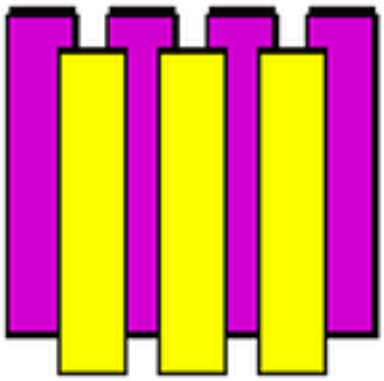

GPCRdb Documentation

Release 1

Vignir Isberg **Christian Munk**
Gloriam Group Members

Sep 22, 2023

User documentation

The GPCRdb contains data, diagrams and web tools for G protein-coupled receptors (GPCRs). Users can browse all GPCR crystal structures and the largest collection of receptor mutants. Diagrams can be produced and downloaded to illustrate receptor residues (snake-plot and helix box diagrams) and relationships (phylogenetic trees). Reference (crystal) structure-based sequence alignments take into account helix bulges and constrictions, display statistics of amino acid conservation and have been assigned generic residue numbering for equivalent residues in different receptors.

The source code and source data are freely available on GitHub.

Below, a table overview of all the different pages and functionalities in GPCRdb grouped by sections, along links to specific documentation pages, associated slides and video demonstrations.

Table 1: **SEQUENCES**

Page name	Video	Slides	Demo	Reference	Short description
Sequence alignments	•	•		Isberg et al. NAR 2016 Kooistra et al. NAR 2021	Provides sequence alignment analyses of receptors.
Generic residue number tables	•	•	•	Isberg et al. TiPS 2015	Displays the generic residue number tables for the single or set of receptors selected.
Genetic variants	•	•	•	Hauser et al. Cell 2018	Section showing the variation coverage, single receptor variants and the estimated economic burden for drugs targeting GPCRs.
Isoforms	•	•	•	Marti-Solano et al. Nature 2020	Info page highlighting the number of unique isoforms detected for each receptor gene.

Table 2: **STRUCTURES**

Page name	Video	Slides	Demo	Reference	Short description
Structure coverage	•	•		Isberg et al. NAR 2016	Overview of the structure information available in GPCRdb. Data is shown using detailed inforgraphics and plots.
Structures	•	•		Isberg et al. NAR 2014 Kooistra et al. NAR 2021	Browser page listing detailed information for all the structures present in GPCRdb.
Structure models	•	•		Pándy-Szekeres et al. NAR 2023	Browser page listing all the AlphaFold-Multistate generated homology models and refined structures present in GPCRdb.
Structure model validation	•	•		Kooistra et al. NAR 2021 Pándy-Szekeres et al. NAR 2023	Browser page listing the RMSD values comparing the latest model before a structure of the same receptor in the same state was published.

Table 3: **STRUCTURE ANALYSIS**

Page name	Video	Slides	Demo	Reference	Short description
Structure comparison tool	•	•		Kooistra et al. Nat. Struct. Mol. Biol 2021	Interactive tool for investigation of single structures or comparison of sets of structures. The tool allows for more than 20 different plot representations and lists available contact position pairs.
Structure similarity trees	•	•		Kooistra et al. Nat. Struct. Mol. Biol 2021	Interactive tool showcasing the similarity trees of selected set of receptors. Generated trees can be further graphically tweaked and downloaded in high resolution format.
Structure superposition	•	•	•	Isberg et al. NAR 2016	Functionality designed for the superpositioning of user defined structures (reference and test), allowing segment based selection.
Generic residue numbering (PDB)	•	•	•	Isberg et al. NAR 2016 Isberg et al. TiPS 2015	GPCRdb generic number annotation of user uploaded PDB files where numbers get deposited in the B-factor column.

Table 4: **STRUCTURE CONSTRUCTS**

Page name	Video	Slides	Demo	Reference	Short description
Construct/Experiment design	•	•	•	Munk et al. Nat. Methods 2019	Section for the design and alignment of experimental constructs and associated browser page listing experiments walkthrough.
Truncation/Fusion analysis	•	•	•	Munk et al. Nat. Methods 2019	Section with pages showing information about truncation sites fusion sites and deletion loops across the available GPCRdb data.
Mutation analysis	•	•	•	Munk et al. Nat. Methods 2019	Section with mutations oriented browsers, listing stabilizing mutations, substitution matrix and a dedicated stabilising mutation analyser.

Table 5: **DETERMINANTS & MUTATIONS**

Page name	Video	Slides	Demo	Reference	Short description
Sequence signature tool	•	•	•	Kooistra et al. NAR 2021	Interactive tool for the investigation of sequence signature across two different user defined set of receptors.
State stabilizing mutation design	•	•	•	Kooistra et al. Nat. Struct. Mol. Biol 2021	Browser page listing detailed state stabilising mutation information for each receptor present in GPCRdb.

Table 6: **LIGANDS & BIOACTIVITIES**

Page name	Video	Slides	Demo	Reference	Short description
Ligand coverage	•	•		Pándy-Szekeres et al. NAR 2018	Overview of the ligands information available in GPCRdb. Data is shown using detailed inforgraphics and plots.
Ligands	•	•		Pándy-Szekeres et al. NAR 2018 Pándy-Szekeres et al. NAR 2023	Section allowing the search for ligands associated to a specific receptor or using ligand name, database ID or chemical information (SMILES or Inchikey).

Table 7: **ENDOGENOUS LIGANDS**

Page name	Video	Slides	Demo	Reference	Short description
Endogenous ligands	•	•		Pándy-Szekeres et al. NAR 2023	Browser page listing detailed information for endogenous ligands present in GPCRdb. Data are derived and maintained by Guide to Pharmacology.

Table 8: **DRUGS & AGENTS IN TRIAL**

Page name	Video	Slides	Demo	Reference	Short description
Drugs targets and indicators	•	•	•	Hauser et al. Nat. Rev. Drug Discov 2017	Browser page listing information about drugs, their targets of action, therapeutic indication and clinical status with associated references.
Drug target tree	•	•	•	Hauser et al. Nat. Rev. Drug Discov 2017	Interactive infographic tree showing drug data information on established or under clinical trials GPCR targets.
Drug statistics	•	•	•	Hauser et al. Nat. Rev. Drug Discov 2017	Overview of the GPCR class as drug target, with focus on receptor family targets, drug molecule types, mode of action and disease indications.

Table 9: **LIGAND SITES**

Page name	Video	Slides	Demo	Reference	Short description
GPCR-ligand interactions	•	•	•	Isberg et al. NAR 2016 Munk et al. COPHAR 2016	Functionality showing the protein-ligand interactions for the available PDB structures in GPCRdb using graphical representation and 3D visualization tools.
Site search	•	•	•	Isberg et al. NAR 2016	Section designed for the investigation of protein-ligand interactions of set of receptors, either using sequence motifs or ligand complex through a user defined PDB file.

Table 10: **LIGAND SITE MUTATIONS**

Page name	Video	Slides	Demo	Reference	Short description
Mutation coverage	•	•	•	Munk et al. COPHAR 2016	Overview of the mutations information available in GPCRdb. Data is shown using detailed inforgraphics.
Mutations	•	•		Munk et al. COPHAR 2016	Browser page listing mutations for user selected receptors and protein segments. Result table also provides information about ligand affinity for the mutation (FoldChange) if available.
Mutation design tool	•	•	•	Munk et al. COPHAR 2016	Section with mutation design tool based either on a PDB code or receptor name. Result page lists suggested mutations with associated structure interactions, supporting ligands, receptors and mutagenesis experiments.
Mutation data submission	•	•	•	Munk et al. COPHAR 2016	Description to the mutation data submission form and guidelines on how to submit your experimental data to GPCRdb.

The documentation is organised into three sections:

- *User documentation*
- *Developer documentation*

- *About GPCRdb*

CHAPTER 1

Receptors and families

The selection page allows users to find a receptor and family by searching or browsing.

The search box displays a list of both families and receptors that match the input keyword. Selecting either a receptor or family will take you to the corresponding receptor/family page.

The browser displays a hierarchical view of the families, and the proteins in each family. Selecting either a receptor or family will take you to the corresponding receptor/family page.

1.1 Receptor pages

The page displays basic information about the selected protein and a sequence viewer, as well as helix box and snake diagrams. The diagrams can be colored by properties, mutant information, or ligand interactions extracted from structures.

1.2 Family pages

The family pages resemble the protein pages, but the sequence shown on a family page is a consensus sequence for the human sequences in the family.

CHAPTER 2

Ligands & Bioactivities

2.1 Ligand coverage

The ligand coverage page is designed to showcase the ligand data present in GPCRdb through a fast visualization tree able to convey the message in a timely fashion. In this page you will find:

1. A table at the top, listing the number of unique ligands per receptor class, divided by each class (family). Moreover, the table lists also the average number of ligands per GPCR in each class and the total count. Lastly, the table lists the number of GPCRs with ligands for each class present in GPCRdb.
2. Visualization phylogenetic trees for each GPCR class, where each branch points to a different ligand family, ending with an inner circle shaded by the amount of ligand data available for that specific GPCR receptor. Coloring of that inner circle is in grayscale, whereas light shades of gray may vary between 10 to 500 ligands, and darker shades of grey up to black add up to more than 1000 compounds targeting that receptor.

Phylogenetic trees be default will show the UniProt name of the different GPCRs, but through a dedicated selection button users can decide to show the IUPHAR names of those GPCRs. Trees can also be downloaded in several different formats (SVG, JPG, TIFF and PNG), allowing for high resolution figures for publication purposes.

2.2 Ligands (ChEMBL, GtP, Ki db)

The ligand query section provides two different alternatives for investigating the ligands available in GPCRdb:

1. Querying by receptor
2. Querying by ligand name, ligand ID, SMILES or Inchikey.

2.2.1 Receptor query

In the receptor based query you will be prompted to select a receptor in the selection table, showing the number of ligand for each receptor, divided also for drugs that have been approved and those in clinical trials. After selecting the receptor target, users can choose between two ligand bioactivity browsers. The ‘Compact (1 row/ligand)’ browser

collates all binding or functional bioactivities of a given ligand and source database on one row by calculating minimum, average and maximum activity across studies. The 'Extended (1 row/activity)' browser instead lists the specific binding affinity or potency value of each study. The latest GPCRdb release has restructured potency and affinity data into separate tabs. It has also added fold selectivity values, along with the underlying number of experiments, allowing ligand selection based on their selectivity for the target of interest relative to all other stored GPCR targets. Both browsers also present information about vendors from which one can purchase the given ligand along with key physicochemical descriptors

2.2.2 Ligand query

In the ligand query page you will be able to input either the ligand name, ligand ID (from either database of origin, being it GPCRdb itself, Guide to Pharmacology, ChEMBL or the PDSP KiDatabase), or structure information (being it SMILES or Inchikey) in the search textbox equipped with autofill. A dropdown menu will list the available ligands adhering to the provided input. By clicking on the selected ligand name users will be redirected to that specific ligand info page.

2.3 Ligand Info page

The ligand info page can be accessed directly by a ligand query or from the results page of the above ligand browser. The top of the ligand info page displays information about ligand structure (2D image, SMILES, InChI key), names (common, and chemical names, and aliases), physicochemical properties (molecular weight, logP and counts for hydrogen bond acceptors/donors, and rotatable bonds), molecule type (small molecule/peptide/protein, drug status and endogenous/surrogate), and database links (internal and external). For ligands that have this information, two additional boxes provide GPCR-ligand crystal/cryo-EM structure complexes and mutations affecting ligand activity. The bottom of the page shows bioactivities for the given ligand across receptor targets. The bioactivity browser allows filtering by the receptor classification, bioactivity, and source database. The information on endogenous ligands and target FDA approval status have been derived from Guide to Pharmacology and DrugBank databases, respectively.

2.4 Endogenous ligand browser

The endogenous ligand-GPCR system spans different relationships ranging (ligand:receptor) 1:1, 1:many, many:1 and many:many relationships. To facilitate browsing across either ligands or receptors, we developed an endogenous ligand browser. This browser contains data for 543 distinct endogenous ligands for 253 human GPCRs, and 157 mouse, rat, or guinea pig receptors. For each receptor, alternative endogenous ligands are classified as principal or secondary, as defined by the nomenclature committee of the International Union of Basic and Clinical Pharmacology and have an additional ranking by potency. For ligand-receptor pairs with multiple potency (pEC50) or affinity (pKi) values, the browser provides minimum, mean and maximum values, with grayscale background aiding comparison. Finally, the browser contains information about the ligand type, ligand name with a direct link to the ligand info page in GPCRdb, receptor information as family, species, IUPHAR and UniProt name and a popup showing the original references for bioactivities.

CHAPTER 3

Signal proteins

3.1 Signal protein page

The selection page allows users to find a signal protein or family (grouped in the 4 main G protein families) by searching or browsing.

The browser displays a hierarchical view of the families, and the proteins in each family. Selecting either a receptor or family will take you to the corresponding receptor/family page. Species orthologs can be selected when toggling the Species button to 'All'

The page displays basic information about the selected protein and a sequence viewer a snake-like diagrams. The diagrams can be colored by properties, receptor interface and barcode information.

3.2 GPCR-G protein coupling

The page shows statistics on known coupling preferences as extracted from Guide to PHARMACOLOGY as:

1. an interactive Venn diagram, which highlights the number of reported receptors for each G protein coupling combination
2. an interactive phylogenetic tree, for which concentric circles illustrate the G protein-coupling selectivity of each GPCR the four dots depict both primary and secondary G protein coupling (from inside to outside: Gαs, Gαi/o, Gαq/11, Gα12/13). Tree nodes can be highlighted and selected to retrieve clade-specific receptor sets, which can be used in dedicated segment specific sequence alignments.

Highlighting and selection of receptors populates a field, which can be used as an input for dedicated segment specific sequence alignments.

3.3 G protein alignments

The "Structure-based alignments" tool allows for alignment of user selected G proteins and sequence segments. Using the tool is a two step process.

1. The user is first presented with a G protein selection page.
2. The user is presented with a sequence segment selection page. The user can select one or more sequence segments, and/or expand each segment to select the residues within it individually.

After completing these two steps, an alignment is displayed. To display the sequence number of an aligned residue, as well as generic numbers (CGN numbering), hover the mouse over it. At the bottom of the page, a consensus sequence as well as conservation statistics for amino acids and chemical features are displayed.

3.4 Interface mapping

Maps the β2-Gαs complex (PDB: 3SN6) interaction interface onto a snake plot of a selected receptor and highlights conserved and accessible interactions.

- Flock, T., Hauser, A. S., Lund, N., Gloriam, D. E., Balaji, S., & Babu, M. M., "Selectivity determinants of GPCR–G-protein binding.", **2017**, *Nature*, May 18;545(7654):317-322 10.1038/nature22070

CHAPTER 4

Sequences

4.1 Structure-based alignments

The "Structure-based alignments" tool allows for alignment of user selected receptors and sequence segments. Using the tool is a two step process.

1. The user is first presented with a receptor selection page. Receptors can be selected individually or by family. The user can select as many receptors as he/she wishes (WARNING: selecting a large number of receptors increases loading time).
2. After receptors have been selected, the user is presented with a sequence segment selection page. The user can select one or more sequence segments, and/or expand each segment to select the residues within it individually. Residues selected individually are grouped into a custom sequence segment.

After completing these two steps, an alignment is displayed. To display the sequence number of an aligned residue, as well as generic numbers, hover the mouse over it. At the bottom of the page, a consensus sequence as well as conservation statistics for amino acids and chemical features are displayed.

4.2 Phylogenetic trees

The phylogenetic tree tool allows for generation of phylogenetic trees based on user selected receptors and sequence segments. Using the tool is a three step process.

1. The user is first presented with a receptor selection page. Receptors can be selected individually or by family. The user can select as many receptors as he/she wishes (WARNING: selecting a large number of receptors increases loading time).
2. After receptors have been selected, the user is presented with a sequence segment selection page. The user can select one or more sequence segments, and/or expand each segment to select the residues within it individually. Residues selected individually are grouped into a custom sequence segment.
3. In the third step, a settings page is displayed. The amount of bootstrapping replicas (0, 10 or 100) and the type of tree (rectangular or circular) are configurable by the user. User are also offered an option to show branch

lengths that represent the evolutionary distance between the nodes, or show the same branch length between every node.

To view an alignment of the sequences used to generate the tree after it has been displayed, click the "View alignment" button.

The trees are generated using PHYLIP and jsPhyloSVG.

4.3 Similarity search - GPCRdb

The GPCRdb similarity search tools allows a user to find the most similar receptors for a reference sequence, out of all GPCRs, or a subset selected by the user. The tools is more accurate than BLAST search, since it uses curated, structure-based alignments, but only works on sequences that are already in the database. Using the tool is a three step process.

1. The user is first presented with a reference receptor selection page.
2. Once a reference receptor has been selected, the user is presented with a sequence segment selection page. The user can select one or more sequence segments, and/or expand each segment to select the residues within it individually. Residues selected individually are grouped into a custom sequence segment.
3. The third step is selecting a comparison receptor set. The selected receptors will be compared to the reference receptor based on the selected sequence segments, and their similarities computed. The user can select as many receptors as he/she wishes (WARNING: selecting a large number of receptors increases loading time).

After completing these three steps, an alignment is displayed, with the receptors in the comparison set ranked by similarity to the reference receptor. The three columns to the right of the receptor ID show three computed properties:

- **Sequence identity (%I)**: The percentage of identical amino acids.
- **Sequence similarity (%S)**: The percentage of similar amino acids (where similar is defined as BLOSUM62 score > 0).
- **Similarity score (S)**: The sum of every position's BLOSUM62 score.

To display the sequence number of an aligned residue, as well as generic number indices, hover the mouse over it.

4.4 Similarity search - BLAST

The BLAST based similarity search is an alternative to the GPCRdb similarity search that works for any user submitted sequence (the query sequence does not have to be in GPCRdb already). The runs a standard BLAST search on a custom BLAST database that contains every sequence from GPCRdb.

The results page show a list of the best BLAST hits for the submitted query sequence.

4.5 Similarity matrix

The similarity matrix tool allows a user to quickly gain an overview of the sequence identity and similarity between all sequences in a receptor family, or a custom selected group of receptors. Using the tool is a two step process.

1. The user is first presented with a receptor selection page. Receptors can be selected individually or by family. The user can select as many receptors as he/she wishes (WARNING: selecting a large number of receptors increases loading time).

2. After receptors have been selected, the user is presented with a sequence segment selection page. The user can select one or more sequence segments, and/or expand each segment to select the residues within it individually. Residues selected individually are grouped into a custom sequence segment.

The results are shown as a table of identities and similarities, color-coded in a red-yellow-green color scale ranging from low to high identity/similarity. Identities are shown in the lower-left half of the table, and similarities in the upper-right half.

CHAPTER 5

Sequence signature tool

A sequence signature is a set of residue positions with amino acids or properties thereof that are distinctly conserved in one of two sets of proteins; and is generated to identify determinants of a biological function differing between these two sets.

The GPCRdb sequence signature tool (available under Receptors – Sequences – Similarity) receives a selection of two sets of GPCRs. These are aligned separately to calculate their individual percent conservation of amino acids or properties thereof within each column of the alignment which is indexed with a generic residue number. For each such residue position, the tool then identifies most distinctly conserved amino acid or property in either dataset and assigns a score representing the difference in percent conservation. For example, a conservation of 87 and 27 percent in the two receptor sets gives a score of 60. Finally, the sequence signature is derived by applying a score cut-off (default is 40) to extract the most distinct residue positions. This sequence signature can be downloaded as an Excel spreadsheet for further analysis, for example design of mutagenesis experiments based on signature scores and amino acids/properties. Furthermore, the sequence signature can be matched against the protein sequences of other receptors to predict which of those share the reflected function (as defined by the two original sets of proteins). This provides a sum of signature position scores that is normalised by dividing by the maximum theoretical score (if the GPCR and signature matches in all positions). All receptors are listed in decreasing score order and can be downloaded as an Excel spreadsheet for example to select GPCRs for validating functional characterization.

CHAPTER 6

Structures

6.1 Structure browser

The structure table shows an annotated list of published GPCR structures. The table can be sorted by each column by clicking on the header. The search fields below each header can be used to filter the structures, e.g. show only those with a co-crystallized agonist or X-ray resolution < 2.5 Å.

To view an alignment of the selected structure sequences, click the "Align" button. Downloading multiple structures at the same time is available with the "Download" button. The "Show representative" button will filter the table by only showing structures that are the representative structures for the given receptor in each state. These structures are the ones which have the most GPCRdb generic numbered positions present in the structure and have high resolution.

6.2 Refined structures

GPCRdb provides regularly updated refined structures where missing segments are modeled in and mutations are reverted back to wild type. For structures where there is no G protein present, missing coordinates are modelled in from our activation state specific AlphaFold-Multistate models (Pándy-Szekeres et al. 2022). For structures where a GPCR is in complex with a G protein, the automated chimeric GPCRdb homology modeling pipeline is used (Pándy-Szekeres et al. 2018). This entails modeling in missing segments (helix ends, loops, H8), reverting mutations to wild type and remodeling distorted regions based on our in-house manual structure curation with the use of alternative structural templates. All refined structures are available on the Structures page (gpcrdb.org/structure), just GPCR refined structures are also on the Structure models pages (gpcrdb.org/structure/homology_models) and just GPCR-G protein complex refined structures are also on the GProteinDb Structure models page (gproteindb.org/structure/complex_models).

6.3 Structure statistics

The statistics page shows a bar graph with the number of structures available by year (and grouped by the endogenous ligand type of the receptors), a bar graph showing the resolution ranges of the available structures, and phylogenetic trees for each receptor class, with receptors with determined structures highlighted.

The graphs are automatically updated when new data is added to GPCRdb, making them ideal for use in publications and presentations.

6.4 Structure models

With every database update GPCRdb builds a homology model repository containing models for every human GPCR in three different activation states (inactive, intermediate, active). Class T is based on class A and class B2 is based on class B1. The models are created based on the GPCRdb homology modeling pipeline (Pándy-Szekeres et al. 2018) that utilizes an automated chimeric modeling approach. For every model a single main template is selected and atomic coordinates from alternative local templates are swapped-in for sections of the model where either the main template is missing coordinates or the algorithm predicts a better template based on multiple criteria. These affect helix ends, loops, H8 and structural anomalies like bulges and constrictions. Furthermore, an in-house rotamer library is applied for side-chains where there is a mismatch between the modeled receptor and the template. The newest version of the homology models can be found on the Structure models page (gpcrdb.org/structure/homology_models) where users can download the older versions as well from the Archive.

6.5 Structure model validation

To assess GPCRdb homology models we provide root-mean-square deviation (RMSD) calculations between the first published experimental structure of a receptor and the latest GPCRdb model before that publish date. The model is in the same activation state as the experimental structure. After a structure is released, manually annotated structural data is added to GPCRdb with the next release. In some cases, there is a delay between the manual annotation and the release of the structure; therefore, some model versions can have a later date than the publication date of the modeled structure; however, these do not contain any information from the modeled structure.

For such structural comparisons the superpositioning method is key. The GPCRdb RMSD calculation workflow employs a sequence-dependent comparison where atoms that are missing from either the structure or the model are excluded and only the 7-transmembrane (7TM) backbone atoms (N, CA, C) determined by GPCRdb sequence alignments are used for the superpositioning. All RMSD calculations, which are also sequence-dependent, are carried out from this superpositioned state. For loop segment RMSD scores, this makes it possible to factor in not only the structural characteristics of the loop but also its position relative to the 7TM bundle. Furthermore, different properties of the models can be assessed based on which atoms we select for the RMSD calculations. The RMSD calculations themselves were done with the following python code:

```
round(np.sqrt(sum(sum((array1[1:]-array2[1:])**2))/array1[1:].shape[0]),1)
```

where array1 and array2 are numpy arrays with atomic coordinates from structure and model respectively.

RMSD categories currently available for GPCRdb models:

- Overall all: all atoms
- Overall backbone: all backbone atoms
- 7TM all: all atoms of 7TM residues
- 7TM backbone: backbone atoms of 7TM residues
- H8: Helix 8 backbone atoms
- ICL1: Intracellular loop 1 backbone atoms
- ECL1: Extracellular loop 1 backbone atoms
- ICL2: Intracellular loop 2 backbone atoms

- ECL2: Extracellular loop 2 backbone atoms
- ECL3: Extracellular loop 3 backbone atoms

6.6 Structure descriptors

- **Activation state definition** The activation state of the structures are defined on the distances between all Cα atoms for residues on the intracellular half of each of the seven transmembrane helices. These distances are also used for the generation of the Structural similarity trees . For each class, structures in complex with a signaling protein are are set as the reference structures for the active state(100% degree activation). Subsequently, structures with a highly closed conformation are set as the reference structures for the inactive state (0% degree activation) based on a maximum distance between 2x46 to 6x37 for all classes except for class F for which the distance between 2x44 to 6x31 is used (maximum distances are 11.9Å, 13Å, 14.5Å, 13Å for classes A, B, C, and F, respectively). The Cα atom distance pairs for each structure are compared to the reference structures and the mean distance to the active structures and the mean distance to the inactive structures are then calculated. If a structure has a low distance to the inactive structures its state is defined as *inactive*, vice versa if a structure has a low distance to the active structures then its state is defined as *active*. However, if both are not the case then the structure is defined as *intermediate*. In some cases, when an unlikely conformation is encountered its state is defined as *other* as is now the case for the structure of the plate-activating factor receptor 5ZKP.

  The degree activation: These distances to the reference structure sets are then converted into an "activation score" by substracting the mean distance to the inactive-state structures from the mean distance to the active-state structures. The activation score is converted into a percentage activation based on the minimum and maximum activation scores for all structures in that class.

- **TM6 tilt** The TM6 tilt measure is defined based on the distance between the Ca atoms for the residues 2x46 and 6x37 for all classes except for class F for which the distance between 2x44 and 6x31 is used. For each structure, this distance (when the residues are present) is converted into a percentage by comparing it to the minimum and maximum distance observed in any other structure for that specific class.

CHAPTER 7

Structure comparison

7.1 Structure comparison tool

The Structure comparison tool allows users to perform powerful analyses of a structure, a set of structures, or even compare two (sets of) structures. When opening op the tool, there are three tabs displayed at the top. Depending on what analysis one wants to perform, one has to choose one of the three:

1. **Single structure** Utilize this tab, when you want to investigate the interactions and properties of a single structure. Click on the *Select Structure* button, select a structure in the list (by ticking the leftmost checkbox, by clicking on a row, or by importing the PDB code at the top) and close the structure selection panel. Then press the green *Go* button and the structure will be analyzed. Optionally, you can adjust the configuration of the tool by clicking on the cogwheel icon before pressing *Go*.

2. **Single set of structures** Utilize this tab, when you want to investigate variations of interactions, distances, and properties of a set of structures. Click on the *Select Structures* button, select the structures in the list (by ticking the leftmost checkbox, by clicking on a row, or by importing the PDB codes at the top) and close the structure selection panel. Then press the green *Go* button and the structure set will be analyzed. Optionally, you can adjust the configuration of the tool by clicking on the cogwheel icon before pressing *Go*.

3. **Two sets of structures** Utilize this tab, when you want to investigate the different or shared interactions, distances, and properties of two sets of structures. First click on the *Select (Set 1)* button, select the structures in the list (by ticking the leftmost checkbox, by clicking on a row, or by importing the PDB codes at the top) and close the structure selection panel. Subsequently, click on the *Select (Set 2)* button. Then press the green *Go* button and the selected structure sets will be analyzed. Optionally, you can adjust the configuration of the tool by clicking on the cogwheel icon before pressing *Go*.

After selecting the structures and pressing the *Go* button, the tool analyzes the structures and collects the data from the analysis, this might take a minute. When completed the gray overlay will disappear and the three plotting panels in the middle of the screen and the data browsers at the bottom are populated. The analyzed data itself is presented in interactive browsers at the bottom in four different tabs:

1. **Contact position pairs** All data relating to the residue-residue pairs that interact in the selected structure(s). This data ranges from interaction frequencies per interaction type to distances, angles, to sequence conservation. It should be noted that filtering action in this browser will impact several of the plotting options

that are discussed below. For example, when showing the 3D structure and then filtering on a column, only those interactions will be shown in the 3D structure or flare plot that are also still displayed in this browser, all other interactions will be filtered out.

2. **Contact position-AA pairs** This browser is an extension of the first browser. Instead of grouping all interaction frequencies for each residue pair, it first groups on unique amino acid pairs and then analyzes the interaction frequencies and sequence conservation for that specific pair of amino acids. This tab can, for example, be used to figure out if a conserved interaction in one structure set is due to the conservation of a specific amino acid pair that is not conserved in the other structure set.
3. **Residue backbone & sidechain movement** In contrast to the first two browsers, the last two browsers focus on each individual residue with a generic residue number. In this browser, multiple properties are analyzed for each individual residue, such as the rotamer angle, (relative) solvent-accessible surface area (SASA and RSA), sequence conservation, and distances. This browser can, for example, be used to identify movements in the backbone and sidechain of specific residues.
4. **Residue helix types, bulges, and constrictions** The last of the browsers also focuses on each individual residue with a generic residue number. In contrast to the previous browser, this one focuses mainly on the backbone conformation and secondary structure described at each of the residues using the secondary structure annotation, tau and theta angles, phi/psi/tau dihedrals, and is complemented again by sequence conservation. If backbone variations occur, this is the browser to identify them in, such as - for example - the constriction at the end of TM7 upon activation.

Finally, in the middle the data can be visualized in a plethora of different ways:

- **TM1-7 segment movement**
  - **Extracellular - Segment plot (2D and 3D)** A 2D or 3D depiction of the movement and rotation of the seven TM helices at the extracellular ends when comparing two different structure sets. The movements are based for each helix, on the most-extracellular three residues with a generic residue number that are present in all structures in both sets.
  - **Middle of membrane - Segment plot (2D and 3D)** A 2D or 3D depiction of the movement and rotation of the seven TM helices at the middle of the membrane (based on the OPM orientation of the GPCR in the membrane) when comparing two different structure sets.
  - **Cytosolic - Segment plot (2D and 3D)** A 2D or 3D depiction of the movement and rotation of the seven TM helices at the intracellular ends when comparing two different structure sets. The movements are based for each helix, on the most-intracellular three residues with a generic residue number that are present in all structures in both sets.
- **Contacts between generic residue positions**
  - **Flare plot** ‡ A GPCRdb-customized version of the flare plot that depicts interacting residues in a circular fashion. With the default enabled option to show interactions between residues from consecutive segments on the outside of the circle and all other interactions on the inside of the circle. For a single structure, the line colors depict the interaction type, for a single set of structures the frequency of interactions for a given residue-residue pair, and for two sets of structures the difference in interaction frequency between the two sets for a given residue-residue pair.
  - **Heatmap** ‡ A heatmap of all residues-against-all-residues depicting all interactions observed in the selected structure(s). For a single structure, the colors depict the interaction type, for a single set of structures the frequency of interactions for a given residue-residue pair, and for two sets of structures the difference in interaction frequency between the two sets for a given residue-residue pair.
  - **Network (2D and 3D)** ‡ These plots show in 2D or 3D how the interacting residues are interconnected in (sets of) connected subnetworks. When right clicking on a specific subnetwork, the plot zooms in on that specific network and by right clicking again it will go back to the full overview.

  - **Structure (3D)** ‡ This plot uses the NGL molecular viewer to show the structure(s) and interaction that are observed between the residue pairs. This viewer is again based on the data shown in first data browser (see above) and interactively changes whenever a filter is applied (or removed) in this data browser.

- **Contacts between segments (TM1-7, H8 & loops)**
  - **Flare plot (segments)** ‡ This option shows flare plots highlighting the number of shared interactions between the structural segments instead of residues. The structural segments are defined as the seven transmembrane helices, helix 8, and the loop regions.
  - **Network (2D and 3D)** ‡ This option, shows 2D or 3D networks highlighting the number of shared interactions between the structural segments instead of residues. The structural segments are defined as the seven transmembrane helices, helix 8 and the loop regions.
- **Contacts frequencies**
  - **Box plot** ‡ A plotting option that shows the interaction frequency differences when comparing two sets of structures.
- **Residue properties**
  - **Box plot (distribution)** ‡ A plotting option that shows the distribution for the different residue properties when comparing two sets of structures.
  - **Heatmap (distance)** This heatmap shows the overall distance (single structure or single set of structures) or change in distance for residue-residue pairs based on the Cα distance pairs.
  - **Scatter plot (correlation)** ‡ Scatter plotting option that allows for the selection of different types of residue properties and maps them against each other to find a correlation.
  - **Snakeplot (2D, topology)** ‡ A highly interactive and customizable version of the well-known snake plot that can be used to map any of the structural properties on to the snakeplot and
  - **Structure (3D, movement)** In the NGL viewer the overall normalized distance changes between two sets of structures are depicted based on the all-against-all Cα distance pairs. The red-white-blue color gradient used on the structure shows where overall an increase or decrease in placement (based on all distances to that specific residue) has been observed.

‡ These plots are based on the data shown in the first data browser (see above) and will therefore change when it is redrawn after a filter is applied or removed in this data browser.

For more information about the different properties that are used in this tool, please check out the preprint describing the Structure analysis platform. All measures used in the structure comparison tool are described in the methods and in "Extended data table 1".

7.2 Structure similarity trees

The Structure similarity trees is a tool that compares the overall conformation of a selected set of GPCR structures and generates a tree based on their shared conformational similarity.

Under the hood, this tool uses calculated distances from all Cα atoms to all other Cα atoms for all residues in the seven transmembrane helices. When a set of structures are being compared, first the average distance is calculated for each Cα atom pair and subsequently, this set of distances is filtered for all residues that are shared by at least 90% of all the structures. After this, for every structure, all the Cα distance pairs are normalized using the average distance for that specific Cα distance pair. After this, the similarity from each structure to each other structure is calculated using the summed absolute differences between all Cα distance pairs for those two structures. Finally, this sum of differences is normalized by the square of the distances divided by the square of the number of shared residues used in the comparison. Finally, the resulting all-against-all distance matrix is hierarchically clustered using average linkage

clustering and a Newick tree is created from the results. The tree is subsequently enriched with additional data about each structure and the receptor.

The online tool, the process is straightforward: after opening up the page click on the "Select Structures" button and select the structures to be compared in the selection panel that opened up (either by clicking on the row of a structure or by ticking the checkbox leftmost of the row). After closing the selection window, click on "Go" and the tree will be calculated. Using the dropdown menus, the tree can subsequently be manipulated by changing the tree type, the way the receptor/structure names are displayed and you can change the markers that are shown on the inside and outside of the structure names (leaves). Note that the internal nodes are colored with a white-to-black gradient (low-to-high) indicating the Silhoutte index , which is a measure of the separation of structures in this node to structures present in the nearest neighboring node. The quantitative value of this index is visible in a tooltip when hovering over the node.

In the top-right corner of the tree panel, three buttons can be found that allow to zoom in and out and recenter tree again. One level higher, in the top-right corner of the clustering panel, there are two icons, one for downloading the tree as a Figure, Newick tree, or the raw distance matrix, and one for making the panel full screen.

Finally, the tree is also interactive and can be used to make selections. When clicking on a node in the tree, the user gets the option to add the structures under that node to "selection set 1" or "selection set 2". Based on this selection, the user can directly start a structural analysis using the structure comparison tool to investigate the similarities and differences for the selected structures.

7.3 Structure superposition

The superposition tool allows users to select two or more GPCR structures (or models) and superpose them based on a user-specified segment selection. Using the tool is a two-step process.

1. **Select the structures to superpose. This can be done using the "structure browser" or by uploading the** PDB files. For the latter, only one reference structure can be uploaded, but multiple structures to be superposed on the reference can be uploaded. To select multiple structures for upload, hold down the Control key (or Command on Mac) while selecting.
2. **After structures have been uploaded or selected, the user is presented with a sequence segment selection page. The user** can select one or more sequence segments, and/or expand each segment to select the residues within it individually. Residues selected individually are grouped into a custom sequence segment.

7.4 Generic residue numbering (PDB)

The PDB file residue numbering tool adds generic residue numbers from GPCRdb to any GPCR structure or model. This can be useful when comparing structures visually.

A user simply uploads her structure and downloads a modified version of that structure, where b factors of certain atoms have been replaced with generic numbers. Note that CA atoms will be assigned a number in GPCRdb notation, and N atoms will be annotated with Ballesteros-Weinstein scheme.

On the structure download page, users can download scripts to visualize the generic numbers in PyMOL and Maestro.

CHAPTER 8

Structure Constructs

8.1 Data

Structure construct data was parsed from PDB files, which were fetched from RCSB PDB (www.rcsb.org). Deleted receptor regions were identified from the missing ranges in the DBREF lines. The PDB file protein sequence was matched to the wild-type sequence as a final validation of deletions and mutations. Residue annotations, including mutations, and additional protein sequences, such as linkers and fusion proteins, were imported from the SIFTS database. Thereafter we fetched data on structure experimental values (resolution, crystal growth, r-factor and method) from the PDBe REST API (https://www.ebi.ac.uk/pdbe/api/pdb/entry/experiment/pdb_code), which also aided the manual annotation effort. We manually annotated a comprehensive collection of GPCR structure experiment methods and reagents from structure publications into a Microsoft Excel file, which was imported to GPCRdb.

Furthermore, inserts not present in the final structure construct and therefore not available from the PDB and the effects of receptor mutations were also annotated and imported in the same way through separate spreadsheets. The completed integrated data can be retrieved at the GPCRdb repository (https://github.com/protwis/gpcrdb_data/tree/master/structure_data/construct_data).

8.2 Construct alignments

Construct alignments allow for quickly browse different constructs. There are three main views: 1. Browser – In a table format it allows for quickly sorting and filtering the dataset. Upon sorting/filtering it is possible to switch back to the schematic views with the subset of data. 2. WT (Wild-type) schematic – This view takes the perspective of the wild-type sequence so it becomes clear to see what parts of the proteins have been deleted and mutated. 3. Construct schematic (default view) – This view show the final construct schematic – here additional inserts are shown (Signal-tags, cleavage tags, fusion proteins).

Furthermore, there are filter options on all views where you can filter by several options to narrow the view to the structures of main interest.

If one wants to analyse the sequences themselves, it is possible to select several constructs and segments and with the "Sequence" button align these to get view shown below.

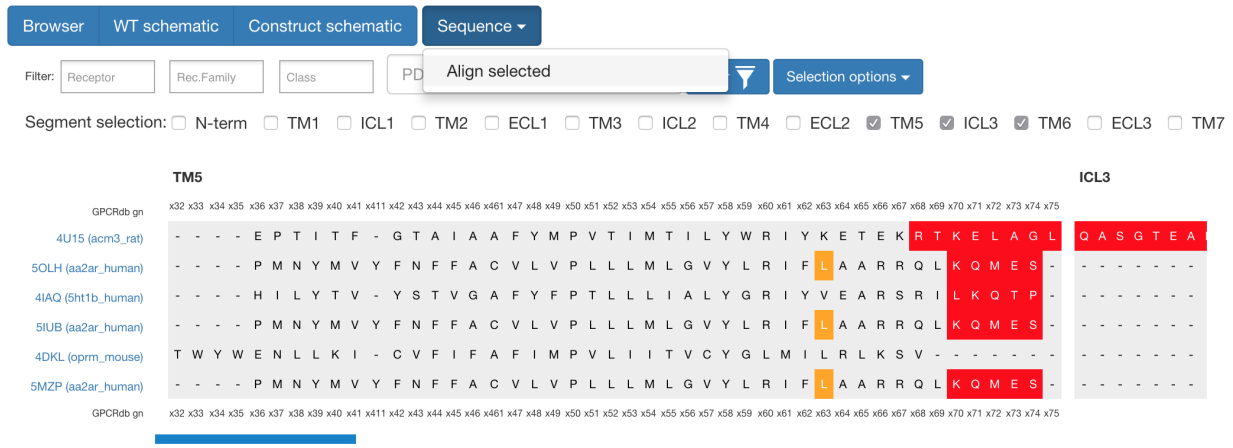

Fig. 1: **Figure 1**. What can be seen here is the alignment around TM5 / ICL3 segments to compare cut-sites among different constructs.

8.3 Construct design tool

When using the Construct design tool on a target receptor, the user first selects where a fusion protein should be inserted. This selection affects the tools design for the given target by only using as a basis the data that were obtained from templates with the same state and region of fusion. For all types of construct modifications the presented suggestions are ranked by target-template homology levels: same receptor, receptor family, or class or a different class (B2 to B1 and Taste 2 receptors to A). Whereas the homology is the default sorting for all types of construct modifications, for targets without a clear closest template users also have the option to instead sort by the frequency that a given modification has been observed among all unique receptors. N- and C-terminal truncation: To allow for cross-receptor comparisons, the lengths of N- and C-termini are defined as the number of residues that have been preserved before the start of the first transmembrane helix (TM1) and after the end of helix 8 (H8), respectively. Data from N-terminal protein fusions, with e.g. T4L or BRIL, are excluded unless the user has selected a fusion in this region. Fusion protein sites: Reflecting the structures obtained hitherto, loop fusions are placed in the second and third intracellular loops for the classes B and C, and A and F, respectively using the GPCRdb generic residue numbering scheme. If the user did not select a loop fusion, long ICL3 loops (>8 residues) are instead assigned suggestions of deletion sites from non-fused and N-terminally fused constructs.

The suggestions of stabilising mutations span a number of specific design rules (https://github.com/protwis/gpcrdb_data/raw/master/structure_data/Mutation_Rules.xlsx) which are both data- and rationale-driven and cover five overall concepts:

1. Homology: This concept infers a mutant position and amino acid if the target is the same receptor or a member of the same receptor family. For the classes B-F, which are smaller and have less data than class A, mutations will also be informed from any member of the same class.
2. Common mutations: These are mutations that have been utilised within several distinct receptor families, but not yet that of the selected receptor target.
3. Conservation: This concept introduces residues that are missing in the target but at least 70% conserved in the receptor family or class. For the positively charged residues H, K and R a lower, 40% conservation is used in order to incorporate multiple position at the tips of the transmembrane helices wherefrom these residues can interact with the polar head groups of the cell membrane. Furthermore, the low-propensity residues G and P are treated separately (below) and cysteine, which can form disulphide bridges, is excluded.
4. Helix propensity: This concept increase helix propensity by replacing Proline and Glycine residues that are present in the target receptor but poorly conserved in the receptor family or class with Alanine. Glycine residues

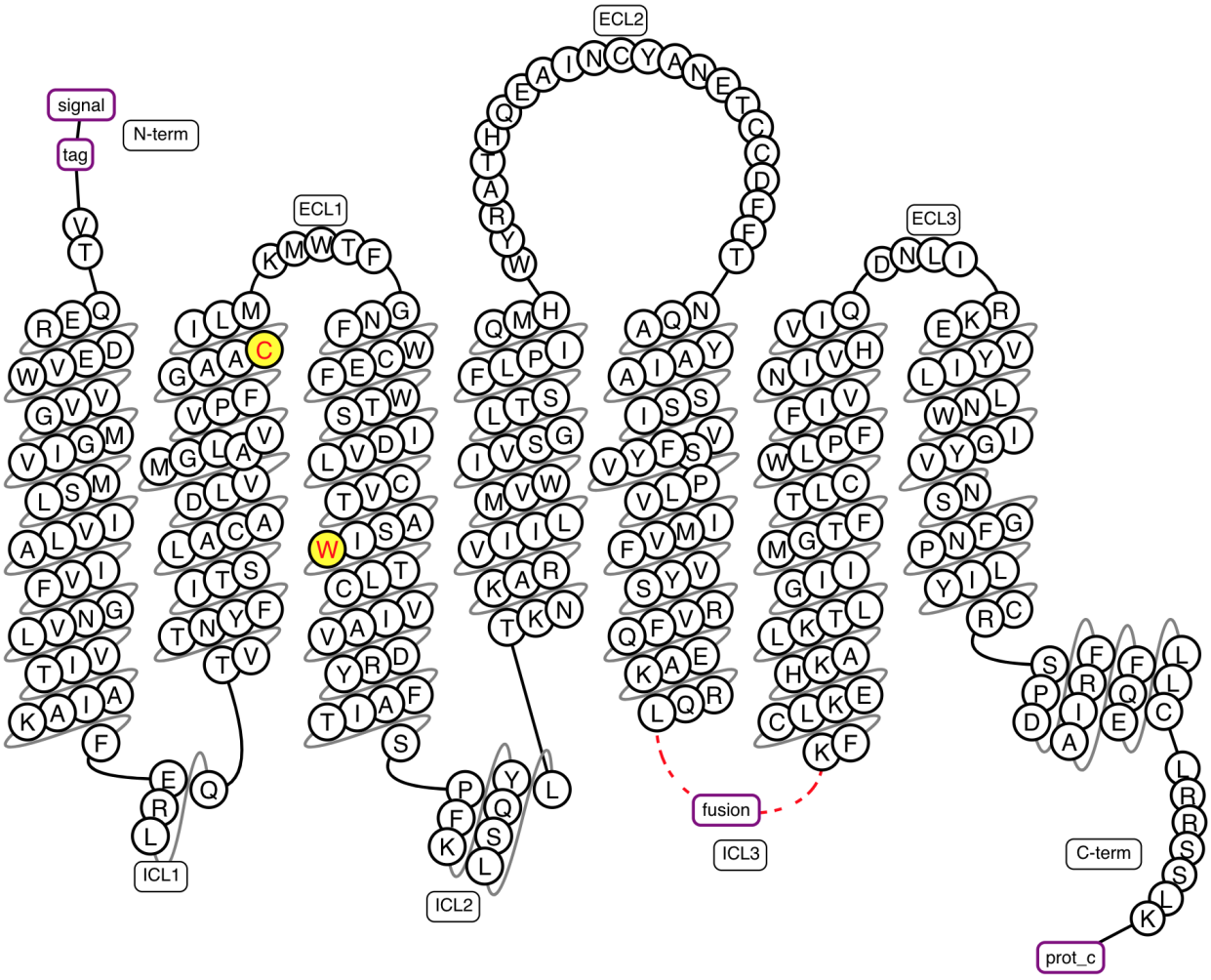

Fig. 2: **Figure 2**. Example snakeplot representation of a construct under construction.

within four positions from a helix end are preserved, as they can be crucial for the transition to a loop structure.

5. State switches: These residues form interactions that are unique for either inactive or active receptor states. The state selected by the user will be targeted by adding and removing residues with such interactions.

Truncation/fusion scan

Total constructs:		3	3			1			1			1			1				2
No	Display snake-plot / Delete other	N-term Inserts	N-term truncation			Fusion site start (last kept Rec res)*			Fusion			Fusion site end (first kept Rec res)*			C-term truncation			C-term Inserts	Add known stabilising mutations to all constructs
			N-term res kept before to TM1*	Pos	AA	BW	Pos	AA	TM5-ICL3 linker (non-wt)	Fusion	ICL3-TM6 linker (non-wt)	BW	Pos	AA	C-term res kept after h8*	Pos	AA		
1	◉	2	24	2	G	5.69	230	L		BRIL (b562RIL)		6.25	263	K	7	348	K	1	E122W 3.41x41,H93C 2.64x63
2	○	2	2	24	V	5.69	230	L		BRIL (b562RIL)		6.25	263	K	7	348	K	1	E122W 3.41x41,H93C 2.64x63
3	○	2	-3	29	D	5.69	230	L		BRIL (b562RIL)		6.25	263	K	7	348	K	1	E122W 3.41x41,H93C 2.64x63

Fig. 3: **Figure 3**. Example design table for a construct under construction.

Using the Construct Design Tool

We have three seperate applications modes, as listed below.

1. Truncation/fusion scan (generates constructs covering all the unique combinations of user-selected truncations and/or fusions).
2. Mutation scan (first designs one reference construct –which can include mutations with known stabilising effect– and then selects a number of stabilising mutations which are individually added to generate as many constructs).
3. Custom constructs (constructs are designed one at the time by repeated custom selection of truncations, fusions and mutations).

The two scanning applications, i.e. 1-2, contain the option to automatically select a given number of top ranking suggested truncations/fusions and mutations, respectively. For all three applications it is possible to edit the generated table of constructs, i.e. modify or remove an existing or add a new construct.

8.4 Experiment browser

The experiment browser shows all the annotated experimental values for constructs. Since there are so many it is possible to toggle the visibility of several types of data (Expression, Solubilization, Purification and Structure determination). Similarly to other views, it is possible to filter by almost any column to narrow down the dataset to the most relevant.

8.5 Truncation & Fusion analysis

These pages are used for detailed analysis for different key features for construct design. It shows the distribution of different cut sites in N-term, C-term and the loops (with and without fusion proteins).

8.6 Mutation analysis

We have three pages for mutation analysis in relation to construct structures.

Mutation Browser Is our manually annotated structure mutations which notes whether a mutation has an effect on thermostability, receptor expression and many other features. As with most of our views it is possible to filter and sort by most columns.

StaMutAnalyser (Stabilising Mutation Analyser) Is a tool developed to analyse mutations that cover several receptor to help determine their rationale. Whether the site is known to participate in either Ligand-binding, Arresting-binding or G-protein-binding can be seen to the right most columns with a count of distinct receptors known to have an interaction at a given position. Functional sites, such the Sodium Ion site, is also annotated to correlate that with the mutation.

Mutation substitution matrix shows the frequencies (no. GPCRs) of thermostabilising mutant amino acid substitutions across all positions of the receptor sequences/structures.

CHAPTER 9

Mutations

9.1 Mutation browser

The mutant browser allows users to view mutant data for a receptor or receptor family and highlight mutants on receptor diagrams. Using the tool is a two step process.

1. The user is first presented with a receptor selection page. Receptors can be selected individually or by family. The user can select as many receptors as he/she wishes (WARNING: selecting a large number of receptors increases loading time).
2. After receptors have been selected, the user is presented with a sequence segment selection page. The user can select one or more sequence segments, and/or expand each segment to select the residues within it individually. Residues selected individually are grouped into a custom sequence segment.

The results page shows a table of mutants for the selected receptors and segments. The table can be ordred and filtered by each column.

Below the table, helix box and snake plots are shown, with the mutated residues highlighted. The sequence in the plots is the consensus sequence of the selected receptors.

Below the plots, a table of every residue in the selected receptors and segments is shown, with the mutated residues highlighted.

9.2 Mutation data submission

The GPCRdb already contains the largest available set of GPCR mutants and the goal is to continuously deposit mutants into GPCRdb, now also capturing the pharmacological effect.

You can contribute to the mutational data available in GPCRdb, e.g. with data from your own lab to increase the visibility and thus the number of citations. You can also contribute with data sets gathered from the literature, which can be put into perspective by comparing to mutational effects in e.g. other GPCR subtypes by use of the visualization tools available in GPCRdb. To capture mutational data in a format that enables comparison of effect on e.g. ligand binding affinity, a standardized Excel spreadsheet has been prepared to collect the data. Please download it here, enter your data by following the instructions included in each cell and email the file to us (info@gpcrdb.org).

A few examples of entered mutant data are available here and if in doubt please contact us via e-mail at info@gpcrdb.org.

9.2.1 Type of mutation data

The current standardized Excel spreadsheet is made for reporting mutational effects on ligand binding and function but additionally contains the possibility to report mutational effect on surface expression, basal activity and Emax.

Future plans for the GPCRdb mutational database includes the possibility to receive and display data for mutations with effect on thermo-stabilization, biased signaling, G-protein binding and dimerization and more may be added.

9.2.2 How will the mutation data be used?

The mutant browser allows users to browse and search the mutation database for e.g. mutations in a given receptor or sub-family of receptors, mutations in a given generic numbering position, mutations with effect on a given ligand or ligand class (and much more). It is also possible to download the mutation data of interest.

Our main focus is on how mutants affect ligand binding and function. Thus, the first visualization tools that are available in the GPCRdb are snake and helix box diagrams for mapping mutated residues on the 7TM domain plus tables for comparing mutated residues across receptor subtypes. Both diagrams and tables can be color-coded according to the fold-effect of the mutation on the desired ligand property (binding or effect – see examples). It is also possible to combine the diagrams and table with information on ligand-interacting residues annotated from experimental structures to give a structural explanation for the observed mutational effect. Furthermore, it is the intention that future versions of GPCRdb tools will additionally offer mapping of mutational data on crystal structures and homology models in 3D.

GPCRdb(A)	GPCRdb(C)	GPRC5B	GPRC5C	GPRC5D	CaS receptor	$GPRC_6$ receptor	$GABA_{B1}$	$GABA_{B2}$	$mGlu_1$ receptor	$mGlu_2$ receptor	$mGlu_3$ receptor	$mGlu_4$ receptor	$mGlu_5$ receptor	$mGlu_6$ receptor	$mGlu_7$ receptor	$mGlu_8$ receptor	TAS1R1	TAS1R2	TAS1
TM1																			
1x35	1.39x39	W55	W49	W22	F612	L593	L591	L481	E592	W567	W576	W587	E579	W585	W590	W583	T566	P565	A568
1x37	1.41x41	I57	I51	I24	I614	I595	I593	S483	I594	V569	I578	V589	I581	A587	V592	V585	W568	I567	L570
1x39	1.43x43	V59	L53	L26	L616	L597	V595	L485	A596	P571	P580	P591	A583	P589	P594	P587	L570	V569	L572
1x42	1.46x46	V62	V56	L29	F619	L600	L598	L488	F599	I574	I583	L594	F586	L592	L597	V590	A573	L572	L575
TM2																			
2x42	2.38x38	L95	T89	T62	L650	V632	L629	M519	L630	L605	L614	L625	L617	L623	L628	L621	L604	M603	L606
2x49	2.45x45	G102	G96	S69	S657	C639	G636	G526	G637	G612	G621	G632	G624	G630	G635	G628	S611	L610	C613
2x53	2.49x49	L106	L100	L73	C661	N643	A640	S530	G641	C616	S625	C636	G628	I634	C639	C632	G615	A614	V617
2x54	2.50x50	F107	F101	F74	F662	F644	L641	Y531	Y642	Y617	Y626	Y637	Y629	Y635	Y640	Y633	S616	Y615	C618
2x55	2.51x51	G108	C102	G75	S663	A645	A642	A532	V643	C618	C627	A638	L630	A636	I641	S634	G617	M616	L619
2x56	2.52x52	L109	L103	L76	S664	S646	A643	S533	C644	M619	M628	T639	C631	I637	I642	I635	S618	V617	S620
2x57	2.53x53	T110	V104	A77	S665	T647	V644	I534	P645	T620	T629	T640	T632	T638	T643	T636	L619	V618	V621
2x59	2.55x55	A112	A106	A79	F667	F649	P646	L536	T647	I622	F631	L642	C634	L640	L645	L638	G621	V620	L623
2x60	2.56x56	F113	C107	F80	F668	F650	L647	F537	L648	F623	F632	M643	L635	M641	M646	M639	F622	Y621	F624
TM3																			
3x27	3.31x31	V124	S118	V91	L679	T661	A665	V555	L659	L634	L643	L654	L646	A652	F657	F650	L633	C632	A635
3x28	3.32x32	R125	R119	R92	R680	R662	R666	R556	Q660	R635	R644	R655	Q647	R653	R658	R651	R634	R633	Q636
3x29	3.33x33	R126	R120	Y93	Q681	Q663	L667	T557	R661	R636	R645	R656	R648	R654	R659	R652	Q635	Q634	Q637
3x30	3.34x34	F127	F121	F94	P682	T664	W668	W558	L662	L637	L646	I657	I649	L655	V660	V653	A636	A635	P638
3x31	3.35x35	L128	L122	L95	A683	M665	L669	I559	L663	G638	G647	F658	G650	F656	F661	F654	L637	L636	L639

Fig. 1: **Figure 1**. Sequence comparison of the 7TM domain binding pocket in the eight mGlu receptor subtypes with all residues that have been mutated. Color-coding: Green indicates increased binding/potency of >5-fold (light green) or >10-fold (dark green), red indicates reduced binding/potency of >5-fold (pink) or >10-fold(red), yellow indicates No/low effect (<5-fold), and grey indicates that no effect is annotated. The first two columns show *generic GPCRdb residue numbers* for each row of residues.

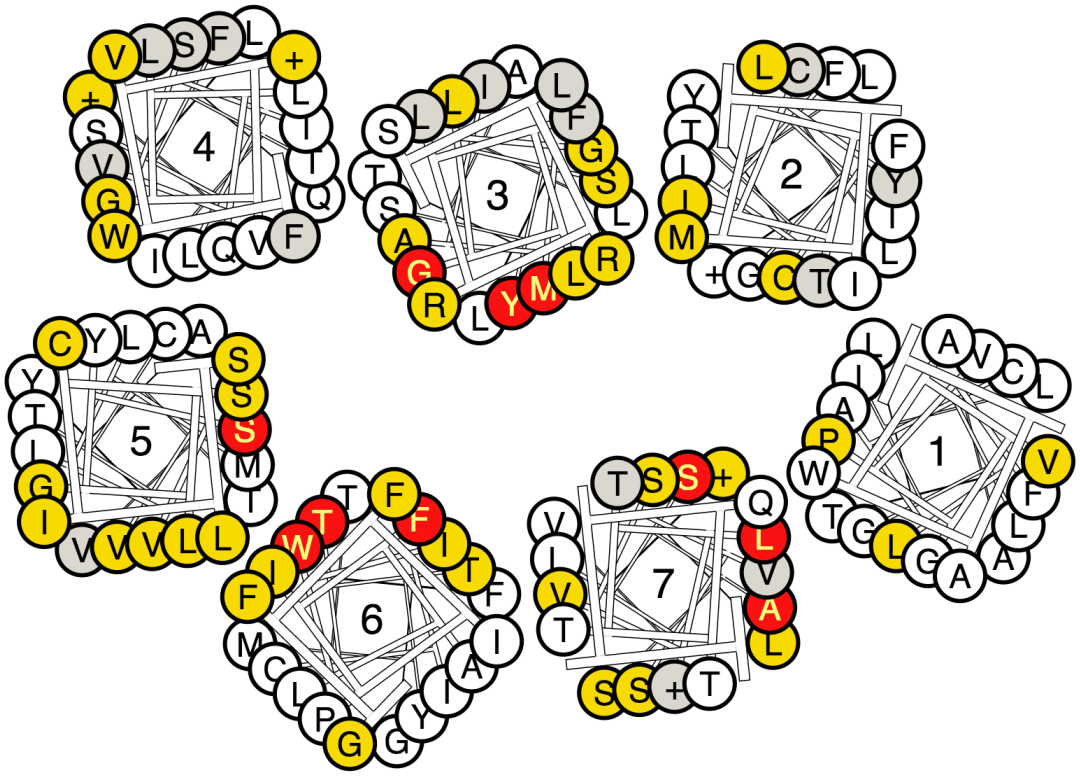

Fig. 2: **Figure 2**. A helix box diagram of the metabotropic glutamate receptors displaying mutated residue positions from the extracellular side with all residues that have been mutated. Color-coding: Green indicates increased binding/potency of >5-fold (light green) or >10-fold (dark green), red indicates reduced binding/potency of >5-fold (pink) or >10-fold(red), yellow indicates No/low effect (<5-fold), and grey indicates that no effect is annotated.

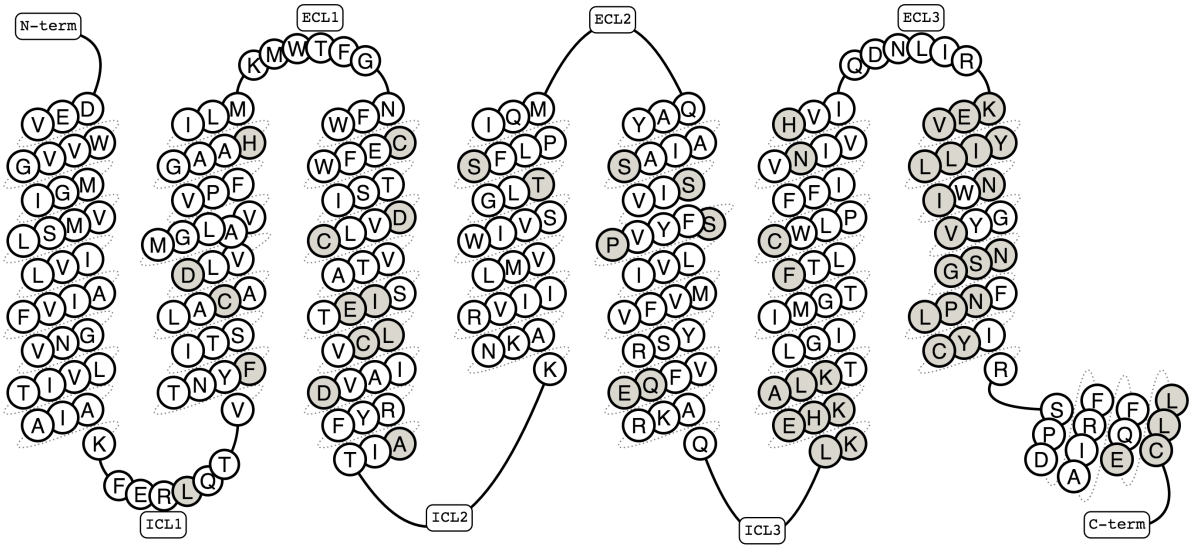

Fig. 3: **Figure 3**. Snake diagram of the human β_2 -adrenoceptor showing all residues (grey) for which mutational experiments have been deposited in the GPCRdb.

CHAPTER 10

Biased signaling

10.1 Data submission

10.1.1 Submission of biased ligands, balanced reference ligands and pathway effects

- **Definitions:** For definitions on biased ligands, see the IUPHAR community guidelines for GPCR ligand bias
- **Support:** In case you need further help, please write to info@gpcrdb.org.
- **GPCRdb:** Your data will be deposited in the next, monthly update of GPCRdb.

10.1.2 Biased ligands

Please submit your ligands tested for bias (even those found to be unbiased) using this annotation file. The file contains pop up explanations and example data, which you can replace with your data.

10.1.3 Balanced ligands (references for pathway-bias)

Please submit your balanced reference ligands (have near-equal signaling in two compared pathways as defined in a bias plot) using annotation file. The file contains pop up explanations and example data, which you can replace with your data. The balanced reference ligands will be used to calculate ligand pathway-bias and avail ligands with pathway-bias in GPCRdb.

10.1.4 Pathway effects

Please submit your data on pathway effects using this annotation file. The file contains pop up explanations and example data, which you can replace with your data.

CHAPTER 11

Sites

11.1 Ligand interactions

The ligand interaction workflow allow a user to upload a PDB file and get an analysis of protein-ligand interactions in the complex.

11.2 Site search - manual

The site search tool allows a user to search a set of receptors for a sequence motif consisting of residue positions and chemical properties. Using the tools is a two step process.

1. The user is first presented with a receptor selection page. Receptors can be selected individually or by family. The user can select as many receptors as he/she wishes (WARNING: selecting a large number of receptors increases loading time).
2. After receptors have been selected, the user is presented with a sequence motif selection page. Site residues should be selected individually. Clicking the down arrow button next to a sequence segment will expand the residues within that segment. Chemical features (Hydrophobic, hydrogen bond donor, etc.) should then be selected for each motif residue. When a feature has been selected, a list of amino acids that match the feature will appear to the right of the residue.

   The selected residues can be organised into separate interactions. An interaction can contain one or more residues. To add an interaction, click the 'Add interaction' button. Selected residues will be added to the currently active interaction (shown in bold text). To change the active interaction, click on the name of the interaction. Within an interaction, the number of residues required to match can be specified in the 'Min. match' selection box.

After completing these two steps, an alignment is displayed. The sequences of the selected receptors are split into "Matching sequences" and "Non-matching sequences", according to their match of the selected site. To display the sequence number of an aligned residue, as well as generic number indices, hover the mouse over it.

11.3 Site search - from pdb complex

This is a variant of the manual site search tool, where the user can upload a PDB structure and have protein-ligand interactions automatically detected and translated into a site search. After interactions have been detected, the user can edit the definition, and continue as in a manual search.

11.4 Pharmacophore generation

The tool is based on the following paper:

K Fidom, V Isberg, A Hauser, S Mordalski, T Lehto, AJ Bojarski, DE Gloriam, "A New Crystal Structure Fragment-Based Pharmacophore Method for G Protein-Coupled Receptors", **2015**, *Methods*, 71, 104–112. 10.1016/j.ymeth.2014.09.009

11.4.1 Abstract

We have developed a new method for the building of pharmacophores for G protein-coupled receptors, a major drug target family. The method is a combination of the ligand- and target-based pharmacophore methods and founded on the extraction of structural fragments, interacting ligand moiety and receptor residue pairs, from crystal structure complexes. We describe the procedure to collect a library with more than 250 fragments covering 29 residue positions within the generic transmembrane binding pocket. We describe how the library fragments are recombined and inferred to build pharmacophores for new targets. A validating retrospective virtual screening of histamine H1 and H3 receptor pharmacophores yielded area-under-the-curves of 0.88 and 0.82, respectively. The fragment-based method has the unique advantage that it can be applied to targets for which no (homologous) crystal structures or ligands are known. 47% of the class A G protein-coupled receptors can be targeted with at least four-element pharmacophores. The fragment libraries can also be used to grow known ligands or for rotamer refinement of homology models. Researchers can download the complete fragment library or a subset matching their receptor of interest using our new tool in GPCRdb.

11.5 Sodium ion site

In many GPCRs, a sodium ion acts as an allosteric modulator stabilising the inactive state. The sodium ion site involves a number of direct and indirect polar interactions. The receptor residues D2x50 and S3x39 are known to interact frequently and directly with the sodium ion, and mutation of these residues causes loss of G protein mediated signalling (White et al., *Structure*, **2018**; 10.1016/j.str.2017.12.013). In the "Structure Browser" under the "Receptors" menu, users can find a "Sodium ion site" column section with two columns. The first shows whether D2x50 and S3x39 are present in the wild type sequence of the structure, the second shows whether there is a sodium ion interacting with the two residues in the given structure.

CHAPTER 12

Generic residue numbering

Sequence-based generic GPCR residue numbering schemes[1] exist for class A (Ballesteros-Weinstein, BW[2]) B (Wootten[3]), C (Pin[4]), and F (Wang[5]). In these systems, the first number denotes the helix (1-7) and the second the residue position relative to the most conserved position, which is assigned the number 50. For example, 6.51 denotes a residue in transmembrane helix 6, one position after the most conserved residue (6.50). The reference helix conserved positions differ between the GPCR classes.

Recent GPCR crystal structures have revealed frequent helix bulges and constrictions in several transmembrane helices[6] . Structural superimposition makes it clear that these cause a gap that offsets all the following residue numbers when compared to an undistorted helix, i.e. the structurally equivalent residues no longer have the same number (Fig. 1).

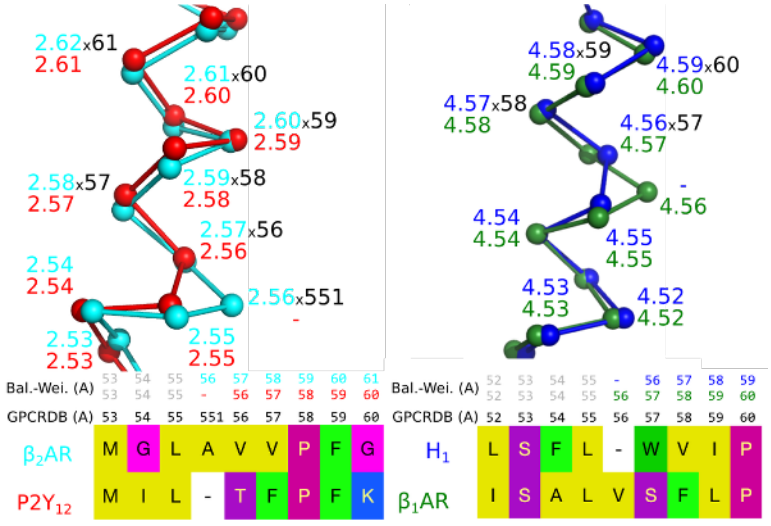

Fig. 1: **Figure 1**. A bulge in helix 2 of the Beta-2 adrenergic receptor (left) and a constriction in helix 4 of the Histamine H1 receptor (right) create offsets in the sequence-based generic numbers when compared to receptors that lack the bulge/constriction.

The GPCRdb numbering scheme[1] is the first that is based on crystal structures and corrects for helix bulges and

constrictions. GPCRdb numbers are distinguished by a unique separator x and may be used alone, e.g. 5x47, or together with one of the sequence-based schemes, e.g. 5.46x47. A bulge residue is assigned the same number as the preceding residue followed by a 1, e.g. 551 for a bulge following position 55.

GPCRdb offers a suite of tools making it easier to use generic residue numbers:

- Structure-based sequence alignments gapped to account for bulges and constrictions
- Lookup tables with receptor-specific and generic residue numbers
- PDB structure numbering along with visualization tools for PyMOL and Maestro

GPCRdb cross-class alignments contain each of the numbering schemes, which may be distinguished in text by appending the letter of the class, e.g. 2x52ax59b. The Lookup tables tool also provides the alternative class A numbering schemes by Oliveira[7] and Baldwin/Schwartz[7,8] .

References

1. V Isberg et al., **2015**, *Trends Pharmacol Sci*, 36(1), 22–31.
2. JA Ballesteros and H Weinstein, **1995**, *Methods Neurosci*, 25, 366–428.
3. D Wootten et al., **2013**, *Proc Natl Acad Sci*, 110(13), 5211-5216.
4. J-P Pin et al., **2003**, *Pharmacol Ther*, 98(3), 325-354.
5. C Wang et al., **2014**, *Nat Commun*, 5, 4355.
6. R van der Kant and G Vriend, **2014**, *Int J Mol Sci*, 15(5), 7841-7864.
7. L Oliveira et al., **1993**, *J Comput Aided Mol Des*, 7(6), 649–658.
8. JM Baldwin, **1993**, *EMBO J*, 12(4), 1693–703.
9. TW Schwartz, **1994**, *Curr Opin Biotechnol*, 5(4), 434–44.

CHAPTER 13

Drugs

13.1 Drug statistics

The statistics page shows several graphs for FDA-approved drugs and agents in clinical trials including:

- Class targets
- Receptor family targets
- Receptor targets
- Drug molecule types
- Mode of action
- Disease indications
- Phase distribution of clinical trial agents
- Approval ovcer time

13.2 Drug target mapping

Established targets have approved drugs as defined in the Drugs@FDA database, and targets of agents in clinical trials were collected from manual annotation of CenterWatch's Drugs in Clinical Trials database, OpenTargets, Drugbank, Pharos and company press releases. Established (red) and phase I–III (green) targets across the G protein-coupled receptor (GPCR) classes, ligand types and receptor families (from the centre to the outer ring) are shown.

The sizes of the circles represent the number of agents. For receptor families, two concentric circles are superimposed: a red circle indicating the number of approved agents (that is, which have an established target in that family) and a green circle indicating the number of agents in trials for the targets in that family. The area over which the two circles overlap is shown in brown. For example:

- for adrenoceptors, there are 117 approved agents and 41 agents in trials, and so the red circle is larger
- for chemokine receptors, there are 2 approved agents and 37 agents in trials and so the green circle is larger

At the family level, agents that modulate multiple receptors in the family are only counted once to determine the circle size. For individual receptors (but not families), different shades of green are used for each trial phase.

GPCRs are listed using the protein name (EntryName removed species tag) in UniProt.

13.3 Drug browser

The structure table shows the complete annotated list of DRUG-GPCR pairs. The table can be sorted by each column by clicking on the header. The search fields below each header can be used to filter the structures.

- Hauser, A. S., Misty, A, Mathias, R., Schiöth, H. B., Gloriam, D. E., "Trends in GPCR drug discovery: new agents, targets and indications", **2017**, *Nature Reviews Drug Discovery*, *in print*, 10.1038/nrd.2017.178

CHAPTER 14

Sales and prescription (NHS)

14.1 NHS sales

Every month, the National Health Service (NHS) in the UK publishes anonymised data about the drugs prescribed by general practitioners. NHS data were retrieved from openprescribing.net (DataLab-EBM, 2017, https://openprescribing.net/) (08/2017) for the list of drugs targeting GPCRs and mapped back to their reported target of therapeutic action. From the 475 queried FDA-approved drugs, data were available for 279 drugs targeting 92 distinct GPCRs (not all FDA-approved drugs are prescribed in the UK due to alternative treatments).

The actual cost is the estimated cost to the NHS, which is usually lower than Net Ingredient Cost ("the basic price of a drug, i.e. the price listed in the Drug Tariff or price lists"). Openprescribing.net provides the actual cost by subtracting the average percentage discount per item received by pharmacists based on the previous month from the Net Ingredient Cost, but adding in the value of a container allowance for each prescription item (DataLab-EBM, 2017).

Indications were grouped according to the British National Formulary (BNF), which is a reference book containing the standard list of medicines prescribed in the UK. Individual drugs can be selected (double click on colored bar chart) in the section page).

Drugs with NHS information can be filtered via the drug browser (column on the far right, NHS=yes).

Sales and variation information per drug is presented within the *Estimated economic burden* page. Here, sales and prescription averages are presented along with aggregated counts of putative and known functional site variants of all receptors targeted by the drug.

14.2 Estimated economic burden

The economic burden estimate was calculated using the following formula:

estimated economic burden per drug (£) = average NHS cost per drug per year (£) **x** *%individuals with a MV in a functional site of the respective drug targets*

where:

- The average NHS cost is the average yearly cost over a 4-year period (2013-2016) per GPCR targeting drug that is listed (n=279). 2012 and 2017 have partial sales data and were not considered.
- % Individuals is the percentage of affected individuals with a missense variant in a functional site of the respective drug target(s) (n=2,504 individuals from 1,000 Genomes Project genotype data as a representative for the UK population; this data includes non-Caucasian populations as well) (Table S9).
- The % of affected individuals was calculated using four different criteria by considering individuals who have a variation in (i) known functional sites in both alleles (homozygous), which is the most conservative, (ii) known functional sites in at least one allele (i.e. homozygous and heterozygous), (iii) known or putative functional sites in both alleles (homozygous), and (iv) known or putative functional sites in at least one allele (i.e. homozygous and heterozygous), which is the least conservative.
- Known functional sites include ligand binding, effector binding, PTM site, sodium binding site and microswitches. Putative functional site include those predicted to be deleterious based on SIFT or PolyPhen (see above).

More specifically, for each drug we collected the respective targets and computed economic burden using the following four criteria above: considering (i) % individuals with homozygous alleles in known functional sites, (ii) % individuals with at least one variant allele in a known functional site, (iii) % individuals with homozygous alleles in known or putative functional sites and (iv) % individuals with at least one variant allele in a known or putative functional sites.

For these estimates, we have incorporated the following considerations (below). The economic burden estimates will vary if one scales/factors these variables differently:

1. We have considered that each prescription (NHS data) is made for a unique individual, due to patient anonymity. Furthermore, information about the dose per prescription, and how this has been altered based on patient response is not explicitly modelled.
2. The effect of known and putative site polymorphisms as well as homozygous/heterozygous conditions are all treated the same way. One could also obtain estimates by weighing these variables differently on a case-by-case basis for each receptor/drug.
3. The focus has been prescription only from GPs. There might be significant additions to the economic burden if one also considers hospital prescriptions.
4. We used the data from 1000 Genomes Project as representative of the UK population, which may vary depending on the receptor.
5. We have not explicitly modeled the age, gender, nature of illness (chronic v/s short-term) and mutations in non-coding regions, which may affect expression level.

- Hauser, A. S., Chavali, S., Masuho, I., Jahn, L. J., Martemyanov, K., Gloriam, D. E., Babu, M. M., "Pharmacogenomics of GPCR drug targets", **2017**, *Cell* 10.1016/j.cell.2017.11.033

CHAPTER 15

Genetic variants

15.1 Variation statistics

The statistics page shows an overview of all non-olfactory receptors with their absolute and relative numbers of reported genetic missense variants. The average density (absolute count divided by length of receptor) is presented for each receptor family, ligand type and class.

15.2 Receptor variant browser

The selection page allows users to find a receptor by searching or browsing. The browser displays all missense and loss of function mutations for the selected receptor. Additional information for each variant is presented such as position, amino acid change, allele counts/frequencies, number of homozygotes, predicted functional impact by SIFT and PolyPhen as well as functional annotation of the specific position (PTM site, ligand-binding site, micro-switch, G protein interaction ..)

All variants are highlighted on helix plot and snake-plot diagrams with predicted deleterious variants (by either SIFT or PolyPhen) highlighted in red or green, respectively.

- Hauser, A. S., Chavali, S., Masuho, I., Jahn, L. J., Martemyanov, K., Gloriam, D. E., Babu, M. M., "Pharmacogenomics of GPCR drug targets", **2017**, *Cell* 10.1016/j.cell.2017.11.033

CHAPTER 16

Web services

Most data in GPCRdb is available programmatically via a REST API.

16.1 API reference

Each endpoint is described in the API reference.

16.2 Examples

16.2.1 Python 3 with requests

This is the recommended approach. Requires installation of the requests library.

```
import requests

# fetch a protein
url = 'https://gpcrdb.org/services/protein/adrb2_human/'
response = requests.get(url)
protein_data = response.json()
print(protein_data)
print(protein_data['sequence'])

# fetch an alignment
url = 'https://gpcrdb.org/services/alignment/protein/adrb1_human,adrb2_human,adrb3_
↪human/TM3,TM5,TM6/'
response = requests.get(url)
alignment_data = response.json()
for protein, sequence in alignment_data.items():
    print(protein)
    print(sequence)
```

16.2.2 Python 3 with urllib

```
from urllib.request import urlopen
import json

# fetch a protein
url = 'https://gpcrdb.org/services/protein/adrb2_human/'
response = urlopen(url)
protein_data = json.loads(response.read().decode('utf-8'))
print(protein_data)
print(protein_data['sequence'])

# fetch an alignment
url = 'https://gpcrdb.org/services/alignment/protein/adrb1_human,adrb2_human,adrb3_
↪human/TM3,TM5,TM6/'
response = urlopen(url)
alignment_data = json.loads(response.read().decode('utf-8'))
for protein, sequence in alignment_data.items():
    print(protein)
    print(sequence)
```

16.2.3 Python 2 with urllib2

```
from urllib2 import urlopen
import json

# fetch a protein
url = 'https://gpcrdb.org/services/protein/adrb2_human/'
response = urlopen(url)
protein_data = json.loads(response.read())
print protein_data
print protein_data['sequence']

# fetch an alignment
url = 'https://gpcrdb.org/services/alignment/protein/adrb1_human,adrb2_human,adrb3_
↪human/TM3,TM5,TM6/'
response = urlopen(url)
alignment_data = json.loads(response.read())
for protein, sequence in alignment_data.iteritems():
    print protein
    print sequence
```

CHAPTER 17

Contributing to the project

We welcome all contributions to the project. If you have an idea for a feature you would like to implement, improvements to make, or data to add/update, please *contact us*.

17.1 As a programmer

We use languages/tools such as Python, Javascript, Django, PostgreSQL, and Git to build GPCRdb. Prior knowledge of these tools is helpful, but not necessary.

17.2 As a data curator

Data curation tasks involve e.g. sequence alignments, analysis of protein structures and collection of mutation data.

CHAPTER 18

Local installation

18.1 For development

To start working on GPCRdb, fork the source code on GitHub, and use Vagrant to set up a development environment.

Developer guide at GitHub

18.2 For internal use

To install GPCRdb for local use at your company or organization, provision a local server with Puppet.

Production guide at GitHub

CHAPTER 19

Coding style

We (mostly) follow the style guide from the the Django project. Unless otherwise specified, follow this guide. Please read this guide, use it, and feel free to point out if existing code does not comply with the style guide.

19.1 Examples

- Max line length is 119 characters
- Indentation is 4 spaces:

```
for protein in proteins:
    print(protein)
```

- Comments start with a # and a single space:

```
# this is a comment
```

- Docstrings use """:

```
"""This is a docstring"""
```

- Use lower case letters and underscores for variable and function names, upper case letters and underscores for constants, and InitialCaps for class names:

```
this_is_a_variable = True

THIS_IS_A_CONSTANT = True

def this_is_a_function():
    pass

class ThisIsAClass:
    __init__(self):
        pass
```

- Class definitions are followed by 2 blank lines:

```
class ThisIsAClass:
   __init__(self):
      pass


class ...
```

- Import statements are grouped in three categories(django, project, and other), separated by one blank line, and followed by 2 blank lines:

```
from django.conf import settings

from protein.models import Protein

import yaml


class ...
```

19.2 Keep your code clean

Before committing, review the changes you have made (using git diff or a GUI like SourceTree) and make sure the code you are committing is working, and relevant. Never commit lines of code that are commented out (comments are for, well, comments), or print statements that you used for debugging.

CHAPTER 20

Recommended git workflow

20.1 Preface

There exist many workflows and guides for using Git, and everyone has their own preferred ways of handling certain aspects of their Git repositories.

Whether or not you follow this guide in detail, or use other methods, it is important that you know what the commands you are using do, and understand basic Git operations such as committing, pulling, merging, pushing and rebasing.

Please refer to the git documentation as needed, and create small demo repositories to test common operations before applying them on to the Protwis repository. If in doubt, you are always welcome to contact members of the *Protwis team*.

20.2 Prerequisites

If you have followed the *setup guide*, you should already have created a fork of the Protwis repository, and cloned the fork to your computer. This means that you have access to three different repositories, all containing the same code. From your perspective, these repositories are referred to as:

- upstream (the main protwis repository, where you have read only access)
- origin (your fork of protwis, where you have full write access)
- local (the repository you work on locally)

20.2.1 Configuring the upstream repository

The upstream repo is currently not connected to your local repo. To connect it, type the following on you local command line (from the repo root dir):

```
git remote add upstream https://github.com/protwis/protwis.git
```

20.3 Workflow

20.3.1 Branches

Before doing any changes to the code, create and check out a new branch:

```
git checkout -b feature/branch-name-describes-ticket
```

You can always see which branch you are on by typing:

```
git branch
```

20.3.2 Committing

Once you have made changes on your branch, add them to the index and commit them:

```
git add my_file.py
git commit -m "Optimized the performance of my_file"
```

ALWAYS add a commit message with the -m flag.

Note that only files that have been added to the index will be committed, and you can add all modified or new files with:

```
git add --all
```

20.3.3 Keeping your branch up to date

While you work on your branch, other developers may push their commits to the master branch. It is important that you keep your repository updated with the latest changes. Do this *DAILY*.

To fetch the latest changes, checkout the master branch (make sure to commit all changes to your branch first) and pull from upstream:

```
git checkout master
git pull upstream master
```

Your local master branch is now up to date, but your feature branch is not. To update it, use the rebase command:

```
git checkout my_feature_branch
git rebase master
```

The rebase will usually go through without issues, but if Git can not merge the changes automatically, a merge conflict will arise.

If this happens, open the conflicted file (Git will tell you which file is conflicted) in a text editor. Conflicts are displayed as two versions of the conflicted code block, one marked "HEAD", and one marked "master". There may be more than one conflict in the same file. Edit the file manually to resolve the conflict(s) (i.e. remove one of the versions, or combine them). Then add the file to the index, and continue the rebase:

```
git add path/to/file
git rebase --continue
```

This will usually complete the rebase. However, it is possible that a new conflict will arise. If this happens, do not worry. Simply follow the same steps as before to resolve the conflict(s), until the rebase is completed.

20.3.4 Merging your branch into master

NOTE! Make sure your master and feature branches are updated before doing this. When the changes on your feature branch are ready, merge them into master:

```
git checkout master
git merge my_feature_branch
```

20.3.5 Pushing changes to Github and sending a pull request

After merging your changes into master, you should push them to your fork on Github (origin) and send a pull request (PR) to the main repository:

```
git push origin master
```

Then go to the main website of your fork and select the "Create pull request" option in the left menu. The PR should be from your fork's master branch, to upstream/master.

CHAPTER 21

Reload database from dump

- Go to the project root directory on your virtual machine:

```
cd /protwis/sites/protwis
```

- Download the newest dump from gpcrdb [Optional when you already have a dump]:

```
curl https://files.gpcrdb.org/protwis_sp.sql.gz > ~/protwis.sql.gz
```

- Delete the current database (password: protwis):

```
dropdb -U protwis -h localhost protwis
```

- Load the dump (Either from default location or a location of your choosing):

```
gunzip -c ~/protwis.sql.gz | psql -U protwis -h localhost -d protwis
```

- **Optional steps:**
  - Restart apache2 to fix adminer behavior in some cases:

    ```
    sudo service apache2 restart
    ```

  - Pull the latest source code from the default branch of protwis (assuming it is set as the default repository):

    ```
    git stash; git pull; git stash pop
    ```

  - Update the static file repository (not for Vagrant setup):

    ```
    <python> manage.py collectstatic
    ```

  - Rebuild the local BLAST database:

    ```
    <python> manage.py build_blast_database
    ```

- Clear the Django caches:

```
<python> manage.py clear_cache
```

CHAPTER 22

Building a local database from source data

- If you have not completed the *local installation* of GPCRdb, please do so before continuing.
- Open up a terminal and clone the gpcrdb_data repository from GitHub:

```
cd ~/protwis_vagrant
git clone https://github.com/protwis/gpcrdb_data.git shared/data/protwis/gpcr
```

- Log into the vagrant VM:

```
vagrant ssh
cd /protwis/sites/protwis
```

- Clean the current database schema (password: protwis):

```
psql -U protwis -h localhost -d protwis -c 'drop schema public cascade; create␣
↪schema public;'
```

- Run migrations:

```
/env/bin/python3 manage.py migrate
```

- Start the build process:

```
/env/bin/python3 manage.py build_all -p 4 -t
```

This will build a test version of the database using only the proteins for which a structure has been determined. For a full build, remove the -t flag from the build_all command (NOTE: a full build takes a long time, and should not be run on the development virtual machine)

CHAPTER 23

About GPCRdb

GPCRdb offers reference data and easy-to-use web tools and diagrams for a multidisciplinary audience investigating GPCR function, drug design or evolution. It stores a manual annotation of all GPCR crystal structures, the largest collections of receptor mutants and reference sequence alignments. The tools run directly in the web browser allowing for swift analysis of structures, sequence similarities, receptor relationships, and ligand target profiles. Diagrams illustrate receptor sequences (snake-plot and helix box diagrams) and relationships (phylogenetic trees). A visual overview can be seen in the GPCRdb poster.

23.1 Background and development

The GPCR database, GPCRdb was started in 1993 by Gert Vriend, Ad IJzerman, Bob Bywater and Friedrich Rippmann. Over two decades, GPCRdb evolved to be a comprehensive information system storing and analysing data. In 2013, the stewardship of GPCRdb was transferred to the David Gloriam group at the University of Copenhagen, backed up by an international team of contributors and developers from the EU COST Action 'GLISTEN'.

CHAPTER 24

Contact

To contact the authors of GPCRdb, please use the e-mail address: info@gpcrdb.org.

CHAPTER 25

Contributors of data and development

University / Institute	Country	Collaborators	Data	Development
MRC Laboratory of Molecular Biology	UK	Madan Babu	G protein alignments	
Paul Scherrer Institute	Switzerland	Xavier Deupi	Crystallisation database	
Philipps-Universität Marburg	Germany	Peter Kolb		Virtual screening
Pompeu Fabra University	Spain	Jana Selent		Molecular dynamics
Radboud University	Netherlands	Gert Vriend		Alignment methods
University of Bonn	Germany	Anke Schiedel	Mutation data	
University of Copenhagen	Denmark	Hans Bräuner-Osborne	Mutation data	
University of Southern California	USA	Ray Stevens		Construct design tool
University of Warsaw	Poland	Slawomir Filipek	Mutation data	
Uppsala University	Sweden	Hugo Gutiérrez-de-Terán	Mutation data	PyMol plugin
VU Amsterdam	Netherlands	Chris de Graaf	Mutation data	Knime workflows

CHAPTER 26

Citing GPCRdb

If you use GPCRdb in your work, please cite one or more of the following:

26.1 Main reference for GPCRdb

- Pándy-Szekeres G, Munk C, Tsonkov TM, Mordalski S, Harpsøe K, Hauser AS, Bojarski AJ, Gloriam DE. GPCRdb in 2018: adding GPCR structure models and ligands. **2017**, *Nucleic Acids Res.*, Nov 16. 10.1093/nar/gkx1109

26.2 Introduction to new users (review)

- Munk, C., Isberg, V., Mordalski, S., Harpsøe, K., Rataj, K., Hauser, A. S., Kolb, P., Bojarski, A. J., Vriend, G. , and Gloriam, D. E. GPCRdb: the G protein-coupled receptor database – an introduction. **2016**, *Br J Pharmacol*, May 8. 10.1111/bph.13509

26.3 Structure-based alignments and generic residue numbering

- V Isberg, C de Graaf, A Bortolato, V Cherezov, V Katritch, F Marshall, S Mordalski, J-P Pin, RC Stevens, G Vriend, DE Gloriam, "Generic GPCR Residue Numbers - Aligning Topology Maps While Minding The Gaps", **2015**, *Trends Pharmacol Sci*, 36(1), 22–31. 10.1016/j.tips.2014.11.001
- R van der Kant, G Vriend, "Alpha-Bulges in G Protein-Coupled Receptors", **2014**, *Int J Mol Sci*, 15(5), 7841-7864. 10.3390/ijms15057841

26.4 GPCR drugs and targets

- Hauser, A. S., Misty, A, Mathias, R., Schiöth, H. B., Gloriam, D. E., "Trends in GPCR drug discovery: new agents, targets and indications", **2017**, *Nature Reviews Drug Discovery*, *in print*, 10.1038/nrd.2017.178

26.5 GPCR-G protein selectivity

- Flock, T., Hauser, A. S., Lund, N., Gloriam, D. E., Balaji, S., & Babu, M. M., "Selectivity determinants of GPCR–G-protein binding.", **2017**, *Nature*, May 18;545(7654):317-322 10.1038/nature22070

26.6 Mutation design tool

- C Munk, K Harpsøe, A Hauser, V Isberg, DE Gloriam, "Integrating structural and mutagenesis data to elucidate GPCR ligand binding", 2016, Curr Opin Pharcol, 30, 51–58. 10.1016/j.coph.2016.07.003

26.7 Crystal structure fragment-based pharmacophore models

- K Fidom, V Isberg, A Hauser, S Mordalski, T Lehto, AJ Bojarski, DE Gloriam, "A New Crystal Structure Fragment-Based Pharmacophore Method for G Protein-Coupled Receptors", 2015, Methods, 71, 104–112. 10.1016/j.ymeth.2014.09.009

26.8 GPCR specific PDF reader

- B Vroling, D Thorne, P McDermott, TK Attwood, G Vriend, S Pettifer, "Integrating GPCR-specific information with full text articles", 2011, BMC Bioinformatics, 12, 362. 10.1186/1471-2105-12-362

26.9 Older GPCRdb articles

- V Isberg, S Mordalski, C Munk, K Rataj, K Harpsøe, AS Hauser, B Vroling, AJ Bojarski, G Vriend, DE Gloriam. "GPCRdb: an information system for G protein-coupled receptors", **2016**, *Nucleic Acids Res.*, 44, D356-D364. 10.1093/nar/gkv1178
- V Isberg, B Vroling, R van der Kant, K Li, G Vriend* and DE Gloriam*, "GPCRDB: an information system for G protein-coupled receptors", **2014**, *Nucleic Acids Res.*, 42 (D1), D422-D425. 10.1093/nar/gkt1255
- B Vroling, M Sanders, C Baakman, A Borrmann, S Verhoeven, J Klomp, L Oliveira, J de Vlieg, G Vriend, "GPCRDB: information system for G protein-coupled receptors", 2011, Nucleic Acids Res., 39(suppl 1), D309-19. 10.1093/nar/gkq1009
- F Horn, E Bettler, L Oliveira, F Campagne, FE Cohen, G Vriend, "GPCRDB information system for G protein-coupled receptors", 2003, Nucleic Acids Res., 31(1), 294-297. 10.1093/nar/gkg103
- F Horn, G Vriend, FE Cohen, "Collecting and harvesting biological data: the GPCRDB and NucleaRDB information systems", 2001, Nucleic Acids Res., 29(1), 346-349. 10.1093/nar/29.1.346
- F Horn, J Weare, MW Beukers, S Hörsch, A Bairoch, W Chen, Ø Edvardsen, F Campagne, G Vriend, "GPCRDB: An information system for G protein-coupled receptors", 1998, Nucleic Acids Res., 26(1), 275-279. 10.1093/nar/26.1.275

CHAPTER 27

Acknowledgements

Welcome to the GPCRdb (G Protein-Coupled Receptor database) acknowledgement page, which has two sections written by the current and former heads of GPCRdb, David E. Gloriam and Gerrit Vriend, respectively.

27.1 GPCRdb versions since 2013

By David E. Gloriam, University of Copenhagen, Denmark

Firstly, we would like to thank the founding father and two-decade protector of GPCRdb, Gerrit Vriend for so generously giving us the database as a gift without reservations. We promise to do the best to build on its legacy, going forward in the spirit of engaging and serving the GPCR community. The former lead developer Bas Vroling played a large role in making the transition of the data and previous codebase possible.

The first Copenhagen version of GPCRdb was the Tools subsite published in NAR, 2014. This sprung mainly from a series of computational drug design data and tools developed by Vignir Isberg during his PhD studies. As a lead developer he has driven the database far beyond anticipation, coordinating a team of international developers with enthusiasm and persistence. The past and current members in the Gloriam group have preserved the cross-fertilisation between developers and users. Kasper Harpsøe has taken a main role in the user expertise and development of the new format for mutant data submissions and storage. With the risk of forgetting someone along the way to the current wholly re-coded current version, we simply refer to the author lists of the various GPCRdb publications.

GPCRdb would not have been where it is without the GLISTEN EU Cost Action, coordinated by Peter Kolb and Chris de Graaf. You brought us into the party – allowing us to have satellite meetings for the international GPCRdb developers and contributors at each of the biannual GLISTEN meetings. The local organisers have so kindly provided room and practical coordination. Furthermore, the GLISTEN financial support made possible a number of in- and outgoing short-term scientific visits to set-up and build collaborations.

A big thanks goes Andrzej Bojarski and his group to whom we own thanks for most of the work behind the GPCRdb structure tools and phylogenetic trees. You generously shared so much of your time, and arranged for short- and long-term visits to facilitate the joint programming.

The whole GuideToPharmacology team is acknowledged for its openness to setting up our collaboration with mutual cross-linking, web services and GPCRdb's adoption of the official receptor nomenclature. You have served as a true

inspiration from a much larger resource that has walked many of the paths of database development and curation before.

Finally, we would like to extend thanks to newly established collaborations. Xavier Deupi and his lab are acknowledged for choosing to work with GPCRdb, while sharing the local expertise. We thank Raymond C. Stevens and Michael Hanson for welcoming GPCRdb as a partner to the GPCR Consortium, which holds great promise to be synergistic resources/initiatives.

27.2 GPCRdb versions 1993-2013

By Gerrit Vriend, Radboud University, Nijmegen, Netherlands

The GPCRdb was started in the early 90's when Bob Bywater, Ad IJzerman, Friedrich Rippmann, and Gert Vriend organized a series of small GPCR workshops at the EMBL. Before the introduction of the first browsers, the GPCRdb worked as an automatic Email answering system that could send sequences, alignments, and homology models to the users.

In 1994 the internet was firmly established in its present form, and money was obtained from the fourth EU framework to set up the GPCRdb. Florence Horn joined us to do this project. When she left us at the end of a four-year post-doc period the GPCRdb was firmly established as the prime source of information for GPCR data.

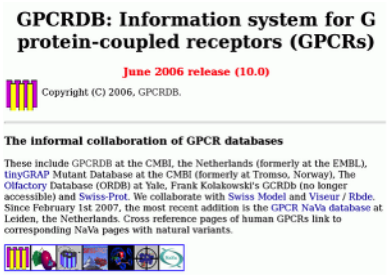

Fig. 1: **Figure 1**. GPCRdb until 2006.

In 2007 TIPharma offered us the possibility to revive the GPCRdb. Bas Vroling joined the team and revived the GPCRdb. We would also like to thank NBIC for their support. This page would not be complete without Laerte Oliveira. Ever since the start of the GPCRdb project Laerte has been our GPCR dictionary. He knows the literature, he knows all sequences by hearth, he is responsible for the alignments, and for a series of innovations. Laerte recently retired, but he is still our full-time adviser.

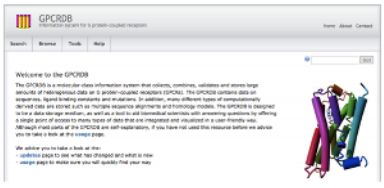

Fig. 2: **Figure 2**. GPCRdb 2007-2013.

Many people have contributed over the years to the shape of the GPCRdb that you see now. Rob Hooft was, and Maarten Hekkelman now is our bit and byte guru. Maarten also wrote the profile BLAST. Fabien Campagne wrote the snake plot software for us. Margot Beukers, Fred Cohen, Oyvind Edvardsen, Kurt Kristiansen, have been involved in the mutant section of the GPCRdb; Oyvind and Kurt made tinyGRAP that now is integrated in the GPCRdb. Wilma Kuipers, Nora vd Wenden, Mike Singer, and Frank Kolakowsky were good colleagues and intellectual sparring partners that helped shape the GPCRdb in its early days. Lisa Holm, Karl Aberer, Amos Bairoch, Nigel Brown, Antonio Paiva, Thure Etzold, and Antoine Daruvar have over the last two decades all contributed to the GPCRdb.

CHAPTER 28

Legal notice

28.1 Formal

The GPCRdb server and data suite is a free data distribution system. The data and the server are distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY without even the implied warranty of merchantability or fitness for a particular purpose. The copyright holders and/or other parties provide the data "AS IS", without warranty of any kind, either expressed or implied. The entire risk as to the quality and performance of the data is with you. Should any data prove defective, you assume the cost of all necessary servicing, repair or correction.

The copyrights for ALL information stored in the GPCRdb belongs to the partners in the GPCRdb project unless stated otherwise. It is free to make unlimited use of the GPCRdb data. The use of data or services provided by the GPCRdb project should be acknowledged the classical way by referring to the articles written by the GPCRdb project partners, and preferably also by mentioning: "We acknowledge the use of the GPCRdb database (https://www.gpcrdb.org).".

28.2 Informal

The complete GPCRdb source code is freely available on GitHub under the Apache 2.0 license. This means that the code can be used, modified and distributed for any use (including commercial use), as long as the original license and copyright notice are included with the code, and any significant changes stated.

The data in GPCRdb is freely available under the Creative Commons Attribution 4.0 International license. This means that the data can be copied, redistributed, remixed, transformed and built upon, as long as appropriate credit is giving, a link to the license provided and any changes stated.

28.3 Privacy

We voluntarily maintain the following privacy rules:

- We collect usage statistics with Google Analytics
- We take all possible measures to ensure that all detailed usage will remain secret.

- Usage data will not be sold or shared with third parties unless for academic purposes, and after we obtained the guarantee that our voluntary privacy rules will be obeyed.
- Detailed usage data will only be used by the site curators for optimisation and error detection purposes.
- We reserve the right to publicise, on our site, the names or internet address of any site using this resource, including frequencies of usage and breakdown of usage data types and tools (not individual receptor targets).
- We reserve the right to occasionally make the usage statistics available to granting organisations but these lists also take the above listed restrictions into account.

CHAPTER 29

Meetings with GPCRdb representation

29.1 2020

- March, 2nd ERNEST meeting Online, hosted from Istambul, Turkey.

29.2 2019

- Dec 1-5, GPCR Workshop in Hawaii, USA.
- Oct 28-30, 1st ERNEST meeting in Belfast, Ireland.

29.3 2018

- Jul, 9th GLISTEN meeting in Berlin, Germany.

29.4 2017

- Mar 29-31, 8th GLISTEN meeting in Porto, Portugal.

29.5 2016

- Sep 26-27, 7th GLISTEN meeting in Prague, Czech Republic.
- Apr 6-8, 6th GLISTEN meeting in Erlangen, Germany.
- Feb 21-25, G Protein-Coupled Receptors: Structure, Signaling and Drug Discovery in Keystone, USA.

29.6 2015

- Dec 5-6, 3rd annual GPCR Forum in Shanghai, China.
- Oct 12-14, 5th GLISTEN meeting in Amsterdam, Netherlands.
- Aug 27-28, 1st Annual Danish Bioinformatics Conference in Odense, Denmark.
- Aug 24-27, Benzon Symposium No. 61 in Copenhagen, Denmark.
- Apr 1-2, 4th GLISTEN meeting in Allschwill, Switzerland.

29.7 2014

- Dec 7, 2nd annual GPCR Forum in Shanghai, China.
- Oct 2-4, 3rd GLISTEN meeting in Budapest, Hungary.
- Jul 13-18, World Congress of Pharmacology in Cape Town, South Africa.
- Apr 28-29, 2nd GLISTEN meeting in Barcelona, Spain.

29.8 2013

- Oct 7-9, 1st GLISTEN meeting in Warsaw, Poland.

CHAPTER 30

Linking to GPCRdb

To link GPCRdb protein pages, download the Uniprot mapping file and use the following link format:

https://gpcrdb.org/protein/{gpcrdb_id}/

For example, for the 5-HT_{2A} receptor, the link is:

https://gpcrdb.org/protein/5ht2a_human/

CHAPTER 31

External GPCR sites

31.1 Modeling servers

31.1.1 GPCRM

GPCRM is a novel method for fast and accurate generation of GPCR models using averaging of multiple template structures and profile-profile comparison. In particular, GPCRM is the first GPCR structure predictor incorporating two distinct loop modeling techniques: Modeller and Rosetta together with the filtering of models based on the Z-coordinate.

31.1.2 scPDB

To assist structure-based approaches in drug design, we have processed the PDB to identify binding sites suitable for the docking of a drug-like ligand and we have so created a database called sc-PDB. The sc-PDB database provides separated MOL2 files for the ligand, its binding site and the corresponding protein chain(s). Ions and cofactors at the vicinity of the ligand are included in the protein.

31.1.3 GPCR-SSFE

The GPCR-Sequence-Structure-Feature-Extractor (SSFE) database provides template suggestions and homology models of the helical regions of 5025 family A GPCRs. SSFE is based on our published workflow for identifying key sequence and structural motifs in family A GPCRs which is used to guide template selection and build homology models.

31.1.4 GOMoDo

GOMoDo (GPCR Online Modeling and Docking) is a web server that allows to seamlessly model GPCR structures and dock the corresponding ligands to these models in a single consistent pipeline. GOMoDo can automatically perform template choice (using HHSearch), homology modeling (with Modeller), binding pocket prediction (with FPocket),

and either blind or information-driven docking (using AutoDock Vina or HADDOCK, respectively). By combining all these state-of-the-art bioinformatic tools, the web server guides the user through the whole procedure, while still permitting manual intervention.

31.1.5 GPCR-ModSim

This server was created to allow any researcher with interest in GPCRs to obtain the most accurate structural and dynamic information for a given receptor. Here, you can generate a homology-based 3D model of your query GPCR sequence, and/or further equilibrate your GPCR structure with our all-atom Molecular Dynamics simulation protocol.

31.1.6 Hybrid MM/CG webserver

Hybrid Molecular Mechanics/Coarse-Grained (MM/CG) simulations have been shown to help refine ligand poses in human G protein-coupled receptors (hGPCRs). The Hybrid MM/CG Webserver takes structures of GPCR/ligand complexes (that can be generated with other GPCR-related webservers) and prepares them for running such multi-scale simulations. This approach allows the description of the ligand, the binding cavity, and the surrounding water molecules at atomistic resolution, while coarse-graining the rest of the receptor to reduce the computational cost. The system, prepared and equilibrated with the Hybrid MM/CG webserver, can be downloaded and the MM/CG simulation can be continued locally.

31.2 Others

31.2.1 Guide to Pharmacology

Founded in 1959 as a section of the International Union of Physiological Sciences, the International Union of Basic and Clinical Pharmacology (IUPHAR) has been independent since 1966. IUPHAR is a member of the International Council for Science (ICSU) and participates in the work of its scientific committees. It receives international recognition, particularly by the United Nations Educational, Scientific and Cultural Organization (UNESCO).

31.2.2 DrGPCR

The Dr. GPCR Ecosystem is a website aiming to connect the general GPCR community through various initiatives such as Cafe and Connect events to promote networking, product and reagent knowledge sharing, as well as talking science and business. Similarly, the GPCR Club on LinkedIn and the very first Podcast dedicated to GPCRs are meant to share information, learn the latest and greatest on GPCR research. Dr. GPCR also seeks to connect GPCR Consultants with clients and generate cohesion in the international GPCR community at large.

31.2.3 GPCRladies

GPCRladies is a list of women scientists working in the field of G-protein coupled receptors (GPCRs). It was set up to help diversify and balance speakers and participants in GPCR-related conferences, boards and committees.