Title
Shooting in the dark: Strategies for discovering magic bullets and magic shotguns for orphan GPCRs

Permalink
https://escholarship.org/uc/item/0dd0c4z2

Author
Karpiak, Joel David

Publication Date
2015

Supplemental Material
https://escholarship.org/uc/item/0dd0c4z2#supplemental

Peer reviewed|Thesis/dissertation

Shooting in the dark: Strategies for discovering magic bullets and
magic shotguns for orphan GPCRs

by

Joel David Karpiak

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

To my family, by blood or over wine

Acknowledgments

The road to this thesis was paved, lit, mapped, and course-corrected through the contributions of many people. Foremost, I would like to thank my advisor, Brian Shoichet. His optimism, which others may term "spin," kept me focused and invested when an experimental result seemed destined to doom my project; before I had the chance to throw it in, he would take the towel out of my hand and convince me to fight Round 2. Brian taught me to ask the right questions, in which any answer would be illuminating for the field, and to not be afraid of my ambition to tackle the big picture. Finally, he advised me to do "as he said" and not "as he did," which has served me well in numerous instances so far, both professionally and personally. The other members of my thesis committee, Matt Jacobson and Andrej Sali, were always available to provide advice to keep my projects on the straight, yet interesting, path. It was a curious and reassuring feeling to go through grad school knowing that my future was in the best hands possible.

There are many members of the Shoichet lab that were instrumental to the completion of this thesis, as well. Oliv Eidam, my fantastic rotation mentor, placed me on this road by not only introducing me to the beautifully frustrating experience known as DOCK, but also by teaching me how to hold a pipette and move tiny amounts of liquids between different tiny containers. John Irwin, Pascal Wassam, Teague Sterling, and Therese Demers kept me from sabotaging myself

actually, is all around. Lekshmi Santhosh, Samarth Keshava, and Jaclyn Opritza, my classmates and life-long friends from Yale, eventually followed me to San Francisco, and every wine bar in the city, along with me, thanks you for this. Thanks to prophecy, I know I will still see you all when I am an old man. Somehow we picked up Tara Marie Vega along the way, and I sincerely doubt I will ever spend as much time connecting with another person on a daily basis. I hope we fall in love. William May's love and belief got me over the finish line. I owe Stacey Cifelli for her constant snark and Spoon Day celebrations, helping me keep my sense of humor intact. To my Alaska-bound group, I would not be the same person I am without you. In order that I met you: Kathleen Maguire, Carolyn Lake, Natalie Blagriff, and Katie Lutz, and Tara Bendig, from kindergarten to AP Biology to my doctorate, you pushed me to conceive, to dream, to want, and ultimately, to achieve more than I ever thought possible for myself. You are my soulmates.

Finally, I thank my brother, sister-in-law, Mom Mom June (who taught me to be good and to be kind), and my parents - both sets of them. I know of your sacrifices and hopes for me that went way beyond anything for yourselves. You gave me everything, and I can only ever continuously attempt to repay that.

This is for you.

The text of Chapter 1 is a reprint of the material as it appears in:

Huang, X.P.*, Karpiak, J.*, Kroeze, W.* et al. Discovering Allosteric Modulators for Orphan GPCRs.

It appears here with permission from the authors. The supplementary material from this paper has been included as Appendix A.1.

The text of Chapter 2 is a variation of the material as it appears in:

Karpiak, J.*, Weiss, D.R.* et al. Structure-based methods for the rational design of polypharmacology in G protein-coupled receptors.

It appears here with permission from the authors. The supplementary material

from this paper has been included as Appendix A.2.

Abstract

Shooting in the dark: Strategies for discovering magic bullets and magic shotguns for orphan GPCRs

Joel David Karpiak

It is now widely accepted that drugs achieve efficacy via interactions with multiple targets simultaneously, a serendipitous biological symphony that is typically only deconvoluted retrospectively. While pan-profiling of these molecules has discovered varying activity against different proteins, efforts to recapitulate therapeutic effects through focused, specific interactions with singular targets have largely failed. This is especially true for G protein-coupled receptors (GPCRs), whose conserved topologies have introduced widespread degeneracy among their ligands. Yet, 38% of GPCRs are orphans, with no elucidated physiological function or no illuminating modulatory chemical matter.

In the first Chapter, we use a robust yeast-screening platform to show that even in a small library of drugs, the NIH Clinical Collection, orphan GPCRs are potent targets for a variety of drugs. We focus on the benzodiazepine lorazepam and show that is also a strong positive allosteric modulator (PAM) of the pH-sensing orphan, GPR68. We generated over 3,000 receptor homology models for virtual screening campaigns that, among other hits, identified ogerin, a PAM that selectively and potently potentiated GPR68-G_s signaling ($\alpha\beta= 27.5$). Ogerin was prioritized to interrogate the *in vivo* function of GPR68, revealing that receptor

activation suppressed recall in fear conditioning in wild-type, but not in GPR68 knockout, mice. Application of the same approach to a second orphan receptor, GPR65, led to the discovery of specific allosteric agonists and negative allosteric modulators.

In the second Chapter, we pick subsets of GPCRs to demonstrate that structure-based virtual screening can be used to directly identify molecules with a desired receptor binding profile. Although only a small number of hits were selective over the anti-target, a result observed in two parallel studies between three aminergic receptors and 2 opioid receptor subtypes, this performance is still better than that of ligand-based methods. Inexhaustive and incorrect treatment of receptor flexibility was subsequently deemed to be the culprit, a potentially receptor-specific nuisance that can be leveraged. At its core, this thesis describes test cases and methods that would ultimately encompass the machinery for a pipeline to generate molecules with a complete, prescribed GPCR polypharmacology.

Table of Contents

List of Tables

List of Figures

Introduction

Five years ago, I entered Brian Shoichet's lab with grandiose visions of myself as a jack of all trades, master of even more. I predicted that I would be a self-sufficient, productive GPCR pipeline: ideas would come in, run through my computer simulations, and out would come a steady stream of crystal structures, IC_{50}s, and papers. I had a keyboard in one hand, a pipette in the other, and, mysteriously, somehow a mug of tea in yet another. This was encouraged by Brian's sabbatical, where he was off to the mythical land of Brian Roth's Psychoactive Drug Screening Program at UNC-Chapel Hill to slay grant proposals, learn magical beta-arrestin assays, and, like Prometheus, bring back the secrets of GPCRs to the ignorant. I even had an entire proposal about subtype specificity holstered until I finished my requisite teaching duties.

Then, Brian returned - empty-handed. It seemed as though trying to set up a mini-PDSP at UCSF was a Sisyphean, and unnecessary, task. However, he had heard of a project idea that seemed entirely plausible to my ambitious sensibilities, especially considering that I now wasn't even running all of the assays myself. This earworm dug and dug and eventually cocooned itself in my brain, and, nourished by my previous interest in selectivity, this thesis emerged from its chrysalis. I was intent on illuminating even the darkest recesses of GPCR pharmacology, but I didn't initially set out on the hunt with the right weapons. Bryan Roth had previously called for designing selectively non-selective drugs ("magic shotguns") that hit a specific pharmacologic profile for efficacy, rather

1

than a molecule that acts on a single-target; a "magic bullet." What I didn't know until much later is that, regardless of the ammunition, one has to know at least where the target is. And to be aware that it is moving.

GPCRs, the largest family of proteins encoded in the human genome, transduce signals for the most diverse endogenous ligands of any single family of receptors. Currently almost a third of all FDA-approved drugs act through GPCRs, but these targets represent only a small percentage of the druggable receptor-ome. As many as 38% of GPCRs are "orphans," with no elucidated physiological function or no illuminating modulatory chemical matter. These orphans, the shadowy and unexplained receptors of interest by the PDSP, could conceivably be the mechanisms for unexplained off-target effects of drugs, or they could even unknowingly contribute to the perceived therapeutic profile. While ligand-based methods had previously been used to discover new side effects for old drugs, these techniques rely on already known sets of molecules to compare. As orphan receptors have no such annotations, in order to gain any traction, we first had to go ligand hunting.

This, however, is not straightforward, and reflects why exactly these receptors are still orphans. The types of molecules to which GPCRs might respond, from photons to proteins, are almost unbounded. How they respond to those signals is also unpredictable - it can be via canonical G protein pathways or non-canonical (arrestin, GRK, etc.) pathways - rendering the design and execution of functional assays problematic. In these assays, there is also a lack of positive control, so it is difficult to determine what is actually a true ligand hit. However,

a yeast-based high-throughput screen of the NIH Clinical Collection of drugs seemed to silence all of these concerns. For one, the readout of the assay was now coupled to the yeast growth pathway; increased colony size in a certain receptor strain indicated that a compound may be an agonist to that receptor. Establishing a matrix by pan-profiling the same library over a group of different receptors weeded out both frequent hitters and toxic compounds, making specific receptor-putative ligand pairs obvious for confirmation in traditional cell-based assays.

With all of this information in hand, we chose to focus on the most efficacious pairing resulting from the screen - the benzodiazepine lorazepam and the orphan pH-sensing receptor, GPR68. Since lorazepam is already a marketed drug for anxiety, acting primarily as a positive allosteric modulator (PAM) of the $GABA_A$ receptor, we would never be able to use it to illuminate GPR68's function *in vivo*. Instead, we decided to use homology modeling and structure-based design as the flashlight, guiding the way in the dark to discover a specific GPR68 probe. With so little information and sequence identity, could we even find a magic bullet using these models? And, if so, could we use them to find ligands and model even more receptors - essentially, could the models also be used to fire magic shotguns?

Answering the first question became Chapter 1 of this thesis. Preliminary evidence from previous studies modeling the dopamine D_3 and chemokine CXCR4 receptors suggested that this docking campaign would be a gambit. While the final D_3 homology model was structurally accurate and performed equally as well as the crystal structure in prospective virtual screens, this was not the case for CXCR4. There, the model not only proposed the incorrect binding site, it also only

returned 1 hit of questionable novelty in a virtual screen, whereas the CXCR4 crystal structure performed almost as well as the D_3 one. This was attributed to a low sequence identity to the modeled template (25%, versus the 42% D_3 was afforded), a lack of reliable mutational data, and a small amount of known ligands - all characteristics that faced GPR68.

The second question had no answer; no one had ever tried to rationally design magic shotguns before using structure-based methods - there was too much SAR, and random luck, to incorporate. So, with the recent publication of the histamine H_1 crystal structure, we decided to attempt to use virtual screens to purposefully discover molecules that were active at both the serotonin 5-HT_{2A} and the dopamine D_2 receptors, yet completely avoided binding to histamine H_1. This is the basis for Chapter 2.

After showing that it was possible to fashion bullets, shotguns, and sometimes nuclear bombs, for GPCRs, we attempted to answer some related questions that were also hiding in the dark - this is the Pandora's Box of Chapter 3. Could we incorporate other types of selectivity, such as functional bias, into this pipeline? Does the template choice pre-determine the model? Do covalent ligands lead to the same modeling outcomes? Can we find probes for every pH-sensing receptor? Can these models transcend their original purpose and be used to inform the endogenous function? Do these conclusions hold up when challenged by other receptors? Finally, can we combine GPCR-ome scale modeling and polypharmacology strategies to essentially create an on-demand GPCR ligand pipeline? These questions are currently being addressed by on-going projects.

Gloss to Chapter 1

How much information is necessary to discover ligands for orphan GPCRs? The following Chapter describes how we attempted to answer this question. As mentioned in the Introduction, we used the identified lorazepam hit from the yeast-based screen as an anchor for structural modeling of GPR68, and the subsequent virtual screen discovered 5 diverse positive allosteric modulators (PAMs) of the receptor's ability to respond to protons. The most efficacious hit was optimized by catalogue to a compound, nicknamed 'ogerin,' that selectively and potently potentiated GPR68-G_s signaling ($\alpha\beta = 27.5$). The receptor model was further vindicated via mutagenesis studies, as well as structure-guided optimization of ogerin to **C2**. These probes were used *in vivo* to discover GPR68's role in contextual fear, validating its high expression in the hippocampus and cerebellum. Finally, the same strategy was used to identify allosteric agonists and negative allosteric modulators of GPR65, a related pH-sensing receptor.

According to this Chapter, the minimum information for a magic bullet is not just knowing one true ligand, but also knowing molecules that are *not* ligands. While the pairing of lorazepam and GPR68 was obviously necessary for the docking, which facilitated model generation and optimization, the negative information from the inactive molecules in the NIHCC allowed model evaluation and selection. There were initially many strikes against this project working: 1) only 1 ligand was known, 2) GPR68 had a low sequence identity to the template, CXCR4 (~29%), 3) there was absolutely no knowledge of the location of the binding

site, and 4) no templates in the active state were available. Accordingly, the negative information was the light shining through the dark. This study also showed that even though a receptor model was created to reproduce the binding mode of a specific molecule, it can be not only generally applicable and accommodating to a wide range of diverse scaffolds, it can also be predictive and guide the structure-based optimization of these compounds. This concept supports the model's representative use in polypharmacological design studies; making magic shotguns out of magic bullets is a concept that we will return to in Chapters 2 & 3.

Chapter 1:

Discovering allosteric modulators for orphan GPCRs

Xi-Ping Huang,[1,2]* Joel Karpiak,[3]* Wesley K. Kroeze,[1]* Hu Zhu,[1] Xin Chen,[4,5] Sheryl S. Moy,[6] Kara A. Saddoris,[6] Viktoriya Nikolova,[6] Martilias S. Farrell,[7] Sheng Wang,[1] Thomas J. Mangano,[1,2] Deepak A. Deshpande,[9] Alice Jiang,[1,2] Raymond B. Penn,[9] Jian Jin,[4,5,10] Beverly H. Koller,[7] Terry Kenakin,[1] Brian K. Shoichet,[3,8†] and Bryan L. Roth[1,2,5†]

[1]Department of Pharmacology and [2]National Institute of Mental Health Psychoactive Drug Screening Program (NIMH PDSP), School of Medicine, [4]Center for Integrative Chemical Biology and Drug Discovery (CICBDD), [5]Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, [6]Department of Psychiatry and Carolina Institute for Developmental Disabilities (CIDD), [7]Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC

[3]Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, CA

[8]Faculty of Pharmacy & Ontario Institute for Cancer Research, University of Toronto, Toronto, Canada

[9]Center for Translational Medicine and Department of Medicine, Thomas Jefferson University, Philadelphia, PA

[10]Current address: Department of Structure and Chemistry Biology, Department of Oncological Sciences, Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY

*These authors contributed equally to this work

[†]Corresponding authors (BKS and BLR)

1.1 Abstract

At least 120 non-olfactory G protein-coupled receptors in the human genome are "orphans" for which endogenous ligands are unknown, and many more have no selective ligands, making elucidation of their biological function and clinical relevance difficult. Here, using the orphan receptor GPR68 (OGR1) as a model, we combine physical and structure-based screening to discover potent and selective GPR68 ligands. Initial GPR68 yeast-based screening identified the benzodiazepine drug lorazepam as a potentially useful, albeit non-selective, GPR68 positive allosteric modulator. We then generated over 3000 GPR68 homology models and refined these for the ability to recognize lorazepam in a putative allosteric site. Docking screens of 3.1 million molecules predicted new GPR68 positive allosteric modulators; many of these were confirmed in cell-based functional assays. Unlike lorazepam, they were specific for GPR68. One, which we call ogerin, potently potentiated proton-mediated GPR68-G_s signaling ($\alpha\beta$ = 27.5) and was prioritized as an *in vivo* tool compound. Because GPR68 is enriched in the hippocampus, we focused on learning- and memory-based behavioral assays, and found that peripheral administration of ogerin suppressed recall in fear conditioning in wild-type, but not in GPR68 knockout mice. Thus, a combination of physical and structure-based approaches facilitated the discovery of novel positive allosteric modulators at an orphan G protein-coupled receptor. Application of the same approach to a second orphan receptor, GPR65, led to the discovery of allosteric agonists and negative allosteric modulators. These results, together with further structure-based optimization of ogerin, suggest that a combination of physical and

structure-based screening may be a practical approach for discovering ligands for orphan G protein-coupled receptors.

1.2 Introduction

G protein-coupled receptors (GPCRs) -- the largest family of proteins encoded in the human genome -- transduce signals for the most diverse endogenous ligands of any single family of receptors. Correspondingly, GPCRs are the most productive drug targets, with over 26% of FDA-approved drugs acting primarily through GPCRs. Astonishingly, of the 356 non-olfactory GPCRs, as many as 38% are "orphans" whose physiological roles and endogenous ligands remain unknown. Given the central role of GPCRs in physiology and disease, and the high conservation of most orphan GPCRs among organisms from worms to humans, orphan GPCRs are likely to emerge as functionally and therapeutically important druggable targets. Indeed, for the few GPCRs deorphanized since 2003[1] (http://www.iuphar-db.org/latestPairings.jsp/), most have investigational new drugs in late stage clinical trials for diseases such as sleep disorders, obesity, and depression[2,3]. As with kinases[4], epigenetic proteins[5,6], proteases[7], and other families of druggable targets[8], ligands specific for orphan GPCRs will illuminate their biology and provide new areas for therapeutic intervention.

A key impediment to GPCR deorphanization is uncertainty about the proteins through which they signal, either via canonical G protein pathways or non-canonical (*e.g.*, arrestin-ergic) pathways, rendering the design and execution of functional assays problematic[9]. Although this problem confronts all genome-

wide function-assignment efforts, for many druggable targets the ligand and reaction space is illuminated by the protein fold family. For instance, efforts to annotate enzymes in the enolase[10], amidohydrolase[11], and short-chain dehydrogenase/reductase superfamilies[12] may draw upon a particular set of reactions, and often substrate types, upon which the family is known to act. Similarly, the range of substrates for lipid kinases, though large, is not unbounded, and whereas protein kinases have a much larger range of targets, these can be addressed by high-throughput proteomics and related approaches[13-15]. GPCRs, conversely, recognize ligands that span the range of biologically active molecules, from protons and photons, to small neurotransmitters like acetylcholine, dopamine, and ATP, to lipids like sphingosine and lysophosphatidic acid, to peptides and folded proteins such as chemokines and proteases[16]. Thus, not only are the molecules to which orphan GPCRs might respond almost unbounded, how they respond to those signals is also unpredictable. Hence, generic functional screens are difficult for orphan GPCRs — one neither knows what class of chemical matter to screen, nor how to screen for it, much less how to demonstrate functional and physiological relevance — thereby explaining the slow progress in determining the roles of orphan GPCRs in signaling and in organismal physiology[3].

The orphan GPCR GPR68 (also known as OGR1) is characteristic of the important roles these orphans are thought to play, and our difficulties in elucidating exactly what these functions might be. GPR68, together with three other receptors GPR4, GPR65 (TDAG8), and GPR132 (G2A), form a subfamily of proton-sensing GPCRs[17]. GPR68 has been shown to couple to multiple signaling

pathways in different cell types through G_q, G_s, $G_{12/13}$, or $G_{i/o}$ proteins[18-21], either in responding to protons or in a pH-independent manner. This receptor is expressed in most cells and tissues, but is most abundant in the cerebellum[22] and hippocampus[23] (Allen Brain Atlas, http://www.brain-map.org/), implying yet-to-be identified roles in brain function. GPR68 has been implicated in many physiological processes including osteoclastogenesis[24-28], insulin secretion[29], hippocampal neurogenesis[23], myocardial function[30], and sodium-hydrogen exchanger 3 (NHE3) and H^+-ATPase activity[31]. In responding to acidic microenvironments, GPR68 appears to regulate inflammatory processes in multiple cell types in which it is expressed, including airway smooth muscle cells (ASMCs)[32-38]. Confoundingly, previous studies with GPR68 knockout (KO) mice uncovered only modest changes in osteoclastogenesis, tumorigenesis, proton-sensing, insulin secretion, and airway smooth muscle inflammation[25,29,31,39], probably because of functional compensations from the other proton-sensing receptors. Although GPR68 was reported to be activated by a family of isoxazoles[30], their weak agonist activity appears to be non-specific[40,41] and could not be reproduced (See **Results** below). Thus, though GPR68 has been considered a proton receptor, and appears to have multiple roles, few of them are well-characterized by the knockout studies and none of its roles are known in the CNS, where it is most highly expressed. Like other targets that lack *bona fide* small molecule ligands[8], which can be independently added to a system (unlike protons), GPR68 remains "pharmacological dark matter". Said another way, without an exogenous ligand with which to modulate its activity, the pharmacology of GPR68 remains opaque.

Here we describe an integrated experimental and computational approach to discover ligands that illuminate and modulate the function of GPR68 *in vitro* and *in vivo*. Initial benzodiazepine hits from an empirical screen were used to anchor structural modeling of the receptor. Against these models, we computationally docked a library of 3.1 million organic molecules, looking for those that fit the modeled ligand binding site. Experimental tests of high-ranking docked molecules revealed a new family of ligands that were modeled to bind to the receptor in a similar manner to the benzodiazepine leads, but that explored new, more potent chemotypes. Further pharmacological characterization of selected lead compounds at GPR68 wild-type and mutant receptors in cell-based assays indicated that these ligands are positive allosteric modulators (PAMs) for the agonist, protons. A lead compound, dubbed ogerin, which functions as a PAM, is exemplified *in vitro* and found to be active *in vivo,* leading to new insights regarding the physiological functions of GPR68. Application of the same method for the discovery of allosteric agonists and negative allosteric modulators for a second orphan GPCR, GPR65, and the further structure-based optimization of the GPR68 PAM ogerin, suggest this approach, notwithstanding important caveats, may be useful to the field.

1.3 Results

I. Initial yeast growth based screening

In an initial de-orphanization campaign with selected orphan or poorly-annotated GPCRs, we developed a screening system in which yeast were engineered to grow

13

in the absence of histidine, but only when the receptor is activated by a ligand[42].

As most GPCRs have constitutive activity even in the absence of agonist, especially

when over-expressed, this system in principle enables the discovery of agonists,

inverse agonists, and allosteric modulators, as predicted by the two-state model of

agonism[43], though this has not been demonstrated at scale. We screened a small

library of approved drugs (NIH Clinical Collection of compounds NCC library;

http://www.nihclinicalcollection.com/) (**Supplementary Figure 1**). To validate

the yeast screening system, we confirmed the activity of short-chain carboxylic

acids at the FFA3 (GPR41) and the FFA2 (GPR43) free fatty acid receptors

(**Supplementary Figure 2a-d**). Further validation was provided using the orphan

receptor GPR39, in which we confirmed the activation of this target by zinc

ions[44,45] (**Supplementary Figure 2e**); interestingly, several other metal ions also

activated GPR39 in this system (**Supplementary Figure 2f-h**), an observation that

could be confirmed in assays in mammalian cells (**Supplementary Figure 2i-k**).

Upon screening the NCC library, we also discovered that the proton pump inhibitor

omeprazole (a racemate), but not its S-enantiomer, esomeprazole, also activated

GPR39 (**Supplementary Figure 2l**). The most striking result of this screen was the

discovery of the apparent agonist activity of lorazepam at GPR68 (**Figure 1**).
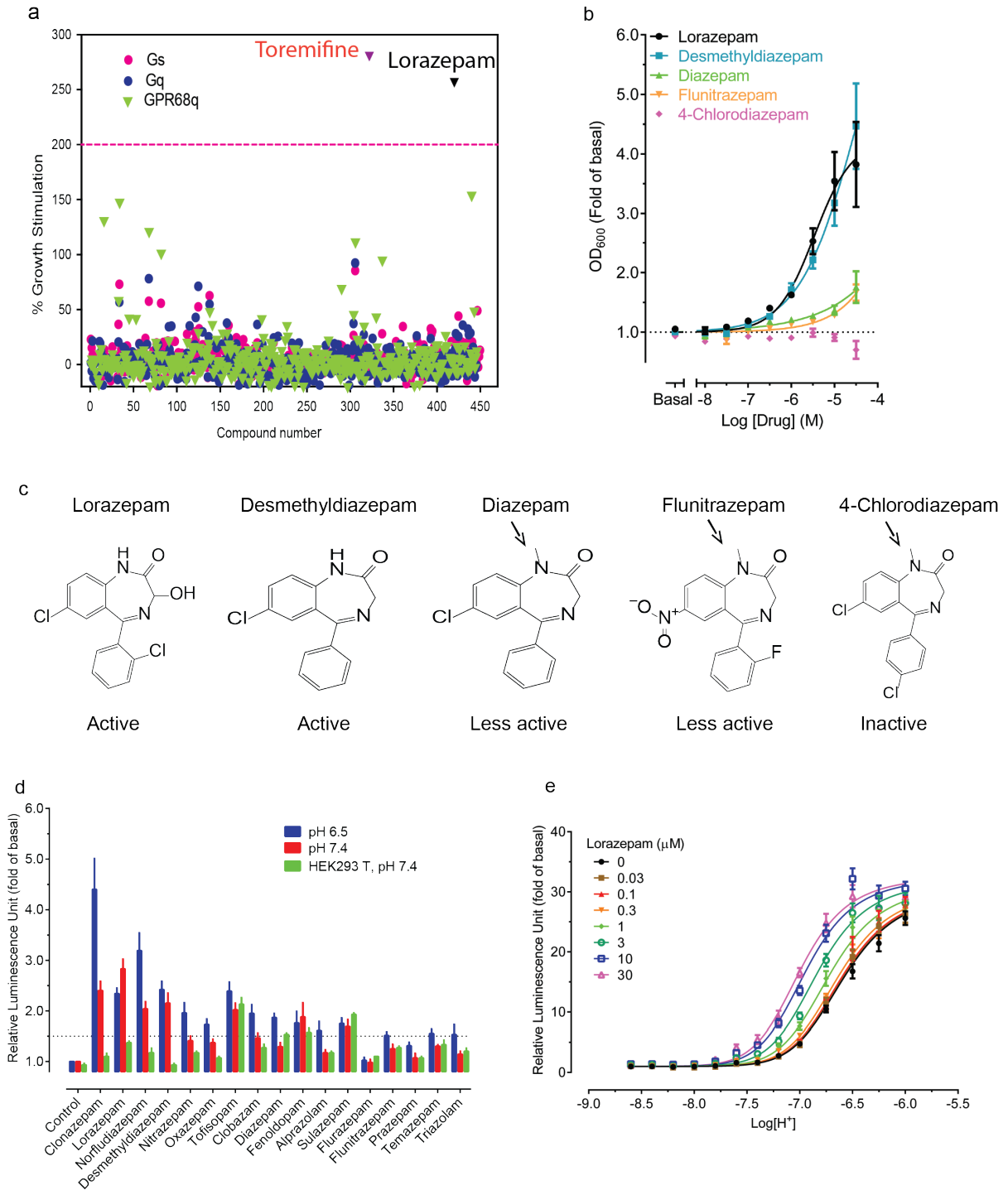
Figure 1.1 Yeast-based screening and identification of lorazepam as an allosteric modulator of proton agonism at GPR68.

15

(**a**) Lorazepam was an activator in yeast growth assays while toremifine was a false positive hit. The drug library was screened at 10 M to stimulate growth of G_s-yeast, G_q-yeast, and GPR68 transfected G_q-yeast. Results were normalized to percentages of growth and were presented as mean ± SEM from three measurements. (**b**) Concentration-dependent stimulation of GPR68 G_q-yeast growth by lorazepam and analogues. Results (fold of basal, OD600) represent mean ± SEM from a minimum of 3 assay done in triplicate and analyzed in Graphpad Prism using the built-in 4-parameter logistic function. (**c**) Representative structures of benzodiazepines (the arrows indicate methyl substituents that greatly reduce the activity of the compounds at GPR68). (**d**) Lorazepam and three other N-unsubstituted benzodiazepines (clonazepam, desmethyldiazepam, norfludiazepam) stimulated cAMP production in a GPR68- and pH-dependent manner. Results (RLU in fold of basal) represent mean ± SEM from a minimum of three assays (each in triplicate) at a final concentration of 10 μM. (**e**) Lorazepam acts as a positive allosteric modulator (PAM) for agonist proton at GPR68. Results (RLU in fold of basal) represent mean ± SEM from three independent assays (each in triplicate) and were analyzed in Graphpad Prism with the built-in 4-parameter logistic function.

II. Lorazepam and *N*-unsubstituted benzodiazepines are GPR68 PAMs

Among the potential hits of the yeast screening, lorazepam--a widely prescribed benzodiazepine anxiolytic--was most active and functioned as an apparent activator of GPR68 signaling, stimulating yeast growth by more than 2-fold

(**Figure 1a**). Further tests with lorazepam and four analogs in the yeast growth assay demonstrated that *N*-unsubstituted benzodiazepines were more efficacious than N-substituted benzodiazepines (**Figure 1b**, **1c**). We then tested 17 benzodiazepines at GPR68-transfected HEK293-T cells, seeking those that stimulated cAMP production. Lorazepam, along with three other *N*-unsubstituted benzodiazepines (clonazepam, desmethyldiazepam, and norfludiazepam, **Figure 1c** and **Supplementary Table 1**) acted as 'agonists' at both pH 6.5 and 7.4 (**Figure 1d**), with lorazepam displaying the most potent ability to shift the H^+ concentration-response profile (*i.e.*, making GPR68 more sensitive to proton activation) (**Figure 1e** and **Supplementary Figure 3a-d**). Importantly, lorazepam had no activity at control cells that did not express GPR68 (**Figure 1d**). The pH dependence of lorazepam's activity suggested that lorazepam functions as a GPR68 PAM. Intriguingly, lorazepam selectively potentiated proton-mediated G_s activation in a concentration-dependent fashion at GPR68 (**Figure 1e**), but not at two related proton-sensing receptors, GPR4 and GPR65 (**Supplementary Figure 4a-b**). Additionally, we profiled lorazepam against a panel of CNS targets and found significant activity only at the benzodiazepine site of the $GABA_A$ ionotropic receptor, its therapeutic target (**Supplementary Figure 5**). Together, these results support a relatively specific interaction between lorazepam and GPR68, with lorazepam acting as a GPR68 PAM.

III. Initial lead compound and molecular modeling of GPR68-lorazepam complex

While analogs around the lorazepam core structure included a few active *N*-unsubstituted benzodiazepines, little improvement in activity or selectivity was achieved. These findings, along with lorazepam's strong activity against its canonical target, the GABA-A receptor, motivated us to turn to structure-based docking screens to interrogate a much larger chemical space for candidate GPR68 PAMs (**Figure 2**). First, we generated 3-D models for the receptor using the CXCR4 chemokine receptor (PDB code: 3ODU) as a template, as it had the highest sequence identity to GPR68 (~29%) of all the then-available GPCR crystal structures. From the sequence alignment that emerged, 407 homology models of GPR68 were built. To these were added another 2,900 models using elastic network modeling (ENM), which sampled backbone and loop conformations (alignment given in **Supplementary Figure 6**).

Against each of the 3,307 models we docked the active benzodiazepines, the 446 compounds of the NCC library and 176 decoy molecules that resembled the benzodiazepines in physical properties, but were topologically dissimilar and thus unlikely to bind[46]. This docking calculation of 622 molecules was conducted against five different sub-sites of each of the 3,307 model structures (**Supplementary Figure 7**). These sites were chosen based on the binding regions of biogenic amine GPCRs, the peptide site of CXCR4 (PDB code: 3OE0), the 1T1t- antagonist site of CXCR4 (PDB code: 3ODU), and the allosteric binding site of the muscarinic M_2 receptor (PDB code: 4MQT). Several iterative cycles of modeling and local optimization of the possible lorazepam binding sites[47] (**Figure 2**), accounting for the benzodiazepine SAR and what is thought to be pH-based activation of

histidines 17, 84, 169 and 269[18], eventually converged to a stable docking pose for lorazepam (**Figure 2f**), and its ranking first in a small docking screen in which it was docked against the 622 inactive and decoy molecules. In this docked geometry, lorazepam is predicted to form hydrogen bonds with residues Glu160, Arg189, Tyr244, and Tyr268, and to form stacking and hydrophobic contacts with Trp77, Leu101, Phe173, His269, and Leu272 (**Figure 3d**).

To test the model, we made substitutions to the Glu160, Arg189, and His269 residues and determined their roles in proton-mediated cAMP production (G_s pathway) and calcium release (G_q pathway) using transiently transfected HEK293-T cells. The results indicated that these residues play important roles in receptor activation (**Figure 3b**, **3c** and **Supplementary Figure 8**). Consistently with a previous report[18], the H269F mutant receptor shifted proton concentration-response curves to the right in both cAMP production and calcium release assays. Substitutions at Arg189 selectively abolished cAMP production, with little effect on calcium release. Intriguingly, substitutions at Glu160 had different effects at downstream signaling pathways that depend on the nature of residue in the position – E160A shifted the proton concentration-response curve to the left together with reduced efficacy in cAMP production, but was inactive in calcium release, while the E160K or E160Q mutations had relatively modest effects in both pathways. Western blot analysis (**Supplementary Figure 8e**) indicated that the inactivity of some mutant receptors was not due to poor expression. It is interesting to note that mutations at residues such as Glu160 and Arg189 at the extracellular interface (**Figure 3a**) resulted in downstream coupling changes in one

but not the other pathway, indicating that these residues play critical roles in GPR68's functions. Therefore, these residues provided potential molecular mechanisms for pathway-selective modulation when modulators interact with these residues. Crucially, these results supported our GPR68 model as a basis for virtual screening.
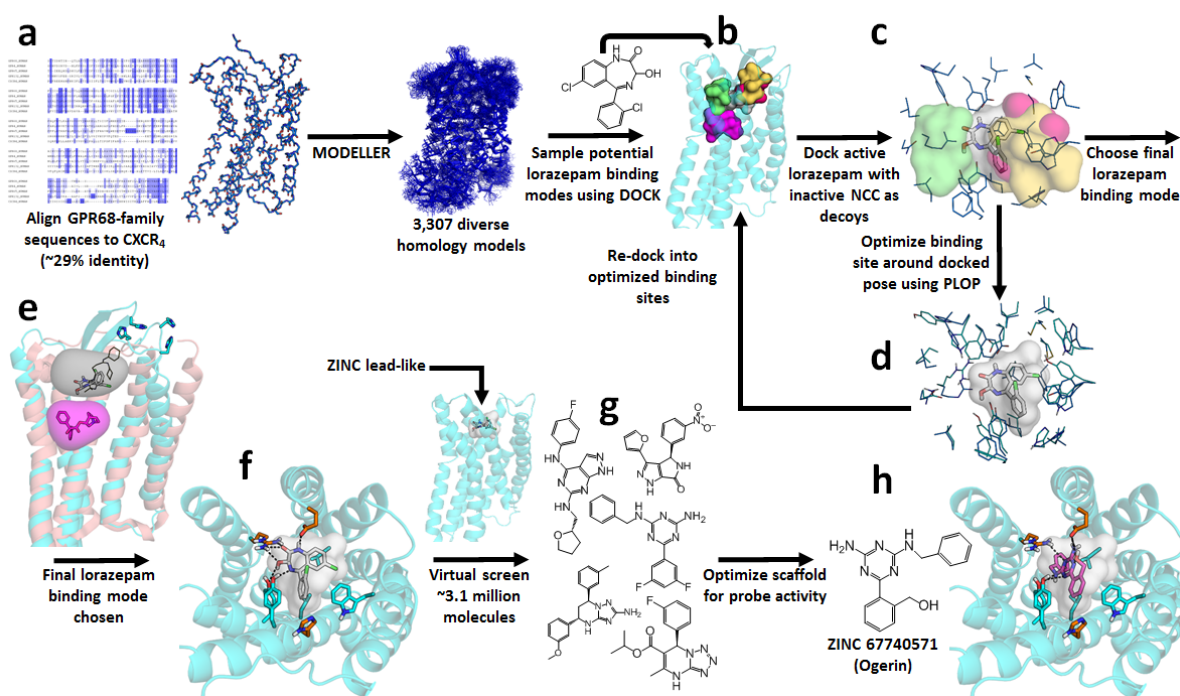
Figure 1.2 Virtual screening workflow and the predicted location of the allosteric binding pocket in GPR68.

(a) The primary sequences of human GPR68 and related receptors GPR4, GPR65, and GPR132 were aligned, initially automatically, to CXCR4 (PDB code: 3ODU) with PROMALS-3D, and then manually adjusted (**Supplementary Figure 3**). **(b)** We docked lorazepam and the rest of the NIH Clinical Collection (NCC) to five distinct binding sites (**Supplementary Figure 4**) in 3,307 conformationally diverse homology models generated by MODELLER and 3K-ENM. **(c)** Models were evaluated by their

favorable ranking of lorazepam vs. decoy molecules. (**d**) After each round of docking, the most favorable predicted lorazepam binding mode was chosen as the starting point for binding site optimization using PLOP. These sites were then re-docked to, and this process continued iteratively. (**e**) A final predicted binding orientation for lorazepam (gray stick) in GPR68 (cyan ribbon) was chosen, which corresponded to a region (in grey) overlapping with the allosteric site of the M_2 muscarinic receptor (shown in salmon ribbon, PDB code: 4MQT) and close to the binding site of the CXCR4 co-crystallized antagonist 1T1t (dark yellow wire). The orthosteric site to which QNB binds in the M_2 receptor is highlighted in magenta. (**f**) Extracellular view of lorazepam in its predicted orientation in the binding pocket (**Figure 3d for detail**). The native GPR68 proton-sensing histidines His17, 20, 84, and 169 are shown in cyan stick models. (**g**) This model served as a starting point for a large virtual screen of the lead-like library of ZINC (~3.1 million molecules). This virtual screen resulted in two sets of predicted hits for functional assays (**Figure 4**), one of which was ZINC 67740571. (**h**) Extracellular view of ZINC 67740571 (magenta stick) in its predicted orientation in the binding pocket of GPR68 (**Figure 3d for detail**).
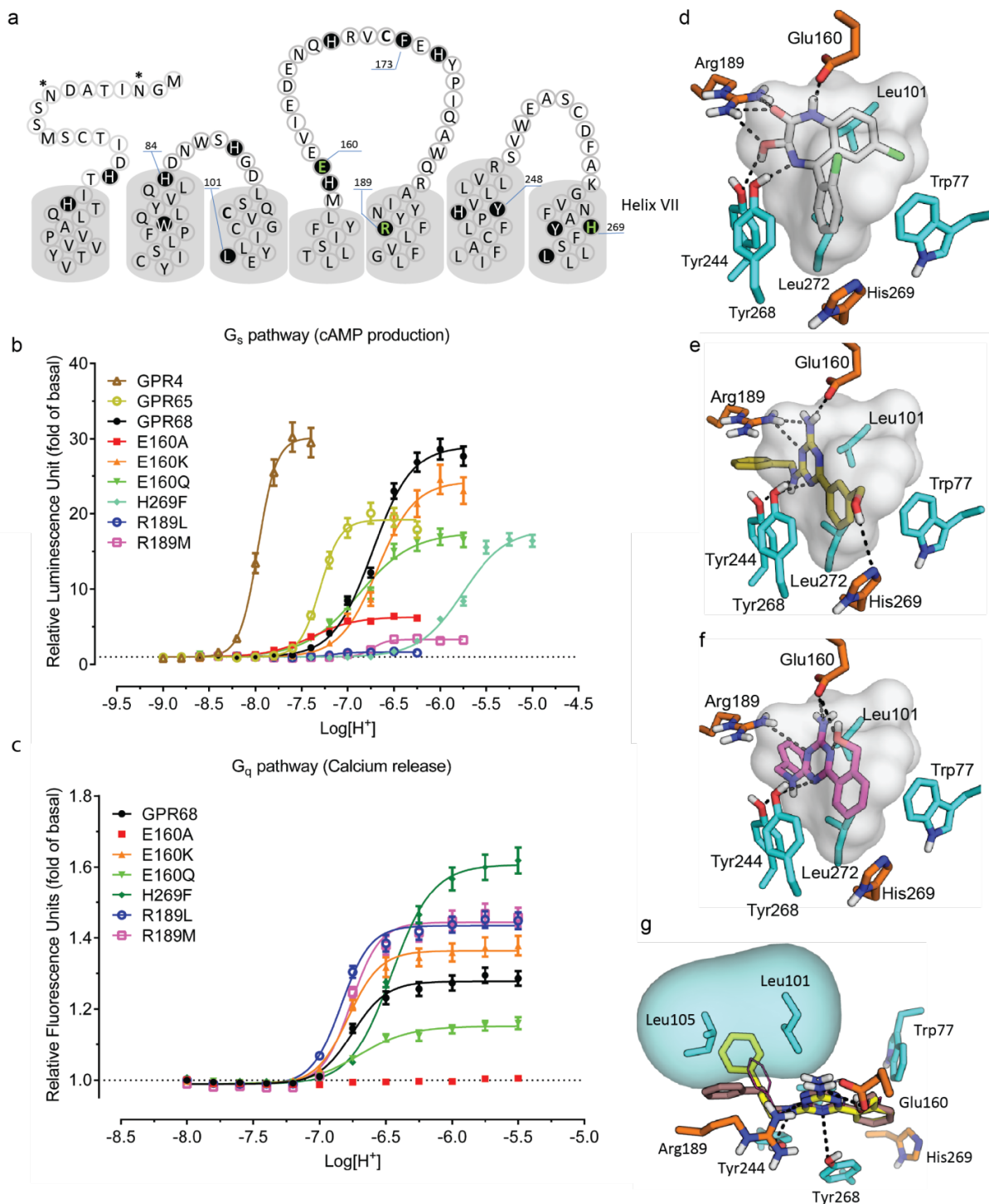
Figure 1.3 Experimental confirmation of important residues in GPR68 activity in transiently transfected HEK293-T cells.

(a) GPR68 snake plot showing extracellular loops and upper portion of transmembrane helices. The plot was made with information from the GPCR database (http://www.gpcr.org/7tm/). Amino acids in dark background were identified in homology modeling and docking studies to be important in ligand binding and receptor activation. Glu160, Arg189, and His269 are the residues mutated in this study. Protons showed concentration-dependent agonist activity at wild-type and mutant receptors in cAMP production measured with a luciferase reporter assay (**b**) and calcium release measured with a FLIPR (**c**) and pharmacological parameters are presented in **Supplementary Figure 7**. Results (fold of basal) represent mean ± SEM from at least 5 independent assays, each in triplicate or quadruplicate, and were analyzed in Prism using the built-in 4-parameter logistic function. (**d**) Lorazepam is predicted to interact via hydrogen-bonding interactions with Glu160, Arg189, Tyr244, Tyr268, and through stacking and hydrophobic contacts with Leu101, His269, Trp77, Leu272, and Phe173 (in front of the picture therefore not visible). ZINC 32547799 (**e**) and ZINC 67740571 (ogerin, **f**) are predicted to hydrogen-bond with Glu160, Arg189, Tyr244, Tyr268, and make stacking and hydrophobic contacts with Leu101, His269, Trp77, Leu272, and Phe173 (not shown). Optimization of ogerin (magenta, thin lines) to **C2** (brown, structure in **Figure 4a**) by insertion of a single methylene is predicted to improve packing in the aryl pocket of the ogerin site, correspondingly improving positive allosterism (increased value by 2.6-fold). Adding a second methylene, thus creating a propyl linker in **C3** (yellow, structure in **Figure 4a**), is predicted to disrupt this packing and thus to reduce the allosteric effect (**g**).

IV. Virtual screening for novel modulators of GPR68 and lead optimization

Seeking new PAMs, we docked the 3.1 million commercially available molecules of the lead-like subset of the ZINC library against the GPR68 model, using DOCK3.6. On average, 1651 orientations were sampled for each molecule, and for each orientation, about 663 conformations; for the 3.1 million molecules, a total of more than 3.3 trillion complexes were calculated in the GPR68 binding site. Each of these was then scored using a physics-based scoring function[46,48]. From among the top 0.1% of the docking-ranked molecules, 17 were purchased for initial testing; in addition to their high docking ranks and scores, these compounds were selected for their recapitulation of the key interactions made by lorazepam in its docked model, for chemical diversity within the set, and for representation of other similar candidate ligands in the top-ranked list (**Supplementary Table 2**). As previously, molecules that scored well by the DOCK3.6 protein-interaction score, but had unfavorable internal conformational energies, or that had unreasonable ionization states owing to errors in pK_a calculations in the ZINC library, were deprioritized in this final triaging step (**Methods**).

The 17 docking hits were tested for functional activity in cAMP accumulation assays in GPR68-transfected HEK293-T cells. Four of the seventeen hits increased cAMP production close to or over 1.5 fold of basal at pH 6.5 (for a 29% hit rate) (**Figure 4a** – 1st batch), and none were active in control cells that did not express GPR68. Though none of the compounds were as active as lorazepam,

two compounds, ZINC 4929116 and 32587282, each with over 1.5 fold of basal activity at pH 6.5, had hundreds of commercially available analogs, reflecting one of the criteria for their selection. These analogs were docked again to the GPR68 model, and 14 analogs of ZINC 4929116 and 11 analogs of ZINC 32587282 were tested (**Figure 4a** – 2nd batch, **Supplementary Table 3**). Three of the compounds had greater activity than lorazepam at pH 7.4, and at pH 6.5, ten additional compounds showed activity at close to or above 1.5 fold of basal. The nature of the pH-dependent agonist activity is a clear sign of allosteric modulation (below). Intriguingly, despite the substantial differences in chemotype between lorazepam and ZINC 67740571, one of the most efficacious analogs at pH 7.4, the new compound docks to form many of the same interactions predicted for lorazepam (**Figure 3d** vs **3f**). In addition, an extra hydrogen bond to Glu160 on the receptor predicted to be formed by the hydroxyl substituent of ZINC 67740571.
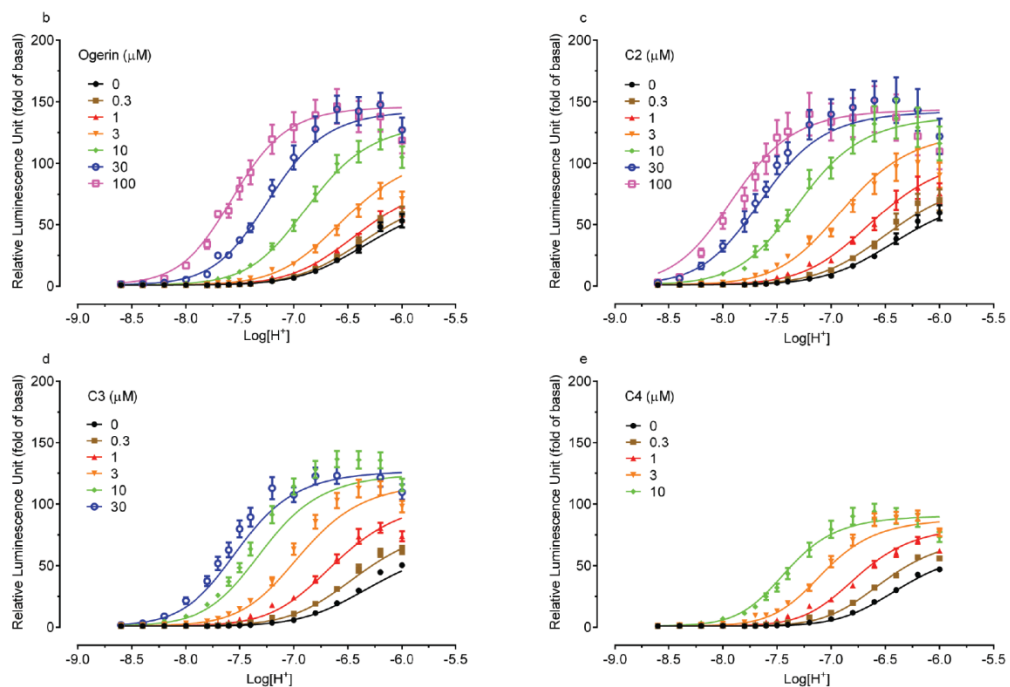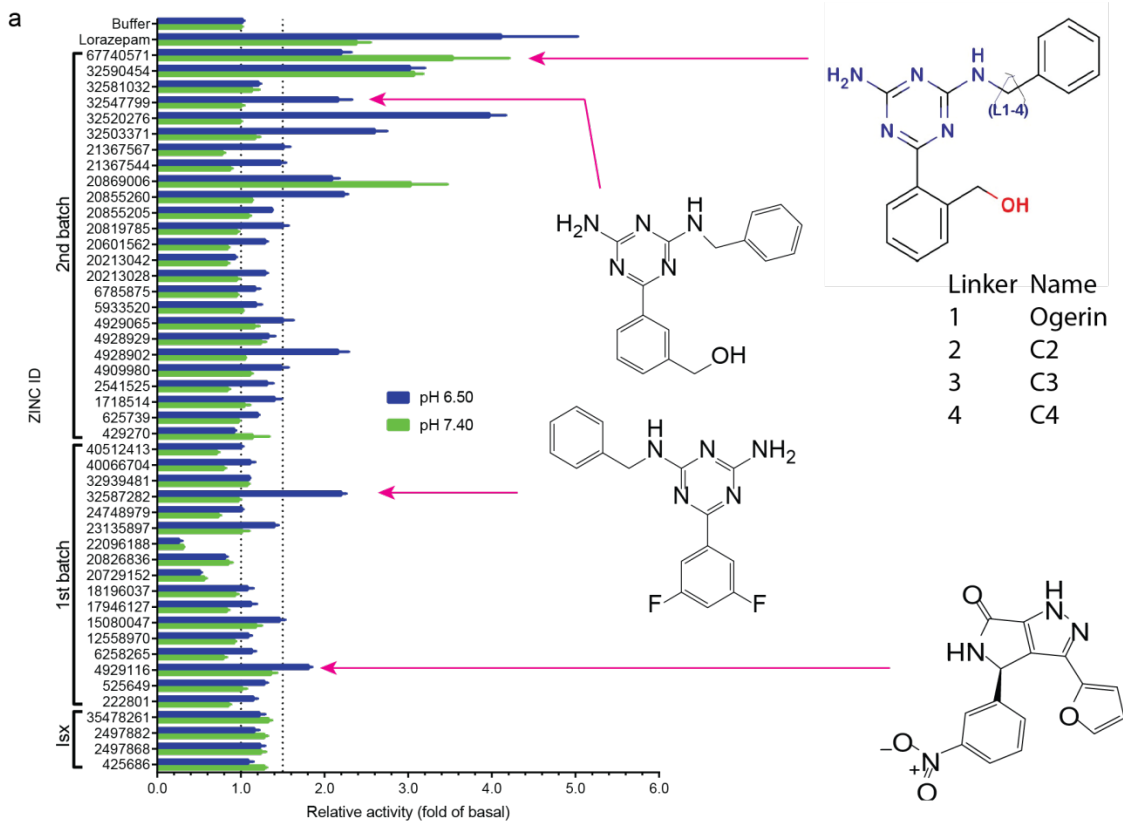
Figure 1.4 Identification, characterization and optimization of positive allosteric modulators of proton at GPR68.

(a) Functional screening of compounds from the structure-based docking screen. GPR68-mediated cAMP production was measured in transiently transfected HEK293-T cells using a luciferase reporter assay. Results of selected compounds (ZINC database numbers) represent normalized (fold of basal) mean ± SEM from a minimum of 4 independent assays, each in quadruplicate, at a single concentration of 10 µM at pH 7.40 and 6.50. Compounds were grouped into a 1st batch from the 1st round of virtual docking, and a 2nd batch from the 2nd round of docking analysis. Compounds labeled Isx are isoxazole analogues. Two lead compounds (32587282 and 4929116) from the 1st batch, as well as ogerin (67740571), its isomer (32547799), and its analogues (**C2**, **C3** and **C4**) with different lengths of linkers, are highlighted. Concentration-response curves of ogerin (**b**), **C2** (**c**), **C3** (**d**), and **C4** (**e**) are shown to illustrate allosteric potentiation of proton concentration responses. Results (fold of basal) represent mean ± SEM from a minimum of 3 independent assays, each in triplicate or quadruplicate; curves were fitted to and analyzed with the standard allosteric operational model in Graphpad Prism, and the allosteric parameters are summarized in **Supplementary Table 4.**

V. Identification of ogerin as an efficacious and selective GPR68 positive allosteric modulator

Ten of the new compounds increased cAMP signaling activity more than 1.5 fold over basal level, and were selected for further study. Like lorazepam, all were PAMs at GPR68: they lacked intrinsic activity at GPR68, but after being exposed to higher concentrations of the proton agonist, shifted the proton concentration-

response curves leftward and typically upward (**Supplementary Figure 10**). The most active of them, ZINC 67740571, had a substantially higher allosteric effect than lorazepam (**Figure 4b vs Figure 1e**), shifting the proton dose-response curve leftward 10-fold (H^+ potency), and increasing the signaling plateau of the receptor by over 2-fold (H^+ efficacy). It is this ability to shift the concentration-response curves leftward and upward that is the most important characteristic of a PAM. We dubbed this molecule "ogerin" (for OGR1 ligand).

As true PAMs that lacked activity on their own (**Supplementary Figure 12a**), we analyzed their modulation of proton signaling with a standard allosteric operational model[49,50] (**Figure 4b** - **4e** and **Supplemental Table 4**). Unlike orthosteric ligands, for which the binding affinity (K_i) is an important parameter, the cooperativity parameters α and β are the key metrics for allosteric modulators in the presence of H^+. The parameter α defines the allosteric cooperativity on ligand binding affinity ($\alpha > 1$ for increasing affinity and $0 < \alpha < 1$ for decreasing affinity), and is responsible for the leftward or rightward shift in the agonist curves, while β, defines the allosteric cooperativity on agonist efficacy (top of the curve, $\beta > 1$ for increasing efficacy and $0 < \beta < 1$ for decreasing efficacy). Conversely, the extrapolated binding affinity parameter (K_B) has little physical meaning in this system, since H^+ exists at functionally relevant concentrations at physiological pH. Even so, the extrapolated average K_B is 8 μM (pK_B of 5.10, ranging from 4.11 – 5.89) for the 10 PAMs in this study (**Supplemental Table 4**), which is comparable to that of the well-known muscarinic M_1 PAM BQCA (benzyl quinolone carboxylic acid), a tool molecule that has a pK_B of 4.72[51].

Ogerin and its six analogues formed a group with the same core structure and provided initial insights into structure-activity relationships (SAR). Among them, ogerin and ZINC 32547799 were isomers, with a hydroxylmethyl at the ortho-position for ogerin and at the meta-position for ZINC 32547799 (**Figure 4a**); each displayed distinct functional activities (Figure 5a - 5c and **Supplementary Figure 15**) and docking poses (**Figure 3e**, **3f**). Thus, the ortho-hydroxylmethyl is predicted to play a critical role in determining PAM activity and receptor activation, perhaps because of its modeled ability to hydrogen-bond with Glu160, which the meta-positioned hydroxylmethyl cannot reach (**Figure 3e**, **3f**); this is consistent with the mutagenesis results at this residue (above). Analogs with other substituent groups at the same meta-position, or other positions on the benzene ring, also displayed lower PAM activity (**Supplementary Table 4**). The receptor mutants H269F and R189L responded to ogerin and its isomer in dramatically different ways (**Figure 5a** – **5c**, and **Supplementary Figure 14**), further supporting the notion that these PAMs modulate proton activity via predicted interactions with residues His269 and Arg189, among others. Ogerin and its isomer (ZINC 32547799) were also tested for their ability to modify GPR68-mediated calcium release (G_q-signaling), as opposed to the cAMP production (G_s signaling) described above. Strikingly, after 10 min incubation with GPR68-transfected cells, ogerin inhibited proton-mediated calcium release – a pathway-specific function rescued by two structure-guided mutations (R189L and H269F) – while the ogerin isomer ZINC 32547799 had essentially no effect on calcium release at either wild-type or mutant receptors (**Figure 5d** – **5f** and **Supplementary Figure 15**). To determine if

29

fast kinetics in the calcium release measurement play a role in the difference between effects on cAMP measurement (under equilibrium) and calcium release (under non-equilibrium), we also carried out phosphatidylinositol (PI) hydrolysis assays, and ogerin showed a small potentiation of proton activity as exemplified by an increased baseline (**Figure 5g**). The ogerin isomer 32547799 had no effect on proton activity in the PI hydrolysis assay, again consistent with its inactivity in the calcium release assay. To determine ogerin's specificity among related receptors, which would be crucial for its use as a biological probe, we also tested it at two other proton-sensing GPCRs, GPR4 and GPR65. Ogerin had minimal PAM activity at either GPR4 or GPR65 (**Supplementary Figure 16a-b**), even though residues such as Glu160, Arg189, and His269 are conserved in one or both receptors (**Supplementary Figure 6**). Therefore, ogerin is a selective GPR68 PAM for the agonist proton.
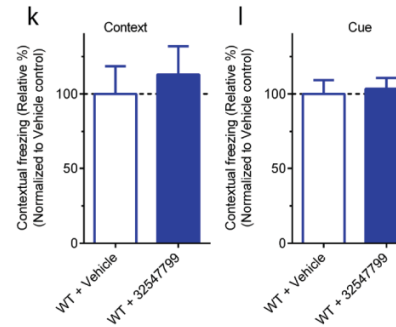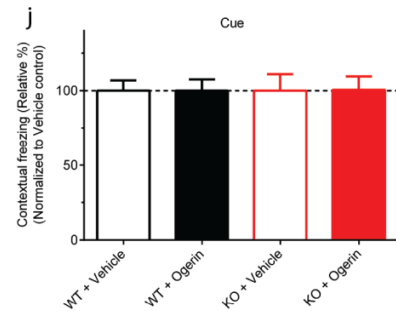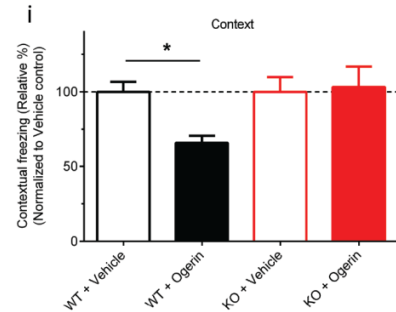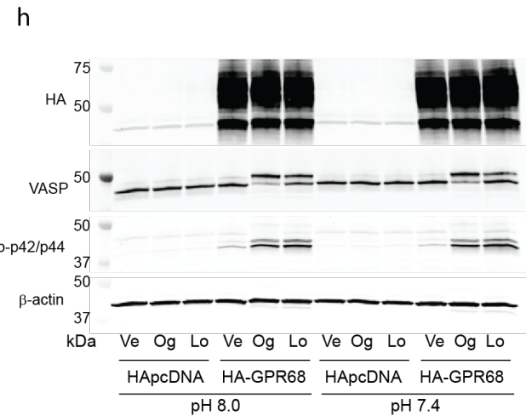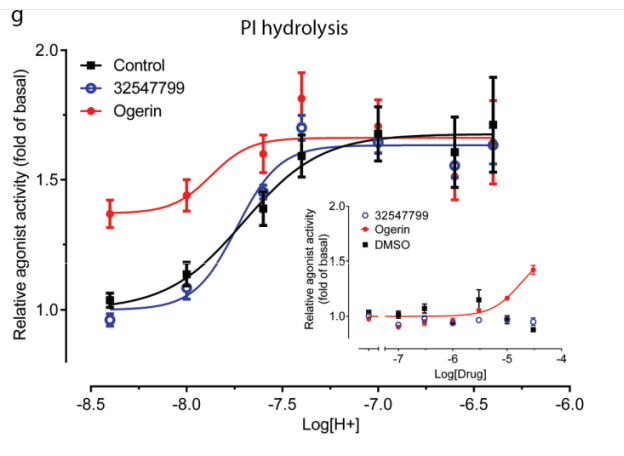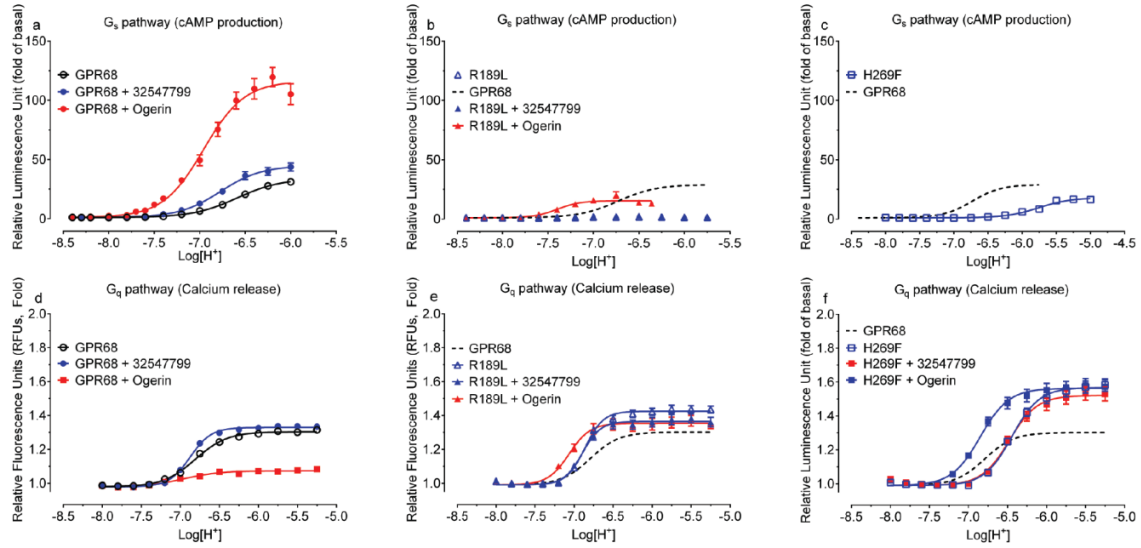
Figure 1.5 Functional comparison of ogerin and its isomer and *in vivo* activity of ogerin on learning and memory.

Effects on responses to protons of ZINC 32547799 (10 μM) and ogerin (10 μM) on GPR68-mediated cAMP production (G_s-pathway) (**a**), GPR68-mediated calcium mobilization (G_q-pathway) (**d**), GPR68 R189L mutant-mediated cAMP production (**b**), GPR68 R189L mutant-mediated calcium mobilization (**e**), GPR68 H269F mutant-mediated cAMP production (c), and GPR68 H269F mutant-mediated calcium mobilization (**f**). Functional assays were carried out with HEK293-T cells transiently transfected with GPR68 wild-type or mutant receptors. For calcium release assays (**d-f**), compounds were first incubated with cells for 10 min before proton-mediated responses were measured. (**g**) Effect of ogerin (10 μM) and its isomer 32547799 (10 μM) on proton-mediated PI hydrolysis in GPR68-transfected HEK293-T cells. Ogerin and 32547799 concentration-response curves at pH 8.40 are shown in the insert of panel **g**. Results in panels **a** – **g** (fold of basal) represent mean ± SEM from a minimum of 3 independent assays, each in triplicate, and were analyzed in Graphpad Prism using the built-in 4-parameter logistic function. (**h**) Ogerin (Og) and lorazepam (Lo) activate PKA and p42/p44 MAP kinase through GPR68. HEK293 cells stably expressing GPR68 (HA-GPR68) or not (HApcDNA) in media at pH 8.0 or 7.4 were stimulated with vehicle (Ve), 50 μM ogerin, or 50 μM lorazepam for 10 min. Immunoblot analysis of harvested lysates demonstrates ogerin- and lorazepam-induced phosphorylation of both p42/p44 and vasodilator-stimulated phosphoprotein (VASP, 50 kDa band is PKA-phosphorylated[37]) in GPR68-expressing cells only, under both pH conditions. (**i** – **l**) Ogerin suppresses learning and memory in wild-type but not in GPR68 KO mice. Ogerin decreased contextual memory retrieval in wild-type mice but not in GPR68 KO mice (**i**, there was no effect of drug

or genotype, however, there was a significant drug x genotype effect, $F_{(1,27)} = 4.71$, $p < 0.05$; Bonferroni's post-hoc test showed that ogerin significantly decreased the freezing in wild-type mice ($p < 0.05$, two-way ANOVA) but not KO mice ($p > 0.05$, two-way ANOVA). Ogerin had no effect on cued memory retrieval in either wild-type or GPR68 KO mice (no main effect of drug ($F_{(1,27)} = 0.0005$, $p = 0.982$), genotype ($F_{(1,27)} = 3.281$, $p = 0.066$) and drug x genotype interaction ($F_{(1,27)} = 0.068$, $p = 0.935$), two-way ANOVA) (**j**). Results were normalized to corresponding vehicle control and raw data are presented in **Supplementary Figures 21c** and **21d**. The ogerin isomer (ZINC 32547799) had no effect on either contextual (**k**) or cued (**l**) memory retrieval at wild-type mice. Results were normalized to vehicle control, and raw data are presented in **Supplementary Figures 22b** and **22c**.

VI. Structure-based optimization for potency

If the ogerin/GPR68 model is relevant and predictive, we should be able to leverage it to optimize the molecule's allosteric effects. We thus designed and generated a virtual library of over 600 synthetically accessible ogerin analogs and docked each into the GPR68 model, prioritizing those that were predicted to form favorable interactions with key recognition residues (**Figure 3f-g**). Thirteen high-scoring molecules were synthesized and three were more active than ogerin (**Supplementary Table 5** and **Supplementary Figure 17**), including the molecule with the best docking rank of the initial 600 analogs (compound 33550) and the most active, compound **C2** (**Figure 4a insert**), ranked 7th in the docking. Addition of a methylene in what had been the benzylamine side chain, converting it to a

phenyethylamine, is modeled to place the phenyl ring deeper into the apolar

pocket of the ogerin site (Figure 3g), resulting in an enhanced allosteric effect (a

2.2-fold higher $\alpha*\beta$ value vs. ogerin), shifting the pH-response curve further

leftward and upward to higher signaling efficacy (**Figure 4b, c,** and

Supplementary Table 4). The addition of one or two methylene residues to make

the longer ogerin analogues, **C3** and **C4** (**Figure 4a insert**), on the other hand, did

not further enhance, but rather reduced, the allosteric effect (**Figure 4d, e,**

Supplementary Table 4, and **Supplementary Figure 18b**), consistent with the

prediction that the longer linkers would cause the phenyl ring to clash with the

pocket formed between Val105 and Leu101, presumably leading to a repositioning

of the ligand and attenuating its allostery. The ability to optimize for more

efficacious analogues of ogerin, and to distinguish them from close mimics with

different degrees of allosteric potentiation, support a structure-based strategy for

the optimization of GPR68 PAMs.

VI. Lead profiling for specificity and probe suitability

While ogerin was specific to GPR68 vs. the related receptors GPR65 and GPR4, its

activity against unrelated targets is also important to determine if it is to be

useful as a biological probe. We first profiled ogerin and its analogs for off-target

activity computationally, using the Similarity Ensemble Approach (SEA), which has

previously been effectively used for this purpose[52,53]. SEA can rapidly interrogate

over 2,800 pharmacologically-relevant targets[54] based on the chemical similarity of

the query ligands to sets of ligands annotated to be active at potential off-targets.

SEA found significant similarity between the GPR68 ligand set and only three GPCR targets: the ghrelin receptor, and the adenosine A_1 and A_{2A} receptors. Subsequent physical profiling against a panel of 58 GPCRs, ion channels, and transporters (**Supplementary Figure 5**) revealed that ogerin had only moderate affinity at two GPCRs, including 5-HT$_{2B}$ (K_i = 736 nM). Consistently with the SEA prediction, the other receptor showing activity was the A_{2A} receptor (K_i = 220 nM), at which ogerin was an inverse agonist (**Supplementary Figure 19**).

VII. Potential connections among proton receptors, adenosine receptors, and GABA channels

We were intrigued by the association between the new GPR68 PAMs and the adenosine receptor antagonists. To probe this further, we computationally screened a library (http://www.tocris.com/dispprod.php?ItemId=5386#.U_s5ZMVdUrU) of 1,120 biological reagents and drugs against the set of GPR68 ligands. Here we used SEA in reverse, with the 10 GPR68 PAMs as the ligand set and each of the 1,120 library molecules as separate queries. From this screen, SLV320, a putatively selective adenosine A_1 receptor antagonist[55], was predicted to be a GPR68 PAM, based on chemical similarity to the 10 GPR68 PAMs. We then physically screened the full library against GPR68 for cAMP production in transiently transfected HEK293-T cells, and confirmed that SLV320 ($\alpha\beta$ = 2.2) was indeed a GPR68 PAM (**Supplementary Figure 20** and **Supplementary Table 4**). Intriguingly, a second adenosine receptor antagonist, CGH2466 ($\alpha\beta$ = 2.1), was also observed to activate

GPR68, as did tracazolate ($\alpha\beta$ = 2.8), a GABA-ergic anxiolytic drug that also antagonizes adenosine receptors[56]. Whereas CGH2466 is the most potent GPR68 PAM thus discovered (K_B= 66 nM), its allosteric effect is lower than the docking-derived PAMs (such as ogerin or **C2**). Taken together, these results point to a previously unknown cross-talk among the *ligands* of GPR68, those of adenosine receptors, and those of the GABA ion channel (**Supplementary Figure 21**). This must be considered when evaluating the pharmacology of what have long been considered specific probes and drugs, such as lorazepam, tracazolate, CGH2466 and SLV320; the GPR68 activity of lorazepam could conceivably explain some of its idiosyncratic side effects, as well as those of non-benzodiazepine drugs (below).

VII. Ogerin as a pharmacological probe for GPR68 biology

With such clear functionally selective activity and specificity, we sought to explore ogerin's downstream signaling pathways and its *in vivo* activity. We first evaluated it in GPR68-expressing HEK293 cells and found that ogerin, as well as our initial lead compound lorazepam, activates the PKA and MAP kinase pathways (**Figure 5h**), mimicking the low pH-induced signaling pathways observed with GPR68 receptors in human ASM cells[37]. The extracellular acidification-mediated GPR68 activation in smooth muscle cells is linked to multiple downstream signaling pathways and biological responses[32-37,39], and a selective allosteric modulator such as ogerin may help delineate GPR68 signaling pathways and biological responses in smooth muscle cells.

Perhaps more compelling were effects in behaviors associated with modulation of the function of the hippocampus, where GPR68 is highly expressed[22] (Allen Brain Atlas). We evaluated the performance of GPR68 KO mice in a classic learning and memory test, fear conditioning, in which the hippocampus plays important roles (**Supplementary Figure 22,** see **Supplementary *In Vivo* Methods and corresponding Supplementary Figures** for initial behavioral profiling of GPR68 KO mice). In wild-type mice, ogerin (10 mg/kg) significantly attenuated learning (**Supplementary Figure 23a**) and contextual-based fear memories (**Figure 5i**), without effect on cue-based memory (**Figure 5j**). The effects we observed with ogerin are comparable to the effects of compounds targeting other GPCRs (such as mGluR1[57], M_1[58], KOR[59] and 5-HT_{1A}[60]), which are also highly expressed in hippocampus, administered either subcutaneously or directly into the hippocampus. Larger effects than those seen here are typically only observed following surgical lesions of the hippocampus[61]. Disrupting these GPCR signaling pathways may alter the neuronal network representing the contextual information in the hippocampus during the memory-encoding process, therefore impairing the hippocampus-dependent contextual memory, while sparing the hippocampus-independent cued memory. Importantly, administration of ogerin (10 mg/kg) had no effect on learning (**Supplementary 23b**) or memory retrieval in GPR68 KO mice (**Figure 5i** and **5j**), indicating that ogerin's *in vivo* effects are dependent on the GPR68 receptor. Furthermore, the less active ogerin isomer, ZINC 32547799, had no measurable effect on learning and memory in wild-type mice at the same dose of 10 mg/kg (**Figure 5k-l** and **Supplementary Figure 24**). Taken together, ogerin,

a novel allosteric modulator specific for the GPR68 receptor, illuminates a potential role for GPR68 PAMs in modulating learning and contextual fear-conditioning.

VIII. General applicability of the approach

We next explored the general usefulness of the computationally-intensive approach for GPCR de-orphanization. Here, we sought ligands for GPR65, another pH-sensing orphan receptor sharing only 37% sequence identity to GPR68. Recently, the compound BTB09089 was reported to be a GPR65 agonist[62] with a potency of 7.8 μM (E_{max} of 7.59 ± 0.33 fold of basal and pEC_{50} = 5.11 ± 0.17, **Figure 6f**); unlike ogerin at GPR68, BTB09089 does not potentiate proton activity at GPR65 (**Figure 6d** and **Supplementary Figure 25a**). BTB09089, therefore, is an allosteric agonist at GPR65. As with the lorazepam-GPR68 pair, we used BTB09089 to anchor modeling of GPR65 and explored 500 GPR65 homology models templated on the GPR68 model, as the sequence identity to any known GPCR crystal structure was too low for accurate modeling. The final docked GPR65-BTB09089 model resembles that of GPR68-ogerin, though with several side-chain substitutions in the putative binding site (**Figure 6a**). To test this model, we made substitutions at three residues predicted to interact with BTB09089 in its docked pose, and a fourth, Asp153, that was nearby but was not predicted to interact with the ligand. With the exception of Asp153, all three substitutions substantially reduced proton activity, shifting proton concentration-responses curves rightwards or downwards or both (**Supplementary Figure 26a**), consistent with the model.

We then docked a 3.1 million compound lead-like subset of the ZINC database against the GPR65 model, purchasing 45 molecules that scored in the top 0.5% of the docking-ranked list for experimental testing. Although these molecules were modeled to bind to the same pocket as BTB09089, they were structurally dissimilar to that molecule (**Supplementary Table 6**) (**Figure 6a-c**). One compound, ZINC 13684400, showed agonist activity of over 2 fold of basal at GPR65, with a potency of 500 nM (E_{max} of 2.18 ± 0.06 fold of basal; pEC_{50} of 6.30 ± 0.08), with no measurable activity at untransfected control cells (**Figure 6d**, **6f**, and **Supplementary Figure 25b-c**). As with BTB09089, ZINC 13684400 did not potentiate proton efficacy at GPR65 (**Figure 6d**). Therefore, ZINC 13684400 is a novel allosteric agonist for GPR65 with no potentiating effect on proton activity, with a higher potency than the known allosteric agonist BTB09089, and a different chemotype. Like BTB09089, the activity of ZINC 13684400 was much reduced at the three GPR65 mutant receptors (**Supplementary Figure 26b, 26c**): R187L, F242A, and Y272A. Intriguingly, while the Asp153→Ala substitution had no effect on BTB09089, it greatly reduced the activity of ZINC 13684400. This is consistent with the hydrogen bond that the latter compound makes with Asp153, and the modeled complex.

Several of the new docking-prioritized compounds, such as ZINC 41613384, 9468042, 62678696, and 78874232, inhibited GPR65 when the receptors were activated by protons or BTB09089 at pH 8.4, but not control cells or under control conditions (ISO-activated signals at GPR65-transfected cells when GPR65 was kept inactive at pH 8.4) (**Supplementary Figure 25g, 25f**). Among these, ZINC

62678696 (**Figure 6e**) was characterized further. ZINC 62678696 showed concentration-dependent partial inhibition of proton activity (**Figure 6g**), suggesting that it is a negative allosteric modulator (NAM) for protons at GPR65. To explore the interaction between the allosteric agonist BTB09089 and the NAM ZINC 62678696, we measured BTB09089 concentration-response curves in the absence and presence of increasing concentrations of ZINC 62678696 (**Figure 6h**). Unexpectedly, the efficacy of BTB09089 was reduced by ZINC 62678696 in a dose-dependent apparently non-competitive manner – an indication that BTB09089 and ZINC 62678696 could bind to the GPR65 simultaneously and allosterically interact. To accommodate such an allosteric mechanism, we looked for a new docking configuration that could accommodate both BTB09089 and 62678696 simultaneously with GPR65, ultimately finding that both compounds could be docked to form a ternary complex with the receptor (**Figure 6i**). In this new ternary configuration, BTB09089 retains its overall position in the site, though its interactions are perturbed, while ZINC 62678696 moves down in the site (**Figures 6a and 6c** vs **6i**). Analysis with the standard allosteric operational model suggested that ZINC 62678696 had a large negative allosteric effect on the efficacy of BTB09089 ($\log\beta$ = -10.93 ± 0.45, β = 1.17E-11) but a small effect on the affinity of BTB09089 ($\log\alpha$ = 0.68 ± 0.29, α = 4.79). We were unable to characterize the interactions between ZINC 62678696 and BTB09089 in more detail without additional tool compounds, such as a radioligand or more efficacious agonists than BTB09089. In summary, two docking-derived novel allosteric modulators for GPR65

have been characterized, an allosteric agonist and a negative allosteric modulator, with several others emerging that merit further study.
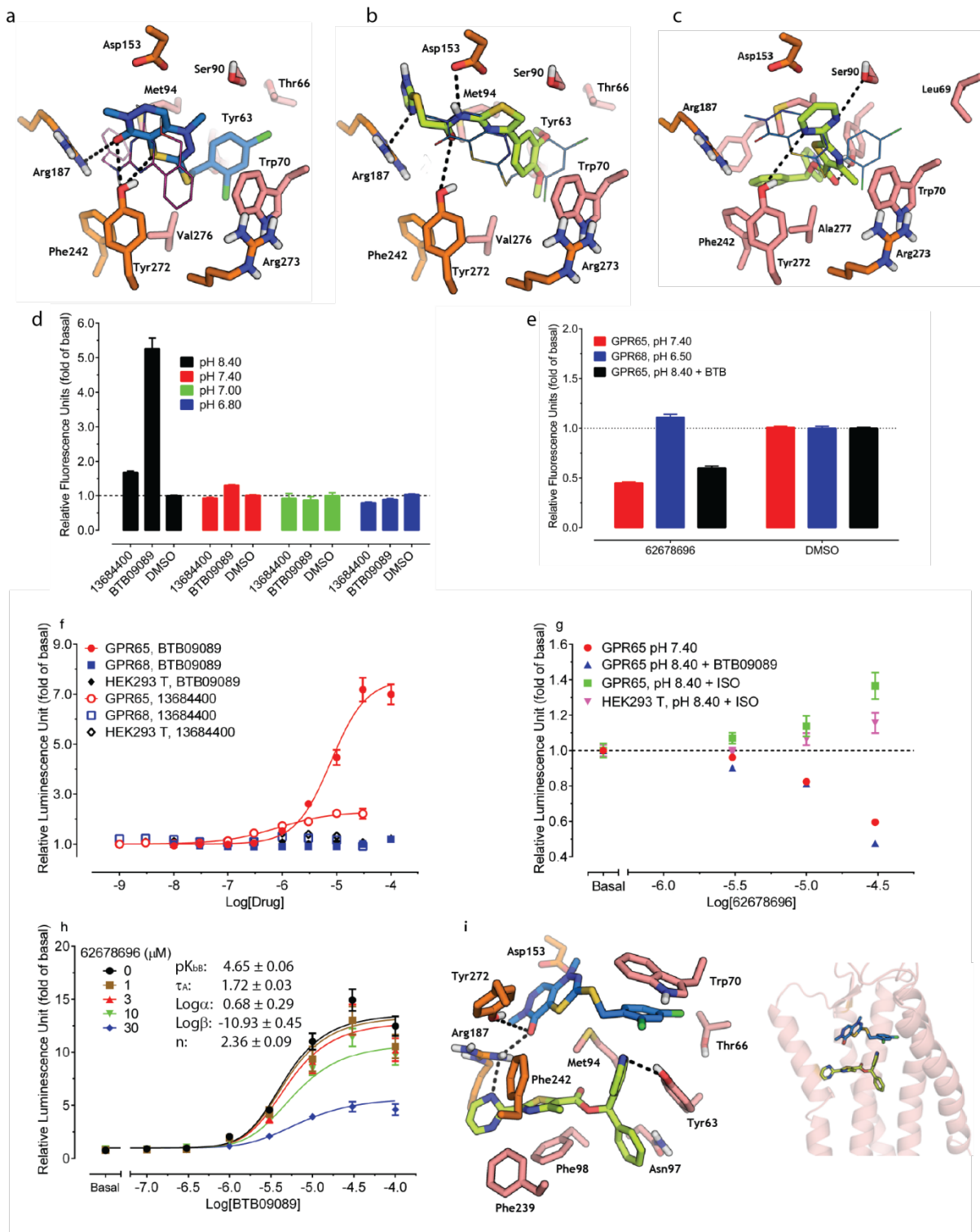
Figure 1.6 Structure-based discovery of allosteric agonist and negative allosteric modulator of GPR65.

BTB09089 (carbons in blue) was leveraged to create a model of GPR65 (carbons in maroon) that could fit this allosteric agonist; the final modeled complex is shown overlaid with the ogerin docked pose from GPR68 (thin magenta lines) (**a**). A docking screen of 3.1 million molecules combined with experimental screening led to the discovery of one allosteric agonist, ZINC 13684400, and several negative allosteric modulators, the most potent of which is ZINC 62678696. The docked pose of ZINC 13684400 (**b**) and ZINC 62678696 (**c**) are also shown. HEK293-T cells were transiently transfected with GPR65 receptors and production of cAMP was measured by a split-luciferase reporter assay (**d**). ZINC 13684400 (30 µM) showed agonist activity at GPR65 at pH 8.40 when the receptor was not activated by protons; no further activation was seen when GPR65 was activated at lower pH conditions. The insert showed that ZINC 13684400 was inactive at control HEK293 T cells. BTB09089 was included as a positive control. (**e**) ZINC 62678696 showed antagonist activity when GPR65 was activated either by proton (pH 7.4) or BTB09089 (30 µM, at pH 8.40), but not at GPR68 receptors at pH 6.50. Results in Panels **d** – **e** (fold of basal) represent mean ± SEM from a minimum of 3 assays (at least a total of 16 measurements). (**f**) ZINC 13684400 showed dose-dependent agonist activity at GPR65 at pH 8.40, but not at GPR68 or control cells. BTB09089 was included as a positive control for GPR65. (**g**) ZINC 62678696 displayed concentration-dependent and GPR65-specific inhibition of cAMP production, but not at control cells or GPR65-transfected cells when cAMP production was stimulated by isoproterenol at pH 8.40. (**h**) ZINC 62678696 shifted BTB09089 concentration-response curves downward at pH 8.40 in a concentration-dependent manner. Allosteric parameters, derived from the

43

allosteric operational model, are inserted into the figure. Results in Panels **f** – **h** (fold of basal) represent mean ± SEM from 3 independent assays, each in triplicate or quadruplicate, and were analyzed in Graphpad Prism using the built-in 4-parameter logistic function (**f** and **h**). (**i**) Ternary complex model illustrating the allosteric effects of ZINC 62678696 on BTB09089.

1.4 Discussion

Here, a combined empirical and structure-based approach has discovered potent PAMs at the orphan GPCR GPR68, with essentially the same strategy revealing an allosteric agonist and NAMs for GPR65, supporting the generality of the approach. Additionally, the empirical screen was exemplified against 24 poorly annotated and orphan GPCRs. This study thus has implications both for deorphanizing the "dark matter" of the GPCR-ome—the 38% of non-olfactory GPCR targets whose ligands and function are unknown—and for explicating the physiology and targeting of the orphan GPCRs GPR68 and GPR65. Methodologically, the combination of a small empirical library screen with a much larger computational screen—followed by the testing of compounds arising from the computational screen—may find wide application for orphan receptors. Whereas truly high-throughput screens are impractical for targets of unknown function, low- to moderate-throughput screens are often feasible. Although the hits that emerge from such a screen may not be suitable as functional probes, they can anchor computational screens for compounds that are probe-like. Correspondingly, we would not ordinarily expect a docking screen to succeed against models of a target that shares only 29%

sequence identity with its nearest template[65]. By calculating several thousand candidate models, and insisting that the relevant ones are those that prioritize the hits from the empirical screen over the inactive molecules, one can distinguish functionally relevant receptor structures from decoys. The ligands that ultimately emerged are not only specific for the relevant target, they are active *in vivo*. This *in vivo* activity, and several canonical *in vitro* readouts of GPR68 activity, support the notion that the new molecules can act as chemical probes for the function of GPR68.

Both the empirical yeast growth and the docking screening strategies may have broad applicability to orphan GPCRs. The empirical screening channels the many possible functions of an orphan receptor into a single cellular readout and, as we demonstrate, can be applied against large numbers of orphan GPCRs in a parallel fashion. Still, one cannot expect to discover an optimized molecule from a small library; indeed, the focus on active drugs ensures that they will retain strong off-target effects, reducing their utility as chemical probes. The structure-based docking screen interrogates a library that is 1000- to 10,000-fold larger, thus increasing the chance of finding efficacious and specific ligands, as well as guiding structure-based optimization (as in the design and synthesis of compound **C2**). As this involves homology modeling to templates that often have low sequence identity to the orphan target, it is crucial that the models distinguish the ligands emerging from the empirical screen over those molecules found to be inactive[66]. With this constraint, the docking screen was successful: 29% of the docking-prioritized molecules tested against GPR68 were active, which is comparable to

docking campaigns against GPCR crystal structures[67-69]. Because these represented families of related molecules, they were readily optimized for efficacy.

The GPR68-specific activity of these molecules is consistent with the models that led to their discovery, as they do not share any properties with ligands of the template CXCR4 receptor, but rather complement particular features of the modeled GPR68. Consistently with this, neither benzodiazepines such as lorazepam, nor ogerin, have activity at the CXCR4 receptor (**Supplementary Figure 27**). Moreover, the activity of these GPR68 PAMs is consistent with the modeling, in which the ligands were docked to a site that overlaps with the known allosteric site of muscarinic receptors, for example (**Figure 2**).

Deorphanizing a receptor can also illuminate it as an off-target for known drugs. The observation that lorazepam and its primary metabolite, desmethyldiazepam, are PAMs at GPR68 may clarify several of the idiosyncratic effects of this widely-used drug. Like the other benzodiazepines, lorazepam and desmethyldiazepam are allosteric potentiators of the $GABA_A$ ion channel, explaining their use as sedative-hypnotics. However, lorazepam is unique among the benzodiazepines in its efficacy against catatonia, an effect that may be explained by its actions on "as yet uncharacterized benzodiazepine receptors"[70]. It is conceivable that GPR68 has a role in this efficacy, as both lorazepam and its metabolite reach low micromolar concentrations in the human plasma during treatment[71].

Pharmacologically, the most unexpected observation to emerge from this study was the activity of GPR68 in learning and memory. Previous studies in GPR68 KO mice revealed only modest phenotypic changes[25,31,39], without reported roles in higher brain function, even though GPR68 is most highly expressed in the brain. Here, we were able to elicit a corresponding transient and reversible response with a GPR68 tool molecule, ogerin, and validate its activity via chemical-genetic epistasis. Thus, administration of the GPR68-selective modulator ogerin reduced learning and contextual-based fear memory in wild-type but not GPR68 KO mice, implicating GPR68 in learning and memory.

As with all computationally-intensive exercises, certain caveats bear airing. Whereas the combination of an empirical and a computational screen is apparently productive for other orphan receptors, it is unlikely to work for all. GPCRs that are poorly expressed or non-functional in yeast or transfected cells (such as HEK293-T cells) will be problematic, and even with a larger library of reagents and tools than employed here, there will be some orphans that simply do not recognize any of the molecules screened in what will remain relatively small libraries. Even with the constraint of identifying active molecules from the empirical screen, some orphans will bear too little similarity to templates of known structure to support accurate model building. Likely those that do work will demand cycles of testing and optimization, which was crucial for both GPR65 and GPR68, where initial docking hits, though active, were often weak, and where new experimental information, such as the apparently allosteric interaction between the two GPR65 ligands, BTB09089 and ZINC 62678696, led to modifications of the models.

These cautions should not obscure the key observations from this study—that combining empirical with structure-based screening led to probe molecules that reveal the function of GPR68 in MAP kinase activation and in context-dependent fear conditioning. Ogerin, in particular, is not only specific for GPR68, and inactive at its most highly-related targets, but retains activity *in vivo* due to its ability to induce large allosteric effects. Though ogerin may be further optimized for allosteric activity, as we have shown with compound **C2**, it is itself already useful to elucidate the role of GPR68 in pharmacology and disease. Most immediately, the finding that ogerin potentiates GPR68 activation and downstream MAP kinase pathways, and previous observations that this receptor mediates airway inflammation (see **Introduction**), enables targeted campaigns for GPR68 functionally-selective PAMs that may reduce inflammatory responses, with potential roles in asthma and other obstructive lung diseases. As PAMs, these compounds would only be active during the clinically relevant spatial and temporal conditions: an asthmatic attack, in which the cells are stressed and placed under acidic conditions. Correspondingly, the role of GPR68 in anxiety offers a new route to treating this disease, and related CNS disorders, an area of pharmacology increasingly in need of new therapeutic modalities[72-75].

The expansion of this strategy to the discovery of allosteric modulators of a second orphan receptor, GPR65, supports its application to illuminating the function of the "dark matter" of the genome, that still large area of pharmacology where targets are known, but function is hidden.

1.5 Methods

I. Homology modeling

The alignment for the construction of the GPR68 models was generated using PROMALS3D, and homology models were built with MODELLER 9v8[76], using the crystal structure of the chemokine CXCR4 receptor (PDB ID: 3ODU) as the template (**Supplementary Figure 3**). This alignment was also used to generate 500 models of GPR65 directly from the final GPR68 model. The initial alignment included both human and mouse sequences of GPR68, as well as those of its closest homolog, GPR4. These were aligned against the whole human C-X-C chemokine receptor family. The alignment was manually edited to: remove the N- and C-termini that extended past the template structure, remove the engineered T4 lysozyme, and create different alignments of the flexible and non-conserved second extracellular loop (the final result is given in the provided alignment, **Supplementary Figure 6**). 407 models were built directly based on the CXCR4 crystal structure, using MODELER-9v8[76], while five more were built from each of 580 elastic network models (ENMs), produced by the program 3K-ENM,[77] for a total of 3,307 models built during each iterative round of model refinement. Models with constraints between pairs of extracellular His residues (His17-His169, His17-His269, His17-His84, His84-His169) to mimic the inactive state of the protein were generated by enforcing a distance constraint of 2.7 Å between the imidazole nitrogens, with a standard deviation of 0.1 Å.

II. Model evaluation

Prior to docking, the second extracellular loop (EL2), between residues 161-177, was removed from each GPR68 model. Models were ranked on the basis of prioritizing active benzodiazepines (lorazepam and desmethyldiazepam) over the rest of the inactive NCC library that was used in the yeast screen, as well as over property-matched decoys. In addition, the docked pose of lorazepam had to form a hydrogen bond from its N-H group to a polar side-chain in GPR68. Five different sites were sampled for possible lorazepam binding, based on the locations of the co-crystallized CXCR4 small molecule antagonist 1T1t (in PDB code: 3ODU), cyclic peptide CVX15 (in PDB code 3OE0), and the positions of the biogenic amines crystallized with the β_2-adrenergic receptor (PDB code: 2RH1) and the dopamine D_3 receptor (PDB code: 3PBL). The entire NCC library was docked to each of the five sub-sites for several rounds of iterative binding site refinement. In each round, the top-ranked models were examined for a binding pose that made hydrophobic and electrostatic interactions with the receptor, including the key N-H hydrogen bond. Residues within 6 Å of the lorazepam pose were minimized around the docked ligand with PLOP[47]. The NCC library was then re-docked into this optimized binding site for each model. This refinement continued for several cycles until the top-ranked models all converged to the same lorazepam pose. Once the final model was chosen, we built the EL2 back onto the receptor using MODELLER 9v8[74] and optimized 1,000 different EL2 conformations around the lorazepam pose with PLOP. Finally, we docked the NCC library back into these

1,000 different EL2-GPR68 structures and chose a final model that retained the previous pose and prioritized the active compounds over the inactives. The GPR65 model was generated similarly, using the pose of BTB09089 as the primary selection criterion, although in this case the EL2 was always present. To determine the ternary complex model of ZINC 62678696 and BTB09089, ZINC 62678696 was docked to the putative binding site in the GPR65 model with BTB09089 present. Then, both ligands were minimized with PLOP. Next, the side chains of the GPR65 binding pocket were allowed to relax, and, finally, BTB09089 and ZINC 62678696 were simultaneously minimized again with PLOP.

III. Virtual screens

We used DOCK3.6 to screen the ZINC database (**Results**). The flexible ligand sampling algorithm in DOCK3.6 superimposes atoms of the docked molecule onto binding site matching spheres, which represent favorable positions for individual ligand atoms. Forty-five matching spheres were used, using the previous refinement round's pose of lorazepam. The degree of ligand sampling is determined by the bin size, bin size overlap, and distance tolerance, set at 0.4 Å, 0.1 Å, and 1.5 Å, respectively, for both the matching spheres and the docked molecules. The complementarity of each ligand pose was scored as the sum of the receptor-ligand electrostatic and van der Waals' interaction energies, and corrected for context-dependent ligand desolvation. Partial charges from the united-atom AMBER force field were used for all receptor atoms; ligand charges and initial solvation energies were calculated using AMSOL[76,77]

(http://comp.chem.umn.edu/amsol/). The best-scoring conformation of each docked molecule was subjected to 100 steps of rigid-body minimization.

IV. Selection of potential ligands for testing

We docked the approximately 3.1 million commercially available molecules of the lead-like subset of the ZINC database to the final GPR68 and GPR65 models. The full hit list was automatically filtered to remove molecules that possess high-internal-energy, non-physical conformations, which are not well modeled by our scoring function. The reported rankings reflect this filtering. From the top 0.1% (~3000 molecules) of the docked ranking list, 17 compounds were chosen for testing based on complementarity to the binding site and presence of electrostatic interactions with Glu160, Arg189, Tyr244, Tyr268, and His269, mimicking those observed with lorazepam. For GPR65, compounds were chosen based on complementarity to the binding site and similarity to the binding pose of BTB09089, interacting with Asp153, Arg187, Tyr272, and aromatic stacking with Trp70.

V. In silico lead profiling

To examine specificity and to discover other potential GPCR targets for the newly discovered GPR68 positive allosteric modulators (PAMs), we used the Similarity Ensemble Approach (SEA)[52,53], which compares individual ligands, and sets of ligands, to the ligand sets for multiple targets; two targets are related, or a particular ligand is predicted to modulate a target, if the ligands are related to

one another. Here, the query set was all of the new GPR68 PAMs, which was screened against either the 2,512 ligand-target sets with activity of 10 μM or better from the ChEMBL12 database[54] or against the Tocris Mini library.

VI. Receptor constructs and yeast growth assays

24 human GPCR plasmids (GPR1, GPR4, GPR15, GPR31, GPR39, GPR41, GR43, GPR45, GPR55, GPR57, GPR58, GPR62, GPR65, GPR68, GPR83, GPR84, GPR87, GPR88, GPR123, GPR132, GPR133, GPR157, GPR161, ADCYAP1R1) were obtained from cdna.org, subcloned into the multiple cloning site of the yeast high copy number plasmid p426GPD78 and were confirmed by full-length sequencing (Eton Bioscience, Durham, NC). The yeast strains used were kindly provided by Mark Pausch (Merck) and have been previously described[79] and used by us[42,80]- MPY578t (G_i yeast), MPY578q5 (G_q yeast) and MPY578s5 (G_s yeast) express chimeric G proteins in which the last five amino acids of the yeast G-alpha protein are replaced with their mammalian G_i, G_q or G_s homologues, respectively. These strains contain the HIS3 gene under the control of the FUS1 promoter. GPCR transformants in yeast were selected and maintained on synthetic defined (SD) media lacking uracil (Clontech). GPR68q indicates the GPR68 paired with G_q-yeast; while GPR4s indicates GPR4 paired with G_s-yeast, and similarly for the other GPCRs. The yeast screening assays were carried out as described previously[42]. Assays were set up in 96-well flat-bottom clear assay plates that contained 50 μl of test compound at 40 μM (final concentration of 10 μM, in triplicate) diluted in SD-His-Ura medium (Clontech), 50 μl of 3-amino-1,2,4-triazole (3-AT) at 4x

concentration diluted in SD-His-Ura medium (pH 5.4), and 100 μl of yeast cell suspension diluted in SD-His-Ura medium to a final OD600 of 0.02. Growth was at 30°C for 2 to 5 days. Before measurement of cell growth, cells were re-suspended by repeated gentle pipetting to ensure uniform suspension of cells. Cell growth was measured by absorbance at 600 nm in a microplate reader (POLARstar Omega, BMG Biotech). After culling of data from obviously contaminated wells, the OD_{600} values of each individual well were adjusted as follows: 100 x (OD_{600} of test well – OD_{600} of plate median value) to give % growth stimulation (positive values), or % growth inhibition (negative value) in the form of means ± SEM of three wells.

To measure and control constitutive activity or leaky HIS expression, each receptor-yeast combination was plated as above in the absence of ligand over a range of concentrations of 3-AT. Concentrations of 3-AT that showed moderate yeast growth (i.e., OD values of 0.2 to 0.6) after 2 days at 30°C were used in assays for drug screening. To measure concentration-dependent activity, various concentrations of cognate ligands diluted in SD-His-Ura medium were incubated with transformed yeast and appropriate concentrations of 3-AT for 2 days at 30°C.

VII. Site-directed mutagenesis

The GPR68 plasmid was obtained from cdna.org. Mutation of E160A, E160K, E160Q, R189L, R189M, and H269F were introduced with Agilent's QuickChange® site-directed mutagenesis kit and confirmed by sequencing. To tag the receptors for comparing receptor expression levels with immunoblotting, FLAG epitope tag was

inserted at the C-terminal of the GPR68 wild-type and mutant receptors, also using the QuickChange site-directed mutagenesis kit. Insertion was confirmed by sequencing.

VIII. Split-luciferase based cAMP reporter assay

GPR4, GPR65, and GPR68 plasmids were obtained from cdna.org. GPR68 mutations were made and confirmed as above. Receptor-mediated G_s activation was measured using a split-luciferase reporter assay (GloSensor cAMP assay from Promega). Briefly, HEK293-T cells were transiently cotransfected with receptor DNA and GloSensor cAMP reporter plasmid (GloSensor 7A). Transfected cells were plated in poly-L-Lys coated 384-well white clear-bottom cell culture plates in DMEM supplemented with 1% dialyzed FBS at a density of 15,000 cells per well in a total of volume of 40 µl for minimum of 6 hours. Before assays, culture medium was removed and cells were incubated with Luciferin (4 mM prepared in drug buffer, pH 8.4) for 90 min at 37°C. The drug buffer was made with 1x HBSS supplemented with 10 mM HEPES and 10 mM MES modified from [19]. TAPS was added to accommodate higher pH values for some assays; no difference was observed between different buffers under the same pH condition. Cells plated at pH 8.4 for 6 hours generated the same H^+ concentration-response curves as those plated at pH 7.4. To make individual pH solutions, the pH was adjusted with NaOH and measured at room temperature with a pH 211 Microprocessor pH meter (Hanna Instruments). To measure modulator activity under different pH conditions, modulator was mixed with pH solutions before adding to cells. To

achieve the goal that drug solutions were delivered at the correct pH values,

luciferin solution was removed from cell plates before addition of drug solutions at

predetermined pH values. To improve solubility for some hydrophobic compounds,

1 mg/ml BSA was added to drug solutions and it had no effect on H^+ concentration-

response curves. For G_s-protein activity (cAMP production), the cell plate was

usually incubated at room temperature for 20 minutes before being counted in a

luminescence counter. Results were analyzed in GraphPad Prism.

IX. Allosteric operational model and data analysis

To estimate allosteric parameters, results were fitted to allosteric operational

model[49,50] as indicated in the following equation:

$$Response = basal + (E_{max} - basal)\frac{\tau_A[A](K_B + \alpha\beta[B])^n}{([A]K_B + K_AK_B + K_A[B] + a[A][B])^n + (\tau_A[A](K_B + \alpha\beta[B]))^n}$$

Where:

　　1. Response is a measured activity in the form of RLUs (Relative Luminescence

Units) for cAMP production measurement. If results are normalized, the 'Response'

is either RLU in fold of basal or a percentage value.

　　2. E_{max} represents the maximal possible responses of the system and the value

was normally constrained to the maximal reading of the corresponding

experiment.

　　3. Basal is the baseline in the absence of testing ligand and is constrained to

the baseline of the corresponding experiment. If results were normalized to

percentage values, basal= 0 and E_{max} = 100%; if results were normalized to fold of basal, the 'Basal' is usually 1.0.

4. [A] and [B] stand for concentrations of the orthosteric and allosteric ligands, respectively. In this case, A is proton.

5. K_A and K_B are the equilibrium dissociation constants (aka binding affinity) of the orthosteric agonist proton (A) and allosteric modulator (B), respectively. To facilitate curve-fitting with the model, K_A is usually fixed to the binding affinity determined from traditional radioligand binding assays under the assumption that the experimentally derived binding affinity is not significantly different from the functional affinity under the condition for corresponding functional assay. Since proton binding affinity is not a measurable parameter in this assay system, the proton K_A is therefore constrained to the corresponding proton EC_{50} value in the absence of allosteric ligand under the assumption that the proton potency is not significantly different from its binding affinity when the cAMP production assay is carried out. Since protons are present at relevant concentrations at physiological pH values, for a proton receptor K_B is largely a fitting parameter without a clear physical meaning.

6. The τ_A is the orthosteric agonist proton efficacy parameter. Since allosteric modulators in this study showed no agonist activity, the allosteric modulator efficacy τ_B is therefore 0 and not included in the function.

7. The term n is the slope factor linking receptor occupancy to response. Steep slopes in this study indicated high cooperativity between proton binding and

receptor activation, probably reflecting the fact that the proton receptors operate within a narrow physiological pH range.

8. The allosteric parameter α defines the mutual effect between the orthosteric agonist A and the allosteric modulator B (α > 1 for increased affinity and α < 1 for reduced affinity); while β defines the allosteric effect on agonist efficacy (β > 1 for increased efficacy and β < 1 for reduced efficacy).

With K_A, Basal, and E_{max} constrained to corresponding values, parameters K_B, τ_A, α, β, and n are globally shared fitting parameters for a family of proton concentration-response curves in the absence and presence of increasing concentrations of a test allosteric modulator. With the above settings, most curves could be easily fitted to generate reasonable parameters. If Prism could not fit the curves but generated 'ambiguous fitting' results, the α value was then manually constrained to an initial fitting value and systematically changed with small increments or decrements until the highest stable high affinity value (K_B) was reached. For GPR65 and GPR68, K_B represents the allosteric binding affinity in the absence of protons, which is unmeasurable and thus has little physical meaning. The product of K_B over α ($K_B/α$) represents the binding affinity of an allosteric ligand in the presence of protons, which could be estimated experimentally. For convenience, we call $K_B/α$ the "Biochemical binding affinity, K_{bB}" (Supplementary Table 4) for an allosteric ligand in the presence of an orthosteric agonist (in this case H^+).

X. PI hydrolysis assay

HEK293-T cells were transfected for 24 hours and plated in polyL-Lys coated 96-well black clear bottom cell culture plates with DMEM supplemented with 10% FBS, at a density of 60,000 cells in 100 µl per well. After 5 hours, cells were washed with inositol-free DMEM once and labeled with ^3H-inositol (1 µCi/well) in inositol-free DMEM supplemented with 5% dialyzed FBS for overnight. On the assay day, labeling medium was removed and cells were washed once with assay buffer (1x HBSS, 10 mM HEPES, 10 mM MES, 20 mM LiCl, pH 8.4). To measure drug concentration responses, then cells were then incubated with drug solutions pH 8.4 for 20 min. To measure proton concentration-responses, the assay buffer was pre-adjusted to desired pH values and supplemented with 20 mM LiCl. To measure the effect of ogerin or its isomer ZINC 32547799 on proton concentration-responses, pH solutions were supplemented with 20 mM LiCl and 10 µM ogerin or ZINC 32547799. The premixed drug solutions were added to cells for 20 min incubation. At the end of incubation, drug solutions were removed and 40 µl per well of 50 mM ice-cold formic acid was added. After incubation at 4°C for 30 min, the acid extracts were transferred to polyethylene terephthalate 96-well sample plates (#1450-401, Perkin Elmer) and mixed with 75 µl (200 µg) YSi RNA binding beads (RPNQ0013, Perkin Elmer). The plate was sealed and further incubated at 4°C for 30 min before being counted on a TriLux MicroBeta counter. Results (cpm/well) were analyzed in Prism.

XI. Anti-HA immunoblots

HEK293 cells were transfected with either pcDNA3 vector containing an HA cassette within the multiple cloning site, or pcDNA3HA-GPR68 encoding human GPR68 with an N-terminal HA tag. Stable lines were generated by selection with 250 μg/ml G418, with >90% of cells expressing HA after 2 weeks as assessed by immunocytochemistry (not shown). Cells were plated into 12 well plates, grown to confluence, and media switched to plain Hams-F12 media, with pH adjusted to pH 8.0 or 7.4, for 1 hour. Cells were then stimulated with either vehicle, 50 μM ogerin, or 50 μM lorazepam for 10 min. Lysates were harvested and subjected to immunoblotting, with blots probed using primary antibodies against HA, total vasodilator-stimulated phosphoprotein (VASP), p-p42/p44, and ß-actin, and secondary antibodies conjugated with infrared fluorophores as per [81].

XII. Generation of GPR68 knockout (KO) mice

To generate GPR68 KO mice, a probe specific for the human GPR68 transcript was generated by PCR amplification of a 450 bp segment of the coding sequence of the final exon of *GPR68* using total placental RNA. The probe was used to identify a clone from a 129 mouse genomic lambda library. The genomic insert was subcloned and a restriction map generated using a panel of enzymes. The targeting construct for the *GPR68* locus consists of a PGK-1 promoter driven neomycin resistance cassette flanked by two arms of homology with the mouse *GPR68* locus. The longer arm of homology was generated using a 7,266 bp PstI fragment extending from the last intron to the beginning of the last exon. This exon contains the entire coding sequence of the *GPR68* gene. The 1335 bp shorter

arm was generated by PCR amplification and extends from the downstream end of the long arm into the 3' untranslated region of the gene. Homologous recombination of the targeting construct with the *GPR68* locus inserts the neomycin resistance cassette into codon 78 of the gene, thereby disrupting expression. Correctly targeted cell lines were identified by Southern blot analysis using a probe consisting of a 1496 bp PstI fragment immediately upstream of the long arm. This probe recognizes a 14,290 bp EcoRV fragment in the endogenous locus and a 7,855 bp fragment in the targeted locus. Genotyping is carried out by PCR with 3 primers. The common (5'GCA GAG GAA GCC CAC GCT GAT GTA3') and endogenous (5'TAA ACG GTA GCT GTG ATT ATT CAA3') primers generate a 516 bp PCR product from the endogenous locus, while the common and targeted (5'AAA TGC CTG CTC TTT ACT GAA GG3') primers generate a 465 bp product from the targeted locus. The chimeras were bred to C57BL/6J mice and pups carrying the mutant allele identified. After ten successive crosses of heterozygous animals to C57BL/6J mice, heterozygous mice were intercrossed and a congenic GPR68 -/- and C57BL/6J breeding colony established.

XIII. In vivo studies

Contextual and cue-dependent learning and memory were evaluated using a Near-Infrared Video Fear Conditioning system (MED Associates, St. Albans, VT). Test chambers (29 x 25 x 25 cm) had transparent walls and metal rod floors, and were enclosed in sound-attenuating boxes. The conditioned fear procedure had 3 phases: training, a test for contextual learning, and a test for cue-dependent

learning. Before each phase, mice were moved to a holding room adjacent to the test room and acclimated for at least 30 min. In the 8-min training phase, mice receive 3 pairings of a 30-sec, 90 dB, 5 kHz tone (the conditioned stimulus, CS) and a 2-sec, 0.6 mA foot shock (the unconditioned stimulus, US), in which the shock was presented during the last 2 sec of the tone. Context-dependent learning was evaluated 24 hours following the training phase. Mice were placed back into the original test chamber, and levels of freezing (immobility) were determined across a 5-min session, without the presence of the CS or US. 48 hours following the training phase, mice were evaluated for associative learning to the auditory cue (the CS) in a final 6-min session. The conditioning chambers were modified using a Plexiglas insert to change the wall and floor surface, and a novel odor (vanilla flavoring) was added to the sound-attenuating box. Baseline behavior was scored for 2 minutes, and then three 30-sec CS tones were presented across a 4 min period. Levels of freezing were automatically measured by the image tracking software (Med Associates, St Albans, VT). Freezing was defined as no movement (below the movement threshold) for 0.5 sec. To evaluate the effect of drug, strain-matched group of animals were given ogerin (10 mg/kg in 10% Tween 80 or saline) 30 min before the training.

Statistical analyses were performed after first assessing the normality of distributions of data sets. Comparisons between groups were made using unpaired t-tests. Welch's corrections were utilized when variances between groups were unequal. Comparisons between groups during conditioning, contextual and cued

memory tests were assessed using two-way ANOVA with p-value < 0.05 being significant.

1.6 Acknowledgments

1.7 Author Contributions

X.-P.H. subcloned GPR68 for yeast screening, made GPR68 mutants, designed, carried out cell-based screening assays, analyzed results, and wrote the paper. J. K. designed and developed homology models, carried out docking screens, analyzed results, and wrote the paper. W. K. K. set up and carried out yeast screening assays, analyzed results, and wrote the paper. H. Z. and M. S. F. designed, carried out *in vivo* fear-conditioning studies, analyzed results, and wrote the paper. M. S. F. and B. L. R. dubbed ZINC 67740571 "ogerin". B. H. K.

created the GPR68 KO mice. S. S. M., K. A. S., and V. N. carried out initial phenotypic characterization, analyzed results, and wrote the paper. X. C. and J. J. synthesized ZINC 32547799, ZINC 67740571 (ogerin), and ogerin analogues (compounds 33548 – 33561, C3 and C4) for functional assays and *in vivo* studies, and wrote the paper. T. J. M. carried out radioligand binding assays. A. J. prepared drug plates and plasmids for initial screening. R. B. P. and D. A. D. designed and carried out anti-HA immunoblot assays, analyzed results, and wrote the paper. S. W. designed primers, prepared Flag-tagged GPR68 wild-type and mutant plasmids, preformed anti-Flag Western blot assays, and analyzed results. T. K. analyzed results and wrote the paper. B. L. R. and B. K. S. coordinated and supervised the project, and with the other authors wrote the paper.

1.8 References

1.   Davenport, A. P. *et al*. International Union of Basic and Clinical Pharmacology. LXXXVIII. G protein-coupled receptor list: recommendations for new pairings with cognate ligands. *Pharmacol Rev* **65,** 967–86 (2013).

2.   Chung, S., Funakoshi, T. & Civelli, O. Orphan GPCR research. *Br J Pharmacol* **153 Suppl 1,** S339–46 (2008).

3.   Civelli, O. *et al*. G Protein–Coupled Receptor Deorphanizations. *Annu. Rev. Pharmacol. Toxicol*. **53,** 127–146 (2013).

4.   Knapp, S. *et al*. A public-private partnership to unlock the untargeted kinome. *Nat. Chem. Biol*. **9,** 3–6 (2013).

5.   Ferguson, F. M. *et al*. Targeting Low-Druggability Bromodomains: Fragment Based Screening and Inhibitor Design against the BAZ2B Bromodomain. *J. Med. Chem*. **56,** 10183–10187 (2013).

6.   Matzuk, M. M. *et al*. Small-Molecule Inhibition of BRDT for Male Contraception. *Cell* **150,** 673–684 (2012).

7.   Leung, D., Hardouin, C., Boger, D. L. & Cravatt, B. F. Discovering potent and selective reversible inhibitors of enzymes in complex proteomes. *Nat. Biotechnol*. **21,** 687–691 (2003).

8.   Edwards, A. M. *et al*. Too many roads not taken. *Nature* **470,** 163–5 (2011).

9.   Allen, J. A. & Roth, B. L. Strategies to discover unexpected targets for drugs active at G protein-coupled receptors. *Annu Rev Pharmacol Toxicol* **51,** 117–44 (2011).

10. Zhao, S. *et al*. Discovery of new enzymes and metabolic pathways by using structure and genome context. *Nature* **502,** 698–702 (2013).

11. Hermann, J. C. *et al*. Structure-based activity prediction for an enzyme of unknown function. *Nature* **448,** 775–779 (2007).

12. Favia, A. D., Nobeli, I., Glaser, F. & Thornton, J. M. Molecular Docking for Substrate Identification: The Short-Chain Dehydrogenases/Reductases. *J. Mol. Biol*. **375,** 855–874 (2008).

13. Allen, J. J. *et al*. A semisynthetic epitope for kinase substrates. *Nat Methods* **4,** 511–6 (2007).

14. Bodenmiller, B. *et al*. Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Sci Signal* **3,** rs4 (2010).

15. Ong, S.-E. *et al*. Identifying the proteins to which small-molecule probes and drugs bind in cells. *Proc. Natl. Acad. Sci*. **106,** 4617–4622 (2009).

16. Kroeze, W. K., Sheffler, D. J. & Roth, B. L. G-protein-coupled receptors at a glance. *J Cell Sci* **116,** 4867–9 (2003).

17. Seuwen, K., Ludwig, M. G. & Wolf, R. M. Receptors for protons or lipid messengers or both? *J Recept Signal Transduct Res* **26,** 599–610 (2006).

18. Ludwig, M. G. *et al*. Proton-sensing G-protein-coupled receptors. *Nature* **425,** 93–8 (2003).

19. Mogi, C. *et al*. Sphingosylphosphorylcholine antagonizes proton-sensing ovarian cancer G-protein-coupled receptor 1 (OGR1)-mediated inositol phosphate production and cAMP accumulation. *J Pharmacol Sci* **99,** 160–7 (2005).

20.   Li, J. *et al*. Ovarian cancer G protein coupled receptor 1 suppresses cell migration of MCF7 breast cancer cells via a Galpha12/13-Rho-Rac1 pathway. *J Mol Signal* **8,** 6 (2013).

21.   Singh, L. S. *et al*. Ovarian Cancer G Protein–Coupled Receptor 1, a New Metastasis Suppressor Gene in Prostate Cancer. *J. Natl. Cancer Inst.* **99,** 1313–1327 (2007).

22.   Regard, J. B., Sato, I. T. & Coughlin, S. R. Anatomical profiling of G protein-coupled receptor expression. *Cell* **135,** 561–71 (2008).

23.   Schneider, J. W. *et al*. Coupling hippocampal neurogenesis to brain pH through proneurogenic small molecules that regulate proton sensing G protein-coupled receptors. *ACS Chem Neurosci* **3,** 557–68 (2012).

24.   Frick, K. K., Krieger, N. S., Nehrke, K. & Bushinsky, D. A. Metabolic acidosis increases intracellular calcium in bone cells through activation of the proton receptor OGR1. *J Bone Min. Res* **24,** 305–13 (2009).

25.   Li, H. *et al*. Abnormalities in Osteoclastogenesis and Decreased Tumorigenesis in Mice Deficient for Ovarian Cancer G Protein-Coupled Receptor 1. *PLoS ONE* **4,** e5705 (2009).

26.   Komarova, S. V., Pereverzev, A., Shum, J. W., Sims, S. M. & Dixon, S. J. Convergent signaling by acidosis and receptor activator of NF-kappaB ligand (RANKL) on the calcium/calcineurin/NFAT pathway in osteoclasts. *Proc Natl Acad Sci U A* **102,** 2643–8 (2005).

27.   Yang, M. *et al*. Expression of and role for ovarian cancer G-protein-coupled receptor 1 (OGR1) during osteoclastogenesis. *J Biol Chem* **281,** 23598–605 (2006).

28.   Kato, K. & Morita, I. Promotion of osteoclast differentiation and activation in spite of impeded osteoblast-lineage differentiation under acidosis: effects of acidosis on bone metabolism. *Biosci Trends* **7,** 33–41 (2013).

29.   Nakakura, T. *et al*. Deficiency of proton-sensing ovarian cancer G protein-coupled receptor 1 attenuates glucose-stimulated insulin secretion. *Endocrinology* **153,** 4171–80 (2012).

30.   Russell, J. L. *et al*. Regulated expression of pH sensing G Protein-coupled receptor-68 identified through chemical biology defines a new drug target for ischemic heart disease. *ACS Chem Biol* **7,** 1077–83 (2012).

31.   Mohebbi, N. *et al*. The proton-activated G protein coupled receptor OGR1 acutely regulates the activity of epithelial proton transport proteins. *Cell Physiol Biochem* **29,** 313–24 (2012).

32.   Tomura, H. *et al*. Prostaglandin I(2) production and cAMP accumulation in response to acidic extracellular pH through OGR1 in human aortic smooth muscle cells. *J Biol Chem* **280,** 34458–64 (2005).

33.   Chen, Y. J., Huang, C. W., Lin, C. S., Chang, W. H. & Sun, W. H. Expression and function of proton-sensing G-protein-coupled receptors in inflammatory pain. *Mol Pain* **5,** 39 (2009).

34.   Ichimonji, I. *et al*. Extracellular acidification stimulates IL-6 production and Ca(2+) mobilization through proton-sensing OGR1 receptors in human airway smooth muscle cells. *Am J Physiol Lung Cell Mol Physiol* **299,** L567–77 (2010).

35.   Liu, J. P. *et al*. Ovarian cancer G protein-coupled receptor 1-dependent and -independent vascular actions to acidic pH in human aortic smooth muscle cells. *Am J Physiol Heart Circ Physiol* **299,** H731–42 (2010).

36.   Matsuzaki, S. *et al*. Extracellular acidification induces connective tissue growth factor production through proton-sensing receptor OGR1 in human airway smooth muscle cells. *Biochem Biophys Res Commun* **413,** 499–503 (2011).

37.   Saxena, H. *et al*. The GPCR OGR1 (GPR68) mediates diverse signalling and contraction of airway smooth muscle in response to small reductions in extracellular pH. *Br J Pharmacol* **166,** 981–90 (2012).

38.   Wang, J., Sun, Y., Tomura, H. & Okajima, F. Ovarian cancer G-protein-coupled receptor 1 induces the expression of the pain mediator prostaglandin E2 in response to an acidic extracellular environment in human osteoblast-like cells. *Int J Biochem Cell Biol* **44,** 1937–41 (2012).

39.   Aoki, H. *et al*. Proton-sensing ovarian cancer g protein-coupled receptor 1 on dendritic cells is required for airway responses in a murine asthma model. *PLoS One* **8,** e79985 (2013).

40.   Mogi, C., Nakakura, T. & Okajima, F. Role of extracellular proton-sensing OGR1 in regulation of insulin secretion and pancreatic beta-cell functions [Review]. *Endocr J* (2013). at <http://www.ncbi.nlm.nih.gov/pubmed/24088601>

41.   Okajima, F. Regulation of inflammation by extracellular acidification and proton-sensing GPCRs. *Cell Signal* **25,** 2263–71 (2013).

42.   Dong, S., Rogan, S. C. & Roth, B. L. Directed molecular evolution of DREADDs: a generic approach to creating next-generation RASSLs. *Nat Protoc* **5,** 561–73 (2010).

43.   Samama, P., Cotecchia, S., Costa, T. & Lefkowitz, R. J. A mutation-induced activated state of the beta 2-adrenergic receptor. Extending the ternary complex model. *J. Biol. Chem.* **268,** 4625–4636 (1993).

44.   Holst, B. *et al.* GPR39 Signaling Is Stimulated by Zinc Ions But Not by Obestatin. *Endocrinology* **148,** 13–20 (2007).

45.   Yasuda, S. *et al.* Isolation of Zn2+ as an endogenous agonist of GPR39 from fetal bovine serum. *J. Recept. Signal Transduct. Res.* **27,** 235–246 (2007).

46.   Mysinger, M. M. & Shoichet, B. K. Rapid Context-Dependent Ligand Desolvation in Molecular Docking. *J. Chem. Inf. Model.* **50,** 1561–1573 (2010).

47.   Jacobson, M. P., Friesner, R. A., Xiang, Z. & Honig, B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *J. Mol. Biol.* **320,** 597–608 (2002).

48.   Meng, E. C., Gschwend, D. A., Blaney, J. M. & Kuntz, I. D. Orientational sampling and rigid-body minimization in molecular docking. *Proteins* **17,** 266–78 (1993).

49.   Leach, K., Sexton, P. M. & Christopoulos, A. Allosteric GPCR modulators: taking advantage of permissive receptor pharmacology. *Trends Pharmacol Sci* **28,** 382–9 (2007).

50.   Christopoulos, A. & Kenakin, T. G protein-coupled receptor allosterism and complexing. *Pharmacol. Rev.* **54,** 323–374 (2002).

51.   Mistry, S. N. *et al.* Synthesis and pharmacological profiling of analogues of benzyl quinolone carboxylic acid (BQCA) as allosteric modulators of the M1 muscarinic receptor. *J Med Chem* **56,** 5151–72 (2013).

52.   Keiser, M. J. *et al.* Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* **25,** 197–206 (2007).

53.   Hert, J., Keiser, M. J., Irwin, J. J., Oprea, T. I. & Shoichet, B. K. Quantifying the Relationships among Drug Classes. *J. Chem. Inf. Model.* **48,** 755–765 (2008).

54.   Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40,** D1100–D1107 (2011).

55.   Kalk, P. *et al.* The adenosine A1 receptor antagonist SLV320 reduces myocardial fibrosis in rats with 5/6 nephrectomy without affecting blood pressure. *Br. J. Pharmacol.* **151,** 1025–1032 (2007).

56.   Thompson, S.-A., Wingrove, P. B., Connelly, L., Whiting, P. J. & Wafford, K. A. Tracazolate Reveals a Novel Type of Allosteric Interaction with Recombinant γ-Aminobutyric AcidA Receptors. *Mol. Pharmacol.* **61,** 861–869 (2002).

57.   Gravius, A., Barberi, C., Schäfer, D., Schmidt, W. J. & Danysz, W. The role of group I metabotropic glutamate receptors in acquisition and expression of contextual and auditory fear conditioning in rats - a comparison. *Neuropharmacology* **51,** 1146–1155 (2006).

58.   Fornari, R. V., Moreira, K. M. & Oliveira, M. G. Effects of the selective M1 muscarinic receptor antagonist dicyclomine on emotional memory. *Learn. Mem. Cold Spring Harb. N* **7,** 287–292 (2000).

59.   Daumas, S. *et al*. Transient activation of the CA3 Kappa opioid system in the dorsal hippocampus modulates complex memory processing in mice. *Neurobiol. Learn. Mem.* **88,** 94–103 (2007).

60.   Stiedl, O., Misane, I., Spiess, J. & Ogren, S. O. Involvement of the 5-HT1A receptors in classical fear conditioning in C57BL/6J mice. *J. Neurosci. Off. J. Soc. Neurosci.* **20,** 8515–8527 (2000).

61.   Phillips, R. G. & LeDoux, J. E. Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behav. Neurosci.* **106,** 274–285 (1992).

62.   Onozawa, Y. *et al*. Activation of T cell death-associated gene 8 regulates the cytokine production of T cells and macrophages in vitro. *Eur J Pharmacol* **683,** 325–31

63.   Paton, W. D. M. & Rang, H. P. The Uptake of Atropine and Related Drugs by Intestinal Smooth Muscle of the Guinea-Pig in Relation to Acetylcholine Receptors. *Proc. R. Soc. Lond. B Biol. Sci.* **163,** 1–44 (1965).

64.   Kenakin, T., Jenkinson, S. & Watson, C. Determining the Potency and Molecular Mechanism of Action of Insurmountable Antagonists. *J. Pharmacol. Exp. Ther.* **319,** 710–723 (2006).

65.   Mysinger, M. M. *et al*. Structure-based ligand discovery for the protein–protein interface of chemokine receptor CXCR4. *Proc. Natl. Acad. Sci.* **109,** 5517–5522 (2012).

66.   Evers, A. & Klebe, G. Successful Virtual Screening for a Submicromolar Antagonist of the Neurokinin-1 Receptor Based on a Ligand-Supported Homology Model. *J. Med. Chem.* **47,** 5381–5392 (2004).

67.   De Graaf, C. *et al*. Crystal Structure-Based Virtual Screening for Fragment-like Ligands of the Human Histamine H1 Receptor. *J. Med. Chem.* **54,** 8195–8206 (2011).

68.   Katritch, V. *et al*. Structure-Based Discovery of Novel Chemotypes for Adenosine A2A Receptor Antagonists. *J. Med. Chem.* **53,** 1799–1809 (2010).

69.   De Graaf, C., Rein, C., Piwnica, D., Giordanetto, F. & Rognan, D. Structure-Based Discovery of Allosteric Modulators of Two Related Class B G-Protein-Coupled Receptors. *ChemMedChem* **6,** 2159–2169 (2011).

70.   Pompeia, S., Manzano, G. M., Tufik, S. & Bueno, O. F. What makes lorazepam different from other benzodiazepines? *J Physiol* **569,** 709; author reply 710 (2005).

71.   Greenblatt, D. J. *et al*. Clinical pharmacokinetics of lorazepam. I. Absorption and disposition of oral 14C-lorazepam. *Clin. Pharmacol. Ther.* **20,** 329–341 (1976).

72.   Amara, S. G., Grillner, S., Insel, T., Nutt, D. & Tsumoto, T. Neuroscience in recession? *Nat Rev Neurosci* **12,** 297–302 (2011).

73.   Schoepp, D. D. Where will new neuroscience therapies come from? *Nat Rev Drug Discov* **10,** 715–6 (2011).

74.   Griebel, G. *et al*. Further evidence for the sleep-promoting effects of 5-HT(2)A receptor antagonists and demonstration of synergistic effects with the hypnotic, zolpidem in rats. *Neuropharmacology* **70,** 19–26 (2013).

75.   Nutt, D. J. & Attridge, J. CNS drug development in Europe - Past progress and future challenges. *Neurobiol Dis* **61,** 6–20 (2014).

76.   Eswar, N. *et al*. in *Current Protocols in Protein Science* (John Wiley & Sons, Inc., 2001). at <http://onlinelibrary.wiley.com/doi/10.1002/0471140864.ps0209s50/abstract>

77.   Yang, Q. & Sharp, K. A. Building alternate protein structures using the elastic network model. *Proteins Struct. Funct. Bioinforma.* **74,** 682–700 (2009).

78.   Li, J., Zhu, T., Cramer, C. J. & Truhlar, D. G. New Class IV Charge Model for Extracting Accurate Partial Charges from Wave Functions. *J. Phys. Chem. A* **102,** 1820–1831 (1998).

79.   Chambers, C. C., Hawkins, G. D., Cramer, C. J. & Truhlar, D. G. Model for Aqueous Solvation Based on Class IV Atomic Charges and First Solvation Shell Effects. *J. Phys. Chem.* **100,** 16385–16398 (1996).

80.   Mumberg, D., Muller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156,** 119-22 (1995).

81.   Erlenbach, I. *et al*. Functional expression of M(1), M(3) and M(5) muscarinic acetylcholine receptors in yeast. *J Neurochem* **77,** 1327–37 (2001).

82.   Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S. & Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proc Natl Acad Sci U A* **104,** 5163–8 (2007).

83.   Horvat, S. J. *et al*. A-kinase anchoring proteins regulate compartmentalized cAMP signaling in airway smooth muscle. *FASEB J* **26,** 3670–9 (2012).

Gloss to Chapter 2

Now that it seemed like that even under the toughest of circumstances we could create a validated model of a receptor producing magic bullets - what could we do with them? What of these magic shotguns? Drugs often have an unknown and under-appreciated polypharmacology that arises fortuitously, but what if a molecule with a specific activity profile could be automatically generated? This is the question that Chapter 2 intended to answer.

A system for designing polypharmacology could benefit the field in two ways. The first is that, if certain targets are desired or known to be beneficial to ameliorate some therapeutic indication, a molecule that is more specific and optimized to not have any toxic off-target effects could be designed. On the other hand, if pan-GPCR assays are not available or are not informative enough to explain the unique clinical effects of some drug, different molecules representing different hypotheses and combinations of active targets can be identified in order to tease out the contributions of individual receptors.

Here, we were concerned with a set of receptors involved in the treatment of psychosis and depression. Antipsychotic drugs concurrently modulate multiple neurotransmitter systems, exerting their therapeutic effect primarily through various bioaminergic GPCR targets, including the serotonin 5-HT$_{2A}$ receptor and the dopamine D$_2$ receptor. Weight gain, a common and serious side effect of most antipsychotic drugs, is thought to be mediated by an unwanted interaction with the histamine H$_1$ receptor, a closely related, and in this case, anti-target GPCR.

Similarly, selective antagonists for the κ-opioid, over the μ-opioid, receptor present a novel anti-depressant mechanism. These stories have different settings, as one involves 3 different receptors and the other 2 closely related subtypes, but both have the same conclusion - that it is difficult to obtain selective molecules via docking, due to incorrect treatment of anti-target receptor flexibility. The magnitude of this flexibility may be inherent to individual receptors, but more test cases are needed.

These narratives show that it is possible to discriminate selective molecules by virtual screens, whereas ligand-based methods fail. In addition, the inability to create more selective analogs supports the claim that the original molecules were chosen for their correctly predicted selectivities. This technique, when combined with the ability to model even the most opaque and intractable of GPCRs, should facilitate the detection of molecules that are tailored for a specific purpose and polypharmacology.

Chapter 2:

Structure-based methods for the rational design of polypharmacology in G protein-coupled receptors

Joel Karpiak,[1]* Dahlia R. Weiss,[1]* Maria F. Sassano,[2]* Xi-Ping Huang,[2] Brian K. Shoichet,[1,3†] and Bryan L. Roth[2†]

[1]Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, CA

[2]Department of Pharmacology and National Institute of Mental Health Psychoactive Drug Screening Program (NIMH PDSP), School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC

[3]Faculty of Pharmacy & Ontario Institute for Cancer Research, University of Toronto, Toronto, Canada

*These authors contributed equally to this work

†Corresponding authors (BKS and BLR)

2.1 Abstract

There is a growing appreciation that many drugs achieve efficacy by acting on multiple targets, however, the discovery of such drugs is almost always serendipitous and the mechanism of action is usually understood only retrospectively. Here, for the first time, we explored the use of structure-based virtual screening as a rational method to directly find compounds that bind to multiple, pre-defined G protein-coupled receptor (GPCR) targets and with selectivity against closely related GPCR anti-targets. An initial screen for binders of the psychiatrically therapeutic targets, the dopamine D_2 (DRD2) and serotonin 5-HT_{2A} (HTR2A) receptors, resulted in several selective compounds, with one high-affinity molecule showing 20-fold selectivity against the anti-target histamine H_1 receptor (HRH1). The screen produced a high rate of true positives, with 71% of the molecules tested binding to at least one of the on-targets, and 30% binding to both. However, the rate of false negatives was also very high, with 65% of the molecules binding to the off-target, including one with a binding affinity of 0.8 nM that is specific to HRH1. A second screen of the kappa opioid receptor (KOR) similarly identified a high number of true positives (41%) and false negatives (45%), but identified only 2 low-affinity compounds with selectivity against the mu opioid receptor (MOR). False negatives in both virtual screens were shown to be due to inadequate handling of receptor flexibility, which persisted in follow-up docking campaigns despite prospectively utilizing approaches such as ensemble and induced-fit docking to account for this problem. These examples involving multiple relevant GPCR targets illustrate the utility and caveats of molecular

docking to quickly and directly find molecules with a prescribed polypharmacology.

2.2 Introduction

The inherent complexity of G protein-coupled receptor (GPCR) signaling networks often works against the "one disease-one target" paradigm of drug discovery, as the efficacy of a drug may depend on interaction with many therapeutic targets (Roth, Sheffler et al. 2004; Fishman and Porter 2005; Frantz 2005; Urban, Clarke et al. 2007; Hopkins 2008; Metz and Hajduk 2010). Conversely, promiscuous actions of a drug on closely related GPCR targets can lead to adverse drug reactions. Many marketed drugs do act through multiple targets, but this mechanism of action has been for the most part serendipitous and usually discovered after-the-fact.

The rational design of GPCR drugs with predefined multi-target specificities has rarely been attempted (Maynard, Bratton et al. 1997; Merlos, Giral et al. 1997; Murugesan, Tellew et al. 2002; Morphy and Rankovic 2005; Morphy and Rankovic 2007; Mohr, Trankle et al. 2010). A recent method relying only on two-dimensional ligand structures used machine learning to automatically and adaptively optimize ligands with desired polypharmacological binding properties (Besnard, Ruda et al. 2012). While this method shows great promise, it necessitates a great deal of structure-activity data for the ligand-target set, and requires a starting scaffold for optimization, transforming only via traditional medicinal chemistry synthetic routes. This limits the amount of chemical space that can be sampled.

Recent advances in GPCR X-ray structure determination have led to a surge in structure-based ligand design, with virtual screens universally finding new and potent ligands for the diverse targeted GPCRs (Kolb, Rosenbaum et al. 2009; Carlsson, Yoo et al. 2010; Katritch, Jaakola et al. 2010; de Graaf, Kooistra et al. 2011; Mysinger, Weiss et al. 2012). The use of virtual screening does not necessitate any prior chemical knowledge, and often leads to novel chemical scaffolds. Here, we use a subset of psychiatrically-relevant GPCRs in order to test, for the first time, the applicability of docking as a method to design GPCR ligands with predefined specificities for multiple targets and anti-targets.

Antipsychotic drugs concurrently modulate multiple neurotransmitter systems, exerting their therapeutic effect primarily through various bioaminergic GPCR targets, including the serotonin 5-HT$_{2A}$ receptor (HTR2A) and the dopamine D$_2$ receptor (DRD2). Weight gain, a common and serious side effect of most antipsychotic drugs, is thought to be mediated by an unwanted interaction with the histamine H$_1$ receptor (HRH1), a closely related bioaminergic GPCR (Kroeze, Hufeisen et al. 2003; Kim, Huang et al. 2007). Despite decades of intensive efforts, the development of "selectively promiscuous" antipsychotic drugs has been extremely difficult and is only recently being appreciated. For example, clozapine, a well-known, highly-prescribed second-generation antipsychotic, has a higher binding affinity for HRH1 than for the therapeutic targets HTR2A and DRD2 (1.2 nM versus 5.4 nM and 256 nM, respectively) (Kroeze, Hufeisen et al. 2003). The relatively new and moderately selective drug ziprasidone is arguably the most selective FDA-approved antipsychotic, displaying affinities of 0.3 nM for HTR2A,

9.7 nM for DRD2, and 43 nM for HRH1, while lurasidone, a novel antipsychotic candidate, shows no affinity for HRH1 below 1 μM yet binds to HTR2A and DRD2 with affinities of 0.5 nM and 1.0 nM, respectively (Kroeze, Hufeisen et al. 2003; Ishibashi, Horisawa et al. 2010). Ziprasidone, however, has issues with blood-brain barrier permeability.

With the publication of the crystal structures of HRH1 and dopamine receptor D_3 (DRD3), this clinically relevant system served as a useful test case for direct design *of* polypharmacology in virtual screening hits. The DRD3 crystal structure, at 78% and 40% transmembrane sequence identity to DRD2 and HTR2A, respectively, is within the range to serve as an accurate template for successful virtual screening at models of both receptors (Carlsson, Coleman et al. 2011). All three bioaminergic receptors are well characterized by extensive medicinal chemistry, and yet HTR2A and DRD2 share only a 37% and 31% transmembrane sequence identity to HRH1, respectively. In our first virtual screen, we asked whether structure-based design methods are able to exploit sequence differences in the orthosteric binding site regions to create selectivity.

Subtype specificity, however, where target and anti-target sequence identities are typically very high, is known to be difficult to directly get from virtual screening (Kolb, Phan et al. 2012). A recent virtual screen targeting the M_2/M_3 subtype muscarinic acetylcholine receptors did not find selective compounds with respect to binding, but did uncover an M_3 subtype agonist that was functionally selective, with no activity at the M_2 subtype (Kruse, Weiss et al. 2013). In our second virtual screening campaign, examining a closely-related target and anti-

target, we docked to the crystal structures of the kappa and mu opioid receptors (KOR and MOR, respectively) to look for selective KOR ligands. KOR-selective ligands are psychiatrically relevant regardless of their functional efficacy; antagonists may serve as novel anti-depressants, and agonists could act as pain relievers that do not activate reward pathways implicated in addiction (Negri, Rives et al. 2013; Spetea, Asim et al. 2013).

Virtual screening has the unique ability to explore a wider amount of chemical space than a traditional high-throughput screen and can subsequently focus optimization efforts onto specific molecules that meet some predefined criteria. Although docking requires the structure of the desired targets and anti-targets, homology modeling makes many more GPCRs amenable to virtual screening campaigns, where novel hits can result directly from the screen. In this study, different subsets of therapeutically important GPCR targets offer practical tests to examine docking as a fast, automated method for the discovery of ligands with a predetermined polypharmacological profile.

2.3 Results

I. Homology modeling and docking to DRD2, HTR2A, and HRH1

First, we required building homology models of DRD2 and HTR2A and validating them with retrospective ligand docking. We used MODELLER9v8 (Sali and Blundell 1993) to generate 400 models for each receptor based on the DRD3 template (PDB ID 3PBL) (Chien, Liu et al. 2010). In order to assess the homology models' ability to recognize known ligands, we selected a set of 68 and 85 diverse, lead-like, high-

affinity HTR2A and DRD2 ligands, respectively, from the ChEMBL10 small molecule binding database (Gaulton, Bellis et al. 2011). Enrichment of both sets of ligands was measured against a property-matched background of decoys (DUD) (Mysinger, Carchia et al. 2012) and experimental non-binders from ChEMBL10 for all homology models of both receptors. To also minimize the likelihood of obtaining hits for the anti-target structure, HRH1, we docked an additional diverse set of 50 HRH1 ligands from ChEMBL10 and their property-matched decoys to the models. Final homology models were chosen for DRD2 and for HTR2A, showing high enrichment, a logAUC of 15 and 11, respectively, of the known ligands to the on-targets, and no enrichment of HRH1 ligands, with logAUC of -2.44 and 0.47 respectively. The anti-target HRH1 crystal structure showed a very high retrospective enrichment with a logAUC of 41 for the diverse set of 50 HRH1 ligands. The utility of this structure has previously been shown in prospective virtual screening, giving a 73% hit rate in a screen of novel, chemically-diverse fragments (de Graaf, Kooistra et al. 2011).

II. Virtual screening for aminergic receptor selectivity

Satisfied with the high retrospective enrichments of both the on- and anti-targets, we proceeded with a virtual screening campaign to identify molecules predicted to bind to HTR2A and DRD2 and not to bind to HRH1. We used DOCK 3.6 to virtually screen the approximately 3.1 million molecules in the lead-like subset of the ZINC database (Irwin and Shoichet 2005; Irwin, Sterling et al. 2012). We then selected molecules that ranked in the top 1% to both HTR2A and DRD2, of which there were

5,862 such molecules. To discourage choosing molecules that would bind to HRH1, we filtered out any compound having a Tanimoto coefficient (Tc) of 0.5 or greater to the top 50,000 ranked molecules docked to HRH1. We further insisted that no molecule had a Tc > 0.35 to any known HRH1 binder in the ChEMBL10 database, ensuring the dissimilarity of the docked hit list to known HRH1 ligands. This left a total of 354 molecules to consider, from which we selected 28 for experimental testing.

Of the 28 molecules tested, 17 (60%) displaced radiolabelled [^{3}H]ketanserin from HTR2A, 10 (35%) displaced radiolabelled [^{3}H]N-methylspiperone from DRD2, and 8 (29%) bound to both, with K_i < 10 μM (**Table 2.1, A.2.1**). We also directly identified a single specific, high-affinity compound, with binding affinities of 55 nM, 334 nM, and 1,144 nM for HTR2A, DRD2 and HRH1 respectively (compound **21**; **Table 2.1**). Docked poses to the three receptors are shown in **Figure 2.1A**. This represents a 21- and 3-fold selectivity, respectively, for the therapeutic targets over the anti-target. Although affinities are 10-fold worse than those of ziprasidone, they are well within the range of known antipsychotics. Four additional compounds were found with micro-molar potency for HTR2A and DRD2 and no affinity for HRH1 up to 10 μM (compounds 5, 14, 19 and 25; **Table 2.1**).

On the other hand, the rate of false positives was also high, with 16/28 molecules (57%) displacing radiolabelled [^{3}H]pyrilamine from HRH1, some with high affinity. The compound with the highest affinity had a K_i of 0.8 nM to HRH1 (compound 6, rank 326,040 to HRH1; **Table 2.1**), with no detectable affinity for DRD2 or HTR2A, representing the tightest binding compound found in a GPCR

virtual screen (Evers and Klabunde 2005; Becker, Dhanoa et al. 2006). Docked

poses for compound **6** are shown in **Figure 2.1B**. In fact, median affinities were

highest for the HRH1 hits (K_i = 3,430 nM, 4,276 nM and 1,042 nM, with median

ranks of 5,784, 2,097, and 1,030,000, for DRD2, HTR2A, and HRH1 respectively).
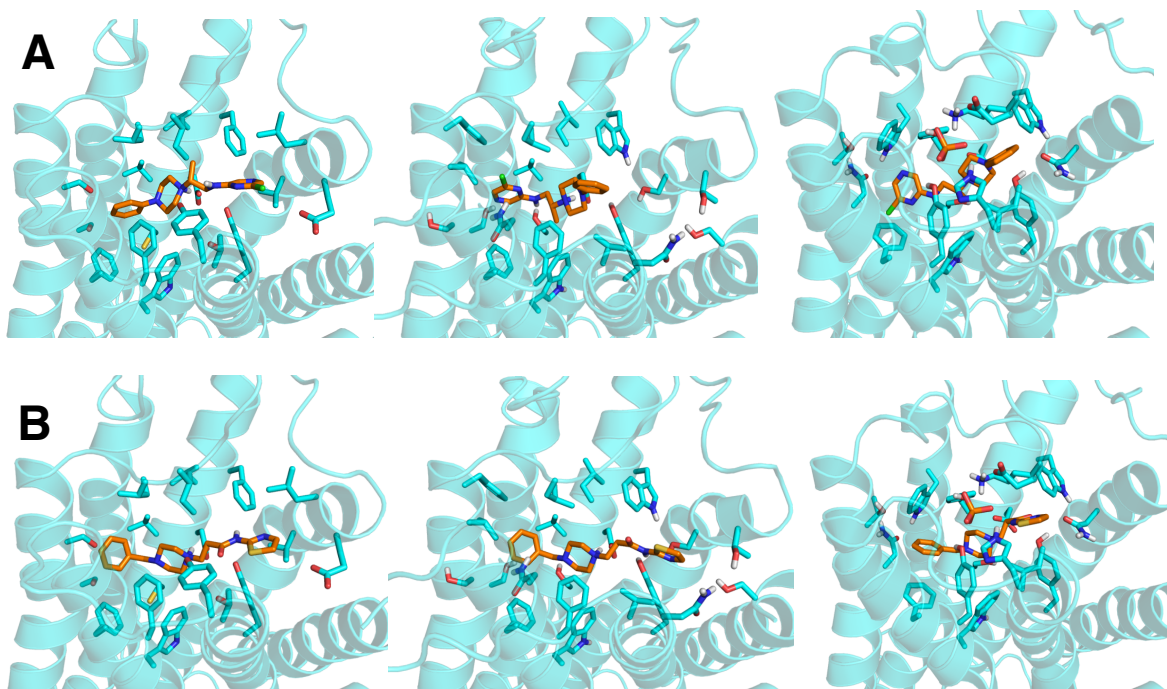
Figure 2.1 The HRH1, HTR2A and DRD2 binding sites. (**A**) Space-filling model of

the most selective compound, compound 21 (orange), docked to HTR2A (left), DRD2

(middle) and HRH1 (right) (receptors in cyan, HRH1 phosphate in orange). In HRH1,

the ligand is unable to make the critical charge-charge interaction with $D^{3.32}$. (**B**)

Space-filling model of the least selective compound, compound **6** (orange), docked

to HTR2A (left), DRD2 (middle) and HRH1 (right). In HRH1, the ligand is again unable

to make the critical charge-charge interaction with $D^{3.32}$. The compound, when

docked into the relatively smaller HRH1 binding site, was predicted not to bind

because of clashes with the receptor, but fit well into the larger HTR2A and DRD2 sites.

III. Virtual screening for opioid receptor selectivity

We next turned to a more extreme test of selectivity - that of discovering subtype-selective molecules for the KOR over the MOR. These receptors share over 60% sequence identity and feature large, solvent-exposed binding sites. In preliminary bench-marking studies, we selected for both the KOR and MOR a set of 60 diverse antagonists to again retrospectively measure enrichment of known binders over property-matched decoys. Unlike docking to the aminergics, for these receptors, the maximum logAUC achieved was 6.5, reflecting both the large diversity of ligands that are already annotated, as well as the large, solvent-exposed binding site's ability to somehow recognize and interact with all of them. As a result of the low-enrichment studies, confidence in prioritization of true positives over true negatives was not as high as for the aminergic campaign. We used DOCK 3.6 to again virtually screen the 3.1 million compound-lead-like subset of ZINC and selected the top 1% of molecules in the KOR screen to study further. We calculated the ratio of each compound's rank in the KOR screen to the compound's rank in the MOR screen, and the binding poses of the molecules with the largest rank ratio were visually inspected for binding site complementarity and favorable electrostatic interactions (**Methods**).

Of the 22 molecules tested, 9 specifically displaced radiolabelled [^{3}H]U69593 from KOR (41% hit rate; **Table 2.2; Table A.2.2, Figure 2.2**). Like the aminergic

screen, however, 7 of these (78%) also hit the MOR anti-target, displacing radiolabelled [^{3}H]DAMGO, with only 2 being specific to the KOR. Indeed, just as many compounds, 3, were actually specific to the MOR, out of a total of 10 hits (45% hit rate). Median affinities were the same for each receptor: 4.67 μM for the KOR vs. 4.39 μM for the MOR, despite having a median rank of 6,842 to the KOR and 251,804 to the MOR. As a previous virtual screen at the KOR crystal structure surprisingly returned some agonist molecules, we tested all of our compounds in G$_i$ functional assays for agonism and antagonism (Negri, Rives et al. 2013). While most compounds were shown to be antagonists, as expected from a screen of an inactive structure (Weiss, Ahn et al. 2013), two compounds were found to be agonists (**Table 2.2**). Despite using this low-enriching system and specifically selecting for interactions that could not be made in the crystal structure of the MOR anti-target, hit rates and affinities for each receptor were comparable.
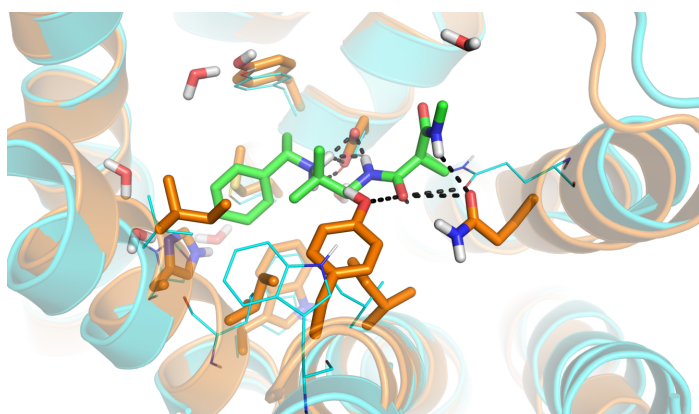
Figure 2.2 The KOR and MOR binding sites. A selective compound, compound **2** (green), is shown in the docked pose to KOR (orange), and to MOR (cyan), where it cannot fit.

IV. Addition of receptor flexibility to aminergic virtual screening protocol

Upon visual inspection of the docking false negatives to HRH1, it became clear that receptor flexibility might play a role in this system as well. For example, the highest affinity HRH1 hit (compound **6**) does make the key salt-bridge to $D^{3.32}$ in the docked pose but has a poor van der Waals (vdW) score (-5.8 DOCK score units, whereas top-ranked molecules in this receptor typically scored around -30) due to several steric clashes with the receptor structure.

To test whether receptor flexibility played a role in predicting false negatives, we turned again to the known ligands in ChEMBL10. We selected 7 chemically diverse ligands with large solvent-accessible surface areas (SASA) in the predicted lowest-energy conformation as an indicator of three-dimensional size. We postulated that these ligands may be too bulky to dock into the HRH1 structure, which presents a small and compact binding site most suitable for fragment-discovery efforts. We then used the program 3K-ENM to generate 3,700 normal mode models with slightly expanded HRH1 binding sites, to perturb the binding site around the crystal structure minimum (Yang and Sharp 2009). For all of the expanded HRH1 normal mode models, we docked the entire set of known ligands from ChEMBL10 and property-matched decoys and known experimental non-binders, and for each of the 7 largest SASA ligands, we selected the model that ranked that ligand most highly. We then combined the docking results of the known ligands and decoys from the crystal structure and those 7 expanded binding site models. We found that adding the normal mode models did not affect enrichment of known binders (logAUC = 35 for the crystal structure only, logAUC = 32 for the combined docking). However, when we tested whether the combined

docking would have predicted HRH1 binding for the false negatives found by our prospective screen, the results were encouraging. While the crystal structure alone did not enrich those prospective hits (logAUC = -1.7), the combined docking did (logAUC = 14). Now, by combining the performance of the crystal structure and the expanded binding site model, we would have filtered out the HRH1 false negatives from the first round. We then asked if all 7 normal mode models were necessary, and, in fact, found that using only one normal mode model was sufficient for the same retrospective enrichment. We selected a model that ranked the ligand with the largest SASA first. Importantly, this model was not selected using any knowledge of our virtual screening hits, but rather on the ability to recognize large, already annotated HRH1 ligands.

To see if adding this expanded model would help triage prospective screening efforts, we repeated the virtual screen of the ZINC lead-like database against the on- and anti-targets, this time against both the HRH1 crystal structure and the expanded HRH1 binding site model. We also filtered the top-ranked molecules more stringently. Again molecules that ranked in the top 30,000 docked to both HTR2A and DRD2 were selected and molecules having a Tc > 0.5 to the top 50,000 to HRH1 were filtered out. In addition, all molecules having a Tc > 0.8 to the top 500,000 ranked molecules to HRH1 were filtered out. In this way, we ensured that all molecules we selected and their close analogs were ranked in the bottom 85% of the database to both the HRH1 crystal structure and the expanded site. From the 91 compounds that passed these filters, we selected 20 to test for binding to the three receptors.

The stricter filtering did not affect the on-target hit rates, with 13 and 9 (65% and 45%) molecules binding to HTR2A and DRD2 respectively, and 6 (30%) dual binders. However, the rate of false negatives actually rose, with 15 (75%) of the molecules binding to HRH1 (**Table 2.1**). Affinities were comparable to the first screen, with median affinities of 1,970 nM, 6,442 nM and 1,271 nM, and median ranking of 14,466, 15,595 and 726,846 to the HTR2A, DRD2, and HRH1 respectively. HRH1 hits were again generally of higher affinity, with 5 sub-micromolar compounds (compared to 3 and 0 for HTR2A and DRD2).

V. Addition of receptor flexibility to opioid virtual screening protocol

Returning to our parallel opioid receptor screening campaign, we compared the crystal structures of the three opioid receptors. Both DOR and MOR were co-crystallized with a morphine-like ligand, and the binding site for these ligands is very close to TM5; consequently, the whole pocket condenses and shrinks slightly around it. JDTic completely extends and lines the bottom of the KOR binding site all the way to TM2, which pushes TM1 outward. As a result, during virtual screening, many ligands could fit in the KOR binding site that were too bulky for the MOR crystal structure. As these receptors are close, both in sequence and ligand space, we reasoned that similar conformations would be thermally accessible at equilibrium. Believing these structures to be ligand-dependent, we wondered what would happen if we added another MOR structure to the selectivity filter – one that had the extended binding site and topology of the KOR

structure. We used MODELLER9v8 to generate 100 MOR homology models based on the KOR crystal structure. Three compounds (**6, 10, 15**) were false negatives in the first round of docking, and so we chose two MOR models that managed to accommodate all three compounds to use as further selectivity filters (**Figure 2.3**).

We screened the ZINC lead-like library again at the KOR and MOR crystal structures, and the MOR models. This time, we calculated the ratio of each compound's rank in the KOR screen to the compound's rank in any of MOR screens. Molecules were sorted based on smallest rank ratios to any of the MOR structures, and the binding poses of the molecules with the biggest minimum rank ratio were visually inspected for interaction quality, as before. Ultimately, only 3 compounds were purchasable, and binding studies revealed that 2 of them were non-selective KOR and MOR ligands (**Supplementary Table A.2.2**).
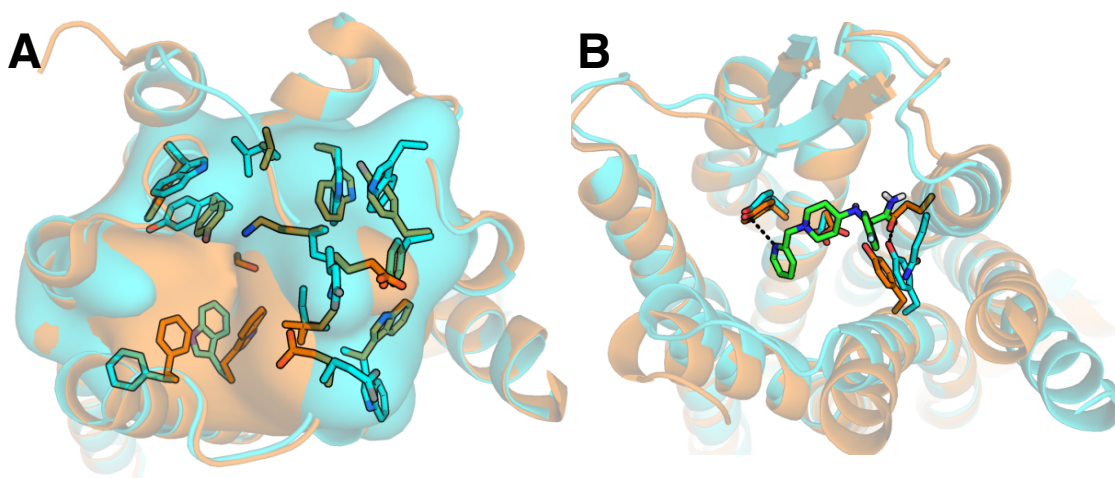
Figure 2.3 Expanded binding sites of HRH1 and MOR flexbile models. (A) The HRH1 ENM (cyan) is aligned with the HRH1 crystal structure (orange), showing the side chains in the orthosteric binding site that change rotamers. The model has a

larger binding site, able to accommodate larger ligands, as shown by the corresponding volume surfaces of each. (**B**) The MOR model (cyan) is aligned with the MOR crystal structure (orange), showing the side chains in the orthosteric binding site. TM1 and TM2 are displaced in the model, as in the KOR crystal structure, allowing **15** (green) to dock.

IV. Ligand-based methods for selectivity

As a comparison, in order to find out if ligand-based methods could just as easily find selective aminergic molecules, we used the Similarity Ensemble Approach (SEA). SEA can rapidly interrogate targets based on the chemical similarity of the query ligands compared to sets of annotated ligands (see **Methods**). Here, we screened the ZINC lead-like database for molecules that significantly resembled HTR2A and DRD2 ligands but also did not resemble HRH1 ligands. After this filtering, the top 15 molecules that were most chemically dissimilar to known HRH1 ligands were selected for purchase and experimental testing, of which, only 7 were actually available. Only 1 molecule displayed weak binding activity at HTR2A (14% hit rate), fortuitously also binding weakly to DRD2, but avoiding HRH1. No molecules showed HRH1 binding. While this suggests that a stricter similarity filter to already known HRH1 ligands would have eliminated potential binders, retrospective analysis of our HRH1 docking hits shows that all but 1 of them either have insignificant SEA scores or Tanimoto coefficients less than 0.3 to annotated ChEMBL ligands; a further similarity filter would not have eliminated these molecules.

2.4 Discussion

Three key conclusions emerge from this study: first, that docking can be used to successfully find compounds with a predefined specificity for a combination of targets and anti-targets. Second, a big caveat to the successful identification of selective compounds was the high incidence of false positives in virtual screening, namely compounds predicted not to bind to anti-targets that did actually bind with high affinity. Finally, complete treatment of receptor flexibility for virtual screening is not currently computationally tractable.

Virtual screening has universally and historically been shown to prospectively predict true positives in GPCR docking with high hit rates (Sabio, Jones et al. 2008; Kolb, Rosenbaum et al. 2009; Katritch, Jaakola et al. 2010; de Graaf, Kooistra et al. 2011). This is true even when the structure is deemed an anti-target and stringent filtering is added as an additional constraint in the virtual screen. We found hit rates for the on-targets of 63%, 40% and 41% for serotonin receptor 5-HT_{2A}, dopamine receptor D_2, and the kappa opioid receptor (HTR2A, DRD2, and KOR respectively), which are comparable to, or higher than, previous screens.

Our polypharmacological screen was able to predict one high-affinity, selective compound directly from docking (compound **21**), and four other selective, low-affinity compounds. This in itself is remarkable, as despite considerable efforts, few histamine receptor H_1 (HRH1)-selective anti-psychotic drugs or drug leads have been found. All purchased analogs, however, failed to increase selectivity; either they lost target affinity, or, in most cases, affinity for HRH1 increased. Not only that, but ligand-based virtually screening performed worse, finding only one

weakly binding molecule. This further illustrates the difficulty of optimizing selectivity against this receptor even when given a starting point in selective-chemical space, although docking did specifically choose the most selective molecule to pass our filters. Due to its ability to directly find novel hits and explore a large amount of chemical space, docking may be useful in order to find novel molecules to serve as starting points for optimization using other *in silico* tools, for example the machine-learning method of Besnard, et al.(Besnard, Ruda et al. 2012). This is especially true given that the screened database was purchasable compounds; there may be analogs easily accessible by medicinal chemistry to confer selectivity that were simply not purchasable or present in the database.

An equally fascinating outcome of this prospective study was the remarkably high hit-rates obtained for the anti-targets. These showed high hit rates of 65% and 45% for HRH1 and the mu opioid receptor (MOR) respectively, despite being actively selected against. Crystal structures are static snapshots of the receptor in one conformation, however, numerous biophysical studies have shown that GPCRs exist in an ensemble of thermally accessible conformational states in the absence, and even in the presence of, a bound ligand (Nygaard, Zou et al. 2013). Theoretically, finding a compound that binds to just one of these myriad conformations counts as a hit, of which the crystal conformation is one of many. For a compound to avoid an anti-target, however, the molecule is required to avoid any possible conformation of the receptor, of which, again, the crystal structure is one of many. When the receptors are closely related and bind
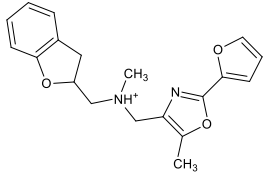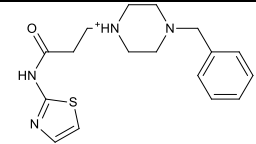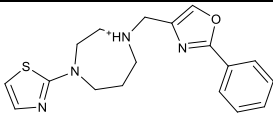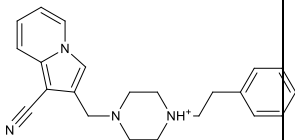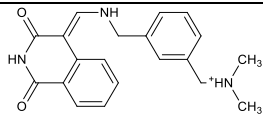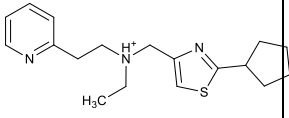
topologically-similar molecules, as is the case for HTR2A, DRD2 and HRH1, it is difficult to address how a potential ligand can avoid all the possible conformations of one receptor, yet fit easily into any one of another. When the receptors are highly homologous subtypes that probably undergo the same thermal motions at room temperature, as is the case for the KOR and the MOR, finding a molecule that binds one receptor and ignores all conformations of the other seems daunting, and the results of this study support that.
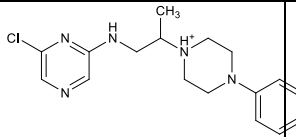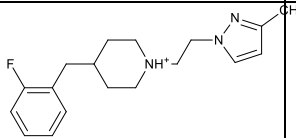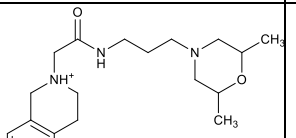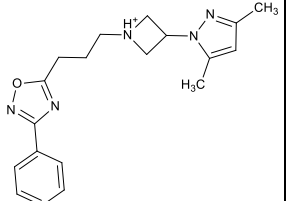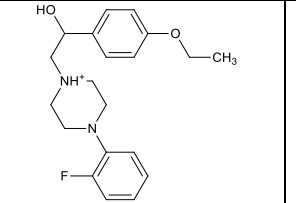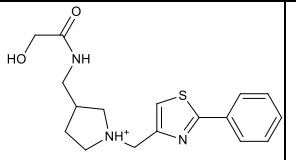
We sought to address receptor flexibility by including alternate anti-target conformations with an expanded binding site that, together with the crystal structures, could accommodate even the largest known ligands. Unfortunately, this resulted in an even higher hit rate on the anti-target. The false negatives were predicted not to bind on the basis of either clashes with both receptor conformations, represented by positive van der Waals scores, or final poses that did not include the crucial salt bridge with $D^{3.32}$. To rationalize this, we further asked if alternative methods, such as induced-fit docking, which involves optimizing the binding site residues around a docked pose and then re-docking the molecule, would have allowed sufficient space in the binding site for these molecules to find acceptable poses. While less than 30% of the false negatives would have been rescued and correctly predicted to be HRH1 binders (including the 0.8 nM HRH1 hit, which was still predicted as a non-binder), 6 true negatives (12.5%) were also shown to find acceptable HRH1 binding poses. This shows that induced-fit docking also would not be useful as a further filter. More stringently tightening our original Tanimoto coefficient or rank filters would have filtered out

the entire database. While this could conceivably mean that there are no molecules in the database with the desired polypharmacology, a real possibility depending on the query and the database, we did end up finding at least 7 molecules across both design campaigns fulfilling our desired goals, showing that, perhaps due to the library's bias towards, and overpopulation with, aminergic ligands, successfully reproducing the designed profile was achievable. Larger databases based on synthetically tractable analogs of ZINC may help alleviate this issue.

Modulation of the neurotransmitter signaling pathways underlying many psychiatric conditions requires a complex and delicate balance. Effective and safe drugs act on therapeutic targets while avoiding the adverse effects of binding to off-targets. While the conceit of this study was not to identify new psychiatric drugs, our results offer several potential starting points for desired specificity and other novel properties. However, these were few among the plentiful hits from the docking. Like previous virtual screens at models and structures of GPCRs, we report high hit rates to our various on-targets. More importantly, explicitly including the anti-target in the structure-based design method did not negatively impact the hit rate, and we implicate insufficient treatments of receptor flexibility as the main deterrent to successful anti-target predictions. Future efforts to accurately incorporate receptor flexibility into the virtual screening protocol would establish docking as a fast, easy method for the direct design of ligands with a predetermined polypharmacological profile.

Table 2.1. Hits found in virtual screening of HTR2A/DRD2.

#	Structure and ZINCID	HTR2A rank	DRD2 rank	HRH1 rank	HTR2A K_i (nM)	DRD2 K_i (nM)	HRH1 K_i (nM)
5	C12585531	1814	6033	571928	1784	5698	>10000
6	C19372191	1936	2024	326040	>10000	>10000	0.8
9	C64402001	2898	1527	1399679	3324	3231	19
12	C58355688	5430	8314	3004881	59	1821	35
14	C32080342	5671	7537	692414	2500	7881	>10000
19	C63994913	9272	3867	170918	793	5947	>10000

#	Structure						
21	C53561601	9893	5772	74924	55	334	1144
23	C19791225	12926	1600	2042077	334	2233	280
25	C25409567	15717	1947	2525004	3430	3195	>10000
31	C71880639	5976	17327	646631	3654	5596	2683
35	C65255278	10478	3342	1078327	2268	1996	718
38	C67817126	13792	66225	840818	4049	4918	2562
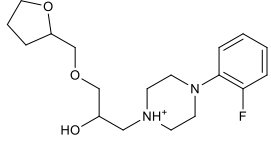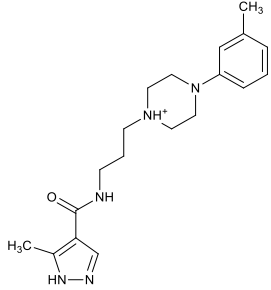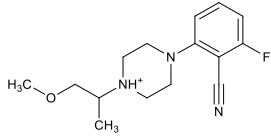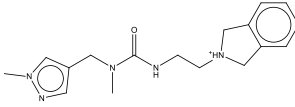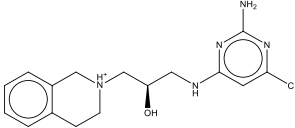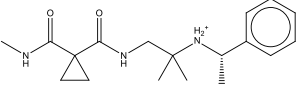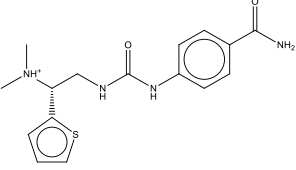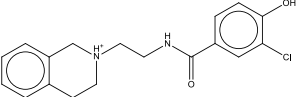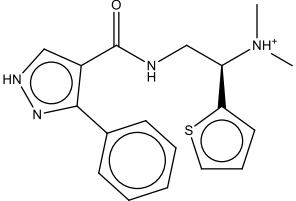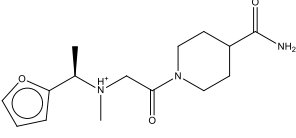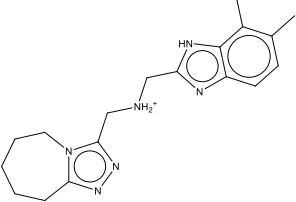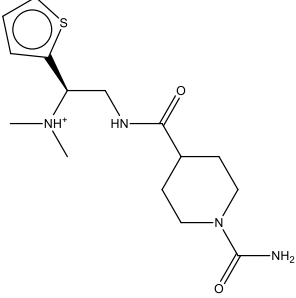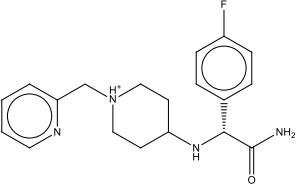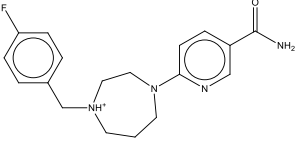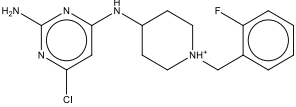
39	C68048611	15140	12396	1281240	1443	7745	4241
42	C63770987	17160	5918	812174	33	1606	44
47	C69896206	26751	9012	603194	3474	6912	1110

Table 2.2. Hits found in virtual screening of KOR

#	Structure and ZINCID	KOR rank	MOR rank	KOR K_i (µM)	MOR K_i (µM)	MOR/KOR ratio
1	C68408255	263	86198	14.4	3.04	0.211
2	C44913867	1166	107428	2.93	10.2	3.48

3		1548	119964	4.67	> 100	> 21.4
	C68387316					
6		4524	201440	1.81 (EC$_{50}$ = 17.0)[a]	0.478 (EC$_{50}$ = 5.21)[a]	0.264
	C43714067					
8		5464	520993	> 100	1.29	< 0.0129
	C45076626					
10		6141	255399	10.5	2.47	0.235
	C69181132					
11		7542	543545	> 100	8.69	0.0869
	C69067765					
13		7995	346294	3.14	6.92	2.20
	C67847468					

14	C47872495	9734	492977	2.69 (N.D.) [a]	2.43 (EC$_{50}$ = 12.3) [a]	0.903
15	C66939732	17057	59182	> 100	45.7	< 0.457
20	C48345561	27360	793832	11.1	5.70	0.513
22	C52469366	28472	248208	5.56	> 100	> 17.9

[a] EC$_{50}$ in functional agonist assay

2.5 Methods

I. Homology modeling and docking

The alignments for the construction of the DRD2 and HTR2A models were generated using PROMALS3D (Pei, Kim et al. 2008), and homology models of the DRD2, HTR2A, and MOR receptors were built with MODELLER 9v8, using the crystal structure of the dopamine D_3 receptor (PDB ID: 3PBL) as the template for the aminergics and the KOR crystal structure (PDB: 4DJH) for the opioid (see **Supplementary Information** for full alignments). Elastic network models for the expanded H_1 binding site were produced by the program 3K-ENM.(Yang and Sharp 2009) We used DOCK 3.6 to screen the ZINC database as described (see **Results**). The flexible-ligand sampling algorithm in DOCK 3.6 superimposes atoms of the docked molecule onto binding site matching spheres, which represent favorable positions for individual ligand atoms. Complementarity of each ligand pose is scored as the sum of the receptor-ligand electrostatic and van der Waals interaction energy and corrected for context-dependent ligand desolvation. Partial charges from the united-atom AMBER force field were used for all receptor atoms except for, in the KOR, Asp138[3.32], Glu209 in ECL2, and Glu297[6.58], for which the dipole moment was increased as previously described to boost electrostatic scores for poses in polar contact with these important residues.(Carlsson, Yoo et al. 2010)

To measure enrichment, we used the metric of adjusted logAUC, which measures the prioritization of known ligands over generated decoy molecules compared to what would be expected at random (an adjusted logAUC of 0

represents random). The log-weighted enrichment emphasizes the highest ranking molecules, which are the most likely to be selected for testing.

The induced-fit docking was performed using a three-step protocol: 1) each molecule was docked into the receptor; 2) for each pose, the receptor side chains and backbone were minimized around the posed ligand using PLOP (Jacobson, Friesner et al. 2002); 3) the compound was redocked into the optimized receptor binding site.

For the aminergic receptors, molecules were visually selected on the basis of the formation of the key salt-bridge to $D^{3.32}$ (Ballesteros-Weinstein numbering (Ballesteros and Weinstein 1995)) in the docked pose to both HTR2A and DRD2, a complementary fit to the HTR2A and DRD2 binding sites and a poor fit to the HRH1 binding site (**Figure 1**), as well as vendor availability and hit diversity (**Table S1**). When selecting molecules for purchasing in the KOR screen, those possessing known high-internal-energy motifs that do not appear in the Cambridge Structural Database were manually discarded. More specifically, opioid molecules were selected on the basis of key salt-bridges to $D^{3.32}$ in the KOR binding pose, a complementary fit to the KOR binding site, and interactions with residues that are either specific to KOR or exist in a different rotamer in the MOR crystal structure (Glu209 in the second extracellular loop, Glu297$^{6.58}$, Tyr312$^{7.35}$, and Thr111$^{2.56}$, Gln115$^{2.60}$, Tyr320$^{7.43}$; **Figure 3**; residue numbering from KOR), thus making a similar binding pose in the MOR impossible.

II. Ligand-based selectivity profiling

In order to determine if ligand-based methods could discover molecules with pre-defined selectivities, we used the Similarity Ensemble Approach (SEA), which compares individual ligands, and sets of ligands, to the ligand sets for multiple targets; two targets are related, or a particular ligand is predicted to modulate a target if the ligands are related to one another. Here, the sets for HRH1, HTR2A, and DRD2 ligands were defined as all the molecules with experimentally measured IC_{50}s or K_is < 10 μM in the ChEMBL16 database. These were the query sets screened against the ZINC lead-like database. If a molecule had a significant E-value (E<-5) to the HRH1 set, it was filtered out. A purchasability filter was added, as was a filter to keep +1 charged ligands. Finally, the remaining molecules were ranked by Tanimoto coefficient similarity (ECFP4 fingerprints) to any annotated HRH1 binder.

III. Binding Affinity and Functional Assays

Radioligand binding and functional (GloSensor, Tango, and FLIPR) assays at the DRD2, HTR2A, HRH1, KOR, and MOR were carried out at the National Institute of Mental Health Psychoactive Drug Screening Program as previously described.(Allen, Yost et al. 2011)

2.6 References

Allen, J. A., J. M. Yost, et al. (2011). "Discovery of beta-arrestin-biased dopamine D2 ligands for probing signal transduction pathways essential for antipsychotic efficacy." Proc Natl Acad Sci U S A **108**(45): 18488-18493.

Ballesteros, J. and H. Weinstein (1995). "Integrated methods for modeling G-protein coupled receptors." Methods Neurosci **25**: 366-428.

Becker, O. M., D. S. Dhanoa, et al. (2006). "An integrated in silico 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT1A agonist (PRX-00023) for the treatment of anxiety and depression." J Med Chem **49**(11): 3116-3135.

Besnard, J., G. F. Ruda, et al. (2012). "Automated design of ligands to polypharmacological profiles." Nature **492**(7428): 215-220.

Carlsson, J., R. G. Coleman, et al. (2011). "Ligand discovery from a dopamine D3 receptor homology model and crystal structure." Nat Chem Biol **7**(11): 769-778.

Carlsson, J., L. Yoo, et al. (2010). "Structure-based discovery of A2A adenosine receptor ligands." Journal of Medicinal Chemistry **53**(9): 3748-3755.

Chien, E. Y., W. Liu, et al. (2010). "Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist." Science **330**(6007): 1091-1095.

1. de Graaf, C., A. J. Kooistra, et al. (2011). "Crystal structure-based virtual screening for fragment-like ligands of the human histamine H(1) receptor." J Med Chem **54**(23): 8195-8206.

Evers, A. and T. Klabunde (2005). "Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor." J Med Chem **48**(4): 1088-1097.

Fishman, M. C. and J. A. Porter (2005). "Pharmaceuticals: a new grammar for drug discovery." Nature **437**(7058): 491-493.

Frantz, S. (2005). "Drug discovery: playing dirty." Nature **437**(7061): 942-943.

Gaulton, A., L. J. Bellis, et al. (2011). "ChEMBL: a large-scale bioactivity database for drug discovery." Nucleic Acids Res.

Hopkins, A. L. (2008). "Network pharmacology: the next paradigm in drug discovery." Nat Chem Biol **4**(11): 682-690.

Irwin, J. J. and B. K. Shoichet (2005). "ZINC--a free database of commercially available compounds for virtual screening." J Chem Inf Model **45**(1): 177-182.

Irwin, J. J., T. Sterling, et al. (2012). "ZINC: A Free Tool to Discover Chemistry for Biology." J Chem Inf Model.

Ishibashi, T., T. Horisawa, et al. (2010). "Pharmacological profile of lurasidone, a novel antipsychotic agent with potent 5-hydroxytryptamine 7 (5-HT7) and 5-HT1A receptor activity." J Pharmacol Exp Ther **334**(1): 171-181.

Jacobson, M. P., R. A. Friesner, et al. (2002). "On the role of the crystal environment in determining protein side-chain conformations." J Mol Biol **320**(3): 597-608.

Katritch, V., V. P. Jaakola, et al. (2010). "Structure-based discovery of novel

    chemotypes for adenosine A(2A) receptor antagonists." J Med Chem **53**(4):

    1799-1809.

Kim, S. F., A. S. Huang, et al. (2007). "From the Cover: Antipsychotic drug-induced

    weight gain mediated by histamine H1 receptor-linked activation of

    hypothalamic AMP-kinase." Proc Natl Acad Sci U S A **104**(9): 3456-3459.

Kolb, P., K. Phan, et al. (2012). "Limits of ligand selectivity from docking to

    models: in silico screening for A(1) adenosine receptor antagonists." PLoS

    One **7**(11): e49910.

Kolb, P., D. M. Rosenbaum, et al. (2009). "Structure-based discovery of beta2-

    adrenergic receptor ligands." Proc Natl Acad Sci U S A **106**(16): 6843-6848.

Kroeze, W. K., S. J. Hufeisen, et al. (2003). "H1-histamine receptor affinity

    predicts short-term weight gain for typical and atypical antipsychotic

    drugs." Neuropsychopharmacology **28**(3): 519-526.

Kruse, A., D. Weiss, et al. (2013). "Muscarinic receptors as model targets and

    antitargets for structure-based ligand discovery." Journal of Molecular

    Pharmacology **submitted**.

Maynard, G. D., L. D. Bratton, et al. (1997). "Synthesis and SAR of 4-(1H-

    benzimidazole-2-carbonyl)piperidin with dual histamine H1/tachykinin NK1

    receptor antagonist activity." Bioorganic & Medicinal Chemistry Letters

    7(22): 2819-2824.

Merlos, M., M. Giral, et al. (1997). "Rupatadine, a new potent, orally active dual antagonist of histamine and platelet-activating factor (PAF)." Journal of Pharmacology and Experimental Therapeutics **280**(1): 114-121.

Metz, J. T. and P. J. Hajduk (2010). "Rational approaches to targeted polypharmacology: creating and navigating protein-ligand interaction networks." Curr Opin Chem Biol **14**(4): 498-504.

Mohr, K., C. Trankle, et al. (2010). "Rational design of dualsteric GPCR ligands: quests and promise." Br J Pharmacol **159**(5): 997-1008.

Morphy, R. and Z. Rankovic (2005). "Designed multiple ligands. An emerging drug discovery paradigm." J Med Chem **48**(21): 6523-6543.

Morphy, R. and Z. Rankovic (2007). "Fragments, network biology and designing multiple ligands." Drug Discov Today **12**(3-4): 156-160.

Murugesan, N., J. E. Tellew, et al. (2002). "Discovery of N-isoxazolyl biphenylsulfonamides as potent dual angiotensin II and endothelin A receptor antagonists." J Med Chem **45**(18): 3829-3835.

Mysinger, M. M., M. Carchia, et al. (2012). "Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking." J Med Chem **55**(14): 6582-6594.

Mysinger, M. M., D. R. Weiss, et al. (2012). "Structure-based ligand discovery for the protein-protein interface of chemokine receptor CXCR4." Proc Natl Acad Sci U S A **109**(14): 5517-5522.

Negri, A., M. L. Rives, et al. (2013). "Discovery of a Novel Selective Kappa-Opioid Receptor Agonist Using Crystal Structure-Based Virtual Screening." J Chem Inf Model.

Nygaard, R., Y. Zou, et al. (2013). "The dynamic process of beta(2)-adrenergic receptor activation." Cell 152(3): 532-542.

Pei, J., B. H. Kim, et al. (2008). "PROMALS3D: a tool for multiple protein sequence and structure alignments." Nucleic Acids Res 36(7): 2295-2300.

Roth, B. L., D. J. Sheffler, et al. (2004). "Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia." Nat Rev Drug Discov 3(4): 353-359.

Sabio, M., K. Jones, et al. (2008). "Use of the X-ray structure of the beta2-adrenergic receptor for drug discovery. Part 2: Identification of active compounds." Bioorg Med Chem Lett 18(20): 5391-5395.

Sali, A. and T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints." J Mol Biol 234(3): 779-815.

Spetea, M., M. F. Asim, et al. (2013). "Current kappa Opioid Receptor Ligands and Discovery of a New Molecular Scaffold as a kappa Opioid Receptor Antagonist using Pharmacophore-Based Virtual Screening." Curr Pharm Des.

Urban, J. D., W. P. Clarke, et al. (2007). "Functional selectivity and classical concepts of quantitative pharmacology." J Pharmacol Exp Ther 320(1): 1-13.

Weiss, D. R., S. Ahn, et al. (2013). "Conformation Guides Molecular Efficacy in

Docking Screens of Activated beta-2 Adrenergic G Protein Coupled

Receptor." <u>ACS Chem Biol</u> **8**(5): 1018-1026.

Wu, H., D. Wacker, et al. (2012). "Structure of the human kappa-opioid receptor in

complex with JDTic." <u>Nature</u> **485**(7398): 327-332.

Yang, Q. and K. A. Sharp (2009). "Building alternate protein structures using the

elastic network model." <u>Proteins</u> **74**(3): 682-700.

Chapter 3:

Ongoing Projects and Future Directions

With these proofs of concept in hand, where do we turn next? Here, I describe the many on-going and future directions that are lurking in the dark. The first section (3.**1**) concerns itself again with pH-sensing receptors. The physical mechanisms responsible for selectivity between GPR68, GPR65, and GPR4 are investigated by mutagenesis, and a GPR4-selective ligand is disclosed. In the second (3.**2**), I introduce a new type of selectivity - functional - and show that models can be used to explain the molecular mechanism of biased signaling by different ligands. In the third (3.**3**), I examine what targets emerge after hunting with a new tool and screening assay. The fourth section (3.**4**) investigates both the utility of a structure-based approach and questions where the important information is stored during the modeling procedure - is it in the template choice? What if the ligand were covalent? Finally, the last section (3.**5**) presents the grand ideas of what is possible, and what connections can be made, with this new toolset.

3.1 Magic bullets for GPR4

Ligand discovery efforts for GPR4, the last remaining pH-sensing receptor of the trio, are currently underway. As GPR4 shares a higher overall sequence identity to GPR68 than GPR65 does (45% vs. 33%), especially in the putative ogerin binding site, it is unsurprising that screening previously tested GPR68 compounds against GPR4 revealed allosteric agonist hits. One compound, ZINC04929116, is actually selective for GPR4 over GPR68 in G_s-signaling assays; its docking mode at GPR4 (with the receptor model built based on GPR68) is shown in **Figure 3.1**.
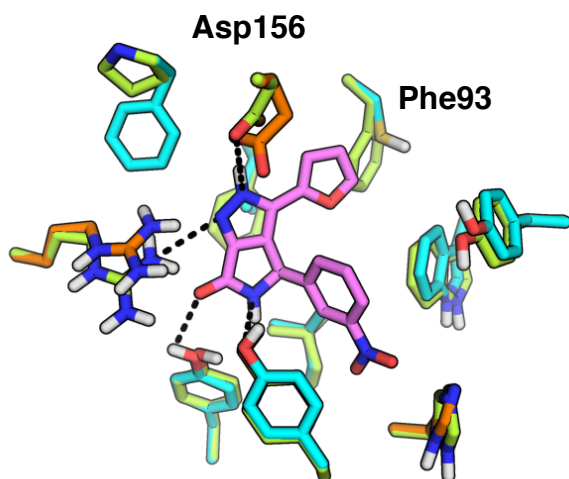
Figure 3.1 Comparison of the binding site for ZINC04929116 in GPR4 and GPR68. ZINC04929116 (magenta) docked at GPR4 (green) is shown overlaid on top of GPR68 (cyan). Typical polar interactions are shown as black dotted lines.

Although most of the binding site residues are conserved between GPR4 and GPR68, the different activities for the ligand might be explained by two different amino acid substitutions. Asp156 is Glu160 in GPR68, and the lack of one methylene carbon forces the ligand to sit a little higher in the cavity – this allows an aromatic interaction between the furan ring of the ligand and Phe93. In GPR68, this residue is Cys97, which precludes any sort of pi-stacking for this

receptor with the ligand. Currently, we are mutating these residues in GPR4 and

GPR68 to their respective counterparts to see if they are responsible for

ZINC04929116's selectivity. Locating these selectivity hotspots would also validate

the binding pose of the ligand, allowing for the design of more selective analogs

and virtual screening for more selective chemotypes.

3.2 The mechanism of functional bias at the D_2 receptor

As discussed in Chapter 2, the dopamine D_2 receptor is a leading target for the development of novel antipsychotics. While Chapter 2 concerned the design of molecules that featured specific GPCR binding profiles, it neglected to take into consideration that molecule's functional action on the receptor; the ligand may bind, but what then? These ligands could behave as agonists and increase a response of downstream signaling pathways, as neutral antagonists, which effectively block other ligands from binding, or as inverse agonists and shut down downstream events. Aripiprazole, a top selling pharmaceutical indicated for schizophrenia and bipolar disorder, exhibits partial agonist activity and functional selectivity at the D_2 receptor. Recently, we discovered other ß-arrestin-biased D_2 ligands which have shown therapeutic potential in animal models of schizophrenia, and we generated several aripiprazole analogs that exhibit ligand bias.[1]

We have recently begun to map out the structure-functional selectivity relationships (SFSR) necessary to manifest D_2 ligand bias, especially as it pertains to ß-arrestin bias. Here, we utilized a series of various substituted aripiprazole analogs to evaluate G protein activation versus ß-arrestin recruitment using several functional assays in order to construct a comprehensive SFSR profile. In order to understand the molecular basis for this ligand bias, we docked these molecules and proposed mutations to selectively abrogate, and understand, this ligand bias.

We previously noted that modifications to the dichlorophenyl portion of aripiprazole influenced activity at both the ß-arrestin and G protein pathways of

the D2 receptor.[2] Exploration of replacement of aripiprazole's dichlorphenyl with a series of *N*-substituted indoles confers a SFSR profile ranging from balanced agonism to ß-arrestin biased, dependent on the *N*-substitution present. Indole was chosen because the nitrogen group can be easily functionalized, and 4-substituted indole-containing aripiprazole ligands have not been previously explored, thus potentially offering a new series of ligands with unique properties. As shown in **Figure 3.2**, interestingly, indole *N*-methyl substitution confers ß-arrestin bias, whereas unsubstituted N-H indole ligands exhibit balanced efficacy in G-protein and ß-arrestin pathways.
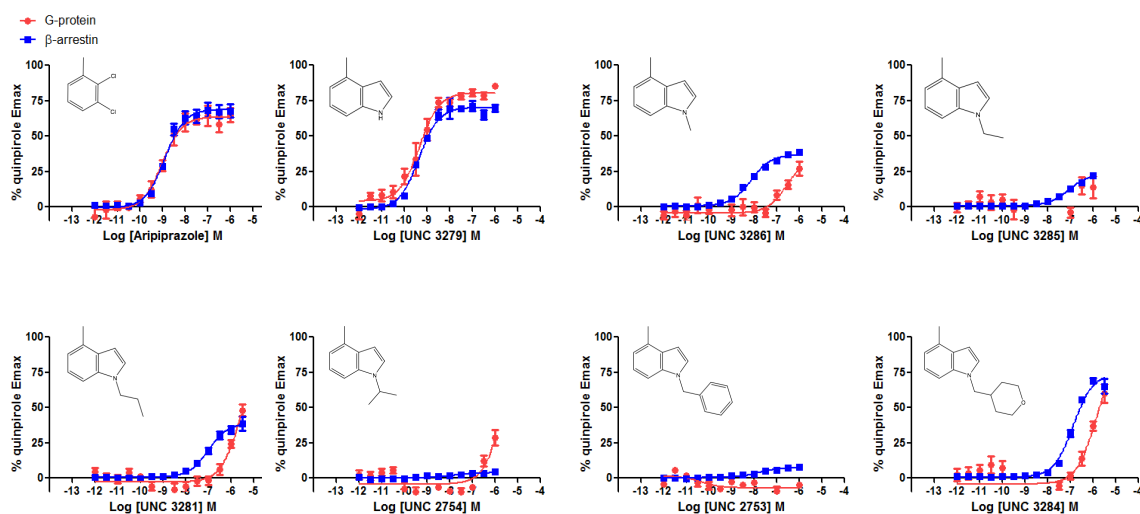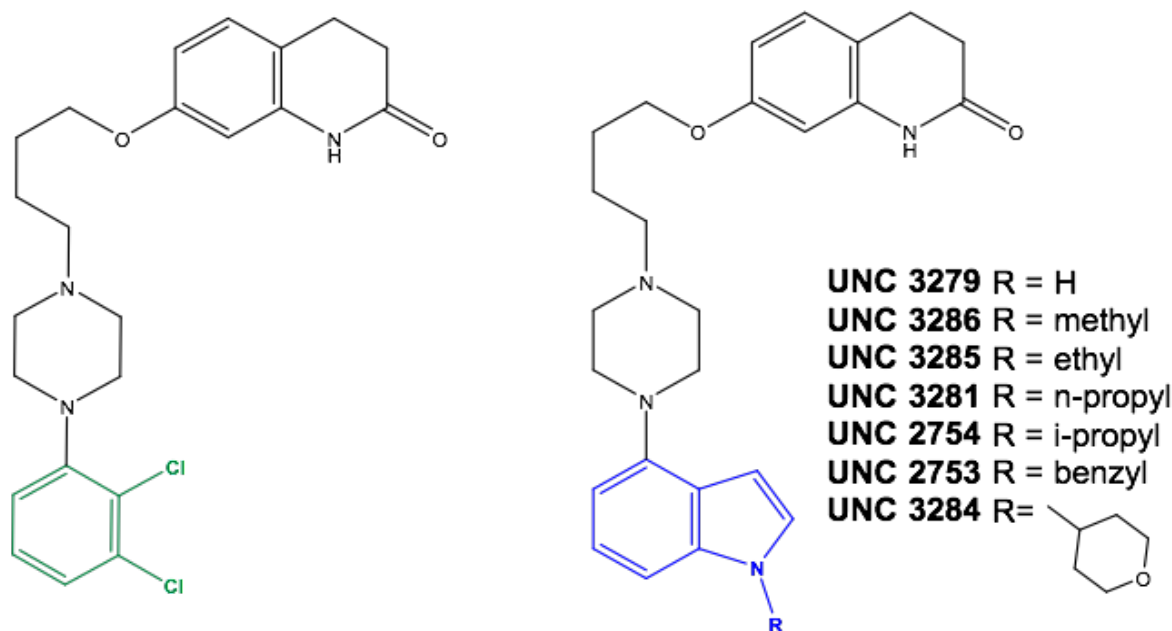
Figure 3.2 Functional efficacy profiles of *N*-indole substituted aripiprazole analogs. Indole analogs (right, blue) of aripiprazole (left, green) were *N*-substituted at the 4 position and tested for G protein and ß-arrestin functional activity. Upon the addition of a methyl substituent, G protein activation is disrupted, suggesting some necessary polar interaction, but ß-arrestin activity still remains. G protein activation is rescued upon addition of the oxane ring, which contains a hydrogen bond acceptor.

Extracellular loop 2 (EL2) lines the binding pocket in the dopamine D_3 crystal structure[3] and the serotonin 5-HT$_{2B}$ receptor crystal structure in complex with the ß-arrestin-biased ligand, ergotamine[4], also reveals that EL2 residues interact extensively with the ligand. Based on this, we sought to investigate mutations of I184 in EL2 of the dopamine D_2 receptor to determine if it was responsible for conferring ß-arrestin bias. Surprisingly, as seen in **Figure 3.3**, mutations here completely abolished the ß-arrestin recruitment by ß-arrestin biased agonists, without affecting G_i signaling or affinity, essentially converting these biased ligands into neutral antagonists. However, EL2 mutations had no effect on the efficacy of other agonists including the unsubstituted indole compounds or aripiprazole itself. The inactivity for the *N*-methyl indole analog, UNC 3286, at I184 mutations was further investigated using radioligand binding assays and in vitro functional antagonist Schild assays. Our results reveal that the affinity of the N-methyl indole analog, UNC 3286, was not affected by the I184A mutation, and in fact, UNC 3286 exhibits potent antagonist activity.
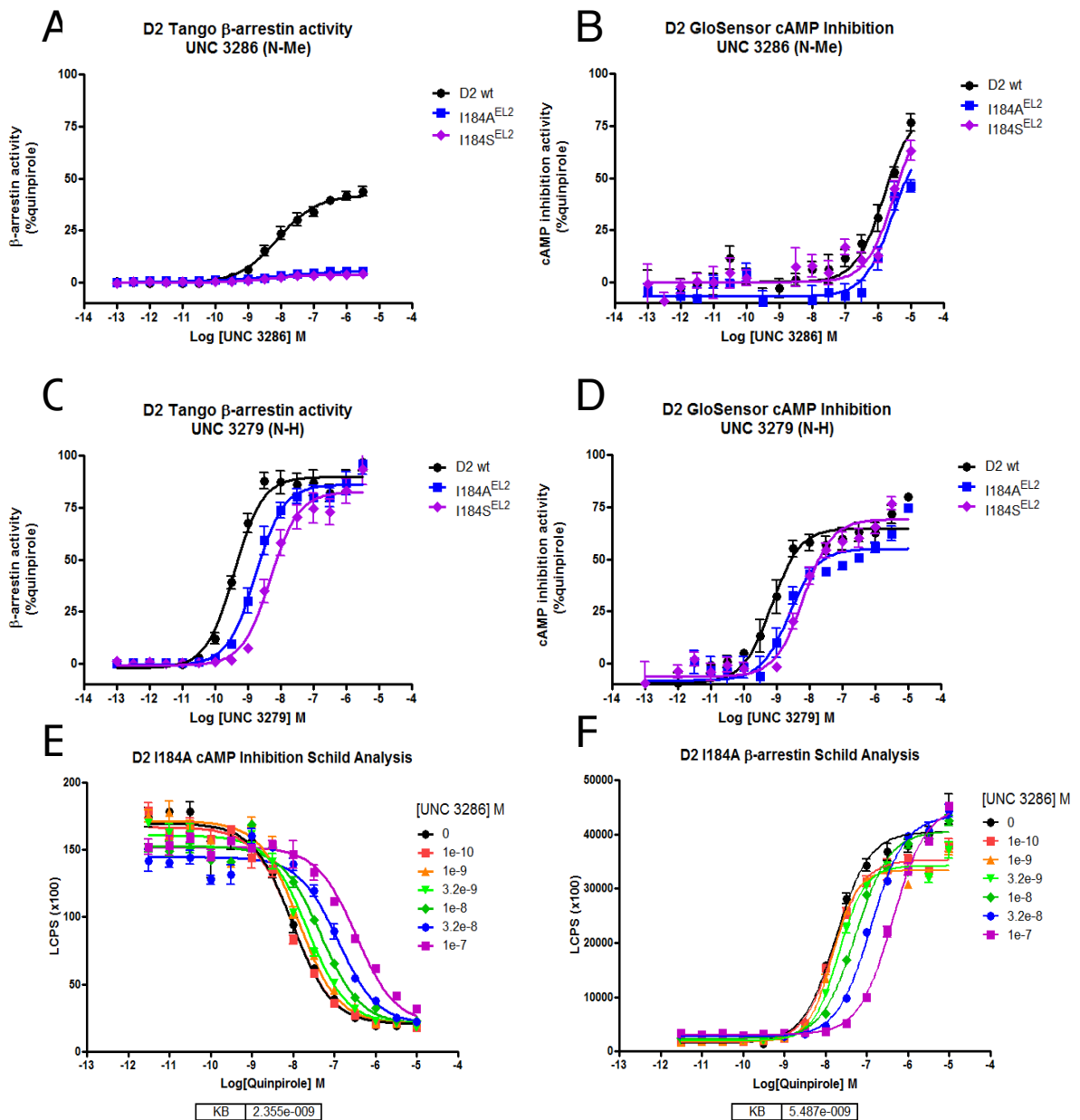
Figure 3.3 Comparison of I184 mutations on *N*-substituted and unsubstituted indole aripiprazole ligands. Although I184A substitution in EL2 does not affect cAMP signaling for neither UNC 3279 nor UNC 3286 (**B** and **D**), this mutation only destroys ß-arrestin signaling for the *N*-methyl substituted UNC3286 (**A** and **C**). Schild analysis of this ligand (**E** and F) shows that affinity for the receptor is unchanged; it is functionally now an antagonist.

Transmembrane 5 (TM5) serines have long been investigated for conferring G protein functional activation, which were noticeably interacting with the ligand in the crystal structure of the agonist-bound state of the ß$_2$-adrenergic receptor. Docking of indole aripiprazole analogs at the D$_2$ receptor model previously used in Chapter 2 (that did reveal some ß-arrestin biased agonists) reveals that the indole interacts with TM5 serines through possible hydrogen bond formation and, importantly, that ß-arrestin biased *N*-methyl analogs preclude hydrogen bond formation with TM5 serines for G protein activation. Mutation of Ser193$^{5.42}$ selectively recovers G protein functional activation for the N-methyl indole analog, thus reversing its ß-arrestin bias (**Figure 3.4**). However, S193A$^{5.42}$ mutation disrupts N-H indole functional activity and affinity, suggestive of hydrogen bond formation to manifest balanced signaling.
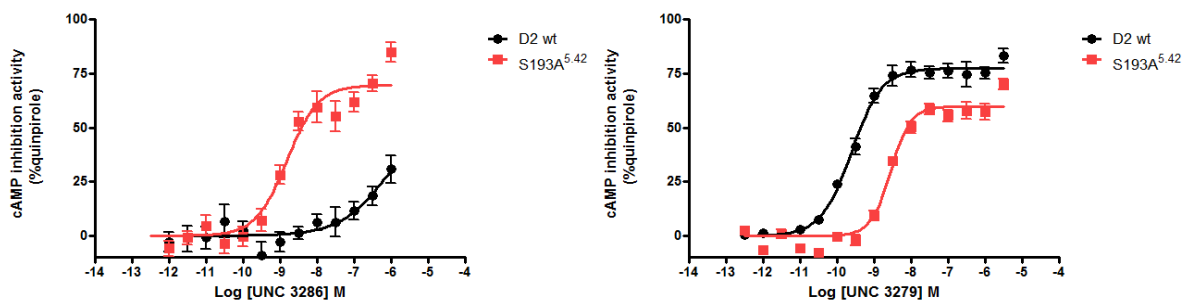
Figure 3.4 Mutational effect of S193A on UNC 3279 and UNC 3286. Mutating S193 to alanine reduces UNC 3279's ability to signal through G proteins, most likely by eliminating a polar interaction with the free indole nitrogen. However, the *N*-methyl analog, UNC 3286, now picks up a hydrophobic interaction with S193A, restoring its G protein activation.

We've shown that slight structural ligand modifications can manifest biased signaling, and in the case of ß-arrestin bias, interaction with EL2 seems to play a prominent role for the D_2 receptor. D_2 ligands that preclude hydrogen bond formation with conserved TM5 serines yet retain interaction with EL2 may show ß-arrestin bias. As shown in **Figure 3.5**, UNC 3279 does not completely lose G protein activity at S193A mutants, because it can also interact with S197, one helical turn below. However, the *N*-methyl moiety of UNC 3286 is solely pointing at S193, and this essentially flipped pose forces the indole ring to interact with I184; one potential reason why these substituted ligands are more sensitive to mutations at this position. Ultimately, these studies suggest common mechanisms of ligand bias and the design of novel functionally selective ligands for other GPCRs.
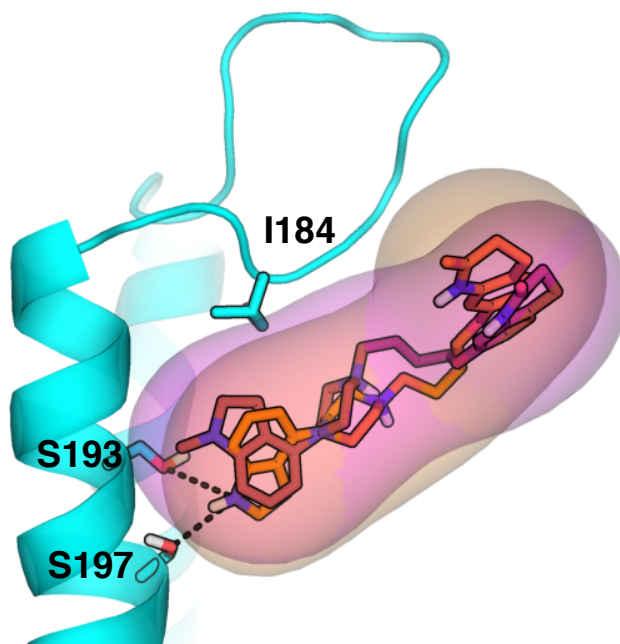
Figure 3.5 Docked poses of UNC 3286 and UNC 3279 at the dopamine D_2 receptor. UNC 3279 (orange) does not interact with I184 due to its hydrogen bonds with S193 and S197; these are responsible for its G protein activity. UNC 3286 (magenta) has an *N*-methyl substitutent that forces the ligand's indole ring to hydrophobically interact with I184. S193A mutation restores G protein activation via new hydrophobic contacts.

3.3 New weapons in the arsenal: β-arrestin screening

As shown in the above section, ligands can have very different effects depending on the functional system measured. Since the yeast high-throughput screen in Chapter 1 was solely based on G protein-coupling, we undertook another, orthogonal high-throughput screen of the orphan GPCR-ome by measuring β-arrestin activity.[5] Whereas there are various G protein-dependent signaling pathways that must be known *a priori* to measure the downstream effects, the vast majority of GPCRs can directly induce arrestin translocation. Here, we screened the same NIH Clinical Collection of 446 compounds at 91 orphan and understudied GPCRs using the transcriptional activation following arrestin translocation (TANGO) assay. Ideal for discovering both agonists and inverse agonists, TANGO has a high signal-to-background ratio and it amplifies relatively small initial inputs into large readout signals.

One of the most efficacious and selective hits was the opioid receptor ligand dextromethorphan for the orphan receptor MRGPRX2 (**Figure 3.6A**). In order to understand the binding mode of these ligands, we generated thousands of homology models of MRGPRX2 based on the κ-opioid receptor, using the same optimization method as that for GPR68. As shown in **Figure 3.7A**, the positive charge of dextromethorphan is predicted to be stabilized primarily through coordination via two negatively-charged amino acids, Glu164$^{4.10}$ and Asp184$^{5.46}$. To test this, we mutated each of these individually to alanine and re-ran a dose-response TANGO assay (**Figure 3.6B** and **C**). Both mutations appear to be necessary for dextromethorphan's β-arrestin activity; there are no positively

charged amino acids nearby, so Glu164 and Asp184 are most likely not involved in required, structural, ionic lock mechanisms, either. Additional mutations of the aromatic residues stabilizing the ligand binding mode are also underway. In analogy to the case of the D_2 receptor, there is a small, hydrophobic amino acid in EL2 (Leu171) that is in contact with the ligand (**Figure 3.7A**); this is also being mutated to see if it affects ß-arrestin activity. In order to investigate the SAR and further characterize the binding site of these opioid ligands, dextromethorphan analogs with N-substitutents, which should point towards EL2, are being currently being tested to see if these affect arrestin signaling and receptor affinity.
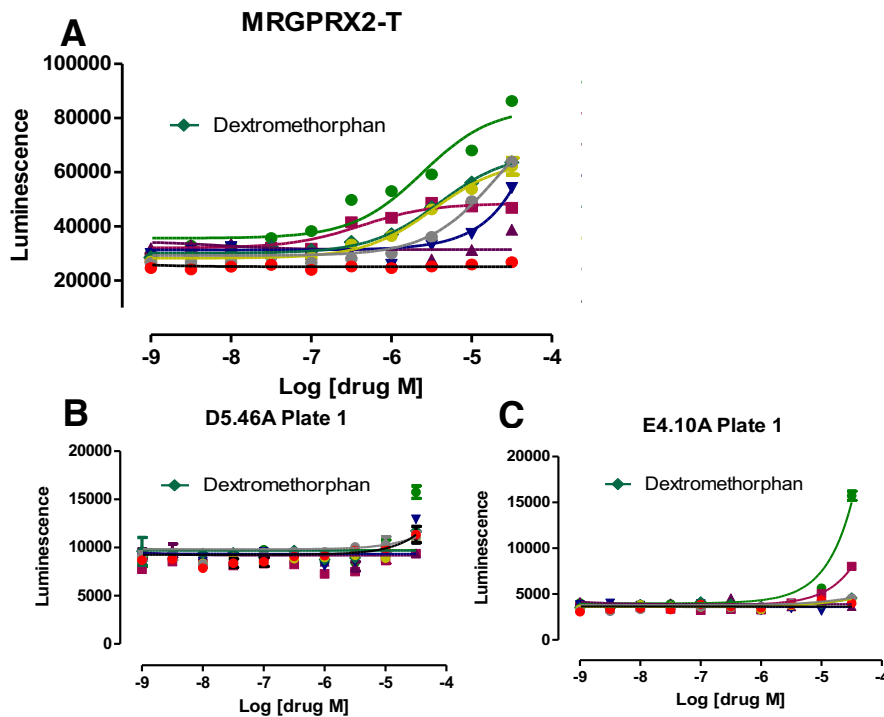
Figure 3.6 TANGO assays of dextromethorphan at wild type, Asp184Ala, and Glu164Ala MRGPRX2 receptors. Dextromethorphan was tested in TANGO assays at wild type (**A**), Asp184Ala (**B**), and Glu164Ala (**C**) receptors for ß-arrestin activity. Each mutation at the negatively charged residues eliminated dextromethorphan activity, indicating that both are important for signaling.

A related receptor, MRGPRX4, actually had the most efficacious hit in the entire high-throughput TANGO screen – the diabetes drug nateglinide, a K_{ATP} channel blocker. These MRGPRX receptors pose a special challenge to the modeling procedure, as the sequence identity to known templates is so low (~20%) and they lack the conserved disulfide bridge between TM3 and EL2. Instead, a putative disulfide bridge was identified between (the especially short) EL2 and TM5 in both receptors, which could provide the same anchoring rigidity seen in other GPCRs. Unlike MRGPRX2, which identified positively charged ligands in the TANGO screen, nateglinide contains a negatively charged carboxylate, potentially coordinated by two arginines, Arg82 and Arg86, on the other side of the binding pocket (**Figure 3.7B**). MRGPRX2, however, has a proline and a tyrosine in these positions, making it impossible for negatively charged ligands to bind. Like MRGPRX2, MRGPRX4 also has the conserved $Glu^{4.10}$ and $Asp^{5.46}$ in TM4 and TM5, and, indeed, positive-charged agonists were identified as ligands in the TANGO screen. The potencies were weak, though, perhaps reflecting the non-ideal coordinating geometries of the side chains shown in **Figure 3.7B**. Mutations of Asp177 and Glu157, as well as Arg82 and Arg86 to confirm the nateglinide binding pose, are currently underway. If successful, we hope to identify new ligands via virtual screens of these receptor models
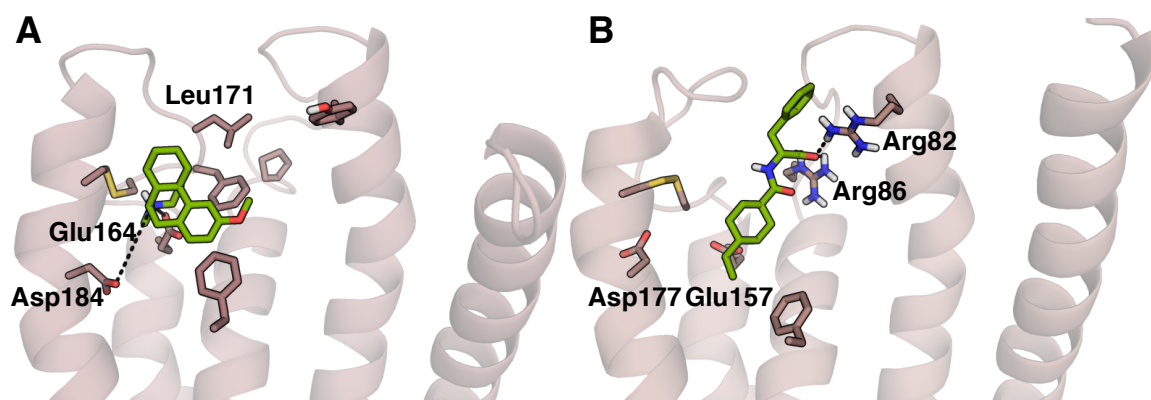
Figure 3.7 Putative binding poses of dextromethorphan and nateglinide at MRGPRX2 and MRGPRX4. (A) The positive charge of dextromethorphan (green) is coordinated between Asp184 and Glu164, while the aromatic rings are stabilized by pi interactions with neighboring phenylalanine residues. In MRGPRX2, there are no TM2 or TM3 arginines to bind negatively charged ligands. Additionally, Leu171 in EL2 provides possible hydrophobic interactions responsible for ß-arrestin activity. **(B)** The binding mode of nateglinide (green) is mainly stabilized by charge-charge coordination between Arg82 and Arg86. The conserved acidic residues, Asp177 and Glu157, are present in MRGPRX4, allowing for the possibility of positively charged ligands also binding.

3.4 Are alternative modeling methods just as successful?

While we had success with this modeling pipeline on at least 2 orphan receptors (and possibly up to 5), the obvious question remains: could the docking results at GPR68 have been replicated, or surpassed by, a simple ligand shape-similarity tool? This is logical to ask, since we are essentially constraining the binding site around the initial optimized ligand. We used three common and powerful ligand-based similarity methods to launch parallel campaigns and see if we could discover ogerin, or related molecules, from the same ZINC lead-like database, using the same benzodiazepine starting point.

The first is a simple 2D structure similarity search, using computed Tanimoto coefficients (Tcs) employing ECFP4 fingerprints, between ogerin and lorazepam. Ogerin bears little similarity to any of the initial four benzodiazepine hits from the empirical screen. In fact, ogerin has a Tc of 0.188 to lorazepam, which is less than the 0.2 Tc one would expect at random. Indeed, by visual inspection lorazepam and ogerin are topologically very different (**Figure 3.8A,B**). No topology-based ligand similarity would find ogerin beginning with lorazepam.

How about something more sophisticated and shape-based, such as a pharmacophore model? A widely-used tool to do so is ZINCPharmer (http://zincpharmer.csb.pitt.edu).[6] We used pharmacophore models of greater and lesser precision based on features found in lorazepam, ranging from a simple model that only defined three hydrophobic rings, a hydrogen bond acceptor and a donor, to a more complex model with multiple hydrophobic areas/rings, 2 acceptors and 2 donors, and hydrophobic areas for the lorazepam halides. These

necessary features represent the amount of knowledge gained from the high-throughput screen that was also used as the starting point for the docking. Against these pharmacophores, we screened the 3.1 million lead-like molecules in ZINC. No molecules matched the more strict pharmacophore, and the only ones that matched the less restrictive pharmacophore closely resembled benzodiazepines in structure; nothing remotely resembling ogerin, or any of the other non-benzodiazepine hits, were accepted by the pharmacophore (**Figure 3.8C**). Different pharmacophores will return different hits, but here we have provided the maximum and minimum of possible results based on the yeast screening data. Pharmacophore models would not have discovered the ogerin family, given the large chemical space that was screened.

Not only is ogerin dissimilar to lorazepam topologically, but it explores both overlapping and different parts of the GPR68 binding pocket, and it is modeled to make different patterns of interactions within that site (**Figure 3.8D**). In addition, structure-based optimization of ogerin to **C2**, and the prediction of the reduced affinity of **C4**, was only enabled by the much greater information and context that structure-based docking affords, and it is impossible for ligand-based methods to recapitulate that when restricted to the information contained only in the benzodiazepines.

The last similarly tool we used was ROCS,[7] a more sophisticated shape-based and chemistry-based ligand similarity method to screen the same ZINC lead-like library, looking for how well ogerin and related molecules ranked against the shape defined by lorazepam. Even using optimized chemical-matching features,

ogerin ranks 268,645 out of 3.1 million. This reflects the same differences seen in

the pharmacophore search—ogerin and lorazepam are not only topologically

distinct, but they occupy only partially overlapping parts of the site, with which

they interact using different groups and even partly different residues (see ROCS-

based superposition in **Figure 3.8E**). Moreover, ogerin is in a completely different

conformation compared to the docked pose - this would not have allowed any type

of the rational optimization that lead to **C2**. This is no impediment to a structure-

based approach, such as docking, which incorporates all the information in the

active site, but it does preclude a ligand-based approach.
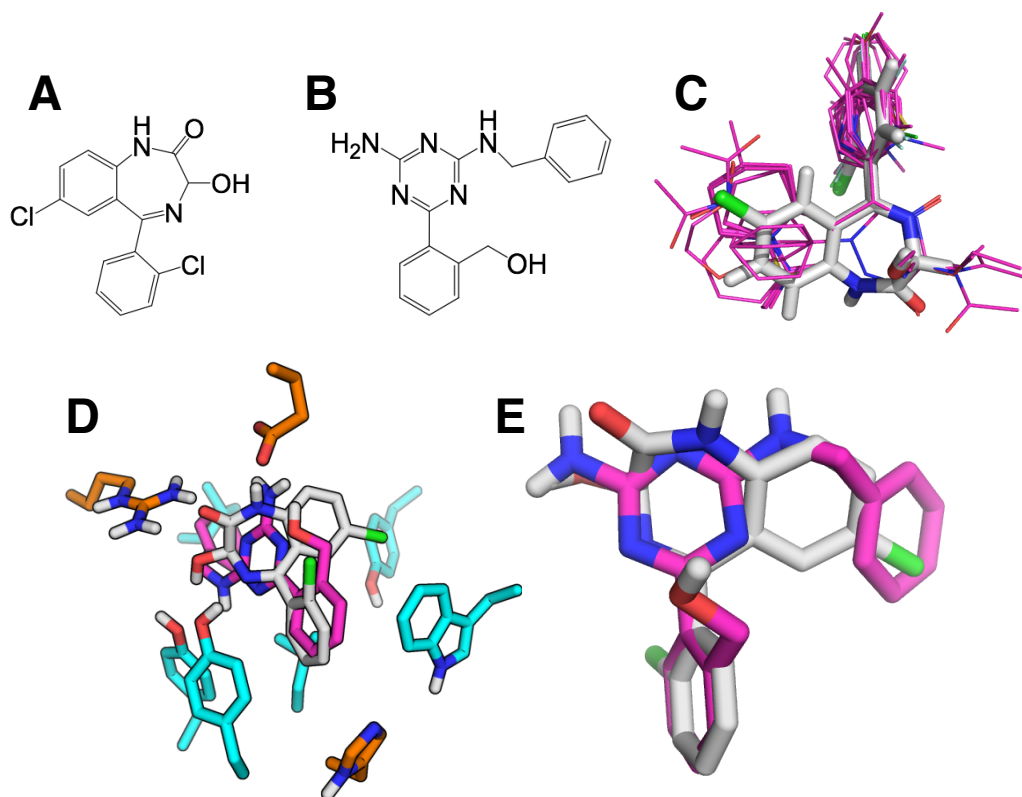
Figure 3.8 Comparison of various ligand-based methods for discovery of ogerin at GPR68. Simple topological comparison using Tanimoto coefficients shows no similarity between lorazepam (**A**) and ogerin (**B**). (**C**) Even loose pharmacophore searching returns hits (magenta) that resemble the benzodiazepine scaffold of lorazepam (gray) extremely closely. (**D**) Overlay of the docking poses of ogerin (magenta) and lorazepam (gray) at GPR68 (cyan) shows that while they partially overlap in the binding site, each explores a different section of the space and interacts with different residues to produce their different effects. (**E**) ROCS-optimized overlay of lorazepam (grey) and ogerin (magenta). Ogerin ranked 268,645 out of 3.1 million compounds, making it impossible to have been discovered in a shape-based virtual screen.

Another question involves the choice of modeling template: would we have wound up with the same results after modeling GPR68 based a template that wasn't CXCR4? To investigate that, we repeated the modeling and docking procedure using the δ-opioid receptor as the template for homology modeling. CXCR4 and the δ-opioid receptor have almost the same sequence identity to GPR68 (28% versus 25%), but distinctly different topologies, especially in the variable EL2 region. The δ-opioid receptor simply wasn't available to use at the time of the original docking campaign, and, with more GPCR crystal structures available, investigating if the choice of template matters is a crucial question for future studies. The final δ-opioid-based model overlaid on top of the CXCR4-based model, with optimized poses of lorazepam and ogerin, is given in **Figure 3.9A** and **B**. Here we see that both models interact with both ligands using many of the same residues, if not in exactly the same binding mode. The only difference is the position of Glu160 – this is due to the fact that, despite the overall seven transmembrane topology being similar between GPCRs, the EL2 is highly variable, both in length and in secondary structure elements. Since the other implicated residues are in very similar positions, this highlights the need for better loop sampling techniques, especially for longer loops.
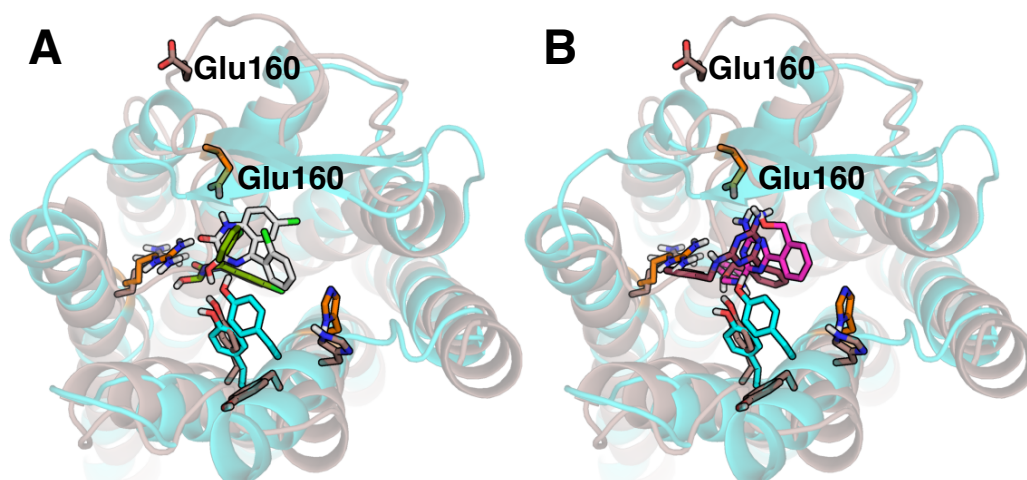
Figure 3.9 Comparison of binding modes of lorazepam and ogerin in CXCR4-based and δ-opioid-based GPR68 models. Lorazepam (**A**) and ogerin (**B**) were docked into the δ-opioid-based model (green and brown, respectively, with the model in brown) binds in the same area as lorazepam in the CXCR4-based model (gray and magenta, respectively, with the model in cyan). All interactions are the same, except for Glu160, which is on a random coil and pointing extracellularly in the δ-opioid-based model.

Finally, throughout this whole study we used non-covelent small molecules, but could covalent docking have also resulted in a successful ligand discovery campaign at an orphan GPCR? The first test is to see if a covalent ligand can be found that binds in the same orientation as its non-covalent analog. The choice of nucleophilic amino acid is based on a position at the top of TM2, which has been shown to be, when mutated to cysteine, useful as an anchoring point across several aminergic GPCRs.[8] In GPR68, this position is Tyr80. **Figure 3.10A** shows the ogerin-GPR68 model aligned to the β_2-adrenergic receptor in complex with a

131

covalent agonist (PDB: 3PDS). In it, ogerin seems to overlap nicely with the covalent agonist, supporting the choice of Tyr80 as a potentially useful nucleophile when mutated to cysteine. We then virtually generated electrophilic analogs of ogerin and covalently docked them, using DOCKovalent[9], to see if any of these molecules would have similar poses to the non-covalent analogs. One such covalent analog, forming a disulfide bond to Y80C, is shown in **Figure 3.10B**, and they overlap nicely. These potential covalent ligands are currently being synthesized and tested for GPR68 activity, as well as for their utility as stabilizing ligands for crystallographic studies.
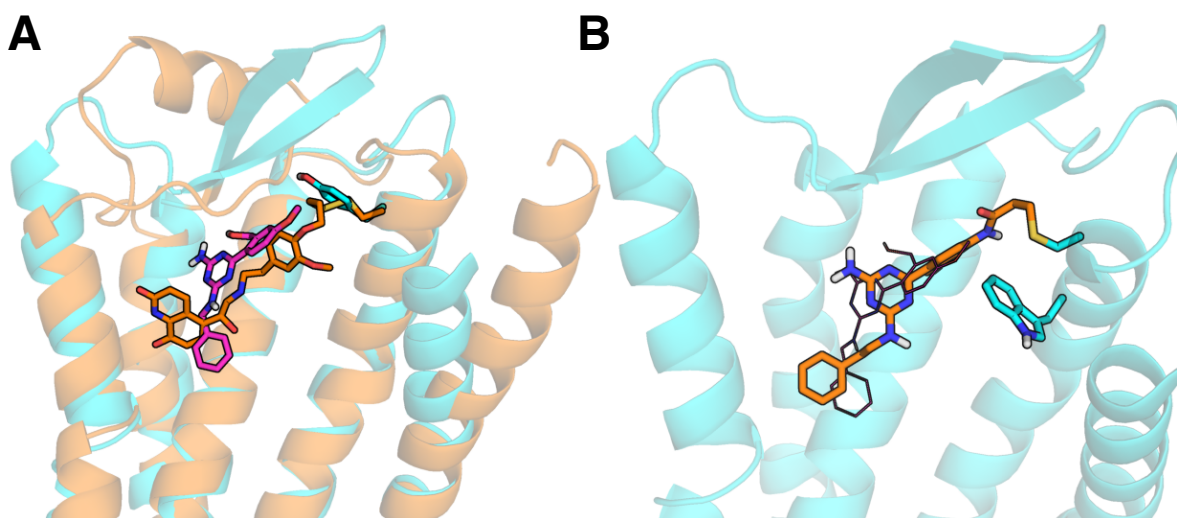
Figure 3.10 Covalent docking strategy for electrophilic ogerin analogs at GPR68. (**A**) The docked ogerin-GPR68 (magenta-cyan) pair is aligned to the β_2-adrenergic receptor in complex with a covalent agonist (orange). Ogerin overlaps with the β_2 agonist at the covalent connection point, identifying a strategy for covalent docking of ogerin analogs with electrophilic groups attached (**B**).

Now that there is some evidence of covalent ligands having similar binding modes as their non-covalent counterparts, what if we had initially done a covalent tethering screen of electrophilic ligands, instead of a yeast-based or TANGO high-throughput screen? Would covalent ligands provide the same amount of information about the binding site and also result in successful docking campaigns? Theoretically, they could be more useful, since it would be possible to easily confirm the covalent attachment point, and thus the binding site, either through MS/MS or mutagenesis studies, instead of running many docking simulations to convergence.

One hypothetical testing case is the chemotactic cytokine complement factor 5a receptor (C5AR1). This receptor has already had a tethering screen done against it in order to find map the binding site of these agonists and antagonists. In addition, C5AR1 also has many non-covalent ligands known. This could be a useful comparison case between models derived from covalent and non-covalent compounds, in order to benchmark and find the screening system that provides the most information and constraints for successful future modeling, and ultimately, successful virtual screens. This endeavor is currently underway.

3.5 Can we infer anything about endogenous orphan GPCR function?

Previously in Chapter 1, we used SEA to discover the reciprocal activity between adenosine receptor antagonists and the newly discovered GPR68 ligands. The obvious question that follows is: can we use these links to other receptors to infer any information about the orphan's endogenous ligand or function? For instance, from this link, we can look at further biologically relevant links: GPR68 and adenosine receptors are significantly differentially co-expressed in ischemic heart failure, Huntington's disease, and several brain and lung cancers, relative to non-disease states (**Supplementary Table 7).** We can also use SEA in reverse, but instead of screening compound libraries like the Tocris one in Chapter 1, we can virtually screen databases of known human metabolites, such as the Human Metabolome Database (HMDB), to look for endogenous ligands.[12] Comparing various combinations of discovered GPR68 ligands to the metabolites in the HMDB produced no significant hits, however.

Abandoning the ligand-based method for docking, however, reveals a new, orthogonal path; instead, we can dock the HMDB directly to the GPR68 model and look for potential endogenous ligands. The GPR68 homology model has already proven itself prospectively in docking virtual screens, both with the ZINC and Tocris databases, demonstrating the capability to recognize and enrich a diverse array of scaffolds. Perhaps unsurprisingly, after screening the HMDB, many of the top-ranked ligands with realistic poses were adenosine or adenosine-like. As seen in **Figure 3.11**, adenosine and lorazepam make similar interactions.

Preliminary results screening a small purinergic library against GPR68, however, suggest that none of these are GPR68 ligands; the hunt continues.
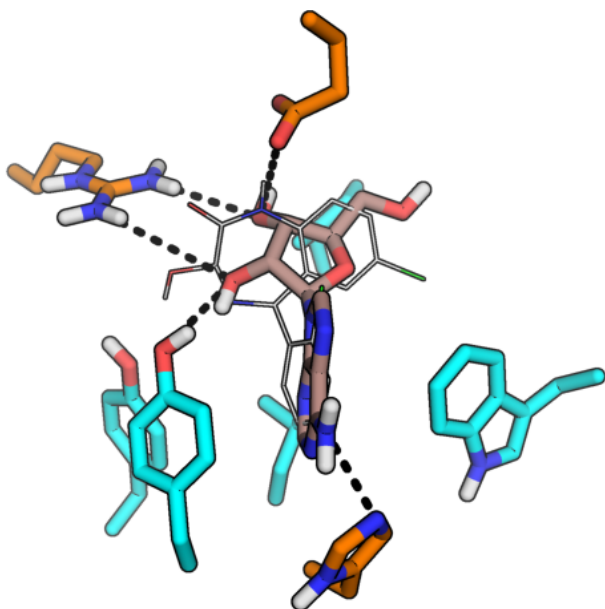
Figure 3.11 Docked pose of adenosine at GPR68. The docked pose of adenosine (brown) makes similar interactions as that of lorazepam (gray). Adenosine-like compounds were among the top-ranked classes of molecules in the HMDB screen.

Ultimately, there is a third option for endogenous ligand discovery: combining ligand-based methods and docking. It has been shown in previous benchmarking studies (*unpublished data*) that it is possible to use docking lists as the input sets to SEA, rather than sets of confirmed ligands for the target. In essence, docking hit lists provide the same information as annotated ligands, assuming a high-quality docking run, and in the case of poorly annotated targets, or orphans, they could be a valuable source of ligand information when no validated small molecule hits exist. Indeed, in aggregate, these clusters of ligands might provide more information for SEA than confirmed ligand singletons. Even if the docking is poor, similar recesptors and binding sites will erroneously enrich the same erroneous ligand chemotypes; pathological errors need only be consistent. Preliminary studies show that comparing docking hit lists with SEA not only

recapitulates known relationships between targets, it also predicts new ones. Novel shared ligands between multiple pairs of targets are currently being tested experimentally.

Instead of using docking hit lists as surrogate ligand sets for both targets, one can also compare a hit list to a set of annotated ligands. While this may assuage or ameliorate fears of false positives being represented as true ligands, using annotated ligands may actually exacerbate problems. Instead of errors from DOCK being consistent across all targets, now misinformation or noise may infiltrate the SEA signal. Also, true ligands may not feature as much diversity, since, depending on the source, they may just represent historical data or medicinal chemistry efforts. Many molecules that might inhabit the ligand chemistry space for these targets have just not been tested; these may appear as enriched in the docked hit list, however.

Brief, preliminary tests have looked into these relationships at GPCRs. Here, we docked the ChEMBL14 database at GPR68 and other available GPCR crystal structures (the neurotensin 1, $\mu/\kappa/\delta$-opioid, B$_1$- and B$_2$-adrenergic, A$_{2A}$ adenosine, H$_1$ histamine, D$_3$ dopamine, CXCR$_4$ chemokine, and M$_2$ and M$_3$ muscarinic receptors) and then we compared both their docked hit lists and sets of annotated ligands with SEA. ChEMBL14 was chosen as the database to dock due to the fact that the ligands included are both diverse and comprehensive, spanning a range of biological targets, although these sets are subject to the caveats mentioned above. Additionally, we used several bin sizes of the molecules that scored the best in order to look for consistency in SEA predictions. We also chose to examine

only the subset of ChEMBL14 that contains molecules with lead-like properties, since: a) it is these that we mostly concern ourselves with in virtual screens, and b) DOCK has a bias towards compounds with higher molecular weights, due to the increased van der Waals component of the overall docking score. Since these receptor docking setups were optimized for high ligand enrichment, as a sanity check it is unsurprising that significant SEA scores were measured between each receptor's hit list and its own annotated ligand set. When various molecular weight-binned docking hit lists of GPR68 were compared to those of the other GPCR crystal structures, the only consistently significant relationship found was again to A_{2A}.

To make this generally useful, however, two hurdles would need to be cleared. The first regards the absence of a ligand to optimize the homology model. While comparing SEA hit lists theoretically does not require knowledge of true ligands, the generation of an accurate, or capable, homology model using the procedure in Chapter 1 relies on it. To get around this, one can potentially compare the ChEMBL docked lists for a multitude of models and choose one that consistently has a significant relationship to another receptor, or, if several exist, choose a representative structure from a cluster that all associate with a particular target.

The second hurdle is that this method requires dockable models for all GPCRs. Not only that, but due to different ligands' abilities to stabilize different conformations in the overall accessible receptor equilibrium, multiple structures would be required for each receptor. For those without crystal structures,

modeling these receptor states is a total necessity, but this cannot be done manually for so many targets. The biggest challenge for an automatic modeling pipeline for these is the subjective, ligand pose-evaluation step, required for final model optimization and choosing. While for GPR68 the top-ranked receptor by benzodiazepine enrichment fortuitously also featured lorazepam bound in a reasonable pose, this was not the case for GPR65. Instead, the final model chosen was actually ranked 20th by enrichment of BTB09089 – the previous 19 models were not ranking BTB09089 highly for the right reasons. So, we tried simple re-scoring metrics in an attempt to rescue the chosen GPR65 model and extrapolate these lessons learned to a new general scoring scheme. These involved accounting for the amount of buried surface area upon ligand binding, penalties for unsatisfied hydrogen bonds, and the number of critical residue contacts as determined by mutagenesis studies. Using these new penalties and bonuses allowed the final, chosen GPR65-BTB09089 model to be ranked 3rd.

Examination of the two models ranked above this one exposed areas where more work has to be done. **Figure 3.12A** shows the top-ranked model by docking enrichment. While it appears to be making favorable interactions with the receptor, it is completely flipped and does not make any hydrogen bonds with residues implicated in the related receptor, GPR68. To fix this, we have made a database of residues implicated, by mutagenesis, in ligand binding to each GPCR, separated into those important for agonist and antagonist binding. During re-scoring, a list of the pose's receptor contacts can be made and then the ligands are filtered based on the presence or absence of these important interactions.

Figure 3.12B shows the second top-ranked model by docking enrichment. While it is quite similar to the final BTB09089 pose, the dichlorobenzyl ring is floating off into solvent, instead of making a nice pi-pi stacking arrangement with Trp70. Interactions beyond typical hydrogen bonds, van der Waals, and charge-charge pairings are lost in the DOCK scoring function; incorporation of these will automatically further enrich ligand poses that manual labor routinely favors.
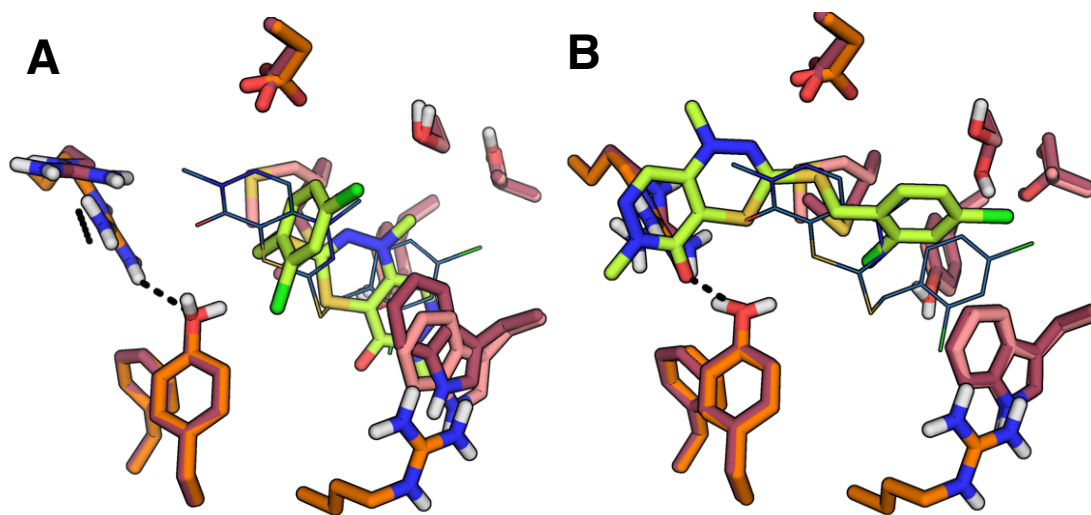
Figure 3.12 Top-scored poses of BTB09089 at GPR65 models. (**A**) The top-ranked GPR65-BTB09089 model is flipped and does not take into consideration interactions with residues presumed to be important for binding. (**B**) The GPR65-BTB09089 model ranked second misses the opportunity for pi-pi stacking between Trp70 and the dichlorobenzyl substituent.

This method offers a tantalizing prospect of an automatic modeling and de-orphanization pipeline (pipedream?). Unlike subjective, manual hit-picking (as used in Chapters 1 and 2) of docking screens to identify potential ligands for testing, based on what, nebulously, looks "good," this type of analysis is both

objective and independent of prior annotations. While fraught with pitfalls, each new test case will bring up more improvements that can be applied. To better understand where other deficiencies lie in the pipeline, we can undertake parallel modeling campaigns and see where the paths diverge.

Finally, these dockable GPCR models can be applied in another way besides de-orphanization and function determination. While grand conjecture, one can envision harmoniously combining the ideas of Chapter 1 and 2 into a transparent, automatic pipeline for polypharmacology. Two uses for this emerge. The first assumes a desired GPCR binding profile, which can be accomplished using the protocol from Chapter 2 and assuming a sufficiently large docking library. Any such imaginary ligand can be discovered for any malady. The second does not require such a preconceived notion. In order to discover the receptor profile responsible for a clinical phenotype, one can design ligands for each hypothetical combination and experimentally test them for clinical efficacy. These are the goals.

Science is magic that works.

3.6 References

1. Allen, J.A. et al. Discovery of ß-arrestin-biased dopamine D_2 ligands for probing signal transduction pathways essential for antipsychotic efficacy. Proc Natl Acad Sci USA 105, 18488-93 (2011).

2. Chen, X. et. al. Structure-functional selectivity relationship studies of ß-arrestin-biased dopamine D_2 receptor agonists. J Med Chem 55, 7141-53 (2012).

3. Chien, E.Y. et. al. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. Science 330, 1091-5 (2010).

4. Liu, W. et. al. Serial femtosecond crystallography of G protein-coupled receptors. Science 342, 1521-4 (2013).

5. Kroeze, W.K. et. al. An open-source resource for the simultaneous and parallel interrogation of the druggable GPCR-ome. In review.

6. Koes, D.R. and Camacho, C.J. ZINCPharmer: pharmacophore search of the ZINC database. Nucleic Acids Res 40, W409-14 (2012).

7. Hawkins, P.C.D et. al. A comparison of shape-matching and docking as virtual screening tools. J Med Chem 50, 74-82 (2007).

8. Weichert, D. et. al. Covalent agonists for studying G protein-coupled receptor activation. Proc Natl Acad Sci USA 111, 19744-8 (2014).

9. London, N. et al. Covalent docking of large libraries for the discovery of chemical probes. Nat Chem Biol 10, 1066-72 (2014).

10. Buck, E. and Wells, J.A. Disulfide trapping to localize small-molecule agonists and antagonists for a G protein-coupled receptor. Proc Natl Acad Sci USA 102, 2719-2724 (2005).

11. Buck, E., Bourne, H., and Wells, J.A. Site-specific disulfide capture of agonist and antagonist peptides on the C5a receptor. J Biol Chem 280, 4009-12 (2005).

12. Wishart, D.S. et. Al. HMDB: the Human Metabolome Database. Nucleic Acids Res 35, D521-6 (2007).

A.1 Supplementary material for Chapter 1

I. Supplementary methods

I.A Functional assays with A_{2A} and CXCR4 receptors

Functional assays with A_{2A} adenosine and CXCR4 chemokine receptors were carried out using a slightly different protocol from that previously described for G_s and G_i receptors (Besnard et al, 2012) and as outlined in the Methods section. Specifically, HEK293-T cells were transfected and plated as in the Methods using regular DMEM supplemented with 1% dialyzed FBS. Before assays, culture medium was removed, and cells were incubated with 20 μl drug solution (prepared in drug buffer 20 mM HEPES, 1x HBSS, pH 7.4) for 15 min at room temperature. To measure agonist activity, 5 μl of 5x luciferin solution (4 mM final concentration) for A_{2A} (G_s coupled GPCRs) or a mixture of luciferin and isoproterenol at a final concentration of 200 nM for CXCR4 (G_i coupled GPCRs) was added and cells were incubated for another 20 min. To measure antagonist activity, test compound was added first for 10 min before a reference agonist at a final of EC_{80} concentration for another 10 min, and then followed by addition of luciferin for A_{2A} or a mixture of luciferin and isoproterenol for CXCR4 as above. Luminescence was counted in a luminescence counter. Results were analyzed in GraphPad Prism.

I.B Calcium mobilization assays

HEK293-T cells were transfected and plated into poly-L-Lys coated 384-well black clear bottom cell culture plates in DMEM supplemented with 1% dialyzed FBS, at a density of 15,000 cells in 40 μl per well for overnight. Before the assay, medium was removed and cells were loaded with Fluo-4 Direct calcium dye (Invitrogen, CA) for 60 min at 37°C in a 5% CO_2 atmosphere. The calcium dye was prepared in drug buffer supplemented with 2.5 mM probenecid (pH 8.0). Proton solutions were made with 1x HBSS, 7 mM HEPES, 7 mM HEPPS, and 7 mM MES and pH was adjusted with NaOH. Drug additions and fluorescence intensity measurement were carried out in a FLIPR^{TETRA}, which was programmed to add drug solutions to cells while recording fluorescence intensity. To measure proton concentration-responses, 10 μl of pH pre-determined solutions were added to each well (with 20 μl Calcium dye) while fluorescence intensity was recorded during and after addition for 4 minutes (one reading per second). The addition procedure was configured in such a way (30 μl per second at height of 10 μl above cells) that local proton concentrations for cells were essentially the same as in the pH working solutions at the moment of addition. Fluorescence intensities reached peak values within 30 seconds after drug addition. To determine modulator effects on proton responses, the protocol was modified slightly. In brief, cells were loaded with calcium dye as above, but only at 15 μl per well. The FLIPR^{TETRA} was programmed to first add 5 μl of 4x test compound (final concentration of 10 μM before addition of 10 μl of pH solutions) prepared with the same drug buffer at pH 8.0 (buffer alone served as a control). After a total of 10 min of reading and incubation, 10 μl

of the pH solutions were added and the fluorescence intensity was recorded exactly the same way as above. Results (fluorescence intensity in fold of basal) were exported and analyzed in GraphPad Prism. For calcium mobilization assays with 5-HT$_{2B}$ receptors, HEK293 cells stably expressing human 5-HT$_{2B}$ receptors were used instead of transiently transfected cells. Cells were set up and tested in the same way as above, with 5-HT serving as an agonist control (3 pM – 30 μM) and with 1 nM 5-HT being used in the second addition to determine the antagonist activity of ogerin.

I.C Radioligand binding assays

Radioligand binding assays with selected CNS targets were carried out as described (Besnard et al, 2012; Keiser et al, 2009) and as detailed in the PDSP protocol book available online (http://pdsp.med.unc.edu/pdspw/binding.php). Briefly, receptor membrane preparations were made from either animal brain tissues, or stable cell lines, or transiently transfected HEK293-T cells. Receptor expression levels and radioligand binding affinities were determined with saturation binding assays. Competition binding assays were performed with membrane aliquots and a fixed concentration of radioligand in 96-well plates in a final volume of 125 μl. Reactions were incubated in the dark and at room temperature (22°C), and terminated by vacuum filtration onto 96-well formatted GF/B filters. Radioactivity on the filters was counted in a beta counter. Results were analyzed in GraphPad Prism.

I.D Anti-Flag immunoblots

HEK293-T cells were transiently transfected in 10-cm dishes with Flag-tagged GPR68 wild-type and mutant receptors. Untransfected HEK293-T cells served as a negative control. After 48 hours, cells were collected, lysed, and sonicated to shear chromatin before being subjected to immunoblotting. Blots were probed with anti-Flag M2-peroxidase antibody (from Sigma). Bands were quantified and normalized to GPR68 wild-type receptor (fold) for bar graphing.

I.E Supplementary Chemistry procedures for ogerin and its analogs

HPLC spectrum was acquired using an Agilent 6110 Series system with UV detector set to 254 nm. Samples were injected (5 µL) onto an Agilent Eclipse Plus 4.6 x 50 mm, 1.8 µM, C18 column at room temperature. A linear gradient from 10% to 100% B (MeOH + 0.1% acetic acid) in 5.0 min was followed by pumping 100% B for another 2 minutes with A being H_2O + 0.1% acetic acid. The flow rate was 1.0 mL/min. Mass spectrum (MS) data was acquired in positive ion mode using an Agilent 6110 single quadrupole mass spectrometer with an electrospray ionization (ESI) source. High-resolution (positive ion) mass spectrum (HRMS) was acquired using a Thermo LTqFT mass spectrometer under FT control at 100,000 resolution. Nuclear Magnetic Resonance (NMR) spectra were recorded at Varian Mercury spectrometer with 400 MHz for proton (^{1}H NMR) and 100 MHz for carbon (^{13}C NMR). Chemical shifts are reported in ppm (δ).[3]
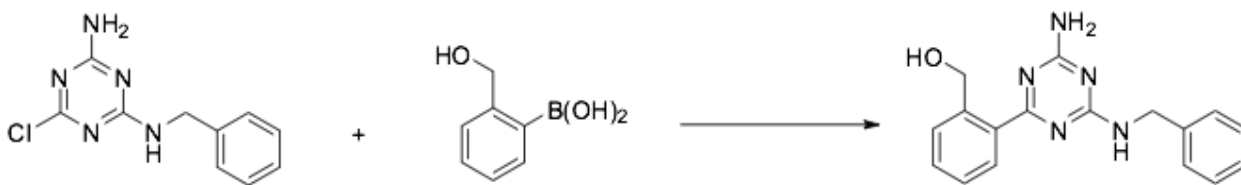
(2-(4-Amino-6-(benzylamino)-1,3,5-triazin-2-yl)phenyl)methanol. To a suspension of N^2-benzyl-6-chloro-1,3,5-triazine-2,4-diamine[4] (1014.3 mg, 4.31 mmol) and (2-(hydroxymethyl)phenyl)boronic acid (1308 mg, 8.6 mmol) in dioxane (4 mL) and H_2O (1mL) was added $Pd(PPh_3)_4$, followed by K_2CO_3. The resulting mixture was then stirred under microwave irradiation at 110^oC for 20 minutes. Removal of solvent gave the residue which was subjected to silica gel column chromatography to give compound (2-(4-amino-6-(benzylamino)-1,3,5-triazin-2-yl)phenyl)methanol as white solid (1179.9 mg, yield 89%). ^{1}H NMR (400 MHz, CD_3OD) δ 7.97 – 7.88 (m, 1H), 7.46 (d, J = 3.9 Hz, 2H), 7.42 – 7.27 (m, 5H), 7.23 (t, J = 7.1 Hz, 1H), 4.65 (s, 2H), 4.62 (s, 2H). ^{13}C NMR (101 MHz, CD_3OD) δ 174.20, 168.19, 166.99, 141.52, 140.58, 138.59, 131.65, 131.39, 130.91, 130.58, 129.46, 128.73, 128.43, 128.03, 64.59, 45.03. HPLC: 99%, RT 4.976 min, MS (ESI) m/z 308.2 $[M + H]^+$. HRMS m/z $[M + H]^+$ calcd for $C_{17}H_{18}N_5O$ 308.1511, found 308.1506. mp: 154-155 ^oC.
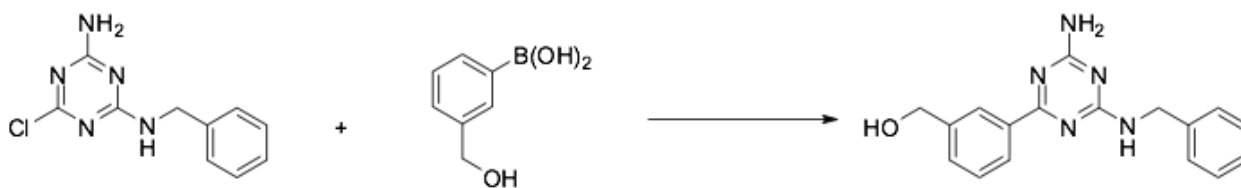
(3-(4-Amino-6-(benzylamino)-1,3,5-triazin-2-yl)phenyl)methanol. The target compound (white solid, 83.7mg, yield 90%) was prepared by the same procedure as preparing (2-(4-amino-6-(benzylamino)-1,3,5-triazin-2-yl)phenyl)methanol from N^2-benzyl-6-chloro-1,3,5-triazine-2,4-diamine[4] (71.1 mg, 0.30 mmol) and (3-

147

(hydroxymethyl)phenyl)boronic acid (68.8 mg, 0.45 mmol). ^{1}H NMR (400 MHz, CD$_3$OD) δ 8.26 (s, 1H), 8.16 (s, 1H), 7.51 (d, J = 7.6 Hz, 1H), 7.46 – 7.26 (m, 5H), 7.22 (t, J = 7.2 Hz, 1H), 4.75 – 4.50 (m, 7H). ^{13}C NMR (101 MHz, CD$_3$OD) δ 172.58, 168.99, 167.78, 142.84, 140.79, 138.27, 130.96, 129.43, 129.26, 128.47, 128.32, 128.21, 127.97, 127.84, 65.07, 45.03. δ 174.20, 168.19, 166.99, 141.52, 140.58, 138.59, 131.65, 131.39, 130.91, 130.58, 129.46, 128.73, 128.43, 128.03, 64.59, 45.03. HPLC: 99%, RT 5.097 min, MS (ESI) m/z 308.2 [M + H]$^+$. HRMS m/z [M + H]$^+$ calcd for C$_{17}$H$_{18}$N$_5$O 308.1511, found 308.1508. mp: 128-129 ^{o}C.
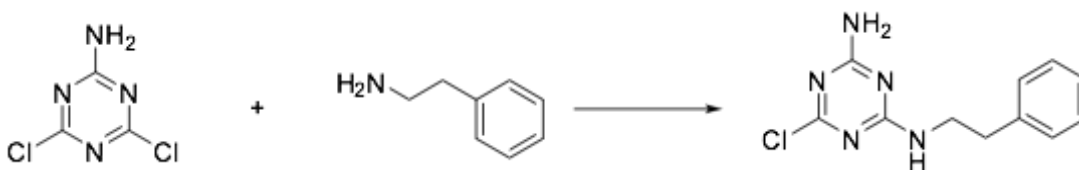
6-Chloro-N^2-phenethyl-1,3,5-triazine-2,4-diamine To a solution of 4,6-dichloro-1,3,5-triazin-2-amine (144.1 mg, 0.87 mmol) in dioxane (5 mL) was added 3-phenylpropan-1-amine (0.12 mL, 0.96 mmol), followed by DIPEA (0.39 mL, 2.18 mmol). The resulting mixture was then refluxed for 4 hours. After cooling down, solvent was removed under vacuum and the residue was subjected to silica gel chromatography to give the target compound as white solid (202.1 mg, yield 93%). ^{1}H NMR (400 MHz, CD$_3$OD) δ 8.25 (s, 1H), 7.71 – 7.60 (m, 1H), 7.61 – 7.50 (m, 1H), 7.48 – 7.17 (m, 3H), 4.76 – 4.55 (m, 2H), 3.37 – 3.30 (m, 2H). HPLC: 99%, RT 4.921 min, MS (ESI) m/z 250.2 [M + H]$^+$.
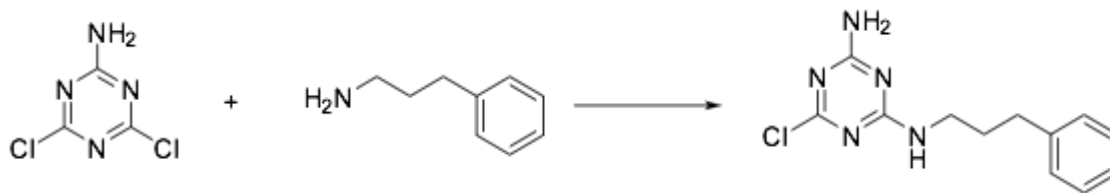
6-Chloro-N^2-(3-phenylpropyl)-1,3,5-triazine-2,4-diamine The target compound (white solid, 293.6 mg, yield 91%) was prepared by the same procedure as preparing 6-chloro-N^2-phenethyl-1,3,5-triazine-2,4-diamine from 4,6-dichloro-1,3,5-triazin-2-amine (201.8 mg, 1.22 mmol), 3-phenylpropan-1-amine (0.19 mL, 1.35 mmol), and DIPEA (0.54 mL, 3.06 mmol) in dioxane (6 mL). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.28 (m, 2H), 7.19 (m, 3H), 5.64 – 5.14 (m, 3H), 3.47 (dd, J = 13.1, 6.5 Hz, 1H), 3.39 (dd, J = 13.5, 6.6 Hz, 1H), 2.75 – 2.62 (m, 2H), 1.90 (m, 1.98 – 1.86, 2H). HPLC: 99%, RT 5.041 min, MS (ESI) m/z 264.2 [M + H]$^+$.
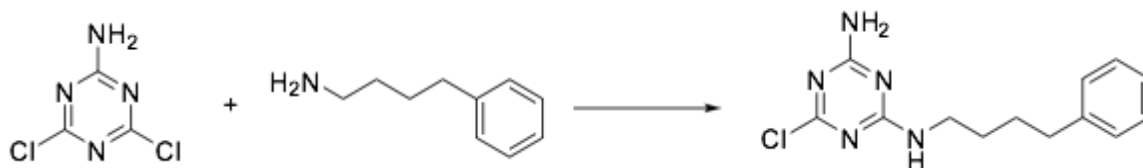
6-Chloro-N^2-(4-phenylbutyl)-1,3,5-triazine-2,4-diamine The target compound (white solid, 295.3 mg, yield 87%) was prepared by the same procedure as preparing 6-chloro-N^2-phenethyl-1,3,5-triazine-2,4-diamine from 4,6-dichloro-1,3,5-triazin-2-amine (200mg, 1.21 mmol), 4-phenylbutan-1-amine (0.21 mL, 1.33 mmol), and DIPEA (0.54 mL, 3.06 mmol) in dioxane (6 mL). ^{1}H NMR (400 MHz, CDCl$_3$) δ δ 7.35 – 7.23 (m, H), 7.18 (t, J = 9.3 Hz, 3H), 5.88 – 5.07 (m, 2H), 3.49 – 3.29 (m, 2H), 3.16 – 3.00 (m, 2H), 2.64 (t, J = 7.3 Hz, 1H), 1.41 (t, J = 7.3 Hz, 1H). HPLC: 99%, RT 5.207 min, MS (ESI) m/z 278.2 [M + H]$^+$.

149
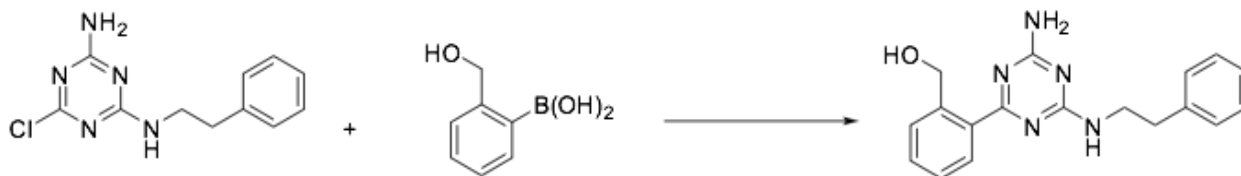
(2-(4-Amino-6-(phenethylamino)-1,3,5-triazin-2-yl)phenyl)methanol   The   target compound (white solid, 56.4 mg, yield 91%) was prepared by the same procedure as preparing (2-(4-amino-6-(benzylamino)-1,3,5-triazin-2-yl)phenyl)methanol from 6-chloro-N^2-phenethyl-1,3,5-triazine-2,4-diamine (65.3 mg, 0.25 mmol) and (3-(hydroxymethyl)phenyl)boronic acid (76 mg, 0.50 mmol). ^{1}H NMR (400 MHz, CD$_3$OD) δ 7.94 (dd, *J* = 50.1, 7.4 Hz, 1H), 7.69 – 7.51 (m, 1H), 7.50 – 7.34 (m, 3H), 7.28 (d, *J* = 6.5 Hz, 4H), 7.18 (s, 1H), 4.68 (d, *J* = 37.8 Hz, 2H), 3.63 (t, *J* = 7.5 Hz, 2H), 2.91 (t, *J* = 7.4 Hz, 2H). HPLC: 99%, RT 5.193 min, MS (ESI) m/z 322.2 [M + H]$^+$.
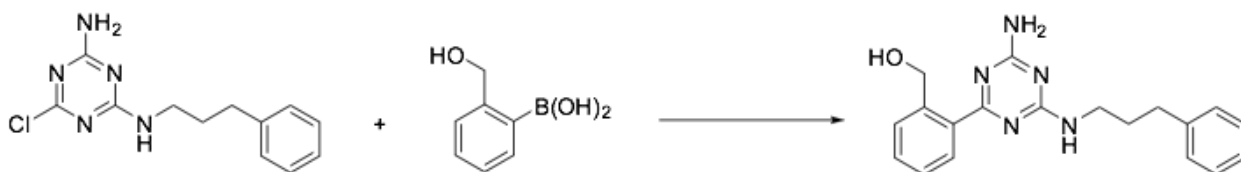
(2-(4-Amino-6-((3-phenylpropyl)amino)-1,3,5-triazin-2-yl)phenyl)methanol   The target compound (white solid, 68.3 mg, yield 92%) was prepared by the same procedure   as   preparing   (2-(4-amino-6-(benzylamino)-1,3,5-triazin-2-yl)phenyl)methanol from 6-chloro-N^2-(3-phenylpropyl)-1,3,5-triazine-2,4-diamine (54.6 mg, 0.21 mmol) and (3-(hydroxymethyl)phenyl)boronic acid (62.8 mg, 0.42 mmol). ^{1}H NMR (400 MHz, CD$_3$OD) δ 7.92 (dd, *J* = 34.4, 7.4 Hz, 1H), 7.46 (d, *J* = 3.8 Hz, 2H), 7.44 – 7.35 (m, 1H), 7.30 – 7.09 (m, 5H), 3.40 (t, *J* = 6.5 Hz, 2H), 3.37 – 3.25 (m, 2H), 2.68 (t, *J* = 7.7 Hz, 2H), 1.99 – 1.85 (m, 2H). ^{13}C NMR (101 MHz, CD$_3$OD)

δ 173.99, 166.73, 143.07, 141.39, 138.53, 131.63, 131.25, 130.89, 129.41, 129.37, 128.76, 126.84, 111.40, 73.48, 68.96, 64.52, 62.22, 54.48, 43.98, 41.23, 34.20, 32.62, 32.16, 18.33. HPLC: 99%, RT 5.286 min, MS (ESI) m/z 336.2 [M + H]$^+$.
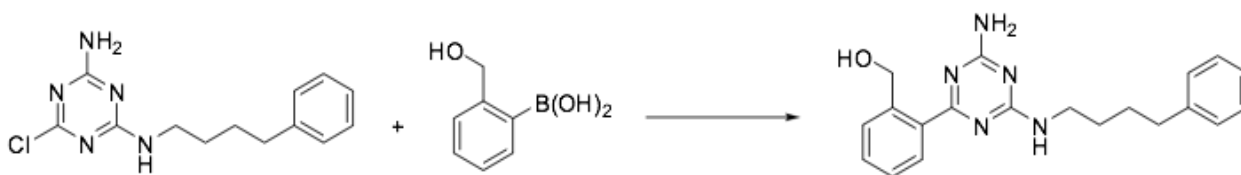
(2-(4-Amino-6-((4-phenylbutyl)amino)-1,3,5-triazin-2-yl)phenyl)methanol   The target compound (white solid, 59.7 mg, yield 85%) was prepared by the same procedure as preparing (2-(4-amino-6-(benzylamino)-1,3,5-triazin-2-yl)phenyl)methanol from 6-chloro-N^2-(3-phenylpropyl)-1,3,5-triazine-2,4-diamine (54.8 mg, 0.20 mmol) and (3-(hydroxymethyl)phenyl)boronic acid (60.5 mg, 0.40 mmol). ^{1}H NMR (400 MHz, CD$_3$OD) δ 7.92 (dd, J = 37.6, 7.4 Hz, 1H), 7.46 (d, J = 4.2 Hz, 2H), 7.39 (dq, J = 8.5, 4.3 Hz, 1H), 7.30 – 7.03 (m, 5H), 3.40 (t, J = 6.5 Hz, 2H), 3.31 (dt, J = 3.2, 1.6 Hz, 2H), 2.64 (t, J = 6.8 Hz, 2H), 1.74 – 1.52 (m, 4H). ^{13}C NMR (101 MHz, CD$_3$OD) δ 173.98, 168.07, 166.68, 143.59, 141.37, 138.53, 131.62, 131.23, 130.88, 130.64, 129.40, 129.27, 128.76, 126.70, 73.48, 64.52, 62.22, 41.57, 41.34, 36.51, 30.26, 29.94, 29.80. HPLC: 99%, RT 5.369 min, MS (ESI) m/z 350.2 [M + H]$^+$.

Brief procedures for other GPR68 ligands. [a]Reagents and conditions: a) ethoxyethene, pyridium p-toluenesulfonate, DCM, 0°C to r.t., 2h, 85-87%; b) i) BuLi, THF, -78°C, 10min; ii) B(OCH$_3$)$_3$, -78°C to r.t., iii) 1N HCl, r.t., 1h, 70-74%; c) NaBH$_4$, CH$_3$OH, r.t., 30 min, 89%; d) Bis(pinacolato)diboron, KOAc, Pd(dppf)$_2$Cl$_2$,

dioxane, reflux, 6h, 78-82%; e) (4-methoxyphenyl)methanamine, DIEA, dioxane, reflux, 4h, 88%; f) (2-(hydroxymethyl)phenyl)boronic acid, Pd(PPh$_3$)$_4$, K$_2$CO$_3$, dioxane, H$_2$O, Microwave, 110°C, 15min, 88-91%; g) (1H-pyrazol-3-yl)methanamine, DIEA, dioxane, reflux, 4h, 83%; h) phenylmethanamine, DIEA, dioxane, reflux, 4h, 89%; i) 3, Pd(PPh$_3$)$_4$, K$_2$CO$_3$, dioxane, H$_2$O, Microwave, 110°C, 15min, 90%; j) (4-(hydroxymethyl)phenyl)boronic acid, Pd(PPh$_3$)$_4$, K$_2$CO$_3$, dioxane, H$_2$O, Microwave, 110°C, 15min, 91%; k) 7, Pd(PPh$_3$)$_4$, K$_2$CO$_3$, dioxane, H$_2$O, Microwave, 110°C, 15min, 82-86%; l) (2-fluorophenyl)methanamine, DIEA, dioxane, reflux, 4h, 82%; m) (2-amino-4,5-difluorophenyl)boronic acid, Pd(PPh$_3$)$_4$, K$_2$CO$_3$, dioxane, H$_2$O, Microwave, 110°C, 15min, 79-81%; n) 9, Pd(PPh$_3$)$_4$, K$_2$CO$_3$, dioxane, H$_2$O, Microwave, 110°C, 15min, 75-78%; o) 11, Pd(PPh$_3$)$_4$, K$_2$CO$_3$, dioxane, H$_2$O, Microwave, 110°C, 15min, 76-80%; p) (3-fluorophenyl)methanamine, DIEA, dioxane, reflux, 4h, 85%.

153
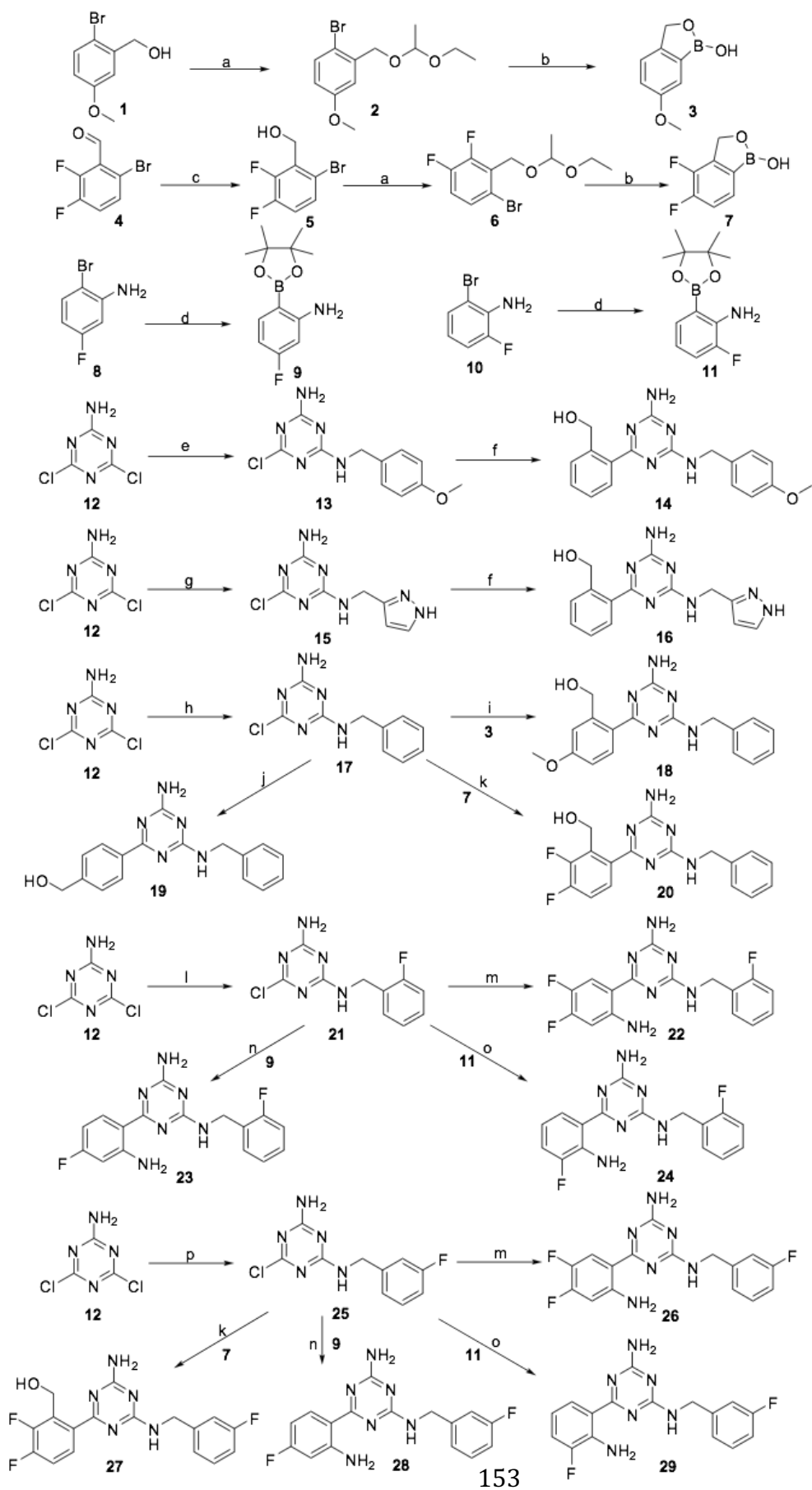
I.F Supplementary in vivo methods and behavioral profiles of GPR68 KO mice

The goal of this study was to determine whether targeted deletion of GPR68 (also known as *Ogr1*; ovarian cancer G protein coupled receptor 1) alters behavioral function in mice. Subjects were 21 wild type (WT) mice (9 males and 12 females) and 18 GPR68 knockout (KO) mice (7 males and 11 females), on a C57BL/6 background. Testing began when animals were 6-7 weeks of age. For each procedure, measures were taken by an observer blind to mouse genotype (WT or KO). Data were analyzed using one-way or repeated measures Analysis of Variance (ANOVA). Fisher's protected least-significant difference (PLSD) tests were used for comparing group means only when a significant F value was determined. Within-group comparisons were conducted to determine side preference in the social behavior tests. For all comparisons, significance was set at $p < 0.05$.

1.F.i Timeline for behavioral tests

Age of mice	Test
6-7 weeks	Elevated plus maze test for anxiety-like behavior.
7-8 weeks	Activity in an open field. Accelerating rotarod (2 tests, 48 hours apart).
8-9 weeks	Three-chamber social approach test. Activity in an open field (re-test).
9-10 weeks	Marble-burying assay.
10-11 weeks	Acoustic startle test. Buried food test for olfactory ability.

11-12 weeks	Visual cue test in the Morris water maze.
12-14 weeks	Hidden platform test for spatial learning.
14-16 weeks	Reversal learning in the Morris water maze.
16-17 weeks	Second acoustic startle test. Hotplate test for thermal sensitivity.

1.F.ii Summary of results

Mice with deletion of GPR68 had normal performance in most of the behavioral tests. No effects of genotype were observed for body weights, activity and anxiety-like behavior in an elevated plus maze or an open field, motor coordination, sociability, prepulse inhibition of acoustic startle responses, or acquisition in the water maze. However, both male and female GPR68 knockout mice had small, significant decreases in acoustic startle responses, suggesting a reduced responsivity to environmental stimuli. Male GPR68 knockout mice also showed significant decreases in marble burying, a test for anxiety-like phenotypes. Overall, the findings indicate that GPR68 might play a role in specific domains of behavior.

1.F.iii Elevated plus maze

This test is used to assess anxiety–like behavior in rodents. The procedure is based on a natural tendency of mice to actively explore a new environment, versus a fear of being in an open area. In the present study, mice were given one five-

155

minute trial on the plus maze, which had two walled arms (the closed arms, 20 cm in height) and two open arms. The maze was elevated 50 cm from the floor, and the arms were 30 cm long. Animals were placed on the center section (8 cm x 8 cm), and allowed to freely explore the maze. Measures were taken of time on, and number of entries into, the open and closed arms. All of the experimental groups showed a strong preference for the closed arms, in comparison to the open arms, of the elevated plus maze. As shown in **Supplementary Table 9**, there were no significant differences between the wild type and GPR68 KO mice for percent time or percent entries on the open arms, or for total entries during the task.

1.F.iv Marble-burying assay

This procedure is used to evaluate anxiety-like behavior and repetitive responses. Mice were tested in a Plexiglas cage located in a sound-attenuating chamber with ceiling light and fan. The cage contained 5 cm of corncob bedding, with 20 black glass marbles (14 mm diameter) arranged in an equidistant 5 X 4 grid on top of the bedding. Animals were given access to the marbles for 30 min. Measures were taken of the number of buried marbles (two thirds of the marble covered by the bedding). A two-way ANOVA indicated a significant genotype x sex interaction [$F(1,35)=7.37$, $p=0.0102$] (**Supplementary Table 9**). Post-hoc comparisons revealed that the male GPR68 KO mice buried significantly fewer marbles than both male WT mice and female KO mice in this task.

1.F.v Buried food test for olfactory function

Several days before the olfactory test, an unfamiliar food (Froot Loops, Kellogg Co., Battle Creek, MI) was placed overnight in the home cages of the mice. Observations of consumption were taken to ensure that the novel food was palatable. Sixteen to twenty hours before the test, all food was removed from the home cage. On the day of the test, each mouse was placed in a large, clean tub cage (46 cm L x 23.5 cm W x 20 cm H), containing paper chip bedding (3 cm deep), and allowed to explore for five minutes. The animal was removed from the cage, and one Froot Loop was buried in the cage bedding. The animal was then returned to the cage and given fifteen minutes to locate the buried food. Measures were taken of latency to find the food reward. As shown in **Supplementary Table 9**, there were no significant differences between the groups in latency to find the buried food.

1.F.vi Hotplate test for thermal sensitivity

Individual mice were placed in a tall plastic cylinder located on a hotplate, with a surface heated to 55°C (IITC Life Science, Inc., Woodland Hills, CA). Reactions to the heated surface, including hindpaw lick, vocalization, or jumping, led to immediate removal from the hotplate. Measures were taken of latency to respond. The maximum test length was 30 sec, to avoid paw damage. A two-way ANOVA indicated a significant main effect of sex [$F(1,1)=8.83$, $p=0.0053$], and genotype x sex interaction [$F(1,35)=4.3$, $p=0.0455$] (**Supplementary Table 9**). Post-hoc

comparisons revealed that the male GPR68 KO mice had significantly lower latencies to respond than female KO mice.

1.F.vii Acoustic startle method

The acoustic startle test can be used to assess auditory function and sensorimotor gating. The test is based on the measurement of the reflexive whole-body flinch, or startle response, that follows exposure to a sudden noise. Mice can be evaluated for levels of startle magnitude and prepulse inhibition, which occurs when a weak prestimulus leads to a reduced startle in response to a subsequent louder noise. For this study, animals were tested with a San Diego Instruments SR-Lab system. Briefly, mice were placed in a small Plexiglas cylinder within a larger, sound-attenuating chamber. The cylinder was seated upon a piezoelectric transducer, which allowed vibrations to be quantified and displayed on a computer. The chamber included a house light, fan,and a loudspeaker for the acoustic stimuli. Background sound levels (70 dB) and calibration of the acoustic stimuli were confirmed with a digital sound level meter (San Diego Instruments). Each session consisted of 42 trials, that began with a five-minute habituation period. There were 7 different types of trials: the no-stimulus trials, trials with the acoustic startle stimulus (40 msec; 120 dB) alone, and trials in which a prepulse stimulus (20 msec; either 74, 78, 82, 86, or 90 dB) occurred 100 ms before the onset of the startle stimulus. Measures were taken of the startle amplitude for each trial across a 65-msec sampling window, and an overall analysis

was performed for each subject's data for levels of prepulse inhibition at each prepulse sound level (calculated as 100 - [(response amplitude for prepulse stimulus and startle stimulus together / response amplitude for startle stimulus alone) x 100].

1.F.viii Results from acoustic startle test

The GPR68 KO mice had decreased startle responses following presentation of acoustic stimuli, in comparison to the WT mice (**Supplementary Figure 27a** and **27b**). A repeated measures ANOVA, conducted on startle response amplitudes, indicated significant main effects of genotype [$F_{(1,35)}=7.22$, $p=0.011$] and sex [$F_{(1,35)}=16.61$, $p=0.0003$], and a genotype x decibel level interaction [$F_{(6,210)}=5.77$, $p<0.0001$]. Separate comparisons confirmed that both male and female KO mice showed significant reductions in startle responses [genotype x decibel level interaction, males, $F_{(6,84)}=2.57$, $p=0.0245$; and females, $F_{(6,126)}=3.48$, $p=0.0032$]. The decreased startle responses and overt sex differences were not associated with changes in pre pulse inhibition (**Supplementary Figure 27c** and **27d**). The significant main effects of genotype on startle were no longer evident during a second acoustic startle test, conducted when mice were 16-17 weeks in age.

1.F.ix Morris water maze visible platform test

The Morris water maze task was used to assess spatial learning and visual function in the mice. The water maze consisted of a large circular pool (diameter = 122 cm) partially filled with water (45 cm deep, 24-26°C), located in a room with numerous visual cues. Mice were first tested using a visible platform. In this case, each animal was given four trials per day, across two days, to swim to an escape platform cued by a patterned cylinder extending above the surface of the water. For each trial, the mouse was placed in the pool at one of four possible locations (randomly ordered), and then given 60 seconds to find the visible platform. If the mouse found the platform, the trial ended, and the animal was allowed to remain 10 seconds on the platform before the next trial began. If the platform was not found, the mouse was placed on the platform for 10 seconds, and then given the next trial. Measures were taken of latency to find the platform via an automated tracking system (Noldus Ethovision). As shown in **Supplementary Table 10**, all groups of mice demonstrated a high degree of proficiency in the visual cue task.

1.F.x Acquisition and reversal learning in the hidden platform test (**Supplementary Figure 28**)

Three days following the visual cue task, mice were tested for their ability to find a submerged, hidden escape platform (diameter = 12 cm). As in the procedure for visual cue learning, each animal was given four trials per day, with one minute per trial, to swim to the hidden platform. The criterion for learning was an average

latency of 15 seconds or less to locate the platform on one day. Mice were tested until the criterion was reached, with a maximum of nine days of testing. When criterion was reached, mice were given a one-minute probe trial in the pool with the platform removed. In this case, selective quadrant search was evaluated by measuring number of crosses over the location where the platform (the target) had been placed during training, and the corresponding areas in the other three quadrants. Following the acquisition phase, mice were tested for reversal learning, using the same procedure as described above. In this phase, the hidden platform was located in a different quadrant in the pool, diagonal to its previous location. As before, measures were taken of latency to find the platform. On the day that the criterion for learning was met, the platform was removed from the pool, and the group was given a probe trial to evaluate reversal learning.

1.G References for supplementary materials

1. Gottlieb, H. E., Kotlyar, V. & Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. J. Org. Chem. **62**, 7512–7515 (1997).

2. Jin, H. et al. Substituted 3-(4-(1,3,5-triazin-2-yl)-phenyl)-2-aminopropanoic acids as novel tryptophan hydroxylase inhibitors. Bioorg. Med. Chem. Lett. **19**, 5229–5232 (2009).

II. Supplementary Figures

Supplementary Figure 1. Yeast based high throughput screening to identify potential ligands for GPR68. Heat map showing growth responses of yeast expressing 24 different GPCRs (horizontal axis) to 446 compounds in the NCC library (vertical axis). Blue color is 0% (or less than 0%) above background; white is 100% above background, red color indicates 200% (or more) above background growth stimulation, *i.e.*, agonist activity. Data were clustered using GENE-E software (http://www.broadinstitute.org/cancer/software/GENE-E/) using complete linkage algorithm. GPR68 is the right-most column in the heat map and details are shown in **Figure 1a**.

Supplementary Figure 2. Validation and confirmation of GPCR activation assays done in yeast. (a-d) Concentration-dependent growth of GPR43-expressing G_s yeast (**a**), GPR43-expressing G_q yeast (**b**), GPR41-expressing G_s yeast (**c**), and GPR41-expressing G_q yeast (**d**) in response to various short-chain fatty acids (SCFAs). **(e)** Concentration-dependent growth of GPR39-expressing G_s yeast (GPR39s) in response to zinc ions; black symbols are responses to $ZnCl_2$, red symbols are responses to $ZnSO_4$, closed symbols are receptor-expressing yeast and open symbols are yeast transformed with empty p426GPD plasmid (p426s). **(f)** Concentration-dependent growth of GPR39-

expressing G_s yeast (GPR39s) in response to chromium ions; black symbols are responses to chromium (III) potassium sulfate dodecahydrate, red symbols are responses to chromium (II) chloride, closed symbols are receptor-expressing G_s yeast and open symbols are G_s yeast transformed with empty p426GPD plasmid (p426s). **(g)** Concentration-dependent growth of GPR39-expressing G_s yeast (GPR39s) in response to cadmium ions; black symbols are responses to $CdSO_4$, red symbols are responses to $CdCl_2$, closed symbols are receptor-expressing G_s yeast and open symbols are G_s yeast transformed with empty p426GPD plasmid (p426s). **(h)** Concentration-dependent growth of GPR39-expressing G_s yeast (GPR39s) in response to iron ions; black symbols are responses to iron(II) chloride heptahydrate, red symbols are responses to iron (II) sulfate heptahydrate, closed symbols are receptor-expressing G_s yeast and open symbols are G_s yeast transformed with empty p426GPD plasmid (p426s). **(i-k)** Concentration-dependent cAMP responses of GPR39-expressing HEK293-T cells to $ZnCl_2$ **(i)**, to $ZnSO_4$ **(j)**, to $CdSO_4$ **(k)** as measured by luciferase cAMP reporter assay. **(l)** Concentration-dependent growth of GPR39-expressing yeast to omeprazole or esomeprazole; black symbols are responses in G_s yeast, red symbols are responses in G_q yeast, closed symbols are responses to omeprazole, open symbols are responses to esomeprazole. All data are expressed as mean ± SEM of n = 3 to 6 replicates; curves were fitted using Graphpad Prism. For responses to metal ions, curves were truncated at high ion concentrations due to toxicity.

164
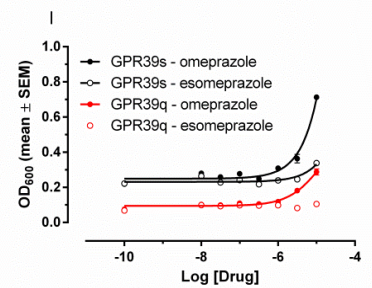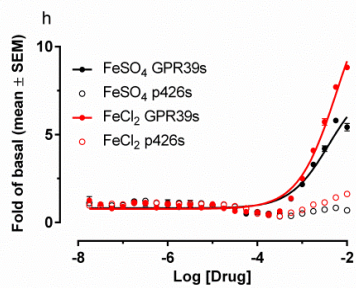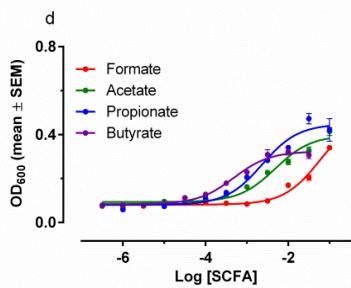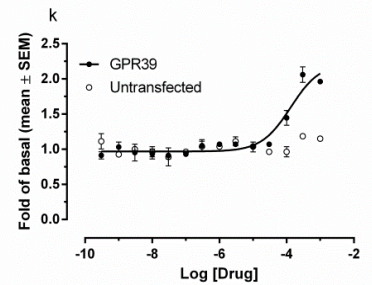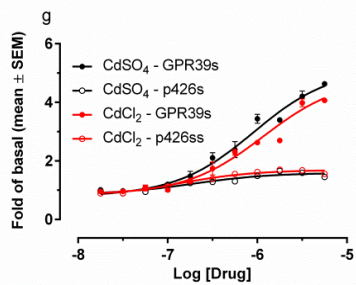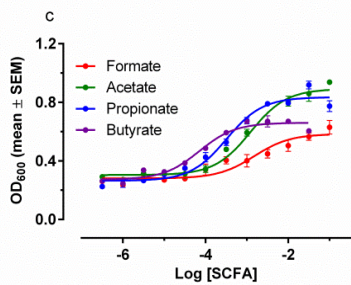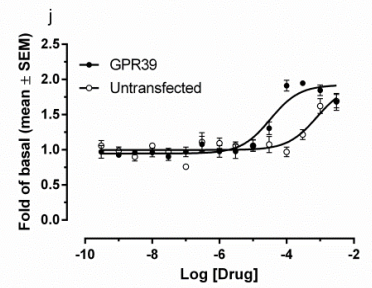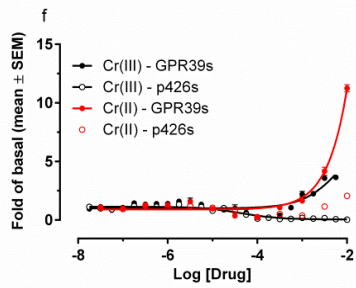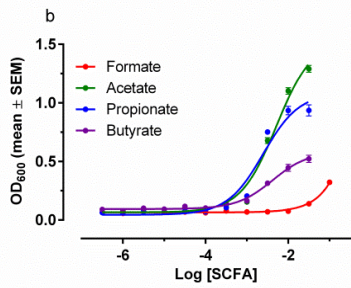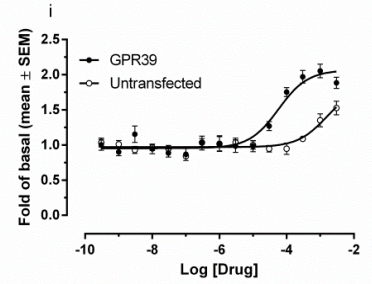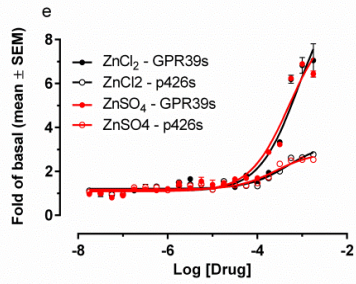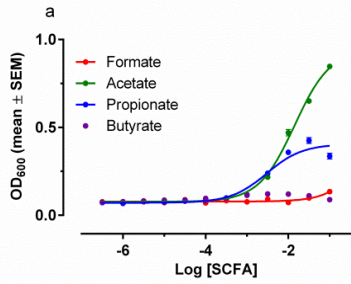
Supplementary Figure 3. Concentration- and pH-dependent activity of benzodiazepines. Representative concentration-response curves of *N*-unsubstituted benzodiazepines lorazepam (**a**), clonazepam (**b**), desmethyldiazepam (**c**), and norfludiazepam (**d**) at pH 6.50 or 7.40 in GPR68-transfected HEK293-T cells (structures in **Supplementary Table 1**). Production of cAMP was measured using a split-luciferase reporter assay in transiently transfected HEK293-T cells. Normalized results (mean ± SEM, fold of basal relative to the corresponding RLU at pH 7.40) represent one of three independent assays, each in triplicate. Lorazepam potency (EC_{50}) is 1.8 µM and 9.1 µM at pH 6.50 and pH 7.40, respectively. Clonazepam potency is 3.2 µM and >10 µM at pH 6.50 and pH 7.40, respectively. Desmethyldiazepam potency is 4.7 µM and 1.7 µM at pH 6.50 and pH 7.40, respectively. Norfludiazepam potency values are 3.2 and 5.5 µM at pH 6.50 and pH 7.40, respectively.
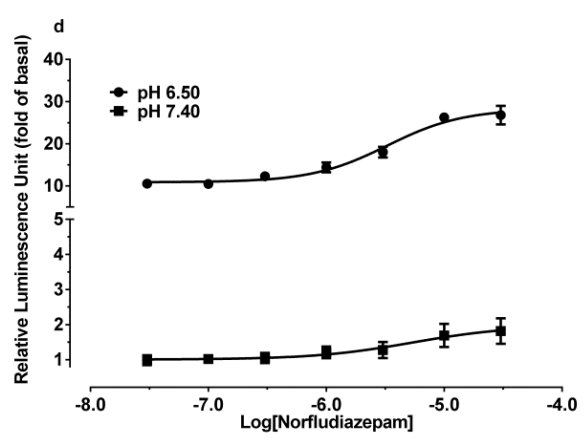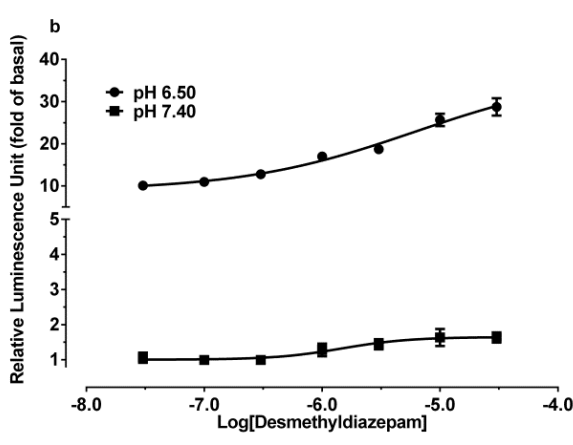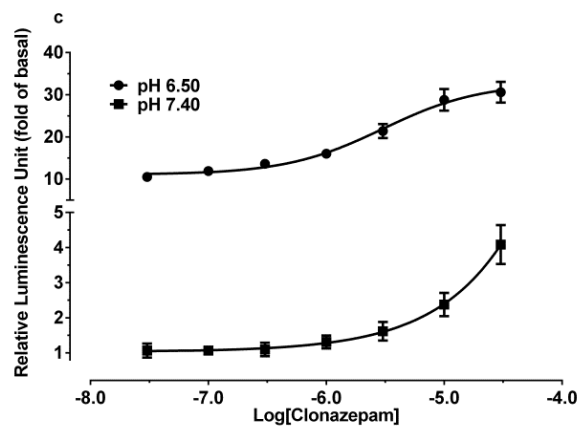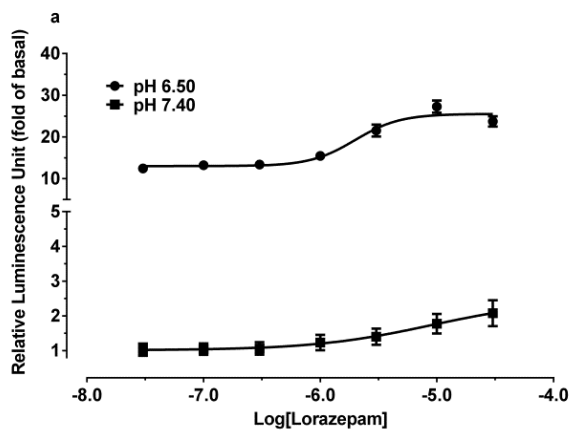
Supplementary Figure 4. Lorazepam has minimal effect on proton-mediated cAMP production at GPR4 and GPR65 receptors. HEK293-T cells were transiently transfected and production of cAMP was measured using a split-luciferase reporter assay. Proton concentration-response curves at GPR4 (**a**) or GPR65 (**b**) were carried out in the absence and presence of varying concentrations of lorazepam. Results (RLU in fold of basal) represent mean ± SEM from a minimum of three independent assays, each in triplicate, and were fitted using Graphpad Prism using the built-in 4-parameter logistic function. Proton potency (pEC$_{50}$) is 7.88 ± 0.08 and the Hill slope is 3.13 ± 1.63 for GPR4; 7.29 ± 0.02 and 3.31 ± 0.36 for GPR65, in the absence of lorazepam.
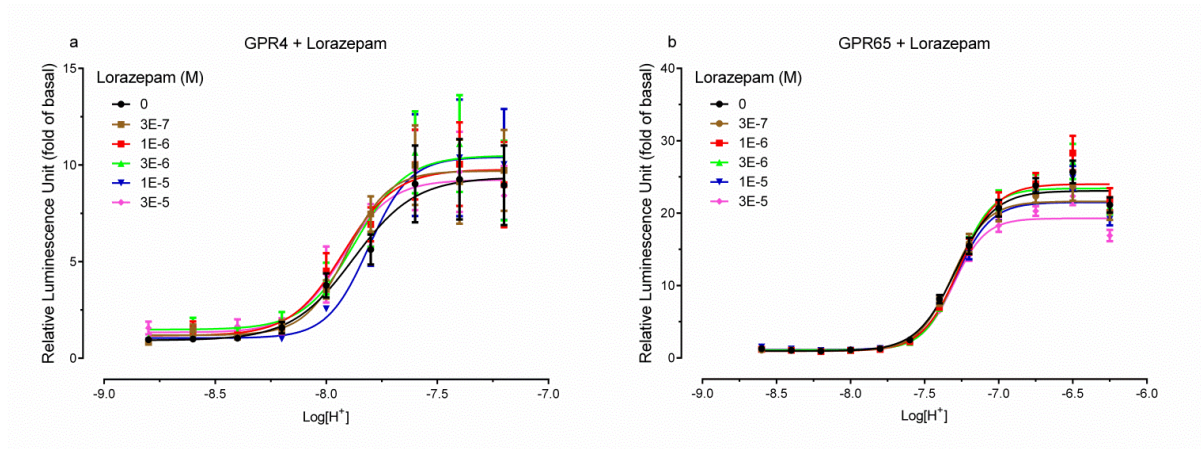
Supplementary Figure 5. Heat map of off-target activities of lead compounds at potential CNS drug targets. Radioligand binding assays were carried out in the National Institute of Mental Health Psychoactive Drug Screening Program (NIMH PDSP) as described previously[1,2] (on-line protocols available at http://pdsp.med.unc.edu/pdspw/binding.php). Values represent mean binding affinities (pK_i) from a minimum of 2 independent assays, each done in triplicate. Affinities lower than a pK_i of 5, or less than 50% inhibition at 10 M (quadruplicate), are shown as a minimum of 5 on the pK_i scale. The hERG inhibition activity was tested in a hERG functional assay as outlined in the PDSP protocol book, and as previously published[3]. ND for not determined; BZP for benzodiazepine receptor; DAT for dopamine transporter; NAT for norepinephrine transporter; SERT for serotonin transporter; DOR for delta ( ) opioid receptor; MOR for mu ( ) opioid receptor; KOR for kappa ( ) opioid receptor; PBR for peripheral benzodiazepine binding site; AMPA for aminomethylphosphonic acid receptor; KA for kainate acid receptor; NMDA for N-methyl-D-aspartate receptor; hERG for human Ether-a-go-go-related Gene (potassium channel Kv11.1).
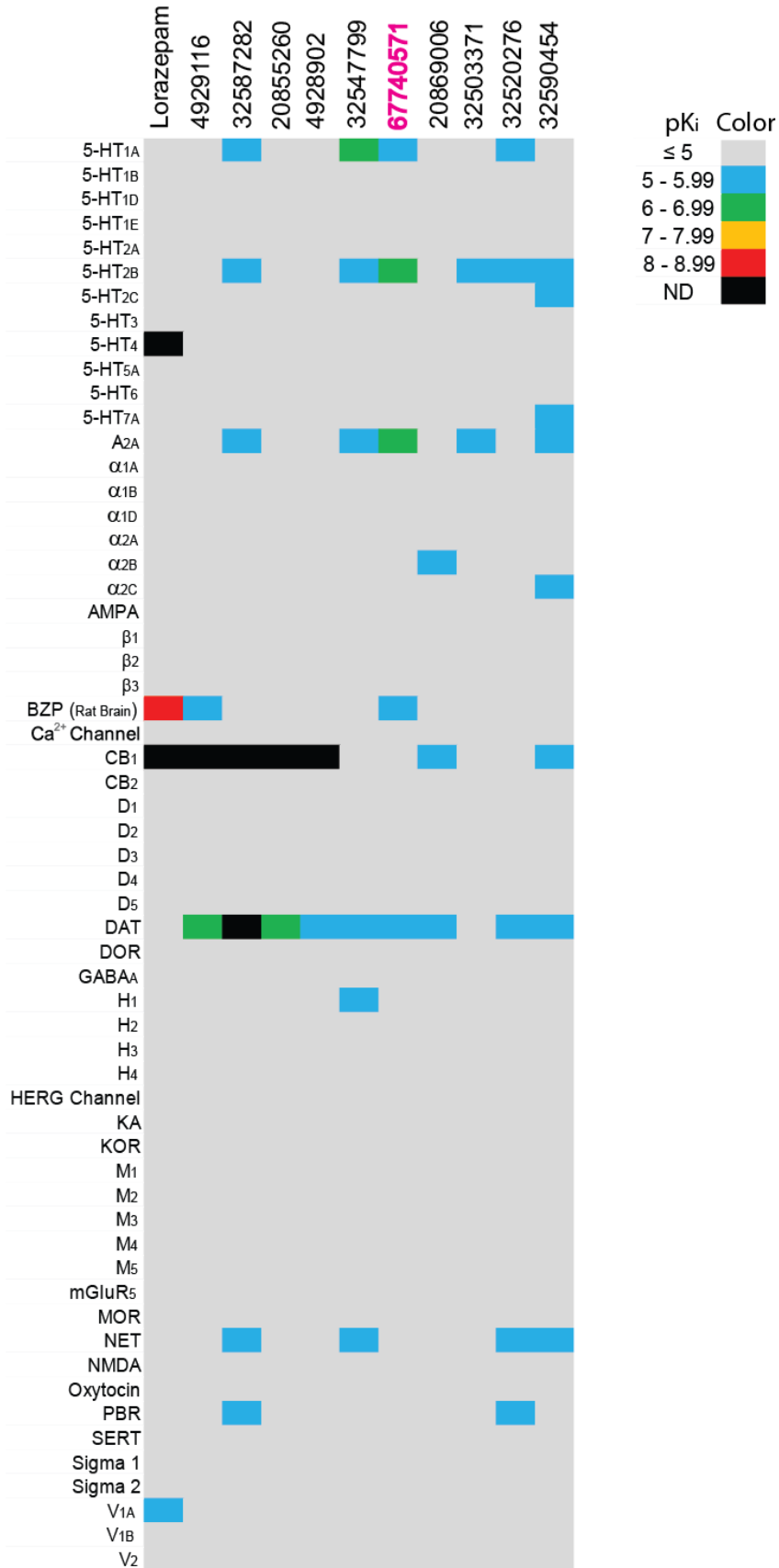
Supplementary Figure 6. Sequence alignment of human proton-sensing receptors to CXCR4. Sequences of human proton-sensing receptors were obtained from the Uniprot database, and they were aligned to CXCR4 (PDB code: 3ODU) automatically using PROMALS-3D. The alignment was manually refined to reduce gaps and to position conserved residues. Black boxes indicate the seven putative transmembrane (TM) regions, along with intracellular (IL) and extracellular (EL) loops, of the proton-sensing receptors. Residues are highlighted in blue by degree of conservation at the position throughout the aligned sequences; darker blue indicates the residue is more conserved. Red boxes indicate residues previously shown by mutagenesis to be crucial to the proton-sensing mechanism of the respective receptor. Red stars (*) indicate residues that were mutated in the current study and shown to be important for proton-mediated GPR68 and GPR65 activation. Numbers indicate the position of the first and last residues in each row.

```
                      TM1                              IL1           TM2
GPR68_HUMAN    12  SCTIDHTIH--QTLAPVVYVTVLVVGFPANCLS-LYFQYLQIKARTELGVYLCNLTVADL  68
GPR4_HUMAN      8  GCHVDSRVD--HLFPPSLYIFVIGVGLPTNCLA-LWAAYRQVQQRTELGVYLMNLSIADL  64
GPR65_HUMAN     4  TC-IEEQHDLDHYLFPIVYIFVIIVSIPANIGS-LCVSFLQAKKESELGIYLFSLSLSDL  61
GPR132_HUMAN   32  TCNNVSFEE-SRIVLVVVYSAVCTLGVPANCLT-AWLALLQVLQGIVLAVYLLCLALCEL  89
CXCR4_HUMAN    27  PCFREENANFNKIFLPTIYSIIFLTGIVGNGLVILVMGY-QKKLRSMTDKYRLHLSVADL  85

                 TM2          EL1              TM3                    IL2
GPR68_HUMAN    69  FYICSLPFWLQYVLQHDNWSHGDLSCQVCGILLYENIYISVGFLCCISVDRYLAVAHPFR 128
GPR4_HUMAN     65  LYICTLPLWVDYFIHHDNWIHGPGSCKLFGFIFYTNIYISIAFLCCISVDRYLAVAHPLR 124
GPR65_HUMAN    62  LYALTLPLWIDYTWNKDNWTFSPALCKGSAFLMYMNFYSSTAFLTCIAVDRYLAVYPLK 121
GPR132_HUMAN   90  LYTGTLPLWVIYIRNQHRWTLGLLACVTAYIFFCNIYVSILFLCCISCDRFVAVYALE  149
CXCR4_HUMAN    86  LFVITLPFWAVDAVA--NWYFGNFLCKAVHVIYTVNLYSSVWILAFISLDRYLAITHATN 143

                 TM4                   *        EL2              TM5
GPR68_HUMAN   129  FHQFRTLKAAVGVSVVIWAKELLTSIYFLMHEEVIE-----DENQHRVCFEHYPSQAWQR 183
GPR4_HUMAN    125  FARLRRVKTAVAVSSVVWATELGANSAPLFHDELFR-----DRYNHTFCFEKFPMEGWVA 179
GPR65_HUMAN   122  FFFLRTRRFALMVSLSIWILETIFNAVMLWEDETVVEYCDAEKSNFTLCYDKYPLEKWQI 181
GPR132_HUMAN  150  SRGRRRRRTAILISACIFILVGIVHYPVFQTED------------KETCFDMLQMDSRIA 197
CXCR4_HUMAN   144  SQRPRKLLAEKVVYVGVWIPALLLTIPDFIFANV-S-----EADDRYICDRFYPNDLWVV 197

                 *       TM5               IL3            TM6
GPR68_HUMAN   184  AINYYRFLVGFLFPICLLLASYQGILRAVRRSHQTQKSRKDQIQRLVLSTVVIFLACFLH 243
GPR4_HUMAN    180  WMNLYRVFVGFLFPWALMLLSYRGILRAVRGSVSTERQEKAKIKRLALSLIAIVLVCFAH 239
GPR65_HUMAN   182  NLNLFRTCTGYAIPLVTILICNRKVYQAVRINKATENKEKKRIIKLLVSITVTFVLCFTP 241
GPR132_HUMAN  198  GYYYARFTVGFAIPLSIIAFTNHRIFRSIKQSMQLSAAQKAKVKHSAIAVVVIFLVCFAP 257
CXCR4_HUMAN   198  VFQFQHIMVGLILPGIVILSCYCIIISKLSHSKGHQK-RK-ALKTTVI-LILAFFACWLP 254

                 TM6                            *      TM7
GPR68_HUMAN   244  YHVLLLVPS--VWEA--S---------CDFAKGVFNAYHFSLLLTSFNCVADPVLYC  287
GPR4_HUMAN    240  YHVLLLSPS--AIYL--GR----PWDCGFEERVFSAYHSSLAFTSLNCVADPILYC  287
GPR65_HUMAN   242  FHVMLLIPC--ILEH--AVN--FEDHSNSGKRTYTMYRITVALTSLNCVADPILYC  291
GPR132_HUMAN  258  YHLVLLVPA--AAFS--YYRGDRNAMCGLEERLYTASVVFLCLSTVNGVADPIIYV  309
CXCR4_HUMAN   255  YYIGISIDSFILLEIIKQG-------CEFENTVHKWISITEALAFFHCCLNPILYA  303
```

Supplementary Figure 7. Sampling regions for lorazepam binding modes in models of GPR68. Each colored surface represents a different region sampled for possible lorazepam binding. The yellow surface contours the binding site of the small molecule antagonist 1T1t in the template CXCR4 crystal structure (PDB code: 3ODU), whereas the grey surface shows how deeply the cyclic peptide CVX15 binds to CXCR4 (PDB code: 3OE0). The green and red surfaces sample across the entire binding cleft. The magenta surface represents the canonical orthosteric pocket for biogenic amine receptors.
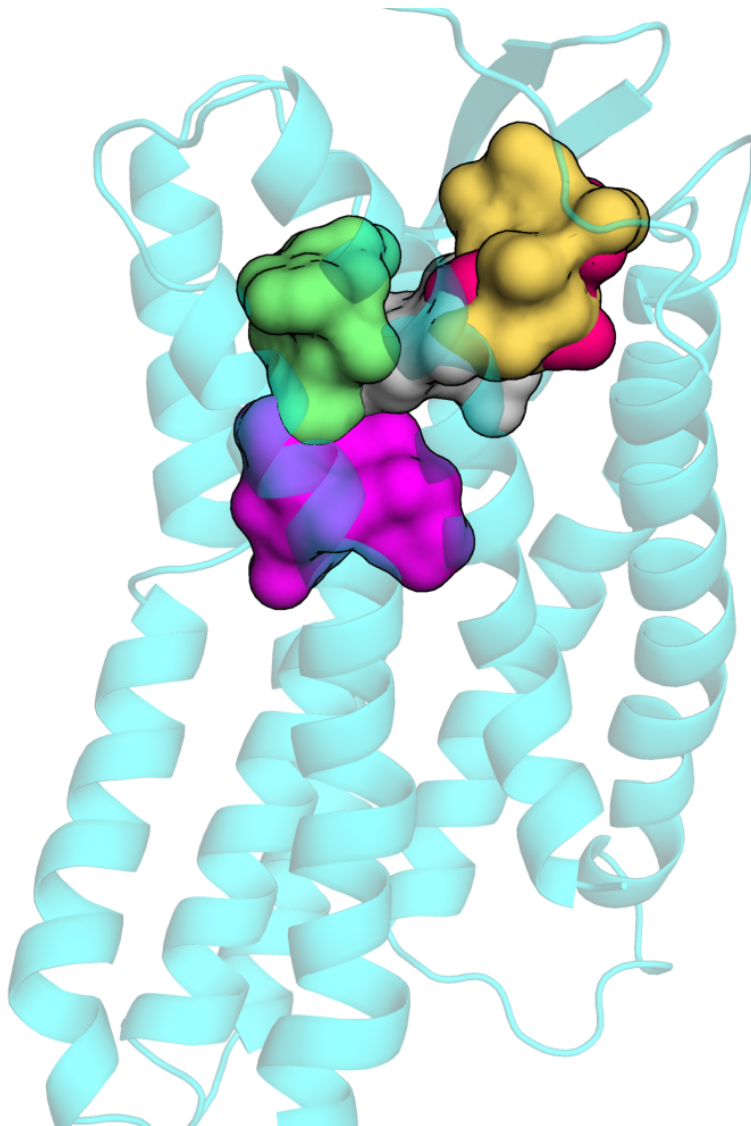
Supplementary Figure 8. Comparisons of proton agonist activity at GPR68 wild-type and mutant receptors. The functional assays were carried out in transiently transfected HEK293-T cells with a split-luciferase reporter assay for cAMP production (G_s-pathway) and calcium mobilization assay on a FLIPR^{TETRA} for the G_q-pathway, analyzed with GraphPad Prism and the normalized curves are presented in **Figure 3b and 3c**. Parameters presented here include proton efficacy (E_{max}) expressed as fold of basal for cAMP production **(a)** and percentage of wild-type for calcium release **(b),** proton potency (pEC_{50}) values **(c)**, and Hill slopes **(d).** For comparison, proton potency values and Hill slopes at GPR4 and GPR65 are also included in panels **c** and **d**. Values represent mean ± SEM from 12 assays for the cAMP production assays or 8 assays for calcium release assays, with each assay done in triplicate or quadruplicate. An example of basal cAMP production is shown in **Supplementary Figure 6**. Relative GPR68 wild-type and mutant receptor expression levels were determined by anti-Flag immunoblotting (**e**). HEK293-T cells were transiently transfected with Flag-tagged (at C-tail) GPR68 wild-type and mutant receptors and cell lysates were subjected to immunoblotting. Blot bands were quantified and normalized to GPR68 wild-type. Results represent mean ± SEM from three independent transient transfections and immunoblots. A representative immunoblot was inserted in panel **e**. N.C. for negative control with untransfected HEK293-T cells. Immunoblottings with commercial anti-GPR68 antibody failed due to high nonspecific activity with control cells (results not shown). Proton-mediated cAMP production in HEK293-T cells transfected with luciferase reporter alone (**f**). Results were normalized to basal at pH 8.40 and are presented as mean ± SEM

173

(fold of basal) from 4 independent assays, each in triplicate. Overall, different pH conditions had little effect on cAMP production. Low pH conditions slightly reduced luciferase signals, in contrast to the large increase in cAMP production when proton receptors were co-expressed and activated under acidic conditions.

a cAMP Production Efficacy

b Calcium Release Efficacy

c pEC$_{50}$

d Hill Slope

e

~35KD

f
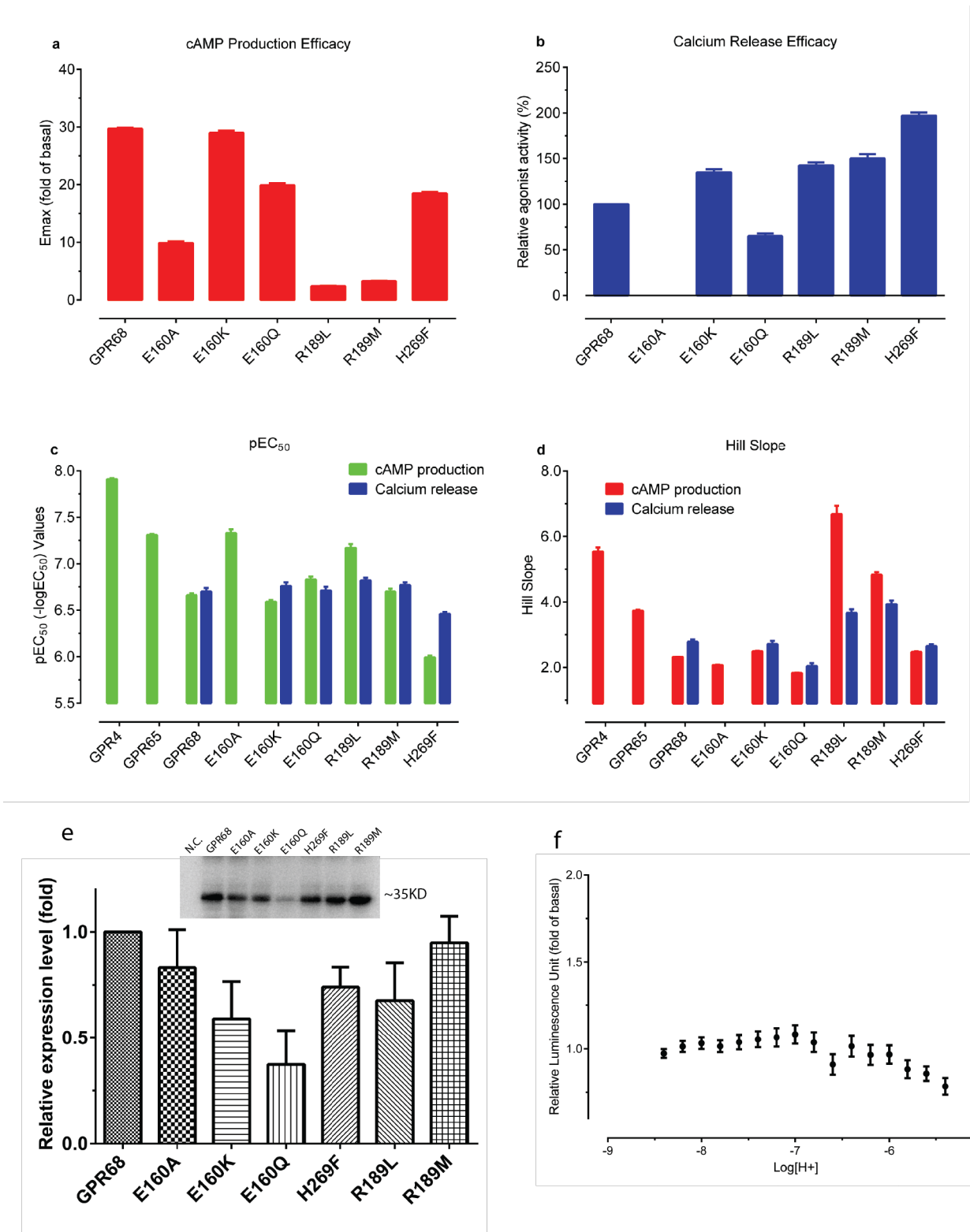
Supplementary Figure 9. Basal activity of GPR68 wild-type and mutant receptors. Production of cAMP was measured in transiently transfected HEK293-T cells using a split-luciferase reporter assay. Results (relative luminescence units in fold of basal at pH 9.0) represent mean ± SEM of a minimum of 6 independent assays, each in triplicate or quadruplicate.
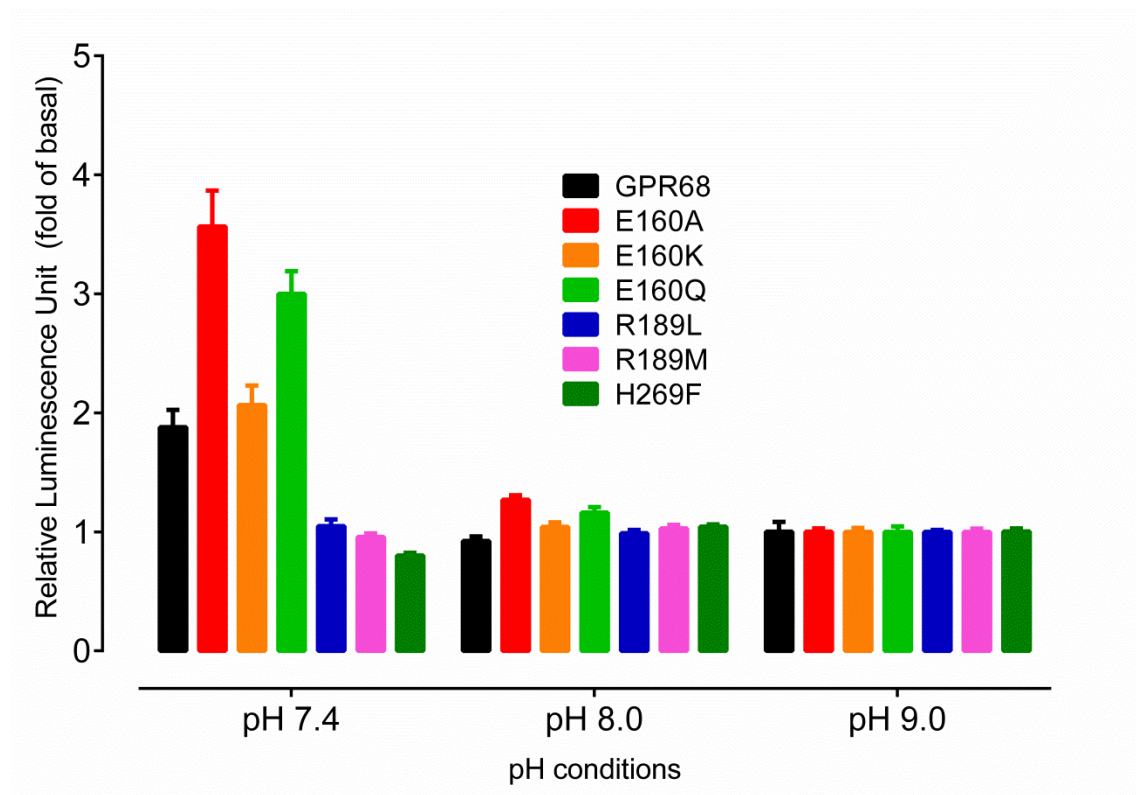
Supplementary Figure 10. Proton concentration-response curves illustrating positive allosteric modulation at wild-type GPR68 receptors. HEK293-T cells were transiently transfected and production of cAMP in the absence and presence of indicated modulators was measured using a split-luciferase reporter assay. Results (RLU in fold of basal) were normalized to corresponding basal activity (counts at the highest pH), and are shown as mean ± SEM from a minimum of 4 independent assays, each in triplicate or quadruplicate. Curves were analyzed in GraphPad Prism with the built-in 4-parameter logistic function (right column) and the standard allosteric operational model (left column). Panel a and a' are for ZINC 4929116; b and b' for ZINC 32587282; c and c' for ZINC 20855260; d and d' for ZINC 4928902; e and e' for ZINC 32547799; f and f' for 20869006; g and g' for ZINC 32503371; h and h' for 32520276, i and i' for 32590454. Allosteric parameters are summarized in **Supplementary Table 4** and plotted in **Supplementary Figure 11a**. For detailed allosteric operational model and curve fitting, see **Supplementary Table 4**. For each pair of fittings, the proton potency value (pEC_{50}) from agonist concentration-response curve (right panel) in the absence of ZINC compound was used as pK_A for the allosteric operational model (left panel); the highest E_{max} value from agonist concentration-response curves (right panel) in the presence of the ZINC compound was used as the E_{max} for the allosteric operational model (left panel).
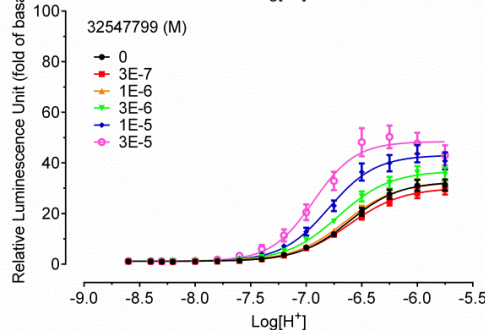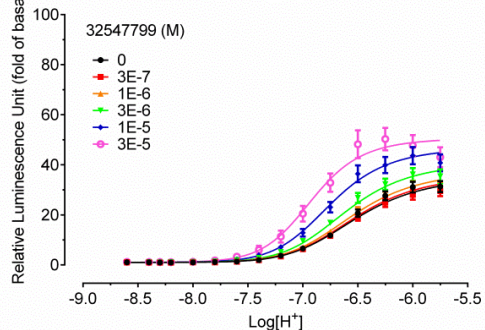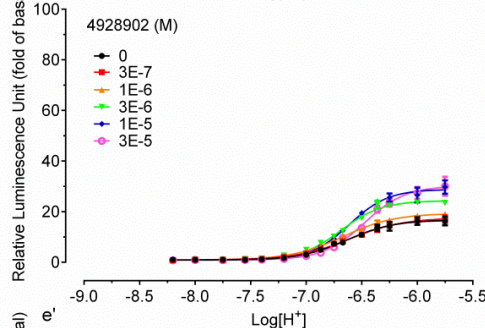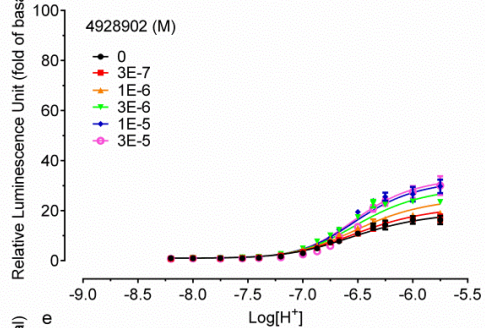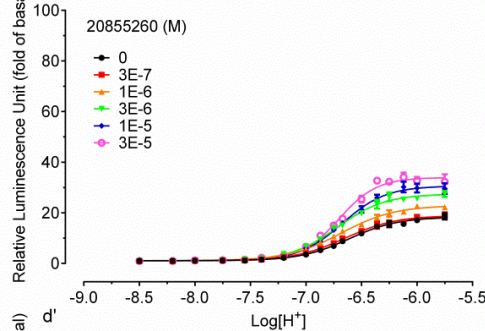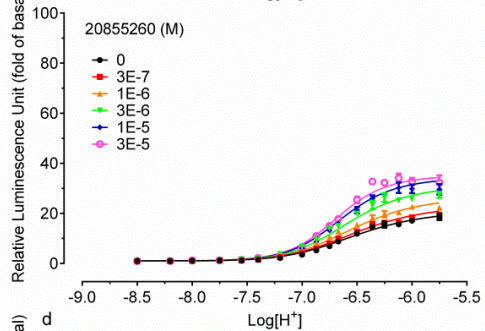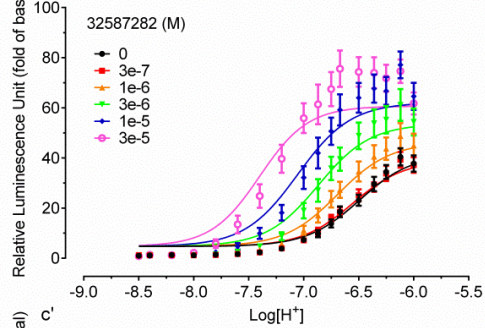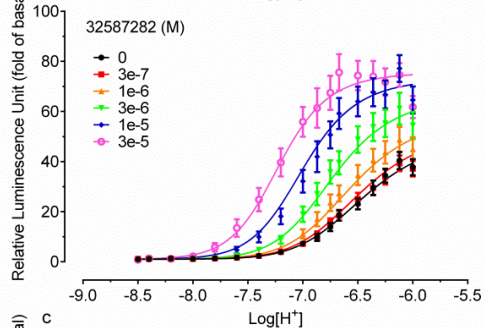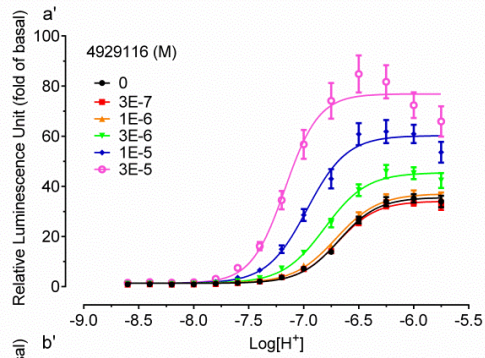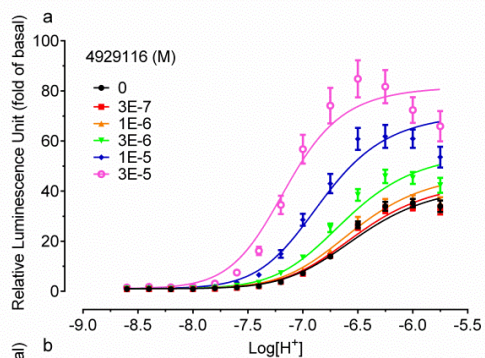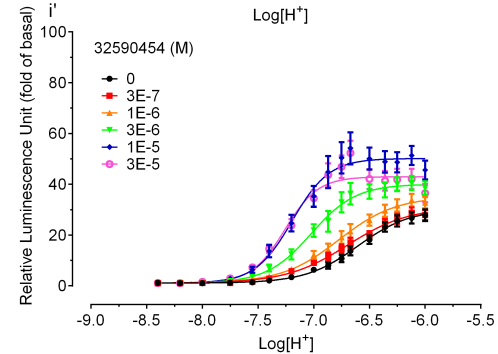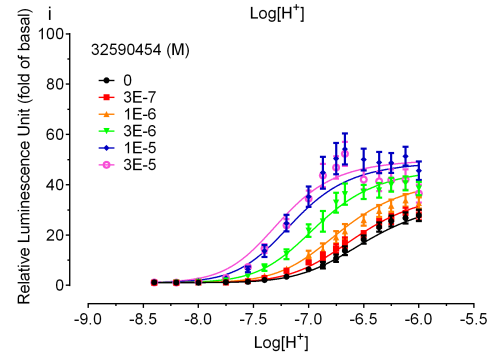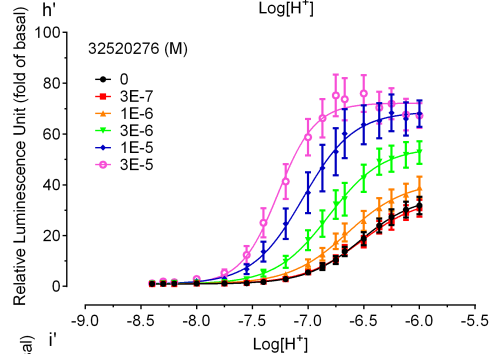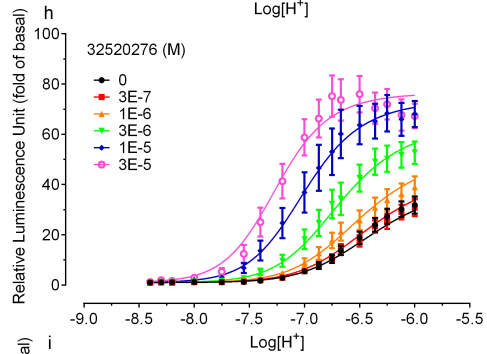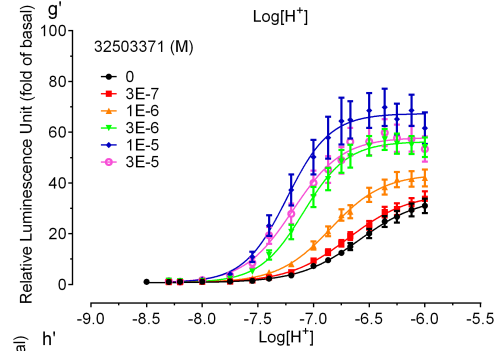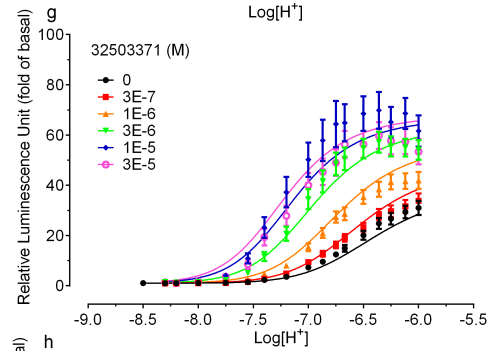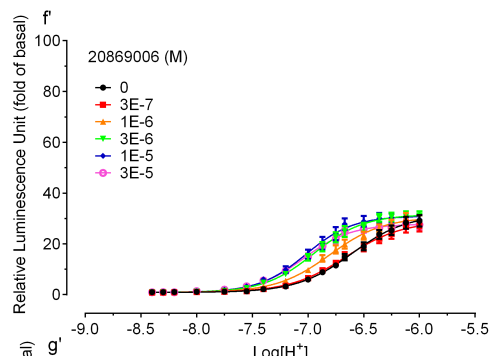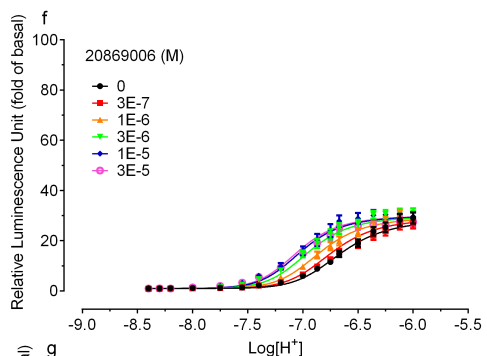
Supplementary Figure 11. Allosteric parameters from Supplementary Table 4 were plotted for comparison. (a) Allosteric parameters of lorazepam and 10 ZINC lead compounds at GPR68 wild-type receptors. ZINC 67740571 is dubbed as ogerin in this study. Representative concentration-response curves are shown in **Supplementary Figure 7**. **(b)** Allosteric parameters of ogerin at GPR68 mutant receptors. Representative concentration-responses curves are shown in **Supplementary Figure 11**.

b
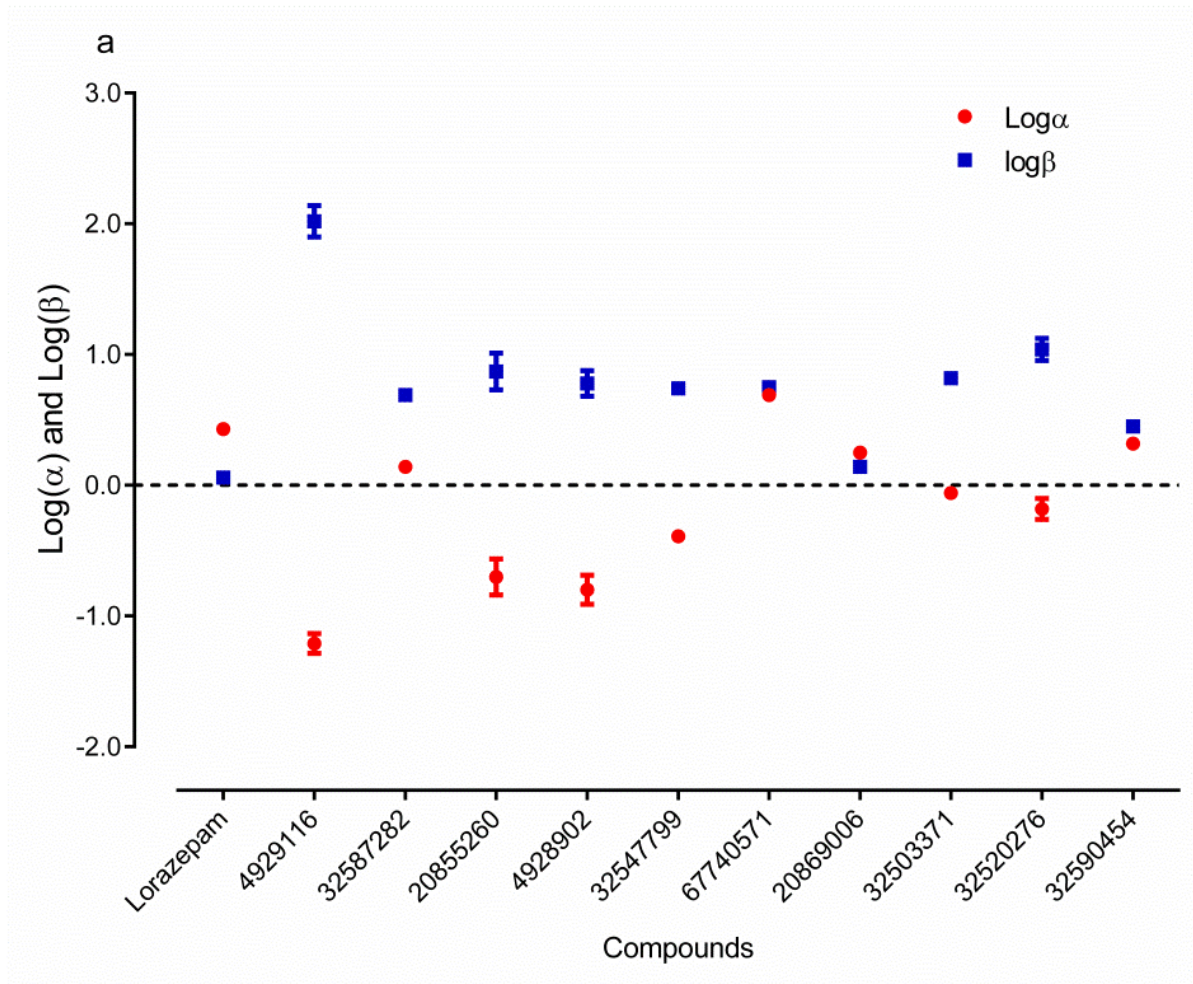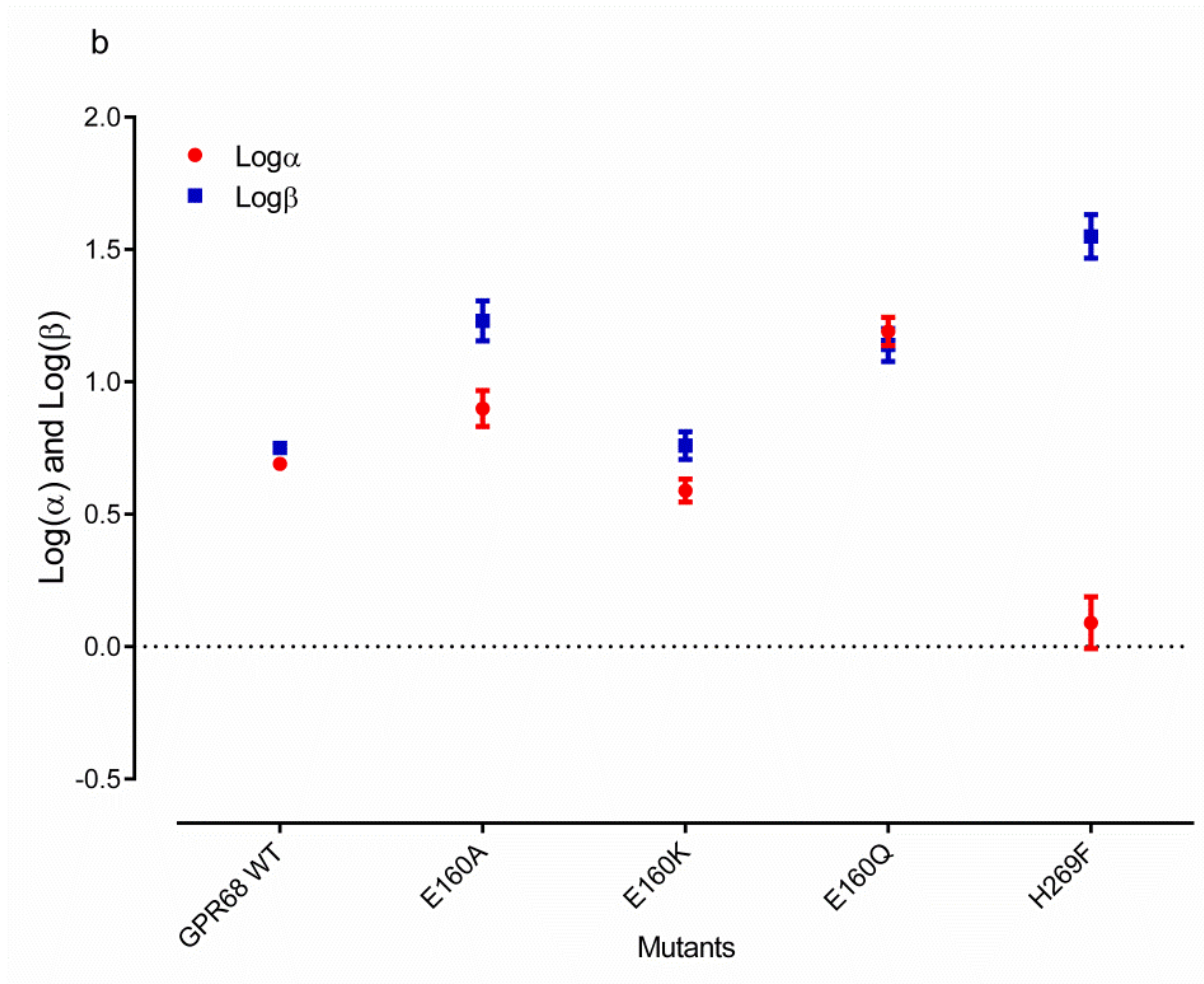
Supplementary Figure 12. Apparent pH-dependent activity of ogerin at GPR68 wild-type (a) and mutant receptors (b). HEK293-T cells were transiently transfected and production of cAMP was measured using a split-luciferase reporter assay. Representative ogerin concentration-response curves (mean ± SEM) under different pH conditions from one of 3 assays, each done in quadruplicate, are shown. Pharmacological parameters are summarized in **Supplementary Table 5**. **(b)** Concentration-responses of ogerin at GPR68 wild-type, E160A, and E160Q mutant receptors**.** A set of ogerin concentration-response curves at pH 9.0 from one of three independent assays (each done in triplicate or quadruplicate) is shown. Results represent mean ± SEM, and were analyzed using GraphPad Prism. The apparent 'agonist' activity of ogerin at high pH conditions at wild-type and mutant receptors E160A and E160Q is best explained by high allosteric effect of ogerin as indicated by large $\log(\alpha{*}\beta)$ values (**Supplementary Table 4**). To determine if high pH conditions could affect endogenous G_s activity, more control experiments were done and are shown in **Supplementary Figure 10**.

a Ogerin at GPR68 under different pH

b Ogerin at GPR68 mutants under pH 9.0
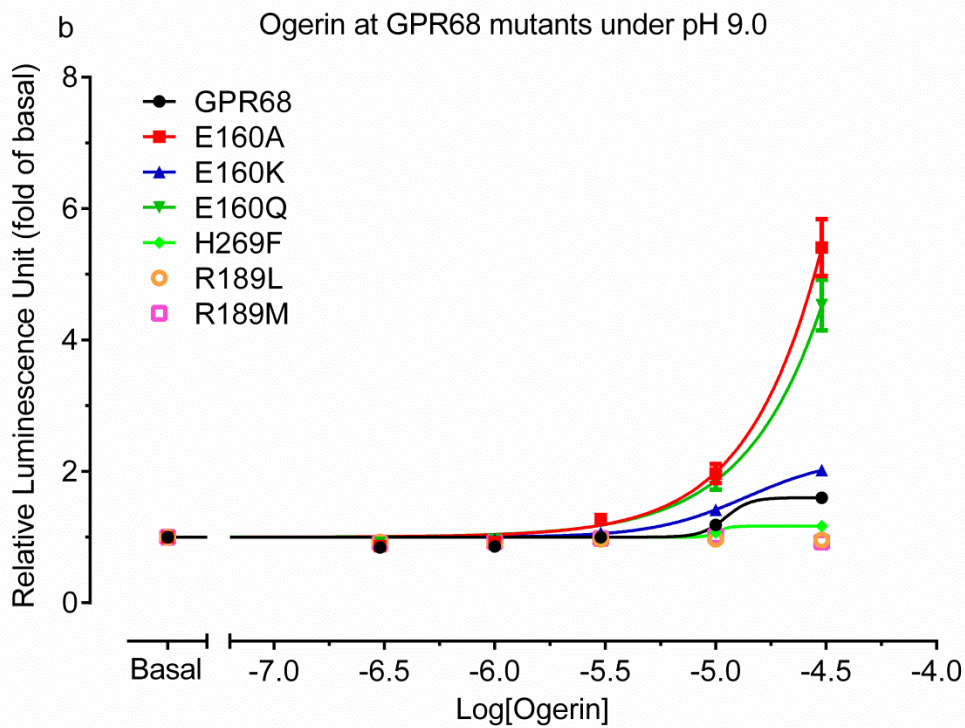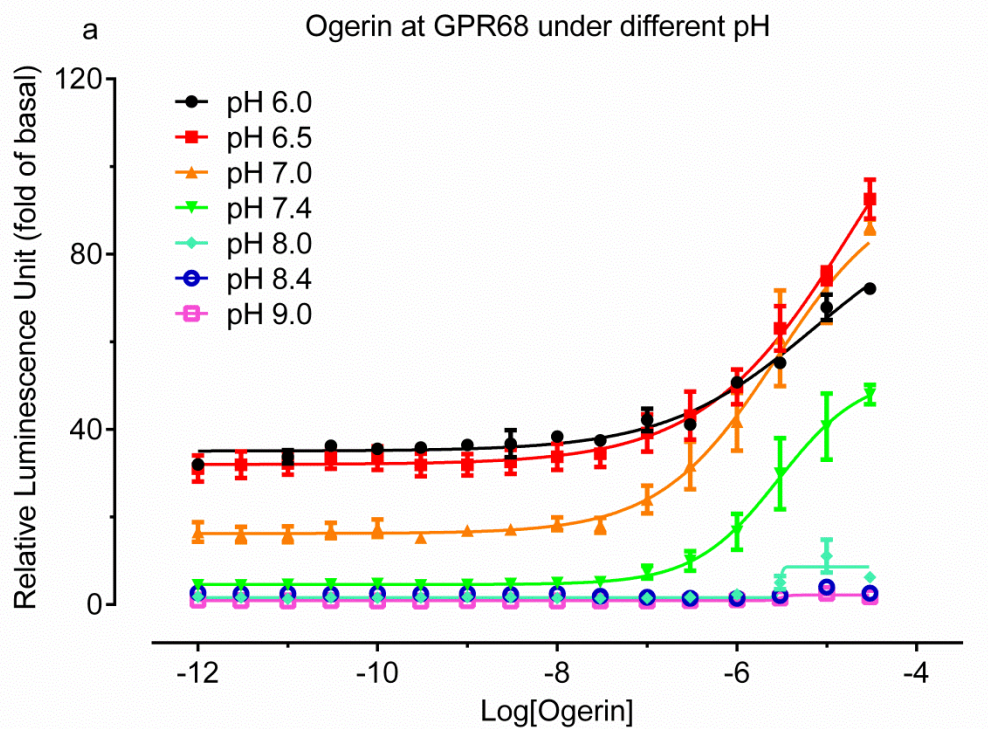
Supplementary Figure 13. Isoproterenol (ISO) mediated endogenous β_2-G_s activation under multiple pH conditions. HEK293-T cells were transiently transfected with GloSensor plasmid and production of cAMP was measured using a split-luciferase reporter assay. Isoproterenol (ISO)-activated endogenous $_2$ adrenergic receptor-G_s activation in control HEK293-T (**a** and **c**) and GPR68-transfected cells (**b** and **d**) at pH 6.50 (red curves), pH 7.40 (green curves), and pH 9.50 (blue curves). Results (RLU in fold of basal at pH 9.5) represent mean ± SEM from a minimum of 6 independent assays, each in triplicate or quadruplicate, and were analyzed in GraphPad Prism using the built-in 4-parameter logistic function. Results in Panels **a** and **b** were normalized to corresponding basal levels (buffer control) to obtain activity in fold of basal, re-analyzed and are presented in Panels **c** and **d**, respectively. The potency of ISO (pEC_{50}) is 8.19 ± 0.14 (6.4 nM), 8.24 ± 0.15 (5.8 nM), and 7.79 ± 0.08 (16.4 nM) at pH 6.50, 7.40, and 9.50, respectively, in control HEK293-T cells. The potency of ISO (pEC_{50}) is 7.24 ± 0.22 (57.7 nM), 7.61 ± 0.15 (24.6 nM), and 7.85 ± 0.06 (14.1 nM) at pH 6.50, 7.40, and 9.50, respectively, in GPR68 wild-type cells. GPR68-transfected cells had increased basal activity at pH 6.50, leading to a much reduced maximal fold-activation by ISO.

185
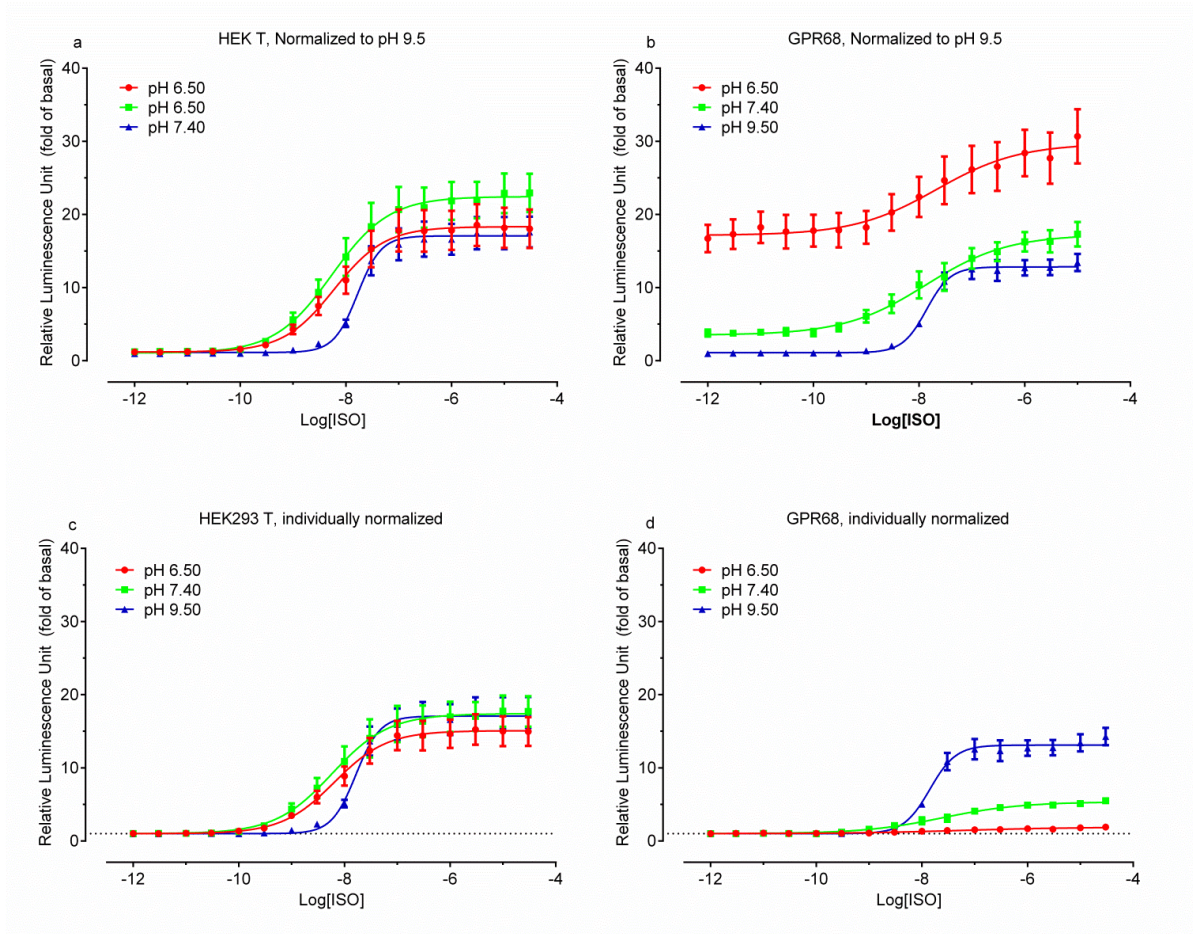
Supplementary Figure 14. Concentration-response curves of ogerin at GPR68 mutant receptors. HEK293-T cells were transiently transfected and production of cAMP was measured using a split-luciferase reporter assay. Proton concentration-response curves were examined in the absence and presence of varying concentrations of ogerin. Results (RLU in fold of basal) represent mean ± SEM from a minimum of 4 independent assays, each in triplicate or quadruplicate, and were analyzed with the standard allosteric operational model in GraphPad Prism. The results are plotted on the same scale to show relative proton activity in the absence and presence of ogerin at different mutant receptors. Results with mutant R189L and R189M were not able to be analyzed with the standard allosteric operational model due to low activity in the absence of ogerin, and were instead analyzed the built-in 4-parameter logistic function in GraphPad Prism. Allosteric parameters are summarized in **Supplementary Table 4** and **Supplementary Figure 11b**.
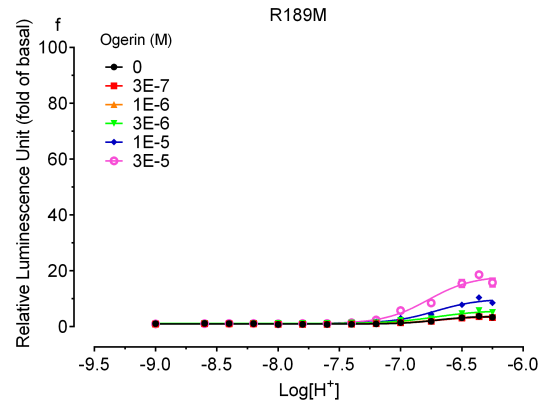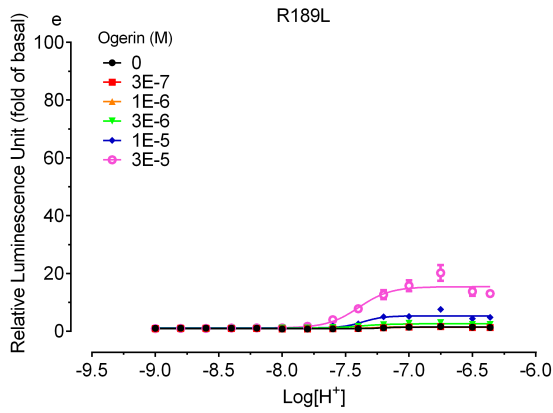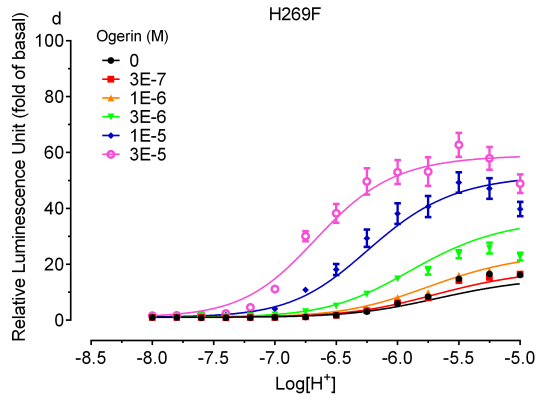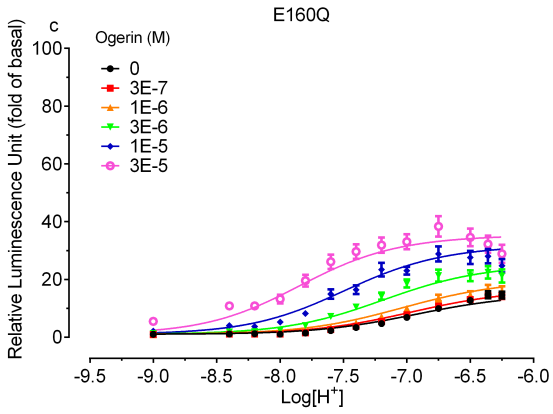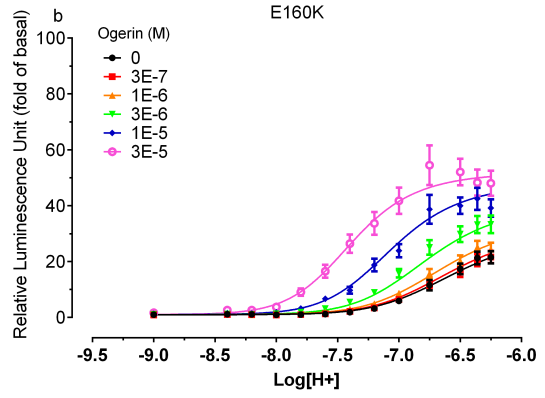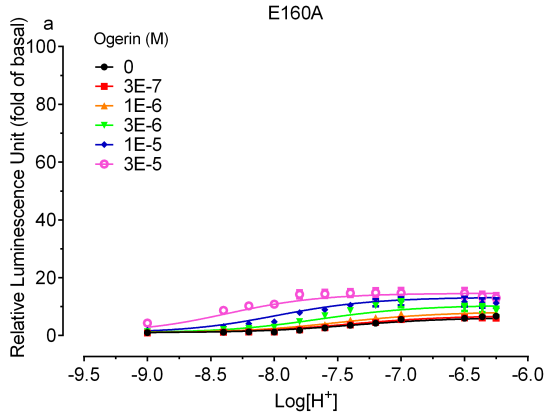
Supplementary Figure 15. Allosteric modulation of proton-mediated calcium release with GPR68 wild-type and mutant receptors. HEK293-T cells were transiently transfected with GPR68 wild-type or mutant receptors, and calcium mobilization was measured using FLIPR according to the procedure in the **Supplementary Methods.** (**a**) Weak nonspecific activity of the lead compounds (10 M at pH 8.00) at GPR68 wild-type and mutant receptors. Normalized (basal = 1, pH 8.0) responses are presented as mean ± SEM from a minimum of 3 independent assays, each in triplicate or quadruplicate. (**b – h**) Effects of selected lead compounds (10 M, ZINC 32587282, 32547799, and 67740571) on proton-mediated calcium release (G_q-pathway activity) at GPR68 wild-type (**b**), E160A (**c**), E160K (**d**), E160Q (**e**), H269F (**f**), R189L (**g**), and R189M (**h**). Compounds (or buffer control) were added and incubated for 10 min before responses to pH solutions were measured. Results (RFU in fold of basal, basal was defined as an average reading before addition of pH solutions) represent mean ± SEM from 3 independent assays, each in triplicate or quadruplicate, and were analyzed with GraphPad Prism using the built-in 4-parameter logistic function.

189
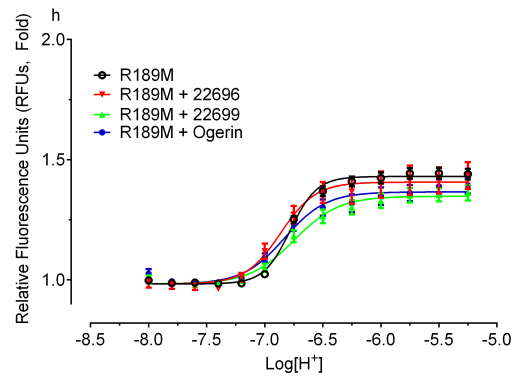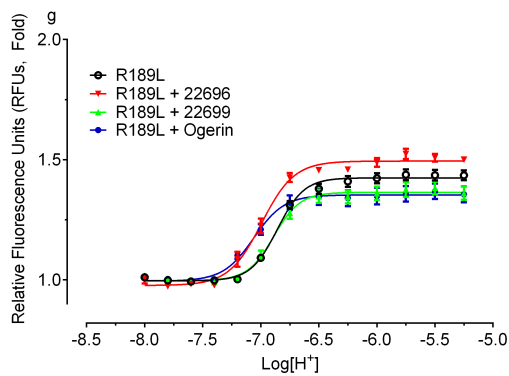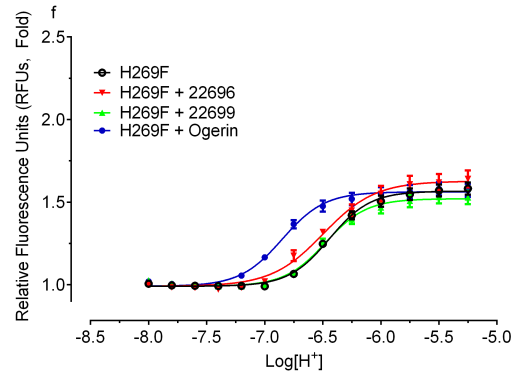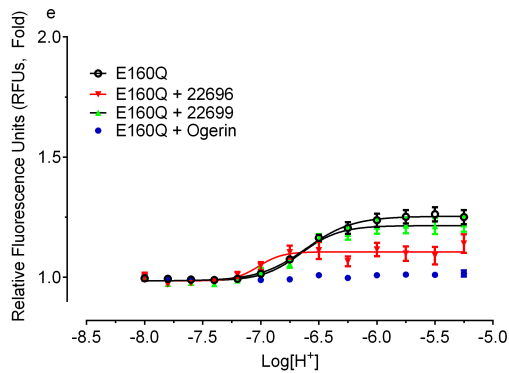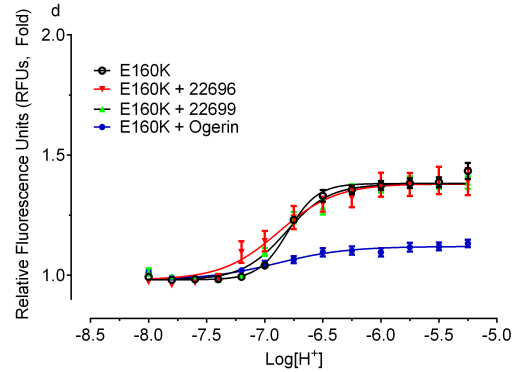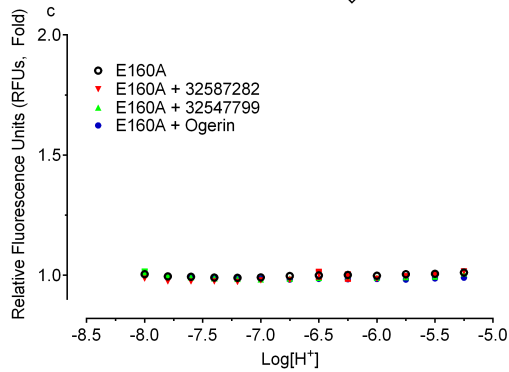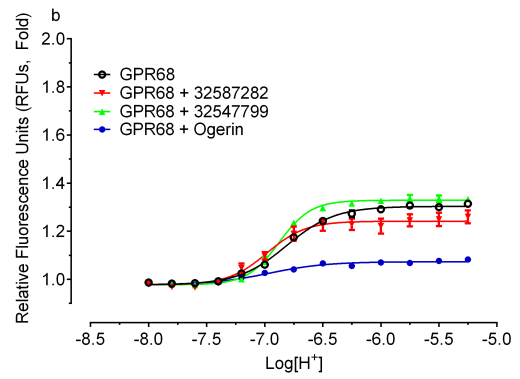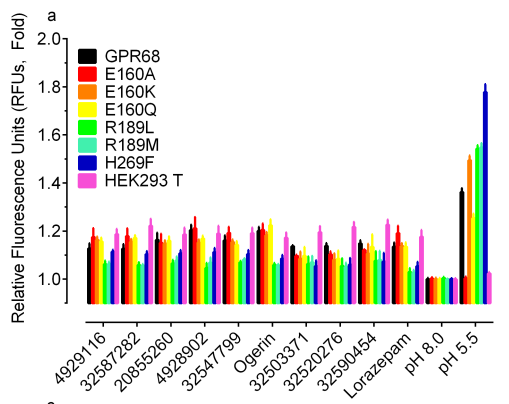
Supplementary Figure 16. Ogerin (ZINC 67740571) has minimal efficacy on proton-mediated cAMP production at GPR4 and GPR65 receptors. HEK293-T cells were transiently transfected and production of cAMP was measured using a split-luciferase reporter assay. Proton concentration-response curves at GPR4 (**a**) or GPR65 (**b**) were measured in the absence and presence of varying concentrations of ogerin. Results (RLU in fold of basal) represent mean ± SEM from a minimum of three independent assays, each in triplicate or quadruplicate, and were analyzed in GraphPad Prism using the built-in 4-parameter logistic function. Proton potency (pEC$_{50}$) and Hill slope values are 7.89 ± 0.05 and 5.02 ± 2.13 for GPR4; 7.27 ± 0.02 and 4.19 ± 0.56 for GPR65, respectively, in the absence of ogerin.

190
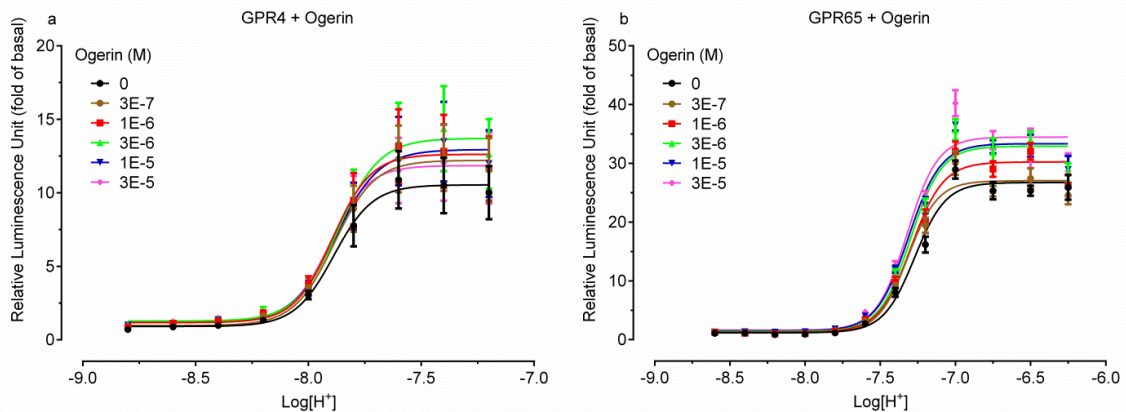
Supplementary Figure 17. Primary screening of 13 ogerin analogues (10 μM) at GPR68 receptors. The 14 ogerin analogues were identified from a virtual library of over 600 ogerin derivatives through docking study and synthesized in the lab (**Chemistry Supplementary**). HEK293-T cells were transiently transfected with GPR68 and production of cAMP was measured using a split-luciferase reporter assay. The functional assays were carried out at 5 different pH conditions (pH 8.4, 7.9, 7.4, 7.0, and 6.5) to reveal pH-dependent potentiation activity. A representative set of results (mean ± SEM, quadruplicate set) is shown and ogerin is included as a reference PAM. Activation of GPR68 receptors is indicated by increasing counts when pH was reduced. Compounds 33549, 33550, and 33554 showed little activity at high pH of 8.4, but became efficacious at pH 7.9 or 7.4. All the other compounds started to show activity when the pH was reduced to 7.0 or 6.5 – indicating that they all had pH-dependent PAM activity at GPR68.

GPR68 at pH 6.5

GPR68 at pH 7.0

GPR68 at pH 7.4

GPR68 at pH 7.9

GPR68 at pH 8.4

192
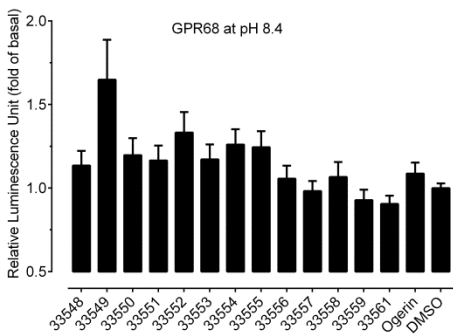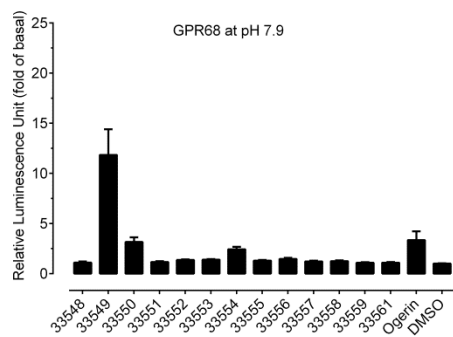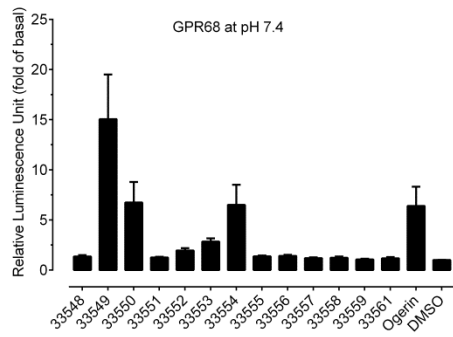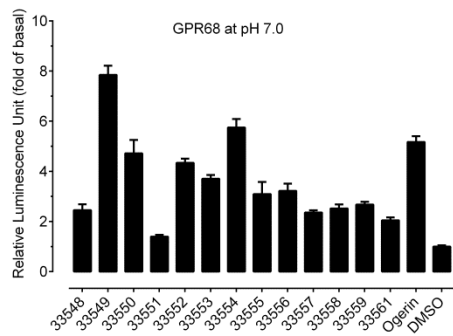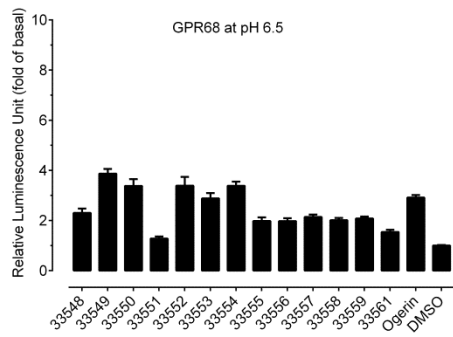
Supplementary Figure 18. Graphic comparisons of allosteric parameters logα and logβ. Estimated allosteric parameters were taken from **Supplementary Table 4** and used for the plots. (**a**) 13 Ogerin analogues (structures in **Supplementary Table 6**) and two additional analogues (structures in **Figure 4**); (**b**) Ogerin and its derivatives with different linker lengths. * indicates significant difference as compared with ogerin (C1), p < 0.05.

b
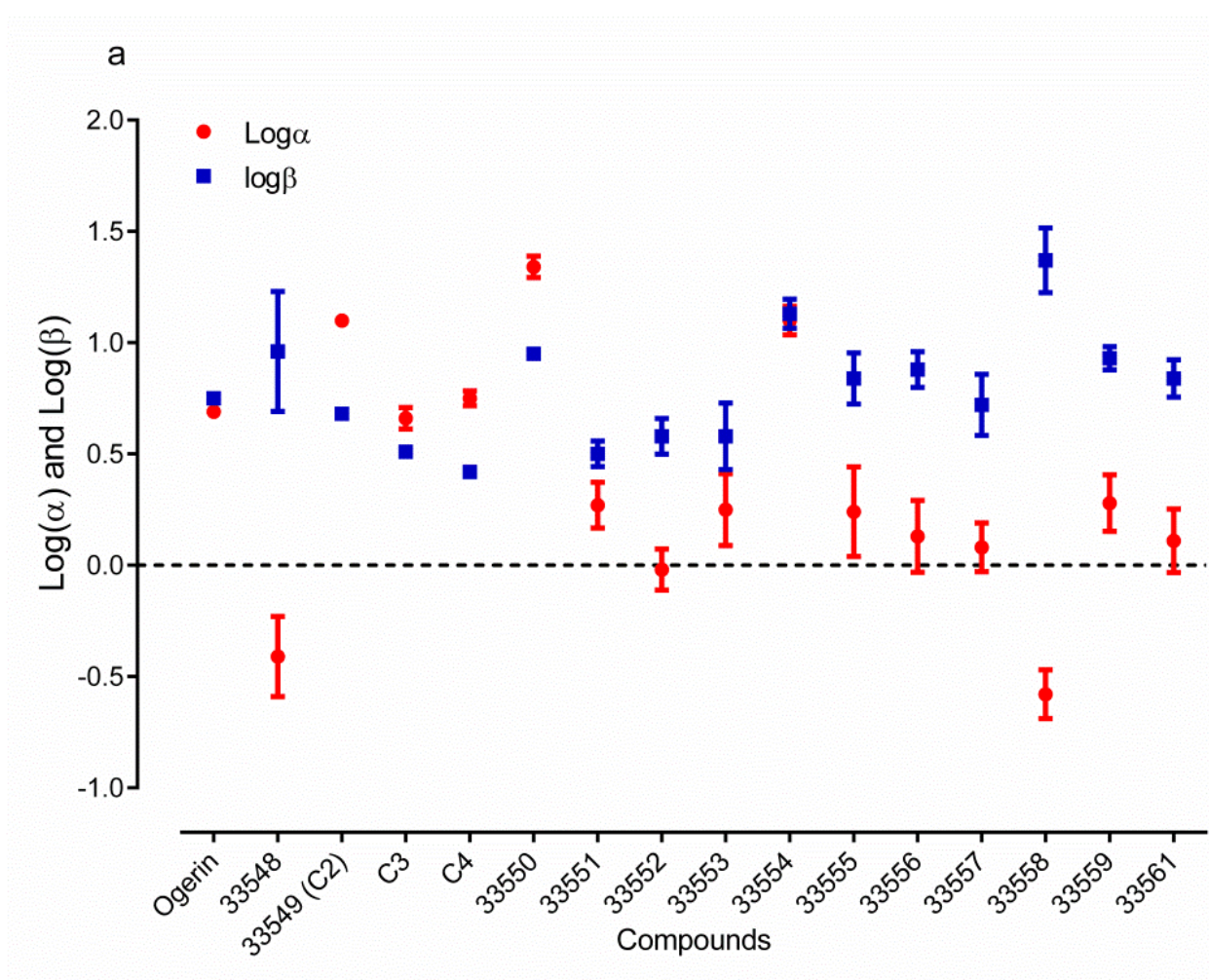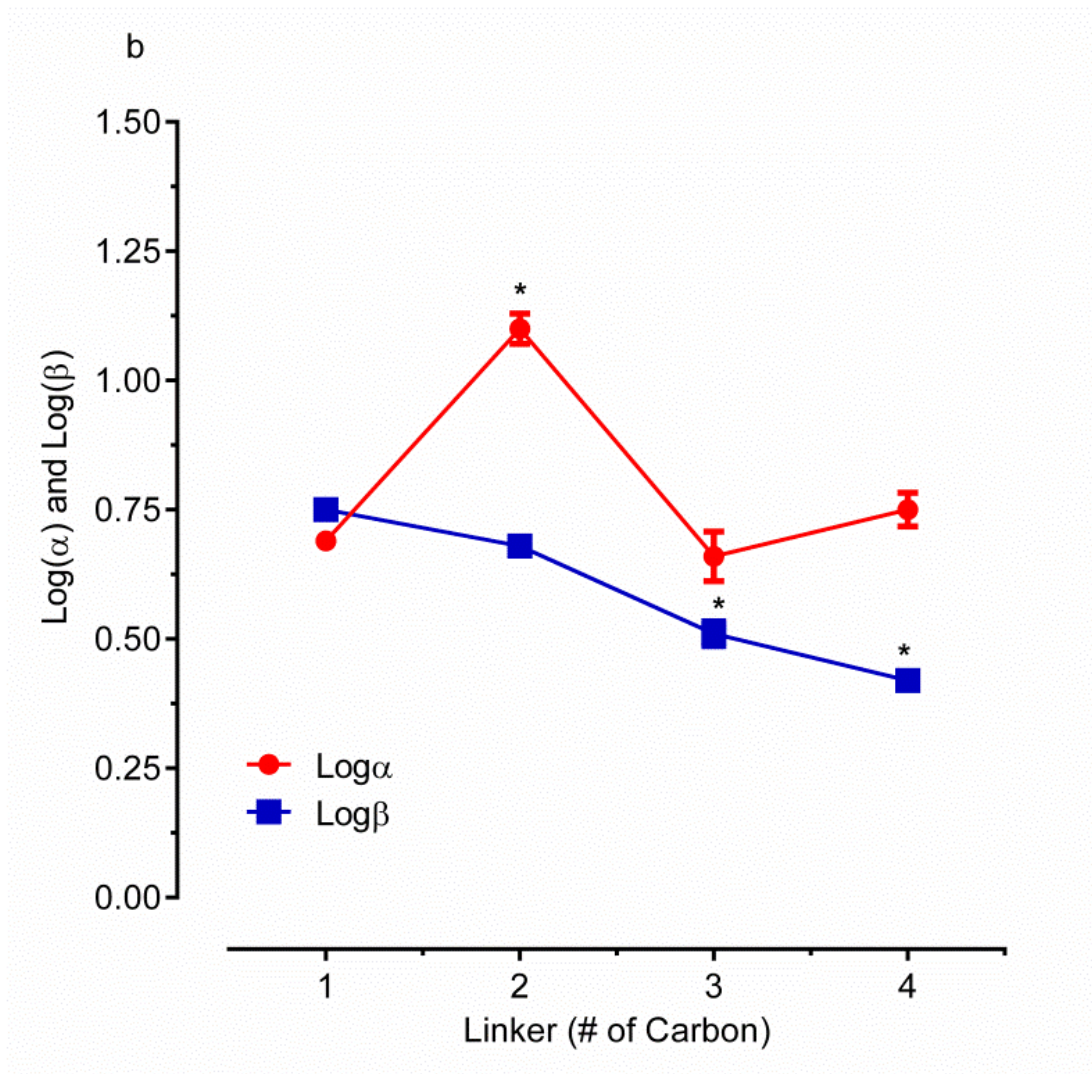
Supplementary Figure 19. Inverse agonist activity of ogerin at A_{2A} adenosine receptors and weak antagonist activity at 5-HT$_{2B}$ serotonin receptors. (a) HEK293-T cells were transiently transfected with the A_{2A} adenosine receptor, and production of cAMP was measured using a split-luciferase reporter assay (**Supplementary Methods**). 5'-N-Ethylcarboxamidoadenosine (NECA) and 2-Chloro-N^6-cyclopentyladenosine (CCPA) served as agonist controls, while CGS15943 served as an inverse agonist control at A_{2A} receptors. Results (RLU or RFU in fold of basal) represent mean ± SEM from a minimum of three independent assays, each done in triplicate, and were analyzed in GraphPad Prism with the built-in 4-parameter logistic function. The EC$_{50}$ of NECA is 1.64 nM; CGS15943 inverse agonist potency is 43.7 nM; ZINC 67740571 inverse agonist potency is 508 nM; CGS15943 antagonist IC$_{50}$ is 233 nM and corresponding K$_i$ is 12.1 nM; ZINC 67740571 antagonist potency is 11.1 M and corresponding K$_i$ is 0.59 μM, which is consistent with the K$_i$ of 0.23 μM from radioligand binding assays using ^{3}H-ZM241385 and stably expressed human A_{2A} receptors from HEK293 cells. **(b)** 5-HT$_{2B}$ stable HEK293 cells were used for ZINC 67740571 functional activity. Calcium mobilization assays were carried out according to the **Supplementary Methods**. Results represent mean ± SEM from one of three independent assays, each done in triplicate. The 5-HT EC$_{50}$ is 0.53 nM, estimated K$_i$ of ZINC 67740571 is 7.43 μM.
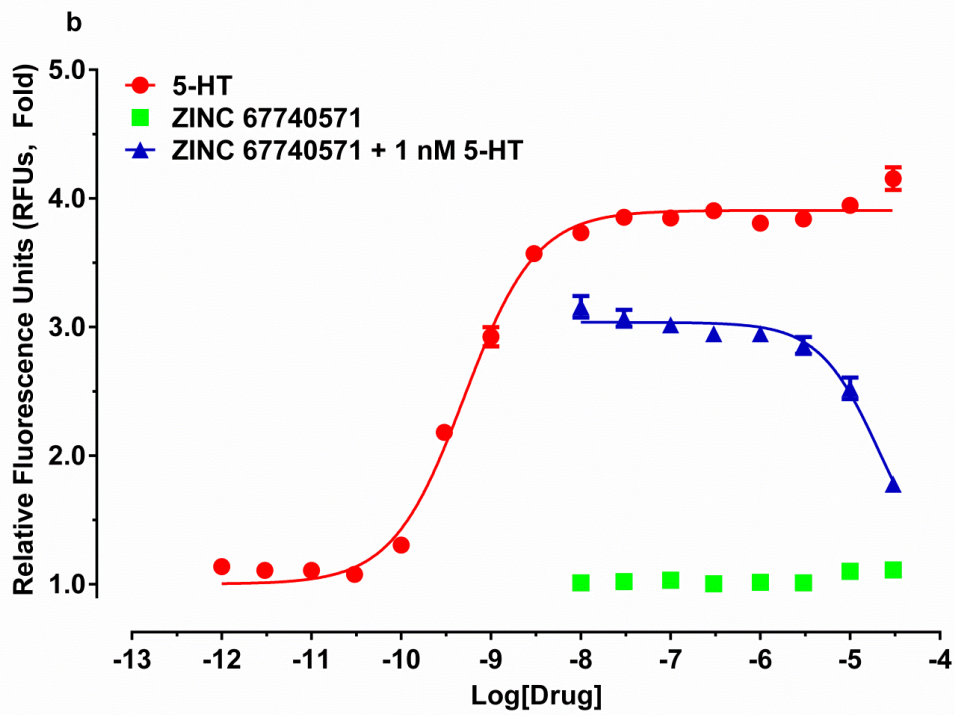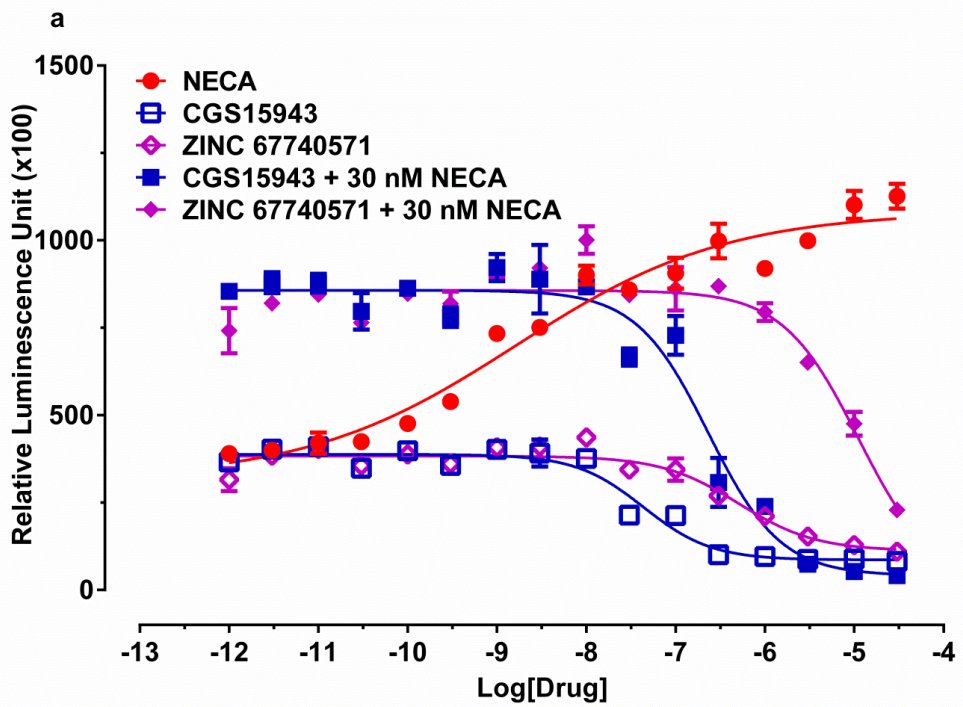
Supplementary Figure 20. New PAMs for GPR68-mediated cAMP production identified from Tocris Mini Library. Concentration-response curves of H^+ in the absence and presence of increasing concentrations of CGH 2466 (**a**), Tracazolate (**b**), SLV 320 (**c**) at GPR68 receptors. HEK293-T cells were transiently transfected with GPR68 and production of cAMP was measured using a split-luciferase reporter assay. Primary screening assays were performed at single final concentration of 10 M at pH 8.4 or pH 6.5 (results not shown). Compounds with potentiating activity at pH 6.5, but with low or no potentiation at pH 8.4, were selected for further characterization. Three compounds were confirmed to have PAM activity at GPR68 receptors. Results (RLU in fold of basal) represent mean ± SEM from a minimum of 3 independent assays, each in triplicate or quadruplicate, and were analyzed in GraphPad Prism using the built-in 4-parameter logistic function (right column, **a'**, **b'**, **c'**) and the standard allosteric operational model (left column, **a, b, c**). Allosteric parameters are summarized in **Supplementary Table 4**. For each pair of fittings, the proton potency value (pEC_{50}) from agonist concentration-response curve (right panel) in the absence of testing compound was used as the pK_A for the allosteric operational model (left panel); the highest E_{max} value from agonist concentration-response curves (right panel) in the presence of test compound was used as the E_{max} for the allosteric operational model (left panel).

GPR68, CGH 2466

GPR68, Tracazolate

GPR68, SLV 320
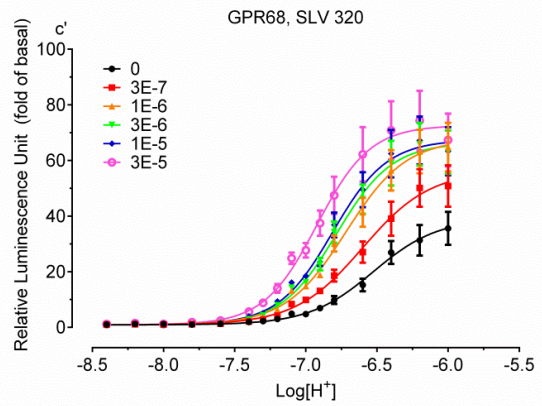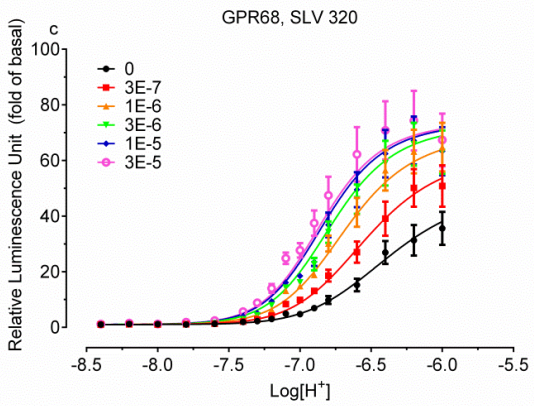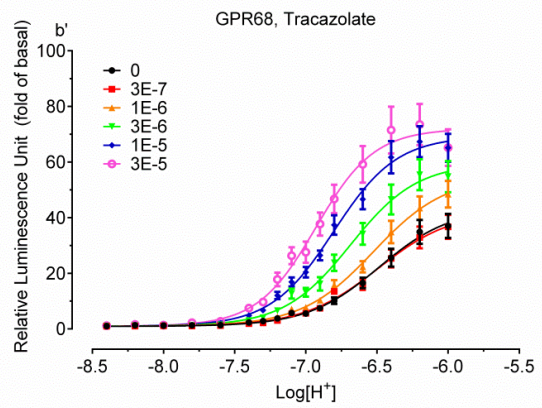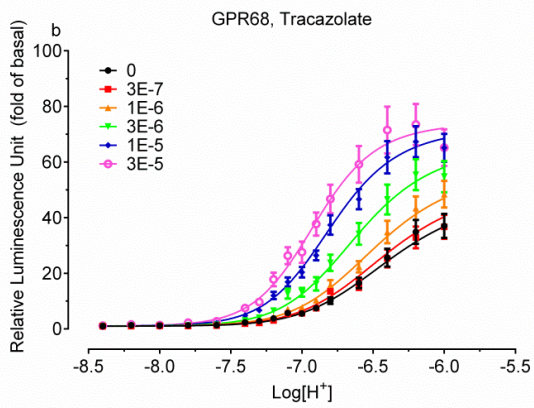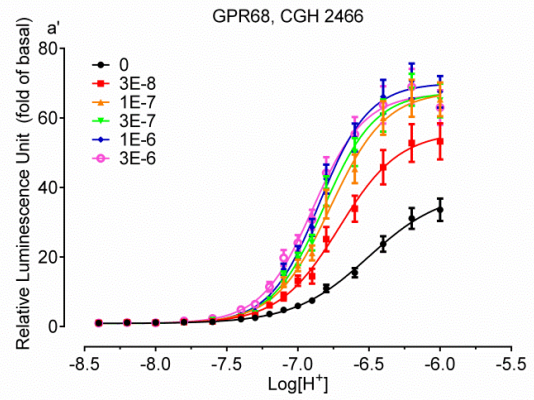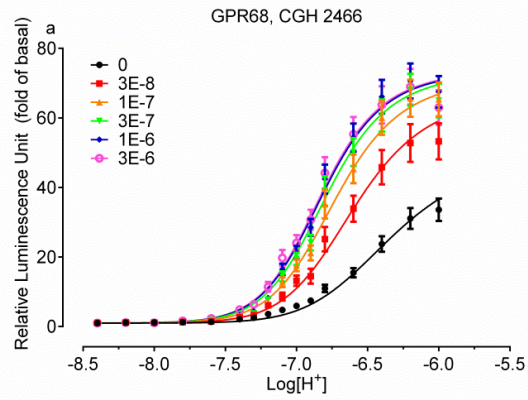
Supplementary Figure 21. Schematic showing shared pharmacology among GABA_A, adenosine GPCRs, and GPR68 ligands. The chemical structures of the endogenous ligands for each of GPR68, GABA_A, and adenosine GPCRs are given above the respective receptor for each. Molecules along each edge of the triangle have been shown to have activity at both targets, while tracazolate, in the middle, shows activity at all three.
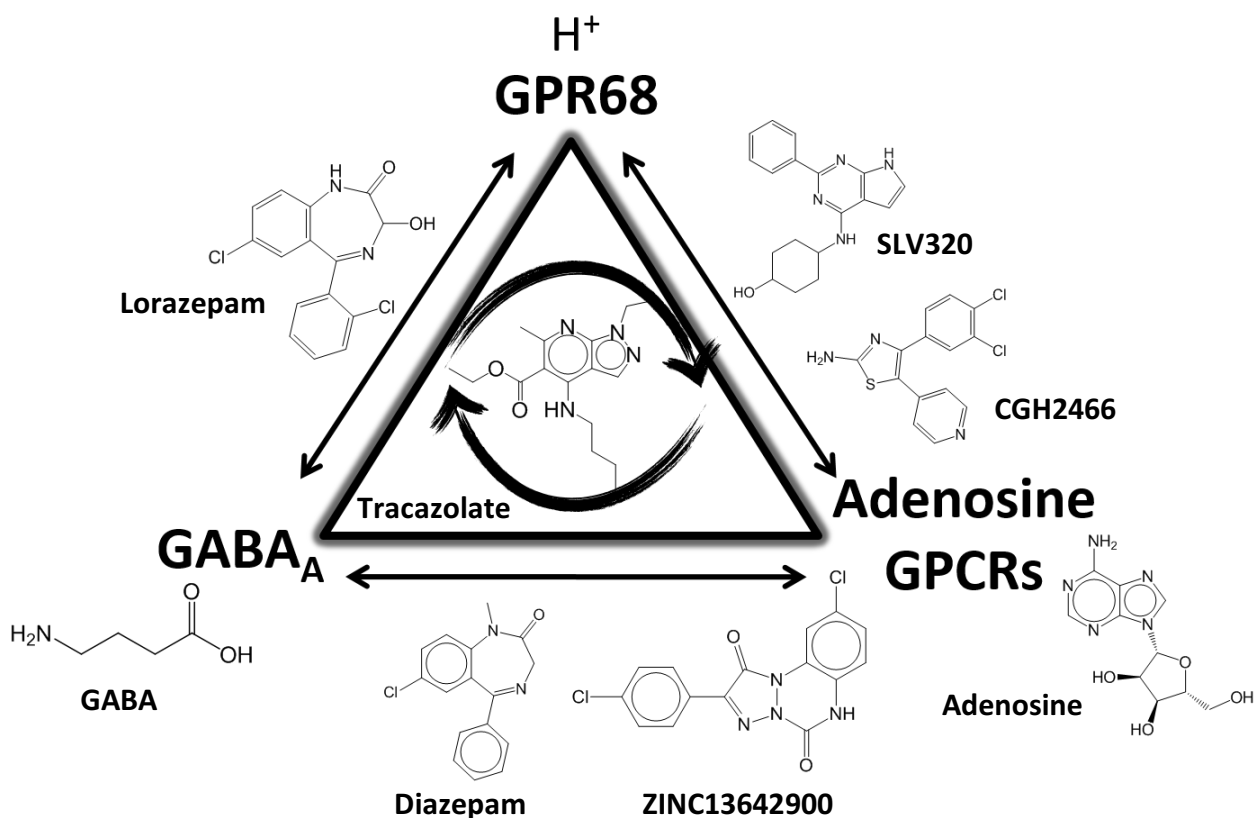
Supplementary Figure 22. GPR68 biology – no effect on learning and memory in GPR68 KO mice. To test learning and memory, animals (WT, n = 8; KO, n = 7) were assessed in a classical fear conditioning paradigm. Animals were presented with three pairings of tone (conditioned stimulus, CS) and foot-shock (unconditioned stimulus, US). Contextual (hippocampus-dependent) and cued memory (hippocampus-independent) were assessed 24 hr and 48 hr later, respectively, using measures of freezing (immobility). (**a - c**) GPR68 KO mice exhibited no difference in freezing during learning (**a**, no main effect of genotype ($F_{(1.52)}$ = 2.335, p =0.133) and genotype x time interaction ($F_{(3.52)}$ = 0.459, p = 0.172), two-way ANOVA), contextual memory retrieval (**b**, p = 0.951, t-test) and cued memory retrieval (**c**, p = 0.962, t-test) as compared to the wild-type mice.

a

Genotype Conditioning

b

Genotype Context

c

Genotype Cue
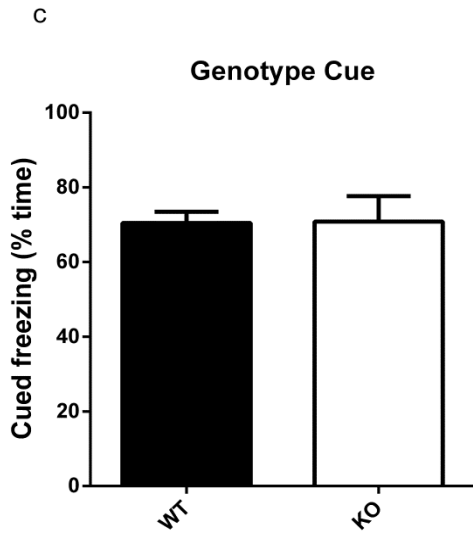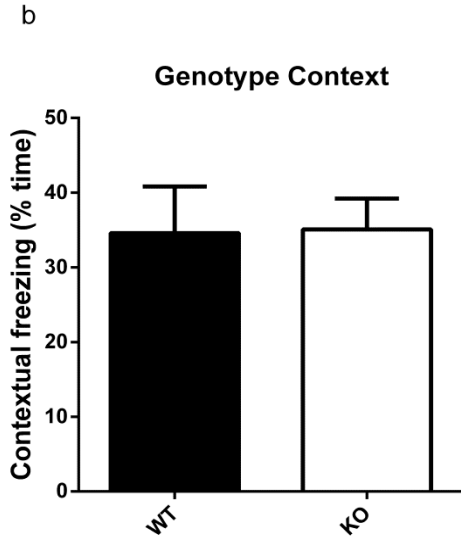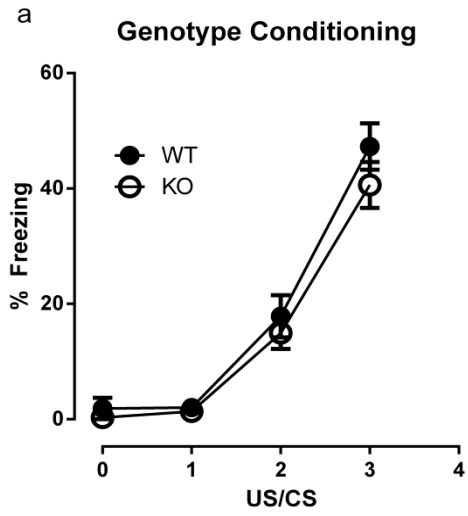
Supplementary Figure 23. GPR68 biology - Ogerin suppresses learning and memory in wild-type but not in GPR68 KO mice. GPR68 KO showed no effect on learning and memory (**Supplementary Figure 22**). To evaluate the effect of ogerin, strain-matched groups of animals (WT + vehicle, n = 7; WT + ogerin, n = 8; KO + vehicle, n = 8; KO + ogerin, n = 8) were given ogerin (10 mg/kg in 10% Tween 90 or saline) 30 min before training. Ogerin decreased the freezing behavior during learning in WT mice (**a**) - there was no main effect of drug alone, however, there was a significant drug x time interaction effect, $F_{(3.39)}$ = 5.86, p < 0.01; Bonferroni's post-hoc test showed a significant decrease of freezing for the WT + ogerin group in the third US/CS training, p < 0.01, two-way ANOVA. Ogerin also reduced contextual memory retrieval in WT mice (**c**), there was no main effect of drug or genotype, however, there was a significant drug x genotype interaction effect, $F_{(1.27)}$ = 4.71, p < 0.05; Bonferroni's post-hoc test showed that ogerin significantly decreased freezing in WT mice (p < 0.05, two-way ANOVA), but not in KO mice (p > 0.05, two-way ANOVA), or cued memory retrieval (**d**, no main effect of drug ($F_{(1.27)}$ = 0.0005, p = 0.982), genotype ($F_{(1.27)}$ = 3.281, p = 0.066) and drug x genotype interaction ($F_{(1.27)}$ = 0.0068, p = 0.935), two-way ANOVA) in wild-type mice. Ogerin had no effect on freezing behavior during learning (**b**, no main effect of drug ($F_{(1.42)}$ = 0.011, p = 0.918) and drug x time interaction ($F_{(3.42)}$ = 0.348, p = 0.79), two-way ANOVA) and did not change contextual (**c**) or cue memory retrieval (**d**) in GPR68 KO mice. Normalized memory retrieval results in panels **c** and **d** are presented in **Figures 5i** and **5j**.
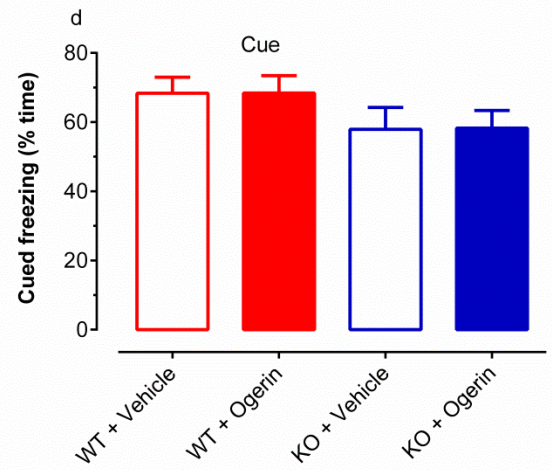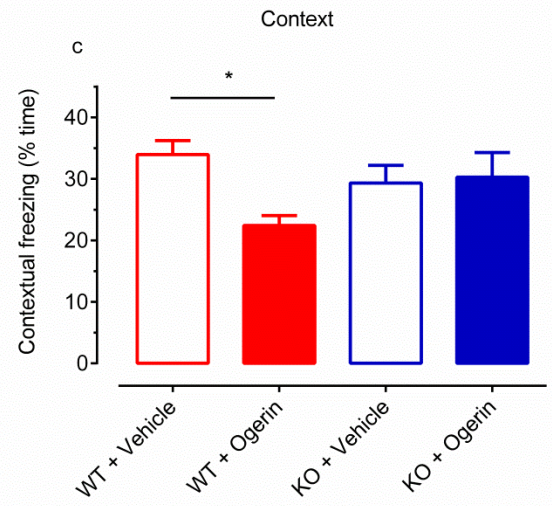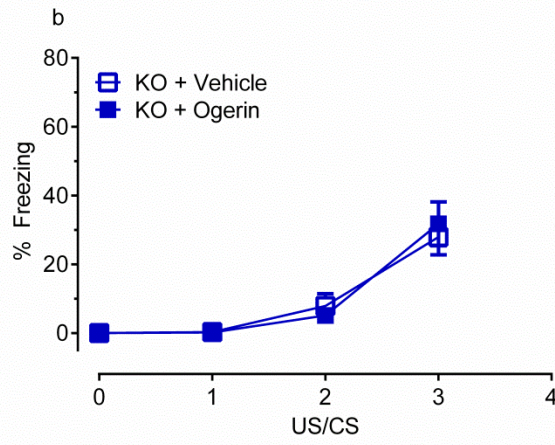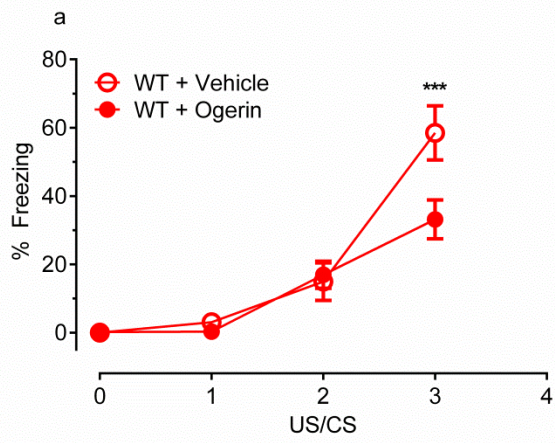
Supplementary Figure 24. The ogerin isomer ZINC 32547799 had no effect on learning and memory in wild-type mice. To test learning and memory, animals were assessed in a classical fear conditioning paradigm. Animals were presented with three pairings of tone (conditioned stimulus, CS) and foot-shock (unconditioned stimulus, US). Contextual and cued memory were assessed 24 hr and 48 hr later, respectively, using measures of freezing (immobility). Results with 10 mg/kg (vehicle, n = 6; drug, n = 7) are shown on the left (**a, b, c**) and 30 mg/kg (vehicle, n = 7; drug n = 8) on the right (**d, e, f**). For the 10 mg/kg dose, there was no significant difference between two groups during learning (**a**, drug x time interaction, $F_{(3.33)}$ = 0.119, p = 0.949; drug $F_{(1.33)}$ = 0.007, p = 0.937, two-way ANOVA), contextual memory retrieval (**b**, p = 0.635, t-test), and cued memory retrieval (**c**, p = 0.774, t-test). Normalized contextual memory retrieval (**b**) and cued memory retrieval (**c**) are presented in **Figure 5k** and **5l**. For the 30 mg/kg dose, there was a significant difference between two groups during learning (**d**, drug x time interaction, $F_{(3.39)}$ = 3.58, p = 0.022; drug alone $F_{(1.39)}$ = 1.19, p = 0.295; Bonferroni post-hoc test revealed a significant effect (p < 0.05) at the third US/CS training point, two-way ANOVA). However, there was no significant difference between two groups in either contextual memory retrieval (**e**, p = 0.615, t-test) or cued memory retrieval (**f**, p = 0.686, t-test).
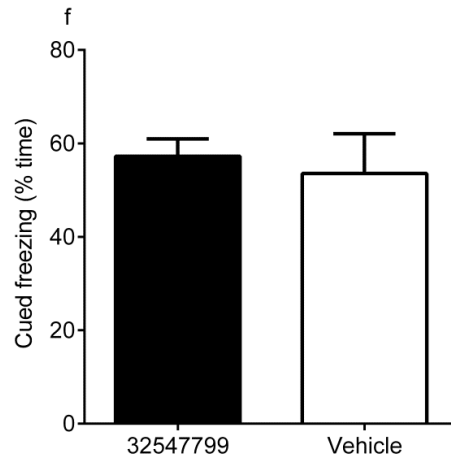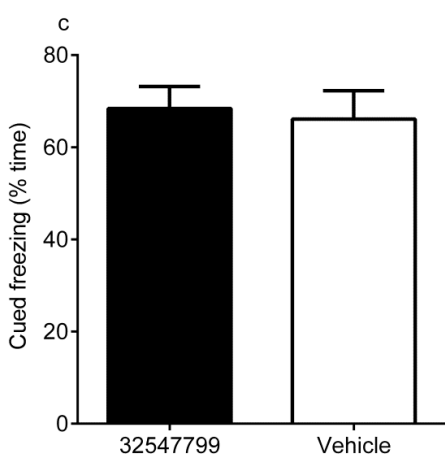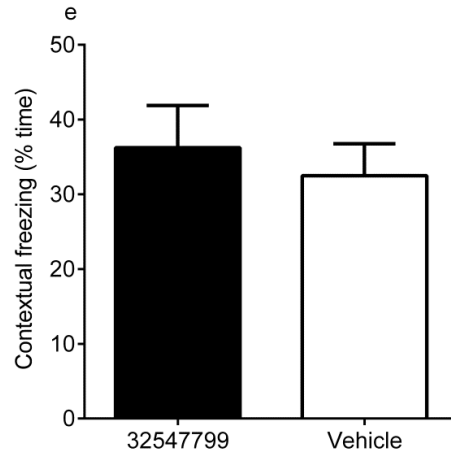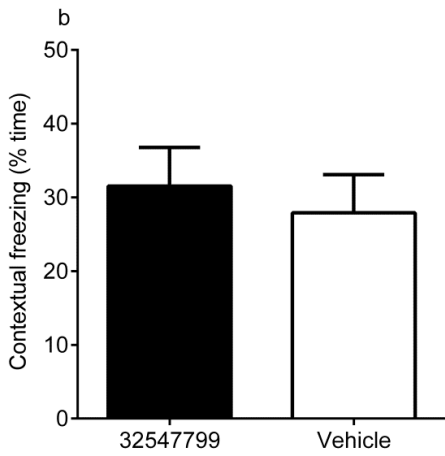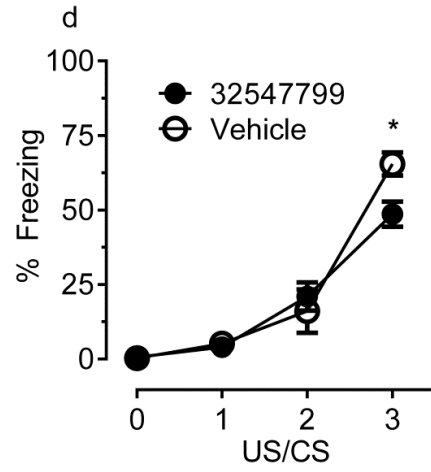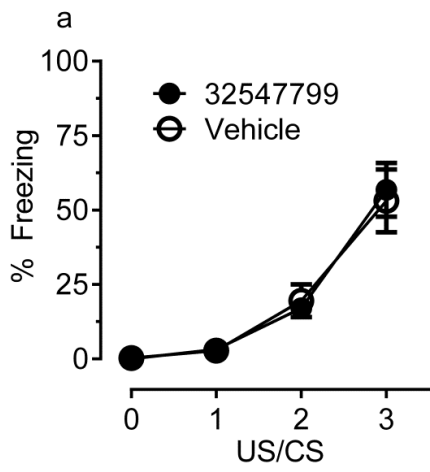
Supplementary Figure 25. Screening of ZINC compounds predicted to be active at GPR65 based on BTB09089 docking using a split luciferase cAMP reporter assay in transiently transfected HEK293-T cells. (**a**) BTB09089 showed weak agonist activity (**Figure 6d, 6f**), but failed to potentiate H$^+$ activity at GPR65. Results were normalized (RLU in fold of basal) with pH 8.40 as the baseline and represent mean ± SEM from 7 independent assays, each in triplicate or quadruplicate, and were analyzed in GraphPad Prism using the built-in 4-parameter logistic function. (**b - g**) Primary screening with ZINC compounds (30 μM) at GPR65 receptors when receptors were kept inactive at pH 8.40 for agonist activity (**b**) and at control HEK293 T cells for nonspecific activity (**c**); at GPR65 receptors when receptors were activated at pH 7.40 for potentiator or antagonist activity (**d**); at GPR65 receptors when receptors were activated by BTB09089 (30 M) at pH 8.40 for antagonist activity (**e**). Compounds with minimum of 20% inhibition at GPR65 at pH 7.40 in panel (**d**) but not at control HEK293 T cells in panel (**c**) were selected for further testing, in which isoproterenol (200 nM ISO at pH 8.40) was used to activate endogenous β_2 adrenergic receptors in GPR65-transfected cells (inactive at pH 8.40) to identify compounds selective for GPR65 receptors (**f**). Several compounds (such as 41613384, 9468042, 62678696, and 78874224) showed GPR65-specific inhibition (**f**). Compounds with a minimum of 20% inhibition against BTB09089 (30 M at pH 8.40) but not at control cells in panel (**c**) were also compared with their effect on ISO-activated signals to identify compounds selective to BTB09089-activated signal at GPR65 (**g**). Results (RLU in fold of basal) represent mean ± SM from a minimum of three assays (each in

minimum of triplicate and a total of ≥ 16 measurements). The red dashed line in

panels **c** – **g** indicated the 20% inhibition line - an arbitrary cut off line.

a

GPR65 pH 8.4

C

Control HEK293 T cells pH 7.4 or 8.4

d

GPR65 pH 7.4

e

GPR65 pH 8.4 + BTB09089
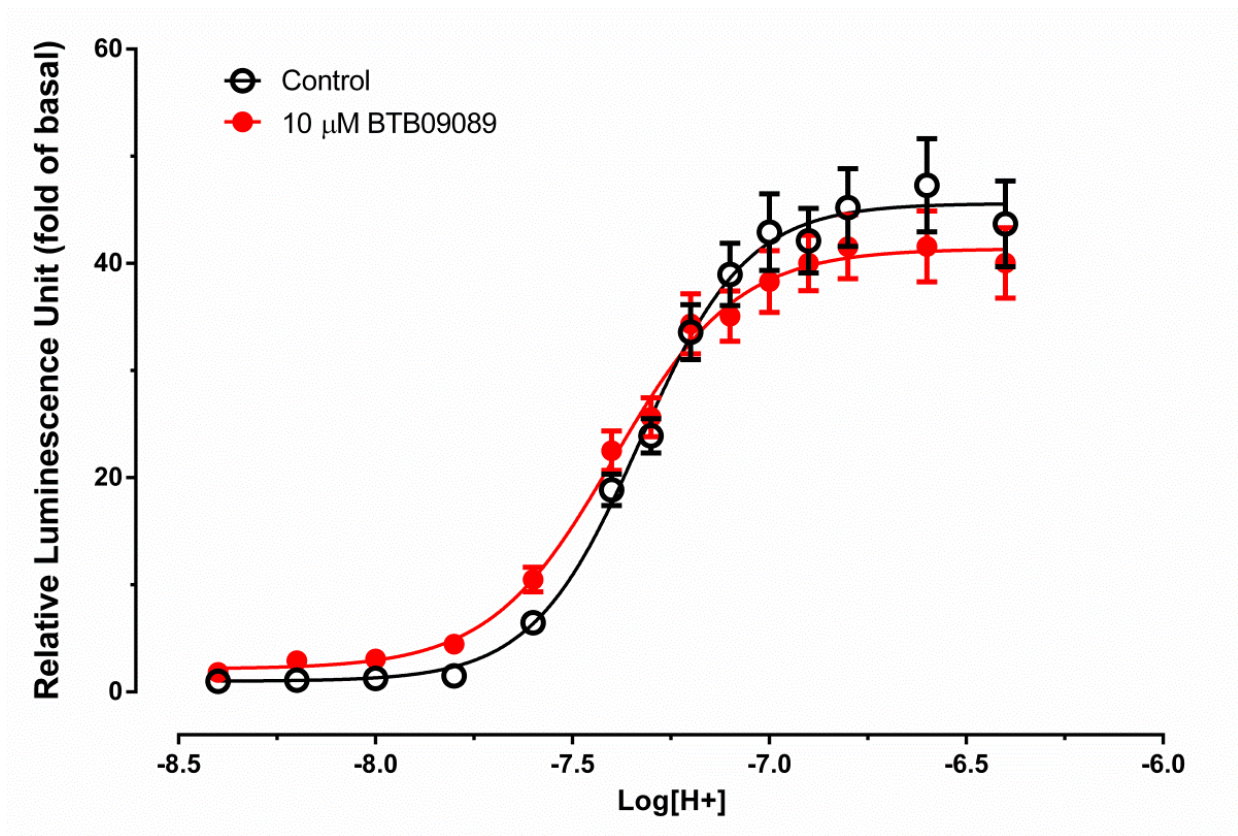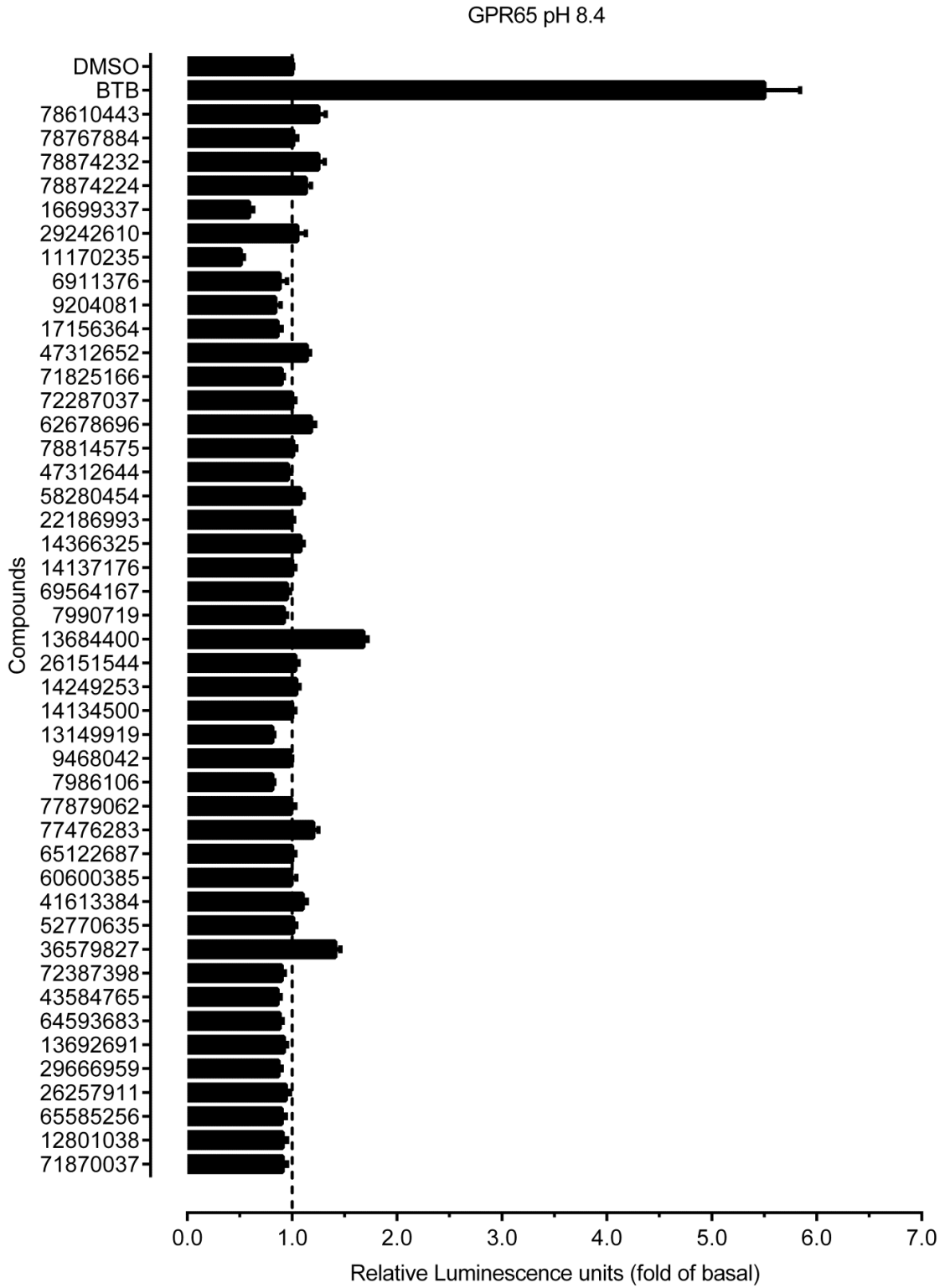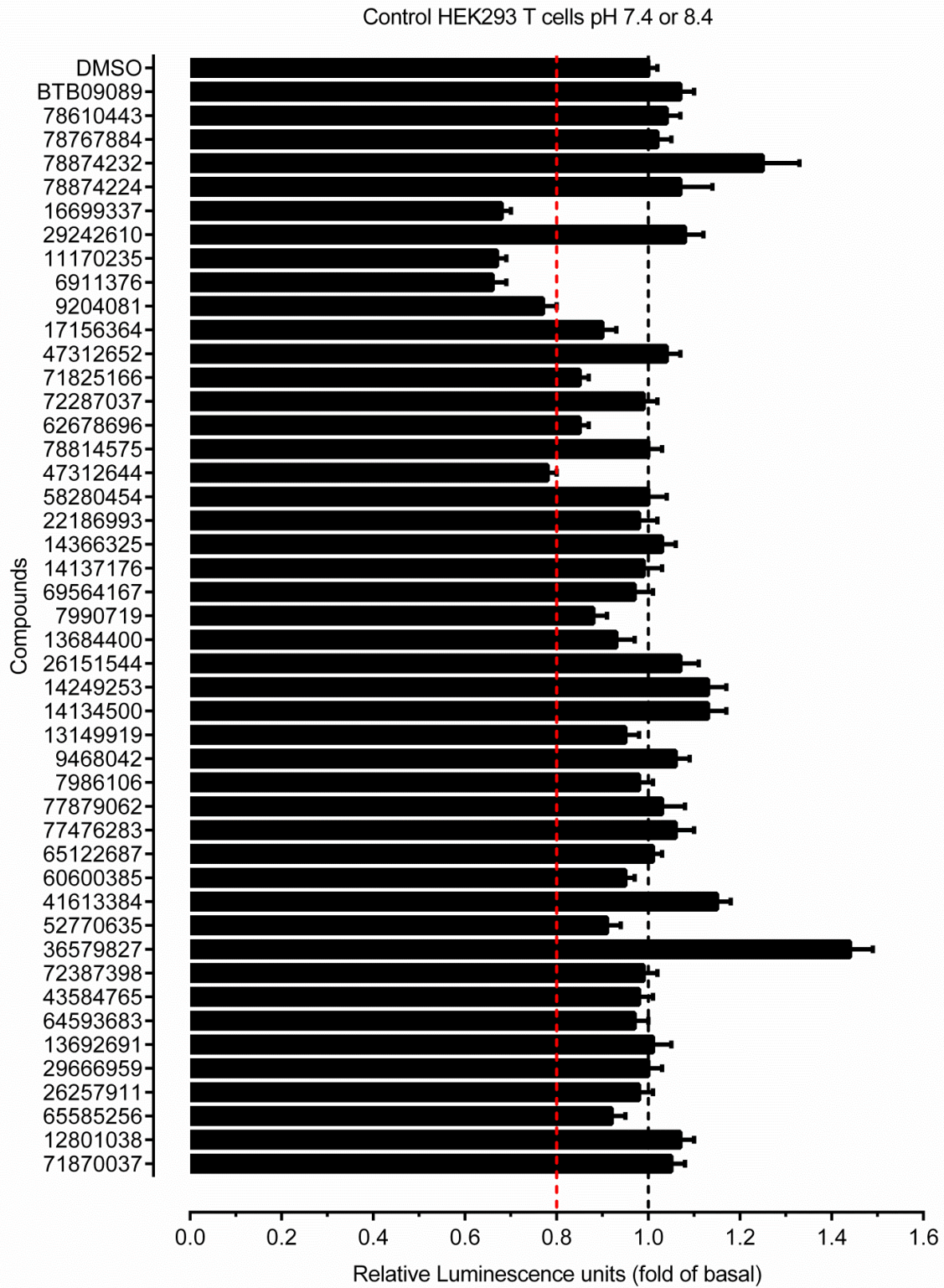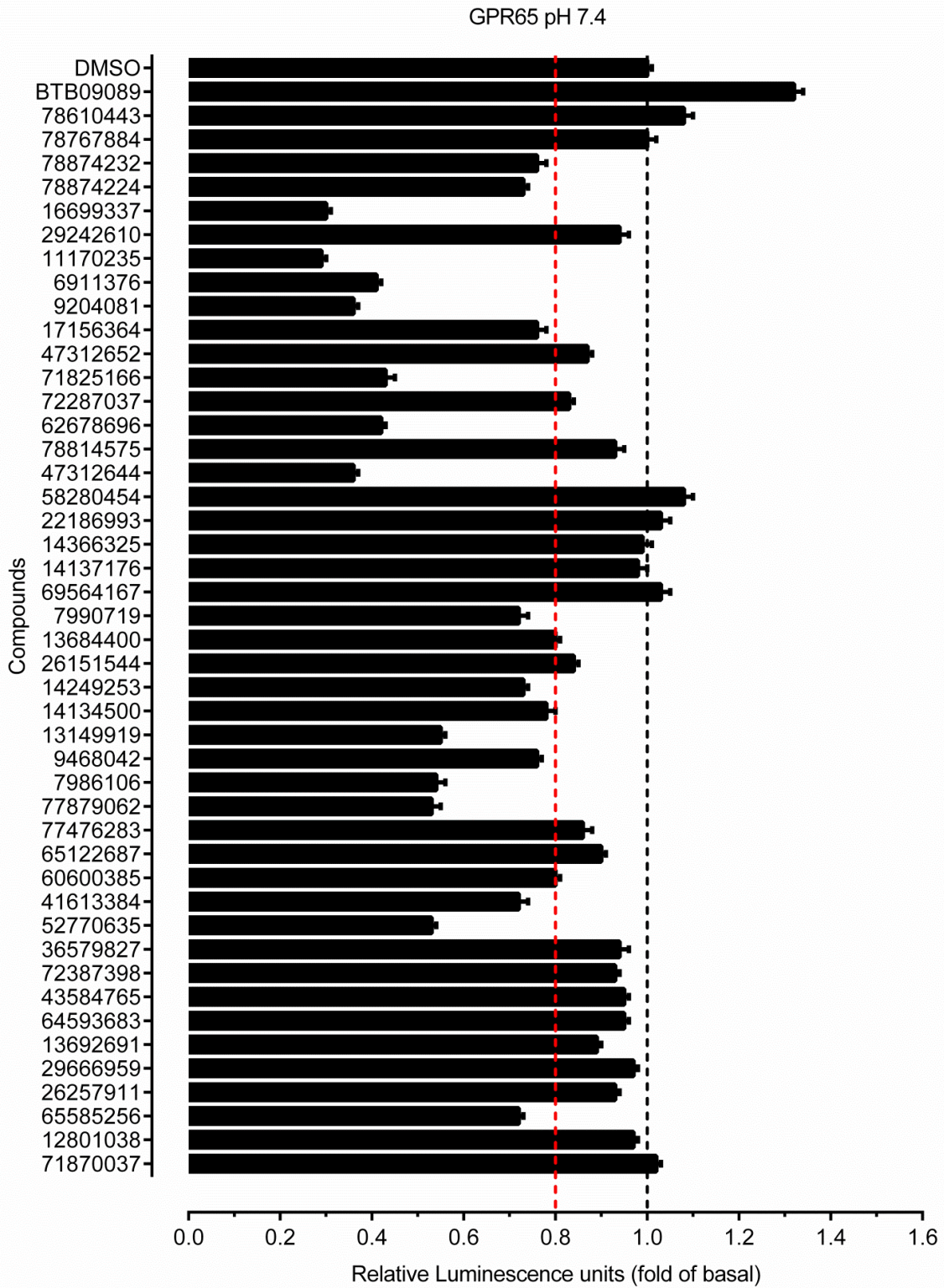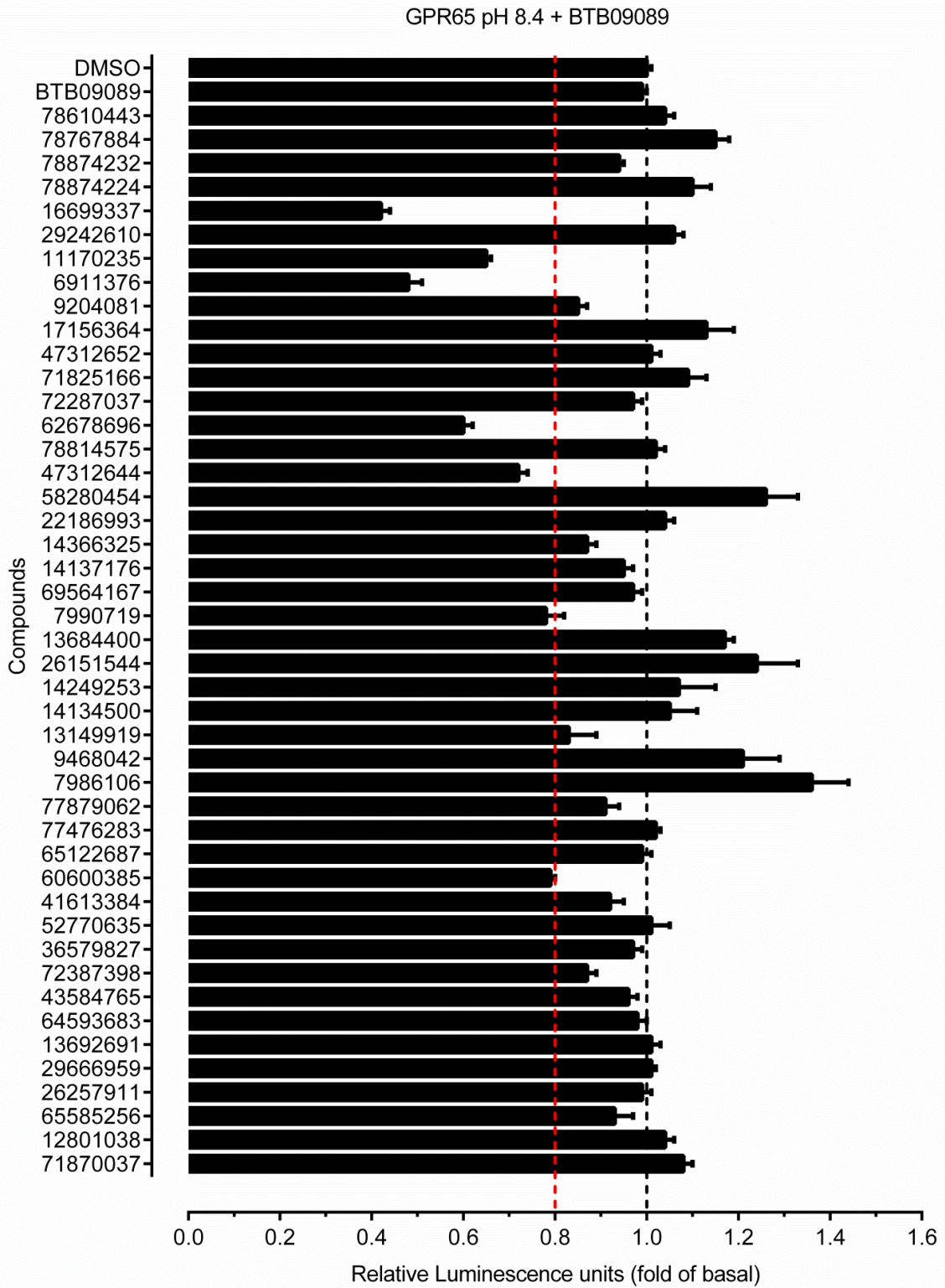
f

g
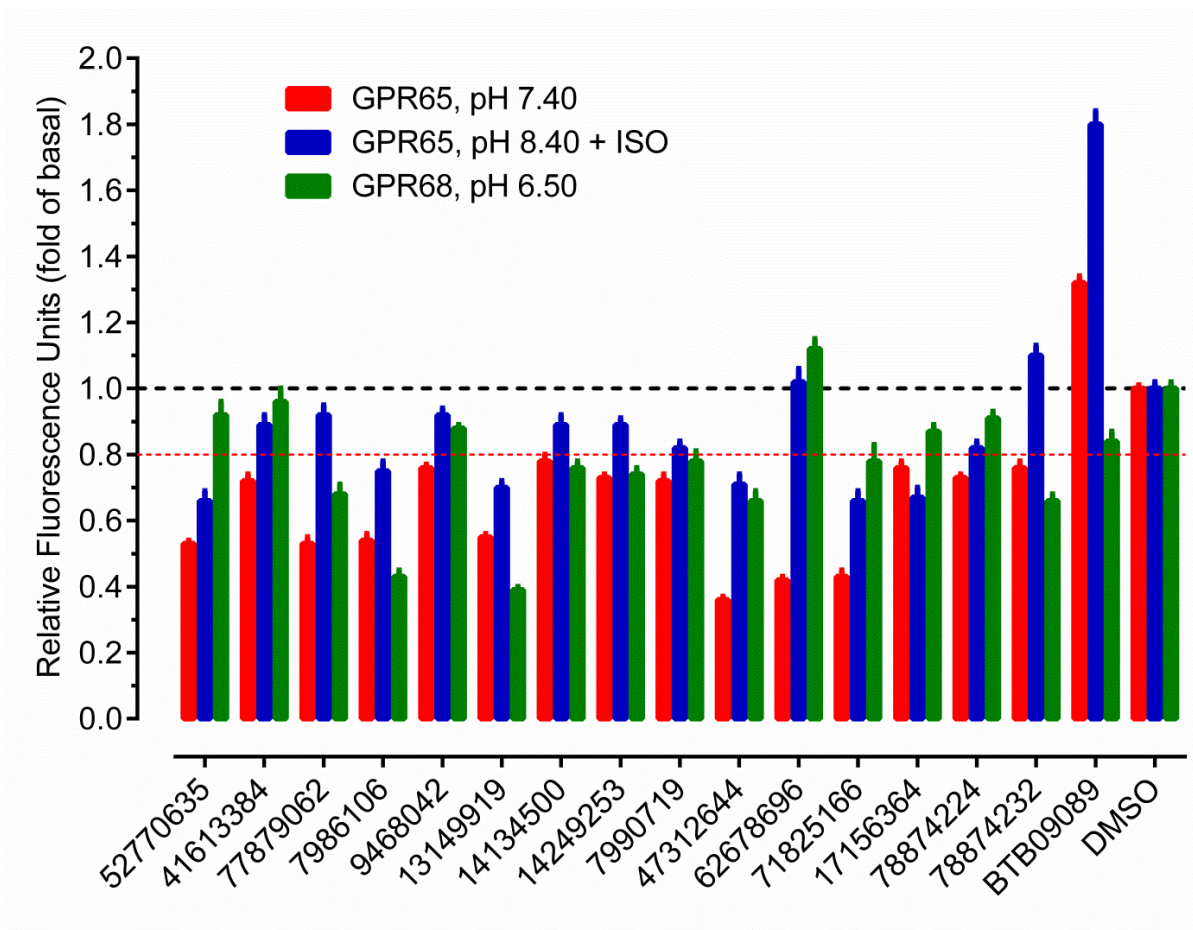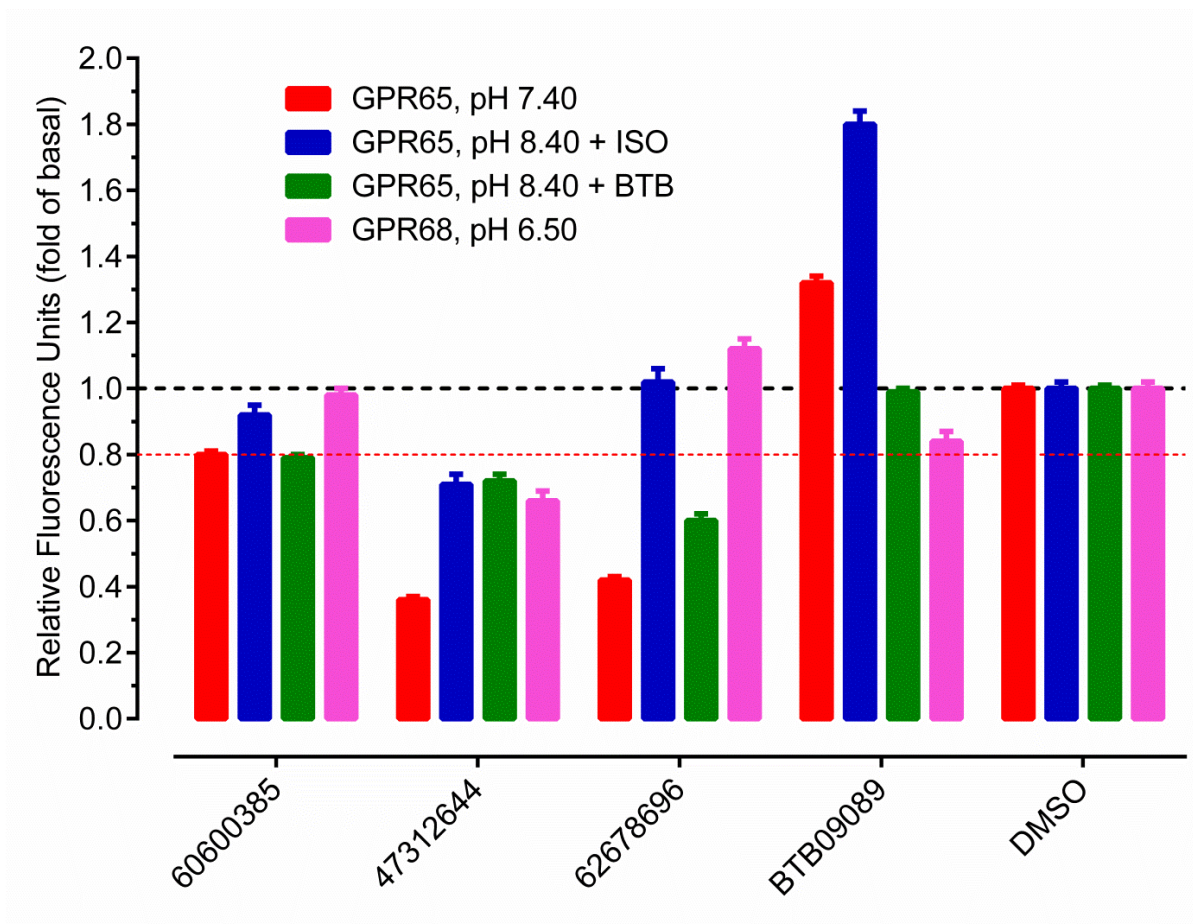
Supplementary Figure 26. Effects of mutations on GPR65 receptor pharmacology. (**a**) Proton concentration-responses, (**b**) BTB09089 concentration-responses, and (**c**) ZINC 13684400 concentration-responses were carried out with transiently transfected HEK293-T cells and production of cAMP was measured using a split-luciferase reporter assay. Results (RLU in fold of basal) represent mean ± SEM from a minimum of 3 assays (each in triplicate or quadruplicate) and were analyzed in GraphPad Prism with a standard 4-parameter logistic function. Corresponding curves of proton (from **Figure 3b**) and BTB09089 and ZINC 13684400 (from **Figure 6f**) at GPR65 wild-type receptors are also included (dashed lines) for comparison. Pharmacological parameters were listed in the **Supplementary Table 9**.
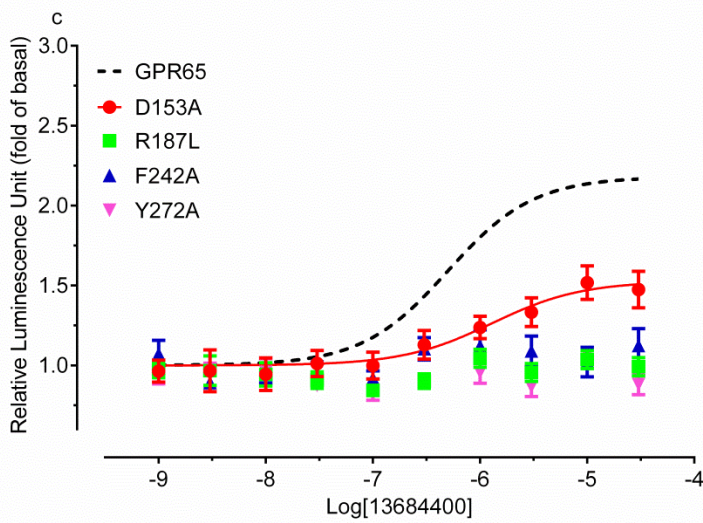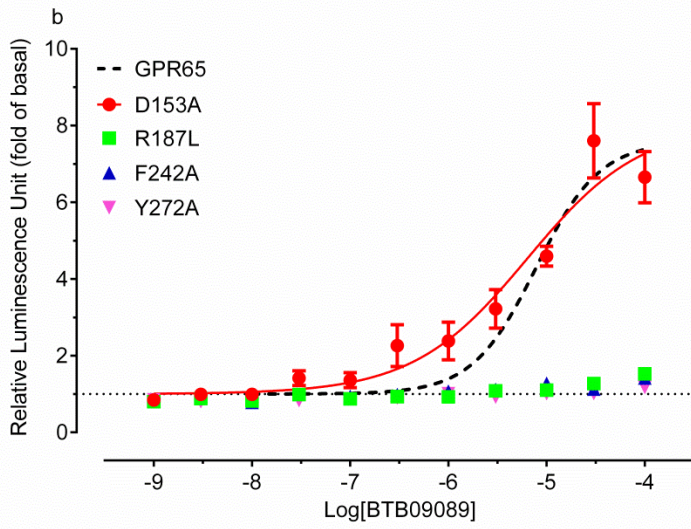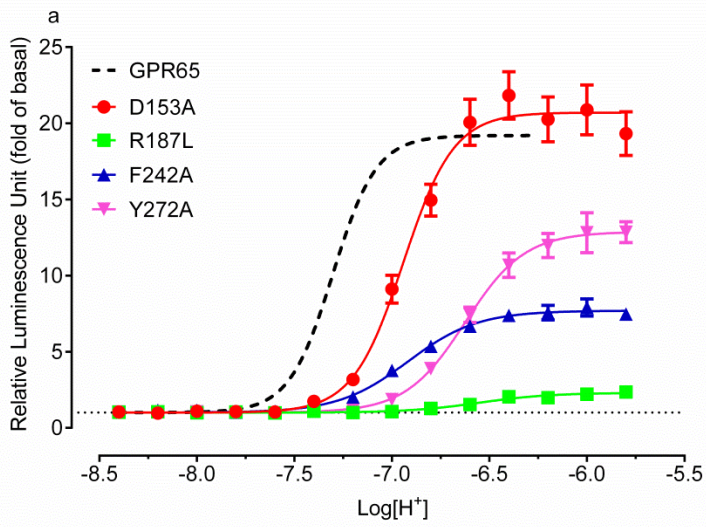
Supplementary Figure 27. Lead compounds showed neither agonist nor antagonist activity at CXCR4 receptors. Compounds (10 µM) were tested for agonist (**a**) and antagonist (**b**) activity at CXCR4 with CXCL12 as an agonist control (3 µM) and AMD 3100 (10 µM) as an antagonist control. HEK293-T cells were transfected with CXCR4 and cAMP production was measured using a split-luciferase reporter assay (**Supplementary Methods**). Agonist control (3 µM CXCL12) activated CXCR4 to reduce isoproterenol (200 nM)-elevated cAMP production; while test compounds showed no agonist activity (**a**). Antagonist control (10 µM AMD 3100) inhibited CXCL12 (50 nM) mediated G_i activation; while test compounds failed to do so (**b**). Results represent mean ± SEM from one representative assay of two independent assays, each done in triplicate.

a

b Antagonist activity: In the presence of 50 nM CXCL12
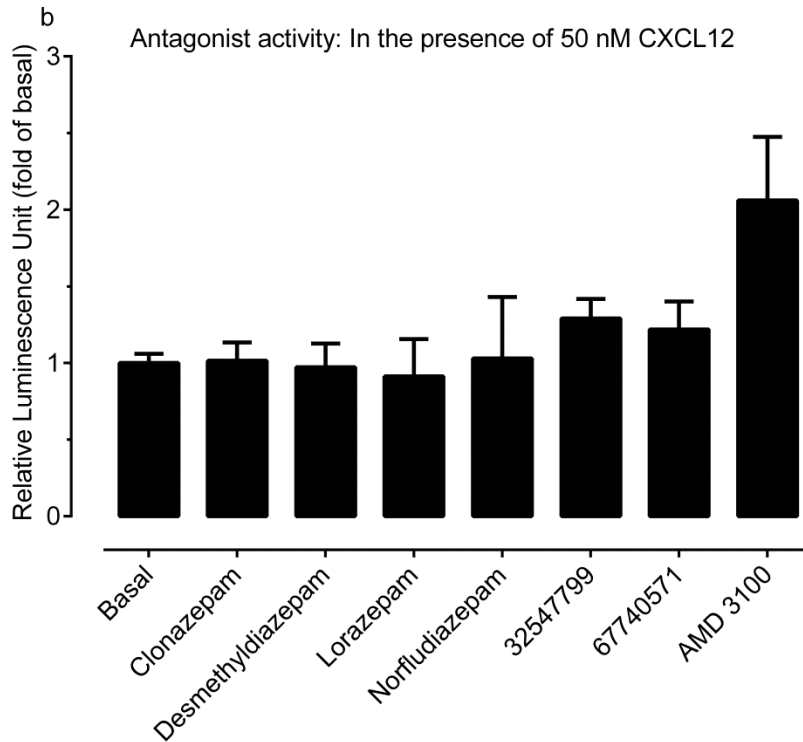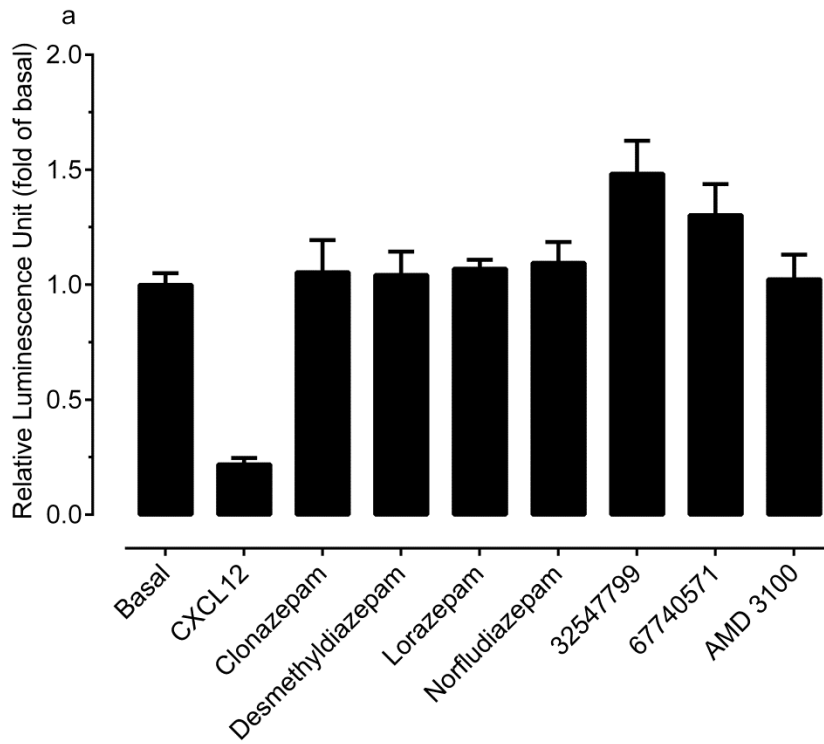
Supplementary Figure 28. No effects of GPR68 deletion on distance traveled in an open field. Data shown are mean ± SEM for each group for a one-hour test session. There were no significant differences between the WT and GPR68 KO mice for distance traveled, or for rearing or center time (data not shown), during the second activity test. A significant (sex x time) interaction was found for the distance measure [two-way ANOVA, $F_{(11,385)} = 2.68$, $p = 0.0025$], reflecting higher levels of activity in the female groups at the beginning of the session. Subject numbers were 9 WT and 7 KO male mice, and 12 WT and 11 KO female mice.
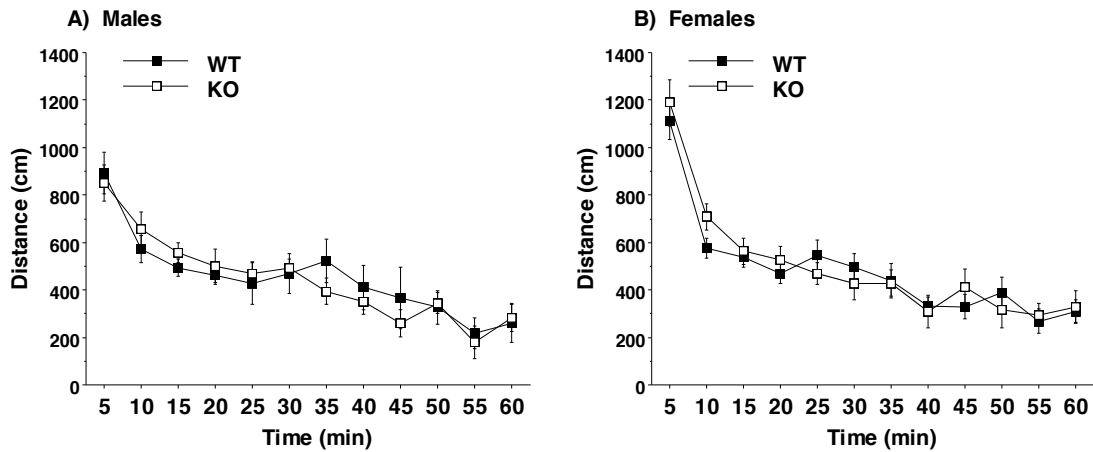
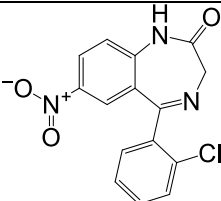**Supplementary Figure 29. Latency to fall from an accelerating rotarod.** Data shown are mean ± SEM for each group. There was a non-significant trend for enhanced performance in the male knockout group (repeated measures ANOVA, genotype x sex interaction, $F_{(1,35)}$ = 3.58, p = 0.0668). Subject numbers were 9 WT and 7 KO male mice, and 12 WT and 11 KO female mice.
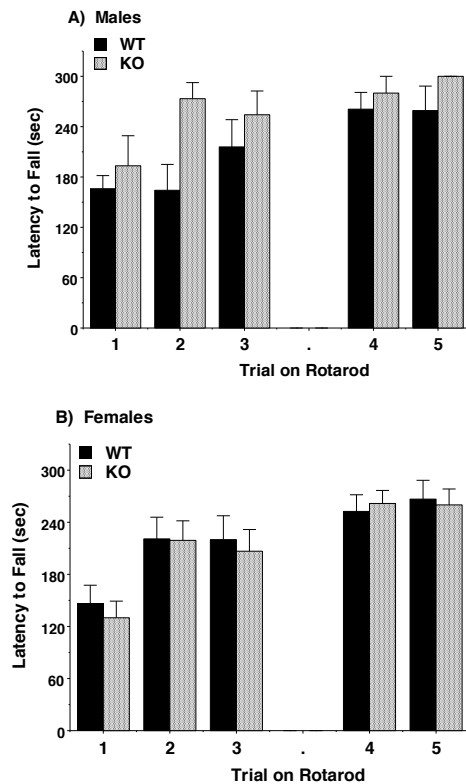
Supplementary Figure 30. Decreased startle responses in GPR68 KO mice following presentation of acoustic stimuli (30a and 30b). Data shown are mean ± SEM for each group. Trials included no stimulus (No S) trials and acoustic startle stimulus (AS) alone trials. No effects of genotype were found for levels of prepulse inhibition (**30c** and **30d**). Data shown are means ± SEM for each group. Subject numbers were 9 WT and 7 KO male mice, and 12 WT and 11 KO female mice. *p<0.05.
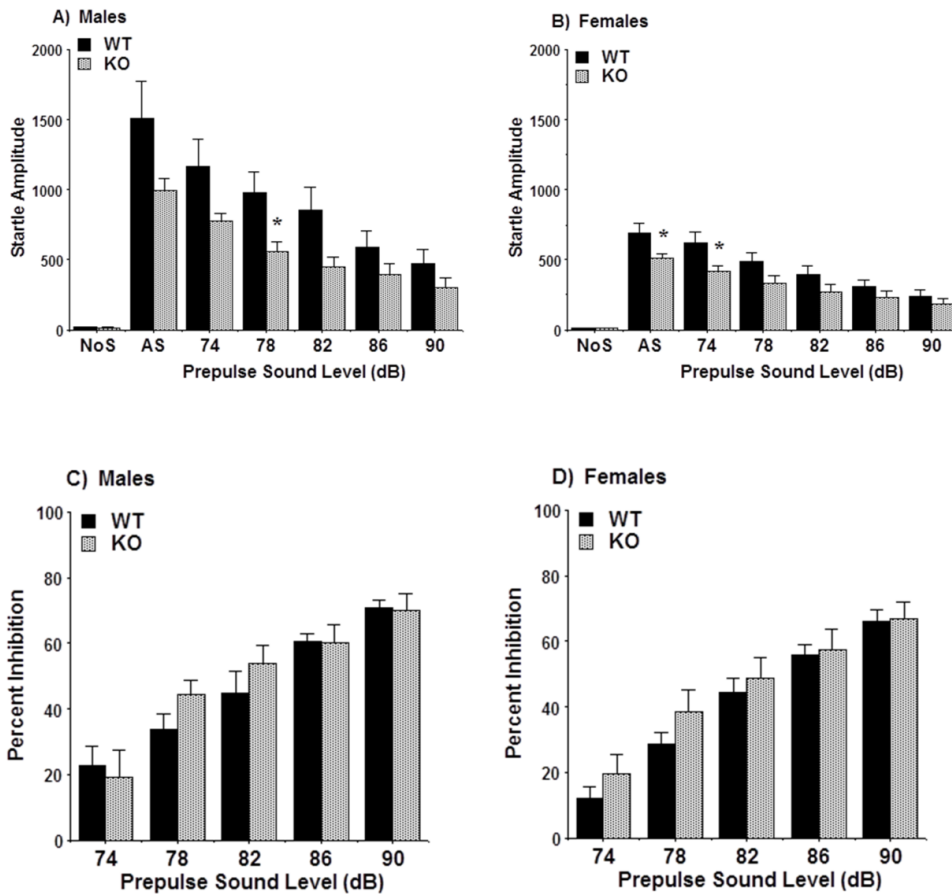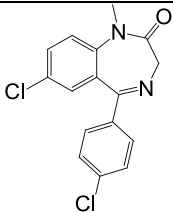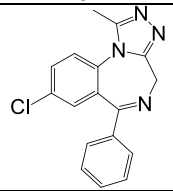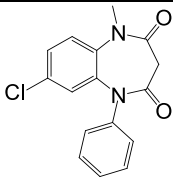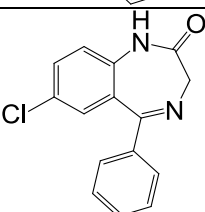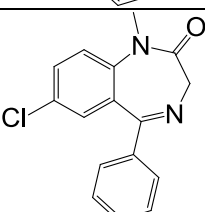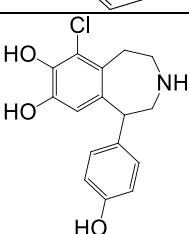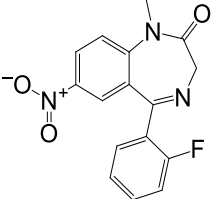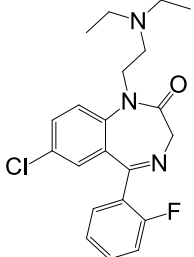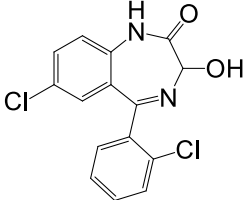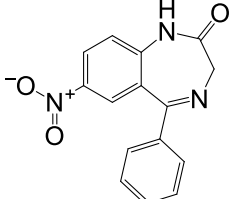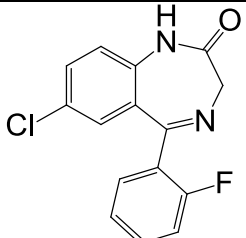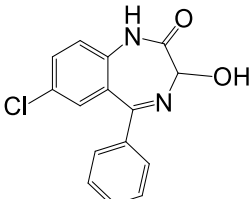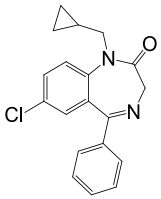
Supplementary Figure 31. Acquisition and reversal learning in the Morris water maze. Data shown are mean ± SEM of four trials per day. There were no significant effects of either genotype or sex on acquisition and reversal learning in the Morris water maze (**Supplementary Figures 31a** and **31b**). Both wild type and GPR68-knockout mice were able to learn the location of a hidden platform. However, during the reversal learning phase (**Supplementary Figures 31c** and **31d**), only the KO mice reached the criterion for learning (a mean of 15 seconds or less to reach the platform) across the 7 days of testing. In this case, the male KO group reached the criterion by day 5, and the female KO groups reached the criterion on day 6. The WT groups never reached the criterion for reversal learning. No effects of genotype were found during the one-minute probe trials (data not shown). Subject numbers were 9 WT and 7 KO male mice, and 12 WT and 11 KO female mice.
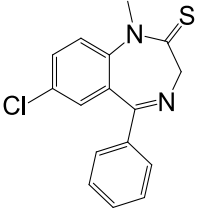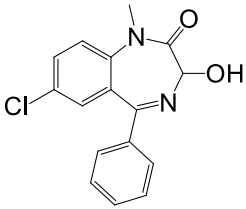
A) Acquisition, Males

B) Acquisition, Females

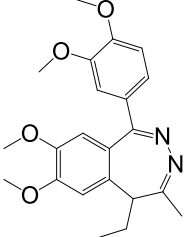C) Reversal, Males

D) Reversal, Females
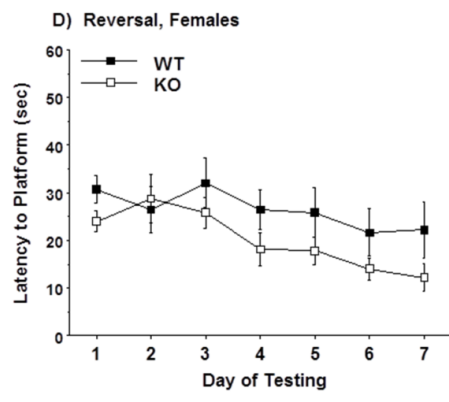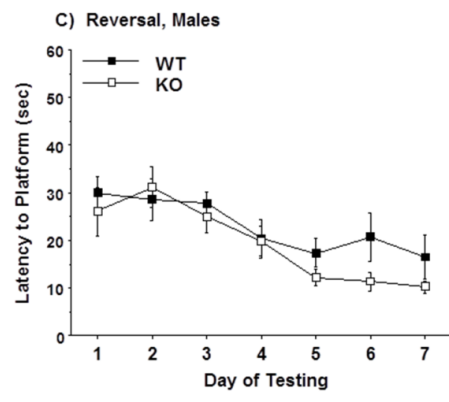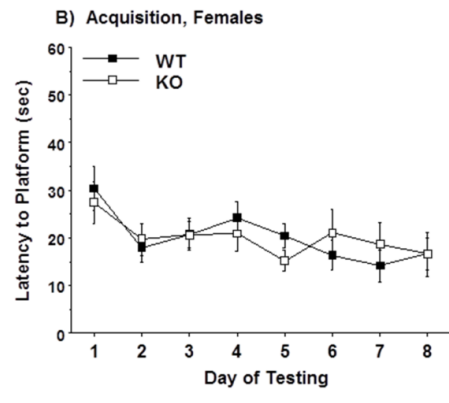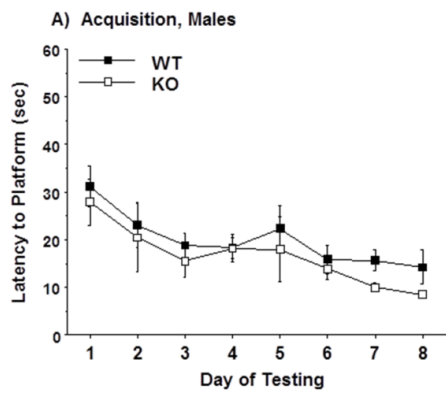
III. Supplementary Tables
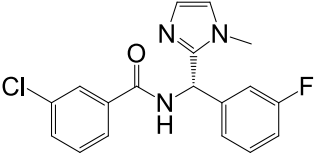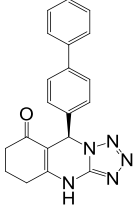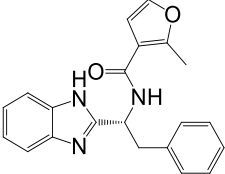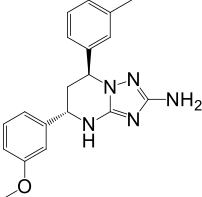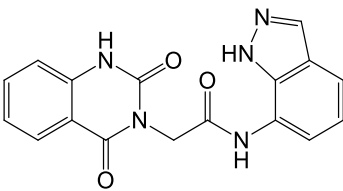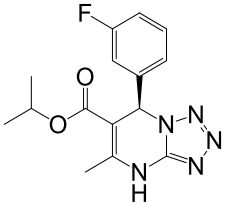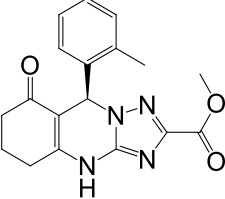
Supplementary Table 1. Structures of experimentally tested benzodiazepines and related molecules. These compounds were purchased from Sigma or Tocris.
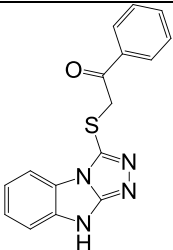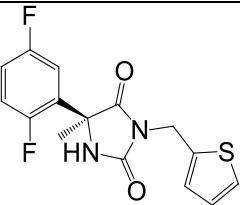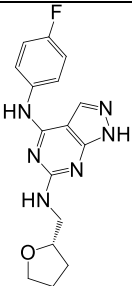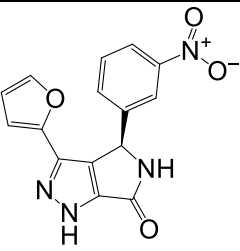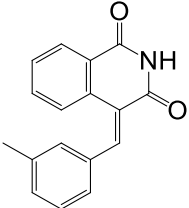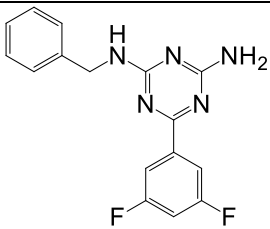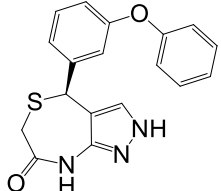
Name	Structure
4-Chlorodiazepam	
Alprazolam	
Clobazam	
Clonazepam	
Desmethyldiazepam	
Diazepam	
Fenoldopam	
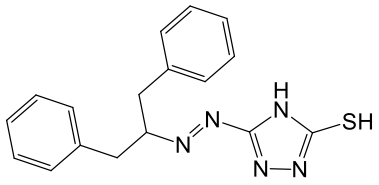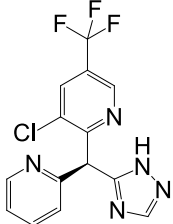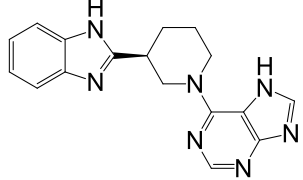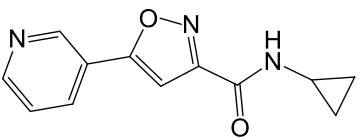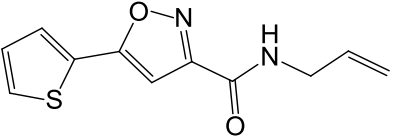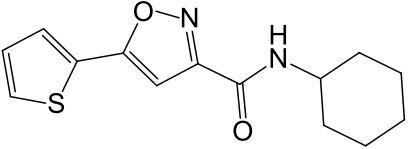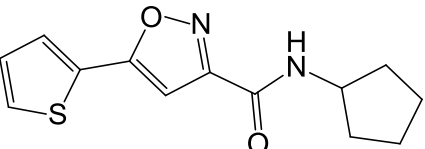
Name	Structure
Flunitrazepam	
Flurazepam	
Lorazepam	
Nitrazepam	
Norfludiazepam	
Oxazepam	
Prazepam	

Name	Structure
Sulazepam	
Temazepam	
Tofisopam	
Triazolam	

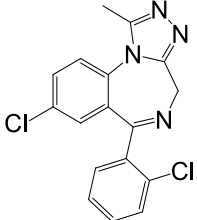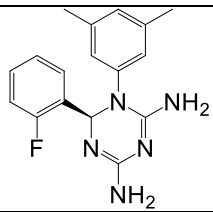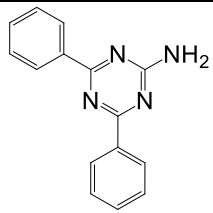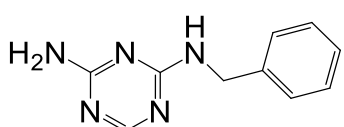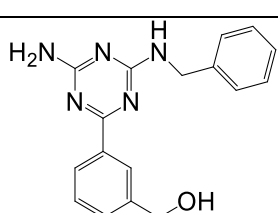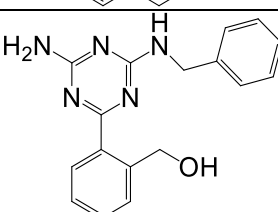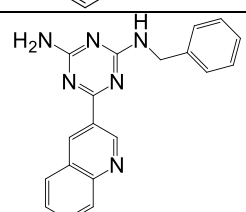**Supplementary Table 2. Structures of molecules chosen from GPR68 virtual screen for experimental testing.** Vendor information is available from ZINC database (http://zinc.docking.org/).
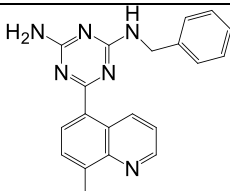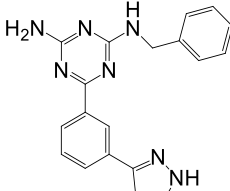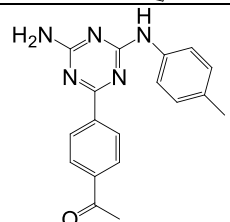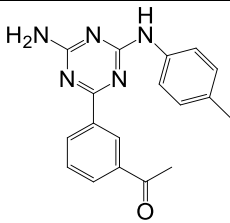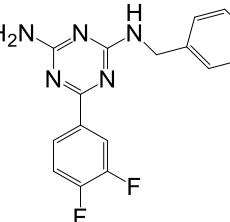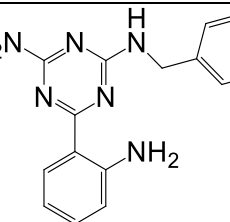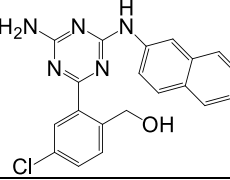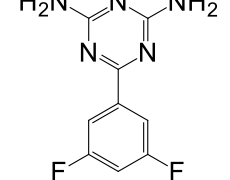
ZINC ID	Structure	Virtual Screen Docking Rank to GPR68 Model
ZINC12558970		133
ZINC18196037		215
ZINC24748979		281
ZINC00525649		284
ZINC32939481		343
ZINC15080047		541
ZINC00222801		673
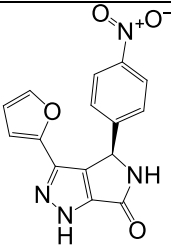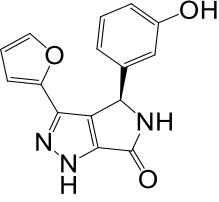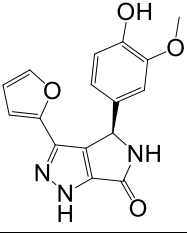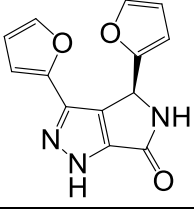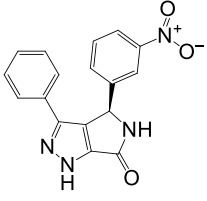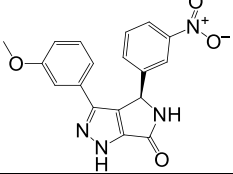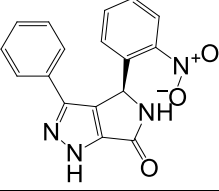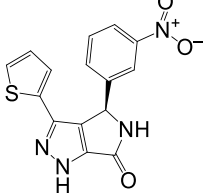
ZINC ID	Structure	Virtual Screen Docking Rank to GPR68 Model
ZINC17946127		1101
ZINC40066704		1364
ZINC23135897		1851
ZINC04929116		1902
ZINC22096188		2343
ZINC32587282		2472
ZINC20729152		2711
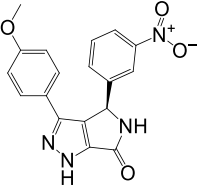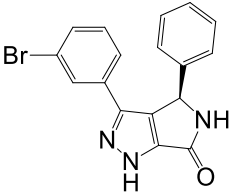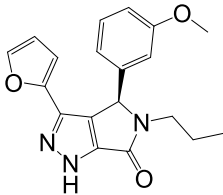
ZINC ID	Structure	Virtual Screen Docking Rank to GPR68 Model
ZINC20826836		3863
ZINC06258265		4127
ZINC40512413		4173
ZINC35478261 (Isx)		NA
ZINC02497882 (Isx)		NA
ZINC02497868 (Isx)		NA
ZINC00425686 (Isx)		NA

Supplementary Table 3. Structures of analogues of ZINC 32587282 and ZINC 04929116 chosen for experimental testing. Vendor information is available from ZINC database (http://zinc.docking.org/). ZINC compounds 32547799 and 67740571 were also synthesized in-house (see **Supplementary methods for chemical synthesis**).

ZINC ID	Structure
ZINC06785875	
ZINC01718514	
ZINC05933520	
ZINC32547799	
ZINC67740571	
ZINC20819785	

ZINC ID	Structure
ZINC20869006	
ZINC20601562	
ZINC32581032	
ZINC32503371	
ZINC32520276	
ZINC32590454	
ZINC00625739	
ZINC02541525	

ZINC ID	Structure
ZINC20855205	
ZINC21367567	
ZINC20855260	
ZINC21367544	
ZINC00429270	
ZINC20213028	
ZINC04928902	
ZINC20213042	

ZINC ID	Structure
ZINC04929065	
ZINC04928929	
ZINC04909980	
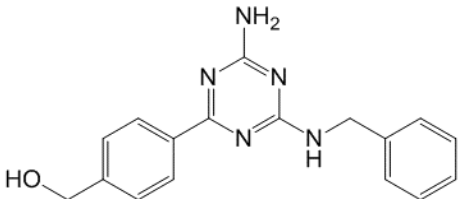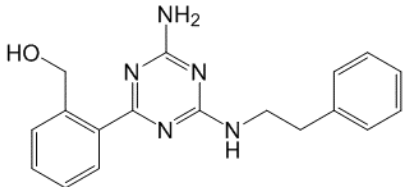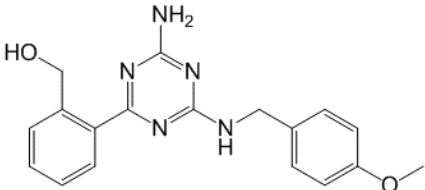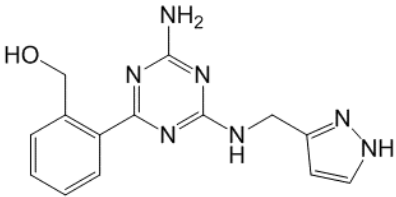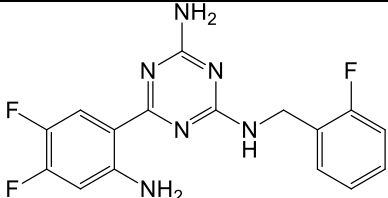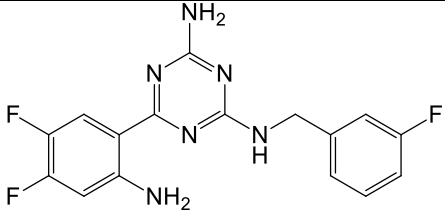
Supplementary Table 4. Allosteric parameters of tested compounds at GPR68 and mutant receptors in this study. HEK293-T cells were transiently transfected and production of cAMP was measured using a split-luciferase reporter assay. Results were analyzed in GraphPad Prism using the standard allosteric operational model (see **Methods** for more detail), with protons as the orthosteric agonist (ligand A) and test compounds as allosteric modulators (ligand B); E_{max} is the maximal possible response of the system; τ_A is the efficacy of protons; α and β are the allosteric cooperativity parameters on agonist affinity and efficacy, respectively; n is the slope parameter. In fitting curves, E_{max} was constrained to the maximal response in fold of basal (basal = 1) for normalized results or the maximal relative luminescence units for un-normalized results; K_A was constrained to be the proton potency in the absence of modulator for that assay; [A] and [B] are the proton concentration and allosteric ligand concentration (column title in molar concentration in Prism), respectively; while all the others (K_B, α, β, τ_A, n) were fitting parameters and set to be globally shared among each family of curves. Most of the curves could be easily fitted. If an assay was not able to be fitted to generate reasonable fitting parameters, the   value was then manually constrained to an initial fitting value and systematically changed with small increments or decrements until the highest affinity (K_B) was reached. Values represent mean ± SEM from N assays, each in triplicate or quadruplicate. Allosteric ligand affinity in the presence of H^+ was defined as "biochemical binding affinity K_{bB}" [$K_{bB} = \frac{K_B}{\alpha}$, pK_{bB} = - log(K_{bB})] to account for the allosteric effect, since K_B alone (defined as a functional binding affinity in the absence of orthosteric agonist H^+ in this case) does not have a physical meaning and is impossible to measure.   Log($\alpha^*\beta$) represents the overall allosteric effect. Values were used in the **Supplementary Figures 11 and 18** for plots.
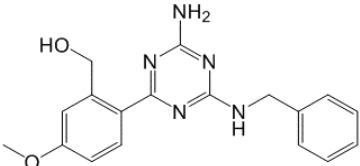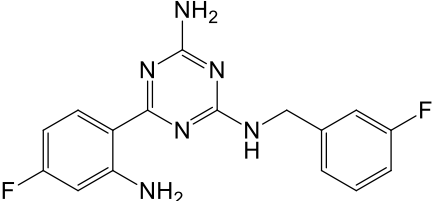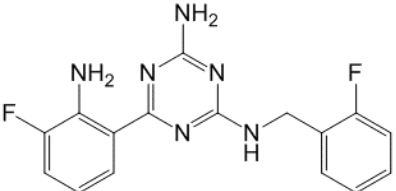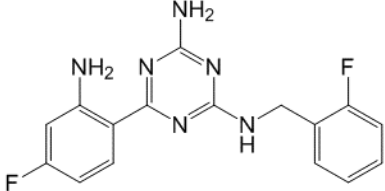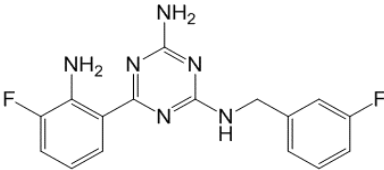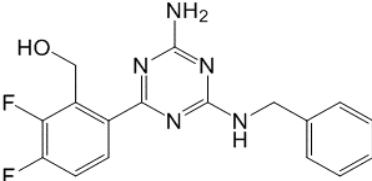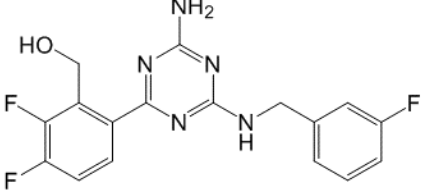
Compounds	τ_A	pK_{bB}	Hill (n)	$Log(\alpha)$	$Log(\beta)$	$Log(\alpha*\beta)$	N
Lorazepam and 10 initial ZINC compounds, Structures in **Supplementary Tables 2** and **3**							
Lorazepam	1.82 ± 0.15	5.58 ± 0.18	3.13 ± 0.37	0.43 ± 0.04	0.06 ± 0.04	0.49 ± 0.04	5
4929116	1.00 ± 0.05	3.13 ± 0.38	2.16 ± 0.07	-1.21 ± 0.24	2.02 ± 0.38	0.81 ± 0.16	10
32587282	1.08 ± 0.08	4.93 ± 0.13	2.30 ± 0.11	0.14 ± 0.16	0.69 ± 0.09	0.83 ± 0.15	11
20855260	1.10 ± 0.09	4.83 ± 0.39	2.18 ± 0.18	-0.70 ± 0.28	0.87 ± 0.28	0.16 ± 0.05	4
4928902	1.23 ± 0.17	4.97 ± 0.31	2.43 ± 0.26	-0.80 ± 0.25	0.78 ± 0.22	-0.03 ± 0.04	5
32547799	1.24 ± 0.04	4.47 ± 0.16	2.39 ± 0.10	-0.39 ± 0.14	0.74 ± 0.10	0.35 ± 0.10	13
67740571 (ogerin)	0.94 ± 0.05	4.78 ± 0.10	1.96 ± 0.06	0.69 ± 0.09	0.75 ± 0.08	1.44 ± 0.09	30
20869006	1.78 ± 0.09	6.13 ± 0.34	3.47 ± 0.47	0.25 ± 0.13	0.14 ± 0.07	0.39 ± 0.08	8
32503371	0.93 ± 0.11	5.39 ± 0.23	2.18 ± 0.17	-0.06 ± 0.11	0.82 ± 0.15	0.76 ± 0.15	8
32520276	0.87 ± 0.06	4.69 ± 0.29	2.24 ± 0.12	-0.18 ± 0.23	1.04 ± 0.24	0.87 ± 0.17	8
32590454	1.11 ± 0.08	5.67 ± 0.28	2.47 ± 0.19	0.32 ± 0.09	0.45 ± 0.10	0.76 ± 0.14	10
Ogerin analogues, Structures in **Supplementary Table 6**							
33548	1.56 ± 0.09	4.21 ± 0.72	2.59 ± 0.13	-0.41 ± 0.36	0.96 ± 0.54	0.55 ± 0.21	4
33549 (**C2**)	0.95 ± 0.04	5.51 ± 0.12	2.04 ± 0.08	1.10 ± 0.12	0.68 ± 0.09	1.78 ± 0.16	17
C3	1.07 ± 0.05	5.69 ± 0.09	2.03 ± 0.07	0.66 ± 0.16	0.51 ± 0.08	1.17 ± 0.19	11
C4	1.31 ± 0.05	5.69 ± 0.08	2.15 ± 0.10	0.75 ± 0.11	0.42 ± 0.06	1.17 ± 0.12	11
33550	1.08 ± 0.11	4.36 ± 0.13	1.86 ± 0.09	1.34 ± 0.11	0.95 ± 0.07	2.29 ± 0.15	5

Compounds	τ_A	pK_{bB}	Hill (n)	Log(α)	Log(β)	Log($\alpha*\beta$)	N
33551	1.70 ± 0.04	3.92 ± 0.25	2.88 ± 0.22	0.27 ± 0.23	0.50 ± 0.13	0.77 ± 0.26	5
33552	1.45 ± 0.14	4.84 ± 0.16	2.81 ± 0.22	-0.02 ± 0.16	0.58 ± 0.14	0.56 ± 0.03	3
33553	1.46 ± 0.14	4.58 ± 0.26	2.84 ± 0.22	0.25 ± 0.28	0.58 ± 0.26	0.83 ± 0.07	3
33554	1.11 ± 0.10	4.20 ± 0.18	1.97 ± 0.11	1.10 ± 0.23	1.13 ± 0.13	2.21 ± 0.22	4
33555	1.42 ± 0.13	3.89 ± 0.31	2.71 ± 0.13	0.24 ± 0.35	0.84 ± 0.20	1.08 ± 0.40	3
33556	1.42 ± 0.14	4.02 ± 0.20	2.63 ± 0.19	0.13 ± 0.28	0.88 ± 0.14	1.02 ± 0.18	3
33557	1.37 ± 0.11	4.19 ± 0.25	2.65 ± 0.21	0.08 ± 0.19	0.72 ± 0.24	0.80 ± 0.06	3
33558	1.32 ± 0.10	3.53 ± 0.37	2.52 ± 0.04	-0.58 ± 0.22	1.37 ± 0.29	0.79 ± 0.25	4
33559	1.35 ± 0.12	3.84 ± 0.06	2.48 ± 0.16	0.28 ± 0.22	0.93 ± 0.09	1.21 ± 0.19	3
33561	1.42 ± 0.08	3.95 ± 0.18	2.42 ± 0.15	0.11 ± 0.32	0.84 ± 0.19	0.92 ± 0.27	5
Three confirmed hits from Tocris Mini Library, Structures in **Supplementary Table 3**							
CGH 2466	1.16 ± 0.06	7.28 ± 0.30	3.83 ± 0.75	-0.13 ± 0.15	0.46 ± 0.14	0.33 ± 0.02	8
Tracazolate	1.11 ± 0.08	5.09 ± 0.09	2.53 ± 0.11	-0.15 ± 0.08	0.58 ± 0.07	0.44 ± 0.02	5
SLV 320	1.15 ± 0.07	5.96 ± 0.07	2.54 ± 0.14	-0.26 ± 0.09	0.59 ± 0.10	0.34 ± 0.02	5
ZINC 67740571 (ogerin) at GPR68 mutant receptors							
Mutant	τ_A	pK_{bB}	Hill (n)	Log(α)	Log(β)	Log($\alpha*\beta$)	N
E160A	0.55 ± 0.07	4.55 ± 0.21	1.38 ± 0.13	0.90 ± 0.18	1.23 ± 0.20	2.13 ± 0.31	7
E160K	0.98 ± 0.08	4.81 ± 0.21	2.02 ± 0.12	0.59 ± 0.13	0.76 ± 0.16	1.35 ± 0.24	9
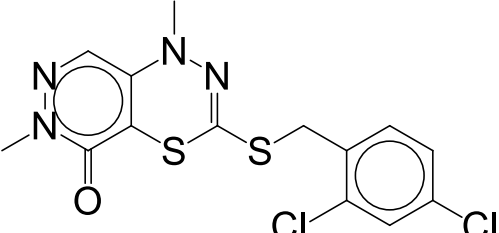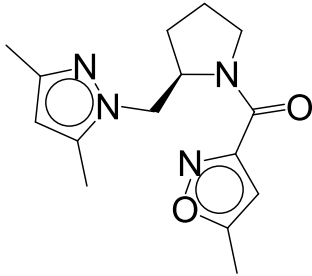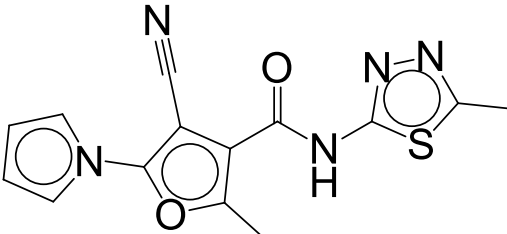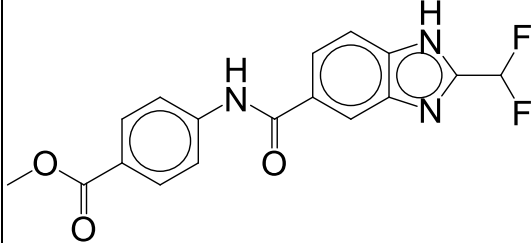
Compounds	τ_A	pK_{bB}	Hill (n)	Log(α)	Log(β)	Log($\alpha*\beta$)	N
E160Q	0.61 ± 0.07	4.87 ± 0.23	1.56 ± 0.11	1.19 ± 0.17	1.14 ± 0.20	2.33 ± 0.34	10
H269F	0.55 ± 0.05	4.00 ± 0.24	1.79 ± 0.06	0.09 ± 0.26	1.55 ± 0.22	1.63 ± 0.16	7
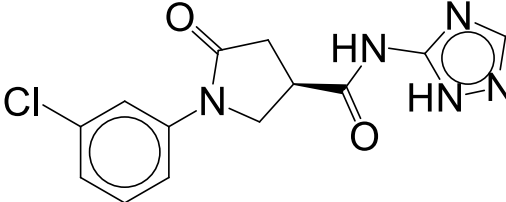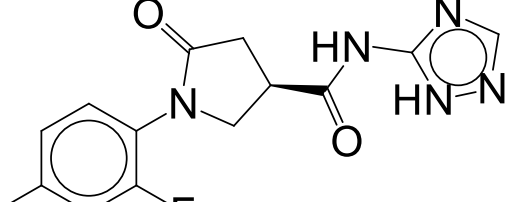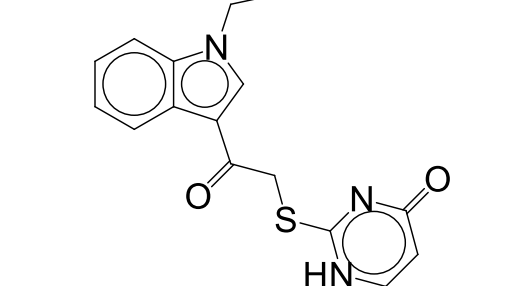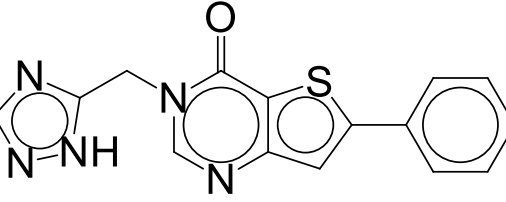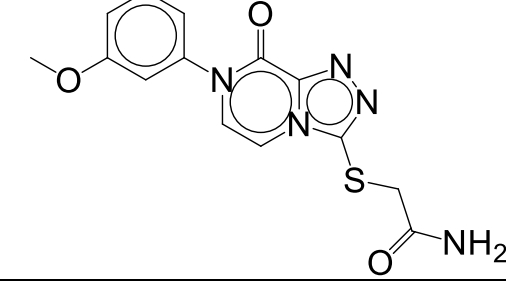
Supplementary Table 5. Structures of ogerin analogs from virtual synthetic library along with their corresponding overall docking ranks. Chemistry synthesis and structural elucidation are in Chemistry supplementary section. Primary screening of these compounds is summarized in **Supplementary Figure 17**.
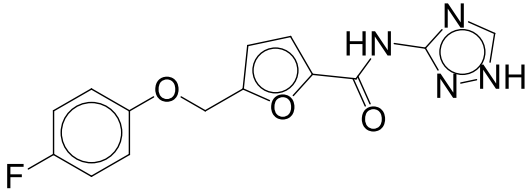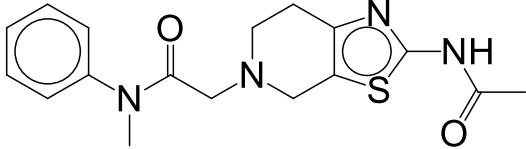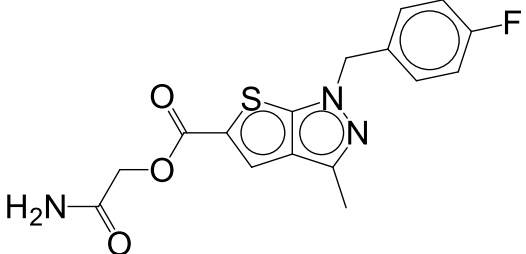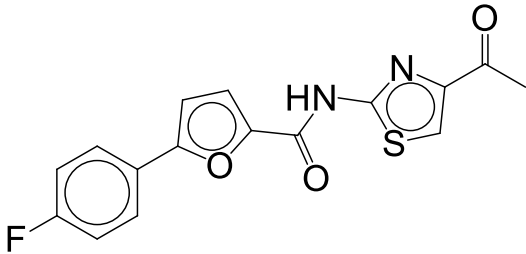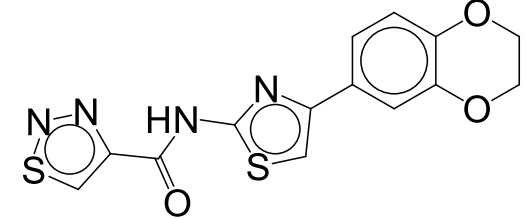
Code name in synthesis procedure	Compounds	Structure	Docking Rank in Virtual Library
19	33548		25
C2 in **Fig 4**	33549 (C2 in **Fig 4**)		7
14	33550		1
16	33551		26
22	33552		370
26	33553		158
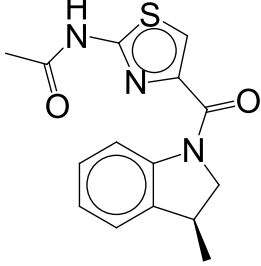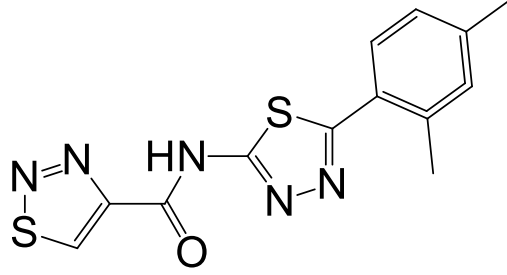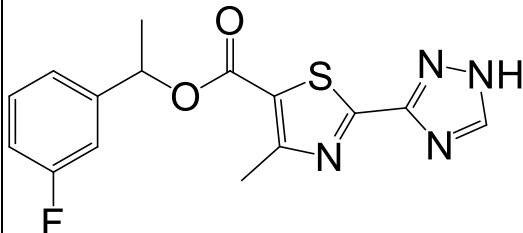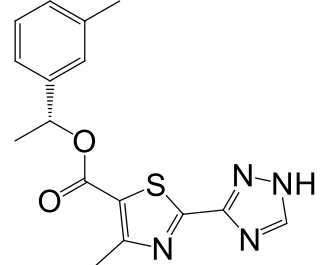
Code name in synthesis procedure	Compounds	Structure	Docking Rank in Virtual Library
18	33554		65
28	33555		31
24	33556		68
23	33557		51
29	33558		78
20	33559		46
27	33561		8

Supplementary Table 6. Structures for 17 ZINC compounds for GPR65 screening.

ZINC ID	Code Number	Structure	Docking Rank in ZINC Lead-like Screen
2143835	BTB09089		
71870037	33718		10695
12801038	33723		14894
65585256	33725		1418

26257911	33730		70
29666959	33732		1591
13692691	33750		3573
32560974	33751		12492
64593683	33753		13415

26262334	33954		268
77141934	33955		14350
7536031	33958		13452
52770635	33959		15727
41613384	33960		15266

60600385	33961		15423
65122687	33963		2370
77476283	33963		14271
77879062	33966		14071

Supplementary Table 7. Anxiety-like behavior on an elevated plus maze and in a marble-burying task, olfactory function in a buried food test, and thermal sensitivity in a hot-plate test.

	Males		Females	
	WT	GPR68 KO	WT	GPR68 KO
Weight at 5 weeks (g)	20 ± 0.3	20 ± 0.7	16 ± 0.3	16 ± 0.2
Elevated plus maze				
Percent open arm time	18 ± 4	22 ± 3	22 ± 3	20 ± 3
Percent open arm entries	27 ± 4	30 ± 2	32 ± 3	32 ± 2
Total number of entries	21 ± 3	25 ± 3	23 ± 2	23 ± 2
Marble-burying test				
Number buried in 30 min	14 ± 1	10 ± 2*	12 ± 1	14 ± 1[#M]
Olfactory test				
Latency to bury food (sec)	294 ± 108	142 ± 32	178 ± 59	238 ± 78
Hotplate test				
Latency to respond (sec)	16 ± 2	12 ± 1	17 ± 2	20 ± 2[#M]

* $p<0.05$, comparison to wild type mice. [#M] $p<0.05$, comparison to male KO mice.

Supplementary Table 8. Latency to find the visible platform. Data shown are time (seconds) in the format of mean ± SEM of four trials.

Day of testing	Males		Females	
	WT	GPR68 KO	WT	GPR68 KO
Day 1	14 ± 2	13 ± 2	13 ± 2	19 ± 3
Day 2	7 ± 1	8 ± 1	8 ± 1	9 ± 2

Supplementary Table 9. Parameters from Supplementary Figure 26.

Receptor	E_{max} (fold of basal)	pEC_{50}	Hill
Proton concentration-responses			
GPR65	19.2 ± 0.5	7.30 ± 0.02	3.88 ± 0.56
D153A	20.7 ± 0.5	6.94 ± 0.02	3.36 ± 0.49
R187L	2.3 ± 0.1	6.55 ± 0.04	2.37 ± 0.45
F242A	7.7 ± 0.2	6.92 ± 0.03	2.43 ± 0.29
Y272A	12.9 ± 0.4	6.62 ± 0.02	2.84 ± 0.37
BTB09089 concentration-responses			
GPR65	7.6 ± 0.3	5.11 ± 0.06	1.34 ± 0.17
D153A	8.1 ± 1.1	5.20 ± 0.25	0.75 ± 0.19
R187L	NA	NA	NA
F242A	NA	NA	NA
Y272A	NA	NA	NA
ZINC 13684400 concentration-responses			
GPR65	2.2 ± 0.1	6.30 ± 0.08	1.08 ± 0.17
D153A	1.5 ± 0.1	5.89 ± 0.35	3.04 ± 0.63
R187L	NA	NA	NA
F242A	NA	NA	NA
Y272A	NA	NA	NA

NA for not applicable.

IIIA. References for Supplementary Results

1. Besnard, J. *et al.* Automated design of ligands to polypharmacological profiles. *Nature* **492,** 215–20 (2012).

2. Keiser, M. J. *et al.* Predicting new molecular targets for known drugs. *Nature* **462,** 175–181 (2009).

3. Huang, X. P. *et al.* Parallel functional activity profiling reveals valvulopathogens are potent 5-hydroxytryptamine(2B) receptor agonists: implications for drug safety assessment. *Mol Pharmacol* **76,** 710–22 (2009).

A.2 Supplementary material for Chapter 2

I. Supplementary Tables

Supplementary Table 1. ZINC IDs, lead-like docking ranks, binding affinities, and selectivity ratios for all compounds tested at HTR2A, DRD2, and HRH1 (NB = nonbinder).

#	ZINC ID	HTR2A rank	DRD2 rank	HRH1 rank	HTR2A K_i (nM)	DRD2 K_i (nM)	DRD2 K_i (nM)	HRH1/HTR2A K_i ratio	HRH1/DRD2 K_i ratio
1	C55405229	211	1808	1915611	7259	>10000	>10000	> 1.4	NB
2	C20207431	500	2025	2740480	1408	>10000	940	0.67	< 0.094
3	C42976658	987	1164	254542	>10000	>10000	1626	< 0.16	< 0.16
4	C63483054	1273	4388	2590373	4590	>10000	>10000	> 2.2	NB
5	C12585531	1814	6033	571928	1784	5698	>10000	> 5.6	> 1.8
6	C19372191	1936	2024	326040	>10000	>10000	0.8	< 0.000080	< 0.000080
7	C63651822	2012	1839	2654099	4476	>10000	2103	0.47	< 0.21
8	C24882708	2354	708	1401070	>10000	>10000	2096	< 0.21	< 0.21
9	C64402001	2898	1527	1399679	3324	3231	19	0.0057	0.0059
10	C00036656	3249	4954	223648	>10000	>10000	6500	< 0.65	< 0.65
11	C20533469	5355	1365	208579	>10000	>10000	>10000	NB	NB
12	C58355688	5430	8314	3004881	59	1821	35	0.59	0.019
13	C48782785	5563	6739	1903962	6176	>10000	6653	1.1	< 0.67
14	C32080342	5671	7537	692414	2500	7881	>10000	> 4.0	> 1.3
15	C12659717	5896	8472	1053134	>10000	5320	356	< 0.036	0.067
16	C19781839	7215	8438	253492	7461	>10000	425	0.057	< 0.042
17	C55253816	7311	2169	1094307	8541	>10000	5920	0.69	< 0.59
18	C12199739	8246	6942	42082	5596	>10000	>10000	> 1.8	NB
19	C63994913	9272	3867	170918	793	5947	>10000	> 12.6	> 1.7
20	C24882377	9444	9674	1006865	>10000	>10000	185	< 0.019	< 0.019
21	C53561601	9893	5772	74924	55	334	1144	21	3.4
22	C63381324	12152	9828	44174	4124	>10000	>10000	> 2.4	NB
23	C19791225	12926	1600	2042077	334	2233	280	0.84	0.13
24	C58138264	14066	988	2139209	>10000	>10000	10000	< 1.0	< 1.0

25	C25409567	15717	1947	2525004	3430	3195	>10000	> 2.9	> 3.1
26	C15076127	18681	868	306823	>10000	>10000	>10000	NB	NB
27	C33028377	19216	1867	842158	>10000	>10000	>10000	NB	NB
28	C57885859	22780	1813	2252846	>10000	8774	>10000	NB	> 1.1
29	C48235265	4717	17156	669739	>10000	8429	>10000	NB	> 1.2
30	C72268948	5092	21229	670369	>10000	>10000	>10000	NB	NB
31	C71880639	5976	17327	646631	3654	5596	2683	0.73	0.48
32	C32853469	7172	16618	1271543	>10000	>10000	>10000	NB	NB
33	C65589455	9300	14449	862013	808	>10000	81	0.1	< 0.0081
34	C70007764	9784	14571	669010	>10000	8404	1706	< 0.17	0.2
35	C65255278	10478	3342	1078327	2268	1996	718	0.32	0.36
36	C69413743	11378	6131	728422	>10000	6442	1271	< 0.13	0.2
37	C69777259	11873	28352	636929	951	>10000	182	0.19	0.018
38	C67817126	13792	66225	840818	4049	4918	2562	0.63	0.52
39	C68048611	15140	12396	1281240	1443	7745	4241	2.9	0.55
40	C29712010	16529	20444	844904	1970	>10000	1003	0.51	< 0.10
41	C19148370	16858	12129	676983	1076	>10000	>10000	> 9.3	NB
42	C63770987	17160	5918	812174	33	1606	44	1.3	0.028
43	C53420081	17493	21596	725269	1561	>10000	271	0.17	< 0.027
44	C32768642	19405	24527	1574526	>10000	>10000	6792	< 0.68	< 0.68
45	C55307836	22540	29218	517824	6580	>10000	4709	0.71	< 0.47
46	C72253360	23594	5396	550131	>10000	>10000	>10000	NB	NB
47	C69896206	26751	9012	603194	3474	6912	1110	0.32	0.16
48	C53420088	28100	13215	1320273	8698	>10000	6540	0.75	< 0.65

Supplementary Table 2. ZINC IDs, lead-like docking ranks, binding affinities, and selectivity ratios for all compounds tested at MOR and KOR (NB = nonbinder).

#	ZINC ID	Kappa Rank	Mu Rank	KOR Ki (µM)	MOR Ki (µM)	MOR/KOR ratio
1	C68408255	263	86198	14.4	3.04	0.211
2	C44913867	1166	107428	2.93	10.2	3.48
3	C68387316	1548	119964	4.67	> 100	> 21.4
4	C68570285	3377	284858	> 100	> 100	NB
5	C68255712	3731	291371	> 100	> 100	NB
6	C43714067	4524	201440	1.81 (EC50 = 17.0)	0.478 (EC50 = 5.21)	0.264
7	C53629583	5101	789761	> 100	> 100	NB
8	C45076626	5464	520993	> 100	1.29	< 0.0129
9	C55359151	5615	292712	>10000	>10000	NB
10	C69181132	6141	255399	10.5	2.47	0.235
11	C69067765	7542	543545	> 100	8.69	0.0869
12	C67983399	7727	502775	> 100	> 100	NB
13	C67847468	7995	346294	3.14	6.92	2.2
14	C47872495	9734	492977	2.69	2.43 (EC50 = 12.3)	0.903
15	C69597222	12266	244121	> 100	> 100	NB
16	C58265334	13870	663142	> 100	> 100	NB
17	C66939732	17057	59182	> 100	45.7	< 0.457
18	C69532007	21290	99087	> 100	> 100	NB
19	C25423862	22855	394837	> 100	> 100	NB
20	C48345561	27360	793832	11.1	5.7	0.513
21	C54119195	28396	71079	> 100	> 100	NB
22	C52469366	28472	248208	5.56	> 100	> 17.9

A.3 Supplementary material for Chapter 3

I. Supplementary Tables

Supplementary Table 1. P-values and log-fold changes for differential expression conditions between GABA$_A$ subunits/GPR68 and adenosine GPCRs and GPR68. P-values in green are <0.05. Log-fold changes are color-scaled according to value, with negative values colored red and positive values colored green.

P-values for differential expression between GABA$_A$ subunits and GPR68							
Differential Expression Condition	**Gene Name**						
	GABRA1	**GABRA2**	**GABRA3**	**GABRA5**	**GABRG2**	**GABRG3**	**GPR68**
idiopathic dilated vs normal : GSE1145	0.087163	0.002228	0.000597	0.001182	0.19145	0.05181	0.000152
ischemic vs normal : GSE1145	0.087163	0.002228	0.000597	0.001182	0.19145	0.05181	0.000152
Huntington Disease vs Control_group : GSE3790	0.116927	1.29E-08	0.145349	0.647019	0.999758	0.017399	0.006973
Hippocampus vs Entorhinal cortex : GSE5281	4.43E-05	0.379327	0.008771	3.38E-06	0.007745	0.008384	8.19E-06
Middle temporal gyrus vs Entorhinal cortex : GSE5281	4.43E-05	0.379327	0.008771	3.38E-06	0.007745	0.008384	8.19E-06
Posterior cingulate gyrus vs Entorhinal	4.43E-05	0.379327	0.008771	3.38E-06	0.007745	0.008384	8.19E-06

Differential Expression Condition	Gene Name						
	GABRA1	GABRA2	GABRA3	GABRA5	GABRG2	GABRG3	GPR68
cortex : GSE5281							
Primary visual cortex vs Entorhinal cortex : GSE5281	4.43E-05	0.379327	0.008771	3.38E-06	0.007745	0.008384	8.19E-06
Superior frontal gyrus vs Entorhinal cortex : GSE5281	4.43E-05	0.379327	0.008771	3.38E-06	0.007745	0.008384	8.19E-06
Adenocarcinoma vs Normal_control _group : bhattacharjee-lung	6.67E-05	0.497762	0.232753	0.148429	0.024915	0.151555	0.004974
Small cell carcinoma of lung vs Normal_control _group : bhattacharjee-lung	6.67E-05	0.497762	0.232753	0.148429	0.024915	0.151555	0.004974
Squamous Cell Carcinoma of Lung vs Normal_control _group : bhattacharjee-lung	6.67E-05	0.497762	0.232753	0.148429	0.024915	0.151555	0.004974
pulmonary carcinoids vs Normal_control _group : bhattacharjee-lung	6.67E-05	0.497762	0.232753	0.148429	0.024915	0.151555	0.004974

P-values for differential expression between GABA$_A$ subunits and GPR68

P-values for differential expression between GABA$_A$ subunits and GPR68							
Differential Expression Condition	**Gene Name**						
	GABRA1	**GABRA2**	**GABRA3**	**GABRA5**	**GABRG2**	**GABRG3**	**GPR68**
Healthy Cerebellum vs Medulloblastoma : pomeroy-embryonal	4.45E-05	1	0.00379	1	1	NA	0.022224
Malignant Glioma vs Medulloblastoma : pomeroy-embryonal	4.45E-05	1	0.00379	1	1	NA	0.022224
Primitive neuroectodermal tumour vs Medulloblastoma : pomeroy-embryonal	4.45E-05	1	0.00379	1	1	NA	0.022224
atypical teratoid/rhabdoid tumours vs Medulloblastoma : pomeroy-embryonal	4.45E-05	1	0.00379	1	1	NA	0.022224
B cell vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Blood and Bone Marrow vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Breast vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
CNS vs Urinary bladder :	1.00E-10	1	1	1	1.46E-06	1	1.00E-10

P-values for differential expression between GABA$_A$ subunits and GPR68							
Differential Expression Condition	**Gene Name**						
	GABRA1	**GABRA2**	**GABRA3**	**GABRA5**	**GABRG2**	**GABRG3**	**GPR68**
ramaswamy-cancer							
Cerebellum vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Colon vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Follicular Stem cells vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Kidney vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Lung vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Melanocytes vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Mesothelium vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Ovary vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10

P-values for differential expression between GABA_A subunits and GPR68							
Differential Expression Condition	**Gene Name**						
	GABRA1	**GABRA2**	**GABRA3**	**GABRA5**	**GABRG2**	**GABRG3**	**GPR68**
Pancreas vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Peripheral blood mononuclear cell vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Prostate vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Uterus vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
WBC vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
Whole Brain vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10
germinal centre vs Urinary bladder : ramaswamy-cancer	1.00E-10	1	1	1	1.46E-06	1	1.00E-10

Log-fold change for differential expression between GABA$_A$ and GPR68							
Differential Expression Condition	**Gene Name**						
	GABRA1	**GABRA2**	**GABRA3**	**GABRA5**	**GABRG2**	**GABRG3**	**GPR68**
idiopathic dilated vs normal : GSE1145	0	0.711448	1.024712	0.870018	0	0	0.785775
ischemic vs normal : GSE1145	0	-0.06273	-0.01333	-0.61616	0	0	0.221
Huntington Disease vs Control_group : GSE3790	0	-0.67874	0	0	0	0.043676	0.097032
Hippocampus vs Entorhinal cortex : GSE5281	-0.68014	0	-2.63103	0.89927	0	-1.53433	-0.58386
Middle temporal gyrus vs Entorhinal cortex : GSE5281	1.413633	0	-1.89001	-0.30736	0	-0.41301	-0.42781
Posterior cingulate gyrus vs Entorhinal cortex : GSE5281	1.177604	0	-1.13868	-0.61935	0	-1.28145	-0.92366
Primary visual cortex vs Entorhinal cortex : GSE5281	0.924649	0	-2.26044	-2.79893	0	-2.13072	-1.73786
Superior frontal gyrus vs Entorhinal cortex : GSE5281	1.925006	0	-1.08579	0.263431	0	-0.04324	0.135823

Log-fold change for differential expression between GABA_A and GPR68							
Differential Expression Condition	**Gene Name**						
	GABRA1	**GABRA2**	**GABRA3**	**GABRA5**	**GABRG2**	**GABRG3**	**GPR68**
Adenocarcinoma vs Normal_control_group : bhattacharjee-lung	1.629798	0	0	0	0	0	0.747755
Small cell carcinoma of lung vs Normal_control_group : bhattacharjee-lung	1.704565	0	0	0	0	0	1.447889
Squamous Cell Carcinoma of Lung vs Normal_control_group : bhattacharjee-lung	0.038047	0	0	0	0	0	0.264071
pulmonary carcinoids vs Normal_control_group : bhattacharjee-lung	2.56347	0	0	0	0	0	1.782032
Healthy Cerebellum vs Medulloblastoma : pomeroy-embryonal	2.578442	0	-0.367	0	2.530974	NA	1.344712
Malignant Glioma vs Medulloblastoma : pomeroy-embryonal	0.633482	0	0.443256	0	2.913316	NA	1.10538

Log-fold change for differential expression between GABA$_A$ and GPR68							
Differential Expression Condition	Gene Name						
	GABRA1	GABRA2	GABRA3	GABRA5	GABRG2	GABRG3	GPR68
Primitive neuroectodermal tumour vs Medulloblastoma : pomeroy-embryonal	0.143268	0	0.496279	0	-1.1927	NA	-0.23125
atypical teratoid/rhabdoid tumours vs Medulloblastoma : pomeroy-embryonal	-0.26164	0	-2.13552	0	-0.239	NA	-0.00923
B cell vs Urinary bladder : ramaswamy-cancer	-0.54804	0	0	0	-0.42818	0	0.57006
Blood and Bone Marrow vs Urinary bladder : ramaswamy-cancer	1.12951	0	0	0	-0.09624	0	3.092719
Breast vs Urinary bladder : ramaswamy-cancer	-0.52175	0	0	0	-1.3244	0	0.334363
CNS vs Urinary bladder : ramaswamy-cancer	-0.35419	0	0	0	0.594359	0	0.797556
Cerebellum vs Urinary bladder : ramaswamy-cancer	2.330058	0	0	0	-1.1671	0	1.933146
Colon vs Urinary bladder : ramaswamy-cancer	-0.23763	0	0	0	-0.82194	0	0.20893

Log-fold change for differential expression between GABA$_A$ and GPR68							
Differential Expression Condition	Gene Name						
	GABRA1	GABRA2	GABRA3	GABRA5	GABRG2	GABRG3	GPR68
Follicular Stem cells vs Urinary bladder : ramaswamy-cancer	-0.57697	0	0	0	0.559337	0	-0.66288
Kidney vs Urinary bladder : ramaswamy-cancer	0.441873	0	0	0	0.511023	0	0.761494
Lung vs Urinary bladder : ramaswamy-cancer	0.111283	0	0	0	-0.06234	0	-0.13902
Melanocytes vs Urinary bladder : ramaswamy-cancer	-1.75393	0	0	0	1.030804	0	-0.61881
Mesothelium vs Urinary bladder : ramaswamy-cancer	-0.72297	0	0	0	-0.20448	0	-0.10944
Ovary vs Urinary bladder : ramaswamy-cancer	0.505625	0	0	0	0.289682	0	0.699752
Pancreas vs Urinary bladder : ramaswamy-cancer	-0.28573	0	0	0	0	0	-0.29397
Peripheral blood mononuclear cell vs Urinary bladder : ramaswamy-cancer	0.960079	0	0	0	0	0	1.62434

Log-fold change for differential expression between GABA_A and GPR68							
Differential Expression Condition	**Gene Name**						
	GABRA1	**GABRA2**	**GABRA3**	**GABRA5**	**GABRG2**	**GABRG3**	**GPR68**
Prostate vs Urinary bladder : ramaswamy-cancer	-0.50497	0	0	0	0	0	0.254173
Uterus vs Urinary bladder : ramaswamy-cancer	-0.01371	0	0	0	0	0	0.400266
WBC vs Urinary bladder : ramaswamy-cancer	0.990094	0	0	0	0	0	2.508593
Whole Brain vs Urinary bladder : ramaswamy-cancer	1.893336	0	0	0	0	0	0.847733
germinal centre vs Urinary bladder : ramaswamy-cancer	-0.91202	0	0	0	0	0	-1.08849

P-values for differential expression between adenosine receptors and GPR68					
	Gene Name				
Differential Expression Condition	ADORA1	ADORA2A	ADORA2B	ADORA3	GPR68
idiopathic dilated vs normal : GSE1145	0.00202	0.000318	0.189357	0.130107	0.000152
ischemic vs normal : GSE1145	0.00202	0.000318	0.189357	0.130107	0.000152
Huntington Disease vs Control_group : GSE3790	1.02E-09	0.697668	0.037732	0.000527	0.006973
Hippocampus vs Entorhinal cortex : GSE5281	0.013006	0.348929	0.047386	0.050767	8.19E-06
Middle temporal gyrus vs Entorhinal cortex : GSE5281	0.013006	0.348929	0.047386	0.050767	8.19E-06
Posterior cingulate gyrus vs Entorhinal cortex : GSE5281	0.013006	0.348929	0.047386	0.050767	8.19E-06
Primary visual cortex vs Entorhinal cortex : GSE5281	0.013006	0.348929	0.047386	0.050767	8.19E-06
Superior frontal gyrus vs Entorhinal cortex : GSE5281	0.013006	0.348929	0.047386	0.050767	8.19E-06
Adenocarcinoma vs Normal_control_group : bhattacharjee-lung	0.997458	2.65E-06	0.006357	0.021156	0.004974
Small cell carcinoma of lung vs Normal_control_group : bhattacharjee-lung	0.997458	2.65E-06	0.006357	0.021156	0.004974
Squamous Cell Carcinoma of Lung vs Normal_control_group : bhattacharjee-lung	0.997458	2.65E-06	0.006357	0.021156	0.004974

P-values for differential expression between adenosine receptors and GPR68					
	Gene Name				
Differential Expression Condition	ADORA1	ADORA2A	ADORA2B	ADORA3	GPR68
pulmonary carcinoids vs Normal_control_group : bhattacharjee-lung	0.997458	2.65E-06	0.006357	0.021156	0.004974
Healthy Cerebellum vs Medulloblastoma : pomeroy-embryonal	3.56E-05	0.010874	0.199435	0.004588	0.022224
Malignant Glioma vs Medulloblastoma : pomeroy-embryonal	3.56E-05	0.010874	0.199435	0.004588	0.022224
Primitive neuroectodermal tumour vs Medulloblastoma : pomeroy-embryonal	3.56E-05	0.010874	0.199435	0.004588	0.022224
atypical teratoid/rhabdoid tumours vs Medulloblastoma : pomeroy-embryonal	3.56E-05	0.010874	0.199435	0.004588	0.022224
B cell vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Blood and Bone Marrow vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Breast vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
CNS vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Cerebellum vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Colon vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10

Differential Expression Condition	ADORA1	ADORA2A	ADORA2B	ADORA3	GPR68
		Gene Name			
Follicular Stem cells vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Kidney vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Lung vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Melanocytes vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Mesothelium vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Ovary vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Pancreas vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Peripheral blood mononuclear cell vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Prostate vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
Uterus vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10
WBC vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.000134	1.00E-10

Table title: P-values for differential expression between adenosine receptors and GPR68

P-values for differential expression between adenosine receptors and GPR68					
Differential Expression Condition		**Gene Name**			
	ADORA1	**ADORA2 A**	**ADORA2 B**	**ADORA3**	**GPR68**
Whole Brain vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.00013 4	1.00E-10
germinal centre vs Urinary bladder : ramaswamy-cancer	1	1.00E-10	1	0.00013 4	1.00E-10

Log-fold change for differential expression between adenosine receptors and GPR68					
	Gene Name				
Differential Expression Condition	**ADORA1**	**ADORA2A**	**ADORA2B**	**ADORA3**	**GPR68**
idiopathic dilated vs normal : GSE1145	0.644785	-0.36998	0	0	0.785775
ischemic vs normal : GSE1145	-0.03592	-0.75743	0	0	0.221
Huntington Disease vs Control_group : GSE3790	0.52575	0	-0.22887	-0.59627	0.097032
Hippocampus vs Entorhinal cortex : GSE5281	-0.81104	0	0.509049	0	-0.58386
Middle temporal gyrus vs Entorhinal cortex : GSE5281	0.210996	0	1.195929	0	-0.42781
Posterior cingulate gyrus vs Entorhinal cortex : GSE5281	-0.68388	0	0.548817	0	-0.92366
Primary visual cortex vs Entorhinal cortex : GSE5281	-1.1151	0	0.377712	0	-1.73786
Superior frontal gyrus vs Entorhinal cortex : GSE5281	-0.38403	0	2.16983	0	0.135823
Adenocarcinoma vs Normal_control_group : bhattacharjee-lung	0	-0.17221	1.361583	0.195044	0.747755
Small cell carcinoma of lung vs Normal_control_group : bhattacharjee-lung	0	0.848598	0.402014	0.05606	1.447889
Squamous Cell Carcinoma of Lung vs Normal_control_group : bhattacharjee-lung	0	-0.48692	-0.26256	-2.32077	0.264071

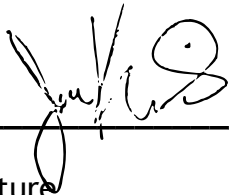| Log-fold change for differential expression between adenosine receptors and GPR68					
	Gene Name				
Differential Expression Condition	**ADORA1**	**ADORA2 A**	**ADORA2 B**	**ADORA3**	**GPR68**
pulmonary carcinoids vs Normal_control_group : bhattacharjee-lung	0	1.58534 5	1.55651 1	0.84826 8	1.782032
Healthy Cerebellum vs Medulloblastoma : pomeroy-embryonal	2.16113 7	1.22664 4	0	1.72871	1.344712
Malignant Glioma vs Medulloblastoma : pomeroy-embryonal	2.00799 9	0.81113 8	0	1.70940 4	1.10538
Primitive neuroectodermal tumour vs Medulloblastoma : pomeroy-embryonal	- 0.07775	- 0.10649	0	0.12529 8	-0.23125
atypical teratoid/rhabdoid tumours vs Medulloblastoma : pomeroy-embryonal	- 0.58335	- 0.36301	0	0.52543 6	-0.00923
B cell vs Urinary bladder : ramaswamy-cancer	0	0.93789 9	0	- 0.85639	0.57006
Blood and Bone Marrow vs Urinary bladder : ramaswamy-cancer	0	2.15088 1	0	1.24225 9	3.092719
Breast vs Urinary bladder : ramaswamy-cancer	0	- 0.06422	0	0.19593 4	0.334363
CNS vs Urinary bladder : ramaswamy-cancer	0	0.19760 5	0	- 0.23299	0.797556
Cerebellum vs Urinary bladder : ramaswamy-cancer	0	1.51371 4	0	1.64820 1	1.933146
Colon vs Urinary bladder : ramaswamy-cancer	0	0.20947 2	0	0.02596	0.20893

Log-fold change for differential expression between adenosine receptors and GPR68					
	Gene Name				
Differential Expression Condition	**ADORA1**	**ADORA2 A**	**ADORA2 B**	**ADORA3**	**GPR68**
Follicular Stem cells vs Urinary bladder : ramaswamy-cancer	0	0.397927	0	-0.18259	-0.66288
Kidney vs Urinary bladder : ramaswamy-cancer	0	0.762487	0	0.580287	0.761494
Lung vs Urinary bladder : ramaswamy-cancer	0	0.379617	0	0.05405	-0.13902
Melanocytes vs Urinary bladder : ramaswamy-cancer	0	-0.83465	0	-0.73739	-0.61881
Mesothelium vs Urinary bladder : ramaswamy-cancer	0	-0.6534	0	0.186126	-0.10944
Ovary vs Urinary bladder : ramaswamy-cancer	0	1.152418	0	0.526519	0.699752
Pancreas vs Urinary bladder : ramaswamy-cancer	0	-0.36398	0	-0.60585	-0.29397
Peripheral blood mononuclear cell vs Urinary bladder : ramaswamy-cancer	0	2.558101	0	1.196608	1.62434
Prostate vs Urinary bladder : ramaswamy-cancer	0	-0.12072	0	0.100777	0.254173
Uterus vs Urinary bladder : ramaswamy-cancer	0	0.055107	0	-0.10414	0.400266
WBC vs Urinary bladder : ramaswamy-cancer	0	1.616454	0	0.789061	2.508593

Log-fold change for differential expression between adenosine receptors and GPR68					
	Gene Name				
Differential Expression Condition	ADORA1	ADORA2A	ADORA2B	ADORA3	GPR68
Whole Brain vs Urinary bladder : ramaswamy-cancer	0	1.229631	0	-0.06392	0.847733
germinal centre vs Urinary bladder : ramaswamy-cancer	0	0.473233	0	-1.08899	-1.08849

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

_____   <u>3/26/2015</u>

Author Signature   Date