Journal of Animal Science, 2023, **101**, 1–11
https://doi.org/10.1093/jas/skac426
Advance access publication 31 December 2022
Lactation and Mammary Gland Biology

Characterization of sow milk N-linked glycoproteome over the course of lactation

Prabha Rajput,[†,‡] Uma K. Aryal,[‖,$] Ketaki Bhide,[¶] Radiah C. Minor,[**] Sairam Krishnamurthy,[‡] and Theresa M. Casey[†,1,]

[†]Department of Animal Sciences, Purdue University, West Lafayette, IN, USA
[‡]Neurotherapeutics Lab, Pharmaceutical Engineering and Technology, Indian Institute of Technology (Banaras Hindu University), Varanasi, UP 221005, India
[‖]Purdue Proteomics Facility, Bindley Bioscience Center, Purdue University, West Lafayette, IN, USA
[$]Department of Comparative Pathobiology, Purdue University, West Lafayette, IN, USA
[¶]Bioinformatics Core, Purdue University, West Lafayette, IN, USA
[**]Department of Animal Sciences, North Carolina A&T University, Greensboro, NC, USA

[1]Corresponding author: theresa-casey@purdue.edu

Abstract

Milk proteins serve as nutrition and affect neonate development and immunity through their bioactivity. Post-translational modifications of proteins affect their bioactivity. Glycosylation is the attachment of sugar moieties to proteins, with attachment of glycans to asparagine indicated as N-linked glycosylation. Our objective was to characterize N-linked glycosylated proteins in homogenate swine milk samples collected from sows (n = 5/6) during farrowing to represent colostrum and on days 3 and 14 post-farrowing to represent transitional and mature milk, respectively. Glycopeptides were isolated with lectin-based extraction and treated with Peptide *N*-glycosidase F (PNGase F) to identify N-linked glycosylation sites. Purified glycopeptides were analyzed by label-free liquid chromatography–tandem mass spectrometry (LC–MS/MS). MaxQuant software was used to align spectra to *Sus scrofa* Uniport database to identify proteins and measure their relative abundances. Analysis of variance and Welch's *t*-test analysis identified glycoproteins differentially abundant between colostrum, transitional, and mature milk (false discovery rate <0.05). Shotgun proteome analysis identified 545 N-linked and glutamine, Q, -linked, glycosylation ($P > 0.75$ for deamidation) sites on 220 glycoproteins in sow milk. Glycoproteins were found across all three phases of swine milk production and varied by number of glycosylation sites (1–14) and in abundance and distribution between colostrum, transitional, and mature milk. Polymeric immunoglobulin receptor was the most glycosylated protein with 14 sites identified. Also highly glycosylated were casein and mucin proteins. These data are described and the relevance of glycosylated milk proteins in neonate development, such as protection against pathogens, is discussed.

Lay summary

Milk is essential for healthy growth and development of neonates, with proteins in milk serving as key nutrients and regulators of these processes. Protein activity is affected by modifications made to their structure including the addition of sugar groups called glycans. Here we present the characterization of sow milk proteins modification with glycan groups on asparagine and glutamine amino acids in colostrum, transitional, and mature milk of pigs. We found 220 high confidences (found in at least two sows on one day) glycoproteins, and that the abundance of glycosylated proteins varied by stage of milk production and number of glycosylated sites.

Key words: glycoproteomes, lactation, milk, *Sus scrofa*

Abbreviations: ACN, acetonitrile; ANOVA, analysis of variance; DAVID, database for annotation visualization, and integrated discovery; DTT, 1,4-dithiothreitol; D0, day 0 postnatal, colostrum; D3, day 3 postnatal, transitional milk; D14, day 14 postnatal, mature milk;FA, formic acid; FDR, false discovery rate; IgA, immunoglobulin A; IgG, immunoglobulin G; IgM, immunoglobulin M; LC-MS/MS, liquid chromatography-tandem mass spectrometry; MFGE8, milk fat globule-EGF factor 8; MFGM, milk fat globule membrane; PCA, principal component analysis; PIGR, polymeric immunoglobulin receptor; PNGase F, Peptide *N*-glycosidase F

Introduction

Neonatal pigs are completely reliant on the sow's milk for a nutrient supply as well as a range of bioactive factors that impact their immunological and physiological development (Theil and Hurley, 2016). Adequate intake of the first milk colostrum and milk yield throughout lactation are the primary factors limiting survival, growth, and development of piglets. Colostrum is the first milk secreted and available to piglets in the first 24–36 h after farrowing. Colostrum is a rich concentrated source of protein reflecting, in part, the relatively high concentration of maternal antibodies needed for the establishment of passive immunity in neonates (Theil et al., 2014). Transitional milk production spans from approximately 34 h to 9 d after parturition and then the period of mature milk production commences which lasts until weaning.

The composition of mammary secretions varies across the course of lactation and reflects the metabolic activity of the gland and the developmental needs of neonates. The concentration of protein drops between the colostrum phase

Received August 30, 2022 Accepted December 30, 2022.

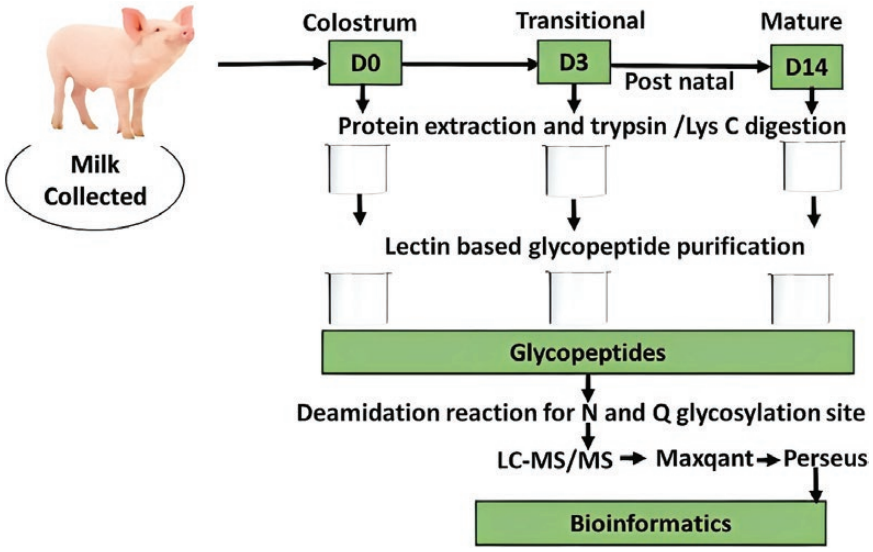

Figure 1. Workflow of the isolation and identification of glycosylated proteins in sow milk. Milk collected during farrowing, day 0 (D0) to represent colostrum, and on day 3 (D3) and 14 (D14) post-farrowing to represent transitional and mature milk.

and transitional phase of milk production, with a concomitant rise in lactose and lipid content, reflecting the onset of secretory activation of the mammary gland (Theil et al., 2014). Our recent analysis of macronutrient and lipid content of sow milk found the greatest change in composition between colostrum and early transitional milk, which was collected on day 3 postnatal, with early transitional milk more similar to mature milk, which was collected on postnatal day 14 (Suarez-Trujillo et al., 2021; Bradshaw et al., 2021) whereas there was little difference in macronutrient composition between late transitional milk, which was collected on day 7 postnatal and mature milk. Milk protein production increases in a pattern similar to the lactation curve and peaks at 700–800 g of protein secreted daily. In parallel to changes in the production level of proteins are changed in component proteins, reflecting the activity of the gland and stage-specific developmental needs of the neonate across the course of lactation (Bradshaw et al., 2021). Milk proteins are generally characterized according to how they segregate during fundamental methods of fractionating milk into lipid rich and skim components, reflecting that milk is a partially stable emulsion of fat globules. Proteins in the lipid fraction are associated with the milk fat globule membrane (**MFGM**). The skim milk fraction contains the casein and whey proteins. The whey proteins are defined as all of the skim milk proteins excluding the caseins, and encompass a wide range of proteins with an extensive array of functions. In order to understand how the proteins of each of these fractions change relative to each other across lactation, we measured the proteome of homogenate sow milk samples using liquid chromatography–tandem mass spectrophotometry (**LC–MS/MS**) (Bradshaw et al., 2021). Milk samples were collected during farrowing to represent colostrum, on day 3 postnatal to represent transitional milk and on day 14 postnatal to represent mature milk. More than 500 proteins were measured across the samples. Of these proteins, 207 were identified as being high-confidence proteins, as they were found in multiple sows within a day, and 81 were common across the three days. The difference between the most and least abundant proteins was more than four orders of magnitude, which indicated that both major and minor milk proteins were measured using our approach.

Post-translational modifications increase the diversity of proteins and affect their bioactivity as well as provide additional sources of nutrients or functional macromolecules like oligosaccharides and lipids. Glycosylation is the attachment of sugar moieties to proteins (Reily et al., 2019), and glycans mediate receptor–ligand interactions, cell communication, host–pathogen interactions, correct protein folding, and protect proteins from digestion (O'Riordan et al., 2014). Studies of milk glycoproteins indicate that such modifications are important to neonate health, as for example the milk fat globule glycoproteins mucins and lactadherin (aka milk fat globule-EGF factor 8, **MFGE8**) inhibit the adhesion of enteric and respiratory microorganisms (Schroten et al., 1992; Yolken et al., 1992; Patton et al., 1995) partly through the glycosyl groups. Although glycosylation of proteins has been implicated in various biological functions; glycoprotein components and interspecies complexity have not yet been elucidated fully in milk proteins. Glycosylation is characterized by various glycosidic linkages, including N-, O-, and C-linked glycosylation, with N-linked glycosylation accounting for approximately 90% of glycosylation types on proteins. The "N" denotes that the glycans are covalently bound to the carboxamide nitrogen on asparagine (N) residues. N-glycosylation occurs cotranslationally by the addition of glycan molecules to the nascent protein as it is being translated and transported into the endoplasmic reticulum. The overall objective of this study was to characterize N-glycosylated and glutamine (Q) glycosylated proteins in homogenate swine milk samples and describe how they changes over the course of lactation. Lectin-based purification was used to isolate glycosylated proteins from homogenate milk samples, and proteins were measured using LC–MS/MS.

Materials and Methods

Experimental animal design

All procedures involving animals were reviewed and approved by the Institutional Animal Care and Use Committee (Purdue

IACUC Protocol #1605001416) prior to the start of the experiments (Figure 1). Milk samples used for this study were collected from six multiparous sows at the Purdue University Animal Research and Education Center (**ASREC**) Swine Unit. Standard farrowing protocols for the ASREC Swine facility were followed, with sows allowed to farrow naturally, after being moved to maternity crates on day 112 of gestation. On postnatal day 1, all litters were standardized to 12–14 piglets.

The collection of samples is previously described (Bradshaw et al., 2021) but also provided here in brief. Colostrum samples were collected from sows 1 h after delivery of the first piglet and designated as day 0 (**D0**). Early transitional milk was collected on day 3 (**D3**). Milk samples collected on day 14 (**D14**) represented mature milk. During colostrum collection, piglets remained with the dam, and sample collection was done during active labor when natural levels of oxytocin are high. On days 3 and 14, at 0600, piglets were removed from the sow for 1 h before milk collection. To stimulate milk let down, just prior to collection, sows were injected with 1 mL oxytocin (VetOne; Boise, ID; 20 USP/mL) intervalvular using a 20-gauge × 1.5-inch needle. If milk was not let down in response to exogenous oxytocin, one or two piglets were returned to sow to further stimulate milk let down. Teats were cleaned with 70% ethanol and wiped clean with sterile gauze then sprayed with water to remove access ethanol left behind and wiped clean again with sterile gauze. All sows were hand milked, and samples were collected into sterile 50 mL polypropylene conical tubes. For colostrum and milk samples, at least 45 mL of milk was collected. Milk and colostrum were collected evenly from thoracic, abdominal, and inguinal mammary glands. Following collection, homogeneous milk samples were aliquoted equally into 2 mL conical tubes, flash frozen in liquid nitrogen, and stored at −80 °C until further analysis.

Sample preparation for glycopeptides enrichment

Isolation of glycosylated proteins and proteome analysis was conducted in the Purdue Proteomics Facility at the Bindley Bioscience Centre, Purdue University. Analysis was conducted on milk samples from the six sows collected each of the three days, except, due to the loss of one D14 sample, only five samples were available for analysis on this day. Frozen milk samples were thawed on ice. One hundred fifty microliters of milk samples were homogenized in the Barocycler NEP2320 (Pressure Biosciences, Inc.) at 4 °C for 90 cycles (45 min, each cycle lasting 20 s at 33,000 psi followed by 10 s at 1 atmosphere). After homogenization, samples were transferred to a new tube and mixed with 3 volumes of 8 M urea and vortexed for 1 min before centrifugation at 13,500 rpm for 10 min at 4 °C. The clear solution was transferred to a new tube and protein concentration was determined using a Pierce™ BCA assay kit (Thermo Scientific). Based on the protein concentration, an aliquot containing 100 µg of protein (equivalent volume) was taken for processing. Before the digestion, protein was precipitated and concentrated from the solution using four volumes of cold acetone (−20 °C) overnight. After drying the precipitated protein pellets in SpeedVac, the protein samples were reduced using 10 µL of 10 mM 1,4-dithiothreitol (**DTT**) followed by alkylation with 10 µL of 2% iodoethanol, 0.5% triethylphosphine in acetonitrile (**ACN**) as described previously (Mohallem and Aryal, 2020), Sequence grade Lys-C/Trypsin (Promega) prepared in 25 mM ammonium bicarbonate buffer was used to enzymatically digest the protein samples, and digestions were carried out in the Barocycler NEP2320 (Pressure Biosciences, Inc.) at 50 °C under 20,000 psi for 1 h (60 cycles, each cycle lasting 50 s at 20,000 psi and 10 s at 1 atmosphere). Digested samples were cleaned using the microspin C18 columns (Nest Group) using manufacturer's protocol, dried, and resuspended in 10 µL of 3% ACN, 0.1% formic acid (**FA**) in water.

Glycopeptides were purified using ProteoExtract Glycopeptide Extraction kit (Millipore Sigma) following the manufacturer's protocol. The ProteoExtract kit is designed to capture high-mannose, hybrid, and complex-type glycosylated peptides. Briefly, peptides reconstituted in binding buffer were mixed with ZIC Glycocapture Resin and incubated with continuous mixing at 1,200 rpm for 20 min. Samples were then washed with washing buffer, and incubated with agitation for 10 min, for a total of three times. Glycopeptides were eluted with elution buffer and dried in a SpeedVac. Glycopeptides were resuspended in 19 µL of 25 mM ammonium bicarbonate buffer and treated with 1 µL Peptide *N*-glycosidase F (**PNGase F**) for 1 h at 37 °C. PNGase F is an amidase of the peptide-N4-(*N*-acetyl-beta-glucosaminyl) asparagine that cleaves between the innermost GlcNAc and asparagine residues, resulting in a deaminated peptide and a free glycan Following PNGase F treatment, samples were dried, resuspended in 10 µL of 97% purified water, 3% ACN, and 0.1% FA.

Liquid chromatography tandem mass spectrometry

Glycoproteome analysis was performed using enriched glycopeptides by label-free shotgun LC–MS/MS. Of the 10 µL peptide sample, 4 µL was injected and analyzed in a Dionex UltiMate 3000 RSLC nano System (Thermo Fisher Scientific, Odense, Denmark) coupled online to an Orbitrap Fusion Lumos Mass Spectrometer (Thermo Fisher Scientific, Waltham, MA, USA), as described previously (Mohallem and Aryal, 2020; Zembroski et al., 2021). Reverse phase peptide separation, using a trap column (300 µm ID × 5 mm) packed with 5 µm 100 Å PepMap C18 medium coupled to a 50-cm long × 75 µm inner diameter analytical column packed with 2 µm 100 Å PepMap C18 silica (Thermo Fisher Scientific). The column temperature was maintained at 50 °C. The sample was loaded to the trap column in a loading buffer (3% ACN, 0.1% FA) at a flow rate of 5 µL/min for 5 min and eluted from the analytical column at a flow rate of 200 nL/min using a 120-min LC gradient as follows: linear gradient of 6.5%–27% of solvent B (80% ACN, 19.9% water, and 0.1% FA) in 82 min, 27%–40% of solvent B in next 8 min, 40%–100% of solvent B in 7 min at which point the gradient was held at 100% of solvent B for 7 min before reverting to 2% of solvent B and held at 2% for 15 min for column equilibration. The column was washed and equilibrated using three 30-min LC gradients before injecting the next sample. All the data were acquired in the Orbitrap mass analyzer using a higher energy collisional dissociation fragmentation scheme. The MS scan range was from 350 to 1,600 m/z at a resolution of 120,000. The automatic gain control target was set at 4 × 105, maximum injection time (50 ms), dynamic exclusion 30 s, and intensity threshold 5.0 × 104. MS data were acquired in data-dependent mode with a cycle time of 5 s/scan. MS/MS data were collected at a resolution of 15,000. MaxQuant software (version 1.6.3.3) was used for protein identification by searching against the *Sus scrofa* database

Table 1. Categories and pathways enriched with non-glycosylated proteins

Category	No. proteins	*P*-value	Proteins
KEGG pathway complement coagulation cascade	7	1.7E−7	C5, C6, C8B, CFB, FGB, PLG, SERPINF2
Innate immunity	6	0.0004	LBP, C5, C8B, C6, CFB, FGB
Membrane attack complex	3	0.0001	C5, C6, C8
Innate immune response	6	0.001	APOE, ACTG1, HSPA8, CDH1, CTNNB1, CAP1,
Complement pathway	3	0.001	C5, C8B, C6
Rap1 signaling pathway	5	0.004	GNAS, ACTG1, ANGPT2, CDH1, CTNNB1
Adherens junction	3	0.02	ACTG1, CTNNB1, CDH1
Calcium ion binding	7	0.003	MCFD2, FBLN2, NUCB1, GSN, SDF4, CDH1, AOAH

downloaded from UniProt (www.uniprot.org). The following parameters were used: precursor mass tolerance of 10 ppm; enzyme specificity of trypsin/Lys-C allowing up to two missed cleavages; oxidation of methionine (M), acetyl N-term, and deamination of asparagine (N) and glutamine (Q) as variable modifications and iodoethanol (C) as a fixed modification. Peptide spectral match and protein identification were set at a false discovery rate (**FDR**) of <0.01. Results were filtered to retain only proteins with non-zero iBAQ values in at least two samples within the days and MS/MS (spectral counts) ≥ 2 for further analysis. Glycoproteins were quantified by unique plus razor peptides. Razor peptides are the non-unique peptides assigned to the most of the proteins The LC–MS/MS raw data files are available in the massive data repository (massive.ucsd.edu) under ID MSV000090185.

Following the removal of decoy and contaminated proteins, chemical deamidation was used to identify N and Q glycosylation sites. A protein was considered as glycosylated if the probability score of deamidation was >0.75 (class 1 glycosites). Data were uploaded to Perseus (v.16.15.0) for downstream analysis. Data were filtered such that at least two samples on any 1 d had a non-zero iBAQ value. For 0 values that remained, imputed values were generated using Perseus tool replace missing values based on normal distribution. iBAQ values were log base 2 transformed, and statistical analysis that included principal component, analysis of variance (**ANOVA**), and Welch's *t*-test analysis was used to determine differentially abundant proteins (Tyanova et al., 2016).

Proteins were identified as differentially abundant between days using an adjusted *P*-value cut-off of 0.05. The UniProt I.D.s were converted into Ensemble I.D.s using Database for Annotation, Visualization, and Integrated Discovery (**DAVID**); (Huang et al., 2009) and Ensemble BioMart (Howe et al., 2021) software. Duplicate Ensemble IDs were removed by selecting the reads with the highest intensity across all time points. To retain robustness of data during functional annotation analysis, *S. scrofa* Ensemble IDs were converted into human Ensemble IDs before uploading into DAVID (Huang et al., 2009). GeneCards (Stelzer et al., 2016) and Uniprot databases were used to determine protein functions. To identify protein common between data sets, Ensemble IDs were uploaded into Venny 2.1 (Oliveros, 2007–2015).

Results

MaxQuant software mapped LC–MS/MS data to 466 unique Unitprot/Swiss Prot I.D.s in the *S. scrofa* database. Of these, 319 were considered high-confidence proteins, as defined by being detected in two samples within a day. After removing redundant proteins, 289 remained, and 69 (21.63%) proteins did not have high-confidence N or Q glycosylation sites (Supplementary Table S1). Functional annotation analysis of the proteins that lack the criteria of glycosylation (the probability score for deamidation >0.75) found seven of these enriched the KEGG pathway complement coagulation cascade, six were involved in innate immunity, and three were involved in membrane attack complex and six as innate immune response. There were also three involved in complement pathway and five enriched the Rap1 signaling pathway (Table 1).

Principal component analysis (**PCA**) and heat map of proteins with N/Q-link glycosylation sites demonstrated distinct clustering of milk samples by day indicating that phase of milk production affects the abundance and distribution of glycosylated proteins (Figure 2a and b). There were 220 non-redundant proteins that had at least one N or Q glycosylation site (probability score of >0.75), and are shown according to the changes in their abundance across the three phases of lactation in Supplementary Table S2. In particular, there 63 glycoproteins differentially abundant between D0 and D3, between D3 and D14 there were 86 glycoproteins, and between D0 and D14, 135 glycoproteins (FDR < 0.05). Thus, as illustrated by the heat map, which shows all glycosylated proteins by individual sow and day, and data in Supplementary Table S2, it is evident that the relative abundance of most glycosylated proteins decreases as lactation progresses from colostrum to transitional and then mature phase of milk production.

Of the 63 proteins differentially abundant between D0 and D3, 46 decreased and 17 increased. Proteins that decreased in abundance included 5 Ig/Ig-like proteins, multiple growth factors (FGF1, TGFB3, MSTN, PDGFC) and their modifiers (NBL1) as well as lysosomal enzymes (DNASE2, GNS, HEXA) vitamin (BTD) and mineral (SELENOP, CP) transporters and extracellular matrix components (LAMA1, LAMC1, THBS1, FMOD). Glycosylated proteins that significantly increased between D0 and D3 included multiple major milk proteins-MFGM proteins (CSN1S1, LALBA, PLIN5, MUC4, SAA2, BTN1A1, LPO) as well as CD209, a pathogen recognition receptor, MAN2A1, which catalyzes the first committed step in the biosynthesis of complex *N*-glycans, and Toll-like receptor-2 (TLR2).

Almost all of the glycoproteins differentially abundant between D0 and D14 were lower on D14 than on D0. Among the proteins that decreased between the days were two apolipoproteins (APOD and APOH), five complement proteins (C2, C3, C4A, C7, and C8G), three carboxypeptidases (CPM, CPB2, and CPQ), four cathepsins (CTSA, CTSB, CTSL, and

CTSZ), two fucosidases (FUCA1 and FUCA2), and three serine protease inhibitors (SERPINC1, SERPIND1, and SERPING1).

Functional annotation analysis of the glycosylated proteins found 23 (10.45%) involved in KEGG pathway of lysosome. Seventeen (7.72%) were involved in negative regulation of endopeptidase activity, and eight (3.63%) were involved in endopeptidase inhibitor activity. Four glycosylated proteins were categorized as being involved in blood coagulation, and eight in proteolysis, including cathepsin D (CTSD). Five were found involved in tissue development and four were involved in iron ion transport (Table 2).

The number of N and Q glycosylation sites per protein was counted and found to range from 14 to 1 (Table 3; Figure 3; Supplementary Table S3). To gain a better understanding of the potential role of glycosylated proteins in milk they were categorized as highly glycosylated (≥6 glycosylation sites), moderately glycosylated (5 to 2), and low glycosylated (1 glycosylation site) categories. Nineteen (8.63%), proteins were defined as being heavily glycosylated. The most highly glycosylated protein found in swine milk was polymeric immunoglobulin receptor (PIGR), with 14 (12N and 2Q) glycosylation sites. There were two casein proteins in the heavily glycosylated category, CSN3 (Kappa-casein) with 12 (1N and 11Q) glycosylation sites, and CSN1S1 (Alpha-S1-casein) having 10 (5N and 5Q) glycosylation sites. Three mucin proteins were considered either heavily or moderately glycosylated proteins. MUC4 had 11, MUC15 had 9, and MUC1 had 3 glycosylation sites, which were respectively 10 N and 1 Q, 8N and 1Q, and 3N.

Among the moderately glycosylated proteins, there were seven with five glycosylation sites, which included transcobalamin 1 (**TCN1**) and lactoferrin (**LTF**). Sixteen (7%) have four glycosylation sites and included the J chain (joining chain of multimeric immunoglobulin A [**IgA**] and immunoglobulin M [**IgM**]). Functional annotation analysis indicated that moderately glycosylated proteins had eight involved in negative regulation of endopeptidase activity, six categorized

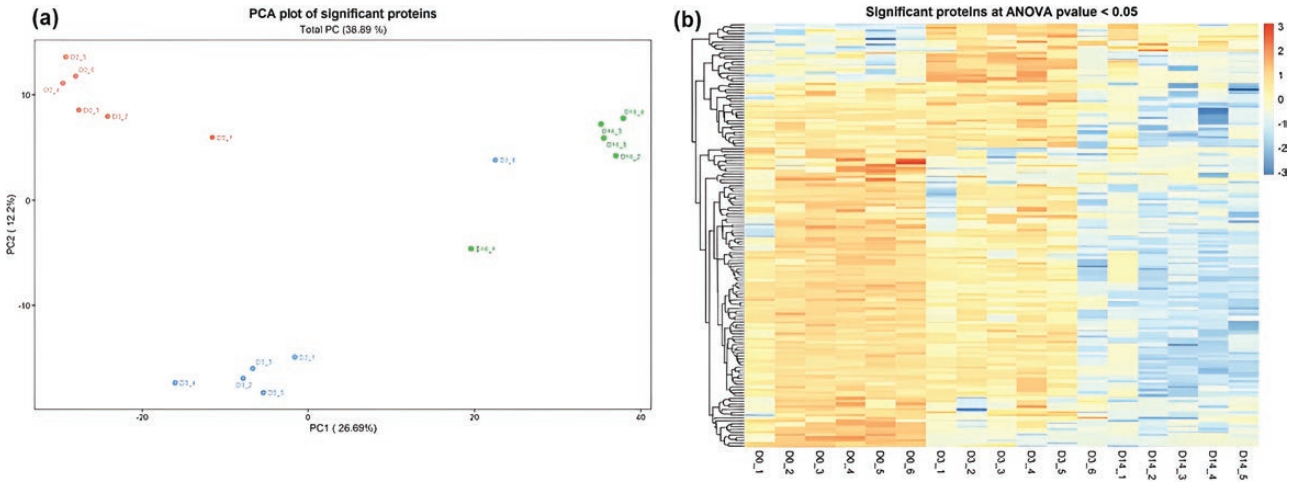

Figure 2. Principal component scores plot of glycosylated proteins collected in the milk samples from sows during farrowing day 0 (D0; red) to represent colostrum, day 3 (D3; blue) postnatal to represent transitional milk, and day 14 (D14; green) postnatal to represent mature milk (a). Heat map of glycoproteins differentially abundant $P < 0.05$ between day 0 (D0), day 3 (D3), and day 14 (D14) (b).

Table 2. Categories and pathways enriched proteins with N/Q-linked glycosylation sites

Category	No. proteins	*P*-value	Proteins
KEGG pathway of lysosome	23	8.2E-20	ATP6AP1, CD63, ASAH1, NPC2, GAA, FUCA1, FUCA2, IDUA, CTSA, CTSB, CTSD, CTSH, CTSL, CTSZ, DNASE2, GNS, GB, HEXA, LGMN, MAN2B1, PPT1, PSAP, SUMF1
Negative regulation of endopeptidase activity	17	2.4E-18	TIMP1, AHSG, C3, FETUB, HRG, ITIH2, ITIH4, KNG1, SERPINB4, F1SCC7, SERPINA3, SERPIN A3-8, SERPINA1, SERPIN, A7, SERPINC1, SERPIND1, SERPING1
Glycosidase	11	9.5E-10	GAA, FUCA1, FUCA2, IDUA, GANAB, GBA, HPX, HEXA, MAN2A1, MAN2B1, AMY2A
Endopeptidase inhibitor activity	8	1.3E-10	AHSG, A2M, C3, C4A, FETUB, HRG, KNG1, PZP
Antigen processing and presentation	4	0.04	CALR, CTSL, LGMN, CTSB
Blood coagulation	4	0.001	FGG, SERPINC1, PROC, KLKB1
Proteolysis	8	0.0001	ANPEP, CPB2, CTSD, CTSH, LTF, LGMN, PROC, TFRC
Tissue development	5	0.0001	POSTN, GAA, LAMA1, LAMC1, LAMB1
Iron ion transport	4	0.003	TFRC, CP, TF, LTF
Developmental protein	6	0.04	DNASE2, ANPEP, SEMA7A, CHRDL2, PDGFC, FGF1

Table 3. Glycosylated proteins categorized by number of N or Q glycosylated sites per protein

No. N/Q-linked sites per protein	No. proteins	Proteins
Highly glycosylated proteins (≥6 sites)		
14	1	PIGR
12	1	CSN3
11	2	LAMA3, MUC4
10	1	CSN1S1
9	4	CEACAM1, SAA2, FMOD, MUC15
8	4	A0A287AGW0, FSTL1, I3LGT0, CLU
7	3	A0A287B5C1, BRICD5, CP
6	3	A0A286ZSX6, MGAM, LALBA
Moderately glycosylated proteins (2–5 sites)		
5	7	TCN1, SLC6A14, LTF, ITIH4, FAT2, CPM, TLR2
4	16	PGLYRP1, APOD, ANPEP, JCHAIN, A0A5G2QXT5, CHRDL2, LAMC1, F1SCD1, ORM1, THBS1, CPN2, TNC, VWDE, P00761, P01846, CSN1S2
3	28	A0A075B7I9, APOH, CTSA, APOB, SLC9A3R1, YBX1, CHL1, SERPINA3-2, GNS, PTPRJ, HPX, LAMB1, FGG, LPO, MUC1, SERPIND1, F1RL06, C4A, FUCA2, F1SCC9, LGMN, ZNF831, FGL2, HAPLN3, PSAP, CEACAM21, NPC2, IGFALS
2	47	A0A286ZJZ7, PTPRK, RNASET2, LGALS3BP, FUCA1, HEXA, PTGDS, A0A287B626, AEBP1, B4GALT1, RFC1, PON1, KNG1, ADAM12, FN1, MERTK, SEMA7A, ADGRG1, C7, SERPINC1, IDUA, A0A5G2R9V5, A0A5G2RDR6, FOLR1, SERPING1, CTSZ, BTN1A1, BTD, KLKB1, F1S1G8, SMPDL3A, GOLM1, SLC34A2, MAN2B1, CREB3L1, LUM, F6Q5X1, I3LHC3, SELENOP, CSF2RA, TF, SPP1, TGFB3, AHSG, GANAB, CTSV, MGP
Low glycosylated proteins (1 site)		
1	103	A0A075B7H9, A0A075B7J0, HYOU1, A0A286ZIM1, TINAGL1, SLC44A2, PPIB, RTN4R, NBL1, SUSD5, SOD3, MYH7, A0A286ZQJ9, ALB, CD82, DAG1, A0A287A4Y3, PTPRF, MFAP4, CD5L, C8G, CD36, A0A287ATT2, ADGRG6, FUT2A, A0A287BF75, FUT11, SLC29A1, A0A287BRF1, CANT1, PRNP, CD63, PPT1, HSPG2, PROC, CPB2, ATP6AP1, POSTN, CFI, MINPP1, HRG, LY96, TIMP1, F11R, ASAH1, LPL, GBA, MAN2A1, PLBD2, PGLYRP4, C2, PLA2G7, C3, MPZL1, MFGE8, F12, SIL1, OAZ1, FOLR2, CTSB, CD209, CTSH, B4GAT1, FAM20A, FGA, CCDC40, B4GALT3, TSPAN1, SLC28A3, CTSL, FGFBP1, C8A, LRG1, SERPINA1, MATN4, DNASE2, FETUB, SUMF1, F1SG31, TP23, PRSS23, PLOD3, HS6ST1, CPQ, PDGFC, PZP, FAM20C, IL18BP, MSTN, CTSD, SAL1, ST3GAL1, F2, HSP90B1, TFRC, SERPINA7, SEMA3F, CREG1, GAA, SLC15A1, ITIH2, MASP2, A2M

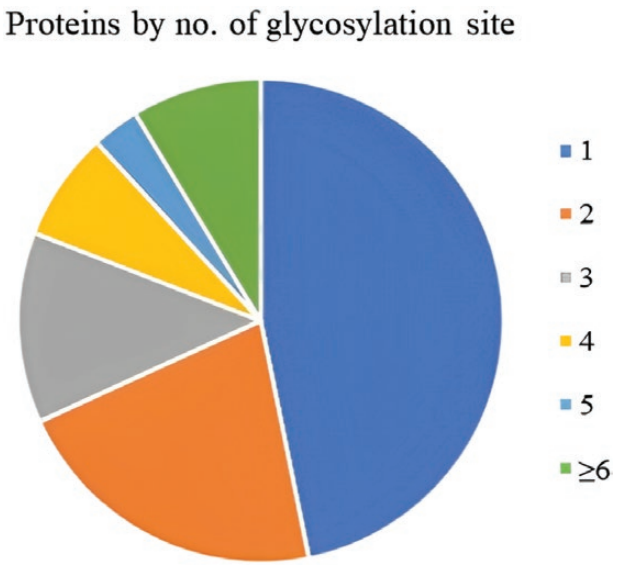

Figure 3. Distribution of glycosylated proteins by total number of N/Q-glycosylation sites.

as serine-type endopeptidase inhibitor activity, eight as glycosidase, and three involved in blood coagulation (Table 4).

There were 103 proteins (47%) identified with a single high-confidence glycosylation site and categorized as low glycosylated proteins. Functional annotation analysis found 7 of them were involved in serine-type endopeptidase inhibitor activity, 10 categorized as lysosome, 5 involved proteolysis, and 5 were involved in endopeptidase inhibitor activity. There were eight proteins involved in negative regulation of endopeptidase activity, nine involved in complement and coagulation cascades, and three were involved in complement pathway (Table 4).

In addition, we compared the high-confidence glycoprotein data set with the global swine milk proteome data set (Bradshaw et al., 2021) and found only 58 (28%) of the 207 high confidence proteins measured in the global proteome were found in common with the data set of high confidence glycoproteins considered in the current analysis (Figure 4). This finding indicates that lectin-based glycoprotein enrichment brings down a unique set of proteins.

Discussion

Lectin-based purification followed by LC–MS/MS proteome analysis identified 545 glycosylation sites on 220 glycoproteins in sow milk. Glycoproteins were found across all three phases of milk production and varied in abundance and

Table 4. Functional annotation analysis categories and pathways enriched with moderate and lowly glycosylated proteins

Category	No.	*P*-value	Proteins
Moderately glycosylated			
Negative regulation of endopeptidase activity	8	2.34E–08	SERPIND1, SERPINC1, AHSG, SERPING1, SERPINA3-8, ITIH4, F1CC7, KNG1
Serine-type endopeptidase inhibitor activity	6	9.00E–05	SERPIND1, SERPINC1, SERPING1, SERPINA3-8, ITIH4, F1CC7
Glycosidase	8	1.67E–08	HPX, HEXA, FUCA2, FUCA1, MAN2B1, AMY2A, GANAB, IDUA
Blood coagulation	3	0.006	FGG, SERPINC1, KLKB1
Cell adhesion	8	0.0002	TNC, FOLR3, ADGRG1, SPP1, CHL1, HAPLN3, THBS1, LGALS3BP
Receptor binding	7	0.0003	TNC, BTN1A1, SLC9A3R1, FGG, FGL2, KNG1, VWDE
Low glycosylated			
Serine-type endopeptidase inhibitor activity	7	1.2E–05	PZP, SERPINA7, HRG, A2M, SERPINA1, SERPINB4, ITIH2
Lysosome	10	3.06E–11	PPT1, DNASE2, CD63, SUMF1, CTSL, FUCA2, GAA, ATP6AP1, CTSD, CTSH
Proteolysis	5	0.002	TFRC, PROC, CPB2, CTSD, CTSH
Endopeptidase inhibitor activity	5	1.26E–06	PZP, A2M, FETUB, C3, HRG
Negative regulation of endopeptidase activity	8	2.83E–08	SERPINA7, FETUB, C3, HRG, TIMP1, SERPINA1, SERPINB4, ITIH2
Complement and coagulation cascades	9	3.78E–08	C2, MASP2, C8G, CPB2, C3, A2M, PROC, SERPINA1, KLKB1
Complement pathway	3	0.001	C2, MASP2, C3

distribution between colostrum, transitional and mature milk. The number of glycosylation sites per protein also varied, ranging from 1 to 14, with 47% having only one high confidence glycosylation site. Although milk protein samples in the current study were enriched for N-glycosylated proteins, 51 glycoproteins were identified as having glycosyl groups on glutamine (Q-linked glycosylation), with 79 of 545 (14.5%) of the total glycosylation sites being Q-linked. The 220 glycoproteins identified accounted for approximately 50% of the milk proteins measured across the two studies, which is similar to the current estimation of 50% of proteins in the body being glycosylated. (Song et al., 2021). Here we discuss the potential role of glycosylation of sow milk proteins and compare our results to similar studies of milk from other species.

The number of proteins measured in the current study was similar to what we found in our analysis of milk proteome across the same phases of sow lactation (Bradshaw et al., 2021). However, these data are largely additive. Lectin-based enrichment resulted in isolation of a majority of proteins unique from the non-enriched global proteome we previously described. In both studies, we isolated proteins from homogenate milk samples. Analysis of homogenized milk resulted in a similar number of proteins when compared to studies that analyzed proteins following fractionation into skim or MFGM components and enabled the querying of the relative abundance of total milk proteins to each other. We identified a greater number of glycoproteins and glycosylation sites per protein in our sow milk samples compared to any one species recently analyzed for N-linked glycosylation sites and glycoproteins in mature MFGM (Yang et al., 2016). The milk samples in Yang et al. (2016) study where from Holstein and Jersey cows, buffaloes, yaks, goats, camels, horses, and humans and number of glycoproteins measured ranged from 83 to 106 by species. Functional annotation profiles of MFGM glycoproteins were similar to what we found for homogenate milk samples from swine,

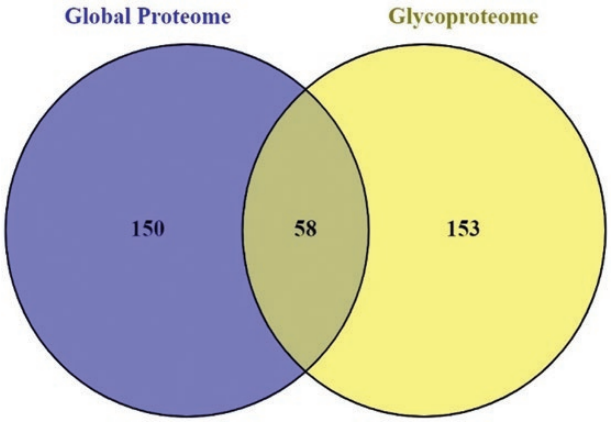

Figure 4. Venn diagram of glycoprotein and global proteome (Bradshaw et al., 2021) data sets, show that only 58 overlaps.

and included the "response to stimulus" gene ontology category, supporting an immunogenic role of glycosylated milk proteins. A study of changes in human MFGM glycoproteome between colostrum and mature milk production phases (Cao et al., 2018) showed similar changes in abundance between colostrum and mature milk phases of synthesis as were measured for homogenate sow milk samples. However, this group found nearly twice the number of glycosylation sites and glycoproteins in human colostrum and mature milk samples compared to our analysis of sow milk. The higher recovery of glycoproteins in human milk may have been due to the pooling of milk samples for analysis and lack of established stringency for identifying a high-confidence glycosylation site.

Similar to what was found for other species (Yang et al., 2016), sow milk glycoproteins enriched categories related

to lysosomes and lysosomal activity. Lysosomal proteins are considered highly glycosylated proteins. This is due, in part, to luminal lysosomal proteins being directed to the lysosome via the mannose 6-phosphate (Man6P) pathway (Dahms et al., 2008). The N-linked carbohydrates receive a Man6*P* modification, which is recognized by Man6*P* receptors that direct the vesicular trafficking of the lysosomal proteins from the *trans-Golgi* to the lysosome. Glycosylation of lysosomal membrane proteins also mediates intracellular homeostasis inclusive of regulation of autophagy (Sudhakar et al., 2020). Involution of the mammary gland following lactation proceeds through lysosomal-mediated cell death and so lysosomes play a central role in mediating mammary epithelial cell homeostasis (Sargeant et al., 2014). Lysosomal proteins are also expressed at relatively high levels during lactation. (Luo et al., 2018). Whether lysosomal proteins and their glycosylation play a role in mediating milk components and in regulating neonate development needs further studies.

The most highly glycosylated protein in sow milk was PIGR. PIGR binds polymeric immunoglobulin molecules, IgA and IgM, at the basolateral surface of epithelial cells, and then transports them across the cell to be secreted at the apical surface. Transport of immunoglobulins across epithelial cells begins with the non-covalent binding of the joining (J) chain to the extracellular domain 1 of PIGR. PIGR bound immunoglobulins are transcytosed to the apical surface where the receptor undergoes endoproteolytic cleavage and disassociates from the membrane-bound domain, forming a secretory component. The secretory component remains bound to the immunoglobulin and forms secretory immunoglobulins (Turula and Wobus, 2018). In vitro studies support that glycosylation of PIGR mediates the transport or release of free PIGR (Matsumoto et al., 2003).

Multiple mucins were found to be high to moderate glycosylated proteins in sow milk. Mucins are a family of heavily glycosylated proteins (glycoconjugates) produced by epithelial tissues. Mucins are the primary constituent of mucus that lines all components of the gastrointestinal tract and the relatively high level of mucin proteins in milk may help in establishing the mucus lining in neonates. Mucus in the small intestine prevents bacteria from reaching the epithelium and the Peyer's patches, and in the large intestine, the inner mucus layer separates the commensal bacteria from the host epithelium (Pelaseyed et al., 2014). The key characteristic of mucin proteins is their ability to form gels that functions as lubrication and in forming chemical barriers (Marin et al., 2007). Mucins bind to pathogens and are considered a component of the innate immune system. Functional studies of milk mucins have demonstrated bactericidal roles of mucin proteins (Li et al., 2012) and antiviral actions mediated by glycan groups (Sargeant et al., 2014). Mucins in the gastrointestinal tract are primary gatekeepers and controllers of bacterial interactions with the host immune system. The glycan groups linked to mucin proteins can be quite complex and cross-linking among them along the gastrointestinal tract prevents degradation of cells by peptidases (Pelaseyed et al., 2014). Also identified as a glycosylated protein was maltase-glucoamylase (MGAM). MGAM is a brush border membrane enzyme that plays a role in the final steps of starch digestion. Presence of MGAM protein in milk likely functions to digest starch as suckling rodents lack all six enzymes needed for the digestion of starch (Nichols et al., 2016).

Milk proteins that included the MFGM protein MFGE8, whey proteins LALBA and LTF, as well as multiple caseins (CSN) were among the glycosylated proteins. Analysis of glycoproteins in cow milk found MFGE8 (aka lactadherin) to have both N and O glycosylation sites (O'Riordan et al., 2014). MFGE8 was also found to be heavily glycosylated in human milk. Our analysis of swine milk only found evidence for one high-confidence N-linked site on MFGE8. LALBA (alpha-lactalbumin) was found to have six (5N and 1Q) glycosylation sites in sow milk. Studies of human colostrum and mature milk also found LALBA to be glycosylated, but only a single glycosylation site was identified (Cao et al., 2017). LTF was also found to be among the glycosylated proteins in human milk. We found 5 N-linked glycosylation sites on LTF in sow milk. Lactoferrin is an iron-binding protein and has been characterized as a heavily glycosylated protein, with one of the glycan groups functioning to mediate iron binding (Karav et al., 2017). CSN3 is a well-characterized glycoprotein in milk. Studies of the role of glycosylation of CSN3 in dairy cow milk found that it varied from 22% to 76% of the captured protein as glycosylated. The variation in level of CSN3 glycosylation found it related to heritability as well as total amount of protein synthesized and efficiency of glycosylation (Bonfatti et al., 2014). Analysis of rennet coagulation time found the more glycosylated CSN3 was, the faster the coagulation time, suggesting the potential for positive selection of protein glycosylation to enable a slow release of casein proteins between suckling bouts of neonates (Huppertz and Chia 2021).

Laminin subunit alpha 1 (LAMA1) was also found to be a glycosylated protein in swine milk. Laminins are a family of extracellular matrix glycoproteins that make up a major component of the basement membrane milk producing alveolar cells are attached to and are needed for induction of milk protein expression (Kariya et al., 2008).

Several proteins considered highly to moderately glycosylated were categorized as being related to immunity. Carcinoembryonic antigen-related cell adhesion molecule five isoform X2 (CEACAM1) with nine N-linked glycosylation sites, belongs to the immunoglobulin superfamily, and found to be a cell–cell adhesion molecule detected on leukocytes, epithelia, and endothelia. Its specific role in milk and neonate development are currently unknown, but recent studies suggest a function in mediating T-cell function through inhibitory actions (Shhadeh et al., 2021). Serum amyloid A protein (SAA2) was also found to have nine (7N and 2Q) glycosylation sites. SAA2 is a member of the serum amyloid A family of apolipoproteins, and activates the inflammasome cascade, it possesses many proinflammatory and cytokine-like properties. In human neutrophils, SAA induces the production of TNF-a, IL-1β, and IL-838 and, in human mast cells, the expression of TNF-a and IL-1β (Eklund et al., 2012). Peptidoglycan recognition protein 1 (PGLYRP1) has four (3Q and 1N) glycosylation sites and is involved in peptidoglycan binding activity and peptidoglycan immune receptor activity. A study of camel milk supports a role of PGLYRP1 in the protection of the lactating mammary gland as well as the passive immunization of the neonate (Kappeler et al., 2004).

There were 69 proteins enriched in samples following lectin-based selection that lacked high-confidence glycosylation sites based on low deamidation scores. Among these were beta-lactalglobulin (BLG), several MFGM proteins, and

Ig-like immunoglobulin proteins. BLG in cow's milk was shown to have low levels of glycation, however, heat treatment induces the Maillard reaction and results in the formation of glycosylation sites on the protein (Perusko et al., 2018). The glycosylated state altered both cellular uptake of BLG and immune response to the protein.

Our study was limited to characterizing milk glycoproteins based on N-linked glycosylation sites. N-glycosylation accounts for only one of the two primary types of protein glycosylation, with O-linked site being the other. Proteins with O-linked sites have a glycan chain covalently attached to the hydroxyl oxygen of a serine or threonine residue. Multiple studies showed that milk glycoproteins have an abundance of O-linked glycans that vary by species and lactation stage (Kwan et al., 2021) and so future studies of sow milk should include enrichment of these for a more comprehensive profile of sow milk glycoproteome.

An uncommon aspect of this study is that we identified glycosylation of not only asparagine residue but also glutamine residues in milk proteins. While N-linked glycosylation is the most common type and most reported, reports on glycosylation of glutamine residue are rare. In this study, about 14% of the glycosylation sites (probability of site-specific glycosylation ≥ 0.75) were Q-linked glycosylation sites. Multiple filtering steps were applied to ensure that these N- and Q-linked glycosylation sites were of high confidence. Thus, this study combined with recent others (Valliere-Douglass et al., 2010; Daubenspeck et al., 2015) may suggest that Q-linked glycosylation may not be as uncommon as previously thought. However, additional investigations are needed to confirm and expand the knowledge base of Q-linked glycosylation.

Our analysis was limited to the difference in abundance of glycoproteins across stages of a sow's lactation and not the changes in glycosylation sites across these stages. We focused on the characterization of changes across lactation, as this is our realm of interest in understanding. We provide data in an open-access forum for further probing of changes in sites by other research groups. The glycan groups attached to the glycoproteins were also not analyzed. This information is important as it likely reflects atopic disease, (Holm et al., 2022), as well as establishes the unique microbiome of the suckling neonate (Bode, 2020). However, the approach is distinct from what was used to enrich and profile the N-linked glycoproteome, (Wilson et al., 2008) and thus, beyond the scope of the current study.

In reviewing the patterns of relative glycoprotein abundance in PCA plots and heat maps, between sows, within a day and across days, a high degree of biological variation is noted. The variation in milk composition and quality has been reported elsewhere, especially in regard to the variability in the quality of colostrum and IgG content (Quesnel 2011). Despite the biological variability, we found a significant difference in the abundance of multiple glycoproteins between the days using a standard, but high stringency, corrected for multiple testing, of FDR < 0.05. These differences reflect the large differences in composition among colostrum, transitional, and mature milk. The differences, as we have previously discussed in relation to the global milk proteome (Bradshaw et al., 2021), reflect the metabolic activity-state of secretory differentiation-activation of the mammary gland, and likely the nutritional and developmental needs of the neonate.

Conclusion

The characterization presented here shows that across the course of a sow's lactation the N-glycosylated proteins and sites varied in abundance by stage of milk production. The majority of the proteins identified as having high confidence N-linked glycans were unique from our previous analysis of global swine milk proteome, indicating that the isolation of glycoproteins with lectin captures a distinct set of proteins, furthering our knowledge of the complexity of milk. Identification of many proteins glycosylated in asparagine as well as glutamine expands our knowledge and understanding of saw milk protein glycosylation. The most highly glycosylated proteins found in the sow milk, e.g. PIGR and mucins, are important bioactive molecules that likely play roles in the protecting neonates from pathogen infection, with the glycan units playing roles in this bioactivity.

Supplementary Data

Supplementary data are available at *Journal of Animal Science* online.

Acknowledgments

This study was supported through grant NCX-254-5-11-120-1 from the National Institute of Food and Agriculture to R.C.M, its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of Food and Agriculture. LC–MS/MS experiments were performed at the Purdue Proteomics Facility in Bindley Bioscience Center, Purdue University. Prabha Rajput received a fellowship from the Science and Engineering Research Board, India Overseas Visiting Doctoral Fellowship (SERB-OVDF), Program.

Conflict of Interest Statement

The authors declare that there is no conflict of interest.

Literature Cited

Bode, L. 2020. Human milk oligosaccharides: structure and functions. *Milk, mucosal immunity and the microbiome: impact on the neonate*. Nestlé Nutr. Inst. Workshop Ser. Basel, Karger, vol. 94; p. 115–123. doi:10.1159/000505339.

Bonfatti, V., G. Chiarot, and P. Carnier. 2014. Glycosylation of κ-casein: genetic and nongenetic variation and effects on rennet coagulation properties of milk. *J. Dairy Sci.* 97:1961–1969. doi:10.3168/jds.2013-7418.

Bradshaw, C. V., A. Suarez Trujillo, S. M. Luecke, L. D. Logan, R. Mohallem, U. K. Aryal, K. R. Stewart, T. M. Casey, and R. C. Minor. 2021. Shotgun proteomics of homogenate milk reveals dynamic changes in protein abundances between colostrum, transitional, and mature milk of swine. *J. Anim. Sci.* 99:skab240.doi:10.1093/jas/skab240

Cao, X., D. Song, M. Yang, N. Yang, Q. Ye, D. Tao, B. Liu, R. Wu, and X. Yue. 2017. Comparative analysis of whey N-glycoproteins in human colostrum and mature milk using quantitative glycoproteomics. *J. Agric. Food Chem.* 65:10360–10367. doi:10.1021/acs.jafc.7b04381.

Cao, Y., V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero. 2018. Unconventional superconductivity in magic-angle graphene superlattices. *Nature* 556:43–50. doi:10.1038/nature26160.

Dahms, N. M., L. J. Olson, and J. P. Kim. 2008. Strategies for carbohydrate recognition by the mannose 6-phosphate receptors. *Glycobiology* 18:664–678. doi:10.1093/glycob/cwn061.

Daubenspeck, J. M., D. S. Jordan, W. Simmons, M. B. Renfrow, and K. Dybvig. 2015. General N-and O-linked glycosylation of lipoproteins in mycoplasmas and role of exogenous oligosaccharide. *PLoS One* 10:e0143362. doi:10.1371/journal.pone.0143362.

Eklund, K. K., K. Niemi, and P. T. Kovanen. 2012. Immune functions of serum amyloid A. *Crit. Rev. Immunol*. 32:335–348. doi:10.1615/critrevimmunol.v32.i4.40.

Holm, M., M. Saraswat, S. Joenväärä, A. Seppo, R. J. Looney, T. Tohmola, J. Renkonen, R. Renkonen, and K. M. Järvinen. 2022. Quantitative glycoproteomics of human milk and association with atopic disease. *PLoS One* 17:e0267967. doi:10.1371/journal.pone.0267967.

Howe, K. L., P. Achuthan, J. Allen, J. Allen, J. Alvarez-Jarreta, M. Ridwan Amode, I. M. Armean, A. G. Azov, R. Bennett, and J. Bhai. 2021. Ensembl 2021. *Nucleic Acids Res*. 49:D884–D891. doi:10.1093/nar/gkaa942.

Huang, D. W., B. T. Sherman, X. Zheng, J. Yang, T. Imamichi, R. Stephens, and R. A. Lempicki. 2009. Extracting biological meaning from large gene lists with DAVID. *Curr. Protoc. Bioinformat*. 27:11–13. doi:10.1038/nprot.2008.211

Huppertz, T., and L. W. Chia. 2021. Milk protein coagulation under gastric conditions: a review. *Int. Dairy J*. 113:104882. doi:10.1016/j.idairyj.2020.104882

Kappeler, S. R., C. Heuberger, Z. Farah, and Z. Puhan. 2004. Expression of the peptidoglycan recognition protein, PGRP, in the lactating mammary gland . *J. Dairy Sci*. 87:2660–2668. doi:10.3168/jds.s0022-0302(04)73392-5

Karav, S., J. Bruce German, C. Rouquié, A. Le Parc, and D. Barile. 2017. Studying lactoferrin N-glycosylation. *Int. J. Mol. Sci*. 18:870. doi:10.1021/acs.jafc.8b04492.

Kariya, Y., R. Kato, S. Itoh, T. Fukuda, Y. Shibukawa, N. Sanzen, K. Sekiguchi, Y. Wada, N. Kawasaki, and J. Gu. 2008. N-Glycosylation of laminin-332 regulates its biological functions: a novel function of the bisecting GlcNAc. *J. Biol. Chem*. 283:33036–33045. doi:10.1074%2Fjbc.M804526200

Kwan, S. H., W. I. Wan-Ibrahim, T. Juvarajah, S. Y. Fung, and P. S. Abdul-Rahman. 2021. Isolation and identification of O-and N-linked glycoproteins in milk from different mammalian species and their roles in biological pathways which support infant growth. *Electrophoresis* 42:233–244. doi:10.1002/elps.202000142

Li, T., B. Liu, M. H. Spalding, D. P. Weeks, and B. Yang. 2012. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol*. 30:390–392. doi:10.1038/nbt.2199

Luo, C., S. Zhao, W. Dai, N. Zheng, and J. Wang. 2018. Proteomic analysis of lysosomal membrane proteins in bovine mammary epithelial cells illuminates potential novel lysosome functions in lactation. *J. Agric. Food Chem*. 66:13041–13049. doi:10.1021/acs.jafc.8b04508

Marin, F., G. Luquet, B. Marie, and D. Medakovic. 2007. Molluscan shell proteins: primary structure, origin, and evolution. *Curr. Top. Dev. Biol*. 80:209–276. doi:10.1016/s0070-2153(07)80006-8

Matsumoto, N., M. Asano, Y. Ogura, N. Takenouchi-Ohkubo, H. Chihaya, W. Chung-Hsing, K. Ishikawa, L. Zhu, and I. Moro. 2003. Release of non-glycosylated polymeric immunoglobulin receptor protein. *Scand. J. Immunol*. 58:471–476. doi:10.1046/j.1365-3083.2003.01325.x

Mohallem, R., and U. K. Aryal. 2020. Regulators of TNFα mediated insulin resistance elucidated by quantitative proteomics. *Sci. Rep*. 10:1–15. doi:10.1038/s41598-020-77914-1

Nichols, B. L., M. Diaz-Sotomayor, S. E. Avery, S. K. Chacko, D. L. Hadsell, S. S. Baker, B. R. Hamaker, L. K. Yan, H. M. Lin, and R. Quezada-Calvillo. 2016. Milk glucosidase activity enables suckled pup starch digestion. *Mol. Cell. Pediatr*. 3:1–6. doi:10.1186/s40348-016-0032-z

O'Riordan, N., M. Kane, L. Joshi, and R. M. Hickey. 2014. Structural and functional characteristics of bovine milk protein glycosylation. *Glycobiology* 24:220–236. doi:10.1093/glycob/cwt162.

Oliveros, J. C. 2007–2015. Venny. An interactive tool for comparing lists with Venn's diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html

Patton, S., S. J. Gendler, and A. P. Spicer. 1995. The epithelial mucin, MUC1, of milk, mammary gland and other tissues. *Biochim. Biophys. Acta* 1241:407–423.

Pelaseyed, T., J. H. Bergström, J. K. Gustafsson, A. Ermund, G. M. H. Birchenough, A. Schütte, S. van der Post, F. Svensson, A. M. Rodríguez-Piñeiro, and E. E. L. Nyström. 2014. The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunol. Rev*. 260:8–20. doi:10.1111/imr.12182

Perusko, M., M. van Roest, D. Stanic-Vucinic, P. J. Simons, R. H. H. Pieters, T. C. Velickovic, and J. J. Smit. 2018. Glycation of the major milk allergen B-lactoglobulin changes its allergenicity by alterations in cellular uptake and degradation. *Mol. Nutr. Food Res*. 62:1800341. doi:10.1002/mnfr.201800341

Quesnel, H. 2011. Colostrum production by sows: variability of colostrum yield and immunoglobulin G concentrations. *Animal* 5:1546–1553. doi:10.1017/s175173111100070x.

Reily, Colin, T. J. Stewart, M. B. Renfrow, and J. Novak. 2019. Glycosylation in health and disease. *Nat. Rev. Nephrol*. 15:346–366. doi:10.1038/s41581-019-0129-4

Sargeant, T. J., B. Lloyd-Lewis, H. K. Resemann, A. Ramos-Montoya, J. Skepper, and C. J. Watson. 2014. Stat3 controls cell death during mammary gland involution by regulating uptake of milk fat globules and lysosomal membrane permeabilization. *Nat. Cell Biol*. 16:1057–1068. doi:10.1038/ncb3043.

Schroten, H., F. G. Hanisch, R. Plogmann, J. Hacker, G. Uhlenbruck, R. Nobis-Bosch, and V. Wahn. 1992. Inhibition of adhesion of S-fimbriated *Escherichia coli* to buccal epithelial cells by human milk fat globule membrane components: a novel aspect of the protective function of mucins in the nonimmunoglobulin fraction. *Infect. Immun*. 60:2893–2899. doi:10.1128/iai.60.7.2893-2899.1992.

Shhadeh, A., J. Galaski, T. Alon-Maimon, J. Fahoum, R. Wiener, D. J. Slade, O. Mandelboim, and G. Bachrach. 2021. CEACAM1 activation by CbpF-expressing *E. coli*. *Front. Cell. Infect. Microbiol*. 700. doi:10.3389/fcimb.2021.699015

Song, Y., F. Zhang, and R. J. Linhardt. 2021. Glycosaminoglycans. In: *The role of glycosylation in health and disease*. Springer; p. 103–16. doi:10.1038%2Fs41581-019-0129-4

Stelzer, G., N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T. I. Stein, R. Nudel, I. Lieder, and Y. Mazor. 2016. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformat*. 54:1–30. doi:10.1002/cpbi.5

Suarez-Trujillo, A., S. M. Luecke, L. Logan, C. Bradshaw, K. R. Stewart, R. C. Minor, C. Ramires Ferreira, and T. M. Casey. 2021. Changes in sow milk lipidome across lactation occur in fatty acyl residues of triacylglycerol and phosphatidylglycerol lipids, but not in plasma membrane phospholipids. *Animal* 15:100280. doi:10.3168/jds.s0022-0302(90)78784-x

Sudhakar, J. N., H.-H. Lu, H.-Y. Chiang, C.-S. Suen, M.-J. Hwang, S.-Y. Wu, C.-N. Shen, Y.-M. Chang, F.-A. Li, and F.-T. Liu. 2020. Lumenal galectin-9-lamp2 interaction regulates lysosome and autophagy to prevent pathogenesis in the intestine and pancreas. *Nat. Commun*. 11:1–17. doi:10.1038/s41467-020-18102-7

Theil, P. K., and W. L. Hurley. 2016. The protein component of sow colostrum and milk. In: *Gigli, I., editor. Milk proteins: from structure to biological properties and health aspects. Rijeka: IntechOpen*; p. 183–198. doi:10.5772/62841

Theil, P. K., C. Lauridsen, and H. Quesnel. 2014. Neonatal piglet survival: impact of sow nutrition around parturition on fetal glycogen deposition and production and composition of colostrum and transient milk. *Animal* 8:1021–1030. doi:10.1017/s1751731114000950

Turula, H., and C. E. Wobus. 2018. The role of the polymeric immunoglobulin receptor and secretory immunoglobulins during mucosal infection and immunity. *Viruses* 10:237. doi:10.3390/v10050237

Tyanova, S., T. Temu, and J. Cox. 2016. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc*. 11:2301–2319. doi:10.1038/nprot.2016.136

Valliere-Douglass, J. F., C. M. Eakin, A. Wallace, R. R. Ketchem, W. Wang, M. J. Treuheit, and A. Balland. 2010. Glutamine-linked and non-consensus asparagine-linked oligosaccharides present in human recombinant antibodies define novel protein glycosylation motifs. *J. Biol. Chem*. 285:16012–16022. doi:10.1074%2Fjbc.M109.096412

Wilson, N. L., L. J. Robinson, A. Donnet, L. Bovetto, N. H. Packer, and N. G. Karlsson. 2008. Glycoproteomics of milk: differences in sugar epitopes on human and bovine milk fat globule membranes. *J. Proteome Res*. 7:3687–3696. doi:10.1021/pr700793k

Yang, Y., N. Zheng, X. Zhao, Y. Zhang, R. Han, J. Yang, S. Zhao, S. Li, T. Guo, and C. Zang. 2016. Metabolomic biomarkers identify differences in milk produced by holstein cows and other minor dairy animals. *J. Proteomics* 136:174–182. doi:10.1016/j.jprot.2015.12.031

Yolken, R. H., D. S. Newburg, J. A. Peterson, S. L. Vonderfecht, E. T. Fouts, and K. Midthun. 1992. Human milk mucin inhibits rotavirus replication and prevents experimental gastroenteritis. *J. Clin. Invest*. 90:1984–1991. doi:10.1172/jci116078.

Zembroski, A. S., K. K. Buhman, and U. K. Aryal. 2021. Proteome and phosphoproteome characterization of liver in the postprandial state from diet-induced obese and lean mice. *J. Proteomics* 232:104072. doi:10.1016/j.jprot.2020.104072