



# Evolutionary association of receptor-wide amino acids with G protein–coupling selectivity in aminergic GPCRs

Berkay Selçuk<sup>1</sup> , Ismail Erol<sup>2,3</sup> , Serdar Durdagi<sup>2</sup> , Ogün Adebali<sup>1,4</sup>

**G protein-coupled receptors (GPCRs) induce signal transduction pathways through coupling to four main subtypes of G proteins ( $G_s$ ,  $G_i$ ,  $G_q$ , and  $G_{12/13}$ ), selectively. However, G protein selective activation mechanisms and residual determinants in GPCRs have remained obscure. Herein, we performed extensive phylogenetic analysis and identified specifically conserved residues for the aminergic receptors having similar coupling profiles. By integrating our methodology of differential evolutionary conservation of G protein–specific amino acids with structural analyses, we identified specific activation networks for  $G_s$ ,  $G_i$ ,  $G_o$ , and  $G_q$ . To validate that these networks could determine coupling selectivity we further analyzed  $G_s$ –specific activation network and its association with  $G_s$  selectivity. Through molecular dynamics simulations, we showed that previously uncharacterized Glycine at position 7x41 plays an important role in receptor activation and it may determine  $G_s$  coupling selectivity by facilitating a larger TM6 movement. Finally, we gathered our results into a comprehensive model of G protein selectivity called “sequential switches of activation” describing three main molecular switches controlling GPCR activation: ligand binding, G protein selective activation mechanisms, and G protein contact.**

DOI [10.26508/lsa.202201439](https://doi.org/10.26508/lsa.202201439) | Received 8 March 2022 | Revised 12 May 2022 | Accepted 13 May 2022 | Published online 25 May 2022

## Introduction

G protein–coupled receptors (GPCRs) constitute a significant group of membrane-bound receptors that contain five different classes (Fredriksson et al, 2003; Rosenbaum et al, 2009). The aminergic subfamily of receptors is present in class A and includes receptors for dopamine, serotonin, epinephrine, histamine, trace amine, and acetylcholine (Vass et al, 2019). With a large number of known coupling profiles, experimental structures, and mutagenesis experiments available, aminergic receptors are by far the most studied subfamily of GPCRs. These receptors can couple with different heterotrimeric G proteins which induce distinct downstream signaling pathways (Wettschureck & Offermanns, 2005). Disruption

of the proper receptor activation is likely to be the cause of diseases such as coronary heart disease (Wang et al, 2018a) or major depression (Catapano & Manji, 2007; Senese et al, 2018). Therefore, understanding the molecular mechanisms of coupling selectivity is crucial for developing better therapeutics and diagnostics.

With the advancement of new methodologies, two recent studies have revealed the G protein–coupling profiles of a large set of receptors. Inoue et al (2019) have used a shedding assay-based method to measure chimeric G protein activity for 11 unique chimeric G proteins representing all human subtypes and 148 human GPCRs. Because they have not managed to find an evident conserved motif determining G protein selectivity between receptors, they have built a machine learning–based prediction tool to identify sequence–based important features for each G protein. Similarly, Avet et al (2022) have used a BRET–based method detecting the recruitment of the G protein subunits to the receptor to reveal coupling profiles for 100 different receptors. The main strength of this study is that it does not require a modified G protein. Although both high–throughput studies largely agree with each other for certain G proteins, there are inconsistencies between the datasets. Thus, these valuable resources should be analyzed together in detail to gain more power in identifying the selectivity–determining factors in G protein coupling.

Several attempts have been made to identify molecular determinants of G protein coupling. Most of these (Chung et al, 2011; Semack et al, 2016; Du et al, 2019; Liu et al, 2019; Okashah et al, 2019) have focused on the G protein–coupling interface by analyzing contacts between receptor and the G protein. The others (Rose et al, 2014; Kang et al, 2018; Van Eps et al, 2018; Wang & Miao, 2019) have highlighted the structural differences between receptors that couple to different G proteins. Flock et al (2017) have analyzed the evolutionarily conserved positions of orthologous and paralogous G proteins and proposed the “lock and key” model. According to their model, G proteins (locks) have evolved with subtype–specific conserved barcodes that have been recognized by different subfamilies of receptors (keys). Because receptors with distinct evolutionary backgrounds can couple to the same G protein, receptors also must have evolved to recognize the existing barcodes. Although the model has explained the selectivity–determining

<sup>1</sup>Molecular Biology, Genetics and Bioengineering Program, Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey <sup>2</sup>Computational Biology and Molecular Simulations Laboratory, Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey <sup>3</sup>Department of Chemistry, Gebze Technical University, Gebze, Turkey <sup>4</sup>TUBITAK Research Institute for Fundamental Sciences, Gebze, Turkey

Correspondence: oadebali@sabanciuniv.edu

interactions between G protein and receptors, we still lack subfamily-specific receptor signaling mechanisms that involves but not limited to the G protein-coupling interface.

Despite the extensive research carried out to identify the determinants of G protein selectivity, selectivity-determining positions within receptors has remained underexplored. Here, we developed a novel methodology to identify a set of specifically conserved residues for the receptors sharing similar coupling profiles through the identification of orthologous receptors. Structural analyses revealed that specifically conserved positions are part of G protein-specific activation pathways that allow receptors to transduce the signal from the ligand-binding pocket to the G protein-coupling interface, induce the necessary conformational changes to get coupled by the relevant G protein subtype.

## Results

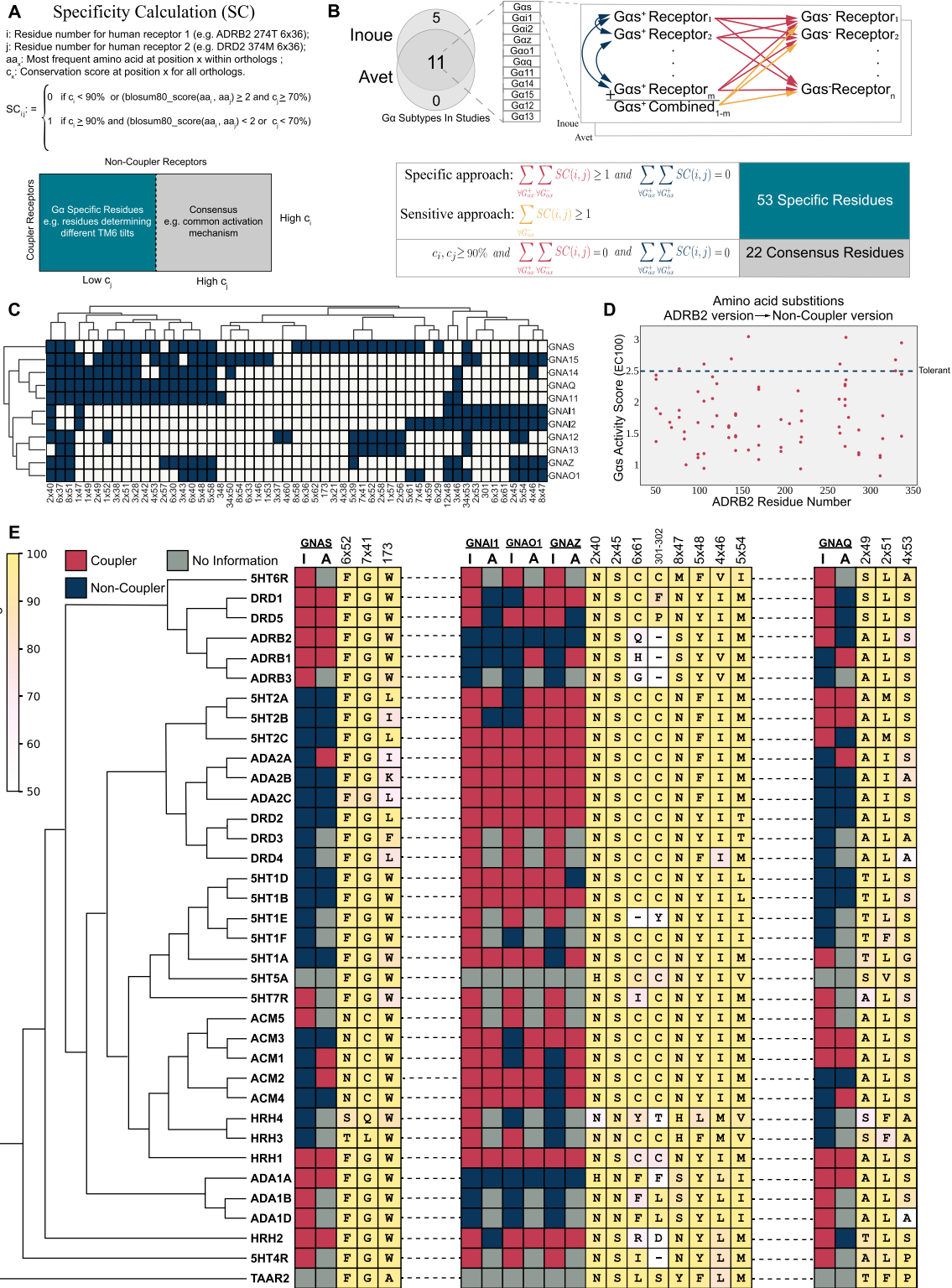
After a gene duplication event, paralogous clades might diverge from each other with respect to their functions. Therefore, evolutionary pressure against paralogous genes might differ. To perform a precise conservation analysis, we aimed to identify the gene duplication nodes in aminergic receptor evolution. We identified receptor subfamilies (orthologous and paralogous sequences) through a meticulous phylogenetic analysis. As we previously proposed (Adebali et al, 2016), the variations that observed in a paralog protein of interest may not be tolerated in the orthologous proteins. In our analyses, we only used orthologous receptors to define a subfamily of interest, members of which are likely to retain the same function. This approach greatly improved the sensitivity of conserved residue assignment for each human GPCR.

To link receptor evolution to its function, we identified residues that are conserved within the functionally equivalent orthologs for each aminergic receptor. For the residues that play a role in common receptor functions, we expect both clades to retain the amino acid residues with similar physicochemical properties. On the other hand, in the positions that serve receptor-specific functions, in our case the coupling selectivity, we expect to see differential conservation (Fig 1A). Therefore, we grouped receptors based on their known coupling profiles for 11 different G proteins (Fig 1B). We termed these groups as couplers (e.g., G<sub>s</sub> coupler receptors) and non-couplers, and performed a two-step enrichment method (Fig 1B) to distinguish specifically conserved residues in couplers from non-couplers. Initially, we used a specific approach to identify evident differentially conserved amino acid residues with high confidence. With the specific approach, residues were labeled as specifically conserved when there was a variation between the coupler and non-coupler receptors but not within coupler receptors (Fig 1B, red and blue arrows). This approach depends solely on the coupling profile datasets (Inoue et al, 2019; Avet et al, 2022) and thus, they may contain false-positive couplings. To tolerate the insensitivity introduced by potential false-positive couplings, we developed and used a sensitive approach enabling to obtain a more complete set of residues for each G protein subtype by allowing minor variations within the coupler receptors. With this method, we used a single comprehensive multiple sequence alignment (MSA) that combined all coupler receptors and their

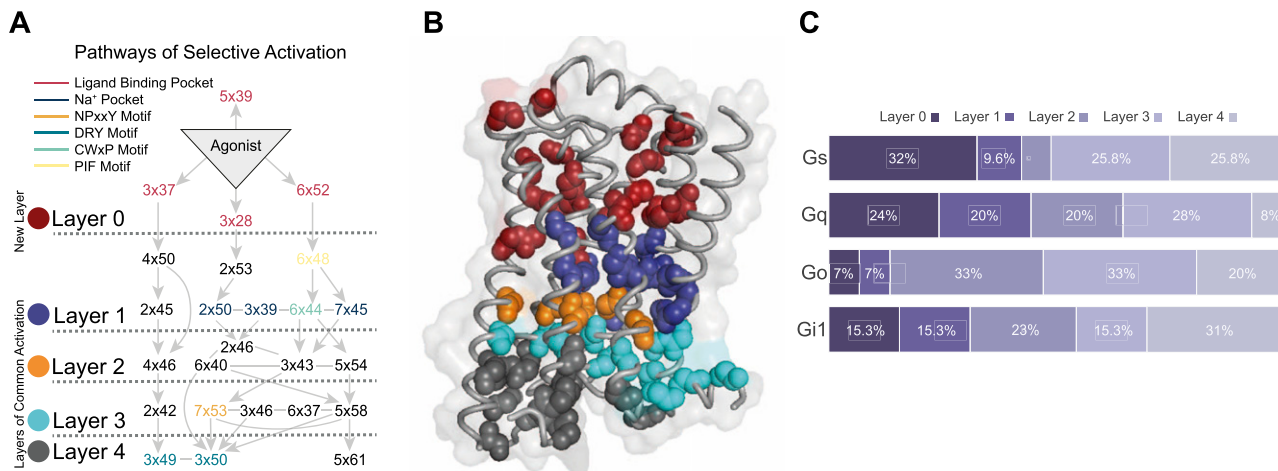
orthologs (Fig 1B, orange arrows), allowing minor variations within a group. We did not apply a sensitive approach to G<sub>12</sub> and G<sub>13</sub> because the low number of coupler receptors would likely cause a high number of false positives. Finally, we compared each aminergic receptor and identified positions that were conserved across all aminergic receptors (consensus) to link the specifically conserved residues to the general mechanism of receptor activation. In total, we identified 53 specifically conserved and 22 consensus residues. The distribution of the specific residues for each G protein is presented in Fig 1C.

We aimed to validate the functional impact of potentially deleterious variants that we observe within non-coupler receptors. Thus, we used a dataset (Jones et al, 2020) containing G<sub>s</sub> activity scores at EC100 for each possible mutation of ADRB2. 31 residues were identified for G<sub>s</sub> and the activity scores of non-coupler variants were plotted (Fig 1D). Non-coupler variants that we identified predominantly decrease G<sub>s</sub> activity when compared to the average activity of tolerant substitutions. Under normal conditions, the decrease in G<sub>s</sub> coupling can be attributed to various reasons including misfolding and decreased cell surface expression. However, the substitutions we proposed are not likely to disrupt general receptor functions because the substituting amino acids are indeed found and tolerated in non-coupler receptors (Fig 1E) having very high sequence and functional similarity. Additional to the G<sub>s</sub> coupling dataset, Kim et al (2020) mutated two of the residues we identified for G<sub>q</sub> coupling (8x47 and 6x37) to alanine and showed a decrease in G<sub>q</sub> activity compared to WT 5HT2A receptor which validates that variations at specifically conserved positions are not well-tolerated.

The experiments we mentioned show that non-coupler variants cause loss of function in receptors. However, losing the coupling function may not be associated with G protein-coupling selectivity. For an amino acid to be involved in G protein-coupling selectivity, it should govern functional G protein-specific roles. These roles can be recognition of G protein, ligand binding and/or establishing allosteric receptor conformations that may favor (or disfavor) the engagement with certain G protein subtypes. Hence, we manually assigned each residue into functional clusters such as coupling interface and ligand binding. For example, our method identified positions that are at the G protein-coupling interface such as 8x47 (Maeda et al, 2019; Kim et al, 2020; Zhuang et al, 2021b) and 6x36 (Rasmussen et al, 2011; Yang et al, 2020; Xiao et al, 2021) with no structural information taken into account. The residues that are in the coupling interface are in line with the model that Flock et al (2017) proposed and are likely important for proper G protein recognition. However, for the residues that we could not directly assign a role in G protein-coupling activity, we hypothesized that they should be a part of a network controlling the signal transduction from ligand-binding pocket to the G protein-coupling interface and establish the required selective structural conformations. To test this hypothesis, we explored the residue-level contact changes upon coupling to a G protein. We used an algorithm that is called Residue-Residue Contact Score (RRCS) which has been proposed to identify the common activation mechanism in class-A GPCRs (Zhou et al, 2019). We calculated ΔRRCS for each interacting residue pair by subtracting the contact scores of the active structure from the inactive structure. All the



**Figure 1. Selectivity-determining residues for each Gα subtype.** (A) The formula for specific residue identification. (B) The schema describes the comparisons between paralogous human receptors to find the specifically conserved residues for each Gα. Arrows represent a single comparison. (C) The distribution of specifically conserved residues for each Gα subtype and hierarchical clustering of them (complete linkage). (D) Possible variants of G<sub>s</sub> specific residues that are observed in non-coupler receptors are compared with the WT activity score. (E) Maximum-likelihood phylogenetic tree of aminergic receptors including coupling profiles, conservation information of selected specifically conserved residues (I, Inoue; A, Avet). The background color scale for each consensus amino acid correlates with their conservation (identity).



**Figure 2. Structural analysis of molecular pathways that are observed upon coupling with a heteromeric G protein complex.**

**(A)** The most common molecular signal transduction pathways from ligand-binding pocket to G protein–coupling interface. The arrows represent a contact change upon coupling to a G protein. The network is summarized and divided into different layers based on their functional relevance. **(B)** Projection of main chains of specifically conserved and consensus residues in different layers of activation on inactive ADRB2 structure (PDB ID 2RH1). **(C)** The distribution of specifically conserved residues for each analyzed G $\alpha$  subtype.

active structures we used contained a heteromeric G protein machinery coupled to a receptor. We filtered out residue pairs with  $|\Delta\text{RRCS}| \leq 0.2$  and only kept residues that are in our pool of conserved residues (75 residues in total). We analyzed the structures of eight different receptors with four different G proteins (see the Materials and Methods section). The structures we used were experimentally characterized except for one state of a single receptor. As we aimed to use the 10 active-state G<sub>s</sub>-coupled structures of DRD1, which lacks an experimental inactive structure, we used a model inactive DRD1 structure (Pándy-Szekeres et al, 2018) retrieved from GPCRdb (Kooistra et al, 2021).

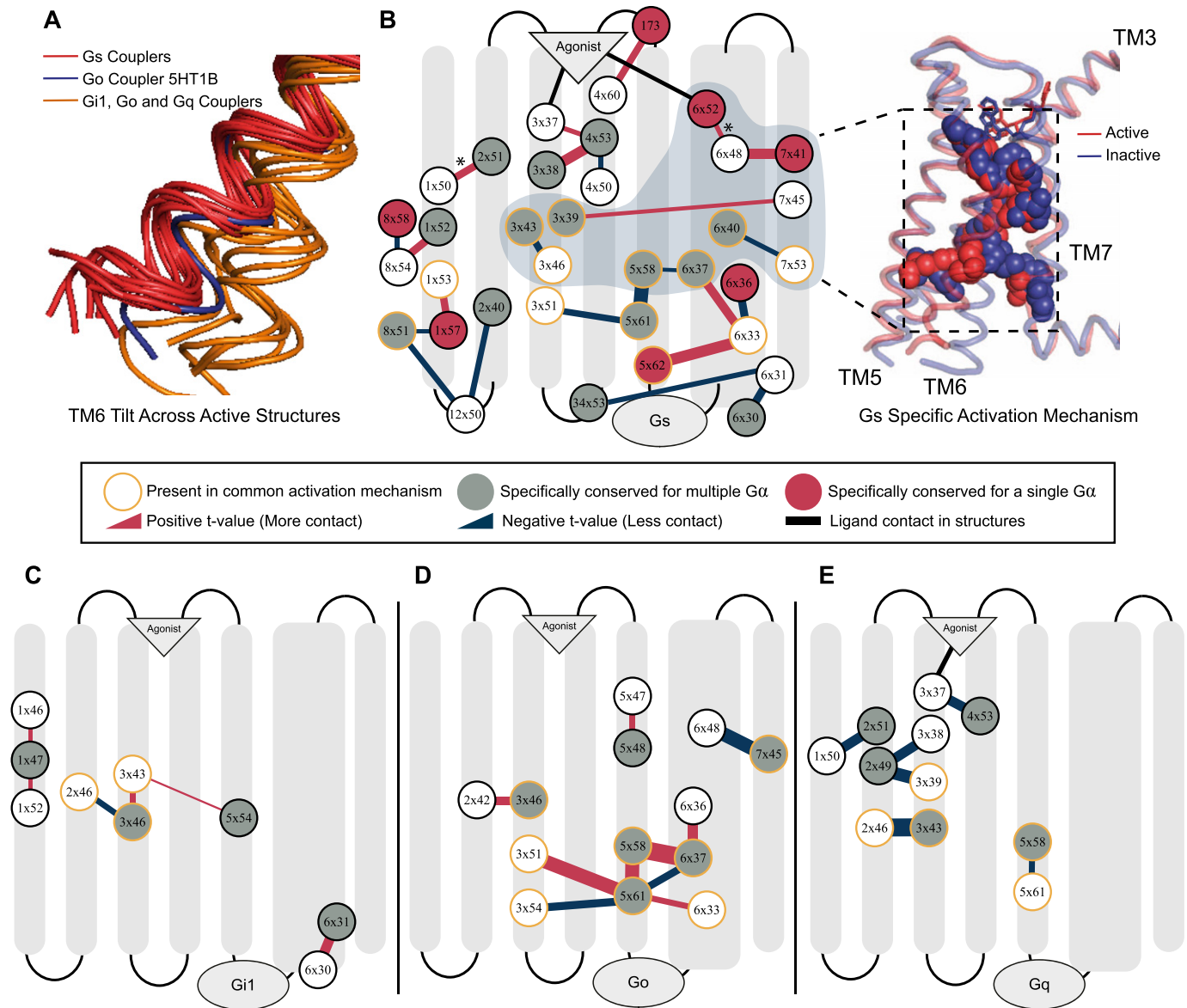
In total, we analyzed 41 pairs of active and inactive structures and identified  $\Delta\text{RRCS}$  values of activation networks. We analyzed each network and detected edges (increase or decrease in contact score) observed at least 36 times regardless of the sign of  $\Delta\text{RRCS}$  value to build a network that would represent all 41 networks. By using this network, we identified the most frequently used signal transduction paths (Fig 2A), connecting ligand-binding pocket to G protein–coupling interface and creating a basis for the routes that can induce coupling selectivity. We divided the receptor into five layers based on the sequential nature of interactions and illustrated the direction of signal transduction between layers. Additional to the four layers (1–4) that were previously proposed in the common activation mechanism (Zhou et al, 2019), we defined “Layer 0” which corresponds to the ligand-binding site. Though most of the signaling paths pass through important motifs such as Na<sup>+</sup> binding pocket and PIF (Katritch et al, 2014), it is remarkable that the novel path starting with a 3x37 does not require the involvement of any of these important motifs. Within the identified network, the signal is transmitted from the ligand-binding pocket to the G protein interface by using mainly TM2, TM3, and TM4. We projected all the residues onto an inactive structure of ADRB2 based on the layers they belong to (Fig 2B) to provide an insight into the locations of different layers.

To determine the contribution of each layer for G<sub>s</sub>, G<sub>i1</sub>, G<sub>o</sub>, and G<sub>q</sub>, we calculated the distribution of specific residues to different

layers (Fig 2C). Layer 0 and Layer 1 are more involved in the coupling for G<sub>s</sub> and G<sub>q</sub> relative to G<sub>i1</sub> and G<sub>o</sub>. For G<sub>o</sub>, 86% of the coupling-related residues are positioned in the layers (2, 3, and 4) closer to the G protein–binding site. Differences in these distributions indicate mechanistic differences between distinct coupling events.

To detect if the specifically conserved residues have differential roles in G protein coupling–related mechanisms, we grouped  $\Delta\text{RRCS}$ s (contact changes upon coupling to a G protein) for the receptors coupled to the same G protein and compared them with the rest by using two-sample t test. This approach yielded interaction changes ( $\Delta\Delta\text{RRCS}$ ) within the receptors that are significantly different ( $P < 0.01$ ) and specific for G<sub>s</sub>, G<sub>i1</sub>, G<sub>o</sub>, and G<sub>q</sub>. Significant contact changes occurring between 75 conserved residues were used to construct G protein–specific activation mechanisms. The constructed networks (Fig 3B–E) support our evolution-driven hypothesis and demonstrate that specifically conserved residues indeed have differential mechanistic roles in G protein coupling. In parallel to Fig 2C, networks for G<sub>s</sub> and G<sub>q</sub> contained ligand contacting residues (Fig 3A and E), whereas networks for G<sub>i1</sub> and G<sub>o</sub> do not. This can indicate that ligand binding could be more important for G<sub>s</sub> and G<sub>q</sub> coupled receptors. Although, G<sub>i1</sub> and G<sub>o</sub> belong to the same subfamily and they share eight of the specifically conserved residues (47% of the specifically conserved residues for G<sub>o</sub> and 62% for G<sub>i1</sub>) of G proteins their networks are totally different from each other. Moreover, even when we grouped the receptors coupled to Gi together, no significant difference in contact scores having a  $P$ -value less than 0.01 was observed for the shared specifically conserved residues (Fig 1C). This suggests that receptors coupling to Gi may not necessarily share a common activation mechanism. Therefore, these differences in activation networks could be one of the factors determining selectivity between G<sub>i1</sub> and G<sub>o</sub> coupled receptors.

Even though residues specifically conserved for the receptors sharing similar coupling profiles are part of G protein–specific activation networks, it is still not clear that these contact changes



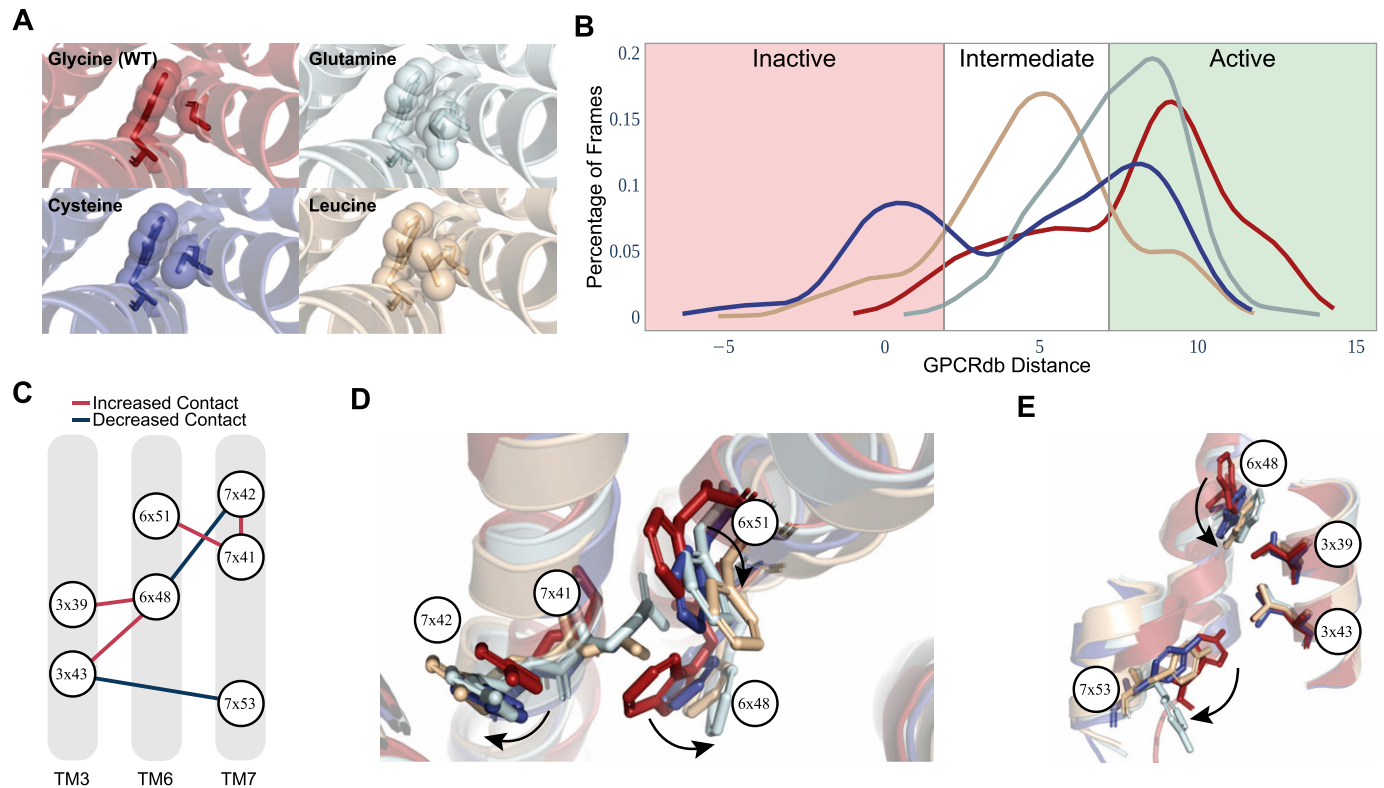
**Figure 3. Specific activation networks for G<sub>s</sub>, G<sub>i1</sub>, G<sub>o</sub> and G<sub>q</sub>.** (A) TM6 tilt comparison between the active receptors we used. Red: G<sub>s</sub> couplers, Orange: G<sub>o</sub>, G<sub>i1</sub>, and G<sub>q</sub>, Blue: 5HT1B G<sub>o</sub> coupler as an exception. (B) Interactions within the receptor that are specific ( $P < 0.01$ ) to G<sub>s</sub>. Red: increasing contact, blue: decreasing contact, orange circle: present in common activation mechanism, red fill: uniquely identified specific residue for G<sub>s</sub>, grey fill: G<sub>o</sub> specific residue. The width of the lines correlates with statistical significance. A group of residues that possibly facilitate in TM6 movement for G<sub>s</sub> coupling was shown on inactive (blue) and active (red) structures. (C, D, E) Specific interaction networks for G<sub>i1</sub>, G<sub>o</sub>, and G<sub>q</sub>.  $P < 0.1$  is used for G<sub>i1</sub>; \*: This interaction is identified only if 5HT1B is neglected from the comparison because of its larger TM6 movement.

are the basis for selective coupling, or they arise due to the physical interaction with a G protein itself. To show that these networks can determine selectivity we further analyzed the activation network for G<sub>s</sub>-coupled receptors. Previously, it was shown that receptors coupled to G<sub>s</sub> achieve a larger TM6 tilt (Rose et al, 2014; Van Eps et al, 2018) than the receptors coupled to other G proteins. Superimposition of the active structures that we used in our analysis (Fig 3A) is also in line with the previous findings. We hypothesized that if differential TM6 movement is a determinant for G<sub>s</sub>, the network we identified can modulate this structural difference. Furthermore, the requirement for a larger TM6 movement can be the reason why G<sub>s</sub>

specific activation mechanism is more complex than the rest (Fig 3B–E). An exception to this is the TM6 position of 5HT1B (García-Nafria et al, 2018) that is coupled to G<sub>o</sub> (Fig 3A, blue structure) because it achieved a slightly larger tilt. Thus, we performed an additional statistical test to reveal possible interactions that can promote larger TM6 movement by excluding the samples for 5HT1B and revealed the 6x52–6x48 interaction indicating the role of 6x48 in differential TM6 movement in G<sub>s</sub>-coupled receptors ( $P = 0.0023$ ).

We projected a part of G<sub>s</sub> specific activation network which we predicted to be associated with the differential TM6 movement onto experimentally resolved active (red, 3SN6) and inactive (blue,





**Figure 4. Analysis of molecular dynamics simulations reveals functional importance of glycine at 7x41.**

(A) Four different MD simulation systems were shown in their initial conformation. (B) For each simulation, distribution of frames with respect to their state of activation was shown, distance in Angstrom. (C) The common pathway representing the impact of the mutations at 7x41. (D, E) The common pathway was represented on average structures that were obtained in all MD trajectories for every mutation and WT. The movements of residues were represented with arrows.

2RH1) ADRB2 structures (Fig 3B). More specifically, we hypothesized that the network containing 6x52 and 7x41 triggers this structural difference because interactions at the upper layers are more likely to be leading a structural change. In agreement with our hypothesis, deep mutational scanning of ADRB2 (Jones et al, 2020), has revealed that 7x41 is the second and 6x48 is the fourth most intolerant residue to any mutations and, to our knowledge, no previous study has investigated the functional role of 7x41 until now. It is expected that a position that is crucial for  $G_s$  coupling to be one of the most intolerant residues for a receptor primarily coupled to  $G_s$ .

To validate our methodology and further understand the mechanistic insight of the relevance of the core transmembrane region in G protein coupling, we studied the glycine at position 7x41 as a test case and performed molecular dynamics (MD) simulations. We applied three different mutations, G315C, G315Q, and G315L, on monomeric active and inactive-state ADRB2 (Fig 4A). We particularly selected variants observed in acetylcholine and histamine receptors (Fig 1E) to validate our hypothesis that variants in non-coupler aminergic receptors at the same position are inactivating. We used two main metrics to assess the molecular impact of these three mutations. First, the comparison active/inactive states based on GPCRdb distances (see the Materials and Methods section) revealed that the WT receptor preserve the active state more than the variants (Fig 4B) and leucine residue was the most inactivating mutation. The significant inactivation through the integration of

leucine mutation is parallel to pre-existing experiments (Arakawa et al, 2011; Jones et al, 2020). Then, to identify the molecular changes in absence of glycine, we evaluated the significant contact differences ( $\Delta$ RRCs) between WT and mutated MD simulation trajectories.

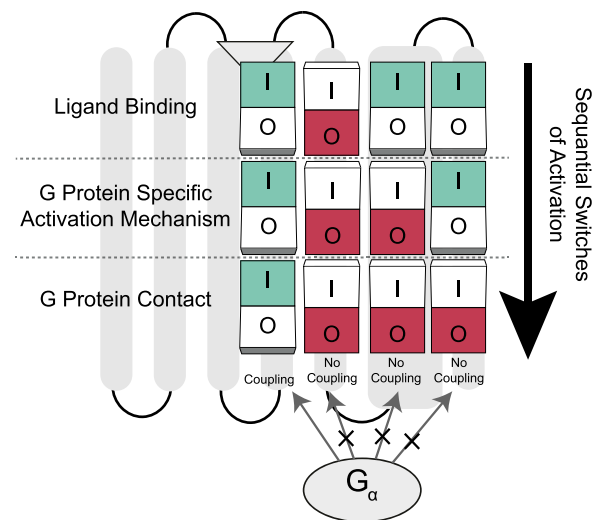
To examine the entire trajectory, we selected 11 frames from each simulation with 50-ns time intervals (in total 500 ns) for each replicate. Thus, we compared RRCs of 77 mutated and 77 WT frames for active-state simulations, whereas we compared 22 mutated and 22 WT frames for inactive-state simulations by using two-sided  $t$  test. For each mutation and activation state, we identified significant contact changes ( $P < 0.01$ ) and intersected common changes that we observed for all of the mutated systems. As a result, we identified 135 residue pairs for active and 83 residue pairs for inactive simulations. When we projected these residue pairs (135 residue pairs) as a contact network, we identified a conserved and highly affected pathway (Fig 4C) connecting ligand-binding pocket to NPxxY motif which showed changes towards the inactivation of the receptor. Then, we projected the identified molecular pathway onto average cluster structures that were produced by using the trajectories from all seven replicates (35,000 frames in total) for each mutation (Fig 4D and E). MD results suggested a pathway (Fig 4C) that explains the importance of G315: An increased bulkiness of the amino acid at 7x41 (by non-glycine amino acids) leads to increased contact with 7x42 and 6x51, whereas 7x41 physically impairs the interaction between 6x48 and 7x42. When 6x48 loses its contact

with 7x42 (Fig 4D), it increases its contact residues at TM3 3x43 and 3x39 (Fig 4E). Increased interactions between TM6 and TM3 loosen TM3-TM7 packing which is an important initiator of the TM6 tilt in class-A GPCRs (Zhou et al, 2019). In addition, it loosens the contacts between TM6 and TM7 through 6x48-7x42, 6x44-7x49, and 6x52-7x45, which explains the increased distance between 7x53 and 3x43 (Fig 4E). Moreover, the simulations of cysteine and leucine variants exhibited an increased contact between 3x43 and 6x40 ( $P < 0.01$ ) inhibiting the receptor activation through restricting outward TM6 movement. When we evaluated the inactive trajectories, we observed similar contact changes between 6x48, 6x51, 7x41, and 7x42 ( $P < 0.01$ ) proving that the simulation results are not biased toward active-state simulations. Thus, analysis of MD trajectories suggests that glycine at 7x41 plays an important role in receptor activation, and it is likely to control selectivity for  $G_s$  coupling by promoting a larger tilt of TM6 which we observe almost exclusively in  $G_s$ -coupled receptors. However, the roles of  $G^{7x41}$  and differential TM6 position in determining  $G_s$  coupling selectivity requires experimental validation.

## Discussion

By integrating our findings and current literature we propose a G protein selectivity model involving a series of modules. As pilots turn on switches in a pre-determined order before the takeoff, GPCRs must turn on their molecular switches for a specific type of G protein coupling to occur. If pilots fail to turn on all the switches properly because of an error, there will be no permission for them to depart. Similarly, all molecular switches must be turned on for receptors to engage with a G protein and induce downstream signaling pathways. For these reasons, we named our model “sequential switches of activation” (Fig 5). We propose the existence of three main switches within a GPCR structure. The first switch checks for binding of the proper agonist which induces conformational changes in the lower layers of the receptors. If an agonist makes the proper contacts with the receptor the first switch turns on. Then as a next step, receptors should be activated through G protein selective activation mechanisms which include multiple micro-switches to turn off the second main switch. Micro-switches represent the arrangement of inner contacts that are specific for G protein subtypes. When inner contacts are established properly the second switch turns on as well. As a third and last checkpoint, receptors should contain the set of residues that can recognize the ridges on G proteins according to the “key and lock” model that Flock et al (2017) suggested. When required contact between G protein and receptor is established, the third switch turns on and the receptor is successfully coupled by a subtype of G proteins. Mutations inducing constitutitional activity can be considered as a “short circuit” because they can bypass switches. On the other hand, mutations that halt the receptor’s ability to turn on a particular switch can prevent coupling. It is important to note that our model is inclusive of and complementary to the model Flock et al (2017) suggested. The combination of these two models gives us a more complete perspective on receptor-level determinants of coupling selectivity.

In our study, we used a novel phylogenetic approach to identify residues that are conserved among groups of receptors coupling to



**Figure 5. Sequential switches of activation model for G protein selectivity.** The model describes that all switches in different layers of receptors must be turned off for receptor activation and coupling of the G protein. If switches at upper layers are halted due to a mutation, the following switches become turned off which inhibits G protein coupling eventually.

a particular G protein. We identified the largest possible set of residues (Fig 1C) by combining sensitive and specific approaches together. Because of our greedy approach, whereas some positions could determine coupling selectivity, others may be “passenger” positions that may modulate core receptor functions. Moreover, the positions we identified are the ones that are shared among all aminergic receptors and lack receptor-level variations. Previous studies on chimeric GPCRs (Wess, 1998, 2021; Wong, 2003) point out the importance of ICL3 in determining coupling selectivity. Although we identified residues that contact with G proteins, our analyses did not reveal any possible determinants at ICL3. This indicates that the determinants at ICL3 are not shared between aminergic receptors and rather be specific to individual receptors. Alternatively, in nature, there may not be a solution for G protein–coupling selectivity determination with ICL3 only. Experimentally constructed chimeric receptor activation should be handled with caution because they cannot be evaluated as a part of receptor evolution. Thus, to identify all selectivity-determining positions, each receptor should be analyzed individually.

Although our study does not include any direct experimental evidence that coupler or non-coupler variants alter coupling selectivity, it provides sufficient evidence to support the existence of receptor-wide selectivity determinants not only at the G protein–coupling site but throughout receptors including the ligand-binding site. A recent analysis (Seo et al, 2021) also shows that serotonin and dopamine receptors contain positions co-evolving with the positions on G proteins they are coupled to. Supporting our conclusions, selectivity-determining positions were shown dispersed throughout the receptor. In our study, we used  $G_s$  coupling data from deep mutational scanning of ADRB2 performed by Jones et al to show that non-coupler variants cause loss of function (Fig 1D) (Jones et al, 2020), their roles in determining coupling selectivity should be clarified further. For that purpose, we used the RRCs

algorithm and revealed the involvement of specifically conserved residues in G protein-specific activation mechanisms (Fig 3B–D) which suggests their role in determining coupling selectivity. We should note that because of the scarcity of  $G_{q-}$ ,  $G_{o-}$ , and  $G_{i1-}$ -coupled structures, the networks we provided could be modified in the future as the number of G protein-coupled experimental structures increases. As a third layer of evidence, we identified the role of a previously uncharacterized  $G^{7x41}$  (Jones et al, 2020) for ADRB2 and  $G_s$ -coupled receptors through MD simulations (Fig 4C). Although we cannot rule out the potential effect of  $G^{7x41}$  in non- $G_s$  activation, we can conclude that it has a critical importance for determining  $G_s$  coupling selectivity. The fact that  $G^{7x41}$  is dispensable for  $G_i$  couplers suggests that it may not be as critical for those GPCRs and  $G_i$  activation. To summarize, multiple layers of evidence suggest that G protein selectivity determinants for aminergic receptors are likely distributed receptor-wide.

The conclusions of this study are limited aminergic receptors only because there has been no supporting evidence for a common selective mechanism that might present for all class-A GPCRs. Therefore, it is necessary to handle each GPCR subfamily separately to identify subfamily-specific selectivity determinants. With such an effort, it may be possible to discover commonalities and differences between different subfamilies of GPCRs. Although different subfamilies of receptors couple with a G protein by having similar structural conformations, underlying mechanisms for achieving a conformation might vary. As the number of solved G protein-coupled receptor structures increases in the protein data bank, it is inevitable that new selectivity determinants and similar mechanisms will be discovered in near future.

## Materials and Methods

### Sequence selection

Sequence selection is the very first step of this study. We used the BLAST+ (Camacho et al, 2009) algorithm to obtain homologous protein sequences from other organisms. We blasted a human target protein to find its homologs. The UniProt (UniProt Consortium, 2019) database is used as a source for the sequences. We retrieved all the sequences until the third human protein from the blast output.

### MSA #1

After sequence selection, the next step is performing MSA for obtained sequences. For this purpose, we used MAFFT (Katoh & Standley, 2013) “einsi” option which allows for large gaps. This option allows us to align multiple homologous regions of different receptors.

### Maximum likelihood (ML) tree #1

The MSA was used to produce a maximum-likelihood (ML) tree. ML trees helped us find relationships between different proteins. ML Tree 1 was used to identify the clade which contains our protein of interest. For ML tree construction we use the IQ-TREE version 2.0.6

(Minh et al, 2020) We used 1,000 ultrafast bootstraps and JTT+I+G4+F substitution model. IQ-TREE is used at this step for mainly its high speed in bootstrapping.

### Obtaining gene clade

For making modifications to the ML trees we use a Python-based tool ETE3 (Huerta-Cepas et al, 2016). To analyze a tree, we first need to root it properly. We chose the third human protein from our BLAST results, as an outgroup. Then, we traversed from our target human leaf node to root until we reached a clade containing another human protein. After each move, we analyzed the species content of the clades we are observing. When a clade contained species that were not observed in previous moves, we included all of the leaf nodes in our analysis. On the other hand, when a clade contains a previously observed species, we exclude that clade from our analysis because seeing a species at lower phylogenetic levels is an indication of a differential gene loss event. We continued with the remaining sequences and produced a MSA with them.

### MSA trimming

MSA trimming is needed to remove some of the noise from the alignment and it speeds up tree reconstruction. MSA trimming removes positions that are misleading for tree production. For example, positions having too many gaps can be removed from the alignment. We used trimAl (Capella-Gutiérrez et al, 2009) with automated1 option which is stated to be the best option for constructing maximum-likelihood trees.

### Maximum likelihood tree #2

ML tree 2 was used to identify the paralogous sequences that we have in our analysis. For ML tree construction, we used the RaxML-NG version 0.9.0 (Kozlov et al, 2019) –search option with JTT+I+G4+F substitution model.

### Paralog trimming

Paralog trimming is a key part of our approach. After gene duplication, one of the paralogous clades tends to diverge more than the other. Unless the diverged clade is removed from our analyses (MSA), it might introduce false divergence signals in conservation calculation. For this reason, we need to exclude diverged paralogs from our analyses. We used the second ML tree for the detection of the diverged paralogs.

We first calculated the global alignment scores (BLOSUM62 is used) of every sequence on the ML tree 2 with respect to our human target sequence. We assessed each internode having two child clades based on the number of leaf nodes and species they contain. When two child clades contained at least one identical species, we looked for a significant divergence between the clades in terms of global alignment scores to label one clade as paralogous. Also, we need those clades to be evolutionarily comparable, thus we compared the taxonomic level of the organisms between two clades. If the clades are comparable with each other, we applied two-sample *t* test for by using the global alignment scores. If one



clade has significantly lower similarity scores ( $P \leq 0.1$ ) that clade is labeled as a diverged paralogous clade. We applied the same approach for detecting the taxonomic level of the organisms and common lineage numbers with *Homo sapiens* were used this time ( $P \leq 0.1$ ). If the clades are evolutionarily comparable and one clade had a significantly lower global alignment score, all of the sequences belonging to that clade were eliminated.

When two of the clades contained less than three sequences each, it was hard to obtain significance. Therefore, for those cases, we compared the average global alignment scores and eliminated the clade with a lower average. For the remaining situations, we do not remove any of the clades.

### Conservation calculation

After obtaining orthologs we used them to calculate the conservation scores for each receptor.

The conservation percentage for a certain residue is calculated as follows:

- (1) Find the most frequent amino acid for a certain position in the MSA.
- (2) After finding the most frequent amino acid, we compared it with other alternatives in that position. When comparing amino acids, we calculated BLOSUM80 score for each of them. If the BLOSUM80 score is higher than 2 we accept it as an “allowed” substitution because it means that these amino acids replace each other frequently and have similar properties.
- (3) The gaps are not included while calculating the conservation percentage.
- (4) If gaps are more than 50%, we categorized that position as a gap.
- (5) The conservation score is equal to the number of most frequently observed and “allowed” amino acids over the number of all non-gap positions

### Specificity calculation

For a position to be specific or consensus the criteria are the following:

- (1) First, we need one alignment of two proteins with their orthologs. Then we split the alignment into two alignments with the same length.
- (2) We label a position as consensus when both alignments are conserved more than the consensus threshold (90%) at that particular position and the most frequent amino acids are similar (BLOSUM80 score is more than 1) to each other.
- (3) We calculated conservation percentages for each alignment. There are two different scenarios in this case. The first one is when the most frequent amino acids of the two of the alignments are not similar (BLOSUM80 score is lower than 2) to each other. If this is the case and the conservation percentage for any alignment is above the specificity threshold (90%) we label that position as specifically conserved for that alignment. The second case is where the most frequently observed amino acids are similar to each other. In this case, for a position to be specific for one alignment first it should satisfy the specificity

threshold and secondly, the conservation percentage of the other alignment should be lower than our lower threshold (70%).

For the steps above, we choose 90% for both specificity and consensus thresholds. 70% is selected for the lower specificity threshold.

### Enrichment of specifically conserved residues

We identified specifically conserved residues with two different approaches:

#### Specific approach

- (1) We divided receptors into two couplers versus non-couplers. Let us assume that we have  $n$  number of couplers and  $m$  number of non-couplers.
- (2) We compare coupler receptors with non-couplers in a pairwise manner. In these comparisons, we count the number of being specific for every residue. In total there are  $n$  times  $m$  comparisons. We divide the obtained counts by the total number of comparisons to get the frequency of a residue being specific for the couplers' group.
- (3) To examine if a residue is generally variable or specific to the coupling event, we compared couplers with themselves. We applied STEP 2 for couplers–couplers comparison as well. This time, we have  $n \times (n - 1)$  comparisons in total. We again calculated the frequencies accordingly.
- (4) For the specific approach, we do not allow any inside variation and this makes the result of STEP 3 zero. On the other hand, for a residue to be labeled as specific, we expect STEP 2 more than zero. When these two conditions are satisfied, we label that residue as specifically conserved

#### Sensitive approach

- (1) We built a comprehensive MSA for the coupler receptors and their orthologs.
- (2) We compared this alignment with non-coupler receptor's MSAs similarly to STEP 2 of the Specific Approach.
- (3) We added newly discovered positions to our analysis as specifically conserved.

### Building the maximum-likelihood phylogenetic tree for aminergic receptors

- (1) We blasted (Camacho et al, 2009) aminergic receptors and obtained the first 50 sequences to generate a fasta file.
- (2) From that fasta file we selected representative sequences by using cd-hit default options.
- (3) MAFFT (Katoh & Standley, 2013) eini algorithm was used to align representative sequences.
- (4) IQ-TREE version 2.0.5 (Minh et al, 2020) was used to create the phylogenetic tree with options: -m JTT+G+I+F -b 100 -tbe

### RRCS and network analysis

We calculated the RRCS score for 20 active (ADRB2: 3SN6, 7DHI; DRD1: 7CKW, 7CKX, 7CKZ, 7CKY, 7CRH, 7JV5, 7JVP, 7JVQ, 7LJC, 7LJD; DRD2: 6VM5,

7JVR; DRD3: 7CMU, 7CMV; 5HT1B: 6G79; ACM2: 6OIK; 5HT2A: 6WHA; HRH1: 7DFL) (Rasmussen et al, 2011; García-Nafria et al, 2018; Maeda et al, 2019; Kim et al, 2020; Xia et al, 2021; Xiao et al, 2021; Xu et al, 2021; Yang et al, 2020; Yin et al, 2020; Zhuang et al, 2021a, 2021b) 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 et al, 2007; Chien et al, 2010; Shimamura et al, 2011; Haga et al, 2012; Wang et al, 2013, 2018b; Huang et al, 2016; Suno et al, 2018; Yin et al, 2018; Ishchenko et al, 2019; Kimura et al, 2019; Fan et al, 2020; Im et al, 2020; Kim et al, 2020; Miyagi et al, 2020). For each receptor, we subtracted inactive RRCS from active RRCS to obtain  $\Delta$ RRCS values for each residue pairs. We wrote a custom python code to obtain files with  $\Delta$ RRCS scores. We combined all of the networks that contain information about the contact changes upon activation to produce the most common molecular signal transduction pathways (Fig S1). For the details of the RRCS algorithm, please read the corresponding article (Zhou et al, 2019).

### Identification of G protein-specific activation networks

After obtaining  $\Delta$ RRCS networks for each active-inactive structure pair we grouped  $\Delta$ RRCS values based on the G protein subtype coupling the receptors. Then we compared  $\Delta$ RRCS values of individual groups (e.g.,  $G_s$ : ADRB2 and DRD1) with the rest of the groups (e.g., Non- $G_s$ : DRD2, DRD3, 5HT1B, ACM2, 5HT2A, HRH1) by using two-sample  $t$  test. Whereas  $P \leq 0.01$  is used for  $G_s$ ,  $G_q$ , and  $G_o$ ,  $P \leq 0.1$  is used for  $G_{i1}$ . We obtained significant contact changes upon coupling to a particular G protein.

### Molecular dynamics simulations

We downloaded inactive and active structures of Beta2 Adrenergic receptor ( $\beta_2$ AR) from PDB (PDB ID: 4GBR, and 3SN6, respectively) (Rasmussen et al, 2011; Zou et al, 2012). Three thermostabilizing mutations, T96M<sup>2x66</sup>, T98M<sup>23x49</sup>, and E187N<sup>ECL2</sup>, were mutated back to the WT in both sequences. Because the used inactive structure of the  $\beta_2$ AR has a short ICL3 that links the TM5 and TM6, we did not introduce additional residues to the ICL3 and used the crystal structure as it is. However, the active structure of the  $\beta_2$ AR lacks ICL3, and we modeled a short loop with GalaxyLoop code (Park et al, 2014). We inserted FHVSKF between ARG239 and CYS265. We introduced three changes at the 315<sup>7x41</sup> position, and one WT and obtained three mutants (namely; G315C, G315L, and G315Q). We used PyMOL to place mutations (PyMOL Molecular Graphics System, Version 2.1.0.). Orientations of proteins in biological membranes were calculated with OPM server (Lomize et al, 2012) and We used CHARMM-GUI web server (Jo et al, 2008; Wu et al, 2014; Lee et al, 2016) to create input files for the MD simulations for Gromacs. Because inactive and active structures start with ASP29<sup>1x28</sup> and GLU30<sup>1x30</sup> and end with LEU342<sup>Cterm</sup> and CYS341<sup>8x59</sup>, respectively, we introduced acetylated N-terminus and methylamidated C terminus to the N- and C-terminal ends. Two disulfide bridges between CYS106<sup>3x25</sup>-CYS191<sup>ECL2</sup> and CYS184<sup>ECL2</sup>-CYS190<sup>ECL2</sup> were introduced. Each lipid leaflet contains 92 (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) POPC biological lipid type (total 192 POPC

molecules in system). Systems were neutralized with 0.15 M NaCl ions (50 Na<sup>+</sup> and 55 Cl<sup>-</sup> ions in total). We used TIP3P water model for the water molecules (Mackerell et al, 1998) and CHARMM36m force field for the protein, lipids, and ions (Huang et al, 2017). One minimization and six equilibration steps were applied to the systems, before production runs (for the equilibration phases 5, 5, 10, 10, 10, and 10 ns MD simulations were run, in total 50 ns). In equilibration phases, both Berendsen thermostat and barostat were used (Berendsen et al, 1984). In production runs, we applied Noose-Hoover thermostat (Nosé & Klein, 1983; Hoover, 1986) and Parrinello-Rahman barostat (Parrinello & Rahman, 1980). 500 ns production simulations were run with Gromacs v2020 (Abraham et al, 2015) and repeated seven times to increase sampling (in total for each system we simulated 3.5  $\mu$ s). 5,000 frames were collected for each run, and for instance for the WT system, we concatenated 35,000 frames to calculate GPCRdb distance distributions (*gmx distance* tool was used for this purpose) and find average structures (Visual Molecular Dynamics code used to find average structure [Humphrey et al, 1996]). To calculate the GPCRdb distance in Class-A GPCR structures, CYS125<sup>3x44</sup>-ILE325<sup>7x52</sup> distance was subtracted from TYR70<sup>2x41</sup>-GLY276<sup>6x38</sup> distance. If the calculated distance is higher than 7.15 Å, lower than 2 Å, and between 2 and 7.15 Å state of the receptors labeled as active, inactive, and intermediate, respectively (Isberg et al, 2015; Shahraki et al, 2021). All figures were generated with PyMOL v2.1.0. To estimate water accessibilities to the internal cavity of the receptors, and sodium ion accessibilities to the ASP79<sup>2x50</sup>, we calculated averaged water and sodium ion densities. Time-averaged three-dimensional water and sodium ion density maps were calculated with GROmaps (Briones et al, 2019).

### Analysis of contact changes within molecular dynamics simulation trajectories

Frames of MD simulation trajectories were selected from 0 to 500 ns with 50 ns gaps for each trajectory and replicate for a mutation. Including the frame at  $t = 0$  ns, for a replicate, we obtained 11 frames to represent the whole trajectory. We have applied the same strategy for all seven active-state replicates and obtained 77 frames for WT and mutated MD trajectories. For each frame, we calculated RRCSs for every residue pair and identified statistically significant ( $P < 0.05$ ) differences between WT and mutated trajectories by applying a two-sided  $t$  test. For the inactive simulations, we had only two replicates; therefore, we compared 22 mutated frames with 22 WT frames.

After applying  $t$  test, we intersected the significant contact changes we observed for each mutational state to observe the common change due to the absence of glycine. In total, we identified 135 common changes for active-state simulations and 83 common changes for inactive-state simulations. We used Cytoscape (Shannon et al, 2003) to visualize the changes as a network. PyMOL was used to visualize the identified pathway on protein structures.

### Data Availability

The open-source code and supplementary data are available at our GitHub repository: <https://github.com/CompGenomeLab/GPCR-coupling-selectivity>.

The MD trajectories have been deposited to: <https://doi.org/10.5281/zenodo.5763490>.

## Supplementary Information

Supplementary Information is available at <https://doi.org/10.26508/lsa.202201439>.

## Acknowledgements

This work is supported by EMBO Installation Grant (to O Adebali 4163) that is funded by TÜBİTAK. We acknowledge partial support from Turkish Academy of Sciences (TÜBA-GEBİP) and Science Academy, Turkey (BAGEP). The molecular dynamics simulations reported in this paper were fully performed at TÜBİTAK ULAKBİM, High Performance and Grid Computing Center (TRUBA resources).

### Author Contributions

B Selçuk: conceptualization, software, formal analysis, investigation, visualization, methodology, and writing—original draft.  
I Erol: investigation, methodology, and writing—review and editing.  
S Durdağ: supervision and writing—review and editing.  
O Adebali: conceptualization, resources, supervision, funding acquisition, project administration, and writing—review and editing.

### Conflict of Interest Statement

The authors declare that they have no conflict of interest.

## References

- Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2: 19–25. doi:10.1016/j.softx.2015.06.001
- Adebali O, Reznik AO, Ory DS, Zhulin IB (2016) Establishing the precise evolutionary history of a gene improves prediction of disease-causing missense mutations. *Genet Med* 18: 1029–1036. doi:10.1038/gim.2015.208
- Arakawa M, Chakraborty R, Upadhyaya J, Eilers M, Reeves PJ, Smith SO, Chelikani P (2011) Structural and functional roles of small group-conserved amino acids present on helix-H7 in the  $\beta(2)$ -adrenergic receptor. *Biochim Biophys Acta* 1808: 1170–1178. doi:10.1016/j.bbame.2011.01.012
- Avet C, Mancini A, Breton B, Le Gouill C, Hauser AS, Normand C, Kobayashi H, Gross F, Hogue M, Lukashcheva V, et al (2022) Effector membrane translocation biosensors reveal G protein and  $\beta$ arrestin coupling profiles of 100 therapeutically relevant GPCRs. *Elife* 11: e74101. doi:10.7554/eLife.74101
- Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81: 3684–3690. doi:10.1063/1.448118
- Briones R, Blau C, Kutzner C, de Groot BL, Aponte-Santamaría C (2019) GROMaps: A GROMACS-based toolset to analyze density maps derived from molecular dynamics simulations. *Biophys J* 116: 4–11. doi:10.1016/j.bpj.2018.11.3126
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10: 421. doi:10.1186/1471-2105-10-421
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973. doi:10.1093/bioinformatics/btp348
- Catapano LA, Manji HK (2007) G protein-coupled receptors in major psychiatric disorders. *Biochim Biophys Acta* 1768: 976–993. doi:10.1016/j.bbame.2006.09.025
- Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, et al (2007) High-resolution crystal structure of an engineered human  $\beta(2)$ -adrenergic G protein-coupled receptor. *Science* 318: 1258–1265. doi:10.1126/science.1150577
- Chien EYT, Liu W, Zhao Q, Katritch V, Won Han G, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V, et al (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* 330: 1091–1095. doi:10.1126/science.1197410
- Chung KY, Rasmussen SGF, Liu T, Li S, Devree BT, Chae PS, Calinski D, Kobilka BK, Woods VL, Sunahara RK (2011) Conformational changes in the G protein Gs induced by the  $\beta(2)$  adrenergic receptor. *Nature* 477: 611–615. doi:10.1038/nature10488
- Du Y, Duc NM, Rasmussen SGF, Hilger D, Kubiak X, Wang L, Bohon J, Kim HR, Wegrecki M, Asuru A, et al (2019) Assembly of a GPCR-G protein complex. *Cell* 177: 1232–1242.e11. doi:10.1016/j.cell.2019.04.022
- Fan L, Tan L, Chen Z, Qi J, Nie F, Luo Z, Cheng J, Wang S (2020) Haloperidol bound D2 dopamine receptor structure inspired the discovery of subtype selective ligands. *Nat Commun* 11: 1074. doi:10.1038/s41467-020-14884-y
- Flock T, Hauser AS, Lund N, Gloriam DE, Balaji S, Babu MM (2017) Selectivity determinants of GPCR-G-protein binding. *Nature* 545: 317–322. doi:10.1038/nature22070
- Fredriksson R, Lagerström MC, Lundin L-G, Schiöth HB (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* 63: 1256–1272. doi:10.1124/mol.63.6.1256
- García-Nafria J, Nehmé R, Edwards PC, Tate CG (2018) Cryo-EM structure of the serotonin 5-HT1B receptor coupled to heterotrimeric Go. *Nature* 558: 620–623. doi:10.1038/s41586-018-0241-9
- Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T, et al (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* 482: 547–551. doi:10.1038/nature10753
- Hoover WG (1986) Constant-pressure equations of motion. *Phys Rev A Gen Phys* 34: 2499–2500. doi:10.1103/PhysRevA.34.2499
- Huang C-Y, Olieric V, Ma P, Howe N, Vogeley L, Liu X, Warshamanage R, Weinert T, Panepucci E, Kobilka B, et al (2016) In meso in situ serial X-ray crystallography of soluble and membrane proteins at cryogenic temperatures. *Acta Crystallogr D Struct Biol* 72: 93–112. doi:10.1107/s2059798315021683
- Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, Grubmüller H, Mackerell AD (2017) CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat Methods* 14: 71–73. doi:10.1038/nmeth.4067
- Huerta-Cepas J, Serra F, Bork P (2016) ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol* 33: 1635–1638. doi:10.1093/molbev/msw046
- Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14: 33–38. doi:10.1016/0263-7855(96)00018-5
- Im D, Inoue A, Fujiwara T, Nakane T, Yamanaka Y, Uemura T, Mori C, Shiimura Y, Kimura KT, Asada H, et al (2020) Structure of the dopamine D2

- receptor in complex with the antipsychotic drug spiperone. *Nat Commun* 11: 6442. doi:[10.1038/s41467-020-20221-0](https://doi.org/10.1038/s41467-020-20221-0)
- Inoue A, Raimondi F, Kadji FMN, Singh G, Kishi T, Uwamizu A, Ono Y, Shinjo Y, Ishida S, Arang N, et al (2019) Illuminating G-protein-coupling selectivity of GPCRs. *Cell* 177: 1933–1947.e25. doi:[10.1016/j.cell.2019.04.044](https://doi.org/10.1016/j.cell.2019.04.044)
- Isberg V, De Graaf C, Bortolato A, Cherezov V, Katritch V, Marshall FH, Mordalski S, Pin J-P, Stevens RC, Vriend G, et al (2015) Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends Pharmacol Sci* 36: 22–31. doi:[10.1016/j.tips.2014.11.001](https://doi.org/10.1016/j.tips.2014.11.001)
- Ishchenko A, Stauch B, Han GW, Batyuk A, Shiriaeva A, Li C, Zatselin N, Weierstall U, Liu W, Nango E, et al (2019) Toward G protein-coupled receptor structure-based drug design using X-ray lasers. *IUCr* 6: 1106–1119. doi:[10.1107/s2052252519013137](https://doi.org/10.1107/s2052252519013137)
- Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem* 29: 1859–1865. doi:[10.1002/jcc.20945](https://doi.org/10.1002/jcc.20945)
- Jones EM, Lubock NB, Venkatakrisnan AJ, Wang J, Tseng AM, Paggi JM, Latorraca NR, Cancilla D, Satyadi M, Davis JE, et al (2020) Structural and functional characterization of G protein-coupled receptors with deep mutational scanning. *eLife* 9: e54895. doi:[10.7554/eLife.54895](https://doi.org/10.7554/eLife.54895)
- Kang Y, Kuybeda O, de Waal PW, Mukherjee S, Van Eps N, Dutka P, Zhou XE, Bartsaghi A, Erramilli S, Morizumi T, et al (2018) Cryo-EM structure of human rhodopsin bound to an inhibitory G protein. *Nature* 558: 553–558. doi:[10.1038/s41586-018-0215-y](https://doi.org/10.1038/s41586-018-0215-y)
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30: 772–780. doi:[10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010)
- Katritch V, Fenalti G, Abola EE, Roth BL, Cherezov V, Stevens RC (2014) Allosteric sodium in class A GPCR signaling. *Trends Biochem Sci* 39: 233–244. doi:[10.1016/j.tibs.2014.03.002](https://doi.org/10.1016/j.tibs.2014.03.002)
- Kim K, Che T, Panova O, DiBerto JF, Lyu J, Krumm BE, Wacker D, Robertson MJ, Seven AB, Nichols DE, et al (2020) Structure of a hallucinogen-activated Gq-coupled 5-HT<sub>2A</sub> serotonin receptor. *Cell* 182: 1574–1588.e19. doi:[10.1016/j.cell.2020.08.024](https://doi.org/10.1016/j.cell.2020.08.024)
- Kimura KT, Asada H, Inoue A, Kadji FMN, Im D, Mori C, Arakawa T, Hirata K, Nomura Y, Nomura N, et al (2019) Structures of the 5-HT<sub>2A</sub> receptor in complex with the antipsychotics risperidone and zotepine. *Nat Struct Mol Biol* 26: 121–128. doi:[10.1038/s41594-018-0180-z](https://doi.org/10.1038/s41594-018-0180-z)
- Kooistra AJ, Mordalski S, Pándy-Szekeres G, Esguerra M, Mamyrbekov A, Munk C, Keserű GM, Gloriam DE (2021) GPCRdb in 2021: Integrating GPCR sequence, structure and function. *Nucleic Acids Res* 49: D335–D343. doi:[10.1093/nar/gkaa1080](https://doi.org/10.1093/nar/gkaa1080)
- Kozlov AM, 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: 4453–4455. doi:[10.1093/bioinformatics/btz305](https://doi.org/10.1093/bioinformatics/btz305)
- Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, Wei S, Buckner J, Jeong JC, Qi Y, et al (2016) CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J Chem Theory Comput* 12: 405–413. doi:[10.1021/acs.jctc.5b00935](https://doi.org/10.1021/acs.jctc.5b00935)
- Liu X, Xu X, Hilger D, Aschauer P, Tiemann JKS, Du Y, Liu H, Hirata K, Sun X, Guixà-González R, et al (2019) Structural insights into the process of GPCR-G protein complex formation. *Cell* 177: 1243–1251.e12. doi:[10.1016/j.cell.2019.04.021](https://doi.org/10.1016/j.cell.2019.04.021)
- Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL (2012) OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res* 40: D370–D376. doi:[10.1093/nar/gkr703](https://doi.org/10.1093/nar/gkr703)
- MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102: 3586–3616. doi:[10.1021/jp973084f](https://doi.org/10.1021/jp973084f)
- Maeda S, Qu Q, Robertson MJ, Skiniotis G, Kobilka BK (2019) Structures of the M1 and M2 muscarinic acetylcholine receptor/G-protein complexes. *Science* 364: 552–557. doi:[10.1126/science.aaw5188](https://doi.org/10.1126/science.aaw5188)
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37: 1530–1534. doi:[10.1093/molbev/msaa015](https://doi.org/10.1093/molbev/msaa015)
- 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. *Sci Rep* 10: 11669. doi:[10.1038/s41598-020-68355-x](https://doi.org/10.1038/s41598-020-68355-x)
- Nosé S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. *Mol Phys* 50: 1055–1076. doi:[10.1080/00268978300102851](https://doi.org/10.1080/00268978300102851)
- Okashah N, Wan Q, Ghosh S, Sandhu M, Inoue A, Vaidehi N, Lambert NA (2019) Variable G protein determinants of GPCR coupling selectivity. *Proc Natl Acad Sci USA* 116: 12054–12059. doi:[10.1073/pnas.1905993116](https://doi.org/10.1073/pnas.1905993116)
- Pándy-Szekeres G, Munk C, Tsonkov TM, Mordalski S, Harpsøe K, Hauser AS, Bojarski AJ, Gloriam DE (2018) GPCRdb in 2018: Adding GPCR structure models and ligands. *Nucleic Acids Res* 46: D440–D446. doi:[10.1093/nar/gkx1109](https://doi.org/10.1093/nar/gkx1109)
- Park H, Lee GR, 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: e113811. doi:[10.1371/journal.pone.0113811](https://doi.org/10.1371/journal.pone.0113811)
- Parrinello M, Rahman A (1980) Crystal structure and pair potentials: A molecular-dynamics study. *Phys Rev Lett* 45: 1196–1199. doi:[10.1103/PhysRevLett.45.1196](https://doi.org/10.1103/PhysRevLett.45.1196)
- Rasmussen SGF, Devree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, et al (2011) Crystal structure of the  $\beta_2$  adrenergic receptor-Gs protein complex. *Nature* 477: 549–555. doi:[10.1038/nature10361](https://doi.org/10.1038/nature10361)
- Rose AS, Elgeti M, Zachariae U, Grubmüller H, Hofmann KP, Scheerer P, Hildebrand PW (2014) Position of transmembrane helix 6 determines receptor G protein coupling specificity. *J Am Chem Soc* 136: 11244–11247. doi:[10.1021/ja5055109](https://doi.org/10.1021/ja5055109)
- Rosenbaum DM, Rasmussen SGF, Kobilka BK (2009) The structure and function of G-protein-coupled receptors. *Nature* 459: 356–363. doi:[10.1038/nature08144](https://doi.org/10.1038/nature08144)
- Semack A, Sandhu M, Malik RU, Vaidehi N, Sivaramakrishnan S (2016) Structural elements in the Gas and Gaq C termini that mediate selective G protein-coupled receptor (GPCR) signaling. *J Biol Chem* 291: 17929–17940. doi:[10.1074/jbc.m116.735720](https://doi.org/10.1074/jbc.m116.735720)
- Senese NB, Rasenick MM, Traynor JR (2018) The role of G-proteins and G-protein regulating proteins in depressive disorders. *Front Pharmacol* 9: 1289. doi:[10.3389/fphar.2018.01289](https://doi.org/10.3389/fphar.2018.01289)
- Seo MJ, Heo J, Kim K, Chung KY, Yu W (2021) Coevolution underlies GPCR-G protein selectivity and functionality. *Sci Rep* 11: 7858. doi:[10.1038/s41598-021-87251-6](https://doi.org/10.1038/s41598-021-87251-6)
- Shahraki A, Işbilir A, Dogan B, Lohse MJ, Durdagi S, Birgul-Iyison N (2021) Structural and functional characterization of allatostatin receptor type-C of *thaumetopoea ptyocampa*, a potential target for next-generation pest control agents. *J Chem Inf Model* 61: 715–728. doi:[10.1021/acs.jcim.0c00985](https://doi.org/10.1021/acs.jcim.0c00985)
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504. doi:[10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303)
- Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, et al (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475: 65–70. doi:[10.1038/nature10236](https://doi.org/10.1038/nature10236)
- Suno R, Lee S, Maeda S, Yasuda S, Yamashita K, Hirata K, Horita S, Tawaramoto MS, Tsujimoto H, Murata T, et al (2018) Structural insights into the



- subtype-selective antagonist binding to the M2 muscarinic receptor. *Nat Chem Biol* 14: 1150–1158. doi:10.1038/s41589-018-0152-y
- UniProt Consortium (2019) UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res* 47: D506–D515. doi:10.1093/nar/gky1049
- Van Eps N, Altenbach C, Caro LN, Latorraca NR, Hollingsworth SA, Dror RO, Ernst OP, Hubbell WL (2018) Gi- and Gs-coupled GPCRs show different modes of G-protein binding. *Proc Natl Acad Sci USA* 115: 2383–2388. doi:10.1073/pnas.1721896115
- Vass M, Podlewska S, De Esch IJP, Bojarski AJ, Leurs R, Kooistra AJ, De Graaf C (2019) Aminergic GPCR–ligand interactions: A chemical and structural map of receptor mutation data. *J Med Chem* 62: 3784–3839. doi:10.1021/acs.jmedchem.8b00836
- Wang C, Jiang Y, Ma J, Wu H, Wacker D, Katritch V, Han GW, Liu W, Huang XP, Vardy E, et al (2013) Structural basis for molecular recognition at serotonin receptors. *Science* 340: 610–614. doi:10.1126/science.1232807
- Wang J, Miao Y (2019) Mechanistic insights into specific G protein interactions with adenosine receptors. *J Phys Chem B* 123: 6462–6473. doi:10.1021/acs.jpcc.9b04867
- Wang J, Gareri C, Rockman HA (2018a) G-Protein-Coupled receptors in heart disease. *Circ Res* 123: 716–735. doi:10.1161/circresaha.118.311403
- Wang S, Che T, Levit A, Shoichet BK, Wacker D, Roth BL (2018b) Structure of the D2 dopamine receptor bound to the atypical antipsychotic drug risperidone. *Nature* 555: 269–273. doi:10.1038/nature25758
- Wess J (2021) Designer GPCRs as novel tools to identify metabolically important signaling pathways. *Front Endocrinol (Lausanne)* 12: 706957. doi:10.3389/fendo.2021.706957
- Wess J (1998) Molecular basis of receptor/G-protein-coupling selectivity. *Pharmacol Ther* 80: 231–264. doi:10.1016/s0163-7258(98)00030-8
- Wetttschureck N, Offermanns S (2005) Mammalian G proteins and their cell type specific functions. *Physiol Rev* 85: 1159–1204. doi:10.1152/physrev.00003.2005
- Wong SK-F (2003) G protein selectivity is regulated by multiple intracellular regions of GPCRs. *Neurosignals* 12: 1–12. doi:10.1159/000068914
- Wu EL, Cheng X, Jo S, Rui H, Song KC, Dávila-Contreras EM, Qi Y, Lee J, Monje-Galvan V, Venable RM, et al (2014) GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* 35: 1997–2004. doi:10.1002/jcc.23702
- 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. *Nat Commun* 12: 2086. doi:10.1038/s41467-021-22427-2
- Xiao P, Yan W, Gou L, Zhong YN, Kong L, Wu C, Wen X, Yuan Y, Cao S, Qu C, et al (2021) Ligand recognition and allosteric regulation of DRD1-Gs signaling complexes. *Cell* 184: 943–956.e18. doi:10.1016/j.cell.2021.01.028
- Xu P, Huang S, Mao C, Krumm BE, Zhou XE, Tan Y, Huang X-P, Liu Y, Shen D-D, Jiang Y, et al (2021) Structures of the human dopamine D3 receptor-Gi complexes. *Mol Cell* 81: 1147–1159.e4. doi:10.1016/j.molcel.2021.01.003
- Yang F, Ling S, Zhou Y, Zhang Y, Lv P, Liu S, Fang W, Sun W, Hu LA, Zhang L, et al (2020) Different conformational responses of the  $\beta$ 2-adrenergic receptor-Gs complex upon binding of the partial agonist salbutamol or the full agonist isoprenaline. *Natl Sci Rev* 8: nwaa284. doi:10.1093/nsr/nwaa284
- Yin J, Chen K-YM, Clark MJ, Hijazi M, Kumari P, Bai X-C, Sunahara RK, Barth P, Rosenbaum DM (2020) Structure of a D2 dopamine receptor-G-protein complex in a lipid membrane. *Nature* 584: 125–129. doi:10.1038/s41586-020-2379-5
- Yin W, Zhou XE, Yang D, de Waal PW, Wang M, Dai A, Cai X, Huang C-Y, Liu P, Wang X, et al (2018) Crystal structure of the human 5-HT<sub>1B</sub> serotonin receptor bound to an inverse agonist. *Cell Discov* 4: 12. doi:10.1038/s41421-018-0009-2
- Zhou Q, Yang D, Wu M, Guo Y, Guo W, Zhong L, Cai X, Dai A, Jang W, Shakhnovich EI, et al (2019) Common activation mechanism of class A GPCRs. *eLife* 8: e50279. doi:10.7554/eLife.50279
- Zhuang Y, Krumm B, Zhang H, Zhou XE, Wang Y, Huang X-P, Liu Y, Cheng X, Jiang Y, Jiang H, et al (2021a) Mechanism of dopamine binding and allosteric modulation of the human D1 dopamine receptor. *Cell Res* 31: 593–596. doi:10.1038/s41422-021-00482-0
- Zhuang Y, Xu P, Mao C, Wang L, Krumm B, Zhou XE, Huang S, Liu H, Cheng X, Huang XP, et al (2021b) Structural insights into the human D1 and D2 dopamine receptor signaling complexes. *Cell* 184: 931–942.e18. doi:10.1016/j.cell.2021.01.027
- Zou Y, Weis WI, Kobilka BK (2012) N-terminal T4 lysozyme fusion facilitates crystallization of a G protein coupled receptor. *PLoS One* 7: e46039. doi:10.1371/journal.pone.0046039



**License:** This article is available under a Creative Commons License (Attribution 4.0 International, as described at <https://creativecommons.org/licenses/by/4.0/>).