

RESEARCH ARTICLE



Unique features of different classes of G-protein-coupled receptors revealed from sequence coevolutionary and structural analysis

Hung N. Do¹ | Allan Haldane² | Ronald M. Levy³ | Yinglong Miao¹

¹The Center for Computational Biology and Department of Molecular Biosciences, The University of Kansas, Lawrence, Kansas, USA

²Department of Physics and Chemistry, Center for Biophysics and Computational Biology, Temple University, Philadelphia, Pennsylvania, USA

³Department of Chemistry, Center for Biophysics and Computational Biology, Temple University, Philadelphia, Pennsylvania, USA

Correspondence

Allan Haldane, Department of Physics, Center for Biophysics and Computational Biology, Temple University, Philadelphia, PA 19122, USA.

Email: allan.haldane@temple.edu

Yinglong Miao, The Center for Computational Biology and Department of Molecular Biosciences, The University of Kansas, Lawrence, KS 66047, USA.

Email: miao@ku.edu

Funding information

National Institute of General Medical Sciences, Grant/Award Numbers: R01GM132572, R35GM132090; National Institute on Aging, Grant/Award Number: 5U54-AG150472; National Science Foundation, Grant/Award Numbers: 193484, ACI-1548562; College of Liberal Arts and Sciences at the University of Kansas; American Heart Association, Grant/Award Number: 17SDG33370094; National Energy Research Scientific Computing Center, Grant/Award Number: M2874

Abstract

G-protein-coupled receptors (GPCRs) are the largest family of human membrane proteins and represent 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 and functional mechanisms of GPCRs remain poorly understood. Here, we have constructed novel sequence coevolutionary models of the A and B classes of GPCRs and compared them with residue contact frequency maps generated with available experimental structures. Significant portions of structural residue contacts were 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. Moreover, we identified distinct residue contacts involving different sets of functional motifs for GPCR activation, such as the Na⁺ pocket, CWxP, DRY, PIF, and NPxxY motifs in the class A and the HETx and PxxG motifs in the class B. Finally, we systematically uncovered critical residue contacts tuned by allosteric modulation in the two classes of GPCRs, including those from the activation motifs and particularly the extracellular and intracellular loops in class A GPCRs. These findings provide a promising framework for rational design of ligands to regulate GPCR activation and allosteric modulation.

KEYWORDS

activation, allosteric modulation, G-protein-coupled receptors, sequence coevolution, structural contacts

1 | INTRODUCTION

G-protein-coupled receptors (GPCRs) comprise the largest and most diverse family of integral membrane proteins in eukaryotes. GPCRs mediate various physiological activities, including vision, olfaction, taste, neurotransmission, endocrine, and immune responses.¹ Due to the critical roles in cellular signaling, approximately 34% of FDA-approved therapeutic agents act on GPCRs.² 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).^{3,4} 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.⁴ 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 and B classes of GPCRs. Class A GPCR activation is triggered by binding of an agonist to the receptor orthosteric pocket located within the 7TM domain.⁵ Upon agonist binding, the receptor intracellular end of TM6 moves outward to open up an intracellular cavity to accommodate and activate the G protein.^{6–9} 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.^{9–11}

Bioinformatics analysis has been previously carried out to identify important residue interactions for GPCR activation. Cvicsek et al. generated a structure-based alignment of 25 GPCRs that were extended to include TM sequences of all human GPCRs.¹² 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.¹² 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.¹³ 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.¹³

GPCRs are also able to bind allosteric ligands at topographically distinct sites, which could induce further conformational changes of the GPCRs.¹⁴ 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.^{15,16} 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.¹⁷ 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.^{18,19} The ICL2 adopted a short helical conformation and the TM5 was shifted along its helical axis toward the extracellular side relative to the TM4.¹⁸ For class B GPCRs, a PAM was found to bind between the extracellular domains of TM1 and TM2 of the GLP-1 receptor.²⁰ In the glucagon receptor, NAM binding restricted the outward movement of the TM6 intracellular domain.²¹ 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.²¹

Despite remarkable advances in structural determination efforts, experimental structures have been resolved for only ~90 unique GPCRs.^{3,4} 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.^{22–28} 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. The direct coupling analysis or “Potts” models have been shown to capture important structural and functional information.^{24,29–32} 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,^{33,34} or be used to predict structural or conformational preferences of individual sequences or subfamilies of sequences.^{35,36}

Here, we have constructed sequence coevolutionary Potts Hamiltonian models for class A and B GPCRs, for which sufficient protein sequences are available. We also generate residue contact frequency maps from available structures of both 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. Several of such exception residue contacts added important missing links for activation and allosteric modulation of the GPCRs. We have also identified distinct residue contacts that are important for activation and allosteric modulation of these two classes of GPCRs. The Ballesteros-Weinstein numbering scheme^{37,38} is used to denote the residue index of GPCRs. The most conserved residue in helix N is assigned N.50 and the others are numbered decreasingly toward the N-terminus and increasingly toward the C-terminus.

2 | METHODS

2.1 | Sequence datasets

We built MSAs of class A and B GPCRs from the corresponding Pfam IDs of PF00001 and PF00002, respectively. The MSAs included 84 481 sequences for class A and 17 804 sequences for class B of GPCRs (Table S1), and the sequences in both MSAs are mainly chordates, with minor amounts of invertebrates and uncategorized eukaryotes, so that the sequences in the MSAs encompass metazoan GPCR sequence diversity. 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 and 12 645 sequences of length 226 residues for class B of GPCRs. The residue sequences of class A and B GPCRs were also aligned using the *hmmalign* function of HMMER³⁹ (Table S2). A number of these sequences, however, were phylogenetically related and sampled with experimental biases.^{24,35,40} 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.^{24,35,40} The correction resulted in a number of effective sequences $N_{\text{eff}} = \sum w$ of 5126 sequences for class A and 902 sequences for class B of GPCRs (Table S1).

2.2 | Sequence coevolutionary Potts model inference

Potts Hamiltonian models of residue covariation were built based on the assumption that pairs of residues with a conserved functional interaction within a protein family or class will mutate in a compensatory and correlated fashion during evolution to maintain that function.⁴⁰ We built separate Potts models for each GPCR class because the classes can be interpreted to represent the main divisions of GPCR subfunction and mechanism, and we aimed to detect and differentiate the interacting pairs related to those mechanisms. Because our sequence dataset was composed of metazoan sequences, we expected the interactions detected by Potts covariation analysis will reflect functions and mechanisms which are conserved across metazoa within each class. The Potts model is constructed as the 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.^{40,41} In this equation, $H(S)$ is the Hamiltonian and takes the form of $H(S) = \sum_{i=1}^L h_i(s_i) + \sum_{i=1}^L \sum_{j=1}^L J_{ij}(s_i, s_j)$, 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 .^{40,41} Strong coupling parameters $J_{ij}(s_i, s_j)$ often correspond to direct physical interactions in protein 3D structures^{24,42}; 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.⁴³ Mi3-GPU can be used to construct Potts models with high statistical precision. A full description of the program can be found in Reference 43.

We applied Mi3-GPU to infer the Potts models of class A and B 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, 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 two²² walkers for four MCMC steps was performed to minimize the finite-sampling error and obtain a model with statistically accurate marginals and negligible residuals.⁴³

2.3 | Residue interaction score using weighted Frobenius norm

A weighted Frobenius norm⁴⁴ 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.⁴⁰ The interaction score can be calculated from the coupling parameters

$J_{\alpha\beta}^{ij}$ using the following formula $i^j = \sqrt{\sum_{\alpha\beta} (w_{\alpha\beta}^{ij} J_{\alpha\beta}^{ij})^2}$, where $w_{\alpha\beta}^{ij}$ is positive and tunable.^{40,43}

2.4 | Construction of residue contact frequency maps

From the Protein Data Bank (PDB)^{45,46} and GPCRdb,^{38,47,48} we collected 283 structures of class A GPCRs (94 active, 185 inactive, and four intermediate) and 31 structures of class B GPCRs (16 active, 14 inactive, and one intermediate) (Table S3). Structures that were poorly aligned to our MSAs were excluded from our datasets. For the remaining structures, the refined models obtained from GPCRdb^{38,47,48} were used for our analysis. GPCRdb used the GPCRdb homology modeling pipeline⁴⁷ to model missing segments, revert mutations and remodel distorted regions in the GPCR structures. The refined residues, however, were primarily located in the diverse loop regions. They did not significantly affect our analysis since most of the residues relevant to GPCR activation and allosteric modulation were identified in the TM domains. Among the active GPCR structures, three class A GPCRs were bound by both the agonist and PAM (PDBs: 4MQT, 6OIK and 6N48), similarly for one class B GPCR (PDB: 6VCB). Among the inactive GPCR structures, three class A GPCRs were bound by both the antagonist and NAM (PDBs: 5X7D, 6OBA, and 5T1A), similarly for one class B GPCR (PDB: 5EE7). A summary of the available GPCR structural data is included in Table S1.

The contact frequencies between conserved residues of class A and B GPCRs were determined. Two residues were considered in contact if their C β -C β atom distance was ≤ 8 Å, unless stated otherwise. For glycine, the C α atom was used. The residue contact frequency maps were built for classes A and B 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 and B GPCRs. Note that a number of these exception residue contacts could result from the arbitrary contact definition since they showed average C β -C β distances close to the 8 Å cutoff (Table S8).

Residue contact frequency difference maps were built for active versus inactive GPCR structures to identify residue contacts involved in activation of the two classes of GPCRs. The concepts of switching and repacking contacts for GPCR activation were adopted from Zhou et al.¹³ 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 two classes of GPCRs. In this context, structures of the same GPCR with very similar or identical orthosteric ligands were selected for comparison. In one instance, we compared residue contacts between two structures of the M2 muscarinic receptor with the same orthosteric ligand, but one with PAM bound (4MQT) and the other without PAM (4MQS).

3 | RESULTS

3.1 | Structural residue contacts of GPCRs were detected in sequence coevolutionary models

We inferred separate sequence-based Potts coevolutionary models for the A and B classes of GPCRs and compared them with

corresponding residue contact frequency maps (Figure 1). The input MSAs included 84 481 sequences for class A and 17 804 sequences for class B 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 and 902 effective sequences for class A and B GPCRs, respectively (Table S1). The residue sequences of class A and B GPCRs were also aligned using the *hmmalign* function of HMMER³⁹ (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 both classes as shown for a model class A GPCR in Figure 1(A) and a model class B GPCR in Figure 1(D). The Potts coupling parameters, which were used to calculate Frobenius norms as residue contact scores, were visualized as a heatmap for each of the two classes of GPCRs. The Potts model was

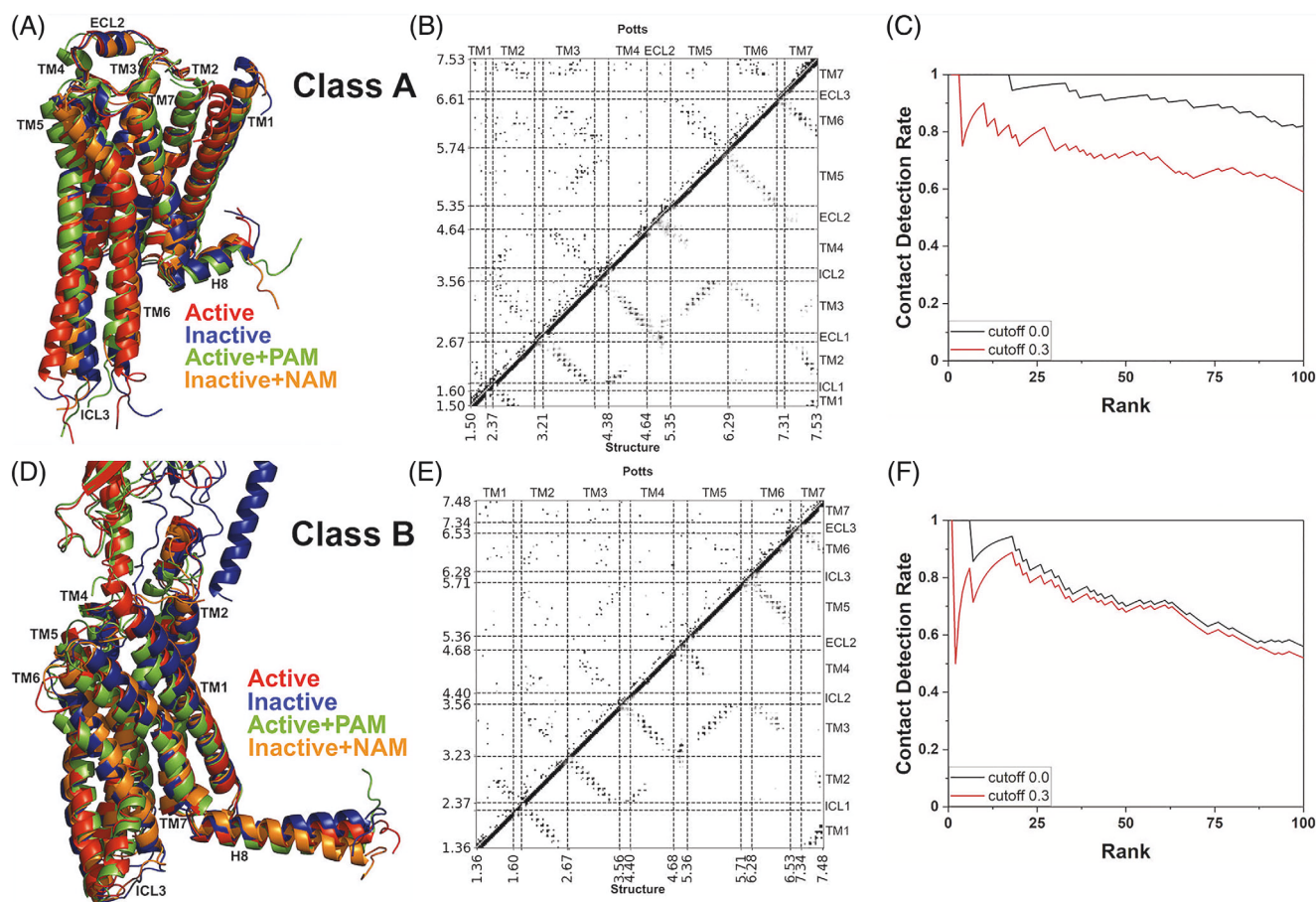


FIGURE 1 Comparison of the Potts model and residue contact frequency maps of class A and B G-protein-coupled receptors (GPCRs). (A) The active (red), inactive (blue), active agonist-positive allosteric modulator (PAM)-bound (green), and inactive antagonist-negative allosteric modulator (NAM)-bound (orange) 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. 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

then directly compared with the residue contact frequency map of class A (Figure 1(B)) and class B (Figure 1(E)) 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 1(B)). 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 1(E)).

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 and ≥ 0.19 for class B) 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 and B of GPCRs in the Figure 1(C,F), respectively. With zero contact frequency cutoff, the residue contact detection rates were 0.82 and 0.56 for the top 100 long-distance Frobenius norms in the Potts models of the classes A and B 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 and 0.52 for the top 100 long-distance Frobenius norms in the Potts models of the classes A and B of GPCRs, respectively. In comparison, the detection rate from Potts model of the class B GPCRs was slightly lower than that of class A GPCRs, due to the relatively smaller number of effective sequences (Table S1). These results are consistent with prior applications of Potts covariation analysis for other protein families, in which contact detection rates for the top 100 contacts observed in the main or predominant conformation for the protein family are typically 0.5–0.7, and this is dependent on the number of effective sequences in the MSA.^{44,49} It has also been established how the remaining Potts interactions, which do not correspond to contacts in the predominant conformation for a protein family, are often found to correspond to contacts formed in alternate conformations, homo-oligomer interfaces, and allosteric interactions upon further structural investigation.^{24,27} Below we will show how some of the predicted Potts interactions in GPCRs which are not contacts according to the 0.3 contact frequency cutoff can be explained structurally in other ways. Based on the above

findings, the 0.3 cutoff of structural contact frequency was used for further analysis.

3.2 | 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 2(A). The exception residue contacts of ranks 14, 28, and 60 are shown in Figure 2(B). Furthermore, we highlighted the top 30 residue contacts with the largest differences of contact frequencies in the inactive and active GPCR structures (Table S5(A)). 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 2(E)) 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 (Figure 2(F–H)). Overall, the residue contacts between TM1-TM7 (Figure 2(C)) and TM3-TM6 (Figure 2(E)) were significantly weakened, while a number of contacts between TM2-TM7 (Figure 2(D)), TM3-TM7 (Figure 2(F)), and TM5-TM6 (Figure 2(G,H)) 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 correlated with the 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 3(A)), similarly for the inactive antagonist-bound class A GPCRs in the absence and presence of NAMs (Figure 3(B)). The residue contacts that were significantly tuned in class A GPCRs in the presence of allosteric modulators were listed in Table S5(B). During binding of PAMs to active class A GPCRs, residues N2.39-R3.50 and

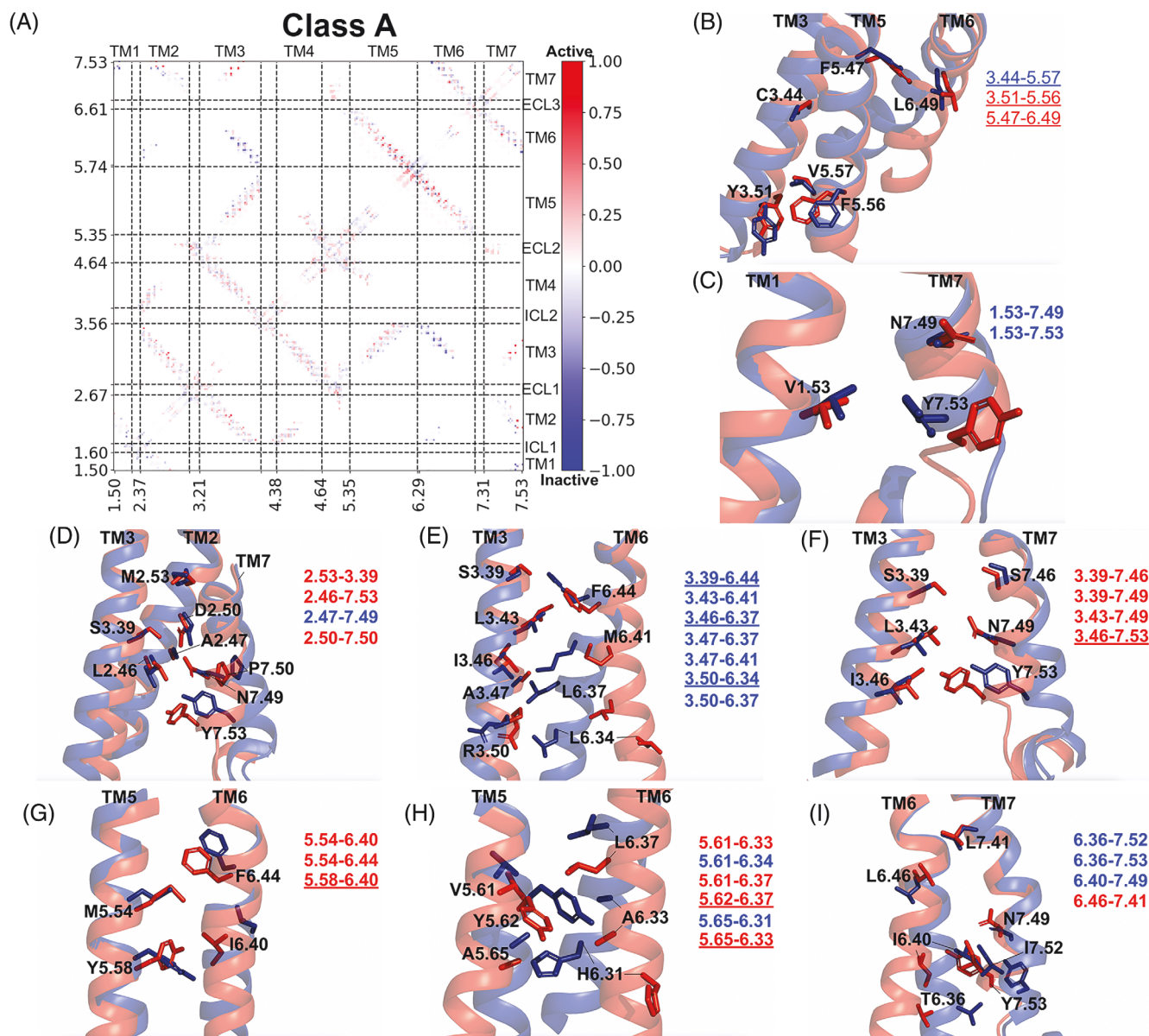


FIGURE 2 Residue contacts that are important for class A G-protein-coupled receptor (GPCR) activation. (A) Comparison of residue contacts between the active and inactive structures of class A GPCRs. The residue contact frequency difference map was calculated by subtracting the residue contact frequencies 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. (B) Exception residue contacts (ranks 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, respectively

N2.39-C3.53 formed new contacts, whereas the exception residue pair A2.49-W4.50 in the Potts model lost contact (Figure 3(C)). 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 (Figure 3(D,E)). With binding of NAMs to the inactive class A GPCRs, residues T2.37-K4.39 formed new contacts between TM2-TM4 (Figure 3(F)), similarly for residue contacts I3.31-S4.57 between TM3-TM4 (Figure 3(F)), contacts A2.49-V3.36 between TM2-TM3 (Figure 3(G)), and contacts A5.39-H6.58 between TM5-TM6 (Figure 3(H)).

3.3 | 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 S6. 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.

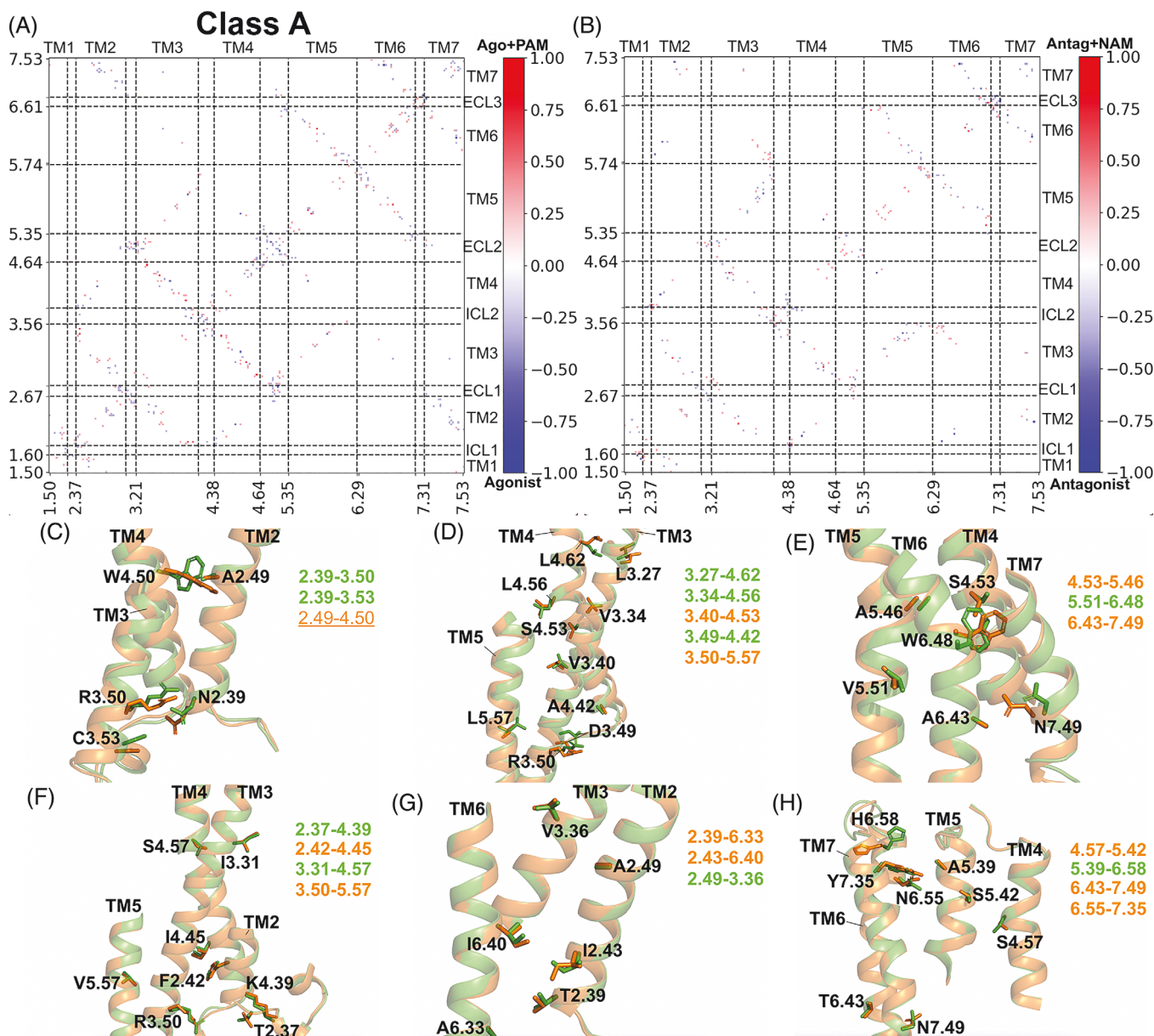


FIGURE 3 Residue contacts tuned by allosteric modulation in class A G-protein-coupled receptors (GPCRs). (A,B) Comparison of residue contacts between the agonist-positive allosteric modulator (PAM) versus agonist bound structures and antagonist-negative allosteric modulator (NAM) versus antagonist bound 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

We computed the residue contact frequency difference map between active and inactive class B GPCR structures as shown in Figure 4(A). The rank 2 exception residue contact is shown in Figure 4(B). We highlighted the top 30 residue contacts with the largest differences in contact frequencies between active and inactive class B GPCR structures (Table S7(A)). 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 (Figure 4(C,F,G)). 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 (Figure 4(D,E,H)).

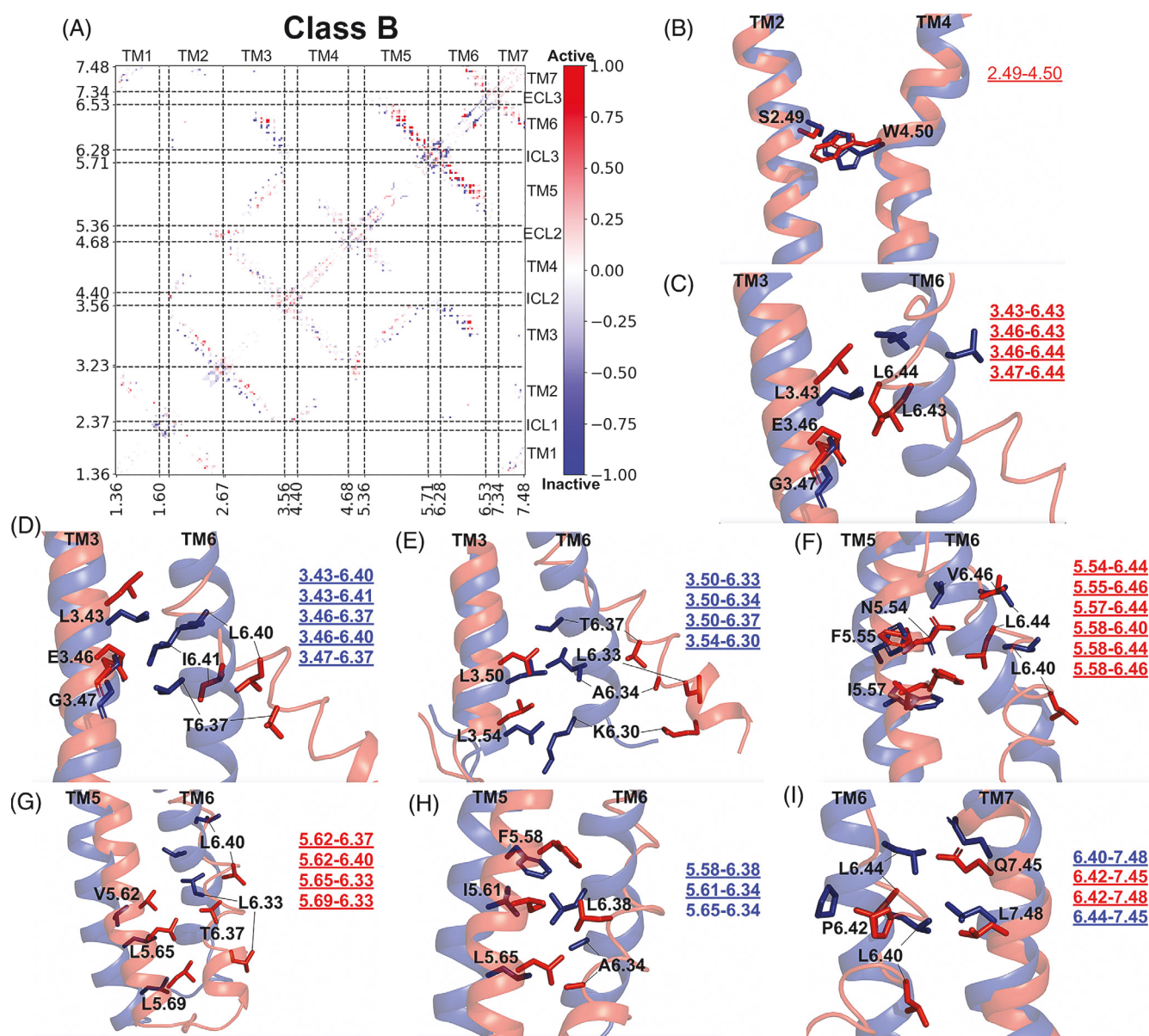


FIGURE 4 Residue contacts that are important for class B G-protein-coupled receptor (GPCR) activation. (A) Comparison of residue contacts between the active and inactive structures of class B GPCRs. The residue contact frequency difference map was calculated by subtracting the residue contact frequencies 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. (B) The exception residue contact (rank 2 in Table S6) 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, respectively

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 5(A)), similarly for the inactive antagonist-bound class B GPCRs in the absence and presence of NAMs (Figure 5(B)). The residue contacts that were significantly tuned in the presence of allosteric modulators were listed in Table S7(B). 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 5(C)), 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 5(D)) and residue contacts H6.47-T7.37, H6.47-T7.42, E6.48-E7.38, and I6.50-E7.38 between TM6-TM7 (Figure 5(F)). 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 5(G)), similarly for contacts W5.40-A6.52, R5.44-E6.48 and F5.48-A6.52 between TM5 and TM6 (Figure 5(H)) and contacts H6.47-Q7.45, V6.50-K7.34, and V6.50-D7.38 between TM6-TM7 (Figure 5(I)).

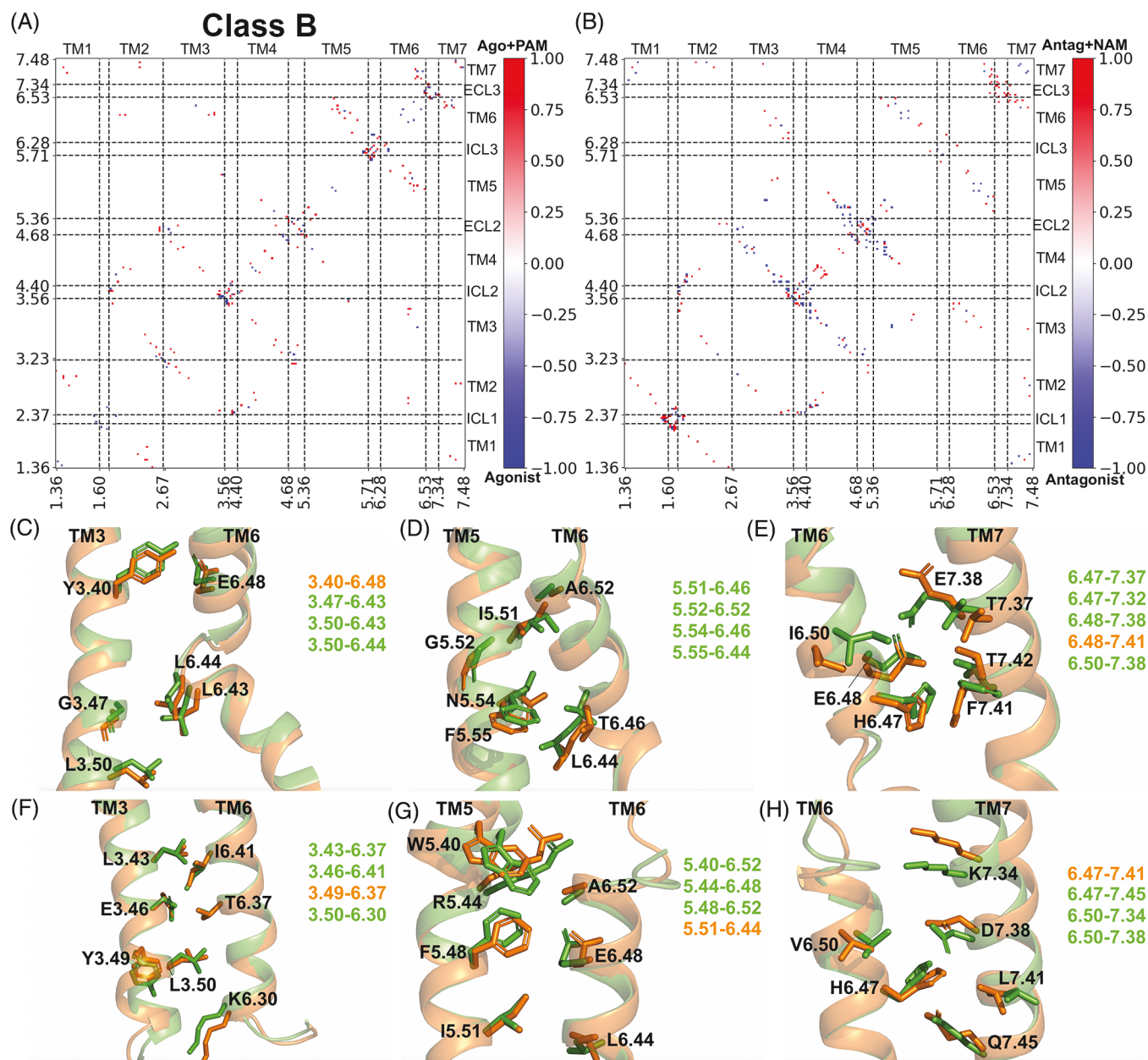


FIGURE 5 Residue contacts tuned by allosteric modulation in class B G-protein-coupled receptors (GPCRs). (A,B) Comparison of residue contacts between the agonist-positive allosteric modulator (PAM) versus agonist bound structures and antagonist-negative allosteric modulator (NAM) and antagonist bound 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

3.4 | “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 (Figure 6).^{33,34} Specifically, 6 and 1 of the top 20 exception contacts resulted from activation and/or allosteric modulation of the class A and B GPCRs, respectively (Tables S4 and S6). In class A GPCRs, exception contacts 3.51–5.56 and 5.47–6.49

augmented the residue network for receptor activation by involving the DRY and CWxP functional motifs,⁵⁰ respectively (Figure 6(A)). 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 (Figure 6(B,C)). Another exception contact between 2.49 (Na⁺ pocket) and 4.50 was important for both allosteric modulation of class A GPCRs^{51–54} and activation of class B GPCRs (Figure 6(B,D)). This was consistent with previous findings that W4.50 is fully conserved in class A and B GPCRs and its

structures.¹³ All these contacts resulted from the global movements of TM6 and TM7 intracellular ends upon activation of class A GPCRs.¹³ Our analysis also revealed 20 additional residue contacts involved in class A GPCR activation (Table S9). 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)^{54,56} and TM6. New residue contacts were formed between TM5-TM6 (involving the extracellular CWxP and PIF motifs) and TM3-TM7 (involving the NPxxY motif)⁵⁶ (Figure 6(A)). It is well established that the TM6 intracellular end moves outward, while the highly conserved NPxxY motif in the TM7 intracellular domain moves toward TM3,^{5,56–62} 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 outward.^{9,11,63} 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 6(D)).

3.6 | Comparison of allosteric modulation of class A and B GPCRs

Our sequence coevolutionary and structural contact analysis further revealed important residue contacts tuned by allosteric modulation of the A and B classes of GPCRs (Figure 6). 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 upon binding of allosteric modulators (Figure 6(B,C)). 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.^{15,64} Furthermore, residue contacts such as 4.53–5.46 and 5.51–6.48 were modified, due to shift of the TM5 toward the extracellular side relative to the TM4.¹⁸ PAM binding also introduced rearrangements of residue contacts in the three extracellular loops and ICL2 of the class A GPCRs (Figure 6(B)). On the other hand, NAM binding induced conformational changes in the TM2 (the Na⁺ binding pocket),^{51–54} TM3 (the DRY motif), TM4, TM5, TM6, and TM7 (the NPxxY motif) domains of the class A GPCRs (Figure 6(C)).^{17,19} For class B GPCRs, residues in the TM6, which is heavily involved in activation of the receptors,^{9,21} were also found to undergo substantial changes of contacts in the presence of allosteric modulators,

especially near the PxxG and HETx motifs (Figure 6(E,F)). Our study, for the first time, systematically identified residue contacts that are tuned by allosteric modulation in different classes of GPCRs. They provide a framework for rationally designing selective allosteric drugs to modulate the structure and function of GPCRs.

4 | DISCUSSION

We have revealed unique features of classes A and B GPCRs from combined sequence coevolutionary and structural contact analysis. We have found plausible explanations for 9 and 2 of the top 20 exception residue contacts in the coevolutionary models for class A and B GPCRs, respectively. They resulted 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 and S6). A certain number of these exceptions could possibly result from inaccuracy of the sequence coevolutionary models, especially for the class B 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.⁴

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. Moreover, only 31 structures of subclass B1 (secretin receptors) were available for class B GPCRs. More experimental structures of GPCRs, especially the class B2 for which the first structure has been obtained only very recently,⁶⁵ will strengthen the analysis. In addition, 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.¹⁴ In addition, more structures associated with new functional mechanisms of GPCRs (e.g., biased agonism⁶⁶) await to be determined. 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. The updated structural analysis will also allow us to identify more general functional mechanisms of GPCRs, notably for the allosteric modulation with currently very limited data.

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^{7,67–69} 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 C, D, E, and F GPCRs are too small to construct accurate sequence coevolutionary models,^{44,49} 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 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. The residue contact frequency map can then be compared with the sequence coevolutionary model to 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 coevolutionary model to study GPCR function. Notably, the coevolutionary model has additional uses in predicting the energetics of individual sequences and particular residue pairs. For instance, the coevolutionary 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.³⁵ The coevolutionary model can also be used to predict functional characteristics of individual residue–residue interactions.⁷⁰ 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.

ACKNOWLEDGMENTS

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. M. L.), NSF (193484 to R. M. L.), and the startup funding in the College of Liberal Arts and Sciences at the University of Kansas (to Y. M.).

CONFLICT OF INTERESTS

The authors declare no conflict of interests.

AUTHOR CONTRIBUTIONS

Allan Haldane, Ronald M. Levy, and Yinglong Miao: Designed the study. Hung N. Do: Performed all calculations. Hung N. Do, Allan Haldane, Ronald M. Levy, and Yinglong Miao: Analyzed data and wrote the manuscript.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/prot.26256>.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supplementary material of this article.

ORCID

Yinglong Miao  <https://orcid.org/0000-0003-3714-1395>

REFERENCES

- Venkatakrishnan AJ, Deupi X, Lebon G, Tate CG, Schertler GF, Babu MM. Molecular signatures of G-protein-coupled receptors. *Nature*. 2013;494:185–194. doi:10.1038/nature11896
- Hauser AS, Chavali S, Masuho I, et al. Pharmacogenomics of GPCR drug targets. *Cell*. 2018;172(1–2):41–54.e19. doi:10.1016/j.cell.2017.11.033
- Stevens RC, Cherezov V, Katritch V, et al. The GPCR network: a large-scale collaboration to determine human GPCR structure and function. *Nat Rev Drug Discov*. 2013;12:25–34.
- Isberg V, Mordalski S, Munk C, et al. GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res*. 2016;44:D365–D364.
- Ye L, Van Eps N, Zimmer M, Ernst OP, Prosser RS. Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature*. 2016;533:265–268. doi:10.1038/nature17668
- Nygaard R, Zou Y, Dror RO, et al. The dynamic process of beta(2)-adrenergic receptor activation. *Cell*. 2013;152:532–542. doi:10.1016/j.cell.2013.01.008
- Dror RO, Arlow DH, Maragakis P, et al. Activation mechanism of the β 2-adrenergic receptor. *Proc Natl Acad Sci U S A*. 2011;108:18684–18689.
- Miao Y, Nichols SE, Gasper PM, Metzger VT, McCammon JA. Activation and dynamic network of the M2 muscarinic receptor. *Proc Natl Acad Sci U S A*. 2013;110:10982–10987. doi:10.1073/pnas.1309755110
- Hilger D, Kumar KK, Hu H, et al. Structural insights into differences in G protein activation by family A and family B GPCRs. *Science*. 2020;369(6503):eaba3373. doi:10.1126/science.aba3373
- Zhang Y, Sun B, Feng D, et al. Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein. *Nature*. 2017;546(7657):248–253. doi:10.1038/nature22394
- Liang YL, Belousoff MJ, Zhao P, et al. Toward a structural understanding of class B GPCR peptide binding and activation. *Molecular Cell*. 2020;77(3):656–668.e5. doi:10.1016/j.molcel.2020.01.012
- Cvick V, Goddard WI, Abrol R. Structure-based sequence alignment of the transmembrane domains of all human GPCRs: phylogenetic, structural, and functional implications. *PLoS Comput Biol*. 2016;12:e1004805.
- Zhou Q, Yang D, Wu M, et al. Common activation mechanism of class A GPCRs. *Elife*. 2019;19:e50279.
- Thal DM, Glukhova A, Sexton PM, Christopoulos A. Structural insights into G-protein-coupled receptor allostery. *Nature*. 2018;559:45–53. doi:10.1038/s41586-018-0259-z

15. Kruse AC, Ring AM, Manglik A, et al. Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature*. 2013;504:101-106.
16. Maeda S, Qu Q, Robertson MJ, Skiniotis G, Kobilka BK. Structures of the M1 and M2 muscarinic acetylcholine receptor/G-protein complexes. *Science*. 2019;364:552-557. doi:10.1126/science.aaw5188
17. Maeda S, Xu J, Kadji FMN, et al. Structure and selectivity engineering of the M1 muscarinic receptor toxin complex. *Science*. 2020;369:161-167. doi:10.1126/science.aax2517
18. Lu J, Byrne N, Wang J, et al. Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40. *Nat Struct Mol Biol*. 2017;24:570-577.
19. Liu H, Kim HR, Deepak RNVK, et al. Orthosteric and allosteric action of the C5a receptor antagonists. *Nat Struct Mol Biol*. 2018;25:472-481. doi:10.1038/s41594-018-0067-z
20. Bueno AB, Sun B, Willard FS, et al. Structural insights into probe-dependent positive allostery of the GLP-1 receptor. *Nat Chem Biol*. 2020;16:1105-1110. doi:10.1038/s41589-020-0589-7
21. Jazayeri A, Doré AS, Lamb D, et al. Extra-helical binding site of a glucagon receptor antagonist. *Nature*. 2016;533:274-277.
22. Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol*. 2003;10:59-69. doi:10.1038/nsb881
23. Dima RI, Thirumalai D. Determination of network of residues that regulate allostery in protein families using sequence analysis. *Protein Sci*. 2006;15:258-268. doi:10.1110/ps.051767306
24. Morcos F, Pagnani A, Lunt B, et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci*. 2011;108:E1293-E1301. doi:10.1073/pnas.1111471108
25. Nichols SE, Hernandez CX, Wang Y, McCammon JA. Structure-based network analysis of an evolved G protein-coupled receptor homodimer interface. *Protein Sci*. 2013;22:745-754. doi:10.1002/pro.2258
26. Levy RM, Haldane A, Flynn WF. Potts Hamiltonian models of protein co-variation, free energy landscapes, and evolutionary fitness. *Curr Opin Struct Biol*. 2017;43:55-62. doi:10.1016/j.sbi.2016.11.004
27. Anishchenko I, Ovchinnikov S, Kamisetty H, Baker D. Origin of coevolution between residues distant in protein 3D structures. *Proc Natl Acad Sci U S A*. 2017;114:9122-9127.
28. Nicoludis JM, Gaudet R. Applications of sequence coevolution in membrane protein biochemistry. *Biochim Biophys Acta - Biomembr*. 2018;1860:895-908. doi:10.1016/j.bbmem.2017.10.004
29. Sulkowska JI, Morcos F, Weigt M, Hwa T, Onuchic J. Genomics-aided structure prediction. *Proc Natl Acad Sci U S A*. 2012;109:10340-10345.
30. Marks DS, Hopf TA, Sander C. Protein structure prediction from sequence variation. *Nat Biotechnol*. 2012;30:1072-1080.
31. Morcos F, Schafer NP, Cheng RR, Onuchic J, Wolynes PG. Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. *Proc Natl Acad Sci U S A*. 2014;111:12408-12413. doi:10.1073/pnas.1413575111
32. Ovchinnikov S, Kinch L, Park H, et al. Large-scale determination of previously unsolved protein structures using evolutionary information. *Elife*. 2015;4:e09248.
33. Figliuzzi M, Jacquier H, Schug A, Tenaillon O, Weigt M. Coevolutionary landscape inference and the context-dependence of mutations in beta-lactamase TEM-1. *Mol Biol Evol*. 2016;33. doi:10.1093/molbev/msv211
34. Hopf TA, Ingraham JB, Poelwijk FJ, et al. Mutation effects predicted from sequence co-variation. *Nature Biotechnology*. 2017;35(2):128-135. doi:10.1038/nbt.3769
35. Haldane A, Flynn WF, He P, Vijayan RSK, Levy RM. Structural propensities of kinase family proteins from a Potts model of residue co-variation. *Protein Sci*. 2016;25:1378-1384. doi:10.1002/pro.2954
36. Sutto L, Marsili S, Valencia A, Gervasio FL. From residue coevolution to protein conformational ensembles and functional dynamics. *Proc Natl Acad Sci U S A*. 2015;112:13567-13572. doi:10.1073/pnas.1508584112
37. Ballesteros J, Weinstein H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci*. 1995;25:366-428. doi:https://www.sciencedirect.com/science/article/abs/pii/S1043947105800497
38. Isberg V, de Graaf C, Bortolato A, et al. Generic GPCR residue numbers- aligning topology maps while minding the gaps. *Trends Pharmacol Sci*. 2015;36:22-31.
39. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol*. 2011;7:e1002195.
40. Haldane A, Flynn WF, He P, Levy RM. Coevolutionary landscape of kinase family proteins: sequence probabilities and functional motifs. *Biophys J*. 2018;114:21-31.
41. Levy RM, Haldane A, Flynn WF. Potts Hamiltonian models of protein covariation, free energy landscapes, and evolutionary fitness. *Curr Opin Struct Biol*. 2017;43:55-62.
42. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci U S A*. 2008;106:67-72.
43. Haldane A, Levy RM. Mi3-GPU: MCMC-based inverting ising inference on GPUs for protein covariation analysis. *Comput Phys Commun*. 2020;260:107312.
44. Ekeberg M, Lökvist C, Lan Y, Weigt M, Aurell E. Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. *Phys Rev E*. 2013;87:012707.
45. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28:235-242.
46. Burley SK, Berman HM, Bhikadiya C, et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res*. 2019;47:D464-D474.
47. Pándy-Szekeres G, Munk C, Tsonkov TM, et al. GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res*. 2018;46:D440-D446.
48. van der Kant R, Vriend G. Alpha-bulges in G protein-coupled receptors. *Int J Mol Sci*. 2014;15:7841-7864.
49. Haldane A, Levy RM. Influence of multiple-sequence-alignment depth on Potts statistical models of protein covariation. *Phys Rev*. 2019;99:032405.
50. Gao ZG, Kim SK, Gross AS, Chen A, Blaustein JB, Jacobson KA. Identification of essential residues involved in the allosteric modulation of the human A(3) adenosine receptor. *Mol Pharmacol*. 2003;63:1021-1031. doi:10.1124/mol.63.5.1021
51. Gutiérrez-de-Terán H, Massink A, Rodríguez D, et al. The role of a sodium ion binding site in the allosteric modulation of the A2A adenosine G protein-coupled receptor. *Structure*. 2013;21(12):2175-2185. doi:10.1016/j.str.2013.09.020
52. Miao Y, Caliman AD, McCammon JA. Allosteric effects of sodium ion binding on activation of the M3 muscarinic G-protein coupled receptor. *Biophys J*. 2015;108:1796-1806. doi:10.1016/j.bpj.2015.03.003
53. Shang Y, LeRouzic V, Schneider S, Bisignano P, Pasternak GW, Filizola M. Mechanistic insights into the allosteric modulation of opioid receptors by sodium ions. *Biochemistry*. 2014;53:5140-5149. doi:10.1021/Bi5006915
54. White KL, Eddy MT, Gao ZG, et al. Structural connection between activation microswitch and allosteric sodium site in GPCR signaling. *Structure*. 2018;26(2):259-269.e5. doi:10.1016/j.str.2017.12.013
55. Peeters MC, Mos I, Lenselink EB, Lucchesi M, IJzerman AP, Schwartz TW. 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*. 2016;30:1836-1848. doi:10.1096/fj.201500110

56. Fleetwood O, Matricon P, Carlsson J, Delemotte L. Energy landscapes reveal agonist control of G protein-coupled receptor activation via microswitches. *Biochemistry*. 2020;59:880-891. doi:10.1021/acs.biochem.9b00842
57. Nygaard R, Frimurer TM, Holst B, Rosenkilde MM, Schwartz TW. Ligand binding and microswitches in 7TM receptor structures. *Trends Pharmacol Sci*. 2009;30:249-259.
58. Rasmussen SGF, DeVree BT, Zou Y, et al. Crystal structure of the β 2 adrenergic receptor-Gs protein complex. *Nature*. 2011;477(7366):549-555. doi:10.1038/nature10361
59. Lebon G, Warne T, Edwards PC, et al. Agonist-bound adenosine A (2A) receptor structures reveal common features of GPCR activation. *Nature*. 2011;474:521-U154. doi:10.1038/nature10136
60. Huang W, Manglik A, Venkatakrishnan AJ, et al. Structural insights into micro-opioid receptor activation. *Nature*. 2015;524:315-321. doi:10.1038/nature14886
61. Venkatakrishnan AJ, Deupi X, Lebon G, et al. Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature*. 2016;536:484-487.
62. Manglik A, Kruse AC. Structural basis for G protein-coupled receptor activation. *Biochemistry*. 2017;56:5628-5634. doi:10.1021/acs.biochem.7b00747
63. Mattedi G, Acosta-Gutiérrez S, Clark T, Grevasio FL. A combined activation mechanism for the glucagon receptor. *Proc Natl Acad Sci U S A*. 2020;117:15414-15422.
64. Gregory KJ, Sexton PM, Christopoulos A. Allosteric modulation of muscarinic acetylcholine receptors. *Curr Neuropharmacol*. 2007;5:157-167. doi:10.2174/157015907781695946
65. Ping YQ, Mao C, Xiao P, et al. Structures of the glucocorticoid-bound adhesion receptor GPR97-Go complex. *Nature*. 2021;589:620-626. doi:10.1038/s41586-020-03083-w
66. Rankovic Z, Brust TF, Bohn LM. Biased agonism: An emerging paradigm in GPCR drug discovery. *Bioorg Med Chem Lett*. 2016;26:241-250. doi:10.1016/j.bmcl.2015.12.024
67. Miao Y, McCammon JA. Graded activation and free energy landscapes of a muscarinic G-protein-coupled receptor. *Proc Natl Acad Sci U S A*. 2016;113:12162-12167. doi:10.1073/pnas.1614538113
68. Seidal L, Zarzycka B, Zaidi SA, Katritch V, Coin I. Structural insight into the activation of a class B G-protein-coupled receptor by peptide hormones in live human cells. *Elife*. 2017;6:e27711. doi:10.7554/eLife.27711
69. Kohlhoff KJ, Shukla D, Lawrenz M, et al. Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nature Chemistry*. 2014;6(1):15-21. doi:10.1038/nchem.1821
70. Coucke A, Uguzzoni G, Oteri F, Cocco S, Monasson R, Weigt M. Direct coevolutionary couplings reflect biophysical residue interactions in proteins. *J Chem Phys*. 2016;145:174102. doi:10.1063/1.4966156

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Do HN, Haldane A, Levy RM, Miao Y. Unique features of different classes of G-protein-coupled receptors revealed from sequence coevolutionary and structural analysis. *Proteins*. 2022;90(2):601-614. doi:10.1002/prot.26256