

The G protein database, GproteinDb

Gáspár Pándy-Szekeres^{1,2}, Mauricio Esguerra¹, Alexander S. Hauser¹, Jimmy Caroli¹, Christian Munk¹, Steven Pilger¹, György M. Keserű², Albert J. Kooistra¹ and David E. Gloriam^{1,*}

¹Department of Drug Design and Pharmacology, University of Copenhagen, 2100 Copenhagen, Denmark and

²Medicinal Chemistry Research Group, Research Centre for Natural Sciences, Budapest H-1117, Hungary

Received August 16, 2021; Revised September 10, 2021; Editorial Decision September 10, 2021; Accepted September 13, 2021

ABSTRACT

Two-thirds of signaling substances, several sensory stimuli and over one-third of drugs act via receptors coupling to G proteins. Here, we present an online platform for G protein research with reference data and tools for analysis, visualization and design of scientific studies across disciplines and areas. This platform may help translate new pharmacological, structural and genomic data into insights on G protein signaling vital for human physiology and medicine. The G protein database is accessible at <https://gproteindb.org>.

INTRODUCTION

Two-thirds of endogenous hormones and neurotransmitters (1), several sensory stimuli and over one-third of the FDA-approved drugs (2) mediate their actions via receptors coupling to G proteins. G proteins are intracellular heterotrimeric proteins consisting of α , β and γ subunits that dissociate to α and $\beta\gamma$ upon activation by the G protein-coupled receptor (GPCR). G proteins are named by their α subunit and are divided into four families which share homology and downstream signaling pathways: G_s (G_s and G_{olf}), $G_{i/o}$ (G_{i1} , G_{i2} , G_{i3} , G_o , G_z , G_{t1} , G_{t2} , G_{gust}), $G_{q/11}$ (G_q , G_{11} , G_{14} and G_{15}) and $G_{12/13}$ (G_{12} and G_{13}). The theoretical G protein ‘couplome’ in human spans the potential interaction of these 16 G proteins with ~800 receptors totaling 12 800 couplings or non-couplings. Recently, breakthroughs in biosensor development (3–6) yielded the first large-scale systematic quantifications of couplings (3,4) which have been unified in a recent meta-analysis (7). The structural elucidation of GPCR–G protein binding currently covers >120 complexes (https://gproteindb.org/structure/gprot_statistics) which make up a majority of the new and nearly all cryo-EM GPCR structures (8). Combined struc-

tural and sequence analysis has uncovered GPCR–G protein selectivity determinants (9).

Despite this information, more is needed to realize huge scientific potential. A structural mechanism, ‘conformational selection’ alters G protein selectivity in ligand-dependent ‘signal bias’ (10,11) but we lack the molecular mechanistic understanding to rationally design probes with functional selectivity and drugs with fewer adverse effects. There are no GPCR structure complexes of the $G_{12/13}$ family and receptors with only weak G protein coupling require more insight into their stabilization. The encoding of selectivity in GPCR and G protein sequences has not been described for determinants in the ligand-binding site (12) most important for drug design. Furthermore, G proteins (13) and their signaling interface (14) have recently been identified as (direct) therapeutic targets.

Fully answering these and many related scientific questions would only be possible through a consolidated community data and analysis infrastructure enabling exploitation of the above and coming data in integrative research. Here, we present an extensive online platform for G protein research. The interactive platform features e.g., a G protein coupling atlas, annotated structural templates, interface interactions and matching, and predicted selectivity determinants for mutagenesis. By providing one-stop reference data and accessible data-driven analysis and visualization tools, this platform may help translate more of the many new data into integrative insights and an actionable foundation to advance G protein research across the scientific disciplines and areas.

MATERIALS AND METHODS

Coding framework

We built the new resource by re-using the GPCRdb framework (15–17) which uses a Django Framework and the packages BioPython (18), NumPy (19), SciPy (20), and MODELLER (21). For all data browsers (i.e. the couplings,

*To whom correspondence should be addressed. Tel: +45 93 56 54 49; Fax: +45 35 33 60 41; Email: david.gloriam@sund.ku.dk
Present addresses:

Mauricio Esguerra, Nostrum Biodiscovery, Carrer de Baldiri Reixac 10-12, 08028 Barcelona, Spain.
Christian Munk, Novozymes A/S, Biologiens Vej 2, 2800 Kongens Lyngby Copenhagen, Denmark.
Steven Pilger, Explorative Biology and Bioinformatics, LEO Pharma A/S, 2750 Ballerup, Denmark.

structures, structures models and coupling determinants) we applied the DataTables.js (<https://datatables.net>) module in conjunction with yadcf.js (<https://yadcf-showcase.appspot.com>) which support sorting and filtering. The visualizations were written in JavaScript with the largest use of the D3.js framework (<https://d3js.org>) to generate SVG figures and animations. While initial versions of some resources were published in (9), new data and functionality have been added here along with many new resources tailored for G protein research.

GPCR–G protein coupling and selectivity resources

G protein couplings were filtered, normalized and aggregated onto families as described in (7) and presented in an interactive browser (<https://gproteindb.org/signprot/couplings>). GPCR–G protein couplings have a confidence filter that by default restricts the quantitative couplings to those supported by a second dataset, while the qualitative (primary versus secondary) couplings in Guide to Pharmacology are instead typically supported by multiple literature references. The requirement for shared coupling is applied on the G protein level—also for families—to avoid the issue of apparent support of family couplings when subtypes differ. G protein coupling selectivity profiles were visualized in a Venn diagram (https://gproteindb.org/signprot/statistics_venn) which, by GPCR class, intersects the receptor sets that couple to the four G protein families. This differs from (9) by adding (i) new datasets ((9) only covered GtP), (ii) a table to select all GPCRs in a class that couple to a given G protein family and (iii) the ability to filter obtained receptor sets based on their classification (lower left in Figure 1). Furthermore, for each GPCR class, we mapped G protein family couplings onto a classification tree of all human receptors by their ligand types (e.g. peptide or aminergic) and receptor families sharing endogenous ligand (<https://gproteindb.org/signprot/statistics.tree>). This differs from (9) which used (i) a phylogenetic classification (which is arbitrary for many cross-class and orphan receptor comparisons), (ii) GtP data only and (iii) a single cross-class tree (for which receptor names are illegible even in double column figures).

GPCR–G protein structure models

The GPCR–G protein structure complex models were built by extending the pipeline described in (17). The same steps are used for the homology modelling of the receptor; except, to ensure correct coupling between the receptor and the signaling protein the main template selection is limited to GPCR–G protein complex structures. Only those complex models are built where there is a structure in the GPCR class and the G protein subfamily. The five main template selection criteria are the same for all complex models: (i) GPCR class, (ii) G protein α subunit, or if that is not available: (iii) G protein subfamily, followed by the highest: (iv) GPCR sequence similarity and (v) resolution. The modelling of the G protein α subunit includes the swap-in of an alternative template for the Helical domain when it is missing from the main template; missing loop coordinates get alternative swap-in templates or are freely modeled, and mutated side chains are reverted to wild type.

GPCR–G protein interface interactions

Pairwise interface interactions were annotated based on geometric rules specified in Supplementary Table S1. Each type of amino acid interaction between a GPCR and a G protein is described on a general and a specific level, where the general level takes the underlying biochemistry into account (e.g. aromatic interaction), while the specific level considers geometric properties (e.g. face-to-edge). Each structural interaction annotation was subsequently enriched with the chemical properties of the participating amino acids. These properties are analyzed to generate an interface interaction fingerprint, a representation of the most conserved properties of the interacting residues across the receptor sequences of all structural templates. This fingerprint can be used to match across all receptors from the same GPCR class to list receptors by decreasing similarity and hence ability to form the same residue and G protein interactions.

Coupling determinant mutation design

The sequence-based coupling determinants were implemented using our recently published tool to identify sequence signatures (15). Based on the user-specified G protein family and receptor of interest, the tool collects all receptors from the same GPCR class for which coupling data is available. Subsequently, two sets of receptors are created, the ‘couplers’ and the ‘non-couplers’. The sequence signature is calculated as previously described (15) for the two receptor sets spanning all residue positions with an associated generic number. Based on the sequence signature, receptor positions lacking a conserved positive property (score > 10%), which potentially contributes to coupling, are proposed as a mutant by introducing the most conserved amino acid of the binding receptors having that positive property. Vice versa for negative properties (score < –10%), which contribute to non-coupling, mutations are proposed for receptor positions into the most conserved amino acid of the binding receptors lacking that negative property. The inverse approach is applied if the user selects to instead obtain mutations to decrease/abolish coupling to a G protein family. In addition to the suggested mutations, information is provided for each residue position about: sequence conservation, known G protein interactions, known ligand interactions, and the availability of mutation data.

RESULTS

GPCR–G protein coupling atlas and selectivity

The ‘G protein couplings’ (<https://gproteindb.org/signprot/couplings>) integrates data from quantitative profiling studies (3–5) and literature annotation from the Guide to Pharmacology database (22) letting any researcher find and compare couplings in one place (Figure 1A). To ensure confidence and comparability of couplings, all data is consistently filtered and normalized ((7) and Methods). Users can modify coupling confidence or coverage using cut-offs counting the number of supporting datasets or the standard deviations from basal signal. Cross-dataset filtering by mean $\log(E_{\max}/EC_{50})$, pEC_{50} and E_{\max} values can differentiate receptors with strong or no/weak coupling to a G

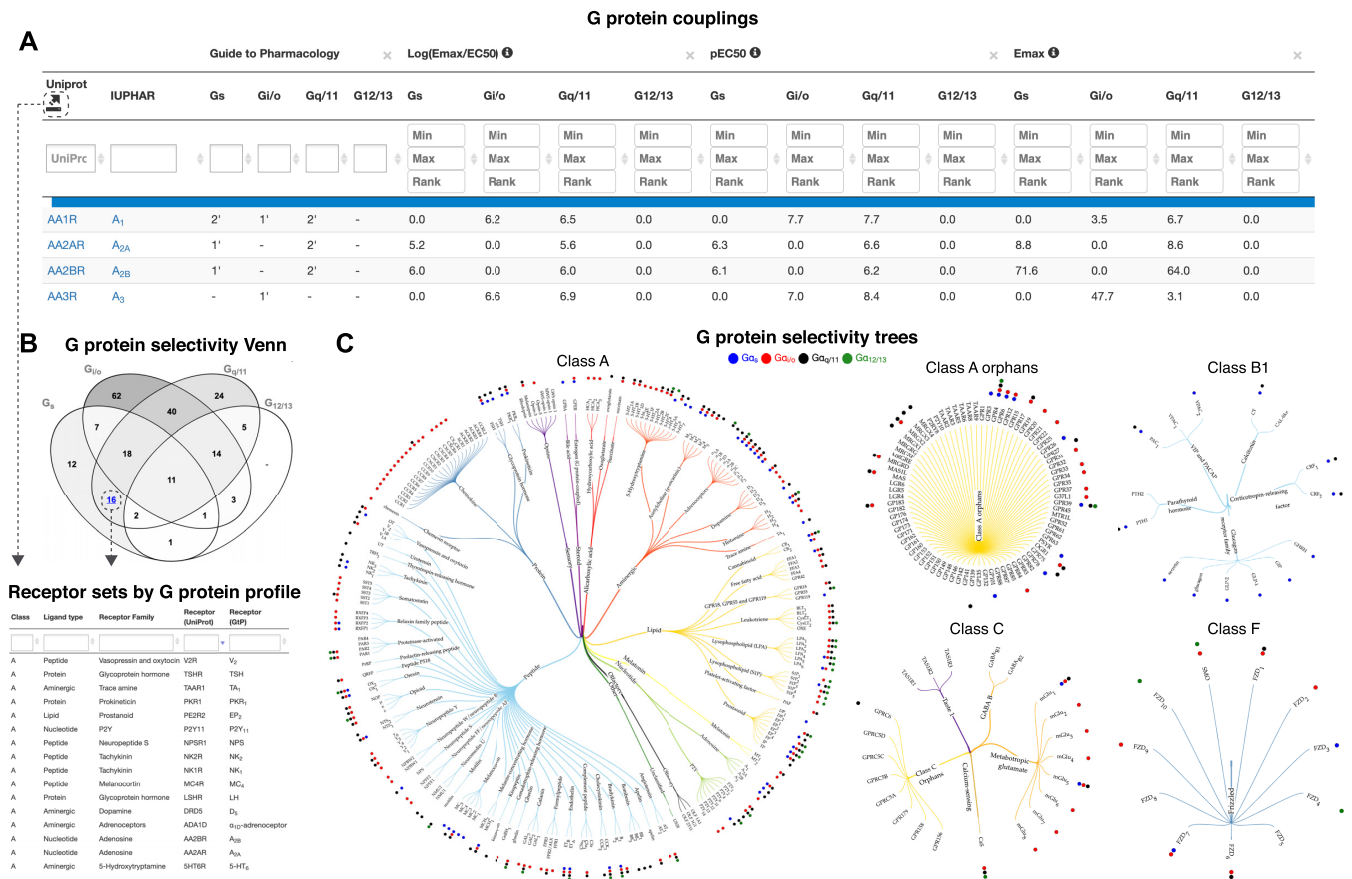


Figure 1. G protein couplings and selectivity. (A) G protein couplings encompass Guide to Pharmacology (GtP) primary/secondary transducers (22) and log(E_{max}/EC_{50}), pEC₅₀ and E_{max} values from profiling studies (3–5) and a means thereof. The confidence and coverage of quantitative couplings can be adjusted based on counts of supporting datasets (default 2) and standard deviations from basal signal (default 1.4). (B) G protein selectivity profiles can be intersected in a Venn diagram counting the receptors in each class. (C), G protein selectivity trees map couplings to trees classifying receptors by class, ligand type and receptor families sharing endogenous ligand(s). (B, C) Dashed arrows illustrate paths to obtain coupling-based selection of receptor sets for further study.

protein subtype or family. Furthermore, a ‘G protein selectivity Venn’ (<https://gproteindb.org/signprot/statistics.venn>) can discriminate the receptors in each GPCR class by their profile of combined G protein family couplings (Figure 1B). Finally, a ‘G protein selectivity tree’ (https://gproteindb.org/signprot/statistics_tree) maps G protein family couplings onto receptor classification trees – one for each GPCR class and further classified through an alphabetic listing of ligand types (e.g. peptide or lipid) and receptor families sharing endogenous ligands (e.g. serotonin receptors) (Figure 1C). These three resources present complementary means to analyze G protein couplings and to select sets of receptors for further study.

Structures

The G protein ‘Structures’ (https://gproteindb.org/structure/g_protein_structure_browser) include all unbound and GPCR complex structures from the Protein Data Bank (23). We additionally provide a refined version in which missing or mutated receptor and G_α regions are re-modelled based on other more complete structures while mutated residues are reverted to wildtype (Figure 2A). Structure selection is guided by information

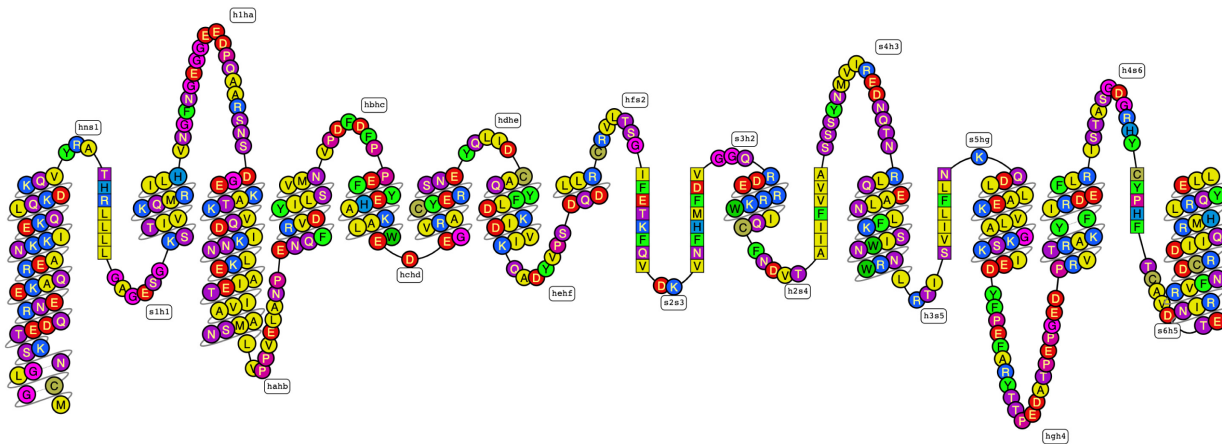
about G protein family and receptor classification, subunit isoform, species origin, ligand name and modality, structure determination method, resolution and author names (Supplementary Figure S1). The obtained templates can be copied (PDB identifiers) for analysis using GPCRdb’s structure comparison tools (24) or exported with their selected data (Excel) for further analysis. Furthermore, this resource is the first to provide ‘Structure models’ (https://gproteindb.org/structure/complex_models) of >3000 GPCR–G protein complexes that are not yet covered by experimental structures but can be modelled based on a similar template from the same GPCR class and G protein family. The structural topology of residue positions can be mapped in a ‘snakeplot’ (Figure 2B) from the ‘G protein page’ (below). The refined and modeled structure complexes expand the GPCR–G protein ‘couplome’ that can be studied across basic and applied structure-based research.

Receptor–G protein interface

The interfaces of GPCR–G protein structure complexes can be analyzed in the ‘Interface interactions and profiling’ (<https://gproteindb.org/signprot/matrix>) to identify residue

A

Snakeplot mapping residue properties



B

Sequence alignment and property conservation



C

Protein sequence with segment annotation

HN	TEDQGNDEKA	QREANKKIEK	QLQKDKLAYK	ATHRLLLLGA	GESGKSTIVK
MGCLGGNSKT	10	20	30	40	50
h1ha	HA	HA	HA	HA	hahb
QMRILHVNGF	70	80	90	100	110
HB	hbhc	HC	HD	HE	hdhe HE
LANPENQFRS	130	140	150	160	170
hehf	HF	hfs2	S2	s2s3	S3
LDKIDSIKLA	190	200	210	220	230
h2s4	S4	s4h3	H3	h3s5	S5
NDVTAIFVVA	250	260	270	280	290
hgh4	H4	H4	H4	H4	h4s6
EKVLGASKI	310	320	330	340	350
S6	s6h5	H5	H5	H5	STASGDGKHY
CYPHPTCAVD	370	380	390		

D

Generic residue number table

Generic residue number	G _{αs2}	G _{αi1}	G _{αq}	G _{α12}
	Human	Human	Human	Human
H5				
G.H5.01	T369	T329	T334	T356
G.H5.02	E370	K330	E335	E357
G.H5.03	N371	N331	N336	N358
G.H5.04	I372	V332	I337	V359
G.H5.05	R373	Q333	R338	R360
G.H5.06	R374	F334	F339	F361
G.H5.07	V375	V335	V340	V362
G.H5.08	F376	F336	F341	F363
G.H5.09	N377	D337	A342	H364
G.H5.10	D378	A338	A343	A365
G.H5.11	C379	V339	V344	V366
G.H5.12	R380	T340	K345	K367
G.H5.13	D381	D341	D346	D368

Figure 3. Sequence data in the online G protein research platform. (A) Snakeplot visualization of amino acid topological positions and properties. All amino acids are colored by property: polarity for most but backbone modifying ability for proline and glycine and disulfide-formation for cysteine. (B) Sequence alignment of the receptor-interacting helix 5 of all human 16 G_α proteins along with conservation measures for amino acids and residue groups with similar property and size, and numeric amino acid descriptors, ‘z-scales’ (29) (from <https://proteindb.org/alignment/gproteinselection>). (C) Protein sequence mapping onto segments by secondary structures (helices, β-sheet and loops). (D) Protein-specific and common residue numbers (25) for G_{αs}, G_{αi1}, G_{αq} and G_{α12} (from <https://proteindb.org/residue/residuetable.gprot>). These number tables can also be downloaded in Excel format or retrieved programmatically via a RESTFUL-API web service to integrate the numbering in any dataset and analysis method. (A, B) Taken from the ‘G protein page’ (<https://proteindb.org/signprot>) for G_{αs}. (A–C), Common residue numbers (25) can be shown by mouse hover.

interactions and their frequencies (Figure 2C). Residue interaction frequencies (% and grayscale in Figure 2C) can be analyzed for distribution across structures and filtered to the desired stringency. All residues have generic residue numbers (25,26) for structurally corresponding positions and can be color-coded by properties or interaction types. The property consensus (27) of receptor residues form an interface ‘fingerprint’ which can be matched to a sequence alignment of all GPCRs in the class to profile receptors by their conservation of the interface while inspecting known couplings (Supplementary Figure S2). The platform also features ‘Interface mutations & chimera’ (https://files.gpcrdb.org/GPCR--Gprotein_Mutations.xlsx) from literature annotations (9). These experimental data span modifications of receptors, G proteins or both and their qualitative and quantitative effects. These resources offer unique means for structure-based identification of GPCR–G protein interfaces and for comparison to experimentally characterized GPCR–G protein coupling profiles and determinants.

Sequence topology, alignment and generic residue numbers

The ‘G protein page’ (<https://gproteindb.org/signprot>) summarizes sequence, structural and mutagenesis data. Residue positions are mapped to a ‘snakeplot’ which can be custom colored or display stored functional data about the receptor interface (below), genetic variants (28), post-translational modification sites and a selectivity barcode (9) (Figure 3A). The full-length G_{α} protein sequence mapped to segments by secondary structures (helices, β -sheets and loops) (Figure 3C). The ‘G protein alignments’ (<https://gproteindb.org/alignment/gproteinselection>) can be customized to cover specific G proteins, sequence segments or common residue number positions (25) (Figure 3B). Each alignment comes with conservation measures for amino acids and residue groups with similar property and size, as well as numeric amino acid descriptors and ‘z-scales’ (29). Furthermore, ‘Generic residue number tables’ (https://gproteindb.org/residue/residuetable_gprot) tabulate G protein-specific and common residue positions (25) (Figure 3D). Together, the sequence alignments, detailed conservation statistics and generic residue numbers provide the foundation for analysis of the sequence basis underlying G protein structure and function, e.g., determinants of molecular interactions and structural conformations.

Coupling determinant mutation design tool

The ‘Coupling determinant mutation design tool’ (https://gproteindb.org/mutations/gprot_coupling) features data-driven prediction of receptor residue determinants of G protein activation (Supplementary Figure S3). This is based on all couplings (above) and distinct conservation among coupling and non-coupling GPCRs, respectively for a G protein family of interest. The conservation measure goes beyond traditional amino acid identities to analyze groups of residues with similar properties and size allowing shared molecular interactions. Depending on the research question, users can choose to either strengthen or weaken G protein coupling. Accordingly, mutations are suggested to

introduce missing and remove conserved consensus amino acids from the receptor set with the desired and undesired coupling status, respectively. Validation is built-in by side-by-side tabulation of suggested mutations with GPCR–G protein interface interactions (above) and effects from literature mutations (15,30), while we invite the research field to feedback new mutagenesis results via a standardized Excel file. Notably, the intersection with interface interactions enables studies focusing on either interface determinants or allosteric modulating G protein binding.

DISCUSSION

Taken together, the online platform integrates diverse G protein sequence, structure and function data and makes accessible sophisticated analysis tools. The G protein coupling atlas opens for one-stop access to consistently normalized reference data from all major datasets with user-defined confidence and coverage cut-offs tailoring to the needs of each study. For example, functional and mechanistic studies will also benefit from the possibility to distil receptor sets with a specific G protein selectivity based on the atlas and the interactive Venn diagram. Given that there are already over 120 GPCR–G protein complexes (https://gproteindb.org/structure/gprot_statistics) and that these make up the majority of receptor cryo-EM structures, the annotated structures presented herein will help substantially to keep track of new structures and selecting study templates. Furthermore, as cryo-EM structure determination is challenging at a high resolution, the refinement based on other experimental templates model-in many missing sidechains or loop segments. While the vast majority of GPCR–G protein complexes lack an experimental structure, our structure models greatly expand the ‘couplome’ for which we can generate structure-based hypotheses across mutagenesis, dynamics, kinetics, molecular mechanistic and drug design studies. The ‘Interface interactions and profiling’ tool may shed new light on how receptors bind G proteins at the interface (9). However, G protein signaling can also be modulated allosterically (12). Therefore, the ‘Coupling determinant mutation design tool’ presents a unique basis to reveal such allosteric determinants through mutagenesis experiments. Of note, its data differs from the PRECOG server (31) by spanning not one but four datasets (3–5,22)—a necessity to remove unsupported couplings (7)—and its unique residue property-based signatures are integrated with known interface interactions and mutation effects. Hence, we expect that the platform presented here will inspire many studies across basic and applied research and disciplines, aiding the elucidation of, e.g., constitutive activity (32), pre-coupling of G proteins (33,34) and ligand-dependent biased G protein signaling (10).

DATA AVAILABILITY

All data is available via the web (Section ‘GproteinDb’ in <https://gproteindb.org>) and GitHub (https://github.com/protwis/gpcrdb_data). Documentation is available at <https://docs.gpcrdb.org>. 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>).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

Nevin Lambert is acknowledged for fruitful discussions. The authors are members of the COST Action CA18133, the European Research Network on Signal Transduction (<https://ernest-gpcr.eu>).

FUNDING

Independent Research Fund Denmark [Natural Sciences [8021-00173B]; Lundbeck Foundation [R218-2016-1266]; Novo Nordisk Foundation [NNF18OC0031226 to D.E.G.]; Hungarian Brain Research Program [2017-1.2.1-NKP-2017-00002 to G.M.K.] Funding for open access charge: Novo Nordisk Foundation [NNF18OC0031226]. *Conflict of interest statement.* After completing their contribution to this article, C.M., M.E. and S.P. moved to Novozymes A/S, Nostrum Biodiscovery and Leo Pharma A/S, respectively.

REFERENCES

- Foster, S.R., Hauser, A.S., Vedel, L., Strachan, R.T., Huang, X.P., Gavin, A.C., Shah, S.D., Nayak, A.P., Haugaard-Kedstrom, L.M., Penn, R.B. *et al.* (2019) Discovery of human signaling systems: pairing peptides to G protein-coupled receptors. *Cell*, **179**, 895–908.
- Hauser, A.S., Attwood, M.M., Rask-Andersen, M., Schioth, H.B. and Gloriam, D.E. (2017) Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.*, **16**, 829–842.
- Avet, C., Mancini, A., Breton, B., Le Gouill, C., Hauser, A.S., Normand, C., Kobayashi, H., Gross, F., Hogue, M., Lukasheva, V. *et al.* (2020) Effector membrane translocation biosensors reveal G protein and B-arrestin profiles of 100 therapeutically relevant GPCRs. bioRxiv doi: <https://doi.org/10.1101/2020.04.20.052027>, 24 April 2020, preprint: not peer reviewed.
- Inoue, A., Raimondi, F., Kadji, F.M.N., Singh, G., Kishi, T., Uwamizu, A., Ono, Y., Shinjo, Y., Ishida, S., Arang, N. *et al.* (2019) Illuminating G-protein-coupling selectivity of GPCRs. *Cell*, **177**, 1933–1947.
- Olsen, R.H.J., DiBerto, J.F., English, J.G., Glaudin, A.M., Krumm, B.E., Slocum, S.T., Che, T., Gavin, A.C., McCorvy, J.D., Roth, B.L. *et al.* (2020) TRUPATH, an open-source biosensor platform for interrogating the GPCR transducerome. *Nat. Chem. Biol.*, **16**, 841–849.
- Masuh, I., Ostrovskaya, O., Kramer, G.M., Jones, C.D., Xie, K. and Martemyanov, K.A. (2015) Distinct profiles of functional discrimination among G proteins determine the actions of G protein-coupled receptors. *Sci. Signal*, **8**, ra123.
- Hauser, A.S., Avet, C., Normand, C., Mancini, A., Inoue, A., Bouvier, M. and Gloriam, D. (2021) GPCR–G protein selectivity – a unified meta-analysis. bioRxiv doi: <https://doi.org/10.1101/2021.09.07.459250>, 08 September 2021, preprint: not peer reviewed.
- Munk, C., Mutt, E., Isberg, V., Nikolajsen, L.F., Bibbe, J.M., Flock, T., Hanson, M.A., Stevens, R.C., Deupi, X. and Gloriam, D.E. (2019) An online resource for GPCR structure determination and analysis. *Nat. Methods*, **16**, 151–162.
- Flock, T., Hauser, A.S., Lund, N., Gloriam, D.E., Balaji, S. and Babu, M.M. (2017) Selectivity determinants of GPCR–G-protein binding. *Nature*, **545**, 317–322.
- Kenakin, T. (2019) Biased receptor signaling in drug discovery. *Pharmacol. Rev.*, **71**, 267–315.
- Smith, J.S., Lefkowitz, R.J. and Rajagopal, S. (2018) Biased signalling: from simple switches to allosteric microprocessors. *Nat. Rev. Drug Discov.*, **17**, 243–260.
- DeVree, B.T., Mahoney, J.P., Velez-Ruiz, G.A., Rasmussen, S.G., Kuzak, A.J., Edwald, E., Fung, J.J., Manglik, A., Masureel, M., Du, Y. *et al.* (2016) Allosteric coupling from G protein to the agonist-binding pocket in GPCRs. *Nature*, **535**, 182–186.
- Campbell, A.P. and Smrcka, A.V. (2018) Targeting G protein-coupled receptor signalling by blocking G proteins. *Nat. Rev. Drug Discov.*, **17**, 789–803.
- Chan, H.C.S., Li, Y., Dahoun, T., Vogel, H. and Yuan, S. (2019) New binding sites, new opportunities for GPCR drug discovery. *Trends Biochem. Sci.*, **44**, 312–330.
- Kooistra, A.J., Mordalski, S., Pándy-Szekeres, G., Esguerra, M., Mamyrbekov, A., Munk, C., Keserü, G.M. and Gloriam, D.E. (2021) GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Res.*, **49**, D335–D343.
- Munk, C., Isberg, V., Mordalski, S., Harpsøe, K., Rataj, K., Hauser, A.S., Kolb, P., Bojarski, A.J., Vriend, G. and Gloriam, D.E. (2016) GPCRdb: the G protein-coupled receptor database - an introduction. *Br. J. Pharmacol.*, **173**, 2195–2207.
- Pándy-Szekeres, G., Munk, C., Tsonkov, T.M., Mordalski, S., Harpsøe, K., Hauser, A.S., Bojarski, A.J. and Gloriam, D.E. (2018) GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.*, **46**, D440–D446.
- Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.
- Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J. *et al.* (2020) Array programming with NumPy. *Nature*, **585**, 357–362.
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J. *et al.* (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*, **17**, 261–272.
- Webb, B. and Sali, A. (2016) Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.*, **86**, 2.9.1–2.9.37.
- Armstrong, J.F., Faccenda, E., Harding, S.D., Pawson, A.J., Southan, C., Sharman, J.L., Campo, B., Cavanagh, D.R., Alexander, S.P.H., Davenport, A.P. *et al.* (2020) The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: extending immunopharmacology content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY. *Nucleic Acids Res.*, **48**, D1006–D1021.
- Armstrong, D.R., Berrisford, J.M., Conroy, M.J., Gutmanas, A., Anyango, S., Choudhary, P., Clark, A.R., Dana, J.M., Deshpande, M., Dunlop, R. *et al.* (2020) PDBE: improved findability of macromolecular structure data in the PDB. *Nucleic Acids Res.*, **48**, D335–D343.
- Gloriam, D., Kooistra, A., Munk, C. and Hauser, A. (2021) An online GPCR structure analysis platform. *Nat. Struct. Mol. Biol.*, **30**, <https://doi.org/10.21203/rs.3.rs-354878/v1>.
- Flock, T., Ravarani, C.N.J., Sun, D., Venkatakrisnan, A.J., Kayikci, M., Tate, C.G., Veprintsev, D.B. and Babu, M.M. (2015) Universal allosteric mechanism for G[alpha] activation by GPCRs. *Nature*, **524**, 173–179.
- Isberg, V., de Graaf, C., Bortolato, A., Cherezov, V., Katritch, V., Marshall, F.H., Mordalski, S., Pin, J.P., Stevens, R.C., Vriend, G. *et al.* (2015) Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.*, **36**, 22–31.
- Kooistra, A., Mordalski, S., Pándy-Szekeres, G., Esguerra, M., Mamyrbekov, A., Munk, C., Keserü, G. and Gloriam, D. (2021) GPCRdb in 2021: Integrating GPCR sequence, structure and function. *Nucleic Acids Res.*, **49**, D335–D343.
- Hauser, A.S., Chavali, S., Masuh, I., Jahn, L.J., Martemyanov, K.A., Gloriam, D.E. and Babu, M.M. (2018) Pharmacogenomics of GPCR Drug Targets. *Cell*, **172**, 41–54.
- Sandberg, M., Eriksson, L., Jonsson, J., Sjöström, M. and Wold, S. (1998) New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J. Med. Chem.*, **41**, 2481–2491.

30. Munk,C., Harpsoe,K., Hauser,A.S., Isberg,V. and Gloriam,D.E. (2016) Integrating structural and mutagenesis data to elucidate GPCR ligand binding. *Curr. Opin. Pharmacol.*, **30**, 51–58.
31. Singh,G., Inoue,A., Gutkind,J.S., Russell,R.B. and Raimondi,F. (2019) PRECOG: PREdicting COupling probabilities of G-protein coupled receptors. *Nucleic Acids Res.*, **47**, W395–W401.
32. Berg,K.A. and Clarke,W.P. (2018) Making sense of pharmacology: inverse agonism and functional selectivity. *Int. J. Neuropsychopharmacol.*, **21**, 962–977.
33. Covicristov,S., Ellisdon,A.M., Suderman,R., Pon,C.K., Evans,B.A., Kleifeld,O., Charlton,S.J., Hlavacek,W.S., Canals,M. and Halls,M.L. (2018) Preassembled GPCR signaling complexes mediate distinct cellular responses to ultralow ligand concentrations. *Sci. Signal*, **11**, eaan1188.
34. Galés,C., Van Durm,J.J., Schaak,S., Pontier,S., Percherancier,Y., Audet,M., Paris,H. and Bouvier,M. (2006) Probing the activation-promoted structural rearrangements in preassembled receptor-G protein complexes. *Nat. Struct. Mol. Biol.*, **13**, 778–786.
35. Rasmussen,S.G., DeVree,B.T., Zou,Y., Kruse,A.C., Chung,K.Y., Kobilka,T.S., Thian,F.S., Chae,P.S., Pardon,E., Calinski,D. *et al.* (2011) Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature*, **477**, 549–555.