LIGAND: chemical database of enzyme reactions

Susumu Goto*, Takaaki Nishioka1 and Minoru Kanehisa

Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan and 1Graduate School of Agricultural Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan

Received September 29, 1999; Accepted October 4, 1999

ABSTRACT

LIGAND is a composite database comprising three sections: ENZYME for the information of enzyme molecules and enzymatic reactions, COMPOUND for the information of metabolites and other chemical compounds, and REACTION for the collection of substrate–product relations. The current release includes 3390 enzymes, 5645 compounds and 5207 reactions. The database is indispensable for the reconstruction of metabolic pathways in the completely sequenced organisms. The LIGAND database can be accessed through the WWW (http://www.genome.ad.jp/dbget/ligand.html) or may be downloaded by anonymous FTP (ftp://kegg.genome.ad.jp/molecules/ligand/).

INTRODUCTION

Recent progress in the transcriptome and proteome analyses has made it possible to examine expression data of whole mRNAs or proteins in a cell and also a large amount of protein–protein interaction data. The information on gene expression and protein interactions is indispensable to predict gene functions from the complete genome sequence and to reconstruct biochemical pathways of an organism. However, for the reconstruction of a specific class of biochemical pathways, namely metabolic pathways, information on chemical compounds and reactions is also required. The LIGAND database has been organized to fill in the gap between genomic information and chemical information and applied to actual reconstruction of metabolic pathways in the completely sequenced organisms in KEGG (2,3).

NEW FEATURES OF COMPOUND

Compounds as interacting objects with proteins

Because chemical compounds in the COMPOUND section have roles in the living cell, they usually have interacting protein partners. At the moment, links are available only to the ENZYME section showing the relationship between chemical compounds and enzyme molecules. This kind of cross-reference information is quite useful to analyze the relationship between proteins and their ligands. Thus, we have added new link information to the PDB (8) and PROMISE (9) databases from the COMPOUND section.

We extract the information on heterogeneous group atoms from the PDB database and make a correspondence table between COMPOUND IDs and PDB HET codes. Then the links are automatically added to the DBLINKS field by the database update program. K. Degtyarenko (European Bioinformatics Institute), who develops the PROMISE database, kindly provided us with the link information between PROMISE and COMPOUND. We have also added it to the DBLINKS field.

Compounds in the ISIS database

For the purpose of substructure search of chemical compounds and for the ease of updating information of chemical compounds, we decided to maintain the COMPOUND section in the form of the ISIS/BASE database. Currently, all the information except for the DBLINKS (other than CAS) field is stored in the ISIS/BASE database. We generate the flat-file version of COMPOUND, which is made publicly available, by extracting the data from the ISIS database and by automatically merging computed link information.

The ENZYME section is based on the nomenclature of enzymes by IUBMB (International Union of Biochemistry and Molecular Biology) (6) and the Enzyme Handbook (7). Information regarding nomenclature by IUBMB is also available from the web at http://www.chem.qmw.ac.uk/iubmb/enzyme/. The COMPOUND section contains a collection of chemical compounds that are found in the ENZYME section and in the KEGG/PATHWAY database, as well as other compounds found in the literature. The REACTION section is a collection of binary relations, namely substrate–product relations extracted from the ENZYME section and the KEGG/PATHWAY database.

The number of entries in the current release is summarized in Table 1.

*To whom correspondence should be addressed. Tel: +81 774 38 3271; Fax: +81 774 38 3269; Email: goto@kuicr.kyoto-u.ac.jp
We also plan to maintain the REACTION section in the ISIS/BASE database.

**Classification of chemical compounds**

Since a hierarchical classification of chemical compounds is useful for searching similar compounds and generic compounds, we started developing a classification scheme for the compounds in the COMPOUND section. A preliminary version of the classification is summarized in Table 2.

**AVAILABILITY**

The LIGAND database is accessible through the WWW at http://www.genome.ad.jp/dbget/ligand.html. The user can then invoke the DBGET/LinkDB system (10,11) to retrieve the COMPOUND and ENZYME sections. Hierarchical classifications of enzymes and compounds can be viewed by the molecular catalog browser in the KEGG system at http://www.genome.ad.jp/kegg/kegg2.html. The periodic table for chemical elements is also available at the same URL.

The LIGAND database can be downloaded via anonymous FTP at ftp://kegg.genome.ad.jp/molecules/ligand/. This directory contains all sections, COMPOUND, ENZYME and REACTION, including GIF image files and MDL-MOL files for compound structures. The same data set is mirrored at the NCBI repository ftp://ncbi.nlm.nih.gov/repository/LIGAND/

The basic concept of the LIGAND database has been published elsewhere (1). The present article reflects the most up-to-date version of the database and should be cited accordingly.

**ACKNOWLEDGEMENTS**

We thank Hiroko Ishida for extraction of reaction data from the KEGG pathway diagrams, Nobue Takeuchi for the input of new chemical structures, and Rumiko Yamamoto for converting the original COMPOUND section into ISIS format. This work was supported by a Grant-in-Aid for Scientific Research on the Priority Area ‘Genome Science’ from the Ministry of Education, Science, Sports and Culture of Japan. The computation time was provided by the Supercomputer Laboratory, Institute for Chemical Research, Kyoto University.
REFERENCES