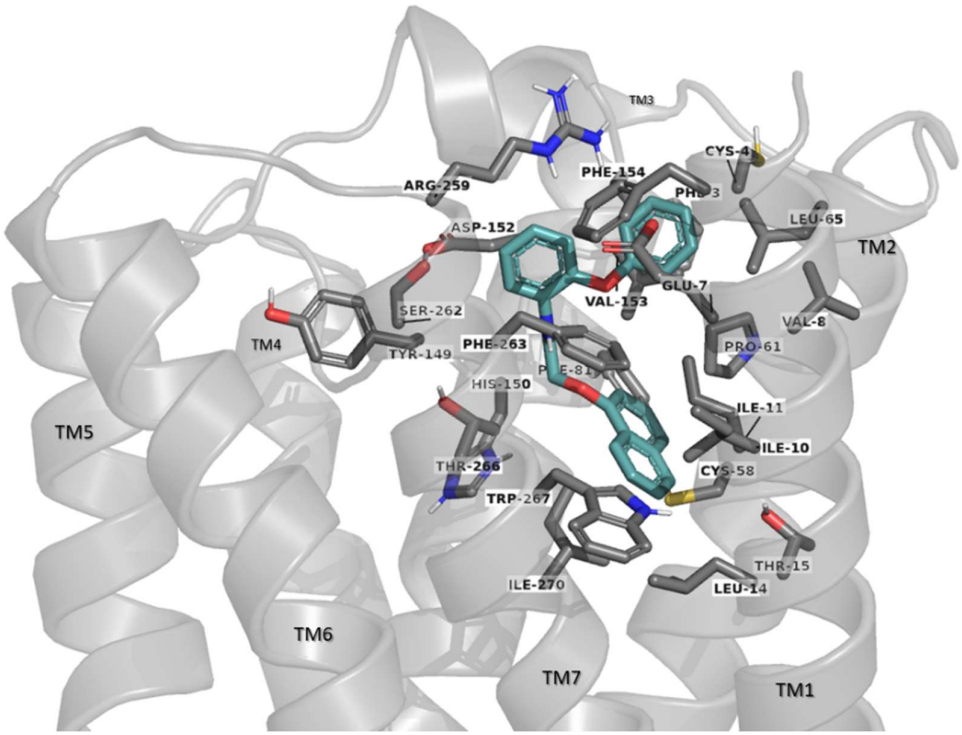

Figure 53: Top pose (1) for GRA2

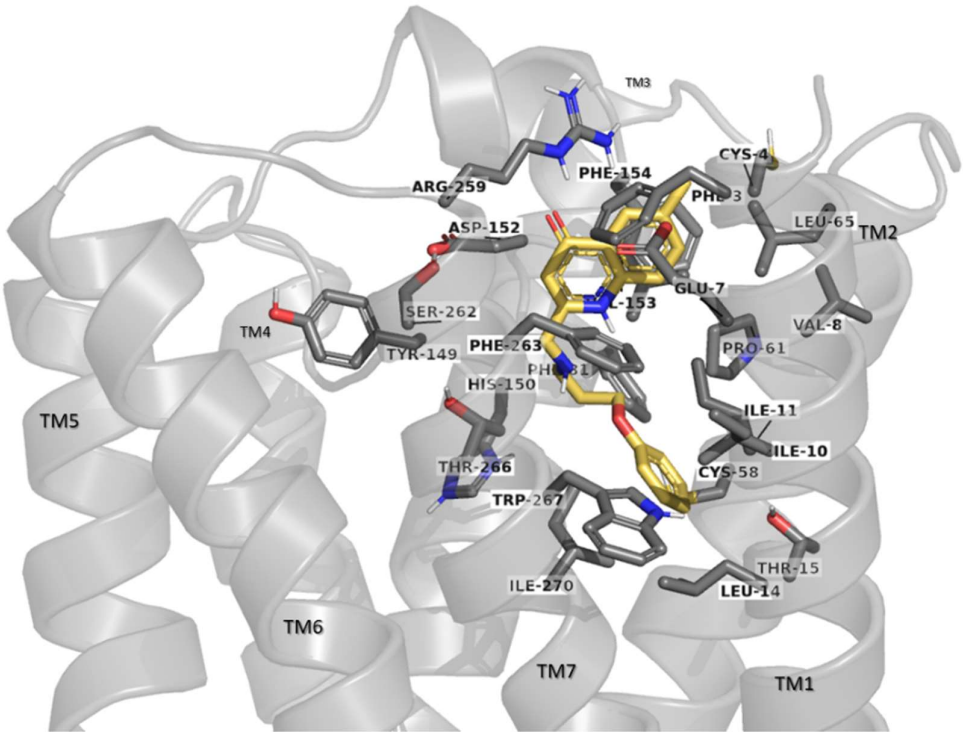

Figure 54: Top pose (1) for UCSF924

When visualizing GPR52 structure with c17 in the side pocket, with the predicted GPR21 rank2 with the top poses of GRA2 and UCSF924, we observe that the two GPR21 ligands take a very similar formation to c17, and they manage to reach a little deeper into the receptor (Fig.55).

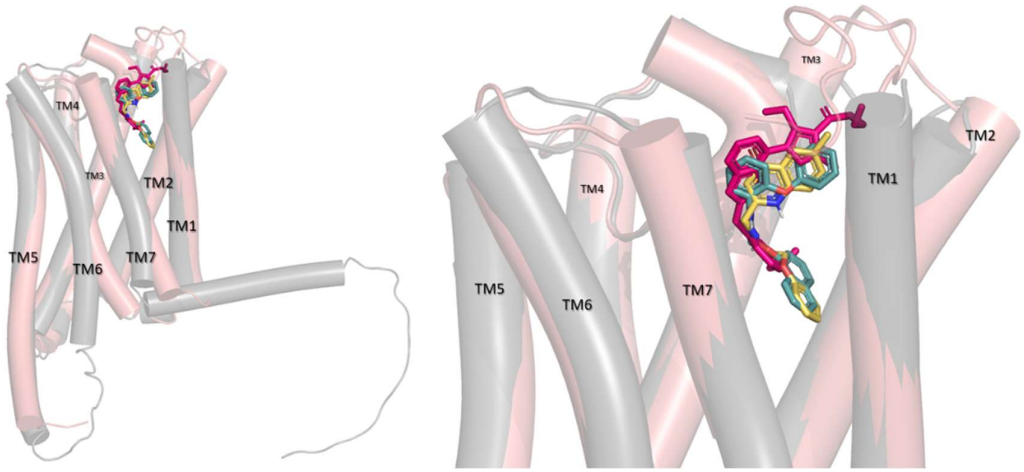

Figure 55: Superposition of GPR21 (gray) and GPR52 (light pink). GPR52 ligand c17 is shown in magenta, GRA2 is in teal and UCSF924 is in yellow.

5. Discussion

This study was designed to explore the interaction mechanisms between the small molecules GRA2 and UCSF924 and the orphan GPR21. In vitro experiments have shown that these compounds have some effect to the receptors activity, however their way of interaction is still not known. A part of the receptor, the ECL2, has been linked with its high basal activity. We produced different refined models of this loop, to investigate whether there could be a loop conformation that allows GRA2 and UCSF924 to interact with GPR21 in the orthosteric pocket area, and consequently interrupt the receptor's self-activation. Additionally, due to the GPR21-GPR52 high structural similarity, we explored the possibility of GPR21 having a side pocket, similar to the GPR52 one, where the two potential ligands could bind to.

Starting from the refined ECL2 models, we found that the cavities present in the whole loop and immersed region refinement had better druggability compared to the cap region. In addition, the whole loop and immersed region refined models gave better docking results with GRA2 and UCSF924. In both cases, the amino acid associated with receptor activation, $H174^{ECL2}$, has been displaced from the orthosteric pocket, whereas it remains inside in the cap region refined models. Combining these observations, we suggest that GRA2 and UCSF924 are more likely to interact with GPR21 in formations that resemble the immersed region refined models 1 and 12 or the whole loop refined models 19 and 4, respectively.

Exploration of the ARG283 rotamers of the 8HMV structure of GPR21 (unmodified) led to the discovery of a cavity with improved druggability over that present in the original structure. The docking scores for both GRA2 and UCSF924 were not better than those produced by the refined models.

The models predicted by Alphafold were of similar quality and confidence. Our goal was to see if we could create a side pocket where GRA2 and UCSF924 could interact with the receptor. The side pocket was expected to be located in the same area as the GPR52 side pocket, between TM1, TM2 and TM7. This expectation was based on the 71% sequence homology between the two GPCRs and the high structural similarity. The predicted GPR21 models without the N-terminus, which were built using the GPR52 structure as a template, performed better in the PAE and lDDT because one area of uncertainty, the N-terminus, was removed. The rank 2 predicted model appeared to have a pocket with at least moderate druggability in the region of interest. The docking poses of both GRA2 and UCSF924 showed great similarity to how c17, the GPR52 ligand, is found in the side pocket. Both GRA2 and UCSF924 reach deep into the cavity and form strong hydrophobic interactions with the side chains of the surrounding amino acids. Therefore, we suggest that the two ligands could bind to the receptor in the side pocket formed between TM1, TM2 and TM7.

To date, knowledge of how the two ligands interact with GPR21 is limited. USCF924 is the only compound linked to GPR21 in ChEMBL as a result of the PRESTO Tango assay that quantifies GPCR signaling (by recruiting β-arrestin2). There is evidence that UCSF924 interacts with GPR21, but the outcome of this interaction is not clear. With respect to GRA2, in vitro experiments have been performed showing that it acts as an inverse agonist for GPR21. There is still no information on how GRA2 and UCSF924 bind to GPR21 and which part of the receptor they interact with. The goal of this work is to propose potential interaction scenarios that the ligands may have with the receptor. The first hypothesis, which incorporates the refined ECL2 models, suggests that the ligands interact with the orthosteric pocket, thereby reducing the high basal activity of the receptor. The second hypothesis is based on the observation that GPR21 has a high degree of structural similarity to GPR52. This similarity suggests the presence of a potential side pocket between TM1, TM2 and TM7, which could serve as a binding site for the ligands. It is postulated that these interactions may ultimately lead to the disruption of the constitutive activity of the receptor.

The proposed models could be further investigated using molecular dynamics simulations to determine the stability of the receptor-ligand complexes. In vitro experiments could be performed with UCSF924 to investigate the effect of this compound on the constitutive activity of GPR21, whether it causes an increase or a decrease.

6. Conclusions

This work suggests two possible interaction mechanisms for GRA2 and UCSF924 with GPR21: the first is by binding within the orthosteric pocket to reduce basal activity, and the second involves interaction with a side pocket that could disrupt the constitutive activity of the receptor. Key findings are presented below:

Druggability of the refined models: The refined models that included the entire loop and the immersed region showed better druggability than those that focused only on the cap region. These models also produced better docking scores for the ligands GRA2 and UCSF924.

Discovery of improved cavity: Exploration of the ARG283 rotamers in the 8HMV structure of GPR21 led to the identification of a cavity with better druggability than the original structure. However, the docking scores of this cavity were not superior to those of the refined models.

Potential for side pocket interactions: Alphafold predicted models, especially those lacking the N-terminus and built using the GPR52 structure as a template, showed promise for forming a side pocket similar to that of GPR52. The docking results suggest that GRA2 and UCSF924 could bind to a side pocket located between TM1, TM2 and TM7 and form strong hydrophobic interactions with the surrounding amino acid side chains.

7. References

[1] A. and R. Vijayan, 'Orphan G protein-coupled receptors: the ongoing search for a home', *Front Pharmacol*, vol. 15, Feb. 2024, doi: 10.3389/fphar.2024.1349097.

[2] N. Tuteja, 'Signaling through G protein coupled receptors', *Plant Signal Behav*, vol. 4, no. 10, pp. 942–947, Oct. 2009, doi: 10.4161/psb.4.10.9530.

[3] J. L. Matthews *et al.*, 'Optimal nutrient exchange and immune responses operate in partner specificity in the cnidarian-dinoflagellate symbiosis', *Proceedings of the National Academy of Sciences*, vol. 114, no. 50, pp. 13194–13199, Dec. 2017, doi: 10.1073/pnas.1710733114.

[4] A. J. Kooistra *et al.*, 'GPCRdb in 2021: integrating GPCR sequence, structure and function', *Nucleic Acids Res*, vol. 49, no. D1, pp. D335–D343, Jan. 2021, doi: 10.1093/nar/gkaa1080.

[5] M. D. Wilkinson *et al.*, 'The FAIR Guiding Principles for scientific data management and stewardship', *Sci Data*, vol. 3, no. 1, p. 160018, Mar. 2016, doi: 10.1038/sdata.2016.18.

[6] B. K. Kobilka, 'G protein coupled receptor structure and activation', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1768, no. 4, pp. 794–807, Apr. 2007, doi: 10.1016/j.bbamem.2006.10.021.

[7] T. Wong *et al.*, 'Cryo-EM structure of orphan G protein-coupled receptor GPR21', *MedComm (Beijing)*, vol. 4, no. 1, Feb. 2023, doi: 10.1002/mco2.205.

[8] R. Fredriksson, M. C. Lagerström, L.-G. Lundin, and H. B. Schiöth, 'The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints', *Mol Pharmacol*, vol. 63, no. 6, pp. 1256–1272, Jun. 2003, doi: 10.1124/mol.63.6.1256.

[9] G. L. Szwabowski, D. L. Baker, and A. L. Parrill, 'Application of computational methods for class A GPCR Ligand discovery', *J Mol Graph Model*, vol. 121, p. 108434, Jun. 2023, doi: 10.1016/j.jmgm.2023.108434.

[10] Q. Zhou *et al.*, 'Common activation mechanism of class A GPCRs', *Elife*, vol. 8, Dec. 2019, doi: 10.7554/eLife.50279.

[11] Q. Zhou *et al.*, 'Common activation mechanism of class A GPCRs', *Elife*, vol. 8, Dec. 2019, doi: 10.7554/eLife.50279.

[12] S. Ali, P. Wang, R. E. Murphy, J. A. Allen, and J. Zhou, 'Orphan GPR52 as an emerging neurotherapeutic target.', *Drug Discov Today*, vol. 29, no. 4, p. 103922, Feb. 2024, doi: 10.1016/j.drudis.2024.103922.

[13] A. S. Hauser *et al.*, 'Pharmacogenomics of GPCR Drug Targets', *Cell*, vol. 172, no. 1–2, pp. 41-54.e19, Jan. 2018, doi: 10.1016/j.cell.2017.11.033.

[14] K. Sriram and P. A. Insel, 'G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs?', *Mol Pharmacol*, vol. 93, no. 4, pp. 251–258, Apr. 2018, doi: 10.1124/mol.117.111062.

[15] E. A. Wold, J. Chen, K. A. Cunningham, and J. Zhou, 'Allosteric Modulation of Class A GPCRs: Targets, Agents, and Emerging Concepts', *J Med Chem*, vol. 62, no. 1, pp. 88–127, Jan. 2019, doi: 10.1021/acs.jmedchem.8b00875.

[16] N. Ye *et al.*, 'Orphan Receptor GPR88 as an Emerging Neurotherapeutic Target', *ACS Chem Neurosci*, vol. 10, no. 1, pp. 190–200, Jan. 2019, doi: 10.1021/acschemneuro.8b00572.

[17] P. Wang, L. Lv, H. Li, C.-Y. Wang, and J. Zhou, 'Opportunities and challenges in drug discovery targeting the orphan receptor GPR12', *Drug Discov Today*, vol. 28, no. 9, p. 103698, Sep. 2023, doi: 10.1016/j.drudis.2023.103698.

[18] A. A. Bolinger, A. Frazier, J.-H. La, J. A. Allen, and J. Zhou, 'Orphan G Protein-Coupled Receptor GPR37 as an Emerging Therapeutic Target', *ACS Chem Neurosci*, vol. 14, no. 18, pp. 3318–3334, Sep. 2023, doi: 10.1021/acschemneuro.3c00479.

[19] E. Jacoby, R. Bouhelal, M. Gerspacher, and K. Seuwen, 'The 7 TM G-Protein-Coupled Receptor Target Family', *ChemMedChem*, vol. 1, no. 8, pp. 760–782, Aug. 2006, doi: 10.1002/cmdc.200600134.

[20] B. J. Melancon *et al.*, 'Allosteric Modulation of Seven Transmembrane Spanning Receptors: Theory, Practice, and Opportunities for Central Nervous System Drug Discovery', *J Med Chem*, vol. 55, no. 4, pp. 1445–1464, Feb. 2012, doi: 10.1021/jm201139r.

[21] M. M. Scharf *et al.*, 'The dark sides of the GPCR tree - research progress on understudied GPCRs', *Br J Pharmacol*, Feb. 2024, doi: 10.1111/bph.16325.

[22] A. Nieto Gutierrez and P. H. McDonald, 'GPCRs: Emerging anti-cancer drug targets', *Cell Signal*, vol. 41, pp. 65–74, Jan. 2018, doi: 10.1016/j.cellsig.2017.09.005.

[23] A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schiöth, and D. E. Gloriam, 'Trends in GPCR drug discovery: new agents, targets and indications', *Nat Rev Drug Discov*, vol. 16, no. 12, pp. 829–842, Dec. 2017, doi: 10.1038/nrd.2017.178.

[24] R. Lappano and M. Maggiolini, 'GPCRs and cancer', *Acta Pharmacol Sin*, vol. 33, no. 3, pp. 351–362, Mar. 2012, doi: 10.1038/aps.2011.183.

[25] F. Demenais *et al.*, 'Association of MC1R Variants and Host Phenotypes With Melanoma Risk in CDKN2A Mutation Carriers: A GenoMEL Study', *JNCI: Journal of the National Cancer Institute*, vol. 102, no. 20, pp. 1568–1583, Oct. 2010, doi: 10.1093/jnci/djq363.

[26] J. Wang and R. Xiao, 'G protein-coupled receptors in energy homeostasis', *Sci China Life Sci*, vol. 57, no. 7, pp. 672–680, Jul. 2014, doi: 10.1007/s11427-014-4694-2.

[27] D. Strassheim *et al.*, 'Metabolite G-Protein Coupled Receptors in Cardio-Metabolic Diseases', *Cells*, vol. 10, no. 12, p. 3347, Nov. 2021, doi: 10.3390/cells10123347.

[28] S. Rehman and S. Sharma, *Biochemistry, G Protein Coupled Receptors*. 2018. Accessed: Sep. 11, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK518966/#:~:text=G%20protein%2Dcoupled%20receptors%20(GPCRs,proteins%2C%20neurotransmitters%2C%20and%20hormones.

[29] O. Civelli, Y. Saito, Z. Wang, H.-P. Nothacker, and R. K. Reinscheid, 'Orphan GPCRs and their ligands', *Pharmacol Ther*, vol. 110, no. 3, pp. 525–532, Jun. 2006, doi: 10.1016/j.pharmthera.2005.10.001.

[30] M. Yoshida, M. Miyazato, and K. Kangawa, 'Orphan GPCRs and Methods for Identifying Their Ligands', 2012, pp. 33–44. doi: 10.1016/B978-0-12-381272-8.00002-7.

[31] X. Tang, Y. Wang, D. Li, J. Luo, and M. Liu, 'Orphan G protein-coupled receptors (GPCRs): biological functions and potential drug targets', *Acta Pharmacol Sin*, vol. 33, no. 3, pp. 363–371, Mar. 2012, doi: 10.1038/aps.2011.210.

[32] A. Fargin, J. R. Raymond, M. J. Lohse, B. K. Kobilka, M. G. Caron, and R. J. Lefkowitz, 'The genomic clone G-21 which resembles a β-adrenergic receptor sequence encodes the 5-HT1A receptor', *Nature*, vol. 335, no. 6188, pp. 358–360, Sep. 1988, doi: 10.1038/335358a0.

[33] J. R. Bunzow *et al.*, 'Cloning and expression of a rat D2 dopamine receptor cDNA', *Nature*, vol. 336, no. 6201, pp. 783–787, Dec. 1988, doi: 10.1038/336783a0.

[34] S. Chung, T. Funakoshi, and O. Civelli, 'Orphan GPCR research', *Br J Pharmacol*, vol. 153, no. S1, Mar. 2008, doi: 10.1038/sj.bjp.0707606.

[35] A. Ozawa, I. Lindberg, B. Roth, and W. K. Kroeze, 'Deorphanization of Novel Peptides and Their Receptors', *AAPS J*, vol. 12, no. 3, pp. 378–384, Sep. 2010, doi: 10.1208/s12248-010-9198-9.

[36] J. Gardner *et al.*, 'G-protein-coupled receptor GPR21 knockout mice display improved glucose tolerance and increased insulin response', *Biochem Biophys Res Commun*, vol. 418, no. 1, pp. 1–5, Feb. 2012, doi: 10.1016/j.bbrc.2011.11.117.

[37] V. Bordano *et al.*, 'G protein–coupled receptor 21 in macrophages: An in vitro study', *Eur J Pharmacol*, vol. 926, p. 175018, Jul. 2022, doi: 10.1016/j.ejphar.2022.175018.

[38] O. Osborn *et al.*, 'G protein–coupled receptor 21 deletion improves insulin sensitivity in diet-induced obese mice', *Journal of Clinical Investigation*, vol. 122, no. 7, pp. 2444–2453, Jul. 2012, doi: 10.1172/JCI61953.

[39] G. K. Kinsella *et al.*, 'GPR21 Inhibition Increases Glucose-Uptake in HepG2 Cells', *Int J Mol Sci*, vol. 22, no. 19, p. 10784, Oct. 2021, doi: 10.3390/ijms221910784.

[40] D. M. Riddy *et al.*, 'Deletion of GPR21 improves glucose homeostasis and inhibits the CCL2-CCR2 axis by divergent mechanisms', *BMJ Open Diabetes Res Care*, vol. 9, no. 2, p. e002285, Nov. 2021, doi: 10.1136/bmjdrc-2021-002285.

[41] S. Leonard, G. K. Kinsella, E. Benetti, and J. B. C. Findlay, 'Regulating the effects of GPR21, a novel target for type 2 diabetes', *Sci Rep*, vol. 6, 2016, doi: 10.1038/srep27002.

[42] S. Schenk, M. Saberi, and J. M. Olefsky, 'Insulin sensitivity: modulation by nutrients and inflammation', *Journal of Clinical Investigation*, vol. 118, no. 9, pp. 2992–3002, Sep. 2008, doi: 10.1172/JCI34260.

[43] J. M. Olefsky and C. K. Glass, 'Macrophages, Inflammation, and Insulin Resistance', *Annu Rev Physiol*, vol. 72, no. 1, pp. 219–246, Mar. 2010, doi: 10.1146/annurev-physiol-021909-135846.

[44] C. N. Lumeng and A. R. Saltiel, 'Inflammatory links between obesity and metabolic disease', *Journal of Clinical Investigation*, vol. 121, no. 6, pp. 2111–2117, Jun. 2011, doi: 10.1172/JCI57132.

[45] S. P. Weisberg *et al.*, 'CCR2 modulates inflammatory and metabolic effects of high-fat feeding', *Journal of Clinical Investigation*, vol. 116, no. 1, pp. 115–124, Jan. 2006, doi: 10.1172/JCI24335.

[46] H. Xu *et al.*, 'Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance', *Journal of Clinical Investigation*, vol. 112, no. 12, pp. 1821–1830, Dec. 2003, doi: 10.1172/JCI19451.

[47] A. E. Obstfeld *et al.*, 'C-C Chemokine Receptor 2 (CCR2) Regulates the Hepatic Recruitment of Myeloid Cells That Promote Obesity-Induced Hepatic Steatosis', *Diabetes*, vol. 59, no. 4, pp. 916–925, Apr. 2010, doi: 10.2337/db09-1403.

[48] N. Lanthier, Y. Horsmans, and I. A. Leclercq, 'Clodronate liposomes: All sites of injection are not equal', *Hepatology*, vol. 51, no. 2, pp. 721–722, Feb. 2010, doi: 10.1002/hep.23455.

[49] M. C. Arkan *et al.*, 'IKK-β links inflammation to obesity-induced insulin resistance', *Nat Med*, vol. 11, no. 2, pp. 191–198, Feb. 2005, doi: 10.1038/nm1185.

[50] D. Cai *et al.*, 'Local and systemic insulin resistance resulting from hepatic activation of IKK-β and NF-κB', *Nat Med*, vol. 11, no. 2, pp. 183–190, Feb. 2005, doi: 10.1038/nm1166.

[51] A. M. Neyrinck, P. D. Cani, E. M. Dewulf, F. De Backer, L. B. Bindels, and N. M. Delzenne, 'Critical role of Kupffer cells in the management of diet-

induced diabetes and obesity', *Biochem Biophys Res Commun*, vol. 385, no. 3, pp. 351–356, Jul. 2009, doi: 10.1016/j.bbrc.2009.05.070.

[52] X. Lin *et al.*, 'Cryo-EM structures of orphan GPR21 signaling complexes', *Nat Commun*, vol. 14, no. 1, p. 216, Jan. 2023, doi: 10.1038/s41467-023-35882-w.

[53] W. K. Kroeze *et al.*, 'PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome', *Nat Struct Mol Biol*, vol. 22, no. 5, pp. 362–369, May 2015, doi: 10.1038/nsmb.3014.

[54] M. Sawzdargo *et al.*, 'Identification and cloning of three novel human G protein-coupled receptor genes GPR52, ΨGPR53 and GPR55: GPR55 is extensively expressed in human brain', *Molecular Brain Research*, vol. 64, no. 2, pp. 193–198, Feb. 1999, doi: 10.1016/S0169-328X(98)00277-0.

[55] K. Kelly, 'Investigating the role of schizophrenia-associated gene expression in the developing human brain using Machine Learning. ', Technological University Dublin, 2021. doi: https://doi.org/10.21427/04vg-8p95.

[56] A. F. Pardiñas *et al.*, 'Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection', *Nat Genet*, vol. 50, no. 3, pp. 381–389, Mar. 2018, doi: 10.1038/s41588-018-0059-2.

[57] H. Komatsu *et al.*, 'Anatomical Transcriptome of G Protein-Coupled Receptors Leads to the Identification of a Novel Therapeutic Candidate GPR52 for Psychiatric Disorders', *PLoS One*, vol. 9, no. 2, p. e90134, Feb. 2014, doi: 10.1371/journal.pone.0090134.

[58] X. Lin *et al.*, 'Structural basis of ligand recognition and self-activation of orphan GPR52', *Nature*, vol. 579, no. 7797, pp. 152–157, Mar. 2020, doi: 10.1038/s41586-020-2019-0.

[59] M. Setoh *et al.*, 'Discovery of the First Potent and Orally Available Agonist of the Orphan G-Protein-Coupled Receptor 52', *J Med Chem*, vol. 57, no. 12, pp. 5226–5237, Jun. 2014, doi: 10.1021/jm5002919.

[60] T. Nakahata *et al.*, 'Design and synthesis of 1-(1-benzothiophen-7-yl)-1H-pyrazole, a novel series of G protein-coupled receptor 52 (GPR52) agonists', *Bioorg Med Chem*, vol. 26, no. 8, pp. 1598–1608, May 2018, doi: 10.1016/j.bmc.2018.02.005.

[61] C. L. Piscitelli, J. Kean, C. de Graaf, and X. Deupi, 'A Molecular Pharmacologist's Guide to G Protein–Coupled Receptor Crystallography', *Mol Pharmacol*, vol. 88, no. 3, pp. 536–551, Sep. 2015, doi: 10.1124/mol.115.099663.

[62] H. M. Berman, 'The Protein Data Bank', *Nucleic Acids Res*, vol. 28, no. 1, pp. 235–242, Jan. 2000, doi: 10.1093/nar/28.1.235.

[63] G. Pándy-Szekeres *et al.*, 'GPCRdb in 2018: adding GPCR structure models and ligands', *Nucleic Acids Res*, vol. 46, no. D1, pp. D440–D446, Jan. 2018, doi: 10.1093/nar/gkx1109.

[64] G. L. Clark and K. E. Corrigan, 'The Crystal Structure of Insulin', *Physical Review*, vol. 40, no. 4, pp. 639–639, May 1932, doi: 10.1103/PhysRev.40.639.

[65] J. D. BERNAL and D. CROWFOOT, 'X-Ray Photographs of Crystalline Pepsin', *Nature*, vol. 133, no. 3369, pp. 794–795, May 1934, doi: 10.1038/133794b0.

[66] J. C. KENDREW *et al.*, 'Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution', *Nature*, vol. 185, no. 4711, pp. 422–427, Feb. 1960, doi: 10.1038/185422a0.

[67] T. Huxford, 'X-Ray Crystallography', in *Brenner's Encyclopedia of Genetics*, Elsevier, 2013, pp. 366–368. doi: 10.1016/B978-0-12-374984-0.01657-0.

[68] C. R. BEDDELL, P. J. GOODFORD, F. E. NORRINGTON, S. WILKINSON, and R. WOOTTON, 'COMPOUNDS DESIGNED TO FIT A SITE OF KNOWN STRUCTURE IN HUMAN HAEMOGLOBIN', *Br J Pharmacol*, vol. 57, no. 2, pp. 201–209, Jun. 1976, doi: 10.1111/j.1476-5381.1976.tb07468.x.

[69] W. G. J. Hol, 'Protein Crystallography and Computer Graphics—toward Rational Drug Design', *Angewandte Chemie International Edition in English*, vol. 25, no. 9, pp. 767–778, Sep. 1986, doi: 10.1002/anie.198607673.

[70] L. Maveyraud and L. Mourey, 'Protein X-ray Crystallography and Drug Discovery', *Molecules*, vol. 25, no. 5, p. 1030, Feb. 2020, doi: 10.3390/molecules25051030.

[71] Z. Dauter and A. Wlodawer, 'Progress in protein crystallography', *Protein Pept Lett*, vol. 23, no. 3, pp. 201–210, Feb. 2016, doi: 10.2174/0929866523666160106153524.

[72] E. Callaway, 'Revolutionary cryo-EM is taking over structural biology', *Nature*, vol. 578, no. 7794, pp. 201–201, Feb. 2020, doi: 10.1038/d41586-020-00341-9.

[73] J. Jumper *et al.*, 'Highly accurate protein structure prediction with AlphaFold', *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, doi: 10.1038/s41586-021-03819-2.

[74] V. Vyas, R. Ukawala, C. Chintha, and M. Ghate, 'Homology modeling a fast tool for drug discovery: Current perspectives', *Indian J Pharm Sci*, vol. 74, no. 1, p. 1, 2012, doi: 10.4103/0250-474X.102537.

[75] M. T. Muhammed and E. Aki-Yalcin, 'Homology modeling in drug discovery: Overview, current applications, and future perspectives', *Chem Biol Drug Des*, vol. 93, no. 1, pp. 12–20, Jan. 2019, doi: 10.1111/cbdd.13388.

[76] A. Šali and T. L. Blundell, 'Comparative Protein Modelling by Satisfaction of Spatial Restraints', *J Mol Biol*, vol. 234, no. 3, pp. 779–815, Dec. 1993, doi: 10.1006/jmbi.1993.1626.

[77] K. Roy, S. Kar, and R. N. Das, 'Other Related Techniques', in *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, Elsevier, 2015, pp. 357–425. doi: 10.1016/B978-0-12-801505-6.00010-7.

[78] V. Salmaso and S. Moro, 'Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview', *Front Pharmacol*, vol. 9, Aug. 2018, doi: 10.3389/fphar.2018.00923.

[79] T. A and B. VA, 'Molecular Docking: From Lock and Key to Combination Lock', *J Mol Med Clin Appl*, vol. 2, no. 1, 2018, doi: 10.16966/2575-0305.106.

[80] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, 'Docking and scoring in virtual screening for drug discovery: methods and applications', *Nat Rev Drug Discov*, vol. 3, no. 11, pp. 935–949, Nov. 2004, doi: 10.1038/nrd1549.

[81] M. Mursal, M. Ahmad, S. Hussain, and M. Faraz Khan, 'Navigating the Computational Seas: A Comprehensive Overview of Molecular Docking Software in Drug Discovery', in *Unravelling Molecular Docking - From Theory to Practice [Working Title]*, IntechOpen, 2024. doi: 10.5772/intechopen.1004802.

[82] S. Rosignoli and A. Paiardini, 'DockingPie: a consensus docking plugin for PyMOL.', *Bioinformatics*, vol. 38, no. 17, pp. 4233–4234, Sep. 2022, doi: 10.1093/bioinformatics/btac452.

[83] A. V. Sadybekov and V. Katritch, 'Computational approaches streamlining drug discovery', *Nature*, vol. 616, no. 7958, pp. 673–685, Apr. 2023, doi: 10.1038/s41586-023-05905-z.

[84] M. Baek *et al.*, 'Accurate prediction of protein structures and interactions using a three-track neural network', *Science (1979)*, vol. 373, no. 6557, pp. 871–876, Aug. 2021, doi: 10.1126/science.abj8754.

[85] J. Abramson *et al.*, 'Accurate structure prediction of biomolecular interactions with AlphaFold 3', *Nature*, vol. 630, no. 8016, pp. 493–500, Jun. 2024, doi: 10.1038/s41586-024-07487-w.

[86] Schrödinger LLC, 'The PyMOL Molecular Graphics System, Version~2.5', Nov. 2015.

[87] S. Sasidharan and P. Saudagar, 'Prediction, validation, and analysis of protein structures: A beginner's guide', in *Advances in Protein Molecular and Structural Biology Methods*, Elsevier, 2022, pp. 373–385. doi: 10.1016/B978-0-323-90264-9.00023-4.

[88] Y. Xu *et al.*, 'CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction', *Nucleic Acids Res*, vol. 46, no. W1, pp. W374–W379, Jul. 2018, doi: 10.1093/nar/gky380.

[89] Y. Yuan, J. Pei, and L. Lai, 'Binding Site Detection and Druggability Prediction of Protein Targets for Structure- Based Drug Design', *Curr Pharm Des*, vol. 19, no. 12, pp. 2326–2333, Feb. 2013, doi: 10.2174/1381612811319120019.

[90] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, 'ColabFold: making protein folding accessible to all', *Nat Methods*, vol. 19, no. 6, pp. 679–682, Jun. 2022, doi: 10.1038/s41592-022-01488-1.

[91] Paulyna Gabriela Magana Gomez and Oleg Kovalevskiy, ‘AlphaFold: A Practical Guide’.

Appendix 1: Scripts for refined models

TABLE 9: Script for loop refinement of whole ECL2 of GPR21, while maintaining the disulfide bond. Highlighted (gray) parts are modified accordingly to produce the desired refined region.

```
from modeller import *
from modeller.automodel import *

env = environ()

class MyLoop(loopmodel):
    def select_loop_atoms(self):
        return selection(self.residue_range('169:A', '186:A'))
    def special_patches(self, aln):
        self.patch(residue_type='DISU',
                   residues=(self.residues['102:A'],
                             self.residues['181:A']))

m = MyLoop(env,
           inimodel='8hmva_disu.pdb',
           sequence='8hmva_disu_whole')
m.loop.starting_model = 1
m.loop.ending_model = 20
m.loop.md_level = refine.very_fast
m.make()


from modeller import *
from modeller.scripts import complete_pdb
```

TABLE 10: Script for DOPE Score of refined whole ECL2 of GPR21. Highlighted (gray) parts are modified accordingly to produce the desired refined region.

```

# Create a MODELLER environment
env = Environ()

# Load the topology and parameters
env.libs.topology.read(file='$(LIB)/top_heav.lib')
env.libs.parameters.read(file='$(LIB)/par.lib')

# List of PDB filenames
pdb_filenames = [
    "8hmva_disu_whole.BL00010001.pdb",
    …
    "8hmva_disu_whole.BL00200001.pdb"
]
```

```
# Create a list to store DOPE scores
dope_scores = []

# Loop through and score each PDB file
for pdb_filename in pdb_filenames:
    # Create a model for scoring
    mdl = complete_pdb(env, pdb_filename)

    # Select all atoms in the first chain
    atmsel = selection(mdl.chains[0])

    # Calculate DOPE score
    score = atmsel.assess_dope()

    # Append the score to the list
    dope_scores.append((pdb_filename, score))

# Output the results to a text file
#output_filename = "dope_scores.txt"
#with open(output_filename, "w") as f:
 #   for filename, score in dope_scores:
  #      f.write(f"File: {filename}, DOPE Score: {score}\n")



# Finalize and clean up
env.io.atom_files_directory = ['./']  # Set the directory where
your PDB files are located
env.io.hetatm = True  # If your PDB files contain HETATM records
env.io.water = True   # If your PDB files contain water molecules
env.io.stop_at_pdb = True

env.cleanup()
```

TABLE 11: Script for loop refinement of cap region of GPR21, while maintaining the disulfide bond. Highlighted (gray) parts are modified accordingly to produce the desired refined region.

```
from modeller import *
from modeller.automodel import *

env = environ()

class MyLoop(loopmodel):
    def select_loop_atoms(self):
        return selection(self.residue_range('179:A', '186:A'))
    def special_patches(self, aln):
        self.patch(residue_type='DISU',
```

```
                    residues=(self.residues['102:A'],
                              self.residues['181:A']))

m = MyLoop(env,
           inimodel='8hmva_disu.pdb',
           sequence='8hmva_disu_cap')
m.loop.starting_model = 1
m.loop.ending_model = 20
m.loop.md_level = refine.very_fast
m.make()


from modeller import *
from modeller.scripts import complete_pdb
```

TABLE 12: Script for DOPE score regarding cap region refined models. Highlighted (gray) parts are modified according to the refined region.

```
# Create a MODELLER environment
env = Environ()

# Load the topology and parameters
env.libs.topology.read(file='$(LIB)/top_heav.lib')
env.libs.parameters.read(file='$(LIB)/par.lib')

# List of PDB filenames
pdb_filenames = [
    "8hmva_disu_cap.BL00010001.pdb",
    …
    "8hmva_disu_cap.BL00200001.pdb"
]

# Create a list to store DOPE scores
dope_scores = []

# Loop through and score each PDB file
for pdb_filename in pdb_filenames:
    # Create a model for scoring
    mdl = complete_pdb(env, pdb_filename)

    # Select all atoms in the first chain
    atmsel = selection(mdl.chains[0])

    # Calculate DOPE score
    score = atmsel.assess_dope()

    # Append the score to the list
```

```
    dope_scores.append((pdb_filename, score))

# Output the results to a text file
#output_filename = "dope_scores.txt"
#with open(output_filename, "w") as f:
 #   for filename, score in dope_scores:
  #      f.write(f"File: {filename}, DOPE Score: {score}\n")



# Finalize and clean up
env.io.atom_files_directory = ['./']  # Set the directory where
your PDB files are located
env.io.hetatm = True  # If your PDB files contain HETATM records
env.io.water = True   # If your PDB files contain water molecules
env.io.stop_at_pdb = True

env.cleanup()
```