An online GPCR structure analysis platform

Albert J. Kooistra, Christian Munk, Alexander S. Hauser and David E. Gloriam

We present an online, interactive platform for comparative analysis of all available G-protein coupled receptor (GPCR) structures while correlating to functional data. The comprehensive platform encompasses structure similarity, secondary structure, protein backbone packing and movement, residue–residue contact networks, amino acid properties and prospective design of experimental mutagenesis studies. This lets any researcher tap the potential of sophisticated structural analyses enabling a plethora of basic and applied receptor research studies.

G-protein coupled receptors (GPCRs) mediate the actions of two-thirds of endogenous hormones and neurotransmitters, several sensory stimuli and over one-third of US Food and Drug Administration-approved drugs. High-resolution structural data for GPCRs is therefore immensely valuable for our understanding of physiological signaling processes and rational drug design. The GPCR field has experienced a structural revolution beginning before and boosted by recent breakthroughs in cryo-electron microscopy leading to >500 molecular structures (https://gpcrdb.org/structure/statistics). The largest comparative structure analyses of GPCRs have uncovered fundamental mechanisms of receptor activation and effector G-protein selectivity (for which structures have also been combined with molecular dynamics simulations, for example). However, such complex analyses are conducted by only a handful of research groups with a core expertise in data analysis and their results are challenging for non-experts to reproduce. Together, this has led to an ‘embarrassment of riches’, as understanding the available structural data—and integrating it with functional data—has become rate-limiting for a myriad of basic and applied research studies.

Here, we provide an online, interactive platform that allows researchers from any discipline to swiftly conduct comparative analysis of all available GPCR structures and to correlate the results to the many functional data already integrated in the GPCRdb hub, such as genetic variants, in vitro mutations, ligand interactions and G-protein couplings. The suite of interactive analysis tools spans structure similarity, secondary structure, protein backbone packing and movement, residue–residue contact networks, amino acid properties and prospective design of experimental mutagenesis studies of the identified functional determinants.

The 'Structure comparison tool' (https://review.gpcrdb.org/structure_comparison/comparative_analysis) is a suite of three analysis modes, each with four browsers of structural features and over 20 visualizations (Fig. 1). The analysis modes are a single structure, structure set (for frequencies and distributions) and structure set pair (for unique/common features). The data browsers support interactive filtering and sorting for: (1) contact position pairs; (2) contact position–amino acid pairs; (3) residue backbone and sidechain movement; and (iv) residue helix types, bulges and constrictions. All quantitative values in a structure set represent means from all receptors and their individual structures, avoiding skewing of comparisons. The visualizations are interactive plots tailored to the type of data: Transmembrane helix 1–7 (TM1–7) segment movement (two-dimensional (2D) and three-dimensional (3D) segment plots), residue contact segments (Flareplot, 2D and 3D network), residue contacts (Flareplot, heatmap, 2D and 3D networks and 3D structure), residue contact frequencies (box plot) and residue properties (box plot, heatmap, scatter plot, snakeplot and 3D structure). Altogether, this tool allows for swift yet powerful analysis of distances, movements, topology, distributions and differences to identify correlations across the macro- to micro-scales, that is, from TM helix bundle contacts to residue backbone kinks, sidechain rotamers and atomic interactions.

The 'Structure similarity trees' (https://review.gpcrdb.org/structure_comparison/structure_clustering) allow for conformational clustering of any GPCR structures through an exhaustive all-to-all Cα–Cα distance pair comparison across structurally equivalent residue positions (~24,000 distances/receptor) and average linkage clustering (Methods). This technique is independent of sequence similarity and the biases of traditional 3D alignment methods with root mean square deviation values, which vary depending on template and superposition region. Clustering receptors based on their conformation instead of sequence aids correlation to receptor function, which is further facilitated by mapping of state, endogenous ligands and G proteins (Fig. 2).

To show the utility of the platform to uncover functional determinants, we applied the Structure comparison tool to list residues that stabilize receptors in an inactive/active state and suggested mutations in a 'State-stabilizing mutation design tool' (https://review.gpcrdb.org/mutations/state_stabilizing, Extended Data Fig. 1a). For each receptor and state, two complementary rationale are presented: removing residues stabilizing the undesired state (alanine mutation) or introducing residues stabilizing the desired state (consensus amino acid from the structures in the desired state). By comparing the overlap of residue positions with those of literature ligand activity-altering and structure construct mutations stored in GPCRdb (>35,000 data points), we find that state determinants have more functional receptor expression, ligand activity (binding and/or efficacy) and thermal stability data than non-determinants (Extended Data Fig. 1b,c). Although this Brief Communication presents an online platform only, we invite the wider community to also deposit new mutagenesis results via the standardized spreadsheets for ligand activity and structural biology experiments in the GPCRdb research hub.

Prime to all comparative structure analysis is selection of the best templates based on quality, diversity and representativeness of the given function or property of interest. The template selection interface inside the Structure similarity tool facilitates this by presenting:

1Department of Drug Design and Pharmacology, University of Copenhagen, Copenhagen, Denmark. 2Present address: Data Tools Department, Novozymes A/S, Copenhagen, Denmark. These authors contributed equally: Albert J. Kooistra, Christian Munk. 3E-mail: albert.kooistra@sund.ku.dk; david.gloriam@sund.ku.dk

NATURE STRUCTURAL & MOLECULAR BIOLOGY | VOL 28 | NOVEMBER 2021 | 875–878 | www.nature.com/nsmb
Brief Communication

Receptor classification by classes (evolutionary relationships), ligand type (for example, peptide or lipid receptors) or receptor families (sharing endogenous ligand); (2) completeness (percentage of sequence); (3) species and homology to the human protein; (4) structure determination method and resolution; (5) receptor activation state; (6) ligand and its modality; (7) signal protein family or subtype; and (8) auxiliary fusion proteins or antibodies. Structures are updated monthly to mirror the Protein Data Bank and authors can add their new structures via collaborations exploiting the analysis tools towards publication. Given the >500 GPCR structures, which are increasing rapidly, this annotated reference source will facilitate selection of the most appropriate structural templates for a range of scientific studies.

The single biggest factor affecting the receptor structure conformation is the activation state. Consequently, misleading conclusions arise from discrepancy between pharmacological states defined by the ligand modality and presence of an effector G protein, and conformational states predominantly based on the outward movement of TM6 opening the effector site upon activation. For example, (1) agonist-stimulated structures without an effector G protein, (2) effector site opened by allosteric modulators, (3) fusion proteins moving cytosolic TM6 and (4) helix 8 repacking to transmembrane helices. Furthermore, the TM6 activation switch behaves differently across GPCR classes.

To address these problems, we present a ‘degree active’ percent measure based on overall similarity with reference structures with consistent pharmacology and structural integrity, and a classification of all GPCR structures into an inactive, intermediate, active or ‘other’ conformational state (Methods, https://review.gpcrdb.org/structure). This should aid all researchers to correctly characterize the conformational state of ambiguous structures to avoid artifacts in the structural basis.

Taken together, the online platform makes sophisticated comparative structure analysis accessible to a broad research community. It features unique classification of receptors, structures states and techniques. The Structure similarity trees based on all-to-all Cα atom distance pairs, unlike superposition approaches comparing root mean square deviation values, allow for more consistent

Fig. 1 | Structure comparison tool. The Structure comparison tool is a comprehensive research tool (https://review.gpcrdb.org/structure/comparative_analysis). The structural templates include all GPCR structures that can be analyzed for variability in one set or differences between two sets. Results can be filtered to identify functionally critical residue contacts, residue backbone and sidechain movements, residue properties and helix types, bulges and constrictions. The tool includes >20 tailored visualizations.

(1) receptor classification by classes (evolutionary relationships), ligand type (for example, peptide or lipid receptors) or receptor families (sharing endogenous ligand); (2) completeness (percentage of sequence); (3) species and homology to the human protein; (4) structure determination method and resolution; (5) receptor activation state; (6) ligand and its modality; (7) signal protein family or subtype; and (8) auxiliary fusion proteins or antibodies. Structures are updated monthly to mirror the Protein Data Bank and authors can add their new structures via collaborations exploiting the analysis tools towards publication. Given the >500 GPCR structures (mean six per receptor), which are increasing rapidly, this annotated reference source will facilitate selection of the most appropriate structural templates for a range of scientific studies.

The single biggest factor affecting the receptor structure conformation is the activation state. Consequently, misleading conclusions arise from discrepancy between pharmacological states defined by the ligand modality and presence of an effector G protein, and conformational states predominantly based on the outward movement of TM6 opening the effector site upon activation. For example, (1) agonist-stimulated structures without an effector G protein, (2) effector site opened by allosteric modulators, (3) fusion proteins moving cytosolic TM6 and (4) helix 8 repacking to transmembrane helices. Furthermore, the TM6 activation switch behaves differently across GPCR classes.

To address these problems, we present a ‘degree active’ percent measure based on overall similarity with reference structures with consistent pharmacology and structural integrity, and a classification of all GPCR structures into an inactive, intermediate, active or ‘other’ conformational state (Methods, https://review.gpcrdb.org/structure). This should aid all researchers to correctly characterize the conformational state of ambiguous structures to avoid artifacts in the structural basis.

Taken together, the online platform makes sophisticated comparative structure analysis accessible to a broad research community. It features unique classification of receptors, structures states and techniques. The Structure similarity trees based on all-to-all Cα atom distance pairs, unlike superposition approaches comparing root mean square deviation values, allow for more consistent
comparisons by avoiding the uncertainty of which template to use and the difficulty of identifying the superimposable substructure that did not move (if any). Unlike previous resources, the Structure comparison tool can compare a group or two sets of structures with respect to backbone secondary structure and residue properties—and visualize results in over 20 interactive plots. Furthermore, the contact percent frequencies presented herein are directly interpretable (unlike the only other available score) and allow identification of determinants differing in opposite structure sets (ref. 17 only supports one set and ref. 16 cannot compare structures). This tool’s built-in anti-skewing averaging of quantitative properties (Methods) eases use and, together with the extensive target selection table including, for example, resolution cut-offs, makes results more robust. Finally, the State-stabilizing mutation design tool has the advantage that it identifies functional determinants, not based on individual pairs but on the net sum of all contacts to other residues. Therefore, the platform could be applied to uncover determinants of, for example, constitutive activity and ligand-dependent biased signaling, allosteric modulation, efficacy or modality, and will grow even stronger as structural biology continues to advance.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-021-00675-6.

Received: 23 March 2021; Accepted: 22 September 2021; Published online: 10 November 2021

**References**

1. Foster, S. R. et al. Discovery of human signaling systems: pairing peptides to G protein-coupled receptors. Cell 179, 895–908 e21 (2019).
2. Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schioth, H. B. & Gloriam, D. E. Trends in GPCR drug discovery: new agents, targets and indications. Nat. Rev. Drug Discov. 16, 829–842 (2017).
3. Munk, C. et al. An online resource for GPCR structure determination and analysis. Nat. Methods 16, 151–162 (2019).
4. Zhou, Q. et al. Common activation mechanism of class A GPCRs. elife 8, e50279 (2019).
5. Hauser, A. S. et al. GPCR activation mechanisms across classes and macro/microscales. Nat. Struct. Mol. Biol. https://doi.org/10.1038/s41594-021-00674-7 (2021).

6. Venkatakrishnan, A. J. et al. Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. Nature 536, 484–487 (2016).

7. Flock, T. et al. Selectivity determinants of GPCR–G-protein binding. Nature 545, 317–322 (2017).

8. Sandhu, M. et al. Conformational plasticity of the intracellular cavity of GPCR–G-protein complexes leads to G-protein promiscuity and selectivity. Proc. Natl Acad. Sci. USA 116, 11956–11965 (2019).

9. Hauser, A. S. et al. Pharmacogenomics of GPCR drug targets. Cell 172, 41–54 e19 (2018).

10. Munk, C., Harpsoe, 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).

11. Kooistra, A. J. et al. GPCRdb in 2021: integrating GPCR sequence, structure and function. Nucleic Acids Res. 49, D335–D343 (2021).

12. Kufareva, I. & Abagyan, R. Methods of protein structure comparison. Methods Mol. Biol. 857, 231–257 (2012).

13. Armstrong, D. R. et al. PDB: improved findability of macromolecular structure data in the PDB. Nucleic Acids Res. 48, D335–D343 (2020).

14. Hilger, D. et al. Structural insights into differences in G protein activation by family A and family B GPCRs. Science 369, eaba3373 (2020).

15. Holm, L. & Sander, C. Mapping the protein universe. Science 273, 595–603 (1996).

16. Kayikci, M. et al. Visualization and analysis of non-covalent contacts using the Protein Contacts Atlas. Nat. Struct. Mol. Biol. 25, 185–194 (2018).

17. Venkatakrishnan, A. J. et al. Uncovering patterns of atomic interactions in static and dynamic structures of proteins. Preprint at bioRxiv https://doi.org/10.1101/840694 (2019).

18. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65 (1987).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021
**Methods**

**Coding framework.** We built the new resource on top of the existing GPCRdb infrastructure which uses a Django Framework \(^1\) and the packages BioPython \(^1\), NumPy \(^2\), SciPy \(^3\) and SciKit-learn (http://scikit-learn.org). The backend calculates and precomputes the data based on the structure selection made by the user. Subsequently, JavaScript functions parse the data in the browser and present the visualization options and data browsers to the user. For all data browsers (that is, the structure selection and the data browsers mentioned above) we applied the *DataTables.js* (https://datatables.net) module in conjunction with yadcf.js (https://yadcf-showcase.appspot.com). This allowed for advanced sorting and filtering of our tables. The visualizations were written in JavaScript and most use the D3.js framework (https://d3js.org) to generate SVG figures and animations. In addition, we utilized D3.js modules for specific plotting options, most notably Flareplot \(^4\), which we substantially customized. For the boxplots we used plotly.js and for 3D representation of receptor structures we utilized NGL \(^5\) with custom additions for superimposing and coloring.

**GPCR structure state classification.** We calculated a ‘degree active’ percent value and added this for all structures in the GPCRdb structure browser (https://review.gpcrdb.org.structure). Zero percent inactive state structures were defined based on a maximum Cα atom distance between class-specific residue pairs: A, 11.9 Å between 2×46 and 6×37; B1, 13.0 Å between 2×53 and 6×42; C, 14.5 Å between 2×43 and 6×39; and F, 13.0 Å between 2×43 and 6×30. These distance cut-offs were defined based on the best correlation with structures reported as inactive in their literature publications. One hundred percent active structure states were defined as G protein or arrestin bound, although it should be noted that the physiologically relevant ‘fully active’ state may differ due to, for example, stabilizing methods used during structure determination. The fully inactive and active state references were then used to measure the conformational similarity (below) classifying all other structures into the same categories or to an intermediate state category if the ‘degree active’ spanned between 50% and 75% for class A and between 25% and 50% for class B1 (classes C and F have no intermediate state structures). The structure of the platelet-activating receptor binding the antagonist forskaparin \(^6\) has a very low similarity to both inactive and active state reference structures and was classified as ‘other’.

**Structure similarity trees.** We developed a webserver for conformational clustering of GPCR structures and mapping of structure state, GPCR class and following TM helix (for example, 45 is ECL2 located between TM4 and TM5). A Ballesteros-Weinstein, B1 (Wootten), C (Pion) and F (Wang), but preserves gaps from a structural alignment of two receptors, caused by a unique helix bulge or constriction, in the sequence alignment, thereby avoiding offset of such and following residues. All schemes assign residue numbers relative to the most conserved amino acid residue, which is given the number 50, and prefixed with the TM helix containing the residue number (e.g. 332 is on TM3 and 18 position refer the reference). This generic residue numbering scheme also uniquely indexes helix 8 and structurally conserved loop segments, which are numbered by the preceding and following TM helix (for example, 45 is ECL2 located between TM4 and TM5).

**Geneic residue numbering.** Corresponding residue positions in each class were indexed with the structure-based GPCRdb generic residue numbering system \(^7\). This builds on the second-based generic residue numbering system (class A (Ballesteros-Weinstein), B1 (Wootten), C (Pion) and F (Wang), but preserves gaps from a structural alignment of two receptors, caused by a unique helix bulge or constriction, in the sequence alignment, thereby avoiding offset of such and following residues. All schemes assign residue numbers relative to the most conserved amino acid residue, which is given the number 50, and prefixed with the TM helix containing the residue number (e.g. 332 is on TM3 and 18 position refer the reference). This generic residue numbering scheme also uniquely indexes helix 8 and structurally conserved loop segments, which are numbered by the preceding and following TM helix (for example, 45 is ECL2 located between TM4 and TM5).

**Structure comparison tool.** We created an extensive integrated online tool for comparative GPCR structure analysis (https://review.gpcrdb.org/structure_comparison/comparative_analysis). Its ‘Structure selection table’ was developed to extend the annotation in the GPCRdb structure browser with additional data aiding selection of representative templates for GPCRs and activation states. Integration with other GPCRdb analysis tools and stand-alone resources was implemented by import and export PDB codes. Separate analysis modes were developed for analysis of a single structure, structure set or structure set pair. We generated separate data browsers for: (1) contact position pairs, (2) contact position- AA pairs, (3) residue backbone and sidechain movement and (4) residue helix types, bulges and constrictions. The data browsers present comprehensive data on contact, helix, helix residue properties (class A (Ballesteros-Weinstein), B1 (Wootten), C (Pion) and F (Wang), but preserves gaps from a structural alignment of two receptors, caused by a unique helix bulge or constriction, in the sequence alignment, thereby avoiding offset of such and following residues. All schemes assign residue numbers relative to the most conserved amino acid residue, which is given the number 50, and prefixed with the TM helix containing the residue number (e.g. 332 is on TM3 and 18 position refer the reference). This generic residue numbering scheme also uniquely indexes helix 8 and structurally conserved loop segments, which are numbered by the preceding and following TM helix (for example, 45 is ECL2 located between TM4 and TM5).

**State-stabilizing contact determination.** We developed two browsers, for any and specific amino acid combinations, respectively, for residue-residue pair contacts in our webservice for comparative structure analysis along with plots to visualize these in a 3D structure \(^8\), Flareplot \(^9\), network (2D and 3D, adapted from https://d3js.org) or heatmap. Contact definitions and defaults for intra- and intersegment sidechain and backbone contacts are explained in the settings menu of the webservice. For each residue in a receptor structure, non-hydrogen atoms in close proximity with non-hydrogen atoms from neighboring residues are taken into account. These putative contacts are further evaluated based on atom and residue types and their distance. For each of the contact types, the default maximum distances are (4.5 Å), polar (4 Å), aromatic (stacking 5.5 Å and cation–π 6.6 Å), hydrophobic (4.5 Å) and Van der Waals contacts (1.1 times their combined Van der Waals radii). Depending on the chosen settings, contact angles for polar and aromatic contacts are also taken into account.

**Transmembrane helix rearrangement plots.** We developed a tool and 2D plot for TM1–7 segment movement at the extracellular end, cytosolic end and membrane mid, respectively, in our webservice for comparative structure analysis (https://review.gpcrdb.org/structure_comparison/comparative_analysis). TM1–7 helix axes are determined on the TM1 side of the three most active TM helix residues of a given receptor. Generic residue positions estimated from the average GPCR structure placement in the membrane according to the Orientations of Proteins in Membranes database \(^2\). The TM helix rotation is calculated as the difference in structure set 1–2 (here inactive and active templates) of the average angle between: (1) a line from the TM1 (least moving TM) residue number, location in TM1 (TM1 with respect to the membrane), through the TM7 bundle axis, calculated as the average axis through all TM helices through the midpoint of the receptor bundle; and (2) lines from the axis of the given TM helix through each Cα atom of the three above residue positions. The representation of the seven TM helices is projected onto a plane where the normal is the average of the vectors from each of the seven TM helix axes.

**Snakeplot topological mapping of determinants to functional sites.** We developed a snakeplot mapping contact residue positions. We integrated all ligand (proteins, peptide, small molecules, and so on) and G protein interacting positions from all 488 structures released before 1 November 2020. These new data points covered 408 and 125 residue positions, respectively, for the GPCR superfamily. Ligand binding positions were considered orthosteric and allosteric if above and below the membrane mid (using Orientations of Proteins in Membranes database, see above), respectively, except in class C where all such positions in the 7TM are allosteric because its orthosteric ligands bind exclusively in the N-terminal domain. For positions with both an allosteric ligand and G protein interaction, precedence is given to the latter in the snakeplot data mapping.

**State-stabilizing mutation design tool.** The ‘State-stabilizing mutation design tool’ (https://review.gpcrdb.org/mutations/state_stabilizing) lists residues with the most frequent state-specific contacts. For each residue determinant, the tool provides a rank as suggested amino acid residue number and suggests whether to increase or repel the given interaction. The suggestions are based on residue contacts observed in inactive versus active state structures of a GPCR class, while uniquely ranking each residue position by all its contacts instead of a single residue pair. For each GPCR class, the 30 residue positions suggested for mutagenesis have the largest difference in contact frequency sums (inactive versus active state) and are therefore hypothesized to stabilize a single state (wet). For each receptor and state, two complementary rationale are presented: (1) remove residues stabilizing the undesired state (alanine mutation), and (2) introduce residues stabilizing the desired state (consensus amino acid from the structure in the desired state). Together, this allows the fine-tuning of receptor activity by exploiting state determinants that stabilize only a single state (not both) and to (re-)introduce consensus amino acid chains that form many state-specific contacts in the GPCR class but are missing in the receptor of interest.

For each suggested residue mutation position, the tool also presents already known experimental mutation effects. This includes literature mutations with effects on ligand activity (>3,468 data points) and structure construct mutations (540 data points) that affect primarily thermostability. It also includes receptor expression-increasing mutations that are subsets of these two datasets (483 and 173 data points, respectively). For each mutant position, the overall supporting data types are presented as the sum of: (1) presence of ligand activity-altering mutations (more than fivefold effect in at least two receptors), (2) thermostabilizing mutations and expression data on homologous (more than fivefold effect in at least two receptors), (3) thermostabilizing mutations and expression data on homologous (more than fivefold effect in at least two receptors), (4) ligand activity studies (>30% increase, minimum of two receptors). The tool automatically incorporates new structural templates, including for the classes C and F, which do not yet have supporting data in GPCRdb.
GPCRdb has previously provided thermostabilizing mutations using rule-based sequence rationale and inference of structure construct mutations, integrated in a structure construct design tool (https://gpcrdb.org/construct/design and ref.). The Katritch and Vaidehi groups have combined similar information (the latter also dynamics) into machine learning predictors. Furthermore, the Vaidehi group earlier combined receptor models with an energy function\(^3\); however, both its approaches limit their mutation suggestions to alanine. Our new state-stabilizing mutation design tool differs by giving direct access to all pregenerated mutation suggestions for all human GPCRs and going further in providing a data-driven rationale for mutations stabilizing a given receptor state, because it is founded on templates from all GPCRs for which a structure has been determined—and new GPCR structures are added to GPCRdb monthly. Whereas established methods focus on only the desired state, this tool also removes residues stabilizing the undesired state. The importance of validating the state-specific nature of a mutation across both states is illustrated by the fact that a majority of agonist-bound crystal structures of GPCR are in the inactive state.\(^1\) It is therefore recommended to measure the effect of a given mutation on an inactive versus active state proxy, such as the effect on binding affinity or thermal stability in the presence of an inverse agonist and agonist, respectively. Furthermore, our analysis of determinant overlap showed that the receptor state is commonly associated with not just thermal stability, but also ligand activity (binding and/or efficacy) and receptor expression at the cell membrane, which benefits from a more stable protein (Extended Data Fig. 1). Hence, state-stabilizing mutations may have an underappreciated utility across pharmacology, biophysics and structural biology, such as dissection of signaling bias determinants\(^2\) or fine-tuning of receptor basal activity\(^3\), ligand assay sensitivity and signal window\(^4\) or complexation with G-proteins or other effectors and receptor activity modulatory proteins.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

All data is available in GPCRdb (https://review.gpcrdb.org) and GitHub (https://github.com/protwis/gpcrdb_data). Documentation is available at https://docs.gpcrdb.org.

**Code availability**

All open-source code can be obtained from GitHub (https://github.com/protwis/protwis) under the permissive Apache 2.0 License (https://www.apache.org/licenses/LICENSE-2.0).

**References**

19. Cock, P. J. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
20. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
21. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
22. Rose, A. S. & Hildebrand, P. W. NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.* **43**, W576–W579 (2015).
23. Cao, C. et al. Structural basis for signal recognition and transduction by platelet-activating-factor receptor. *Nat. Struct. Mol. Biol.* **25**, 488–495 (2018).
24. Isberg, V. et al. Generic GPCR residue numbers—aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.* **36**, 22–31 (2015).
25. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
26. Popov, P., Kozlovskii, I. & Katritch, V. Computational design for thermostabilization of GPCRs. *Curr. Opin. Struct. Biol.* **55**, 25–33 (2019).
27. Muk, S. et al. Machine learning for prioritization of thermostabilizing mutations for G-protein coupled receptors. *Biophys. J.* **117**, 2228–2239 (2019).
28. Bhattacharya, S., Lee, S., Grishammer, R., Tate, C. G. & Vaidehi, N. Rapid computational prediction of thermostabilizing mutations for G-protein-coupled receptors. *J. Chem. Theory Comput.* **10**, 5149–5160 (2014).
29. Schonegge, A. M. et al. Evolutionary action and structural basis of the allosteric switch controlling beta2AR functional selectivity. *Nat. Commun.* **8**, 2169 (2017).
30. Seifert, R. & Wenzel-Seifert, K. Constitutive activity of G-protein-coupled receptors: cause of disease and common property of wild-type receptors. *Naunyn-Schmiedeberg’s Arch. Pharmacol.* **366**, 381–416 (2002).
31. Walldhoer, M. et al. The carbosyl terminus of human cytomegalovirus-encoded 7 transmembrane receptor US28 camouflages agonism by mediating constitutive endocytosis. *J. Biol. Chem.* **278**, 19473–19482 (2003).

**Acknowledgements**

This work was supported by the Lundbeck Foundation (grant nos R163-2013-16327 and R218-2016-1266), the Novo Nordisk Foundation (grant no NNF18OC001226) and Independent Research Fund Denmark | Natural Sciences (grant no 8021-00173B) to D.E.G.

**Author contributions**

D.E.G. conceptualized the study. D.E.G., C.M. and A.J.K. developed the methodology. A.J.K. curated the data. A.J.K. and D.E.G. carried out the investigation. A.J.K. and D.E.G. validated the data. D.E.G. wrote the original draft of the manuscript. A.J.K and A.S.H. reviewed and edited the manuscript. D.E.G., A.S.H. and A.J.K. visualized the study. D.E.G. acquired the funding. A.J.K. and C.M. developed the software. D.E.G. supervised the study.

**Competing interests**

After the completion of this study, C.M. moved to become an employee of Novozymes A/S.

The other authors declare no competing interests.

**Additional information**

Extended data is available for this paper at https://doi.org/10.1038/s41594-021-00675-6. Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-021-00675-6.

Correspondence and requests for materials should be addressed to Albert J. Kooistra or David E. Gloriam.

Peer review information Nature Structural & Molecular Biology thanks Peter Hildebrand and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Florian Ulrich was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.
**State-stabilising mutation design tool**

| Generic residue number | Contact frequency sum | Contact frequency difference | Ligand activity altering | Expression-increasing (ligand activity studies) | Expression-increasing (structure constructs) | Thermostabilising | State stabiliser type |
|-----------------------|-----------------------|------------------------------|--------------------------|-----------------------------------------------|----------------------------------------------|------------------|-----------------------|
| 3x50                  | 155                   |                              |                          |                                               | Switch                                        | Activator        | Suggested state-stabilising position |
| 7x50                  | 150                   |                              |                          |                                               | Switch                                        | Activator        | Suggested state-stabilising position |
| 5x58                  | 125                   |                              |                          |                                               | Inactivator                                    | Activator        | Suggested state-stabilising position |
| 7x42                  | 99                    |                              |                          |                                               | Inactivator                                    | Activator        | Suggested state-stabilising position |
| 3x41                  | 97                    |                              |                          |                                               | Inactivator                                    | Activator        | Suggested state-stabilising position |
| 7x46                  | 94                    |                              |                          |                                               | Inactivator                                    | Activator        | Suggested state-stabilising position |
| 7x56                  | 87                    |                              |                          | Switch                                        | Activator                                      | Activator        | Suggested state-stabilising position |
| 2x40                  | 80                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 2x56                  | 80                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 3x43                  | 79                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 2x42                  | 77                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 6x40                  | 75                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 4x49                  | 74                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 6x38                  | 74                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 2x39                  | 74                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 1x49                  | 74                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 5x61                  | 71                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 6x36                  | 71                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 8x54                  | 71                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 2x50                  | 70                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 1x50                  | 70                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 3x28                  | 69                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 2x60                  | 69                    |                              |                          | Inactivator                                    | Activator                                      | Activator        | Suggested state-stabilising position |
| 6x47                  | 67                    |                              |                          | Inactivator                                    | Activator                                      | Inactivator      | Suggested state-stabilising position |
| 6x51                  | 64                    |                              |                          | Inactivator                                    | Activator                                      | Inactivator      | Suggested state-stabilising position |
| 2x45                  | 64                    |                              |                          | Inactivator                                    | Activator                                      | Inactivator      | Suggested state-stabilising position |
| 6x34                  | 61                    |                              |                          | Inactivator                                    | Activator                                      | Inactivator      | Suggested state-stabilising position |
| 3x29                  | 60                    |                              |                          | Inactivator                                    | Activator                                      | Inactivator      | Suggested state-stabilising position |
| 1x39                  | 59                    |                              |                          | Inactivator                                    | Activator                                      | Inactivator      | Suggested state-stabilising position |
| 5x36                  | 57                    |                              |                          | Inactivator                                    | Activator                                      | Inactivator      | Suggested state-stabilising position |

**Extended Data Fig. 1** | See next page for caption.
Extended Data Fig. 1 | State-stabilizing mutation design tool, residue positions and experimental data. a. The ‘State-stabilizing mutation design tool’ presents data-driven suggestions of mutagenesis experiments for all human GPCRs (https://review.gpcrdb.org/mutations/state_stabilizing). The tool ranks receptor positions by calculating a net sum of residue contacts expected to be gained or removed upon mutation. b. Suggested state-stabilizing positions for classes A and B1, respectively. These are limited to the 30 generic residue positions with the largest inactive/active state contact sum difference. The rightmost column indicates state stabilizers with high-frequency contacts. c. Percent coverage of suggested state-stabilizing versus all other generic residue positions by experimentally determined mutations that are ligand activity-altering (>5-fold effect), thermostabilizing (540 data points) or expression increasing (100% would mean that all determinants or non-determinants, respectively are covered by experimental mutations). For ligand activity mutations (34,648 data points in GPCRdb), we required an effect in at least two receptors. For class A GPCRs, 27/30 residue positions have experimental support (avg. 1.8 functional associations). In class B1, 8 positions are supported by functional data (avg. 0.3 associations). We compared the percentages of residue positions covered by experimental effects for the class A and B1 determinants suggested in the mutation design tool (top 30) versus all other generic residue positions. This shows a near double representation of such data for suggested determinants than other generic residue positions in class B1. For class A GPCRs, we find stronger determinant overlaps spanning 2.1-, 2.9-, 3.1 and 8.8-fold ratios for mutations shown to influence thermostability, expression in ligand activity studies, ligand activity and expression of structure constructs, respectively. Notably, the top and third positions for class A GPCRs are two well-characterized residue microswitches, R3x50 and Y5x58 and the second position is a conserved proline causing the hinge of TM7. Notably, in both classes 13 out of 30 (43%) of suggested mutagenesis positions are unique from this tool (not in5.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a  Confirmed

☑ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☑ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☑ The statistical test(s) used AND whether they are one- or two-sided

Only common tests should be described solely by name; describe more complex techniques in the Methods section.

☑ A description of all covariates tested
☑ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☑ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☑ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

Give P values as exact values whenever suitable.

☑ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☑ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☑ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

GPCRdb data: https://github.com/protwis/gpcrdb_data [release April 2020]
Guide to Pharmacology: https://www.guidetopharmacology.org/webServices.jsp [release June 2017]
RCSB Protein Data Bank: https://www.rcsb.org [release 1st April 2020]
UniProt: https://www.uniprot.org [version 2020_03]

Data analysis

GPCRdb web resource: https://gpcrdb.org [release November 2020]
Microsoft Office 365: www.microsoft.com
NGl structure viewer: https://github.com/arose/ngl [version 2.0.0-dev.36]
FlarePlot: https://github.com/GPCRViz/flareplot [release December 2017]
DSSP: https://github.com/cmb/dssp [version 3.1.2]
FreeSASA: https://github.com/mttinatten/freesasa [version 2.0.3]
BioPython: https://github.com/biopython/biopython [version 1.74]
D3.js: https://github.com/d3/d3 [versions 3.5.17 and 5.14.2]
plotly.js: https://github.com/plotly/plotly [version v1.52.1]
jQuery: https://github.com/jquery/jquery [version v3.1.1]

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. Github). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All relevant data are integrated into the web resource in the GPCR database at http://www.gpcrdb.org] and are available at GitHub (https://github.com/protwis/gpcrdb_data). All other data that support the findings of this study are provided as Extended Data.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- [ ] Life sciences  [ ] Behavioural & social sciences  [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | The platform for comparative structure analysis includes all available crystal and cryo-EM structures from the Protein Data Bank. |
| Data exclusions | No structures crystal or cryo-EM structures were excluded by the authors. Readers wishing to selected the best templates for their given study can do so using a comprehensive annotation in the structure selection table included in the platform. |
| Replication | The platform was applied for identification of GPCR activation state determinants in a separate analysis study (referenced and submitted as a related manuscript file). During the course of that study, new structural templates emerged and were added. In two rounds of updates, all values were updated and results obtained were similar (e.g. only few state-specific contacts changed) and none of the overall conclusions/findings in the manuscript changed. This supports that the methods implemented in the platform are robust. |
| Randomization | This study did not produce new experimental data, rather analyzed existing data (structures, in vitro mutations, sequence conservation etc.) and therefore this field is not applicable. |
| Blinding | Blinding was not possible because of the selection and analysis of samples (structures) that required intervention. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a  Involved in the study
- [ ] Antibodies
- [ ] Eukaryotic cell lines
- [ ] Palaeontology and archaeology
- [ ] Animals and other organisms
- [ ] Human research participants
- [ ] Clinical data
- [ ] Dual use research of concern

Methods

n/a  Involved in the study
- [ ] ChIP-seq
- [ ] Flow cytometry
- [ ] MRI-based neuroimaging