IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels

Anthony J. Harmar1,*, Rebecca A. Hills1, Edward M. Rosser1, Martin Jones2, O. Peter Buneman3, Donald R. Dunbar1, Stuart D. Greenhill1, Valerie A. Hale1, Joanna L. Sharman1, Tom I. Bonner4, William A. Catterall5, Anthony P. Davenport6, Philippe Delagrange7, Colin T. Dollery6, Steven M. Foord9, George A. Gutman10, Vincent Laudet11, Richard R. Neubig12, Eliot H. Ohlstein13, Richard W. Olsen14, John Peters15, Jean-Philippe Pin16, Robert R. Raffolo17, David B. Sears18, Mathew W. Wright19 and Michael Spedding7

1Centres for Cardiovascular Science and Neuroscience Research, The Queen’s Medical Research Institute, 2Institute of Evolutionary Biology, Ashworth Labs, 3School of Informatics, University of Edinburgh, Edinburgh, UK, 4Laboratory of Genetics, National Institute of Mental Health, Bethesda, MD 20892-4405, USA, 5Department of Pharmacology, University of Washington, Seattle, WA 98195, USA, 6Clinical Pharmacology Unit, University of Cambridge, Cambridge, CB2 2QQ, UK, 7Institut de Recherches Servier, 92150 Suresnes, France, 8Management Division, GlaxoSmithKline, Harlow, CM19 5AW, UK, 9GlaxoSmithKline Research and Development, Stevenage, Hertfordshire, UK, 10Department of Microbiology and Molecular Genetics, University of California, Irvine, CA 92697, USA, 11Molecular Zoology Group, Institut de Génomique Fonctionnelle de Lyon, Lyon, France, 12Department of Pharmacology, University of Michigan, Ann Arbor, MI, USA, 13Venuvics Pharmaceuticals, Glenmoore, PA, USA 14Department of Molecular & Medical Pharmacology, University of California, Los Angeles, CA 90095-1735, USA, 15Neurosciences Institute, The University of Dundee, Dundee, DD1 9SY, UK, 16Centre National de la Recherche Scientifique, Montpellier, France, 17Wyeth Research, Collegeville, PA 19426, 18GlaxoSmithKline Pharmaceuticals, King of Prussia, PA 19406, USA and 19HGNC, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK

Received August 11, 2008; Revised September 30, 2008; Accepted October 1, 2008

ABSTRACT

The IUPHAR database (IUPHAR-DB) integrates peer-reviewed pharmacological, chemical, genetic, functional and anatomical information on the 354 nonsensory G protein-coupled receptors (GPCRs), 71 ligand-gated ion channel subunits and 141 voltage-gated-like ion channel subunits encoded by the human, rat and mouse genomes. These genes represent the targets of approximately one-third of currently approved drugs and are a major focus of drug discovery and development programs in the pharmaceutical industry. IUPHAR-DB provides a comprehensive description of the genes and their functions, with information on protein structure and interactions, ligands, expression patterns, signaling mechanisms, functional assays and biologically important receptor variants (e.g. single nucleotide polymorphisms and splice variants). In addition, the phenotypes resulting from altered gene expression (e.g. in genetically altered animals or in human genetic disorders) are described. The content of the database is peer reviewed by members of the International Union of Basic and Clinical Pharmacology Committee on Receptor Nomenclature and Drug Classification (NC-IUPHAR); the data are provided through manual curation of the primary literature by a network of over 60 subcommittees of NC-IUPHAR. Links to other bioinformatics resources, such as NCBI, Uniprot, HGNC and the rat and mouse genome databases are provided. IUPHAR-DB is freely available at http://www.iuphar-db.org.

INTRODUCTION

One-third of the medicinal drugs in current use and many drugs of abuse target members of three protein superfamilies: nonsensory G protein-coupled receptors (GPCRs),
voltage-gated-like ion channels (VGICs) and ligand-gated ion channels (LGICs) (1). These proteins, encoded by ~570 genes, encompass an estimated 20% of all likely drug targets and are, therefore, a focus of intense research in academia and in industry. The International Union of Basic and Clinical Pharmacology Committee on Receptor Nomenclature and Drug Classification (NC-IUPHAR; http://www.iuphar.org/nciuphar.html) has, since 1992, issued guidelines for receptor and ion channel nomenclature and for classifying the major receptor and ion channel systems. More recently, an important part of the mission of NC-IUPHAR has been to facilitate the characterization of new functional receptors and ion channels identified by sequencing of the human, rat and mouse genomes. In addition to publishing a series of reviews on these issues (http://www.iuphar.org/nciuphar_arti.html), the committee has, since 2000, worked to create a database of GPCRs and ion channels containing peer-reviewed information on the pharmacology, genetics, function and distribution of these proteins.

GPCRs, also known as seven-transmembrane domain (7TM) receptors because of their characteristic topology, comprise one of the largest protein superfamilies in mammals (2) and are found in a range of eukaryote taxa. The characteristic 7TM topology but for which no endogenous ligand has yet been identified. ‘Reverse pharmacology’ (3) is progressively allowing ‘deorphanisation’ of these receptors by assigning endogenous ligands and physiological functions to them (4).

The LGIC superfamily includes 10 families that share a common cation-selective pore-forming module composed of two transmembrane segments and an intervening P loop (5). The voltage-gated Na⁺ (Naᵥ) and Ca²⁺ (Caᵥ) channels are the most structurally complex. These single ion channel subunits have four repeating domains that each contains six transmembrane segments yielding 24 transmembrane segments in all. In each domain, segments S1–S4 comprise a regulatory module that confers voltage sensitivity, and segments S5 and S6 and the P loop line the central pore. In addition to their primary regulation by voltage on the millisecond time scale, these channels have slower secondary regulation by numerous signaling pathways. Members of the two-pore channel family (TPC), whose functional properties are unknown, are composed of two separate subunits that each have two linked domains with six transmembrane segments, similar to the homologous domains of Naᵥ and Caᵥ channels. Five ion channel families are tetramers of subunits that each has a structure homologous to one domain of a Naᵥ or Caᵥ channel. Voltage-gated K⁺ channels (Kᵥ) are primarily regulated by voltage and secondarily by G proteins and second-messenger signaling pathways. Ca²⁺-activated K⁺ channels (KᵥCa), transient receptor potential channels and hyperpolarization- and cyclic-nucleotide-gated ion channels are jointly regulated by voltage, membrane lipids and intracellular ligands. Cyclic-nucleotide-gated ion channels are primarily regulated by cyclic nucleotides, even though their S1–S4 segments are similar in structure to the voltage-gated channels. Finally, the two structurally simplest VGICs have only pore-forming domains. Inwardly rectifying K⁺ channels (Kᵢ) are composed of tetramers of subunits having two transmembrane segments with an intervening P loop. Two-P K⁺ channels (KᵥP) are composed of dimers of subunits having two linked pore-forming motifs, each similar to the Kᵢ₆ channels. Kᵢ₆ and KᵥP channels are regulated by membrane lipids and intracellular ligands, including G proteins and small molecules such as Mg²⁺, polyamines and ATP. Diversity within this large protein superfamily is increased by association of the principal pore-forming subunits with one or more auxiliary subunit and by formation of heterooligomers of the pore-forming subunits of the family members that function as tetramers.

LGICs, unlike GPCRs, incorporate the ligand-binding site and effector (i.e. ion channel) within a common multimeric complex. They are the mediators of fast, phasic, synaptic transmission in the nervous system and at the skeletal neuromuscular junction. In addition, some LGICs underlie a tonic form of synaptic transmission in the central nervous system and their distribution and functions are not limited to excitable cells. The LGICs form three superfamilies on the basis of homology in the amino acid sequences and topology of their component subunits, namely the pentameric Cys-loop, and the cation-selective ionotropic glutamate and P2X receptors, which assemble as tetramers and trimers, respectively. The Cys-loop receptors comprise the cation-selective nicotinic acetylcholine and 5-hydroxytryptamine type-3 (5-HT₃) receptors (with 17 and 5 subunits, respectively, encoded by distinct genes) and the anion-selective GABAₐ (19 subunits) and glycine receptors (five subunits) (6–9). A cation-selective zinc-activated channel forms an additional member of the Cys-loop superfamily (10). Ionotropic glutamate receptors comprise the N-methyl-D-aspartate (seven subunits), α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (four subunits) and kainate (five subunits) receptor classes. Two orphan subunits have also been cloned (11). The P2X receptor subunits are P2X1 through to P2X7 (12).

Several other databases contain information on GPCRs and ion channels. The NIMH-PDSP K Database (http://pdsp.med.unc.edu/pdsp.php), DrugBank (13) and BindingDB (14) each contain data on the interaction of small molecule ligands with known or potential drug targets.
PharmGKB (15) is a pharmacogenetics and pharmacogenomics resource with coverage of many of the drugs and drug targets in IUPHAR database (IUPHAR-DB). GLIDA (GPCR-ligand database) (16) contains biological information on GPCRs and chemical information on their known ligands. GPCRdb (17) contains cDNA and amino acid sequences, multiple sequence alignments, phylogenetic trees and structural models of GPCRs, the GPCR NaVa database (18) describes naturally occurring sequence variation in human GPCRs and gpDB (G protein database) (19) is a database of G proteins and their interactions with GPCRs and effector molecules. The Endogenous GPCR List (http://www.tumor-gene.org/GPCR/) tabulates the GPCRs expressed endogenously in various cell lines. There are two databases of olfactory receptors (the largest multigene family in multicellular organisms): HORDE (The Human Olfactory Data Explorer) (20) and the Olfactory Receptor Database (21); these are sensory GPCRs and are, therefore, outside the scope of IUPHAR-DB. There are few databases concerned with LGICs and VGICs in the public domain. VKCDB, the voltage-gated potassium channel database (22), contains protein sequences and electrophysiological and pharmacological data on voltage-gated potassium channels and LGICdb (23) contains nucleic acid and protein sequences, multiple sequence alignments, phylogenetic trees and structural information on LGIC subunits.

IUPHAR-DB complements existing databases by providing a richly curated overview of the biology of GPCRs and ion channels, underpinned by rigorous peer-review by NC-IUPHAR and its network of ~60 subcommittees of international experts.

CONSTRUCTION AND CONTENT

IUPHAR-DB is implemented as a MySQL relational database (http://www.mysql.com) containing information on GPCRs and a PostgreSQL relational database (http://www.postgresql.org) holding data on VGICs and LGICs. Data are submitted and edited using an in-house editing tool written in Java and using JDBC to map the biological objects to the database. To ensure that data remain consistent between databases, the editing tool retrieves citations directly from the PubMed database (http://www.ncbi.nlm.nih.gov/pubmed). The public web interface uses Java servlets, Java Server Pages and JDBC to provide free online access to the entire public database. The public interface runs in the Tomcat servlet container on a Linux platform.

IUPHAR-DB contains information on the GPCRs, VGICs and LGICs (‘drug targets’) encoded by the human, mouse and rat genomes. In accordance with IUPHAR guidelines, GPCRs and LGICs are grouped into families according to their endogenous ligands (24,25), whereas VGICs are grouped according to a phylogenetically based classification scheme (5). NC-IUPHAR has a network of over 60 expert subcommittees, each committee being responsible for developing the nomenclature for their receptor. The compilation of the data submitted to the database is, in most cases, coordinated by members of the relevant subcommittee. Where no relevant subcommittee exists, data are captured by the curators or individual experts and peer reviewed by at least two external expert referees. Data are sourced from and referenced to the primary literature (original articles in peer-reviewed publications rather than review articles), with links to citations in PubMed. Wherever possible, data are supported by more than one literature source. After review by the curators to ensure accuracy and consistency with the rest of the information in the database, the data are added to the production server and transferred to the public database, after approval by NC-IUPHAR (updates normally take place twice a year following NC-IUPHAR executive committee meetings). Data are reviewed at regular intervals (at least yearly) by subcommittees and other contributors and updated as necessary.

Data on each gene in the database can be displayed on an individual page containing the following information:

1. Approved IUPHAR nomenclature alongside alternative or outdated names for the receptor or channel.
2. Structural and genomic data linked to its sources in the HUGO Gene Nomenclature Committee (HGNC) (26), Mouse Genome Informatics (MGI) (27), Rat Genome (28) and Refseq protein (29) databases.
3. Links to papers describing the first cloning of each receptor or ion channel cDNA and gene.
4. Links to other databases including Entrez Gene (30), GeneCards (31) and OMIM (32).

The database contains support for heterooligomers composed of two or more GPCR (33) or ion channel subunits, for example GABA_B receptor heterodimers (34) and for complexes of receptors or channels with accessory proteins, for example the multiple pharmacologically distinct receptor subtypes that are generated by association of receptor activity-modifying proteins with the calcitonin-receptor-like receptor and the calcitonin receptor (35–37). On pages describing such complexes, there is a subunit table with links to database pages describing the properties of the constituent subunits or accessory proteins.

The following curated and peer-reviewed information is provided for all genes in the database (with the exception, at present, of some orphan GPCRs):

1. Tissue distribution of gene expression at the levels of mRNA, protein and radioligand binding, focusing on the adult.
2. Tissue function (physiological responses mediated by the receptor or ion channel).
3. Functional assays (whole tissue or isolated cell systems in which a pharmacological response can be firmly attributed to the function of a defined receptor or ion channel).
4. Physiological consequences of altering gene expression (e.g. in knockout and transgenic animals).
5. Functionally important receptor variants (e.g. polymorphisms, mutations and splice variants, which have been demonstrated to alter receptor function).
6. Tables of affinity data for selected ligands.

Only selected ligands are displayed in the database. These groups include drugs that are potent and selective and/or...
used as prescription medications, endogenous ligands (e.g. 5-HT as an endogenous substance that acts through 5-HT receptors) and radiolabeled substances that can be employed in radioligand binding studies. In addition, ligands important for understanding structure–activity relationships for a drug target or a family of receptors or ion channels are included. Where possible, data for these index compounds are included for all members of the receptor or ion channel family.

When possible, the database cites estimates of the equilibrium dissociation constant (Kᵰ) for each ligand at the cloned human receptor or channel expressed in a transfected cell line, determined in a binding assay using a radiolabeled antagonist as tracer. When such data are lacking, results obtained from other species and/or from other assay systems are reported. If possible, data are obtained from more than one publication and a range of values is displayed together with multiple citations. Where available, alternative ligand names are displayed and links are provided to the relevant entry in the PubChem compound database (38) and to a table listing the potency of the ligand at all other reported receptors and channels in IUPHAR-DB. Tables of ligands can be sorted, for example, alphabetically by name, by activity, by affinity or by the units in which affinity data are provided.

Online Supplementary Figure 1 is an interactive PDF file illustrating the main features of a database page for a GPCR. These pages contain tables of ligands classified as agonists (full or partial), antagonists, inverse agonists or allosteric regulators (positive, negative or neutral) (39). Details of primary and secondary transduction mechanisms are given, listing the G proteins involved and the downstream response to receptor activation.

Although there is evidence that some VGICs may exist as heterooligomeric complexes with distinct pharmacological properties, IUPHAR-DB presently provides information on individual VGIC genes and the properties of homomeric channels. Quantitative data are included for the voltage dependence of activation and inactivation, single-channel conductance and binding of drugs and neurotoxins (classified as activators, gating inhibitors or pore blockers), focusing on agents that are widely used and that are diagnostic of channel identity and function.

In the case of the LGICs, which, in contrast to VGICs, do not generally assemble and function as homomers, the characteristics of individual receptors of defined subunit composition are presented. The data presented comprise entries concerning ion selectivity, conductance and voltage dependence, agonists, antagonists, channel blockers, allosteric regulators and functional assays.

The web interface includes a comprehensive search facility to enable text-based searches of drug targets and ligands. Searches are, by default, across all fields in the database; alternatively, searches can be focused by selecting one or more fields (e.g. searching for a receptor name or alias). Alphabetical lists of receptor and ion channel families are available via a menu on the page sidebar, which links to individual receptor or channel database pages.

**ADDITIONAL FILES**

There is an introductory article for each receptor and ion channel family, which reviews the properties, nomenclature and classification of each receptor and channel family. Curated lists of the genes encoding human, mouse and rat GPCRs, VGICs and LGICs can be viewed as HTML pages that list the approved IUPHAR nomenclature, have links to the receptor or channel page in the database, give the name of the primary endogenous ligand or physiological ion, human, rat and mouse gene names and provide links to the Entrez Gene database for all three species. Alternatively, the gene lists can be downloaded as Microsoft Excel spreadsheets, which additionally include the HGNC, RGD and MGI identifiers, genomic location, RefSeq nucleotide and protein accession numbers and SwissProt and Entrez Gene identifiers for human, rat and mouse.

The Evolving Pharmacology subcommittee of NC-IUPHAR monitors the literature for reports of new ‘pairings’ of GPCRs with endogenous ligands, for example, the recent emergence of estrogen as an endogenous ligand for GPR30 (40). A webpage (http://www.iuphar-db.org/latestPairings.jsp) gives details of recent developments in this area. A ‘hot topics’ page (http://www.iuphar-db.org/hotTopics.jsp) contains brief summaries of important developments in receptor and ion channel pharmacology. An RSS feed (http://www.iuphar-db.org/feed.xml) provides details of the latest items added to the database.

**DISCUSSION AND FUTURE DIRECTIONS**

The aim of the IUPHAR-DB project is to create a richly annotated resource giving pharmacological, genetic, functional and anatomical information on a subset of drug targets that are of particular importance in the treatment of disease and in the development of new medicines. Uniquely, the content of the database is peer reviewed by international experts and the drug targets are defined unambiguously using IUPHAR-approved nomenclature. Only drugs that are potent and selective, used clinically or are important in understanding structure–activity relationships are included.

The future development of IUPHAR-DB will include the addition of further classes of drug targets, such as the nuclear receptors (41) and receptor tyrosine kinases. Refinements to the database will include the provision of more sophisticated search tools and the development of ‘ligand-centered’ pages, which will aggregate information on the pharmacology of individual drugs, as an alternative entry point to the database.

Although there are other databases that document aspects of the molecular biology and pharmacology of GPCRs and ion channels, IUPHAR-DB is the first in the public domain to integrate curated structural, physiological, pathophysiological and quantitative pharmacological data across a wide range of drug targets. Intended as an international resource for students, scientists and the interested public, the website receives over 3500 unique visitors from ~80 countries each month.
DATA AVAILABILITY

IUPHAR-DB is freely available at http://www.iuphar-db.org with online help at http://www.iuphar-db.org/helpPage.jsp. SQL dumps of the datasets can be supplied on request to curators@iuphar-db.org.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the Executive Committee of IUPHAR for its sustained support of the project and the NC-IUPHAR Subcommittee members and individual experts for their contributions to the content of the database.

FUNDING

British Pharmacological Society (through their Anniversary Strategic Initiatives Fund); UNESCO (through the ICSU Grants Programme); Incyte; GlaxoSmithKline; Merck; Novartis; Servier; Wyeth. Funding for Open Access Publication charges were waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES


