Differences in post-translational modifications of proteins in milk from early and mid-lactation dairy cows as studied using total ion chromatograms from LC-ESI/MS

Martin N. Thesbjerg a, b, *, Marianne Johansen c, Lotte B. Larsen a, Nina A. Poulsen a

a Department of Food Science, Aarhus University, Agro Food Park 48, DK-8200, Aarhus N, Denmark
b Sino-Danish College (SDC), University of Chinese Academy of Sciences
c Department of Animal Science, AU Foulum, Aarhus University, Blichers Allé 20, DK-8830, Tjele, Denmark

A R T I C L E   I N F O
Article history:
Received 26 July 2021
Received in revised form 9 October 2021
Accepted 20 October 2021
Available online 3 November 2021

A B S T R A C T
Liquid chromatography top-down mass-spectrometry combined with an in-house library with 5180 entries of known milk protein masses was used to identify common and rare isoforms of major milk proteins from individual cow's milk samples from early- and mid-lactating (EL and ML, respectively) cows. The milk samples were from 24 Danish Holstein-Friesians, with 12 cows each in EL and ML. A Latin square design with four periods was used to obtain one composite milk sample from each cow per period. The deconvoluted masses were used for identifying common and rare isoforms of the major milk proteins (α s1-CN, α s2-CN, β-CN, κ-CN, α-La, and β-LG). α s2-CN splice variants were found in all samples, α s2- and κ-CN isoforms differed between EL and ML, κ-CN post-translational modifications were found to relate to κ-CN genotype. This approach could identify 66 isoforms of major milk proteins, where 15 were of the α s2-CNs and 39 from κ-CN.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Bovine milk is a complex mixture of minor and major proteins, even when focusing on the four caseins (α s1-CN, α s2-CN, β-CN and κ-CN) and two major whey proteins (α-La and β-LG) alone, a wide range of isoforms are present. One part of the complexity is caused by the different genetic variants of these proteins (Farrell et al., 2004), while another is related to different combinations of post-translational modifications (PTMs). Within the Bos genus, a total of 49 common or rare genetic variants of the major milk proteins have been identified (8 for α s1-CN, 4 for α s2-CN, 12 for β-CN, 11 for κ-CN, 11 for α-La and 3 for β-LG) (Farrell et al., 2004). PTMs are added to these proteins during biosynthesis, including disulphide bond formation, phosphorylations and glycosylations, which are all of known importance for milk protein functionalities (Bijl, Holland, & Boland, 2020; Le, Deeth, & Larsen, 2017). Natural casein glycosylations are found as O-glycosylation on κ-CN, and five different structural glycosylation forms, denoted a, b, c, d and e are commonly found in bovine milk (Saito & Itoh, 1992). These include:

- a = GalNAc; b = Gal[(1→3)GalNAc; c = NeuAcα[(2→3)]Gal[(1→3)]GalNAc; d = Gal[(1→3)]NeuAcα[(2→6)]GalNAc; e = NeuAcα[(2→3)]Gal[(1→3)]NeuAcα[(2→6)]GalNAc.

In relation to phosphorylation of Ser/Thr residues in the milk proteins, the casein kinase motif has been identified as Ser/Thr-X-Glu/Asp/SerP, where X can be any amino acid (Mercier, 1981). Genetic polymorphisms may influence the PTMs attached, as exemplified by the α s2-CN C and D variants (Farrell et al., 2004). The α s2-CN C variant has a Thr residue at residue 47, which grants the C variant an additional potential site for phosphorylation matching this required motif. In the α s2-CN D variant, residues 51–59 have been deleted comprising three of the known sites of phosphorylation (Ser56, Ser57 and Ser58). For α s1-CN, it has been shown that different genetic marker regions are linked to the expression of the α s1-CN 8P and α s1-CN 9P isoforms, possibly reflecting the involvement of different CN kinases (Bijl, van Valenberg, Huppertz, van Hooijdonk, & Bovenhuis 2014b).

Even though genetic factors are crucial governing milk protein heterogeneity, environmental and physiological factors, like lactation stage and dry period, have also been linked to glycosylation type and glycosylation degree (GD) of κ-CN (Bonfatti, Chiarot, & Carnier, 2014; Maciel et al., 2017; Poulsen, Giagnoni, Johansen, Lund, & Larsen, 2021; Poulsen, Jensen, & Larsen, 2016). GD of κ-
CN has been associated with CN micelle size (Bijl, de Vries, van Valenberg, Huppertz, & van Hooijdonk, 2014a) and rennet-induced coagulation properties (Poulsen et al., 2016). Furthermore, the phosphorylation degrees (PD) of both S2-CN and S2-CN have been shown to be affected by lactation stage (Fang et al., 2017). Using LC-MS, Fang et al. (2017) demonstrated the presence of as many as six phosphorylation isoforms of S2-CN (9–15P) along with the two major phosphorylation isoforms of S2-CN (8 and 9P). The relative abundance of S2-CN 9P and of S2-CN 13 and 14P was further found to increase over lactation, with the corresponding relative abundances of S2-CN 8P and of S2-CN 10 and 11P to decrease, thereby significantly affecting their PDs.

A common approach for the detailed analysis of isoforms of the major milk proteins and how they may vary is to use LC coupled with either a UV (normally at 214 nm) or a mass spectrometer (MS), or both (Jensen et al., 2012; Nilsson et al., 2020). This allows for the detection of the UV and/or the total ion chromatogram (TIC) data from top-down MS data. Quantification based on UV is widely used, but has its limitations, as it is not always possible to chromatographically separate PTM isoforms and genetic variants, like, e.g., the S2-CN phosphorylation isoforms, complex glycosylation patterns of κ-CN or even rare splice variants, which have recently been identified and quantified using the TIC by top-down MS using a reverse phase LC coupled to an ESI-TOF (Miranda, Bianchi, Krupova, Trossat, & Martin, 2020).

The aim of the present study was to build on this approach and investigate presence and amount of isoforms of the six major proteins in individual cow’s milk based on the TICs from liquid chromatography electrospray mass spectrometry (LC-ESI/MS) in relation to lactation stage (EL and ML).

2. Materials and methods

2.1. Milk samples

In total 24 Danish Holstein cows, 6 primiparous and 6 multiparous in EL, and 12 multiparous in ML, were subjected to an experimental feeding trial in a 6 × 4 incomplete Latin square design with 4 periods (Giagnoni, Lund, Sehested, & Johansen, 2021). At the beginning of the experiment, EL cows were 23.3 ± 6.7 (mean ± SD) days in milk (DIM) and ML cows were 176 ± 15 DIM. In the current study, cows are only assigned to lactation group and MS data are not explored relative to feeding, which was explored by Poulsen et al. (2021). The cows were milked twice daily (at 06:00 and 17:00 h) in a herringbone milking parlour, where milk yield was recorded at each individual milking. By the end of each period, an afternoon and a morning milk sample were collected from each cow and, immediately after morning milking, a representative sample was prepared by mixing the 2 milk samples proportional to milk yield. These compositied samples were used for further milk analyses giving a total of 96 individual cow’s samples. After mixing the morning and afternoon samples, skim milk samples were prepared by removal of fat after centrifugation at 2643 × g for 30 min at 4 °C. Skim milk samples were frozen for later analysis (–80 °C).

2.2. Protein compositional analysis by LC-ESI/MS

As previously described by Jensen et al. (2012), the protein composition of the skim milk samples was determined using reversed phase LC-ESI/MS Single Q. Proteins were separated using an HPLC 1100 system (Agilent Technologies, Santa Clara, CA) with a Jupiter C4 column (250 mm × 2 mm, 5 μm particle size, 300 Å pores; Phenomenex, Torrance, CA) operated at 40 °C with a G1315A diode-array detector for UV detection at 214 nm coupled to a mass selective detector for identification and relative quantification of the milk proteins. The LC gradient used was a linear gradient from 33% B (B = 0.05% trifluoroacetic acid (TFA) in acetonitrile (ACN), A = 0.05% TFA in LCMS grade water) to 50% B over 25 min.

A database was constructed using amino acid sequences representing S2-CN, S2-CN, β-CN, κ-CN, as well as α-LA and β-LG, as downloaded from Uniprot.org (accessed on 27-04-2020) and modified to match and represent the different genetic variants previously identified in Danish Holstein (Jensen et al., 2012). Post-processing of imported sequences included removal of signal peptides, representation of single point mutations (Farrell et al., 2004) and inclusion of pyroGlu as the N-terminal of κ-CN. As all samples were reduced with DTE prior to LC-ESI/MS, cysteine bridges were not included, and all masses were based on reduced forms. The database contained the known possible phosphorylation variants: S2-CN (7–10P), S2-CN (9–15P), β-CN (4–6P), κ-CN (0–3P), and α-LA (0–1P) (Fang et al., 2017; Farrell et al., 2004; Miranda et al., 2020). For κ-CN, up to 3 glycosylations were allowed, representing possible combinations of the five reported glycosylation forms: a, b, c, d, and e (Saito & Itoh, 1992). The masses of the c and d glycosylated ptms are identical, therefore they are referred to as c/d glycosylated ptms. Since lactosylations on β-LG have been reported in milk (Miranda et al., 2020), one lactosylation (324 Da each) for β-LG was included in the database. In addition, earlier reported fragments of plasmin cleavage (S2-CN 1P fragment 80–199) or splice variants (S2-CN 8P – exon 12, S2-CN 9P – exon 12, S2-CN 8P – Q59 or Q78 and S2-CN 9P – Q59 or Q78) of S2-CN were included as well (Gaiaschi et al., 2000; Martin, Bianchi, Cebo, & Miranda, 2013a; Martin, Cebo, & Miranda, 2013b). In total, 518 different masses were included in the database. To allow for common method-induced modifications, the database was further expanded to include loss of up to three water molecules, deamidations or sodium adducts per mass. The database was thus expanded to 5180 entries.

MassHunter 10 Bioconfirm software (Agilent) was used to generate the deconvoluted masses from the TICs of the recorded data (a representative TIC is shown in Supplementary material Fig. S1). The default data processing method for intact reduced proteins was used with the following modifications: charge stages allowed from 7 to 40, deconvolution mass range 10,000 to 35,000 Da, mass step 0.05 Da, 50% of peak height was selected for calculation of average mass, and mass tolerance ±1.5 Da. Results were exported to Excel, parts per million (ppm) error and abundance of the deconvoluted masses of identified biomolecules relative to total area of each sample were calculated. The milk samples were genotyped based on deconvoluted masses of observed protein variants using the constructed database. To make isoforms comparable across homozygotes and heterozygotes, the relative abundances of homozygotes were divided by two.

2.3. Statistical analysis

To investigate if differences in relative abundance were significant between EL and ML for a specific isoform, the following model was applied in R.

\[
y_{ijkl} = \mu + L_i + P_j + A_k + C_t + e_{ijkl}
\]  

here \(y_{ijkl}\) is the dependent variable, \(\mu\) is the overall mean, \(L\) is the fixed effect of lactation stage (\(i = EL, ML\)), \(P\) is the fixed effect of period (\(j = 1\) to \(4\)), \(A\) was the fixed effect of parity (\(k =\) primiparous, multiparous), \(C\) is the random effect of the cow (\(t = 1\) to 27, as 3 cows were substituted during the experiment) and \(e_{ijkl}\) is the random residual error. Only isoforms with at least five observations
in both EL and ML were tested. To test for the effect of genotype for some isoforms, the model was expanded to include genotypes.

\[ Y_{ijklmn} = \mu + t_i + P_j + A_k + C_l + G_m + e_{ijklmn} \]  

(2)

here \( G \) is the fixed effect of genotype (\( m = 1 \text{ to } 3 \)). In all cases when genotypes were compared, the former division of relative abundances by two for homozygotes was first reversed to make abundances comparable.

### 3. Results

After assessment of the data, six samples were removed, resulting in 90 samples being used for further analysis. Across the 90 skimmed milk samples, the obtained TIC from the LC-ESI/MS revealed a total of 170 unique isoforms. These represented genetic variants, splice variants, natural PTMs, like phosphorylation and glycosylation, as well as losses of water, deamidation and sodium adducts. In total, 1870 identifications could be assigned to these isoforms across all milk samples. Of these, 195 identifications — assigned as method-induced modifications (loss of water, deamidations and sodium adducts) — were discarded due to very low relative abundances. Thus, a total of 1675 identifications, corresponding to 66 isoforms of the six major milk proteins were kept and included in the analysis. An overview of all isoforms and their distribution between EL and ML can be found in Supplementary material Table S1.

The whole palette of these unique 66 isoforms identified, representing genetic variants and their natural PTMs (glycosylated and phosphorylated isoforms), splice- and cleavage-variants of \( \alpha_{51-CN} \), as well as lactosylations of \( \beta-LG \), as shown in Fig. 1. Furthermore, the range of chromatographic RT covering these are shown, including not only expected variations in RT between isoforms, but also some variations within isoforms. Overall, the first protein to elute was \( \kappa-CN \), with its various glycosylation and phosphorylation isoforms (4–10 min). In total, 39 isoforms were assigned to \( \kappa-CN \). The first \( \kappa-CN \) molecules to elute are the heavily glycosylated \( \kappa-CN \) A and B isoforms with up to three O-glycans attached. Of these, based on the mass identifications, the first isoforms to elute are rich in mainly e glycoforms and some c/d glycoforms. These are followed by \( \kappa-CN \) isoforms with glycosylations of the a and b type. Then, the non-glycosylated \( \kappa-CN \) A with 1 or 2P elutes in the same time range as \( \kappa-CN \) B with one or two glycans, one is the e or c/d, glycoforms, while the other can be of the a/b glycoforms with 0, 1 and 2P. These were followed by \( \kappa-CN \) B with one glycosylation, then \( \kappa-CN \) A without any PTMs, \( \kappa-CN \) B 0 or 1P with 1–3 glycoforms of the a, b and c/d type and then by non-glycosylated \( \kappa-CN \) B with 1 or 2P. Next is \( \alpha_{52-CN} \) (10–12 min), with all \( \alpha_{52-CN} \) isoforms representing genetic variant A, resulting in six isoforms assigned to \( \alpha_{52-CN} \), overall, in the order of lower to higher number of phosphorylations. This is followed by \( \alpha_{51-CN} \) isoforms (13–16 min), all representing \( \alpha_{51-CN} \) B, including some splice variants and a plasmin cleavage product. The first \( \alpha_{51-CN} \) molecule to elute was the \( \alpha_{51-CN} \) B 1P plasmin cleavage product 80–199, then \( \alpha_{51-CN} \) B 7 and 8P, followed by the splice variants \( \alpha_{51-CN} \) Q59 or Q78, then \( \alpha_{51-CN} \) B 9P, followed by splice variants \( \alpha_{51-CN} \) B 8P –exon12 and \( \alpha_{51-CN} \) B 9P –exon 12, and lastly \( \alpha_{51-CN} \) B 10P. Thus, nine isoforms were assigned to \( \alpha_{51-CN} \).

The last CN to elute was \( \beta-CN \) (17–19 min), the genetic variants of which eluted in the order: \( \beta-CN \) B, A1, A2 and then \( \beta-CN \) was found as the major 5P isoform and a minor 4P isoform. The 4P isoform was found only for \( \beta-CN \) A1 and A2, where they eluted before their 5P isoforms. In total, six isoforms were assigned to \( \beta-CN \). The CNs were followed by \( \beta-LG \) and \( \alpha-LA \); \( \beta-LG \) eluted first (19–22 min), with the \( \beta-LG \) genetic variant B before A. In addition, \( \beta-LG \) with addition of one lactose (+ 324 Da) was identified for both variants; in both cases the lactosylated isoform eluted prior to the non-lactosylated form, resulting in a total of four isoforms being assigned to \( \beta-LG \). The final protein to elute (21–22 min) was \( \alpha-LA \) variant B, where a low abundant \( \alpha-LA \) 1P isoform eluted after \( \alpha-LA \) B.

### 3.1. Presence and abundance of \( \alpha_{51-CN} \) isoforms and effect of lactation stage

Across the 90 milk samples, 438 observations were assigned to the nine isoforms of \( \alpha_{51-CN} \), of these 207 represented the four different phosphorylation isoforms (7–10 P). The relative abundances and percentage of samples with an observed isoform within each lactation group are displayed in Fig. 2. The major

![Fig. 1. Retention time for all identified isoforms across the 90 samples.](image-url)
phosphorylation isoforms S51-CN B 8P and 9P were found in all samples. In addition, the low abundant S51-CN B 7P and 10P isoforms were observed, though with a lower frequency (27.7% and 2.2%, respectively). When comparing the relative abundances between lactation groups, S51-CN B 8P (P < 0.05) were significantly more abundant in EL, S51-CN B 10P was found only in ML.

Apart from the four phosphorylation isoforms of S51-CN B, five additional masses, matching masses of S51-CN B splice- or plasmin cleavage-products were found. These splice or cleavage products were observed 231 times in the dataset. The identified plasmin cleavage product was a fragment of S51-CN, corresponding to a loss of the first 79 amino acids, i.e., S51-CN B 1P fragment (80–199) with a mass of 14107.95 Da. The splice products were S51-CN B of 8 and 9P isoforms, where exon 12 was missing, i.e., S51-CN B 8P –exon 12 with a mass of 21876.73 Da and S51-CN B 9P –exon 12 with mass of 21956.72 Da. Furthermore, S51-CN B 8P and 9P isoforms, where a single glutamine (Q or Gln) residue at 59 or 78 are missing, i.e. S51-CN B 8P –Q59 or 78Q, with mass of 23486.76 Da or S51-CN B 9P –Q59 or 78Q with a mass of 23566.75 Da were also observed. The relative abundances of these splice variants were relatively low. The splice products with the highest relative abundances were those with a missing Gln at positions 59 or 78. In addition, splice products representing S51-CN B 8P were more abundant and frequent than those representing S51-CN B 9P. The most abundant splice product was S51-CN B 8P –Q59 or 78Q, which was slightly more abundant than S51-CN B 7P, while the lowest abundant splice product was S51-CN B 9P –exon 12. Statistical analysis of these products revealed that S51-CN B 8P –exon 12 was significantly more abundant in EL than ML (P < 0.001). Further analysis of the splice- and cleavage-products (results not shown) tracked these to all but one sample. Thus, these splice- and cleavage-products were expressed alongside an otherwise intact S51-CN B protein profile and each sample typically contained two of these products.

3.2. Presence and abundance of S52-CN isoforms and effect of lactation stage

In total, 312 of the observed masses were assigned to either of the six S52-CN isoforms across the dataset, comprising S52-CN variant A 10–15P (Fig. 3). The most frequent of the S52-CN A isoforms were those with 11 and 12P, which were also the isoforms of highest abundance. However, also the 13P form, and, to some extent the 10P form, were both present in relatively high abundance, and were identified in more than half of the samples. The 14 and 15P isoforms were present, but in relatively low abundance, as well as with a lower frequency (14P EL = 6.98%, 14P ML = 25.53%, 15P EL = 2.33% and 15P ML = 2.13%). Only S52-CN A 11P and S52-CN A 12P had enough observations to be considered for statistical analysis, these were both significantly more abundant in EL (P11P < 0.01 and P12P < 0.05). In terms of frequency, it seems that S52-CN’s with higher P counts are more frequent in ML while the lower P counts are more frequent in EL.

3.3. Presence and abundance of β-CN isoforms and effect of lactation stage

A total of 134 observations was assigned to the six β-CN isoforms identified, representing four different genetic variants (A1, A2, B and I). These were found as six different genotypes in the dataset; A2A2, A1A2, A1I, A1I, II and A1B. The β-CN B was found only in the ML group, and, in general, the genetic variants were not balanced between the EL and ML groups, whereas the low abundant 4P isoform was found only in the ML group for the A1 and A2 variants only (Fig. 4). Due to the number of observations, only the β-CN A2 5P isoform qualified for
statistical analysis, and here no significant difference was observed between EL and ML.

3.4. Presence and abundance of \( \kappa \)-CN isoforms and effect of lactation stage

A total of 512 observations were assigned to 39 different \( \kappa \)-CN isoforms, including both \( \kappa \)-CN A (19 isoforms) and \( \kappa \)-CN B (20 isoforms), and were represented as the genotypes \( \kappa \)-CN AA, \( \kappa \)-CN BB and \( \kappa \)-CN AB. This by far makes \( \kappa \)-CN the protein with the largest isoform heterogeneity, as expected. Each genetic variant is found in isoforms with up to two phosphorylations and up to three glycosylations using the previously described glycoforms (a, b, c/d and e). Furthermore, among the \( \kappa \)-CN isoform molecules, only the A variant was found in the unmodified form. For \( \kappa \)-CN AA milk, 4.15 isoforms were, on average, observed per sample, whereas 5.70
isoforms on average were detected per κ-CN BB milk sample and 6.26 isoforms for κ-CN AB milk samples. The most frequent PTM on κ-CN is the 1P modification, followed by 2P, both without glycosylations. Among the glycoforms, the e glycoform was confirmed to be most frequent, while the most frequent isoform with P and G was 1P 2G (e, e), as observed for both κ-CN A and B genetic variants (Fig. 5). Among the many isoforms of κ-CNs, not many of the isoforms had enough observations to qualify for statistical analysis. The isoforms that qualified were: κ-CN A 1P, κ-CN A 2P, κ-CN A 1P 2G (e, e), κ-CN B 1P, κ-CN B 2P, κ-CN 1P 1G (e), κ-CN B 1P 2G (c/d, e).

Fig. 5. Relative abundances and frequency observations for all isoforms of κ-CN within either EL or ML. The primary y-axis shows the relative abundance in percentage of total protein identified (bars with the first and third quartile as whiskers), while the secondary y-axis indicates observations of each specific isoform (dots) for EL and ML as a frequency of all available samples within either EL and ML. Light grey represents EL while dark grey represents ML. An asterisk denotes the level of significance in abundance between lactation groups: *, P ≤ 0.05.

Fig. 6. Relative abundances and observations for all isoforms of κ-CN within the combined groups of phosphorylations and glycosylation. The primary y-axis shows the relative abundance in percentage of total protein identified (bars with the first and third quartile as whiskers), while the secondary y-axis indicates observations of each specific isoform (dots) as a frequency of all available samples within the combined groups of phosphorylation and glycosylation. Light grey, grey and dark grey represents the genotypes BB, AA and AB respectively. Asterisks denote the level of significance in abundance between genotypes: *, P ≤ 0.05; ***, P ≤ 0.001.
k-CN B 1P 2G (e, e) and k-CN 1P 3G (e, e, e). Among these only k-CN A 2P and k-CN B 1P 1G (e) turned out to have significantly different relative abundances between EL and ML ($P_{k-CN\;A\;2P} < 0.05$ and $P_{k-CN\;B\;1P\;1G\;(e)} < 0.05$), thus k-CN A 2P was more abundant in EL while k-CN B 1P 1G (e) were more abundant in ML (Fig. 5).

To explore the presence of phosphorylations and glycosylations across k-CNs genotypes (AA, AB and BB), the former division by two for the homozygotes are reversed to represent actual conditions in milk. In addition, the isoforms were merged into groups representing the different combinations of P and G (Fig. 6). From this it
became clear that the PTM profiles of κ-CN were different between genotypes. Among these groups, only κ-CN 1P, κ-CN 1P 1G, κ-CN 1P 2G, κ-CN 1P 3G and κ-CN 2P had enough observations for statistical analysis. The relative abundances of the total sum of glycosylations across all groupings were strongly affected by genotype (P < 0.001), while the comparison of total κ-CN among genotypes was insignificant. Comparing genotypes for groups of κ-CN with 1G, 2G and 3G, respectively, were all significant (P1G < 0.05, P2G < 0.001 and P3G < 0.01). Combined this indicates that the total abundance of κ-CN is similar between genotypes, the difference between genotypes is to be found in the abundance of the glycosylated κ-CN, among these glycosylated κ-CNs, the 2G group seems to vary the most between genotypes. To explore if these differences in glycosylation were linked to any particular combination of phosphorylation and glycosylation, groupings of P and G content were generated (Fig. 6).

It was found that genotype was of significance for all the glycosylated groups of κ-CN with 1P (P1P 1G < 0.05, P1P 2G < 0.001 and P1P 3G < 0.05). It is thus clear that differences between κ-CN genotypes are to be found in the pool of glycosylated isofoms, specifically those with 1P and 2G. Interestingly, for all glycosylated 1P groups, it seems as if the BB genotype is the most frequent and abundant, followed by AB and AA. The only exception is the abundance of 1P 1G, where the order is BB > AA > AB (Fig. 6).

3.5. Presence and abundance of major whey protein isoforms and effect of lactation stage

For α-LA, a total of 139 observations were assigned. All were representing the B variant, which was found in two isoforms, a major non-phosphorylated isofom, identified in all samples, and a minor singly phosphorylated form, observed in 54.44% of the samples (Fig. 7). Based on relative abundances, α-LA was significantly more abundant in the EL group compared with the ML group (P < 0.01).

A total of 140 observations was assigned to four different isoforms of β-LG. These were assigned to the two genetic variants, β-LG A and B representing three genotype combinations, β-LG AA, β-LG AB and β-LG BB (Fig. 8). Of the 140 observations, some were β-LG with a single lactosylation (+324Da), for β-LG A and B, 25.80% and 3.33% of samples contained lactosylations, respectively. It is interesting to note that the majority of lactosylations were linked to β-LG A. In addition, by comparing the non-lactosylated β-LG genotypes (AA, AB and BB) a significant difference was found (P < 0.001), and thus the genotypes are expressed in different relative quantities (data not shown).

4. Discussion

In this study, isoform heterogeneity of major milk proteins, including some less studied low abundant PTM isoforms and splice and cleavage variants, were investigated in relation to the lactation stages EL (DIM at the beginning of the experiment of 23.3 ± 6.7) and ML (DIM at the beginning of the experiment of 176 ± 15). Using deconvoluted data from the TIC of the LC-ESI/MS, isoforms eluting at similar RTs could be identified, which otherwise would be difficult or impossible to distinguish. In total, this approach allowed us to identify a total of 66 different isoforms with 1675 observations throughout the 90 samples. From the observed RP-HPLC RT profiles (Fig. 1), it is clear that most of the isoforms cluster around the same RT. However, some of the isoforms belonging to κ-CN seem to have a wider distribution than those of the other proteins, which is consistent with the larger PTM heterogeneity of κ-CN. Similar observations were made by Jensen et al. (2015), who suggested that one of the explanations behind this could be that the PTMs in question are attached to different sites on the protein backbone. This could also be an explanation for some of the RT variations observed here. In addition, Fig. 1 demonstrates that a higher ACN concentration is needed to elute isoforms with higher levels of phosphorylation. This is particularly evident for β-CN, β-CN, β-CN and α-LA, and to a lesser degree for κ-CN. The RP-HPLC RT profiles also demonstrate that glycosylated/lactosylated isoforms elute at lower ACN concentrations. This was observed for β-LG and κ-CN. For free glycans, their RT have been correlated with NeuAC content (Dallas et al., 2011); however, for attach glycan, this relation was reported to be less clear (Goonatilleke et al., 2019). The combined effect of glycosylation(s), types thereof, phosphorylation(s), sites of PTMs and genotypes, could explain the deviations observed for κ-CN relative to these general patterns.

4.1. Phosphorylation isoforms of β-CN and κ-CN

From our data, multiple phosphorylation isoforms were observed. This was particularly true for β-CNs, where isoforms with 7 to 10P were observed for β-CN and 10 to 15P for β-CN. Since the analysis was carried out using top-down LC-ESI/MS proteomics, the specific sites of phosphorylation in the protein backbones are not known, and are expected to be within the commonly accepted casein phosphorylation motifs Thr-X-Glu/Asp/Ser/P, with Ser/P referring to an already phosphorylated Ser residue (Bjil et al., 2020; Fang et al., 2016; Mercier, 1981). Mercier (1981) used a database of known phosphorylation sites to investigate known kinases, and it was suggested that the kinase(s) involved in casein phosphorylation would have different preferences for specific motifs. The following order was proposed: Ser-X-Glu/Ser/P > Thr-X-Glu/Ser/P > Ser-X-Asp/Ser/P > Thr-X-Asp/Ser/P. One of the suspected primary kinases behind phosphorylation was isolated from the Golgi apparatus and identified as the FAM20C kinase (Tagliabracci et al., 2012). This kinase was responsible for adding phosphorylations to the Ser-X-Glu/Ser/P motif, but not to the Ser/Thr-X-Asp or Thr-X-Glu/Ser/P motifs. This could indicate that more than one kinase is involved in the complete phosphorylation of milk proteins, and possibly related to different phosphorylated isoforms. This seemingly differential preference for phosphorylation sites could indicate, that different kinases are involved in the generation of the low and high phosphorylation isoforms. In the present study, different frequencies of phosphorylation isoforms of β-CN were observed. These frequencies could be the product of different kinase compositions or activities, and thus that different motifs could be preferred at different lactation stages. This may also explain the differences in abundance observed between specific phosphorylation isoforms within the same protein (Fang et al., 2016, 2017). Mercier (1981) suggested that the phosphorylation sites should be relatively close to or part of a β-turn. This would ensure that the site is present near the surface of the protein, and thus that a kinase could interact with the sequence. A consequence of this could be that a phosphorylation remains on the outside of a protein and is thus available for interactions. Combining this would allow for up to 10P for β-CN, 15P for β-CN and 6P for β-CN (Bjil et al., 2020; Fang et al., 2016; Miranda et al., 2020; Nilsson et al., 2020). In spite of these theoretical possibilities, neither β-CN 16P nor β-CN 6P have so far, to our knowledge, been observed.

For β-CN, multiple masses corresponding to splice products or a plasmic cleavage product have earlier been reported (Galasisi et al., 2000; Martin et al., 2013a,b). Not all these earlier reported isoforms were observed here. The plasmic cleavage product observed in this work corresponded to what would be a loss of the first 79 AA in β-CN. This loss was also observed by Galasisi et al. (2000) and Miranda et al. (2020). As these first 79 AA in β-CN contain all phosphorylation sites, except one, the remaining fragment [β-CN 1P fragment (80–199)] holds only one possible...
phosphorylation. The two observed splice products of z-CN-CN corresponded to a loss of either exon 12 or to a deletion of a Gln residue at position 59 or 78. According to Martin et al. (2013a,b), these splice products originate from two different events. The z-CN splice variant form corresponds to the loss of exon 12 (representing deletion of residues 96–109). The amino acid stretch represented by exon 12 does not contain any potential phosphorylation sites, and therefore this exon skipping does not affect the phosphorylation status. The loss of Gln at either position 59 or 78 is caused by a cryptic splice event. However, simultaneous loss of both these Gln residues was not observed in our data. The deletion of Gln at either position 59 or 78 does not interfere with any phosphorylation motif, and these splice variants are thus found in both the 5P and 1P forms.

Also, for α-CN-CN, considerable variation within phosphorylation forms was observed throughout the dataset. Our database allowed identification of isoforms from 9 to 15P, and the dataset was later searched for 16P as well. This resulted in the observation of α-CN-CN with 10 to 15P. The measured relative abundances of α-CN-CN isoforms observed in this study are comparable with that of Nilsson et al. (2020), where α-CN-CN 9–15P forms were observed. As also observed by Fang et al. (2016), not all possible isoforms were present in all samples. In the present study, the 11 and 12P isoforms were both the most frequent and the most abundant of the α-CN-CN isoforms. According to Mercier (1981), a total of 18P should in theory be possible for α-CN-CN; however, as they also point out, some sites and use of the motifs have never been observed, and thus the actual maximum is probably closer to 15P, as described by Farrell et al. (2004).

For β-CN, 4 and 5P isoforms were observed. It was shown, as expected, that the 4P isoform was low abundant, compared with the 5P counterpart. According to Mercier (1981), a total of 6P is possible, as at least 6 matches to the Ser/Thr-X-Glu/Asp/SerP motif are in theory possible. However, one of these theoretical matches is to the Thr-X-Asp motif, which has not been confirmed as an actual used site of phosphorylation. Based on the casein motif (Ser/Thr-X-GluAsp/SerP), and the preference of a nearby β-turn, the most logical sequence therefore seems to be that the cluster of Ser residues at positions 15, 17, 18 and 19 are phosphorylated in conjunction, while the more isolated Ser at position 35 is phosphorylated on its own (Mercier, 1981). Considering that Ser at position 35 is relatively close to a β-turn, this only highlights this as an option for the fifth and final phosphorylation site for β-CN.

4.2. κ-CN isoforms

As expected, a large number of κ-CN isoforms was observed (total of 39, Fig. 5). The majority of observed κ-CN isoforms were glycosylated (34 of 39). Of these, most contained NeuAc (30 of 39) as part of their glycosylation, as observed earlier (Holland, Deeth, & Alewood, 2004, 2006; Miranda et al., 2020; Nilsson et al., 2020; Poulsen et al., 2016). In Nilsson et al. (2020), the precise number of identified isoforms was not reported, therefore a direct comparison with the present study is not possible. From Nilsson et al. (2020), it is interesting to note, that up to 11 different isoforms of κ-CN could be identified in one single milk sample. In the present study, on average the samples contained 5.86 different κ-CN isoforms, with a maximum of 11. The sample with 11 identified isoforms represented an AB milk, where three of the isoforms were assigned to variant A, and the remaining eight were assigned to variant B. The genotype representing milk with the highest number of isoforms from Nilsson et al. (2020) was not given. Notably, Miranda et al. (2020) accepted masses matching glycoforms of κ-CN eluting later than non-glycosylated κ-CN variants. In the present study, where a gradient from low to high hydrophobicity was used (see Section 2.2), such matches were discarded because it was considered unlikely that addition of a glycosylation would increase the overall hydrophobicity, as demonstrated in Fig. 1. In Holland, Deeth, and Alewood (2006), 2D gels were used to separate milk proteins according to their masses and pI values with subsequent MS/MS analysis of spot digests. However, comparing results of Holland et al. (2006) with those in the present study, surprisingly few glycosylated κ-CN isoforms without NeuAc were found here.

From Fig. 1, it is evident that some κ-CN masses appeared at multiple RTs. Similar findings were also observed by Jensen et al. (2015) in a study of κ-CN macroprotein. As suggested by Jensen et al. (2015), this could be due to different attachment sites on the protein backbone for the PTMs in question. In the work of Holland et al. (2006), a total of 17 different isoforms was described for κ-CN B alone, some with up to six attached glycans, and it is thus possible, that the heterogeneity of κ-CN, especially for κ-CN B, could be much greater than indicated in the present study, where a maximum of three attached glycans was allowed. In addition, Holland et al. (2006) proposed, that the fourth glycosylated amino acid could be at position 145, being a Thr residue, whereas no suggestions for the fifth and sixth glycosylation was made. As LC-ESI/MS was used here, it would be of interest to study differences between microbial milks to further validate these positions, as suggested by both Holland et al. (2004) and Jensen et al. (2015).

The average GD (defined as total glycosylated κ-CN relative to total κ-CN) was calculated here, and found to be 10.85% (result not shown), which is lower than most GD values reported in other studies (Bijl et al., 2020; Poulsen et al., 2016). However, these studies were using UV or gels and not LC-ESI/MS to determine the GD, and therefore the reported difference may relate to methodological differences. The value obtained here complies well with that of Nilsson et al. (2020), who also used LC-ESI/MS as well as a to our study comparable database for matching theoretical and observed masses. A calculation based on their reported results gives a GD of 9.90%. However, in future studies, it should be considered to allow for more potential glycosylations per κ-CN molecule, as suggested by Holland et al. (2006), even though some of these expectedly may turn out to be of very low abundance or even undetectable.

From the κ-CN data it was further observed that the κ-CN isoforms differed in relative abundances between genotypes. As also observed by Poulsen et al. (2016), in the present study, different total relative abundances of κ-CN isoforms present in κ-CN AA, κ-CN AB and κ-CN BB milks were observed (Fig. 6). In our dataset the total relative abundance of κ-CN BB > κ-CN AB > κ-CN AA, which is the same order as reported earlier in milk from Danish Holstein by Poulsen et al. (2016). Even though GDs from UV-based studies and MS based studies are not directly comparable, the broad pattern of GD reported here reflected the expected order of κ-CN BB > κ-CN AB > κ-CN AA (Poulsen et al., 2016).

4.3. Whey proteins

For the whey proteins, the number of isoforms found was relatively low; though, notably, a 1P form of α-LA was found. The existence of such a 1P isoform of α-LA was reported recently (Miranda et al., 2020); though, to the best of our knowledge, the site of phosphorylation has not been reported. Considering the Ser/Thr-X-GluAsp/SerP phosphorylation motif suggested by Mercier (1981), two positions seem to be possible, at positions 47 or 75. According to Farrell et al. (2004) none of these positions are altered between known genotypes. Given that residue 47 matches the Ser in the Ser-X-Glu motif, whereas the residue at position 75 matches the Ser in the Ser-X-Asp motif, the residue at 47 would match the most frequently used motif.
For β-LG, it was noted, that lactosylation was more frequent on β-LG variant A compared with β-LG variant B (25.80% versus 3.33% of samples, see Fig. 8). Without MS/MS data, it can only be speculated what drives this. One explanation may be related to different chemical properties of Asp and Gly, which are present at position 64 of β-LG variant A and B, respectively (Farrell et al., 2004). In contrast to Gly, Asp is negatively charged, which is expected to affect the structural properties of the protein. Since this position is not far from known lactosylation sites at residues 60 and 69 (Le et al., 2017), it could be speculated that these structural differences between the A and B variants could affect at least one of these positions, thus making variant A more accessible for lactosylation. The presence of lactosylations in fresh milk, which was used in the present study, is somewhat surprising, as this is normally a result of processing (Le et al., 2017), but shows that this can be observed even in cold-stored milk, as also reported by Miranda et al. (2020).

4.4. Effect of stage of lactation on PTM patterns

The physiological process of lactation has been reported to potentially affect PTM decoration of the milk proteins. Phosphorylation degree and/or phosphorylation isoforms of the S2-CN have been reported to be potentially affected by lactation stage (Fang et al., 2017; Poulsen et al., 2016). In earlier studies, both GD of CN, as well as total content of NeuAc in its glycans, were found to increase with advance of lactation (Bonfatti et al., 2014; Maciel et al., 2017; Poulsen et al., 2016; Robitaille, Ng-Kwai-Hang, & Monardes, 1991). In the present study, both S2-CN and κ-CN, including their many isoforms, were significantly affected by lactation stages. This indicates, that the level or presence of kinases and glycosyltransferases in the endoplasmatic reticulum (ER) and Golgi apparatus (GA) can be affected by lactation. Here, the S1-CN–10P was found only in ML at quite low frequency (2.22%). Considering the motif from Mercier (1981) (Ser/Thr-X-Glu/Asp/SerP), 9 matches can be found in the S2-CN amino acid sequence corresponding to the Ser/Thr-X-Glu/SerP motif, the 10th site belonging to a Thr-X-Asp motif at residue 49, which has never been observed before. Considering the rarity of S2-CN–10P and that it was observed only in the ML group, this could indicate that the composition of kinases in the ER and/or GA could be affected by stage of lactation. For β-CN, 4P was found in ML only, together these findings could indicate that the composition of kinases of the ER and GA are indeed affected by lactation.

5. Conclusion

In the present study, it was shown that the use of deconvoluted TIC data from an LC-ESI/MS system, combined with a database of theoretical masses, was a powerful tool for identifying and quantifying isoforms of bovine milk. By this approach, it was possible to identify and quantify PTM differences relative to lactation stage. Some of these changes were found in the phosphorylation stages of S2-CN, ranging from 7 to 10P for S2-CN, and 10 to 15P for S2-CN. In addition, some low abundant splice products for S2-CN were identified and quantified, alongside an otherwise normal S2-CN profile. For κ-CN, 39 different isoforms were identified and quantified with different P and G contents; the most common glycosylated isoform being κ-CN 1P 2G (e.e.). Some of the κ-CN isoforms were observed to cluster around different RT. For β-CN, the rare 4P PTM form was observed, but only in ML. Taken together, the presence of S2-CN–10P and β-CN–4P only in milk from cows in mid lactation could indicate that the composition of kinases in the ER and/or GA changes during lactation. The differences in RT of some of the glycosylated isoforms of κ-CN could further indicate that the glycans are attached to different locations. This may reflect that the glycosyltranferases found in the ER and/or GA could be affected by lactation as well. These findings would require further studies to enlighten possible mechanisms behind these observations. For this purpose, an MS/MS approach could be an approach to identify the locations of unknown phosphorylations, glycosylations and to investigating the presence of different kinases in bovine milk.

Declaration of competing interest

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.

Acknowledgements

The present study was part of the project “Indicators in milk for nitrogen and phosphorus excretion” financed by the Danish Milk Levy Fund (Aarhus N, Denmark). From the Department of Food Science, Aarhus University, Denmark, the authors would like to thank Gitte Hald Kristiansen for excellent technical assistance, Søren Drud–Heydary Nielsen for initial discussions about the MS database and Gayani Madushani Sirinayake Lokuje for excellent help and discussions regarding the statistical analysis and how to apply it.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idairyj.2021.105262.

References


