Beyond canonical binding sites: A receptorome-wide atlas of GPCR surface cavities and co-resolved binding hotspots

Supplementary Information

Janik B. Hedderich[1,*], Peter Kolb[1,*]

[1]Philipps-Universität Marburg, Faculty of Pharmacy, Department of Pharmaceutical Chemistry, Marbacher Weg 8, 35032 Marburg, Germany

[*]To whom correspondence should be addressed:
janik.hedderich@uni-marburg.de (J.B.H.), peter.kolb@uni-marburg.de (P.K.)

Supplementary Results

Location clustering

Bayesian optimization selected the upper bound of the cluster range ($K = 60$), achieving a best silhouette score of 0.3932. The optimized distance weighting strongly emphasized the segment component: with $\alpha = 0.01$, $\beta = 0.01$, and $\gamma = 1.0$, the effective weights after renormalization were $\alpha' \approx 0.0098$, $\beta' \approx 0.0098$, and $\gamma' \approx 0.9804$ (i.e., $\sim$98% segment). Smoothing parameters indicated minimal smoothing for region and orientation ($\sigma_R = \sigma_O = 0.01$) and substantially stronger smoothing for segments ($\sigma_S = 1.0$). Similarities were computed as $S = \exp(-\tau D^2)$ with $\tau = 0.1$ and sparsified with a kNN graph using $k = 30$. The resulting clustering was highly stable across random seeds (mean ARI = 1.0000) and remained consistent under 80% subsampling (bootstrapped mean ARI = 0.9220). Although the model produced 60 distinct clusters, only 48 unique human-readable cluster labels were obtained from medoid-derived labeling, indicating that multiple clusters mapped to the same label (likely due to overlapping feature distributions or identical medoid label profiles under the applied labeling thresholds).

A comprehensive database and coordinate archive support GPCR cavity and binding site exploration

All data associated with receptors and detected cavities is readily accessible through a comprehensive SQLite database. The database schema is structured to enable efficient querying and downstream analyses, supporting rapid exploration of candidate allosteric sites across the GPCR landscape. The database integrates structure- and protein-level annotations from widely used resources, including the UniProt Knowledgebase (UniProtKB),[1] the Protein Data Bank (PDB),[2, 3] and GPCRdb.[4] Furthermore, we provide three separate PDB files per structure: receptor coordinates, co-resolved chemical entity coordinates, and dummy atoms representing the detected cavities. These files can be readily visualized in molecular viewers such as PyMOL[5] or Chimera.[6] This allows users to quickly inspect cavity locations and properties in the context of receptor structures. The database and cavity coordinate

files can be downloaded from our repository[7] and are also available for direct download through the web interface described below.

Activation-state differences in class A cavity volumes

Besides investigating state-specific differences in cavity occurrence, we additionally analyzed differences in cavity-volume distributions for all location clusters observed in at least 30 active and 30 inactive structures. Supplementary Figure S5 shows the resulting boxplots for the clusters hosting cavities with the most significant volume differences between active and inactive states (Mann–Whitney U test,[8] BH-FDR corrected, $q < 0.05$).

Across all location clusters with significant state-specific differences, cavity volumes were generally larger in inactive-state structures. Moreover, these volume shifts were consistent with the enrichment analysis above, with cavities assigned to clusters that were more enriched in inactive-state structures also tending to exhibit larger inactive-state volumes. Although the distribution for TM-CP/EH-NH/H1-H7-H8 is shifted towards larger volumes in inactive-state structures, active-state structures can still display substantial cavities in this region. This may explain why a PAM was observed to be assigned to this location cluster despite its overall enrichment in inactive-state structures. Notably, the G protein binding site cluster TM-CP/IH-EH/H3-H5-H6 showed larger volumes in the active state, but was excluded from Supplementary Figure S5 due to the presence of several very large cavities that skewed the scale. This observation is consistent with the known opening of this site upon activation to accommodate G protein binding.

In summary, the volume analysis supports the patterns observed in the state-stratified cavity occurrence analysis, indicating that activation modulates not only where cavities occur but also their size at specific locations.

Cavities show distinct location preferences across receptor classes

In the following section, we analyzed the distribution of detected cavities across GPCR classes. Class D was excluded due to the presence of only a single structure in our dataset. Furthermore, we combined classes B1 and B2 into a single class B to enable a more robust statistical analysis. For this analysis, we did not account for the activation state of the structures, which may bias cavities associated with specific states.

In Supplementary Figure S2A, a heatmap summarizing the prevalence of each location cluster across GPCR classes A, B, C, and F is shown. Prevalence is defined as the fraction of structures (within a given receptor class) that contain at least one detected cavity from the respective location cluster. This analysis therefore highlights locations that commonly contain a cavity within each receptor class (i.e., locations that represent a structually conserved cavity). Overlaid star markers indicate class-cluster combinations for which so far no functional ligands have been assigned to a particular location cluster in the ligand-annotated dataset, i.e., as-of-yet untargeted cavities.

From the heatmap, we can immediately observe that cavities assigned to two location clusters, EC-TM/IH-EH/H3-E2-H7 and TM-CP/IH-EH/H3-H5-H6, are highly prevalent across all receptor classes. These clusters correspond to the intrahelical binding site (canonically defined as orthosteric for class A GPCRs) and the G protein binding site, respectively.

While the high prevalence of cavities assigned to the former cluster is expected for class A receptors, it is notable that this location cluster is also common in class C receptors, where this pocket is known to act as an allosteric site. The high prevalence of the G protein binding site cluster across all classes is expected given its universal role in GPCR signaling. While the overall prevalence of cavities assigned to extracellular location clusters (cluster labels starting with 'EC') is low for classes A and B, it is notably higher for class C and F receptors.

By calculating Spearman correlations between class-specific prevalence profiles, we found that location-cluster profiles of detected cavities are overall quite distinct across receptor classes. The highest correlation was observed between classes B and F (0.63), whereas classes A and C showed the lowest correlation (0.43) — that is, classes B and F have many cavities in common, whereas A and C do not. The corresponding correlation matrix is shown in Supplementary Figure S6. With respect to functional ligands, the largest fraction of molecules that have experimentally been shown to bind outside the canonical intrahelical binding site are found in class A receptors. This likely reflects that class A receptors are the most extensively studied GPCRs in structural biology (Supplementary Table S1). Nonetheless, our analysis suggests substantial potential for discovering novel allosteric modulators targeting cavities assigned to the "orphan" location clusters (denoted with a star in Supplementary Figure S2A), especially in classes B, C, and F.

While this analysis provides a comprehensive overview of cavity location preferences across GPCR classes, it does not address whether cavities assigned to specific clusters are significantly more prevalent in one class over the other classes. To test this, we performed a statistical enrichment analysis using Fisher's exact test[9] for each cluster-class combination. P-values were adjusted for multiple testing using the Benjamini-Hochberg procedure,[10] and the results are summarized in SupplementaryFigure S2B.

We found that cavities in the transmembrane region ('TM' is part of the cluster label) within the helical bundle ('IH' in cluster label) were predominantly enriched in class A receptors. In contrast, extracellular cavities ('EC' in cluster label) near the N-terminal segments ('NT' in cluster label) and/or extracellular loops ('E1-E3' in cluster label) were more enriched in classes C and F. For class B receptors, cavities were mainly enriched in the extrahelical region ('EH' in cluster label) of the cytoplasmic half of the transmembrane bundle ('CP' in cluster label) compared with other classes, while we also observed enrichment of a cluster in the extracellular domain (EC/NH/NT).

Of note, we found that in a large fraction of non-class-A structures the transmembrane bundle exhibits wider openings at both the extracellular and cytoplasmic ends, which leads to the detection of large cavities that span multiple clusters. In these cases, the intrahelical binding site, the G protein interface as well as the respective loop and termini segments were often connected, resulting in a single continuous cavity extending through the transmembrane bundle (as shown in Supplementary Figure S7). As a result, the observed prevalences and enrichments of cavities assigned to clusters located within the transmembrane bundle or near the loops and termini in non-class-A receptors may be influenced by this trend.

In summary, we observed distinct preferences for cavity locations across GPCR classes. While, as expected, the intrahelical binding site and the G protein interface were the most prevalent and conserved location clusters, overall cavity profiles differed along the transmembrane axis: extracellular cavities were more prevalent in classes C and F, whereas

transmembrane cavities were more common in class A. The correlation analysis supported these findings by indicating low similarity between cavity location profiles across classes. Finally, enrichment analysis highlighted location clusters significantly overrepresented in specific classes, pointing to potentially class-specific binding sites as opportunities for ligand discovery.

Supplementary Discussion

Location clustering

The selected hyperparameters indicate that segment composition dominates the learned grouping structure, while region and orientation provide only minor refinements. Consistently, the optimized weights allocate nearly all influence to the segment component, and the smoothing strengths imply that both region and orientation are treated almost discretely ($\sigma_R = \sigma_O = 0.01$), whereas segments benefit from stronger smoothing ($\sigma_S = 1.0$), likely reflecting the larger and noisier segment vocabulary.

The observation that $K = 60$ clusters yielded only 48 unique medoid-derived cluster labels suggests that the clustering resolves finer subdivisions than are captured by the downstream labeling rules. In practice, distinct clusters can share the same dominant region/orientation/segment terms (or even the same medoid label profile after thresholding), leading to label collisions. This can be interpreted as either (i) meaningful substructure within a broader category (e.g., subtle differences in secondary terms or in the similarity graph) or (ii) partial overlap between clusters in the chosen feature space. The high seed stability (mean ARI = 1.00) and strong subsample stability (ARI $\approx$ 0.92) argue that these subdivisions are not numerical artifacts, but rather reproducible partitions of the data. If a one-to-one mapping between clusters and labels is desired, the label extraction thresholds (minimum share, secondary ratio, segment cap) can be tightened or expanded, or alternatively multiple medoids per cluster can be summarized to better reflect within-cluster heterogeneity.

Cavities show distinct location preferences across receptor classes

Across GPCR classes, the highest-prevalence clusters correspond to the intrahelical, i.e. the class A canonical orthosteric, site and the G protein interface, indicating that our localization framework recapitulates conserved functional binding sites. At the same time, prevalence profiles and enrichment patterns differ markedly between classes, with class A enriched for intrahelical transmembrane cavities and classes C and F showing comparatively higher representation of extracellular clusters near N-terminal and loop segments. The apparent depletion of cavity clusters in the intrahelical transmembrane region of non-class-A receptors may be explained by our observation that cavities in this region are often merged into a single large cavity that is assigned to a different cluster. These class-dependent patterns are consistent with known architectural and pharmacological differences across the superfamily and suggest that cavity landscapes encode family-specific opportunities for ligand discovery. Notably, a class A-orthosteric-like intrahelical cluster is also prevalent in class C receptors, where the corresponding location is known to act as an allosteric site. This supports the hy-

pothesis that spatially similar cavity locations can be functionally repurposed across receptor classes, and it motivates cross-class comparisons of ligandability even when pharmacological roles diverge.

Methodological considerations and limitations

Several limitations should be considered when placing these findings in context. First, structural coverage remains uneven across GPCR classes, with class A dominating available structures and ligand-bound complexes; this likely inflates apparent richness of non-orthosteric binding events in class A and leaves many clusters in other classes effectively unexplored (see Supplementary Table S1 and Supplementary Figure S2). Also, even within well-covered classes, available structures represent a biased subset of the conformational landscape, with overrepresentation of certain receptor subtypes, ligand types, and activation states. Additionally, we did not apply structure- or residue-level resolution cutoffs and instead used receptor coordinates as deposited. As a result, some structures contain missing residues or unresolved parts, which may affect cavity detection and clustering. Given these coverage limitations, we refrained from generalized ligandability predictions and instead focused on providing a comprehensive atlas of cavity and binding-site profiles, together with statistically robust analyses to help guide future research efforts. Second, our geometry-based cavity detection is inherently structure-centric and does not assess druggability or capture cryptic pockets that may only appear transiently in dynamic ensembles. Furthermore, we used a single parameter set for both cavity detection and clustering. While this enables a consistent analysis across the entire structural receptorome, it may also introduce artifactual differences between structure subsets (for example, large intrahelical cavities in non-class-A receptors may be assigned to a single cluster, whereas analogous regions are resolved as separate cavity clusters in class A structures). Although such artifacts may be partially mitigated by the fact that most structures in our dataset belong to class A, it is important to view clustering primarily as a pragmatic framework for comparing cavity locations across the receptorome rather than as a definitive partitioning of all cavities into unique sites. Future work could explore alternative parameter settings or clustering algorithms to assess the robustness of the identified clusters and their functional interpretations. Third, co-resolved lipids, steranes and ions can reflect detergents, crystallization conditions, or other experimental factors, and thus may not always represent physiological occupancy. Finally, the discrete active/inactive annotations extracted from the GPCRdb may mask intermediate conformations and context-dependent signaling states that influence cavity formation.

Supplementary Methods

Statistical analysis

All downstream analyses were performed by querying the SQLite database and processing the resulting tables in Python. We combined structure metadata, predicted cavity tables, and co-resolved chemical entity tables by PDB identifier and propagated receptor class and activation-state annotations from GPCRdb. Unless stated otherwise, class D was excluded due to insufficient coverage, and classes B1 and B2 were combined into class B.

Cluster prevalence. Cluster prevalence across receptor classes and activation states was computed by treating a cluster c as *present* in a structure u if at least one predicted cavity in that structure was assigned to c. Formally, letting $z(u)$ be the set of clusters assigned to cavities in structure u, we define the presence indicator

$$I_{u,c} \;=\; \mathbf{1}[\, c \in z(u) \,]. \tag{S1}$$

For a group g (receptor class or activation state), let U_g be the set of unique structures in that group. The prevalence of cluster c in group g was defined as the fraction of structures in g that contained c:

$$\mathrm{prev}(c, g) \;=\; \frac{1}{|U_g|} \sum_{u \in U_g} I_{u,c}. \tag{S2}$$

Similarity of prevalence profiles between receptor classes was quantified using Spearman rank correlation[11] (as implemented for pandas dataframes). For two classes g_1, g_2, letting $\mathbf{v}_g = (\mathrm{prev}(c_1, g), \ldots, \mathrm{prev}(c_m, g))$ be the vector of prevalences over all clusters, Spearman's ρ was computed as the Pearson correlation of the ranks:

$$\rho(g_1, g_2) \;=\; \mathrm{corr}\Big(\mathrm{rank}(\mathbf{v}_{g_1}), \mathrm{rank}(\mathbf{v}_{g_2})\Big). \tag{S3}$$

Enrichment of clusters in receptor classes and activation states. To identify clusters enriched in specific receptor classes or activation states, we performed Fisher's exact tests[9] (SciPy implementation) for each cluster–group pair (c, g) using structure-level presence and absence. For each pair we formed a 2×2 contingency table over structures:

	cluster present ($I_{u,c} = 1$)	cluster absent ($I_{u,c} = 0$)
in group g	A	B
not in group g	C	D

where

$$A = \sum_{u \in U_g} I_{u,c}, \quad B = |U_g| - A, \quad C = \sum_{u \in U_{\neg g}} I_{u,c}, \quad D = |U_{\neg g}| - C, \tag{S4}$$

and $U_{\neg g}$ denotes all structures not in g. Fisher's exact test was applied to this table to obtain a two-sided p-value.

Effect sizes were reported as $\log_2$ odds ratios. The odds ratio (OR) for the table above is

$$\mathrm{OR} \;=\; \frac{A\,D}{B\,C} \tag{S5}$$

To avoid undefined odds ratios in sparse settings, we applied the Haldane–Anscombe correction[12, 13] by adding 0.5 to all cells prior to OR calculation:

$$\mathrm{OR}_{\mathrm{HA}} \;=\; \frac{(A+0.5)(D+0.5)}{(B+0.5)(C+0.5)}, \tag{S6}$$

P-values were adjusted for multiple testing using the Benjamini–Hochberg procedure[10] (BH-FDR, statsmodels implementation) and considered significant at $q < 0.05$.

Entity-type composition and enrichment. For entity-type analyses, we focused on functional ligands, ions, lipids, and sterane-containing molecules and excluded water and remaining other entities. Let $\mathcal{T}$ denote the set of included entity types. For each cluster c and type $t \in \mathcal{T}$, we counted the number of unique structures in which at least one chemical entity of type t mapped to cluster c. Let $T(u,c) \subseteq \mathcal{T}$ denote the set of entity types observed in structure u that map to cluster c, and define the indicator

$$E_{u,c,t} \;=\; \mathbf{1}[\, t \in T(u,c)\,] \tag{S7}$$

After filtering and de-duplication by structure, cluster, and type, the structure count for type t in cluster c is

$$n_{c,t} \;=\; \sum_{u \in U} E_{u,c,t} \tag{S8}$$

with U the set of all structures considered in this analysis. To summarize entity-type composition within a cluster, we computed the (non-exclusive) share of type t as

$$\mathrm{share}(c,t) \;=\; \frac{n_{c,t}}{\sum_{t' \in \mathcal{T}} n_{c,t'}} \tag{S9}$$

Because a structure can contribute to multiple types within the same cluster, these shares describe composition rather than exclusive occupancy.

To test entity-type enrichment, we compared the frequency of a given type t in cluster c against its frequency across all other clusters using Fisher's exact tests (SciPy implementation). Let

$$G \;=\; \sum_{c} \sum_{t \in \mathcal{T}} n_{c,t}, \qquad N_t \;=\; \sum_{c} n_{c,t}, \qquad N_{\neg t} \;=\; G - N_t \tag{S10}$$

denote the grand total across all cluster–type structure counts and the corresponding totals for type t and non-t, respectively. For each cluster–type pair (c,t) we formed the 2×2 contingency table

	type t	not type t
cluster c	A	B
other clusters	C	D

where

$$A = n_{c,t}, \quad B = \sum_{t' \neq t} n_{c,t'}, \quad C = N_t - A, \quad D = N_{\neg t} - B. \tag{S11}$$

Fisher's exact tests were applied to these tables (one-sided enrichment), reporting $\log_2$ odds ratios with Haldane–Anscombe correction and BH-FDR adjusted significance ($q < 0.05$).

State-dependent cavity size differences (class A). To quantify state-dependent differences in cavity size for class A, we compared cavity volume distributions between active and inactive structures for each cluster. Only clusters observed in at least 30 active and 30 inactive structures were considered. Volumes were compared using two-sided Mann–Whitney U tests[8] (as implemented in SciPy) followed by BH-FDR correction ($q < 0.05$).

Ligand modulatory effect (class A). For class A, we summarized functional ligand mappings by modulatory effect by counting, for each cluster and effect label, the number of unique structures with at least one mapped ligand. These counts were visualized as heatmaps, with empty cells indicating no mapped ligands for the corresponding cluster–effect combination.

Supplementary Figures

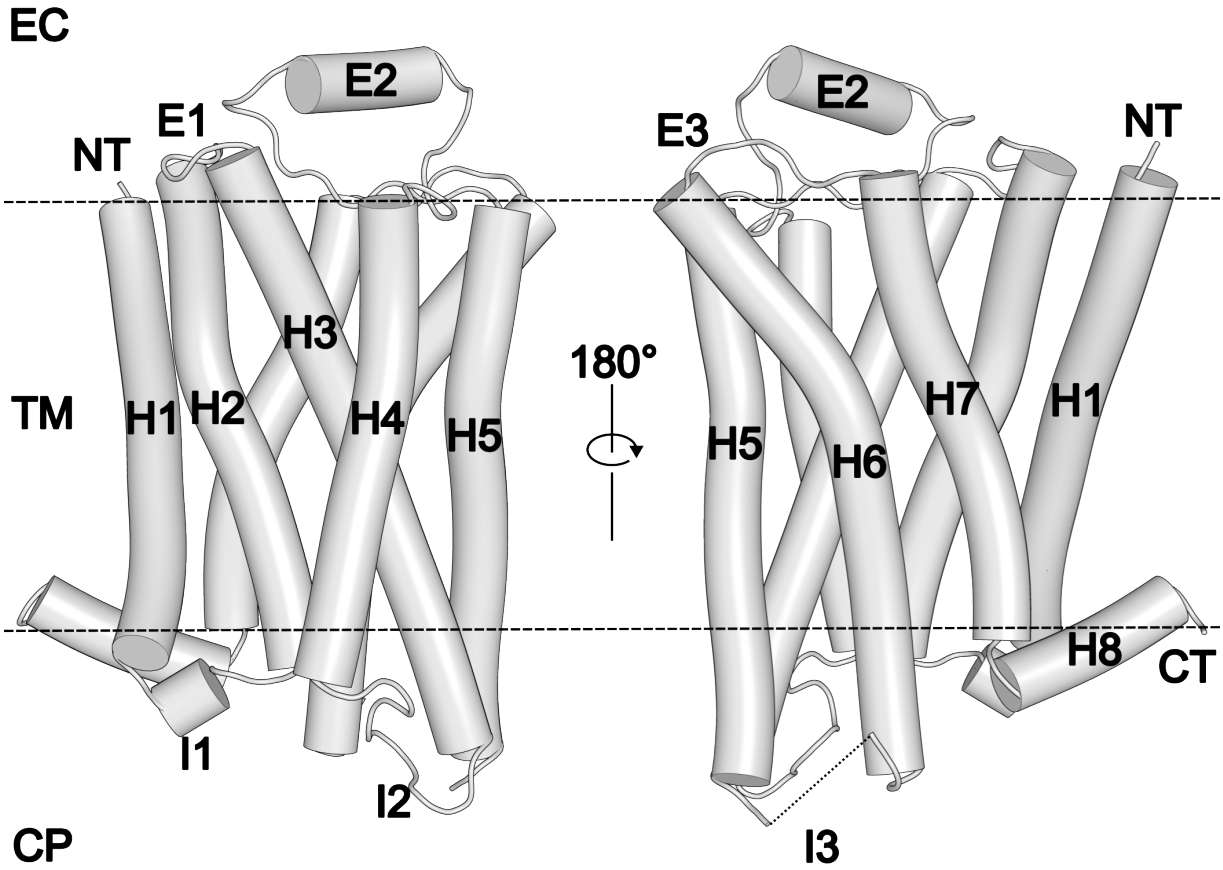

Figure S1: Overview of the regions and segments used in the cavity and ligand labeling scheme. Shown is a structure of the β_2-adrenergic receptor (PDB ID: 6OBA[14]). Note that intracellular loop 3 (I3) and the N- and C-termini (NT and CT) are not visible in this representation due to deletions and the presence of inserted fusion proteins, which are not part of the native receptor sequence and are therefore not shown.

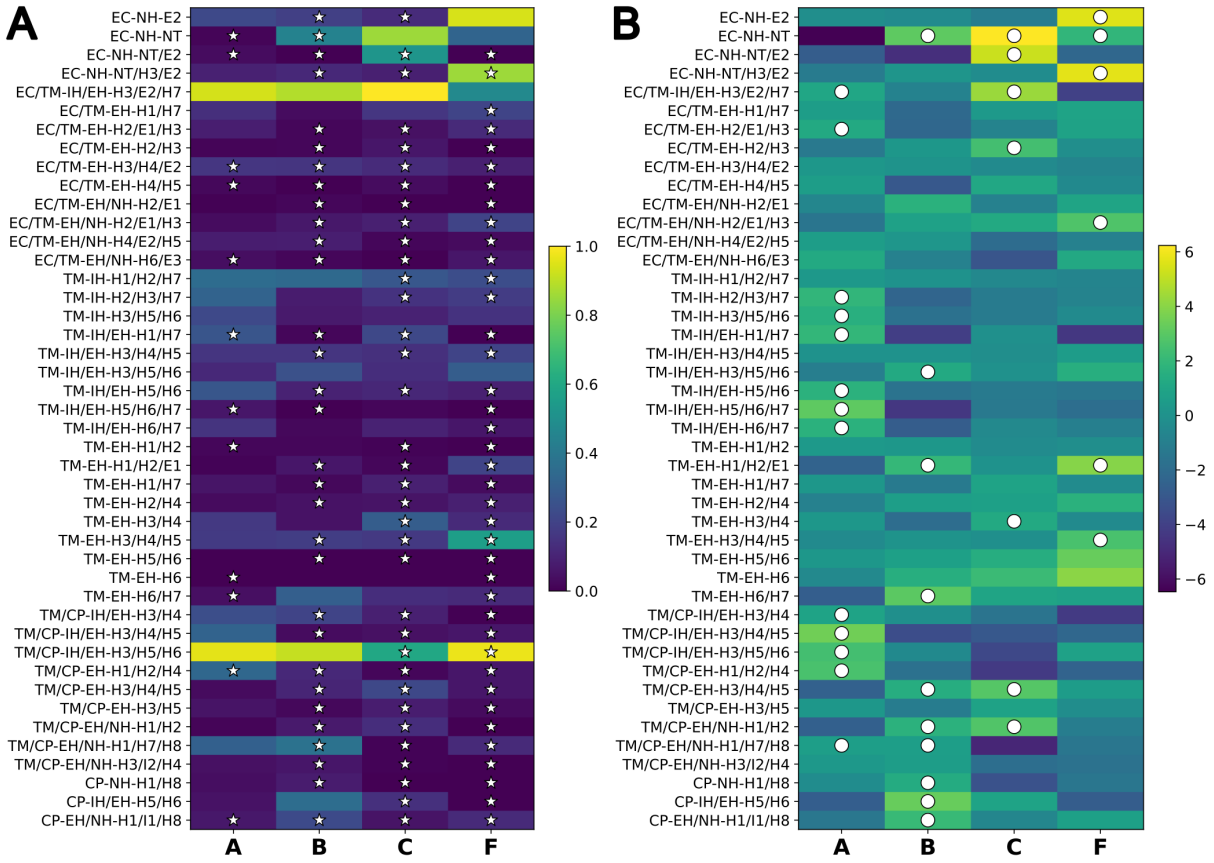

Figure S2: (A) Heatmap showing the prevalence (that is, the fraction of structures within a class that contain at least one cavity located within the respective cluster) of each location cluster across GPCR classes A, B, C, and F. Larger values indicate that the cluster is observed in a larger fraction of structures in the corresponding class. Star markers indicate class–cluster combinations for which no functional ligands have been experimentally determined yet. (B) Heatmap summarizing receptor-class enrichment of detected cavities across location clusters. Cells show $\log_2$ odds ratios from Fisher's exact tests comparing cavity presence in a given cluster for one receptor class versus all other classes. Filled circle markers indicate significant enrichment after multiple-testing correction (BH-FDR, $q < 0.05$). Positive values indicate enrichment in the corresponding class, whereas negative values indicate depletion.

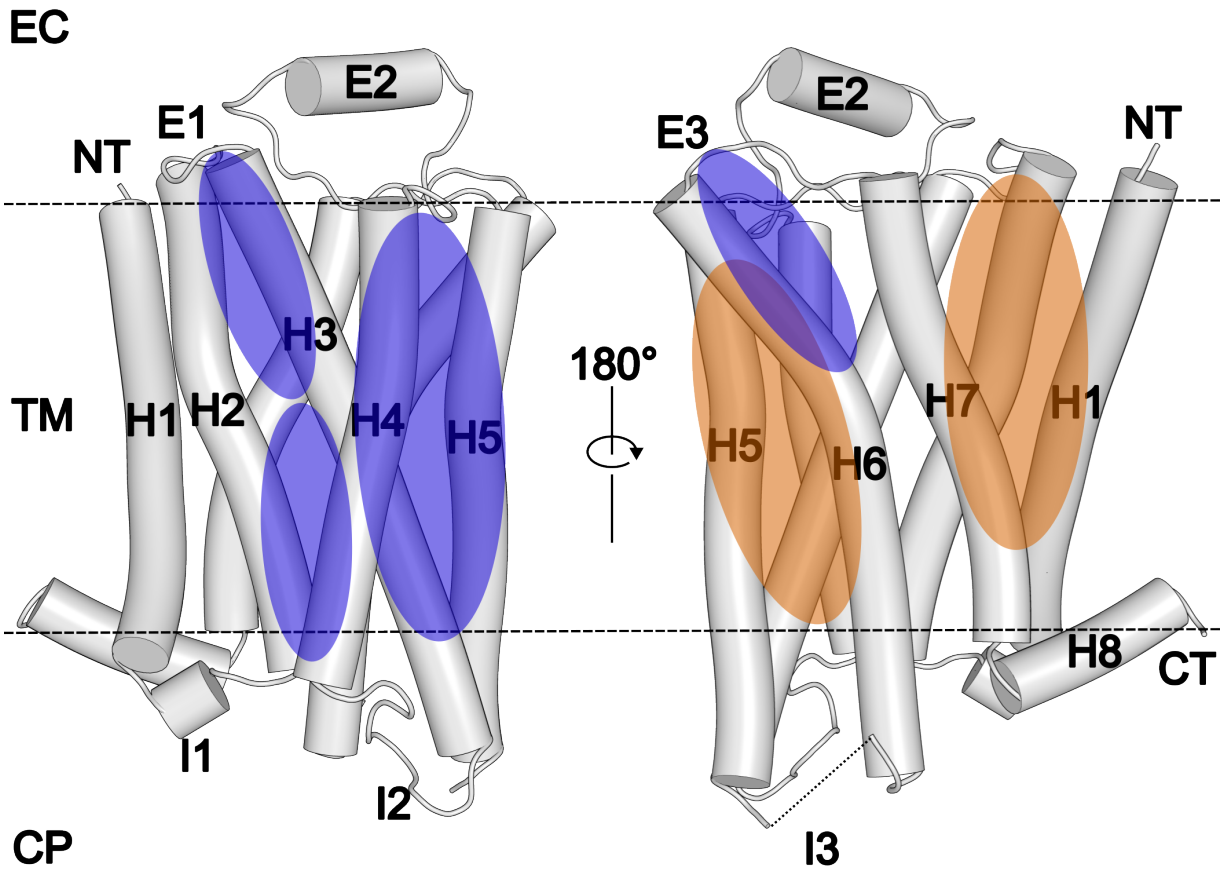

Figure S3: Frequently observed lipid insertion sites (orange) and interfaces preferentially occupied by sterane-containing molecules (blue) on GPCR surfaces. Shown is a structure of the β_2-adrenergic receptor (PDB ID: 6OBA[14]). Note that intracellular loop 3 (I3) and the N- and C-termini (NT and CT) are not visible in this representation due to deletions and the presence of inserted fusion proteins, which are not part of the native receptor sequence and are therefore not shown.

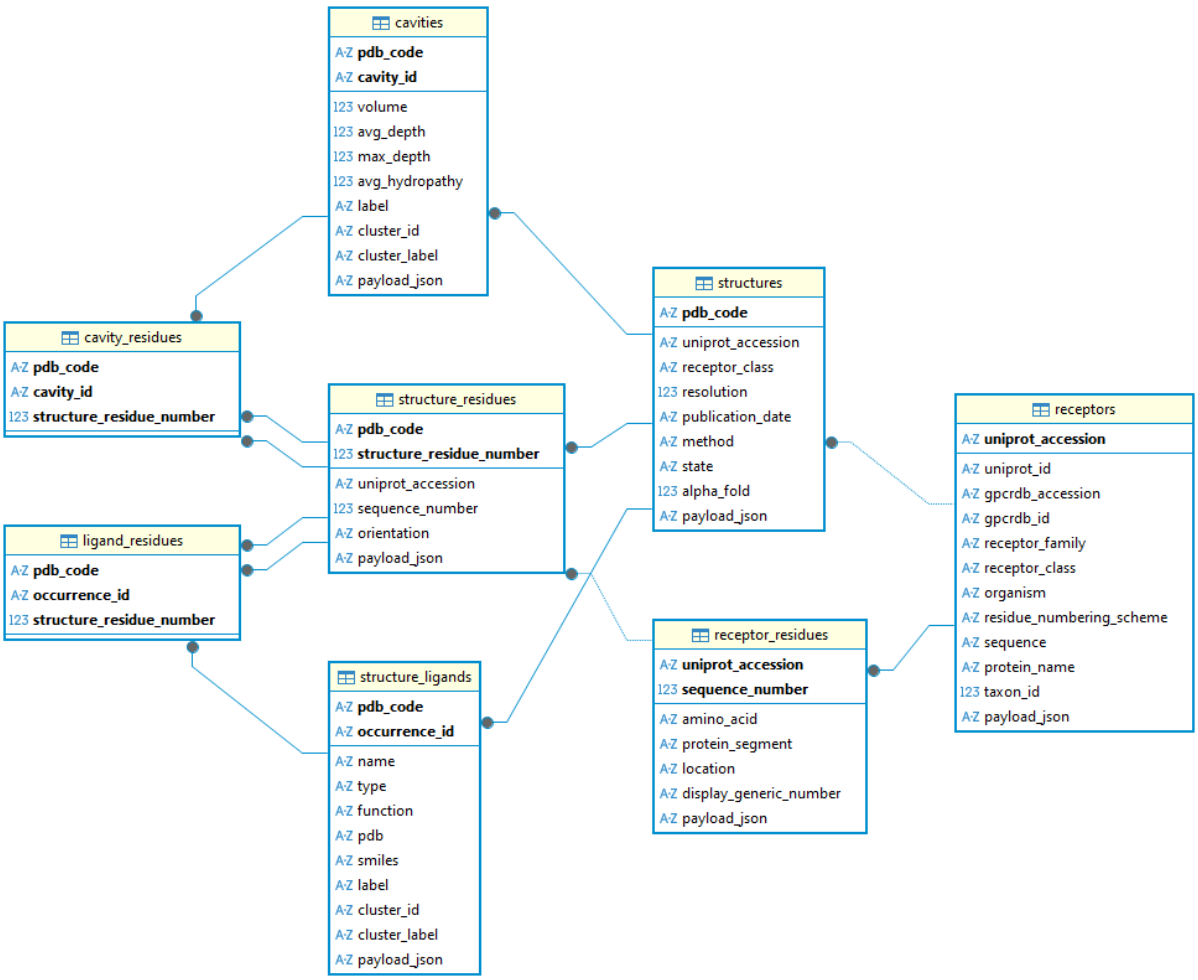

Figure S4: Entity-relationship (ER) diagram of the SQLite database schema used for storing structure-level and (receptor) protein-level payloads. The diagram was generated using DBeaver 26.0.0.[15]

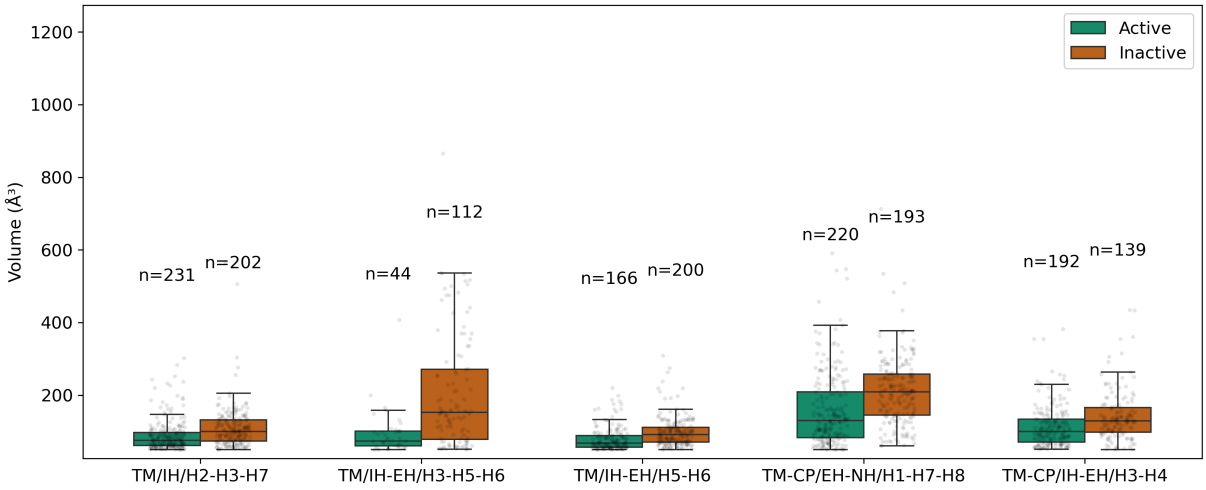

Figure S5: Boxplots showing the distribution of cavity volumes for location clusters with significant differences between active (green) and inactive (orange) states in class A GPCRs. Single data points are shown as grey, semi-transparent dots. Only clusters observed in at least 30 active and 30 inactive structures are shown. Cavity cluster TM-CP/IH-EH/H3-H5-H6 (G protein binding site) was excluded due to large volumes that skewed the scale. Significant differences were determined using the Mann–Whitney U test with BH-FDR correction ($q < 0.05$).

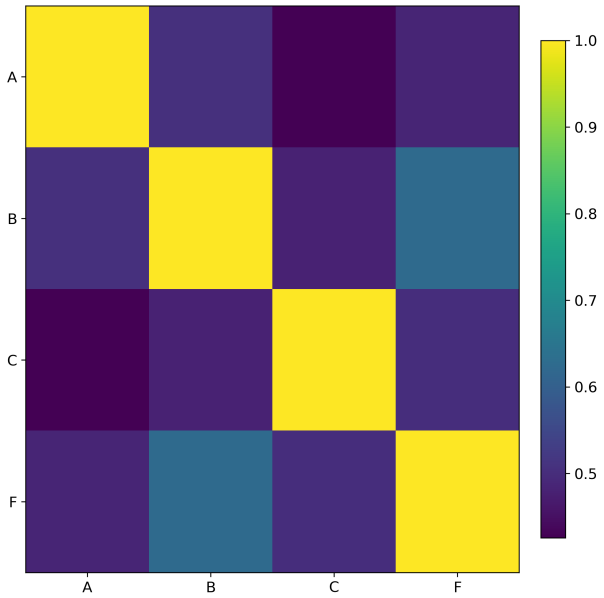

Figure S6: Spearman correlation matrix showing the similarity of cavity location cluster prevalence profiles between GPCR classes A, B, C, and F.

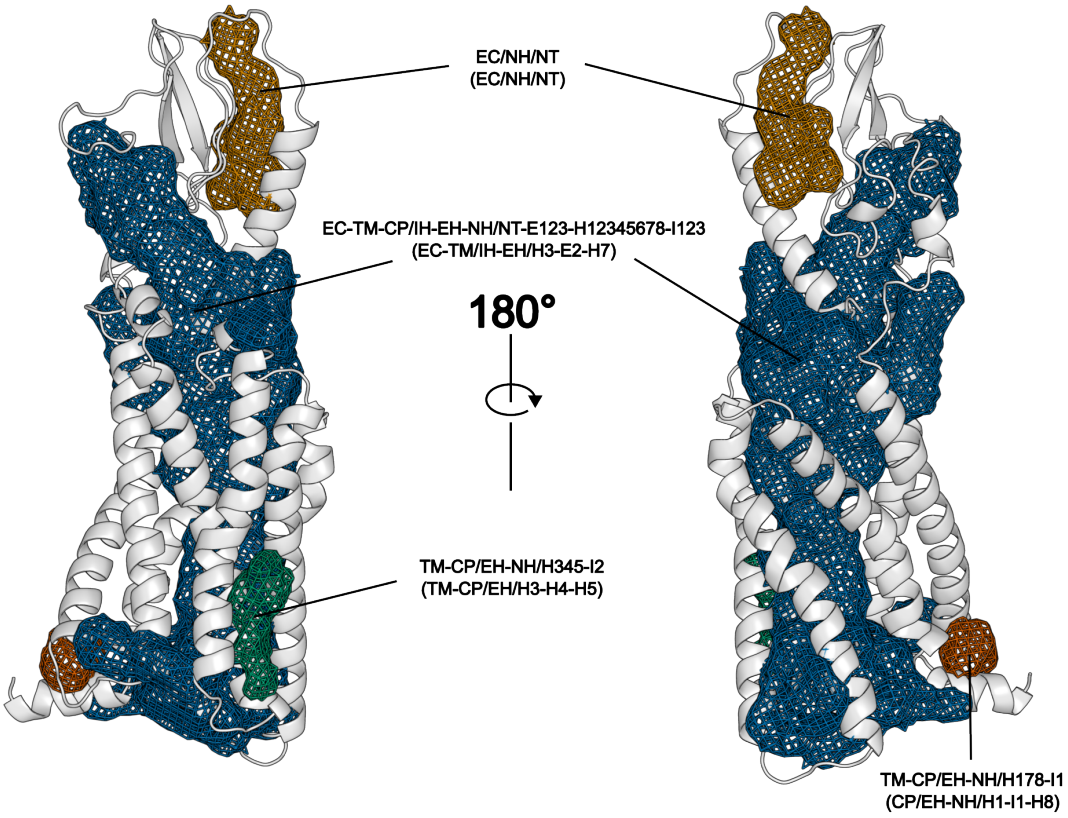

Figure S7: Detected cavities (colored meshes) on the GLP-1 receptor (PDB ID: 5NX2[16], cartoon representation) along with their corresponding location labels. The upper label indicates the full residue-based location label, whereas the lower label in parentheses indicates the location cluster a particular cavity (label) has been assigned to. This is an example of a non-class-A receptor structure in which one large cavity spans the transmembrane bundle, connecting the extracellular and cytoplasmic ends.

Supplementary Tables

Table S1: Overview of the receptor class and activation state distribution of all GPCR structures analyzed in this study. Numbers in parentheses indicate unique receptor counts. Classes B1 and B2 have been combined into one class.

Receptor Class	Active-state structures	Inactive-state structures	Intermediate-state structures	Other	Total counts
A	988 (194)	374 (80)	21 (6)	1 (1)	1384 (209)
B	159 (26)	21 (7)	1 (1)	0 (0)	181 (26)
C	47 (8)	58 (13)	10 (6)	0 (0)	115 (14)
D	1 (1)	0 (0)	0 (0)	0 (0)	1 (1)
F	20 (8)	14 (5)	0 (0)	0 (0)	34 (9)
ALL	1215 (237)	467 (105)	32 (13)	1 (1)	1715 (259)

Table S2: Abbreviations used for location labeling of cavities and co-resolved chemical entities. Note that the abbreviations H, E, and I in the "Segment" category are followed by a number indicating the specific helix, intra-, and extracellular loop, respectively.

Component	Abbreviations
Region	EC (extracellular), TM (transmembrane), CP (cytoplasmic)
Orientation	IH (intra-helical), EH (extra-helical), NH (non-helical)
Segment	H (helix), E-I (extracellular- or intracellular loop), NT-CT (termini)

Table S3: SMARTS patterns used for detecting co-resolved lipids and sterane-containing molecules.

Entity type	SMARTS
Lipid	`[CR0]~[CR0]~[CR0]~[CR0]~[CR0]~[CR0]~[CR0]~[CR0]~[CR0](=O)[$([OX2H1]),$([OX1-]),$([OX2]~*),$([NX3]~*)]`
Sterane	`[CR]1~[CR]~[CR]~[CR]2~[CR]~[CR]~[CR]3~[CR]4~[CR]~[CR]~[CR]~[CR]4~[CR]~[CR]~[CR]3~[CR]2~[CR]1`

References

[1] *Nucleic Acids Res.* **2025**, *53*, D609–D617.

[2] B. Vallat, Y. Rose, D. W. Piehl, J. M. Duarte, S. Bittrich, C. Bi, J. Segura, A. Zalevsky, M. R. Sekharan, B. M. Webb, et al., *Nucleic Acids Res.* **2026**, *54*, D489–D498.

[3] M. Querino Lima Afonso, I. Pidruchna, S. Nair, A. Midlik, D. Lawal, M. Vollmar, S. D. Appasamy, P. Choudhary, I. R. Kunnakkattu, D. Bertoni, et al., *Nucleic Acids Res.* **2026**, *54*, D440–D444.

[4] L. P. T. Herrera, S. N. Andreassen, J. Caroli, I. Rodríguez-Espigares, A. A. Kermani, G. M. Keserű, A. J. Kooistra, G. Pándy-Szekeres, D. E. Gloriam, *Nucleic Acids Res.* **2025**, *53*, D425–D435.

[5] W. L. DeLano et al., *CCP4 Newsl. Protein Crystallogr.* **2002**, *40*, 82–92.

[6] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.

[7] J. B. Hedderich, P. Kolb, GPCR pocketome 2026: Database.

[8] H. B. Mann, D. R. Whitney, *Ann. Math. Stat.* **1947**, 50–60.

[9] R. A. Fisher et al., *The world of mathematics* **1956**, *3*, 1514–1521.

[10] Y. Benjamini, Y. Hochberg, *J. R. Stat. Soc. Series B Stat. Methodol.* **1995**, *57*, 289–300.

[11] C. Spearman, **1961**.

[12] J. Haldane, *Ann. Hum. Genet.* **1956**, *20*, 309–311.

[13] F. J. Anscombe, *Biometrika* **1956**, *43*, 461–464.

[14] X. Liu, J. Kaindl, M. Korczynska, A. Stößel, D. Dengler, M. Stanek, H. Hübner, M. J. Clark, J. Mahoney, R. A. Matt, et al., *Nat. Chem. Biol.* **2020**, *16*, 749–755.

[15] The DBeaver development team, *DBeaver: Free Open-Source Database Management Tool*, **2026**.

[16] A. Jazayeri, M. Rappas, A. J. Brown, J. Kean, J. C. Errey, N. J. Robertson, C. Fiez-Vandal, S. P. Andrews, M. Congreve, A. Bortolato, et al., *Nature* **2017**, *546*, 254–258.