



Published in final edited form as:

*Anal Chem.* 2012 September 4; 84(17): 7456–7462. doi:10.1021/ac301398h.

## Employment of tandem MS for the accurate and specific identification of oligosaccharide structures

Shuai Wu<sup>1</sup>, Juli Salcedo<sup>2</sup>, Ning Tang<sup>2</sup>, Keith Waddell<sup>2</sup>, Rudolf Grimm<sup>2,3</sup>, J. Bruce German<sup>4</sup>, and Carlito B. Lebrilla<sup>1,5,\*</sup>

<sup>1</sup>Department of Chemistry, University of California, Davis, CA 95616

<sup>2</sup>Agilent Technologies, Inc., Santa Clara, CA 95051

<sup>3</sup>Robert Mondavi Institute for Wine and Department of Food Science and Technology, University of California, Davis, CA 95616

<sup>4</sup>Department of Food Science and Technology, University of California, Davis, CA 95616

<sup>5</sup>Department of Biochemistry and Molecular Medicine, School of Medicine, University of California, Davis, CA 95616

### Abstract

A method is described for the rapid identification of oligosaccharides employing a library of tandem MS spectra. Identification is aided by software that compares the sample tandem MS to those in the library. The method incorporates quadrupole time-of-flight mass spectrometry along with an annotated oligosaccharide (OS) structure library and the MassHunter Personal Compound Database and Library (PCDL) software. With an automated spectra search, OS structures in different samples are readily identified. This method is shown to be useful in the study of milk oligosaccharides but can be readily applied to oligosaccharide pools in other biological tissues.

### Keywords

Glycobioinformatics; Spectral matching; MassHunter PCDL; oligosaccharides

### Introduction

The importance of glycosylation in protein function has led to the development of new methods for the analysis of oligosaccharides. Mass spectrometry has become the most sensitive and rapid method for characterizing oligosaccharides.<sup>1–13</sup> Recent developments in separation science by using different stationary phases<sup>14–18</sup> when coupled with mass spectrometry (MS) and tandem MS techniques improve these methods further yielding isomer specific analysis and even rapid structural identification.<sup>19–21</sup>

The combination of liquid chromatography and mass spectrometry has also made it easier to determine the extent or the size of the glycome. For example, mathematically there could be thousands of different human milk oligosaccharides (HMOs) based on the combinations of possible monosaccharide composition.<sup>22, 23</sup> However, LC-MS analysis shows that within a five order of magnitude range there are less than 150 for each individual, while there may be less than 500 different structures combined for all humans. Similarly, a separate systematic

To whom correspondence should be addressed: Carlito B. Lebrilla, cblebrilla@ucdavis.edu; Tel: +1-530-752-6364; Fax: +1-530-752-8995.

N-glycan study of serum showed using a neural network that there may be less than 400 compositions. When coupled with the empirical results of nanoLC-MS analyses, it can be estimated that there are probably less than 5000 structures.<sup>24</sup>

The realization that the glycome of specific biological systems is relatively small and finite will change our approach towards developing rapid methods for structural analysis of oligosaccharides. Rather than determining the structure *de novo* each time, a structure may be assigned a number of characteristics that will allow rapid identification once the structure is fully characterized. In this way, glycomics becomes more like metabolomics than proteomics. To this end, a functional database for the identification of oligosaccharides will have considerable value.

There are currently several major carbohydrate databases including the US Consortium for Functional Glycomics (CFG),<sup>25, 26</sup> the Kyoto Encyclopedia of Genes and Genomes (KEGG) Glycan,<sup>27</sup> and Glycosciences.de.,<sup>28–30</sup> which can be linked to mass spectra to provide potential structures. Additionally, there are also software packages that are used for semi-automated interpretation of MS spectra. They include Glycofragments from Glycoscience.de,<sup>31, 32</sup> GlycoPeakfinder<sup>22</sup> and GlycoworkBench<sup>33</sup> from Eurocarbdb, and GlycoMod and GlycanMass from Expasy.org.,<sup>34</sup> GlycosidiQ<sup>35</sup> and Cartoonist.<sup>36</sup> Software such as Glycofragments, GlycoPeakfinder, GlycosidiQ, and Glycoworkbench predict a structure based on the MS/MS data that is compared to extensive databases of published structures. Cartoonist predicts structures based on a biologically curated list of possible structures. While these approaches are potentially useful, the results are still mainly conjectures that need further rigorous verification.

An alternative approach is to identify known structures using a set of characteristics. These methods can be classified as structural identification rather than structural elucidation. They provide a rapid analysis of oligosaccharides but are bound to the known structures in a library. A successful method developed by Rudd and co-workers employ normal phase chromatography to construct a database of structures that are identified based on retention times.<sup>38</sup> This method employs a glucose ladder to assign a retention time that is independent of the specific HPLC brand but is specific to the chromatographic method. An alternative approach developed in this laboratory employed nanoLC retention times and high mass accuracy spectra to identify structure.<sup>39, 40</sup> A more recent study reported that groups of N- and O- linked glycans could be identified by comparing the “characteristic signal intensity profiles” to the multistage tandem MS spectra acquired from the same synthesized glycans.<sup>41</sup> The study compiled tandem MS spectra of neutral OS into the library—acidic glycans were de-sialylated prior to analysis. A hybrid quadrupole ion trap/time-of-flight (TOF) MS was used to generate multistage fragmentation. However, multistage fragmentation strategies are not amenable to rapid analysis, particularly in conjunction with liquid chromatography methods.

In this report, we examine the use of MS/MS data to identify oligosaccharide compounds. A reference library is employed that includes nearly 100 HMO structures that were previously elucidated.<sup>19, 20</sup> A commercial software, Personal Compound Database and Library (PCDL), was used to generate a library of structures characterized by tandem MS. The software MassHunter (Agilent Tech. Inc.) is used to compare an “unknown” spectrum to those in the library. The program includes a scoring system to quantify how closely the library spectra match the sample spectra allowing accurate identification of OS mixtures in human milk.

## Experimental Section

### Milk oligosaccharide extraction and reduction

Milk samples, previously stored in  $-80^{\circ}\text{C}$ , were completely thawed, and  $100\ \mu\text{L}$  was diluted with  $100\ \mu\text{L}$  nanopure water and centrifuged at  $4,000\ \text{g}$  at  $4\ ^{\circ}\text{C}$  for 30 min. After the top fat layer was removed, four volumes of chloroform/methanol (2:1 v/v) were added to the defatted sample. After centrifugation at  $4,000\ \text{g}$  for an additional 30 min at  $4\ ^{\circ}\text{C}$ , the upper organic layer was carefully removed. The supernatant (aqueous phase containing the milk oligosaccharide-rich fraction) was freeze-dried with a SpeedVac. The milk oligosaccharide samples (reconstituted in  $250\ \mu\text{L}$  nanopure water) were reduced by  $250\ \mu\text{L}$  of  $1.0\ \text{M}$  sodium borohydride aqueous solution at  $65\ ^{\circ}\text{C}$  for 1.5 hours. The resulting product was desalted and purified by solid phase extraction (SPE) using graphitized carbon cartridge (GCC).<sup>19</sup>

### HPLC-Chip/Q-TOF MS Analysis

The HPLC-Chip/Q-TOF MS instrument (Agilent Technologies, Santa Clara, CA) is equipped with an 1200 series nano-LC system and an 6520 Q-TOF MS coupled by a chip interface. The nano-LC has a capillary pump for sample enrichment and a nanoflow pump for separation. It is also equipped with microwell-plate auto-sampler maintained at  $6\ ^{\circ}\text{C}$  by a thermostat. The LC chip consisted of an enrichment column with a volume of  $40\ \text{nL}$  and an analytical column  $43 \times 0.075\ \text{mm}$  i.d., which were both packed with PGC having  $5\ \mu\text{m}$  pore size. Both pumps use binary solvent: **A** 3.0% ACN/water (v/v) with 0.1% formic acid and **B** 90% ACN/water (v/v) with 0.1% formic acid. A  $4\text{-}\mu\text{L}/\text{min}$  flow rate of solvent **A** was used for sample loading with a  $1\text{-}\mu\text{L}$  injection volume. A 45 minute gradient delivered by a nanoflow pump with a flow rate of  $0.3\ \mu\text{L}/\text{min}$  was used for separation: 2.5–20.0 min, 0–16% **B**; 20.0–30.0 min, 16–44% **B**; 30.0–35.0 min, 44–100% **B**; 35.0–45.0 min, 100% **B**; and a 20 minute equilibration time at 0% **B**. The data was collected in the positive ion mode and calibrated by a dual nebulizer electrospray source with calibrant ions over a broad mass to charge ( $m/z$ ) range. Mass accuracies are typically  $< 5\ \text{ppm}$  for MS and  $< 20\ \text{ppm}$  for MS/MS experiments.

The collision energy (CE) applied was based on the  $m/z$  of the ion with higher energy for larger ions. For this instrument, the collision energy (CE) was varied according to the equation:

$$CE(V) = \frac{\left(\frac{m}{z}\right)}{100(Da)} \times k + b$$

where  $k$  is the slope and  $b$  is the y-intercept of the equation (both of which can be adjusted by the users),  $m/z$  is the mass-to-charge ratio of the precursor ion. The equation was empirically determined by the manufacturer. In this study, five optimized values of slope and y-intercept were used to fragment OS with multiple CEs for more specific structure identification. Specifics are described with the data.

### Automated MS/MS search and compound identification

The automated compound identification was performed in the MassHunter Qualitative software. The MassHunter software offers two different spectral search algorithms – forward search and reverse search. These spectral match algorithms initially filter by applying a precursor ion filter to find the list of spectra that have the same precursor ion. For library spectra where the precursor ions match the unknown spectrum precursor ion, the search routine checks for matching ions between the unknown spectrum and the library spectra. For peaks that are in both spectra the algorithm uses a dot product calculation that

compares the relative intensities between the matched ions to determine the match score. The scoring for library searching uses a very similar algorithm as the NIST Library Search.<sup>42, 43</sup> When a matching peak is found, the program considers each mass peak to be a one dimensional vector and calculates the dot product between the two. The two spectra being matched are normalized. A number is returned between 0 (no match at all) and 1 (complete agreement). The sum of the dot products is used to create a score. This score is normalized to between 0 and 100 and reported as the match score.

The forward search compares all of the peaks in the unknown spectrum with all of the peaks in a library spectrum. A peak that is not in the library spectrum, or has a different relative peak intensity than a matching peak in the library spectrum, will be penalized in the scoring calculation. If there are peaks in the library spectrum that are not in the unknown spectrum, the score is also reduced. The forward search score is a result of a comparison of all the peaks in the unknown spectrum with all the peaks in the library spectrum (both  $m/z$  values and relative intensities). Forward search is useful for strong signals and minimizes false positives. For weak or noisy signals, forward search may yield lower scores since ions that are in the unknown spectra will not find matches in the library.

Reverse search takes peaks in the library spectra and matches them against the peaks in the unknown spectrum. The reverse search will penalize for a peak in the library spectrum that is not present in the unknown spectrum, and for ions of the same  $m/z$  value that are in both spectra but have different relative intensities. However, unlike forward searches, peaks that are in the unknown spectrum but not in the library spectrum are not considered in the calculation for the reverse search. The reverse search is useful for weak and noisy signals when trying to verify that the right ions were present. Reverse search may generate false positives but minimizes false negatives. The software allows searching the data using the combination of both forward and reverse searches and reports scores for each search type.

## Results and Discussion

A database of tandem MS composed of nearly 100 structures containing both neutral and sialylated human milk oligosaccharides (HMOs) was constructed. The structures used in this study were elucidated previously and characterized in earlier reports.<sup>19, 20</sup> For the analyses, HMOs were reduced, enriched and analyzed by nano-LC Q-TOF MS as described above. Figure S1 shows the chromatograms of representative HMO samples illustrating the distribution of the peaks across retention times and masses.

The ability to distinguish structures by tandem MS is illustrated with two compositions corresponding to hepta- and octasaccharides. Figure 1 shows the extracted ion chromatograms (EIC) of isomers with  $m/z$  611.25 (Figure 1a) and  $m/z$  756.78 (Figure 1b) in an HMO sample. Both masses show three major peaks in the chromatogram corresponding to three isomers for both  $m/z$  611.25 and  $m/z$  756.78. The structures were assigned previously by retention time and accurate mass (Table 1). In order to determine whether the correct structures could be correctly determined using the method, the MS/MS spectra of each peak were compared to the ones identified via PCDL search. Figure 2a, 2c, 2e are the respective tandem MS spectrum of 4120a, MFLNH III, IFLNH III from the library (red), while Figure 2b, 2d, 2f are tandem MS spectrum from the corresponding OS found in the milk sample (blue). In Figure 2a and 2b, the same fragment peaks were found in both the library spectra and the sample spectra. The ion intensity scale from the two spectra are slightly different ( $2.0 \times 10^3$  cps in Figure 2a and  $3.0 \times 10^3$  cps in Figure 2b). However, the relative intensities of the fragment ions appeared highly similar.

The PCDL search results for  $m/z$  611.2 with the spectrum in Figure 2b yielded four compounds shown in Table 2 (Please see also Figure S2). Two types of searches were used.

The “Match (Forward)” score was calculated by searching peaks in the unknown spectrum against the library spectrum. The “Reverse” score was calculated by searching peaks in the library spectrum against the compound spectrum. In both cases, compound 4120a (Table 2) has the highest matching scores among the four isomers (Figure S2). Furthermore, the tandem MS is also visibly closest to the reference compound (Figure 2a). Similarly, the search results of MFLNH III and IFLNH III are listed in Table S1a and S1b, respectively. In both cases, the correct isomer yielded the highest scores. For IFLNH III, only one compound was selected.

To determine the repeatability of the method, the same sample was run in five consecutive runs. Table S2a–e list the automated search results of the sample from five consecutive runs. The data from five runs were searched against the HMO structure library file separately. Over 30 structures were identified by tandem MS in each of the five runs. There are typically approximately 100 structures from individual subjects, with approximately 30 structures representing nearly 99% of the total abundances. Without exception, all of the same structures were identified in each of the five runs.

### Multiple Collision Energies

A potentially important parameter in obtaining unique fragmentation patterns is the collision energy (CE) of the tandem MS experiments. The energetics of collision-induced dissociation (CID) and fragment ion yield depends on the size of the ion with higher energy generally required to obtain useful fragment ions for larger precursor ions.<sup>44–46</sup> For these experiments, the CE was adjusted based on the  $m/z$  value according to the expression described in Table 3. To determine whether the compound identification is sensitive to CE, a systematic study was performed where the sample was subjected to a broad CE range. Figure 3 shows the tandem MS spectra of FS-LNnH I obtained from milk samples and under five different CEs. The compound was identified and validated by retention times and accurate mass of a commercial standard. As the CE increases, the intensity of the precursor ion decreases and the relative intensities between fragment ions vary accordingly. Each spectrum was searched with MassHunter Qual against a library with spectra corresponding to the same five CEs. The results tabulated in Table 4a correspond to the use of Figure 3a as the sample spectrum. The top four results correspond to the correct compound FS-LNnH I, however the top hit does not necessarily correspond to the same energy (6.3V and 5.6V), respectively. Indeed, the top result for all the spectra correspond to the entry 6.3V except for the highest energy where the sample CE of 9.2V scores the best for the library entry of 9.2 V. Incorrect entries such as FS-LNH I in Table 4a do not score well compared to the correct entries. In general, the correct compound was obtained as confirmed by the accurate mass of the precursor ion and the nanoLC retention times.

### Application of the library search to different types of samples

The program was validated using samples of milk from nonhuman species, namely orangutan milk. Because human and primates are genetically similar, there is considerable overlap between HMOs and primate milk oligosaccharides (PMOs) (Figure S3)<sup>47</sup>. The primate samples were prepared and analyzed in exactly the same way as the human milk samples using nanoLC-MS. The library search strategy using reproducible retention times and accurate masses has previously been applied to identify oligosaccharides in primate milk.<sup>47</sup>

The results of the analysis are summarized in Table S3. MS scans provided accurate masses of PMOs that were used to assign composition, using the in-house software – “oligosaccharide calculator”. In the orangutan milk samples, 49 compounds along were obtained, including five containing NeuGc (Table S3a). The software MassHunter Qual was

used to perform an automated library search by comparing the MS/MS spectra from the sample in the PCDL. The results produced 24 HMOs that were verified based with the method employing accurate masses and reproducible retention times (Table S3b). Comparison of the tandem MS with this previous method showed there were no false positive determinations in this analysis.

## Conclusion

From the above results, automated spectral matching is an efficient and reliable method to identify oligosaccharide structures using an MS/MS spectral library containing known structures. The advantages of this method include:

1. Nano-LC using PGC as stationary phase provides excellent isomer separation for OS. Q-TOF MS provides high mass accuracy, high resolution and high-speed data collection. The combination of highly reproducible nano-LC, high mass accuracy MS, and tandem MS provides orthogonal dimensions that allow highly accurate compound identification.
2. Program assisted structural identification is fast, taking only a few minutes to identify compounds in an LC/MS run. This compares well to the traditional method that often requires extensive chemical analysis with enzymatic digestion requiring extensive sample preparation and time-consuming data analysis to identify glycans in new samples. By employing program assisted spectral matching, the time for the whole process can be reduced from weeks or even months to only minutes.
3. Once the library is constructed, the program is simple to use. This allows users who are not experts in carbohydrate structural analysis to use the program for their study. The spectra matching method can be applied to other types of carbohydrates such as N-linked and O-linked structures.

This strategy has certain limitations. The user may require a single instrument platform to match the conditions so that effective searches can be performed. However, as better bioinformatics tools emerge, it may be possible in the future to search tandem MS spectra from different databases. To be effective, this method also requires a much larger number of tandem MS spectra from fully or partially characterized glycans. This work is currently in progress with oligosaccharides in serum and other tissues to be available soon. Once these tasks are completed, a method for identifying oligosaccharides incorporating accurate masses, retention times, and tandem MS will be available so that even novices can study oligosaccharides and glycobiology will be greatly enhanced..

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Supported in part by grants from the Eunice K. Shriver National Institute of Child Health and Human Development Grant (HD059127, HD061923, GM049077), the National Center for Research Resources, a component of the National Institutes of Health, (UL1 RR024146). Dr. Pascal Gagneux from the University of California – San Diego and Smithsonian Zoological Park is gratefully acknowledged for providing the primate milk samples. Dr. Frank Kuhlmann and Dr. Joe Roark from Agilent Technologies, Inc. – Santa Clara, CA are gratefully acknowledged for providing technique support for MassHunter and PCDL software and comments on the manuscript.

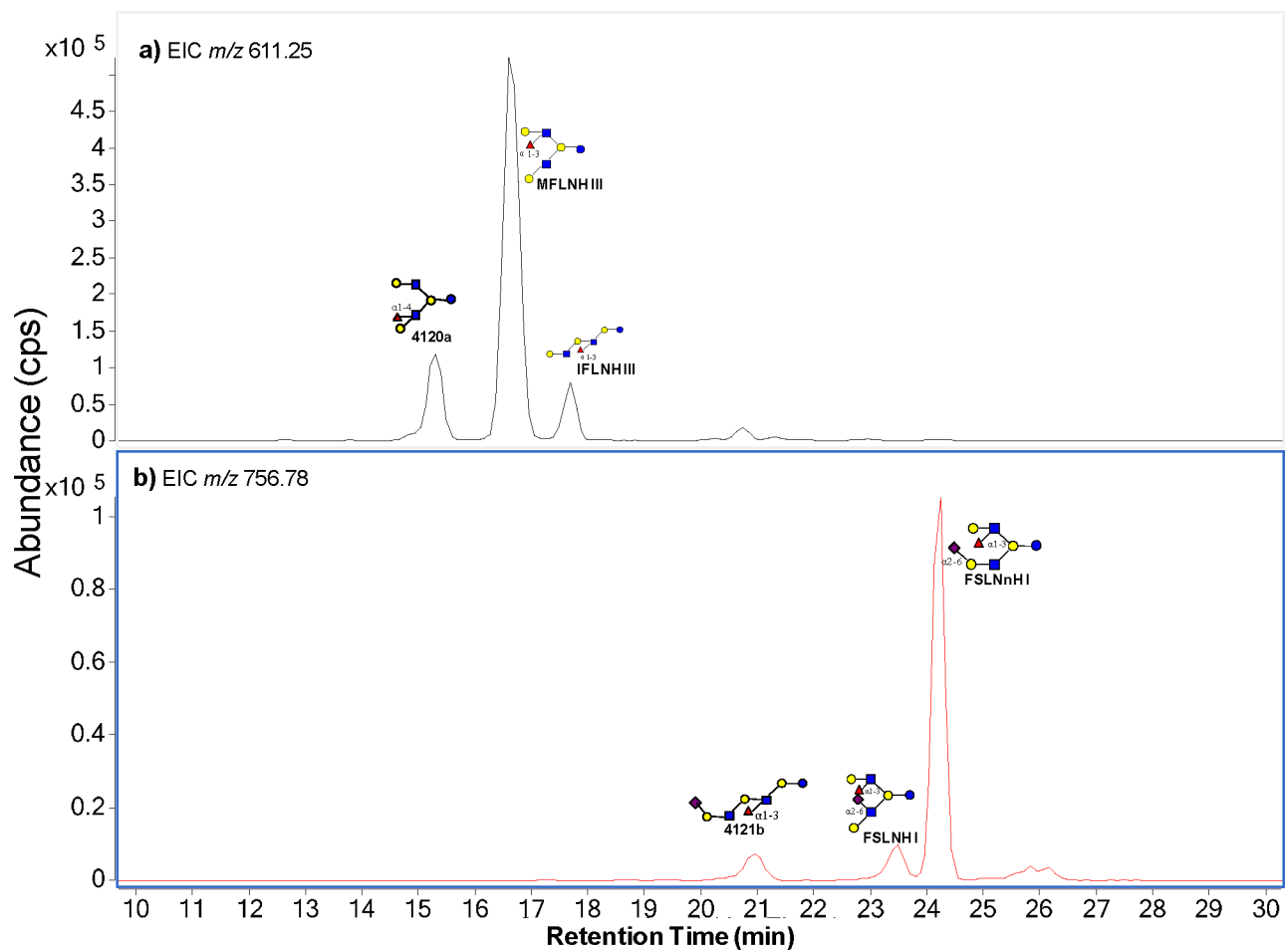
## References

1. Penn SG, Cancilla MT, Lebrilla CB. *Anal Chem.* 1996; 68(14):2331–2339. [PubMed: 8686926]

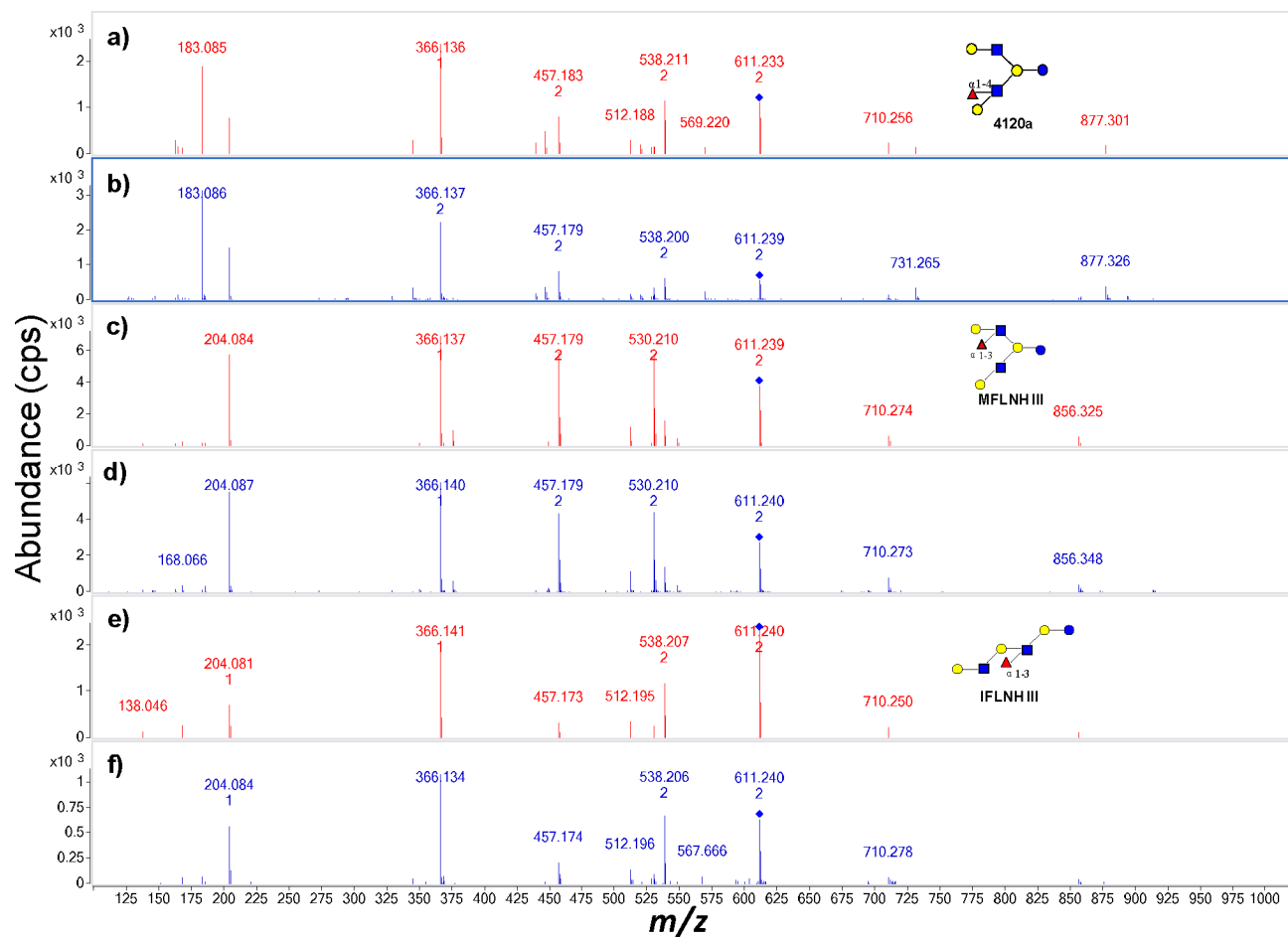
2. Harvey DJ. *J Mass Spectrom.* 2000; 35(10):1178–1190. [PubMed: 11110090]
3. Pfenninger A, Karas M, Finke B, Stahl B. *J Am Soc Mass Spectr.* 2002; 13(11):1331–1340.
4. Pfenninger A, Karas M, Finke B, Stahl B. *J Am Soc Mass Spectr.* 2002; 13(11):1341–1348.
5. Mechref Y, Novotny MV, Krishnan C. *Anal Chem.* 2003; 75(18):4895–4903. [PubMed: 14674469]
6. Zaia J. *Mass Spectrom Rev.* 2004; 23(3):161–227. [PubMed: 14966796]
7. Zhang JH, Schubotho K, Li BS, Russell S, Lebrilla CB. *Anal Chem.* 2005; 77(1):208–214. [PubMed: 15623298]
8. Pikulski M, Hargrove A, Shabbir SH, Anslyn EV, Brodbelt JS. *J Am Soc Mass Spectr.* 2007; 18(12):2094–2106.
9. Adamson JT, Hakansson K. *Anal Chem.* 2007; 79(7):2901–2910. [PubMed: 17328529]
10. Adamson JT, Hakansson K. *J Am Soc Mass Spectr.* 2007; 18(12):2162–2172.
11. Wolff JJ, Laremore TN, Aslam H, Linhardt RJ, Amster IJ. *J Am Soc Mass Spectr.* 2008; 19(10):1449–1458.
12. Wolff JJ, Laremore TN, Busch AM, Linhardt RJ, Amster IJ. *J Am Soc Mass Spectr.* 2008; 19(6):790–798.
13. Amano J, Sugahara D, Osumi K, Tanaka K. *Glycobiology.* 2009; 19(6):592–600. [PubMed: 19240273]
14. Koizumi K. *J Chromatogr A.* 1996; 720(1–2):119–126. [PubMed: 8601188]
15. Pabst M, Altmann F. *Anal Chem.* 2008; 80(19):7534–7542. [PubMed: 18778038]
16. Ruhaak LR, Deelder AM, Wuhrer M. *Anal Bioanal Chem.* 2009; 394(1):163–174. [PubMed: 19247642]
17. Guile GR, Rudd PM, Wing DR, Prime SB, Dwek RA. *Anal Biochem.* 1996; 240(2):210–226. [PubMed: 8811911]
18. Staples GO, Bowman MJ, Costello CE, Hitchcock AM, Lau JM, Leymarie N, Miller C, Naimy H, Shi X, Zaia J. *Proteomics.* 2009; 9(3):686–95. [PubMed: 19137549]
19. Wu S, Tao N, German JB, Grimm R, Lebrilla CB. *J Proteome Res.* 2010; 9(8):4138–51. [PubMed: 20578730]
20. Wu S, Grimm R, German JB, Lebrilla CB. *J Proteome Res.* 2011; 10(2):856–68. [PubMed: 21133381]
21. Aldredge D, An HJ, Tang N, Waddell K, Lebrilla CB. *J Proteome Res.* 2012; 11(3):1958–68. [PubMed: 22320385]
22. Maass K, Ranzinger R, Geyer H, von der Lieth CW, Geyer R. *Proteomics.* 2007; 7(24):4435–44. [PubMed: 18072204]
23. Cooper CA, Harrison MJ, Wilkins MR, Packer NH. *Nucleic Acids Res.* 2000; 29(1):332–335. [PubMed: 11125129]
24. Yang ZR, Chou KC. *Bioinformatics.* 2004; 20(6):903–8. [PubMed: 14751965]
25. Raman R, Raguram S, Venkataraman G, Paulson JC, Sasisekharan R. *Nat Methods.* 2005; 2(11):817–24. [PubMed: 16278650]
26. Raman R, Venkataraman M, Ramakrishnan S, Lang W, Raguram S, Sasisekharan R. *Glycobiology.* 2006; 16(5):82R–90R.
27. Hashimoto K, Goto S, Kawano S, Aoki-Kinoshita KF, Ueda N, Hamajima M, Kawasaki T, Kanehisa M. *Glycobiology.* 2006; 16(5):63R–70R.
28. von der Lieth CW, Lutteke T, Frank M. *Biochim Biophys Acta.* 2006; 1760(4):568–77. [PubMed: 16459020]
29. Lutteke T. *Chembiochem.* 2008; 9(13):2155–60. [PubMed: 18693281]
30. Lutteke T, Bohne-Lang A, Loss A, Goetz T, Frank M, von der Lieth CW. *Glycobiology.* 2006; 16(5):71R–81R.
31. Lohmann KK, von der Lieth CW. *Proteomics.* 2003; 3(10):2028–35. [PubMed: 14625865]
32. Lohmann KK, von der Lieth CW. *Nucleic Acids Res.* 2004; 32(Web Server issue):W261–6. [PubMed: 15215392]
33. Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. *J Proteome Res.* 2008; 7(4):1650–9. [PubMed: 18311910]

34. Cooper CA, Gasteiger E, Packer NH. *Proteomics*. 2001; 1(2):340–9. [PubMed: 11680880]
35. Joshi HJ, Harrison MJ, Schulz BL, Cooper CA, Packer NH, Karlsson NG. *Proteomics*. 2004; 4(6): 1650–64. [PubMed: 15174134]
36. Goldberg D, Sutton-Smith M, Paulson J, Dell A. *Proteomics*. 2005; 5(4):865–75. [PubMed: 15693066]
37. von der Lieth CW, Bohne-Lang A, Lohmann KK, Frank M. *Brief Bioinform*. 2004; 5(2):164–78. [PubMed: 15260896]
38. Mattu TS, Royle L, Langridge J, Wormald MR, Van den Steen PE, Van Damme J, Opendakker G, Harvey DJ, Dwek RA, Rudd PM. *Biochemistry-US*. 2000; 39(51):15695–704.
39. Royle L, Campbell MP, Radcliffe CM, White DM, Harvey DJ, Abrahams JL, Kim YG, Henry GW, Shadick NA, Weinblatt ME, Lee DM, Rudd PM, Dwek RA. *Anal Biochem*. 2008; 376(1):1–12. [PubMed: 18194658]
40. Campbell MP, Royle L, Radcliffe CM, Dwek RA, Rudd PM. *Bioinformatics*. 2008; 24(9):1214–6. [PubMed: 18344517]
41. Kameyama A, Kikuchi N, Nakaya S, Ito H, Sato T, Shikanai T, Takahashi Y, Takahashi K, Narimatsu H. *Anal Chem*. 2005; 77(15):4719–25. [PubMed: 16053281]
42. Stein SE, Scott DR. *J Am Soc Mass Spectr*. 1994; 5(9):859–866.
43. Stein SE. *J Am Soc Mass Spectr*. 1994; 5(4):316–323.
44. Cancilla MT, Wong AW, Voss LR, Lebrilla CB. *Anal Chem*. 1999; 71(15):3206–18. [PubMed: 10450162]
45. Laskin J, Futrell JH. *Mass Spectrom Rev*. 2005; 24(2):135–67. [PubMed: 15389858]
46. Sleno L, Volmer DA. *J Mass Spectrom*. 2004; 39(10):1091–112. [PubMed: 15481084]
47. Tao N, Wu S, Kim J, An HJ, Hinde K, Power ML, Gagneux P, German JB, Lebrilla CB. *J Proteome Res*. 2011; 10(4):1548–57. [PubMed: 21214271]

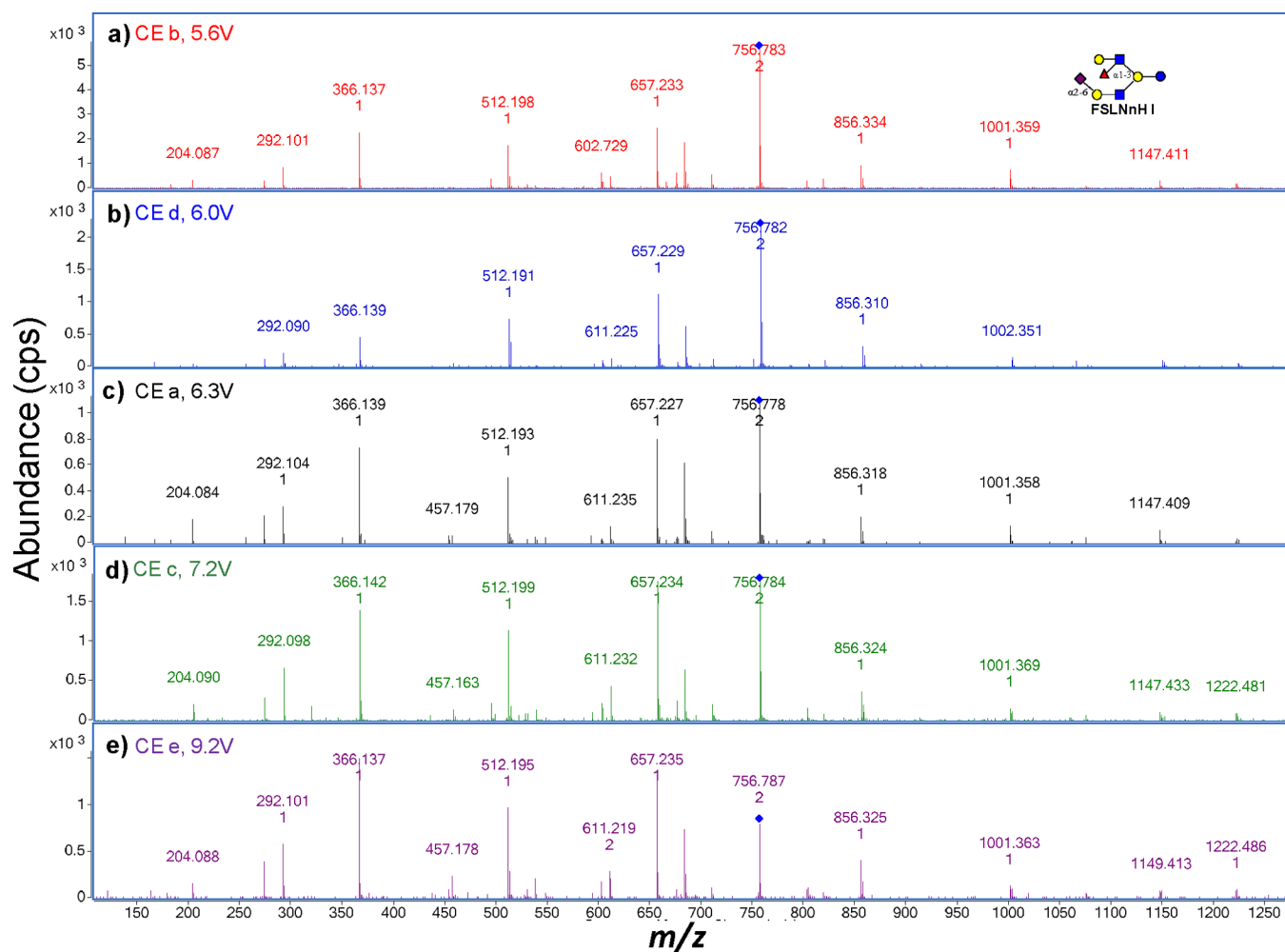




**Figure 1.** The extracted ion chromatogram (EIC) of isomers with  $m/z$  611.25 (a) and  $m/z$  756.78 (b) from a sample of human milk.



**Figure 2.** Comparison of tandem MS between sample and library spectra of three isomers – 4120a, MFLNH III and IFLNH III with  $m/z$  611.25. Spectra (a), (c), (e) are from the library while (b), (d), (f) are from the sample.



**Figure 3.** Tandem MS spectra of FS-LNnH I under five collision energies derived from conditions and equations illustrated in Table 3.

Table 1

Isomers with neutral mass 1220.454 and 1511.550 determined by Chip/Q-TOF.

a) Isomers with neutral mass (1220.454)							
Name	<i>m/z</i>	Hex	Fuc	HexNAc	NeuAc	RT (min)	Abundance
4120a	611.250	4	1	2	0	15.3	2.49*10 <sup>6</sup>
MFLNH III	611.252	4	1	2	0	16.6	1.30*10 <sup>7</sup>
IFLNH III	611.249	4	1	2	0	17.7	1.04*10 <sup>6</sup>

b) Isomers with neutral mass (1511.550)							
Name	<i>m/z</i>	Hex	Fuc	HexNAc	NeuAc	RT (min)	Abundance
4121b	756.785	4	1	2	1	21.0	1.96*10 <sup>5</sup>
FSLNH I	756.782	4	1	2	1	23.5	1.85*10 <sup>5</sup>
FSLNnH I	756.787	4	1	2	1	24.2	1.58*10 <sup>6</sup>

**Table 2**

The PCDL search result for isomer 4120a, including the  $m/z$  of the precursor ion, collision energy, forward and reverse matching scores and the neutral mass in the library.

Name	Precursor	CE	Match Score	Reverse Score	Mass (Library)
4120a	611.2	4.5	75.6	79.1	1220.5
IFLNHI	611.2	4.5	74.6	77.5	1220.5
MFpLNHIV	611.2	4.5	71.3	72	1220.5
4120b unknown	611.2	4.5	67	65.8	1220.5

\$watermark-text

\$watermark-text

\$watermark-text

**Table 3**

The summary of slope ( $k$ ) and intercept ( $b$ ) applied for collision energy equation in this study.

	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>	<b>e</b>
$k$	1.3	1	1.15	1.15	1.15
$b$	-3.5	-2	-1.5	-2.75	0.5

$$CE(V) = \frac{(m/z)}{100(Da)} \times k + b$$

**Table 4**  
The PCDL search results for FS-LNnHI using five collision energies as applied in Table 3.

<b>Table 4a</b>					
Name	Precursor	CE	Match Score	Reverse Score	Mass (Library)
FS-LNnHI	756.8	6.3	70.1	83.4	1511.5
FS-LNnHI	756.8	6.0	68.5	79.2	1511.5
FS-LNnHI	756.8	7.2	67.4	77.7	1511.5
FS-LNnHI	756.8	5.6	66.0	78.0	1511.5
FS-LNH I	756.8	6.3	65.0	76.3	1511.5
FS-LNH I	756.8	5.6	61.0	67.4	1511.5

<b>Table 4b</b>					
Name	Precursor	CE	Match Score	Reverse Score	Mass (Library)
FS-LNnHI	756.8	6.3	72.2	83.2	1511.5
FS-LNnHI	756.8	7.2	70.4	77.0	1511.5
FS-LNH I	756.8	6.3	69.1	77.4	1511.5
FS-LNnHI	756.8	6.0	69.0	77.5	1511.5
FS-LNnHI	756.8	5.6	67.6	76.7	1511.5
FS-LNH I	756.8	5.6	61.8	66.1	1511.5
412.1a	756.8	6.3	60.8	58.8	1511.5

<b>Table 4c</b>					
Name	Precursor	CE	Match Score	Reverse Score	Mass (Library)
FS-LNnHI	756.8	6.3	82.3	82.5	1511.5
FS-LNnHI	756.8	7.2	78.8	75.0	1511.5
FS-LNnHI	756.8	6.0	77.9	74.7	1511.5
FS-LNnHI	756.8	5.6	76.0	73.4	1511.5
FS-LNH I	756.8	6.3	68.4	64.0	1511.5
FS-LNH I	756.8	5.6	66.7	59.5	1511.5
412.1a	756.8	6.3	66.3	60.9	1511.5
412.1a	756.8	6.0	61.8	50.7	1511.5

**Table 4c**

Name	Precursor	CE	Match Score	Reverse Score	Mass (Library)
4121a	756.8	7.2	61.6	51.3	1511.5

**Table 4d**

Name	Precursor	CE	Match Score	Reverse Score	Mass (Library)
FS-LNnHI	756.8	6.3	73.5	83.8	1511.5
FS-LNnHI	756.8	7.2	70.4	75.6	1511.5
FS-LNH I	756.8	6.3	69.2	76.4	1511.5
FS-LNnHI	756.8	9.2	69.1	73.8	1511.5
FS-LNnHI	756.8	6.0	68.1	74.0	1511.5
FS-LNnHI	756.8	5.6	66.9	73.8	1511.5
FS-LNH I	756.8	5.6	65.3	68.1	1511.5

**Table 4e**

Name	Precursor	CE	Match Score	Reverse Score	Mass (Library)
FS-LNnHI	756.8	9.2	72.6	74.4	1511.5
FS-LNnHI	756.8	7.2	69.5	71.8	1511.5