

Orphan receptor ligand discovery by pickpocketing pharmacological neighbors

Tony Ngo^{1,2}, Andrey V Ilatovskiy³, Alastair G Stewart^{2,4}, James L J Coleman^{1,2}, Fiona M McRobb³, R Peter Riek¹, Robert M Graham^{1,2}, Ruben Abagyan³, Irina Kufareva^{3*} & Nicola J Smith^{1,2*}

Understanding the pharmacological similarity of G protein-coupled receptors (GPCRs) is paramount for predicting ligand off-target effects, drug repurposing, and ligand discovery for orphan receptors. Phylogenetic relationships do not always correctly capture pharmacological similarity. Previous family-wide attempts to define pharmacological relationships were based on three-dimensional structures and/or known receptor–ligand pairings, both unavailable for orphan GPCRs. Here, we present GPCR-CoINPocket, a novel contact-informed neighboring pocket metric of GPCR binding-site similarity that is informed by patterns of ligand-residue interactions observed in crystallographically characterized GPCRs. GPCR-CoINPocket is applicable to receptors with unknown structure or ligands and accurately captures known pharmacological relationships between GPCRs, even those undetected by phylogeny. When applied to orphan receptor GPR37L1, GPCR-CoINPocket identified its pharmacological neighbors, and transfer of their pharmacology aided in discovery of the first surrogate ligands for this orphan with a 30% success rate. Although primarily designed for GPCRs, the method is easily transferable to other protein families.

Class A G protein-coupled receptors (GPCRs) are well-established therapeutic targets, yet a subgroup of them remains largely unexplored—*orphan GPCRs*. One of the main obstacles for understanding the therapeutic potential of orphan GPCRs is the absence of a cognate ligand, which is identified through ‘*deorphanization*’¹. While finding the endogenous ligand is of paramount importance, much can be learned by the identification of even low-potency small-molecule surrogate ligands. The propensity of ligands to display ‘*off-target*’ activities provides rationale for surrogate ligand searches by screening clinically tested small molecules^{2–4}. Nevertheless, screening campaigns remain expensive and often inaccessible to academic research, while structure-based drug design at orphan GPCRs is still in its infancy⁵. Thus, alternative approaches to surrogate ligand identification are required.

Given that a number of homology-related receptor subgroups, such as the muscarinic or adrenergic receptors, share endogenous and often small-molecule ligands, pharmacologically characterized receptors that are homologous to an orphan GPCR in question represent an obvious source of surrogate ligand candidates. Such homologs are traditionally determined by sequence alignment and calculation of amino acid identity or similarity at each position. Phylogenetic relationships are largely sensible⁶, and underlie classification systems for the wider GPCR superfamily⁷; however, they are not always predictive of pharmacological similarities. Phylogenetically unrelated GPCRs frequently share ligands; for example, astemizole, an antagonist of the bioaminergic histamine H₁ receptor, also antagonizes the somatostatin peptide receptor (sst) 5, sst₅ (refs. 8,9). On the other hand, homology-related receptors frequently differ substantially in their ligand selectivity (exemplified by the chemokine receptor subfamily¹⁰).

A breakthrough in understanding pharmacological relatedness of GPCRs came from arranging GPCRs by their ligand similarities¹¹. This chemogenomics approach not only explained known ‘*off-target*’ effects of certain ligands, but also predicted relationships

unseen by phylogeny alone. These newly predicted relationships were experimentally validated, extending previous computational efforts to organize GPCRs by their ligand similarity¹². Although a major advance in the field, the approach cannot identify pharmacological neighbors for orphan GPCRs because it relies upon existing ligand–receptor pairings.

Intuitively, pharmacological similarity of GPCRs implies similar binding-site features. However, various attempts to organize GPCRs by binding-site comparisons largely reproduced phylogenetic relationships^{11–15}. This may reflect shortcomings in the definitions of the binding site, which lack sufficient texture to reveal the true ligand interaction pharmacophores. Here we sought to further refine the comparisons of GPCR binding sites by analyzing both the persistence and the strength of residue–ligand interactions observed across the GPCR Pocketome—the set of annotated GPCR binding pockets that have been crystallographically characterized to date¹⁶. By enriching the binding-site comparisons with ligand contact strengths, a method we term GPCR-CoINPocket (GPCR contact-informed neighboring pocket), we organized and clustered all class A GPCRs into a hierarchical structure that closely reproduced previously characterized pharmacological relationships. Furthermore, we discovered the first surrogate ligands for the orphan receptor GPR37L1 by transferring pharmacology of its neighbors identified by GPCR-CoINPocket. Hence, our approach not only complements the phylogenetic and pharmacological organization of GPCRs, but also uniquely enables the discovery of surrogate ligands for orphan receptors.

RESULTS

Ligand interaction patterns across the GPCR Pocketome

Phylogenetic similarities of GPCRs, even those based on partial alignments of carefully selected pocket residues¹⁵, usually do not recapitulate the chemogenomic organization of GPCRs¹¹. We hypothesized that alignment-based recognition of pharmacological

¹Molecular Cardiology and Biophysics Division, Victor Chang Cardiac Research Institute, Darlinghurst, New South Wales, Australia. ²St Vincent's Clinical School, University of New South Wales, Darlinghurst, New South Wales, Australia. ³Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California, USA. ⁴Molecular, Structural and Computational Biology Division, Victor Chang Cardiac Research Institute, Darlinghurst, New South Wales, Australia. *e-mail: n.smith@victorchang.edu.au or ikufareva@ucsd.edu

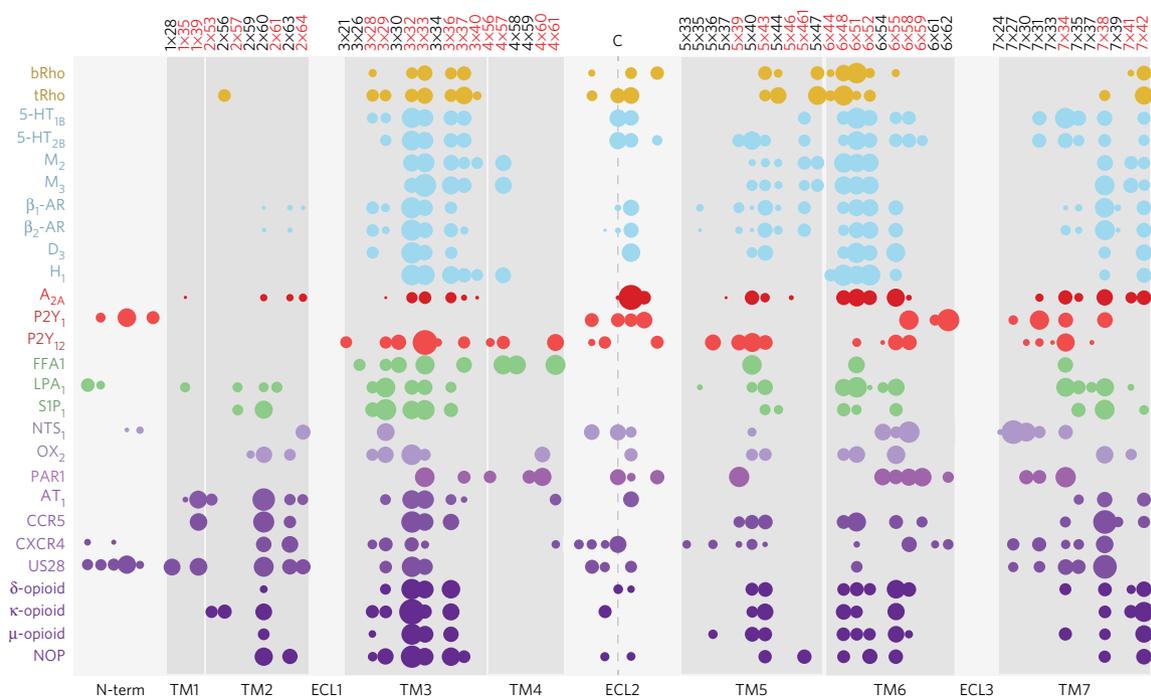


Figure 1 | Ligand contact map of class A GPCRs. Ligand contact-strength fingerprints of receptor residues located in the transmembrane (TM) and extracellular loop two (ECL2) regions of crystallized class A GPCRs, as collated in the Pocketome. The area of the circles represents the relative strength of the ligand contact and residue position is numbered according to the GPCRdb numbering scheme¹⁸. Residue positions overlapping the binding site in Gloriam *et al.*¹⁵ are highlighted in red. Ligand contacts within three residues of the highly conserved ECL2 cysteine have been portrayed (LPA1 receptor ECL2 contacts excluded). Contacts from residue backbone atoms (except for glycine), ECL1 and ECL3 were also excluded. Colors of the circles represent the chemical class of the GPCR's endogenous ligand: gold, retinal; blue, biogenic amines; red, nucleosides; green, lipids; purple, peptides.

similarities can be improved by accounting for the importance and/or strength of individual residue–ligand interactions. Informed by the growing number of published GPCR crystal structures that are also annotated in the Pocketome¹⁶, we calculated contact strengths between ligands and receptor residue side chains for each co-crystallized GPCR (see Online Methods). As calculated, residue contact strengths (1) are distance based, not energy based; (2) represent a coarse-grain approximation of actual interaction energies; (3) have improved signal-to-noise ratios through analysis of multiple ligands and crystallographic conformational ensembles; and (4) provide a reasonable compromise between a binary definition of interatomic contact based on a cut-off interatomic distance and the complexities and ambiguities of accurate energy calculations for conformationally variable crystallographic ensembles.

The calculation resulted in identification of a ‘cloud’ of 61 residue positions that outline the transmembrane (TM) domain (N termini and extracellular loops (ECLs) were excluded) across the 27 unique class A GPCR entries in the Pocketome (Fig. 1). The chemical diversity of endogenous and co-crystallized ligands of GPCRs is reflected in the variation of amino acid composition of their binding pockets. However, we found that the core pattern of residue positions involved in ligand interactions is largely conserved, similar to previous observations¹⁷; for example, positions 3×32, 3×33, 3×36, 5×43, 6×48, 6×51, 6×55 and 7×38 (GPCRdb numbering¹⁸) are involved in ligand binding in at least 70% of class A GPCR Pocketome entries. Furthermore, these positions consistently had prominent ligand contact strengths (Fig. 1). Although the majority of sites fell within a similar spatial region, the expected diversity amongst GPCR binding pockets was still evident, particularly within peptide binding receptors where TM2 contacts were also strong (Fig. 1).

To facilitate a comparison with previous work, we defined a consensus GPCR binding site by aggregating the obtained fingerprints. The aggregation was based on both the frequency and the

magnitude of ligand contacts in conserved positions. The aggregated GPCR binding site largely overlaps with the Gloriam *et al.* binding-site residues (32 residues out of 43) (ref. 15); however, quantitation of the relative role of each position provides an important methodological distinction. Furthermore, the present approach to such quantitation bears resemblance to previously described weighted protein–ligand interaction fingerprints¹⁹ (PLIFs) that utilize the frequency of residue–ligand interactions to define ligand pharmacophore profiles²⁰; however, our method is advanced in that it accounts for interaction magnitudes and in ‘inverting’ the perspective from ligand to binding site.

Organization of class A GPCRs using GPCR-CoINPocket

Given a sequence alignment of two GPCRs, their similarity can now be calculated with a greater emphasis on important binding site residues. Here, using a global alignment of GPCR TM domains, we calculated the pairwise sequence similarities for all GPCRs using the ligand interaction patterns observed in all Pocketome entries. We then normalized these similarities to balance the contribution of pockets located within evolutionarily conserved regions of the TM domains and averaged the results to generate a final similarity score for every pair of GPCRs (Supplementary Results, Supplementary Data Set 1). This similarity of contact-strength-profiled binding sites is designed to capture the pharmacological similarities of GPCRs by focusing on residues that provide specific and important ligand interaction pharmacophores. We have named this method GPCR-CoINPocket: GPCR contact-informed neighboring pocket.

For visualization and comparison purposes, we produced a heat map based on the unweighted TM sequence similarities across all class A GPCRs (Fig. 2a). Receptors were clustered and ordered according to the unweighted pair group method with arithmetic mean (UPGMA) algorithm²¹ (Fig. 2a). In this organization, homologous receptors group together as expected. Even when restricted

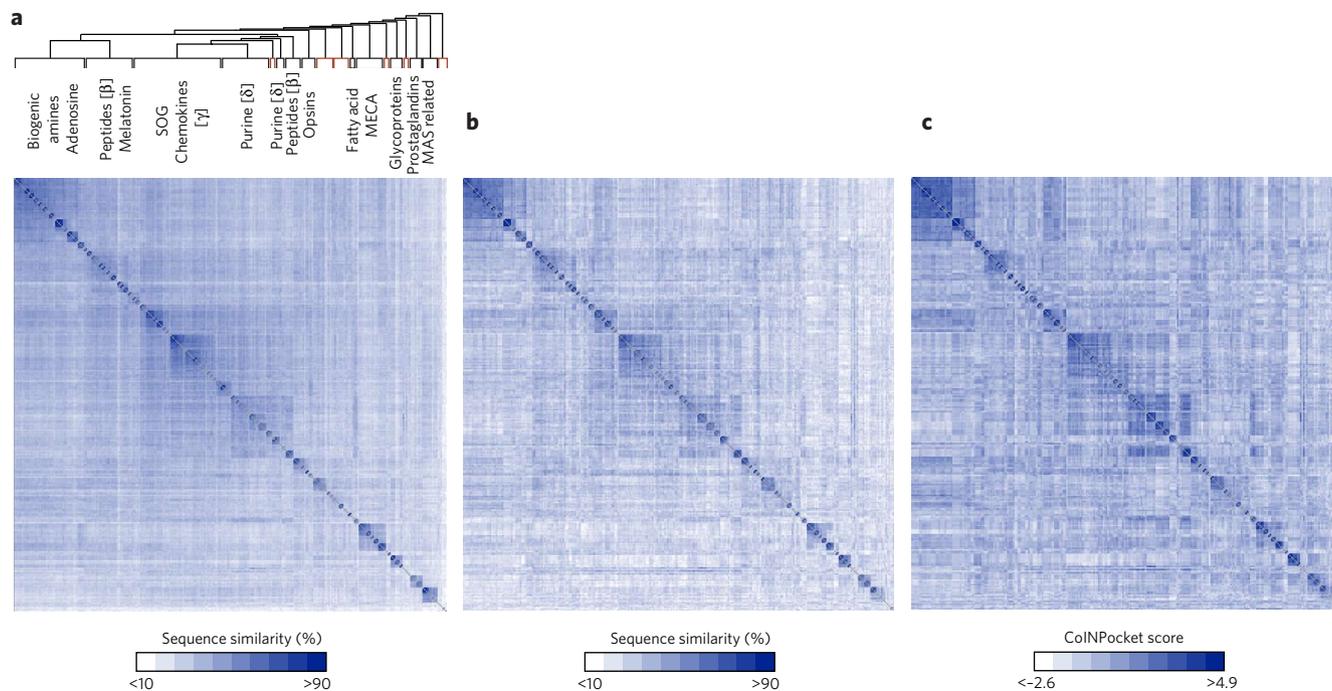


Figure 2 | Comparison matrix of class A GPCR binding site and transmembrane sequence similarities compared to GPCR-CoINPocket. Heat map representation of GPCR sequence similarities based on transmembrane (TM) domains (**a**), unweighted generic GPCR binding site residues (**b**), or contact profiling of GPCR binding sites using GPCR-CoINPocket (**c**). In the heat maps, the GPCRs have been ordered according to TM sequence similarity clustering, with orphan clusters colored in red. The matrix diagonals are colored in gray. SOG, somatostatin, opioids, and galanin. The source data to generate the heat maps can be found in **Supplementary Data Set 1**.

to the uniformly defined generic GPCR binding site (in the spirit of Gloriam *et al.*¹⁵, but taking into account the new GPCR crystal structure contacts), amino acid comparison leads to an organization where receptors display very little similarity outside of their subfamily (**Fig. 2b**). In contrast, in the GPCR-CoINPocket heat map, some GPCRs share high similarities outside of their clustered subgroup (**Fig. 2c**), likely reflecting pharmacological relationships.

We probed the ability of GPCR-CoINPocket to capture established pharmacological similarities, particularly those identified chemogenomically¹¹. GPCR-CoINPocket correctly predicted the relationship between sst_5 and various biogenic amine receptors with very high scores (**Table 1**), reflecting the fact that an aspartic acid at position 3 \times 32 is not only conserved across these receptors but is also one of the strongest contacts found in GPCRs (**Fig. 1**). Likewise, a phenylalanine residue at position 6 \times 52 is also conserved across these receptors, perhaps explaining off-target activity of H_1 receptor ligands at sst_5 receptors^{8,9,11}. A high degree of similarity was detected between the melatonin MT_2 and neuropeptide Y_5

receptors, between the muscarinic M_1 and melanin-concentrating hormone MCH_1 receptors, and between M_2 and CCR5 (**Table 1**). Additionally, the opioid, CCR3 and other chemokine receptors consistently ranked as pharmacological neighbors of the muscarinic receptors, confirming their previously described pharmacological similarities^{11,22} (**Table 1**; **Supplementary Fig. 1**). Finally, despite the peptide melanocortin receptors (MCR) being phylogenetically related to lysophospholipid receptors, they exhibited a high degree of similarity with other peptide receptors, such as the oxytocin and vasopressin receptors; this corroborates the recent discovery of activity of MCR agonists on oxytocin-mediated signaling pathways²³ (**Table 1**, **Supplementary Fig. 2**). These observations demonstrate the ability of GPCR-CoINPocket to identify pharmacological similarities of GPCRs without any knowledge of their ligands.

Next, we sought to validate the ability of GPCR-CoINPocket to identify known pharmacological similarities on a larger scale and in an unbiased manner using a data set of GPCR pairs with shared ligands, as annotated in the ChEMBL database²⁴. Different metrics

Table 1 | Examples of known pharmacological similarities correctly predicted by GPCR-CoINPocket.

GPCR1 ^a	GPCR2 ^a	Pharmacological similarity reference(s)	CoINPocket score ^b	Key conserved contacts
sst_5	D_2 ; 5HT _{2A} ; H_1	8,9	2.05; 1.83; 1.15;	D(3 \times 32), F(5 \times 39), F(5 \times 47), W(6 \times 48), F(6 \times 52), Y(7 \times 42)
MT_2	Y_5	11	1.24	V(3 \times 36), T(5 \times 40), W(6 \times 48), L(6 \times 51)
M_1	MCH_1	11	0.90	D(3 \times 32), T(5 \times 40), W(6 \times 48), Y(6 \times 51), Y(7 \times 42)
M_2	CCR5	42	0.89	T(5 \times 40), W(6 \times 48), Y(6 \times 51), N(6 \times 52), N(6 \times 58)
δ -opioid	M_{1-5}	22	1.35-1.67	L(3 \times 29), D(3 \times 32), W(6 \times 48), V(6 \times 55), Y(7 \times 42)
OT	MC_3 ; MC_4	23	0.93; 0.98;	I(5 \times 40), W(6 \times 48), F(6 \times 51), F(6 \times 52), F(7 \times 34)
CXCR4	ACKR3	27	3.17	Y(1 \times 39), W(2 \times 60), H(3 \times 29), W(6 \times 48), Y(6 \times 51), D(6 \times 58)

^aReceptors have been named and abbreviated in accordance with the IUPHAR/BPS Concise Guide to PHARMACOLOGY. ^bThe final profiled similarity score (GPCR-CoINPocket) is given as the average of the individual Z-scores based on the 27 GPCR binding-site profiles.

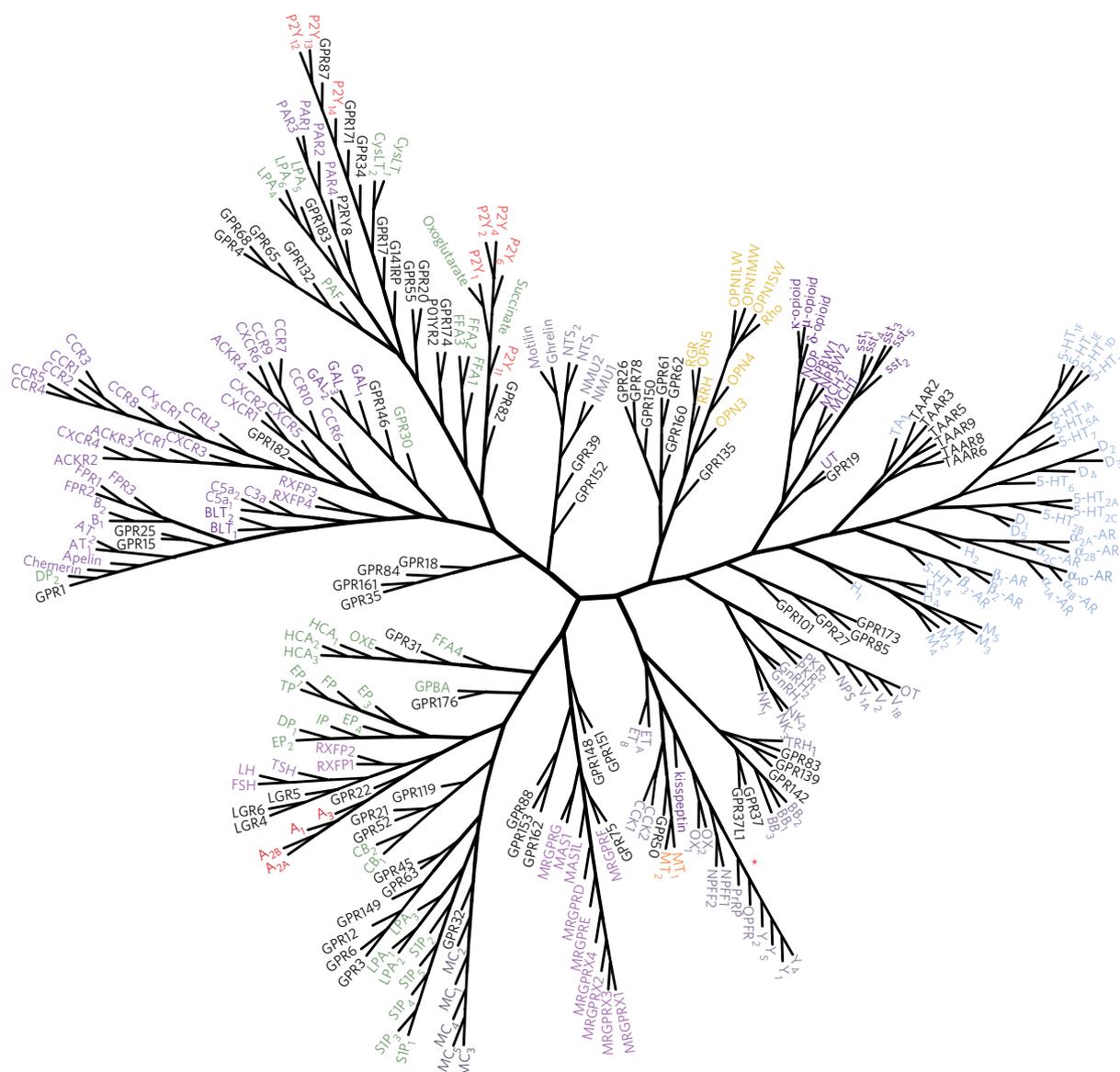


Figure 3 | Organization of class A GPCRs based on GPCR-CoINPocket. Dendrogram of class A GPCRs clustered by contact-strength-profiled binding-site sequence similarities (GPCR-CoINPocket). All receptors have been named and abbreviated in accordance with the IUPHAR/BPS Concise Guide to PHARMACOLOGY (<http://www.guidetopharmacology.org>) except for the adrenoceptors where the suffix AR has been used for figure clarity. GPR37L1 is marked with a red asterisk. Receptor names are colored according to the chemical class of the GPCR's endogenous ligand: gold, retinal; blue, biogenic amines; red, nucleosides; green, lipids; purple, peptides; orange, melatonin; black, orphans.

were tested in their ability to discriminate such receptor pairs from all other pairs, with the performance quantified as the area under the receiver operating characteristic curve (ROC AUC). When applied globally, GPCR-CoINPocket identified these GPCR pairs with the same level of accuracy as TM and the unweighted binding-site similarity (**Supplementary Fig. 3a**). This is not surprising, as the ChEMBL data is inherently biased toward closest phylogenetic relatives because of the common practice of testing ligand selectivity profiles only on homologous receptors. Consequently, for the majority of receptor pairs sharing ligands in ChEMBL, all three of TM similarity, GPCR-CoINPocket score, and unweighted binding-site similarity are trivially high. Yet the overarching aim of our work was to identify surrogate ligands for orphan GPCRs, which have low sequence similarity to either crystal structures or well-characterized receptors⁵. Surrogate ligands are most needed for these unexplored receptors, as they enable both the biochemical characterization of the receptors *in vitro* and ligand-guided homol-

ogy model refinement *in silico*. Thus, it was essential to benchmark the performance of GPCR-CoINPocket when the TM similarity of the annotated GPCR pairs was between 20% and 35%, outside the cutoff for high-precision homology modeling, but high enough to construct an accurate sequence alignment^{25,26}. In this range, we found that GPCR-CoINPocket outperformed TM similarity and the unweighted binding site measure with a ROC AUC of 71.9, 62.0 and 56.3, respectively (**Supplementary Fig. 3b**). This data further supports the ability of GPCR-CoINPocket to explain pharmacological similarities seen between dissimilar GPCRs retrospectively.

We next portrayed the hierarchical clustering of the GPCRs via unweighted binding-site similarities or via GPCR-CoINPocket scores as dendrograms (**Fig. 3; Supplementary Fig. 4**). Whilst dendrograms provide a simple way to visualize GPCR relationships, be they chemogenomic, TM sequence based, or contact-strength profiled, many interesting and subtle observations seen in **Table 1** are lost. In our GPCR-CoINPocket dendrogram, many known receptor subfamilies

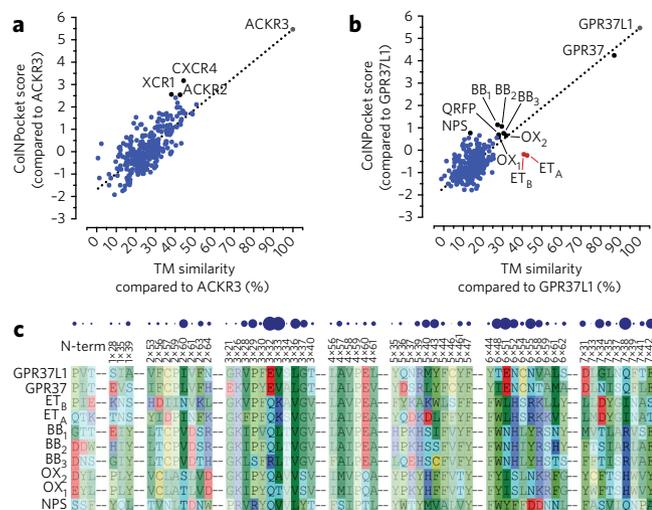


Figure 4 | The pharmacological neighborhood of orphan receptors, ACKR3 and GPR37L1. GPCR–CoINPocket and transmembrane (TM) similarity of receptors, relative to ACKR3 (a) and GPR37L1 (b) were plotted against each other. The top ranking binding-site-related receptors are highlighted in black while in b, the phylogenetically related endothelin receptors (ET) are highlighted in red. Dotted lines represent a linear regression of the data. For ACKR3, GPCR–CoINPocket correctly identified the pharmacological similarity with the distantly related CXCR4 receptor. For GPR37L1, receptors had rather moderate GPCR–CoINPocket scores. (c) The respective binding-site residues are shown for GPR37L1 and its related receptors as defined by contact-strength fingerprints. The aggregated contact strength across the class A GPCR Pocketome entries is represented by blue circles and by the color intensity of the alignment in the respective positions. Receptors have been named according to the IUPHAR/BPS Concise Guide to PHARMACOLOGY.

(for example, biogenic amine or chemokine receptors) were clustered together (Fig. 3), similar to what was seen in the previous study by Gloriam *et al.*¹⁵ and our own unweighted binding-site relationships (Supplementary Fig. 4). However, receptors in larger phylogenetic subgroups frequently diverged from one another in our GPCR–CoINPocket dendrogram, reflecting their diverse pharmacology. For example, we recapitulated the shift of opioid, MCH and sst receptors toward the biogenic amines¹¹, though we did not see separation of the biogenic amine receptors themselves (for example, muscarinic and β -adrenergic receptors¹¹). Overall, the new GPCR relationships that we identify appear to be intermediate between those based on phylogeny and/or TM sequence similarity, generic binding-site definitions and those obtained through the chemogenomic approach¹¹. These observations also support the incorporation of contact strength into our binding-site similarity calculation as the differentiating factor from previous GPCR binding-site definitions.

The pharmacological neighborhood of an orphan GPCR

We reasoned that the newly defined binding-site similarity of GPCRs would aid in finding pharmacological neighbors for orphan GPCRs (Supplementary Data Set 1), and through these neighbors—in the identification of surrogate and/or endogenous ligands for these orphans. Retrospectively, GPCR–CoINPocket correctly identified pharmacological similarity between two phylogenetically distant chemokine receptors, CXCR4 and ACKR3 (now deorphanized), which share the same endogenous ligand, CXCL12 (refs. 1,27) (Fig. 4a)—a relationship that was not identified either in the Gloriam *et al.* binding site or by chemogenomic approaches. For prospective utility of GPCR–CoINPocket, we focused our efforts on the orphan GPCR, GPR37L1, which has been linked to cardiovascular homeostasis and cerebellum development^{28–30}.

Phylogenetically, GPR37L1 (and its closest relative, the orphan GPR37) is most closely related to the endothelin (ET) receptors; therefore, it was originally named the endothelin B receptor-like protein 2 (refs. 30–32). However, GPR37L1 does not bind to, nor is it activated by, ET or related ligands^{31,32}. According to GPCR–CoINPocket, but not to the unweighted GPCR binding site definition, GPR37L1 actually clusters closer to the bombesin (BB) and several orphan GPCRs, while ET receptors are more distant (Fig. 3). This contrasts the high TM similarity between the ET receptors and GPR37L1 (Fig. 4b). Moreover, while orexin (OX), pyroglutamylated RFamide peptide (QRFP) and neuropeptide S (NPS) receptors do not normally cluster with GPR37L1 and GPR37 by phylogeny, they displayed a relatively high degree of similarity to GPR37L1 according to GPCR–CoINPocket (Fig. 4b).

Comparing the binding-site residues of GPR37L1 with those of ET and the newly identified pharmacological neighbors, BB, OX and NPS receptors (Fig. 4c), it became clear that ET binding sites are indeed very dissimilar to GPR37L1, especially in key residue positions heavily weighted by contact strength (Fig. 4c). ET receptors have a lysine at position 3 \times 33, a leucine at 6 \times 51, an arginine at 6 \times 55, an aspartate at 7 \times 34, an isoleucine at 7 \times 38 and a serine or a threonine at 7 \times 42—all dramatic non-conserved substitutions with respect to the Val, Glu, Asn, Gly, Gln, and Phe found at the corresponding key positions of GPR37L1. The basic nature of several key residues in the ET receptors reflects its pharmacology, in which both the C terminus of the ET peptide and the majority of synthetic ET ligands carry a negatively charged carboxyl group³³; the corresponding features are completely absent from GPR37L1. In contrast to ET, OX, BB and NPS receptors have milder, less dramatic substitutions in key positions. Because their key pocket residues are not identical to those of GPR37L1, the obtained GPCR–CoINPocket scores between these receptors and GPR37L1 are only moderate (for example, in comparison to high GPCR–CoINPocket scores for ACKR3 relative to CXCR4; Fig. 4a). However, OX, BB and NPS receptors represent the closest pharmacological neighbors of GPR37L1 among all examined receptors.

Transferring pharmacology identifies ligands for GPR37L1

Although no class A GPCR reproduces the GPR37L1–GPR37 binding site features with high similarity, we sought to examine known ligands of the closest pharmacological neighbors of GPR37L1, the BB, OX and NPS receptors. We purchased eight OX receptor antagonists with varying selectivity for OX₁ and OX₂ receptors, one non-selective BB₁/BB₂ receptor antagonist, PD176252 (peptide agonists for the BB receptors did not bind to GPR37L1 (refs. 31,32)) and one NPS receptor antagonist SHA-68. Our previous studies established that GPR37L1 is a constitutively active receptor coupled to the G α_s –cAMP pathway³⁴, so we tested each ligand for either agonism or antagonism using a CRE–luciferase assay³⁵. Several OX receptor antagonists, namely ACT 335827, TCS 1102, JNJ 10397049 and SB 674042, and the NPS receptor antagonist SHA-68, showed promise as GPR37L1 inverse agonists, reducing GPR37L1-mediated CRE–luciferase levels (Fig. 5a).

Each of the five ligands demonstrated concentration-dependent inhibition of GPR37L1 constitutive activity, albeit with low potency. However, their chemical compositions (Fig. 5b) suggested the possibility of non-specific inhibition through colloidal aggregation³⁶. Thus, we used both dynamic light scattering and centrifugation to further confirm the specificity of each ligand. SHA-68, JNJ 10397049 and SB 674042 had similar light scattering properties to that of buffer alone (Supplementary Fig. 5a), while TCS 1102 and ACT 335827 were visible only in the region where the positive aggregator control, quercetin³⁶, was observed (Supplementary Fig. 5b). The distinction between aggregating and non-aggregating ligands was further validated biologically by examining their half-maximal inhibitory concentration (IC₅₀) values either with or without

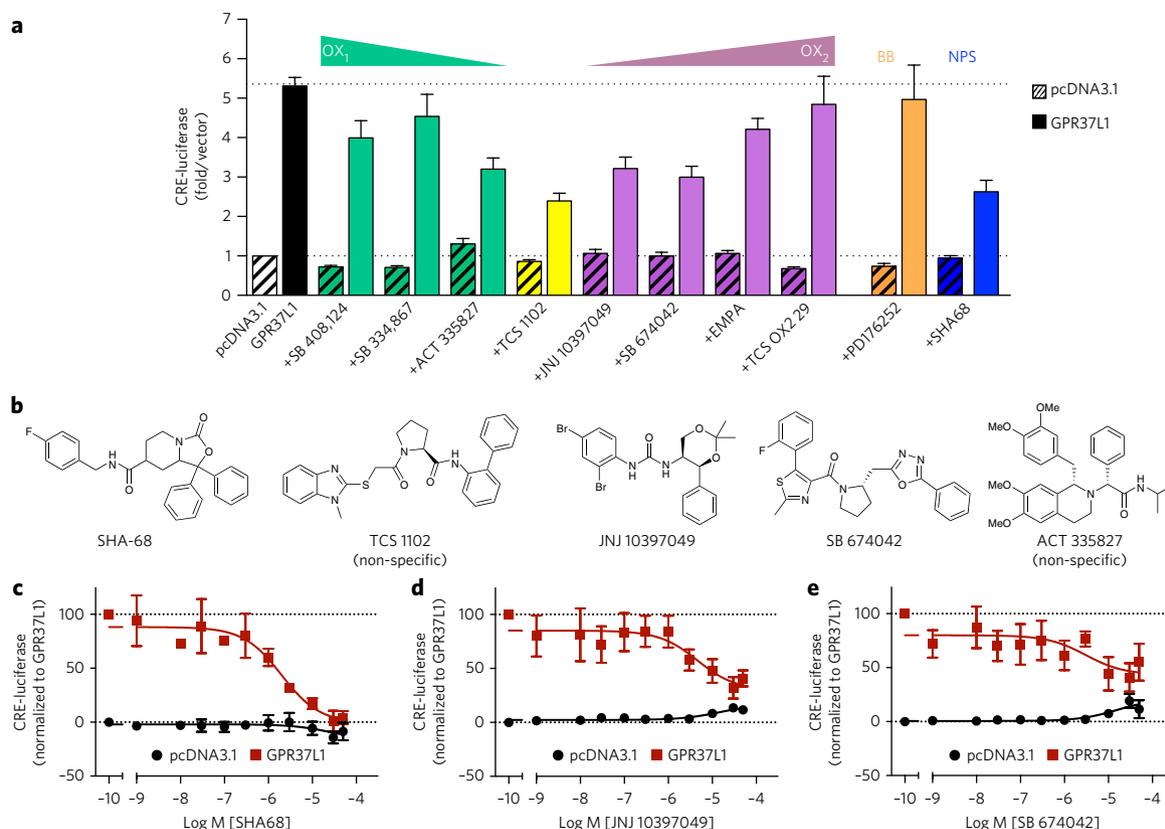


Figure 5 | Screening for GPR37L1 surrogate ligands. HEK293 cells were transiently transfected with pcDNA3 or GPR37L1 and a cAMP response element luciferase reporter (CRE-luciferase) 24 h before the addition of ligands. **(a)** Initial single concentration screen of various bombesin (BB), orexin (OX) and neuropeptide S receptor (NPS) antagonists, with triangles indicating increasing or decreasing selectivity for one OX receptor subtype over the other. $n = 3$ independent biological replicates. **(b)** Chemical structures of the five ligands that reduced GPR37L1-mediated CRE-luciferase levels. Concentration-response curves for SHA-68, $n = 3$, **(c)**, JNJ 10397049, $n = 4$ **(d)** and SB 674042, $n = 4$ **(e)**, following confirmation of GPR37L1 specificity by ligand aggregation counter screening. The mean and s.e.m. is reported for each point.

centrifugation (**Supplementary Fig. 5c**). Through these control experiments, three compounds were confirmed to be specific for GPR37L1; SHA-68 and JNJ 10397049 were efficacious inverse agonists (SHA-68 $pIC_{50} = 5.57 \pm 0.23$, $I_{max} = 100\%$, $n = 3$; JNJ 10397049 $pIC_{50} = 5.34 \pm 0.49$, $I_{max} = 70.3\%$, $n = 4$), while SB 674042 was a partial inverse agonist (**Fig. 5c**). TCS 1102 and ACT 335827 inverse agonist activity was attributed to a non-specific aggregation mechanism (**Supplementary Fig. 5d**). None of the reported compounds significantly affected vector-transfected cells, further indicating GPR37L1 specificity. These results highlight the utility of GPCR-CoINPocket, in which we rationally narrowed down our surrogate ligand search to 10 candidates, with a 30% success rate (3 out of 10 ligands). In further validation, we extended previous endogenous ligand screens^{31,32} to synthetic ET receptor ligands and a reported GPR37 ligand, head activator³⁷ (**Supplementary Fig. 6**), which did not yield any hits.

DISCUSSION

Here we describe GPCR-CoINPocket, a new measure of GPCR binding-site similarity. This metric, which is informed by structural data but can be applied to GPCRs without known structures, takes into account the expected relative importance of pocket residues for ligand binding and correctly predicts known pharmacological similarities of GPCRs without relying on the knowledge of their ligands. Using GPCR-CoINPocket, we characterized the pharmacological neighborhood of the orphan GPR37L1, and despite moderate numerical similarity scores, identified the first surrogate ligands for GPR37L1 by adopting the known ligands of its neighbors. We

predicted that the binding cavity of GPR37L1 is more closely related to the BB, OX and NPS receptors than to its nearest full-length sequence homologs, the ET receptors. This was confirmed experimentally by the discovery that two OX receptor antagonists and one NPS receptor antagonist were efficacious partial or full inverse agonists at GPR37L1.

Searching for endogenous ligands of orphan receptors is a daunting task. A more sensible approach is to identify surrogate ligands—small molecules that act at the orphan receptor but are not its cognate ligand; they can then be used as tools to probe both the pharmacology and physiology of the receptor. But the question is, where to start? Initially, one could simply test ligands from phylogenetically related GPCRs^{6,38}. However, the chemogenomic rearrangement of GPCRs demonstrates that pharmacological relatedness of GPCRs can deviate drastically from phylogenetic similarity¹¹. This work prompted us to revisit the comparison of GPCR binding sites. Previously, the general GPCR binding site defined by Gloriam *et al.* used a reference set of 43 different ligand-accessible residues within the TM binding region from the published crystal structures at the time (bovine and squid rhodopsin, ligand-free opsin, β_1 - and β_2 -adrenoceptors and the A_{2A} receptor)¹⁵. This did not lead to drastic changes in GPCR organization, in contrast to the chemogenomic approaches^{11,15}. For our binding-site definition, not only did we include new ligand-receptor interaction positions as observed in recent receptor X-ray structures, but we also used the crystallographically observed contact strengths to emphasize the relative importance of specific residue positions for receptor-ligand interactions. A hierarchical organization of GPCRs built on such similarities

recapitulates features from both the previous binding-site- and ligand-based dendrograms. These observations build confidence in the general applicability of GPCR–CoINPocket to GPCRs, including to orphan receptors.

Identification of a surrogate ligand may be the first step toward real deorphanization of a receptor in the sense of finding the cognate endogenous ligand. Previously, pharmacophores built from GPR139 surrogate agonists helped identify L-Trp and L-Phe amino acids as the endogenous ligands for this orphan GPCR^{39,40}. Alternatively, surrogate ligands themselves can provide a stepping stone toward drug discovery and understanding *in vivo* pharmacology, as recently exemplified for two orphan GPCRs, GPR68 and GPR65. For GPR68, the surrogate ligands identified were used for homology model optimization, followed by virtual screening, leading to identification of novel GPR68 chemical scaffolds⁴¹. This was followed by scaffold analog searches, resulting in the development of a potent GPR68 ligand with *in vivo* efficacy⁴⁰.

Despite the success of GPCR–CoINPocket in identifying pharmacological similarities of GPCRs, it could not predict all known relationships, for example, that of the Y₅ receptor with the CB₂ receptor¹¹ (Supplementary Data Set 1) and those from the ChEMBL benchmarking (Supplementary Fig. 3b), perhaps because of binding-mode diversity, which is overlooked when defining a consensus binding site. Currently, the segregation of binding sites or modalities is limited by the number of available GPCR crystal structures, as many more will be required to provide the combinatorial power that allows unambiguous identification of agonist, antagonist or allosteric binding contacts. A further limitation is the reliance of GPCR–CoINPocket on good TM sequence alignment. In fact, we have excluded contact sites within the ECL and N termini because of ambiguities in aligning these regions across the family, even though they represent strong ligand-contacting regions (Fig. 1). To mitigate these limitations, we included at least the first or last three residues of each TM to capture additional extracellular contacts. Still, we could not propagate contacts at 5×33, 7×24, 7×27 and 7×30 onto our alignment because of differing lengths of the TM from which these contacts were derived. Finally, because it is based on sequence alignment, GPCR–CoINPocket is limited to transfer of ligands from other GPCRs (rather than between different protein families). However, GPCR–CoINPocket should be easy to translate to other classes of GPCRs and other receptors, including non-mammalian targets.

In summary, surrogate ligands for orphan GPCRs are typically identified by costly *in vitro* high-throughput screening (HTS) campaigns, often the remit of large companies. However, with much of the ‘low hanging fruit’ now exhausted, novel knowledge-driven approaches are needed for orphan GPCR ligand discovery. One notable advance has been the move to computer-based, virtual ligand screening (VLS) using both 3D receptor and ligand chemical-field models. We envisage that GPCR–CoINPocket may complement HTS by guiding and narrowing the selection of potential surrogate ligands to test at an orphan receptor of interest. Here, we identified surrogate ligands for GPR37L1 as an example. The identified surrogates can then be used to refine and validate high-quality homology models to be used for VLS, with additional chemistry to follow, as well as to facilitate parallel *in vitro* and *in vivo* studies into orphan GPCR pharmacology and pathophysiology.

Received 28 February 2016; accepted 11 October 2016;
published online 19 December 2016

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

References

- 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**, 967–986 (2013).
- Southern, C. *et al.* Screening β -arrestin recruitment for the identification of natural ligands for orphan G-protein-coupled receptors. *J. Biomol. Screen.* **18**, 599–609 (2013).
- Jenkins, L. *et al.* Identification of novel species-selective agonists of the G-protein-coupled receptor GPR35 that promote recruitment of β -arrestin-2 and activate G α 13. *Biochem. J.* **432**, 451–459 (2010).
- Kroeze, W.K. *et al.* PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat. Struct. Mol. Biol.* **22**, 362–369 (2015).
- Ngo, T. *et al.* Identifying ligands at orphan GPCRs: current status using structure-based approaches. *Br. J. Pharmacol.* **173**, 2934–2951 (2016).
- Fredriksson, R., Lagerström, M.C., Lundin, L.G. & Schiöth, H.B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogue groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
- Schiöth, H.B. & Fredriksson, R. The GRAFS classification system of G-protein coupled receptors in comparative perspective. *Gen. Comp. Endocrinol.* **142**, 94–101 (2005).
- 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.* **50**, 6295–6298 (2007).
- Martin, R.E., Green, L.G., Guba, W., Kratochwil, N. & Christ, A. Discovery of the first nonpeptidic, small-molecule, highly selective somatostatin receptor subtype 5 antagonists: a chemogenomics approach. *J. Med. Chem.* **50**, 6291–6294 (2007).
- Scholten, D.J. *et al.* Pharmacological modulation of chemokine receptor function. *Br. J. Pharmacol.* **165**, 1617–1643 (2012).
- Lin, H., Sassano, M.F., Roth, B.L. & Shoichet, B.K. A pharmacological organization of G protein-coupled receptors. *Nat. Methods* **10**, 140–146 (2013).
- 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* **11**, 316 (2010).
- Surgand, J.S., Rodrigo, J., Kellenberger, E. & Rognan, D. A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins* **62**, 509–538 (2006).
- Jacoby, E. A novel chemogenomics knowledge-based ligand design strategy—application to G protein-coupled receptors. *Mol. Inform.* **20**, 115–123 (2001).
- Gloriam, D.E., Foord, S.M., Blaney, F.E. & Garland, S.L. Definition of the G protein-coupled receptor transmembrane bundle binding pocket and calculation of receptor similarities for drug design. *J. Med. Chem.* **52**, 4429–4442 (2009).
- Kufareva, I., Ilatovskiy, A.V. & Abagyan, R. Pocketome: an encyclopedia of small-molecule binding sites in 4D. *Nucleic Acids Res.* **40**, D535–D540 (2012).
- Venkatakrishnan, A.J. *et al.* Molecular signatures of G protein-coupled receptors. *Nature* **494**, 185–194 (2013).
- Isberg, V. *et al.* Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.* **36**, 22–31 (2015).
- Nandigam, R.K., Kim, S., Singh, J. & Chuaqui, C. Position specific interaction dependent scoring technique for virtual screening based on weighted protein–ligand interaction fingerprint profiles. *J. Chem. Inf. Model.* **49**, 1185–1192 (2009).
- Marcou, G. & Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **47**, 195–207 (2007).
- Sokal, R.R. & Michener, C.D. A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* **38**, 1409–1438 (1958).
- Keiser, M.J. *et al.* Predicting new molecular targets for known drugs. *Nature* **462**, 175–181 (2009).
- Modi, M.E. *et al.* Melanocortin receptor agonists facilitate oxytocin-dependent partner preference formation in the prairie vole. *Neuropsychopharmacology* **40**, 1856–1865 (2015).
- Bento, A.P. *et al.* The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* **42**, D1083–D1090 (2014).
- Kufareva, I., Katritch, V., Stevens, R.C. & Abagyan, R. Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. *Structure* **22**, 1120–1139 (2014).
- Kufareva, I., Rueda, M., Katritch, V., Stevens, R.C. & Abagyan, R. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* **19**, 1108–1126 (2011).
- Thelen, M. & Thelen, S. CXCR7, CXCR4 and CXCL12: an eccentric trio? *J. Neuroimmunol.* **198**, 9–13 (2008).

28. Min, K.D. *et al.* Identification of genes related to heart failure using global gene expression profiling of human failing myocardium. *Biochem. Biophys. Res. Commun.* **393**, 55–60 (2010).
29. Marazziti, D. *et al.* Precocious cerebellum development and improved motor functions in mice lacking the astrocyte cilium-, patched 1-associated Gpr3711 receptor. *Proc. Natl. Acad. Sci. USA* **110**, 16486–16491 (2013).
30. Smith, N.J. Drug discovery opportunities at the endothelin B receptor-related orphan G protein-coupled receptors, GPR37 and GPR37L1. *Front. Pharmacol.* **6**, 275 (2015).
31. Leng, N., Gu, G., Simerly, R.B. & Spindel, E.R. Molecular cloning and characterization of two putative G protein-coupled receptors which are highly expressed in the central nervous system. *Brain Res. Mol. Brain Res.* **69**, 73–83 (1999).
32. Valdenaire, O. *et al.* A new family of orphan G protein-coupled receptors predominantly expressed in the brain. *FEBS Lett.* **424**, 193–196 (1998).
33. Shihoya, W. *et al.* Activation mechanism of endothelin ET_B receptor by endothelin-1. *Nature* **537**, 363–368 (2016).
34. Coleman, J.L. *et al.* Metalloprotease cleavage of the N terminus of the orphan G protein-coupled receptor GPR37L1 reduces its constitutive activity. *Sci. Signal.* **9**, ra36 (2016).
35. Ngo, T., Coleman, J.L. & Smith, N.J. Using constitutive activity to define appropriate high-throughput screening assays for orphan G protein-coupled receptors. *Methods Mol. Biol.* **1272**, 91–106 (2015).
36. Sassano, M.F., Doak, A.K., Roth, B.L. & Shoichet, B.K. Colloidal aggregation causes inhibition of G protein-coupled receptors. *J. Med. Chem.* **56**, 2406–2414 (2013).
37. Rezzaoui, M. *et al.* The neuropeptide head activator is a high-affinity ligand for the orphan G-protein-coupled receptor GPR37. *J. Cell Sci.* **119**, 542–549 (2006).
38. Civelli, O. *et al.* G protein-coupled receptor deorphanizations. *Annu. Rev. Pharmacol. Toxicol.* **53**, 127–146 (2013).
39. Isberg, V. *et al.* Computer-aided discovery of aromatic L- α -amino acids as agonists of the orphan G protein-coupled receptor GPR139. *J. Chem. Inf. Model.* **54**, 1553–1557 (2014).
40. Liu, C. *et al.* GPR139, an orphan receptor highly enriched in the habenula and septum, is activated by the essential amino acids L-tryptophan and L-phenylalanine. *Mol. Pharmacol.* **88**, 911–925 (2015).
41. Huang, X.P. *et al.* Allosteric ligands for the pharmacologically dark receptors GPR68 and GPR65. *Nature* **527**, 477–483 (2015).
42. Tagat, J.R. *et al.* Piperazine-based CCR5 antagonists as HIV-1 inhibitors. I: 2(S)-methyl piperazine as a key pharmacophore element. *Bioorg. Med. Chem. Lett.* **11**, 2143–2146 (2001).

Acknowledgments

This work was supported in part by an Endeavour Research Fellowship from the Australian Government to T.N., an NHMRC & NHF CJ Martin Fellowship (N.J.S.), NHMRC Program Grants 573732 and 1074386 (R.M.G.), and Australian Postgraduate Awards (T.N. & J.L.J.C.). A.G.S. is supported by NHMRC Early Career Fellowship 1090408. R.A., I.K. and A.V.I. are supported by the NIH Grant R01 GM071872. I.K. is supported by R01 AI18985, R01 GM117424, R21 AI121918 and R21 AI122211. N.J.S. thanks the Mostyn Family Foundation for philanthropic support.

Author contributions

T.N., F.M.M., R.A., I.K. and N.J.S. conceived the study; T.N. and I.K. wrote the code for the weighted alignment scores with assistance from A.V.I. and R.A.; A.V.I. generated the ChEMBL data set, which was analyzed by T.N., A.V.I. and I.K.; A.G.S. performed dynamic light scattering experiments and analyzed the data; T.N. and N.J.S. designed and performed reporter gene assay experiments; R.M.G. and J.L.J.C. provided key reagents and ideas and R.P.R. provided the initial transmembrane alignments of all class A GPCRs; N.J.S. provided overall supervision of the project; T.N., I.K. & N.J.S. wrote the manuscript. All authors have provided input and approved the final version of the manuscript.

Competing financial interests

The authors declare competing financial interests: details accompany the [online version of the paper](#).

Additional information

Any supplementary information, chemical compound information and source data are available in the [online version of the paper](#). Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to N.J.S. or I.K.

ONLINE METHODS

Receptor nomenclature. All receptors have been named and abbreviated in accordance with the IUPHAR/BPS Concise Guide to PHARMACOLOGY⁴³.

Determination of ligand contact-strength profiles for Pocketome GPCRs. The Pocketome (<http://www.pocketome.org>) is an encyclopedia of annotated ligand binding site ensembles as seen in co-crystallized protein structures from the Protein Data Bank¹⁶. At the time of our study, the Pocketome contained 27 unique class A GPCRs (release 15.12) that were used to determine the relative roles of different residue positions for contact profiling. The entries ranged from having a single co-crystallized ligand to up to 10 unique ligands, with both small molecules and peptides represented (**Supplementary Data Set 2**).

For each Pocketome entry, ligand contact-strength fingerprints for the receptor residues were calculated as previously described^{26,44,45}. Briefly, for each pair of non-hydrogen ligand and receptor atoms separated by interatomic distance d , contact strength was assigned to 1 for $d < d_{\min} = 3.23 \text{ \AA}$, 0 for $d > d_{\max} = 4.63 \text{ \AA}$, and decreased from 1 to 0 as a linear function of d for $d_{\min} < d < d_{\max}$. The contacts were separately aggregated over the backbone (C α , C, O, N) and over side chain atoms of each residue where only side chain atoms contacts were considered in the final consensus fingerprint of each class A GPCR. For glycine residues, backbone contacts were treated as side chain.

The calculated side chain contact strengths were averaged over multiple occurrences of the same ligand, and multiplied by a weighting factor that ranged from 0 to 1 and represented the inherent flexibility of the residue as determined by the distribution of its observed r.m.s. deviation between pairs of structures within the entry. In entries with a single structure, the factor was automatically assigned to 1 for all residues.

Where a protein had more than one co-crystallized ligand, the unique ligands were clustered based on their residue fingerprints, clusters were ordered from largest to smallest, and residues making strong (>0.5) contacts with the ligands in the top 80% of the list were included in the consensus fingerprint. Residue contact-strength values were multiplied by their relative frequency in the ligand ensemble to arrive at the final fingerprint associated with the given Pocketome entry and thus with the given class A GPCR (**Fig. 1**).

Family-wide comparison of class A GPCRs using the calculated ligand contact-strength profiles. Pairwise comparison of projected binding pockets in class A GPCR sequences was performed as follows. Given a pairwise sequence alignment and a projected binding site positional fingerprint vector (one of the 27 fingerprints calculated from the Pocketome class A GPCR entries as described), a vector of pairwise per-residue similarities was first calculated using $S_{ij} = M_{ij} / \sqrt{(M_{ii} \times M_{jj})}$ where i and j are amino acids at the given position in the two sequences and M is the Gonnet residue comparison matrix⁴⁶. The similarity was then calculated as the sum of this vector multiplied, element-wise, by the binding site fingerprint vector of contact strengths. When calculated like this, the similarities of different sequences to themselves ranged from 14.19 to 35.18, depending on the receptor and the binding site fingerprint used, while the similarities between non-identical binding site sequences reached -19.67 at a minimum (for very dissimilar sequences and pockets) and 34.87 at a maximum (based on the binding pockets of US28 and LPA1 respectively; **Supplementary Data Set 3**).

The procedure was applied across an alignment of the transmembrane (TM) domains of 285 non-olfactory class A GPCRs, anchored around the highly conserved residues within each TM and guided by the GPCRdb numbering scheme¹⁸. Due to the different lengths of the TM domains observed in the crystal structures, contacts at residues numbered 5 \times 33, 7 \times 24, 7 \times 27 and 7 \times 30 (CXCR4, NTS1, PAR1, US28 contacts) could not be propagated. The obtained family-wide distribution of pairwise similarities varied for different binding site fingerprints, reflecting their size and the degree of overall conservation within the class A family; for example, the average pairwise similarity within the family was significantly higher for the fingerprints derived from the deep, well-conserved ligand-binding pocket of bRho than for the superficial P2Y₁. To avoid this kind of bias, the profiled similarity scores were standardized into Z-scores for each binding-site fingerprint (**Supplementary Fig. 7**). The final profiled similarity score between two class A GPCR sequences was calculated as the average of the Z-scores across the 27 GPCR binding site

profiles calculated from the Pocketome (values ranged from -2.89 to 5.47). For pairwise comparison between class A GPCR TM domains and a generic GPCR binding site definition, residue positions were projected onto the alignment and sequence similarities were calculated using the Gonnet residue comparison matrix⁴⁶, without an additional multiplication of a binding site fingerprint vector.

The similarity scores were converted into distances and used to cluster all class A GPCRs using the unweighted pair group method with arithmetic mean (UPGMA) algorithm²¹, implemented in ICM⁴⁷ (Molsoft LLC, La Jolla, CA). Unrooted dendrogram representation was generated using the Analysis of Phylogenetics and Evolution (ape) package⁴⁸ in R and nodes were expanded in Adobe Illustrator for clarity.

ChEMBL data set generation and benchmarking. Ligand sets for human class A GPCRs were extracted from the ChEMBL database²³, (release 21), with the following filters: activities from binding assays with single protein targets with a high confidence score (9), without known validity issues (as specified in data_validity_comment field), that are not duplicates and have a pChEMBL value of 6 or better. After receptors without at least 10 distinct ligands were discarded, there were 163 unique class A GPCRs.

For ligand set comparison, the chemical distance d between two ligands was calculated as the Tanimoto distance between chemical fingerprints as implemented in ICM. Similarity of ligand set A to ligand set B (S_{AB}) was defined as the weighted sum of chemical distances between ligands of set A and most similar ligands from set B, using a weight function $\exp(-200 \cdot d^4)$ as previously described⁴⁹. Two receptors were considered to have common ligands if both S_{AB} and S_{BA} were ≥ 20 (1,125 pairs), and to be dissimilar if S_{AB} and $S_{BA} < 1$ (8,520 pairs).

TM domain similarity, unweighted binding site sequence similarity and GPCR-CoINPocket were assessed for their ability to discriminate those receptor pairs that share ChEMBL ligands among the remaining pairs and its subset with TM similarity from 20 to 35% (4,796 pairs). Performance was quantified using the area under the receiver operating characteristic (ROC AUC).

Materials. HEK293 cells were freshly obtained from the ATCC (ATCC CRL-1573) and have been authenticated and certified mycoplasma free by the ATCC. Tissue culture reagents were purchased from Life Technologies (Carlsbad, CA). Compounds were purchased from various suppliers, as listed: ACT 335827, BQ-123, CI 1020, EMPA, JNJ 10397049, PD 176252, SB 674042, SHA 68 and TCS 1102 (Tocris Bioscience, Bristol, UK), Ambrisentan, SB 334867 and Zibotentan (AdooQ Bioscience, Irvine, CA), Bosentan (Cayman Chemical, Ann Harbor, MI), Head Activator (Bachem, Bubendorf, CH), SB 408124 (Selleck Chemicals, Houston, TX) and TCS OX2 29 (Abcam, Cambridge, UK). All compounds were >98% pure as determined by the manufacturer and were resuspended in DMSO at stock concentration of 10 mM and stored at -20 °C until required.

Luciferase reporter gene assay. We have previously described the capacity of GPR37L1 to constitutively signal downstream of G α_x using cAMP-response element (CRE) luciferase reporter gene assays³⁴; this allows the detection of both agonism and inverse agonism. The pcDNA3-GPR37L1 construct was generated as described previously and verified by DNA sequencing³⁴. To assess ligand activity at GPR37L1, transient transfection of HEK293 cells and luciferase reporter gene assays were performed using the Dual-Luciferase Reporter Assay System (Promega, Madison, WI) as previously described³⁵. Compounds were initially tested at a single concentration of 10 μ M with a maximum concentration of 1% DMSO. Compounds that substantially affected receptor basal activity at this concentration were validated with concentration-response curves. Samples were processed as per manufacturer's instructions and luciferase activity detected using a BMG PHERAstar FS.

Counter screen for non-specific receptor inhibition. To rule out colloidal aggregation as the mechanism for non-specific GPR37L1 inhibition, dynamic light scattering and centrifugation treatment experiments were performed as previously described³⁶. Briefly, from concentrated 10 mM stocks, compounds were diluted into serum-free DMEM to 50 μ M (0.5% DMSO final

concentration). Samples were run in quadruplicate and averaged for analysis (Supplementary Fig. 5). All measurements were made at room temperature using a DynaPro NanoStar (Wyatt Technology, Santa Barbara, CA) with a laser wavelength of 658 nm. The laser power was 100%, and the detector angle was 90°. To validate the formation of ligand aggregates, we also performed centrifugation experiments where the ligands were spun at 13,000g for 20 min and the resulting supernatant was used to stimulate the cells.

Data analysis and reporting. All molecular modeling and sequence analysis was carried out using ICM⁴⁷. All experimental data were analyzed in GraphPad Prism 6 (GraphPad Software, Inc., La Jolla, CA) and are expressed as mean \pm s.e.m., unless otherwise stated. Data was normalized to vector alone (0%) and GPR37L1 basal response (100%) to detect non-specific effects. Concentration–response curves were fitted to three-parameter nonlinear regression (fixed Hill slope). Based upon our previous experience, we performed all reporter assays a minimum of three separate times, in triplicate. Neither randomization nor blind studies were used.

Code availability. The contact-strength calculation utility (Ilatovskiy *et al.*, under review) is available online at <http://www.pocketome.org/basilico>. The code

for calculation of the weighted alignment scores (CoINPocket) is written in ICM and is available from the authors upon request.

Data availability. The source data to generate the dendrograms and heat maps is available in **Supplementary Data Set 1**.

43. Alexander, S.P. *et al.* The concise guide to PHARMACOLOGY 2015/16: G protein–coupled receptors. *Br. J. Pharmacol.* **172**, 5744–5869 (2015).
44. Kufareva, I. & Abagyan, R. in *Homology Modelling: Methods and Protocols* Vol. 857 (eds Orry, A.J.W. & Abagyan, R.) Ch. 10 (Humana Press, 2012).
45. McRobb, F.M., Sahagún, V., Kufareva, I. & Abagyan, R. In silico analysis of the conservation of human toxicity and endocrine disruption targets in aquatic species. *Environ. Sci. Technol.* **48**, 1964–1972 (2014).
46. Gonnet, G.H., Cohen, M.A. & Benner, S.A. Exhaustive matching of the entire protein sequence database. *Science* **256**, 1443–1445 (1992).
47. Abagyan, R., Totrov, M. & Kuznetsov, D. ICM—A new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **15**, 488–506 (1994).
48. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
49. Abagyan, R., Chen, W. & Kufareva, I. in *Computational Approaches to Nuclear Receptors* (eds. Cozzini, P. & Kellogg, G.E.) 84–109 (The Royal Society of Chemistry, 2012).