Yu et al, Supplemental Document

Human Milk Contains Novel Glycans That Are Potential Decoy Receptors for Neonatal Rotavirus

Ying Yu[1], Yi Lasanajak[1], Xuezheng Song[1], Liya Hu[2], Sasirekha Ramani[3], Megan L. Mickum[1], David J. Ashline[4], B. V. Venkataram Prasad[2,3], Mary K. Estes[3], Vernon N. Reinhold[4], Richard D. Cummings[1], and David F. Smith[1]

[1]Department of Biochemistry and the National Center for Functional Glycomics, Emory University School of Medicine, Atlanta, GA 30322; [2]Verna and Marrs McLean Department of Biochemistry and Molecular Biology and [3]Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX; and [4]Glycomics Center, University of New Hampshire, Durham, NH

There are two sets of glycans on the RV-MAGS subarray used for structural analysis (**Table 1**). The controls are 14 defined glycans (glycans 33-46) that include many of the typical structural motifs in human milk shown as in **Fig. S4**. The results of lectin and antibody binding to these glycans can be used not only to confirm structural predictions but also to monitor the behavior of reagents. The panel of unknowns, containing 32 human milk glycans (glycans 1-32 in **Table 1**) was selected because they were ligands for rotavirus attachment proteins. The RV-MAGS subarray for structural analysis was interrogated with lectins and antibodies before and after treatment by the exoglycosidases (**Table 1**). From the analyses of the large amount of binding data shown in Table 1 and Supplemental Table S7, we were able to predict the structures (**Fig. 6**) for most of the unknown glycans as described below. The identification numbers of HMGs containing structures not reported in recent reviews (1-3) are highlighted in blue.

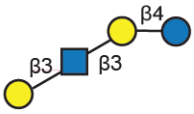
Defined Glycans

With the exception of #43 LDFT (Le^y) and #45 (Le^y-Le^x), all of the glycans printed on the RV-MAGS subarray used for structural analysis are AEAB derivatives of human milk glycans (known and unknown) and were generated by reductive amination. The nomenclature used below will include a reducing terminal glucose, but this glucose is reduced in the actual structure of the AEAB derivatives. The LDFT (#43) and Le^y-Le^x glycan (#45) were not from human milk, but were synthesized from a glycoside and

obtained from the CFG. The glycan identified as LNFP-IV (#42), which has not been described in human milk free glycans, was enzymatically synthesized from LNnT by the action of a commercial preparation of FUT1 (R & D Systems) with GDP-Fucose.

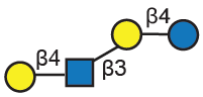
Glycan 33 (Agal LNT, Lacto-N-Triose) – GlcNAcβ1-3Galβ1-4Glc, is a trisaccharide prepared from LNnT by digestion with Galβ1-4-specific galactosidase. It was chosen as a positive control for GSL-II binding because GlcNAc is located at the non-reducing end, and among the lectins and antibodies tested, only GSL-II binds to this glycan.

Glycan 34 (LNT, lacto-N-tetraose) – Galβ1-3GlcNAcβ1-3Galβ1-4Glc, does not bind any lectin as there is no lectin available for the type 1 motif (Galβ1-3GlcNAc). Weak binding by RCA-I at 10 µg/ml is

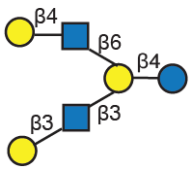


observed, but it does not bind ECL at all. This is consistent with the strict specificity of ECL for Galβ1-4GlcNAc and that the low affinity of RCA-I for terminal Gal in a β1-3 linkage to GlcNAc. The structure of LNT is identified by an anti-type 1 glycan antibody (6A7), and the binding of GSL-II is observed only after treatment by the β1-3 specific galactosidase, but not after Galβ1-4-specific galactosidase digestion.

Glycan 35 (LNnT, lacto-N-neotetraose) - Galβ1-4GlcNAcβ1-3Galβ1-4Glc, is a type 2 (Galβ1-4GlcNAc) glycan and is bound by RCA-I and ECL. The anti-Type 1 antibody (6A7) does not bind this glycan indicating the strict specificity of this antibody. In contrast to LNT, Galβ1-3- galactosidase does not change the binding properties of LNnT, while the Galβ1-4- galactosidase treatment generates GSL-II binding due to the loss of terminal galactose and exposure of the terminal GlcNAcβ1-3Gal.

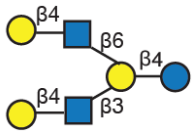


Glycan 36 (LNH, lacto-N-hexose) – Galβ1-4GlcNAcβ1-6(Galβ1-3GlcNAcβ1-3)Galβ1-4Glc), is a

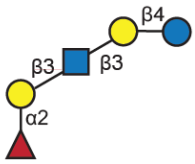


branched glycan with a type 2 sequence on the 6-branch and a type 1 sequence on the 3-branch. Therefore, this glycan is strongly bound by RCA-I and ECL due to the presence of the type 2 branch and by the anti-Type 1 antibody due to the presence of the type 1 branch. Consistent with this structure, both Galβ1-3-galactosidase and Galβ1-4- galactosidase treatment result in GSL-II binding. However, the signal from Galβ1-4- galactosidase digestion is much lower than the signal from Galβ1-3- galactosidase because GSL-II prefers GlcNAc on the 3-branch and binds weakly to the GlcNAc on the 6-branch.

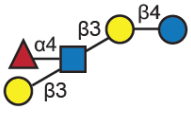
Glycan 37 (LNnH, lacto-N-neohexose) – Galβ1-4GlcNAcβ1-6(Galβ1-4GlcNAcβ1-3)Galβ1-4Glc), is a branched glycan with the type 2 sequence on both the 6-branch and the 3-branch. Therefore, this glycan is strongly bound by RCA-I and ECL and is not bound by the anti-Type 1 antibody. In comparison with glycan #36 (LNH), the ECL binding for LNnH at low concentration (1 µg/ml) is significantly higher than that of LNH, which indicates that ECL has lower affinity for the type 2 structure on the 6-branch. In addition, Galβ1-4- galactosidase of glycan #37 treatment leads to strong GSL-II binding.



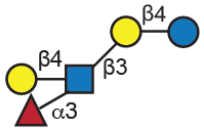
Glycan 38 (LNFP I, lacto-N-fucopentaose I) – Fucα1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glc, shows strong binding by AAL, indicating the presence of Fucose. Strong binding by the anti-H1 antibody is consistent with the known structure of #38. Interestingly, this glycan is also bound by the anti-type 1 antibody indicating that this antibody is not affected by the 2-substitution of a type 1 glycan, which is important in interpreting the structures defined by these reagents. No other binding is observed for LNFP I.



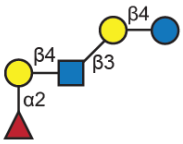
Glycan 39 (LNFP II, lacto-N-fucopentaose II) – Galβ1-3(Fucα1-4)GlcNAcβ1-3Galβ1-4Glc is a Lewis^{a} pentose, which is strongly bound by AAL and an anti-Le^{a} antibody, but is also weakly bound by the anti-Le^{b} and anti-Le^{x} antibodies due to their cross-reactivity. The structure is galactosidase resistant due to the presence of Fucose on the penultimate GlcNAc.



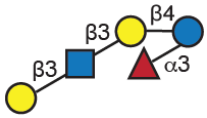
Glycan 40 (LNFP III, lacto-N-fucopentaose III) – Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ1-4Glc is a Lewis^x pentose. It is strongly bound by AAL and the two Lewis x-recognizing antibodies. The anti-Le^x (5F1) antibody is a monoclonal antibody generated in our lab and the CD15 is a commercially available antibody. The specificity of these two antibodies were screened on the CFG array and displayed in **Fig. S5.**



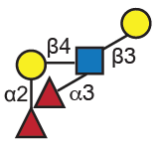
Glycan 41 (LNFP IV, lacto-N-fucopentaose IV) – Fucα1-2Galβ1-4GlcNAcβ1-3Galβ1-4Glc is an H type 2 pentose and is recognized by two lectins AAL and UEA-I. The binding to ECL is due to the cross-reactivity of ECL as this has been observed in CFG array as well. This glycan has not been described in human milk and we designated it lacto-N-fucopentaose IV**.**



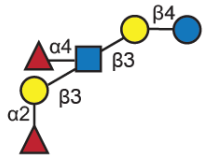
Glycan 42 (LNFP V, lacto-N-fucopentaose V) – Galβ1-3GlcNAcβ1-3Galβ1-4(Fucα1-3)Glc is a pentaose with an α1-3-fucose linked to the glucose at the reducing end. In agreement with its structure, this glycan is bound by AAL and the anti-type 1 antibody. Similar to the LNT, weak binding to RCA-I is also observed and GSL-II binding is revealed after Galβ1-3-galactosidase digestion but not after Galβ1-4-galactosidase digestion.



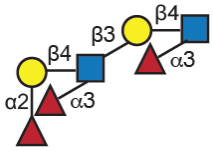
Glycan 43 (LDFT or Le^y) The Lewis^y motif, Fucα1-2Galβ1-4(Fucα1-3)GlcNAcβ1-3Gal, does not have lactose at the reducing end like the milk oligosaccharides and is not found in human milk. This structure, obtained from the CFG, contains two fucoses and therefore shows strong binding by AAL. Glycan #43 binds lectin UEA-1, which is specific for Fucα1-2Galβ1-4GlcNAc. No other lectin or antibody binding is observed for this glycan.



Glycan 44 (LNDFH I, lacto-N-difucohexaose I) – Fucα1-2Galβ1-3(Fucα1-4)GlcNAcβ1-3Galβ1-4Glc, contains a Lewis^{b} moiety and is bound by AAL as expected. However, it also consistently shows low binding toward GSL-II, which is known to recognize the terminal GlcNAcβ1-3Gal structure. Since there is no exposed GlcNAc in the LNDFH-I, we assume the GSL-II binding is due to the cross reaction of the lectin. The anti-Le^{a} antibody also has a strong cross reaction with this glycan, which is not surprising since the Le^{b} structure contains a Le^{a} determinant. However, the anti-Le^{b} antibody does not cross react with Glycan #39, the Le^{a} glycan.

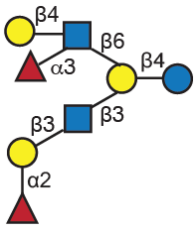


Glycan 45 (Le^{y}-Le^{x}) – Le^{y}-Le^{x}, Fucα1-2Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ1-4(Fucα1-3)Glc, has a Le^{y} moiety attached to a Le^{x} moiety. It shows strong binding toward AAL, which is consistent with the two



Fucose residues. Although UEA-1 is reported as a Le^{y}-recognizing lectin, it binds to this glycan with very low signal. However, the anti-Ley antibody binds to this structure with strong signal (**Table S7**). In addition, the underlying Le^{x} structure is bound by our anti-Lex antibody at high concentration, but not at low antibody concentration (**Table S7**).

Glycan 46 (DFLNH) - Galβ1-4(Fucα1-3)GlcNAcβ1-6(Fuca1-2Galβ1-3GlcNAcβ1-3)Galβ1-4Glc is a branched structure with a Lewis x motif on the 6-branch and an H type-1 determinant on the 3-branch. DFLNH is strongly bound by AAL, as well as the anti-Type 1 antibody due to its cross-reactivity with H type 1. Noticeably, the H-type 1 structure is not bound by the anti-H1 antibody at low concentration (1:100 dilution). Weak and moderate binding is only observed when the antibody concentration is increased to 1:10 and 1:2 dilution, respectively (**Table S7**), which suggested that this antibody either does not prefer the H1 structure on the 3-branch or the presence of a nearby Le^{x} structure affects the antibody

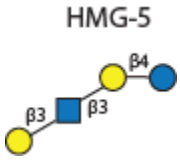


recognition. Similarly, the Le^x structure on the 6-branch is only bound by the anti-Le^x antibody (5F1) at very high concentration (no dilution, **Table S7**), also indicating that these antibodies do not prefer the Le^x on a branched structure.

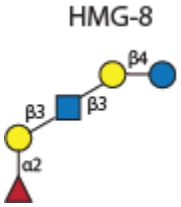
Unknown Glycans

The 32 unknown HMGs include seven glycans bound by VP8*N155 (HMG-13, 14, 18, 21, 27, 28 and 33), four glycans bound by VP8*N155 after they were treated by α1-2-fucosidase digestion (HMG-29, 34, 51 and 60), eleven glycans bound by VP8*RV3 (HMG-5, 8, 37, 41, 48, 49, 55, 56, 62, 69 and 76) and ten glycans bound by VP8*B223 (HMG-16, 20, 23, 31 45, 47, 54, 65, 66 and 67). Many of these glycans have not been previously described as free glycans in human milk based on recent reviews [1-3], and the structures as determined based on the descriptions below are represented in **Fig. 6**.

Glycan 1 (HMG-5) – HMG-5 has a molecular mass of 871.416 [M+H]+ that is consistent with a tetrasaccharide containing three Hex and one HexNAc (H3N1). Because there are only two possible structures for H3N1 composition in the free glycans of human milk, LNT and LNnT, and since this glycan displays the same binding profile as the control glycan #34, HMG-5 was assigned the structure of LNT, lacto-N-tetraose.



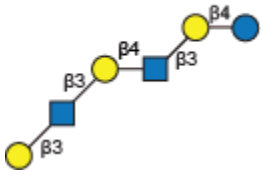
Glycan 2 (HMG-8) – Since HMG-8 has the same molecular mass 1017.506 [M+H]+ and binding profile as control glycan #38, HMG-8 was assigned the structure of LNFP I, lacto-N-fucopentaose I.



Glycan 3 (HMG-13) – HMG-13 has a molecular mass of 1236.828 [M+H]+ that matches a composition of Hex4HexNAc2, and its HPLC profile shows a symmetrical peak (**Fig. S3**). For human milk glycans, a structure with a H4N2 composition may be either a linear or a branched hexasaccharide (LNH, glycan #36 and LNnH, glycan #37). HMG-13 is strongly bound by the anti-type 1 antibody and very weakly by RCA-I and not at all by ECL, indicating that this glycan only has terminal type 1 and no terminal type 2.

This assumption is also supported by the gain of GSL-II binding after Galβ1-3-galactosidase digestion. Because branched HMGs always have a type 2 structure on the 6-branch, the lack of a terminal type 2 determinant indicates HMG-13 is a linear glycan with a terminal type 1determinant. The β1-4 linkage of the internal Gal to GlcNAc is revealed by the specific endo-β1-4-galactosidase digestion, which results in the gain of GSL-II binding after internal cleavage. Based on these results, we predict HMG-13 is the linear hexasaccharide. This structure was confirmed by MS^n analysis (**see companion report, Ashline et al.**) to be *para*-Lacto-N-hexaose (*p*-LNH) [1-3].

HMG-13



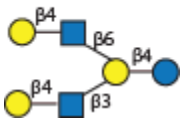
Glycan 4 (HMG-14) – HMG-14 also has a composition of H4N2 based on its molecular ion 1248.665[M+H]+. Unlike HMG-13, however, it is bound by RCA-I and ECL, and not bound at all by the anti-type 1 antibody. Galβ1-4-galactosidase digestion, but not Galβ1-3-galactosidase digestion reveals GS-II binding to a terminal GlcNAc. Taken together these data indicate that this glycan only has terminal type 2 determinant(s); not type 1. This H4N2 HMG must be either a linear structure terminating in a type 2 determinant or a branched structure containing two type 2 determinants (LNnH). However, glycan #37 (authentic LNnH) is not bound by viral protein N155 and it is resistant to the endo-β1-4-galactosidase digestion, while HMG-14 is a N155 binder and is bound by GSL-II after endo-β1-4-galactosidase treatment. Thus, we predict HMG-14 is a linear hexasaccharide with a type 2 terminal determinant, and our prediction is in agreement with the result from tandem MS analysis. This structure was confirmed by MS^n analysis (**see companion report, Ashline et al.**) to be *para*-Lacto-N-*neo*hexaose [1, 3].

HMG-14

Glycan 5 (HMG-16) – HMG-16 is another glycan with a composition of H4N2 (1236.746 [M+H]+). Its retention time on HPLC and its binding profile are different from the other two H4N2 glycans (HMG-13 and 14). It is bound by RCA-I and ECL, as well as by GSL-II after β1-4-galactosidase digestion, but not after β1-3-galactosidase digestion, indicating the exclusive presence of type 2 termini. As discussed above, such a glycan is either a linear hexasaccharide or a branched LNnH. The negative binding to GSL-II after endo-β1-4-galactosidase treatment further excludes the possibility of an HMG-14 structure. Therefore,

HMG-16 is either a linear glycan with an internal Galβ1-3GlcNAc structure or LNnH. Since a linear hexaose with an internal Galβ1-3GlcNAc structure has never been described in human milk and since the binding profile of HMG-16 is the same as the control LNnH we predicted, this abundant (171 nmol, **Table S3**) glycan is LNnH, lacto-N-*neo*hexose. This structure was confirmed by MS^n analysis (**see companion report, Ashline et al.**) to be Lacto-N-*neo*Hexaose (LNnH) [1-3].

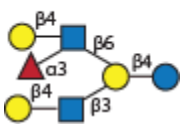
HMG-16



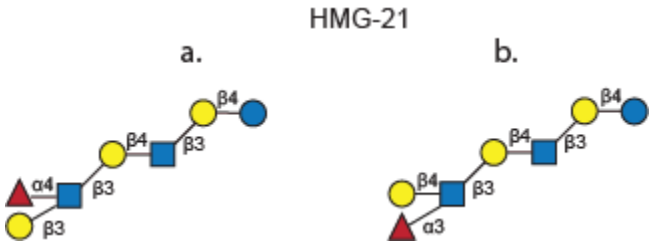
Glycan 6 (HMG-18) – HMG-18 displays the same molecular ion (1236.521 [M+H]+), HPLC retention time and lectin binding profile as HMG-14. We therefore predicted it to have the same structure as HMG-14. This structure was confirmed by MS^n analysis (**see companion report, Ashline et al.**) to be *para*-Lacto-N-*neo*hexaose [1, 3].

Glycan 7 (HMG-20) – HMG-20 has a molecular mass of 1382.806 [M+H]+ consistent with a composition of Hex4HexNAc2Fuc1 (H4N2F1) and the HPLC profile shows a single symmetrical peak. This glycan is bound by AAL, consistent with the mass data. The strong binding to RCA-I and ECL indicate the presence of a terminal type 2 determinant, which is also confirmed by the GSL-II binding after β1-4-galactosidase digestion. This latter observation also indicates that the type 2 determinant is located on the 3-arim since GS-II does not bind to a terminal GlcNAc on the 6-arm. Additionally, the anti-Le^x antibody (5F1) binds HMG-20 weakly at low concentration (1:20 dilution, **Table 1**) and stronger at high concentration (no dilution, **Table S7**) indicating a Le^x structure on the 6-branch. This conclusion is based on the preference of this antibody for terminal Le^x determinants and the similar binding pattern of this antibody to control glycan DFLNH (**Table S7**). Therefore, we predicted HMG-20 has an LNnH core with Le^x located on the 6-arm. This structure was confirmed by MS^n analysis (**see companion report, Ashline et al.**) to be a fucosylated LNnH, but the MS^n analysis could not confirm the location of the Le^x determinant on the 6-arm. This isomer of F-Lacto-N-*neo*hexaose has not been previously identified due to the inability to determine the position of the Fucose among the 7 previously described HMGs with this intact mass [1-3].

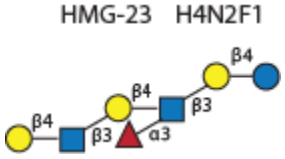
HMG-20



Glycan 8 (HMG-21) – The MALDI profile shows HMG-21 has a single mass 1405.311[M+Na]+ that is consistent with H4N2F1 composition. However, HPLC profile shows an asymmetrical peak, which is presumably due to the presence of other isomers. The positive binding by anti-Le^a antibody indicates there is a terminal Le^a determinant, while the binding by anti-Le^x antibodies 5F1 and CD15 indicate there is also a terminal Le^x determinant. Because MALDI analysis of HMG-21 indicates only one Fucose, the Le^a and Le^x structures must be from two distinct isomers, which is consistent with the HPLC data. Since the negative binding by ECL and anti-type 1 antibody suggest the absence of terminal type 1 and type 2 motif, we predicted that HMG-21 represents an mixture of linear isomers a and b, below. This prediction was confirmed by MS^n analysis (**see companion report, Ashline et al.**) and identified isomer a. as F-*para*-LNH II [1] and isomer b. as a previously unidentified HMG among 7 previously described HMGs with this intact mass [1-3].

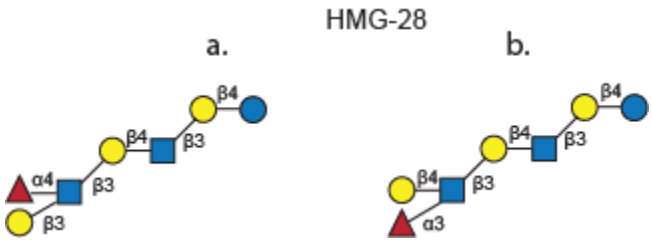


Glycan 9 (HMG-23) – HMG-23 also has a composition of H4N2F1 (intact mass of 1405.151[M+Na]+) and the HPLC profile shows a major peak with a minor contaminating glycan(s). The positive binding of AAL is consistent with the presence of Fucose. Strong binding to RCA-I and ECL, as well as the binding to GSL-II after β1-4-galactosidase digestion indicates the presence of a terminal type-2 determinant. The binding to the anti-type 1 antibody is very weak, which exclude the possibility of a terminal type-1 structure. The weak anti-type 1 antibody binding is probably due a minor contaminating peak. Therefore, HMG-23 could either be a branched structure with an LNnH core or a linear structure. However, the presence of a Fucose is inconsistent with a branched structure because no anti-H1, anti-Le^a or anti-Le^x binding is observed. Thus, we conclude that HMG-23 is a linear structure with an internal Le^x as the anti-Le^x antibodies 5F1 and CD15 could not detect an internal Le^x. These data are consistent with the previously described milk glycan F-*para*-Lacto-N-*neo*hexaose I [1, 3] or IF-LNH III [2].



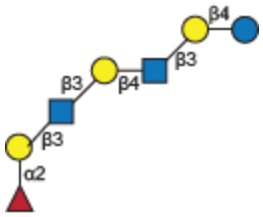
Glycan 10 (HMG-27) – HMG-27 displays the same molecular ion (1248.4 [M+H]+), HPLC retention time and lectin binding profile as HMG-13, therefore it is considered to have the same structure as HMG-13, although its purity is lower than the HMG-13 as indicated by MALDI and HPLC profile. This structure was confirmed by MS^n analysis (**see companion report, Ashline et al.**) to be *para*-Lacto-N-hexaose [1].

Glycan 11 (HMG-28) – HMG-28 also has a composition of H4N2F1 (1404.470 [M+Na]+) and its binding profile is similar to that of HMG-21. It is bound by the anti-Le^a antibody and the anti-Le^x antibodies 5F1 and CD15 indicating the presence of at least two isomers as discussed for HMG-21. The absence of ECL and anti-type 1 antibody binding excludes terminal type 1 and type 2 determinants. The anti-H1 binding at high concentration could be the result of an impurity and the weak anti-Le^b antibody binding is probably due to its cross-reactivity with the Le^a determinant. Taken together, we predict that HMG-28 contains the same isomers **a.** and **b.** as HMG-21 and the significant lower binding signal to CD15 may suggest that the Le^x-containing isomer **(b.)** in HMG-28 is less abundant than it is in HMG-21. This prediction was confirmed by MS^n analysis (**see companion report, Ashline et al.**).



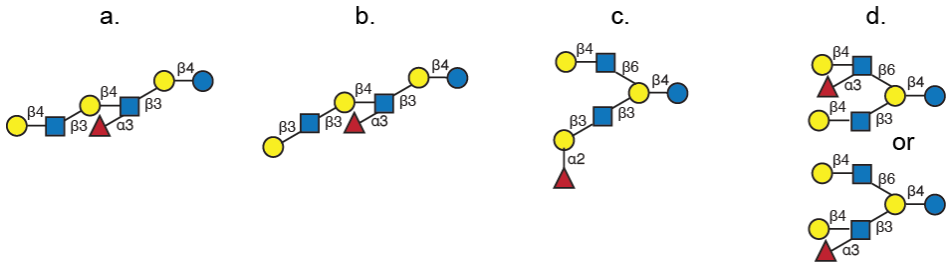
Glycan 12 (HMG-29) – HMG-29 is another H4N2F1 glycan (1404.465 [M+Na]+) and the HPLC profile showed a pure symmetrical peak. This glycan is only bound by three reagents AAL, anti-type 1 antibody and anti-H1 antibody, suggesting the presence of terminal type 1 and terminal H-type 1 determinants. Since branched HMGs with a hexaose core structure all have a type 2 LacNAc on the 6-branch, the anti-type 1 binding is due to its the cross-reactivity to the H-type 1 determinant as seen with LNFP I (see Glycan #38). This interpretation is confirmed by the lack of GSL-II binding after β1-3-galactosidase digestion. Thus, we predicted that HMG-29 is a linear hexaose terminating with an H-type 1 determinant. This prediction was confirmed by MS^n analysis (**see companion report, Ashline et al.**) to be IFLNH I [2].

HMG-29



Glycan 13 (HMG-31) – HMG-31 is another H4N2F1 (1404.506 [M+Na]+) but its HPLC profile shows an asymmetrical peak, indicating it is a mixture. This glycan is bound by AAL, RCA-I, ECL and anti-type 1 antibody, indicating the presence of both type 2 and type 1 termini. Since there is no anti-H-1, anti-Le^{a} or anti-Le^{b} binding, the Fucose must be an internal Le^{x} structure. Based on these data, we predicted the HMG-31 is a mixture of linear structures with either type-2 (structure **a.**) or type-1 (structure **b.**) determinants at the non-reducing end. The **a.** isomer is likely more abundant because GSL-II binding after β1-4-galactosidase digestion is much stronger than that of after β1-3-galactosidase digestion. The **a.** structure is the same as HMG-23, which is in agreement with the retention time on HPLC profile. This prediction was confirmed by MS^{n} analysis, which also indicated that this fraction had minor contaminants with branched hexaose (isomers **c.** and **d.**) cores that were identified as F-Lacto-N-Hexaose I [1, 3] and a F-LNnH (not previously identified as a free glycan in human milk) that was related to HMG-20 and shown below (**see companion report, Ashline et al.**). Isomer **a.** has the same structure as HMG-23 above, while isomer **b.** has been previously identified as F-*para*-Lacto-N-Hexaose I [1, 3] or MFpLNH IV [2].

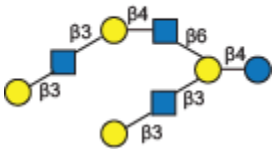
HMG-31



Glycan 14 (HMG-33) – HMG-33 has a molecular mass of 1623.123 [M+Na]+ that is consistent with a composition of Hex5HexNAc3 (H5N3), and the HPLC profile displays a single symmetrical peak. The weak RCA-I binding and strong anti-type 1 antibody binding, as well as the GSL-II binding after β1-3-galactosidase digestion, all indicate the presence of a type-1 terminal. No other binding data was observed with any of the lectins and antibodies tested. Therefore, this glycan is either a linear structure with a type 1 terminal or a branched structure terminated with two type 1 chains. Because the known linear HMGs

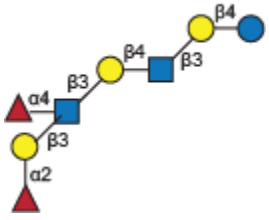
are always an extension of a type 2 LacNAc; i.e., the LNnT (Galβ1-4GlcNAcβ1-3Galβ1-4Glc) and possess an internal Galβ1-4-linkage, the negative binding of GSL-II after endo-β1-4-galactosidase treatment excluded the linear structure. Thus, we predicted the most likely structure for HMG-33 is the branched structure with two terminal type-1 determinants, *iso*-Lacto-N-octaose. This prediction was confirmed by MS^n analysis (**see companion report, Ashline et al.**) and is *iso*lacto-N-octaose. The two type-1 termini were further confirmed by MALDI analysis of the β1-3-galactosidase treated HMG-33, which indicated that two galactose residues were lost and no change was observed after β1-4-galactosidase. This is the first report of a free octaose (not substituted with Fucose or sialic acid) among the HMGs [1].

HMG-33



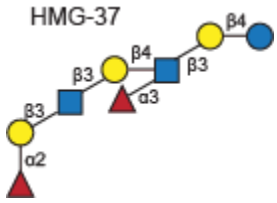
Glycan 15 (HMG-34) – HMG-34 has a molecular mass of 1551.103 [M+Na]+ that is consistent with a composition of H4N2F2, and the HPLC profile displays a single symmetrical peak. This glycan is also bound by only two reagents; AAL and the anti-Le^b antibody, indicating a terminal Le^b determinant. Upon treatment by α1-2-fucosidase, it becomes recognized by the anti-Le^a antibody due to the loss of the terminal α1-2-fucose. Since there is no evidence for the presence of type 1 and type 2 terminal determinants, we predicted HMG-34 is a linear Le^b glycan, which has not been identified among the 6 previously reported difucosylated hexaoses in HMG [1-3]. This prediction was confirmed by MS^n analysis (**see companion report, Ashline et al.**).

HMG-34

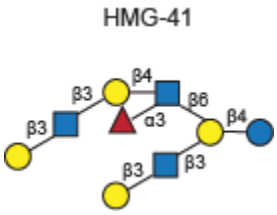


Glycan 16 (HMG-37) – HMG-37 has a molecular mass (1550.309 [M+Na]+) that is consistent with a composition of H4N2F2, and the HPLC profile suggests a relative pure peak. Its binding profile is very similar to HMG-29 (glycan 12), both bind AAL, anti-type 1 antibody and anti-H1 antibody. Following the same rationale, HMG-37 should also be a linear structure with a terminal H type 1; however, this glycan contains an additional Fucose that is not recognized by the anti-Le^a, anti-Le^b or anti-Le^x antibody.

Since this glycan has a single H type 1 terminus, we presumed it to be linear with the second Fucose likely associated with a Le^x determinant on the internal LNnT. This prediction was confirmed by MS^n analysis (**see companion report, Ashline et al.**) and identified as a glycan not previously described among the 6 previously reported difucosylated hexaoses in HMGs [1-3].



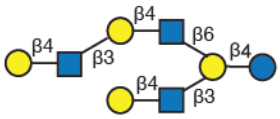
Glycan 17 (HMG-41) – HMG-41 has a molecular mass of 1769.379 [M+Na]+ that is consistent with a composition of H5N3F1, and the HPLC profile indicated a single symmetrical peak. The weak RCA-I binding, the strong anti-type 1 antibody binding and the GSL-II binding after β1-3-galactosidase digestion were all consistent with the presence of type 1 terminus. The very low ECL binding and the weak anti-Le^a antibody binding at high concentration are probably due to the small amount of impurity. Since this glycan has an octasaccharide core that is resistant to endo β1-4-galactosidase, we presumed it to be branched. The AAL binding is consistent with the presence of Fucose, which is not a component an H type 1 or Le^a determinant since neither anti-H1 antibody nor anti-Le^a antibody was bound by this glycan. We predicted, therefore, that the single Fucose would be a component of an internal Le^x determinant, which is not detected with any of the anti-Le^x reagents we have available. This prediction was confirmed by MS^n analysis (**see companion report, Ashline et al.**) and identified as a previously described [1-3] fucosylated *iso*lacto-N-octaose (F-isoLNO). The two type-1 termini were further confirmed by MALDI analysis of the β1-3-galactosidase treated HMG-41, which indicated that two galactose residues were lost and no change was observed after β1-4-galactosidase.



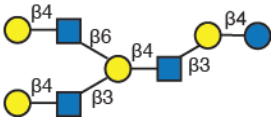
Glycan 18 (HMG-45) – HMG-45 has a molecular mass of 1601.389 [M+H]+ that is consistent with a composition of H5N3 and the HPLC profile displays a single symmetrical peak. It binds only RCA-I and ECL; and after β1-4-galactosidase treatment (but not after β1-3-galactosidase or endo-β1-4galactosidase treatment) it binds GSL-II. These data indicate the exclusive presence of type 2 termini, and the

resistance to endo-β1-4-galactosidase confirms the branched structure (the linear isomer would be cleaved by endo-β1-4-galactosidase). Thus, we predicted this glycan to be a previously undescribed octaose [1-3], which would have the trivial name, *iso*Lacto-N-*neo*octaose (see below). The two type-2 termini were confirmed by MALDI analysis on galactosidase treated HMG-45. Two galactoses were lost after β1-4-galactosidase digestion and the majority remained intact after β1-3-galactosidase, suggesting a minor contamination of type-1 containing glycan. However, detailed structural analysis by MS^n (**see companion report, Ashline et al.**) indicated that the structure had an unsubstituted lactose, which places the branch point on the in the position shown in HMG-45, which is also a previously undescribed core structure.

*iso*Lacto-N-*neo*octaose

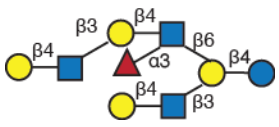


HMG-45



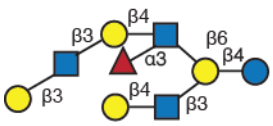
Glycan 19 (HMG-47) – HMG-47 has a molecular mass of 1769.437 [M+Na]+ that is consistent with a composition of H5N3F1, and the HPLC profile displayed a single symmetrical peak. Except for the positive AAL binding, this glycan displayed terminal type 2 determinants as observed in Glycan #18 (HMG-45). While AAL binding is consistent with the presence of Fucose, the Fucose is not a component an H type 1 or Le^a determinant since neither anti-H1 antibody nor anti-Le^a antibody was bound by this glycan. We predicted, therefore, that the single Fucose would be a component of an internal Le^x determinant, which is not detected with any of the anti-Le^x reagents we have available. However, the presence of this Fucose prevented us from excluding the possibility that this glycan could be linear since an internal Le^x may block the endo-β1-4-galactosidase cleavage. Thus, we were unable to assign a structure based on composition and binding data alone. More detailed MS^n analyses differentiated these two possibilities and confirmed that HMG-47 was the fucosylated Lacto-N-*neo*octaose shown below (**see companion report, Ashline et al.**) that has not been previously described [1-3].

HMG-47

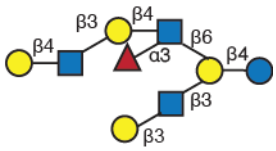


Glycan 20 (HMG-48) – HMG-48 also has a composition of H5N3F1 (1769.457 [M+Na]+), and the HPLC profile displays a single symmetrical peak. This glycan contains both type 1 and type 2 termini as indicated by the binding of anti-type 1 antibody, RCA-I and ECL, as well as by sensitivity to digestion with β1-3- and β1-4-galactosidase. The Fucose is likely a component of an internal Le^x structure, which is not detected with any of the Fucose-binding reagents we have available. Therefore, we predict HMG-48 contains either the Lacto-N-*neo*octaose or Lacto-N-octaose core and is one of the two glycans below. One is fucosylated Lacto-N-*neo*octaose, F-LNnO [1, 3] and the other is Fucosylated Lacto-N-octaose, F-LNO [2]. To distinguish between these isomers, HMG-48 was treated with β1-3- and β1-4-galactosidase and examined by MALDI. One galactose was removed by β1-3-galactosidase while no change was introduced by β1-4-galactosidase, which is consistent with F-LNnO, because a β1-4-linked galactose directly on a branch point (as in F-LNnO and LNnH) is resistant to β1-4-galactosidase digestion under our conditions, but it is released if it is on an extended chain (like F-LNO). F-LNnO has been previously identified as a human milk glycan [1, 3].

HMG-48



F-LNnO

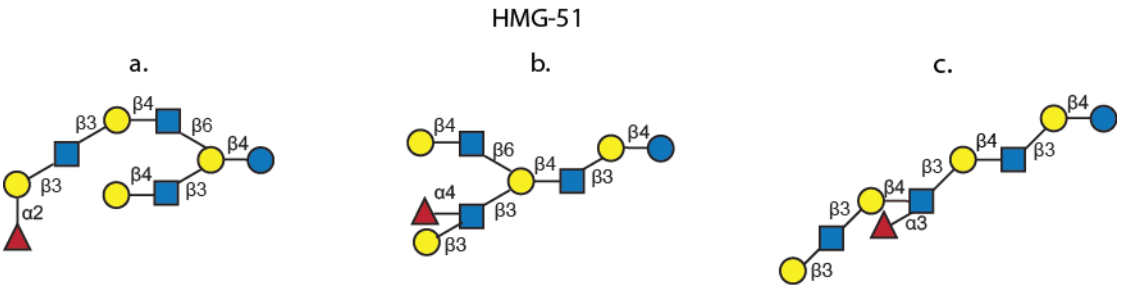


F-LNO

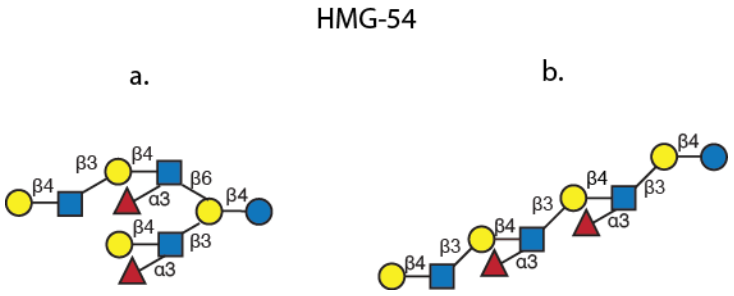
Glycan 21 (HMG-49) – HMG-49 displays the same molecular ion (1769.920 [M+Na]+), HPLC retention time and lectin binding profile as HMG-41. We predicted, therefor, that HMG-49 has the same structure as HMG-41. This prediction was confirmed by MS^n analysis (**see companion report, Ashline et al.**). The two type-1 termini were further confirmed by MALDI analysis of the β1-3-galactosidase treated HMG-49, which indicated that two galactose residues were lost and no change was observed after β1-4-galactosidase.

Glycan 22 (HMG-51) – HMG-51 has a composition of H5N3F1 (1769.602 [M+Na]+), and the HPLC profile gives a single peak. However, the binding data suggest it is actually a mixture of at least two isomers. AAL binding is consistent with the presence of Fucose and the composition indicates only 1 Fucose per glycan. Strong binding by anti-Le^a and weak binding by anti-H1 and anti-type 1 antibodies suggest at least two distinct Fucose-containing glycans are present in this mixture. The positive binding of RCA-I and ECL, suggest the presence of type 2 determinants. The data for the presence of a terminal type 1 is equivocal since the anti-type 1 antibody binding is weak and can also bind H-type 1. In order to

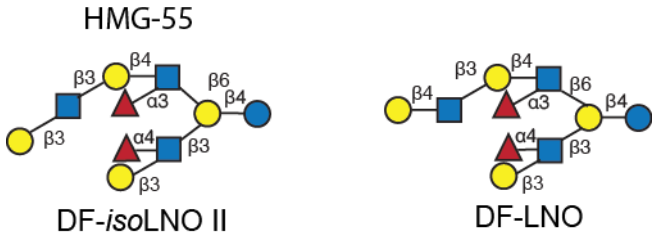
determine the structures of the glycans in this mixture, we subjected this fraction to MS^n analysis. The complex data from this analysis (**see companion report, Ashline et al.**) indicated a mixture of 3 isomers a., b., and c. The binding of VP8* from N155 to HMG-51 after α1-2 Fucosidase digestion is consistent with this interpretation based on the similarity of the 6-branch of HMG-51a and HMG-29, which also became a binder for VP8* from N155 after α1-2Fucosidase digestion. Thus, HMG-51a, b, and c are mono-fucosylated octaoses. Glycan a. is a monofucosylated Lacto-N-*neo*octaose; glycan b. is a monofucosylated derivative of a novel octaose core structure that is related to HMG-45, and glycan c. is a previously undescribed monofucosylated *para*-Lacto-N-octaose [1-3]. None of these glycans have been previously identified as free glycans in human milk.



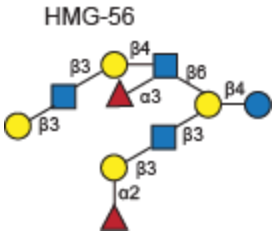
Glycan 23 (HMG-54) – HMG-54 has an intact mass of 1915.946 [M+Na]+ that is consistent with a composition of H5N3F2, and the HPLC profile indicated this glycan was relatively pure. The binding profile of HMG-54 is almost the same as the HMG-47, except for very weak binding to anti-Le^x antibody at a high antibody concentration (**Table S7**). Therefore, we predict it has the same structure as HMG-47, but with a terminal Le^x determinant, which is consistent with a difucosylated derivative of the novel *iso*Lacto-N-*neo*octaose [1, 2]. However, the location of the Le^x determinants could not be placed with the binding data alone. The MS^n analysis (**see companion report, Ashline et al.**) of this glycan indicated the presence of two structures a. and b. in this fraction. Neither of these glycans has been previously described in human milk [1-3], and HMG-54b is a fucosylated derivative of a novel linear core structure.



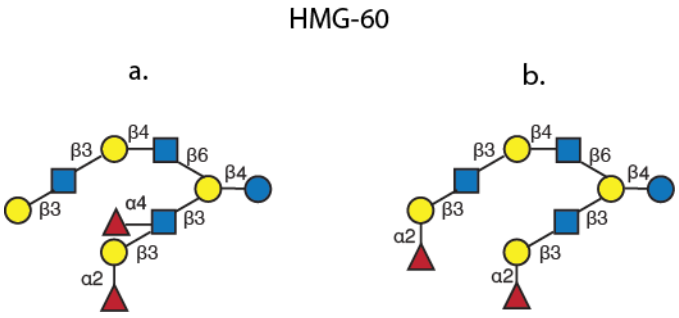
Glycan 24 (HMG-55) – HMG-55 has a composition of H5N3F2 (1916.142 [M+Na]+). The HPLC profile shows a major fraction and a minor fraction. This glycan is strongly bound by anti-type 1 antibody and anti-Le^a antibody, suggesting the presence of a type 1 branch and a Le^a branch. The low RCA-I binding and weak GSL-II binding after β1-4-galactosidase digestion is probably from the type 2 terminus of a minor component with a terminal Gabβ1-4GlcNAc on the 6-branch. The second Fucose most likely belongs to an internal Le^x structure that was not detected by our anti-Le^x antibodies. Taken together, we predict the major glycan in HMG-55 is below contaminated with a small amount of the difucosylated Lacto-N-octaose DF-LNO. This prediction was confirmed by MS^n analysis (**see companion report, Ashline et al.**) and the structure is consistent with the difucosylated *iso*Lacto-N-octaose, DF-*iso*LNO II below that was previously undescribed among the 9 known difucosylated octaoses [1-3]. The Difucosylated Lacto-N-octaose shown below was previously reported as Difucosylated Lacto-N-octaose II [2].



Glycan 25 (HMG-56) – HMG-56 is another glycan with a composition of H5N3F2 (1916.090 [M+Na]+), and the HPLC profile shows a major peak with impurities. This glycan is bound by anti-type 1 antibody without treatment and by GSL-II after β1-3-galactosidase treatment (but not after β1-4-galactosidase treatment), indicating the presence of only terminal type 1 determinants. Moderate binding to anti-Le^b antibody, as well as low binding to anti-H1 antibody at high concentration, indicates this glycan also contains Le^b and H1 determinants. Based on only MALDI and binding data, we are unable to predict an unequivocal structure for HMG-56. However, tandem MS analysis (**see companion report, Ashline et al.**) indicated that this fraction contained at least 3 isomeric structures but identified an H1 determinant and a monofucosylated diLacNAc fragment with an apparent internal Le^x determinant. In addition, the MS^n data indicated no terminal Le^x or Le^a determinants, no H structure, but identified a terminal LacNAc that was all Gal1-3GlcNAc. These data were consistent with the structure shown below being the major glycan in HMG-56, which is consistent with the previously described difucosylated *iso*Lacto-N-octaose II [1, 3].



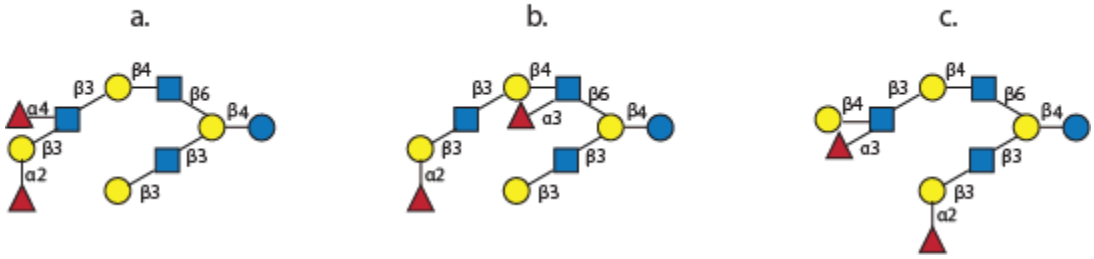
Glycan 26 (HMG-60) – HMG-60 also has a composition of H5N3F2 (1916.078 [M+Na]+), and the HPLC profile shows one major peak with two minor peaks, which is reflected by the complexity of binding data with positive binding by anti-H1 and anti-Le^b. This glycan is strongly bound by AAL due to the two Fucose residues. The weak binding of RCA-I, ECL and anti-type 1 antibody indicate that the terminal type 2 and type 1 determinants are fucosylated, which is consistent with the resistance to β-galactosidase digestion and the high binding of anti-H1 and anti-Le^b. Thus, H type 1 and Le^b are likely the major determinants and the Le^b determinant is confirmed by the anti-Le^a binding observed after α1,2-fucosidase digestion. Based on the robust binding by anti-Le^b and anti-type 1 antibodies, we predict HMG-60 to be comprised of two glycans, but the binding data did not allow us to identify which arm of the core octasaccharide carried the Le^b determinant. Analysis of this fraction by tandem MS confirmed that this fraction was comprised of two glycans a. and b. and that the Le^b determinant was not located on the diLacNAc on the 6-arm of the parent Lacto-N-hexaose (**see companion report, Ashline et al.**). These glycans have not been previously described among the 9 known difucosylated octaoses in human milk [1-3].



Glycan 27 (HMG-62) – HMG-62 has a composition of H5N3F2 (1916.083 [M+Na]+), and HPLC clearly shows a mixture. The binding profile of this glycan is similar to HMG-60, but presented stronger binding by anti-type 1 antibody and was susceptible to digestion with β1-3-galactosidase, but not the β1-4-galactosidase indicating that it had more terminal type 1 determinants than HMG-60. When MS^n analysis was carried out on this glycan (see companion report, Ashline et al.), terminal LacNAc appeared to be all type 1 glycan (Galβ1-3GlcNAc), the monofucosylated LacNAc was predominantly H1 with a trace of Le^x,

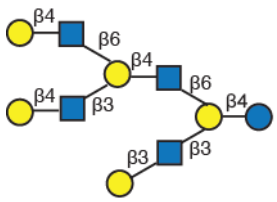
and an internal Le^x determinant was identified. Based on the binding profiles and the determinants identified by MS^n analysis, HMG-62 **a.** and HMG-62 **b.** are the major glycans with HMG-62 **c.** being a minor component. HMG-62 **b**. was previously identified as glycan 5230a [2], and HMG-62 **a**. and HMG-62 **c**. are previously undescribed glycans among the 9 known difucosylated octaoses in human milk [1-3].

HMG-62



Glycan 28 (HMG-65) – HMG-65 has a molecular mass of 1988.360 [M+Na]+ that is consistent with a composition of Hex6HexNAc4, which is an unsubstituted decaose. Among the 13 core structures known to occur in HMGs only 2 branched decaoses have been described; Lacto-N-decaose and Lacto-N-*neo*decaose [1, 3]. The AAL binding (Fucose) and weak anti-Le^a binding does not agree with the composition but may be attributed to the impurities seen in HPLC profile. The RCA-I, ECL binding and GSL-II binding after β1-4-galactosidase digestion indicate the presence of terminal type 2, while the anti-type 1 antibody binding indicates the presence of terminal type 1. Thus, HMG-65 could be either of the two known decaoses in Human milk, which are extensions of Lacto-N-hexaose. When this glycan was treated by specific galactosidase and analyzed by MALDI, it was observed that one galactose was removed by β1-3-galactosidase, which is consistent with lacto-N-*neo*decaose structure. No change was introduced by the β1-4-galactosidase as the β1-4-linked galactose on a branch is very difficult to remove under the condition we used. Glycan 30 (HMG-67) is another isomer and the composition and binding data are also consistent with it being one of these isomers (Lacto-N-decaose or Lacto-N-*neo*decaose). Tandem MS^n analysis (**see companion report, Ashline et al.**) indicated that this fraction was predominantly lacto-N-*neo*decaose.

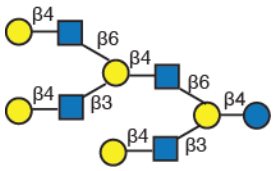
HMG-65



lacto-N-neodecaose

Glycan 29 (HMG-66) – HMG-66 also has a molecular mass of 1988.385 [M+Na]+ with a composition of H6N4, which is another decaose isomer. The weak AAL and weak anti-H1 binding at high concentrations of lectin or antibody may be due to the contamination with Fucose-containing glycans. Interestingly, only type 2 terminal determinants are present based the strong binding to RCA-I and ECL, the strong GSL-II binding after β1-4-galactosidase digestion (not after β1-3-galactosidase digestion), and the absence of anti-type 1 antibody binding. These data are inconsistent with the two known structures of decaose core glycans, but consistent with an undescribed decaose based on LNnH precursor shown below [1-3]. The lack of the type-1 terminal was further confirmed by the MALDI analysis on β1-3-galactosidase treated

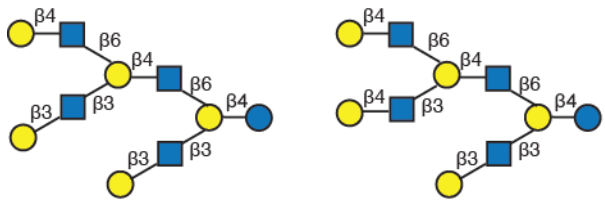
HMG-66



HMG-66 as no change was observed on mass, and by tandem MS analysis (**see companion report, Ashline et al.**).

Glycan 30 (HMG-67) – HMG-67 has the same molecular ion (1988.478 [M+Na]+) as HMG-64/65 and is one of 3 isomers of decaoses found in the TGL prepared from human milk. Based on the observation that it contains both type 1 and type 2 terminal determinants, this glycan presumably represents one of the known isomers of decaoses as mentioned in the description of Glycan 28 (HMG-65) mentioned above. When this glycan was treated by specific galactosidase and analyzed by MALDI, it was observed that one galactose was removed by β1-3-galactosidase, which is consistent with lacto-N-neodecaose structure. No change was introduced by the β1-4-galactosidase as the β1-4-linked galactose on a branch is very difficult to remove under the condition we used. Tandem MS analysis (**see companion report, Ashline et al.**) confirmed that the major structures in this fraction was lacto-N-*neo*decaose, with some lacto-N-decaose present based on the detection of terminal Galβ1-4GlcNAc in the branched 6 arm.

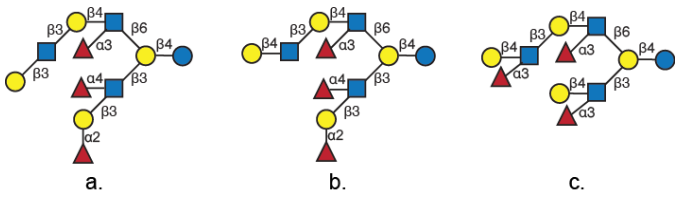
HMG-67



lacto-N-decaose lacto-N-neodecaose

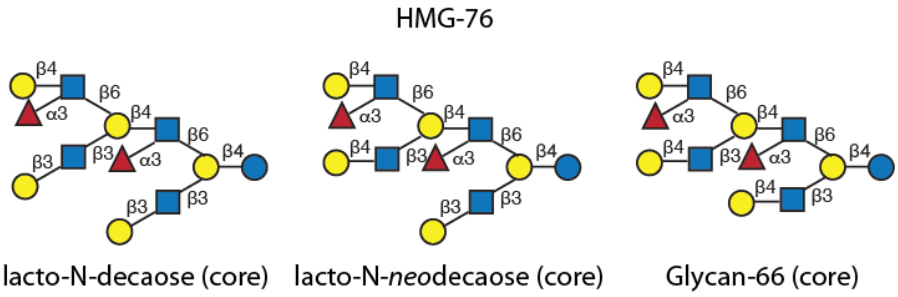
Glycan 31 (HMG-69) – HMG-69 has a molecular mass of 2061.456 [M+Na]+ that is consistent with the composition of H5N3F3, and the HPLC profile is very symmetrical suggesting a relatively homogeneous fraction. This glycan is weakly bound by RCA-I but not bound by ECL, indicating the absence of terminal type 2 determinants; however, the weak RCA-I binding and the slightly positive binding to GSL-II observed after β1-4-galactosidase digestion suggest a low level of type 2 determinant. Relatively weak anti-type 1 antibody binding and weak but significant binding to GSL-II after β1-3-galactosidase digestion suggests presence of type 1 terminal determinant, and taken together these data suggest the majority terminal sequences are type 1 and that most of the glycans have fucosylated terminal sequences. This observation is consistent with the strong binding by anti-Le^b, anti-Le^x 5F1 and CD15 antibodies. Weak anti-Le^a binding is also observed and this binding is significantly increased after α1-2-fucosidase digestion, suggesting Le^a may not a major determinant in this glycan and the anti-Le^a binding could be a slight cross reactivity with the Le^b determinant. In addition, internal Le^x is also possible to exist in a highly fucosylated glycan. This glycan was subjected to MS^n analysis (**see companion report, Ashline et al.**), which indicated the presence of terminal Le^b, terminal Le^x, a di-LacNAc arm with an internal Le^x determinant and type 1 and 2 terminal LacNAc [Galβ1-3/4GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAc], and a difucosly-diLacNAc with both a terminal Le^x and an internal Le^x [Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAc]. Arrangement of the determinants detected by MS^n and antibody and lectin binding suggests that HMG-69 is a mixture of isomers with the most likely structures shown below, and there are several possibilities for the structures. None of these glycans have been described among the 5 known trifucosylated octaoses from human milk [1-3].

HMG-69

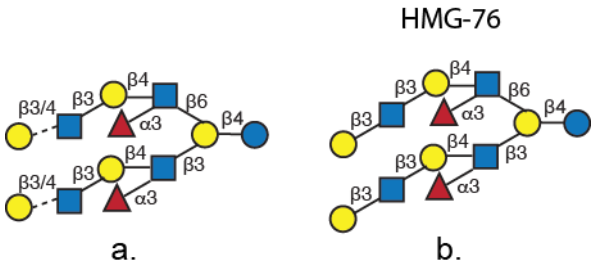


Glycan 32 (HMG-76) – HMG-76 has a molecular mass of 2280.822 [M+Na]+ that matches a composition of H6N4F2 and HPLC profile shows a major peak with a very small shoulder. The weak ECL binding and moderate RCA-I binding indicate the terminal type 2 is not the major structure, and the binding to anti-type 1 antibody is also not very strong; however, strong GSL-II binding is gained after β1-3-galactosidase digestion. Binding by anti-H1, anti-Le^a and anti-Le^b antibodies is weak suggesting that the Fucose residues are present as internal Le^x determinants, which are poorly recognized by the 5F1 and CD-15 anti-Le^x antibodies. Based on the structures of the two known decaoses in human milk and the binding

data, we initially predict this glycan is most likely a mixture of difucosylated Lacto-N-decaose and Lacto-N-*neo*decaose and possibly a difucosylated Glycan 66 (the novel decaoses described above).



However, analysis using MS^n resulted in the identification of extremely different structures (**see companion report, Ashline et al.)**, possessing novel core decaoses that have not been previously reported [1-3] as shown in **a**. below. The presence of two type-1 termini were confirmed by MALDI analysis of galactosidase treated HMG-76. Two galactoses were lost after β1-3-galactosidase digestion and a very minor fraction lost one galactose after β1-4-galactosidase, indicating the major component of HMG-76 is **b**.



References

1. Kobata, A., *Structures and application of oligosaccharides in human milk.* Proceedings of the Japan Academy. Series B, Physical and biological sciences, 2010. **86**(7): p. 731-47.
2. Wu, S., et al., *Development of an annotated library of neutral human milk oligosaccharides.* J Proteome Res, 2010. **9**(8): p. 4138-51.
3. Urashima, T., Kitaoka, M., Terabayashi, T., Fukuda, K., Ohnishi, M., and Kobata, A, *Milk Oligosaccharides*, in *Oligosaccharides: Sources, Properties and Applications*, N.S. Gordon, Editor. 2011, Nova Science Publishers, Inc.: New York, NY. p. 1-58.