Unique Features of Different Classes of G-Protein-Coupled Receptors Revealed from Sequence Coevolutionary and Structural Analysis

Hung N Do[1], Allan Haldane[2,*], Ronald M Levy[2], Yinglong Miao[1,*]

[1]The Center for Computational Biology and Department of Molecular Biosciences, The University of Kansas, Lawrence, Kansas 66047
[2]Department of Chemistry, Center for Biophysics and Computational Biology, Institute for Computational Molecular Science, Temple University, Philadelphia, Pennsylvania 19122

*Corresponding authors: allan.haldane@temple.edu and miao@ku.edu

Abstract

G-protein-coupled receptors (GPCRs) are the largest family of human membrane proteins and serve as the primary targets of about one third of currently marketed drugs. Despite the critical importance, experimental structures have been determined for only a limited portion of GPCRs. Functional mechanisms of GPCRs remain poorly understood. Here, we have constructed sequence coevolutionary models of the A, B and C classes of GPCRs and compared them with residue contact frequency maps generated with available experimental structures. Significant portions of structural residue contacts have been successfully detected in the sequence-based covariational models. "Exception" residue contacts predicted from sequence coevolutionary models but not available structures added missing links that were important for GPCR activation and allosteric modulation. Our combined coevolutionary and structural analysis revealed unique features of the different classes of GPCRs. First, we provided evidence from coevolutionary couplings that dimerization is required for activation of class C GPCRs, but not for activation of class A and B GPCRs. Second, we identified distinct residue contacts involving different sets of functional motifs for activation of the class A and B GPCRs. Finally, we uncovered critical residue contacts

tuned by allosteric modulation in the three classes of GPCRs. These findings provide a promising framework for designing selective therapeutics of GPCRs.

Keywords: G-protein-coupled receptors, Sequence Coevolution, Structural Contacts, Activation, Oligomerization, Allosteric modulation.

Introduction

G-protein-coupled receptors (GPCRs) comprise the largest and most diverse family of integral membrane proteins in eukaryotes. At least four percent of the entire human genome is coded for GPCRs (Bjarnadóttir, 2006). GPCRs mediate various physiological activities, including vision, olfaction, taste, neurotransmission, endocrine, and immune responses (A. J. Venkatakrishnan et al., 2013). Due to the critical roles in cellular signaling, approximately 34% of FDA-approved therapeutic agents act on GPCRs (Hauser, 2018). On the basis of sequence homology and functional similarity, GPCRs are classified into six different classes, four of which are present in the human body: class A (Rhodopsin-like), class B (secretin receptors), which is further divided into subclasses of B1 (classical hormone receptors), B2 (adhesion GPCRs) and B3 (methuselah-type receptors), class C (metabotropic glutamate receptors) and class F (frizzled/TAS2 receptors) (V. Isberg et al., 2016; Stevens et al., 2013). The other two classes include D (fungal mating pheromone receptors) and E (cyclic AMP receptors). In comparison, class A is the largest with 701 known receptors and by far the most extensively studied class of GPCRs (V. Isberg et al., 2016). GPCRs share a characteristic structural fold of seven transmembrane (TM) α-helices (TM1-TM7) connected by three extracellular loops (ECL1-ECL3) and three intracellular loops (ICL1-ICL3). The extracellular and intracellular domains are typically

important for binding of the ligands and G proteins, respectively. Consequently, the loop regions are notably diverse in sequences and structures.

Activation varies among the A, B and C classes of GPCRs. Class A GPCR activation is triggered by binding of an agonist to the receptor orthosteric pocket located within the 7TM domain (Peeters et al., 2014; Peeters, Li, van Westen, & IJzerman, 2012; Ye, Van Eps, Zimmer, Ernst, & Prosser, 2016). Upon agonist binding, the receptor intracellular end of TM6 moves outwards to open up an intracellular cavity to accommodate and activate the G protein (Draper-Joyce et al., 2018; Dror et al., 2011; Garcia-Nafria, Lee, Bai, Carpenter, & Tate, 2018; Hilger et al., 2020; Maeda, Qu, Robertson, Skiniotis, & Kobilka, 2019; Manglik & Kruse, 2017; Miao, Nichols, Gasper, Metzger, & McCammon, 2013; R. Nygaard et al., 2013; S. G. F. Rasmussen et al., 2011; Vilardaga, Bünemann, Krasel, Castro, & Lohse, 2003; Ye et al., 2016). Activation of class B GPCRs requires binding of both the agonist and G protein, as well as disruption of the TM6 helix with a sharp kink (Hilger et al., 2020; A. Jazayeri et al., 2017; Y.-L. Liang et al., 2018; Y. L. Liang et al., 2020; Seidal, Zarzycka, Zaidi, Katritch, & Coin, 2017; Zhang et al., 2017). The orthosteric pockets of class C GPCRs are located in the conserved Venus flytrap (VFT) region in large extracellular domains (ECD) (Chun, 2012; Koehl, 2019). Activation of class C GPCRs require dimerization of two receptor subunits (Grushevskyi et al., 2019; Koehl, 2019). The dimer interface of class C GPCRs compresses, bringing the cysteine rich domains (CRDs) in the ECDs closer, and interactions between CRDs and ECL2 reposition the 7TM domains of two receptors for signal transduction (Koehl, 2019).

Bioinformatics analysis has been previously carried out to identify important residue interactions for GPCR activation. Cvicek et al. generated a structure-based alignment of 25 GPCRs that was extended to include transmembrane sequences of all human GPCRs (Cvicek, 2016). The

final sequence-structure alignment revealed 40 interhelical contacts that were common to class A GPCRs, 23 of which were conserved among class B, C, and F GPCRs. Furthermore, by comparing the active and inactive structures of class A receptors, they identified 15 Native ACtivation "Hot-spOt" residues (NACHOs) for class A GPCR activation (Cvicek, 2016). In 2019, Zhou et al. discovered a common activation pathway of class A GPCRs through an analysis residue-residue contact scores of 235 available class A GPCR structures (Zhou, 2019). A four-layer activation pathway that connected the extracellular to intracellular regions was characterized at the residue level. Changes in critical residue contacts were identified during global movements of TM6 and TM7 in class A GPCR activation (Zhou, 2019).

GPCRs are also able to bind allosteric ligands at topographically distinct sites, which could induce further conformational changes of the GPCRs (Christopoulos, 2002; Goblyos & Ijzerman, 2011; Jeffrey Conn, Christopoulos, & Lindsley, 2009; Keov, 2011; May, Leach, Sexton, & Christopoulos, 2007; Nawaratne, Leach, Felder, Sexton, & Christopoulos, 2010). Allosteric ligands often include the positive allosteric modulator (PAM) and negative allosteric modulator (NAM) of GPCR activation. For class A GPCRs, binding of a PAM in the M2 muscarinic receptor was shown to induce slight contraction of the receptor extracellular pocket, which was pre-formed in the active agonist-bound structure (Kruse, 2013; Maeda et al., 2019). Binding of a muscarinic toxin NAM to the inactive antagonist-bound M1 muscarinic receptor induced conformational changes in the receptor ECL2, TM1, TM2, TM6 and TM7 extracellular domains, as well as the TM2 and TM6 intracellular domains (Maeda et al., 2020). In the free fatty acid receptor GPR40 and the C5a receptor, PAM binding in a lipid-facing pocket formed by TM3-TM4-ICL2 induced conformational changes in the ICL2, TM4 and TM5 of the active receptor (Liu et al., 2018; Lu,

2017). The ICL2 adopted a short helical conformation and the TM5 was shifted along its helical axis towards the extracellular side relative to the TM4 (Lu, 2017).

For class B GPCRs, a PAM was found to bind between the extracellular domains of TM1 and TM2 of the GLP-1 receptor (Bueno et al., 2020). In the glucagon receptor, NAM binding restricted the outward movement of the TM6 intracellular domain (A. Jazayeri, Doré, A.S., Lamb, D., Krishnamurthy, H., Southall, S.M., Baig, A.H., Bortolato, A., Koglin, M., Robertson, N.J., Errey, J.C., Andrews, S.P., Teobald, I., Brown, A.J.H., Cooke, R.M., Weir, M., Marshall, F.H., 2016). The ECL2 stretched to the central axis of the TM helical bundle, allowing for interactions from TM3 to TM6 and TM7 in the inactive class B GPCRs (A. Jazayeri, Doré, A.S., Lamb, D., Krishnamurthy, H., Southall, S.M., Baig, A.H., Bortolato, A., Koglin, M., Robertson, N.J., Errey, J.C., Andrews, S.P., Teobald, I., Brown, A.J.H., Cooke, R.M., Weir, M., Marshall, F.H., 2016). For class C GPCRs, TM3 is packed against TM2, TM4, TM5, TM6 and ICL3 within inactive monomers of the metabotropic $GABA_B$ receptor (Kim, 2020; Papasergi-Scott et al., 2020; Shaye, 2020). Binding of the agonist and PAM led to significant conformational changes in the ICL3, TM3, TM4, TM5 and TM6 domains, with the intracellular half of TM3 bent substantially away from TM5 (Kim, 2020; Shaye, 2020).

Despite remarkable advances in structural determination efforts, experimental structures have been resolved for only ~90 unique GPCRs (V. Isberg et al., 2016; Stevens et al., 2013). Functional mechanisms of many GPCR classes related to activation and allosteric modulation remain poorly understood at the residue level. Recent developments in methods for residue-covariation analysis have shown that another source of functional and structural information is in observed patterns of mutational covariation in multiple sequence alignments (MSAs) constructed from diverse protein sequences from a protein family (Anishchenko, Ovchinnikov, Kamisetty, &

Baker, 2017; R. M. Levy, Haldane, & Flynn, 2017; Faruck Morcos et al., 2011). Residue covariation analysis methods infer a global probability model of sequences in the MSA which crucially captures the covariation of different columns of the MSA, while disentangling direct from indirect mutational covariation through inference of the underlying functional couplings which generated the observed covariation. These "Potts" models have been shown to capture important structural and functional information (Marks, Hopf, & Sander, 2012; Faruck Morcos et al., 2011; F. Morcos, Schafer, Cheng, Onuchic, & Wolynes, 2014; Ovchinnikov et al., 2015; Sulkowska, Morcos, Weigt, Hwa, & Onuchic, 2012). The models have further applications through the use of the probability model as a scoring function for individual sequences, for instance Potts models can predict effects of mutations to a sequence (Hopf et al., 2017; M, H, A, O, & M, 2016), or be used to predict structural or conformational preferences of individual sequences or subfamilies of sequences (A. Haldane, Flynn, He, Vijayan, & Levy, 2016; Sutto, Marsili, Valencia, & Gervasio, 2015).

Here, we have constructed sequence coevolutionary Potts Hamiltonian models for class A, B and C GPCRs, for which sufficient protein sequences are available. We also generate residue contact frequency maps from available structures of the three classes of GPCRs. Residue pairs that exhibit strong coevolutionary couplings but low structural contact frequencies are referred to as "exceptions" from the Potts model predictions. A number of such exception residue contacts added important missing links for activation and allosteric modulation of the GPCRs. We have provided evidence that dimerization is required for activation of class C GPCRs, but not for activation of class A and B GPCRs. We have also identified distinct residue contacts that are important for activation and allosteric modulation of the three classes of GPCRs.

Results

Structural residue contacts of GPCRs were detected in sequence coevolutionary models

We inferred separate sequence-based Potts coevolutionary models for the A, B and C classes of GPCRs and compared them with corresponding residue contact frequency maps (**Figure 1**). The input multiple sequence alignments (MSAs) included 84,481 sequences for class A, 17,804 sequences for class B, and 12,602 sequences for class C of GPCRs. Any sequences and columns in the MSAs with more than 10% gaps were removed, and through the correction to downweight phylogenetically similar sequences, we obtained 5126, 902 and 813 effective sequences for class A, B, and C GPCRs, respectively (**Table S1**). The residue sequences of class A, B and C GPCRs were also aligned using the *hmmalign* function of HMMER (Eddy, 2011) (**Table S2**). Moreover, residue contact frequency maps were built from all available structures of each class of GPCRs (**Table S3**). The structures included the active, inactive, active agonist-PAM-bound and inactive antagonist-NAM-bound GPCRs for all three classes as shown for a model class A GPCR in **Figure 1A**, a model class B GPCR in **Figure 1D** and a model class C GPCR in **Figure 1G**. The Potts coupling parameters, which were used to calculate Frobenius norms as residue contact scores, were visualized as a heatmap for each of the three classes of GPCRs. The Potts model was then directly compared with the residue contact frequency map of class A (**Figure 1B**), class B (**Figure 1E**) and class C (**Figure 1H**) of GPCRs.

The correspondence between the Potts model and residue contact frequency map of class A GPCRs was evident in residue interactions of the TM6-TM7, TM5-TM6, TM3-TM6, TM3-TM5, TM3-TM4, TM2-TM7, TM2-TM4, TM2-TM3, TM1-TM7 and TM1-TM2 domains (**Figure 1B**). Overall, the number of predicted residue contacts from Potts model of class B GPCRs was lower than that of class A GPCRs. Nevertheless, the Potts model and residue contact frequency

map of class B GPCRs were in agreement for residue interactions in the TM6-TM7, TM5-TM6, TM3-TM5, TM3-TM4, TM2-TM4, TM2-TM3, TM1-TM7 and TM1-TM2 domains (**Figure 1E**). For class C GPCRs, agreement between the Potts model and residue contact frequency map was observed for residue interactions in the TM6-TM7, TM5-TM6, TM3-TM6, TM3-TM4, TM2-TM7, TM2-TM3, TM1-TM7 and TM1-TM2 domains (**Figure 1H**). However, the number of predicted residue contacts from Potts model of class C GPCRs was further reduced compared with the classes A and B GPCRs.

Furthermore, we extracted the top 100 long-distance Frobenius norms, consisting of residue pairs that were ≥ 5 residues apart in the numbering scheme, from the Potts model of each GPCR class and examined whether the corresponding residue contacts were observed in the experimental GPCR structures. The residue contacts predicted from the sequence-based Potts models were deemed true positives if they were present in available GPCR structures above a percentage cutoff (e.g., 30% or 0.3). Those residue pairs with significantly high Frobenius norms (particularly ≥0.16 for class A, ≥0.19 for class B and ≥0.17 for class C) but with low contact frequencies below the percentage cutoff in the GPCR structures were considered exceptions in the Potts model predictions. A number of them were found to form important contacts for GPCR activation and allosteric modulation. With the cutoffs of 0.0 and 0.3 for the structural contact frequency, the residue contact detection rates of our Potts models were plotted for classes A, B and C of GPCRs in the **Figures 1C**, **1F** and **1I**, respectively. With zero contact frequency cutoff, the residue contact detection rates were 0.82, 0.56 and 0.42 for the top 100 long-distance Frobenius norms in the Potts models of the classes A, B and C of GPCRs, respectively. The detection rates decreased with increasing cutoff of the structural contact frequency. At the 0.3 contact frequency cutoff, the residue contact detection rates decreased to 0.59, 0.52 and 0.32 for the top 100 long-

distance Frobenius norms in the Potts models of the classes A, B and C of GPCRs, respectively. In comparison, the detection rate from Potts model of the class C GPCRs was lower than that of the A and B classes of GPCRs, likely due to the relatively smaller number of effective sequences (**Table S1**). This was consistent with previous studies that contact detection rate in the Potts model correlated with the number of effective sequences in the MSA (Ekeberg, 2013; A. Haldane, Levy, R.M., 2019). Based on the above findings, the 0.3 cutoff of structural contact frequency was used for further analysis.

Activation and Allosteric Modulation of Class A GPCRs

For class A GPCRs, contact exceptions with the top 20 long-distance Frobenius norms are summarized in **Table S4**. Notably, residue pair C3.44-V5.57 (rank 14) had the contact frequency increased to 0.38 in only the inactive class A GPCR structures, while residue pairs Y3.51-F5.56 (rank 28) and F5.47-L6.49 (rank 60) had the contact frequencies increased to 0.38 and 0.49 in only the active class A GPCR structures, respectively. Since these residue contacts showed significantly higher contact frequencies in one of the GPCR functional states (active or inactive), they were considered contacts that were important for GPCR activation or inactivation.

We built residue contact frequency maps for the active and inactive class A GPCRs and calculated their difference by subtracting residue contact frequencies in the inactive GPCR structures from those in the active structures. The residue contact frequency difference map between active and inactive class A GPCR structures is shown in **Figure 2A**. The exception residue contacts of rank 14, 28 and 60 were shown in **Figure 2B**. Furthermore, we highlighted the top 30 residue contacts with the largest differences of contact frequencies in the inactive and active GPCR structures (**Table S5A**). The seven switching residue contacts in the list included S3.39-F6.44,

I3.46-L6.37 and R3.50-L6.34 that were present in only the inactive GPCR structures (**Figure 2E**) and I3.46-Y7.53, Y5.58-I6.40, Y5.62-L6.37 and A5.65-A6.33 that were present in only the active GPCR structures (**Figures 2F-2H**). Overall, the residue contacts between TM1-TM7 (**Figure 2C**) and TM3-TM6 (**Figure 2E**) were significantly weakened, while a number of contacts between TM2-TM7 (**Figure 2D**), TM3-TM7 (**Figure 2F**) and TM5-TM6 (**Figures 2G-2H**) were strengthened upon class A GPCR activation.

In addition to activation, we examined residue contacts that were tuned by allosteric modulation of class A GPCRs. Among the list of top ranked exception residue contacts (**Table S4**), the Y3.51-S3.56 residue pair (rank 4) had the contact frequency decreased by 0.33 upon binding of NAMs to the inactive GPCRs. With binding of PAMs to the active GPCRs, two residue pairs A2.49-W4.50 (rank 11) and D3.49-Y89^{ICL2} (rank 30) had the contact frequency decreased by 0.67 and 0.33, respectively, while the F5.47-L6.49 residue pair (rank 60) had the contact frequency increased by 0.33. These exception residue contacts in the Potts model thus resulted from allosteric modulation of class A GPCRs.

We generated additional residue contact frequency maps for the active agonist-bound class A GPCRs in the absence and presence of PAMs and calculated their difference of residue contact frequencies (**Figure 3A**), similarly for the inactive antagonist-bound class A GPCRs in the absence and presence of NAMs (**Figure 3B**). The residue contacts that were significantly tuned in class A GPCRs in the presence of allosteric modulators were listed in **Table S5B**. During binding of PAMs to active class A GPCRs, residues N2.39-R3.50 and N2.39-C3.53 formed new contacts, whereas the exception residue pair A2.49-W4.50 in the Potts model lost contact (**Figure 3C**). Moreover, residues L3.27-L4.62, V3.34-L4.56, D3.49-A4.42 and V5.51-W6.48 also formed new contacts in the active class A GPCRs (**Figures 3D-3E**). With binding of NAMs to the inactive class A GPCRs,

residues T2.37-K4.39 formed new contacts between TM2-TM4 (**Figure 3F**), similarly for residue contacts I3.31-S4.57 between TM3-TM4 (**Figure 3F**), contacts A2.49-V3.36 between TM2-TM3 (**Figure 3G**) and contacts A5.39-H6.58 between TM5-TM6 (**Figure 3H**).

Finally, we performed an exhaustive search of residue contacts at the protein interface in experimental oligomer structures of class A GPCRs, including the rhodopsin, adenosine A1 receptor and CXCR4 receptors. Here, a residue contact was defined by ≤ 6.5Å between any side chain heavy atoms of the two residues from neighboring protein subunits. Only residues that were ≥ 5 residues apart in the numbering scheme were considered. In general, all of these residue contacts had negligible Frobenius norms, with the highest being 0.02, suggesting that residue contacts between class A oligomers lack coevolutionary basis (**Table S6**).

Activation and Allosteric Modulation of Class B GPCRs

For class B GPCRs, exception residue contacts with the top 20 long-distance Frobenius norms are summarized in **Table S7**. The contact frequency of residue pair S2.49-W4.50 (rank 2) increased to 0.31 in the active class B GPCR structures, while only 0.14 in the inactive structures. Therefore, this exception residue contact predicted in the Potts model could be explained by activation of class B GPCRs.

We computed the residue contact frequency difference map between active and inactive class B GPCR structures as shown in **Figure 4A**. The rank 2 exception residue contact is shown in **Figure 4B**. We highlighted the top 30 residue contacts with the largest differences in contact frequencies between active and inactive class B GPCR structures (**Table S8A**). A sharp kink was formed in the middle of the TM6 helix, leading to switching contacts including L3.43-L6.43, E3.46-L6.43, E3.46-L6.44, G3.47-L6.44, N5.54-L6.44, F5.58-L6.44, V5.62-T6.37/L6.40, L5.65-

6.33 and L5.69-6.33 in the active class B GPCR structures (**Figures 4C, 4F** and **4G**). Meanwhile, the TM6 intracellular end moved away from the TM3, TM5 and TM7 helices during class B GPCR activation, losing a number of key residue contacts that were formed in the inactive receptor structures, including L3.50-L6.33/A6.34/T6.37, L3.54-K6.30, L3.43-6.40/I6.41, E3.46-T6.37/L6.40, I5.61-A6.34, L6.40-L7.48 and L6.44-Q7.45 (**Figures 4D, 4E** and **4H**).

We also computed the residue contact frequency difference map for the active agonist-bound class B GPCRs in the absence and presence of PAMs (**Figure 5A**), similarly for the inactive antagonist-bound class B GPCRs in the absence and presence of NAMs (**Figure 5B**). The residue contacts that were significantly tuned in the presence of allosteric modulators were listed in **Table S8B**. With the binding of PAMs in active class B GPCR structures, residue pairs G3.47-L6.43, L3.50-L6.43 and L3.50-L6.44 formed new contacts between TM3-TM6 (**Figure 5C**), similarly for residue contacts I5.51-V6.46, L5.52-A6.52, N5.54-V6.46 and P5.55-L6.44 between TM5-TM6 (**Figure 5D**) and residue contacts H6.47-T7.37, H6.47-T7.42, E6.48-E7.38 and I6.50-E7.38 between TM6-TM7 (**Figure 5F**). With the binding of NAMs in the inactive class B GPCR structures, residue pairs L3.43-T6.37, E3.46-I6.41 and L3.50-K6.30 formed new contacts between TM3-TM6 (**Figure 5G**), similarly for contacts W5.40-A6.52, R5.44-E6.48 and F5.48-A6.52 between TM5-TM6 (**Figure 5H**) and contacts H6.47-Q7.45, V6.50-K7.34 and V6.50-D7.38 between TM6-TM7 (**Figure 5I**).

Dimerization, Activation and Allosteric Modulation of Class C GPCRs

For class C GPCRs, the exception residue contacts with the top 20 long-distance Frobenius norms are summarized in **Table S9**. Strikingly, exception residues S2.39-N3.48 (rank 4) had the contact frequency increased to 0.4 in the active class C GPCR structures, while residues S3.39-

Y6.42 (rank 11) had the contact frequency increased to 0.33 in the inactive C GPCR structures. Therefore, these two exception residue contacts in the Potts model resulted from activation of class C GPCRs.

We computed the residue contact frequency difference map between active and inactive class C GPCR structures as shown in **Figure 6A**. We highlighted the top 30 residue contacts with the largest differences in contact frequencies between the active and inactive class C GPCR structures (**Table S10A**). In particular, residue contacts between TM2-TM3 were mostly strengthened (**Figure 6C**), whereas contacts between TM2-TM4 (**Figure 6C**), TM2-TM7, TM3-TM7 (**Figure 6D**), TM3-TM4 (**Figure 6E**), TM3-TM5 (**Figure 6F**), TM4-TM5 (**Figure 6G**) and TM5-TM6 (**Figure 6H**) were mostly weakened.

In addition, we performed an exhaustive search of residue contacts at the protein interface in the available active structure of class C metabotropic glutamate receptor 5 (PDB:6N51) (**Figure 6B**). The criteria for consideration were the same as for class A dimers. We identified two residue contacts P6.53-F7.40 and Y6.42-V7.48 at the receptor dimer interface, with Frobenius norms of 0.12 for both. It was necessary to determine if this Frobenius norm value was sufficiently high to suggest that the two residue pairs coevolve and form contacts. In this context, the Frobenius norms of 230 long-distance residue contacts in one monomeric subunit of the 6N51 PDB structure were retrieved, and the average was calculated as 0.09 ± 0.04, with a median value of 0.09. These statistics thus provided coevolutionary basis for residue contacts P6.53-F7.40 and Y6.42-V7.48 in the active class C metabotropic glutamate receptor 5.

We also examined residue contacts tuned by allosteric modulation of class C GPCRs. Upon binding of PAMs to the active GPCRs, exception residue pair Y2.54-T7.47 (rank 1) had the contact frequency decreased by 0.5, whereas residue pairs S2.39-N3.48 (rank 4), I3.32-G5.45 (rank 10),

S3.39-Y6.42 (rank 11), T3.45-F768^{ICL3} (rank 23) and I6.37-V7.48 (rank 40) had the contact frequency increased by 1.0, 0.5, 0.5, 0.5 and 0.5, respectively. With binding of NAMs to the inactive GPCRs, residue pair T3.45-F768^{ICL3} (rank 23) had the contact frequency decreased by 1.0, while residue pair I3.53-N763^{ICL3} (rank 28) had the contact frequency increased by 1.0. These exception residue contacts in Potts model of class C GPCRs resulted from allosteric modulation of the receptors.

We further computed the residue contact frequency difference map for the active agonist-bound class C GPCRs in the absence and presence of PAMs (**Figure 7A**), similarly for the inactive antagonist-bound class C GPCRs in the absence and presence of NAMs (**Figure 7B**). We highlighted a number of residue contacts with the largest differences in contact frequencies in the class C GPCR structures upon binding of PAMs and NAMs (**Table S10B**). With the binding of PAMs to the active class C GPCR structures, residues F1.56-F7.54 and N1.57-F7.54 formed new contacts between TM1-TM7 (**Figure 7C**), similarly for residue contacts G3.41-L5.50, F3.44-L5.57, T3.47-A5.58, W3.48-L5.57 and H3.51-E5.60 between TM3-TM5 (**Figure 7D**) and residue contacts G5.54-I6.37, A5.58-I6.37 and R5.62-K6.35 between TM5-TM6 (**Figure 7E**). With the binding of NAMs to the inactive class C GPCR structures, residues F1.56, N1.57 and N1.60 in TM1 formed new contacts with residue K7.57 of TM7 (**Figure 7F**), similarly for residue contacts S2.53-C7.45 and S2.56-I7.42 between TM2-TM7 (**Figure 7G**) and residue contacts A3.39-V6.44 and F3.40-V6.41 between TM3-TM6 (**Figure 7H**).

Discussion

"Exception" residue contacts added important missing links for GPCR signaling

A number of exception contacts predicted from the Potts models added missing links that were critical for the activation and allosteric modulation of GPCRs (**Figures 8** and **9**) (Hopf et al., 2017; M et al., 2016). Specifically, 6, 1 and 7 of the top 20 exception contacts resulted from activation and/or allosteric modulation of the class A, B and C GPCRs, respectively (**Tables S4, S7** and **S9**). In class A GPCRs, exception contacts 3.51-5.56 and 5.47-6.49 augmented the residue network for receptor activation of by involving the DRY and CWxP functional motifs (Gao et al., 2003), respectively (**Figure 8A**). Exception contacts 3.49-149^{ICL2} and 3.51-3.56, both of which involved the DRY motif, were tuned off upon binding of PAMs and NAMs to the class A GPCRs, respectively (**Figures 9A** and **9D**). Another exception contact between 2.49 (Na^+ pocket) and 4.50 was important for both allosteric modulation of class A GPCRs (Gutierrez-de-Teran et al., 2013; Miao, Caliman, & McCammon, 2015; Shang et al., 2014; White et al., 2018) and activation of class B GPCRs (**Figures 8B** and **9A**). This was consistent with previous findings that W4.50 is fully conserved in class A and B GPCRs and its mutation significantly affected activation of the class B GPCRs (Peeters et al., 2016; A. J. Venkatakrishnan et al., 2013). In class C GPCRs, two exception contacts connecting TM3 to TM2 (3.48-2.39) and TM6 (3.39-6.42) were turned on and off upon receptor activation, respectively (**Figure 8C**). A significant number of exception residue contacts in class C GPCRs were also tuned by binding of PAMs (including 2.54-7.47, 3.32-5.45, 6.37-7.48) and NAMs (3.53-763^{ICL3}) (**Figure 9**). Notably, the 3.45-768^{ICL3} exception contact was consistently tuned on upon PAM binding (**Figure 9C**), but off upon NAM binding (**Figure 9F**). In addition, certain exception residue contacts in the Potts models appeared to result from arbitrary contact definition. They showed average Cβ-Cβ distances close to the 8Å contact cutoff, including 3 residue pairs T3.42-W4.50, D3.32-Y7.43 and L2.46-S7.46 in class A GPCRs, 1 residue pair I3.36-S7.42 in class B GPCRs and 1 residue pair V1.53-G7.51 in class C GPCRs (**Table S11**).

Comparison of activation pathways of class A, B and C GPCRs

We have calculated residue contact frequency difference maps between active versus inactive structures to identify important residue contacts for GPCR activation. These maps systematically uncovered residue contacts that are important for activation of class A, B and C GPCRs. For the most extensively studied class A GPCRs, our results were mostly consistent with previous findings. Among our 30 residue pairs that showed the largest differences in contact frequencies between the active and inactive class A GPCR structures, 10 of them involved the previously identified NACHOs (Cvicek, 2016) and 9 of them were found in the common activation pathway of class A GPCRs (Zhou, 2019) (**Table S12**). Five of the previous switching residue contacts (Zhou, 2019) became repacking contacts in this analysis provided a significant increase in the number of available experimental structures. Our analysis also revealed 20 additional residue contacts involved in class A GPCR activation (**Table S12**). Upon activation of class A GPCRs, residue contacts were lost between the intracellular domains of TM6 and TM3 (involving the highly conserved DRY motif), as well as for the conserved NPxxY motif in TM7 with the TM1, TM2 (involving the Na^+ binding pocket) (Fleetwood, Matricon, Carlsson, & Delemotte, 2020; White et al., 2018) and TM6. New residue contacts were formed between TM5-TM6 (involving the extracellular CWxP and PIF motifs) and TM3-TM7 (involving the NPxxY motif) (Fleetwood et al., 2020) (**Figure 8A**). It is well established that the TM6 intracellular end moves outwards, while the highly conserved NPxxY motif in the TM7 intracellular domain moves towards TM3 (Fleetwood et al., 2020; Huang et al., 2015; Lebon et al., 2011; Manglik & Kruse, 2017; R. Nygaard, Frimurer, T.M., Holst, B., Rosenkilde, M.M., Schwartz, T.W., 2009; S. G. Rasmussen, 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., Lyons, J.A., Caffrey, M., Gellman, S.H., Steyaert, J., Skiniotis, G., Weis, W.I., Sunahara, R.K., Kobilka, B.K., 2011; A. J. Venkatakrishnan, Deupi, X., Lebon, G., Heydenreich, F.M., Flock, T., Miljus, T., Balaji, S., Bouvier, M., Veprintsev, D.B., Tate, C.G., Schertler, G.F., Babu, M.M., 2016; Ye et al., 2016), which resulted in the switching TM3 residue contacts from TM6 to TM7. Our structural analysis determined the residue contacts that were weakened between TM1-TM7 and TM3-TM6 and strengthened between TM3-TM7 in this global movement.

For class B GPCR activation, all 30 top-ranked residue contacts involved residues from TM6, which is known to distort significantly to the point of losing helical properties and form a sharp kink in the middle with the intracellular domain pointing outwards (Hilger et al., 2020; Y. L. Liang et al., 2020; Mattedi, 2020). Due to dramatic conformational change of TM6, new switching contacts were formed near its sharp kink for the conserved PxxG motif, while residue contacts were lost for the TM3 (the HETx motif) and the TM6 intracellular domain (residues K6.30-I6.41) with TM5 and TM7 (**Figure 8B**).

For class C GPCRs, the major difference for activation is the requirement of dimerization between two receptor subunits (Kim, 2020; Koehl, 2019; Shaye, 2020). The activation leads to significant rotation of one receptor subunit relative to the other and formation of a new TM6-TM6 dimer interface. The TM3 and TM5 straighten substantially on the intracellular side (Kim, 2020; Shaye, 2020). Our calculations identified changes of residue contacts between notably TM2-TM3, TM2-TM7, TM3-TM4, TM3-TM5, TM3-TM7, TM4-TM5 and TM5-TM6 within the receptor monomer upon activation of class C GPCRs (**Figure 8C**) (**Figure 6**). Furthermore, we determined the long-distance residue contacts at the interfaces of class A and C dimeric structures. Based on the corresponding Frobenius norms from the Potts model, our analysis suggested that residue

contacts at the dimer interfaces of class A GPCRs lacked coevolutionary basis, whereas residue contacts at the dimer interfaces of class C coevolved.

Comparison of allosteric modulation of class A, B and C GPCRs

Our sequence coevolutionary and structural contact analysis further revealed important residue contacts tuned by allosteric modulation of the three A, B and C classes of GPCRs (**Figure 9**). Overall, the TM helices that were found important for activation could be also allosterically modulated in GPCRs. For class A GPCRs, residue contacts between TM2-TM3, TM2-TM4, TM3-TM4, TM3-TM5, TM4-TM5, TM5-TM6 and TM6-TM7 changed significantly upon binding of allosteric modulators (**Figures 9A** and **9D**). Residue pairs including 3.27-4.62, 3.34-4.56 and 5.51-6.48 (toggle switch in the CWxP motif) formed new contacts upon binding of PAMs, being consistent with the previous finding that PAM binding induced slight contraction of the receptor extracellular mouth (Gregory, Sexton, & Christopoulos, 2007; Kruse, 2013). Furthermore, residue contacts such as 4.53-5.46 and 5.51-6.48 were modified, due to shift of the TM5 towards the extracellular side relative to the TM4 (**Figure 9A**) (Lu, 2017). On the other hand, NAM binding induced conformational changes in the TM2 (the Na^+ binding pocket) (Gutierrez-de-Teran et al., 2013; Miao et al., 2015; Shang et al., 2014; White et al., 2018), TM3 (the DRY motif), TM4, TM5, TM6 and TM7 (the NPxxY motif) domains of the class A GPCRs (**Figure 9D**) (Liu et al., 2018; Maeda et al., 2020). For class B GPCRs, residues in the TM6, which is heavily involved in activation of the receptors (Hilger et al., 2020; A. Jazayeri, Doré, A.S., Lamb, D., Krishnamurthy, H., Southall, S.M., Baig, A.H., Bortolato, A., Koglin, M., Robertson, N.J., Errey, J.C., Andrews, S.P., Teobald, I., Brown, A.J.H., Cooke, R.M., Weir, M., Marshall, F.H., 2016), were also found to undergo substantial changes of contacts in the presence of allosteric modulators, especially near

the PxxG and HETx motifs (**Figures 9B** and **9E**). For class C GPCRs, PAM binding was able to tune residue contacts in TM helices near the dimer interface within the receptor monomers (**Figures 9C** and **9F**), which resulted from substantial shifts of receptor TM3 and TM5 intracellular domains (Kim, 2020; Shaye, 2020). Our study, for the first time, systematically identified residue contacts that are tuned by allosteric modulation in GPCRs. They provide a framework for rationally designing selective allosteric drugs to modulate the structure and function of GPCRs.

Conclusion

We have revealed unique features of classes A, B and C GPCRs from combined sequence coevolutionary and structural contact analysis. A total number of 9, 2 and 8 of the top 20 exception residue contacts in the Potts models have been explained for class A, B and C GPCRs, respectively. They result from activation and allosteric modulation of GPCRs, as well as the arbitrary contact definition. However, the remaining exception residue contacts are still unexplained (**Tables S4, S7** and **S9**). A certain number of these exceptions could possibly result from inaccuracy of the sequence coevolutionary models, especially for the class B and C GPCRs that have relatively smaller numbers of effective sequences. Further explanations of the other exception residue contacts, however, can be potentially obtained with more GPCR structures that are yet to be determined in the future (V. Isberg et al., 2016).

The number of inactive experimental structures of class A GPCRs is about twice of their active structures (**Table S1**). More of the latter are thus needed, which has been recently boosted by remarkable advances in the cryo-EM technique for solving the active GPCR-G protein complex structures. More experimental structures are also needed for the class B and C GPCRs, especially the class B2 GPCRs for which the first structure has been obtained only very recently (Ping et al.,

2021). Only a handful of structures are currently available for the GPCRs bound by allosteric modulators. Because the allosteric modulators are advantageous over traditional agonists and antagonists for providing more selective therapeutic drugs of GPCRs, a significantly larger number of allosteric modulator-bound structures may be expected in the near future. In addition, more structures associated with new functional mechanisms of GPCRs (e.g., biased agonism (Rankovic, Brust, & Bohn, 2016)) await to be determined. We need to obtain GPCR structures bound by the biased agonists. When more of the above structures become available, they should be periodically added for updated structural contact analysis. Different residue contacts can be potentially identified and used to explain the remaining exception contacts in Potts models of the GPCRs.

Compared with the inactive and active structures, intermediate structures have been rarely determined for GPCRs (**Table S1**). This largely results from the highly dynamic nature and relatively short lifetime of the GPCR intermediates, which have proven difficult to characterize in experiments. In this context, computational molecular dynamics simulations that have captured activation pathway and intermediate conformations of the GPCRs (Dror et al., 2011; Kohlhoff et al., 2014; Miao & McCammon, 2016; Seidal et al., 2017) provide a promising approach to address the challenge. The simulation derived intermediate structures can be also used for additional structural contact analysis. Meanwhile, it will enable a more complete sequence-structure-dynamics-function relationship analysis of the GPCRs.

Furthermore, while the number of sequences for class D, E and F GPCRs are too low to construct accurate sequence coevolutionary Potts models (Ekeberg, 2013; A. Haldane, Levy, R.M., 2019), we can still perform structural contact analysis and dynamics simulations of these GPCRs. These studies will help us to identify important residue contacts and understand functional mechanisms of these even less studied or orphan GPCRs. Finally, it is worthy constructing a

sequence coevolutionary Potts model for the GPCR superfamily. When enough representative structures are obtained for each class of GPCRs, we can include all the structures for more complete analysis. Then the residue contact frequency map can be compared with the sequence coevolutionary model uncover common features of the different GPCR classes in the entire superfamily.

This analysis demonstrates how the combination of residue covariation and structural statistical analysis can supplement each other when applied to GPCR datasets. It establishes the basic connections between the sequence covariations and GPCR structural contacts, setting the stage for more advanced uses of the Potts model to study GPCR function. Notably, the Potts model has additional uses in predicting the energetics of individual sequences and particular residue pairs. For instance, the Potts coupling values for residue pairs in a single target sequence have been used in "threading" calculations with protein structures to determine the propensity of that sequence to take on particular functionally active or inactive conformations (A. Haldane et al., 2016). The Potts model can also be used to predict the functional characteristics of individual residue-residue interactions (Coucke et al., 2016). Applying this sequence-based energetics view to the GPCR family may further uncover details of GPCR function. These analyses and detailed understanding of the functional mechanisms are expected to greatly facilitate rational structure-based drug design of the pharmaceutically important GPCRs.

Methods

Sequence datasets

We built multiple sequence alignments (MSAs) of class A, B, and C GPCRs from the corresponding Pfam IDs of PF00001, PF00002, and PF00003, respectively. The MSAs included

84,481 sequences for class A, 17,804 sequences for class B, and 12,602 sequences for class C of GPCRs (**Table S1**). Any sequences and columns in the MSAs with more than 10% gaps were removed, leaving 37,471 sequences of length 235 residues for class A, 12,645 sequences of length 226 residues for class B, and 10,127 sequences of length 223 residues for class C of GPCRs. The residue sequences of class A, B and C GPCRs were also aligned using the *hmmalign* function of HMMER (Eddy, 2011) (**Table S2**). A number of these sequences, however, were phylogenetically related and sampled with experimental biases (A. Haldane, Flynn, W.F., He, P., Levy, R.M., 2018; A. Haldane, Flynn, W.F., He, P., Vijayan, R.S., Levy, R.M., 2016; F. Morcos, Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T., Weigt, M., 2011). One common correction was to downweight similar sequences by assigning a weight $w = {}^{1}/_{n}$ to each sequence, with n being the number of sequences that are more than 40% similar to the target sequence (A. Haldane, Flynn, W.F., He, P., Levy, R.M., 2018; A. Haldane, Flynn, W.F., He, P., Vijayan, R.S., Levy, R.M., 2016; F. Morcos, Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T., Weigt, M., 2011). The correction resulted in a number of effective sequences $N_{eff} = \sum w$ of 5126 sequences for class A, 902 sequences for class B1, and 813 sequences for class C of GPCRs (**Table S1**).

Sequence coevolutionary Potts model inference

Potts Hamiltonian models of residue covariation was built based on the assumption that two interacting residues in a protein have correlated mutations during evolution (A. Haldane, Flynn, W.F., He, P., Levy, R.M., 2018; A. Haldane, Flynn, W.F., He, P., Vijayan, R.S., Levy, R.M., 2016). The Potts model seeks to construct a least biased sequence probability distribution $P(S) \propto e^{-H(S)}$ that reproduces one-site $\langle s_i \rangle$ and two-site $\langle s_i s_j \rangle$ mutational probabilities of a

protein MSA (A. Haldane, Flynn, W.F., He, P., Levy, R.M., 2018; A. Haldane, Flynn, W.F., He, P., Vijayan, R.S., Levy, R.M., 2016; R. M. Levy, Haldane, A., Flynn, W.F., 2017). In this equation, *H(S)* is the Hamiltonian and takes the form of $H(S) = \sum_{i=1}^{L} h_i(s_i) + \sum_{i=1}^{L} \sum_{i<j}^{L} J_{ij}(s_i, s_j)$, in which *S* is a sequence of amino acid types (s) at each of L positions, $h_i(s_i)$, or "fields", refers to the single point contribution to the statistical energy of residue s_i at position i, and $J_{ij}(s_i,s_j)$, or "couplings", stands for the energy contribution of a position-pair i and j (A. Haldane, Flynn, W.F., He, P., Levy, R.M., 2018; A. Haldane, Flynn, W.F., He, P., Vijayan, R.S., Levy, R.M., 2016; R. M. Levy, Haldane, A., Flynn, W.F., 2017). Strong coupling parameters $J_{ij}(s_i,s_j)$ often correspond to direct physical interactions in protein 3D structures (A. Haldane, Flynn, W.F., He, P., Vijayan, R.S., Levy, R.M., 2016; F. Morcos, Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T., Weigt, M., 2011; Weigt, 2008); therefore, the model is of great interest in the field of structure determination. Mi3-GPU is a newly developed software that solves for real-valued coupling parameters $J_{ij}(s_i,s_j)$ in the Hamiltonian H(S) with few approximations using Markov-Chain Monte-Carlo (MCMC) methods with quasi-Newton optimization (A. Haldane, Levy, R.M., 2020). Mi3-GPU can be used to construct Potts models with high statistical precision. A full description of the program can be found here (A. Haldane, Levy, R.M., 2020).

We applied Mi3-GPU to infer the Potts models of class A, B, and C GPCRs. A 21-letter alphabet was used for Potts covariation analysis, including the 20 amino acids plus gap. The damping parameter was set to 0.01 and an l1 regularization of 0.002 throughout the inference. Three rounds of inference were performed for each Potts model. In the first round, 2^{18} walkers with 64 MCMC steps were used to reduce the sum-of-square-residuals (SSR), average bivariate marginal relative error (Ferr) and the covariance energy (X). Since our datasets were from Pfam,

the X values were not yet fully leveled off after the first round, a second round of inference was thus performed using an increased number of 2^{20} walkers with 32-64 MCMC steps to fully level off X values. One final round with 2^{22} walkers for four MCMC steps was performed to minimize the finite-sampling error and obtain a model with statistically accurate marginals and negligible residuals (A. Haldane, Levy, R.M., 2020).

Residue interaction score using weighted Frobenius norm

A weighted Frobenius norm (Ekeberg, 2013) was used to obtain a residue pair interaction score from resulting Potts model parameters to control and reduce the contribution of marginals with a large sampling error(A. Haldane, Flynn, W.F., He, P., Levy, R.M., 2018). The interaction score can be calculated from the coupling parameters $J_{\alpha\beta}^{ij}$ using the following formula $I^{ij} = \sqrt{\sum_{\alpha\beta}(w_{\alpha\beta}^{ij} J_{\alpha\beta}^{ij})^2}$, where $w_{\alpha\beta}^{ij}$ is positive and tunable(A. Haldane, Flynn, W.F., He, P., Levy, R.M., 2018; A. Haldane, Levy, R.M., 2020).

Construction of residue contact frequency maps

From the Protein Data Bank (PDB) (Berman, 2000; Burley, 2019) and GPCRdb (V. Isberg, 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; Pándy-Szekeres, 2018; van der Kant, 2014), we collected 283 structures of class A GPCRs (94 active, 185 inactive, and four intermediate), 31 structures of class B GPCRs (16 active, 14 inactive, and one intermediate), and 11 structures of class C GPCRs (five active and six inactive) (**Table S3**). Among the active GPCR structures, three class A GPCRs were bound by both the agonist and positive allosteric modulator (PAM) (PDBs: 4MQT, 6OIK and 6N48), similarly for one class B GPCR (PDB: 6VCB) and two class C GPCRs

(PDBs: 7C7Q and 6UO8). Among the inactive GPCR structures, three class A GPCRs were bound by both the antagonist and negative allosteric modulator (NAM) (PDBs: 5X7D, 6OBA and 5T1A), similarly for one class B GPCR (PDB: 5EE7) and one class C GPCR (PDB: 7CUM). A summary of the available GPCR structural data is included in **Table S1**.

The contact frequencies between conserved residues of class A, B and C GPCRs were determined. Two residues were considered in contact if their Cβ-Cβ atom distance was ≤8 Å, unless stated otherwise. For glycine, the Cα atom was used. The residue contact frequency maps were built for classes A, B and C of GPCRs using all the available structures for comparison with the sequence coevolutionary Potts models. Exceptions of residue contacts that showed significantly high Frobenius norms but low structural contact frequencies were identified in the Potts models of classes A, B and C GPCRs. Note that a number of these exception residue contacts could result from the arbitrary contact definition since they showed average Cβ-Cβ distances close to the 8Å cutoff (**Table S11**).

Residue contact frequency difference maps were built for active versus inactive GPCR structures to identify residue contacts involved in activation of the three classes of GPCRs. The concepts of switching and repacking contacts for GPCR activation were adopted from Zhou et al. (Zhou, 2019). Switching contacts were defined as residue contacts that were present in only one GPCR functional state (active or inactive). Repacking contacts were residue contacts that were present in both GPCR functional states but showed notable changes in contact frequencies. Furthermore, residue contact frequency difference maps were built for active agonist-bound GPCRs in the presence versus absence of PAMs, as well as inactive antagonist-bound GPCRs in the presence versus absence of NAMs, to identify residue contacts tuned by allosteric modulation in the three classes of GPCRs.

Acknowledgements

This work used supercomputing resources with allocation awards TG-MCB180049 through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation (NSF) grant ACI-1548562 and project M2874 through the National Energy Research Scientific Computing Center (NERSC), which is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231, and computational resources provided by the Research Computing Cluster at the University of Kansas. This work was supported in part by the American Heart Association (Award 17SDG33370094 to Y.M.), the National Institutes of Health (R01GM132572 to Y.M. and 5U54-AI150472 and R35GM132090 to R.L.), NSF (193484 to R.L.) and the startup funding in the College of Liberal Arts and Sciences at the University of Kansas (to Y.M.).

Competing interests

The authors declare that no competing interests exist.

References:

Anishchenko, I., Ovchinnikov, S., Kamisetty, H., & Baker, D. (2017). Origin of coevolution between residues distant in protein 3D structures. *Proc Natl Acad Sci U S A, 114*, 9122.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Research, 28*, 235-242.

Bjarnadóttir, T. K., Gloriam, D.E., Hellstrand, S.H., Kristiansson, H., Fredriksson, R., Schiöth, H.B. (2006). Comprehensive repertoire and phylogenetic analysis of the G-protein-coupled receptors in human and mouse. *Genomics, 88*(3), 263-273.

Bueno, A. B., Sun, B., Willard, F. S., Feng, D., Ho, J. D., Wainscott, D. B., . . . Sloop, K. W. (2020). Structural insights into probe-dependent positive allosterism of the GLP-1 receptor. *Nat Chem Biol, 16*(10), 1105-1110. doi:10.1038/s41589-020-0589-7

Burley, S. K., Berman, H.M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Dalenberg, K., Duarte, J.M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D.S., Green, R.K., Guranovic, V., Guzenko, D., Hudson, B.P., Kalro, T., Liang, Y., Lowe, R., Namkoong, H., Peisach, E., Periskova, I., Prlic, A., Randle, C., Rose, A., Rose, P., Sala, R., Sekharan, M., Shao, C., Tan, L., Tao, Y.P., Valasatava, Y., Voigt, M., Westbrook, J., Woo, J., Yang, H., Young, J., Zhuravleva, M., Zardecki, C. (2019). RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Research, 47*(D), D464-D474.

Christopoulos, A. (2002). Allosteric binding sites on cell-surface receptors: Novel targets for drug discovery. *Nature Reviews Drug Discovery, 1*(3), 198-210. doi:Doi 10.1038/Nrd746

Chun, L., Zhang, W.H., Liu, J.F. (2012). Structure and ligand recognition of class C GPCRs. *Acta Pharmacologica Sinica, 33*, 312-323.

Coucke, A., Uguzzoni, G., Oteri, F., Cocco, S., Monasson, R., & Weigt, M. (2016). Direct coevolutionary couplings reflect biophysical residue interactions in proteins. *Journal of Chemical Physics, 145*(17), 174102. doi:10.1063/1.4966156

Cvicek, V., Goddard, W.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.

Draper-Joyce, C. J., Khoshouei, M., Thal, D. M., Liang, Y.-L., Nguyen, A. T. N., Furness, S. G. B., . . . Christopoulos, A. (2018). Structure of the adenosine-bound human adenosine A1 receptor–Gi complex. *Nature, 558*, 559–563. doi:10.1038/s41586-018-0236-6

Dror, R. O., Arlow, D. H., Maragakis, P., Mildorf, T. J., Pan, A. C., Xu, H., . . . Shaw, D. E. (2011). Activation mechanism of the β2-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A., 108*(46), 18684-18689. Retrieved from https://www.pnas.org/content/pnas/108/46/18684.full.pdf

Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Computational Biology, 7*(10), e1002195.

Ekeberg, M., Lövkvist, C., Lan, Y., Weigt, M., Aurell, E. (2013). Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. *Physcial Review E, 87*(1), 012707.

Fleetwood, O., Matricon, P., Carlsson, J., & Delemotte, L. (2020). Energy Landscapes Reveal Agonist Control of G Protein-Coupled Receptor Activation via Microswitches. *Biochemistry, 59*(7), 880-891. doi:10.1021/acs.biochem.9b00842

Gao, Z. G., Kim, S. K., Gross, S., Chen, A. S., Blaustein, J. B., & Jacobson, K. A. (2003). Identification of essential residues involved in the allosteric modulation of the human A(3)

adenosine receptor. *Molecular Pharmacology, 63*(5), 1021-1031. doi:DOI 10.1124/mol.63.5.1021

Garcia-Nafria, J., Lee, Y., Bai, X., Carpenter, B., & Tate, C. G. (2018). Cryo-EM structure of the adenosine A2A receptor coupled to an engineered heterotrimeric G protein. *Elife, 7*. doi:10.7554/eLife.35946

Goblyos, A., & Ijzerman, A. P. (2011). Allosteric modulation of adenosine receptors. *Biochimica Et Biophysica Acta-Biomembranes, 1808*(5), 1309-1318. doi:10.1016/j.bbamem.2010.06.013

Gregory, K. J., Sexton, P. M., & Christopoulos, A. (2007). Allosteric modulation of muscarinic acetylcholine receptors. *Curr Neuropharmacol, 5*(3), 157-167. doi:10.2174/157015907781695946

Grushevskyi, E. O., Kukaj, T., Schmauder, R., Bock, A., Zabel, U., Schwabe, T., . . . Lohse, M. J. (2019). Stepwise activation of a class C GPCR begins with millisecond dimer rearrangement. *Proc Natl Acad Sci U S A, 116*(20), 10150-10155. doi:10.1073/pnas.1900261116

Gutierrez-de-Teran, H., Massink, A., Rodriguez, D., Liu, W., Han, G. W., Joseph, J. S., . . . Stevens, R. C. (2013). The Role of a Sodium Ion Binding Site in the Allosteric Modulation of the A Adenosine G Protein-Coupled Receptor. *Structure, 21*(12), 2175-2185. doi:10.1016/j.str.2013.09.020

Haldane, A., Flynn, W. F., He, P., Vijayan, R. S., & Levy, R. M. (2016). Structural propensities of kinase family proteins from a Potts model of residue co-variation. *Protein Sci, 25*(8), 1378-1384. doi:10.1002/pro.2954

Haldane, A., Flynn, W.F., He, P., Levy, R.M. (2018). Coevolutionary Landscape of Kinase Family Proteins: Sequence Probabilities and Functional Motifs. *Biophysical Journal, 114*, 21-31.

Haldane, A., Flynn, W.F., He, P., Vijayan, R.S., Levy, R.M. (2016). Structural propensities of kinase family proteins from a Potts model of residue covariation. *Protein Sciences, 25*(8), 1378-1384.

Haldane, A., Levy, R.M. (2019). Influence of multiple-sequence-alignment depth on Potts statistical models of protein covariation. *Physcial Review, 99*(3), 032405(032415).

Haldane, A., Levy, R.M. (2020). Mi3-GPU: MCMC-based Inversing Ising Inference on GPUs for protein covariation analysis. *Computer Physics Communications, 107312*.

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*, 41-54.

Hilger, D., Kumar, K. K., Hu, H., Pedersen, M. F., O'Brien, E. S., Giehm, L., . . . Kobilka, B. K. (2020). Structural insights into difference in G protein activation by family A and family B GPCRs. *Science, 369*(6503), eaba3373.

Hopf, T. A., Ingraham, J. B., Poelwijk, F. J., Schärfe, C. P. I., Springer, M., Sander, C., & Marks, D. S. (2017). Mutation effects predicted from sequence co-variation. *Nature Biotechnology, 35*(2), 128-135. doi:doi:10.1038/nbt.3769

Huang, W., Manglik, A., Venkatakrishnan, A. J., Laeremans, T., Feinberg, E. N., Sanborn, A. L., . . . Kobilka, B. K. (2015). Structural insights into micro-opioid receptor activation. *Nature*. doi:10.1038/nature14886

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.

Isberg, V., Mordalski, S., Munk, C., Rataj, K., Harpsøe, K., Hauser, A. S., . . . Gloriam, D. E. (2016). GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Research, 44*(D1), D365-364.

Jazayeri, A., Doré, A.S., Lamb, D., Krishnamurthy, H., Southall, S.M., Baig, A.H., Bortolato, A., Koglin, M., Robertson, N.J., Errey, J.C., Andrews, S.P., Teobald, I., Brown, A.J.H., Cooke, R.M., Weir, M., Marshall, F.H. (2016). Extra-helical binding site of a glucagon receptor antagonist. *Nature, 533*, 274-277.

Jazayeri, A., Rappas, M., Brown, A. J. H., Kean, J., Errey, J. C., Robertson, N. J., . . . Marshall, F. H. (2017). Crystal structure of the GLP-1 receptor bound to a peptide agonist. *Nature, 546*(7657), 254-258. doi:10.1038/nature22800

Jeffrey Conn, P., Christopoulos, A., & Lindsley, C. W. (2009). Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nat Rev Drug Discov, 8*(1), 41-54. Retrieved from http://dx.doi.org/10.1038/nrd2760

Keov, P., Sexton, P., Christopoulos, A. (2011). Allosteric modulation of G protein-coupled receptors: A pharmacological perspective. *Neuropharmacology, 60*(1), 24-35.

Kim, Y., Jeong, E., Jeong, J-H., Kim, Y., Cho, Y. (2020). Structural Basis for Activation of the Heterodimeric GABAB Receptor. *Journal of Molecular Biology, 432*(32), 5966-5984.

Koehl, A., Hu, H., Feng, D., Sun, B., Zhang, Y., Robertson, M.J., Chu, M., Kobilka, T.S., Laeremans, T., Steyaert, J., Tarrasch, J., Dutta, S., Fonseca, R., Weis, W.I., Mathiesen, J.M., Skiniotis, G., Kobilka, B.K. (2019). Structural insights into the activation of metabotropic glutamate receptors. *Nature, 567*(7747), E10.

Kohlhoff, K. J., Shukla, D., Lawrenz, M., Bowman, G. R., Konerding, D. E., Belov, D., . . . Pande, V. S. (2014). Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nature Chemistry, 6*(1), 15-21. doi:Doi 10.1038/Nchem.1821

Kruse, A. C., Ring, A.M., Manglik, A., Hu, J., Hu, K., Eitel, K., Hübner, H., Pardon, E., Valant, C., Sexton, P.M., Christopoulos, A., Felder, C.C., Gmeiner, P., Steyaert, J., Weis, W.I., Garcia, K.C., Wess, J., Kobilka, B.K. (2013). Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature, 504*(7478), 101-106.

Lebon, G., Warne, T., Edwards, P. C., Bennett, K., Langmead, C. J., Leslie, A. G. W., & Tate, C. G. (2011). Agonist-bound adenosine A(2A) receptor structures reveal common features of GPCR activation. *Nature, 474*(7352), 521-U154. doi:10.1038/nature10136

Levy, R. M., Haldane, A., & Flynn, W. F. (2017). Potts Hamiltonian models of protein co-variation, free energy landscapes, and evolutionary fitness. *Current Opinion in Structural Biology, 43*, 55-62. doi:https://doi.org/10.1016/j.sbi.2016.11.004

Levy, R. M., Haldane, A., Flynn, W.F. (2017). Potts Hamiltonian models of protein covariation, free energy landscapes, and evolutionary fitness. *Current Opinion in Structural Biology, 43*, 55-62.

Liang, Y.-L., Khoshouei, M., Glukhova, A., Furness, S. G. B., Zhao, P., Clydesdale, L., . . . Wootten, D. (2018). Phase-plate cryo-EM structure of a biased agonist-bound human GLP-1 receptor–Gs complex. *Nature, 555*, 121. doi:10.1038/nature25773

https://www.nature.com/articles/nature25773#supplementary-information

Liang, Y. L., Belousoff, M. J., Zhao, P., Koole, C., Fletcher, M. M., Truong, T. T., . . . Wootten, D. (2020). Toward a Structural Understanding of Class B GPCR Peptide Binding and Activation. *Mol Cell, 77*(3), 656-668 e655. doi:10.1016/j.molcel.2020.01.012

Liu, H., Kim, H. R., Deepak, R., Wang, L., Chung, K. Y., Fan, H., . . . Zhang, C. (2018). Orthosteric and allosteric action of the C5a receptor antagonists. *Nat Struct Mol Biol, 25*(6), 472-481. doi:10.1038/s41594-018-0067-z

Lu, J., Byrne, N., Wang, J., Bricogne, G., Brown, F.K., Chobanian, H.R., Colletti, S.L., Di Salvo, J., Thomas-Fowlkes, B., Guo, Y., Hall, D.L., Hadix, J., Hastings, N.B., Hermes, J.D., Ho, T., Howard, A.D., Josien, H., Kornienko, M., Lumb, K.J., Miller, M.W., Patel, S.B., Pio, B., Plummer, C.W., Sherborne, B.S., Sheth, P., Souza, S., Tummala, S., Vonrhein, C., Webb, M., Allen, S.J., Johnston, J.M., Weinglass, A.B., Sharma, S., Soisson, S.M. (2017). Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40. *Nature Structural and Molecular Biology, 24*(7), 570-577.

M, F., H, J., A, S., O, T., & M, W. (2016). Coevolutionary Landscape Inference and the Context-Dependence of Mutations in Beta-Lactamase TEM-1. *Molecular biology and evolution, 33*(1). doi:10.1093/molbev/msv211

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. doi:10.1126/science.aaw5188

Maeda, S., Xu, J., FM, N. K., Clark, M. J., Zhao, J., Tsutsumi, N., . . . Kobilka, B. K. (2020). Structure and selectivity engineering of the M1 muscarinic receptor toxin complex. *Science, 369*(6500), 161-167. doi:10.1126/science.aax2517

Manglik, A., & Kruse, A. C. (2017). Structural Basis for G Protein-Coupled Receptor Activation. *Biochemistry, 56*(42), 5628-5634. doi:10.1021/acs.biochem.7b00747

Marks, D. S., Hopf, T. A., & Sander, C. (2012). Protein structure prediction from sequence variation. *Nature Biotechnology, 30*, 1072-1080.

Mattedi, G., Acosta-Gutiérrez, S., Clark, T., Grevasio, F.L. (2020). A combined activation mechanism for the glucagon receptor. *Proc Natl Acad Sci USA, 117*(27), 15414-15422.

May, L. T., Leach, K., Sexton, P. M., & Christopoulos, A. (2007). Allosteric modulation of G protein-coupled receptors. *Annual Review of Pharmacology and Toxicology, 47*, 1-51. doi:10.1146/annurev.pharmtox.47.120505.105159

Miao, Y., Caliman, A. D., & McCammon, J. A. (2015). Allosteric Effects of Sodium Ion Binding on Activation of the M3 Muscarinic G-Protein Coupled Receptor. *Biophysical Journal, 108*(7), 1796–1806. doi:10.1016/j.bpj.2015.03.003

Miao, Y., & McCammon, J. A. (2016). Graded activation and free energy landscapes of a muscarinic G-protein–coupled receptor. *Proc Natl Acad Sci U S A, 113*(43), 12162–12167. doi:10.1073/pnas.1614538113

Miao, Y., Nichols, S. E., Gasper, P. M., Metzger, V. T., & McCammon, J. A. (2013). Activation and dynamic network of the M2 muscarinic receptor. *Proc Natl Acad Sci U S A, 110*(27), 10982-10987. doi:10.1073/pnas.1309755110

Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., . . . Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences, 108*(49), E1293. doi:10.1073/pnas.1111471108

Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T., Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci USA, 108*(49), E1293-1301.

Morcos, F., Schafer, N. P., Cheng, R. R., Onuchic, J., & Wolynes, P. G. (2014). Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. *Proc Natl Acad Sci U S A, 111*(34), 12408-12413.

Nawaratne, V., Leach, K., Felder, C. C., Sexton, P. M., & Christopoulos, A. (2010). Structural determinants of allosteric agonism and modulation at the M4 muscarinic acetylcholine receptor: identification of ligand-specific and global activation mechanisms. *J Biol Chem, 285*(25), 19012-19021. doi:10.1074/jbc.M110.125096

Nygaard, R., Frimurer, T.M., Holst, B., Rosenkilde, M.M., Schwartz, T.W. (2009). Ligand binding and microswitches in 7TM receptor structures. *Trends in Pharmacological Sciences, 30*(5), 249-259.

Nygaard, R., Zou, Y., Dror, R. O., Mildorf, T. J., Arlow, D. H., Manglik, A., . . . Kobilka, B. K. (2013). The dynamic process of beta(2)-adrenergic receptor activation. *Cell, 152*(3), 532-542. doi:10.1016/j.cell.2013.01.008

Ovchinnikov, S., Kinch, L., Park, H., Liao, Y., Pei, J., Kim, D. E., . . . Baker, D. (2015). Large-scale determination of previously unsolved protein structures using evolutionary information. *Elife, 4*, e09248.

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.

Papasergi-Scott, M. M., Robertson, M. J., Seven, A. B., Panova, O., Mathiesen, J. M., & Skiniotis, G. (2020). Structures of metabotropic GABAB receptor. *Nature, 584*(7820), 310-314. doi:10.1038/s41586-020-2469-4

Peeters, M. C., Li, Q., Elands, R., van Westen, G. J., Lenselink, E. B., Muller, C. E., & AP, I. J. (2014). Domains for activation and inactivation in G protein-coupled receptors--a mutational analysis of constitutive activity of the adenosine A2B receptor. *Biochem Pharmacol, 92*(2), 348-357. doi:10.1016/j.bcp.2014.08.022

Peeters, M. C., Li, Q. L., van Westen, G. J. P., & IJzerman, A. P. (2012). Three "hotspots" important for adenosine A(2B) receptor activation: a mutational analysis of transmembrane domains 4 and 5 and the second extracellular loop. *Purinergic Signalling, 8*(1), 23-38. doi:10.1007/s11302-011-9251-x

Peeters, M. C., Mos, I., Lenselink, E. B., Lucchesi, M., AP, I. J., & Schwartz, T. W. (2016). Getting from A to B-exploring the activation motifs of the class B adhesion G protein-coupled receptor subfamily G member 4/GPR112. *FASEB J, 30*(5), 1836-1848. doi:10.1096/fj.201500110

Ping, Y. Q., Mao, C., Xiao, P., Zhao, R. J., Jiang, Y., Yang, Z., . . . Sun, J. P. (2021). Structures of the glucocorticoid-bound adhesion receptor GPR97-Go complex. *Nature*. doi:10.1038/s41586-020-03083-w

Rankovic, Z., Brust, T. F., & Bohn, L. M. (2016). Biased agonism: An emerging paradigm in GPCR drug discovery. *Bioorganic & Medicinal Chemistry Letters, 26*(2), 241-250. doi:10.1016/j.bmcl.2015.12.024

Rasmussen, S. G., 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., 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 β2 adrenergic receptor-Gs protein complex. *Nature, 477*(7366), 549-555.

Rasmussen, S. G. F., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., . . . Kobilka, B. K. (2011). Crystal structure of the β2 adrenergic receptor–Gs protein complex. *Nature, 477*, 549. doi:10.1038/nature10361
https://www.nature.com/articles/nature10361#supplementary-information

Seidal, L., Zarzycka, B., Zaidi, S. A., Katritch, V., & Coin, I. (2017). Structural insight into the activation of a class B G-protein-coupled receptor by peptide hormones in live human cells. *Elife, 6*. doi:ARTN e27711
10.7554/eLife.27711

Shang, Y., LeRouzic, V., Schneider, S., Bisignano, P., Pasternak, G. W., & Filizola, M. (2014). Mechanistic Insights into the Allosteric Modulation of Opioid Receptors by Sodium Ions. *Biochemistry, 53*(31), 5140-5149. doi:10.1021/Bi5006915

Shaye, H., Ishchenko, A., Lam, J.H., Han, G.W., Xue, L., Rondard, P., Pin, J.P., Katritch, V., Gati, C., Cherezov, V. (2020). Structural basis of the activation of a metabotropic GABA receptor. *Nature, 584*(7820), 298-303.

Stevens, R. C., Cherezov, V., Katritch, V., Abagyan, R., Kuhn, P., Rosen, H., & Wüthrich, K. (2013). The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. *Nature Reviews Drug Discovery, 12*(1), 25-34.

Sulkowska, J. I., Morcos, F., Weigt, M., Hwa, T., & Onuchic, J. (2012). Genomics-aided structure prediction. *Proc Natl Acad Sci U S A, 109*(26), 10340-10345.

Sutto, L., Marsili, S., Valencia, A., & Gervasio, F. L. (2015). From residue coevolution to protein conformational ensembles and functional dynamics. doi:10.1073/pnas.1508584112

van der Kant, R., Vriend, G. (2014). Alpha-bulges in G protein-coupled receptors. *International Journal of Molecular Sciences, 15*(5), 7841-7864.

Venkatakrishnan, A. J., Deupi, X., Lebon, G., Tate, C. G., Schertler, G. F., & Babu, M. M. (2013). Molecular signatures of G-protein-coupled receptors. *Nature, 494*(7436), 185-194.

Venkatakrishnan, A. J., Deupi, X., Lebon, G., Heydenreich, F.M., Flock, T., Miljus, T., Balaji, S., Bouvier, M., Veprintsev, D.B., Tate, C.G., Schertler, G.F., Babu, M.M. (2016). Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature, 536*(7617), 484-487.

Vilardaga, J.-P., Bünemann, M., Krasel, C., Castro, M., & Lohse, M. J. (2003). Measurement of the millisecond activation switch of G protein–coupled receptors in living cells. *Nat Biotechnol, 21*(7), 807-812. doi:10.1038/nbt838

Weigt, M., White, R.A., Szurmant, H., Hoch, J.A., Hwa, T. (2008). Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci USA, 106*(1), 67-72.

White, K. L., Eddy, M. T., Gao, Z. G., Han, G. W., Lian, T., Deary, A., . . . Stevens, R. C. (2018). Structural Connection between Activation Microswitch and Allosteric Sodium Site in GPCR Signaling. *Structure, 26*(2), 259-269 e255. doi:10.1016/j.str.2017.12.013

Ye, L., Van Eps, N., Zimmer, M., Ernst, O. P., & Prosser, R. S. (2016). Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature, 533*(7602), 265-268. doi:10.1038/nature17668

Zhang, Y., Sun, B., Feng, D., Hu, H., Chu, M., Qu, Q., . . . Skiniotis, G. (2017). Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein. *Nature, 546*, 248. doi:10.1038/nature22394
https://www.nature.com/articles/nature22394#supplementary-information

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, 19*(8), e50279.

Figure Captions

Figure 1. Comparison of the Potts model and residue contact frequency maps of class A, B and C GPCRs. (A) The active (red), inactive (blue), active agonist-PAM-bound (green) and inactive antagonist-NAM-bound (orange) structures of a model class A GPCR (PDBs: 6IBL, 5A8E, 6OIK and 6OBA). **(B)** Comparison of the Potts model (upper triangle) and residue contact frequency map (lower triangle) of class A GPCRs. **(C)** Structural contact detection rate of residue pairs in class A GPCRs with the top 100 long-distance Frobenius norms. **(D)** The active (red), inactive (blue), active agonist-PAM-bound (green) and inactive antagonist-NAM-bound (orange) structures of a model class B GPCR (PDBs: 6LMK, 5XEZ, 6VCB and 5EE7). **(E)** Comparison of the Potts model (upper triangle) and residue contact frequency map (lower triangle) of class B GPCRs. **(F)** Structural contact detection rate of residue pairs in class B GPCRs with the top 100 long-distance Frobenius norms. **(G)** The active (red), inactive (blue), active agonist-PAM-bound (green) and inactive antagonist-NAM-bound (orange) structures of a model class C GPCR (PDBs: 6N51, 6N52, 6UO8 and 7CUM). **(H)** Comparison of the Potts model (upper triangle) and residue contact frequency map (lower triangle) of class C GPCRs. **(I)** Structural contact detection rate of residue pairs in class C GPCRs with the top 100 long-distance Frobenius norms. The GPCR seven transmembrane (TM) helices, three intracellular loops (ICL) and three extracellular loops (ECL) are labeled. Only residue pairs with at least five residues apart were considered to calculate the contact detection rates using cutoffs of 0.0 and 0.3 for the contact frequency.

Figure 2. Residue contacts that are important for class A GPCR activation. **(A)** Comparison of residue contacts between the active and inactive structures of class A GPCRs. The residue contact frequency difference map was calculated by subtracting the residue contact frequencies in inactive structures from active structures. The residue contacts are colored by their differences in

contact frequency in a blue (-1.00) – white (0.00) – red (1.00) color scale. The GPCR seven transmembrane (TM) helices, three intracellular loops (ICL) and three extracellular loops (ECL), are labeled. **(B)** Exception residue contacts (rank 14, 28 and 60 in **Table S4**) between TM3-TM5 and TM5-TM6 that are changed during class A GPCR activation. **(C-I)** Residue contacts between TM1-TM7, TM2-TM7, TM3-TM6, TM3-TM7, TM5-TM6 and TM6-TM7 that showed the largest differences in contact frequencies between active and inactive class A GPCR structures. The exception, repacking and switching residue contacts are shown in underlined regular, bold and underlined bold fonts.

Figure 3. Residue contacts tuned by allosteric modulation in class A GPCRs. **(A-B)** Comparison of residue contacts between the agonist-PAM vs. agonist bound structures and antagonist-NAM vs antagonist bound structures of class A GPCRs. The residue contact frequency difference maps were calculated by subtracting the residue contact frequencies in structures without from structures with modulators**. (C-E)** Residue contacts between TM2-TM3, TM2-TM4, TM3-TM4, TM3-TM5, TM4-TM5, TM5-TM6 and TM6-TM7 that change upon binding of PAMs to the active agonist-bound class A GPCR structures. **(F-H)** Residue contacts between TM2-TM3, TM2-TM4, TM2-TM6, TM3-TM4, TM3-TM5, TM4-TM5, TM5-TM6 and TM6-TM7 that change upon binding of NAMs to the inactive antagonist-bound class A GPCR structure. Residue contacts that are strengthened and weakened are highlighted in green and orange, respectively. The exception residue contact is shown in underlined regular font. The GPCR seven transmembrane (TM) helices, three intracellular loops (ICL) and three extracellular loops (ECL), are labeled.

Figure 4. Residue contacts that are important for class B GPCR activation. **(A)** Comparison of residue contacts between the active and inactive structures of class B GPCRs. The residue contact frequency difference map was calculated by subtracting the residue contact frequencies in

inactive structures from active structures. The residue contacts are colored by their differences in contact frequency in a blue (-1.00) – white (0.00) – red (1.00) color scale. The GPCR seven transmembrane (TM) helices, three intracellular loops (ICL) and three extracellular loops (ECL), are labeled. **(B)** The exception residue contact (rank 2 in **Table S5**) between TM2-TM4 that are changed during class B GPCR activation. **(C-I)** Residue contacts between TM3-TM6, TM5-TM6 and TM6-TM7 that showed the largest differences in contact frequencies between active and inactive class B GPCR structures. The exception, repacking and switching residue contacts are shown in underlined regular, bold and underlined bold fonts.

Figure 5. Residue contacts tuned by allosteric modulation in class B GPCRs. **(A-B)** Comparison of residue contacts between the agonist-PAM vs agonist bound structures and antagonist-NAM and antagonist bound structures of class B GPCRs. The residue contact frequency difference maps were calculated by subtracting the residue contact frequencies in structures without from structures with modulators**. (C-E)** Residue contacts between TM3-TM6, TM5-TM6 and TM6-TM7 that change upon binding of PAMs to the active agonist-bound class B GPCR structures. **(F-H)** Residue contacts between TM3-TM6, TM5-TM6 and TM6-TM7 that change upon binding of NAMs to the inactive antagonist-bound class B GPCR structures. Residue contacts that are strengthened and weakened are highlighted in green and orange, respectively. The GPCR seven transmembrane (TM) helices, three intracellular loops (ICL) and three extracellular loops (ECL) are labeled.

Figure 6. Residue contacts that are important for dimerization and activation of class C GPCRs. **(A)** Comparison of residue contacts between the active and inactive structures of class C GPCRs. The residue contact frequency difference map was calculated by subtracting the residue contact frequencies in inactive structures from active structures. The residue contacts are colored

by their differences in contact frequency in a blue (-1.00) – white (0.00) – red (1.00) color scale. The GPCR seven transmembrane (TM) helices, three intracellular loops (ICL) and three extracellular loops (ECL), are labeled. **(B)** The metabotropic glutamate receptor 5 (PDB: 6N51) was used as a model class C GPCR. The closest distance between any side-chain heavy atoms in the residues of different receptor subunits that is ≤6.5Å is labeled. *FB* is the Frobenius norms calculated from the MSA by Potts model. **(C-H)** Residue contacts between TM2-TM3, TM2-TM4, TM2-TM7, TM3-TM4, TM3-TM5, TM3-TM7, TM4-TM5 and TM5-TM6 that showed the largest differences in contact frequencies between active and inactive class C GPCR structures. The repacking and switching residue contacts are shown in bold and underlined bold fonts.

Figure 7. Residue contacts tuned by allosteric modulation in class C GPCRs. **(A-B)** Comparison of residue contacts between the agonist-PAM vs agonist bound structures and antagonist-NAM and antagonist bound structures of class C GPCRs. The residue contact frequency difference maps were calculated by subtracting the residue contact frequencies in structures without from structures with modulators**. (C-E)** Residue contacts between TM1-TM7, TM3-TM5 and TM5-TM6 that change upon binding of PAMs to the active agonist-bound class C GPCR structures. **(F-H)** Residue contacts between TM1-TM7, TM2-TM7 and TM3-TM6 that change upon binding of NAMs to the inactive antagonist-bound class C GPCR structures. Residue contacts that are strengthened and weakened are highlighted in green and orange, respectively. The GPCR seven transmembrane (TM) helices, three intracellular loops (ICL) and three extracellular loops (ECL) are labeled.

Figure 8. Pathways of important residue contacts for activation of (A) class A, (B) class B and (C) class C GPCRs. The conserved CWxP, PIF, DRY and NPxxY functional motifs in class A and the PxxG and HETx motifs in class B GPCRs are labeled. Contacts that are important for

activation and inactivation are colored red and blue, respectively. The line thickness is proportional to the magnitude of differences in residue contact frequencies between the inactive and active GPCR structures (**Tables S7, S9** and **S10**).

Figure 9. Summary of residue contacts tuned by allosteric modulation in class A, B and C GPCRs. (**A-C**) Residue contacts that are tuned upon binding of PAMs in (A) class A, (B) class B and (C) class C GPCRs. (**D-F**) Residue contacts that are tuned upon binding of NAMs in (D) class A, (E) class B and (F) class C GPCRs. The conserved CWxP, PIF, DRY and NPxxY functional motifs in class A and the PxxG and HETx motifs in class B GPCRs are labeled. Contacts that are formed in the presence and absence of allosteric modulators are colored green and orange, respectively. The line thickness is proportional to the magnitude of differences in residue contact frequencies between the modulator-free and modulator-bound GPCR structures (**Tables S7, S9** and **S10**).

Figure 1

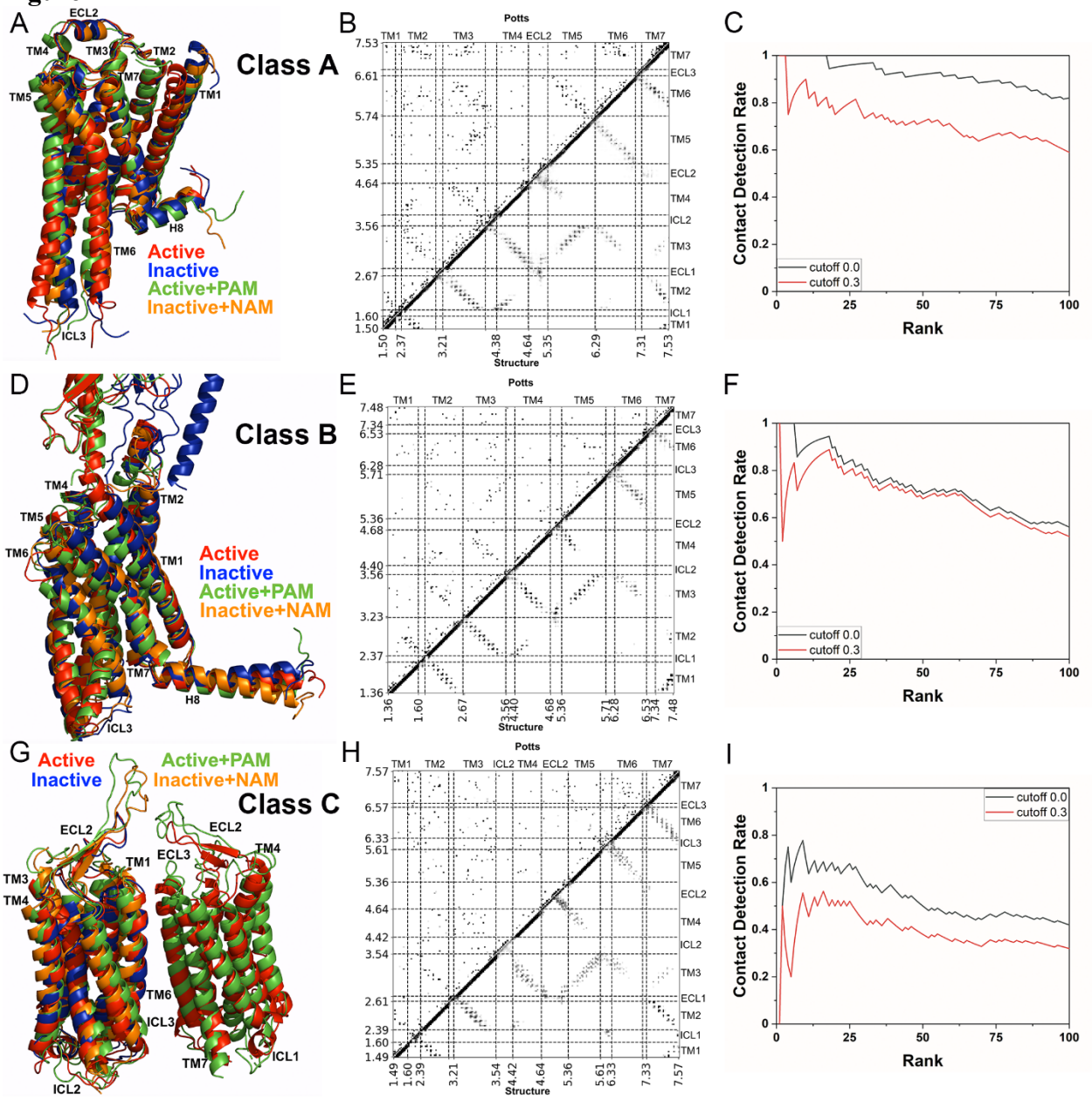

A
Class A
ECL2
TM4
TM3
TM2
TM7
TM5
TM1
H8
TM6
ICL3
Active
Inactive
Active+PAM
Inactive+NAM
B
Potts
TM1 TM2 TM3 TM4 ECL2 TM5 TM6 TM7
Structure
C
Contact Detection Rate
Rank
cutoff 0.0
cutoff 0.3
D
Class B
TM4
TM2
TM5
TM6
TM1
TM7
H8
ICL3
Active
Inactive
Active+PAM
Inactive+NAM
E
Potts
TM1 TM2 TM3 TM4 TM5 TM6 TM7
Structure
F
Contact Detection Rate
Rank
cutoff 0.0
cutoff 0.3
G
Class C
Active
Inactive
Active+PAM
Inactive+NAM
ECL2
TM1
ECL3
TM4
TM3
TM4
TM6
ICL3
TM7
ICL1
ICL2
H
Potts
TM1 TM2 TM3 ICL2 TM4 ECL2 TM5 TM6 TM7
Structure
I
Contact Detection Rate
Rank
cutoff 0.0
cutoff 0.3

Figure 2

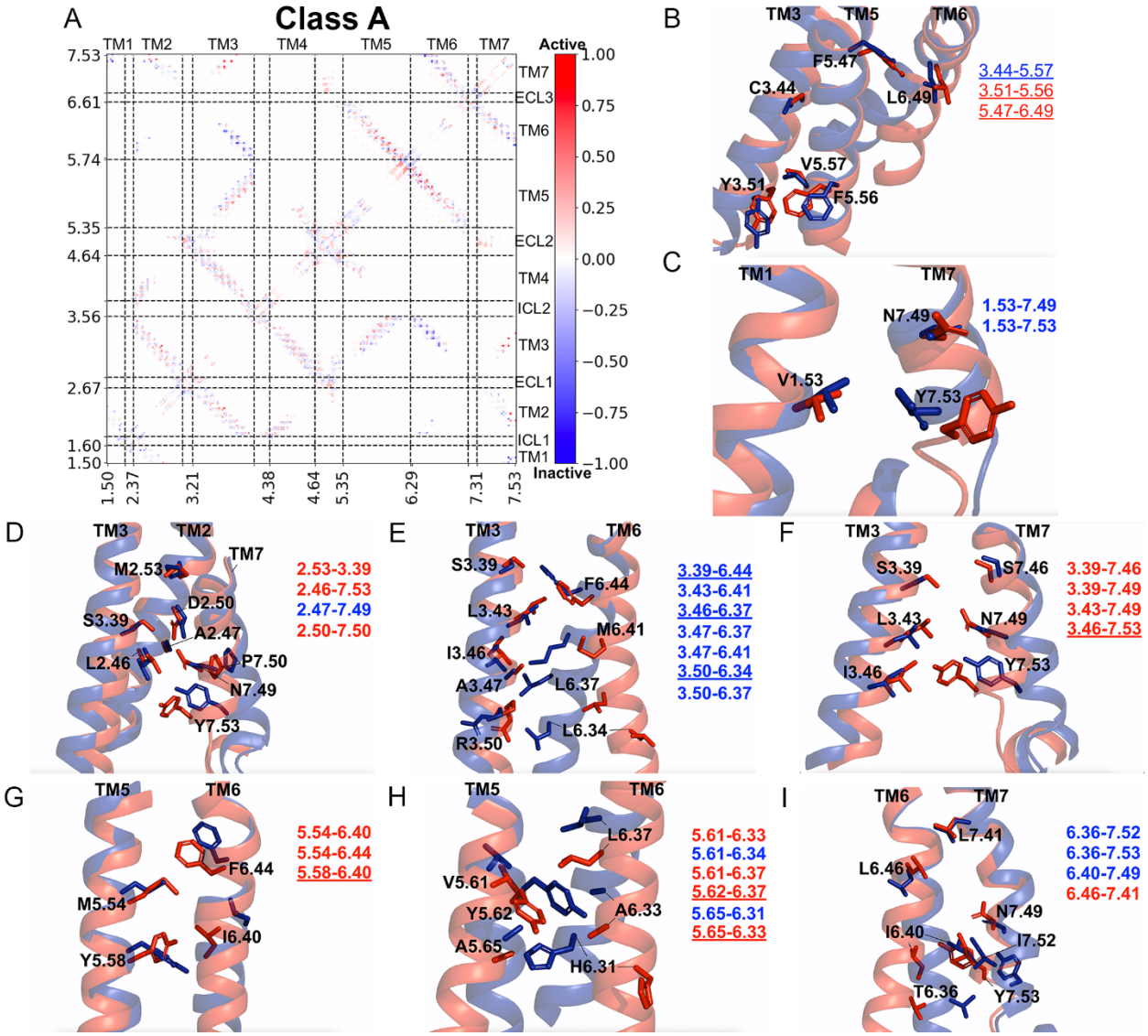

A
Class A
TM1 TM2 TM3 TM4 TM5 TM6 TM7
7.53
6.61
5.74
5.35
4.64
3.56
2.67
1.60
1.50
1.50 2.37 3.21 4.38 4.64 5.35 6.29 7.31 7.53
TM7
ECL3
TM6
TM5
ECL2
TM4
ICL2
TM3
ECL1
TM2
ICL1
TM1
Active
Inactive
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
B
TM3 TM5 TM6
F5.47
C3.44
L6.49
V5.57
Y3.51
F5.56
3.44-5.57
3.51-5.56
5.47-6.49
C
TM1 TM7
N7.49
V1.53
Y7.53
1.53-7.49
1.53-7.53
D
TM3 TM2 TM7
M2.53
D2.50
S3.39
A2.47
L2.46
P7.50
N7.49
Y7.53
2.53-3.39
2.46-7.53
2.47-7.49
2.50-7.50
E
TM3 TM6
S3.39
F6.44
L3.43
M6.41
I3.46
A3.47
L6.37
R3.50
L6.34
3.39-6.44
3.43-6.41
3.46-6.37
3.47-6.37
3.47-6.41
3.50-6.34
3.50-6.37
F
TM3 TM7
S3.39
S7.46
L3.43
N7.49
I3.46
Y7.53
3.39-7.46
3.39-7.49
3.43-7.49
3.46-7.53
G
TM5 TM6
F6.44
M5.54
I6.40
Y5.58
5.54-6.40
5.54-6.44
5.58-6.40
H
TM5 TM6
L6.37
V5.61
Y5.62
A6.33
A5.65
H6.31
5.61-6.33
5.61-6.34
5.61-6.37
5.62-6.37
5.65-6.31
5.65-6.33
I
TM6 TM7
L7.41
L6.46
N7.49
I6.40
I7.52
T6.36
Y7.53
6.36-7.52
6.36-7.53
6.40-7.49
6.46-7.41

Figure 3

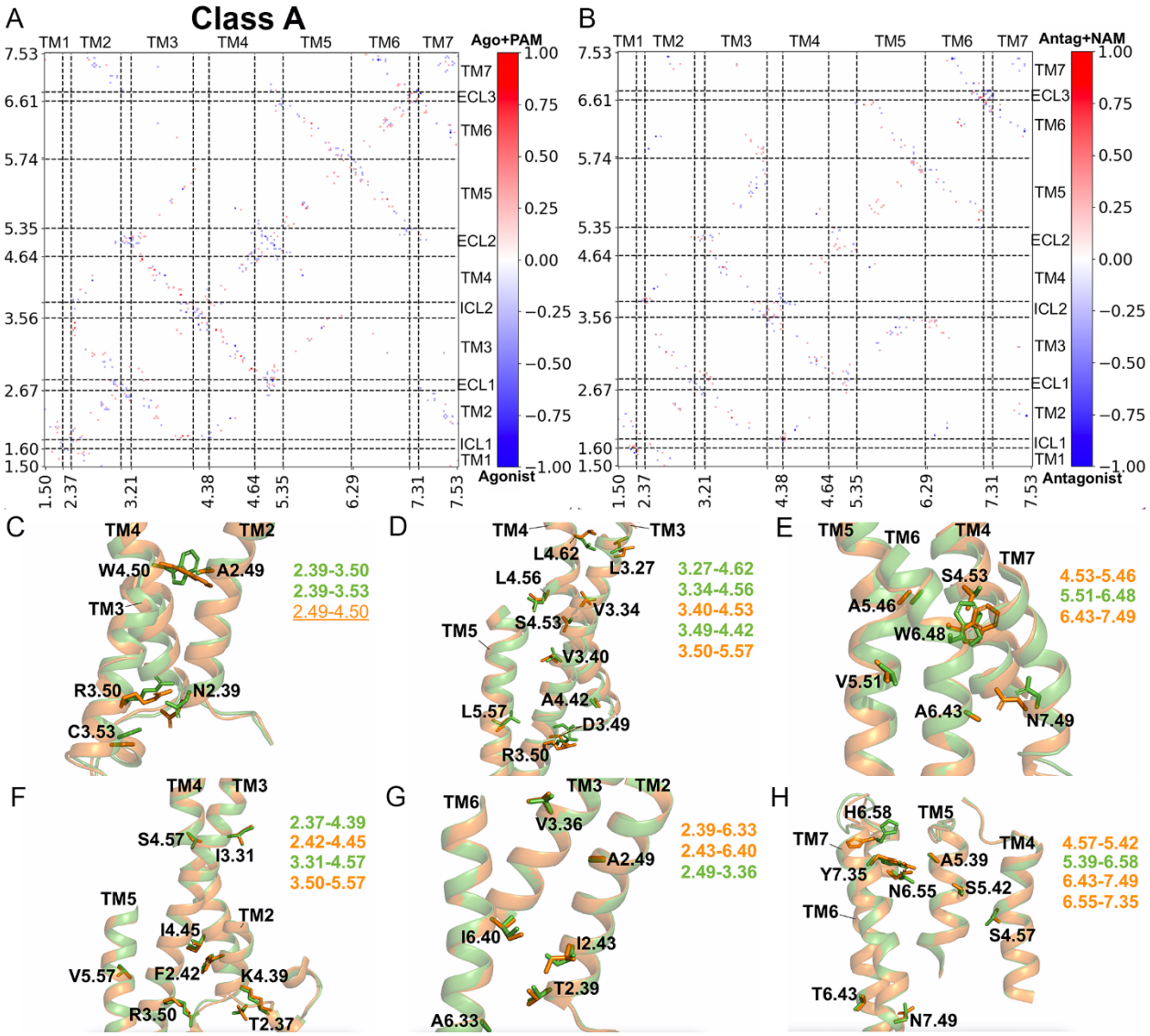

A
Class A
Ago+PAM
Agonist
B
Antag+NAM
Antagonist
TM1 TM2 TM3 TM4 TM5 TM6 TM7
ECL3 ECL2 ECL1 ICL2 ICL1
7.53 6.61 5.74 5.35 4.64 3.56 2.67 1.60 1.50
1.50 2.37 3.21 4.38 4.64 5.35 6.29 7.31 7.53
1.00 0.75 0.50 0.25 0.00 −0.25 −0.50 −0.75 −1.00
C
TM4 TM2 W4.50 A2.49 TM3 R3.50 N2.39 C3.53
2.39-3.50
2.39-3.53
2.49-4.50
D
TM4 TM3 L4.62 L3.27 L4.56 V3.34 S4.53 TM5 V3.40 A4.42 L5.57 D3.49 R3.50
3.27-4.62
3.34-4.56
3.40-4.53
3.49-4.42
3.50-5.57
E
TM5 TM6 TM4 TM7 S4.53 A5.46 W6.48 V5.51 A6.43 N7.49
4.53-5.46
5.51-6.48
6.43-7.49
F
TM4 TM3 S4.57 I3.31 TM5 TM2 I4.45 V5.57 F2.42 K4.39 R3.50 T2.37
2.37-4.39
2.42-4.45
3.31-4.57
3.50-5.57
G
TM6 TM3 TM2 V3.36 A2.49 I6.40 I2.43 T2.39 A6.33
2.39-6.33
2.43-6.40
2.49-3.36
H
H6.58 TM5 TM7 TM4 A5.39 Y7.35 N6.55 S5.42 TM6 S4.57 T6.43 N7.49
4.57-5.42
5.39-6.58
6.43-7.49
6.55-7.35

Figure 4

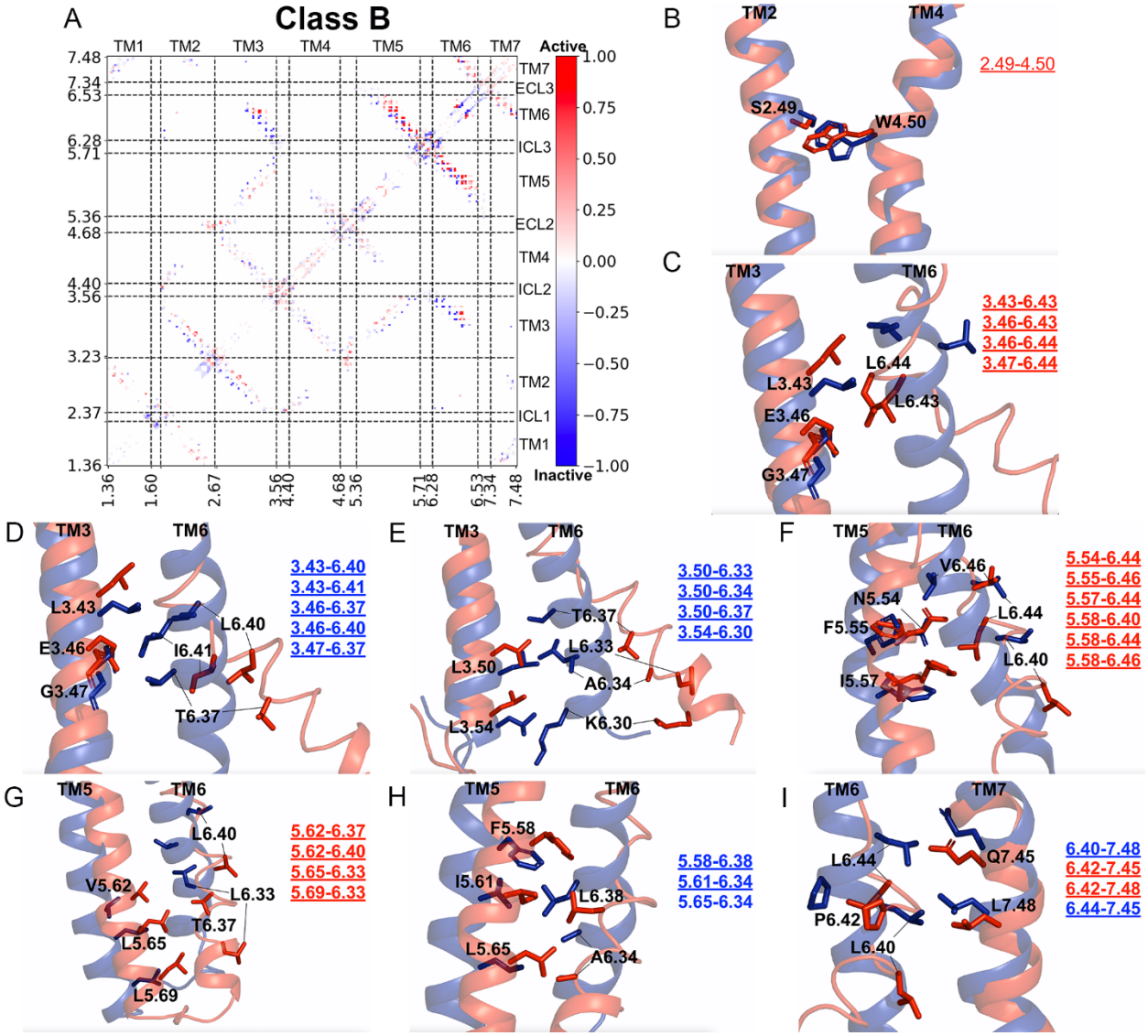

A
Class B
TM1 TM2 TM3 TM4 TM5 TM6 TM7
Active
Inactive
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
B
TM2
TM4
2.49-4.50
S2.49
W4.50
C
TM3
TM6
3.43-6.43
3.46-6.43
3.46-6.44
3.47-6.44
L3.43
L6.44
L6.43
E3.46
G3.47
D
TM3
TM6
3.43-6.40
3.43-6.41
3.46-6.37
3.46-6.40
3.47-6.37
L3.43
L6.40
E3.46
I6.41
G3.47
T6.37
E
TM3
TM6
3.50-6.33
3.50-6.34
3.50-6.37
3.54-6.30
T6.37
L6.33
L3.50
A6.34
L3.54
K6.30
F
TM5
TM6
5.54-6.44
5.55-6.46
5.57-6.44
5.58-6.40
5.58-6.44
5.58-6.46
V6.46
N5.54
L6.44
F5.55
L6.40
I5.57
G
TM5
TM6
5.62-6.37
5.62-6.40
5.65-6.33
5.69-6.33
L6.40
V5.62
L6.33
T6.37
L5.65
L5.69
H
TM5
TM6
5.58-6.38
5.61-6.34
5.65-6.34
F5.58
I5.61
L6.38
L5.65
A6.34
I
TM6
TM7
6.40-7.48
6.42-7.45
6.42-7.48
6.44-7.45
L6.44
Q7.45
P6.42
L7.48
L6.40

Figure 5

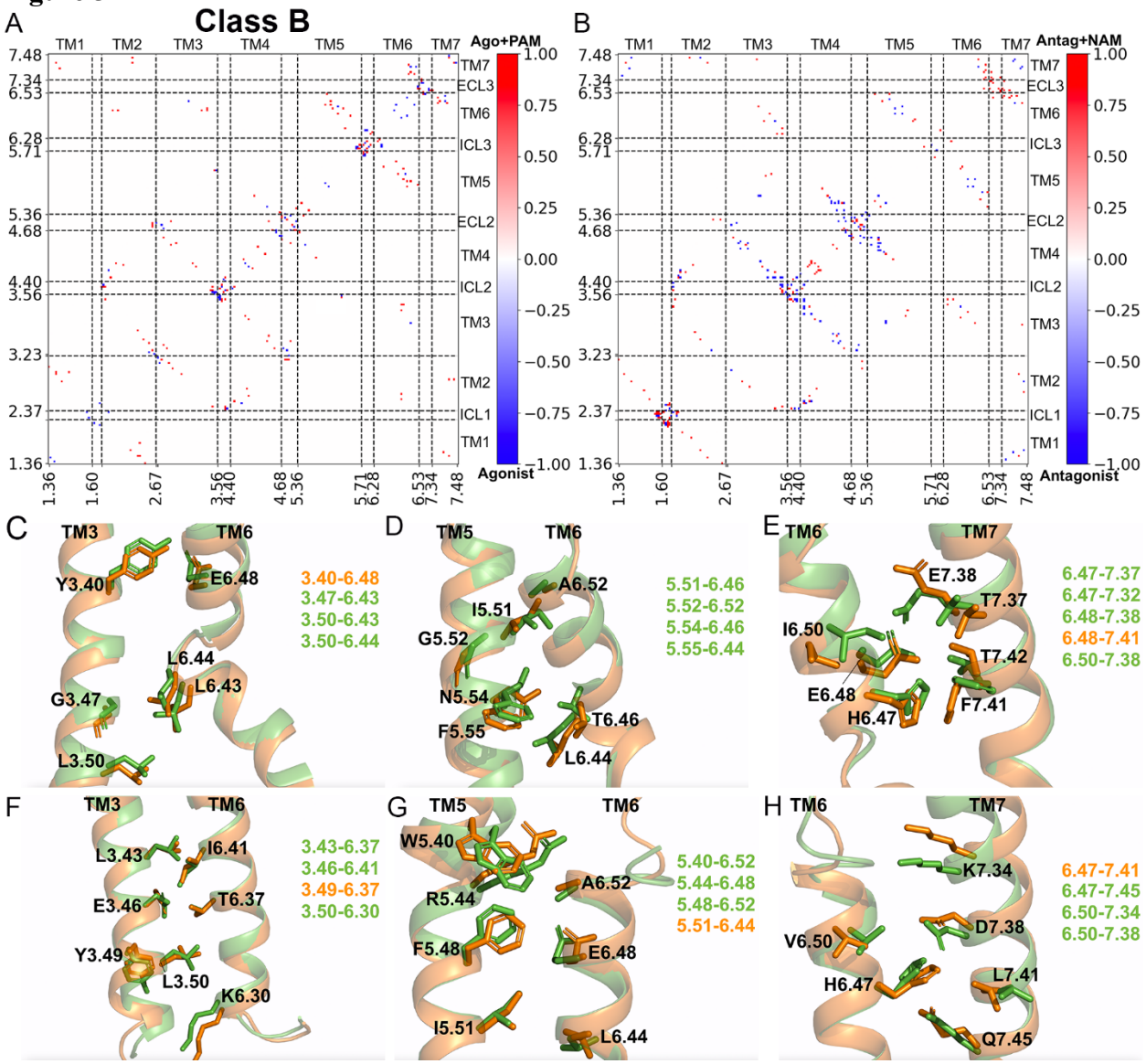

A
Class B
Ago+PAM
Agonist
B
Antag+NAM
Antagonist
TM1 TM2 TM3 TM4 TM5 TM6 TM7
ECL3 ICL3 ECL2 ICL2 ICL1
1.00 0.75 0.50 0.25 0.00 −0.25 −0.50 −0.75 −1.00
C
TM3 TM6
Y3.40 E6.48 L6.44 L6.43 G3.47 L3.50
3.40-6.48
3.47-6.43
3.50-6.43
3.50-6.44
D
TM5 TM6
A6.52 I5.51 G5.52 N5.54 F5.55 T6.46 L6.44
5.51-6.46
5.52-6.52
5.54-6.46
5.55-6.44
E
TM6 TM7
E7.38 T7.37 I6.50 T7.42 E6.48 H6.47 F7.41
6.47-7.37
6.47-7.32
6.48-7.38
6.48-7.41
6.50-7.38
F
TM3 TM6
L3.43 I6.41 E3.46 T6.37 Y3.49 L3.50 K6.30
3.43-6.37
3.46-6.41
3.49-6.37
3.50-6.30
G
TM5 TM6
W5.40 R5.44 A6.52 F5.48 E6.48 I5.51 L6.44
5.40-6.52
5.44-6.48
5.48-6.52
5.51-6.44
H
TM6 TM7
K7.34 D7.38 V6.50 H6.47 L7.41 Q7.45
6.47-7.41
6.47-7.45
6.50-7.34
6.50-7.38

Figure 6

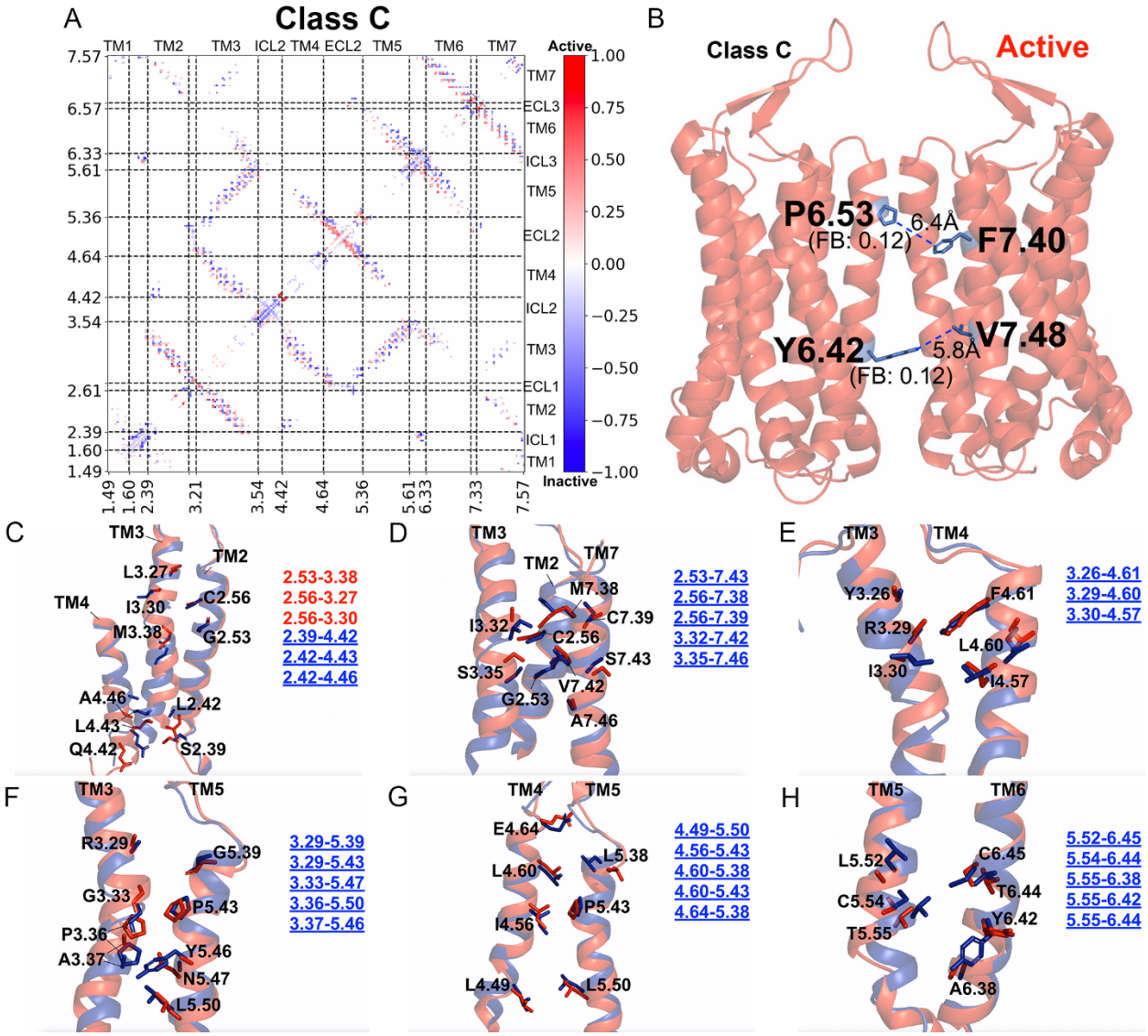

A
Class C
TM1 TM2 TM3 ICL2 TM4 ECL2 TM5 TM6 TM7
7.57 6.57 6.33 5.61 5.36 4.64 4.42 3.54 2.61 2.39 1.60 1.49
TM7 ECL3 TM6 ICL3 TM5 ECL2 TM4 ICL2 TM3 ECL1 TM2 ICL1 TM1
1.49 1.60 2.39 3.21 3.54 4.42 4.64 5.36 5.61 6.33 7.33 7.57
Active
Inactive
1.00 0.75 0.50 0.25 0.00 −0.25 −0.50 −0.75 −1.00
B
Class C
Active
P6.53
(FB: 0.12)
6.4Å
F7.40
Y6.42
(FB: 0.12)
5.8Å
V7.48
C
TM3 TM2 L3.27 TM4 I3.30 C2.56 M3.38 G2.53 A4.46 L2.42 L4.43 Q4.42 S2.39
2.53-3.38
2.56-3.27
2.56-3.30
2.39-4.42
2.42-4.43
2.42-4.46
D
TM3 TM2 TM7 M7.38 I3.32 C7.39 C2.56 S3.35 S7.43 V7.42 G2.53 A7.46
2.53-7.43
2.56-7.38
2.56-7.39
3.32-7.42
3.35-7.46
E
TM3 TM4 Y3.26 F4.61 R3.29 L4.60 I3.30 I4.57
3.26-4.61
3.29-4.60
3.30-4.57
F
TM3 TM5 R3.29 G5.39 G3.33 P5.43 P3.36 A3.37 Y5.46 N5.47 L5.50
3.29-5.39
3.29-5.43
3.33-5.47
3.36-5.50
3.37-5.46
G
TM4 TM5 E4.64 L5.38 L4.60 I4.56 P5.43 L4.49 L5.50
4.49-5.50
4.56-5.43
4.60-5.38
4.60-5.43
4.64-5.38
H
TM5 TM6 L5.52 C6.45 C5.54 T6.44 T5.55 Y6.42 A6.38
5.52-6.45
5.54-6.44
5.55-6.38
5.55-6.42
5.55-6.44

Figure 7

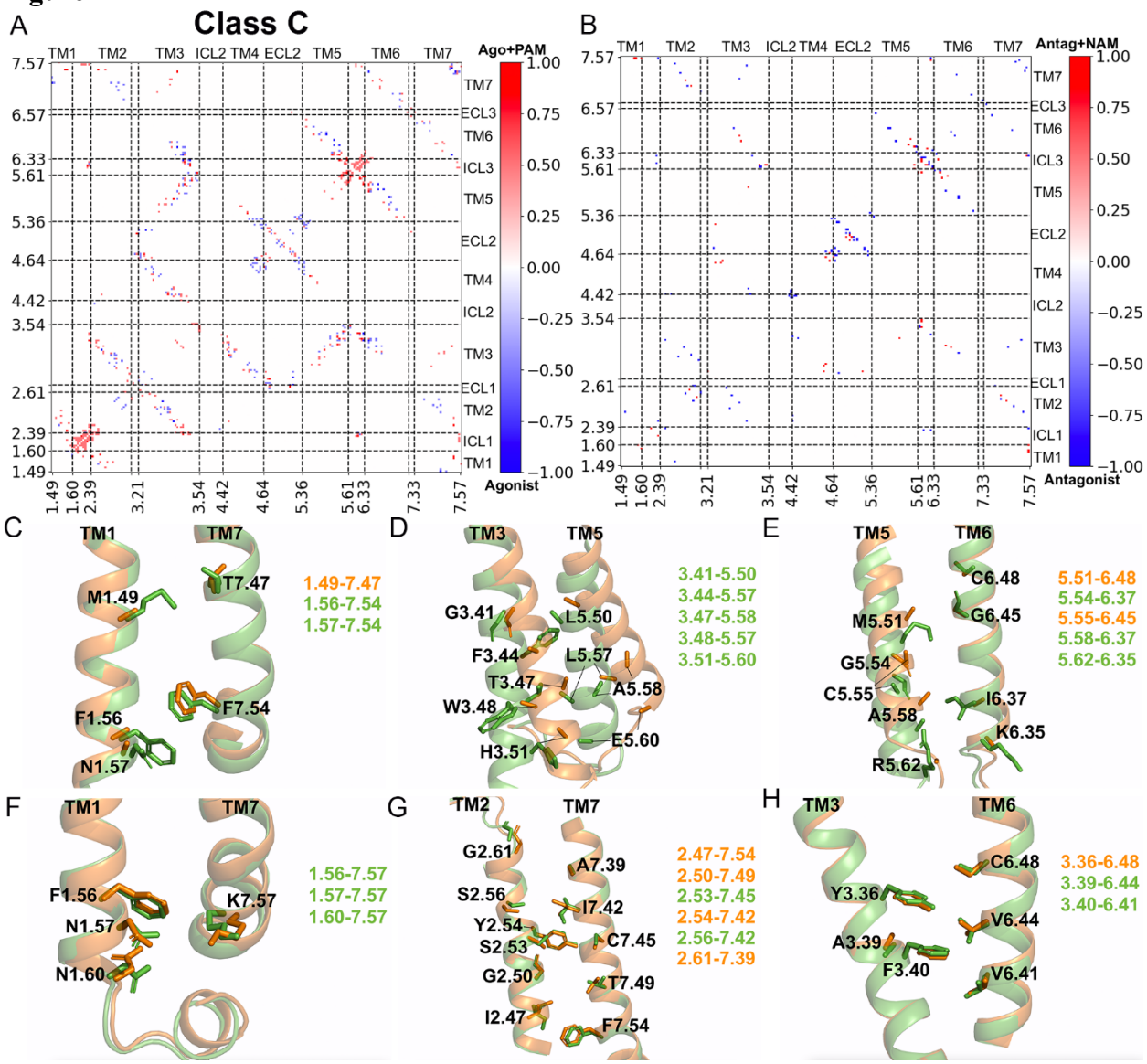

A
Class C
Ago+PAM
Agonist
B
Antag+NAM
Antagonist
C
M1.49
T7.47
F7.54
F1.56
N1.57
1.49-7.47
1.56-7.54
1.57-7.54
D
G3.41
L5.50
F3.44
L5.57
T3.47
A5.58
W3.48
E5.60
H3.51
3.41-5.50
3.44-5.57
3.47-5.58
3.48-5.57
3.51-5.60
E
C6.48
M5.51
G6.45
G5.54
C5.55
I6.37
A5.58
K6.35
R5.62
5.51-6.48
5.54-6.37
5.55-6.45
5.58-6.37
5.62-6.35
F
F1.56
K7.57
N1.57
N1.60
1.56-7.57
1.57-7.57
1.60-7.57
G
G2.61
A7.39
S2.56
I7.42
Y2.54
S2.53
C7.45
G2.50
T7.49
I2.47
F7.54
2.47-7.54
2.50-7.49
2.53-7.45
2.54-7.42
2.56-7.42
2.61-7.39
H
C6.48
Y3.36
V6.44
A3.39
F3.40
V6.41
3.36-6.48
3.39-6.44
3.40-6.41

Figure 8

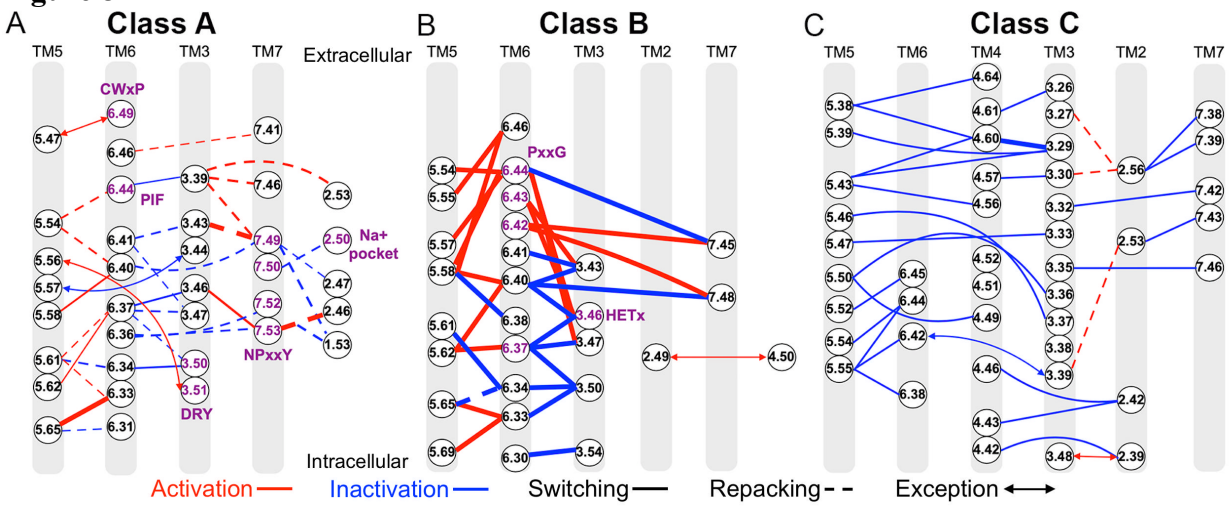
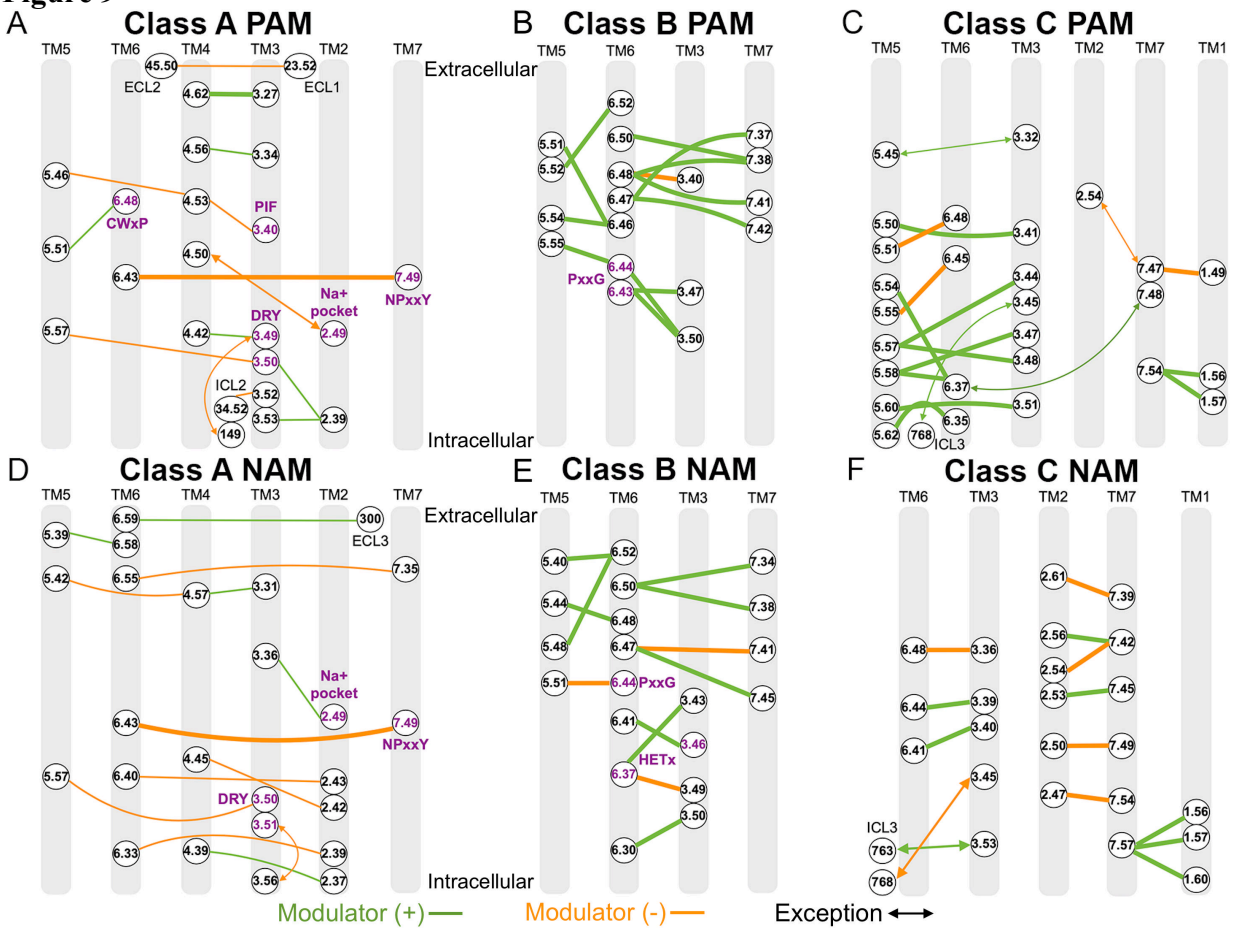

A
Class A
TM5
TM6
TM3
TM7
Extracellular
CWxP
PIF
Na+ pocket
NPxxY
DRY
Intracellular
B
Class B
TM5
TM6
TM3
TM2
TM7
PxxG
HETx
C
Class C
TM5
TM6
TM4
TM3
TM2
TM7
Activation
Inactivation
Switching
Repacking
Exception

Figure 9


A
Class A PAM
B
Class B PAM
C
Class C PAM
D
Class A NAM
E
Class B NAM
F
Class C NAM
Extracellular
Intracellular
ECL1
ECL2
ECL3
ICL2
ICL3
CWxP
PIF
Na+ pocket
NPxxY
DRY
PxxG
HETx
Modulator (+)
Modulator (-)
Exception