analytical chemistry

Article

Cite This: *Anal. Chem.* 2018, 90, 8977−8988

pubs.acs.org/ac

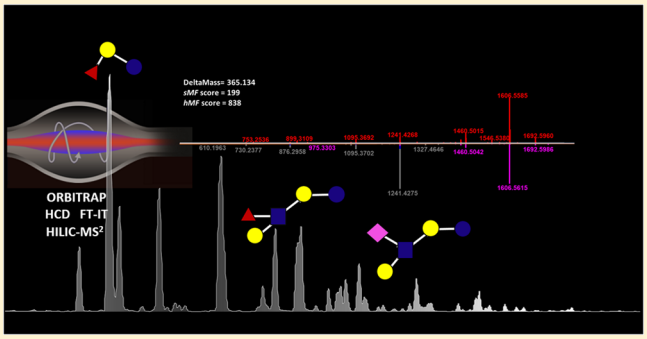
Creating a Mass Spectral Reference Library for Oligosaccharides in Human Milk

Connie A. Remoroza,* Tytus D. Mak, Maria Lorna A. De Leoz, Yuri A. Mirokhin, and Stephen E. Stein

Mass Spectrometry Data Center, Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8362, United States

Supporting Information

ABSTRACT: We report the development and availability of a mass spectral reference library for oligosaccharides in human milk. This represents a new variety of spectral library that includes consensus spectra of compounds annotated through various data analysis methods, a concept that can be extended to other varieties of biological fluids. Oligosaccharides from the NIST Standard Reference Material (SRM) 1953, composed of human milk pooled from 100 breastfeeding mothers, were identified and characterized using hydrophilic interaction liquid chromatography electrospray ionization tandem mass spectrometry (HILIC-ESI-MS/MS) and the NIST 17 Tandem MS Library. Consensus reference spectra were generated, incorporated into a searchable library, and matched using the newly developed hybrid search algorithm to elucidate unknown oligosaccharides. The NIST hybrid search program facilitates the structural assignment of complex oligosaccharides especially when reference standards are not commercially available. High accuracy mass measurement for precursor and product ions, as well as the relatively high MS/MS signal intensities of various oligosaccharide precursors with Fourier transform ion trap (FT-IT) and higher energy dissociation (HCD) fragmentation techniques, enabled the assignment of multiple free and underivatized fucosyllacto- and sialyllacto-oligosaccharide spectra. Neutral and sialylated isomeric oligosaccharides have distinct retention times, allowing the identification of 74 oligosaccharides in the reference material. This collection of newly characterized spectra based on a searchable, reference MS library of annotated oligosaccharides can be applied to analyze similar compounds in other types of milk or any biological fluid containing milk oligosaccharides.



Human milk is the gold standard for healthy human infant feeding. Human milk contains unique bioactive oligosaccharides that play a significant role in brain development and increased immunity to infection in infants.[1,2] Milk oligosaccharides are typically composed of three to ten monosaccharide units, consisting of glucose (Glc), galactose (Gal), *N*-acetyl-glucosamine (GlcNAc), fucose (Fuc), and sialic acid (Neu5Ac). The core group present at the reducing end of milk oligosaccharides is either lactose (Galβ1−4Glc) or *N*-acetyl-lactosamine (Galβ1−4GlcNAc).[3] The most common oligosaccharides in human milk have a lactose unit (Galβ1−4Glc) at the reducing end.

Because milk oligosaccharides are highly polar, appropriate separation and identification techniques are required to characterize unknown compounds. Enzyme digestion by exoglycosidases combined with size exclusion chromatography and capillary electrophoresis (CE)[4] or porous graphitized carbon (PGC) separation techniques have been used frequently.[5] Several studies have shown that hydrophilic interaction liquid chromatography electrospray ionization coupled to mass spectrometry and/or fluorescence detection (HILIC-ESI-MS/FLD) enable sufficient separation for the characterization of neutral and acidic oligosaccharides from plant materials[6,7] and mammalian milk.[8] The elucidation of unknown oligosaccharides by HILIC-tandem mass spectrometry (MS/MS) alone can be challenging especially because reference standards for many of these oligosaccharides are not commercially available.

One objective of this work is to facilitate analysis in the identification of these oligosaccharides through the development of a library of tandem mass spectra. Mass spectral (MS) libraries enable the tentative identification of unknown compounds in complex matrices by matching known fragmentation patterns of electrospray derived ions present in tandem MS libraries.[9] Recently, this method was enhanced by matching the unknown and MS library spectra with different parent ions based upon consistently mass shifted peaks. This strategy is termed the Hybrid Library Search[10] and simplifies the recognition of unknown compounds in the sample based on their similarity to known and well-characterized reference spectra.

Received: March 15, 2018
Accepted: July 3, 2018
Published: July 3, 2018

DOI: 10.1021/acs.analchem.8b01176
Anal. Chem. 2018, 90, 8977−8988

Another objective is to develop methods for creating libraries that include recurring spectra of components not commercially available but identifiable to a meaningful extent using current data analysis methods. Such "material-oriented" libraries are needed to deal with complex biological samples analyzed by mass spectrometry.

This study reports the creation of a MS library of oligosaccharides from the NIST Standard Reference Material on human milk (SRM 1953), which is significant because there is little published characterization of the highly polar and complex composition of oligosaccharides in SRM 1953. Additionally, we describe the process of the structural assignment for isomeric oligosaccharides using the NIST 17 Tandem MS Library[11] and hybrid search method, a process of general applicability.

MATERIALS AND METHODS

Standard Reference Material (SRM) 1953 was obtained from the National Institute of Standards and Technology (Gaithersburg, MD). SRM 1953 is a human milk pool from one hundred breastfeeding mothers (https://www-s.nist.gov/srmors/view_detail.cfm?srm+1953). The human milk sample was stored in a sterile container and kept frozen (−80 °C) until use. Water used in the sample preparation was LC-MS grade. All other chemicals used were of analytical grade.

Extraction and Purification of Human Milk Oligosaccharides. Oligosaccharides from SRM1953 (2 mL) were extracted and purified by solid phase extraction (SPE) followed by drying as previously described.[12] Briefly, 2 mL of SRM 1953 was centrifuged at 14 000*g*, 4 °C for 30 min. The liquid layer was transferred by pipet and mixed with four volumes of 2:1 v/v of chloroform−methanol solvent and centrifuged at 14 000*g*, 4 °C for 30 min. Proteins were precipitated overnight at 4 °C by adding two volumes of ethanol into the mixture and then centrifuged at 14 000*g*, 4 °C for 30 min. The decanted liquid was evaporated to dryness prior to solid phase extraction and HILIC-ESI-MS analyses (Supporting Information S2).

UHPLC-HILIC-MS/MS Analysis of Oligosaccharides. An Ultimate 3000 UHPLC system (Thermo Scientific) coupled to an Orbitrap mass spectrometer (Thermo Scientific Orbitrap Fusion Lumos) was used for the analysis of the samples. The chromatographic separation was performed on an ACQUITY Glycoprotein BEH Amide column, 300 Å (1.7 μm, 2.1 mm × 150 mm, Waters Corporation, Milford, MA, U.S.A.). The acquisition time was 65 min, and the mobile phase had a flow rate of 400 μL/min, pH 4.5, and a column oven temperature of 35 °C. The injection volume was 10 μL. The composition of the two mobile phases was 10 mmol/L ammonium formate with 0.1% (v/v) formic acid (A) and 99.9% (v/v) ACN with 0.1% (v/v) formic acid (B). The elution program was performed as follows: 1.5 min isocratic 95% (v/v) B; 8.5 min linear gradient from 95% (v/v) to 80% (v/v) B; 50 min linear gradient from 80% (v/v) to 50% (v/v) of B followed by 5 min of column washing with a linear gradient from 50% (v/v) to 2% (v/v) B including column re-equilibration with 95% (v/v) B. During the column washing, the flow rate was set at 250 μL/min. The electrospray MS detection was performed in positive and negative detection mode for both neutral and acidic oligosaccharides with the ion source voltage set to ±3.5 kV; the capillary temperature of 250 °C; sheath gas of 15 (arbitrary units); auxiliary gas of 10 (arbitrary units). Spectra were acquired using both ion trap and beam-type collision cell fragmentation, both with spectrum measurement at high mass accuracy in the orbitrap mass analyzer. The former is referred as Fourier transform ion trap (FT-IT), in which all spectra were acquired at the "Normalized Collision Energy" setting (NCE) of 35% and *Q* value of 0.25. The latter is called "HCD" (higher energy dissociation) by the instrument maker, although fragmentation patterns are equivalent to those of most triple quadrupole and QTOF instruments at comparable collision energies.[13] These spectra were acquired at NCE values of 10, 15, 20, 25, 30, 40, and 50. Each sample was analyzed in triplicate.

Unidentified Spectra of Oligosaccharides for Mass Spectral Matching. The acquired FT-IT and HCD MS^2 spectra from the raw HILIC-MS/MS data were sorted and clustered into consensus spectra using NIST algorithms[14] to create a library of unknown consensus spectra. A consensus spectrum is a weighted average of the similar spectra having the same precursor ion. Each spectrum must have a minimum match factor (MF) score of 999 based on similarity in peak relative intensities and fragment masses (Supporting Information S2).

Methods for Identification and Annotation of Unknown Spectra. The following describes the systematic analysis of neutral and acidic oligosaccharides in SRM 1953. Identification of oligosaccharides in the unknown MS library of SRM 1953 was done manually, based on the literature and using results of searches against the NIST Tandem spectral library. Spectra were first examined individually; then, corresponding entries were matched in the unknown MS library. Nine considerations in making these identifications are described below. The first three of these used NIST MS Search 2.3 software[15] while the last six were done manually (Supporting Information S2).

Library MS Search Results. Spectra from the unknown MS library were searched against the NIST 17 Tandem MS Library using the NIST MS Search 2.3 software[15] using Simple and MS/MS hybrid search methods. Search parameters such as precursor m/z and product ion masses were set to error tolerances of 10×10^{-6} and 50×10^{-6} mg/kg, respectively. The search software generates the possible oligosaccharide structures based on the similarity of peak intensity and masses between the consensus spectrum and the library reference oligosaccharide spectrum.

Hybrid Search Peak Alignment. The hybrid search method matches query spectra with library spectra that differ by discrete chemical groups. The basic principle of the search is that, when two precursor ions differ only in a single modification that does not greatly affect the fragmentation mechanism, each product ion peak in one spectrum of one precursor corresponds to a peak created by exactly the same fragmentation in the other precursor spectrum.[10] This is done by shifting library peaks by the difference in the query and library mass (DeltaMass). For example, reduction of carbohydrates with a reducing group modifies the core unit lactose to lactitol, thereby adding two H atoms (m/z 2.004) to the precursor ion m/z. This method is incorporated into the freely available NIST MS Search 2.3 software.[15] The software is intended for the mass spectral matching of unknown and library tandem mass spectra (high accuracy), qualitative characterization, and illustration of fully annotated MS^2 spectra of oligosaccharides in human milk.

Fragment Annotation. The corresponding B/Y and C/Z type ions were used to annotate and distinguish isomeric structures as previously described[16] and by using SimGlycan.[17]

DOI: 10.1021/acs.analchem.8b01176
Anal. Chem. 2018, 90, 8977−8988

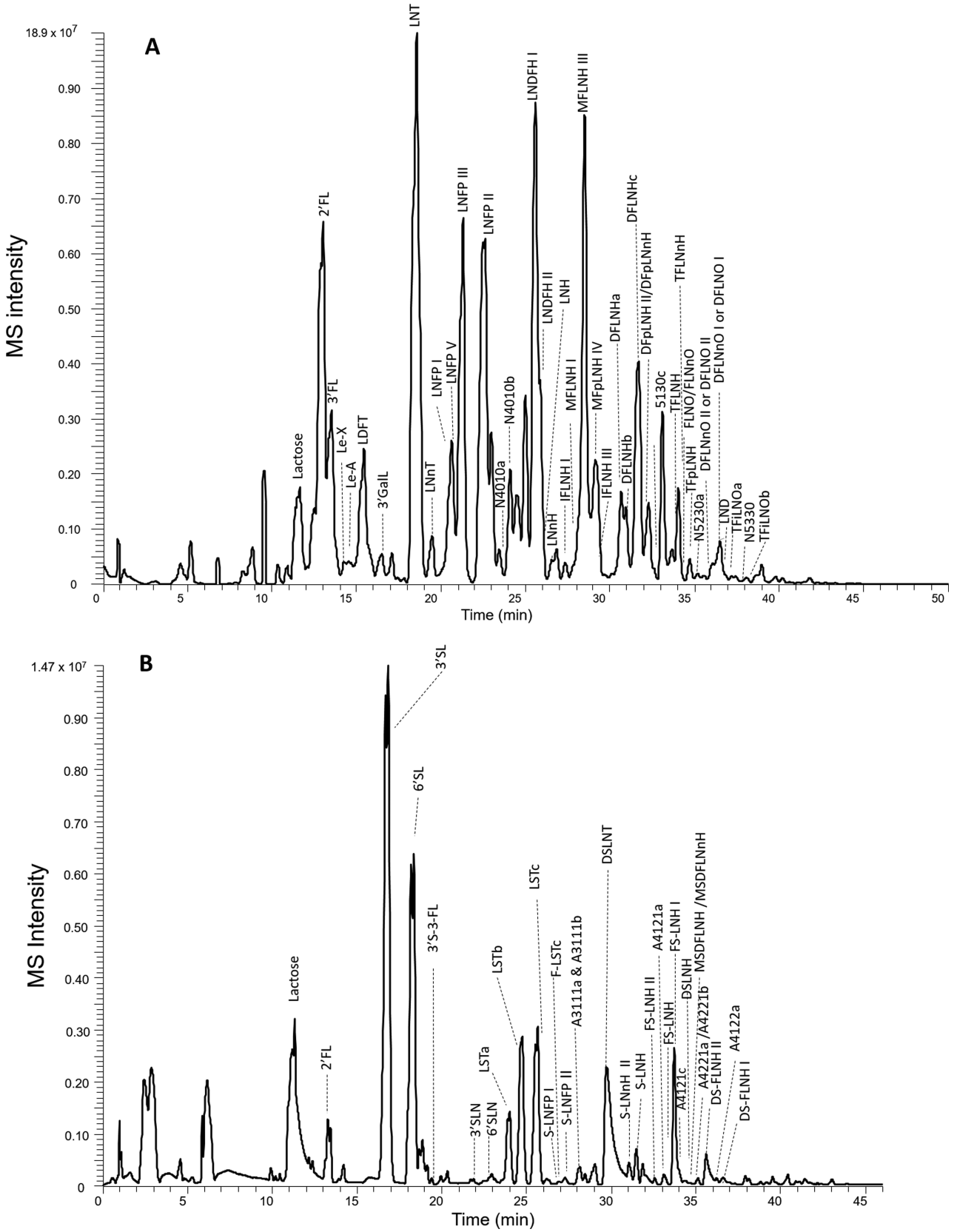


Figure 1. Base peak HILIC-MS elution profile of free oligosaccharides in SRM 1953 human milk in positive detection mode. (A) Neutral oligosaccharides. (B) Acidic oligosaccharides. Refer to Table 2 for the description and annotation of peaks. Annotation number: A4121 means 4 hexose; 1 fucose; 2 GlcNAc; 1 Neu5Ac.

Utilizing these fragmentation rules, fragment annotation was comprehensively conducted for each spectrum as described (Supporting Information S1). Initially, all possible fragments were acquired in Glypy 0.11.3[18] for the spectrum when

DOI: 10.1021/acs.analchem.8b01176
Anal. Chem. 2018, 90, 8977−8988

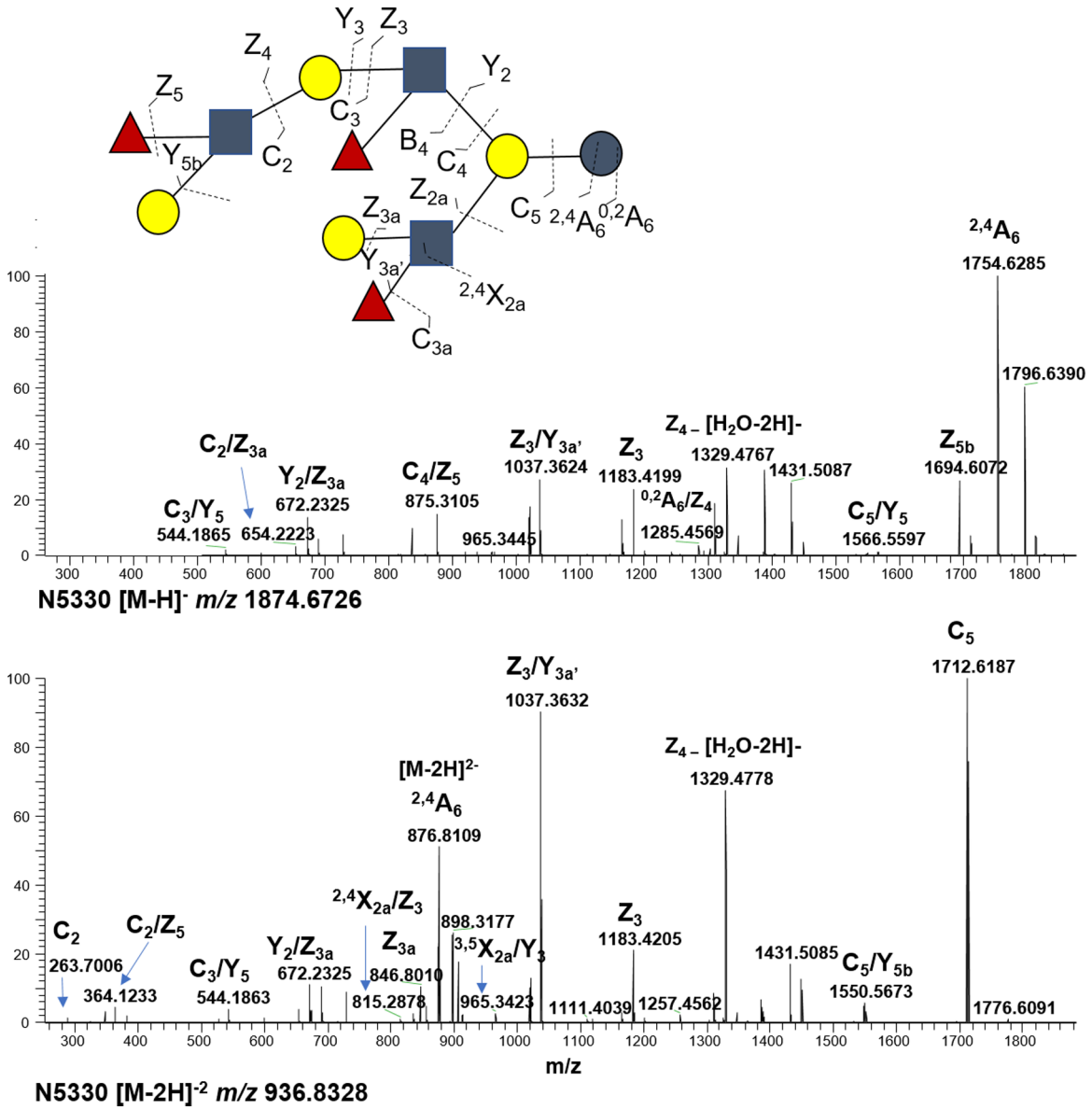


Figure 2. Negative ion FT-IT MS^2 fragmentation pattern of trifucosyl *iso*-lacto-*N*-octaose. Annotation of deprotonated singly $[M - H]^-$ and doubly $[M - 2H]^{-2}$ charged states molecular ions enabled one to distinguish isomeric structures. Annotation number: N5330 denotes N: neutral, 5 hexose; 3 fucose; 3 GlcNAc; 0 Neu5Ac. C_3/Y_5 means glycosidic−glycosidic linkage; $^{2,4}X_{2a}/Z_3$ means cross-ring−glycosidic linkage.

considering all single and double cleavages that could occur with the associated glycan structure. Theoretical *m/z* values were then combinatorially generated from these neutral fragment masses when considering all common adduct types, water losses and gains, and isotopic shifts. Finally, annotations were assigned to a peak if the theoretical *m/z* value matched the experimental *m/z* value to within 10×10^{-6} mg/kg. This annotation provides information on the chemical composition, branching site, and sequence type present in the oligosaccharide. Specific illustrations of these ideas are discussed in the following sections.

RESULTS AND DISCUSSION

Elution Profile of Milk Oligosaccharides. Elution of different oligosaccharides in the sample is caused by HILIC using different commercially available milk oligosaccharides (Table S1). Base peak chromatograms of neutral and acidic fractions are displayed in Figure 1.

Figure 1A displays the elution profile of fucosyllactose (2′FL), lacto-*N*-tetraose (LNT), lacto-*N*-difucohexaose (LNDFH), and monofucosyllactose-*N*-hexaose (MFLNH), known to be the most abundant neutral oligosaccharides in human milk. The milk oligosaccharides showed an excellent separation on a HILIC column related to their size. The elution of 2-FL prior to Neu5Ac-lactose (3-SL) illustrates the selectivity of HILIC. Sialylated oligosaccharides display increased polarity relative to fucosylated oligosaccharides because they contain an additional carboxylate ion (COO−). The clear distinction in the elution pattern of fucosylated and/or sialylated oligosaccharides confirmed the HILIC separation, which is based on both size and polarity. Neutral oligosaccharides, FL, LNFP, and F-LNH isomers (Figure 1A), and sialylated oligosaccharides, SL, LST, and S-LNFP isomers (Figure 1B), were distinguished.

MS/MS Analysis. Structural assignment of isomeric oligosaccharides is challenging due to the heterogeneity and

DOI: 10.1021/acs.analchem.8b01176
Anal. Chem. 2018, 90, 8977−8988

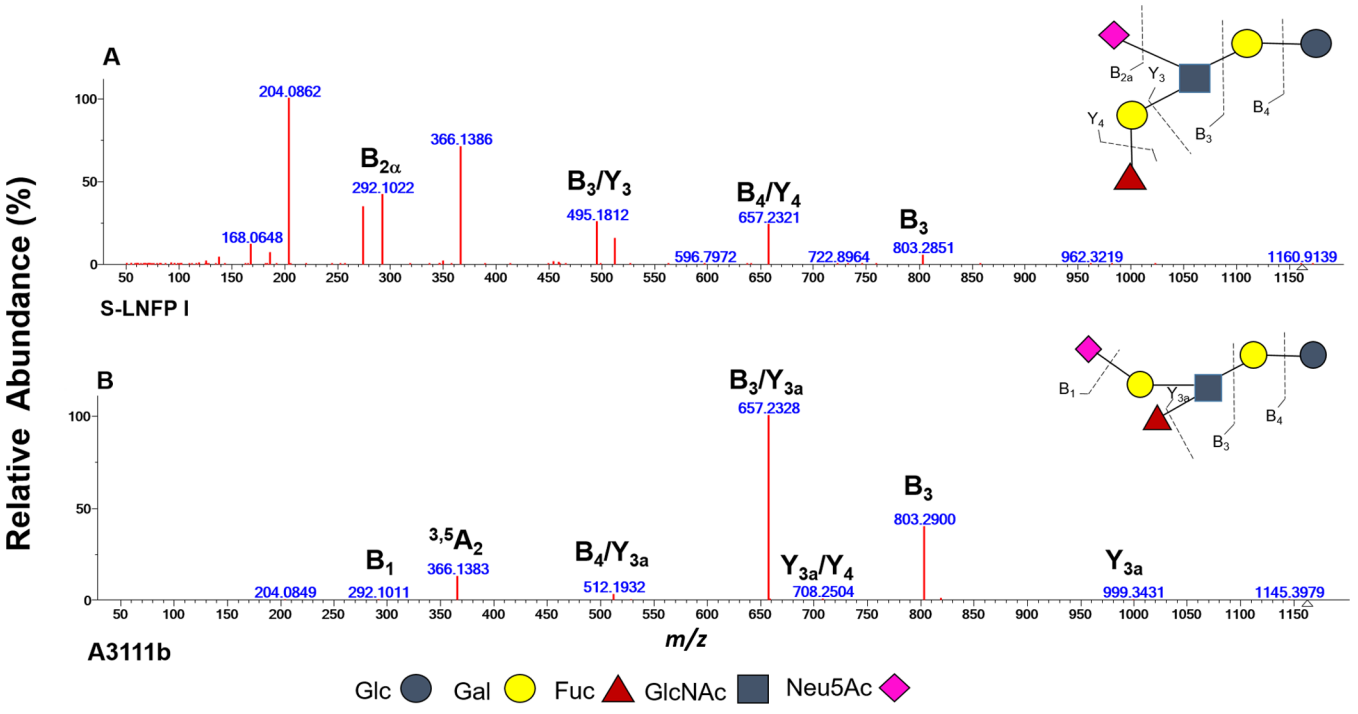


Figure 3. Identification and annotation of sialylated lacto-*N*-pentaose isomers with precursor ions of m/z 1162.436. (A) S-LNFP I and previously unreported (B) S-LNFP (A3111b). B_3/Y_{3a} means glycosidic−glycosidic linkage.

complexity of monosaccharide composition and linkages. The following sections described the annotation of unknown experimental HCD and FT-IT MS^2 spectra in the sample. The chemical composition constraints and glycosidic bonds and cross-ring fragmentation patterns rules were applied in the annotation of similar precursor ions (isomers), singly and doubly charged precursor ions of neutral and sialylated oligosaccharides. This method demonstrates the recognition of different branching patterns and linkages in the structures of fucosylated and sialylated oligosaccharides.

Neutral Milk Oligosaccharides. Monofucosylated lacto-*N*-hexaose isomers (MFLNH III and MFpLNH IV) eluted at different retention times with precursor ions m/z 1219.4 $[M + H]^+$ have product ions that are found to be useful to distinguish branched from linear structure (Figure S2.1). This technique was then used to predict the possible chemical structures of three trifucosyl *iso*-lacto-*N*-octaose isomers eluting at longer retention time (37 to 39 min; Figure 1A).

It is known that singly charged state ions in FT-IT spectra do not allow the trapping of fragments with m/z values lower than one-third of the precursor mass.[19] Figure 2 illustrates the MS^2 spectra of singly $[M - H]^-$ (m/z 1874.67) and doubly $[M - 2H]^{-2}$ (m/z 936.83) charged for oligosaccharide eluted at 38.56 min. C, Z, and A type fragment ions are dominant. The unknown spectra for N5330 isomer has the peak signal at m/z 1037.3632, evidence of the internal fucose residue at the β1−6 branch and consistent with previously reported TF*i*LNO b.[20] Furthermore, product ions of $[M - 2H]^{-2}$ such as m/z 364.1233, m/z 544.1863, and m/z 672.2325 indicate the Gal(β1−4)Fuc(α1−3)GlcNAc sequence at the terminal β1−3 branch with two cross-ring-glycosidic linkages $^{2,4}X_{2a}/Z_3$ (m/z 815.2878) and $^{3,5}X_{2a}/Y_3$ (m/z 965.3423) illustrating the internal ring cleavage of GlcNAc at the terminal β1−3 branch. The information provided by the low mass ions from a doubly charge spectra is important in the structural assignment of trifucosyl octasaccharides[20] TF*i*LNOa, TF*i*LNOb (Figure S2.2), and the proposed new structure of N5330 oligosaccharide.

Sialylated Milk Oligosaccharides. The HILIC elution profile of sialylated oligosaccharides is shown in Figure 1B. The 3-SL having a terminal Neu5Ac(α2−3) linked to a lactose unit elutes before 6-SL with a Neu5Ac(α2−6) linkage. It was observed that LSTa having a terminal Neu5Ac(α2−3) linked to a lactose unit elutes before LSTb with Neu5Ac(α2−6)GlcNAc and LSTc with a Neu5Ac(α2−6) linkage, respectively.

Moreover, unknown precursor ions m/z 1162.436 of oligosaccharides were identified as sialyl-lacto-*N*-fucopentaose (S-LNFP) isomers eluted at 26.5 to 28.8 min (Figure 3A,B). S-LNFP I and S-LNFP II were previously identified in human milk.[21] Note that fragment ions of ammoniated precursor are protonated due to loss of ammonia.[22] As expected, S-LNFP isomer exhibits prominent B and Y type ions in positive mode detection. Diagnostic ions such as m/z 495.1797 (S-LNFP I) and m/z 454.1570 (S-LNFP II) characterize the linkages Neu5Ac(α2−6) and Neu5Ac(α2−3), respectively. The m/z 495.1797 ion is evidence that the Neu5Ac residue links to GlcNAc residue as previously observed.[23] Product ions of m/z 454.1570 and m/z 495.1797 are not present in A3111b as observed with the product ions of LSTc indicating that the Neu5Ac residue is α2−6 linked to a terminal galactose. These observations suggest that S-LNFP I and S-LNFP II (Table 2) elute from HILIC prior to the proposed isomeric structure A3111b. The latter oligosaccharide was reported to be the conjugate glycan of glycoprotein or glycolipid group belonging to the sialyltransferase gene family[24] but has not been reported previously as free oligosaccharide in human milk.

MS Library Aided Identification of Milk Oligosaccharides. MS library searches could produce high scores when library and the unknown spectra have similar or different

DOI: 10.1021/acs.analchem.8b01176
Anal. Chem. 2018, 90, 8977−8988

Table 1. MS/MS Direct Search of the Consensus MS^2 Spectra in Milk SRM 1953 against NIST 17 Tandem MS Library for Glycans

name	RT	theoretical m/z	experimental m/z	precursor type	collision energy[a] (NCE)	MF score[b]	RMF score[c]
			Neutral Oligosaccharides				
2′fucosyllactose, 2′FL	13.17	511.1633	511.1624	$[M + Na]^+$	25	965	988
Lex-X trisaccharide, Le-X	13.46	552.1893	552.1899	$[M + Na]^+$	*d*	811	914
Le-A trisaccharide, Le-A	25.11	512.1974	512.1980	$[M + H - H_2O]^+$	20	987	993
lacto-*N*-tetraose, LNT	18.79	730.2376	730.2380	$[M + Na]^+$	30	909	909
lacto-*N*-neotetraose, LNnT	19.18	730.2376	730.2399	$[M + Na]^+$	30	945	978
lacto-*N*-fucotetraose I, LNFP I	21.77	446.6384	446.6376	$[M + H + K]^{2+}$	15	933	980
lacto-*N*-fucotetraose III, LNFP III	22.75	876.2955	876.2956	$[M + Na]^+$	40	942	870
difucolacto-*N*-hexaose c, DFLNHc	30.21	1387.4856	1387.4862	$[M + Na]^+$	*d*	850	882
difucoparalacto-*N*-hexaose II, DFpLNH II	32.43	1387.4856	1387.4856	$[M + Na]^+$	*d*	927	957
trifucoparalacto-*N*-hexaose, TFpLNH	34.92	1533.5435	1533.5435	$[M + Na]^+$	*d*	835	891
			Acidic Oligosaccharides				
3′sialyllactose, 3′SL	17.01	656.2009	656.2033	$[M + Na]^+$	20	899	994
6′sialyllactose, 6′SL	18.20	656.2009	656.1980	$[M + Na]^+$	20	900	919
3′-α-sialyl-*N*-acetyllactosamine, 3′SLN	23.70	657.2349	657.2360	$[M + H - H_2O]^+$	15	900	919
6′-α-sialyl-*N*-acetyllactosamine, 6′SLN	25.47	657.2349	657.2354	$[M + Na]^+$	15	992	995
sialyllacto-*N*-tetraose b, LSTb	24.55	1021.3330	1021.3399	$[M + Na]^+$	40	857	871

[a]Normalized collision energy (NCE) = 10 to 50 eV. [b]Match factor (MF) score. [c]Reverse match factor (RMF) score (nonmatching peaks in query spectra are ignored). [d]FT-ITMS = 35%.

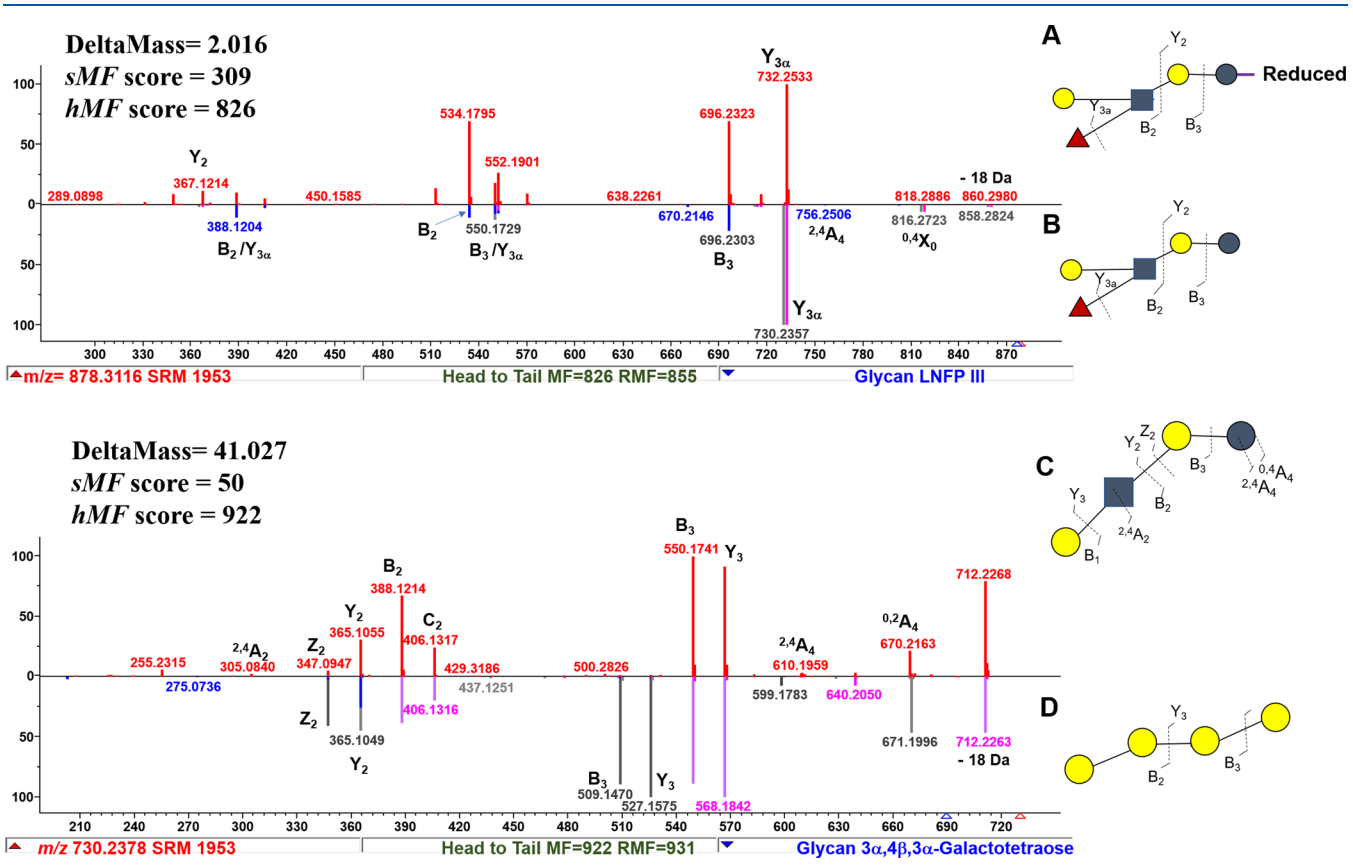
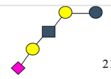
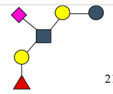
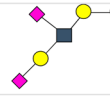
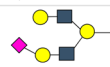
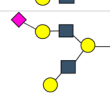
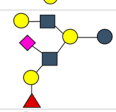
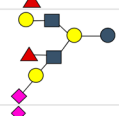
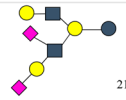


Figure 4. Hybrid MS library search identifications. Precursor m/z values of adduct ion $[M + Na]^+$ (A) m/z 878.3116 (reduced); (B) LNFP III (m/z 876.2963); (C) m/z 730.2378; (D) 3α,4β,3α-galactotetraose (m/z 689.2111). The head-to-tail plot shows the spectral matching of product ions of the unknown (red) against the known ions (blue) in NIST Tandem MS Library 2017. Shifted peak (gray line); inserted/predicted peak (pink line). Original match factor score (*sMF*) and hybrid match factor (*hMF*). DeltaMass is the difference of precursor m/z values between the unknown consensus spectrum and the NIST 17 library spectrum.

precursor m/z values (DeltaMass). For mass spectral matching, HILIC-MS^2 data of milk oligosaccharides in SRM 1953 were processed and clustered into high mass accuracy HCD/FT-IT MS^2 of unknown consensus spectra using a nearest-neighbor clustering algorithm[14] with the following constraints: first, two spectra cannot belong to the same cluster if they have different

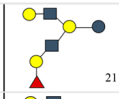
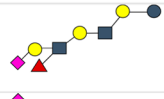
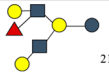
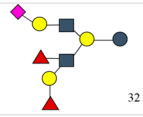
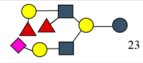
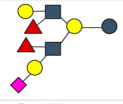
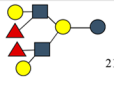
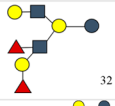
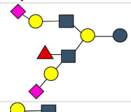
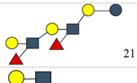
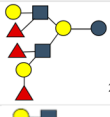
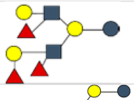
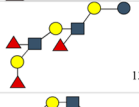
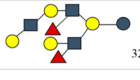
DOI: 10.1021/acs.analchem.8b01176
Anal. Chem. 2018, 90, 8977−8988

Table 2. Annotation of Peaks in Figure 2 Derived from SRM 1953 Human Milk Sample by HILIC-MS/MS Using HCD and FT-IT Fragmentation Techniques

Name	Proposed Structure[‡]	RT (min)	Precursor Type		Collision Energy [#]NCE	Name	Proposed Structure[‡]	RT (min)	Precursor Type		Collision Energy [#]NCE
			+	-					+	-	
Neutral Oligosaccharides						**Acidic Oligosaccharides**					
2'-FL		12.91	a, b, c, d	f, g, h, i	20-50, *	3'-SL		15.82	a, b, c		10-40, *
3'-FL		13.99	b	g	25, *	6'-SL		17.18	a, b, c		15-50, *
H-Tri		12.26	a, c		*	3'-S-3-FL		19.96	b, c	f, g	10-50, *
Le-X Tri		13.80	a, c		10, 40, *	3'-SLN		23.71	l		10-50, *
Le-A Tri		22.95	l		20	6'-SLN		25.47	l		10-40, *
LDFT		15.61	b, c	g, i	10-50, *	LSTa		23.92	a, b	f	10-50, *
3'GalL		16.04	b, c		10-50, *	LSTb		24.73	a, b, c		10-50, *
LNT		18.75	b, c	f, g, I, j	10-50, *	LSTc		25.23	a, b, c	f	20-25, *
LNnT		19.16	a, c	f, g	10-50, *	S-LNFP I		26.10	a, b	f	*
Le B tetra		19.29	a, c		10, *	F-LSTc		26.10	c		10-20, *
LNFP I		21.77	e	f, g	10-30, *	S-LNFP II		26.55	a, b	f	10-25, *
LNFP V		22.09	c	f, g	10-40, *	A3111a		28.30	a, b	f	*
LNFP III		22.71	a, b, c	f, g	10-50, *	A3111b		28.45	c		10-40, *
LNFP II		22.99	a, b, c	f, g	10-50, *	DSLNT		29.59	a, b, e	f	10-20, *
N4010a		23.46	a, b, c	f	10-30, *	S-LNnH II		30.00	a, b	f	10-25, *
N4010b		24.00	a, b	f	10, *	S-LNH		31.40	e	f, g	*
LNDFH I		26.10	a, b, c	g	10-30, *	A4121a		31.48	a	f	10-40, *
LNDFH II		26.60	c	g, h	*	A4121b		32.30	a, b		10-20, *
LNH		26.10	a, b	f, g	10-30, *	FS-LNH		33.33		f	20, *
LNnH		26.65		f, g	20, *	DSLNH		33.40	a, b		*
IFLNH I		27.85	a		*	FS-LNH I		33.72		f	*

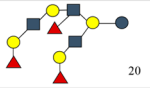
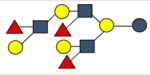
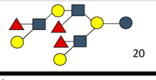
DOI: 10.1021/acs.analchem.8b01176
Anal. Chem. 2018, 90, 8977−8988

Table 2. continued

Name	Proposed Structure[‡]	RT (min)	Precursor Type +	Precursor Type -	Collision Energy [#]NCE	Name	Proposed Structure[‡]	RT (min)	Precursor Type +	Precursor Type -	Collision Energy [#]NCE
Neutral Oligosaccharides						**Acidic Oligosaccharides**					
MFLNH I		28.11	b	f	10-25, *	A4121c		33.86	a		*
MFLNH III		28.59	a, b, e	f, g, j	10-30, *	MSDFLNH		34.73	c	f, g	10-25, *
MFpLNH IV		29.54	a, b	f	10-15, *	MSDFLNnH		34.95	c	f, g	15, 30, *
IFLNH III		29.71	a, b	f	10-20, *	A4221a		35.03	a, b	g	20, *
DFLNH a		31.20	b, e	f, g	10-20, *	A4221b		35.30	a, b	g	20, *
DFLNH b		32.09	b	g	10, 15	DS-FLNH II		34.65		j	*
DFLNH c		32.29	a, b, c	g	10-30, *	A4122a		35.42	a, b		*
DFpLNH II		32.53	a, b, c	f, g	15-30, *	DS-FLNH I		35.77	e	j	20-25, *
DFpLNnH		32.65	c		*						
TFLNH		34.07	a, b, c	k	20-40, *						
TFLNnH		34.77	a, b	g	20, 25, *						
TFpLNH		34.98	a, b	g	*						
5130c		33.60		J	15-30, *						
FLNO		34.32	a, b, c	f, g	10-30, *						
FLNnO		34.66	a, b	g, j	*						
N5230a		34.97	a, b	f, k	*						
DFLNO II		35.95	a	f, k	*						
DFLNnO II		36.20	a, e	k	*						
LND		35.63	e		*						

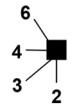
DOI: 10.1021/acs.analchem.8b01176
Anal. Chem. 2018, 90, 8977−8988

Table 2. continued

Name	Proposed Structure[‡]	RT (min)	Precursor Type +	Precursor Type -	Collision Energy [#]NCE
Neutral Oligosaccharides					
DFLNO I		36.69	a, b, e	f, k	*
DFLNnO I		37.05	a, b, e	f, k	*
TFiLNO a		37.26		f, j	*
N5330		38.56		f, j	*
TFiLNO b		39.03		f, j	*
Acidic Oligosaccharides					

Annotation code
A3111 denotes N=Neutral, A=Acidic, 3 Hexoses, 1 Fucose, 1 GlcNAc, 1 Neu5Ac
Positive mode:
[a][M+H]$^+$ [b][M+NH$_4$]$^+$ [c][M+Na]$^+$ [d][2M+Na]$^+$ [e][M+2NH4]$^{2+}$ / [M+H+K]$^{2+}$
Neutral loss [l][M-H$_2$O-H]$^+$
Negative mode:
[f][M-H]$^-$ [g][M+H$_2$CO$_2$ -H]$^-$ [h][M+HC$_2$F$_3$O$_2$ -H]$^-$ [i][2M-H]$^-$
[j][M-2H]$^{2-}$ [k][M+2H$_2$CO$_2$ -2H]$^{2-}$
*IT-FTMS = 35%
[#] Normalized Collision Energy (NCE) = 10, 15, 20, 25, 30, 40, 50 eV
[‡] Reference

Linkages
α 1-6/α 1-2
β 1-4
β1-3/β1-6
α 2-3/α 2-6



charges, NCE, or library ID. Second, differences in precursor m/z values and retention time have a threshold of 20×10^{-6} mg/kg and ±0.3 min, respectively.

Unknown spectra were then matched by direct and hybrid search against the NIST 17 Tandem MS Library using an automated search software NIST Mass Spectral Search Program (version v2.3). The program uses a modified vector dot product to calculate a match factor[9,10,25,26] that ranges from 0 (no peaks in common) to 999 (identical spectra). As shown in Table 1, the direct search matched oligosaccharides in the NIST 17 library spectra with values ranging from 811 to 965 with a median of 909. This strategy is complementary to the previously reported chromatography-retention time[8,27] based experiments.

Direct Identification. Oligosaccharides that matched compounds in the NIST 17 library were tentatively identified using the conventional direct search, where both the precursor m/z charge and spectrum must match. A good consensus spectrum match typically had a match factor (MF) score of >800. Reverse match factors (RMF) treat peaks not in the library as possible contaminants and yield high scores. Since different isomers may have indistinguishable spectra, it is essential to use other factors to assign isomeric structures. Table 1 shows 15 milk oligosaccharides with various precursor ions and normalized collision energies that matched NIST 17 library entries with MF ranging from 850 to 992 while the reverse match factor (RMF) varies from 871 to 995. Neutral and sialylated oligosaccharides present in the sample such as lacto-*N*-tetraose (LNT), lacto-*N*-fucopentaose (LNFP), and sialyllactose (SL) produced MF most often above 900.

Hybrid Search Identification. MS reference library requires an in-depth characterization of available data, especially when the experimentally acquired spectra produce match factor scores below 800, indicating lack of identity with available MS libraries. The hybrid search method can assist in the identification of several varieties of oligosaccharides. The m/z difference between the consensus spectrum and mass library spectrum (DeltaMass) was previously described for hybrid search identification of peptides[10] and fentanyl-related compounds.[26] We now described for the first time the extension of this method in the identification of reduced oligosaccharides or oligosaccharides differing with a single or multiple sugar units.

One example is its ability to link reduced to nonreduced spectra, by shifting peaks containing the reduced group by two Da. The reduction is often used in carbohydrate analyses to simplify oligosaccharide chromatographic analysis.[27] This is illustrated in Figure 4A where one unknown spectrum (m/z 878.312) was searched against the NIST 17 MS library, finding the nonreduced glycan LNFP III. As expected, the calculated mass difference is due to the DeltaMass between the m/z values of the Y type fragment ions of the nonreduced versus reduced oligosaccharides. With the direct search, the simple match factor (*sMF*) score is 309 (relative to a maximum of 999); however, shifting the Y type fragment ions (gray line) by two Da (pink line) produced a hybrid match factor (*hMF*) of 826 (Figure 4B).

Another example is the ability of the hybrid search to link glycans differing by a single sugar unit to confirm the direct MS search of LNT. This is illustrated in Figures 4C,D and S2.9, where the unknown spectrum (m/z 730.2378) matches with the library spectrum of 3α,4β,3α galactotetraose with a DeltaMass of 41.027 Da. This is consistent with the unknown compound containing a GlcNAc sugar unit instead of the Gal residue in the library compound. This is demonstrated in Figure 1D, where fragment ions m/z 509.1470, m/z 527.1575, and m/z 671.1996 are shifted by 41.027 Da to m/z 388.1214 (B_2), m/z 406.1317 (C_2), m/z 550.1741 (B_3), m/z 568.1842 (Y_3), and m/z 712.2263. Note that LNT is one of the most abundant neutral and nonfucolactosylated milk oligosacchar-

DOI: 10.1021/acs.analchem.8b01176
Anal. Chem. 2018, 90, 8977−8988

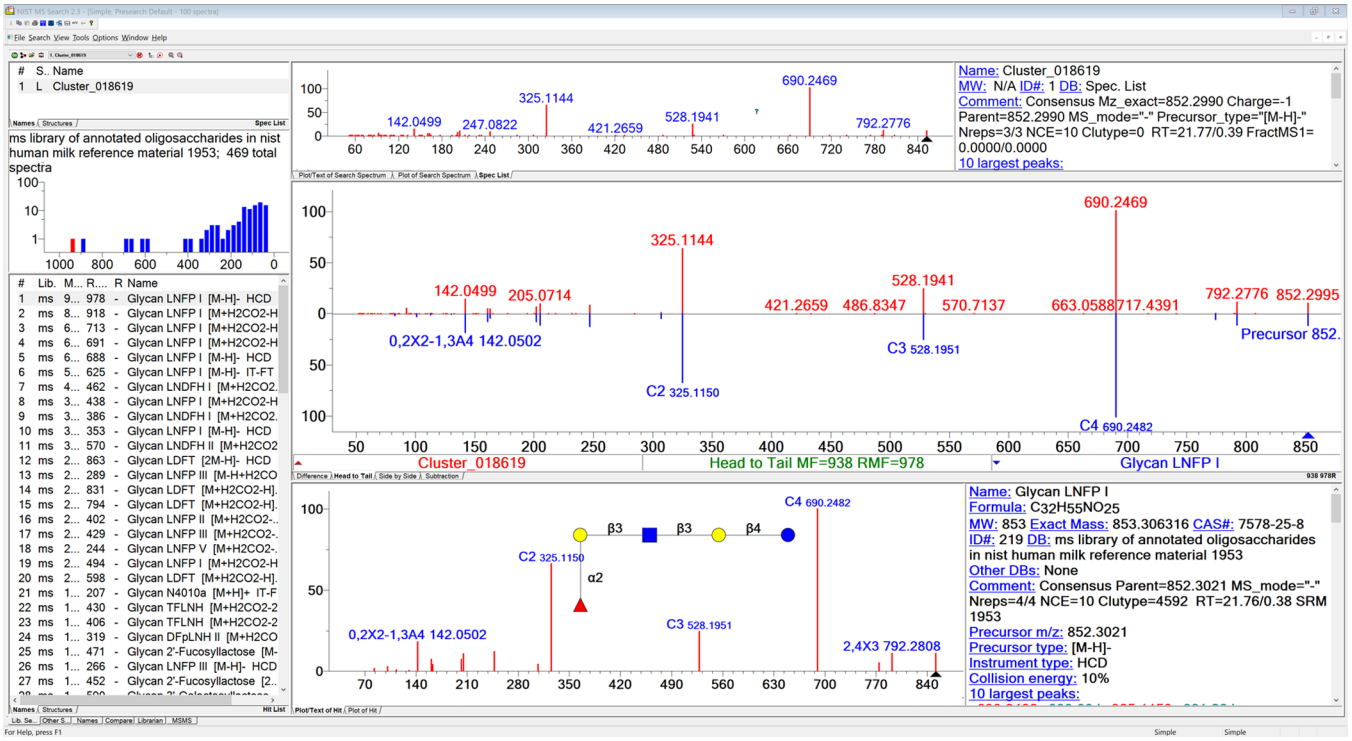


Figure 5. Overview of NIST MS search interface illustrating the data information and comparison of the MS^2 spectra between the unknown (top) and the annotated peaks of LNFP I (bottom). The head-to-tail comparison of the unknown spectrum (cluster 018619) and LNFP I (middle). Match factor (MF) score = 938; Reverse MF score = 978.

ides in the sample, illustrating how the hybrid search strategy aids the identification of unknown spectra. This strategy may be useful in the identification of permethylated and other derivatized oligosaccharides.

MS Library of Annotated Oligosaccharides in NIST Human Milk Reference Material. Raw HILIC MS/MS data of 196 runs were processed, clustered to create consensus spectral files, and searched using the hybrid search and Tandem MS Library[11] as described above to create a reference material-based library.

Table 2 displays the list of 74 oligosaccharides that were identified and elucidated in the human milk sample, of which 45 are neutral and 29 are sialylated oligosaccharides. The MS library of identified and annotated oligosaccharides has different adduct ions and normalized collision energies using HCD and FT-IT fragmentation techniques. Among the precursor and product ions, positive adduct ions of $[M + NH_4]^+$ and $[M + H]^+$ were abundant. Precursor ions $[M - H]^-$ and $[M - H + H_2CO_2]^-$ are the common adduct ions observed in negative detection mode. Annotation of the negative HCD/FT-IT MS^2 spectra allows the distinction of various fucosyl (α1−2, α1−3/α1−4) glycosidic linkages and cross ring cleavages (A type ions) present in the oligosaccharide structure.

The hybrid search technique enabled identification of reduced known oligosaccharides and 12 previously unreported free oligosaccharides in human milk. The high mass accuracy and high signal resolution of the HCD/FT-IT spectra confirmed most of the oligosaccharides using both positive and negative detection. Extensive analysis of the FT-IT spectra enabled resolution and annotation of about 30% of the precursor ions in the higher mass region ($<m/z$ 2000) using the dual charge state fragmentation strategy.

So far, the remaining partially identified oligosaccharides at longer retention time (11−12 sugar units) require additional analysis for identification because of the ambiguity in structural assignment of terminal fucose, galactose, and sialic acid linkages (Figure S2.12).

Searching Unknown Spectra Using the MS Library of Annotated Milk Oligosaccharides. The MS library of oligosaccharides in this study derived from human milk SRM 1953 is available online[34] along with search software and can be readily applied to other bovine milk samples and biological fluids (Supplementary Figures S2.6 and S2.7). The MS library consists of 469 positive and negative ion spectra having 45 neutral and 29 acidic oligosaccharides. All fragments ions of MS^2 spectra are comprehensively annotated. Figure 5 illustrates the library search of an unknown spectrum against the mass spectral database of milk oligosaccharides. The head-to-tail comparison between the unknown spectrum and LNFP I shows the similarity of fragment ions in terms of peak intensity and their m/z values as interpreted by the MF score of 938.

CONCLUSIONS

Human milk reference material SRM 1953 containing neutral and acidic oligosaccharides was analyzed on HILIC coupled to electrospray ionization with an Orbitrap-based mass spectrometer using HCD and FT-IT fragmentation techniques. Consensus library spectra of MS^2 ions were generated from the raw HILIC-MS data. The NIST Tandem MS Library with its new hybrid search algorithm facilitated the identification of unknown spectra of both nonreduced and reduced oligosaccharides. This strategy enabled spectral matching between the consensus and library spectra despite shifts in m/z values. Moreover, the high mass accuracy for both precursor and

DOI: 10.1021/acs.analchem.8b01176
Anal. Chem. 2018, 90, 8977−8988

product ions was sufficient to enable the assignment of 74 oligosaccharides including novel glycan structures. A clear distinction could be made in the elution pattern of isomeric fucosylated and sialylated oligosaccharides using previously established elution orders related to size and polarity. The reduction of oligosaccharides by sodium borohydride increased product chromatographic retention times and peak quality resulting to an addition of two hydrogen atoms to the precursor ions. The hybrid search can reliably find reduced oligosaccharides in comparison to nonreduced analogs, and vice versa, enabling an increase in match factors. The SRM 1953 milk oligosaccharide MS library has been demonstrated to aid in identifying oligosaccharides in other biological or milk samples.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.8b01176.

Supporting Information S1: Milk SRM 1953 glycan MS library (https://chemdata.nist.gov/srmd-1953-glycan/spectra). For download of the Milk SRM 1953 library (https://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:lib:oligosaccharides); Supporting Information S2: detailed methods on sample preparation and processing of spectra and structural assignment of neutral and acidic oligosaccharides; Table S1: list of commercially available oligosaccharides in the study; Table S2: diagnostic ions used for the identification and annotation of the consensus spectrum in the MS library of unidentified spectra (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: concepcion.remoroza@nist.gov.

ORCID

Connie A. Remoroza: 0000-0003-1540-1635

Notes

The commercial instruments and materials are used for the experimental part of the study. Such identification does not intend recommendation or endorsement by the National Institute of Standards and Technology, nor does it intend that the materials or instruments used are necessarily the best available for the purpose.

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Sanford Markey for detailed discussion. We also thank Drs. Dmitrii Tchekhovskoi, Arun Moorthy, Meghan Burke, and Yuxue Liang for fruitful discussions and technical support on NIST MS Search software, hybrid search method, and LC-MS instrument settings. Finally, we thank Joshua Klein for his contribution to the fragment ion annotation using *glypy* Python package.

■ REFERENCES

(1) Claps, S.; Di Napoli, M. A.; Sepe, L.; Caputo, A. R.; Rufrano, D.; Di Trana, A.; Annicchiarico, G.; Fedele, V. *Small Ruminant Research* **2014**, *121* (1), 116−119.

(2) Donovan, S.; Glenn, G.; Newburg, D. *Prebiotics in Infant Nutrition*; Mead Johnson Nutrition: Chicago, IL, 2009.

(3) Manz, C.; Pagel, K. *Curr. Opin. Chem. Biol.* **2018**, *42*, 16−24.

(4) Difilippo, E.; Willems, H. A. M.; Vendrig, J. C.; Fink-Gremmels, J.; Gruppen, H.; Schols, H. A. *J. Agric. Food Chem.* **2015**, *63* (19), 4805−4814.

(5) Zhou, S.; Huang, Y.; Dong, X.; Peng, W.; Veillon, L.; Kitagawa, D. A. S.; Aquino, A. J. A.; Mechref, Y. *Anal. Chem.* **2017**, *89* (12), 6590−6597.

(6) Remoroza, C.; Cord-Landwehr, S.; Leijdekkers, A. G. M.; Moerschbacher, B. M.; Schols, H. A.; Gruppen, H. *Carbohydr. Polym.* **2012**, *90* (1), 41−48.

(7) Leijdekkers, A. G.; Huang, J.-H.; Bakx, E. J.; Gruppen, H.; Schols, H. A. *Carbohydr. Res.* **2015**, *404*, 1−8.

(8) Albrecht, S.; Lane, J. A.; Marino, K.; Al Busadah, K. A.; Carrington, S. D.; Hickey, R. M.; Rudd, P. M. *Br. J. Nutr.* **2014**, *111* (07), 1313−1328.

(9) Stein, S. *Anal. Chem.* **2012**, *84* (17), 7274−7282.

(10) Burke, M. C.; Mirokhin, Y. A.; Tchekhovskoi, D. V.; Markey, S. P.; Heidbrink Thompson, J.; Larkin, C.; Stein, S. E. *J. Proteome Res.* **2017**, *16* (5), 1924−1935.

(11) *NIST/EPA/NIH Mass Spectral Library NIST*, version 17; June 6, 2017; https://www.nist.gov/srd/nist-standard-reference-database-1a-v17.

(12) De Leoz, M. L. A.; Gaerlan, S. C.; Strum, J. S.; Dimapasoc, L. M.; Mirmiran, M.; Tancredi, D. J.; Smilowitz, J. T.; Kalanetra, K. M.; Mills, D. A.; German, J. B.; Lebrilla, C. B.; Underwood, M. A. *J. Proteome Res.* **2012**, *11* (9), 4662−4672.

(13) Rohmer, M.; Baeumlisberger, D.; Stahl, B.; Bahr, U.; Karas, M. *Int. J. Mass Spectrom.* **2011**, *305* (2), 199−208.

(14) Yang, X.; Neta, P.; Stein, S. E. *Anal. Chem.* **2014**, *86* (13), 6393−6400.

(15) *The NIST Mass Spectral Search Program for the NIST/EPA/NIH Mass Spectral Library*, NIST version 2.3; June 6, 2017 http://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:nistmssearch.

(16) Domon, B.; Costello, C. E. *Glycoconjugate J.* **1988**, *5* (4), 397−409.

(17) Apte, A.; Meitei, N. S. Bioinformatics in Glycomics: Glycan Characterization with Mass Spectrometric Data Using SimGlycan. In *Functional Glycomics: Methods and Protocols*; Li, J., Ed.; Humana Press: Totowa, NJ, 2010; pp 269−281.

(18) Klein, J. *Glypy 0.11.3*; https://pypi.python.org/pypi/glypy/0.11.3; 2018.

(19) Guo, T.; Gan, C. S.; Zhang, H.; Zhu, Y.; Kon, O. L.; Sze, S. K. *J. Proteome Res.* **2008**, *7* (11), 4831−4840.

(20) Kogelberg, H.; Piskarev, V. E.; Zhang, Y.; Lawson, A. M.; Chai, W. *Eur. J. Biochem.* **2004**, *271* (6), 1172−1186.

(21) Thurl, S.; Munzert, M.; Boehm, G.; Matthews, C.; Stahl, B. *Nutr. Rev.* **2017**, *75* (11), 920−933.

(22) Gu, H.; Deng, Y.; Wang, J.; Aubry, A.-F.; Arnold, M. E. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2010**, *878* (25), 2319−2326.

(23) Wu, S.; Grimm, R.; German, J. B.; Lebrilla, C. B. *J. Proteome Res.* **2011**, *10* (2), 856−868.

(24) Montgomery, A. P.; Xiao, K.; Wang, X.; Skropeta, D.; Yu, H. Chapter Two - Computational Glycobiology: Mechanistic Studies of Carbohydrate-Active Enzymes and Implication for Inhibitor Design. In *Advances in Protein Chemistry and Structural Biology*; Karabencheva-Christova, T., Ed.; Academic Press: New York, 2017; Vol. *109*, pp 25−76.

(25) Wallace, W. E.; Ji, W.; Tchekhovskoi, D. V.; Phinney, K. W.; Stein, S. E. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (4), 733−738.

(26) Moorthy, A. S.; Wallace, W. E.; Kearsley, A. J.; Tchekhovskoi, D. V.; Stein, S. E. *Anal. Chem.* **2017**, *89* (24), 13261−13268.

(27) Wu, S.; Tao, N.; German, J. B.; Grimm, R.; Lebrilla, C. B. *J. Proteome Res.* **2010**, *9* (8), 4138−4151.

(28) Shen, Z.; Warren, C. D.; Newburg, D. S. *Anal. Biochem.* **2000**, *279* (1), 37−45.

(29) Naarding, M. A.; Ludwig, I. S.; Groot, F.; Berkhout, B.; Geijtenbeek, T. B. H.; Pollakis, G.; Paxton, W. A. *J. Clin. Invest.* **2005**, *115* (11), 3256−3264.

DOI: 10.1021/acs.analchem.8b01176
Anal. Chem. 2018, 90, 8977−8988

(30) Zivkovic, A. M.; German, J. B.; Lebrilla, C. B.; Mills, D. A. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (Supplement 1), 4653−4658.
(31) National Center for Biotechnology Information. *PubChem Compound Database*; CID = 189088; https://pubchem.ncbi.nlm.nih.gov/compound/189088#section=Top; 2018.
(32) Charbonneau, M. R.; O'Donnell, D.; Blanton, L. V.; Totten, S. M.; Davis, J. C. C.; Barratt, M. J.; Cheng, J.; Guruge, J.; Talcott, M.; Bain, J. R.; Muehlbauer, M. J.; Ilkayeva, O.; Wu, C.; Struckmeyer, T.; Barile, D.; Mangani, C.; Jorgensen, J.; Fan, Y.-m.; Maleta, K.; Dewey, K. G.; Ashorn, P.; Newgard, C. B.; Lebrilla, C.; Mills, D. A.; Gordon, J. I. *Cell* **2016**, *164* (5), 859−871.
(33) Bereman, M. S.; Young, D. D.; Deiters, A.; Muddiman, D. C. *J. Proteome Res.* **2009**, *8* (7), 3764−3770.
(34) Remoroza, C. A.; Mak, T.; Stein, S. E. *Milk SRM 1953 Glycan MS Library*; https://chemdata.nist.gov/srmd-1953-glycan/; 2018.

DOI: 10.1021/acs.analchem.8b01176
Anal. Chem. 2018, 90, 8977−8988