British Journal of Pharmacology

British Journal of Pharmacology (2016) **173** 2195–2207 2195

REVIEW ARTICLE

GPCRdb: the G protein-coupled receptor database – an introduction

Correspondence D E Gloriam, Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. E-mail: david.gloriam@sund.ku.dk

Received 19 January 2016; **Revised** 18 March 2016; **Accepted** 24 April 2016

C Munk[1], V Isberg[1], S Mordalski[2], K Harpsøe[1], K Rataj[2], A S Hauser[1], P Kolb[3], A J Bojarski[2], G Vriend[4] and D E Gloriam[1]

[1]Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [2]Department of Medicinal Chemistry, Institute of Pharmacology, Polish Academy of Sciences, Krakow, Poland, [3]Department of Pharmaceutical Chemistry, Philipps-University Marburg, Marburg, Germany, and [4]Centre for Molecular and Biomolecular Informatics, Radboudumc, Nijmegen, The Netherlands

GPCRs make up the largest family of human membrane proteins and of drug targets. Recent advances in GPCR pharmacology and crystallography have shed new light on signal transduction, allosteric modulation and biased signalling, translating into new mechanisms and principles for drug design. The GPCR database, GPCRdb, has served the community for over 20 years and has recently been extended to include a more multidisciplinary audience. This review is intended to introduce new users to the services in GPCRdb, which meets three overall purposes: firstly, to provide reference data in an integrated, annotated and structured fashion, with a focus on sequences, structures, single-point mutations and ligand interactions. Secondly, to equip the community with a suite of web tools for swift analysis of structures, sequence similarities, receptor relationships, and ligand target profiles. Thirdly, to facilitate dissemination through interactive diagrams of, for example, receptor residue topologies, phylogenetic relationships and crystal structure statistics. Herein, these services are described for the first time; visitors and guides are provided with good practices for their utilization. Finally, we describe complementary databases cross-referenced by GPCRdb and web servers with corresponding functionality.

Abbreviations

7TM, seven transmembrane; PDB, Protein Data Bank

Table of Links

TARGETS		
5-HT_{1B} receptor	CCR5	NTS_1 receptor
β_1-adrenoceptor	D_3 receptor	OX_2 receptor
β_2-adrenoceptor	FFA1 receptor	PAR1
δ receptor	H_3 receptor	$P2Y_1$ receptor
μ receptor	LPA_1 receptor	$S1P_1$ receptor
A_{2A} receptor	M_2 receptor	
AT_1 receptor	$mGlu_1$ receptor	

This table lists key protein targets in this article which are hyperlinked to corresponding entries in http://www.guidetopharmacology.org, the common portal for data from the IUPHAR/BPS Guide to PHARMACOLOGY (Southan *et al.*, 2016) and are permanently archived in the Concise Guide to PHARMACOLOGY 2015/16 (Alexander *et al.*, 2015).

DOI:10.1111/bph.13509

Introduction

GPCRs

The GPCR family comprises about 800 members in humans making it the largest membrane protein family (Venter *et al.,* 2001). A bit more than half of the GPCRs sense exogenous signals: odours, tastes, light or pheromones (Mombaerts, 2004); whereas ~350 receptors regulate a plethora of physiological processes spanning nervous and endocrine systems. Their abundance, regulation of pathophysiology in diverse disease areas, accessibility at the cell surface and druggable binding sites have made GPCRs the largest drug target family. GPCRs make up ~19% of targets for drugs on the market and are one of the largest families in clinical trials; however, the majority are still unexploited in therapies or trials (Rask-Andersen *et al.,* 2014).

The human GPCRs can be classified into six classes, and as many unique (other) receptors (Table 1). Two overlapping classification systems denote the classes A–F (Kolakowski, 1994) or by their prototypical members, glutamate, rhodopsin, adhesion, frizzled and secretin, and are based on sequence homology and phylogenetic analysis (Fredriksson *et al.,* 2003) respectively. The taste type 2 receptors were recently placed as a separate sixth class having evolved from class A (Nordstrom *et al.,* 2011). The classes are further grouped into receptor families by pharmacological classification of their endogenous ligands that span ions, neurotransmitters, lipids, carbohydrates, nucleotides, amino acids, peptides and proteins (Southan *et al.,* 2016). The pharmacological receptor families mirror the evolutionary subfamilies, with a few exceptions.

Structurally, all GPCRs share a common core of seven transmembrane (7TM) helices that form the machinery for signal transduction across the cell membrane. Classes A and B1 bind their endogenous ligands in the 7TM, which serves as a site for allosteric modulation of class B2, C and F GPCRs that instead have their orthosteric site in the extracellular domain (Lagerstrom and Schioth, 2008). Crystal structures are now available for the 7TM of all classes but B2/adhesions and T/taste2 and have revealed common conformational changes during receptor activation, allosteric modulation by ions, lipids, cholesterol, and water; as well as G protein binding (Katritch *et al.,* 2013; Venkatakrishnan *et al.,* 2013; Tehan *et al.,* 2014). This wealth of information has sparked great activity in the GPCR field to understand the structural mechanisms for signal transduction, allosteric modulation and biased signalling; and to exploit these new templates and principles for drug design.

Furthermore, as much as one third of the non-sensory GPCRs are still orphan receptors with unknown endogenous ligands and/or function (IUPHAR/BPS Guide to PHARMACOLOGY). Most of these were identified only relatively recently after the mapping of the human genome. Thus, significant work remains to be done in the elucidation of their basic pharmacology and physiology, and in the longer term, disease validation and exploitation as novel clinical targets.

The GPCR database, GPCRdb

The GPCR database, GPCRdb, is well-established in the GPCR community with more than 20 years of service and over a thousand citations (Horn *et al.,* 1998; Horn *et al.,* 2003; Vroling *et al.,* 2011; Isberg *et al.,* 2014, 2016). GPCRdb was started in 1993 by Gert Vriend, Ad IJzerman, Bob Bywater and Friedrich Rippmann. At this time a growing number of receptor sequences were identified through gene cloning. As web browsers had not yet been introduced, the GPCRdb was originally an automated email answering system that could send sequences, alignments and homology models to its users. Over two decades, the GPCRdb has evolved to be a comprehensive information system storing and analysing data.

In 2013, the stewardship of the 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 European COST Action 'GLISTEN'. It has since been extended to increase its utility for a greater number of GPCR researches and studies. The manually annotated and derived data focus on crystal structures, sequence alignments,

Table 1

The human GPCRs can be classified into six classes, and as many unique (other) receptors

Class	GRAFS Family	Note	# Human members
A	*Rhodopsin*	Incl. 390 olfactory and 5 vomeronasal 1 receptors	689
B	*Secretin*	Also referred to as B1	15
	Adhesion	Also referred to as B2	33
C	*Glutamate*	Incl. 3 Taste 1 receptors	22
D	—	Fungal mating pheromone receptors	0
E	—	cAMP receptors	0
F	*Frizzled*	—	11
T	*Taste 2*	Previously grouped with Class F but later redefined as a separate class evolved from class A	25
O	*Other*	7TM receptors not belonging to any of the above classes	6

Two overlapping classification systems denote the classes A–F (Kolakowski, 1994) or by their prototypical members *glutamate, rhodopsin, adhesion, frizzled and secretin* (GRAFS) and are based on sequence homology and phylogenetic analysis (Fredriksson *et al.,* 2003) respectively. The taste type 2 receptors were recently placed as a separate sixth class having evolved from class A (Nordstrom *et al.,* 2011) abbreviated with T. The GPCRdb provides both classifications.

and receptor mutations; can be visualized in interactive diagrams; and form the basis of a suite of online analysis tools. Below, we introduce new users to the key GPCRdb data and analysis and visualization tools, complemented by good practice guidelines for their application.

Receptor mutations and residue diagrams

GPCRdb holds a large collection of manually annotated mutations from published studies that have served to pinpoint ligand binding sites, specific ligand-receptor interactions or the residues responsible for subtype selectivity (Beukers *et al.,* 1999; Isberg *et al.,* 2016). In the past, only minimal information was stored: the receptor, residue number, wild type and mutation amino acids and a reference. In the last year, this has been expanded to also include the effect (qualitative or quantitative) on ligand affinity or potency, as well as influence on receptor surface expression or basal activity. Thus, this resource has shifted from a source of publications to direct access to generated results and meta-comparison of multiple studies or receptors. Pharmacologists and other researchers can submit their data to generate diagrams for publication (below) and to increase dissemination. The major considerations for user-based data deposition are that to receive sufficient data the submission has to be fast, while its utility requires all critical data to be captured in a uniform representation. In the GPCRdb, mutation data are submitted in an excel file with one mutation per row, described using standardized controlled vocabulary terms.

The mutation data are described in three ways. (i) The mutation browser mirrors the Excel file submitted but combines data from multiple sources and allows for sorting and filtering. (ii) The helix-box and snake-plot diagrams (Figure 1A, B) are 2D receptor topology plots that map the position of binding site residues as seen from the extracellular and membrane sides respectively. (iii) The residue table (Figure 1C) is a side-by-side comparison of mutations to their equivalent sequence/structure positions (see generic residue numbering below) across multiple receptor subtypes or species. Together, the three representations allow users to map common hotspots for ligand binding or receptor subtype selectivity, as well as to delineate specific receptor interactions for a selected ligand. All representations apply a uniform color-coding of mutants according to their fold effect on ligand binding/activity (the specific type can be seen in the browser or table download) and can be downloaded for further analysis or publication.

Good practices for the interpretation and design of mutagenesis experiments

Alanine scanning, mutation of a series of non-glycine residues to alanine, is often used to replace larger sidechains with the minimal methyl group in the identification of functionally important residues (Morrison and Weiss, 2001). However, alanine scanning has drawbacks: drastic mutations, such as from large or charged residues, are more likely to perturb the surface expression of a receptor or its basal activity. Furthermore, it does not give sufficient information to distinguish between the different molecular interactions that could be mediated by a single residue, for example Tyr can exhibit aromatic stacking, hydrophobic van der Waals contacts, (π)–cation interactions or serve as a hydrogen bond acceptor or donor. Hence, a more ideal strategy is to make the most conservative mutation, while removing one molecular functionality at a time.

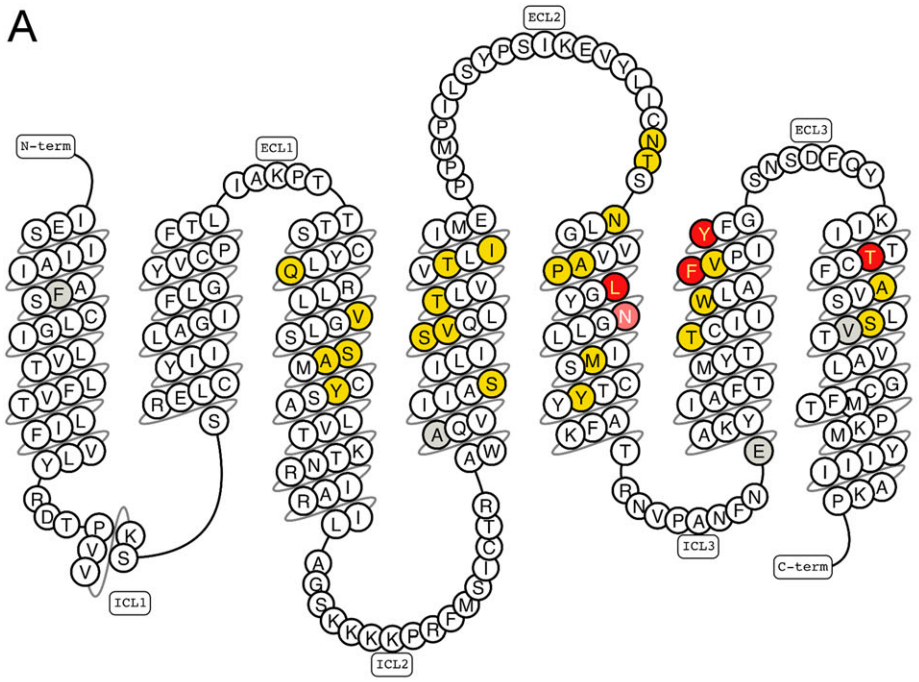
When locating an unknown site, it is desirable to cover many parts of the receptor pocket; optionally, this could be achieved with fewer mutants by mutation to larger sidechains that block ligand binding. Prioritization is typically towards the residue mutations that are expected to have the strongest direct effect. Ligand affinity is expected to decrease more upon removal of stronger interactions, that is, in the order of charged, polar, aromatic and van der Waals contacts. Furthermore, the charged and polar interactions provide higher resolution as they are more spatially defined (distances and angles), and the hydrophobic interactions, especially aromatic stacking, more often give rise to indirect effects.

In delineation of a specific receptor–ligand interaction, it is typically desired to generate complementary mutants for a residue. For example, Tyr to Phe mutation removes only the hydrogen bonding hydroxyl functionality, while a Tyr to Leu mutant also removes the aromaticity. However, there are special cases that warrant broader strategies: less conservative mutants are used in selectivity studies wherein subtype-specific residues are exchanged, and when several proximal residues can interchangeably mediate the same ligand interaction, double or even triple mutants may be required.

For convenience a quick reference guide has been collated in the GPCRdb documentation that tabulates 'good practice' mutations for all of the 20 amino acids and their primary interaction types. Furthermore, a beta-version is available as an online tool to design new mutagenesis experiments based on previously annotated mutations, ligand interactions extracted from crystal structures and, optionally, a user-uploaded receptor–ligand complex in Protein Data Bank (PDB) format.

Structure statistics, browsing and processing

In recent years, technological breakthroughs (Salon *et al.,* 2011) have led to an exponential increase in the number of GPCR crystal structures. The first structure of a ligand-activated receptor, the β_2-adrenoceptor, was presented in 2007 by the Nobel laureate Brian Kobilka and today, more than 142 structures have been reported for 73 unique GPCRs. These have revealed the molecular sites and mechanisms for ligands, lipids, G proteins, water networks as well as conformational changes upon receptor activation (Katritch *et al.,* 2013; Venkatakrishnan *et al.,* 2013; Tehan *et al.,* 2014). Thus, GPCR researchers are presented with a range of receptors and complexes, and alternative templates may be selected in studies of, for example, ligand binding, allosteric modulation, signal transmission, signal protein activation and biased signalling. As indicated by the three community-wide 'GPCR Dock' assessments, the increasing number of structural templates has led to higher precision in homology modelling (Michino *et al.,* 2009; Kufareva *et al.,* 2011; Kufareva *et al.,* 2014).



C

GPCRdb(A)	GPCRdb(C)	mGlu$_1$ receptor Human	mGlu$_1$ receptor Rat	mGlu$_1$ receptor Mouse	mGlu$_1$ receptor C. elegans
TM1					
1x42	1.46x46	F599	F599	F599	L688
TM3					
3x28	3.32x32	Q660	Q660	Q660	K749
3x32	3.36x36	V664	V664	V664	I753
3x36	3.40x40	S668	S668	S668	F757
3x37	3.41x41	A669	A669	A669	S758
3x40	3.44x44	Y672	Y672	Y672	Y761
TM4					
4x41	4.31x31	A705	A705	A705	S790
4x47	4.37x37	S711	S711	S711	A796
4x51	4.41x41	S715	S715	S715	G800
4x52	4.42x42	V716	V716	V716	V801
4x55	4.45x45	T719	T719	T719	I804
4x59	4.49x49	T723	T723	T723	I808
4x61	4.51x51	I725	I725	I725	L810
TM5					
5x37	5.37x37	N750	N750	N750	D835
5x42	5.42x42	A755	A755	A755	Y840
5x43	5.43x43	P756	P756	P756	S841
5x44	5.44x44	L757	V757	V757	L842
5x47	5.47x47	N760	N760	N760	D845
5x52	5.52x52	M765	M765	M765	V850
5x56	5.56x56	Y769	Y769	Y769	T854
TM6					
6x33	6.35x35	E783	E783	E783	E868
6x44	6.46x46	T794	T794	T794	T879
6x48	6.50x50	W798	W798	W798	W883
6x51	6.53x53	F801	F801	F801	W886
6x52	6.54x54	V802	V802	V802	I887
6x55	6.57x57	Y805	Y805	Y805	F890
TM7					
7x39	7.32x33	T815	T815	T815	L904
7x42	7.35x36	A818	A818	A818	S907
7x46	7.39x40	S822	S822	S822	S911
7x47	7.40x41	V823	V823	V823	A912

Increased binding/potency: >5-fold, >10-fold; Reduced binding/potency: >5-fold, >10-fold; No/low effect (<5-fold); and N/A

Figure 1

(A) Snake and (B) helix box diagrams depict the receptor topology as seen from the side and above respectively. (C) Residue Tables give a side-by-side comparison of receptor subtype residues lined up by their common generic residue number. Figure A covers the overall and B–C the 7TM sequence, respectively; the first two are a consensus representation and the latter all metabotropic glutamate 1 (mGlu$_1$) receptor species orthologues in the GPCRdb. The colour scheme indicates the fold effect of mutation on ligand binding, as described in the label. The same residue diagrams and tables can also be used to highlight ligand interactions from crystal structure complexes and residue physico-chemical properties.

An overview is provided as structure statistics for all GPCR entries within the PDB (Rose *et al.*, 2015). Bar diagrams can show either the increase in structures over the years or a break down according to endogenous ligand types, for example, amine, lipid and peptide receptors (Figure 2). A tree shows the structural coverage of all human receptors, except for the Class A orphan receptors. Crystallized members are flagged, highlighted with a red circle (Figure 3).

Structure retrieval is facilitated by a structure browser allowing for sorting and filtering based on receptor class, crystallized and endogenous ligands, signalling protein, resolution and completeness. Users that wish to identify the most suitable template for a certain target may use the template selection tool. This is identical to the browser, but has a preceding step to select a reference target of interest and adds its sequence similarity to crystallized receptors.

Processing of multiple structures can be performed by superposition on any substructure, that is, the whole, a segment (e.g. one or more transmembrane helices), or a set of residues

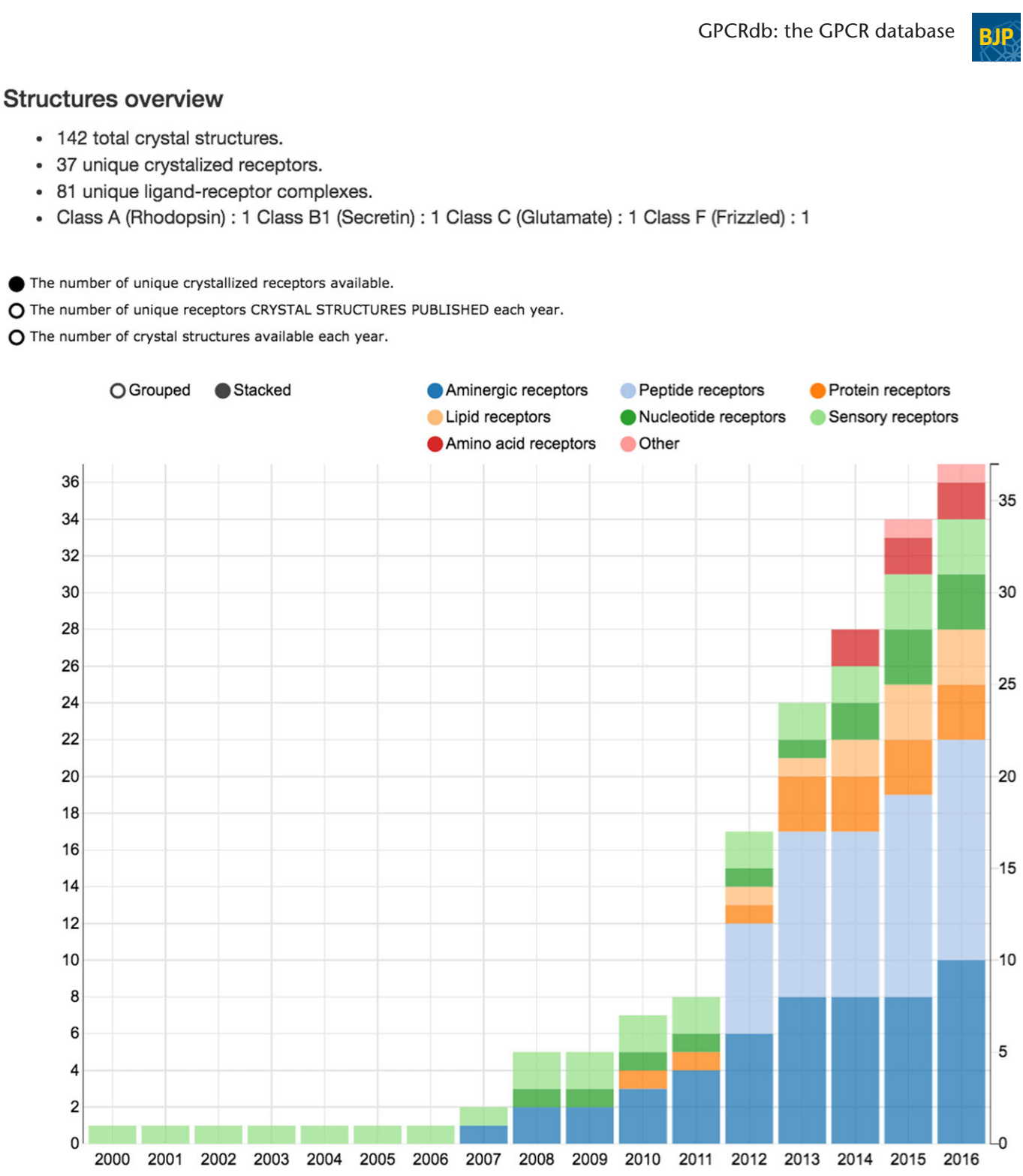


Figure 2

Bar diagrams on the structure statistics page plot the number of unique or total crystallized GPCRs in the PDB by year, and the colours indicate their type of endogenous ligand, such as amine, lipid and peptide receptors.

defined by generic numbers (below). The latter gives the unique option to focus the comparison on a functional domain, such as binding sites of ligands, signalling proteins and dimers, or microswitches for receptor activation. Users may download a batch of receptor structures or substructures to continue the analysis on a local computer.

Good practices for selection of a structural template

When the primary interest is in the receptor itself, the standard procedure is typically to sort receptor templates by their sequence identity/similarity and resolution. Crystal structures are now available for at least one receptor in the

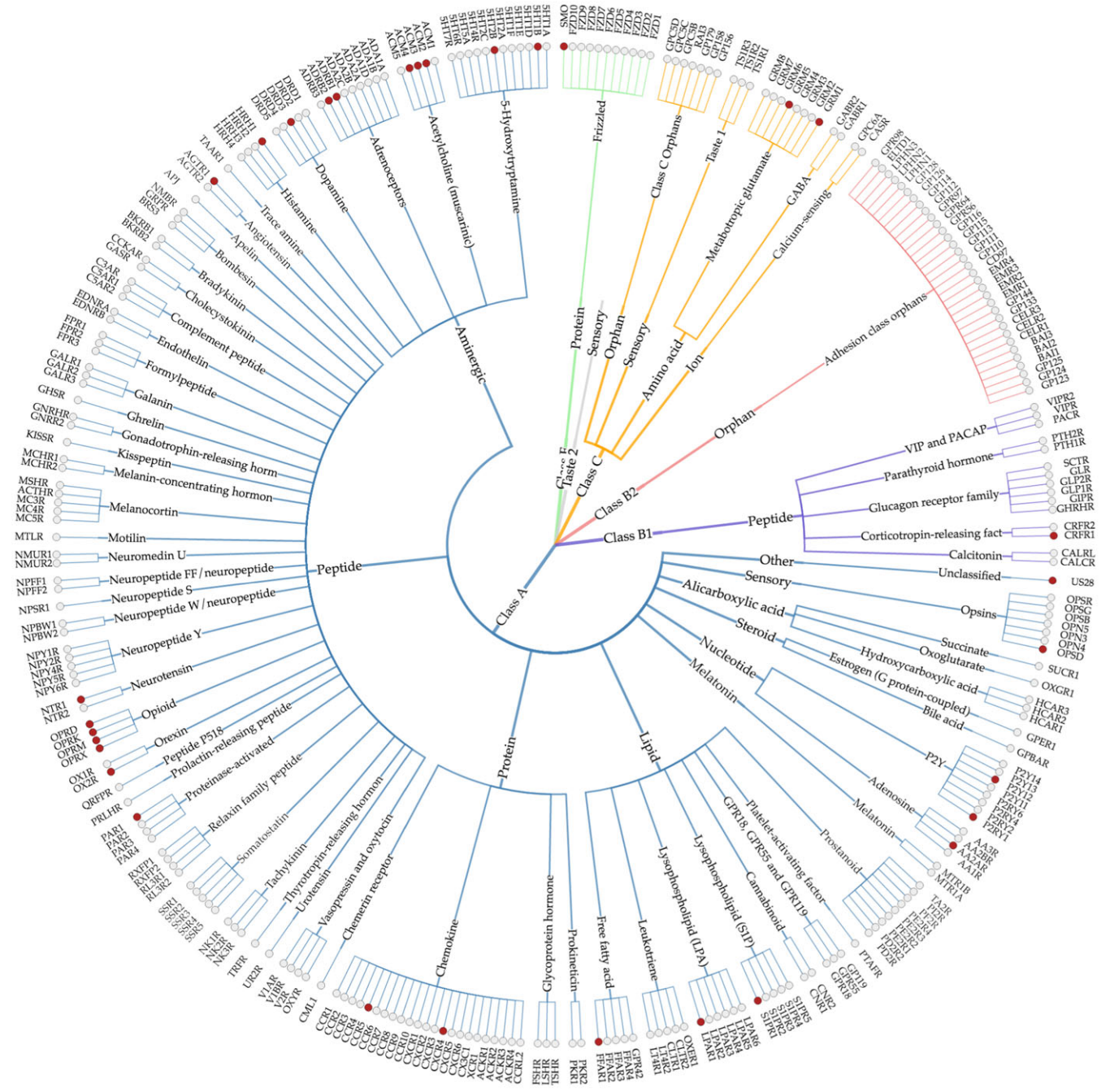


Figure 3

Tree from the structure statistics page depicting crystallized receptors with a red circle. The tree can be navigated from its centre on the levels of class, (endogenous) ligand type, receptor family and receptor. The receptor families are listed as defined by the nomenclature committee of the International Union of Pharmacology (Southan *et al.,* 2016), whereas gene names are used in order to fit all receptors.

major classes A–C and F of human GPCRs, although representative structures are still missing for the adhesion and taste 2 receptors, which although related to classes A and B, respectively, in the GRAFS are considered to have evolved into separate families (Nordstrom *et al.,* 2011). A template from the same overall family provides sufficient conservation of the structure, that is, conformation of the 7TM bundle, and sequence, that is, the minimum similarity and conserved motifs, needed to make a correct template–target alignment. The 7TM backbone displays very modest movement upon only agonist-binding, but significant changes upon full activation through simultaneous binding of a signalling protein. However, structures of fully activated receptors are so far limited to complexes of β_2-adrenoceptor to Gs (Rasmussen *et al.,* 2011), opsin (activated rhodopsin) to β-arrestin (Kang *et al.,* 2015), or a G_t α-subunit C-terminus (Scheerer *et al.,* 2008), and the μ

opioid receptor in complex with a nanobody G protein-mimic (Huang *et al.,* 2015).

In drug design, special consideration is given to the ligand structure and activity, as these have a large influence on the rotamers of contact residues and may give some precedence over receptor template–target sequence similarity. An agonist study may also apply an antagonist-bound template if the structural similarity of the two ligands is higher than to the closest agonist complex. Antagonist-bound receptors typically also have a slightly larger binding cavity, which eases ligand docking when sidechain rotamers are not conserved and thus modelled with lower precision. Users may filter the available structures based on the above parameters in the GPCRdb structure browser. When multiple structures are available for the same receptor, it is possible to retrieve a representative, which is selected based on resolution, completeness (number of intact segments) and integrity (soundness of the obtained coordinates).

Sequence alignments and similarities

The GPCRdb contains a manually curated 7TM sequence alignment of all human non-olfactory receptors extended by automatic alignment of all species orthologues in Swiss-Prot and TrEMBL (>18.000). Custom alignment retrieval allows for selective extraction of any subset of receptors and sequence segments: TM helices, termini, loops or individual residues (defined by generic residue numbers). Predefined sets are available for crystallized receptors and for ligand-accessible residues in Class A GPCRs (Gloriam *et al.,* 2009). When assessing sequence similarities, it is often interesting to look at the local conservation of a particular amino acid or property. To this end, each aligned position has a consensus sequence, followed by the percentage conservation of the 20 amino acids, as well as residue properties, such as aromaticity, charge and hydrogen bonding ability (Figure 4).

To facilitate optimal receptor comparisons, the GPCRdb alignments are structure-based. This means that residues are aligned in sequence only if they are positioned in the equivalent position in the receptor structures – as defined by superposition of the two most homologous crystal structures. The structures are used to manually annotate the start and stop of each helix, helix 8, as well as some conserved loop segments; and their relative alignment between receptors, especially where there is a lack of conserved sequence motifs. Furthermore, the alignments are assigned a single-residue gap where a helix bulge or constriction has caused a gap in the structure, as described under generic residue numbers below (Isberg *et al.,* 2015).

Sequence similarity is often used to deduct protein homology and functions. The GPCRdb can be queried using BLAST (Johnson *et al.,* 2008), and a specialized similarity search that utilizes custom segment selections (above) of the reference structure-based alignments. The sequences of a receptor selection, such as receptor family subtypes, can be plotted in a similarity matrix that lists all pairwise sequence identities and similarities. Receptor relationships may be visualized in phylogenetic trees that are generated on the fly for custom receptor and segment selections (Felsenstein, 1989). The trees can be shown in circular and ladder representations, or downloaded for visualization in other software. Annotation of GPCR class, ligand type and receptor family can be added next to the receptor labels, making it easy to correlate them to the sequence-based groups. The names of crystallized receptors are highlighted with a blue background, and any background colour can be assigned to indicate custom labelling schemes.

Good practices for receptor similarity and relationship analyses

Alignments constitute the core of all sequence-based comparisons (e.g. receptor evolution, ligand selectivity, activation microswitches/domains), inferences (e.g. prediction of orphan receptor agonists and function); and interpretation of experimental data (e.g. mutagenesis experiments). Purely sequence-based alignments might be better suited in evolutionary analyses where one-codon insertion or deletion should not be gapped, whereas structure-based alignments are made to ensure that residue positions are structurally equivalent making them more ideal for sequence/structure–function studies.

Likewise, phylogenetic analyses based on complete alignments are suitable for evolutionary and general receptor studies, whereas a specific receptor (sub-)function may be separated by grouping in trees based on only the underlying sequence/structure site. For example, a tree based on a particular ligand binding site may be more applicable to rationalize selectivity, off-targets, polypharmacology or to predict the endogenous ligands of orphan receptors. When a particular receptor is the sole interest of the study, it is preferable to run a similarity search instead of a phylogenetic analysis, as this will yield numeric similarity measurements and reveal similarity to receptors from multiple groups/branches.

Generic residue numbers for equivalent receptor residues

To facilitate receptor residue comparisons, schemes have been developed that assign a generic residue number to a given sequence/structure position (Isberg *et al.,* 2015). A generic GPCR residue number is composed of two numbers: firstly the transmembrane helix, 1–7, and secondly the residue position relative to a reference, number 50, which is the most conserved position (within the given class A–F). For example, 5.42 is the position in TM5 that is located eight residues before the most-conserved residue, Pro5.50. GPCRdb also provides generic numbers for structurally ordered segments within the first two intracellular and extracellular loops respectively. Generic numbers are often combined with receptor-specific numbers by placing one of them in superscript, for example, $S348^{5.47}$ or $S5.47^{348}$.

In 2015, the GPCRdb, NC-IUPHAR, Heptares and Raymond Stevens lab published the first structure-based generic numbering scheme for GPCRs (Isberg *et al.,* 2015). The GPCRdb numbers are an evolution of the most advanced scheme, that of Ballesteros and Weinstein (Ballesteros and Weinstein, 1995), but also provides a solution for a limitation

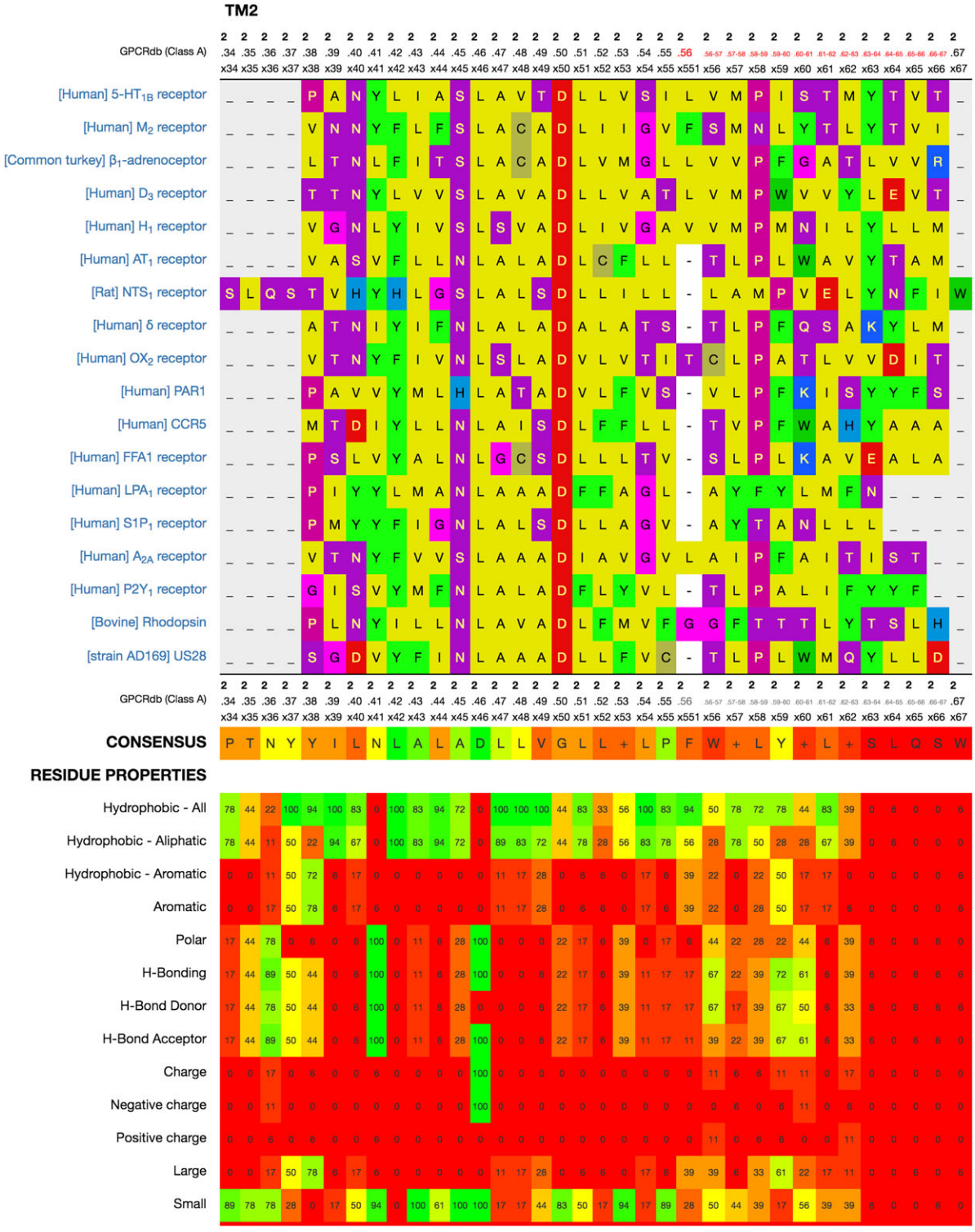


Figure 4

Structure-based sequence alignment of the transmembrane helix 2 in representative (the first in each receptor family) crystallized class A GPCRs, followed by the consensus sequence, and statistics on residue and property conservation. Of note, the GPCRdb numbers are structure-based and take into account the bulge in position 2.56x551. This avoids the offset seen in the sequence-based (here Ballesteros and Weinstein) numbers of gapped versus un-gapped receptors and is in agreement with two highly conserved flanking Asp (D) and Pro (P) residues.

of the sequence-based schemes that only became evident in the structural era. Specifically, GPCR structures contain frequent helix bulges and constrictions that offset the alignment and numbering. The GPCRdb scheme solves this by placing gaps in 7TM helix alignments (according to structure superposition). The single bulge residue that protrudes the furthest is assigned the same number as the preceding residue followed by a 1, for example, a bulge after residue 46 is given the number 461. The position lacking in a constriction is simply skipped in the residue numbering. To distinguish the GPCRdb scheme, it uses a unique separator, x (e.g. 5x46), to denote that it is based predominantly on X-ray structures.

With the recent structural templates, it also became possible to make correct alignments between the GPCR classes. Such comparisons across classes use the class A numbers as the common key and optionally append that of the other class, for example Y7.53a.57b. Reference cross-class alignments, based on the crystal structures available, are present in the GPCRdb.

Most researchers are expected to use generic numbers without having to think about their generation by simply retrieving them from the GPCRdb. The reference sequence alignments (Figure 4) contain the structure-based GPCRdb numbers, as well as the sequence-based number for the given classes. Residue look-up tables (Figure 1C) list receptor-specific and generic residue numbers side-by-side facilitating swift translation. Generic residue numbers can also be assigned to GPCR structures selected from the structure browser or uploaded as PDB files. The numbers are stored in the backbone B-factors, and scripts are provided for visualization in the most popular software.

Binding sites – ligand interactions, site-based target profiling and pharmacophores

Users may select any of the GPCR-ligand structure complexes from the PDB or upload a receptor model containing a docked ligand to retrieve information about its ligand interaction residues and types. The predefined binding sites from PDB receptor–ligand structure complexes currently cover 828 ligand interactions from 32 receptors with 80 ligands. The ligand, receptor residues and interaction types can be browsed and visualized in schematic 2D and interactive 3D diagrams (Figure 5). As with the mutations, the topologies of binding residues are illustrated in helix box and snake diagrams (Figure 1A, B), and compared in residue tables (Figure 1C) across receptor families and subtypes.

A binding site or sub-site may be shared by more than one receptor; proteins that share a given site are also expected to share its associated function. For this reason, target profiling is key to the investigation of selectivity issues, desired polypharmacology, and ligand inference. By uploading a receptor–ligand complex in PDB format, researchers can conduct a site search against the GPCRdb reference alignments to retrieve the matching receptors. Each contact residue position is also allowed to match other amino acids that are able to mediate the same molecular interaction, for example, hydrogen bond donor or aromatic stacking (this is different from standard evolutionary amino acid substitution matrices). The definition of residue positions and their allowed amino acids can be manually fine-tuned, for example to focus the search on a ligand fragment. Of note, the site can be manually defined and then applied to any (aligned) receptor site, for example, the binding sites of G proteins or dimers, or structural microdomains stabilizing an (in)active receptor conformation (Congreve *et al.*, 2011).

Pharmacophores have a widespread use in drug design for ligand identification through virtual screening, and in the later optimization of leads (Leach *et al.*, 2009). A pharmacophore represents the 3D map of shared chemical functionalities (charge, aromatic, hydrogen bonding, etc.) across different ligand chemotypes that interact with complementary residues within the biological target. The GPCRdb allows users to upload a receptor PDB file to automatically retrieve ligand fragments, which are placed by superposition of their (conserved) interacting receptor residue (Fidom *et al.*, 2015). The fragments, each consisting of one such moiety and receptor residue, have been previously extracted from all GPCR ligand-receptor

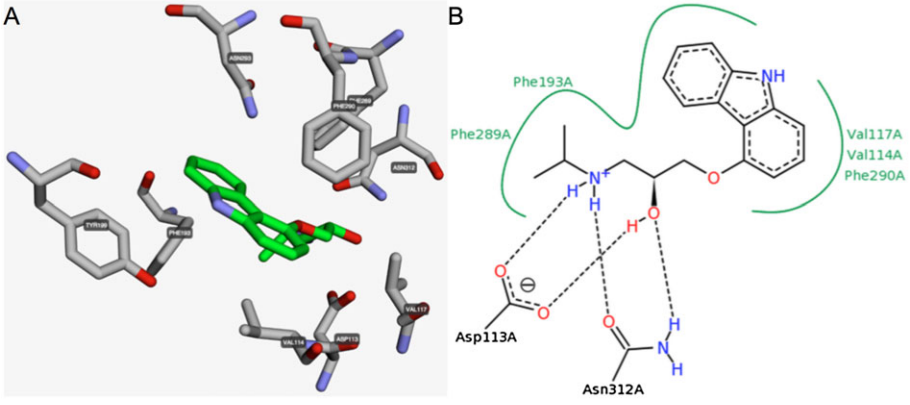


Figure 5

Users may select any of the GPCR–ligand structure complexes from the PDB or upload a receptor model containing a docked ligand to visualize the molecular interactions in (A) interactive 3D and (B) schematic 2D diagrams.

complexes in the PDB. The download provides the option to select between all ligand fragments, which are used for a complete sampling, or one representative (from the most homologous receptor), which can be used to place pharmacophore elements directly in an external software (Figure 6).

Good practices for definition of ligand interactions, site-based target profiling and pharmacophore construction

The definition of receptor–ligand interactions often contains borderline cases, and the GPCRdb has prioritized incorporation of the reliable overall data points. The definitions of molecular interaction distances and angles are provided in the GPCRdb documentation. Some ambiguity is also introduced in the fit of the molecular structure to the electron density, especially at lower resolutions. Thus, structure-based ligand design that relies heavily upon a specific structure often involves expert assessment of its electron density to explore alternative configurations of binding site residues.

The site search is a relatively new method but is the preferred target-profiling technique for well-defined sites. It offers higher precision by focusing on only the interacting residue positions and matches the amino acids that can mediate the same molecular interaction. Search on a large and selective ligand should yield few off-targets, unless it exhibits polypharmacology, whereas the querying of fragments and leads results in broader receptor profiles. Extra care needs to be taken in cases where more than one proximal residue can mediate the same molecular interaction to the ligand. These residues could be identified from the receptor–ligand complex and appended to the interaction group in the site definition.

The GPCRdb pharmacophore method is also new, in that it builds on fragments inferred from other receptor structures. A unique advantage is that it can be applied for targets lacking a good structure template or known ligands. The pharmacophores have experimental support from previous interactions but are not as complete as an overall model of the receptor binding cavity. The placement of pharmacophore elements can be done on either the single representative fragments, from the most homologous receptor, or in the centre of the density of all available fragments. As a rule-of-thumb the first should offer a fast and relevant placement, whereas the second can be checked to correct for outliers or provide complementary versions of the pharmacophore.

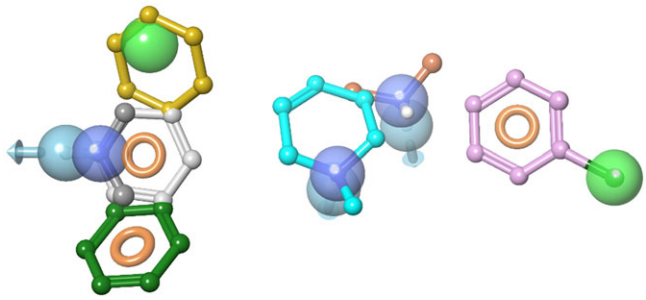


Figure 6

Histamine H_3 receptor pharmacophore constructed based on ligand fragments inferred from GPCR–ligand complex crystal structures. Ligand fragments can be automatically matched and superimposed in the GPCRdb. Pharmacophore elements: orange, aromatic; green, hydrophobic; blue, cationic; and light blue, hydrogen bond donor, were here assigned with Phase (Dixon *et al.,* 2006).

GPCRdb development focus and integration with GPCR communities

Annotation and development focus

The GPCRdb team develops data browsers, analysis tools and visualization diagrams of use for their own research, while making them easily accessible to all researchers. The main challenge of database maintenance is to continuously offer data that is up-to-date, while of sufficient quality and quantity. The GPCRdb focuses the manual annotation on the core data: crystal structures, sequence alignments and receptor mutations, whereas the derived data are automatically updated. To be accepted by users, the representation and visualization of the data have to be intuitive and in agreement with community traditions and guidelines. The diagrams that are available in the GPCRdb, displaying receptor residue topologies, phylogenetic relationships and crystal structure statistics, are frequently used in GPCR publications. Many of the analysis tools in the GPCRdb are intended to make routinely conducted analyses swifter and more accessible by non-experts, such as the generation of phylogenetic trees without the need to supply own sequence alignments or install software. Other tools offer more advanced functions not available elsewhere, such as the efficient browsing of structures and site search to predict the target profile of a ligand.

For external databases and server developers, the GPCRdb offers programmatic access (REST web services) of sequence alignments, mutations, structures and residue diagrams (http://www.gpcrdb.org/services). The database can also be retrieved as a virtual machine, which facilitates development and private in-house deployment. The back-end and user interface uses a common programming language, Python, making it easier for new programmers to get started. The complete source code is freely available at https://bitbucket.org/gpcr/protwis.

Finally, specialized web servers exist for GPCR homology modelling; GPCRM (Latek *et al.,* 2013), GoMoDo (Sandal *et al.,* 2013) and SSFE (Worth *et al.,* 2011); and molecular dynamics: GPCR Mod-Sim (Rodríguez *et al.,* 2012). The GPCRdb is working together with these servers towards exchange of resources and competences.

Good practices for developing a sustainable specialized database

Specialized databases have the advantages that they can go more in-depth, and often cross-analyse more heterogeneous data types, than the general databases. Many specialized databases have appeared within the biosciences community, but time has shown that few are sustainable long-term (Attwood *et al.,* 2015). In contrast to major databases, hosted by large

infrastructure centres like NCBI and EMBL-EBI, they typically do not have long-term stable funding and depend on just one or a few developers. The GPCRdb has survived for over two decades, and has seen four different lead developers in this time. The GPCRdb was recently invited to a meeting where the heads of specialized databases came together to share ideas and discuss how their systems can serve the communities better and longer (Babbitt *et al.,* 2015). Herein, the former head of the GPCRdb Gerrit Vriend presented '10 rules for good database practices':

(i) *Longevity*: Only start a database if you know you can maintain it for at least 10 years. (ii) *Users*: To gain and keep users, the services should be prioritized based on their relevance and ease of use and have a system for answering questions. (iii) *Publish*: Target publications to the readership, with a particular focus to introduce new users. (iv) *Collaborate and be open*: Apart from extending the services and the user population, this may be what gives longevity. (v) *Give credit*: Be generous with credits. (vi) *Automate*: Everything that can be automatic should be, while strictly prioritizing the manual work to where it is needed to achieve sufficient quality of prioritized core data. (vii) *Follow community standards*: Use the recognized nomenclature, classifications and procedures. (viii) *Keep it simple*: Prioritize function and avoid too many distractions. (ix) *Visibility*: Build a network within the community, present at the relevant conferences and use the same logo throughout the years. (x) *Exit strategy*: Make a retirement plan in time to ensure a smooth transition.

Integration with GPCR communities and resources

GLISTEN (short for GPCR–ligand interactions, structures and transmembrane signalling: a European Research Network) is a network of researchers funded through the European Union's COST scheme. The network was created in May 2013 and has since attracted researchers at all levels of experience from 30 countries in Europe. Currently, the network has about 190 member labs and approximately 100 researchers meet biannually to discuss new developments and collaborations in the field of GPCRs – with a dedicated satellite meeting for the GPCRdb development team.

GLISTEN is organized into four working groups: (i) structures and dynamics; (ii) biased signalling; (iii) interactions with membrane lipids and other membrane-bound proteins; and (iv) identification of novel modulators with potential for drug design or chemical biology. In addition, one of the strongest networking tools within GLISTEN is Short-Term Scientific Missions, which allow researchers to join a different lab for up to 3 months. GPCRdb development has been greatly facilitated by such exchanges. While GLISTEN funding will cease to flow after 4 years (fall 2017), the connections that have been built during that time will continue to exist, maybe most obviously manifested in the existence and continuous development of the GPCRdb.

The IUPHAR/BPS GuideToPharmacology database (GoToPdb) holds very comprehensive information on the pharmacology of GPCRs, as well as all other major human drug target families. GoToPdb is expert-curated with target-focused committees involving more than 600 experts. The GPCRdb uses the official NC-IUPHAR receptor nomenclature and has made available its receptor residue diagrams (snakeplot and helix box) and mutation data for direct visualization in the GoToPdb. The receptor pages of the two databases are mutually cross-linked to facilitate easy browsing between the resources (for info on linking to GPCRdb, visit http://www.gpcrdb.org/pages/linking).

The GPCRdb has initiated collaboration with a GPCR Consortium initiated by the Professor Raymond Stevens lab, which has produced the largest number of GPCR structures, and involves nine pharmaceutical companies. The GPCR Consortium aims to publish the structures of an unprecedented number of validated GPCR drug targets (Parmley, 2014). The GPCRdb team, which already has ongoing joint international development of new structure database services, will contribute to the GPCR Consortium by producing tailored tools for GPCR structural biology.

Conclusions and future directions

In conclusion, the GPCRdb provides experimental and derived data, visualization diagrams and analysis tools for the wider GPCR community. The GPCRdb places focus on structures, receptor mutations, ligand interactions and offers the first structure-based sequence alignments. Complementary data types can be assigned to receptor residue positions using generic numbering and visualized within uniform residue diagrams and tables. This can help to provide a structural explanation of pharmacological effects observed, and users may submit new mutation data to facilitate comparison to the data already included. It is expected that many more structure complexes will continue to be solved, and these will provide further insights into the molecular mechanisms of GPCR activation, allosteric modulation, signal transduction, signalling protein binding and receptor dimerization. The GPCRdb is ready to aid the GPCR research community in the dissemination of these data and it will continue to be improved for the development of future services.

Acknowledgements

GPCRdb is supported by the COST Action CM1207 'GLISTEN'. D.E.G., V.I. and K.H. received financial support from the Lundbeck Foundation (R163-2013-16327) and the European Research Council (DE-ORPHAN 639125). S.M. was supported by the Polish National Science Centre (Etiuda 2). Furthermore, the GPCRdb team would also like to acknowledge all the hard work from its co-founders Ad Ijzermann, Robert Bywater and Friedrich Rippmann and long-term lead developers Florence Horn and Bas Vroling.

Author contributions

C.M. and V.I. contributed equally (shared first authors).

Conflict of interest

The authors declare no conflicts of interest.

References

Alexander SPH, Davenport AP, Kelly E, Marrion N, Peters JA, Benson HE *et al.* (2015). The Concise Guide to PHARMACOLOGY 2015/16: G protein-coupled receptors. Br J Pharmacol 172: 5744–5869.

Attwood TK, Agit B, Ellis LBM (2015). Longevity of biological databases. EMBnet.journal 21: e803.

Babbitt PC, Bagos PG, Bairoch A, Bateman A, Chatonnet A, Chen MJ *et al.* (2015). Creating a specialist protein resource network: a meeting report for the protein bioinformatics and community resources retreat. Database 2015 .bav063

Ballesteros JA, Weinstein H (1995). Integrated methods for the construction of three-dimensional models and computational probing of structure–function relations in G protein-coupled receptors. Methods Neurosci 25: 366–428.

Beukers MW, Kristiansen K, Ijzerman AP, Edvardsen O (1999). TinyGRAP database: a bioinformatics tool to mine G-protein-coupled receptor mutant data. Trends Pharmacol Sci 20: 475–477.

Congreve M, Langmead CJ, Mason JS, Marshall FH (2011). Progress in structure based drug design for G protein-coupled receptors. J Med Chem 54: 4283–4311.

Dixon S, Smondyrev A, Knoll E, Rao S, Shaw D, Friesner R (2006). PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J Comput Aided Mol Des 20: 647–671.

Felsenstein J (1989). PHYLIP – Phylogeny Inference Package (Version 3.2). Cladistics 5: 164–166.

Fidom K, Isberg V, Hauser AS, Mordalski S, Lehto T, Bojarski AJ *et al.* (2015). A new crystal structure fragment-based pharmacophore method for G protein-coupled receptors. Methods 71: 104–112.

Fredriksson R, Lagerström MC, Lundin L-G, Schiöth HB (2003). The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol 63: 1256–1272.

Gloriam DE, Foord SM, Blaney FE, Garland SL (2009). Definition of the G protein-coupled receptor transmembrane bundle binding pocket and calculation of receptor similarities for drug design. J Med Chem 52: 4429–4442.

Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G (2003). GPCRDB information system for G protein-coupled receptors. Nucleic Acids Res 31: 294–297.

Horn F, Weare J, Beukers MW, Horsch S, Bairoch A, Chen W*et al.* (1998). GPCRDB: an information system for G protein-coupled receptors. Nucleic Acids Res 26: 275–279.

Huang W, Manglik A, Venkatakrishnan AJ, Laeremans T, Feinberg EN, Sanborn AL *et al.* (2015). Structural insights into [micro]-opioid receptor activation. Nature 524: 315–321.

Isberg V, Vroling B, van der Kant R, Li K, Vriend G, Gloriam D (2014). GPCRDB: an information system for G protein-coupled receptors. Nucleic Acids Res 42: D422–D425.

Isberg V, Mordalski S, Munk C, Rataj K, Harpsoe K, Hauser AS *et al.* (2016). GPCRdb: an information system for G protein-coupled receptors. Nucleic Acids Res 44: D356–D364.

Isberg V, de Graaf C, Bortolato A, Cherezov V, Katritch V, Marshall FH *et al.* (2015). Generic GPCR residue numbers – aligning topology maps while minding the gaps. Trends Pharmacol Sci 36: 22–31.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL (2008). NCBI BLAST: a better web interface. Nucleic Acids Res 36: W5–W9.

Kang Y, Zhou XE, Gao X, He Y, Liu W, Ishchenko A *et al.* (2015). Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. Nature 523: 561–567.

Katritch V, Cherezov V, Stevens RC (2013). Structure–function of the G protein-coupled receptor superfamily. Annu Rev Pharmacol Toxicol 53: 531–556.

Kolakowski LF Jr (1994). GCRDb: a G-protein-coupled receptor database. Receptors Channels 2: 1–7.

Kufareva I, Rueda M, Katritch V, Stevens RC, Abagyan R (2011). Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. Structure 19: 1108–1126.

Kufareva I, Katritch V, Participants of GD, Stevens RC, Abagyan R (2014). Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. Structure 22: 1120–1139.

Lagerstrom MC, Schioth HB (2008). Structural diversity of G protein-coupled receptors and significance for drug discovery. Nat Rev Drug Discov 7: 339–357.

Latek D, Pasznik P, Carlomagno T, Filipek S (2013). Towards improved quality of GPCR models by usage of multiple templates and profile–profile comparison. PLoS One 8: e56742.

Leach AR, Gillet VJ, Lewis RA, Taylor R (2009). Three-dimensional pharmacophore methods in drug discovery. J Med Chem 53: 539–558.

Michino M, Abola E, Brooks CL 3rd, Dixon JS, Moult J, Stevens RC (2009). Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. Nature reviews. Drug Discov 8: 455–463.

Mombaerts P (2004). Genes and ligands for odorant, vomeronasal and taste receptors. Nat Rev Neurosci 5: 263–278.

Morrison KL, Weiss GA (2001). Combinatorial alanine-scanning. Curr Opin Chem Biol 5: 302–307.

Nordstrom KJ, Sallman Almen M, Edstam MM, Fredriksson R, Schioth HB (2011). Independent HHsearch, Needleman–Wunsch-based, and motif analyses reveal the overall hierarchy for most of the G protein-coupled receptor families. Mol Biol Evol 28: 2471–2480.

Parmley S (2014). GPCRs' grand plans. Science-Business eXchange 7.

Rask-Andersen M, Masuram S, Schioth HB (2014). The druggable genome: evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. Annu Rev Pharmacol Toxicol 54: 9–26.

Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS *et al.* (2011). Crystal structure of the beta2 adrenergic receptor-Gs protein complex. Nature 477: 549–555.

Rodríguez D, Bello X, Gutiérrez-de-Terán H (2012). Molecular modelling of G protein-coupled receptors through the web. Mol Inform 31: 334–341.

Rose PW, Prlic A, Bi C, Bluhm WF, Christie CH, Dutta S *et al.* (2015). The RCSB protein data bank: views of structural biology for basic and applied research and education. Nucleic Acids Res 43: D345–D356.

Salon JA, Lodowski DT, Palczewski K (2011). The significance of G protein-coupled receptor crystallography for drug discovery. Pharmacol Rev 63: 901–937.

Sandal M, Duy TP, Cona M, Zung H, Carloni P, Musiani F *et al.* (2013). GOMoDo: a GPCRs online modeling and docking webserver. PLoS One 8: e74092.

Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe H-W*et al.* (2008). Crystal structure of opsin in its G-protein-interacting conformation. Nature 455: 497–502.

Southan C, Sharman JL, Benson HE, Faccenda E, Pawson AJ, Alexander SP *et al.* (2016). The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. Nucleic Acids Res 44: D1054–D1068.

Tehan BG, Bortolato A, Blaney FE, Weir MP, Mason JS (2014). Unifying family A GPCR theories of activation. Pharmacol Ther 143: 51–60.

Venkatakrishnan AJ, Deupi X, Lebon G, Tate CG, Schertler GF, Babu MM (2013). Molecular signatures of G-protein-coupled receptors. Nature 494: 185–194.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG *et al.* (2001). The sequence of the human genome. Science 291: 1304–1351.

Vroling B, Sanders M, Baakman C, Borrmann A, Verhoeven S, Klomp J *et al.* (2011). GPCRDB: information system for G protein-coupled receptors. Nucleic Acids Res 39: 309–319.

Worth CL, Kreuchwig A, Kleinau G, Krause G (2011). GPCR-SSFE: a comprehensive database of G-protein-coupled receptor template predictions and homology models. BMC Bioinformatics 12: 185.