

RESEARCH REPOSITORY

This is the author's final version of the work, as accepted for publication following peer review but without the publisher's layout or pagination. The definitive version is available at:

https://doi.org/10.1021/acs.jproteome.7b00486

Liao, Y., Weber, D., Xu, W., Durbin-Johnson, B.P., Phinney, B.S. and Lönnerdal, B. (2017) Absolute quantification of human milk caseins and the whey/casein ratio during the first year of lactation. Journal of Proteome Research, 16 (11). pp. 4113-4121.

https://researchrepository.murdoch.edu.au/id/eprint/39174

Copyright: © 2017 American Chemical Society It is posted here for your personal use. No further distribution is permitted.

Absolute quantification of human milk caseins and the whey/casein ratio during the first year of lactation

Yalin Liao^{1,2}, Darren Weber³, Wei Xu⁴, Blythe P Durbin-Johnson⁵, Brett S. Phinney³, Bo Lönnerdal¹*.

- 1. Department of Nutrition, University of California, Davis, 95616, USA
- John Curtin School of Medical Research, Australian National University, Canberra ACT 2601, Australia
- Genome Center Proteomics Core Facility, University of California, Davis, 95616, USA
- School of Veterinary & Life Sciences, Murdoch University, Murdoch WA 6150, Australia
- Division of biostatistics, Department of Public Health Sciences, School of Medicine, University of California, Davis, 95616, USA

* Correspondence: bllonnerdal@ucdavis.edu (B.L.)

Abstract

Whey proteins and caseins in breast milk provide bioactivities and also have different amino acid composition. Accurate determination of these two major protein classes will therefore provide a better understanding of human milk composition and function and also aid in developing improved infant formulas based on bovine whey proteins and caseins. In this study, we implemented a label-free LC-MS/MS quantitative analysis based on the APEX spectral counting quantitative pipeline to estimate absolute concentrations of alpha-casein, beta-casein and kappa-casein in human milk samples (n=88) collected between day 1 and day 360 post partum. Total protein concentration measured by ninhydrin-based amino acid analysis ranged from 2.03 to 17.52 with a mean of 9.37±3.65 g/L. Casein subunits ranged from 0.06 to 0.82 g/L (alpha-), 0.06 to 2.23 g/L (beta-), and 0.13 to 1.30 g/L (kappa-), with beta-casein having the highest average concentration among the three subunits. Calculated whey/casein ratio ranged from 73:27 to 97:3. Linear regression analyses show significant decreases in total protein, kappa-casein, and total casein during the course of lactation. Our study presents a novel and accurate analysis of human milk casein content, demonstrating a lower casein content than earlier believed, which has implications for improved infants formulas.

Key words

Human milk, alpha-casein, beta-casein, kappa-casein, total protein, whey/casein ratio, label-free LC-MS/MS, amino acid analysis, APEX

Introduction

Proteins from mothers' milk provide approximately 8% of the energy required by infants, and more importantly, provide the building blocks for protein synthesis for the rapid growth of infants during early life ¹. Proteins in human milk are present in two major compartments, whey and casein micelles, as well as other, but minor compartments, such as cells, milk fat globule membranes, and exosomes ^{2 3 4}. Human milk casein micelles are water-insoluble high molecular weight molecules mainly made up of three casein subunits, alpha-S1-casein, beta-casein, and kappa-casein ⁵. Casein micelles can be physically separated from the whey fraction, which contains a diverse profile with more than a hundred proteins ⁶, and largely remains in the aqueous phase even at an environment as acidic as pH 4.3 ⁷.

Biological functions of milk caseins are extensive; for example, caseins are the primary source of phosphate and calcium in human milk, because of the highly phosphorylated nature of beta-casein and alpha-S1-casein, and the requirement for calcium in forming the aggregates of casein micelles ⁸. Longitudinal changes in beta-casein phosphorylation status during the first two months lactation have been found to be different in term (mostly decrease, n=8) and preterm milk (stay unchanged or increase, n=16), suggesting that this post-translational modification of beta-casein may also play meaningful regulatory roles ⁹. Kappa-casein is known to have antioxidant, ACE-inhibitory, and anti-bacterial activities ¹⁰. Casein glycomacropeptide (CGMP) hydrolysates have endotoxin-binding activity, and therefore negatively regulates TLR4-mediated inflammatory responses following LPS stimulation in cell culture ¹¹. In ulcerative colitis, the disease-modifying effect of CGMP was similar to that of mesalamine dose escalation ¹². Caseino-phosphopeptides (CPP) formed during casein digestion can influence calcium and zinc absorption ¹³.

It is well recognized that considerable differences in milk protein biology exist between bovine and human milk, e.g. alpha-S1-casein is the most abundant casein subunit in bovine milk, whereas it is a minor one in human milk. Further, bovine milk is very high in casein (about 82% of total protein), whereas human milk is wheypredominant, with a conventionally estimated whey:casein ratio of 60:40. During the manufacturing process for cow-milk based infant formula, skim milk powder (high in casein) is mixed with whey protein concentrate in order to make the whey:casein ratio more similar to that of human milk. Assessment of true casein concentration and whey/casein ratio is therefore a key consideration when attempting to make infant formula protein composition as similar to that of human milk as possible. Caseins and whey proteins are also differentially expressed during the course of lactation; hence there is no "fixed ratio" of whey and casein ², which should be considered when using a "staging" approach, i.e. the composition of the formula varies with the age of the infant ¹⁴.

In this study, we improved the traditional approaches which isolate the casein fraction first, followed by a variety of biochemical methods to assess casein protein quantity. We implemented a label-free LC-MS/MS quantitative data analysis pipeline based on the APEX spectral counting software to estimate absolute amounts of total protein and caseins at the same time, and were able to obtain absolute concentrations of alpha-S1-casein, beta-casein, and kappa-casein in a large number of human milk samples.

Experimental Section

Human milk sample collection

All human milk collection procedures were approved by the Institutional Review Board (IRB) at the University of California, Davis. Milk donors were healthy

volunteer mothers with term delivery (38-42 weeks), exclusively breast-fed to approximately 6 months. Samples from four mothers were collected at each of the twenty-two time points. Breast milk samples were collected from one breast (at least 2-4 h after prior nursing) into a sterile 50 ml tube by either manual expression (up to 7 days) or using a breast pump. All samples were immediately placed on ice, transported to the laboratory and stored at -20°C until analyses.

SDS-PAGE of human milk

Protein concentration (from a Day 30 human milk sample) was determined by BCA protein assay (Pierce, Rockford, IL); a total of 30 µg protein was then reduced and electrophoresed on a Novex 10% Tris-Glycine gel (Thermo Fisher Scientific, Waltham, MA) for 50 min at 180V. The gel was stained by 0.1% Coomassie Blue R250 in 10% acetic acid, 50% methanol, 40% H₂O for 1 h. The gel was then destained in 10% acetic acid, 50% methanol, 40% H₂O until background was minimal and protein bands were clearly visualized. The SDS–PAGE gel is used to illustrate the protein composition of human milk in Figure 1D.

Mass Spectrometry Sample Preparation and In-gel Digestion

10 µl of human milk were added to 10 µl of LDS sample buffer (Thermo Fisher Scientific, Waltham, MA), and all 20 µl were loaded into a Novex 10% Tris-Glycine gel (Thermo Fisher Scientific, Waltham, MA) and run for 20 min at 125 V to produce a gel smear. The gel was stained for 30 min using Instant Blue Protein Stain (Expedeon Inc. San Diego, CA) and the gel "blobs" were excised and digested as follows: Gel pieces were cut into \sim 1 mm³ pieces, washed three times with 50 mM ammonium bicarbonate (AmBic), pH 8, and then chemically dried twice with 100% acetonitrile (ACN). Gel pieces were then reduced in 15 mM dithiothreitol for 30 min at 56°C, chemically dried twice with 100% ACN, and then alkylated with 20 mM iodoacetamide for 20 min in the dark. The gel pieces were washed twice more with 50 mM AmBic, chemically dried twice with 100% ACN, and then mechanically dried using vacuum centrifugation. Next, 3 µl of PNGase F (New England BioLabs, Ipswich, MA) in 50 mM AmBic was added to each gel and incubated overnight at 37°C to remove any N-linked glycans. The following day, the supernatant was removed and discarded and each gel was chemically dried twice with 100% ACN, and then mechanically dried using vacuum centrifugation. Trypsin (Promega, Madison, WI) in 50 mM AmBic was added at a 1:30 ratio and digested overnight at 37°C. Peptide extraction proceeded the next day by collecting the supernatant and adding 60% ACN in 0.1% trifluoroacetic acid to the gel pieces, sonicating for 10 min, and centrifuging for 5 min. The supernatant was collected and added to the supernatant of the previously collected supernatant. The supernatant was then vacuum-centrifuged.

LC-MS/MS Methods

LC-MS/MS analysis was performed on a standard top 15 method using a Thermo Scientific Q Exactive orbitrap mass spectrometer in conjunction with a Paradigm MG4 HPLC (Michrom Bio Resources, Auburn, CA). The digested peptides were loaded onto a Michrom C18 trap and desalted before they were separated using a Michrom 200 μ m × 150 mm Magic C18AQ reverse phase column. A flow rate of 2 μ l/min was used. Peptides were eluted using a 90-min gradient with 2% B to 35% B over 70 min, 35% B to 80% B for 5 min, 80% B for 2 min, a decrease from 80% to 2% B in 1 min, and held at 2% B for 12 min. (A= 0.1% formic acid, B= 100%

acetonitrile). A spray voltage of 2.0 kV was used with a transfer capillary temperature of 200°C.

Amino Acid Analysis (AAA)

Samples were dried via speed vacuum concentration (LabConco Centrivap Concentrator, Kansas City, MO) and then acid-hydrolyzed using 6N HCl (Pierce, Rockford, IL), 1% phenol (Sigma-Aldrich, St. Louis, MO) at 110°C for 24 h. The resulting acid hydrolysates were dried under vacuum and subsequently hydrated in Sodium Diluent (Pickering Laboratories, Mountain View, CA) containing norleucine (40 nmol/ml, CalBioChem, La Jolla, CA). Hydrolyzed samples were analyzed using ion-exchange chromatography (Hitachi 8800 Amino Acid Analyzer, Tokyo, Japan) with a post-column ninhydrin reaction. Calibration of the Hitachi 8800 was performed using amino acid standards (Sigma-Aldrich, St. Louis, MO) in conjunction with the National Institute of Standards and Technology's (NIST) amino acid standard reference data. Absorbance was measured and the response factor determined for each individual amino acid to quantify levels relative to the known amino acid standards. The included reference standard (norleucine) was used to correct for any variances in injection volume due to the autosampler.

Mass Spectrometry Data Analysis

APEX

Absolute protein expression (APEX), an open-sourced software ¹⁵, was used to determine absolute casein amounts in replicate milk samples similar to the approach presented in previous reports ^{15 16}. The APEX user-generated normalization factor, used in this study was the absolute amount of milk protein (μ g) loaded onto the

column for LC-MS/MS. The absolute amount of milk protein loaded on the column was calculated by taking the absolute amount of crude milk protein loaded on the gel and correcting for protein loss during the in-gel digestion step. The absolute amount of milk protein (in μ g) loaded onto the gel was calculated by analyzing each replicate milk sample by amino acid analysis. To correct for protein loss during the in-gel digestion were also analyzed by amino acid analysis. The amount of protein lost during the in-gel digestion step was calculated by taking the total amount of protein loaded on the gel, and then subtracting the total amount of protein extracted from the gel post in-gel digestion. This allowed us to calculate the absolute amount of total milk protein loaded on the column for every LC-MS/MS analysis.

Database Searching

Tandem mass spectra were identified using the search engine X! Tandem . X! Tandem (Sledgehammer (2013.09.01.1)) was set up to search the Uniprot Reference Human database (May 2012, 20252 entries) with an equal number of reverse sequences and 47 contaminant proteins, assuming the digestion enzyme trypsin. X! Tandem was searched with a fragment ion mass tolerance of 20 ppm and a parent ion tolerance of 20 ppm. Carbamidomethyl of cysteine was specified in X! Tandem as a fixed modification. Glu->pyro-Glu of the N-terminus, ammonia-loss of the Nterminus, gln->pyro-Glu of the N-terminus, deamidation of asparagine and glutamine, oxidation of methionine and tryptophan, dioxidation of methionine and tryptophan and acetylation of the N-terminus were specified in X! Tandem as variable modifications.

Criteria for Protein Identification

Scaffold (version Scaffold_4.2.1, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 50.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 92.0% probability and contained at least 1 identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm ¹⁷. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

Statistical Analysis

Linear regression was used to analyze the changes in amounts of milk protein over time. Analyses were conducted using the statistical software environment R, version 3.2.5 (R Core Team, 2016).

Results and Discussion

Absolute quantification of human milk caseins

Most studies exploring human milk casein concentration have involved separation of casein from the rest of milk matrix, either by centrifugation, with or without acid precipitation which our laboratory was among the first to implement for human milk in the late 1980's ¹⁸; or by 2D electrophoresis ^{19 20}; or by freeing caseins with EDTA chelation, followed by reconstructing the casein micelle with a new calcium core away from whey proteins ²¹. Kjeldahl nitrogen distribution analysis or colorimetric methods can be used downstream of these compartmentalizations, each yielding

different results¹. Direct casein measurement bypassing compartmentalization include immunoassay²², and a very recently described mass spectrometry based method²³. Altendorfer et al. applied LC pseudo-MRM Q-TOF mass spectrometry to analyse alpha-casein (CSN1S1) in milk collected within the first week post-partum, based on detection of the peptide LQNPSESSEPIPLESR²³. Prior to this report, we also attempted an MRM targeted MS approach to assay human milk alpha-casein, and the same peptide as the one used by Altendorfer et al.²³ was tested, as well as two additional peptides; EEYMNGNR and LNEYNQLQLQAAHAQEQR. We found that random and frequent internal post-translational amino acid modifications, including at least phosphorylation and deamidation, on the targeted peptides made it very difficult to obtain absolute quantification of caseins (data not shown). Further analyses revealed similar issues when quantifying beta-casein by MRM, based on detection of the peptides DTVYTK, VPIPQQVVPYPQR, and SPTIPFFDPQIPK (data not shown). Since only kappa-casein was considered accurate and reproducible by MRM (QYLPNSHPPTVVR, TYYANPAVVRPHAQIPQR) in our endeavour of absolute casein quantification, we did not further pursue MRM for investigating human milk whey/casein ratios.

We settled on a label-free LC-MS/MS approach using the APEX spectral counting quantitative software for whole milk samples without compartmentalization. The Absolute Protein Expression (APEX) software is a well-established method based on spectral counting and a machine learning derived correction factor (Oi) that takes into account the likelihood of each peptide to be detected by a mass spectrometer. APEX was chosen because it was easy to implement and has been shown to be comparable to precursor-based intensity methods such as IBAQ and TOP3, especially for highly abundant proteins ²⁴ ¹⁶. In addition, due to the large number of chemical and post-

translation modifications seen in our preliminary MRM analysis of the casein proteins we hypothesized that a spectral counting method would be more accurate, especially for highly abundant proteins such as caseins which seemed to be heavily deamidated. As illustrated in Figure 1A, each milk sample was separated into two aliquots, one was used to determine casein quantity by LC-MS/MS (top route), and the other was used to determine absolute total protein quantity (bottom route), which allowed us to accurately calculate the percentage of protein loss during the multi-step mass spectrometry sample preparation (top route), therefore obtaining an accurate measurement of each casein in the original milk sample. Upon establishment, this method was applied to all eighty-eight samples in this study (see Figure 1B for detailed time points), and we determined total casein as the sum of three measured individual caseins, and estimated total whey proteins by subtracting total casein from total proteins (Figure 1C, see also results in the following sections). Figure 1D shows a representative protein profile of whole human milk prior to sample processing as shown in Figure 1A. Prior to analysis of our samples, we used 1D-SDS-PAGE to remove lipids, sugars, minerals, etc. components that interfere with mass spectrometry processing.

Human milk total protein concentration decreases over time

The average total protein concentration of all eighty-eight samples was 9.37 ± 3.65 g/L (see Table S1 for individual numeric values, ranging from 2.03 to 17.52 g/L), and broadly concurs with most of the literature ³. Using FPLC with an anion-exchange column and polyacrylamide gradient gel electrophoresis techniques to analyze casein subunit composition, our analysis of milk samples within the first three months of lactation, revealed a total protein of 7.5-15.6 g/L (12.72±2.57 g/L), and this was done

by summing casein (FPLC or Kjeldahl) and whey protein measured separately ⁷. Our later analysis separating casein and whey by acid precipitation at pH4.3, with subsequent Kjeldahl analysis showed an average total protein concentration of 14.3 g/L at 0-3 days, 14.5 g/L at 6-10 days, 12.9 g/L at 11-30 days, 9.7 g/L at 31-60 days, 8.8 g/L at 61-120 days, 8.1 g/L at 121-240 days, 7.3 g/L at 241-365 days, and 8.9 g/L for samples beyond 365 days ²⁵. By acid precipitation to separate casein for Bradford protein assay, the Hartmann group reported 13.5 ± 2.1 g/L of total protein in skim milk from 25 mothers between 1 and 8 months lactation ²⁶.

Our data are shown in Figure 2A to visualize the estimated absolute and relative abundance across the lactation period. Considerable inter-individual variance at each time points is obvious; at 12 of the 22 time points, the highest value is more than twice that of the lowest (Table S1). On average, the ratio of highest/lowest reading among the four biological replicates at each point is 2.54, and all but one (D150) points have standard deviations more than 15% of the average. The average total protein at each point is shown in Figure 2B, with the highest value of 13.19 g/L at Day 1 (See also Table S2). The total protein levels following Day 1 do fluctuate, but display a significant decrease over time, and more steeply after approximately 3 months (p=0.002) (Figure 2C).

Large inter-individual variations in alpha-, beta-, kappa-casein exist at all lactation stages

We next analysed the absolute concentrations of the three casein subunits, alpha-S1casein (abbreviated as alpha-casein), beta-casein, kappa-casein, each encoded by a single gene on human chromosome 4. In bovine milk, two alpha-casein subtypes exist, alpha-S1-casein and alpha-S2-casein encoded by two different genes on chromosome 6, and the S1 subtype is the most abundant casein ²⁷. In contrast, alpha-casein is a minor casein component in human milk, hence much less information is available in literature. A recent study by Altendorfer et al.²³ utilized a tryptic alpha-S1 casein peptide as reference in MRM mass spectrometry to evaluate human milk collected during the first week after birth. Ranging from 0.03±0.01 to 0.54±0.02 g/L, alpha-S1 casein averaged ~0.127 g/L, with median of ~0.088 g/L (n=20) 23 . We also noted that the total protein concentration reported in this study is quite high, even when taking into account that samples were collected during the very first week of lactation, ranging from 36.00 ± 4.60 to 111.68 \pm 17.70 g/L, with an average of ~47.91 g/L and a median of ~43.30 g/L. In our study, the alpha-case in concentration of all samples is shown in Figure 3 (top), and ranges from 0.06 to 0.82 g/L, with an average of 0.23 g/L and a median of 0.21 g/L. Beta-casein shown in Figure 3 (middle) and kappa-casein in Figure 3 (bottom) are generally expressed at higher absolute quantities than alpha-casein. When calculating data from all 88 samples, beta-casein ranged from 0.06 to 2.23 g/L, with an average of 0.61 g/L and a median of 0.60 g/L; kappa-casein ranged from 0.13 to 1.30 g/L, with an average of 0.54 g/L and a median of 0.53 g/L. (See Table S1 for individual measurements).

Alpha- and beta-caseins, but not kappa-casein, are stably expressed throughout lactation

The average casein concentrations at each lactation stage were also further analysed and are shown in Figure 4 (see Table S2 for numeric values). Similar to total protein, quantitation of three casein subunits exhibited large inter-individual variations (Figure 4B, 4C, 4E). During the course of the 22 study time points, alpha-casein ranged from 0.14 ± 0.05 g/L to 0.39 ± 0.37 g/L, beta-casein from 0.39 ± 0.22 g/L to 1.00 ± 0.88 g/L, and kappa-casein from 0.40 ± 0.09 g/L to 0.71 ± 0.28 g/L. All three casein subunits show a trend of decreasing expression towards late lactation; however, linear regression analysis confirmed significant changes only for kappa-casein over time (Figure 4B, 4D, 4F).

In terms of total casein summed from alpha-, beta-, and kappa-caseins, it ranged from 0.31 to 4.36 g/L, with an average of 1.38 ± 0.61 g/L and a median of 1.38 g/L (Figure 5A, Table S3). This is lower than the 3.4 ± 0.97 g/L reported by Khan et al. ²⁶, in which pH 4.3 acid precipitation was performed followed by ultraspin to isolate the casein fraction for Bradford protein assay; Our earlier analysis of nine mature human milk gave 2.33 ± 1.69 g/L by isoelectric precipitation (pH 4.6), 1.80 ± 0.48 g/L by ultraspin sedimentation, and 2.96 ± 1.08 g/L by indirect micro-Kjeldahl nitrogen analysis ²⁸. We later applied a more acidic pH of 4.3 to precipitate human milk casein ^{18 7}, and obtained a total casein value of 4.46 ± 0.76 g/L ⁷. These studies all compartmentalized the casein fraction before quantification, and consequently the measurements would have over-estimated the true casein values. As shown by mass spectrometry, nearly 80 soluble proteins remain associated with acid precipitated casein micelles ⁶. Our current study implicates that the proportion of non-casein proteins in the casein and non-casein proteins simultaneously in the same milk samples.

We found a small but significant decrease in total casein quantity over the course of lactation (Figure 5B); however, when taking into account the accompanying change in total protein quantity, the percentage of casein in human milk total protein remains

relatively stable (Figure 5D), with an average value of 15.29±4.31%, ranging from 5.12% to 26.69% (Figure 5C, Table S3).

Re-evaluation of the whey/casein ratio in human milk

Another important objective of this study was to advance our knowledge of the human milk whey/casein ratio through this detailed analysis of a large number of samples. Since the casein level is very high in cow milk, the whey/casein ratio is a critical consideration when adapting cow milk based infant formula to make it more similar to human milk ^{29 2}, together with other important considerations such as the level of bioactive proteins/peptides³. In this study, we accurately measured casein subunits, but not the "casein fraction" as has often been done in the past. We found the whey/casein ratio to be considerably higher than previously believed, ranging from 73:27 to 95:5 in all 88 samples tested. The average as well as the median whey/casein ratio was 86:14 (Supplemental Table 1). The whey/casein ratio tends to be highest (91:9) at Day 1 and gradually decreases, although with some fluctuations, towards late lactation (Figure 6B, Supplemental Table 3). However, this decreasing trend did not reach significance (Figure 6C). Consistent with other measurements, the inter-individual variation is considerable. There are most likely several reasons for the previous over-estimation of the whey/casein ratio in human milk. Firstly, as mentioned previously, all precipitation/sedimentation methods used for separating casein will result in significant contamination with whey proteins. Secondly, all gelbased methods (PAGE, etc.) include staining of the proteins and the staining intensity of each protein varies considerably due to their amino acid composition, posttranslation modifications, etc., which makes accurate quantitation very difficult. Thirdly, chromatography methods are based on the UV absorbance of each separated

protein/subunit, which also varies with amino acid composition and post-translational modifications. We therefore believe that our approach provides a more accurate determination of the true whey/casein ratio of human milk.

Conclusions

Label-free LC-MS/MS quantitation gave us an opportunity to make significant progress on more accurate determination of human milk casein subunits. The experimental approach in this study does not require compartmentalization of milk samples, and measures casein concentrations individually and directly, therefore vastly improving the accuracy. Our analysis show generally lower measurements of caseins compared to previous reports. The whey/casein ratio ranged from 73:27 to 97:3 without significant changes throughout the lactation.

Limitations and Outlook

Although the scope of this study is quite wide, sampling twenty-two time points during the first year of lactation, it has a few limitations: 1. A sample size of four at each time point is rather small; however, it illustrates the feasibility of the approach. For large scale recommendations to be made, far more number of subjects are needed to account for inter-individual variability. 2. Technical replicates of each sample are desired. 3. Increased resolution of study time points is desired, especially during the second half of the twelve-month lactation duration. 4. Independent experiments are needed to support and validate the absolute value of caseins. 5. Subject clinical data are needed to evaluate confounding and covariate factors, e.g., age, gender of infants, dietary patterns, physical activity, ethnicity, stress), that may relate to the changes in milk protein abundances. Tracking subjects across time to identify the subjects

contributing each sample would also be beneficial in evaluating inter-individual variability.

Supporting information

Supplemental Table 1. Measured values (alpha casein, beta casein, kappa casein, total protein) and calculated values (total casein, percentage of total casein in total protein, whey/casein ratio) for all individual subjects.

Supplemental Table 2. Average protein concentrations at all studied lactation time points.

Supplemental Table 3. Average values of total casein and total casein in total human milk protein at all studied lactation time points.

References

1. Butte, N. F.; Lopez-Alarcon, M. G.; Garza, C., Nutrient adequacy of exclusive breastfeeding for the term infant during the first six months of life. *World Health Organization* **2002**. Page 16.

2. Lönnerdal, B., Nutritional and physiologic significance of human milk proteins. *Am J Clin Nutr* **2003**, 77, (6), 15378-15438.

3. Lönnerdal, B., Bioactive proteins in human milk: health, nutrition, and implications for infant formulas. *J Pediatr* **2016**, 173 Suppl, S4-9.

4. Lönnerdal, B., Infant formula and infant nutrition: bioactive proteins of human milk and implications for composition of infant formulas. *Am J Clin Nutr* 2014, 99, (3), 712S-7S.

5. McMahon, D. J.; Oommen, B. S., Supramolecular structure of the casein micelle. *J Dairy Sci* **2008**, 91, (5), 1709-21.

Liao, Y.; Alvarado, R.; Phinney, B.; Lönnerdal, B., Proteomic characterization of specific minor proteins in the human milk casein fraction. *J Proteome Res* 2011, 10, (12), 5409-15.

7. Kunz, C.; Lönnerdal, B., Human-milk proteins: analysis of casein and casein subunits by anion-exchange chromatography, gel electrophoresis, and specific staining methods. *Am J Clin Nutr* **1990**, *5*1, (1), 37-46.

8. Poth, A. G.; Deeth, H. C.; Alewood, P. F.; Holland, J. W., Analysis of the human casein phosphoproteome by 2-D electrophoresis and MALDI-TOF/TOF MS reveals new phosphoforms. *J Proteome Res* **2008**, *7*, (11), 5017-27.

9. Molinari, C. E.; Casadio, Y. S.; Hartmann, B. T.; Arthur, P. G.; Hartmann, P. E., Longitudinal analysis of protein glycosylation and beta-casein phosphorylation in term and preterm human milk during the first 2 months of lactation. *Br J Nutr* **2013**, 110, (1), 105-15.

10. López-Expósito, I.; Quirós, A.; Amigo, L.; Recio, I., Casein hydrolysates as a source of antimicrobial, antioxidant and antihypertensive peptides. *Dairy Science and Techonology* **2007**, 87, (4-5), 241-249.

 Cheng, X.; Gao, D.; Chen, B.; Mao, X., Endotoxin-Binding Peptides Derived from Casein Glycomacropeptide Inhibit Lipopolysaccharide-Stimulated Inflammatory Responses via Blockade of NF-kappaB activation in macrophages. *Nutrients* 2015, 7, (5), 3119-37.

12. Hvas, C. L.; Dige, A.; Bendix, M.; Wernlund, P. G.; Christensen, L. A.; Dahlerup, J. F.; Agnholt, J., Casein glycomacropeptide for active distal ulcerative colitis: a randomized pilot study. *Eur J Clin Invest* **2016**, 46, (6), 555-63. 13. Hansen, M.; Sandstrom, B.; Lönnerdal, B., The effect of casein phosphopeptides on zinc and calcium absorption from high phytate infant diets assessed in rat pups and Caco-2 cells. *Pediatr Res* **1996**, 40, (4), 547-52.

14. Lönnerdal, B.; Hernell, O., An Opinion on "Staging" of Infant Formula: A
Developmental Perspective on Infant Feeding. *J Pediatr Gastroenterol Nutr* 2016, 62, (1), 9-21.

Braisted, J. C.; Kuntumalla, S.; Vogel, C.; Marcotte, E. M.; Rodrigues, A. R.;
 Wang, R.; Huang, S. T.; Ferlanti, E. S.; Saeed, A. I.; Fleischmann, R. D.; Peterson, S.
 N.; Pieper, R., The APEX Quantitative Proteomics Tool: generating protein
 quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics* 2008, 9, 529.

Arike, L.; Valgepea, K.; Peil, L.; Nahku, R.; Adamberg, K.; Vilu, R.,
 Comparison and applications of label-free absolute proteome quantification methods
 on Escherichia coli. *J Proteomics* 2012, 75, (17), 5437-48.

17. Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003, 75, (17), 4646-58.

18. Kunz, C.; Lönnerdal, B., Human milk proteins: separation of whey proteins and their analysis by polyacrylamide gel electrophoresis, fast protein liquid chromatography (FPLC) gel filtration, and anion-exchange chromatography. *Am J Clin Nutr* **1989**, 49, (3), 464-70.

19. Goldfarb, M., Two-dimensional electrophoresis and computer imaging: quantitation of human milk casein. *Electrophoresis* **1999**, 20, (4-5), 870-4.

20. Goldfarb, M. F., Analysis of casein using two-dimensional electrophoresis, western blot, and computer imaging. *Adv Exp Med Biol* **2001**, 501, 535-9.

21. Morcol, T.; He, Q.; Bell, S. J., Model process for removal of caseins from milk of transgenic animals. *Biotechnol Prog* **2001**, 17, (3), 577-82.

22. Cuilliere, M. L.; Tregoat, V.; Bene, M. C.; Faure, G.; Montagne, P., Changes in the kappa-casein and beta-casein concentrations in human milk during lactation. *J Clin Lab Anal* **1999**, 13, (5), 213-8.

23. Altendorfer, I.; Konig, S.; Braukmann, A.; Saenger, T.; Bleck, E.; Vordenbaumen, S.; Kubiak, A.; Schneider, M.; Jose, J., Quantification of alphaS1casein in breast milk using a targeted mass spectrometry-based approach. *J Pharm Biomed Anal* **2014**, 103C, 52-58.

Ahrne, E.; Molzahn, L.; Glatter, T.; Schmidt, A., Critical assessment of proteome-wide label-free absolute abundance estimation strategies. *Proteomics* 2013, 13, (17), 2567-78.

25. Kunz, C.; Lönnerdal, B., Re-evaluation of the whey protein/casein ratio of human milk. *Acta Paediatr* **1992**, 81, (2), 107-12.

26. Khan, S.; Casadio, Y. S.; Lai, C. T.; Prime, D. K.; Hepworth, A. R.; Trengove, N. J.; Hartmann, P. E., Investigation of short-term variations in casein and whey proteins in breast milk of term mothers. *J Pediatr Gastroenterol Nutr* **2012**, *55*, (2), 136-41.

27. Eigel, W. N.; Butler, J. E.; Ernstrom, C. A.; Farrell, H. M.; Harwalkar, V. R.; Jenness, R.; Whitney, R. M., Nomenclature of proteins of cow's milk: fifth revision. *Journal of Dairy Science* **1984**, 67, (8), 1599-1631.

28. Lönnerdal, B.; Forsum, E., Casein content of human milk. *Am J Clin Nutr*1985, 41, (1), 113-20.

29. Goedhart, A. C.; Bindels, J. G., The composition of human milk as a model for the design of infant formulas: recent findings and possible applications. *Nutr Res Rev* **1994**, 7, (1), 1-23.

Acknowledgements

The authors thank Tina Du for assistance in sample processing, Hardip Patel for the advise of using R. W.X. is an Australian Research Council (ARC) Development Early Career Research Awards (DECRA) fellow (DE160100382). The study was funded in part by an unrestricted grant from Mead Johnson Nutrition to B.L.

Conflict of Interest Disclosure

The authors declare no conflict of interests.

Data Availability

The mass spectrometry proteomics data have been deposited to the University of

California, San Diego MassIVE datasets.

MassIVE ID: MSV000080058

Dataset Password: jp2NQ4bmvp2t

Please access via the link below:

http://massive.ucsd.edu/ProteoSAFe/status.jsp?task=9a781ca339b34783a96a4a1a846 51c2d

Figure legends

Figure 1. Experimental outline for human milk casein and total protein quantification.

- A. Schematic of sample processing for mass spectrometry.
- B. Human milk sample collection schedule. n=4 for each time point.
- C. Calculation strategy of this study.
- D. SDS-PAGE of whole human milk.

Figure 2. Total protein concentration in human milk.

- A. Dot plot depicting absolute total protein concentrations of all eighty-eight samples.
 Colors indicate four biological replicates.
- B. Scatter plot of average total protein concentration arranged by time of milk sample collection. Values are means±SD, n=4.
- C. Scatter plot of individual total protein concentration arranged by time of milk sample collection. The solid line depicts the linear regression fit, and the dashed lines show the 95% confidence intervals for the linear regression fit. Regression slope (change in protein per day)=0.0121, p=0.002.

Figure 3. Casein subunits concentration of all studied human milk samples. Dot plot depicting absolute casein concentration of all eighty-eight samples. Color scheme is as in Figure 2A. Top, alpha-casein; middle, beta-casein; bottom, kappacasein.

Figure 4. Casein subunits average concentration and trend of expression during lactation.

- A. Scatter plot of average alpha-casein concentration arranged by time of milk sample collection. Values are means±SD, n=4.
- B. Scatter plot of individual alpha-casein concentration arranged by time of milk sample collection. The solid line depicts the linear regression fit, and the dashed

lines show the 95% confidence intervals for the linear regression fit. Regression slope=-2.00E-04, p=0.090.

- C. Scatter plot as in A for beta-casein.
- D. Scatter plot as in B for beta-casein. Regression slope=-5.00E-04, p=0.134.
- E. Scatter plot as in A for kappa-casein.
- F. Scatter plot as in B for kappa-casein. Regression slope=-7.00E-04, p=0.004.

Figure 5. Total casein in human milk.

- A. Scatter plot of average total casein concentration arranged by time of milk sample collection. Values are means±SD, n=4.
- B. Scatter plot of individual total casein concentration arranged by time of milk sample collection. The solid line depicts the linear regression fit, and the dashed lines show the 95% confidence intervals for the linear regression fit. Regression slope=-0.0014, p=0.037.
- C. Scatter plot as in A for the percentage of total casein in total milk protein (weight/weight, or w/w).
- D. Scatter plot as in B for the percentage of total casein in total milk protein. Regression slope=5.00E-04, p=0.924.

Figure 6. Ratio of whey/casein in human milk calculated as (total proteincasein)/casein.

- A. Dot plot depicting whey/casein ratio of all eighty-eight samples. Colors scheme is as in Figure 2A.
- B. Scatter plot of whey/casein ratio arranged by time of milk sample collection.
 Values are means±SD, n=4.
- C. Scatter plot of individual whey/casein ratio arranged by time of milk sample collection. The solid line depicts the linear regression fit, and the dashed lines

show the 95% confidence intervals for the linear regression fit. Regression slope =0.0022, p=0.431.

Supplemental Table 1. Values of measured concentration for alpha-casein, betacasein, kappa-casein, and total protein. Set V, N, P, Y indicate biological replicates.

Supplemental Table 2. Average protein concentrations at all studied lactation time points. Average concentrations of measured parameters, total protein, alpha-casein, beta-casein, and kappa-casein at all twelve time points are presented.

Supplemental Table 3. Average values of total casein, and total casein in total human milk protein at all studied lactation time points. The calculated parameters, average total casein concentration, and the proportion of total casein in total protein at all twelve time points are presented.

For TOC only







Figure 2





Figure 4



Figure 5



Figure 6

Supporting Information

Absolute quantification of human milk caseins and the whey/casein ratio during the first year of lactation

Yalin Liao^{1,2}, Darren Weber³, Wei Xu⁴, Blythe P Durbin-Johnson⁵, Brett S. Phinney³, Bo Lönnerdal¹*.

- 1. Department of Nutrition, University of California, Davis, 95616, USA
- John Curtin School of Medical Research, Australian National University, Canberra ACT 2601, Australia
- Genome Center Proteomics Core Facility, University of California, Davis, 95616, USA
- School of Veterinary & Life Sciences, Murdoch University, Murdoch WA 6150, Australia
- Division of biostatistics, Department of Public Health Sciences, School of Medicine, University of California, Davis, 95616, USA
- * Correspondence: bllonnerdal@ucdavis.edu (B.L.)

Table of Content

Supplemental Table 1. Measured values (alpha casein, beta casein, kappa casein, total protein) and calculated values (total casein, percentage of total casein in total protein, whey/casein ratio) for all individual subjects.

Supplemental Table 2. Average protein concentrations at all studied lactation time points.

Supplemental Table 3. Average values of total casein and total casein in total human milk protein at all studied lactation time points.

Time	Total Protein	alpha-casein	beta-casein	kappa-casein			
(post partum day)		g/L, n=4					
D1	13.20±2.37*	0.24±0.11	0.47±0.27	0.6±0.15			
D2	11.71±5.80	0.26±0.15	0.55±0.40	0.59±0.31			
D3	11.26±3.73	0.30±0.20	0.62±0.35	0.65±0.25			
D4	10.09±3.67	0.22±0.10	0.55±0.17	0.50±0.14			
D5	12.43±3.32	0.30±0.13	0.78±0.25	0.69±0.12			
D6	9.72±4.57	0.21±0.07	0.62±0.23	0.60±0.19			
D7	9.09±2.10	0.21±0.01	0.53±0.10	0.56±0.11			
D8	10.63±3.86	0.27±0.11	0.67±0.20	0.65±0.21			
D9	10.63±2.34	0.26±0.10	0.64±0.14	0.66±0.15			
D10	12.06±5.34	0.31±0.10	0.86±0.37	0.72±0.28			
D11	8.90±4.97	0.24±0.12	0.70±0.42	0.55±0.25			
D12	9.02±3.69	0.24±0.07	0.69±0.29	0.57±0.21			
D13	7.38±4.61	0.15±0.06	0.39±0.22	0.40±0.19			
D14	9.92±3.99	0.26±0.08	0.75±0.20	0.60±0.24			
D30	8.86±1.54	0.22±0.07	0.69±0.17	0.55±0.16			
D60	7.24±2.13	0.19±0.11	0.58±0.30	0.44±0.14			
D90	6.92±2.15	0.22±0.11	0.64±0.30	0.47±0.20			
D120	8.13±5.86	0.40±0.38	1.00±0.88	0.56±0.51			
D150	7.53±0.76	0.19±0.08	0.48±0.18	0.43±0.14			
D180	6.64±1.53	0.17±0.08	0.43±0.15	0.41±0.09			
D270	7.04±1.56	0.18±0.07	0.46±0.22	0.41±0.12			
D360	7.77±1.33	0.16±0.08	0.44±0.30	0.42±0.15			

Supplemental Table 2. Average protein concentrations at all studied lactation time points.

* Mean±SD.

Time	Total casein #	Total casein in total human milk protein
(post partum day)	g/L, n=4	w/w, n=4
D1	1.31±0.49*	0.10±0.04
D2	1.39±0.83	0.12±0.03
D3	1.56±0.77	$0.14{\pm}0.04$
D4	1.26±0.39	0.13±0.03
D5	1.75±0.47	0.15±0.04
D6	1.41 ± 0.48	0.16±0.03
D7	1.29±0.19	$0.15{\pm}0.03$
D8	1.58±0.48	0.16±0.03
D9	1.54±0.37	0.15±0.03
D10	1.87±0.73	0.17±0.04
D11	1.48 ± 0.76	$0.18{\pm}0.07$
D12	1.49±0.55	0.17±0.05
D13	0.93 ± 0.45	$0.15{\pm}0.04$
D14	1.60 ± 0.48	0.18±0.04
D30	1.45±0.34	0.17±0.05
D60	1.20±0.48	0.17±0.06
D90	1.32±0.59	0.19±0.05
D120	1.94±1.73	0.22±0.07
D150	1.09±0.36	$0.15{\pm}0.05$
D180	1.00±0.29	0.15±0.03
D270	1.04±0.39	0.15±0.03
D360	1.01±0.52	0.13±0.06

Table S3. Average values of total casein and total casein in total human milk protein at all studied lactation time points.

See Figure 1C for scheme of calculation.* Mean±SD.

	Time (neet	Measured Values				
	lime (post					
	partum dov)	alpha	beta casein	kappa	Total	Total Caseins
	uay)	casein (g/L)	(g/L)	casein (g/L)	Protein (g/L)	(g/L)
	D1	0.13	0.07	0.39	11.15	0.59
	D2	0.07	0.06	0.17	3.60	0.31
	D3	0.13	0.19	0.33	6.06	0.65
	D4	0.14	0.39	0.37	7.89	0.91
	D5	0.25	0.78	0.74	12.53	1.78
	D6	0.21	0.64	0.58	9.20	1.44
	D7	0.20	0.58	0.54	9.00	1.32
	D8	0.41	0.78	0.78	13.16	1.97
	D9	0.30	0.62	0.71	10.41	1.64
	D10	0.44	1.33	0.99	14.71	2.76
Sot V	D11	0.12	0.14	0.31	4.53	0.57
Serv	D12	0.16	0.35	0.33	5.65	0.84
	D13	0.15	0.24	0.39	5.62	0.78
	D14	0.34	0.90	0.84	12.70	2.08
	D30	0.23	0.74	0.67	10.04	1.64
	D60	0.21	0.47	0.59	9.15	1.27
	D90	0.30	0.90	0.70	9.48	1.90
	D120	0.59	0.96	0.41	8.36	1.96
	D150	0.19	0.44	0.54	8.31	1.16
	D180	0.13	0.22	0.37	4.48	0.72
	D270	0.22	0.61	0.52	8.43	1.35
	D360	0.18	0.39	0.44	6.23	1.01
	D1	0.38	0.59	0.62	11.75	1.59
	D2	0.41	0.62	0.64	11.78	1.67
	D3	0.58	1.02	0.91	13.57	2.52
	D4	0.32	0.66	0.51	8.81	1.48
	D5	0.47	1.04	0.77	11.71	2.27
	D6	0.18	0.43	0.42	5.79	1.03
	D7	0.20	0.47	0.43	6.22	1.10
	D8	0.18	0.40	0.37	4.99	0.95
	D9	0.32	0.75	0.77	10.41	1.84
	D10	0.21	0.44	0.39	5.22	1.03
Set N	D11	0.19	0.65	0.37	4.74	1.21
	D12	0.26	0.70	0.48	6.31	1.44
	D13	0.07	0.19	0.13	2.03	0.39
	D14	0.17	0.49	0.29	4.11	0.94
	D30	0.30	0.89	0.56	7.58	1.74
	D60	0.33	1.01	0.49	7.68	1.83
	D90	0.32	0.89	0.53	7.04	1.74
	D120	0.82	2.23	1.30	16.32	4.36
	D150	0.25	0.49	0.40	6.65	1.14

	D180	0.27	0.55	0.52	8.07	1.34
	D270	0.18	0.30	0.35	6.34	0.83
	D360	0.07	0.07	0.22	7.10	0.36
	D1	0.25	0.64	0.71	16.42	1.61
	D2	0.20	0.48	0.64	14.55	1.31
	D3	0.27	0.67	0.65	14.28	1.60
	D4	0.28	0.70	0.69	15.55	1.68
	D5	0.28	0.81	0.71	16.75	1.81
	D6	0.29	0.91	0.86	16.24	2.06
	D7	0.20	0.42	0.57	10.01	1.19
	D8	0.27	0.83	0.83	13.08	1.92
	D9	0.26	0.73	0.68	13.69	1.67
	D10	0.27	0.81	0.58	10.79	1.66
Sot D	D11	0.38	1.13	0.83	12.44	2.35
SelF	D12	0.21	0.64	0.64	10.72	1.49
	D13	0.17	0.44	0.51	9.10	1.12
	D14	0.28	0.70	0.68	10.58	1.66
	D30	0.16	0.50	0.33	7.48	0.99
	D60	0.09	0.35	0.28	4.20	0.73
	D90	0.12	0.36	0.26	4.23	0.74
	D120	0.11	0.54	0.27	3.62	0.92
	D150	0.21	0.70	0.53	7.98	1.44
	D180	0.17	0.53	0.41	7.10	1.10
	D270	0.20	0.67	0.49	8.21	1.36
	D360	0.24	0.80	0.58	8.72	1.62
				-		
	D1	0.20	0.56	0.66	13.45	1.42
	D2	0.33	1.02	0.91	16.90	2.27
	D3	0.21	0.57	0.69	11.11	1.46
	D4	0.12	0.41	0.42	8.08	0.95
	D5	0.16	0.45	0.52	8.73	1.13
	D6	0.13	0.47	0.51	7.64	1.10
	D7	0.22	0.63	0.69	11.12	1.53
	D8	0.21	0.67	0.58	11.28	1.46
	D9	0.12	0.45	0.44	8.01	1.00
	D10	0.30	0.83	0.89	17.52	2.01
Set Y	D11	0.24	0.85	0.67	13.89	1.76
	D12	0.32	1.03	0.81	13.39	2.16
	D13	0.19	0.67	0.55	12.76	1.41
	D14	0.24	0.90	0.56	12.29	1.70
	D30	0.19	0.61	0.62	10.31	1.41
	D60	0.11	0.45	0.39	7.91	0.96
	D90	0.12	0.41	0.36	6.91	0.89
	D120	0.06	0.25	0.22	4.22	0.53
	D150	0.08	0.28	0.24	7.15	0.60

D180	0.09	0.41	0.31	6.90	0.81
D270	0.08	0.24	0.27	5.18	0.59
D360	0.14	0.49	0.41	9.02	1.05

Calculated Values	
Total casein in total	
human milk protein (%	Whey/casein
w/w)	ratio
5.30%	82:18
8.51%	89:11
10.69%	92:8
11.49%	92:8
14.19%	94:6
15.61%	95:5
14.62%	94:6
14.97%	94:6
15.76%	94:6
18.77%	96:4
12.58%	93:7
14.84%	94:6
13.91%	94:6
16.37%	95:5
16.33%	95:5
13 89%	94.6
20.02%	96:4
23.47%	97:3
13 98%	94.6
16.18%	94.0
16.05%	95.5
16.19%	95:5
10.1370	00.0
13.50%	86:14
14.18%	86:14
18.56%	81:19
16.83%	83.17
19.41%	81.19
17 81%	82.18
17.5%	82.10
10 12%	81.10
17.66%	82.18
10.82%	80·20
19.02 /0	00.20 75:25
23.40%	75.25
22.80%	11.23
19.15%	01:19
22.89%	77:00
23.01%	11:23
23.82%	/6:24
24.69%	75:25
26.69%	73:27
17.11%	73:27

16.57%	73:27
13.13%	87:13
5.12%	95:5
9.80%	90:10
9.03%	91:9
11.22%	89:11
10.77%	89:11
10.80%	89:11
12.70%	87:13
11.90%	88:12
14.69%	85:15
12.19%	88:12
15.41%	85:15
18.87%	81:19
13.93%	86:14
12.34%	88:12
15.70%	84:16
13.21%	87:13
17.30%	83:17
17 53%	82.18
25.34%	75:25
18 01%	82.18
15.01%	85:15
16.55%	83.17
18 53%	81.10
18:55 %	01.19
10.54%	80.11
13 / 1%	87·13
13.41%	87·13
11 78%	<u> </u>
12,00%	00.12
13.00%	07.13
14.43%	00.14
13.78%	80:14
12.95%	87:13
12.53%	87:13
11.50%	88:12
12.66%	87:13
16.16%	84:16
11.05%	89:11
13.80%	86:14
13.71%	86:14
12.11%	88:12
12.92%	87:13
12.44%	88:12
8.36%	92:8

11.68%	88:12
11.34%	89:11
11.65%	88:12