HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2013 August 01.

Published in final edited form as:

Nat Methods. 2013 February ; 10(2): 140–146. doi:10.1038/nmeth.2324.

A Pharmacological Organization of G Protein-coupled Receptors

Henry Lin[1,3], **Maria F. Sassano**[2,3], **Bryan L. Roth**[2,*], and **Brian K. Shoichet**[1,*]

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

[2]Department of Pharmacology, University of North Carolina Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA

Abstract

Protein classification typically uses structural, sequence, or functional similarity. Here we introduce an orthogonal method that organizes proteins by ligand similarity, focusing here on the class A G protein-coupled receptor (GPCR) protein family. Comparing a ligand-based dendogram to a sequence-based one, we sought examples of GPCRs that were distantly linked by sequence but neighbors by ligand similarity. Experimental testing of compounds predicted to link three of these new pairs confirmed the predicted association, with potencies ranging from the low-nanomolar to low-micromolar. We then identified hundreds of *non*-GPCRs closely related to GPCRs by ligand similarity, including the CXCR2 chemokine receptor to Casein kinase I, the cannabinoid receptors to epoxide hydrolase 2, and the α_2 adrenergic receptor to phospholipase D. These, too, were confirmed experimentally. Ligand similarities among these targets may reflect a chemical integration in the time domain of molecular signaling.

Introduction

Since the molecular biology revolution, proteins have been related to each other bioinformatically by either sequence or structural similarities.[1,2] When we seek to understand the ligand recognition of a protein or the specificity of a drug or a reagent, we typically consider those proteins that are related structurally, functionally[3], or by sequence[4]. Correspondingly, methods and databases of protein families such as Pfam[5] and TRIBE-MCL[6] rely on multiple sequence alignments and machine learning to classify protein families.

*Corresponding Authors: Brian Shoichet: shoichet@cgl.ucsf.edu, phone: 415-514-4126, fax: 415-514-4126. Bryan Roth: bryan_roth@med.unc.edu, phone: 919-966-7535, fax 919-843-5788.

[3]These authors contributed equally to this work.

Author Contributions

H.L performed the calculations. M.F.S performed experiments. B.L.R. reviewed experimental observations. H.L. and B.K.S drafted the manuscript. M.F.S and B.L.R extensively edited the manuscript. B.K.S. and B.L.R. were provoked by editorial comments to explore the chemo-evolutionary constraints and time-domain signaling implied by this work.

Competing Financial Interest

BKS declares a competing financial interest: he is the founder of SeaChange Pharmaceuticals, which uses chemoinformatics for target prediction. All other authors declare no competing interests.

Ligand recognition does not always respect such molecular biology metrics. For instance, acetylcholine and serotonin signal both through G protein-coupled receptors (GPCRs) and ion channels, which are unrelated by sequence or structure. Both ligands are also recognized by specific transporters, which are, in turn, are unrelated to GPCRs and ion channels. In addition, drugs like alosetron, which target the ionotropic serotonin receptors (HTR3), also modulate the metabotropic serotonin receptors (e.g. HTR2B, HTR4),[7,8] while serotonergic GPCR-targeting drugs also modulate the serotonin transporter.[9] Ligands that modulate bile acid nuclear hormone receptor (NR1H4) also modulate the G protein-coupled bile acid receptor (GPBAR1)[10]. Inhibitors of enzymes, from reverse transcriptases to kinases to proteases, can also modulate GPCRs and nuclear hormone receptors.[9,11–13]

We thus wondered how a quantitative ligand-based organization of pharmacological targets might differ from the more familiar sequence- and structure-based approaches. It is easy, after all, to build a “just-so” story with a few selected cases, such as acetylcholine and serotonin, but to understand whether a ligand-based relationship among targets will substantially differ from a sequence-based one, the two schemes must be compared globally and quantitatively. Since sequence and structure comparisons are restricted to targets within a single, evolutionarily related target family, we will focus our attention on class A (rhodopsin-like) GPCRs. These targets are recommended by their abundance--about 700 genes in the human genome[14]— and the substantial number that have annotated ligands.

Here we ask the following questions: how different is a sequence-based organization of the class A GPCRs from one based on ligand similarity? Do the differences explain non-obvious aspects of target pharmacology and drug discovery? Can we use the ligand-based organization *prospectively*, to predict and test new associations among previously unrelated targets? Whereas we and others[15–17] have used ligand-based metrics to predict the activities of individual drugs against off-targets, this is, to our knowledge, the first effort to compare pharmacological relationships across an entire family of targets. The associations that emerge are startling: some GPCRs that are distant by sequence identity become neighbors by ligand similarity, while others that are neighbors by sequence are pushed far apart by the dissimilarity of their ligand sets. The ligand-based target similarities also suggest new associations among receptors that are, for the first time, predicted and demonstrated to share ligands. Because these associations are based on ligand similarities, they may be expanded to explore the polypharmacology between GPCRs and *non*-GPCRs, which are wholly unrelated by sequence and structure. An emergent property of these associations is that they recapitulate the activities of the cognate primary messengers, which also cross major target boundaries. This may reflect relationships in the time domain of molecular signaling, where ligand chemistry, not receptor sequence, is conserved.

Results

There were 146 class A GPCRs with at least six ligands in the ChEMBL database, which annotates ligands to targets based on literature reports.[18] On average each GPCR had 608 ligands, with a median of 380. Whereas this list captures a minority of the roughly 700 class A GPCR members[14], all of the major sub-families are included, such as the biogenic amine

receptors, the peptide-receptors, the lipid-activated GPCRs, and receptors responding to protein ligands.

These GPCRs were organized by sequence and by ligand similarity. To focus on the part of the sequence most implicated in ligand binding, we only used those residues previously mapped to one of 43 orthosteric sites.[19] Sequence distances between any pair of targets was measured using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm,[20] and rendered using FigTree (http://tree.bio.ed.ac.uk/software/figtree/) (Fig. 1a); notwithstanding the focus on sequence identity in the binding site, the relationships that emerge resemble those based on dendograms using full receptor sequence identity (e.g., http://gpcr.scripps.edu/).

To associate receptors by ligand similarity, ligands were represented by topological fingerprints, which are bit strings that reflect the presence or absence of chemotypes and their chemical environment in the ligand. The similarity of these fingerprints was compared for all pairs of molecules in each ligand-set for each pair of receptors, and the overall similarity of these sets was compared to an expected random similarity using machinery drawn from the sequence algorithm Basic Local Alignment Search Tool (BLAST). Similarly, an expectation value (E-value) can be calculated, using the Similarity Ensemble Approach (Methods).[11,21] Ligand-based dendrograms, too, were calculated using FigTree, with the distance between pairs of GPCRs quantified by the cosine angle of their SEA E-values (Fig 1b).

In the sequence-based dendrogram, the relationships among the GPCRs are as expected. The biogenic amine receptors, including the adrenergic, dopaminergic and serotonergic GPCRs largely cluster together, as do peptidic receptors such as the chemokines and melanocortins, as do the lipid responding GPCRs. At a finer resolution, some peculiar divergences and associations begin to appear. For instance, the cysteinyl leukotriene and leukotriene B4 receptors are separated from not only the other lipid-recognizing GPCRs, but also from each other, even though they are a part of the same 5-lipoxygenase pathway involved in airway inflammation.

Compared to the sequence-based organization, the ligand-based dendrogram seems victimized by almost grotesque rearrangements. The muscarinic receptors shift away from the other biogenic amine GPCRs and toward the chemokine receptors, with which they share very little orthosteric site sequence identity (9–21% identity across all subtypes). Equally perplexing, the β-adrenergic receptors separate from the α-adrenergic receptors and indeed other biogenic amine GPCRs, moving closer to the cannabinoid lipid receptors and melatonin receptors. Other rearrangements, though covering just as much distance, seem easier to reconcile with the biology they control. Thus, the cysteinyl leukotriene and leukotriene B4 receptors move much closer to each other than they were by sequence, and now cluster with other lipid GPCRs, consistent with their roles in the same leukotriene inflammatory pathway.

Though these rearrangements seem superficially perplexing, their basis may be grasped by comparing the ligands that bind to these targets. Many GPCRs that are dissimilar by

orthosteric site sequence bind similar ligands, to the point where exactly the same ligands are sometimes shared between them (Supplementary Table 1). For example, the opioid and somatostatin receptors shift closer to the biogenic amine receptors. Despite their sequence differences, these peptidic receptors often bind aminergic molecules. The SSTR5 somatostatin 5 receptor and the HRH1 histamine H1 receptor, for instance, share only 33% sequence identity in their binding sites, even though their ligand sets resemble one another (E-value of 9.9×10^{-8}; Supplementary Table 1). Indeed, the two receptors are modulated by several identical ligands[22] (Supplementary Table 1).

Conversely, some receptors, like the muscarinic, the β-adrenergic, and the chemokine families, separate from apparently cognate GPCRs. Based on ligand similarity, the muscarinic receptors move closer to peptidic GPCRs, such as neuropeptide Y and chemokine receptors, and to lipid GPCRs, like sphingosine phosphate and prostaglandin receptors, and *away* from the biogenic amine receptors. Thus, whereas the CHRM1 muscarinic acetylcholine receptor M1 and the MCHR1 Melanin-containing hormone receptor 1 share only 26% sequence identity in the binding site, their SEA E-value is 8.3×10^{-7} and they share several sub-micromolar ligands[23] (Supplementary Table 1). Meanwhile the muscarinics share few ligands, and little ligand-set similarity, with most bioaminergic receptors. The separation of the β- and α-adrenergics is explained by the divergence of their ligand sets. The two classes of receptors share adrenaline and noradrenaline as primary messengers, and have sequence identities ranging from 49% to 63%, but once past the small catecholamines their ligands diverge: the β adrenergic ligands largely resemble isoproterenol, while the α adrenergic antagonists vary widely, often characterized by larger compounds with disparate scaffolds. Meanwhile, the chemokine receptors, which form an essentially contiguous family by sequence, are split into two groups by ligand similarity. One group, characterized by CXCR4, CCR1, CCR2, and CCR5, move closer to the biogenic amine receptors, while CCR3, CCR8 and CXCR3 move closer to the muscarinics and the neuropeptide Y receptors. For instance, though CCR5 and the CHRM2 muscarinic acetylcholine receptor M2 share only 16% sequence identity in the binding site, they share over 30 antagonists in several different ligand series (Supplementary Table 1).

Emboldened by these observations, we asked if the new associations *predict* crosstalk between targets not formerly known to share ligands. Many of the new neighbors in the ligand-based dendrogram share not even a single ligand, neither in ChEMBL nor in the literature, but nevertheless are highly related by the SEA E-values of their ligand lists. One such was the link between the OPRK κ opioid receptor and the HTR2B 5-HT_{2B} serotonin receptor ligands, which resemble each other with a SEA E-value of 9.9×10^{-8} though their sites share only 28% sequence identity. A SEA-screen of the ZINC database[24] suggested that compound **1** was similar to both the OPRK and HTR2B ligands. Upon *in vitro* testing, compound **1** had a K_i of 0.9 μM to HTR2B and 1.0 μM to OPRK (Fig. 2, Table 1). We note that after these experiments were concluded, another series of compounds were found by some of us, in an unrelated project, that also inhibited both targets. The chemical series that did so is unrelated to that described here[25].

If there have been many previous examples of ligand crosstalk between peptide and bioamine GPCRs, there are many fewer between peptide- and lipid-recognizing GPCRs. We were therefore interested to observe an association between the NPY5R neuropeptide Y receptor 5 and CNR2 cannabinoid receptor 2. Whereas their binding sites share only 7% identity, they had a SEA E-value of 1.1×10^{-9}. A particular CNR2 agonist, compound **2**, resembled NPY5R ligands and was commercially available (Table 1). Compound **2** was found bind to NPY5R with an IC_{50} of 190 nM (K_i = 8.5 nM), similar to its CNR2 potency (EC_{50} = 140 nM)[26]. NPY5R was also linked to the MTNR1B Melatonin receptor 1B, in yet another GPCR sub-clade, with a SEA E-value of 5.3×10^{-13}. Here too, we found a particular MTNR1B agonist (EC_{50} = 14 μM), compound **3,** that we measured to antagonize NPY5R with a K_i of 1.9 μM.

We next asked how many of the GPCRs were strongly related by ligand similarity to a sequence-unrelated target. Interrogating all of the ChEMBL ligand sets, there were 485 *non*-GPCRs that resembled at least one GPCR in our dendrogram with an E-value of 1×10^{-10} or better (lower). Similarity values ranged from this level, for the ligand sets of the EDG7 lysophosphatidic acid GPCR and the enzyme Arachidonate 12-lipoxygenase (LOX12), to 3×10^{-314} for the ligand sets of the NTSR1 neurotensin 1 GPCR and Sortilin (SORT). These *non*-GPCR targets covered most protein families including ion channels, enzymes, kinases, and glycoproteins. Indeed, there were so many *non*-GPCR to GPCR links that clarity only allowed us to show up to two for any given GPCR (Fig. 3).

Here again, many highly-related pairs shared no single ligand between them, and for a few we predicted and tested ligands that would bind to both targets. We started with CXCR2 and casein kinase 1 (KC1G1), linked by a SEA E-value of 1.3×10^{-15}, and identified an inhibitor of the kinase that resembled the CXCR2 ligands. Compound **4** was tested and found to be an agonist for CXCR2 with an EC_{50} of 254 nM (Table 2, Fig. 4). More ambitiously, we searched for a compound that can inhibit a GPCR and an enzyme in the same pathway. One such link was between the α2 adrenergic receptors and their downstream phospholipase D1 and D2 (PLD1 and PLD2) enzymes[27]. Compound **5**, a known phospholipase inhibitor, was tested against three α_2 adrenergic receptor subtypes and had a K_d of 556 nM to the α_{2c} sub-type (Table 2, Fig. 4).

Finally, we sought targets implicated not only in the same pathway, but also in a similar clinical indication. Among these were the cannabinoid receptors and the enzyme epoxide hydralase 2 (HYES), whose ligand sets have an E-value of 1.3×10^{-18}. Intriguingly, both proteins are cardioprotectant targets and both are in the endocannabinoid pathway (epoxide hydrolase 2 deactivate epoxidated endocanniboids).[28] We identified compound **6**, an HYES inhibitor, as a potential CNR2 cannabinoid receptor 2 ligand. On testing, compound **6** had K_i values of 3.6 and 2.3 μM against CNR1 and CNR2, respectively (Table 2, Fig. 4).

Discussion

Relationships among targets are typically visualized by sequence-based family trees, and it is common to infer from these trees both on- and off-target pharmacology[29]. A key observation from this study is that when GPCRs are compared by ligand similarity, the

arborization of the family tree changes dramatically. Targets that are neighbors by sequence are separated, while targets that are distant by sequence become neighbors. This is reflected in targets that unexpectedly respond to the same drugs and reagents, and can predict sequence-distant neighbors that will share ligands where none were previously known. The predicted and confirmed cross-activity of ligands against the κ opioid and serotonin receptors, the cannabinoid and neuropeptide Y receptors, and the neuropeptide Y and melatonin receptors, is doubly unexpected. These pairs of targets not only share little residue identity in their orthosteric sites, from 7% to 28%, but they cross target boundaries among the GPCRs: from peptide to biogenic amine, lipid to peptide, and peptide to neutral small molecule. More startling still is the observation that many *non*-GPCRs strongly resemble GPCRs by ligand similarity (Fig. 3, Table 2). Whereas some of this undoubtedly reflects the conservatism of medicinal chemistry, it is impossible to look at the penumbra of *non*-GPCRs that are strongly associated with GPCRs (Fig. 3) without wondering whether a more basic principle might be at work.

As sequence similarities reflect the action of evolution on proteins, the ligand-based dendograms may reflect the chemical pressures *against which* the receptors have evolved. Many primary signaling molecules themselves target receptors unrelated by sequence or structure. For instance, serotonin modulates both the HTR3 receptor, an ion channel, and the HTR1-2,4-7receptors, which are GPCRs. Acetylcholine targets the nicotinic receptors (ion channels) and the muscarinic receptors (GPCRs). Glutamate and GABA similarly both signal ionotropically and metabatropically. Leukotriene B4 activates GPCRs and the nuclear hormone PPARs. Estrogen binds to not only its eponymous nuclear hormone receptor but also to GPR30 (Supplementary Table 2).[30]

The promiscuity of primary signaling molecules reflects two constraints in biological signaling. First, cells respond to signals in multiple time domains: the millisecond, the second to minute, and the hour-to-day. To achieve this temporal resolution, they will often use ion channels, GPCRs, and nuclear hormone receptors, respectively. Second, these responses are evoked by a small repertoire of chemical messengers; once the machinery to synthesize, degrade, and regulate molecules like serotonin, acetylcholine, and estrogen is created, it is costly to change and becomes fixed.[31] On the other hand, it is relatively easy for evolution to repurpose an ion channel to recognize serotonin or acetylcholine, or a GPCR to recognize glutamate. Thus, the ability of receptors across major sequence and fold boundaries to recognize related ligands, which is captured in the ligand-based dendograms, may reflect a core chemo-evolutionary constraint in molecular signaling. If true, then probe and drug polypharmacology is neither epiphenomenal nor capricious, but reflects the evolution of signaling relationships in the time domain. Pragmatically, the associations among unrelated targets, revealed in the ligand-based dendograms, may suggest joint targets for a single molecule. Known examples are drugs that bind to both ionotropic and metabatropic serotonin receptors, like alosetron, or that bind to both muscarinic receptors and acetylcholinesterase, like flaxedil (Fig. 3). Meanwhile, the discovery that compound **6** modulates both cannabinoid GPCRs and epoxide hydrolase 2 is consistent with a role for this enzyme in the degradation pathway of the endocannabinoids, potentially arresting their signaling.[28]

Several weaknesses in this approach merit airing. Most prominently, a ligand-based view remains inference-based: targets for which no ligands are known are invisible to it, and even when ligands *are* known they can never be known perfectly, unlike the protein sequence. Mechanically, SEA remains imperfect, here, as previously,[7,12,32] the method had a 50% false-positive rate, with six of twelve predictions falsified by experiment (Supplementary Table 3). Pharmacologically, finding a ligand to modulate a GPCR and an enzyme *in vitro* does not guarantee intracellular enzyme inhibition *in vivo*, though GPCR activity of an enzyme inhibitor may be more likely. Also, we do not currently distinguish among agonists and antagonists, nor even between allosteric and orthosteric ligands; the conflation of these for a single receptor weakens the signal on which SEA operates. Meanwhile, in some protein families, such as the kinases, ligand-based and sequence-based dendrograms may resemble each other more closely than do the GPCRs, since the binding site environments are more similar and the proteins bind a single or closely related native ligand.

These cautions should not obscure the central observation from this study: a systematic and comprehensive ligand-based receptor organization differs startlingly from the more familiar sequence-based view. If this approach is weakened by ligand-based inference, it is also true that at least one other chemoinformatic approach, using only partially overlapping ligands and GPCRs, results in a dendogram with receptor associations and disassociations that resemble those observed here[29]. Pragmatically, ligand-based organizations of receptors offer a guide to the off-targets of tool and therapeutic molecules that is orthogonal to, but sometimes as illuminating as, the sequence-based view. More broadly, the association of 485 *non*-GPCRs with GPCRs by ligand similarity suggests a model for polypharmacology that reflects to the roles of primary messengers in cellular signaling. A virtue of this model is that it leads naturally to testable hypotheses, articulated through the very molecules that are the basis of the ligand-based organization. Some of these are suggested by the dendrograms investigated here (Fig. 3).

Online Methods

Sequences and structural alignment

The initial transmembrane sequence alignments were downloaded and filtered for human sequences only. The 43 binding site residues described by Gloriam DE et. al.[19] were then extracted for all human sequences, maintaining the sequence alignments.

Annotated ligands

The ligands and affinity data were downloaded from ChEMBL (version 7) and filtered by their binding affinity values to create sets of ligands for targets if their IC_{50}, K_d, K_i or EC_{50} were 10μM or less. Ligands were also filtered by molecular weight (under 700), nitrogen count (fewer than eight) and oxygen count (fewer than eight) to remove large molecules and peptides. The ChEMBL database does not explicitly differentiate between agonists and antagonists for its ligands and here we combine both into the same ligand-set for each GPCR without differentiating their functional activity. 146 human GPCR sets and 2090 *non*-GPCR protein sets were assembled that each contained at least five annotated ligands and were used to compare using SEA.

Protein similarity calculations

Binding site sequence alignments were used to calculate relative distances between all 146 GPCRs that had ligand sets associated with them. The Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm was used in MEGA 3.1[33] to produce the pairwise distance matrix between all GPCRs. Similarities ranged from just 5% to 88% identity in the binding sites with an average of 23%.

The ligand sets were also used to calculate relative distances between all 146 GPCRs by using SEA[11] to obtain E-values between each GPCR. Each ligand was broken down into molecular fingerprints; here Extended Connectivity Fingerprints (ECFPs) [34] were used. Briefly, ECFPs are circular topological fingerprints that represent molecular structures by small atom neighborhoods or substructures, along with their physical chemical properties. The similarity between any pair of bit strings (molecules) is quantified by the bits they share in common divided by the total number of bits, via the Tanimoto coefficient (Tc)[35]. The sum of all Tc values over a certain cutoff between all the molecules in the two target-ligand sets is then calculated and compared to what we would expect for two sets of ligands, of the same set size, randomly drawn from ChEMBL. The ratio of the observed sum of Tc values to that expected at random is divided by the standard deviation of the random similarity to give a Z-score; when plotted against an extreme value distribution, this gives an expectation value (E-value). The E-values were then logged and used to calculate the pairwise cosine angle. The cosine angle was used as the distance metric, since E-values are not necessarily completely correlative with similarity, rather we use them as more of a binary measure with E-values less (better) than 1×10^{-5} as significant and anything greater (worse) taken as insignificant. Therefore, using the cosine angle, the magnitude of the E-value is not over-weighed such that E-values of 1×10^{-300} and 1×10^{-20} are treated about the same since they are both significant E-values. Similarity between GPCRs and *non*-GPCRs were calculated in the same way, using the annotated ligand sets as surrogates for the protein to calculate SEA E-values. The two lowest E-values between each GPCR and the *non*-GPCRs were retained.

Dendrograms

Using the similarity distance matrices of the binding site sequences and ligand sets, dendrograms were constructed using FigTree. The distance matrices were inputted in Newick format and a radial tree format was used for the layout. The spread was increased to better distinguish the proteins that are highly similar to each other. Nodes were further expanded out in Adobe Illustrator for legibility and color coded based on the chemistry of their endogenous ligands, e.g. peptide, bioamine, lipid, and so forth. The two *non*-GPCRs with the lowest E-value was drawn on using Adobe Illustrator and linked to their respective GPCRs. The *non*-GPCRs in bold and italicized represent those *non*-GPCRs that have a known shared ligand with the GPCR.

Radioligand competition binding assays

Standard techniques were used [36] at the NIMH Psychoactive Drug Screening Program.

CXCR2 β-Arrestin Recruitment Tango Assay

Recruitment of β-arrestin to agonist-stimulated CXCR2 receptors was performed using a previously described "Tango"-type assay. [37] Briefly, HTLA cells stably expressing β-arrestin-TEV protease and a tetracycline transactivator-driven luciferase were plated in 10-cm dishes in DMEM containing 10% FBS and transientlytransfected (via calcium phosphate) with 10 μg of a CXCR2-V_2-TCS-tTA construct. The next day, cells were plated in white, clear-bottom, 384-well plates (Greiner; 15,000 cells/well, 50 μL/well) in DMEM containing 1% dialyzed FBS and incubated overnight at 37°C. The following day, cells were challenged with 10 μL/well of reference agonist or CXCR2 test ligand (CXCL6 and CXCL8) at evenly distributed concentrations that ranged from 6 pM to 60 μM prepared in HBSS, 20 mM Hepes, pH 7.4, and 6% DMSO (final ligand concentrations are 1 pM to 10 μM, final DMSO concentration is 1%). After 18 h, the medium was removed and replaced with 1× BriteGlo reagent (Promega), and luminescence per well was read using a TriLux plate reader (1 s/well). Data were normalized to vehicle (0%) and reference compound (100%) controls and regressed using the sigmoidal dose-response function built into GraphPad Prism 5.0.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by NIH grant GM71896 (to B.K.S. & J. Irwin) and by the National Institutes of Mental Health Psychoactive Drug Screening Program (to B.L.R.). We thank E. Gregori-Puigjane, M. Keiser and R. Coleman for reading this manuscript. B.K.S. is grateful to I.D. Kuntz and H. Bourne for illuminating conversations about time domains in molecular signaling.

References

1. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins. 1991; 9:56–68. [PubMed: 2017436]
2. Orengo CA, et al. CATH-a hierarchic classification of protein domain structures. Structure. 1997; 5:1093–1109. [PubMed: 9309224]
3. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25:25–29. [PubMed: 10802651]
4. Hopf, Thomas A., et al. Three-dimensional structures of membrane proteins from genomic sequencing. Cell. 2012; 149:15.
5. Bateman A, et al. The Pfam protein families database. Nucleic Acids Res. 2004; 32:D138–D141. [PubMed: 14681378]
6. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002; 30:1575–1584. [PubMed: 11917018]
7. Lounkine E, et al. Large-scale prediction and testing of drug activity on side-effect targets. Nature. 2012; 486:361–367. [PubMed: 22722194]
8. Morphy R, Kay C, Rankovic Z. From magic bullets to designed multiple ligands. Drug Discov Today. 2004; 9:641–651. [PubMed: 15279847]
9. Roth BL, Sheffler DJ, Kroeze WK. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. Nat Rev Drug Discov. 2004; 3:353–359. [PubMed: 15060530]

10. Watanabe M, et al. Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. Nature. 2006; 439:484–489. [PubMed: 16400329]
11. Keiser MJ, et al. Relating protein pharmacology by ligand chemistry. Nat Biotechnol. 2007; 25:197–206. [PubMed: 17287757]
12. Keiser MJ, et al. Predicting new molecular targets for known drugs. Nature. 2009; 462:175–181. [PubMed: 19881490]
13. Hopkins AL, Mason JS, Overington JP. Can we rationally design promiscuous drugs? Curr Opin Struct Biol. 2006; 16:127–136. [PubMed: 16442279]
14. Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol. 2003; 63:1256–1272. [PubMed: 12761335]
15. Bender A, et al. Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. ChemMedChem. 2007; 2:861–873. [PubMed: 17477341]
16. Lengauer T, Lemmen C, Rarey M, Zimmermann M. Novel technologies for virtual screening. Drug Discov Today. 2004; 9:27–34. [PubMed: 14761803]
17. Paolini GV, Shapland RHB, van Hoorn WP, Mason JS, Hopkins AL. Global mapping of pharmacological space. Nat Biotechnol. 2006; 24:805–815. [PubMed: 16841068]
18. Gaulton A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. 2012; 40:D1100–D1107. [PubMed: 21948594]
19. Gloriam DE, Foord SM, Blaney FE, Garland SL. Definition of the G protein-coupled receptor transmembrane bundle binding pocket and calculation of receptor similarities for drug design. J Med Chem. 2009; 52:4429–4442. [PubMed: 19537715]
20. Sokal RR, Michener CD. A statistical method for evaluating systematic relationships. Univ Kans Sci Bull. 1958; 28:1409–1438.
21. Hert, Jrm; Keiser, MJ.; Irwin, JJ.; Oprea, TI.; Shoichet, BK. Quantifying the relationships among drug classes. J Chem Inf Model. 2008; 48:755–765. [PubMed: 18335977]
22. Guba W, et al. From Astemizole to a novel hit series of small-molecule somatostatin 5 receptor antagonists via GPCR affinity profiling. J Med Chem. 2007; 50:6295–6298. [PubMed: 18020391]
23. Balavoine F, et al. Design and synthesis of novel hydantoin-containing melanin-concentrating hormone receptor antagonists. Bioorg Med Chem Lett. 2007; 17:3754–3759. [PubMed: 17448659]
24. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: A free tool to discover chemistry for biology. J Chem Inf Model. 2012; 52:12.
25. Frankowski KJ, et al. N-Alkyl-octahydroisoquinolin-1-one-8-carboxamides: Selective and Nonbasic κ-Opioid Receptor Ligands. ACS Med Chem Lett. 2010; 1:189–193. [PubMed: 20729985]
26. Cheng Y, et al. Discovery and Optimization of a Novel Series of N-Arylamide Oxadiazoles as Potent, Highly Selective and Orally Bioavailable Cannabinoid Receptor 2 (CB2) Agonists. J Med Chem. 2008; 51:5019–5034. [PubMed: 18680277]
27. Scott SA, et al. Design of isoform-selective phospholipase D inhibitors that modulate cancer cell invasiveness. Nat Chem Biol. 2009; 5:108–117. [PubMed: 19136975]
28. Snider NT, Walker VJ, Hollenberg PF. Oxidation of the endogenous cannabinoid arachidonoyl ethanolamide by the cytochrome P450 monooxygenases: Physiological and pharmacological implications. Pharmacol Rev. 2010; 62:136–154. [PubMed: 20133390]
29. van der Horst E, et al. A novel chemogenomics analysis of G protein-coupled receptors (GPCRs) and their ligands: a potential strategy for receptor de-orphanization. BMC Bioinformatics. 2010; 11:316. [PubMed: 20537162]
30. Revankar CM, Cimino DF, Sklar LA, Arterburn JB, Prossnitz ER. A transmembrane intracellular estrogen receptor mediates rapid cell signaling. Science. 2005; 307:1625–1630. [PubMed: 15705806]
31. Tomkins GM. The metabolic code. Science. 1975; 189:760–763. [PubMed: 169570]
32. Gregori-Puigjane E, et al. Identifying mechanism-of-action targets for drugs and probes. Proc Natl Acad Sci. 2012; 109:11178–11183. [PubMed: 22711801]

33. Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 2004; 5:150–163. [PubMed: 15260895]

34. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010; 50:742–754. [PubMed: 20426451]

35. Rogers DJ, Tanimoto TT. A computer program for classifying plants. Science. 1960; 132:1115–1118. [PubMed: 17790723]

36. Jensen NH, et al. N-Desalkylquetiapine, a potent norepinephrine reuptake inhibitor and partial 5-HT1A agonist, as a putative mediator of quetiapine's antidepressant activity. Neuropsychopharmacology. 2007; 33:2303–2312. [PubMed: 18059438]

37. Chen X, et al. Structure functional selectivity relationship studies of β-arrestin-biased dopamine D2 receptor agonists. J Med Chem. 2012; 55:7141–7153. [PubMed: 22845053]

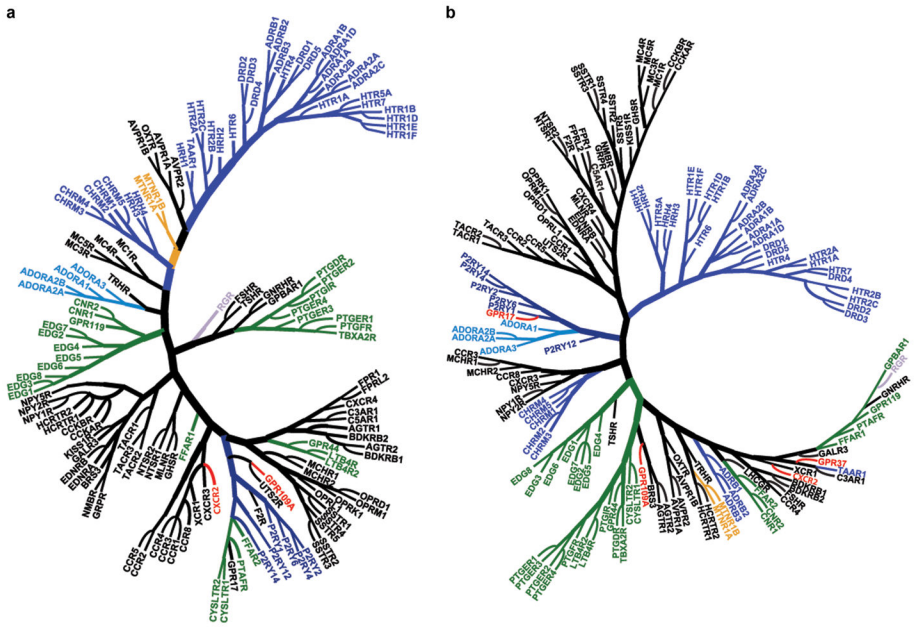


Figure 1.

Dendrograms of human GPCRs with annotated ligands from ChEMBL. Organization based on (a) sequence similarity in the binding site and (b) ligand set similarity based on SEA E-values. Color coding is based on chemistry of their endogenous ligands (i.e. Bioamines (blue), melatonins (gold), lipids (green), peptides (black), purinergics (dark blue), adenosines (light blue), orphans (red).).

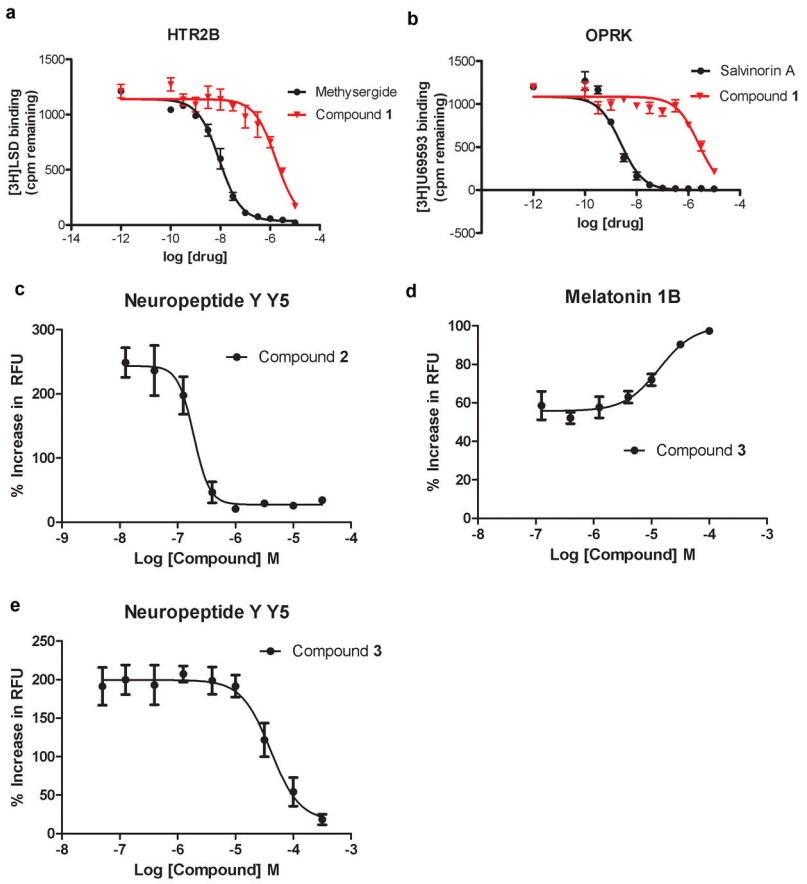


Figure 2.

Dose-response curves of new GPCR cross-activities. (**a–e)** Radioligand competition binding assay: compound **1** at HTR2B (**a**) and OPRK (**b**), compound **2** at NPY5R (**c**), compound **3** at MTR1B (**d**) and NPY5R (**e**). Data represent mean values ± s.e.m, performed on triplicate experiments.

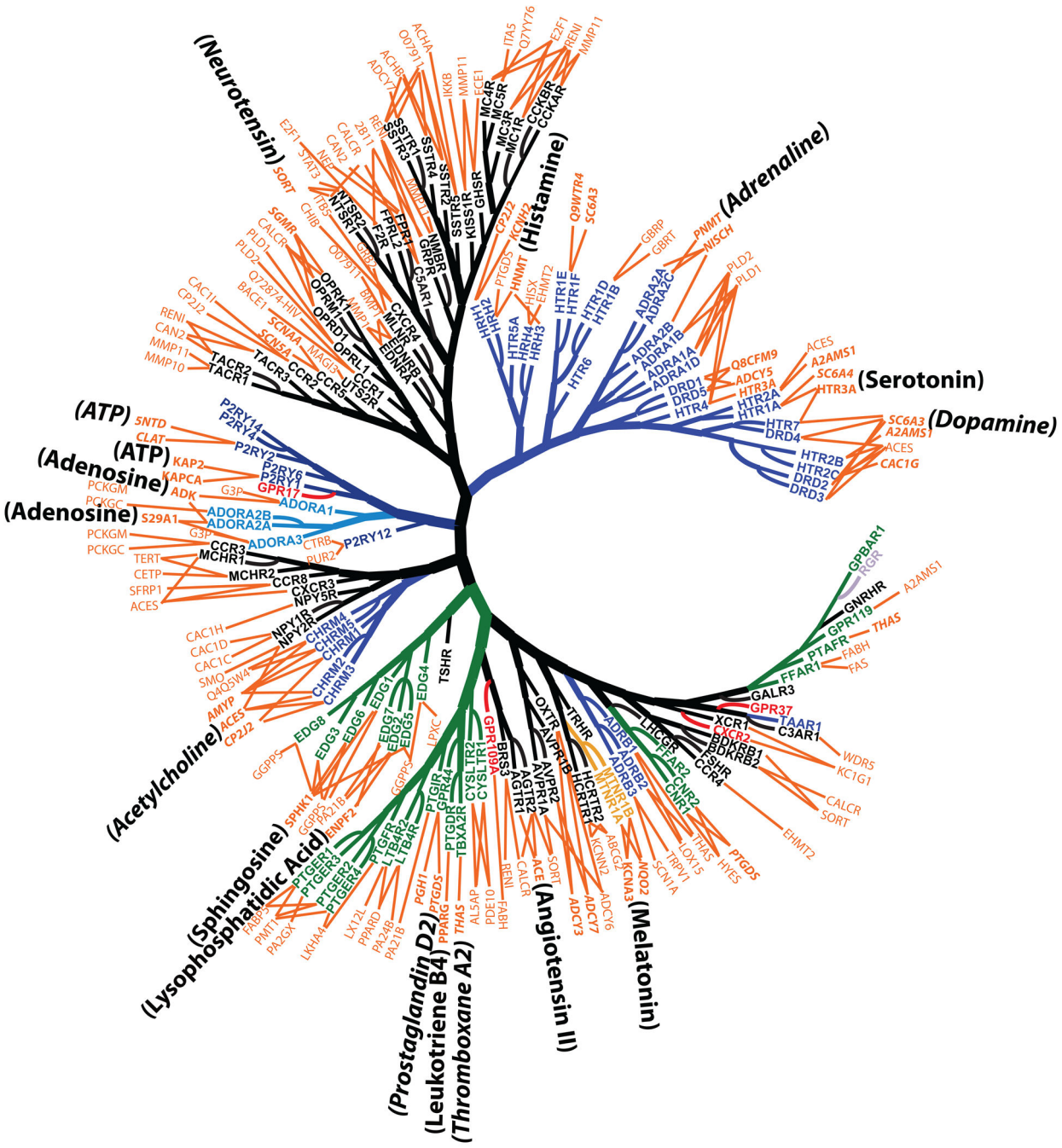


Figure 3.

Non-GPCRs (orange) highly-related to particular GPCRs by ligand similarity (color code is as in Figure 1). Bolded targets have known ligands that bind to both the GPCR and non-GPCR target. Links that share known messengers are labeled in black in parenthesis.

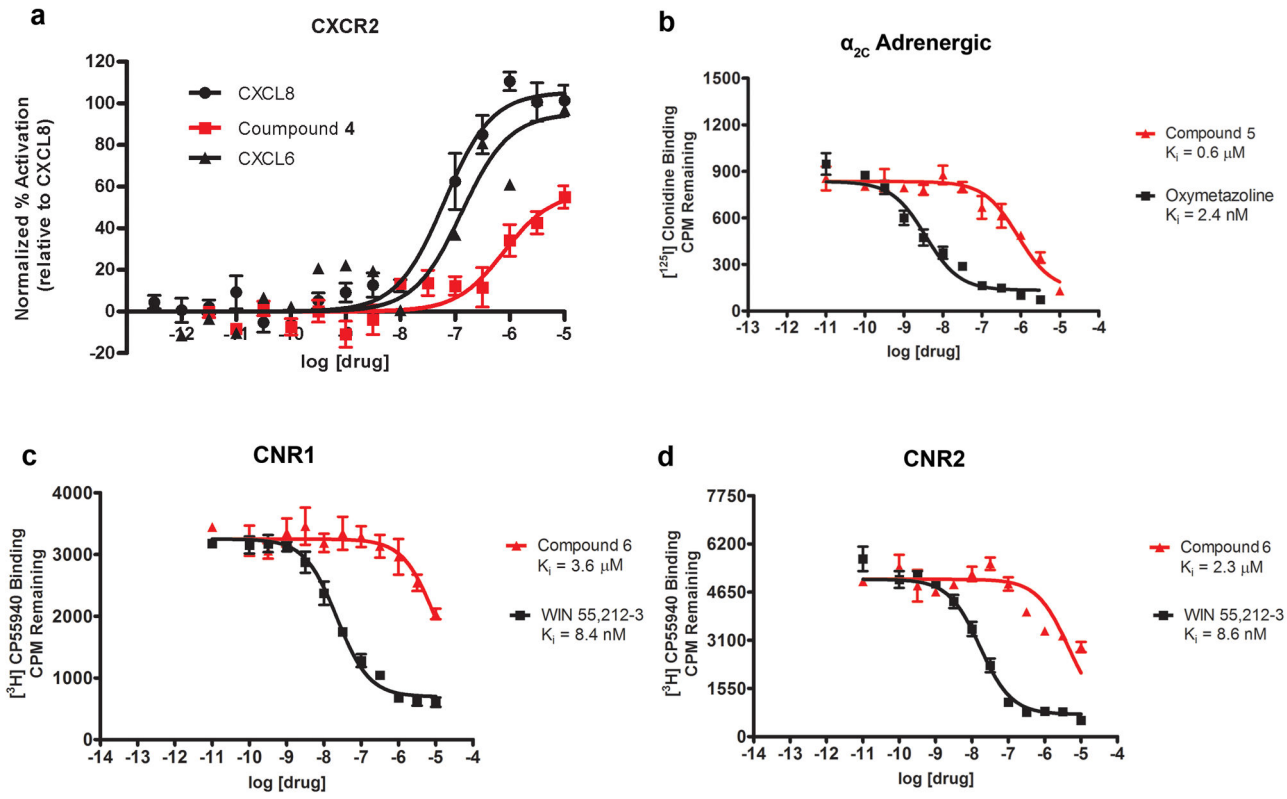


Figure 4.

Dose-response curves of new GPCR cross-activities with non-GPCRs. Testing new GPCR cross-activities with non-GPCRs. β-Arrestin Recruitment Tango Assay: compound **4** at CXCR2 (**a**), competition binding assay: compound **5** at α_{2c} Adrenergic receptor (**b**), compound **6** at CNR1 (**c**) and CNR2 (**d**). Data represent mean values ± s.e.m, performed on triplicate experiments.

Table 1

Predicted and confirmed ligand associations between GPCRs with lowsequence identities.

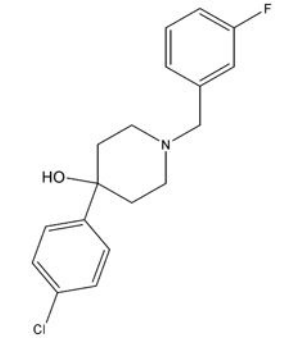
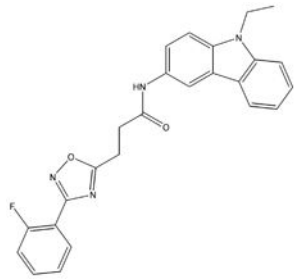
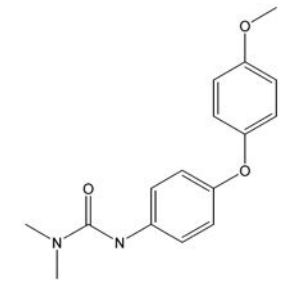
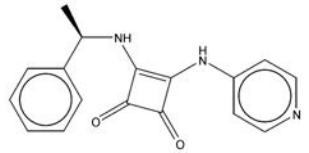
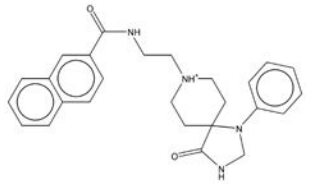
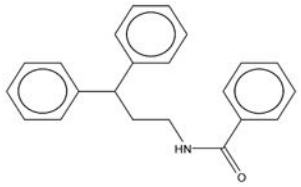
Compound	GPCR1	GPCR1 affinity of predicted ligand (uM)	GPCR2	GPCR2 affinity of predicted ligand (uM)	Sequence Identity (shared residues)	SEA E-value between ligand sets	Predicted Ligand
1	5-HT_{2B} serotonin (HTR2B)	K_i = 0.9	Opioid receptor (OPRK)	K_i = 1.0	28% (12)	9.88×10^{-8}	
2	Neuropeptide Y Y5 receptor (NPY5R)	K_i = 0.0085	Cannabinoid receptor 2 (CNR2)	EC_{50} = 0.11	7% (3)	1.08×10^{-9}	
3	Neuropeptide Y receptor 5 (NPY5R)	K_i = 1.9	Melatonin 1B receptor (MTNR1B)	EC_{50} = 15	21% (9)	5.33×10^{-13}	

Table 2

New GPCR associations with *non*-GPCR proteins linked by tested ligands.

Compound	GPCR	GPCR affinity (μM)	*Non*-GPCR	*Non*-GPCR affinity (uM)	SEA E-value	Ligand
4	C-X-C chemokine receptor type 2 (CXCR2)	0.78[a]	Casein kinase I isoform gamma-1 (KC1G1)	0.25 [c]	1.3×10^{-15}	
5	α_{2c} Adrenergic receptor (ADA2C)	0.6 [b]	Phospholipase D2 (PLD2)	0.11 [c]	9.22×10^{-15}	
6	Cannabinoid receptor 1 (CNR1) Cannabinoid receptor 2 (CNR2)	3.6 [b] 2.3 [b]	Epoxide hydralase 2 (HYES)	0.005 [c]	1.3×10^{-8}	

[a] EC_{50}

[b] K_i

[c] IC_{50}