Anal. Chem. **1999,** *71,* 3747−3754

Catalog-Library Approach for the Rapid and Sensitive Structural Elucidation of Oligosaccharides

Ken Tseng,[†] Jerry L. Hedrick,[‡] and Carlito B. Lebrilla*,[†]

Department of Chemistry and Section of Molecular and Cellular Biology, University of California, Davis, California 95616

We obtained the nearly complete structural elucidation of oligosaccharide components, including sequence, linkage, and even stereochemistry in the picomolar levels. The "catalog-library" approach is used for elucidating the structures of minor components in a mixture of oligosaccharides. Oligosaccharides released from a family of glycoproteins are often composed of a small finite set of monosaccharides. In this regard, the numerous oligosaccharide species are analogous to the products found in syntheses involving combinatorial libraries. The great structural diversity in the library is the result of the nearly infinite combinations in which even a small number of monosaccharides can be arranged. Fortunately, structural similarities exist between different oligosaccharides, as specific substructural motifs are preserved among different compounds. We propose that a catalog of substructural motifs can be identified and characterized by collision-induced dissociation mass spectrometry. The catalog is constructed from a set of known compounds that have been fully structurally elucidated by, for example, nuclear magnetic resonance. The catalog consists of the characteristic fragmentation patterns belonging to a set of specific substructural motifs. Collision-induced dissociation is used to determine the presence of these motifs and reconstruct the structures of less abundant components.

The ability to accurately sequence oligosaccharides has revealed the remarkable structural complexity and functional diversity of these molecules.[1−3] Oligosaccharides are involved in a host of biological functions including cell−cell and cell−matrix recognition, hormonal actions, inter- and intracellular trafficking, and protection.[4] However, unlike DNA and proteins where sequence provides nearly all the primary structure, oligosaccharides are characterized by their sequence, linkage, and stereochemistry. Additionally, the large diversity in the monosaccharides due to chemical modification and isomerism, the labile nature of the glycosidic bonds, and the poor intrinsic basicity all combine to make the structural elucidation of oligosaccharides significantly more difficult than that of other biopolymers. There is currently no analogous method for determining oligosaccharide structures with the sensitivity, reliability, and accuracy of the Edman degradation for peptides.

The lack of a rapid method for the complete structural elucidation of oligosaccharides remains a major barrier for understanding structure−function relationships of this important class of compounds. The structural elucidation of oligosaccharides is a difficult task even under the best circumstances. There is a wide array of tools ranging from chemical to spectroscopic that provide structural determination, but often these methods require long and tedious separation procedures while still limited in sensitivity. The complete structural elucidation of complex oligosaccharides requires minimally micromoles of material when often only picomoles or less are available. This is further complicated by the heterogeneity of the compounds as their release from a single glycoprotein produces a collection of compounds. We will call this collection of compounds a library since the collection is composed of combinations involving a small finite set of residues. Often, it is possible to determine the structures of the more abundant components but the structure of the minor components remain somewhat inaccessible.

We propose an approach to the structural elucidation of minor components in oligosaccharide libraries based on the construction of a catalog of specific substructural motifs that can be used to rebuild the total structure of the unknown components. The method begins with the release (or synthesis) of the library. The structures of the most abundant components are then determined by a combination of methods including chemical degradation, nuclear magnetic resonance, and mass spectrometry. This represents the most difficult and time-consuming step. The components whose structures have been elucidated are characterized further by collision-induced dissociation (CID) mass spectrometry. CID, or tandem MS, has been shown to be highly sensitive to both linkage and sequence in oligosaccharides. For example, König and Leary showed that isomers of lacto-*N*-fucopentaoses produce distinct fragmentation patterns when coordinated to cobalt.[5] Similar observations were reported by Viseux et al.[6] and Weiskopf et al.[7] during the CID of permethylated oligosaccharides.

[†] Department of Chemistry.

[‡] Section of Molecular and Cellular Biology.

(1) Ivatt, R. J. *The Biology of Glycoproteins*; Plenum Press: New York, 1984.

(2) Schauer, R. *Sialic Acids. Chemistry, Metabolism and Function*; Springer-Verlag: New York, 1982.

(3) Lennarz, W. J. *The Biochemistry of Glycoproteins and Proteoglycans*; Plenum Press: New York, 1980.

(4) Varki, A. *Glycobiology* **1993**, *3*, 97−130.

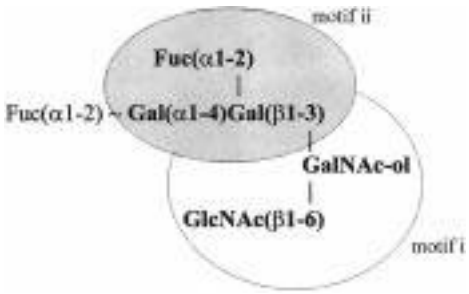
(5) König, S.; Leary, J. A. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 1125−34.

(6) Viseux, N.; Hoffmann, E. D.; Domon, B. *Anal. Chem.* **1998**, *70*, 4951−9.

(7) Weiskopf, A. S.; Vouros, P.; Harvey, D. J. *Anal. Chem.* **1998**, *70*, 4441−7.

10.1021/ac990095r CCC: $18.00
Published on Web 07/13/1999

Scheme 1. Example of an Oligosaccharide with Two Substructural Motifs Found in the Catalog



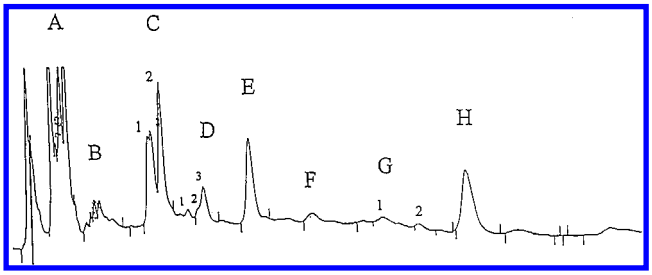


Figure 1. HPLC chromatograph of neutral oligosaccharides isolated from the egg jelly of *Xenopus laevis*. See text for details regarding conditions.

Viseux et al.[6] further used the CID spectra of a set of reference compounds to elucidate the structure of unknown components.

Substructural motifs that are found in several oligosaccharide structures are identified by their distinct fragmentation patterns. The motifs are cataloged according to their CID fragmentation patterns. Similar CID experiments are then performed on the unknown components. The CID mass spectrum of the unknown is evaluated for the presence of patterns that are similar to those in the catalog. By combining motifs found in the unknown, the structure is elucidated. For example in the structure presented in Scheme 1, the compound is composed of two motifs that are all identified by CID.

It must be emphasized that this method is not the same as the "fingerprinting" method commonly used in, for example, gas chromatography/mass spectrometry (GC/MS) analysis where an unknown compound is compared to a vast library to obtain its identity. Instead, the catalog-library method is used to determine structures of totally unknown compounds. This approach has several advantages. It is nonlinear in that as the catalog becomes larger, it becomes easier to characterize new structures. Structural determination is performed on the mass spectrometry scale requiring picomoles or less of material compared to micro- or nanomoles for NMR.[8−15] The procedure is relatively fast, requiring only a few minutes compared to the days and months of traditional techniques.

This approach is illustrated with the neutral oligosaccharides found in the jelly coat of the South African toad *Xenopus laevis*. Their structures play a key role in fertilization. The O-linked oligosaccharides are released from the egg jelly using alkaline sodium borohydride. A representative HPLC chromatogram of the neutral oligosaccharides is shown in Figure 1.[16] The most abundant components have been characterized by Strecker et al. using NMR.[17] The structures of 12 oligosaccharides have been elucidated (Chart 1); their mass assignments and location in the HPLC trace are shown in Table 1. To obtain the structures by NMR, the mixtures were further separated by an additional HPLC procedure. For example, peak C1 contains four major components that were separated to perform the NMR analysis. We have confirmed these assignments with mass spectrometry by obtaining the accurate masses. For mass spectrometry, each species is readily mass resolved and identified by collision-induced dissociation (CID). In addition to the 12, we have identified 7 other oligosaccharides whose abundances were too small to be analyzed by NMR. The masses and compositions of the unknown components are also given in Table 2. The structures of these compounds will be elucidated using only mass spectrometry.

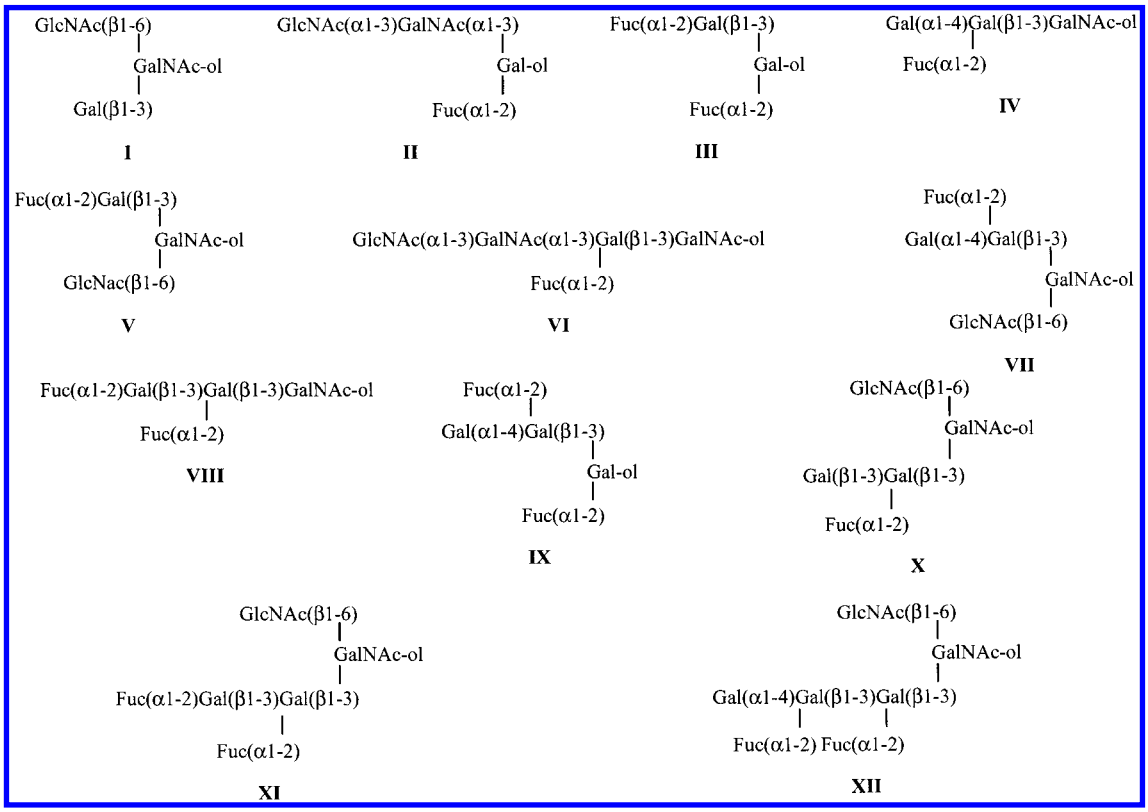
EXPERIMENTAL SECTION

Methods for the release and the purification of the oligosaccharides are described in detail in an earlier publication.[16] Briefly, approximately 348 mg of lyophilized egg jelly was treated with a solution of 100 mM NaOH and 1.0 M $NaBH_4$ at 57 °C for 23 h. Hydrogen-form Dowex 50 was added to the solution until bubbles ceased forming and the reaction stopped. The solution was filtered and adjusted to pH 6.5 with a 0.1 M NaOH aqueous solution. The volume was reduced, and methanol distillation was performed five times to remove boric acid. A Bio-Gel P2 column was used to desalt the solution. The collected fractions were tested for oligosaccharides by the phenol−sulfuric acid assay using galactose as the standard. The combined fractions was treated with COO^--form Dowex to remove acidic oligosaccharide alditols and filtered. It was found that the remaining neutral still contained excessive amounts of boric acid. Removal of residual boric acid is important for obtaining clean mass spectra. Borates complex readily to oligosaccharides and to the matrix, producing a series of borate complexes in the mass spectra. To eliminate the borate, methanol distillation was performed a total of seven times. The purified mixture of the neutral oligosaccharide alditols was separated by HPLC on a primary amine-bonded silica (25.0 cm; Supelco, Bellefonte, PA) using a 75:25 acetonitrile:water mixture. A total of 70 fractions were collected and analyzed by MALDI−FTMS.

The external source MALDI−FTMS (IonSpec Corp., Irvine, CA) used in this study has also been described.[13,14] The instrument has a 4.7 T superconducting magnet and a nitrogen laser operating at 337 nm. The sample was prepared by concentrating 20 μL of the sample on the probe tip with 2- or 3-μL aliquots. One microliter of a 0.01 M NaCl was added to the probe tip to enrich the Na^+

(8) Harvey, D. J.; Rudd, P. M.; Bateman, R. H.; Bordoli, R. S.; Howes, K.; Hoyes, J. B.; Vickers, R. G. *Org. Mass Spectrom.* **1994**, *29*, 753−66.
(9) Harvey, D. J.; Bateman, R. H.; Green, M. R. *J. Mass Spectrom.* **1997**, *32*, 167−87.
(10) Naven, T. J. P.; Harvey, D. J.; Brown, J.; Critchley, G. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1681−6.
(11) Reinhold: V. N.; Reinhold: B. B.; Costello, C. E. *Anal. Chem.* **1995**, *67*, 1772−84.
(12) Juhasz, P.; Costello, C. E. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 785−96.
(13) Penn, S. G.; Cancilla, M. T.; Lebrilla, C. B. *Anal. Chem.* **1996**, *68*, 2331−9.
(14) Cancilla, M. T.; Penn, S. G.; Lebrilla, C. B. *Anal. Chem.* **1998**, *70*, 663−72.
(15) Cancilla, M. T.; Penn, S. G.; Carroll, J. A.; Lebrilla, C. B. *J. Am. Chem. Soc.* **1996**, *118*, 6736−45.
(16) Tseng, K.; Lindsay, L. L.; Penn, S. G.; Hedrick, J. L.; Lebrilla, C. B. *Anal. Biochem.* **1997**, *250*, 18−28.
(17) Strecker, G.; Wieruszeski, J.-M.; Plancke, Y.; Boilly, B. *Glycobiology* **1995**, *5*, 137−46.

Chart 1. Structures of 12 Known Neutral Oligosaccharides Released from the Jelly Coats of the Eggs of *Xenopus laevis*[a]



[a] The structures were determined by NMR. See ref 14.

Table 1. HPLC Location with Theoretical and Experimental *m/z* Values of 12 Known Compounds Determined by NMR[a]

structure	HPLC location	(*m/z*) theoret	(*m/z*) exptl	structure	HPLC location	(*m/z*) theoret	(*m/z*) exptl
I	C1	611.23	611.23	**VII**	E	919.34	919.34
II	C1	757.29	757.29	**VIII**	E	862.32	862.32
III	C1	659.24	659.24	**IX**	E	821.29	821.30
IV	C1	716.26	716.26	**X**	F	919.34	919.35
V	C2	757.29	757.29	**XI**	H	1065.40	1065.40
VI	D2	960.37	960.37	**XII**	not shown	1227.46	1227.46

[a] The masses correspond to $[M + Na]^+$.

Table 2. Seven Unknown Neutral Oligosaccharides Found in the Alditol-Released Mixtures of the Egg Jelly Coats of *Xenopus laevis*[a]

unknown	HPLC location	(*m/z*) theoret	(*m/z*) exptl	composition Hex:HexNAc:Fuc
1	D1	903.35	903.34	1:2:2
2	D2	903.35	903.34	1:2:2
3	F	1065.40	1065.40	2:2:2
4	G1	919.34	919.34	2:2:1
5	G1	1024.36	1024.37	3:1:2
6	G1	1389.48	1389.50	4:2:2
7	G2	1065.40	1065.40	2:2:2

[a] The masses correspond to $[M + Na]^+$.

concentration and produce primarily sodiated species. One microliter of 0.4 M 2,5-dihydroxybenzoic acid (DHB) was added as the matrix.

All collision-induced dissociation experiments were performed in the off-resonance mode.[13,18] The conditions were optimized to produce the greatest amount of fragment ions. Two argon pulses were used during the CID event to maintain a base pressure of 10^{-5} Torr. The ions were excited +1000 Hz of their cyclotron frequency for 1000 ms at 2−7 V (base to peak). The fragmentation patterns were readily identifiable with this range of amplitudes.

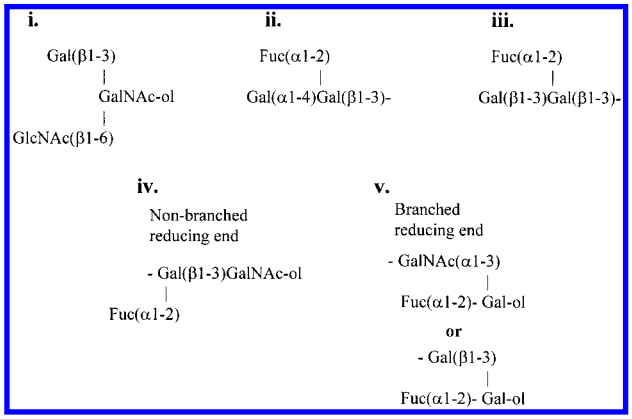
RESULTS

The neutral oligosaccharides released by the method described above were all O-linked. A library consisting of at least 19 neutral oligosaccharides was obtained. As discussed previously, the 12 most abundant compounds have been isolated and characterized using NMR by Strecker et al.[17] The mass assignments of the 12 structures were recently published by Tseng et al.[16] The typical mass spectra of compounds isolated by HPLC were also provided in the previous publication.

The CID of all 12 known compounds was performed to construct the catalog. Two stages of MS were performed on all 12 compounds with some requiring MS^3 or MS^4. From the 12 known structures, five substructural motifs were identified (Chart 2). The group of five motifs makes up the entire catalog. The description of each motif and its characteristic fragmentation follows. For clarity, the known compounds are designated by capital Roman numerals while the structural motifs are designated by small Roman numerals. The unknown compounds are designated by Arabic numbers.

(18) Gauthier, J. W.; Trautman, T. R.; Jacobson, D. B. *Anal. Chim. Acta* **1991**, *246*, 211−25.

Chart 2. Five Structural Motifs Identified by Mass Spectrometry from the 12 Known Structures of Neutral Oligosaccharides Found in the Jelly Coats of the Eggs of *Xenopus laevis*



Identification of Substructural Motifs from the Collision-Induced Dissociation of Compounds I−XII. The entire compound **I** is a substructural motif (motif **i**, Chart 2) found in several other compounds. This trisaccharide structure, consisting of a Gal-(β1−3) and a GlcNAc(β1−6) linked to a reduced GalNAc, was incorporated into several larger compounds, including **V**, **VII**, **X**, **XI**, and **XII**. The MALDI−FTMS spectrum of **I** showed only the sodium-coordinated parent as the quasimolecular ion (spectrum not shown). To ensure a single quasimolecular ion, all samples were doped with NaCl. The CID spectrum of **I** is shown in Figure 2a. Note that the motifs will be in bold and in larger fonts in the inset structures of this section. To produce this spectrum, the sodiated molecular ion was isolated using an arbitrary waveform generator. The group of fragment ions between *m*/*z* 350 and *m*/*z* 450 is unique to this motif and was found with other compounds that contained this motif. The fragments, in order of decreasing ion abundance with their corresponding cleavages, are *m*/*z* 408 ($Y_{1\alpha}$), 449 ($Y_{1\beta}$), 431 ($Z_{1\beta}$), 390 ($Z_{1\alpha}$), and 413 ($Z_{1\beta}$ − H_2O). An additional fragment, the smallest fragment at *m*/*z* 228, corresponds to [GalNAc-ol − H_2O + Na]$^+$, which was the reducing terminus converted to the alditol. In general, the alditol residue was the smallest fragment ion obtained for most of the neutral alditols, indicating that the alkali metal ions were strongly bound to this residue. This is not observed with oligosaccharides, where the reducing end remains a pyranose.[13] Apparently, the open reduced saccharide is a stronger chelator of the alkali metal than the closed pyranose precursor.

CID of the larger compound containing motif **i**, for example structure **VII**, showed a similar fragmentation pattern in the same mass range. The CID spectrum of **VII** (isolated) is shown in Figure 2b. At the higher mass, fragments corresponding to loss of fucose (*m*/*z* 773) and loss of the (α1−4)-linked Gal (*m*/*z* 757) were observed. Below *m*/*z* 500, the fragmentation pattern matches closely that of compound **I**. The fragment at *m*/*z* 228 was observed as before along with the distinct pattern described above between *m*/*z* 350 and 450 (insert). CID performed on other structures that contain motif **i**, including **V**, **X**, **XI**, and **XII**, showed this distinctive fragmentation pattern with only slight variations in the relative intensity. For example with compound **V** *m*/*z* 390 was slightly more abundant than *m*/*z* 413 while, for **X**, *m*/*z* 449 was the most abundant peak.

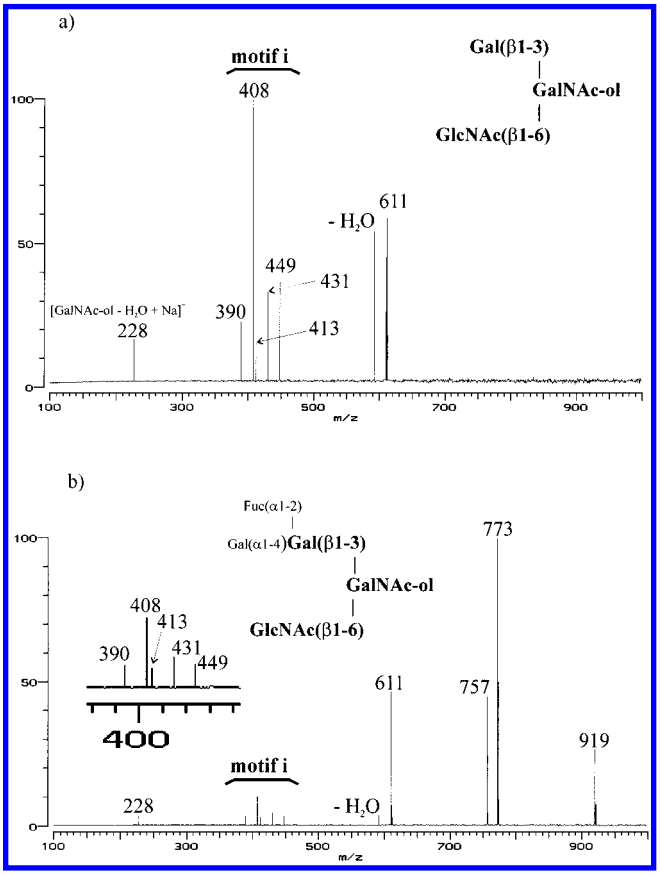


Figure 2. (a) The CID−FTMS spectrum of compound **I** ([M + Na]$^+$, *m*/*z* 611). The fragmentation pattern particularly between *m*/*z* 350 and 450 designates the pattern of motif **i**. (b) The CID−FTMS spectrum of compound **VII** ([M + Na]$^+$, *m*/*z* 919) which contains motif **i** (bold in inset structure). This spectrum illustrates how the fragmentation pattern of the motif is preserved in different compounds.

The second and third substructural motifs (**ii** and **iii**, Chart 2) consisted of isomeric trisaccharide units that differed by only one linkage. Motif **ii** consisted of a Fuc(α1−2) and with a Gal(α1−4) connected to a galactose. In motif **iii**, the Gal(α1−4) was replaced by Gal(β1−3). Motif **ii** is found in structures **VII**, **IX**, and **XII**. The CID spectrum of **VII** is shown again in Figure 3a with the peaks labeled to designate the fragments associated with motif **ii**. The characteristic fragmentation of motif **ii** involved losses of Fuc (*m*/*z* 773) and Gal (*m*/*z* 757) from the quasimolecular ion with the loss of Gal half as intense as that of Fuc. The losses of a Fuc and a Gal from the quasimolecular ion combine to yield an abundant peak at *m*/*z* 611. The fragmentation pattern of motif **iii** was considerably different from that of motif **ii** despite the difference of only a single linkage in the entire molecule. Motif **iii**, found in structures **X**, **XI**, and **XII**, produced the loss of Fuc (*m*/*z* 773) primarily from the quasimolecular ion (*m*/*z* 919) (structure **X**, Figure 3b). The loss of Gal from the quasimolecular ion was not observed. However, the combination of losses of a single Gal and a Fuc was similarly intense (*m*/*z* 611) as were the losses of two Gal and a Fuc (*m*/*z* 449). We note that the Gal(β1−3)Gal was a particularly strong linkage and did not readily dissociate before fucose under CID conditions. In Figure 3b, the loss of two Gal groups from *m*/*z* 773 produced the most abundant fragment with *m*/*z* 449. With other compounds containing the Gal(β1−3)Gal unit, e.g., **XI** and **XII**, the loss of one Gal (*m*/*z* 611)

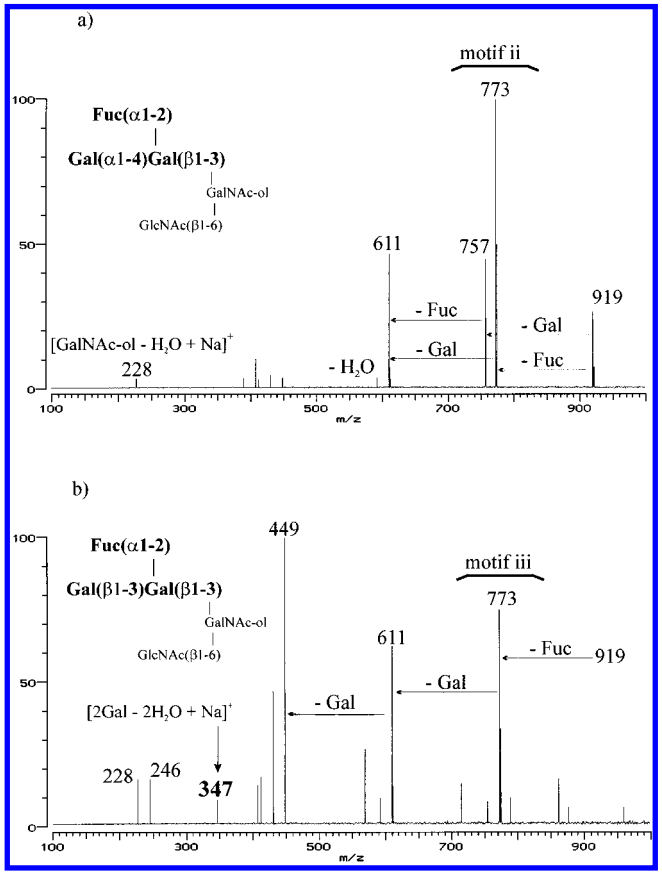


Figure 3. (a) Replot of Figure 2b ([M+Na]$^+$, *m*/*z* 919) with the fragments and the residue loss corresponding to motif **ii**. (b) The CID−FTMS spectrum of compound **X** ([M + Na]$^+$, *m*/*z* 919). The sequential losses of Fuc, Gal, and Gal are characteristic of motif **iii**. Note also the presence of the disaccharide fragment with *m*/*z* 347 corresponding to the Gal(β1−3)Gal fragment. We find this combination is particularly stable. The *m*/*z* 611 fragment, corresponding to the core trisaccharide, is observed only at high collision energies.

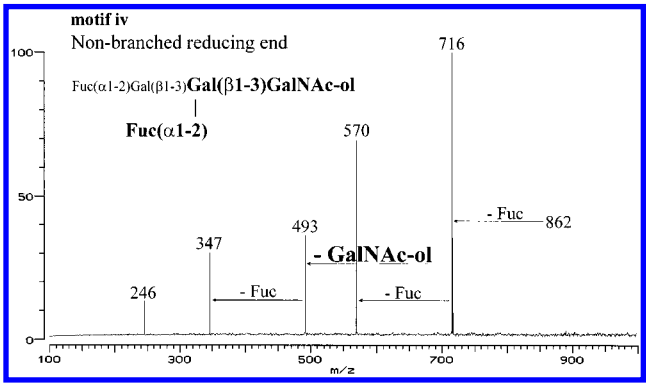


Figure 4. CID−FTMS spectrum of compound **VIII** ([M + Na]$^+$, *m*/*z* 862). The loss of the alditol group (GalNAc-ol) from the reducing end is characteristic of motif **iv**—a nonbranched reducing end. See text for additional details.

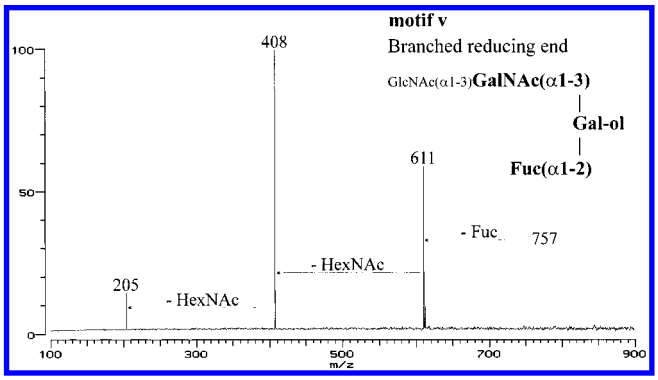


Figure 5. CID−FTMS (MS^2) spectrum of compound **II** ([M + Na]$^+$, *m*/*z* 757). This structure contains motif **v**. The absence of the alditol loss even at high translational energies is characteristic of motif **v**—a branched reducing end. In general, branched reducing ends do not produce loss of alditol residues as fragments; see, for example, Figure 3a. See also text for details.

was only a minor component while the loss of two Gal units (*m*/*z* 449) was dominant. The presence of *m*/*z* 347 in Figure 3b further supported this notion. This disaccharide fragment was composed of a sodiated Gal(β1−3)Gal species. The large differences in the fragmentation pattern between motifs **ii** and **iii** was important and illustrated that even variations in a single linkage were differentiated using this method.

The fourth and fifth substructural motifs (**iv** and **v**) refer to the branching at the reducing-end residue. A nonbranched reducing-end residue (motif **iv**) was found in **IV**, **VI**, and **VIII**. In each case, the loss of reducing end (GalNAc-ol) was observed in the CID spectrum as the ion [M + Na − GalNAc-ol]$^+$. Figure 4 shows the MS/MS spectrum of **VIII**. Loss of Fuc (*m*/*z* 716) was the first fragment from the quasimolecular ion followed by either an additional loss of fucose (*m*/*z* 570) or a loss of GalNAc-ol (*m*/*z* 493). In structures where the reducing end was branched, loss of the reducing end residue was not observed. For example, a compound that contained motif **i** (e.g., Figure 2a) did not produce the loss of GalNAc-ol even under the most vigorous (high rf amplitude) CID conditions. The absence of the loss of GalNAc-ol in **VII** suggested again that the alkali metal ion is strongly bound to the branched reducing end, sequestering the charge at that position. In contrast, the loss of GalNAc-ol in **VIII**, which has a nonbranched reducing end, suggested that Na$^+$ was not sequestered there, allowing the residue to be lost as a neutral fragment. Another type of branching at the reducing end was designated as motif **v**, where one of the branching saccharides is a Fuc(α1−2). This motif was found in structures **II**, **III**, and **IX**. This motif was differentiated from motif **i** by the lack of a satellite peak shifted 18 mass units lower that accompanied several of the major peaks in the mass spectrum. For example, Figure 5 shows the CID spectrum of compound **II**, which contained motif **v**. Neither loss of the reducing end Gal-ol nor the mass shift of 18 mass units with the major glycosidic bond cleavage fragments was observed in the mass spectra. The shift of 18 mass units corresponded to a fragmentation pathway that included the glycosidic oxygen in the fragment lost. The two fragmentation pathways were designated by B- and C-cleavages according to the Domon−Costello formalism.[19] In summary, the presence of the loss of the alditol means a nonbranched reducing end.

Structural Elucidation of Unknown Oligosaccharide Alditols Using the Catalog. Seven additional neutral oligosaccharides were found with mass spectrometry that could not be analyzed by NMR. The location in the HPLC trace, theoretical mass, experimental mass, and monosaccharide compositions are provided in Table 2. The compositions were obtained strictly from

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

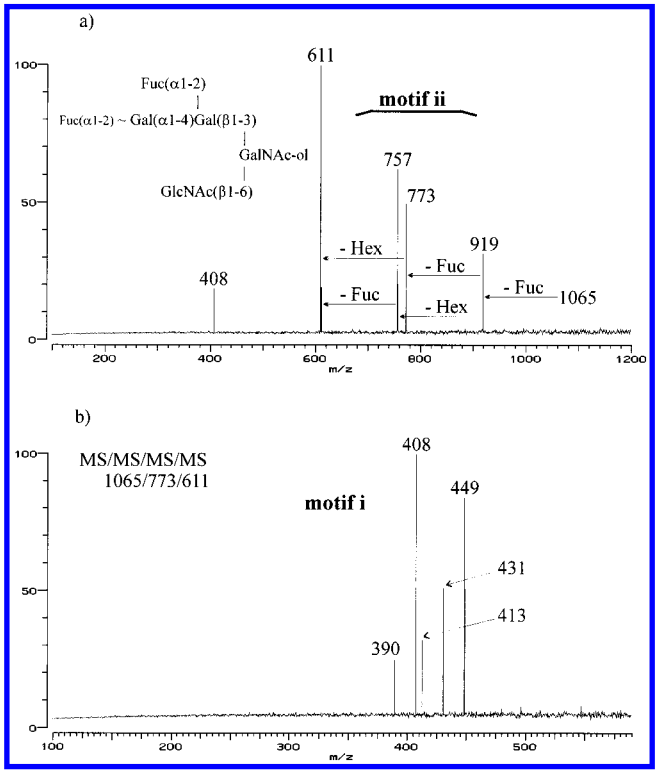


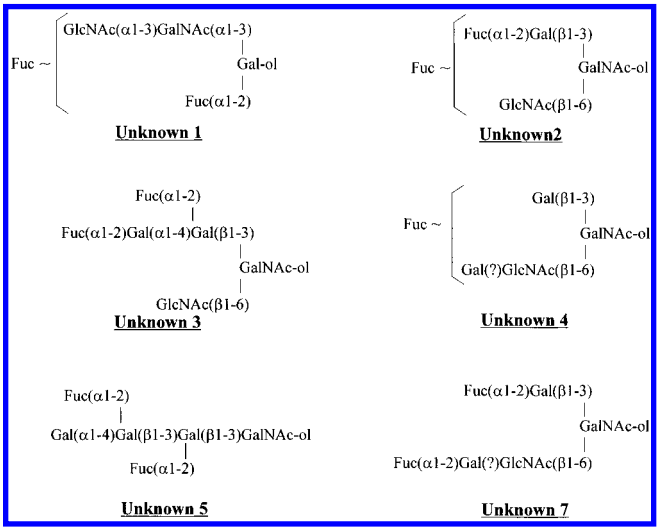
Figure 6. (a) CID (MS^2) spectrum of unknown **3** ($[M + Na]^+$, *m*/*z* 1065). The fragmentation pattern observed corresponds to motif **ii**. (b) PCID (MS^4, *m*/*z* 1065/757/611) spectrum of unknown **3**. The fragmentation pattern observed corresponds to motif **i**. The two spectra yield the structure shown. Note that this compound is also illustrated in Scheme 1.

the experimental molecular weight. The exact identity of the hexoses (Hex) and the *N*-acetylhexoses (HexNAc) cannot be determined by the molecular weight alone. However, on the basis of the 12 known structures, the hexoses are likely to be galactose while the *N*-acetylhexosamines are either *N*-acetylgalactosamines (GalNAc) or *N*-acetylglucosamines (GlcNAc). The discussion below will begin with the most abundant compound and end with the least abundant.

The unknown oligosaccharide **3** (peak F, Figure 1) had an *m*/*z* of 1065.40 which corresponded to (accounting for Na^+) two hexoses, two fucoses, and two *N*-acetylhexosamines. The MS/MS spectrum (Figure 6a) showed an initial loss of one fucose to yield *m*/*z* 919 and additional losses of a hexose (Hex) and a fucose (*m*/*z* 757 and 773, respectively). Both ionic species in turn lose a Fuc and a Hex, respectively, to produce a single product at *m*/*z* 611. Note the strong similarities between Figure 6a and Figure 3a. The major products were identical in both spectra. The major difference is found in the relative intensities of *m*/*z* 611 which is larger in Figure 6a than in Figure 2. Nonetheless, the fragmentation pattern is easily recognizable and corresponds to motif **ii**.

To obtain the structural motifs corresponding to the remainder of the molecule, an MS^4 experiment was performed by following the sequence *m*/*z* 1065 → 773 → 611. This method involves the sequential excitation of three frequencies corresponding to *m*/*z* 1065, 773, and 611. The purpose of this technique is to enhance fragments smaller than *m*/*z* 611 by driving the CID of *m*/*z* 1065, 773, and 611. During the CID event, no isolation was performed as the other products did not interfere with the analysis.

Chart 3. Proposed Structures of Six of Seven Unknown Neutral Oligosaccharides Determined by the Catalog-Library Approach[a]



[a] The compounds were released from the egg jelly coats of *Xenopus laevis*.

Furthermore, ions with masses far away from 1065, 773, and 619 were generally unaffected by the excitation frequencies so that the fragmentation was due solely to the CID of the three ionic species. This method allowed us to perform multiple stages of tandem MS without the significant ion loss that usually accompanies ion isolation. This approach, which we will refer to as "progressive CID" (or PCID), was used throughout this work to obtain MS^n where $n > 2$.

The PCID ($n = 4$, or MS^4) of unknown **3** is shown in Figure 6b. The fragmentation pattern was easily identified as that corresponding to motif **i**. The relative intensities of the ions were nearly identical (compare Figures 2a and 6b) with the *m*/*z* 449 peak slightly greater for unknown **3**. Combining motif **i** and **ii** produced a structure that was one fucose less than the mass of unknown **3** (see also Scheme 1). The proposed structure is depicted in the inset of Figure 6 and in Chart 3. Note that the proposed structures are inset in all the figures in this section. The only ambiguity was the position of the second fucose; however, it was assigned to the terminal galactose because of the absence of the loss of a Hex unit from the quasimolecular ion. This analysis required less than 10 pmoles material and less than 1 h to perform.

Unknowns **1** and **2** were found in adjacent peaks of the HPLC trace, with unknown **1** eluting slightly ahead of unknown **2** (peaks D1 and D2, Figure 1). The proximity of the two peaks and the identical masses could easily be interpreted to mean that the two compounds were identical. However, CID clearly showed that the two are distinct isomers (Figures 7 and 8a, respectively). The smallest fragment obtained for unknown **1** (*m*/*z* 205) corresponded to a hexose alditol, while for unknown **2** (*m*/*z* 228) it corresponded to an *N*-acetylhexose alditol. The MS/MS spectrum of unknown **1** (Figure 7) matched that of motif **v** and specifically structure **II** (Figure 5) with the exception of an additional fucose. It shows not only that the Fuc was attached to the reducing end but that the rest of the molecule was a GlcNAc(α1−3)GalNAc-(α1−3) combination (Chart 3). The spectrum further suggested

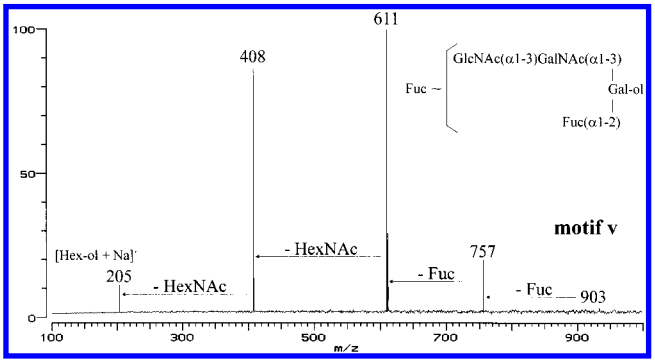


Figure 7. CID−FTMS spectrum of unknown **1** ([M + Na]$^+$, *m*/*z* 903). The fragment *m*/*z* 205 yields the reducing end as a reduced hexose. The fragmentation pattern is consistent with motif **v**. The spectrum yields the structures shown. Only the connectivity of the second fucose is undetermined.

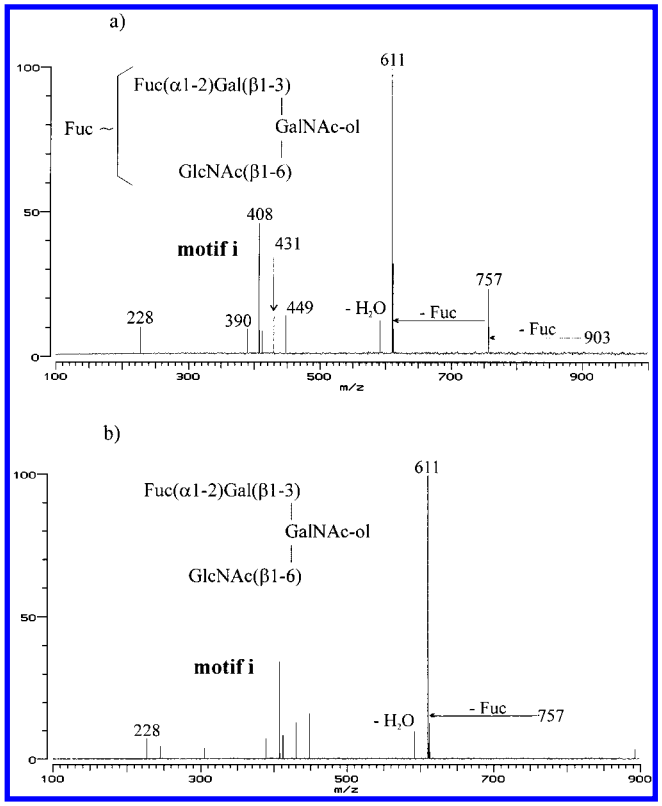


Figure 8. (a) CID−FTMS (MS^2) spectrum of unknown **2** ([M + Na]$^+$, *m*/*z* 903). The fragmentation pattern of motif **i** is observed. (b) CID−FTMS (MS^2) spectrum of compound **V** ([M + Na]$^+$, *m*/*z* 757).

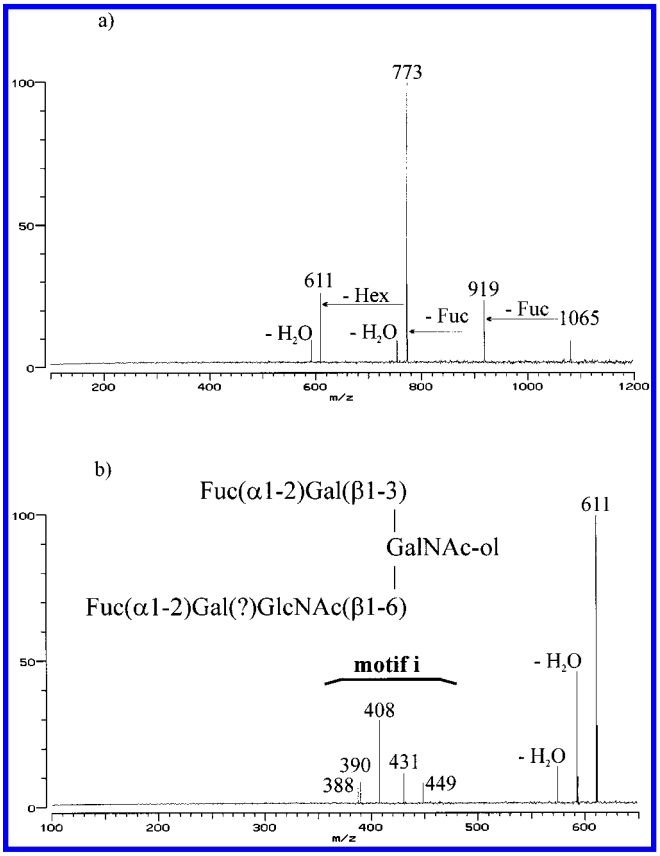


Figure 9. (a) CID−FTMS (MS^2) spectrum of unknown **7** ([M + Na]$^+$, *m*/*z* 1065). (b) CID (MS^4) spectrum of unknown **7**. The two spectra are consistent with the structure shown. The fragment at *m*/*z* 388 corresponds to a Hex−HexNAc unit.

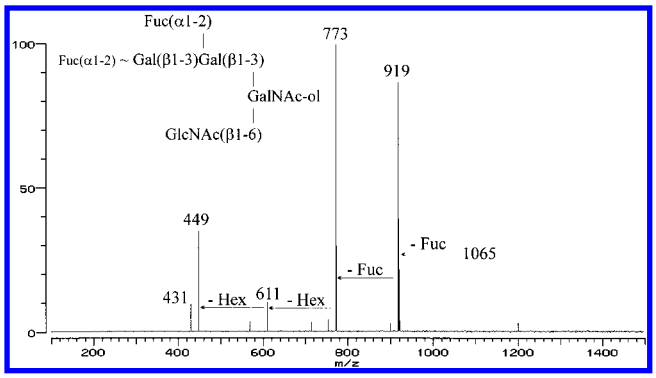


Figure 10. CID−FTMS (MS^2) spectrum of compound **XI** ([M + Na]$^+$, *m*/*z* 1065).

that unknown **1** and structure **II** were similar compounds with an additional fucose on unknown **1**. On the other hand, the CID of unknown **2** was consistent with motif **i**. Unknown **2** had two more fucose than **I** and one more than **V**. Indeed, the comparison of the CID spectra of unknown **2** and compound **V** (Figure 8b) showed distinct similarities between the two compounds. The proposed structure of unknown **2** is provided (inset and Chart 3); we propose that unknown **2** and compound **V** have analogous structures. The position of the second fucose could not be readily determined. We estimate that the total amount of material used for the analysis is less than 10 pmol.

Unknown **7** was composed of two Hex, two HexNAc, and two Fuc units (peak G2, Figure 1). The MS/MS spectrum was consistent with the loss of two Fuc units and a Hex unit to yield *m*/*z* 611 (Figure 9a). This compound is isomeric with structure **XI**, whose MS/MS spectrum is shown in Figure 10, and unknown **3** (Figure 6), but comparison of the MS/MS spectra indicates that it does not share the same Gal(α1−4)Gal or Gal(*β*1−3)Gal substructures found in unknown **3** and structure **XI**, respectively. The MS^4 spectrum (Figure 9b), using the path *m*/*z* 1065/919/773, yielded a familiar pattern corresponding to motif **i** so that the reducing end is readily assigned. We need to account for the additional Hex and the two Fuc units. The fucoses are likely to be linked to the two Hex units as is generally observed in Chart 1. The second Hex may be bound either to the Gal or to the GlcNAc. However, the presence of *m*/*z* 388, which corresponds to a Hex−HexNAc combination, suggests that the second Hex is bound to GlcNAc, as shown in the inset structure.

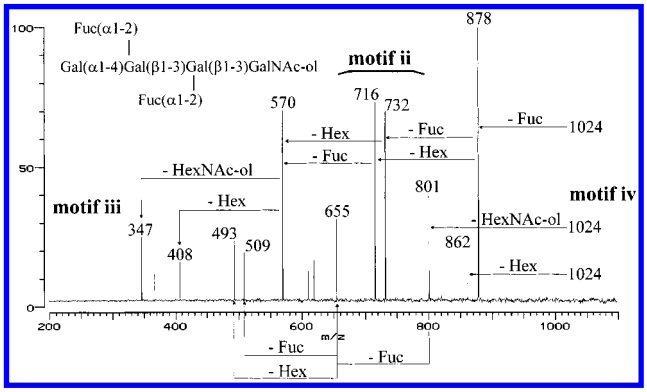


Figure 11. CID−FTMS (MS^2) spectrum of unknown **5** ($[M + Na]^+$, *m*/*z* 1024). The single spectrum yields three substructural motifs consistent with the structure shown.

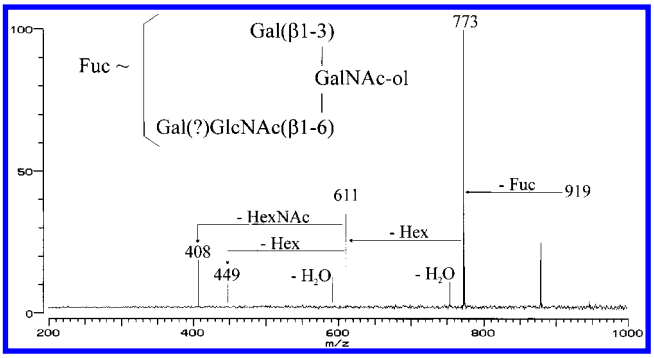


Figure 12. CID−FTMS (MS^2) spectrum of unknown **4** ($[M + Na]^+$, *m*/*z* 919).

Unknowns **4**, **5**, and **6** are the least abundant and all eluted at approximately the same time (G1 in the HPLC, Figure 1). Unknown **5** was the most abundant of the three while unknown **6** was the least abundant. There was not sufficient amount of unknown **6** to perform the analysis. For unknown **5**, loss of Fuc and Hex (*m*/*z* 878 and 862, respectively) from the quasimolecular ion (*m*/*z* 1024) were observed followed by the complementary fragmentation resulting in the total loss of a Fuc and a Hex (*m*/*z* 716) consistent with motif **ii** (Figure 11). The quasimolecular ion also dissociates to lose an *N*-acetylhexose alditol (*m*/*z* 801), denoting the reducing end and consistent with motif **iv**. This was followed by a loss of a Fuc (*m*/*z* 655) and an additional Hex (*m*/*z* 493), consistent with the proposed sequence at the reducing end. The presence of *m*/*z* 347 denoted the Gal(β1−3)Gal structure, consistent with motif **iii**. On the basis of the fragmentation pattern, we constructed the structure for **5** (inset, Figure 11). This structure was analogous to that of the known structure **XII** with the loss of a GlcNAc(β1−6). It had further similarities with structures **VIII** and **IX**.

Unknown **4** was isomeric with structures **VII** and **X** (*m*/*z* 919, quasimolecular ion). The MS/MS spectrum of the quasimolecular ion (Figure 12) showed a loss of Fuc followed by a loss of Hex to produce the peak at *m*/*z* 611. A comparison of Figure 12 with the MS/MS spectrum of unknown **7** (Figure 9a) showed strong similarities, with the exception of the difference of one fucose. Note that the fragments between *m*/*z* 550 and 950 have very similar relative abundances in the two spectra. Fragments at *m*/*z* 408 and 409 in Figure **12** were consistent with motif **i** and were more apparent for unknown **7** (Figure 9b). On the basis of the spectrum and the comparison with unknown **7**, we propose that unknown **4** has the same structure as **7** with the exception of one less fucose (Chart 3). Unknown **4** was mass limited, and MS^n beyond $n = 2$ could not be attempted. For this reason, the peak at *m*/*z* 388, present in Figure 9b, was not observed. There is further an ambiguity regarding the position of the fucose. It may be attached to any of the two galactoses.

CONCLUSION

By constructing a catalog of substructural motifs, we obtained nearly the complete structures of six of seven unknowns. The advantages of this method are readily apparent. It is rapid, requiring only a few minutes to complete a structure once the catalog has been developed, and it requires a small amount of material—only enough to perform CID. It is further a nonlinear approach. The structural elucidation of each species becomes easier as more of the compounds are known and the catalog has more entries. We are currently verifying these structures with exoglycosidases. In general, the catalog-library method can be used for the more effective selection of exoglycosidases.

The general application of this method is aided by the capabilities of the MS analyzer and the unique fragmentation behavior of alditol oligosaccharides. The distinct fragmentation behavior of the alditol oligosaccharides in the cation mode contrasts to that of the nonreduced aldehyde species. The aldehyde oligosaccharides in the cation mode produce alkali metal coordinated species. However, these typically do not produce structurally informative fragment ions under conditions of low-energy CID. When fucoses are present, the major fragments observed are those corresponding to losses of fucoses. No other cleavages are observed until the ion is totally defucosylated. The differences in the fragmentation behavior between the alditol and the aldehyde are due to the coordination of the alkali metal ion. We posit that, in the alditol species, the alkali metal ions are coordinated more strongly by the open alditol species. This sequesters the alkali metal ion in the reducing end, making it more difficult to fragment the ions. The increased activation barrier necessitates higher translational energies for CID, which also allows access to other fragmentation channels.

The ability of MALDI−FTMS to allow multiple stages of MS is essential to produce the necessary fragmentation patterns. In time-of-flight analyzers, tandem MS is still a nontrivial operation. Typically, postsource decay is used to affect fragmentation. Postsource decay may produce the same distinct fragmentation pattern as MALDI−CID; however, this possibility needs to be explored. MALDI with a quadrupole ion trap mass analyzer may also produce similar fragmentation patterns when the method becomes more readily available.

ACKNOWLEDGMENT

Funding Provided by the National Institutes of Health (Grant R01 GM49077-06A1) is gratefully acknowledged.

Received for review January 29, 1999. Accepted May 27, 1999.

AC990095R