

## Unusual milk oligosaccharides and the biosynthetic pathways

Wei-Chien Weng<sup>a,b</sup>, Hun-En Liao<sup>a,c</sup>, Shih-Pei Huang<sup>a</sup>, Shang-Ting Tsai<sup>a</sup>, Hsu-Chen Hsu<sup>a</sup>, Chia-Yen Liew<sup>a,d</sup>, Veeranjanyulu Gannedi<sup>e</sup>, Shang-Cheng Hung<sup>e</sup>, Chi-Kung Ni<sup>\*a,b</sup>

<sup>a</sup> Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, 10617, Taiwan

<sup>b</sup> Molecular Science and Technology, International Graduate Program, Academia Sinica and National Tsing Hua University, Hsinchu, 30013, Taiwan

<sup>c</sup> Department of Applied Chemistry, National Yang Ming Chiao Tung University, Hsinchu, 300, Taiwan

<sup>d</sup> International Graduate Program of Molecular Science and Technology, National Taiwan University and Taiwan International Graduate Program of Molecular Science and Technology, Academia Sinica, Taipei, 10617, Taiwan

<sup>e</sup> Genomic Research Center, Academia Sinica, Taipei, 11529, Taiwan

\*Corresponding author, e-mail address: ckni@po.iam.s.sinica.edu.tw

## **Abstract**

Free oligosaccharides are abundant macronutrients in milk and involved in prebiotic functions and antiadhesive binding of pathogenic bacteria to colonocytes. Despite the importance of these oligosaccharides, structural determination of oligosaccharides is challenging, and milk oligosaccharide biosynthetic pathways remain unclear.

Oligosaccharide structures are conventionally determined using a combination of chemical reactions, exoglycosidase digestion, nuclear magnetic resonance spectroscopy, and mass spectrometry. Most reported free oligosaccharides are highly abundant and have lactose at the reducing end, and current oligosaccharide biosynthetic pathways in human milk are proposed based on these oligosaccharides. In this study, a new mass spectrometry technique, which can identify linkages, anomericities, and stereoisomers, was applied to determine the structures of free oligosaccharides in human, bovine, and caprine milk. Oligosaccharides that do not follow the current biosynthetic pathways and are not synthesized by any discovered enzymes were found, indicating the existence of undiscovered biosynthetic pathways and enzymes. New biosynthetic pathways were proposed.

Free oligosaccharides are among the most abundant macronutrients in milk<sup>1-2</sup>. Milk oligosaccharides are a class of indigestible carbohydrates. The majority of free oligosaccharides consumed by infants from their mother's milk are not digested and absorbed within the small intestine, and most of these oligosaccharides reach the colon of infants. The biological functions of free oligosaccharides present in milk have been of scientific interest for many years. Recent studies have suggested that milk oligosaccharides are involved in prebiotic functions and the antiadhesive binding of pathogenic bacteria to colonocytes<sup>2,3</sup>.

Most free milk oligosaccharides contain lactose at the reducing end, and this disaccharide is assumed to be the initial substrate in biosynthesis.  $\alpha$ -Lactalbumin, which is only found in lactating mammary glands and milk, plays a crucial role in the biosynthesis of oligosaccharides in milk. The biosynthesis of oligosaccharides starts from lactose, which is synthesized from UDP-Gal and glucose through transgalactosylation catalyzed by a complex consisting of  $\beta$ -1,4-galactosyltransferase I and  $\alpha$ -lactalbumin. The further glycosylation of lactose by glycosyltransferases produces various oligosaccharides<sup>4-9</sup>.

Although the structures of free oligosaccharides present in milk have been widely studied<sup>10-15</sup>, the non-template-based biosynthesis of oligosaccharides and the large number of oligosaccharide isomers make structural identification challenging even today. The full structures of oligosaccharides were conventionally determined through wet chemistry<sup>16-20</sup> and nuclear magnetic resonance spectroscopy (NMR)<sup>21-26</sup>. However, these methods are time consuming and require large amount of samples. These methods are suitable only for highly abundant oligosaccharides. The sensitivity of mass spectrometry is three to four orders of magnitude higher than that of wet chemistry and NMR. Since the invention of matrix-assisted laser

desorption/ionization and electrospray ionization, which can easily generate ionized biomolecules in the gas phase, mass spectrometry has been widely used for the structural determination of oligosaccharides<sup>27-29</sup>. However, only part of the structures of oligosaccharides, including compositions, sequences, and linkage positions, can be identified through conventional mass spectrometry. Conventional mass spectrometry does not provide information on anomericity (differentiation of the  $\alpha$  and  $\beta$  anomeric configuration of glycosidic bonds) and monosaccharide stereoisomers (e.g., differentiation of glucose, galactose, and mannose or N-acetylglucosamine and N-acetylgalactosamine). Although structures obtained through conventional mass spectrometry can be further identified according to possible biosynthetic pathways, using exoglycosidase digestion, or through comparison with oligosaccharide standards from an oligosaccharide mass spectrum library<sup>30-34</sup>, the results obtained are limited by current knowledge on biosynthetic pathways, the available enzymes, and existing oligosaccharide standards in libraries, prohibiting the discovery of new oligosaccharides.

Recently, a new mass spectrometry method, namely logically derived sequence (LODES) multistage tandem mass spectrometry ( $MS^n$ ) was developed for the full structural determination of oligosaccharides<sup>35-38</sup>. This method can be used to determine oligosaccharide structures including linkage positions, anomericities, and monosaccharide stereoisomers. The method has exhibited high sensitivity and can be combined with high-performance liquid chromatography (HPLC) for the online structural identification of small oligosaccharides. In this study, we applied LODES/ $MS^n$  for the structural determination of free oligosaccharides extracted from human, bovine, and caprine milk. Many new oligosaccharides were identified, some of them do not follow the current biosynthetic pathways of free oligosaccharides in

human milk and are not synthesized by any discovered enzymes, pointing toward the existence of further undiscovered biosynthetic pathways and enzymes. The present study compared free oligosaccharide biosynthetic pathways in human milk versus bovine milk and proposed new biosynthetic pathways.

## Results

Free oligosaccharides extracted from milk were separated into several groups by using a size exclusion column followed by HPLC with an amide-80 column. Each fraction of eluents obtained after amide-80 column separation was further separated into individual isomer through another HPLC with a porous graphitic carbon (PGC) column. Eluents obtained from the PGC column were transported into a linear ion trap mass spectrometer for structural determination by using LODS/MS<sup>n</sup>. Only neutral trisaccharides extracted from human, bovine, and caprine milk and neutral tetrasaccharides extracted from human and bovine milk consisting glucose, galactose, mannose, N-acetylglucosamine, and N-acetylgalactosamine were examined in this study. Details of the experimental methods are described in Method and Supplementary Information, section A.

### (A) Trisaccharides

Fig. 1a–c, and d–f illustrate the chromatograms of ions  $m/z$  527 [sodium adducts of (Hex)<sub>3</sub>] and  $m/z$  568 [sodium adducts of (Hex)<sub>2</sub>HexNAc], respectively. Each chromatogram represents oligosaccharide separation performed using a PGC column from a fraction of eluents separated using the amide-80 column for human milk oligosaccharides (HMO), bovine milk oligosaccharides (BMO), or caprine milk oligosaccharides (CMO). The fraction of eluents separated by the amide-80 column was chosen such that all oligosaccharides identified in this study could be shown in a

single chromatogram for compact display in Fig. 1. Fractions consisting of fewer oligosaccharides are not shown. Therefore, the relative intensity shown in the chromatogram in Fig. 1 does not represent the relative abundance of these oligosaccharides in milk.

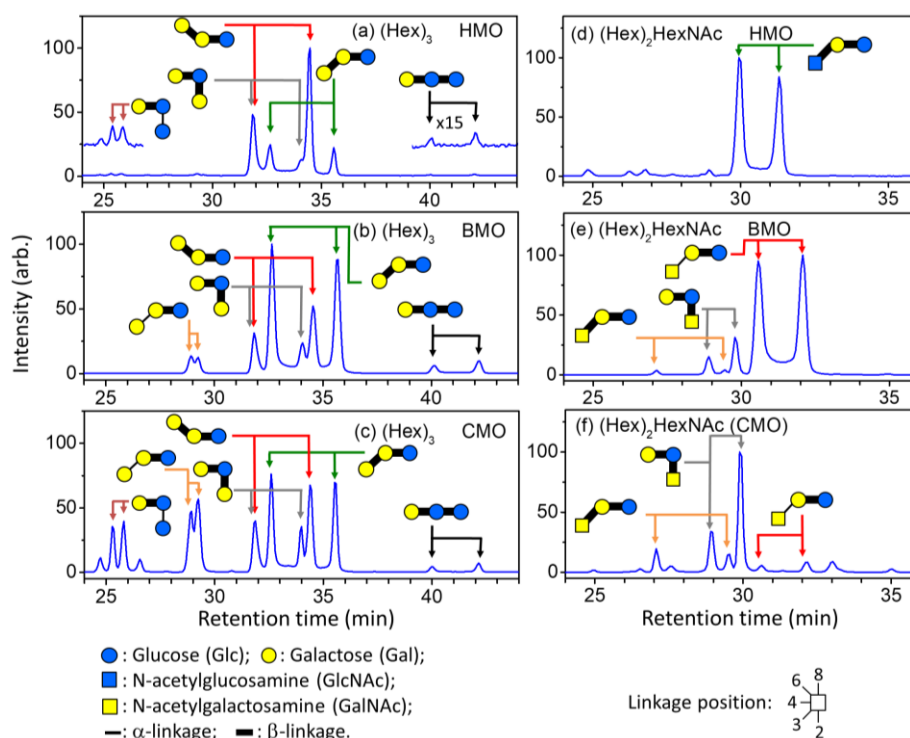


Fig. 1. PGC chromatograms of oligosaccharides separated using a PGC column from a fraction of eluents separated using an amide-80 column for human milk oligosaccharides (HMO), bovine milk oligosaccharides (BMO), or caprine milk oligosaccharides (CMO). (a)–(c) ion  $m/z$  527  $(\text{Hex})_3$  sodium adducts, and (d)–(f) ion  $m/z$  568  $(\text{Hex})_2\text{HexNAc}$  sodium adducts. The fraction collected from the amide-80 column was chosen so that all oligosaccharides found in this study could be shown in a single chromatogram for compact display, therefore, the relative intensity in this chromatogram does not represent the relative abundance of these oligosaccharides in milk. The oligosaccharide structure for each peak in chromatograms was determined using LODS/MS<sup>n</sup>. Details of structural determination by using LODS/MS<sup>n</sup> are

described in text and Supplementary Information. The HPLC retention times of chemically synthesized compounds illustrated in Supplementary Information were used to double check the structures determined by LODS/MS<sup>n</sup>.

Notably, oligosaccharides extracted from milk in this study were not reduced at the reducing end. They were intact, and each oligosaccharide had two anomers ( $\alpha$  and  $\beta$  of the sugar at the reducing end), which coexisted in the solution. The PGC column could separate these two anomers for most oligosaccharides, resulting in two peaks for each oligosaccharide in the chromatogram. Although the chromatogram became complicated when oligosaccharides were not reduced, the use of intact oligosaccharides has several advantages. First, no loss of sample during the reduction process; this increased the sensitivity for detecting less abundant oligosaccharides. Second, no products generated from potential unintended reactions, such as the peeling reaction, during reduction interfered in the oligosaccharide structural determination. Third, the structure of each oligosaccharide could be identified twice by using these two peaks in the chromatogram, providing a crosscheck for structural identification. Fourth, the retention times and MS<sup>n</sup> mass spectra of two anomers belonging to one oligosaccharide isomer were not likely to be the same as those of two anomers belonging to another oligosaccharide isomer. The probability of the retention times and MS<sup>n</sup> mass spectra of one isomer being the same as those of another isomer was considerably low, particularly after they were separated through multidimensional HPLC. These advantages enabled us to determine oligosaccharide isomer structures unambiguously.

LODS/MS<sup>n</sup> involves the sequential collision induced dissociation (CID) of oligosaccharide sodium (or lithium) adducts in a mass spectrometer. The sequences of CID are guided by LODS which is derived from carbohydrate dissociation

mechanisms<sup>39-41</sup>. The mechanisms of oligosaccharide sodium adducts used in this study are summarized as three propensity rules as follows.

- (1) Dehydration mainly occurs at the reducing end of oligosaccharides.
- (2) Cross-ring dissociation mainly occurs at the reducing end of oligosaccharides and follows the retro-aldol reaction. Fragmentation patterns from cross-ring dissociation can be used to determine the linkage position of the sugar at the reducing end. Details of fragmentation patterns are illustrated in Supplementary Information.
- (3) The cleavages of the glycosidic bond to produce B, C, Y and Z ions occur at any glycosidic bond (i.e., not limited to the reducing end). The notations B, C, Y, and Z were used according to the nomenclature of Domon and Costello<sup>42</sup>.

The dissociation mechanism of lithium adducts is similar to that of sodium adducts; however, the dehydration and cross-ring dissociation occurring at the nonreducing end was not neglectable. Thus, the O1 atoms of the monosaccharide at the reducing end labeled by <sup>18</sup>O in some oligosaccharides are necessary when lithium adducts are used for structural determination.

In this study, we used the trisaccharide  $\beta$ -Gal-(1 $\rightarrow$ 4)- $\beta$ -Glc-(1 $\rightarrow$ 4)-Glc as an example to illustrate how the structures of oligosaccharides are determined using LODS/MS<sup>n</sup>. The CID spectra of the oligosaccharide of the peak at the retention time 40.1 min in Fig. 1b are shown in Fig. 2. The mass spectrum, presented on the left side of Fig. 2a, shows the fragments produced from CID of the precursor ion  $m/z$  527 [sodium adduct of (Hex)<sub>3</sub>]. The loss of neutral  $m = 60$  from the precursor ion resulting in the fragment ion  $m/z$  467 represented cross-ring dissociation at the reducing end (rule 2). The trisaccharide must be linear with a linkage of 1 $\rightarrow$ 4 at the reducing end or branched with 1 $\rightarrow$ 6 and 1 $\rightarrow$ 4 linkages at the reducing end, according to the



retro-aldol reaction (details of fragmentation patterns are provided in Supplementary Information). The CID sequence and structures of fragments are illustrated in the middle of Fig. 2a; the possible precursor structures derived from these fragments are illustrated on the right side of Fig. 2a. Ion  $m/z$  347 found in the CID sequence 527→467→fragments (left side of Fig. 2b) indicated that the trisaccharide was linear because a branched trisaccharide with 1→6 and 1→4 linkages cannot produce the fragment ion  $m/z$  347 from this CID sequence (middle of Fig. 2b). The CID sequence 527→509→365 represented dehydration at the reducing end (rule 1), followed by glycosidic bond cleavage (rule 3), as illustrated in the middle of Fig. 2c. Ion  $m/z$  365 produced from this CID sequence was the disaccharide at the nonreducing end of the trisaccharide. The CID spectrum of this disaccharide, as illustrated on the left side of Fig. 2c, showed a high intensity of ion  $m/z$  305 (the loss of neutral  $m = 60$  from ion  $m/z$  365), indicating that the linkage of the disaccharide is 1→4 according to fragmentation patterns obtained from the retro-aldol reaction (rule 2). These three CID spectra of sodium adducts (Fig. 2a–c) suggested that the trisaccharide is Hex-(1→4)-Hex-(1→4)-Hex. To determine the stereoisomer of each monosaccharide, the O1 atom of hexose at the reducing end of the trisaccharide was  $^{18}\text{O}$  labeled, and the CID spectra of  $^{18}\text{O}$  labelled trisaccharide lithium adducts were investigated. The hexose lithium adducts, with ion  $m/z$  187 or 189, produced from the CID sequences 513→451→331→187, 513→351→187, and 513→351→189 represented the hexose at the nonreducing end, center, and reducing end of the trisaccharide, respectively. The CID spectra of these hexose monosaccharides are compared to the monosaccharide database provided in Supplementary Information. Spectrum similarities were calculated in the comparison. Method of spectrum similarity calculations for structural identification has been described in our previous report<sup>37</sup>. The similarity scores are shown in Fig. 2. The monosaccharide (glucose, galactose, or mannose) which has the

highest similarity score is identified as the stereoisomer of the monosaccharide. The results show that the hexose at the nonreducing end, center, and reducing end of the trisaccharide were  $\beta$ -Gal,  $\beta$ -Glc, and Glc, respectively. Consequently, the entire trisaccharide was determined to be  $\beta$ -Gal-(1 $\rightarrow$ 4)- $\beta$ -Glc-(1 $\rightarrow$ 4)-Glc. This trisaccharide,  $\beta$ -Gal-(1 $\rightarrow$ 4)- $\beta$ -Glc-(1 $\rightarrow$ 4)-Glc, has not been found before, and its structure is unusual: it does not contain a lactose at the reducing end. In addition to the presence in bovine milk, we found that this trisaccharide was present in human milk and caprine milk as well, as illustrated in the chromatograms of Fig. 1a and c. We crosschecked the structure of this trisaccharide by using the chromatogram retention time of the chemically synthesized  $\beta$ -Gal-(1 $\rightarrow$ 4)- $\beta$ -Glc-(1 $\rightarrow$ 4)-Glc, as shown in Supplementary Information.

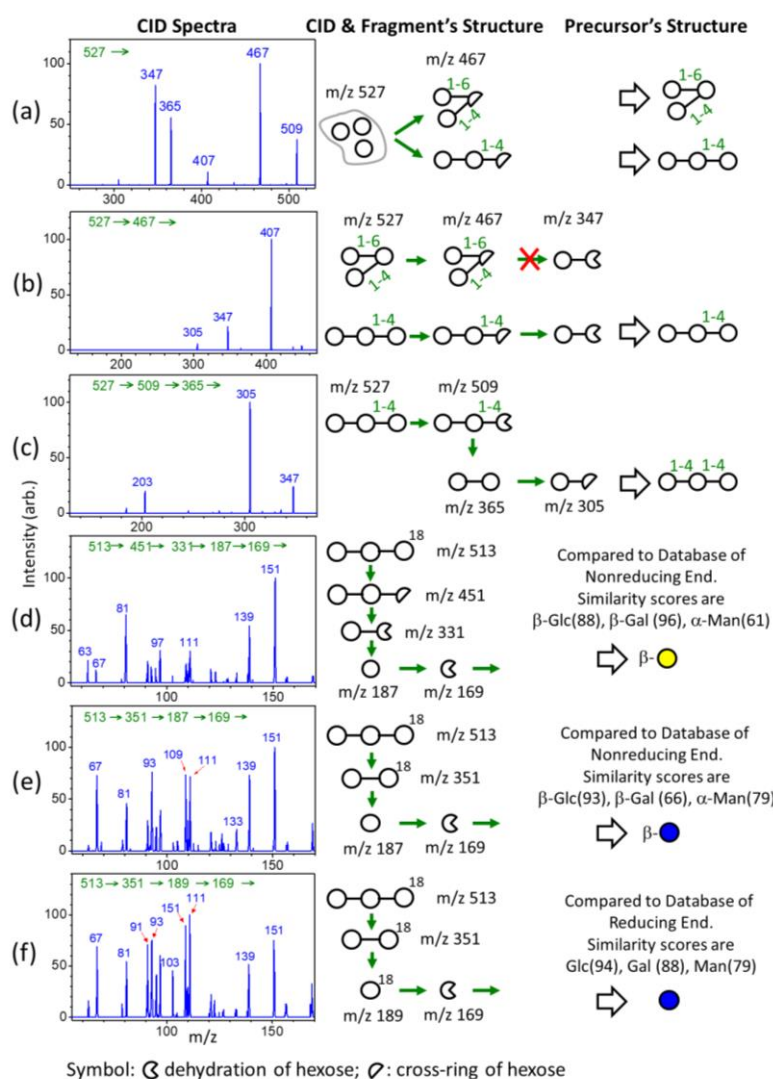


Fig. 2. Structural determination of the peak at retention time 40.1 min shown in Fig. 1b by using LODS/MS<sup>n</sup>. CID spectra, CID sequence, and possible fragment and precursor structures are illustrated in the right, middle, and left columns, respectively. Hexoses circled by gray line in (a) represent the linkage between these hexoses are not determined. Green arrows represent CID.

Five (Hex)<sub>3</sub> isomers were found in human and bovine milk and six (Hex)<sub>3</sub> isomers were discovered in caprine milk (Fig. 1a–c). We did not find the trisaccharide  $\beta$ -Gal-(1→4)- $\beta$ -Gal-(1→4)-Glc in human, bovine, or caprine milk. The existence of this trisaccharide in bovine milk has been reported previously<sup>31,32,43,44</sup>. Rudd et al used enzyme digestion and mass spectrometry to identify the structure of

$\beta$ -Gal-(1 $\rightarrow$ 4)- $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc in bovine milk<sup>31</sup>; however, these authors did not report the discovery of this oligosaccharide in bovine milk in their later report<sup>33</sup>. We examined the presence of this trisaccharide by using chemically synthesized  $\beta$ -Gal-(1 $\rightarrow$ 4)- $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc. The retention times of the trisaccharide  $\beta$ -Gal-(1 $\rightarrow$ 4)- $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc in the chromatogram were located at  $t = 21.6$  and  $24.0$  min (Supplementary Information), but no signal was observed in the chromatograms of human, bovine, or caprine milk at the same retention time.

The trisaccharide  $\beta$ -GlcNAc-(1 $\rightarrow$ 3)- $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc, which is a crucial precursor in human milk oligosaccharides for subsequent synthesis, is the dominant species of (Hex)<sub>2</sub>HexNAc in human milk. We did not find this trisaccharide in bovine and caprine milk, as observed by a comparison of the Hex<sub>2</sub>HexNAc chromatograms of human, bovine, and caprine milk in Fig. 1d, e and f, although the existence of this trisaccharide in bovine milk was reported in previous studies<sup>31,33,34</sup>. The trisaccharide  $\beta$ -GlcNAc-(1 $\rightarrow$ 6)- $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc was not found in human, bovine, or caprine milk<sup>31,33</sup>. This trisaccharide was found in bovine milk in one previous study<sup>32</sup>, however, it was not reported to be present as an oligosaccharide in bovine milk by the same authors in their later study<sup>34</sup>. We used chemically synthesized  $\beta$ -GlcNAc-(1 $\rightarrow$ 6)- $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc to double-check whether this trisaccharide was present. No signals at the same retention times of  $\beta$ -GlcNAc-(1 $\rightarrow$ 6)- $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc were observed in the chromatograms of human, bovine, or caprine milk (Supplementary Information). Three other trisaccharides, namely  $\alpha$ -GalNAc-(1 $\rightarrow$ 3)- $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc,  $\beta$ -GalNAc-(1 $\rightarrow$ 3)- $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc, and  $\beta$ -GalNAc-(1 $\rightarrow$ 2)-[ $\beta$ -Gal-(1 $\rightarrow$ 4)]-Glc, were identified in bovine milk in the present study. The presence of the first trisaccharide in bovine milk has been reported in many studies<sup>25, 31-34</sup>, the second trisaccharide was only reported in one study<sup>45</sup>, and the third

trisaccharide was not reported before.

The complete CID spectra used for the structural identification of these trisaccharides by using LODS/MS<sup>n</sup> and the approach for crosschecking these trisaccharide structures using the chromatogram retention time of chemically synthesized trisaccharides are illustrated in Supplementary Information.

## **(B) Tetrasaccharides**

Fig. 3 shows the chromatograms of tetrasaccharide ion  $m/z$  730 [sodium adducts of (Hex)<sub>3</sub>HexNAc] and  $m/z$  771 [sodium adducts of (Hex)<sub>2</sub>(HexNAc)<sub>2</sub>]. In addition to the tetrasaccharides  $\beta$ -Gal-(1→4)- $\beta$ -GlcNAc-(1→3)- $\beta$ -Gal-(1→4)-Glc, and  $\beta$ -Gal-(1→3)- $\beta$ -GlcNAc-(1→3)- $\beta$ -Gal-(1→4)-Glc, which have been found in human milk<sup>18,19,22,24,46</sup>, we found two unreported tetrasaccharides, namely

$\beta$ -GalNAc-(1→4)- $\beta$ -GlcNAc-(1→6)- $\beta$ -Gal-(1→4)-Glc and GalNAc-(1→3)-GlcNAc-(1→3)- $\beta$ -Hex-(1→4)-Hex, in human milk in this study.

Among the five tetrasaccharides we found in bovine milk, namely

$\beta$ -Gal-(1→4)- $\beta$ -GlcNAc-(1→3)- $\beta$ -Gal-(1→4)-Glc,

$\beta$ -Gal-(1→4)- $\beta$ -GlcNAc-(1→6)- $\beta$ -Gal-(1→4)-Glc,

$\beta$ -Gal-(1→3)- $\alpha$ -GalNAc-(1→3)- $\beta$ -Gal-(1→4)-Glc,

$\beta$ -GalNAc-(1→4)- $\beta$ -GlcNAc-(1→6)- $\beta$ -Gal-(1→4)-Glc, and

$\alpha$ -Man-(1→6)- $\beta$ -Man-(1→4)- $\beta$ -GlcNAc-(1→4)-GlcNAc, only the first one had been reported previously<sup>31,33,34</sup>. The CID spectra used for the structural identification of these tetrasaccharides by using LODS/MS<sup>n</sup> and the crosschecking of part of these tetrasaccharide structures by using the chromatogram retention time of synthesized tetrasaccharides are shown in Supplementary Information.

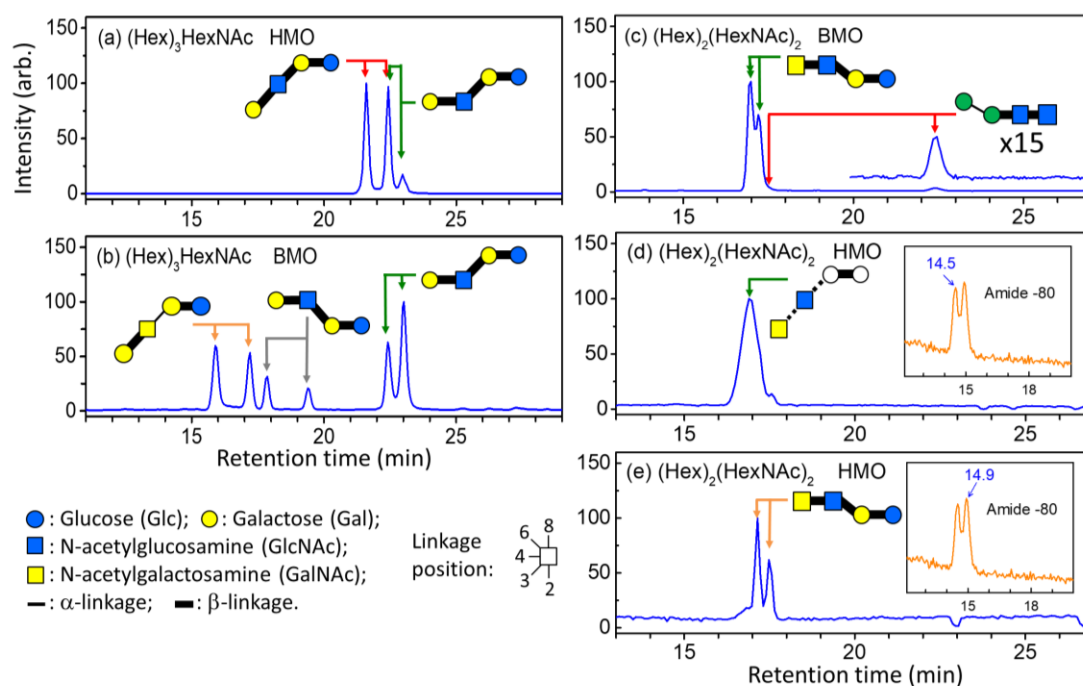


Fig. 3. PGC chromatograms of tetrasaccharides (a) and (b) ion  $m/z$  730 [sodium adducts of  $(\text{Hex})_3\text{HexNAc}$ ]; (c)-(e) ion  $m/z$  771 [sodium adducts of  $(\text{Hex})_2(\text{HexNAc})_2$ ] for human milk oligosaccharides (HMO) and bovine milk oligosaccharides (BMO). The inserts in (d) and (e) are the amide-80 chromatograms of ion  $m/z$  771, in which eluents at the retention time of 14.5 and 14.9 min were collected separately and injected into a PGC column to obtain the chromatograms shown in (e) and (f), respectively. The fraction collected from the amide-80 column was chosen such that all oligosaccharides found in this study are shown in a single chromatogram for compact display. The relative intensity in this chromatogram does not represent the relative abundance of these oligosaccharides in milk. The oligosaccharide structure for each peak in chromatograms was determined using LODS/MS<sup>n</sup>. Some of tetrasaccharide structures were cross checked using the chromatogram retention time of synthesized oligosaccharides. Details of the structural determination using LODS/MS<sup>n</sup> are described in Supplementary Information.

## Discussion

The current biosynthesis pathways of HMOs were proposed on the basis of the observation of lactose at the reducing end and linkages found in free oligosaccharides in human milk. The elongation of lactose to generate various free oligosaccharides in milk was throughout the synthesis by the following four enzymes: iGnT ( $\beta$ -GlcNAc-(1 $\rightarrow$ 3)-Gal extension), IGnT ( $\beta$ -GlcNAc-(1 $\rightarrow$ 6)-Gal branching),  $\beta$ 3Gal-T ( $\beta$ -Gal-(1 $\rightarrow$ 3)-GlcNAc type 1 termination), and  $\beta$ 4Gal-T ( $\beta$ -Gal-(1 $\rightarrow$ 4)-GlcNAc type 2 formation)<sup>2,6,9</sup>, as illustrated in Fig. 4a. Among these four enzyme, glycotransferase IGnT ( $\beta$ -GlcNAc-(1 $\rightarrow$ 6)-Gal branching) only transfers GlcNAc onto Gal after the completion of  $\beta$ -GlcNAc-(1 $\rightarrow$ 3)-Gal extension by glycotransferase iGnT. Most free oligosaccharides in human milk reported in previous studies can be synthesized in a systematic manner by using these four enzymes, except the two reported trisaccharides, namely  $\beta$ -Gal-(1 $\rightarrow$ 3)- $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc and  $\beta$ -Gal-(1 $\rightarrow$ 6)- $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc. By contrast, the number of free oligosaccharides, the structures of which have been identified in previous studies, is considerably lower in bovine or caprine milk than in human milk. Nevertheless, most of the free oligosaccharides in bovine and caprine milk reported in previous studies had lactose at the reducing end.

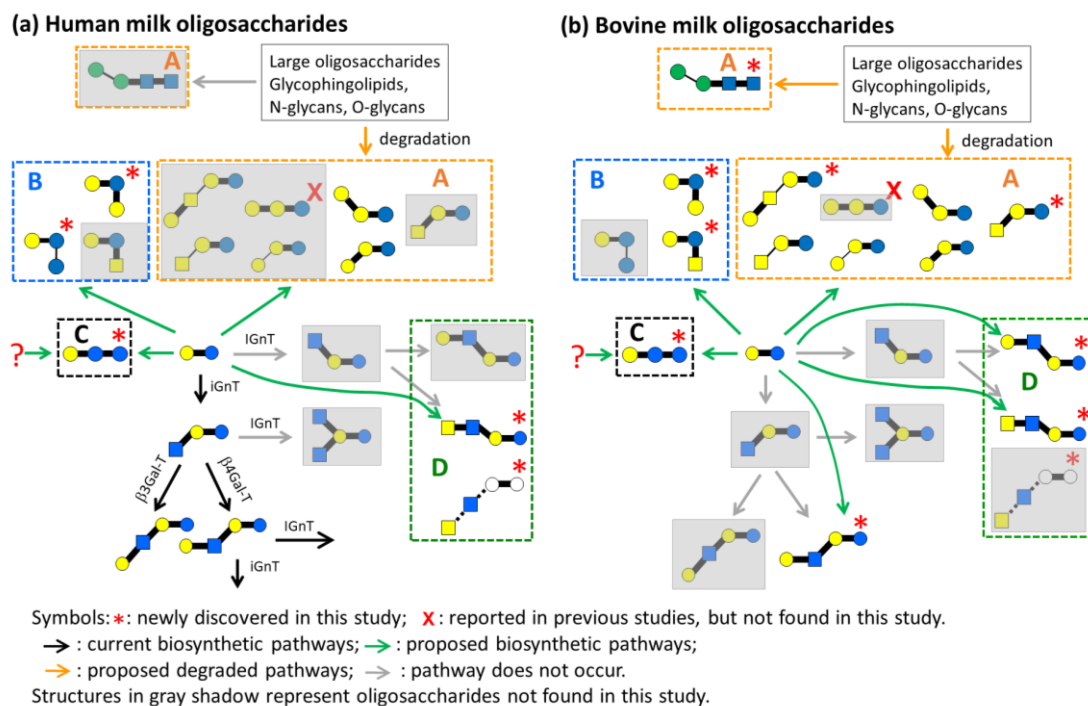


Fig. 4. Biosynthetic pathways of neutral trisaccharides and tetrasaccharides consisting glucose, galactose, mannose, N-acetylglucosamine, and N-acetylgalactosamine. (a) human milk oligosaccharides and (b) bovine milk oligosaccharides. To easily compare human and bovine milk oligosaccharides, all oligosaccharides discussed in this study are illustrated, and they are shown at the same positions in both biosynthetic pathways.

Although most oligosaccharides in milk have lactose at the reducing end, milk oligosaccharides lacking lactose at the reducing end have been reported. A trisaccharide was found in  $\alpha$ -Neu5Ac-(2→6)- $\beta$ -Gal-(1→4)-GlcNAc in caprine colostrum<sup>47</sup>; two disaccharides,  $\beta$ -Gal-(1→4)-GlcNAc and  $\beta$ -GalNAc-(1→4)-Glc, in bovine colostrum were identified<sup>48</sup>; a sialyl oligosaccharide phosphate,  $\alpha$ -NeuAc-(2→6)- $\beta$ -Gal-(1→4)-GlcNAc-P, was found in bovine colostrum<sup>49</sup>; and a tetrasaccharide,  $\alpha$ -NeuAc-(2→3)- $\beta$ -Gal-(1→3)-[ $\alpha$ -Fuc-(1→4)]-GlcNAc, and a pentasaccharide  $\alpha$ -NeuAc-(2→3)- $\beta$ -Gal-(1→3)-[ $\alpha$ -Fuc-(1→4)]- $\beta$ -GlcNAc-(1→3)-Gal were found in



human milk<sup>50</sup>. The discovery of these oligosaccharides raises the question as to how these oligosaccharides without a lactose moiety at the reducing end are synthesized. Kitagawa et al suggested two possibilities<sup>50</sup>. One possibility is that they are synthesized through the sequential transfer of various monosaccharides to N-acetylglucosamine or galactose, catalyzed by the aforementioned glycosyltransferases. Notably, except disaccharides  $\beta$ -GalNAc-(1 $\rightarrow$ 4)-Glc, all these oligosaccharides lacking lactose at the reducing end have linkages related to the aforementioned four enzymes. The other possibility is that the oligosaccharides lacking lactose are the degradation products of large free oligosaccharides by respective endoglycosidases.

The synthesis of several oligosaccharides discovered in this study cannot be explained by the two possibilities proposed by Kitagawa et al. Trisaccharides and tetrasaccharides in human milk found in this study and the corresponding biosynthetic pathways are illustrated in Fig. 4a. Oligosaccharides that follow the current biosynthetic pathway are illustrated at the end of black arrows in Fig. 4a. Other oligosaccharides that cannot be produced by the current biosynthetic pathways, as illustrated at the end of green and orange arrows in Fig. 4a, can be classified into four groups. Oligosaccharides in group A (circulated by orange dash line), which have been found in previous studies, have lactose at the reducing end. The oligosaccharides in group A can be present as moieties in other large glycans (e.g., N-glycans, O-glycans, and glycosphingolipids). They can be produced either by the elongation of lactose by undiscovered enzymes or the degradation from large glycans by other undiscovered enzymes. Oligosaccharides in group B (circulated by blue dash line) were identified in this study. Analogous to oligosaccharides in group A, oligosaccharides in group B have lactose at the reducing end. However, the linkages

of the extension from lactose is unusual: oligosaccharides in group B have a branch with the linkage of  $\beta$ -Gal-(1 $\rightarrow$ 2)-Glc or  $\alpha$ -Glc-(1 $\rightarrow$ 2)-Glc. These types of linkages are not found in any large oligosaccharides in milk, N-glycans, O-glycans, or glycosphingolipid, indicating they are not the degradation products of large oligosaccharides. A trisaccharide,  $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc- $\beta$ -Glc-(1 $\rightarrow$ 4)-Glc, found in this study is particular unusual and was classified into group C (circulated by black dash line). This trisaccharide does not have lactose at the reducing end, and it is not the moiety of any large oligosaccharides. This trisaccharide could be an extension of glucose to lactose at the reducing end or it might not be related to the extension of lactose (e.g., it starts from the extension at the nonreducing end of cellobiose instead of lactose). Group D consists of tetrasaccharides resulting from the extension of  $\beta$ -GalNAc-(1 $\rightarrow$ 4)-GlcNAc from lactose. The tetrasaccharide GalNAc-(1 $\rightarrow$ 3)-GlcNAc-(1 $\rightarrow$ 3)-Hex- $\beta$ -(1 $\rightarrow$ 4)-Hex could be the elongation of the trisaccharide  $\beta$ -GlcNAc-(1 $\rightarrow$ 3)-Gal- $\beta$ -(1 $\rightarrow$ 4)-Glc by an undiscovered enzyme. Notably, the other tetrasaccharide  $\beta$ -GalNAc-(1 $\rightarrow$ 4)- $\beta$ -GlcNAc-(1 $\rightarrow$ 6)-Gal- $\beta$ -(1 $\rightarrow$ 4)-Glc was found; however, the trisaccharide  $\beta$ -GlcNAc-(1 $\rightarrow$ 6)-Gal- $\beta$ -(1 $\rightarrow$ 4)-Glc was not found. These findings indicate that tetrasaccharide  $\beta$ -GalNAc-(1 $\rightarrow$ 4)- $\beta$ -GlcNAc-(1 $\rightarrow$ 6)-Gal- $\beta$ -(1 $\rightarrow$ 4)-Glc was not synthesized through trisaccharide  $\beta$ -GlcNAc-(1 $\rightarrow$ 6)-Gal- $\beta$ -(1 $\rightarrow$ 4)-Glc. Instead of the addition of only one monosaccharide in each step, transfer of the disaccharide  $\beta$ -GalNAc-(1 $\rightarrow$ 4)-GlcNAc onto lactose occurs.

Trisaccharides and tetrasaccharides in bovine milk and the corresponding biosynthetic pathways are shown in Fig. 4(b). The biosynthetic pathways of oligosaccharides found in human milk considerably differ from those of free oligosaccharides in bovine milk. Fig. 4(b) shows that the

trisaccharide  $\beta$ -GlcNAc-(1 $\rightarrow$ 3)-Gal- $\beta$ -(1 $\rightarrow$ 4)-Glc, which is an important precursor trisaccharide for subsequent synthesis in human milk free oligosaccharides, was not found in bovine milk. Many tetrasaccharides found in bovine milk are not synthesized through trisaccharides. Instead of the step-by-step addition of monosaccharides, they are likely to be synthesized through the addition of disaccharides onto the nonreducing end of lactose. Many “unusual” free oligosaccharides in human milk were found in bovine and caprine milk. In particular, oligosaccharides in group A of bovine and caprine milk have more varieties than do those in human milk. The observation of these unusual oligosaccharides suggests that many undiscovered glycosyltransferases and glycosidases are involved in biosynthetic pathways. The structures of these oligosaccharides found in this study provide crucial clues relevant to the search for undiscovered enzymes and the modification of current biosynthetic pathways.

## **Online Methods**

The method used to extract free oligosaccharides from milk was similar to that described in our previous study<sup>37</sup>. Bovine milk was purchased from Experimental Farm, College of Biosources and Agriculture, National Taiwan University, Taiwan. Caprine milk was purchased from a local market, and human milk was obtained from donors. Folch solution and ethanol were used to remove fat and proteins in milk, respectively. First, Folch solution was added to milk in a centrifuge tube and mixed using a vortex mixer (VTX-3000, Mixer Uzusio, Tokyo, Japan) for 1 min. The mixture was centrifuged for 30 min at 4000 $\times$  g and 4°C (High-speed Micro Refrigerated Centrifuge, CF15RN, Hitachi, Japan). Subsequently, the top layer of the aqueous solution containing oligosaccharides was collected, and ethanol (−20°C) was added to the collected aqueous solution. The solution of ethanol/aqueous mixture was

maintained at  $-20^{\circ}\text{C}$  overnight. The mixture was centrifuged for 30 min at  $4000\times g$  and  $4^{\circ}\text{C}$ , and the top oligosaccharide-rich layer was collected and dried in vacuum. Subsequently, the samples were further purified using two solid phase extractions (SPEs). The first SPE involved using C18 cartridges (C18 SPE column, 2000 mg/12 mL, S\*Pure Pte. Ltd, Singapore) to remove remaining lipids and proteins, and the second SPE involved using PGC (porous graphitized carbon, 1000 mg/15 mL, S\*Pure Pte, Ltd, Singapore) cartridges to remove lactose. In the second SPE, oligosaccharides without sialic acid were collected by eluting the cartridge by using one column volume of 20% ACN in deionized (DI) water. The collected sample was dried in vacuum and then dissolved in DI water for subsequent separation.

After SPE, oligosaccharides were size-selected through size exclusion chromatography (TOYOPEARL HW-40F, Tosoh Bioscience GmbH, Griesheim, Germany) followed by separation using an HPLC system (Dionex Ultimate 3000, Thermo Fisher Scientific, Waltham, MA USA) with a TSKgel amide-80 column ( $150\text{ mm} \times 2.0\text{ mm}$ , particle size of  $5\text{ }\mu\text{m}$ ; Tosoh Bioscience GmbH, Griesheim, Germany). Finally, each fraction collected from the amide-80 column was separately injected into another HPLC with a PGC Hypercarb column ( $2.1\text{ mm} \times 100\text{ mm}$ , particle size of  $3\text{ }\mu\text{m}$ ; Thermo Fisher Scientific, Waltham, MA, USA) for further separation. For trisaccharides, methanol was added to the eluents from the PGC column, and the mixture was sent into a linear ion trap mass spectrometer (LTQ XL, Thermo Fisher Scientific, Waltham, MA USA) directly for online structural determination. For tetrasaccharides, eluents from the PGC column were sent into a fraction collector (FC204, Gilson, Middleton, WI, USA). Fractions collected from the fraction collector were vacuum dried. Subsequently, the sample was dissolved in a 50:50 (vol/vol) water/methanol mixture and sent into a nanoelectrospray mass spectrometer for

structural determination. Detailed settings of the mass spectrometer are described in Supplementary Information.

## References

1. Kunz, C., Rudloff, S., Baier, W., Klein, N. & Strobel, S. Oligosaccharides in human milk: Structural, functional, and metabolic aspects. *Ann. Rev. Nutr.* **20**, 699–722 (2000).
2. Bode, L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22**, 1147–1162 (2012).
3. Newburg, D.S., Ruiz-Palacios, G.M. & Morrow, A.L. Human milk glycans protect infants against enteric pathogens. *Ann. Rev. Nutr.* **25**, 37–58 (2005).
4. Brew, K. & Hill, R.L. Lactose biosynthesis. *Rev Physiol Biochem Pharmacol.* **72**, 105–58 (1975).
5. Ramakrishnan, B., Boeggeman, E., Qasba, P.K.  $\beta$ 1,4-galactosyltransferase and lactose synthase: Molecular mechanical devices. *Biochem Biophys Res Commun.* **291**, 1113–1118 (2002).
6. Kobata A. Possible application of milk oligosaccharides for drug development. *Chang Gung Med. J.* **26**, 621–636 (2003).
7. Urashima, T., Asakuma, S. & Messer, M. Milk oligosaccharide. *Comprehensive Glycolscience From Chemistry To Systems Biology*. (Elsevier Ltd. Oxford, 2007).
8. Blank, D., Dotz, V., Geyer, R. & Kunz, C. Human milk oligosaccharides and Lewis blood group: individual high-throughput sample profiling to enhance conclusions from functional studies. *Adv. Nutr.* **3**, 440S–449S (2012).
9. Urashima, T., Hirabayashi, J., Sato, S. & Kobata, A. *Trends Glycosci. Glycotech.* **30**, SE51–SE65 (2018).
10. Ninonuevo, M.R. et al. A strategy for annotating the human milk glycome. *J. Agric. Food Chem.* **54**, 7471–7480 (2006).
11. Mehra, R. & Kelly, P. Milk oligosaccharides: structural and technological aspects. *Int. Dairy J.* **16**, 1334–1340A (2006).

12. Ruhaak, L.R. & Lebrilla, C.B. Advances in analysis of human milk oligosaccharides. *Adv. Nutr.* **3**, 406S–414S (2012).
13. Urashima, T., Taufik, E., Fukuda, K. & Asakuma, S. Recent advances in studies on milk oligosaccharides of cow and other domestic farm animals. *Biosci. Biotechnol. Biochem.* **77**, 455–466 (2013).
14. Oliveira, D., Wilbey, R.A., Grandison, A.S. & Iro, L.R. Milk oligosaccharides: A review. *Int. J. Dairy Tech.* **68**, 305–321 (2015).
15. Dong, X., Zhou, S. & Mechref, Y. LC-MS/MS analysis of permethylated free oligosaccharides and N-glycans derived from human, bovine, and goat milk samples. *Electrophoresis* **37**, 1532–1548 (2016).
16. Kuhn, R. & Baer, H.H. Die Konstitution der Lacto-N-tetraose. *Chem. Ber.* **89**, 504–511 (1956).
17. Kuhn, R. & Gauhe, A. Die Konstitution der Lacto-N-neotetraose. *Chem. Ber.* **95**, 518–522 (1962).
18. Kuhn, R., Baer, H. H., and Gauhe, A. Kristallisierte Fucosido-lactose. *Chem. Ber.* **89**, 2513–2513 (1956).
19. Montreuil, J. Structure de deux triholosides isoles du lai de femme. *C. R. Acad. Sci.* **242**, 192–193 (1956).
20. Yamashita, K., and Kobata, A. Oligosaccharides of human milk V. Isolation and characterization of a new trisaccharide, 6'-galactosyllactose. *Arch. Biochem. Biophys.* **161**, 164–170 (1974).
21. Duus, J.O., Gotfredsen, C.H. & Bock, K. Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem. Rev.* **100**, 4589–4614 (2000).
22. Dabrowski, U., Egge, H. & Dabrowski, J. Proton-nuclear magnetic resonance study of peracetylated derivatives of ten oligosaccharides isolated from human

- milk. *Arch. Biochem. Biophys.* **224**, 254–260 (1983).
23. Saito, T., Ito, T. & Adachi, S. Chemical structure of three neutral trisaccharides isolated in free form from bovine colostrum. *Carbohydr. Res.* **165**, 43–51 (1987).
  24. Donald, A.S.R. & Feeney, J. Separation of human milk oligosaccharides by recycling chromatography. First isolation of lacto-N-neo-difucohexose II and 3' galactosyllactose from this source. *Carbohydr. Res.* **178**, 79–91. (1988).
  25. Urashima, T., Saito, T., Ohmisya, K. & Shimazaki, K. Structural determination of three neutral oligosaccharides in bovine (Holstein-Friesian) colostrum, including the novel trisaccharide; GalNAc $\alpha$ 1-3Gal $\beta$ 1-4Glc. *Biochim. Biophys. Acta* **1073**, 225–229 (1991).
  26. Urashima, T., Bubb, W.A., Messer, M., Tsuji, Y. & Taneda, Y. Studies of the neutral trisaccharides of goat colostrum and of the one- and two-dimensional <sup>1</sup>H and <sup>13</sup>C NMR spectra of 6'-N-acetylglucosaminylactose. *Carbohydr. Res.* **269**, 173–84 (1994).
  27. Dell, A. & Morris, H.R. Glycoprotein structure determination by mass spectrometry. *Science* **291**, 2351–2356 (2001).
  28. Ashline, D., Singh, S., Hanneman, A. & Reinhold, V. Congruent strategies for carbohydrate sequencing. 1. Mining structural details by MS<sub>n</sub>. *Anal. Chem.* **77**, 6250–62 (2005).
  29. Pfenninger, A., Karas, M., Finke, B. & Stahl, B. Structural analysis of underivatized neutral human milk oligosaccharides in the negative ion mode by nano-electrospray MS<sub>n</sub>. *J. Am. Soc. Mass Spectrom.* **13**, 1331–1340 (2002).
  30. Tao, N., DePeters, E.J., Freeman, S., German, J.B., Grimm, R. & Lebrilla, C.B. Bovine Milk Glycome. *J. Dairy Sci.* **91**, 3768–3778 (2008).
  31. Marino, K. et al. Method for milk oligosaccharide profiling by 2-aminobenzamide labeling and hydrophilic interaction chromatography. *Glycobiology* **21**,



1317–1330 (2011).

32. Aldredge, D.L. et al. Annotation and structural elucidation of bovine milk oligosaccharides and determination of novel fucosylated structures. *Glycobiology* **23**, 664–676 (2013).
33. Albrecht, S. et al. A comparative study of free oligosaccharides in the milk of domestic animals. *Br. J. Nutr.* **111**, 1313–1328 (2014).
34. Lee, H., Cuthbertson, D.J., Otter, D.E. & Barile, D. Rapid screening of bovine milk oligosaccharides in a whey permeate product and domestic animal milks by accurate mass database and tandem mass spectral library. *J. Agricultural Food Chem.* **64**, 6364–6374 (2016).
35. Hsu, C.H., Liew, C.Y., Huang, S.P., Tsai, S.T. & Ni, C.K. Simple approach for de novo structural identification of mannose trisaccharides. *J. Am. Soc. Mass Spectrom.* **29**, 470–480 (2018).
36. Hsu, C. H., Liew, C. Y., Huang, S. P., Tsai, S. T. & Ni, C. K. Simple method for de novo structural determination of underivatized glucose oligosaccharides. *Sci. Rep.* **8**, 5562 (2018).
37. Tsai, S.T., et al. Automatic full glycan structural determination through logically derived sequence tandem mass spectrometry. *ChemBioChem* **20**, 2351–2359 (2019).
38. Hsu, C.H., Huang, S.P., Liew, C.Y., Tsai, S.T. & Ni, C.K. De novo structural determination of mannose oligosaccharides by using a logically derived sequence for tandem mass spectrometry. *Anal. Bioanal. Chem.* **411**, 3241–3255 (2019).
39. Chen, J.L., et al. Collision-induced dissociation of sodiated glucose and identification of anomeric configuration. *Phys. Chem. Chem. Phys.* **19**, 15454–15462 (2017).
40. Hai, T.H., et al. Collision-induced dissociation of sodiated glucose, galactose, and

- mannose, and the identification of anomeric configuration. *Phys. Chem. Chem. Phys.* **20**, 19614–19624 (2018).
41. Chiu, C.C., et al. Unexpected dissociation mechanism of sodiated N-acetylglucosamine and N-acetylgalactosamine. *J. Phys. Chem. A* **123**, 3441–3453 (2019).
  42. Domon, B. & Costello, C.E. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **5**, 397–409 (1988).
  43. Kimura, K., Watanabe, Y., Matsumoto, K. & Miyagi, A. *Yakult Reports* **17**, 1–7 (1997).
  44. Sugawara, M. & Idota, T. *Proceedings of the annual meeting of Japan society for bioscience, biotechnology, and agrochemistry*. Sapporo, Japan, p.132 (1995).
  45. Watanabe, Y., Kimura, K., Nakamura, T., Kume, S. & Urasima T. *Proceedings of the 23rd International Carbohydrate Symposium*. Whistler, Canada, p.78 (2006).
  46. Kuhn, R., & Gauhe, A. Über die Lacto- Difuco- Tetraose Der Frauenmilch. Ein Beitrag zur Strukturspezifität der Blutgruppensubstanz Le. *Justus Liebigs Ann. Chem.* **611**, 249–252 (1958).
  47. Urashima, T., Murata, S. & Nakamura, T. Structural determination of monosialyl trisaccharides obtained from caprine colostrum. *Comp. Biochem. Physiol.* **116B**, 431–435 (1997).
  48. Saito, T., Itoh, T. & Adachi, S. Presence of two neutral disaccharides containing N-acetylhexosamine in bovine colostrum as free forms. *Biochim. Biophys. Acta* **801**, 147–150 (1984).
  49. Parkkinen, J. & Finne, J. Occurrence of N-acetylglucosamine 6-phosphate in complex. *J. Biol. Chem.* **260**, 10971–10975 (1985).
  50. Kitagawa, H. et al. Occurrence of tetra- and pentasaccharides with the sialyl-Le<sup>a</sup>

structure in human milk. *J. Biol. Chem.* **265**, 4859–4862 (1990).

## **Acknowledgements**

This work was financially supported by thematic research project (AS-107-TP-A08) and iMATE (AS-iMATE-109-32), Academia Sinica, Taiwan.

## **Author Information**

### **Affiliations**

Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, 10617, Taiwan

Wei-Chien Weng, Hun-En Liao, Shih-Pei Huang, Shang-Ting Tsai, Hsu-Chen Hsu,

Chia Yen Liew, Chi-Kung Ni

Molecular Science and Technology, International Graduate Program, Academia Sinica and National Tsing Hua University, Hsinchu, 30013, Taiwan

Wei-Chien Weng, Chi-Kung Ni

International Graduate Program of Molecular Science and Technology, National Taiwan University and Taiwan International Graduate Program of Molecular Science and Technology, Academia Sinica, Taipei, 10617, Taiwan

Chia Yen Liew

Department of Applied Chemistry, National Chiao Tung University, Hsinchu, 300, Taiwan

Hun-En Liao

Genomic Research Center, Academia Sinica, Taipei, 11529, Taiwan

Veeranjaneyulu Gannedi, Shang-Cheng Hung

### **Contributions**

W.C.W. extracted and analyzed bovine and human milk oligosaccharides, Hun-En Liao extracted and analyzed caprine milk oligosaccharides; S.T.T. built the

monosaccharide database; S.P.H., H.S.H., and C.Y.L. built the disaccharide database. C.K.N. designed the experiment, conceived the LODS/MS<sup>n</sup> method, analyzed data, wrote the manuscript and supplementary information; V.G. and S.C.H. synthesized the trisaccharides,  $\beta$ -Gal-(1 $\rightarrow$ 2)-[ $\beta$ -Gal-(1 $\rightarrow$ 4)]-Glc,  $\beta$ -Gal-(1 $\rightarrow$ 4)- $\beta$ -Glc-(1 $\rightarrow$ 4)-Glc,  $\alpha$ -Glc-(1 $\rightarrow$ 2)-[ $\beta$ -Gal-(1 $\rightarrow$ 4)]-Glc,  $\beta$ -GalNAc-(1 $\rightarrow$ 2)-[ $\beta$ -Gal-(1 $\rightarrow$ 4)]-Glc, and  $\beta$ -GlcNAc-(1 $\rightarrow$ 6)- $\beta$ -Gal-(1 $\rightarrow$ 4)-Glc, and measured and assigned the corresponding NMR spectra.

Corresponding authors

Corresponding to Chi-Kung Ni

### **Competing Interests**

S.T.T., H.C.H., S.P.H., C.Y.L. and C.K.N. are co-inventors of a United States patent (US 10,796,788 B2) that part of the method described in the patent to determine the carbohydrate structure was used in this work. All other authors declare no competing interests.