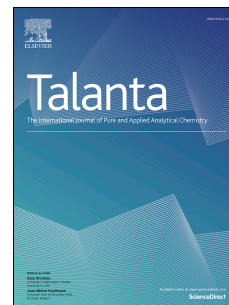


Journal Pre-proof

Systematic examination of protein extraction, proteolytic glycopeptide enrichment and MS/MS fragmentation techniques for site-specific profiling of human milk N-glycoproteins

Bum Jin Kim, David C. Dallas



PII: S0039-9140(20)31102-4

DOI: <https://doi.org/10.1016/j.talanta.2020.121811>

Reference: TAL 121811

To appear in: *Talanta*

Received Date: 7 August 2020

Revised Date: 22 October 2020

Accepted Date: 23 October 2020

Please cite this article as: B.J. Kim, D.C. Dallas, Systematic examination of protein extraction, proteolytic glycopeptide enrichment and MS/MS fragmentation techniques for site-specific profiling of human milk N-glycoproteins, *Talanta*, <https://doi.org/10.1016/j.talanta.2020.121811>.

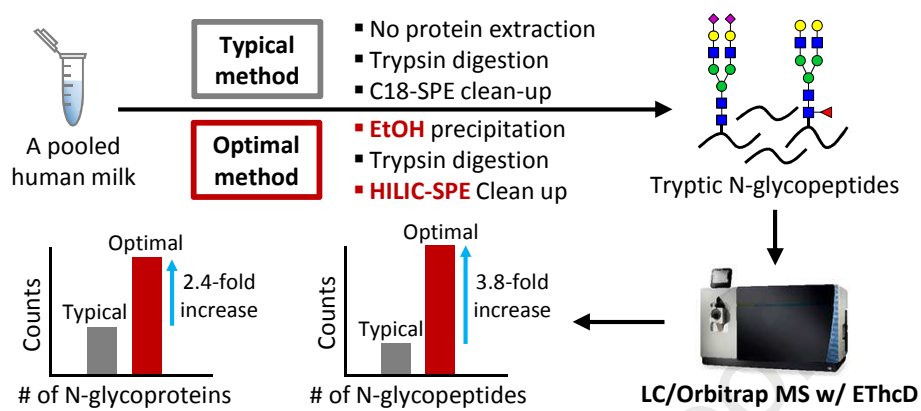
This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier B.V.

CRediT authorship contribution statement

Bum Jin Kim: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Roles/Writing - original draft, Writing - review & editing. **David C. Dallas:** Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing - review & editing.

Graphical Abstracts



**Systematic examination of protein extraction, proteolytic glycopeptide
enrichment and MS/MS fragmentation techniques for site-specific profiling of
human milk N-glycoproteins**

Bum Jin Kim, David C. Dallas*

*Nutrition Program, School of Biological and Population Health Sciences, College of Public Health and
Human Sciences, Oregon State University, Corvallis, OR 97331, USA*

*To whom correspondence should be addressed: David C. Dallas (e-mail:
Dave.Dallas@oregonstate.edu), 118G, Milam Hall, 2520 SW Campus Way, Oregon State
University, Corvallis, OR 97331, USA

ABSTRACT

Human milk contains numerous N-glycoproteins with functions that provide protection to the infant. Increasing understanding of the functional role of human milk glycoproteins within the infant requires toolsets to comprehensively profile their site-specific glycosylation patterns. However, optimized methods for site-specific glycosylation analysis across the entire human milk proteome are not available. Therefore, we performed a systematic analysis of techniques for profiling the sites and compositions of N-glycans in human milk using liquid chromatography/mass spectrometry. To decrease interference from non-target molecules, we compared techniques for protein extraction, including ethanol (EtOH) precipitation, trichloroacetic acid precipitation, molecular weight cut-off filtration and techniques for tryptic glycopeptide enrichment, including C18-, porous graphitized carbon and hydrophilic interaction liquid chromatography (HILIC)-solid phase extraction (SPE) and acetone precipitation. We compared the capacity of higher-energy collision dissociation, electron-transfer dissociation and electron-transfer/higher-energy collision dissociation (ETHcD) to produce fragment ions that would enable effective identification of the glycan composition, peptide sequence and glycosylation site. Of these methods, a combination of EtOH precipitation, HILIC-SPE and ETHcD-fragmentation was the most effective for human milk N-glycopeptide profiling. This optimized approach significantly increased the number of N-glycopeptides and precursor N-glycoproteins (293 N-glycopeptides from 29 glycoproteins) compared with a more common extraction approach with no protein extraction and C18 clean-up (77 N-glycopeptides from 12 glycoproteins). The advancement in methods for human milk N-glycoproteins provided by this study represents a key step for better understanding the function of glycoproteins within the breast milk-fed infant.

35 **Keywords**

36 Human milk N-glycoproteomics; Nano LC/Orbitrap MS; Electron-transfer/higher-energy collision

37 dissociation; Ethanol precipitation; Hydrophilic interaction chromatography; Solid phase extraction

1. Introduction

Breast milk is a complex biofluid that includes an array of proteins. These proteins provide a digestible source of amino acids as a primary nutrient for the infant. Some human milk proteins, however, survive intact or partially intact as milk traverses the infant digestive system, and these surviving proteins can exert bioactivities that support infant health [1-4]. Human milk contains about 10 g L⁻¹ of proteins [5], most of which are glycosylated with N- and/or O-linked oligosaccharides (over 50% of milk proteins by concentration) [6]. As the human digestive system does not secrete glycosidases that can cleave the specific bond structures of protein-linked glycans [7], these glycoproteins are more likely than non-glycosylated proteins to survive intact in the digestive system to exert their functions. Some milk glycoproteins have anti-pathogenic actions that help decrease pathogen infection risk [8]. In many instances, the glycan component of the protein acts as a target decoy that prevents the binding of pathogens to gastrointestinal epithelial cells, a first step in enteric infection [9]. The glycan components of glycoproteins may also provide a substrate for the growth of commensal bacteria in the intestine [10].

Comprehensive profiling of human milk N-glycoproteins is essential to understanding their diverse functions within the neonate. Several studies [11-14] profiled human milk N-glycoproteins via various combinations of enzyme digestion, enrichment techniques, and liquid chromatography (LC)/mass spectrometry (MS) and MS/MS, yet none provided a method for identification of the glycan composition, glycosylation site and protein of origin for the overall human milk proteome. For instance, we previously applied porous graphitized carbon (PGC)-nano LC chip/Q-TOF MS to identify the composition of N-linked glycans in human milk after removing them from proteins using PNGase F [11]. Picariello et al. [12] extracted human milk proteins via strong acid/acetone precipitation, enriched the tryptic digested N-glycopeptides using hydrophilic interaction liquid chromatography (HILIC)-solid phase extraction (SPE),

61 removed the N-linked glycans with PNGase F treatment and characterized the deglycosylated peptides
62 by MS and MS/MS analysis. Cao et al. [13] used centrifugation to isolate whey, tryptic digestion and C18-
63 SPE to extract the resulting tryptic peptides, lectin affinity chromatography to isolate glycopeptides,
64 PNGase F to deglycosylate the glycopeptides and LC/MS/MS to identify previously glycosylated peptides.
65 These PNGase F-based methods identified either the glycan compositions or the sites of glycosylation
66 but not the glycan composition at specific sites. Huang et al. [14] performed site-specific profiling of N-
67 glycopeptides from isolated immunoglobulin A after non-specific pronase digestion. However, pronase
68 digestion leads to short peptide sequences (1-3 amino acids) that are not long enough to identify the
69 parent protein sequence in complex biological samples. Thus, developing a method for site-specific
70 profiling across the whole N-glycoproteome was needed.

71 Extraction of human milk proteins is essential to minimize interference from other abundant
72 components, such as lipids, carbohydrates and endogenous peptides, in subsequent protease digestion
73 and LC/MS analysis. Ethanol (EtOH) precipitation was previously used to purify and precipitate intact
74 proteins from human milk and remove de-glycosylated proteins after N-glycan release using PNGase F
75 [11]. Increasing the percentage of EtOH, which is more non-polar than the water that comprises the
76 majority of milk, decreases protein solubility and increases precipitation. Strong acid (trichloroacetic
77 acid, TCA) was used to remove proteins and extract naturally occurring peptides in human milk [15]. The
78 strong acid condition triggers proteins to strongly interact with each other, rather than the solution
79 medium, and to precipitate out of solution. Molecular weight cut-off (MWCO) filtration was also used to
80 extract proteins from milk [16-17]. The molecular weights of casein and whey proteins in human milk
81 are over 10 kDa (e.g., caseins, 19–24 kDa; α -lactalbumin, 14 kDa; lactoferrin (LF), 80 kDa; and secretory
82 Immunoglobulin A (sIgA), 150 kDa) [6], therefore, theoretically, a 10-kDa MWCO filter should effectively
83 extract proteins from human milk.

For bottom-up glycoproteomics, after protease digestion, enrichment and purification of glycopeptides are required to avoid interference with binding and elution from the LC column and MS signal suppression caused by co-eluting compounds, including compounds inherent to the sample, substances added during sample preparation and co-digested non-glycosylated peptides. C18 [18], PGC [19], HILIC [20] and liquid-liquid extraction (LLE) using acetone [21] were used previously to enrich glycopeptides after protease digestion of biological fluids. The C18-based glycopeptide enrichment method is based on hydrophobic interaction between non-polar peptide moieties and the C18 chains of the stationary phase [22]. Glycopeptide binding to the PGC column is affected by polar interactions between the glycan moiety and the graphite solid phase, the hydrophobicity of the eluent and planar-planar interactions between the glycan moiety and the graphite surface [23]. The theory behind the HILIC-based glycopeptide enrichment method is that the glycan moiety of glycopeptides will allow their greater retention on the polar stationary phase (via dipole-dipole, ionic and hydrogen interactions) than retention of non-glycosylated peptides [24]. The LLE method using acetone is based on the lower solubility of the generally more polar glycopeptides in the more non-polar acetone compared with non-glycosylated peptides [25].

Though a variety of methods exist for protein extraction from milk and proteolytic glycopeptide enrichment, no systematic comparison has been performed to evaluate which method is most effective for extraction of human milk proteins and proteolytic glycopeptides.

To disambiguate between potential glycopeptide compositions based on intact mass alone, MS/MS fragmentation is needed. Collision-induced dissociation and higher-energy collision dissociation (HCD) fragmentation of glycopeptides digested by site-specific proteolysis such as trypsin and chymotrypsin digestion produce mainly glycan cleavages, limiting the potential to verify the peptide sequence and glycosylation site. Electron-transfer dissociation (ETD) mainly induces fragmentation of the peptide

backbone, limiting the potential to verify the glycan composition [26]. Electron-transfer/higher-energy collision dissociation (ETHcD) is a hybrid fragmentation technique that combines ETD and HCD. In this method, precursor ions are fragmented using ETD and the resulting product ions, as well as the remaining precursor ions, undergo HCD fragmentation. Thus, ETHcD provides both peptide product ions (c- and z-) and glycan product ions (B- and Y-) derived from the glycopeptide within a single tandem spectrum. ETHcD fragmentation-based N-glycoproteomics allows simultaneous determination of glycan composition, peptide sequence and glycosylation site(s) of the N-glycopeptide from a single MS/MS scan [27], therefore, this emerging fragmentation technique could enable the simultaneous site-specific profiling of the array of human milk glycoproteins.

We systematically examined human milk protein extraction methods, proteolytic N-glycopeptide enrichment methods and tandem MS fragmentation techniques to identify an optimal approach to analyze human milk N-glycoprotein. ETHcD, HCD and ETD were compared for glycopeptide identification. EtOH precipitation, TCA precipitation and MWCO filtration were compared for human milk protein extraction. C18-, PGC- and HILIC-based SPE and acetone precipitation were compared for N-glycopeptide enrichment capacity.

2. Material and methods

2.1. Human milk samples

A pooled human milk sample (from four individuals) was donated by the Northwest Mother's Milk Bank. Milk donors provided informed consent for their milk to be used in research. The milk was centrifuged at 4,000 ×g at 4 °C for 30 min and the infranate (between the upper lipid (cream) layer and the sediment (somatic milk cells)) was collected. Lipid and cell removal prevents their interference

during ample preparation steps and MS analysis. Collected skimmed milk samples were frozen at -80°C until further sample preparation steps were performed.

2.2. Human milk protein extraction

Aliquots ($30\ \mu\text{L}$) of the pooled skimmed milk were used to compare three methods for human milk protein extraction.

2.2.1. EtOH precipitation

Skimmed milk was mixed with $150\ \mu\text{L}$ of chilled (-20°C) EtOH (Sigma Aldrich, St. Louis, MO, USA) and placed at -20°C for 1 h. Samples were centrifuged at $16,000\times g$ at 4°C for 20 min to precipitate the proteins. The supernatant was collected by pipette into a new tube. The supernatant and pellet fractions were dried by vacuum centrifugation. Dried samples were dissolved in $100\ \mu\text{L}$ of $50\ \text{mmol L}^{-1}$ ammonium bicarbonate before trypsin digestion.

2.2.2. TCA precipitation

Thirty microliters of 24% TCA (Sigma Aldrich) were added to skimmed milk and centrifuged at $1,800\times g$ at 4°C for 10 min. Supernatants were collected by pipette into a new tube and supernatants and pellets were dried separately by vacuum centrifugation. Dried samples were dissolved in $100\ \mu\text{L}$ of $50\ \text{mmol L}^{-1}$ ammonium bicarbonate before trypsin digestion.

2.2.3. MWCO-based centrifugal filtration

Proteins in skimmed milk were concentrated using a centrifugal filtration device (10-kDa MWCO, Sigma Aldrich). A filter was cleaned with nanopure water and samples were loaded in the filter. These were washed with nanopure water and concentrated 5 times by centrifugation at $14,000\times g$ for 10 min. The retentate including high-molecular weight materials was transferred to a collection tube. The

permeate and retentate were collected separately and dried by vacuum centrifugation. Dried samples were dissolved in 100 μL of 50 mmol L^{-1} ammonium bicarbonate before trypsin digestion.

2.3. Trypsin digestion

Thirty microliters of skimmed milk were mixed with 100 μL of 50 mmol L^{-1} ammonium bicarbonate (Thermo Scientific, Waltham, MA, USA), followed by the addition of 2 μL of 550 mmol L^{-1} dithiothreitol (Promega, Madison, WI, USA), and incubated at 50 °C for 50 min to reduce disulfide bonds. After incubation, 4 μL of 450 mmol L^{-1} iodoacetamide (Sigma Aldrich) were added and incubated at room temperature for 60 min in the dark to alkylate the thiol groups. To digest the denatured proteins into peptides, 2 μL of 1 μg μL^{-1} trypsin (Thermo Scientific) were added and mixtures were incubated at 37 °C with shaking at 300 rpm overnight.

2.4. Tryptic human milk n-glycopeptide enrichment

Aliquots (30 μL) of skimmed milk that were digested by trypsin without protein extraction were used to explore four methods for tryptic N-glycopeptide enrichment.

2.4.1. C18-SPE

C18 cartridges (5-mL tube volume, 500 mg bed weight, 45 μm particle size, Sigma Aldrich) were washed and reconditioned with 5 mL of 80% acetonitrile (ACN, Thermo Scientific), 0.1% trifluoroacetic acid (TFA, Sigma Aldrich) and 5 mL of nanopure water prior to loading the samples. After sample loading, the cartridges were washed with 5 mL of nanopure water to remove salts and interfering substances. A 5-mL aliquot of 80% ACN, 0.1% TFA was added to elute the glycopeptides. Each eluate was dried by vacuum centrifugation.

2.4.2. PGC-SPE

PGC cartridges (5-mL tube volume, 500-mg bed weight, 120–400-mesh particle size, Sigma Aldrich) were conditioned and samples were loaded, washed and eluted as described above for C18 cartridges.

2.4.3. HILIC-SPE

A HILIC cartridge (3-mL tube volume, 200-mg bed weight, 50- μ m particle size, Hilicon, Sweden) was washed and reconditioned with 3 mL of nanopure water with 1% TFA and 3 mL of 90% ACN prior to sample loading. A 1-mL aliquot of ACN was added and a sample dissolved in 100 μ L of nanopure water was loaded in the cartridge. The sample was washed with 3 mL of 90% ACN to remove salts and interfering substances. A 3-mL aliquot of nanopure water with 1% TFA was added to elute the glycopeptides. Eluates were dried by vacuum centrifugation.

2.4.4. Acetone precipitation

A glycopeptide enrichment method using acetone (Sigma Aldrich) precipitation was performed as described previously [25]. Briefly, a 5-fold volume of chilled acetone was added to a tryptic digested sample and placed at -20°C overnight. The sample was centrifuged at $12,000 \times g$ at 20°C for 10 min. The liquid layer was discarded and the precipitated glycopeptides were collected, followed by vacuum centrifugation drying.

2.5. LC/MS and MS/MS analysis of tryptic human milk N-glycopeptides

Glycopeptides were analyzed using a Waters nanoACQUITY UPLC (Waters, Milford, MA) with an Orbitrap FusionTM LumosTM TribridTM mass spectrometer (Thermo Scientific, Waltham, MA). Dried samples were reconstituted with 30 μ L of nanopure water and diluted 10-fold. One microliter of each sample was loaded onto a C18 180 μ m $\times$ 20 mm, 5 μ m bead nanoAcquity UPLC trap column (Waters) for enrichment and desalting, and separated with a 100 μ m $\times$ 100 mm, 1.7 μ m bead Acquity UPLC Peptide BEH C18 column (Waters) over 60 min. Glycopeptides were separated at a flowrate of $0.5 \mu\text{L min}^{-1}$ with

a gradient elution using solvent A (100% nanopure water with 0.1% formic acid) and solvent B (100% ACN with 0.1% formic acid): 3% to 11.5% B, 0 min to 10 min; 11.5% to 20% B, 10 min to 31 min; 20% to 30% B, 32 min to 36 min; 30% to 95% B, 36 min to 45 min; 95% B, 45 min to 54.5 min, 95–3% B over 0.5 min then the column was re-equilibrated with 97% A for 5 min.

Glycopeptides were ionized with an electrospray voltage of 2,350 V and an ion-transfer tube temperature of 300 °C. Full scan MS spectra were acquired in positive ionization mode over an m/z range of 300–2,000 with a resolution of 60,000. The automatic gain control target was set to 4.0×10^5 , with a maximum injection time of 50 ms. The MS cycle time was set to 3 s.

Three tandem MS techniques—HCD, ETD and ETHcD—were tested with data dependent analysis to compare fragmentation patterns of glycopeptides in MS/MS spectra acquired from each fragmentation type. Following an MS scan, precursor compounds were automatically selected for MS/MS analysis by the acquisition software based on the following criteria: ion-intensity threshold 5.0×10^4 , charge state 2–8 and exclusion time 60 s. For HCD, normalized collision energy was set to 25% for all precursor ions. The ETD reaction times were set depending on charge state (2+, 130 ms; 3+, 70 ms; 4+, 50 ms; 5+, 40 ms; 6+ to 8+, 20 ms) for ETD and ETHcD. Supplemental HCD activation was performed with 25% of collision energy for ETHcD. All MS/MS spectra were acquired in the positive ionization mode over an m/z range of 300–2,000 by the Orbitrap at resolution of 30,000. The automatic gain control target was set to 5.0×10^4 for all fragmentation techniques.

2.6. Data processing

Identification of glycopeptides and peptides based on database searching using an in-house human milk protein sequence database ($n = 432$), which can be accessed at <http://www.dallaslab.org/resources>, was directly performed using Byonic v.3.4.0 (Protein Metrix Inc., Cupertino, CA, USA) from raw files

without processing. Cleavage sites were set to C-terminal arginine (R) and lysine (K), and digestion specificity was selected as fully specific with a maximum of two missed cleavages. The precursor mass tolerance was set to 10 ppm with fragment mass tolerance of 20 ppm. Potential modifications included phosphorylation of serine and threonine and oxidation of methionine. Carbamidomethylation of cysteine was specified as a fixed modification. N-glycan searches used a database of 142 glycans (listed in Table S1) comprised of the human milk N-glycans identified in our previous publication [11] and common human N-glycans provided in the Byonic software. Outputs of proteins were filtered at a 1% false discovery rate as calculated by Byonic. Outputs of peptides were automatically filtered within at 0–5% false discovery rate based on the protein filtering outputs (automatic score cut-off). As an additional filter to ensure quality data, glycopeptide-spectra matches with PEP 2D (protein-aware posterior error probability) < 0.01 (is equal to |Log Prob| > 2) and Score > 75 were retained for the reported results. These filtering conditions were based on previous studies [27–29]. Glycopeptide assignments were further confirmed via manual analysis of each tandem MS spectrum. Of the retained glycopeptides, those with a tandem MS spectrum with insufficient fragment ions to identify the composition of the precursor ion were excluded in the final report.

Byonic search results were merged with using Byologic (Protein Metrix Inc., Cupertino, CA). For comparing MS/MS results acquired from HCD, ETD and ETHcD fragmentation, the merged result was extracted and investigated from Byologic without additional in-silico processing. For the comparison of protein enrichment methods, glycopeptide extraction methods and typical and optimal approaches, in-silico processing using "Add missing via existing peptides" was performed in Byologic. This process searched the MS analysis results for each sample against the complete library of MS/MS-identified glycopeptides from all samples and added in glycopeptide results for compounds that were not identified by MS/MS results in the particular sample but were present according to the MS data. This

extraction used a 10 ppm mass error threshold and only extracted compounds within the retention time window and charge state of the MS/MS-identified compound.

Glycopeptides detected in multiple charge states were grouped into a single glycopeptide for counting the number of N-glycopeptides identified in a sample. Abundance reported in the results is a measure of the area under the curve of the eluted peak based on ion intensity (based on extracted ion chromatograms). All abundance information provided in this manuscript is based on relative quantification. Relative abundances of each N-glycopeptide were calculated by the following equation: chromatogram peak area of individual glycopeptide $\div$ chromatogram peak area of all glycopeptides $\times$ 100. All LC/MS analyses were performed in technical duplicates and the relative abundances of peptides were averaged in the reported data. Pearson correlations coefficients (R) were applied to evaluate variations within instrument measurements, variations within multiple extractions for a single investigator ($n = 2$) and variations between extractions performed by multiple investigators ($n = 3$).

3. Results and discussion

An overview of the workflow and variables tested in this study are shown in Figure 1. We evaluated MS/MS results acquired from EThcD-fragmentation technique through comparison of HCD- and ETD-derived MS/MS spectra, the efficiency of human milk protein extraction using EtOH precipitation, TCA precipitation and filtration using the 10-kDa MWCO filter and the efficiency of tryptic N-glycopeptide enrichment using C18-, PGC- and HILIC-based SPE and acetone precipitation. All samples were measured in duplicate and average abundances obtained in each measurement were used to examine their MS profiles.

3.1. Evaluation of N-glycopeptide MS/MS fragmentation methods

To explore differences in MS/MS analysis of human milk N-glycopeptides using EThcD, HCD and ETD, skimmed human milk proteins were digested using trypsin without protein pre-extraction and tryptic glycopeptides were enriched via C18-SPE. The sample was analyzed by HCD-, ETD- and EThcD-based full MS/MS scans separately. As examples of the detailed results, three MS/MS spectra recorded from the same precursor with each fragmentation method confirmed the peptide sequence, glycan composition and glycosylation site. The interpretation of tandem MS spectra of the sialylated glycopeptide, TPLTANITK with N-acetyl hexosamine (HexNAc)₅–hexose (Hex)₅–fucose (Fuc)₁–N-acetyl neuraminic acid (NeuAc)₁ (calculated m/z 1074.4640, $z=3$) and the neutral glycopeptide, TPLTANITK with HexNAc₅Hex₄Fuc₁ (calculated m/z 923.4166, $z=3$) are shown in Figures 2 and 3. These glycopeptides derived from slgA2 at glycosylation site ²⁰⁵N.

Oxonium ions from fragmentation of the glycan moiety of the glycopeptide were present and abundant at the lower mass range in the HCD-tandem MS spectra (Figures 2a and 3a). For example, these glycopeptide precursor ions both generated representative oxonium ions—138 (HexNAc -2H₂O - CH₂O), 168 (HexNAc -2H₂O), 204 (HexNAc) and 366 (HexNAcHex)—of complex type glycans, and the sialylated glycopeptide additionally included the diagnostic fragment ions of sialic acid such as 274 (NeuAc -18), 292 (NeuAc) and 657 (HexNAcHexNeuAc). HCD fragmentation induced the cleavage of glycosidic bonds without peptide fragmentation, allowing determination of glycan composition and structure. Singly charged ions at m/z 1307 (HexNAcFuc + Peptide) and 1510 (HexNAc₂Fuc + Peptide) were indicative of core fucosylation and m/z 1729 (HexNAc₃Hex + peptide) and 1875 (HexNAc₃HexFuc + peptide) indicated complex type N-glycan containing bisecting N-acetyl glucosamine attached to the mannose of the core. HCD provided some fragment ions from the peptide backbone such as y1 and y2, but the peak intensities were low and the fragments were too few to assign the entire peptide sequence.

Whereas HCD resulted in spectra dominated by fragment ions from glycan degradation, the ETD spectrum shows the diagnostic fragment ions of the peptide backbone as annotated in Figures 2b and 3b. In ETD spectra, most fragment ions were singly and doubly charged c- and z-peptide fragment ions. The ETD spectra also included intact glycopeptide ions with and without the neutral losses of 42 Da from the acetyl moiety of N-acetylhexosamine [30] and 17 Da from the ammonia (NH₃) moiety of the N-terminal amine [31] at multiple charge states. Though ETD fragmentations provided the information required for glycosylation site determination, they did not produce any diagnostic ions useful for interpreting the glycan composition and structure.

ETHcD spectra of each precursor glycopeptide (Figures 2c and 3c) show diagnostic fragment ions for assigning both the glycan structure and peptide sequence. Representative oxonium ions of glycans such as m/z 168, 204, 274, 292 and 366 were also observed in high abundance in the ETHcD spectra. Y-ions displayed in ETHcD spectra, such as m/z 1307 and 1510 indicating core fucosylation and m/z 865 ($z=2$, HexNAc₃Hex + peptide), and 938 ($z=2$, HexNAc₃HexFuc + peptide) corresponding to Y ions with bisecting N-acetyl glucosamine provided structural information to elucidate structures of precursor glycopeptides. These B- and Y-ions yielded by glycan fragmentation mainly matched those observed in the HCD fragmentation spectra. Similar to the ETD fragmentation spectra, c- and z-ions—corresponding to peptide backbone fragmentation and the precursor glycopeptide with and without NH₃ and acetyl moiety losses—were present in the ETHcD spectra. These findings indicate that ETHcD produced a wide range of both glycan fragments and peptide backbone fragments that allowed confident glycopeptide identification.

ETHcD-based fragmentation allowed determination of the exact composition of glycopeptides when multiple compositions were possible for a single m/z value. One glycopeptide (measured m/z 1048.1060, $z=3$) were assigned as both HexNAc₅Hex₄NeuAc₁ + ENISDPTSPLR (calculated m/z 1048.1028,

Immunoglobulin J chain) and HexNAc₅Hex₄NeuAc₁ + VEIANDQGNR (calculated m/z 1048.1065, Heat shock cognate 71 kDa protein) within 5 ppm error tolerance (Figure S1a). Another glycopeptide (measured m/z 910.7328, $z=3$) had two possible glycopeptide compositions as HexNAc₅Hex₃ + ENISDPTSPLR (calculated m/z 910.7289, Immunoglobulin J chain) and HexNAc₅Hex₃ + VEIANDQGNR (calculated m/z 910.7326, Heat shock cognate 71 kDa protein) (Figure S1b). In both interpretations, the assigned glycan composition was identical, but the peptides sequences differed based on accurate mass profiling. Therefore, tandem MS spectra peaks that provided peptide sequence information were necessary to determine the accurate glycopeptide composition. Spectral annotation identified products corresponding to z -ions of the peptide backbones matching to "ENISDPTSPLR", not "VEIANDQGNR." Indeed, Immunoglobulin J chain includes the N-glycosylation site at Asn 71, whereas there is no N-glycolyl modification in Heat shock cognate 71 kDa protein according to previous reports in Uniprot (P01591 and P11142).

The effectiveness of HCD-, ETD- and ETHcD-based MS/MS analysis in identifying N-glycopeptides was evaluated using the number of MS/MS spectra identified (Table 1). The total number of tandem MS spectra extracted from HCD-, ETD- and ETHcD-raw files was 12,568, 7,857 and 7,667, respectively. Of all HCD-, ETD- and ETHcD-tandem MS spectra, 1,156, 654 and 757 peptide spectral matches and 54, 17 and 36 N-glycopeptide spectral matches were identified by Byonic searching. After filtering these N-glycopeptides as described in the Experimental section (Score > 75 and PEP-2D < 0.01), 20, 9 and 25 N-glycopeptide-spectral matches remained for HCD-, ETD- and ETHcD- tandem MS spectra, respectively. Though ETHcD showed relatively lower MS/MS events than HCD and ETD, the number of N-glycopeptide spectral matches that remained after filtering was higher than for HCD and ETD due to their higher scores and lower PEP-2D values (Figure S2).

Yu et al. [27] reported that EThcD provided more informative spectra and allowed more confident determination of human serum N-glycopeptides compared with HCD and ETD. Though this finding was demonstrated with human serum, no previous study had systematically examined the capacity of EThcD to analyze the complex array of glycoproteins in human milk. In our study, we also found that EThcD provided more informative spectra for glycopeptide identification than HCD or ETD. The data of Yu et al. [27] demonstrated a more pronounced advantage of EThcD over HCD in terms of the number of glycopeptides identified compared with our data. This difference is likely due to the longer ETD reaction times used in our study, which lowered the duty cycle and resulted in fewer MS/MS spectra.

Implementing EThcD fragmentation in the Orbitrap MS can lead to improved identification of both the glycan composition and peptide sequence of human milk N-glycopeptides. However, there is a need to optimize instrumental conditions. ETD reaction times should be minimized to allow quality fragmentation while increasing the number of MS/MS spectra collected. In the subsequent experiments herein comparing protein enrichment and glycopeptide extraction methods, EThcD-based MS/MS analysis was performed and in-silico processing was performed to identify glycopeptides based on MS-level results identified by MS/MS in other samples based on mass and retention time matching.

3.2. Human milk protein extraction—comparison of EtOH precipitation, TCA precipitation and MWCO filtration

Human milk contains various components such as lipids, oligosaccharides, proteins and peptides. These components can interfere with the retention of glycopeptides on the stationary phase of the LC column and cause low ionization efficiency and signal suppression of glycopeptides during MS analysis. To extract human milk proteins from complex milk components, we employed liquid precipitation methods using EtOH, and TCA and filtration using a 10-kDa MWCO filter. In our previous studies profiling

the endogenous peptides in human milk, 24% TCA was used to precipitate intact proteins [32]. Chilled EtOH precipitation is commonly used for precipitating deglycosylated proteins and non-glycosylated proteins after N-glycan release using enzyme digestion from body fluids such as serum [33] and saliva [34]. Molecular weights of the major human milk proteins are over 10-kDa [35]; meanwhile, other major components of human milk such as lactose, most oligosaccharides and most endogenous peptides are less than 10-kDa [36-37]. For these reasons, we selected strong acid precipitation using 24% TCA, chilled EtOH precipitation and 10-kDa MWCO filtration for protein extraction from human milk.

Proteins were extracted by three methods from the pooled human milk samples. We collected both the pellet/retentate and the supernatant/permeate after TCA and EtOH precipitation and filtration in order to monitor the loss of proteins during protein extraction. Trypsin digestion was performed in each fraction and tryptic glycopeptides/peptides were cleaned by C18-SPE prior to MS analysis. Each extraction method was evaluated in terms of the number of detected N-glycopeptides and total abundances of N-glycopeptides identified.

Counts of total tryptic N-glycopeptides in the pooled human milk sample are shown in Figure 4a. A total of 113, 53 and 88 N-glycopeptides were extracted from human milk proteins by EtOH and TCA precipitations and 10-kDa MWCO filtration, respectively. Comparison of the relative peak area values of total glycopeptides from each protein extraction method indicates that EtOH precipitation yielded more glycopeptides than strong acid precipitation and filtration (Figure 4b). Protein loss from human milk using each protein extraction method was investigated by comparing abundances of total N-glycopeptides in the pellet/retentate and the supernatant/permeate. Over 99% of N-glycopeptides were present in pellet after EtOH precipitation and the retentate after 10-kDa MWCO filtration (Figure S3). In contrast, only 26% of the total abundance of N-glycopeptides were present in the pellet compared with the supernatant after TCA precipitation. TCA precipitation showed the lowest counts and abundances of

N-glycopeptides compared with the other protein extraction methods. Additionally, the total peak abundances of N-glycopeptides from the pellet after TCA precipitation were lower than those from the supernatant portion. Possibly, the proteins precipitated using TCA were so strongly aggregated with each other [38] that the efficiency of re-dissolving is low and limits trypsin digestion.

EtOH precipitation and filtration were more efficient for human milk protein extraction than for TCA precipitation. EtOH precipitation of the resulting pellet yielded higher counts and total abundance of N-glycopeptides than in the retentate produced by MWCO filtration. The filter membrane may be clogged by various components in milk such as lactose, oligosaccharides, lipids, cells and the proteins themselves, resulting in incomplete filtration [39]. Thus, other biological components that remained in the retentate could interfere with the subsequent sample preparation steps and MS analysis, reducing the number and abundance of detected glycopeptides.

Concentrations of human milk proteins extracted using each method were quantified by the bicinchoninic acid (BCA)-based protein assay to evaluate the efficiency of EtOH and TCA precipitations and filtration for the extent of protein extraction (measured concentrations and calibration curves in Figure S4). Compared with the total protein concentration in the milk sample without protein extraction, as measured by BCA-assay, 55%, 40% and 5% of the total protein was in the EtOH precipitated, filtered and acid precipitated samples, respectively. There was a loss of BCA-detectable protein after protein extraction regardless of the extraction methods used. This may be due to incomplete re-dissolving of aggregated proteins. Nevertheless, human milk protein extraction using EtOH precipitation yielded the highest protein recovery among the methods tested.

Though our study mainly focused on analysis of intact N-glycoproteins in human milk, endogenous peptides present in supernatant/permeate separated from pellet/retentate were analyzed to assess which protein extraction method most effectively separated endogenous peptides from intact proteins

in human milk. Endogenous peptides were most abundant (in both counts and abundances) in TCA precipitation supernatants (Figure S5). EtOH precipitation more efficiently enriched endogenous peptides than filtration but yielded relatively lower enrichment than TCA precipitation. In particular, the extents of counts and abundances of phosphopeptides were higher in the TCA supernatant than in the EtOH supernatant. The phosphate groups of the phosphopeptides possess a negative charge under the acidity of 24% TCA, therefore increasing their overall hydrophilicity and extracting higher levels of phosphopeptides compared with the other methods [40]. Endogenous peptides were lower in the filtration permeate than those yielded by the other extraction methods, likely due to clogging of the filter membrane by complex human milk components, preventing complete transfer of endogenous peptides through the membrane pores.

3.3. Glycopeptide enrichment – comparison of C18-, PGC- and HILIC-SPE and acetone precipitation

Enrichment and purification of tryptic glycopeptides is necessary prior to MS analysis to increase sensitivity for glycopeptide detection and avoid ion suppression induced by the co-presence of non-target molecules. For this comparison of glycopeptide enrichment strategies, human milk proteins were digested using trypsin without pre-extraction of proteins, and tryptic N-glycopeptides were enriched and purified using SPE with three different cartridges—C18, PGC and HILIC. These enrichment techniques have been widely used for enriching proteolytic glycopeptides [19, 41-42]. A recently reported LLE method using acetone for separating glycopeptides from co-present peptides [43] was employed to enrich human milk N-glycopeptides. The eluate/pellet of each technique was analyzed via EThcD-based nano LC/Orbitrap MS. Counts and total abundances (log-scale) of identified N-glycopeptides were compared to determine the most effective method for enriching tryptic human milk N-glycopeptides.

Eluates from C18- and PGC-SPE had similar counts ($n = 79$ and 81) of tryptic N-glycopeptides (see the line chart in Figure 5a). The counts of N-glycopeptides identified after HILIC-SPE and acetone-LLE ($n = 192$ and 165) were significantly higher than after C18- and PGC-SPE. Counts of N-glycoproteins (based on the identified N-glycopeptides) were also two-fold higher after HILIC-SPE and acetone-LLE than after C18- and PGC-SPE (Figure S6). The total abundance of tryptic N-glycopeptides was significantly higher after HILIC-SPE than after C18- and PGC-SPE and acetone-LLE (see the bar graph in Figure 5a). Though we performed the sample preparation in this stage, no protein extraction was involved prior to trypsin digestion. Thus, other milk compounds such as carbohydrates, lipids and cells also were loaded into the SPE cartridges. These matrix compounds may have interfered with the retention and elution of glycopeptides from the SPE cartridges.

Acetone-based LLE provided higher enrichment of N-glycopeptides (both counts and abundances) than C18- and PGC-SPE but lower enrichment (in terms of abundance) than HILIC-SPE. The lower abundance of glycopeptides identified after acetone precipitation could have resulted from incomplete re-dissolving of the pellet. Moreover, the precipitation of the glycopeptides may need to be optimized. Mancera-Artu et al. [44] reported that glycopeptides bearing high sialic acids and structurally large glycan moieties were highly enriched at a five-fold acetone:sample water ratio, whereas glycopeptide enrichment for glycopeptides with smaller glycan components and longer peptide chains (> 15 amino acids) required a higher proportion ($> eight-fold$) of acetone. However, more non-glycosylated peptides also can be precipitated with glycopeptides as the proportion of acetone increases. Thus, optimization of the proportion of acetone to sample water, considering glycan/peptide sizes of tryptic N-glycopeptides, is needed to maximize the precipitation of N-glycopeptides and minimize co-precipitation of peptides. Among the methods tested, HILIC-SPE provided the best enrichment, as demonstrated by the counts and abundance of glycopeptides detected.

Because proteolytic glycopeptides are in relatively low abundance and have low ionization efficiency compared with non-glycosylated peptides, tryptic peptides co-eluted/precipitated with glycopeptides in SPE and LLE could suppress the signal of glycopeptides in MS analysis. Thus, separation of peptides from glycopeptides during SPE and LLE is necessary to improve glycopeptide identification. To evaluate the extent of the separation of glycopeptides and peptides in each enrichment method, we compared the counts and total abundances of tryptic peptides that were enriched together with N-glycopeptides in samples. The counts of peptides were relatively lower after C18- and PGC-SPE ($n = 873$ and 856) than after HILIC-SPE and acetone-LLE ($n = 1,243$ and $1,087$) (Figure 5b). The abundance of peptides was somewhat lower after acetone-LLE than after C18-, PGC- and HILIC-SPE. Tryptic peptides were co-present with N-glycopeptides in significant abundance in the sample regardless of the glycopeptide extraction method tested. These peptide profiles indicated that none of the tested methods, including C18-, PGC- and HILIC-SPE and acetone-LLE, separated the tryptic N-glycopeptides from peptides released from human milk protein digestion. Thus, a method for separating tryptic peptides from glycopeptides remains elusive in human milk glycoproteomics.

3.4. Determination of the optimized method for human milk N-glycoprotein analysis

In our comparative study of fragmentation techniques, EThcD-based fragmentation enabled assignment of glycosylation site, glycan composition and peptide sequence of human milk N-glycopeptides within a single MS/MS scan. This detailed fragmentation information enables disambiguation between multiple compositional possibilities for glycopeptide structure. Among the protein extraction and glycopeptide enrichment methods compared, EtOH precipitation and HILIC-SPE most efficiently extracted proteins in human milk and enriched tryptic N-glycopeptides, respectively. For these reasons, we determined that the combination of EtOH precipitation, HILIC-SPE and EThcD-based

MS analysis were optimal for profiling human milk N-glycoproteins. Indeed, this optimal method significantly improved human milk N-glycopeptide profiling compared with the “typical” method using only C18-SPE clean-up without protein extraction prior to trypsin digestion.

Compared with the “typical” extraction method, this optimized method enabled identification of a higher number of N-glycopeptides (246 vs. 62) and a higher number of originating glycoproteins (29 vs. 11) (Figure 6a). All N-glycopeptides identified by the typical and optimal methods are listed in Table S2 and S3, respectively. Of unique N-glycopeptides identified only in the sample prepared via the optimal method, tandem MS spectra of 116 abundant N-glycopeptides (making up 95% of the total abundance of N-glycopeptides unique to the optimal method)) are shown in Figure S7.

To investigate potential differences in the N-glycopeptide profiles between the typical and optimal methods, N-glycopeptides were grouped according to biosynthetic classes; high mannose, complex/hybrid type without fucose and sialic acid, complex/hybrid type with fucose, complex/hybrid type with sialic acid and complex/hybrid type with fucose and sialic acid. N-glycopeptides with fucose and/or sialic acid represented the majority of glycopeptides identified by both the typical (48 of 62 N-glycopeptides) and optimal (213 of 246 N-glycopeptides) approaches (Figure 6b). In terms of relative abundance, N-glycopeptides consisting only of Hex and HexNAc were relatively more abundant (44%) in the sample treated with the typical method than in the optimized method (10%) (Figure 6c). N-glycopeptides with both fucose and sialic acid were more abundant (50%) after the optimal method than after the typical method (19%). Nwosu et al. [45] characterized N-glycans enzymatically released from a pooled human milk sample by PGC-chip LC/Q-TOF MS analysis. According to their results, 6%, 37% and 57% of relative abundances of all human milk N-glycans were high mannose, neutral N-glycans and sialylated N-glycans, respectively. N-Glycopeptide profiles using the optimal method had significantly similar relative distributions in each glycan group; 9% high mannose, 34% neutral glycans and 57%

sialylated glycans. Additionally, relative abundance of total fucosylated N-glycopeptides (76%) was consistent with their previous observation that 74% of human milk N-linked glycans are fucosylated [45].

In terms of individual glycoproteins measured with each method, LF and sIgA components were detected in high abundance (based on the sum of abundance of N-glycopeptides detected from each protein) in samples prepared using both the typical and optimal methods (Figure 6d). Previous glycoproteomic investigators also reported that LF and sIgA are predominant N-glycoproteins in human milk [46]. A previous study using MALDI-TOF/TOF MS found that 93% of relative abundances of N-glycans from human milk LF were fucosylated and/or sialylated [47]. Our findings aligns with this previous report as for the optimized method, over 99% of human milk LF N-glycopeptides were fucosylated and/or sialylated based on relative abundance. A previous PGC-chip LC/Q-TOF MS analysis found that by relative abundance human milk sIgA N-glycans were 74% fucosylated and/or sialylated, 17% truncated and 9% high mannose [48]. Our results from the optimized method indicating that human milk sIgA is 59% fucosylated and/or sialylated, 22% truncated and 19% high mannose are in alignment with those reported results [48].

α -Lactalbumin is one of the most abundant proteins in human milk and contains two N-glycosylation sites on the triplet⁶⁴ Asn-Glu-⁶⁶Ser and unusually, ⁹⁰Asn-Ile-⁹²Cys [49]. To the best of our knowledge, a few studies have analyzed the glycosylation of α -lactalbumin but only via MS analysis, which does not allow confirmation of the glycopeptide composition and structure. MALDI-TOF MS was used to detect trypsin-digested glycopeptides [49] and determine the glycosylation sites via deglycosylated peptide analysis [12]. Using the optimized method herein, we found a novel tryptic N-glycopeptide with one phosphorylation (³³Asp-⁶⁴Asn-⁷⁷Leu with HexNAc₅Hex₄Fuc₁NeuAc₁, *m/z* 1768.7608, *z*=4) derived from α -lactalbumin and elucidated its glycan composition and peptide sequence (tandem MS spectrum in Figure S7). Although some studies identified the glycosylation of α -lactalbumin, phosphorylation has, as

far as we are aware, not yet been identified in the analysis of human milk α -lactalbumin. As this finding is novel, we confirmed its accuracy via manual annotation of the tandem MS spectrum. Tryptic N-glycopeptides of α -lactalbumin includes relatively extensive peptides sequence due to absence of trypsin cleavage sites (Arg and Leu) between ^{32}Leu and ^{77}Leu . Establishing an ideal method for α -lactalbumin N-glycopeptide analysis will therefore require optimizing protein digestion parameters (including perhaps using additional proteolytic enzymes) and data processing conditions that enable coverage of a wider range of glycopeptide masses.

Casein proteins make up 20–40% of human milk proteins [50]. Most previous glycosylation studies focused on the O-glycosylation of κ -casein [51-53]. Recently, Caval et al. [54] identified a monofucosylated biantennary N-glycopeptide ($\text{HexNAc}_4\text{Hex}_5\text{Fuc}_1 + \text{DTR}^{69}\text{NESTQNCVVAEPEK}$) from human milk α -S1-casein by nano LC/Orbitrap MS combined with ETHcD-fragmentation. The optimized method developed herein allowed identification of abundant N-glycopeptides (7% of total N-glycopeptides) derived from α -S1-casein. In addition to the glycan-glycosite combination previously described [54], we identified an additional 22 different N-glycan compositions ($\text{HexNAc}_{3-6}\text{Hex}_{4-7}\text{Fuc}_{0-2}\text{NeuAc}_{0-2}$) on the same glycosylation site (^{69}N) in α -S1-casein (Table S3). Of total α -S1-casein N-glycopeptides, 86% of N-glycopeptides included fucose and sialic acid residues both on $\text{HexNAc}_{3-6}\text{Hex}_{4-7}$ backbone and 12% and <2% of N-glycopeptides contained either fucose or sialic acid, respectively.

A limitation of our method is that it did not detect several glycoproteins previously identified in human milk. For example, Froelich et al. identified bile-salt stimulated lipase, tenascin and xanthine dehydrogenase from human milk samples [55]. The lack of detection of these compounds is likely due to our attempt to analyze the array of milk glycoproteins within a single analysis, unlike Froelich et al. who first used SDS-PAGE separation to isolate individual glycoproteins. Analysis of the array of milk glycoproteins simultaneously is potentially limited by the large number of co-eluting peptides and

glycopeptides which can suppress the signal of low abundance glycopeptides and a lack of selection of low abundance glycopeptides for MS/MS fragmentation due to time limitations (limited by MS/MS acquisition rate). Furthermore, a limitation of this study is that centrifugation was performed to remove lipids, thus decreasing detection of milk fat globule membrane-associated proteins [56]. We selected to remove lipids in this manner to prevent clogging of the LC column and contamination in the mass spectrometer. Additional work is needed to apply the methods developed herein with an approach to extract glycoproteins from the milk fat globule membrane to enable their analysis.

3.5. Validation of the optimized method

The reproducibility of the optimized method was assessed in terms of variation in the relative abundance of each detected N-glycopeptide between instrumental replicates, between extractions within a single investigator (intra-individual) and between extractions performed by multiple investigators (inter-individual). To create all of these comparisons, samples were extracted by three investigators who each extracted the samples on two occasions and measured the extracted N-glycopeptides by LC/MS/MS in duplicate.

Comparison of each of the instrumental duplicates demonstrated a high reproducibility (Pearson correlation $R > 0.96$ for all duplicates, Figure S8a). Comparison of the results of duplicate extractions by a single investigator ($n = 2$) demonstrated high reproducibility ($R > 0.93$ for all comparisons across duplicate extractions for all duplicate measures, Figure S8b). Comparison of the results of extractions across multiple investigators ($n = 3$) also demonstrated high reproducibility ($R > 0.93$ for all comparisons across investigators for all extractions and all duplicate measures, Figure S8c). Experimental reproducibility was relatively lower across both intra-individual and inter-individual sample preparation replicates than across instrument measurement replicates, particularly for minor N-glycopeptides. This

finding could be due to individual variation in pipetting for separating the pellet from the supernatant and re-dissolving the EtOH precipitated proteins.

4. Conclusion

Human milk protein extraction methods, proteolytic N-glycopeptide enrichment methods and tandem MS fragmentation techniques were systematically examined to determine an optimal approach to analyze human milk N-glycoprotein in this study. The optimized method (EtOH precipitation, HILIC SPE and EThcD fragmentation) dramatically enhanced the capacity for reliable site-specific profiling of both minor and major human milk N-glycoproteins. We anticipate that this optimized approach can be further improved via additional experimentation with factors such as the conditions in EThcD (e.g., collision energy, reaction time and automatic gain control target), conditions to improve redissolution of precipitated proteins after EtOH extraction (e.g., increasing vortex time, adding sonication) and HILIC-SPE conditions (e.g., adjusting the pH and compositions of the wash and elution solvents). Moreover, we will consider applying glycopeptide-specific enrichment techniques based on lectin array, boronic acid resins or hydrazide-activated resins for improving separation of glycopeptides from non-glycosylated peptides in future analyses. This study significantly advances methods for the characterization of human milk N-glycoproteins, which is essential to better understand the bioactivity of human milk glycoproteins in the breast milk-fed infant.

Declaration of competing interest

The authors declare that they have no conflict of interest.

CRediT authorship contribution statement

Bum Jin Kim: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Roles/Writing - original draft, Writing - review & editing. **David C. Dallas:** Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing - review & editing.

Acknowledgements

This study was supported by the K99/R00 Pathway to Independence Career Award, Eunice Kennedy Shriver Institute of Child Health & Development of the National Institutes of Health (R00HD079561) (D.C.D.), the Gerber Foundation (2017-1586) (D.C.D) and the USDA National Institute of Food and Agriculture (2018-67017-27521) (D.C.D.). We acknowledge the Mass Spectrometry Center at Oregon State University, which is supported in part by the National Institute of Health grant (NIH # 1S10OD020111-01).

References

- [1] B. Lonnerdal, Nutritional and physiologic significance of human milk proteins, *Am. J. Clin. Nutr.* 77 (2003) 1537S-1543S. <http://doi.org/10.1093/ajcn/77.6.1537S>.
- [2] O. Ballard, A. L. Morrow, Human milk composition: Nutrients and bioactive factors, *Pediatr. Clin. North. Am.* 60 (2013) 49-74. <http://doi.org/10.1016/j.pcl.2012.10.002>.
- [3] B. Lonnerdal, Human milk proteins: Key components for the biological activity of human milk, *Adv. Exp. Med. Biol.* 554 (2004) 11-25.
- [4] D. C. Dallas, M. A. Underwood, A. M. Zivkovic, J. B. German, Digestion of protein in premature and term infants, *J. Nutr. Disord. Ther.* 2 (2012) 112. <http://doi.org/10.4172/2161-0509.1000112>.
- [5] Institute of Medicine (US) Committee on Nutritional Status During Pregnancy and Lactation, 1991. Nutrition during lactation. National Academies Press (US): Washington (DC). <http://doi.org/10.17226/1577>.
- [6] D. S. Newburg, Glycobiology of human milk, *Biochemistry.* 78 (2013) 771-85. <http://doi.org/10.1134/S0006297913070092>.
- [7] D. C. Dallas, D. Sela, M. A. Underwood, J. B. German, C. B. Lebrilla, Protein-linked glycan degradation in infants fed human milk, *J. Glycomics Lipidomics. Suppl* 1 (2012) 002. <http://doi.org/10.4172/2153-0637.S1-002>.
- [8] B. Liu, D. S. Newburg, Human milk glycoproteins protect infants against human pathogens, *Breastfeed Med.* 8 (2013) 354-62. <http://doi.org/10.1089/bfm.2013.0016>.
- [9] R. Peterson, W. Y. Cheah, J. Grinyer, N. Packer, Glycoconjugates in human milk: Protecting infants from disease, *Glycobiology.* 23 (2013) 1425-38. <http://doi.org/10.1093/glycob/cwt072>.
- [10] D. Garrido, D. C. Dallas, D. A. Mills, Consumption of human milk glycoconjugates by infant-associated bifidobacteria: Mechanisms and implications, *Microbiology.* 159 (2013) 649-664. <http://doi.org/10.1099/mic.0.064113-0>.
- [11] D. C. Dallas, W. F. Martin, J. S. Strum, A. M. Zivkovic, J. T. Smilowitz, M. A. Underwood, M. Affolter, C. B. Lebrilla, J. B. German, N-linked glycan profiling of mature human milk by high-performance microfluidic chip liquid chromatography time-of-flight tandem mass spectrometry, *J. Agric. Food Chem.* 59 (2011) 4255-63. <http://doi.org/10.1021/jf104681p>.

- [12] G. Picariello, P. Ferranti, G. Mamone, P. Roepstorff, F. Addeo, Identification of N-linked glycoproteins in human milk by hydrophilic interaction liquid chromatography and mass spectrometry, *Proteomics*. 8 (2008) 3833-47. <http://doi.org/10.1002/pmic.200701057>.
- [13] X. Cao, D. Song, M. Yang, N. Yang, Q. Ye, D. Tao, B. Liu, R. Wu, X. Yue, Comparative analysis of whey N-glycoproteins in human colostrum and mature milk using quantitative glycoproteomics, *J. Agric. Food Chem.* 65 (2017) 10360-10367. <http://doi.org/10.1021/acs.jafc.7b04381>.
- [14] J. Huang, A. Guerrero, E. Parker, J. S. Strum, J. T. Smilowitz, J. B. German, C. B. Lebrilla, Site-specific glycosylation of secretory immunoglobulin A from human colostrum, *J. Proteome Res.* 14 (2015) 1335-49. <http://doi.org/10.1021/pr500826q>.
- [15] D. C. Dallas, A. Guerrero, N. Khaldi, P. A. Castillo, W. F. Martin, J. T. Smilowitz, C. L. Bevins, D. Barile, J. B. German, C. B. Lebrilla, Extensive in vivo human milk peptidomics reveals specific proteolysis yielding protective antimicrobial peptides, *J. Proteome Res.* 12 (2013) 2295-304. <http://doi.org/10.1021/pr400212z>.
- [16] S. Karav, A. Le Parc, J. M. Leite Nobrega de Moura Bell, S. A. Frese, N. Kirmiz, D. E. Block, D. Barile, D. A. Mills, Oligosaccharides released from milk glycoproteins are selective growth substrates for infant-associated bifidobacteria, *Appl. Environ. Microbiol.* 82 (2016) 3622-3630. <http://doi.org/10.1128/AEM.00547-16>.
- [17] D. C. Dallas, V. Weinborn, J. M. de Moura Bell, M. Wang, E. A. Parker, A. Guerrero, K. A. Hettinga, C. B. Lebrilla, J. B. German, D. Barile, Comprehensive peptidomic and glycomic evaluation reveals that sweet whey permeate from colostrum is a source of milk protein-derived peptides and oligosaccharides, *Food Res. Int.* 63 (2014) 203-209. <http://doi.org/10.1016/j.foodres.2014.03.021>.
- [18] X. Cao, Y. Zheng, S. Wu, N. Yang, J. Wu, B. Liu, W. Ye, M. Yang, X. Yue, Characterization and comparison of milk fat globule membrane N-glycoproteomes from human and bovine colostrum and mature milk, *Food Funct.* 10 (2019) 5046-5058. <http://doi.org/10.1039/c9fo00686a>.
- [19] J. Liu, F. Wang, J. Zhu, J. Mao, Z. Liu, K. Cheng, H. Qin, H. Zou, Highly efficient N-glycoproteomic sample preparation by combining C18 and graphitized carbon adsorbents, *Anal. Bioanal. Chem.* 406 (2014) 3103-9. <http://doi.org/10.1007/s00216-014-7716-9>.
- [20] M. Thaysen-Andersen, I. B. Thogersen, H. J. Nielsen, U. Lademann, N. Brunner, J. J. Enghild, P. Hojrup, Rapid and individual-specific glycoprofiling of the low abundance N-glycosylated protein tissue inhibitor of metalloproteinases-1, *Mol. Cell. Proteomics*. 6 (2007) 638-47. <http://doi.org/10.1074/mcp.M600407-MCP200>.

- [21] L. Cao, Y. Qu, Z. Zhang, Z. Wang, I. Prytkova, S. Wu, Intact glycopeptide characterization using mass spectrometry, *Expert Rev. Proteomics*. 13 (2016) 513-22.
<http://doi.org/10.1586/14789450.2016.1172965>.
- [22] M. Rogeberg, H. Malerod, H. Roberg-Larsen, C. Aass, S. R. Wilson, On-line solid phase extraction-liquid chromatography, with emphasis on modern bioanalysis and miniaturized systems, *J. Pharm. Biomed. Anal.* 87 (2014) 120-9. <http://doi.org/10.1016/j.jpba.2013.05.006>.
- [23] C. West, C. Elfakir, M. Lafosse, Porous graphitic carbon: A versatile stationary phase for liquid chromatography, *J. Chromatogr. A*. 1217 (2010) 3201-16.
<http://doi.org/10.1016/j.chroma.2009.09.052>.
- [24] M. Wuhrer, A. R. de Boer, A. M. Deelder, Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry, *Mass Spectrom. Rev.* 28 (2009) 192-206.
<http://doi.org/10.1002/mas.20195>.
- [25] D. Takakura, A. Harazono, N. Hashii, N. Kawasaki, Selective glycopeptide profiling by acetone enrichment and LC/MS, *J. Proteomics*. 101 (2014) 17-30.
<http://doi.org/10.1016/j.jprot.2014.02.005>.
- [26] W. R. Alley, Jr., Y. Mechref, M. V. Novotny, Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data, *Rapid Commun. Mass Spectrom.* 23 (2009) 161-70. <http://doi.org/10.1002/rcm.3850>.
- [27] Q. Yu, B. Wang, Z. Chen, G. Urabe, M. S. Glover, X. Shi, L. W. Guo, K. C. Kent, L. Li, Electron-transfer/higher-energy collision dissociation (ET_hCD)-enabled intact glycopeptide/glycoproteome characterization, *J. Am. Soc. Mass Spectrom.* 28 (2017) 1751-1764.
<http://doi.org/10.1007/s13361-017-1701-4>.
- [28] M. Bern, Y. J. Kil, C. Becker, Byonic: Advanced peptide and protein identification software, *Curr. Protoc. Bioinformatics*. Chapter 13 (2012) Unit13 20.
<http://doi.org/10.1002/0471250953.bi1320s40>.
- [29] S. M. Totten, C. L. Feasley, A. Bermudez, S. J. Pitteri, Parallel comparison of N-linked glycopeptide enrichment techniques reveals extensive glycoproteomic analysis of plasma enabled by sax-erlic, *J. Proteome Res.* 16 (2017) 1249-1260. <http://doi.org/10.1021/acs.jproteome.6b00849>.
- [30] Z. Darula, K. F. Medzihradszky, Analysis of mammalian O-glycopeptides-we have made a good start, but there is a long way to go, *Mol. Cell Proteomics*. 17 (2018) 2-17.
<http://doi.org/10.1074/mcp.MR117.000126>.

- [31] M. I. Catalina, C. A. Koeleman, A. M. Deelder, M. Wührer, Electron transfer dissociation of n-glycopeptides: Loss of the entire N-glycosylated asparagine side chain, *Rapid Commun. Mass Spectrom.* 21 (2007) 1053-61. <http://doi.org/10.1002/rcm.2929>.
- [32] R. L. Beverly, M. A. Underwood, D. C. Dallas, Peptidomics analysis of milk protein-derived peptides released over time in the preterm infant stomach, *J. Proteome Res.* 18 (2019) 912-922. <http://doi.org/10.1021/acs.jproteome.8b00604>.
- [33] S. Ozcan, D. A. Barkauskas, L. Renee Ruhaak, J. Torres, C. L. Cooke, H. J. An, S. Hua, C. C. Williams, L. M. Dimapasoc, J. Han Kim, M. Camorlinga-Ponce, D. Rocke, C. B. Lebrilla, J. V. Solnick, Serum glycan signatures of gastric cancer, *Cancer Prev. Res. (Phila)*. 7 (2014) 226-35. <http://doi.org/10.1158/1940-6207.CAPR-13-0235>.
- [34] B. J. Kim, C. Han, H. Moon, J. Kwon, I. S. Jang, S. K. Lim, K. W. Park, J. S. Choi, H. J. An, Monitoring of post-mortem changes of saliva N-glycosylation by nano LC/MS, *Anal. Bioanal. Chem.* 410 (2018) 45-56. <http://doi.org/10.1007/s00216-017-0702-2>.
- [35] E. Armaforte, E. Curran, T. Huppertz, C. A. Ryan, M. F. Caboni, P. M. O'Connor, R. P. Ross, C. Hirtz, N. Sommerer, F. Chevalier, A. L. Kelly, Proteins and proteolysis in pre-term and term human milk and possible implications for infant formulae, *Int. Dairy J.* 20 (2010) 715-723. <http://doi.org/10.1016/j.idairyj.2010.03.008>.
- [36] S. D. Nielsen, R. L. Beverly, D. C. Dallas, Peptides released from foremilk and hindmilk proteins by breast milk proteases are highly similar, *Front. Nutr.* 4 (2017) 54. <http://doi.org/10.3389/fnut.2017.00054>.
- [37] S. Wu, N. Tao, J. B. German, R. Grimm, C. B. Lebrilla, Development of an annotated library of neutral human milk oligosaccharides, *J. Proteome Res.* 9 (2010) 4138-51. <http://doi.org/10.1021/pr100362f>.
- [38] D. Rajalingam, C. Loftis, J. J. Xu, T. K. Kumar, Trichloroacetic acid-induced protein precipitation involves the reversible association of a stable partially structured intermediate, *Protein Sci.* 18 (2009) 980-93. <http://doi.org/10.1002/pro.108>.
- [39] L. L. Manza, S. L. Stamer, A. J. Ham, S. G. Codreanu, D. C. Liebler, Sample preparation and digestion for proteomic analyses using spin filters, *Proteomics*. 5 (2005) 1742-5. <http://doi.org/10.1002/pmic.200401063>.
- [40] G. Han, M. Ye, H. Zou, Development of phosphopeptide enrichment techniques for phosphoproteome analysis, *Analyst*. 133 (2008) 1128-38. <http://doi.org/10.1039/b806775a>.

- [41] Y. Huang, Y. Nie, B. Boyes, R. Orlando, Resolving isomeric glycopeptide glycoforms with hydrophilic interaction chromatography (HILIC), *J. Biomol. Tech.* 27 (2016) 98-104. <http://doi.org/10.7171/jbt.16-2703-003>.
- [42] Y. Xue, J. Xie, P. Fang, J. Yao, G. Yan, H. Shen, P. Yang, Study on behaviors and performances of universal N-glycopeptide enrichment methods, *Analyst.* 143 (2018) 1870-1880. <http://doi.org/10.1039/C7AN02062G>.
- [43] L. Turiak, S. Sugar, A. Acs, G. Toth, A. Gomory, A. Telekes, K. Vekey, L. Drahos, Site-specific N-glycosylation of hela cell glycoproteins, *Sci. Rep.* 9 (2019) 14822. <http://doi.org/10.1038/s41598-019-51428-x>.
- [44] M. Mancera-Arteu, E. Gimenez, F. Benavente, J. Barbosa, V. Sanz-Nebot, Analysis of O-glycopeptides by acetone enrichment and capillary electrophoresis-mass spectrometry, *J. Proteome Res.* 16 (2017) 4166-4176. <http://doi.org/10.1021/acs.jproteome.7b00524>.
- [45] C. C. Nwosu, D. L. Aldredge, H. Lee, L. A. Lerno, A. M. Zivkovic, J. B. German, C. B. Lebrilla, Comparison of the human and bovine milk N-glycome via high-performance microfluidic chip liquid chromatography and tandem mass spectrometry, *J. Proteome Res.* 11 (2012) 2912-24. <http://doi.org/10.1021/pr300008u>.
- [46] J. T. Smilowitz, S. M. Totten, J. Huang, D. Grapov, H. A. Durham, C. J. Lammi-Keefe, C. B. Lebrilla, J. B. German, Human milk secretory immunoglobulin a and lactoferrin N-glycans are altered in women with gestational diabetes mellitus, *J. Nutr.* 143 (2013) 1906-12. <http://doi.org/10.3945/jn.113.180695>.
- [47] T. Yu, C. Guo, J. Wang, P. Hao, S. Sui, X. Chen, R. Zhang, P. Wang, G. Yu, L. Zhang, Y. Dai, N. Li, Comprehensive characterization of the site-specific N-glycosylation of wild-type and recombinant human lactoferrin expressed in the milk of transgenic cloned cattle, *Glycobiology.* 21 (2011) 206-24. <http://doi.org/10.1093/glycob/cwq151>.
- [48] E. Goonatilleke, J. T. Smilowitz, K. V. Marino, B. J. German, C. B. Lebrilla, M. Barboza, Immunoglobulin a N-glycosylation presents important body fluid-specific variations in lactating mothers, *Mol. Cell. Proteomics.* 18 (2019) 2165-2177. <http://doi.org/10.1074/mcp.RA119.001648>.
- [49] M. G. Giuffrida, M. Cavaletto, C. Giunta, B. Neuteboom, A. Cantisani, L. Napolitano, V. Calderone, J. Godovac-Zimmermann, A. Conti, The unusual amino acid triplet Asn-Ile-Cys is a glycosylation consensus site in human alpha-lactalbumin, *J. Protein Chem.* 16 (1997) 747-53. <http://doi.org/10.1023/a:1026359715821>.

- [50] B. Lonnerdal, E. Forsum, Casein content of human milk, *Am. J. Clin. Nutr.* 41 (1985) 113-20.
<http://doi.org/10.1093/ajcn/41.1.113>.
- [51] A. M. Fiat, J. Jolles, J. P. Aubert, M. H. Loucheux-Lefebvre, P. Jolles, Localisation and importance of the sugar part of human casein, *Eur. J. Biochem.* 111 (1980) 333-9.
<http://doi.org/10.1111/j.1432-1033.1980.tb04946.x>.
- [52] N. Azuma, S. Kaminogawa, K. Yamauchi, Properties of glycomacropeptide and para- κ -casein derived from human κ -casein and comparison of human and bovine κ -caseins as to susceptibility to chymosin and pepsin, *Agric. Biol. Chem.* 48 (1984) 2025-2031.
<http://doi.org/10.1080/00021369.1984.10866443>.
- [53] C. Kunz, B. Lönnerdal, Human-milk proteins: Analysis of casein and casein subunits by anion-exchange chromatography, gel electrophoresis, and specific staining methods, *Am. J. Clin. Nutr.* 51 (1990) 37-46. <http://doi.org/10.1093/ajcn/51.1.37>.
- [54] T. Caval, J. Zhu, A. J. R. Heck, Simply extending the mass range in electron transfer higher energy collisional dissociation increases confidence in N-glycopeptide identification, *Anal. Chem.* 91 (2019) 10401-10406. <http://doi.org/10.1021/acs.analchem.9b02125>.
- [55] 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. <http://doi:10.1021/jf100112x>.
- [56] Y. Liao, R. Alvarado, B. Phinney, B. Lönnerdal, Proteomic characterization of human milk fat globule membrane proteins during a 12 month lactation period, *J. Proteome Res.* 10 (2011) 3530-3541. <https://doi.org/10.1021/pr200149t>.

Table 1. Summary of MS/MS spectra identified from the pooled human milk sample via HCD-, ETD- and EThcD-based full MS/MS scan.

Category	HCD	ETD	EThcD
Counts of total MS/MS spectra	12,568	7,857	7,667
Counts of peptide MS/MS spectra	1156	654	757
Counts of N-glycopeptide MS/MS spectra (before data filtering)	54	17	36
Counts of N-glycopeptide MS/MS spectra (after data filtering) ^a	20	9	25

^a N-Glycopeptide MS/MS spectra identified by Byonic searching were filtered based on Score (> 75) and PEP-2D (< 0.01) values.

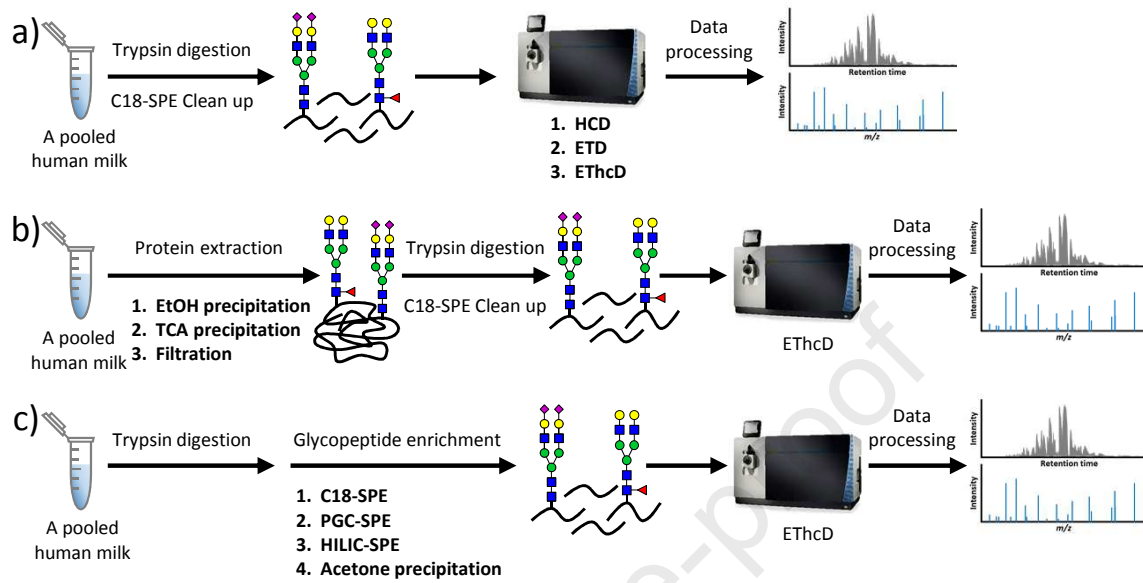
Figure 1

Figure 2

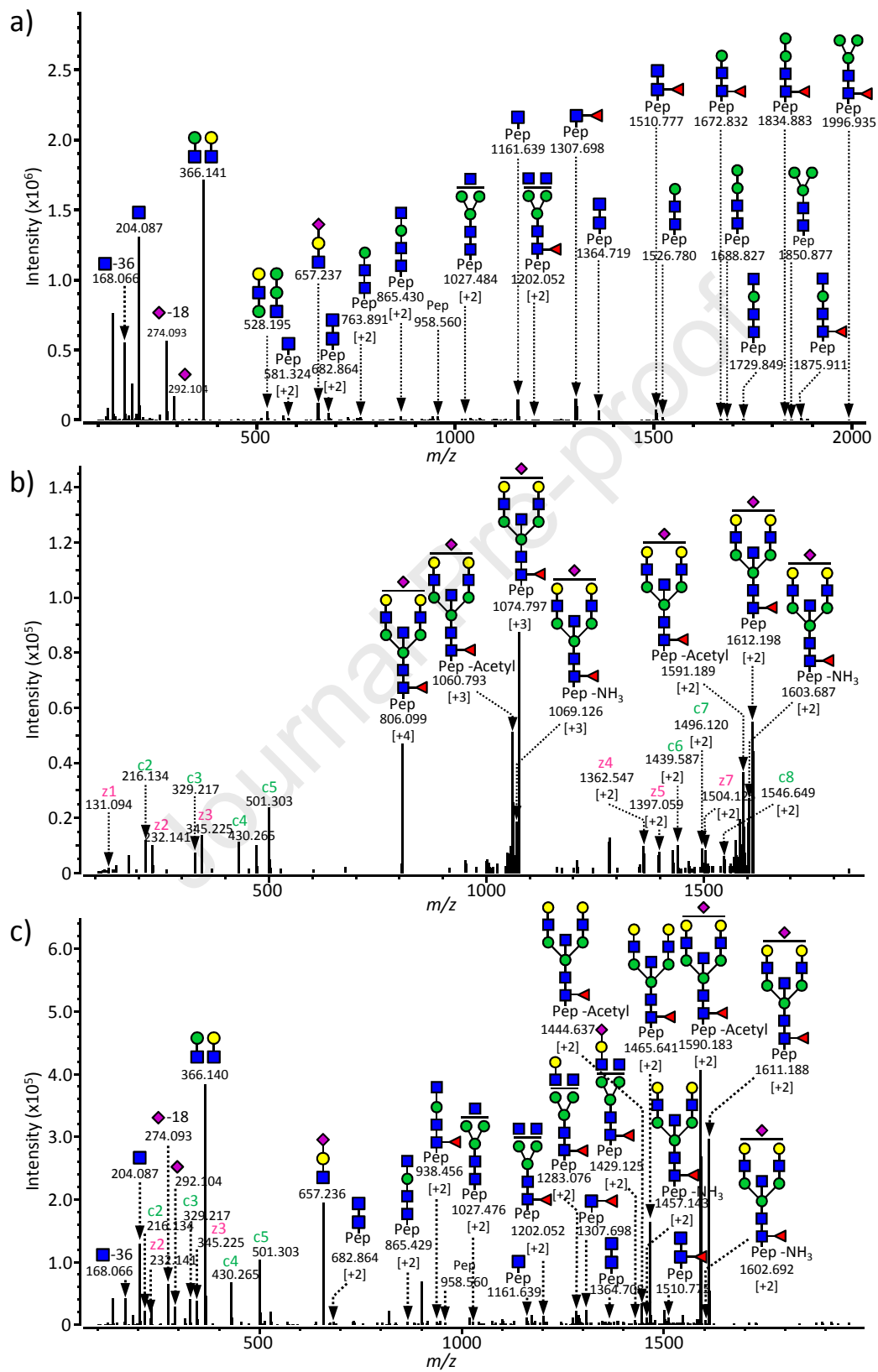


Figure 3

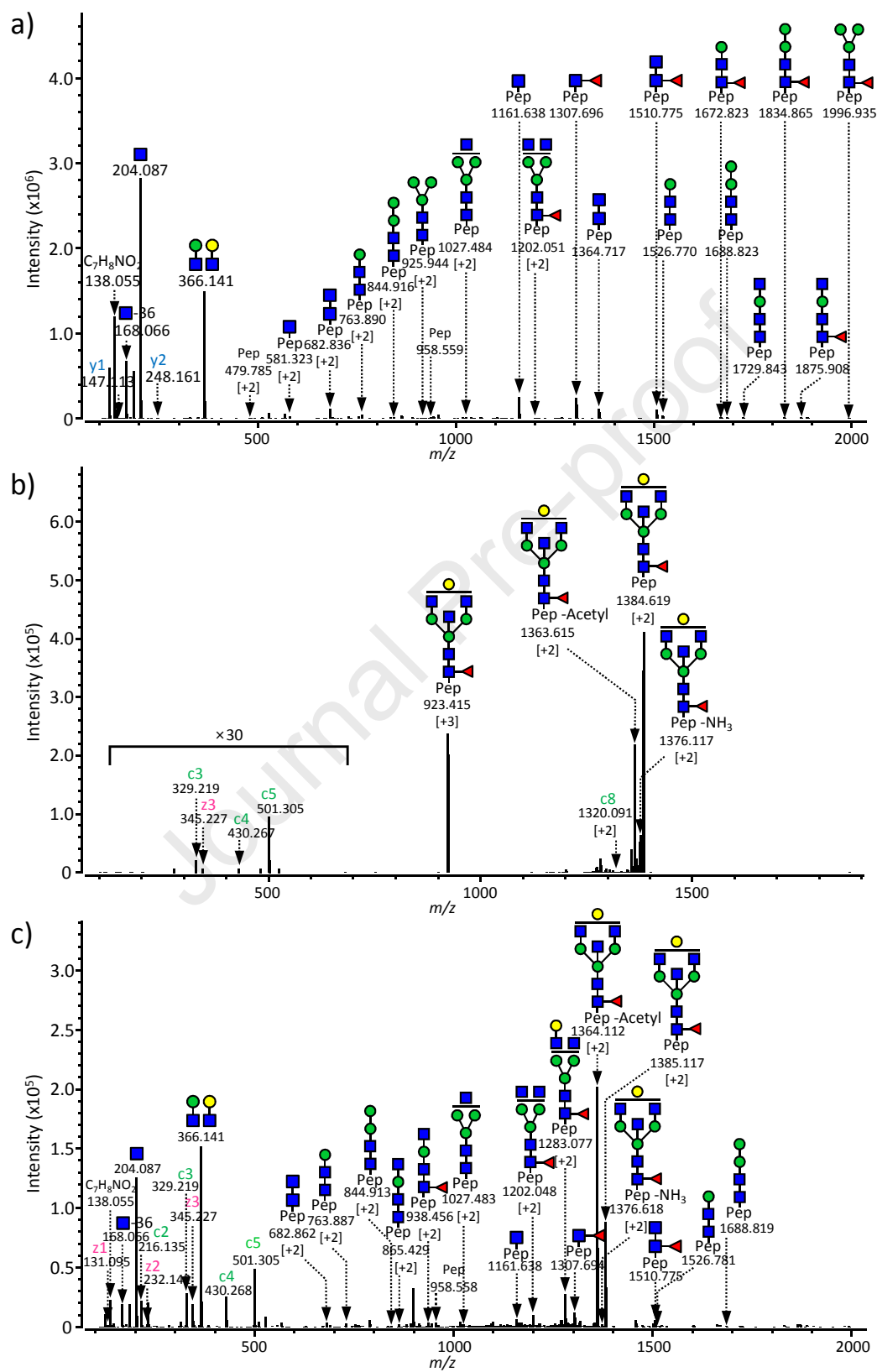


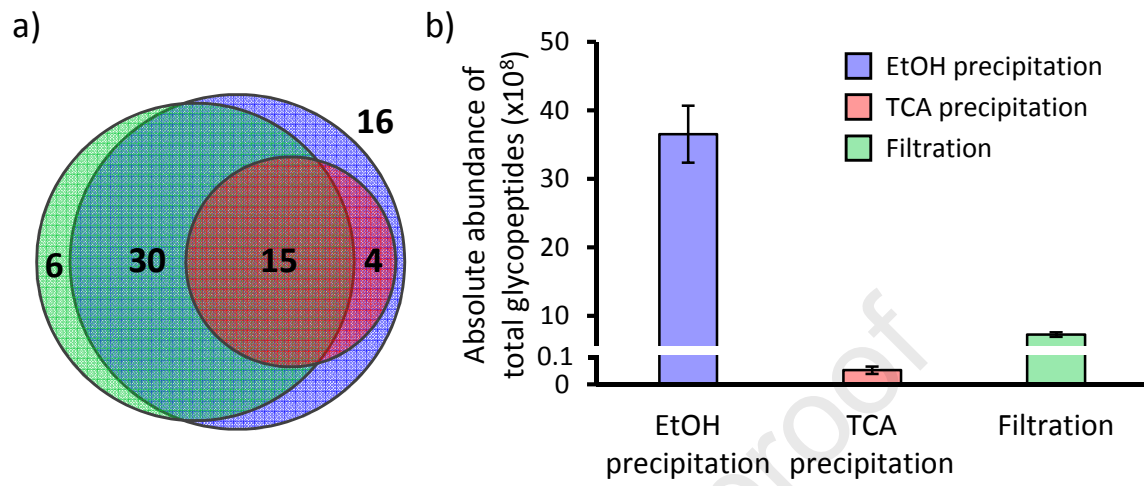
Figure 4

Figure 5

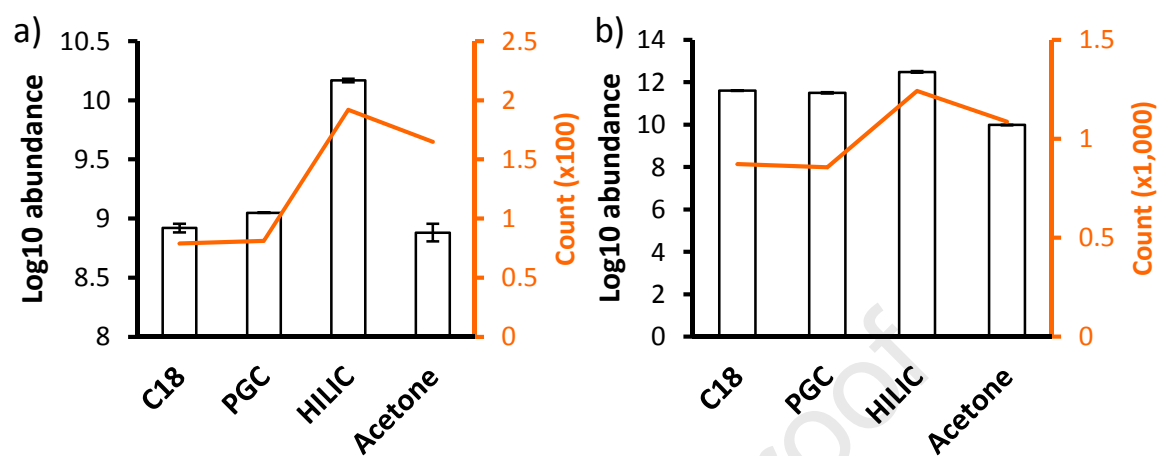
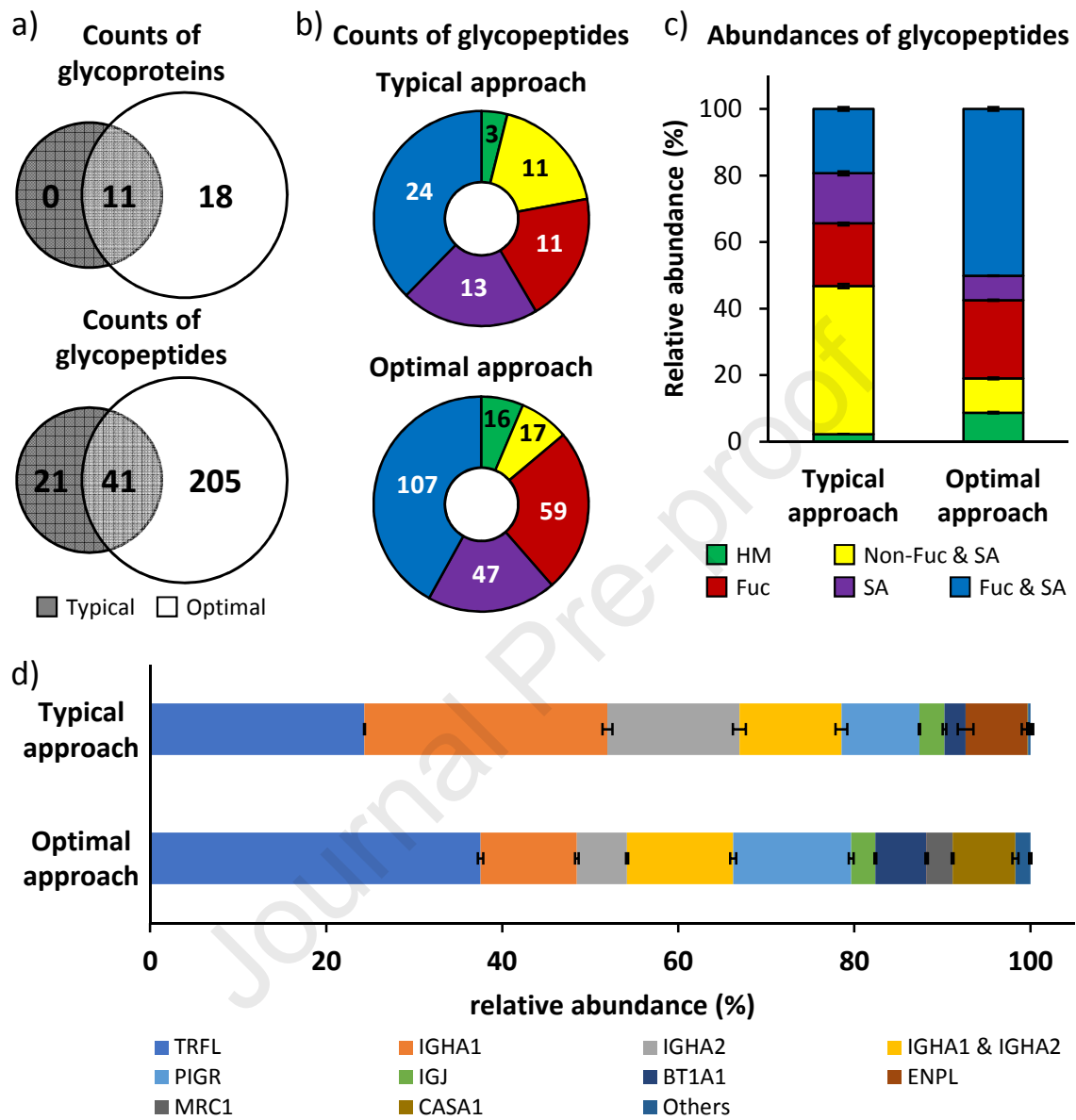


Figure 6



Highlights

- Both glycan and peptide moieties of N-glycopeptides were confirmed using EThcD.
- EtOH precipitation was most effective to enrich proteins from human milk.
- HILIC-SPE most efficiently extracted tryptic human milk N-glycopeptides.
- The optimized method dramatically enhanced human milk N-glycopeptide analysis.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: