Automated Spectrum Annotation and Structure Disambiguation of Released N-Glycans

Protein Metrics

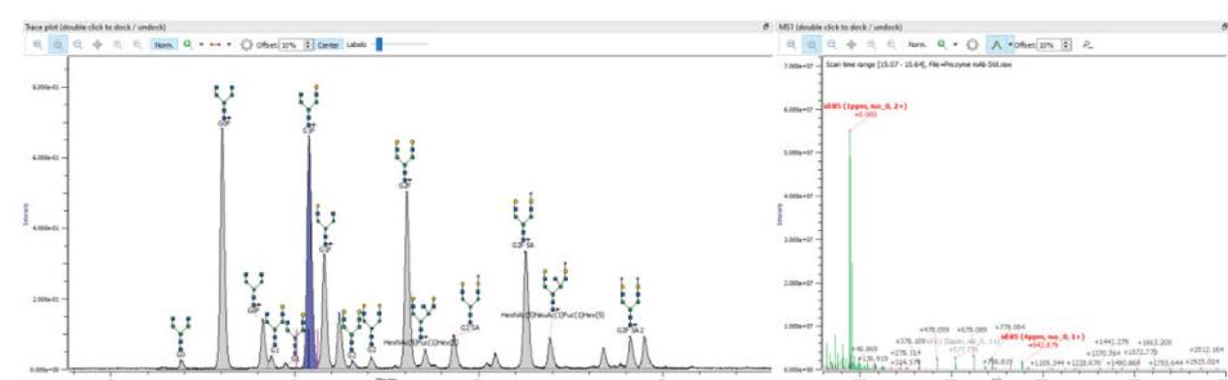
Gary M. Wilson[1], Anastasia Chernykh[2], Rebecah Kawahara[2], St. John Skilton[1], Morten Thaysen-Andersen[2], and Marshall W. Bern[1]

[1]Protein Metrics, Cupertino, CA, [2]Macquarie University, Sydney, Australia

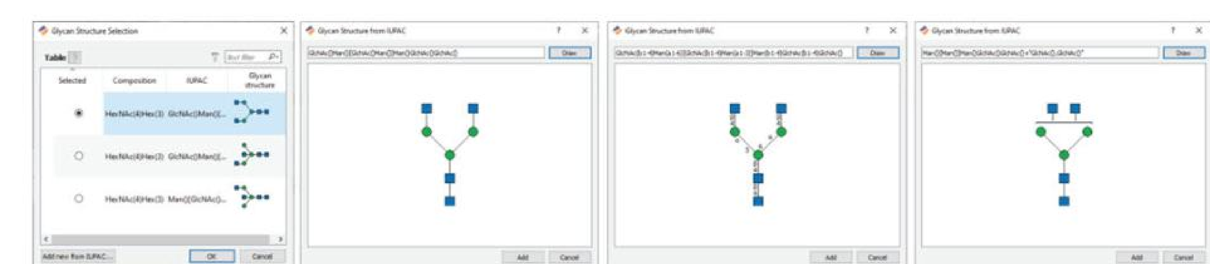
Introduction

Glycosylation is one of the most common and biologically important classes of protein modifications, yet it is also one of the most difficult to study. Exact structure determination currently requires multiple methods, such as mass spectrometry (MS), liquid or gas chromatography, enzyme deconstruction, immunoassays, etc. There is, however, a need for all-MS methods, even if these methods provide less than complete determination, in order to make glycosylation analysis available to non-specialist laboratories. Monosaccharide composition is often discernible directly from accurate molecular mass (MS[1]); however, topology (monosaccharide connectivity) requires MS[2], and glycosidic linkage information and anomericity requires, at least, MS[3] if not combined with orthogonal insight.[1] Here we report on disambiguation of topological isomers directly from tandem MS spectra.

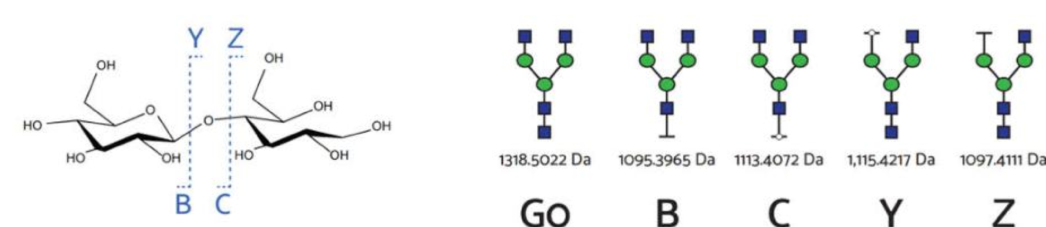
Background



Released glycans are mapped to elution peaks by accurate mass matching at the MS1 level, providing a confident assignment of the glycan composition. For fast comprehension, a cartoon representation of the glycans structure is available to be rendered in the standard notation for glycans (SNFG). While, MS1 alone does not inform of the glycan's fine structure, insight can be afforded by orthogonal evidence, such as the analyses of standards or retention time indexing to a dextran ladder (glucose units, GUs).[2]



Alternative representations of the glycan cartoon annotations is afforded to users by selection from a list of preset options (left). One may also provide new structures in condensed IUPAC format (middle left). Varying degrees of specificity are available. For example, if the user wishes to display complete specificity, edges can be annotated with anomeric and linkage information (middle right). Or, alternatively, if less specificity is needed, brackets denoting ambiguous connectivity can be rendered (right).



Additional evidence to a glycan's fine structure can be ascertained from a tandem mass spectrum following gas-phase fragmentation. Fragmentation at the glycosidic bond between monosaccharides produces B, C, Y, and Z-type fragments.[3] Cross-ring fragmentation, producing A and X-type fragments, may also occur, from which glycan linkage can be determined.[1]

Methods

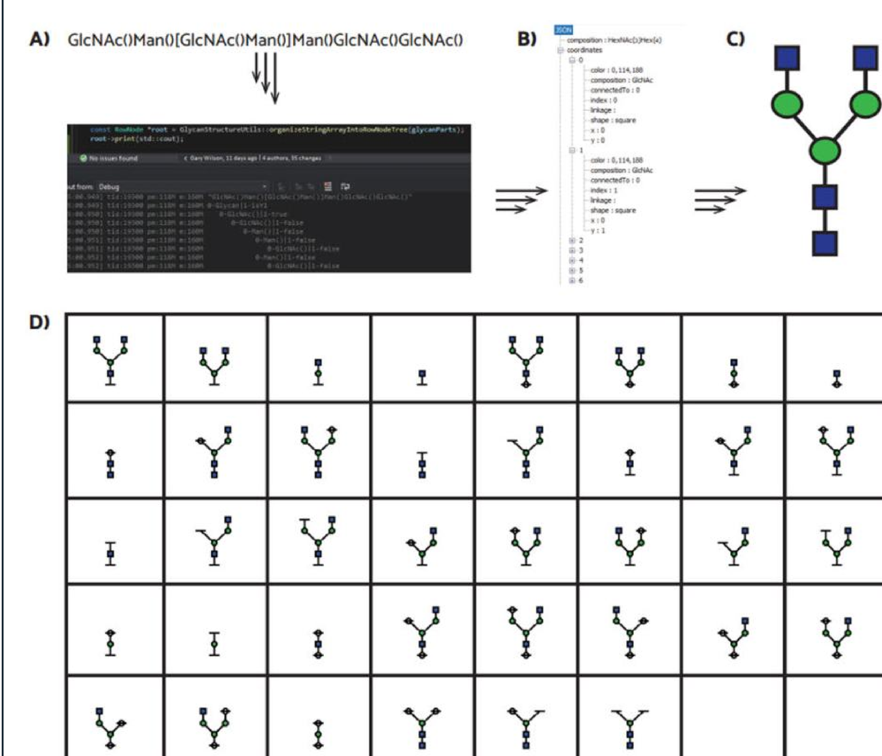


Figure 1: Conversion of the condensed IUPAC string to a non-binary tree data structure is a requisite step in the process of rendering to SNFG **(A)**.[4,5] A JSON object is subsequently generated from the tree structure, which holds cardinal coordinates, shape, color, and connectivity for the cartoon rendering **(B,C)**. The intermediate data structure utilized in this method thus enables the generation of connectivity-specific fragment ions by performing a tree traversal, which systematically produces all subgraphs associated with the glycan structure, in this case, considering a maximum of 2 bond breaks **(D)**. Candidate fragment ions are subsequently matched against spectrum peaks for annotation and probability scoring.[6] The data used for this study comes from released N-linked glycans from human plasma collected in negative-mode following ion trap CID on a Velos Pro (Thermo Scientific).[7]

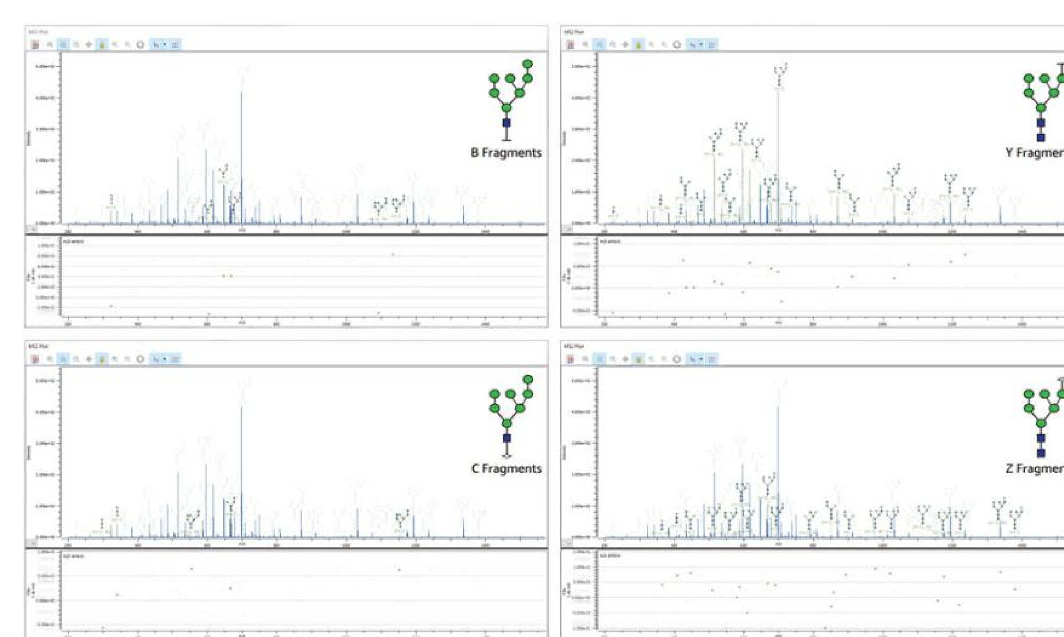


Figure 2: Here we show a representative tandem MS spectra of Man7 collected in negative mode following resonant excitation in a low pressure linear ion trap (CID). Released glycan spectra are annotated with B, C, Y, and Z-type fragment ions produced during in silicofragment generation, described in Figure 1. Cross-ring fragments are not yet considered as these types would require linkage-aware subgraph generation. Despite this, we often see comprehensive spectrum annotation when only considering glycosidic bond breaks. In general, Y-ions explain most prominent peaks (top right).

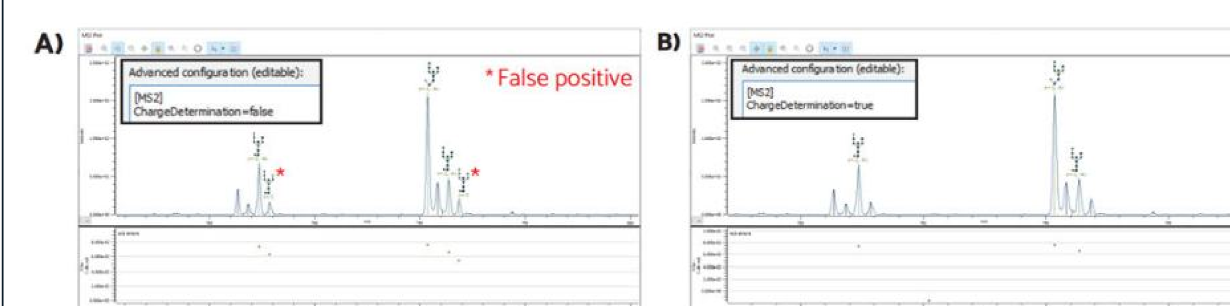


Figure 3: The use of spectrum peaks' charge and isotope number improves the accuracy and reliability of fragment ion identification, enabling more accurate structural elucidation of glycans. This problem is pronounced in glycan spectral annotation, compared to peptide fragment matching, as candidates differing by the mass of 1 or 2 protons is common. In the example to the left, we see that without spectral preprocessing, two glycan fragments at 782.8 and 791.8 m/z are falsely matched at +1 isotope peaks **(A)**. When charge and monoisotope identification is enabled, the false positive assignments are successfully culled while retaining accurate fragment matching at 782.3, 790.3, and 791.3 m/z **(B)**.

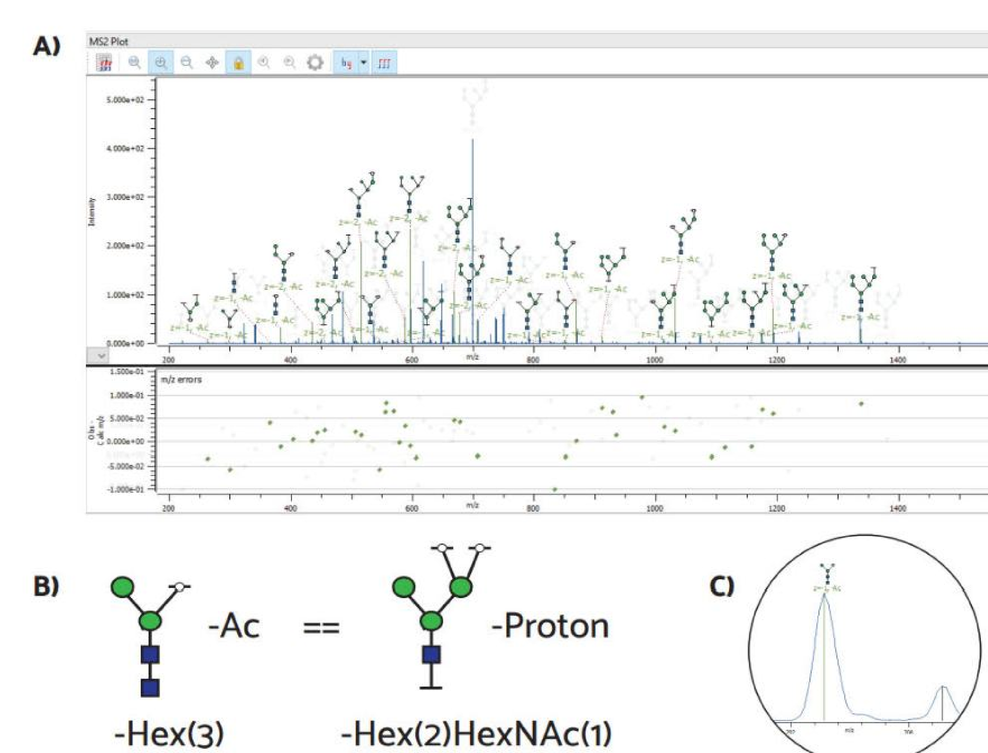


Figure 4: We observe that a significant proportion of the total ion current of glycan MS/MS spectra can be attributed to the neutral loss of acetate from glycan fragments **(A)**. This ion can be conflated for another fragment that differs by one proton, containing one more hexose and one less N-acetyl hexosamine **(B)**. Since the acetate loss fragment ions are prominant at monoisotopic peaks, we propose these fragments to be the more likely fragmentation product **(C)**.

Identification Summary

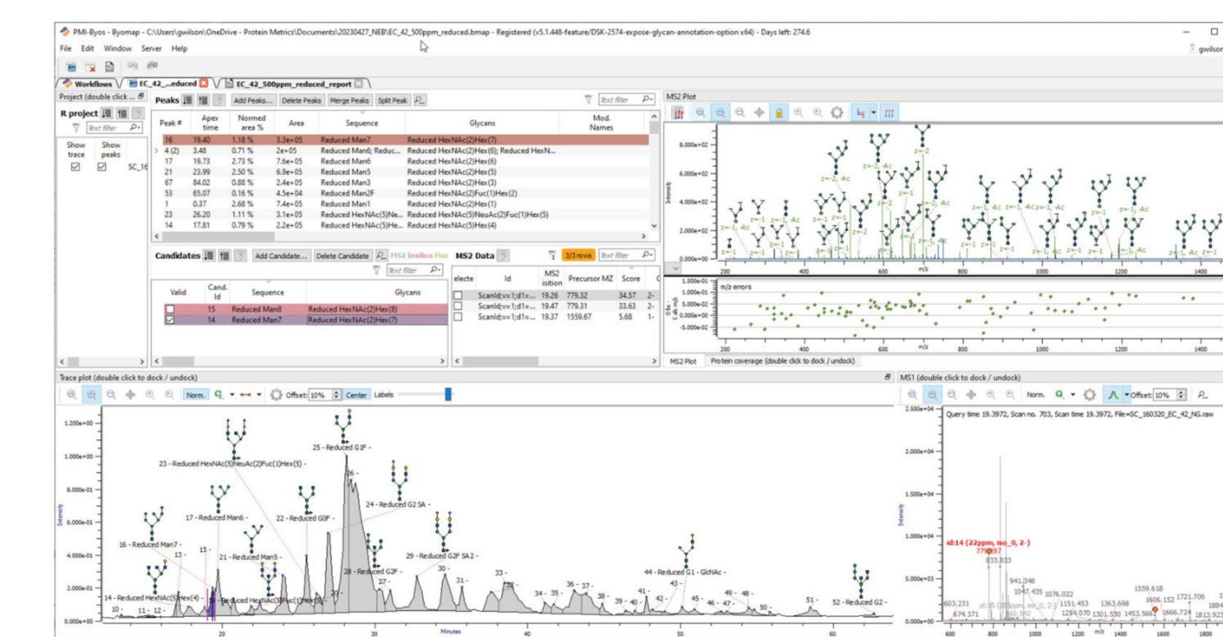


Figure 5: Released glycan workflows within Byos® generate chromatogram mapping projects for the analysis of complex biomolecule samples. The processing steps include integration of traces across various detector types, peak and feature finding, and peak matching to biomolecule candidates. With the addition of tandem MS features, including the MS2 Data table and MS2 Plot widgets, users can now use fragmentation data to better assist in the assignment of biomolecules to elution peaks. For the case of released glycans, the user can utilize these features to view fragment-matched and scored spectra to explore different glycan fine structures.

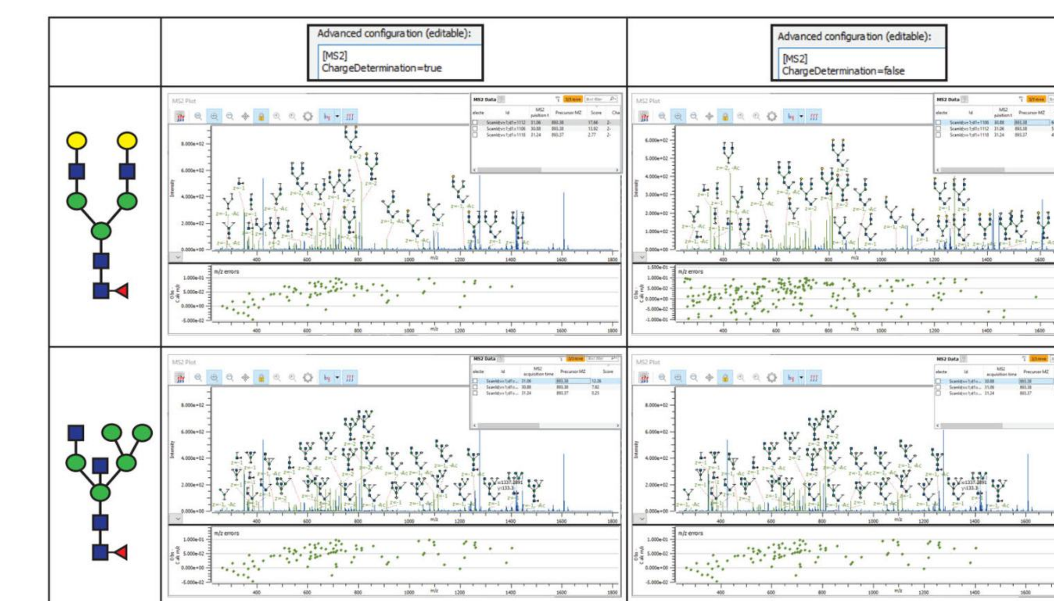


Figure 6: Glycan fragmentation can help deduce glycan topology, meaning the way in which monosaccharides are linked together in the tree structure. However, when performing the fragment matching to spectrum peaks, it is requisite to consider only monoisotopic peaks and to assess charge state or else false positive matches can mislead the interpretation (Figure 2). Here we see that an unlikely bisecting N-acetyl glucosamine structure (bottom) scores better than G2F (top) without this check (right). However, the more likely structure scores higher when charge and monoisotope matching is enabled (left).

References

1. Zaia, J. (2004). Mass spectrometry of oligosaccharides. Mass Spectrometry Reviews, 23(3), 161-227.
2. Ashwood, C., Pratt, B., MacLean, B. X., Gundry, R. L., & Packer, N. H. (2019). Standardization of PGC-LC-MS-based glycomics for sample specific glycotyping. Analyst, 144(11), 3601-3612.
3. Domon, B., & Costello, C. E. (1988). A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. Glycoconjugate Journal, 5(4), 397-409.
4. Mehta, A. Y., & Cummings, R. D. (2020). GlycoGlyph: A glycan visualizing, drawing and naming application. Bioinformatics, 36(11), 3613-3614.
5. Cheng, K. et al., 'DrawGlycan-SNFG: a robust tool to render glycans and glycopeptides with fragmentation information', Glycobiology 27(3): 200-205, 2017.
6. Perkins, D. N., Pappin, D. J. C., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data Several. Electrophoresis, 20(1), 3551-3567.
7. Chatterjee, S., Kawahara, R., Tjondro, H. C., Shaw, D. R., Nenke, M. A., Torpy, D. J., and Thaysen-Andersen, M. (2021). Serum N-glycomics stratifies bacteremic patients infected with different pathogens. Journal of Clinical Medicine. 10(3). 1-18.