

# GPCRDB: an information system for G protein-coupled receptors

Vignir Isberg<sup>1</sup>, Bas Vroling<sup>2</sup>, Rob van der Kant<sup>3</sup>, Kang Li<sup>1</sup>, Gert Vriend<sup>3,\*</sup> and David Gloriam<sup>1</sup>

<sup>1</sup>Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark, <sup>2</sup>Bio-Product B.V., Castellastraat 116, 6512 EZ, Nijmegen, The Netherlands and <sup>3</sup>CMBI, NCMLS, Radboudumc Nijmegen Medical Centre, Geert Grooteplein Zuid 26-28, 6525 GA, Nijmegen, The Netherlands

Received October 16, 2013; Revised and Accepted November 11, 2013

## ABSTRACT

For the past 20 years, the GPCRDB (G protein-coupled receptors database; <http://www.gpcr.org/7tm/>) has been a 'one-stop shop' for G protein-coupled receptor (GPCR)-related data. The GPCRDB contains experimental data on sequences, ligand-binding constants, mutations and oligomers, as well as many different types of computationally derived data, such as multiple sequence alignments and homology models. The GPCRDB also provides visualization and analysis tools, plus a number of query systems. In the latest GPCRDB release, all multiple sequence alignments, and >65 000 homology models, have been significantly improved, thanks to a recent flurry of GPCR X-ray structure data. Tools were introduced to browse X-ray structures, compare binding sites, profile similar receptors and generate amino acid conservation statistics. Snake plots and helix box diagrams can now be custom coloured (e.g. by chemical properties or mutation data) and saved as figures. A series of sequence alignment visualization tools has been added, and sequence alignments can now be created for subsets of sequences and sequence positions, and alignment statistics can be produced for any of these subsets.

## INTRODUCTION

G protein-coupled receptors (GPCRs) constitute a large family of cell surface receptors. They regulate a wide range of cellular processes, including those associated with taste, smell and vision, and they control myriad intracellular systems, ranging from neurotransmission to hormone signalling. GPCRs are major targets for the

pharmaceutical industry, as reflected by the fact that more than a quarter of all FDA-approved drugs act on a GPCR (1). At present, only ~30 of the ~350 genes that code for non-olfactory receptors in the human species (2) are truly validated therapeutic targets (3), indicating this family's immense potential for future drug development. An increasing number of drugs have been found to display polypharmacology, i.e. activity through multiple receptor targets (4). However, endogenous ligands for ~135 of the so-called orphan receptors have so far eluded researchers.

Early releases of the GPCRDB (5–8) focused on the compilation and homogeneous presentation of many types of heterogeneous data, with the aim of providing the four main facilities needed in an information system: browsing, querying, retrieval and inference. The first three of these facilities have been available ever since the start of the project, but received a major boost when the GPCRDB was coupled to an intelligent PDF reader (9) that puts all relevant aspects of the GPCRDB non-intrusively in a side-bar in the PDF reader window. Inference really only started when a number of interactive tools were added (10), enabling bioinformaticians to interact with multiple sequence alignments, together with derived data such as entropy and variability scores, in an integrated environment. For example, these tools were successfully applied in the 2010 GPCR-Dock competition (11). In the past, computing facilities were aimed at expert GPCR bioinformaticians. In contrast, the new interactive tools are readily accessible to non-expert users and allow faster execution of visualization and analysis tasks.

## STRUCTURAL DATA AND TOOLS

### Crystal structure browser

In recent years, X-ray crystallography of GPCRs has revealed the sites and mechanisms for binding ligands, lipids, and G proteins, as well as the conformations of

\*To whom correspondence should be addressed. Tel: +31 24 3619390; Fax: +31 24 3619395; Email: vriend@cmbi.ru.nl, Gerrit.Vriend@radboudumc.nl

activated states. The GPCRDB structure browser includes manually annotated key data for all ligand–receptor complexes and at least one (the highest resolution) representative file for receptors solved only as apo structures. The structure browser can select crystal structures based on a series of filters for receptors, ligands, activation states, G protein presence, PDB and Pubmed identifiers, resolution, structure completeness and so forth. Sequence similarities can be retrieved by specifying a receptor reference, facilitating template selection for homology modelling.

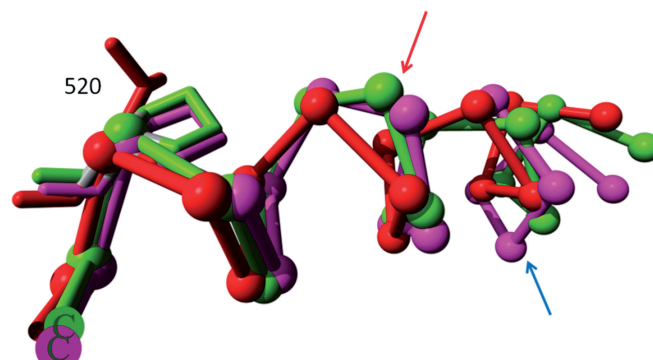
### Structure-based sequence alignments and homology models

GPCR transmembrane (TM) helices are known to contain many irregularities. However, it was only recently realized, with the availability of new crystallographic data, that GPCRs contain many more  $\alpha$ -bulges than those in helix II and helix V. The  $\alpha$ -bulges in GPCRs are relatively non-conserved. As an example of this non-conservation, Figure 1 shows the  $\alpha$ -bulges in helix V of rhodopsin and the adenosine 2A receptor, and Figure 2 shows the alignment of the middle part of helix II in 60 trace amine receptors. The receptors are on average 70% sequence identical, but the bulge is only present in around half of all family members.

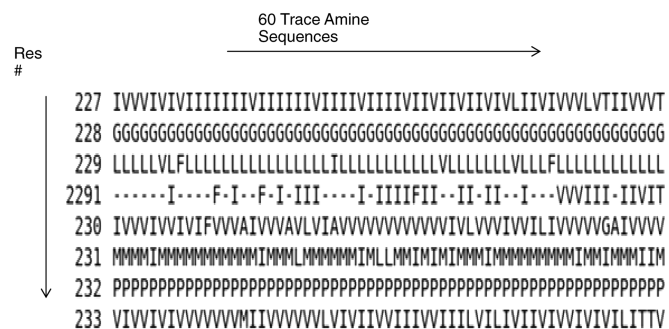
The abundance of  $\alpha$ -bulges required a novel residue numbering approach that involved rewriting the DSSP secondary structure analysis software (15) and the GPCR-specific alignment software. This, in turn, meant that >65 000 homology models had to be constructed again. All GPCR sequences are now modelled twice—once using as template the most sequence-similar inactive form, and once using the most similar active form. Modelling was done with YASARA (16), aligning the model with the template contained in the GPCRDB and using default values for all other parameters and options. All GPCR alignment profiles were manually updated to reflect our latest knowledge about the  $\alpha$ -bulges, and all 1272 alignments were regenerated. It seems likely that this exercise will need to be repeated in the coming years as new GPCR structure data reveal new  $\alpha$ -bulge patterns.

### Translation of generic and receptor-specific residue numbers

We wanted to take proper care of the  $\alpha$ -bulges that are widely present in six of the seven GPCR TM helices without excessively changing the commonly used generic residue numbering schemes. The Oliveira numbering (17) and the B&W numbers (18) were maintained as far as possible, whereas bulge residues were given the same number as the residue directly N-terminal in the sequence but with a digit added that reflects the number of the bulge. The Utopia-GPCRDB PDF reader automatically takes these new numbers into account, whereas our new sequence indexing tool provides computational access to all generic and receptor-specific residue numbers for selected receptors.



**Figure 1.** The area around the bulge in helix V. The SIP lipid receptor [red, PDBid = 3v2y (12)] does not have  $\alpha$ -bulges in helix V and is provided as a reference. Rhodopsin [green, PDBid = 1f88 (13)] and the adenosine-2A receptor [purple, PDBid = 3eml (14)] have an  $\alpha$ -bulge (red arrow) between positions 516<sup>5,46</sup> and 517<sup>5,47</sup>. The adenosine-2A receptor has an extra bulge (blue arrow) between positions 511<sup>5,41</sup> and 512<sup>5,42</sup>. Rhodopsin and the adenosine-2A receptor have a proline at position 520<sup>5,50</sup>. The SIP lipid receptor, which does not have bulges in helix V, does not have a proline at position 520<sup>5,50</sup>. Time will tell whether this correlation is accidental or causal. Residues are numbered using the GPCRDB scheme with the B&W numbers given as superscripts.



**Figure 2.** Alignment of the middle part of TM helix II in 60 trace amine receptors. The residues 227<sup>2,57</sup>–233<sup>2,63</sup> (GPCRDB numbers with B&W numbers as superscripts) are shown running vertically using the GPCRDB numbering scheme.

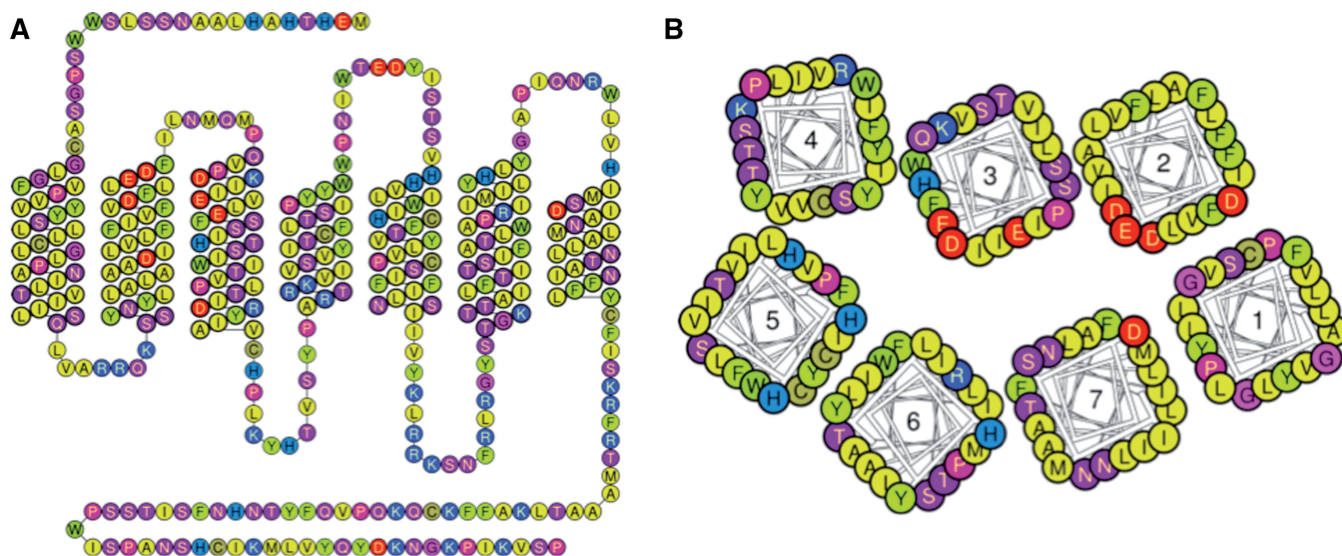
## RECEPTOR SEQUENCE DIAGRAMS

### Snake-like and helix box diagrams

Snake-like diagrams now include the full loops and terminal sequences (Figure 3A). New helix box diagrams present the TM helices as seen from ‘above’ (Figure 3A). These diagrams are similar to previously used helical wheel plots but orient the amino acids in better agreement with the 3D structures. Hovering the computer mouse pointer over the TM amino acids displays their residue numbers. Amino acids can be coloured to illustrate their physicochemical properties, or the presence of mutation data or the mutation effects. The diagrams can be downloaded as picture files or in scalable vector graphics format to allow further editing.

### Residue conservation and property statistics

In functional or evolutionary studies of specific amino acids or their properties, it is typically relevant to find



**Figure 3.** Snake-like (A) and helix box (B) diagrams depict GPCRs from the side and top, respectively.

out which additional receptors share the conserved amino acids or properties (e.g. specific charge or hydrophobicity). Therefore, sequence alignments are augmented with a series of statistics. For each position, the following data are listed: the consensus sequence, the percentage of each of the 20 amino acids present, and percentages for relevant properties such as aromaticity, acidity or hydrogen bonding capability.

### (BINDING) SITE-SPECIFIC RECEPTOR SIMILARITIES

#### Alignment sub-site/domain selections

The alignment and similarity tools integrated into GPCRDB offer users the ability to select arbitrary combinations of helices, residue positions or predefined sets of residues, for example, the amino acids in the TM binding cavity (19). By focusing on a given functional site rather than the full sequence, the receptor similarities will better reflect the structural features involved in, for example, receptor dimerization, ligand binding or G protein binding.

#### Similarity search with a reference receptor (one-to-all similarities)

Similarity searches are conducted by specifying a target of interest, a set of receptors and the residue positions of interest. Results are presented as a sequence alignment, in which the target is followed by a list of hits in order of sequence identity, similarity or alignment score. The data can be downloaded as either an alignment file or a spreadsheet.

#### Trees (all-to-all similarities)

Neighbour-joining trees (20) can be generated based on any sub-site/domain and set of receptors. Trees can be calculated with up to 100 bootstraps, displayed in

circular and ladder representations, and downloaded in Newick format for use with stand-alone tree software.

### Sequence motif search (conserved and non-conserved separation)

The sequence motif search tool generates more precise and discriminative results, by allowing residues to be matched for relevant amino acid properties (21), e.g. their hydrophobicity, hydrogen bond donor capability or size. Relevant applications for this tool include rationalization of observed polypharmacology, receptor panel selection for off target screening and ligand inference from old to new targets.

### CONCLUSIONS

The 20th yearly release of the GPCRDB includes a large number of novel discoveries. The solved structures (see <http://gpcr.scripps.edu/>) reveal the presence of many  $\alpha$ -bulges that are not conserved among or even within GPCR subfamilies. We have updated all the alignments and homology models, together with the residue numbering schemes, to ensure agreement between the contents of the GPCRDB and new insights obtained by studying all the available structure data. Additionally, the new GPCRDB release includes a powerful yet user-friendly computational toolbox that provides users with crystal structure browser, receptor visualization and alignment analysis tools, plus options to study receptor similarity both quantitatively and graphically.

### ACKNOWLEDGEMENTS

Andrius Senulis, Frantisek Sudzina and Magnus Vesterager Larsen are acknowledged for programming contributions and Kimberley Fidom for the annotation of GPCR crystal structures.

## FUNDING

Lundbeck Foundation [R54-A5441 to V.I.]; the Carlsberg Foundation [R77-A6854 to D.G.] and the European Union NewProt project [289350 to G.V.]. Funding for open access charge: The publication will be paid by the CMBI running costs.

*Conflict of interest statement.* None declared.

## REFERENCES

- Overington, J.P., Al-Lazikani, B. and Hopkins, A.L. (2006) How many drug targets are there? *Nat. Rev. Drug Discov.*, **5**, 993–996.
- Davenport, A.P., Alexander, S.P., Sharman, J.L., Pawson, A.J., Benson, H.E., Monaghan, A.E., Liew, W.C., Mpamhanga, C.P., Bonner, T.I., Neubig, R.R. *et al.* (2013) International Union of Basic and Clinical Pharmacology. LXXXVIII. G protein-coupled receptor list: recommendations for new pairings with cognate ligands. *Pharmacol. Rev.*, **65**, 967–986.
- Klabunde, T. and Hessler, G. (2002) Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem*, **3**, 928–944.
- Reddy, A.S. and Zhang, S. (2013) Polypharmacology: drug discovery for the future. *Exp. Rev. Clin. Pharmacol.*, **6**, 41–47.
- Vroiling, B., Sanders, M., Baakman, C., Borrmann, A., Verhoeven, S., Klomp, J., Oliveira, L., de Vlieg, J. and Vriend, G. (2011) GPCRDB: information system for G protein-coupled receptors. *Nucleic Acids Res.*, **39**, 309–319.
- Horn, F., Bettler, E., Oliveira, L., Campagne, F., Cohen, F.E. and Vriend, G. (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.*, **31**, 294–297.
- Horn, F., Vriend, G. and Cohen, F.E. (2001) Collecting and harvesting biological data: the GPCRDB and NucleaRDB information systems. *Nucleic Acids Res.*, **29**, 346–349.
- Horn, F., Weare, J., Beukers, M.W., Horsch, S., Bairoch, A., Chen, W., Edvardson, O., Campagne, F. and Vriend, G. (1998) GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res.*, **26**, 275–279.
- Vroiling, B., Thorne, D., McDermott, P., Attwood, T.K., Vriend, G. and Pettifer, S. (2011) Integrating GPCR-specific information with full text articles. *BMC Bioinformatics*, **12**, 362.
- Sanders, M.P., Verhoeven, S., de Graaf, C., Roumen, L., Vroiling, B., Nabuurs, S.B., de Vlieg, J. and Klomp, J.P. (2011) Snooker: a structure-based pharmacophore generation tool applied to class A GPCRs. *J. Chem. Inf. Model.*, **51**, 2277–2292.
- Kufareva, I., Rueda, M., Katritch, V., Stevens, R.C. and Abagyan, R. (2011) Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure*, **19**, 1108–1126.
- Hanson, M.A., Roth, C.B., Jo, E., Griffith, M.T., Scott, F.L., Reinhart, G., Desale, H., Clemons, B., Cahalan, S.M., Schuerer, S.C. *et al.* (2012) Crystal structure of a lipid G protein-coupled receptor. *Science*, **335**, 851–855.
- Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le, T.I., Teller, D.C., Okada, T., Stenkamp, R.E. *et al.* (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. *Science*, **289**, 739–745.
- Jaakola, V.-P., Griffith, M.T., Hanson, M.A., Cherezov, V., Chien, E.Y.T., Lane, J.R., Ijzerman, A.P. and Stevens, R.C. (2008) The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science*, **322**, 1211–1217.
- Joosten, R.P., te Beek, T.A., Krieger, E., Hekkelman, M.L., Hooft, R.W., Schneider, R., Sander, C. and Vriend, G. (2011) A series of PDB related databases for everyday needs. *Nucleic Acids Res.*, **39**, D411–D419.
- Krieger, E., Nabuurs, S.B. and Vriend, G. (2003) Homology modeling. *Methods Biochem. Anal.*, **44**, 509–523.
- Oliveira, L., Paiva, A.C.M. and Vriend, G. (1993) A common motif in G-protein-coupled seven transmembrane helix receptors. *J. Comput. Aided Mol. Des.*, **7**, 649–658.
- Ballesteros, J.A. and 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.
- Gloriam, D.E., Foord, S.M., Blaney, F.E. and Garland, S.L. (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.
- Felsenstein, J. (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*, **5**, 164–166.
- Garland, S. and Gloriam, D. (2011) Methods for the successful application of chemogenomics to GPCR drug design. *Curr. Top. Med. Chem.*, **11**, 1870–1871.