Prediction of Glycan Structures from the Glycan Related Microarray Expression Profiles

Shin Kawano kawano@kuicr.kyoto-u.ac.jp
Yasushi Okuno okuno@pharm.kyoto-u.ac.jp
Kosuke Hashimoto khashimo@kuicr.kyoto-u.ac.jp
Takashi Miyama miyama@kuicr.kyoto-u.ac.jp
Susumu Goto goto@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: glycobiology, glycosyltransferase, microarray, glycan structure

1 Introduction

Glycans, which attach to some lipids and to Asn/Ser/Thr residues of proteins, draw attention as the third biological chains next to DNA and protein, since they play a key role in embryogenesis, immunity and diseases. Glycans consist of carbohydrate sugars and their derivatives such as glucose (Glc), mannose (Man), *N*-acetyl-glucosamine (GlcNAc) and sialic acid (Neu5Ac), and form linear and branched structures. Because the sugars in glycans have similar properties and have various linkage patterns, it has been difficult to determine the exact glycan structures (primary sequence). However, recent advances of NMR and MASS technologies have made it possible to determine many new glycan structures, and a public glycan structural database, KEGG/GLYCAN [3,5], was released. This allows us to apply bioinformatics to glycobiology.

While polynucleotide chains (DNA or RNA) and polypeptide chains (protein) are continuously synthesized by the single machinery using DNA or mRNA as a template, glycan chains are synthesized by several kinds of glycosyltransferases, each catalyzing formation of a glycosidic-bond between the glycan precursor as an acceptor and the nucleotide-activated sugar as a donor, without template. Therefore, glycan structure is determined by the combination of glycosyltransferases, glycosidases (glycan degradation enzyme) and supply of the nucleotide-activated sugars. In this study, we construct a pattern library consisting of bond-formation patterns of glycosyltransferase reactions in human. Using the glycan database and the library, we try to predict the repertoire of possible glycan structures from the expression data of human glycosyltransferase genes.

2 Methods

To construct a reaction pattern library, the gene set of human glycosyltransferase was collected from the GENES database 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. Glycan data was extracted from the KEGG/GLYCAN database and the entries with sugars not used in human (mono-sugars not present in the pattern library), were eliminated. The expression data of glycosyltransferase genes was obtained from chorionic carcinoma cell, BeWo, compared with normal placenta cell observed by cDNA microarrays containing glyco-chain related genes [4]. If the gene expression level is changed compared to normal cells, the corresponding glycosyltransferase is searched against the pattern library and the donor-acceptor pairs are identified. These pairs are then searched against the GLYCAN database using KCaM [1], a similar glycan structure search program.

3 Results and Discussion

About 160 glycosyltransferase genes are annotated in human and the pattern library contained 41 donor-acceptor pairs. The 6353 entries of human glycans were extracted from the KEGG/GLYCAN database (10640 entries). Thirty-five of up-regulated glycan structures were predicted from microarray data, some examples are shown in Fig. 1.

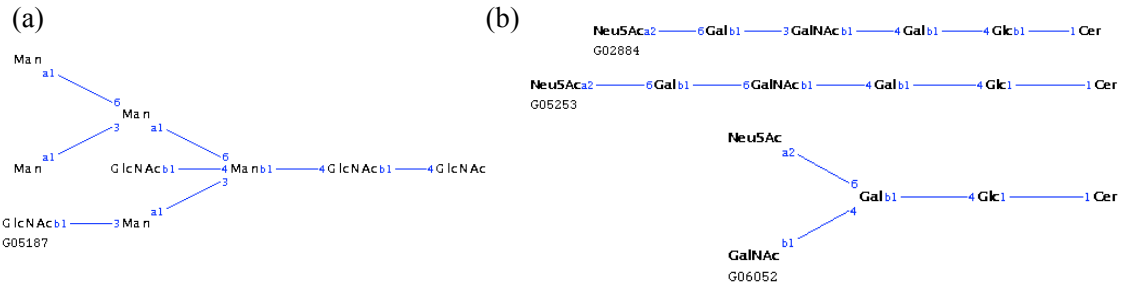

Figure 1: Examples of discriminative glycan structures of chorionic carcinoma cell, BeWo. (a): glycan containing bisecting GlcNAc (GlcNAc b1-4 Man), (b): glycolipids containing sialic acid(s). G number is the glycan id number in KEGG/GLYCAN. Gal: galactose, GalNAc: *N*-acetyl-galactosamine, Cer: ceramide.

It is known that cancer cells produce different glycans than those of normal cells, and the presence of these glycans is actually taken as a marker of cancer. In this study, we try to predict discriminative glycans in chorionic carcinoma using glycosyltransferase expression profiles. Thirty-five glycan structures were identified, including N-type glycan with bisecting GlcNAc (Fig. 1a.). It has been experimentally demonstrated that this type of glycan was produced by BeWo cell [2]. This suggests that it is possible to predict the glycan structures synthesized actually by this method. In addition, glycans that have sialic acid in their non-reducing end were predicted (Fig. 1b). It is known that the sialic acid content is increased cancer cells, this prediction result agrees with this fact.

Acknowledgments

We thank Masami Hamajima, Tomomi Kamiya, Yuriko Matsuura, Kana Matsumoto, Atsuko Yano, Ami Matsuzawa and Fujitsu Kyushu System Engineering Ltd., for development and maintenance of KEGG/GLYCAN database. We also thank Harumi Yamamoto, Dr. Hiromu Takematsu and Prof. Yasunori Kozutsumi for provision of Microarray data. Computational 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] Hard, K., Damm, J.B., Spruijt, M.P., Bergwerff, A.A., Kamerling, J.P., Van Dedem, G.W., and Vliegenthart, J.F., The carbohydrate chains of the beta-subunit of human chorionic-gonadotropin produced by the choriocarcinoma cell-line BeWo – novel O-linked and novel bisecting-GlcNAc-containing N-linked carbohydrates, *Eur. J. Biochem.*, 205:785-798, 1992.

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

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

[5] http://glycan.genome.ad.jp/