Databases and ontologies

gpDB: a database of GPCRs, G-proteins, effectors and their interactions

Margarita C. Theodoropoulou[1], Pantelis G. Bagos[1,2], Ioannis C. Spyropoulos[1] and Stavros J. Hamodrakas[1,*]

[1]Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Athens 157 01 and [2]Department of Informatics with Applications in Biomedicine, University of Central Greece, Lamia 35 100, Greece

ABSTRACT

Summary: gpDB is a publicly accessible, relational database, containing information about G-proteins, G-protein coupled receptors (GPCRs) and effectors, as well as information concerning known interactions between these molecules. The sequences are classified according to a hierarchy of different classes, families and subfamilies based on literature search. The main innovation besides the classification of G-proteins, GPCRs and effectors is the relational model of the database, describing the known coupling specificity of GPCRs to their respective alpha subunits of G-proteins, and also the specific interaction between G-proteins and their effectors, a unique feature not available in any other database.

Availability: http://bioinformatics.biol.uoa.gr/gpDB

Contact: shamodr@biol.uoa.gr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

G-proteins, through their interaction with G-protein coupled receptors (GPCRs), act as switches for signal transduction from the extracellular space into the cell. G-proteins form heterotrimers composed of Gα, Gβ and Gγ subunits, and they also possess a binding site for a nucleotide (GTP or GDP). They are named after their α-subunits, which on the basis of their amino acid similarity and function are grouped into four families (Gαs, Gαi/o, Gαq, Gα12)(Cabrera-Vera *et al.*, 2003).

GPCRs form the major group of receptors in eukaryotes and they possess seven transmembrane α-helical domains. GPCRs are usually classified into several classes, according to the sequence similarity shared by the members of each class. The major GPCR classes are presented in the classification of GPCRDB (Horn *et al.*, 2003). Furthermore, there are a number of putative classes of newly discovered GPCRs, whose nomenclature has not been accepted yet from the scientific community (Kristiansen, 2004; Pierce *et al.*, 2002).

The stimulation of GPCRs leads to the activation of G-proteins, which dissociate into Gα and G$\beta\gamma$ subunits. The subunits then activate several effector molecules that lead to

many kinds of cellular and physiological responses (Oldham and Hamm, 2007). Effectors form a diverse group of proteins, through their interaction with G-proteins that either act as second messengers, or lead directly to a cellular and physiological response. Many proteins act as effectors, such as tubulins, adenylate cyclases, ion channels and others (Kristiansen, 2004).

The great importance of GPCRs and the corresponding signal transduction pathways is indicated by the fact that almost 50% of the current prescription drugs target GPCRs (Attwood, 2001). gpDB is a publicly available database that contains data about G-proteins and GPCRs, classified into different classes, families and subfamilies, and also, information about the coupling specificity of GPCRs to G-proteins (Elefsinioti *et al.*, 2004). We now, extend gpDB in order to contain the information about the interactions between G-proteins and their effectors.

2 METHODS

In order to collect recent data concerning interactions between G-proteins and effectors in gpDB, we performed an extensive literature search. The initial sequence information for effectors was retrieved from the UNIPROT database (Wu *et al.*, 2006). The entries were acquired using suitable scripts written in Perl in order to parse the DE (description), the GN (gene) or the DR (database cross reference) field in the respective database entry. The datasets were then checked in order to eliminate duplicates. We used user-written Perl scripts to manipulate the data, whereas annotations regarding the interaction between G-proteins and effectors, the effect of the particular interaction and the corresponding references were appended manually in a spreadsheet. For G-protein/effector interactions, we now provide links to PUBMED corresponding to original articles reporting the association.

The data has been organized on the basis of a relational model and is stored in a PostgreSQL database system. The user has supervisory access through our Apache web server interferential software, which was developed in Java for database manipulation. This software tends to settle any web server's query. In order to extend gpDB, we modified the initial relational scheme, which is the main innovation of this database. Figure 1 in Supplementary Material shows the relational scheme of gpDB. The interactions between G-proteins and GPCRs are shown at the subfamily level, whereas, G-protein subfamilies are interacting with specific effector types. Both G-protein to GPCR and G-protein to effector interactions are not one-to-one functions.

G-proteins and GPCRs are classified in classes, families, subfamilies and types, whereas effectors are classified in families, subfamilies and types. The classification of effectors is based on their exact biological

*To whom correspondence should be addressed.

function, which is a new innovation of gpDB. Each database entry contains the following fields: gpDB name, gpDB id, Uniprot accession number, protein description and classification, sequence, species, organism common name, taxonomy, links to other databases, coupling preference for G-proteins and GPCRs (if existent) and interaction between G-proteins and effectors. For GPCRs there is information for accessory proteins, as well as for homodimerization and/or hetero-dimerization (if existent). All this information is accompanied by links to original articles.

3 RESULTS

gpDB currently contains data concerning 391 G-proteins, 2738 GPCRs with known coupling preference and 1390 effectors, knowing to interact with specific G-proteins. The data is classified in classes, families, subfamilies and types for G-proteins and GPCRs, while for effectors in families, subfamilies and types. In particular, effectors are categorized in 21 families, 31 subfamilies and 68 types, based on their biological function. Figure 2 in Supplementary Material shows an entry of gpDB.

The application possesses a user-friendly environment, through which, the user may retrieve the necessary information, find available resources and cross-references and perform additional tasks such as run predictive algorithms, perform alignments, etc. In the main page of gpDB the user may find links for the following tools: Navigation, Text Search, BLAST Search and Pattern Search. In the entry page for GPCRs the user may also find additional tools: HMMTM, an algorithm for the prediction of the topology of transmembrane proteins using HMMs (Bagos *et al.*, 2006); PRED-GPCR, a tool for the classification of GPCRs (Papasaikas *et al.*, 2004); PRED-COUPLE2, a tool for the prediction of the coupling specificity of GPCRs to four families of G-proteins (Sgourakis *et al.*, 2005) and TMRPres2D, a tool for the visual representation of transmembrane protein models (Spyropoulos *et al.*, 2004). There is also an extensive user's manual page, describing in detail the available tools.

The information on the interaction between G-proteins and effectors is presented in each entry of both of these molecules. In a G-protein entry the user can find information concerning the biological effect of the interaction with specific effectors, accompanied with a link to the original article. The interactions are also shown at the subfamily level for G-proteins, using the yellow link button, which presents a list of all the effector types, which interact with the specific G-protein subfamily. At the effectors' type-level, the user can retrieve a list of all G-protein subfamilies, which interact with the specific effector type, by using the red link button, respectively.

4 DISCUSSION

The database that we present here has some innovative and unique features not available in any other publicly accessible resource. The relational scheme, on which the database is organized, is especially designed to capture the coupling preferences of G-proteins to GPCRs and the interaction between G-proteins and effectors according to the reported data in the scientific literature. General sequence databases (i.e. UniProt) in case they contain such information, this is at the free-text field of FUNCTION, while all the other GPCR-specific databases are focusing mainly in classifying GPCRs

(Horn *et al.*, 2003). However, gpDB includes information for G-proteins, GPCRs and effectors, and, at the same time provides information regarding the coupling specificity of G-proteins and GPCRs, the result of the G-protein/effector interaction, the accessory proteins interacting with GPCRs and information about the dimerization of GPCRs, all accompanied by links to original research articles from which the information was retrieved, features not available in any other publicly accessible resource. gpDB does not aim at being a universal resource for GPCRs, but will be acting complementary to the existing GPCR-related databases such as GPCRDB (Horn *et al.*, 2003) or RINGdb (Fang *et al.*, 2006). Integrating the various GPCR-related databases and constructing ontologies for GPCRs will provide useful directions for future research (Skrabanek *et al.*, 2007). All the information provided in gpDB will be updated on a yearly basis and may be used in the future to develop algorithms predicting the coupling specificity of GPCRs to G-proteins, for predicting the biological effect of effector molecules and/or help in the construction of protein interaction networks representing the signal transduction pathway.

REFERENCES

Attwood,T.K. (2001) A compendium of specific motifs for diagnosing GPCR subtypes. *Trends Pharmacol. Sci.*, **22**, 162–165.

Bagos,P.G. *et al.* (2006) Algorithms for incorporating prior topological information in HMMs: application to transmembrane proteins. *BMC Bioinformatics*, **7**, 189.

Cabrera-Vera,T.M. *et al.* (2003) Insights into G protein structure, function, and regulation. *Endocr. Rev.*, **24**, 765–781.

Elefsinioti,A.L. *et al.* (2004) A database for G proteins and their interaction with GPCRs. *BMC Bioinformatics*, **5**, 208.

Fang,Y.C. *et al.* (2006) RINGdb: an integrated database for G protein-coupled receptors and regulators of G protein signaling. *BMC Genomics*, **7**, 317.

Horn,F. *et al.* (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.*, **31**, 294–297.

Kristiansen,K. (2004) Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol. Ther.*, **103**, 21–80.

Oldham,W.M. and Hamm,H.E. (2008) Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat. Rev. Mol. Cell Biol.*, **9**, 60.

Papasaikas,P.K. *et al.* (2004) PRED-GPCR: GPCR recognition and family classification server. *Nucleic Acids Res.*, **32**, W380–W382.

Pierce,K.L. *et al.* (2002) Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.*, **3**, 639–650.

Sgourakis,N.G. *et al.* (2005) Prediction of the coupling specificity of GPCRs to four families of G-proteins using hidden Markov models and artificial neural networks. *Bioinformatics*, **21**, 4101–4106.

Skrabanek,L. *et al.* (2007) Requirements and ontology for a G protein-coupled receptor oligomerization knowledge base. *BMC Bioinformatics*, **8**, 177.

Spyropoulos,I.C. *et al.* (2004) TMRPres2D: high quality visual representation of transmembrane protein models. *Bioinformatics*, **20**, 3258–3260.

Wu,C.H. *et al.* (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.*, **34**, D187–D191.