Development of Bioinformatics Support for High Throughput Isomeric Separation and the Structural Identification of Glycans by LC-MS

Ningombam Sanjib Meitei[1], Arun Apte[2], Udayanath Aich[3], Julian Saba[4]
[1]PREMIER Biosoft, Indore, India, [2]PREMIER Biosoft, Palo Alto, U.S.A, [3]Thermo Fisher Scientific, Sunnyvale, CA, [4]Thermo Fisher Scientific, San Jose, CA
Corresponding author e-mail: sanjib@premierbiosoft.com

Introduction

The separation of glycans by chromatography prior to MS analysis can reduce sample complexity, minimize ion suppression, and increase dynamic range and separation of structural isomers. Recent developments in mixed-mode columns and faster scanning mass spectrometers have increased the number of glycans resolved and identified by LC-MS. However, these advancements lead to large data sets. Additionally, chromatograms of isomeric glycans are complex with some isomers co-eluting under a single peak. Manual deconvolution of such complex chromatograms, identification of isotopic peak components, identifying MS/MS scans for detected compounds and selection of correct precursor m/z values from the isotope cluster for MS/MS data analysis is time consuming. In addition to these challenges, the accuracy of rapid identification of glycans in high throughput manner has been hampered by lack of glycan templates. Therefore, we have developed a software tool to streamline this process. We also have developed web based software modules which facilitates users to store glycans along with retention times.

Methods

Sample Preparation

N-Linked glycans were released from glycoproteins (Bovine Fetuin) with PNGase F enzyme (New England BioLabs). The released glycans were labeled with 2-aminobenzamide (2AB) with slight modification from the reported procedure of Bigge et. al., [1]. Prior to analysis, samples were dissolved in 100 µL D.I. water in a 250 µL auto sampler vial.

Liquid Chromatography

All glycans were separated on a Thermo Scientific™ GlycanPac™ AXR-1 (1.9 µm, 2.1 × 150 mm) column [2] by a Thermo Scientific™ Dionex™ Ultimate™ 3000 UHPLC instrument with either a fluorescence or MS detector.

Mass Spectrometry

MS analysis was performed using a Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass spectrometer and a Q Exactive™ mass spectrometer in negative ion mode. LC-MS^2 experiments were conducted for structural elucidation.

Data Analysis

SimGlycan® 5.4 software (PREMIER Biosoft) was used for LC-MS and MS/MS data analysis.
Thermo Scientific™ Xcalibur™ software is also used to visualize raw LC-MS data.

Results and Discussion

Figure 1 shows the schematic representation of SimGlycan software.

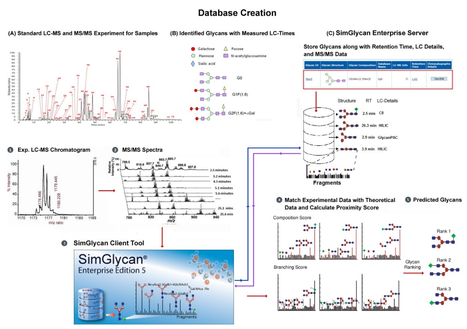

Figure 1: Structure of the TG(18:1(9Z)/18:1(9Z)/18:1(9Z))

Scoring Mechanism

SimGlycan predicts and scores candidate glycans in order of most probable match by using a proprietary scoring algorithm that is based on structure specific diagnostic ions observed in the MS/MS data [3]. The typical procedure is that for each MS/MS scan, SimGlycan creates a list of candidate structures by using precursor m/z. The program compares the *in silico* fragment ions of each candidate against the observed product ions which provides the basis of the scoring algorithm.

Custom LC-MS/MS Glycan Templates

In addition to the scoring algorithm, in an effort towards rapid identification of glycans, we have developed web based software modules which facilitates users to store glycans along with retention time.

Users can add glycans into database by importing corresponding KEGG Chemical Function (KCF) format file in batch mode (Figure 2). In case KCF file is not readily available, users can draw the structure in SimGlycan software's "Draw" module and save the corresponding KCF data (Figure 3).

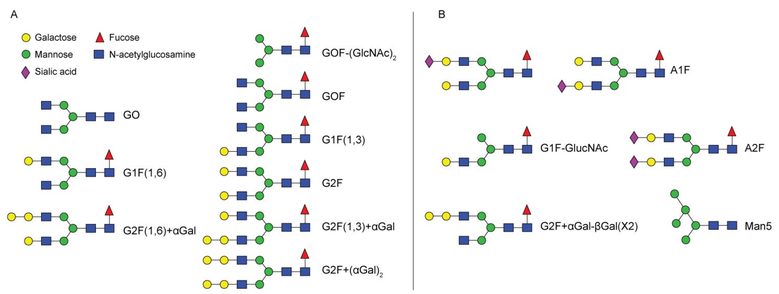

Figure 2: A prototype template of glycans

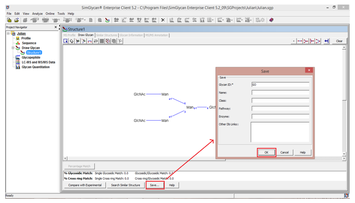

Figure 3: A typical SimGlycan software "Draw" interface in which glycans can be drawn and structure can be saved as KCF file format

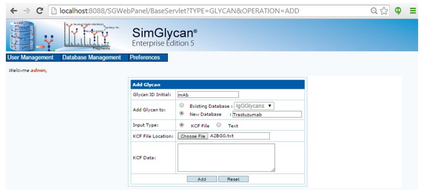

Figure 4: SimGlycan software interface to add glycans into an existing or new database

Additional functionality such as browse glycans by searching from taxonomy or structure search, move/copy data from a database to another database, edit information of existing glycans etc. are allowed. Figure 4 shows a typical SimGlycan web browser showing a list of searched glycans. In order to store retention times corresponding to glycans, just click "Add" (Figure 5) and enter (multiple) detailed information from LC-experiment/s (Figure 6).

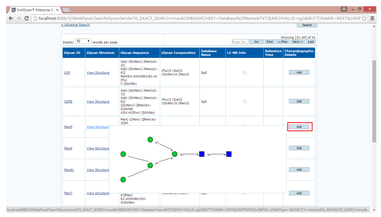

Figure 5: A typical SimGlycan software web page displaying searched glycans

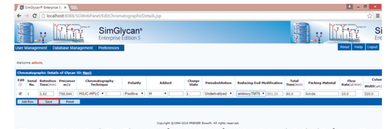

Figure 6: A typical SimGlycan software interface to store detailed information from LC-Experiments

Once the database is constructed, users can restrict database search to custom database/s that contains curated data, thereby increasing the confidence of identification (Figure 7).

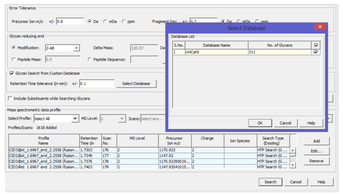

Figure 7: A typical SimGlycan software interface facilitating users to specify retention times as initial database search predicate

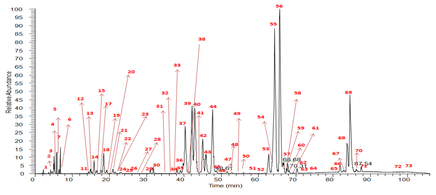

Figure 8: LC-MS analysis of 2AB-labeled *N*-linked glycans from bovine fetuin by GlycanPac AXR-1 column with MS detection

Support LC-MS and MS/MS Workflows

Software modules were developed for automatic detection of compounds, deconvolution of chromatograms to separate glycan isomers, identification of isotope clusters and MS/MS scans corresponding to detected compounds and precursor *m/z* selection. MS/MS data was subjected to the program for automatic structural identification of the detected LC peaks. The accuracy of the results was also tested on a Q Exactive mass spectrometer (data not shown). Key features of the software are explained using screenshots of the graphical user interfaces and results that we have obtained from the LC-MS^2 data analysis.

Highest Number of Resolved Peaks

The LC-MS profile of the GlycanPac AXR-1 column showed the highest number of resolved peaks (≥ 70) for bovine fetuin glycans ever achieved (Figure 8), as compared to the commercially available columns . Generally, a single LC peak can have many structural isomers. In many instances, the use of commercially available columns with poor capability to resolve isomers results in mixed MS2 spectrum that contains fragment ions from multiple glycans making it extremely difficult to assign correct structures.

Need for Sophisticated Bioinformatics Tool

The GlycanPac AXR-1 column with its ability to resolve structural isomers introduces complexity to analysis. Namely,

- **Higher number of MS/MS scans**: Far more MS/MS spectra are triggered in a single LC-MS^2 analysis.
- **Complex chromatogram peaks**: Even with higher LC peak resolving capability with the GlycanPac AXR-1 column, isomeric glycans still exhibit complex LC peaks with some isomers co-eluting under a single peak (e.g., peaks observed after 65 minutes in Figure 9(C)).

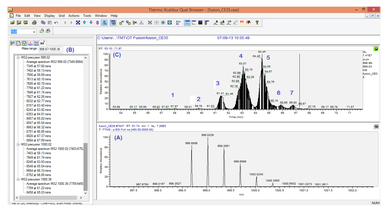

Figure 9: Xcalibur software window showing (A) Isotope cluster with base peak at M+1 peak; (B) MS/MS scans with precursor *m/z* values corresponding to M+1, M+3 and M+4 and (C) LC-peaks that may be isomers of the of 2-AB labelled *N*-Glycan with precursor m/z value 998.690 (charge: 3-)

- **MS/MS scans triggered for higher isotopes:** Unlike isotope cluster for peptides, glycan isotopic clusters do not necessarily have base peak at monoisotopic *m/z* (Figure 9 (A)). For example, Figure 9 (B) shows the Xcalibur software windows displaying MS/MS scans acquired for the 2-AB labelled *N*-Glycan with mass of 2897.011Da. All the MS/MS scans have been triggered for the *m/z* values corresponding to M+1 i.e., *m/z* 999.023, M+3 i.e., *m/z* 1000.023and M+4 i.e., *m/z* 1000.363 peaks of the isotope cluster. Precursor *m/z* needs to be corrected before performing MS/MS data analysis**.**

LC MS and MS/MS Data Pre-processing

SimGlycan software simplifies the LC-MS data processing by completing it within a few clicks of buttons. Just select the instrument series e.g., Orbitrap Fusion from the template name and click OK (Figure 10). Entire process will be completed within a few minutes and a peaklist will be generated (Figure 11). Figure 12 shows a typical SimGlycan software interface showing Search Parameter dialog. The list of MS/MS scans with corrected precursor m/z values is shown in the dialog. After performing MS/MS database search, identified glycans are listed in the result pane. Identified glycan structures can be sorted based on retention time, intensity and other parameters. Structures, fragments and other information can be viewed in a single workspace (Figure 13, top left). The MS/MS spectra will be automatically annotated with the fragment ions of identified structures using three modes of annotation namely Domon and Costelo fragmentation nomenclature, successive loss of monosaccharide residues and cartons showing moveable fragment structures (Figure 13).

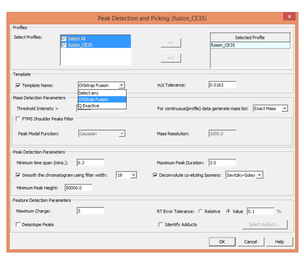

Figure 10: A typical SimGlycan software interface to specify LC-MS data processing parameters

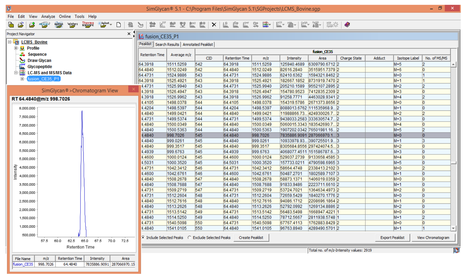

Figure 11: A typical SimGlycan software interface showing generated peaklist

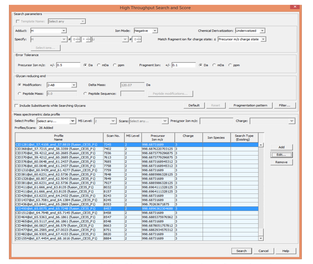

Figure 12: A typical SimGlycan software interface showing Search Parameter dialog. The list of MS/MS scans with corrected precursor m/z values is shown in the dialog.

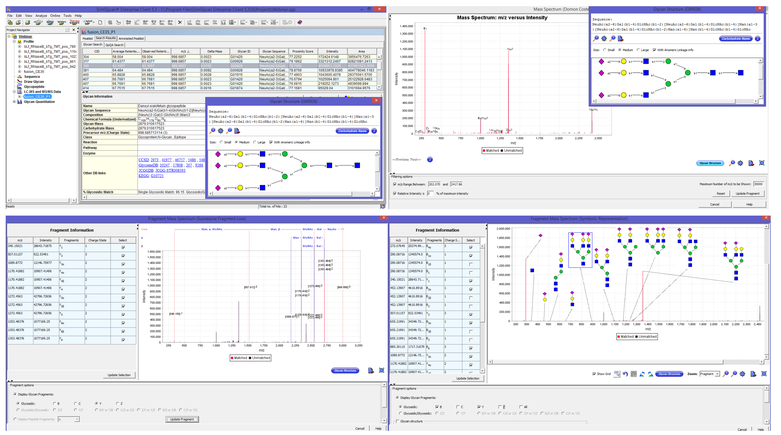

Figure 13: SimGlycan software search result pane (top left); MS/MS peak annotation with Domon and Costelo fragment nomenclature (top right); MS/MS peak annotation showing successive loss of monosaccharide residues (bottom right); MS/MS peak annotation with cartoons (bottom right)

Portable Reports

Spreadsheet based reports facilitate easy reviewing of results for further verification, downstream analysis and dynamic information sharing. One major challenge with spreadsheet based report format is to save glycan structures into spreadsheet cells so that information such as retention time, precursor *m/z*, glycan ID etc. can be processed using spreadsheet operations to further organize the results. SimGlycan software generates report in MS excel file wherein glycan structures are also exported along with other structure specific information (Figure 14).

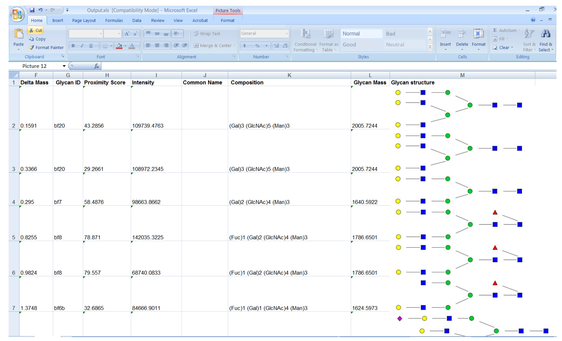

Figure 14: Typical MS excel file containing results outputted by SimGlycan software after performing MS/MS database search

Conclusion

- SimGlycan® 5.4 software provides informatics support for LC-MS and MS/MS data analysis by enabling users to create LC-MS glycan templates.
- SimGlycan® 5.4 software facilitates analysis of 10000 MS/MS scans in a batch for structural identification of glycans. Multiple batch searches can be triggered simultaneously. Finally, results including glycan structures can be exported into MS excel file facilitating easy review of results as well as dynamic sharing of information for further post identification data analysis.

Reference

1. Bigge, J. C. et al., Anal. Biochem. 1995, 230, 229-238.
2. Thermo Scientific GlycanPac AXR-1 Product Specification: http://www.dionex.com/en-us/webdocs/114170-PS-GlycanPac-AXH1-Column-PS20695_E.pdf
3. Apte A., Meitei N. S. (2010) In Functional Glycomics, Humana Press; p. 269-81.
4. Chester, T.L., Anal. Chem. 2013, 85 (2), 579-589.
5. Ruhaak, L.R. et al., Anal. Bioanal. Chem. 2010, 397, 3457-3481.
6. Application Note 20786: Structural Analysis of Labeled N-Glycans from Proteins by LC-MS/MS Separated Using a Novel Mixed-Mode Stationary Phase.
7. Application Note 20827: Structural Analysis of Native N-Glycans Released from Proteins Using a Novel Mixed-Mode Column and a Hybrid Quadrupole-Orbitrap Mass Spectrometer.