Glycobiology, 2024, **34**, cwae062
https://doi.org/10.1093/glycob/cwae062
Advance access publication date 8 August 2024
Original Article

OXFORD

Editor's Choice

LacdiNAc to LacNAc: remodelling of bovine α-lactalbumin *N*-glycosylation during the transition from colostrum to mature milk

Inge Gazi[1,2,*], Karli R. Reiding[1,2], André Groeneveld[3], Jan Bastiaans[3], Thom Huppertz[3,4], Albert J.R. Heck[1,2]

[1]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, Utrecht 3584 CH, The Netherlands, [2]Netherlands Proteomics Center, Padualaan 8, Utrecht 3584 CH, The Netherlands, [3]Research and Development, FrieslandCampina, Stationsplein 4, Amersfoort 3818 LE, The Netherlands, [4]Department of Agrotechnology and Food Sciences, Wageningen University, Bornse Weilanden 9, Wageningen 6708 WG, The Netherlands
*Corresponding author: Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, Utrecht 3584 CH, The Netherlands. Email: i.gazi@uu.nl

α -Lactalbumin, an abundant protein present in the milk of most mammals, is associated with biological, nutritional and technological functionality. Its sequence presents *N*-glycosylation motifs, the occupancy of which is species-specific, ranging from no to full occupancy. Here, we investigated the *N*-glycosylation of bovine α-lactalbumin in colostrum and milk sampled from four individual cows, each at 9 time points starting from the day of calving up to 28.0 d post-partum. Using a glycopeptide-centric mass spectrometry-based glycoproteomics approach, we identified *N*-glycosylation at both Asn residues found in the canonical Asn-Xxx-Ser/Thr motif, i.e. Asn45 and Asn74 of the secreted protein. We found similar glycan profiles in all four cows, with partial site occupancies, averaging at 35% and 4% for Asn45 and Asn74, respectively. No substantial changes in occupancy occurred over lactation at either site. Fucosylation, sialylation, primarily with *N*-acetylneuraminic acid (Neu5Ac), and a high ratio of *N,N'*-diacetyllactosamine (LacdiNAc)/*N*-acetyllactosamine (LacNAc) motifs were characteristic features of the identified *N*-glycans. While no substantial changes occurred in site occupancy at either site during lactation, the glycoproteoform (i.e. glycosylated form of the protein) profile revealed dynamic changes; the maturation of the α-lactalbumin glycoproteoform repertoire from colostrum to mature milk was marked by substantial increases in neutral glycans and the number of LacNAc motifs *per* glycan, at the expense of LacdiNAc motifs. While the implications of α-lactalbumin *N*-glycosylation on functionality are still unclear, we speculate that *N*-glycosylation at Asn74 results in a structurally and functionally different protein, due to competition with the formation of its two intra-molecular disulphide bridges.

Key words: bovine milk N-glycoproteome; glycan structure; glycoproteomics; LacdiNAc; LacNAc.

Introduction

Colostrum and milk are multifunctional biological fluids meant to nourish and support the immune, digestive and cognitive development of new-borns and developing young infants. McGrath et al. (2016) described the compositional and physical–chemical changes of bovine colostrum during the initial days of lactation, with comparisons to mature milk. Colostrum is the mammary secretion produced in the first 3.0 d post-partum and is a concentrated fluid rich in proteins with immune function (e.g. immunoglobulins, lactoferrin, cytokines and growth factors), supporting the development of the immune and digestive systems of the neonate. Mature milk is produced from 28.0 d onwards. Compared to colostrum, it is a less viscous fluid, with lower concentrations of protein and fat, compositionally dominated by caseins, β-lactoglobulin and α-lactalbumin, supporting the growth and development of the infant. The composition of the mammary secretions gradually transitions from colostrum to mature milk during which time they are referred to as transitional milk (3.0–28.0 d postpartum).

Mammary secretions contain many classes of bioactive components, one of them being the glycoprotein fraction (Gopal and Gill 2000). Protein glycosylation is of interest due to its function to modulate protein interactions and activity, increase stability against proteolytic degradation, bind pathogens, and promote gut health and a balanced microbiome (Krištić and Lauc 2017). Previous studies have shown differences at the bovine *N*-glycome level, where *N*-glycolylneuraminic acid (Neu5Gc) sialylation was gradually lost during the transition from colostrum to mature milk (Takimori et al. 2011). The changes in the *N*-glycome observed during the first month of lactation could primarily be attributed to the associated changes in protein composition. Nevertheless, while *N*-glycosylation is protein- and site-specific, it has also been shown to undergo compositional (e.g. sialylation, fucosylation) and structural (e.g. bisection, antennarity) dynamic changes throughout lactation (Zhu et al. 2020; Gazi et al. 2023). These glycosylation changes are likely driven by hormonal changes to glycosyltransferase gene expression and may be linked to specific functions in relation to the protein. E.g., in our previous study we identified *N*-acetylneuraminic acid (Neu5Ac) sialylation as a hallmark of bovine colostrum IgG, concluding that the high occurrence of Neu5Ac likely played a role in the transepithelial transport,

Received: March 5, 2024. **Revised:** August 5, 2024. **Accepted:** August 7, 2024

resistance to digestive proteases and longer half-life of bovine IgG in providing passive immunity to the calf (Gazi et al. 2023).

α-Lactalbumin is one of the most abundant functional proteins in mammary secretions. It is exclusively produced in the lactating mammary gland, and it is present in the colostrum and milk of most mammals (Brew 2013). In bovine milk, α-lactalbumin is the second-most abundant whey protein and makes up approximately 3.5% (*m/m*) of total protein content, constituting also a source of bioactive peptides and essential amino acids (Layman et al. 2018). By contrast, in human milk, α-lactalbumin is a more abundant protein, constituting approximately 22% (*m/m*) of the total protein content (Layman et al. 2018). For this reason, bovine α-lactalbumin is also a protein of interest in the production of bovine milk-based infant formula. α-Lactalbumin is a globular protein, with its tertiary structure stabilized by four intramolecular disulphide bridges (Vanaman et al. 1970) and the binding of one Ca^{2+} ion (Chrysina et al. 2000). In the mammary gland, α-lactalbumin regulates the synthesis of lactose in a heterodimeric complex with β-1,4-galactosyltransferase 1 (B4GALT1) (Brodbeck et al. 1967). While B4GALT1 normally catalyses the reaction between uridine diphosphate galactose (UDP-Gal) and *N*-acetylglucosamine (GlcNAc), with the formation of *N*-acetyllactosamine (LacNAc), the presence of α-lactalbumin changes the affinity of B4GALT1 from GlcNAc to glucose (Glc), leading to the synthesis of lactose (Lac) (Brew et al. 1968).

Bovine α-lactalbumin contains two putative *N*-glycosylation motifs, containing the Asn residues Asn45 (Asn-Asp-Ser) and Asn74 (Asn-Ile-Ser), and a third non-canonical motif containing Asn71 (Asn-Ile-Cys). While *N*-glycosylation of bovine α-lactalbumin at Asn74 has never been reported, partial occupancy of Asn45 has been shown in previous studies, primarily with complex diantennary *N*-glycans containing LacNAc and/or *N,N′*-diacetyllactosamine (LacdiNAc) motifs, fucosylation and Neu5Ac sialylation, with on average 10% of total α-lactalbumin occurring as a glycoprotein (Barman 1970; Hindle and Wheelock 1970; Hopper and McKenzie 1973; Chandrika 1999; Slangen and Visser 1999; Valk-Weeber et al. 2020). However, the differences between bovine colostrum and mature milk α-lactalbumin *N*-glycosylation remain to be elucidated, both in terms of site occupancy and qualitative heterogeneity.

We here present the *N*-glycoproteome context of bovine colostrum, transitional and mature milk, in which α-lactalbumin plays a central role. We further dived into a detailed characterisation of α-lactalbumin *N*-glycosylation, including that of glycan localisation on the amino acid sequence (i.e. *N*-glycosylation site identification), occupancy (i.e. percentage of molecules that are *N*-glycosylated at a given site) and composition, as well as the maturation of these features in the transition from colostrum to mature milk. Via our glycoproteomics approach we could independently identify the glycosylation of Asn45 and Asn74. Of note, while α-lactalbumin glycosylation indeed changed when transitioning from colostrum to mature milk, overall site occupancy remained constant in time and between individual animals, averaging at 35% for Asn45 and 4% for Asn74. The change in glycosylation was marked by a decrease in the extent of Neu5Ac sialylation, which remained present in the mature milk, and by a clear transition from LacdiNAc to LacNAc antennary motifs. To further understand the impact of the identified *N*-glycosylation on the structure of bovine α-lactalbumin and its functionality in the lactose synthase complex, we modelled the 3D structures of the doubly-glycosylated protein and of its interaction with the B4GALT1 catalytic domain.

Results

The general *N*-glycoproteomes of bovine colostrum and milk

To study the characteristics and changes during lactation of α-lactalbumin *N*-glycosylation in the context of the bovine colostrum and milk proteome and *N*-glycoproteome, we worked with the sample set described in Gazi et al. (2023), on which we performed new analyses in the current study. Within this, we analysed 36 individual bovine colostrum (0.5, 1.0, 1.5, 2.0, 2.5 and 3.0 d after calving), transitional (7.0 and 14.0 d after calving) and mature milk (28.0 d after calving) samples collected from 4 individual Dutch Friesian-Holstein dairy cows, referred to as Cow 1, Cow 2, Cow 3 and Cow 4.

To gain greater insights into not only into bovine α-lactalbumin as an individual protein, but also to its position in relation to the other colostrum and milk (*N*-glyco)proteins, we adopted a holistic approach, investigating the full range of (*N*-glyco)proteins present in the samples. The relative abundances of the *N*-glycoproteins in the total proteome *per* lactation time point are shown in Fig. 1A and are expressed as percentages of total intensity-Based Absolute Quantification (iBAQ) values, which are proportional to the molar ratios of the proteins. This revealed α-lactalbumin as a main component of bovine mammary secretions, making up 11% of the total protein molar concentration in colostrum at 0.5 d, which increased to 21% in the mature milk at 28.0 d. Our glycoproteomics approach was based on *N*-glycopeptide-centric analysis of the samples and allowed us to connect each identified *N*-glycan composition and structure to the glycoprotein it occurred on (Fig. 1). Within the glycoproteomics data, a total of 18 *N*-glycoproteins could be identified. These 18 *N*-glycoproteins made up 34% of the total protein molar concentration in the 0.5 d colostrum, gradually decreasing to 25% in the 7.0 d transitional milk and maintaining this value until the 28.0 d mature milk. Of these, 12 *N*-glycoproteins (including all immunoglobulin heavy chain constants, the JCHAIN and PIGR) were significantly (t-test, 0.01 false discovery rate, 250 permutations, s0 = 0.1) more abundant in the proteome of 0.5 d colostrum and 4 *N*-glycoproteins (including α-lactalbumin and GLYCAM1) were significantly more abundant in the 28.0 d mature milk. The remaining 2 *N*-glycoproteins, i.e. the milk fat globule membrane proteins MFGE8 and CD36, did not substantially differ in their abundances across milk maturation. The remainder of the proteome where no *N*-glycosylation was detected was primarily made up of the major milk proteins LGB, CSN1S1, CSN1S2, CSN2 and CSN3, the proportions of which gradually increased in the proteome during the transition from the 0.5 d colostrum to the 28.0 d mature milk. The greater part of *N*-glycopeptides in colostrum (0.5–3.0 d after calving) originated from the *N*-glycoproteins that were also significantly more abundant in the 0.5 d colostrum than in the 28.0 d mature milk (Fig. 1B). In the transitional (7.0 and 14.0 d) and mature (28.0 d) milk samples, over half of the *N*-glycopeptides originated

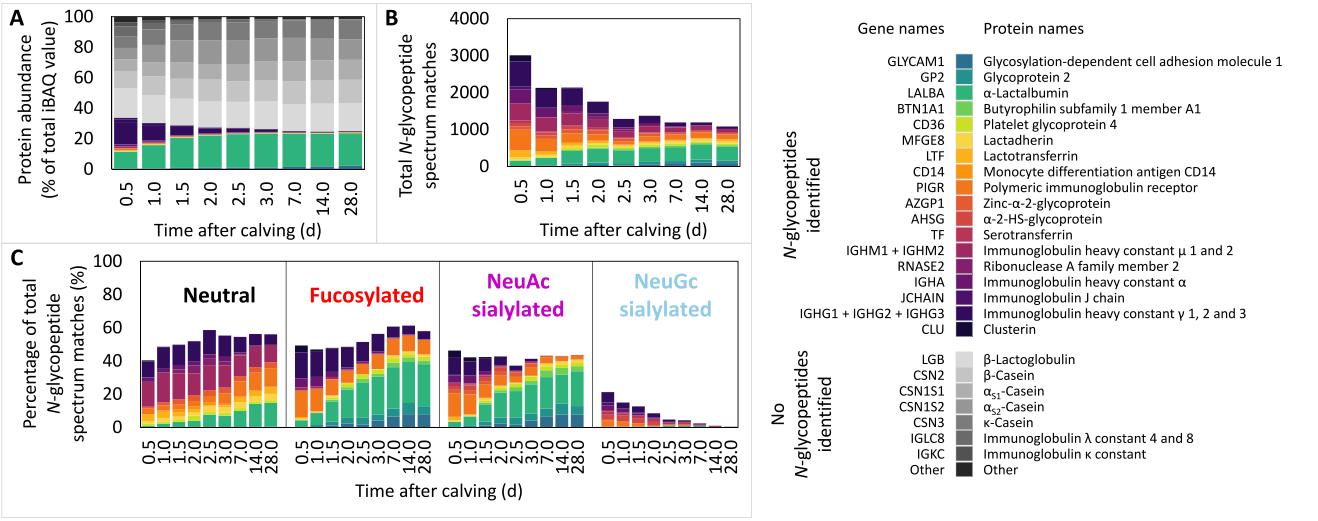

Fig. 1. Changes in the *N*-glycoproteome of bovine mammary secretions during lactation at 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 7.0, 14.0 and 28.0 d after calving. All data is summed across the four individual cows and illustrates compositional changes in time normalised to the total protein content. The identified proteins are indicated by their gene names. The *N*-glycoproteins are indicated in colour and ordered based on the differences in their abundance between the 28.0 d mature milk and 0.5 d colostrum proteomes (Supplementary Table S1). The remaining non-glycoproteins are indicated in greyscale. A) Protein abundances calculated as the intensity-based absolute quantification (iBAQ) value of a given protein expressed as the percentage of total iBAQ-values *per* lactation time point. Due to high sequence similarity, IGHM1 and IGHM2 could not be distinguished and were quantified together. Due to sharing an identical *N*-glycopeptide sequence, the abundances of IGHG1, IGHG2 and IGHG3 were summed. B) Absolute numbers of *N*-glycopeptide spectrum matches *per* identified *N*-glycoprotein and lactation time point. C) Glycan features, i.e. neutral (black), fucosylated (red), and sialylated with *N*-acetylneuraminic acid (Neu5Ac; purple) and *N*-glycolylneuraminic acid (Neu5Gc; light blue), of the *N*-glycans as a function of the glycoprotein they belong to and during lactation. The results are calculated as numbers of *N*-glycopeptide spectrum matches of a given glycoprotein expressed as a percentage of total *N*-glycopeptide spectrum matches *per* time point. The source data for these figures are provided in Supplementary Table S1 for protein abundances and in Supplementary Table S2 for *N*-glycopeptide abundances.

from *N*-glycoproteins significantly more abundant in the mature milk, primarily α-lactalbumin. Across the whole *N*-glycoproteome, when transitioning from colostrum to mature milk, we observed a general decrease in sialylation, attributed in particular to the decrease of Neu5Gc sialylation, along with the corresponding increase in neutral glycans (Fig. 1C). The neutral glycans in colostrum were primarily found on immunoglobulin heavy chain constant region glycopeptides, particularly from IGHG1, IGHG2 and IGHG3 and IGHM1 and IGHM2, whereas their gradual increase throughout lactation originated primarily from α-lactalbumin and PIGR glycopeptides. Fucosylated glycopeptides in colostrum were primarily found on PIGR and IGHG1-IGHG3; with the changes in protein composition during lactation, α-lactalbumin became the main contributor of fucosylated glycopeptides in the mature milk. Despite the levels of total Neu5Ac sialylation remaining constant throughout lactation, the origins of the Neu5Ac sialylated *N*-glycopeptides were different. Neu5Ac sialylated *N*-glycopeptides in colostrum primarily originated from the immunoglobulin heavy chain constants, the JCHAIN and PIGR. By contrast, the main contributors of Neu5Ac sialylated *N*-glycopeptides in the mature milk were α-lactalbumin, GLYCAM1 and GP2. The percentage of total *N*-glycopeptide spectrum matches carrying Neu5Gc sialylation, as also indicated in Fig. 1C, was found to decrease from a 21% in the 0.5 d colostrum to <1% in the 28.0 d mature milk. Neu5Gc sialylation was almost exclusively associated with proteins significantly more abundant in colostrum than in mature milk and in particular the immunoglobulin heavy chain constants, the JCHAIN and PIGR. A minor contribution of up to 1% of the total Neu5Gc sialylated *N*-glycopeptides was made by α-lactalbumin.

Remodelling of α-lactalbumin *N*-glycosylation during the transition from colostrum to mature milk

The amino acid sequence of bovine α-lactalbumin contains two Asn residues found in *N*-glycosylation motifs, i.e. Asn45 and Asn74, as highlighted in Supplementary Table S3. In our glycoproteomics study, we primarily detected *N*-glycosylation of bovine α-lactalbumin at Asn45 and we demonstrated for the first time the occurrence of *N*-glycosylation also at Asn74. The identification of *N*-glycosylation at these sites was confirmed by fragmentation spectra rich in ions providing both peptide backbone and attached glycan composition and sequence information (e.g. Supplementary Figs S1–S7, illustrating fragmentation spectra of *N*-glycopeptides Trp26-Glu49 and Asp63-Lys79, covering *N*-glycosylation sites Asn45 and Asn74, respectively). While we identified *N*-glycosylation at both Asn45 and Asn74, we found *N*-glycosylation site occupancy to differ between these sites, averaging at 35% spectral counts for Asn45 and 4% for Asn74; no substantial trend during lactation or differences between the individual cows were observed with respect to site occupancy (Supplementary Fig. S8). We hypothesise that the observed partial occupancies were caused by the amino acid residues present in the *N*-glycosylation motif and/or the proximity of *N*-glycosylation sites to structural features such as disulphide bridges and Ca^{2+} ion-binding residues. To visualise these features on the primary protein sequence, we performed an inter-species amino acid sequence alignment of all intact, mammalian α-lactalbumin sequences in UniProt KB, shown in Supplementary Table S3. Asn45 was conserved in 75 of the 114 investigated species. Of these, the second amino acid residue in the *N*-glycosylation motif was Gly in 50 species, followed by Asp in 18 species (including bovines

and other dairy ruminants) and Glu in the remaining 7 species (including humans and related hominids). The third amino acid residue in the *N*-glycosylation motif containing Asn45 was Ser in all 75 species. The influence of the second amino acid residue of the *N*-glycosylation motif and site occupancy will be examined in the *Discussion* section of this study, based on inter-species comparison of α-lactalbumin *N*-glycosylation site occupancy information from literature. With respect to Asn74, it was conserved in an *N*-glycosylation motif in only 15 of the 114 interrogated species, where in 14 species, including bovine and other dairy ruminants, it occurred in the Asn-Ile-Ser motif, whereas in the 15th species it was found in the Asn-Ile-Thr motif. In all species, this *N*-glycosylation motif was bordered by Cys73 and Cys77 and was located in close proximity on the primary protein sequence to the amino acid residues involved in Ca^{2+} ion binding, i.e. Lys79, Asp82, Asp84, Asp87 and Asp88. The proximity of the *N*-glycosylation motif containing Asn74 to these structural elements may explain the low occupancy observed at this site due to presumable competition between *N*-glycosylation and structure formation.

Due to the considerably higher *N*-glycosylation occupancy occurring at Asn45, we were able to identify a broad repertoire of 30 compositionally and structurally different *N*-glycans. While Byonic, the glycoproteomics database search software, identified glycopeptides with glycan compositions, it did not identify glycan structures. Nevertheless, by manual analysis of the fragmentation spectra and particularly of the ions resulting from the fragmentation of glycosidic linkages, we were able to curate the glycan compositions and derive structural information on the identified *N*-glycans. A detailed overview of the *N*-glycosylation identified at Asn45 *per* individual cow and lactation time point is shown in Fig. 2A. These results indicated that the common features of α-lactalbumin glycans at site Asn45 were 1) LacdiNAc motifs, 2) LacNAc motifs, 3) Neu5Ac sialylation and 4) fucosylation, all of which showed changes during lactation. These compositional and structural glycan features were verified by manual curation of the *N*-glycopeptide fragmentation spectra. The 2HexNAc oxonium ion (*m/z* = 407.1660 Th), observed in the fragmentation spectra shown in Supplementary Figs S1–S5, formed exclusively as a result of fragmentation of a LacdiNAc motif-containing *N*-glycan. The presence of the LacNAc motif was indicated by a relatively high intensity of the HexNAcHex oxonium ion (*m/z* = 366.1395 Th; Supplementary Figs S1 and S3). This ion was also present at lower intensity in fragmentation spectra of *N*-glycopeptides not containing the LacNAc motif, where it instead originated from the common oligosaccharide core structure shared by all *N*-glycans (e.g. Supplementary Figs S2 and S4–S6). The occurrence of Neu5Ac sialylation was confirmed by the formation of Neu5Ac (*m/z* = 292.1027 Th), (Neu5Ac-18) (*m/z* = 274.0921 Th) and (Neu5Ac-36) (*m/z* = 256.0816 Th) oxonium ions observed in Supplementary Figs S3 and S4. Additionally, the HexNeu5Ac (*m/z* = 454.1555 Th) and HexNAcNeu5Ac (*m/z* = 495.1821 Th) oxonium ions indicated Neu5Ac sialylation of LacNAc (Supplementary Fig. S3) and LacdiNAc motifs (Supplementary Fig. S4), respectively. Fucosylation occurred mostly on the *N*-glycan core, evidenced by the (Peptide +2HexNAcFuc) ion (*m/z* = 1103.4735 Th, $[M + 3H]^{3+}$) that can be observed in Supplementary Figs S1–S5. To a lesser extent, antennary fucosylation was also identified, as illustrated by the (2HexNAc + Hex + Fuc) oxonium ion (*m/z* = 715.2768 Th) in Supplementary Fig. S2. Supplementary Figs S5 and S6 additionally illustrate representative fragmentation spectra of minor *N*-glycopeptide species containing a glycan sialylated with Neu5Gc and a phosphorylated high-mannose glycan, respectively, while depicted in Supplementary Fig. S9 is the fragmentation spectrum of the non-glycosylated, unmodified, peptide covering Asn45.

All of the above described structural and compositional features of the *N*-glycans identified at Asn45 showed changes during the transition from the 0.5 d bovine colostrum to the 28.0 d mature milk, which are summarised in Fig. 2B. Fucosylation and Neu5Ac sialylation decreased from 1.0 ± 0.1 (mean ± standard deviation) and 0.8 ± 0.1 residues/glycan in the colostrum (0.5–3.0 d) samples, to 0.7 ± 0.1 and 0.6 ± 0.1 residues/glycan in the 28.0 d mature milk, respectively. Neu5Gc sialylation was detected, albeit at very low levels of up to 0.1 residues/glycan across all analysed samples. The number of LacdiNAc motifs decreased from on average 1.8 ± 0.1 motifs/glycan in the 0.5 d colostrum to 1.3 ± 0.0 motifs/glycan in the 28.0 d mature milk, simultaneously with the increase in the number of LacNAc motifs from 0.3 ± 0.1 to 0.8 ± 0.1 motif/glycan.

Due to the very low occupancy of 4% at Asn74, both the total number of identifications (146 *N*-glycopeptide spectra across 4 cows, 9 time points and 3 replicates) and the number of *N*-glycoforms identified (i.e. 4) at this site were considerably lower than those identified for the higher occupied (i.e. 35%) Asn45 (i.e. 30 *N*-glycoforms identified by 2,603 *N*-glycopeptide spectra across all individual animals, time points and replicates). Consequently, there was not sufficient information available on *N*-glycosylation compositional and structural changes over time at Asn74. The only confident observation we could make was that, albeit at considerably occupancy, the detectable *N*-glycosylation of Asn74 exhibited similar features as those described above for Asn45, with N6H3F1A1 being the glycan composition most abundantly identified at each of these sites. Supplementary Figs S4 and S7 illustrate representative fragmentation spectra of *N*-glycopeptides modified with N6H3F1A1 at Asn45 and Asn74, respectively.

From structural modelling of the doubly *N*-glycosylated bovine α-lactalbumin occupied with N6H3A1F1 at both Asn45 and Asn74 (Fig. 3A), it was apparent that the two 2.1 kDa potential glycan moieties constituted a large portion of the protein size and mass, as the protein backbone (without signal peptide) has a molecular weight of 14.2 kDa. Furthermore, while both Asn45 and Asn74 appeared to be located in surface-accessible loops, Asn74 was found on the primary structure in proximity of cysteine residues that engage in intra-molecular disulphide bridges, i.e. Cys73-Cys91 and Cys61-Cys77 (Supplementary Table S3), as well as on the tertiary structure in proximity of the amino acid residues involved in the binding of the Ca^{2+}, with distances of 12.51–16.98 Å measured with ChimeraX v1.5 on the structure shown in Fig. 3A between the side chain of Asn74 and the side chains of Lys79, Asp82, Asp84, Asp87 and Asp88. In order for α-lactalbumin to play a role in lactose synthesis, it needs to be able to form a heterodimer with B4GALT1. We modelled the structure of the doubly *N*-glycosylated bovine α-lactalbumin into the lactose synthase heterodimer complex (Fig. 3B), to verify whether it was possible for the complex to form. From the structure shown in Fig. 3B, it became apparent that neither

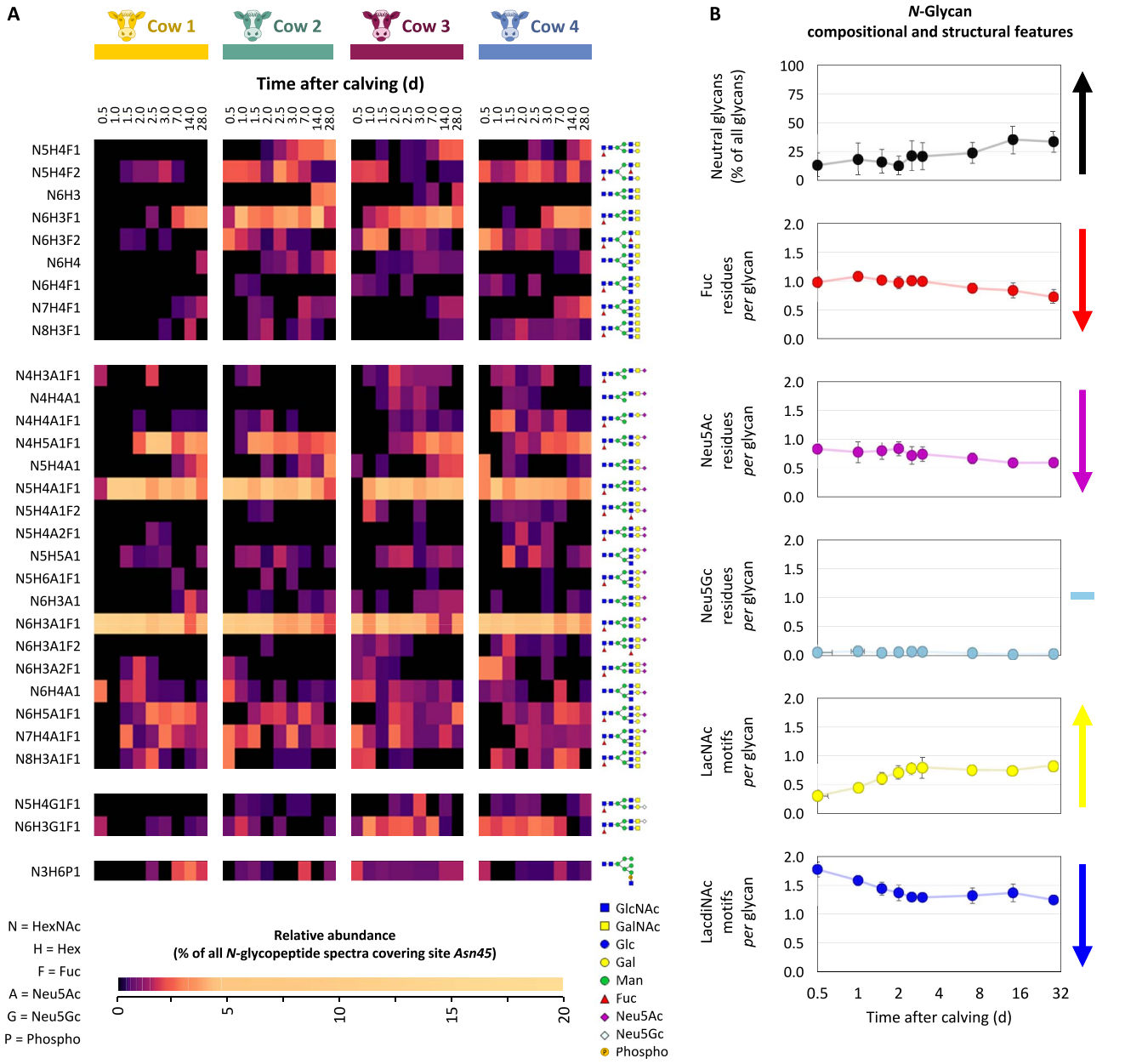

Fig. 2. Changes in the *N*-glycosylation profiles of bovine α-lactalbumin site Asn45 in the interval 0.5–28.0 d after calving in the bovine milk of the four cows, i.e. cow 1, cow 2, cow 3 and cow 4. A) Heatmap depicting the macro- and micro-heterogeneity of the *N*-glycosylation determined based on spectral counts. Clustering was performed on *N*-glycan structural features, from top to bottom: Neutral, sialylated with Neu5Ac, sialylated with Neu5Gc, and phosphomannose. The glycan composition is indicated to the left of the heatmap and to the right are proposed corresponding glycan structures. B) Dynamics of the *N*-glycan compositional and structural features during lactation averaged across the four cows. From top to bottom, the panels describe changes in neutral glycans, fucosylation, sialylation with Neu5Ac or Neu5Gc, and quantification of LacNAc and LacdiNAc motifs, respectively. The error bars represent the standard deviation of values from the four cows. The increasing, decreasing and no change trends were determined based on Student's t-test probability determined between the values at 0.5 and 28.0 d at a significance level of 0.05. Abbreviations: Gal = galactose; Glc = glucose; man = mannose; hex = hexose; GalNAc = *N*-acetylgalactosamine; GlcNAc = *N*-acetylglucosamine; HexNAc = *N*-acetylhexosamine; Fuc = fucose; Neu5Ac = *N*-acetylneuraminic acid; Neu5Gc = *N*-glycolylneuraminic acid; LacNAc = *N*-acetyllactosamine; LacdiNAc = *N,N'*-diacetyllactosamine; Phospho = phosphorylation.

of the α-lactalbumin *N*-glycans was located at the interaction interface between the two proteins and no clashes were visible between the structures of the two proteins. This indicated that, provided correct forming of the tertiary structure of the doubly *N*-glycosylated α-lactalbumin, this protein should be able to interact with B4GALT1 and be active in the synthesis of lactose, similarly to non-glycosylated α-lactalbumin.

Discussion

In this study, we monitored the changes in bovine α-lactalbumin *N*-glycosylation during lactation in the context of the *N*-glycoproteome, analysed directly from the mammary gland secretions, in the range of colostrum (0.5 d after calving) to mature milk (28.0 d). Overall, our study found that protein *N*-glycosylation was much more diverse in colostrum

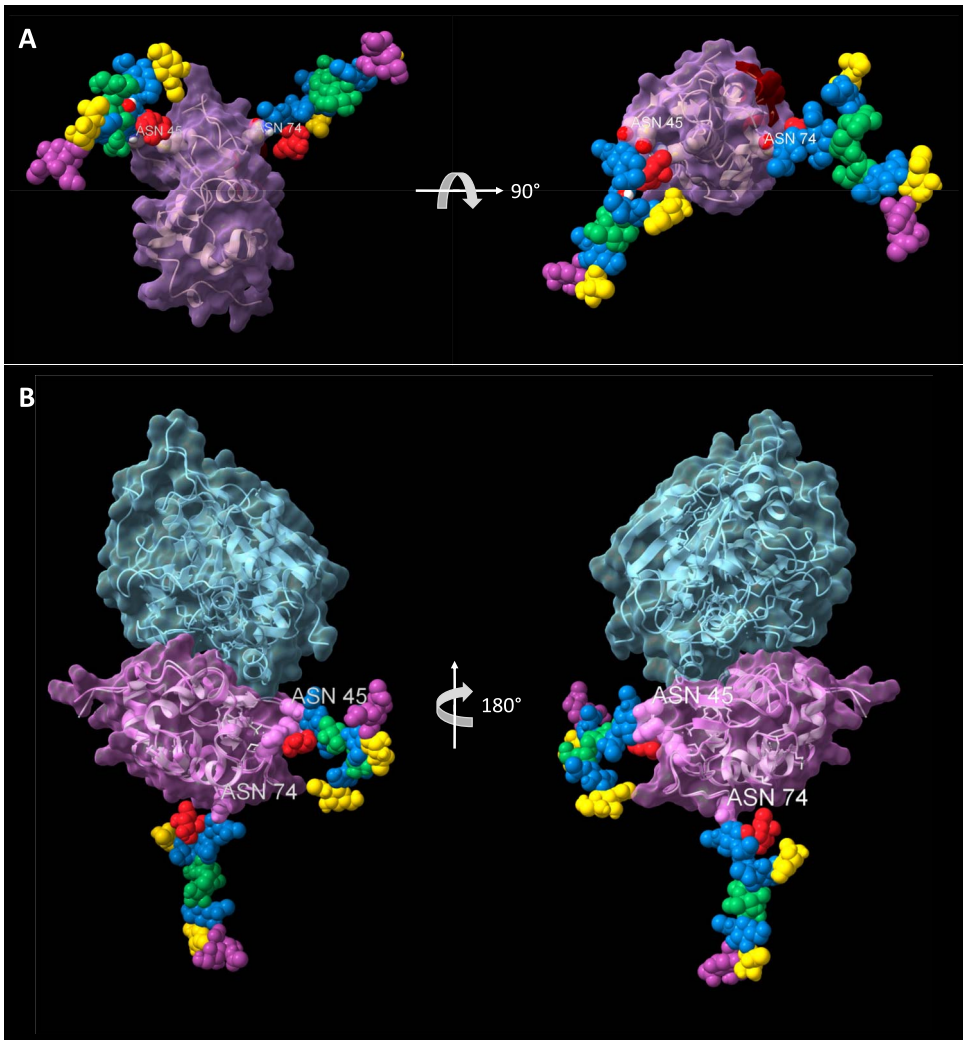

Fig. 3. Structural models of doubly-*N*-glycosylated bovine α-lactalbumin. A) Glycoprotein monomer occupied with the N6H3A1F1 glycan at Asn45 and Asn74 modelled using the glycoprotein builder tool of GLYCAM-web (glycam.org) based on the 1F6S PDB structure of the protein (Chrysina et al. 2000). Highlighted in red are the amino acid residues responsible for the binding of the Ca^{2+} ion. B) Structural model of the bovine lactose synthase heterodimer complex between doubly-*N*-glycosylated bovine α-lactalbumin and the catalytic domain of bovine B4GALT1. Depicted in light blue is the catalytic domain of B4GALT1 from the 2FYC PDB structure (Ramakrishnan et al. 2006). The recombinant mouse α-lactalbumin from the original 2FYC PDB structure is shown in silver; overlayed in pink is the doubly-*N*-glycosylated bovine α-lactalbumin modelled in panel (A). The surfaces of the bovine proteins are shown at 70% transparency. The monosaccharides in the glycan structures are coloured according to Varki et al. (2009): *N*-acetylglucosamine (N) – Blue, *N*-acetylgalactosamine (N) – Yellow, mannose (H) – Green, *N*-acetylneuraminic acid (A) – Purple, and fucose (F) – Red.

than in mature milk, and that the *N*-glycoprofile of the samples was to a greater extent influenced by the changes in protein composition during lactation and to a lesser extent by the protein-specific changes in *N*-glycosylation. Notably, Neu5Gc sialylation was contributed exclusively by signature proteins of colostrum, particularly immunoglobulin heavy chain constants, the JCHAIN and PIGR, while the mature milk proteins, including α-lactalbumin, only carried Neu5Ac sialylation (Fig. 1). The decrease in Neu5Gc sialylation during the transition from colostrum to mature milk was therefore directly caused by the decrease in the abundance of *N*-glycoproteins carrying this type of sialylation.

The *N*-glycosylation of a protein is directly influenced by the available glycosylation machinery of the cell producing the protein. This can be observed, for instance, in differences between the *N*-glycosylation of wild type and recombinant proteins expressed in different host cells (Yu et al. 2010; Croset et al. 2012; Lin et al. 2018). With respect to our data, the differences in *N*-glycosylation, particularly Neu5Gc sialylation, and protein composition between colostrum and mature milk (Fig. 1) support the fact that signature colostrum and signature mature milk *N*-glycoproteins likely had different origins and were produced by different cells. It is well-known that the immunoglobulins, including the JCHAIN, are produced by B-cells (Max and Korsmeyer 1985; Hoffman et al. 2016), whereas milk *N*-glycoproteins such as α-lactalbumin, GLYCAM1 and the milk fat globule membrane proteins (including BTN1A1, CD36 and MFGE8) are produced by mammary epithelial cells (Dowbenko et al. 1993; Groenen et al. 1995; Spitsberg et al. 1995; Banghart et al. 1998; Riley et al. 2008; Cebo and Martin 2012; Brew 2013), therefore the differences observed in their Neu5Gc sialylation can be further substantiated by their different origins. In this study, we identified Neu5Gc as the signature feature of bovine colostrum *N*-glycosylation in general (Fig. 1C). The sialylation presented by α-lactalbumin was almost exclusively

Neu5Ac (Fig. 2B), differentiating it as a protein produced directly by the mammary epithelial cells (Riley et al. 2008; Brew 2013), rather than being transported into milk following secretion by different cells.

Bovine α-lactalbumin has often been described as being partially glycosylated, with up to 10% of the protein harbouring *N*-glycosylation (Barman 1970; Hopper and McKenzie 1973; Chandrika 1999; Slangen and Visser 1999; Valk-Weeber et al. 2020), whereas we here report on average 35% site occupancy of Asn45 based on spectral counts (Supplementary Fig. S8). This discrepancy can be explained by the different methods used for determining *N*-glycosylation site occupancy. Our glycoproteomics LC–MS/MS method was designed to selectively fragment glycopeptides, allocating more time to their fragmentation relative to the fragmentation of non-glycosylated peptides. The resulting increased number of glycopeptide fragmentation spectra could therefore be responsible for overestimating *N*-glycosylation site occupancy determined based on spectral counts. While no information was found in literature with respect to the *N*-glycosylation site occupancy of bovine α-lactalbumin in colostrum or transitional milk, we demonstrated here that occupancy was constant during lactation and between individual animals, averaging at 35% for Asn45 and 4% for Asn74.

While the putative *N*-glycosylation site at Asn45 was conserved in the α-lactalbumin orthologues of 75 out of 114 interrogated species (Supplementary Table S3), considerable inter-species differences in site occupancy have been reported in literature, ranging from no occupancy to full occupancy. Interestingly, these differences in Asn45 occupancy appeared to correlate with the identity of the amino acid occurring at site 46 (Hopp and Woods 1979). Human α-lactalbumin, containing Glu46, was not *N*-glycosylated at Asn45 (Giuffrida et al. 1997). Ruminant α-lactalbumin from buffalo, cow, and goat where residue 46 is Asp, exhibited low *N*-glycosylation occupancy (Macgillivray et al. 1979; Kim and Jimenez-Flores 1994; Chianese et al. 2004; D'Ambrosio et al. 2008). Cat, mouse, rat and rabbit α-lactalbumins, each containing Gly46 exhibit high or full *N*-glycosylation occupancy at Asn45 (Hopp and Woods 1979; Prasad et al. 1979; Nagamatsu and Oka 1980; Prasad et al. 1982; Halliday et al. 1990). This is in line with studies investigating the Asn-Xxx-Ser *N*-glycosylation motif, showing high glycosylation efficiency when Xxx = Gly and decreased glycosylation efficiency when Xxx = Glu or Xxx = Asp (Bause and Legler 1981; Mononen and Karjalainen 1984; Shakin-Eshleman et al. 1996; Kasturi et al. 1997; Breuer et al. 2001; Rao and Bernd 2010; Malaby and Kobertz 2014). Since α-lactalbumin could occur in the full range of no to full occupancy at Asn45 across different species, we speculate that the protein is active in the synthesis of lactose irrespective of its *N*-glycosylation occupancy or microheterogeneity at Asn45. In support of this hypothesis, previous studies have shown that the non-glycosylated and *N*-glycosylated bovine α-lactalbumin fractions exhibited equal lactose synthase activities in complex with B4GALT1 (Barman 1970; Proctor et al. 1974). Furthermore, our structural modelling showed that it is conceptually possible to form the bovine lactose synthase complex with *N*-glycosylated α-lactalbumin, with the *N*-glycans neither located at the interaction interface, nor clashing with the catalytic domain of B4GALT1 (Fig. 3B).

The detectable, albeit very low, occupancy identified here at Asn74 (Supplementary Fig. S7, Supplementary Fig. S8) may be explained by the localization of this site on the protein, despite its high calculated surface availability. The Asn74-Ile-Ser motif is bordered by two cysteine residues that are each engaged in two different intra-molecular bonds in the structure of α-lactalbumin (Vanaman et al. 1970). We hypothesise that *N*-glycosylation of Asn74 competes with the formation of the disulphide bridges. Brune et al. (2023) showed that preventing the formation of disulphide bridges during the production of β-lactoglobulin resulted in disordered structure. By similarity, if bovine α-lactalbumin Asn74 is glycosylated in competition with the formation of disulphide bridges, the consequent absence of the Cys61-Cys77 and Cys73-Cys91 bridges likely results in the formation of a protein with altered structure. Consequently, functionalities of α-lactalbumin that depend on protein structure and on the presence of disulphide bridges, such as its activity in lactose synthesis or the bactericidal activity of disulphide-linked α-lactalbumin-derived peptides (Pellegrini et al. 1999), will either be altered or lost. Furthermore, due to the amino acid residues which bind the Ca^{2+} ion being located in the primary sequence between Cys73 and Cys91 (Chrysina et al. 2000), this structurally-altered protein, glycosylated at Asn74 and lacking the Cys61-Cys77 and Cys73-Cys91 disulphide bridges, may also lack the ability of binding the Ca^{2+} ion. Therefore, *N*-glycosylation at Asn74 may be undesirable.

We are not the first to analyse the glycosylation of α-lactalbumin. Initial research on glycosylated α-lactalbumin detected the glycosylated protein by its altered migration on native polyacrylamide gel electrophoresis, identifying glycosylation by analysis of chemically-released monosaccharides (Barman 1970; Hindle and Wheelock 1970; Hopper and McKenzie 1973). Nearly three decades later, Slangen and Visser (1999) and Chandrika (1999) analysed the intact mass of glycosylated α-lactalbumin and of its enzymatically-released *N*-glycans by LC–MS, inferring glycan compositions and structures solely based on intact glycan mass and identifying the *N*-glycosylation site Asn45 by indirect analysis. In a more recent study, Valk-Weeber et al. (2020) analysed the *N*-glycosylation of whey proteins, including bovine α-lactalbumin, combining the intact mass analysis of released *N*-glycans with sequential exoglycosidase digestion of the glycans for composition and structure analysis. In these studies combined, a series of ~15 *N*-glycan compositions corresponding to complex di- and triantennary structures with LacNAc and LacdiNAc motifs, with or without monofucosylation and/or monosialylation with Neu5Ac were identified. Chandrika (1999) identified the main sialylated and neutral glycan compositions as N6H3A1F1 and N6H3F1, respectively. Slangen and Visser (1999) found that the sialylated glycoforms (notably N6H3A1F1, N5H4A1F1 and N4H5A1F1) were more abundant than the neutral glycoforms (notably N6H3F1, N5H5F1 and N4H5F1). In the bovine α-lactalbumin glycan fingerprints reported by Valk-Weeber et al. (2020), N6H3A1F1 and N5H4A1F1 were the main sialylated glycoforms identified, while N6H3F1 was the most abundant neutral glycoform. In comparison with previous studies, our LC–MS/MS-based glycoproteomics approach at glycopeptide level ensured, through the oxonium-ion triggered hybrid dissociation methods, optimal fragmentation and sequencing of both the amino acid backbone and the glycan structure, hereby providing direct and unbiased information on glycan localization, composition and structure, which facilitated for the first time the identification of *N*-glycosylation also

occurring at the Asn74 site in bovine α-lactalbumin. Our study expands substantially and importantly beyond the existing knowledge of bovine α-lactalbumin *N*-glycosylation with the localization-specific identification of a broader repertoire of *N*-glycans, including difucosylated and disialylated species, as well as glycans sialylated with Neu5Gc, and a minor species of phosphorylated high-mannose glycan (Fig. 2), with all identifications manually curated at fragmentation spectrum level. The increased sensitivity of newer-generation instrumentation in part enabled the identification of a higher diversity of glycoforms in our study. In line with previous studies, here we also found N6H3A1F1, N5H4A1F1 and N4H5A1F1 as the overall most abundant glycoforms at Asn45 in mature milk α-lactalbumin, followed to a lesser extent by N6H3F1 and N5H4F1. While all of these previous findings analysed glycosylated α-lactalbumin in bovine milk, information in literature is lacking on the glycosylation of the protein in colostrum or during the transition from colostrum to mature milk. Protein-specific compositional and structural changes in *N*-glycosylation (i.e. Neu5Ac sialylation, proportion of neutral glycans, fucosylation, occurrence of LacNAc and LacdiNAc motifs) during lactation were evident in the detailed analysis of the *N*-glycosylation of α-lactalbumin. Our study shows the maturation in *N*-glycosylation microheterogeneity of α-lactalbumin during the transition from colostrum to mature milk in four individual cows (Fig. 2). Notably, bovine colostrum α-lactalbumin *N*-glycosylation was characterised by higher levels of Neu5Ac sialylation and higher ratios of LacdiNAc to LacNAc motifs. Consequently, the N6H3A1F1 glycan composition, containing two LacdiNAc motifs and a Neu5Ac moiety, was the single most abundant glycoform identified in the 0.5 d colostrum, followed to a lower extent by the N5H4A1F1 composition, containing one LacdiNAc and one LacNAc motif and a Neu5Ac moiety, while the N4H5A1F1 composition with two LacNAc motifs was not detected at all.

While in the case of bovine IgG, where the change in Neu5Ac to Neu5Gc sialylation during the transition of colostrum to mature milk may have served a role in the protection and transport of the IgG for the crucial passive systemic immunity of the calf (Gazi et al. 2023), the same could not be said about the changes in bovine α-lactalbumin glycosylation. If, as speculated above, the *N*-glycosylation of Asn45 is inconsequential to the activity of the protein and the minor *N*-glycosylation of Asn74 is undesirable for the structural and functional integrity of the protein, then the identified glycosylation and changes thereof during lactation were likely representative of changes in the glycosyltransferase repertoire of the mammary epithelial cells, but not reflective of transitions in the functionality of α-lactalbumin during lactation.

In conclusion, owing to advances in mass spectrometry and glycoproteomics technology, we were able to consolidate and expand the existing knowledge of *N*-glycosylation changes in bovine α-lactalbumin from colostrum to mature milk, also proving for the first time the occurrence of *N*-glycosylation at Asn74. Our research forms the basis for future studies investigating the relation between the here described α-lactalbumin *N*-glycosylation and the biological functionality of the protein.

Materials and methods

Here, we analysed the peptide-centric mass spectrometry data, acquired for the whole bovine mammary gland secretions described in Gazi et al. (2023), for the glycosylation of α-lactalbumin.

Chemicals and reagents

Tris(2-carboxyethyl)phosphine (TCEP), Tris, chloroacetamide (CAA), sodium deoxycholate (SDC) and formic acid (FA) were purchased from Sigma-Aldrich (Darmstadt, Germany). The Pierce BCA Protein Assay Kit (23225) for the bicinchoninic acid (BCA) assay was sourced from Thermo Scientific, Rockford, Illinois, USA. Trifluoroacetic acid (TFA) was purchased from Fisher Scientific (Landsmeer, The Netherlands). PepMap Trap Cartridges (5 μm C18 300 μm X 5 mm) were sourced from Thermo Fisher Scientific (Germering, Germany). Ultrapure MilliQ water was prepared with the system sourced from Merck Millipore, Darmstadt, Germany. Sequencing-grade endoproteinase GluC and broad range protease inhibitor cocktail (cOmplete, EDTA-free) were sourced from Roche Diagnostics (Mannheim, Germany). Sequencing-grade trypsin was sourced from Promega (Madison, Wisconsin, USA). Oasis PRiME HLB 96-well plates (10 mg sorbent *per* well) for solid phase extraction (SPE) were sourced from Waters (Etten-Leur, The Netherlands). C18 column material (Poroshell 120 EC-C18, 2.4 μm) was sourced from Agilent Technologies (Amstelveen, The Netherlands).

Bovine milk samples

A set of 36 samples from four individual Dutch Holstein-Friesian dairy cows, referred to as Cow 1, Cow 2, Cow 3 and Cow 4, and sampled at 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 7.0, 14.0 and 28.0 d after calving, were obtained from a farm in The Netherlands. Following collection, the samples were frozen and stored at −20 °C. The samples collected at 0.5–3.0 d, at 7.0–14.0 d and at 28.0 d after calving were further referred to as colostrum, transitional and mature milk, respectively.

Protein content

The protein content of the bovine colostrum and milk samples was determined by the BCA assay using the Pierce BCA Protein Assay Kit and 96-well plate method, according to manufacturer's instructions.

Analysis of bovine colostrum and milk proteome and *N*-glycoproteome

Amounts of 10 μg of protein of each whole bovine colostrum and milk sample were denatured, reduced and alkylated in a buffer containing 10 mM TCEP, 100 mM Tris, 40 mM CAA and 1% (*m/v*) SDC at pH 8.5, followed by proteolytic digestion with GluC (1:50 *m/m* GluC:protein, 4 h at room temperature) and trypsin (1:50 *m/m* GluC:protein, overnight at 37 °C). TFA was added to a final concentration of 1% (*v/v*) to precipitate SDC and inhibit the proteolytic reactions. The resulting peptides were extracted by SPE using Oasis PriME HLB 96-well plates according to the instructions of the manufacturer, followed by drying in a vacuum concentrator and storage at −20 °C until further analysis.

The proteolytic digests were further analysed by LC–MS/MS. Amounts of 800 ng of digested protein (based on the initial quantity digested) were injected on an Ultimate 3,000 UHPLC system (Thermo Fisher Scientific, Germering, Germany) coupled online to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA). Separation of the peptides was performed by reversed-phase chromatography using PepMap Trap

Cartridges and C18 analytical columns (50 cm length, 50 μm inner diameter). Each sample was injected in triplicate with identical chromatographic and MS methods, but with differing fragmentation methods. The fragmentation methods employed included the regular higher-energy collisional dissociation (HCD; normalised collision energy of 29%) and two hybrid, oxonium ion-triggered fragmentation methods, dedicated for the analysis of glycopeptides (Reiding et al. 2018), i.e. product-dependent stepped HCD (normalised collision energy steps of 10, 25 and 40%) and product-dependent electron-transfer/higher-energy collisional dissociation (EthcD). The chromatography gradient and mass spectrometer settings have been described in detail and can be found in Gazi et al. (2023).

Database search and identification of bovine milk protein *N*-glycosylation

Database searches were performed using MaxQuant v 1.5.3.30 (Tyanova et al. 2016a) against an optimised protein database, consisting of the non-redundant *Bos taurus* (taxon ID 9913) reference proteome (ID UP000009136) from UniProt.org, the bovine immunoglobulin sequences from the reference directory of the international ImMunoGeneTics information system (IMGT RefSeq, https://www.imgt.org/vquest/refseqh.html) and the contaminants FASTA file from the installation folder of MaxQuant software, from which the bovine sequences were excluded, as described in Gazi et al. (2023). The abundance of a given *N*-glycopeptide in the total proteome was calculated as the intensity-Based Absolute Quantification (iBAQ) value of the *N*-glycoprotein expressed as a percentage of the total iBAQ-values *per* lactation time point. Significant differences in protein abundances between the proteomes of the 0.5 d colostrum and 28.0 d mature milk samples across the four individual cows were established by statistical analysis using Perseus v 1.6.13.0 (Tyanova et al. 2016b). The significantly-different proteins were determined based on a t-test at a 0.01 false discovery rate level based on a 250-permutation-based calculation and s0 = 0.1. The results were visualised in the form of a volcano plot and presented in Gazi et al. (2023).

In addition to the work described in Gazi et al. (2023), here we performed the analysis of the full glycoproteome detectable within this experimental approach. Based on the results of the bovine proteome identification, a small protein database containing only the identified sequences was created for the *N*-glycoprotein search. The database search for *N*-glycoproteins was performed with Byonic v4.5.2 (Protein Metrics, Cupertino, California, USA). A decoy database was created by reversing the protein sequences of the target database. The *N*-glycan database used was the one containing 2,440 compositions, including high mannose, hybrid and complex glycan antennae containing LacNAc, LacdiNAc and all combinations thereof, as well as sialylation with Neu5Ac and the non-human Neu5Gc, as well as all combinations thereof. Proteolytic cleavage sites were defined C-terminal of arginine, lysine, and aspartic and glutamic acid residues with 3 missed cleavages, but digestion specificity was set to semi-specific. Fragmentation type was set either to higher-energy collisional dissociation (HCD) for the samples analysed with HCD or product ion-triggered stepping HCD, or to both HCD and electron-transfer/higher-energy collisional dissociation (EThcD) for the samples analysed with product ion-triggered EThcD. Cysteine carbamidomethylation was set as a fixed modification. Methionine and tryptophan oxidation, N-terminal cyclisation of glutamine and glutamic acid to pyroglutamic acid, and serine, threonine and tyrosine phosphorylation were all searched as rare variable modifications, whereas *N*-glycosylation was searched as a common variable modification. A maximum of one rare and one common variable modification were allowed *per* peptide. Further post-processing included filtering of the data based on |Log Prob| ≥ 1 and score ≥ 150.

Quantification results were calculated based on spectral counts. Normalisation was performed *per* cow and *per* lactation time point relative to all spectra covering the glycosylation site, whereby the sum of all glycoforms and the non-glycosylated site amounts to 100% *per* time point.

The mass spectrometry raw data and complete MaxQuant and Byonic search results of the glycoproteomics analyses on the bovine colostrum, transitional and mature milk samples have been deposited to the ProteomeXchange Consortium (Deutsch et al. 2020) via the PRIDE (Perez-Riverol et al. 2022) partner repository with the dataset identifier PXD050331.

Based on the identified glycan composition, proposed intact glycan, oxonium ion and glycopeptide fragment ion structures were built using GlycoWorkBench 2.1 build 146, according to the symbol nomenclature for glycan representation of the Consortium for Functional Glycomics (Varki et al. 2009).

Amino acid sequence alignment of α-lactalbumin across different species

Protein sequences orthologous to bovine α-lactalbumin were downloaded from UniProt Knowledge Base (uniprot.org) on July 12th, 2024 under the conditions: class = Mammalia, gene name = LALBA, sequence status = complete and one sequence *per* species. Amino acid sequence alignment of bovine α-lactalbumin against its orthologs from different mammalian species was performed with the multiple sequence alignment program Clustal Omega on the European Bioinformatics Institute (EBI) mirror (https://www.ebi.ac.uk/jdispatcher/msa/clustalo). The simplified taxonomic rank of the different organisms was determined using the CommonTree tool of the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi). The different protein sequences were identified by their UniProt header. All cysteine residues were highlighted in black. Highlighted in red were the amino acid residues responsible for binding a Ca^{2+} ion *per* protein, which were annotated based on sequence identity to the Ca^{2+} ion-binding residues in the bovine protein, as described in UniProt. The *N*-glycosylation motifs observed in the primary sequence were surrounded by purple borders, with the putatively *N*-glycosylated Asn residues highlighted in purple.

3D structure modelling of doubly *N*-glycosylated bovine α-lactalbumin and lactose synthase heterodimer complex

The modelling of the doubly *N*-glycosylated bovine α-lactalbumin 3D structure was based on the 1F6S PDB structure of the protein, to which the structure of the abundantly-identified N6H3A1F1 glycan was linked to Asn45 and Asn74 using the Glycoprotein Builder tool of GLYCAM-Web (glycam.org). The surface accessible surface areas (SASAs) calculated by NAccess within Glycoprotein Builder were 108.8+ and 99.8+ for Asn45 and Asn74, respectively, indicating high surface availability of both sites.

The structure of the doubly *N*-glycosylated α-lactalbumin was visualised with ChimeraX v1.5. Modelling of the doubly *N*-glycosylated α-lactalbumin into the lactose synthase complex was also performed using ChimeraX v1.5. For this purpose, the modelled structure of the glycosylated bovine α-lactalbumin was transposed onto the structure of the recombinant mouse α-lactalbumin in the 2FYC PDB structure (Ramakrishnan et al. 2006) and possible clashes of the glycosylated bovine α-lactalbumin with the catalytic domain of B4GALT1 were verified.

Author contributions

I.G., T.H. and A.J.R.H. conceptualised the project; I.G. performed all experiments; I.G. performed data curation and analysis with support from K.R.R.; I.G. was responsible for visualisation of the results and writing of the manuscript; all authors contributed to the interpretation of the findings and provided feedback for editing; A.J.R.H. and A.G. secured the funding for the project. All authors have read and agreed to the final version of the manuscript.

Inge Gazi (Conceptualization [equal], Data curation [lead], Formal analysis [lead], Investigation [lead], Methodology [equal], Visualization [lead], Writing—original draft [lead], Writing—review & editing [lead]), Karli R. Reiding (Data curation [supporting], Formal analysis [supporting], Methodology [equal], Supervision [supporting], Writing—review & editing [supporting]), André Groeneveld (Funding acquisition [supporting], Writing—review & editing [supporting]), Jan Bastiaans (Writing—review & editing [equal]), Thom Huppertz (Conceptualization [equal], Resources [supporting], Supervision [supporting], Writing—review & editing [supporting]), and Albert Heck (Conceptualization [equal], Funding acquisition [lead], Resources [lead], Supervision [lead], Writing—review & editing [supporting])

Supplementary material

Supplementary material is available at *Glycobiology Journal* online.

Funding

We acknowledge support from the Dutch Research Council (NWO) in the framework of the Innovation Fund for Chemistry (SATIN project 731.017.202). K.R.R. acknowledges support from NWO Veni project VI.Veni.192.058.

Conflict of interest statement. No conflicts of interest to declare.

References

Banghart LR, Chamberlain CW, Velarde J, Korobko IV, Ogg SL, Jack LJW, Vakharia VN, Mather IH. Butyrophilin is expressed in mammary epithelial cells from a single-sized messenger RNA as a type I membrane glycoprotein. *J Biol Chem*. 1998:**273**(7):4171–4179.

Barman T. Purification and properties of bovine milk glyco-α-lactalbumin. *Biochim Biophys Acta*. 1970:**214**(1):242–244.

Bause E, Legler G. The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr (Ser) for the N-glycosylation step during glycoprotein biosynthesis. *Biochem J*. 1981:**195**(3):639–644.

Breuer W, Klein RA, Hardt B, Bartoschek A, Bause E. Oligosaccharyltransferase is highly specific for the hydroxy amino acid in Asn-Xaa-Thr/Ser. *FEBS Lett*. 2001:**501**(2–3):106–110.

Brew K. α-Lactalbumin. In: McSweeney PLH, Fox PF, editors. *Advanced dairy chemistry: volume 1A: proteins: basic aspects*. 4th ed. Springer US: Boston, MA; 2013. pp. 261–273

Brew K, Vanaman TC, Hill RL. The role of alpha-lactalbumin and the A protein in lactose synthetase: a unique mechanism for the control of a biological reaction. *Proc Natl Acad Sci*. 1968:**59**(2):491–497.

Brodbeck U, Denton W, Tanahashi N, Ebner KE. The isolation and identification of the B protein of lactose synthetase as α-lactalbumin. *J Biol Chem*. 1967:**242**(7):1391–1397.

Brune SE, Hoppenreijs LJG, Kühl T, Lautenbach V, Walter J, Peukert W, Schwarz K, Imhof D, Boom RM, Krull R, et al. Precision fermentation as a route to modify β-lactoglobulin structure through substitution of specific cysteine residues. *Int Dairy J*. 2023:**147**:105772.

Cebo C, Martin P. Inter-species comparison of milk fat globule membrane proteins highlights the molecular diversity of lactadherin. *Int Dairy J*. 2012:**24**(2):70–77.

Chandrika UG. *Glycosylation of bovine α-lactalbumin: a thesis presented in partial fulfilment of the requirements for the degree of master of philosophy in biochemistry, at Massey University, New Zealand*. Palmerston North, New Zealand: Massey University; 1999.

Chianese L, Caira S, Lilla S, Pizzolongo F, Ferranti P, Pugliano G, Addeo F. Primary structure of water buffalo α-lactalbumin variants A and B. *J Dairy Res*. 2004:**71**(1):14–19.

Chrysina ED, Brew K, Acharya KR. Crystal structures of Apo- and Holo-bovine α-Lactalbumin at 2.2-Å resolution reveal an effect of calcium on inter-lobe interactions. *J Biol Chem*. 2000:**275**(47):37021–37029.

Croset A, Delafosse L, Gaudry J-P, Arod C, Glez L, Losberger C, Begue D, Krstanovic A, Robert F, Vilbois F, et al. Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells. *J Biotechnol*. 2012:**161**(3):336–348.

D'Ambrosio C, Arena S, Salzano AM, Renzone G, Ledda L, Scaloni A. A proteomic characterization of water buffalo milk fractions describing PTM of major species and the identification of minor components involved in nutrient delivery and defense against pathogens. *Proteomics*. 2008:**8**(17):3657–3666.

Deutsch EW, Bandeira N, Sharma V, Perez-Riverol Y, Carver JJ, Kundu DJ, García-Seisdedos D, Jarnuczak AF, Hewapathirana S, Pullman BS. The ProteomeXchange consortium in 2020: enabling 'big data'approaches in proteomics. *Nucleic Acids Res*. 2020:**48**(D1):D1145–D1152.

Dowbenko D, Kikuta A, Fennie C, Gillett N, Lasky LA. Glycosylation-dependent cell adhesion molecule 1 (GlyCAM 1) mucin is expressed by lactating mammary gland epithelial cells and is present in milk. *J Clin Invest*. 1993:**92**(2):952–960.

Gazi I, Reiding KR, Groeneveld A, Bastiaans J, Huppertz T, Heck AJ. Key changes in bovine milk immunoglobulin G during lactation: NeuAc sialylation is a hallmark of colostrum immunoglobulin G N-glycosylation. *Glycobiology*. 2023:**33**(2):115–125.

Giuffrida MG, Cavaletto M, Giunta C, Neuteboom B, Cantisani A, Napolitano L, Calderone V, Godovac-Zimmermann J, Conti A. The unusual amino acid triplet Asn-Ile-Cys is a glycosylation consensus site in human α-Lactalbumin. *J Protein Chem*. 1997:**16**(8):747–753.

Gopal PK, Gill HS. Oligosaccharides and glycoconjugates in bovine milk and colostrum. *Br J Nutr*. 2000:**84**(S1):69–74.

Groenen MAM, Dijkhof RJM, van der Poel JJ. Characterization of a GlyCAM1-like gene (glycosylation-dependent cell adhesion molecule 1) which is highly and specifically expressed in the lactating bovine mammary gland. *Gene*. 1995:**158**(2):189–195.

Halliday JA, Bell K, McKenzie HA, Shaw DC. Feline whey proteins: identification, isolation and initial characterization of α-lactalbumin, β-lactoglobulin and lysozyme. *Comp Biochem Physiol B*. 1990:**95**(4):773–779.

Hindle E, Wheelock J. Carbohydrates of bovine alpha-lactalbumin preparations. *Biochem J*. 1970:**119**(3):14P.

Hoffman W, Lakkis FG, Chalasani G. B cells, antibodies, and more. *Clin J Am Soc Nephrol*. 2016:**11**(1):137–154.

Hopp TP, Woods KR. Primary structure of rabbit. Alpha.-lactalbumin. *Biochemistry*. 1979:**18**(23):5182–5191.

Hopper K, McKenzie H. Minor components of bovine α-lactalbumin A and B. *Biochim Biophys Acta Protein Struct*. 1973:**295**(1):352–363.

Kasturi L, Chen H, Shakin-Eshleman SH. Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify

Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. *Biochem J*. 1997:**323**(2):415–419.

Kim HHY, Jimenez-Flores R. Comparison of milk proteins using preparative isoelectric focusing followed by polyacrylamide gel electrophoresis. *J Dairy Sci*. 1994:**77**(8):2177–2190.

Krištić J, Lauc G. Ubiquitous importance of protein glycosylation. In: Lauc G, Wuhrer M, editors. *High-throughput Glycomics and Glycoproteomics: methods and protocols*. Springer New York: New York, NY; 2017. pp. 1–12

Layman DK, Lönnerdal B, Fernstrom JD. Applications for α-lactalbumin in human nutrition. *Nutr Rev*. 2018:**76**(6):444–460.

Lin Y-H, Franc V, Heck AJR. Similar albeit not the same: In-depth analysis of Proteoforms of human serum, bovine serum, and recombinant human Fetuin. *J Proteome Res*. 2018:**17**(8): 2861–2869.

Macgillivray RT, Brew K, Barnes K. The amino acid sequence of goat α-lactalbumin. *Arch Biochem Biophys*. 1979:**197**(2):404–414.

Malaby HL, Kobertz WR. The middle X residue influences cotranslational N-glycosylation consensus site skipping. *Biochemistry*. 2014:**53**(30):4884–4893.

Max EE, Korsmeyer SJ. Human J chain gene. Structure and expression in B lymphoid cells. *J Exp Med*. 1985:**161**(4):832–849.

McGrath BA, Fox PF, McSweeney PL, Kelly AL. Composition and properties of bovine colostrum: a review. *Dairy Sci Technol*. 2016:**96**(2): 133–158.

Mononen I, Karjalainen E. Structural comparison of protein sequences around potential N-glycosylation sites. *Biochim Biophys Acta*. 1984:**788**(3):364–367.

Nagamatsu Y, Oka T. Purification and characterization of mouse α-lactalbumin and preparation of its antibody. *Biochem J*. 1980:**185**(1):227–237.

Pellegrini A, Thomas U, Bramaz N, Hunziker P, von Fellenberg R. Isolation and identification of three bactericidal domains in the bovine α-lactalbumin molecule. *Biochim Biophys Acta Gen Subj*. 1999:**1426**(3):439–448.

Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, Kundu DJ, Prakash A, Frericks-Zipper A, Eisenacher M. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res*. 2022:**50**(D1):D543–D552.

Prasad R, Hudson BG, Butkowski R, Hamilton JW, Ebner KE. Resolution of the charge forms and amino acid sequence and location of a tryptic glycopeptide in rat alpha-lactalbumin. *J Biol Chem*. 1979:**254**(21):10607–10614.

Prasad RV, Butkowski RJ, Hamilton JW, Ebner KE. Amino acid sequence of rat α-lactalbumin: a unique α-lactalbumin. *Biochemistry*. 1982:**21**(7):1479–1482.

Proctor SD, Wheelock JV, Davies DT. Heterogeneity of bovine α-bactalbumin. *Biochem Soc Trans*. 1974:**2**(4):621–622.

Ramakrishnan B, Ramasamy V, Qasba PK. Structural snapshots of β-1,4-galactosyltransferase-I along the kinetic pathway. *J Mol Biol*. 2006:**357**(5):1619–1633.

Rao RSP, Bernd W. Do N-glycoproteins have preference for specific sequons? *Bioinformation*. 2010:**5**(5):208–212.

Reiding KR, Bondt A, Franc V, Heck AJ. The benefits of hybrid fragmentation methods for glycoproteomics. *TrAC Trends Anal Chem*. 2018:**108**:260–268.

Riley LG, Sheehy PA, Williamson P, Wynn PC. The role of native bovine α-lactalbumin in bovine mammary epithelial cell apoptosis and casein expression. *J Dairy Res*. 2008:**75**(3):319–325.

Shakin-Eshleman SH, Spitalnik SL, Kasturi L. The amino acid at the X position of an Asn-X-Ser Sequon is an important determinant of N-linked Core-glycosylation efficiency (∗). *J Biol Chem*. 1996:**271**(11): 6363–6366.

Slangen CJ, Visser S. Use of mass spectrometry to rapidly characterize the heterogeneity of bovine α-lactalbumin. *J Agric Food Chem*. 1999:**47**(11):4549–4556.

Spitsberg VL, Matitashvili E, Gorewit RC. Association and Coexpression of fatty-acid-binding protein and glycoprotein CD36 in the bovine mammary gland. *Eur J Biochem*. 1995:**230**(3):872–878.

Takimori S, Shimaoka H, Furukawa JI, Yamashita T, Amano M, Fujitani N, Takegawa Y, Hammarström L, Kacskovics I, Shinohara Y. Alteration of the N-glycome of bovine milk glycoproteins during early lactation. *FEBS J*. 2011:**278**(19):3769–3781.

Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc*. 2016a:**11**(12):2301–2319.

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 2016b:**13**(9):731–740.

Valk-Weeber RL, Deelman-Driessen C, Dijkhuizen L, Eshuis-de Ruiter T, van Leeuwen SS. In depth analysis of the contribution of specific glycoproteins to the overall bovine whey N-linked glycoprofile. *J Agric Food Chem*. 2020:**68**(24):6544–6553.

Vanaman TC, Brew K, Hill RL. The disulfide bonds of bovine α-lactalbumin. *J Biol Chem*. 1970:**245**(17):4583–4590.

Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Marth JD, Bertozzi CR, Hart GW, Etzler ME. Symbol nomenclature for glycan representation. *Proteomics*. 2009:**9**(24):5398–5399.

Yu T, Guo C, Wang J, Hao P, Sui S, Chen X, Zhang R, Wang P, Yu G, Zhang L, et al. Comprehensive characterization of the site-specific N-glycosylation of wild-type and recombinant human lactoferrin expressed in the milk of transgenic cloned cattle. *Glycobiology*. 2010:**21**(2):206–224.

Zhu J, Lin Y-H, Dingess KA, Mank M, Stahl B, Heck AJ. Quantitative longitudinal inventory of the N-glycoproteome of human milk from a single donor reveals the highly variable repertoire and dynamic site-specific changes. *J Proteome Res*. 2020:**19**(5):1941–1952.