Present Status of LIGAND Chemical Database for Enzyme Reactions

Takaaki Nishioka [1] nishioka@scl.kyoto-u.ac.jp
Susumu Goto [2] gogo@kuicr.kyoto-u.ac.jp
Minoru Kanehisa [2] kanehisa@kuicr.kyoto-u.ac.jp

[1] Graduate School of Agricultural Sciences, Kyoto University
Kyoto 606-01, Japan

[2] Institute of Chemical Research, Kyoto University
Gokasho, Uji, Kyoto 611, Japan

1 Introduction

LIGAND Chemical Database for Enzyme Reactions was originally designed to link structural aspects of enzymatic proteins to chemical aspects of enzymatic reactions . It has accumulated all known enzymes classified according to the EC numbers given by the International Union of Biochemistry and Molecular Biology (IUBMB) and has been integrated within the DBGET/LinkDB system with protein structure databases. Enzymatic reaction catalyzed by each enzyme is given by the chemical equation and by the chemical names of the substrates, products, and inhibitors of the reaction. LIGAND has provided a various approach to the analysis of structure-function relationships of enzymes such as the extraction of the amino acid sequence motifs responsible for the substrate specificity of the enzymes (Suyama *et al.*, 1995). The present paper show the present status of the LIGAND.

2 Result and Discussion

We have made two important extensions of the LIGAND database to reconstruct metabolic pathway from a set of genes coding enzyme which are produced by genome sequencing projects. Metabolic pathway is a network of chemical compounds called metabolites, any consecutive pair of which are substrate and product of an enzymatic reaction. Such reconstruction requires a database that is a collection of metabolites with known enzymatic reactions, through which they are converted to other metabolites. LIGAND has been extended to include chemical data of metabolites and to show a list of enzymatic reactions for each metabolites.

First, LIGAND consists of two sections; ENZYME and COMPOUND. The ENZYME section succeeded most part of the original LIGAND; definition and classification of enzymes and links to protein gene and structural databases. This section newly added two fields; the PATHWAY and GENES fields. The PATHWAY field describes the metabolic pathway in which the enzyme appears and it is a link to the PATHWAY database which is a metabolic and regulatory pathway database in the KEGG project (Goto et al., 1997; Kanehisa 1997). The PATHWAY field has 1,456 links to KEGG/PATHWAY (October 1997). The GENES field contains gene names that encode the enzyme for a number of organisms and they are linked to the KEGG/GENE database which is a gene catalog for each organism. This field has 958 links. The COMPOUND section is actually a new one. This is a collection of chemical and enzymatic data about known metabolites. Chemical data include their chemical names, structure including stereochemistry as a connection table in the MDL MOL file format, figure of structure, atomic component, and CAS registry number. Enzymatic data give a list of enzymes for each metabolite, which is defined as a substrate or a product in their known catalyzed

reactions. The COMPOUND section has 5,075 metabolites with 3,476 chemical formula and 4,337 structures. Metabolites have 4,566 links to the enzymes in the ENZYME section.

Second, all known enzymatic reactions were added to the REACTION field of the ENZYME section. In the previous version of the LIGAND database, the field only contains a single, IUBMB-defined enzymatic reaction, in which substrate and product are definitely declared in most case. In actual metabolism, however, a single enzyme apparently catalyzes a reaction with other substrates and products than those declared by IUBMB. The REACTION field intends to collect all the known enzymatic reactions with any different substrates and products. The REACTION field has extended to contain 2,976 enzymatic reactions, which are increasing with the progress of the construction of the PATHWAY database. With these two extensions, LIGAND is a major component of the KEGG and DBGET/LinkDB systems and serves as an interface between chemical aspects of metabolites and biological aspects of metabolism. Computation of pathway reconstruction can be performed with the extended REACTION field data. The connection of two neighboring enzymes on the metabolic pathway is the result of the common compound that is both the product of the first reaction and the substrate of the second reaction. Thus, the network of enzymes can be computed by generating networks of chemical compounds from a set of substrate-product relationships. Recently, we have developed a path computation tool, PathComp, which computes possible reaction paths for a given set of substrate-product binary relations extracted from the ENZYME entries (Goto et al., 1997).

In conjunction with these efforts, we are implementing the COMPOUND section in the ISIS/BASE database to enable chemical structure and substructure search. The present version of LIGAND still contains several problems to be solved. Genome sequencing projects have revealed that genomes include several copies of genes with sequence similarity (paralogous genes) and that about one quarter of the ORFs code the products functionally unknown. One of the major problems is related to paralogous gene products. Paralogous proteins are known to contribute to tissue- and developmental stage-specific metabolisms or to the biosynthesis of secondary metabolites which characterize biological species. They catalyze the reactions with similar, but different substrates and products in different metabolic reactions in a single organism. Following the classification by IUBMB, however, they belong to a single enzyme species with a common EC number. To reconstruct metabolic pathway, paralogous proteins should be grouped into several subclasses according to their substrate specificity and sequence similarity.

Acknowledgments

This work is supported by a Grant-in-Aid (09272101) for Scientific Research on Priority Areas, "Genome Science", from the Ministry of Education, Science, and Culture of Japan.

References

[1] S.Goto, H.Bono, H.Ogata, W.Fujibuchi, T.Nishioka, T.Sato, M.Kanehisa, "Organizing and computing metabolic pathway data in terms of binary relations," In: *Pacific Symposium on Biocomputing '97*, (Ed by R.B.Altman, A.K.Dunker, L.Hunter, T.E.Klein), pp.175–186, Maui, World Scientific, 1997.

[2] M.Kanehisa, "A database for post-genome analysis," *Trend in Genetics*, 13:375–376, 1997.

[3] M.Suayam, T.Nishioka, J.Oda, "Searching for common sequence patterns among distantly related proteins," *Prot. Engineering*, 147:195–197, 1995.