Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An online GPCR structure analysis platform

David Gloriam (david.gloriam@sund.ku.dk)
University of Copenhagen https://orcid.org/0000-0002-4299-7561
Albert Kooistra
University of Copenhagen https://orcid.org/0000-0001-5514-6021
Christian Munk
University of Copenhagen
Alexander Hauser
University of Copenhagen

Brief Communication

Keywords: Interactive Platform, Functional Data, Sophisticated Structural Analyses, Applied Research Studies

Posted Date: April 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-354878/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 Structural & Molecular Biology on November 10th, 2021. See the published version at https://doi.org/10.1038/s41594-021-00675-6.

Brief communication

An online GPCR structure analysis platform

Albert J. Kooistra[1,3*], Christian Munk[1,2,3], Alexander S. Hauser[1] and David E. Gloriam[1*]

[1]*Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen, Denmark.*

[2]*Present address: Novozymes A/S, Biologiens Vej 2, 2800 Kongens Lyngby Copenhagen, Denmark*

[3]*These authors contributed equally*

**Correspondence: albert.kooistra@sund.ku.dk (A.J.K.) and david.gloriam@sund.ku.dk (D.E.G.)*

We present an online, interactive platform for comparative analysis of all available G protein-coupled receptor 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 research studies.

G protein-coupled receptors (GPCRs) mediate the actions of two-thirds of endogenous hormones and neurotransmitters[1], several sensory stimuli and over one-third of the FDA-approved drugs[2]. 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 the recent breakthroughs in cryo-electron microscopy[3] leading >500 molecular structures (https://gpcrdb.org/structure/statistics). The largest comparative structure analyses of GPCRs have uncovered fundamental mechanisms of receptor activation[4-6] and effector G protein selectivity[7,8]. However, such complex analyses are conducted only by a handful of research groups with a core expertise in data analysis/science and their results are challenging to reproduce by non-experts. 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[9], *in vitro* mutations[10], ligand interactions[10] and G protein couplings[11]. The suite of interactive analysis tools span 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 range 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 i) Contact position pairs, ii) Contact position-amino acid pairs, iii) Residue backbone & 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: TM1-7 segment movement (2D and 3D segment plots), receptor segment contacts (flare plot, 2D and 3D network), residue contacts (flare plot, 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 i.e., from TM helix bundle contacts to residue backbone kinks, sidechain rotamers and atomic interactions.

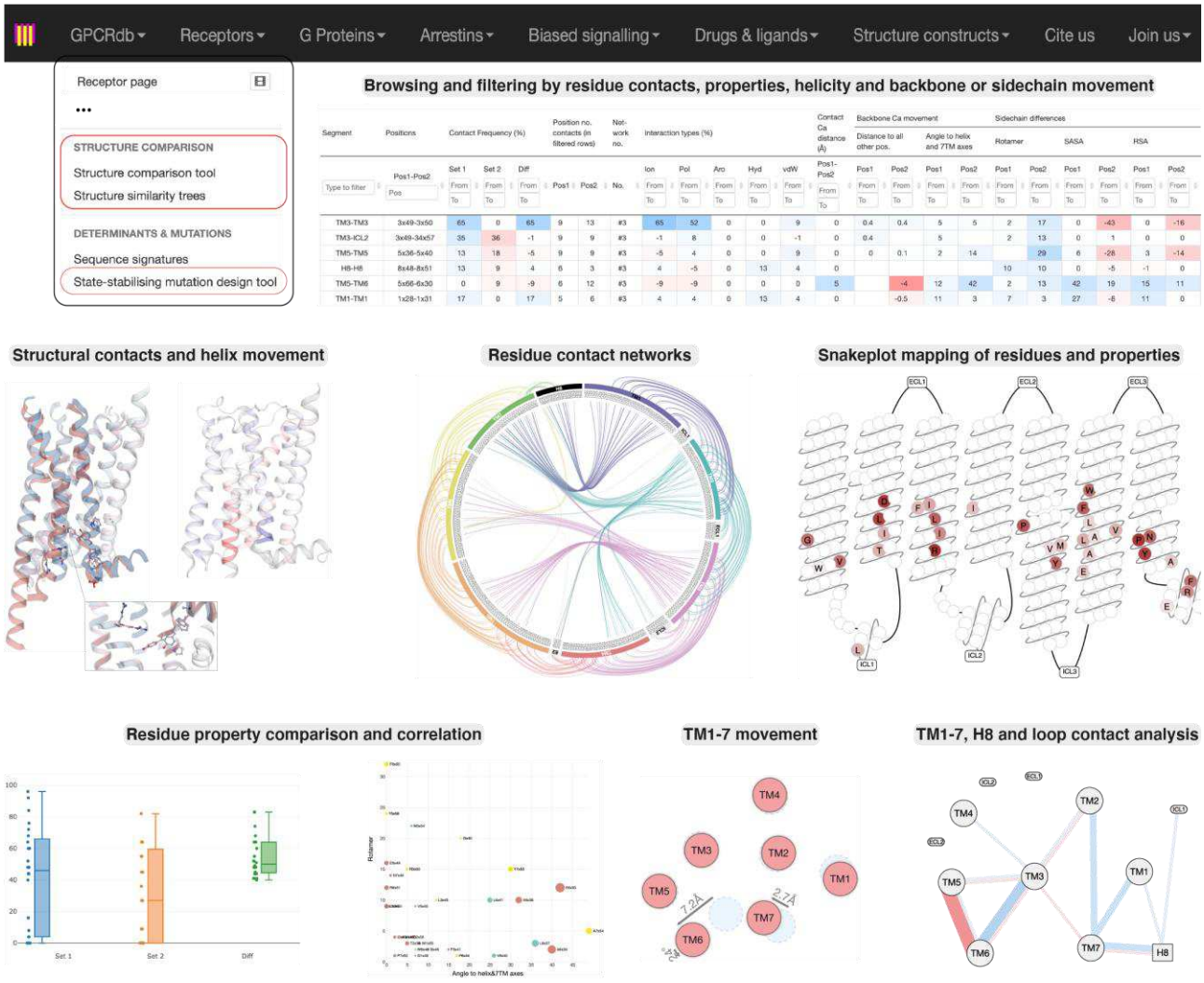


Fig. 1 | Structure comparison tool. The 'Structure comparison tool' is a comprehensive research tool (https://review.gpcrdb.org/structure_comparison/comparative_analysis). The structural templates include all GPCR structures which can be analyzed for variability in one set or differences between two 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.

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[12]. 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).

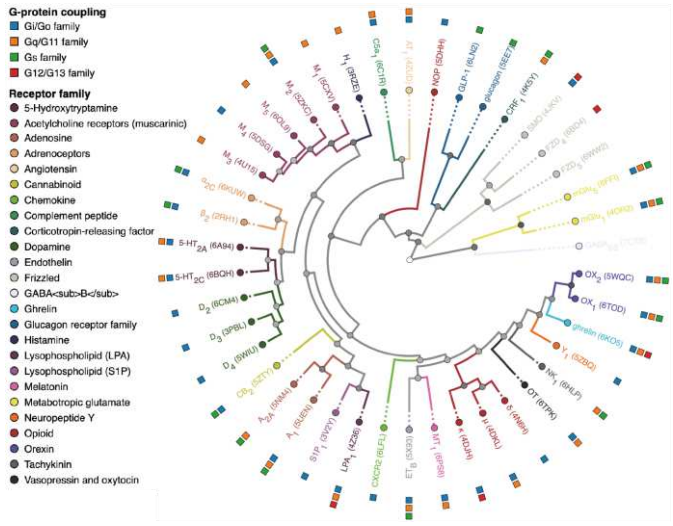


Fig. 2 | Structure similarity trees. The structure similarity trees cluster receptor structures by their conformation of the common seven transmembrane helix bundle and can assign annotations, such as inactive/intermediate/active state, G protein coupling and phylogenetic and pharmacological classifications (http://review.gpcrdb.org/structure_comparison/structure_clustering). Tree nodes have grey scale coloring illustrating the separation of clades based on a silhouette score[13] (Methods) shown upon mouse hovering.

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[10,11] (>35,000 datapoints), we find that the 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). Whereas this Brief Communication presents an online platform only, we invite the wider community to deposit also 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 table in the *Structure Similarity tool* facilitates this by presenting: i) Receptor classification by classes (evolutionary relationships), ligand type (e.g. peptide or lipid receptors) or receptor families (sharing endogenous ligand), ii) completeness (% of sequence), iii) species and homology to the human protein, iv) structure determination method and resolution, v) receptor activation state, vi) ligand and its modality, vii) signal protein family or subtype and vii) auxiliary fusion proteins or antibodies. Structures are updated monthly to mirror the Protein Data Bank[14] and authors can add their novel structures via collaborations exploiting the analysis tools towards publication. Given the >500 GPCR structures (mean 6/receptor), which is increasing rapidly, this annotated reference source will facilitate the 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[4,6,15]. For example, i) agonist-stimulated structures

without an effector G protein, ii) effector site opened by allosteric modulators, iii) fusion proteins moving cytosolic TM6, and iv) helix 8 repacking to transmembrane helices. Furthermore, the TM6 activation switch behaves differently across GPCR classes[5,15]. 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, http://review.gpcrdb.org/structure). This should aid all researchers to correctly characterize the conformational state of ambiguous structures to avoid artefacts 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[16] avoids inconsistencies arising from template and substructure selection in traditional protein superposition and root mean square deviation (RMSD) approaches. Unlike previous resources[3,17,18], the *Structure comparison tool* can compare a group or two sets of structures w.r.t. 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[4]) and allows identification of determinants differing in opposite structure sets ([18] only supports one set and [17]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 e.g. 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[4,6,17,18], but on the net-sum of all contacts to other residues. Therefore, the platform could be applied to uncover determinants of e.g., 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, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at [URL].

References

1 Foster, S. R. *et al.* Discovery of Human Signaling Systems: Pairing Peptides to G Protein-Coupled Receptors. *Cell* **179**, 895-908 e821, doi:10.1016/j.cell.2019.10.010 (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, doi:10.1038/nrd.2017.178 (2017).

3 Munk, C. *et al.* An online resource for GPCR structure determination and analysis. *Nat. Methods* **16**, 151-162, doi:10.1038/s41592-018-0302-x (2019).

4 Zhou, Q. *et al.* Common activation mechanism of class A GPCRs. *Elife* **8**, e50279, doi:10.7554/eLife.50279 (2019).

5 Kooistra, A. J., Munk, C., Hauser, A. S., Babu, M. M. & Gloriam, D. E. GPCR activation mechanisms across classes and macro/microscales. *Submitted*.

6 Venkatakrishnan, A. J. *et al.* Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature* **536**, 484-487, doi:10.1038/nature19107 (2016).

7 Flock, T. *et al.* Selectivity determinants of GPCR-G-protein binding. *Nature* **545**, 317-322, doi:10.1038/nature22070 (2017).

8 Sandhu, M. *et al.* Dynamic Spatiotemporal Determinants Modulate GPCR:G protein Coupling Selectivity and Promiscuity. *Submitted*.

9 Hauser, A. S. *et al.* Pharmacogenomics of GPCR Drug Targets. *Cell* **172**, 41-54 e19, doi:10.1016/j.cell.2017.11.033 (2018).

10 Munk, C., Harpsoe, K., Hauser, A. S., Isberg, V. & Gloriam, D. E. Integrating structural and mutagenesis data to elucidate GPCR ligand binding. *Current opinion in pharmacology* **30**, 51-58, doi:10.1016/j.coph.2016.07.003 (2016).

11 Kooistra, A. J. *et al.* GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Res.* **49**, D335-D343, doi:10.1093/nar/gkaa1080 (2021).

12 Kufareva, I. & Abagyan, R. Methods of protein structure comparison. *Methods Mol. Biol.* **857**, 231-257, doi:10.1007/978-1-61779-588-6_10 (2012).

13 Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**, 53-65, doi:10.1016/0377-0427(87)90125-7 (1987).

14 Armstrong, D. R. *et al.* PDBe: improved findability of macromolecular structure data in the PDB. *Nucleic Acids Res.* **48**, D335-D343, doi:10.1093/nar/gkz990 (2020).

15 Hilger, D. *et al.* Structural insights into differences in G protein activation by family A and family B GPCRs. *Science* **369**, eaba3373, doi:10.1126/science.aba3373 (2020).

16 Holm, L. & Sander, C. Mapping the protein universe. *Science* **273**, 595-603, doi:10.1126/science.273.5275.595 (1996).

17 Kayikci, M. *et al.* Visualization and analysis of non-covalent contacts using the Protein Contacts Atlas. *Nat. Struct. Mol. Biol.* **25**, 185-194, doi:10.1038/s41594-017-0019-z (2018).

18 Venkatakrishnan, A. J. *et al.* Uncovering patterns of atomic interactions in static and dynamic structures of proteins. *bioRxiv*, 840694, doi:10.1101/840694 (2019).

Methods

Coding framework. We built the new resource on top of the existing GPCRdb infrastructure which uses a Django Framework[11] and the packages BioPython[19], NumPy[20], SciPy[21] and Scikit-learn (http://scikit-learn.org). The backend calculates and prepares 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 (i.e., 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 of 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[18] which we significantly customized. For the boxplots we used plotly.js and for 3D representation of receptor structures we utilized NGL[22] 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 (http://review.gpcrdb.org/structure). G protein and arrestin bound structures were set to 100% active, and structures with a distance between Cα atoms of a specific residue pair in TM2 and TM6 below a fixed cut-off (see below) were set as reference for 0% active. The aforementioned residue pairs were selected for each class based on a high correlation between their distance and the state of the structures as reported in their literature publications. The accompanying distance cut-offs were set to a strict maximum value within which only inactive states structures were selected while adding a margin to account for possible variability in intermediate and active state structures. The selected residue pairs and distance cut-offs for each of the classes are A: 2x46 and 6x37 with a cut-off of 11.9Å, B1: 2x53 and 6x42 with a cut-off of 13.0Å, C: 2x43 and 6x39 with a cut-off of 14.5Å, and F: 2x43 and 6x30 with a cut-off of 13.0Å. All other structures were assigned a degree active value based on conformational similarity (see next section) to these references. Intermediate state structures in each class have degree active spans of; A: 50-75%, B1: 25-50% (for class C and F there are currently no intermediate state structures available). "Other" state structures have a high dissimilarity to both the active and the inactive state structures and are therefore not assigned any degree active value.

Structure similarity trees. We developed a webserver for conformational clustering of GPCR structures and mapping of structure state, GPCR class and receptor families sharing endogenous ligand (http://review.gpcrdb.org/structure_comparison/structure_clustering). Structure pair similarities are calculated based on the sum of all differences between all all-to-all Cα-Cα distance pairs of TM1-7 generic residue positions present in all structures (this avoids skewing). The distances are normalized by dividing them by the mean values for each residue pair. Structures as subsequently grouped based on the acquired all -structures-against-all structures distance matrix using average linkage clustering. Tree nodes have grey scale coloring illustrating the separation of clades based on a silhouette score[13], which represents the score for all structures under this node compared to all structures under the next closest node. This score spans quantifies the separation from -1 (wrong, light red) via 0 (non-significant, white) to 1 (perfect, black).

Generic residue numbering. Corresponding residue positions in each class were indexed with the structure-based GPCRdb generic residue numbering system[23]. This builds on the sequence-based generic residue numbering systems for class A (Ballesteros-Weinstein), B1 (Wootten), C (Pin) 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 number (e.g. 3x32 is on TM3 and 18 positions before 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 (e.g. 45 is ECL2 located between TM4-5).

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 i) Contact position pairs, ii) Contact position-AA pairs, iii) Residue backbone & sidechain movement and iv) Residue helix types, bulges and constrictions. The data browsers present comprehensive data on contact, helix and residue properties (Extended Data Table 1). To mitigate skewing upon comparison of sets of few versus many diverse structural templates, each quantitative value in a structure set was calculated based on the mean from all receptors, each of which is in its turn represented by a mean value from all its structures. For angles we use a mean of circular quantities. In contrast, values are not averaged across structure states (inactive, intermediate, active and other) which can instead optionally be separated into two different structure sets before analysis. Integrated into the tool, we developed >20 interactive visualizations e.g., snakeplot (topology), scatter plot (correlations), box plot (distributions and differences), heatmap (distances or movements) and structure (movements).

State stabilizing contact determination. We developed two browsers, for any and specific amino acid combinations, respectively, for residue-residue pair contacts in our webserver for comparative structure analysis along with plots to visualize these in a 3D structure[22], flare plot[18], 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 webserver. For each residue in a receptor structure, non-hydrogen atoms in close proximity of non-hydrogen atoms from neighboring residues are taken into account. These potential contacts are further evaluated based on atom and residue types and their distance. For each of the contact types, the default maximum distances are ionic (4.5Å), polar (4Å), aromatic (stacking 5.5Å and cation-π 6.6Å), hydrophobic (4.5Å) and Van der Waals contacts (1.1 times their combined VdW radii). Depending on the chosen settings, also contact angles are taken into account for polar and aromatic contacts.

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 webserver for comparative structure analysis (http://review.gpcrdb.org/structure_comparison/comparative_analysis). TM1-7 helix axes are defined based on the three most terminal (of the two ends) or membrane mid generic residues positions estimated from the average GPCR structure placement in the membrane according to the Orientations of Proteins in Membranes (OMP) database[24]. The TM helix rotation is calculated as the difference in structure set 1-2 (here inactive and active templates) of the average angle between: i) a line from the TM1 (least moving TM) axis at position 1x46 (class A residue number, located near the middle of TM1 with respect to the membrane), through the 7TM bundle axis, calculated as the average axis through all TM helices through the midpoint of the receptor bundle and ii) lines from the axis of the given TM helix through each Cα atom of 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 etc.) and Gα protein interacting positions from all 488 structures released before Nov 1st, 2020. These new datapoints 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 OMP, see above), respectively, except in class C where all such positions in the 7TM are allosteric, as 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 well as one or more suggested amino acid(s) to remove or repel the given interaction. The suggestions are based on residue contacts observed in inactive vs. active state structures of a GPCR class, while uniquely ranking each residue position by all its contacts instead of in a single residue pair. For each GPCR class, the 30 residue positions suggested for mutagenesis have the largest difference in contact frequency sums (inactive vs. active state) and are therefore hypothesized to stabilize a single state the most. For each receptor and state, two complementary rationale are presented: i) remove residues stabilizing the undesired state (alanine mutation) and ii) 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 acids 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 effect on ligand activity (>34,648 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 datapoints, respectively). For each mutant position, the overall supporting data types is presented as the sum of i) presence of ligand activity altering mutations (>5-fold effect in at least two receptors), ii) thermostabilizing mutations and expression-increasing mutations from iii) structure constructs (non-quantified data) or iv) 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[3]). The Katritch[25] and Vaidehi[26] 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[27], however both its approaches limit their mutation suggestions to alanine. Our new state stabilizing mutation design tool differs by giving direct access to all pre-generated mutation suggestions for all human GPCRs and going further in providing a data-driven rationale for mutations stabilizing a given receptor state, as it is founded on templates from all GPCRs for which a structure has been determined – and new GPCR structures are added to GPCRdb monthly. While 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[3]. 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[28] or fine-tuning of receptor basal activity[29], ligand assay sensitivity and signal window[30] or complexation with G proteins[3] 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).

Acknowledgements

This work was supported by the Lundbeck Foundation (grants R163-2013-16327 and R218-2016-1266), the Novo Nordisk Foundation (grant NNF18OC0031226) and Independent Research Fund Denmark | Natural Sciences (grant 8021-00173B) to D.E.G.

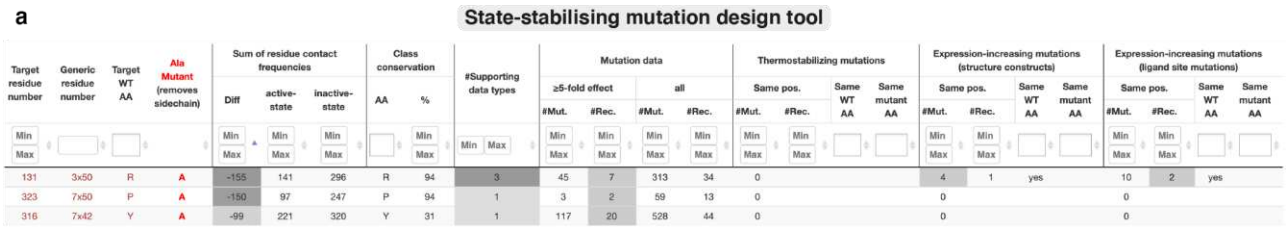
Author Contributions

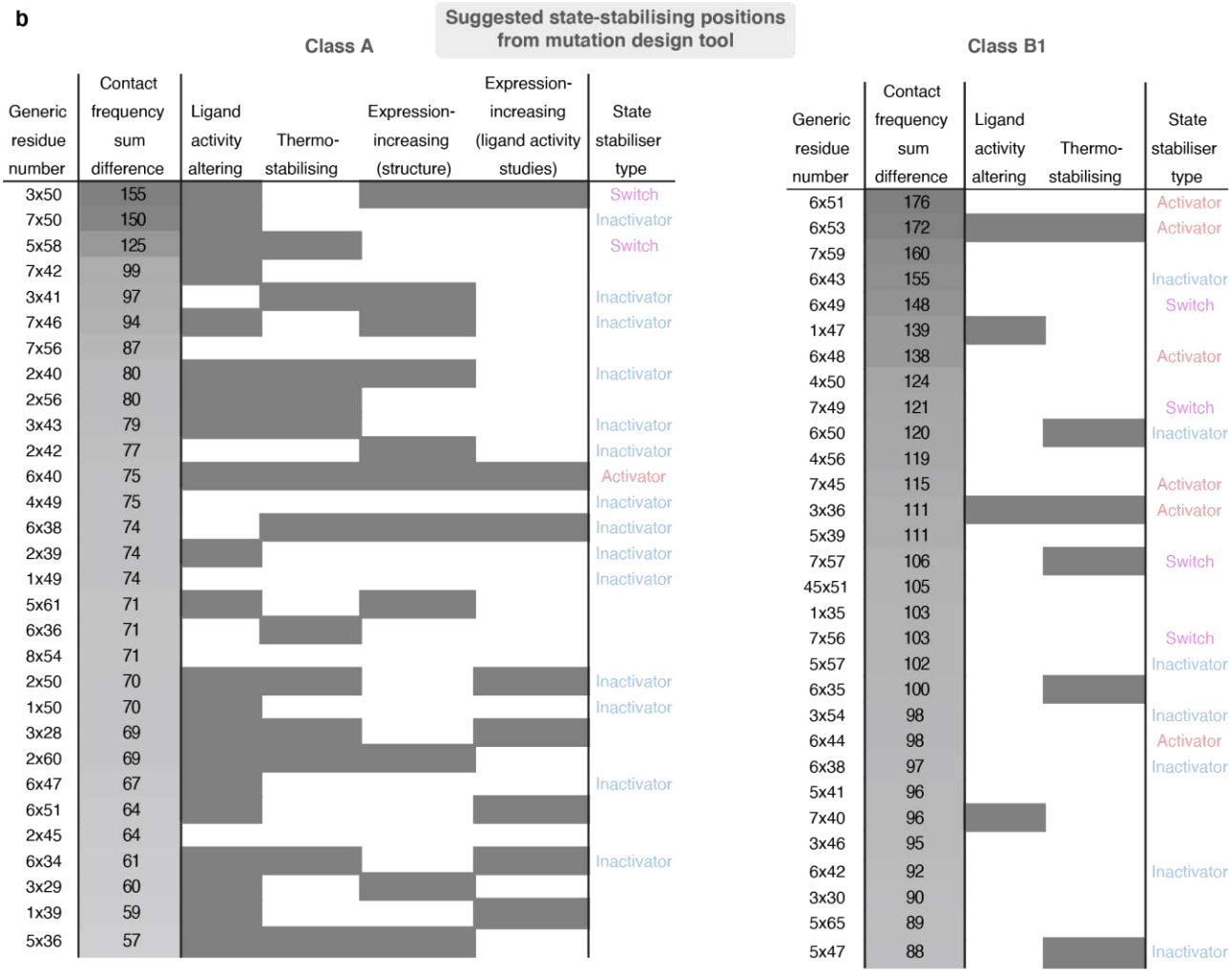
Conceptualization, D.E.G.; Methodology, D.E.G., C.M. and A.J.K.; Data Curation, A.J.K.; Investigation, A.J.K. and D.E.G.; Validation: A.J.K. and D.E.G.; Writing – Original Draft, D.E.G.; Writing – Review & Editing, A.J.K and A.S.H.; Visualization, D.E.G., A.S.H. and A.J.K., Funding Acquisition, D.E.G.; Software, A.J.K. and C.M.; Supervision, D.E.G.

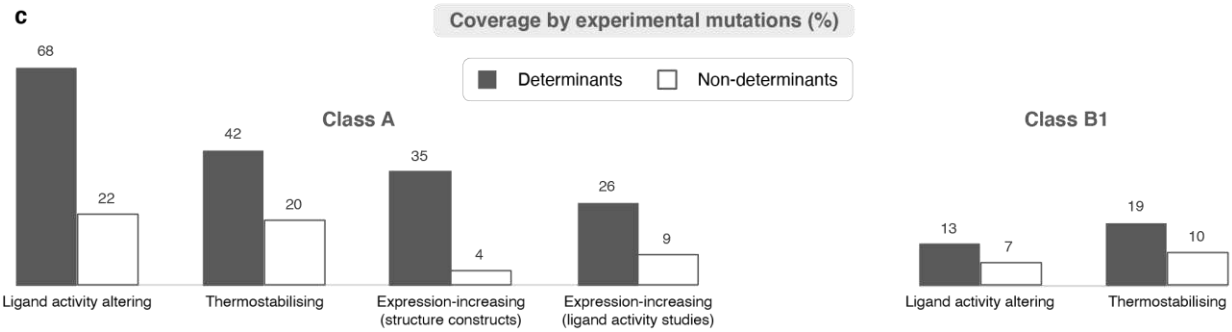
Competing interests

The authors declare no competing interests.

Extended Data Figures and Tables







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 and would therefore strengthen or weaken the targeted inactive or active state. **b**, Suggested state-stabilizing positions for classes A and B1. The suggested mutations are limited to the 30 generic residue positions with the largest contact sum difference between inactive and active state structures. The rightmost column indicates state stabilizers part of high-frequency contacts[5]. **c**, Percent coverage of suggested state-stabilizing (A) versus all other generic residue positions by experimentally determined mutations that are ligand activity-altering (>5-fold effect on affinity or potency), thermostabilizing (540 data points) or expression increasing (100% would mean that all determinants or non-determinants, respectively are covered by experimental mutations). For the large dataset of ligand activity mutations (34,648 data points in GPCRdb), we required an effect in at least two receptors. For class A GPCRs, 27 out of the 30 residue positions have experimental support (mean of 1.8 functional associations per position). In class B1, 8 positions are supported by functional data (mean of 0.3 associations / position). To adjust the assessment in relation to the available data, 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 have an effect on thermostability, expression in ligand activity studies, ligand activity (>5-fold) 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, as their high contact frequency sum difference is not evident from a state stabiliser based on a single high-frequency contact[5].

Extended Data Table 1 | Contact, residue and helix properties in the *Structure comparison tool.*

Type	Property	Definition
Contact	Contact AA pair sequence conservation in class (%)	Conservation (%) of any of the contact's amino acid pairs observed in structures across all receptor paralogues in the given class extracted from the sequence alignments in GPCRdb[31].
Contact	Contact Cα distance (Å)	Cα-Cα distance between two residues that form a contact.
Contact	Contact frequency (%)	Percent frequency of a residue-residue contact. Average for receptors which are first averages across structures.
Contact	Interaction types (%)	Percent frequency of ionic (Ion), polar (Pol), aromatic (Aro), hydrophobic (Hyd) and van der Waals (vdW) interactions.
Contact	Network number	Contract network number (1 …). Used to sort contacts by network.
Contact	Position no. contacts	A count of contacts for a residue position in filtered rows (excluding removed rows).
General	Positions	Generic residue positions using the GPCRdb scheme[23].
General	Segment	Transmembrane helices 1-7 (TM1-7), the first two intra- (ICL1-2) or extracellular loops (ECL1-2) or helix 8 (H8).
Helix	Next tau dihedral	Cα(+1)-Cα-Cα(-1)-Cα(-2)
Helix	Tau dihedral	Cα(+2)-Cα(+1)-Cα-Cα(-1)
Helix	Theta angle	(Cα(+1)-Cα-Cα(-1)
Residue	Angle to helix and 7TM axes	Angle (˚) to the axes of the given TM helix and the 7TM axis (only calculated for residues in TM1-7).
Residue	Cα half-sphere exposure (Å²)	A measure of how "buried" a residue is by counting the number of Cα-atoms in half of the sphere (12 Å radius) perpendicular to the Cβ-Cα vector[32].
Residue	Class seq consensus	Consensus amino acid (AA) and conservation (%) across all receptor paralogues in each class extracted from the sequence alignments in GPCRdb[31].
Residue	Distance to all other pos.	Change in average distance (Å) to all other generic residue positions.
Residue	No 3D coordinates	Percent of structures in the set in which the given residue lacks coordinates.
Residue	No generic number (gap pos)	Percent of structures in the set in which the given residue has a gap in the structure (e.g., due to a shorter segment or a local helix constriction).
Residue	Phi dihedral	N(+1)-C-Cα-N atom dihedral.
Residue	Position presence (%)	Percent of structures in the set in which the given residue has a conserved secondary structure.
Residue	Psi dihedral	C-Cα-N-C(-1) atom dihedral.
Residue	Rotamer	To enable comparison of residue rotamers across all amino acids (including glycine by virtually placing a Cβ atom), rotation was measured as the angle from their terminal heavy atom to the Cα and N atoms.
Residue	RSA	Relative surface area (RSA) (Å²) calculated as the percentage of the calculated SASA for each residue compared to the empirical maximum SASA values for each residue as reported in[33].
Residue	SASA	Solvent-accessible surface area (SASA) (Å²) calculated using FreeSASA[34].
Residue	Secondary structure	Secondary structure information was calculated with DSSP 3.1.2 (which has a highly improved recognition of π-helices)[35].
Residue	Sequence consensus	Consensus amino acid (AA) and conservation (%) in each structure set.
Residue	Tau angle	N-Cα-C atom angle.

References

3 Munk, C. *et al.* An online resource for GPCR structure determination and analysis. *Nat. Methods* **16**, 151-162, doi:10.1038/s41592-018-0302-x (2019).

5 Kooistra, A. J., Munk, C., Hauser, A. S., Babu, M. M. & Gloriam, D. E. GPCR activation mechanisms across classes and macro/microscales. *Submitted*.

11 Kooistra, A. J. *et al.* GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Res.* **49**, D335-D343, doi:10.1093/nar/gkaa1080 (2021).

13 Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**, 53-65, doi:10.1016/0377-0427(87)90125-7 (1987).

18 Venkatakrishnan, A. J. *et al.* Uncovering patterns of atomic interactions in static and dynamic structures of proteins. *bioRxiv*, 840694, doi:10.1101/840694 (2019).

19 Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423, doi:10.1093/bioinformatics/btp163 (2009).

20 Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357-362, doi:10.1038/s41586-020-2649-2 (2020).

21 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261-272, doi:10.1038/s41592-019-0686-2 (2020).

22 Rose, A. S. & Hildebrand, P. W. NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.* **43**, W576-579, doi:10.1093/nar/gkv402 (2015).

23 Isberg, V. *et al.* Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends in pharmacological sciences* **36**, 22-31, doi:10.1016/j.tips.2014.11.001 (2015).

24 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**, D370-376, doi:10.1093/nar/gkr703 (2012).

25 Popov, P., Kozlovskii, I. & Katritch, V. Computational design for thermostabilization of GPCRs. *Current opinion in structural biology* **55**, 25-33, doi:10.1016/j.sbi.2019.02.010 (2019).

26 Muk, S. *et al.* Machine Learning for Prioritization of Thermostabilizing Mutations for G-Protein Coupled Receptors. *Biophys. J.* **117**, 2228-2239, doi:10.1016/j.bpj.2019.10.023 (2019).

27 Bhattacharya, S., Lee, S., Grisshammer, R., Tate, C. G. & Vaidehi, N. Rapid Computational Prediction of Thermostabilizing Mutations for G Protein-Coupled Receptors. *J. Chem. Theory Comput.* **10**, 5149-5160, doi:10.1021/ct500616v (2014).

28 Schonegge, A. M. *et al.* Evolutionary action and structural basis of the allosteric switch controlling beta2AR functional selectivity. *Nat Commun* **8**, 2169, doi:10.1038/s41467-017-02257-x (2017).

29 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 archives of pharmacology* **366**, 381-416, doi:10.1007/s00210-002-0588-0 (2002).

30 Waldhoer, M. *et al.* The carboxyl terminus of human cytomegalovirus-encoded 7 transmembrane receptor US28 camouflages agonism by mediating constitutive endocytosis. *The Journal of biological chemistry* **278**, 19473-19482, doi:10.1074/jbc.M213179200 (2003).

31 Isberg, V. *et al.* GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res.* **44**, D356-364, doi:10.1093/nar/gkv1178 (2016).

32 Hamelryck, T. An amino acid has two sides: a new 2D measure provides a different view of solvent exposure. *Proteins* **59**, 38-48, doi:10.1002/prot.20379 (2005).

33 Tien, M. Z., Meyer, A. G., Sydykova, D. K., Spielman, S. J. & Wilke, C. O. Maximum allowed solvent accessibilites of residues in proteins. *PLoS One* **8**, e80635, doi:10.1371/journal.pone.0080635 (2013).

34 Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Res* **5**, 189, doi:10.12688/f1000research.7931.1 (2016).

35 Touw, W. G. *et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* **43**, D364-368, doi:10.1093/nar/gku1028 (2015).

Figures

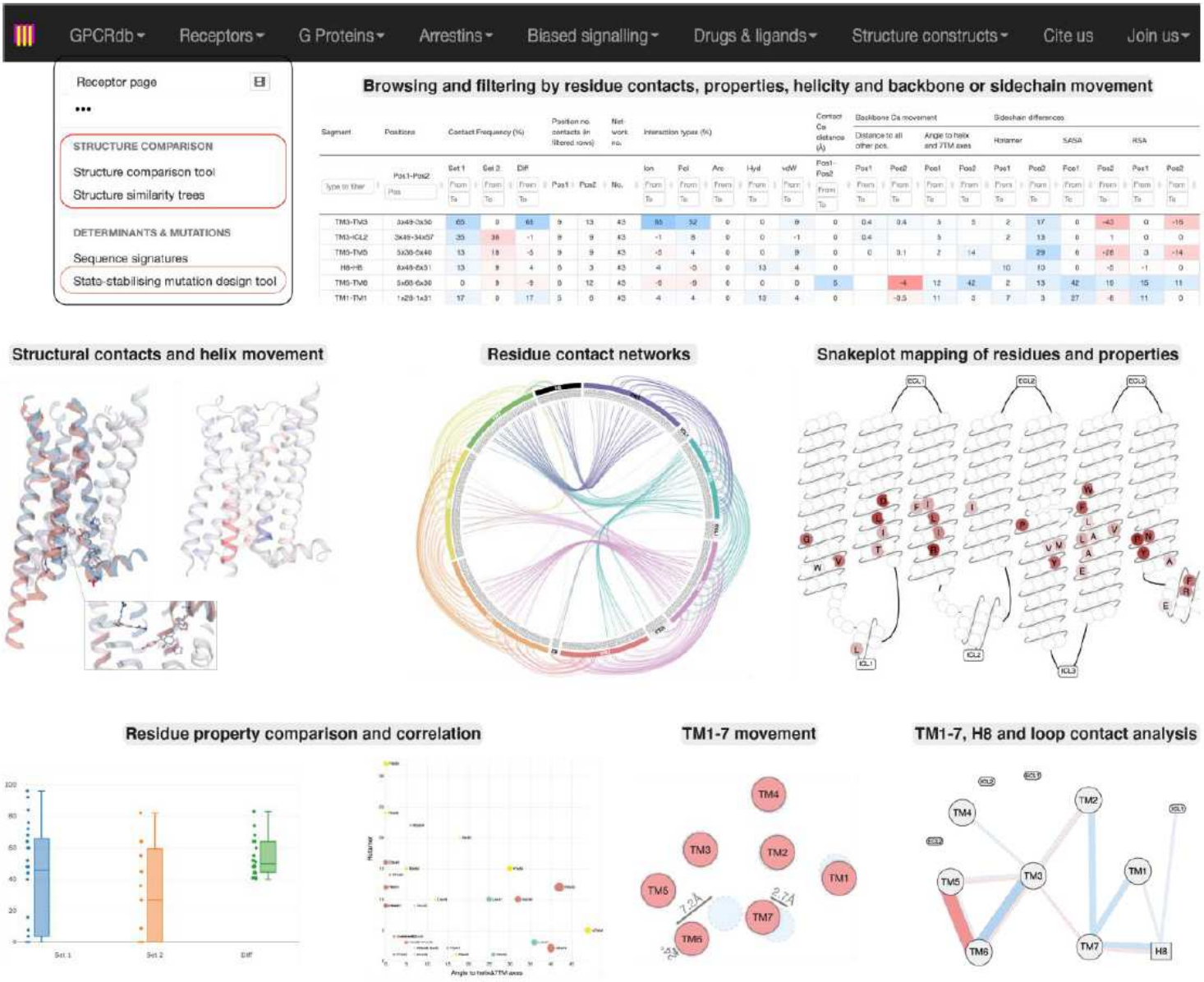


Figure 1

Structure comparison tool. The ‘Structure comparison tool’ is a comprehensive research tool (https://review.gpcrdb.org/structure_comparison/comparative_analysis). The structural templates include all GPCR structures which 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.

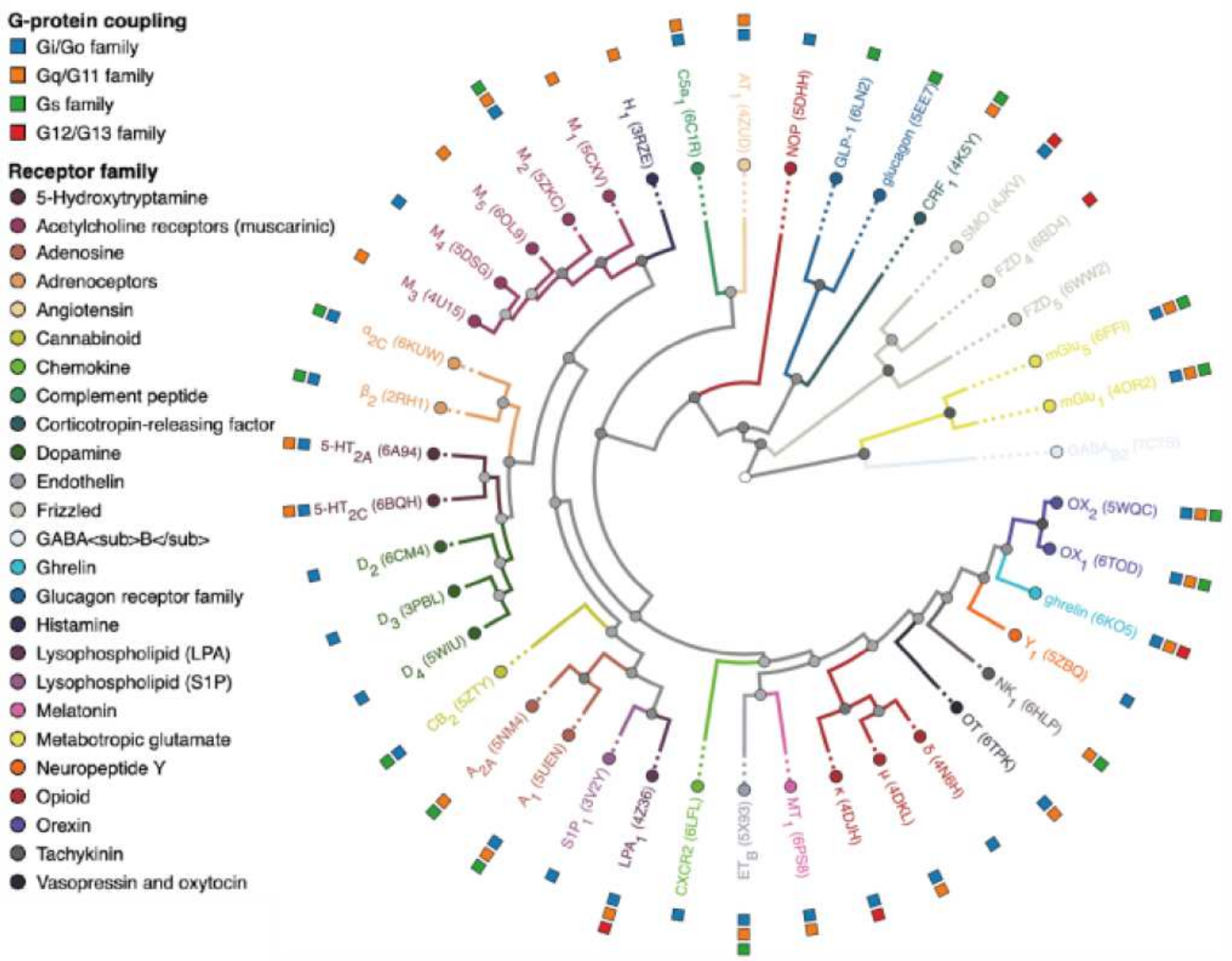


Figure 2

Structure similarity trees. The structure similarity trees cluster receptor structures by their conformation of the common seven transmembrane helix bundle and can assign annotations, such as inactive/intermediate/active state, G protein coupling and phylogenetic and pharmacological classifications (http://review.gpcrdb.org/structure_comparison/structure_clustering). Tree nodes have grey scale coloring illustrating the separation of clades based on a silhouette score13 (Methods) shown upon mouse hovering.