Annotation and Structural Analysis of Sialylated Human Milk Oligosaccharides

Shuai Wu, Rudolf Grimm, J. Bruce German, and Carlito B. Lebrilla

Department of Chemistry, University of California, Davis, California 95616, United States, Department of Biochemistry and Molecular Medicine, School of Medicine, University of California, Davis, California 95616, United States, Agilent Technologies Inc., Santa Clara, California 95051, United States, and Robert Mondavi Institute for Wine and Department of Food Science and Technology, University of California, Davis, California 95616, United States

Received October 4, 2010

Sialylated human milk oligosaccharides (SHMOs) are important components of human milk oligosaccharides. Sialic acids are typically found on the nonreducing end and are known binding sites for pathogens and aid in neonates’ brain development. Due to their negative charge and hydrophilic nature, they also help modulate cell–cell interactions. It has also been shown that sialic acids are involved in regulating the immune response and aid in brain development. In this study, the enriched SHMOs from pooled milk sample were analyzed by HPLC-Chip/QTOF MS. The instrument employs a microchip-based nano-LC column packed with porous graphitized carbon (PGC) to provide excellent isomer separation for SHMOs with highly reproducible retention time. The precursor ions were further examined with collision-induced dissociation (CID). By applying the proper collision energy, isomers can be readily differentiated by diagnostic peaks and characteristic fragmentation patterns. A set of 30 SHMO structures with retention times, accurate masses, and MS/MS spectra was deduced and incorporated into an HMO library. When combined with previously determined neutral components, a library with over 70 structures is obtained allowing high-throughput oligosaccharide structure identification.

Keywords: sialylated human milk oligosaccharides • HPLC-Chip/QTOF MS • CID • structure identification

Introduction

Human milk oligosaccharides (HMOs) have been known to have many biological functions including as prebiotics for stimulating the growth of beneficial intestinal bacteria, as receptor analogs to inhibit the binding of pathogens, and as components involved in modulating the immune system. Sialic acids are important components of HMOs and are often found on the nonreducing termini of HMOs. Sialylated HMOs (SHMOs) constitute about 20% of all HMOs. They serve as binding sites for specific pathogens and toxins. For example, in vitro studies have shown that SHMOs significantly decrease the binding of leukocyte to endothelial cells while the neutral HMOs had no effect. Sialic acids, due to their negative charge and hydrophilic nature, help modulate the cell–cell interaction. It is also believed that sialic acids serve as ligands for lectin binding involved in regulating the immune response. In addition, the brain is the organ with the highest level of sialic acids where it plays an important role in facilitating neuronal sprouting and plasticity. SHMOs are therefore believed to play a key role in postnatal brain development.

Mass spectrometry provides the most sensitive and rapid method for characterizing SHMOs. Structure elucidation of SHMOs is generally more difficult than neutrals because sialic acids readily dissociate under many ionization conditions. A labile glycosidic bond is caused by the adjacent proton from the carboxylic acid group. In addition, the signal of SHMOs can be strongly suppressed by neutral HMOs. Derivatization of the acid group with methyl or other groups tended to stabilize the NeuAc moiety and enhance the sensitivity while decreasing fragmentation. The use of high performance liquid chromatography (HPLC) coupled to mass spectrometry (MS) eliminates many of the difficulties in analyzing sialylated oligosaccharides (OS). The separation of sialylated OS from the neutral minimizes the ion suppression. Furthermore, isomer separation with HPLC in conjunction with electrospray ionization, a soft ionization source, enhances the analysis of SHMO. Reversed phase column was studied for separating the permethylated OS. Normal phase chromatography was also used for separating derivatized or underivatized OS. High pH anion-exchange chromatography (HPAEC) is also an option and provides adequate separation of OS, although it is generally not amenable for coupling to mass spectrometry (MS). However, we generally find porous graphitized carbon (PGC) to be the best stationary phase for separating native OS
isomers.\textsuperscript{6,36–40} LC–MS provides composition and structural information during OS profiling\textsuperscript{6,41} and LC–MS/MS or tandem MS provides efficient differentiation of isomeric species.\textsuperscript{26,38,42,43} Fragmentation methods for analyzing sialylated OS currently include collision induced dissociation (CID),\textsuperscript{29,44–47} infrared multiphoton dissociation (IRMPD),\textsuperscript{28,48,49} ultraviolet photodissociation (UVPD),\textsuperscript{50} electron capture dissociation (ECD),\textsuperscript{51} electron detachment dissociation (EDD)\textsuperscript{52} and high energy CID in tandem time-of-flight (TOF/TOF) instruments.\textsuperscript{23,53,54}

This laboratory is in the process of creating an annotated library of HMOs for the rapid identification of HMO structures. We have already reported a set of neutral HMOs.\textsuperscript{36} In this report, we catalog the sialylated oligosaccharides. HPLC-Chip/TOF MS instrument is used for analyzing enriched SHMO samples. The instrument employs microchip based nano-LC column packed with PGC that provides excellent isomer separation for SHMOs with highly reproducible retention time (RT). The precursor ions were further fragmented with CID. By applying proper collision energy (CE), isomers can be readily differentiated by diagnostic peaks and characteristic fragmentation patterns.

**Experimental Section**

**Reagents and Materials.** The OS used in this study are from pooled human milk provided by milk banks in San Jose, CA and Austin, TX. HMOs were isolated from the milk using a previously described procedure involving defatting, chloroform/methanol extraction, ethanol precipitation, and evaporation.\textsuperscript{6} A sample enrichment step with solid phase extraction (SPE) employing graphitized carbon cartridge (GCC) was used before the analysis. GCC (150 mg bed weight, 4 mL volume) were purchased from Alltech (Deerfield, IL). Sodium borohydride (90%) and 2, 5-dihydroxybenzoic acid (DHB) were obtained from Sigma-Aldrich (St. Louis, MO). Standard HMOs were purchased from Dextra Laboratories (Earley Gate, UK). \(\alpha(1\rightarrow2)\)-Fucosidase was obtained from EMD CALBIOCHEM (La Jolla, CA), \(\beta(1\rightarrow3)\)-galactosidase and \(\alpha(2\rightarrow3)\)-neuraminidase from New England Biolab (Beverly, MA), \(\beta(1\rightarrow4)\)-galactosidase from ProZyme (San Leandro, CA), and \(\alpha(1\rightarrow3,4)\)-fucosidase from Sigma-Aldrich (St. Louis, MO). The nonselective sialidase was purified and provided by Prof. David Mills from the Department of Viticulture and Enology in UC Davis. All reagents are of analytical or HPLC grade.

**Oligosaccharide Reduction and Enrichment.** The pooled HMO sample (50 mg in 250 \(\mu\)L nanopure water) was reduced by 250 \(\mu\)L of 1.0 M sodium borohydride in a water bath at 65 °C for 1.5 h. The resulting product was desalted and purified by SPE-GCC. The SPE cartridge was first conditioned by 6 mL of 80% acetonitrile (ACN) with 0.1% trifluoroacetic acid (TFA, v/v) and then 6 \(\mu\)L nanopure water. The desalting was performed by loading 1-mg sample to each cartridge and washed with 30 \(\mu\)L nanopure water. The HMOs were eluted with 6 \(\mu\)L 5% ACN in water (v/v), 6 \(\mu\)L 10% ACN in water (v/v), 6 \(\mu\)L 20% ACN in water (v/v) and 6 \(\mu\)L 40% ACN with 0.05% TFA. SHMOs were enriched by collecting the 40% eluent only. The sample was dried in vacuo and reconstituted with nanopure water before MS analysis.

**Separation of SHMOs by HPLC.** The enriched SHMOs were separated on an Agilent 1100 series HPLC instrument with Hypercarb PGC column (100 mm \(\times\) 3.0 mm, 5 \(\mu\)m particle size) and detected with photodiode array detector at 206 and 254 nm. A 20 \(\mu\)L sample (about 0.05 mg/\(\mu\)L) was injected and eluted by 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 at the flow rate 0.35 mL/min and with a gradient of 0.0–1.0 min, 0% B; 1.1–6.0 min, 8% B; 6.0–60 min, 8–18% B; 60–70 min, 18–100% B; 70–80 min, 100% B. Sialylated HMOs were collected into 80 fractions with one minute per tube. The samples were dried in vacuo and reconstituted with 25 \(\mu\)L water before MALDI MS analysis.

**MALDI FTICR MS.** The HiRes MALDI FTICR (IonSpec, Irvine, CA) is composed of an external MALDI source with a pulsed 355 nm Nd:YAG laser, a hexapole ion guide and a FTICR cell with a 7.0 T shielded superconducting magnet. 2, 5-Dihydroxybenzoic acid (DHB) was used as matrix (8 mg/160 \(\mu\)L in 50% ACN/water (v/v)) in both positive and negative mode. The HMOs (0.5 \(\mu\)L) were spotted on a 100-sample stainless steel plate with 0.5 \(\mu\)L matrix solution. In the positive mode, 0.25 \(\mu\)L, 0.01 \(\mu\)M NaCl solution was added as a cation dopant, while in the negative mode, no NaCl solution was added. The sample plate was dried in the vacuum chamber before MS analysis.

**HPLC-Chip/TOF MS Analysis.** The SHMOs samples were analyzed using the Agilent 6200 HPLC-Chip/TOF MS instrument (Agilent Technologies, Santa Clara, CA) with Agilent 1200 nano-LC and 6210 TOF MS. A detailed procedure was described in our previous paper.\textsuperscript{36} The nano-LC was equipped with a capillary pump as the loading pump for sample enrichment, a nanopump as the analytical pump for sample separation, a microwell-plate autosampler maintained at 6 °C by a thermostat, an Agilent HPLC-Chip cube interface and Agilent 6210 TOF MS. The micro-Chip consisted of an enrichment column with a volume of 40 nL and an analytical column 43 \(\times\) 0.075 mm i.d., which were both packed with PGC having 5 \(\mu\)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 \(\mu\)L/min flow rate of solvent A was used for sample loading with 1 \(\mu\)L injection volume. A 45 min gradient delivered by a nanoflow pump with a flow rate of 0.3 \(\mu\)L/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 min equilibration time at 0% B.

Data analysis was performed with the Agilent MassHunter Qualitative Analysis software. The deconvoluted data file was calculated by an in-house software “Oligosaccharide Calculator,” written in Igor Pro (Wavemetrics, Inc.). The output file included measured mass, calculated mass with mass error, and composition of each oligosaccharide sorted based on retention times and abundances (Supplementary Table 1, Supporting Information).

Twelve commercial OS standards were reduced and introduced into Chip/TOF under the identical condition in order to match the retention times and accurate masses with the corresponding OS in SHMO sample.\textsuperscript{36}

**HPLC-Chip/TOFMS Analysis.** The Agilent HPLC-Chip/TOF MS instrument (Agilent Technologies, Palo Alto, CA) is equipped with an Agilent 1200 series nano-LC system and Agilent 6520 QTOF coupled with a chip cube interface. The nano-LC system and the PGC chip used has been described above. Using the same injection volume, binary solvent and gradient, the resulting LC chromatograms in the Chip/QTOF were nearly identical to those of the Chip/TOF with minor variations of retention times. For both instruments, the same internal calibration technique was used to yield <5 ppm mass accuracy for MS and <20 ppm for MS/MS experiments.

In the positive ion mode, the data stored during the QTOF run included both centroid and profile. The instrument settings
were adjusted during autotune and set to: fragmentor voltage 175 V, skimmer 65 V, and the octopole 1 RF voltage 750 V. The drying gas was heated at 325 °C with a flow rate of 5 L/min. The data acquisition was set to auto MS/MS with 2 spectra/s for MS scan and 1 spectra/s for MS/MS scan. The precursor ion was selected based on abundances with doubly charged ion being given the first priority followed by singly charged ion, triply charged, and then other multiply charged ions. A 4-m/z isolation window was used. All the calibrant ions were excluded from the isolation. The collision energy applied was based on the mass-to-charge (m/z) ratio of the ion with higher energy for larger ions. For this instrument, the collision energy was varied according to the equation:

\[ CE(V) = \frac{(m/z)}{100(Da)} \times 1.3 - 3.5 \]

where 1.3 is the slope and −3.5 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, CE is the collision energy. The equation was empirically determined by the manufacturer. In this study, the value of the slope and y-intercept were optimized to yield fragmentation that produced the most distinguishing features for each HMO isomer. The optimal collision energy for singly- and doubly charged ions are expectedly different. For example, the collision energy applied to singly charged ion of FS-LNH isomers (m/z 1512.6) is 16.2 V, while 6.3 V is used for the doubly charged ion (m/z 756.8(+2)).

**Exoglycosidase Digestion.** Reaction buffer solutions were prepared with glacial acetic acid and 0.1 M ammonium acetate solution to the desired pH.55,56 Enzymes were used without further purification. In a 0.2 mL PCR tube, 3 μL buffer solution was added followed by 1 μL OS sample and 1 μL enzyme solution (mole ratio of protein to OS is about 1:100–200). The reaction mixture was incubated at 37 °C in a water bath. The reaction time used for each enzyme is tabulated in Supplementary Table 2 (Supporting Information).

**Results**

**Enrichment and MALDI MS Profile of Sialylated OS from Human Milk.** Sialylated OS make up about 20% of the total OS in human milk.6,7 To enrich the sialylated species, HMO mixtures where fractionated by SPE - GCC. Sialylated OS tend to elute with higher ratio of organic solvent, specifically 40% aqueous acetonitrile. Negative MALDI profile of the enriched mixture showed deprotonated ions and their monosaccharide compositions (Supplementary Figure 1, Supporting Information). The compositions listed in the spectrum were assigned using an in-house software “Oligosaccharide Calculator” written in Igor (based on a mass accuracy <5 ppm). The information provided by accurate mass showed that most of the SHMOs are monosialylated while only a small percentage is disialylated. The majority of the OS are also fucosylated with the number of fucoses (Fuc) ranging from one to four.

**Chip/TOF MS Profile of Sialylated Components.** Chip/TOF MS provides the online LC separation and MS detection with high mass accuracy. The PGC stationary phase is packed in a nanoflow column incorporated into a microchip with the electrospray tip integrated to minimize dead volume. The chip has been shown to yield excellent separation of OS isomers with highly reproducible retention times.5,36 Since the separation is highly effective, α- and β- anomers at the reducing end have distinct retention times. To eliminate this complexity, the SHMOs were reduced to the alditol.6,36

Figure 1a is the base peak chromatogram (BPC) of the enriched SHMOs from a sample of pooled milk. Supplementary Table 1 (Supporting Information) shows the 70 SHMOs found along with their compositions and retention times. Most of the SHMOs are monosialylated and eluted after 20 min. SHMOs are generally eluted later than the neutral HMO. Isomers are effectively separated as illustrated in Figure 1b, which shows the extracted ion chromatogram (EIC) for the neutral mass 1511.6 (m/z 756.8 (+2), MS inset). For the convenience of this discussion, doubly charged ions will be marked as +2 after the m/z, while the charge state for singly charged ions will not be annotated. Supplementary Table 1 (Supporting Information) contains six different species (# 18 to 23) found by Chip/TOF with their monosaccharide compositions. Also included are the respective abundances sorted by RT. The neutral mass 1511.6 corresponds to one Fuc and one NeuAc on a lacto-N-hexaose core (FS-LNH). The composition belongs to seven isomers that are also the most abundant components (Table 1).

The detailed procedure for elucidating their structures follows. To simplify the analysis, the mixture was separated using a standard HPLC into smaller pools of glycans, which were further analyzed individually. The HPLC column (100.0 mm × 3.0 mm) employed an 80-min gradient (experimental section) allowing separation of components that were not resolved in the conditions for the Chip/TOF MS. Each HPLC fraction was analyzed further to perform the complete structural elucidation.

Figure 2 is the EIC for FS-LNH isomers found in several HPLC fractions (Figure 2a–g). For reference, Figure 1b is duplicated in the first panel of Figure 2. The figure illustrates the general approach for elucidating the glycome as well as the efficacy of the methods for separation and structural elucidation. Figures 2a–g show each component as separated by standard HPLC with the HPLC fractions examined by Chip/TOF MS. The inset structures were determined by one of the two methods. (1) For compounds where standards were available retention times and tandem MS were compared with the standards. (2) For the remainder, which was the vast majority, structures were determined by a combination of tandem MS and targeted exoglycosidase digestion. From the group in Figure 2a–g, two compounds were determined by standards obtained commercially, FS-LNH (Figure 2d) and FS-LNh I (Figure 2f). Among all sialylated OS, 12 standards were identified in this manner. Two new structures were obtained in this group, 4121a (Figure 2a) and 4121b (Figure 2b). The nomenclature is based on the number of hexose (Hex), Fuc, GlcNac, and NeuAc, respectively. Three structures were elucidated and found to correspond to published structures (Figure 2c, e and g) with the names previously assigned.

With PGC as stationary phase, the chip based nano-LC column (43 mm) is generally sufficient to separate isomers using the 45 min gradient. However, under the chosen conditions some isomers such as 4121b and FS-LNH III have similar retention times with 23.9 and 24.0 min, respectively. In this situation, the off-line HPLC column (100 mm) with an 80-min gradient allowed separation of the isomers into different fractions. The isomers in the fractions were injected into Chip/TOF, which showed two distinct structures based on tandem MS.
Tandem MS Analysis of Sialylated OS. Figure 3a and b are MS/MS spectra of a commercial standard FS-LNH. The singly charged ion \( m/z 1512.6 \) yields the tandem MS spectrum in Figure 3a, The doubly charged ion \( m/z 756.8 \) yields the spectrum in Figure 3b. As seen clearly, the doubly charged species generally provides more structurally informative fragment ions. The following discussions of all the FS-LNH isomers will therefore focus on the MS/MS of the doubly charged ions.

Fragmentation for this ion starts from both nonreducing termini when the HMO has two antennae at the lactose core. In Figure 3b, \( m/z 756.8 \) (+2) loses a Fuc to generate \( m/z 683.8 \) (+2) and a galactose to generate \( m/z 602.7 \) (+2). The subsequent loss of GlcNAc generates \( m/z 1001.4 \). This product ion further loses the lactose core to generate \( m/z 657.2 \). An alternative fragmentation pathway starts with the loss of NeuAc to generate \( m/z 1221.5 \), which is weakly abundant. The further loss of a lactosamine unit [Gal + GlcNAc] from \( m/z 1221.5 \) generates \( m/z 856.3 \). The following loss of the lactose core generates \( m/z 512.2 \). Figure 3c is the tandem spectrum of FS-LNH in HPLC fraction 36 from the milk sample. The FS-LNH from two different sources, with different amounts, yielded nearly identical spectra. Figure 3b and c show that even though the total ion intensities (or concentrations) for the two samples are different, the overall features of the tandem MS are nearly identical. In general, when the same collision energy is applied, the resulting tandem MS spectra are highly reproducible.

Structure Elucidation. Elucidation of individual structures begins with the tandem MS. Figure 4 shows the MS/MS spectra for three of FS-LNH isomers, a new glycan 4121a, and two structures that were elucidated and found to correspond to previously published ones, FS-LNH I and FS-LNH II. In Figure 4a, the precursor ion of 4121a \( m/z 756.8 \) (+2) loses lactose to yield \( m/z 1168.4 \). The fragment ion \( m/z 1168.4 \) readily loses a Fuc to generate \( m/z 1022.4 \). In Figure 4b and c, the precursor ions of the two other isomers did not yield the analogous fragment ions. Based on this information, it is believed that the new compound 4121a has a linear core structure while the other two both have branched cores, which cannot lose lactose from the quasimolecular ion. In contrast, both structures FS-LNH I and FS-LNH II loses Fuc and a Gal to generate \( m/z 602.7 \) (+2) (weakly abundant in Figure 4c) from the precursor ion, consistent with the presence of Lewis a and x epitopes at one terminus of each compound. The loss of these residues further confirms the position of NeuAc as being on the other terminus.

Table 1. FS-LNH Isomers Detected by Chip/TOF from Enriched SHMOs

<table>
<thead>
<tr>
<th>SHMOs</th>
<th>mass (exptl)</th>
<th>mass (calcd)</th>
<th>error (Da)</th>
<th>Hex</th>
<th>Fuc</th>
<th>GlcNAc</th>
<th>NeuAc</th>
<th>RT (min)</th>
<th>abundance*</th>
</tr>
</thead>
<tbody>
<tr>
<td>4121a</td>
<td>1511.547</td>
<td>1511.550</td>
<td>-0.003</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>23.6</td>
<td>255,511</td>
</tr>
<tr>
<td>4121b</td>
<td>1511.553</td>
<td>1511.550</td>
<td>0.003</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>23.9</td>
<td>226,932</td>
</tr>
<tr>
<td>FS-LNH III</td>
<td>1511.554</td>
<td>1511.550</td>
<td>0.003</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>24.8</td>
<td>771,332</td>
</tr>
<tr>
<td>FS-LNH</td>
<td>1511.554</td>
<td>1511.550</td>
<td>0.002</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>26.2</td>
<td>286,110</td>
</tr>
<tr>
<td>FS-LNH I</td>
<td>1511.554</td>
<td>1511.550</td>
<td>0.004</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>26.6</td>
<td>2,883,226</td>
</tr>
<tr>
<td>FS-LNH II</td>
<td>1511.550</td>
<td>1511.550</td>
<td>0.000</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>29.2</td>
<td>206,722</td>
</tr>
</tbody>
</table>

* The abundance is from HPLC-Chip/TOF counts per second (cps).
The same fragment ion is not observed in Figure 4a since 4121a has a linear core and NeuAc at the nonreducing end must fragment before Gal. The differences between FS-LNH I and FS-LNH II are readily determined by several other diagnostic peaks. The fragment peak \( m/z 495.2 \) is only found in FS-LNH I, and in none of the other isomers, because sialylation is directly on the GlcNAc (Figure 4b). The Y type ion\(^{57} m/z 1059.4 \) in FS-LNH II is generated by losing NeuAc (\( m/z 1221.5 \)) and then Gal. The corresponding B ion \( m/z 454.2 \) also indicates the differences in connectivity of NeuAc in both FS-LNH II and FS-LNH I (Figure 4c).

Two isomers with similar retention times, the new OS 4121b and FS-LNH III are readily distinguished based on their tandem MS (Figure 5). The linear compound 4121b yields the fragment ion \( m/z 1168.4 \) (and \( m/z 584.7 (+2) \)) due to the loss of the lactose. The subsequent loss of a Fuc produces \( m/z 1022.4 \) (and \( m/z 511.7 (+2) \)) (Figure 5a). An alternative pathway further yields the intense peak \( m/z 1022.4 \) (and \( m/z 511.7 (+2) \)) due to the loss of the reducing end glucose (Glc) from the quasimolecular ion - also indicating that the Fuc is not on the reducing end. The lack of the diagnostic ion \( m/z 495.2 \) (NeuAc-GlcNAc), as shown with FS-LNH I, is indicative that the NeuAc is attached to Gal.

The fragmentation behavior of FS-LNH III also follows several pathways due to its branched core structure (Figure 5b). One of the fragmentation pathways begins with the loss of NeuAc to generate \( m/z 1221.5 \). The subsequent loss of a lactosamine unit yielded \( m/z 856.3 \). The fragment ion \( m/z 856.3 \) further loses lactose to produce \( m/z 512.2 \). An alternate pathway starting with the loss of a Fuc yielded \( m/z 683.7 (+2) \), followed by a Gal loss to form the product ion \( m/z 602.7 (+2) \). The subsequent fragmentation generates \( m/z 1001.4 \) by losing a GalNAc and lactose loss to yield \( m/z 657.2 \). The complemen-
Tertiary information from these two fragmentation pathways also confirm that NeuAc and Fuc are on different branches. However, the generation of $m/z$ 675.8 (+2) and $m/z$ 1147.4 is not consistent with the structure. These ions are the results of...
the rearrangement of the Fuc within the protonated species. The exact linkages between the monosaccharides were confirmed with exoglycosidase digestion.

**Linkage Elucidation with Exoglycosidase Digestion.** The detailed procedure for performing the exoglycosidase digestions is described in previous publications. The conditions in this study for the specific enzyme have been modified and summarized in Supplementary Table 2 (Supporting Information). For convenience the reaction was monitored by MALDI MS. In so doing, the full structure is elucidated by combining MS, tandem MS and the exoglycosidase digestion.

Compounds where no standard is available, the vast majority of the structures, need to be fully elucidated even if the structures have been previously published. The digestion of FS-LNH III (in HPLC fraction 35) is described to provide a more complicated example than 4121b, as the linkages need to be confirmed for both antennae. The results of the digestion for fraction 35 are summarized in Figure 6. Figure 6a is the MALDI MS of HPLC fraction 35 in the positive ion mode before enzyme digestion. As can be seen, the spectrum of FS-LNH III yielded the quasimolecular ion $m/z$ 1556.5 $[M+2Na-H]^+$ and a fragment ion, $m/z$ 1243.4 from the loss of NeuAc. The incubation with $\alpha(2\rightarrow3)$-neuraminidase for 1 h showed that the presence of $a_{2\rightarrow6}$ linkage of NeuAc in this isomer because $m/z$ 1556.5 was not digested (Figure 6b). Figure 6c is the result from the reaction with a nonspecific sialidase. After cleaving off the NeuAc from the Gal, a second digestion step with $\beta(1\rightarrow4)$-galactosidase generated a new peak, $m/z$ 1081.4, that confirmed the $\beta1\rightarrow4$ linkage of Gal bound to GlcNAc (Figure 6e). The result from this step also indicated that Fuc is present on the other branch. The digestion from the other branch started with $\alpha(1\rightarrow3,4)$-fucosidase that generated a new $[M+2Na-H]^+$ ion with $m/z$ 1410.5 (Figure 6d). The result from this step showed that Fuc has $\alpha1\rightarrow3,4$ linkage attached to the GlcNAc and not the Gal. In addition, the disappearance of $m/z$ 1243.4 in Figure 6d confirmed that it was a fragment peak from $m/z$ 1556.5 in Figure 6a. The further digestion of $m/z$ 1410.5 by $\beta(1\rightarrow3)$-galactosidase generated a new peak $m/z$ 1248.4, which is still a sialylated species (Figure 6f). The result proved a Lewis a epitope on the fucosylated antenna. This method was applied to all the other structures.

**Negative Mode MS/MS Analysis.** In some occasions, it has been necessary to supplement the positive ion MS/MS with negative ion analysis. In the positive ion mode, the fragmentation of the collision activated ions generated primarily $B, Y$ type ions. In certain cases, such as FS-LNH and FS-LNnH I, MS/MS yielded the same fragments albeit with great variation in intensities. In this way, tandem MS in the positive mode can be used to identify specific oligosaccharides but not elucidating similar structures (Supplementary Figure 2, Supporting Information). However, the fragmentation in the negative ion mode using the Chip/QTOF MS was useful for producing different pathways. In the negative ion mode, our solvent system generated primarily deprotonated ions and formylated ions. For all FS-LNH isomers, $m/z$ 754.8($-2$) $[M-2H]^+$ and $m/z$ 777.8($-2$) $[M-H+HCOO]^+$ were both seen in the negative ion mode. The precursor ion $m/z$ 777.8($-2$) readily loses formic acid (HCOOH) to generate $m/z$ 754.8($-2$) and subsequent
fragment ions (Figure 7). In Figure 7a, $m/z$ 754.8(-2) first loses a Fuc to generate a $Z_{4a}$ ion with $m/z$ 1346.5. The subsequent loss of Gal yields $Z_{3a}$ with $m/z$ 1184.4. While in Figure 7b, $m/z$ 754.8(-2) loses a Fuc to produce $Z_{3b}$ then a following Gal loss yielded $Z_{3a}$'. Since Fuc and Gal both link to GalNAc, the Gal loss may occur before Fuc loss to generate the $Z_{3a}$', which is not found in Figure 7a since Fuc links to Gal at the reducing end in FS-LNH. In this case, the fragmentation in the negative ion mode provides a better way to resolve the connectivity between the two isomers. The structure information from the negative mode MS/MS was therefore, at times, a necessary compliment to the positive ion mode.

**Discussion**

Thirty SHMO structures with retention time and mass are listed in Table 2. This table along with the tandem MS is valuable for rapidly identifying known structures. Comparison of the retention time and the tandem MS provides a highly specific method for identifying oligosaccharides. When combined with previously determined neutral HMOs, a library with over 70 structures is constructed which incorporates nano-LC retention time, accurate mass, and MS/MS spectra.

On the basis of the structures, we find a few general observations regarding sialylation in HMOs. These observations are based on a pool of samples from five mothers.

1. The majority of the SHMOs are monosialylated (90.9%) with the smaller fraction being disialylated (9.1%) (Supplementary Figure 3, Supporting Information).
2. The more abundant linkage corresponds to $\alpha_{2-6}$ (56.9%), while only 4.8% SHMOs were found with $\alpha_{2-3}$ NeuAc so far (Supplementary Figure 3, Supporting Information).
3. The $\alpha_{2-3}$ NeuAc is usually bound to a $\beta_1-3$ galactose, except for sialyl lactose.
4. The $\alpha_{2-6}$ NeuAc is usually bound to $\beta_1-4$ galactose.
5. Sialylation on the GlcNAc is only found as $\alpha_{2-6}$ linkage as previously reported. However, this arrangement is even more specifically found only on the $\beta_1-3$ branch from the core lactose.

The method for elucidating structure employing tandem MS and exoglycosidase is generally successful with a caveat. Examination of the different fragmentation pathways from various protonated isomers showed rearrangements that can affect the interpretation. For example, the compound FS-LNH generates fragment ion $m/z$ 675.8(+2) by losing a Gal, which is not consistent with the structure (Figure 3b and c). This rearrangement is caused by the long-range migration of the Fuc, most likely to a hydroxyl group of the reducing end residue. Similar rearrangements are not observed for sodiated ions in MALDI FTICR when performing CID or IRMPD experiments as noted previously. To confirm the connectivity and linkages of monosaccharides, exoglycosidase digestion was necessary to validate the structural information obtained from tandem MS.

The future work will be inputting all the retention times and MS/MS spectra with all the elucidated structures into Agilent Masshunter software. Auto MS/MS search will be performed.
<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Mass (exptl)</th>
<th>Composition</th>
<th>TOF RT (min)</th>
<th>QTOF RT (min)</th>
<th>Structure</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5'SL</td>
<td>635.229</td>
<td>2001</td>
<td>13.6</td>
<td>14.6</td>
<td>α 2-6</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>3'SL</td>
<td>635.227</td>
<td>2001</td>
<td>22.9</td>
<td>22.6</td>
<td>α 2-3</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>6'SLN</td>
<td>676.254</td>
<td>1011</td>
<td>13.1</td>
<td>14.5</td>
<td>α 2-6</td>
<td>61</td>
</tr>
<tr>
<td>4</td>
<td>3'SLN</td>
<td>676.252</td>
<td>1011</td>
<td>23.1</td>
<td>23.7</td>
<td>α 2-3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3'Sle*</td>
<td>822.314</td>
<td>1111</td>
<td>14.6</td>
<td>12.9</td>
<td>α 1-4</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>LSTc</td>
<td>1000.365</td>
<td>3011</td>
<td>24.5</td>
<td>23.8</td>
<td>α 2-6</td>
<td>62</td>
</tr>
<tr>
<td>7</td>
<td>LSTb</td>
<td>1000.365</td>
<td>3011</td>
<td>25.5</td>
<td>24.7</td>
<td>α 2-6</td>
<td>62</td>
</tr>
<tr>
<td>8</td>
<td>LSTA</td>
<td>1000.362</td>
<td>3011</td>
<td>28.1</td>
<td>26.8</td>
<td>α 2-3</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>F-LSTc</td>
<td>1146.419</td>
<td>3111</td>
<td>24.3</td>
<td>23.5</td>
<td>α 2-6</td>
<td>63</td>
</tr>
<tr>
<td>10</td>
<td>DSLNT</td>
<td>1291.446</td>
<td>3012</td>
<td>31.2</td>
<td>33.7</td>
<td>α 2-3</td>
<td>64</td>
</tr>
<tr>
<td>11</td>
<td>S-LNH</td>
<td>1365.495</td>
<td>4021</td>
<td>26.2</td>
<td>25.3</td>
<td>α 2-6</td>
<td>65</td>
</tr>
<tr>
<td>12</td>
<td><strong>4021a</strong></td>
<td>1365.494</td>
<td>4021</td>
<td>27.1</td>
<td>26.1</td>
<td>α 2-6</td>
<td>66</td>
</tr>
<tr>
<td>13</td>
<td>S-LNnH II</td>
<td>1365.494</td>
<td>4021</td>
<td>27.3</td>
<td>26.3</td>
<td>α 2-6</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Continued

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Mass (exptl)</th>
<th>Composition</th>
<th>TOF RT (min)</th>
<th>QTOF RT (min)</th>
<th>Structure</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>4021b</td>
<td>1365.491</td>
<td>4021</td>
<td>29.0</td>
<td>27.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>4121a</td>
<td>1511.547</td>
<td>4121</td>
<td>23.6</td>
<td>22.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>4121b</td>
<td>1511.553</td>
<td>4121</td>
<td>23.9</td>
<td>23.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>FS-LNH III</td>
<td>1511.553</td>
<td>4121</td>
<td>24.0</td>
<td>23.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>FS-LNH</td>
<td>1511.554</td>
<td>4121</td>
<td>24.8</td>
<td>24.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>FS-LNH I</td>
<td>1511.552</td>
<td>4121</td>
<td>26.2</td>
<td>25.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>FS-LNh I</td>
<td>1511.554</td>
<td>4121</td>
<td>26.6</td>
<td>25.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>FS-LNH II</td>
<td>1511.550</td>
<td>4121</td>
<td>29.2</td>
<td>27.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>DFS-LNH</td>
<td>1657.612</td>
<td>4221</td>
<td>24.6</td>
<td>23.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>DFS-LNh</td>
<td>1657.608</td>
<td>4221</td>
<td>28.6</td>
<td>27.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>5031a</td>
<td>1730.622</td>
<td>5031</td>
<td>29.1</td>
<td>27.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Continued

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Mass (exptl)</th>
<th>Composition</th>
<th>TOF RT (min)</th>
<th>QTOF RT (min)</th>
<th>Structure</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>FS-LNO</td>
<td>1875.682</td>
<td>5131</td>
<td>27.2</td>
<td>26.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>5131a</td>
<td>1875.682</td>
<td>5131</td>
<td>27.9</td>
<td>26.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>5231a</td>
<td>2022.741</td>
<td>5231</td>
<td>26.5</td>
<td>25.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>5231b</td>
<td>2022.739</td>
<td>5231</td>
<td>27.3</td>
<td>26.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>6041a</td>
<td>2096.757</td>
<td>6041</td>
<td>29.3</td>
<td>27.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>5331a</td>
<td>2168.803</td>
<td>5331</td>
<td>26.1</td>
<td>25.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

New structures are indicated in red, for example 4021a. Monosaccharide composition 4Hex:0Fuc:2HexNAc:tNeuAc represented as 4021 with a, b, c... as the order of OS from LC.
by the software for the unknown HMO samples and compared with the spectra in the database. An output of the compound list will include all the identified structures with the matching scores. This library will eventually provide a high-throughput method for oligosaccharide analysis.

Acknowledgment. Funding provided by the National Institutes of Health (HD059127 and HD061923), UC Discovery, and the California Dairy Research Fund is gratefully acknowledged.

Supporting Information Available: MALDI FTICR-MS spectrum of enriched SHMOs in the negative ion mode with the list of monosaccharide compositions. MS/MS spectra of FS-LNHy and FS-LNHy I in the positive ion mode. Quantitative analysis of sialylated HMOs based on the abundances from HPLC Chip/TOF MS. The deconvoluted data of 70 SHMOs analysis of sialylated HMOs based on the abundances from FS-LNH and FS-LNnH I in the positive ion mode. Quantitative spectrum of enriched SHMOs in the negative ion mode with the software for the unknown HMO samples and compared to the library. This library will eventually provide a high-throughput method for oligosaccharide analysis.

References
research articles


PR101006U