Trends in Pharmacological Sciences

Review



Chemical Diversity in the G Protein-Coupled Receptor Superfamily

Márton Vass,^{1,8} Albert J. Kooistra,^{1,8} Dehua Yang,² Raymond C. Stevens,^{3,4,5,*} Ming-Wei Wang,^{2,4,6,7,*} and Chris de Graaf^{1,*}

G protein-coupled receptors (GPCRs) are the largest family of cell signaling transmembrane proteins that can be modulated by a plethora of chemical compounds. Systematic cheminformatics analysis of structurally and pharma-cologically characterized GPCR ligands shows that cocrystallized GPCR ligands cover a significant part of chemical ligand space, despite their limited number. Many GPCR ligands and substructures interact with multiple receptors, providing a basis for polypharmacological ligand design. Experimentally determined GPCR structures represent a variety of binding sites and receptor-ligand interactions that can be translated to chemically similar ligands for which structural data are lacking. This integration of structural, pharmacological, and chemical information on GPCR-ligand interactions enables the extension of the structural GPCR-ligand interactome and the structure-based design of novel modulators of GPCR function.

Pharmacological and Structural G Protein-Coupled Receptor-Ligand Interactome

G protein-coupled receptors (GPCRs) (see Glossary) are the largest family of transmembrane proteins and they play several essential roles in cell signaling [1–3]. The function of GPCRs can be modulated by a plethora of chemical compounds, ranging from large proteins and peptides to small molecule ligands with a variety of physiochemical properties [4]. Due to their affinity for a diversity of chemical modulators, as well as their important roles in cellular processes and disease pathologies, GPCRs represent the most important class of drug targets, accounting for 12% of all human protein drug targets and the therapeutic effects of 33% of small molecule drugs [1,5]. In fact, many drugs (e.g., clozapine), while originally designed to interact with only one protein target, have been retrospectively shown to exert their clinical action by modulating several GPCR proteins, exhibiting a specific polypharmacological profile [6]. Bioactivity data of more than one million unique GPCR-ligand combinations for 321 GPCRs, and recent breakthroughs in GPCR structural biology [7] have so far resulted in 229 high-resolution structures of 136 distinct GPCR-ligand complexes of 48 different GPCRs (Table 1 and Figure 1, Key Figure). These structures provide opportunities for structure-based drug design [8,9]. However, despite the rapid accumulation of chemical, biological, and structural data, rational design of new molecules with well-defined GPCR pharmacological and selectivity profiles is still challenging for several reasons. Firstly, GPCR selectivity data are still limited compared with the high number of reported GPCR ligands. For example, 96 287 distinct active ligands (Ki/Kp/ IC₅₀/EC₅₀ <1 µM) have been reported for 321 of 419 nonolfactory GPCRs in ChEMBL [3,10] (Box 1 and Table 1), but bioactivity data have been reported for only 3% of the 31 million GPCRligand combinations covered by this GPCR and ligand data set. Secondly, the structures of only 11% of nonolfactory GPCRs (representing 41% of current GPCR drug targets) have been solved, and for most crystallized receptors only one or two different GPCR-ligand complexes

Highlights

The structural coverage of chemical GPCR ligand space can be efficiently extended by systematic structural cheminformatics analyses.

The integration of GPCR bioactivity data, structural receptor–ligand interaction analyses, and chemical similarity searches yields a comparative map of the pharmacological and structural GPCR–ligand interactome.

The systematic analysis of multitarget ligands and ligand substructures as well as receptor binding site similarities provides insights into opportunities and limitations for the identification and design of polypharmacological GPCR ligands.

The combination of GPCR bioactivity data mining, structural receptor–ligand interaction analyses, molecular similarity assessments, and chemical scaffold evaluations offers building blocks for the construction of a structurebased GPCR medicinal chemistry toolbox.

¹Division of Medicinal Chemistry, Faculty of Sciences, Amsterdam Institute for Molecules, Medicines and Systems (AIMMS), Vrije Universiteit Amsterdam, De Boelelaan 1108, Amsterdam 1081 HZ, The Netherlands ²The National Center for Drug Screening and CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 189 Guo Shou Jing Road, Pudong, Shanghai 201203, China

³iHuman Institute, ShanghaiTech University, 393 Hua Xia Zhong Road, Shanghai 201210, China
⁴School of Life Science and

Trends in Pharmacological Sciences



have been reported, limiting the possibilities for structure-based drug design acting on novel or specific combinations of GPCRs. In addition, it should be noted that bioactivity data are limited to less than 10 molecules for more than 40% of human GPCRs. Therefore, the design of polypharmacological drug molecules and the identification of chemical modulators of understudied orphan GPCRs requires a better understanding and exploitation of the chemical space of ligands that are compatible with the molecular interaction features of specific (combinations of) GPCR binding sites. However, structural chemogenomics analyses, based on the assumption that similar ligands interact with similar binding sites (and vice versa), can provide important insights for potential extrapolation of bioactivity and structural data to other protein–ligand complexes. The current analysis aims to provide a structure-based medicinal chemistry and chemical biology perspective of the pharmacological and structural GPCR–ligand interactome, covering all currently available structural, pharmacological, and chemical information on GPCR–ligand interactions.

Coverage of Crystallized G Protein-Coupled Receptor Ligand Chemical Space

In order to assess the impact of the increasing amount of GPCR structure data on GPCR-wide knowledge of receptor–ligand binding modes, the chemical structures of all 111 small molecule ligands bound to GPCR crystal structures were compared with all known small molecule ligands in ChEMBL (96 287 GPCR ligands with $K_i/K_D/IC_{50}/EC_{50} \leq 1 \ \mu$ M, of which 51 665 cover crystallized GPCRs, Box 1). The combined bioactivity data mining and **chemical similarity** analyses, described in Box 2 and presented in Table 1 and Figure 1A–C, show that:

- (i) 4523 (9%) of the 51 665 ligands of crystallized GPCRs are similar to the already crystallized GPCR ligands, considering each crystallized receptor individually;
- (ii) 7655 (15%) of the 51 665 ligands of crystallized GPCRs are similar to the already crystallized GPCR ligands, considering all crystallized receptors altogether;
- (iii) 7866 (8%) of the 96 287 known small molecule ligands of all 419 nonolfactory GPCRs are similar to any cocrystallized GPCR ligand.

This analysis indicates that the structural information from currently known GPCR-ligand complexes can be extended to cover a similar number of GPCRs with yet unknown structures. However, for most GPCR ligands, structural receptor interaction information is still limited and the modeling of these ligands in complex with their GPCR targets remains challenging. In particular, the cocrystallized ligands of purinergic P2Y₁ (58% coverage, Table 1) [11], chemokine CCR9 (52%) [12], protease-activated PAR1 (41%) [13], adenosine A1/2A (27%), angiotensin AT₁ (23%) [11], δ-opioid (23%) [14–16], β_{1/2} adrenergic (17%), endothelin ET_B (20%) [17], and free fatty acid FFA1 (19%) [18,19] receptors cover a relatively large part of the chemical ligand space of the corresponding receptors (Figure 1C). For these receptors, the binding mode of a relatively large number of ligands can be confidently predicted using computational techniques such as molecular docking and ligand shape-based alignments, although small differences in ligand structure may affect the overall binding mode [20,21]. By contrast, the cocrystallized small molecule ligands of the dopamine $D_{3/4}$ [22], histamine H₁ [23], muscarinic M_{1/3/4} [24–26], serotonin 5-HT_{1B/2B} [27–31], complement peptide C5a₁ [32], chemokine CCR2 [33], lysophosphatidic acid LPA₁ [34], sphingosine-1-phosphate S1P₁ [35], leukotriene BLT₁ [36], protease-activated PAR2 [37], corticotropin-releasing factor CRF₁ [37,38], and glucagonlike peptide-1 GLP-1 [39] receptors are similar to only small portions of the ligands of the respective receptors (<3% Figure 1C). There are several GPCRs for which an X-ray or electron microscopy structure has only been solved bound to larger peptide-like ligands (apelin [40], chemokine US28 [41], and calcitonin-like [19,42,43] receptors) or without any ligand (LPA₆) [44]. Clearly, cocrystallization of these receptors with chemically diverse ligands would greatly

Technology, ShanghaiTech University, 393 Hua Xia Zhong Road, Pudong, Shanghai 201210, China ⁵Bridge Institute, Department of Chemistry and Biology, University of Southern California, Los Angeles, CA 90089, USA ⁶University of Chinese Academy of Sciences, Number 19A Yuquan Road, Beijing 100049, China ⁷School of Pharmacy, Fudan University, 826 Zhangheng Road, Shanghai 201203, China ⁸M.V. and A.J.K. contributed equally to this review

*Correspondence: stevens@shanghaitech.edu.cn (R.C. Stevens), mwwang@simm.ac.cn (M.-W. Wang), and c.de.graaf@vu.nl (C. de Graaf).

Trends in Pharmacological Sciences



benefit the drug discovery efforts in these subfamilies. It should be noted that the cocrystallized ligands of several of these receptors share conserved substructures and/or a conserved shape/pharmacophore with larger portions of known ligands. For instance, in the 5-HT_{1B} receptor, the tryptamine substructure of the cocrystallized (dihydro-)ergotamine is present in 23% of known 5-HT_{1B} ligands [27]. Similarly, the phosphonic acid group of the cocrystallized ML056 [35] and carboxylic acid group of the cocrystallized ligands of LPA₁ [34] are present in 14% of S1P₁ and 58% of LPA₁ ligands, respectively. The antagonist CP-376395 [37,38] cocrystallized with CRF₁ shares perpendicularly oriented *N*-heterocyclic and hydrophobic aromatic rings with most CRF₁ ligands. The cocrystallized doxepin [23] in H₁, QNB [45] in M₂, and tiotropium [24–26] in M₁₋₄ ligands share an amine, with two aromatic rings oriented in a butterfly shape, with many other H₁ and M₁₋₄ ligands [46]. The crystallized ligands of C5a₁ [32], PAR2 [37], CCR2 [33], LPA₁ [34], BLT₁ [36], and GLP-1R [39], are chemically distinct from most known ligands for these receptors, and therefore can provide useful templates to extend novel chemical ligand space using structure-based drug discovery and design approaches.

About 70% of the currently known GPCR ligand space interacts with receptors for which no structures have been reported (Figure 1B). Several receptors that bind to large numbers of different ligands share high sequence similarity with crystallized receptors (53%-72% sequence identity), including dopamine D_2 (4758 ligands, similar to D_3), cannabinoid CB_2 (3687, similar to CB₁), serotonin 5-HT_{1A} (3271), 5-HT_{2A} (3113), and 5-HT_{2C} (2298) receptors (similar to 5-HT_{1B} and 5-HT_{2B}), and adenosine A_3 (2728, similar to A_1 and A_{2A}). The structural homologues and similar crystallized ligands provide structural and chemical templates to model and design receptor-ligand interactions of these GPCRs. Conversely, several receptors that bind to large numbers of different ligands share relatively low sequence similarity with any of the crystallized GPCRs (23%-36% sequence identity), including melanin-concentrating hormone MCH₁ (3068 known ligands), neurokinin NK₁ (2372), melanocortin MC₄ (2244), ghrelin GHSR (1272), and gonadotropin-releasing hormone GnRH₁ (1208) receptors (Figure 1B). Only 1.5%–4.5% of the ligands of these receptors are similar to cocrystallized GPCR ligands. The structures of these five peptide receptors would therefore create great potential for structure-based drug design in novel therapeutic areas [1], extending the coverage of structural GPCR ligand space by about 20%. A similarity analysis of small molecule ligands of noncrystallized receptors shows that structures of ligands of neuropeptide and prostanoid receptors would result in the largest extension of the currently known structural GPCR-ligand interactome (Figure 1C). In short, neuropeptide receptor familyspecific peptide ligands such as Bz-Arg-Phe-NH₂ are similar to ligands of several other peptide receptors, including NPFF_{1/2}, NPY_{1/4}, QRFP, C3a, NTS₂, FPR₁, NMU_{1/2}, and BKRB₂, covering 9% of the ligand space of these receptors. Laropiprant is chemically similar to ligands of the DP₁, EP₃, TP, and IP receptors, and covers 5% of the ligand space of these receptors. It should be noted that the number of publicly available GPCR structures and ligands, and thus the coverage of the ligand space, is constantly increasing. For example, in the past 2 years (from January 2016 to January 2018), the experimental structures of 52 new GPCR-ligand combinations have been deposited in the Protein Data Bank (62% increase), covering 27 different receptors, including 17 for which structures had not yet been solved (M1, M₄, D₄, AT₂, APJ, C5a₁, ET_B, OX₁, PAR2, CCR2, CCR9, BLT₁, LPA₆, CB₁, A₁, CT, and GLP-1R; 55% increase). To enable timely update and reuse of the data reported in Table 1 we have provided a transferrable KNIME workflowⁱ.

G Protein-Coupled Receptor Ligand Polypharmacology

The 'one drug-one target-one disease' paradigm has led to the discovery of many successful drugs. However, in recent years it has been increasingly recognized that complex diseases

Glossary

Allosteric modulator: a ligand that binds to a nonoverlapping site away from the orthosteric binding site, and either enhances (positive allosteric modulator, PAM) or reduces (negative allosteric modulator, NAM) the signaling of orthosteric agonists. **Chemical fingerprint:** a binary representation of molecules encoding the presence or absence of predefined groups (key-based fingerprint, e.g., MACCS keys) or enumerated substructures (hashed fingerprint, e.g., ECFP-4) in the formula.

Chemical similarity: the similarity of two chemical structures based on the presence and connectivity of their substructures, most often evaluated as the similarity of chemical fingerprints using the Tanimoto similarity metric. Cheminformatics: the use of computer and information technology to study chemical problems, including chemical data and compound management and analysis, compound enumeration, substructure and similarity searching, prediction of compound properties,

virtual screening, etc. Extracellular loops (ECL): the amino acid sequences connecting the seven transmembrane helices in GPCRs facing the space outside of the plasma membrane.

Extracellular vestibule (ECV): an allosteric binding site postulated for several GPCRs formed by the extracellular ends of the seven transmembrane helices and the three extracellular loops.

G protein-coupled receptor (GPCR): a superfamily of

transmembrane receptor that are activated by an extracellular ligand (light, ions, small molecules, lipids, peptides, or proteins), and transmit signals to the intracellular compartment via activation of intracellular pathways primarily coupling to G proteins but also to other mediators such as arrestins, kinases, or lipases.

 $K_{I}/K_{D}/IC_{50}/EC_{50}$: different endpoint measurements of a ligand's affinity or efficacy for its target, the concentration at which 50% of the receptor binding sites are occupied by the ligand from inhibition/ saturation experiments, or at which

Trends in Pharmacological Sciences



such as neurological disorders [48-50] or cancer [51] require synergistic modulation of several protein targets in a network of interacting proteins and signaling pathways [6]. As GPCRs are key players in many different disease-related pathways, elucidation of novel GPCR structures is indispensable for the rational design of drugs with a specific polypharmacological profile [1,6]. Structural knowledge of specific receptors will provide insight into the binding of known multitarget ligands and this in turn will enable the structure-based design of novel therapeutics with tailored effects. An analysis of all 96 287 GPCR ligands in ChEMBL with $K_i/K_D/IC_{50}/EC_{50}$ \leq 1 μ M reveals that 24 475 (25%) ligands are associated with at least two GPCR targets, 9271 (10%) with at least three targets, and 945 (1%) with five or more targets. Many more known ligands likely also have polypharmacological profiles, but only 3% of all possible GPCR-ligand combinations have been evaluated. An analysis of pair-wise target association data demonstrates that most information is available on aminergic receptor polypharmacology (Figure 2A). Serotonin receptors share the highest number of multitarget ligands with other receptor subtypes, including 1853 ligands with dopamine receptors (activity at both families $\leq 1 \mu M$). Other aminergic receptor subfamilies share 72-486 ligands between each pair. Aminergic receptors share 183 ligands with the chemokine and 43 ligands with the opioid receptor family (Figure 2B). Opioid receptors share 77 ligands with tachykinin receptors and 38 ligands with cholecystokinin receptors. The number of polypharmacological ligands that interact with different GPCR families (first two digits of the GPCRdb slug, Box 1 [52-55]), but that do not interact with aminergic receptors are scarce: only the chemokine-tachykinin and melatonin-melanocortin receptor subfamily pairs share more than five ligands. Figure 2C shows several druglike molecules, and diverse chemical structures can possess such polypharmacology profiles. For example, astemizole and thioridazine act on a variety of aminergic receptor subfamilies as well as μ/κ opioid receptors (astemizole and thioridazine) and the tachykinin NK₂ receptor (astemizole). Loperamide and naloxone are selective for the δ , κ , and μ opioid receptors, but also act on dopamine D₃ and adenosine A₃ receptors, respectively. Most of these ligands have been retrospectively shown to exert their therapeutic action in pain relief by modulating several GPCRs [50,56]. The increasing structural information on GPCRs provides an important resource for the prospective design of novel ligands with specific polypharmacological profiles [48]. Several protein structure-based ligand repurposing studies have been reported between different GPCR subfamilies (e.g., OX1/OX2, BB1/BB2, and NPS antagonists repurposed as GPR37L1 ligands [57]), and between GPCRs and other protein families (e.g., kinase p38 inhibitor repurposed as 5HT_{2A} ligand [58]). Furthermore, several ligand-based GPCR polypharmacology design studies have been reported for closely related receptors (e.g., aminergic receptors [59]). By contrast, structure-based GPCR ligand design studies have so far been limited to the design of molecules with increased receptor selectivity (e.g., dopamine D₃ vs. D₂ receptors [60]), rather than the design of ligands with well-defined polypharmacological profiles. The design of molecules that target unique combinations of structural interactions in specific sets of evolutionary nonrelated receptors (that are not present in more closely related off-target receptors) is highly challenging. The combination of structures and selectivity profiles of receptor-ligand complexes with different GPCR polypharmacology is required to provide more detailed insights into structure-selectivity relationships in order to guide future GPCR ligand design.

Chemistry of G Protein-Coupled Receptor-Ligand Interactions

Figures 3 and 4 show how similar chemical **scaffolds** of GPCR ligands can adopt conserved or distinct binding modes in different GPCR binding pockets. The increasing availability of information about structural GPCR–ligand interaction patterns will stimulate structure-based drug discovery [61]. Most ligands (95%) bind the ancestral GPCR binding cavity comprised of residues in the seven **transmembrane (TM) helices** and three **extracellular loops** (ECL),

50% of the downstream signal is inhibited/elicited.

Orthosteric pocket: the binding site of the natural endogenous agonists of receptors.

Polypharmacology: the rational design and use of a single drug molecule able to simultaneously and specifically interact with multiple biological targets, characterized by an improved efficacy when compared with a highly selective pharmacological agent.

Structure-activity relationship (SAR): the observed relationship

between the three-dimensional chemical structure of a molecule and its biological activity, enabling the rational inclusion of specific functional groups responsible for evoking a biological effect at the desired target.

Scaffold: the fixed part of a chemical series of bioactive compounds on which functional groups are added or exchanged. Transmembrane (TM) helix: the characteristic membrane spanning alpha helical domains of membraneembedded proteins, such as the seven transmembrane helices of GPCRs or the 6–24 transmembrane helices of ligand-gated ion channels.

Trends in Pharmacological Sciences



Table 1. Coverage of Chemical GPCR Ligand Space of 122 Small Molecule and Peptide Ligands in 229 GPCR Crystal Structures (Cyan, Orange, and Green Columns, Box 2) and GPCR Ligand Polypharmacology (Red Columns). Data in the table were calculated using the KNIME workflow described in Box 2ⁱ

Receptor ^a	Number of structures ^b	Number of ligands (X-ray) ^c	Number of ligands (rec.) ^d	Number of ligands (rec.) similar to ligands (rec. X-ray) ^e	Number of ligands (rec.) similar to ligands (GPCR X-ray) ^f	Number of ligands (other GPCR) similar to ligands (rec. X-ray) ⁹	Number of ligands active at other GPCRs ^h	Number of ligands active at other non-subfamily GPCRs ⁱ
All (46/419)	229	122 (11)	96287					
Class A (39/299)	204	104 (8)	89315					
Aminergic (10/37)	58	33	26515					
5-HT _{1B}	2	2	966	2	79	68	888	90
5-HT _{2B}	4	2	1122	15	44	72	966	270
M ₁	1	1	1312	10	69	19	930	136
M ₂	3	3	1336	48	80	102	974	102
M ₃	4	2	1384	26	137	25	855	149
M ₄	1	1	466	4	46	19	387	87
β1	18	12	888	110	211	724	789	74
β2	21	11	1099	237	324	258	783	168
D ₃	1	1	3436	23	302	83	2993	1206
D ₄	2	1	1982	55	217	191	1265	281
H ₁	1	1	963	5	10	4	600	565
Peptide (14/77)	39	24 (4)	31775					
AT ₁	2	2	656	151	203	65	223	29
AT ₂	3	2	346	37	64	188	195	1
APJ	1	1 (1)	76	-	0	-	0	0
C5a ₁	1	1	310	0	5	6	3	2
ETB	4	3 (1)	527	108	110	141	459	2
NTS ₁	8	1 (1)	133	-	0	-	53	0
δOR	4	2 (1)	2588	437	629	411	1685	103
кOR	2	2	2872	177	582	307	1896	56
μOR	2	2	2785	285	505	590	2385	70
NOP	3	3	1301	59	104	613	720	7
OX ₁	2	2	1354	115	216	144	1138	12
OX ₂	3	2	1823	133	318	171	1128	2
PAR1	1	1	547	224	225	1	1	1
PAR2	3	2	42	0	0	0	0	0
Protein (4/29)	12	9 (4)	6980					
CCR2	1	2	1346	29	57	86	81	9
CCR5	2	2 (1)	1743	113	160	35	75	28
CCR9	1	1	159	83	84	9	0	0
CXCR4	6	3 (2)	286	15	31	1	2	0
US28 (viral)	2	1 (1)	2	-	0	-	1	1

Trends in Pharmacological Sciences



Table 1. (continued)

Receptor ^a	Number of structures ^b	Number of ligands (X-ray) ^c	Number of ligands (rec.) ^d	Number of ligands (rec.) similar to ligands (rec. X-ray) ^e	Number of ligands (rec.) similar to ligands (GPCR X-ray) ^f	Number of ligands (other GPCR) similar to ligands (rec. X-ray) ⁹	Number of ligands active at other GPCRs ^h	Number of ligands active at other non-subfamily GPCRs ⁱ
Lipid (6/37)	14	12	14472					
FFA1	3	3	625	117	117	0	18	1
BLT ₁	1	1	287	0	1	0	0	0
LPA1	3	3	106	0	0	1	45	0
LPA ₆	1	0	11	-	0	-	11	0
S1P1	2	1	1739	43	43	46	580	26
CB ₁	4	4	2877	413	451	227	1221	19
Nucleotide (4/12)	42	23	7299					
A ₁	2	2	2605	102	734	212	2091	6
A _{2A}	35	16	3020	811	848	1540	2026	11
P2Y1	2	2	314	181	212	11	13	1
P2Y ₁₂	3	3	907	99	126	280	4	2
Sensory (1/7)	39	2	0					
Rhodopsin	39	2	0	0	0	2	0	0
Class B (4/21)	12	7 (3)	3598					
Calcitonin	1	0	10	-	0	-	0	0
CRF1	2	1	1687	5	27	0	29	1
GLP-1R	4	4 (2)	303	1	1	34	55	0
GCGR	5	3 (1)	828	115	135	22	117	4
Class C (2/23)	4	4	2928					
mGluR ₁	1	1	535	18	21	3	102	1
$mGluR_5$	3	3	1481	74	76	16	82	1
Class F (1/11)	9	7	480					
Smoothened	9	7	480	46	53	346	3	3

See color coding also in Box 2 and Figures 1 and 2.

^aReceptors classified according to GPCR database [52-55].

^bX-ray and electron microscopy structures.

^cUnique cocrystallized ligands; peptide ligands (molecular weight > 800 Da) indicated in brackets.

^dNumber of unique ligands in ChEMBL with binding affinity (K/IC₅₀) and/or functional potency (EC₅₀/IC₅₀) of at least 1 µM.

^eNumber of ligands of the specified receptor (rec.) that is chemically similar to any cocrystallized ligand of the same receptor (blue column). Molecules are determined to be chemically similar if their ECFP-4 Tanimoto similarity is ≥ 0.4 or MACCS Tanimoto similarity is ≥ 0.8 (Box 2).

^fNumber of ligands of the specified receptor (rec.) that is chemically similar to any cocrystallized ligand of any GPCR (orange column).

⁹Number of ligands of other GPCRs chemically similar to any cocrystallized ligand of the specified receptor (rec.) (green column).

^hNumber of ligands in ChEMBL with binding affinity (K₁/IC₅₀) and/or functional potency (EC₅₀/IC₅₀) of at least 1 µM at the specified receptor and another GPCR (red). Number of ligands in ChEMBL with binding affinity (K₁/IC₅₀) and/or functional potency (EC₅₀/IC₅₀) of at least 1 µM at the specified and another non-subfamily GPCR (red).

that is accessible from the extracellular side of the membrane or from the membrane bilayer (such as for lipid receptors). This ancestral binding cavity represents the **orthosteric** pocket in class A and B1 GPCRs and an allosteric pocket in class C GPCRs [62]. The ancestral binding cavity can be divided into a minor pocket (between TM helices 1, 2, 3, and 7), major pocket (between TM helices 3, 4, 5, 6, and 7), and an **extracellular vestibule** (ECV; between the N terminus, the ECLs, and the top of TM helices 1–7) [62,63]. Endogenous small molecule ligands

Trends in Pharmacological Sciences



Key Figure

Structural Coverage of Chemical G Protein-Coupled Receptor Ligand Space



Trends in Pharmacological Sciences

Figure 1. (A) G protein-coupled receptor (GPCR) phylogenetic tree with GPCRs color-coded according to number of unique complexes from X-ray and electron microscopy structures. (B) GPCR phylogenetic tree, color-coded according to number of unique active ligand counts in ChEMBL. (C) GPCR phylogenetic tree with ECFP-4/MACCS keys fingerprint similarity to any cocrystallized ligand (Box 2), color-coded according to the percentage of ligands similar to any cocrystallized ligand. Cocrystallized GPCR ligands with largest coverage of GPCR ligand space indicated in blue insets (similarity of active receptor ligands to own cocrystallized ligands). Cocrystallized GPCR ligand classes with a large coverage of the ligand space of related GPCRs indicated in orange insets (similarity of active receptor ligands to all GPCR cocrystallized ligands). Noncrystallized GPCR ligands that would contribute most to the coverage of the GPCR phylogenetic tree indicated in grey insets (similarity of all GPCR ligands to noncrystallized ligands) according to ECFP-4/MACCS keys fingerprint similarity, see Box 2.

Trends in Pharmacological Sciences



Box 1. Resources for Pharmacological and Structural GPCR-Ligand Data Mining

Different resources were used to collect, analyze, and integrate GPCR–ligand interaction data presented in this review, including the ChEMBL (https://www.ebi.ac.uk/chembl/) [10] and GPCRdb (http://gpcrdb.org/) [53,55] databases, and 3D-e-Chem (http://dd-e-chem.github.io/) [52,54], MACCS [109] (RDKit implementation [110]), and ECFP-4 [111] (CDK implementation [112]) **cheminformatics** tools integrated in KNIME analytical platform [113]. ChEMBL is an open large-scale bioactivity database performing automatic extraction of data from medicinal chemistry literature and manual curation and providing a web interface as well as web services for data mining [10]. GPCRdb is an open specialized database providing reference data and analysis tools related to GPCRs, including classification, sequence, structural, mutation, ligand, protein–ligand interaction data, structural models, alignments, customizable structure-sequence data mapping, and construct design [53,55]. The 3D-e-Chem cheminformatics tools facilitate the integrated analysis of all these data in the graphical environment of KNIME [113] and allow the creation of customizable and transferrable ligand- and structure-based data analysis workflows [52,54]. Methodological details of the chemical similarity search of GPCR ligand space performed in the current work are provided in Box 2.

of class A receptors predominantly bind in the major pocket (e.g., retinal in rhodopsin, adrenaline in β_2 [64], adenosine in A_{2A} [65]), or at the interface of the major and minor pockets (e.g., natural ligand analogs 2MeSATP in P2Y12 [66], ML056 in S1P1 [35], and AM841 and AM11542 in CB₁ [67]). None of the currently reported GPCR crystal structures contain an endogenous ligand bound to the ECV, but combined structural, mutation, and modeling studies suggest that leukotriene B_4 , for example, binds in the ECV of the BLT₁ receptor [36]. Endogenous peptide and protein ligands such as apelin, endothelin, neurotensin, and chemokines occupy the major and minor pockets and the ECV simultaneously. Most cocrystallized ligands primarily target the major binding pocket in 54 aminergic, 37 adenosine, and 11 opioid receptor crystal structures. In some GPCR crystal structures more than 30% of ligandreceptor contacts involve both major and minor binding pockets, such as for the CCR5 antagonist maraviroc [68] (Figure 3) and the AT1/2 antagonists, olmesartan and ZD7155 [11] (Figure 4). In only a few structures is the minor pocket primarily involved in ligand recognition, including IT1t [69] (CXCR4, Figure 3) and AZ8838 [37] (PAR2). Several cocrystallized ligands that predominantly bind the major/minor TM binding pocket also target the ECV, including adenosine A1/2A [47,70–74] and 5-HT1B/2B [27–29,31] ligands. Less than a handful of cocrystallized small molecule ligands mainly target the ECV, including LY2119620 [64] [negative allosteric modulator (NAM) of the muscarinic M₂ receptor] and MRS2500 [11] (P2Y₁ receptor agonist).

In 46 of the 136 crystallized GPCR-ligand complexes, the interaction between cationic, basic amines, and negatively charged residues play an important role (Figure 3). The conserved anionic residues [D^{3.32} in aminergic (5-HT_{1B/2B}, $\beta_{1/2}$, D_{3/4}, H₁, M₁₋₄) and opioid (δ OR, μ OR, κOR, NOP) receptor structures and E^{7.39} in chemokine receptor (CCR2, CCR5, CXCR4) structures] are located at the interface of both minor and major binding pockets and play important roles in ligand binding to these receptor subfamilies [46,75,76]. Most of the aminergic (86%), opioid (90%), and chemokine (74%) receptor ligands contain a cationic group, including 28 aminergic, 9 opioid, and 19 chemokine receptors for which no crystal structure is currently available. Furthermore, 30 peptide receptors predominantly bind basic ligands (more than 50% of their known ligands are positively charged, see Table S1 in the Supplemental Information online). Structure-activity relationship (SAR) studies indicate that these cationic groups are essential for receptor binding and activity [46,75,76]. Ligand similarity assessments (Table 1) and mutation studies [46,75,76] suggest that many ligands of aminergic, opioid, chemokine, and peptide binding receptors likely share the conserved ionic interaction features observed in the available crystal structures (Figure 2), providing a structural basis for the polypharmacological action of ligand at combinations of these receptors. The basic amine moieties of NAM LY2119620 and the antagonist AZ8838 target negatively charged residues in the ECVs of muscarinic M₂ (E^{45.46}) and PAR2 (D^{45.52}), respectively [37,64]. Interestingly, the BIIL260 inverse

Trends in Pharmacological Sciences



Box 2. Chemical Similarity Search of GPCR Ligand Space

A combination of the complementary molecular fingerprint methods, MACCS [109] (RDKit [110]) and ECFP-4 [111] (CDK [112]) was used to assess the similarity of 96 287 distinct ligands ($K_{/}K_{D}/IC_{50}/EC_{50} \le 1 \mu$ M) of 321 nonolfactory GPCRs in ChEMBL [10] and 111 small molecule cocrystallized ligands in X-ray crystal structures of 48 different GPCRs (see Table 1 and Figure 1 in main text). Established similarity Tanimoto coefficient (T_c) cutoffs [114] were used to determine whether molecules are chemically similar (ECFP-4 $T_c \ge 0.4$ or MACCS $T_c \ge 0.8$) and the following were assessed:

- (i) the number of ligands of a specified crystallized GPCR similar to its cocrystallized ligands (column 5 of Table 1, cyan Venn diagram in Figure I). The modeling of these GPCRs in complex with molecules similar to cocrystallized ligands should be feasible using state-of-the-art modeling tools assisted by experimental information (e.g., SAR and mutation data).
- (ii) the number of ligands of a specified crystallized GPCR similar to cocrystallized ligands of any GPCR (Table 1 column 6, orange Venn diagram in Figure I). This analysis indicates to which extent structural GPCR-ligand interactions for a specific receptor can be derived from the accumulated information of all experimentally determined GPCR structures.
- (iii) the number of ligands of other GPCRs that are similar to the cocrystallized ligands of a specified receptor (Table 1 column 7, green Venn diagram in Figure I). This assessment shows to which extent structural protein–ligand interaction information of a specified receptor can potentially be transferred to other receptors.
- (iv) the number of ligands of other GPCR subfamilies that are similar to the cocrystallized ligands of a specified receptor (Table 1 columns 8–9). This analysis shows potential cross-receptor subfamily polypharmacology of the cocrystallized ligands of a specified receptor. Receptor subfamilies were defined based on the first nine digits of the GPCRdb slug for GPCRs [52–55], consistent with Guide to Pharmacology [3].
- (v) the number of ligands that are similar to cocrystallized allosteric GPCR ligands using an iterative similarity ensemble search approach (see Figure 6 in main text). In this approach a combination of higher ECFP-4 ($T_c \ge 0.6$) and MACCS ($T_c \ge 0.9$) cutoff Tanimoto values were used in consecutive rounds of similarity searches against increasing sets of chemically similar ligands until no new molecules were identified. This analysis provides an overview of ensembles of ligands that are likely to bind the same allosteric binding sites as cocrystallized GPCR ligands.

The KNIME workflow to perform assessments (i)–(iv) and to calculate the values in Table 1 is shown in Supplemental Information Figure S2, and is available online (see Resources)¹.



Trends in Pharmacological Sciences

Figure I. Chemical Similarity Search of GPCR Ligand Space. Schematic example of chemical fingerprint-based similarity search using co-crystallised adenosine A_{2A} receptor ligands, for example, ZM241385 [47]. Fifteen of the co-crystallised A_{2A} ligands are chemically similar to 811 A_{2A} ligands (cyan Venn diagram), 24 of the co-crystallised GPCR ligands are similar to 848 A_{2A} ligands (orange Venn diagram), and 15 of the co-crystallised A_{2A} ligands are similar to 1540 GPCR ligands (green Venn diagram).

agonist mimics the sodium ion-centered water cluster with the conserved $D^{2.50}$ inside the seven-TM domain of BLT₁ [36]. The crystal structures represent a variety of different basic moieties that play a role in receptor binding (Figure 3), including:

- (i) linear aliphatic primary (β₂, δOR), secondary (β_{1/2}, SMO), tertiary (H₁, CCR2), and quaternary (M₂) amines;
- (ii) monocyclic amines, including pyrrolidines (D_{3/4}, NOP), piperidines (NOP), piperazines (β_1);
- (iii) multicyclic, bridged amine ring structures, including azabicyclo-octane (CCR5) azabicy-
- clo-octanium (M₁₋₄), quinuclidine (M₂), and multicyclic opioids (δ OR, μ OR);
- (iv) ergolines (5HT $_{1B/2B}$);

Trends in Pharmacological Sciences





Trends in Pharmacological Sciences

Figure 2. Polypharmacology of G Protein-Coupled Receptor Ligands. (A) Flare plot (https://github.com/GPCRviz/flareplot) of G protein-coupled receptors (GPCRs) with shared ligands (see Table 1 columns 8 and 9). GPCRs are organized according to GPCR database (db) [52–55] around the circle and a line is drawn between pairs of GPCRs if they share at least five active ligands in ChEMBL (pActivity \geq 6). The line width is proportional to the number of shared ligands, and families shown in panel B are highlighted. (B) Active ligand overlap of selected receptor pairs. (C) Representative ligands with polypharmacological profiles from panel B. (D) Receptor affinity profiles of known polypharmacological drugs acting on opioid receptors. Structures of selected ligands (indicated by stars) shown in panel C.

- (v) heterocyclic ring systems such as imidazole (PAR2) and imidazothiazole (CXCR4);
- (vi) isothioureas (CXCR4);
- (vii) benzamidines (BLT₁).

Several receptors of other GPCR subfamilies for which structures have not yet been reported also recognize amine ligands via conserved residues in the extracellular TM binding pocket, including melanocortin (D^{3.29}) and FMLP-related receptors (D^{3.33}), as well as class C GPCRs such as calcium-sensing receptor CASR, GPR158, and GPRC6A (E^{7.39}).

Figure 4 shows anionic GPCR ligands interacting with cationic residues in receptor binding sites. In 19 of the 136 crystallized GPCR–ligand complexes, the interaction between anionic, acidic groups and positively charged residues play an important role. The structures of S1P₁ ($K^{NT}/R^{3.28}$)

Trends in Pharmacological Sciences





Trends in Pharmacological Sciences

Figure 3. Cationic G Protein-Coupled Receptor Ligands Interacting with Anionic Residues in Receptor Binding Sites. (A) Chemical structures of cationic crystallized G protein-coupled receptor (GPCR) ligands. Polar interactions with receptor residues are depicted schematically. (B) Structural binding modes of representative cationic crystallized GPCR ligands shown in panel A. The locations of the overlaid binding pockets are shown in the scheme on the left.

Trends in Pharmacological Sciences





Figure 4. Anionic G Protein-Coupled Receptor Ligands Interacting with Cationic Residues in Receptor Binding Sites. (A) Chemical structures of anionic crystallized G protein-coupled receptor (GPCR) ligands. Polar interactions with receptor residues are depicted schematically. (B) Structural binding modes of representative anionic crystallized GPCR ligands shown in panel A. The relative locations of the binding pockets are shown in the scheme on the left.

Trends in Pharmacological Sciences



[35], LPA1 (K^{NT}/R^{3.28}/K^{7.36}) [34], AT1/2 (R^{4.64}/K^{5.42}) [11,40], ETB (K^{5.38}/R^{6.55}) [17], FFA1 (R^{5.39}/ K^{7.35}) [18,19], P2Y₁ (K^{1.27}/R^{45.43}/R^{7.39}) [11], and P2Y₁₂ (R^{3.21}/K^{45.54}/R^{6.55}/K^{7.35}) [66,77] receptors exhibit a diverse combination of positions of cationic residues that can interact with anionic ligand moieties in minor and major binding pockets of GPCRs. Cocrystallized negatively charged ligands furthermore interact with polar subpockets in the allosteric extrahelical binding sites of FFA1 (Y^{2.42}) [19], as well as class B GPCRs GCGR and GLP-1R (R^{6.37b}/S^{6.41}/N^{8.47}/K^{8.48}) [39,43,78,79]. The GPCR structures demonstrate how a variety of different anionic groups, including carboxylic acids (FFA1, LPA1, GCGR, GLP-1R), phosphonic acid (S1P1, P2Y1/12), tetrazole (AT_{1/2}, GCGR, GLP-1R), and acidic sulfonamides (ET_B, P2Y₁₂) play important roles in receptor binding, consistent with SAR studies for these and homologous receptors (e.g., 79-82]). Large portions of the lipid (44%) and nucleotide (11%) receptor ligands contain an anionic group, including 32 lipid and 8 nucleotide receptors for which no crystal structure is currently available. More than 50% of the ligands of several other GPCR subfamilies for which no structures have been solved contain acidic moieties, including cholecystokinin CCK_{1/2}, hydroxycarboxylic acid HCA_{2/3}, and DP_{1/2}, EP₁₋₄, FP, IP, and TP prostanoid receptors (Table S1 in the Supplemental Information online), and these anionic ligand features are found to be important for receptor binding [83,84]. Ionic interactions between anionic ligand moieties and cationic residues have been shown to play an important role in ligand binding of brain gut peptide receptors GHSR ($\mathbb{R}^{6.55}$), MTLR (R^{6.55}), and NTS_{1/2} (R^{6.54}) [85–87]. Ligand similarity assessments (Table 1 and Figure 1) and mutation studies indicate that most of these ligands very likely share the conserved interaction features observed in the available crystal structures.

G Protein-Coupled Receptor Ligand Scaffold Diversity

The deconstruction of GPCR ligands using predefined fragmentation rules [52,88] yielded 101 494 unique substructures of which 15 212 occur in ligand sets of three or more different GPCRs (Table S3 in the Supplemental Information online). These computationally derived, multitarget substructures include 4872 fragment-like (rule of 3 [89]) and 11 571 druglike (rule of 5 [90]) scaffolds (Figure 5). Frequently occurring substructures are:

- (i) amines in combination with an adjacent aromatic ring, including phenyl- and benzylpiperazines and -piperidines. These substructures are present in 5% of all known GPCR ligands of 89 different receptors, including aminergic, chemokine, opioid, and peptide receptors. These scaffolds are present in nine cocrystallized ligands interacting with negatively charged residues (D^{3.32}) and clusters of aromatic residues (F/Y^{3.33}, F/Y^{6.51}, F/H^{6.52}) in the binding sites of β_1 , δ OR, μ OR, κ OR, and NOP opioid receptors (Figure 3) [91–94]. An example of a more complex promiscuous amine scaffold is ergoline, which primarily binds aminergic receptors (Figure 5B);
- (ii) purine moleties present in the endogenous ligands of adenosine and purinergic receptors. These scaffolds are present in nine cocrystallized ligands interacting with conserved asparagine residues ($N^{6.55}$ in A₁ and A_{2A}; $N^{5.40}$ in P2Y₁₂; $N^{6.58}$ in P2Y₁; Figure 5B) that have been shown to play an important role in multiple adenosine and purinergic receptors [11,66,95];
- (iii) phenyl rings attached to indoles, triazoles, and bioisosteres of esters, amides, and phenols (e.g., phenyloxadiazoles, quinolinones, benzoxazinones, benzo-amidazolones). These moieties are compatible with polar residues in a wide variety of GPCR binding sites, including orphan GPCRs, and are represented, for example, in adrenergic β₂ and adenosine A_{1/2A} crystal structures [96–98];
- (iv) sulfonamide or urea linkers between aromatic groups, facilitating ligand binding in narrow parts of the intracellular pocket of CCR9 chemokine receptor [12] (Figure 5B) and the orthosteric pocket in OX₂ [99], or to interact with the helical backbone at the helixmembrane interface, observed in the P2Y₁ crystal structure [11];

Trends in Pharmacological Sciences



Trends in Pharmacological Sciences

Figure 5. Chemical Scaffold Analysis of G Protein-Coupled Receptor Ligands. (A) Number of chemical scaffolds (containing at least two ring structures and at least one H-bond donor and/or acceptor) shared by a specific number of G protein-coupled receptors (GPCRs) (top) and a specific number of GPCR ligands (bottom). (B) Chemical structures of selected scaffolds frequently occurring in GPCR ligands. (C) Selected abundant scaffolds in GPCR ligands that are also present in cocrystallized ligands and their structural binding modes in representative crystal structures.

(v) phenyl rings attached to carboxylic acid isosteres such as tetrazoles, which are present in ligands for several different GPCRs, including chemoattractant and brain gut receptors. The tetrazole of olmesartan and analogues interact with $R^{4.64}$ and $R^{5.42}$ in AT_1 and AT_2 crystal structures (Figure 5B) [11,40] and is proposed to interact with alternative cationic residues in, for example, GHSR ($R^{6.55}$), bombesin BB₃ ($R^{3.32}$, $R^{7.39}$), and bradykinin B_{1/2} receptors ($K^{3.33}$, $R^{5.38}$) [85,100–102].

Exploring the Chemistry of Novel Allosteric Binding Pockets

As more and more GPCR crystal structures became available, it was recognized that ligands not only bind the ancestral binding pocket within the extracellular portion of the TM bundle but that various other binding sites also exist for specific GPCRs, such as deep or intracellular sites within the TM bundle, or lipid-exposed binding sites on the outside of the TM bundle (Figure 6A) [4,62,103–105]. The cocrystallized NAMs of P2Y₁ (BPTU [11]), PAR2 (AZ3451 [37]), FFA1 (AP8

Trends in Pharmacological Sciences





Figure 6. Allosteric Extrahelical and Intracellular G Protein-Coupled Receptor Ligands. (A) Different small molecule ligand binding sites observed in G protein-coupled receptor (GPCR) structures within the transmembrane bundle (left) and exposed to solvent/lipid (right). (B) Ratio of exposed/buried ligand surface area versus total ligand surface area buried by the receptor for all cocrystallized small molecule GPCR ligands; (C) Fraction of ligands within subfamily that are chemically similar to GPCR ligands interacting with noncanonical binding pockets. The cocrystallized ligands were compared with all known actives from ChEMBL (pActivity \geq 6) for the same receptor using ECFP-4 (cutoff 0.6) and MACCS (cutoff 0.9) molecular fingerprints (Box 2). The identified similar ChEMBL ligands were iteratively used as input for similarity comparisons using the same approach until no new ligands were identified. (D) Representative ligands identified to be potential membrane site binders in P2Y₁ and GCGR based on the iterative chemical similarity assessment.

Trends in Pharmacological Sciences



[19]), C5a1 (NDT9513727 [32]), GCGR (NNC0640 [43], MK-0893 [78]), and GLP-1R (PF-06372222 and NNC0640 [39]) lock the receptor in an inactive conformation by targeting the extrahelical receptor-membrane interface [62]. The agonist TAK875 interacts with the orthosteric and extrahelical binding sites of the FFA1 receptor [18]. The observed binding sites are receptor dependent, involving: (i) the extracellular half of helices III, IV, and V (FFA1, C5a1); (ii) the midregion of helices I, II, and III (P2Y1); (iii) the midregion helices II, III, and IV (PAR2); and (iv) the intracellular portion of helices V, VI, and VII (GCGR, GLP-1R). The exposed versus buried surface ratio of these ligands is significantly larger than the other cocrystallized GPCR ligands (Figure 6B) and their receptor affinity is determined not only by interactions with buried receptorbinding pockets, but also by favorable hydrophobic interactions with the membrane. The intracellular binding pockets of the adrenergic β_2 (CMPD-15) [106] and chemokine CCR2 (RA-[R]) [33] and CCR9 (vercirnon) [12] receptors are more buried and partially overlap with the G-protein binding pocket (Figures 6A,B). Large fractions of the ligands reported for FFA1 (77%), P2Y₁ (64%), GCGR (54%), and CCR9 (48%) are chemically similar to the cocrystallized extrahelical and intracellular ligands of these receptors (Figure 6C). Most of the reported CCR9 ligands share an N-phenylbenzenesulfonamide scaffold with the cocrystallized antagonist vercirnon or contain a carboxylic acid moiety and thus are likely to target the backbone of helix VIII in the intracellular binding site of CCR9 (Figures 5 and 6A). The cocrystallized intracellular ligand of CCR2 is not chemically similar to other CCR2 ligands reported in ChEMBL, but 40 CCR2 ligands contain the same N-phenylbenzenesulfonamide scaffold as the CCR9 ligand vercimon [12] and it can be hypothesized that they also bind the intracellular binding pocket. The cocrystallized GLP-1R ligands are similar only to weakly potent GLP-1R ligands [107] but share similarity with 54% of the known GCGR ligands. Eighteen (6%) GLP-1R ligands act as positive allosteric modulators that covalently bind C^{6.36b} and target an extrahelical binding site that partially overlaps with the binding site of the crystallized NAMs. The extracellular M₂ positive allosteric modulator LY2119620 shares the same thienopyridine-2carboxamide scaffold as 51 (11%) muscarinic M₄ receptor ligands [108]. Finally, the cocrystallized intracellular β_2 ligand, and the PAR2, FFA1, and C5a₁ membrane site binders are dissimilar from any ligands of these receptors. In the case of FFA1 and C5a1, several ligands of the receptors clearly share the pharmacophores with the respective cocrystallized ligands, but they do not reach the specified cutoffs for chemical similarity.

Concluding Remarks

The comparative analysis of structurally and pharmacologically characterized GPCR ligands shows that cocrystallized GPCR ligands cover a significant part of chemical GPCR ligand space. The accumulated structural GPCR-ligand interactome contains several ligands and their substructures that interact across multiple receptor subfamilies. Experimentally determined GPCR structures represent diverse orthosteric and allosteric binding pockets, ligand binding modes, and receptor-ligand interactions, that can be translated into a variety of chemically similar GPCR ligands. The integrated structural cheminformatics and chemogenomics analyses suggest that the currently solved GPCR structures already provide efficient templates for modeling a significant part of GPCR-ligand complexes. However, the systematic structural cheminformatics assessment identifies several GPCRs and GPCR ligand chemotypes that require experimentally determined GPCR structural templates. Solving structures of such key GPCR-ligand complexes would allow significant extension of the structural GPCRligand interactome. The current analysis indicates that the combination of structural, biological, and chemical information on GPCR-ligand interactions enables the investigation of as yet unexplored GPCR-ligand complexes and the structure-based design of novel chemical modulators of GPCR polypharmacology (see Outstanding Questions).

Outstanding Questions

How many more experimental structures of GPCR-ligand complexes are required to extend the coverage of chemical GPCR ligand space from 11% to 50%?

Will the extended coverage of chemical ligand space of new GPCR structures hold the same pace as the extension of chemical GPCR ligand space by (computer-aided) medicinal chemistry efforts?

How many known GPCR ligands have a cross-GPCR subfamily polypharma-cological interaction profile (i.e., ${\cal K}_i/{\cal K}_D/$ IC₅₀/EC₅₀ $\leq 1~\mu M$ for members of at least two different GPCR subfamilies)?

Are the currently known allosteric binding sites exclusively identified in specific receptor crystal structures conserved between receptors, allowing ligand repurposing?

How many more druggable binding sites (in addition to the ones so far identified in GPCR crystal structures) offer new opportunities for structurebased GPCR ligand design?

Do GPCR oligomer interfaces provide druggable binding sites, and will it be possible to capture the structures of such complexes experimentally by cryo-electron microscopy (cryo-EM) studies?

What are the structural dynamics and binding kinetics of transient extrahelical GPCR binding sites, and will it be possible to identify these dynamic binding sites by (serial) crystallography studies?

Do privileged GPCR ligand scaffolds reflect receptor binding site similarity, or are they the result of the prioritization of preferred synthetic medicinal chemistry approaches?

To what extent will structure-based design accelerate the enrichment of the GPCR medicinal chemistry toolbox with new chemical scaffolds, including macrocycles and new heteroaromatic ring systems?

Trends in Pharmacological Sciences



Acknowledgments

This work was supported by the National Natural Science Foundation of China grant 81573479 (D.Y.), the Personalized Medicine–Molecular Signature-based Drug Discovery and Development Strategic Priority Research Program of the Chinese Academy of Sciences grants XDA12020347 (M.-W.W.) and XDA12040308 (D.Y.), the National Health and Family Planning Commission grants 2012ZX09304-011, 2013ZX09401003-005, 2013ZX09507001, and 2013ZX09507-002 (M.-W.W.), the Novo Nordisk–Chinese Academy of Sciences Research Fund (M.-W.W. and D.Y.), the Shanghai Science and Technology Development Fund 15DZ2291600 (M.-W.W.), the Shanghai local government (R.C.S.), the GPCR Consortium (R.C.S., A.J.K., C.d.G., M.V.), the Netherlands eScience Center (NLeSC)/NWO Enabling Technologies project: 3D-e-Chem, grant 027.014.201 (A.J.K., C.d.G., M.V.).

Resources

ⁱhttps://zenodo.org/record/1163106

Supplemental Information

Supplemental information associated with this article can be found, in the online version, at https://doi.org/10.1016/j.tips. 2018.02.004.

References

- Hauser, A.S. et al. (2017) Trends in GPCR drug discovery: new agents, targets and indications. Nat. Rev. Drug Discov. 16, 829–842
- Katritch, V. et al. (2013) Structure-function of the G proteincoupled receptor superfamily. Annu. Rev. Pharmacol. Toxicol. 53, 531–556
- Alexander, S.P. *et al.* (2017) The concise guide to pharmacology 2017/18: G protein-coupled receptors. *Br. J. Pharmacol.* 174 (Suppl 1), S17–S129
- 4. Wacker, D. et al. (2017) How ligands illuminate GPCR molecular pharmacology. Cell 170, 414–427
- 5. Santos, R. *et al.* (2017) A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* 16, 19–34
- Peters, J.U. (2013) Polypharmacology foe or friend? J. Med. Chem. 56, 8955–8971
- Piscitelli, C.L. *et al.* (2015) A molecular pharmacologist's guide to G protein-coupled receptor crystallography. *Mol. Pharmacol.* 88, 536–551
- Mason, J.S. et al. (2012) New insights from structural biology into the druggability of G protein-coupled receptors. *Trends Pharmacol. Sci.* 33, 249–260
- Cooke, R.M. et al. (2015) Structures of G protein-coupled receptors reveal new opportunities for drug discovery. Drug Discov. Today 20, 1355–1364
- Bento, A.P. et al. (2014) The ChEMBL bioactivity database: an update. Nucleic Acids Res. 42 (Database issue), D1083–D1090
- Zhang, H. *et al.* (2015) Structure of the angiotensin receptor revealed by serial femtosecond crystallography. *Cell* 161, 833–844
- 12. Oswald, C. et al. (2016) Intracellular allosteric antagonism of the CCR9 receptor. Nature 540, 462–465
- Zhang, C. *et al.* (2012) High-resolution crystal structure of human protease-activated receptor 1. *Nature* 492, 387–392
- Fenalti, G. et al. (2014) Molecular control of delta-opioid receptor signalling. Nature 506, 191–196
- Fenalti, G. et al. (2015) Structural basis for bifunctional peptide recognition at human delta-opioid receptor. Nat. Struct. Mol. Biol. 22, 265–268
- 16. Granier, S. et al. (2012) Structure of the delta-opioid receptor bound to naltrindole. *Nature* 485, 400–404
- Shihoya, W. et al. (2017) X-ray structures of endothelin ETB receptor bound to clinical antagonist bosentan and its analog. *Nat. Struct. Mol. Biol.* 24, 758–764

- Srivastava, A. et al. (2014) High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature* 513, 124–127
- Liang, Y.L. et al. (2017) Phase-plate cryo-EM structure of a class B GPCR-G-protein complex. Nature 546, 118–123
- Drwal, M.N. et al. (2017) Multi-target fragments display versatile binding modes. Mol. Inform. Published online October 2017. http://dx.doi.org/10.1002/minf.201700042
- de Graaf, C. *et al.* (2013) Small and colorful stones make beautiful mosaics: fragment-based chemogenomics. *Drug Discov. Today* 18, 323–330
- Chien, E.Y. *et al.* (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* 330, 1091–1095
- Shimamura, T. et al. (2011) Structure of the human histamine H1 receptor complex with doxepin. Nature 475, 65–70
- Thal, D.M. et al. (2016) Crystal structures of the M1 and M4 muscarinic acetylcholine receptors. Nature 531, 335–340
- Kruse, A.C. et al. (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. Nature 482, 552–556
- Thorsen, T.S. et al. (2014) Modified T4 lysozyme fusion proteins facilitate G proteincCoupled receptor crystallogenesis. Structure 22, 1657–1664
- Wang, C. *et al.* (2013) Structural basis for molecular recognition at serotonin receptors. *Science* 340, 610–614
- Wacker, D. et al. (2013) Structural features for functional selectivity at serotonin receptors. Science 340, 615–619
- Liu, W. et al. (2013) Serial femtosecond crystallography of G protein-coupled receptors. Science 342, 1521–1524
- Wacker, D. et al. (2017) Crystal structure of an LSD-bound human serotonin receptor. Cell 168, 377–389
- Ishchenko, A. et al. (2017) Structural insights into the extracellular recognition of the human serotonin 2B receptor by an antibody. Proc. Natl. Acad. Sci. U. S. A. 114, 8223–8228
- Robertson, N. et al. (2018) Structure of the complement C5a receptor bound to the extra-helical antagonist NDT9513727. *Nature* 553, 111–114
- Zheng, Y. et al. (2016) Structure of CC chemokine receptor 2 with orthosteric and allosteric antagonists. Nature 540, 458–461
- 34. Chrencik, J.E. et al. (2015) Crystal structure of antagonist bound human lysophosphatidic acid receptor 1. Cell 161, 1633–1643
- Hanson, M.A. et al. (2012) Crystal structure of a lipid G proteincoupled receptor. Science 335, 851–855



- Hori, T. *et al.* (2018) Na(+)-mimicking ligands stabilize the inactive state of leukotriene B4 receptor BLT1. *Nat. Chem. Biol.* 14, 262–269
- Cheng, R.K.Y. et al. (2017) Structural insight into allosteric modulation of protease-activated receptor 2. Nature 545, 112–115
- Hollenstein, K. et al. (2013) Structure of class B GPCR corticotropin-releasing factor receptor 1. Nature 499, 438–443
- Song, G. *et al.* (2017) Human GLP-1 receptor transmembrane domain structure in complex with allosteric modulators. *Nature* 546, 312–315
- Zhang, H. et al. (2017) Structural basis for selectivity and diversity in angiotensin II receptors. Nature 544, 327–332
- Burg, J.S. et al. (2015) Structural biology. Structural basis for chemokine recognition and activation of a viral G protein-coupled receptor. Science 347, 1113–1117
- 42. Jazayeri, A. et al. (2017) Crystal structure of the GLP-1 receptor bound to a peptide agonist. *Nature* 546, 254–258
- 43. Zhang, H. *et al.* (2017) Structure of the full-length glucagon class B G-protein-coupled receptor. *Nature* 546, 259–264
- Taniguchi, R. *et al.* (2017) Structural insights into ligand recognition by the lysophosphatidic acid receptor LPA6. *Nature* 548, 356–360
- Haga, K. *et al.* (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* 482, 547–551
- Kooistra, A.J. et al. (2013) A structural chemogenomics analysis of aminergic GPCRs: lessons for histamine receptor ligand design. Br. J. Pharmacol. 170, 101–126
- Jaakola, V.P. et al. (2008) The 2. 6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* 322, 1211–1217
- 48. Jacobson, K.A. (2015) New paradigms in GPCR drug discovery. Biochem. Pharmacol. 98, 541–555
- Rosini, M. *et al.* (2016) Multitarget strategies in Alzheimer's disease: benefits and challenges on the road to therapeutics. *Future Med. Chem.* 8, 697–711
- Manglik, A. et al. (2016) Structure-based discovery of opioid analgesics with reduced side effects. Nature 537, 185–190
- 51. Klaeger, S. et al. (2017) The target landscape of clinical kinase drugs. *Science* 358, eaan4368
- McGuire, R. *et al.* (2017) 3D-e-Chem-VM: structural cheminformatics research infrastructure in a freely available virtual machine. *J. Chem. Inf. Model.* 57, 115–121
- Pandy-Szekeres, G. et al. (2018) GPCRdb in 2018: adding GPCR structure models and ligands. Nucleic Acids Res. 46, D440–D446
- Kooistra, A.J. et al. (2018) 3D-e-Chem: structural cheminformatics workflows for computer-aided drug discovery. Chem-MedChem Published online January 16, 2018. http://dx.doi. org/10.1002/cmdc.201700754
- Isberg, V. et al. (2016) GPCRdb: an information system for G protein-coupled receptors. Nucleic Acids Res. 44, D356–D364
- Turnaturi, R. et al. (2016) Multitarget opioid ligands in pain relief: new players in an old game. Eur. J. Med. Chem. 108, 211–228
- Ngo, T. *et al.* (2017) Orphan receptor ligand discovery by pickpocketing pharmacological neighbors. *Nat. Chem. Biol.* 13, 235–242
- Lin, X. et al. (2012) Life beyond kinases: structure-based discovery of sorafenib as nanomolar antagonist of 5-HT receptors. J. Med. Chem. 55, 5749–5759
- 59. Besnard, J. *et al.* (2012) Automated design of ligands to polypharmacological profiles. *Nature* 492, 215–220
- Vass, M. et al. (2014) Multiple fragment docking and linking in primary and secondary pockets of dopamine receptors. ACS Med. Chem. Lett. 5, 1010–1014

- Vass, M. *et al.* (2016) Molecular interaction fingerprint approaches for GPCR drug discovery. *Curr. Opin. Pharmacol.* 30, 59–68
- Congreve, M. et al. (2017) Applying structure-based drug design approaches to allosteric modulators of GPCRs. Trends Pharmacol. Sci. 38, 837–847
- Rosenkilde, M.M. et al. (2010) The minor binding pocket: a major player in 7TM receptor activation. *Trends Pharmacol. Sci.* 31, 567–574
- Kruse, A.C. et al. (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. Nature 504, 101–106
- Lebon, G. et al. (2011) Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* 474, 521–525
- Zhang, J. et al. (2014) Agonist-bound structure of the human P2Y12 receptor. Nature 509, 119–122
- 67. Hua, T. et al. (2017) Crystal structures of agonist-bound human cannabinoid receptor CB1. Nature 547, 468–471
- Tan, Q. et al. (2013) Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. Science 341, 1387–1390
- Wu, B. et al. (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330, 1066–1071
- Glukhova, A. *et al.* (2017) Structure of the adenosine A1 receptor reveals the basis for subtype selectivity. *Cell* 168, 867–877
- Cheng, R.K.Y. et al. (2017) Structures of human A1 and A2A adenosine receptors with xanthines reveal determinants of selectivity. Structure 25, 1275–1285
- Segala, E. et al. (2016) Controlling the dissociation of ligands from the adenosine A2A receptor through modulation of salt bridge strength. J. Med. Chem. 59, 6470–6479
- Dore, A.S. *et al.* (2011) Structure of the adenosine A(2A) receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* 19, 1283–1293
- 74. Xu, F. et al. (2011) Structure of an agonist-bound human A2A adenosine receptor. *Science* 332, 322–327
- Surgand, J.S. et al. (2006) A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. Proteins 62, 509–538
- Arimont, M. et al. (2017) Structural analysis of chemokine receptor-ligand interactions. J. Med. Chem. 60, 4735–4779
- Zhang, K. et al. (2014) Structure of the human P2Y12 receptor in complex with an antithrombotic drug. *Nature* 509, 115–118
- Jazayeri, A. et al. (2016) Extra-helical binding site of a glucagon receptor antagonist. Nature 533, 274–277
- de Graaf, C. et al. (2017) Extending the structural view of class B GPCRs. Trends Biochem. Sci. 42, 946–960
- Parrill, A.L. (2008) Lysophospholipid interactions with protein targets. *Biochim. Biophys. Acta* 1781, 540–546
- Hudson, B.D. et al. (2013) The therapeutic potential of allosteric ligands for free fatty acid sensitive GPCRs. Curr. Top. Med. Chem. 13, 14–25
- Karnik, S.S. et al. (2015) International Union of Basic and Clinical Pharmacology. XCIX. Angiotensin receptors: interpreters of pathophysiological angiotensinergic stimuli corrected. *Pharma*col. Rev. 67, 754–819
- Jones, R.L. et al. (2009) Prostanoid receptor antagonists: development strategies and therapeutic applications. Br. J. Pharmacol. 158, 104–145
- Foucaud, M. et al. (2008) Insights into the binding and activation sites of the receptors for cholecystokinin and gastrin. Regul. Pept. 145, 17–23
- Feighner, S.D. et al. (1998) Structural requirements for the activation of the human growth hormone secretagogue receptor by peptide and nonpeptide secretagogues. *Mol. Endocrinol.* 12, 137–145



- Matsuura, B. *et al.* (2002) Differential determinants for peptide and non-peptidyl ligand binding to the motilin receptor. Critical role of second extracellular loop for peptide binding and action. *J. Biol. Chem.* 277, 9834–9839
- Labbe-Jullie, C. *et al.* (1998) Mutagenesis and modeling of the neurotensin receptor NTR1. Identification of residues that are critical for binding SR 48692, a nonpeptide neurotensin antagonist. *J. Biol. Chem.* 273, 16351–16357
- Wood, D.J. et al. (2012) Pharmacophore fingerprint-based approach to binding site subpocket similarity and its application to bioisostere replacement. J. Chem. Inf. Model. 52, 2031–2043
- Congreve, M. et al. (2003) A 'rule of three' for fragment-based lead discovery? Drug Discov. Today 8, 876–877
- Lipinski, C.A. et al. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 46, 3–26
- 91. Wu, H. et al. (2012) Structure of the human kappa-opioid receptor in complex with JDTic. Nature 485, 327–332
- Thompson, A.A. *et al.* (2012) Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* 485, 395–399
- Miller, R.L. *et al.* (2015) The importance of ligand-receptor conformational pairs in stabilization: spotlight on the N/OFQ G protein-coupled receptor. *Structure* 23, 2291–2299
- Christopher, J.A. *et al.* (2013) Biophysical fragment screening of the beta1-adrenergic receptor: identification of high affinity arylpiperazine leads using structure-based drug design. *J. Med. Chem.* 56, 3446–3455
- Jespers, W. et al. (2018) Structural mapping of adenosine receptor mutations: ligand binding and signaling mechanisms. *Trends Pharmacol. Sci.* 39, 75–89
- Congreve, M. et al. (2012) Discovery of 1,2,4-triazine derivatives as adenosine A(2A) antagonists using structure based drug design. J. Med. Chem. 55, 1898–1903
- Warne, T. et al. (2011) The structural basis for agonist and partial agonist action on a beta(1)-adrenergic receptor. Nature 469, 241–244
- Rasmussen, S.G. et al. (2011) Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. Nature 469, 175–180
- Suno, R. et al. (2018) Crystal structures of human orexin 2 receptor bound to the subtype-selective antagonist EMPA. Structure 26, 7–19
- Gbahou, F. et al. (2010) Molecular basis for agonism in the BB3 receptor: an epitope located on the interface of transmembrane-III, -VI, and -VII. J. Pharmacol. Exp. Ther. 333, 51–59

- Leeb-Lundberg, L.M. *et al.* (2005) International union of pharmacology. XLV. Classification of the kinin receptor family: from molecular mechanisms to pathophysiological consequences. *Pharmacol. Rev.* 57, 27–77
- Bondensgaard, K. et al. (2004) Recognition of privileged structures by G-protein coupled receptors. J. Med. Chem. 47, 888–899
- 103. Gentry, P.R. et al. (2015) Novel allosteric modulators of G protein-coupled receptors. J. Biol. Chem. 290, 19478–19488
- 104. Kooistra, A.J. *et al.* (2014) The receptor concept in 3D: from hypothesis and metaphor to GPCR-ligand structures. *Neurochem. Res.* 39, 1850–1861
- 105. Christopoulos, A. (2014) Advances in G protein-coupled receptor allostery: from function to structure. *Mol. Pharmacol.* 86, 463–478
- 106. Liu, X. et al. (2017) Mechanism of intracellular allosteric beta2AR antagonist revealed by X-ray crystal structure. Nature 548, 480–484
- 107. Filipski, K.J. et al. (2012) A novel series of glucagon receptor antagonists with reduced molecular weight and lipophilicity. *Bioorg. Med. Chem. Lett.* 22, 415–420
- 108. Le, U. et al. (2013) Discovery of a selective M(4) positive allosteric modulator based on the 3-amino-thieno[2,3-b]pyridine-2-carboxamide scaffold: development of ML253, a potent and brain penetrant compound that is active in a preclinical model of schizophrenia. *Bioorg. Med. Chem. Lett.* 23, 346–350
- 109. Durant, J.L. et al. (2002) Reoptimization of MDL keys for use in drug discovery. J. Chem. Inf. Comput. Sci. 42, 1273–1280
- 110. Landrum, G. RDKit: Open-source cheminformatics. http:// www.rdkit.org
- 111. Rogers, D. and Hahn, M. (2010) Extended-connectivity fingerprints. J. Chem. Inf. Model. 50, 742–754
- 112. Clark, A.M. et al. (2014) New target prediction and visualization tools incorporating open source molecular fingerprints for TB Mobile 2.0. J. Cheminform. 6, 38
- Berthold, M.R. et al. (2007) KNIME: the Konstanz information miner. In Data Analysis, Machine Learning and Applications, pp. 319–326, Springer
- 114. Jasial, S. et al. (2016) Activity-relevant similarity values for fingerprints and implications for similarity searching. F1000Research Published online 6 April 2016. http://dx.doi.org/10.12688/ f1000research.8357.2