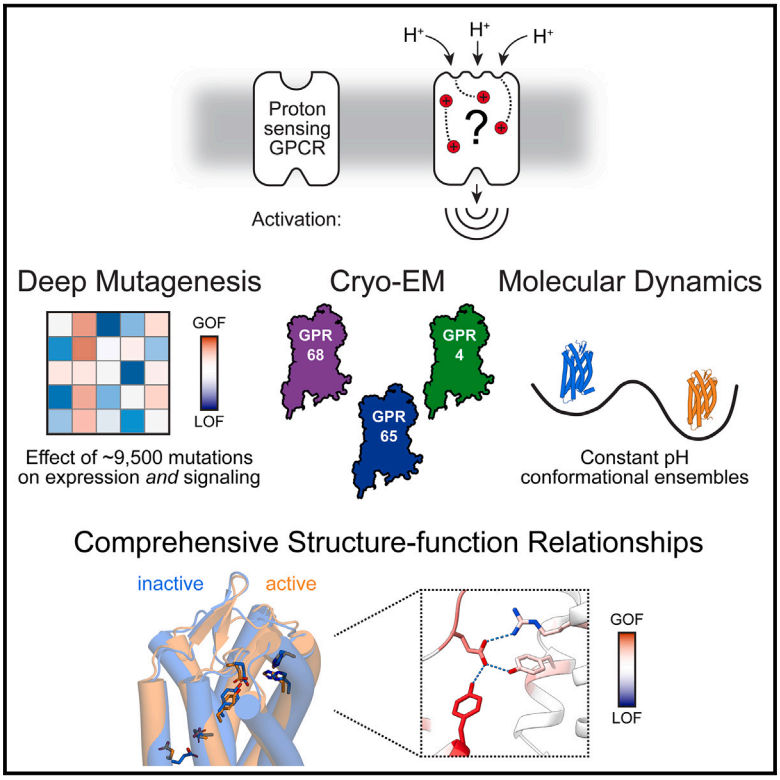
Cell

Molecular basis of proton sensing by G protein-coupled receptors

Graphical abstract



Highlights

- Cryo-EM structures of three human proton-sensing GPCRs: GPR4, GPR65, and GPR68
- Multiphenotypic deep mutational scanning uncovers functional hotspots in GPR68
- Constant-pH molecular dynamics simulations of GPR68 link protonation and conformation
- Structures, mutational scanning, and simulations illuminate the proton-sensing mechanism

Authors

Matthew K. Howard, Nicholas Hoppe, Xi-Ping Huang, ..., Justin G. English, Willow Coyote-Maestas, Aashish Manglik

Correspondence

willow.coyote-maestas@ucsf.edu (W.C.-M.), aashish.manglik@ucsf.edu (A.M.)

In brief

Howard et al. combine mechanistic deep mutational scanning, cryo-EM, and constant-pH molecular dynamics simulations to provide a holistic view of proton activation in human pH-sensitive GPCRs.

Howard et al., 2025, Cell *188*, 671–687
February 6, 2025
https://doi.org/10.1016/j.cell.2024.11.036

Cell

CellPress

Article

Molecular basis of proton sensing by G protein-coupled receptors

Matthew K. Howard,[1,2,3,13] Nicholas Hoppe,[2,4,13] Xi-Ping Huang,[5,14] Darko Mitrovic,[6,14] Christian B. Billesbølle,[2] Christian B. Macdonald,[3] Eshan Mehrotra,[1,2,7] Patrick Rockefeller Grimes,[3] Donovan D. Trinidad,[8] Lucie Delemotte,[6] Justin G. English,[9] Willow Coyote-Maestas,[3,10,11,*] and Aashish Manglik[2,10,11,12,15,*]

[1]Tetrad graduate program, University of California, San Francisco, San Francisco, CA 94143, USA
[2]Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94143, USA
[3]Department of Bioengineering and Therapeutic Science, University of California, San Francisco, San Francisco, CA 94143, USA
[4]Biophysics graduate program, University of California, San Francisco, San Francisco, CA 94143, USA
[5]Department of Pharmacology and the National Institute of Mental Health Psychoactive Drug Screening Program (NIMH PDSP), The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[6]Science for Life Laboratory, Department of Applied Physics, KTH Royal Institute of Technology, 12121 Solna, Stockholm, Stockholm County 114 28, Sweden
[7]Medical Scientist Training Program, University of California, San Francisco, San Francisco, CA 94143, USA
[8]Department of Medicine, Division of Infectious Disease, University of California, San Francisco, San Francisco, CA 94143, USA
[9]Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84112, USA
[10]Chan Zuckerberg Biohub, San Francisco, CA 94148, USA
[11]Quantitative Biosciences Institute, University of California, San Francisco, San Francisco, CA 94143, USA
[12]Department of Anesthesia and Perioperative Care, University of California, San Francisco, San Francisco, CA 94115, USA
[13]These authors contributed equally
[14]These authors contributed equally
[15]Lead contact
*Correspondence: willow.coyote-maestas@ucsf.edu (W.C.-M.), aashish.manglik@ucsf.edu (A.M.)
https://doi.org/10.1016/j.cell.2024.11.036

SUMMARY

Three proton-sensing G protein-coupled receptors (GPCRs)—GPR4, GPR65, and GPR68—respond to extracellular pH to regulate diverse physiology. How protons activate these receptors is poorly understood. We determined cryogenic-electron microscopy (cryo-EM) structures of each receptor to understand the spatial arrangement of proton-sensing residues. Using deep mutational scanning (DMS), we determined the functional importance of every residue in GPR68 activation by generating ~9,500 mutants and measuring their effects on signaling and surface expression. Constant-pH molecular dynamics simulations provided insights into the conformational landscape and protonation patterns of key residues. This unbiased approach revealed that, unlike other proton-sensitive channels and receptors, no single site is critical for proton recognition. Instead, a network of titratable residues extends from the extracellular surface to the transmembrane region, converging on canonical motifs to activate proton-sensing GPCRs. Our approach integrating structure, simulations, and unbiased functional interrogation provides a framework for understanding GPCR signaling complexity.

INTRODUCTION

Homeostatic acid-base balance is vital for cellular and tissue physiology and relies on sensing of extracellular pH. Diverse cell surface proteins respond to changes in extracellular pH. While most cell surface proton sensors are ion channels,[1–5] three G protein-coupled receptors (GPCRs) respond to extracellular pH: GPR4, GPR65, and GPR68.[6,7] These receptors are expressed in diverse cells that regulate central pH homeostasis,[8] pH sensing in the immune system,[9–11] and vascular responses to pH.[12]

Given the relevance of proton-sensing GPCRs to pH-dependent physiology, it is important to understand how these receptors work at the molecular level. For pH-sensing ion channels and transporters, a defined cluster of polar and charged residues is often ascribed as the proton recognition site.[1–8,10,13] This view is controversial because it is often challenging to completely abolish proton sensitivity with targeted mutagenesis.[14,15] There are several proposed models for proton recognition by proton-sensing GPCRs. Proton-sensing GPCRs harbor many extracellular histidine residues that likely titrate at physiologically relevant pH levels, and initial studies ascribed these histidines as critical for proton sensing.[7,16] However, mutational studies in GPR68 showed that removal of all extracellular histidines does not fully abolish proton-driven activation.[14,15] A recent study

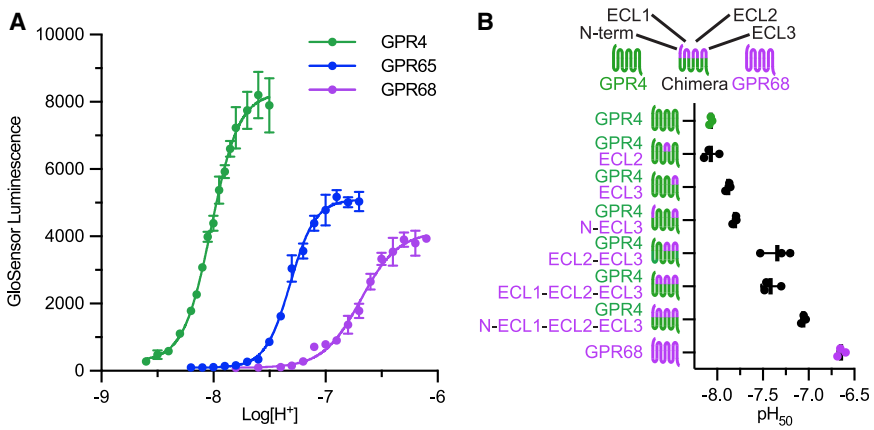


Figure 1. Chimeric pH sensors reveal distributed proton sensing

(A) GloSensor cAMP accumulation assay demonstrates GPR4, GPR65, and GPR68, respond to decreasing pH.

(B) GloSensor cAMP accumulation assay of GPR4-GPR68 chimeric receptors. Extracellular segments of GPR68 were grafted onto GPR4. A sequence alignment of GPR4 and GPR68 indicating swapped segments is available in Figure S1. Data shown in (A) and (C) are from three or four independent biological replicates ± SD. See also Table S1 and Figure S1.

employing parallel mutagenesis of titratable residues in GPR68 identified a conserved triad of buried acidic residues in proton-sensing GPCRs.[15] Here, too, neutral mutations to these sites shift pH_{50} but do not abolish proton-mediated receptor activation. How protons activate proton-sensing GPCRs remains poorly defined.

Structural biology combined with molecular dynamics simulations has revealed molecular recognition of diverse GPCR stimuli ranging from light, ions, small molecules, peptides, and large proteins.[17–24] However, individual protons are not readily resolved by modern X-ray crystallography and cryogenic-electron microscopy (cryo-EM) approaches, and the precise positions of protons remain challenging to determine by conventional molecular dynamics simulation approaches. An ideal alternative would be comprehensive data for how every single residue contributes to proton sensation. Unfortunately, conventional mutagenesis does not scale to unbiased testing of all residues within a GPCR.

Deep mutational scanning (DMS) has emerged as a powerful method to probe protein function.[25] In DMS, comprehensive mutagenesis is combined with a sequencing-based pooled assay to systematically measure how individual substitutions at every single position in a protein affect function. A key requirement for DMS is a robust phenotype that can provide a mechanistic readout in a pooled cellular assay. DMS has uncovered the molecular basis of protein function, folding, and allostery.[26–29] For GPCRs, previous DMS studies have uncovered important residue-level contributions to cell surface expression[30–32] or signaling.[33,34] Conventional mutagenesis studies, however, routinely highlight that GPCR mutations influence both cell surface expression (e.g., due to changes in synthesis, folding, or trafficking) and cellular signaling (either via direct effects on stimulus recognition, allosteric communication, or signaling effector coupling). To quantify how mutations influence signaling requires untangling the contribution of mutational effects on surface expression vs. signaling.

Here, we integrate structural biology, molecular dynamics simulations, and DMS to understand how protons activate proton-sensing GPCRs. We identify critical residues responsible for proton sensing and for allosteric activation of G protein signaling. This combined approach yielded a comprehensive structure-function model for how a stimulus activates a GPCR.

RESULTS

Receptor chimeras reveal distributed proton sensing

We first investigated whether a single, conserved site confers proton sensitivity in proton-sensing GPCRs analogous to proton-sensing channels and transporters.[1,3,4,35–37] In HEK293 cells, GPR4, GPR65, and GPR68 activate cyclic AMP (cAMP) signaling with distinct pH_{50} values of 8.0, 7.4, and 6.7, respectively (Figure 1A). If a single site is responsible for proton sensing, swapping segments of one pH-sensing receptor for another would yield concordant changes in pH_{50}. We chose GPR4 and GPR68 for this experiment as they have the highest sequence identity (44%) but the largest difference in pH_{50}.

We designed chimeric receptors by grafting segments of the GPR68 extracellular regions onto GPR4 (Figures 1B, S1A, and S1B). Each linear segment was tested individually and combined with other segments in a cAMP accumulation assay (Figure S1C). Out of the 15 constructs, 6 failed to show any proton-dependent signaling response (Figure S1C), even when expression normalized to wild-type constructs (Figures S1D and S1E).

Chimeric constructs bearing single segments of GPR68 had little effect on pH_{50} (Figures 1B and S1C; Table S1). Introducing two segments of GPR68 into GPR4 also had little effect on pH_{50}, except for the ECL2/ECL3 chimera, which shifts the pH_{50} from 8.0 to 7.5 (Figures 1B and S1C; Table S1). Addition of the GPR68 ECL1 to this ECL2/ECL3 construct did not yield a further shift in pH_{50} (Figures 1B and S1C; Table S1). Paradoxically, the addition of the GPR68 N terminus to the ECL2/ECL3 construct restored pH_{50} to 8.0 (Figure S1C; Table S1). A final construct bearing the entire extracellular region of GPR68 grafted onto GPR4 yielded a pH_{50} of 7.1 (Figures 1B and S1C; Table S1).

These chimeric receptor experiments challenge a single-site model of proton sensing in proton-sensing GPCRs. Because swapping the entire extracellular region of GPR68 is required for a pH_{50} that approaches that of native GPR68, we conclude that a network of proton-sensitive sites encompassing multiple extracellular surface and transmembrane (TM) regions is likely important for receptor activation.

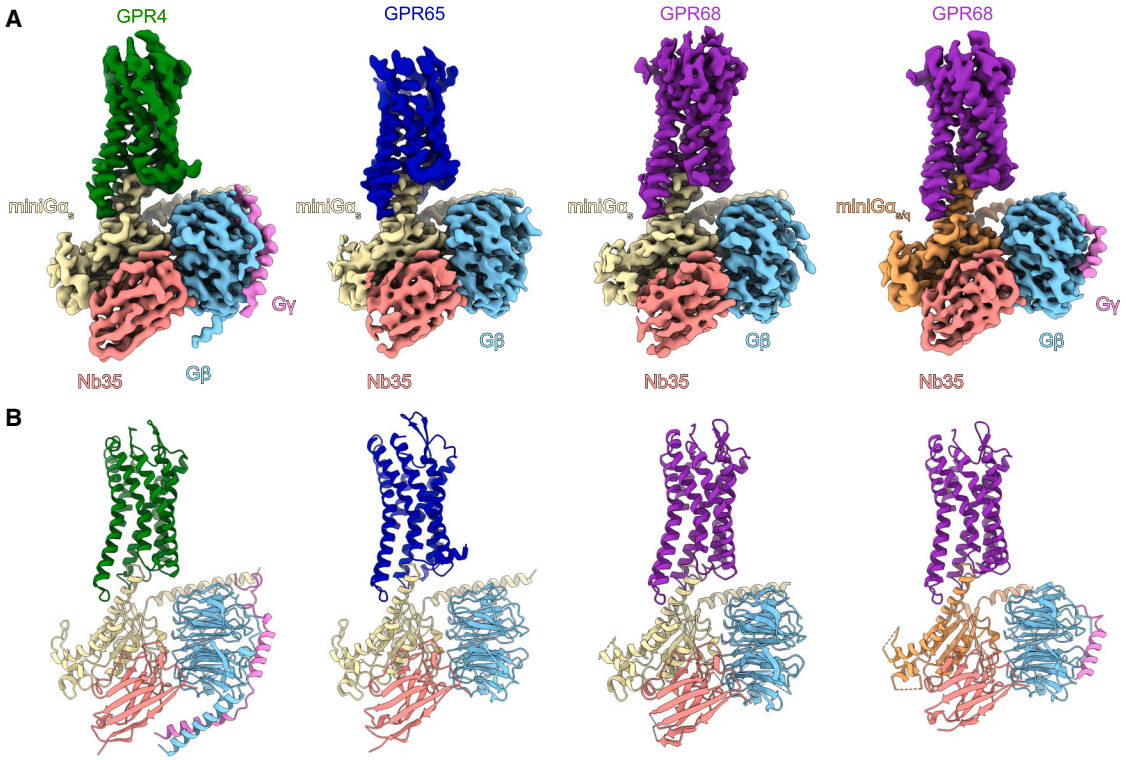


Figure 2. Chimeric pH sensors reveal distributed proton sensing and cryo-EM structures of proton-sensing GPCRs

(A) Cryo-EM density maps of GPR4-miniGα_s, GPR65-miniGα_s, GPR68-miniGα_s, and GPR68-miniG$\alpha_{s/q}$. Each is complexed with G$\beta\gamma$ and Nb35. See Figures S2 and S3 and Table S2.

(B) Model of GPR4, GPR65, and GPR68 G protein complexes. See Figures S2 and S3 and Table S2.

Cryo-EM structures of proton-sensing GPCRs

We next determined cryo-EM structures of human GPR4, GPR65, and GPR68 in complex with heterotrimeric G protein signaling subunits (Figures 2A, S2, and S3). To overcome poor expression, we generated constructs of each receptor fused C-terminally to miniGα proteins.[38] GPR4 and GPR65 drive cAMP production via Gα_s,[7,16,38] and we used miniGα_s to stabilize these receptors. GPR68 signals via Gα_q and Gα_s.[7,39] We therefore used both miniGα_s and a chimeric miniG$\alpha_{s/q}$ construct to obtain structures of GPR68. We also screened different pH values for optimal high-resolution reconstruction of receptor-G protein complexes. Although each receptor activates at distinct pH_{50} values when expressed heterologously in HEK293 cells, we found that purification at pH 6 enabled the best resolution for each receptor during single-particle cryo-EM reconstruction. For GPR68-miniGα_s, we included the positive allosteric modulator MS48107,[14,40] in biochemical preparations. However, our structures did not reveal density for this ligand. Single-particle reconstructions yielded nominal resolutions between 2.8 and 3.0 Å for the complexes (Table S2). To improve reconstructions in the receptor extracellular regions, we performed focused refinements on the 7TM domains (Figures S2 and S3). The resulting maps enabled us to model each proton-sensing GPCR (Figure 2B).

Structures of GPR4, GPR65, and GPR68 bound to miniGα_s revealed similar active-state conformations across the 7TM domains (root-mean-square deviation [RMSD] < 1.5 Å) despite sequence identities of 30%–44% (Figure 3A). The conformation of GPR68 is similar between miniGα_s and miniG$\alpha_{s/q}$, with an RMSD of 0.9 Å. Comparison to inactive and active structures of the prototypical class A GPCR, the β_2-adrenergic receptor (β2AR), shows all proton receptors are captured in fully active conformations (RMSD of each proton receptor to active β2AR is <1.2 Å in the TMs) (Figure 3A). This is reflected in a similar conformation of the common "$P^{5.50}I^{3.40}F^{6.44}$" motif (Figures 3B and 3C), which has slight sequence variations in GPR4, GPR65, and GPR68 (Figures 3B and 3C). In the proton sensors, the conserved GPCR "$N^{7.49}P^{7.50}xxY^{7.53}$" motif harbors an aspartate at the 7.49 position, a substitution shared with ~18% of all human class A GPCRs (Figures 3B and 3D). Finally, each proton receptor substitutes a phenylalanine in the conserved "$C^{6.47}W^{6.48}xP^{6.50}$" motif in TM6 (Figure 3B). These motifs adopt a similar conformation to active β2AR (Figures 3C and 3D). While protons are a non-canonical stimulus, the activation pathway linking proton recognition to active conformational change is conserved between the proton-sensing GPCRs and the broader class A GPCR family.

Each proton-sensing receptor harbors an extracellular-facing pocket that is lined by many polar and charged residues (Figure 3E). Despite the presence of such cavities, our structures do not resolve density for potential activating ligands or metabolites that co-purify with the activated receptor. Given the size of these pockets, however, it is possible that endogenous

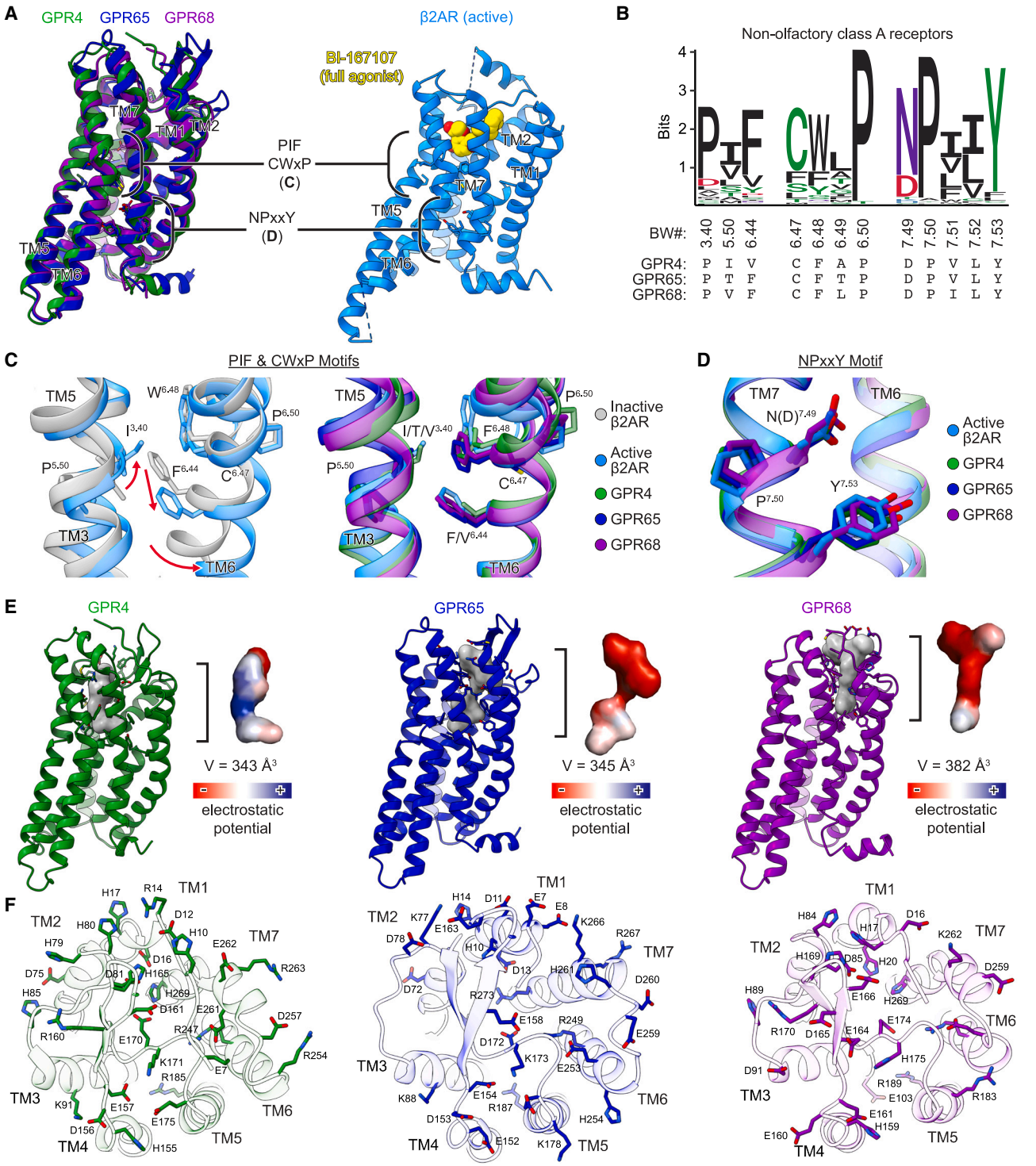


Figure 3. Structural features of human proton-sensing GPCRs

(A) Alignment of 7TM domain of GPR4-miniGα_s, GPR65-miniGα_s, and GPR68-miniG$\alpha_{s/q}$ compared with β2AR activated by BI-167107 (PDB: 3SN6).

(B) Sequence logo of PIF, CWxP, and NPxxY motifs among non-olfactory class A GPCRs.

(C) Zoom views of the PIF and CWxP motifs within the connector region in β2AR, GPR4, GPR65, and GPR68 show high similarity. Activation is associated with an outward movement of TM6 (PDB: 2RH1, 4LDO).

(legend continued on next page)

metabolites or lipids may act as agonists or allosteric modulators for each of the three proton-sensing GPCRs.

A unique feature of proton-sensing GPCRs is many proton-titratable residues in the extracellular domain of the receptor (Figures 3E and 3F). In addition to several histidine residues, all proton-sensing GPCRs harbor acidic and basic residues that engage in an extended network. Each receptor has a distinct network, although there are structurally similar positions harboring proton-titratable residues (Figure 3F). Many of these surround the canonical class A GPCR orthosteric site. Collectively, these residues may coordinate protonation network(s) extending from the extracellular receptor surface that terminate at buried titratable residues.[15,41,42] Without inactive-state structures or the ability to directly see protons at cryo-EM resolutions, it is challenging to determine which titratable residues are important for proton sensing. Nevertheless, these structures reveal the organization of putative proton-sensing residues in the active state of each receptor.

DMS of GPR68 pH response

We next aimed to identify which residues are important for receptor activation by protons. The scale of mutagenesis required for a comprehensive and unbiased profiling of proton activity is not feasible with conventional approaches. We therefore employed high-throughput DMS for unbiased sequence-to-function insight.[25]

We first devised a sensitive assay to enable DMS for GPCR signaling. Paramount to any successful DMS is a high-throughput assay capable of discerning small differences in phenotype. Prior work establishing DMS of GPCR activation used the prototypical β2AR to profile the effect of missense substitutions at every residue.[33] This pioneering GPCR-DMS was enabled by a cAMP-dependent transcriptional readout for receptor activation but with limitations of low-dynamic range and high basal activity at baseline.[33] We engineered a fluorescence-activated cell sorting (FACS)-seq approach with higher signal to noise. In this system, receptor activation of $G\alpha_s$ triggers cAMP production that acts via a transcriptional reporter to produce eGFP (Figure 4A). To improve sensitivity, we used a synthetic cAMP response element (CRE) sequence architecture recently discovered by massively parallel profiling of transcriptional response element architectures.[43] GPCR basal activity often saturates transcriptional reporters, leading to low signal to noise. We used two approaches to circumvent this issue. First, we precisely titrated receptor cell surface expression. Second, we fused the eGFP reporter to a dihydrofolate reductase (DHFR) degron that is stabilized with the small molecule trimethoprim (TMP).[44] The addition of TMP simultaneously with GPCR ligands causes eGFP accumulation only in the presence of stimulus.

We used β2AR to benchmark and optimize our assay's sensitivity and dynamic range. Forskolin treatment saturated eGFP fluorescence by directly stimulating adenylyl cyclase cAMP production (Figure 4B). The full agonist BI-167107 produced a robust eGFP signal similar to forskolin (Figure 4B). A weak partial agonist, alprenolol, modestly increased eGFP signal (Figure 4B).[46] Inverse agonist ICI-118,551 reduced eGFP signal (Figure 4B). These results provided confidence that this assay measures $G\alpha_s$-coupled receptor activation with the high dynamic range required for an effective GPCR-DMS.

We next determined whether this transcriptional reporter assay can detect pH activation of proton-sensing GPCRs. Cell lines expressing GPR4 and GPR65 revealed significant basal signal at standard pH values required for cell culture (∼pH 7.4). Attempts to increase the pH to decrease signaling were constrained by cell viability. By contrast, the eGFP signal for GPR68 is low at pH 7.4 and increased ∼30-fold at pH 5.5, a change similar to forskolin (Figures 4C and 4D). The flow cytometry transcriptional reporter assay was performed without phosphodiesterase inhibitors and thus measures cAMP production not accumulation. As in others' work, the pH_{50} of GPR68 is shifted ∼1 log unit to 5.8, compared with cAMP accumulation assays (Figures 1A, 4C, and 4D).[14,42] We surmised that the transcriptional reporter provides a platform for DMS of GPR68.

Using the DIMPLE pipeline, we designed and generated a GPR68 DNA library containing all possible single missense mutations, a single synonymous mutation, as well as 1–3 amino acid insertions and deletions (indels) at each position (Figures 4E, S4A, and S4B).[47] This library of 9,464 variants was used to generate a pool of stable HEK293T cell lines where each cell contains one GPR68 variant, enabling genotype-phenotype linkage (Figures 4E, S4A, and S4B).[47,48] With this pooled cell line library, we performed a screen at pH 5.5 (active) and pH 6.5 (inactive) (Figures 4C, 4D, S4C–S4E, and S5). To determine phenotype and genotype, the cell line library for each pH condition was sorted based on eGFP intensity into four bins using FACS (Figure 4E). The resulting subpopulations were sequenced, and a "fitness" score was calculated for each variant based on its distribution relative to synonymous mutations (see STAR Methods and Figure 4F).[45] These scores indicate whether a given variant is deleterious, beneficial, or neutral for pH-dependent GPR68 activation and are plotted as heatmaps of the full-length receptor for both the pH 5.5 (Figures 4G and S5) and pH 6.5 screens (Figure S5).

Several results from the DMS provide confidence in the approach. In the DMS at pH 6.5, mutations have little effect on fitness scores (Figure 4F), consistent with relatively little eGFP signal at the inactivating condition (Figures S4E and S5). At pH 5.5, we observe reduced fitness in regions essential for function across the broader GPCR family (Figures 4F and 4G). Specifically, substitutions in the TM regions are poorly tolerated, while the amino and carboxy termini are less constrained (Figures 4G and S5). Additionally, substitution of cysteine residues known to form disulfide bonds between ECL2 and TM3 (residues 94 and 172) and the N terminus and TM7 (residues 13 and 258) are universally deleterious (Figures 4G and S5). Substitutions to

(D) Zoom view of the NPxxY motif in active-state GPR4, GPR65, GPR68, and β2AR activated by adrenaline (PDB: 4LDO).
(E) GPR4, GPR65, and GPR68 each contain a charged pocket in the orthosteric site. Pockets were calculated using CavitOmix, and electrostatic surfaces were calculated using PyMol.
(F) Extracellular regions of GPR4, GPR65, and GPR68 contain numerous titratable residues. Titratable residues are shown as sticks.

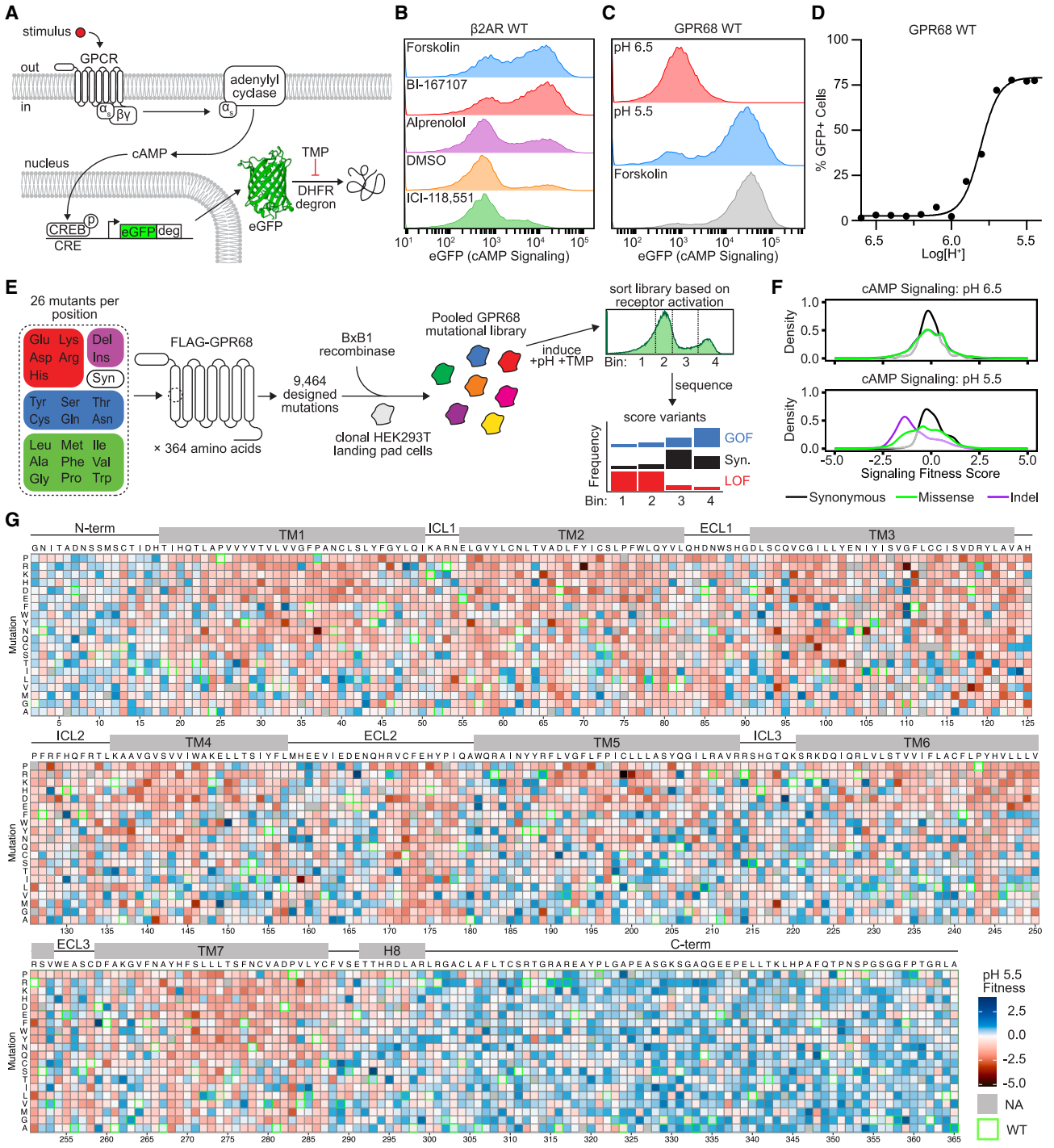


Figure 4. Deep mutational scanning for GPR68 activation

(A) Schematic of cAMP transcriptional reporter assay. GPR68 activation triggers cAMP production leading to transcription of eGFP downstream of an engineered cAMP response element.[43] A dihydrofolate reductase (DHFR) degron eliminates background signal prior to stimulation.

(B) Representative flow cytometry traces of β2AR treated with forskolin (10 μM), which directly stimulates adenylyl cyclase, BI-167107 (10 μM, full agonist), alprenolol (10 μM, antagonist), DMSO (0.1% v/v, vehicle), and ICI-118,551 (10 μM, inverse agonist).

(C) Representative flow cytometry traces of GPR68 at pH 6.5 (inactive), pH 5.5 (active), and at pH 7.5 with forskolin (25 μM).

(D) Representative pH dose-response curve for GPR68 WT. Arrows indicate pH condition shown in (C).

(E) Overview of GPR68 mutational library generation and FACS-seq pipeline for GPCR-DMS.

(legend continued on next page)

conserved GPCR motif positions and the G protein interface are also mostly deleterious. We conclude that DMS of GPR68 activation using a transcriptional reporter of G_s signaling provides a comprehensive map of mutational effects.

Multi-phenotypic DMS

Many mutational effects in our signaling DMS likely stem from changes in receptor surface expression. To deconvolve mutational effects on expression vs. pH-dependent activation, we screened the GPR68 library for surface expression using antibody labeling of an N-terminal FLAG tag (Figures 4A, S4, and S5). Others have similarly measured surface expression of variant libraries of many membrane and secreted proteins.[31,32,47,49–52]

We compared mutational effects on GPR68 activation and cell surface expression. Synonymous mutations have little effect on GPR68 signaling or surface expression (Figures 4F and 5B). Indels have significant deleterious effects (Figures 4F and 5B). Missense mutations cause more distributed effects in surface expression and signaling (Figures 4F and 5B). At pH 6.5, mutational effects are minimal because the receptor is inactive; rare missense mutations activate GPR68 (Figures 4F and S5).

We calculated an expression-adjusted functional score for each variant to identify GPR68 mutations specifically important for pH-dependent activation. We first compared synonymous mutations in the pH 5.5 cAMP signaling and surface expression screens. Synonymous mutations should have minimal effects on expression or function. Thus, we used the synonymous variants to define a regression line describing how changing expression influences signaling (Figure 5C). Missense mutations above this line are gain of function (“GOF”), due to having higher signaling activity at the same expression level. Conversely, mutations below the line are loss of function (“LOF”). We quantified change as the Euclidean distance of mutations to the regression fit (Figure 5D).

To identify residues important for pH-dependent activation, we separated negative and positive scores and averaged mutational scores per position (Figures 5E and 5F). Grouping mutation scores by physicochemistry (e.g., size, hydrophobicity, charge) in each screen reveals similar trends. Enriched in the 5% highest GOF and 5% lowest LOF scores are hallmark class A GPCR motifs such as the DRY, N(D)PxxY, CW(F)xP, and residues that contact the Gα protein (Figures 5E and 5F).[53] Intriguingly, numerous high-scoring mutations map to GPR68’s extracellular proton-titratable residues.

Molecular dynamics simulations of GPR68

Mutations in GPR68 may influence residue-level contacts in active, inactive, or intermediate conformations. We were unable to determine a structure of inactive GPR68 despite trying numerous strategies.[54–56] We instead used computational approaches to model inactive GPR68. While AlphaFold2 predictions of GPR68 provide a potential snapshot of inactive GPR68, such methods have limited confidence in the alternative states identified. By contrast, molecular dynamics simulations can estimate relative populations of states across conditions.[57,58] We used simulations to identify GPR68 conformations enriched in low proton concentrations. We used a coevolution-driven conformational exploration method[59] that reliably matches experimental alternative conformations of receptors and transporters.[60,61] We started with coevolutionary scores for all residue pairs in class A GPCRs (Figure 6A). This model provides constraints for conserved contacts important for maintaining active, inactive, and intermediate conformations across class A GPCRs. We next performed simulations of GPR68 initiating from the active-state cryo-EM model. To explore alternative states, we generated transition trajectories via enhanced sampling along unsatisfied coevolutionary constraints, employed accelerated weight histogram sampling to discourage simulations from revisiting already explored conformations, and estimated the free energy landscape for transitions. The transition from GPR68 active to intermediate and inactive-like states across these simulations is described by a reduced-dimension collective variable (CV) that captures the essential features of the conformational transitions.

The pH sensitivity of GPR68 presented a novel hurdle. Typically, molecular dynamics simulations consider a fixed protonation state for each amino acid and do not allow changes in response to the environment. Our central question is: how does protonation at distinct sites influence conformation? Therefore we needed an approach to correlate protonation and conformational states. We combined our enhanced sampling with constant-pH simulations to construct a technique that enables individual protonation states to change with conformation (see STAR Methods).

Using this workflow, we performed 10 ns of simulations during the exploration phase. Additionally, we conducted 2 sets of 4-replica single-walker accelerated weight histogram simulations, one set at pH 6.0 and the other at pH 7.0, with each replica running for 500–600 ns. We reached convergence under both pH conditions (Figure S6). The converged free energy landscapes show GPR68 explores multiple conformations at pH 7.0 (Figure 6B). Simulations at pH 6.0 reveal a narrowing of the conformational landscape, with the active conformation observed by cryo-EM at low pH becoming more populated (Figure 6B). The appropriate response to pH in simulations provides confidence in the simulated conformations of GPR68. In addition to thermodynamic validation, the simulations capture canonical GPCR allosteric switches such as the DRY, tyrosine-tyrosine lock, and CW(F)xP motifs in conformations aligning with experimental structures.[62] This corroborates the conformational dynamics of GPR68 at different pH conditions (Figure S6).

(F) Distributions of variant effects on GPR68 signaling at pH 6.5 and pH 5.5. Fitness scores are relative to the mean synonymous mutant score (set to 0) and were calculated using Enrich2.[45] See Figures S4 and S5.

(G) Heatmap of cAMP signaling fitness scores for GPR68 mutational library at pH 5.5. GPR68 WT sequence is shown above each section of heatmap, and mutations are indicated on the left axis of each section, and the amino acid position is indicated by the numbers below each section. Positions and mutations with no data are shown as gray. The WT amino acid at each position is boxed in green. Transmembrane helix cutoffs were determined using our GPR68 structure. Blue indicates increased cAMP signaling relative to synonymous mutants, and red indicates decreased cAMP signaling relative to synonymous mutants. Data are fitness values from three biologically independent deep mutational scans. See Figures S4 and S5.

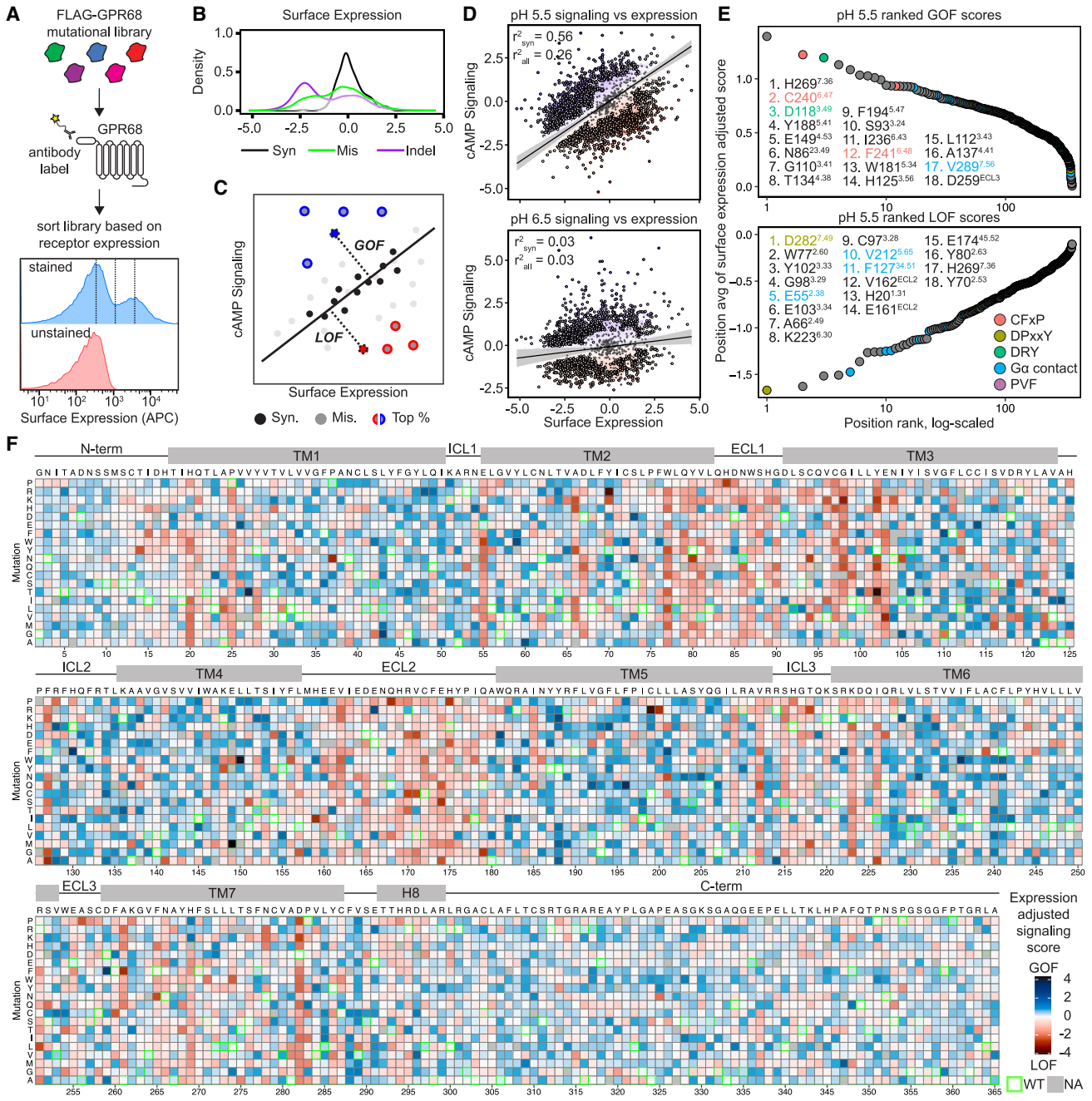


Figure 5. Integrating phenotypes reveals residues critical for GPR68 activation

(A) Overview of GPR68 surface expression DMS assay. Representative flow cytometry traces of stained (blue) and unstained (red) GPR68 mutational library.

(B) Distribution of variant effects of GPR68 surface expression. Fitness scores are relative to synonymous mutations. See Figures S4 and S5.

(C) Each mutation's cAMP signaling score at each pH condition screened was plotted against its' surface expression score. The Euclidean distance of each mutant was calculated to a line fit to the population of synonymous mutations. Black points are synonymous variants, and each missense mutant is colored by its Euclidean distance on a red-white-blue scale (same as F). Gain-of-function ("GOF") mutants are those with higher signaling scores than synonymous variants at a given level of expression. Loss-of-function ("LOF") mutants have lower signaling score than synonymous variants at a given level of expression.

(D) Scatter plot of surface expression vs. cAMP signaling scores at pH 5.5 and pH 6.5. R^2 values are shown for the synonymous (R^2_{syn}) and full missense (R^2_{all}) mutational library.

(E) Surface expression-adjusted GOF and LOF pH 5.5 cAMP signaling scores are plotted in rank order. Positions are colored by sequence motif. Superscript corresponds to each residue's Ballesteros-Weinstein number.

(legend continued on next page)

Mechanism of GPR68 activation by protons

We integrated DMS with structures and molecular dynamics simulations to build a structure-function map of GPR68. We reasoned that LOF mutations are likely to disrupt key interactions stabilized in the active conformation and therefore visualized LOF scores on the active-state GPR68 cryo-EM structure (Figure 6C). Conversely, GOF mutations likely destabilize the inactive conformation. We therefore visualized GOF scores using the inactive structure from the pH 7.0 MD simulation (CV ~1.2) (Figure 6D).

Many GOF and LOF scoring mutations map to established class A GPCR motifs. For example, mutations in D118$^{3.49}$ of the DRY motif, which stabilizes the inactive state of GPCRs, lead to increased GPR68 signaling (Figure 6D).[53,62] Mutations to C240$^{6.47}$ and F241$^{6.48}$ in the CW(F)xP motif increased GPR68 signaling, consistent with their role as allosteric switches (Figure 6D).[53,62] All mutations in D282$^{7.49}$ in the N(D)PxxY motif decreased signaling, supporting a key role of TM7 in receptor activation and G protein activation (Figures 6C and S7D). Additionally, mutation of F127$^{34.51}$ in ICL2, which interacts directly with Gα, leads to a LOF (Figures 6C and S7E).

Many LOF and GOF sites are located adjacent to the extracellular cavity in GPR68 (Figure 6C). This pocket is near orthosteric sites in class A GPCRs.[63,64] Several LOF residues, including H20$^{1.31}$, H269$^{7.36}$, Y102$^{3.33}$, and E174$^{45.52}$ line this electronegative cavity, suggesting this region is important for GPR68 activation (Figure 6C). We first looked at the extracellular histidine residues of GPR68, previously proposed as critical determinants of proton-induced activation.[14,41] Our mutational scan provides an unbiased view on the relative importance of each histidine residue in GPR68 activity. Furthermore, testing every amino acid substitution provides direct insight into how the charge state, hydrogen bonds, and van der Waals of each residue influence GPR68 activity. Two histidine residues in the extracellular region, H20$^{1.31}$ and H269$^{7.36}$, emerged as positions with high LOF scores in this global analysis (Figures 6E and 6F). Other extracellular histidines had fewer LOF effects (Figures 6E and 6F). Surprisingly, substitutions in H20$^{1.31}$ and H269$^{7.36}$ revealed mutations cause both GOF and LOF with positively charged substitutions leading to increased activity and negatively charged substitutions resulting in decreased activity (Figure 6F). For H20$^{1.31}$, hydrophobic substitutions are additionally LOF (Figure 6F). In both positions, we used a cAMP GloSensor assay to validate the DMS and further understand how substitutions influence proton activation (Figures 7B and 7D). In H20$^{1.31}$ and H269$^{7.36}$, the amino acid sidechain pK_a correlates with proton potency, in agreement with the DMS scores. Lysine substitution increases proton potency while aspartate decreases proton potency. This suggests these residues' charge and local electrostatic environment are key to receptor activity. The Hill slope of proton activation decreases from 4.2 in wild-type receptor to 2.6–2.9 in mutant receptors, suggesting perturbed cooperative proton binding. Simulations at pH 7 reveal the protonation state of H20$^{1.31}$ and H269$^{7.36}$ is highly correlated with the conformational state (Figures 7C and 7E), and reach the pH 6 protonation level once in the fully active state (Figures 7C and 7E). Our mutagenesis screen and simulations therefore highlight that H20$^{1.31}$ and H269$^{7.36}$ play central roles in pH activation. More broadly, we observe large variability in the effect of specific amino acid substitutions on each extracellular histidine, suggesting protonation of other histidine residues is likely less important for GPR68 activation (Figures 6E and 6F).

Using the DMS and simulations, we identified a network of residues connecting the extracellular cavity to the receptor core (Figures 7A and S7A–S7C). These interactions change during the MD simulations to accommodate protonation associated with activating pH conditions. Under pH 6 conditions, important titratable residues remained fully protonated across states, suggesting the energetic effect of pH on GPR68 stems from the inability to accommodate inactive-state-specific protonation patterns, rather than active-state stabilization (Figures 7C, 7E, and 7K). An extensive set of ionic and hydrogen-bonding interactions in active GPR68 engage LOF residues H269$^{7.36}$ and E174$^{45.52}$ (Figures 7F and 7G). These interactions connect extracellular-facing residues to two key residues in the core of GPR68: E103$^{3.34}$ and E149$^{4.53}$, which have strong LOF and GOF scores, respectively (Figures 7F, 7G, 7L, and 7M). According to our MD simulations, activation of GPR68 is associated with a movement of E174$^{45.52}$, which engages Y102$^{3.33}$ and R251$^{6.58}$ in the active state (Figures 7A and S7G). Mutation of E174$^{45.52}$, Y102$^{3.33}$, or R251$^{6.58}$ leads to a significant decrease in proton potency (Figures 7H, 7I, and S7F), supporting this interaction's importance in GPR68 activation. The E174$^{45.52}$:Y102$^{3.33}$ interaction rotates and upwardly displaces TM3, which is relayed to E103$^{3.34}$ and E149$^{4.53}$ (Figure 7A). In both our MD-derived and cryo-EM active states, E149$^{4.53}$ points toward the centrally located N104$^{3.35}$, leading to a twist in TM3 resulting in CW(F)xP motif rearrangement and an outward movement of TM6 (Figures 7A, 7L, and 7M). E149$^{4.53}$ is fully protonated in our pH 6 MD simulations (Figure 7K). Indeed, the E149Q$^{4.53}$ mutation, which mimics the protonated state, is more easily activated by protons (Figure 7J). Thus, our results confirm a prior study that identified E149$^{4.53}$ as a critical activity-associated residue in GPR68,[15] but provide key structural context.

The combination of DMS, simulations, and structural analysis reveals protonation along key residues in the extracellular cavity trigger conformational rearrangements in GPR68 with TM3 as a central conduit. This relay, initiating with H269$^{7.36}$ and H20$^{1.31}$ and terminating at E149$^{4.53}$, converges to the conserved connector region in class A GPCRs that coordinates TM helices rearrangement to allow G protein binding and activation (Figure 7A).

Tuning pH sensitivity in the proton sensor family

We next examined whether GPR4 and GPR65 sense protons through a similar network to GPR68. If they do, we expect the

(F) Heatmap of GPR68 mutational library surface expression-adjusted pH 5.5 cAMP signaling scores. WT sequence is shown above each section of heatmap, and mutations are indicated on the left axis of each section, and the amino acid position is indicated by the numbers below each section. Positions and mutations with no data are shown as gray. The WT amino acid at each position is boxed in green. Transmembrane helix cutoffs were determined using our GPR68 structure. Blue indicates higher activity relative to synonymous mutants, and red indicates lower activity relative to synonymous mutants. See also Figures S4 and S5.

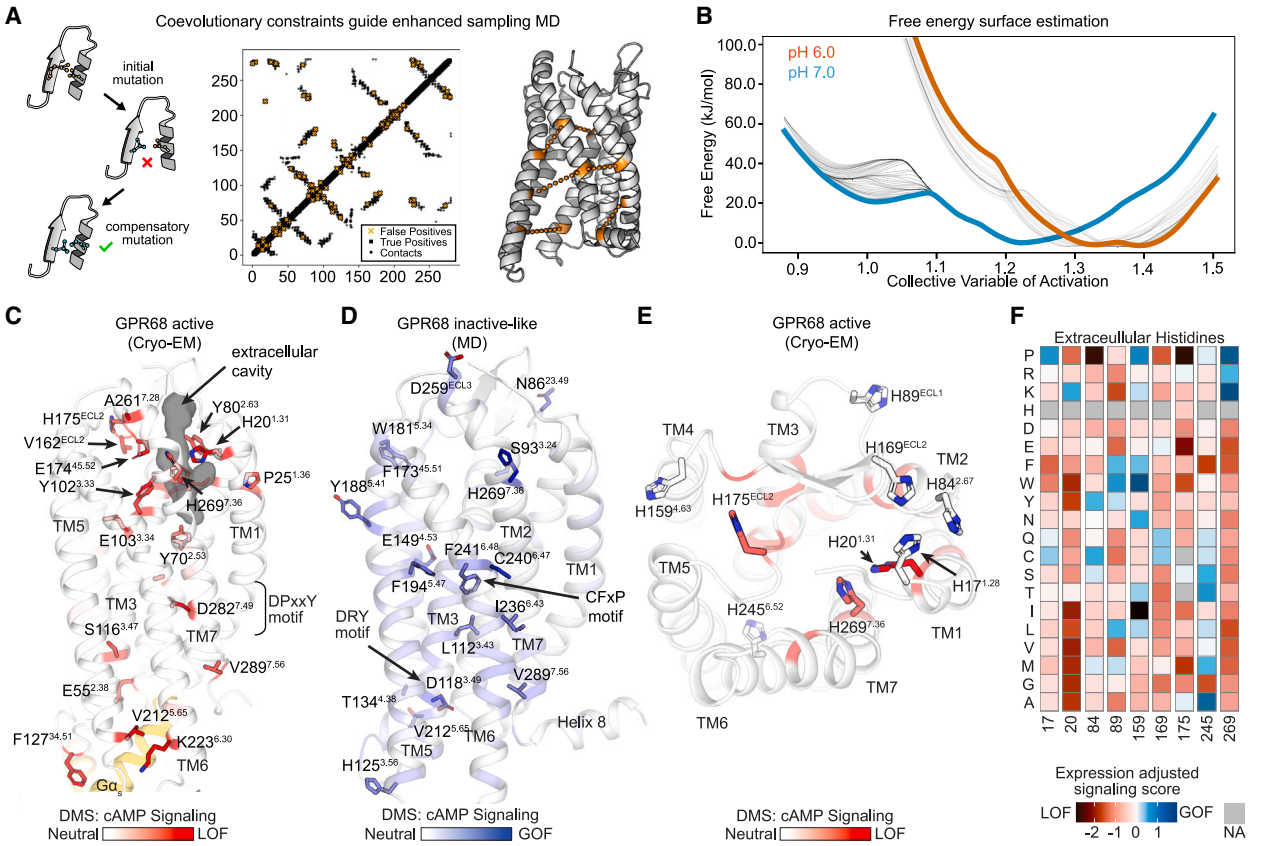


Figure 6. Mapping mutational effects across the conformational landscape of GPR68

(A) A multiple sequence alignment of all class A GPCRs was used to identify pairs of coevolving residues, displayed as a matrix. Strongly coevolving residue pairs tend to be interacting at some point during the life of the protein. Upon comparison with a 3D structure, most coevolving residue pairs are in contact (true positives, black), while a small subset of highly coevolving residue pairs are far away in 3D space (false positives, orange). Such pairs can be used to identify putative alternative conformations not seen in single-state experimental models. These coevolutionary constraints (orange dotted lines) were used to guide enhanced sampling molecular dynamics simulations using our experimental GPR68 cryo-EM structure as a starting point to generate a conformational ensemble of GPR68. See Figure S6.

(B) The free energy landscape of GPR68 at pH 7.0 and pH 6.0. At pH 7.0 (blue), GPR68 explores a comparatively wide array of conformations along the collective variable of activation. The free energy estimations of the replicates show the uncertainty and possible range of the energetic landscape (black). Additionally, the weighted, normal, and Boltzmann averages are shown in gray. The energetic minimum occurs at approximately a CV of 1.2. Upon acidification at pH 6.0 (orange), the conformational landscape narrows and shifts toward more active CV values, with the energetic minimum at approximately a CV of 1.4. See Figure S6.

(C) Residues where mutations result in decreased cAMP signaling activity are shown as sticks on cryo-EM active-state structure of GPR68. The extracellular cavity of GPR68 is shown as a gray surface.

(D) Residues where mutations result in increased cAMP signaling activity are shown as sticks on the MD-derived representative inactive basin snapshot of GPR68. Common class A GPCR activation motifs are indicated.

(E) Top, extracellular, view of our active structure of GPR68, where all extracellular histidine residues are shown as stick and colored by their LOF score.

(F) Subset heatmap from Figure 5F for each of the histidine residues shown in (E).

positions critical for GPR68 activation to be important for GPR4 and GPR65 activation. Across the family, there are structurally conserved proton-titratable residues (Figure 3F).

We first investigated the conserved acidic residue within ECL2 (E170 in GPR4, D172 in GPR65, and E174 in GPR68) (Figures S7H, S7L, and S7P). Like GPR68, alanine and neutralizing mutations decrease cooperativity across receptors (Figures S7I and S7M; Tables S4 and S5). The effect on proton potency is diminished in GPR65 and negligible in GPR4. These effects highlight that stabilization of ECL2 is important for receptor activation across the family.

GPR4 and GPR65 have charged extracellular-facing cavities similar to GPR68. Having uncovered the charge dependence of GPR68 H269$^{7.36}$, which is positioned at the top of this cavity, we tested the homologous set of mutations for GPR4 (H269$^{7.36}$) and GPR65 (R273$^{7.36}$) (Figures S7H, S7L, and S7P). Indeed, at this site, we see that negatively charged mutations universally decrease proton potency (Figures S7J and S7N; Tables S4 and S5). Positively charged mutations in GPR4 and GPR65 have less pronounced effects, perhaps because this position is already protonated at more basic pH conditions.

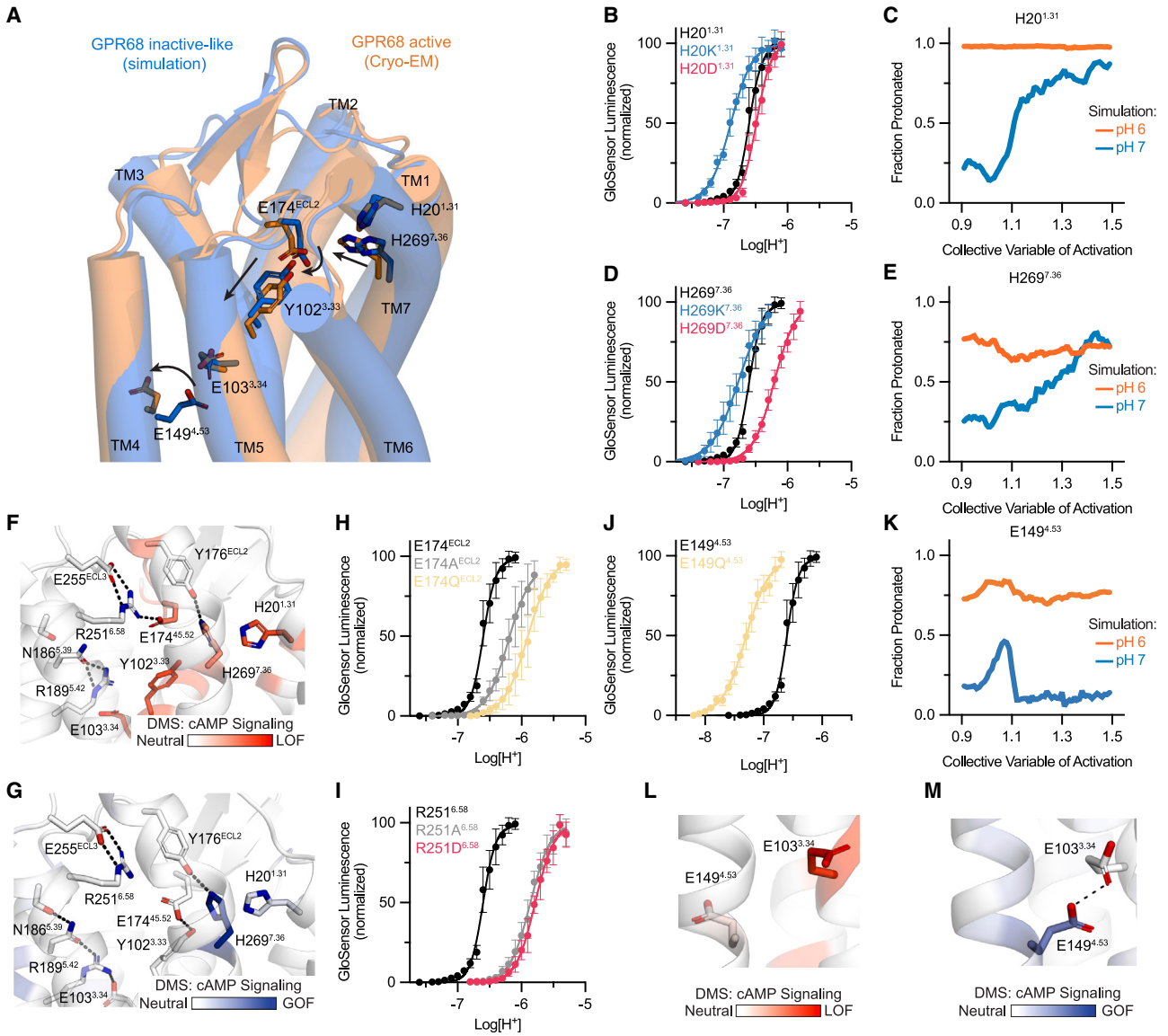


Figure 7. An integrated model of GPR68 activation

(A) Overlay of GPR68 active (cryo-EM, orange) and inactive-state (simulation, blue) models. Sticks and arrows are shown for residues involved in conformational change upon activation by protons.

(B) cAMP GloSensor signaling assay testing the effect of acidic and basic mutations at $H20^{1.31}$.

(C) $H20^{1.31}$ fraction protonated vs. collective variable of activation from constant-pH molecular dynamics simulations demonstrating H20 titrates as a function of receptor conformation.

(D) cAMP GloSensor signaling assay testing the effect of acidic and basic mutations at $H269^{7.36}$.

(E) $H269^{7.36}$ fraction protonated vs. collective variable of activation from constant-pH molecular dynamics simulations demonstrating H269 titrates as a function of receptor conformation.

(F) Zoom view of (A) showing key extracellular residues and hydrogen bonds in proton-activated GPR68 cryo-EM structure. Residues are colored by their respective LOF score from the pH 5.5 cAMP signaling mutational scanning experiment.

(G) Zoom view of (A) showing key extracellular residues and hydrogen bonds in inactive-like GPR68 model from simulation. Residues are colored by their respective GOF score from the pH 5.5 cAMP signaling mutational scanning experiment.

(H) cAMP GloSensor signaling assay testing the effect of neutralizing and alanine mutations at $E174^{ECL2}$.

(I) cAMP GloSensor signaling assay testing the effect of acidic and alanine mutations at $R251^{6.58}$.

(J) cAMP GloSensor signaling assay testing the effect neutralizing mutation at $H269^{7.36}$.

(K) $E149^{4.53}$ fraction protonated vs. collective variable of activation from constant-pH molecular dynamics simulations.

(L) Zoom view of (A) showing $E149^{4.53}$ conformation in proton-activated GPR68 cryo-EM structure. Residues are colored by their respective LOF score from the pH 5.5 cAMP signaling mutational scanning experiment.

(legend continued on next page)

Finally, we examined the conserved glutamate, $E^{4.53}$, which is in the middle of TM4 and buried far from the extracellular solvent and intracellular G protein binding pocket in each receptor. This position potentially serves as a link between the proton-sensing network and canonical activation motifs.[41] We hypothesized what we found in GPR68 may transfer to GPR4 and GPR65. In agreement with previous work, we observe an increase in proton potency for each receptor upon mutation to glutamine: GPR4 $E145Q^{4.53}$ increases pH_{50} by ~0.25. GPR65 $E142Q^{4.53}$ increases pH_{50} by ~0.1, and GPR68 $E149Q^{4.53}$ increases pH_{50} by nearly a full pH unit (Figures 7J, S7K, S7O, and S7P; Tables S4 and S5).[15] Interestingly, we see basal activity increases with each $E^{4.53}Q$ mutation, suggesting this position governs baseline receptor signaling (Tables S3 and S5).

With these studies, we conclude all receptors in the family share a common buried acidic residue at which protonation likely drives activation. Furthermore, each receptor has a similar sensing mechanism on the extracellular side of the receptor, but the exact identity of the residues comprising these networks and resulting nuances in the local electrostatic environment differ between them.

DISCUSSION

Our structural, simulation, and DMS studies suggest a general model for how protons activate the proton-sensing GPCRs. Using GPR68 as a prototype of the subfamily, we find a network of residues that connects an extracellular-facing cavity to a conserved charged residue buried in the TM core of the receptor. Protonation likely drives conformational changes in ECL2, which stabilizes movement of TM3 and a series of rearrangements that connect the extracellular-facing cavity to $E149^{4.53}$, a residue uniquely conserved in the proton-sensing GPCRs. Several observations support this distributed network of proton sensing. First, for each proton-sensing receptor, cAMP assays reveal a pH Hill slope >4, suggesting significant cooperativity between distinct proton binding sites in proton-dependent activation. Second, our chimeric receptor constructs between GPR4 and GPR68 suggest that multiple distributed regions within the extracellular portions of the receptors define the pH setpoint. Third, our MD simulations show clear state-dependent protonation patterns associated with stabilizing active conformations under pH 6.0 conditions. Finally, our DMS experiments do not identify a single cluster for GPR68, but instead many distinct residues in the extracellular and TM regions with functional consequences. Although there are nuances to the proton-sensing domain of each proton-sensing receptor, these networks converge upon hallmark GPCR motifs that link ligand binding to a conformational change allowing G protein binding. This provides an activation pathway from extracellular proton binding to G protein activation, and it points toward a conserved model across the family where GPCR proton sensing is not localized to a single site.

The distributed model for proton-sensing-based GPCR activation contrasts with mechanisms proposed for other membrane protein proton sensors and transporters. For many of these membrane proteins, proton-driven activation has been ascribed to single or small subsets of amino acids.[1–3,35–37,65] By contrast, we demonstrate that many distributed protonatable residues contribute to proton-dependent activation in GPR68. A similar distributed network is likely important for GPR4 and GPR65.

Our approach to analyze the functional consequence of each amino acid in GPR68 provides key advances in DMS to understand GPCR function. The cAMP-driven transcriptional reporter assay is transferable to a large number of GPCRs that modulate cAMP levels. An additional advance is an engineered system that only integrates cAMP-driven transcriptional output in the presence of the receptor stimulus; this overcomes fundamental challenges with basal signaling suppressing the signal to noise of transcriptional readouts of GPCR activation. Perhaps the most important advance we introduce here is accounting for receptor surface expression while evaluating the effect of any mutation on cAMP production. In the absence of such normalization, many LOF or GOF mutations simply reflect changes in receptor biogenesis or trafficking to the cell surface. By developing a way to integrate mutational scanning for multiple phenotypes, we unambiguously identified residues critical for GPR68 activation by protons.

In addition to these advances, the use of constant-pH MD simulations introduced another dimension to our analysis by simultaneously sampling the full conformational landscape while capturing the pH-dependent free energy landscape. This approach allows us to not only map out the dynamic structural transitions of GPR68 but also to retain atomic-level detail, including critical interaction networks that govern receptor behavior. By incorporating both conformational flexibility and precise protonation states, we gain a deeper understanding of how pH modulates receptor activity, providing insights into the mechanistic underpinnings of GPCR function.

Integration of structural biology, simulations, and DMS provides a holistic approach to understand GPCR function. Future GPCR-DMS could assess how mutations influence signaling through different G protein and β-arrestin pathways, receptor internalization, location-dependent signaling, and receptor biogenesis. We envision that DMS interrogation of GPCR structures will reveal determinants of orthosteric and allosteric ligand binding, novel allosteric sites, and regions of receptors important in engaging signal transducers and regulatory complexes. Such integration of structure and function could reveal how genetic variation in GPCRs influences human disease and variation in drug efficacy.

Limitations of the study

Several caveats are notable. First, the presented cryo-EM structures represent stable, active conformations of each receptor.

(M) Zoom view of (A) showing $E149^{4.53}$ conformation in inactive-like GPR68 model from simulations. Residues are colored by their respective GOF score from the pH 5.5 cAMP signaling mutational scanning experiment.
Data in (B), (D), and (H)–(J) are from 3 to 4 biological replicates ± SD (see Table S3).
See also Figure S7.

While we used advances in molecular simulation to model the conformational ensemble of GPR68, experimental inactive or partially inactive structures of each receptor would illuminate residues and interactions that are uniquely essential for activation of each receptor. An additional caveat is that we do not assess how pH influences the G protein components in our cryo-EM experiments. Second, compared with GPR68, we do not observe cholesterol bound to either GPR4 or GPR65. Although cholesterol may play a role in GPR68 function, our DMS data did not reveal a clear mutational effect for residues lining the cholesterol pocket. The role of lipid interactions was not probed during our MD simulations, which may impact the precise energetic balance of inactive and active states. Future studies may reveal the functional importance of this lipid interaction. Third, we cannot fully explain why each receptor has a different pH_{50}. We uncover key residues and interactions important for GPR68 activation by protons, some of which also modulate GPR4 and GPR65 activity, but further studies will be needed to define the exact mechanisms dictating pH_{50}.

These limitations do not interfere with the central advances of this study, which provide structural and dynamic resolution into proton sensing by GPCRs, an innovative GPCR mutational scanning approach, and a way to integrate structural and functional data at scale.

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to lead contact, Aashish Manglik (aashish.manglik@ucsf.edu).

Materials availability

Stable reagents generated within this study will be shared upon request to the lead contact. The pLP_mod_mNeonGreen11_P2A_puroR_CRE-GFP plasmid used in the DMS study has been deposited to Addgene: https://www.addgene.org/229227/.

Data and code availability

- Coordinates for the GPR4-G_s, GPR65-G_s, GPR68-G_s, and GPR68-$G_{s/q}$ complexes have been deposited in the RCSB Protein Data Bank (PDB: 9BIP, 9BHL, 9BHM, and 9BI6, respectively). EM density maps for GPR4-G_s, GPR65-G_s, GPR68-G_s, and GPR68-$G_{s/q}$ complexes have been deposited in the Electron Microscopy Data Bank under accession codes EMDB-44597, EMDB-44549, EMDB-44550, and EMDB-44560, respectively. EM density maps for transmembrane domain-focused refinements of GPR4, GPR65, and GPR68 have been deposited in the Electron Microscopy Data Bank under accession codes EMDB-44596, EMDB-44548, and EMDB-44561, respectively. Sequencing data from the GPR68 deep mutational scan have been deposited in the NCBI Sequence Read Archive (SRA) under Bioproject PRJNA1062987. GPR68 deep mutational scan scores for each screen are deposited on maveDB experiment #00001207.
- The code and pipelines used to process deep mutational scan sequencing data along with raw variant counts and fitness scores have been deposited on GitHub: https://github.com/odcambc/GPR68_processing, https://github.com/odcambc/GPR68_DMS_QC and Zenodo: https://doi.org/10.5281/zenodo.10999528, https://doi.org/10.5281/zenodo.10999523. Scripts used to analyze data and prepare figures have been deposited on GitHub: https://github.com/matthewkarlhoward/GPR68_DMS_Analysis and Zenodo: https://zenodo.org/doi/10.5281/zenodo.11056584. The pipelines and scripts used for generating, analyzing, and post-processing the MD data have been deposited to Zenodo: https://zenodo.org/doi/10.5281/zenodo.13785406.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank all members of the Coyote-Maestas and Manglik labs for their helpful feedback and discussion as we conducted this project. Funding: National Institutes of Health Ruth L. Kirschstein Predoctoral Fellowship F31HL164045 (N.H.), National Institutes of Health Ruth L. Kirschstein Postdoctoral Fellowship 1F32GM152977 (C.B.M.), National Institutes of Health Ruth L. Kirschstein Predoctoral Fellowship 1F31AI157438 (D.D.T.), National Institutes of Health National Institute of General Medical Sciences 5T32GM139786 (M.K.H.), National Institutes of Health National Institute of General Medical Sciences T32GM141323 (E.M.), National Institutes of Health National Institute of Mental Health 1R21MH120422-01 (X.-P.H.), cryo-EM equipment at UCSF is partially supported by National Institutes of Health grants S10OD020054 and S10OD021741, Stanford-SLAC Cryo-EM Center (S2C2), which is supported by the National Institutes of Health Common Fund Transformative High-Resolution Cryo-Electron Microscopy program (U24 GM129541), National Cancer Institute Cryo-Electron Microscopy Facility, and Pacific Northwest Center for Cryo-EM, Edward Mallinckrodt, Jr. Foundation (A.M.), Vallee Foundation (A.M.), Chan Zuckerberg San Francisco Biohub (A.M. and W.C.-M.), Howard Hughes Medical Institute Hanna Gray Fellowship (W.C.-M.), UCSF Quantitative Biosciences Institute Fellow Funding (W.C.-M.), Knut and Alice Wallenberg Foundation (L.D.), Göran Gustafsson Foundation (L.D.), Swedish Research Council (VR 2019 - 02433) (L.D.), and PDC Center for High-Performance Computing (L.D.).

AUTHOR CONTRIBUTIONS

M.K.H., P.R.G., D.D.T., and W.C.M. generated and cloned the GPR68 deep mutational library. M.K.H. designed and performed deep mutational scan experiments with input and assistance from N.H., J.G.E., A.M., and W.C.-M. C.B.M. processed raw next-generation sequencing data. M.K.H. analyzed DMS datasets with input from N.H., A.M., and W.C.-M. N.H. cloned, expressed, and biochemically optimized the purification of proton sensor constructs for structural studies. N.H. performed cryo-EM data collection and data processing with help from cryo-EM facilities. N.H., C.B.B., E.M., and A.M. built and refined models of the proton sensors. M.K.H., N.H., and X.-P.H. generated receptor constructs, performed signaling studies, and analyzed the data. D.M. performed and analyzed molecular dynamics simulations with input from L.D. M.K.H., N.H., and D.M. prepared figures with input from W.C.-M. and A.M. M.K.H., N.H., D.M., W.C.-M., and A.M. wrote the manuscript, with edits and approval from all authors. W.C.-M. and A.M. supervised the overall project.

DECLARATION OF INTERESTS

A.M. is a founder of Epiodyne and Stipple Bio, consults for Abalone, and serves on the scientific advisory board of Septerna. D.M. is the founder and chief scientific officer of Dioclea Labs. L.D. is the chief scientific advisor of the Dioclea Labs advisory board.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODE AND STUDY PARTICIPANT DETAILS
  - Cell lines
- METHOD DETAILS
  - GloSensor cAMP assays
  - GPR4-GPR68 chimeric receptor design
  - GPR68 Deep Mutational Scan

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.11.036.

Received: May 19, 2024
Revised: September 23, 2024
Accepted: November 21, 2024
Published: January 2, 2025

REFERENCES

1. Jasti, J., Furukawa, H., Gonzales, E.B., and Gouaux, E. (2007). Structure of acid-sensing ion channel 1 at 1.9 A resolution and low pH. Nature *449*, 316–323. https://doi.org/10.1038/nature06163.
2. Jordt, S.E., Tominaga, M., and Julius, D. (2000). Acid potentiation of the capsaicin receptor determined by a key extracellular site. Proc. Natl. Acad. Sci. USA *97*, 8134–8139. https://doi.org/10.1073/pnas.100129497.
3. Yang, J., Chen, J., Del Carmen Vitery, M., Osei-Owusu, J., Chu, J., Yu, H., Sun, S., and Qiu, Z. (2019). PAC, an evolutionarily conserved membrane protein, is a proton-activated chloride channel. Science *364*, 395–399. https://doi.org/10.1126/science.aav9739.
4. Li, B., Rietmeijer, R.A., and Brohawn, S.G. (2020). Structural basis for pH gating of the two-pore domain K+ channel TASK2. Nature *586*, 457–462. https://doi.org/10.1038/s41586-020-2770-2.
5. Rahman, M.F., Askwith, C., and Govindarajan, R. (2017). Molecular determinants of acidic pH-dependent transport of human equilibrative nucleoside transporter 3. J. Biol. Chem. *292*, 14775–14785. https://doi.org/10.1074/jbc.M117.787952.
6. Choi, J.W., Lee, S.Y., and Choi, Y. (1996). Identification of a putative G protein-coupled receptor induced during activation-induced apoptosis of T cells. Cell. Immunol. *168*, 78–84. https://doi.org/10.1006/cimm.1996.0051.
7. Ludwig, M.-G., Vanek, M., Guerini, D., Gasser, J.A., Jones, C.E., Junker, U., Hofstetter, H., Wolf, R.M., and Seuwen, K. (2003). Proton-sensing G-protein-coupled receptors. Nature *425*, 93–98. https://doi.org/10.1038/nature01905.
8. Kumar, N.N., Velic, A., Soliz, J., Shi, Y., Li, K., Wang, S., Weaver, J.L., Sen, J., Abbott, S.B.G., Lazarenko, R.M., et al. (2015). Physiology. Regulation of breathing by CO_2 requires the proton-activated receptor GPR4 in retrotrapezoid nucleus neurons. Science *348*, 1255–1260. https://doi.org/10.1126/science.aaa0922.
9. Robert, R., and Mackay, C.R. (2018). Gαs-coupled GPCRs GPR65 and GPR174. Downers for immune responses. Immunol. Cell Biol. *96*, 341–343. https://doi.org/10.1111/imcb.12027.
10. Lassen, K.G., McKenzie, C.I., Mari, M., Murano, T., Begun, J., Baxt, L.A., Goel, G., Villablanca, E.J., Kuo, S.-Y., Huang, H., et al. (2016). Genetic coding variant in GPR65 alters lysosomal pH and links lysosomal dysfunction with colitis risk. Immunity *44*, 1392–1405. https://doi.org/10.1016/j.immuni.2016.05.007.
11. Wirasinha, R.C., Vijayan, D., Smith, N.J., Parnell, G.P., Swarbrick, A., Brink, R., King, C., Stewart, G., Booth, D.R., and Batten, M. (2018). GPR65 inhibits experimental autoimmune encephalomyelitis through CD4+ T cell independent mechanisms that include effects on iNKT cells. Immunol. Cell Biol. *96*, 128–136. https://doi.org/10.1111/imcb.1031.
12. Xu, J., Mathur, J., Vessières, E., Hammack, S., Nonomura, K., Favre, J., Grimaud, L., Petrus, M., Francisco, A., Li, J., et al. (2018). GPR68 senses flow and is essential for vascular physiology. Cell *173*, 762–775.e16. https://doi.org/10.1016/j.cell.2018.03.076.
13. Reeh, P.W., and Steen, K.H. (1996). Chapter 8. Tissue acidosis in nociception and pain. In Progress in Brain Research, T. Kumazawa, L. Kruger, and K. Mizumura, eds. (Elsevier), pp. 143–151.
14. Huang, X.-P., Kenakin, T.P., Gu, S., Shoichet, B.K., and Roth, B.L. (2020). Differential roles of extracellular histidine residues of GPR68 for proton-sensing and allosteric modulation by divalent metal ions. Biochemistry *59*, 3594–3614. https://doi.org/10.1021/acs.biochem.0c00576.
15. Rowe, J.B., Kapolka, N.J., Taghon, G.J., Morgan, W.M., and Isom, D.G. (2021). The evolution and mechanism of GPCR proton sensing. J. Biol. Chem. *296*, 100167. https://doi.org/10.1074/jbc.RA120.016352.
16. Wang, J.-Q., Kon, J., Mogi, C., Tobo, M., Damirin, A., Sato, K., Komachi, M., Malchinkhuu, E., Murata, N., Kimura, T., et al. (2004). TDAG8 is a proton-sensing and psychosine-sensitive G-protein-coupled receptor. J. Biol. Chem. *279*, 45626–45633. https://doi.org/10.1074/jbc.M406966200.
17. Ring, A.M., Manglik, A., Kruse, A.C., Enos, M.D., Weis, W.I., Garcia, K.C., and Kobilka, B.K. (2013). Adrenaline-activated structure of β2-adrenoceptor stabilized by an engineered nanobody. Nature *502*, 575–579. https://doi.org/10.1038/nature12572.
18. Yu, J., Kumar, A., Zhang, X., Martin, C., Raia, P., Koehl, A., Laeremans, T., Steyaert, J., Manglik, A., Ballet, S., et al. (2023). Structural basis of μ-opioid receptor-targeting by a nanobody antagonist. Preprint at bioRxiv. https://doi.org/10.1101/2023.12.06.570395.
19. Billesbølle, C.B., de March, C.A., van der Velden, W.J.C., Ma, N., Tewari, J., del Torrent, C.L., Li, L., Faust, B., Vaidehi, N., Matsunami, H., et al. (2023). Structural basis of odorant recognition by a human odorant receptor. Nature *615*, 742–749. https://doi.org/10.1038/s41586-023-05798-y.
20. Koehl, A., Hu, H., Maeda, S., Zhang, Y., Qu, Q., Paggi, J.M., Latorraca, N.R., Hilger, D., Dawson, R., Matile, H., et al. (2018). Structure of the μ-opioid receptor-Gi protein complex. Nature *558*, 547–552. https://doi.org/10.1038/s41586-018-0219-7.
21. Papasergi-Scott, M.M., Pérez-Hernández, G., Batebi, H., Gao, Y., Eskici, G., Seven, A.B., Panova, O., Hilger, D., Casiraghi, M., He, F., et al. (2023). Time-resolved cryo-EM of G protein activation by a GPCR. Preprint at bioRxiv. https://doi.org/10.1101/2023.03.20.533387.
22. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., et al. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. Science *289*, 739–745. https://doi.org/10.1126/science.289.5480.739.
23. Liu, W., Chun, E., Thompson, A.A., Chubukov, P., Xu, F., Katritch, V., Han, G.W., Roth, C.B., Heitman, L.H., IJzerman, A.P., et al. (2012). Structural basis for allosteric regulation of GPCRs by sodium ions. Science *337*, 232–236. https://doi.org/10.1126/science.1219218.
24. Latorraca, N.R., Venkatakrishnan, A.J., and Dror, R.O. (2017). GPCR dynamics: structures in motion. Chem. Rev. *117*, 139–155. https://doi.org/10.1021/acs.chemrev.6b00177.
25. Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807. https://doi.org/10.1038/nmeth.3027.
26. Yee, S.W., Macdonald, C., Mitrovic, D., Zhou, X., Koleske, M.L., Yang, J., Silva, D.B., Grimes, P.R., Trinidad, D., More, S.S., et al. (2023). The full spectrum of OCT1 (SLC22A1) mutations bridges transporter biophysics

to drug pharmacogenomics. Preprint at bioRxiv. https://doi.org/10.1101/2023.06.06.543963.

27. Weng, C., Faure, A.J., Escobedo, A., and Lehner, B. (2024). The energetic and allosteric landscape for KRAS inhibition. Nature *626*, 643–652. https://doi.org/10.1038/s41586-023-06954-0.
28. Faure, A.J., Domingo, J., Schmiedel, J.M., Hidalgo-Carcedo, C., Diss, G., and Lehner, B. (2022). Mapping the energetic and allosteric landscapes of protein binding domains. Nature *604*, 175–183. https://doi.org/10.1038/s41586-022-04586-4.
29. Tsuboyama, K., Dauparas, J., Chen, J., Laine, E., Mohseni Behbahani, Y., Weinstein, J.J., Mangan, N.M., Ovchinnikov, S., and Rocklin, G.J. (2023). Mega-scale experimental analysis of protein folding stability in biology and design. Nature *620*, 434–444. https://doi.org/10.1038/s41586-023-06328-6.
30. Heredia, J.D., Park, J., Brubaker, R.J., Szymanski, S.K., Gill, K.S., and Procko, E. (2018). Mapping interaction sites on human chemokine receptors by deep mutational scanning. J. Immunol. *200*, 3825–3839. https://doi.org/10.4049/jimmunol.1800343.
31. Penn, W.D., McKee, A.G., Kuntz, C.P., Woods, H., Nash, V., Gruenhagen, T.C., Roushar, F.J., Chandak, M., Hemmerich, C., Rusch, D.B., et al. (2020). Probing biophysical sequence constraints within the transmembrane domains of rhodopsin by deep mutational scanning. Sci. Adv. *6*, eaay7505. https://doi.org/10.1126/sciadv.aay7505.
32. Park, J., Selvam, B., Sanematsu, K., Shigemura, N., Shukla, D., and Procko, E. (2019). Structural architecture of a dimeric class C GPCR based on co-trafficking of sweet taste receptor subunits. J. Biol. Chem. *294*, 4759–4774. https://doi.org/10.1074/jbc.RA118.006173.
33. Jones, E.M., Lubock, N.B., Venkatakrishnan, A.J., Wang, J., Tseng, A.M., Paggi, J.M., Latorraca, N.R., Cancilla, D., Satyadi, M., Davis, J.E., et al. (2020). Structural and functional characterization of G protein-coupled receptors with deep mutational scanning. eLife *9*, e54895. https://doi.org/10.7554/eLife.54895.
34. Howard, C.J., Abell, N.S., Osuna, B.A., Jones, E.M., Chan, L.Y., Chan, H., Artis, D.R., Asfaha, J.B., Bloom, J.S., Cooper, A.R., et al. (2024). High resolution deep mutational scanning of the melanocortin-4 receptor enables target characterization for drug discovery. Preprint at bioRxiv. https://doi.org/10.1101/2024.10.11.617882.
35. Vullo, S., Bonifacio, G., Roy, S., Johner, N., Bernèche, S., and Kellenberger, S. (2017). Conformational dynamics and role of the acidic pocket in ASIC pH-dependent gating. Proc. Natl. Acad. Sci. USA *114*, 3768–3773. https://doi.org/10.1073/pnas.1620560114.
36. Sun, D., Liu, S., Li, S., Zhang, M., Yang, F., Wen, M., Shi, P., Wang, T., Pan, M., Chang, S., et al. (2020). Structural insights into human acid-sensing ion channel 1a inhibition by snake toxin mambalgin1. eLife *9*, e57096. https://doi.org/10.7554/eLife.57096.
37. Ruan, Z., Osei-Owusu, J., Du, J., Qiu, Z., and Lü, W. (2020). Structures and pH-sensing mechanism of the proton-activated chloride channel. Nature *588*, 350–354. https://doi.org/10.1038/s41586-020-2875-7.
38. Nehmé, R., Carpenter, B., Singhal, A., Strege, A., Edwards, P.C., White, C.F., Du, H., Grisshammer, R., and Tate, C.G. (2017). Mini-G proteins: novel tools for studying GPCRs in their active conformation. PLoS One *12*, e0175642. https://doi.org/10.1371/journal.pone.0175642.
39. Mogi, C., Tomura, H., Tobo, M., Wang, J.-Q., Damirin, A., Kon, J., Komachi, M., Hashimoto, K., Sato, K., and Okajima, F. (2005). Sphingosylphosphorylcholine antagonizes proton-sensing ovarian cancer G-protein-coupled receptor 1 (OGR1)-mediated inositol phosphate production and cAMP accumulation. J. Pharmacol. Sci. *99*, 160–167. https://doi.org/10.1254/jphs.fp0050599.
40. Yu, X., Huang, X.-P., Kenakin, T.P., Slocum, S.T., Chen, X., Martini, M.L., Liu, J., and Jin, J. (2019). Design, synthesis, and characterization of Ogerin-based positive allosteric modulators for G protein-coupled receptor 68 (GPR68). J. Med. Chem. *62*, 7557–7574. https://doi.org/10.1021/acs.jmedchem.9b00869.
41. Kapur, B., Baldessari, F., Lazaratos, M., Nar, H., Schnapp, G., Giorgetti, A., and Bondar, A.-N. (2023). Protons taken hostage: dynamic H-bond networks of the pH-sensing GPR68. Comput. Struct. Biotechnol. J. *21*, 4370–4384. https://doi.org/10.1016/j.csbj.2023.08.034.
42. Matsingos, C., Howell, L.A., McCormick, P.J., and Fornili, A. (2023). Elucidating the activation mechanism of the proton-sensing GPR68 receptor. Preprint at bioRxiv. https://doi.org/10.1101/2023.12.08.570878.
43. Zahm, A.M., Owens, W.S., Himes, S.R., Fallon, B.S., Rondem, K.E., Gormick, A.N., Bloom, J.S., Kosuri, S., Chan, H., and English, J.G. (2024). A massively parallel reporter assay library to screen short synthetic promoters in mammalian cells. Nat. Commun. https://doi.org/10.1038/s41467-0240545-9.
44. Iwamoto, M., Björklund, T., Lundberg, C., Kirik, D., and Wandless, T.J. (2010). A general chemical method to regulate protein stability in the mammalian central nervous system. Chem. Biol. *17*, 981–988. https://doi.org/10.1016/j.chembiol.2010.07.009.
45. Rubin, A.F., Gelman, H., Lucas, N., Bajjalieh, S.M., Papenfuss, A.T., Speed, T.P., and Fowler, D.M. (2017). A statistical framework for analyzing deep mutational scanning data. Genome Biol. *18*, 150. https://doi.org/10.1186/s13059-017-1272-5.
46. Wisler, J.W., DeWire, S.M., Whalen, E.J., Violin, J.D., Drake, M.T., Ahn, S., Shenoy, S.K., and Lefkowitz, R.J. (2007). A unique mechanism of β-blocker action: carvedilol stimulates β-arrestin signaling. Proc. Natl. Acad. Sci. USA *104*, 16657–16662. https://doi.org/10.1073/pnas.0707936104.
47. Macdonald, C.B., Nedrud, D., Grimes, P.R., Trinidad, D., Fraser, J.S., and Coyote-Maestas, W. (2023). DIMPLE: deep insertion, deletion, and missense mutation libraries for exploring protein variation in evolution, disease, and biology. Genome Biol. *24*, 36. https://doi.org/10.1186/s13059-023-02880-6.
48. Matreyek, K.A., Stephany, J.J., Chiasson, M.A., Hasle, N., and Fowler, D.M. (2020). An improved platform for functional assessment of large protein libraries in mammalian cells. Nucleic Acids Res. *48*, e1. https://doi.org/10.1093/nar/gkz910.
49. Coyote-Maestas, W., Nedrud, D., Suma, A., He, Y., Matreyek, K.A., Fowler, D.M., Carnevale, V., Myers, C.L., and Schmidt, D. (2021). Probing ion channel functional architecture and domain recombination compatibility by massively parallel domain insertion profiling. Nat. Commun. *12*, 7114. https://doi.org/10.1038/s41467-021-27342-0.
50. Coyote-Maestas, W., Nedrud, D., He, Y., and Schmidt, D. (2022). Determinants of trafficking, conduction, and disease within a K+ channel revealed through multiparametric deep mutational scanning. eLife *11*, e76903. https://doi.org/10.7554/eLife.76903.
51. McKee, A.G., Kuntz, C.P., Ortega, J.T., Woods, H., Most, V., Roushar, F.J., Meiler, J., Jastrzebska, B., and Schlebach, J.P. (2021). Systematic profiling of temperature- and retinal-sensitive rhodopsin variants by deep mutational scanning. J. Biol. Chem. *297*, 101359. https://doi.org/10.1016/j.jbc.2021.101359.
52. Popp, N.A., Powell, R.L., Wheelock, M.K., Zapp, B.D., Holmes, K.J., Sheldon, K.M., Fletcher, S.N., Wu, X., Fayer, S., Rubin, A.F., et al. (2024). Multiplex, multimodal mapping of variant effects in secreted proteins. Preprint at bioRxiv. https://doi.org/10.1101/2024.04.01.587474.
53. Zhou, Q., Yang, D., Wu, M., Guo, Y., Guo, W., Zhong, L., Cai, X., Dai, A., Jang, W., Shakhnovich, E.I., et al. (2019). Common activation mechanism of class A GPCRs. eLife *8*, e50279. https://doi.org/10.7554/eLife.50279.
54. Zhang, K., Wu, H., Hoppe, N., Manglik, A., and Cheng, Y. (2022). Fusion protein strategies for cryo-EM study of G protein-coupled receptors. Nat. Commun. *13*, 4366. https://doi.org/10.1038/s41467-022-32125-2.
55. Tsutsumi, N., Mukherjee, S., Waghray, D., Janda, C.Y., Jude, K.M., Miao, Y., Burg, J.S., Aduri, N.G., Kossiakoff, A.A., Gati, C., et al. (2020). Structure of human Frizzled5 by fiducial-assisted cryo-EM supports a heterodimeric mechanism of canonical Wnt signaling. eLife *9*, e58464. https://doi.org/10.7554/eLife.58464.

56. Robertson, M.J., Papasergi-Scott, M.M., He, F., Seven, A.B., Meyerowitz, J.G., Panova, O., Peroto, M.C., Che, T., and Skiniotis, G. (2022). Structure determination of inactive-state GPCRs with a universal nanobody. Nat. Struct. Mol. Biol. *29*, 1188–1195. https://doi.org/10.1038/s41594-022-00859-8.

57. Brown, B.P., Stein, R.A., Meiler, J., and Mchaourab, H.S. (2024). Approximating projections of conformational Boltzmann distributions with AlphaFold2 predictions: opportunities and limitations. J. Chem. Theor. Comput. *20*, 1434–1447. https://doi.org/10.1021/acs.jctc.3c01081.

58. Sala, D., Engelberger, F., Mchaourab, H.S., and Meiler, J. (2023). Modeling conformational states of proteins with AlphaFold. Curr. Opin. Struct. Biol. *81*, 102645. https://doi.org/10.1016/j.sbi.2023.102645.

59. Mitrovic, D., McComas, S.E., Alleva, C., Bonaccorsi, M., Drew, D., and Delemotte, L. (2023). Reconstructing the transport cycle in the sugar porter superfamily using coevolution-powered machine learning. eLife *12*, e84805. https://doi.org/10.7554/eLife.84805.

60. Yee, S.W., Macdonald, C.B., Mitrovic, D., Zhou, X., Koleske, M.L., Yang, J., Buitrago Silva, D., Rockefeller Grimes, P., Trinidad, D.D., More, S.S., et al. (2024). The full spectrum of SLC22 OCT1 mutations illuminates the bridge between drug transporter biophysics and pharmacogenomics. Mol. Cell *84*, 1932–1947.e10. https://doi.org/10.1016/j.molcel.2024.04.008.

61. Mitrovic, D., Chen, Y., Marciniak, A., and Delemotte, L. (2023). Coevolution-driven method for efficiently simulating conformational changes in proteins reveals molecular details of ligand effects in the β2AR receptor. J. Phys. Chem. B *127*, 9891–9904. https://doi.org/10.1021/acs.jpcb.3c04897.

62. Hauser, A.S., Kooistra, A.J., Munk, C., Heydenreich, F.M., Veprintsev, D.B., Bouvier, M., Babu, M.M., and Gloriam, D.E. (2021). GPCR activation mechanisms across classes and macro/microscales. Nat. Struct. Mol. Biol. *28*, 879–888. https://doi.org/10.1038/s41594-021-00674-7.

63. Manglik, A., Kruse, A.C., Kobilka, T.S., Thian, F.S., Mathiesen, J.M., Sunahara, R.K., Pardo, L., Weis, W.I., Kobilka, B.K., and Granier, S. (2012). Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. Nature *485*, 321–326. https://doi.org/10.1038/nature10954.

64. Rasmussen, S.G.F., DeVree, B.T., Zou, Y., Kruse, A.C., Chung, K.Y., Kobilka, T.S., Thian, F.S., Chae, P.S., Pardon, E., Calinski, D., et al. (2011). Crystal structure of the β2 adrenergic receptor-Gs protein complex. Nature *477*, 549–555. https://doi.org/10.1038/nature10361.

65. Wang, L., Hall, C., Li, J., Choi, E., and Bai, X.-C. (2023). Structural basis of the alkaline pH-dependent activation of insulin receptor-related receptor. Nat. Struct. Mol. Biol. *30*, 661–669. https://doi.org/10.1038/s41594-023-00974-0.

66. Bushnell, B. (2014). BBTools Software Package. https://sourceforge.net/projects/bbmap.

67. Van der Auwera, G.A., and O'Connor, B.D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (O'Reilly Media, Inc.).

68. Grant, B.J., Rodrigues, A.P., ElSawy, K.M., McCammon, J.A., and Caves, L.S. (2006). Bio3D: An R package for the comparative analysis of protein structures. Bioinformatics *22*, 2695–2696.

69. Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J.H., Lindahl, E., and Scheres, S.H.W. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife *7*, e42166. https://doi.org/10.7554/eLife.42166.

70. Grant, T., Rohou, A., and Grigorieff, N. (2018). cisTEM, user-friendly software for single-particle image processing. eLife *7*, e35383. https://doi.org/10.7554/eLife.35383.

71. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D *66*, 213–221. https://doi.org/10.1107/S0907444909052925.

72. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D *60*, 2126–2132. https://doi.org/10.1107/S0907444904019158.

73. Asarnow, D., Palovcak, E., and Cheng, Y. UCSF pyem v0.5 https://doi.org/10.5281/zenodo.3576630.

74. Croll, T.I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallogr. D Struct. Biol. *74*, 519–530. https://doi.org/10.1107/S2059798318002425.

75. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82. https://doi.org/10.1002/pro.3943.

76. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

77. Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. J. Comput. Chem. *29*, 1859–1865. https://doi.org/10.1002/jcc.20945.

78. Zheng, S.Q., Palovcak, E., Armache, J.-P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods *14*, 331–332. https://doi.org/10.1038/nmeth.4193.

79. Ballesteros, J.A., and Weinstein, H. (1995). Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In Methods in Neurosciences, S.C. Sealfon, ed. (Academic Press), pp. 366–428.

80. Rao, J., Xin, R., Macdonald, C., Howard, M., Estevam, G.O., Yee, S.W., Wang, M., Fraser, J.S., Coyote-Maestas, W., and Pimentel, H. (2023). Rosace: a robust deep mutational scanning analysis framework employing position and mean-variance shrinkage. Preprint at bioRxiv. https://doi.org/10.1101/2023.10.24.562292.

81. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. J. Open Source Softw. *4*, 1686. https://doi.org/10.21105/joss.01686.

82. Kassambara, A. (2023). ggpubr: "ggplot2" Based Publication Ready Plots. https://rpkgs.datanovia.com/ggpubr/.

83. Zeileis, A., Fisher, J.C., Hornik, K., Ihaka, R., McWhite, C.D., Murrell, P., Stauffer, R., and Wilke, C.O. (2020). Colorspace: A toolbox for manipulating and assessing colors and palettes. J. Stat. Softw. *96*, 1–49. https://doi.org/10.18637/jss.v096.i01.

84. Xiao, N. (2024). ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for "ggplot2.". https://github.com/nanxstats/ggsci.

85. Pedersen, T.L. (2024). patchwork: The Composer of Plots. https://github.com/thomasp85/patchwork.

86. Venables, W.N., and Ripley, B.D. (2002). Modern applied statistics with S (Springer).

87. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296. https://doi.org/10.1038/nmeth.4169.

88. Mastronarde, D.N. (2003). SerialEM: A program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. Microsc. Microanal. *9*, 1182–1183. https://doi.org/10.1017/S1431927603445911.

89. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

90. Schüttelkopf, A.W., and van Aalten, D.M.F. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr. D *60*, 1355–1363. https://doi.org/10.1107/S0907444904011679.

91. Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. Nucleic Acids Res. *42*, D222–D230. https://doi.org/10.1093/nar/gkt1223.

92. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., et al. (2023). InterPro in 2022. Nucleic Acids Res. *51*, D418–D427. https://doi.org/10.1093/nar/gkac993.

93. Ovchinnikov, S., Kamisetty, H., and Baker, D. (2014). Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. eLife *3*, e02030. https://doi.org/10.7554/eLife.02030.

94. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926–935. https://doi.org/10.1063/1.445869.

95. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H., and MacKerell, A.D., Jr. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat. Methods *14*, 71–73. https://doi.org/10.1038/nmeth.4067.

96. Aho, N., Buslaev, P., Jansen, A., Bauer, P., Groenhof, G., and Hess, B. (2022). Scalable constant pH molecular dynamics in GROMACS. J. Chem. Theor. Comput. *18*, 6148–6160. https://doi.org/10.1021/acs.jctc.2c00516.

97. Buslaev, P., Aho, N., Jansen, A., Bauer, P., Hess, B., and Groenhof, G. (2022). Best practices in constant pH MD simulations: accuracy and sampling. J. Chem. Theor. Comput. *18*, 6134–6147. https://doi.org/10.1021/acs.jctc.2c00517.

98. Jansen, A., Aho, N., Groenhof, G., Buslaev, P., and Hess, B. (2024). Phbuilder: A tool for efficiently setting up constant pH molecular dynamics simulations in GROMACS. J. Chem. Inf. Model. *64*, 567–574. https://doi.org/10.1021/acs.jcim.3c01313.

99. Nosé, S. (1984). A unified formulation of the constant temperature molecular dynamics methods. J. Chem. Phys. *81*, 511–519. https://doi.org/10.1063/1.447334.

100. Hess, B. (2008). P-LINCS: A parallel linear constraint solver for molecular simulation. J. Chem. Theor. Comput. *4*, 116–122. https://doi.org/10.1021/ct700200b.

101. Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. J. Chem. Phys. *98*, 10089–10092. https://doi.org/10.1063/1.464397.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
SureLight® APC Anti-DDDDK tag antibody [M2]	Abcam	Cat# ab72569; RRID: AB_1310127
Bacterial and virus strains		
E. coli: MegaX DH10B $T1^R$ Electrocompetent Cells	ThermoFisher	Cat# C640003
E. coli: XL1-Blue chemically competent cells	QB3 MacroLab, UC Berkeley	N/A
E. coli: BL21 Rosetta	QB3 MacroLab, UC Berkeley	N/A
Chemicals, peptides, and recombinant proteins		
PirmeStar GXL DNA polymerase	Takara	Cat# R050A
NEBridge Golden Gate Assembly Kit (BsaI-HF® v2)	NEB	Cat# E1601S
NEBridge Golden Gate Assembly Kit (BsmBI-HF® v2)	NEB	Cat# E1602S
AP1903	MedChemExpress	Cat# HY-16046
ProNex size-selective purification system	Promega	Cat# NG2001
Lipofectamine 3000	ThermoFisher	Cat# L3000008
Puromycin	Life Technologies Corporation	Cat# NC2380542
ExpiFectamine 293 transfection kit	ThermoFisher	Cat# A14524
Ro-20-1724	Millipore Sigma	Cat# 557502
Doxycycline Hyclate	Sigma Aldrich	Cat# D9891
L-MNG	Anatrace	Cat# NG310
Cholesteryl hemisuccinate	Steraloids	Cat# C6823
GDN	Anatrace	Cat# GDN101
MS48104	MedChemExpress	Cat# HY-134494
Critical commercial assays		
Nextera XT DNA Library Preparation Kit	Illumina	Cat# FC-131-1096
IDT for Illumina Nextera DNA Unique Dual Indexes Set C	Illumina	Cat# 20042666
Agilent D5000 high sensitivity screen tape	Agilent	Cat# 5067-5592
Qubit 1x dsDNA high sensitivity assay	ThermoFisher	Cat# Q33230
Quick-DNA Microprep Plus Kit	Zymo Research	Cat# D4074
Q5 Site-Directed Mutagenesis Kit	NEB	Cat# E0554S
GloSensor cAMP Assay	Promega	Cat# E2301
Deposited data		
GPR68 deep mutational scan fitness scores	This study	MAVEdb: 00001207
Raw GPR68 DMS sequencing reads	This study	NCBI SRA: PRJNA1062987
GPR68 molecular dynamics data, setup, and processing files	This study	Zenodo: https://doi.org/10.5281/zenodo.13785406
Coordinates: GPR4-miniGs complex	This study	PDB: 9BIP
Coordinates: GPR65-miniGs complex	This study	PDB: 9BHL
Coordinates: GPR68-miniGs complex	This study	PDB: 9BHM
Coordinates: GPR68-miniGs/q complex	This study	PDB: 9BI6
Experimental models: Cell lines		
Human HEK293T	ATCC	Cat# CRL-3216
Human: HEK-293T LLP-iCasp9-Blast Clone 4	Matreyek et al.[48]	N/A
Human Expi293F inducible cells	ThermoFisher	Cat# A39241
Trichoplusia ni Hi5 insect cells	Expression Systems	N/A
Spodoptera frugiperda Sf9 insect cells	Expression Systems	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
Cell_line_5_for: GGAGCAATTCCACAACACTTTTGTC	Yee et al.[60]	NA
Puro_rev: CGGCGAAAGCAGCTGCC	Yee et al.[60]	NA
Primers and oligos for GPR68 mutational library generation see Table S6	This study	NA
Recombinant DNA		
pLP_mod_mNeonGreen11_P2A_puroR_CRE-GFP	This study	Addgene: 229227
pCAG-NLS-BxB1	Matreyek et al.[48]	Addgene: 51271
pcDNA3.1-zeo-tetO-GPR4	This study	N/A
pcDNA3.1-zeo-tetO-GPR65	This study	N/A
pcDNA3.1-zeo-tetO-GPR68	This study	N/A
pcDNA3.1-zeo-tetO-GPR4-miniGs	This study	N/A
pcDNA3.1-zeo-tetO-GPR65-miniGs	This study	N/A
pcDNA3.1-zeo-tetO-GPR68-miniGs	This study	N/A
pcDNA3.1-zeo-tetO-GPR68-miniGs/q	This study	N/A
Software and algorithms		
BBTools	Bushnell[66]	https://jgi.doe.gov/data-and-tools/software-tools/bbtools/
GATK	Van der Auwera et al.[67]	https://gatk.broadinstitute.org/hc/en-us
Enrich2	Rubin et al.[45]	https://github.com/FowlerLab/Enrich2
Bio3D R package	Grant et al.[68]	http://thegrantlab.org/bio3d/
CryoSparc v3	Structura Biotechnology	https://structura.bio/
Relion	Zivanov et al.[69]	https://www3.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page
cisTEM	Grant et al.[70]	https://cistem.org/
Phenix	Adams et al.[71]	http://www.phenix-online.org/
Coot	Emsley and Cowtan[72]	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Pyem	Asarnow et al.[73]	https://github.com/asarnow/pyem
ISOLDE	Croll[74]	https://tristanic.github.io/isolde/
UCSF ChimeraX	Pettersen et al.[75]	https://www.cgl.ucsf.edu/chimerax/
Pymol	Schrodinger	https://pymol.org/
Prism v10	Graphpad	https://www.graphpad.com/
GROMACS	Abraham et al.[76]	https://www.gromacs.org/
CHARMM-GUI	Jo et al.[77]	https://www.charmm-gui.org/
UCSF MotionCor2	Zheng et al.[78]	https://emcore.ucsf.edu/ucsf-software
Other		
Code for DMS sequencing processing pipeline	This study	Zenodo: https://doi.org/10.5281/zenodo.10999528
Code for DMS sequencing QC	This study	Zenodo: https://doi.org/10.5281/zenodo.10999522
Code for DMS data analysis and figure generation	This study	Zenodo: https://doi.org/10.5281/zenodo.11056584
Titan Krios	ThermoFisher	N/A
BioQuantum electron detector	Gatan	N/A
K3 direct electron detector	Gatan	N/A
Holey Gold Supports: UltraAUFoil (R 1..2/1.3) Au300 mech	Quantfoil Micro Tools	NA

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Code for GPR68 molecular dynamics simulation setup and processing	This study	Zenodo: https://doi.org/10.5281/zenodo.13785406
Cytoflex SRT	Beckman Coulter	N/A
Tapestation 4200	Agilent	N/A
Novaseq 6000	Illumina	N/A

EXPERIMENTAL MODE AND STUDY PARTICIPANT DETAILS

Cell lines

HEK293T cells used for GloSensor studies were obtained from ATCC. HEK293T LLP-iCasp9 cells used in the DMS study were a gift from Doug Fowler.[48] All HEK cells were cultured at 37 °C with 5% CO_2. Exact media formulations and pH conditions are noted in the method details section for each experiment. Several cell lines were used as expression systems to obtain components for GPCR complexes in cryo-EM studies. Expi293F-TetR cells were obtained from ThermoFisher, *Trichoplusia ni* Hi5 insect cells were obtained from Expression Systems, and BL21 Rosetta *Escherichia coli* cells were obtained from UC Berkeley QB3 MacroLab. Each cell line was cultured under their recommended conditions with any specific modifications noted in the purification method details.

METHOD DETAILS

GloSensor cAMP assays

Proton-sensing GPCR G_s activation and cAMP production were determined using the GloSensor cAMP assay. The following method was adopted from a previously published procedure with modifications.[40] In detail, HEK293T cells were maintained at 37 °C with 5% CO_2 and cotransfected with receptor DNA and GloSensor cAMP reporter plasmids in DMEM containing 10% FBS. Overnight transfected cells were plated in poly-l-lysine coated 384-well white clear-bottom plates in DMEM supplemented with 1% dialyzed fetal bovine serum (dFBS), about 15,000 cells in 40 μL per well, for a minimum of 6 h up to 24 h. Assay buffers were prepared in 1x Calcium- and Magnesium-free HBSS supplemented with different organic buffer agents for different pH ranges, 20 mM MES for pH 5.00–6.60, 20 mM HEPES for pH 6.70–8.20, and 20 mM TAPS for pH 8.30–8.60. pH was adjusted with KOH at room temperature. PDE inhibitor Ro 20-1724 at final concentration of 10 μM was added to working solutions, just before the assays. To stimulate cells with desired pH solutions, cells were first removed of medium (gently shaking off) and stimulated with desired pH solutions (25 μl/well) supplemented with 2 mM luciferin. The cell plate was incubated at room temperature for 20-30 min before luminescence was counted. For stimulation solutions with pH below 6.0, cells (medium was not removed) first received 10 μl pH 7.4 assay buffer containing luciferin (final 2 mM) and Ro 20-1724 (final of 10 μM) for a minimum of 30 min. After luciferin loading, medium and luciferin solutions were removed; cells were then stimulated with desired pH solutions containing 2 mM luciferin and 10 μM Ro 20-1724 as above. The cell plate was incubated at room temperature for 20 - 30 min before counting. Data presented in Figures here has been normalized to % max response or fold of basal, pooled for analysis using the built-in 4 parameter logistic function in the GraphPad Prism V10. Full tables of pharmacologic parameters can be found in Tables S1 and S3–S5. Statistical analyses were performed using one-way ANOVA followed by Dunnett's multiple comparison test between all mutants and their respective wildtype receptor using GraphPad Prism version 10.3.1.

GPR4-GPR68 chimeric receptor design

A sequence alignment of GPR4 and GPR68 amino acid sequences was generated using Clustal Omega (Figure S1). Grafting points were chosen by matching the final Ballesteros-Weinstein position ("BW", superscript numbers, indicate common GPCR numbering where first value is the helix and second value is residue position) position where GPR4 and GPR68 shared residue identity before diverging on the extracellular side.[79] Linear segments of GPR68 were then transferred to GPR4 alone and in combination with other segments. This led to a set of chimeric constructs that contain portions of GPR68 spanning extracellular loops (ECL) and the extracellular portions of the transmembrane (TM) helices.

GPR68 Deep Mutational Scan

GPR68 mutational library generation

Our DIMPLE platform was used to generate the GPR68 deep mutational library.[47] We designed the library to contain all missense mutations at each position in GPR68. We additionally included synonymous mutations, insertions of G/GS/GSG, and deletions of one, two, and three amino acids at each position. These mutations were encoded in oligos with flanking BsaI sites and then ordered as a SurePrint Oligonucleotide library (Agilent Technologies)(Table S6). Sublibrary-specific primers to amplify oligonucleotides from the pool were ordered from IDT (Table S6). Oligonucleotides were resuspended in TE buffer to 10 ng/μL. Sublibrary fragments were

amplified using PrimeStar GXL DNA polymerase and fragment-specific primers (Table S6). PCR conditions were as follows: 20 cycles of 98 °C for 10 sec, 55 °C for 10 sec, and 68 °C for 30 sec. These reactions were subjected to PCR cleanup using Zymo Clean and Concentrate-5 kits. The cDNA sequence of GPR68 WT was synthesized by Twist Bioscience in their High Copy Number kanamycin backbone, BsmBI and BsaI cutsites were removed using synonymous mutations. For each library fragment, this plasmid was amplified to add BsaI sites. PCR conditions were as follows: 20 cycles of 98 °C for 10 sec, 55 °C for 10 sec, and 68 °C per kb. Bands were gel purified using Zymo Gel DNA Recovery kits. The WT backbone fragment and corresponding oligo sublibrary were assembled using BsaI-mediated Golden Gate assembly. Golden Gate reactions were scaled to 40 μL and run in a thermocycler for 35 cycles of 37 °C for 5 min followed by 16 °C for 10 min. These reactions were cleaned using Zymo Clean and Concentrate-5 kits (eluting with 10 μL water). Following this, 3 μL was transformed into electrocompetent MegaX DH10B cells, recovered, and added to 30 mL LB + Kanamycin and grown while shaking until they reached OD 0.6-0.7. DNA was isolated using a Zymo Zyppy Plasmid Miniprep kit. Each sublibrary was quantified using Invitrogen Qubit dsDNA HS assay kit and pooled in equimolar ratios. This pooled library was then assembled into a landing pad compatible cAMP-GFP reporter vector containing a GSGSGS-P2A-PuroR cassette for positive selection.

GPR68 library cell line generation

The HEK293T LLP-iCasp9 cells used in this study were a gift from Doug Fowler (UW).[48] Cells were cultured in "D10" media (DMEM, 10% dialyzed FBS, 1% sodium pyruvate, and 1% penicillin/streptomycin) inside humidified incubators at 37 °C and 5% CO_2. Cell lines for GPR68 WT and the GPR68 mutational library were generated as follows. 1 μg of DNA was cotransfected with 1 μg BxB1 recombinase (pCAG-NLS-BxB1, Addgene #51271) using 3.75 μL lipofectamine 3000 and 5 μL P3000 reagent per well of a 6 well dish. For GPR68 and β2AR WT, 2 wells were transfected and pooled following selection. For the GPR68 mutational library, 18 wells were transfected in parallel and pooled following selection. The landing pad in the cell line contains a Tet-on promoter upstream of the BxB1 recombination site and a split rapamycin analog inducible dimerizable Casp-9. Two days after transfection, expression off the landing pad is induced with doxycycline hyclate (2 μg/mL). After two days of induction, cells are treated with a combination of 2 μg/mL doxycycline and 10 nM AP1903. Successfully recombined cells will have shifted the iCasp-9 cassette out of frame while unrecombined cells will express the cassette and upon treatment with AP1903 die from iCasp-9 induced apoptosis. Cells were selected for 2 days in AP1903 after which they were transitioned back to D10 supplemented with doxycycline. After two days of recovery, cells were transitioned to D10 supplemented with both doxycycline and puromycin (2 μg/mL) to select for cells that have proper in-frame, full-length assemblies. Following puromycin selection for two days, cells were transitioned to D10 and expanded before freezing down or using in subsequent assays. The GPR68 mutational library pooled cell line was passaged and/or frozen at >200x coverage to maintain diversity.

Fluorescence activated cell sorting

For flow-based assays and cell sorting, frozen stocks of cells were thawed and allowed to recover for several days in D10 media. 48 h prior to starting the experiment, cells were split into an appropriate sized dish such that they reach ~75% confluency by the start of the sort. 36 h prior to starting the assay, cells were induced with doxycycline hyclate (2ug/mL). Doxycycline was subsequently washed out after 24 h and cells were maintained in D10 for the remaining 12 h prior to starting the cAMP assay or staining for surface expression. For the cAMP signaling assays, the pH of D10 media was adjusted using NaOH/HCl on the same day as the assay. The cAMP assay was run as follows: cells were swapped to D10 (at indicated pH) with trimethoprim (10 μM) to stabilize the DHFR degron for 8 h. After this incubation, cells were detached using TrpyLE Express (Gibco), washed with PBS + 3% FBS, and resuspended in BD FACS buffer (BD). The surface expression assay was run similarly, cells were detached using TrypLE Express after induction, washed, incubated with M2 FLAG APC-Surelight antibody (1:500, Abcam) for 1 h, washed 3x with PBS + 3% FBS, and then resuspended in BD FACS buffer and kept covered on ice prior to sorting. Cell sorting was performed using a Cytoflex SRT. Briefly, cells were gated on FSC-A and SSC-A for HEK293T cells, then FSC-A and FSC-H for singlets. For the cAMP assay, we assessed activity using eGFP on the FITC-A channel, and for surface expression assays, the APC-A channel. For the cAMP sorting experiments, the population was split into four roughly equal populations (% cells) based on the most active condition, pH 5.5 + 30 μM Ogerin. These gates were maintained for all subsequent samples. For surface expression assays, the population was bimodal. Gates were set at the peaks of each distribution and the intervening trough. For sorting experiments we aimed to collect cells equal or greater than 100x the expected number of variants in our library.

Genomic DNA extraction and sequencing

Following cell sorting, genomic DNA (gDNA) was immediately extracted from cells using Quick-DNA Zymo Microprep Plus kits. A plot of the amount of gDNA extracted vs estimated genome equivalents recovered for each sample is included in Figure S4. All resultant gDNA was used as template for PCR using PrimeStar GXL to generate amplicons of the target gene using cell_line_for_5 and Puro_rev primers. PCR conditions were as follows: 20 cycles of 98 °C for 10 sec, 55 °C for 15 sec, 68 °C for 90 sec. PCR reactions were stored at 4 C prior to further processing. PCR reactions were pooled for each sample then concentrated using Zymo DNA Clean and Concentrator-25 kits, mixed with NEB Purple Loading dye (6x, no SDS) and run on a 1% agarose 1x TBE gel. Target amplicons were excised and purified using Zymo Gel DNA Recovery kits. Amplicon DNA concentrations were then quantified using Invitrogen Qubit dsDNA HS assay kit. Libraries were prepared for deep sequencing using the Illumina Nextera XT DNA Library prep kit. Libraries were indexed using the IDT for Nextera Unique Dual Indexes set C. Then, the lengths of indexed libraries were quantified using the

Agilent TapeStation HS D5000 assay and concentrations were determined using Invitrogen Qubit dsDNA HS assay kit. Samples were normalized and pooled equimolar and then paired-end sequenced at ~200x coverage (SP) on a NovaSeq6000.

Next generation sequencing data processing

Sequencing files were obtained from the sequencing core as fastq.gz after demultiplexing. The experiment was processed using a DMS-specific pipeline we have developed.[80] The pipeline implemented the following steps: first, adapter sequences and contaminants were removed using BBDuk, then paired reads were error corrected with BBMerge and mapped to the reference sequence using BBMap with 15-mers (all from BBTools[66]). Variants in the mapped SAM file were called using the AnalyzeSaturationMutagenesis tool in GATK v4.[67] The output of this tool is a CSV containing the genotype of each distinct variant as well as the total number of reads for each sample. This was then further processed using a python script which filtered out sequences that were not part of the designed variants.

Variant scoring

Variant scores were calculated from the collected, processed files using weighted least squares and normalized using synonymous sequences. Briefly, by sorting into 4 bins and sequencing we are able to determine the frequency of each variant across the range of fluorescence intensities being sorted into each bin. These frequencies are used to construct the score for each variant. To do this we used a statistical package, Enrich2[45] that calculates a score and standard error for each variant based on a fitted weighted-least-squares regression across a series of selection points (e.g. different sort bins).[45] Mathematically this is expressed as: $M_{v,b} = \log_{10}[(c_{v,b} + 0.5) / (c_{syn,b} + 0.5)]$. Where v is the variant, b is the sort bin number, c is the variant count. In practice, 0.5 is added to each count to assist with very small counts. Each variant's score is defined as the $\log_{10}$-fold transform of the slope of the regression line, which is weighted to wildtype/synonymous frequencies. This results in synonymous variants being centered at a value of zero. Positive scores/slopes represent enrichment greater than synonymous. Negative scores/slopes represent enrichment lower than synonymous. In the context of our assays, this means that variants with higher signaling or surface expression than synonymous have positive scores, and variants with lower signaling or surface expression than WT/synonymous have negative scores. The final scores (Table S7) were then processed and plotted using R. A copy of this processing pipeline, sequencing counts, and scores has been deposited in the Github/Zenodo repositories listed in the data availability section. Additionally, fitness scores have been deposited in MaveDB, entry details are in the data availability section.

GPR68 Deep Mutational scanning data analysis

Deep mutational scanning data were analyzed in R. In particular, we would like to acknowledge use of the following packages: Tidyverse,[81] ggpubr,[82] colorspace,[83] ggsci,[84] patchwork,[85] gridExtra, cowplot, bio3d,[68] and MASS.[86] Heatmaps of cAMP signaling pH 5.5 (Figures 4G and S5), cAMP signaling pH 6.5 (Figure S5), surface expression (Figure S5), and associated standard error heatmaps (Figure S5) are generated from the direct output of the Enrich2 scores described above. Expression adjusted signaling scores (Figures 5C–5F) were calculated as follows. Raw scores from Enrich2 were filtered to remove insertion and deletion mutations. A linear regression model used to describe the relationship between the surface expression and cAMP signaling scores of the synonymous variants. The Euclidian distance of each variant to the synonymous regression line was calculated. Variants with a higher signaling score than synonymous at a given level of expression were classified as having positive distances, gain-of-function ("GOF"). Variants with a lower signaling score than synonymous at a given level of expression were classified as having negative distances, loss-of-function ("LOF"). At each position the average value of all positive distances corresponds to the positional GOF score. Similarly, the average value of all negative distances corresponds to the positional LOF score. These positional scores denoted as "expression adjusted signaling scores" for each the pH 5.5 and pH 6.5 screens. We have highlighted the highest and lowest 5% of positions in Figure 5E (pH 5.5). Scripts used to make figures have been deposited in a Github/Zenodo repository listed in the data availability section.

Proton senor expression and purification

The human GPR4, GPR65, and GPR68 genes with an N-terminal influenza hemagglutinin signal sequence and Flag epitope tag were cloned into a pcDNA3.1/Zeo vector containing a tetracycline inducible cassette. The miniG proteins ($miniG_{s399}$ for GPR4 and GPR65 and $miniG_{s/q70}$ for GPR68) were fused to the C terminus of each proton sensor preceded by a glycine/serine linker and rhinovirus 3C protease recognition site.[38] The resulting fusion constructs were transfected into inducible Expi293F-TetR cells (Thermo Fisher) using the ExpiFectamine transfection reagent per manufacturer instructions. After 18 h, protein expression was induced with 1 μg/mL doxycycline hyclate for 24 h before collection by centrifugation. Pelleted cells were washed with 50 mL phosphate buffered saline, pH 7.5 before storage at −80 °C. For receptor purification, frozen cells were hypotonically lysed in 20 mM MES, pH 6, 1 mM EDTA, 160 μg/mL benzamidine, 2 μg/mL leupeptin for 10 min at 25 °C. The membrane fraction was collected by centrifugation, and the fusion protein was extracted with 20 mM MES, pH 6, 300 mM NaCl, 1% (w/v) lauryl maltose neopentyl glycol (L-MNG, Anatrace), 0.1% (w/v) cholesteryl hemisuccinate (CHS, Steraloids), 2 mM $MgCl_2$, 2 mM $CaCl_2$, 160 μg/mL benzamidine, 2 μg/mL leupeptin with dounce homogenization and incubation with stirring for one hour at 4 °C. The soluble fraction was separated from the insoluble fraction by centrifugation and was incubated in batch for 1 h at 4 °C with homemade M1–Flag antibody-conjugated Sepharose beads. Sepharose resin was then washed extensively with 20 mM MES, pH 6, 150 mM NaCl, 0.1% (w/v) L-MNG, 0.01% (w/v)

CHS, 2 mM $MgCl_2$, 2 mM $CaCl_2$ and then with 20 mM MES, pH 6, 150 mM NaCl, 0.0075% (w/v) L-MNG, 0.00075% (w/v) CHS, 2 mM $MgCl_2$, 2 mM $CaCl_2$ prior to elution with 20 mM MES, pH 6, 150 mM NaCl, 0.0075% (w/v) L-MNG, 0.00075% (w/v) CHS, 5 mM EDTA, 0.2 mg/mL Flag peptide. Eluted protein was concentrated in a 100 kDa MWCO Amicon spin concentrator, and injected onto a Superdex200 Increase 10/300GL (Cytiva) gel filtration column equilibrated in 20 mM MES, pH 6, 150 mM NaCl, 0.0075% (w/v) L-MNG, 0.0025% glyco-diosgenin (GDN, Anatrace), and 0.0005% CHS. Monodisperse fractions were complexed with $G_{\beta1\gamma2}$ heterodimer and Nb35 at 2 molar excess overnight at 4℃. The next day, the heterotrimeric complex was concentrated with a 100 kDa MWCO spin concentrator and excess $G_{\beta1\gamma2}$ and Nb35 was removed via size-exclusion chromatography, using a Superdex200 Increase 10/300 GL column (GE Healthcare) equilibrated in 20 mM MES pH 6, 150 mM NaCl, 0.00075% (w/v) L-MNG, 0.00025% (w/v) GDN, and 0.0001% CHS. Resulting heterotrimeric complex was concentrated with a 100 kDa MWCO spin concentrator for preparation of cryo-EM grids. For GPR68-miniGs structure we attempted to solve with positive allosteric modulator MS48107 bound, the compound was included in all buffers at a concentration of 40μM.

Expression and purification of $G_{\beta1\gamma2}$

Human $G_{\beta1\gamma2}$ heterodimer was expressed in *Trichoplusia ni* Hi5 insect cells (Expression Systems) using a single baculovirus generated in *Spodoptera frugiperda* Sf9 insect cells (Expression Systems). A bicistronic pVLDual construct contained the $G_{\beta1}$ subunit with a N-terminal 6 × His tag, and an untagged human $G_{\gamma2}$ subunit. For expression, Hi5 insect cells were transduced with baculovirus at a density of ~3.0 × 10^6 cells per mL, grown with 27 °C shaking at 130 rpm. 48 h post-transduction, cells were collected and washed in a hypotonic buffer containing 20 mM HEPES, pH 8.0, 5 mM β-mercaptoethanol (β-ME), and protease inhibitors (20 μg/mL leupeptin, 160 μg/mL benzamidine). The membrane fraction was then separated by centrifugation and solubilized with 20 mM HEPES pH 8.0, 100 mM sodium chloride, 1.0% sodium cholate, 0.05% dodecylmaltoside (Anatrace), and 5 mM β-mercaptoethanol (β-ME). Solubilized $G_{\beta1\gamma2}$ heterodimer was then incubated with HisPur Ni-NTA resin (Thermo Scientific) in batch. Bound $G_{\beta1\gamma2}$ heterodimer was washed extensively and detergent was slowly exchanged to 0.1% (w/v) lauryl maltose neopentyl glycol (L-MNG, Anatrace) and 0.01% CHS before elution with 20 mM HEPES pH 7.5, 100 mM NaCl, 0.1% L-MNG, 0.01% CHS, 270 mM imidazole, 1 mM dithiothreitol (DTT), and protease inhibitors. Eluted $G_{\beta1\gamma2}$ heterodimer was pooled and rhinovirus 3C protease was added to cleave the N-terminal 6 × His tag during overnight dialysis in 20 mM HEPES pH 7.5, 100 mM NaCl, 0.02% L-MNG, 0.002% CHS, 1 mM DTT, and 10 mM imidazole. To remove uncleaved $G_{\beta1\gamma2}$, dialysed material was incubated with HisPur Ni-NTA resin in batch. The unbound fraction was then incubated for 1 h at 4 °C with lambda phosphatase (New England Biolabs), calf intestinal phosphatase (New England Biolabs), and Antarctic phosphatase (New England Biolabs) for dephosphorylation. Final anion exchange chromatography was performed using a MonoQ 4.6/100 PE (Cytiva) column to purify only geranylgeranylated heterodimer. The resulting protein was pooled and dialysed overnight in 20 mM HEPES pH 7.5, 100 mM NaCl, 0.02% L-MNG, and 100 μM TCEP, and concentrated with a 3 kDa centrifugal concentrator to a final concentration of 162 μM. Glycerol was added to a final concentration of 20%, and the protein was flash frozen in liquid nitrogen and stored at −80 °C until further use.

Expression and purification of Nb35

A pET-26b vector containing the Nb35 sequence with a carboxy-terminal Protein C affinity tag was transformed into BL21 Rosetta Escherichia coli cells (UC Berkeley QB3 MacroLab) and inoculated into 8 L of Terrific Broth supplemented with 0.1% glucose, 2 mM $MgCl_2$, and 50 μg/mL kanamycin. Cells were induced with 400 μM IPTG at A600 of 0.6 and allowed to express at 20 °C for 21 h. Collected cells were incubated SET Buffer (200 mM Tris pH 8.0, 500 mM sucrose, 0.5 mM EDTA) in the presence of protease inhibitors (20 μg/mL leupeptin, 160 μg/mL benzamidine) and benzonase. To initiate hypotonic lysis, two volumes of deionized water were added to the cell mixture after 30 min of SET buffer mixing. Following lysis, NaCl was added to 150 mM, $CaCl_2$ was added to 2 mM, and $MgCl_2$ was added to 2 mM and lysate was centrifuged to remove the insoluble fraction. Supernatant was incubated with homemade anti-Protein C antibody-coupled Sepharose. Nb35 was eluted with 20 mM HEPES pH 7.5, 100 mM NaCl, and 2 mM $CaCl_2$, 0.2 mg/mL protein C-peptide, and 5 mM EDTA pH 8.0, concentrated in a 10 kDa MWCO Amicon filter and injected over a Superdex S75 Increase 10/300 GL column (Cytiva) size-exclusion chromatography column equilibrated in 20 mM HEPES pH 7.5, 100 mM NaCl. Monodisperse Nb35 fractions were pooled, concentrated, and flash frozen in liquid nitrogen for storage at −80 °C until further use.

Cryo-EM Methods

GPR4-G_s pH 6 complex sample preparation

The GPR4-G_s pH 6 complex was concentrated to 14 mg/mL and supplemented with 0.05% CHAPS (Thermo Fisher) to improve the distribution of protein molecules in ice, and 3 μL was applied onto a glow-discharged 300 mesh 1.2/1.3 gold grid covered in a holey gold film (UltrAufoil). Excess sample was removed with a blotting time of 4 s and a blotting force of 1 at 4 °C prior to plunge freezing into liquid ethane using a Vitrobot Mark IV (Thermo Fisher).

GPR4-G_s pH 6 complex data collection

A total of 9,018 movies were recorded with a K3 detector (Gatan) on a Titan Krios (Thermo Fisher) microscope operated at 300 keV with a BioQuantum post-column energy filter set to a zero-loss energy selection slit width set of 20 eV. Movies were recorded using dose-fractionated illumination at a nominal magnification of 105,000x (physical pixel size of 0.86 Å/pixel) and a defocus range of -1 to -2.1 μm for a total dose of 50.7 e^-/Å.^{2} Exposure areas were acquired with image shift collection using EPU (Thermo Fisher).

GPR4-G_s pH 6 complex data processing

Movies of the GPR4-G_s pH 6 complex were motion-corrected and dose-fractionated using UCSF MotionCor2.[78] Corrected micrographs were imported into cryoSPARC v3[87] for CTF estimation via the Patch Estimation job. Micrographs with estimated CTF fit resolution > 5 Å were removed before further processing. Templates for particle picking were generated from the same complex reconstructed from a previous 200 keV imaging session. Particle picking templates were low-pass filtered to 20 Å and used to pick 8,608,607 particles. After picking, particles were extracted in a 288 pixel box and Fourier cropped to 48 pixels before 3D classification with alignment using a 20 Å low-pass filtered reconstruction and three random reconstructures generated from a prematurely truncated ab initio reconstruction job, called "garbage collectors," with the Heterogeneous Refinement job type. Two rounds of Heterogeneous Refinement yielded 2,501,915 particles that were re-extracted in the same box size cropped to 72 pixels and classified in a third Heterogeneous Refinement job. The resulting 1,453,906 particles were re-extracted in the same box cropped to 144 pixels. A fourth round of Heterogeneous Refinement and 2D classification, yielded 878,077 particles that were extracted without cropping. A final round of Heterogeneous Refinement yielded 439,296 particles that were refined using the Non-Uniform Refinement job type giving the final full-particle map. Finally, local refinement using an inclusion mask covering the 7TM domain was performed, using poses/shift Gaussian priors with standard deviation of rotational and shift magnitudes limited to 3° and 2 Å, respectively.

GPR65-G_s pH 6 complex sample prep

The GPR65-G_s pH 6 complex was concentrated to 11 mg/mL and supplemented with 0.05% CHAPS (Thermo Fisher) to improve the distribution of protein molecules in ice, and 3 μL was applied onto a glow-discharged 300 mesh 1.2/1.3 gold grid covered in a holey gold film (UltrAufoil). Excess sample was removed with a blotting time of 4 s and a blotting force of 1 at 4 °C prior to plunge freezing into liquid ethane using a Vitrobot Mark IV (Thermo Fisher).

GPR65-G_s pH 6 complex data collection

A total of 8,294 movies were recorded with a K3 detector (Gatan) on a Titan Krios (Thermo Fisher) microscope operated at 300 keV with a BioQuantum post-column energy filter set to a zero-loss energy selection slit width set of 20 eV. Movies were recorded using dose-fractionated illumination at a nominal magnification of 105,000x (physical pixel size of 0.835 Å/pixel) and a defocus range of -1 to -2.1 μm for a total dose of 46 $e^-/Å^2$. Exposure areas were acquired with image shift collection using SerialEM 3.8.[88]

GPR65-G_s pH 6 complex data processing

Movies of the GPR65-G_s pH 6 complex were motion-corrected and dose-fractionated using UCSF MotionCor2.[78] Corrected micrographs were imported into cryoSPARC v3.1 for CTF estimation via the Patch Estimation job.[87] Micrographs with estimated CTF fit resolution > 5 Å were removed before further processing. Templates for particle picking were generated from the same complex reconstructed from a previous 200 keV imaging session. Particle picking templates were low-pass filtered to 20 Å and used to pick 8,673,428 particles. After picking, particles were extracted in a 288 pixel box and Fourier cropped to 48 pixels before 3D classification with alignment using a 20 Å low-pass filtered reconstruction and "garbage collectors" with the Heterogeneous Refinement job type. Two rounds of Heterogeneous Refinement yielded 2,588,765 particles that were re-extracted in the same box size cropped to 74 pixels and classified in two Heterogeneous Refinement jobs. The resulting 1,637,819 particles were re-extracted in the same box cropped to 150 pixels and further classified with two rounds of Heterogeneous Refinement and 2D classification. The resulting 1,055,443 particles were refined using the Non-Uniform Refinement job type. Particles were exported using csparc2star.py from the pyem script package, and a mask covering the 7TM domain of GPR65 was generated using the Segger tool in UCSF ChimeraX and the Volume Tools utility in cryoSPARC.[73,75] The particles and mask were imported into Relion v3.0 and classified in 3D without alignment through three separate iterations.[69] Particles comprising the three highest resolution classes were reimported into cryoSPARC for Non-Uniform Refinement. Finally, particles were exported into cisTEM for 7TM local refinements using the Manual Refinement job type and low-pass filtering outside of the mask.[70]

GPR68-$G_{s/q}$ pH 6 complex sample prep

The GPR68-G_q pH 6 complex was concentrated to 4 mg/mL and 3 μL was applied onto a glow-discharged 300 mesh 1.2/1.3 gold grid covered in a holey carbon film (Quantifoil). Excess sample was removed with a blotting time of 4 s and a blotting force of 1 at 4 °C prior to plunge freezing into liquid ethane using a Vitrobot Mark IV (Thermo Fisher).

GPR68-$G_{s/q}$ pH 6 complex data collection

A total of 6,650 movies were recorded with a K3 detector (Gatan) on a Titan Krios (Thermo Fisher) microscope operated at 300 keV with a BioQuantum post-column energy filter set to a zero-loss energy selection slit width set of 20 eV. Movies were recorded using dose-fractionated illumination at a nominal magnification of 105,000x (physical pixel size of 0.873 Å/pixel) and a defocus range of -1 to -2.1 μm for a total dose of 50 $e^-/Å^2$. Exposure areas were acquired with image shift collection using EPU (Thermo Fisher).

GPR68-$G_{s/q}$ pH 6 complex data processing

Movies of the GPR68-G_q pH 6 complex were motion-corrected and dose-fractionated using UCSF MotionCor2.[78] Corrected micrographs were imported into cryoSPARC v3.1 for CTF estimation via the Patch Estimation job.[87] Micrographs with estimated CTF fit resolution > 5 Å were removed before further processing. Templates for particle picking were generated from the same complex reconstructed from a previous 200 keV imaging session. Particle picking templates were low-pass filtered to 20 Å and used to pick 6,764,523 particles. After picking, particles were extracted in a 288 pixel box and Fourier cropped to 72 pixels before 3D classification with alignment using a 20 Å low-pass filtered reconstruction and "garbage collectors" with the Heterogeneous Refinement job type. Two rounds of Heterogeneous Refinement yielded 2,774,555 particles that were re-extracted in the same box size cropped to 192 pixels and classified in an additional Heterogeneous Refinement job. The resulting 1,144,750 particles were refined using the

Non-Uniform Refinement job type. Particles were exported using csparc2star.py from the pyem script package, and a mask covering the 7TM domain of GPR68 was generated using the Segger tool in UCSF ChimeraX and the mask.py pyem script.[69,70,75] The particles and mask were imported into Relion v3.0 and classified in 3D without alignment.[69] Particles comprising the highest resolution class were reimported into cryoSPARC for Non-Uniform Refinement. Finally, particles were exported into cisTEM for 7TM local refinements using the Manual Refinement job type and low-pass filtering outside of the mask.[70]

GPR68-G_s pH 6 complex sample prep

The GPR68-G_s pH 6 complex was concentrated to 4 mg/mL and 3 μL was applied onto a glow-discharged 300 mesh 1.2/1.3 gold grid covered in a holey carbon film (Quantifoil). Excess sample was removed with a blotting time of 4 s and a blotting force of 1 at 4 °C prior to plunge freezing into liquid ethane using a Vitrobot Mark IV (Thermo Fisher).

GPR68-G_s pH 6 complex data collection

A total of 6,812 movies were recorded with a K3 detector (Gatan) on a Titan Krios (Thermo Fisher) microscope operated at 300 keV with a BioQuantum post-column energy filter set to a zero-loss energy selection slit width set of 20 eV. Movies were recorded using dose-fractionated illumination at a nominal magnification of 105,000x (physical pixel size of 0.83 Å/pixel) and a defocus range of -1 to -2.1 μm for a total dose of 49 $e^-/Å^2$. Exposure areas were acquired with image shift collection using SerialEM 3.8.[88]

GPR68-G_s pH 6 complex data processing

Movies of the GPR68-G_s pH 6 complex were imported into cryoSPARC v3.1 for motion-correction, dose-fractionation, and CTF estimation.[87] Micrographs with estimated CTF fit resolution > 5 Å were removed before further processing. Templates for particle picking were generated from the same complex reconstructed from a previous 200 keV imaging session. Particle picking templates were low-pass filtered to 20 Å and used to pick 7,064,401 particles. After picking, particles were extracted in a 288 pixel box and Fourier cropped to 48 pixels before 3D classification with alignment using a 20 Å low-pass filtered reconstruction and "garbage collectors" with the Heterogeneous Refinement job type. Two rounds of Heterogeneous Refinement yielded 2,524,876 particles that were re-extracted in the same box size cropped to 144 pixels and classified in an Heterogeneous Refinement job. The resulting 804,228 particles were refined using the Non-Uniform Refinement job type. Particles were exported using csparc2star.py from the pyem script package, and a mask covering the 7TM domain of GPR68 was generated using the Segger tool in UCSF ChimeraX and the mask.py pyem script.[73,75] The particles and mask were imported into Relion v3.0 and classified in 3D without alignment. Particles comprising the highest resolution classes were reimported into cryoSPARC for Non-Uniform Refinement.[69] Finally, particles were exported into cisTEM for two local refinements using the Manual Refinement job type and low-pass filtering outside of masks.[70] In the first local refinement, the previous 7TM mask was used, and the second local refinement used a full-particle mask.

Model building and refinement

Model building and refinement began with the Alphafold2 predicted structures as the starting models, which were fitted into the experimental cryoEM maps using UCSF ChimeraX.[89] The model was iteratively refined with real space refinement in Phenix and manually in Coot and Isolde.[71,72,74] The cholesteryl hemisuccinate model and rotamer library were generated with the PRODRG server, docked using Coot, and refined in Phenix and Isolde.[90] Final map-model validations were carried out using Molprobity and EMRinger in Phenix.

Molecular Dynamics Methods

Coevolutionary analysis

The coevolutionary analysis at the basis for seeding alternative conformations is based on the entire class A family of GPCRs, as defined by the PFAM family with ID 00001.[91] Available on the InterPro database,[92] the family with this accession number contains 264689 already aligned sequences. Using the multiple sequence alignment (MSA) as a training set for fitting a Potts-like model we expressed a pseudo-maximum likelihood, commonly known as the GREMLIN method.[93] The Potts model contains pairwise and column-wise parameters, which correspond to couplings and column-wise importances, respectively. Once trained, the 20x20 matrices containing the mutational maps at each residue-residue pair were then processed using an L2-norm and average product corrections (APC). For comparable results, the resulting single scores for each residue-residue pair were standardized to a N(0,1) normal distribution. Subsequently, the top 500 coevolving residue pairs were identified and used for further analysis.

System preparation and equilibration

While the active state structure of GPR68 provides an excellent starting point for molecular simulations, the local environment of the receptor, such as membrane, solvent and neutralizing ions need to be added to simulate the system as close to the physiologically relevant environment as possible. All simulations were done using a modified GROMACS package,[76] the source-code of which is available in the associated Zenodo upload (https://zenodo.org/doi/10.5281/zenodo.13785406). The final system was prepared using CHARMM-GUI[77] and embedded in a POPC bilayer in a 90x90x100 $Å^3$ box solvated with the TIP3 water model[94] and 10 mM KCl for periodic boundary conditions (PBC). The forcefield used for the entire system was CHARMM36m,[95] chosen for the ability to accurately model lipid-protein interactions.

Since the goal of this MD study was to calculate the energetic effect of pH on the activation pathway, we first needed to include a way to dynamically sample the protonation states of all titratable residues. This was achieved using the recently developed constant pH algorithm,[96] where the dynamic protonation state of a residue is expressed using a lambda coordinate. Each potential along the lambda coordinates needs to be calibrated such that the appropriate pKa is achieved in solution. By adding this potential with the non-bonded forcefield terms surrounding the residue, the microenvironment around the residue can be taken into consideration to shift the local

propensity of protonation of each residue. All titratable residues were coupled to one lambda coordinate representing the protonation state of the residue, using the parameters calibrated in the pH-builder package.[97] The notable exception was histidine residues, which require 3 lambda coordinates to model the protonation states at the epsilon and delta positions. Initially, every titratable residue was coupled to lambda coordinates, but due to border instabilities the lambda coordinates of 3 constantly protonated residues were set to 1 in the pH 6 condition. By employing the replacement method for water molecules, implemented in the pH-builder package,[98] we could add an appropriate number of buffer molecules to counter any electrostatic changes in the titratable residues.

The equilibration consisted of 2 rounds, the first following the standard CHARMM-GUI[77] NVT equilibration protocol, and the second appended to equilibrate the lambda coordinates in both pH 6.0 and pH 7.0. Generally, for the production phase, simulations used a 2-femtosecond timestep, employing a Nose-Hoover thermostat[99] with separate coupling for the receptor, solvent, and the membrane. The pressure was controlled using the C-rescale barostat, with a compressibility of 4.5e-5 and a relaxation time of 5.0. All hydrogen-involved bonds were constrained using the Lincs algorithm,[100] and long-range electrostatic interactions were calculated with the particle mesh Ewald (PME) method using a 12 Å cutoff.[101] Lennard-Jones (LJ) interactions were treated with a neighbor-list cutoff and a switching function at 10 Å, and a van der Waals (rvdw) cutoff of 12 Å.

Collective variable generation and pulling simulations

The collective variable generation consisted of two steps, part of the coevolution-driven conformational exploration method as described in Mitrovic et al.[61] In the first step, only evolutionary information combined with the experimentally determined structure was used to seed conformational change in MD simulations by non-equilibrium pulling. In the second step, the conformational change was analyzed to yield a mathematical description capable of distinguishing the start- and end-states. According to our previously published methods, GPCRs generally require only one round of non-equilibrium pulling thanks to their relatively simple functional landscape containing only two metastable states and a single transition state, as opposed to other proteins featuring more complex transition pathways.[59,60]

For the first step, a contact map was calculated based on the GPR68 active state conformation, where a minimum distance <4 Å constituted a contact. Then, the coevolution scores from the coevolutionary analysis were compared to the physical contacts in 3D space. Specifically, false positives, or coevolving residue pairs not in contact, were assembled in three sets of linear combinations with equal weights, clustered according to their coevolution strength. The reason behind this was that different aspects of the conserved features of a protein family, such as protein-protein interactions, conformational states, or important folding contacts, are expected to be coupled to differing degrees to function, thus resulting in different coevolutionary coupling strengths. With the constant pH module turned off, we applied an umbrella potential at 0.5 nm summed value with a force constant ramping up in steps at 1000, 2000, 5000 and 10000 kJ/mol•Å^2. This procedure generated inactive-like conformations and transition pathways, which were equilibrated for 10 ns and used as training data for the second step.

In the second step, all coevolving residue-pair distances (consisting of both false and true positives) were extracted and used as features for a support-vector machine (SVM). The decision function aimed to distinguish the active-like conformations from inactive-like conformations as consisting of the equilibrium simulations of the start- and end-states. For the SVM, a linear kernel was used such that the projection along the decision function could be made using the weights vector. This linear projection was reduced to only the 20 highest weighted distances to reduce the influence of degeneracy when applying a bias along the coordinate. All code used in these procedures are available in the associated MD Zenodo repository (https://zenodo.org/doi/10.5281/zenodo.13785406).

Accelerated Weight Histogram simulations

To estimate the free energy landscape along the final collective variable of activation, the accelerated weight histogram (AWH) method was employed in conjunction with the previously described constant pH algorithm. This approach required a non-straightforward merging of the constant pH code with modern versions of the AWH code that support transformation pull coordinates, necessary for adaptive biasing along the collective variable of activation.

Contrary to Metadynamics-like approaches, AWH applies a bias in the extended ensemble of a reference coordinate, which is coupled to the real system via a harmonic potential to, in this case, the collective variable of activation. The force constant of this harmonic potential determines the strength of the coupling, and the resolution of the free energy landscape. For the simulations, a force constant of 120,000 kJ/mol•Å^2 was used. The AWH method consists of two phases: first, an initial phase in which the bias update size gradually decreases with each covering, controlled by a growth factor of 3.0 (default). After a specific criterion—comparing the histogram of weights and the actual number of samples—is satisfied, the system switches to the Wang-Landau algorithm to reach convergence. The convergence was determined as the point when the change of the free energy estimation changes less than 1 kJ/mol per 100 ps (Figure S6).

Since the degree of inactivation under physiological conditions could not be known a priori, the bounds of the sampling interval were extended beyond what was physically credible, for instance conformations in which the TM4 helix exhibited exotic breaking. The final.mdp files for the single-walker replicates have been uploaded to the Zenodo repository (https://zenodo.org/doi/10.5281/zenodo.13785406).

QUANTIFICATION AND STATISTICAL ANALYSIS

Details of statistical analyses are in the text, figure legends, and or methods description for corresponding experiments.

Supplemental figures

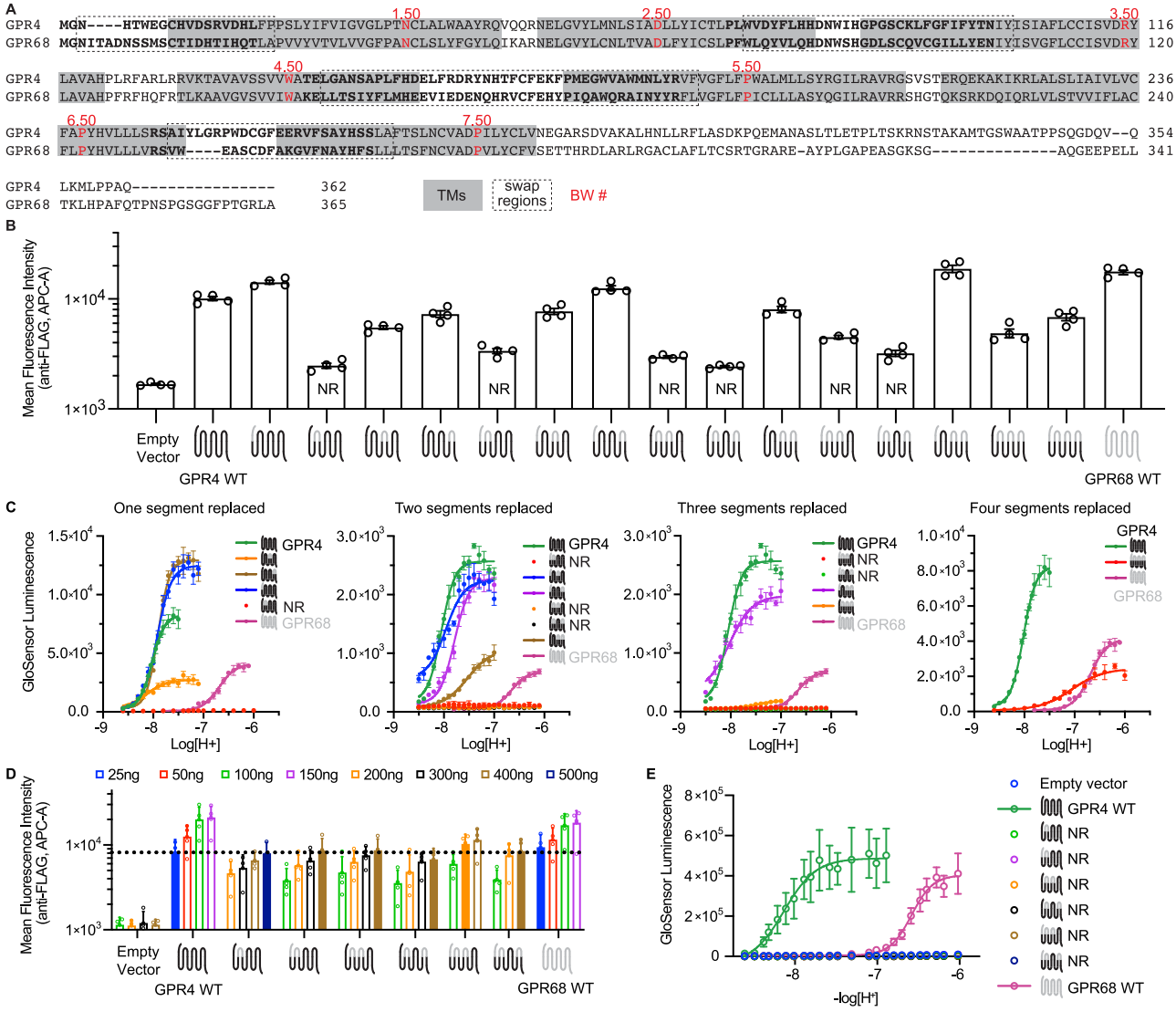


Figure S1. Chimeric receptor strategy to discern proton sensing, related to Figure 1

(A) Alignment of GPR4 and GPR68 generated using Clustal Omega. Gray indicates transmembrane helix regions. Dotted regions represent sequence segments that were grafted from GPR68 onto GPR4 to generate chimeric proton sensors.

(B) Expression level of chimeric receptors used in cAMP GloSensor assays. Mean fluorescence intensity ± SD (n = 4 biologic replicates) of anti-FLAG-APC staining is plotted for each receptor. Flow cytometry was performed with identical gating across all cell lines tested. "NR" indicates constructs for which no pH response was observed in signaling assays.

(C) GloSensor cAMP accumulation assay showing proton activation for GPR4-GPR68 chimeras. Data are representative technical replicates from three independent biological replicates ± SD. Fits are shown for chimeras that retain proton sensitivity. Chimeras that do not demonstrate proton sensitivity are indicated by "NR" for no response, and no fit line is shown. A full table of pharmacologic parameters is available in Table S1.

(D) DNA titrations of pH-insensitive chimeric receptors and WT GPR4 and GPR68 vs. expression level. Mean fluorescence intensity ± SD (n = 3–4 biologic replicates) of anti-FLAG-APC staining is plotted for each receptor. Flow cytometry was performed with identical gating across all cell lines tested. Filled bars closest to the dotted horizontal line indicate the amount of DNA used in signaling assays in (E) for matched receptor expression.

(E) GloSensor cAMP accumulation assay showing proton activation for expression matched constructs from (D). Data from three independent biological replicates ± SD. Fits are shown for WT receptors. pH-insensitive chimeras show similar results to Figure S1C.

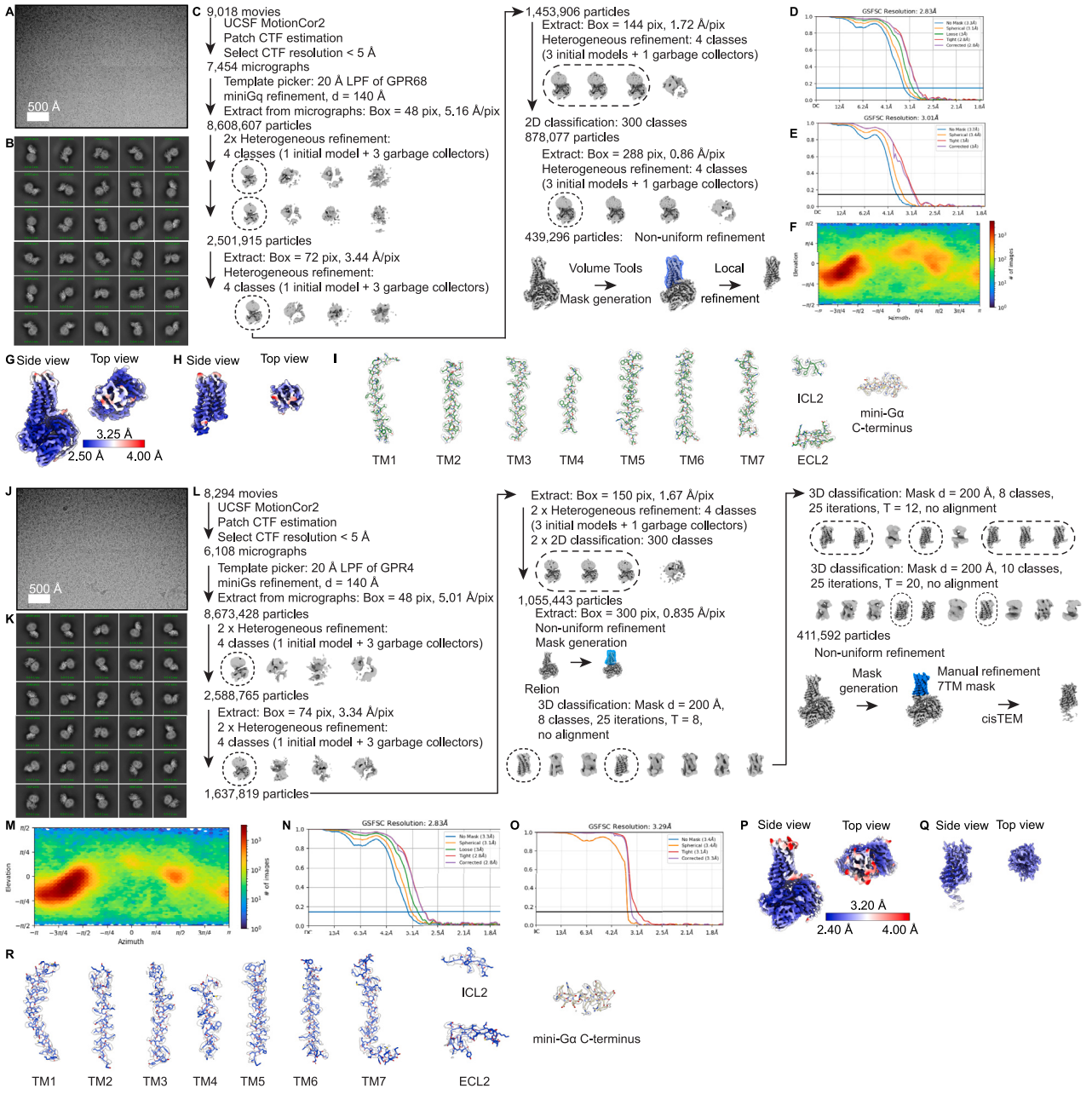


Figure S2. Cryogenic-electron microscopy processing of GPR4 miniG$_s$ and GPR65 miniG$_s$ pH 6, related to Figure 2

(A) A representative motion-corrected cryogenic-electron microscopy (cryo-EM) micrograph obtained from a Titan Krios microscope. Scale bar, 500 Å.

(B) A subset of highly populated, reference-free 2D-class averages.

(C) Schematic showing the cryo-EM data processing workflow. Initial processing was performed using UCSF MotionCor2 and cryoSPARC. Particles were selected using iterative heterogeneous refinement jobs followed by 2D classification. Finally, particles were processed using the local refinement job type with a 7TM mask. Dashed boxes indicated selected classes.

(D) Gold-standard Fourier shell correlation (GSFSC) curve for final full-particle map computed in cryoSPARC.

(E) GSFSC curve for final 7TM region map computed in cryoSPARC.

(F) Euler angle distribution of final full-particle map computed in cryoSPARC.

(G) Side view and top view of local resolution for the final full-particle map of GPR4-G$_s$ pH 6 complex computed with local resolution in cryoSPARC.

(H) Side view and top view of local resolution for the focused 7TM map of GPR4-G$_s$ pH 6 complex computed with local resolution in cryoSPARC.

(I) Expanded view of each helix and loop with the electron density overlaid.

(J) A representative motion-corrected cryogenic-electron microscopy (cryo-EM) micrograph obtained from a Titan Krios microscope. Scale bar, 500 Å.

(K) A subset of highly populated, reference-free 2D-class averages.

(legend continued on next page)

(L) Schematic showing the cryo-EM data processing workflow. Initial processing was performed using UCSF MotionCor2 and cryoSPARC. Particles were transferred using the pyem script package to RELION for alignment-free 3D classification. Finally, particles were processed in cisTEM using the manual refinement job type with a 7TM mask followed by a full-particle mask. Dashed boxes indicated selected classes.
(M) Euler angle distribution of final full-particle map computed in cryoSPARC.
(N) GSFSC curve for final full-particle map computed in cryoSPARC.
(O) GSFSC curve for final 7TM region map computed in cryoSPARC.
(P) Side view and top view of local resolution for the final full-particle map of GPR65-G_s pH 6 complex computed with local resolution in cryoSPARC.
(Q) Side view and top view of local resolution for the focused 7TM map of GPR65-G_s pH 6 complex computed with local resolution in cryoSPARC.
(R) Expanded view of each helix and loop with the electron density overlaid.

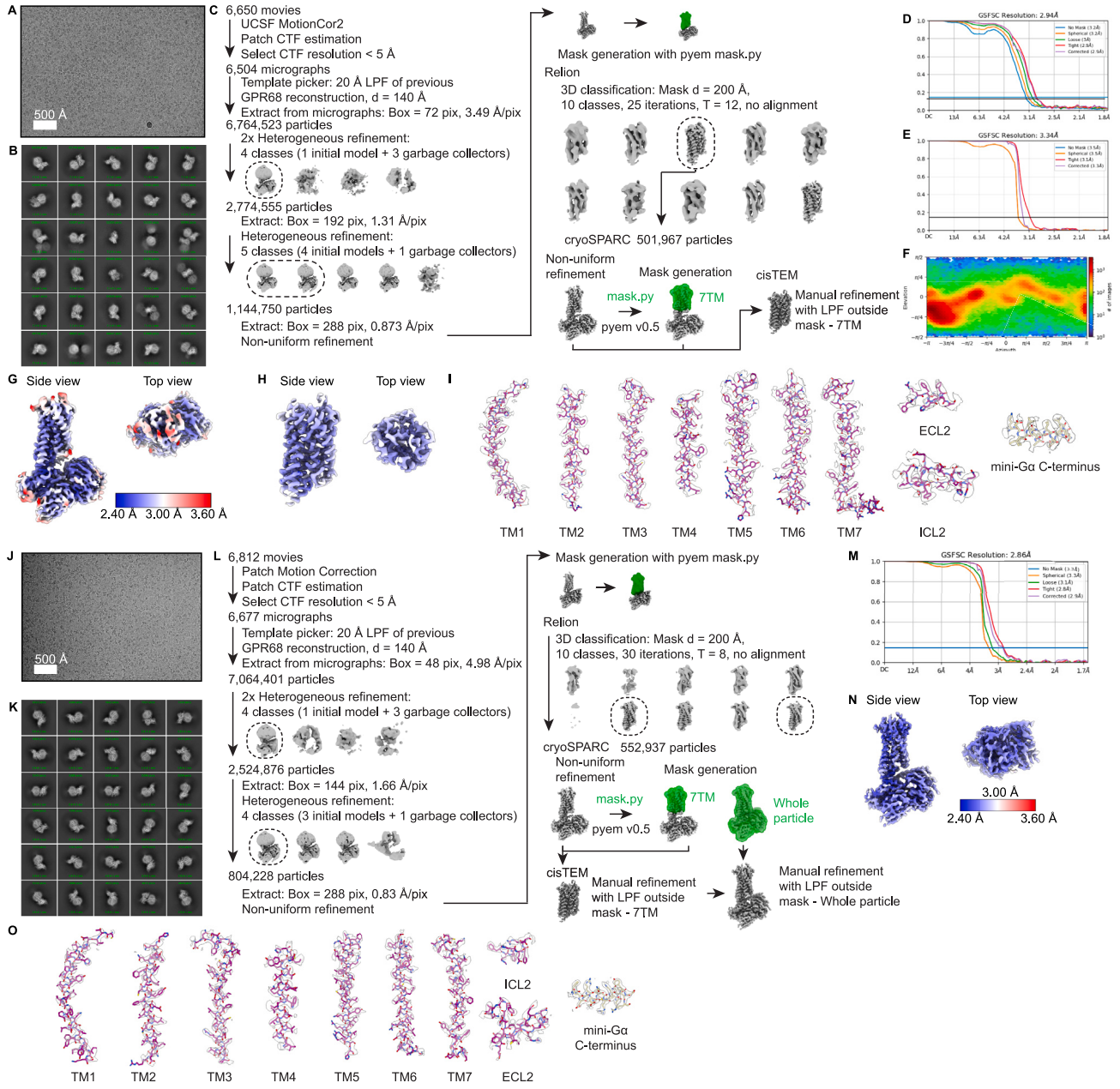


Figure S3. Cryogenic-electron microscopy processing of GPR68 miniG$_{s/q}$ pH 6 and GPR68 miniG$_s$ pH 6 with positive allosteric modulator MS48107, related to Figure 2

(A) A representative motion-corrected cryogenic-electron microscopy (cryo-EM) micrograph obtained from a Titan Krios microscope. Scale bar, 500 Å.

(B) A subset of highly populated, reference-free 2D-class averages.

(C) Schematic showing the cryo-EM data processing workflow. Initial processing was performed using UCSF MotionCor2 and cryoSPARC. Particles were transferred using the pyem script package to RELION for alignment-free 3D classification. Finally, particles were processed in cisTEM using the manual refinement job type with a 7TM mask. Dashed boxes indicated selected classes.

(D) GSFSC curve for final full-particle map computed in cryoSPARC.

(E) GSFSC curve for final masked 7TM region map computed in cryoSPARC.

(F) Euler angle distribution of final full-particle map computed in cryoSPARC.

(G) Side view and top view of local resolution for the final full-particle map of GPR68-G$_q$ pH 6 complex computed with local resolution in cryoSPARC.

(H) Side view and top view of local resolution for the focused 7TM map of GPR68-G$_q$ pH 6 complex computed with local resolution in cryoSPARC. Expanded view of each helix and loop with the electron density overlaid.

(J) A representative motion-corrected cryo-EM micrograph obtained from a Titan Krios microscope. Scale bar, 500 Å.

(K) A subset of highly populated, reference-free 2D-class averages.

(legend continued on next page)

(L) Schematic showing the cryo-EM data processing workflow. Initial processing was performed using UCSF MotionCor2 and cryoSPARC. Particles were transferred using the pyem script package to RELION for alignment-free 3D classification. Finally, particles were processed in cisTEM using the manual refinement job type with a 7TM mask followed by a full-particle mask. Dashed boxes indicated selected classes.
(M) GSFSC curve for final full-particle map computed in cryoSPARC.
(N) Side view and top view of local resolution for the final full-particle map of GPR68-G_s pH 6 complex computed with local resolution in cryoSPARC.
(O) Expanded view of each helix and loop with the electron density overlaid.

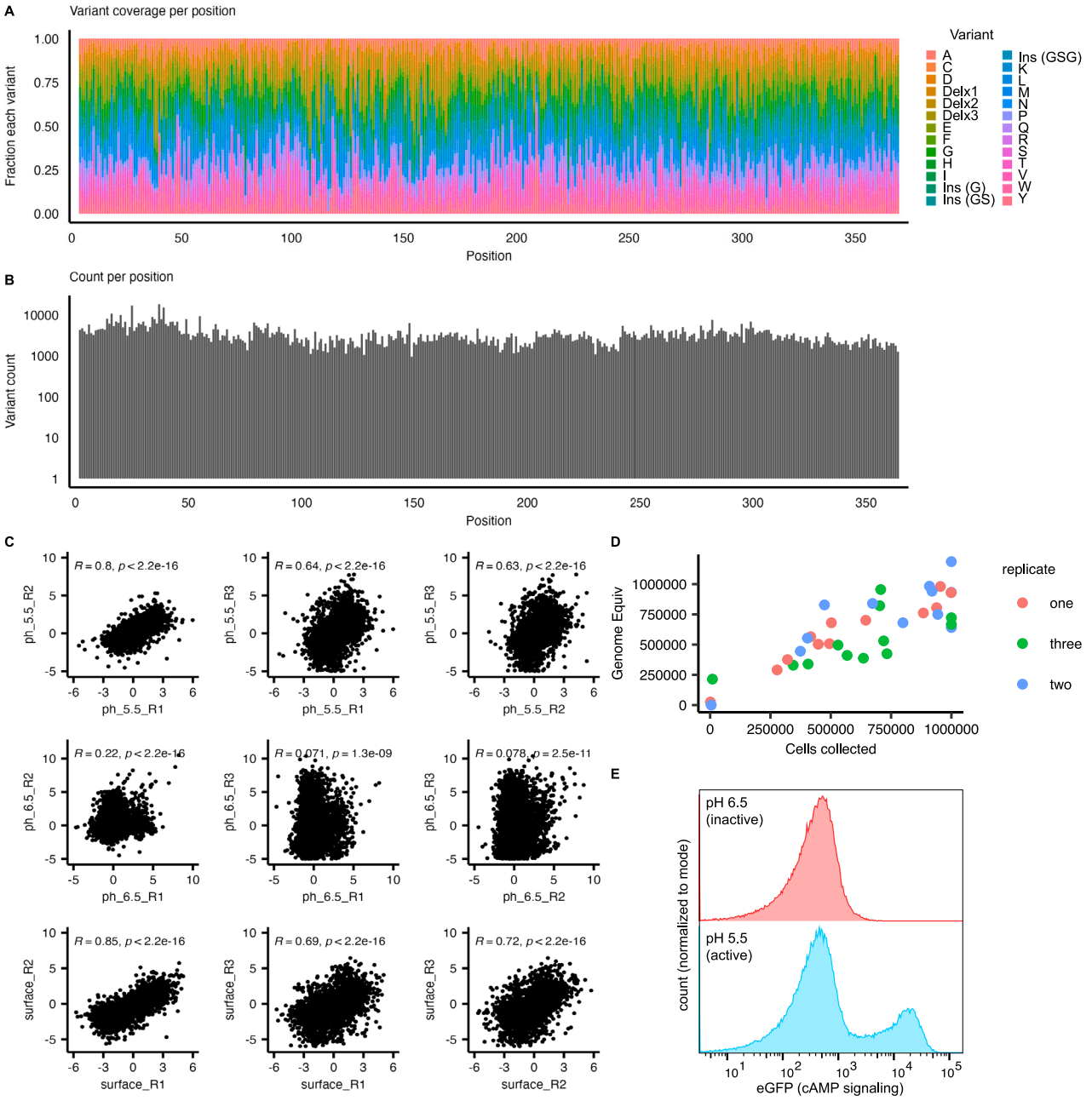


Figure S4. Baseline GPR68 DMS library, replicate correlations, cells collected vs. recovered genome equivalents, example FACS traces, related to Figures 4 and 5

(A) Normalized counts of each variant per position in the GPR68 deep mutational library demonstrate a relatively even overage of each possible mutation at each position.

(B) Total counts per position in the GPR68 deep mutational library demonstrate relatively even total coverage for each position.

(C) Replicate correlations for each DMS screening condition. Spearman coefficient (*R*) and *p* values are indicated for each plot.

(D) Scatter of number of cells collected for each DMS sample compared with estimated number of genome equivalents DNA recovered after extraction, stratified by replicate.

(E) Representative plot of GPR68 mutational library eGFP distribution from cAMP signaling assay at both pH 5.5 and pH 6.5.

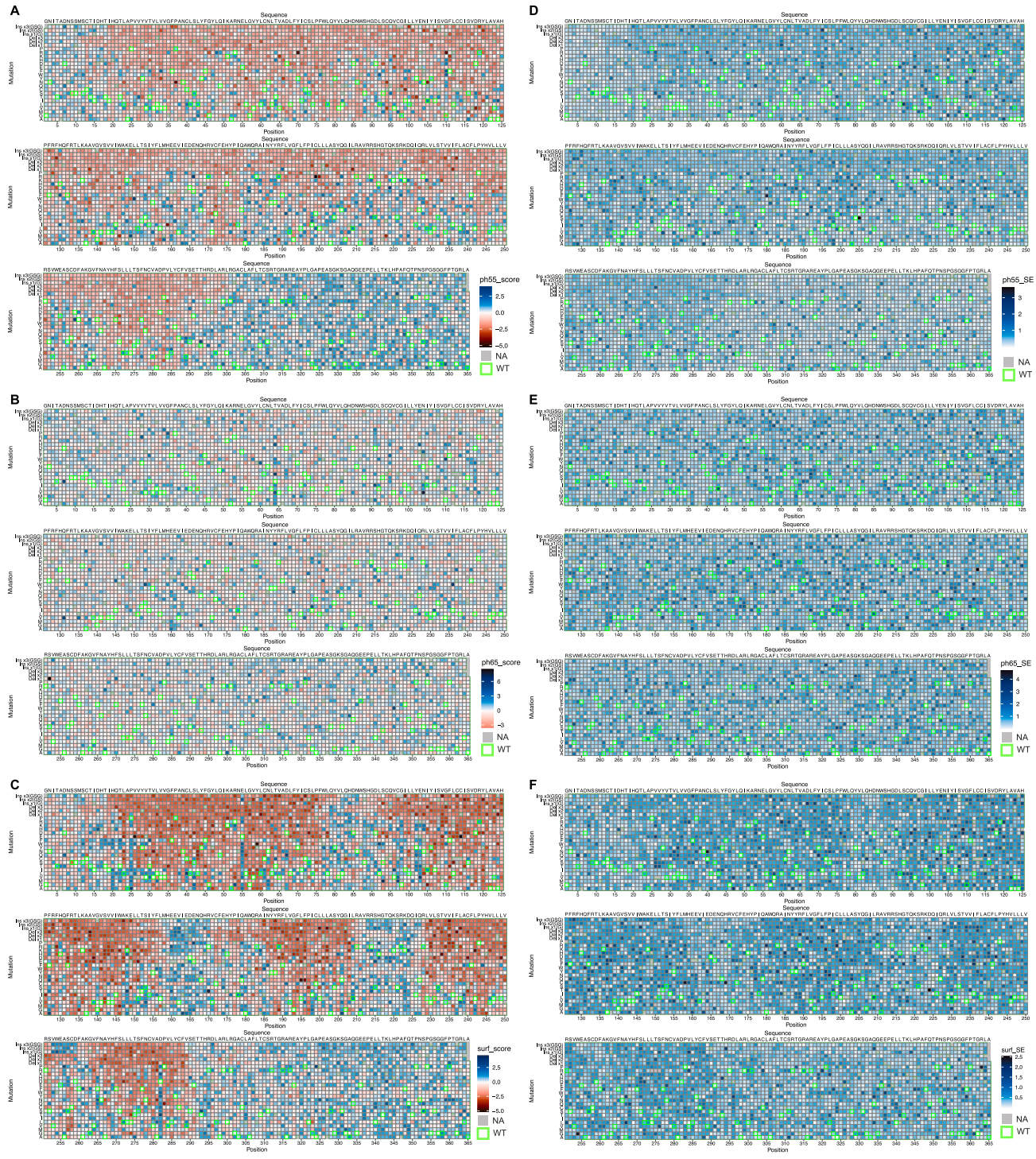


Figure S5. Full GPR68 DMS fitness score and SE heatmaps, related to Figures 4 and 5

(A–C) Full heatmaps of DMS fitness values for (A) GPR68 cAMP signaling at pH 5.5, (B) GPR68 cAMP signaling at pH 6.5, and (C) surface expression. (D–F) Standard error heatmaps of DMS fitness scores for (D) GPR68 cAMP signaling at pH 5.5, (E) GPR68 cAMP signaling at pH 6.5, and (F) surface expression. For all heatmaps, the WT sequence is shown above each section, mutations are indicated on the left axis, and the amino acid position is indicated by the numbers below. Positions and mutations with no data are shown as gray, and the WT amino acid at each position has a green border. Fitness scores are relative to the

(legend continued on next page)

mean synonymous mutant score (set to 0) and were calculated using Enrich2.[45] Blue indicates increased cAMP signaling or surface expression relative to synonymous mutants, and red indicates decreased cAMP signaling or surface expression relative to synonymous mutants. Data are fitness values from three biologically independent deep mutational scans. Standard error (SE) heatmaps have a similar layout, an SE of zero is white, and a higher SE is indicated by increasingly dark shades of blue.

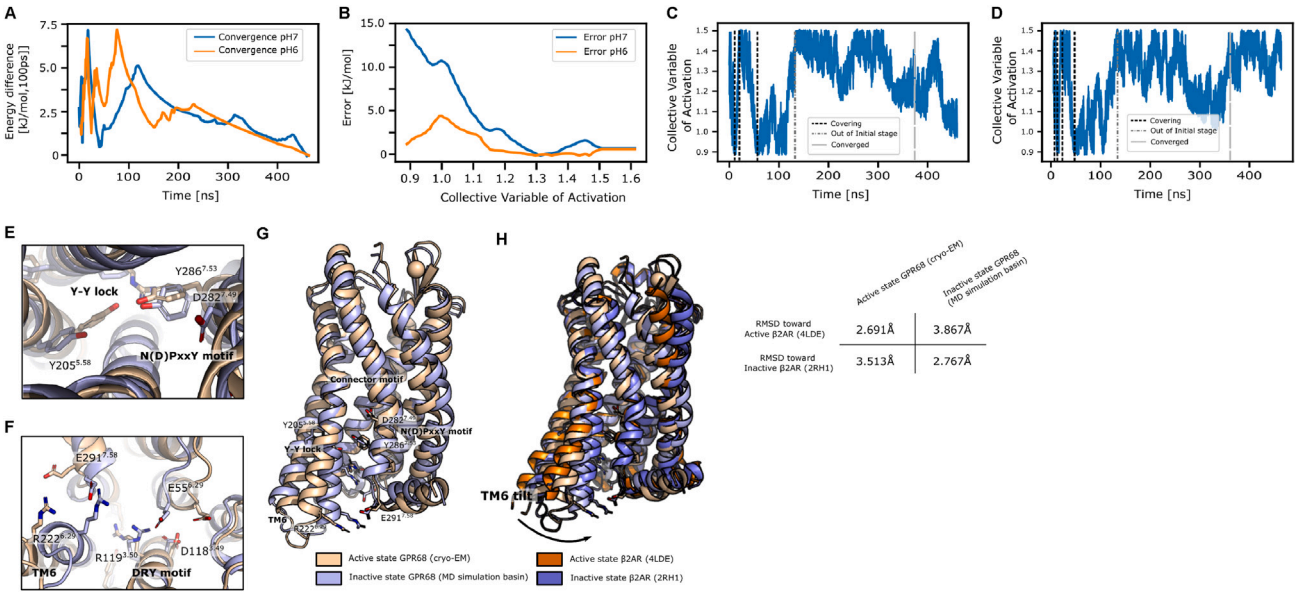


Figure S6. MD simulation convergence, coverage, and structural validation, related to Figure 6

(A) Convergence of the accelerated weight histogram (AWH) simulations. For both conditions, the y axis indicates the maximum point-wise difference in the free energy estimation over 100 ps, with the time of the simulations on the x axis.

(B) Error plots. The standard error of the free energy estimations corresponding to Figure 6B, calculated on the basis of 4 replicates. Generally, the error is significantly lower in low-energy regions (CV values between 1.1 and 1.4) as compared with high-energy regions (CV values between 0.9 and 1.1).

(C and D) The time evolution of the collective variable for pH 7 (C) and pH 6 (D) conditions. The dashed black lines indicate when a covering of the entire range has occurred, which reduces the update size by the growth factor. Once a sufficient number of samples have been accrued, the algorithm exits the initial stage (dashed gray line), and the update size is adaptively reduced based on prior sampling. Upon reaching a free energy difference below 1.0 kJ/mol every 100 ps, the simulation is deemed converged. Note that high-energy areas are rarely revisited in the final stage, resulting in a higher error in those regions (B).

(E–G) Comparison of the inactive GPR68 state derived from MD simulations and the GPR68 active-state structure as determined by cryo-EM. Structural alignment enabled close analysis of the downstream microswitches common for class A GPCRs, such as the Y-Y lock and the N(D)PxxY motifs (E and G) and the DRY motif along with the canonical TM6 tilt (F and G). All microswitches displayed their previously described behavior in accordance with the allosteric pathways of activation. The analysis was limited to the intracellular microswitches due to the non-canonical activation mechanism of GPR68.

(H) Comparison between the β2AR inactive (PDB: 2RH1) and β2AR active (PDB: 4LDE) conformations and the conformations found in this study. While the active and inactive states of this study proved less extreme than the ones found in β2AR, the lowest RMSD was consistently found with the corresponding state in β2AR, showing that the MD inactive basin is comparable to the inactive state in β2AR.

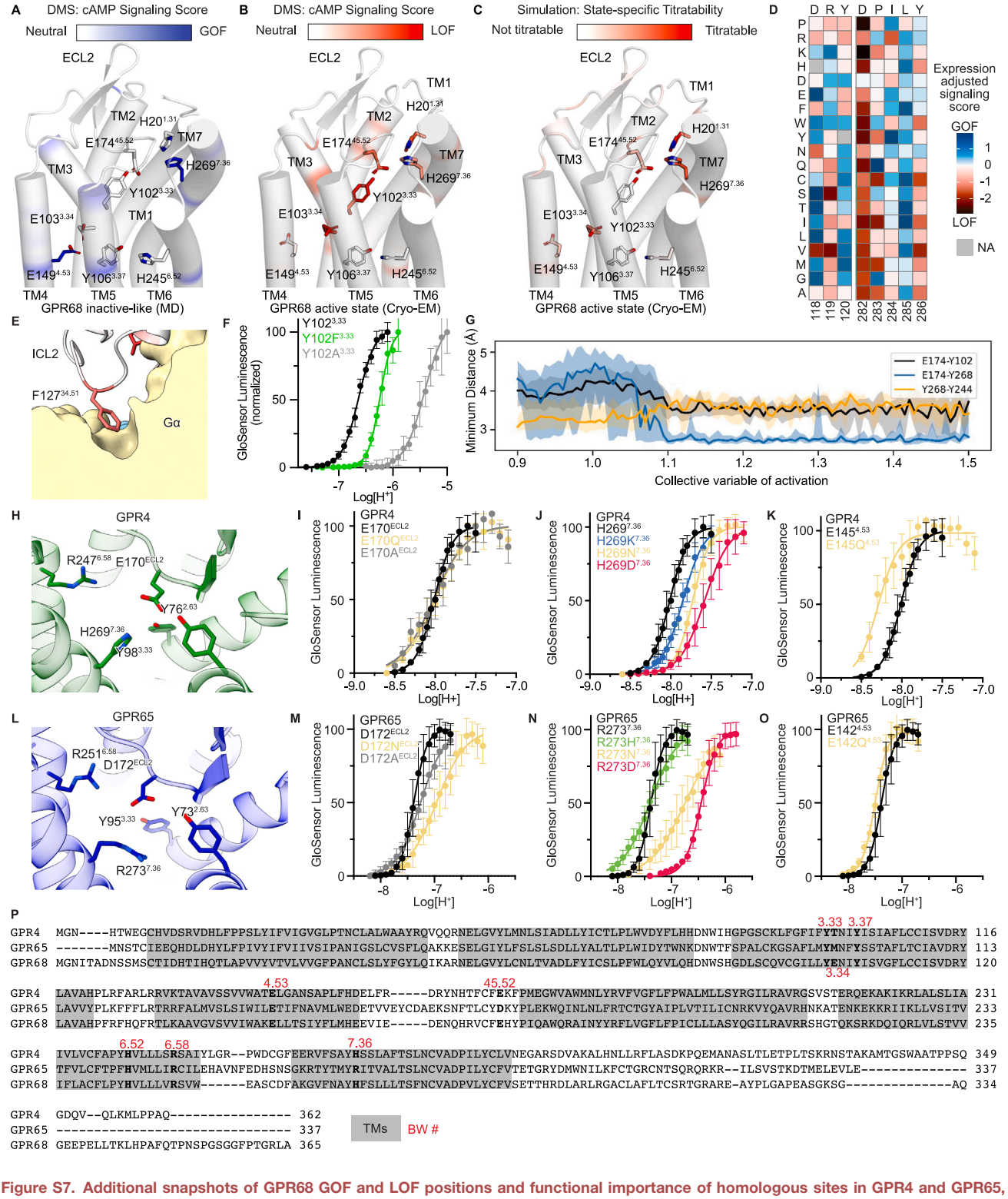


Figure S7. Additional snapshots of GPR68 GOF and LOF positions and functional importance of homologous sites in GPR4 and GPR65, related to Figure 7

(A) Mapping of GOF DMS scores onto the extracellular potion of GPR68 inactive state from MD simulations.

(B) Mapping of LOF DMS scores onto the extracellular portion of GPR68 active-state cryo-EM model.

(C) Mapping of state-specific protonation score from constant-pH MD simulations onto the extracellular portion of GPR68 active-state cryo-EM model.

(legend continued on next page)

(D) Heatmap of surface expression-adjusted pH 5.5 cAMP signaling scores for the DRY and DPxxY motifs GPR68. Mutations are indicated on the left axis, and the amino acid position is indicated by the numbers below each. Positions and mutations with no data are shown as gray. Blue indicates higher activity relative to synonymous mutant mean score, and red indicates lower activity relative to synonymous mutant mean score.
(E) Zoom view of GPR68 F127 within ICL2, which is critical for G protein activation. GPR68 residues are colored by LOF score as in (A).
(F) cAMP accumulation GloSensor assays testing impact of phenylalanine and alanine mutations to Y102$^{3.33}$ on proton potency in GPR68. Data shown are from three or four independent biological replicates ± SD.
(G) Distance measurements between several pairs of residues on the extracellular side of GPR68 monitored across the collective variable of activation in MD simulations. Across the range, CV 1.2–1.4, which we consider the transition from inactive to active when changing pH from 7.0 to 6.0, residue distances are relatively stable.
(H) Zoom view of residues linking ECL2 of GPR4.
(I–K) GloSensor data for structurally conserved ECL2 residue mutants in GPR4.
(L) Zoom view of residues linking ECL2 of GPR65.
(M–O) GloSensor data for structurally conserved ECL2 residue mutants in GPR65. Curves in (B)–(D) and (F)–(H) are from three independent biological replicates ± SD.
(P) Sequence alignment of GPR4, GPR65, and GPR68 with transmembrane helices indicates and BW labels of residues tested in (I)–(K) and (M)–(O), and those that correspond to residues in the GPR68 activation network in Figure 6.