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 Graaf1,*

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 pharmacologically 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 physicochemical 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 disease modulators of GPCR function. 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 (K_i/K_D/IC_50/EC_50 ≤ 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 GPCR–ligand 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.
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], β1/2 adrenergic (17%), endothelin ETB (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 D3/4 [22], histamine H1 [23], muscarinic M1/3/4 [24–26], serotonin 5-HT1B/2B [27–31], complement peptide C5a, 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 glucagon-like 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
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-HT₁B receptor, the tryptamine substructure of the cocrystallized (dihydro-)ergotamine is present in 23% of known 5-HT₁B ligands [27]. Similarly, the phosphonic acid group of the cocrystallized ML056 [33] 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 M₁,₂ ligands share an amine, with two aromatic rings oriented in a butterfly shape, with many other H₁ and M₁,₂ 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 D₂ (4758 ligands, similar to D₃), cannabinoid CB₂ (3687, similar to CB₁), serotonin 5-HT₁A (3271), 5-HT₂A (3113), and 5-HT₂C (2298) receptors (similar to 5-HT₁B and 5-HT₂B), and adenosine A₁ (2728, similar to A₁ and A₂A). 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 melatonin-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 GPCR ligands, and thus the coverage of the ligand space of these receptors, and therefore can provide useful templates to extend novel chemical ligand space using structure-based drug discovery and design approaches.

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
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/IC50/EC50 ≤ 1 µM 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 ≤ 1 µM). 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 melanocortin–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 also 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 SHT2A 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.

Scaffolds: 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 membrane-embedded proteins, such as the seven transmembrane helices of GPCRs or the 6–24 transmembrane helices of ligand-gated ion channels.

References [48–50,52–55,57–58,60–61]
### 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 2.

| Receptor | All (46/419) | Class A (39/299) | Aminergic (10/37) | B 
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of structures</td>
<td>229</td>
<td>204</td>
<td>58</td>
<td>18</td>
</tr>
<tr>
<td>Number of ligands (X-ray)</td>
<td>122 (11)</td>
<td>104 (8)</td>
<td>33</td>
<td>12</td>
</tr>
<tr>
<td>Number of ligands (rec.)</td>
<td>96287</td>
<td>89315</td>
<td>26515</td>
<td>6980</td>
</tr>
<tr>
<td>Number of ligands (rec.) similar to ligands (rec. X-ray)</td>
<td>96</td>
<td>72</td>
<td>69</td>
<td>31</td>
</tr>
<tr>
<td>Number of ligands (rec.) similar to ligands (GPCR X-ray)</td>
<td>68</td>
<td>72</td>
<td>69</td>
<td>31</td>
</tr>
<tr>
<td>Number of ligands (other GPCR) similar to ligands (rec. X-ray)</td>
<td>287</td>
<td>258</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>Number of ligands (other GPCR) similar to ligands (rec. X-ray)</td>
<td>223</td>
<td>195</td>
<td>95</td>
<td>1</td>
</tr>
<tr>
<td>Number of ligands active at other GPCRs</td>
<td>65</td>
<td>78</td>
<td>183</td>
<td>698</td>
</tr>
<tr>
<td>Number of ligands active at other non-subfamily GPCRs</td>
<td>29</td>
<td>1</td>
<td>28</td>
<td>9</td>
</tr>
</tbody>
</table>

### Peptide (14/77)

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Number of structures</th>
<th>Number of ligands (X-ray)</th>
<th>Number of ligands (rec.)</th>
<th>Number of ligands (rec.) similar to ligands (rec. X-ray)</th>
<th>Number of ligands (rec.) similar to ligands (GPCR X-ray)</th>
<th>Number of ligands (other GPCR) similar to ligands (rec. X-ray)</th>
<th>Number of ligands (other GPCR) similar to ligands (rec. X-ray)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT&lt;sub&gt;1&lt;/sub&gt;</td>
<td>2</td>
<td>2</td>
<td>666</td>
<td>151</td>
<td>203</td>
<td>65</td>
<td>223</td>
</tr>
<tr>
<td>AT&lt;sub&gt;2&lt;/sub&gt;</td>
<td>3</td>
<td>2</td>
<td>346</td>
<td>37</td>
<td>64</td>
<td>188</td>
<td>195</td>
</tr>
<tr>
<td>APJ</td>
<td>1</td>
<td>1</td>
<td>76</td>
<td>–</td>
<td>0</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>C&lt;sub&gt;5a&lt;/sub&gt;A</td>
<td>1</td>
<td>1</td>
<td>310</td>
<td>0</td>
<td>5</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>ET&lt;sub&gt;B&lt;/sub&gt;</td>
<td>4</td>
<td>3</td>
<td>527</td>
<td>108</td>
<td>110</td>
<td>141</td>
<td>459</td>
</tr>
<tr>
<td>NTS&lt;sub&gt;1&lt;/sub&gt;</td>
<td>8</td>
<td>1</td>
<td>133</td>
<td>–</td>
<td>0</td>
<td>–</td>
<td>53</td>
</tr>
<tr>
<td>δOR</td>
<td>4</td>
<td>2</td>
<td>2588</td>
<td>437</td>
<td>629</td>
<td>411</td>
<td>1685</td>
</tr>
<tr>
<td>kOR</td>
<td>2</td>
<td>2</td>
<td>2872</td>
<td>177</td>
<td>582</td>
<td>307</td>
<td>1806</td>
</tr>
<tr>
<td>μOR</td>
<td>2</td>
<td>2</td>
<td>2785</td>
<td>285</td>
<td>505</td>
<td>590</td>
<td>2385</td>
</tr>
<tr>
<td>NOP</td>
<td>3</td>
<td>3</td>
<td>1301</td>
<td>59</td>
<td>104</td>
<td>613</td>
<td>720</td>
</tr>
<tr>
<td>OX&lt;sub&gt;1&lt;/sub&gt;</td>
<td>2</td>
<td>2</td>
<td>1354</td>
<td>115</td>
<td>216</td>
<td>144</td>
<td>1138</td>
</tr>
<tr>
<td>OX&lt;sub&gt;2&lt;/sub&gt;</td>
<td>3</td>
<td>2</td>
<td>1823</td>
<td>133</td>
<td>318</td>
<td>171</td>
<td>1128</td>
</tr>
<tr>
<td>PAR1</td>
<td>1</td>
<td>1</td>
<td>547</td>
<td>224</td>
<td>225</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PAR2</td>
<td>3</td>
<td>2</td>
<td>42</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Protein (4/29)

<table>
<thead>
<tr>
<th>Protein</th>
<th>Number of structures</th>
<th>Number of ligands (X-ray)</th>
<th>Number of ligands (rec.)</th>
<th>Number of ligands (rec.) similar to ligands (rec. X-ray)</th>
<th>Number of ligands (rec.) similar to ligands (GPCR X-ray)</th>
<th>Number of ligands (other GPCR) similar to ligands (rec. X-ray)</th>
<th>Number of ligands (other GPCR) similar to ligands (rec. X-ray)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR2</td>
<td>1</td>
<td>2</td>
<td>1346</td>
<td>29</td>
<td>57</td>
<td>86</td>
<td>81</td>
</tr>
<tr>
<td>CCR5</td>
<td>2</td>
<td>2</td>
<td>1743</td>
<td>113</td>
<td>160</td>
<td>35</td>
<td>75</td>
</tr>
<tr>
<td>CCR9</td>
<td>1</td>
<td>1</td>
<td>159</td>
<td>83</td>
<td>84</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>CXCR4</td>
<td>6</td>
<td>3</td>
<td>286</td>
<td>15</td>
<td>31</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>US28 (viral)</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>–</td>
<td>0</td>
<td>–</td>
<td>1</td>
</tr>
</tbody>
</table>
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

---

**Table 1. (continued)**

<table>
<thead>
<tr>
<th>Receptor</th>
<th>Number of structuresb</th>
<th>Number of ligands (X-ray)c</th>
<th>Number of ligands (rec.d)</th>
<th>Number of ligands (rec.) similar to ligands (rec. X-ray)b</th>
<th>Number of ligands (GPCR X-ray)</th>
<th>Number of ligands active at other GPCRs</th>
<th>Number of ligands active at other non-subfamily GPCRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipid (6/37)</td>
<td>14</td>
<td>12</td>
<td>14472</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFA1</td>
<td>3</td>
<td>3</td>
<td>625</td>
<td>117</td>
<td>0</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>BLT1</td>
<td>1</td>
<td>1</td>
<td>287</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LPA1</td>
<td>3</td>
<td>3</td>
<td>106</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>LPA2</td>
<td>1</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S1P1</td>
<td>2</td>
<td>1</td>
<td>1739</td>
<td>43</td>
<td>43</td>
<td>46</td>
<td>580</td>
</tr>
<tr>
<td>CB1</td>
<td>4</td>
<td>4</td>
<td>2877</td>
<td>413</td>
<td>451</td>
<td>227</td>
<td>1221</td>
</tr>
<tr>
<td>Nucleotide (4/12)</td>
<td>42</td>
<td>23</td>
<td>7299</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>2</td>
<td>2</td>
<td>2605</td>
<td>102</td>
<td>734</td>
<td>212</td>
<td>2091</td>
</tr>
<tr>
<td>A2A</td>
<td>35</td>
<td>16</td>
<td>3020</td>
<td>811</td>
<td>848</td>
<td>1540</td>
<td>2026</td>
</tr>
<tr>
<td>P2Y1</td>
<td>2</td>
<td>2</td>
<td>314</td>
<td>181</td>
<td>212</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>P2Y12</td>
<td>3</td>
<td>3</td>
<td>907</td>
<td>99</td>
<td>126</td>
<td>280</td>
<td>4</td>
</tr>
<tr>
<td>Sensory (1/7)</td>
<td>39</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhodopsin</td>
<td>39</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Class B (4/21)</td>
<td>12</td>
<td>7 (3)</td>
<td>3598</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcitonin</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CRF</td>
<td>2</td>
<td>1</td>
<td>1687</td>
<td>5</td>
<td>27</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>GLP-1R</td>
<td>4</td>
<td>4 (2)</td>
<td>303</td>
<td>1</td>
<td>1</td>
<td>34</td>
<td>55</td>
</tr>
<tr>
<td>GCGR</td>
<td>5</td>
<td>3 (1)</td>
<td>828</td>
<td>115</td>
<td>135</td>
<td>22</td>
<td>117</td>
</tr>
<tr>
<td>Class C (2/23)</td>
<td>4</td>
<td>4</td>
<td>2928</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mGluR1</td>
<td>1</td>
<td>1</td>
<td>535</td>
<td>18</td>
<td>21</td>
<td>3</td>
<td>102</td>
</tr>
<tr>
<td>mGluR5</td>
<td>3</td>
<td>3</td>
<td>1481</td>
<td>74</td>
<td>76</td>
<td>16</td>
<td>82</td>
</tr>
<tr>
<td>Class F (1/11)</td>
<td>9</td>
<td>7</td>
<td>480</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoothened</td>
<td>9</td>
<td>7</td>
<td>480</td>
<td>46</td>
<td>53</td>
<td>346</td>
<td>3</td>
</tr>
</tbody>
</table>

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

*Receptors classified according to GPCR database \[52–55\].

*X-ray and electron microscopy structures.

*Unique cocrystallized ligands; peptide ligands (molecular weight > 800 Da) indicated in brackets.

*Number of unique ligands in ChEMBL, with binding affinity (Ki/IC50) and/or functional potency (EC50/IC50) of at least 1 µM.

*Number 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).

*Number of ligands of other GPCRs chemically similar to any cocrystallized ligand of the specified receptor (red), determined as above.
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.
of class A receptors predominantly bind in the major pocket (e.g., retinal in rhodopsin, adrenaline in A2A [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 B4, 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 ligand–receptor 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 D3.32 in aminergic (5-HT1B/2B, β1/2, D3/4, H1, M1-4) and opioid (δOR, μOR, κOR, NOP) receptor structures and E7.39 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 (E45.46) and PAR2 (D65.52), respectively [37,64]. Interestingly, the BIIL260 inverse

---

**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://3d-e-chem.github.io/) [52,54], MACCS [109] (RDKit implementation [110]), and ECPF-4 [111] (CDK 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 toGPCRs, 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.
agonist mimics the sodium ion-centered water cluster with the conserved D2.50 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 (b2, dOR), secondary (b1/2, SMO), tertiary (H1, CCR2), and quaternary (M2) amines;

(ii) monocyclic amines, including pyrrolidines (D3/4, NOP), piperidines (NOP), piperazines (b1);

(iii) multicyclic, bridged amine ring structures, including azabicyclo-octane (CCR5) azabicyclo-octanium (M1-4), quinuclidine (M2), and multicyclic opioids (dOR, uOR);

(iv) ergolines (5HT1B/2B);

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 96287 distinct ligands (Ki/KD/IC50/EC50 ≤ 1 μM) 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 (Tc) cutoffs [114] were used to determine whether molecules are chemically similar (ECFP-4 Tc ≥ 0.4 or MACCS Tc ≥ 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 (Tc ≥ 0.6) and MACCS (Tc ≥ 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).

Figure I. Chemical Similarity Search of GPCR Ligand Space. Schematic example of chemical fingerprint-based similarity search using co-crystallised adenosine A2A receptor ligands, for example, ZM241385 [47]. Fifteen of the co-crystallised A2A ligands are chemically similar to 811 A2A ligands (cyan Venn diagram), 24 of the co-crystallised GPCR ligands are similar to 848 A2A ligands (orange Venn diagram), and 15 of the co-crystallised A2A ligands are similar to 1540 GPCR ligands (green Venn diagram).
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 (D3.29) and FMLP-related receptors (D3.33), as well as class C GPCRs such as calcium-sensing receptor CASR, GPR158, and GPRC6A (E7.39).

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})
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.
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.
LPA1 (KNT/R3.28/K7.36) [34], AT1/2 (R4.64/K5.42) [11,40], ETB (K5.38/R6.55) [17], FFA1 (R5.39/K7.35) [18,19], P2Y1 (K1.27/R5.54/R6.55/K7.35) [11], and P2Y12 (R5.39/K5.54/R6.55/K7.35) [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. Co-crystallized negatively charged ligands furthermore interact with subpockets in the allosteric extrahelical binding sites of FFA1 (Y2.42) [19], as well as class B GPCRs GCGR and GLP-1R (R6.37b/S6.41/N8.47/K8.48) [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 (AT1/2, GCGR, GLP-1R), and acidic sulfonamides (ETB, 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 CCK1/2, hydroxycarboxylic acid HCA2/3, and DP1/2, 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 (R6.55), MTLR (R6.55), and NTS1/2 (R6.54) [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, multtarget 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 benzyl-piperazines 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 co-crystallized ligands interacting with negatively charged residues (D3.52) and clusters of aromatic residues (F/Y3.33, F/Y6.51, F/H6.53) in the binding sites of α1, δOR, μOR, κOR, 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 moieties present in the endogenous ligands of adenosine and purinergic receptors. These scaffolds are present in nine co-crystallized ligands interacting with conserved asparagine residues (N6.55 in A1 and A2A; N5.40 in P2Y12; N6.58 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 A1/2A 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 [89], or to interact with the helical backbone at the helix–membrane interface, observed in the P2Y1 crystal structure [11];
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 R4.64 and R5.42 in AT1 and AT2 crystal structures [11,40] and is proposed to interact with alternative cationic residues in, for example, GHSR (R6.55), bombesin BB3 (R3.32, R7.39), and bradykinin B1/2 receptors (K3.33, R5.38) [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 (AP8LSD, (dihydro-)ergotamine 5HT1B, 5HT2B

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.
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 ≥ 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.
ligands [107], but share similarity with 54% of the known GCGR ligands. Eighteen (6%) GLP-1R receptor
interactions with the membrane. The intracellular binding pockets of the adrenergic β2 (CMPD-15) [106]
and chemokine CCR2 (RA-[R]) [33] and CCR9 (vercirnon) [12] receptors are more buried and
extrahelical and intracellular ligands of these receptors (Figure 6C). Most of the reported
extrahelical GPCR binding sites by (serial) crystallography [107] but share similarity with 54% of the known GCGR ligands. Eighteen (6%) GLP-1R ligands act as positive allosteric modulators that covalently bind C6-[36b] and target an extrahelical binding site that partially overlaps with the binding site of the crystalized NAMs. The extracellular M2 positive allosteric modulator LY2119620 shares the same thienopyridine-2-carboxamide scaffold as 51 (11%) muscarinic M4 receptor ligands [108]. Finally, the cocrystralized intracellular β2 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 cocryrstallized 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 cocryrstallized 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 GPCR–ligand 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 GPCR binding sites exclusively identified in specific receptor crystal structures offer new opportunities for structure-based GPCR ligand design?

What are the structural dynamics and binding kinetics of transient extrahelical GPCR binding sites, and will it be possible to capture the structures of such complexes experimentally by cryo-electron microscopy (cryo-EM) studies?

To what extent will structure-based design accelerate the enrichment of privileged GPCR ligand scaffolds that reflect receptor binding site similarity, or are they the result of the prioritization of preferred synthetic medicinal chemistry approaches?
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 20122X09304-011, 20132X09401003-005, 20132X09507001, and 20132X09507-002 (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
https://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
Trends in Pharmacological Sciences