

# Improvement of the LIGAND Chemical Database

**Rumiko Yamamoto**

rumiko@scl.kyoto-u.ac.jp

**Tomoko Komeno**

ktomoko@scl.kyoto-u.ac.jp

**Susumu Goto**

goto@kuicr.kyoto-u.ac.jp

**Yasushi Okuno**

okuno@kuicr.kyoto-u.ac.jp

**Masahiro Hattori**

hattori@kuicr.kyoto-u.ac.jp

**Minoru Kanehisa**

kanehisa@kuicr.kyoto-u.ac.jp

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho,  
Uji, Kyoto 611-0011, Japan

**Keywords:** chemical compound, chemical reaction, enzyme, database

## 1 Introduction

LIGAND is a composite database comprising three sections: COMPOUND for the information about metabolites and other chemical compounds, REACTION for the collection of substrate-product relations representing metabolic and other biochemical reactions, and ENZYME for the information about enzyme molecules [1, 2]. The concept of the database is a creation of a bridge between bioinformatics and cheminformatics in the postgenome era. We report here the current status of the LIGAND database where efforts are being made to update day by day. The current release (as of September, 2002) includes 10,384 compounds, 5,240 reactions and 3,663 enzymes.

## 2 Current Contents of COMPOUND Entries

The COMPOUND section is manually created by extracting chemical compounds from the metabolic pathways of the KEGG/PATHWAY database [3], and the ENZYME section. Efforts have been made to add more new entries from literature, such as drug-related chemicals including pharmaceuticals or pesticides and phytochemical compounds corresponding to plant secondary metabolites. Table 1 shows the current contents of the COMPOUND section, where the ratios of the first two categories are increasing daily. In addition, we also plan to develop a classification scheme for the compounds based on the new structure similarity measure [5].

Table 1: Contents of the COMPOUND section.

Content	Number
Drug-related chemicals	1,196
Phytochemical compounds*	2,624
Metabolites** and others	6,564
Total	10,384

\* mostly secondary metabolites

\*\* mostly primary metabolites

## 3 Improvement of REACTION Section

Because the LIGAND is a chemical information database involving biological processes, the chemical validation is essential in addition to the biological validation of metabolic pathways. The quality of the REACTION section in the current release has improved through the following two steps based on the law of mass conservation. The improvement as of September 2002 is summarized in Table 2.

### (Step 1) Equivalence of the number of atoms between the substrates and the products

The total numbers of atoms other than hydrogens were compared between the substrates and the products of each reaction. The calculation was automatically performed using the entire RXN files. We have identified and revised nonequivalent reactions such as incorrectly entered chemical structures and insufficient reaction formula with missing or superfluous substrates and the products.

**(Step 2) Definition of the compound binary relationships between the substrates and the products**

We manually defined every possible binary relationship between the substrates and the products in the entire reaction entries [4]. These works allow us to find a number of duplicated reaction entries, such as two similar reactions with the differences of existence or absence of proton. All of the duplicates were merged. Further efforts have been made to update other reaction entries by manually checking compound structures and reaction formula. The structures of the chemical compounds in the COMPOUND section are also substantially improved as a byproduct.

Table 2: The number of updated entries in the REACTION section.

Description	Step 1	Step 2
Reactions updated by the correction of chemical structures	185	**
Merged reactions	*	129
Reactions updated by the correction of reaction formula	261	**
Total	446	129

\* included in the number of reactions updated by the correction of reaction formula

\*\* currently checking

Table 3 shows the current statistics of the COMPOUND and REACTION sections.

Table 3: Statistics of LIGAND (September, 2002).

COMPOUND section		REACTION section	
Content	Number	Content	Number
Entries	10,384	Entries	5,240
Entries with molecular structures	9,612	Links to ENZYME	4,767
Links to ENZYME	4,228	Links to PATHWAY	3,626
Links to PATHWAY	2,471		
Links to REACTION	4,472		

The LIGAND database is accessible through the WWW at: <http://www.genome.ad.jp/ligand/>

**Acknowledgments**

We thank Prof. Takaaki Nishioka for his helpful suggestions, Masaaki Kotera for checking the reaction mechanisms and Yoshinobu Igarashi and Sanae Asanuma for inputting the compound data. We also thank Koichiro Tonomura for developing the interface for searching and updating COMPOUND and REACTION. This work was supported by the grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan, Japan Society for the Promotion of Science, and Japan Science and Technology Corporation.

**References**

- [1] Goto, S., Nishioka, T., and Kanehisa, M., LIGAND: Chemical database for enzyme reactions, *Bioinformatics*, 14:591–599, 1998.
- [2] Goto, S., Okuno, Y., Hattori, M., Nishioka, T., and Kanehisa, M., LIGAND: Database of chemical compounds and reactions in biological pathways, *Nucleic Acids Res.*, 30:402–403, 2002.
- [3] Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A., The KEGG databases at GenomeNet, *Nucleic Acids Res.*, 30:42–46, 2002.
- [4] Kotera, M., Okuno, Y., Hattori, M., Goto, S., and Kanehisa, M., Classification and analysis of KEGG/REACTION database, *Genome Informatics*, 13:490–491, 2002.
- [5] Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M., manuscript in preparation.