



GPCRdb in 2018

adding GPCR structure models and ligands

Pándy-Szekeres, Gáspár; Munk, Christian; Tsonkov, Tsonko M; Mordalski, Stefan; Harpsøe, Kasper; Hauser, Alexander S; Bojarski, Andrzej J; Gloriam, David E

Published in:
Nucleic Acids Research

DOI:
[10.1093/nar/gkx1109](https://doi.org/10.1093/nar/gkx1109)

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY](#)

Citation for published version (APA):
Pándy-Szekeres, G., Munk, C., Tsonkov, T. M., Mordalski, S., Harpsøe, K., Hauser, A. S., ... Gloriam, D. E. (2018). GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Research*, 46(D1), [gkx1109]. <https://doi.org/10.1093/nar/gkx1109>

GPCRdb in 2018: adding GPCR structure models and ligands

Gáspár Pándy-Szekeres^{1,2,†}, Christian Munk^{1,†}, Tsonko M. Tsonkov¹, Stefan Mordalski², Kasper Harpsøe¹, Alexander S. Hauser¹, Andrzej J. Bojarski² and David E. Gloriam^{1,*}

¹Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, DK-2100, Copenhagen, Denmark and ²Department of Medicinal Chemistry, Institute of Pharmacology, Polish Academy of Sciences, Smetna 12, 31-343 Krakow, Poland

Received October 01, 2017; Revised October 22, 2017; Editorial Decision October 23, 2017; Accepted October 25, 2017

ABSTRACT

G protein-coupled receptors are the most abundant mediators of both human signalling processes and therapeutic effects. Herein, we report GPCRome-wide homology models of unprecedented quality, and roughly 150 000 GPCR ligands with data on biological activities and commercial availability. Based on the strategy of 'Less model – more Xtal', each model exploits both a main template and alternative local templates. This achieved higher similarity to new structures than any of the existing resources, and refined crystal structures with missing or distorted regions. Models are provided for inactive, intermediate and active states—except for classes C and F that so far only have inactive templates. The ligand database has separate browsers for: (i) target selection by receptor, family or class, (ii) ligand filtering based on cross-experiment activities (min, max and mean) or chemical properties, (iii) ligand source data and (iv) commercial availability. SMILES structures and activity spreadsheets can be downloaded for further processing. Furthermore, three recent landmark publications on GPCR drugs, G protein selectivity and genetic variants have been accompanied with resources that now let readers view and analyse the findings themselves in GPCRdb. Altogether, this update will enable scientific investigation for the wider GPCR community. GPCRdb is available at <http://www.gpcrdb.org>.

INTRODUCTION

G protein-coupled receptors represent both the largest membrane and drug target protein families, accounting for 800 (4%) human genes and 34% of FDA-approved drug targets (1). The human GPCRs have been divided into six

classes (or families): A (Rhodopsin), B1 (Secretin), B2 (Adhesion), C (Glutamate), F (Frizzled) and Taste 2 based on evolutionary homology (2,3). All GPCRs share a common heptahelical transmembrane domain carrying the signal from the extracellular ligand to the intracellular signalling protein, typically one of four families of G proteins or β -arrestin, which can also internalise the receptor. Recently, structural biology and pharmacology breakthroughs have opened up new GPCR research fields exploring structural mechanisms and complexes, and novel principles to achieve a more selective action through 'biased agonism' preferentially activating a specific signal protein or 'allosteric modulation' of the natural ligand response from an alternative binding site. The GPCR database, GPCRdb (4–6) serves the wide GPCR community, currently ~1800 monthly users, with reference data, web server analysis tools and dynamic visualisation of data and statistics. Current data ranges all 398 human non-olfactory GPCRs and 16 G proteins, over 14 000 species orthologues, 30 328 binding site mutations, all 218 experimental structures, and 10 059 extracted ligand interactions.

GPCR crystal structures have now been determined for the major human classes A–C, F (D and E fungal and amoeba receptors, respectively), but not yet for the Adhesion (B1) and Taste 2 classes, which have atypical pharmacology in that they are not activated by an endogenous agonist. In total, structures are available for 46 unique receptors, however close to 90% of the non-olfactory receptors still lack a crystal structure (<http://gpcrdb.org/structure/statistics>). Furthermore, only eleven receptors have a structure in the active state. Hence, many GPCR studies utilise homology models for virtual screening, lead optimisation, elucidation of ligand binding sites, off-target rationalisation, modeling of complexes, and molecular dynamics (7). The latest GPCR Dock assessment (8) showed that the accuracy of such models has greatly increased, and the best models utilize more than one template. Our 5-HT_{1B} model achieved the best RMSD compared to the crystal structure

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

†These authors contributed equally to this work as first authors.

and was based on five backbone templates and a position-specific rotamer library derived from all GPCR structures. This strategy of '*Less model—more Xtal*' covers more of the model with experimental data by extending the main template with alternative local templates for missing or distorted segments. Herein, we present the first resource with pre-generated multi-template models of all human non-olfactory GPCRs in their inactive, intermediate and active states (except for classes C and F for which active and intermediate templates are still missing).

GPCR ligands have previously been collated in a dedicated resource, GLIDA database (9), but maintenance was discontinued in October 2010 when the database contained 30,399 ligands. Today, the Guide to PHARMACOLOGY database offers curated reference ligands for a large number of protein families, including GPCRs (10), however this is just a small subset of the publically available ligands. The most comprehensive resource for biological activities is the ChEMBL database, which has compiled data from publications, patents and commercial sources (11). However, ChEMBL data is not presented and filtered to be directly accessible by many users. A further resource is the Ki database of the psychoactive drug screening programme containing broad target profiling, however most of which are single-point values (12). Herein, we present a new GPCR ligand resource that applies the same quality requirements as Guide to PHARMACOLOGY, but encompasses ligands from ChEMBL, GLIDA and the PDSP Ki database. It gives users novel possibilities, such as commercial availability, target selectivity and selection of ligands suitable for crystallography experiments.

METHODS

Technical features of the homology modeling pipeline

The new GPCRdb homology modeling pipeline (Figure 1) updates the templates and models upon each database update. It was developed in Python in the Django web framework, utilizes several tools from Biopython (13), saves data in a PostgreSQL database, builds the final models with the help of MODELLER 9.18 (14) and uses NGL (15) for model visualisation. It leverages on integration of several unique manually annotated GPCRdb resources, including structure-based sequence alignments with generic residue numbers (16), GPCR structures (all in PDB) with defined preferred chain, helix segment borders, and distorted regions; and a position-specific sidechain rotamer library (17).

Pairwise sequence alignment

Target-template pairwise sequence alignments are fetched from GPCRdb.

Main template selection

For each receptor with a structure at least one (the most complete) representative template (currently totalling 169) has been manually annotated to define aligned segments (TM1–7, H8, ICL1–2 and ECL1–2), bulges, constrictions

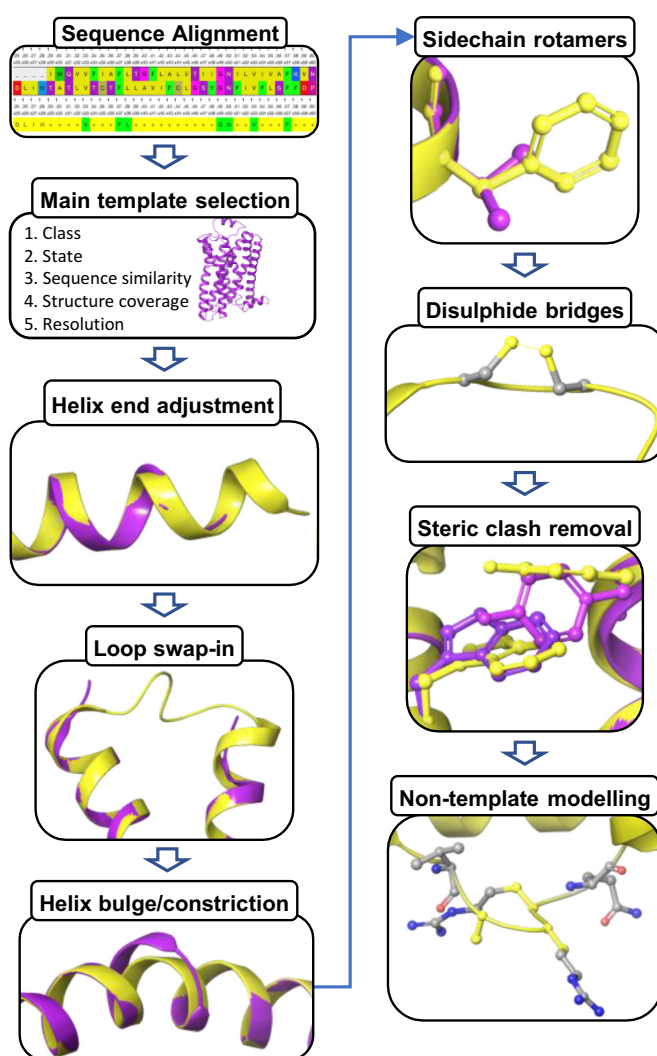


Figure 1. The GPCRdb homology model pipeline builds on the principle of '*Less model – more Xtal*' meaning that more of the target is covered by structural templates, as an incomplete or partly distorted main template is complemented with alternative local templates. All templates and models are automatically updated upon each database update. Models are produced for all the human non-olfactory GPCRs in the inactive, and where templates exist (classes A and B1), also intermediate and active states.

and any missing or distorted regions. A model main template is selected based on: (i) class (B2 to B1 and Taste2 to A), (ii) state (active/intermediate/inactive), (iii) receptor sequence similarity, (iv) structure coverage (no. generic residue positions (16) with coordinates) and (v) resolution. The template states are defined based on the extend of outward and inward movement of TM6 and TM5, respectively, on the intracellular, G protein binding side (<http://docs.gpcrdb.org/structures.html#structure-state>).

Helix end adjustment

Transmembrane helix ends are in some places (typically in ICL3) altered due to protein fusions, H8 is sometimes missing or truncated, and both can be distorted due to crystal packing effects. Non-native parts of helices are removed from the initial model and missing segments are modeled

based on a local alternative template, which is superposed onto the last helical turn of the main template.

Loop modeling

As a unique feature, GPCRdb has alignments and generic residue numbers for loop regions with a conserved secondary structure; ECL1: three-residue turn motif with terminal amino acids packing between TM2-TM3, ICL1: single-turn helix, ECL2: cysteine forming a disulphide bridge to TM3 plus two following residues, and ICL2: two-turn helix. This allows our pipeline to model these loops based on alternative templates, where needed. Structurally disordered loops are also modeled based on an alternative template if the main template is incomplete, distorted or has a different length than the target. An alternative template is chosen based on length (as many residues as the target), sequence similarity, structure coverage and resolution. The template is superposed on the adjacent helix ends, except for ECL2 for which a first and latter half are anchored on the backbone of the conserved cysteine (position 45x50). If the main template structure lacks the backbone coordinates of the conserved ECL2 cysteine, these are inferred from an alternative template that is superposed using the last TM4 residue, first TM5 residue and the conserved cysteine in TM3 (3x25) that makes up the other half of the disulphide bridge. Altogether, this ensures that only a few remaining loop stretches need to be modeled without a template.

Helix bulge and constriction treatment

When the target or template has a unique 7TM helix bulge or constriction (a local turn of π - or 3_{10} -helix, respectively), a swap-in template with the same configuration as the target is applied. Such helix distortions are associated with characteristic sequence motifs, typically involving proline or glycine residues, and have been uniquely accounted in the GPCRdb alignments and residue numbers (16).

Sidechain rotamer modeling

Amino acids not conserved between the target and main template are defined from a position-specific rotamer library extracted from all experimental GPCR structures (17). A conserved amino acid from the same structure/sequence position is superposed onto the backbone atoms of the original residue. In practice, this generates a chimeric structural template with a significantly higher sequence identity, often over 20%, to the target than a single template. Also, template mutations, introduced to facilitate expression and crystallization, are reverted to the wildtype residue using the rotamer library.

Disulphide bridge introduction

Two conserved disulphide bridges are modeled: (i) one cross-class between the top of TM3 (3x25) and ECL2 (45x50) and (ii) another present in some class A receptors that links two cysteines in ECL3 or ECL3 to the top of TM6.

Non-template-based modeling of remainder

Residues not covered by a structural template get their missing backbone and/or sidechains modeled using MODELLER (14), which also re-models non-conserved sidechains with a steric clash.

Building of a GPCR ligand database with information about commercial availability

Ligands and biological activities were imported from ChEMBL_23 using the ChEMBL protein IDs and API (11). Records were filtered to only include the standardised/translated pChEMBL values and the binding and functional assay types, while excluding those with disqualifying statuses in 'data_validity_comment' and 'activity_comment' (see <https://www.ebi.ac.uk/chembl/faq>). Ligand chemical (molecular weight, no. hydrogen bond donors/acceptors and rotatable bonds) and commercial availability data (vendor and catalogue number) were retrieved from PubChem (18) after building a translation routine between ChEMBL and PubChem compound identifiers.

RESULTS

Recent resources for GPCR drugs, G protein coupling selectivity and natural genetic variants

Three recent high-impact publications have been accompanied with tailored resources that integrate biologically and clinically relevant data and let researchers make their own analyses online at GPCRdb. The first new section, 'Drugs', supplement an analysis of the new trends for GPCR drugs, targets and indications covering FDA-approved drugs and agents in clinical trials (1). The first item is Drug statistics concerning, e.g. targets, drug molecule types and modes of action, disease indications and clinical phase distributions. Secondly, a Drug target mapping page shows the distribution of in-trial and approved clinical agents onto a target tree classification of GPCR classes, ligand types and receptor families to give a rapid overview for targets of interest. Finally, a Drug browser allows for navigation and filtering of complete information for drugs, targets, indications, clinical progression and references.

The second new section, 'Signal Proteins', was added as part of a landmark study on GPCR-G protein selectivity (19). This features selection of human receptor sets by their GPCR-protein coupling data, G protein sequence alignments and residue topology diagrams, structures, interface mapping and search and in vitro mutations of GPCRs and G proteins with effect on signalling. The third new (sub)section Genetic variation (found in the Receptors section) was added as part of a comprehensive report on pharmacogenomics of GPCR drug targets (20). This subsection features Variation statistics with counts and densities of missense and loss of function polymorphisms across receptors, receptor families, ligand types and GPCR classes; a Receptor variant browser tabulating missense variant data and functional annotations (such as ligand and effector protein binding sites, post-translational modification sites, sodium ion pocket and micro-switches) by target(s) and

| OVERALL | CCR2 | AA1R | PAR2 | APJ | Avg |
|-------------------------|-------------|-------------|-------------|------------|------------|
| GPCRdb | 1,8 | 2,7 | 3,1 | 3,7 | 2,8 |
| GPCR-I-TASSER | 1,9 | 3,4 | 4,3 | 3,3 | 3,2 |
| SwissModel | 1,8 | 2,8 | 5,8 | 3,7 | 3,5 |
| GoMoDo | 3,5 | 2,9 | 4,6 | 3,7 | 3,7 |
| GPCR-ModSim | 2,7 | 3,2 | 3,9 | 4,9 | 3,7 |
| GPCRM | 3,7 | 4,0 | 3,9 | - | 3,9 |
| GPCR-SSFE | 1,9 | 6,0 | 5,2 | 8,6 | 5,4 |
| Avg | 2,5 | 3,6 | 4,4 | 4,7 | 3,7 |
| 7TM | CCR2 | AA1R | PAR2 | APJ | Avg |
| GPCRdb | 1,6 | 1,8 | 2,6 | 3,0 | 2,3 |
| GPCRM | 2,8 | 2,0 | 2,4 | - | 2,4 |
| GPCR-I-TASSER | 1,6 | 2,7 | 3,3 | 2,6 | 2,6 |
| GPCR-ModSim | 1,8 | 2,4 | 2,9 | 3,5 | 2,7 |
| SwissModel | 1,6 | 1,9 | 4,3 | 2,9 | 2,7 |
| GoMoDo | 3,2 | 2,1 | 3,6 | 3,3 | 3,1 |
| GPCR-SSFE | 1,7 | 3,5 | 3,6 | 4,8 | 3,4 |
| Avg | 2,0 | 2,3 | 3,2 | 3,4 | 2,7 |
| LIGAND SITE (4Å) | CCR2 | AA1R | PAR2 | APJ | Avg |
| GPCRdb | 1,2 | 1,7 | 1,4 | 2,7 | 1,8 |
| GPCRM | 2,5 | 1,7 | 4 | - | 2,7 |
| SwissModel | 1,1 | 1,4 | 4,1 | 4,6 | 2,8 |
| GoMoDo | 2,6 | 1,7 | 3,7 | 3,7 | 2,9 |
| GPCR-I-TASSER | 1,8 | 3,5 | 3,1 | 3,4 | 3,0 |
| GPCR-ModSim | 1,3 | 1,6 | 3,3 | 5,8 | 3,0 |
| GPCR-SSFE | 1,4 | 5,9 | 7,0 | 10,8 | 6,3 |
| Avg | 1,7 | 2,5 | 3,8 | 5,2 | 3,2 |
| ECL2 | CCR2 | AA1R | PAR2 | APJ | Avg |
| GPCRdb | 2,7 | 5,1 | 3,0 | 3,8 | 3,7 |
| GPCR-ModSim | 2,3 | 5,1 | 3,0 | 6,0 | 4,1 |
| GoMoDo | 2,2 | 5,1 | 5,2 | 4,2 | 4,2 |
| GPCR-I-TASSER | 2,1 | 6,4 | 5,0 | 3,7 | 4,3 |
| SwissModel | 2,4 | 5,8 | 5,2 | 4,8 | 4,6 |
| GPCR-SSFE | 2,7 | 6,6 | 8,4 | 8,2 | 6,5 |
| GPCRM | 4,3 | 7,6 | 7,7 | - | 6,5 |
| Avg | 2,7 | 6,0 | 5,4 | 5,1 | 4,8 |
| MAIN TEMPLATE | CCR2 | AA1R | PAR2 | APJ | Avg |
| Seq similarity (%) | 88 | 68 | 55 | 54 | 66,3 |
| Seq identity (%) | 80 | 53 | 36 | 34 | 50,8 |
| Receptor | ccr5 | aa2ar | par1 | agtr1 | |
| PDB ID | 4MBS | 5IU4 | 3VW7 | 4ZUD | |

Figure 2. RMSD (Å) values of database/server homology models calculated directly upon release of the first structure for four GPCRs (AA1R:

mapping their topology in receptor residue diagrams; as well as an Estimated economic burden caused by genetic variation along with UK NHS sales and prescription information per drug.

Ready-to-use models of several states for all non-olfactory human GPCRs

GPCRdb contains inactive state models of all human non-olfactory GPCRs and intermediate as well as active state models for classes (families) A (Rhodopsin), B1 (Secretin), B2 (Adhesion) and T (Taste 2). This includes also refined experimental structures filling in missing and replacing distorted regions and reverting mutated amino acids (added to the Structure Browser). Models are listed on the receptor pages, but also available from a new homology model browser (http://gpcrdb.org/structure/homology_models) allowing for multiple structures to be filtered, superposed, assigned generic residue numbers and downloaded. Finally, each receptor model has a page with embedded visualisation and details of which templates were utilised for which segment.

Homology model accuracy

Figure 2 shows RMSD values of database/server homology models calculated directly upon release of the first structure for four GPCRs. The GPCRdb homology models have the best average RMSD values in all four categories: overall (2.8 Å), 7TM (2.3 Å), ligand site (1.8 Å) and ECL2 (3.7 Å). Among other databases with pre-generated models (blue background), GPCR-I-TASSER performs the best, especially in the overall and 7TM. In contrast, three modeling web servers (gray background) perform best in different categories; SWISS-MODEL: overall, GPCRM: 7TM and ligand site, and GPCR-ModSim: ECL2. As could be expected, the RMSDs of the overall and 7TM models correlate directly with the sequence similarities and identities to the main template (bottom of figure). Altogether, the comparisons confirm the technological advantages of the GPCRdb

←
5UEN, CCR2: 5T1A, PAR2: 5NDD and APJ: 5VBL). Notably, the GPCRdb homology models have the best average RMSD values in all four categories. Superimposition was performed on backbone heavy atoms. For comparability, RMSD calculations were restricted to residues present in all of the models. SWISS-MODEL settings: default, used main template listed on top. GOMoDo settings: Blast – 2 search rounds, MODELLER options—two models, global alignment (no realign templates), no loop refinement, chose model with best normalized DOPE score (PAR2 model was sixth best, ones before were incomparable due to a shift in sequence numbering). GPCRM settings (kindly provided by the developers): beta version (under development), ECL2 disulphide bridge specified, Task mode Auto, Set of templates inactive, Lysozyme do not add from template, Rosetta loop-modeling yes, fast, additional options left default. Selected r01 model. GPCR-ModSim settings (kindly provided by the developers): group inactive, alignment was edited, number of models 10, selected one with best DOPE score, no Lenard-Jones restraints, select top templates – default, ECL2 disulphide bridge specified, short loops ICL1, ECL1, ICL2 and ECL3 were modeled with loop modeling, number of models 5, selected one with best DOPE score, no MD. GPCR-i-TASSER model selection from repository: Model 1. GPCR-SSFE model selection from repository: EntireModel1, loop versions 0.

Example model: Inactive state GPR75

Based on 10 backbone and 27 rotamer templates

| Segment | Start pos | End pos | Backbone Template |
|---------|-----------|---------|-------------------|
| TM1 | 35 | 38 | 3P0G |
| TM1 | 39 | 70 | 4U15 |
| ICL1 | 71 | 72 | None |
| ICL1 | 73 | 76 | 4U15 |
| TM2 | 77 | 107 | 4U15 |
| ECL1 | 108 | 108 | None |
| ECL1 | 109 | 109 | 4PXZ |
| ECL1 | 110 | 110 | None |
| TM3 | 111 | 113 | 4PXZ |
| TM3 | 114 | 149 | 4U15 |
| ICL2 | 150 | 150 | None |
| ICL2 | 151 | 155 | 5DYS |
| ICL2 | 156 | 156 | None |
| TM4 | 157 | 183 | 4U15 |
| ECL2 | 184 | 188 | None |
| ECL2 | 189 | 191 | 4U15 |
| ECL2 | 192 | 192 | None |
| ECL2 | 193 | 197 | 3VW7 |
| ECL2 | 198 | 198 | None |
| TM5 | 199 | 199 | 4XT1 |
| TM5 | 200 | 209 | 4U15 |
| TM5 | 210 | 211 | 3V2Y |
| TM5 | 212 | 231 | 4U15 |
| TM5 | 232 | 240 | 3SN6 |
| ICL3 | 241 | 245 | None |
| ICL3 | 305 | 309 | None |
| TM6 | 310 | 313 | 5TVN |
| TM6 | 314 | 347 | 4U15 |
| ECL3 | 348 | 348 | None |
| ECL3 | 349 | 352 | 2YDV |
| ECL3 | 353 | 353 | None |
| TM7 | 354 | 379 | 4U15 |
| H8 | 380 | 391 | 4U15 |
| H8 | 392 | 392 | 3SN6 |

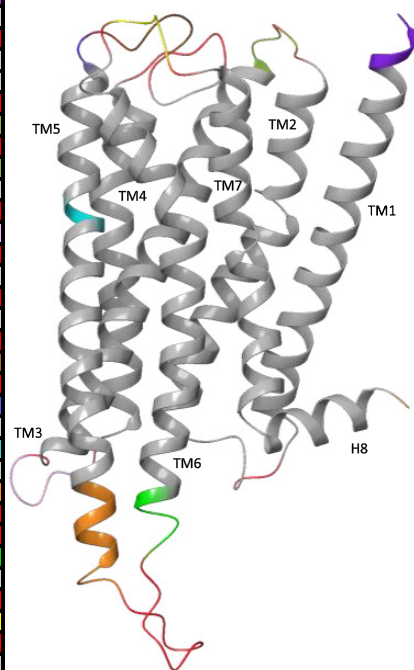


Figure 3. Example model of the inactive state GPR75. The colors indicate the use of 10 different backbone template to fill in lacking coordinates, elongate helices, add missing secondary structure segments to loops and to remove a main template helix bulge not shared by GPR75. Furthermore, the use of 27 sidechain templates from the GPCR position-specific rotamer library increased the percentage of residues that could be based on an identical amino acid from 15.7% to 58.2%.

homology modeling pipeline, which gives all researchers access to ready high-quality models without the need for own configuration.

Example model

Figure 3 shows an example model, the inactive state GPR75, which was based on 10 backbone and 27 sidechain templates. Alternative backbone templates were used to extend helices that were too short in the main template. These were TM1 start (3P0G), TM3 start (4PXZ), TM5 start (4XT1), TM5 end (3SN6), TM6 start (5TVN) and H8 end (3SN6). This feature is especially useful in the TM5-ICL3-TM6 region, where most GPCR crystal structures have non-native conformations due to fusion proteins. Alternative templates were also used for loops that differed in length or had missing coordinates. These were ECL1 (4PXZ), ICL2 (5DYS), ECL2 (3VW7) and ECL3 (2YDV). In TM5, an alternative template was used to remodel a main template bulge not present in the GPR75 alignment, as this does not share the associated sequence motif. Finally, by applying our GPCR position-specific rotamer library, the per-

centage of residues that could be modeled from an identical amino acid increased from 15.7% to 58.2%.

The GPCRdb ligand database and statistics

This first instalment of the GPCRdb ligand database contains 144 773 ligands, 597 targets, 29,530 assay experiments amounting to ~320 000 activity records. A dedicated ligand and statistics page has been added to provide continuously updated distributions of ligands across the GPCR classes. Overall statistics is provided in a table listing the numbers of: ligands, average ligands per receptor and receptors that have a ligand. Specific receptors and receptor families are shown in trees, as also illustrated in Figure 4. The vast majority of ligands (91%) are found for class A GPCRs with great abundance (>1000 ligands) for aminergic, adenosine, opioid receptors, but no or very few ligands for the many understudied orphan receptors.

Ligand Browser

The ligand browser spans four separate pages for target selection, ligand filtering, ligand source data and commercial availability; all of which allow for filtering directly in the column headers. The first page allows for selection of targets on the level of receptors, receptor families or classes, while displaying the numbers of associated ligands. The second page shows the min, max and mean activities for ligand-receptor-species triads across all their experiments; hence provide a non-redundant overview suitable for filtering of ligands by their activities, commercial availability and chemical properties, such as logP. The negative logarithmic (p) values are used to compare the large span of affinities/activities, which are also provided in nanomolar units. A third page allows users to retrieve the exact ligand activities in all separate assay experiments. The fourth and final page adds vendor names and identifiers to facilitate ligand purchases. The three ligand-containing pages have the option to download ligands smiles (with ChEMBL identifies) and activity spreadsheets (csv format) to facilitate further processing.

CONCLUSIONS AND OUTLOOK

In the last two years, GPCRdb has enabled and disseminated high-impact biological analyses (1,19,20), through resources that also allow scientists to analyse the data themselves. Herein, we have reported homology models of unprecedented quality for all human non-olfactory GPCRs and across activity states. The models will stay up to date, as the integrated pipeline incorporates new templates upon database re-building. The structural completeness and precise rotamers of the GPCRdb homology models make them uniquely suited for molecular dynamics and docking, respectively. The GPCRdb ligand database allows users to find relevant tool compounds and drugs faster and to know where to buy them, thereby facilitating experimental studies across many GPCR disciplines. Going forward it would be very valuable to expand from ChEMBL to additional ligand sources and to implement a ligand selectivity filter. Furthermore, the structural revolution within the GPCR

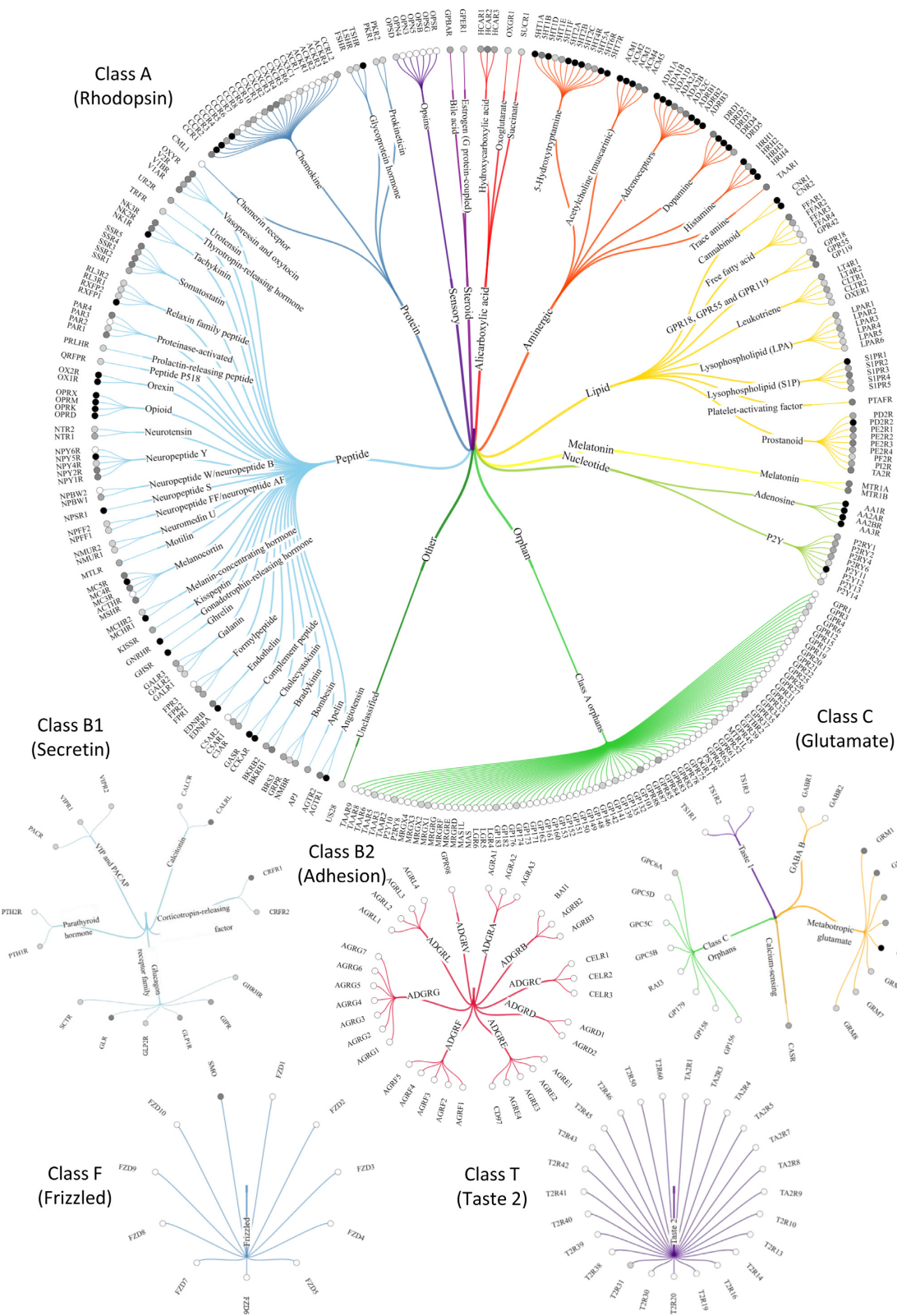


Figure 4. Trees showing the human GPCR classes (GRAFS family): A (Rhodopsin), B1 (Secretin), B2 (Adhesion), C (Glutamate), F (Frizzled) and T: Taste 2. Each tree is sorted alphabetically by ligand type (only class A) and receptor families which share the same physiological ligand. The tree branches are color-coded by ligand types, and the gray-scale circles before receptor names indicate their number of ligands (white: 0, light gray: >100, gray: >500 and black: >1000).

field should now allow for analysis of the successful constructs and data-driven design toward more structures. Finally, GPCR signalling is a hot topic warranting additional resources, such as β -arrestin data, GPCR-G protein complex models and a database of published biased agonists.

AVAILABILITY

GPCRdb is available at <http://www.gpcrdb.org> and can also be accessed via REST web services, as exemplified at http://docs.gpcrdb.org/web_services.html. The open source code and a virtual machine is available at <https://github.com/protwis/protwis>.

ACKNOWLEDGEMENTS

The COST Action CM1207 ‘GLISTEN’ has generously hosted the biannual GPCRdb meetings and supported multiple research visits. The GPCRmd consortium (gpcrmd.org) and Jens Carlsson are acknowledged for their homology model feedback.

FUNDING

European Research Council [639125]; Lundbeck Foundation [R163–2013-16327]; and Polish National Science Centre [2015/18/M/NZ2/00423]. Funding for open access charge: European Research Council [639125].

Conflict of interest statement. None declared.

REFERENCES

- Hauser,A.S., Attwood,M.M., Rask-Andersen,M., Schiöth,H.B. and Gloriam,D.E. (2017) Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.*, doi:10.1038/nrd.2017.178.
- Fredriksson,R., Lagerström,M.C., Lundin,L.-G. and Schiöth,H.B. (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.*, **63**, 1256–1272.
- Kolakowski,L.F. Jr (1994) GPCRdb: a G-protein-coupled receptor database. *Receptors Channels*, **2**, 1–7.
- Munk,C., Isberg,V., Mordalski,S., Harpsoe,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.
- Isberg,V., Mordalski,S., Munk,C., Rataj,K., Harpsoe,K., Hauser,A.S., Vroiling,B., Bojarski,A.J., Vriend,G. and Gloriam,D.E. (2016) GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res.*, **44**, D356–D364.
- Isberg,V., Vroiling,B., van der Kant,R., Li,K., Vriend,G. and Gloriam,D. (2014) GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res.*, **42**, D422–D425.
- Tautermann,C.S. and Gloriam,D.E. (2016) Editorial overview: New technologies: GPCR drug design and function-exploiting the current (of) structures. *Curr. Opin. Pharmacol.*, **30**, vii–x.
- Kufareva,I., Katritch,V., Participants of,G.D., Stevens,R.C. and Abagyan,R. (2014) Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. *Structure*, **22**, 1120–1139.
- Okuno,Y., Tamon,A., Yabuuchi,H., Nijima,S., Minowa,Y., Tonomura,K., Kunimoto,R. and Feng,C. (2008) GLIDA: GPCR–ligand database for chemical genomics drug discovery–database and tools update. *Nucleic Acids Res.*, **36**, D907–D912.
- Harding,S., Sharman,J.L., Faccenda,E., Southan,C., Pawson,A.J., Ireland,S., Gray,A.D.J., Bruce,L., Alexander,S.P., Anderton,S. *et al.* (2017) The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: Updates and expansion to encompass the new Guide to IMMUNOPHARMACOLOGY. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1121.
- Bento,A.P., Gaulton,A., Hersey,A., Bellis,L.J., Chambers,J., Davies,M., Krüger,F.A., Light,Y., Mak,L., McGlinchey,S. *et al.* (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.*, **42**, D1083–D1090.
- Roth,B.L., Kroeze,W.K., Patel,S. and Lopez,E. (2000) The Multiplicity of Serotonin Receptors: Uselessly diverse molecules or an embarrassment of riches? *The Neuroscientist*, **6**, 252–262.
- Cock,P.J.A., 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.
- Webb,B. and Sali,A. (2016) Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Protein Sci.*, **86**, 2.9.1–2.9.37.
- Rose,A.S. and Hildebrand,P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.*, **43**, W576–W579.
- 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.
- Fidom,K., Isberg,V., Hauser,A.S., Mordalski,S., Lehto,T., Bojarski,A.J. and Gloriam,D.E. (2015) A new crystal structure fragment-based pharmacophore method for G protein-coupled receptors. *Methods*, **71**, 104–112.
- Kim,S., Thiessen,P.A., Bolton,E.E., Chen,J., Fu,G., Gindulyte,A., Han,L., He,J., He,S., Shoemaker,B.A. *et al.* (2016) PubChem Substance and Compound databases. *Nucleic Acids Res.*, **44**, D1202–D1213.
- 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.
- Hauser,A.S., Chavali,S., Masuho,I., Jahn,L.J., Martemyanov,K.A., Gloriam,D.E. and Babu,M.M. (2017) Pharmacogenomics of GPCR drug targets. *Cell*, in press.