EVOLUTIONARY IDENTIFICATION OF RESIDUES RELATED TO G PROTEIN COUPLING SELECTIVITY IN AMINERGIC GPCRS

by
BERKAY SELÇUK

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfilment of
the requirements for the degree of
Master of Science

Sabancı University
July 2022

ABSTRACT

EVOLUTIONARY IDENTIFICATION OF RESIDUES RELATED TO G PROTEIN COUPLING SELECTIVITY IN AMINERGIC GPCRS

BERKAY SELÇUK

Molecular Biology, Genetics And Bioengineering M.Sc. THESIS, July 2022

Thesis Supervisor: Dr. Ogün Adebali

Keywords: GPCRs, G protein, coupling selectivity, comparative genomics, selectivity determinants

G protein-coupled receptors (GPCRs) are coupled to four different subfamilies of G proteins namely G_s, Gi, G_q, and $G_{12/13}$, in a selective way. However, receptor-wide determinants of selective G protein coupling and G protein specific activation mechanisms remained under investigated. Here, we analyzed aminergic receptors and its orthologs from different phylogenetic levels to reveal positions that are specifically evolved and conserved for the receptors that are coupled to similar G proteins. By co-analyzing these evolutionary conserved amino acids with available structures of the receptors, we revealed specific activation mechanisms for G_s, G_{i1}, G_o, and G_q. To verify that these pathways could play role in determining coupling selectivity we investigated G_s-specific activation mechanism further. By the help of molecular dynamics simulations, we demonstrated that G7x41 is important for receptor activation and it might determine G_s coupling selectivity by promoting outwards movement of TM6. Lastly, we summarized our findings into a model of G protein selectivity called "sequential switches of activation" explaning three major molecular switches controlling GPCR activation: ligand binding, G protein selective activation mechanisms, and G protein contact.

ÖZET

AMİNERJİK RESEPTÖRLERDE G PROTEİN KENETLENME SEÇİCİLİĞİYLE ALAKALI POZİSYONLARIN EVRİMSEL OLARAK ORTAYA ÇIKARILMASI

BERKAY SELÇUK

MOLEKÜLER BİYOLOJİ, GENETİK ve BİYOMÜHENDİSLİK YÜKSEK LİSANS TEZİ, TEMMUZ 2022

Tez Danışmanı: Dr. Ogün Adebali

Anahtar Kelimeler: GPCR, G protein, kenetlenme seçiciliği, karşılaştırmalı genetik, seçicilik belirleyicileri

G protein kenetli reseptörler (GPCR) dört farklı G protein ailesiyle seçici bir şekilde etkileşme geçip aktifleşirler (G_s, G_i, G_q ve $G_{12/13}$). Fakat, reseptörlerde bu seçiciliği belirleyen pozisyonlar ve G protein spesifik aktivasyon mekanizmaları yeterince ortaya çıkarılmamışır. Biz bu çalışmada aminerjik reseptörler ve onların ortologlarını evrimsel olarak inceleyip her bir G protein için özel olarak korunmuş pozisyonları tespit ettik. Ortaya çıkardığımız bu pozisyonları yapısal olarak incelediğimizde G_s, G_{i1}, G_o ve G_q için spesifik aktivasyon mekanizmaları olduğunu tespit ettik. Bulduğumuz aktivasyon mekanizmalarının G protein seçiciliği ile ilişkili olduğunu göstermek için Gs spesifik mekanizmayı daha detaylı inceledik. Moleküler dinamik simülasyonlarının yardımıyla G7x41'in reseptör aktivasyonu ve G_s kenetlenme seçiciliğini TM6'nın dışa doğru hareketini destekleyerek kontrol edebileceğini gösterdik. Son olarak, bulgularımızı sıralı aktivasyon anahtarları isimli, G protein ve reseptör arasındaki seçiciliği özetleyen bir modelde birleştirdik. Modelimiz GPCR aktivasyonunu açıkayan üç farklı anahtarı açıklıyor: ligand bağlanması, G protein seçici aktivasyon mekanizması ve G protein kontağı.

ACKNOWLEDGEMENTS

I would like to thank all the people who supported my journey at Sabancı University for the last three years. First of all, I would like to thank my advisor Ogün Adebali who took me as an undergraduate student and turned me into a researcher. He always shared my enthusiasm when I discovered something exciting no matter the time and pushed me to produce the best work possible. I believe that he has all the necessary qualities that a person can ask from an advisor.

Besides my advisor, I am lucky that I have great colleagues. I would like to thank Cem Azgari for being a supportive friend and research buddy; Aylin Bircan and Nurdan Kuru for listening all my research struggles; Onur Dereli, Sezgi Kaya, Ümit Akköse, Veysel Oğulcan Kaya, Burak İşlek, and Ömer Fatih Konar for creating the best lab environment possible. I am also thankful that I had a chance to guide a brilliant future scientist Tuana Aksu during her early career stages. I am sure that she will become a successful researcher in the future. As a part of this thesis, I had the pleasure to collaborate with Prof. Dr. Serdar Durdağı and İsmail Erol.

Lastly, I would like to mention my family and friends who supported me emotionally. I would like to thank my mother Gülay Uçar and brother Orçun Selçuk who always encouraged me to choose academia as my career path. I would like to extend my sincere thanks to my dear ex-girlfriend Halenur Ayrım and my best friend İlayda Kayapınar for keeping my motivation high when I needed it.

To my supportive family and friends

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

5HT1B: 5-Hydroxytryptamine receptor 1B
5HT2A: 5-Hydroxytryptamine receptor 2A
ACM2: Muscarinic acetylcholine receptor M2
ADRB2: Beta-2 adrenergic receptor
DNA: Deoxyribonucleic acid
DRD1: Dopamine Receptor D1
DRD2: Dopamine Receptor D2
DRD3: Dopamine Receptor D3
FDA: Food and Drug Administration
GEF: Guanine exhange factor
GPCR: G protein-coupled receptor
GPCRdb: G protein-coupled receptors database
GTP: Guanosine three phosphate
HRH1: Histamine H1 receptor
ML: Maximmum-likelihood
MSA: Multiple Sequence Alignment
Na+: Sodium ion
PDB: Protein data bank
POPC 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
RGS: Regulators of G protein signaling
RRCS: Residue-residue contact score
TM: Transmembrane domain
WT: Wild-type

1. INTRODUCTION

1.1 Activation Mechanisms of GPCRs

G protein-coupled receptors (GPCRs) are major regulators of signal transduction for the eukaryotes. These receptors might be activated through various external stimuli including light, hormones, or proteins. There are more than 800 human GPCRs discovered making hem largest group of transmembrane receptors in humans (Fredriksson, Lagerström, Lundin & Schiöth, 2003). This high number indicates for their importance as biological tools for maintaining functionality of the organism and inventing novel functions. Thus, the amount and diversity of receptors are fine-tuned for each organism (Insel, Snead, Murray, Zhang, Yokouchi, Katakia, Kwon, Dimucci & Wilderman, 2012; Niimura, 2009; Schiöth, Nordström & Fredriksson, 2007). Human GPCRs are divided into four major classes as A,B,C and F. Within the scope of this thesis we analyzed a receptor group belonging to class A which contain receptors for sensing light (rhodopsin (**?**)), monomaines such as adrenaline and dopamine (aminergic receptors (Vass, Podlewska, de Esch, Bojarski, Leurs, Kooistra & de Graaf, 2019)), chemokines (Mollica Poeta, Massara, Capucetti & Bonecchi, 2019), opioids (Waldhoer, Bartlett & Whistler, 2004), cannabinoids (Mackie, 2008), and many other types of ligands making it the largest class of GPCRs. Around one third of the drugs approved by Food and Drug Administration (FDA) target GPCRs (Hauser, Chavali, Masuho, Jahn, Martemyanov, Gloriam & Babu, 2018). Therefore, analysis of these receptors is important to understand the human physiology, reveal disease causing mutations, and invent novel therapeutic approaches targeting these receptors.

GPCR activation involves interaction of multiple protein families in a organized way. GPCRs induce downstream signaling pathways by coupling to four different subfamilies and 16 subtypes of G proteins. Although receptors induce four major signaling

pathways the physiological output is optimized for the cell types they are expressed and the stimuli they receive. Furthermore, G protein's activity is controlled by proteins named as regulators of G protein signaling (RGS). The GPCR activation mechanisms can be divided into two sub-categories as common and selective activation (Selçuk & Adebali, 2022). Common activation mechanisms are the ones that are shared mainly within a class of GPCRs, while selective mechanisms can highly vary from receptor to receptor due to their functional differences such as binding to a different ligand or engaging with a different set of G proteins (Fay & Wu, 2003; Gu, 2003). After a gene duplication if the new copy of the gene shows a functional difference this is called as functional divergence. When this functional divergence results in invention of new functions this is called as neofunctionalization. If the new copy only performs a subset of functions of the ancestral copy this is called as subfunctitonalization. Finally, if the gene duplication results in formation of a pseudogene this is called as non-functionalization. By analyzing changes in receptor amino acid sequence by comparing species from diverse phylogenetic complexity it is possible to infer commonalities and differences between receptor functions. While differences in receptor evolution provides an understanding on selective processes such as RGS-G protein binding, or G protein coupling selectivity, similarities provide information about shared functions between receptors. To fully comprehend the GPCR activation, one should integrate receptor-specific and shared functions.

1.1.1 Shared activation

GPCRs have diversified to recognize various ligands and induce different biological pathways. The main reason why GPCRs can easily invent new functions is that they contain a class-specific working activation machinery. The new functions can be considered additions to core receptor functions. From an evolutionary perspective each class can be considered as a successful solution to the problem of receptor activation ensuring the proper engagement of G protein and the receptor. Recently, shared class specific activation mechanisms have been investigated for each of the human GPCR classes by analyzing existing experimental structures(Selçuk & Adebali, 2022).

In 2019 a common activation mechanism for the class A GPCRs was revealed (Zhou, Yang, Wu, Guo, Guo, Zhong, Cai, Dai, Jang, Shakhnovich, Liu, Stevens, Lambert, Babu, Wang & Zhao, 2019). Researchers proposed a new measure to quantify the contact between two closely located residues called "residue-residue contact score"

(RRCS) and analyzed 234 experimental structures. Common trends in contact changes (ΔRRCS) observed upon activation of the receptor were identified. These common increases or decreases in contact scores were basis for the common activation pathway connecting the bottom of the ligand-binding pocket to G protein coupling. The common mechanism included important motifs including CWxP, DRY, and NPxxY (Filipek, 2019). The proposed activation mechanism included 4 main layers which connects the bottom of the ligand binding pocket to the G protein coupling interface. The layers are connected to each other in extracellular to intracellular direction. No interaction was identified within the ligand binding pocket due to high sequence variability between receptors and diversity of ligands they are bound to. The main finding of this work is that it shows the molecular rearrangements leading to outward movement of transmembrane helix 6. Although the TM6 tilt was accepted as an important indicator of receptor activation for class A GPCRs the underlying molecular mechanism was revealed through this study.

A similar work conducted by Hauser et. al. in 2021, similar mechanisms were identified for the classes A, B1, C and F by using a different methodological approach (Hauser, Kooistra, Munk, Heydenreich, Veprintsev, Bouvier, Babu & Gloriam, 2021). According to their study contacts were classified as active state or inactive state stabilizers according to their existence frequency within inactive and active state receptor structures. When a pair of residues interacting more frequently in inactive state receptors it is classified as inactive state stabilizing interaction. The same logic is also applicable for the active state stabilizers. It should be noted that the number of existing experimental structures in different class of GPCRs vary. Because this affect the confidence of identified pathways, researchers used different frequency threshold each class. Using an approach based on thresholds allowed identification of contacts within the ligand binding pocket. Moreover, it is much more likely to observe mechanistic similarities withing the ligand binding pocket for classes B, F, and C because there is less diversity in terms of ligands which receptors recognize.

When the mechanisms of different classes compared it can be observed that class B contain less inactivating contacts then other classes. This can be associated with the larger TM6 observed within active-state class B receptors. Less number of inactive state stabilizers might lower the energy barrier required for activation of class B receptors. This should be supported with further research.

Helical rearrangements were also investigated in this study. It has been shown that while the cytosolic movement of TM6 is a common feature for all classes, extracellular movement is not shared across classes. More specifically, no extracellular

movement of TM6 is observed for the receptors within class C. This contradicts with the dogma that TM6 tilt is a universal indicator of GPCR activation.

It has been shown that class A and class B1 receptors share four state determining switch residues. No residual switch was identified for class C GPCRs. This can be attributed to the fact that class C GPCRs require dimerization for switching to active state. In this case dimerization itself might work as a molecular switch. The other classes of GPCR can be activated in monomer form which might have forced them to acquire switches of activation that are triggered upon sensing external stimuli. Although class C has been presented as the earliest class (Cvicek, Goddard & Abrol, 2016; Flock, Hauser, Lund, Gloriam, Balaji & Babu, 2017), more precise phylogenetic analyses are required to establish the true GPCR evolution history.

To summarize, by analyzing available experimental structures, researchers identified commonalities and differences across different classes of GPCRs. 12 common residue pairs, 23 residues, and six state-determining residues were shared between two major studies for the class A activation mechanism. Both of the studies supports the idea that TM3 plays a very important role for receptor activation and maintaining signature 7TM fold the receptors for all classes. Furthermore, TM4 was shown to be making the least amount of connections among all transmembrane helices. While GPCRs contain class-inherited functions, they also invent new function. These novel receptor functions are responsible for the selective events such as G protein coupling or ligand binding that we will explain the following chapter.

1.1.2 Selective processes in G protein coupling

GPCRs can couple to a single or multiple G protein subtypes (Fig 1.1A). Moreover, GPCRs with distinct evolutionary backgrounds can couple to same G protein subtypes (Fig 1.1B). receptors share a similar mechanism to couple to the same G protein. On the other hand, while a receptors couples to a particular G protein subtype, the other receptor may not (Fig 1.1C). Thus, it is crucial to identify the causes of selectivity between the receptor and the G protein.

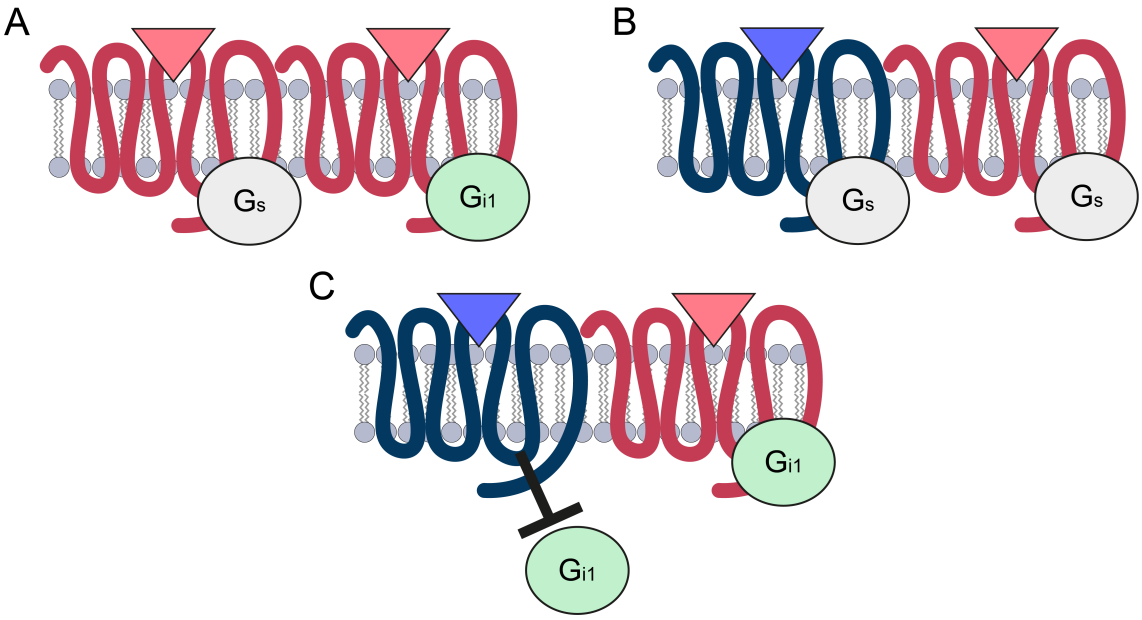

Figure 1.1 Selective Nature of G Protein Coupling
(A) Single receptor can couple to multiple G proteins. (B) Different receptors can couple to the same G protein. (C) While a receptor can couple to a particular G protein subtype, another receptor may disfavor it.

There were numerous attempts to identify G protein coupling selectivity determinants. Most of these (Chung, Rasmussen, Liu, Li, Devree, Chae, Calinski, Kobilka, Woods & Sunahara, 2011; Du, Duc, Rasmussen, Hilger, Kubiak, Wang, Bohon, Kim, Wegrecki, Asuru, Jeong, Lee, Chance, Lodowski, Kobilka & Chung, 2019; Liu, Xu, Hilger, Aschauer, Tiemann, Du, Liu, Hirata, Sun, Guixà-González, Mathiesen, Hildebrand & Kobilka, 2019; Okashah, Wan, Ghosh, Sandhu, Inoue, Vaidehi & Lambert, 2019; Semack, Sandhu, Malik, Vaidehi & Sivaramakrishnan, 2016) have mainly investigated the contacts within the known coupling interface. The remaining studies (Kang, Kuybeda, de Waal, Mukherjee, Van Eps, Dutka, Zhou, Bartesaghi, Erramilli, Morizumi, Gu, Yin, Liu, Jiang, Meng, Zhao, Melcher, Ernst, Kossiakoff, Subramaniam & Xu, 2018; Rose, Elgeti, Zachariae, Grubmüller, Hofmann, Scheerer & Hildebrand, 2014; Van Eps, Altenbach, Caro, Latorraca, Hollingsworth, Dror,

Ernst & Hubbell, 2018; Wang & Miao, 2019) have focused on the conformational differences between receptors coupled to G proteins. In 2017, researchers analyzed the co-evolution between G proteins and the receptors within the G protein coupling interface for 16 human G protein subtypes (Flock et al., 2017). They have revealed the selective nature of the interactions between G protein and the receptor. It has been suggested that receptors evolved ridges that can recognize G proteins. Because the number of receptors are much more higher than the number of G proteins, receptors should have developed features to recognize existing G protein set. This model is explained the selective receptor-G protein interactions by using a key and lock analogy. The shape of each key (GPCR) determines which locks (G proteins) it can open. As single key can open multiple keys a GPCR can be coupled to multiple subtypes of G proteins. During evolution, each receptor develops a certain structural shape which determines its ability to recognize G protein subtypes. No common solution for recognition of G protein was revealed so far. This indicates that each receptor come up with a receptor specific solution to recognize G protein specific barcodes. Although the key and lock model has explained the basis of selectivity within the known coupling interface a broader approach is needed to identify positions receptor-wide.

1.2 Recent Advancement in Identification of Coupling Profiles of GPCRs

In order to reveal positions that could determine G protein coupling selectivity the coupling profiles of the receptors should be known. In past, Guide to the Pharmacology database (Harding, Jane, Faccenda, Southan, Alexander, Davenport, Adam, Spedding & Jamie, 2021) was used for this purpose because they have gathered previously performed profiling experiments in a single database. Because experiments were performed by separate research groups making comparison across receptors and G proteins were not possible. Moreover, because there were not direct methods to detect G protein activity for all G protein subtypes some of the results needed further validation. Recently, two major studies overcame this problem (Avet, Mancini, Breton, Le Gouill, Hauser, Normand, Kobayashi, Gross, Hogue, Lukasheva, St-Onge, Carrier, Héroux, Morissette, Fauman, Fortin, Schann, Leroy, Gloriam & Bouvier, 2022; Inoue, Raimondi, Kadji, Singh, Kishi, Uwamizu, Ono, Shinjo, Ishida, Arang, Kawakami, Gutkind, Aoki & Russell, 2019) and identified coupling preferences of more than 100 human GPCRs. Firstly, Inoue et al. developed a methodology that is based on producing chimeric G proteins which C-terminal tail of the G protein is swapped for the G protein of interest. G_q was used as a backbone and a the activity of a G_q induced pathway that was measured by a colorimetric method called as shedding assay. The intensity of the color is correlated the degree of activation for the G protein of interest. This method takes advantage of the fact that C-terminal tail of the G protein is an important determinant of the coupling selectivity (Okashah et al., 2019). 11 different chimeric G proteins were produced that could represent 16 human G protein subtypes. The main drawback of using a chimeric G protein is that it neglects potential selectivity determining positions outside of the C-terminus. Moreover, because G_q was used as backbone of these chimeric G proteins, it may promote or inhibit interactions with certain receptors. Coupling preferences of 148 human GPCRs were identified through this methodology. No conserved motif was identified for the receptors coupling to same G proteins. This shows that there is no common solution exist for successful coupling event. This supports the idea that there is no universal solution for determining selectivity.

Second, Avet et al. invented a novel methodology that uses unmodified G proteins (Avet et al., 2022). They have developed biosensors measuring the disasocation of the proteins upon G protein activation such as p63-RhoGEF and PDZ-RhoGEF. They have quantified the amount of dsassociation by using a bioluminescence resonance energy transfer–based methodology. In their work they have profiled 100 GPCRs against 12 G proteins and three Beta-arrestins. They have demonstrated novel receptor-G protein engagements that wasn't shown before (Harding et al., 2021). Because most of the receptors they have looked for coupled either G_{15} (81%) or G_z (73%) they have proposed a G_{15}/G_z biosensor that can universally be used to measure receptor activity. This biosensor can be used to deorphanize a receptor or perform drug screenings. Analyzing datasets of Inoue et al. and Avet et al. together it is possible decipher the mysteries behind G protein coupling selectivity.

2. THE SCOPE OF THE THESIS

G protein-coupled receptors are receive external stimuli and induce four major downstream pathways through four subfamilies of G proteins (G_s,G_q,G_i, and $G_{12/13}$). In total there 16 human subtypes of G proteins. Understanding the nature of engagement between GPCRs and G proteins are important to create better therapeutic approaches, and detect polymorphisms that can be disease causing for the organism. Although researchers have been investigating possible scenarios to explain the differences in G protein repertoires of the receptors a complete understanding of receptor-level determinants is still required. Here, in this thesis we have developed a new approach to reveal evolutionary conserved residues for the receptors with similar G protein coupling preferences. We have retrieved protein sequences of human aminergic GPCRs and its functionally identical homologs from other organisms. We used these sequences to identify evolutionary conserved positions of each receptor. Combining evolutionary conservation with existing profiling datasets (Inoue, Avet) a pool of possible selectivity determining positions were revealed. We investigated our potential selectivity determining candidates in light of structural analysis. We have analyzed existing experimental structures of active and inactive-state GPCRs to demonstrate their potential mechanistic roles in G protein coupling mechanisms. Lastly, by the help of our collaborators we have performed molecular dynamics simulations to understand the role of G7x41 in G_s coupling mechanism and general receptor activation by analyzing non-G_s coupler variants at the same position.

The results and discussion part of this thesis are adapted from my first author publication "Evolutionary association of receptor-wide amino acids with G protein–coupling selectivity in aminergic GPCRs" (Selçuk, Erol, Durdağı & Adebali, 2022). Molecular dynamics simulations were performed by our collaborator Ismail Erol under the supervision of Prof. Dr. Serdar Durdagi.

3. MATERIALS AND METHODS

3.1 METHODS

3.1.1 GPCRdb Numbering System

In this study we used GPCRdb numbering to refer the homologous residues at different receptors. GPCRdb number of a residue consists of two main components seperated by an "x". The number at the left hand side indicates at which helix the residue is present. The number at the right hand side indicates the position of the residue with respect to the most conserved residue in that given helix. The most conserved position for each helix is given the number of 50 as Ballesteros-Weinstein suggested previously (Ballesteros & Weinstein, 1995). On top of that, GPCRdb numbering scheme also takes structural information into consideration. For example, when there is a kink observed within the structure of the receptor that is receptor-specfic instead of going from 55 to 56, that residue was numbered as 551. So simply GPCRdb takes gaps and insertions into account to refer residues having both structural and sequence homology. (Kooistra, Mordalski, Pándy-Szekeres, Esguerra, Mamyrbekov, Munk, Keserű & David, 2021)

3.1.2 Initial sequence selection

As a very first step of the evolutionary analysis, we blasted human aminergic GPCR sequences using BLAST+ (Camacho, Coulouris, Avagyan, Ma, Papadopoulos, Bealer & Madden, 2009) and retrieved every sequence until we observe third human GPCR sequence. We used UniProt (architecture & applications, 2019) database for the protein sequences.

3.1.3 First Multiple Sequence Alignment

The sequences retrieved from blast search are aligned by using MAFFT einsi option (Katoh & Standley, 2013).

3.1.4 First Maximum Likelihood (ML) Tree

We used the initial MSA for the production of the first maximum likelihood tree that includes sequences from multiple receptors. At this step we isolate the clade that belongs the homologs of our target protein. IQ-TREE version 2.0.6 (Minh, Schmidt, Chernomor, Schrempf, Woodhams, Von Haeseler & Lanfear, 2020) was used at this step with the evolutionary model of JTT+I+G4+F.

3.1.5 Isolation of the homologous clade

We used ETE3 python library (Huerta-Cepas, Serra & Bork, 2016) to isolate the clade belongs to our protein of interest. The third human protein hit from the blast search was used as an outgroup. Starting from our target protein, we moved to the root of the tree. We after each move we only kept clades containing not previously observed species. We stopped taking new clades if we see a human protein in a clade. We produced the second multiple sequence alignment with the direct homologs of our target protein.

3.1.6 Multiple Sequence Alignment Trimming

For trimming the second multiple sequence alignment we used trimAl (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009). automated1 option was used.

3.1.7 Second Maximum Likelihood Tree

RaxML-NG version 0.9.0 (Kozlov, Darriba, Flouri, Morel & Stamatakis, 2019) was used with the –search option and JTT+I+G4+F substitution model.

3.1.8 Obtaining Functionally Equivalent Orthologs

In this part we try to detect paralogous gene clades that are diverged and may not share the original function of the receptor. For this purpose we first calculated a pairwise global alignment score by using BLOSUM62 matrix. We scored each pair with respect to our human target protein. After that we compared each sister clade with similar taxonomic profiles and removed the ones that have a lower average similarity score. We used two sample t test to identify clades (sharing at least one identical species) that are significantly more diverged ($P<0.01$) from their sister clade. For the cases we have less than three leaf nodes for each sister clade, we removed the clade having an average lower similarity score.

3.1.9 Calculating Conservation Score by Using MSA

For each position in a given multiple sequence alignment we calculate the conservation as following:

First, detect the most frequent amino acid at that position. Second, if there are other amino acids that are similar (BLOSUM80 score > 2) to the most frequent amino acid we also add their frequency. Third, if the gaps are not the majority they are removed from the frequency calculation, otherwise that position is accepted as a gap. The conservation score is the frequency of most common and similar amino acids times 100, because we want to deal with percentages.

3.1.10 Identification of Specifically Conserved Amino Acids

For a position to be specifically conserved it must fit to a certain criteria. Given an MSA of two or more receptors and their functionally equal homologs, we calculate the conservation score for two of the receptors we are comparing. If a position is conserved more or equal than 90% we call that position as conserved. For a position to be specifically conserved for a receptor it should satisfy the conservation threshold and the same position of the compared receptor can be not conserved or a non-similar amino acid might be conserved.

3.1.11 Detection Potential Selectivity Determining Positions

We used two different approaches to determine position that have a potential to be a selectivity determinant:

3.1.11.1 Specific Approach

According to the specific approach we first divide receptors as into two separate groups as couplers and non-couplers. Secondly, we compare each coupler receptor with each non-coupler receptor and we kept the list of specifically residues. Thirdly, we compared coupler receptors with each other and from the previously obtained list, we removed the ones that are does not appear in our second comparison..

3.1.11.2 Sensitive Approach

In contrary to the specific approach in this case we build a single comprehensive multiple sequence alignment for the coupler receptors and calculate the conservation based on that alignment. Then, we identify specifically conserved residues for couplers against each non-coupler receptors.

The total list of potential selectivity residues are obtained after combining the results of specific and sensitive approaches for each G protein subtype we analyzed.

3.1.12 Construction of the Phylogenetic Tree of Aminergic Receptors

We used BLAST+ to retrieve top 50 hits for each aminergic receptor. We formed a large fasta file including top 50 hits of every aminergic receptor and used cd-hit with default options to get representative sequences only. We aligned the sequences by using MAFFT einsi option and produced the phylogenetic tree by using IQ-TREE version 2.0.5 with -m JTT+G+I+F -b 100 –tbe options.

3.1.13 Analysis of Residue-Residue Contact Scores

Residue-residue contact scores (Zhou et al., 2019) were calculated for 20 active-state structures: (ADRB2: 3SN6, 7DHI; DRD1: 7CKW, 7CKX, 7CKZ, 7CKY, 7CRH, 7JV5, 7JVP, 7JVQ, 7LJC, 7LJD; DRD2: 6VMS, 7JVR; DRD3: 7CMU, 7CMV; 5HT1B: 6G79; ACM2: 6OIK; 5HT2A: 6WHA; HRH1: 7DFL) (García-Nafría, Nehmé, Edwards & Tate, 2018; Kim, Che, Panova, DiBerto, Lyu, Krumm, Wacker, Robertson, Seven, Nichols, Shoichet, Skiniotis & Roth, 2020; Maeda, Qu, Robertson, Skiniotis & Kobilka, 2019; Rasmussen, Devree, Zou, Kruse, Chung, Kobilka, Thian, Chae, Pardon, Calinski, Mathiesen, Shah, Lyons, Caffrey, Gellman, Steyaert, Skiniotis, Weis, Sunahara & Kobilka, 2011; Xia, Wang, Xu, Lu, Song, Zhang, Guo & He, 2021; Xiao, Yan, Gou, Zhong, Kong, Wu, Wen, Yuan, Cao, Qu, Yang, Yang, Xia, Hu, Zhang, He, Zhang, Zhang, Hou, Liu, Zhu, Fu, Yang, Rosenbaum, Sun, Du, Zhang, Yu & Shao, 2021; Xu, Huang, Mao, Krumm, Zhou, Tan, Huang, Liu, Shen, Jiang, Yu, Jiang, Melcher, Roth, Cheng, Zhang & Xu, 2021; Yang, Ling, Zhou, Zhang, Lv, Liu, Fang, Sun, Hu, Zhang, Shi & Tian, 2020; Yin, Chen, Clark, Hijazi, Kumari, Bai, Sunahara, Barth & Rosenbaum, 2020; Zhuang, Krumm, Zhang, Zhou, Wang, Huang, Liu, Cheng, Jiang, Jiang, Zhang, Yi, Roth, Zhang & Xu, 2021; Zhuang, Xu, Mao, Wang, Krumm, Zhou, Huang, Liu, Cheng & Huang, 2021) and 24 inactive structures (ADRB2: 2RH1, 6PS2, 6PS3, 5D5A; DRD1: GPCRdb inactive model; DRD2: 6CM4, 6LUQ, 7DFP; DRD3: 3PBL; 5HT1B: 4IAQ, 4IAR, 5V54, 7C61; ACM2: 3UON, 5YC8, 5ZK3, 5ZKB, 5ZKC; 5HT2A: 6A93, 6A94, 6WH4, 6WGT; HRH1: 3RZE) (Cherezov, Rosenbaum, Hanson, Rasmussen, Thian, Kobilka, Choi, Kuhn, Weis, Kobilka & Stevens, 2007; Chien, Liu, Zhao, Katritch, Won Han, Hanson, Shi, Newman, Javitch, Cherezov & Stevens,

2010; Haga, Kruse, Asada, Yurugi-Kobayashi, Shiroishi, Zhang, Weis, Okada, Kobilka, Haga & Kobayashi, 2012; Huang, Olieric, Ma, Howe, Vogeley, Liu, Warshamanage, Weinert, Panepucci, Kobilka, Diederichs, Wang & Caffrey, 2016; Im, Inoue, Fujiwara, Nakane, Yamanaka, Uemura, Mori, Shiimura, Kimura, Asada, Nomura, Tanaka, Yamashita, Nango, Tono, Kadji, Aoki, Iwata & Shimamura, 2020; Ishchenko, Stauch, Han, Batyuk, Shiriaeva, Li, Zatsepin, Weierstall, Liu, Nango, Nakane, Tanaka, Tono, Joti, Iwata, Moraes, Gati & Cherezov, 2019; Kim et al., 2020; Kimura, Asada, Inoue, Kadji, Im, Mori, Arakawa, Hirata, Nomura, Nomura, Aoki, Iwata & Shimamura, 2019; Miyagi, Asada, Suzuki, Takahashi, Yasunaga, Suno, Iwata & Saito, 2020; Shimamura, Shiroishi, Weyand, Tsujimoto, Winter, Katritch, Abagyan, Cherezov, Liu, Han, Kobayashi, Stevens & Iwata, 2011; Suno, Lee, Maeda, Yasuda, Yamashita, Hirata, Horita, Tawaramoto, Tsujimoto, Murata, Kinoshita, Yamamoto, Kobilka, Vaidehi, Iwata & Kobayashi, 2018; Wang, Jiang, Ma, Wu, Wacker, Katritch, Han, Liu, Huang, Vardy, McCorvy, Gao, Zhou, Melcher, Zhang, Bai, Yang, Yang, Jiang, Roth, Cherezov, Stevens & Xu, 2013; Wang, Gareri & Rockman, 2018; Yin, Zhou, Yang, de Waal, Wang, Dai, Cai, Huang, Liu, Wang, Yin, Liu, Zhou, Wang, Liu, Caffrey, Melcher, Xu, Wang, Xu & Jiang, 2018). For each receptor, we calculated ΔRRCS scores of each contacting residue pair by subtracting active-state contacts from inactives. Each active and inactive pair of structures resulted in formation of a single contact change (ΔRRCS) network. From 41 individual networks we identified residue pairs that have a contact change more or equal than 36 times and revealed most common paths (Figure 3.1) where signal is transferred within the receptor. each receptor, we subtracted inactive RRCS from active RRCS to obtain ΔRRCS values for each residue pairs.

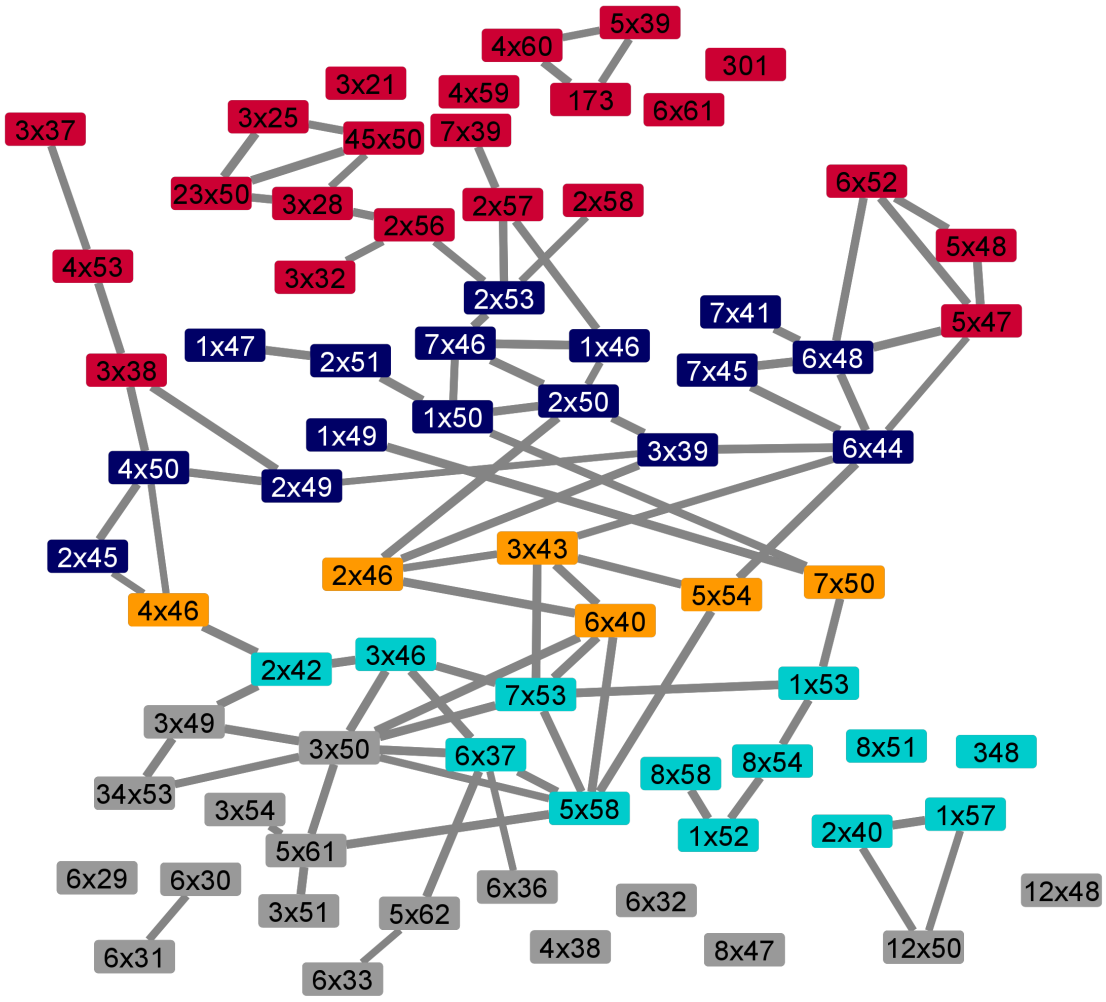

Figure 3.1 Most Frequently Used Signal Transduction Pathways Upon G protein Coupling
The extended network for Figure 3.2A including all conserved residues

3.1.14 Identification of G protein specific activation networks

To identify G protein specific activation networks, we grouped ΔRRCS values for coupler and non-coupler receptors. By using two sided t test we compared coupler and non-coupler groups and identified residue pairs exhibiting a higher or lower magnitude of contact change within coupler group compared to non-couplers. $P \leq 0.01$ is used as a threshold, but due to scarcity of G_{i1} network we projected the network for $P \leq 0.1$.

3.1.15 Analysis of Molecular Dynamics Simulations

We used structures of active and inactive ADRB2 (PDB ID: 3SN6, 4GBR Rasmussen et al. (2011); Zou, Weis & Kobilka (2012)). The sequences of the structures were converted to WT receptor by applying three different mutations to the sequences from the Protein Data Bank. Because the active structure of the ADRB2 lacked ICL3, we completed it by using GalaxyLoop (Park, Lee, Heo & Seok, 2014). We used PyMOL to introduce three mutations; G315C, G315L, and G315Q . The position of the receptors on lipibilayer membrane was calculated with OPM server (Lomize, Pogozheva, Joo, Mosberg & Lomize, 2012) CHARMM-GUI web server (Jo, Kim, Iyer & Im, 2008; Lee, Cheng, Swails, Yeom, Eastman, Lemkul, Wei, Buckner, Jeong, Qi, Jo, Pande, Case, Brooks, MacKerell, Klauda & Im, 2016; Wu, Cheng, Jo, Rui, Song, Dávila-Contreras, Qi, Lee, Monje-Galvan, Venable, Klauda & Im, 2014) were used to create inital files the simulations that are performed on Grmocas. We performed N-terminus acetlyation and C-terminus methylamidation. We introduced disulfide bridges between 106-191 and 184-190. Lipid leaflets contain 92 (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) POPC biological lipid type (total 192 POPC molecules in system). 0.15 M NaCl ions were used for neutralization. TIP3P water model (MacKerell, Bashford, Bellott, Dunbrack, Evanseck, Field, Fischer, Gao, Guo, Ha, Joseph-McCarthy, Kuchnir, Kuczera, Lau, Mattos, Michnick, Ngo, Nguyen, Prodhom, Reiher, Roux, Schlenkrich, Smith, Stote, Straub, Watanabe, Wiórkiewicz-Kuczera, Yin & Karplus, 1998) and CHARMM36m force field were used (Huang, Rauscher, Nawrocki, Ran, Feig, De Groot, Grubmüller & Mackerell, 2017). The simulation contianed a single minimization and six equilibration before the production for 50 ns. We used Berendsen thermostat and barostat (Berendsen, Postma, Van Gunsteren, Dinola & Haak, 1984). Noose–Hoover thermostat (Hoover, 1986; Nosé & Klein, 1983) and Parrinello–Rahman barostat (Parrinello & Rahman, 1980) were applied during the production. We used Gromacs v2020 (Abraham, Murtola, Schulz, Páll, Smith, Hess & Lindahl, 2015) and run 500ns long production. We run each simulation system seven times.

For the calculation of GPCRdb distances 5000 frames were retrieved from each run.

GPCRdb distance= (Y2x41-G6x38) - (C3x44-I7x52)

The structure is considered active if the GPCRdb distance is higher than 7.15 angstrom and inactive if lower than 2 angstrom (Isberg, De Graaf, Bortolato, Cherezov, Katritch, Marshall, Mordalski, Pin, Stevens, Vriend & Gloriam, 2015; Shahraki, Işbilir, Dogan, Lohse, Durdagi & Birgul-Iyison, 2021).

3.1.16 Residue-residue contact score analysis of MD Simulation Trajectories

11 frames from each simulation run were selected as a representative of the all simulation run. RRCSs were calculated for each frame and seven replicates. In total 77 frames of variant and WT simulations were compared by using two sample

t-test. Contacts that are significantly different were identified ($P < 0.05$). For the inactive-state simulations we had two replicates for each system. Thus, we used 22 data points for each system.

The common significant changes were identified in non-coupler variants. Overall, contacts between 135 residue pairs for the active receptor and 83 for the inactive receptor changed significantly. Cytoscape was used for the visualization of the network (Figure 3.1).

3.1.17 Data Availability

The codes and the data used in this project are available at:
https://github.com/CompGenomeLab/GPCR-coupling-selectivity
https://doi.org/10.5281/ zenodo.5763490.

4. RESULTS

After each DNA duplication the genes present on the newly synthesized DNA can diverge from their ancestral version. These changes create basis for natural selection and they can cause speciation if they are accumulated significantly. Sometimes a duplication within species can cause formation of a paralogous copy of the gene which may not be performing the same function with its ancestral version. Thus, the evolutionary forces acting on the paralog protein of interest might differ from the original. It has been previously shown (Adebali, Reznik, Ory & Zhulin, 2016) that the substitutions observed in a paralogous clades may not be tolerated for its orthologous ones. To infer the correct conservation frequency for a position these paralogous proteins should be excluded from the analysis. Thus, I identified functionally equivalent homologous receptors for every aminergic receptor and identified positions that are conserved above a certain threshold.

Conserved positions can be divided into two major categories: The ones that that are related to the core receptor functions such as maintaining proper fold and related to the more specific functions such as sensing a ligand or coupling to a G protein (Fig 4.1A). Therefore, to reveal positions that might be associated to G protein coupling selectivity, we grouped GPCRs having similar coupling profiles for 11 G protein subtypes and investigated differential conservation of amino acids between coupler and non-coupler groups (Fig 4.1B). Firstly, I used coupling information from two different datasets and identified positions that are strictly conserved withing the coupler receptor group and not the non-couplers. I called this as specific approach (Fig 4.1B, red and blue arrows). This approach does not tolerate variations in coupler receptors (Avet et al., 2022; Inoue et al., 2019) and accept labels as true positives. However, both of the datasets can contain false negatives which can affect our findings. Furthermore, selectivity determinants may not be too stringent and tolerate small variation. Therefore, I also applied a secondary approach that is named as "sensitive approach". In this method, instead of analyzing conservation in each coupler receptor individually, I treated them as a single group of receptors and formed a large inclusive multiple sequence alignment for all coupler receptors and their functionally equivalent orthologs (Fig 4.1B, orange arrows). I didn't apply sensitive approach for G proteins G12 and G13. This is because they contained low number of coupler receptors which can lead discovery of less specific positions. Lastly, I have identified positions that are conserved for all aminergic receptors. Finally, I analyzed each aminergic receptor separately and revealed residues that are conserved in all of the aminergic subfamily (consensus). As I hypothesized, these residues should be related to core receptor functions such as maintaining 7TM fold and performing a TM6 tilt upon activation. As a result, 22 consensus positions and

53 potential selectivity determinants were revealed. Fig 4.1C shows the distribution of the potential determinants for each G protein subtype.

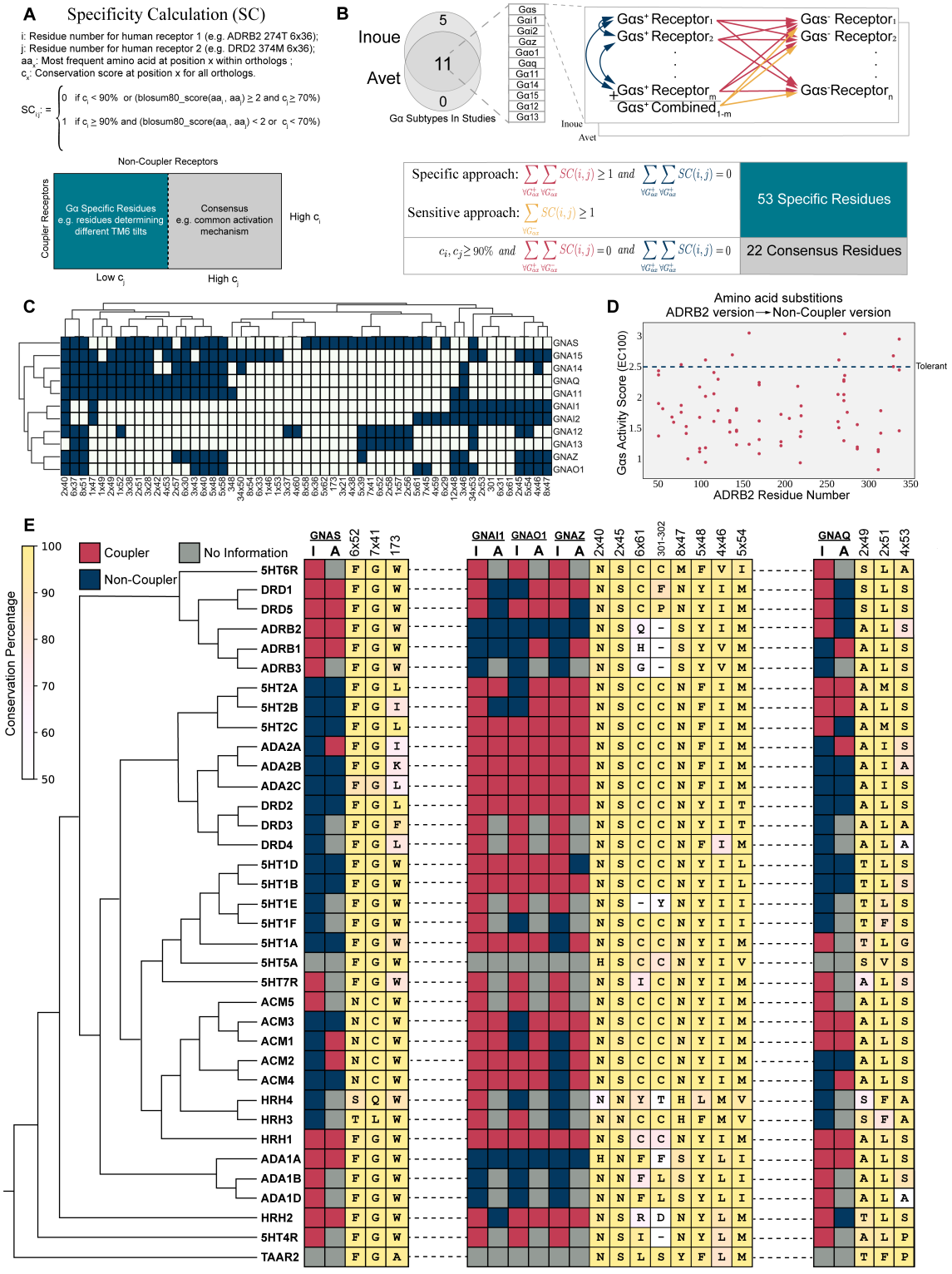

Figure 4.1 Potential Selectivity Determining Positions
(A) Specifically conserved residue finding formula. (B) The summary of comparisons for specific and sensitive approaches. Each arrow represents a single comparison. (C) The distribution of potential selectivity determinants for each G protein subtype. The residues and G proteins are clustered based on complete linkage. (D) G_s activity scores of non-coupler variants compared to tolerant mutation threshold. (E) Phylogenetic tree of aminergic receptors with coupling profiles, and conservation of some potential variants.(I, Inoue; A, Avet). The color intensity for each amino acid is correlated with its conservation.

We wanted to explore that if non-coupler variants at the positions that we identified as potential determinants are deleterious or not. Therefore, I have benefited the mutational data of ADRB2 based on its G_s activation at different ligand concentrations (Jones, Lubock, Venkatakrishnan, Wang, Tseng, Paggi, Latorraca, Cancilla, Satyadi, Davis, Babu, Dror & Kosuri, 2020). I have plotted activity scores of non-coupler mutations for the 31 residues that was identified for G_s. In Figure 4.1D it can be seen that the majority of non-coupler variants exhibit a lower receptor activity then what is considered as tolerant mutation. Because there can be many reasons explaining the decrease in G_s signal such as ER retention or misfolding, we argue that it is more likely to be due to decreased coupling. This is because the variants that I looked are already present in close homologs of ADRB2 (Fig 4.1E). If these non-coupler receptors can tolerate these amino acids, we can argue that they are more likely to be related to G protein coupling selectivity rather than core receptor functions.

In theory, if a position determines coupling selectivity it should be involved in G protein specific functions. It can be part of the coupling interface or could help receptor to achieve a G protein specific conformation which can promote or inhibit coupling. To understand how each residue can contribute to selectivity I grouped them based on their place on the receptors such as ligand binding or coupling interface. For instance, positions 8x47 (for GPCRdb numbering see methods) and 6x36 were previously shown to be interacting with the G protein directly (Kim et al., 2020; Maeda et al., 2019; Rasmussen et al., 2011; Xiao et al., 2021; Yang et al., 2020; Zhuang et al., 2021). It should be noted that we have identified G protein contacting residues without using a prior structural information which validates the success of our methodology. Moreover, our findings are in accordance with the key and lock model proposed by Flock et al (2017). The positions that we cannot group into a functional category can either play role in allosteric regulation of the receptor or may be false positives. To assign a structural function to the residues I have analyzed molecular insights of receptor activation through G protein coupling. We have benefited from Residue–Residue Contact Score (RRCS) (Zhou et al., 2019) algorithm to quantify the contact between two neighbouring residues. For each residue pair we calculated differences in RRCSs (ΔRRCS) to reveal residue pairs exhibiting a contact change upon activation. For this purpose I calculated ΔRRCS for each interacting residue pair by subtracting the contact scores of the active structure from the inactive structure. Because the active structures that I used contained a G protein I wanted to compare changes observed for different G proteins and identify G protein specific mechanistic changes for the evolutionary conserved positions that I identified in the first step. In total, I used structures from eight different GPCRs coupled to G_s, G_q, G_{i1} and G_o. Because we wanted to use 10 active G_s-coupled DRD1 structures I included a model inactive DRD1 structure to our analysis which was not available experimentally (Kooistra et al., 2021; Pándy-Szekeres, Munk, Tsonkov, Mordalski, Harpsøe, Hauser, Bojarski & Gloriam, 2018).

As a part of our structural analysis I analyzed 41 active and inactive structure pairs. I have built a common signal transduction network based on the frequency of information change observed in these 41 pairs. I have kept residue pairs that were changing contacts upon activation more than 87% of the pairs. I projected this path connecting ligand binding pocket to coupling interface in Figure 4.2A. By

using the layers of the previously proposed common activation mechanism for Class A GPCRs (Zhou et al., 2019) as our reference, I built a five layer network showing possible signal transduction routes within aminergic receptor. As an addition to the common activation mechanism I included Layer 0 containing residues that are inside the ligand binding pocket and closer to the extracellular part of the receptor. Despite the fact that most of the routes include important motifs such as PIF and Na+ binding pocket (Katritch, Fenalti, Abola, Roth, Cherezov & Stevens, 2014), it should be emphasized that the path starting with 3x37 does not include any known motifs. We demonstrated all the layers on experimental structures of ADRB2 and colored each layer differently (Fig 4.2B).

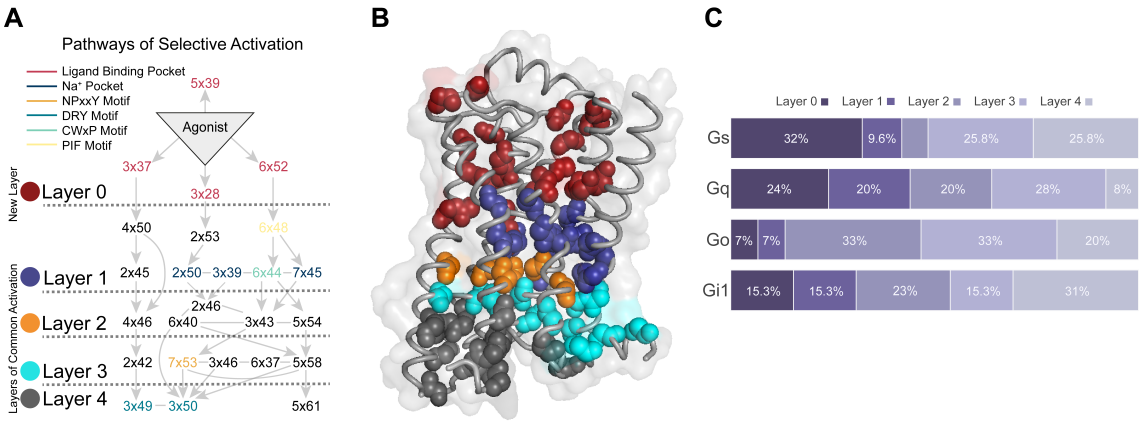

Figure 4.2 Frequently Used Signal Transduction Paths and Layers Upon G Protein Coupling
(A) The central route for molecular selective signal transduction which connects ligand binding pocket to the G protein coupling interface. The network is divided into five different layers. (B) Demonstration of different layers on inactive ADRB2 structure (PDB id: 2RH1) (C) The specifically conserved residue distribution of G_s, G_q, G_o, and G_{i1} for different layers of the activation pathway.

In Figure 4.2C I plotted the distribution of specifically conserved residues for G_s, G_q, G_{i1} and G_o to understand where they are accumulated mostly. It is evident that more residues are present at Layers 0 and 1 for G_s and G_q coupled receptors, while more residues were present for G_{i1} and G_o at the lower layers. Because these layers are closer to the coupling interface we can infer that physical contact with the G protein may be more crucial for the receptors coupled to G_{i1} and G_o.

To reveal any mechanistic difference between receptors coupled to different G proteins I compared changes in contact score ΔRRCS with each other by using t-test. ΔRRCS in this case represents changes in contact upon coupling to a G protein. I built G protein specific networks for G_s, G_q, G_{i1} and G_o by using residue pairs that show differential contact change (Δ Δ RRCS) which is not equal to zero ($p<0.01$) (Fig 4.3B–E). The existence of significant mechanistic differences support our idea that conserved residues involve in G protein specific activation mechanisms. In Figure 4.3A it can be seen that G_s and G_q networks include positions contacting to a ligand. d ligand contacting residues (Fig 4.3A and E). It can be inferred that ligand binding might be more important for G_s and G_q than G_{i1} and G_o in terms of determining coupling selectivity. It is interesting that, despite the fact that G_{i1} and

G_o have a high sequence similarity, they share 47% of their potential determinants for G_o and 62% for G_{i1}. Also, their networks are totally different from each other. Although I tried to group receptors as a G_{i1} and G_o group and compared with the rest I did not observe any common mechanistic feature that is statistically significant. This shows that G proteins belonging to same subfamily do not necessarily induce similar mechanistic changes within receptors. This can be one of the factors contributing to the selectivity for the aminergic receptors that are coupled either of the G proteins.

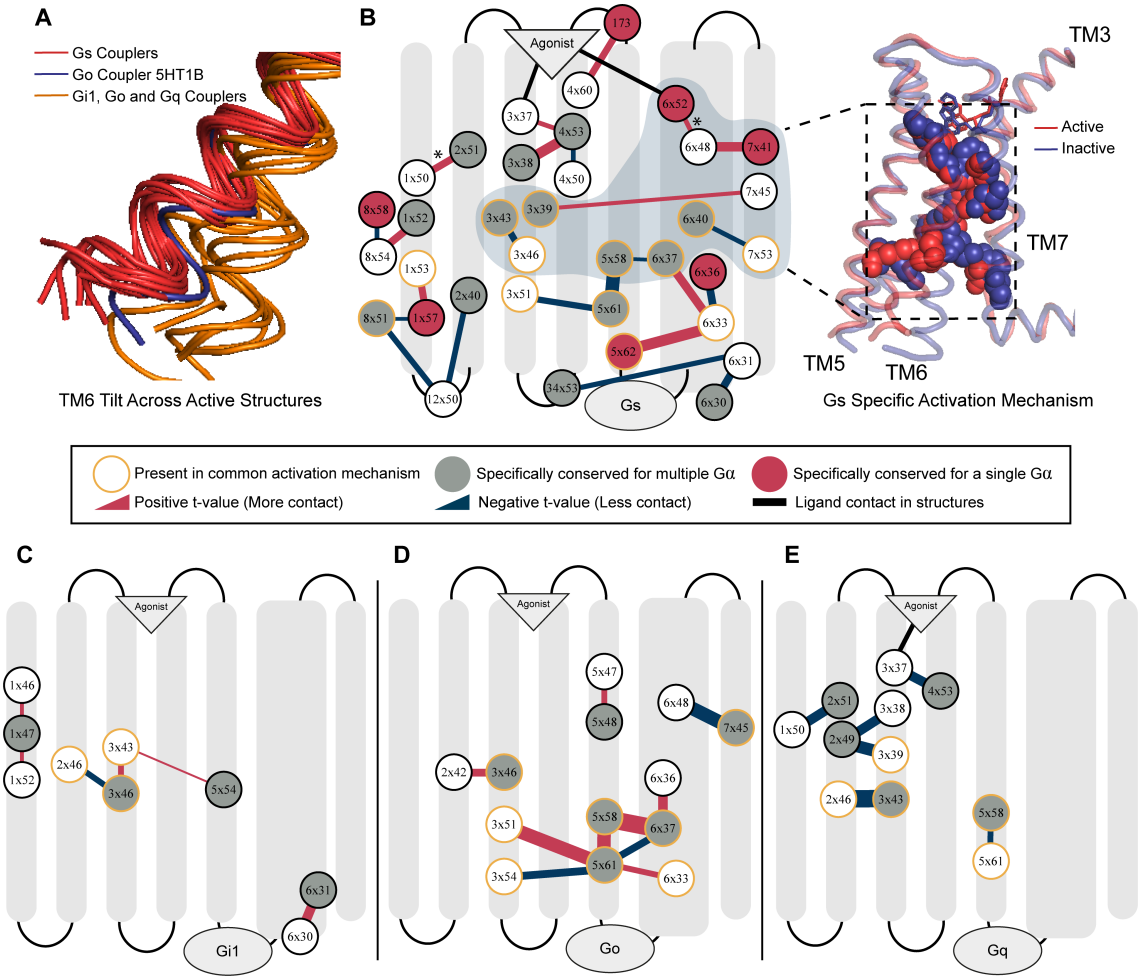

Figure 4.3 G Protein Specific Activation Networks
(A) The comparison of TM6 oreientation between active state aminergic receptors.Red G_s coupler, orange G_o-G_{i1}-G_q coupler, blue 5HT1B G_o coupler. (B) G_s specific activation network. Red edge: increase in contact score, blue edge: decrease in contact score, orange circle: part of the common activation mechanism of class A, red fill: only part of the G_s network, grey fill: specifically conserved for G protein of interest. Line width correlates with the t-value. Residues that are closely located to TM6. The highlihted residues were shown on inactive (blue) and active-state (red) structures of ADRB2. (C, D, E) G protein specific activation networks of G_{i1}, G_o, and G_q. $P < 0.1$ is used for G_{i1}. *: This interaction appears as significant when 5HT1B removed from the comparisons.

Even though we have showed that conserved residues are part of G protein specific networks we still cannot label any residue as a selectivity determinant because we are

not confident that these G protein specific pathways determine coupling selectivity. These contact changes can also arise as a result of the G protein coupling. To understand the impact of our networks on selectivity, I investigated G_s specific activation mechanism and its relationship with the differential TM6 tilt observed for G_s coupled receptors (Rose et al., 2014; Van Eps et al., 2018). I superimposed active state structures that I used analysis and observed that the structures I benefited follow a similar trend as well (Fig 4.3A). Furthermore, the reason that G_s specific mechanism is more complex than others (Fig 4.3B–E) can be due to necessity to achieve a larger TM6 movement. G_o coupled serotonin receptor 5HT1B (Fig 4.3A, blue structure) can be an exception to this (García-Nafría et al., 2018), because it exhibits a more open TM6 position. Therefore, to remove the impact of 5HT1B I repeated the analysis for G_s without using samples of it. As a result, I obtained 6x52-6x48 interaction ($P = 0.0023$) hinting the possible role of 6x48 in controlling TM6 position.

I demonstrated residues that are potentially associated with the structural difference in G_s coupled receptors onto active and inactive state ADRB2 structures (Fig 4.3B) and hypothesized that network including 6x48, 6x52 and 7x41 might trigger this change from upper layers of the network and lead changes at the lower layers. Previously, G7x41 have been identified as one of the most mutational intolerant residues for G_s coupled ADRB2 (Jones et al., 2020).

To further investigate the molecular impact of G7x41 we benefited from molecular dynamics simulation performed by our collaborators. We built three different systems by mutating glycine to cysteine, leucine, and glutamine (Fig 4.4A). We specifically chose these variants because they are already present on non-coupler close homologs of ADRB2 (Figure 4.1E) within aminergic subfamily. We performed simulations for both active and inactive states of the receptor to eliminate state specific biases. We interpreted the impact of the mutations based on two different criteria. Firstly, based on GPCRdb distances wild-type receptor conserves its active state more than other three systems (Figure 4.4B). The major downregulation of G315L mutation is in accordance with the previous experiments (Arakawa, Chakraborty, Upadhyaya, Eilers, Reeves, Smith & Chelikani, 2011; Jones et al., 2020). As a next step we wanted to explore the molecular changes based on changes in contact scores.

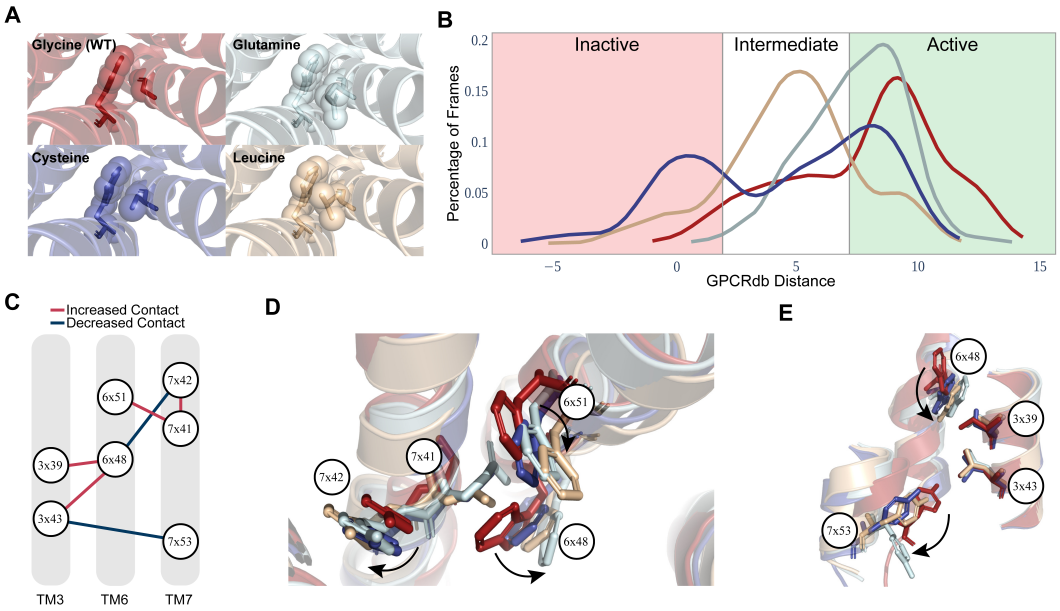

Figure 4.4 Molecular Dynamics Simulations Show Importance of G7x41
(A) Initial conformation of WT and non-coupler variants. (B) GPCRdb distance distribution of the simulation trajectories. (C) Common changes observed in absence of glycine a 7x41. (D, E) The molecular demonstration of the interactions in part C. Average cluster structures were used for representation. The arrows represent the movements of the residues.

11 frames with from each simulation were selected as representatives of the whole trajectory. Because we run 7 replicates for each variant we had 77 data point indicating the contact score of each residue pair at different moments of the simulation. Then we compared 77 mutated and 77 WT contact scores to identify the impact caused in absence of glycine by using two sided t test ($P < 0.01$). To find the impact of not having glycine we identified significantly altered interactions for all systems with mutations. In total, 135 residue pairs were identified for active-state simulations and 83 for inactive-state simulations. When we viewed residue pairs as an interaction network we identified an evolutionary conserved route that inactivates the active receptor through NPxxY motif (Fig 4.4C). We demonstrated that route on average simulation frames for each simulation system (Fig 4.4D and E). Based on molecular dynamics simulation we are able to explain the importance of G7x41. When three non-coupler variants cysteine, leucine and glutamine introduced to ADRB2 the bulkiness of 7x41 increases which physically blocks the interaction between 7x42 and 6x48 (Fig 4.4D). Following that 6x48 makes more contact with TM3 (3x43 and 3x39) and tightens the interactions between TM3 and TM6 (Fig 4.4E). Increased TM3-TM6 interactions weakens the contacts between TM3 and TM7 and reverses important interactions required for TM6 tilt (Zhou et al., 2019). Moreover, the increased distance between 3x43 and 7x53 can be explained weakening of TM6 and TM7 contacts (Fig 4.4E). Supporting our hypothesis that glycine is important for achieving a larger TM6 tilt, we observed that cysteine and leucine mutations results in stronger 3x43-6x40 interaction ($P < 0.01$) inhibiting TM6 movement. To understand if the changes we observed are specific to active-state simulations, we checked same residue pairs in inactive-state simulations as well. The contacts made by 7x42, 7x41, 6x51, and 6x48 changed in a similar manner. To summarize, MD

simulations suggest that all three variants makes the receptor less active and two of these variants (cysteine and leucine) inhibits the outward movement of TM6 compared to wild-type. Because larger TM6 movement is almost exclusively observed for G_s coupled receptors it is highly likely that G7x41 controls G_s selectivity by controlling TM6 conformation. However, further experimentation is required to understand the role of TM6 tilt and G7x41 in determining G_s coupling selectivity.

5. DISCUSSION

In light of our results and previous literature we created a model for G protein coupling selectivity combines series of factors. As in an electrical circuit there should be no disconnection in any part of the wire for electricity to be transmitted and all switches must be turned on accordingly. We propose existence of three main switches within GPCRs that are required to be sequentially turned on for receptors to be activated. Thus, we call our model as "sequential switches of activation" (Fig 6.1). Based on our model GPCRs contain three main switches for ligand binding, G protein specific activation mechanisms and G protein contact. Ligand binding switch is turned on if the receptor successfully binds to its ligand and leads structural changes at the lower layers of the receptor. The switch for the G protein specific activation mechanism checks if the receptor established a G protein specific conformation. Lastly, G protein and the receptor needs to make sufficient contacts and induce GTPase activity. Some mutations can lead to constitutional activation, which can be considered as a short circuit which ignores existing molecular switches. In contrary, mutations that damage molecular switches can inhibit coupling to a particular G protein. Because the model that we proposed includes all parts the receptor, it is complementary to the key and lock model (Flock et al., 2017). By the help of evolutionary analysis we know have a more complete view of determinants of the G protein coupling.

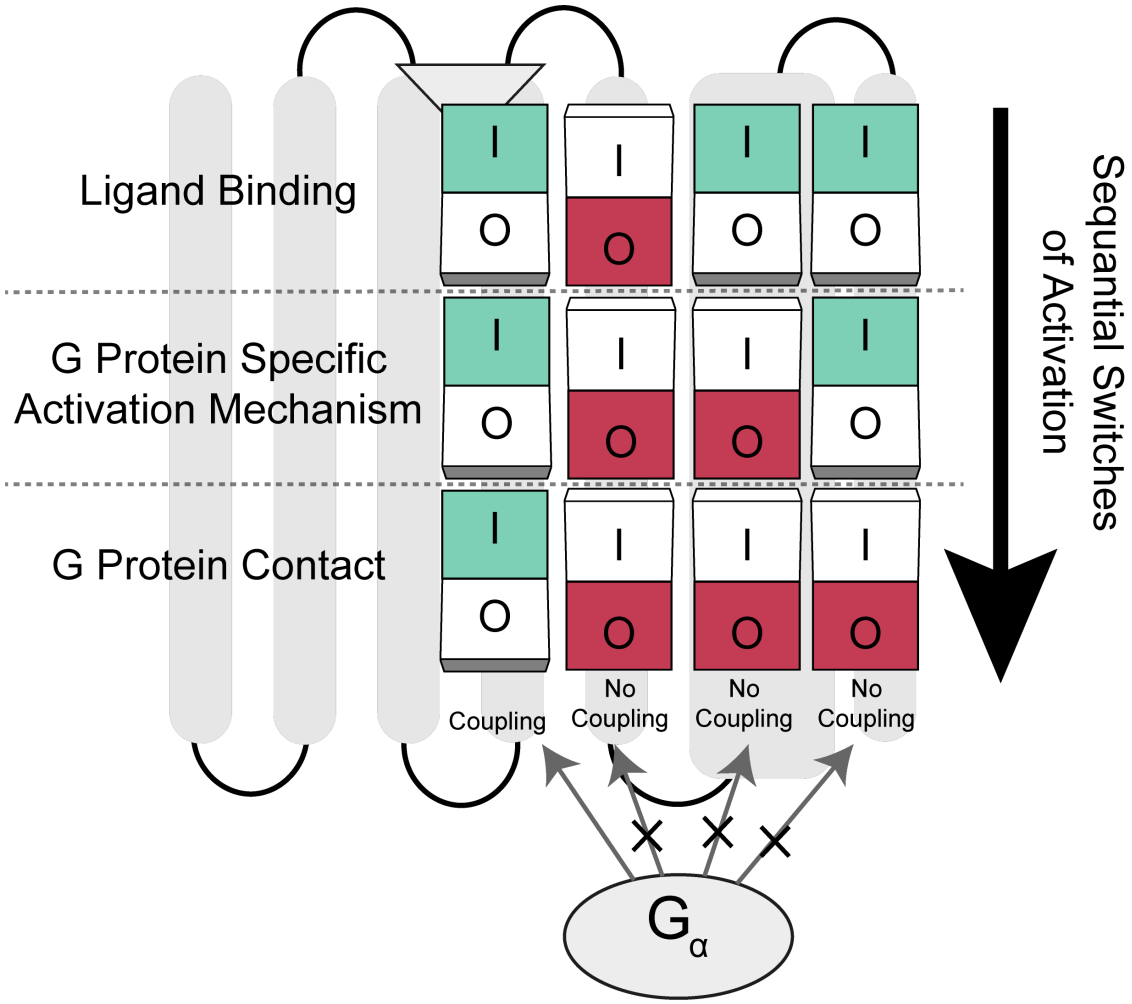

Figure 5.1 A Comprehensive Model for G protein Coupling Selectivity: Sequential Switches of Activation
The model proposes existence of three molecular switches that all should be turned on in an extracellular to intracellular direction for a successful G protein engagement.

In this work, we identified positions that are specifically conserved within aminergic receptors having similar coupling profiles. We used two different approaches to obtain as much as position we could (Fig 4.1C). Because we didn't perform any wet-lab experiments verifying positions we identified, some of the potential determinants can have a correlation relationship rather than cause and effect. Furthermore, the we did not include receptor level conservation into consideration and looked for the positions that are shared for all aminergic receptors. Thus, to identify all selectivity-determining positions, each receptor should be analyzed individually.

Although we did not provide any experimental results validating our findings, the level of evidence is enough to claim the existence of receptor-wide determinants. A recent evolution based analysis (Seo, Heo, Kim, Chung & Yu, 2021) also identified co-evolving residues on G proteins and receptors. These positions were not limited to the coupling interface only. We benefited from the previously published G_s activity data produced by Jones et al to demonstrate that non-coupler substitutions decrease receptor activity (Fig 4.1D)(Zhou et al., 2019). Furthermore, we showed that some of these conserved residues are part of G protein specific activation mech-

anisms (Fig 4.3B–D). It should be noted that as the number of structures increase, it is likely that the networks for G_q, G_o, and G_{i1} can change as well. As a next step, we investigated the molecular importance of G7x41 by analyzing molecular dynamics simulation trajectories of non-coupler variants. Because some non-coupler receptors also conserve glycine at 7x41, we cannot ignore its impact on general receptor activation. However, it might additionally control G_s selectivity by controlling the movement of TM6. Experimental validation could be an important last step that is required to label any position as selectivity determinant truly. Overall, the amount of evidence is enough to propose that the positions controlling coupling selectivity is dispersed through out the receptor.

Within the scope of this thesis, we analyzed receptor activation in monomer form only. However, we know that GPCRs frequently form dimers and heterodimers (Terrillon & Bouvier, 2004). Because possible dimerization of the receptor can change its structural conformation, it may promote or inhibit G protein coupling (Charles, Mostovskaya, Asas, Evans, Dankovich & Hales, 2003; George, Fan, Xie, Tse, Tam, Varghese & O'Dowd, 2000). Therefore, these structural changes may alter the coupling preferences of the receptors. The impact of dimerization should be more pronounced for the receptors belonging to class C, because dimerization is a necessity for these receptors (Kniazeff, Prézeau, Rondard, Pin & Goudet, 2011). In this view, residues within the dimerization interface, or molecular changes occurring after a dimerization can also be analyzed in terms of coupling selectivity in the future.

As we emphasized that there is no universal solution for achieving coupling selectivity, we should highlight that our findings are limited to the aminergic GPCRs. To reveal all selectivity determinant positions phylogenetic history of each receptor should be investigated individually. By doing that it is possible to discover differences and similarities between different receptors and receptor groups. Despite the fact that evolutionary distant receptors achieve similar structural conformations when coupling to a particular G protein, it does not mean that the underlying mechanism is the same. In the future, we are certain that similar mechanisms will be proposed as the number of available GPCR structures increase.

BIBLIOGRAPHY

(2019). Uniprot: a worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515.

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1-2*, 19–25.

Adebali, O., Reznik, A. O., Ory, D. S., & Zhulin, I. B. (2016). Establishing the precise evolutionary history of a gene improves prediction of disease-causing missense mutations. *Genetics in Medicine*, *18*(10), 1029–1036.

Arakawa, M., Chakraborty, R., Upadhyaya, J., Eilers, M., Reeves, P. J., Smith, S. O., & Chelikani, P. (2011). Structural and functional roles of small group-conserved amino acids present on helix-h7 in the (2)-adrenergic receptor. *Biochim Biophys Acta*, *1808*(4), 1170–8.

Avet, C., Mancini, A., Breton, B., Le Gouill, C., Hauser, A. S., Normand, C., Kobayashi, H., Gross, F., Hogue, M., Lukasheva, V., St-Onge, S., Carrier, M., Héroux, M., Morissette, S., Fauman, E. B., Fortin, J.-P., Schann, S., Leroy, X., Gloriam, D. E., & Bouvier, M. (2022). Effector membrane translocation biosensors reveal g protein and arrestin coupling profiles of 100 therapeutically relevant gpcrs. *eLife*, *11*, e74101.

Ballesteros, J. A. & Weinstein, H. (1995). [19] integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in g protein-coupled receptors. In S. C. Sealfon (Ed.), *Receptor Molecular Biology*, volume 25 of *Methods in Neurosciences* (pp. 366–428). Academic Press.

Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, *81*(8), 3684–3690.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). Blast+: architecture and applications. *BMC Bioinformatics*, *10*(1), 421.

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973.

Charles, A. C., Mostovskaya, N., Asas, K., Evans, C. J., Dankovich, M. L., & Hales, T. G. (2003). Coexpression of δ-opioid receptors with μ receptors in gh3 cells changes the functional response to μ agonists from inhibitory to excitatory. *Molecular pharmacology*, *63*(1), 89–95.

Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K., & Stevens, R. C. (2007). High-resolution crystal structure of an engineered human 2-adrenergic g protein-coupled receptor. *Science*, *318*(5854), 1258–1265.

Chien, E. Y. T., Liu, W., Zhao, Q., Katritch, V., Won Han, G., Hanson, M. A., Shi, L., Newman, A. H., Javitch, J. A., Cherezov, V., & Stevens, R. C. (2010). Structure of the human dopamine d3 receptor in complex with a d2/d3 selective antagonist. *Science*, *330*(6007), 1091–1095.

Chung, K. Y., Rasmussen, S. G. F., Liu, T., Li, S., Devree, B. T., Chae, P. S., Calinski, D., Kobilka, B. K., Woods, V. L., & Sunahara, R. K. (2011). Conformational changes in the g protein gs induced by the 2 adrenergic receptor. *Nature*, *477*(7366), 611–615.

Cvicek, V., Goddard, William A., I., & Abrol, R. (2016). Structure-based sequence alignment of the transmembrane domains of all human gpcrs: Phylogenetic, structural and functional implications. *PLOS Computational Biology*, *12*(3), e1004805.

Du, Y., Duc, N. M., Rasmussen, S. G. F., Hilger, D., Kubiak, X., Wang, L., Bohon, J., Kim, H. R., Wegrecki, M., Asuru, A., Jeong, K. M., Lee, J., Chance, M. R., Lodowski, D. T., Kobilka, B. K., & Chung, K. Y. (2019). Assembly of a gpcr-g protein complex. *Cell*, *177*(5), 1232–1242.e11.

Fay, J. C. & Wu, C.-I. (2003). Sequence divergence, functional constraint, and selection in protein evolution. *Annual Review of Genomics and Human Genetics*, *4*(1), 213–235.

Filipek, S. (2019). Molecular switches in gpcrs. *Current Opinion in Structural Biology*, *55*, 114–120.

Flock, T., Hauser, A. S., Lund, N., Gloriam, D. E., Balaji, S., & Babu, M. M. (2017). Selectivity determinants of gpcr-g-protein binding. *Nature*, *545*(7654), 317–322.

Fredriksson, R., Lagerström, M. C., Lundin, L. G., & Schiöth, H. B. (2003). The g-protein-coupled receptors in the human genome form five main families. phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol*, *63*(6), 1256–72.

García-Nafría, J., Nehmé, R., Edwards, P. C., & Tate, C. G. (2018). Cryo-em structure of the serotonin 5-ht1b receptor coupled to heterotrimeric go. *Nature*, *558*(7711), 620–623.

George, S. R., Fan, T., Xie, Z., Tse, R., Tam, V., Varghese, G., & O'Dowd, B. F. (2000). Oligomerization of μ-and δ-opioid receptors: generation of novel functional properties. *Journal of Biological Chemistry*, *275*(34), 26128–26135.

Gu, X. (2003). Functional divergence in protein (family) sequence evolution. *Genetica*, *118*(2-3), 133–41.

Haga, K., Kruse, A. C., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., Zhang, C., Weis, W. I., Okada, T., Kobilka, B. K., Haga, T., & Kobayashi, T. (2012). Structure of the human m2 muscarinic acetylcholine receptor bound to an antagonist. *Nature*, *482*(7386), 547–551.

Harding, S. D., Jane, Faccenda, E., Southan, C., Alexander, S. P. H., Davenport, A. P., Adam, Spedding, M., & Jamie (2021). The iuphar/bps guide to pharmacology in 2022: curating pharmacology for covid-19, malaria and antibacterials. *Nucleic Acids Research*.

Hauser, A. S., Chavali, S., Masuho, I., Jahn, L. J., Martemyanov, K. A., Gloriam, D. E., & Babu, M. M. (2018). Pharmacogenomics of gpcr drug targets. *Cell*, *172*(1), 41–54.e19.

Hauser, A. S., Kooistra, A. J., Munk, C., Heydenreich, F. M., Veprintsev, D. B., Bouvier, M., Babu, M. M., & Gloriam, D. E. (2021). Gpcr activation mechanisms across classes and macro/microscales. *Nature Structural Molecular Biology*, *28*(11), 879–888.

Hoover, W. G. (1986). Constant-pressure equations of motion. *Physical Review A*,

34(3), 2499–2500.

Huang, C.-Y., Olieric, V., Ma, P., Howe, N., Vogeley, L., Liu, X., Warshamanage, R., Weinert, T., Panepucci, E., Kobilka, B., Diederichs, K., Wang, M., & Caffrey, M. (2016). In meso in situserial x-ray crystallography of soluble and membrane proteins at cryogenic temperatures. *Acta Crystallographica Section D Structural Biology*, *72*(1), 93–112.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., Grubmüller, H., & Mackerell, A. D. (2017). Charmm36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, *14*(1), 71–73.

Huerta-Cepas, J., Serra, F., & Bork, P. (2016). Ete 3: Reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution*, *33*(6), 1635–1638.

Im, D., Inoue, A., Fujiwara, T., Nakane, T., Yamanaka, Y., Uemura, T., Mori, C., Shiimura, Y., Kimura, K. T., Asada, H., Nomura, N., Tanaka, T., Yamashita, A., Nango, E., Tono, K., Kadji, F. M. N., Aoki, J., Iwata, S., & Shimamura, T. (2020). Structure of the dopamine d2 receptor in complex with the antipsychotic drug spiperone. *Nature Communications*, *11*(1).

Inoue, A., Raimondi, F., Kadji, F. M. N., Singh, G., Kishi, T., Uwamizu, A., Ono, Y., Shinjo, Y., Ishida, S., Arang, N., Kawakami, K., Gutkind, J. S., Aoki, J., & Russell, R. B. (2019). Illuminating g-protein-coupling selectivity of gpcrs. *Cell*, *177*(7), 1933–1947.e25.

Insel, P. A., Snead, A., Murray, F., Zhang, L., Yokouchi, H., Katakia, T., Kwon, O., Dimucci, D., & Wilderman, A. (2012). Gpcr expression in tissues and cells: are the optimal receptors being used as drug targets? *British journal of pharmacology*, *165*(6), 1613–1616.

Isberg, V., De Graaf, C., Bortolato, A., Cherezov, V., Katritch, V., Marshall, F. H., Mordalski, S., Pin, J.-P., Stevens, R. C., Vriend, G., & Gloriam, D. E. (2015). Generic gpcr residue numbers – aligning topology maps while minding the gaps. *Trends in Pharmacological Sciences*, *36*(1), 22–31.

Ishchenko, A., Stauch, B., Han, G. W., Batyuk, A., Shiriaeva, A., Li, C., Zatsepin, N., Weierstall, U., Liu, W., Nango, E., Nakane, T., Tanaka, R., Tono, K., Joti, Y., Iwata, S., Moraes, I., Gati, C., & Cherezov, V. (2019). Toward g protein-coupled receptor structure-based drug design using x-ray lasers. *IUCrJ*, *6*(6), 1106–1119.

Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). Charmm-gui: A web-based graphical user interface for charmm. *Journal of Computational Chemistry*, *29*(11), 1859–1865.

Jones, E. M., Lubock, N. B., Venkatakrishnan, A., Wang, J., Tseng, A. M., Paggi, J. M., Latorraca, N. R., Cancilla, D., Satyadi, M., Davis, J. E., Babu, M. M., Dror, R. O., & Kosuri, S. (2020). Structural and functional characterization of g protein–coupled receptors with deep mutational scanning. *eLife*, *9*.

Kang, Y., Kuybeda, O., de Waal, P. W., Mukherjee, S., Van Eps, N., Dutka, P., Zhou, X. E., Bartesaghi, A., Erramilli, S., Morizumi, T., Gu, X., Yin, Y., Liu, P., Jiang, Y., Meng, X., Zhao, G., Melcher, K., Ernst, O. P., Kossiakoff, A. A., Subramaniam, S., & Xu, H. E. (2018). Cryo-em structure of human rhodopsin bound to an inhibitory g protein. *Nature*, *558*(7711), 553–558.

Katoh, K. & Standley, D. M. (2013). Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and*

evolution, *30*(4), 772–780.

Katritch, V., Fenalti, G., Abola, E. E., Roth, B. L., Cherezov, V., & Stevens, R. C. (2014). Allosteric sodium in class a gpcr signaling. *Trends Biochem Sci*, *39*(5), 233–44.

Kim, K., Che, T., Panova, O., DiBerto, J. F., Lyu, J., Krumm, B. E., Wacker, D., Robertson, M. J., Seven, A. B., Nichols, D. E., Shoichet, B. K., Skiniotis, G., & Roth, B. L. (2020). Structure of a hallucinogen-activated gq-coupled 5-ht2a serotonin receptor. *Cell*, *182*(6), 1574–1588.e19.

Kimura, K. T., Asada, H., Inoue, A., Kadji, F. M. N., Im, D., Mori, C., Arakawa, T., Hirata, K., Nomura, Y., Nomura, N., Aoki, J., Iwata, S., & Shimamura, T. (2019). Structures of the 5-ht2a receptor in complex with the antipsychotics risperidone and zotepine. *Nature Structural Molecular Biology*, *26*(2), 121–128.

Kniazeff, J., Prézeau, L., Rondard, P., Pin, J.-P., & Goudet, C. (2011). Dimers and beyond: The functional puzzles of class c gpcrs. *Pharmacology Therapeutics*, *130*(1), 9–25.

Kooistra, A. J., Mordalski, S., Pándy-Szekeres, G., Esguerra, M., Mamyrbekov, A., Munk, C., Keserű, G. M., & David (2021). Gpcrdb in 2021: integrating gpcr sequence, structure and function. *Nucleic Acids Research*, *49*(D1), D335–D343.

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). Raxml-ng: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455.

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L., MacKerell, A. D., Klauda, J. B., & Im, W. (2016). Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field. *Journal of Chemical Theory and Computation*, *12*(1), 405–413.

Liu, X., Xu, X., Hilger, D., Aschauer, P., Tiemann, J. K. S., Du, Y., Liu, H., Hirata, K., Sun, X., Guixà-González, R., Mathiesen, J. M., Hildebrand, P. W., & Kobilka, B. K. (2019). Structural insights into the process of gpcr-g protein complex formation. *Cell*, *177*(5), 1243–1251.e12.

Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I., & Lomize, A. L. (2012). Opm database and ppm web server: resources for positioning of proteins in membranes. *Nucleic Acids Research*, *40*(D1), D370–D376.

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., & Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, *102*(18), 3586–3616.

Mackie, K. (2008). Cannabinoid receptors: Where they are and what they do. *Journal of Neuroendocrinology*, *20*(s1), 10–14.

Maeda, S., Qu, Q., Robertson, M. J., Skiniotis, G., & Kobilka, B. K. (2019). Structures of the m1 and m2 muscarinic acetylcholine receptor/g-protein complexes. *Science*, *364*(6440), 552–557.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). Iq-tree 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530–1534.

Miyagi, H., Asada, H., Suzuki, M., Takahashi, Y., Yasunaga, M., Suno, C., Iwata, S., & Saito, J.-I. (2020). The discovery of a new antibody for bril-fused gpcr structure determination. *Scientific Reports*, *10*(1).

Mollica Poeta, V., Massara, M., Capucetti, A., & Bonecchi, R. (2019). Chemokines and chemokine receptors: New targets for cancer immunotherapy. *Frontiers in Immunology*, *10*.

Niimura, Y. (2009). Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents. *Human Genomics*, *4*(2), 107.

Nosé, S. & Klein, M. L. (1983). Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, *50*(5), 1055–1076.

Okashah, N., Wan, Q., Ghosh, S., Sandhu, M., Inoue, A., Vaidehi, N., & Lambert, N. A. (2019). Variable g protein determinants of gpcr coupling selectivity. *Proceedings of the National Academy of Sciences*, 201905993.

Park, H., Lee, G. R., Heo, L., & Seok, C. (2014). Protein loop modeling using a new hybrid energy function and its application to modeling in inaccurate structural environments. *PLoS ONE*, *9*(11), e113811.

Parrinello, M. & Rahman, A. (1980). Crystal structure and pair potentials: A molecular-dynamics study. *Physical Review Letters*, *45*(14), 1196–1199.

Pándy-Szekeres, G., Munk, C., Tsonkov, T. M., Mordalski, S., Harpsøe, K., Hauser, A. S., Bojarski, A. J., & Gloriam, D. E. (2018). Gpcrdb in 2018: adding gpcr structure models and ligands. *Nucleic Acids Research*, *46*(D1), D440–D446.

Rasmussen, S. G. F., Devree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T. A., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K., & Kobilka, B. K. (2011). Crystal structure of the 2 adrenergic receptor–gs protein complex. *Nature*, *477*(7366), 549–555.

Rose, A. S., Elgeti, M., Zachariae, U., Grubmüller, H., Hofmann, K. P., Scheerer, P., & Hildebrand, P. W. (2014). Position of transmembrane helix 6 determines receptor g protein coupling specificity. *Journal of the American Chemical Society*, *136*(32), 11244–11247.

Schiöth, H. B., Nordström, K. J., & Fredriksson, R. (2007). Mining the gene repertoire and ests for g protein-coupled receptors with evolutionary perspective. *Acta Physiol (Oxf)*, *190*(1), 21–31.

Selçuk, B., Erol, I., Durdağı, S., & Adebali, O. (2022). Evolutionary association of receptor-wide amino acids with g protein–coupling selectivity in aminergic gpcrs. *Life Science Alliance*, *5*(10).

Selçuk, B. & Adebali, O. (2022). Common and selective signal transduction mechanisms of gpcrs. Progress in Molecular Biology and Translational Science. Academic Press.

Semack, A., Sandhu, M., Malik, R. U., Vaidehi, N., & Sivaramakrishnan, S. (2016). Structural elements in the gs and gq c termini that mediate selective g protein-coupled receptor (gpcr) signaling. *Journal of Biological Chemistry*, *291*(34),

17929–17940.

Seo, M. J., Heo, J., Kim, K., Chung, K. Y., & Yu, W. (2021). Coevolution underlies gpcr-g protein selectivity and functionality. *Scientific Reports*, *11*(1).

Shahraki, A., Işbilir, A., Dogan, B., Lohse, M. J., Durdagi, S., & Birgul-Iyison, N. (2021). Structural and functional characterization of allatostatin receptor type-c of thaumetopoea pityocampa, a potential target for next-generation pest control agents. *Journal of Chemical Information and Modeling*, *61*(2), 715–728.

Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katritch, V., Abagyan, R., Cherezov, V., Liu, W., Han, G. W., Kobayashi, T., Stevens, R. C., & Iwata, S. (2011). Structure of the human histamine h1 receptor complex with doxepin. *Nature*, *475*(7354), 65–70.

Suno, R., Lee, S., Maeda, S., Yasuda, S., Yamashita, K., Hirata, K., Horita, S., Tawaramoto, M. S., Tsujimoto, H., Murata, T., Kinoshita, M., Yamamoto, M., Kobilka, B. K., Vaidehi, N., Iwata, S., & Kobayashi, T. (2018). Structural insights into the subtype-selective antagonist binding to the m2 muscarinic receptor. *Nature Chemical Biology*, *14*(12), 1150–1158.

Terrillon, S. & Bouvier, M. (2004). Roles of g-protein-coupled receptor dimerization. *EMBO reports*, *5*(1), 30–34.

Van Eps, N., Altenbach, C., Caro, L. N., Latorraca, N. R., Hollingsworth, S. A., Dror, R. O., Ernst, O. P., & Hubbell, W. L. (2018). G(i)- and g(s)-coupled gpcrs show different modes of g-protein binding. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(10), 2383–2388.

Vass, M., Podlewska, S., de Esch, I. J. P., Bojarski, A. J., Leurs, R., Kooistra, A. J., & de Graaf, C. (2019). Aminergic gpcr–ligand interactions: A chemical and structural map of receptor mutation data. *Journal of Medicinal Chemistry*, *62*(8), 3784–3839. PMID: 30351004.

Waldhoer, M., Bartlett, S. E., & Whistler, J. L. (2004). Opioid receptors. *Annual Review of Biochemistry*, *73*(1), 953–990. PMID: 15189164.

Wang, C., Jiang, Y., Ma, J., Wu, H., Wacker, D., Katritch, V., Han, G. W., Liu, W., Huang, X. P., Vardy, E., McCorvy, J. D., Gao, X., Zhou, X. E., Melcher, K., Zhang, C., Bai, F., Yang, H., Yang, L., Jiang, H., Roth, B. L., Cherezov, V., Stevens, R. C., & Xu, H. E. (2013). Structural basis for molecular recognition at serotonin receptors. *Science*, *340*(6132), 610–614.

Wang, J., Gareri, C., & Rockman, H. A. (2018). G-protein–coupled receptors in heart disease. *Circulation Research*, *123*(6), 716–735.

Wang, J. & Miao, Y. (2019). Mechanistic insights into specific g protein interactions with adenosine receptors. *The Journal of Physical Chemistry B*, *123*(30), 6462–6473.

Wu, E. L., Cheng, X., Jo, S., Rui, H., Song, K. C., Dávila-Contreras, E. M., Qi, Y., Lee, J., Monje-Galvan, V., Venable, R. M., Klauda, J. B., & Im, W. (2014). Charmm-guimembrane buildertoward realistic biological membrane simulations. *Journal of Computational Chemistry*, *35*(27), 1997–2004.

Xia, R., Wang, N., Xu, Z., Lu, Y., Song, J., Zhang, A., Guo, C., & He, Y. (2021). Cryo-em structure of the human histamine h1 receptor/gq complex. *Nature Communications*, *12*(1).

Xiao, P., Yan, W., Gou, L., Zhong, Y. N., Kong, L., Wu, C., Wen, X., Yuan, Y., Cao, S., Qu, C., Yang, X., Yang, C. C., Xia, A., Hu, Z., Zhang, Q., He, Y. H.,

Zhang, D. L., Zhang, C., Hou, G. H., Liu, H., Zhu, L., Fu, P., Yang, S., Rosenbaum, D. M., Sun, J. P., Du, Y., Zhang, L., Yu, X., & Shao, Z. (2021). Ligand recognition and allosteric regulation of drd1-gs signaling complexes. *Cell*, *184*(4), 943–956.e18.

Xu, P., Huang, S., Mao, C., Krumm, B. E., Zhou, X. E., Tan, Y., Huang, X.-P., Liu, Y., Shen, D.-D., Jiang, Y., Yu, X., Jiang, H., Melcher, K., Roth, B. L., Cheng, X., Zhang, Y., & Xu, H. E. (2021). Structures of the human dopamine d3 receptor-gi complexes. *Molecular Cell*, *81*(6), 1147–1159.e4.

Yang, F., Ling, S., Zhou, Y., Zhang, Y., Lv, P., Liu, S., Fang, W., Sun, W., Hu, L. A., Zhang, L., Shi, P., & Tian, C. (2020). Different conformational responses of the 2-adrenergic receptor-gs complex upon binding of the partial agonist salbutamol or the full agonist isoprenaline. *National Science Review*.

Yin, J., Chen, K.-Y. M., Clark, M. J., Hijazi, M., Kumari, P., Bai, X.-C., Sunahara, R. K., Barth, P., & Rosenbaum, D. M. (2020). Structure of a d2 dopamine receptor–g-protein complex in a lipid membrane. *Nature*, *584*(7819), 125–129.

Yin, W., Zhou, X. E., Yang, D., de Waal, P. W., Wang, M., Dai, A., Cai, X., Huang, C.-Y., Liu, P., Wang, X., Yin, Y., Liu, B., Zhou, Y., Wang, J., Liu, H., Caffrey, M., Melcher, K., Xu, Y., Wang, M.-W., Xu, H. E., & Jiang, Y. (2018). Crystal structure of the human 5-ht(1b) serotonin receptor bound to an inverse agonist. *Cell discovery*, *4*, 12–12.

Zhou, Q., Yang, D., Wu, M., Guo, Y., Guo, W., Zhong, L., Cai, X., Dai, A., Jang, W., Shakhnovich, E. I., Liu, Z.-J., Stevens, R. C., Lambert, N. A., Babu, M. M., Wang, M.-W., & Zhao, S. (2019). Common activation mechanism of class a gpcrs. *eLife*, *8*.

Zhuang, Y., Krumm, B., Zhang, H., Zhou, X. E., Wang, Y., Huang, X.-P., Liu, Y., Cheng, X., Jiang, Y., Jiang, H., Zhang, C., Yi, W., Roth, B. L., Zhang, Y., & Xu, H. E. (2021). Mechanism of dopamine binding and allosteric modulation of the human d1 dopamine receptor.

Zhuang, Y., Xu, P., Mao, C., Wang, L., Krumm, B., Zhou, X. E., Huang, S., Liu, H., Cheng, X., & Huang, X.-P. (2021). Structural insights into the human d1 and d2 dopamine receptor signaling complexes. *Cell*, *184*(4), 931–942. e18.

Zou, Y., Weis, W. I., & Kobilka, B. K. (2012). N-terminal t4 lysozyme fusion facilitates crystallization of a g protein coupled receptor. *PLoS ONE*, *7*(10), e46039.