

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 (1) 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).

The LIGAND database is a composite database comprising three sections: ENZYME for information on enzyme molecules and enzymatic reactions, COMPOUND for information on metabolites and other chemical compounds, and REACTION for the collection of substrate–product relations. We report here the current status of the LIGAND database and the new features of the COMPOUND section.

CURRENT STATUS OF LIGAND

LIGAND is constructed as a flat-file database and the data format of each section is similar to those of GenBank (4) and PIR (5) flat-files, a fixed number of columns are assigned to specify each field of entry (1).

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.

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.

*To whom correspondence should be addressed. Tel: +81 774 38 3271; Fax: +81 774 38 3269; Email: goto@kuicr.kyoto-u.ac.jp

Table 1. The number of entries in release 19.0 (October 1999) of the LIGAND database

Section	Content	Number
ENZYME	Entries	3390
	Entries with reaction formula in chemical equations	2906
	Links to KEGG/PATHWAY (metabolic pathways)	1734
	Links to KEGG/GENES (gene catalogs)	1099
	Links to OMIM (human genetic disorders) (12)	469
	Links to PROSITE (proteins sequence motifs) (13)	977
COMPOUND	Entries	5645
	Entries with chemical formula	3755
	Entries with molecular structure	4768
	Links to ENZYME	4536
	Links to ENZYME as reactants	4365
	Links to ENZYME as cofactors	82
	Links to ENZYME as inhibitors	154
	Links to ENZYME as effectors	33
	Links to CAS	1537
REACTION	Entries	5200
	Reactions defined in ENZYME	3084
	Reactions with known enzymes in KEGG/PATHWAY	3102
	Reactions with unknown enzymes in KEGG/PATHWAY	303
	Non-enzymatic reactions in KEGG/PATHWAY ^a	385

^aNon-enzymatic reactions include reactions where it is not known whether enzymes are involved in catalysis.

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.

Table 2. Classification of chemical compounds and the number of entries in each class in release 19.0 (October 1999) of the LIGAND database.

Class	Subclass	Number		
Carbohydrates	Monosaccharides	Aldoses	21	
		Ketoses	12	
		Deoxysugars	7	
		Aminosugars	12	
		Uronates	11	
	Disaccharides		6	
		Polysaccharides	11	
		Lipids	Fatty acids	21
			Fats	6
			Phospholipids	11
Nucleic acids	Glycolipids	6		
	Steroid hormones	7		
	Other steroids	5		
	Eicosanoids	26		
	Bases	5		
Peptides	Nucleosides	9		
	Nucleotides	27		
	Cyclic nucleotides	10		
	Common amino acids	20		
	Other amino acids	35		
Others	Amines	11		
	Peptide hormones	5		
	Neurotransmitters	12		
	Fat-soluble vitamins	7		
	Water-soluble vitamins	11		

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

1. Goto,S., Nishioka,T. and Kanehisa,M. (1998) *Bioinformatics*, **14**, 591–599.
2. Kanehisa,M. and Goto,S. (2000) *Nucleic Acids Res.*, **28**, 27–30.
3. Kanehisa,M. (1997) *Trends Genet.*, **13**, 375–376.
4. Benson,D.A., Boguski,M.S., Lipman,D.J., Ostell,J., Ouellette,B.F., Rapp,B.A. and Wheeler,D.L. (1999) *Nucleic Acids Res.*, **27**, 12–17. Updated article in this issue: *Nucleic Acids Res.* (2000), **28**, 15–18.
5. Barker,W.C., Garavelli J.S., McGarvey,P.B., Marzec,C.R., Orcutt,B.C., Srinivasarao,G.Y., Yeh,L.S.L., Ledley,R.S., Mewes,H.W., Pfeiffer,F., Tsugita,A. and Wu,C. (1999) *Nucleic Acids Res.*, **27**, 39–43.
6. IUBMB (1992) *Enzyme Nomenclature: Recommendations (1992) of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology*. Academic Press, New York, NY.
7. Schomburg,D. (ed.) (1990) *Enzyme Handbook*. Springer-Verlag, Heidelberg, Germany, pp. 1–16.
8. Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.F., Brice,M.B., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.
9. Degtyarenko,K.N., North,A.C.T. and Findlay,J.B.C. (1999) *Nucleic Acids Res.*, **27**, 233–236.
10. Fujibuchi,W., Goto,S., Migimatsu,H., Uchiyama,I., Ogiwara,A., Akiyama,Y. and Kanehisa,M. (1998) *Pacific Symp. Biocomput.*, 683–694.
11. Kanehisa,M. (1997) *Trends Biochem. Sci.*, **22**, 442–444.
12. Pearson,P., Francomano,C., Foster,P., Bocchini,C., Li,P. and McKusick,V. (1994) *Nucleic Acids Res.*, **22**, 3470–3473.
13. Bairoch,A, Bucher,P. and Hofmann,K. (1999) *Nucleic Acids Res.*, **27**, 215–219.