

The pocketome of G protein-coupled receptors reveals previously untargeted allosteric sites

Peter Kolb (✉ peter.kolb@uni-marburg.de)

Philipps-University Marburg <https://orcid.org/0000-0003-4089-614X>

Janik Hedderich

Philipps-University Marburg

Margherita Persechino

Philipps-University Marburg

Katharina Becker

Philipps-University Marburg

Franziska Heydenreich

Stanford University School of Medicine <https://orcid.org/0000-0002-8049-4383>

Torben Gutermuth

Philipps-University Marburg

Michel Bouvier

University of Montreal <https://orcid.org/0000-0003-1128-0100>

Moritz Bünemann

Philipps-University Marburg

Article

Keywords: pocketome, g-protein coupled receptors, GPCR structures

Posted Date: September 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-870555/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on May 10th, 2022. See the published version at <https://doi.org/10.1038/s41467-022-29609-6>.

The pocketome of G protein-coupled receptors reveals previously untargeted allosteric sites

Janik B. Hedderich¹, Margherita Persechino¹, Katharina Becker²,
Franziska M. Heydenreich^{3,4}, Torben Gutermuth¹,
Michel Bouvier⁴, Moritz Bünemann², Peter Kolb^{1,*}

September 2, 2021

¹Department of Pharmaceutical Chemistry, Philipps-University Marburg, Marburg, Germany

²Department of Pharmacology & Clinical Pharmacy, Philipps-University Marburg, Marburg, Germany

³Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, USA

⁴Institute for Research in Immunology and Cancer, Department of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, QC, Canada

*to whom correspondence should be addressed: peter.kolb@uni-marburg.de

Abstract

G protein-coupled receptors do not only feature the orthosteric pockets, where most endogenous agonists bind, but also a multitude of other allosteric pockets that have come into the focus as potential binding sites for synthetic modulators. We have investigated 557 GPCR structures to better characterise such pockets by exhaustively docking small molecular probes *in silico* and converting the ensemble of binding locations to pocket-defining volumes. Our analysis confirmed all previously identified pockets and revealed nine previously untargeted sites. In order to test for the feasibility of functional modulation of receptors through binding of a ligand to such sites, we mutated residues in two sites in two model receptors. Moreover, we analysed the correlation of inter-residue contacts with the activation states of receptors and showed that contact patterns closely correlated with activation indeed coincide with these sites.

13 G protein-coupled receptors (GPCRs) have evolved to transduce signals from the outside
14 of a cell to the inside, thereby allowing the cell to respond to changes in its environment.¹ As a
15 consequence of their role as transducers, GPCRs feature at least two interaction sites: one on
16 the extracellular side, sensing the signalling agents (from photons to peptides), the other on
17 the intracellular side, providing a place for the effector proteins to bind.² As the repertoires of
18 extracellular signalling agents and intracellular effector proteins are quite limited, these sites
19 are oftentimes conserved within a receptor subclass. This can pose a challenge to ligand and
20 drug discovery efforts when the treatment of an ailment requires the selective targeting of a
21 particular receptor subtype. An example of such a challenge are the β_1 - and β_2 -adrenergic
22 receptors (β_1 - and β_2 AR), which differ only by a Phe/Tyr substitution in their orthosteric
23 sites. Blockade of the β_1 AR in heart by beta-blockers is desired for cardiovascular disease,
24 but antagonising the β_2 AR in lung tissue is detrimental for asthma. Conversely, stimulation
25 of the β_2 AR helps asthma patients but potentially damages their heart through concomitant
26 agonism of the β_1 AR.

27 As a possible way of circumventing this challenge of highly similar pockets, the targeting
28 of allosteric pockets is billed as a sensible alternative.³ Due to the nature of GPCRs as
29 bundles of seven transmembrane helices that are only relatively loosely coupled,⁴ one could
30 indeed expect that a ligand binding to one of these pockets is able to modulate the response
31 of a receptor. Moreover, it is generally claimed – but has never been shown – that these
32 alternative pockets share lower sequence homology.³ There are examples of individual ligands
33 binding to non-orthosteric sites on a few receptors (e.g. refs⁵⁻⁸), but it is currently unknown
34 to what extent such binding sites exist across the receptorome and how different or similar
35 they are in shape and sequence.

36 In this work, we therefore identified and analysed the ensemble of all discernible pockets
37 – the “pocketome” – of 557 GPCR structures of 113 different receptors. We discovered
38 potential pockets by exhaustive docking of small molecular probes, taking into account the
39 different electrostatics of the solvent-exposed and transmembrane parts of the receptors, and
40 compared these data across all receptors.

41 Based on class A and B1 structures in active and inactive conformations, we computed
42 residue contacts including both backbone and side chain atoms. In doing so, we identified
43 interhelical residue contacts crucial for an active or inactive state of both class A and class
44 B1 GPCRs (we follow the nomenclature in IUPHAR’s “Guide to Pharmacology” and refer
45 to “classes” of GPCRs rather than “families”). We were then able to show that known and
46 orphan allosteric pockets contain such contacts of importance, speaking to the likelihood
47 of their functional relevance. These computational investigations were strengthened with
48 experimental studies of two model class A receptors, the muscarinic acetylcholine receptor

49 M₃ (M₃R) and the β₂AR. Through mutations of two pockets that have not been targeted by
50 a synthetic ligand before, we demonstrate that the residues forming these pockets are indeed
51 involved in receptor activation. Last, but not least, we compare the sequence similarity of
52 the most frequently occurring pockets, thereby providing a quantitative assessment of their
53 overall selectivity potential.

54 This represents the currently most exhaustive analysis of the GPCR pocketome, span-
55 ning receptors from classes A, B1, B2, C, D1, and F. In the present manuscript, we will
56 first describe the receptor alignment and quality control procedure. Second, the docking of
57 the small probes as well as the aggregation of all probe poses to densities will be explained.
58 We will then describe several of the known and as-of-yet-untargeted (“orphan”) sites (ab-
59 breviated as KS and OS, respectively, in the following), and show mutational data speaking
60 to the biological relevance of two of the OS. Finally, we will discuss the implications of our
61 findings and possible avenues for ligand design.

62 Results

63 Probe Docking & Conversion to Volumes

64 Our definition of a pocket is based on the computational docking of small molecules (“probes”;
65 while the probes we used are probably too small to bind strongly to a receptor by them-
66 selves, they represent chemical moieties that are typical for GPCR ligands) to the surface
67 of each GPCR structure individually. We therefore first show the results of our docking
68 calculations and the conversion to volumes before turning to the identified hotspots (the
69 pockets) themselves. Please note that we provide a list of all analysed structures together
70 with additional information as a [.csv] file. Exhaustively docking the 40 small, chemically
71 diverse molecular probes (see Methods and Tab. S1) into 557 structures from 113 distinct
72 receptors, we obtained 1 621 367 poses in total. On average, 2 500-3 300 poses were obtained
73 for each of the receptor structures we docked to. While the lowest number of poses was
74 obtained for PDB 2J4Y (1 776), the highest number was obtained for PDB 4IAR (3 808).
75 This indicates a number of probe poses for all receptors sufficient for further analysis. In a
76 probe-based view, benzimidazole showed the highest average number of poses per structure,
77 while favourable poses for the trimethyl ammonium cation were extracted least frequently
78 (Tab. S1 and Fig. S1). No poses were found for methanamine in PDB 4PY0 and the tetram-
79 ethyl ammonium cation in PDB 4OO9. At the other end, the highest number of poses (200)
80 was obtained for benzimidazole in PDB 6FUF, β -lactam in PDB 6UP7 and indoline in PDB
81 6FUF. All receptor structures were able to host both apolar and polar probes, with no clear
82 trends discernable.

83 To analyse the vast number of docked molecules in a statistical manner, we used our novel
84 volumetric averaging algorithm (see Methods) in order to transform the poses of each docking
85 into visualisable probe density maps. These maps are divided into equal volume elements,
86 each of them giving information about how often a probe atom occupied a particular region.
87 On average, each of the obtained maps consisted of 1 000 000 up to 3 500 000 volume elements.
88 Since we wanted to investigate the density maps for trends across the different receptor
89 classes, maps of individual receptors were added up for each class to yield a single map with
90 higher populations overall.

91 The Pocketome

92 General Distribution of Pockets

93 The class-specific density maps provided with this work can be visualised using Pymol (see
94 supplementary Pymol density [.dx] files and README) and might aid a reader with the

95 following description. Said density maps reveal multiple contiguous regions that represent
96 common cavities on the surface of all GPCRs analysed in this study (Fig. 1). Particularly
97 for class A GPCRs, these pockets are distributed in a notably symmetric manner: both at
98 the intra- and extracellular end of the 7TM bundle, pockets can be seen between each pair of
99 adjacent helices. The density maps for the other classes are somewhat less well-defined and
100 more scattered overall. This is owed to the lower numbers of structures and therefore poorer
101 statistics, as individual structures – and possible deviations in them – carry a relatively
102 higher weight than for the more numerous class A structures.

103 Here, we present only those pockets that we will discuss and examine in depth, whereas
104 the rest of them is described in the Supplementary Information. We chose to focus on three
105 of the largest and – by our analysis – best-defined orphan sites and contrast them with an
106 equal number of known sites, which we picked because they are clearly defined and because
107 they host synthetic ligands. While the vast majority of sites defined by the densities is
108 located at the outward-facing receptor portion (i.e. receptor residues in contact with the
109 membrane), we also were able to identify regions of density inside the 7TM bundle. Of note,
110 in each class, the orthosteric (ORTHO) and adjacent secondary binding pockets (ORTHO1
111 and ORTHO2) can clearly be discerned, making it plausible that the other pockets identified
112 in this work could also host ligands. By aligning our density maps with each other, one can
113 see that the average orthosteric pocket for class C receptors is significantly deeper than for
114 the one of class B1, which again is slightly deeper than the one in class A. This is perfectly
115 consistent with experimental evidence.⁹ Due to the overall higher flexibility and thus often
116 worse resolution of extra- and intracellular loops, pockets found within these regions will not
117 be further analysed or discussed.

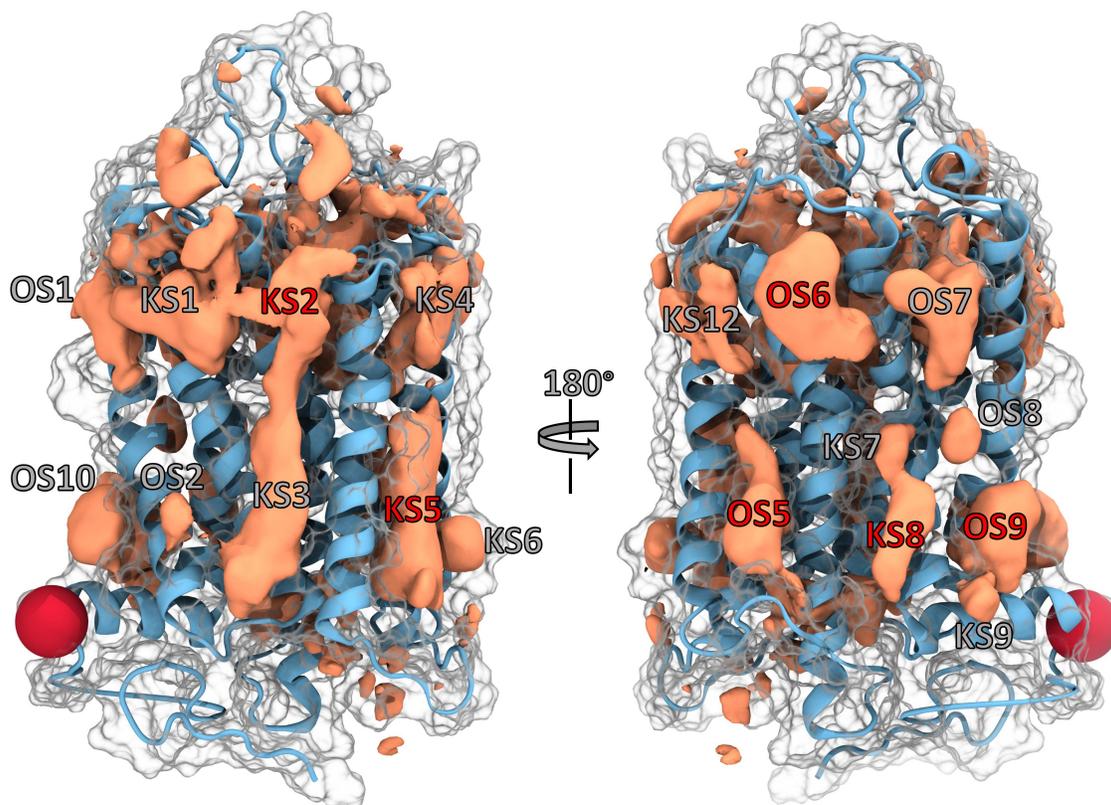


Figure 1: Representative depiction of the GPCR pocketome. Cumulative densities for all class A GPCRs (orange volumes) are shown projected on the structure with PDB 1F88 (ice blue ribbon). The surface is indicated white transparent. Visible hotspots (pockets) located at the lipid-facing receptor portion around the 7TM bundle are labelled either as “OS” (Orphan Site) or “KS” (Known Site). A more detailed description of their location is provided in the text and Tab. 1. We note that OS3 was described in the most recent X-ray structure PDB 7M3J¹⁰ during writing of this manuscript. We therefore re-labeled this site to KS12. Furthermore, OS4 was not found in the class A densities. Three known and three orphan pockets (red labels) are discussed in more detail in the text. The red sphere indicates the tip of HVIII and has been included for ease of orientation.

118 Comparing the densities on the outward-facing receptor portion for all analysed GPCR
119 classes, we assigned pocket identifiers to several volumes that appeared well-defined and
120 clearly distinct from their neighbouring densities. This facilitated later analysis and provided
121 the means for a common orientation and discussion. However, since not only the GPCR
122 structures themselves but also the density map shapes differ across the classes, the reader's
123 view on whether a particular region is an individual pocket might differ from ours. That being
124 said, our general conclusions are independent of any such small differences in definitions. The
125 full list of pockets is presented in Tab. 1. Going around the 7TM bundle, one can observe
126 regions of density at the upper and lower ends between helices V and VI. These sites are
127 referred to as KS12 and OS5, respectively. For some classes, another separated hotspot
128 resides right between these two sites (OS4). At the lower end of the 7TM bundle, OS5 shows
129 a large spot for classes A and F. When directly compared to class A, the density of class B1
130 is subdivided into multiple regions. While for classes B2 and C a small hotspot is visible,
131 class D1 only shows some fragmented density in front of helix V.

132 Another larger spot is visible between helices I and VII above helix VIII for classes A,
133 B1, C and F (OS9). The classes B2 and D1 maps only show a small spot in this region,
134 which might be due to the lack of helix VIII in the available structures.

135 Encouragingly, we identified density near the region of the sodium binding pocket (SODIUM)
136 for some classes. While classes A, B1 and F show somewhat weaker densities, the orthosteric
137 pocket of class C extends down into this region, which makes it clearly defined.

138 Lastly, two regions of density were found at the intracellular portion of the 7TM bundle
139 for all classes. Here, one spot could be identified as the G Protein binding site between
140 helices II, III, V and VI (GPROT). Adjacent to it, density for KS11 resides between helices
141 I, II, VII and VIII.

142 **GPCR States can be Described by their Residue Contact Network**

143 In order to provide evidence that it is possible to achieve modulation of receptor function
144 with a ligand binding to one of the allosteric pockets, we investigated to what extent these
145 pockets are formed by residues that also participate in contact patterns specific for an active
146 or inactive conformation of the receptor. The rationale is that residues which are involved
147 in crucial state-specific contacts are more susceptible to interference by a ligand. In Fig. 2
148 and S2, the principal component analyses of the class A and B1 residue contacts are shown,
149 respectively. Here, we decided to focus on the first two components, since they contributed
150 the most to the overall variance as shown in Fig. S3 and revealed a clear separation of
151 activation states.

152 Across the diagonal of the PC1 vs. PC2 plot for class A, we identified a distribution of

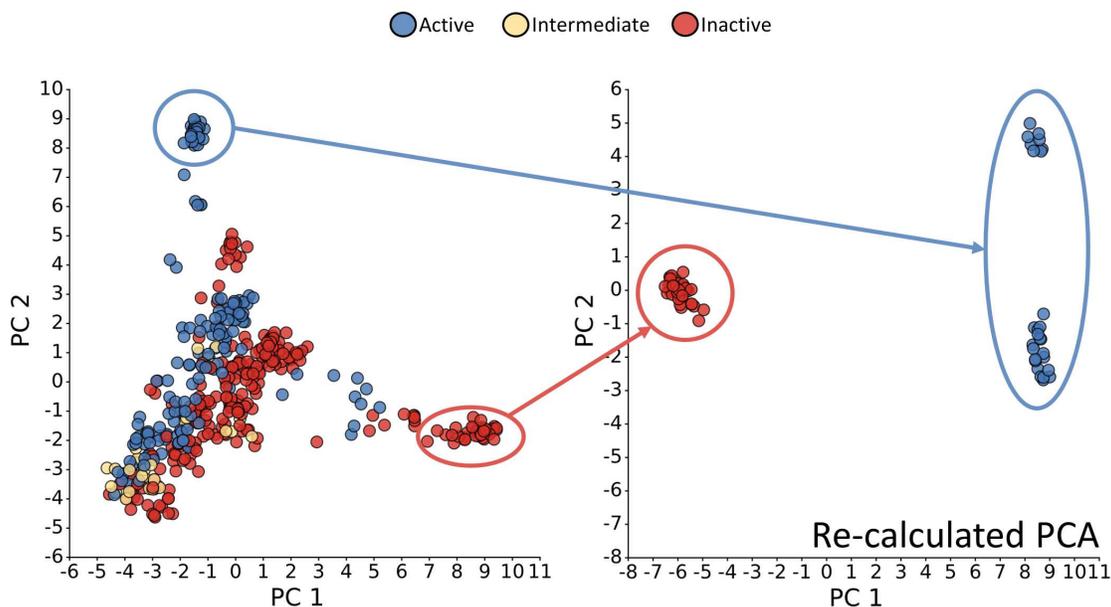


Figure 2: Plot of the first two principal components (PCs) of the residue contact analysis for class A structures. Each point represents a PDB structure. They are colored according to the GPCRdb classification into active (blue), intermediate (yellow), and inactive (red).¹¹ The right panel shows the data re-calculated based on the points in the clearly separated active and inactive clusters. Each PC value for each PDB can be seen as a linear combination of variables (i.e. contacts) that represents the residue contact landscape of a structure in a condensed manner. The first two PCs shown here explained most of the variance across all structural data, hence represent the most interesting PCs for investigating differences between receptor states on the residue contact level.

153 states ranging all the way from structures classified as active to those classified as inactive,
 154 with intermediate structures positioned inbetween, congruent with the assignment of states
 155 in GPCRdb.¹¹ To a certain degree, the large accumulation of structures in the bottom left
 156 shows a mixture of the three classifications. Interestingly, structures classified as inactive are
 157 spread over a wide range of values of PC1, with only small differences in PC2, while active
 158 and intermediate structures display greater variance along PC2. Our contact-map-based
 159 PCA seems to indicate a slightly different view of activation compared to the assignment
 160 in GPCRdb which is based on TM2-TM6 distance cutoffs, the presence of G protein or
 161 arrestin and further similarity measurements. The re-calculated PCA for those points that
 162 are clearly active or inactive according to our measures shows that one principal component
 163 is sufficient in order to explain the difference between the residue contacts of clearly active
 164 and inactive structures.

165 The PCA for class B1 contacts (Fig. S2) shows that the structures classified as active
 166 or inactive are separated along the second principal component. Notably, four structures

167 are separated from the others across the first principal component. As, by the time of this
168 analysis, only one B1 structure with an assignment as an intermediate conformation in the
169 GPCRdb was available, it was not included in the PCA. As for class A, the class B1 PCA
170 was re-calculated considering only those structures belonging to the groups of points clearly
171 classifiable as active or inactive. The four outliers described before were not considered in
172 this recalculation. As expected, the PCA now shows a separation of the states across the
173 most important first principal component.

174 Based on our analysis for two GPCR classes, we show that the structural state of a recep-
175 tor by GPCRdb definition is closely linked to its entire residue contact network. However,
176 we point out that a non-negligible number of class A GPCR structures with a GPCRdb-
177 assignment as active or inactive would fall into the intermediate classification by our contact
178 map categorisation (351 out of 417 class A structures). Hence, the residue contact map of
179 a given structure might provide additional information on top of the GPCRdb definition of
180 a conformational state based on interhelical distances and the type of co-crystallised ligand.
181 Finally, we used the well-separated groups of structures from the re-calculated PCA (Fig. 2,
182 right panel) to extract the most important and conserved active- and inactive-state-specific
183 contacts for each of the sites of interest. We focused on contacts formed between residues of
184 two distinct helices, since such contacts could potentially be targeted by a ligand.

185 **Identification of Known Pockets**

186 As mentioned in the general description (above and Supplementary Information), we found
187 all the allosteric binding sites already known from crystallographic experiments (e.g. refs⁵⁻⁸),
188 which can be considered an excellent validation of the general applicability of our docking-
189 based approach (see Tab. 1). In this section, we want to focus on one exemplary site, describe
190 its conservation across the receptorome and explain possible modes of action by using our
191 residue contact data. Two more sites are discussed in more detail in the Supplementary
192 Information.

193 This known pocket, KS2, is located at the outward-facing residues of the upper ends
194 of helices III and IV. While this site is only known for two class A GPCRs, namely the
195 free fatty acid receptor FFAR1 and protease-activated receptor-2 (PAR2), our density maps
196 show that it seems to be conserved across all GPCR classes. In order to further validate
197 this finding, we analysed the receptorome-wide sequence identity and similarity of residues
198 forming this site by using the definition of Tab. 2. While the matrices in Fig. S4 and S5
199 show that the overall identity is considerably low, the similarity based on physicochemical
200 properties is much higher with an average value above 50% (Fig. S6).

201 We then investigated the interactions of known ligands with KS2 and compared them to

202 our residue contact analysis for class A and class B1 GPCRs. Two cases are known from
203 the available structural data: In the case of the FFAR1 (PDB: 4PHU¹², 5TZR¹³, 5TZY¹³),
204 the agonists fasiglifam and MK-8666 penetrate between the upper ends of helices III and IV
205 coming from the inner portion of the receptor. While being anchored by polar contacts in
206 the orthosteric region, hydrophobic interactions are dominant in KS2. A structure for the
207 PAR2 (PDB: 5NDZ¹⁴) reveals a different mode of binding. Here, the allosteric antagonist
208 AZ3451 stacks against the outward-facing portion of helices III and IV while only making
209 one polar contact to residue 3.30. Instead of pushing the two helices apart, this allosteric
210 ligand seems to hold them together, mainly through hydrophobic interactions. Our class
211 A contact analysis for known sites shown in Fig. S7 revealed multiple helix III-IV contacts
212 crucial for an inactive conformation of the receptor such as 3.23-4.61, 3.27-4.61, 3.30-4.60 and
213 3.34-4.58. Furthermore, this analysis indicated one highly conserved active state contact,
214 3.30-4.61.

Site	Location ^a	Class	PDB ID
OS1	UP OF I,II	A, B1, B2, C, D1, F	
OS2	LP OF I,II	A, B1, B2, C, D1	
OS4	MP OF V,VI	B1, B2, C, D1, F	
OS5	MP-LP OF V,VI	A, B1, B2, C, D1, F	
OS6	UP OF VI,VII	A, B1, C, F	
OS7	UP OF I,VII	A, B1, B2, C, D1, F	
OS8	MP OF I,VII	A, B1, B2, C, F	
OS9	LP OF I,VII,VIII	A, B1, B2, C, D1, F	
OS10	LP OF I,VIII	A, B1, C, F	
ORTHO	UP-MP IF	A, B1, C, D1, F	
ORTHO1	UP IF Above ORTHO IV,V, VI	A, B2, C	A (4MQT, 4XNW, 6RZ5, 6RZ8, 6RZ9, 6U1N)
ORTHO2	UP IF Next to ORTHO I,II,III,VII	A, B1, C, D1	A (3ODU, 3OE6, 3OE8, 3OE9, 3REY, 5NDD, 5T1A, 5U09, 5X33, 5YHL, 5YWY, 6AK3, 6GPS, 6GPX, 6IIU, 6IIV, 6LI0, 6OIK, 6RNK, 6U1N, 6Z10), B1 (6ORV, 6VCB, 6X19, 6X1A, 6XOX, 7C2E)
SODIUM	MP IF I,II,III,VI,VII	A, B1, C, D1, F	A (5X33), C (4OO9, 5CGC, 5CGD, 6FFH, 6FFI)
GPROT	LP IF II,III,V,VI	A, B1, C, D1, F	
KS1	UP OF II,III	A, B1, B2, C, D1, F	A (4XNV)
KS2	UP OF III,IV	A, B1, B2, C, D1, F	A (4PHU, 5NDZ, 5TZR, 5TZY)
KS3	MP-LP OF II,IV	A, B1, B2, C, D1, F	A (6KQI)

KS4	UP OF IV,V	A, B1, B2, C, D1, F	A (6RZ5, 6RZ8, 6RZ9)
KS5	MP-LP OF III,IV,V	A, B1, B2, C, D1, F	A (5KW2, 5O9H, 5TZY, 6C1Q, 6C1R, 6N48, 6OBA, 7CKZ, 7LJC, 7LJD)
KS6	LP OF IV,V	A, B1, C, D1, F	A (6C1R)
KS7	MP OF VI,VII	A, B1, B2, C, F	B1 (6KJV)
KS8	LP OF VI,VII	A, B1, C, D1, F	B1 (5EE7, 5VEW, 5VEX, 5XEZ, 5XF1, 6KJV, 6KK1, 6KK7, 6LN2), C (7CA3)
KS9	LP OF VII,VIII	A, B1, C, D1	A (5X7D), B1 (5EE7, 5VEW, 5VEX, 5XEZ, 5XF1, 6KK1, 6KK7, 6LN2)
KS10	MP IF III,V,VI	B1, B2	B1 (4K5Y, 4Z9G)
KS11	LP IF I,II,VII,VIII	A, B1, B2, C, D1, F	A (5LWE, 5T1A, 5X7D, 6LFL, 6QZH)
KS12	UP OF V,VI	A, B1, B2, C, D1, F	C (7M3J)

^aLP: lower portion, UP: upper portion, MP: middle portion, IF: inward-facing, OF: outward-facing

Table 1: List of all pockets observed after probe docking, their approximate locations and structures in which they are visible. We note that OS3 was described in the most recent X-ray structure PDB 7M3J¹⁰ during writing of this manuscript. We therefore re-labeled this site to KS12.

Site	Residues
OS5	5.51, 5.54, 5.55, 5.58, 5.61, 5.62, 6.35, 6.38, 6.39, 6.41, 6.42, 6.45, 6.46, 6.49
OS6	6.47, 6.50, 6.53, 6.54, 6.57, 7.33, 7.34, 7.37, 7.38, 7.41
OS9	1.45, 1.49, 1.52, 1.53, 1.56, 7.47, 7.50, 7.51, 7.54, 7.55, 8.48, 8.50, 8.51, 8.54
KS2	3.23, 3.26, 3.27, 3.30, 3.31, 3.34, 4.54, 4.57, 4.58, 4.61, 4.62
KS5	3.41, 3.44, 3.45, 3.48, 3.52, 4.41, 4.44, 4.45, 4.48, 4.52, 4.55, 5.45, 5.46, 5.49, 5.50, 5.53, 5.57
KS8	6.35, 6.36, 6.39, 6.40, 6.42, 6.43, 7.48, 7.51, 7.52, 7.56

Table 2: Position of residues making up each of the sites discussed in more detail in the text and Supplementary Information (Ballesteros-Weinstein numbering¹⁵ was chosen for an overall better comparability). The conservation of these residues across all analysed receptors is shown in Fig. S4 and S5.

215 Identification of Orphan Pockets

216 In this section, we will focus on two orphan sites that could represent binding sites for al-
 217 losteric modulators. They are among the best-defined and largest-volume sites that emerged
 218 from our analysis. A third site is discussed in the Supplementary Information. Since no
 219 structural data of ligands binding to these regions is known yet, we will describe the pockets
 220 based on their amino acid sequences and our class A and class B1 residue contact analysis.

221 Similar to the three known sites described in more detail in this work, these pockets
 222 also reside in the outward-facing portion of the receptor. As shown in Fig. S8, they mainly
 223 consist of hydrophobic residues. This is expected, since most of their volume lies within the
 224 membrane portion of the GPCRs. The first orphan site discussed here, OS5, is located at
 225 the lower portion of the 7TM bundle between helices V and VI. Density in this region was
 226 conserved across all GPCR classes. However, our receptorome-wide sequence similarity anal-
 227 ysis reveals that the physicochemical properties slightly differ across the classes. While class
 228 A and B1 receptors share a high sequence similarity with each other in this region, GPCRs
 229 belonging to classes C and F only show high sequence similarity within their respective sub-
 230 class. This fact might explain the different shapes of the densities. In direct comparison
 231 to the other classes, the class F density is shifted more towards the intracellular side of the
 232 GPCR. Hence, the receptorome-wide definition of OS5 (Tab. 2) might not be suitable for
 233 class F receptors. In contrast, the position and shape of OS9 between helices I, VII and VIII
 234 is highly conserved across all GPCR structures. The physicochemical properties of this site
 235 are more conserved across classes A, B1 and C, with an average sequence similarity above
 236 50%. Again, one exception is class F with a much lower homology to the other classes.

237 In order to identify the impact of OS5 and OS9 on the biological function of the receptors,

238 we constructed double or quadruple mutants, where two or four, respectively, of the residues
 239 that form these pockets in the M₃R were changed (residues mutated in a particular mutant
 240 are connected by a grey vertical bar in Fig. 3). These mutants were tested in a BRET-based
 241 G protein activation assay as well as in a FRET-based β -arrestin2 recruitment assay. The
 242 summary of the resulting data is shown in Fig. 3.

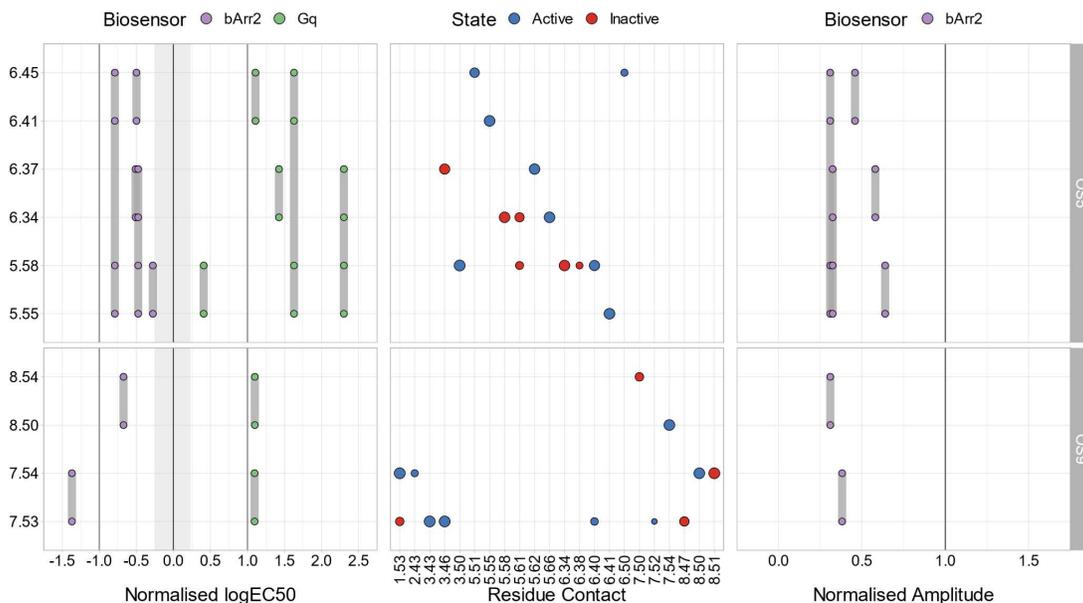


Figure 3: Experimental data for OS5- and OS9-mutants in the M₃R. The y-axis depicts the position of the pocket residues (Ballesteros-Weinstein numbering¹⁵). Grey lines linking the points connect the respective double and quadruple mutants. The left panel shows normalised (to the wt) logEC₅₀ values, and the right panel the normalised amplitude of the extent to which β -arrestin2 was recruited. While in the left plot a value of 0 corresponds to no potency changes, a value of 1 in the right plot corresponds to no changes in efficacy compared to wt. The grayed out area indicates minimal and maximal potency shifts of multiple G_{αq} wt measurements. The central panel depicts interhelical residue contacts of each of the mutated residues that are important in active (blue) or inactive (red) conformations of class A GPCRs in a similar manner to Fig. S9 (normalised PCA coefficient cut-off: 0.5). The point size correlates with the normalised PCA coefficient for a given contact and can be seen as a direct measurement of importance for a given state.

243 Regarding OS5, our results show a clear increase in the logEC₅₀ values of the concentration-
 244 response curve of acetylcholine-induced G_{αq} activation relative to M₃R wt (Fig. 3 left and S10A).
 245 Of note, except for one double mutation, all others resulted in shifts of at least one log unit.
 246 In addition, a marked decrease in the efficacy of acetylcholine to recruit β -arrestin2 to the
 247 mutant M₃Rs was observed, with only a small effect on the logEC₅₀ value. This is consistent
 248 with our finding that residues forming this pocket are involved in rearrangements important
 249 for receptor activation as shown in the middle panel of Fig. 3.

250 The same is true for OS9, where a similar right shift of $G_{\alpha q}$ activation in the mutants
 251 occurs (Fig. 3 and S10B). Similarly, the extent of the recruitment of β -arrestin2 is substan-
 252 tially reduced for all mutations of OS9 (Fig. 3 and S11), without major effects on the $\log EC_{50}$
 253 values, suggesting a major impact of the mutants on agonist efficacy. By demonstrating that
 254 the partial agonist arecoline led to a greatly diminished G protein activation even under
 255 saturating conditions (Fig. S10C), we confirmed that the mutations indeed strongly affect
 256 the efficacy of muscarinic agonists.

257 To further support our findings, we also mutated individual OS5 and OS9 residues in the
 258 β_2 AR. The results are summarised in Fig. 4 and full curves are shown in Fig. S12.

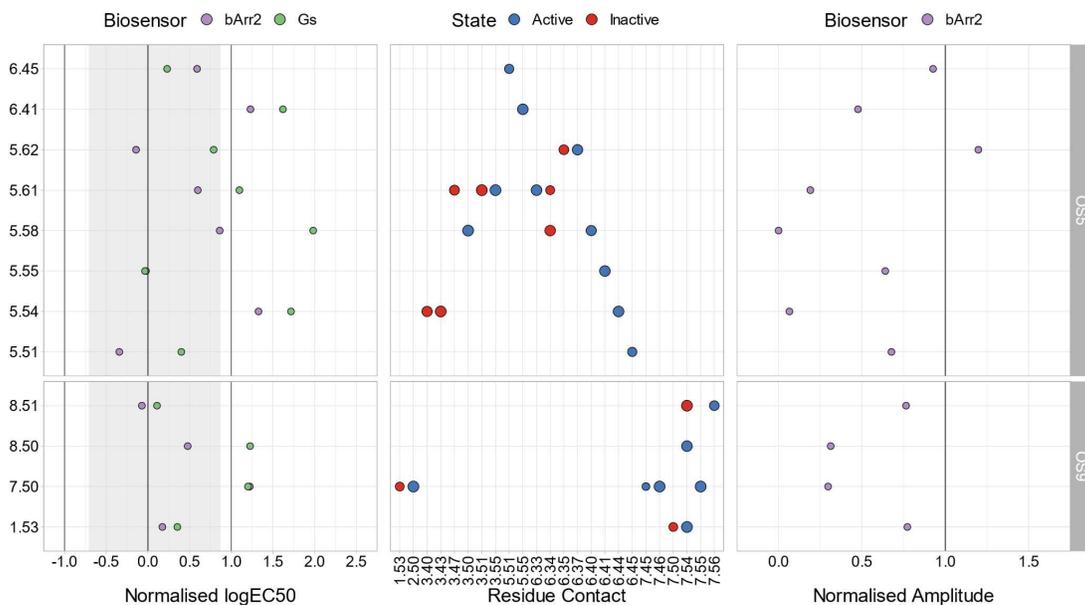


Figure 4: Experimental data for OS5- and OS9-mutants in the β_2 AR. The y-axis depicts the position of the pocket residues (Ballesteros-Weinstein numbering¹⁵). The left panel shows normalised (to the wt) $\log EC_{50}$ values, and the right panel the normalised amplitude of the extent to which β -arrestin2 was recruited. While in the left plot a value of 0 corresponds to no potency changes, a value of 1 in the right plot corresponds to no changes in efficacy compared to wt. The grayed out area indicates minimal and maximal potency shifts of multiple $G_{\alpha s}$ wt measurements. The central panel depicts interhelical residue contacts of each of the mutated residues that are important in active (blue) or inactive (red) conformations of class A GPCRs in a similar manner to Fig. S9 (normalised PCA coefficient cut-off: 0.7). The point size correlates with the normalised PCA coefficient for a given contact and can be seen as a direct measurement of importance for a given state.

259 Similar to our findings for the M_3 R, our results show an increase in the $\log EC_{50}$ values of
 260 adrenaline-induced activation of two different biosensors ($G_{\alpha s}$ and β -arrestin2) and a decrease
 261 in efficacy relative to the β_2 AR wt. Our β_2 AR signalling results show that the $\log EC_{50}$ is

262 increased for either $G_{\alpha s}$ or β -arrestin2 or both for half the mutations across the OS5 and OS9
263 pockets. Similarly, for about half of the mutations, the efficacy for β -arrestin2 recruitment
264 is reduced. As was the case for the M_3R , these results further strengthen the assumption
265 that OS5 and OS9 are indeed physiologically relevant.

266 In order to obtain deeper insight into the possible reasons behind the interference of
267 residues in OS5 and OS9 with GPCR activation, we compared the mutational data with our
268 class A contact analysis (middle panel of Fig. 3 and 4). For OS5, we found that the mutated
269 residues were frequently involved in conserved active and inactive state contacts. Starting
270 with residue 5.54, our analysis revealed two contacts (3.43-5.54 and 5.54-6.44) important
271 for an inactive and active state of the receptor, respectively. The decrease of function upon
272 mutating 5.54 indicates that 5.54-6.44 might be a contact crucial for receptor activation. The
273 same holds true for residues 5.58, 5.61 and 6.41. In our contact analysis, we found that the
274 region formed by these residues includes numerous contacts important for an active state.
275 While the microswitch contacts 3.50-5.58 and 5.58-6.40 are known for their importance for
276 an active conformation of the receptor, 5.61 shows two active-state contacts with 3.55 and
277 6.33, respectively. Furthermore, residue 6.41 makes exactly one crucial active-state contact,
278 namely to 5.55. Overall, our mutagenesis experiments and residue contact analysis suggest
279 that a considerable fraction of the residues of OS5 are involved in key active-state contacts.
280 These residues could therefore be addressed by a synthetic allosteric modulator in order to
281 reduce receptor activity.

282 A largely similar picture holds true for OS9. Here, mutating the conserved residues 7.50
283 and 8.50 led to a decrease of receptor function. Per our contact analysis, both residues are
284 involved in several important active-state contacts such as 2.50-7.50 and 7.54-8.50. Finally,
285 we mutated two residues (6.34 and 6.37) that are not part of the two sites described here
286 by themselves, but are involved in relevant active-state contacts to OS5 and OS9 residues,
287 namely 6.34-5.66 and 5.62-6.37. These mutations led to a decrease in $G_{\alpha q}$ activation in the
288 M_3R , speaking to the occurrence of a second-shell effect.

289 Occupancy of Known Allosteric Pockets

290 In order to obtain deeper insight into the properties of the allosteric pockets, we analysed the
291 known sites for the occurrence of molecules besides (synthetic) ligands designed for them, i.e.
292 focusing on crystallisation additives and co-purified substances. We tabulated all structure
293 determination adjuvants resolved in any of the 557 investigated in this work. Tab. S2 shows
294 the complete results of the analysis.

295 This analysis revealed that in around 35% of the investigated structures, the known

296 allosteric binding sites host additional chemical compounds besides the added orthosteric and
297 allosteric ligands. In particular, a recurrence of cholesterol (or cholesteryl hemisuccinate),
298 oleic acid and glyceryl monooleate can be noticed, all used as adjuvants in the purification
299 and/or structure determination process. Binding is more frequent to the more superficial
300 pockets. To avoid bias introduced by the fact that not all adjuvants are present in all buffers,
301 we also calculated the background distribution for all substances.

302 We found that for 12.4% (69 out of 557) of the structures the known pockets are occu-
303 pied by at least one type of crystallisation adjuvant. In order to avoid a too fine-grained
304 analysis, we grouped the additives into five categories: surfactants, steroids (cholesterol and
305 derivatives), fatty acids, polymers (predominantly PEG and PPG), and anions. This allows
306 us to deduce a preference of the pockets for certain types of components. Of course, not
307 every known pocket was occupied by a component, and, as mentioned above, only a rela-
308 tively small set of structures out of the total 557 contained at least one occupied pocket.
309 Hence, we assume that the presence of a particular chemical moiety can be interpreted as
310 a preference of a pocket rather than a random occurrence attributable to the crystallisation
311 conditions. In other words, if a component was stable enough in order to be resolved in a
312 pocket (and given that the electron density was sufficient to determine this), it constitutes
313 a binding event.

314 A more detailed analysis of the chart showing the number of times a component from
315 one of the five categories was found in a pocket (Fig. 5) elucidates a rather clear picture:
316 The preference of KS11 for anions (26%, or 6 out of 23) is evident, probably fueled by the
317 interactions with R3.50 in this pocket. The next most frequent category are polymers (17.4%
318 or 4 out of 23). In contrast, KS5 hosts a striking abundance of fatty acids, with approximately
319 70% occurrence (45 out of 64), followed by the substantially smaller percentage of $\approx 6\%$ (4
320 out of 64) of steroids.

321 Known pockets KS7 and KS8 are also mostly populated by fatty acids, but less frequently
322 than in the previous case, as they are only observed up to $\approx 43\%$ (or 13 out of 30). In both
323 cases, the preference for this category is favoured by the presence of several aromatic and
324 hydrophobic residues in the middle sections of TM3 and TM4. KS2 is the pocket with the
325 highest value of occupancy overall (12.4% or 69 out of 557), and in this case as well, the
326 most recurrent category are fatty acids ($\approx 48\%$ or 33 out of 69), followed by steroids ($\approx 32\%$
327 or 22 out of 69). KS10 and KS1 are populated the least, only 3 structures each (i.e. less
328 than 1%) contain a component — fatty acids in both cases (33% or 1 out of 3 and 100% or
329 3 out of 3 of the occupied structures, respectively).

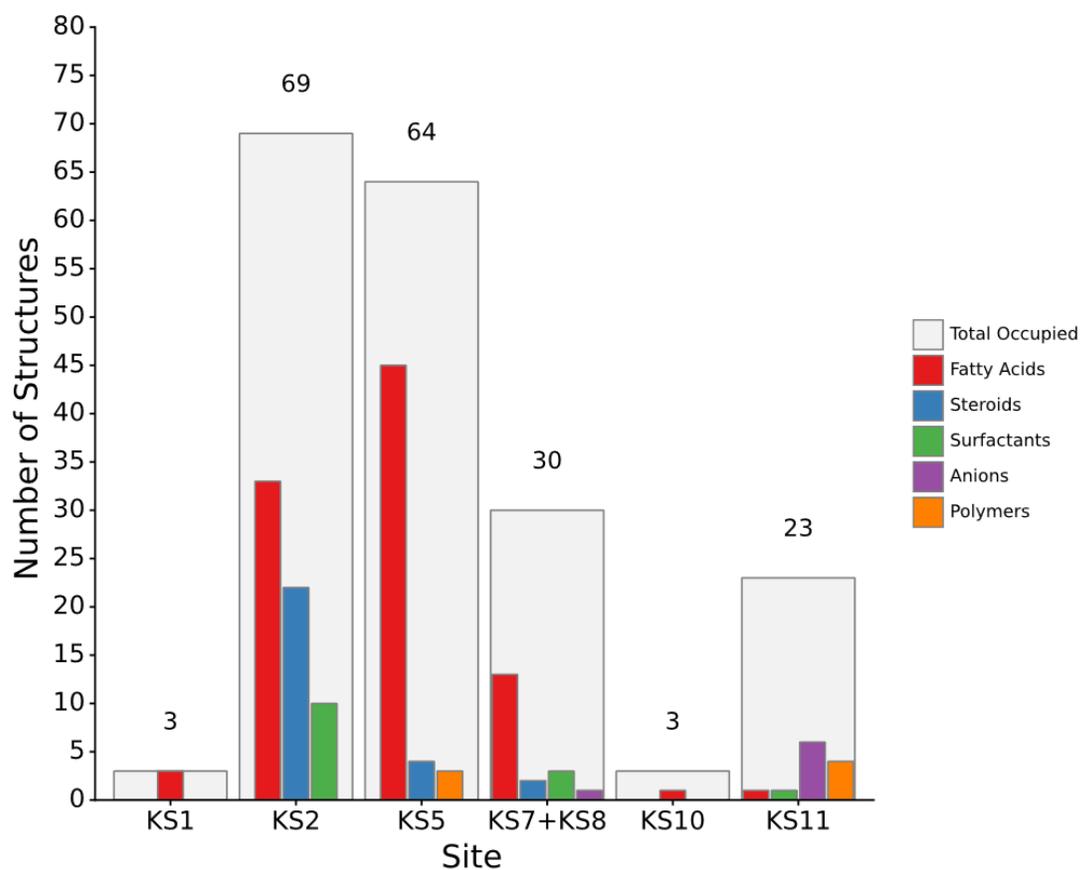


Figure 5: Absolute number of observed instances of components other than synthetic ligands in the KS. Shown are the total number of structures in which at least one component was resolved in each site (light grey bars) and the respective numbers for surfactants (green), anions (purple), fatty acids (red), steroids (blue), and polymers (orange).

330 Discussion

331 In our study, we have determined the occurrence of pockets across the largest part of the
332 currently available structural G protein-coupled receptorome. Because the ultimate goal
333 of this research is to identify ligands for these pockets through which receptors might be
334 modulated in their activity, we chose the docking of small molecular probes, i.e. chemically
335 valid compounds, as a method, rather than definitions of a pocket based on the protein
336 surface. Thus, we optimised our definition towards what a potential ligand would “see”.
337 We explicitly accounted for the different environments of the various receptor portions, *viz.*
338 the membrane-embedded core and the solvent-exposed ends, by using a docking method
339 where the dielectric constant of the environment can be set to the appropriate values, thus
340 avoiding artefacts. A novel aggregation method allowed us to average over the probe docking
341 calculations to all 557 GPCR structures. Hence, the pockets we identified exist in the
342 majority of receptors and the pocketome we present in Fig. 1 thus constitutes a representation
343 of the shape and distribution of frequently observable cavities. In addition, we were able
344 to make statements on the conservation of pockets across receptor classes. Our findings
345 strongly suggest that while the pockets are quite dissimilar in sequence space, they are more
346 similar than generally assumed when focusing on the properties of the amino acids rather
347 than their identity. This will merit attention when designing selective ligands, so as not to
348 rely too much on nondirectional interactions. Despite this overall conservation, we can show
349 that some sites are more conserved regarding their shape and physicochemical properties
350 than others. This is also borne out at the level of the presence of crystallisation additives in
351 the pockets. These characteristics therefore should be exploitable for the design of class- or
352 type-specific allosteric modulators, which has implications when designing ligands for such
353 pockets in order to avoid unintended polypharmacology.

354 We are convinced that the number of structures investigated is such that the general
355 trends observed, at least for class A and class B1, will hold even as new structures become
356 available. In fact, we reran our analysis pipeline shortly before submission, approximately
357 nine months after the first time, and did not observe noticeable changes in the pocket defi-
358 nitions. During this time, the number of class A and B1 structures available increased from
359 404 to 455 and 46 to 55, respectively.

360 While we found all pockets that have previously been localised through structure determi-
361 nation with a ligand, we also identified several that have not yet successfully been targeted,
362 at least according to publicly available data. To demonstrate that the two most prominent
363 “orphan pockets” OS5 and OS9 have potential as target sites for small-molecule modulators,
364 we mutated several of the residues lining these two pockets in two model GPCRs, the M₃R

365 and the β_2 AR. In both cases, mutations had a robust effect on both G protein activation as
366 well as β -arrestin recruitment. Of note, we did not only observe substantial right-shifts of
367 up to more than 10-fold of the concentration-response curves for G protein activation, indi-
368 cating the need for higher ligand concentrations to achieve similar levels of stimulation, but
369 also decreases in the maximum level of response for β -arrestin. These results indeed point
370 towards the modulatory potential of ligands binding at these sites. Moreover, given the fact
371 that $\log EC_{50}$ -values were shifted in opposite directions for $G_{\alpha q}$ and β -arrestin2 at the M_3R ,
372 one might speculate that pathway-selective ligands could be designed for these pockets. A
373 receptor region similar to OS9 has also recently been investigated in the angiotensin II type
374 1 receptor,¹⁶ where it has been termed a cryptic site. As mutations of amino acids led to
375 changes in $G_{\alpha q}$ and β -arrestin2 responses similar to what we show for our receptors, this
376 lends further credibility to our finding that OS9 is a pan-class A pocket.

377 In addition, we compared the experimental findings to our contact analysis, which is based
378 on the residue contact maps of 557 GPCR structures. We found that ligands addressing the
379 two pockets could potentially act as negative allosteric modulators (NAMs) by disrupting
380 contacts crucial for an active state. This was expected, since both OS5 and OS9 reside
381 near the G protein binding site, which is known for undergoing profound rearrangements
382 upon receptor activation. However, it also seems plausible that by strategically targeting
383 the inactive-state contacts or stabilising active-state contacts, one could potentially design
384 positive allosteric modulators (PAMs) for both pockets.

385 Our probe docking allows us to make observations also for several of the known pockets.
386 In the case of KS2, the available structural data from the FFAR1 and the PAR2 suggests
387 that either positive or negative modulation of agonism at a receptor could be achieved in this
388 pocket by separating or keeping in place, respectively, the upper ends of helices III and IV
389 with a small molecule ligand. This hypothesis is supported by our contact analysis shown in
390 Fig. S7. KS2 contains highly conserved inactive-state contacts between helices III and IV, e.g.
391 3.23-4.61, 3.27-4.61, 3.30-4.60, and 3.34-4.58, and one crucial active-state contact, 3.30-4.61.
392 By either breaking these contacts or keeping them intact with hydrophobic interactions, an
393 allosteric ligand binding to KS2 could modulate the activation state of a GPCR.

394 From the analysis of the occurrence of crystallisation additives in the known sites, we were
395 able to draw conclusions about the preferences of each. Overall, this analysis showed that
396 fatty acids, in particular monoolein and oleic acid, are the most prevalent type of additives
397 occupying the analysed known pockets and that KS5 is the pocket where they are located
398 most frequently. This was to be expected, as the pockets are located in the membrane, but
399 might on the other hand point towards a role of certain pockets in the regulation of function
400 through lipid bilayer composition. A notable exception is KS11, which prefers anions.

401 Similar rationales as for KS2 above can be applied to the design of ligands for the orphan
402 sites. Of course, in these cases, the challenge will be to unequivocally determine the binding
403 locations of such ligands and to design assays that are fast yet precise enough to be utilised
404 in their optimisation. We certainly hope that the three-dimensional atlas laid out in this
405 work will aid the community in achieving this goal.

406 **Online Methods**

407 **In Silico Pocket Identification**

408 If not stated otherwise, all operations in this workflow were scripted using python 3.7 and
409 bash. The python packages `requests` (version 2.25)¹⁷ and `urllib3` (version 1.25.11) were
410 used in order to access the REST API of websites listed below. The retrieved data was han-
411 dled using `pandas` (version 1.1.4).¹⁸ All protein structures and sequence data were handled
412 in `Biopython` (version 1.78)¹⁹ and `BioPandas` (version 0.2.7).²⁰ Mathematical operations
413 were carried out using `NumPy` (version 1.19.4).²¹ Open-source PyMOL (version 2.3.0)²² and
414 Visual Molecular Dynamics (VMD, version 1.9.3)²³ were used for the visualisation and fur-
415 ther editing of protein structures and volumes. Any other type of data was visualised using
416 `plotnine` (version 0.8.0)²⁴ and RStudio 1.4.1717.²⁵

417 **Collection of Structural Information**

418 Information about all available GPCR structures was fetched from the GPCRdb,²⁶ UniProt,²⁷
419 and the Protein Data Bank²⁸ by accessing their REST API. The data most relevant for our
420 work was extracted from GPCRdb and included the PDB identification code, UniProt en-
421 try name, class, activation state and preferred chain of each GPCR structure. This data
422 was enriched by fetching the accession numbers and canonical amino acid sequences from
423 UniProt. In order to correctly assign solvent-accessible and intra-membrane regions at a
424 later stage of the workflow, information about positioning of GPCR residues relative to the
425 membrane (inside or outside) was also included. For a more convenient and uniform han-
426 dling of GPCR amino acid sequences, the canonical amino acid sequences were mapped to
427 their respective class-specific GPCR numbering scheme. By accessing the GPCRdb generic
428 residue number tables,²⁹ the Ballesteros-Weinstein¹⁵, Wootten³⁰, Pin³¹, Wang,³² and fungal
429 numbering schemes were utilised for class A, B, C, F and D1 GPCRs, respectively. Finally,
430 the PDB-formatted structures were retrieved from the Protein Data Bank.

431 **Preparation of Structures for Docking**

432 For each structure, the transmembrane portion and adjacent motifs belonging to the GPCR
433 were separated from all non-native insertions by using the information about the preferred
434 chain retrieved from the GPCRdb and the DBREF tag in the PDB file. Residues listed in the
435 SEQADV section as expression tags and insertions were not considered. For residues that
436 were resolved in multiple conformations, only the first conformation was extracted. Struc-
437 tures that contained a faulty or non-uniform DBREF or SEQADV section were manually

438 corrected before extraction. After visually inspecting the extracted portions, the structures
439 were prepared by using the Molecular Operating Environment (MOE, version 2020.09) soft-
440 ware.³³ Here, incomplete residues were built utilising the “Structure Preparation” function.
441 Termini and chain breaks that contained only one atom were removed. The built-in method
442 “Protonate3D” was used to assign protonation states to histidine and cysteine residues. For
443 consistency, all other residues were assigned their most frequent protonation state under
444 physiological conditions.

445 Preparations were continued using CHARMM together with the CHARMM36 protein
446 force field.³⁴ Termini and breaks were capped by adding ACE and NME caps to the N- and
447 C-terminal ends, respectively. Hydrogen atoms were placed with the HBUILD command. In
448 order to remove too close van-der-Waals contacts, an energy minimisation was carried out for
449 each structure with a short 20-step steepest-descent optimisation followed by an adopted-
450 basis Newton-Raphson optimisation until convergence. In order to keep as much original
451 structural information as possible, only the side chains of formerly incomplete residues and
452 the backbone and caps of terminal residues were allowed to move. Hydrogen atoms were
453 rebuilt using HBUILD again after all previous operations. Then, structures were aligned by
454 using the “cealign” algorithm as implemented in Pymol. Finally, structures were converted
455 to MOL2 file format using UCSF Chimera.³⁵ The correct CHARMM atom types and charges
456 were reassigned based on the information from the CHARMM PSF output file.

457 **Preparation of Molecular Probes for Docking**

458 In order to exhaustively scan the receptors for possible binding sites, a diverse set of small
459 molecular probes was assembled. Diversity was achieved by including probes with differ-
460 ent physicochemical properties such as size, charge and hydrogen bond acceptor/donor
461 distribution. Forty probes were selected as representatives of different functional groups
462 (Tab. S1) and their protonation states were calculated at physiological condition using the
463 ChemAxon Software Solution (Calculator Plugins, Marvin 20.10).³⁶ MOL2 3D-conformers
464 were generated with OpenEye’s OMEGA2 and default settings.³⁷ Next, CGenFF4.0 pa-
465 rameters were generated for each probe by using the CGenFF webservice accessible via
466 <https://cgenff.umaryland.edu/>.³⁸ In order to update the MOL2 files with the CGenFF pa-
467 rameters and prepare a SEED 4.1.2-ready library, scripts from the SEED 4.1.2 repository³⁹
468 were used.

469 Docking Calculations with SEED

470 For each structure, two docking calculations were carried out using SEED 4.1.2.³⁹ For the
471 first docking calculation, only the intramembranous residues were considered and the di-
472 electric constant of the surrounding medium was set to 3.0 in order to better reflect the
473 lipid bilayer. The second docking calculation only considered the solvent-accessible residues
474 and the solvent dielectric constant was set to 78.5, the value for water. The SEED search
475 algorithm works by exhaustively matching multiple copies of each molecular probe to the
476 polar and apolar portions of the defined region. The poses are then spatially clustered and
477 evaluated with energy models that also account for receptor and fragment desolvation. The
478 maximum number of allowed clusters per probe was set to 2000 and only the best-ranked
479 pose per cluster was considered for the output. All other parameters and settings were used
480 with their default values.

481 Extraction of Molecular Features

482 In order to aggregate and average the information from the SEED³⁹ docking calculations
483 to volumes, a custom software was developed and applied. Within this tool, docking poses
484 are searched for substructures relevant for protein:ligand interactions using RDKit⁴⁰ and
485 the cartesian coordinates, atom types, molecule identity, and substructure are stored. The
486 substructures are hydrogen bond donors, hydrogen bond acceptors, aromatic atoms, halogen
487 atoms, basic substructures, acidic substructures, aliphatic rings and an everything substruc-
488 ture (SMARTS are listed in Tab. S3). The docking poses output by SEED were used to
489 construct three-dimensional grids of a user-specified voxel spacing s_v (0.5 Å in this work),
490 encompassing all molecules. For each substructure investigated, a separate grid was con-
491 structed. To reduce the influence of arbitrary parameters such as the precise grid placement
492 and voxel boundaries, each occurrence of a substructure was not only recorded in the grid
493 voxel it was directly located in, but also – with a fractional value – in neighbouring voxels. A
494 distance-dependent dampening factor ensured that the majority of the change introduced in
495 the grid is still recorded close to the grid voxel the substructure was primarily located in. In
496 practice, each recording operation will affect four different types of grid voxels: The center
497 voxel (in which the substructure is located); six directly adjacent voxels sharing a surface
498 with the center voxel, at a distance d of s_v ; twelve voxels at $d = \sqrt{2}s_v$; and eight voxels at
499 $d = \sqrt{3}s_v$. In each grid voxel of a type, an equal change v is introduced, which is multiplied
500 by a dampening factor t and a distance penalty of $1/d$ — except for the center voxel, where no
501 dampening factor is applied. The change v is chosen such that the overall change introduced
502 in the grid is equal to 1, i.e. $\sum_{voxels} v \cdot t \cdot 1/d = 1$. In the present work, the variables used led

503 to 83.34 % of each change being applied to the neighboring voxels and 16.66 % to the center
504 voxel. Using this data, it is possible to average and visualise the areas in which each feature
505 is frequently represented for any number of docking calculations. The grids can either be
506 exported as a PDB file containing dummy atoms at the voxel centers that correspond to a
507 user-given percentage of the sum of each grid or by exporting a grid file using GridDataFor-
508 mats (<https://griddataformats.readthedocs.io/en/latest/gridData/formats.html>) which can
509 be opened in commonly used molecular visualisation tools. It is also possible to calculate and
510 save the grids of a single docking calculation and then combine multiple docking calculation
511 outputs. The potential problem of grids that are not aligned is solved by constructing a
512 master grid that is encompassing all individual grids. The values of grid voxels in the single
513 grids are then added to the master grid using the volume overlap of the grid voxel in the
514 eight respective grid voxels of the master grid. In this way, we were able to calculate average
515 grids and volumes across arbitrary combinations of structures, e.g. for each GPCR class.

516 **Definition of Allosteric Pockets**

517 In the following, our approach of obtaining a generalised, receptorome-wide definition for
518 each site discussed in this work is described. First, reference structures were selected for
519 each class (A: 1F88, B1: 5EE7, C: 7CA3, F: 4JKV). Since class B2 and D1 structures
520 were heavily underrepresented in our data set, they were not considered for this generalised
521 definition. In addition to our density maps, all structures that have an allosteric ligand
522 were visualised. Then, all structures and maps were aligned to rhodopsin (PDB: 1F88).
523 The region around each density and allosteric ligand was examined and matched with the
524 residues of the reference structures. For better comparability and in order to obtain a
525 receptor-wide definition, the site definitions for each class were converted to Ballesteros-
526 Weinstein numbers¹⁵ using the GPCRdb residue tables. For each site, only the residues that
527 occurred for at least two classes were used for the final definition.

528 **Sequence Analysis**

529 The amino acid sequence of known and orphan allosteric pockets described here was analysed
530 across the receptorome in order to determine the degree of conservation of these pockets in
531 the GPCR spectrum. Only sequences of receptors that are structurally resolved were taken
532 into account. For each of the sites discussed in this work, the amino acid sequence was
533 extracted for each receptor by using the site definitions described above and the GPCRdb
534 residue tables. For each pair of receptors, the sequence identity and sequence similarity were
535 calculated. In order to determine the sequence similarity, the following classifications were

536 used: polar, apolar, positively or negatively charged, aromatic. Furthermore, the overall site
537 polarity was calculated by averaging the ratio of polar and apolar amino acids of each pocket
538 across all receptors analysed.

539 **Occupancy of Known Allosteric Pockets**

540 In order to verify whether the known allosteric sites identified here are also occupied by
541 other types of compounds that could influence our results, e.g. crystallisation additives, an
542 alignment of all 557 investigated GPCR structures was performed, using rhodopsin as main
543 template. Every binding site occupied by a known allosteric ligand was visually inspected.
544 The occurrence of the different components in the selected known pockets was then collected,
545 grouped, and analysed. A text-based analysis of the crystallisation conditions stated in all
546 pdb files was also performed to retrieve the background distribution and use it as reference.

547 **Biological Relevance of Pockets — M₃R**

548 **Materials**

549 DMEM, penicillin/streptomycin, FCS, L-glutamine, PBS and trypsin-EDTA were purchased
550 from Capricorn Scientific GmbH, Ebsdorfergrund, Germany. Poly-L-Lysine hydrobromide,
551 PEI and acetylcholine iodide were acquired from Sigma-Aldrich, Merck KGaA, Darmstadt,
552 Germany. Arecoline hydrobromide was purchased from TCI Chemicals, Eschborn, Germany.
553 Coelenterazine h was obtained from NanoLight Technologies, Pinetop, USA.

554 **Plasmids**

555 cDNAs for G_{αq}-YFP,⁴¹ Gβ1,⁴² GRK2,⁴³ β-arrestin2-mTurq,⁴⁴ and M₃-mCit⁴⁵ were described
556 previously. The human M₃R was obtained from the Missouri S&T cDNA Resource Center.
557 pNluc-Gγ2 was a kind gift from Dr. N. Lambert (Augusta University, Georgia, USA). The
558 mCit-β-arrestin2-mTurq was analogously cloned as described for similar reference constructs
559 in Dorsch *et al.*⁴⁶ The M₃R mutants were generated from these plasmids by mutagenesis
560 using the following primers:

5' CAAGGAGAAGAAAGCGGGCCAGACCCGCCAGTGCGATCTTGCTTG 3' (M3_A490G_L493A)
5' ACCCTCAGTGCGATCGCGCTTGCCTTCGCCATCACTTGGACCCC 3' (M3_L497A_I501A)
561 5' GCCTGTCACCATATGGCTATTTTAGCCTGGAGGATCTATAAGG 3' (M3_T248A_Y251A)
5' GTGAACCCCGTGTGCGCTGGTCTGTGCAACAAAACA 3' (M3_Y544A_A545G)
5' CTGTGCAACAAAACAGCCAGAACCACTGCCAAGATGCTGCTGCTG 3' (M3_F551A_F555A)

562 The M₃Rs containing four mutations were cloned analogously in two steps. The M₃R-mCit
563 mutants were generated in the same way.

564 Cell Culture and Transfection

565 All experiments were performed in HEK293T cells. Cells were cultured at 37 °C and 5 % CO₂
566 in Dulbecco's modified Eagle's medium (4.5 g/L glucose), supplemented with 100 units/mL
567 penicillin, 0.1 mg/mL streptomycin, 2 mM L-glutamine and 10 % FCS. The cells were tran-
568 siently transfected in a 6 cm dish using linear polyethylenimine (PEI) 25 kDa as the transfect-
569 ing agent. For the G_{αq} activation, HEK293T cells were transfected with 1.5 μg M₃R wt/M₃R
570 mutants, 2.4 μg G_{αq}-YFP, 0.75 μg G_{β1}, 0.75 μg GRK2 and 0.3 μg pNluc-G_{γ2}. For β-arrestin2
571 recruitment, the cells were transfected with 1.5 μg M₃R-mCit/M₃R-mCit mutants, 1.5 μg
572 β-arrestin2-mTurq and 0.75 μg GRK2. For the quantification of relative expression levels
573 1.5 μg mCit-β-arrestin2-mTurq were transfected. The ratio of DNA and PEI was determined
574 as 1 to 3. For 1 μg DNA 50 μL DMEM w/o FCS were added to the DNA and PEI solutions.
575 Both solutions were mixed, incubated at 20 °C for 30 minutes, being protected against light
576 and were afterwards added to the HEK293T cells in a 6 cm dish. For the BRET-based G_{αq}
577 activation, cells were counted after 24 h and 16000 cells/well were seeded into a poly-L-lysine
578 coated 96-well plate (Greiner 96 Flat White). For the FRET-based β-arrestin2 recruitment
579 the cells were plated into six-well plates with poly-L-lysine coated 25 mm coverslips 24 h after
580 transfection. All experiments were performed 48 h after transfection at room temperature.

581 BRET-based Measurements

582 Transiently transfected adherent HEK293T cells were measured in a 96-well plate with a
583 Spark 20M Multimode Microplate Reader (Tecan), using the luciferase reporter Nluc.⁴⁷
584 G_{αq} activation was assessed with G_{αq}-YFP/G_{β1}/pNluc-G_{γ2} biosensors in the presence of
585 M₃R wt/M₃R mutants.⁴⁸ Fluorescence and luminescence intensities were acquired using the
586 Spark-Control application and the BRET emission ratio was calculated as the YFP signal
587 (light emission between 520 nm and 700 nm) divided by the Nluc signal (light emission be-
588 tween 415 nm and 485 nm). In general, sixteen wells were measured in one round. Cells
589 were washed once with extracellular buffer (137 mM NaCl, 5.4 mM KCl, 2 mM CaCl₂, 1 mM
590 MgCl₂, 10 mM HEPES, pH 7.3) and 80 μL of a 3.07 μM solution of Coelenterazine h in buffer
591 were added to every well. After 10 minutes of incubation, 10 cycles of baseline measure-
592 ment were performed with a duration of about 6.5 minutes all together. The measurement
593 was paused shortly, 20 μL buffer or agonist in buffer were added and 10 cycles of agonist
594 measurement were performed. Afterwards 20 μL agonist in a saturating concentration was
595 added and 10 cycles of BRET measurement were performed once again. The agonist-induced
596 change in BRET emission ratio was calculated as the difference in average values of the third
597 cycles before and after adding the agonist. The additional change in BRET emission ratio

598 induced by a saturating concentration of acetylcholine was calculated as the difference in
599 average values of the third cycles before and after adding the saturating concentration of
600 acetylcholine. The maximum change in BRET emission ratio was calculated as the sum of
601 the agonist-induced change and the additional change induced by the saturating concentra-
602 tion of acetylcholine. The agonist-induced change was normalized to the maximum change
603 in BRET emission ratio for every well. Concentration response curves were fitted by Graph-
604 Pad Prism 8.3 with variable slopes. The bottom constrained to 0 and the top to 1 were
605 determined for the concentration response curves of acetylcholine. In order to calculate the
606 cut-off area shown in Fig. 3, all single concentration response curves of M₃R wt were plotted
607 individually and the minimal and maximal EC₅₀ values of wt measurements were identified.

608 **Single-Cell FRET Imaging**

609 A FRET-based assay was used to measure the agonist-induced interaction between β -arrestin2-
610 mTurq and M₃R-mCit wt/mutant sensors.⁴⁹ The measurements were performed as previ-
611 ously described by Milde *et al.*,⁵⁰ except where declared otherwise, using an inverted flu-
612 orescence microscope (Eclipse Ti, Nikon, Germany). The cells were excited with an LED
613 excitation system (pE-2; CoolLED, UK) at 425 nm and 500 nm. The intensity of both LEDs
614 was set to 2%. The fluorescence intensity was measured using the software NIS-Elements
615 advanced research (Nikon Corporation) and the image recording frequency was set to 2 Hz.
616 FRET emission ratio was calculated as the ratio of mCitrine intensity divided by mTurquoise
617 intensity upon excitation of mTurquoise at 425 nm by plotting over time. All fluorescence
618 data were corrected for background fluorescence, bleed-through and false mCitrine excitation.
619 The measurements were additionally baseline-corrected for photobleaching, using Origin Pro
620 9.5 (Originlab, USA). The cells were constantly superfused with either extracellular buffer
621 (described in BRET-based Measurement) or acetylcholine. Every cell was stimulated for 30
622 seconds with each concentration of acetylcholine. The concentration-dependent change in
623 FRET emission ratio were calculated as the average value of the last 5 seconds of stimulat-
624 ing with each concentration of acetylcholine. Concentration response curves were fitted by
625 GraphPad Prism 8.3 with variable slopes.

626 **Quantification and Correction of Relative Expression Levels**

627 The relative expression level of M₃R-mCit and β -arrestin2-mTurquoise was corrected, using
628 the construct mCit- β -arrestin2-mTurq for calibration of the stoichiometry. mTurquoise was
629 excited with 425 nm whereas mCitrine was excited with 500 nm. The fluorescence intensity
630 was measured and corrected for background fluorescence. The calibration factor was calcu-

631 lated as $F_{\text{mCitrine}}/F_{\text{mTurquoise}}$. For each single-cell measurement, the factor was calculated in
632 the same way. Due to the influence on the extent of FRET signal, the relative expression level
633 of M₃R-mCit and β -arrestin2-mTurquoise was corrected for an equal stoichiometry. There-
634 fore, the factor of every single-cell FRET-measurement was divided by the calibration factor
635 (Fig. S15) and every measurement was multiplied with its individual reciprocal (Fig. S11).

636 **Biological Relevance of Pockets — β_2 AR**

637 **Plasmids and Mutagenesis**

638 Human β_2 AR and all biosensor constructs were assembled in pcDNA3.1. The β_2 AR was
639 codon-optimised and a SNAP tag and an N-terminal signal sequence were attached at the
640 N-terminus. The biosensor plasmids are based on Renilla luciferase (RlucII) and a GFP, ei-
641 ther GFP10 or Renilla GFP (rGFP) as described previously.^{51–53} Single-point mutants of the
642 β_2 AR were generated as described in an earlier work.⁵⁴ Non-alanine amino acids were mu-
643 tated to alanine, native alanine residues were mutated to glycine. Primers were designed us-
644 ing custom software as described earlier⁵⁵ (available at: <https://github.com/dmitryveprintsev/AAScan>).

645 **BRET-based Signaling Assays**

646 All assays were using human embryonic kidney (HEK)-293 SL cells (a gift from Stephane
647 Laporte). Cells were grown at 37 °C with 5% CO₂ in DMEM with 4.5 g/L glucose, L-
648 glutamine, and 10% newborn calf serum (NCS, Wisent BioProducts, Canada) and penicillin-
649 streptomycin (PS, Wisent BioProducts, Canada). Two days prior to measurements, cells
650 were transfected using polyethyleneimine (PEI, Polysciences Inc., Canada, No. 23966), with
651 a ratio between PEI and DNA of 3:1. Afterwards, 20 000 cells per well were seeded into
652 white Cellstar PS 96-well cell culture plates (Greiner Bio-One, Germany). On the day of
653 the measurement, medium was removed and Tyrode's buffer (137 mM NaCl, 0.9 mM KCl,
654 1 mM MgCl₂, 11.9 mM NaHCO₃, 3.6 mM NaH₂PO₄, 25 mM Hepes, 5.5 mM glucose, 1 mM
655 CaCl₂, pH 7.4) was added, followed by incubation at 37 °C of at least 30 min. Ten minutes
656 before measurement, adrenaline was added, with concentrations ranging from 31.6 nM to
657 3.16 mM in half-log steps, as well as a buffer control. At 5 min prior to measurement,
658 coelenterazine 400a (DeepBlueC, Nanolight Technology) was added for a final concentration
659 of 5 μ M. Coelenterazine 400a was initially dissolved in DMSO and diluted into Tyrode's
660 buffer with 1% Pluronic F-127 for increased solubility. BRET was measured in a Synergy
661 Neo microplate reader (Biotek) using detection at 410 nm and 515 nm. All experiments were
662 done at least in biological triplicates. Cut-off areas shown in Fig 4 were calculated in a
663 manner similar to the M₃R.

664 **Residue Contact Analysis**

665 **Calculation of Residue Contact Maps**

666 For all class A and class B1 structures, residue contact maps were calculated only considering
667 the transmembrane portions. A contact was defined to occur when the distance between any
668 two atoms of two distinct residues was smaller than the sum of their van der Waals radii plus a
669 buffer distance of 0.5 Å. In order to prevent sampling of local contacts which might introduce
670 noise at later stages of the analysis, contacts between residues less than four positions apart
671 in sequence and where one of the atoms involved was in the backbone were not considered.

672 **Creation of Contact Fingerprints**

673 In order to describe the residue contact distribution of each GPCR in a simplistic manner,
674 a class-specific contact fingerprint was calculated for every structure. First, for each class of
675 GPCRs, the set of all residue-residue contacts that occurred in at least one of its members
676 was compiled. Only contacts that were observable in all activation states were considered.
677 Further, this set of contacts was treated as a fingerprint in which the individual bits were set
678 to either 1 or 0 depending whether or not a particular contact occurred in a structure. Finally,
679 for each of our 557 structures, the appropriate class-specific fingerprint was calculated, the
680 aggregate of which was then used to determine activation networks.

681 **Principal Component Analysis**

682 A class-specific principal component analysis was carried out based on the contact finger-
683 prints using the python package `Scikit-learn` (version 0.23.2).⁵⁶ Here, each structure can
684 be seen as a sample while each contact can be seen as a variable. The contribution of each
685 of the first 10 principal components to the overall variance of the data was plotted and eval-
686 uated. Then, PCA plots were created for the principal components that explained most of
687 the variance. The data points, each of them representing one PDB structure, were colored
688 according to their activation state as ascribed in the GPCRdb. Principal components that
689 showed a clear separation between the active and the inactive state were used in order to
690 classify contacts important for the respective state. For each contact, the sign and the abso-
691 lute value of the pertaining principal component coefficient gave the necessary information
692 about the state and importance, respectively.

693 The procedures described above and the PCA were carried out again on those structures
694 to which a clear inactive or active state could be assigned in the first PCA. This ensured that
695 no intermediate structures were considered for the following analysis. For the class A PCA,

696 we decided on structures with PC1 values larger than 7 and PC2 values larger than 7.5 for
697 the inactive and active state, respectively. For class B1, structures with PC1 values larger
698 than 3 were not considered for the re-calculation. Since the re-calculated class B1 PCA still
699 showed two outliers, we decided on eliminating them (PDB: 6NIY and 6P9X) from the PCA
700 before continuing with the network analysis.

701 **Contact Analysis**

702 For both classes, the PCA coefficients were used in order to estimate the importance of a
703 contact for a certain receptor state. Here, we focused on those contacts that were formed
704 between residues in KS2, KS5, KS8, OS5, OS6 and OS9. Contacts between two residues
705 that belong to the same helix were not considered. By investigating the re-calculated PCA
706 plots (Fig. 2 and S2), each contact was considered either as an active or inactive contact
707 depending on the sign of its corresponding PCA coefficient. The PCA coefficients were
708 normalised to their highest absolute value such that they ranged from 0 (not important) to
709 ± 1 (important). For each pocket, the residue contacts together with their normalised PCA
710 coefficient were plotted.

711 **Acknowledgements**

712 We thank Daniel Hilger and Anthony Watts for helpful discussions. P.K. thanks the German
713 Research Foundation DFG for Heisenberg Professorship KO4095/5-1. The doctoral theses
714 of J.B.H., M.P., and K.B. are funded by the LOEWE project “GLUE” (GPCR Ligands for
715 Underexplored Epitopes). F.M.H. is funded by a Marie Skłodowska-Curie Individual Fel-
716 lowship from the European Union’s Horizon 2020 research and innovation programme under
717 grant agreement No 844622 and was previously funded by the American Heart Associa-
718 tion’s Grant #19POST34380839. M.B. (Michel Bouvier) is funded by a Canadian Institute
719 of Health Research Foundation Grant # 148431 and holds the Canada Research Chair in
720 Signal Transduction and Molecular Pharmacology. F.M.H., M.B. (Moritz Bünemann), and
721 P.K. are members of COST Action CA18133 “ERNEST” (European Research NEtwork on
722 Signal Transduction).

723 **Author Contributions**

724 J.B.H., M.P., F.M.H., T.G., M.B. (Moritz Bünemann) and P.K. designed research; J.B.H.,
725 M.P., K.B., and F.M.H. performed calculations and experiments; J.B.H., M.P., K.B., F.M.H.,
726 M.B. (Michel Bouvier), M.B. (Moritz Bünemann) and P.K. analysed data; F.M.H., M.B.
727 (Michel Bouvier), M.B. (Moritz Bünemann) and P.K. acquired funding; M.B. (Michel Bou-
728 vier), M.B. (Moritz Bünemann) and P.K. supervised research; all authors wrote the manuscript.

729 **Competing Interests**

730 M.B. (Michel Bouvier) is the president of the scientific advisory board of Domain Therapeu-
731 tics, a biotechnology company which licenses BRET-based biosensors for commercial use.
732 The biosensors used in the present study are freely available for academic research through
733 material transfer agreements. All other authors have no conflicts of interest to declare.

734 **Supplementary Information**

735 Supplementary Results and Discussion, as well as Tables and Figures are available as a pdf.
736 Separate pymol sessions for each class are provided, as well as the raw data to display in a
737 different viewer. A list of the 557 structures is provided in .csv format. All files are online
738 at . . .

739 Data Availability

740 The datasets generated and analysed during the current study (e.g. aligned receptor struc-
741 tures) are available from the corresponding author on reasonable request.

742 Code Availability

743 The structure retrieval pipeline is available from the corresponding author upon reasonable
744 request. SEED 4.1.2 is available from <http://www.biochem-caflisch.uzh.ch/download>.
745 The code for the volumetric averaging software is available at
746 <https://github.com/torbengutermuth/volumetricaveraging>.

References

- [1] Hilger, D., Masureel, M. & Kobilka, B. K. Structure and dynamics of GPCR signaling complexes. *Nat. Struct. Mol. Biol.* **25**, 4–12 (2018).
- [2] Weis, W. I. & Kobilka, B. K. The molecular basis of G protein-coupled receptor activation. *Annu. Rev. Biochem.* **87**, 897–919 (2018).
- [3] Christopoulos, A. Advances in G protein-coupled receptor allostery: From function to structure. *Mol. Phys.* **86**, 463–478 (2014).
- [4] Hilger, D. The role of structural dynamics in GPCR-mediated signaling. *FEBS J.* **288**, 2461–2489 (2021).
- [5] Oswald, C. *et al.* Intracellular allosteric antagonism of the CCR9 receptor. *Nature* **540**, 462–465 (2016).
- [6] Zheng, Y. *et al.* Structure of CC chemokine receptor 2 with orthosteric and allosteric antagonists. *Nature* **540**, 458–461 (2016).
- [7] Liu, X. *et al.* Mechanism of intracellular allosteric β 2AR antagonist revealed by X-ray crystal structure. *Nature* **548**, 480–484 (2017).
- [8] Song, G. *et al.* Human GLP-1 receptor transmembrane domain structure in complex with allosteric modulators. *Nature* **546**, 312–315 (2017).
- [9] Munk, C., Harpsøe, K., Hauser, A. S., Isberg, V. & Gloriam, D. E. Integrating structural and mutagenesis data to elucidate GPCR ligand binding. *Curr. Opin. Pharmacol.* **30**, 51–58 (2016).
- [10] Gao, Y. *et al.* Asymmetric activation of the calcium-sensing receptor homodimer. *Nature* 1–5 (2021).

- [11] Kooistra, A. J. *et al.* GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Res.* **49**, D335–D343 (2020).
- [12] Srivastava, A. *et al.* High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature* **513**, 124–127 (2014).
- [13] Lu, J. *et al.* Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40. *Nat. Struct. Mol. Biol.* **24**, 570–577 (2017).
- [14] Cheng, R. K. *et al.* Structural insight into allosteric modulation of protease-activated receptor 2. *Nature* **545**, 112–115 (2017).
- [15] Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In Sealfon, S. C. (ed.) *Receptor Molecular Biology*, vol. 25 of *Methods in Neurosciences*, chap. 19, 366–428 (Academic Press, San Diego, CA, 1995).
- [16] Lu, S. *et al.* Activation pathway of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design. *Nat. Commun.* **12**, 4721 (2021).
- [17] Reitz, K. *Requests: HTTP for Humans* (2021). URL <https://docs.python-requests.org>.
- [18] pandas development team, T. pandas-dev/pandas: Pandas (2020). URL <https://doi.org/10.5281/zenodo.3509134>.
- [19] Cock, P. J. *et al.* Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
- [20] Raschka, S. Biopandas: Working with molecular structures in pandas dataframes. *The Journal of Open Source Software* **2** (2017). URL <http://dx.doi.org/10.21105/joss.00279>.
- [21] Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
- [22] DeLano, W. *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA, USA, 2002).
- [23] Humphrey, W., Dalke, A. & Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
- [24] Kibirige, H. *et al.* has2k1/plotnine: v0.8.0 (2021). URL <https://doi.org/10.5281/zenodo.4636791>.
- [25] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA (2020). URL <http://www.rstudio.com/>.
- [26] Munk, C. *et al.* GPCRdb: the G protein-coupled receptor database—an introduction. *Br. J. Pharmacol.* **173**, 2195–2207 (2016).

- [27] The UniProt Consortium. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
- [28] Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
- [29] Isberg, V. *et al.* GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res.* **44**, D356–D364 (2015).
- [30] Wootten, D., Simms, J., Miller, L. J., Christopoulos, A. & Sexton, P. M. Polar transmembrane interactions drive formation of ligand-specific and signal pathway-biased family BG protein-coupled receptor conformations. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5211–5216 (2013).
- [31] Pin, J.-P., Galvez, T. & Prézeau, L. Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. *Pharmacol. Ther.* **98**, 325–354 (2003).
- [32] Wang, C. *et al.* Structural basis for Smoothed receptor modulation and chemoresistance to anticancer drugs. *Nat. Commun.* **5**, 1–11 (2014).
- [33] Chemical Computing Group ULC 1010 Sherbooke St. West Suite 910, C. H. R., Montreal QC. Molecular Operating Environment (MOE) (2018). URL <https://www.chemcomp.com>.
- [34] Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
- [35] Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- [36] ChemAxon. *Calculator Plugins* 16.1.11. URL <http://www.chemaxon.com>. (accessed Jan 11, 2016).
- [37] Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *J. Chem. Inf. Model.* **50**, 572–584 (2010).
- [38] Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
- [39] Majeux, N., Scarsi, M. & Caffisch, A. Efficient electrostatic solvation model for protein-docking. *Proteins: Struct., Funct., Bioinf.* **42**, 256–268 (2001).
- [40] Landrum, G. RDKit: Open-source cheminformatics. URL <http://www.rdkit.org>. (accessed Nov 15, 2014).
- [41] Hughes, S. J. *et al.* Fragment based discovery of a novel and selective PI3 kinase inhibitor. *Bioorg. Med. Chem. Lett.* **21**, 6586–6590 (2011).

- [42] Bünemann, M., Frank, M. & Lohse, M. J. Gi protein activation in intact cells involves subunit rearrangement rather than dissociation. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 16077–16082 (2003).
- [43] Winstel, R., Freund, S., Krasel, C., Hoppe, E. & Lohse, M. J. Protein kinase cross-talk: membrane targeting of the beta-adrenergic receptor kinase by protein kinase C. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 2105–2109 (1996).
- [44] Krasel, C., Bünemann, M., Lorenz, K. & Lohse, M. J. β -arrestin binding to the β 2-adrenergic receptor requires both receptor phosphorylation and receptor activation. *J. Biol. Chem.* **280**, 9528–9535 (2005).
- [45] Jelinek, V., Mösslein, N. & Bünemann, M. Structures in G proteins important for subtype selective receptor binding and subsequent activation. *Commun. Biol.* **4**, 1–11 (2021).
- [46] Dorsch, S., Klotz, K.-N., Engelhardt, S., Lohse, M. J. & Bünemann, M. Analysis of receptor oligomerization by FRAP microscopy. *Nat. Methods* **6**, 225–230 (2009).
- [47] Hall, M. P. *et al.* Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *J. Am. Chem. Soc.* **7**, 1848–1857 (2012).
- [48] Thomsen, A. R. *et al.* GPCR-G protein- β -arrestin super-complex mediates sustained G protein signaling. *Cell* **166**, 907–919 (2016).
- [49] Rinne, A., Mobarec, J. C., Mahaut-Smith, M., Kolb, P. & Bünemann, M. The mode of agonist binding to a G protein-coupled receptor switches the effect that voltage changes have on signaling. *Sci. Signal.* **8**, ra110 (2015).
- [50] Milde, M., Rinne, A., Wunder, F., Engelhardt, S. & Bünemann, M. Dynamics of G α i1 interaction with type 5 adenylate cyclase reveal the molecular basis for high sensitivity of Gi-mediated inhibition of cAMP production. *Biochem. J.* **454**, 515–523 (2013).
- [51] Galés, C. *et al.* Real-time monitoring of receptor and G-protein interactions in living cells. *Nat. Methods* **2**, 177–184 (2005).
- [52] Galés, C. *et al.* Probing the activation-promoted structural rearrangements in preassembled receptor-G protein complexes. *Nat. Struct. Mol. Biol.* **13**, 778–786 (2006).
- [53] Namkung, Y. *et al.* Monitoring G protein-coupled receptor and β -arrestin trafficking in live cells using enhanced bystander BRET. *Nat. Commun.* **7**, 12178 (2016).
- [54] Heydenreich, F. M. *et al.* High-throughput mutagenesis using a two-fragment PCR approach. *Sci. Rep.* **7**, 6787 (2017).
- [55] Sun, D. *et al.* AAscan, PCRdesign and MutantChecker: a suite of programs for primer design and sequence analysis for high-throughput scanning mutagenesis. *PLOS ONE* **8**, e78878 (2013).
- [56] Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [classb2.dx.zip](#)
- [63.gpcrpocketome.readme.20210902.pdf](#)
- [receptors.csv](#)
- [classb1.dx.zip](#)
- [63.gpcrpocketome.si.20210902.pdf](#)
- [classa.dx.zip](#)
- [classc.dx.zip](#)
- [classd1.dx.zip](#)
- [classf.dx.zip](#)
- [63.gpcrpocketome.si.20210902.pdf](#)
- [63.gpcrpocketome.readme.20210902.pdf](#)