Databases and ontologies

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

Margarita C. Theodoropoulou¹, Pantelis G. Bagos¹,², Ioannis C. Spyropoulos¹ and Stavros J. Hamodrakas¹,*
¹Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Athens 157 01 and
²Department of Informatics with Applications in Biomedicine, University of Central Greece, Lamia 35 100, Greece

Received on December 26, 2007; revised on March 26, 2008; accepted on April 23, 2008
Advance Access publication April 25, 2008
Associate Editor: Jonathan Wren

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.

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
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 heterodimerization (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 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.

ACKNOWLEDGEMENTS

The authors would like to thank Antigoni Elefsinioti for her technical assistance during the extension of the database, and the anonymous reviewers for their valuable comments.

Funding: P.G.B. would like to acknowledge the State Scholarships Foundation of Greece (SSF), for financial support of the project ‘Machine Learning Algorithms for Bioinformatics’.

Conflict of Interest: none declared.

REFERENCES