Development of Bioinformatics support for Quantitative Glycomics using Tandem Mass Tags

Ningombam Sanjib Meitei[1]; Arun Apte[2]; Sergei Snovida[3]; Julian Saba[4]; John Rogers[3]
[1]PREMIER Biosoft, Indore, India; [2]PREMIER Biosoft, Palo Alto, CA, USA; [3]Thermo Fisher Scientific, Rockford, IL, USA; [4]Thermo Fisher Scientific, San Jose, CA, USA

PREMIER Biosoft

Introduction

AminoxyTMT reagents enable relative quantitation of glycans in a MS/MS spectrum. However, there are no bioinformatics tools to automate the qualitative/quantitative analysis. We have implemented features within SimGlycan® software that enables identification, quantitation of glycans and differential analysis across biological samples from complex mixture.

Methods

Tryptic digests of several standard glycoproteins and monoclonal antibodies were treated with PNGase F glycosidase to release N-glycans. Following reversed-phase and graphitized carbon extraction/desalting steps, the glycans were labeled with aminoxyTMT reagents (Figure 1) according to an optimized protocol.

Labeled glycans were analyzed using Velos Pro mass spectrometer (Thermo Scientific™) in the positive ion mode to acquire trap-HCD MS/MS fragmentation. Data was subjected to SimGlycan to identify structure of the sample glycans. Post identification, we assigned TMT 130.13 ion as the control. The program allows custom setting of reporter ion correction factors, calculates sum/average/ median of the reporter ion intensities, number of MS/MS spectra which identifies a glycan. Different types of chart facilitate glycan differential analysis across samples.
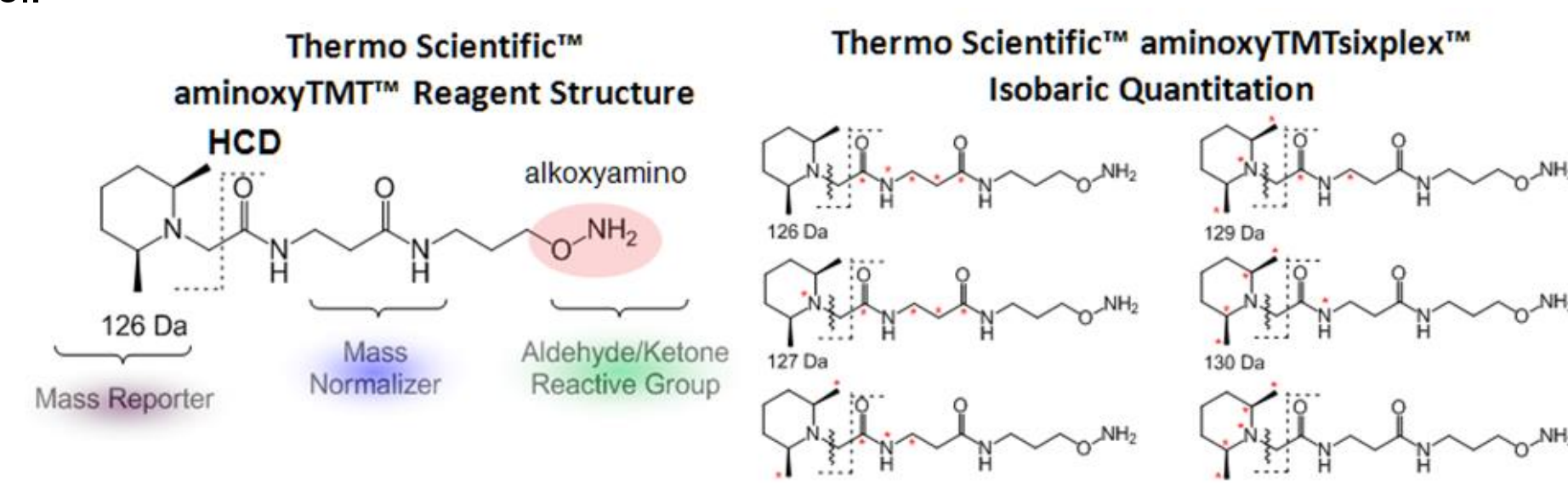

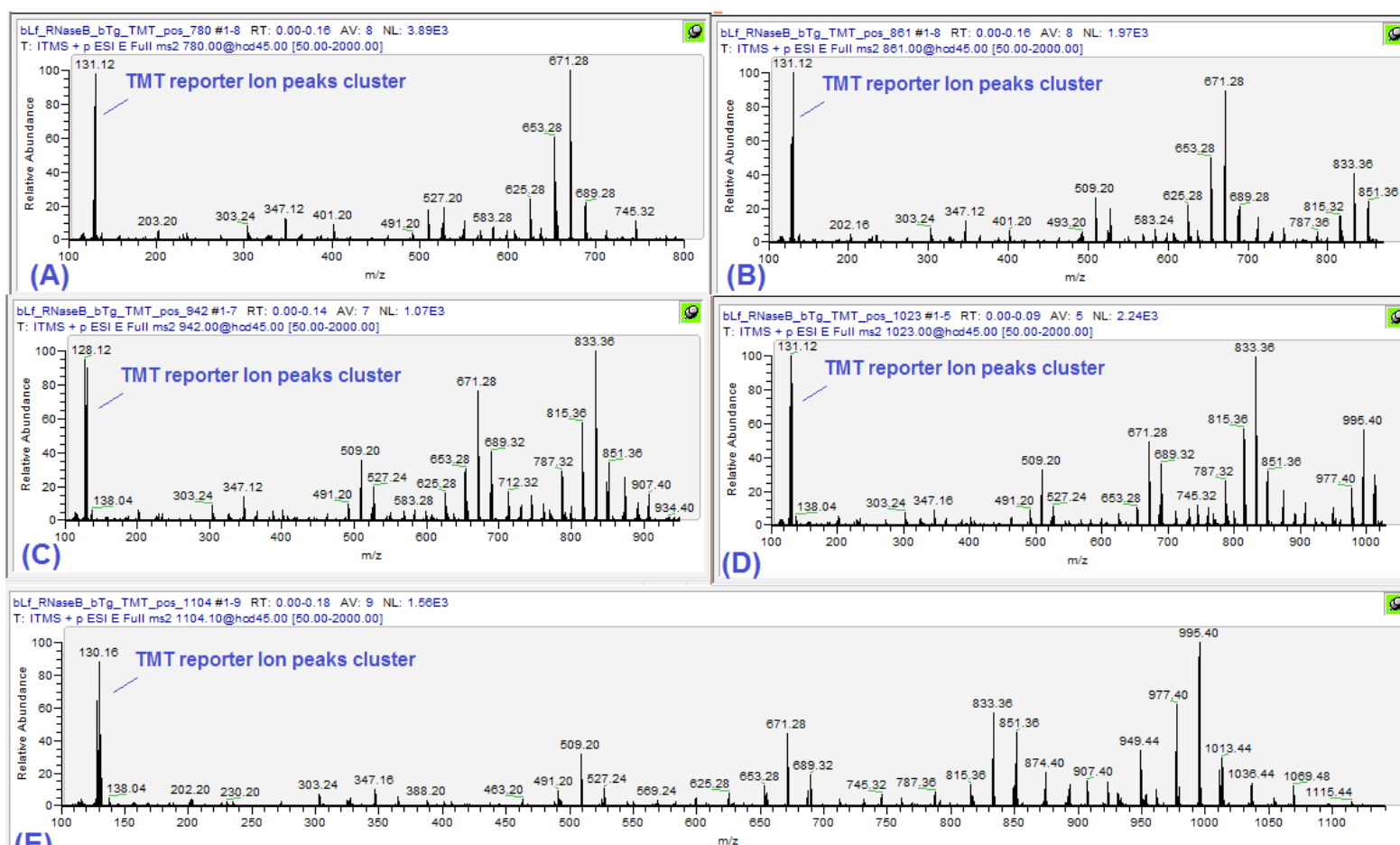Figure 1: aminoxyTMT Reagent Structure and corresponding reporter ions.

Results

HCD spectra were acquired on five high-mannose glycan precursors m/z 780, 861, 942, 1023 and 1124 m/z. Figure 2 shows the averaged MS/MS spectra for each precursor. TMT reporter ion peak cluster was one of the most intense features in the MS2 spectra for all doubly charged precursors that contained one sodium and one proton adduct. The MS/MS data was subjected to SimGlycan software for structural identification in batch mode. Searches were performed using a 0.2 Da precursor ion tolerance while product ion tolerance was set to 0.1 Da. Additional filter conditions were set in order to limit the glycan search to glycoprotein N-glycans with intact core.

Figure 2: Prominent TMT reporter ion peaks feature in all the averaged MS/MS spectra for the precursors (A) m/z 780 with 8 MS/MS scans (B) m/z 861 with 8 MS/MS scans (C) m/z 942 with 7 MS/MS scans (D) m/z 1023 with 5 MS/MS scans and (E) m/z 1104 with 9 MS/MS scans.

Rapid Glycan Identification using Templates

For rapid identification of glycans, we have developed web based software modules to facilitate construction of custom databases wherein researcher can easily store glycan structures, and other related information such as glycan class, biological sources, pathways, reaction, enzymes and links to other public databases.

Figure 3: SimGlycan can restrict database search to custom database/s that contains curated data, thereby increasing the confidence of identification.

Users can add glycans into database by importing corresponding KEGG Chemical Function (KCF) format file. Adding multiple KCF files in batch is also allowed (shown in figure 3). 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 3 shows a typical SimGlycan web browser showing a list of searched glycans.

The program identifies N-glycan with carbohydrate residues (GlcNAc)2(Man)5 for the precursor m/z value 780, N-glycan (GlcNAc)2(Man)6 for m/z 861, N-glycan (GlcNAc)2(Man)7 for m/z 942, N-glycan (GlcNAc)2(Man)8 for m/z 1023 and N-glycan with (GlcNAc)2(Man)9 for m/z 1124, which is the correct assignment for all high-mannose glycans present in the sample. Figure 4 shows the typical search result pane displaying identified glycans in the SimGlycan Software.

The navigation window on the left displays the file name and MS/MS scans. Search results are displayed in two panes: The Search Results display pane and the Annotated Peaklist pane. Here the results for the MS/MS of the precursor m/z 780 are shown. SimGlycan ranks them on the basis of the peaks observed in the MS/MS data that correspond to diagnostic ions. For an identified glycan, the structure along with corresponding information such as name, sequence, composition, glycan mass, carbohydrate mass, class, links to reactions, pathways and other databases is provided. The Annotated Peaklist pane displays a table containing MS/MS spectrum data list (m/z, intensity) labeled with matched fragments, charge state and adduct. Structure of matched fragments can be viewed.

Figure 4: SimGlycan Software Interface: Structure of the (GlcNAc)2(Man)5 glycan with precursor m/z 780 is displayed. In silico fragments and corresponding structures can also be viewed.

Figure 5 shows the MS/MS spectrum for the precursor m/z 780 (same as the MS/MS spectrum in Figure 2(A)) annotated with fragments from the identified (GlcNAc)2(Man)5 glycan. It is observed that majority of the intense peaks correspond to the TMT reporter ions, as well as Z/Y/B/A-type ions. These ions are generally the outcome of two glycosidic cleavages towards reducing end with intact TMT tag and carry Na adduct. Besides this, all these intense peaks are singly charged ions. This observation is consistent for all the MS/MS spectra corresponding to (GlcNAc)2(Man)6, (GlcNAc)2(Man)7, (GlcNAc)2(Man)8 and (GlcNAc)2(Man)9 glycans (Figures 5-9).
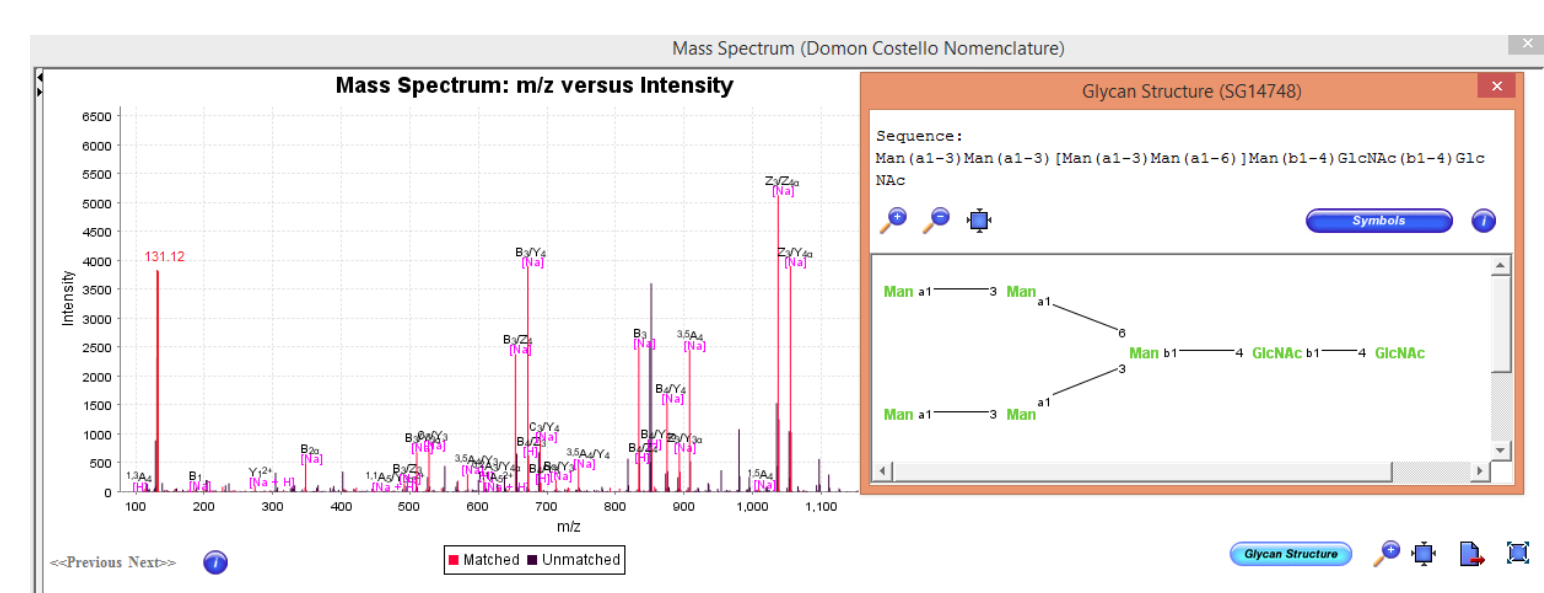

Figure 5: SimGlycan Software Interface: MS/MS spectrum for the precursor m/z 780 annotated with fragments of the (GlcNAc)2(Man)5 glycan.
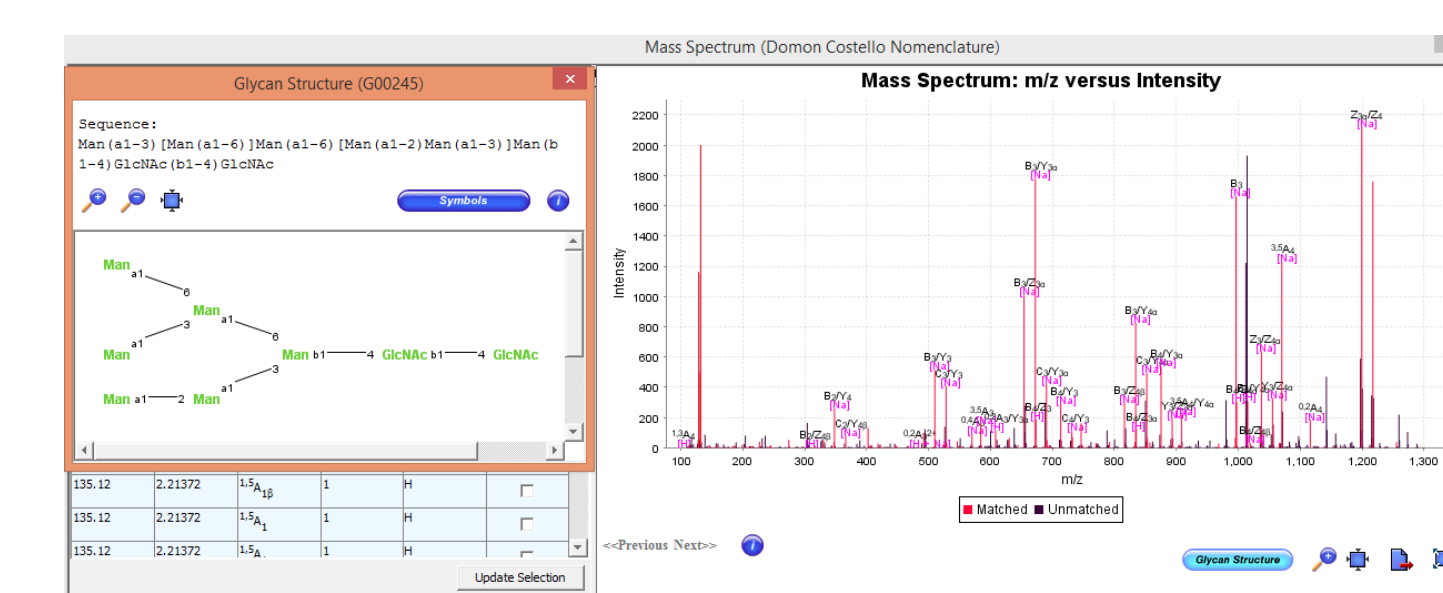

Figure 6: SimGlycan Software Interface: MS/MS spectrum for the precursor m/z 861 annotated with fragments of the (GlcNAc)2(Man)6 glycan.
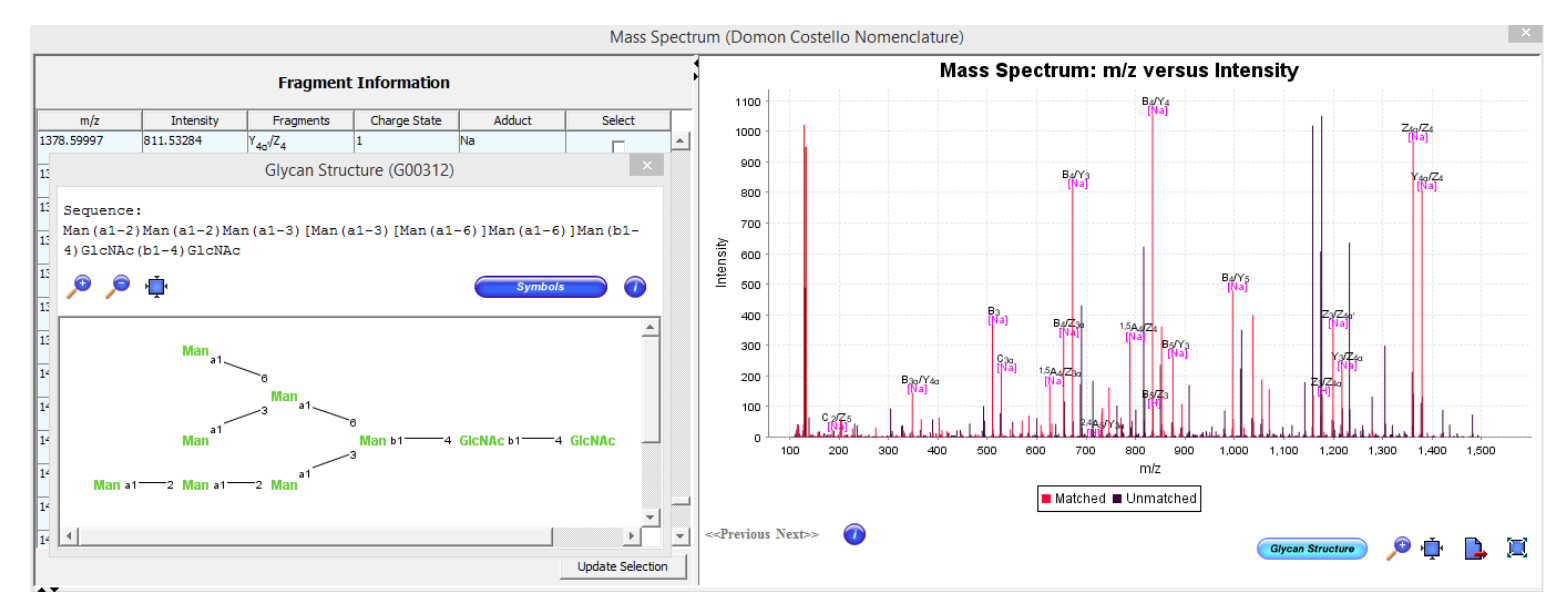

Figure 7: SimGlycan Software Interface: MS/MS spectrum for the precursor m/z 942 annotated with fragments of the (GlcNAc)2(Man)7 glycan.
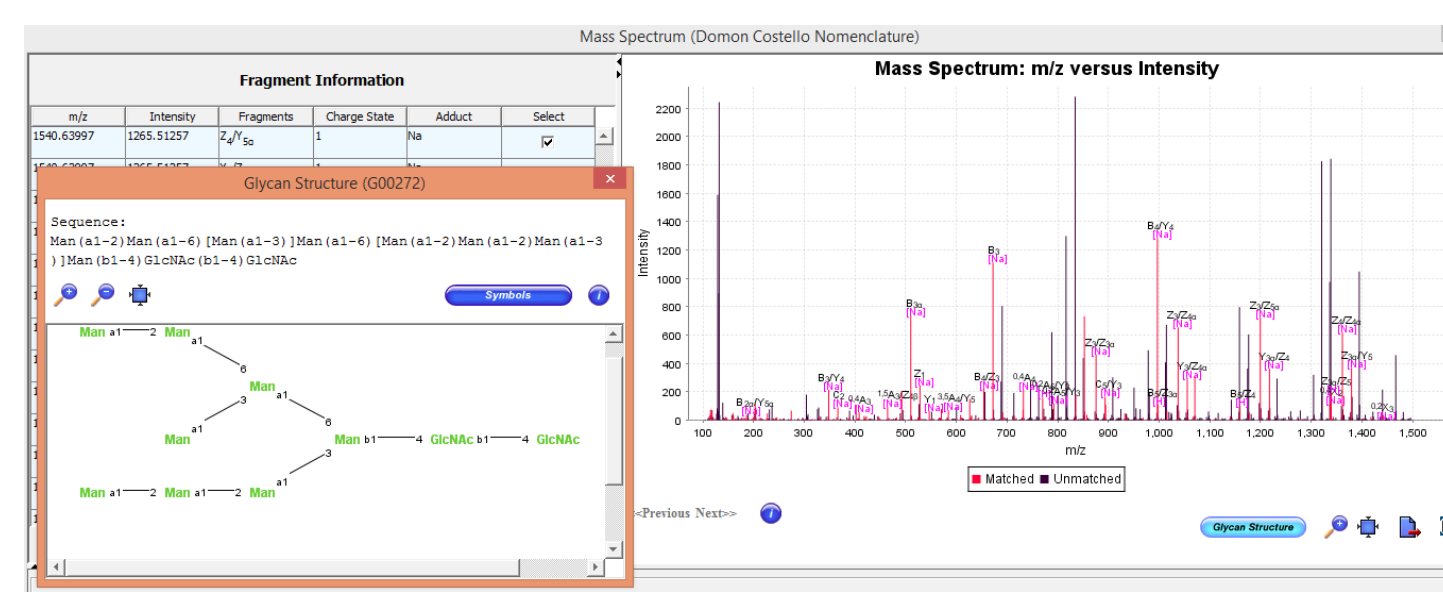

Figure 8: SimGlycan Software Interface: MS/MS spectrum for the precursor m/z 1023 annotated with fragments of the (GlcNAc)2(Man)8 glycan.

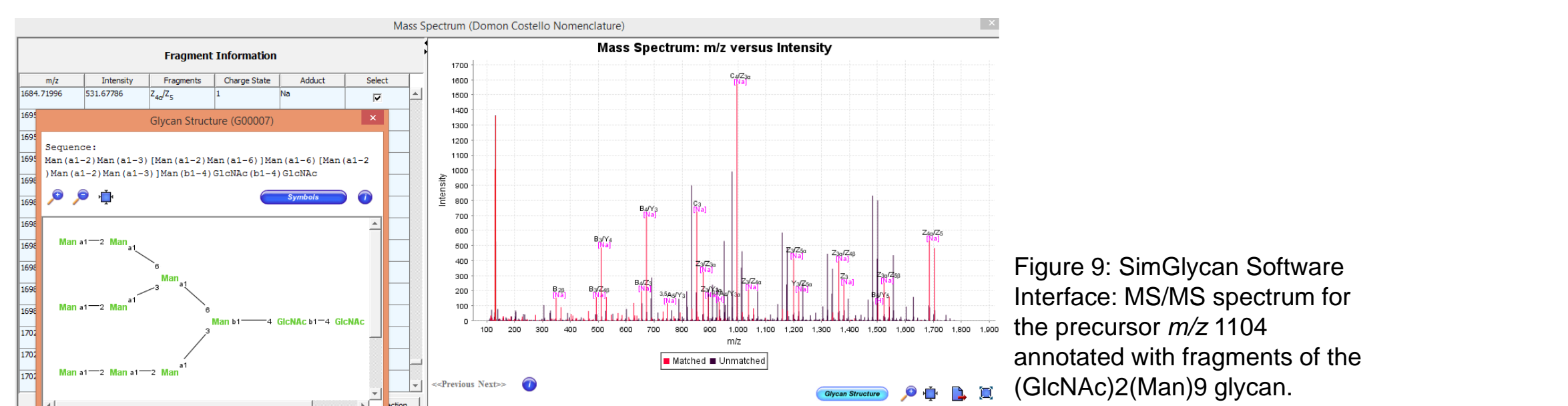Figure 9: SimGlycan Software Interface: MS/MS spectrum for the precursor m/z 1104 annotated with fragments of the (GlcNAc)2(Man)9 glycan.

Figure 10: SimGlycan Software Interface: Specify parameters to quantify glycans based on reporter ion intensity values.

Quantitative Data Analysis

The program enables the quantitation by measuring reporter ion peak intensities. Figure 10 shows the SimGlycan Software Interface wherein parameters for quantitative analysis are specified. We set the search space for each reporter ion to a tolerance of 0.2 Da to prevent overlap between the isobaric reporter ions. The program automatically extracts the intensity from all the MS/MS spectra for each TMT channel. Following extraction, the TMT reporter ion signals were adjusted to account for isotopic impurities in each TMT variant as specified in Figure 11.

Figure 11: SimGlycan Software Interface: Specify reporter ion isotopic distributions.

For glycan quantitation and comparison, the program identifies the glycan-spectral matches. Reporter ion intensities from different MS/MS spectra are then assembled into glycans. Glycans were then quantified by calculating the sum/average/median reporter ion intensities across all glycan-spectral matches. The program automatically ignores glycan-spectral matches with poor quality MS/MS spectra - more than 4 TMT channels missing - from quantitation.

As analysis summary, the program displays the following statistics as a measure of the relative amount of each identified glycan. The statistics constitute sum/average/median/standard deviation of the reporter ion intensities, ratio sum/average/median intensity of each TMT ion and the control TMT 130.13 ion, Log2 relative expressions of these ratios, number of MS/MS spectra which identifies each glycan (glycan spectral match/count).

Results of the quantitative data analysis are displayed in three different result panes. Glycan View, Scan View and Annotated Peaklist. In the Glycan View (Figure 12), all the glycans identified for precursor m/z values are listed along with the following information in each column:

m/z: Observed precursor m/z.

Adduct: Ion species of the precursor m/z.

Composition: Number of different carbohydrate residues present in the glycan.

Mass: Monoisotopic mass of the identified glycan in Da.

GSC: Glycan spectral count i.e., number of MS/MS spectra in which the glycan is identified.

Avg. Score/Avg. Rank: Average of the scores/ranks the glycan received across spectra.

<Total/Average/Median> Ion Intensity: Each sub-column displays the Total/Average/Median intensity observed for the corresponding reporter ion across MS/MS spectra. In case of multiple replicates, it is the average of the Total/Average/Median intensity observed in each replicate. In case of more than one experiment, it is the average across experiments.

Normalized <Total/Average/Median> Ion Intensity: The Total/Average/Median intensity for each reporter ion is calculated based on the average of Total/Average/Median intensity values across observed glycans.

Log2 (TMT Ion/ <Control Reporter Ion>): It is calculated as follows:
Log2 [(Total/Average/Median Intensity of the reporter Ion)/Intensity of the Control Ion]

Additional information such as class, link to other databases, reaction, pathways and enzymes, can also be viewed at the lower left pane of the glycan view. The 'Filter Results' pane provides filters to refine glycans for quantitative analysis while the columns to be displayed can be selected using the lower right pane.

Figure 12: SimGlycan Software Interface: 'Glycan View' displaying identified glycans and their corresponding information.

Bar chart (Figure 13) displays the total ion intensity for each glycan at different TMT channels. In the TMT 131, (GlcNAc)2(Man)5 is the most abundant followed by (GlcNAc)2(Man)6, (GlcNAc)2(Man)7, (GlcNAc)2(Man)8 and (GlcNAc)2(Man)9 in decreasing order. Similarly, observations for other TMT ions are also displayed in the bar chart.

Using the cluster dot plot feature in SimGlycan Software (Figure 14), we perform comparative and differential analysis of the identified glycans in different TMT channels. This graph shows the ratio of the normalized intensities of TMT reporter ions and the control reporter ion in Log2 scale. The intensities are normalized to the average of the intensities so as to nullify the effect of under sampling or oversampling of glycans in any of the TMT channels. This graph displays the change in glycan expression level across different TMT reporter ion channels with respect to the control TMT 130. The absence of glycans in TMT 129 is evident from the graph. The quantity of (GlcNAc)2(Man)6, (GlcNAc)2(Man)7 and (GlcNAc)2(Man)8 is higher in TMT126, TMT127, TMT128 and TMT131 than that in the TMT130 while the quantity of (GlcNAc)2(Man)5 and (GlcNAc)2(Man)9 glycans is lower than that in the control.

The performance of the software for several standard glycoproteins and monoclonal antibodies was also examined in these studies. (Data not shown).

Conclusion

We have investigated the effectiveness of the recently introduced aminoxyTMT Reagents, which enable efficient relative quantitation of carbohydrates, improved labeled-glycan ionization efficiency and increased analytical throughput, for quantitative analysis of native N-glycans by LC coupled ESI-MS. Labeled glycans were analyzed using Velos Pro mass spectrometer (Thermo Scientific™) in the positive ion mode to acquire trap-HCD MS/MS fragmentation.

Large data was generated in the process of analyzing labeled-glycoforms derived from complex mixtures in mass spectrometry workflow. SimGlycan is redesigned to address problems faced by glycomics researchers due to the lack of informatics tools to automate the mass spectrometry based qualitative and quantitative analysis protocols. The fragmentation patterns of the identified glycans in MS/MS spectra, ion species of the product ions etc. are interpreted using SimGlycan. Glycans present in each TMT reporter ion channel are then quantified. Various features such as bar chart, cluster dot plot, heat map etc. facilitate comparative and differential analysis of glycans present in different TMT reporter ion channels.

References

1. Gong, B., Hoyt, et al. (2013). N-Glycosylamine-mediated isotope labeling for mass spectrometry-based quantitative analysis of N-linked glycans. Analytical and bioanalytical chemistry, 405(17), 5825-5831.

2. Yang, Shuang, et al. (2013). Glycan Analysis by Isobaric Aldehyde Reactive Tags and Mass Spectrometry. Analytical chemistry 85(17) : 8188-8195.

3. Viner, R. I., et al (2010). A Novel Workflow for Glycopeptide Analysis Using Cellulose-Based Separation Cartridges, TMT-Labeling and LTQ Orbitrap ETD. Journal of biomolecular techniques: JBT, 21(3 Suppl), S25.

4. Atwood Iii, J. A., et al (2007). Quantitation by isobaric labeling: applications to glycomics. Journal of proteome research, 7(01), 367-374.

5. Apte, A., & Meitei, N. S. (2010). Bioinformatics in Glycomics: Glycan Characterization with Mass Spectrometric Data Using SimGlycan. In Functional Glycomics (pp. 269-281). Humana Press.

6. Meitei, N. S., Apte, A., Snovida, S. I., Rogers, J. C., & Saba, J. (2015). Automating Mass Spectrometry-Based Quantitative Glycomics using Aminoxy Tandem Mass Tag Reagents with SimGlycan. Journal of proteomics.
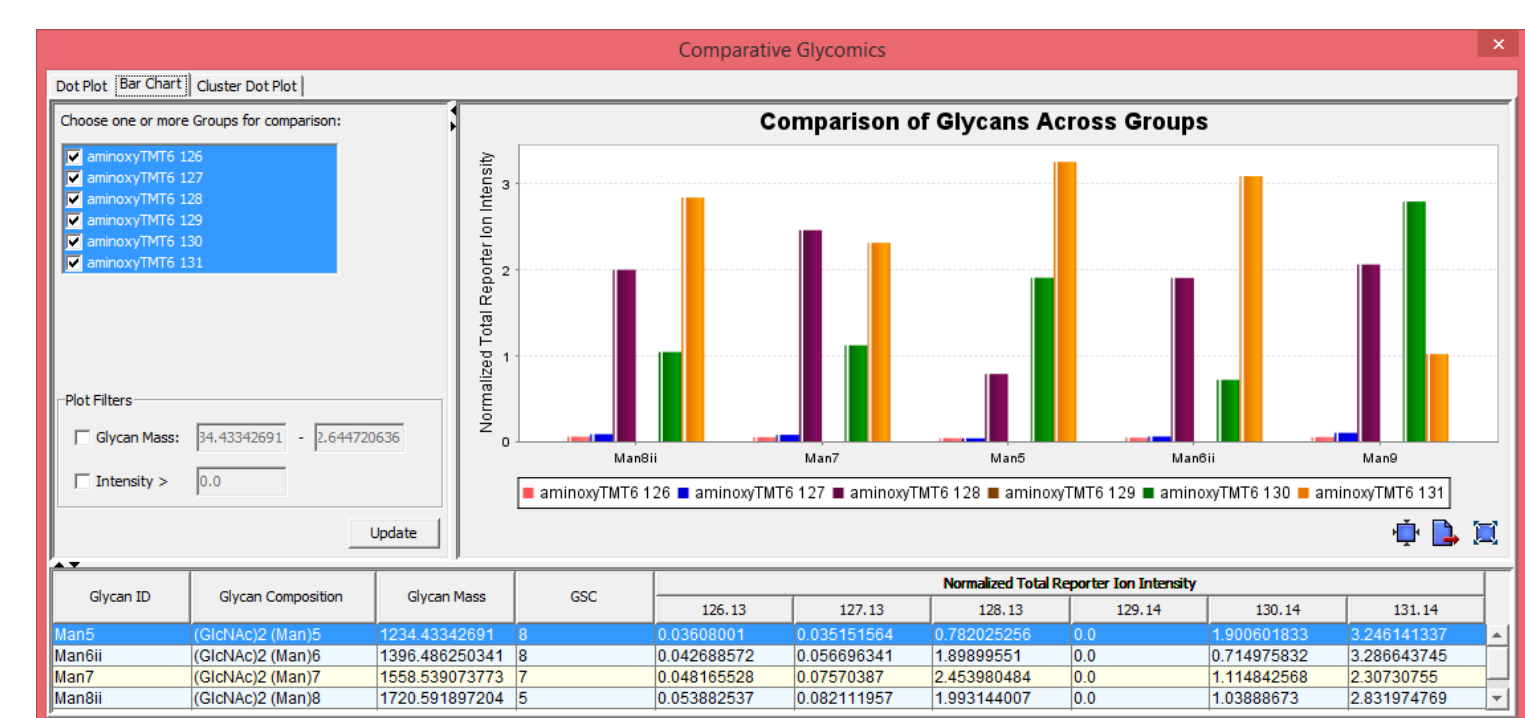
Figure 13: SimGlycan Software Interface: Bar Chart displaying the total ion intensity for glycans in different TMT reporter ions.
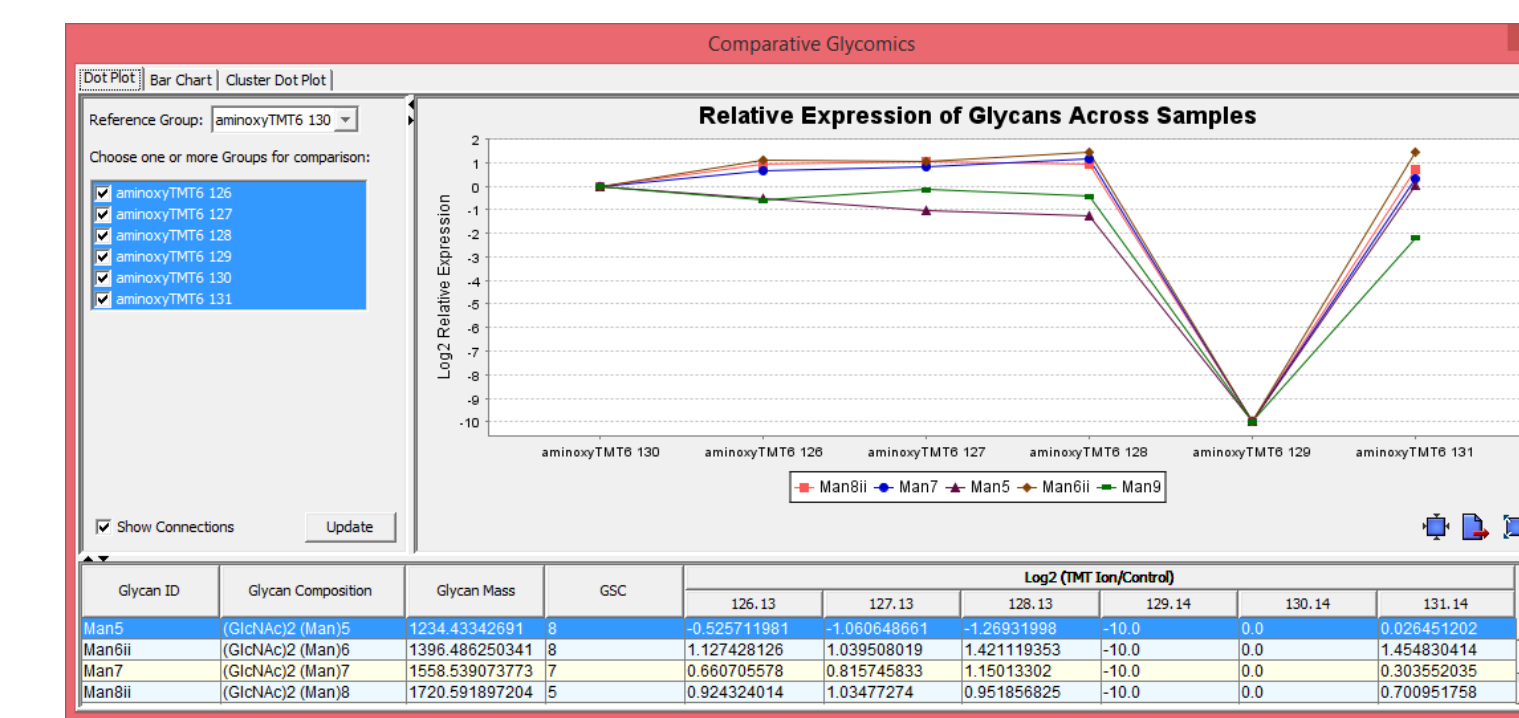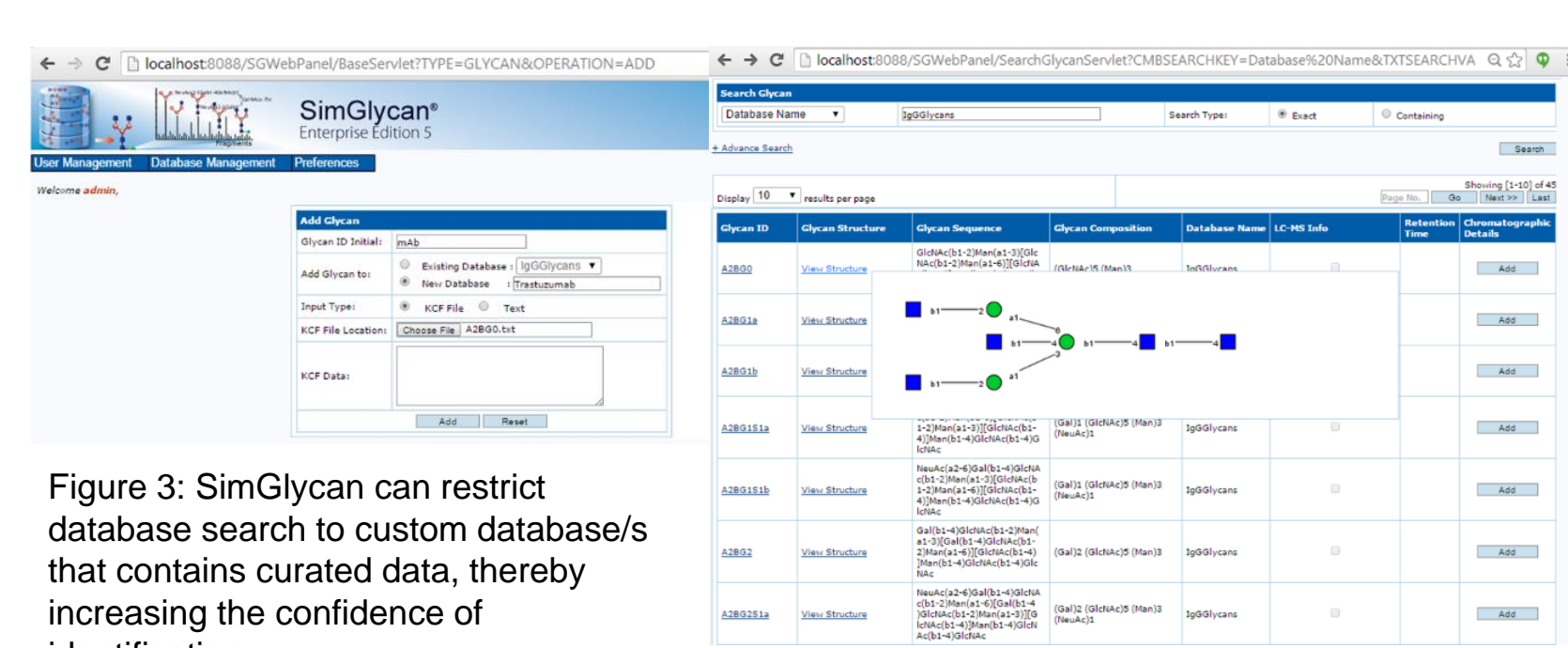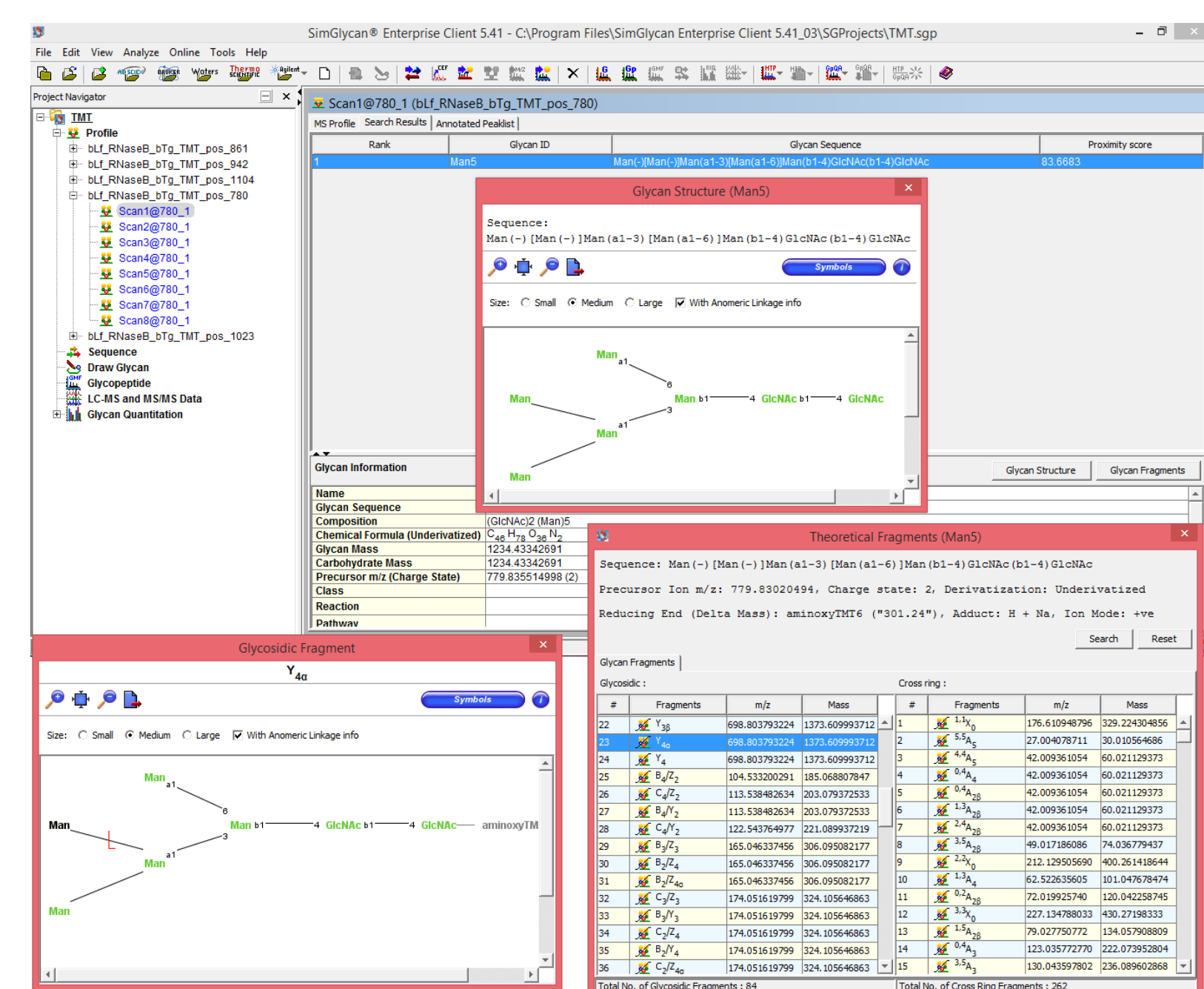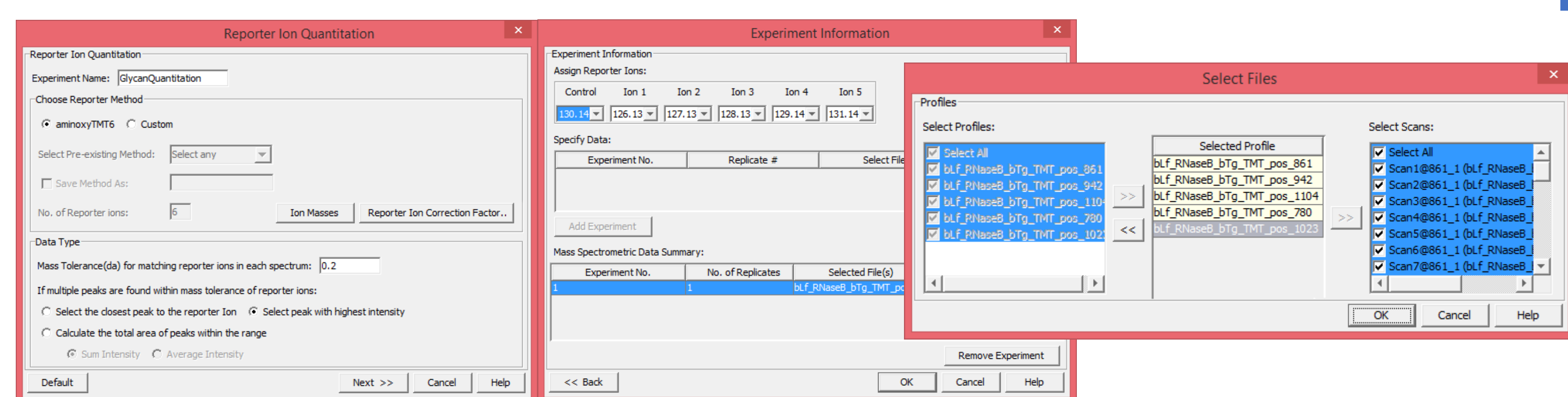

Figure 14: SimGlycan Software Interface: Cluster dot plot depicting fold changes across TMT channels.