O-linked glycosylations in human milk casein and major whey proteins during lactation

Thesbjerg, Martin Nørmark; Poulsen, Katrine Overgaard; Astono, Julie; Poulsen, Nina Aagaard; Larsen, Lotte Bach; Nielsen, Søren Drud-Heydary; Stensballe, Allan; Sundekilde, Ulrik Kræmer

Published in:
International Journal of Biological Macromolecules

DOI (link to publication from Publisher):
10.1016/j.ijbiomac.2024.131613

Publication date:
2024

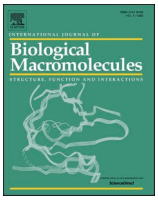
Document Version
Publisher's PDF, also known as Version of record

International Journal of Biological Macromolecules 267 (2024) 131613

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac



O-linked glycosylations in human milk casein and major whey proteins during lactation

Martin Nørmark Thesbjerg [a,b,*], Katrine Overgaard Poulsen [a,b], Julie Astono [a], Nina Aagaard Poulsen [a], Lotte Bach Larsen [a], Søren Drud-Heydary Nielsen [a], Allan Stensballe [c,d], Ulrik Kræmer Sundekilde [a]

[a] Department of Food Science, Aarhus University, Agro Food Park 48, DK-8200 Aarhus N, Denmark
[b] Sino-Danish College (SDC), University of Chinese Academy of Science, Huairou District, Beijing 101408, China
[c] Department of Health Science and Technology, Aalborg University, Selma Lagerløfsvej 249, DK-9260 Gistrup, Denmark
[d] Clinical cancer center, Aalborg University Hospital, 9000 Aalborg, Denmark

ARTICLE INFO

Keywords:
Human milk
O-linked glycosylation
Whey protein
Casein
β-CN

ABSTRACT

As glycosylations are difficult to analyze, their roles and effects are poorly understood. Glycosylations in human milk (HM) differ across lactation. Glycosylations can be involved in antimicrobial activities and may serve as food for beneficial microorganisms. This study aimed to identify and analyze O-linked glycans in HM by high-throughput mass spectrometry. 184 longitudinal HM samples from 66 donors from day 3 and months 1, 2, and 3 postpartum were subjected to a post-translational modification specific enrichment-based strategy using TiO_2 and ZrO_2 beads for O-linked glycopeptide enrichment. β-CN was found to be a major O-linked glycoprotein, additionally, α_{S1}-CN, κ-CN, lactotransferrin, and albumin also contained O-linked glycans. As glycosyltransferases and glycosidases are involved in assembling the glycans including O-linked glycosylations, these were further investigated. Some glycosyltransferases and glycosidases were found to be significantly decreasing through lactation, including two O-linked glycan initiator enzymes (GLNT1 and GLNT2). Despite their decrease, the overall level of O-linked glycans remained stable in HM over lactation. Three different motifs for O-linked glycosylation were enriched in HM proteins: Gly-Xxx-Xxx-Gly-Ser/Thr, Arg-Ser/Thr and LysSer/Thr. Further O-linked glycan motifs on β-CN were observed to differ between intact proteins and endogenous peptides in HM.

1. Introduction

Human milk (HM) has been evolutionarily optimized to support the immune system, brain development, and nutritional needs of newborns and growing infants during the first stages of life. HM is made up of complex components, including proteins, lipids, carbohydrates, vitamins, minerals, and bioactive factors [1]. The protein component of HM comprises casein (α_{S1}-CN, β-CN, and κ-CN) and whey proteins. The main human whey proteins include α-lactalbumin (α-LA), lactotransferrin (TRFL), and albumin (ALB), as well as several immunoglobulins, among which IgAs are notably abundant in milk [2]. Post-translational modifications (PTMs), including protein phosphorylation and glycosylation are crucial determinants of the immunological and bioactive properties of HM proteins [3–5]. The glycoproteome and individual proteoforms of HM have not been studied in detail, and their roles in the immune system and nutritional impact are largely unknown. In HM, glycosylation has been observed to change between lactation periods [6–9]. Further, it plays a role in antimicrobial defense mechanisms [4,5,10], and glycosylated HM components are metabolized by beneficial microorganisms [11–14]. Furthermore, O-linked glycans are known to play a major role in specific proteins, and in milk, κ-CN is a classic example of an O-linked glycoprotein. κ-CN harbors a heavily glycosylated region known as the casein-glycomacropeptide (cGMP), and in HM, the cGMP has up to ten O-linked glycosylated sites [15]. The O-linked glycans in κ-CN make up approximately 40 % of the mass of the mature protein, equating to approximately 13 kDa of the 33 kDa protein [16]. Some of the structures

* Corresponding author at: Agro Food Park 48, DK-8200 Aarhus N, Denmark.
E-mail address: mnt@food.au.dk (M.N. Thesbjerg).

https://doi.org/10.1016/j.ijbiomac.2024.131613
Received 1 December 2023; Received in revised form 10 April 2024; Accepted 13 April 2024
Available online 18 April 2024

of the O-linked glycans in κ-CN have been isolated and identified [17,18]. Additionally, a recent study demonstrated that human β-CN was also glycosylated with O-linked glycans, and the O-linked glycan patterns in the intact β-CN differed from those of naturally occurring β-CN peptides [19].

One of the reasons for our limited understanding of N-linked and O-linked glycosylations and non-enzymatic glycations is that they represent a substantial analytical challenge, as they are highly complex and structurally labile. The large diversity of sugar structures, variation in polarity of the individual glycan units, and poor ionization efficiencies of glycans, collectively make analysis by mass spectrometry (MS) substantially more difficult than that for any other biopolymer [20–23]. In the context of O-linked glycosylation, understanding and annotating specific sites become more challenging due to the absence of a single consistent motif for O-linked glycans; instead, a diverse array of motifs is observed [24].

O-linked glycans are covalently attached to Ser or Thr amino acid (AA) residues by a class of enzymes called polypeptide *N*-acetylgalactosaminyltransferases (GALNTs); this class of enzymes includes GalNAc-Ts and ppGalNTase-T. In humans, this protein family consists of 20–27 different enzymes [25,26]. O-linked glycans are added to proteins as they mature through the endoplasmic reticulum (ER) and golgi apparatus (GA), where the GALNTs deposit the initial *N*-acetylgalactosamine (GalNAc) onto the target Ser or Thr residues; this is followed by the addition of subunits to the initial GalNAc by other enzymes. [27,28]. The final glycans are assembled from an array of subunits other than GalNAc, namely *N*-acetylglucosamine (GlcNAc), fucose (Fuc), sialic acid (NeuAc), galactose (Gal), glucose (Glc), phosphorylation (Phosphor), and sulfatations (Sulf). Thus, the final and observed glycans are influenced by factors, such as the available glycosyltransferases and glycosidases, their capacity to be secreted or retained in the ER and GA, and available substrates, minerals, and cofactors, among others. As tandem MS-based analysis cannot distinguish between Glc and Gal, the term hexose (Hex) is used for both monosaccharides, while the term *N*-acetylhexosamine (HexNAc) is used for both GlcNAc and GalNAc.

Although analysis of O-linked glycans is challenging owing to their lability and different motifs, they exhibit numerous potential biological functions. κ-CN can inhibit the adhesion of *Haemophilus influenzae, Streptococcus pneumonia* [4], and *Helicobacter pylori* [5] via its O-linked glycans. Glycans are believed to function as molecular decoys that prevent bacteria from binding to the intestinal wall of the infant gut. Similarly, whey proteins such as TRFL have been shown to interfere with bacterial growth via their N-linked glycans [10], even though the overall glycosylation of TRFL seems to be decreased by lactation [29]. Finally, O-linked glycans can serve as a carbohydrate source for beneficial bacteria in the infant gut microbiome [12], such as *Bifidobacterium* species, whose subspecies have been observed to have distinct preferences [12,14].

The objective of this study was to explore the natural variability of O-linked glycome in HM, this was done by utilizing 184 HM samples from a longitudinal setup. The samples originated from 66 donors and were selected from a representative biobank of HM with samples from day 3 and months 1, 2, and 3 postpartum. The main goal was to identify the relative abundance of O-linked glycans in HM during the first 90 days of lactation. For this purpose, a comprehensive and thorough analytical strategy was employed, thereby enabling the assessment of thousands of O-linked glycans in the same LC-TandemMS analysis. Furthermore, a detailed analysis of the main proteins in HM (α_{S1}-CN, β-CN, κ-CN α-LA, TRFL, and ALB) and how their O-linked glycans were affected by lactation, donor secretor status, and the body mass index (BMI) of the mother was performed. Ultimately, these findings provide insights into the research gap regarding O-linked glycans in HM, and the methodology used in this study could be beneficial for future MS-based investigations on O-linked glycomes.

2. Materials and methods

2.1. Human milk samples

In total, 184 HM samples from 66 donors were used in this study. HM samples were collected from participants included in the MainHealth cohort [30], registered in clinicaltrials.gov (identification number: NCT05111990). The Central Denmark Region Committee on Health Research Ethics approved this study protocol (J-nr. 1–10–72-296-18). Before inclusion, all participants signed an informed consent form and deputy informed consent form in accordance with the Declaration of Helsinki II. For this study, various meta-data were used for the samples, such as lactation stage, secretor status, and the BMI of the mother. Samples were obtained at four main lactation stages, i.e., day 3 and months 1, 2, and 3, with 32, 56, 55, and 41 donor milk samples obtained for each stage, respectively. Based on a nuclear magnetic resonance (NMR) phenotyping methodology established in a previous study [31], the HM samples were assessed for various human milk oligosaccharides (HMOs), such as 2-fucosyllactose (2'FL), lacto-difucosyllactose (LDFT), and others, which are unique to secretors. Thus, 47 and 19 donors were identified as secretors and non-secretors, respectively. Two BMI groups were established, one for normal weight, with BMI ranging from 18.66 to 24.91, comprising 26 of the 66 donors, and the other was a combined group for overweight and obese participants, with BMI ranging from 25.09 to 46.45, comprising 40 donors. Finally, 30 infants were male and 36 were female. The HM was frozen at −20 °C and stored at each participant's home until collection and transported on dry ice to the department for storage at −80 °C until further use.

2.2. Sample allocation and preparation

A semi-automated sample preparation protocol for the fractionation of the HM endogenic peptidome, proteome, and O-glycome was developed and employed using robotic sample preparation with an Opentrons OT-2 pipetting robot equipped with Heater-Shaker and Magnetic modules (Opentrons, Queens, New York, USA) (a detailed description is provided in Supplementary 1). Briefly, 10 μL of milk from each sample was aliquoted into separate wells of a 96-well plate (AB0800 from Thermo Fisher). Given that HM contains approximately 1 % protein [32,33], the 10 μL milk aliquot contains approximately 100 μg protein.

2.3. Separating the proteome and endogenic peptidome

Prior to separating naturally occurring peptides (referred to as the endogenic peptidome throughout this article) from intact proteins (hereafter referred to as the proteome), the samples were reduced and alkylated using previously described methods, with minor adjustments [34,35] (see Supplementary 1 for details). Briefly, the proteome and endogenic peptidome were separated by adding an equal volume of 20 % trichloroacetic acid (TCA) (A1431, PanReact AppliChem) to the samples. To pellet the intact proteins, the samples were centrifuged for 15 min at 1300 ×*g* and 4 °C. To separate the endogenic peptidome and the proteome, the supernatant was transferred to another 96-well plate and stored at −20 °C until further use.

2.4. Digesting the intact proteins

To prepare the proteome for digestion, any remaining TCA was neutralized by redissolving the dried pellet in 10 μL of 1.0 M triethylammonium bicarbonate buffer (TEAB) (T7408, Sigma Alrich) by shaking at 1500 rpm for 30 s. Seventy microliters of ice-cold acetonitrile (ACN) (45605.01, Serva) was added to each sample. The samples were shaken at 1500 rpm for 2 min before incubation at −20 °C for 1 h, followed by centrifugation for 15 min at 1300 ×*g* and 4 °C. Following this, the supernatant was discarded. The pellet was redissolved in 10 μL of 6.0 M urea (51456, Sigma Aldrich), and the samples were shaken at

1500 rpm for 30 min. Sixty microliters of 50 mM TEAB buffer containing 0.8 μg of trypsin (90057, Thermo Scientific) was added to each sample and mixed, and the samples were incubated overnight at room temperature.

2.5. TiO_2 and ZrO_2 fractionation

To enrich glycosylated peptides, TiO_2 and ZrO_2 beads were used to fragment the sample into enriched fractions, as described previously [36,37]. Briefly, magnetic TiO_2 (TIP002, ResynBio) and ZrO_2 (ZHP002, ResynBio) beads were used with an Opentrons OT-2 (the detailed automation protocol is provided in Supplementary 1 and example code is available in Supplementary 2). Enrichment was performed twice per digest, once using TiO_2 beads and once using ZrO_2 beads. All fractions containing neutral glycopeptides and phosphor- and glycopeptides from TiO_2 and ZrO_2 were merged into the same wells. The fractions containing unmodified peptides were placed in a separate 96-well plate owing to volume constraints. Prior to subsequent processing, the plates were dried using a vacuum centrifuge.

2.6. Preparing the endogenic peptidomes

Before desalting and loading the prepared endogenic peptidomes, the samples were thawed and filtered using a 10 kDa spin filter. Following this, 10 μL of solvent A (water, 0.1 % formic acid, LS1181–1, Thermo Scientific) was added to acidify each sample, and the entire volume was transferred to a 96-well format 10 kDa filter (8034, Pall Corporation). The filter was placed on top of a 96-well plate for collection and centrifuged at 2280 $\times g$ until as much material as feasible passed through the filter.

2.7. Loading and desalting the samples onto evotips

Before proceeding, the samples were redissolved in solvent A. The fractions of non-modified peptides were redissolved in 40 μL of solvent A, while the fractions of neutral glycopeptides and phosphor- and glycopeptides were redissolved in 20 μL of solvent A. To ensure proper redissolving, the samples were shaken at 1500 rpm for 15 min.

To load the samples onto the C18 Evotip (Evotip Pure, Odense, Denmark), an Evosep sample loading protocol was employed with minor modifications to ensure effective sample loading and desalting. A detailed protocol is available in Supplementary 1, and example code for the Opentrons OT-2 pipetting robot is available in Supplementary 2.

2.8. Analysis using Evosep One-timsTOF Pro2

For LC-MS/MS analysis, a setup consisting of a timsTOF PRO2 (Bruker, Bremen, Germany) and an Evosep One (Evosep, Odense, Denmark) was employed, as described previously (Supplementary 1). Briefly, an 8-cm Evosep performance column (EV1109, Evosep) was used with a standard gradient, enabling the analysis of 60 samples per day (60 SPD). The acquisition queue was configured such that after every 10 samples, an Evosep column wash was performed. Every second cycle was followed by a system QC combined with a commercial HeLa tryptic digest sample (200 ng peptides injected, Thermo Fisher) supplemented with iRT (1 uL, Biognosys).

2.9. Data handling

The acquired data were searched using a pre-released version of FragPipe (v. 19.2-build2) [38,39] (Supplementary 1). A glycan database was prepared (available in Supplementary 3) together with a series of glycan ion features (Table S1), allowing FragPipe to identify glycopeptides in the scans with glycan ion features, as previously described [40]. The human reference proteome (UPID ID5640; approximately 20,600 gene products) was retrieved from UniProt.org with the addition of the FragPipe built-in reverse decoys and a list of contaminants (acquired on 06-05-2023). Cysteine carbamidomethylation (C) was used as the fixed modification, and methionine oxidation (M) was used as the variable modification. To ensure accurate identification, peptides and glycans were false discovery rate (FDR)-corrected, as described previously [40,41]. To identify similar precursor ions across the samples, matches between runs were performed as described previously [42].

2.9.1. Relative abundance, data filtering, and grouping

The relative abundances of all individual proteins and peptides were calculated as a fraction of the whole: (intensity of protein or peptide/sum of proteins or peptides in that sample) × 100 [35]. All observed glycopeptides in the proteome and endogenic peptidome were extracted, resulting in 52,071 and 21,050 different O-linked glycopeptides from the proteome and endogenic peptidome, respectively. These glycopeptides were classified in four different ways to explore different parameters; these included classification of O-linked glycopeptides based on the following: 1) glycosylation location, termed the "position" (e.g., all glycans at β-CN-224), 2) the glycan itself, termed "glycan" (e.g., all HexNAc(2)Hex(2) glycans), 3) a combination of the previous two, termed "glycan-position" (e.g., all β-CN-224-HexNAc(2)Hex(2)), and 4) glycoproteins (list of unique proteins from the list of positions). A schematic representation of the position, glycan, and glycan-position is presented in Fig. 1. To ensure sufficient observations for statistical analysis, the proteins and above-described groups were filtered such that they had to be present in at least X of all 184 samples to be included in the final analysis; X was set to 20 and 10 samples for the proteome and endogenic peptidome, respectively. After filtration, the proteome contained 438 O-linked glycoproteins, 492 O-linked positions, 323 O-linked glycans, and 444 O-linked glycan-positions. In the endogenic peptidome, there were 10 O-linked glycoproteins, 17 O-linked positions, 206 O-linked glycans, and 11 O-linked glycan-positions.

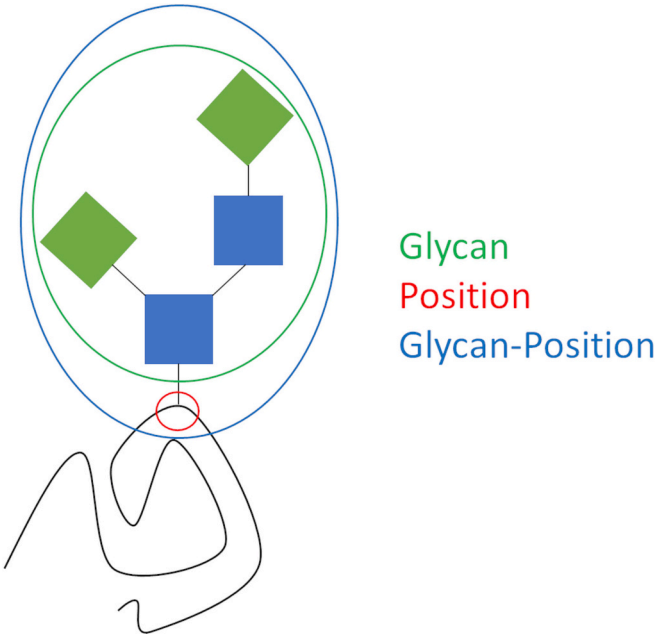


Fig. 1. A schematic representation of a **glycan**, exemplified with the glycan HexNAc(2)Hex(2) glycosylation attached to a **position** (Ser or Thr) in a protein (proteome) or peptide (endogenic peptidome). The combined information of the **glycan** and **position** is referred to as the **glycan-position**. The blue square represents HexNAc, while the green square represents Hex. The shown configuration of the HexNAc and Hex units is only one possibility among several for the HexNAc(2)Hex(2) glycan. The black line at the bottom represents the protein. The different circles highlight different features of the glycosylation. The green circle highlights the **glycan**, the red circle highlights the **position**, and the blue circle highlights the **glycan-position**.

2.10. Clustering of glycosyltransferases and glycosidases

Hierarchical clustering was performed using the pheatmap package in R (ver. 4.2.2). To generate hierarchical clustering, the data were first normalized using z-score normalization. Next, clustering was restricted for grouping samples along a single axis (x-axis), whereas the other (y-axis) was arranged based on the lactation stage (day 3 and months 1, 2, and 3). This allowed the generation of both heat maps and dendrograms.

2.11. Motif analysis

Motif analysis was performed using the rmotifx package [43] in R (version 4.2.2). The analysis encompassed all identified 492 glycan locations from the proteome and 17 glycan locations from the endogenic peptidome; these comprised the foreground datasets. For the background dataset, we extracted sequences centered on Serine (Ser) and Threonine (Thr) from the human proteome used in our data search. The resulting motifs were visualized using the ggseqlogo package in R.

2.12. Statistical analysis

To analyze changes in the relative abundance of proteins and glycan features (position, glycan, and glycan-position), linear mixed models in R (ver. 4.2.2) were employed for the data using the lmerTest package using the following equation:

$$Y_{lSC} = \mu + L_l + S_s + B_b + (LSB)_{lsb} + C_c + e_{lsbc}$$

where Y is the dependent variable, μ is the overall mean of the population, L is the fixed effect of lactation time, S is the fixed effect of secretor status, B is the fixed effect of BMI group, C is the random effect of the donor, and e is the random residual error, which was assumed to be independent and normally distributed with constant variance. The results were corrected for multiple testing by applying FDR correction to the obtained *P*-values at an FDR level of 0.01.

This approach was employed to compare the relative abundances of proteins and glycan features within the proteome and the endogenic peptidome.

3. Results

The HM proteome and endogenic peptidome contained 4076 and 355 unique proteins, respectively; corresponding to approximately 20 % of the human proteome. Of the 4076 proteins in the proteome, 412 were found to have significantly different relative abundances within the first 90 days of lactation. Of the 355 proteins in the endogenic peptidome, 5 were found to have significantly different relative abundances during the lactation period investigated. A list of the proteins that were significantly regulated during lactation is shown in Supplementary 4. In terms of relative abundance, the O-linked glycopeptides accounted for 6.26 ± 1.25 % and 19.99 ± 12.03 % of the relative abundance of the HM proteome and endogenic peptidome, respectively, with non-significant fluctuations during lactation as shown in Table 1. Therefore, the endogenous peptidome had a greater relative abundance of O-linked glycopeptides than the proteome.

After filtering, 438 and 10 proteins were identified as O-linked glycoproteins in the proteome and the endogenic peptidome, respectively. Of the 444 glycan positions in the proteome, 28 were found to be significantly altered during lactation, whereas no glycan positions in the endogenic peptidome were found to be different. Of the 323 and 206 glycans in the proteome and endogenic peptidome, respectively, 3 and 2 were significantly altered during lactation. Furthermore, of the 492 positions in the proteome, 21 were significantly changed during lactation, whereas no positions in the endogenic peptidome were different. The significant features (glycoprotein, glycan-position, glycans, and position) of the proteome and endogenic peptidomes are summarized in Supplementary 5.

A series of annotated O-linked glycopeptide spectra is shown in Fig. 2. The spectra show the C-terminus of β-CN (AA 190–226) without any O-linked glycans (Fig. 2 A), with a HexNAc(1)Hex(1) glycan attached to Thr-217 (Fig. 2 B), and a HexNAc(2)Hex(2) glycan also attached to Thr-217 (Fig. 2 C). For the two scans with O-linked glycosylation, the O-linked glycan-derived ions are highlighted in blue and are among the provided oxonium ions (Table S1). In the case of the larger glycan (Fig. 2 C), the diagnostic ions reflect only parts of the glycan. Therefore, for HexNAc(2)Hex(2), the diagnostic ions at 407.18 *m/z*, 528.18 m/z, and 569.22 m/z indicate the glyco fragments HexNAc(2), HexNAc(1)Hex(2), and HexNAc(2)Hex(1), respectively. When comparing the two glycans (Fig. 2 B and C), these fragments are unique to the HexNAc(2)Hex(2) glycan only (Fig. 2 C), thereby confirming the presence of the identified glycan. Furthermore, all O-linked glycan scans were validated by FDR calculations of individual scans [40]. Individual scans were visualized using the built-in PDV viewer in FragPipe and manually assessed for spectral quality [44].

3.1. The major milk proteins in human milk

The relative abundances of major proteins were monitored throughout lactation (FigS. 1). In the proteome, the average relative abundance of the major proteins, casein (α_{S1}-CN, β-CN, and κ-CN) and whey (TRFL, α-LA, and ALB), was 49.48 ± 1.59 % of the total proteins. The caseins constituted 16.09 ± 0.78 %, while whey proteins made up the remaining 33.39 ± 0.88 % (FigS. 1 A,B). In the proteome, the change in the relative abundance of any of the caseins was not significant during lactation; however, the relative abundance of the main whey proteins was significantly affected by lactation. Specifically, TRFL was downregulated throughout lactation ($P = 5.09E\text{-}9$), while ALB and α-LA were upregulated ($P = 3.58E\text{-}9$ and $P = 1.68E\text{-}5$, respectively). In the endogenic peptidome, the summarized relative abundance of the major proteins was 55.08 ± 15.65 % of the total proteins, of which caseins accounted for 54.40 ± 15.84 %. Furthermore, in the endogenic peptidome, only the relative abundance of β-CN was significantly upregulated throughout lactation ($P = 7.65E\text{-}9$) (FigS. 1 C,D).

In the HM proteome, O-linked glycans were observed at one or more positions in all the main milk proteins, except for α-LA, while β-CN was the major O-linked glycoprotein observed in the endogenic peptidome. As these milk proteins constitute a major portion of the proteins in HM (FigS. 1), we focused on them and their O-linked glycans. β-CN was observed with O-linked glycans in both the proteome and endogenic peptidome (Fig. 3), with nine different O-linked positions and eighteen different glycans attached to these positions. Of these, five sites were observed in the proteome (Ser120, Thr122, Thr207, Thr214, and Ser225) with twelve different glycans attached (Fig. 3 B), whereas in the endogenic peptidome, another five sites (Thr18, Ser21, Ser24, Ser28, and Thr214) and nine different glycans (Fig. 3 A) were identified. Therefore, one site (Thr214) and three glycans (Hex(2), HexNAc(1), and HexNAc(1)Hex(1)) were observed in both the proteome and the endogenic peptidome. In the proteome, four of the glycan-positions on β-CN were observed to be significantly changed throughout lactation. Specifically, HexNAc(1)Hex(1) on Ser120 as well as HexNAc(2)Hex(2)Fuc(1) on Thr207 and Thr214 were significantly upregulated throughout lactation ($P = 0.022$, $P = 0.006$, and $P = 0.012$, respectively), while HexNAc(2)Hex(2)NeuAc(1) on Thr214 was significantly downregulated ($P = 0.020$). In the endogenic peptidome, none of the individual O-linked glycan-positions on β-CN were observed to be significantly affected by lactation; however, the position at Thr18 was significantly upregulated throughout lactation ($P = 0.015$).

For the other caseins (α_{S1}-CN and κ-CN), O-linked glycans were only observed in the proteome (FigS. 2). Specifically, for κ-CN, the two O-linked glycans HexNAc(1)Hex(1) and HexNAc(2)Hex(2) were observed at two positions, namely Ser102 and Thr106 (FigS. 2 B), and no significant regulation was observed throughout lactation. For α_{S1}-CN, the O-

linked glycan Hex(2) was observed at Thr61 without any significant regulation (FigS. 2A).

For the major whey proteins (TRFL, α-LA, and ALB), O-linked glycans were found on TRFL and ALB (FigS. 3). In TRFL, three O-linked glycans, Hex(2), HexNAc(1)Hex(3), and HexNAc(4)Hex(5)NeuAc(1)Fuc(1), were observed at five positions, i.e., Ser122, Thr141, Thr150, Thr158, Ser271, and Thr676 (FigS. 3 B). In ALB, the O-linked glycan Hex(2) was observed at Thr551 (FigS. 3 A). No significant regulation of these O-linked glycans was observed in any of the whey proteins.

3.2. Other O-linked glycoproteins and positions in human milk

In addition to the main milk proteins (α_{S1}-CN, β-CN, κ-CN, TRFL, α-LA, and ALB), a variety of other O-linked glycoproteins were observed, especially within the proteome fraction. The most notable ones included; polymeric immunoglobulin receptor (PIGR), with two O-linked positions at Ser(s) 179 and 414, as well as perilipin 2 (PLIN2), containing an O-linked position at Ser246. Additionally, macrophage mannose receptor 1 (MRC1) was observed with nine different glycans at eight different positions, namely, Thr(s) 633, 640, 641, 1093, 1366, 1371, and 1372, and Ser1215. Finally, immunoglobulin lambda-like polypeptide 5 (IGLL5) was observed with an O-linked position at Thr198. These described positions are novel, as they are not documented on UniProt.org.

3.3. Glycosyltransferases and glycosidases in human milk

The 4076 proteins detected in the HM proteome, were searched for known glycosyltransferases and glycosidases. A total of 13 glycosyltransferases and 18 glycosidases were observed in HM, and the relative abundance of 9 changed significantly during lactation (Fig. 4). To further explore these enzymes, a heatmap with hierarchical clustering was generated (Fig. 4 A). This enabled the clustered groups to be plotted on the x-axis (proteins), while the lactation time (day 3 and months 1, 2, and 3) was plotted on the y-axis (samples). Based on the heatmap, we identified four clusters (as shown in Fig. 4, delineated by red lines). Notably, the first cluster—comprising eight enzymes (four glycosyltransferases and four glycosidases)—exhibited a higher relative abundance on day 3 but became less abundant in subsequent sampling time points. Six of these enzymes, namely, β-1,4-galactosyltransferase (B4GALT1, $P = 1.96E^{-8}$), polypeptide *N*-acetylgalactosaminyltransferase 2 (GALNT2, $P = 0.007$), mannosyl-oligosaccharide 1,2-α-mannosidase IA (MAN1A1, $P = 9.48E^{-8}$), multifunctional procollagen lysine hydroxylase and glycosyltransferase LH3 (PLOD3, $P = 2.68E^{-16}$), lysosomal acid glucosylceranmidase (GBA1, $P = 4.67E^{-11}$), and polypeptide *N*-acetylgalactosaminyltransferase 1 (GALNT1, $P = 8.20E^{-8}$) were significantly downregulated throughout lactation. In the second cluster, no general pattern could be observed, as some proteins were upregulated while others were downregulated, and only one protein (tissue α-L-fucosidase (FUCA1)) was significantly upregulated throughout lactation in this cluster ($P = 0.003$). In the third cluster, the general pattern indicated a slight increase in abundance over time; however, not all proteins followed this trend, and only one of the proteins (Lysozyme C (LYZ)) was significantly elevated during lactation ($P = 6.53E^{-15}$). In the fourth and last cluster, the overall tendency was less apparent; however, the overall trend seemed to indicate an increase in protein from day 3 to month 1, followed by a decline over the next two months. In this cluster, only 4-galactosyl-*N*-acetylglucosaminide 3-α-L-fucosyltransferase (FUT6) was significantly regulated ($P = 0.010$); this protein followed a pattern of higher abundance at day 3 and month 1, followed by decreased levels at months 2 and 3.

3.4. O-linked glycosylation motif analysis

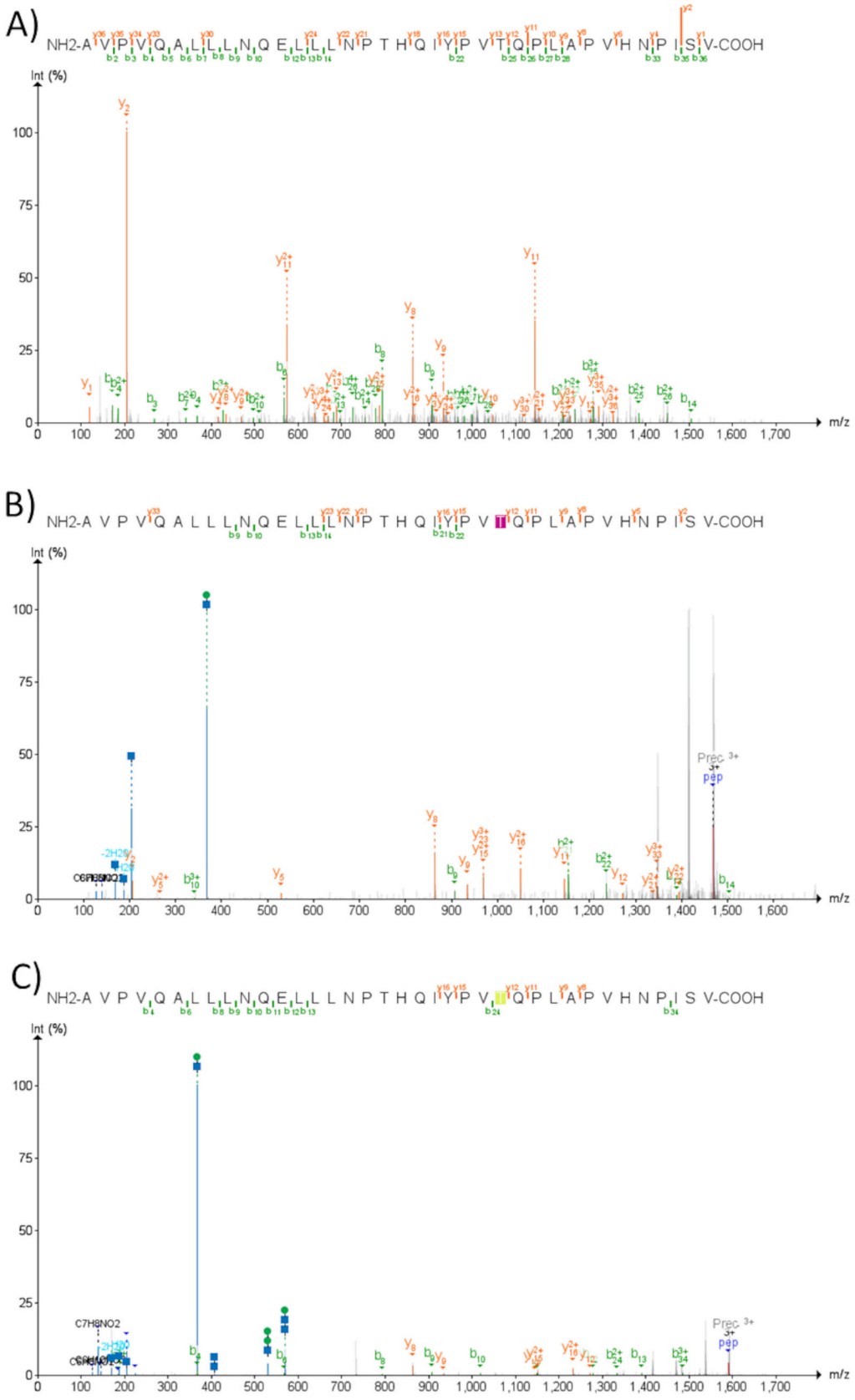
All 492 and 17 O-linked positions of the proteome and endogenic peptidome, respectively, were included in the motif analysis, although the proteome and endogenic peptidome were searched separately. Analysis revealed three prevalent motifs in the proteome. In the endogenic peptidome, no prevalent motif was identified, probably because there were insufficient observations. The most prevalent motifs in the proteome were Gly-Xxx-Xxx-Gly-Ser/Thr, with a 7.65-fold enrichment over the background database. This was followed by the Arg-Ser/Thr motif (enriched 2.43-fold) and Lys-Ser/Thr motif (enriched 2.28-fold). Centralized data for all sequences are shown, along with the probability of encountering a specific AA at a given position (Fig. 5). Apart from the motifs, the analysis revealed that polar and small AAs, such as Ser, Thr, and Gly, are highly likely along the entire sequence of O-linked glycans, with the exception of the third AA toward the N-terminal, where a Glu residue is predominant. Furthermore, basic AAs such as Arg and Lys appeared to be common on the first to third AA toward the N-terminal; similarly, Pro appeared to be very probable at the site (two AAs to either side of this position, especially toward the N-terminal).

4. Discussion

In the present study, intact O-linked glycans in HM were identified and characterized. O-linked glycans are described as either the glycan itself (reported as glycans), its position (reported as position), or a combination of a glycan and position (reported as glycan-position).

Given that the relative abundance of some of the glycosyltransferases and glycosidases (Fig. 4) significantly changed throughout lactation while the overall level of O-linked glycation remained unchanged (Table 1), it is suggested that at least some of the implicated enzymes exhibit altered activity during lactation to compensate for the altered abundance while still producing the same product. The overall carbohydrate content of HM is relatively stable during the first year of lactation [33]. Furthermore, lactose is synthesized in milk [45,46], indicating that carbohydrates are acted on by enzymes within milk. In the present study, both B4GALT1 and α-LA were observed; however, only the relative abundance of B4GALT significantly decreased with lactation (section. 3.3). Taken together, these observations indicate that the activity of glycosyltransferases and glycosidases changes throughout lactation to maintain the observed consistent level of O-linked glycans, despite the observed decline in the relative amount and number of enzymes available. Similar adaptations have been observed and speculated for other HM enzymes, such as plasmin [35].

Among the major proteins in HM, β-CN has previously been reported with O-linked glycans by Dinges et al. (2021) [19]. Dinges et al. assessed HM from two donors throughout the first 16 weeks of lactation and measured the phosphorylation and O-linked glycosylation levels of β-CN in both the proteome and endogenic peptidome [19]. In summary, they observed nine glycans at 18 positions. Further differences between the proteome and endogenic peptidome of the two donors were observed, with a relatively stable degree of glycosylation at the individual O-linked positions. Interestingly, the O-linked positions they observed are very similar to those observed in this present study. Specifically, the sites shared between Dinges et al. [19] and the present study were Thr18, Ser21, Ser24, Ser28, Ser120, Thr122, Thr207, Thr214, and Ser225. Additionally, Dinges et al. (2021) reported nine additional sites. Upon closer inspection of our own data (not shown), many of these sites were identified. However, owing to the filters applied in the present study, specifically the statistical analysis criteria mandating that the glycan-position needs to be observed in 20 or 10 samples in the proteome or endogenic peptidome, respectively, most of these sites were excluded as they had insufficient consistent observations. Moreover, the specific glycans reported by Dinges et al. (2021) [19] and the present study do not exactly match. This could simply indicate that the HM is more complex; alternatively, it could indicate methodological differences or a combination of both. Further studies are required to elucidate these differences. In summary, the two studies share similar findings, indicating that human β-CN is indeed glycosylated with O-linked glycans and that some of these are consistent, while several others appear to be



(caption on next page)

Fig. 2. Three scans of an O-linked glycopeptide. In every instance, the peptide is the same and located in the C-terminal of β-CN (amino acids 190–226 of β-CN). The b- and y-ions identified by PDV are marked in green and orange, respectively. Panel A) shows the peptide without any glycosylation; panels B) and C) show the peptide with an O-linked glycan at Thr-214, with the diagnostic ions shown in blue with a miniature glycan above them (green circle = Hex; blue square = HexNAc). For panel B), a HexNAc(1)Hex(1) glycan is identified at Thr-214. For panel C), a HexNAc(2)Hex(2) glycan is identified at Thr-214. For panels B) and C), the diagnostic ions provide evidence for their individual glycans since there are no ions indicating larger glycans, and the identified diagnostic ions represent fractions of the parent glycan. In all cases, owing to the size of the peptide, many of the ions shown have multiple charges, and the precursor has three charges.

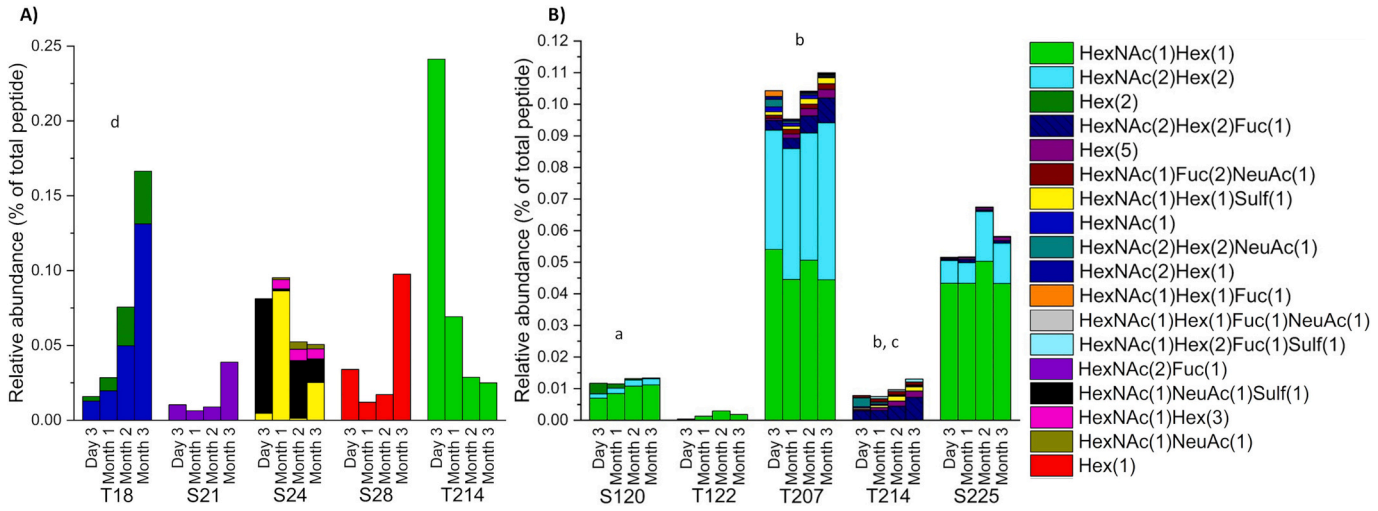


Fig. 3. Relative abundance of the 18 different O-linked glycans on human β-CN in the endogenic peptidome (panel A) and proteome (right B) throughout lactation. The y-axis indicates the relative abundance as a % of the total amount of peptides. The x-axis represents the different positions of the O-linked glycan within β-CN. For each position, four stacked bars represent the different lactation stages (day 3 and months 1, 2, and 3), while the section of each of the stacked bars represents the average of each of the individual 18 glycans. The colors of the different sections indicate different O-linked glycans, as explained by the legend on the right. Lowercase letters above the bars indicate significant changes at that position throughout lactation: a = significant up-regulation of HexNAc(1)Hex(1) on Ser120, b = significant up-regulation of HexNAc(2)Hex(2)Fuc(1) on Thr(s) 207 and 214, c = significant down-regulation of HexNAc(2)Hex(2)NeuAc(1) on Thr214, and d = significant up-regulation in the overall use of Thr18 as an O-linked position.

Table 1
An overview of the summarized relative abundance of the O-linked glycopeptides in the proteome and peptidome of human milk throughout lactation with an average value and the calculated level of significance across lactation.

	Day 3	Month 1	Month 2	Month 3	Average	*P*-value
Proteome	6.12 ± 0.98	6.12 ± 1.35	6.29 ± 1.21	6.51 ± 1.33	6.26 ± 1.25	0.84
Peptidome	20.33 ± 12.23	24.35 ± 11.87	16.63 ± 11.77	18.25 ± 10.52	19.99 ± 12.03	0.18

donor-dependent.

Few studies have reported how individual glycans in HM inhibit the growth of various microorganisms [4,5,10]. Aniansson et al. (1990) [4] identified GlcNAcβ1-3Gal as the glycan exhibiting an inhibitory effect on *Streptococcus pneumoniae* and demonstrated that *S. pneumoniae* was bound to the casein portion of HM and not the whey portion. Strömqvist et al. (1995) [5] revealed that fucosylated glycans from human κ-CN are unique to HM and that they play a key role in the inhibition of *Helicobacter pylori*. Lastly, Barboza et al. (2012) [10] demonstrated that removing the terminal fucose units from TRFL increased the overall ability of bacteria to adhere to and invade host cells. As demonstrated by Strömqvist et al. (1995) [5] for HM κ-CN, the high fucose and sialic acid content in N-linked HM TRFL glycans has recently been recognized as a distinctive feature of HM TRFL [47]. In the present study, fucosylated glycans were observed to constitute a part of the O-linked glycans in the main HM proteins, namely β-CN and TRFL. The abundance of one of the structures, HexNAc(2)Hex(2)Fuc(1), was significantly increased on β-CN with lactation, and β-CN was observed with a range of other fucosylated glycans (Fig. 3). Therefore, in the present study, fucosylated O-linked glycans were observed on β-CN and TRFL, which validates the previous observations by Strömqvist et al. (1995) [5], Barboza et al. (2012) [10], and Zhang et al. (2024) [47], all of which describe the uniqueness and importance of fucose in HM glycans.

The three glycan antigens sialyl T, asialyl T, and Tn (translated to the following glycan formulas in the present study: HexNAc(1)Hex(1)NeuAc (1), HexNAc(1)Hex(1), and HexNAc(1), respectively) in HM were measured using a lectin-based ELISA [48] and found to play a role in interfering with bacterial adhesion. Although we were unable to distinguish between individual glycan structures in the present study, the overall O-linked glycan levels were stable; notably, certain previously described glycan features were observed in the present study. For instance, HexNAc(1)Hex(1), or possibly asialyl T, were among the major O-linked glycans in the main proteins (Fig. 3 and FigS. 2). In addition to acting as functional units, some glycans serve as food for beneficial bacteria in the infant gut microbiome [11,12]. This could be important for ensuring proper colonization of the infant gut (similar to the role of HMOs [49]). Thus, it is remarkable that some O-linked glycans seem to have a dual function in preventing bacterial invasion by pathogens [4,5,10], while also facilitating the colonization of the intestine by beneficial bacteria [11,12].

Analysis and identification of glycans, particularly O-linked glycans, via MS is challenging because the sugar-protein linkage is extremely labile [22] without adhering to any single consensus sequence or motif [24]. Furthermore, collision-induced dissociation (CID) favors glycan fragmentation, usually without simultaneous peptide fragmentation [50,51]. As demonstrated in the present study, a partial solution was to engage “ms

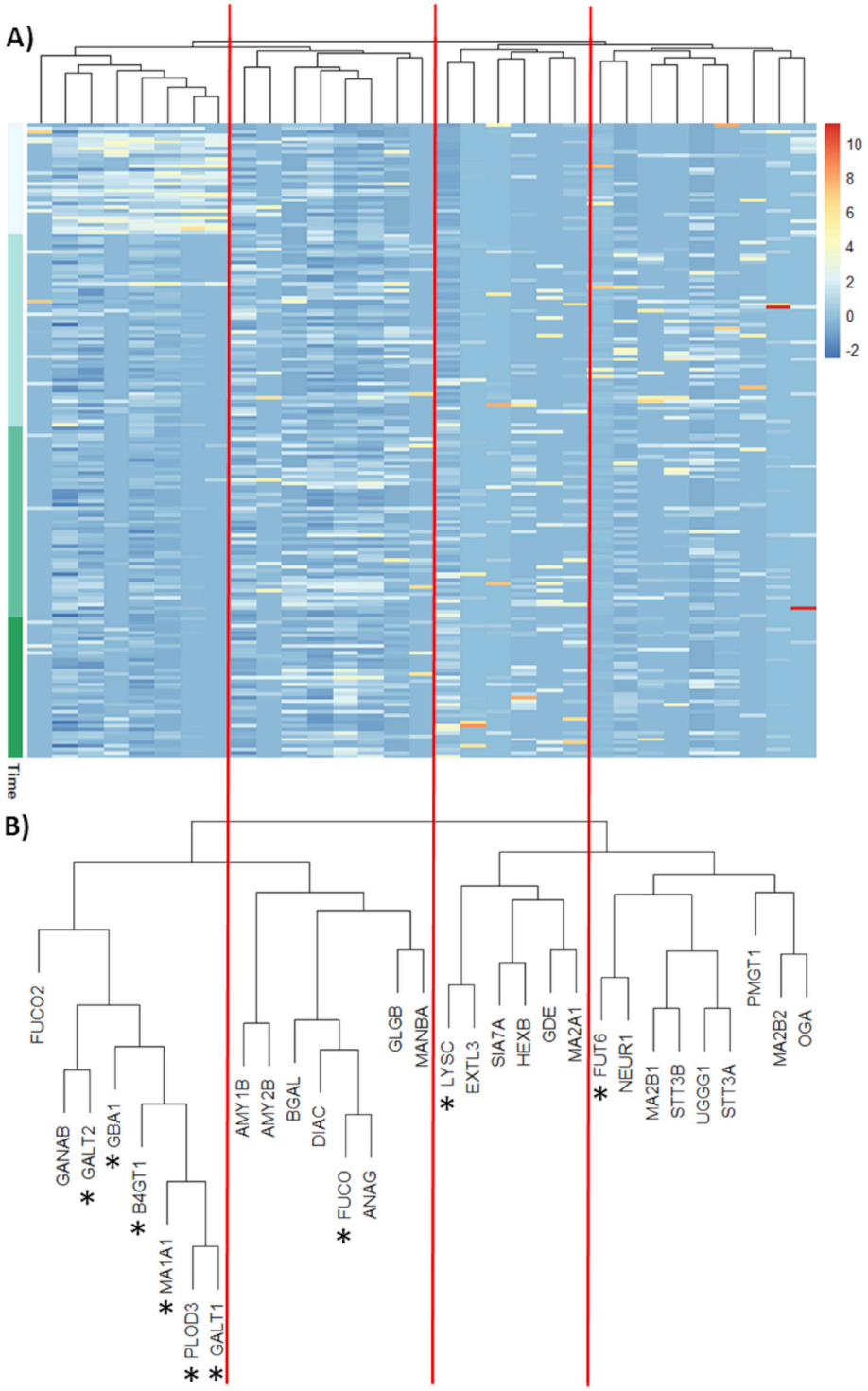


Fig. 4. Heatmap (panel A) and dendrogram (panel B) of the 31 glycosyltransferases and glycosidases observed in human milk. The proteins were clustered (x-axis), while the y-axis was arranged according to the lactation stage (day 3 = 0 (white) and months 1 = 1 (light green), 2 = 2 (green), and 3 = 3 (dark green)). The red lines indicate the division between the four groups (groups 1, 2, 3, and 4 from left to right) generated by the clustering. The colour of the heatmap ranges from dark blue (low) to white (average) to red (high). For the dendrogram, the gene names are used to represent the individual proteins. * = Significantly changed throughout lactation.

stepping," as demonstrated by Mukherjee et al. (2023) [52]. Other chemical methods have proven useful for identifying O-linked glycan motifs, namely, BEMAP [53] or chemical modification of the O-linked structure [54], which was used in *E. coli* to identify O-linked glycan motifs. However, both of these processes destroy the O-linked glycans and replace them with features, such as phosphorylation in the case of BEMAP; this process could be modified to add a stable side chain instead of a phosphorylation [54]. To the best of our knowledge, this is the first study to generate a motif for O-linked glycans in HM using intact glycopeptides analyzed by CID induced MS. Three different motifs were found to be dominant: Gly-Xxx-Xxx-Gly-Ser/Thr, Arg-Ser/Thr, and Lys-Ser/Thr. The last two motifs, containing Arg and Lys, contradicted the findings of

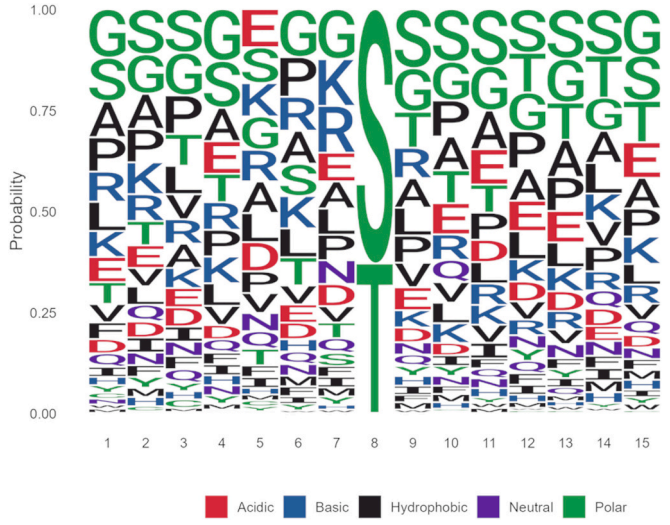


Fig. 5. Motifs of the attachment sites from O-linked glycans in the proteome, categorized by the number of observations. The y-axis represents the probability of a position being a specific amino acid. The x-axis shows the individual positions from 1 to 15, with the center position at eight representing the position where the O-linked glycan is attached. Additionally, the colour of the individual amino acids indicates their chemical properties. The three major motifs identified are Gly-Xxx-Xxx-Gly-Ser/Thr, Arg-Ser/Thr, and Lys-Ser/Thr, in that order.

Wilson et al. (1991) [55], who observed that Arg and Lys were less frequent around the glycosylation site, whereas they were more prevalent in the present study. These differences may be attributed to the different sample types and, consequently, different glycosyltransferase/glycosidase profiles. As shown in Fig. 5, basic AAs were relatively frequent at the three AAs on the N-terminal side of the glycosylation site, even though they were only significantly enriched at the first AA on the N-terminal side of the glycosylation site. Interestingly, none of the motifs seem to match the sites identified on β-CN, which was identified as a major O-linked glycoprotein in HM, featuring nine O-linked positions (Thr18, Ser21, Ser24, Ser28, Ser120, Thr122, Thr207, Thr214, and Ser225). Based on the AA sequences of β-CN: $^{16-}$RETIESLSSSEESITE^{-31}, $^{116-}$PVLKSPTIPFF^{-126}, and $^{205-}$NPTHQIYPVTQPLAPVHNPISV^{-226}, it can be seen that the sites of the proteome (Thr18, Ser21, Ser24, Ser28, and Thr214) and endogenic peptidome (Ser120, Thr122, Thr207, Thr214, and Ser225) follow different patterns. According to literature, some known O-linked position motifs are Xxx-Pro-Xxx-Xxx (where one Xxx is a glycosylated Thr), Ser/Thr-Xxx-Xxx-Xxx (where the first Ser/Thr is glycosylated and one Xxx is another Ser/Thr), and Xxx-Xxx-Thr-Xxx (where one Xxx is Lys or Arg) [55,56]. The O-linked glycans at Thr122, Thr207, Thr214, and Ser225 are in close proximity to a Pro residue, thereby matching the Xxx-Pro-Xxx-Xxx motif reported by Pisano et al. (1993) [56] and Wilson et al. (1991) [55]. The Thr18 and Ser120 residues are close to a basic AA, as indicated by two of the motifs identified by us (Fig. 5). The last sites at Ser21, Ser24, and Ser28 did not follow any of the previoly described motifs; however, Ser21 and Ser24 followed the Ser/Thr-Xxx-Xxx-Xxx motif described by Pisano et al. (1993) [56]. Thus, only Ser28 follows an unknown motif. Furthermore, Ser28 has only one type of glycan, namely, Hex(1). Ser21, Ser24, and Ser28 were all close to other polar AAs, primarily Ser and Thr, and these positions are unique to the endogenic peptidome.

The method employed in the present study is not foolproof as it has some limitations. The major limitation is that FragPipe only allows for the presence of one glycan per peptide; combining this with the use of trypsin for the generation of peptides, as in the present study, could produce peptides with multiple O-linked glycans. This would thus affect the detection of O-linked glycans on κ-CN, which is heavily glycosylated, especially in the cGMP fragment, as indicated by earlier Edman sequencing and PAS staining studies [15,16]. Because there are only a few trypsin cleavage sites in cGMP, it generates peptides with multiple attached glycans, thus rendering them unidentifiable using FragPipe. One solution could be to employ other proteases, such as elastase, chymosin, pepsin, Glu-C, and mucinase, that can potentially cleave between the glycan positions. Exploring these enzymes and their ability to handle cGMP and the proteome requires additional studies. Despite these limitations, we successfully identified intact O-linked glycans in HM using bottom-up strategies, including two novel positions on κ-CN at Ser 102 and T106, which are not part of the cGMP, starting at AA 118, according to UniProt.org.

5. Conclusion

In the present study, O-linked glycans were identified as intact glycopeptides in HM using an automated TiO_2 and ZrO_2 enrichment procedure, followed by LC-TandemMS characterization. A total of 552 fractions of 184 samples from 66 donors at different lactation stages (day 3 and months 1, 2, and 3) were used in the present study. In the proteome, 6.26 ± 1.25 % of the relative abundance was found to be O-linked glycopeptides, corresponding to 323 glycans, 492 positions, and 444 glycan-positions included in the present study. For the endogenic peptidome, 19.99 ± 12.03 % of the relative abundance was found to be accounted for by O-linked glycopeptides, corresponding to 206 glycans, 17 positions, and 11 glycan-positions included in the study. β-CN, a major O-linked glycoprotein in HM, was observed in both the proteome and endogenic peptidome. In total, β-CN was observed with 18 different glycans at 9 positions, and among these, 1 position was shared between the proteome and endogenic peptidome. The nine sites on β-CN from the present study were also observed by Dinges et al. [19]. Moreover, in the present study, HexNAc(2)Hex(2)Fuc(1) at Thr207 and Thr214 of β-CN were found to be significantly increased throughout lactation, and this could be speculated to be relevant for the defense against bacterial invasions of the infant gastrointestinal tract.

31 glycosyltransferases and glycosidases were observed in HM, nine of which were significantly regulated during lactation; however, the overall O-linked glycosylation levels remained stable throughout lactation. Three enriched motifs of O-linked glycans were identified in the HM: Gly-Xxx-Xxx-Gly-Ser/Thr, Arg-Ser/Thr, and Lys-Ser/Thr. Furthermore, matches with the known Xxx-Pro-Xxx-Xxx and Ser-Xxx-Xxx-Xxx motifs were identified in HM. On β-CN, the Ser-Xxx-Xxx-Xxx motif was unique to the endogenic peptidome, as the O-linked positions of β-CN in the endogenic peptidome were typically near other Ser residues.

Further research into the functional aspects and consequences of O-linked glycosylation in HM is needed to comprehend their functionalities and, thus, their importance. Furthermore, there is a need to find a solution to the limitations of the current study, which could involve selective digestion of the proteins as well as the ability to identify multiple glycans within the same peptides.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2024.131613.

CRediT authorship contribution statement

Martin Nørmark Thesbjerg: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Katrine Overgaard Poulsen:** Writing – review & editing, Resources. **Julie Astono:** Writing – review & editing, Resources. **Nina Aagaard Poulsen:** Writing – review & editing, Supervision. **Lotte Bach Larsen:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Søren Drud-Heydary Nielsen:** Writing – review & editing, Supervision, Funding acquisition. **Allan Stensballe:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. **Ulrik Kræmer Sundekilde:** Writing – review & editing, Writing – original draft, Supervision, Resources, Funding

acquisition, Conceptualization.

Declaration of competing interest

S.D-H.N. and K.O.P. currently hold positions in Arla Foods Ingredients. The remaining authors declare no conflicts of interest. The funders had no role in the design of the study; the collection, analysis, or interpretation of data; the writing of the manuscript; or the decision to publish the results.

Acknowledgements

The authors would like to thank Daniel A. Polasky from the Nesvhzhskii Laboratory, University of Michigan, for assistance with the proper configuration of FragPipe and its many modules, as well as interpretation of the generated output.

Funding

This study was funded by the Sino-Danish Centre for Education and Research (SDC). This work was partly funded by Arla Food for Health (Viby J, Denmark). The Danish National Mass Spectrometry Platform for Functional Proteomics (PRO-MS; grant no. 5072-00007B), Svend Andersen Foundation, and SparNord Foundation are acknowledged for grants to the analytical platform, which enabled parts of this study.

Informed consent statement

Informed consent was obtained from all subjects involved in the study. The Central Denmark Region Committee on Health Research Ethics approved this study protocol (J-nr. 1-10-72-296-18). Before inclusion, all participants signed an informed consent form and deputy informed consent in accordance with the Declaration of Helsinki II.

References

[1] R.G. Jensen, Introduction, in: R.G.B.T.-H. of M.C. Jensen (Ed.), Handb. Milk Compos, Elsevier, San Diego, 1995, pp. 1–3, https://doi.org/10.1016/B978-012384430-9/50003-2.

[2] S.M. Donovan, Human Milk proteins: composition and physiological significance, in: S.M. Donovan, J.B. German, B. Lönnerdal, A. Lucas (Eds.), Hum. Milk Compos. Clin. Benefits Futur. Oppor. 90th Nestlé Nutr. Inst. Work. Lausanne, October–November 2017, S.Karger AG, Basel, 2019, pp. 93–101, https://doi.org/10.1159/000490298.

[3] T.A. Kroening, J.H. Baxter, S.A. Anderson, R.G. Hards, L. Harvey, P. Mukerji, Concentrations and anti-Haemophilus influenzae activities of beta-casein phosphoforms in human milk, J. Pediatr. Gastroenterol. Nutr. 28 (1999) 486–491, https://doi.org/10.1097/00005176-199905000-00008.

[4] G. Aniansson, B. Andersson, R. Lindstedt, C. Svanborg, Anti-adhesive activity of human casein against Streptococcus pneumoniae and Haemophilus influenzae, Microb. Pathog. 8 (1990) 315–323, https://doi.org/10.1016/0882-4010(90)90090-D.

[5] M. Strömqvist, P. Falk, S. Bergström, L. Hansson, B. Lönnerdal, S. Normark, O. Hernell, S.B. Lennart Hansson, B. Lönnerdal, S. Normark, O. Hernell, Human milk kappa-casein and inhibition of helicobacter pylori adhesion to human gastric mucosa, J. Pediatr. Gastroenterol. Nutr. 21 (1995) 288–296, https://doi.org/10.1097/00005176-199510000-00006.

[6] M. Elwakiel, S. Boeren, J.A. Hageman, I.M. Szeto, H.A. Schols, K.A. Hettinga, Variability of serum proteins in Chinese and Dutch human milk during lactation, Nutrients 11 (2019) 499, https://doi.org/10.3390/nu11030499.

[7] M. Elwakiel, E.J. Bakx, I.M. Szeto, Y. Li, K.A. Hettinga, H.A. Schols, Serum protein N-Glycans in colostrum and mature milk of Chinese mothers, J. Agric. Food Chem. 68 (2020) 6873–6883, https://doi.org/10.1021/acs.jafc.0c02161.

[8] Z. Wang, N. Zhang, W. Wang, Y. Li, I.M. Szeto, H. Qin, Y. Jin, M. Ye, Glycoproteomics analysis reveals differential expression of site-specific glycosylation in human Milk whey during lactation, J. Agric. Food Chem. 69 (2021) 6690–6700, https://doi.org/10.1021/acs.jafc.0c07998.

[9] J. Zhu, Y.-H. Lin, K.A. Dingess, M. Mank, B. Stahl, A.J.R. Heck, 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. 19 (2020) 1941–1952, https://doi.org/10.1021/acs.jproteome.9b00753.

[10] M. Barboza, J. Pinzon, S. Wickramasinghe, J.W. Froehlich, I. Moeller, J. T. Smilowitz, L.R. Ruhaak, J. Huang, B. Lönnerdal, J.B. German, J.F. Medrano, B. C. Weimer, C.B. Lebrilla, Glycosylation of human milk lactoferrin exhibits dynamic changes during early lactation enhancing its role in pathogenic bacteria-host interactions, Mol. Cell. Proteomics 11 (2012), https://doi.org/10.1074/mcp.M111.015248. M111.015248.

[11] N. Kirmiz, R.C. Robinson, I.M. Shah, D. Barile, D.A. Mills, Milk Glycans and their interaction with the infant-gut microbiota, Annu. Rev. Food Sci. Technol. 9 (2018) 429–450, https://doi.org/10.1146/annurev-food-030216-030207.

[12] K.J. González-Morelo, M. Vega-Sagardía, D. Garrido, Molecular insights into O-linked glycan utilization by gut microbes, Front. Microbiol. 11 (2020), https://doi.org/10.3389/fmicb.2020.591568.

[13] D. Garrido, C. Nwosu, S. Ruiz-Moyano, D. Aldredge, J.B. German, C.B. Lebrilla, D. A. Mills, Endo-β-N-acetylglucosaminidases from infant gut-associated bifidobacteria release complex N-glycans from human milk glycoproteins, Mol. Cell. Proteomics 11 (2012) 775–785, https://doi.org/10.1074/mcp.M112.018119.

[14] K. Sercan, L.P. Annabelle, L.N. de M.B.J. Maria, F.S. A., K. Nina, B.D. E., B. Daniela, M.D. A., Oligosaccharides Released from Milk Glycoproteins Are Selective Growth Substrates for Infant-Associated Bifidobacteria, Appl. Environ. Microbiol. 82 (2016) 3622–3630. doi:10. 1128/AEM.00547–16.

[15] A.-M. Fiat, J. Jolles, J.-P. Aubert, M.-H. Loucheux-Lefbvre, P. Jolles, Localisation and importance of the sugar part of human casein, Eur. J. Biochem. 111 (1980) 333–339, https://doi.org/10.1111/j.1432-1033.1980.tb04946.x.

[16] K. Yamauchi, N. Azuma, H. Kobayashi, S. Kaminogawa, Isolation and properties of human k-casein, J. Biochem. 90 (1981) 1005–1012, https://doi.org/10.1093/oxfordjournals.jbchem.a133552.

[17] T. Saito, T. Itoh, S. Adachi, Chemical structure of neutral sugar chains isolated from human mature milk κ-casein, Biochim. Biophys. Acta 964 (1988) 213–220, https://doi.org/10.1016/0304-4165(88)90169-9.

[18] H. van Halbeek, J.F.G. Vliegenthart, A.-M. Fiat, P. Jollès, Isolation and structural characteri- zation of the smaller-size oligosaccharides from desialylated human κ-casein, FEBS Lett. 187 (1985) 81–88, https://doi.org/10.1016/0014-5793(85)81219-9.

[19] K.A. Dingess, I. Gazi, H.W.P. van den Toorn, M. Mank, B. Stahl, K.R. Reiding, A.J. R. Heck, Monitoring human milk β-casein phosphorylation and O-glycosylation over lactation reveals distinct differences between the proteome and endogenous peptidome, Int. J. Mol. Sci. 22 (2021) 8140, https://doi.org/10.3390/ijms22158140.

[20] D.A. Polasky, D.J. Geiszler, F. Yu, K. Li, G.C. Teo, A.I. Nesvizhskii, MSFragger-labile: a flexible method to improve labile PTM analysis in proteomics, Mol. Cell. Proteomics 22 (2023) 100538, https://doi.org/10.1016/j.mcpro.2023.100538.

[21] M. Sanda, J. Benicky, R. Goldman, Low collision energy fragmentation in structure-specific Glycoproteomics analysis, Anal. Chem. 92 (2020) 8262–8267, https://doi.org/10.1021/acs.analchem. 0c00519.

[22] S.J. North, P.G. Hitchen, S.M. Haslam, A. Dell, Mass spectrometry in the analysis of N-linked and O-linked glycans, Curr. Opin. Struct. Biol. 19 (2009) 498–506, https://doi.org/10.1016/j.sbi.2009.05. 005.

[23] L.R. Ruhaak, G. Xu, Q. Li, E. Goonatilleke, C.B. Lebrilla, Mass spectrometry approaches to Glycomic and Glycoproteomic analyses, Chem. Rev. 118 (2018) 7886–7930, https://doi.org/10.1021/acs.chemrev.7b00732.

[24] Z. Zhu, X. Su, D.F. Clark, E.P. Go, H. Desaire, Characterizing O-linked Glycopeptides by Electron transfer dissociation: fragmentation rules and applications in data analysis, Anal. Chem. 85 (2013) 8403–8411, https://doi.org/10.1021/ac401814h.

[25] J. Raman, Y. Guan, C.L. Perrine, T.A. Gerken, L.A. Tabak, UDP-N-acetyl-α-D-galactosamine: polypeptide N-acetylgalactosaminyltransferases: completion of the family tree, Glycobiology 22 (2012) 768–777, https://doi.org/10.1093/glycob/cwr183.

[26] K.G. Ten Hagen, T.A. Fritz, L.A. Tabak, All in the family: the UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferases, Glycobiology 13 (2003) 1R–16R, https://doi.org/10.1093/glycob/ cwg007.

[27] A. Hassinen, A. Rivinoja, A. Kauppila, S. Kellokumpu, Golgi N-glycosyltransferases form both Homo- and heterodimeric enzyme complexes in live cells, J. Biol. Chem. 285 (2010) 17771–17777, https://doi.org/10.1074/jbc.M110.103184.

[28] M. Takeuchi, M. Yoshikawa, R. Saski, H. Chiba, Purification and characterization of UDP-N-Acetylgalactosamine: κ-casein polypeptide N-Acetylgalactosaminyltransferase from mammary gland of lactating cow, Agric. Biol. Chem. 49 (1985) 1059–1069, https://doi.org/10.1271/bbb1961.49.1059.

[29] J.W. Froehlich, E.D. Dodds, M. Barboza, E.L. McJimpsey, R.R. Seipert, J. Francis, H.J. An, S. Freeman, J.B. German, C.B. Lebrilla, Glycoprotein expression in human Milk during lactation, J. Agric. Food Chem. 58 (2010) 6440–6448, https://doi.org/10.1021/jf100112x.

[30] K.O. Poulsen, J. Astono, R.R. Jakobsen, N. Uldbjerg, J. Fuglsang, D.S. Nielsen, U. K. Sundekilde, Influence of maternal body mass index on human milk composition and associations to infant metabolism and gut colonisation: MAINHEALTH – a study protocol for an observational birth cohort, BMJ Open 12 (2022) e059552, https://doi.org/10.1136/bmjopen-2021-059552.

[31] G. Praticò, G. Capuani, A. Tomassini, M.E. Baldassarre, M. Delfini, A. Miccheli, Exploring human breast milk composition by NMR-based metabolomics, Nat. Prod. Res. 28 (2014) 95–101, https://doi.org/10.1080/14786419.2013.843180.

[32] Y. Liao, D. Weber, W. Xu, B.P. Durbin-Johnson, B.S. Phinney, B. Lönnerdal, Absolute quantification of human Milk caseins and the whey/casein ratio during the first year of lactation, J. Proteome Res. 16 (2017) 4113–4121, https://doi.org/10.1021/acs.jproteome.7b00486.

[33] K.F. Michaelsen, L. Skafte, J.H. Badsberg, M. Jørgensen, Variation in Macronutrients in Human Bank Milk, J. Pediatr. Gastroenterol. Nutr. 11 (1990) 229–239, https://doi.org/10.1097/00005176-199008000-00013.

[34] D.C. Dallas, S.D.-H. Nielsen, Milk Peptidomics to identify functional peptides and for quality control of dairy products, in: M. Schrader, L. Fricker (Eds.), Methods

Mol. Biol, Springer, New York, New York, NY, 2018, pp. 223–240, https://doi.org/10.1007/978-1-4939-7537-2_15.
[35] M.N. Thesbjerg, S.D.-H. Nielsen, U.K. Sundekilde, N.A. Poulsen, L.B. Larsen, Fingerprinting of proteases, protease inhibitors and indigenous peptides in human milk, Nutrients 15 (2023) 4169, https://doi.org/10.3390/nu15194169.
[36] M.N. Melo-Braga, M. Ibáñez-Vea, K. Kulej, M.R. Larsen, Comprehensive protocol to Simulta- neously study protein phosphorylation, acetylation, and N-linked Sialylated glycosylation, in: A. Posch (Ed.), Proteomic Profiling Methods Protoc, 1295th ed., Methods in molecular biology, New York, NY, 2021, pp. 55–72, https://doi.org/10.1007/978-1-0716-1186-9_5.
[37] C. Koenig, A. Martinez-Val, G. Franciosa, J.V. Olsen, Optimal analytical strategies for sensitive and quantitative phosphoproteomics using TMT-based multiplexing, Proteomics 22 (2022) 2100245, https://doi.org/10.1002/pmic.202100245.
[38] A.T. Kong, F.V. Leprevost, D.M. Avtonomov, D. Mellacheruvu, A.I. Nesvizhskii, MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics, Nat. Methods 14 (2017) 513–520, https://doi.org/10.1038/nmeth.4256.
[39] D.J. Geiszler, A.T. Kong, D.M. Avtonomov, F. Yu, F. da V. Leprevost, A.I. Nesvizhskii, PTM-Shepherd: Analysis and summarization of post-translational and chemical modifications from open search results, Mol. Cell. Proteomics. 20 (2021) 100018. doi:https://doi.org/10.1074/mcp.TIR120.002 216.
[40] D.A. Polasky, F. Yu, G.C. Teo, A.I. Nesvizhskii, Fast and comprehensive N- and O-glycopro- teomics analysis with MSFragger-Glyco, Nat. Methods 17 (2020) 1125–1132, https://doi.org/10.1038/s41592-020-0967-9.
[41] F. da Veiga Leprevost, S.E. Haynes, D.M. Avtonomov, H.-Y. Chang, A. K. Shanmugam, D. Mellacheruvu, A.T. Kong, A.I. Nesvizhskii, Philosopher: a versatile toolkit for shotgun proteo- mics data analysis, Nat. Methods 17 (2020) 869–870, https://doi.org/10.1038/s41592-020-0912-y.
[42] F. Yu, S.E. Haynes, A.I. Nesvizhskii, IonQuant enables accurate and sensitive label-free quantification with FDR-controlled match-between-runs, Mol. Cell. Proteomics 20 (2021) 100077, https://doi.org/10.1016/j.mcpro.2021.100077.
[43] O. Wagih, N. Sugiyama, Y. Ishihama, P. Beltrao, Uncovering phosphorylation-based specificities through functional interaction networks, Mol. Cell. Proteomics 15 (2016) 236–245, https://doi.org/10.1074/mcp.M115.052357.
[44] K. Li, M. Vaudel, B. Zhang, Y. Ren, B. Wen, PDV: An integrative proteomics data viewer, Bioinformatics 35 (2019) 1249–1251, https://doi.org/10.1093/bioinformatics/bty770.
[45] P.G. Arthur, J.C. Kent, P.E. Hartmann, Metabolites of lactose synthesis in Milk from women during established lactation, J. Pediatr. Gastroenterol. Nutr. 13 (1991) 260–266, https://doi.org/10.1097/00005176-199110000-00004.
[46] B. Lönnerdal, E.L. Lien, Nutritional and physiologic significance of α-Lactalbumin in infants, Nutr. Rev. 61 (2003) 295–305, https://doi.org/10.1301/nr.2003.sept.295-305.
[47] Y. Zhang, L. Zhang, S. Xia, Y. Hou, T. Wu, P. Zhou, Site-specific glycoproteomic analysis of purified lactoferrin from human and animal milk, Int. J. Biol. Macromol. 254 (2024) 127766, https://doi.org/10.1016/j.ijbiomac.2023.127766.
[48] M. Orczyk-Pawiłowicz, M. Berghausen-Mazur, L. Hirnle, I. Kątnik-Prastowska, O-glycosylation of α-1-acid glycoprotein of human milk is lactation stage related, Breastfeed. Med. 10 (2015) 270–276, https://doi.org/10.1089/bfm.2015.0049.
[49] L. Bode, Human milk oligosaccharides: every baby needs a sugar mama, Glycobiology 22 (2012) 1147–1162, https://doi.org/10.1093/glycob/cws074.
[50] Z. Darula, R.J. Chalkley, A. Lynn, P.R. Baker, K.F. Medzihradszky, Improved identification of O-linked glycopeptides from ETD data with optimized scoring for different charge states and cleavage specificities, Amino Acids 41 (2011) 321–328, https://doi.org/10.1007/s00726-010-0692-2.
[51] J. Nilsson, A. Halim, A. Grahn, G. Larson, Targeting the glycoproteome, Glycoconj. J. 30 (2013) 119–136, https://doi.org/10.1007/s10719-012-9438-6.
[52] S. Mukherjee, A. Jankevics, F. Busch, M. Lubeck, Y. Zou, G. Kruppa, A.J.R. Heck, R. A. Scheltema, K.R. Reiding, Oxonium Ion–Guided Optimization of Ion Mobility–Assisted Glycoproteomics on the timsTOF Pro, Mol. Cell. Proteomics 22 (2023) 100486. doi:https://doi.org/10.1016/j.mcpro.2022.100486.
[53] A. Boysen, G. Palmisano, T.J. Krongh, I.G. Duggin, M.R. Larsen, J. Møller-Jensen, A novel mass spectrometric strategy "BEMAP" reveals extensive O-linked protein glycosylation in entero- toxigenic Escherichia coli, Sci. Rep. 6 (2016) 32016, https://doi.org/10.1038/srep32016.
[54] S.K. Knudsen, A. Stensballe, M. Franzmann, U.B. Westergaard, D.E. Otzen, Effect of glycosylation on the extracellular domain of the Ag43 bacterial autotransporter: enhanced stability and reduced cellular aggregation, Biochem. J. 412 (2008) 563–577, https://doi.org/10.1042/BJ20071497.
[55] I.B. Wilson, Y. Gavel, G. von Heijne, Amino acid distributions around O-linked glycosylation sites, Biochem. J. 275 (Pt 2) (1991) 529–534, https://doi.org/10.1042/bj2750529.
[56] A. Pisano, J.W. Redmond, K.L. Williams, A.A. Gooley, Glycosylation sites identified by solid-phase Edman degradation: O-linked glycosylation motifs on human glycophorin a, Glycobiology 3 (1993) 429–435, https://doi.org/10.1093/glycob/3.5.429.
[57] D.J. Geiszler, D.A. Polasky, F. Yu, A.I. Nesvizhskii, Mining for ions: Diagnostic feature detection in MS/MS spectra of post-translationally modified peptides, BioRxiv. (2022) 2022.09.12.507 594. doi:https://doi.org/10.1101/2022.09.12.507594.