

# GLYCAN: The Database of Carbohydrate Structures

**Kosuke Hashimoto**<sup>1</sup>      **Masami Hamajima**<sup>1</sup>      **Susumu Goto**<sup>1</sup>  
khashimo@kuicr.kyoto-u.ac.jp      masami@scl.kyoto-u.ac.jp      goto@kuicr.kyoto-u.ac.jp  
**Shigekazu Masumoto**<sup>2</sup>      **Masayuki Kawasima**<sup>2</sup>      **Minoru Kanehisa**<sup>1</sup>  
masumoto@fqs.fujitsu.com      kawasima@fqs.fujitsu.com      kanehisa@kuicr.kyoto-u.ac.jp

<sup>1</sup> Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

<sup>2</sup> Fujitsu Kyushu System Engineering Ltd., 2-2-1 Momochihama, Sawara-ku, Fukuoka 814-8589, Japan

**Keywords:** carbohydrate structure, carbohydrate database

## 1 Introduction

The structural roles of carbohydrates are important in constructing complex multicellular organs and organisms, which requires interactions of cells with one another and with the surrounding matrix.

In recent years, many structures and functions of complex carbohydrates have become clear. Although it has been necessary to collect and classify these data, there has been nothing that fills the demand. There was a well known database for complex carbohydrates called CarbBank/CCSD (Complex Carbohydrate Structure Database) [2]. Since the fund for the database was discontinued, the database has not been updated and maintained. Under the KEGG project we released the GLYCAN database for carbohydrate structures and associated information in July 2003. GLYCAN is a new addition to the LIGAND collection which also contains COMPOUND, REACTION and ENZYME databases [3, 6]. We have also developed a new drawing tool of branched carbohydrate structures, as well as software to search substructures and similar structures [1]. We report here an overview of the GLYCAN database where efforts are being made to enter and daily update different types of data. The current release (as of October, 2003) includes 10385 entries.

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

## 2 Overview of the Current GLYCAN Entries

Entries in GLYCAN were obtained from three data sources, unique structures computationally identified in CarbBank/CCSD [2], KEGG/PATHWAY database [5] and literature, and by adding new information. Each entry of GLYCAN contains information on the name, the composition of mono sugars, the class of molecules, the links to metabolic pathways, reactions, bound proteins or lipids, lectins, and others. Table 1 shows the numbers of attributes annotated in the entries of GLYCAN.

Table 1: Annotated entries in the Glycan database.

Attribute	Number
NAME	1078
COMPOSITION	10385
CLASS	6325
PATHWAY	202
BINDING	742
REACTION	241

## 3 Drawing Carbohydrate Structures and Searching Similar Structures

### 3.1 Representation of Carbohydrates and their Drawing Tool

Instead of text representation of carbohydrate structures, we have extended KCF (KEGG Chemical Function) format [4] to carbohydrate structures. KCF represents each carbohydrate structure as a

graph and, in this case, mono sugars and glycosidic linkages are represented by nodes and edges, respectively. We developed a carbohydrate structure drawing tool using Java based on this representation (Fig.1). It is possible to enter and edit mono sugars and glycosidic linkages in a ChemDraw or ISIS/Draw style.

### 3.2 Structure Search

GLYCAN provides two graph algorithms for searching similar carbohydrate structures. The users via WWW can select Approximate Match and Exact Match [1] with some options. Fig. 2 shows an example of local Exact Match result with the options for both nodes & edges and the recursive matches. The similarity-score represents the number of matching edges between the query structure and searched one.

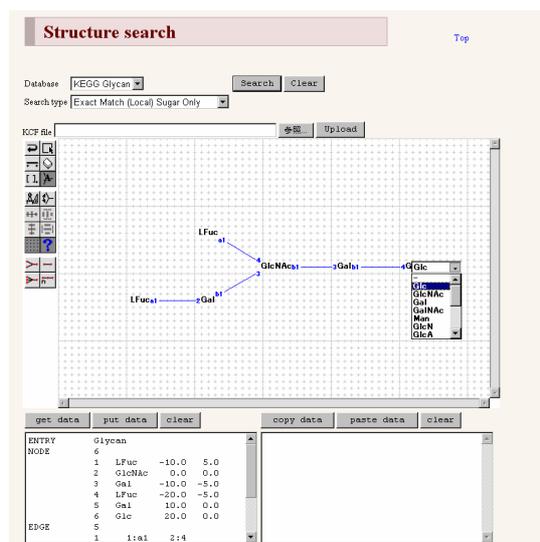


Figure 1: Drawing tool.

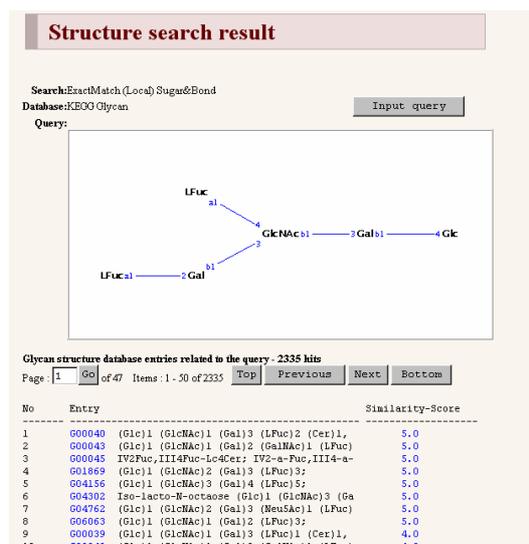


Figure 2: Result of the structure search.

## 4 Acknowledgments

We thank Prof. Tatsuya Akutsu and Dr. Kiyoko F. Aoki for the development of the structural search program, Dr. Yasushi Okuno and Dr. Shin Kawano for their helpful suggestions, and Tomomi Kamiya, Yuriko Matsuura and Junko Nishida for inputting the glycan data. 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] Aoki, F.K., Yamaguchi, A., Okuno, Y., Akutsu, T., Ueda, N., Kanehisa, M., and Mamitsuka, H., Efficient tree-matching methods for accurate carbohydrate database queries, *Genome Informatics*, 14:134–143, 2003.
- [2] Doubet, S. and Albersheim, P., CarbBank, *Glycobiology*, 6:505, 1992.
- [3] Goto, S., Nishioka, T., and Kanehisa, M., LIGAND: chemical database for enzyme reactions, *Bioinformatics*, 14:591-599, 1998.
- [4] Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M., 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, 2003.
- [5] Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A., The KEGG databases at GenomeNet, *Nucleic Acids Res.*, 30:42-46, 2002.
- [6] Yamamoto, R., Komeno, T., and Kanehisa, M., Improvement of the LIGAND chemical database, *Genome Informatics*, 13:492-493, 2002.