Prediction of Glycan Structures from Glycosyltransferase Expression Profiles

Shin Kawano[1]
kawano@kuicr.kyoto-u.ac.jp

Yasushi Okuno[1]
Okuno@kuicr.kyoto-u.ac.jp

Kosuke Hashimoto[1]
khashimo@scl.kyoto-u.ac.jp

Harumi Yamamoto[2]
harumiy@pharm.kyoto-u.ac.jp

Hiromu Takematsu[3]
htakema@pharm.kyoto-u.ac.jp

Yasunori Kozutsumi[2,3]
yasu@pharm.kyoto-u.ac.jp

Susumu Goto[1]
Goto@kuicr.kyoto-u.ac.jp

Minoru Kanehisa[1]
kanehisa@kuicr.kyoto-u.ac.jp

[1] Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan
[2] RIKEN Frontier Research System, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
[3] Graduate School of Biostudies, Kyoto University, Yoshida Shimo-Adachi, Sakyo-ku, Kyoto 606-8501, Japan

Keywords: glycobiology, glycosyltransferase, microarray, glycan structure

1 Introduction

Glycans, which attach to some lipids and Asn/Ser/Thr residues of proteins, draw attentions as the third biological macromolecule next to DNA and protein, since glycans play key roles in the embryogenesis, immunity and diseases. Glycans consist of carbohydrate sugars and their derivatives such as glucose (Glc), mannose (Man), *N*-acetyl-glucosamine (GlcNAc) and *N*-acetyl neuraminic acid (Neu5Ac), and form linear and branched structures. Recent advances of NMR and MASS technologies have made it possible to determine more and more structures of glycans, and our laboratory provides a public glycan structural database, KEGG/GLYCAN [2, 4].

Glycans are synthesized by several kinds of glycosyltransferases, each catalyzing formation of a glycosidic-bond between the glycan precursor as an acceptor and the nucleotide sugar as a donor. In this study, we construct a pattern library consisting of bond-formation patterns of glycosyltransferase reactions in human. Using the database and the library, we try to predict the repertoire of possible glycan structures from the expression data of human glycosyltransferase genes. This is a first attempt of glycoinformatics research linking genome to glycome.

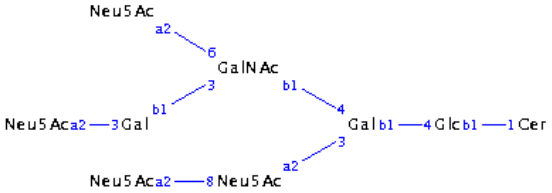
2 Methods

The gene set of human glycosyltransferases was collected from the GENES database in KEGG [5] and each glycosyltransferase reaction was characterized by the three features: the acceptor monosaccharide residue in the glycan chain, the donor monosaccharide and the bond information between them. The time course data of glycosyltransferase gene expressions were from Ramos cells Burkitt lymphoma, stimulated by anti-IgM, and observed by DNA microarrays containing glyco-chain related genes [3]. If the gene expression level is changed compared to control cells, the corresponding glycosyltransferase is checked against the pattern library and the donor-acceptor pairs are identified. These pairs are then searched against KEGG/GLYCAN by KCaM [1], which is the similar glycan structure search program.

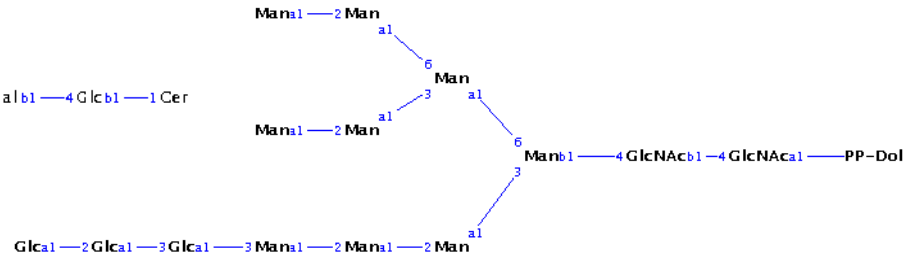
3 Results and Discussion

About 160 glycosyltransferase genes were now annotated in human and the pattern library contained 41 donor-acceptor pairs. Up-regulated glycan structures were predicted from microarray data, 1.5 h, 3.0 h and 5.0 h after the stimulation with anti-IgM. Multiple glycan structures were predicted (1.5 h: 76 structures, 3.0 h: 94 structures, 5.0 h: 83 structures), and they can be classified according to their similarities. Predicted structures from the data of 1.5 h, 3.0 h and 5.0 h after the stimulation can be classified into certain types of glycolipids, glycosamino glycans and N-linked glycans, respectively (Fig. 1). Thus, we conclude we can predict glycan structure profiles in the cell from microarray gene expression data, linking the genome to the glycome.

(a) G00129 (1.5 h)



(c) G00008 (5.0 h)



(b) G00165 (3.0 h)

Figure 1: Example of predicted glycan structures (up-regulate) from microarray data, 1.5 h (a), 3.0 h (b) and 5.0 h (c) after the stimulation with anti-IgM. G number is the glycan structural id number in KEGG/GLYCAN. Gal: galactose, GalNAc: *N*-acetyl-galactosamine, GlcA: glucuronic acid, Xyl: xylose, Cer: ceramide, PP-Dol: dolichol di-phosphate.

Acknowledgments

We thank Masami Hamajima, Tomomi Kamiya, Yuriko Matsuura and Fujitsu Kyushu System Engineering Ltd. for development and maintenance of KEGG/GLYCAN database. Computation time was provided by the Supercomputer Laboratory, Institute for Chemical Research, Kyoto University. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

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] Hashimoto, K., Hamajima, M., Goto, S., Masumoto, S., Kawasima, M., and Kanehisa, M., GLYCAN: the database of carbohydrate structure, *Genome Informatics*, 14:649–650, 2003.

[3] Yamamoto, H., and Kozutsumi, Y., DNA microarrays of glyco-related genes (in Japanese), *Protein, Nucleic Acid and Enzyme*, 48(8):1200–1205, 2003.

[4] `http://www.genome.ad.jp/`

[5] `http://www.genome.ad.jp/ligand/`