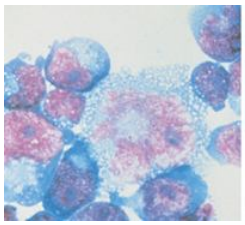
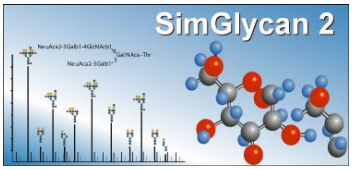


TECHNICAL NOTE

SimGlycan™ Software*: A New Predictive Carbohydrate Analysis Tool for MS/MS Data

Automated Data Interpretation for Glycan Characterization



Jenny Albanese1, Matthias Glueckmann2 and Christof Lenz2
1Applied Biosystems, Foster City, CA, USA and 2Applied Biosystems, Darmstadt, Germany

Challenge of Carbohydrate Analysis by Mass Spectrometry

The analysis of carbohydrates is of great importance in modern biochemistry and is often tackled using various analytical techniques, including mass spectrometry. However, the structural analysis of carbohydrates by mass spectrometry is not routine and often requires tedious, time consuming manual spectral interpretation. Additionally, the analysis requires the knowledge of various carbohydrate sequences, anomeric linkages between the monosaccharides and their fragmentation patterns under different experimental conditions such as permethylation, reducing end modifications, adducts used, and the ionization mode.

Here we present novel software for the characterization of carbohydrates. SimGlycan Software predicts the structure of a glycan from the MS/MS data acquired by mass spectrometry, facilitating the study of glycosylation and post translational modifications. It supports MS/MS profiles of both glycopeptides and released glycans and can deduce carbohydrate and glycan composition and structure directly from low and high energy CID MS/MS data. SimGlycan accepts experimental MS/MS data generated by all AB Sciex mass spectrometers, including the 4000 QTRAP® System, QSTAR® Elite System, and the 4800 *Plus* MALDI TOF/TOF™ Analyzer. SimGlycan matches these experimental MS/MS data with its own database containing theoretical fragmentation of over 8,000 glycans and generates a list of probable glycan structures. Each structure is scored to reflect how closely it matches the experimental data (Figure 1). SimGlycan provides for every probable glycan result, the glycan fragments, structure, sequence, composition, glycan mass, class, reaction, pathway, enzyme and other database links, allowing easy access to all published glycan information without referring to multiple sources.

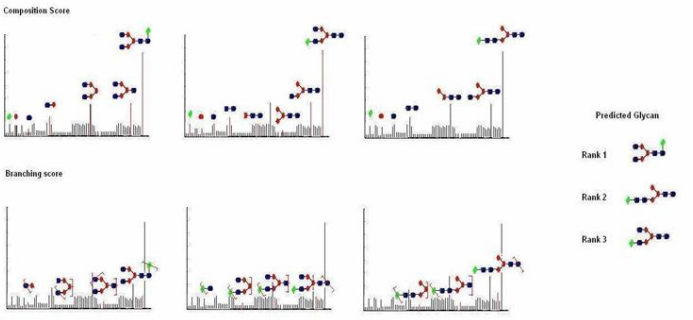


Figure 1. Matching, Scoring and Glycan Ranking. SimGlycan verifies the monosaccharide composition of the unknown glycan using diagnostic ions calculated for various glycans in the database. A composition score is calculated for every candidate entry. Further, a branching score is calculated to identify the linkages and branching patterns of the unknown glycan using the internal and cross ring cleavages of theoretical glycans. The glycan with the highest composition and branching scores should reflect the actual structure.

Key Features of SimGlycan Software

- Imports experimental MS/MS data generated by all AB Sciex mass spectrometers and can be launched directly from Analyst® QS 2.0 and Analyst 1.5 Software using the companion software feature.
- Matches glycan MS/MS data with its own database of theoretical fragmentation of over 8,000 glycans by using a proprietary ranking and scoring algorithm.
- Database can be searched using ID, sequence, composition or mass.
- References other biological information for the probable glycan structures such as glycan class, reaction, pathway and enzyme.
- Structural analysis of glycopeptides can be performed by specifying the sequence or mass of the attached peptide moiety.
- Enables searching for glycans with chemical derivatives used for reducing end modifications even if the derivative is not available in the program's database.

**SimGlycan™ Software is a product of PREMIER Biosoft International, Palo Alto, CA*

Materials and Methods

***MALDI Sample Preparation*:** Carbohydrate samples in different concentrations were prepared using DHB matrix (20 mg/mL in deionized water). Dried droplets were prepared with carbohydrate solutions at concentrations of 10^{-4} to 10^{-5} mol/L.

Mass Spectrometry: Samples were analyzed using the 4800 *Plus* MALDI TOF/TOF™ Analyzer in reflector mode. Positive ion MS/MS data of sodiated ions were collected using 2kV collision energy and air or Argon as collision gas. Operating pressure of ion source was 10^{-7} Torr.

For the 4000 QTRAP® System, MS/MS and MS^3 experiments were performed using a NanoSpray® Source (800 V) and medium nanospray needles (Protana). Samples were sprayed from water:methanol (1:1, v:v). Positive ion spectra were obtained with added sodium acetate. MS/MS spectra were acquired from *N*-linked glycans at 1000 amu/sec. The MS/MS and MS^3 collision and excitation energies were set to be appropriate for the mass of the compounds. The Q1 resolution was low to allow transmission of the isotopic envelope from each ion and the declustering potential (30 V) was low to prevent in-source decay.

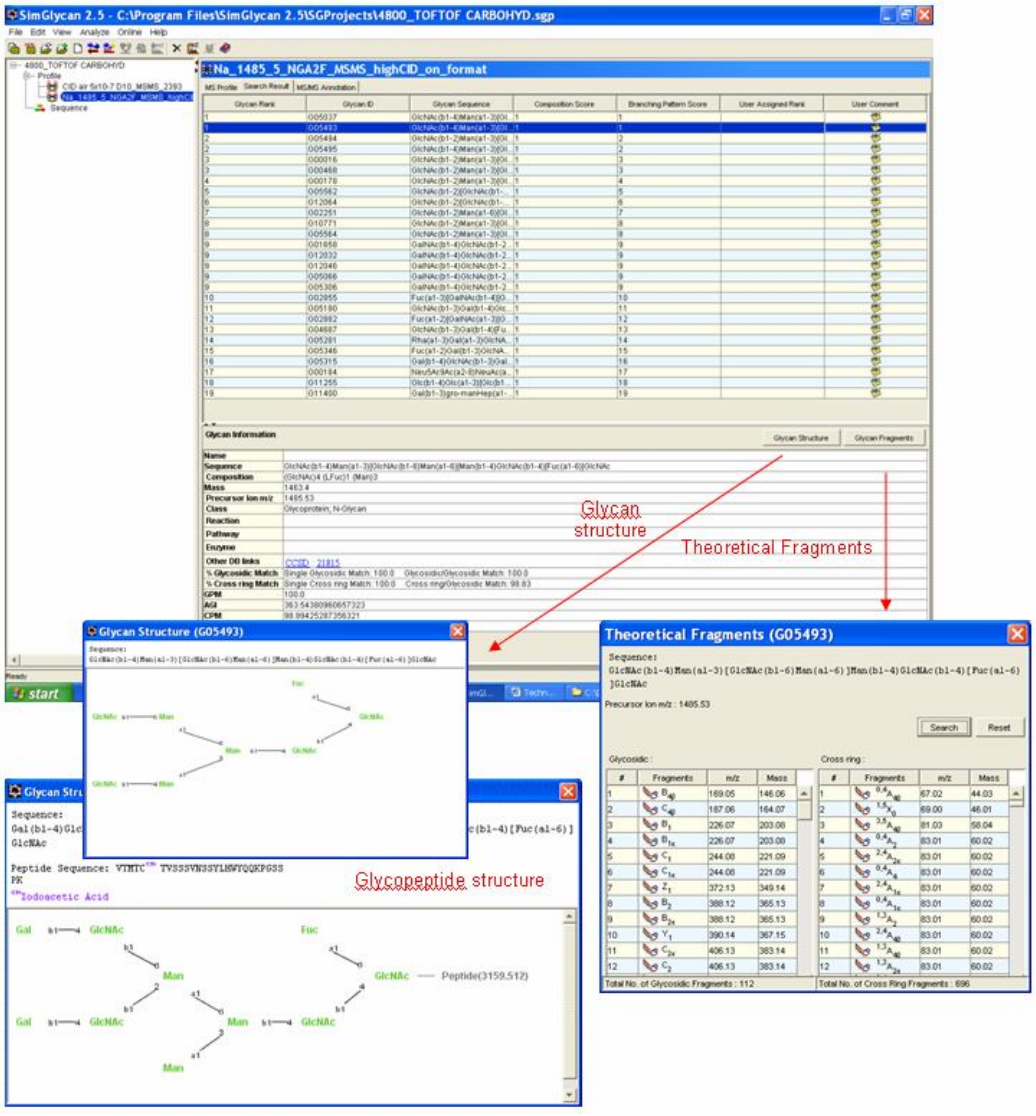


Figure 2. SimGlycan Software User Interface. The search results window of SimGlycan software provides an intuitive user interface for data review. In the upper pane, search results from a high energy CID MS/MS spectrum are displayed. In the lower pane, the glycan information is listed. The Glycan Structure button provides the sequence and linkage information. An example of a glycopeptide search is shown below that. The Glycan Fragments button provides the sequence, m/z, and theoretical glycosidic and cross ring fragment ions. These can be further searched by name, mass or m/z value.

SimGlycan Settings: SimGlycan accepts experimental *m/z* and intensity values (MS/MS data) of glycans. The following settings were used for the data: Precursor ion tolerance of 0.2 Da, precursor ion *m/z* as indicated on the figures, spectrum peak *m/z* tolerance of 0.2 Da, no glycan derivatization, ion mode was MALDI positive for the 4800 *Plus* MALDI TOF/TOF™ Analyzer and ESI positive for the 4000 QTRAP® System, and the adduct was sodiated carbohydrates as $[M+Na]^+$.

Intuitive User Interface for Efficient Data Review

The interface is organized such that important information is at the user's finger tips (Figure 2). The navigation window on the left makes it easy to move from one MS/MS spectrum to the next within each project. Search results are displayed in two panes: the results display pane and the glycan information pane. The Glycan Structure button displays the glycan structure along with its sequence. If glycopeptide MS/MS data are used for prediction, then clicking the Glycan Structure button shows the molecular structure of the glycan along with the specified peptide sequence/mass, including any modifications. On clicking the Glycan Fragment button, a list of glycan fragments is displayed. Specific fragments can be searched by name, mass or m/z value.

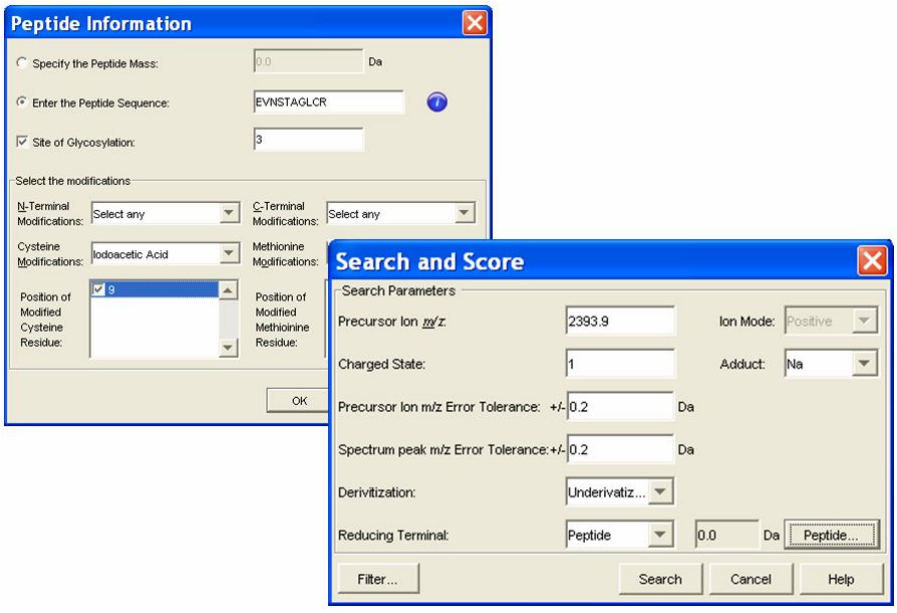


Figure 3. Carbohydate and Glycopeptide Search Parameters. The Search and Score window displays the required parameters for a carbohydrate search. An example of data input for a MALDI MS/MS spectrum containing sodium adducts is shown. In the case of a glycopeptide search, additional peptide information is required.

Data Processing of Carbohydrates and Glycopeptides with SimGlycan Software

In addition to being easy to use, SimGlycan software is extremely flexible for importing and exporting data. SimGlycan accepts MS/MS or MS^3 data in text format or Microsoft Excel format containing m/z and intensity. It directly imports data from the 4000 QTRAP® System, QSTAR® Elite System and the 4800 *Plus* MALDI TOF/TOF™ Analyzer. For files containing multiple spectra such as LC/MS/MS experiments, the user decides on the glycan or glycopeptide spectrum of choice in the acquisition software and then selects the data file and specific spectrum within SimGlycan software. SimGlycan then imports a peak list from this particular spectrum.

To search and score carbohydrate MS/MS or MS^3 data, simply fill in all the required parameters, precursor ion m/z, precursor ion m/z error tolerance, and spectrum peak m/z error tolerance. Other key settings included "Glycan Derivatization, "Reducing Terminal", "Ion mode" , and "Adduct ". For glycopeptide interpretation, the workflow is similar except the "Reducing Terminal" field is not applicable and additional parameters are employed such as peptide mass, sequence, and peptide modification (Figure 3). The results are viewed in the Search Result tab. Selecting the MS/MS Annotation tab displays matched m/z values in blue and unmatched m/z values in yellow. The Plot Mass Spectrum button creates an MS/MS spectrum of all matched and unmatched fragment ions (Figure 4).

A comprehensive glycan search report can be generated for lab notebooks, publication, or sharing information with colleagues. The report includes the glycan characteristics such as the identification, sequence, name, composition, mass, precursor ion m/z, class, reaction, and pathway, along with a graphical view of the glycan and its possible theoretical fragments (Figure 5). SimGlycan can save search results at any desired location by selecting the Export Results option. The results are exported to any supported file format for printing or further processing.

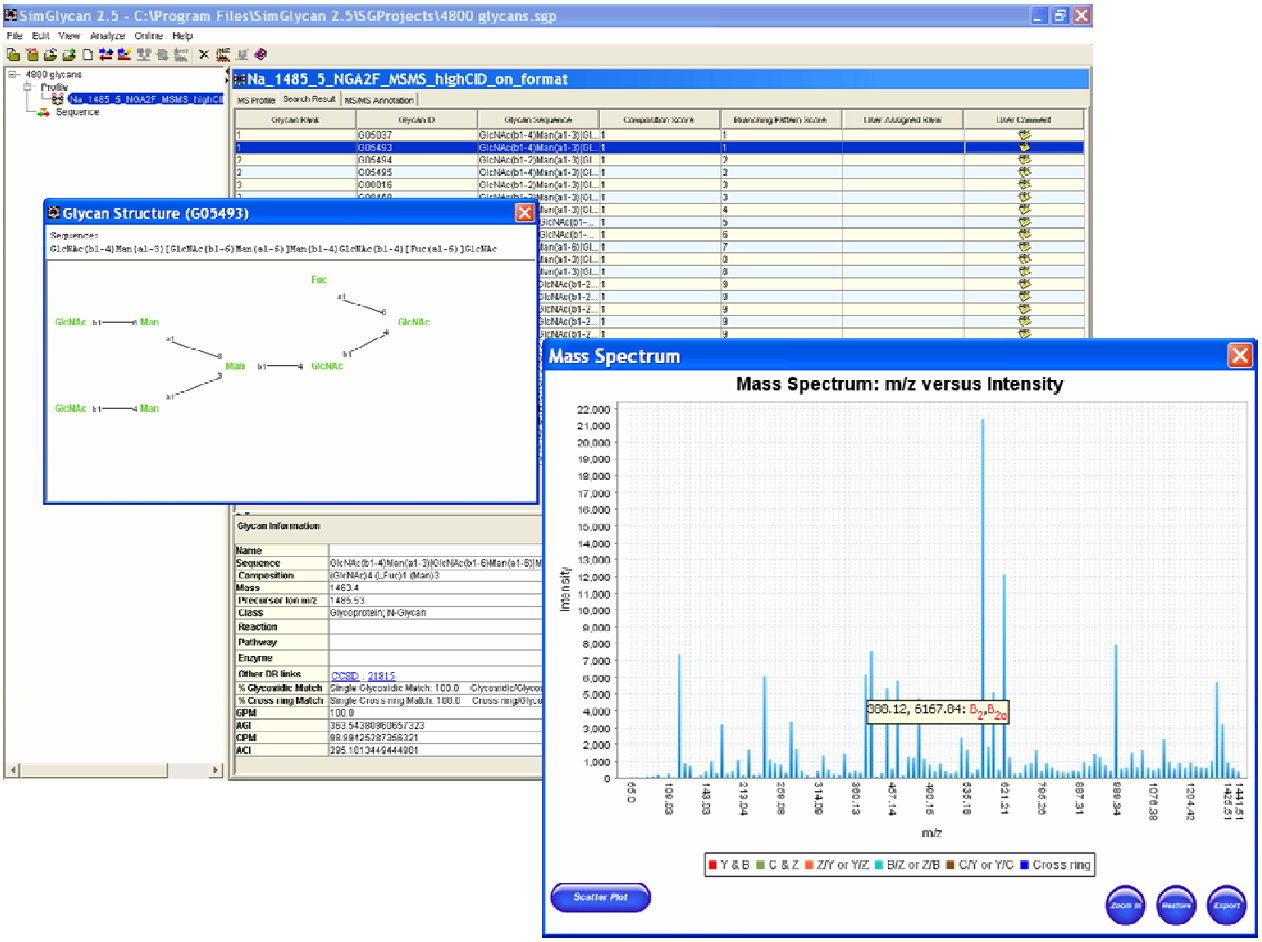


Figure 4. Search Results and MS/MS Annotation of m/z 1485.8 Precursor Generated by the 4800 *Plus* MALDI TOF/TOF™ Analyzer. Dragging the mouse over each annotated peak displays the fragment ion label using the nomenclature of Domon and Costello[1].

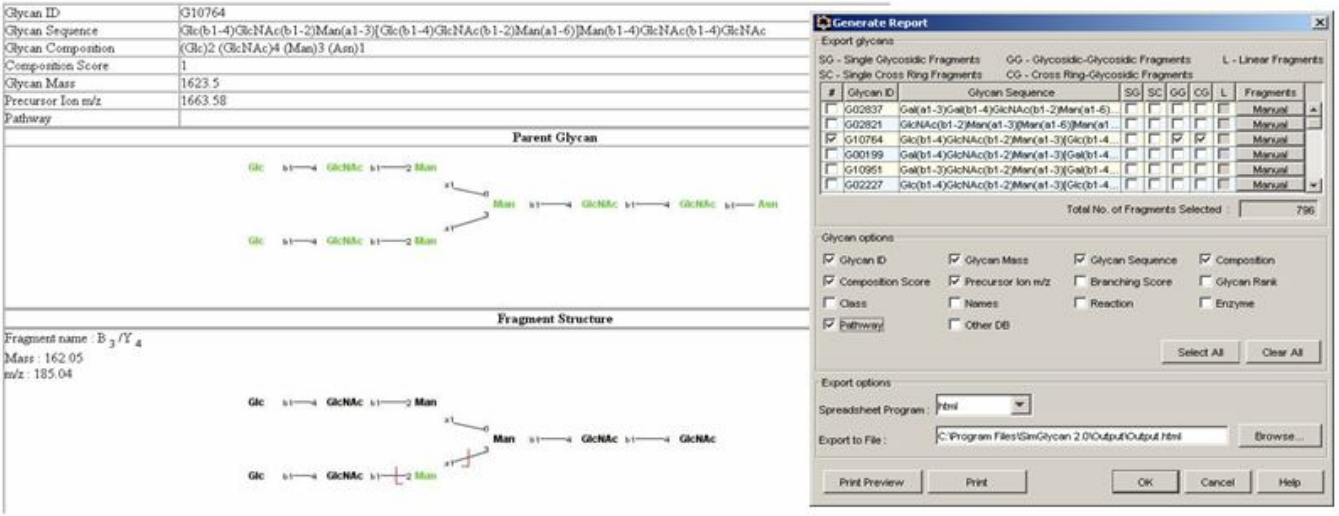


Figure 5. Generate Report and Report View. The Generate Report window provides a variety of display and reporting options.

4800 *Plus* MALDI TOF/TOF™ Analyzer for Carbohydrate Analysis

Because of its large mass range, high sensitivity, and soft desorption and ionization, MALDI time of-flight (MALDI-TOF) mass spectrometry is widely used in the molecular weight determination of underivatized oligo- and polysaccharides[2,3]. Since the introduction of the post-source decay (PSD) technique for MALDI TOF mass spectrometers, several studies concerning the PSD analysis of oligosaccharides have shown that much information can be obtained by MALDI-PSD-TOF experiments[4,5]. PSD (metastable or unimolecular decay) spectra of sodiated ions from neutral carbohydrates are dominated by glycosidic and internal cleavages, providing information related to sequence and branching; however, a lack of abundant cross-ring cleavages limits the linkage information that can be deduced from such experiments.

Cross-ring fragmentation can be induced by high energy collision-induced dissociation (CID) on MALDI TOF/TOF™ instruments[6,7]. Figure 6 shows a comparison at different instrument conditions (Collision induced dissociation (CID) vs. metastable decay) and their SimGlycan results for each MS/MS spectrum of Maltoheptaose $[M+Na]^+$, *m/z* 1175.3. As shown in Figure 6, higher energy CID provides cross-ring fragmentation and allows the assignment of the correct linkage, whereas PSD alone incorrectly identifies the glycan as a branched carbohydrate.

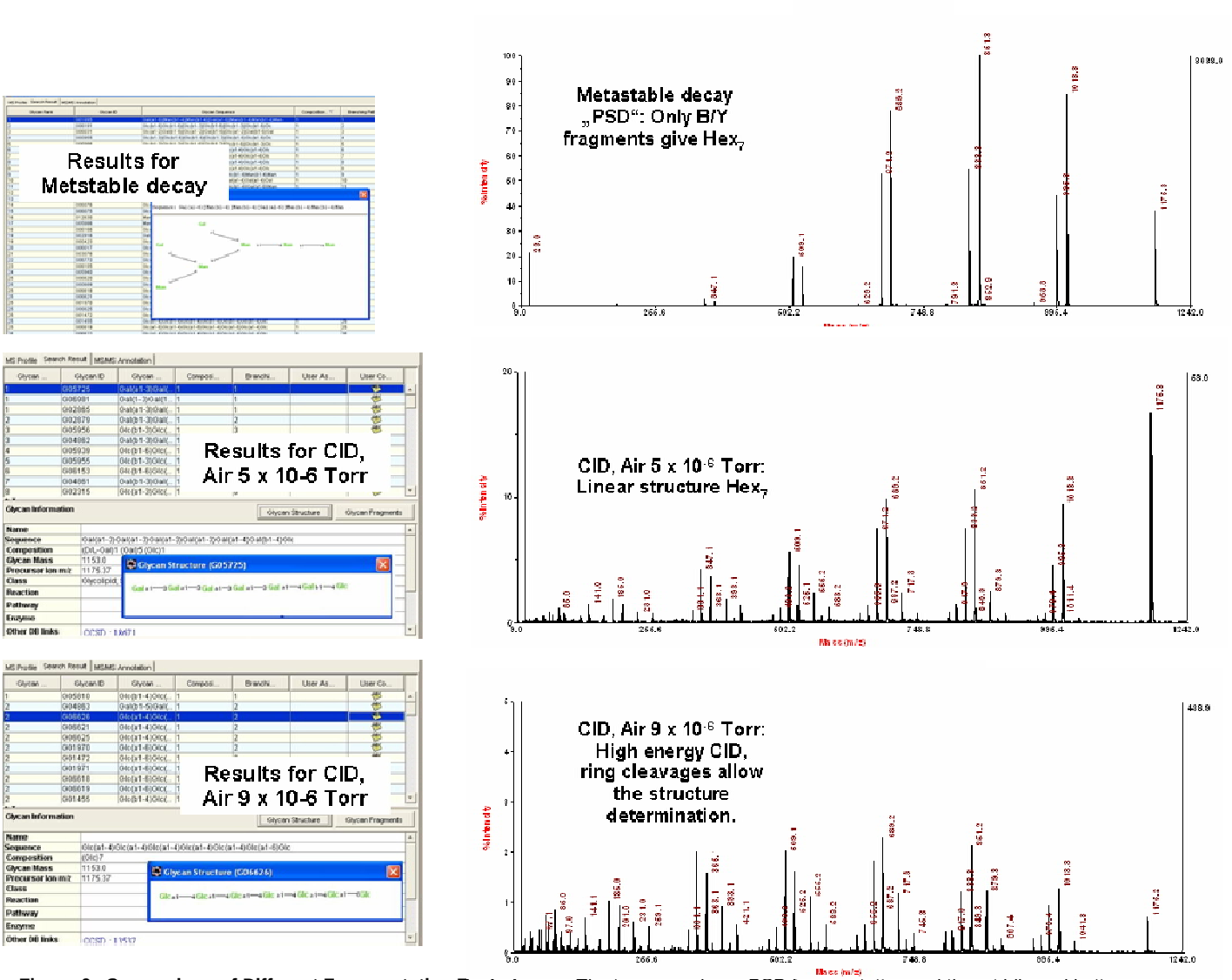


Figure 6. Comparison of Different Fragmentation Techniques. The top pane shows PSD fragmentation and the middle and bottom panes show true high energy CID using different collision gases. As shown, the CID spectra generate important cross ring fragments that allow full characterization of the saccharide sequence and linkage.

4000 QTRAP® System for Carbohydrate Analysis

QTRAP® systems (hybrid triple quadrupole linear ion trap instruments) are unique in their ability to perform post translational modification (PTM) discovery experiments. Coupling the triple quadrupole scan functions (precursor ion scan, MRM) with the high sensitivity ion trap MS/MS and MS^3 scans enables sensitive and selective experiments to be performed on glycan structure and connectivity. Figure 7 shows MS/MS and MS^3 data for a Neutral *N*-glycan. The major singly and doubly charged ions in the presence of sodium acetate were $[M + Na]^+$. The SimGlycan search results of the fragment data obtained from both the MS/MS and MS^3 spectra lead to the correct glycan identification. An MS^3 spectrum helped to confirm the glycan ID (Figures 7, 8). Alternatively, sequential exoglycosidase digests of complex glycans should be used to confirm anomeric linkage.

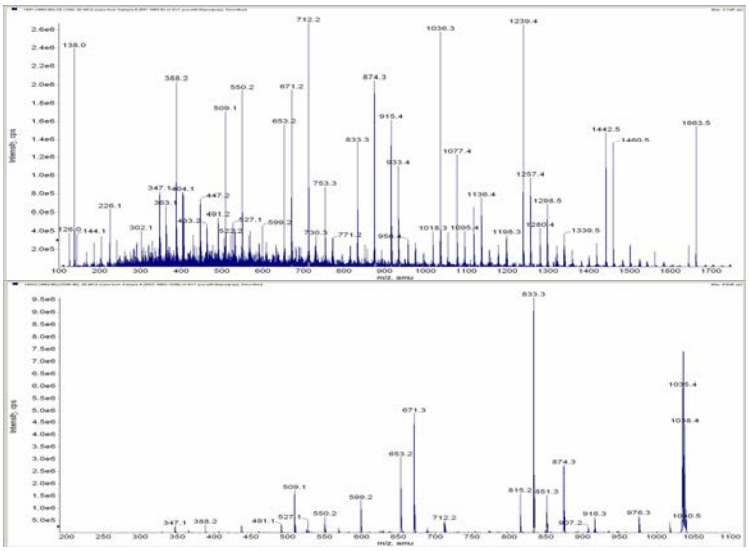


Figure 7. Glycan Spectra Generated by the 4000 QTRAP® System. Top: Glycan MS/MS *m/z* 1663.5+. Bottom: Glycan MS^3 *m/z* 1663.5+ > 1036.4+. Fragment ions produced in the MS^3 spectrum helped to confirm the glycan sequence identification.

QSTAR® Elite System for Carbohydrate Analysis

QSTAR Systems (QqTOF instruments) equipped with a NanoSpray® or oMALDI™ 2 Source can be successfully used for glycan analysis, performing MS/MS experiments in positive ion mode and delivering additional structural information in negative ion mode[8]. Additionally, the oMALDI 2 Source allows rapid MS detection of the intact mass and fast compositional analysis of the glycans by MS/MS. The combination of exoglycosidase digests with a QSTAR system allows complete glycan structure analysis in a relatively short period (hours) with small amounts of material (picomoles)[9]. Exoglycosidases produce highly specific glycosidic bond cleavages that are sensitive to the linkage, the stereochemistry, and even the configuration of the anomeric carbon. This workflow involves digestion of the oligosaccharide with defined exoglycosidases, followed by analysis of the digestion products. The sequence of the oligosaccharide and the specific linkages between the individual sugar moieties are deduced from the pattern of observed carbohydrate fragments and the specificity of the enzymes. SimGlycan Software can predict possible glycan IDs, which can be further confirmed by the MS profiles and MS/MS data of sequential exoglycosidase digests.

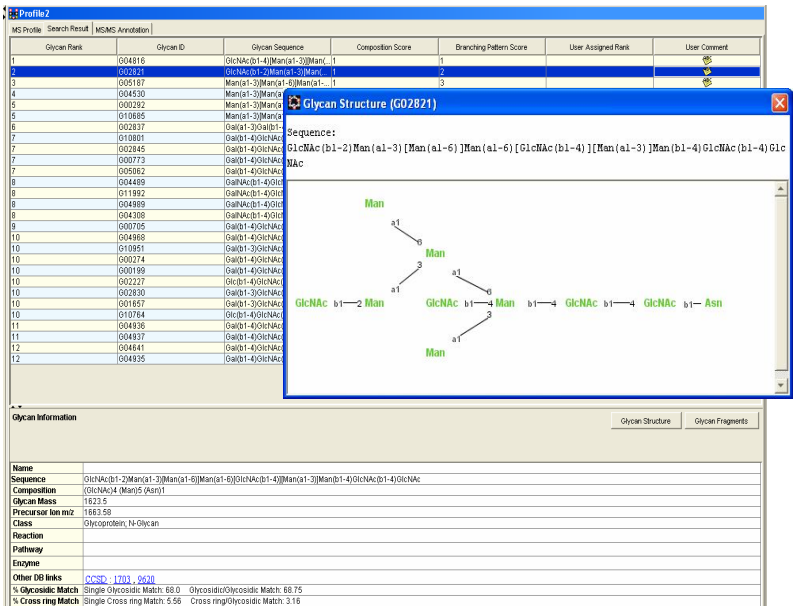
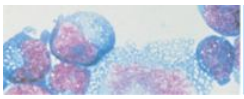


Figure 8: Search Results for 4000 QTRAP® System Neutral N-glycan Data. The top search result is the correct identification. Shown in the inset is the glycan structure for the top result. MS^3 data helped to confirm the identity of the glycan.



Conclusions

SimGlycan Software is a powerful program for the analysis of carbohydrates in general and glycans in particular. For every probable glycan structure, SimGlycan provides glycan fragments, structure, sequence, composition, glycan mass, class, reaction, pathway, enzyme and other database links. All the possible glycan structures are ranked based on SimGlycan's proprietary search and scoring algorithm.

In high energy MALDI CID MS/MS spectra generated by the 4800 *Plus* MALDI TOF/TOF™ Analyzer, a considerably greater number of fragments are found for the fragmentation of carbohydrates in comparison to other mass spectrometry fragmentation techniques. Although this could complicate any type of manual interpretation, it permits acquisition of more information on linkage position and points of branching, allowing the automated search routine in SimGlycan Software to correctly identify and characterize glycans without any secondary experiments.

On the other hand, the QSTAR® Elite and 4000 QTRAP® Systems generate low energy CID spectra. MS/MS data from sequential glycosidase digests should be used to confirm anomeric linkage. For the 4000 QTRAP System, a comparison was performed between collision cell MS/MS and ion trap MS/MS mode and the fragment ions observed were found to be very similar. The major fragments were B- and Y-type glycosidic cleavages with small amounts of cross-ring fragments defining linkage. Because of the unique instrument configuration in the QTRAP system, the MS/MS spectra do not show the low mass cut-off experienced by 3D-ion traps. Subsequent MS^3 fragmentation spectra were obtained with good signal/noise allowing mechanistic studies to be performed on the diagnostic fragments to confirm a predictive SimGlycan search result. In addition, the use of negative ions generated by either a QSTAR or QTRAP system give very specific fragmentation and details of the fine structure of the glycans (data are not shown).

References

1. Domon, Costello. Glycoconj. J. 5 (1988) 397-409.
2. Karas, Bachmann, Bahr, Hillenkamp. Int. J. Mass Spectrom. Ion Processes 1988; 78: 53.
3. Garozzo, Spina, Sturiale, Montaudo, Rizzo. Rapid Commun. Mass Spectrom. 1994; 8: 358.
4. Spengler, Kirsch, Kauffmann, Lemoine. J. Mass Spectrom. 1995; 30: 782.
5. Kauffmann, Chaurand, Kirsch, Spengler. Rapid Commun. Mass Spectrom. 1996; 10: 1199.
6. Mechref, Novotny, Krishnan. Anal. Chem. 2003, 75: 4895.
7. Spina, Sturiale, Romeo, Impallomeni, Garozzo, Waidelich, Glueckmann, Rapid Commun. Mass Spectrom. 2004, 18, 392-398.
8. Harvey, Thomas-Oates, Thomas, Waidelich, Main, Lenz, Hornshaw. ASMS2004.
9. Colangelo, Orlando. Anal. Chem., 1999, 71, (7), pp 1479–1482.

Founded in 1994, PREMIER Biosoft is a group of computer scientists and molecular biologists dedicated to producing cutting-edge intuitive software for the molecular biologist. The goal of the company is to study the most recent innovations in molecular biology and translate them into software products that aid biologists. Additional information about PREMIER Biosoft International is available at http://www.PremierBiosoft.com.

Literature Stock Number: 115TN14-02

Applied Biosystems

Headquarters
850 Lincoln Centre Drive | Foster City, CA 94404 USA
Phone 650.638.5800 | Toll Free 800.345.5224
www.appliedbiosystems.com

International Sales
For our office locations please call the division headquarters or refer to our Website at
www.appliedbiosystems.com/about/offices.cfm