

GLIDA: GPCR-ligand database for chemical genomic drug discovery

Yasushi Okuno*, Jiyeon Yang, Kei Taneishi, Hiroaki Yabuuchi and Gozoh Tsujimoto

Department of Genomic Drug Discovery Science, Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida-Shimo-Adachi-cho, Sakyo-ku, Kyoto 606-8501, Japan

Received August 15, 2005; Revised and Accepted September 22, 2005

ABSTRACT

G-protein coupled receptors (GPCRs) represent one of the most important families of drug targets in pharmaceutical development. GPCR-Ligand Database (GLIDA) is a novel public GPCR-related chemical genomic database that is primarily focused on the correlation of information between GPCRs and their ligands. It provides correlation data between GPCRs and their ligands, along with chemical information on the ligands, as well as access information to the various web databases regarding GPCRs. These data are connected with each other in a relational database, allowing users in the field of GPCR-related drug discovery to easily retrieve such information from either biological or chemical starting points. GLIDA includes structure similarity search functions for the GPCRs and for their ligands. Thus, GLIDA can provide correlation maps linking the searched homologous GPCRs (or ligands) with their ligands (or GPCRs). By analyzing the correlation patterns between GPCRs and ligands, we can gain more detailed knowledge about their interactions and improve drug design efforts by focusing on inferred candidates for GPCR-specific drugs. GLIDA is publicly available at <http://gdds.pharm.kyoto-u.ac.jp:8081/glida>. We hope that it will prove very useful for chemical genomic research and GPCR-related drug discovery.

INTRODUCTION

The superfamily of G-protein coupled receptors (GPCRs) forms the largest class of cell surface receptors. These molecules regulate various cellular functions responsible for physiological responses (1). GPCRs represent one of the most important families of drug targets in pharmaceutical development (2). A large majority of human-derived GPCRs still

remain 'orphans' with no identified natural ligands or functions, and thus a key goal of GPCR research related to drug design is to identify new ligands for such orphan GPCRs.

With the unprecedented accumulation of the genomic information, databases and bioinformatics have become essential tools to guide GPCR research. The GPCRDB (<http://www.gpcr.org/7tm/>) (2) and IUPHAR (<http://iuphar-db.org/iuphar-rd/index.html>) (3) receptor databases are representatives of widely used public databases covering GPCRs. These databases, which provide substantial data on the GPCR proteins and pharmacological information on receptor proteins containing GPCRs, are mainly focused on biological aspects of the gene products or proteins. In spite of the significance of ligand compounds as drug leads, the relationships between GPCRs and their ligands and/or chemical information on the ligands themselves are not yet fully covered.

On the other hand, there is increasing interest in collecting and applying chemical information in the post-genome era. This new trend is called 'chemical genomics', in which biological information and chemical information are integrated on the genome scale (4,5). PubChem (<http://pubchem.ncbi.nlm.nih.gov/>) (6), KEGG/LIGAND (<http://www.genome.jp/kegg/ligand.html>) (7) and ChEBI (<http://www.ebi.ac.uk/chebi/>) (8) have been developed as databases related to chemical genomics. KEGG/LIGAND and ChEBI contain primarily biochemical information on reported enzymatic reactions. Recently, NIH (the National Institutes of Health) opened PubChem, a public database providing information on the chemical structures of small molecules. However, one cannot retrieve direct information relating these chemical structures to gene or protein entries. Although chemical genomic approaches have thrown new light on relationships between receptor sequences and compounds that interact with particular receptors, the GPCR-ligand information is not well represented in these large-scale databases for chemical genomics.

There are still very few publicly available databases or tools for GPCR-specialized drug discovery from the viewpoint of chemical genomics. Herein, we have developed a novel relational database, GLIDA (GPCR-Ligand Database) (9).

*To whom correspondence should be addressed. Tel: +81 75 753 9264; Fax: +81 75 753 4544; Email: okuno@pharm.kyoto-u.ac.jp

Table 1. The current numbers of GLIDA ligands and GPCRs and their respective links

Information item	Number of entries
GPCR entries	3738
Links to Entrez Gene	3073
Links to GPCRDB	3738
Links to UniProt	3738
Links to IUPHAR	389
Links to KEGG	595
Ligand entries	649
Cas registry number	320
Lolecular structure	364
Links to PubChem	242
Links to ChEBI	28
Links to KEGG	109
GPCR-ligand pair entries	1989
GPCR entries	281
Ligand entries	632

GLIDA contains biological information on GPCRs and chemical information on their ligand compounds. Furthermore, it provides various analytical data on GPCR-ligand correlations by incorporating bioinformatics and chemoinformatics methods, and thus it should prove very useful for chemical genomic research in GPCR-related drug discovery.

DATA CONTENTS

GLIDA contains three types of primary data: biological information on GPCRs, chemical information on their ligands and information on binding of specific GPCR-ligand pairs. The GPCR entries were acquired from the deposits of human, mouse and rat entries in the GPCRDB because these three species include sufficient information regarding ligands, and rats and mice are representative model animals for drug discovery. The ligand information was manually collected and curated using various public web sites and commercial DBs, such as the IUPHAR Receptor Database, PubMed, PubChem and MDL ISIS/Base 2.5. Table 1 indicates the size and scope of the GLIDA database.

GPCR and ligand data

The database lists general information on GPCR and ligand data, respectively. The general information table of GPCR contains gene names, family names, protein sequences and links to other biological databases, such as GPCRDB, UniProt, IUPHAR, Entrez Gene and KEGG. The ligand result page provides a general information table containing names, molecular structures, CAS registry numbers, formulas, molecular weights, MOLfiles and links to the other chemical databases KEGG, PubChem and ChEBI.

Information on binding of GPCR-ligand pairs

The correlation information relating GPCRs to particular ligands, a key issue for GPCR-related drug discovery, is stored in a relational database. GLIDA allows users to retrieve GPCR-ligand binding information dynamically and continuously. When users retrieve a GPCR (or ligand) entry, its result page displays all entries showing the corresponding ligands (or GPCR entries) with their binding activity types, as well as

references. The references are hyperlinked with the corresponding PubMed literature or the IUPHAR pages that were used to collect the information regarding GPCR-ligand binding. The activity types include agonist, inverse agonist, antagonist and so on. An agonist will bind to and activate the corresponding GPCRs, whereas an antagonist will bind to and block the activity of the corresponding GPCRs. An inverse agonist binds to GPCRs and reduces the fraction of them that are in an active conformation, and a partial agonist is an agonist that in a given tissue, under specified conditions, cannot elicit as large an effect as another agonist acting through the same GPCRs in the same tissue can.

WEB INTERFACE AND APPLICATION

GLIDA was constructed on the LAMP (Linux, Apache, MySQL and PHP) platform. GLIDA is available at <http://gdds.pharm.kyoto-u.ac.jp:8081/glida>. The web interface of GLIDA includes a GPCR search page (Figure 1a) and a ligand search page (Figure 1b). Each page consists of a classification table and a keyword search box. The user can search a GPCR (or ligand) manually from the guide-tree of the classification table, or automatically by using the keyword search function of MySQL. Every GPCR (or ligand) has its own result page (Figure 1c or d) containing a general information table for a GPCR (or ligand), a table of its correlated ligands (or GPCRs) and a button to carry out a similarity search and correlation analysis. Clicking the button starts the calculation, and an analytical report page (Figure 1e) then appears with a list of the top 25 entries that are most similar to the GPCR (or ligand) and a correlation map of the 25 GPCRs (or ligands) and their corresponding binding pairs. A search starting from ligand retrieval proceeds in the same way.

Hierarchical classification

The GPCR classification table on the search page was adapted from the phylogenetic tree of the GPCRDB information system (<http://www.gpcr.org/7tm/phylo/phylo.html>). As for the ligand classification table, GLIDA offers an original one (Figure 1b) that is based on a cluster analysis of the ligand structures as follows. We converted the structural images of the ligands into computational MDL Mol files using ISIS/Draw software. Next, we calculated distance metrics among all of the ligands using the frequency profiles of the atoms and the bonds of the KEGG atom types (10), and carried out complete-linkage clustering. We manually defined sub-clusters based on their common structural skeletons. Both the GPCR and ligand classification tables display the entries of the corresponding GPCRs or ligands at the end of the tree, and these are hyperlinked with their respective result pages.

Similarity search and GPCR-LIGAND correlation maps

GLIDA has a structure similarity search function on its result pages. Alignment scores of protein sequences generated by the BLAST algorithm provide similarity measures for GPCRs. Ligand similarity is defined by the dissimilarity (distance) of frequency profile patterns generated from the constitutive atoms and bonds of the chemical structure, using the KEGG atom types (10,11). From this similarity search, the 25 most

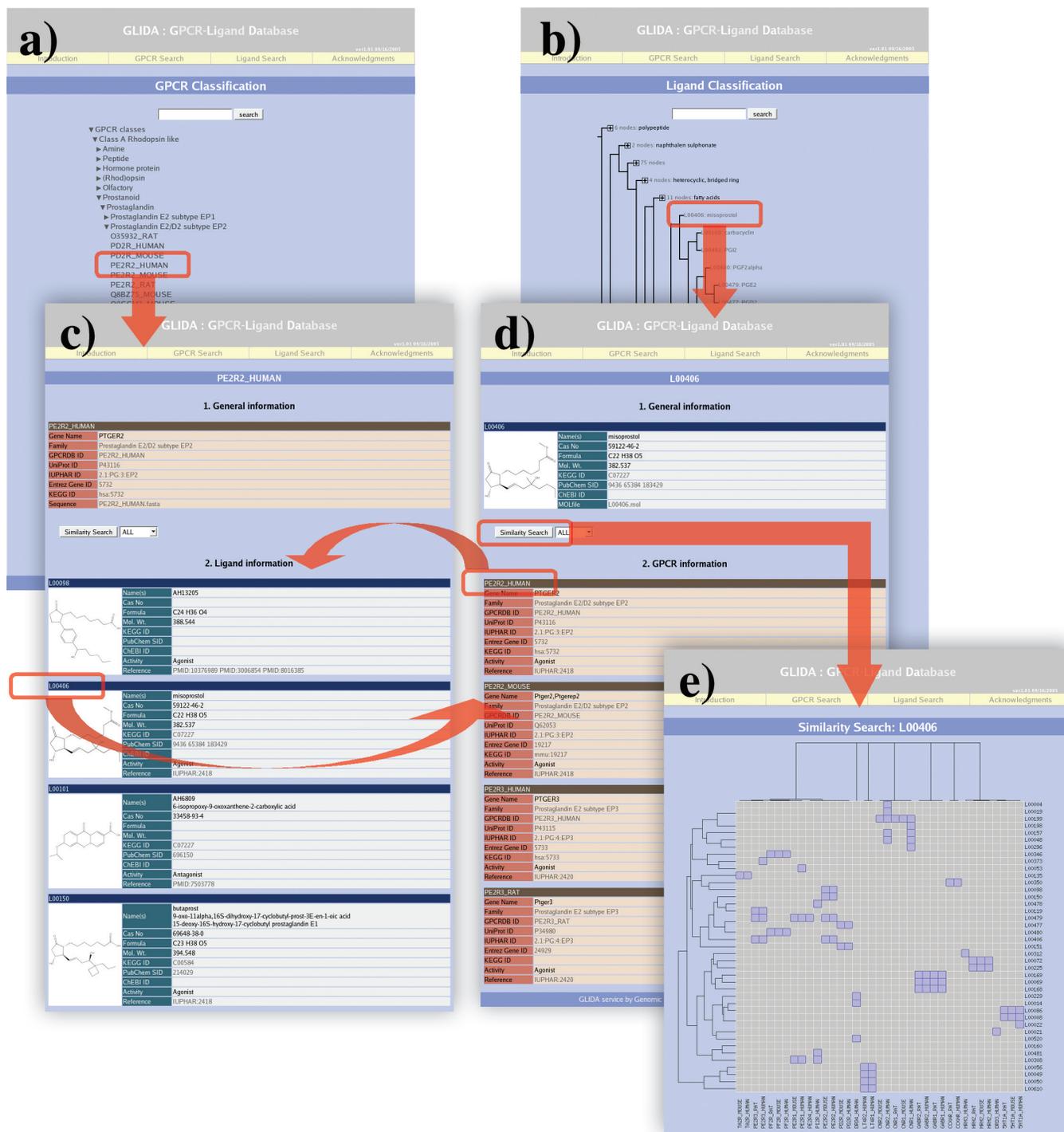


Figure 1. A screenshot of GLIDA showing its linked relations among search pages (a, b), result pages (c, d) and an analytical report page (e).

similar GPCRs (or 40 ligands) are retrieved and listed with their similarity scores on an analytical report page.

As the similarity search calculation is proceeding, GLIDA illustrates the correlation map (Figure 2e) showing the homologous GPCRs (or ligands) and their ligands (or GPCRs) that are retrieved. This map shows spots that match the GPCRs and their ligands in a two-dimensional matrix. The ordering along the x-axis and the y-axis are calculated respectively by

two-way clustering of the GPCRs and the ligands based on their similarities. In particular, the ordering along the x- and y-axis allows users to evaluate information regarding similarities and correlations between GPCRs and ligands simultaneously. By analyzing the correlation patterns between GPCRs and ligands that are illustrated by these maps, we can gain detailed knowledge about their interactions and utilize this information to infer possible candidates for development

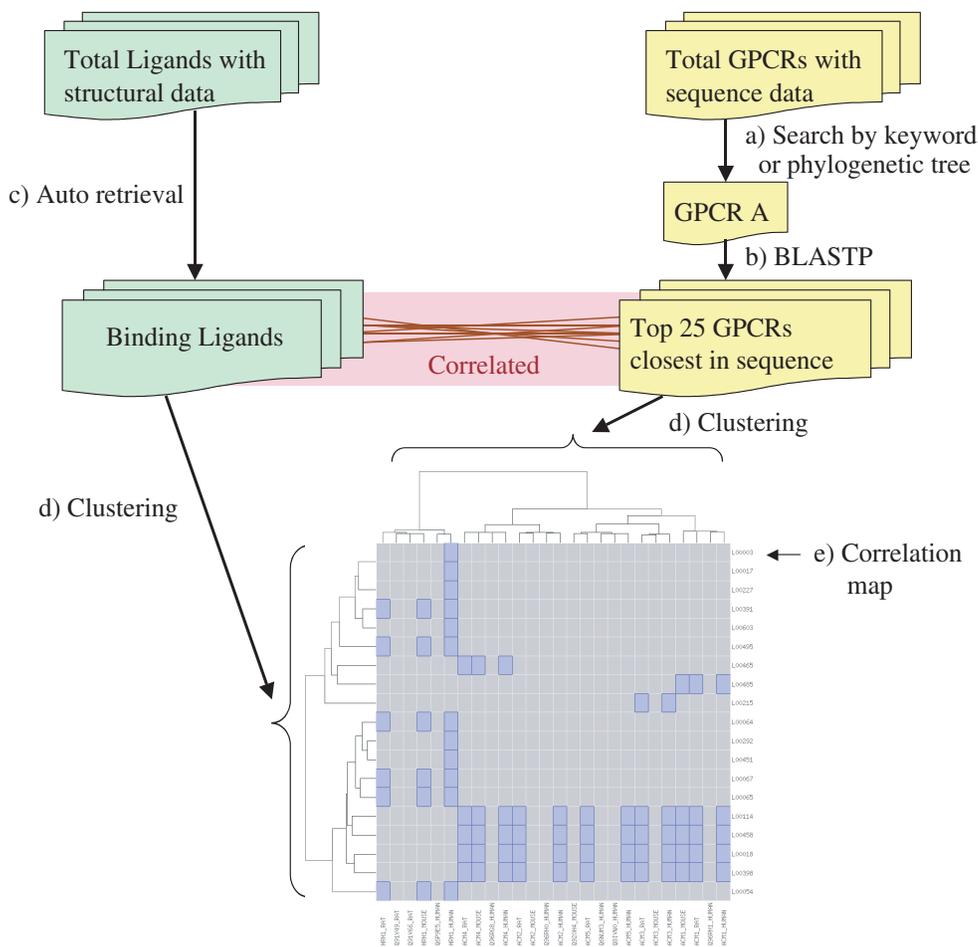


Figure 2. A schematic example of the search and analysis process showing GPCR-ligand correlations produced from a GPCR query using GLIDA. (a) If GPCR A is selected using a keyword search or a guide-tree search on the GPCR search page, its retrieved data will be displayed in its result page, (b) By clicking an analysis button on the result page, a list of the top 25 GPCRs that are most similar in sequence, including GPCR A, are obtained by the BLASTP calculation. (c) The server retrieves a list of corresponding ligands, which are respectively correlated with the 25 GPCRs. (d) Finally, a map is displayed to help visualize the matching spots linking GPCRs with particular ligands. The x-axis and y-axis respectively indicate the clustering results for GPCRs and ligands, calculated using sequence alignment scores among the GPCRs and structural profile distances among the ligands.

of GPCR-specific drugs. Figure 2 shows an example of the GPCR-ligand search and analysis process starting from a GPCR query using GLIDA.

DISCUSSION AND FUTURE DIRECTIONS

GLIDA provides a unique database for GPCR-related chemical genomic research and drug discovery. GLIDA is distinct from other public chemical genomic databases because it contains original, GPCR-specific chemical entries, although the total scale of its contents is not yet large (Table 1). GLIDA provides several advantages over other databases, in that a search can be started either from a GPCR or from a ligand. Thus, searches may be carried out in a dynamic and user-friendly way. GLIDA's coverage of chemical and biological information simultaneously also provides an advantage to users by saving them the time and labor required to search multiple databases. The ligand search page is another distinct characteristic of GLIDA in that it displays the structural distribution of ligands, and thereby facilitates research on

GPCR-related drugs by incorporating structural aspects of the ligand compounds. The analytical report pages resulting from the calculated structural similarities of GPCRs and ligands can give the user deep insights into the GPCR-ligand relationships. The lists of neighboring ligands (or GPCRs) and the correlation maps are useful visualizing tools for analyzing correlations among their structural features and their GPCR-ligand binding properties. Because the GLIDA algorithms can be applied to proteins other than the GPCR family, it may also be considered as a promising database for chemical genomics research.

GLIDA will be updated continuously. In particular, we are planning to computationally extract GPCR-ligand information from the literature and from patents using a text-mining tool, and to increase the number of ligand entries immediately. Further information on ligands from various computable chemical descriptors is currently being incorporated, and GLIDA will be combined with a system for predicting novel ligands of orphan GPCRs in the future. Furthermore, we also plan to carry out XML publication of GLIDA.

ACKNOWLEDGEMENTS

This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan, from the Ministry of Health, Labor and Welfare of Japan and from the 21st Century COE program 'Knowledge information infrastructure for Genome Science'. Funding to pay the Open Access publication charges for this article was provided by the Ministry of Health, Labor and Welfare of Japan.

Conflict of interest statement. None declared.

REFERENCES

1. George, S.R., O'Dowd, B.F. and Lee, S.P. (2002) G-protein-coupled receptor oligomerization and its potential for drug discovery. *Nature Rev. Drug Discov.*, **1**, 808–820.
2. 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.
3. Fredholm, B.B., Fleming, W.W., Vanhoutte, P.M. and Godfraind, T. (2002) The role of pharmacology in drug discovery. *Nature Rev. Drug Discov.*, **1**, 237–248.
4. Lipinski, C. and Hopkins, A. (2004) Navigating chemical space for biology and medicine. *Nature*, **432**, 855–861.
5. Dobson, C.M. (2004) Chemical space and biology. *Nature*, **432**, 824–828.
6. Zerhouni, E. (2003) Medicine: the NIH roadmap. *Science*, **302**, 63–72.
7. Goto, S., Okuno, Y., Hattori, M., Nishioka, T. and Kanehisa, M. (2002) LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic. Acids Res.*, **30**, 402–404.
8. Brooksbank, C., Cameron, G. and Thornton, J. (2005) The European Bioinformatics Institute's data resources: towards systems biology. *Nucleic Acids Res.*, **33**, D46–D53.
9. Yang, J., Okuno, Y. and Tsujimoto, G. (2004) GLIDA: GPCR and Ligand Database. *Genome Informatics*, **15**, P057.
10. Hattori, M., Okuno, Y., Goto, S. and Kanehisa, M. (2003) Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.*, **125**, 11853–11865.
11. Kotera, M., Okuno, Y., Hattori, M., Goto, S. and Kanehisa, M. (2004) Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions. *J. Am. Chem. Soc.*, **126**, 16487–16498.