Structural Identification, and Quantitation of Glycans Labeled with Dual Modifications Strategies from Complex Mixtures using LC-MS Workflows with SimGlycan® Software

Ningombam Sanjib Meitei[1, 2], Haiying Li[3], Sohini Saha[1], Rupanjan Goswami[1], Arun Apte[2], Richard S. Lee[3]
[1]PREMIER Biosoft, Indore, India; [2]PREMIER Biosoft, CA; [3]Department of Urology, Boston Children's Hospital and Harvard Medical School, Boston, MA
Corresponding author e-mail: sanjib@premierbiosoft.com

INTRODUCTION

Glycans play a critical role in the therapeutic efficacy and toxicity of monoclonal antibody proteins. Recently developed integrated chemical strategy, Dual Reactions for Analytical Glycomics (DRAG), enables quantitative comparison between neutral and sialylated glycans simultaneously using mass spectrometry methods [1]. This strategy has been proved to overcome a major technical bottleneck in MS-based glycomics – differences in ionization efficiency among neutral and sialylated glycans prevent direct quantitative comparison by their respective mass spectrometric signals – facilitating qualitative and quantitative studies of neutral, and sialylated glycans from complex biological mixtures using advanced LC-MS workflows.

Currently, no informatics solution exists that can analyze the MS/MS data generated from such methods. We have redeveloped SimGlycan software [2] to streamline the data analysis process.

METHODS

SAMPLE PREPARATION: N-glycans were released from a mixture of five glycoprotein standards by PNGase F. The released N-glycan underwent dual-reaction based on a previously published method. Briefly, two equivalent N-glycan aliquots were respectively labeled with a pair of "light" and "heavy" 2-AA reagents, which were then mixed together and purified by cellulose beads. The purified 2-AA labeled glycans were further methylamidated followed by another cellulose purification (Figure 1).

LC-MS METHODS: The purified N-glycans derived from 1.5µg protein mixture were injected into an Ekspert nanoLC 415 coupled to a Thermo Q-Exactive mass spectrometer.

DATA ANALYSIS: SimGlycan software [2] was used for (i) LC-MS data processing, (ii) identification of glycans treated with stable isotope labeling and methylamidation; and methylamidation of fixed/variable number of acidic carbohydrate residues in a glycan.

DATABASE CURATION

A custom database containing 65 curated compositional structures from five standard glycoproteins mixtures was created using SimGlycan server (Table 1). This custom database was used as the MS/MS database search template to facilitate rapid and accurate identification of glycans with 10 ppm tolerance for both the precursor and product ions. The validation of the results generated by newly developed software modules was done by comparing the manually annotated vs SimGlycan annotated compositional structures of the LC-MS data.

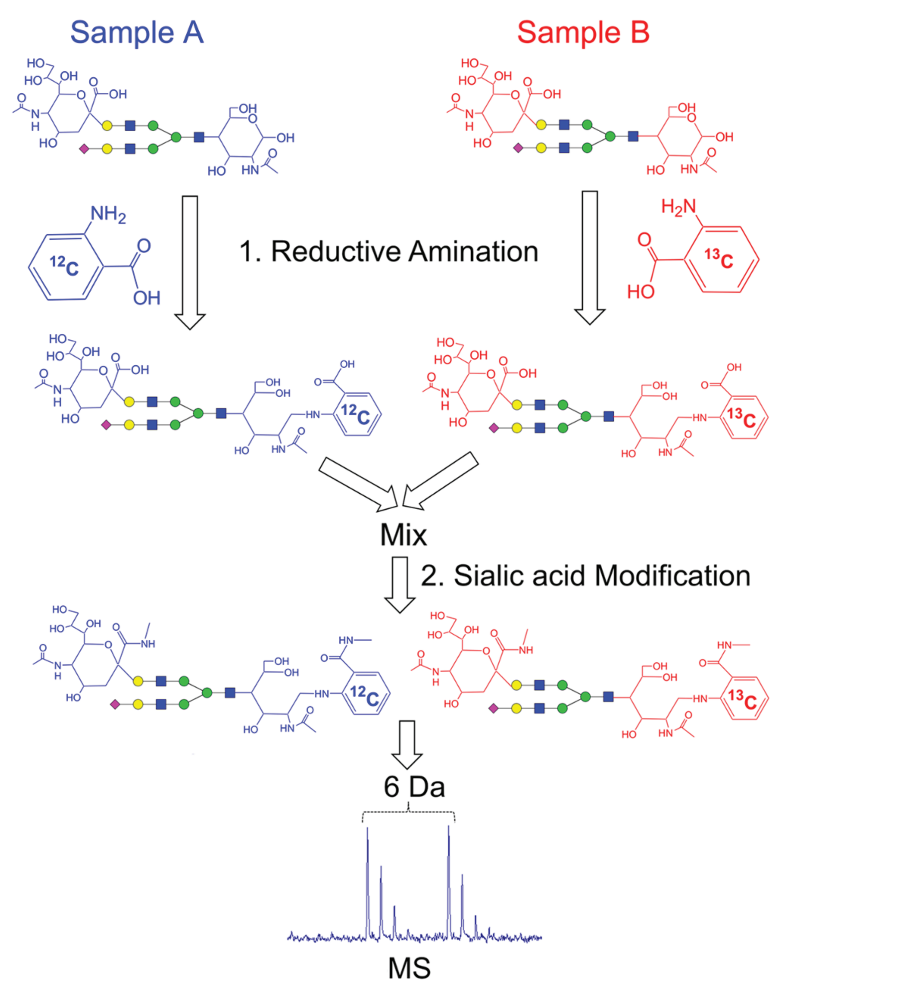

Figure 1: Workflow of the Dual Reactions for Analytical Glycome (Zhou H., et al (2014))

1	HexNAc2	18	Hex5HexNAc4	35	Hex3HexNAc5Fuc1	52	Hex6HexNAc5NeuAc2
2	Hex1HexNAc2	19	Hex5HexNAc4Fuc1	36	Hex3HexNAc5Fuc2NeuAc1	53	Hex6HexNAc5NeuAc3
3	Hex2HexNAc2	20	Hex3HexNAc4Fuc1	37	Hex4HexNAc5	54	Hex6HexNAc5NeuAc4
4	Hex3HexNAc2	21	Hex4HexNAc4	38	Hex4HexNAc5Fuc1	55	Hex4HexNAc6Fuc2NeuAc1
5	Hex4HexNAc2	22	Hex4HexNAc4Fuc1	39	Hex4HexNAc5Fuc1NeuAc1	56	Hex4HexNAc6Fuc2NeuAc2
6	Hex5HexNAc2	23	Hex4HexNAc4Fuc1NeuAc1	40	Hex5HexNAc5	57	Hex4HexNAc6Fuc2NeuAc3
7	Hex5HexNAc4NeuGc1	24	Hex5HexNAc4Fuc1NeuAc1	41	Hex5HexNAc5Fuc1	58	Hex4HexNAc6Fuc3NeuAc2
8	Hex6HexNAc2	25	Hex5HexNAc4Fuc1NeuAc2	42	Hex5HexNAc5Fuc1NeuAc1	59	Hex7HexNAc6Fuc1NeuAc4
9	Hex7HexNAc2	26	Hex5HexNAc4Fuc2NeuAc1	43	Hex5HexNAc5Fuc1NeuAc2	60	Hex7HexNAc6NeuAc1
10	Hex8HexNAc2	27	Hex5HexNAc4NeuAc1	44	Hex5HexNAc5Fuc1NeuAc3	61	Hex7HexNAc6NeuAc2
11	Hex9HexNAc2	28	Hex5HexNAc4NeuAc1NeuGc1	45	Hex5HexNAc5NeuAc1	62	Hex7HexNAc6NeuAc3
12	Hex6HexNAc3	29	Hex5HexNAc4NeuAc2	46	Hex5HexNAc5NeuAc2	63	Hex7HexNAc6NeuAc4
13	Hex5HexNAc3	30	Hex5HexNAc4NeuAc3	47	Hex6HexNAc5Fuc1NeuAc2	64	Hex7HexNAc6Fuc1NeuAc2
14	Hex4HexNAc3	31	Hex5HexNAc4NeuGc2	48	Hex6HexNAc5Fuc1NeuAc3	65	Hex8HexNAc7NeuAc4
15	Hex3HexNAc3Fuc1	32	Hex6HexNAc4Fuc1NeuAc3	49	Hex6HexNAc5Fuc2NeuAc2	66	Hex5HexNAc7Fuc2NeuAc3
16	Hex6HexNAc3Fuc1	33	Hex7HexNAc4Fuc2NeuAc2	50	Hex6HexNAc5Fuc2NeuAc3	67	Hex6HexNAc8Fuc2NeuAc1
17	Hex6HexNAc3Fuc2NeuAc1	34	Hex3HexNAc5	51	Hex6HexNAc5NeuAc1	68	Hex4HexNAc3Fuc1

Table 1: List of the 68 curated compositional glycan structures derived from five standard glycoproteins mixtures

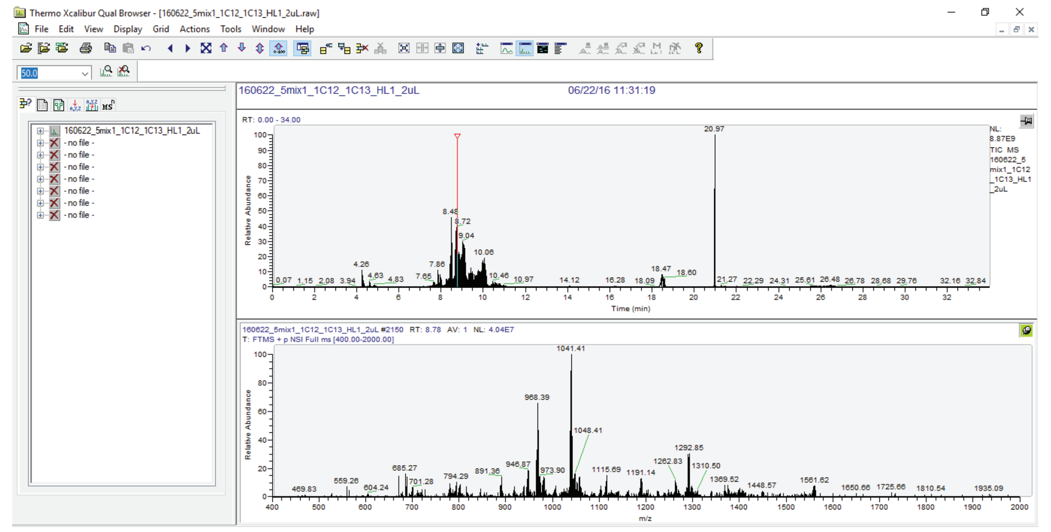

Figure 2: Typical Thermo Xcalibur Software graphical user interface showing total ion chromatogram, and mass spectrum

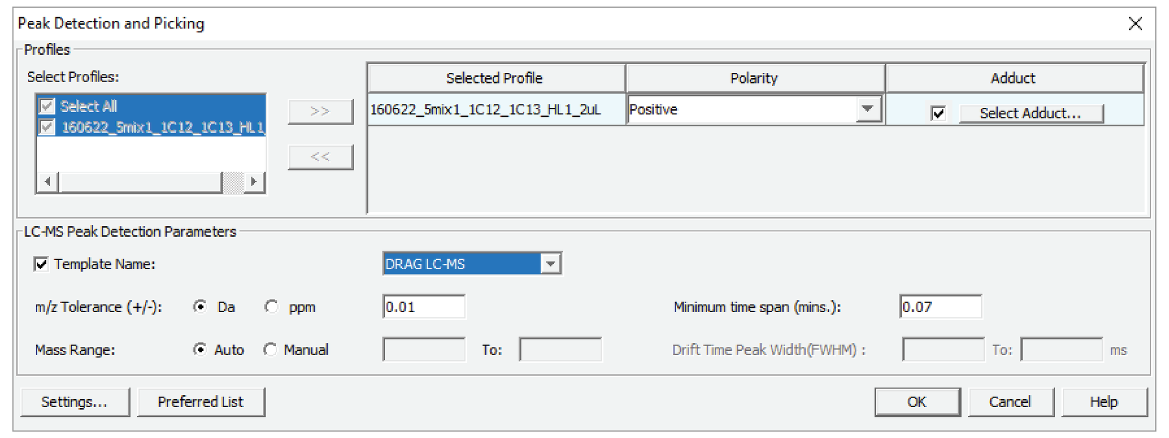

Figure 3(A): Typical GUIs of SimGlycan software showing Peak Detection and Peak Picking parameters using the DRAG LC-MS template

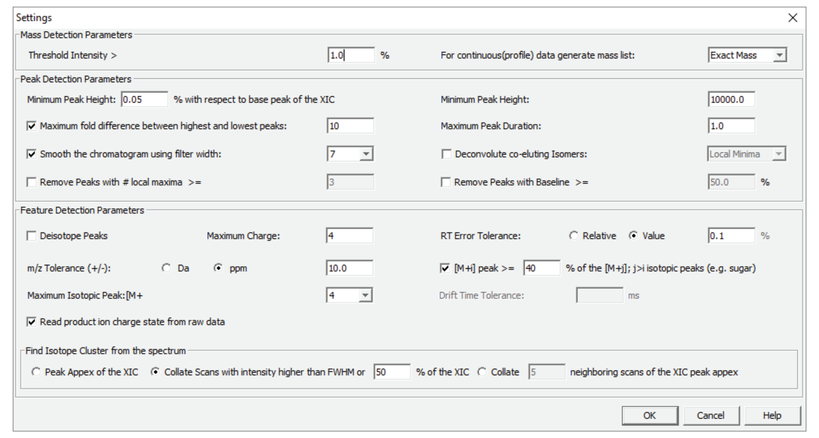

Figure 3(B): Typical GUI of SimGlycan software showing peak detection, and molecular feature finding parameters for the DRAG LC-MS template

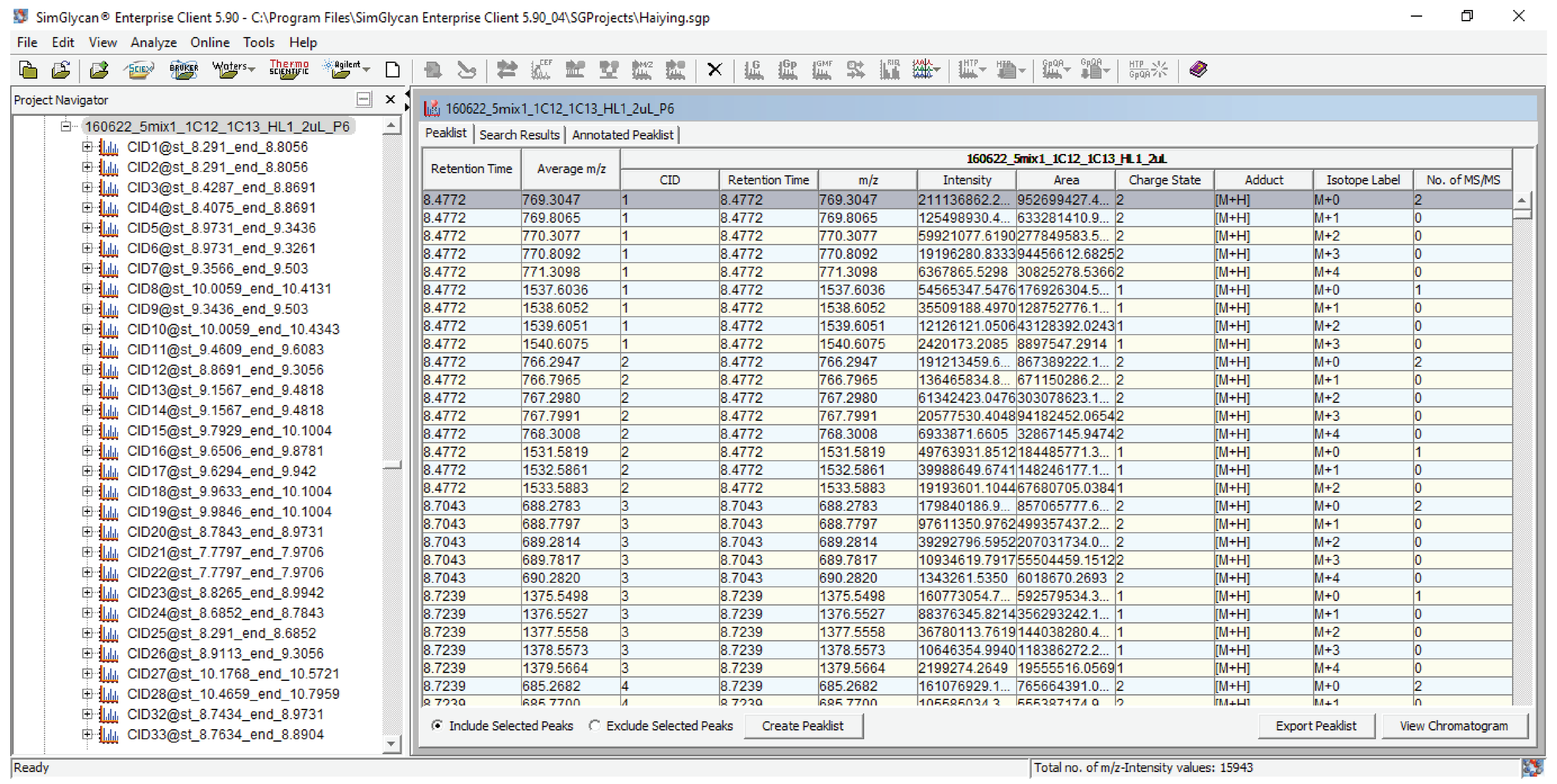

Figure 4: Typical SimGlycan software GUIs showing the LC-compounds

RESULTS AND DISCUSSION

The raw data from the LC-MS experiment (Figure 2) was subjected to SimGlycan software for peak detection, molecular feature finding, the association of MS/MS data to LC-compounds, and assigning correct precursor m/z values – the m/z at monoisotopic peak of the identified isotopic envelopes – for the MS/MS scans. A specific LC-MS peak picking template named "DRAG LC-MS" was created (Figure 3). The molecular feature finding algorithm reported 5058 LC-compounds out of which 1982 has MS/MS data (Figure 4).

The MS/MS scans with only H adduct were subjected to SimGlycan database search for glycan annotation using the parameter settings displayed in the Figure 5. The search was performed on the custom database containing 68 curated compositional structures using 10 ppm tolerance for the precursor and product ions.

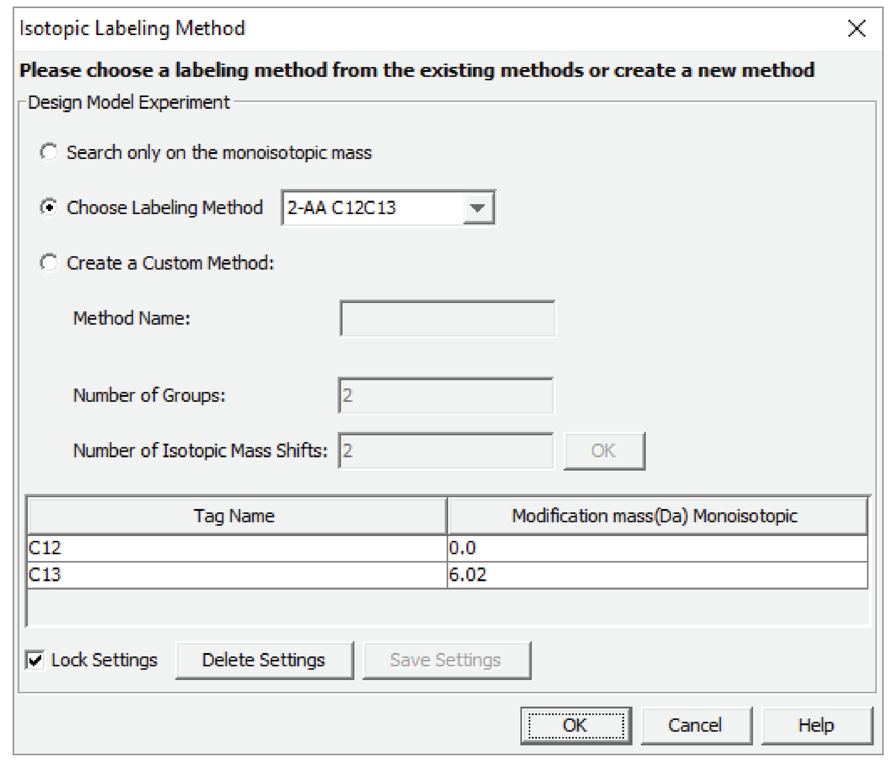

Figure 5(A): Typical SimGlycan software search window showing (A) the database search parameters for the DRAG LC-MS workflows

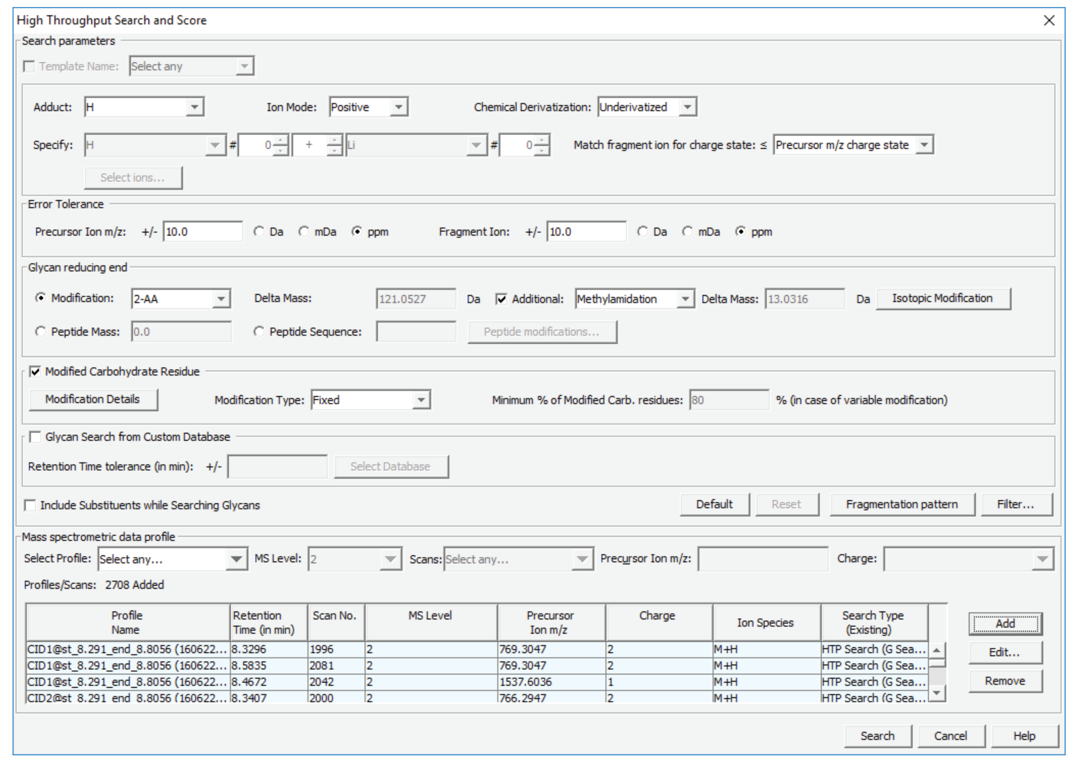

Figure 5(B): Specify stable isotope labeling method

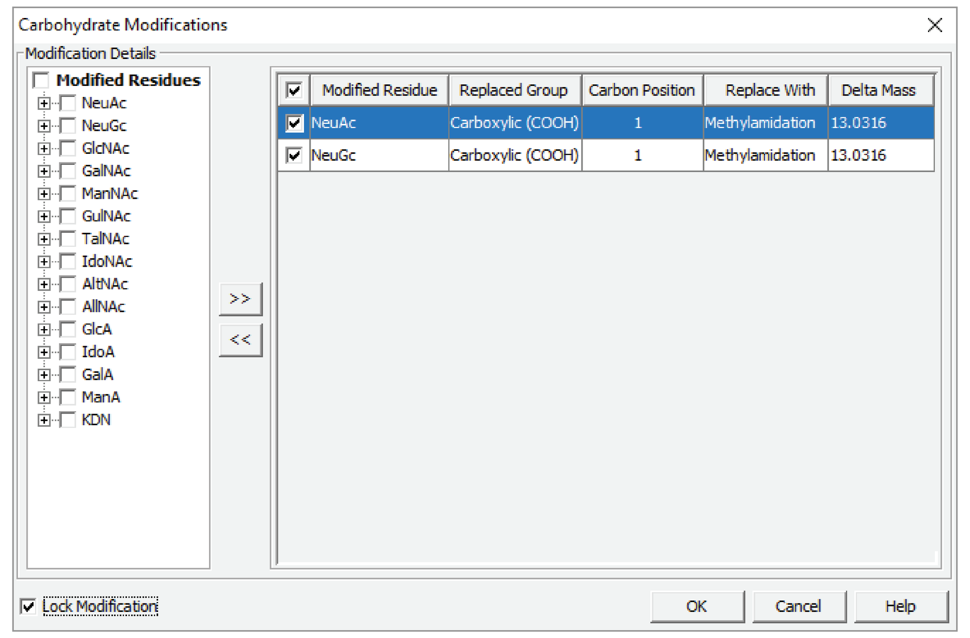

Figure 5(C): Carbohydrate modifications e.g., methylamidation of sialic acid residues

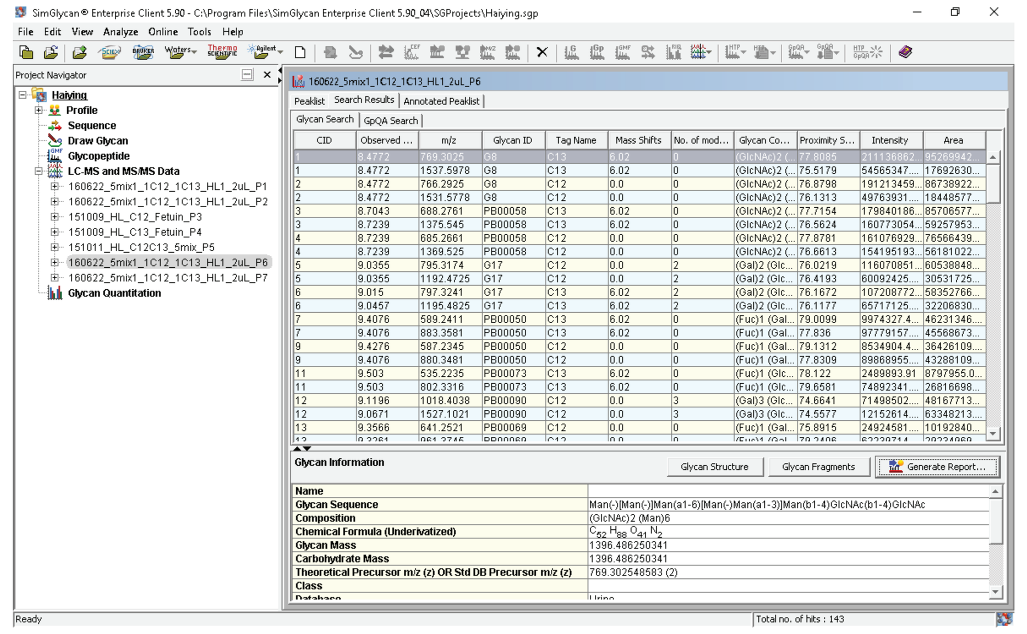

Figure 6: Typical GUIs of SimGlycan software showing the list of identified glycan structures along with their isotope labels and mass shift information

The molecular feature finding algorithm of SimGlycan software correctly reported 234 LC-compounds out of the 241 manually detected LC-compounds. The MS/MS database search correctly annotated 227 LC-compounds with correct compositional structures i.e., 94% correct assignment.

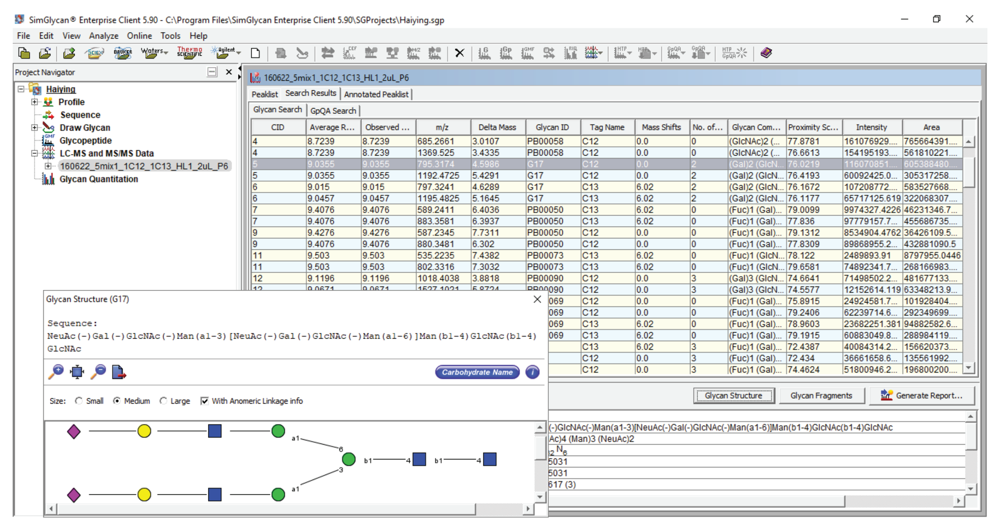

Figure 7(A): Typical SimGlycan interface showing identification of 2-AA[12C] labeled (Gal)2 (GlcNAc)4 (Man)3 (NeuAc)2 as the compositional structure at retention time 9.035 minute*

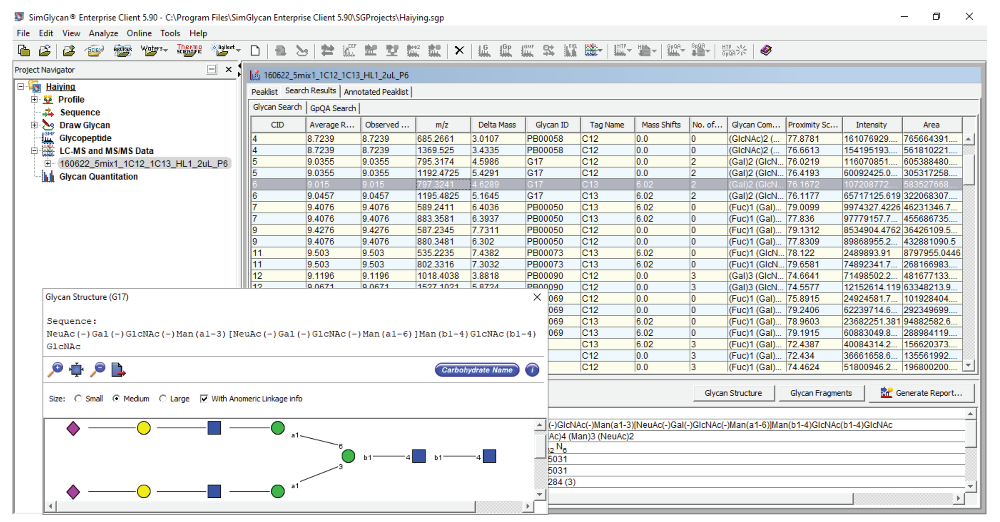

Figure 7(B): Typical SimGlycan interface showing identification of 2-AA[13C] labeled (Gal)2 (GlcNAc)4 (Man)3 (NeuAc)2 glycan at retention time 9.015 minute*

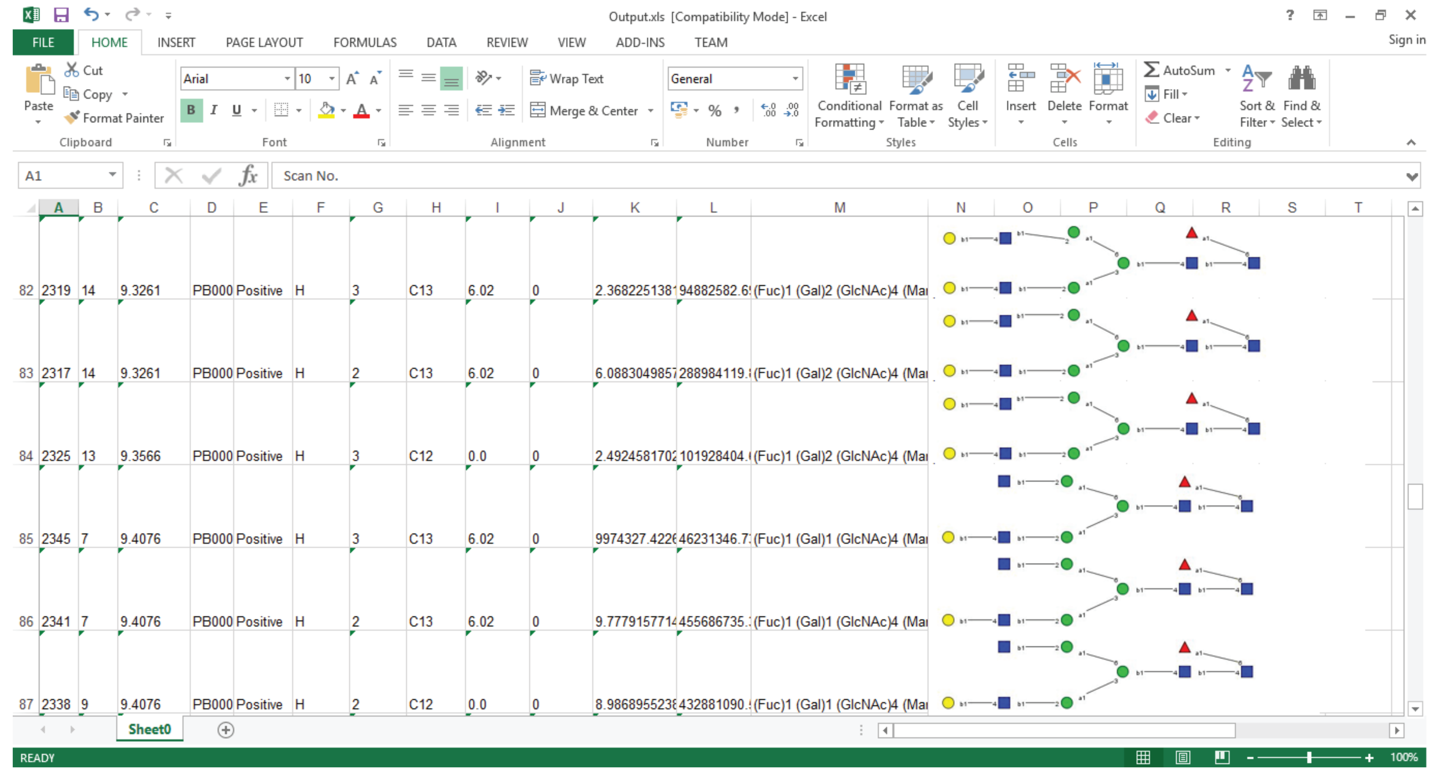

Figure 8: Typical SimGlycan software generated MS Excel format report file showing detailed information of the identified glycans*

* The topology of the reported glycan structures are incorrect and displayed to demonstrate the software functionality. Only the glycan compositions are meaningful.

The identified glycans were then mapped between the biological groups – Group A: labeled with the light 2-AA reagent, and Group B: labeled with the heavy 2-AA reagent. The area under the curve or peak height of extracted ion chromatograms was used as their measure of quantity in the samples.

CONCLUSION

New software modules of SimGlycan v.5.90 provides the following key features:

1. High resolution accurate mass (HRAM) LC-MS data processing – accurate peak picking, and charge state identification
2. Identification of glycans using MS based DRAG workflow

REFERENCES

1. Zhou, Hui, et al. Analytical Chemistry 86.13 (2014): 6277-6284.
2. Meitei, NS, et al. Journal of Proteomics 127 (2015): 211-222.