Supplementary Information for

De novo structural determination of oligosaccharide isomers in glycosphingolipids using logically derived sequence tandem mass spectrometry

Chia Yen Liew,[1,2,3] Chieh-Kai Chan,[4] Shih-Pei Huang,[1] Yu-Ting Cheng,[4] Shang-Ting Tsai,[1] Hsu Chen Hsu,[1] Cheng-Chung Wang,[4] Chi-Kung Ni*[1,5]

Affiliations

1. Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, 10617, Taiwan.
2. International Graduate Program of Molecular Science and Technology, National Taiwan University (NTU-MST) Taipei, 10617, Taiwan
3. Taiwan International Graduate Program (TIGP) of Molecular Science and Technology (MST), Academia Sinica, Taipei, 10617, Taiwan.
4. Institute of Chemistry, Academia Sinica, Taipei, 11529, Taiwan.
5. Department of Chemistry, National Tsing Hua University, Hsinchu 30013, Taiwan. E-mail: ckni@po.iams.sinica.edu.tw

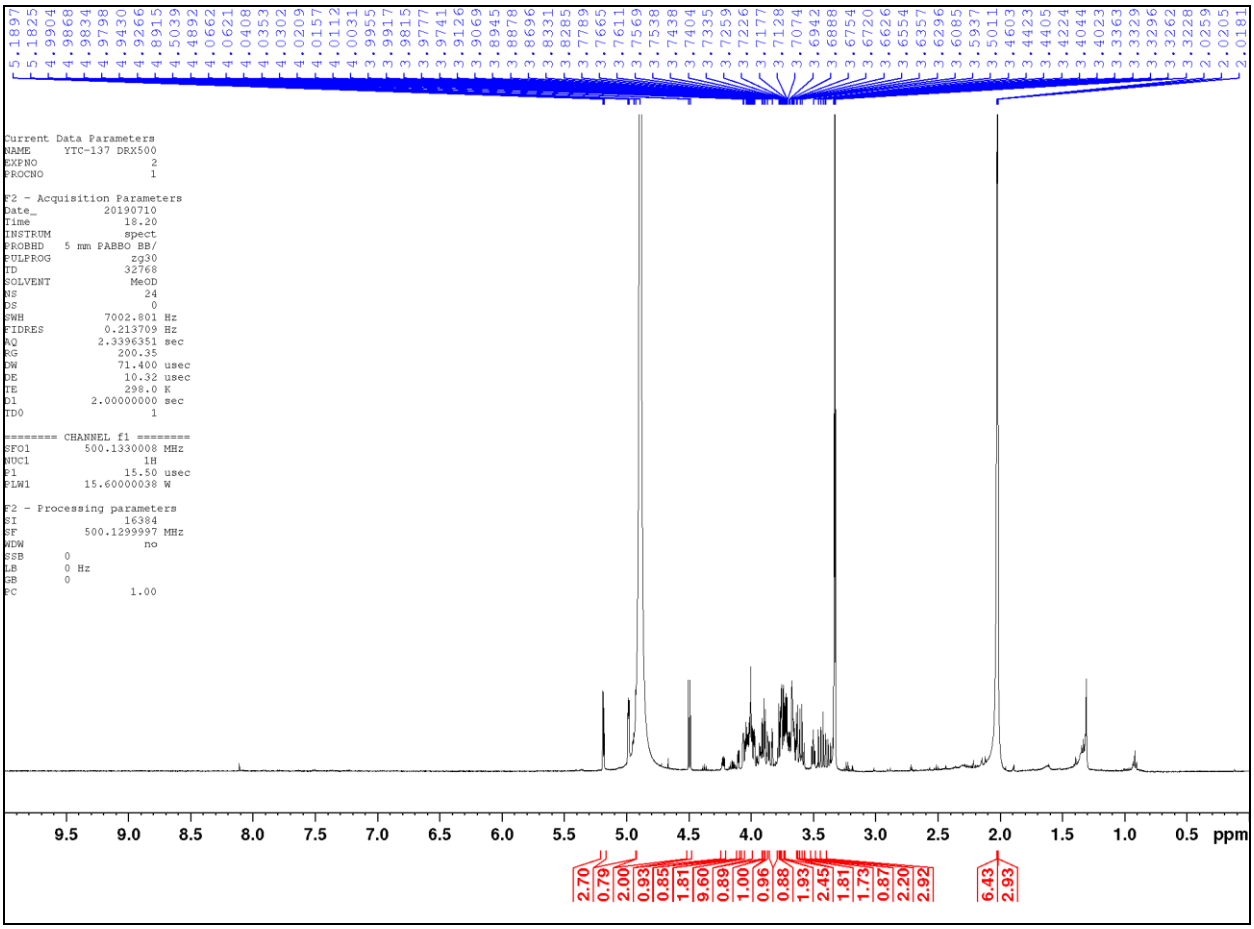

Fig. S1. ^{1}H NMR (500 MHz, MeOD) of disaccharide α-GlcNAc-(1→3)-Gal synthesized in our laboratory. δ 5.19 (d, *J* = 1.9 Hz, 1H, H-1'-α), 5.19 (d, *J* = 3.8 Hz, 1H, H-1'-β), 4.92 (d, *J* = 3.7, 1H, H-1-α), 4.50 (d, *J* = 7.4 Hz, 1H, H-1-β), 4.22 (dd, *J* = 3.3, 5.9 Hz, 1H, H-2'-α), 4.10 (dd, *J* = 2.1, 3.5 Hz, 1H, H-3'-α), 4.06 (dd, *J* = 0.8, 2.6 Hz, 1H, H-2'-β), 4.05-3.96 (m, 4H+2H, H-3'-β, H-4'-β, H-2-β, H-3-β+H-2-α, H-3-α), 3.91 (d, *J* = 3.1 Hz, 1H, H-4-α), 4.89 (d, *J* = 3.4 Hz, 1H, H-4'-α), 3.87-3.83 (m, 4H, H-6a-α, H-6b-α, H-6a'-α, H-6b'-α), 3.79 (d, *J* = 5.1, H-5'-α), 3.77-3.72 (m, 4H, H-6-a-β, H-6-b-β, H-6-a'-β, H-6-b'-β), 3.63 (d, *J* = 3.1 Hz, 1H, H-4-β), 3.58 (d, *J* = 7.6 Hz, 1H, H-5-α), 3.50 (t, 1H, *J* = 6.09 Hz, H-5'-β), 3.42 (t, *J* = 9.4 Hz, 1H, H-5-β), 2.03 (s, 3H, CH_3), 2.01 (s, 3H, CH_3)

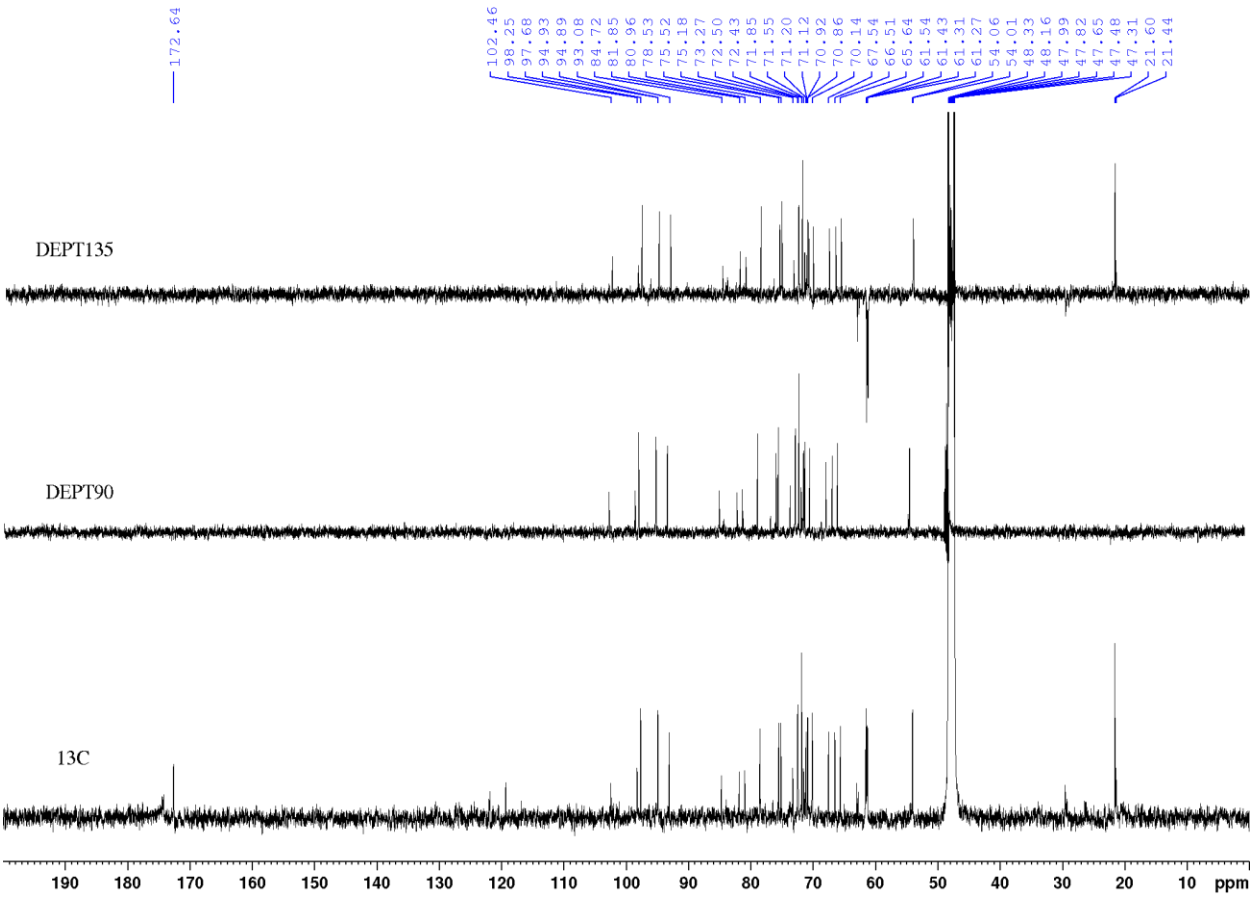

Fig. S2. ^{13}C NMR (125 MHz, MeOD) of disaccharide α-GlcNAc-(1→3)-Gal synthesized in our laboratory. δ 172.64 (C), 102.46 (CH-α), 98.25 (CH-α), 97.68 (CH-β), 94.93 (CH-β), 94.89 (CH-β), 93.08 (CH-β), 84.72 (CH-α), 81.85 (CH-α), 80.96 (CH-α), 78.53 (CH-β), 75.53 (CH-β), 75.18 (CH-β), 73.27 (CH-α), 72.50 (CH-β), 72.43 (CH-β), 71.85 (CH-β), 71.55 (CH-α), 71.20 (CH-α), 71.12 (CH-β), 72.92 (CH-β), 70.86 (CH-β), 70.14 (CH-β), 67.54 (CH-β), 66.51 (CH-β), 66.64 (CH-β), 61.54 (CH_2-β), 61.43 (CH_2-α), 61.31 (CH_2-β), 61.27 (CH_2-α), 54.06 (CH-β), 54.01 (CH-β), 21.60 (CH_3-β), 21.44 (CH_3-α)

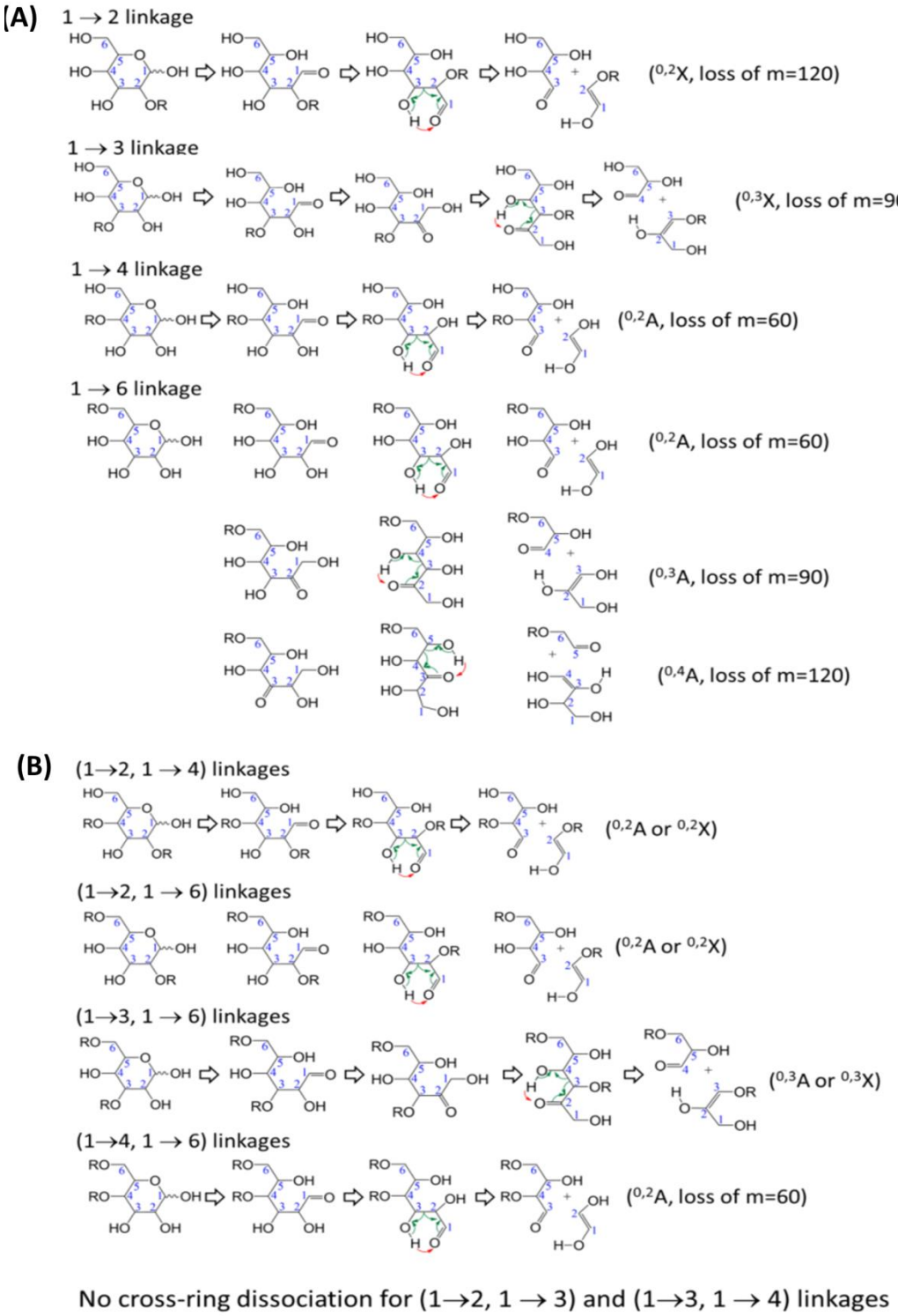

Fig. S3. Cross-ring dissociation mechanisms of oligosaccharides with Hex at the reducing end. (A): linear, (B): branched oligosaccharides, according to retro-aldol reaction.

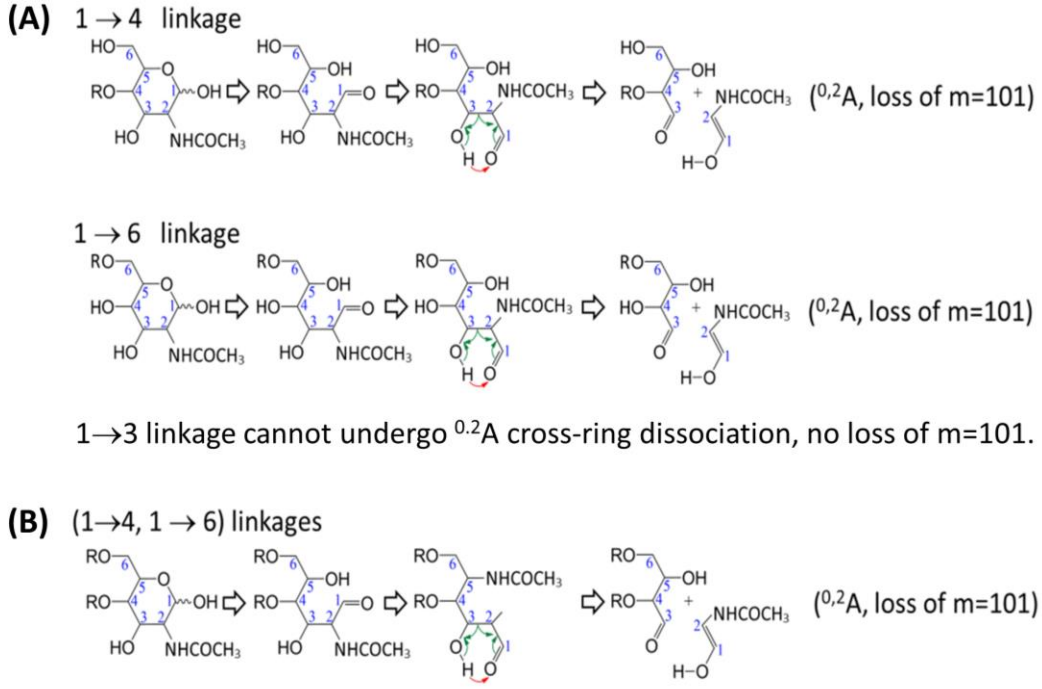

Fig. S4. Cross-ring dissociation mechanisms of oligosaccharides with HexNAc at the reducing end. (A): linear, (B): branched oligosaccharides, according to retro-aldol reaction. according to retro-aldol reaction.

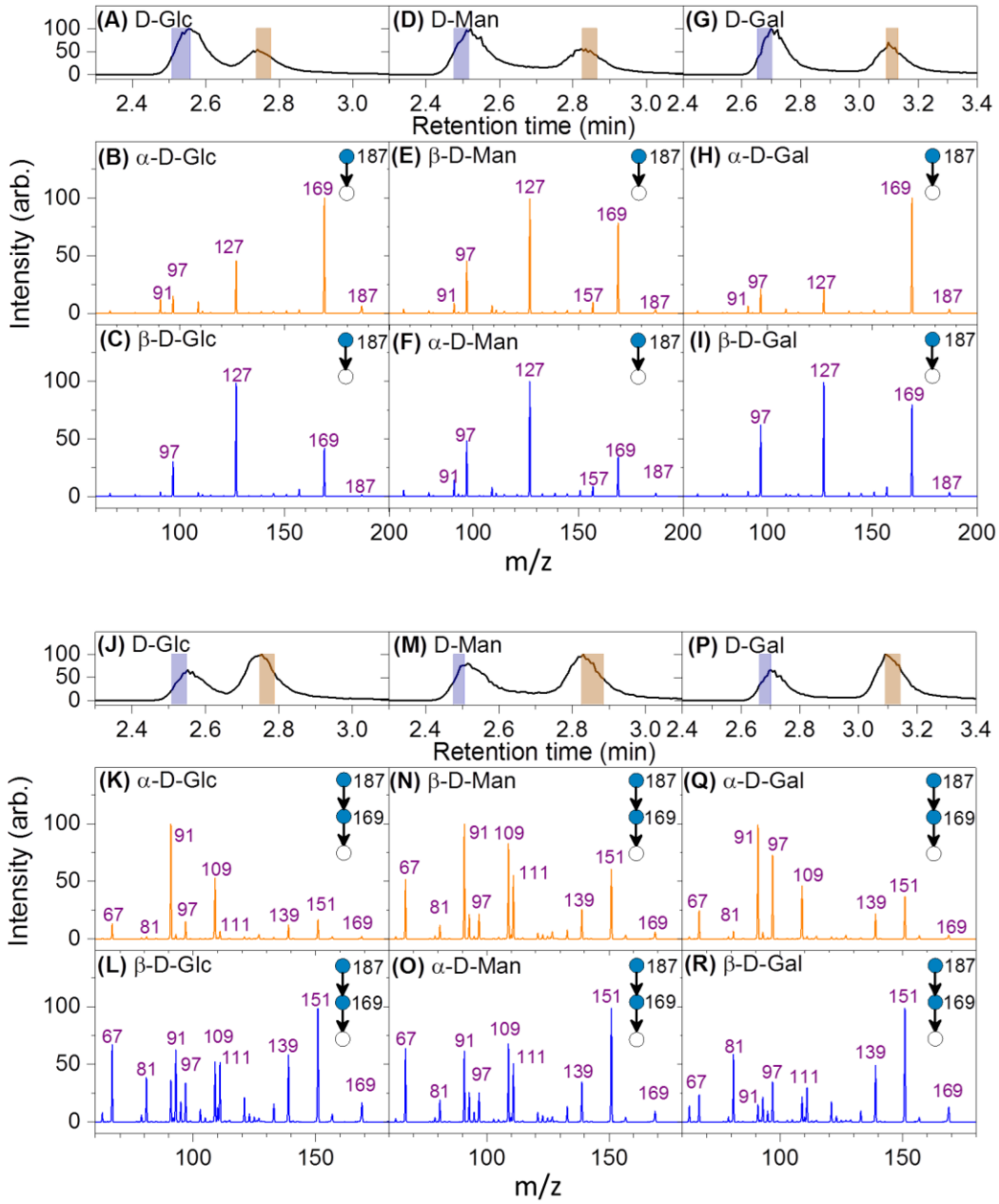

Fig. S5. Monosaccharide database for the Hex not at the reducing end of oligosaccharide. HPLC chromatograms and CID spectra of monosaccharide lithium adducts: 187→fragments, and 187→169→fragments. Each anomer was measured immediately after the separation of two anomers through HPLC. The blue and orange bars surrounding each peak in the chromatograms represent the period in which the CID spectra were measured. The corresponding spectra are displayed in blue and orange, respectively. These CID spectra were collected to serve as the monosaccharide database for the Hex not at the reducing end.

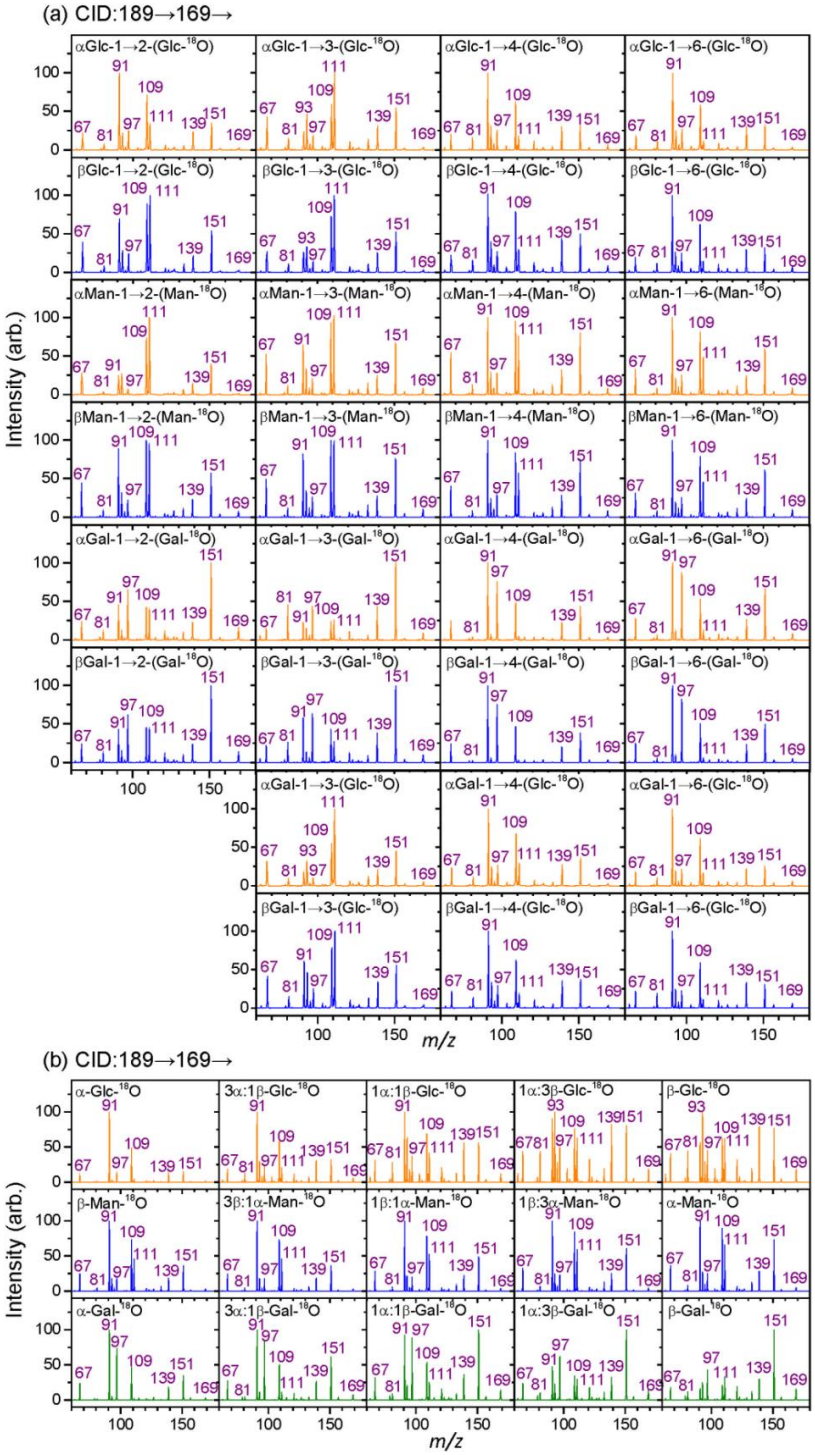

Fig. S6. Monosaccharide database for the Hex at the reducing end of oligosaccharide. ^{18}O was labeled at O1 atom of monosaccharides or the sugar at the reducing end of hexose disaccharides. (a) CID spectra of lithium adducts: 351→189→169→fragments. In (b), the CID spectra of lithium adducts: 189→169→fragments of α-Glc-^{18}O, β-Man-^{18}O, α-Gal-^{18}O, β-Glc-^{18}O, α-Man-^{18}O and β-Gal-^{18}O were measured immediately after the separation of two anomers of ^{18}O labeled monosaccharides by HPLC. The other CID spectra were the sum of CID spectra of separated anomers with different ratios (3:1, 1:1, or 1:3).

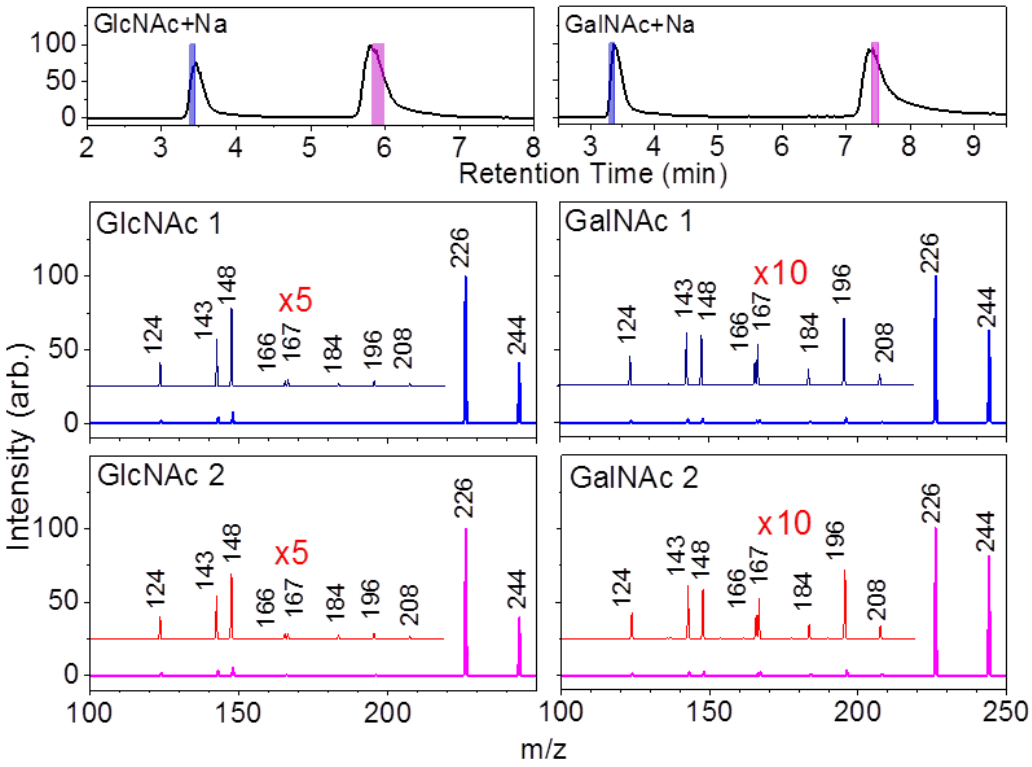

Fig. S7. HPLC chromatograms and CID spectra of HexNAc sodium adducts. The CID spectrum of each anomer was measured immediately after the separation of two anomers through HPLC. The blue and pink bars surrounding each peak in the chromatograms represent the period in which the CID spectra were measured. The corresponding spectra are displayed in blue and pink, respectively. These CID spectra were collected to serve as the monosaccharide database. The major difference between GlcNAc and GalNAc is the relative intensities of fragment ions m/z 143, 148, and 196 which are used in spectrum similarity calculations. The method of spectrum similarity calculations is the same as that of Hex, which was described in our previous reports.

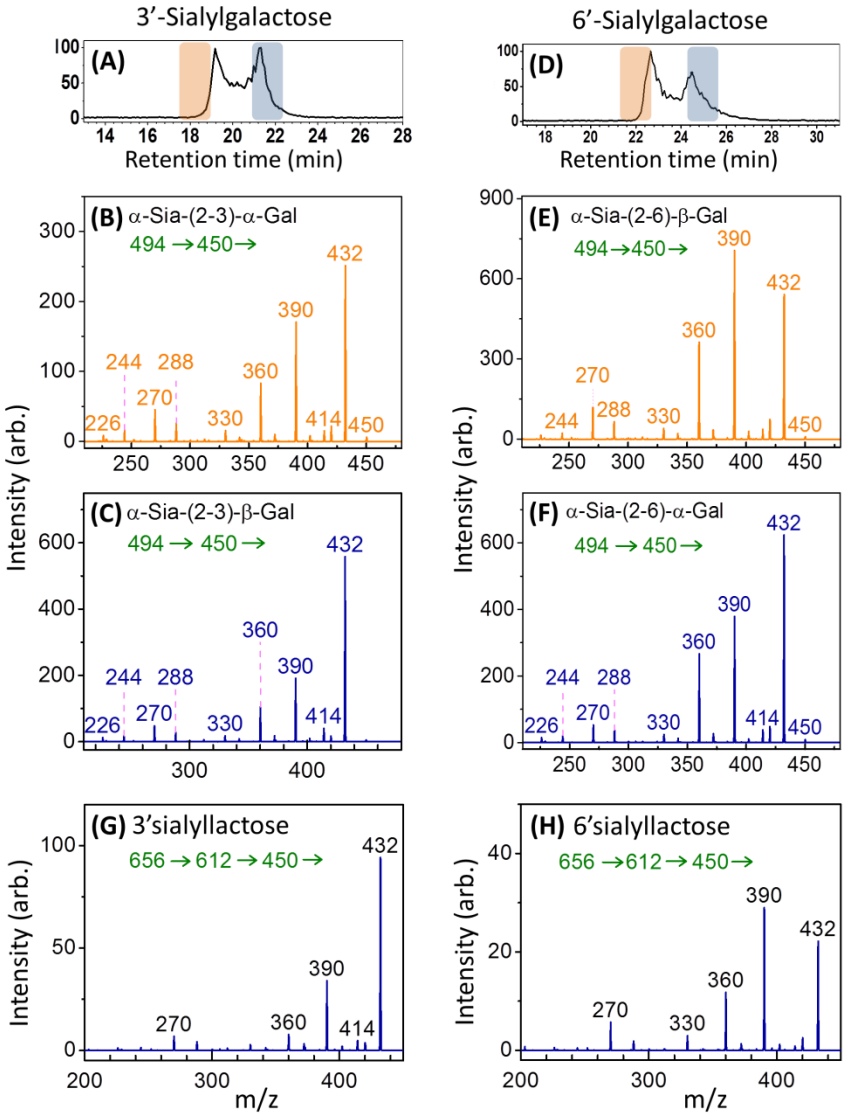

Fig. S8. Chromatograms of sialylgalactose and CID spectra of sialylgalactose and sialyllactose using precursor ion of sodium adduct ($M+Na^+$). Two anomeric configurations at the reducing end of each sialygalactose (3' sialygalactose or 6' sialygalactose) which coexisted in nature were separated by HPLC and then measured the corresponding CID spectra immediately after separation. The chromatograms are illustrated in (A) and (D), and the corresponding CID spectra are shown in (B), (C), (E), and (F), respectively. Assignment of anomericity to the CID spectra was made by comparison to the CID spectra of sialyllactose in (G) and (H). The CID sequence of sialyllactose, 656→612→450, generate the sodium adduct of α-Sia-(2-x)-β-Gal without CO_2. Comparison of (G) to (B) and (C) shows that (C) is α-Sia-(2→3)-β-Gal, and thus the other spectrum (B) is α-Sia-(2→3)-α-Gal. The analogous method was used to assign the spectra of (E) and (F), and the CID spectra of sialygalactose using precursor ion of sodium adduct ($M-H^++2Na^+$).

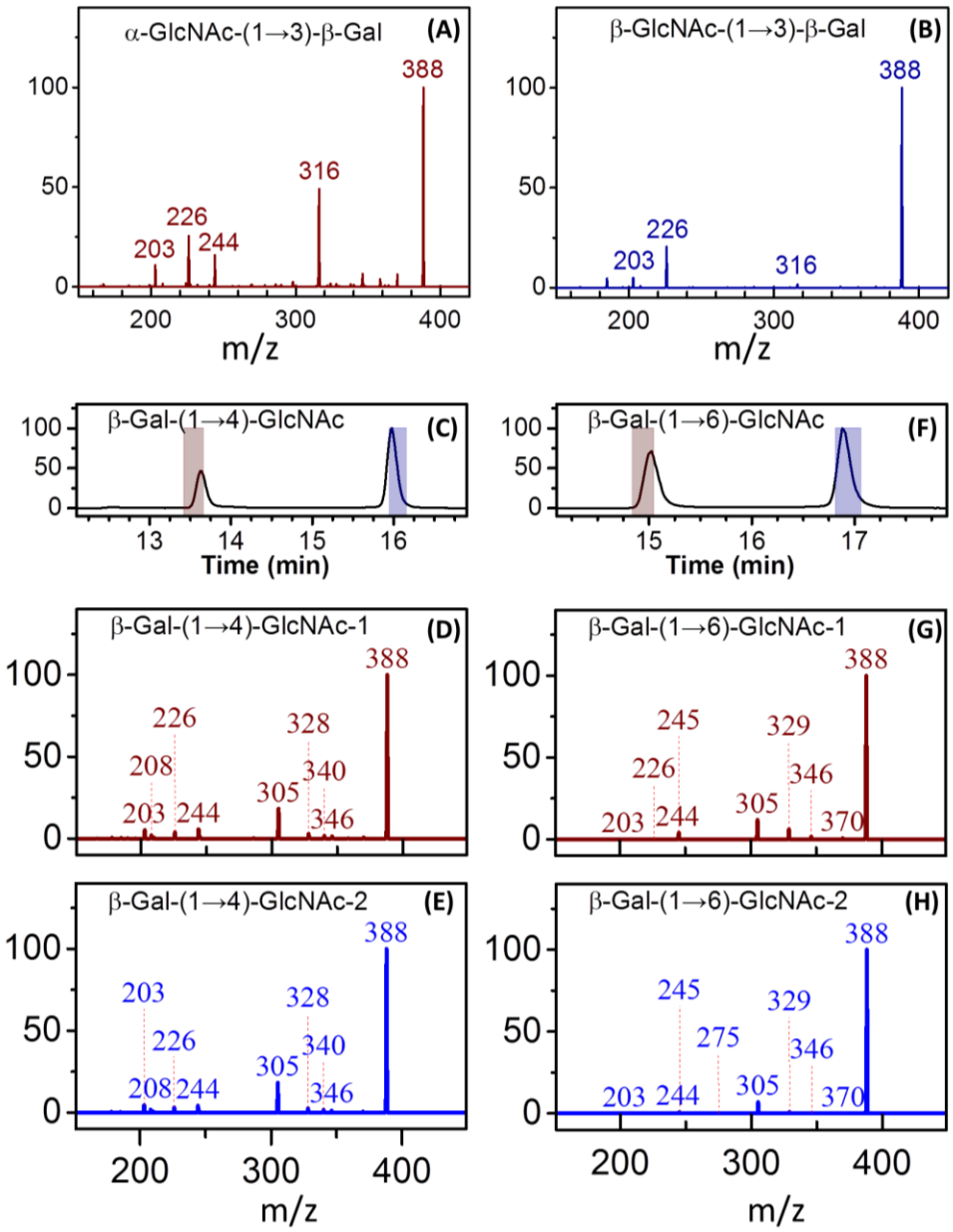

Fig. S9. Part of disaccharide database. CID spectra of GlcNAc-(1→3)-Gal disaccharide sodium adducts ((A) and (B)). Chromatograms ((C) and (F)) and CID spectra ((D), (E), (G), (H)) of β-Gal-(1→4)-GlcNAc and β-Gal-(1→6)-GlcNAc disaccharide sodium adducts. The difference of the CID spectra between β-Gal-(1→4)-GlcNAc and β-Gal-(1→6)-GlcNAc is the fragments *m/z* 328 (only occurs for 1→4 linkage) and 329 (only occurs for 1→6 linkage). These spectra were used in the structural determination of LSTa and LSTc in this study.

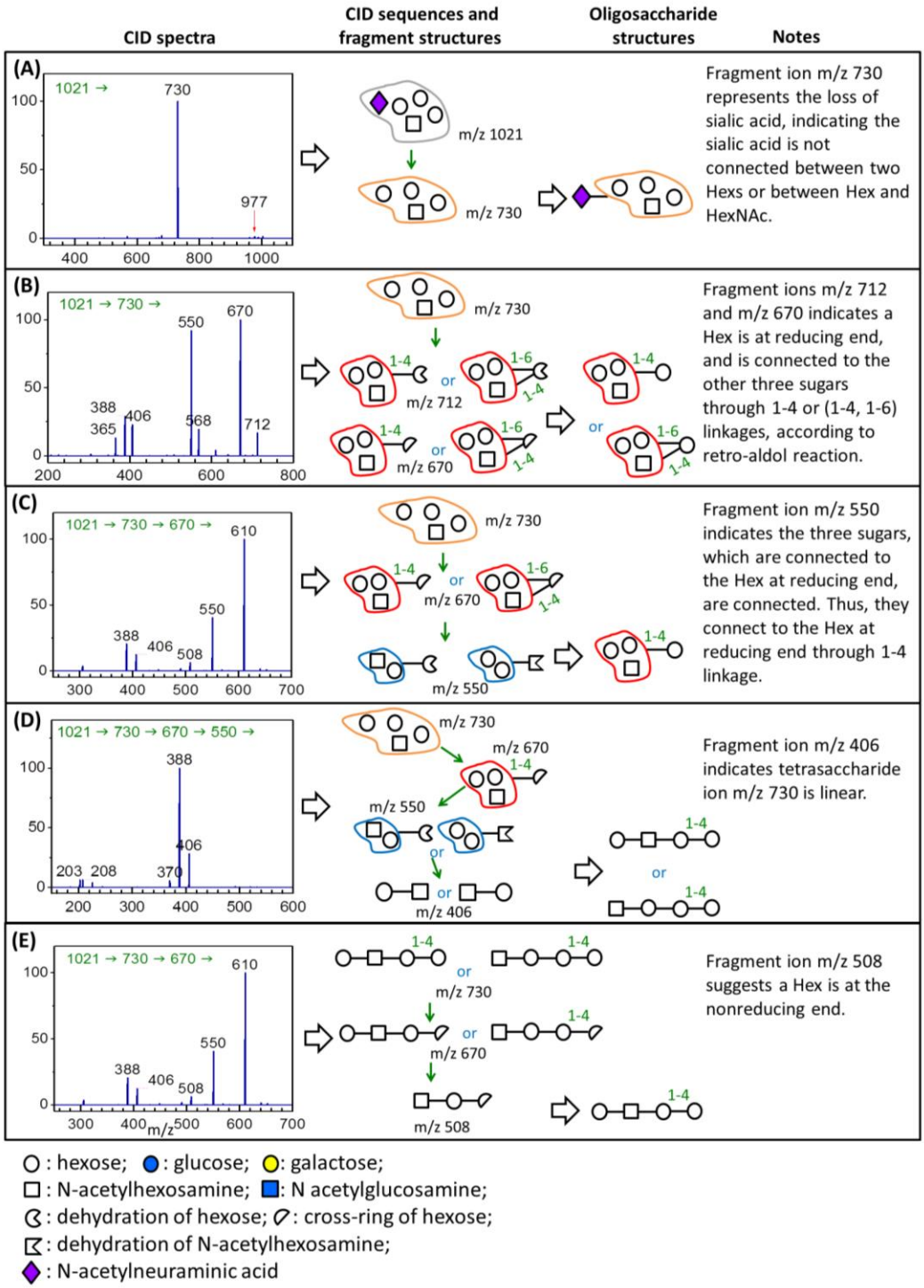

Fig. S10. Part of the structural analysis of LSTa in step 1(a). Parts of the linkage positions were determined in this step. Precursor ion *m/z* 1021 ($M+Na^+$) was used in these CID spectra. The precursor ion (*m/z* 1021) is the precursor molecule (molecular weight: 998 Da) attached to one sodium ion, representing an oligosaccharide containing one N-acetylneuraminic acid (Neu5Ac), three Hexs, and a HexNAc. The CID spectrum of the precursor ion *m/z* 1021, as illustrated in the right side of (A), revealed that the major fragment was ion *m/z* 730, representing the loss of Neu5Ac from the precursor ion. Observation of the ion *m/z* 730 indicates that Neu5Ac is neither

between two Hexs nor between Hex and HexNAc. The CID sequence and structures of important fragments are illustrated in the middle part of (A), and the possible oligosaccharide structures derived from the fragments are illustrated in the left side of (A). The orange circle around three Hexs and one HexNAc in the middle part of (A) indicates that the glycosidic bonds between these four sugars had not been determined. The CID spectrum of tetrasaccharide ion *m/z* 730 is illustrated in the left part of (B). Ion *m/z* 670 represents the fragments produced from the cross-ring dissociation (^{0,2}A, loss of neutral m = 60 Da) of the Hex at the reducing end of the tetrasaccharide ion *m/z* 730, suggesting that the Hex at the reducing end of ion *m/z* 730 was connected to the other three sugars [two Hexs and one HexNAc, circled in red in (B)] through one glycosidic bond with a 1→4 linkage or through two glycosidic bonds with 1→4 and 1→6 linkages. Possible structures of these fragments are illustrated in the middle part of (B), and the possible tetrasaccharide structures derived from these fragments are illustrated in the right part of (B). In the CID spectrum of 1021→730→670→fragments [the left part of (C)], the fragment ion *m/z* 550 suggested that the Hex at the reducing end of tetrasaccharide ion *m/z* 730 was connected to the other three sugars through one glycosidic bond with the 1→4 linkage. The possible oligosaccharide structure is illustrated in the right part of (C). The CID spectrum of 1021→730→670→550→fragments is displayed in the left part of (D). The fragment ion *m/z* 406 represents a disaccharide consisting of a Hex and a HexNAc. The observation of this ion indicates the ion *m/z* 730 is linear and a disaccharide with Hex-HexNAc or HexNAc-Hex is located at the terminal nonreducing end. Fragment ion *m/z* 508 in (E) indicates a Hex at the terminal nonreducing end of the ion *m/z* 730.

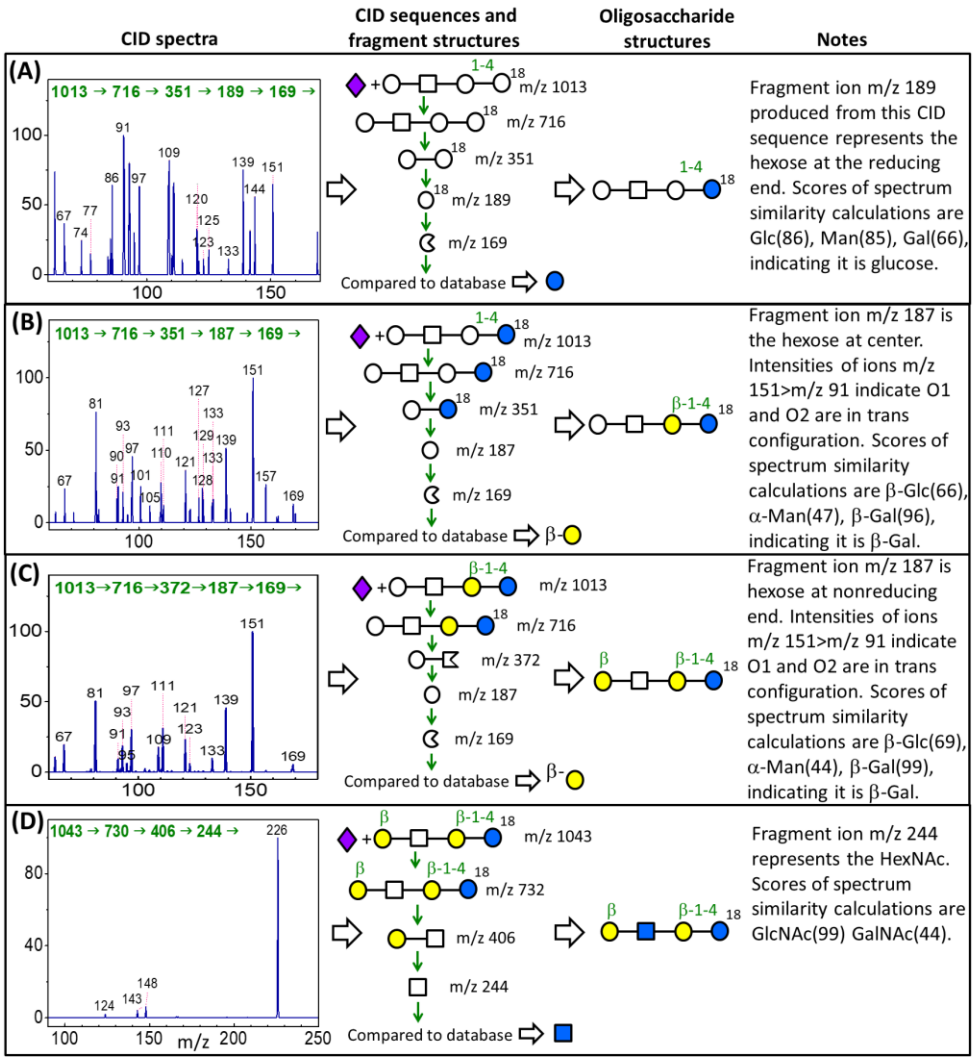

Fig. S11. Structural analysis of LSTa in step 1(b) and 1(c). Stereoisomers and parts of the anomericities were determined in these steps. Precursor ion *m/z* 1013 ($M+Li^+$) was used in (A)-(D) and precursor ion *m/z* 1043 ($M+2Na^+-H^+$) was used in (D). The selected Hexs were produced from the special sequences of CID of ^{18}O-labelled glycan, as displayed in the middle of each subfigure. The corresponding CID spectra, as illustrated in the left part (A)-(C), were compared to the CID spectra in our Hex monosaccharide database in Figure S5 and S6 in Supplementary Information for structural identification. The monosaccharides were identified as Glc, β-Gal, and β-Gal for the Hexs at the reducing end, middle, and terminal nonreducing end of the tetrasaccharide, respectively. Comparing the CID spectrum in (D) and HexNAc monosaccharide database indicates the HexNAc is GlcNAc.

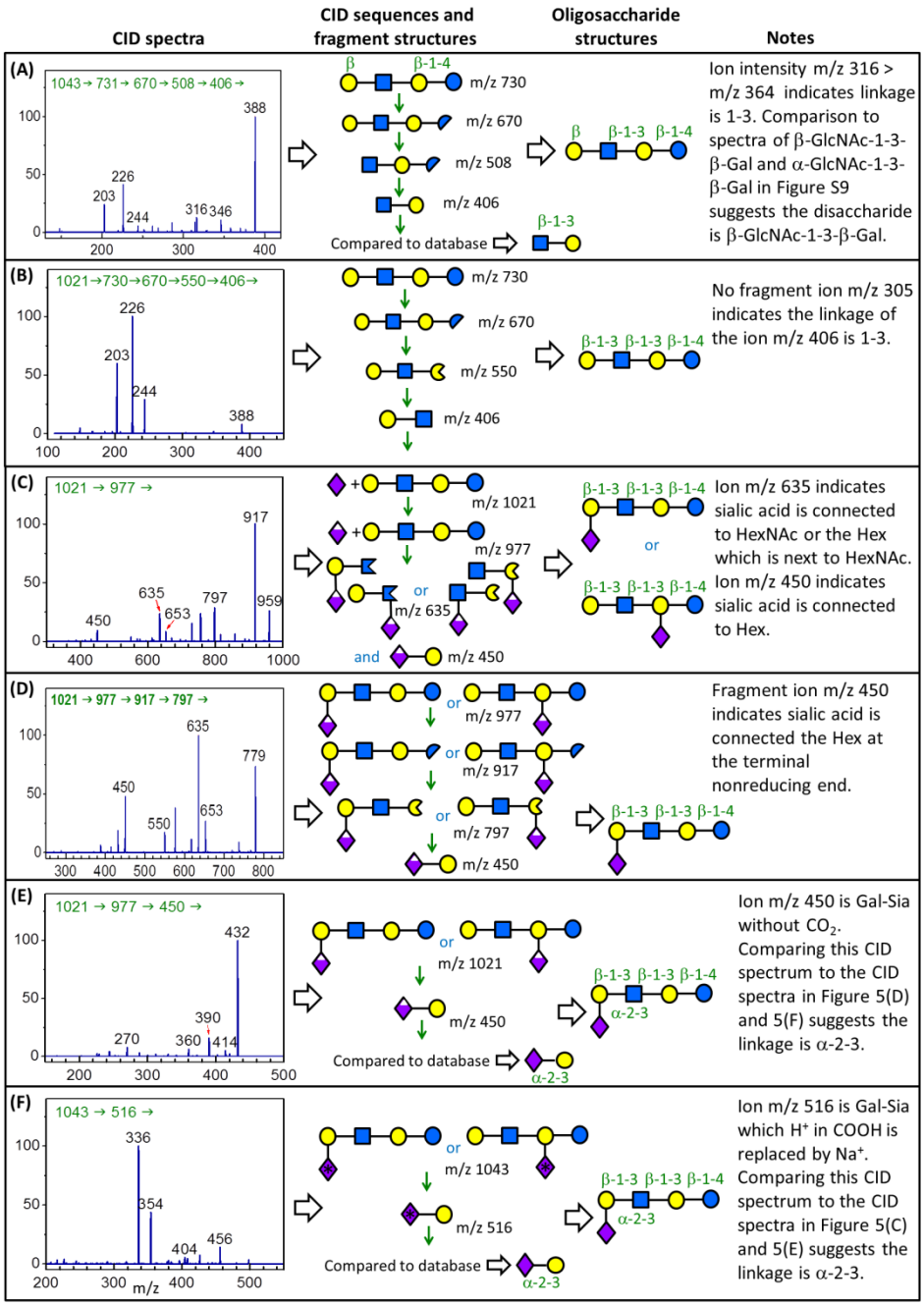

Fig. S12. Part of the structural analysis of LSTa in step 1(a) and step 2. (A) and (B): Parts of the linkage positions and anomericities were determined in this step. (C)-(F): Structural analysis of LSTa in step 2. Precursor ions *m/z* 1021 ($M+Na^+$) was used to determine the linkage of sialic acid in (C)-(E), and precursor ions *m/z* 1043 ($M+2Na^+-H^+$) was used in (F) to double check the linkage of sialic acid. Ion *m/z* 406 produced according to the CID sequence in (A) represents the disaccharide HexNAc-Hex at the center of the tetrasaccharide ion *m/z* 730. Ion *m/z* 316 in the CID spectrum of ion *m/z* 406 in (A) indicates the linkage is 1→3. Comparing the CID spectrum in (A) and GlcNAc-(1→3)-Gal disaccharide database indicates it is β-GlcNAc-(1→3)-Gal. Ion

m/z 406 produced according to the CID sequence in (B) represents the disaccharide Hex-HexNAc at the terminal nonreducing end. Ion *m/z* 305 not observed in CID spectrum of (B) suggest the disaccharide Hex-HexNAc at the terminal nonreducing end has a 1→3 linkage. Ion *m/z* 977 in (C) is the loss of CO_2 from the carboxylic group of precursor ion m/z 1021. Ion *m/z* 635 in (C) indicates sialic acid is not connected to the hexose at the reducing end. Ion *m/z* 450 in (C) indicates sialic acid is connected to the hexose. Ion *m/z* 450 in (D) indicates sialic acid is connected to the Hex at the terminal nonreducing end. Comparison of the CID spectrum of ion *m/z* 450 in (D) and sialic acid–galactose disaccharide database suggests the linkage is α-2→3. The linkage of sialic acid–galactose was double checked using the other sialic acid-galactose disaccharide, ion *m/z* 516, produced from the other precursor ion *m/z* 1043. Comparison of the CID spectrum of ion *m/z* 516 in (D) and sialic acid–galactose disaccharide database suggests the linkage is α-2→3.

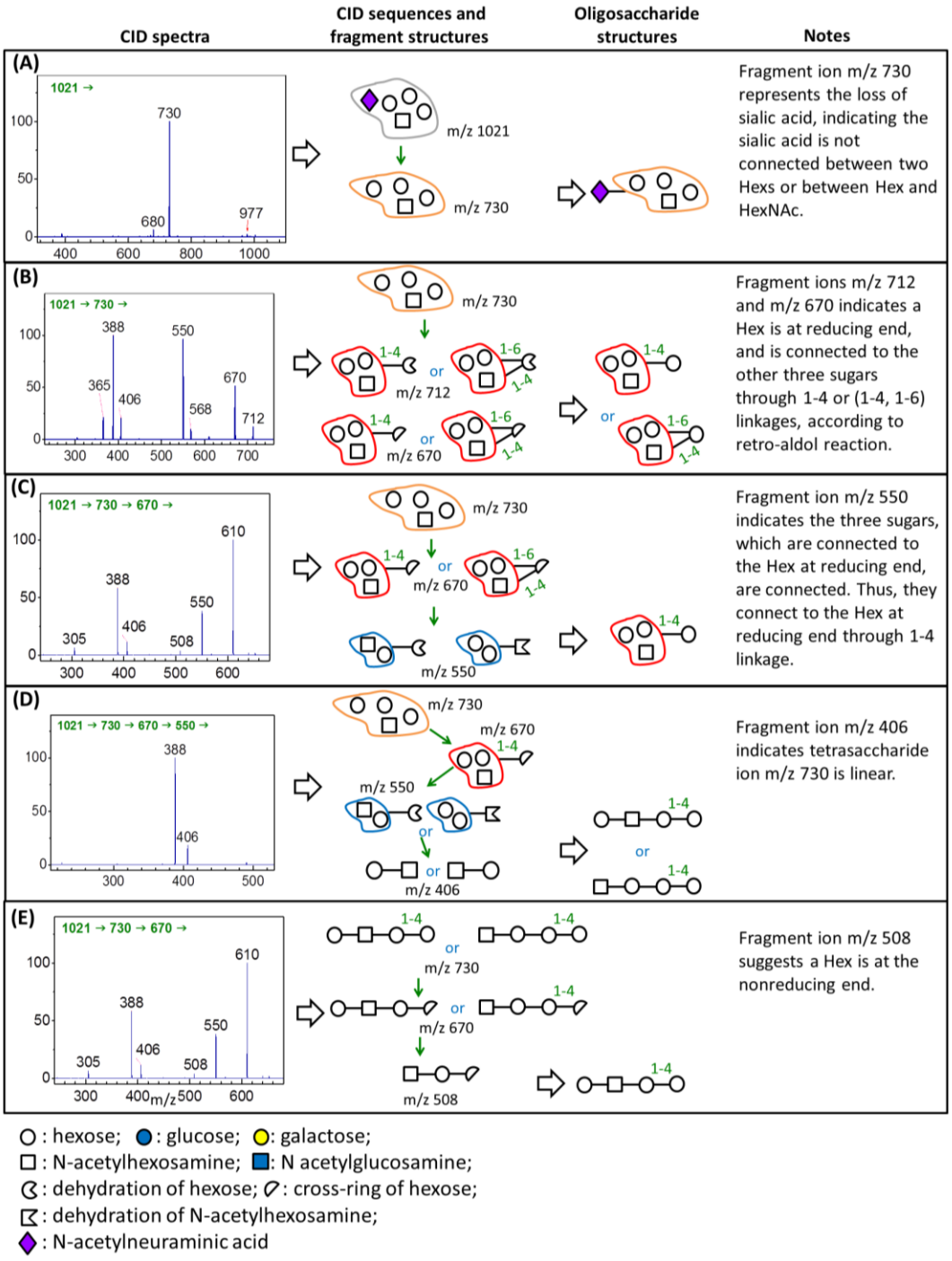

Fig. S13. Part of the structural analysis of LSTc in step 1(a). Parts of the linkage positions were determined in this step. The precursor ion (*m/z* 1021) is the precursor molecule (molecular weight: 998 Da) attached to one sodium ion, representing an oligosaccharide containing one N-acetylneuraminic acid (Neu5Ac), three Hexs, and a HexNAc. The CID spectrum of the precursor ion *m/z* 1021, as illustrated in the right side of (A), revealed that the major fragment was ion *m/z* 730, representing the loss of Neu5Ac from the precursor ion. Observation of the ion *m/z* 730 indicates that Neu5Ac is neither between two Hexs nor between Hex and HexNAc. The CID sequence and structures of important fragments are illustrated in the middle part of (A), and the

possible oligosaccharide structures derived from the fragments are illustrated in the left side of (A). The orange circle around three Hexs and one HexNAc in the middle part of (A) indicates that the glycosidic bonds between these four sugars had not been determined. The CID spectrum of tetrasaccharide ion *m/z* 730 is illustrated in the left part of (B). Ion *m/z* 670 represents the fragments produced from the cross-ring dissociation (^{0,2}A, loss of neutral m = 60 Da) of the Hex at the reducing end of the tetrasaccharide ion *m/z* 730, suggesting that the Hex at the reducing end of ion *m/z* 730 was connected to the other three sugars [two Hexs and one HexNAc, circled in red in (B)] through one glycosidic bond with a 1→4 linkage or through two glycosidic bonds with 1→4 and 1→6 linkages. Possible structures of these fragments are illustrated in the middle part of (B), and the possible tetrasaccharide structures derived from these fragments are illustrated in the right part of (B). In the CID spectrum of 1021→730→670→fragments [the left part of (C)], the fragment ion *m/z* 550 suggested that the Hex at the reducing end of tetrasaccharide ion *m/z* 730 was connected to the other three sugars through one glycosidic bond with the 1→4 linkage. The possible oligosaccharide structure is illustrated in the right part of (C). The CID spectrum of 1021→730→670→550→fragments is displayed in the left part of (D). The fragment ion *m/z* 406 represents a disaccharide consisting of a Hex and a HexNAc. The observation of this ion indicates the ion *m/z* 730 is linear and a disaccharide with Hex-HexNAc or HexNAc-Hex is located at the terminal nonreducing end. Fragment ion *m/z* 508 in (E) indicates a Hex is located at the terminal nonreducing end of the ion *m/z* 730.

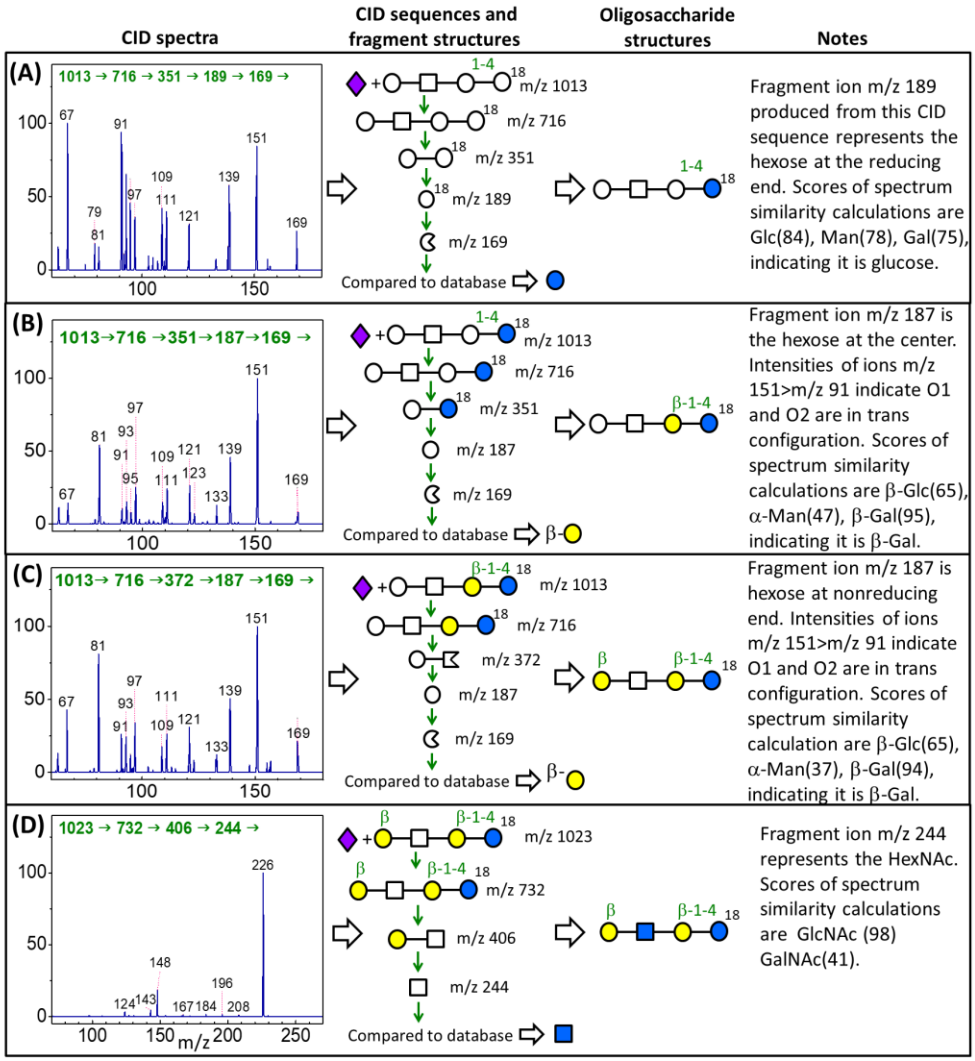

Fig. S14. Part of the structural analysis of LSTc in step 1(b) and 1(c). Stereoisomers and parts of the anomericities were determined in these steps. The selected Hexs were produced from the special sequences of CID of ^{18}O-labelled glycan, as displayed in the middle of each subfigure. The corresponding CID spectra, as illustrated in the left part (A)-(C), were compared to CID spectra in our Hex monosaccharide database in Figure S5 and S6 in Supplementary Information for structural identification. The monosaccharides were identified as Glc, β-Gal, and β-Gal for the Hexs at the reducing end, middle, and terminal nonreducing end of the tetrasaccharide, respectively. Comparing the CID spectrum in (D) and the HexNAc monosaccharide database indicates the HexNAc is GlcNAc.

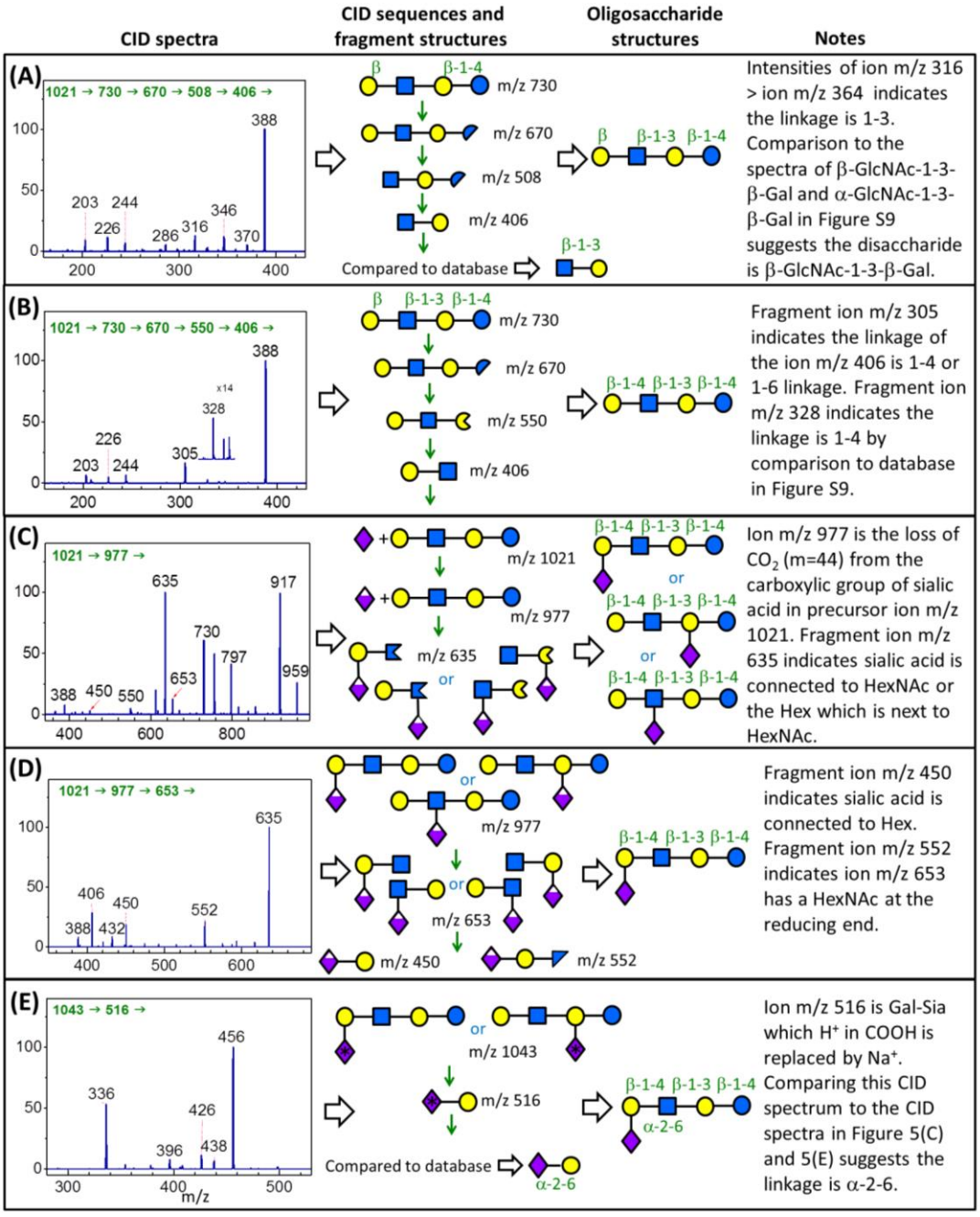

Fig. S15. Part of the structural analysis of LSTc in step 1(a) and step 2. (A) and (B): Parts of the linkage positions and anomericities were determined in this step. (C)-(E): Structural analysis of LSTc in step 2. Linkage of sialic acid was determined in this step. Ion *m/z* 406 produced according to the CID sequence in (A) represents the disaccharide HexNAc-Hex at the center of ion m/z 730. Ion *m/z* 316 in the CID spectrum of ion *m/z* 406 in (A) indicates the linkage is 1→3. Comparing the CID spectrum in (A) and GlcNAc-(1→3)-Gal disaccharide database indicates it is β-GlcNAc-(1→3)-Gal. Ion *m/z* 406 produced according to the CID sequence in (B) represents the disaccharide Hex-HexNAc at the terminal nonreducing end. Ions *m/z* 305 and 328 in CID spectrum of (B) suggest the disaccharide Hex-HexNAc at non-reducing end has a 1→4 linkage.

Ion m/z 977 in (C) is the loss of CO_2 from the carboxylic group of precursor ion *m/z* 1021. Ion *m/z* 635 in (C) indicates sialic acid is not connected to the hexose at the reducing end. Ion *m/z* 450 in (D) indicates sialic acid is connected to a hexose. Ion *m/z* 552 in (D) represent the loss of neutral m=101, indicating ion m/z 653 in (D) has a HexNAc at the reducing end. Consequently, sialic acid is determined to connect to the hex at the terminal nonreducing end. Comparing the CID spectrum in (E) to the CID spectra of sialic acid-galactose disaccharide database suggests the linkage is α-2$\rightarrow$6.

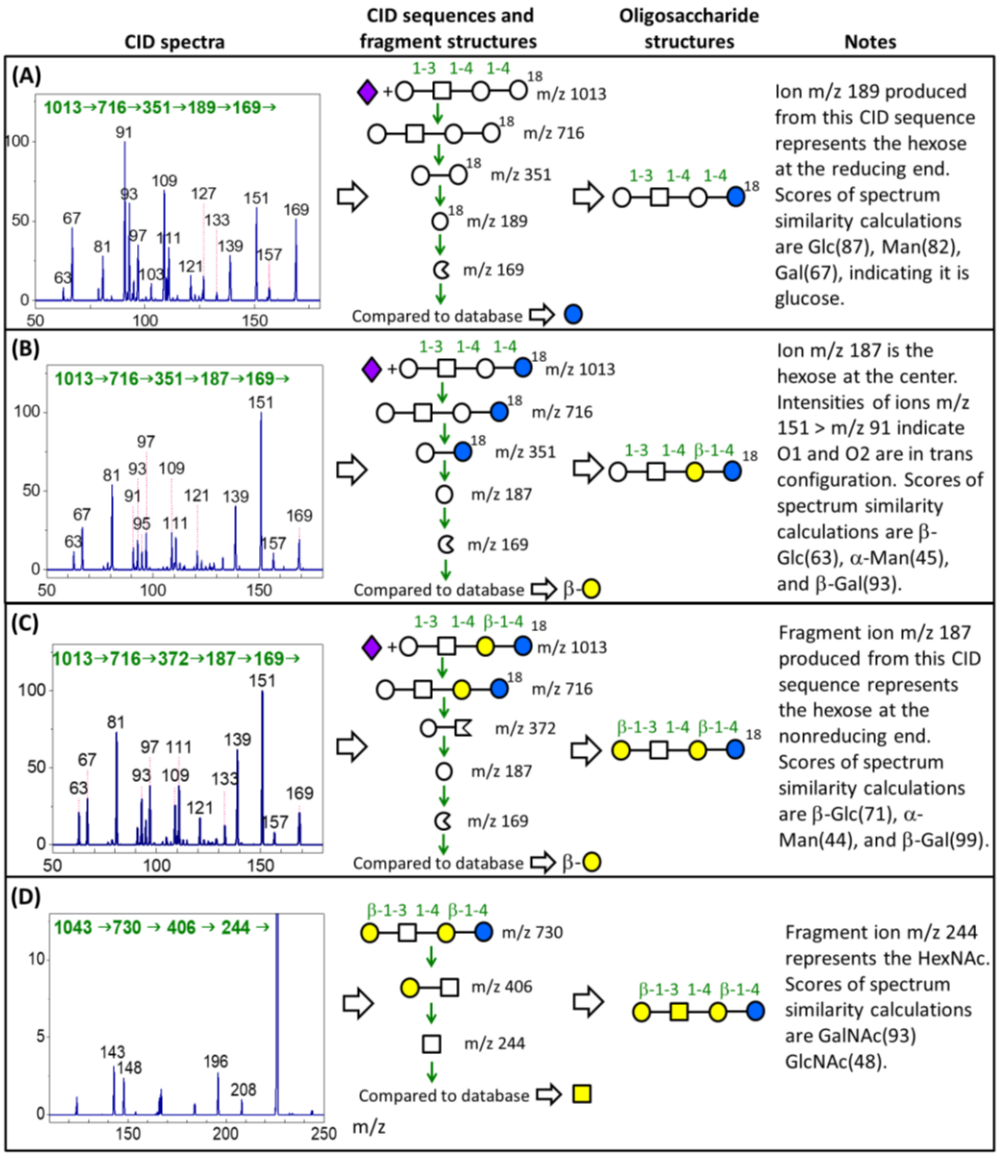

Fig. S16. Structural analysis of GM1a in step 1(b) and 1(c). (A)-(C): Lithium adducts of the ^{18}O labeled precursor ion m/z 1013 ($M+2Li^+-H^+$) was used in step 1(b). The Hexs at the reducing end, middle, and terminal nonreducing end were generated through the CID sequences illustrated in the middle parts of (A), (B), and (C), respectively. A comparison of the CID spectra to the CID spectra in the monosaccharide database (Supporting Information) revealed that they were Glc, β-Gal, and β-Gal. The spectrum similarity scores are shown in the parentheses. (D): Sodium adducts of precursor ion m/z 1043 was used in step 1(c). A comparison of the CID spectra to the CID spectra in the monosaccharide database (Supporting Information) revealed that it was GalNAc.

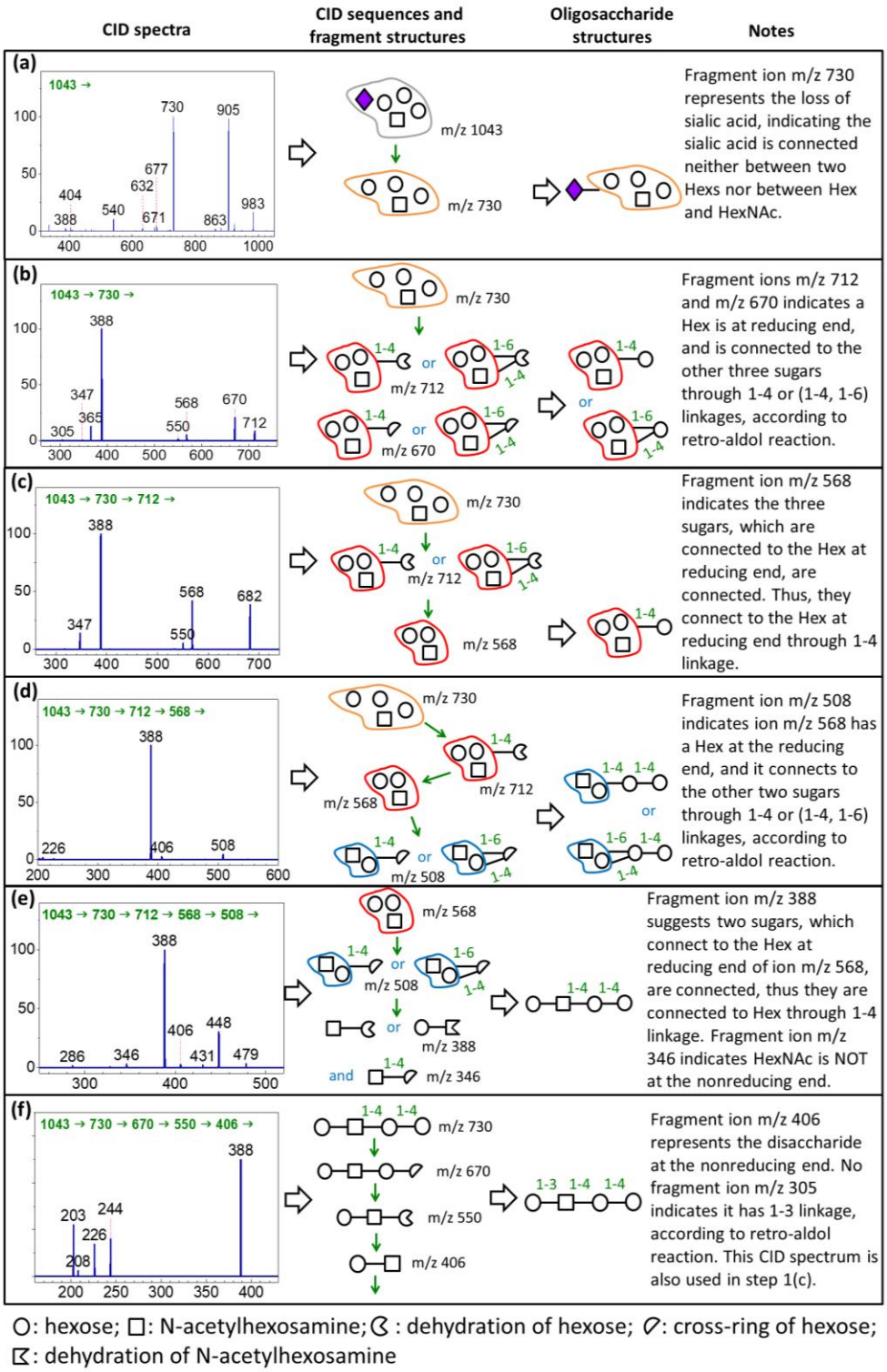

Fig. S17. Structural analysis of GM1a in step 1(a), analogous to the steps in Figure 3 for GM1b.