

# Molecular Determinant Underlying Selective Coupling of Primary G-Protein by Class A GPCRs

Qingya Shen, Xinyan Tang, Xin Wen, Shizhuo Cheng, Peng Xiao, Shao-Kun Zang, Dan-Dan Shen, Lei Jiang, Yanrong Zheng, Huibing Zhang, Haomang Xu, Chunyou Mao, Min Zhang,\* Weiwei Hu,\* Jin-Peng Sun,\* Yan Zhang,\* and Zhong Chen\*

G-protein-coupled receptors (GPCRs) transmit downstream signals predominantly via G-protein pathways. However, the conformational basis of selective coupling of primary G-protein remains elusive. Histamine receptors H<sub>2</sub>R and H<sub>3</sub>R couple with G<sub>s</sub>- or G<sub>i/o</sub>-proteins respectively. Here, three cryo-EM structures of H<sub>2</sub>R-G<sub>s</sub> and H<sub>3</sub>R-G<sub>i</sub> complexes are presented at a global resolution of 2.6–2.7 Å. These structures reveal the unique binding pose for endogenous histamine in H<sub>3</sub>R, wherein the amino group interacts with E206<sup>5,46</sup> of H<sub>3</sub>R instead of the conserved D114<sup>3,32</sup> of other aminergic receptors. Furthermore, comparative analysis of the H<sub>2</sub>R-G<sub>s</sub> and H<sub>3</sub>R-G<sub>i</sub> complexes reveals that the structural geometry of TM5/TM6 determines the primary G-protein selectivity in histamine receptors. Machine learning (ML)-based structuromic profiling and functional analysis of class A GPCR–G-protein complexes illustrate that TM5 length, TM5 tilt, and TM6 outward movement are key determinants of the G<sub>s</sub> and G<sub>i/o</sub> selectivity among the whole Class A family. Collectively, the findings uncover the common structural geometry within class A GPCRs that determines the primary G<sub>s</sub>- and G<sub>i/o</sub>-coupling selectivity.

in the human genome.<sup>[1]</sup> GPCR-mediated signaling pathways are involved in virtually every physiological function and many pathologies, thus representing the targets of approximately one-third of all medications, including antihistamines, antipsychotics, and opioid painkillers. In response to myriad extracellular stimuli, GPCRs exert adequate intracellular signaling predominantly through the engagement and activation of four major G-protein families: G<sub>s</sub>, G<sub>i/o</sub>, G<sub>q/11</sub>, and G<sub>12/13</sub>.<sup>[1b,2]</sup> Various G-protein subtypes elicit distinct signaling pathways and trigger versatile physiological effects by regulating different key effectors, such as adenylate cyclase, phospholipase C, and mitogen-activated protein kinases.<sup>[2b]</sup> While GPCR activation by agonists typically triggers multiple G-protein signaling pathways, many GPCRs exhibit pronounced bias and can preferentially activate certain downstream responses.<sup>[3]</sup> In many circumstances, the beneficial effects of a drug come from one signaling pathway and the adverse effects from other pathways.<sup>[4]</sup> Understanding the molecular mechanism underlying the selective coupling of an agonist-stimulated GPCR to an appropriate G-protein is therefore crucial for transmembrane signal transduction.

## 1. Introduction

Comprised of over 800 members, G-protein-coupled receptors (GPCRs) constitute the largest family of membrane proteins

Q. Shen, S. Cheng, S.-K. Zang, D.-D. Shen, H. Zhang, H. Xu, C. Mao, Y. Zhang  
Department of Pharmacology and Department of Pathology of Sir Run Run Shaw Hospital & Liangzhu Laboratory  
Hangzhou 310058, China  
E-mail: zhang\_yan@zju.edu.cn

Q. Shen, S. Cheng, S.-K. Zang, D.-D. Shen, H. Zhang, H. Xu, Y. Zhang  
MOE Frontier Science Center for Brain Research and Brain-Machine Integration  
Zhejiang University School of Medicine  
Hangzhou 310058, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/advs.202310120>

© 2024 The Authors. Advanced Science published by Wiley-VCH GmbH. This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/advs.202310120

X. Tang, L. Jiang, W. Hu, Z. Chen  
Department of Pharmacology and Department of Pharmacy of the Second Affiliated Hospital  
NHC and CAMS Key Laboratory of Medical Neurobiology  
School of Basic Medical Sciences  
Zhejiang University School of Medicine  
Hangzhou 310058, China  
E-mail: huww@zju.edu.cn; chenzhong@zju.edu.cn

X. Wen, P. Xiao, J.-P. Sun  
Advanced Medical Research Institute  
Meili Lake Translational Research Park  
Cheeloo College of Medicine  
Shandong University  
Jinan 250012, China  
E-mail: sunjinpeng@sdu.edu.cn

X. Wen, P. Xiao, J.-P. Sun  
Department of Biochemistry and Molecular Biology  
Shandong University School of Medicine  
Jinan 250012, China

An increasing number of studies suggested that many GPCRs exhibit a general promiscuity in their interactions with various G-protein subfamilies.<sup>[2a,5]</sup> The inherent dynamic nature of GPCRs enables them to sample multiple transition states between inactive and active states, contributing to the promiscuity of G-protein selectivity.<sup>[5b,6]</sup> Primary signaling stands out due to its remarkable coupling efficiency and rapid kinetics, emerging as the predominant signaling route for GPCRs among the spectrum of G-protein signaling interactions. According to the GPCR database (GPCRdb), among 265 non-olfactory GPCRs annotated on the G-protein family level, 228 (86%) have the potential to modulate receptor-specific cell activities by regulating adenylyl cyclase<sup>[2a,5a,7]</sup> (Figure 1A). Notably, among 183 GPCRs that predominantly couple with G<sub>s</sub>- or G<sub>i/o</sub>-proteins, 169 GPCRs (92.3% of the total) exhibit a high degree of selectivity for either G<sub>s</sub>- (48 GPCRs) or G<sub>i/o</sub>- (121 GPCRs) proteins.<sup>[2a]</sup> Therefore, G<sub>s</sub>- and G<sub>i/o</sub>-proteins are the most prevalent G-proteins among all the subtypes, leading to inverse cellular signaling responses through the stimulation or inhibition of adenylyl cyclase, respectively.

GPCR structures are indispensable tools, providing snapshots at local energy minimum that aid in uncovering the determinants for the primary signaling of G<sub>s</sub>- and G<sub>i/o</sub>-protein selectivity.<sup>[8]</sup> The resolution revolution in cryo-electron microscopy (cryo-EM) has led to an explosion of GPCR–G-protein complex structures, particularly the G<sub>s</sub>- and G<sub>i/o</sub>-coupled receptors. These structures have unveiled a common mode of GPCR–G-protein recognition, in which the C-terminal helix ( $\alpha 5$ ) of G $\alpha$  inserts into the cytosolic transmembrane (TM) cavity of GPCR.<sup>[1b,9]</sup> The coupling determinants for GPCR in G $\alpha$  proteins are widely accepted, with the selectivity barcode being the  $\alpha 5$  helix, the  $\alpha N$ - $\beta 1$  hinge, and the  $\beta 2$ - $\beta 3$  loop of G $\alpha$ .<sup>[1b,9c,d]</sup> However, the molecular basis of

primary G-protein-coupling selectivity of GPCRs at the receptor side remains intricate. Most investigations suggest that GPCRs coupled with different G-proteins generally adopt nearly identical conformations,<sup>[10]</sup> with the exception of the 5HT<sub>4</sub>,<sup>[11]</sup> resulting in the widespread doubts of the existence of complementary determinants at the GPCR level.<sup>[12]</sup> Additionally, the absence of evolved consensus sequences for G-protein recognition at GPCRs makes it difficult to draw a unified conclusion regarding primary G-protein selectivity.<sup>[1b,12,13]</sup> Previous structural studies have underscored the importance of ICL2, ICL3, TM5, and TM6 of GPCRs in G<sub>s</sub>/G<sub>i/o</sub>-protein selectivity.<sup>[9d,14]</sup> However, the limitation is that these studies often focused on specific receptors or receptor subfamilies, and the observed characteristics are not universal across class A GPCR families, rendering the primary G-protein selectivity mechanism still promiscuous and challenging.

Histamine, a biogenic amine, plays a pivotal role in various physiological and pathophysiological processes, such as allergic and inflammatory reactions, gastric acid secretion, wake-sleep disorders, and neurotransmission in the central nervous system.<sup>[15]</sup> The histamine receptor system comprises four members (H<sub>1</sub>R, H<sub>2</sub>R, H<sub>3</sub>R, and H<sub>4</sub>R), all of which belong to the Class A GPCR family.<sup>[16]</sup> Drugs targeting H<sub>1</sub>R, H<sub>2</sub>R, and H<sub>3</sub>R have been approved to treat allergic disorders,<sup>[17]</sup> gastroduodenal ulcers,<sup>[18]</sup> and narcolepsy,<sup>[19]</sup> respectively. The four subtypes of histamine receptors deliver downstream signaling by coupling with different heterotrimeric G-proteins: H<sub>1</sub>R mainly couples with G<sub>q</sub>-protein; H<sub>2</sub>R primarily couples with G<sub>s</sub>-protein and could alternatively couple with G<sub>q</sub>-protein; H<sub>3</sub>R and H<sub>4</sub>R are both capable of mediating downstream signaling via G<sub>i/o</sub>-protein activation primarily. Given their diverse coupling profiles, histamine receptors serve as an excellent model system for investigating the structural basis of primary G-protein-coupling and the preferential profile of GPCRs.

To date, structures of the inactive states of H<sub>1</sub>R, H<sub>2</sub>R, and H<sub>3</sub>R, as well as the active states of H<sub>1</sub>R and H<sub>4</sub>R, are available.<sup>[20]</sup> However, the absence of active-state structures of H<sub>2</sub>R and H<sub>3</sub>R has impeded the understanding of the molecular basis of H<sub>2</sub>R and H<sub>3</sub>R signaling and the structural-based drug design of histamine receptors. In this study, we report three cryo-EM structures of H<sub>2</sub>R-G<sub>s</sub> and H<sub>3</sub>R-G<sub>i</sub> complexes at a global resolution of 2.6–2.7 Å. These structures reveal the structural basis for agonist recognition and G-protein selectivity of H<sub>2</sub>R and H<sub>3</sub>R. A comparative analysis of the H<sub>2</sub>R-G<sub>s</sub> and H<sub>3</sub>R-G<sub>i</sub> complexes reveals that the conformation of TM5 and TM6 determines G-protein-selectivity in histamine receptors. Furthermore, machine learning (ML)-based structural profiling and functional analysis of publicly available and home-generated class A GPCR–G-protein complexes indicate that TM5 length, TM5 tilt, and TM6 outward movement determine G<sub>s</sub> and G<sub>i/o</sub> selectivity. To further enhance our understanding, we develop a ML-derived architecture classifier for G<sub>s</sub>- and G<sub>i/o</sub>-coupled receptors using homology models from GPCRdb as a training dataset. Remarkably, the classifier achieves an accuracy up to 91%, as validated by experimentally determined GPCR–G complex structures. Collectively, our results reveal the common structural geometries of class A GPCRs determine primary G<sub>s</sub>- and G<sub>i/o</sub>-coupling selectivity.

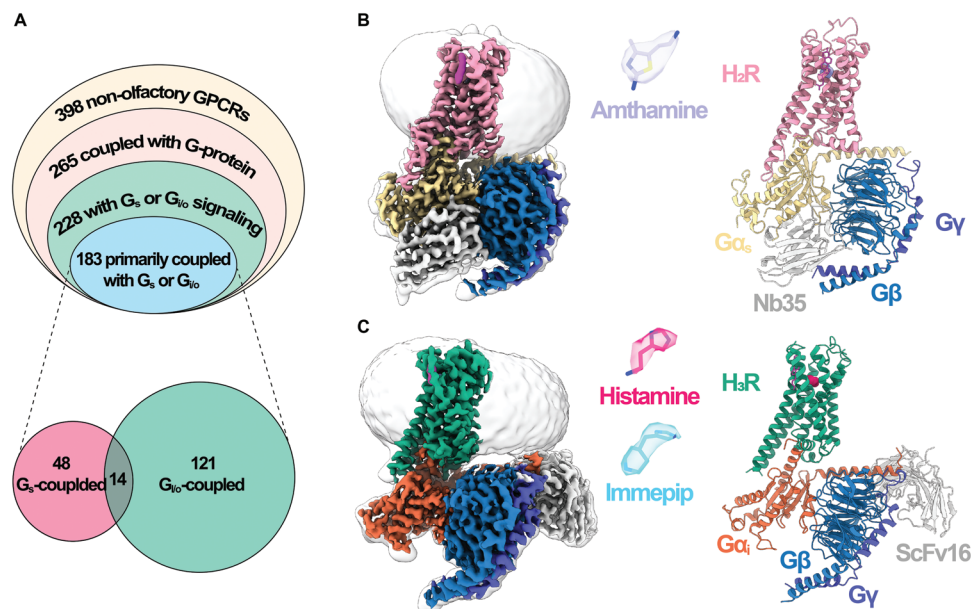
S. Cheng, M. Zhang  
College of Computer Science and Technology  
Zhejiang University  
Hangzhou 310027, China  
E-mail: [min\\_zhang@zju.edu.cn](mailto:min_zhang@zju.edu.cn)

Y. Zheng, Z. Chen  
Key Laboratory of Neuropharmacology and Translational Medicine of Zhejiang Province  
Zhejiang Chinese Medical University  
Hangzhou 310053, China

C. Mao  
Department of General Surgery  
Sir Run Run Shaw Hospital  
Zhejiang University School of Medicine  
Hangzhou, Zhejiang 310016, China

C. Mao  
Zhejiang Research and Development Engineering Laboratory of Minimally Invasive Technology and Equipment  
Zhejiang University  
Hangzhou 310016, China

J.-P. Sun  
Department of Physiology and Pathophysiology, School of Basic Medical Sciences  
Peking University  
Key Laboratory of Molecular Cardiovascular Science  
Ministry of Education  
Beijing 100191, China



**Figure 1.** Cryo-EM structures of H<sub>2</sub>R-G<sub>s</sub> and H<sub>3</sub>R-G<sub>i</sub> complexes. A) (Up) Proportion of G<sub>s</sub>- or G<sub>i/o</sub>-coupled receptors in the non-olfactory GPCRs. (Down) Number of receptors primarily coupling with G<sub>s</sub>- or G<sub>i/o</sub>-proteins. B) The cryo-EM density map (left) and atomic model (right) of amthamine-bound H<sub>2</sub>R-G<sub>s</sub> complex. The amthamine is depicted as stick within a transparent EM density map. C) The cryo-EM density map (left) and atomic model (right) of H<sub>3</sub>R-G<sub>i</sub> complex. The ligands histamine and immepip are depicted as sticks within a transparent EM density map.

## 2. Results

### 2.1. Cryo-EM Structures of the Agonist-Bound H<sub>2</sub>R-G<sub>s</sub> and H<sub>3</sub>R-G<sub>i</sub> Complexes

To investigate the structure of the H<sub>2</sub>R-G<sub>s</sub> complex, we co-expressed full-length H<sub>2</sub>R with the human heterotrimeric G<sub>s</sub> protein in Sf9 insect cells. A flag tag was fused to the N-terminus of H<sub>2</sub>R to aid purification. The H<sub>2</sub>R-G<sub>s</sub> complex was assembled in the presence of a highly selective agonist, amthamine, and further stabilized by a camelid antibody, Nb35.<sup>[21]</sup> Utilizing single-particle cryo-EM analysis, the purified H<sub>2</sub>R-G<sub>s</sub> complex was solved at a nominal resolution of 2.7 Å (Figure 1B; Figures S1, S2 and Table S1, Supporting Information). The H<sub>3</sub>R-G<sub>i</sub> complex was obtained through a co-expression strategy, similar to that of the H<sub>2</sub>R-G<sub>s</sub> complex. In addition to the N-terminal flag tag, a C-terminal tandem maltose-binding protein (MBP) tag was also introduced to H<sub>3</sub>R. A NanoBiT tethering strategy was used to improve the stability of the H<sub>3</sub>R-G<sub>i</sub> complex.<sup>[22]</sup> The endogenous agonist histamine and synthetic agonist immepip were used separately to generate the agonist-bound H<sub>3</sub>R-G<sub>i</sub> complexes, and the complexes were further stabilized by a single-chain antibody scFv16.<sup>[23]</sup> The histamine-bound H<sub>3</sub>R-G<sub>i</sub> complex and immepip-bound H<sub>3</sub>R-G<sub>i</sub> complex were solved at a global resolution of 2.7 Å and 3.0 Å, respectively (Figure 1C; Figures S1, S2 and Table S1, Supporting Information).

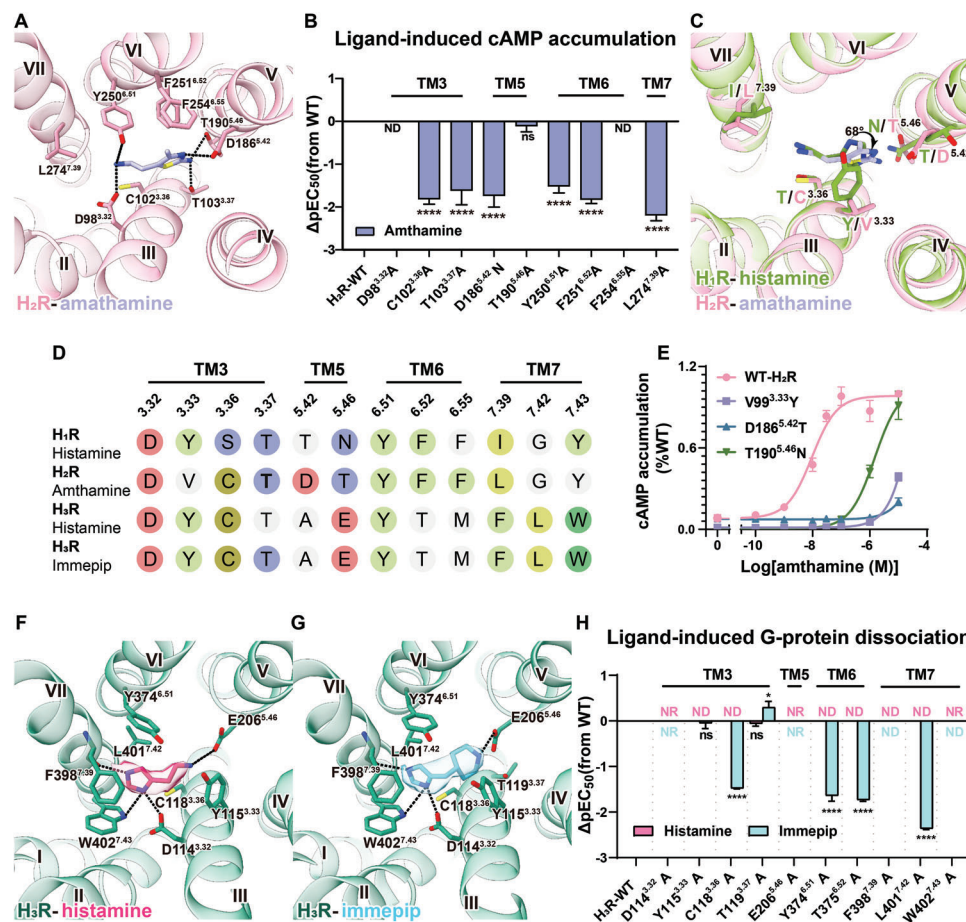
All three GPCR-G-protein complex structures are resolved at near-atomic resolution, and the EM density maps were clear enough to build the model of the receptors, heterotrimeric G-protein, antibody, and the bound ligand in the receptor orthosteric pockets (Figure S2 and Table S1, Supporting Information). The N- and C-termini of both receptors, as well as the ICL3 of H<sub>3</sub>R and the α-helical domains of Gα, exhibit poor resolution

in the cryo-EM density maps, consistent with previous cryo-EM solved GPCR-G-protein complexes due to the high flexibility of these regions. Notably, the density of cholesterol between TM1 and TM7 is observed in both H<sub>2</sub>R-G<sub>s</sub> and H<sub>3</sub>R-G<sub>i</sub> complexes, highlighting the critical role of cholesterol molecules in maintaining conformational stability and signaling activity of GPCR.

### 2.2. Ligand Recognition of Histamine Receptors

In the H<sub>2</sub>R-G<sub>s</sub> complex, the agonist amthamine occupies the conventional orthosteric pocket of H<sub>2</sub>R. As depicted in Figure S3A (Supporting Information), this pocket is defined by TM3, 5, 6, and 7, and it is capped by the extracellular loop 2 (ECL2) of H<sub>2</sub>R. The binding of amthamine is stabilized through a combination of polar and hydrophobic interactions (Figure 2A). Specifically, the aminoethyl group of amthamine is positioned between D98<sup>3.32</sup> and Y250<sup>6.51</sup> (we use the Ballesteros-Weinstein numbering system<sup>[24]</sup>), engendering a charge-charge interaction with D98<sup>3.32</sup>—a conserved “salt-bridge” also observed in other available monoamine receptor structures (Figure 2A).<sup>[20a,25]</sup> Additionally, the two amines present in the thiazolamine moiety of amthamine establish polar interactions with T103<sup>3.37</sup>, D186<sup>5.42</sup>, and T190<sup>5.46</sup> of H<sub>2</sub>R (Figure 2A). Consequently, the polar network effectively restrains both the aminoethyl group (head) and thiazolamine group (tail) of amthamine. Consistent with these observations, alanine mutagenesis of the majority of these residues markedly attenuated the cAMP accumulation response of H<sub>2</sub>R to amthamine (Figure 2B; Figure S4A and Table S2, Supporting Information). It’s important to note that the T190<sup>5.46</sup>A mutant exhibits an anomalous behavior, possibly attributed to compensatory effects involving T103<sup>3.37</sup>. The thiazole ring of amthamine is sandwiched between V99<sup>3.33</sup> on one side and F251<sup>6.52</sup>/F254<sup>6.55</sup> on the other side, facilitating crucial hydrophobic interactions





**Figure 2.** Ligand recognition of histamine receptors. A) Detailed interaction of amthamine with H<sub>2</sub>R. Dashed lines represent hydrogen bonds. B) Amthamine induced cAMP accumulation in HEK293 cells expressing H<sub>2</sub>R mutants of the residues in orthosteric pocket ( $n = 3$ , ordinary one-way ANOVA,  $***P < 0.0001$ , ND, not determinable, which refers to cannot be established over the tested concentration range, ns refers to no significance between the WT and mutant). C) Structural comparisons of ligand recognition between histamine-H<sub>1</sub>R and amthamine-H<sub>2</sub>R structures. D) Sequence alignment of the orthosteric pocket of histamine receptors. Residues that interact with ligands are highlighted with colored circles. E) Dose-dependent curves for amthamine induced cAMP accumulation in HEK293 cells expressing the H<sub>2</sub>R mutants (V99<sup>3.33</sup>Y, D186<sup>5.42</sup>T, T190<sup>5.46</sup>N) that are not conserved in H<sub>1</sub>R ( $n = 3$ ). F-G) Detailed interactions of histamine F) and immepip G) with H<sub>3</sub>R. H) Agonists induced G<sub>i</sub> dissociation in HEK293 cells expressing H<sub>3</sub>R mutants of the residues in orthosteric pocket by NanoBiT assays ( $n = 3$ , ordinary one-way ANOVA,  $*P = 0.0456$ ,  $***P < 0.0001$ , NR refers to no response to the ligand, ND, not determinable, which refers to cannot be established over the tested concentration range, ns refers to no significance between the WT and mutant).

(Figure 2A). The methyl group position on the thiazole ring intensifies the hydrophobic interaction with F254<sup>6.55</sup>. Moreover, amthamine establishes van der Waals interactions with C102<sup>3.36</sup> and L274<sup>7.39</sup> (Figure 2A). Our mutational studies targeting these residues corroborate the substantial import of these hydrophobic interactions (Figure 2B; Figure S4A and Table S2, Supporting Information).

H<sub>2</sub>R displays the greatest similarity to H<sub>1</sub>R among the four histamine receptors, exhibiting an approximate 73% sequence homology in the 7TM region. As anticipated, the residues within the binding pocket of both H<sub>1</sub>R and H<sub>2</sub>R manifest a high degree of conservation. Among the 12 residues within the H<sub>2</sub>R binding pocket, seven are identical and three show homogeneous to their counterparts in H<sub>1</sub>R (Figure 2D; Figure S3F, Supporting Information). Notably, despite this conservation, discernible differences emerge in the ligand pocket shapes of H<sub>1</sub>R and H<sub>2</sub>R (Figure S3B,C, Supporting Information). Y108<sup>3.33</sup> and N198<sup>5.46</sup>,

possessing substantial steric bulk, serve to constrict the ligand pocket in H<sub>1</sub>R (Figure 2C). In contrast, the structurally equivalent residues V99<sup>3.33</sup> and T190<sup>5.46</sup> in H<sub>2</sub>R do not introduce such steric hindrance. As a consequence, the docking of amthamine into H<sub>1</sub>R would inevitably give rise to significant steric clashes, particularly involving the thiazole ring and the primary amine moiety of the ring with Y108<sup>3.33</sup> and N198<sup>5.46</sup> (Figure 2C). Furthermore, the substitution of D186<sup>5.42</sup> in H<sub>2</sub>R with the corresponding T194<sup>5.42</sup> residue in H<sub>1</sub>R diminished the polar interactions with amthamine (Figure 2C). Collectively, our structures provide a clear explanation for the high selectivity of amthamine as an H<sub>2</sub>R agonist. Furthermore, mutations of these three residues of H<sub>2</sub>R to the corresponding residues in H<sub>1</sub>R significantly decreased the potency and efficacy of amthamine, further supporting our conclusion (Figure 2E; Table S2, Supporting Information).

The H<sub>3</sub>R-G<sub>i</sub> complexes, bound with histamine and immepip, exhibit highly analogous overall structures, featuring a

remarkably low root-mean-square deviation (RMSD) of 0.3 Å for the H<sub>3</sub>R C $\alpha$  atoms. Both ligands reside within the orthosteric pocket, a characteristic shared with H<sub>2</sub>R and other aminergic receptors, as depicted in Figure 2. In the histamine-bound H<sub>3</sub>R-G<sub>i</sub> complex, the amino group of histamine engages a pivotal hydrogen bond with E206<sup>5,46</sup> (Figure 2F). Simultaneously, the N $\pi$  atom within the imidazole ring establishes van der Waals forces with C118<sup>3,36</sup>, while also forming hydrogen bonds with D114<sup>3,32</sup> and W402<sup>7,43</sup> (Figure 2F). Furthermore, the N $\tau$  atom establishes a hydrogen bond with the backbone of F398<sup>7,39</sup> (Figure 2F). Through diligent alanine mutagenesis investigations, we affirmatively ascertained the indispensability of these interactions for histamine binding and the ensuing receptor activation. Notably, the alanine mutations targeting E206<sup>5,46</sup>, D114<sup>3,32</sup>, and W402<sup>7,43</sup> unequivocally abolish the capacity for G<sub>i</sub> protein activation (Figure 2H; Figure S4B and Table S2, Supporting Information). Additionally, the EC<sub>50</sub> value of the C118<sup>3,36</sup>A mutant experiences an approximately 150-fold reduction in comparison to wild-type H<sub>3</sub>R. Furthermore, the imidazole ring of histamine securely embeds itself within the hydrophobic core constituted by Y374<sup>6,51</sup>, L401<sup>7,42</sup>, W402<sup>7,43</sup>, and F398<sup>7,39</sup> (Figure 2F). This interaction is particularly evident in the formation of a face-to-face  $\pi$ - $\pi$  stacking association with F398<sup>7,39</sup>. The veracity of these observations is endorsed by alanine substitutions targeting the residues implicated in this critical hydrophobic core, further underscoring their pivotal role in facilitating histamine recognition (Figure 2H; Figure S4B and Table S2, Supporting Information).

Within the H<sub>3</sub>R orthosteric pocket, the imidazole rings of histamine and immepip overlap and share many interactions with H<sub>3</sub>R in a nearly identical position (Figure 2G). Immepip, a synthetic agonist derived from histamine, replaces the aminoethyl moiety with a piperidine ring. A notable departure emerges in the positioning of the amine within the piperidine ring, situated a mere 1.3 Å closer to E206<sup>5,46</sup> than the primary amine in histamine, thus intensifying its polar interactions with E206<sup>5,46</sup>. The bulky piperidine ring also establishes hydrophobic and van der Waals interactions with Y115<sup>3,33</sup> and T119<sup>3,37</sup>, conferring enhanced compatibility of immepip with the dumbbell-shaped pocket characteristic of H<sub>3</sub>R (Figure S3D,E, Supporting Information). Consequently, the binding interface of immepip with H<sub>3</sub>R spans  $\approx 290$  Å<sup>2</sup>, an appreciable 20 Å<sup>2</sup> more expansive than that of histamine. Notably, despite modifications introduced to Y115<sup>3,33</sup> and T119<sup>3,37</sup>, the resulting effect on the pEC<sub>50</sub> values to immepip is negligible, potentially attributable to the compensatory amplification of polar interactions between the piperidine moiety and E206<sup>5,46</sup> (Figure 2H; Figure S4C and Table S2, Supporting Information).

### 2.3. Unique Histamine Binding Mode of H<sub>3</sub>R

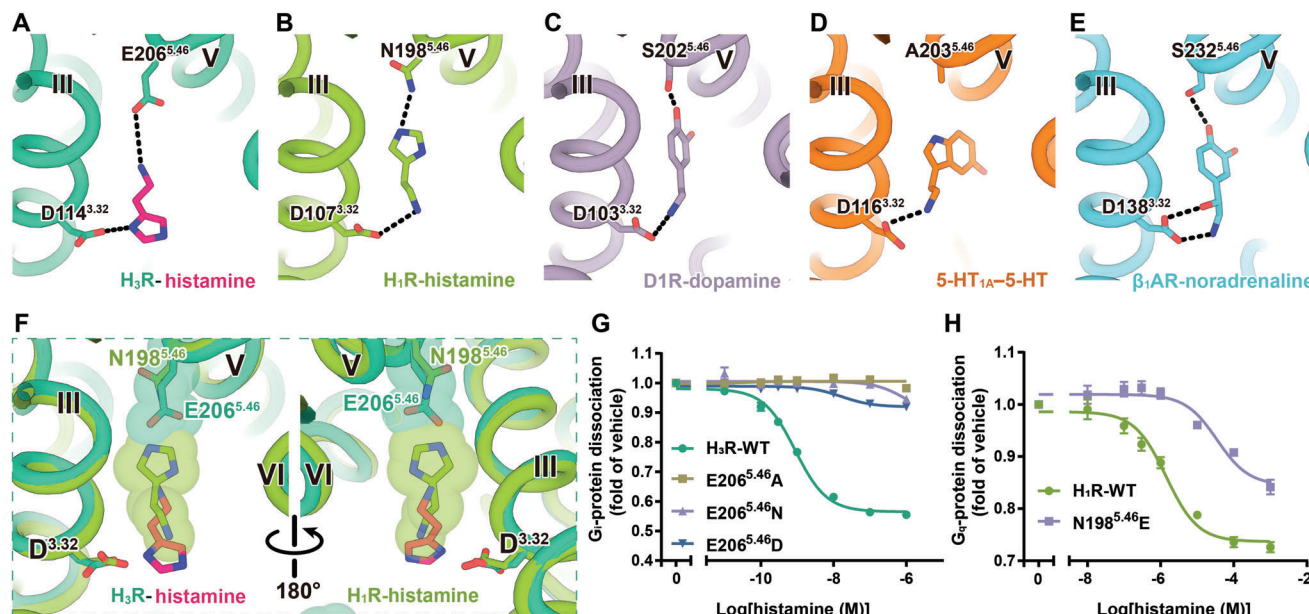
Recent reports have detailed the structures of aminergic receptors bound to their endogenous ligands, including H<sub>1</sub>R with histamine,<sup>[20a]</sup> D1R with dopamine,<sup>[25c]</sup> 5-HT<sub>1A</sub> with 5-HT,<sup>[25a]</sup> and  $\beta_1$ AR with noradrenaline.<sup>[26]</sup> Comparison of these structures reveals that the biogenic monoamine ligands adopt a nearly identical binding pose, situated at the bottom of the receptor's orthosteric binding pockets. Notably, the amino group of these

monoamine ligands forms a salt-bridge with the residue D<sup>3,32</sup>, which is highly conserved among aminergic receptors. Intriguingly, our high-resolution cryo-EM Map unveils a distinctive binding pose of histamine in H<sub>3</sub>R, in which the amino group of histamine forms a salt-bridge with E206<sup>5,46</sup> rather than interacting with D114<sup>3,32</sup> in other aminergic receptors (Figure 3A–F; Figure S5, Supporting Information). To further validate this surprising observation, we constructed an alternative model with histamine adopting the classical pose observed in other aminergic receptors. This alternative model, along with the non-classical model revealed by cryo-EM map, were subjected to molecular dynamics (MD) simulation. As anticipated, histamine in our non-classical pose for H<sub>3</sub>R remained stable throughout a 200 ns simulation (Figure S5B, Supporting Information). In contrast, histamine in the classic pose of aminergic receptors, as seen in H<sub>1</sub>R, exhibited high instability after a 200 ns simulation, with two considerably distinct conformations in the trajectory (Figure S5C, Supporting Information). These findings conclusively establish that histamine adopts a unique pose in H<sub>3</sub>R compared to other aminergic receptors.

Sequence alignment of all aminergic receptors reveals that the 5.46 position is mainly occupied by small side-chain residues that are either uncharged polar or hydrophobic, such as threonine, serine, and alanine (Table S3, Supporting Information). This allows sufficient space to accommodate the bulky conjugated group of monoamine ligands. Consequently, the relatively compact imidazole group of histamine enables the inclusion of residues with bulkier side chains in the orthosteric binding pocket, as exemplified by the presence of asparagine in H<sub>1</sub>R. However, compared to A<sup>5,46</sup> in 5-HT<sub>1A</sub> and S<sup>5,46</sup> in D1R, the elongated glutamine residue in H<sub>3</sub>R causes severe steric hindrance with the imidazole group in the classic pose. Consequently, histamine rotates 180-degree along the bilayer plane, resulting in the amino group of histamine interacting with E206<sup>5,46</sup> instead of D114<sup>3,32</sup> (Figure 3A–F; Figure S5, Supporting Information). A comparison of the orthosteric ligand binding pockets of H<sub>1</sub>R and H<sub>3</sub>R reveals that five of the eight residues constituting the H<sub>3</sub>R ligand pocket are not conserved with H<sub>1</sub>R (Figure 2D; Figure S3F, Supporting Information). This indicates that the shape and size of histamine binding pockets are significantly distinct between the two histamine receptors. The unique histamine binding of H<sub>3</sub>R is further confirmed by the differential effects of mutations of H<sub>1</sub>R and H<sub>3</sub>R at the 5.46 position (Figure 3G,H; Table S4, Supporting Information). This position contributes to the different binding capacities to the endogenous ligand histamine, enabling the bulk discovery of selective agonists and antagonists targeting H<sub>3</sub>R.

### 2.4. Structural Differences Reveal the Basis of G-Protein Selectivity in Histamine Receptors

Histamine receptors H<sub>1</sub>R, H<sub>2</sub>R, and H<sub>3</sub>R primarily couple with G<sub>q</sub>-, G<sub>s</sub>-, or G<sub>i</sub>-proteins, respectively. Our structures of H<sub>2</sub>R-G<sub>s</sub> and H<sub>3</sub>R-G<sub>i</sub> complexes, along with the recently determined H<sub>1</sub>R-G<sub>q</sub> structures,<sup>[20a]</sup> enable structural comparisons among histamine receptor–G-protein signaling complexes. Such comparative analysis holds the potential to unveil the intricate molecular determinants underlying G-protein-coupling



**Figure 3.** Endogenous ligand binding poses of aminergic receptors. A–E) Binding poses of H<sub>3</sub>R (A), H<sub>1</sub>R (B), D1R (C), 5-HT<sub>1A</sub> (D), and β<sub>1</sub>AR (E) with their endogenous ligands. F) Superimposition of H<sub>3</sub>R-histamine and H<sub>1</sub>R-histamine structures. G, H) Effects of histamine-induced G-protein dissociation in HEK293 cells expressing histamine receptor mutants at 5.46 position. Dose-response curves for H<sub>3</sub>R E206<sup>5.46</sup> mutants (G) and H<sub>1</sub>R N198<sup>5.46</sup>E mutant (H) were measured using the NanoBiT assay (*n* = 3).

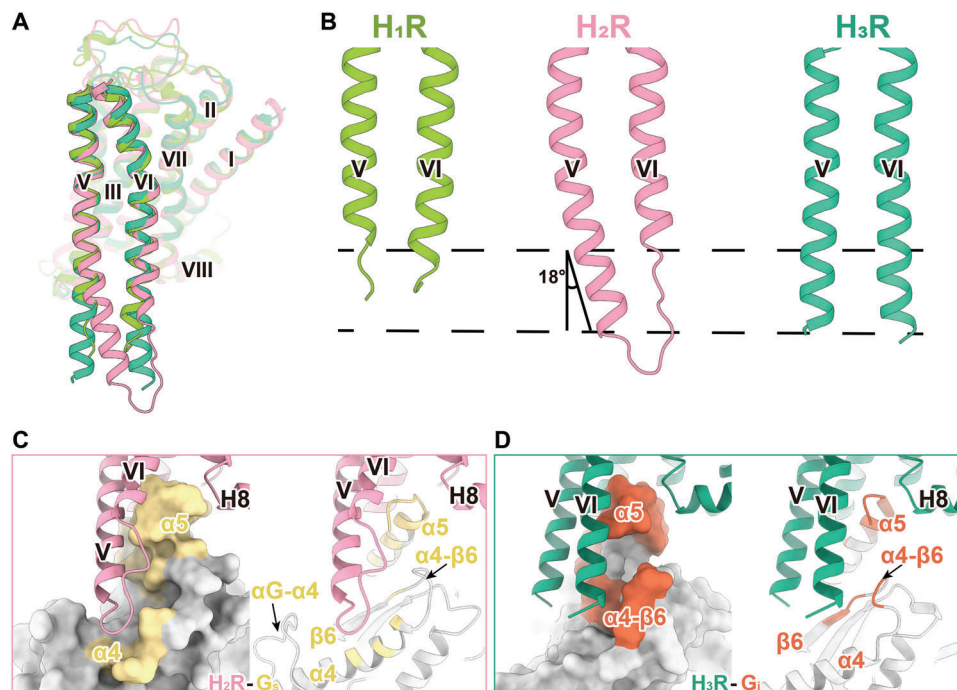
preferences within this receptor subfamily, offering insights that could extend to other GPCRs. Upon superimposing the three structures focused on the receptor, we observe a congruence in the alignment of TM1, TM2, TM3, TM7, and ICL1 on the cytoplasmic side. Conversely, disparities emerge in the configuration of TM4-6 and ICL2-3 (Figure S6A,B, Supporting Information). This observation highlights that while the α5 of Gα inserts into the intracellular cavity of the receptors, contributing to the classical α5-insertion interface within the complex structures, the specific manner in which the G-protein binds diverges significantly across the three structures.

In both H<sub>1</sub>R-G<sub>q</sub> and H<sub>3</sub>R-G<sub>i</sub> structures, TM6 exhibits comparable outward displacements at their cytoplasmic ends. However, in the H<sub>2</sub>R-G<sub>s</sub> complex, a significant TM6 movement of 4.7 Å further outward than in H<sub>3</sub>R, as measured at Cα atoms of residue 6.30 (Figure S6C, Supporting Information). Notably, this represents the most significant TM6 outward displacement among the histamine receptor subfamilies. Consequently, the C-terminus of α5 in Gα<sub>s</sub> shifts by 4.0 Å toward TM6 in comparison to Gα<sub>i</sub> within the H<sub>3</sub>R-G<sub>i</sub> complex, as measured at the conserved Cα atoms of residue H5.25 (common Gα numbering system)<sup>[27]</sup> in Gα (Figure S6C, Supporting Information). Additionally, beyond the dynamic TM6, we observed significant variations in the conformation of TM4, as well as their corresponding ICL2 regions, on the cytoplasmic side. Notably, the intracellular tip of TM4 in H<sub>3</sub>R exhibits a noteworthy 4.9 Å shift toward TM2 in comparison to H<sub>2</sub>R, as determined by measuring the Cα atoms of residue 4.40 (Figure S6B, Supporting Information). This specific movement induces the ICL2 of H<sub>3</sub>R to rotate approximately 64-degree compared to H<sub>2</sub>R (Figure S6B, Supporting Information). Besides, our H<sub>2</sub>R-G<sub>s</sub> and H<sub>3</sub>R-G<sub>i</sub> structures also reveal that the ICL2s form

polar and hydrophobic interactions with the αN helix, the β2-β3 loop, and the α5 helix of Gα, which are critical for G-protein coupling.

The most notable distinctions among the three GPCR-G-protein complexes are prominently evident within the TM5/TM6 regions (Figure 4A,B). In the H<sub>2</sub>R-G<sub>s</sub> complex, the length TM5 of H<sub>2</sub>R extends ≈2.5 helical turns (from position 5.67 to 5.75, encompassing nine residues) compared with H<sub>1</sub>R (Figure 4B; Figure S2, Supporting Information). Furthermore, the extended TM5 in H<sub>2</sub>R tilts toward TM6 from the Y202<sup>5.58</sup> residue for an 18-degree angle (Figure 4B). Subsequently, the cytosolic tip of TM5 inserts into the cavity of Gα<sub>s</sub>, which is formed by the αG-α4 loop, the α4 helix, and the α4/β6 turn, resulting in additional interactions with the α4 of Gα<sub>s</sub> (Figure 4C). In the H<sub>3</sub>R-G<sub>i</sub> structure, both TM5 and TM6 of H<sub>3</sub>R exhibit unexpected elongations comprising nine and eight residues (2.5 and 2.2 helical turns) compared to H<sub>1</sub>R, respectively (Figure 4B). This elongation of TM5/TM6 regions in H<sub>3</sub>R establishes close contacts with the α4/β6 turn of Gα<sub>i</sub>, thereby strengthening the H<sub>3</sub>R-G<sub>i</sub> coupling effectiveness (Figure 4D). Consequently, the interfaces of H<sub>2</sub>R-G<sub>s</sub> (1610 Å<sup>2</sup>) and H<sub>3</sub>R-G<sub>i</sub> (1381 Å<sup>2</sup>) are both significantly larger than that of H<sub>1</sub>R-G<sub>q</sub> (1144 Å<sup>2</sup>). In conclusion, the elongations observed in TM5/TM6 regions within H<sub>2</sub>R and H<sub>3</sub>R constitute additional yet distinctive interfaces between the receptor and G-proteins, hereafter referred to as non-canonical receptor-Gα interface. We anticipated that the differential non-canonical receptor-Gα interactions resulting from the extended TM5/TM6 could be the key determinants for the primary G<sub>s</sub> or G<sub>i/o</sub> selectivity. To validate our hypothesis, we analyzed the structural geometries of TM5/TM6 among available GPCR-G-protein structures. This exploration aimed to unravel potential the correlations between the G<sub>s</sub>- or G<sub>i/o</sub>-coupling selectivity within class A receptors.





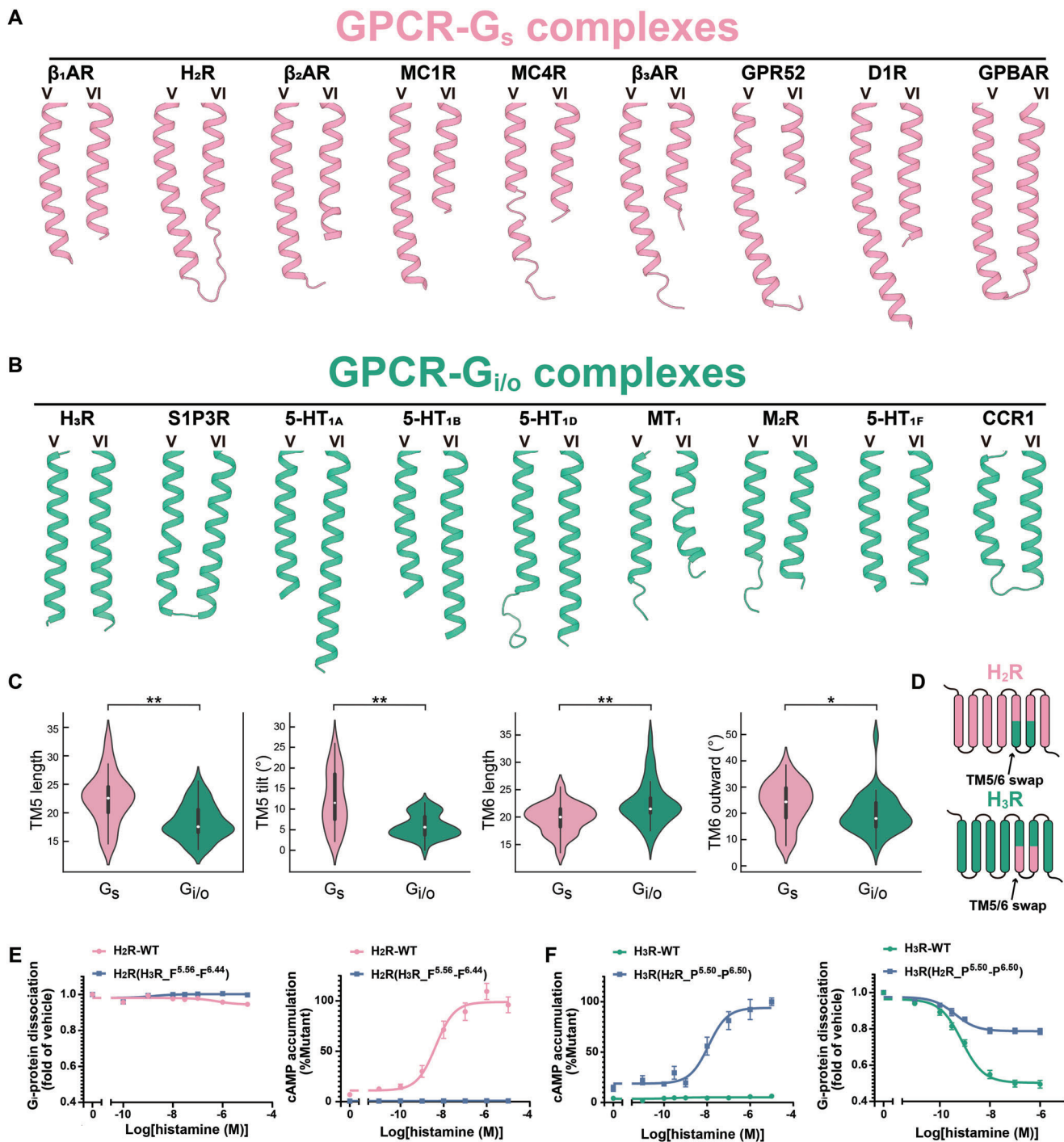
**Figure 4.** Structural comparison of histamine receptor–G-protein complexes. A) Structural superposition of the active models of H<sub>1</sub>R, H<sub>2</sub>R, and H<sub>3</sub>R. B) Comparison of TM5 and TM6 in H<sub>1</sub>R–G<sub>q</sub>, H<sub>2</sub>R–G<sub>s</sub>, and H<sub>3</sub>R–G<sub>i</sub> complexes. C,D) Surface (left) and cartoon (right) representation of binding area of H<sub>2</sub>R (C) and H<sub>3</sub>R (D) with their respective coupled G $\alpha$  subunit. Residues within 4Å of H<sub>2</sub>R or H<sub>3</sub>R in G $\alpha_s$  or G $\alpha_i$  are highlighted in yellow or orange, respectively.

## 2.5. TM5 and TM6 are Responsible for G<sub>s</sub> and G<sub>i/o</sub> Selectivity

Plentiful GPCR–G-protein structures have been determined with increasing speed since we resolved the first two cryo-EM structures of GPCR–G-protein complexes in 2017.<sup>[9a,b]</sup> Currently,  $\approx 170$  structures of class A GPCRs coupled to their corresponding G-protein are available (as of July 2022). Among them, 13 G<sub>s</sub>- and 30 G<sub>i/o</sub>-coupled signaling complex structures (Table S5, Supporting Information) of unique class A receptors, whose structures have been resolved with their corresponding primary G-protein, are chosen for further analysis. We found that the extended TM5/TM6 structural geometries observed in H<sub>2</sub>R–G<sub>s</sub> and H<sub>3</sub>R–G<sub>i</sub> complexes are widespread in above structures. More than half of receptors in G<sub>s</sub>-coupled state, the length of TM5 is remarkably longer than TM6 (the median length of TM5 and TM6 is 25 and 21, respectively (Table S5, Supporting Information). Helix lengths of the intracellular region are measured from the residue x.50 to the cytosolic tip of TM5 or TM6), including three  $\beta$ -adrenergic receptors, D1R, MC1R, MC4R, GPBAR, GPR52, and H<sub>2</sub>R (Figure 5A).<sup>[14c,28]</sup> In addition, similar to H<sub>2</sub>R, the cytosolic half of TM5 in G<sub>s</sub>-coupled receptors exhibits a curved conformation toward TM6, with a range of tilting angles spanning from 8- to 26-degree (Figure 5A; Table S5, Supporting Information). Nevertheless, the situation for G<sub>i/o</sub>-coupled structures is more complex. For instance, certain receptors, including H<sub>3</sub>R,  $\mu$ OR, S1PR3, 5-HT<sub>1A</sub>, 5-HT<sub>1B</sub>, and 5-HT<sub>1D</sub>, possess a TM6 length exceeding 26 residues.<sup>[14e,25a,29]</sup> Notably, for 5-HT<sub>1A</sub>, 5-HT<sub>1B</sub>, and 5-HT<sub>1D</sub>, the length of TM6 is 35, 31, and 31 residues, respectively. A significant proportion of GPCR–G<sub>i/o</sub> structures, including MT<sub>1</sub>, M<sub>2</sub>R, 5-HT<sub>1F</sub>, chemokine receptors, etc., exhibit short

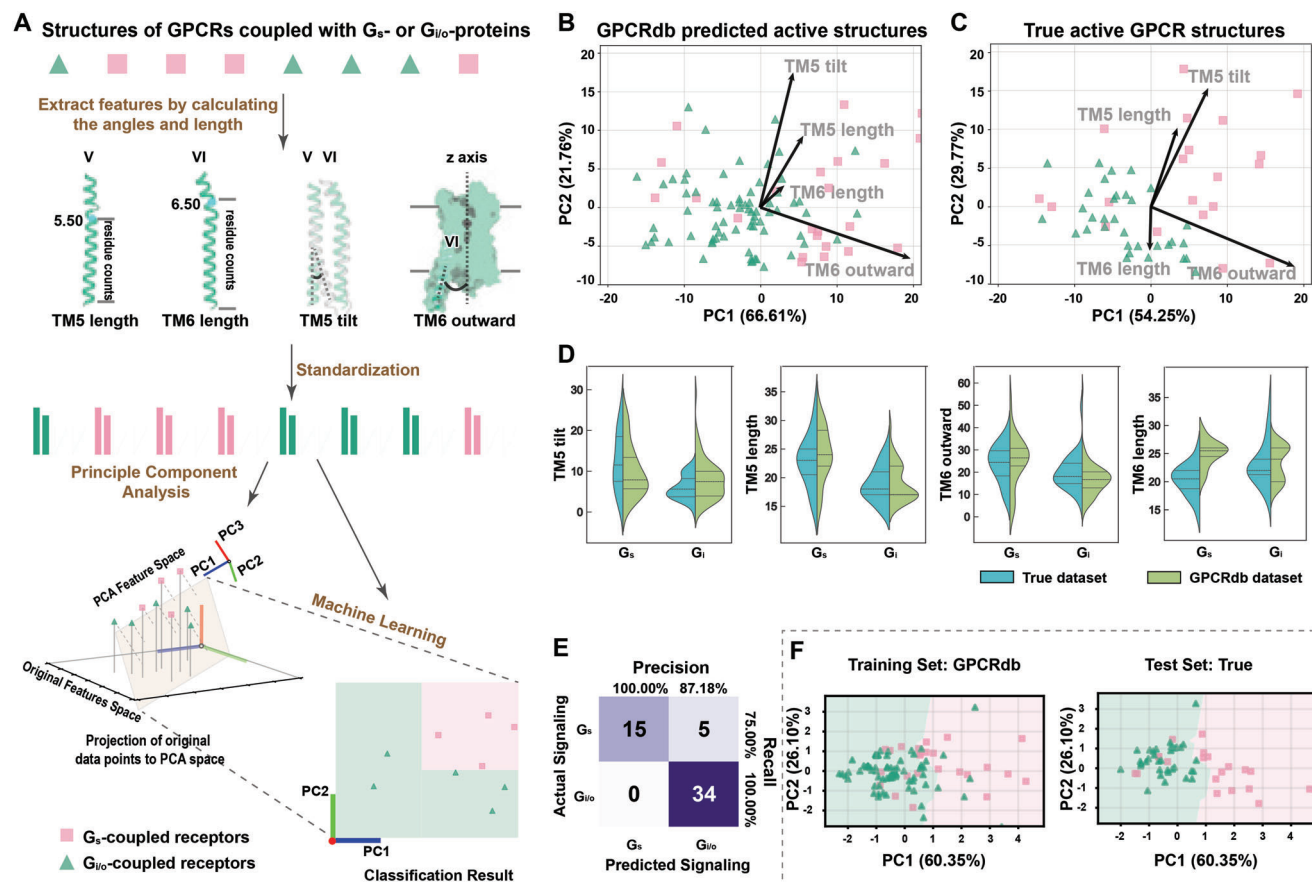
helices of both TM5 and TM6, with the length of TM6 similar or only slightly longer than TM5 (Figure 5B).<sup>[30]</sup> It is noteworthy that in the majority of G<sub>i/o</sub>-coupled receptor structures, the TM5 is straight, as in H<sub>3</sub>R, indicating that the cytosolic ends of TM5 are not curved toward TM6, as in H<sub>2</sub>R in G<sub>s</sub>-coupled structures (Figure 5A,B). Therefore, the length of TM5 and TM6, together with the tilt of the TM5 cytosolic end, exhibits significant differences between G<sub>s</sub>- and G<sub>i/o</sub>-coupled receptors, likely responsible for G<sub>s</sub> and G<sub>i/o</sub> selectivity.

Previous research has established that the outward movement of TM6 is the hallmark of receptor activation, thus plays a crucial role in receptor G<sub>s</sub>/G<sub>i/o</sub>-coupling selectivity. Specifically, G<sub>s</sub>-coupled structures exhibit larger outward movements of TM6 compared to G<sub>i/o</sub>-coupled receptors.<sup>[9c]</sup> To evaluate the significance of four structural geometries of TM5 and TM6 in G<sub>s</sub> and G<sub>i/o</sub> selectivity (the lengths of TM5 and TM6, the tilt of the TM5 cytosolic end, and the outward movement of TM6), we analyzed currently available G<sub>s</sub>- and G<sub>i/o</sub>-coupled structures of class A receptors, along with our unpublished structures (including a total of 20 G<sub>s</sub>- and 34 G<sub>i/o</sub>-coupled receptors) (Table S5, Supporting Information). The lengths of TM5 and TM6 were measured from the cytosolic tip to 5.50 and 6.50, respectively, and the lengths and angles of TM5/TM6 of receptors were extracted as illustrated in Figure 6A (see Experimental Section). Consistent with our speculation, significant differences were observed in the geometries of TM5 and TM6 between G<sub>s</sub>- and G<sub>i/o</sub>-coupled receptors. The average length of TM5 in G<sub>s</sub>-coupled receptors is four residues longer than that of G<sub>i/o</sub>-coupled receptors ( $23.1 \pm 4.6$  and  $18.9 \pm 3.1$ , respectively) (Figure 5C; Table S5, Supporting Information). The tilt of TM5 in G<sub>s</sub>-coupled structures



**Figure 5.** TM5 and TM6 are responsible for G<sub>s</sub> and G<sub>i</sub> selectivity. A) Representative of the TM5/TM6 structural geometry of G<sub>s</sub>-coupled receptors. B) Representation of the TM5/TM6 structural geometry of G<sub>i</sub>-coupled receptors. C) Violin plots depicting four distinct features of TM5/TM6 in G<sub>s</sub>- and G<sub>i/o</sub>-coupled receptors. The white dot denotes the median. The interquartile range is shown by the broad black bar in the middle. Except for points considered to be “outliers” using an interquartile range-based technique, the thin line reflects the remainder of the distribution.  $n = 54$ ,  $n^s P > 0.05$ ,  $*P < 0.05$ ,  $**P < 0.01$  by Mann-Whitney U test. D) Schematic representation of the “TM5/TM6 swap” experiments. E–F) Dose-dependent curves for histamine induced G-protein dissociation and cAMP accumulation in HEK293 cells expressing chimera receptors ( $n = 3$ ). E) Replacement of F<sup>5.56</sup>-F<sup>6.44</sup> of H<sub>2</sub>R with F<sup>5.56</sup>-F<sup>6.44</sup> of H<sub>3</sub>R did not confer the ability to dissociate G<sub>i</sub>-protein (left), but resulted in loss of the ability to activate the G<sub>s</sub> signal pathway (right). F) Replacement of P<sup>5.50</sup>-P<sup>6.50</sup> of H<sub>3</sub>R with P<sup>5.50</sup>-P<sup>6.50</sup> of H<sub>2</sub>R resulted in gain of the ability to activate the G<sub>s</sub> signal pathway (left), but loss of the ability to dissociate G<sub>i</sub>-protein (right).





**Figure 6.** Machine learning-based classification of GPCRs into  $G_s$  and  $G_i$  signaling pathway. A) Workflow for GPCR classification utilizing machine learning and feature pre-processing. The length and tilt of TM5, as well as the length and outward movement of TM6 are extracted from active GPCR structures obtained from GPCRdb homology models and True structures. The resulting data of these four geometries are standardized, and then subjected to PCA. Subsequently, a Random Forest classifier is employed to classify GPCRs based on the PCA results. B,C) PCA biplots were generated for the four geometries of GPCRdb homology models (B) and True GPCR structures (C) individually. The contributions of each variable to the principal components are depicted as vectors on the plot, where the vertical component of a vector on a given PC illustrates the respective contribution of that variable to the PC. The angle between two vectors reflects the correlation between the corresponding features, while the length of a vector represents the significance of the corresponding feature. D) Comparison of violin plots depicting four geometries of TM5/TM6 in GPCRdb structures and True GPCR structures. Here the median and interquartile range are depicted as dashes. E) Confusion matrix for the True GPCR structures given by the best classifier. F) Decision boundary visualization via PCA given by the best classifier and scatter plot of GPCRdb homology models (left) and True GPCR structures (right) based on GPCRdb homology models.

was distributed into two major clusters ( $12.3 \pm 7.1^\circ$ ), while the majority of the  $G_{i/o}$ -coupled receptors exhibit smaller tilts ( $5.90 \pm 2.9^\circ$ ) (Figure 5C; Table S5, Supporting Information). The average outward movement of TM6 in  $G_s$ -coupled receptors was slightly larger than that of  $G_{i/o}$ -coupled structures ( $23.8 \pm 8.4^\circ$  and  $19.7 \pm 7.6^\circ$ , respectively), consistent with previous findings (Figure 5C; Table S5, Supporting Information). However, the TM6 lengths of  $G_s$ -coupled receptors and  $G_{i/o}$ -coupled receptors were indistinguishable, with comparable length distributions and medians (Figure 5C; Table S5, Supporting Information). In conclusion, the structural geometries of TM5/TM6, particularly the length of TM5, the tilt of cytosolic TM5, and the outward movement of TM6, may be responsible for primary  $G_s$  and  $G_{i/o}$  selectivity.

To validate the hypothesis that the preference selection of G-proteins for class A receptors is determined by TM5/TM6, we conducted “TM5/TM6 swap” experiments to reverse the  $G_s$  and

$G_{i/o}$  selectivity of receptors (Figure 5D). In the first set of experiments, we replaced the intracellular half of TM5, TM6, and ICL3 ( $P^{5.50}$  to  $P^{6.50}$ ) of  $G_{i/o}$ -coupled receptors with that of  $G_s$ -coupled receptors. As expected, these chimeric receptors induced cAMP accumulation, including  $H_3R(H_2R_{P^{5.50}.P^{6.50}})$ ,  $H_3R(D1R_{P^{5.50}.P^{6.50}})$ ,  $H_3R(MC4R_{P^{5.50}.P^{6.50}})$ , and  $H_3R(\beta 2AR_{P^{5.50}.P^{6.50}})$ , and  $D2R(H_2R_{P^{5.50}.P^{6.50}})$  (Figure 5E; Figure S7A and Table S6, Supporting Information). In the second set of experiments, we attempted to switch  $G_s$ -coupled receptors to  $G_{i/o}$ -coupled receptors. Although these engineered receptors could not recruit  $G_i$  proteins, their ability to stimulate cAMP accumulation was almost lost (Figure 5F; Figure S7B and Table S6, Supporting Information). In conclusion, these “TM5/TM6 swap” experiments could alter the  $G_s$ - or  $G_{i/o}$ -coupling selectivity of receptors, supporting the notion that the conformational architecture of TM5/TM6 determines the  $G_s$  and  $G_{i/o}$  selectivity of a GPCR.

## 2.6. Classification of $G_s$ - and $G_{i/o}$ -Coupled Receptors by Machine Learning

Can the structural determinants of  $G_s$  and  $G_{i/o}$  selectivity identified in this study be generalized to all Class A GPCRs? To address this question, a machine learning approach was utilized to categorize primary  $G_s$ - and  $G_{i/o}$ -coupled receptors, based on the structural geometries of TM5/TM6 of receptors (Figure 6A). To augment the generalization of the predictive model, we incorporated predicted structures in the structural datasets (see Experimental Section). The training set comprised 98 (24  $G_s$ - and 74  $G_{i/o}$ -coupled receptors) GPCRdb-predicted homology models (Table S7, Supporting Information),<sup>[31]</sup> while the test set consisted of 54 (20  $G_s$ - and 34  $G_{i/o}$ -coupled receptors) true (experimental) structures (Table S5, Supporting Information). In light of potential correlations among the four structural geometries of TM5/TM6, we opted to apply Principal Component Analysis (PCA)<sup>[32]</sup> to effectively decorrelate and compress these characterizes, thereby facilitating a reduction of the TM5/TM6 geometry space from a 4D to a 2D construct (Figure 6B,C). The first two Principle Component (PCs) are reserved for further analysis (PC1 and PC2), contributing more than 86% of the variance (Figure 6B,C). The individual biplots of the training and true datasets revealed that the TM5 length, TM5 tilt, and TM6 outward movement are the primary contributors to the principal components (The contributions of each variable to the principal components are depicted as vectors on the plot, where the vertical component of a vector on a given PC illustrates the respective contribution of that variable to the PC) (Figure 6B,C). Specifically, the TM6 outward movement significantly influences PC1, whereas the length and tilt of TM5 play a critical role in PC2 (Figure 6B,C). Moreover, these three vectors exhibit similarities in the principal component plots between the training and true datasets, particularly in terms of TM5 tilt and TM6 outward movement (Figure 6B,C). The significance of TM5 length, TM5 tilt, and TM6 outward movement in determining the selectivity of  $G_s$ - and  $G_{i/o}$ -protein coupling in class A receptors is suggested by these findings.

Remarkably, the biplots indicate that between the two datasets, the TM6 length has the largest discrepancy: its eigenvector directions are almost opposite in the two biplots, whereas the remaining vectors have similar directions (Figure 6B,C). In the training dataset, the impact of the TM6 length vector on each principal component appears to be minimal and positively correlated to both PCs (Figure 6B,C). In contrast, it exerts a negligible effect on PC1 and displays a negative correlation with PC2 in the true dataset (Figure 6B,C). Subsequent violin plots reveal a symmetry of the distributions of TM5 tilt and TM6 outward movement in the two datasets (Figure 6D). Nevertheless, there is a notable disparity in the distribution pattern of TM6 length between the two datasets (Figure 6D). Further correlation analysis of four geometries of TM5/TM6 elucidates similar inconsistency. The TM6 length from the training dataset reveals a mildly positive correlation of 0.20 for the TM5 tilt (Figure S8, Supporting Information). Conversely, the TM6 length from the true dataset unveils a negative correlation for TM5 tilt, with a magnitude of  $-0.29$  (Figure S8, Supporting Information). In conclusion, these results imply that the TM6 length poses a challenge for model training, concurrently indicating that TM6 length marginally contributes to G-protein selectivity.

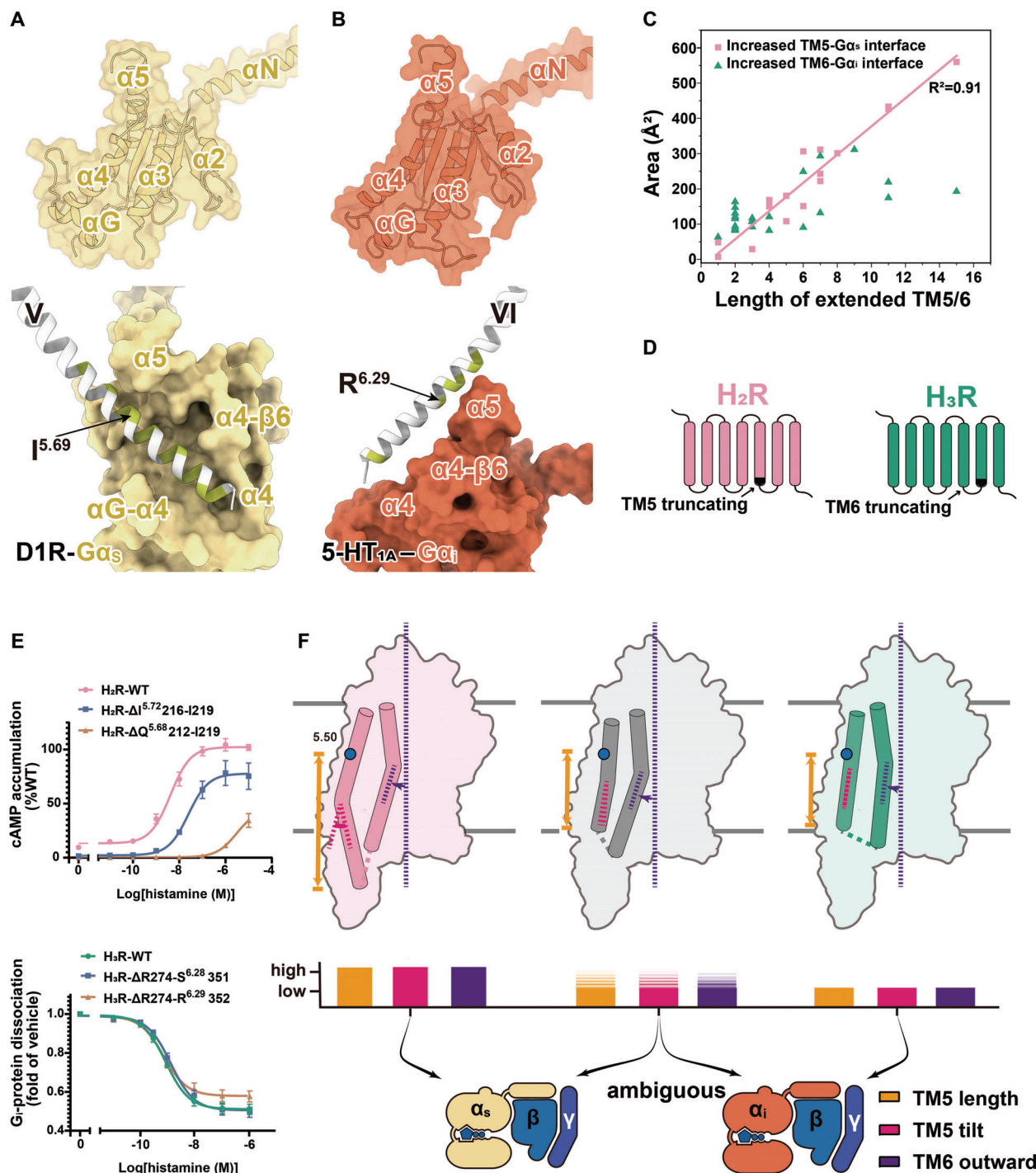
Aimed at deriving a more precise conclusion about the underlying mechanism, we trained a classifier for primarily  $G_s$ - and  $G_{i/o}$ -coupled receptors through a machine-learning approach directly on the three standardized geometries of TM5/TM6 (TM6 length excluded). We employed a variant of the Random Forest Classifier<sup>[33]</sup> with constraints to the number of estimators and their maximum depth to decrease the risk of overfitting. Validation of the model using the true dataset achieved an accuracy of 90.74%, F1-score of 0.8943, Area Under Curve (AUC) of 0.8162, and Matthews Correlation Coefficient (MCC) of 0.8086 (Figure 6E; Tables S5,S8, Supporting Information). This further illustrates the effectiveness of these three geometries for selectivity. Interestingly, certain falsely predicted receptors (PE2R2, PE2R4, V2R) in our classifier exhibited comparable  $G_s$  and  $G_{i/o}$ -protein selectivity, and LSHR had limited proof to support its primary coupling according to the GPCRdb/GproteinDb.<sup>[31]</sup> In brief, the predictive model has fulfilled good performance.

Besides, with the aid of PCA visualization, we can depict a probable decision boundary given by our best classifier to derive a simpler formula (Figure 6F). The decision boundary elucidates that GPCRs with extended TM5 length, significant TM5 tilt, and prominent TM6 outward movement are indicative of  $G_s$ -coupling receptors. In contrast, GPCRs featuring shorter TM5 length, minimal TM5 tilt, and constrained TM6 outward movement are more likely associated with  $G_{i/o}$ -coupling receptors (Figure 7F).

## 2.7. TM6 Length is not Significant for $G_s/G_{i/o}$ Selectivity

Structuromic analysis and ML analysis of GPCR-G-protein complexes suggested that the length of TM6 was not responsible for  $G_s$  and  $G_{i/o}$  selectivity. Therefore, further investigation is required to elucidate the mechanisms underlying the differential effects of TM5 and TM6 length in G-protein selectivity. In the GPCR- $G_s$  complex, TM5 extends to insert into the groove formed by the  $\alpha G\text{-}\alpha 4$  loop, the  $\alpha 4$  helix, and the  $\alpha 4\text{-}\beta 6$  loop of  $G\alpha_s$  (Figure 7A), leading to increased interactions with  $G\alpha_s$ . To achieve this, extended TM5 tilts toward TM6 to fit into the groove in  $G\alpha_s$ . Failure to accomplish this could result in severe steric hindrance with the  $\alpha G\text{-}\alpha 4$  loop of  $G\alpha_s$ , as observed in D1R (Figure 7A),<sup>[28d]</sup> indicating a correlation between the length of TM5 and its tilt in the  $G_s$ -coupling receptors. Consistent with this, biplot analyses of GPCRdb-predicted structures and true structures consistently reveal a high correlation between the length and tilt of TM5 in  $G_s$ -coupling receptors, with coefficients of 0.54 and 0.63, respectively (Figure S8, Supporting Information). However, the groove in  $G\alpha_s$  is absent in  $G\alpha_{i/o}$  due to the absence of 13 residues in the  $\alpha G\text{-}\alpha 4$  loop (Figure 7B). Consequently, in the GPCR- $G_{i/o}$  complex, the extended TM5 and TM6 can only rest on the flattened surface of  $G\alpha_{i/o}$ , resulting in smaller contact areas and fewer interactions compared to the  $G_s$ -coupled structure.

To validate this hypothesis, we investigated whether the extended TM5 in  $G_s$ -coupled structures or extended TM6 in  $G_{i/o}$ -coupled structures contributed equivalently to the increase in the contact area. Among the 54 true GPCR-G structures, the median length of the entire TM5 is 23 in 20  $G_s$ -coupled structures and 18 in 34  $G_{i/o}$ -coupled structures (Table S5, Supporting Information). The median length of TM6 is 20.5 in  $G_s$ -coupled re-



**Figure 7.** Contributions to G-protein signaling of extended TM5/TM6. A) (Up) Surface and cartoon representations of  $G_{\alpha_s}$  are shown, with the surface of  $G_{\alpha_s}$  depicted as transparent. (Down) The binding surface of TM5 with  $G_{\alpha_s}$  of D1R is shown, with TM5 beyond I<sup>5.69</sup> noted as extended helix. Residues that interact with  $G_{\alpha}$  are highlighted in green. B) (Up) Surface and cartoon representations of  $G_{\alpha_i}$  are shown, with the surface of  $G_{\alpha_i}$  depicted as transparent. (Down) The binding surface of TM6 with  $G_{\alpha_{i/o}}$  of 5-HT<sub>1A</sub>. TM6 beyond R<sup>6.29</sup> was considered as extended helix. C) The additional contact areas of  $G_s$ - or  $G_{i/o}$ -coupled receptors between the extended TM5/TM6 with  $G_{\alpha_s}$ / $G_{\alpha_{i/o}}$ , respectively. D) Schematic representation of the TM5- (up) or TM6- (down) breaking mutants. E) Dose-dependent curves for histamine induced cAMP accumulation or G-protein dissociation in HEK293 cells expressing the truncated receptors ( $n = 3$ ). (Up) The TM5 of H<sub>2</sub>R was respectively truncated by four amino acids and eight amino acids. The cAMP assay validated that the length of TM5 was closely related to the activation of the  $G_s$  signal pathway. (Down) The TM6 of H<sub>3</sub>R was respectively truncated by eight amino acids and nine amino acids. The NanoBit assay validated that the length of TM6 did not correlate with the dissociation of the  $G_i$ -protein. F) Schematic representation of the geometry determinants for G-protein selectivity of GPCRs.



ceptors and 22 in  $G_{i/o}$ -coupled structures. Consequently, we considered TM5 longer than 18 (below 5.69) in  $G_s$ -coupled receptors and TM6 longer than 20.5 (below 6.29) in  $G_{i/o}$ -coupled receptors as extended TMs. Interestingly, we found that the additional contact area between the extended TM5 and  $G\alpha_s$  exhibited a linear correlation with the extended residue amount,  $\approx 40 \pm 3.0 \text{ \AA}^2$  per residue ( $R^2 = 0.91$ ) (Figure 7C; Figure S9 and Table S9, Supporting Information). For example, the contact area between the 15-residue-extended TM5 of D1R with  $G\alpha_s$  reached  $561 \text{ \AA}^2$  (Figure 7A). In contrast, the increased contact area between TM6 and  $G\alpha_{i/o}$  did not show a significant correlation with the length of the extended TM6. For instance, the contact area between the 15-residue-extended TM6 of 5-HT<sub>1A</sub> and  $G\alpha_i$  was only  $192.9 \text{ \AA}^2$  (Figure 7B).

Visualizing cryo-EM density map of  $G_{i/o}$ -coupled complexes further revealed that the extended TM6 is not significant for G-protein coupling. The densities for the extended TM6 in several  $G_i$ -coupled complex including 5-HT<sub>1A</sub>, 5-HT<sub>1B</sub>, and 5-HT<sub>1D</sub> are dramatically weaker and lack high-resolution features compared with the rest of the receptor, suggesting the instable interactions between the extended TM6 and  $G_i$ -protein (Figure S10, Supporting Information).<sup>[25a,29b]</sup> Consistently, the removal of the extended TM6 of H<sub>3</sub>R posed negligible effect on its  $G_i$  activity even when nine residues were removed from the cytosolic extremity of TM6 (Figure 7D,E; Table S10, Supporting Information). In contrast, cAMP accumulation experiments confirmed that the removal of the extended TM5 of H<sub>2</sub>R significantly reduced its  $G_s$ -protein activity (Figure 7D,E; Table S10, Supporting Information). In conclusion, our findings indicate that a long TM5 is crucial for  $G_s$ -coupled receptors, whereas the TM6 length is not significant for  $G_s$  and  $G_{i/o}$  selectivity.

### 3. Discussion

The four histamine receptors serve essential roles in pathophysiological and signaling events, rendering them significant drug targets. In this study, we presented the structures of one H<sub>2</sub>R- $G_s$  and two H<sub>3</sub>R- $G_i$  complexes, offering profound insights into the orthosteric pocket of histamine receptors and G-protein selectivity. Through our investigation of the amthamine-bound H<sub>2</sub>R structure and mutagenesis studies, we revealed the discrepancy in the ligand pocket of H<sub>1</sub>R and H<sub>2</sub>R that contributed to H<sub>2</sub>R selectivity for the agonist amthamine. Notably, our findings uncovered a unique binding pose for endogenous histamine in H<sub>3</sub>R, deviating from other monoamine receptors, as the amino group interacts with E206<sup>5,46</sup> in H<sub>3</sub>R instead of the conserved D114<sup>3,32</sup> found in other aminergic receptors. This observation coincided with the behavior of the antipsychotic drug quetiapine, which exhibits high or modest affinities for many aminergic receptors with the exception of H<sub>3</sub>R and H<sub>4</sub>R.<sup>[34]</sup> Importantly, the structural geometry differences observed in TM5/TM6 between H<sub>2</sub>R- $G_s$  and H<sub>3</sub>R- $G_i$  structures led us to propose the determinants of primary  $G_s$ - and  $G_{i/o}$ -coupling selectivity.

GPCRs predominantly regulate downstream signaling through the engagement and activation of four major G-protein subtypes. Among these four members,  $G_s$  and  $G_{i/o}$  are two principal signaling pathways, for about 86% GPCRs are annotated to couple with  $G_s$ - or  $G_{i/o}$ -protein.<sup>[2a]</sup> The selective

coupling of GPCRs to the G-proteins is crucial for effective signal transduction, especially those primarily coupled with  $G_s$ - and  $G_{i/o}$ -proteins with remarkable coupling efficiency and rapid kinetics. The explosion of GPCR-G-protein structures in recent years has significantly enhanced our comprehension on GPCR coupling mechanisms. Receptors and G-proteins interact through a common mode, wherein the ligand-activated GPCR opens its intracellular cavity and accommodates the  $\alpha 5$  of  $G\alpha$ , which is conserved through the family/class A, B1, B2, and F GPCRs.<sup>[9c,35]</sup> From the perspective of the G-protein, the selectivity-determining positions in  $G\alpha$  predominantly attribute to the  $\alpha 5$  helix, along with the  $\alpha N$ - $\beta 1$  loop, the  $\beta 2$ - $\beta 3$  loop, and the  $\alpha 4$  helix of  $G\alpha$ .<sup>[1b]</sup> However, the selectivity signatures of receptors are more complicated due to the great divergence in the amino acid sequences and structural conformations of GPCRs. In contrast to the 16 members of  $G\alpha$  proteins which adopt similar conformations, the intracellular parts of GPCRs are more diverse, particularly in the regions of TM5, TM6, H8, intracellular loops, and the receptor C-terminal tail. By combining reported GPCR-G-protein structures with unpublished structures from our laboratory, we obtained a number of non-redundant  $G_s$ -coupled or  $G_{i/o}$ -coupled structures of class A receptors, totaling 20 and 34, respectively. These structures revealed that the selectivity of  $G_s$  and  $G_{i/o}$  is determined by TM5/TM6 geometries, including the length and tilt of TM5, and the outward movement of TM6. The length and tilt of TM5 exhibit a high correlation in true GPCR- $G_s$  complexes, consistent with the extended and tilted TM5 (e.g., H<sub>2</sub>R) inserting into the groove formed by the  $\alpha G$ - $\alpha 4$  loop, the  $\alpha 4$  helix, and the  $\alpha 4$ - $\beta 6$  loop of  $G\alpha_s$ . However, the  $\alpha G$ - $\alpha 4$  loop is 13 residues shorter in  $G\alpha_{i/o}$  than in  $G\alpha_s$ , which cannot constitute the same groove as  $G\alpha_s$  in  $G\alpha_{i/o}$ . Hence, the extended TM6 (e.g., H<sub>3</sub>R) leads to less contact with  $G\alpha_i$  in  $G_{i/o}$ -coupled structures. This phenomenon indicates that the interactions of GPCR- $G_s$  are stronger than that of GPCR- $G_{i/o}$  complexes, which is consistent with Matic et al.'s report that the binding energy of  $G_s$  complexes is stronger than that of  $G_{i/o}$  complexes.<sup>[13]</sup> Furthermore, our "TM5/6-breaking" experiments also confirmed that TM6 length is not significant for  $G_s$  and  $G_{i/o}$  selectivity.

The intricate molecular mechanism underlying primary G-protein coupling selectivity in GPCRs remains complex and not fully elucidated. The structures of GPCR-G-protein complexes indicate that the ICL2, ICL3, TM5, and TM6 of receptors are associated with G-protein-coupling selectivity. However, these cytoplasmic regions of receptors rarely exhibit common patterns in terms of the sequence or amino-acid properties related to G-protein selectivity. In our prior study, we proposed that the TM5 and TM6 helices synergistically alternate lengths to determine the selectivity between  $G_s$ - and  $G_{i/o}$ -proteins, defined as "a macro-switch".<sup>[11]</sup> Comparative sequence analyses revealed that specific amino acids function as "micro-switches" uniquely localized in the complementary pocket. However, these two switches only partially determine the selectivity between  $G_s$  and  $G_{i/o}$ . Aided by the drastic increasing number and improved qualities of the structures of GPCRs in complexes with different G-proteins in house and in the public data server, in the present study, we conducted a systematic evaluation to thoroughly explore the intricate relationship between the four geometries of TM5/TM6 and their impact on primary G-protein selectivity by using

machine learning-based structuromic profiling. Using the GPCRdb homology models as the training dataset, we generated a  $G_s$ - or  $G_{i/o}$ -coupling classifier that achieved an impressive 91% accuracy in the true dataset based on the TM5/TM6 geometries. Based on our classifier, we find that GPCRs with a long TM5 length and large TM5 tilt and TM6 outward movement are  $G_s$ -coupling receptors, whilst GPCRs with a short TM5 length and small TM5 tilt and TM6 outward movement are more likely to be  $G_{i/o}$ -coupling receptors. We anticipate that GPCRs that fail to satisfy the aforementioned conditions probably have comparable or promiscuous activity in stimulating  $G_s$  and  $G_{i/o}$  signaling (Figure 7F). Interestingly, class B1 receptors, primarily coupled to  $G_s$ -protein, exhibit large TM5 tilt and TM6 outward movement. These structural geometries also conform to the G-protein selectivity features observed in Class A receptors, further demonstrating the universality of our proposed G-protein selectivity mechanism (Table S11, Supporting Information). In summary, our studies revealed that the structural geometry of TM5/TM6 serves as the determinant of the primary G-protein signaling selectivity in GPCRs. These findings have the potential to aid efforts in engineering GPCR selectivity, especially by modifying promiscuous GPCRs to improve affinity to specific G-proteins.

## 4. Experimental Section

**Constructs:** The wild-type human  $H_2R$  gene was cloned into the pFast-Bac1 vector. The N-terminus of  $H_2R$  was fused with the hemagglutinin signal peptide (HA) to enhance receptor expression, followed by a Flag tag (DYKDDDDK) to facilitate complex purification. The wild-type human  $H_3R_{445}$  isoform was cloned into a modified pFastBac1 vector with a hemagglutinin (HA) signal sequence at the N-terminus and a PreScission protease site followed by a Flag tag. The C-terminus of  $H_3R$  was fused with the LgBiT,<sup>[22]</sup> followed by a TEV protease site and a double MBP tag to facilitate expression and purification. The dominant-negative bovine  $G\alpha_s$  ( $DNG\alpha_s$ ) and human  $G\alpha_{i1}$  ( $DNG\alpha_{i1}$ ) were generated by site-directed mutagenesis, as previously described, to stabilize the interaction with the  $\beta\gamma$  subunits,<sup>[36]</sup> and were cloned into the pFastBac1 vector. Human  $G\beta 1$  was fused with a 10 $\times$  His tag at the N-terminus and HiBiT at the C-terminus, and together with  $G\gamma 2$ , was cloned into the pFastBac dual vector.

**Protein Complex Expression and Purification:** For the  $H_2R$ - $G_s$  complex, the  $H_2R$ ,  $DNG\alpha_s$ , and  $G\beta 1\gamma 2$  constructs were co-expressed in *Spodoptera frugiperda* (Sf9) cells using the Bac-to-Bac Baculovirus Expression System (Invitrogen). Cells were infected at a density of  $2.4 \times 10^6$  cells per mL and then co-infected with three separate viruses at a ratio of 2:1:1 for  $H_2R$ ,  $DNG\alpha_s$ , and  $G\beta 1\gamma 2$ . Cells were collected 48 hours post-infection and stored at  $-80^\circ\text{C}$  until use. Sf9 cell pellets expressing  $H_2R$  and G-protein trimer were resuspended in a buffer containing 20 mM HEPES pH 7.4, 100 mM NaCl, 3 mM  $MgCl_2$ , 5 mM  $CaCl_2$ , 2.5 mg mL<sup>-1</sup> leupeptin, and 0.2 mg mL<sup>-1</sup> benzamidine. The complex was assembled on the membrane by adding 100  $\mu\text{M}$  amthamine, 10 mg mL<sup>-1</sup> Nb35, 25 mU mL<sup>-1</sup> apyrase, and further incubated at RT for 2 h. The complex was then extracted from the cell membrane using 0.5% (w/v) lauryl maltose neopentylglycol (LMNG) (Anatrace) and 0.1% (w/v) cholesteryl hemisuccinate (CHS) (Anatrace) at  $4^\circ\text{C}$  for 2 h. After that, the supernatant was collected by centrifugation at 25 000 rpm for 30 min, and the solubilized complex was incubated with M1 anti-FLAG resin for 2.5 h at  $4^\circ\text{C}$ . Afterwards, the resin was collected and washed with a buffer containing 20 mM HEPES pH 7.4, 100 mM NaCl, 3 mM  $MgCl_2$ , 5 mM  $CaCl_2$ , 0.1% (w/v) LMNG, 0.02% (w/v) CHS, 100  $\mu\text{M}$  amthamine, 2.5 mg mL<sup>-1</sup> leupeptin, and 0.2 mg mL<sup>-1</sup> benzamidine. Subsequently, the detergent was changed to 0.01% (w/v) LMNG, 0.005% (w/v) CHS. Then, the  $H_2R$ - $G_s$  complex was eluted with 20 mM HEPES pH 7.4, 100 mM NaCl, 0.01% (w/v) LMNG, 0.001% (w/v)

CHS, 5 mM EGTA, 0.2 mg mL<sup>-1</sup> FLAG peptide, 100  $\mu\text{M}$  amthamine, 2.5 mg mL<sup>-1</sup> leupeptin, and 0.2 mg mL<sup>-1</sup> benzamidine. The eluate was collected and injected onto a Superdex 6 Increase 10/300 GL column (GE Healthcare) with buffer containing 20 mM HEPES pH 7.4, 100 mM NaCl, 0.00025% (w/v) LMNG, 0.0002% (w/v) CHS, 0.00075% (w/v) GDN (Anatrace), and 100  $\mu\text{M}$  amthamine. The complex fractions were concentrated with a 100 kDa MWCO Millipore concentrator to 8 mg mL<sup>-1</sup> for making a cryo-EM grid.

For the  $H_3R$ - $G_i$  complex, the  $H_3R_{445}$  isoform,  $DNG\alpha_{i1}$ , and  $G\beta 1\gamma 2$  constructs were co-expressed in *Spodoptera frugiperda* (Sf9) cells using the Bac-to-Bac Baculovirus Expression System (Invitrogen). Cells were infected at a density of  $2.4 \times 10^6$  cells per mL and then co-infected with three separate viruses at a ratio of 1:1:1 for  $H_3R$ ,  $DNG\alpha_{i1}$ , and  $G\beta 1\gamma 2$ . Cells were collected 48 h post-infection and stored at  $-80^\circ\text{C}$  until use. The co-expressed  $H_3R$ - $G_i$  complex cell pellets were lysed in a buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, 2 mM  $MgCl_2$ , adequate agonist (1 mM histamine or 10  $\mu\text{M}$  immpip), and supplemented with EDTA-free protease inhibitor cocktail (Bimake) using a Dounce homogenizer. The complex formation was initiated by adding 20 mg mL<sup>-1</sup> scFv16 and 25 mU mL<sup>-1</sup> apyrase (NEB) and subsequently incubated for 1.5 h at RT. Cell membranes were solubilized by 1% (w/v) n-Dodecyl- $\beta$ -D-Maltopyranoside (DDM, Anatrace) and 0.2% (w/v) cholesterol hemisuccinate (CHS, Anatrace) for 2.5 h at  $4^\circ\text{C}$ . After centrifugation at 30 000 g for 30 min, the supernatant was isolated and incubated with amylose resin (NEB) for 1 hour at  $4^\circ\text{C}$ . Then, the resin was washed with a wash buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, 2 mM  $MgCl_2$ , 0.1% (w/v) LMNG, 0.02% (w/v) CHS, and agonist (1 mM histamine or 10  $\mu\text{M}$  immpip). Subsequently, the detergent was changed to 0.01% (w/v) LMNG, 0.005% (w/v) CHS, and eluted with an elution buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, 2 mM  $MgCl_2$ , 0.01% (w/v) LMNG, 0.005% (w/v) CHS, 10 mM Maltose, and agonist (1 mM histamine or 10  $\mu\text{M}$  immpip). The double MBP tag of  $H_3R$  was removed by incubating with TEV protease for 1 hour at RT. Then, the protein sample was further purified by size-exclusion chromatography with a Superose 6 10/300 GL column (GE Healthcare) equilibrated with a running buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, 2 mM  $MgCl_2$ , 0.00225% (w/v) LMNG, 0.0006% (w/v) CHS, 0.00075% (w/v) GDN (Anatrace), and agonist (1 mM histamine or 10  $\mu\text{M}$  immpip).

**Cryo-Grid Preparation and EM Data Collection:** Three microliters of the purified GPCR-G-protein complexes were applied onto a glow-discharged holey carbon grid (Quantifoil R1.2/1.3) at  $\approx 20$  mg mL<sup>-1</sup>. The grids were plunge-frozen in liquid ethane using Vitrobot Mark IV (Thermo Fisher Scientific). The frozen grids were then transferred to liquid nitrogen and stored for data collection.

Cryo-EM imaging was performed on a Titan Krios at 300 kV using the Gatan K2 Summit detector at the Center of Cryo-Electron Microscopy, Zhejiang University (Hangzhou, China). Movies were recorded at a dose rate of approximately 8.0 e/Å<sup>2</sup>/s with a defocus ranging from  $-1.0$  to  $-2.2$   $\mu\text{m}$  using the SerialEM software in counting mode for the three receptor-G-protein complexes, respectively. The total exposure time was 8 seconds and 40 frames were recorded per micrograph. A total of 3122, 3008, and 3079 movies were collected for the amthamine- $H_2R$ - $G_s$ , histamine- $H_3R$ - $G_i$ , and immpip- $H_3R$ - $G_i$  complexes, respectively, for structure reconstruction.

**Image Processing and 3D Reconstructions:** Image stacks were aligned using MotionCor 2.1.<sup>[37]</sup> Contrast transfer function (CTF) parameters were estimated using Gctf v1.18.<sup>[38]</sup> The following data processing was performed using RELION-3.0-beta2.

For the amthamine- $H_2R$ - $G_s$  complex, automated particle picking yielded 3361930 particles. The particles extracted from the dataset were downsampled two times and subjected to 2D classification. The map of the GPBAR- $G_s$  complex (EMD-30344),<sup>[14c]</sup> low-pass filtered to 40 Å, was used as an initial reference model for two rounds of 3D classification, resulting in a well-defined subset of 857906 particles. The particles were then re-extracted with the original pixel size and performed further 3D classification focusing on the receptor- $G_s$  complex, which produced a good subset accounting for 368447 particles. After 3D refinement, CTF refinement, and Bayesian polishing of the final particles, the final refinement map was

generated with a global resolution of 2.7 Å at a Fourier shell correlation of 0.143.

For the histamine-H<sub>3</sub>R-G<sub>i</sub> complex, template-based particle selection produced 2487329 particles. After 2D classification and two rounds of 3D classification using the 5-HT<sub>1D</sub>-G<sub>i</sub> complex (EMD-30974)<sup>[25a]</sup> low-pass filtered map as an initial reference model, a well-defined subset with 742110 particles was selected. Further 3D classifications focusing on the alignment of the receptor-G<sub>i</sub> complex produced a good subset accounting for 234203 particles, which were subsequently subjected to 3D refinement, CTF refinement, and Bayesian polishing. The final refinement generated a map with an indicated global resolution of 2.7 Å at a Fourier shell correlation of 0.143.

For the immepip-H<sub>3</sub>R-G<sub>i</sub> complex, template-based particle selection produced 2050983 particles. After 2D classification and two rounds of 3D classification using the histamine-H<sub>3</sub>R-G<sub>i</sub> complex low-pass filtered map as an initial reference model, two well-defined subsets with 697323 particles were selected. Further 3D classifications focusing on alignment of the receptor produced a good subset accounting for 330449 particles, which were subsequently subjected to 3D refinement, CTF refinement, and Bayesian polishing. The final refinement generated a map with an indicated global resolution of 3.0 Å at a Fourier shell correlation of 0.143.

**Model Building and Refinement:** For the H<sub>2</sub>R-G<sub>s</sub> complexes, the initial model of H<sub>2</sub>R was downloaded from the activated homology models of H<sub>2</sub>R from GPCRdb.<sup>[31]</sup> The initial G<sub>s</sub> and Nb35 complex was generated from the GPBAR-G<sub>s</sub> complex (PDB ID: 7CFM).<sup>[14c]</sup> For the H<sub>3</sub>R-G<sub>i</sub> complex, the initial model of H<sub>3</sub>R was downloaded from GPCRdb as H<sub>2</sub>R. The initial G<sub>i</sub> and scFv16 complex was generated from the 5-HT<sub>1D</sub>-G<sub>i</sub> complex (PDB ID: 7E32).<sup>[25a]</sup> Agonist and lipid coordinates and geometry restraints were generated using phenix.elbow. Then, the models were docked into the cryo-EM density map using Chimera. After the initial docked models were refined using Rosetta, the models were subjected to iterative rounds of manual adjustment and auto-refinement in Coot and Phenix, respectively. The final refinement scores were validated by the module “comprehensive validation (cryo-EM)” in Phenix. Structure figures were prepared using PyMOL, Chimera, and ChimeraX.

**Molecular Dynamics Simulations:** First, the histamine bound to H<sub>3</sub>R model was subtracted from histamine-H<sub>3</sub>R-G<sub>i</sub> complex. Second, the histamine was also positioned in a similar direction to the histamine-H<sub>1</sub>R structure and fitted to the cryo-EM density. The orientations of the receptors were calculated by the Positioning of Proteins in Membranes (PPM) Web Server.<sup>[39]</sup> Following these steps, the whole systems were prepared using the CHARMM-GUI and embedded in a bilayer consisting of 200 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids using replacement methods.<sup>[40]</sup> The membrane systems were then solvated into a periodic TIP3P water box supplemented with 0.15 M NaCl. The CHARMM36m Force Field was used to model protein molecules, and the CHARMM General Force Field (CGenFF) was used for the agonist histamine.

Subsequently, two systems were subjected to minimization for 10 000 steps using the conjugate gradient algorithm. They were then heated and equilibrated at 310.13 K and 1 atm for 200 ps with 10.0 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic restraints in the NAMD 2.13 software. This was followed by 5 cycles of equilibration for 2 ns each at 310.13 K and 1 atm, during which the harmonic restraints were sequentially reduced to 5.0, 2.5, 1.0, 0.5, 0.1 kcal mol<sup>-1</sup> Å<sup>-2</sup>. Production simulations were run at 310.13 K and 1 atm in the NPT ensemble using the Langevin thermostat and Nose-Hoover method for 250 ns. Electrostatic interactions were calculated using the particle mesh Ewald (PME) method with a cutoff of 12 Å. Throughout the final stages of equilibration and production, 5.0 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic restraints were placed on the backbone of the receptors due to the flexibility of the second extracellular loop (ICL2) of H<sub>3</sub>R. Trajectories were visualized and analyzed using Visual Molecular Dynamics (VMD) version 1.9.3.

**GloSensor cAMP Assay:** The agonist-induced cAMP accumulation was measured using the GloSensor™ cAMP assay kit (Promega). HEK293T cells were co-transfected with a 3:1 ratio of WT or mutant and the pGloSensor™–22F plasmid by using LIPO (YEASON). After a minimum of 6 h, transfected cells were seeded onto cell adherent reagent (Applygen) coated 384-well culture plates and incubated for more than 12 h at 37 °C in

5% CO<sub>2</sub>. The culture medium was removed and PBS was added to starve the cells. After 20 min later, the PBS was removed, and the cells were incubated with the equilibration medium (CO<sub>2</sub>-Independent Medium (Gibco) with 10% FBS) containing a 4% dilution of the GloSensor™ cAMP reagent stock solution for 30 minutes at 37 °C and then 10 min at RT. To obtain the dose-response curves, serially diluted agonists were added to each well to stimulate the cells. Luminance signal was measured using 200 ms intervals then (TECAN, 25 °C). Dose-responses were generated from the peak response. cAMP accumulation was analyzed by a standard dose-response curve using GraphPad Prism 8.0 (GraphPad Software). EC50 and pEC50± SEM were calculated using nonlinear regression (curve fit). Data are means ± SEM from at least three independent experiments performed in technical triplicates.

**NanoLuc Binary Technology (NanoBIT) Assay:** For the measurement of G-protein dissociation, HEK293T cells were co-transfected with WT or mutant, Gα<sub>i</sub>-LgBit, Gβ, and Gγ-SmBit plasmids at a 6:2:5:5 ratio. Following transfection, cells were seeded onto 96-well culture plates and incubated for over 12 h at 37 °C in 5% CO<sub>2</sub>. Then, the cells were rinsed twice with D-PBS and incubated with 4 nM coelenterazine-400a (Maokangbio) in HBSS supplemented with 5 mM HEPES pH 7.4 and 0.1% BSA for 1 h. The baseline luminance signal was read immediately for 5 cycles (TECAN, 25 °C). To obtain the dose-response curves, serially diluted agonists were added to each well to stimulate the cells. The luminance signal was then measured for another 15 cycles using 500 ms intervals (TECAN, 25 °C). Dose-responses were generated from the 5th response, and the row data was standardized by the baseline and the vehicle group. G-protein activation was analyzed by a standard dose-response curve using GraphPad Prism 8.0 (GraphPad Software). EC50 and pEC50± SEM were calculated using nonlinear regression (curve fit). Data are means ± SEM from at least three independent experiments performed in technical triplicates.

**Enzyme Linked Immunosorbent (ELISA) assay:** To confirm the cell surface expression of H<sub>3</sub>R and its mutants, the ELISA was performed 24 h after transfection using cells plated on 96-white plates. The cells were fixed with 4% formaldehyde for five minutes at RT and then washed once with PBS. Following fixation, the cells were blocked with blocking buffer (1% FBS in PBS) for 2 h at RT. Afterwards, the plates were incubated with a 1:10000 dilution of monoclonal ANTI-FLAG M2-Peroxidase (HRP) antibody (Sigma Aldrich) in blocking buffer for another 0.5 h at RT. After careful washing, 80 μL per well of SuperSignal ELISA Femto Maximum Sensitivity Substrate (ThermoFisher Scientific) solution was added. The luminance signal was measured using 500 ms intervals. The data are presented as means ± SEM from at least three independent experiments performed in technical triplicates.

**Collections of Activated GPCR Data Sets:** For the true GPCR datasets, structures were collected from reported distinct class A GPCR–G-protein complex structures that are primarily coupled to G<sub>s</sub>- or G<sub>i/o</sub>-proteins. Several G<sub>s</sub>- or G<sub>i/o</sub>-coupled receptors that are primarily coupled to G<sub>q</sub> protein, such as CCK<sub>1</sub>R, CCK<sub>2</sub>R, GASR, NK<sub>1</sub>R, and NTSR1, were excluded. The true datasets were enlarged to 20 G<sub>s</sub>- and 34 G<sub>i/o</sub>-coupled structures by adding the four G<sub>s</sub>-coupled and one G<sub>i</sub>-coupled complexes from our unpublished data.

For the GPCRdb homology model datasets, the primary G<sub>s</sub>- or G<sub>i/o</sub>-coupled activated class A GPCR homology models were collected from GPCRdb. However, the accuracy of orphan class A receptors homology models is poor due to the lack of true homologous structures. After removing the orphan receptors, 98 (24 G<sub>s</sub>- and 74 G<sub>i/o</sub>-coupled receptors) GPCRdb homology models were used as training datasets.

**Extractions of Length and Angle:** To conveniently extract the length and angle geometries of receptors, the reference H<sub>3</sub>R model was positioned on the lipid bilayer using the PPM server. The 7TMs of H<sub>3</sub>R were perpendicular to the membrane plane. Then, the structures of true GPCR dataset and GPCRdb homology model dataset were superimposed onto H<sub>3</sub>R. The Z-axis of all receptor models was perpendicular to the membrane plane. The lengths of TM5 and TM6 were measured from the cytosolic tip to 5.50 and 6.50, respectively. The angle of TM5 and TM6 was calculated between the orientation of a helix and the reference direction. The orientation of a helix was obtained by averaging the C = O bond orientation of each amino acid on the helix segment. Considering that an alpha helix turn consists of



3.6 amino acids, we selected 11 amino acids from the cytoplasmic tip of TM5 or TM6 to calculate the average orientation and eliminate the component perpendicular to the helix axis. The reference orientation for TM5 angle calculation is the TM5 orientation of H<sub>3</sub>R, in which the average C = O bond orientation is 22 amino acids instead of 11. As the TM6 angle indicates the TM6 outward movement during receptor activation, the reference orientation for the TM6 angle is the central axis of the receptors, which is the Z-axis that is perpendicular to the membrane plane.

**Data Pre-Processing and Machine Learning Classification:** We utilize the Sklearn package<sup>[41]</sup> to perform the data pre-processing and machine learning tasks. First, the TM6 length is excluded due to the statistical reason previously stated, followed by standardizing, namely, centering and scaling the three-geometries dataset. From above, we construct a training set of (3,n) dimensions where  $n = 98$  (G<sub>s</sub>:24, G<sub>i</sub>:74) is the number of data points in the training set.

For the cross-validation, the stratified 10-fold validation is used on the GPCRdb set. After that, we train the Random Forest classifier on the whole training set with the parameters as follows (unspecified parameters set as default): RandomForestClassifier(max\_depth = 1, n\_estimators = 15, max\_features = 1, class\_weight = {0:74, 1:24}), in which the bagging strategy is used with sample ratio as 0.7. Note that the standardization fit with the training dataset is applied to the true dataset as well. The best model is trained with random seed as 2784 for both bagging and Random Forest.

Subsequently, we apply PCA to fit the training dataset and get a PCA transformation for visualization. The technique of linear dimensionality reduction, known as Principal Component Analysis (PCA), employs the Singular Value Decomposition method to project the data into a subspace of reduced dimensions. We use the PCA implementation in sklearn.decomposition and it can be used with the default parameters. The first two principal components are reserved, which explain over 87% variance of the data.

The resultant principle can be distilled from the decision boundary for approximate categorization. The comprehensive transformation equation transitioning from the raw domain to Principal Component (PC) space is as follows:

$$X = 0.649 \cdot \frac{1}{4} \cdot 1.64 \cdot (\text{tm5\_len} - 20.276) + 0.569 \cdot \frac{1}{5} \cdot 2.64 \cdot (\text{tm5\_tilt} - 8.474) + 0.506 \cdot \frac{1}{9} \cdot 1.56 \cdot (\text{tm6\_outward} - 18.749)$$

$$Y = -0.081 \cdot \frac{1}{4} \cdot 1.64 \cdot (\text{tm5\_len} - 20.276) + -0.609 \cdot \frac{1}{5} \cdot 2.64 \cdot (\text{tm5\_tilt} - 8.474) + 0.789 \cdot \frac{1}{9} \cdot 1.56 \cdot (\text{tm6\_outward} - 18.749)$$

Upon abstracting standardization, we obtain:

$$X = 0.649 \cdot \text{tm5\_len\_std} + 0.569 \cdot \text{tm5\_tilt\_std} + 0.506 \cdot \text{tm6\_outward\_std}$$

$$Y = 0.081 \cdot \text{tm5\_len\_std} + 0.609 \cdot \text{tm5\_tilt\_std} - 0.789 \cdot \text{tm6\_outward\_std}$$

Herein, [0.649 0.569 0.506] and [-0.081 -0.609 0.789] are the corresponding eigenvectors for PC1 and PC2.

The following formula facilitates the direct calculation of the corresponding values:

$$X = 0.156 \cdot \text{tm5\_len} + 0.108 \cdot \text{tm5\_tilt} + 0.055 \cdot \text{tm6\_outward} - 5.110$$

$$Y = -0.020 \cdot \text{tm5\_len} - 0.116 \cdot \text{tm5\_tilt} + 0.086 \cdot \text{tm6\_outward} - 0.239$$

When  $X > 0.7$ , it is highly probable that GPCR will couple with G<sub>s</sub>, indicative of an elongated TM length, a substantial TM tilt, and a significant outward shift in TM6 orientation. Conversely when  $X$  falls below or equals 0.7, there's an increased likelihood of GPCR coupling with G<sub>i</sub> which comprises instances that are either Y.

**Statistical Analysis:** Data were represented as the mean ± SEM values of ≥3 independent experiments. The raw data of the NanoBiT assay was standardized by the baseline and the vehicle group, while the raw data of cAMP and ELISA assay was standardized by the corresponding WT group. The sigmoidal curves of the reporter luciferase assay were analyzed using a standard dose-response curve in GraphPad Prism version 8.02 (GraphPad Software, San Diego, CA, USA). Data sets with two groups were analyzed by Student's t-tests. Data with ≥3 groups were analyzed by one-way analysis of variance (ANOVA), while the violin plots depicting four distinct features of TM5/TM6 in G<sub>s</sub>- and G<sub>i/o</sub>-coupled receptors are calculated using Mann-Whitney U test taking advantage of scikit learn python package. A significance level of  $p < 0.05$  was considered, and the details are provided in the figure legends.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

The authors thank the Cryo-Electron Microscopy Center of Liangzhu laboratory for the cryo-EM data collection. Cryo-EM specimens were examined with the help of Shenghai Chang at the Center of Cryo-Electron Microscopy, Zhejiang University. Sample preparation for cryo-EM studies was supported by Cheng Ma from the Protein Facility, Zhejiang University School of Medicine. This project was partially supported by the Ministry of Science and Technology (2019YFA050880 to Y.Z.), the National Natural Science Foundation of China (92353303 and 32141004 to Y.Z., 32000845 to Q.S., U21A20418 to Z.C., 81973302 to W.H., and 62202426 to M.Z.), National Science Fund for Distinguished Young Scholars Grant (81825022 to J.-P.S.), ST12030-Major Projects (2022ZD0205400 to Q.S.), the "Pioneer" and "Leading Goose" R&D Program of Zhejiang (2024C03147 to Y.Z.), the Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (2020R01006 to Y.Z.), the Key R&D Projects of Zhejiang Province (2021C03039 to Y.Z.), and the Fundamental Research Funds for the Central Universities (Y.Z.).

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

Q.S., X.T., X.W., and S.C. contributed equally to this work. Y.Z. initiated and conceived the study. Y.Z., Z.C., J.-P.S., W.H., and M.Z. supervised the project. Q.S. designed the expression constructs, optimized the sample purification, and prepared the final samples for cryo-EM studies for the H<sub>3</sub>R-G<sub>i</sub> complexes. X.W., P.X., and J.-P.S. designed the expression constructs and prepared the final samples for cryo-EM studies for the H<sub>2</sub>R-G<sub>s</sub> complex. C.M., D.-D.S., and H.Z. evaluated the specimen by negative-stain EM, screened the cryo-EM conditions, prepared the cryo-EM grids, and collected cryo-EM images. Q.S. performed density map calculations, model building, refinement of final models. Y.Z. and Q.S. analyzed the structure, designed the mutations, and functional assays. X.T., X.W., P.X., H.X., L.J., and Y.Z. performed the functional assays. M.Z., Y.Z., S.C., and Q.S. performed machine learning analyses. Q.S. and S.-K.Z. performed molecular dynamics simulations. Q.S., X.T., S.C., and X.W. prepared the figures. Y.Z. and Q.S. wrote the manuscript. Z.C., J.-P.S., W.H., and M.Z. provided important discussions and essential revisions.

## Data Availability Statement

The Electron Microscopy Data Bank accession numbers and Protein Data Bank identifiers are EMD-39582 and PDB ID 8YUT for the amphetamine-H<sub>2</sub>R-G<sub>s</sub> complex, EMD-39583 and PDB ID 8YUU for the histamine-H<sub>3</sub>R-G<sub>i</sub> complex, and EMD-39584 and PDB ID 8YUV for the immpip-H<sub>3</sub>R-G<sub>i</sub> complex, respectively. Other data that support the findings of this study are available in the supplementary material of this article.

## Keywords

cryo-EM structure, G protein selectivity, GPCR, H<sub>2</sub>R, H<sub>3</sub>R, machine learning, signaling complex

Received: December 22, 2023

Revised: April 2, 2024

Published online: April 22, 2024

- [1] a) T. K. Bjarnadottir, D. E. Gloriam, S. H. Hellstrand, H. Kristiansson, R. Fredriksson, H. B. Schioth, *Genomics* **2006**, *88*, 263; b) T. Flock, A. S. Hauser, N. Lund, D. E. Gloriam, S. Balaji, M. M. Babu, *Nature* **2017**, *545*, 317.
- [2] a) A. S. Hauser, C. Avet, C. Normand, A. Mancini, A. Inoue, M. Bouvier, D. E. Gloriam, *Elife* **2022**, *11*, 74107; b) S. R. Neves, P. T. Ram, R. Iyengar, *Science* **2002**, *296*, 1636.
- [3] D. Wacker, R. C. Stevens, B. L. Roth, *Cell* **2017**, *170*, 414.
- [4] J. W. Wisler, H. A. Rockman, R. J. Lefkowitz, *Circulation* **2018**, *137*, 2315.
- [5] a) A. Inoue, F. Raimondi, F. M. N. Kadji, G. Singh, T. Kishi, A. Uwamizu, Y. Ono, Y. Shinjo, S. Ishida, N. Arang, K. Kawakami, J. S. Gutkind, J. Aoki, R. B. Russell, *Cell* **2019**, *177*, 1933; b) M. Sandhu, A. Cho, N. Ma, E. Mukhaleva, Y. Namkung, S. Lee, S. Ghosh, J. H. Lee, D. E. Gloriam, S. A. Laporte, M. M. Babu, N. Vaidehi, *Nat. Commun.* **2022**, *13*, 7428.
- [6] a) Y. Du, N. M. Duc, S. G. F. Rasmussen, D. Hilger, X. Kubiak, L. Wang, J. Bohon, H. R. Kim, M. Wegrecki, A. Asuru, K. M. Jeong, J. Lee, M. R. Chance, D. T. Lodowski, B. K. Kobilka, K. Y. Chung, *Cell* **2019**, *177*, 1232; b) S. Anantakrishnan, A. N. Naganathan, *Nat. Commun.* **2023**, *14*, 128.
- [7] a) C. Avet, A. Mancini, B. Breton, C. Le Gouill, A. S. Hauser, C. Normand, H. Kobayashi, F. Gross, M. Hogue, V. Lukasheva, S. St-Onge, M. Carrier, M. Heroux, S. Morissette, E. B. Fauman, J. P. Fortin, S. Schann, X. Leroy, D. E. Gloriam, M. Bouvier, *Elife* **2022**, *11*, 74101; b) S. D. Harding, J. F. Armstrong, E. Faccenda, C. Southan, S. P. H. Alexander, A. P. Davenport, A. J. Pawson, M. Spedding, J. A. Davies, I. Nc, *Nucleic Acids Res.* **2022**, *50*, D1282.
- [8] N. R. Latorraca, A. J. Venkatakrisnan, R. O. Dror, *Chem. Rev.* **2017**, *117*, 139.
- [9] a) Y. Zhang, B. Sun, D. Feng, H. Hu, M. Chu, Q. Qu, J. T. Tarrasch, S. Li, T. Sun Kobilka, B. K. Kobilka, G. Skiniotis, *Nature* **2017**, *546*, 248; b) Y. L. Liang, M. Khoshouei, M. Radjainia, Y. Zhang, A. Glukhova, J. Tarrasch, D. M. Thal, S. G. B. Furness, G. Christopoulos, T. Coudrat, R. Danev, W. Baumeister, L. J. Miller, A. Christopoulos, B. K. Kobilka, D. Wootton, G. Skiniotis, P. M. Sexton, *Nature* **2017**, *546*, 118; c) A. Glukhova, C. J. Draper-Joyce, R. K. Sunahara, A. Christopoulos, D. Wootton, P. M. Sexton, *ACS Pharmacol. Transl. Sci.* **2018**, *1*, 73; d) J. Garcia-Nafria, C. G. Tate, *Mol. Cell. Endocrinol.* **2019**, *488*, 1.
- [10] a) C. Mao, P. Xiao, X. N. Tao, J. Qin, Q. T. He, C. Zhang, S. C. Guo, Y. Q. Du, L. N. Chen, D. D. Shen, Z. S. Yang, H. Q. Zhang, S. M. Huang, Y. H. He, J. Cheng, Y. N. Zhong, P. Shang, J. Chen, D. L. Zhang, Q. L. Wang, M. X. Liu, G. Y. Li, Y. Guo, H. E. Xu, C. Wang, C. Zhang, S. Feng, X. Yu, Y. Zhang, J. P. Sun, *Science* **2023**, *380*, eadd6220; b) Y. Ji, J. Duan, Q. Yuan, X. He, G. Yang, S. Zhu, K. Wu, W. Hu, T. Gao, X. Cheng, H. Jiang, H. E. Xu, Y. Jiang, *Nat. Commun.* **2023**, *14*, 1268; c) Q. Liu, D. Yang, Y. Zhuang, T. I. Croll, X. Cai, A. Dai, X. He, J. Duan, W. Yin, C. Ye, F. Zhou, B. Wu, Q. Zhao, H. E. Xu, M. W. Wang, Y. Jiang, *Nat. Chem. Biol.* **2021**, *17*, 1238; d) K. O. Alegre, N. Paknejad, M. Su, J. S. Lou, J. Huang, K. D. Jordan, E. T. Eng, J. R. Meyerson, R. K. Hite, X. Y. Huang, *Nat. Struct. Mol. Biol.* **2021**, *28*, 936; e) X. Zhang, C. He, M. Wang, Q. Zhou, D. Yang, Y. Zhu, W. Feng, H. Zhang, A. Dai, X. Chu, J. Wang, Z. Yang, Y. Jiang, U. Sensfuss, Q. Tan, S. Han, S. Reedtz-Runge, H. E. Xu, S. Zhao, M. W. Wang, B. Wu, Q. Zhao, *Nat. Chem. Biol.* **2021**, *17*, 1230.
- [11] S. Huang, P. Xu, D. D. Shen, I. A. Simon, C. Mao, Y. Tan, H. Zhang, K. Harpsøe, H. Li, Y. Zhang, C. You, X. Yu, Y. Jiang, Y. Zhang, D. E. Gloriam, H. E. Xu, *Mol. Cell* **2022**, *82*, 2681.
- [12] M. Seyedabadi, M. Gharghabi, E. V. Gurevich, V. V. Gurevich, *Trends Biochem. Sci.* **2022**, *47*, 570.
- [13] M. Matic, P. Miglionico, M. Tatsumi, A. Inoue, F. Raimondi, *Nat. Commun.* **2023**, *14*, 4361.
- [14] a) F. Sadler, N. Ma, M. Ritt, Y. Sharma, N. Vaidehi, S. Sivaramakrishnan, *Nature* **2023**, *615*, 734; b) X. Teng, S. Chen, Q. Wang, Z. Chen, X. Wang, N. Huang, S. Zheng, *Sci. Adv.* **2022**, *8*, eabo4158; c) F. Yang, C. Mao, L. Guo, J. Lin, Q. Ming, P. Xiao, X. Wu, Q. Shen, S. Guo, D. D. Shen, R. Lu, L. Zhang, S. Huang, Y. Ping, C. Zhang, C. Ma, K. Zhang, X. Liang, Y. Shen, F. Nan, F. Yi, V. C. Luca, J. Zhou, C. Jiang, J. P. Sun, X. Xie, X. Yu, Y. Zhang, *Nature* **2020**, *587*, 499; d) K. K. Kumar, M. Shalev-Benami, M. J. Robertson, H. Hu, S. D. Banister, S. A. Hollingsworth, N. R. Latorraca, H. E. Kato, D. Hilger, S. Maeda, W. I. Weis, D. L. Farrens, R. O. Dror, S. V. Malhotra, B. K. Kobilka, G. Skiniotis, *Cell* **2019**, *176*, 448; e) A. Koehl, H. Hu, S. Maeda, Y. Zhang, Q. Qu, J. M. Paggi, N. R. Latorraca, D. Hilger, R. Dawson, H. Matile, G. F. X. Schertler, S. Granier, W. I. Weis, R. O. Dror, A. Manglik, G. Skiniotis, B. K. Kobilka, *Nature* **2018**, *558*, 547.
- [15] a) Y. Hattori, K. Hattori, N. Matsuda, *Handb Exp Pharmacol* **2017**, *241*, 239; b) H. L. Haas, O. A. Sergeeva, O. Selbach, *Physiol. Rev.* **2008**, *88*, 1183.
- [16] I. Obara, V. Telezhkin, I. Alrashdi, P. L. Chazot, *Br. J. Pharmacol.* **2020**, *177*, 580.
- [17] F. E. Simons, *N. Engl. J. Med.* **2004**, *351*, 2203.
- [18] L. Jiang, L. Cheng, H. Chen, H. Dai, D. An, Q. Ma, Y. Zheng, X. Zhang, W. Hu, Z. Chen, *J. Exp. Med.* **2021**, *218*.
- [19] a) Y. Dauvilliers, C. Bassetti, G. J. Lammers, I. Arnulf, G. Mayer, A. Rodenbeck, P. Leher, C. L. Ding, J. M. Lecomte, J. C. Schwartz, H. I. s. group, *Lancet Neurol.* **2013**, *12*, 1068; b) J. S. Lin, Y. Dauvilliers, I. Arnulf, H. Bastuji, C. Anaclet, R. Parmentier, L. Kocher, M. Yanagisawa, P. Leher, X. Ligneau, D. Perrin, P. Robert, M. Roux, J. M. Lecomte, J. C. Schwartz, *Neurobiol Dis* **2008**, *30*, 74.
- [20] a) R. Xia, N. Wang, Z. Xu, Y. Lu, J. Song, A. Zhang, C. Guo, Y. He, *Nat. Commun.* **2021**, *12*, 2086; b) X. Peng, L. Yang, Z. Liu, S. Lou, S. Mei, M. Li, Z. Chen, H. Zhang, *Nat. Commun.* **2022**, *13*, 6105; c) M. J. Robertson, M. M. Papesergi-Scott, F. He, A. B. Seven, J. G. Meyerowitz, O. Panova, M. C. Peroto, T. Che, G. Skiniotis, *Nat. Struct. Mol. Biol.* **2022**, *29*, 1188; d) T. Shimamura, M. Shiroishi, S. Weyand, H. Tsujimoto, G. Winter, V. Katritch, R. Abagyan, V. Cherezov, W. Liu, G. W. Han, T. Kobayashi, R. C. Stevens, S. Iwata, *Nature* **2011**, *475*, 65; e) D. Im, J. I. Kishikawa, Y. Shiimura, H. Hisano, A. Ito, Y. Fujita-Fujiharu, Y. Sugita, T. Noda, T. Kato, H. Asada, S. Iwata, *Nat. Commun.* **2023**, *14*, 6538.
- [21] S. G. Rasmussen, B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski, J. M. Mathiesen, S. T. Shah, J. A. Lyons, M. Caffrey, S. H. Gellman, J. Steyaert, G. Skiniotis, W. I. Weis, R. K. Sunahara, B. K. Kobilka, *Nature* **2011**, *477*, 549.
- [22] J. Duan, D. D. Shen, X. E. Zhou, P. Bi, Q. F. Liu, Y. X. Tan, Y. W. Zhuang, H. B. Zhang, P. Y. Xu, S. J. Huang, S. S. Ma, X. H. He, K. Melcher, Y. Zhang, H. E. Xu, Y. Jiang, *Nat. Commun.* **2020**, *11*, 4121.
- [23] S. Maeda, A. Koehl, H. Matile, H. Hu, D. Hilger, G. F. X. Schertler, A. Manglik, G. Skiniotis, R. J. P. Dawson, B. K. Kobilka, *Nat. Commun.* **2018**, *9*, 3712.
- [24] J. A. Ballesteros, H. Weinstein, in *Methods in Neurosciences*, Vol. 25, Elsevier, Amsterdam, Netherlands **1995**, pp. 366–428.
- [25] a) P. Xu, S. Huang, H. Zhang, C. Mao, X. E. Zhou, X. Cheng, I. A. Simon, D. D. Shen, H. Y. Yen, C. V. Robinson, K. Harpsøe, B. Svensson, J. Guo, H. Jiang, D. E. Gloriam, K. Melcher, Y. Jiang, Y. Zhang, H. E. Xu, *Nature* **2021**, *592*, 469; b) D. Yuan, Z. Liu, J. Kaindl, S. Maeda, J. Zhao, X. Sun, J. Xu, P. Gmeiner, H. W. Wang, B. K. Kobilka, *Nat. Chem. Biol.* **2020**, *16*, 507; c) P. Xiao, W. Yan, L. Gou, Y. N. Zhong, L. Kong, C. Wu, X. Wen, Y. Yuan, S. Cao, C. Qu, X. Yang, C. C. Yang, A. Xia, Z. Hu, Q. Zhang, Y. H. He, D. L. Zhang, C. Zhang, G. H. Hou, H. Liu, L. Zhu, P. Fu, S. Yang, D. M. Rosenbaum, J. P. Sun, Y. Du, L. Zhang, X. Yu, Z. Shao, *Cell* **2021**, *184*, 943.
- [26] X. Xu, J. Kaindl, M. J. Clark, H. Hubner, K. Hirata, R. K. Sunahara, P. Gmeiner, B. K. Kobilka, X. Liu, *Cell Res.* **2021**, *31*, 569.
- [27] T. Flock, C. N. J. Ravarani, D. Sun, A. J. Venkatakrisnan, M. Kayicki, C. G. Tate, D. B. Veprintsev, M. M. Babu, *Nature* **2015**, *524*, 173.

- [28] a) X. Lin, M. Li, N. Wang, Y. Wu, Z. Luo, S. Guo, G. W. Han, S. Li, Y. Yue, X. Wei, X. Xie, Y. Chen, S. Zhao, J. Wu, M. Lei, F. Xu, *Nature* **2020**, 579, 152; b) H. Zhang, L. N. Chen, D. Yang, C. Mao, Q. Shen, W. Feng, D. D. Shen, A. Dai, S. Xie, Y. Zhou, J. Qin, J. P. Sun, D. H. Scharf, T. Hou, T. Zhou, M. W. Wang, Y. Zhang, *Cell Res.* **2021**, 31, 1163; c) S. Ma, Y. Chen, A. Dai, W. Yin, J. Guo, D. Yang, F. Zhou, Y. Jiang, M. W. Wang, H. E. Xu, *Cell Res.* **2021**, 31, 1061; d) Y. Zhuang, P. Xu, C. Mao, L. Wang, B. Krumm, X. E. Zhou, S. Huang, H. Liu, X. Cheng, X. P. Huang, D. D. Shen, T. Xu, Y. F. Liu, Y. Wang, J. Guo, Y. Jiang, H. Jiang, K. Melcher, B. L. Roth, Y. Zhang, C. Zhang, H. E. Xu, *Cell* **2021**, 184, 931; e) M. Su, L. Zhu, Y. Zhang, N. Paknejad, R. Dey, J. Huang, M. Y. Lee, D. Williams, K. D. Jordan, E. T. Eng, O. P. Ernst, J. R. Meyerson, R. K. Hite, T. Walz, W. Liu, X. Y. Huang, *Mol. Cell* **2020**, 80, 59; f) F. Yang, S. Ling, Y. Zhou, Y. Zhang, P. Lv, S. Liu, W. Fang, W. Sun, L. A. Hu, L. Zhang, P. Shi, C. Tian, *National Science Review* **2020**, 8; g) C. Nagiri, K. Kobayashi, A. Tomita, M. Kato, K. Kobayashi, K. Yamashita, T. Nishizawa, A. Inoue, W. Shihoya, O. Nureki, *Mol. Cell* **2021**, 81, 3205.
- [29] a) C. Zhao, L. Cheng, W. Wang, H. Wang, Y. Luo, Y. Feng, X. Wang, H. Fu, Y. Cai, S. Yang, P. Fu, W. Yan, Z. Shao, *Cell Res.* **2022**, 32, 218; b) J. Garcia-Nafria, R. Nehme, P. C. Edwards, C. G. Tate, *Nature* **2018**, 558, 620.
- [30] a) H. H. Okamoto, H. Miyauchi, A. Inoue, F. Raimondi, H. Tsujimoto, T. Kusakizako, W. Shihoya, K. Yamashita, R. Suno, N. Nomura, T. Kobayashi, S. Iwata, T. Nishizawa, O. Nureki, *Nat. Struct. Mol. Biol.* **2021**, 28, 694; b) S. Maeda, Q. Qu, M. J. Robertson, G. Skiniotis, B. K. Kobilka, *Science* **2019**, 364, 552; c) S. Huang, P. Xu, Y. Tan, C. You, Y. Zhang, Y. Jiang, H. E. Xu, *Cell Res.* **2021**, 31, 1036; d) Z. Shao, Q. Shen, B. Yao, C. Mao, L. N. Chen, H. Zhang, D. D. Shen, C. Zhang, W. Li, X. Du, F. Li, H. Ma, Z. H. Chen, H. E. Xu, S. Ying, Y. Zhang, H. Shen, *Nat. Chem. Biol.* **2022**, 18, 264; e) Z. Shao, Y. Tan, Q. Shen, L. Hou, B. Yao, J. Qin, P. Xu, C. Mao, L. N. Chen, H. Zhang, D. D. Shen, C. Zhang, W. Li, X. Du, F. Li, Z. H. Chen, Y. Jiang, H. E. Xu, S. Ying, H. Ma, Y. Zhang, H. Shen, *Cell Discov.* **2022**, 8, 44.
- [31] G. Pandey-Szekeress, J. Caroli, A. Mamyrbekov, A. A. Kermani, G. M. Keseru, A. J. Kooistra, D. E. Gloriam, *Nucleic Acids Res.* **2023**, 51, D395.
- [32] S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.* **1987**, 2, 37.
- [33] H. T. Kam, in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, **1995**, 1, pp. 278–282.
- [34] N. H. Jensen, R. M. Rodriguiz, M. G. Caron, W. C. Wetsel, R. B. Rothman, B. L. Roth, *Neuropsychopharmacology* **2008**, 33, 2303.
- [35] a) Y. Q. Ping, C. Mao, P. Xiao, R. J. Zhao, Y. Jiang, Z. Yang, W. T. An, D. D. Shen, F. Yang, H. Zhang, C. Qu, Q. Shen, C. Tian, Z. J. Li, S. Li, G. Y. Wang, X. Tao, X. Wen, Y. N. Zhong, J. Yang, F. Yi, X. Yu, H. E. Xu, Y. Zhang, J. P. Sun, *Nature* **2021**, 589, 620; b) X. Qi, L. Friedberg, R. De Bose-Boyd, T. Long, X. Li, *Nat. Chem. Biol.* **2020**, 16, 1368.
- [36] Y. L. Liang, P. Zhao, C. Draper-Joyce, J. A. Baltos, A. Glukhova, T. T. Truong, L. T. May, A. Christopoulos, D. Wootten, P. M. Sexton, S. G. B. Furness, *ACS Pharmacol. Transl. Sci.* **2018**, 1, 12.
- [37] S. Q. Zheng, E. Palovcak, J. P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, *Nat. Methods* **2017**, 14, 331.
- [38] K. Zhang, *J. Struct. Biol.* **2016**, 193, 1.
- [39] M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, A. L. Lomize, *Nucleic Acids Res.* **2012**, 40, D370.
- [40] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks 3rd, A. D. MacKerell Jr., J. B. Klauda, W. Im, *J. Chem. Theory Comput.* **2016**, 12, 405.
- [41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. J. t. J. o. m. L. r. Dubourg, **2011**, 12, 2825.