

Probability-based Glycosphingolipid Prediction from Mass Spectrometry Data

Kazuhito YOKOI¹
gsltools@zuhito.com

Saki ITONORI²
itonori@sue.shiga-u.ac.jp

Mutsumi SUGITA²
sugita@sue.shiga-u.ac.jp

Masahiro ITO¹
maito@sk.ritsumei.ac.jp

¹ Graduated School of Science and Engineering, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

² Faculty of Arts and Education, Shiga University, 2-5-1 Hiratsu, Otsu, Shiga 520-0802, Japan

Abstract

We propose an improvement computer method based on a scoring algorithm for predicting the glycosphingolipid structure from MALDI-TOF MS spectra in PSD mode. This algorithm is probability-based, which has several advantages, e.g. simple rules, simple comparison and simple optimization. Tests of the theoretical and observed data showed high accuracy. Furthermore, the permethylated glycosphingolipid was shown to be very effective for this method. This method which can easily predict a GSL structure from a very little sample could be a powerful tool for genome-wide analysis of the glycosphingolipid structures.

Keywords: glycosphingolipid, chemical structure, MALDI-TOF MS, metabolism

Availability: The website is freely accessible at
<http://www.gsltools.sk.ritsumei.ac.jp/>

1 Introduction

Glycosphingolipid (GSL) is a large group of biomolecules containing two basic structural units: a hydrophilic carbohydrate chain and a non-polar ceramide moiety. GSLs are membrane components and play important roles in a wide variety of cellular functions, for instance cell-cell recognition, cell-cell interaction, cell growth, differentiation, transmembrane signaling and apoptosis.

GSLs are secondary products of genes, which are biosynthesized through metabolism. The general pathways for its metabolism are reasonably well characterized and their regulation is beginning to be understood through the development of specific inhibitors and identification of the genes for many of these enzymes. Although there is very little GSL content in living organisms, high-throughput analysis of GSLs is necessary to understand their metabolism network and their functions. Recently, such analysis has been enabled by the rapid progress of mass spectrometry equipment. With the development of computer tools for identification from mass spectra data as well as the above-mentioned progress of mass spectrometry, proteomics has attracted attention. However, in glycomics, the computer tools have been poor. Therefore, we developed computer tools for GSL structure prediction from MALDI-TOF MS data [1].

We propose an improvement computer algorithm for prediction of GSLs from matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) spectra in their post source decay (PSD) mode. The scoring algorithm is probability-based, which has several advantages, e.g. simple rules, simple comparison and simple optimization. The tests of theoretical and observation data showed high accuracy. This method could be a powerful tool for genome-wide analysis of the GSL structures.

2 Method

2.1 Algorithm

The GSL structure is predicted by probability-based scoring. The procedure is as follows: (1) Candidate compositions of GSL, carbohydrate chain and ceramide, are predicted from the precursor mass in MALDI-TOF MS in the reflector mode. Here, we adopted two conditions (parameters) as follows: (a) Long

chain base is composed only the di- or tri-hydroxyl type, and (b) Fatty acid is composed only the normal or hydroxyl type (normal or hydroxyl). (2) Candidate structures of GSL are found from candidate compositions. (3) All fragment structures are calculated from each candidate structure. A predicted mass spectrum simulates by the appearance probabilities of each fragment. (4) In each candidate structure, theoretical masses from the fragment structure are matched with observed masses by MALDI-TOF MS in the PSD mode. After this matching, candidate structures are scored. Finally, all candidate structures are sorted in descending order.

2.2 Samples

The samples were six GSLs, two of a branch and four of linear carbohydrate chain containing GSLs. The accuracy of this method was evaluated using these samples.

2.3 MALDI-TOF MS

MALDI-TOF MS analysis is performed using an Applied Biosystems/Voyager DE STR™ Biospectrometer with a nitrogen laser (337 nm). Samples are analyzed in delayed extraction mode, post source decay (PSD), using an acceleration voltage of 20 kV, and a grid voltage of 78%.

2.4 Permethylated of GSL

The GSL was permethylated using the Ciucanu-Kerek method [2]. The GSL was permethylated with NaOH and methyl iodide in DMSO for 2 min.

3 Results

3.1 Structural prediction of intact GSL

This method was tested using six intact GSLs. As a result, all correct GSL structures were ranked at the top on the predicted structures. However, in two GSLs, L5 and L7, several structures include a correct structure were predicted as the same scoring. This may be because any structurally specific fragments were not observed by MALDI-TOF MS analysis, especially in GSL structure with a long linear carbohydrate chain.

3.2 Structural prediction of permethylated GSL

In the permethylation of GSL, the number of the methyl group is different at the monosaccharide position in the carbohydrate chain. For example, non-reduced terminal and branched hexose are substituted to four methyl groups, hexose in the middle of linear carbohydrate chain to three, and trunk hexose to two, respectively. Therefore, we decided to apply this method to the permethylated GSL to obtain specific fragments. This method was tested for two structures, L5 and L3, which were not predicted with intact GSL alone. As a result, these structure predictions were at the top ranked, respectively.

4 Discussions

In this study, we developed a novel method for predicting GSL structure from MALDI-TOF MS data automatically. Furthermore, we showed that the analysis of permethylated GSL is very effective. This method is able to identify a branched monosaccharide and a trunk monosaccharide with the molecular weight, such as a methyl deoxy-hexose and a hexose, which was impossible in intact GSL because the permethylation adds methyl groups depending on the number of hydroxyl groups on a monosaccharide.

This present method could use the permethylated GSL as an intermediate material in the alditol acetate method. These two methods are complementary, overcoming the disadvantages of each method. Therefore, this method could be a powerful tool for high throughput put data analysis such as genome-wide analysis of the GSL structures.

References

- [1] Matsumuro, Y, Itonori, S., Sugita, M. Ito, M. Prediction of glycosphingolipid chemical structure from MALDI-TOF MS data, The 17th international conf. Genome informatics 2006, P114-1-2
- [2] Ciucanu, I, Kerek, F, A simple and rapid method for the permethylation of carbohydrates, Carbohydrate research, (1984) 131, 209-217.