Informatics Support for Isomeric Separation and the Structural Identification of Labeled *N*-Glycans from Proteins

Ningombam Sanjib Meitei1; Arun Apte2; Udayanath Aich3; Julian Saba4

1PREMIER Biosoft, Indore, India; 2PREMIER Biosoft, Palo Alto, US; 3Thermo Fisher Scientific, Sunnyvale, CA; 4Thermo Fisher Scientific, San Jose, CA

Corresponding author e-mail: sanjib@premierbiosoft.com

1

Overview

Purpose: Develop a software to facilitate LC-MS and MS/MS glycan data that have been acquired for separation and identification of various complex *N*-linked glycans from proteins using a novel high resolution mixed-mode column and an Orbitrap Fusion Tribrid mass spectrometer.

Introduction

The separation of glycans by chromatography prior to MS analysis can provide number of benefits. Primarily, separation can reduce sample complexity, minimize ion suppression, increase dynamic range of analysis and provide separation of structural isomers. Recent developments in mixed–mode column chemistries and faster scanning mass spectrometers have increased the number of glycans resolved and identified by LC-MS workflows. However, the increase in peaks resolution and detected compounds can 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 peaks 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 task. Therefore, we have developed a software tool (SimGlycan®) to streamline this process.

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-MS2 experiments were conducted for structural elucidation.

Data Analysis

SimGlycan® 5.1 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.

2

Results and Discussion

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 1), as compared to the commercially available columns . Orbitrap Fusion with its wide dynamic range and ultra high mass resolution of makes it the ideal platform for looking deeper into the glycome and confidently identifying low-abundance glycans. Overall, 135 unique glycan structures were identified using a combination of GlycanPac AXR-1 column and Orbitrap Fusion mass spectrometer.

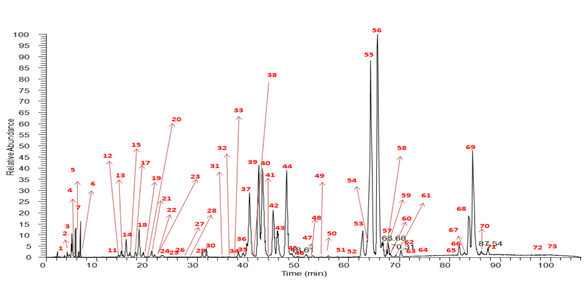

FIGURE 1. LC-MS analysis of 2AB-labeled *N*-linked glycans from bovine fetuin by GlycanPac AXR-1 column with MS detection.

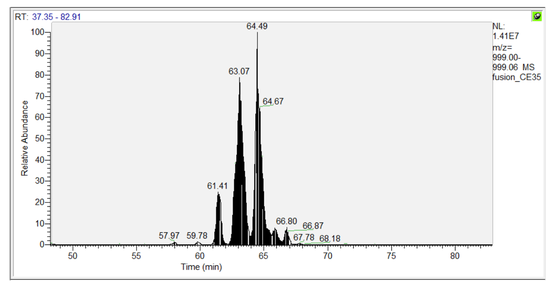

Figure 2: Xcalibur software showing extracted ion chromatogram of 2-AB labelled *N*-Glycan with precursor m/z value 998.68 (charge: 3-).

Resolving Structural Isomers

The GlycanPac AXR-1 column can resolve structural isomers (Figure 2) much better than most existing commercial stationary phases. 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 contain 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 need to be triggered in a single LC-MS2 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 3(C)).

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 3 (A)). For example, Figure 3 (B) shows the Xcalibur software windows displaying that MS/MS scans acquired for the 2-AB labelled *N*-Glycan with mass of 2897.0106 Da. 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 1000.023-and 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.

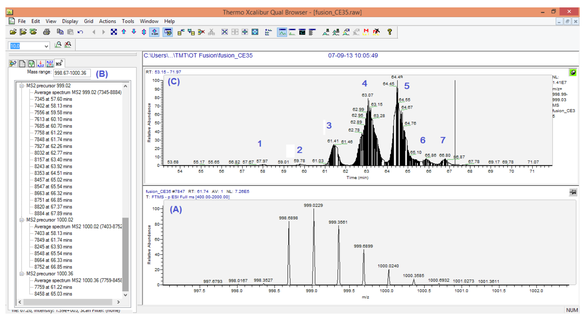

Figure 3: Xcalibur software window showing (A) Isotope cluster with base peak at M+1 peak; (B) MS/MS scans with precursor *m/z* values corresponding to M+1, M+3 and M+4 and (C) LC-peaks that may be isomers of the of 2-AB labelled *N*-Glycan with precursor m/z value 998.68 (charge: 3-).

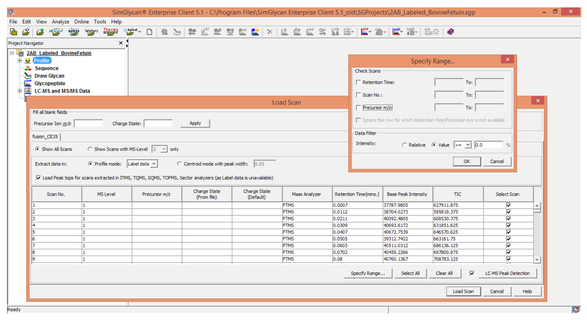

Figure 4: Shows a typical graphical user interface of SimGlycan® software facilitating loading of data from selected ranges of scans, retention time, precursor *m/z* values and intensity threshold values.

SimGlycan® 5.1 Software to 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 were also tested on a Q Exactive mass spectrometer (data not shown). All isomeric glycans correctly detected, separated and identified by the program were manually validated for these experiments.

Key features of the software are explained using screenshots of the graphical user interfaces and results that we have obtained from the LC-MS2 data analysis.

Import Data

User can directly import raw data from native Thermo Scientific Xcalibur RAW files (Figure 4).

4

LC MS and MS/MS Data Pre-processing

SimGlycan provides a simplified module for pre-processing the LC-MS and MS/MS glycan data. The pre-processing steps includes peak detection, chromatogram deconvolution to separate structure isomers which are observed under single LC peak, identification of isotopic envelops, clustering of MS/MS scans corresponding to LC peaks and selection of accurate precursor *m/z* values for MS/MS scans from isotopic envelop. Just select the instrument series e.g., Orbitrap Fusion from the template name and click OK. All the process will be completed within few minutes.

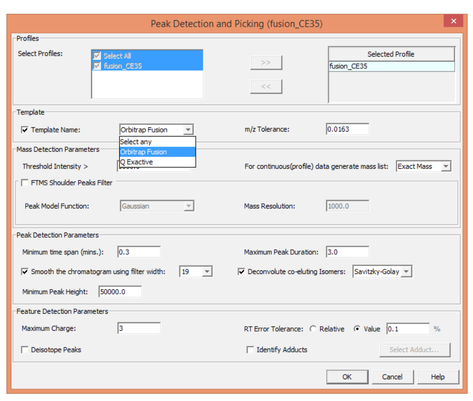

Figure 5: Shows a typical graphical user interface of SimGlycan software facilitating selection of peak detection and picking templates based on instruments.

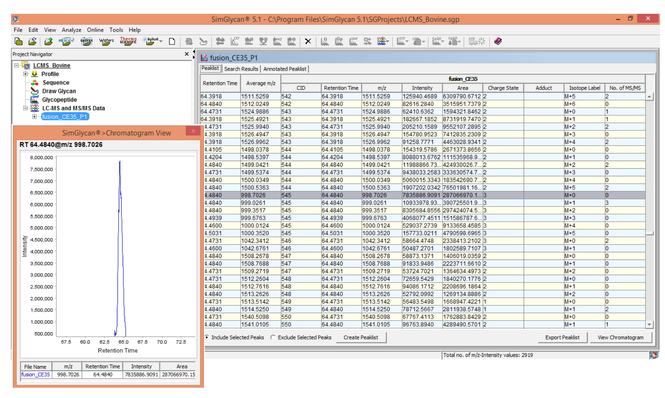

Figure 6: SimGlycan software showing detected LC-peaks; information such as retention time, *m/z*, compound ID (CID), peak height (intensity), area, adduct, charge state, isotope level and number of MS/MS scans triggered for each detected peak is displayed.

On completion of the data pre-processing, a peak list table will be generated (Figure 6). All the LC peaks corresponding to an isotope envelop will be automatically clustered under a compound ID number (Figure 6). Besides, MS/MS scans that were triggered for peaks corresponding to higher isotopes will have their precursor *m/z* values replaced by monoisotopic peak *m/z* values. Figure 7 shows the corrected precursor *m/z* values for scans with incorrect precursor *m/z* values that are shown in Figure 3(B).

5

MS/MS Data Analysis in Batch Mode

SimGlycan can perform MS/MS database search for 10000 scans in a batch. Furthermore, multiple batches can be triggered simultaneously. User has to specify the search parameters (Figure 7) and also specify filters (Figure 8) to narrow down the search. On completion of the search, user can view identified glycan structures (Figure 9). Results can be sorted on the basis of any of the columns in result pane. MS/MS spectra are automatically annotated using various annotation techniques (Figure 10).

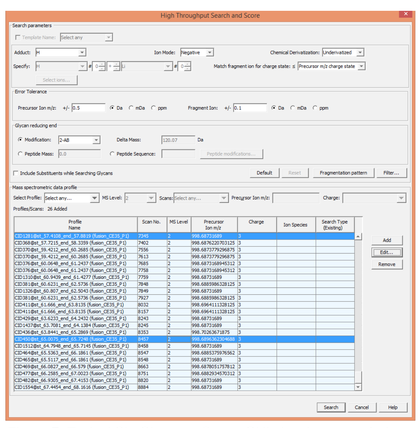

Figure 7: Typical SimGlycan software GUI showing corrected precursor *m/z* values for the MS/MS scans triggered for higher isotopic peaks. All the MS/MS scans with incorrect precursor *m/z* values in Figure 3(B) now have 998.683- as the precursor *m/z* value.

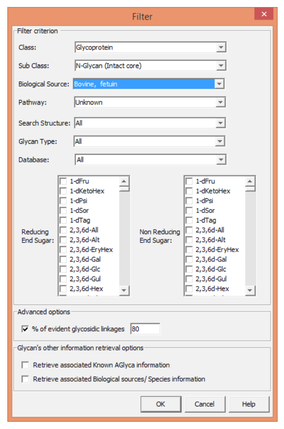

Figure 8: SimGlycan software dialog showing filters to reduce false positives.

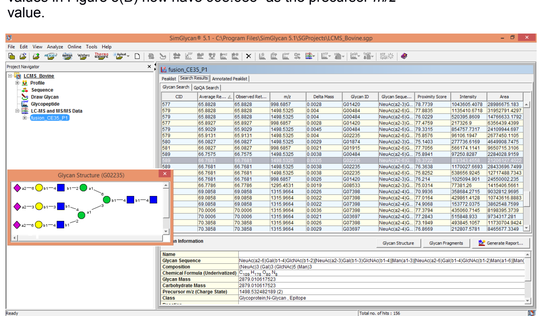

Figure 9: SimGlycan software search result pane; results sorted based on retention time. Identified glycan structures, corresponding fragment ions can be viewed besides corresponding database information.

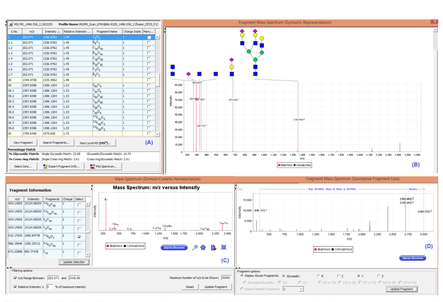

Figure 10: SimGlycan software MS/MS spectrum list/spectral annotation; (A) MS/MS spectrum list annotated with matched fragment structures, (B) MS/MS spectrum annotated with symbolic structures of fragment ions, (C) MS/MS spectrum annotated with name of fragment ions using Domon-Costello nomenclature, and (D) MS/MS spectrum annotated with successive loss of carbohydrate residues between peaks.

6

Portable Reports

Spreadsheet based reports facilitates 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 to information such as retention time, precursor *m/z*, glycan ID etc. can be processed using spreadsheet operations to further organize the results.

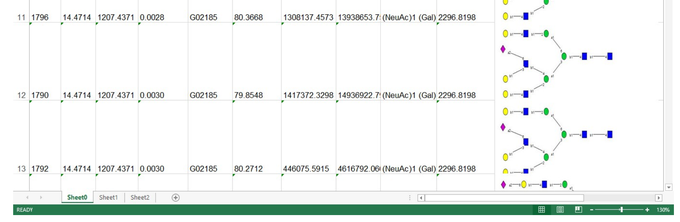

Figure 11: Report in MS excel file wherein glycan structures are also exported along with other structure specific information.

Conclusion

- The GlycanPac AXR-1 column separates glycans based on charge, size, polarity and isomeric structure, providing a greater number of resolved peaks compared to commercial amide HILIC columns for 2AB-labeled bovine fetuin *N*-linked glycans.
- Faster Orbitrap enables higher scan rates at higher resolution. This translates to increased sensitivity and better quality MS/MS data for both abundant and low abundance glycans.
- SimGlycan® 5.1 software provides informatics support for LC-MS and MS/MS data analysis. It facilitates peak detection, separation of isomers that elute under a LC-peak, precursor m/z selection from isotope cluster, identification of glycan structures using MS/MS database search.
- SimGlycan® 5.1 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.

References

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.Chester, T.L., Anal. Chem. 2013, 85 (2), 579-589.
4.Ruhaak, L.R. et al., Anal. Bioanal. Chem. 2010, 397, 3457-3481.
5.Chen, X. et al., Anal. Biochem. 2007, 370, 147-161.
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.